



...is



suitable enough for biostatisticians

involved in Clinical Research & Evidence-Based Medicine?

...and can it replace SAS in this area?





Welcome to my journey through the world of R!

- ✓ Become familiar with the *lingua franca* of statistics
- ✓ Discover applications of R in Evidence-Based Medicine
- ✓ Rock, squeeze and explore your data deeply - for free
- ✓ Find 13 reasons why you will love R! [cannot wait? jump now!](#)
- ✓ Enhance your skills...

...and start using R today!



DISCLAIMER

All trademarks, logos of companies and names of products

used in this document

are the sole property of their respective owners

and are included here

for informational, illustrative purposes only *nominative fair use*



This presentation is based exclusively on information

publicly available on the Internet under provided hyperlinks.



If you think I violate your rights, please email me: r.clin.res@gmail.com



- [Brief introduction](#)
- [The R family](#)
- [7 cool facts about R](#)
- [List of R users](#)
- [Is R a popular software?](#)
- [Demonstrative screenshots](#)
- [13 reasons why you will love GNU R](#)
- [Further important issues](#)
- [Conclusion](#)



Agenda: 13 reasons why you will love GNU R

- I [R is \(extremely\) cheap. In fact - it's free :\)](#)
- II [R has \(extremely\) wide range of capabilities](#)
- II 1/3 :) [R is easy to maintain!](#)
- II 2/3 :) [R is not resource consuming!](#)
- III [R is supported by the world of science](#)
- IV [R is supported by the community](#)
- IV 1/2 :) [Books](#)
- V [R is supported by the business](#)
- V 1/2 :) [R and SAS](#)
- VI [R is able to read data in many formats](#)
- VI 1/3 :) [R and relational databases](#)
- VI 2/3 :) [Advanced data manipulation](#)
- VII [Interoperability is easy to achieve](#)
- VIII [R is truly cross-platform](#)
- IX [R offers numerous ways of presenting data](#)
- IX 1/2 :) [Graphics – all have waited for this moment :\)](#)
- X [There are many options to optimize the code](#)
- XI [R is able to handle large amount of data](#)
- XII [R has a set of fancy tools and IDEs](#)
- XIII [FDA: R is OK for drug trials! - it should start from this slide :\)](#)



Agenda: Further impoRtant issues

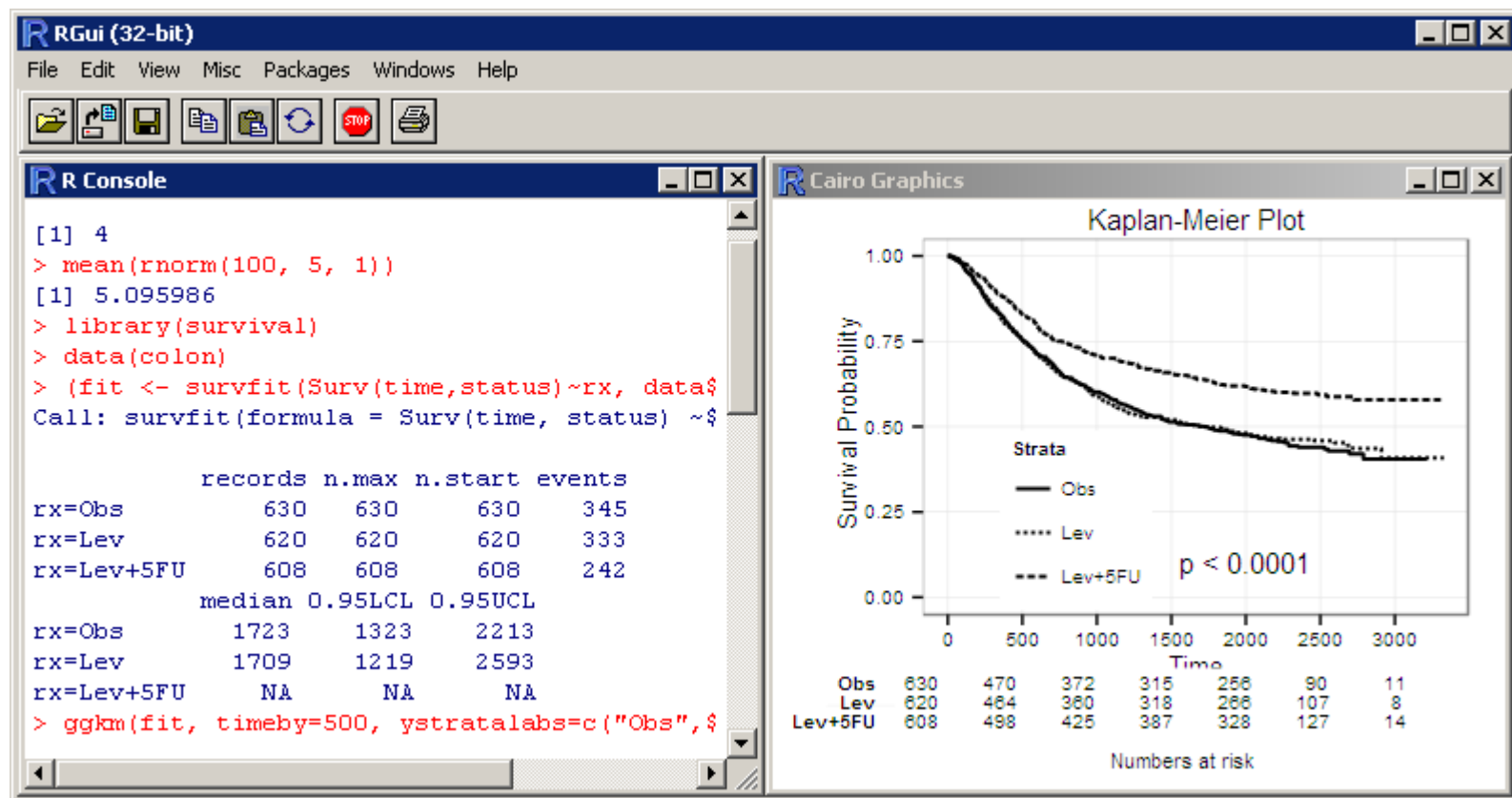
- I [Handling metadata](#)
- II [Handling ODM and CDA metadata](#)
- III [Issue with multilingual data](#)
- IV [Differences between SAS and R](#)
- V [Implementation of useful SAS functions](#)



IntRoduction

In simply words, R is a free software environment for statistical computing, data manipulation and charting, widely used in the scientific world. It is also the name of a high-level, interpreted programming language.

R (the successor of S) was created in 1992 by [Ross Ihaka](#) and [Robert Gentleman](#) at University of Auckland. Now it is developed by the [R Development Core Team](#).





is it a popular softwa**Re** ?

[...] R is also the name of a popular programming language used by a growing number of data analysts inside corporations and academia. It is becoming their *lingua franca* partly because data mining has entered a golden age, whether being used to set ad prices, find new drugs more quickly or fine-tune financial models.

Companies as diverse as Google, Pfizer, Merck, Bank of America, the InterContinental Hotels Group and Shell use it.

Data Analysts Captivated by R's Power (The New York Times)

<http://tinyurl.com/captivated-by-r>





Short characteristics:

- **Description:** computational environment + programming language
- **Developer:** R Development Core Team
- **Operating systems:** cross-platform: Windows, Unix, Linux, Linux-based “mobile” OS: (Android, Maemo, Raspbian), Mac OS X
- **Form:** command line + third-party IDEs and editors RStudio, RCommander, etc.
- **Infrastructure:** R core library + shell + libraries (base and third-party)
- **Model of work:** 1) standalone application, 2) standalone server, 3) server process
- **Programming lang.:** interpreted, high-level with dynamic typing; debugger onboard
- **Paradigm:** 1) array, 2) object-oriented, 3) imperative, 4) functional, 5) procedural, 6) reflective
- **Source of libraries:** central, mirrored repository – CRAN; users' private repositories, third-party repositories (Github, Rforge), other sources
- **License of the core:** GNU General Public License ver. 2
- **License of libraries:** 99.9% open-source, rest – licensed (free for non-commercial use)



The R Project for Statistical Computing

[\[Home\]](#)

Download

[CRAN](#)

R Project

[About R](#)

[Contributors](#)

[What's New?](#)

[Mailing Lists](#)

[Bug Tracking](#)

[Conferences](#)

[Search](#)

R Foundation

[Foundation](#)

[Board](#)

[Members](#)

[Donors](#)

[Donate](#)

Documentation

[Manuals](#)

[FAQs](#)

[The R Journal](#)

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News

- **R 3.2.1 (World-Famous Astronaut) prerelease versions** will appear starting June 8. Final release is scheduled for 2015-06-18.
- **R version 3.2.0** (Full of Ingredients) has been released on 2015-04-16.
- **R version 3.1.3** (Smooth Sidewalk) has been released on 2015-03-09.
- **The R Journal Volume 6/2** is available.
- **useR! 2015**, will take place at the University of Aalborg, Denmark, June 30 - July 3, 2015.
- **useR! 2014**, took place at the University of California, Los Angeles, USA June 30 - July 3, 2014.



In 1998, S became the first statistical system to receive the Software System Award, the top software award from the ACM



AT&T/ Bell Laboratories

Date of foundation: 1976, 1998
John Chambers, Rick Becker, Allan Wilks

<http://ect.bell-labs.com/sl/S>

Spotfire S+
formerly **S-PLUS**

TIBCO Software Inc.

Date of foundation: 2010
License: Commercial

<http://spotfire.tibco.com>
[What's new in version 8.1](#)

GNU R

University of Auckland

Date of foundation: 1993
Robert Gentleman, Ross Ihaka
License: GNU GPL v2

<http://www.r-project.org>

REVOlution

(bought by Microsoft in 2015)
Commercial + Free (GPL v2)

<http://www.revolutionanalytics.com>

RStudio

Commercial
+ Free (AGPL v3)

<http://www.rstudio.com>

Oracle R

Commercial + Free

<http://www.oracle.com...>



Seven quick (and cool) facts about R April 2014

- (1) R is the **highest paid** IT skill ¹
- (2) R is **most-used** data science language **after SQL** ²
- (3) R is used by **70%** of data miners ³
- (4) R is **#15** of all programming languages ⁴
- (5) R is **growing faster** than any other data science language ⁵
- (6) R is **the #1** Google Search for Advanced Analytic software ⁶
- (7) R has **more than 2 million** users worldwide ⁷

1. [Dice.com survey, January 2014](#)

2. [O'Reilly survey, January 2014](#)

3. [Rexer survey, October 2013](#)

4. [RedMonk language rankings, January 2014](#)

5. [KDNuggets survey, August 2013](#)

6. [Google Trends, March 2014](#)

7. [Oracle estimate, February 2012](#)



List of R users



They use(d) R



European Environment Agency



<http://www.revolutionanalytics.com/companies-using-r>

<http://www.rstudio.com/>



A word about the list of UseRs

The list is built based **exclusively on publicly available** information:

- lists of users provided by [Revolution](#), [RStudio](#) and others
- articles ([example](#), [example](#)) and interviews ([example](#))
- published documents in which a name of a company is visible ([example](#))
- job advertisements
- names of companies supporting / organizing events (conferences, courses)

That is to say, a logo of a company is included in the list only if there is a strong evidence that the company uses or supports (or used or supported) R, based on information shared on the Internet – and thus available for everyone.

Please note, that I am not aware if all listed companies are still using any version of R at the time the presentation is being viewed.



"We use R for adaptive designs frequently because it's the fastest tool to explore designs that interest us. Off-the-shelf software, gives you off-the-shelf options. Those are a good first order approximation, but if you really want to nail down a design, R is going to be the fastest way to do that."

Keaven Anderson

Executive Director Late Stage Biostatistics,
Merck

<http://www.revolutionanalytics.com/content/merck-optimizes-clinical-drug-development...>





They use **R**

Using R For Flexible Modell... x +

www.r-project.org/nosvn/conferences/useR-2009/slides/Harbron.pdf

Search

Page: 1 of 23

Automatic Zoom

Using R For Flexible Modelling Of Pre-Clinical Combination Studies

Chris Harbron

Discovery Statistics

AstraZeneca

AstraZeneca 
life inspiring ideas

They use R



Revolution R at Pfizer | Rev... x +

www.revolutionanalytics.com/content/revolution-r-pfizer


Get Technical Support

REVOLUTION ANALYTICS

Community Applications Products AdviseR Resources Company

Home » Resources » Case Study

Revolution R at Pfizer



Downloads: [Download the case study in pdf.](#)

Background

"De facto, R is already a significant component of Pfizer core technology. Access to a supported version of R will allow us to keep pace with the growing use of R in the organization, and provides a path forward to use of R in regulated applications."

James A. Rogers Ph.D., Associate Director, Nonclinical Statistics Group, Pfizer Global Research and Development

R is an implementation of the S language which, "forever altered how people analyze visualize, and manipulate data" (excerpt from the citation accompanying the Association for Computing Machinery



They use R

The screenshot shows a web browser window with the following elements:

- Browser Tab:** "CardioDX Uses Revolution ... x +"
- Address Bar:** "www.revolutionanalytics.com/news-events/cardioidx-uses-revolution-analytics-develop-first-n" with a search icon and a search input field containing "Search".
- Navigation Icons:** Star, list, download, home, ABP, and chat icons.
- Header:** "REVOLUTION ANALYTICS" logo on the left, "Get Technical Support" with a search icon on the right, and a menu of links: "Community", "Applications", "Products", "AdviseR", "Resources", "Company".
- Breadcrumbs:** "Home"
- Section Header:** "CardioDX Uses Revolution Analytics to Develop First Non-Intrusive Test for Predicting Coronary Artery Disease"
- Sub-Header:** "Scientists Use Revolution R to Develop Corus CAD(r) Test, Recognized as One of 'Top Ten Medical Breakthroughs of 2010' by TIME Magazine"
- Date:** "January 18th, 2011"
- Text:** "Revolution Analytics, the leading provider of commercial software and support for the popular open source R statistics language, today announced the successful implementation of its signature product, Revolution R Enterprise, by biostatisticians at CardioDX, a genomic research firm. Revolution R was used to design the Corus CAD(r) test, the first test of its kind to analyze genomic data and identify at-risk patients for coronary artery disease (CAD) without requiring an invasive procedure. The Corus CAD(r) test was recently honored as one of TIME Magazine's 'Top Ten Medical Breakthroughs of 2010'."
- Text:** "Revolution R was used to help develop the algorithm that powers the Corus CAD(r) test. To do so, biostatisticians at CardioDX analyzed over 10 gigabytes of genomic data to ultimately identify 23 predictive genes for CAD. In addition to using open source R, biostatisticians at CardioDX employ..."



- Project Management
- Quality Assurance
- Regulatory Affairs
- Medical Affairs
- Clinical Monitoring
- Site set up & Contracting
- Biometrics**

→ BIOMETRICS

KCR's Biometrics (BMX) department includes **Biostatistics (BS)** and **Data Management (DM)** and is designed to provide a full data analysing and cleaning service. A team of Biostatisticians, Statistical Programmers, Clinical Data Managers and Clinical Data Associates will be appointed to make sure if all protocol requirements and industry guidelines are followed and the timelines set are met.

potential issues. Analyses are performed with the use of the most recognizable and acclaimed statistical packages – SAS and GNU R. In special cases the use of an additional and specialized software may be considered.

Comprehensive statistical analysis – including interim evaluations – to the final report, our experienced Biostatisticians and Statistical Programmers maximize efforts to provide the highest quality product in full compliance with legal regulations and industry guidelines. In order to achieve this objective, KCR's Biostatisticians are constantly enhancing their skills. They cooperate seamlessly with trial monitors and data managers to provide scientific support during the entire analysis process, including formulation of hypotheses and delivery of reports on

potential issues. Analyses are performed with the use of the most recognizable and acclaimed statistical packages – SAS and GNU R. In special cases the use of an additional and specialized software may be considered.

Data Management services are built around a solid and proven EDC technology designed to maximize transparency of the study progress across different stages of a trial. The technology contains multiple modules (e.g. IWRS, Central Lab data loading tool, Encoding Tool, CTMS) included in one system to minimize costs and



They use R

R Core Team (2013). — Eur... x +

www.eea.europa.eu/data-and-maps/indicators/oxygen-consuming-substances-in-rive

Search [star] [list] [down] [home] [ABP] [smiley] [hamburger]

Networks Subscriptions Mobile Contact us EEA homepage in your language Log in



Search Europe's environment... Search
Advanced search A-Z Glossary

Topics Data and maps Indicators Publications Media About EEA The EEA is an agency of the European Union [EU flag]

You are here: Home / Data and maps / Indicators / Oxygen consuming substances in rivers / R Core Team (2013).

R Core Team (2013).

Methodology Reference

R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Title:
R Core Team (2013).

Description:
R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

URL:
http://www.R-project.org

Geographic coverage

Share with others



Data and maps

Menu

Global search

Datasets


In many discussions some praise SAS, others STATA or SPSS.
But most of them refer to **R** with respect.

Firefox

R or SAS: which one is the best statistical... +

www.researchgate.net/post/R_or_SAS_which_one_is_the_best_statistical_software_used_in_medical... Google

ResearchGate Already a member? Log in


 **Deepak k s** · Amrita Institute of Medical Sciences

Question

R or SAS: which one is the best statistical software used in medical field?


I am on the starting stage to study a software. I know spss only. I want to study a software either R or SAS. Some of them say R is better and others SAS. I am totally confused. Which one is best and easily understanding and helpful in medical field?

TOPICS

 **Juan P Steibel** · Michigan State University

I could not imagine working without one of those two. Both are good and any applied biostatistician should have access to both. That is my experience and opinion.


Nov 4, 2013

 **Jason Leung** · The Chinese University of Hong Kong

I used SAS for many years and find it very powerful and well developed. R will be more and more popular and I think in the near future, R will be the top one.


R is not only free, it is more powerful than SAS, SPSS and STATA
<http://www.analyticbridge.com/group/productreviews2/forum/topics/product-reviews-comparing-r-matlab-sas-stata-spss>

Mar 25, 2013

 **Alain-Claude Roudot** · Université de Bretagne Occidentale

The only advantage of SAS over R is its hotline if problems occur. With R you only have the R community, but it can be as powerful as the SAS hotline! so, if you are rich take SAS, in any other case R.

May 13, 2013

 **Jeffrey E. Jarrett** · University of Rhode Island

This is definitely a dumb controversy.

Apr 25, 2014

ResearchGate is the professional network for scientists and researchers. ⌵



Is R a popular software?

R is near the top 10 most popular languages TIOBE, 2015



TIOBE Software: Tiobe Index x + <http://www.tiobe.com>

www.tiobe.com/index.php/content/paperinfo/tpci/index.html Search

Jun 2015	Jun 2014	Change	Programming Language	Ratings	Change
1	2	▲	Java	17.822%	+1.71%
2	1	▼	C	16.788%	+0.60%
3	4	▲	C++	7.756%	+1.33%
4	5	▲	C#	5.056%	+1.11%
5	3	▼	Objective-C	4.339%	-6.60%
6	8	▲	Python	3.999%	+1.29%
7	10	▲	Visual Basic .NET	3.168%	+1.25%
8	7	▼	PHP	2.868%	+0.02%
9	9		JavaScript	2.295%	+0.30%
10	17	▲▲	Delphi/Object Pascal	1.869%	+1.04%
11	-	▲▲	Visual Basic	1.839%	+1.84%
12	12		Perl	1.759%	+0.28%
13	23	▲▲	R	1.524%	+0.85%
14	-	▲▲	Swift	1.440%	+1.44%
15	19	▲▲	MATLAB	1.436%	+0.66%
16	13	▼	Ruby	1.359%	-0.03%
17	26	▲▲	PL/SQL	1.229%	+0.74%
18	31	▲▲	COBOL	0.948%	+0.54%



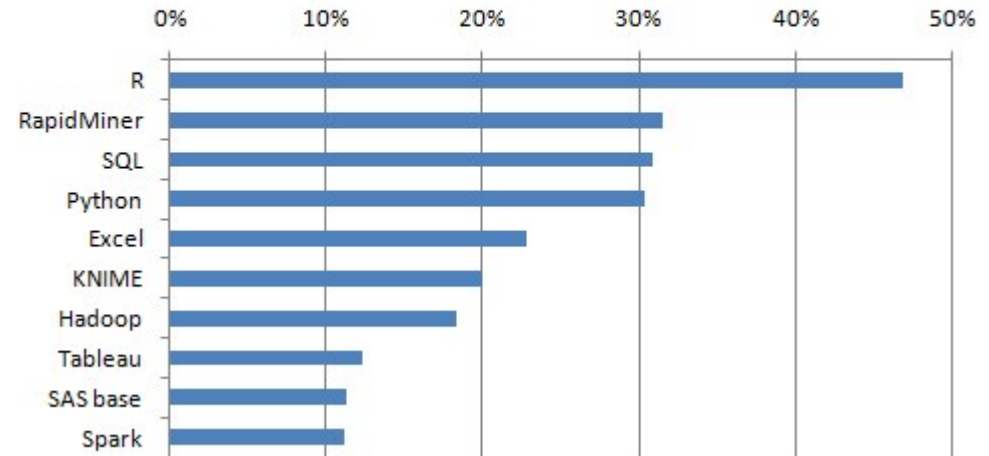
The 16th annual KDNuggets Software Poll²⁰¹⁵

What Analytics, Big Data, Data mining, Data Science software you used in the past 12 months for a real project? [2759 voters]

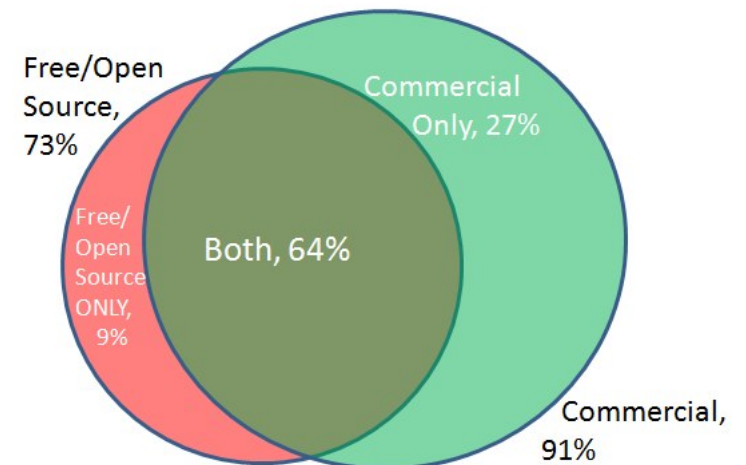
Legend:	
Red: Free/Open Source tools	% users in 2015
Green: Commercial tools	% users in 2014
Fuchsia: Hadoop/Big Data tools	% users in 2013

R (1293), 3.6% alone	46.9%	38.5%	37.4%
RapidMiner (870), 13.7% alone	31.5%	44.2%	39.2%
SQL (853), 0% alone	30.9%	25.3%	na
Python (837), 0% alone	30.3%	19.5%	13.3%
Excel (631), 0% alone	22.9%	25.8%	28.0%
KNIME (553), 6.7% alone	20%	15.0%	5.9%
Hadoop (507), 0% alone	18.4%	12.7%	9.3%
Tableau (341), 0% alone	12.4%	9.1%	6.3%
SAS base (313), 0.6% alone	11.3%	10.9%	10.7%
Spark (311), 0% alone	11.3%	2.6%	na
Weka (310), 0% alone	11.2%	17.0%	14.3%
SAS Enterprise Miner (302), 3.6% alone	10.9%	7.2%	5.9%

Top Analytics, Data Mining, Data Science software used, 2015



Analytics, Data Mining, Data Science Software Usage, 2015



<http://www.kdnuggets.com/2015/05/poll...>



The 15th annual KDNuggets Software Poll²⁰¹⁴

What Analytics, Big Data, Data mining, Data Science software you used in the past 12 months for a real project? [3285 voters]

Legend: Red: Free/Open Source tools Green: Commercial tools Fuchsia: Hadoop-related tools	% users in 2014 % users in 2013 na - not included in poll in 2013.
RapidMiner (1453), 35.1% alone	44.2% 39.2%
R (1264), 2.1% alone	38.5% 37.4%
Excel (847), 0.1% alone	25.8% 28.0%
SQL (832), 0.1% alone	25.3% na
Python (639), 0.9% alone	19.5% 13.3%
Weka (558), 0.4% alone	17.0% 14.3%
KNIME (492), 10.6% alone	15.0% 5.9%
Hadoop (416), 0% alone	12.7% 9.3%
SAS base (357), 0% alone	10.9% 10.7%
Microsoft SQL Server (344), 0% alone	10.5% 7.0%
Revolution Analytics R (300), 13.3% alone	9.1% 4.5%
Tableau (298), 1.3% alone	9.1% 6.3%
MATLAB (277), 0% alone	8.4% 9.9%
IBM SPSS Statistics (253), 0.4% alone	7.7% 8.7%
SAS Enterprise Miner (235), 1.3% alone	7.2% 5.9%
SAP (including BusinessObjects/Sybase/Hana) (225), 0% alone	6.8% 1.4%



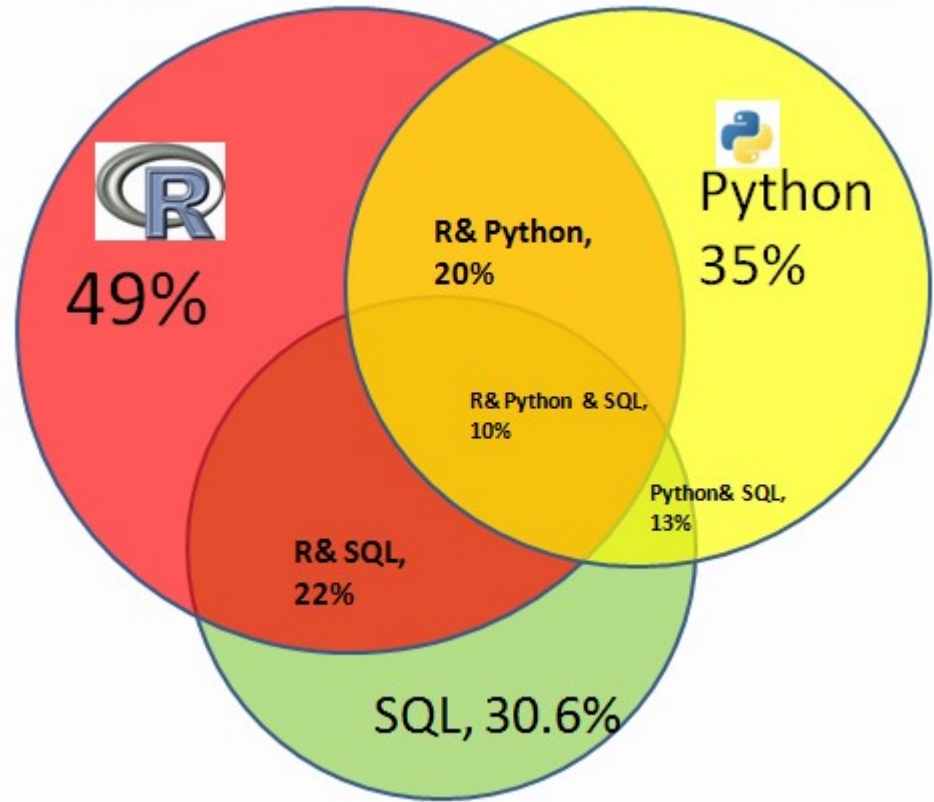
Region, Avg N tools	% users who only use commercial tools % users who only use free tools % users who use both
US/Canada (39%), 3.8	33% 17% 49%
Europe (36%), 3.5	14% 38% 47%
Asia (12%), 4.2	15% 31% 53%
Latin America (6.1%), 3.8	8.5% 39% 53%
Africa/MidEast (3.2%), 3.6	12% 40% 48%
Australia/NZ (3.0%), 4.2	23% 19% 58%

<http://www.kdnuggets.com/2014/08...>

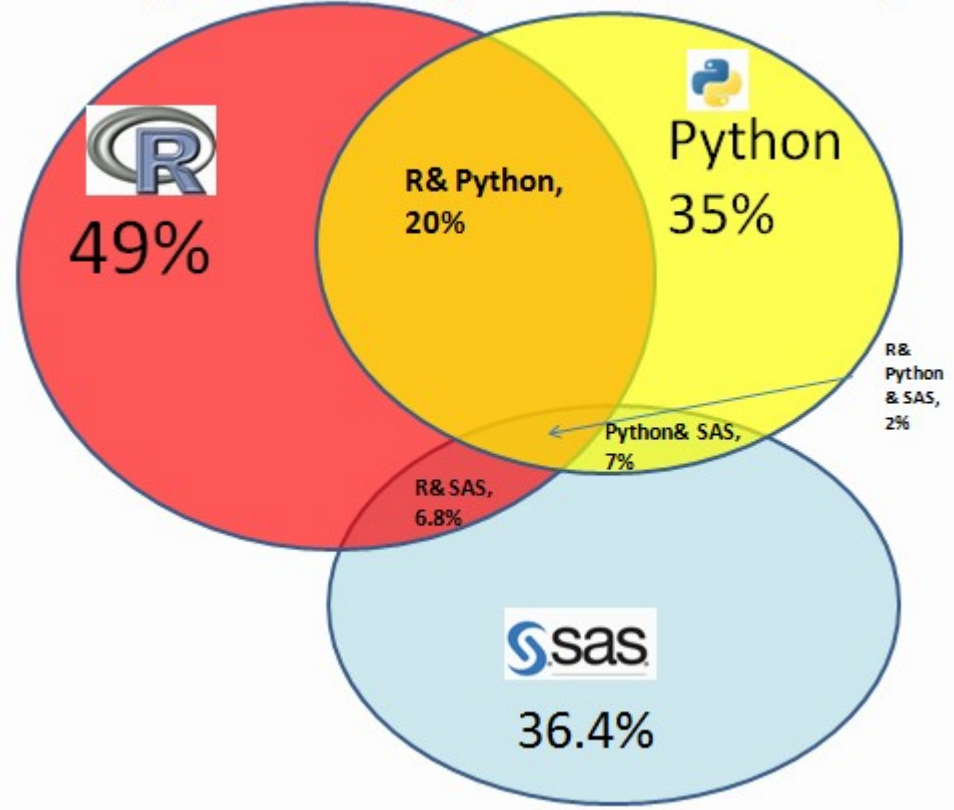


KDnuggets²⁰¹⁴: R vs. Python vs. SAS vs. SQL

KDnuggets 2014 Poll: Languages used for Analytics/Data Mining



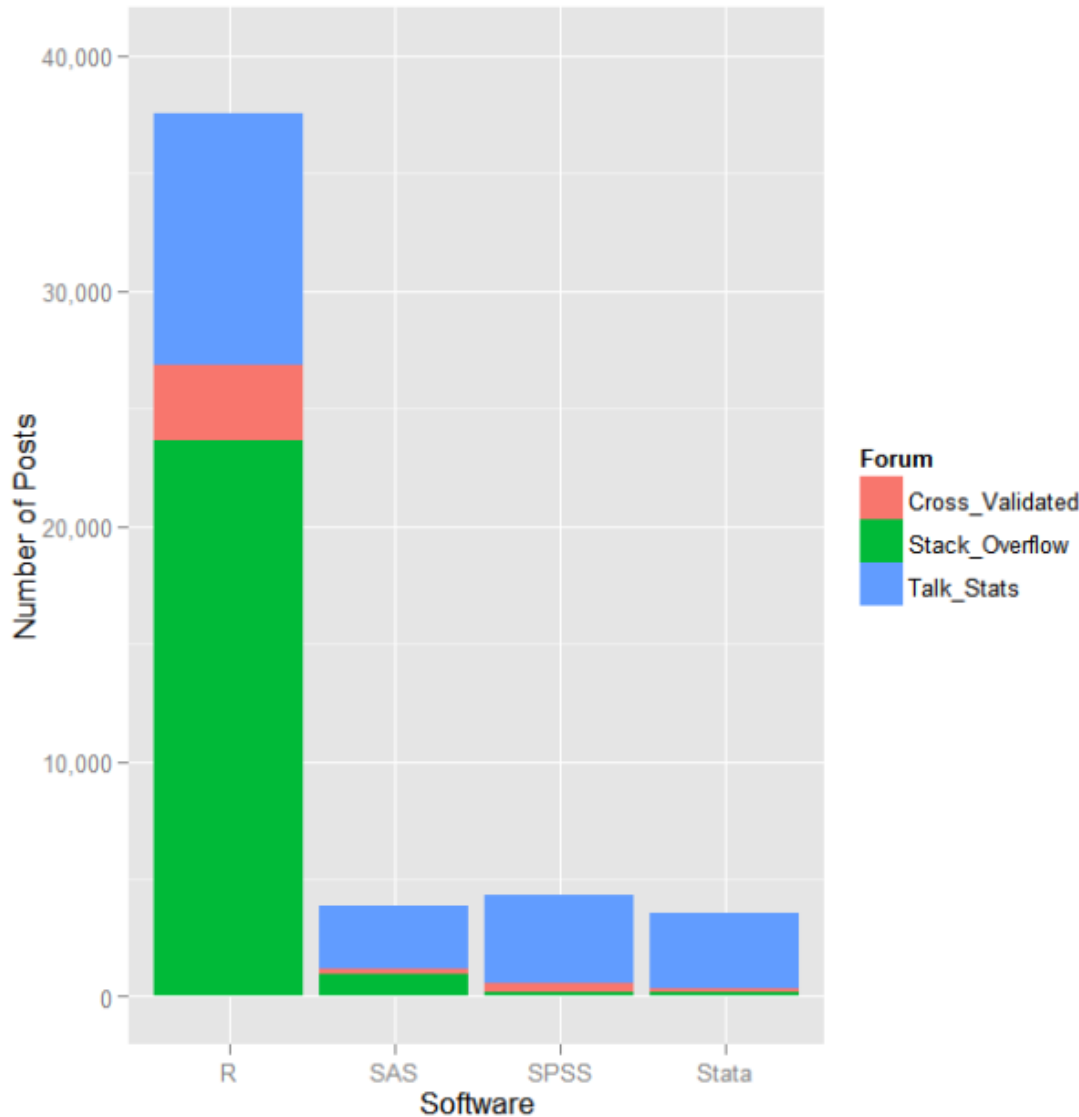
KDnuggets 2014 Poll: Languages used for Analytics/Data Mining, 2



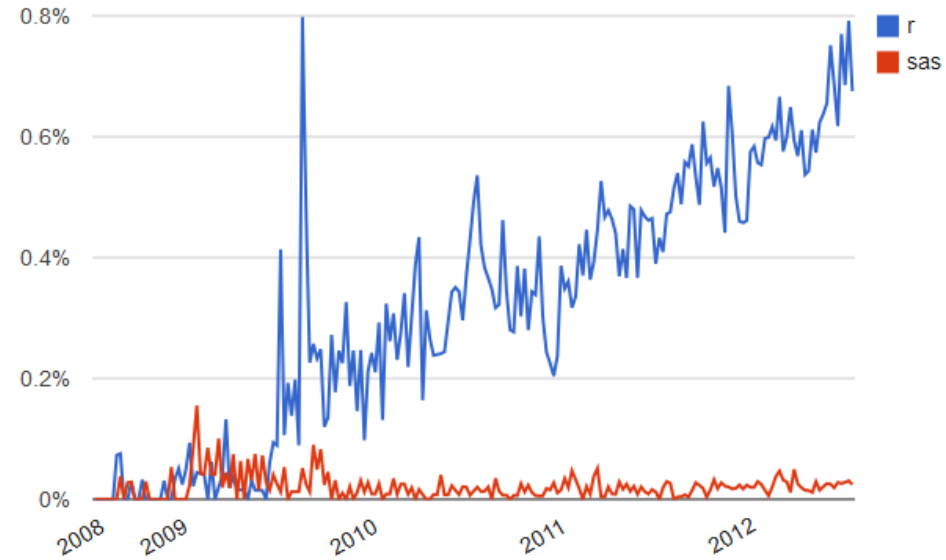


R quickly gains **high position** in statistics 2011-2013

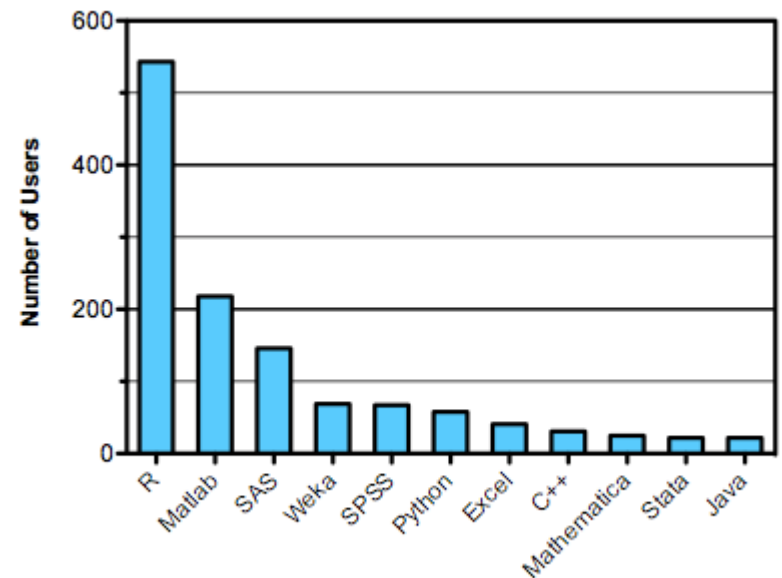
<http://r4stats.com/articles/popularity/>



Number of posts per software on each forum on 2/10/2013



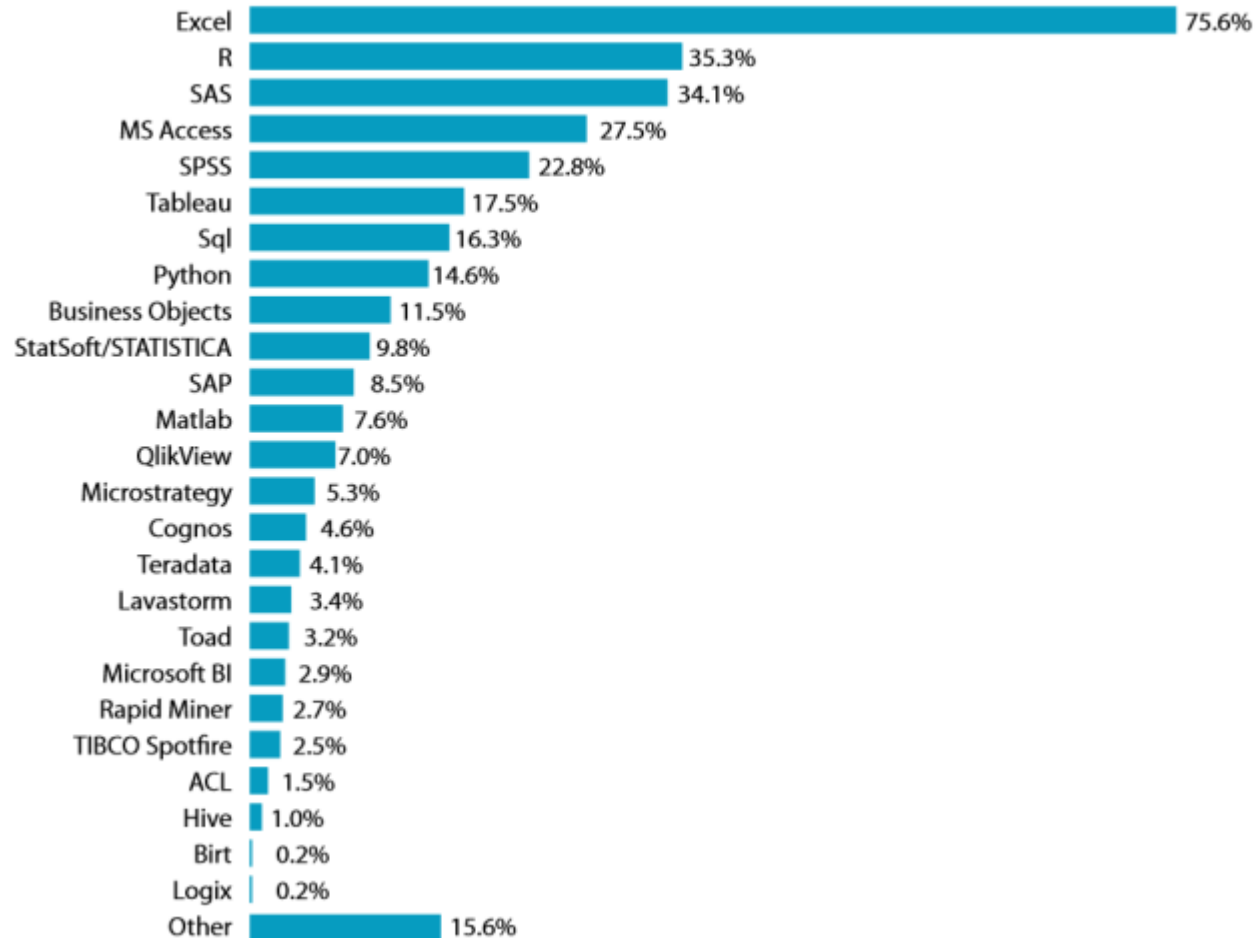
Number of R- or SAS-related posts to Stack Overflow by week



Software used in data analysis competitions in 2011.

R quickly gains **high position** in statistics²⁰¹³

What self-service analytic tool are you currently using?

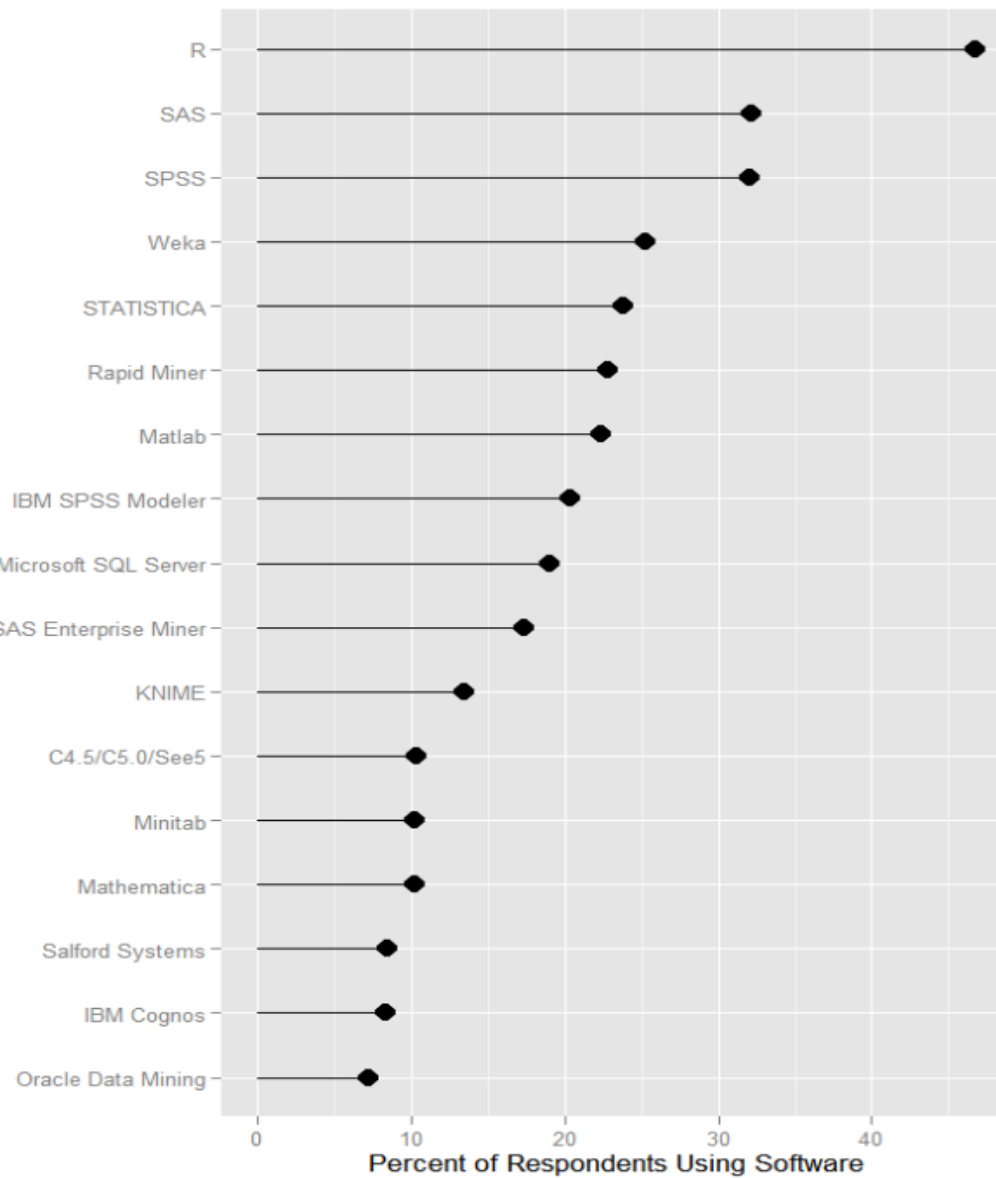


<http://r4stats.com/articles/popularity/>



Rexer Analytics Survey²⁰¹⁰

What Data mining/analytic tools did you use in 2010?



What programming languages you used for data mining / data analysis in the past 12 months? [570 voters]

R (257)	45%
SQL (184)	32%
Python (140)	25%
Java (139)	24%
SAS (121)	21%
MATLAB (83)	15%
C/C++ (73)	13%
Unix shell/awk/gawk/sed (59)	10%
Perl (45)	7.9%
Hadoop/Pig/Hive (35)	6.1%
Lisp (4)	0.7%
Other (70)	12.0%
None (7)	1.2%

Soon, they'll be speaking R on the subway

Rexer Analytics Survey 2010 results for data mining/analytic tools

<http://r4stats.com/articles/popularity/>

Michael Rennie
(giving 'Kudos to the R support team')
Fortunes(68)



Ancient history: R was already popular in 2004

[R] Kudos to the R support team

Berton Gunter gunter.bernton@gene.com

Thu Jul 29 17:36:06 CEST 2004

- Previous message: [\[R\] Kudos to the R support team](#)
- Next message: [\[R\] Question on getting a data from dataframe](#)
- Messages sorted by: [\[date \]](#) [\[thread \]](#) [\[subject \]](#) [\[author \]](#)

May I add to Michael's comment:

1) There is no formal service; the R Core team members and other regular contributors who do yeo-persons' service do so entirely voluntarily and at their individual discretion..

2) The whole cultural/sociological phenomenon of R strikes me as remarkable. Granting that S and S-Plus provided the essential framework and template, it still strikes me as amazing that such a powerful flexible software system could be constructed, maintained, and documented with such high quality and consistency by such a geographically separated team, especially as they did/do it "part-time" and voluntarily. Surely this is testimony to both their wisdom and hard work.

3). Given their efforts on our behalf, I think it behooves us to be as considerate and solicitous as possible by first following the advice to read Help files, FAQ's and posting guides before wasting their time with queries that do not require their sage advice.

--

Bert Gunter

Non-Clinical Biostatistics
Genentech
MS: 240B
Phone: 650-467-7374



Demonstrative screenshots



fiRst insight: the main program window

RGui (64-bit)

File Edit View Misc Packages Windows Help

R Console

```
> data <- c(rnorm(98, mean=5, sd=1), NA, NA)
> (stats <- DescrStats(data))
```

Descriptive statistics of: data

N:	98	IQR:	0.7853
NA:	2	SD:	1.0009
Min:	2.1586	QD:	0.3927
Q1:	4.2707	SE:	0.1011
Median:	5.056	95%CI mean from:	4.7693
Mean:	4.9699	95%CI mean to:	5.1706
Trimmed mean 5%:	4.97	95%CI median from:	4.8002
Mid-mean:	4.9824	95%CI median to:	5.2649
Winsor mean 5%:	4.9842	Var. coef.:	0.2014
Tukey trimean:	4.99	Kurtosis:	-0.1189
Q3:	5.056	Skewness:	-0.0293
Max:	7.2582	Shapiro pval:	0.9009
Range:	5.0996		

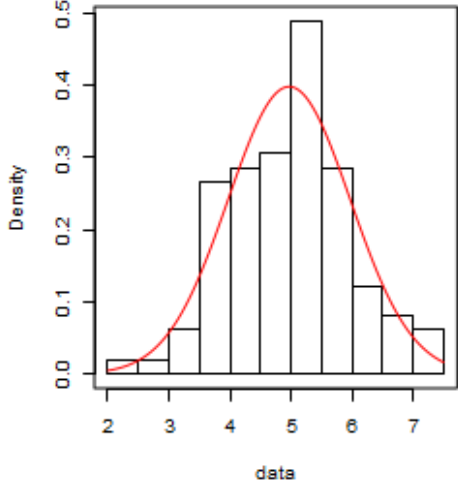
```
> layout(matrix(1:4, nrow=2))
> hist(data, main="Histogram of data", prob=TRUE); box();
> curve(dnorm(x, stats$Mean, stats$SD), add=T, col="red")
> qqnorm(data, main="Quantile-Quantile plot of data")
> boxplot(data, main="Boxplot of data")
> plot(data, main="Scatterplot of data")
> t.test(data, mu=5)
```

One Sample t-test

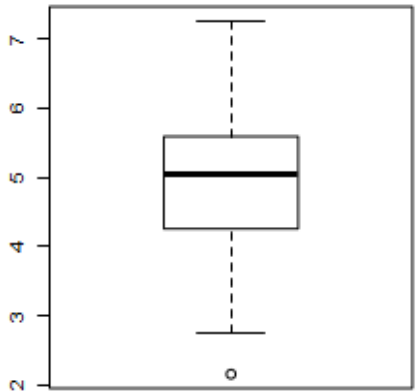
data: data
t = -0.2974, df = 97, p-value = 0.7668
alternative hypothesis: true mean is not equal to 5
95 percent confidence interval:
4.769266 5.170601
sample estimates:
mean of x
4.969933

Cairo Graphics

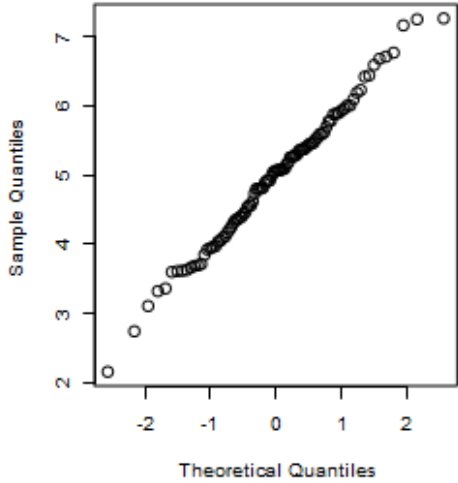
Histogram of data



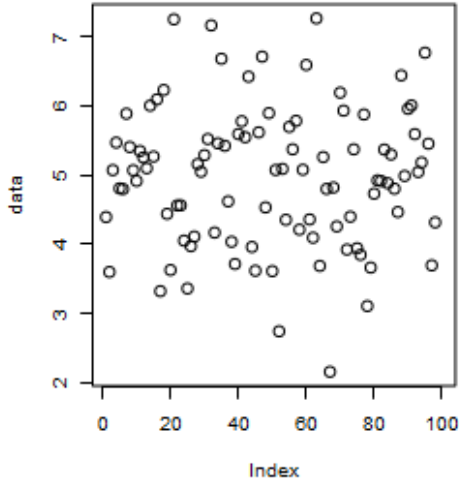
Boxplot of data



Quantile-Quantile plot of data

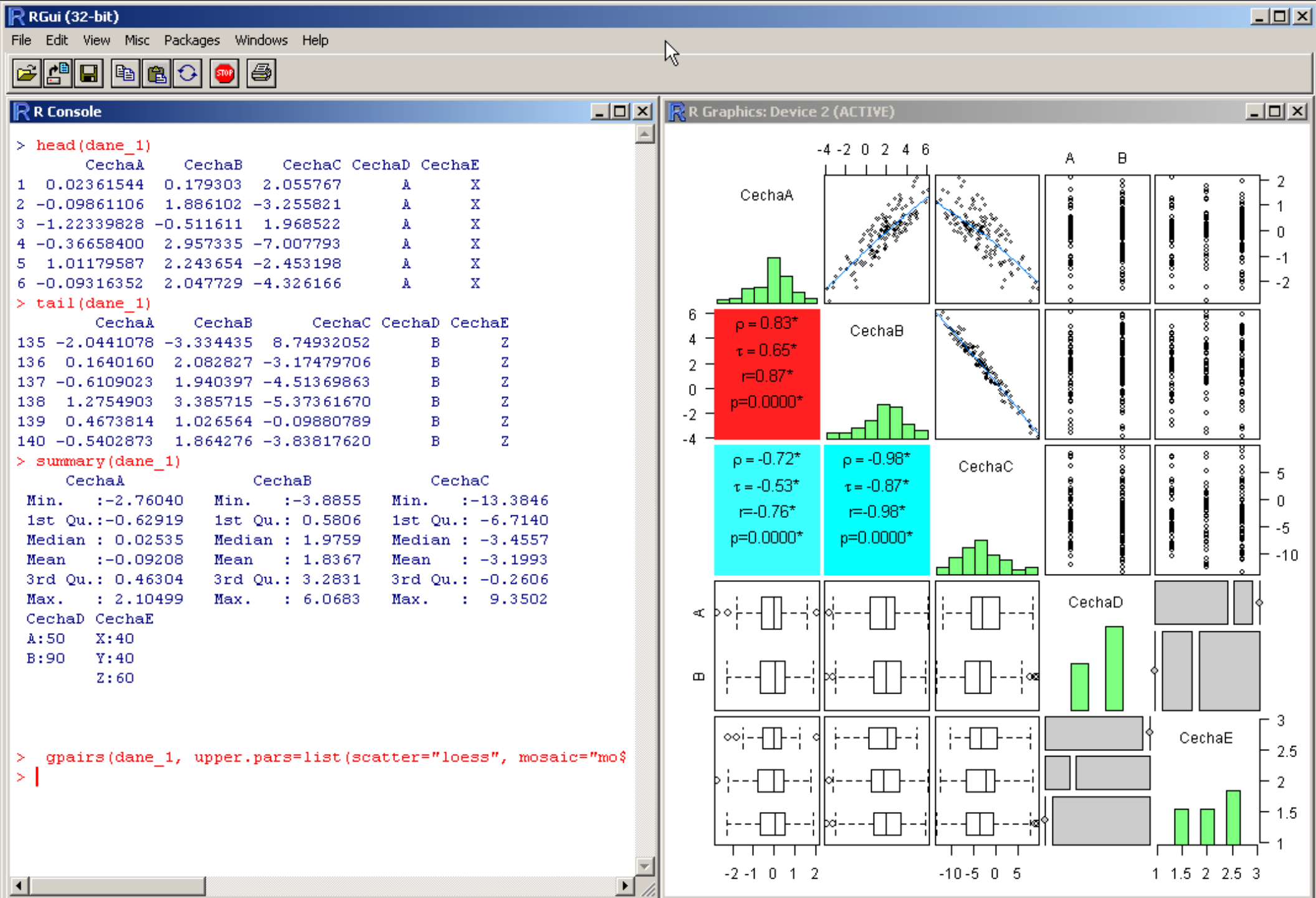


Scatterplot of data

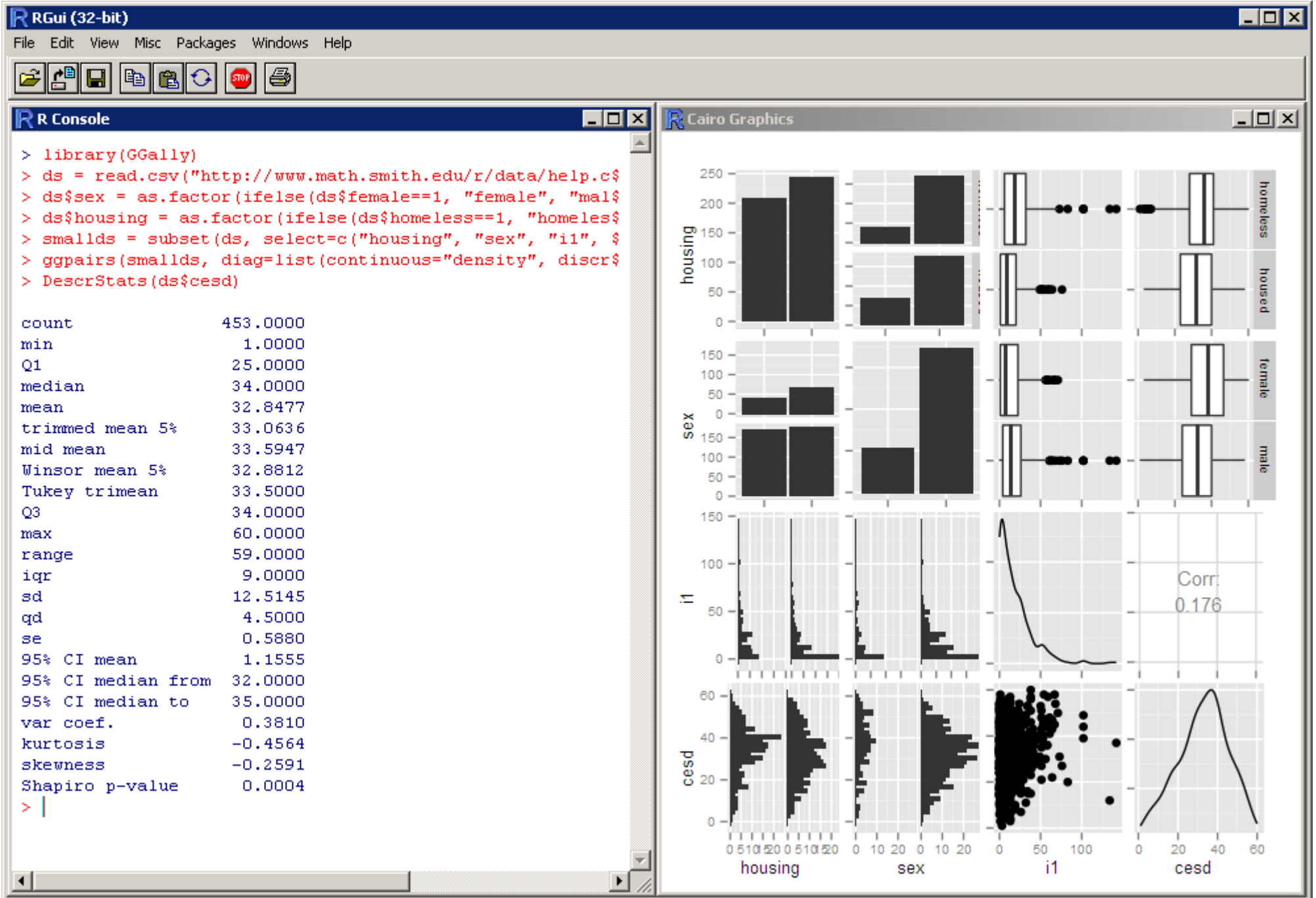




...some quick'n'dirty data inspection...



...quick'n'dirty data inspection other way...





...some linear modeling...

RGui (32-bit)

File Edit View Misc Packages Windows Help



R Console

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-13.89983	0.69406	-20.027	< 2e-16 ***
x1	2.39060	0.24896	9.602	1.46e-12 ***
x2	2.91550	0.26772	10.890	2.52e-14 ***
x3	0.12943	0.06418	2.017	0.0496 *

< Linear regression

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.902 on 46 degrees of freedom
 Multiple R-squared: 0.8211, Adjusted R-squared: 0.8094
 F-statistic: 70.38 on 3 and 46 DF, p-value: < 2.2e-16

> (summ <- summary(model <- aov(dep~indep)))

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
indep	3	607	202.34	166	<2e-16 ***
Residuals	196	239	1.22		

< ANOVA

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> (summ_pairs <- summary(pairs <- glht(model, linfct=mcp(ind\$

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = dep ~ indep)

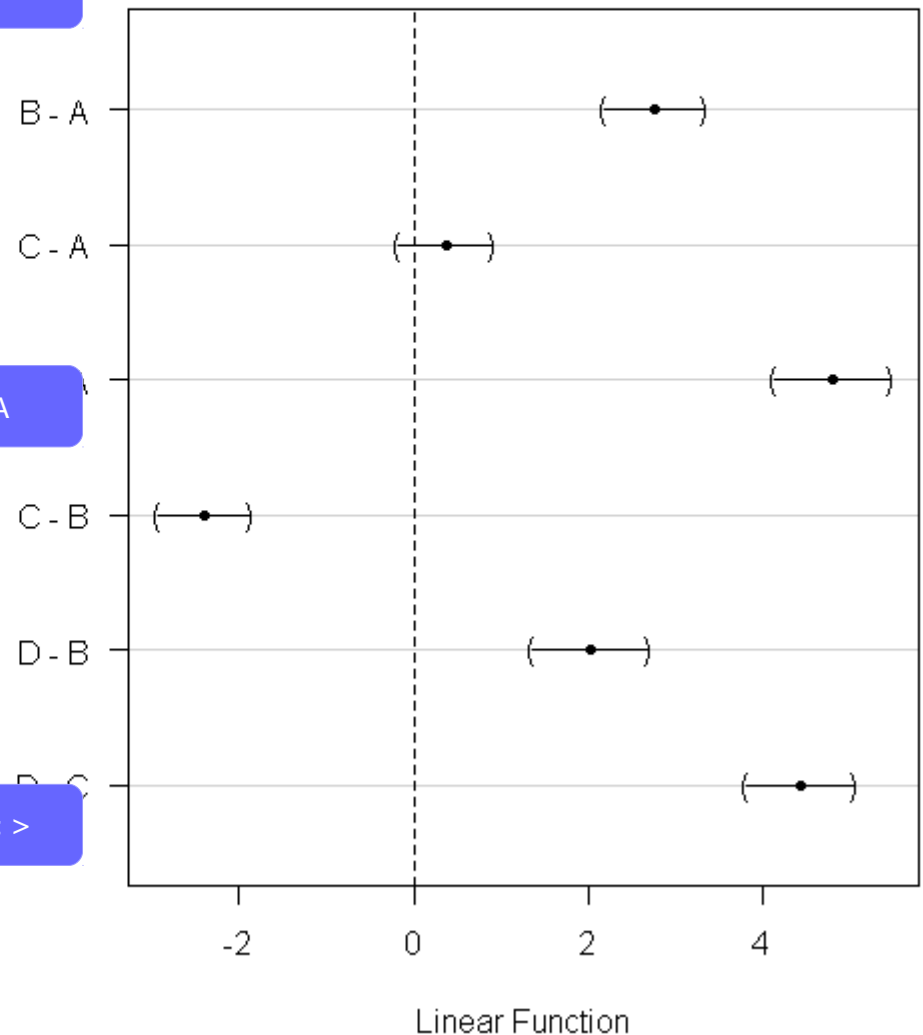
Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t)
B - A == 0	2.7640	0.2208	12.516	<1e-04 ***
C - A == 0	0.3626	0.2045	1.773	0.287
D - A == 0	4.7881	0.2550	18.776	<1e-04 ***
C - B == 0	-2.4014	0.2045	-11.745	<1e-04 ***
D - B == 0	2.0241	0.2550	7.937	<1e-04 ***
D - C == 0	4.4255	0.2410	18.366	<1e-04 ***

< Post-hoc >

R Graphics: Device 2 (ACTIVE)

95% family-wise confidence level



...some linear modeling...

RGui (64-bit)

File Edit View Misc Packages Windows Help

R Console

```
      Df Sum of Sq  RSS   AIC
- hp   1    9.219 169.29 61.307
<none>          160.07 61.515
- qsec 1   20.225 180.29 63.323
- am   1   25.993 186.06 64.331
- wt   1   78.494 238.56 72.284

Step:  AIC=61.31
mpg ~ wt + qsec + am

      Df Sum of Sq  RSS   AIC
<none>          169.29 61.307
- am   1   26.178 195.46 63.908
- qsec 1  109.034 278.32 75.217
- wt   1  183.347 352.63 82.790
> summary(stepmodel)
```

Call:

```
lm(formula = mpg ~ wt + qsec + am, data = mtcars)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.4811	-1.5555	-0.7257	1.4110	4.6610

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.6178	6.9596	1.382	0.177915
wt	-3.9165	0.7112	-5.507	6.95e-06 ***
qsec	1.2259	0.2887	4.247	0.000216 ***
am	2.9358	1.4109	2.081	0.046716 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 2.459 on 28 degrees of freedom
Multiple R-squared: 0.8497, Adjusted R-squared: 0.8336
F-statistic: 52.75 on 3 and 28 DF, p-value: 1.21e-11

Cairo Graphics

Residuals vs Fitted

Normal Q-Q

Scale-Location

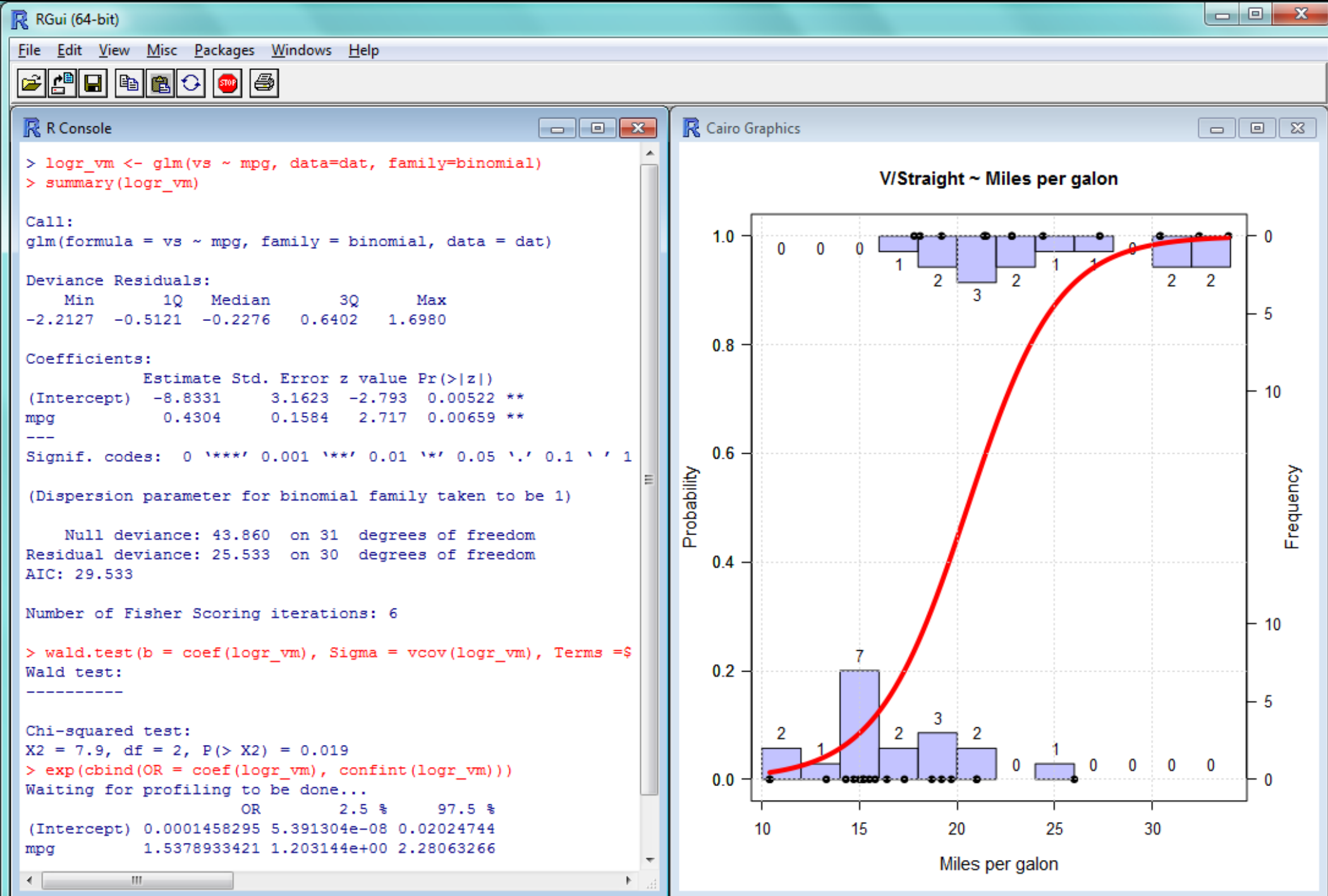
Residuals vs Leverage

< stepwise regression

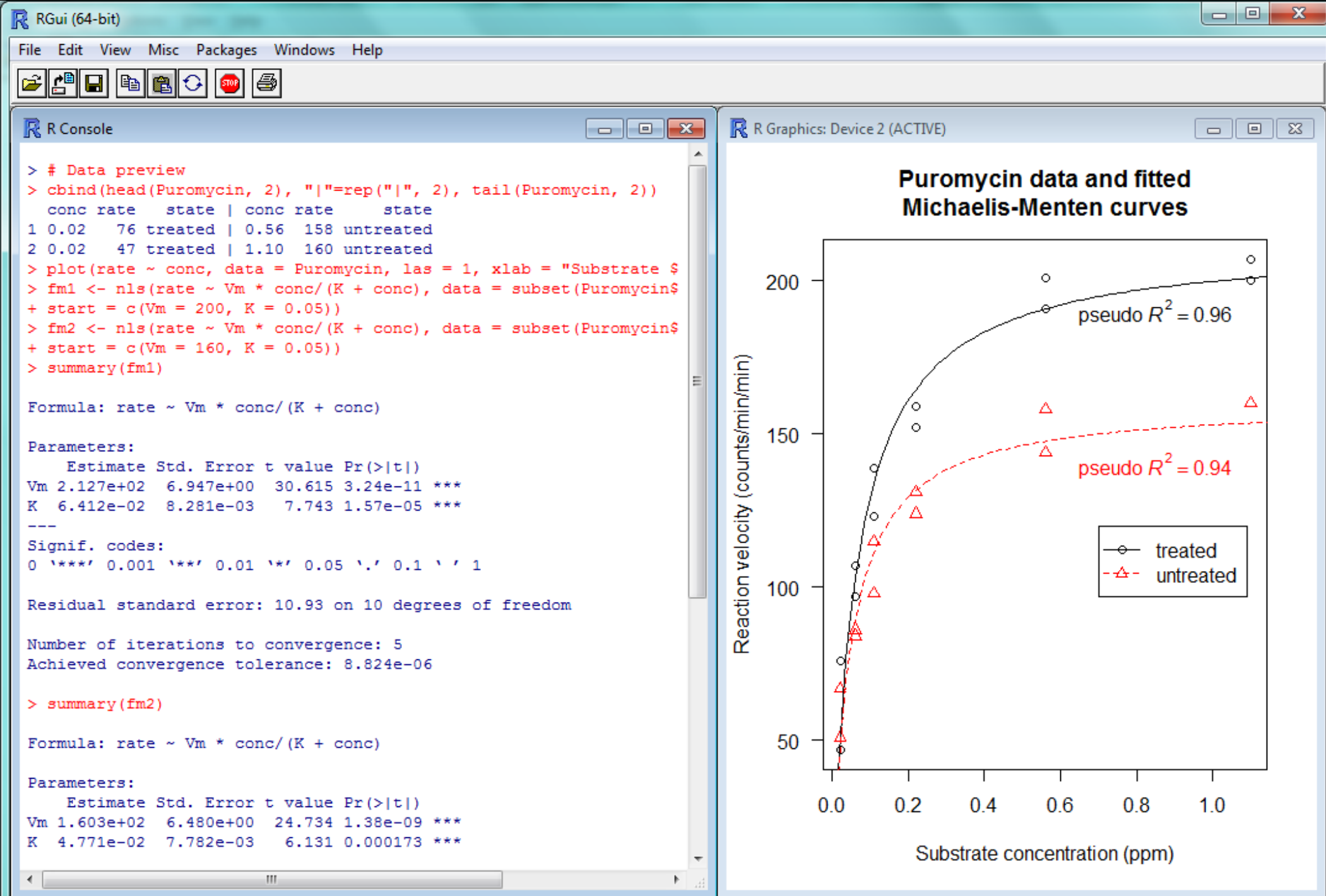
< best model

diagnostics >

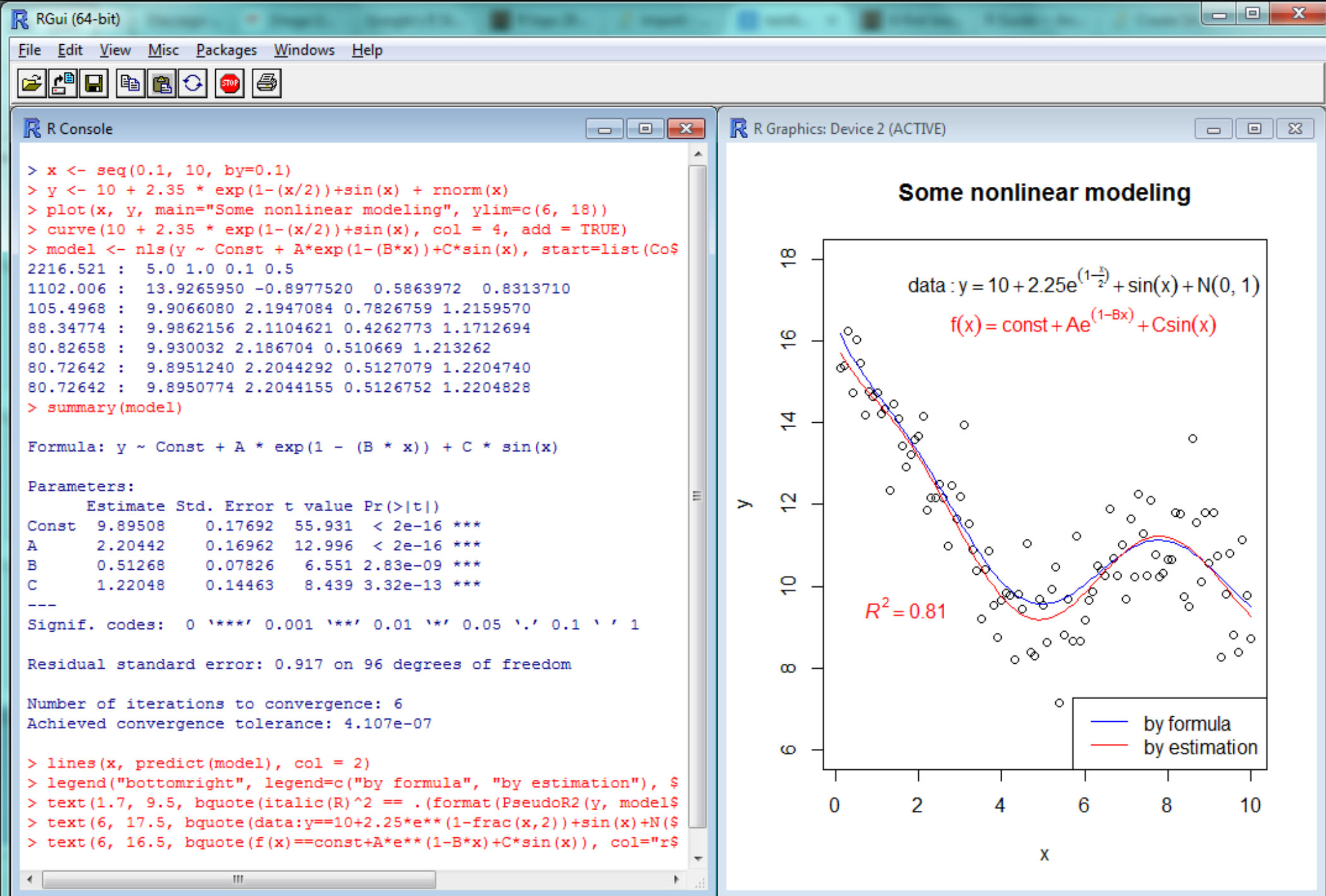
...some GLM modeling...



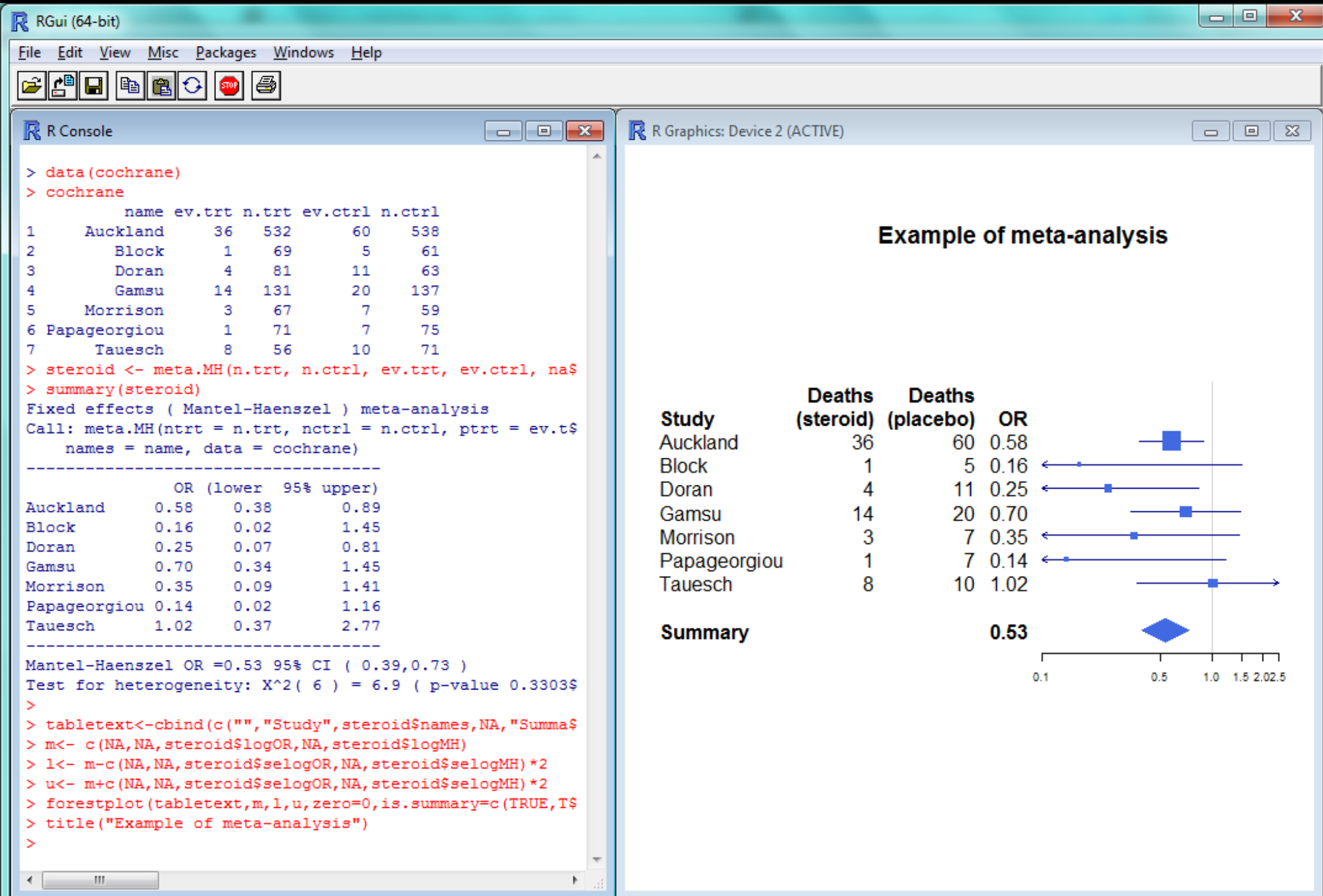
...and nonlinear modeling...



...and nonlinear modeling...

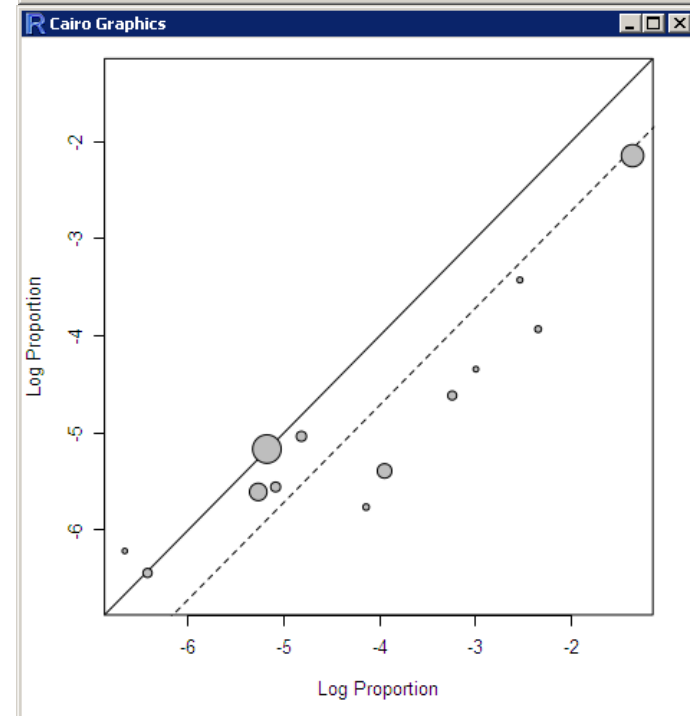
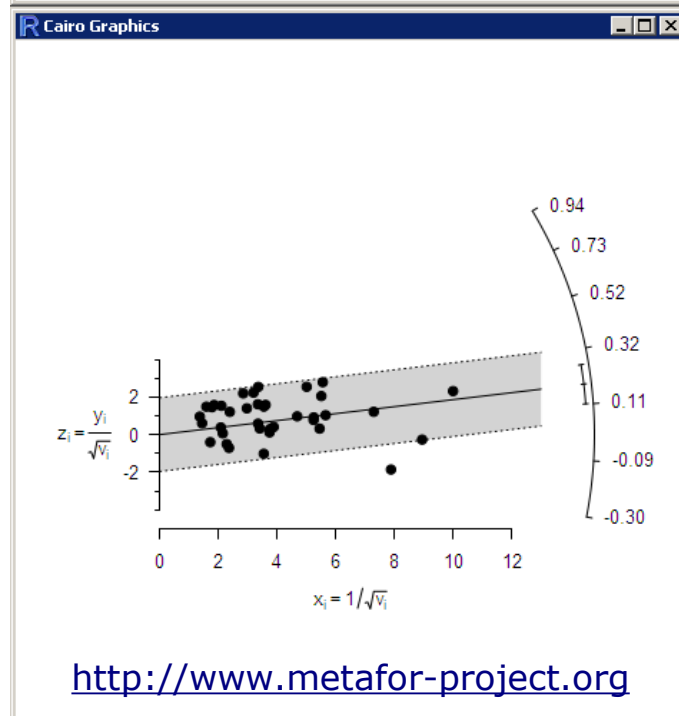
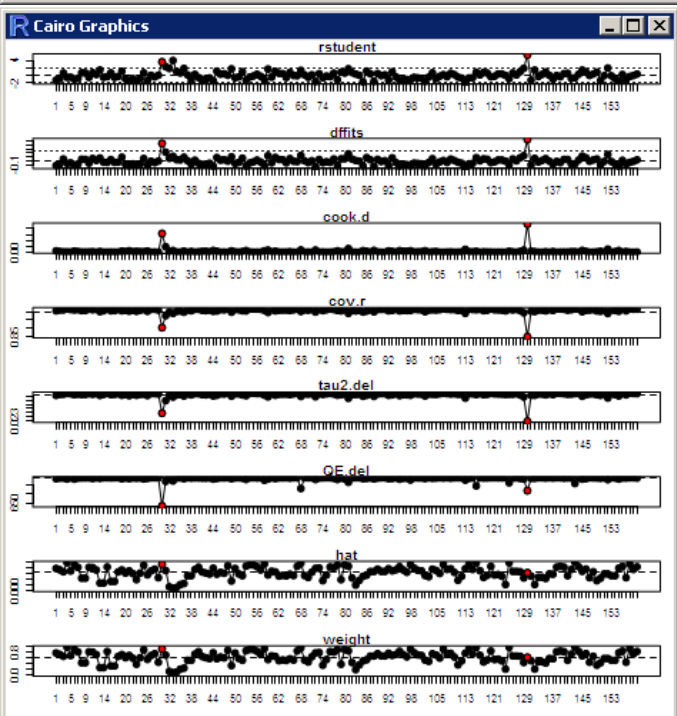
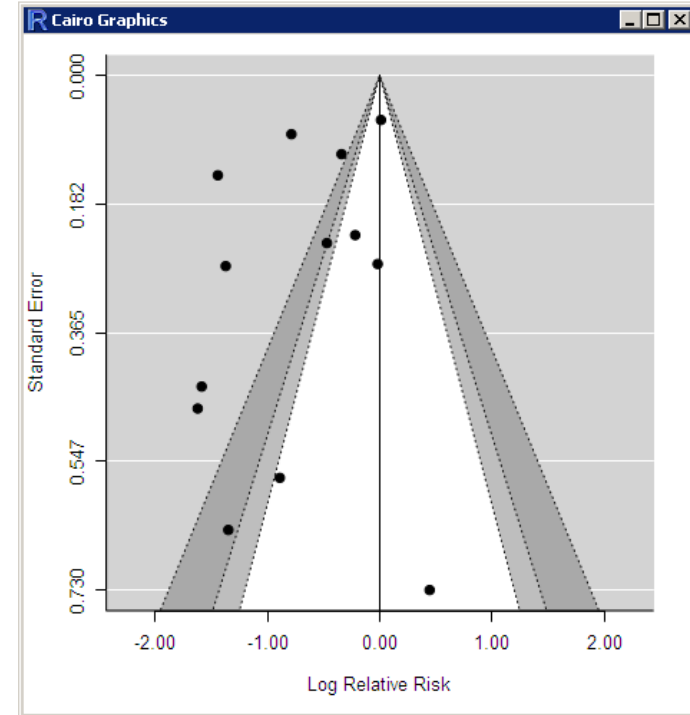
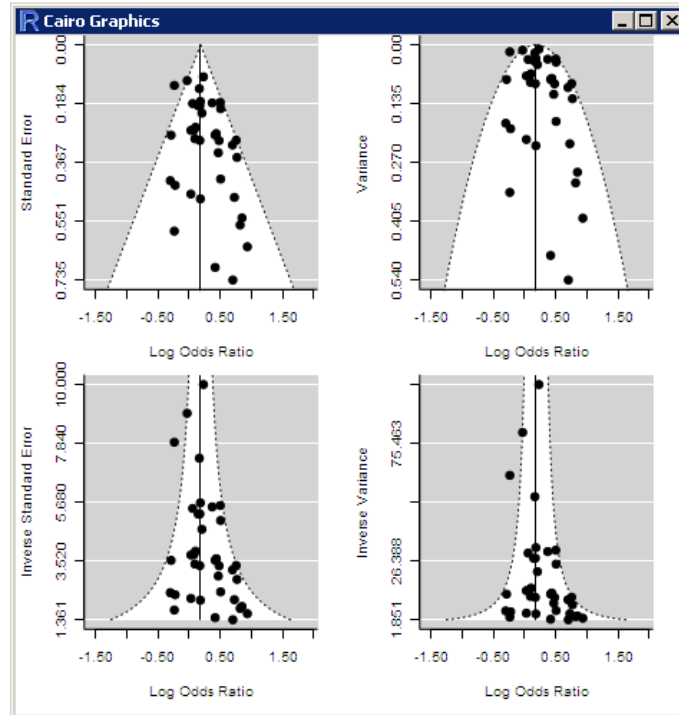
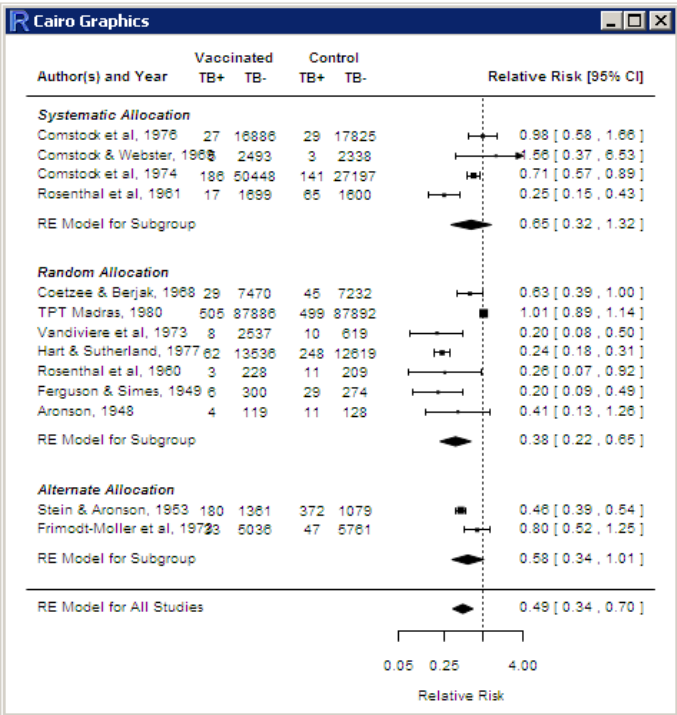


...and meta-analysis...



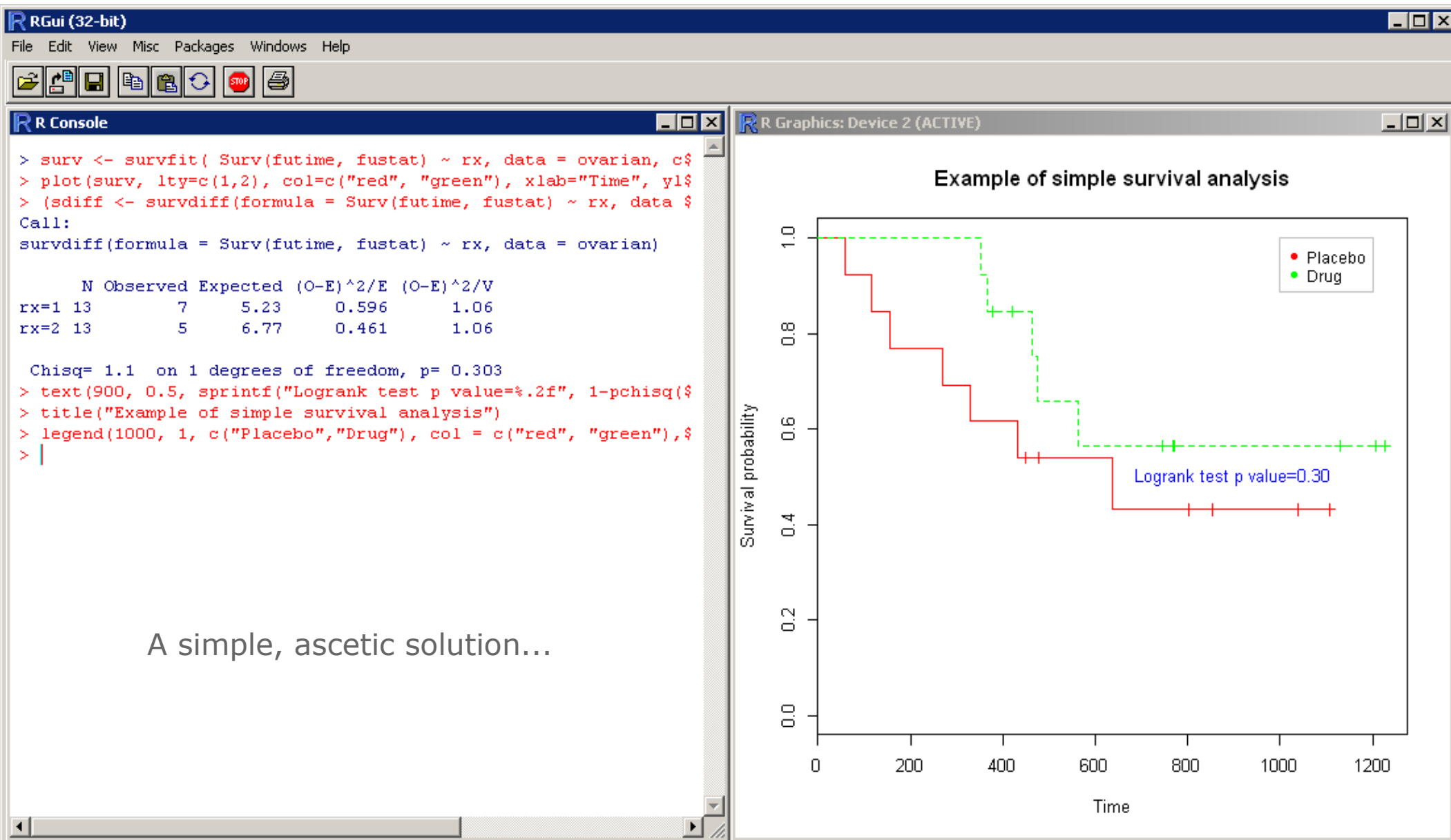


...a bit more detailed output...





...and survival analysis...



A simple, ascetic solution...



...and survival analysis...

RGui (32-bit)

File Edit View Misc Packages Windows Help

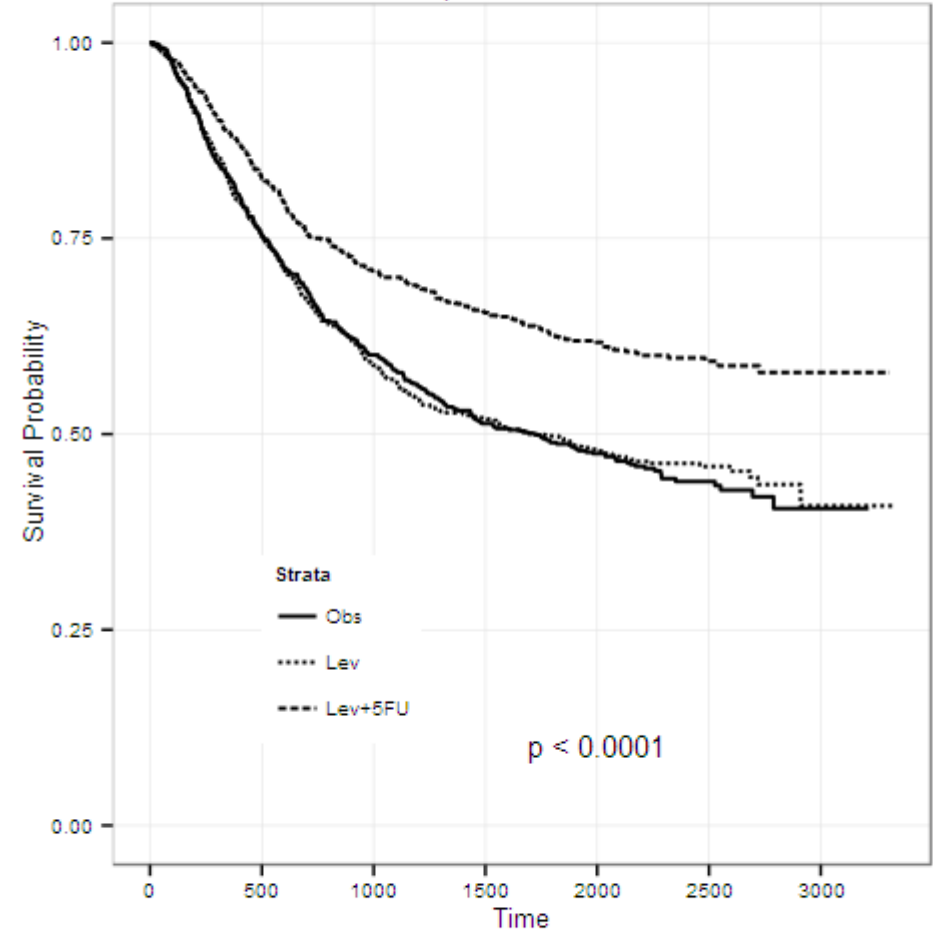


R Console

```
> CairoWin()
> library(survival)
> data(colon)
> fit <- survfit(Surv(time,status)~rx, data=colon)
> ggkm(fit, timeby=500, ystratalabs=c("Obs","Lev","Lev+5FU")$
> |
```

R Cairo Graphics

Kaplan-Meier Plot

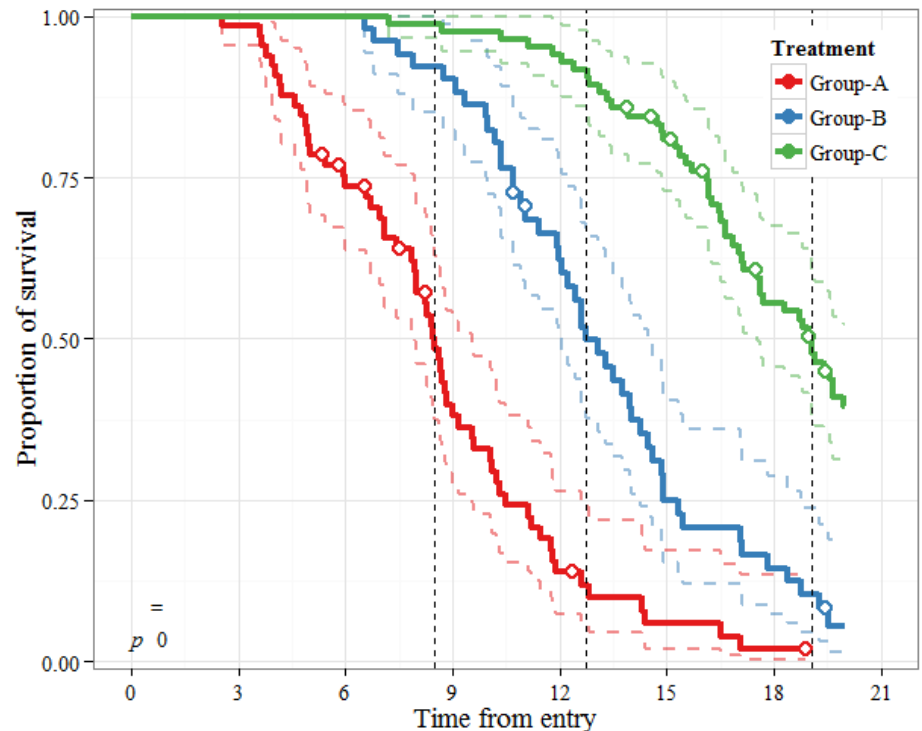


Obs	630	470	372	315	256	90	11
Lev	620	464	360	318	266	107	8
Lev+5FU	608	498	425	387	328	127	14

Numbers at risk

R Graphics: Device 2 (ACTIVE)

Sample Kaplan-Meier curves



Group-A	65	64	46	22	8	3	1	0
Group-B	51	51	51	46	30	12	7	0
Group-C	84	84	84	82	79	66	43	0



...and dose-response analysis...

RGui (32-bit)

File Edit View Misc Packages Windows Help



R Console

```

> library(drc)
> modell1 <- drm(SLOPE~DOSE, CURVE, fct=LL.4($
> summary(modell1)

```

Model fitted: Log-logistic (ED50 as parameter)

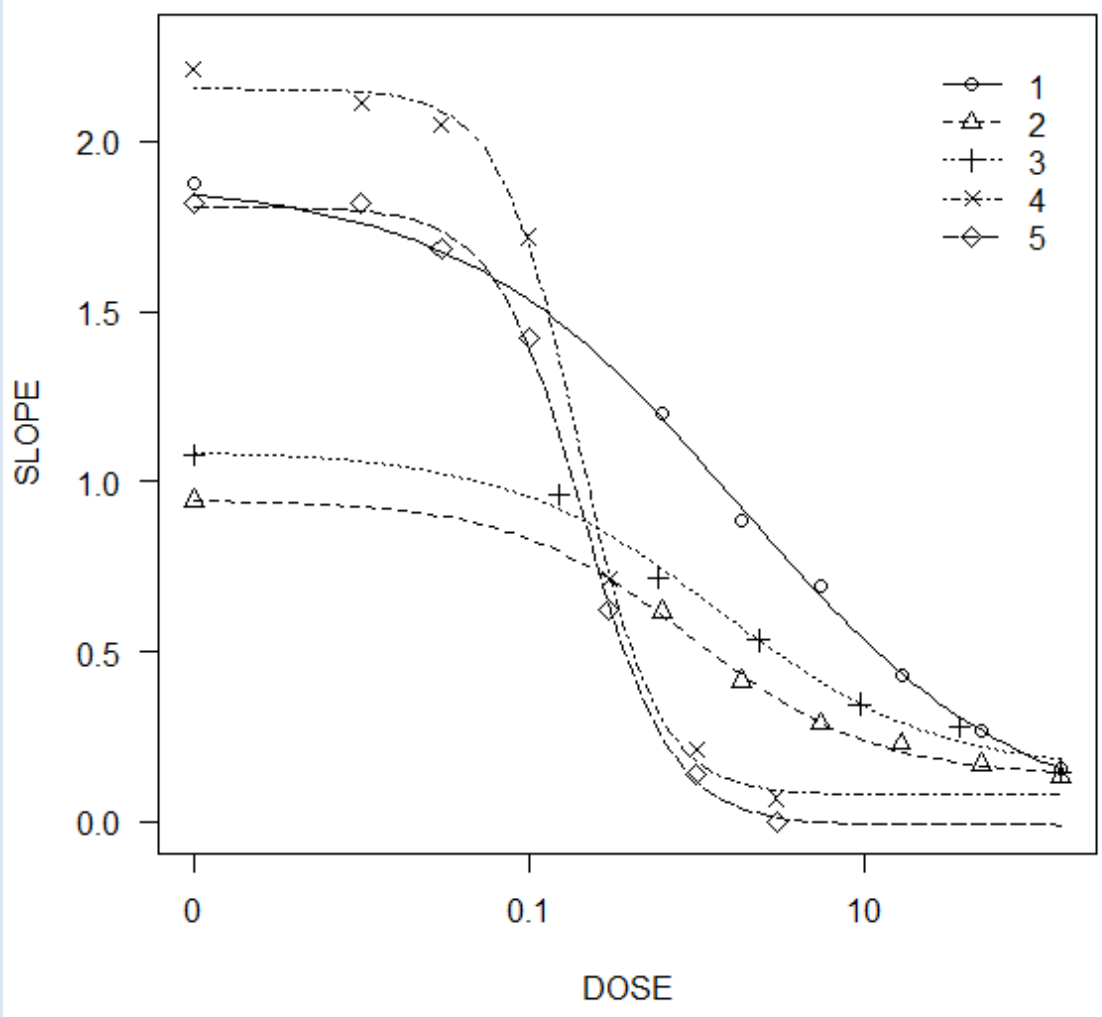
Parameter estimates:

	Estimate	Std. Error	t-value
Slope:1	0.5195192	0.0763600	6.8035513
Slope:2	0.8007959	0.2256794	3.5483791
Slope:3	0.6819134	0.1285568	5.3043748
Slope:4	1.8448094	0.1663521	11.0897851
Slope:5	1.6507576	0.1758293	9.3884079
Lower:1	-0.0165952	0.1078254	-0.1539078
Lower:2	0.1325890	0.0471932	2.8094922
Lower:3	0.1464061	0.0604288	2.4227879
Lower:4	0.0795516	0.0394596	2.0160251
Lower:5	-0.0090656	0.0443536	-0.2043946
Upper:1	1.8795534	0.0423710	44.3594092
Upper:2	0.9460003	0.0422667	22.3816998
Upper:3	1.0903215	0.0405604	26.8813972
Upper:4	2.1535780	0.0281853	76.4079202
Upper:5	1.8062825	0.0292460	61.7616182
ED50:1	1.7949548	0.4782321	3.7533130
ED50:2	0.9455299	0.2494933	3.7898008
ED50:3	1.3730228	0.4526848	3.0330656
ED50:4	0.1973263	0.0101895	19.3656923
ED50:5	0.2107935	0.0138248	15.2475365

	p-value
Slope:1	0.0000
Slope:2	0.0006
Slope:3	0.0000
Slope:4	0.0000
Slope:5	0.0000

R Graphics: Device 2 (ACTIVE)

Simple Dose-Response analysis





...and ROC curves...

RGui (32-bit)

File Edit View Misc Packages Windows Help

R Console

```
> plot.roc(aSAH$outcome, aSAH$s100b,
+ main="Confidence interval of a threshold",
+ ci=TRUE, of="thresholds", thresholds="best$

Call:
plot.roc.default(x = aSAH$outcome, predictor$

Data: aSAH$s100b in 72 controls (aSAH$outcom$
Area under the curve: 73.14%
95% CI (2000 stratified bootstrap replicates$
thresholds sp.low sp.median sp.high se.low
0.205 70.83 80.56 88.89 48.78
se.median se.high
63.41 78.05

>
> plot.roc(aSAH$outcome, aSAH$s100b, # data
+ percent=TRUE, # show all values in percent
+ partial.auc=c(100, 90), partial.auc.correc$
+ print.auc=TRUE, #display pAUC value on the$
+ print.auc.pattern="Corrected pAUC (100-90%$
+ auc.polygon=TRUE, auc.polygon.col="#1c61b6$
+ max.auc.polygon=TRUE, max.auc.polygon.col=$
+ main="Partial AUC (pAUC)")

Call:
plot.roc.default(x = aSAH$outcome, predictor$

Data: aSAH$s100b in 72 controls (aSAH$outcom$
Corrected partial area under the curve (spec$
>
> plot.roc(aSAH$outcome, aSAH$s100b,
+ percent=TRUE, add=TRUE, type="n", # add to$
+ partial.auc=c(100, 90), partial.auc.correc$
+ partial.auc.focus="se", # focus pAUC on th$
+ print.auc=TRUE, print.auc.pattern="Correct$
```

R Graphics: Device 2 (ACTIVE)

Confidence intervals of specificity/sensiti

Confidence interval of a threshold

Statistical comparison

Partial AUC (pAUC)



13 reasons why you will love GNU R



13 reasons why **you will** love GNU R

I R is (extremely) cheap. In fact - it's free :)

II R has (extremely) wide range of capabilities

III R is supported by the world of science

IV R is supported by the community

V R is supported by the business

VI R is able to read data in many formats

VII Interoperability is easy to achieve

VIII R is truly cross-platform

IX R offers numerous ways of presenting data

X There are many options to optimize the code

XI R is able to handle large amount of data

XII R has a set of fancy tools and IDEs

XIII FDA: R is OK for drug trials! 



GNU R is a **free software**.

One can legally use it, even commercially, at no cost.

Some companies provide their own, both free and commercial, customized versions of R along with **support and integration services**. Well known providers are: **Revo**lution (since 2015 part of **Microsoft**), **RStudio** and **Oracle**.

REVOLUTION
ANALYTICS

<http://buy.revolutionanalytics.com/>

<http://blogs.microsoft.com/blog/2015/01/23/...>



<http://www.rstudio.com...>

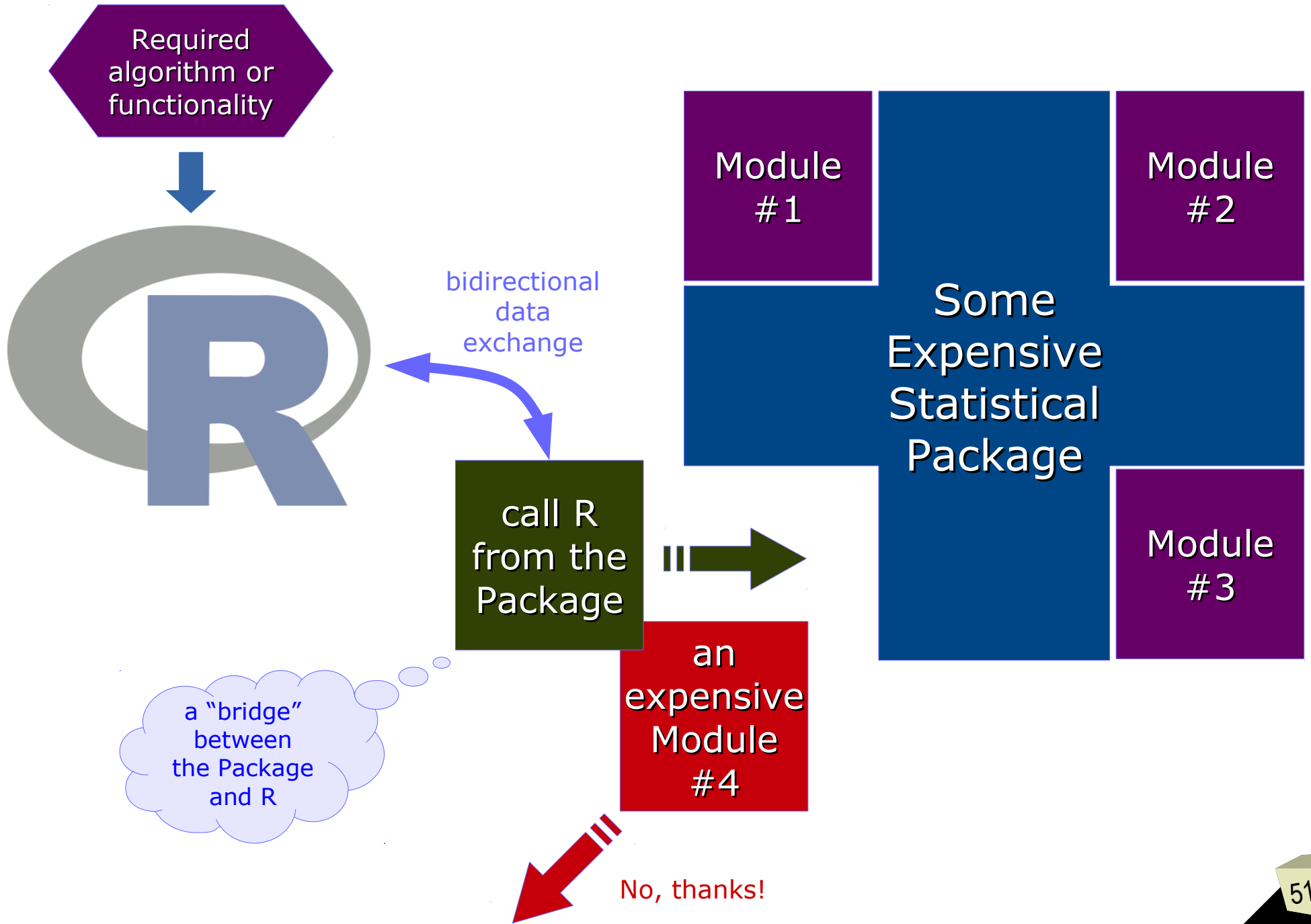
R Technologies from Oracle
Bringing the Power of R to the Enterprise



<http://www.oracle.com/technetwork/...>



save your money with R!

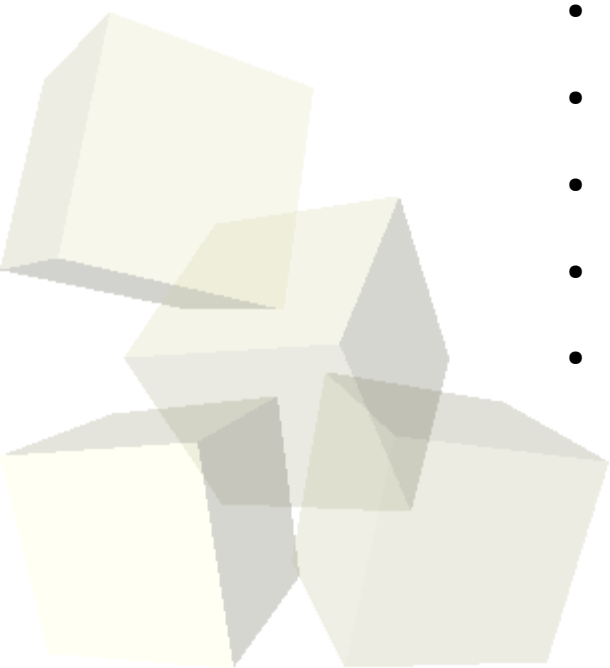




The following licenses are in use for R and associated software:

- GNU **A**ffero **GPL** v.3
- **A**rtistic **L**icense" v.2.0
- **BSD** 2-clause
- **BSD** 3-clause
- GNU **GPL** v.2
- GNU **GPL** v.3
- GNU **L**ibrary **GPL** v.2
- GNU **L**esser **GPL** v.2.1
- GNU **L**esser **GPL** v.3
- **M**assachusetts **I**nstitute of **T**echnology (X11)

R as a package is licensed under **GPLv2**





13 reasons why **you will** love GNU R

- I R is (extremely) cheap. In fact - it's free :)
- II R has (extremely) wide range of capabilities**
- III R is supported by the world of science
- IV R is supported by the community
- V R is supported by the business
- VI R is able to read data in many formats
- VII Interoperability is easy to achieve
- VIII R is truly cross-platform
- IX R offers numerous ways of presenting data
- X There are many options to optimize the code
- XI R is able to handle large amount of data
- XII R has a set of fancy tools and IDEs
- XIII FDA accepted using R for drug trials!



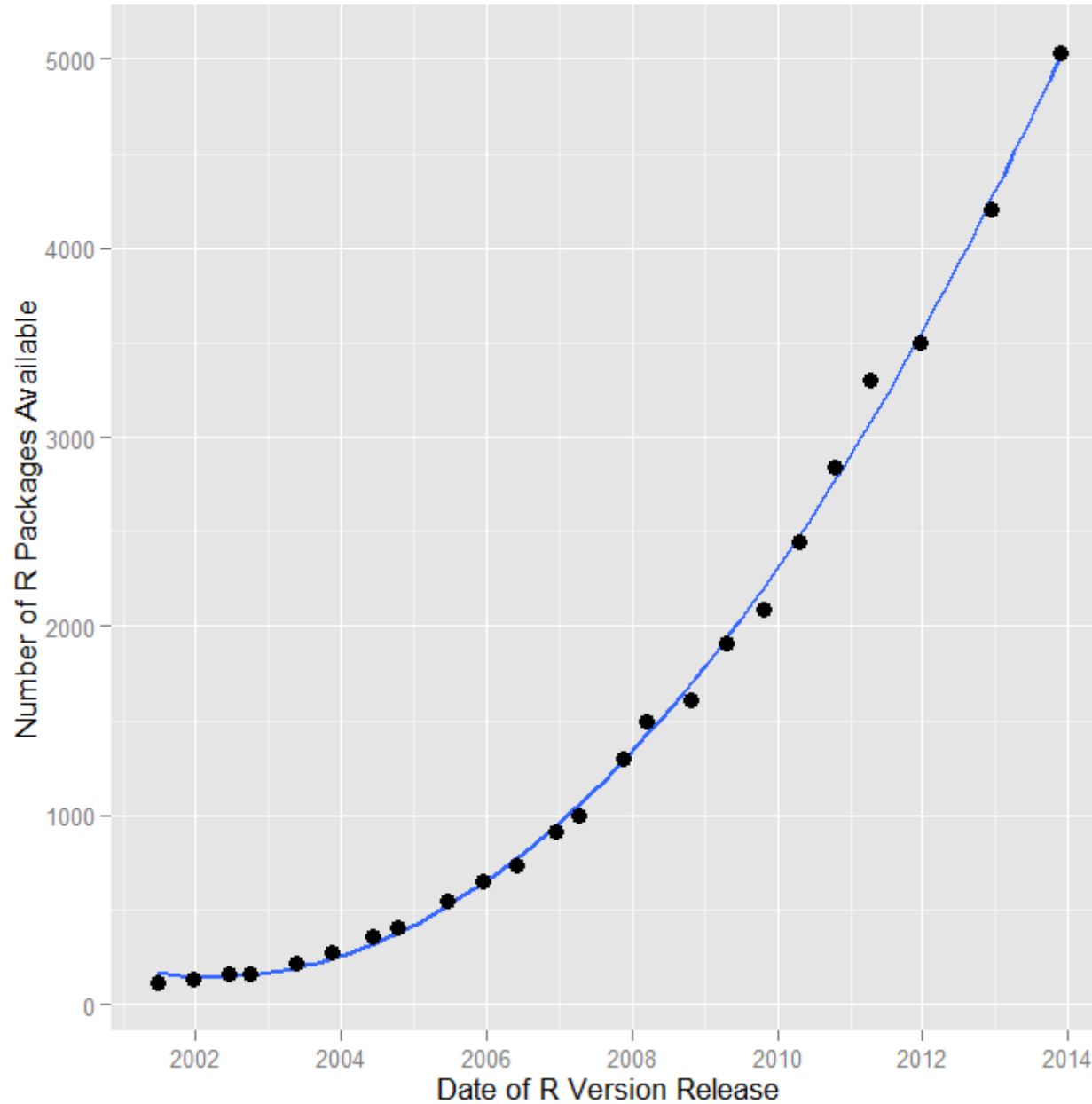
Sit down and hold tight! [CRAN](http://cran.r-project.org) holds more than **6600** `nrow(available.packages())` packages.

Just describe your problem or ask me for a statistical test or procedure and I will give you the right package(s): <http://cran.r-project.org/web/views>

- Linear models of ANY kind. (M)AN(C)OVA, regression (linear, logistic, hierarchical, etc.)
- Post-factum analysis and planned comparisons
- Nonlinear models (NLS, Generalized Additive Models with a rich set of smoothers), trees
- Robust methods: Regularized methods, M-estimators
- Models with fixed and random effects (mixed models). Variance components
- Monte Carlo methods (permutational, bootstrap). Exact methods.
- [Survival](#) • [PK/PD](#) • [Superior / Non inferior. / Equiv. / Bioequiv. trials](#) • [Meta-analysis](#)
- [Design of experiments](#) – including those applicable in clinical research
- Structural equations. Time series. Forecasting.
- Methods for analyzing multidimensional data: NN, SVM, LDA/QDA, PCA, FA, CC, MS, KLT, CA, MART, POLYMARS, PINDIS, PPR, ACE, AVAS, K-means, KN, and lots of more!
- Trees (CART, CHAID). Random forests. Aggregating (boosting, bagging)
- Reproducible research. Graphical User Interfaces. Widely understood interoperability.



Growth in the number of packages





A list of [thematic sections](#) covered by the CRAN repository:

- Bayesian Inference
- Chemometrics and Computational Physics
- Clinical Trial Design, Monitoring, and Analysis
- Cluster Analysis & Finite Mixture Models
- Differential Equations
- Probability Distributions
- Computational Econometrics
- Analysis of Ecological and Environmental Data
- Design of Experiments (DoE) & Analysis of Experimental Data
- Empirical Finance
- Statistical Genetics
- Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization
- High-Performance and Parallel Computing with R
- Machine Learning & Statistical Learning
- Medical Image Analysis
- Meta-Analysis
- Multivariate Statistics
- Natural Language Processing
- Numerical Mathematics
- Official Statistics & Survey Methodology
- Optimization and Mathematical Programming
- Analysis of Pharmacokinetic Data
- Phylogenetics, Especially Comparative Methods
- Psychometric Models and Methods
- Reproducible Research
- Robust Statistical Methods
- Statistics for the Social Sciences
- Analysis of Spatial Data
- Handling and Analyzing Spatio-Temporal Data
- Survival Analysis
- Time Series Analysis
- Web Technologies and Services
- gRaphical Models in R



What kind of analyses common in clinical research can be done in R?

- Descriptive statistics, summaries (demographic, recruitment)
- Advanced, linear and nonlinear modeling (models of any type)
- Comparisons of treatments
- PK / PD analysis
- Analysis of bio-equivalence, non-inferiority, superiority
- Design of experiments
- Time-to-event analysis (survival analysis)
- Analysis of data from longitudinal trials
- Sample size determination and power analysis
- Meta-analysis
- Bayesian analyzes
- Analysis of Adverse Events
- Analysis of DNA micro-arrays
- ROC curves



<http://cran.r-project.org/web/views/ClinicalTrials.html>

<http://cran.r-project.org/web/views/Survival.html>

<http://cran.r-project.org/web/views/ExperimentalDesign.html>

<http://cran.r-project.org/web/views/Pharmacokinetics.html>

The screenshot shows a Firefox browser window with the address bar containing cran.r-project.org/web/views/ClinicalTrials.html. The page content includes:

CRAN Task View: Clinical Trial Design, Monitoring, and Analysis

Maintainer: Ed Zhang and Harry G. Zhang
Contact: Ed.Zhang.jr at gmail.com
Version: 2014-02-02

This task view gathers information on specific R packages for design, monitoring and analysis of data from clinical trials. It focuses on including packages for clinical trial design and monitoring in general plus data analysis packages for a specific type of design. Also, it gives a brief introduction to important packages for analyzing clinical trial data. Please refer to task views [ExperimentalDesign](#), [Survival](#), [Pharmacokinetics](#) for more details on these topics. Please feel free to e-mail me regarding new packages or major package updates.

Design and Monitoring

- [TrialSize](#) This package has more than 80 functions from the book *Sample Size Calculations in Clinical Research* (Chow & Wang & Shao, 2007, 2nd ed., Chapman & Hall/CRC).
- [asd](#) This Package runs simulations for adaptive seamless designs using early outcomes for treatment selection.
- [bcrm](#) This package implements a wide variety of one and two-parameter Bayesian CRM designs. The program can run interactively, allowing the user to enter outcomes after each cohort has been recruited, or via simulation to assess operating characteristics.
- [blockrand](#) creates randomizations for block random clinical trials. It can also produce a PDF file of randomization cards.
- [conf.design](#) This small package contains a series of simple tools for constructing and manipulating confounded and fractional factorial designs.

CRAN packages:

- adaptTest
- AGSDest
- asd
- asypow
- bcrm
- bifactorial
- blockrand
- clinfun
- clinsig
- coin
- conf.design
- copas
- CRM
- CRTSize
- dfcrm
- DoseFinding
- epibasix
- epicalc
- experiment
- FrF2
- GroupSeq
- gsDesign
- HH
- Hmisc
- ldbounds
- longpower
- MChtest
- MCPMod
- meta
- metafor
- metaLik
- multcomp
- nppbib
- PIPS
- PowerTOST
- pwr
- PwrGSD
- qtIDesign
- rmeta
- samplesize
- seqmon
- speff2trial
- ssanv
- survival
- tdm
- TEQR
- TrialSize



Package: TrialSize

"More than 80 functions in this package are widely used to calculate sample size in clinical trial research studies.

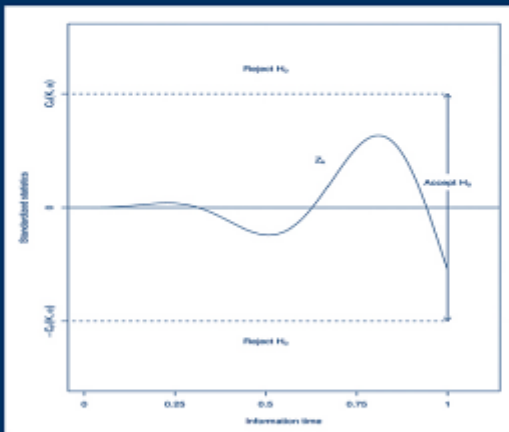
This package covers the functions in Chapter 3,4,6,7,9,10,11,12,14,15 of the reference book"

- AB.withDescalation
- AB.withoutDescalation
- ABE
- ANOVA.Repeat.Measure
- Carry.Over
- Cochran.Armitage/Trend
- Cox.Equality
- Cox.Equivalence
- Cox.NIS
- CrossOver.ISV.Equality
- CrossOver.ISV.Equivalence
- CrossOver.ISV.NIS
- Dose.MinE.ffect
- Dose.Response.binary
- Dose.Response.Linear
- Dose.Response.time.to.event
- gof.Pearson
- gof.Pearson.twoway
- IBE
- InterSV.Equality
- InterSV.NIS
- ISCV.Equality
- ISCV.Equivalence
- ISCV.NIS
- ISV.Equality
- ISV.Equivalence
- ISV.NIS
- McNemar.Test
- MeanWilliamsDesign.Equality
- MeanWilliamsDesign.Equivalence
- MeanWilliamsDesign.NIS
- Multiple.Testing
- Nonpara.Independ
- Nonpara.One.Sample
- Nonpara.Two.Sample
- OneSampleMean.Equality
- OneSampleMean.Equivalence
- OneSampleMean.NIS
- OneSampleProportion.Equality
- OneSampleProportion.Equivalence
- OneSampleProportion.NIS
- OneSide.fixEffect
- OneSide.varyEffect
- OneWayANOVA.pairwise
- OneWayANOVA.PairwiseComparison
- PBE
- Propensity.Score.nostrata
- Propensity.Score.strata
- QOL
- QT.crossover
- QT.parallel
- QT.PK.crossover
- QT.PK.parallel
- RelativeRisk.Equality
- RelativeRisk.Equivalence
- RelativeRisk.NIS
- RelativeRiskCrossOver.Equality
- RelativeRiskCrossOver.Equivalence
- RelativeRiskCrossOver.NIS
- Sensitivity.Index
- Stuart.Maxwell.Test
- TwoSampleCrossOver.Equality
- TwoSampleCrossOver.Equivalence
- TwoSampleCrossOver.NIS
- TwoSampleMean.Equality
- TwoSampleMean.Equivalence
- TwoSampleMean.NIS
- TwoSampleProportion.Equality
- TwoSampleProportion.Equivalence
- TwoSampleProportion.NIS
- TwoSampleSeqCrossOver.Equality
- TwoSampleSeqCrossOver.Equivalence
- TwoSampleSeqCrossOver.NIS
- TwoSampleSurvival.Conditional
- TwoSampleSurvival.Equality
- TwoSampleSurvival.Equivalence
- TwoSampleSurvival.NIS
- TwoSide.fixEffect
- TwoSide.varyEffect
- Vaccine.CEM
- Vaccine.ELDI
- Vaccine.RDI
- Vitro.BE
- WilliamsDesign.Equality
- WilliamsDesign.Equivalence
- WilliamsDesign.NIS

 Chapman & Hall/CRC Biostatistics Series

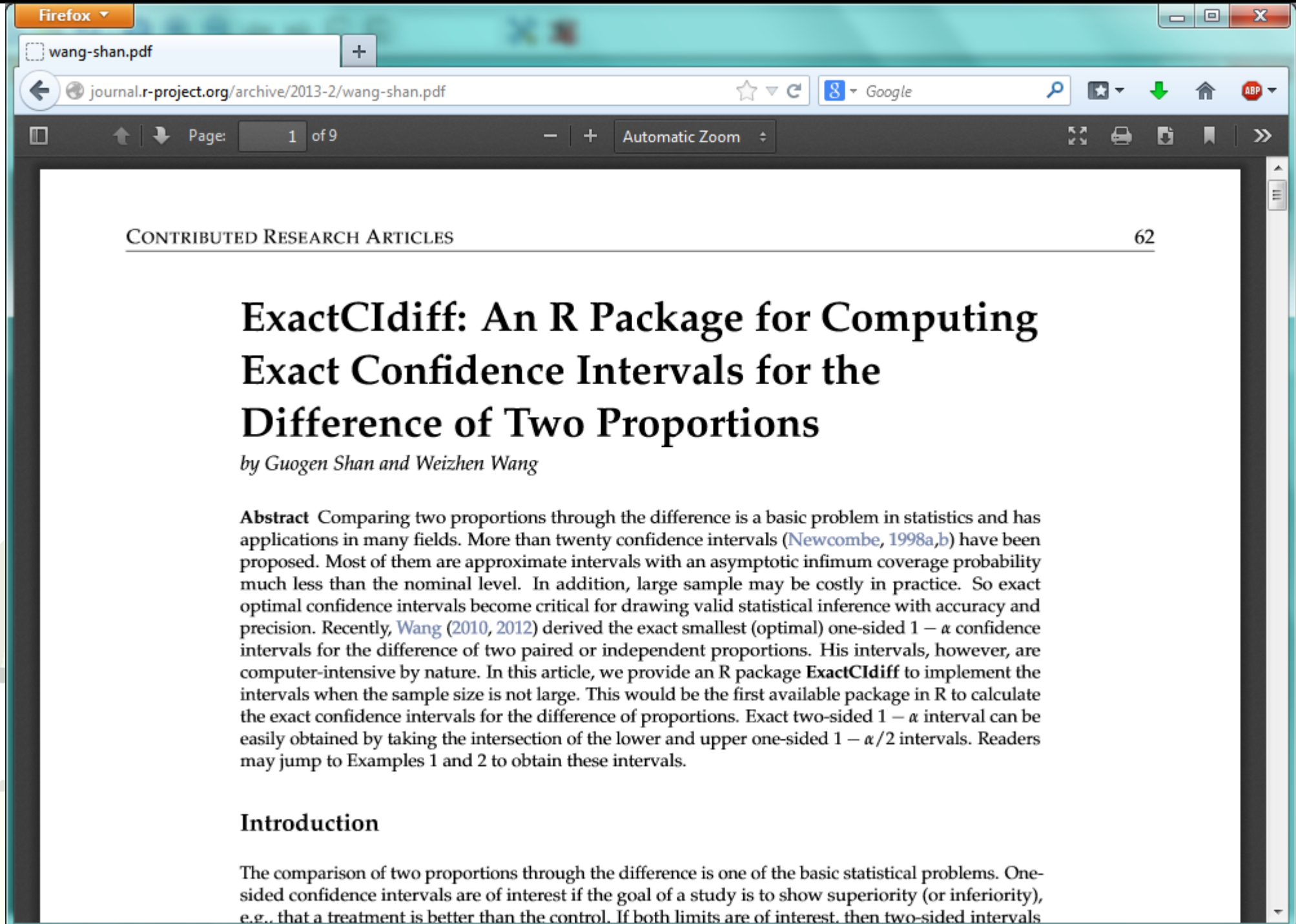
SAMPLE SIZE CALCULATIONS IN CLINICAL RESEARCH

Second Edition



Shein-Chung Chow
Jun Shao
Hansheng Wang

 Chapman & Hall/CRC
Taylor & Francis Group



Firefox

wang-shan.pdf

journal.r-project.org/archive/2013-2/wang-shan.pdf

Page: 1 of 9 Automatic Zoom

CONTRIBUTED RESEARCH ARTICLES 62

ExactCIdiff: An R Package for Computing Exact Confidence Intervals for the Difference of Two Proportions

by Guogen Shan and Weizhen Wang

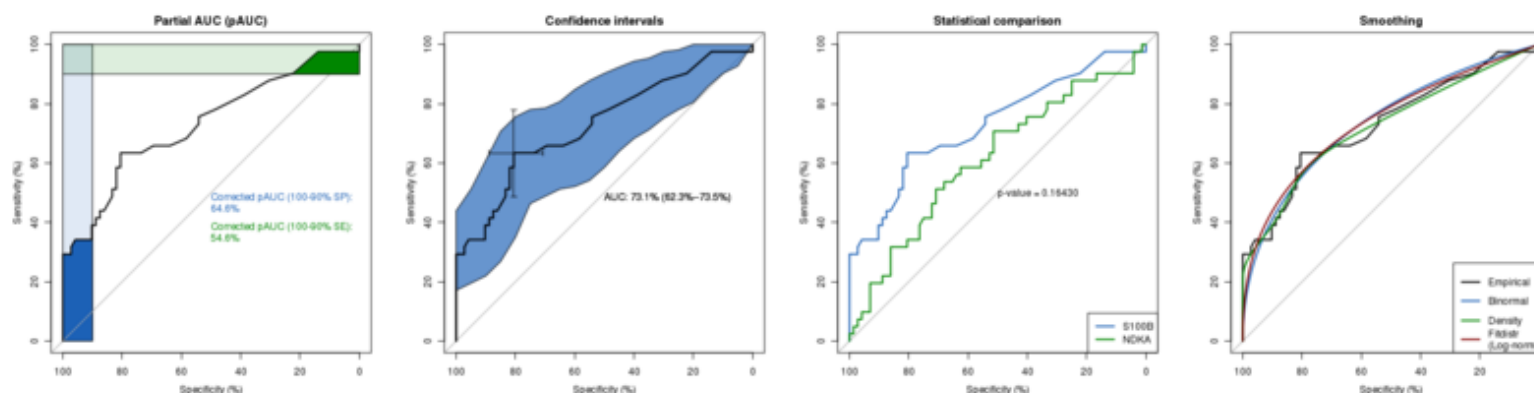
Abstract Comparing two proportions through the difference is a basic problem in statistics and has applications in many fields. More than twenty confidence intervals (Newcombe, 1998a,b) have been proposed. Most of them are approximate intervals with an asymptotic infimum coverage probability much less than the nominal level. In addition, large sample may be costly in practice. So exact optimal confidence intervals become critical for drawing valid statistical inference with accuracy and precision. Recently, Wang (2010, 2012) derived the exact smallest (optimal) one-sided $1 - \alpha$ confidence intervals for the difference of two paired or independent proportions. His intervals, however, are computer-intensive by nature. In this article, we provide an R package **ExactCIdiff** to implement the intervals when the sample size is not large. This would be the first available package in R to calculate the exact confidence intervals for the difference of proportions. Exact two-sided $1 - \alpha$ interval can be easily obtained by taking the intersection of the lower and upper one-sided $1 - \alpha/2$ intervals. Readers may jump to Examples 1 and 2 to obtain these intervals.

Introduction

The comparison of two proportions through the difference is one of the basic statistical problems. One-sided confidence intervals are of interest if the goal of a study is to show superiority (or inferiority), e.g., that a treatment is better than the control. If both limits are of interest, then two-sided intervals

pROC: display and analyze ROC curves in R and S+

pROC is a set of tools to visualize, smooth and compare receiver operating characteristic (ROC curves). (Partial) area under the curve (AUC) can be compared with statistical tests based on U-statistics or bootstrap. Confidence intervals can be computed for (p)AUC or ROC curves.



[More screenshots and examples...](#)

If you use pROC in published research, please cite the following paper:

Xavier Robin, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez and Markus Müller (2011). [pROC: an open-source package for R and S+ to analyze and compare ROC curves](#). *BMC Bioinformatics*, 12, p. 77. DOI: [10.1186/1471-2105-12-77](https://doi.org/10.1186/1471-2105-12-77).

Authors:

Xavier Robin, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez and Markus Müller

Contact:

[Xavier Robin](#)

License:

GPLv3

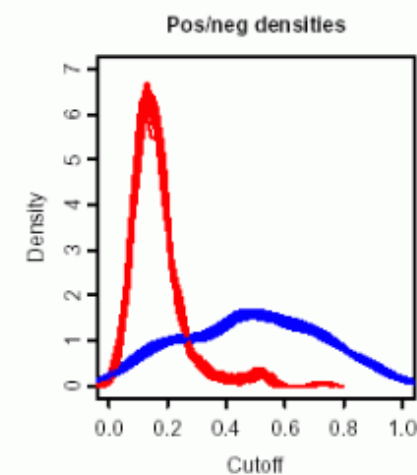
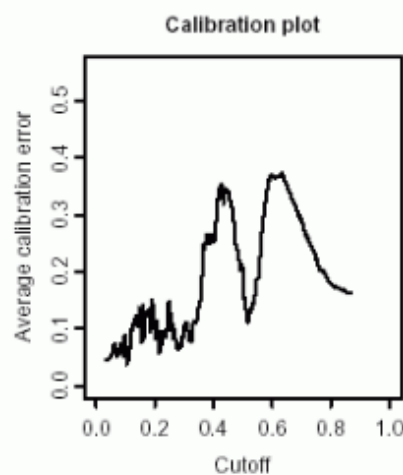
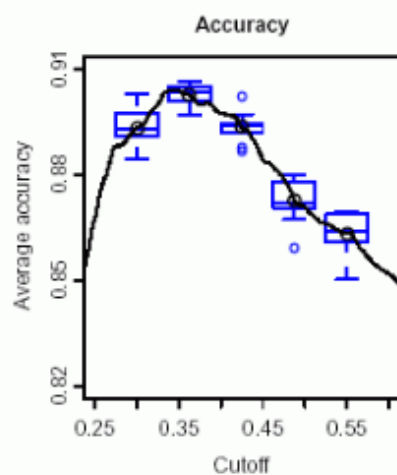
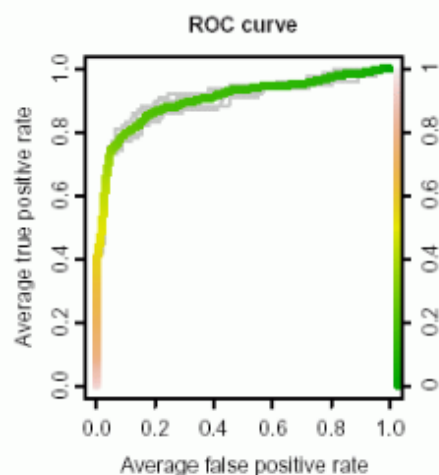
Find us on:



ROCR

prediction - performance - plot

[visualizing classifier performance in R, with only 3 commands]



Performance measures that ROCR knows:

Accuracy, error rate, true positive rate, false positive rate, true negative rate, false negative rate, sensitivity, specificity, recall, positive predictive value, negative predictive value, precision, fallout, miss, phi correlation coefficient, Matthews correlation coefficient, mutual information, chi square statistic, odds ratio, lift value, precision/recall F measure, ROC convex hull, area under the ROC curve, precision/recall break-even point, calibration error, mean cross-entropy, root mean squared error, SAR measure, expected cost, explicit cost.

ROCR features:

ROC curves, precision/recall plots, lift charts, cost curves, custom curves by freely selecting one performance measure for the x axis and one for the y axis, handling of data from cross-validation or bootstrapping, curve averaging (vertically, horizontally, or by threshold), standard error bars, box plots, curves that are color-coded by cutoff, printing threshold values on the curve, tight integration with R's plotting facilities (making it easy to adjust plots or to combine multiple plots), fully customizable, easy to use (only 3 commands).



<http://cran.r-project.org/web/views/ClinicalTrials.html>

<http://cran.r-project.org/web/views/Survival.html>

<http://cran.r-project.org/web/views/ExperimentalDesign.html>

<http://cran.r-project.org/web/views/Pharmacokinetics.html>

CRAN Task View: Survival Analysis

Maintainer: Arthur Allignol and Aurelien Latouche

Contact: arthur.allignol@uni-ulm.de

Version: 2014-01-19

Survival analysis, also called event history analysis in social science, or reliability analysis in engineering, deals with time until occurrence of an event of interest. However, this failure time may not be observed within the relevant time period, producing so-called censored observations.

This task view aims at presenting the useful R packages for the analysis of time to event data.

Please let the [maintainers](#) know if something is inaccurate or missing.

Standard Survival Analysis

Estimation of the Survival Distribution

- **Kaplan-Meier:** The `survfit` function from the `survival` package computes the Kaplan-Meier

CRAN packages:

- AdapEnetClass
- AER
- ahaz
- AIM
- asbio
- aster
- aster2
- BaSTA
- bayesSurv
- BayHaz
- BGPhazard
- Biograph
- BMA
- boot
- bpcp
- censReg
- changeLOS
- clinfun
- cmprsk
- coarseDataTools
- coin
- compeir
- complex.surv.dat.sim
- compound.Cox
- concreg
- condGEE
- CoxBoost
- coxme
- coxphf
- coxphw
- CoxRidge
- coxrobust
- CPE
- CPHshape
- CR
- crrSC
- crrstep
- currentSurvival
- DAAG
- dblcens
- DPPackage
- DTDA
- dynpred
- dynamSurv
- eha
- emplik
- Epi
- epiR
- etm
- exactRankTests
- fastcox
- FHtest
- fitdistrplus
- flexSurv
- frailtyHL
- frailtypack
- gamlss.cens
- gbm
- gems
- genSurv
- glmpath
- globalboosttest
- glrt
- gof
- gss
- ICE
- intcox
- interval
- ipdmeta
- ipred
- ipw
- jackknifeKME
- JM
- JMbayes
- jmec
- JMLSD
- joineR
- JPSurv
- kaps
- km.ci
- kmc
- kmconfband
- kmi
- KMSurv
- lava.tobit
- lbiassurv
- LearnBayes
- lmecc
- locfit
- logconcens
- LogicReg
- LogrankA
- logspline
- lpc
- lss
- MAMSE
- maxstat
- mboost
- MCMCglmm
- MCMCpack
- Mets
- mfpm
- mixAK
- mixPHM
- MLEcens
- MRSurv
- msm
- msSurv
- mstate
- muhaz
- multcomp
- mvna
- mvpart
- NADA
- NestedCohort
- NPHMC
- NPMLEcmprsk
- OISurv
- OrdFacReg
- OutlierDC
- pamr
- parfm
- party
- pec
- penalized
- peperr
- PermAlgo
- phmm
- plsRcox
- polspline
- powerSurvEpi
- prodlim
- psbcGroup
- pseudo
- quantreg
- randomForestSRC
- RandomSurvivalForest
- rankhazard
- relSurv
- rhosp
- riskRegression
- risksetROC
- rms
- RobustAFT
- ROct
- rpart
- rsig
- saws
- SemiCompRisks
- SemiMarkov
- SGL
- simexaft
- simMSM
- simPH
- SMIR
- SmoothHazard
- smoothHR
- smoothSurv
- SMPracticals
- spatstat
- spatsurv
- superpc
- Surv2sampleComp
- SurvAUC
- SurvBayes
- SurvC1
- Survexp.fr
- survey
- Survgini
- SurvIDINRI
- survival
- survivalBIV
- survivalROC
- survJamda
- survMisc
- survPresmooth
- survrec
- SurvRegCensCov
- survsim
- survSNP
- SvyNom
- TBSSurvival
- TestSurvRec
- timereg
- timeROC
- tlmec
- TPmsm
- tpr
- TraMineR
- TSHRC
- uniCox
- VGAM
- wtcrsk
- YPmodel



<http://cran.r-project.org/web/views/ClinicalTrials.html>

<http://cran.r-project.org/web/views/Survival.html>

<http://cran.r-project.org/web/views/ExperimentalDesign.html>

<http://cran.r-project.org/web/views/Pharmacokinetics.html>

Firefox

CRAN Task View: Design of Experiments (Do... +

cran.r-project.org/web/views/ExperimentalDesign.html

CRAN Task View: Design of Experiments (DoE) & Analysis of Experimental Data

Maintainer: Ulrike Groemping
Contact: groemping at bht-berlin.de
Version: 2013-03-20

This task view collects information on R packages for experimental design and analysis of data from experiments. Please feel free to suggest enhancements, and please send information on new packages or major package updates if you think they belong here. Contact details are given on my [Web page](#).

Experimental design is applied in many areas, and methods have been tailored to the needs of various fields. This task view starts out with a section on the most general packages, continues with specific sections on agricultural and industrial experimentation, computer experiments, and experimentation in the clinical trials contexts, and closes with a section on various special experimental design packages that have been developed for other specific purposes. Of course, the division into fields is not always clear-cut, and some packages from the more specialized sections can also be applied in general contexts. You may also notice that my own experience is mainly from industrial experimentation (in a broad sense), which may explain a somewhat biased view on things.

Experimental designs for general purposes

There are a few packages for creating and analyzing experimental designs for general purposes: First of all, the standard (generalized) linear model functions in the base package stats are of course very important for analyzing data from designed experiments (especially functions `lm()`, `aov()` and the methods and functions for the resulting linear model objects). These

CRAN packages:

- agricolae
- AlgDesign
- asd
- BatchExperiments
- BHH2
- blockTools
- BsMD
- conf.design
- crossdes
- dae
- desirability
- DiceDesign
- DiceEval
- DiceKriging
- DiceView
- displayHTS
- DoE.base
- DoE.wrapper
- DoseFinding
- dynaTree
- experiment
- FrF2
- FrF2.catlg128
- GAD
- granova
- gsbDesign
- gsDesign
- IdDesign
- lhs
- mixexp
- mkssd
- mxkssd
- odprism
- osDesign
- planor
- plgp
- qtlDesign
- qualityTools
- RcmdrPlugin.DoE
- rsm
- SensoMineR
- support.CEs
- TEQR
- tgp
- Vdgraph



Some R packages are real gems! Meet **gsDesign**, a package for deriving and describing group sequential designs, created by **Keaven Anderson** (Merck) and *REVOlution Computing* to optimize Merck's clinical trial process...

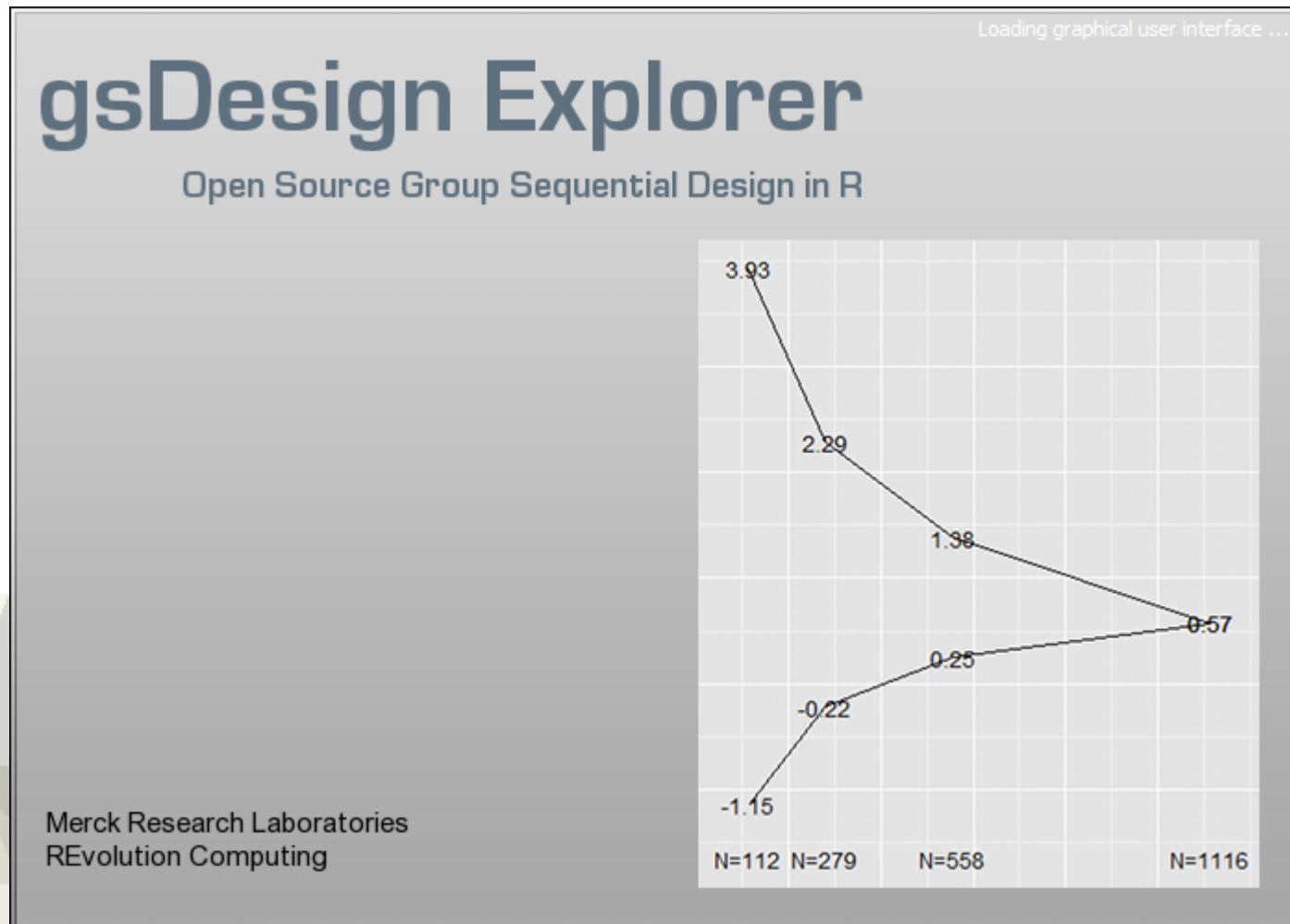
<http://www.amstat.org/sections/sbiop/webinars/2011/AndersonWebinar2-23-2011.pdf>

Group Sequential Design Basics
with Application Using the
gsDesign R package and its GUI

Keaven M. Anderson
Merck Research Laboratories
February 23, 2011

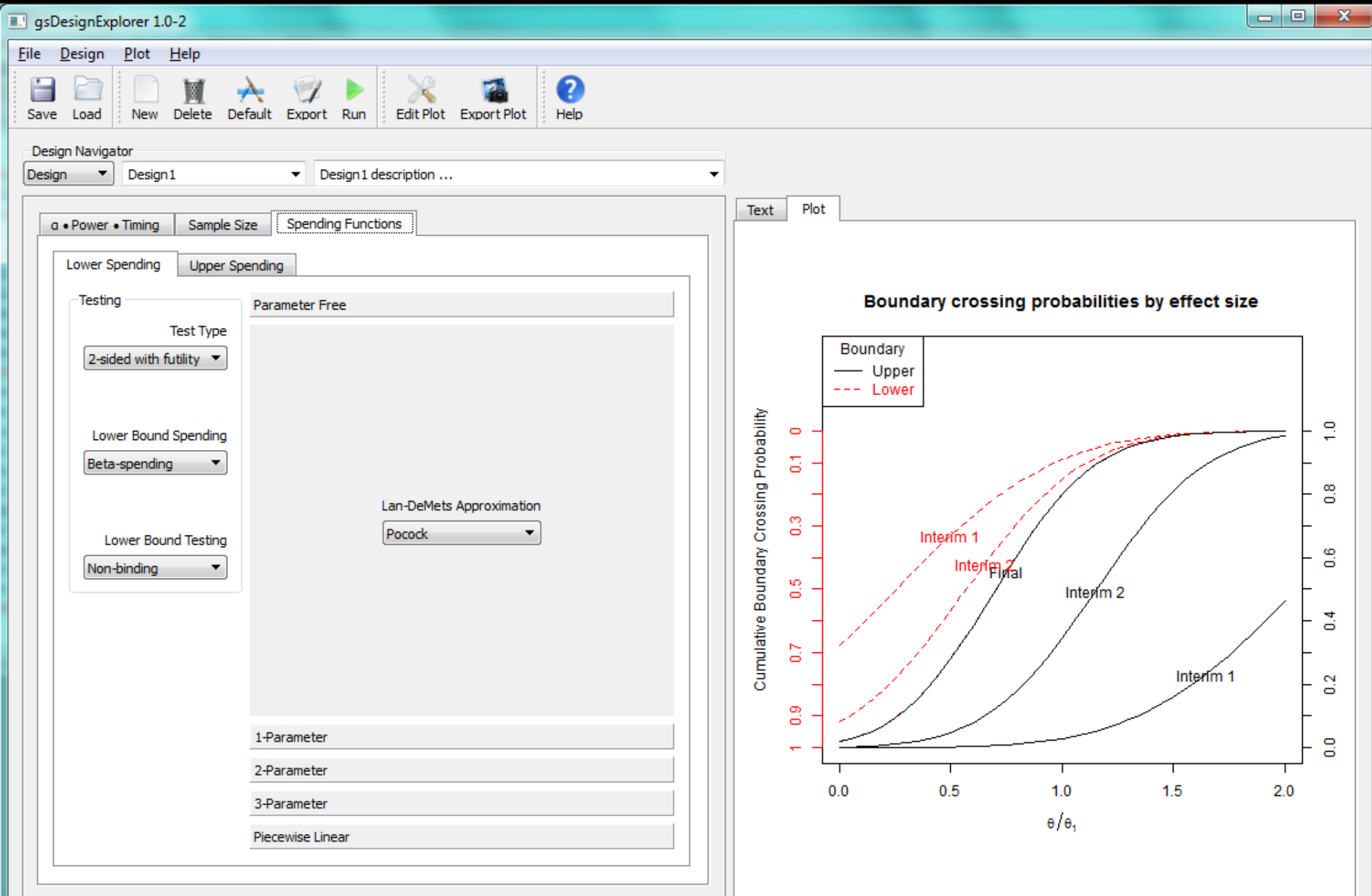


... and [gsDesignExplorer](#) - a nice GUI for gsDesign. Unfortunately it works only with 2.x version of R but, luckily, old "portable" versions of R are [available](#).





gsDesignExplorer





<http://cran.r-project.org/web/views/ClinicalTrials.html>

<http://cran.r-project.org/web/views/Survival.html>

<http://cran.r-project.org/web/views/ExperimentalDesign.html>

<http://cran.r-project.org/web/views/Pharmacokinetics.html>

CRAN Task View: Analysis of Pharmacokinetic Data

Maintainer: Suzette Blanchard
Contact: sblanchard at coh.org
Version: 2013-05-21

The primary goal of pharmacokinetic (PK) data analysis is to determine the relationship between the dosing regimen and the body's exposure to the drug as measured by the nonlinear concentration time curve or related summaries (e.g. the area under the curve). Base R contains `nls` which can be used to calculate nonlinear least-squares estimates of the parameters from a PK model. It returns an object of the class "nls" having methods `coef()`, `formula()`, `resid()`, `print()`, `summary()`, `AIC()`, `fitted()` and `vcov()`. Four packages are available in CRAN that directly aid in PK data analysis, including; the packages [PK](#), [PKfit](#), [nlmeODE](#) which incorporates [nlme](#) and [deSolve](#), and the package [PKtools](#). While [PK](#) provides basic pharmacokinetics functions which implement non-compartmental analysis methods, the latter three packages focus on modeling methods. In addition, this task view includes a discussion of a few related packages including the package [lattice](#) for trellis graphics, and the package [MASS](#) and the package [drc](#) that aid in the analysis of dose response curves.

[PK](#) is a basic pharmacokinetics package providing functions to aid in non-compartmental analysis (Gibaldi and Perrier, 1982) of PK data. This package includes a function, `AUC()`, to calculate area under the curve, and the area under the first moment curve and two new functions `auc.ssd()` and `eqv.ssd()` for estimating and comparing AUCs in the case of serial sampling. It also includes functions for half-life estimation for a biexponential model, `biexp()`, and a two phase linear regression, `lee()` as well as an associated plot function, `plot.halfLife()`.

The package [PKfit](#) is a nonlinear regression program designed for the analysis of pharmacokinetic data. This package

CRAN packages:

- deSolve
- drc
- lattice
- MASS
- nlme
- nlmeODE
- PK
- PKfit
- PKPDmodels
- PKtools



Tools on R for Dose-Response curves analysis

Chantal THORIN

2009 July 8th

UPSP 5304 : Physiopathologie Animale et Pharmacologie Fonctionnelle

ENV Nantes France



Analysis of **dose-Response** curves

- actimL Model function for the universal response surface approach (URSA) for the quantitative assessment of drug interaction
- anova.drc ANOVA for dose-response model fits
- AR.2 Asymptotic regression model
- AR.3 Asymptotic regression model
- baro5 The modified baro5 function
- BC.4 The Brain-Cousens hormesis models
- BC.5 The Brain-Cousens hormesis models
- bcl3 The Brain-Cousens hormesis models
- bcl4 The Brain-Cousens hormesis models
- boxcox.drc Transform-both-sides Box-Cox transformation
- braincousens The Brain-Cousens hormesis models
- cedergreen The Cedergreen-Ritz-Streibig model
- coef.drc Extract Model Coefficients
- comped Comparison of effective dose values
- compParm Comparison of parameters
- confint.drc Confidence Intervals for model parameters
- CRS.4a The Cedergreen-Ritz-Streibig model
- CRS.4b The Cedergreen-Ritz-Streibig model
- CRS.4c The Cedergreen-Ritz-Streibig model
- CRS.5a Cedergreen-Ritz-Streibig dose-reponse model for describing hormesis
- CRS.5b Cedergreen-Ritz-Streibig dose-reponse model for describing hormesis
- CRS.5c Cedergreen-Ritz-Streibig dose-reponse model for describing hormesis
- CRS.6 The Cedergreen-Ritz-Streibig model
- diagnostics Information on estimation
- drm Fitting dose-response models
- drmc Sets control arguments
- ED Estimating effective doses
- ED.drc Estimating effective doses
- ED.mrdrc Estimating effective doses
- estfun.drc Bread and meat for the sandwich
- EXD.2 Exponential decay model
- EXD.3 Exponential decay model
- fitted.drc Extract fitted values from model
- FPL.4 Fractional polynomial-logistic dose-response models
- fplogistic Fractional polynomial-logistic dose-response models
- G.2 Mean function for the Gompertz dose-response or growth curve
- G.3 Mean function for the Gompertz dose-response or growth curve
- G.3u Mean function for the Gompertz dose-response or growth curve
- G.4 Mean function for the Gompertz dose-response or growth curve
- genBliss Model function for the universal response surface approach (URSA) for the quantitative assessment of drug interaction
- genBliss2 Model function for the universal response surface approach (URSA) for the quantitative assessment of drug interaction
- genLoewe Model function for the universal response surface approach (URSA) for the quantitative assessment of drug interaction
- genLoewe2 Model function for the universal response surface approach (URSA) for the quantitative assessment of drug interaction
- genursa Model function for the universal response surface approach (URSA) for the quantitative assessment of drug interaction
- getInitial Showing starting values used
- getMeanFunctions Display available dose-response models
- gompertz Mean function for the Gompertz dose-response or growth curve
- gompertzd The derivative of the Gompertz function
- gompGrowth.1 Gompertz growth models
- gompGrowth.2 Gompertz growth models
- gompGrowth.3 Gompertz growth models
- iceLoewe.1 Model function for the universal response surface approach (URSA) for the quantitative assessment of drug interaction
- iceLoewe2.1 Model function for the universal response surface approach (URSA) for the quantitative assessment of drug interaction
- isobole Creating isobolograms
- L.3 The logistic model
- L.4 The logistic model
- L.5 The logistic model
- I2 The two-parameter log-logistic function
- I3 The three-parameter log-logistic function
- I3u The three-parameter log-logistic function
- I4 The four-parameter log-logistic function
- I5 The five-parameter log-logistic function
- lin.test Lack-of-fit test for the mean structure based on cumulated residuals
- LL.2 The two-parameter log-logistic function
- LL.3 The three-parameter log-logistic function
- LL.3u The three-parameter log-logistic function
- LL.4 The four-parameter log-logistic function
- LL.5 The five-parameter log-logistic function
- LL2.2 The two-parameter log-logistic function
- LL2.3 The three-parameter log-logistic function
- LL2.3u The three-parameter log-logistic function
- LL2.4 The four-parameter log-logistic function
- LL2.5 The five-parameter log-logistic function
- llogistic The log-logistic function
- llogistic2 The log-logistic function
- LN.2 Log-normal dose-response model
- LN.3 Log-normal dose-response model
- LN.3u Log-normal dose-response model
- LN.4 Log-normal dose-response model
- Inormal Log-normal dose-response model
- logistic The logistic model
- logLik.drc Extracting the log likelihood
- maED Estimation of ED values using model-averaging
- MAX Maximum mean response
- mixture Fitting binary mixture models
- mI3a The Cedergreen-Ritz-Streibig model
- mI3b The Cedergreen-Ritz-Streibig model
- mI3c The Cedergreen-Ritz-Streibig model
- mI4a Cedergreen-Ritz-Streibig dose-reponse model for describing hormesis
- mI4b Cedergreen-Ritz-Streibig dose-reponse model for describing hormesis
- mI4c Cedergreen-Ritz-Streibig dose-reponse model for describing hormesis
- MM.2 Michaelis-Menten model
- MM.3 Michaelis-Menten model
- modelFit Assessing the model fit



Analysis of **d**ose-**R**esponse **c**urves [cont'd]

- `mr.test` Mizon-Richard test for dose-response models
- `mrdrm` Model-robust dose-response modelling
- `mselect` Model selection
- `NEC` Dose-response model for estimation of no effect concentration (NEC).
- `NEC.2` Dose-response model for estimation of no effect concentration (NEC).
- `NEC.3` Dose-response model for estimation of no effect concentration (NEC).
- `NEC.4` Dose-response model for estimation of no effect concentration (NEC).
- `neill.test` Neill's lack-of-fit test for dose-response models
- `plot.drc` Plotting fitted curves for a 'drc' or 'mrdrm' object
- `plot.mrdrm` Plotting fitted curves for a 'drc' or 'mrdrm' object
- `PR` Expected or predicted response
- `predict.drc` Prediction
- `predict.mrdrm` Prediction
- `print.drc` Printing key features
- `print.mrdrm` Printing key features
- `print.summary.drc` Printing summary of non-linear model fits
- `rdrm` Simulating a dose-response curve
- `relpot` Comparing selectivity indices across curves
- `residuals.drc` Extracting residuals from model
- `ryegrass` Effect of ferulic acid on growth of ryegrass
- `S.capricornutum` Effect of cadmium on growth of green alga
- `searchdrc` Searching through a range of initial parameter values to obtain convergence
- `secalonic` Root length measurements
- `SI` Comparing selectivity indices across curves
- `simDR` Simulating ED values under various scenarios
- `summary.drc` Summarising non-linear model fits
- `twophase` Two-phase dose-response model
- `ucedergreen` The Cedergreen-Ritz-Streibig model
- `UCRS.4a` The Cedergreen-Ritz-Streibig model
- `UCRS.4b` The Cedergreen-Ritz-Streibig model
- `UCRS.4c` The Cedergreen-Ritz-Streibig model
- `UCRS.5a` Cedergreen-Ritz-Streibig dose-reponse model for describing hormesis
- `UCRS.5b` Cedergreen-Ritz-Streibig dose-reponse model for describing hormesis
- `UCRS.5c` Cedergreen-Ritz-Streibig dose-reponse model for describing hormesis
- `uml3a` The Cedergreen-Ritz-Streibig model
- `uml3b` The Cedergreen-Ritz-Streibig model
- `uml3c` The Cedergreen-Ritz-Streibig model
- `uml4a` Cedergreen-Ritz-Streibig dose-reponse model for describing hormesis
- `uml4b` Cedergreen-Ritz-Streibig dose-reponse model for describing hormesis
- `uml4c` Cedergreen-Ritz-Streibig dose-reponse model for describing hormesis
- `update.drc` Updating and re-fitting a model
- `ursa` Model function for the universal response surface approach (URSA) for the quantitative assessment of drug interaction
- `vcov.drc` Calculating variance-covariance matrix for objects of class 'drc'
- `W1.2` The two-parameter Weibull functions
- `W1.3` The three-parameter Weibull functions
- `W1.3u` The three-parameter Weibull functions
- `W1.4` The four-parameter Weibull functions
- `w2` The two-parameter Weibull functions
- `W2.2` The two-parameter Weibull functions
- `W2.3` The three-parameter Weibull functions
- `W2.3u` The three-parameter Weibull functions
- `W2.4` The four-parameter Weibull functions
- `W2x.3` The three-parameter Weibull functions
- `W2x.4` The four-parameter Weibull functions
- `w3` The three-parameter Weibull functions
- `w4` The four-parameter Weibull functions
- `weibull1` Weibull model functions
- `weibull2` Weibull model functions
- `weibull2x` Weibull model functions
- `yieldLoss` Calculating yield loss parameters

Analysis of **d**ose-**R**esponse **c**urves [cont'd]

DraftDrcManual.pdf

www.bioassay.dk/index-filer/start/DraftDrcManual.pdf

Search

Page: 1 of 51

110%

Dose response curves and other nonlinear curves in Weed Science and Ecotoxicology with the add-on package *drc* in **R**

Christian Ritz & Jens C. Streibig

March 5, 2012

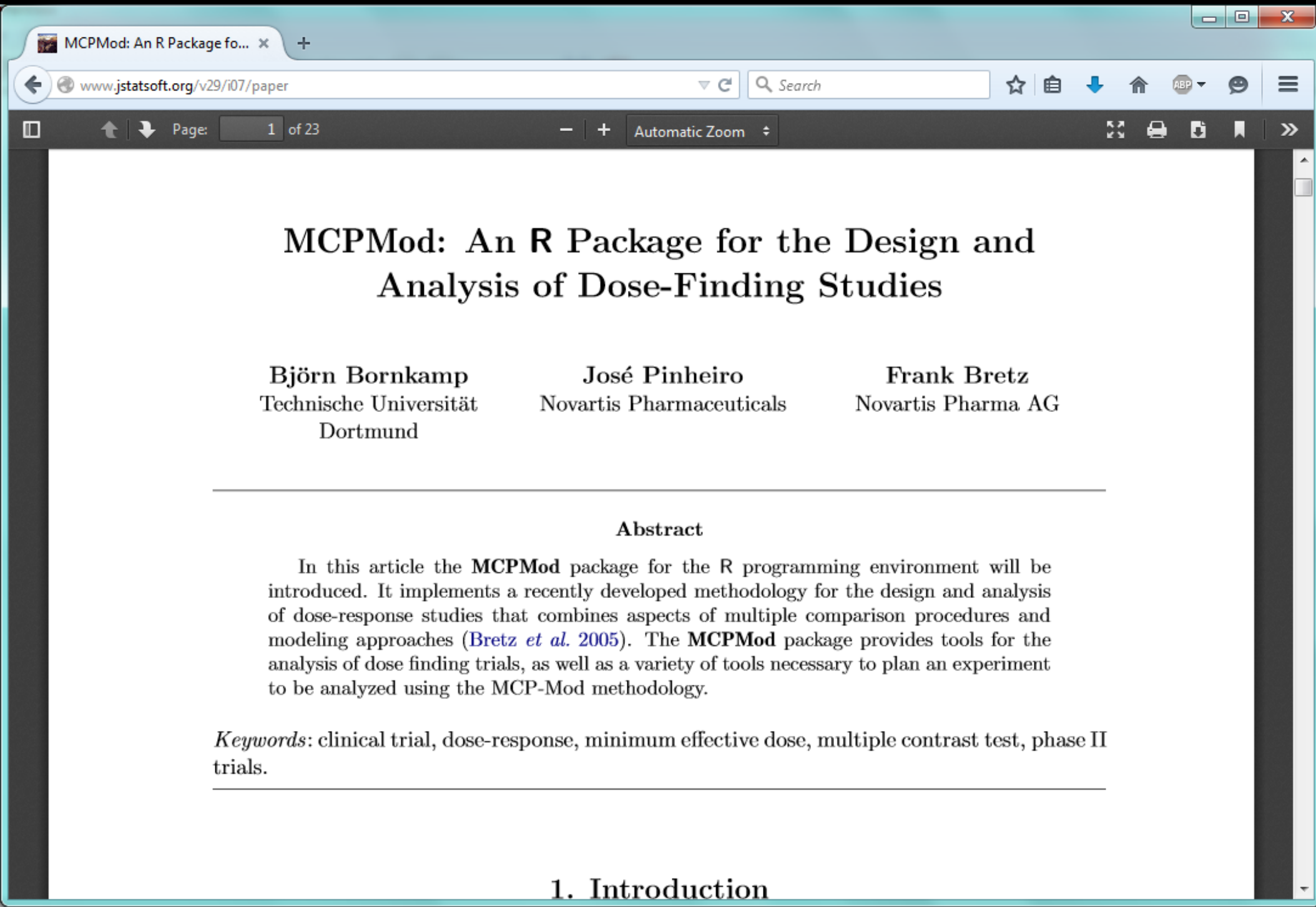
Contents

1 Introduction	2
2 Install <i>drc</i> load it and and getting help	4
3 The <i>drc</i> engine	5
4 Experiments with replication	9
5 Comparing response curves	14
6 Fitting multiple dose-response curves and model reduction	15
6.1 Taking charge of the upper and lower limits	25
7 Hormetic curves	31
8 Exponential and Michaelis-Mentens curves	33
8.1 Exponential curves	33
8.2 Michaelis-Mentens curves	36
9 Non linear regression with binomial endpoint	39
9.1 An example with an upper limit	41
9.2 Multiple dose-response curves	41
10 Statistical considerations	47
10.1 Heterogeneous variance	47

<http://www.bioassay.dk/index-filer/start/DraftDrcManual.pdf>



They use **R**



MCPMod: An R Package for the Design and Analysis of Dose-Finding Studies

Björn Bornkamp
Technische Universität
Dortmund

José Pinheiro
Novartis Pharmaceuticals

Frank Bretz
Novartis Pharma AG

Abstract

In this article the **MCPMod** package for the R programming environment will be introduced. It implements a recently developed methodology for the design and analysis of dose-response studies that combines aspects of multiple comparison procedures and modeling approaches (Bretz *et al.* 2005). The **MCPMod** package provides tools for the analysis of dose finding trials, as well as a variety of tools necessary to plan an experiment to be analyzed using the MCP-Mod methodology.

Keywords: clinical trial, dose-response, minimum effective dose, multiple contrast test, phase II trials.

1. Introduction



Advanced PK modeling with mixed effects

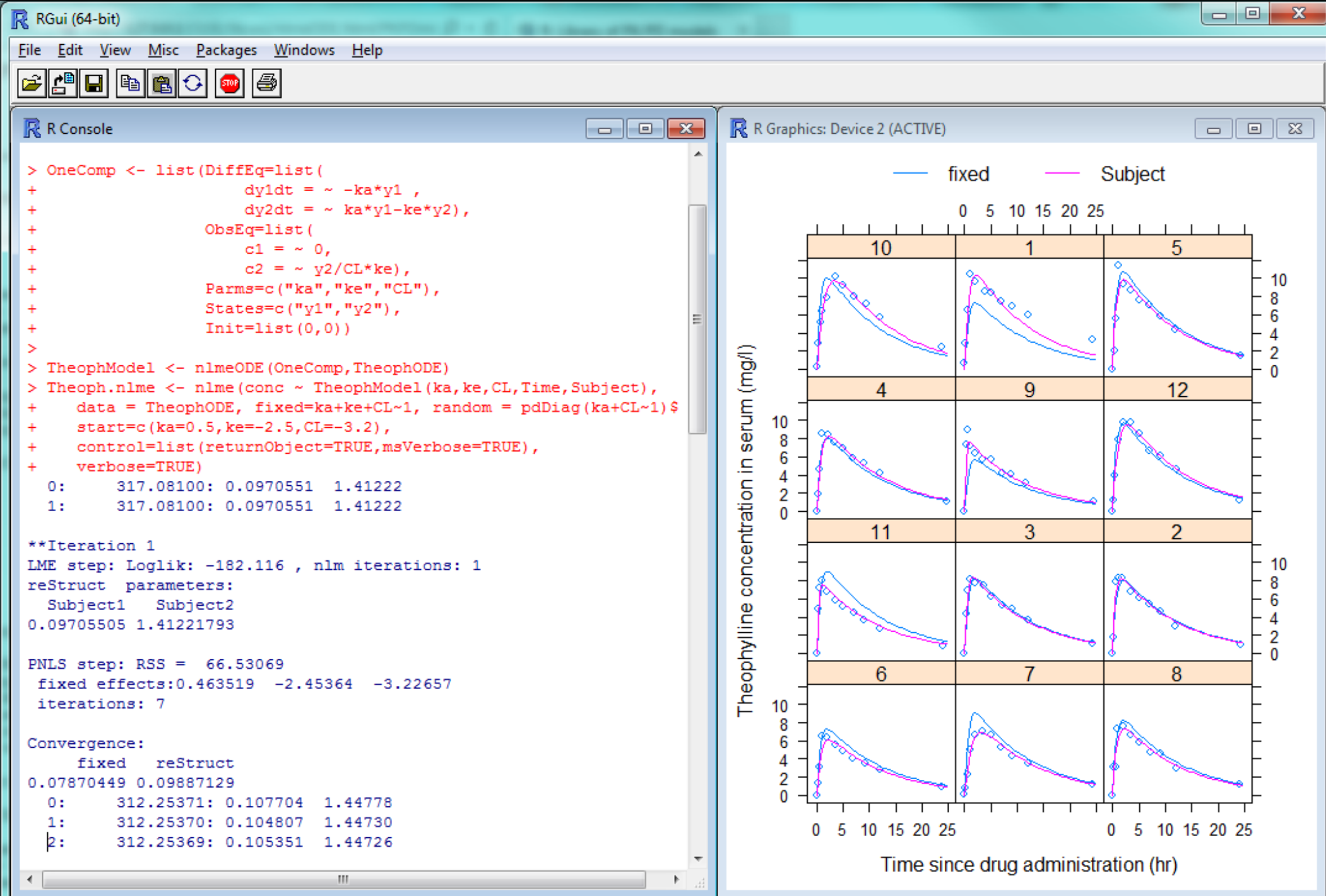
[nlmeODE](#) is a powerful combination of the famous [nlme](#) package with [deSolve](#) (an ordinary differential equations solver) for non-linear, mixed-effects modelling, designed especially for pharmaceutical research.

The package also includes several PK examples such as one- and two-compartment models with multiple doses and infusions.

- Pharmacokinetics of Theophylline
- Pharmacokinetics of Indomethacine
- Absorption model with estimation of time/rate of infusion
- Simulation and simultaneous estimation of PK/PD data
- Minimal Model of Glucose and Insulin
- Minimal Model of Glucose using observed insulin as forcing function

<http://www.inside-r.org/packages/cran/nlmeODE/docs/PKPDmodels>

Advanced PK modeling with mixed effects





R + ADMB (automatic differentiation!)

Welcome — ADMB Project

www.admb-project.org

Search

Site Map Accessibility Contact

Home Downloads Documentation Examples Courses Tools Developers Users

Log in

You are here: Home

- [FAQ](#)
- [Download](#)
- [Documentation](#)
- [API-Function reference](#)
- [User Community](#)
- [Users mailing lists](#)
- [Courses](#)
- [Repository](#)
- [Issue tracker](#)
- [Donate](#)

Welcome

AD Model Builder, or ADMB, is a powerful software package for the development of state-of-the-art nonlinear statistical models. ADMB is built around the AUTODIF Library, a C++ language extension which implements reverse mode automatic differentiation. A closely related software package, ADMB-RE, implements random effects in nonlinear models. ADMB was created by [David Fournier](#) and now continues to be developed by the ADMB Project, a creation of the non-profit [ADMB Foundation](#). ADMB is free, open source, and available for Windows, Linux, MacOS, and Sun/SPARC.

How to cite ADMB:

[Fournier, D.A., H.J. Skaug, J. Ancheta, J. Ianelli, A. Magnusson, M.N. Maunder, A. Nielsen, and J. Sibert. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optim. Methods Softw. 27:233-249.](#)

[ADMB project background](#)

News

- ADMB-11.4 released
May 29, 2015
- ADMB-11.3 released
May 22, 2015
- ADMB Users and Developers lists moved to Google Groups
Apr 01, 2015
- Issue tracker (<http://www.admb-project.org/redmine/>) has been retired.
Jan 15, 2015

More news...



R + ADMB (automatic differentiation!)

R — ADMB Project x +

www.admb-project.org/tools/r

You are here: [Home](#) > [Tools](#) > [R](#)

R

Working with ADMB and R

- [ADMB2R](#)
- ADMB output that can be read directly into R**
- [Emacs support for R and ADMB](#)
- Efficient platform for working with R and ADMB**
- [FLR](#)
- Fisheries Library in R**
- [PBSadmb](#)
- Organize and run ADMB model from R**
- [Interfacing R and ADMB](#)
- Document by Mark Maunder, written for the La Jolla meeting in March 2010**
- [Reading report files using R](#)
- [r2admb](#)
- An interface to the AD Model Builder system, simplifying and streamlining the use of ADMB from within R by Ben Bolker**
- [scape](#)
- R package for plotting fisheries stock assessment data and model fit**
- [scapeMCMC](#)
- R package for plotting multipanel MCMC diagnostic plots**
- [Specifications for R-ADMB interface \(creating dat and pin files\) with example](#)

News

- ADMB-11.4 released**
May 29, 2015
- ADMB-11.3 released**
May 22, 2015
- ADMB Users and Developers lists moved to Google Groups**
Apr 01, 2015
- Issue tracker**
(<http://www.admb-project.org/redmine/>) has been retired.
Jan 15, 2015

[More news...](#)

Navigation Menu:

- ADMB-IDE
- Editors
- flex
- gcc
- gdb
- R**
- ADMB2R
- Emacs support for R and ADMB
- FLR
- PBSadmb
- Interfacing R and ADMB
- Reading report files using R
- r2admb
- scape
- scapeMCMC
- Specifications for R-ADMB interface (creating dat and pin



Genomic data analysis



Search:

[Home](#) [Install](#) [Help](#) [Developers](#) [About](#)

BioC2015

Join us for morning talks from distinguished speakers and community members, afternoon workshops to hone your skills, and poster sessions and social activities to get to know members of the Bioconductor community at our [Annual Conference](#), July 20 (Developer Day), 21 and 22 in Seattle, WA.

About Bioconductor

Bioconductor provides tools for the analysis and comprehension of high-throughput genomic data. Bioconductor uses the R statistical programming language, and is open source and open development. It has two releases each year, [1024 software packages](#), and an active user community. Bioconductor is also available as an [AMI](#) (Amazon Machine Image) and a series of [Docker](#) images.

Install »

Get started with *Bioconductor*

- [Install Bioconductor](#)
- [Explore packages](#)
- [Get support](#)
- [Latest newsletter](#)
- [Follow us on twitter](#)
- [Install R](#)

Learn »

Master *Bioconductor* tools

- [Courses](#)
- [Support site](#)
- [Package vignettes](#)
- [Literature citations](#)
- [Common work flows](#)
- [FAQ](#)
- [Community resources](#)
- [Videos](#)

Use »

Create bioinformatic solutions with *Bioconductor*

- [Software](#), [Annotation](#), and [Experiment](#) packages
- [Amazon Machine Image](#)
- [Latest release announcement](#)
- [Support site](#)

Develop »

Contribute to *Bioconductor*

- [Use Bioc 'devel'](#)
- 'Devel' [Software](#), [Annotation](#) and [Experiment](#) packages
- [Package guidelines](#)
- [New package submission](#)
- [Developer resources](#)
- [Build reports](#)



Log on



Journals

Gateways



Genome Biology

IMPACT
FACTOR
10.5

Search for

[Advanced search](#)

Home

Articles

Authors

Reviewers

About this journal

My Genome Biology

Top

Abstract

Rationale

Basic usage

Plotting trac...

Genomic overv...

Specialized p...

Biological ex...

Low-level gra...

Materials and
methods

Discussion

Abbreviations

Competing
interests

Authors'
contributions

Acknowledgements

References

Software

Highly accessed

Open Access

ggbio: an R package for extending the grammar of graphics for genomic data

Tengfei Yin¹, Dianne Cook² and Michael Lawrence³*

* Corresponding author: Michael Lawrence lawrence.michael@gene.com [Author Affiliations](#)

¹ Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA 50011, USA

² Department of Statistics, Iowa State University, Ames, IA 50011, USA

³ Department of Bioinformatics, Genentech, 1 Dna Way South San Francisco, CA 94080, USA

For all author emails, please [log on](#).

Genome Biology 2012, **13**:R77 doi:10.1186/gb-2012-13-8-r77

The electronic version of this article is the complete one and can be found online at:
<http://genomebiology.com/content/13/8/R77>

Received: 8 June 2012

Revisions received: 17 July 2012

Accepted: 31 August 2012

Published: 31 August 2012

© 2012 Yin et al.; licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Genome Biology

Volume 13

Issue 8

Viewing options

[Abstract](#)

Full text

[PDF \(886KB\)](#)

[Additional files](#)

Associated material

[PubMed record](#)

[Article metrics](#)

[Readers' comments](#)

Related literature

[Cited by](#)

[Google blog search](#)

Other articles by authors

▶ [on Google Scholar](#)

▶ [on PubMed](#)

Related articles/pages

[on Google](#)

[on Google Scholar](#)

[on PubMed](#)

Tools

[Download references](#)

[Download XML](#)

[Order reprints](#)



The screenshot shows a web browser window with the address bar containing 'https://sites.google.com/site/therepiproject/'. The page title is 'The R-epi project'. The main content area features a large heading 'Home' and a welcome message: 'Welcome to The R-epi project!'. Below this, a paragraph describes the project's resources for infectious disease epidemiology using R software, including R packages, documentation, cross-packages tutorials, and forums of discussion. A yellow highlight box contains a 'Highlight news' section with two items: 'Hackout 2: GRINDER is over!' and 'outbreaker 1.1-4 has been released!'. A left sidebar provides navigation links for Home, Contributors, Forum & contact, News, and Other resources (including About the R-epi project, EpiJson, Misc links, Other R packages, and Other software). A bottom right corner contains a 'Przetłumacz' button.

The R-epi project

Navigation

- Home
- Contributors
- Forum & contact
- News
- Other resources
 - About the R-epi project
 - EpiJson: the unified epi format
 - Misc links
 - Other R packages
 - Other software
- R packages
 - EpiEstim
 - epinet
 - episerve
 - expoTree
 - outbreaker
 - OutbreakTools
 - rsatscan
 - seedy
 - surveillance
 - TreePar
 - TreeSim
 - vetsyn

Home

Welcome to The R-epi project!

The *epi project* lists resources for *infectious disease epidemiology* using the R software. These resources include an overview of R packages, alongside dedicated documentation, cross-packages tutorials, and links to relevant forums of discussion.

Highlight news

Hackout 2: GRINDER is over! Many news exciting packages for outbreak analysis are coming! Click [here](#) to see what happened more information about the event.

outbreaker 1.1-4 has been released! This version will be significantly faster than previous versions, and scale better too, thanks to a new local likelihood evaluation.



R for laboratory diagnostics

RGui (64-bit) - [Data Editor]

	Measurement	ConventionalUnit	Conversion	SIUnit
1	Acetaminophen	g/mL	6.62	mol/L
2	Acetoacetic acid	mg/dL	0.098	mmol/L
3	Acetone	mg/dL	0.172	mmol/L
4	Alanine	mg/dL	112.2	mol/L
5	Albumin	g/dL	10	g/L
6	Aldosterone	ng/dL	0.0277	nmol/L
7	Aluminum			
8	Aminobutyric acid			
9	Amitriptyline			
10	Ammonia (as NH3)			
11	Androstenedione			
12	Angiotensin I			
13	Angiotensin II			
14	Anion gap			
15	Antidiuretic hormone			
16	Antithrombin III			
17	alpha-Antitrypsin			
18	Apolipoprotein A			
19	Apolipoprotein B			
20	Arginine			
21	Asparagine			
22	Bicarbonate			
23	Bilirubin			
24	Bromide			
25	C-peptide			
26	C1 esterase inhibitor			
27	C3 complement			
28	C4 complement			
29	Calcitonin			

RGui (64-bit) - [R Console]

```
> ConvertMedUnits( 27.5, "Creatin", to="US")
Error in ConvertMedUnits(27.5, "Creatin", to = "US") :
  More than one matching row. Please use 'exact=TRUE'
  and supply one of these matching strings:
      "Creatine"
      "Creatinine"
      "Creatinine clearance"
> ConvertMedUnits( 27.5, "Creatinine", to="US", exact=TRUE)
[1] 0.311086
> GlucoseSI = c(5, 5.4, 5, 5.1, 5.6, 5.1, 4.9, 5.2, 5.5) # in SI Units
> GlucoseUS = ConvertMedUnits( GlucoseSI, "Glucose", to="US" )
> cbind(GlucoseSI,GlucoseUS)
      GlucoseSI GlucoseUS
[1,]         5.0  90.09009
[2,]         5.4  97.29730
[3,]         5.0  90.09009
[4,]         5.1  91.89189
[5,]         5.6 100.90090
[6,]         5.1  91.89189
[7,]         4.9  88.28829
[8,]         5.2  93.69369
[9,]         5.5  99.09910
>
```



Process DICOM data with **oro.dicom**

Firefox

CMBX12 - dicom.pdf

cran.r-project.org/web/packages/oro.dicom/vignettes/dicom.pdf

Google

Page: 1 of 14 Automatic Zoom

Working with the DICOM Data Standard in R

Brandon Whitcher
Pfizer Worldwide R&D

Volker J. Schmid
Ludwig-Maximilians Universität München

Andrew Thornton
Cardiff University

Abstract

The package **oro.dicom** facilitates the interaction with and manipulation of medical imaging data that conform to the DICOM standard. DICOM data, from a single file or single directory or directory tree, may be uploaded into R using basic data structures: a data frame for the header information and a matrix for the image data. A list structure is used to organize multiple DICOM files. The conversion from DICOM to ANALYZE/NIFTI is straightforward using the capabilities of **oro.dicom** and **oro.nifti**.



Process DICOM data with **oro.dicom**

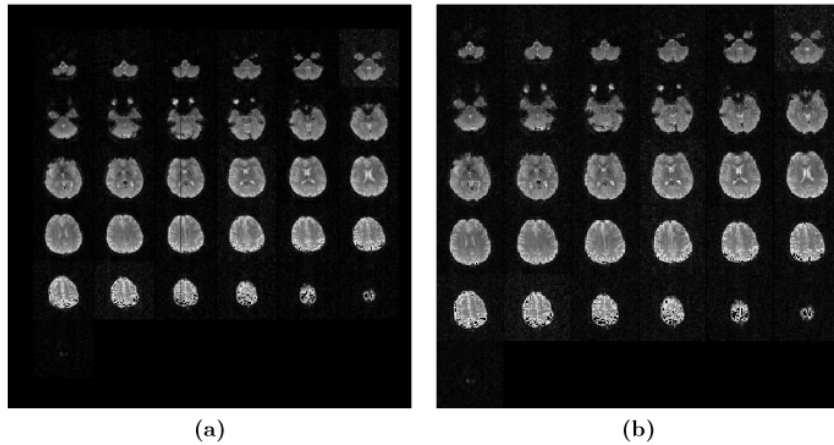


Figure 3: (a) Single MOSAIC image as read in from `readDICOMFile`. (b) Lightbox display of three-dimensional array of images after processing via `create3D`.

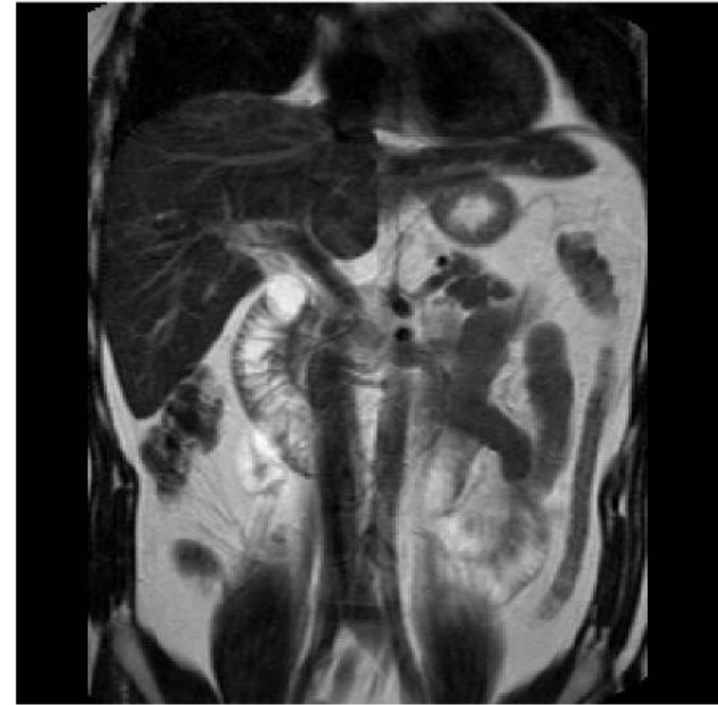


Figure 2: Coronal slice of the abdomen viewed in *neurological* convention (left is right and right is left).

```
R> tail(abdo$hdr)
```

group	element	name	code	length	value	sequence
79	0028	0101	BitsStored	US	2	12
80	0028	0102	HighBit	US	2	11
81	0028	0103	PixelRepresentation	US	2	0
82	0028	1050	WindowCenter	DS	4	530
83	0028	1051	WindowWidth	DS	4	1052
84	7FE0	0010	PixelData	OB	131072	PixelData



Process NIfTI data with **oro.nifti**

Firefox

CMBX12 - nifti.pdf

cran.r-project.org/web/packages/oro.nifti/vignettes/nifti.pdf

Page: 1 of 18 Automatic Zoom

Working with the NIfTI Data Standard in R

Brandon Whitcher
Mango Solutions

Volker J. Schmid
Ludwig-Maximilians Universität München

Andrew Thornton
Cardiff University

Abstract

The package **oro.nifti** facilitates the interaction with and manipulation of medical imaging data that conform to the ANALYZE, NIfTI and AFNI formats. The S4 class framework is used to develop basic ANALYZE and NIfTI classes, where NIfTI extensions may be used to extend the fixed-byte NIfTI header. One example of this, that has been implemented, is an XML-based “audit trail” tracking the history of operations applied to a data set. The conversion from DICOM to ANALYZE/NIfTI is straightforward using the




The Origin Of Things - RForge

Firefox

R-Forge: Software Map

https://r-forge.r-project.org/softwaremap/trove



Home

Software Map

Tag cloud | Project Tree | Project List

Project tree

Topic

- Bayesian Statistics (53 projects)
- Bioinformatics (148 projects)
- Biostatistics & Medical Statistics (68 projects)
- Cheminformatics (17 projects)
- Cluster Analysis (41 projects)
- Computational Physics (12 projects)
- Connectivity (20 projects)
- Database (27 projects)
- Datasets (28 projects)
- Design of Experiments & Analysis of Experiments (10 projects)
- Econometrics (63 projects)
- Education (27 projects)
- Environmetrics (44 projects)
- Finance (69 projects)
- Genetics (67 projects)
- Graphical Models (9 projects)
- Graphical User Interface (21 projects)
- Graphics (77 projects)
- High Performance Computing (28 projects)
- Machine Learning (71 projects)
- Marketing & Business Analytics (1 projects)
- Multivariate Statistics (138 projects)
- Numerical Analysis (18 projects)
- Optimization (56 projects)
- Other/Nonlisted Topic (75 projects)
- Regression Models (80 projects)
- Robust Statistics (22 projects)
- Social Sciences (76 projects)
- Software Development (54 projects)
- Spatial Data & Statistics (133 projects)
- Text (37 projects)
- Time Series (62 projects)

Firefox

R-Forge: Software Map

https://r-forge.r-project.org/softwaremap/full_list.php

Google

ADRminer - Adverse Drug event Reporting systems miner: An R package for the automated generation of drug safety signals from spontaneous reporting databases.

This project has not yet categorized itself in the [Trove Software Map](#)

Register Date: 2013-03-07 15:55

AFLP - A package for the normalisation on classification of AFLP (Amplification Fragment Length Polymorphism) data.

- Development Status : 4 - Beta
- Environment : Console (Text Based)
- Intended Audience : End Users/Desktop
- License : OSI Approved : GNU General Public License (GPL)
- Natural Language : English
- Operating System : OS Independent
- Programming Language : R
- Topic : Genetics : Population Genetics

Register Date: 2011-03-02 11:24

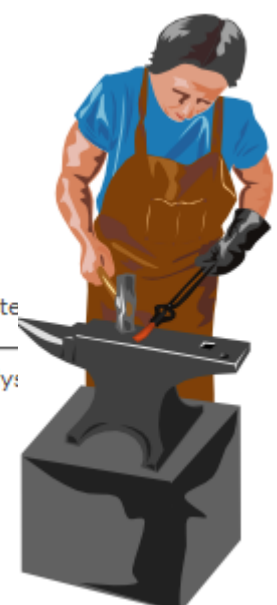
AICTS I - Unit root and cointegration tests encountered in applied econometric analysis.

- Development Status : 5 - Production/Stable
- Environment : Console (Text Based)
- Intended Audience : Developers
- Intended Audience : End Users/Desktop
- Intended Audience : Other Audience
- License : OSI Approved : GNU General Public License (GPL)
- Natural Language : English
- Operating System : OS Independent
- Programming Language : R
- Topic : Econometrics : Time Series Modelling
- Topic : Finance : Time Series
- Topic : Time Series

Register Date:

AICTS II - VAR, SVAR, VECM and SVECM models: Estimation, prediction, impulse response analysis: variance decomposition, diagnostic testing.

- Development Status : 5 - Production/Stable
- Environment : Console (Text Based)
- Intended Audience : Developers
- Intended Audience : End Users/Desktop
- Intended Audience : Other Audience
- License : OSI Approved : GNU General Public License (GPL)
- Natural Language : English





The Origin Of Things - RForge

Welcome to the RForge

User: Password: Login

RForge strives to provide a collaborative environment for R package developers. The ultimate goal is to offer SourceForge-like services (such as SVN repository, place for documentation, downloads, mailing lists, bugzilla, wiki etc.) without the annoying look and feel but with additional features specific to R package development, such as `make check` on-commit, nightly builds of packages, testing on various platforms and full CRAN-like repository access. The focus is on R-specific features that are not offered by SourceForge or GForge. It is complementary to sites like GitHub with which it can integrate as R package back-end.

Current projects

Name	Title	Rev	Last SVN change
Acinonyx	iPlots eXtreme: next-generation interactive graphics	169	2013-04-08 21:30:34
actogram	Single and double-plotted conditional actograms	1	2010-11-07 12:42:39
affinity	Functions and datasets for Web-mining social-based related knowledge	1	2008-08-01 08:38:48
ALA4R	Atlas of Living Australia (ALA) data and resources in R	7	2014-01-04 21:09:11
animation	The animation package		
AntBioR			
Arduino	Arduino projects	4	2010-02-17 11:09:29
audio	Audio Interface for R	32	2011-09-04 10:35:37
background	Background processing and asynchronous callbacks	3	2014-03-07 15:14:36
base64enc	Tools for base64 encoding	11	2014-06-26 09:47:42
brew	Mixing text and R code output for report generation	27	2011-03-11 11:23:41
CADStat	Provides a GUI to several statistical methods for making biological inferences	241	2010-03-25 07:03:36
Cairo	Graphics device using cairographics library for creating bitmap images on the fly	130	2012-11-05 15:02:06
canvas	R graphics device for the HTML canvas element	24	2010-04-26 16:50:19






The Origin Of Things - GitHub

Browser tabs: davidgohel (David Gohel) x +

Address bar: GitHub, Inc. (US) | https://github.com/davidgohel

Search: Search GitHub

Navigation: Pull requests Issues Gist



David Gohel
davidgohel

Lysis-consultants
Paris
<http://davidgohel.github.io>
Joined on 3 May 2013

16 Followers **9** Starred **6** Following


Contributions Repositories Public activity Unfollow

Popular repositories

- ReporteRs** 70 ★
ReporteRs is an R package for creating MS ...
- R2DOCX** 16 ★
- rtable** 2 ★
tabular reporting from R with package Report...
- R2DOC** 1 ★
- davidgohel.github.io** 0 ★

Repositories contributed to

- user2014/user2014.github.io**
UseR 2014 Web Pages
- isomorphisms/hire-an-r-progr...**
list of R developers
- jverzani/gWidgets2**
Rewrite of gWidgets



Public contributions

	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
M												
W												
F												

Summary of Pull Requests, issues opened, and commits. [Learn more.](#) Less [color scale] More

Contributions in the last year

702 total

Longest streak

6 days

Current streak

0 days



The Origin Of Things - **Bioconductor**

Bioconductor - BiocViews x +

www.bioconductor.org/packages/release/BiocViews.html#__Genetics

Search:

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Home Install Help Developers About

Home » BiocViews

All Packages

Bioconductor version 3.1 (Release)

Autocomplete biocViews search:

Packages found under Genetics:

Show entries Search table:

Package	Maintainer	Title
aCGH	Peter Dimitrov	Classes and functions for Array Comparative Genomic Hybridization data.
AGDEX	Cuilan Iani Gao	Agreement of Differential Expression Analysis
AllelicImbalance	Jesper R Gadin	Investigates allele specific expression
BaseSpaceR	Jared O'Connell	R SDK for BaseSpace RESTful API
BEAT	Kemal Akman	BEAT - BS-Seq Epimutation Analysis Toolkit
biomvRCNS	Yang Du	Copy Number study and Segmentation for multivariate biological data
Biostrings	H. Pages	String objects representing biological sequences, and matching algorithms
BiSeq	Katja Hebestreit	Processing and analyzing bisulfite sequencing data
BSgenome	H. Pages	Infrastructure for Biostrings-based genome data packages

- CellBiology (24)
- Cheminformatics (7)
- FunctionalGenomics (3)
- Genetics (111)**
- Lipidomics (1)
- Metabolomics (17)
- Metagenomics (3)
- Pharmacogenetics (5)
- Pharmacogenomics (5)
- Proteomics (71)
- SystemsBiology (17)
- StatisticalMethod (293)
- Bayesian (17)
- Classification (68)
- Clustering (102)
- DecisionTree (5)



13 reasons why **you will** love GNU R

- I R is (extremely) cheap. In fact - it's free :)
- II R has (extremely) wide range of capabilities
- II 1/3 :) **R is easy to maintain!**
- IV R is supported by the community
- V R is supported by the business
- VI R is able to read data in many formats
- VII Interoperability is easy to achieve
- VIII R is truly cross-platform
- IX R offers numerous ways of presenting data
- X There are many options to optimize the code
- XI R is able to handle large amount of data
- XII R has a set of fancy tools and IDEs
- XII FDA accepted using R for drug trials!



R is easy to maintain!

R is easy to maintain. Managing installed packages and keeping them up-to-date becomes a piece of cake:

- Packages are available in [CRAN](#), [GitHub](#), [BioConductor](#) and [Omegahat](#) repositories. CRAN Repositories are [mirrored](#) in more than 50 countries.
- Dependencies between packages are resolved automatically. Just type `install.packages("package_name")` and R download necessary packages, validate ^{MD5} and install them.
- Packages in a binary form can be installed from local `.zip` archives
- Packages can be also [built from sources](#) (common approach on Linux)
- Stay up-to-date with `update.packages()`. This could not be easier.
- By the use of [installr](#) package entire environment can be updated at once.
- One can have [multiple libraries](#) of packages in order to organize them



R is easy to maintain!

Firefox

CRAN - Package RcmdrPlugin.KMggplot2

cran.r-project.org/web/

RcmdrPlugin.KMggplot2: An Rcmdr Plug-In for Kaplan-Meier Plots and Other Plots by using the ggplot2 Package

This package is an R Commander plug-in for Kaplan-Meier plots by using the ggplot2 package.

Version: 0.2-0

Depends: R (≥ 2.15.2), stats, methods, grid

Imports: ggthemes (≥ 1.3.1), gtable (≥ 0.1.1), Rcmdr (≥ 1.9-3), RColorBrewer (≥ 1.1-2), survival (≥ 2.37-2), tcltk2 (≥ 2.0-0)

Published: 2013-01-23

Author: Triad sou. and Kengo NAGASHIMA

Maintainer: Triad sou. <triadsou at gmail.com>

License: GPL-2

NeedsCompilation: no

Materials: NEWS

CRAN checks: RcmdrPlugin.KMggplot2 results

Downloads:

Reference manual: RcmdrPlugin.KMggplot2.pdf

Package source: RcmdrPlugin.KMggplot2_0.2-0.zip

Windows binaries: r-devel: RcmdrPlugin.KMggplot2_0.2-0.zip, r-release: RcmdrPlugin.KMggplot2_0.2-0.zip, r-oldrel: RcmdrPlugin.KMggplot2_0.2-0.zip

OS X Snow Leopard binaries: r-release:

RGui (32-bit)

File Edit View Misc Packages Windows Help

R Console

```
> install.packages("RcmdrPlugin.KMggplot2")
Installing package into 'C:/Users/.../Documents/R/win-library/3.0'
(as 'lib' is unspecified)
--- Please select a CRAN mirror for use in this session ---
also installing the dependency 'ggthemes'

trying URL 'http://r.meteo.uni.wroc.pl/bin/windows/contrib/3.0/ggthemes_1.7.0.zip'
Content type 'application/zip' length 182333 bytes (178 Kb)
opened URL
downloaded 178 Kb

trying URL 'http://r.meteo.uni.wroc.pl/bin/windows/contrib/3.0/RcmdrPlugin.KMggplot2_0.2-0.zip'
Content type 'application/zip' length 520569 bytes (508 Kb)
opened URL
downloaded 508 Kb

package 'ggthemes' successfully unpacked and MD5 sums checked
package 'RcmdrPlugin.KMggplot2' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:\Users\...\AppData\Local\Temp\Rtmp0wuFZv\downloaded_packages
```



With the [miniCRAN](#) package one can build his own, private mini repository of R packages. This is a perfect solution for creating dedicated, in-house production environments for the following reasons:

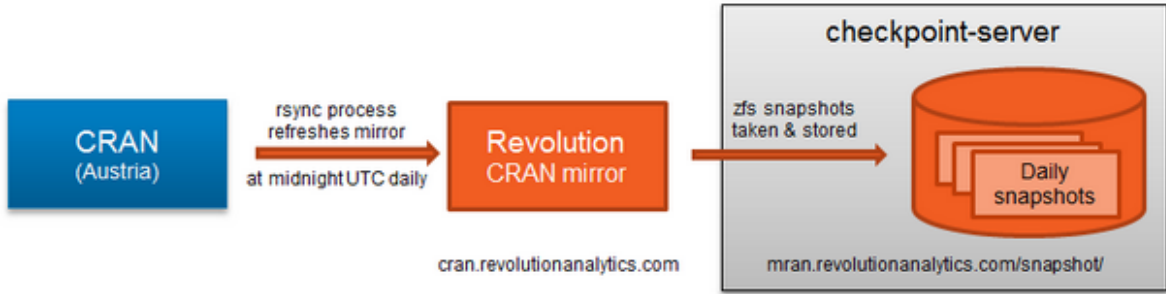
- *You may wish to mirror only a subset of CRAN, for security, legal compliance or any other in-house reason*
- *You may wish to restrict internal package use to a subset of public packages, to minimize package duplication, or other reasons of coding standards*
- *You may wish to make packages available from public repositories other than CRAN, e.g. BioConductor, r-forge, OmegaHat, etc.*
- *You may wish to add custom in-house packages to your repository*

www.r-bloggers.com/introducing-minicran-an-r-package-to-create-a-private-cran-repository/

Do not forget to visit the [quick introduction to miniCRAN](#).



Versioned CRAN – meet Revolution::checkpoint



- [announcements](#)
- [applications](#)
- [beginner tips](#)
- [big data](#)
- [courses](#)
- [current events](#)
- [data science](#)
- [developer tips](#)
- [events](#)
- [finance](#)
- [government](#)
- [graphics](#)
- [high-performance computing](#)
- [life sciences](#)
- [open source](#)
- [other industry](#)
- [packages](#)
- [popularity](#)
- [predictive analytics](#)
- [profiles](#)
- [R](#)
- [R is Hot](#)
- [random](#)
- [Revolution](#)
- [Rmedia](#)
- [roundups](#)
- [sports](#)
- [statistics](#)
- [user groups](#)

checkpoint package

The goal of the `checkpoint` package is to solve the problem of package reproducibility in R. Since packages get updated on CRAN all the time, it can be difficult to recreate an environment where all your packages are consistent with some earlier state. To solve this issue, `checkpoint` allows you to install packages locally as they existed on a specific date from the corresponding snapshot (stored on the checkpoint server) and it configures your R session to use only these packages. Together, the `checkpoint` package and the checkpoint server act as a "CRAN time machine", so that anyone using `checkpoint` can ensure the reproducibility of scripts or projects at any time.

```
> install.packages("checkpoint")
```



```
1 #myscript.R
2 require(checkpoint)
3 checkpoint("2014-09-17")
4
5 require("foreach")
```

Use `checkpoint()` to install and use packages from 2014-09-17





13 reasons why **you will** love GNU R

- I R is (extremely) cheap. In fact - it's free :)
- II R has (extremely) wide range of capabilities
- II 2/3 :) **R is not resource consuming!**
- IV R is supported by the community
- V R is supported by the business
- VI R is able to read data in many formats
- VII Interoperability is easy to achieve
- VIII R is truly cross-platform
- IX R offers numerous ways of presenting data
- X There are many options to optimize the code
- XI R is able to handle large amount of data
- XII R has a set of fancy tools and IDEs
- XII FDA accepted using R for drug trials!



R is not resouRce consuming!

R-3.1.3 Properties

General | Sharing | Security | Previous Versions | Customize

R-3.1.3

Type: File folder (.3)

Location: C:\Program Files\R

Size: 115 MB (121 497 612 bytes)

Size on disk: 127 MB (133 419 008 bytes)

Contains: 4 534 Files, 709 Folders

Created: 13 kwietnia 2015, 02:35:03

Attributes: Read-only (Only applies to files in folder) Hidden

OK Cancel Apply

```
C:\windows\system32\cmd.exe
C:\>tasklist /fi "imagename eq Rgui.exe"
Image Name                PID Session Name        Session#    Mem Usage
-----
Rgui.exe                   8316 Console                1           34 848 K
C:\>_
```

Fresh installation:

- Disk space: ~ 130MiB
- RAM: ~ 35MiB

```
adrian@aaolsz: ~
Plik Edycja Widok Wyszukiwanie Terminal Pomoc
top - 16:59:16 up 33 min, 3 users, load average: 0,06, 0,06, 0,12
Tasks: 1 total, 0 running, 1 sleeping, 0 stopped, 0 zombie
%Cpu(s): 0,3 us, 0,0 sy, 0,0 ni, 99,7 id, 0,0 wa, 0,0 hi, 0,0 si, 0,0 st
KiB Mem: 4058944 total, 1165472 used, 2893472 free, 111888 buffers
KiB Swap: 731132 total, 0 used, 731132 free. 506700 cached Mem

  PID USER      PR  NI   VIRT   RES   SHR S  %CPU  %MEM    TIME+  COMMAND
 1922 adrian    20   0  94660 34000 7052 S   0,0   0,8   0:00.27 R
```

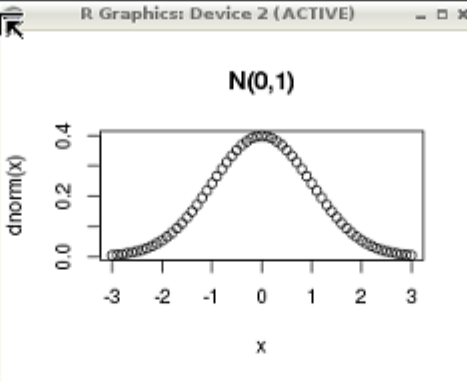



R is not resouRce consuming!

pi's X desktop (raspberrypi:1) - TightVNC Viewer

Shutdown Task Manager R Wicd Network Manager WiFi Config Up-169.254.10.10 Root Terminal LXTerminal LeafPad File Manager Midnight Commander Xarchiver calculator Chromium Web Browser NetSurf Web Browser LibreOffice Document Viewer GNU Image Manipulation ... IDLE 3 MonoDevelop RapidSVN pgAdmin III Start PostgreSQL VLC media player

R Graphics: Device 2 (ACTIVE)



LXTerminal

```
File Edit Tabs Help
> x <- seq(-3, 3, by=0.1)
> plot(x, dnorm(x), main="N(0,1)")
>
```

pi@raspberrypi: ~

```
File Edit Tabs Help
top - 02:39:06 up 22 min, 3 users, load average: 0.31, 0.39, 0.43
Tasks: 1 total, 0 running, 1 sleeping, 0 stopped, 0 zombie
%Cpu(s): 10.7 us, 6.7 sy, 0.0 ni, 82.6 id, 0.0 wa, 0.0 hi, 0.0 si, 0.0 st
KiB Mem: 382840 total, 282344 used, 100496 free, 21744 buffers
KiB Swap: 102396 total, 0 used, 102396 free, 123076 cached
```

PID	USER	PR	NI	VIRT	RES	SHR	S	%CPU	%MEM	TIME+	COMMAND
3365	pi	20	0	65332	27m	8072	S	0.0	7.5	0:06.00	R





13 reasons why **you will** love GNU R

- I R is (extremely) cheap. In fact - it's free :)
- II R has (extremely) wide range of capabilities
- III R is supported by the world of science**
- IV R is supported by the community
- V R is supported by the business
- VI R is able to read data in many formats
- VII Interoperability is easy to achieve
- VIII R is truly cross-platform
- IX R offers numerous ways of presenting data
- X There are many options to optimize the code
- XI R is able to handle large amount of data
- XII R has a set of fancy tools and IDEs
- XIII FDA accepted using R for drug trials!



At the center of the R Community is the [R Core Group](#) of approximately 20 developers who maintain R and guide its evolution. They are [experienced](#) statisticians, [well-known](#) in the world of science, with significant achievements.

Each [thematic section](#) has its own [academic supervisor](#).

[Robert Gentleman](#)'s articles about R and Bioconductor are among the most cited in bioinformatics: over **5200 citations** according to [Google Scholar](#) April 2014

There are numerous [e-learning materials](#) on the Internet which provide valuable assistance in data analysis with R. These are often provided by prestigious academic centers like [Princeton](#), [Stanford](#) or [Hopkins](#).

The R project is closely linked with "[the R Journal](#)" which is a mine of scientific knowledge of using R for professional data analysis.



...supported by the science

Welcome. The R Journal

Browser navigation bar with back, forward, search, and utility icons.

The R Journal

 [RSS Feed](#)
ISSN: 2073-4859

Home

About The R Journal

Current Issue

The R Journal is the open access, refereed journal of the [R project](#) for statistical computing. It features short to medium length articles covering topics that might be of interest to users or developers of R, including

Accepted Articles

- Add-on packages:** short introductions to R extension packages.
- Programmer's Niche:** hints for programming in R.
- Help Desk:** hints for newcomers explaining aspects of R that might not be so obvious from reading the manuals and FAQs.
- Applications:** demonstrating how a new or existing technique can be applied in an area of current interest using R, providing a fresh view of such analyses in R that is of benefit beyond the specific application.

Archive

Submissions

Editorial Board

The R Journal intends to reach a wide audience and have a fast-track but thorough review process. Papers are expected to be reasonably short, clearly written, not too technical, and of course focused on R. Authors of refereed articles should take care to:

- put their contribution in context, in particular discuss related R functions or packages;
- explain the motivation for their contribution;
- provide code examples that are reproducible.

Continuing from [R News](#), *The R Journal* also has a news section, including information on:

Changes in R: new features of the latest release.



There are hundreds of places containing e-learning materials devoted to R.

Good places to start learning R:



1. <http://cran.r-project.org/manuals.html>
2. <http://cran.r-project.org/other-docs.html>
3. <http://cran.r-project.org/doc/manuals/r-patched/R-intro.html>



4. <http://stats.stackexchange.com/questions/138/resources-for-learning-r>



5. http://www.youtube.com/results?search_query=R+learn+statistics

6. <http://www.statmethods.net> - **Quick R**

7. <http://adv-r.had.co.nz> - **Advanced R** by Hadley Wickham

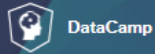
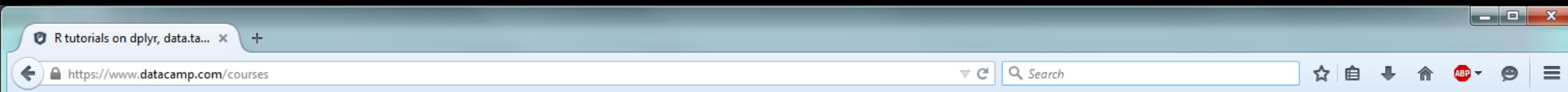
8. <http://www.cookbook-r.com>

9. <http://rseek.org/?q=learn+R>

10. <http://www.google.pl/search?q=learn+R>



Take a cou**R**se!



Courses Pricing Teams [Create free account](#) [Sign in](#)

Discover all courses & tutorials on **R & Data Science**

Build a solid foundation in data science, and strengthen your R programming skills. Learn how to do data manipulation, visualization and more with our R tutorials.

[Start Learning R](#)

Step One - Get To Know R

2 Courses

R is the leading open-source programming language in data science and statistics. It is used by students, academics and professionals to perform data analysis, and for building data-driven solutions and applications.

Introduction to R

Write your first R code, and discover vectors, matrices, data frames and lists.



Intermediate R

Take your R programming skills to the next level with loops apply.



Step Two - Statistics with R

7 Courses

A selection of courses by prof. Andrew Conway (Princeton University) that provide a comprehensive yet friendly introduction to fundamental concepts of statistics.

In collaboration with



Courses given by



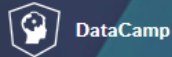
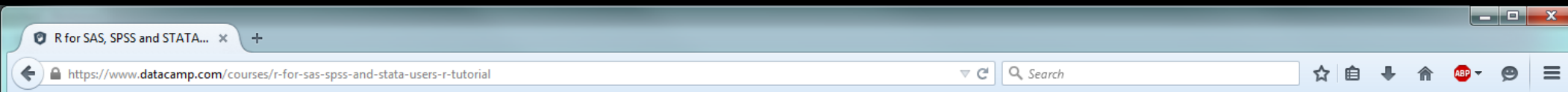
Jonathan
Cornelissen



Andrew
Conway



Take a course!



Courses Pricing Teams [Create free account](#) [Sign in](#)

Course

R for SAS, SPSS and STATA Users

If you already know SAS, SPSS or Stata, you don't need to spend time learning how to analyze data; you need a course that focuses on translating your knowledge into R. This comprehensive course introduces R jargon using the language you're familiar with.

- Difficulty:** Beginner
- Duration:** 16 hours
- Participating:** 2688

[Start Course for Free](#)

[Or, take a subscription](#)

Chapter 1: Introduction

This section introduces R and describes how it integrates the five main parts of SAS, SPSS and Stata into a powerful, comprehensive system.

Number of videos: 1 Number of exercises: 2



Chapter 2: Installing & Maintaining R

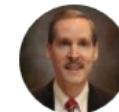
The software you're familiar with is a complete software package. However, R is downloaded and installed in pieces. This chapter tells you how to find parts of R that match your current software and how to install them.

Number of videos: 1 Number of exercises: 3



Course given by:

Bob Muenchen



Robert A. Muenchen is the author of R for SAS and SPSS Users and, with Joseph M. Hilbe, R for Stata Users. He is a consulting statistician with over 30 years of experience and is currently the manager of the Research Computing Support (formerly the Statistical Consulting Center) at the University of Tennessee. He holds a B.A. in Psychology and an M.S. in Statistics. Bob has conducted research for a variety of public and private organizations and has assisted on more than 1,000 graduate theses and dissertations. He has written or coauthored over 70 articles published in scientific journals and conference proceedings.

Bob has served on the advisory boards of the SAS Institute, SPSS Inc, the Statistical



Look! This library contains pRiceless resources!

The screenshot shows a Firefox browser window with the address bar displaying `cran.r-project.org/other-docs.html`. The page title is "Contributed Documentation". Below the title, there are links for "English" and "Other Languages". The main text explains that manuals, tutorials, etc. are provided by users of R, and the R core team does not take responsibility for contents. It encourages contributions and provides instructions on how to submit. A note mentions using a "directory listing" to sort documents. Below this, there is a section titled "English Documents" and a list of documents with more than 100 pages.

Contributed Documentation

[English](#) --- [Other Languages](#)

Manuals, tutorials, etc. provided by users of R. The R core team does not take any responsibility for contents, but we appreciate the effort very much and encourage everybody to contribute to this list! To submit, follow the submission instructions on the [CRAN main page](#). All material below is available directly from CRAN, you may also want to look at the list of [other R documentation](#) available on the Internet.

Note: Please use the [directory listing](#) to sort by name, size or date (e.g., to see which documents have been updated lately).

English Documents

Documents with more than 100 pages:

- “Using R for Data Analysis and Graphics - Introduction, Examples and Commentary” by John Maindonald ([PDF](#), data sets and scripts are available at [JM's homepage](#)).
- “Practical Regression and Anova using R” by Julian Faraway ([PDF](#), data sets and scripts are available at the [book homepage](#)).
- The [Web Appendix](#) to the book “An R and S-PLUS Companion to Applied Regression” by John Fox contains information about using S (R and S-PLUS) to fit a variety of regression models.
- “An Introduction to S and the Hmisc and Design Libraries” by Carlos Alzola and Frank E. Harrell, especially of interest to SAS users, users of the Hmisc or Design packages, or R users interested in data manipulation, recoding, etc. ([PDF](#))
- “Statistical Computing and Graphics Course Notes” by Frank E. Harrell, includes material on S, LaTeX, reproducible research, making good graphs, brief overview of computer languages, etc. ([PDF](#)).
- “An Introduction to R: Software for Statistical Modelling & Computing” by Petra Kuhnert and Bill Venables ([ZIP 3.8MB](#)): A 360 page PDF document of lecture notes in combination with the data sets and R scripts used in the manuscript.
- “Introduction to the R Project for Statistical Computing for Use at the ITC” by David Rossiter ([PDF](#), 2012-08-20, 141 pages).
- “Analysis of Epidemiological Data Using R and Epicalc” by Virasakdi Chongsuvivatwong ([PDF](#)).
- “Statistics Using R with Biological Examples” by Kim Seefeld and Ernst Linder ([PDF](#)).
- “Icebreaker” by Andrew Robinson ([PDF](#), 2008-05-08).
- “Applied Statistics for Bioinformatics Using R” by Wim Krijnen ([PDF](#), 2009-11-17, 278 pages).
- “An Introduction to R” by Longhow Lam ([PDF](#), 2010-10-28, 212 pages).
- “R and Data Mining: Examples and Case Studies” by Yanchang Zhao ([PDF](#), 2013-04-26, 160 pages).



...thousands of valuable paper**R**s at our fingertips

Firefox


Bioassay analysis using R - paper

www.jstatsoft.org/v12/i05/paper

Google

Strona: 1 z 22

Skala automatyczna



Journal of Statistical Software

January 2005, Volume 12, Issue 5. <http://www.jstatsoft.org/>

Bioassay Analysis using R

Christian Ritz
Royal Veterinary and
Agricultural University

Jens C. Streibig
Royal Veterinary and
Agricultural University

Abstract

We describe an add-on package for the language and environment R which allows simultaneous fitting of several non-linear regression models. The focus is on analysis of dose response curves, but the functionality is applicable to arbitrary non-linear regression models. Features of the package is illustrated in examples.

Keywords: dose response data, multiple curves, non-linear regression.



...thousands of valuable paper**R**s at our fingertips

R-intro.pdf

www.bendixcarstensen.com

Page: 1 of 54 70%

A short introduction to for Epidemiology

June 2014
Version 4

Compiled Friday 27th June, 2014, 09:48
from: C:/Bendix/undervis/SPE/Intro/R-intro.tex

Michael Hills Retired
Highgate, London

Martyn Plummer International Agency for Research on Cancer, Lyon
plummer@iarc.fr

Bendix Carstensen Steno Diabetes Center, Gentofte, Denmark
& Department of Biostatistics, University of Copenhagen
bxc@steno.dk
www.pubhealth.ku.dk/~bxc

Edition 2014 by Bendix Carstensen

all2012_epir-...

...net/docs/ph251d_fall2

Page: 1 of 215 80%

Tomás J. Aragón

Applied Epidemiology Using R

An Open Access Book—Chapters 1–3

September 23, 2012

University of California, Berkeley School of Public Health, and
San Francisco Department of Public Health

Applied Epidemiology Using R 23-Sep-2012 © Tomás J. Aragón (www.medepi.com)



Where to seaRch? → Rseek.org

The screenshot shows a web browser window with the Rseek.org search engine. The search term 'dose response' is entered in the search bar, and the results are sorted by 'Relevance'. The search returned approximately 4,470,000 results in 0.37 seconds. The results list includes several entries related to R packages and articles on dose-response analysis.


Rseek.org logo:

Search bar: [X] [Search]

Navigation tabs: All (selected), Packages, For Beginners, Articles, Support, Books

Results summary: About 4,470,000 results (0.37 seconds) | Sort by: Relevance

Powered by Google™ Custom Search

- Package 'drfit'**
cran.r-project.org/web/packages/drfit/drfit.pdf
File Format: PDF/Adobe Acrobat
Jul 28, 2014 ... Title **Dose-response** data evaluation. Imports MASS, RODBC, drc. Description drfit provides basic and easy-to-use functions for fitting.
Labeled [Packages](#)
- Package 'drc'**
cran.r-project.org/web/packages/drc/drc.pdf
File Format: PDF/Adobe Acrobat
Package 'drc'. April 14, 2015. Version 2.5-12. Date 2015-04-14. Title Analysis of **Dose-Response** Curves. Author Christian Ritz <ritz@bioassay.dk>, Jens C.
Labeled [Packages](#)
- IsoGene: An R Package for Analyzing **Dose-response** Studies in ...**
journal.r-project.org/archive/.../RJournal_2010-1_Pramana-et-al.pdf
File Format: PDF/Adobe Acrobat
Dose-response Studies in Microarray. Experiments by Setia Pramana, Dan Lin, Philippe Haldermans, Ziv. Shkedy, Tobias Verbeke, Hinrich Göhlmann, An De ...
- Modeling **Dose-Response** Microarray Data in Early Drug Deve...**
www.amazon.com/Modeling-Dose-Response.../dp/3642240062
 This book focuses on the analysis of **dose-response** microarray data in pharmaceutical settings, the goal being to cover this important topic for early drug ...
Labeled [Books](#)
- Package 'DoseFinding'**
cran.r-project.org/web/packages/DoseFinding/DoseFinding.pdf



Where to seaRch? → R Site Search

R site search: <non inferior... x +

finzi.psych.upenn.edu/cgi-bin/namazu.cgi?query=non+inferiority&max=100&result=normal&sort=s Search

R Site Search

Query: non inferiority Search! [\[How to search\]](#)

Display: 100 Description: normal Sort: by score

Target:

- Functions
- Vignettes
- Task views

For problems WITH THIS PAGE (not with R) contact baron@psych.upenn.edu.

Results:

References:

- views: [non: 27] [inferiority: 0] [TOTAL: 0]
- vignettes: [non: 1515] [inferiority: 4] [TOTAL: 4]
- functions: [non (Too many documents hit. Ignored)] [inferiority: 75] [TOTAL: 75]

Total 79 documents matching your query.

1. [R: Design and Analysis of Three-armed Clinical Non-inferiority...](#) (score: 26)
Author: unknown
Date: Wed, 07 Jan 2015 09:37:05 -0500
Design and Analysis of Three-armed Clinical **Non-inferiority** or Superiority Trials with Active and Placebo Control Description Author(s) References page for ThreeArmedTrials-package {ThreeArmedTrials
<http://finzi.psych.upenn.edu/R/library/ThreeArmedTrials/html/ThreeArmedTrials-package.html> (2,389 bytes)
2. [R: Sample size for the non-inferiority t-test](#) (score: 25)
Author: unknown
Date: Sat, 07 Feb 2015 08:00:56 -0500
Sample size for the **non-inferiority** t-test Description Usage Arguments Details Value Warning Author(s) References See Also Examples page for sampleN.noninf {PowerTOST} sampleN.noninf {PowerTOST} R D



Where to seaRch? → R Documentation

http://www...on=&author= x +

www.rdocumentation.org/advanced_search?utf8=✓&q=non-linear&package_name=&function_name=&title=&description=&author=

Search

Start searching the documentat

Documentation Discussion About Rdocumentation package

Advanced Function Search

All Fields

Package Name

Function Name

Title

Description

Author(s)

Start search

Filter by packages

- stats (2)
- MASS (2)
- DEoptimR (1)
- ade4 (1)
- deSolve (2)
- TSA (3)
- lava (2)
- pastecs (2)
- HistData (1)
- Rsolnp (5)

- ez (1)
- nleqslv (3)
- MatrixModels (2)
- gamlss (3)
- TIMP (1)
- rPortfolio (1)
- drc (6)
- asbio (1)
- circular (1)
- tsDyn (7)

Found 269 results

1 2 3 4 5 ... Next > Last >

nlm stats 3.0.1

Non-Linear Minimization

This function carries out a minimization of the function f using a Newton-type algorithm. See the references for ...

summary.nls stats 3.0.1

Summarizing Non-Linear Least-Squares Model Fits

summary method for class "nls".

sammon MASS 7.3-35

Sammon's Non-Linear Mapping

One form of non-metric multidimensional scaling.

rms.curv MASS 7.3-35

Relative Curvature Measures for Non-Linear Regression

Task Views: Bayesian, ChemPhys, ClinicalTrials, Cluster, DifferentialEquations, Distributions, Econometrics, Environmetrics, ExperimentalDesign, Finance, Genetics, gR, Graphics, HighPerformanceComputing, MachineLearning, MedicalImaging, MetaAnalysis, Multivariate, NaturalLanguageProcessing, NumericalMathematics, OfficialStatistics, Optimization, Pharmacokinetics, Phylogenetics, Psychometrics, ReproducibleResearch, Robust, SocialSciences, Spatial, SpatioTemporal, Survival, TimeSeries, WebTechnologies

DataCamp: Learn Data Science with R. \$25/month. Data Manipulation, Data Visualization, R Programming, Big Data, and much more. Discover All Courses

Aggregating packages from: CRAN, Bioconductor, GitHub



13 reasons why **you will** love GNU R

- I R is (extremely) cheap. In fact - it's free :)
- II R has (extremely) wide range of capabilities
- III R is supported by the world of science
- IV R is supported by the community**
- V R is supported by the business
- VI R is able to read data in many formats
- VII Interoperability is easy to achieve
- VIII R is truly cross-platform
- IX R offers numerous ways of presenting data
- X There are many options to optimize the code
- XI R is able to handle large amount of data
- XII R has a set of fancy tools and IDEs
- XII FDA accepted using R for drug trials!



...supported by the **community**

R is a global community of **more than 2 million** and still growing! users (2012, Oracle) and developers who voluntarily contribute their time and technical expertise to maintain, support, **test** and extend the R language and its environment, tools and infrastructure (e.g. CRAN). Among them are **experienced statisticians**, often employed in the well-known pharmaceutical companies like *Merck* or *Amgen*.

Hundreds of bloggers maintain their **webpages**, take active part in **communities** both online and offline through hundreds of forums and **mailing lists**, building knowledge base (visit **rseek.org**). **UseRs** also organize **meetings** and **conferences**.

There are dozens of valuable books written both by academics, researchers and “regular” R users. These books are issued by the prestigious publishing houses like **Springer Verlag** and **Wiley & Sons**.



*The size of the R user community (difficult to define precisely, because there are no sales transactions, but conservatively estimated as being in the tens of thousands, with some independent estimates in the hundreds of thousands), **provides for extensive review of source code and testing in "real world"** settings outside the confines of the formalized testing performed by R Core.*

*This is a key distinction, related to product quality, between R and similar software that is only available to end users in a binary, executable format. In conjunction with detailed documentation and references provided to end users, the size of the R user community, all having full access to the source code, **enables a superior ability to anticipate and verify R's performance and the results produced by R.***



...supported by the **community**

PAZUR

Poznański Akademicki
Złot Użytkowników R



Organizatorzy:

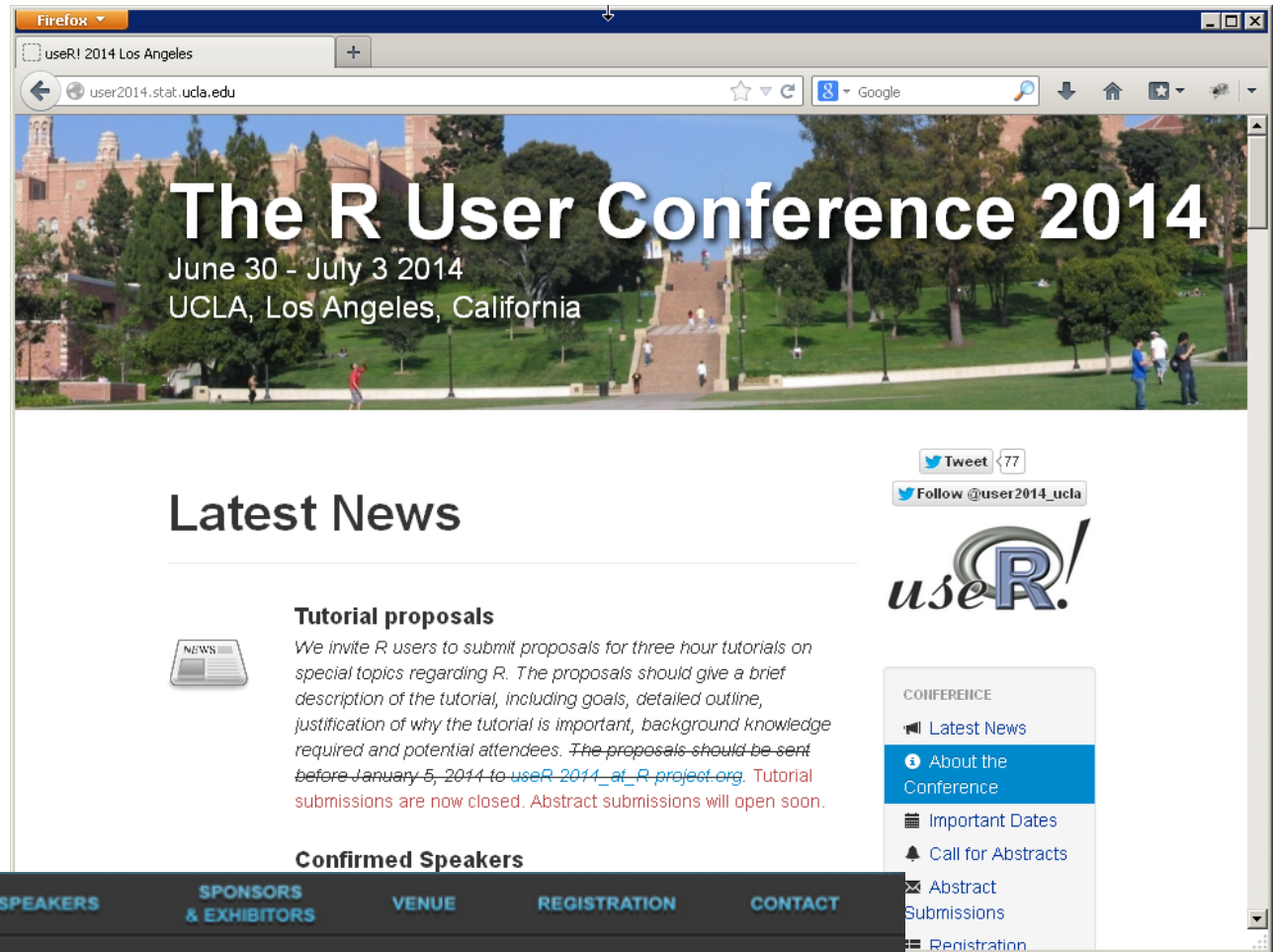
Uniwersytet Przyrodniczy w Poznaniu

ESTYMATO R
STUDENCKIE KOŁO NAUKOWE
Uniwersytet Ekonomiczny
w Poznaniu

Sponsor:
REVOLUTION
ANALYTICS

13 grudnia (piątek) o 18:00
Zoo Coworking
Zwierzyniecka 20, II piętro
Szczegóły:
thinking-in-r.blogspot.com

ZOO POZNAŃ
COWORKING SPACE



Firefox

userR! 2014 Los Angeles

user2014.stat.ucla.edu

The R User Conference 2014

June 30 - July 3 2014
UCLA, Los Angeles, California

[Tweet](#) <77

[Follow @user2014_ucla](#)

Latest News

Tutorial proposals

We invite R users to submit proposals for three hour tutorials on special topics regarding R. The proposals should give a brief description of the tutorial, including goals, detailed outline, justification of why the tutorial is important, background knowledge required and potential attendees. The proposals should be sent before January 5, 2014 to [userR_2014_at_R_project.org](#). Tutorial submissions are now closed. Abstract submissions will open soon.

Confirmed Speakers

CONFERENCE

- Latest News
- About the Conference**
- Important Dates
- Call for Abstracts
- Abstract Submissions
- Registration

ABOUT AGENDA SPEAKERS SPONSORS & EXHIBITORS VENUE REGISTRATION CONTACT

EARL 2015

EFFECTIVE APPLICATIONS OF THE R LANGUAGE
LONDON 14 - 16 SEPTEMBER



 @earlconf



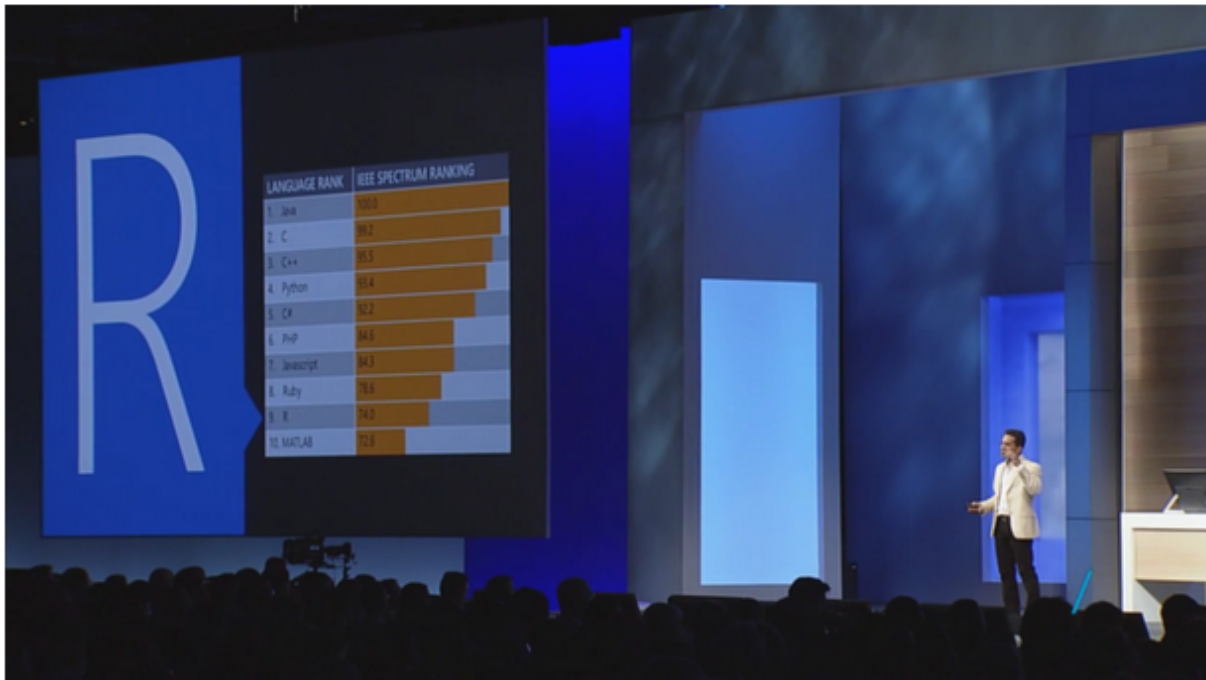
R at Microsoft's BUILD 2015 conference say "wow!"

R Any R code as a cloud servi... x +

blog.revolutionanalytics.com/2015/06/r-build-keynote.html Search

Any R code as a cloud service: R demonstration at BUILD

At last month's [BUILD conference](#) for Microsoft developers in San Francisco, [R](#) was front-and-center on the keynote stage.



In the keynote, Microsoft CVP Joseph Sirosh introduced the "language of data": [open source R](#). Sirosh encouraged the audience to learn R, saying "if there is a single language that you choose to learn today .. let it be R".

The keynote featured a demonstration of genomic data analysis using R. The analysis was based on the [1000 genomes data set](#) stored in the [HDInsight Hadoop-in-the-cloud](#) service.

- [About Categories](#)
- [About the Authors](#)
- [R Community Calendar](#)
- [Local R User Group Directory](#)



Free to download, use, and share.

GET IT NOW

Got comments or suggestions for the blog editor? Email [David Smith](#).

- Follow David on Twitter: [@revodavid](#)
- [+David Smith](#)

[Subscribe to this blog's feed](#)

Categories

- [academia](#)
- [advanced tips](#)



...supported by the **community**



R news and tutorials contributed by (552) R bloggers

Revolutions

Learn more about using open source R for big data analysis, predictive modeling, data science and more from the staff of Revolution Analytics.

- Home
- About
- add your blog!
- Contact us
- RSS
- R jobs

WELCOME!

Here you will find daily news and tutorials about R, contributed by over 555 bloggers. You can subscribe for e-mail updates:

Your e-mail here

Subscribe

13863 readers

And get updates to your Facebook:



17,457 people like R

Fun with the Raspberry Pi

January 21, 2014

By Markus Gesmann

Like Share 103 Tweet 10 +1 2

(This article was first published on mages' blog, and kindly contributed to R-bloggers)

Since Christmas I have been playing around with a Raspberry Pi. It is certainly not the fastest computer, but what a great little toy! Here are a few experiences and online resources that I found helpful.



statistics

September 18, 2014

Comparing machine learning models in R

by Joseph Rickert

While preparing for the DataWeek R Bootcamp that I conducted this week I came across the following gem. This code, based directly on a Max Kuhn presentation of a couple years back, compares the efficacy of two machine learning models on a training data set.

```

#-----
# SET UP THE PARAMETER SPACE SEARCH GRID
ctrl <- trainControl(method="repeatedcv", # use repeated 10
                    repeats=5, # do 5 repetition
                    summaryFunction=twoClassSummary, # Use AUC to pick
                    classProbs=TRUE)
# Note that the default search grid selects 3 values of each tuning param
#
grid <- expand.grid(.interaction.depth = seq(1,7,by=2), # look at tree de
                  n.trees=seq(10,100,by=5) # let iterations

```

Information

- [About this blog](#)
- [Comments Policy](#)
- [About Categories](#)
- [About the Authors](#)
- [R Community Calendar](#)
- [Local R User Group Directory](#)



Got comments or suggestions for this Email [David Smith](#).

Follow David on Twitter: [@revolutions](#)

+David Smith

blog.RDataMining.com

R and Data Mining



Home About

RDataMining Slides Series

Posted on September 14, 2014

by Yanchang Zhao, [RDataMining.com](#)

I have made a series of slides on R and data mining, based on my book titled [R and Data Mining – Examples and Case Studies](#). The slides will be used at my presentations at seminars to graduate students at Universidad Juárez Autónoma de Tabasco (UJAT), prior to my keynote speech on *Analysing Twitter Data with Text Mining and Social Network Analysis* at the CONAIS 2014 conference in Mexico in October 2014.

The slides cover seven topics below. Click the links to download them in PDF files.

- Introduction to Data Mining with R and Data Import/Export in R <http://www.rdatamining.com/docs/RDataMining-slides-introduction-data-import-export.pdf>

Search

Recent Posts

- RDataMining Slides Series
- Slides of 12 tutorials at ACM SIGKDD 2014
- RDataMining group having 6000 members today
- Currency Exchange Rate Forecasting with ARIMA and STL
- Step-by-Step Guide to Setting Up an R-Hadoop System

Recent Comments

- bipin chandra on A Sequence of 9 Courses on Dat...
- emaasit on RDataMining Slides Series

- Home
- Interface
- Input
- Manage
- Stats
- Adv Stats
- Graphs
- Adv Graphs
- Blog

Quick-R

accessing the power of R

Advanced Statistics

[Generalized Linear Models](#)

[Discriminant Function](#)

[Time Series](#)

[Factor Analysis](#)

[Correspondence Analysis](#)

[Multidimensional Scaling](#)

[Cluster Analysis](#)

[Tree-Based Models](#)

[Bootstrapping](#)

[Matrix Algebra](#)

Cluster Analysis

R has an [amazing variety](#) of functions for [cluster analysis](#). In this section, I will describe three of the many approaches: hierarchical agglomerative, partitioning, and model based. While there are no best solutions for the problem of determining the number of clusters to extract, several approaches are given below.

Data Preparation

Prior to clustering data, you may want to remove or estimate missing data and rescale variables for comparability.

```

# Prepare Data
mydata <- na.omit(mydata) # listwise deletion of missing
mydata <- scale(mydata) # standardize variables

```

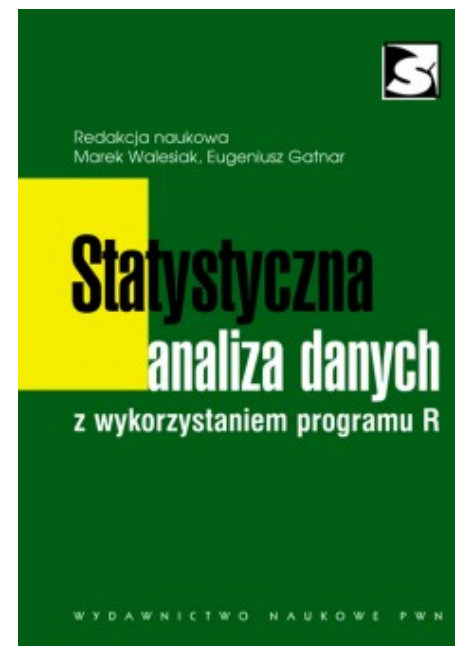
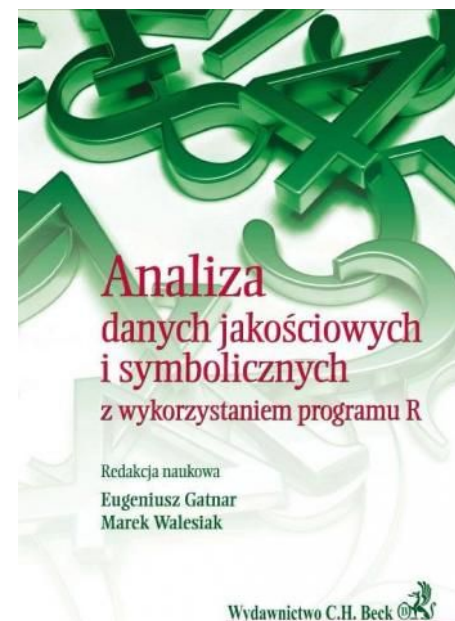
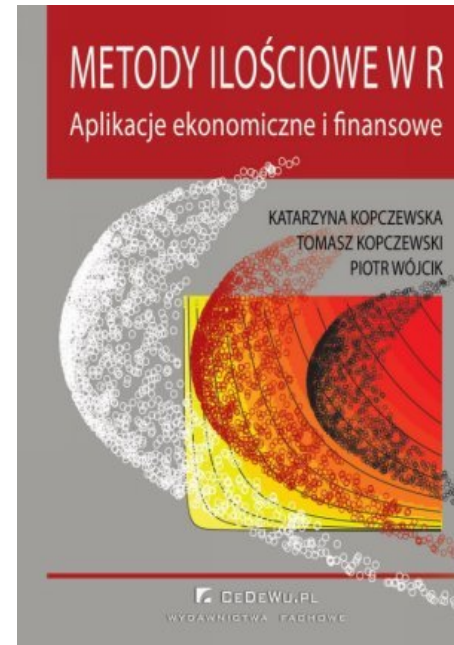
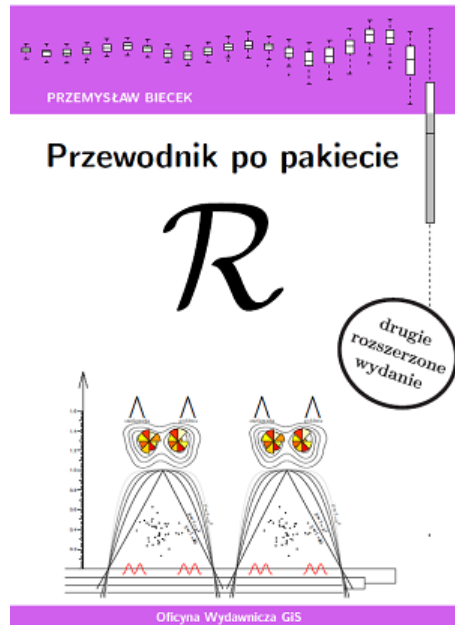
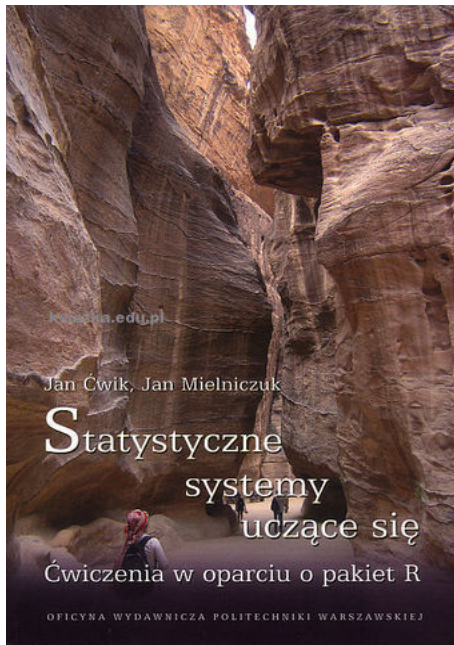


13 reasons why **you will** love GNU R

- I R is (extremely) cheap. In fact - it's free :)
- II R has (extremely) wide range of capabilities
- III R is supported by the world of science
- IV ½ :) Books**
- V R is supported by the business
- VI R is able to read data in many formats
- VII Interoperability is easy to achieve
- VIII R is truly cross-platform
- IX R offers numerous ways of presenting data
- X There are many options to optimize the code
- XI R is able to handle large amount of data
- XII R has a set of fancy tools and IDEs
- XII FDA accepted using R for drug trials!

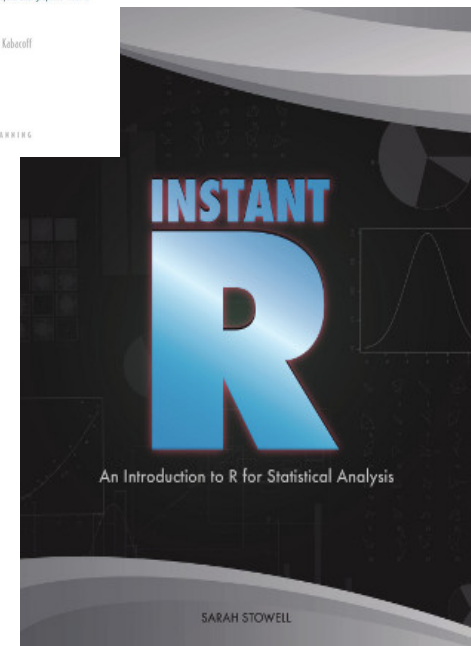
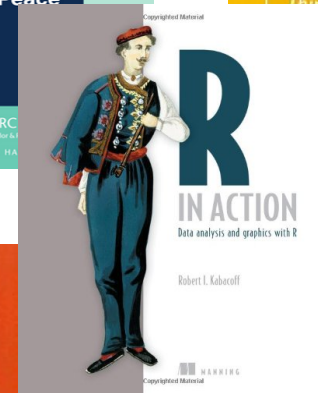
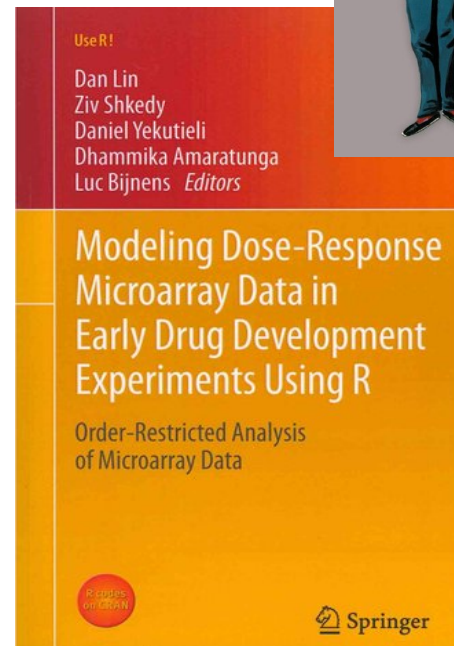
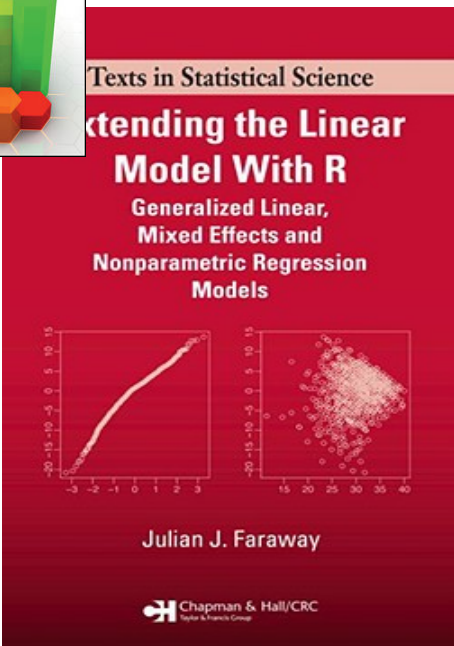
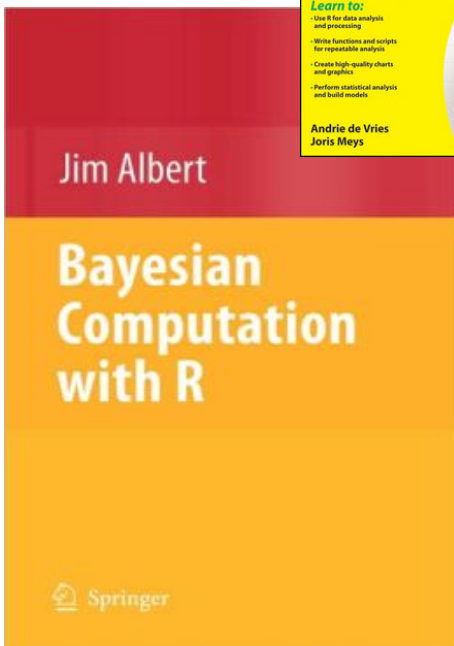
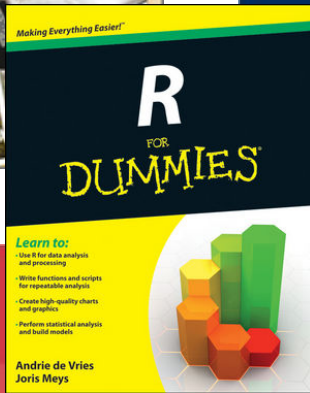
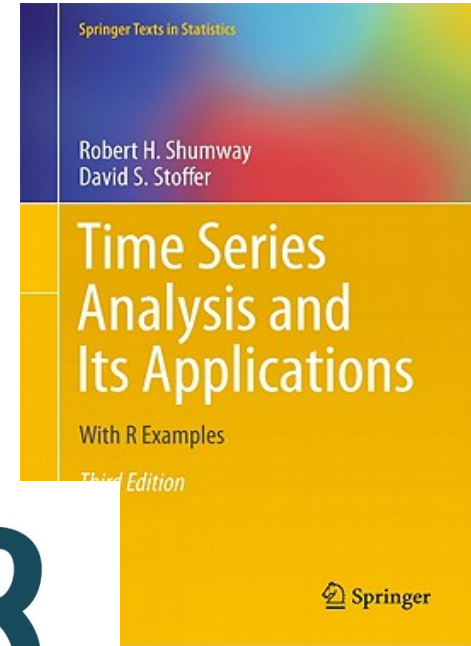
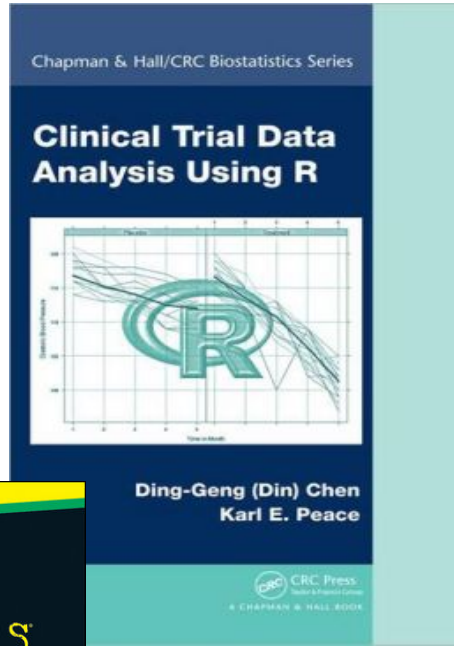
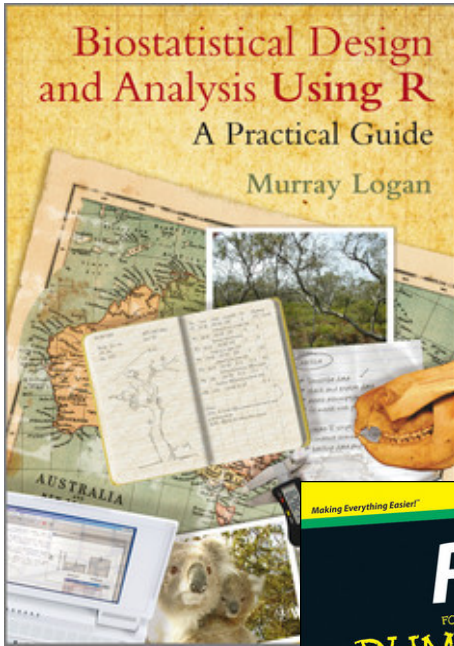


...some polish books





...some foReign books...





...and courses provided by their authors...

The screenshot shows a Firefox browser window with the address bar displaying `www.statistics.com/Clinical-Trials-R/`. The page title is "Biostatistics in R: Clinical Trial Applications" and it is taught by Din Chen and Karl Peace. The page includes a "Register" button and a list of course topics. The browser's address bar shows the URL `www.statistics.com/Clinical-Trials-R/` and the search engine is set to Google.

Biostatistics in R: Clinical Trial Applications
taught by **Din Chen** and **Karl Peace**

Register

Aim of Course:

This course covers the implementation in R of statistical procedures important for the clinical trial statistician. Students completing the course will learn how to use R to compare treatments, incorporate covariates into the analysis, analyze survival (time-to-event) trials, model longitudinal data, and analysis of bioequivalence trials.

This course may be taken individually (one-off) or as part of a [certificate program](#).

Course Program:

WEEK 1: Treatment Comparisons

- R fundamentals associated with clinical trials
- A simple simulated clinical trial
- Statistical models for treatment comparisons
- Incorporating covariates

WEEK 2: Survival Analysis

- Time-to-event data structure
- Statistical models for survival data
- Right-censored data analysis
- Interval-censored data analysis

Course Topics (with checkboxes):

- Introductory
- Bayesian Statistics
- Biostatistics
- Clinical Trial Statistics
- Data / Text Analytics
- Engineering
- Environment
- Predictive Modeling
- Social Sciences
- Survey Statistics
- Using R

Submit

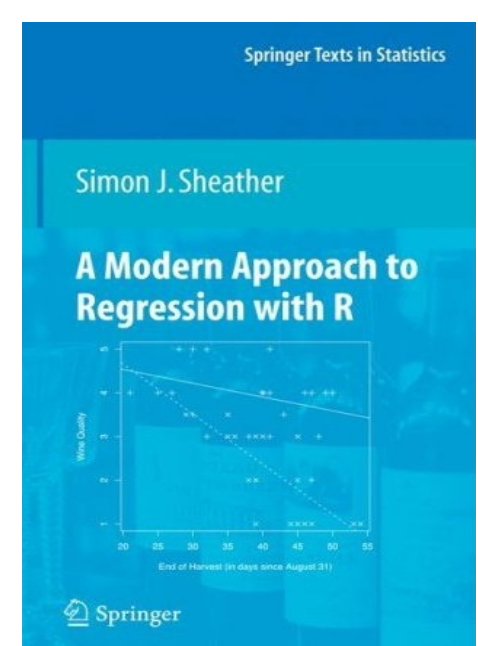
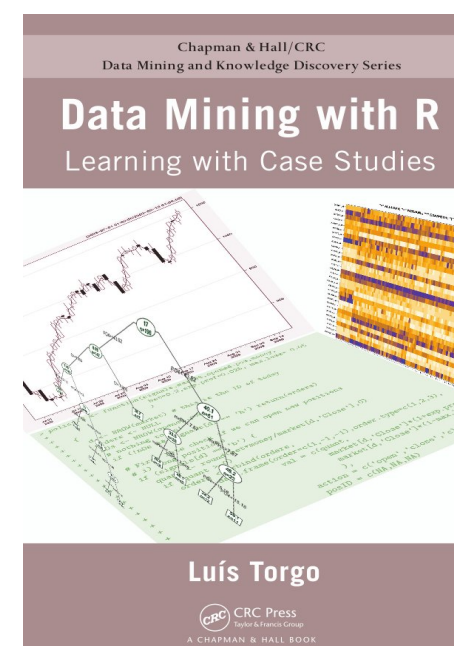
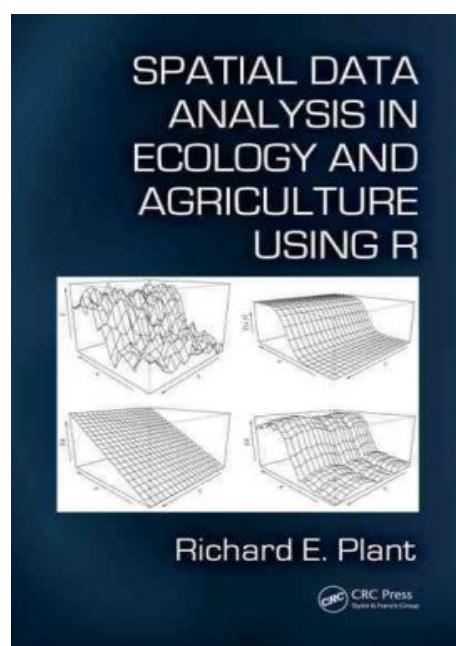
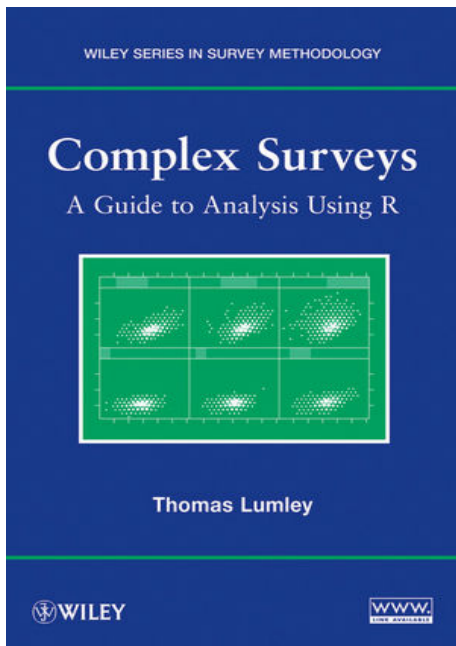
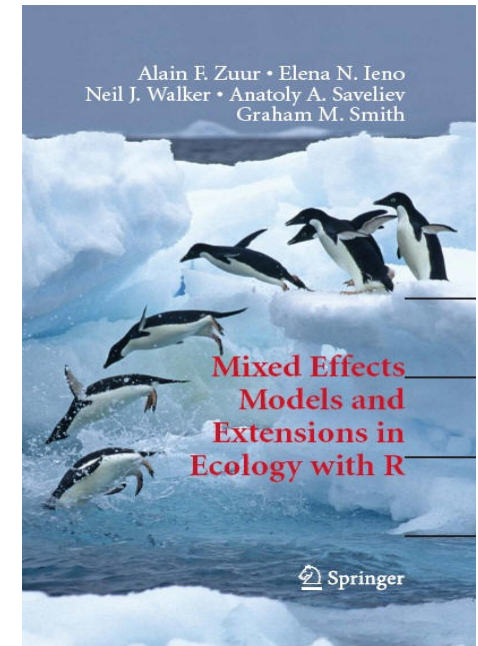
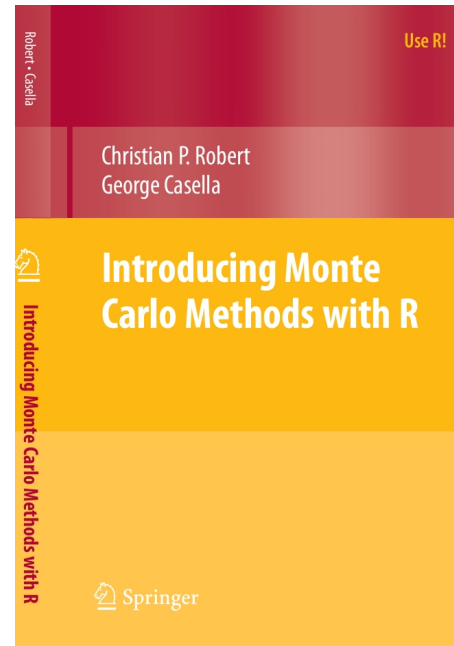
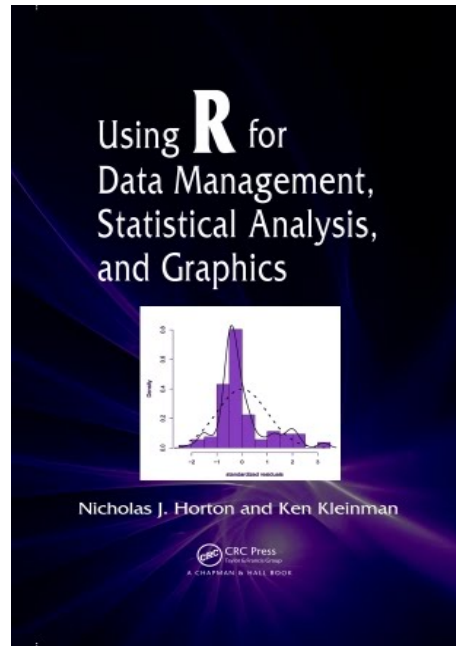
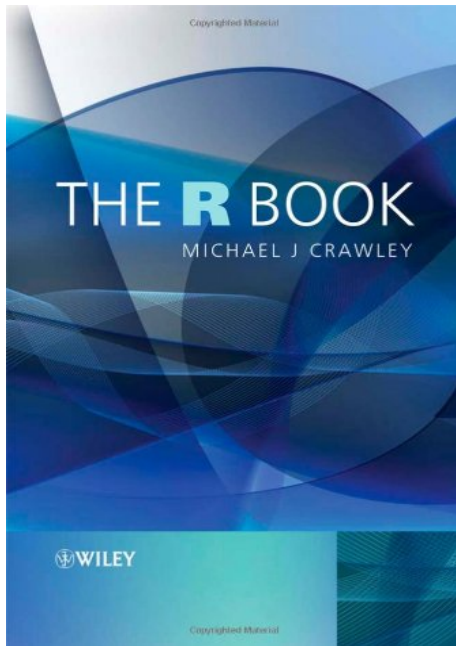
> TOUR A COURSE
> HOW COURSES WORK
> SUGGEST A COURSE

See also the following related courses:

- > [Biostatistics 1](#)
- > [Biostatistics 2](#)
- > [Introduction to Statistical Issues in Clinical Trials](#)

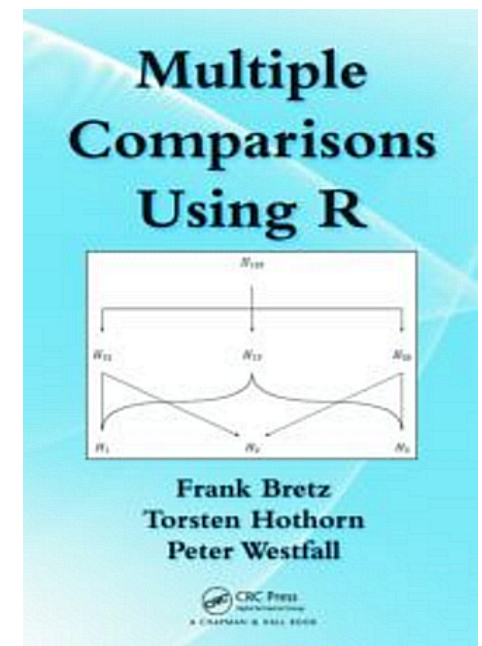
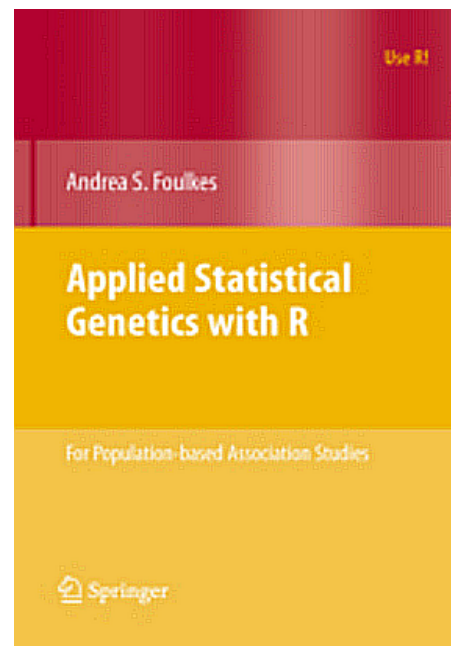
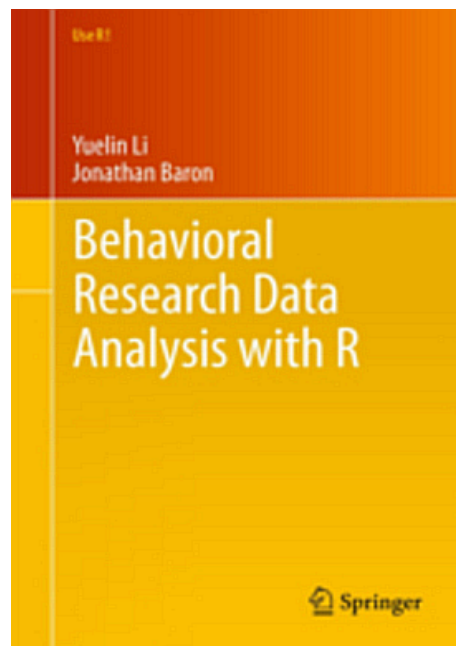
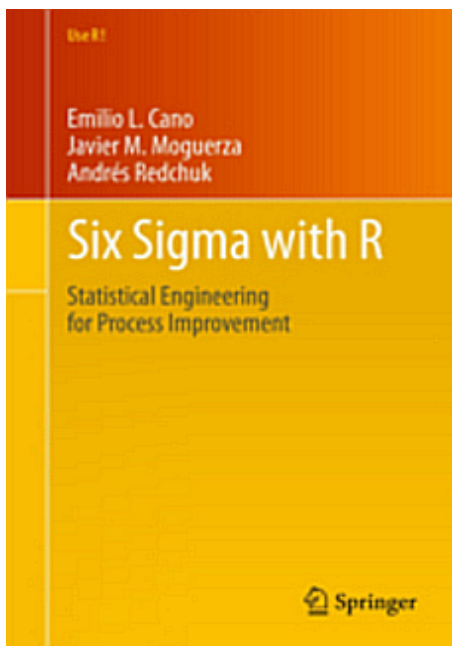
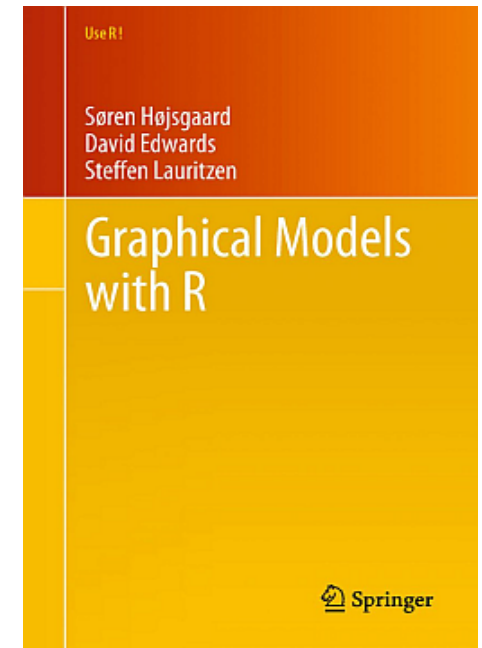
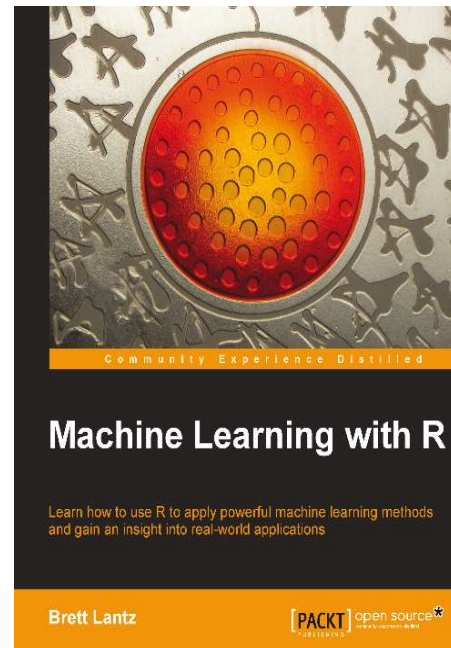
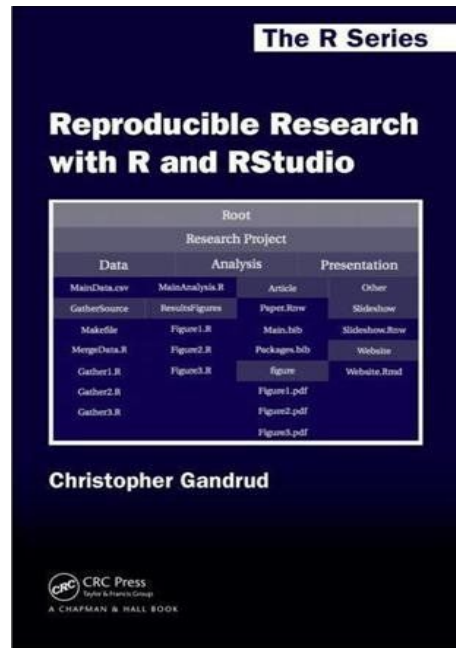
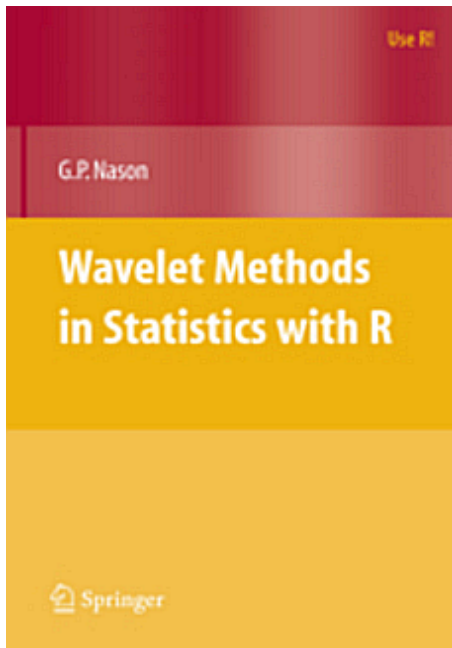


...more foreign books...



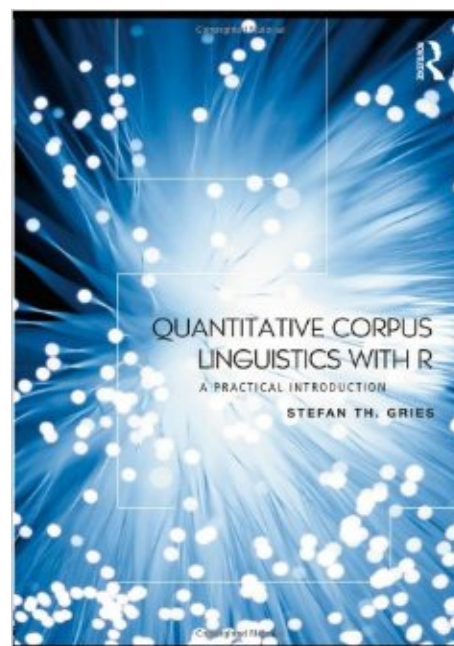
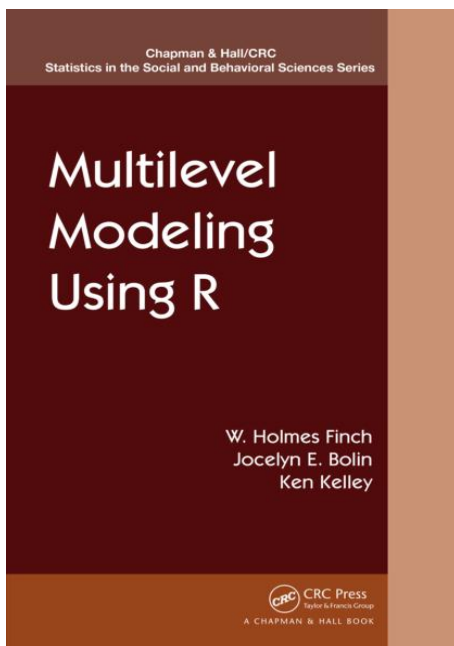
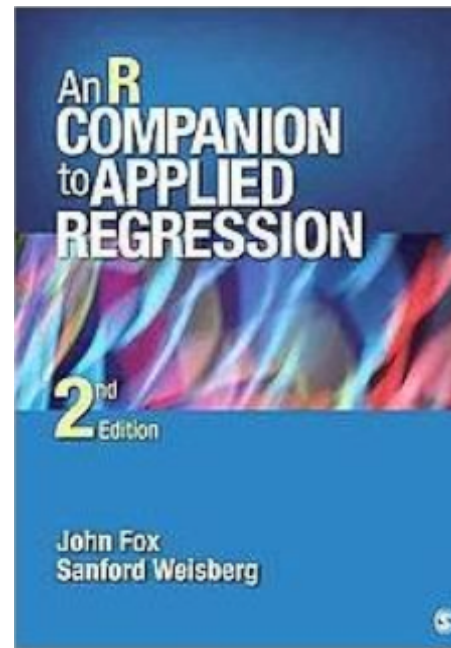
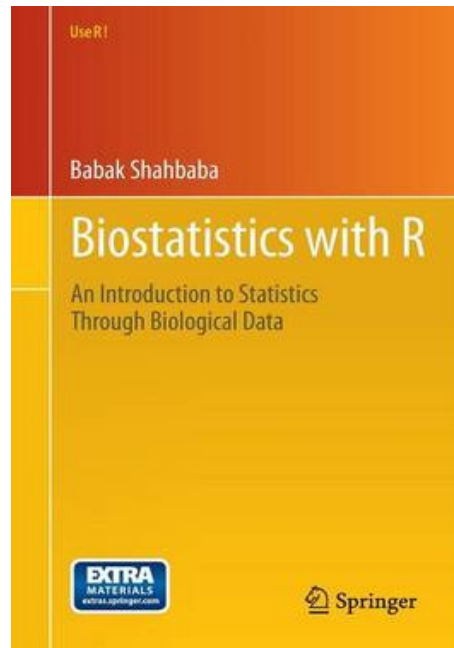
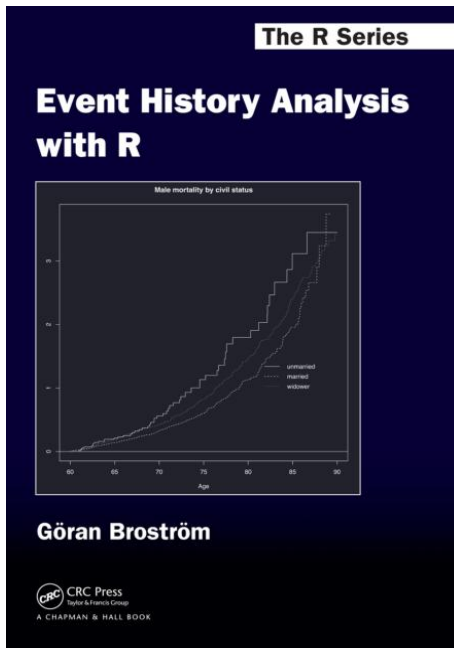


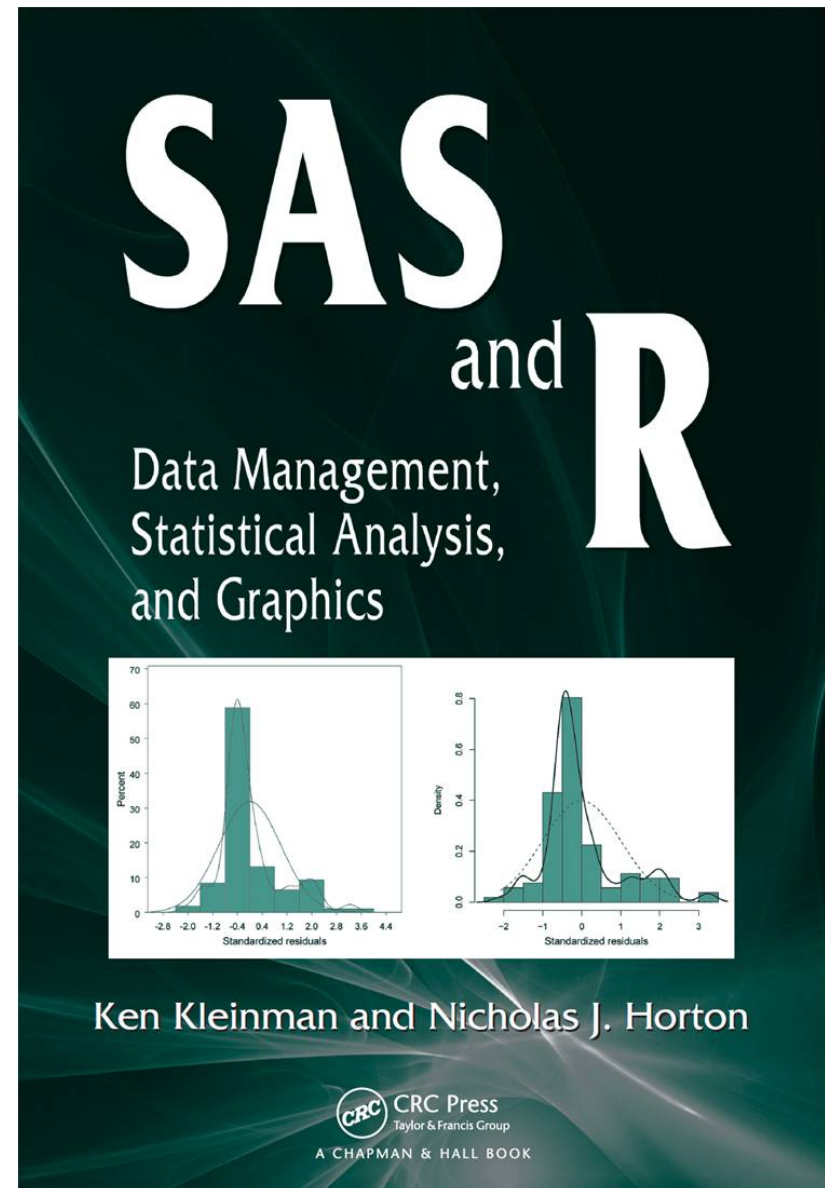
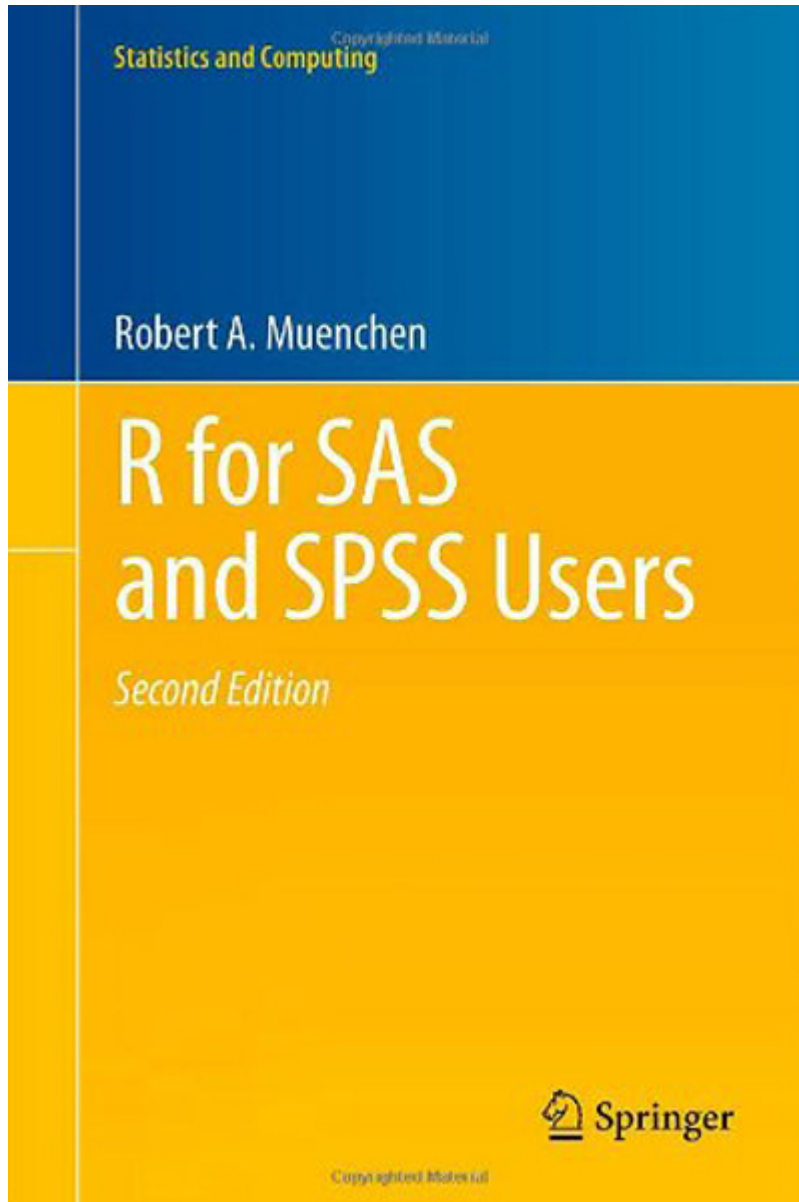
...and more foreign books...





...and more foreign books...







...and even mobile tutorials...

The image displays a series of mobile application screens for R tutorials, organized into a grid. The top row shows the main navigation menu with categories: Home, Videos, Quick Guide, and Which Test?. The middle row lists various topics: Fundamentals (importing data, data classes, ...), I want to... (merge, export, subset, ...), Data summarization (tapply(), functions, arithmetic, ...), and Plotting (Base plots, ggplot2). The bottom row shows Statistical tests (parametric, non-parametric, ...), Plots, Statistical tests, and MANOVA. The MANOVA screen includes a table and R code for performing a MANOVA test.

C	5	5	10
...

```

Code:
Y <- matrix(my.data[2:4], ncol=3) # create a matrix of response variables
grps <- my.data[,1] # Grouping factor
grps <- as.factor(grps) # Converting group to factor

Model1 <- manova(Y ~ grps)
summary(Model1) # Test summary
summary.aov(Model1) # Significance of each variable

Note: There are four common test statistics used in MANOVA: Wilk's lamda, Pillai-Bartlett trace, Hotelling-Lawley trace

```



13 reasons why **you will** love GNU R

- I R is (extremely) cheap. In fact - it's free :)
- II R has (extremely) wide range of capabilities
- III R is (widely) supported by the world of science
- IV R is supported by the community
- V R is (increasingly) supported by the business**
- VI R is able to read data in many formats
- VII Interoperability is easy to achieve
- VIII R is truly cross-platform
- IX R offers numerous ways of presenting data
- X There are many options to optimize the code
- XI R is able to handle large amount of data
- XII R has a set of fancy tools and IDEs
- XIII FDA accepted using R for drug trials!



...supported by the **business**

The biggest tycoons in the market of statistical analysis and data mining recognize the potential of the R and develop specialistic packages as well as provide methods of easy integration between their products and R. These include companies such as [Oracle](#), [StatSoft](#), [IBM](#) (SPSS), [Teradata](#), [Merck](#), [Tibco](#), [Sybase](#), [RapidMiner](#), [SAS](#) and others.

The situation is well described by the two following articles:

1. [Adoption of R by large Enterprise Software Vendors](#)
2. [R integrated throughout the enterprise analytics stack](#)



...supported by the **business**

FC Why The R Programming L... x +

www.fastcompany.com/3030063/why-the-r-programming-language-is-good-for-business Search

WHY THE R PROGRAMMING LANGUAGE IS GOOD FOR BUSINESS

THANKS TO ONE COMPANY, THE SAME CODE THAT IS REVOLUTIONIZING THE SCIENTIFIC COMMUNITY IS NOW MOVING UP THE RANKS OF THE BUSINESS WORLD.

BY TINA AMIRTHA

With terabytes of data at hand, every business is trying to figure out the best way to understand information about their customers and themselves. But simply using Excel pivot tables to analyze such quantities of information is absurd, so many companies use the commercially available tool [SAS](#) to cull business intelligence.

But SAS is no match for the open-source language that pioneering data scientists use in academia, which is simply known as [R](#). The R programming language leans more frequently to the cutting edge of data science, giving businesses the latest data analysis tools. The problem: With loose standards and scores of diverse contributors, it is shaky ground for business. Will that ever change?

THE R EVANGELISTS

At least one company thinks R is ready for commercial prime time. Like RedHat is to [Linux](#) and Cloudera is to [Hadoop](#), [Revolution Analytics](#) is to the R language in the

This story contains interviews with [David Smith](#), chief community officer at Revolution Analytics; [Casey Herron](#), data scientist at Revolution Analytics; [Tess Nesbitt](#), director of analytics at DataSong; and [Solomon Messing](#), data scientist at Facebook.

TRENDING

HAPPENING NOW

- 1 The Top Jobs In 10 Years Might Not Be What You Expect
- 2 The Highest-Paying Jobs Of The Future Will Eat Your Life
- 3 This 28-Year-Old Startup Founder's Life Hack: Don't Waste Time Dating
- 4 Coffee Alternatives That Are Better For Productivity
- 5 How Two Liberal Arts Students Created A Million-Dollar, Rainforest-Saving Tea Company



May 15, 2015

In-database R coming to SQL Server 2016

R is coming to SQL Server. SQL Server 2016 ([which will be in public preview this summer](#)) will include new real-time analytics, automatic data encryption, and the ability to run R within the database itself.

For deeper insights into data, SQL Server 2016 expands its scope beyond transaction processing, data warehousing and business intelligence to deliver advanced analytics as an additional workload in SQL Server with proven technology from Revolution Analytics. We want to make advanced analytics more accessible and increase performance for your advanced analytic workloads by bringing R processing closer to the data and building advanced analytic capabilities right into SQL Server. Additionally, we are building PolyBase into SQL Server, expanding the power to extract value from unstructured and structured data using your existing T-SQL skills. With this wave, you can then gain faster insights through rich visualizations on many devices including mobile applications on Windows, iOS and Android.

With this update, data scientists will no longer need to extract data from SQL server via ODBC to analyze it with R. Instead, you will be able to take your R code to the data, where it will be run inside a sandbox process within SQL Server itself. This eliminates the time and storage required to move the data, and gives you all the power of R and CRAN packages to apply to your database.

At last week's Microsoft Ignite conference in Chicago, SQL Server program managers Lindsey Allen and Borko Novakovic demonstrated a prototype of running R within SQL Server. (A description of the integration begins at 57:00, and the demo at 1:05:00, in the [video](#) below.) In the demo, Lindsey applies a Naive Bayes classification model (from the [e1071 R package](#)) to the famous Iris data, using the same R code used in [this Azure ML Studio experiment](#).

- [About this blog](#)
- [Comments Policy](#)
- [About Categories](#)
- [About the Authors](#)
- [R Community Calendar](#)
- [Local R User Group Directory](#)





Free to download, use, and share.



GET IT NOW

Got comments or suggestions for the blog editor? Email [David Smith](#).

 Follow David on Twitter: [@revodavid](#)
 [+David Smith](#)

[Subscribe to this blog's feed](#)



Revolutions

Learn more about using open source R for big data analysis, predictive modeling, data science and more from the staff of Revolution Analytics.

[« R for more powerful clustering](#) | [Main](#) | [The new science journalism and open science](#) »

April 22, 2015

Microsoft hiring engineers for R projects

Are you a talented software engineer who would like to build out the R ecosystem and help more companies access the power of R? Microsoft ([Revolution Analytics' parent](#)) is [hiring a new team](#) to do just that:

Our mission is to empower enterprises to easily and cost-effectively build high-scale analytics solutions leveraging R.

Exponential growth has transformed data into a new natural resource. Every industry has focused on exploiting data analytics for competitive advantage.

Business applications of advanced analytics abound: consumer companies doing [big data analytics](#), financial institutions using [R for risk management](#), manufacturers using [R for predictive maintenance](#), and many more.

For the uninitiated, R is an open source programming language & environment for statistical computing. More importantly, R is an innovation engine, with applications that run the gamut from quantitative finance to bioinformatics to machine learning. Over the past several years, R has enjoyed tremendous growth in usage & mindshare in the data science community, reaching a user count in the millions.

Information

- [About this blog](#)
- [Comments Policy](#)
- [About Categories](#)
- [About the Authors](#)
- [R Community Calendar](#)
- [Local R User Group Directory](#)



Free to download, use, and share.

OPEN
REVOLUTION R OPEN

GET IT NOW

Got comments or suggestions for the blog editor?



R Technologies from Oracle

Bringing the Power of R to the Enterprise



Oracle has adopted R as a language and environment to support Statisticians, Data Analysts, and Data Scientists in performing statistical data analysis and advanced analytics, as well as generating sophisticated graphics. In addressing the enterprise and the need to analyze Big Data, Oracle provides R integration through four key technologies:

- Oracle R Distribution** Oracle's supported redistribution of open source R, provided as a free download from Oracle, enhanced with dynamic loading of high performance linear algebra libraries.
- Oracle R Enterprise** Integration of R with Oracle Database. A component of the [Oracle Advanced Analytics Option](#). Oracle R Enterprise makes the open source R statistical programming language and environment ready for the enterprise with scalability, performance, and ease of production deployment.
- Oracle R Advanced Analytics for Hadoop** High performance native access to the Hadoop Distributed File System (HDFS) and MapReduce programming framework for R users. Oracle R Advanced Analytics for Hadoop is a component of Oracle Big Data Connectors software suite.
- ROracle** An open source R package, maintained by Oracle and enhanced to use the Oracle Call Interface (OCI) libraries to handle database connections - providing a high-performance, native C-language interface to Oracle Database.

Why Oracle for Advanced Analytics?

If you're an enterprise company, chances are you have your data in an Oracle database. You chose Oracle for its global reputation at providing the best software products (and now engineered systems) to support your organization. Oracle database is known for stellar performance and scalability, and Oracle delivers world class support.

R in the world of **big UNIX machines** :)

Platform	ORE 1.4.1 Documentation	ORE 1.4 Documentation	ORE 1.3.1 Documentation
Windows	<ul style="list-style-type: none"> Client (6M) Supporting (4M) Server (16M) 	<ul style="list-style-type: none"> Client (6M) Supporting (2M) Server (9M) 	<ul style="list-style-type: none"> Client (4M) Supporting (1M) Server (6M)
Linux 64-bit	<ul style="list-style-type: none"> Client (5M) Supporting (4M) Server (102M) 	<ul style="list-style-type: none"> Client (5M) Supporting (1M) Server (91M) 	<ul style="list-style-type: none"> Client (4M) Supporting (1.2M) Server (92M)
AIX 64-bit	<ul style="list-style-type: none"> Client (5M) Supporting (4M) Server (5M) 	<ul style="list-style-type: none"> Client (5M) Supporting (1M) Server (5M) 	<ul style="list-style-type: none"> Client (4M) Supporting (1M) Server (4MB)
Solaris SPARC 64-bit	<ul style="list-style-type: none"> Client (5M) Supporting (5M) Server (5M) 	<ul style="list-style-type: none"> Client (5M) Supporting (1M) Server (5M) 	<ul style="list-style-type: none"> Client (4M) Supporting (1M) Server (4MB)
Solaris x86 64-bit	<ul style="list-style-type: none"> Client (4M) Supporting (4M) Server (5M) 	<ul style="list-style-type: none"> Client (5M) Supporting (1M) Server (5M) 	<ul style="list-style-type: none"> Client (4M) Supporting (1M) Server (4MB)



R feat. **Statistica**

STATISTICA - Workbook1* - [Estimated Ratios of Effective Doses]

File Server Edit View Insert Format Statistics Data Mining Graphs Tools Data Workbook Enterprise Window Help

Arial 10 B I U

DoseResponse.r

```

1 # Dose Response: http://www.bioassay.dk
2
library(drc) # if script fails here, call "install.packages('drc')" i

# dataset: use 'PestSci' data (part of 'drc' package) or 'ActiveDataSet'
# (select Spinach.sta, STATISTICA equivalent of 'PestSci')

DR <- multdrc(SLOPE ~ DOSE, CURVE, data = ActiveDataSet)

# transfer results to STATISTICA

RouteOutput(coefficients(summary(DR)), "Results / Summary", "Parameter:
RouteOutput(coef(DR), name = "Coefficients", header = "DR coef
RouteOutput(residuals(DR), "Residuals")
RouteOutput(anova(DR), "ANOVA: Results") # inherits the default

```

Leerlauf. 12

Workbook1* - Estimated Ratios of Effective Doses

Workbook1*

- R Integration (DoseResponse.r)
 - Results / Summary
 - Coefficients
 - Residuals
 - ANOVA: Results
 - Covariances
 - Log-Likelihood
 - Relative Effective Doses (10% / 50%
 - Estimated Ratios of Effective Doses
 - Plot 1
 - Console

	Estimated Ratios of Effective Doses		
	1	2	3
	Estimate	Std. Error	t-value
1/2:90/10	2025,80475	2311,43649	0,875994108
1/3:90/10	2250,60975	2216,78443	1,01480763
1/4:90/10	2054,87546	1711,04427	1,20036372
1/5:90/10	2191,62442	1834,71378	1,19398701
2/3:90/10	268,438836	268,931697	0,994448921
2/4:90/10	245,092858	208,956099	1,16815378
2/5:90/10	261,40343	224,00382	1,16249549
3/4:90/10	573,898728	472,750484	1,21184165
3/5:90/10	612,09085	506,9784	1,20535875
4/5:90/10	11,546881	2,18720991	4,8220708

Plot 1*

Dose Response Curves

Rate of Oxygen Evolution

Dose

Legend: 1 (black circles), 2 (red triangles), 3 (green pluses), 4 (blue crosses), 5 (cyan diamonds)

Run Analysis or Report

- STATISTICA Enterprise
 - _SampleData
 - Präsentationsbeispiele
 - R-Beispiele
 - DoseResponse
 - Lynx

<http://www.statsoft.de/unternehmen...>



R feat. RapidMiner

Firefox

RapidMiner and R | RapidMiner

rapidminer.com/2013/06/18/rapidminer-an

Google

rapidminer

RapidMiner and R

By Ingo Mierswa Posted June 18th, 2013 In RapidMiner Blog

RapidMiner brings the power of predictive analytics to the business user, with no programming required.

However, analysts sometimes want to add custom code to their analytical processes. This brief presentation reviews how to add custom R scripts to the enterprise-ready environment of RapidMiner, using an R extension.

RapidMiner and R are both widely used. Our R integration offers the best of both worlds in enterprise environments. Here you can also see the R extension in action:

RapidMiner R Extension

Edit Parameter Text: script
This script will be executed on one of the available R servers.

Contact Sales

<http://rapidminer.com/2013/06/18/rapidminer-and-r/>

<http://www.simafore.com/blog/bid/204923/Combining-power-of-R>

Firefox

Combining power of R and RapidMiner for ti...

www.simafore.com/blog/bid/204923/Combining

Google

76	10/1/2013	2923	1522	1985
77	11/1/2013	4173	2503	1293

Configuring *Execute R Script* operator:

There are 3 steps here. First, provide the names of the input variables being sent to R. If you are sending the entire dataset (as a data frame), then type the name of this data frame in the second box of the parameters tab (*inputs: Edit Enumeration*). If you are sending only a few attributes from your data set, each attribute name has to be entered separately as shown below.

Execute Script (R)

script [Edit Text (643 ch...)]

inputs [Edit Enumeration...]

results [Edit List (3)...]

Edit Parameter List: inputs

Edit Parameter List: inputs
This assigns each input port a variable name. If the type of input object is supported by the R translation, it will be accessible under this variable name.

name of variable

WT1

WT2

Date

Add Entry Remove Entry OK Cancel

The second step is to write the R script in the "script: Edit text" which is the first box of the parameter tab. Make sure that you reference the names of the variables selected above exactly (or rename the variables within R, as done below). In this case we are only using the attribute WT1 for forecasting along with the



Firefox

IBM Calling R from SPSS

Using R from SPSS

The R integration plug-in does two things: It opens communication between SPSS and R which to translate SPSS data structures into R objects.

Hello R!

Open a syntax file, and type the following lines. Select and run the command

```
BEGIN PROGRAM R.  
cat("\t\tHello R!\n")  
END PROGRAM.
```

The line `BEGIN PROGRAM R.` launches R and loads the requisite library of functions that override any options that you might set in your `.First()` function.

The first and last lines here follow the conventions of SPSS syntax code and are treated as R code and must obey the rules of R syntax, so no period marks the end of a line.

When SPSS meets the `END PROGRAM.` statement, it interprets subsequent lines as R code and variables that an R chunk creates are available to subsequent R chunks during the same session.

Reading data into R and returning changes to SPSS

R chunks that are called from SPSS can read and write data from external sources and you want access to an SPSS database. I created a simple test database to illustrate the lines in [Listing 1](#).

Listing 1. Read and write a database

```
BEGIN PROGRAM R.  
# Pull the data into a data frame  
testData = spssdata.GetDataFromSPSS()  
  
# Pull the data dictionary into another data frame  
testDict = spssdictionary.GetDictionaryFromSPSS()
```

Firefox

lebebr01/SPSStoR · GitHub

SPSS to R

build passing

A set of functions that takes *SPSS* syntax as input and outputs *R* commands to do the same analysis or data management tasks.

Current Features

- Read in Data
 - Get for sav files
 - Get for csv, txt, xls, xlsx files with get data.
- Data Manipulation
 - Sort Cases
- Descriptives
 - Aggregate
 - Correlations
 - Crosstab
 - Descriptives
 - Frequencies
 - Graphics
- Models
 - One sample t-test
 - Independent sample t-test
 - One-way ANOVA - with oneway routine

Upcoming Features

- Dataset commands
- Value Labels
- Further arguments for descriptives
- Modeling functions
 - t-test (two sample with cut score and paired)
 - analysis of variance - more complicated designs
 - regression
 - generalized models
- if else statements



Firefox

Gretl - R

Econometric analysis of the general linear model with R

Interaction between R and Gretl

Index

[Homepage](#)

[Download and installation of R](#)

[Estimation, validation and analysis the performance of the basic assumptions of heteroscedasticity and uncorrelated of an econometric model](#)

[Interaction between R and Gretl](#)

Gretl has long had the ability to save the current data set in a format suitable for analysis with R, or the option of launching an R session with the current gretl data set automatically loaded into the Rworkspace. However, since version 1.7.5 of Gretl, its interaction with R is much more sophisticated: it is now possible to incorporate R scripts within Gretl scripts and both programs will transfer the data and results in a relatively transparent manner. The following video shows this interaction.

```
$`Normalidad residuos`  
      One-sample Kolmogorov-Smirnov test  
data:  residuos  
D = 0.0925, p-value = 0.9827  
alternative hypothesis: two-sided  
  
$`Heteroscedasticidad residuos`  
Non-constant Variance Score Test  
Variance formula: ~ fitted.values  
Chisquare = 9.53477    Df = 1    p = 0.002016154  
  
$`Autocorrelación residuos`  
      Durbin-watson test  
data:  modelo  
Dw = 2.3038, p-value = 0.7679  
alternative hypothesis: true autocorrelation is greater than 0  
  
$`Multicolinealidad variables`  
Renta  Deuda  Hijos
```





13 reasons why **you will** love GNU R

- I R is (extremely) cheap. In fact - it's free :)
- II R has (extremely) wide range of capabilities
- III R is (widely) supported by the world of science
- IV R is supported by the community

V ½ :) R and SAS

- VI R is able to read data in many formats
- VII Interoperability is easy to achieve
- VIII R is truly cross-platform
- IX R offers numerous ways of presenting data
- X There are many options to optimize the code
- XI R is able to handle large amount of data
- XII R has a set of fancy tools and IDEs
- XIII FDA accepted using R for drug trials!



R feat. SAS

SAS/IML Studio - rtest.sx

File Edit View Program Graph Analysis Tools Window Help

rtest.sx

```
submit / R;
  suppressMessages(library(ggplot2))
  suppressMessages(library(psych))

  data <- data.frame(norm = rnorm(500, mean=10, sd=1))
  describe(data)

  ggplot(data, aes(x=norm)) +
    geom_histogram(aes(y=..density..), binwidth=.5,
                  colour="black", fill="white") +
    geom_density(alpha=.2, fill="#FF6666") +
    ggtitle("Histogram of 'norm'")
endsubmit;
```

R Graphics: Device 2 (ACTIVE)

Plik Historia Zmień rozmiar

Histogram of 'norm'

density

norm

HTML Output

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
norm	1	500	10.02	1	10.02	10.03	1	6.93	13.31	6.38	-0.08	-0.06	0.04

rtest

Ready

0 Error(s) 0 Warning(s)



R feat. SAS

```
RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help
> PrintTextFile("c:/tmp/testcsv.sas")
ods csv file="c:\tmp\test.csv";
ods listing;

proc means data=sashelp.class noprint;
  var age height;
  output out=new(drop=_type_ _freq_);
run;

proc transpose data=new out=out;
  id _stat_;
run;

proc print data=out 1 noobs;
  var _name_ n mean std min max ;
  label mean="Mean"
         std="Std Dev"
         min="Minimum"
         max="Maximum"
         _name_="Variable";
run;
ods _all_ close;
>
> system('"C:/Program Files/SASHome/SASFoundation/9.4/sas.exe"
+ -sysin c:/tmp/testcsv.sas
+ -nosplash
+ -log c:/tmp/test.log
+ -print c:/tmp/test.lst')
>
> (data <- read.csv("c:/tmp/test.csv", header=T, check.names=F))
  Variable  N   Mean Std Dev Minimum Maximum
1     Age 19 13.3158 1.49267   11.0      16
2   Height 19 62.3368 5.12708   51.3      72
> PrintTextFile("c:/tmp/test.lst")

                The SAS System                02:01 Friday, June 12, 2015  1

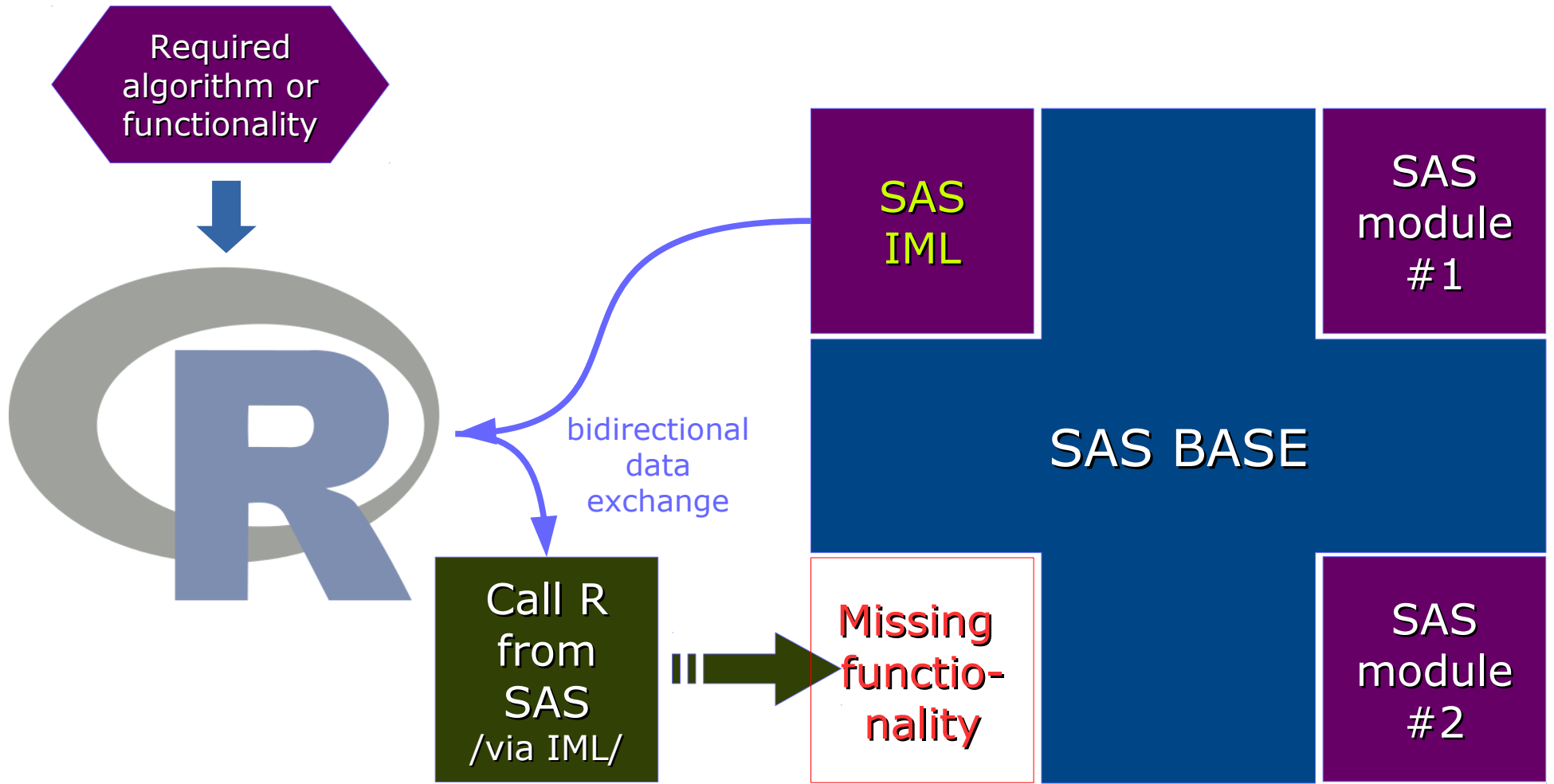
      Variable      N      Mean      Std Dev      Minimum      Maximum
      Age           19    13.3158    1.49267     11.0         16
      Height        19    62.3368    5.12708     51.3         72
>
```

calling SAS
in batch
mode

reading
created CSV



Enhance SAS with R via IML





A set of factors makes the cooperation between SAS and R easier:

- Package [sas7bdat](#) enables R reading SAS [SAS7BDAT](#) datasets without having SAS installed
- Package [SASxport](#) enables R both reading and writing SAS Transport files ([XPT](#)) without having SAS installed.

It is removed from CRAN yet still available in the [CRAN Archive](#). I use this package extensively and have never been disappointed.

- Package [foreign](#) is another option for exchanging data with SAS
- For several statistical methods SAS-compliant variants are created:
 - Contrast: [contr.SAS](#)
 - Quantiles: type 3 → *nearest even order statistic*
 - Type-III Sum of Square is available in R
 - Both SAS and R share the same numbers rounding algorithm
- Both R and SAS can call each other in batch mode from a command line
- R can also be accessed conveniently from SAS via IML module



SAS – R companion

✓ SAS® and R Working Toge... x + <http://www.lexjansen.com/nesug/nesug12/bb/bb10.pdf>

www.lexjansen.com/nesug/nesug12/bb/bb10.pdf Search

Page: 1 of 9 Automatic Zoom

NESUG 2012 **Beyond the Basics**

SAS® and R Working Together

Matthew Cohen, Wharton Research Data Services

ABSTRACT

This paper will explore the ways SAS and R can work together. It will cover data transfers using SAS Transport and ASCII files and how to call R directly from within SAS. Rather than focusing on the pros and cons of each language, I will assume that some people want to use both.

INTRODUCTION

Below is a scenario of how someone might use both SAS and R. Even if this scenario does not apply to you, it's becoming more and more likely that you as a SAS programmer will be asked to provide data to someone using R or receive data from an R user to be integrated back into the SAS environment.

Both SAS and R can perform data management and create subsets. It appears to be more common to do this in SAS. One reason may be that SAS does not need to load the entire dataset into memory before creating the subset, and there may be other reasons as well. This paper will assume the data management is done in SAS.

Both SAS and R can run basic statistical procedures and create graphs. You could argue that one language is better than the other for this, or at least that specific procedures are better on the SAS or R side. But we know for sure that some people prefer to run graphs and/or statistics in R. For purposes of illustration, this paper will assume that the data will be transferred to R at this point: after the data processing is complete and the analysis is ready to start.

Since research is often an iterative process, this paper will assume that the results of the analysis in R must be transferred back to SAS.



Let's MoRtal Kombat begin!

348-2013: The SAS® Versus R Debate in Industry and Academia (SECURED) - Adobe Reader

File Edit View Window Help

1 / 6 95,9%

Tools Sign Comment

SAS Global Forum 2013 Quick Tips

Paper 348-2013

The SAS® Versus R Debate in Industry and Academia

Chelsea Loomis Lofland, University of California, Santa Cruz;
Rebecca Ottesen, California Polytechnic State University, San Luis Obispo

ABSTRACT

Despite industry being heavily dominated by SAS, R is used widely in academia due to being free and open source software that is structured around users being able to write and share their own functions. However, this disconnect leaves many students who are pursuing analytical degrees struggling to get a job with less SAS experience than desired by companies. Alternatively they could face the struggle of transitioning everything they learned in university from R to SAS. Ideally one would know every possible programming language and use the one that best suits the situation. This is rather unrealistic. Our goal is to show the benefits of these two very different software packages and how to leverage both of their strengths together. We also address many of the misconceptions we often see with R users who have left the SAS programming language years ago and have not kept up with current SAS trends and advances.

INTRODUCTION

We chose to focus on SAS and R as they currently dominate the programming language choices in the statistics field. We've noticed that there is a disconcerting trend in that R is being used heavily in Academia which seems to be at odds with what is going on within industry where SAS is primarily utilized. However the mastery of both packages is critical to a young person's success as each language plays an important role in analysis. Therefore taking a one sided preference for either software based on common misconceptions does a disservice to students and needs to be addressed.

There are several updated technologies in SAS that statistical programmers from other languages seem to have yet to discover because they are not keeping current with software. Graphics are a good example of an area SAS has


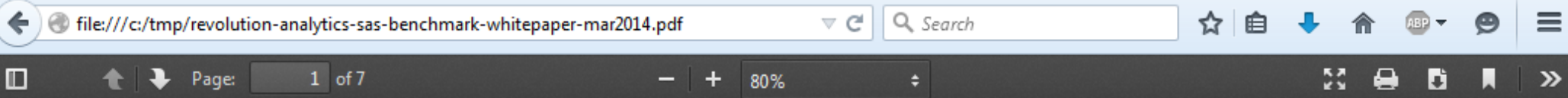


Let's MoR**tal Kombat begin!**

revolution-analytics-sas-bench... x +

file:///c:/tmp/revolution-analytics-sas-benchmark-whitepaper-mar2014.pdf Search

Page: 1 of 7 80%




White Paper

Revolution R Enterprise: Faster Than SAS

Benchmarking Results

by Thomas W. Dinsmore and Derek McCrae Norton





Let's MoRtal Kombat begin!



We performed all benchmark tests sequentially, with no other operations running concurrently.

The actual SAS 9.4 and RRE 7 programs used for this test are freely available to anyone at GitHub: <https://github.com/RevolutionAnalytics/Benchmark>. We invite readers to test these scripts in any environment and compare your results to those we have published below.

Results

The table below shows results for the larger data set of five million records. Using SAS 9.4, the complete script takes on average 5,192 seconds (about 1.5 hours) to complete in the benchmark environment. The same tasks performed in Revolution R Enterprise 7 take 124 seconds (about two minutes) to complete. Table 2 shows the performance of SAS 9.4 and RRE 7 for each of the 10 components of the script.

Table 2: Benchmark Results

Task	Runtime (Seconds)		RRE 7 Speed Multiple
	RRE 7	SAS 9.4	
n = 5,000,000			
Descriptive statistics	1.2	247.3	213X
Median and deciles	1.4	249.6	185X
Frequency distribution	0.8	262.7	350X
Linear regression with 20 numeric predictors	6.8	267.2	39X
Linear regression with 20 mixed predictors	7.3	269.6	37X
Stepwise linear regression, 100 numeric predictors	13.9	262.8	18X
Logistic regression with 20 numeric predictors	16.9	980.7	58X
Generalized linear model, 20 numeric predictors	32.7	573.6	18X
k-means clustering, 20 active variables	10.1	1,025.9	101X
k-means clustering, 100 active variables	32.5	1,053.0	32X
Total, all tasks	123.6	5,192.4	42X



13 reasons why **you will** love GNU R

- I R is (extremely) cheap. In fact - it's free :)
- II R has (extremely) wide range of capabilities
- III R is (widely) supported by the world of science
- IV R is supported by the community
- V R is (increasingly) supported by the business
- VI R is able to read data in many formats**
- VII Interoperability is easy to achieve
- VIII R is truly cross-platform
- IX R offers numerous ways of presenting data
- X There are many options to optimize the code
- XI R is able to handle large amount of data
- XII R has a set of fancy tools and IDEs
- XIII FDA accepted using R for drug trials!



Does **R** speak XYZ ?

R is able to read data in many formats: MS Excel, OpenOffice Calc, Gnumeric, DICOM, SPSS, Weka, Systat, Stata, EpiInfo, **SAS datasets**, **SAS Transport**, Systat, Minitab, Octave, Matlab, DBF (dBase, FoxPro), CSV/TSV, XML, HTML, JSON and...

One can access data stored in various databases via native drivers (MySQL, PostgreSQL, SQLite, MS Access) as well as **ODBC** and **JDBC** interfaces. This gives easy access to any database for which appropriate drivers are available: **Oracle**, **SQL Server**, **MySQL**, **dBase**, **PostgreSQL**, **SQLite**, **DB/2**, **Informix**, **Firebird**, etc.

SCREENING NO. 012345 FIRST NAME S.Y.D. LAST 06/10/2001 SCREENING DATE 03/15/01

DATE INFORMED CONSENT WAS SIGNED 07/23/00

DEMOGRAPHICS

DATE OF BIRTH: 11/24/1945	SEX: Male	RACE: White
---------------------------	-----------	-------------

ALCOHOL / TOBACCO / CAFFEINE USAGE

DOES THE PATIENT USE ALCOHOL? (CHECK ONE)

DOES THE PATIENT USE TOBACCO? (CHECK ONE)

DOES THE PATIENT USE CAFFEINE? (CHECK ONE)

CHILD BEARING POTENTIAL

IF PATIENT IS FEMALE, IS SHE OF REPRODUCTIVE AGE? YES NO

IF YES, SPECIFY THE FORM OF CONTRACEPTION THE PATIENT WILL USE THROUGHOUT THE COURSE OF THE STUDY. CHECK ALL THAT APPLY

Handwritten notes: "this visit in 1963 - STD" and "Occasionally"





Does R speak XYZ ?

```
> read.csv("c:/tmp/data.csv", header=TRUE, sep=";")
```

```
Column.1 Column.2 Column.3
1         1   First 2011-01-03
2         2  Second 2011-01-04
3         3   Third 2011-01-05
4         4  Fourth 2011-01-06
5         5   Fifth 2011-01-07
```

CSV

```
> sqldf::read.csv.sql("c:/tmp/data.csv", sql="SELECT * FROM file f WHERE f.`Column.1` BETWEEN 2 AND 3", sep=";")
```

```
Column.1 Column.2 Column.3
1         2  Second 2011-01-04
2         3   Third 2011-01-05
```

CSV^{SQL}

```
> openxlsx::read.xlsx("c:/tmp/data.xlsx", sheet=1, colNames=TRUE, detectDates = TRUE)
```

```
Column.1 Column.2 Column.3
1         1   First 2011-01-03
2         2  Second 2011-01-04
3         3   Third 2011-01-05
4         4  Fourth 2011-01-06
5         5   Fifth 2011-01-07
```

Excel

```
> data <- read.ods("c:/tmp/data.ods", sheet=1)
```

```
> colnames(data) <- data[1, ]
```

```
> data[-1, ]
```

```
Column 1 Column 2 Column 3
2         1   First 2011-01-03
3         2  Second 2011-01-04
4         3   Third 2011-01-05
5         4  Fourth 2011-01-06
6         5   Fifth 2011-01-0
```

Calc



Does R speak XYZ ?

```
> read.xport("c:/tmp/data.xpt")
```

	COLUMN1	COLUMN2	COLUMN3
1	1	First	2011-01-03
2	2	Second	2011-01-04
3	3	Third	2011-01-05
4	4	Fourth	2011-01-06
5	5	Fifth	2011-01-07

```
> data=read.sas7bdat("c:/tmp/data.sas7bdat")
```

```
> data$COLUMN4 <- as.Date(data$COLUMN3, origin = "1960-01-01")
```

```
> data
```

	COLUMN1	COLUMN2	COLUMN3
1	1	First	2011-01-03
2	2	Second	2011-01-04
3	3	Third	2011-01-05
4	4	Fourth	2011-01-06
5	5	Fifth	2011-01-07

	COLUMN1	COLUMN2	COLUMN3
1	1	First	2011-01-03
2	2	Second	2011-01-04
3	3	Third	2011-01-05
4	4	Fourth	2011-01-06
5	5	Fifth	2011-01-07

SAS
Transport

SAS 7

```
> xml <- xmlParse("c:/tmp/data.xml")
```

```
> xmlToDataFrame(xml, homogeneous = TRUE, collectNames = TRUE, colClasses = c("integer", "character", "character"))
```

```
> data$Column3 <- as.Date(data$Column3)
```

```
> data
```

	Column1	Column2	Column3
1	1	First	2011-01-03
2	2	Second	2011-01-04
3	3	Third	2011-01-05
4	4	Fourth	2011-06-03
5	5	Fifth	2011-01-07

```

Lister - [C:\tmp\data.xml]
File Edit Options Encoding Help
<?xml version="1.0"?>
<root>
  <DATA>
    <Column1>1</Column1>
    <Column2>First</Column2>
    <Column3>2011-01-03T00:00:00.000</Column3>
  </DATA>
  <DATA>
    <Column1>2</Column1>
    <Column2>Second</Column2>
  
```

XPath,
DTD & XSD
supported

XML



Does R speak XYZ ?

```
> readHTMLTable("c:/tmp/data.html", header =T, which = 1)
```

Column 1	Column 2	Column 3
1	1 first	2011-01-03
2	2 second	2011-01-04
3	3 third	2011-01-05
4	4 fourth	2011-01-06
5	5 fifth	2011-01-07

HTML

```
> json
```

```
[1] [{"Column 1": ["1", "2", "3", "4", "5"], "Column 2": ["First", "Second", "Third", "Fourth", "Fifth"], "Column 3": ["2011-01-03", "2011-01-04", "2011-01-05", "2011-01-06", "2011-01-07"]}]"
```

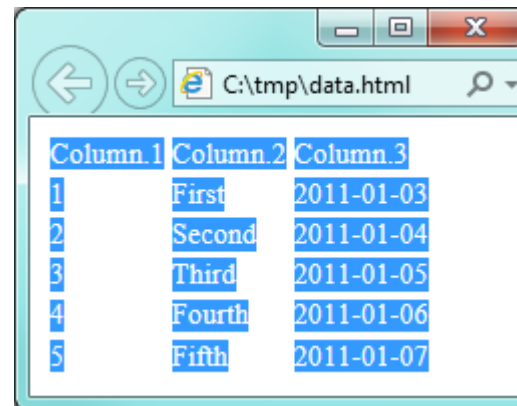
```
> as.data.frame(fromJSON(json))
```

Column.1	Column.2	Column.3
1	1 First	2011-01-03
2	2 Second	2011-01-04
3	3 Third	2011-01-05
4	4 Fourth	2011-01-06
5	5 Fifth	2011-01-07

JSON

```
> read.table(file = "clipboard", sep = " ", header=TRUE) (# with some artefacts)
```

Column.1	Column.2	Column.3	X
1	1 First	2011-01-03	NA
2	2 Second	2011-01-04	NA
3	3 Third	2011-01-05	NA
4	4 Fourth	2011-01-06	NA
5	5 Fifth	2011-01-07	NA



Clipboard



Processing Excel spreadsheets

- [RODBC](#) - in SQL manner
- [XLConnect](#) - requires Java
- [xlsx](#) - requires Java
- [openxlsx](#) - Java-free, fast, requires ZIP
- [gdata](#) - portable, requires PERL
- [RExcel](#) - COM based, licensed
- [By using external code](#)
- From clipboard via `read.table("clipboard")`



13 reasons why **you will** love GNU R

- I R is (extremely) cheap. In fact - it's free :)
- II R has (extremely) wide range of capabilities
- III R is (widely) supported by the world of science
- IV R is supported by the community
- V R is (increasingly) supported by the business

VI 1/3 :) R and relational databases

- VII Interoperability is easy to achieve
- VIII R is truly cross-platform
- IX R offers numerous ways of presenting data
- X There are many options to optimize the code
- XI R is able to handle large amount of data
- XII R has a set of fancy tools and IDEs
- XIII FDA accepted using R for drug trials!



The R and SQL companion

The screenshot shows Microsoft SQL Server Management Studio with the following components:

- Object Explorer:** Shows a database named 'test' with tables 'dbo.RR' and 'dbo.Subject'.
- Database Diagrams:** Shows a diagram 'dbo.ERD_1' with a relationship between 'Subject.ID' and 'RR.SubjID'.
- Table Schemas:**
 - Subject:**

Column Name	Data Type	Allow Nulls
ID	int	<input type="checkbox"/>
Name	nvarchar(50)	<input type="checkbox"/>
Surname	nvarchar(50)	<input type="checkbox"/>
Birth_date	date	<input type="checkbox"/>
 - RR:**

Column Name	Data Type	Allow Nulls
ID	int	<input type="checkbox"/>
SubjID	int	<input type="checkbox"/>
RR_SYS	int	<input type="checkbox"/>
RR_DIA	int	<input type="checkbox"/>
BPM	int	<input checked="" type="checkbox"/>
Time	datetime	<input type="checkbox"/>
- SQLQuery2.sql:**

```
SELECT s.ID, s.Name + ' ' + s.Surname AS Subject,
       rr.Time, rr.RR_SYS, rr.RR_DIA, rr.BPM
FROM Subject s
INNER JOIN RR rr ON s.ID = rr.SubjID
```
- Results:**

ID	Subject	Time	RR_SYS	RR_DIA	BPM
1	Adrian Olszewski	2014-01-02 10:10:00.000	138	78	76
2	Adrian Olszewski	2014-01-02 15:30:00.000	142	82	99
3	Adrian Olszewski	2014-01-02 18:25:00.000	137	78	82
4	Adrian Olszewski	2014-01-03 10:45:00.000	142	88	87
5	Adrian Olszewski	2014-01-03 15:30:00.000	138	74	76
6	Adrian Olszewski	2014-01-04 11:30:00.000	146	91	91
7	Sheryll Smith	2014-01-02 10:20:00.000	127	68	72
- ODBC Data Source Administrator:** Shows 'Test1' as a User DSN using the ODBC Driver 11 for SQL Server.



Querying data sources via ODBC ...

RGui (32-bit)

File Edit View Misc Packages Windows Help

R Console

```
> library(RODBC)
> con <- odbcConnect("test1")
>
> query <- "SELECT s.ID,
+           s.Name + ' ' + s.Surname AS Subject,
+           rr.Time,
+           rr.RR_SYS, rr.RR_DIA, rr.BPM
+ FROM Subject s
+ INNER JOIN RR rr ON s.ID = rr.SubjID"
>
> result <- sqlQuery(con, query)
> odbcClose(con)
>
> result$Time <- as.POSIXct(strptime(result$Time, "%Y-%m-%d %H:%S"))
> (sOlszewski <- subset(result, ID==1))
  ID Subject      Time RR_SYS RR_DIA BPM
1  1 Adrian Olszewski 2014-01-02 10:10:00 138 78 76
2  1 Adrian Olszewski 2014-01-02 15:30:00 142 82 99
3  1 Adrian Olszewski 2014-01-02 18:25:00 137 78 82
4  1 Adrian Olszewski 2014-01-03 10:45:00 142 88 87
5  1 Adrian Olszewski 2014-01-03 15:30:00 138 74 76
6  1 Adrian Olszewski 2014-01-04 11:30:00 146 91 91
>
>
> par(mar=c(6, 4, 4, 2) + 0.5)
> with(sOlszewski, plot(Time, RR_SYS, col="blue", type="o", ylim=c(60, 150)))
> with(sOlszewski, lines(Time, RR_DIA, col="magenta", type="o"))
> title(xlab='Time', line=5)
> grid(nx=NA, ny=NA)
> Sys.setlocale(category = "LC_TIME", locale="C")
[1] "C"
> tck <- axis(1, at=sOlszewski$Time, labels=F)
> text(tck, par("usr")[3], labels=format(sOlszewski$Time, "%b-%d %H:%S"))
> abline(v=tck, col = "lightgray", lty = "dotted", lwd = par("lwd")[1])
> abline(h=axis(2, at=seq(60, 150, 10), labels=F), col = "lightgray", lty = "dotted", lwd = par("lwd")[1])
> legend("bottomright", c("SYS", "DIA"), col=c("blue", "magenta"))
```

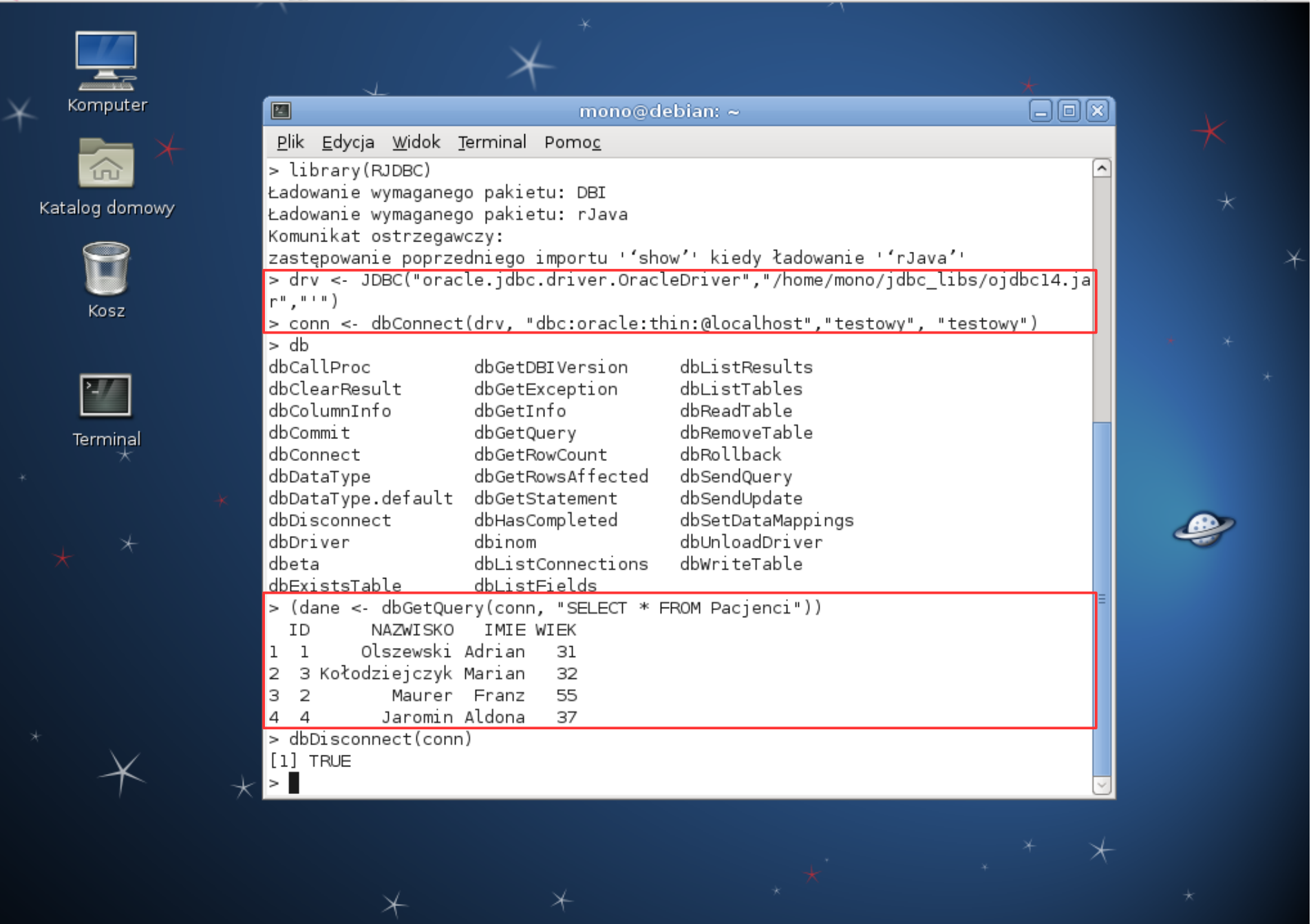
R Graphics: Device 2 (ACTIVE)

Blood pressure of: Olszewski Adrian, #1

Time	SYS	DIA
Jan-02 10:10	138	78
Jan-02 15:30	142	82
Jan-02 18:25	137	78
Jan-03 10:45	142	88
Jan-03 15:30	138	74
Jan-04 11:30	146	91

Time

Legend: SYS (blue), DIA (magenta)



```
mono@debian: ~
Plik Edycja Widok Terminal Pomoc
> library(RJDBC)
Ładowanie wymaganego pakietu: DBI
Ładowanie wymaganego pakietu: rJava
Komunikat ostrzegawczy:
zastępowanie poprzedniego importu 'show' kiedy ładowanie 'rJava'
> drv <- JDBC("oracle.jdbc.driver.OracleDriver", "/home/mono/jdbc_libs/ojdbc14.jar", "")
> conn <- dbConnect(drv, "dbc:oracle:thin:@localhost", "testowy", "testowy")
> db
dbCallProc          dbGetDBIVersion     dbListResults
dbClearResult       dbGetException      dbListTables
dbColumnInfo        dbGetInfo           dbReadTable
dbCommit            dbGetQuery          dbRemoveTable
dbConnect           dbGetRowCount       dbRollback
dbDataType          dbGetRowsAffected   dbSendQuery
dbDataType.default dbGetStatement      dbSendUpdate
dbDisconnect        dbHasCompleted     dbSetDataMappings
dbDriver            dbinom              dbUnloadDriver
dbeta               dbListConnections  dbWriteTable
dbExistsTable       dbListFields
> (dane <- dbGetQuery(conn, "SELECT * FROM Pacjenci"))
  ID  NAZWISKO  IMIE  WIEK
1  1  Olszewski  Adrian  31
2  3  Kołodziejczyk  Marian  32
3  2  Maurer      Franz   55
4  4  Jaromin     Aldona  37
> dbDisconnect(conn)
[1] TRUE
>
```



... or diRect ...

Widok arkusza danych.

```
> library(RODBC)
> con <- odbcConnectAccess("c:/tmp/17sep.mdb")
> data <- sqlQuery(con, "
+ SELECT cntr.COUNTRY,
+       (
+         SELECT COUNT(*) + Int(100 * Rnd)
+         FROM
+         (
+           SELECT DISTINCT sv.COUNTRYID, sv.SITEID &'-'& sv.SUBJID
+           FROM SV sv
+           WHERE sv.VISITID=1
+         ) sv1
+         WHERE sv1.COUNTRYID = cntr.COUNTRY
+       ) AS Screened
+ FROM
+ (
+   SELECT DISTINCT sit.COUNTRY
+   FROM SITES sit
+   WHERE sit.COUNTRY IS NOT NULL
+ ) cntr")
> knitr::kable(data)
```

COUNTRY	Screened
	151
	53
	231
	32
	214
	105
	64

R version 3.0.3 (2014-03-06)



Querying **R** data.frames with **sqldf**

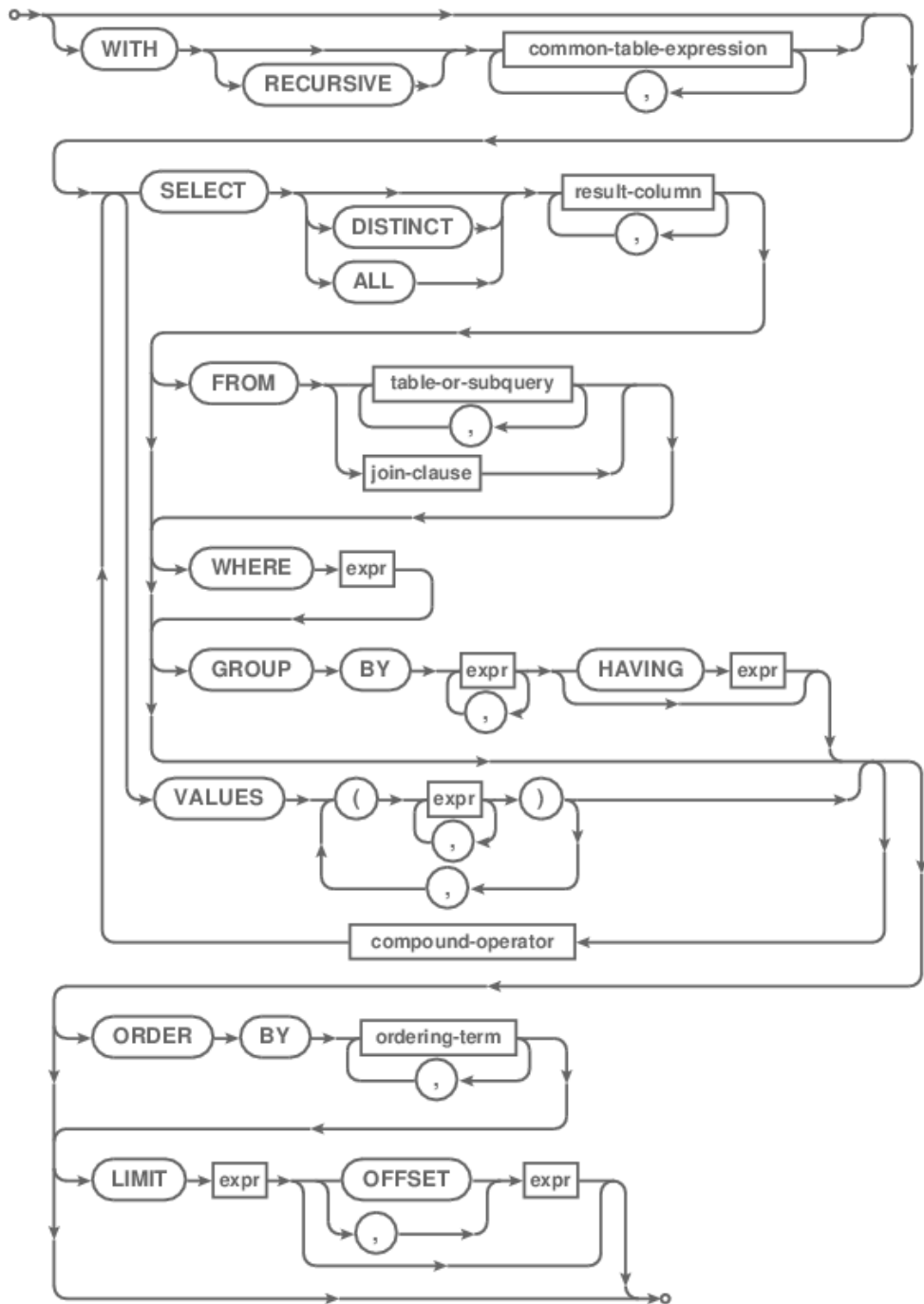
With the [sqldf](#) library data.frames can be queried with a **full blown SQL** like any regular database tables. Available syntax depends on the chosen engine: SQLite ^{default}, H2, PostgreSQL or MySQL.

```
RGui (32-bit) - [R Console]
File Edit View Misc Packages Windows Help
[Icons: Home, Run, Save, Print, Refresh, Stop, Copy]

> library(sqldf)
>
> ( data <- data.frame(ID=1:10, Name=c(rep("John Doe",3), rep("Sue Kowalski", 5),
+ rep("Adrian Olszewski", 2)), SYS=round(runif(10, 120, 150),0), DIA=round(runif(10, 60, 90),0)) )
  ID      Name  SYS  DIA
1  1   John Doe 128   68
2  2   John Doe 143   65
3  3   John Doe 129   76
4  4 Sue Kowalski 147   66
5  5 Sue Kowalski 135   68
6  6 Sue Kowalski 146   87
7  7 Sue Kowalski 137   78
8  8 Sue Kowalski 124   85
9  9 Adrian Olszewski 142  84
10 10 Adrian Olszewski 149  62
>
> sqldf("SELECT ID, d.Name, SUBSTR(d.SYS, 1,3) || '/' || SUBSTR(d.DIA, 1,2) AS BP
+ FROM data d WHERE SYS > 135 ORDER BY d.name, d.SYS DESC")
  ID      Name      BP
1 10 Adrian Olszewski 149/62
2  9 Adrian Olszewski 142/84
3  2      John Doe 143/65
4  4 Sue Kowalski 147/66
5  6 Sue Kowalski 146/87
6  7 Sue Kowalski 137/78
> |
```



sqldf – a full blown SQL at your fingertips!



- ✓ Write any complex SQL queries using your favourite syntax and engine
- ✓ Use complex sub-queries in:
 - ✓ SELECT
 - ✓ FROM / JOIN
 - ✓ WHERE
- ✓ Nest up to 255 levels of sub-queries
- ✓ Use lots of engine-specific functions
- ✓ Limit results with LIMIT / OFFSET
- ✓ Query R data frames with spaces and dots in column names
- ✓ ... and much more!



Do I really need SQL for querying data.frames?

Why would one bother using SQL when R is capable itself of sub-setting (filtering) data.frames, merging them, ordering, summarizing in sub-groups and reshaping? There are various packages that make it easy!

```
> attach(mydata)
> fit <- aov(y~A*B)
> summary(fit)
> layout(matrix(c(1:4), 2))
> plot(fit)
> TukeyHSD(fit)
> interaction.plot(A, B, y)
```

R is perfect for **analysing** data

SQL is ideal for **querying** data

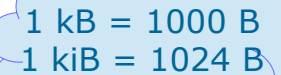
```
SELECT Grp, Sex, COUNT(*)
FROM Table1 t1
INNER JOIN Table2 t2
      ON t1.ID=t2.ID
WHERE t1.Age >=
      (SELECT AVG(t.Age)
       FROM Table1 t
       WHERE t.ID=t1.ID)
GROUP BY Grp, Sex
HAVING COUNT(*) > 1
```

Let everyone do what he does best!



SQLite – it's quite fast!

The default, **SQLite** engine makes **sqldf** really fast.



1 kB = 1000 B
1 kiB = 1024 B

Three simple benchmarks were run. Results are quite satisfying:

- I. Reading a **1.7 GiB** ($4^{\text{columns}} \times 40,000,000^{\text{rows}}$) CSV file into R data.frame
 - via sqldf: **4m:20sec**
 - via data.table::fread, the fastest method available in R: **3m:45sec**

- II. Querying this **~1.1 GiB** dataset.
 - via sqldf: **5.6 sec**
 - via the fastest, native sub-setting method: **7.4 sec**

- III. Querying a **~17 MiB** ($22^{\text{columns}} \times 100,000^{\text{rows}}$) data.frame
 - via sqldf: **0.7 sec** – one won't notice any lags in everyday practice.



SQLite – reading a huge CSV

```
RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help
[Icons: Home, Copy, Paste, Print, Refresh, Stop, Print]

> path <- "bigdf.csv"
> paste( round(file.info(path)$size / (1024*1024*1024),2), "GiB" )
[1] "1.76 GiB"
> library(data.table)
data.table 1.9.4 For help type: ?data.table
*** NB: by=.EACHI is now explicit. See README to restore previous behaviour.
> system.time(data <- fread(path))
Read 40000000 rows and 4 (of 4) columns from 1.756 GB file in 00:03:45
  user system elapsed
113.43   5.51  224.43
> f <- file(path)
> library(sqldf)
Loading required package: gsubfn
Loading required package: proto
Loading required package: RSQLite
Loading required package: DBI

> rm(data)
> system.time(data <- sqldf("select * from f", dbname = tempfile(), file.format = list(header = T, row.names = F)))
Loading required package: tcltk
  user system elapsed
236.68  16.31  255.74
>
```

← fread

← sqldf

SQLite – performance of querying huge dataset

```
RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help
> sqldf() # persistent in-memory session
<SQLiteConnection>
> paste( round(file.info("bigdf.csv")$size / (1024*1024*1024),2), "GiB")
[1] "1.76 GiB"
> f <- file("bigdf.csv")
> system.time(result <- sqldf("SELECT COUNT(*) FROM f WHERE dim = 'k'")) # 1st time it reads the
Loading required package: tcltk
  user system elapsed
202.87    2.73  253.19
> rm(f)
> system.time(result <- sqldf("SELECT COUNT(*) FROM f WHERE dim = 'k'"))
  user system elapsed
  5.57    0.02    5.61
Warning message:
closing unused connection 3 (bigdf.csv)
> system.time(result <- sqldf("SELECT COUNT(*) FROM f WHERE dim = 'k'"))
  user system elapsed
  5.48    0.00    5.52
> fR <- sqldf("SELECT * FROM f") # copying data into R
> object.size(fR)
1120002232 bytes
> system.time(nrow(fR[fR$dim=='k',]))
  user system elapsed
  7.08    0.22    7.39
> system.time(nrow(fR[fR$dim=='k',]))
  user system elapsed
  7.03    0.28    7.32
> system.time(nrow(subset(fR, dim=='k')))
  user system elapsed
  8.19    0.29    8.54
```

reading dataset

querying via sqldf

querying via native R



SQLite – performance of querying typical dataset

```
RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help
[Icons: File Explorer, Print, Save, Copy, Paste, Refresh, Stop, Print]

> dim(data)
[1] 100000 22
> head(data, 1)
      Col1          Col2      Col3
1 1.024371 Some string of a typical length 1.024371
      Col4      Col5          Col6
1 Some string of a typical length 1.024371 Some string of a typical length
      Col7          Col8      Col9
1 1.024371 Some string of a typical length 1.024371
      Col10     Col11          Col12
1 Some string of a typical length 1.024371 Some string of a typical length
      Col13     Col14     Col15
1 1.024371 Some string of a typical length 1.024371
      Col16     Col17          Col18
1 Some string of a typical length 1.024371 Some string of a typical length
      Col19     Col20     Col21
1 1.024371 Some string of a typical length 1.024371
      Col22
1 Some string of a typical length
> paste(round(object.size(data) / (1024^2), 2), "MiB")
[1] "16.79 MiB"
> system.time(sqldf("SELECT COUNT(*) FROM data WHERE Col2 = Col4"))
Loading required package: tcltk
  user  system elapsed
 0.66   0.03   0.74
> system.time(sqldf("SELECT COUNT(*) FROM data WHERE Col2 = Col4"))
  user  system elapsed
 0.61   0.02   0.64
```



...much closer cooperation...

R can be invoked directly from within [PostgreSQL](#) and [Microsoft SQL Server](#):

- In [PostgreSQL](#) one can use the [PL/R language](#) ([example](#))
- In [SQL Server](#) there are three options:
 - write a CLR function which wraps invocations of a chosen R \leftrightarrow .NET “connector”: R.NET, RserveCLI, StatConnector, etc.
 - create and use DCOM objects directly (sp_OACreate)
 - **NEW**: use native T-SQL calls to run R directly from SQL Server

This enables the user to easily perform statistical analyses directly under the database engine and employ complex validation rules in triggers.





With R and a rich set of database drivers (ODBC, JDBC, native) it is easy to transfer data between various sources with only few lines of code.

Actually, this is even easier than in C# :)

```
tables <- list()
tables[["tab1"]] <- read.csv("c:/tmp/table1.csv")           # source 1
tables[["tab2"]] <- read.xlsx("c:/tmp/mysheet.xlsx", sheet=2) # source 2
tables[["tab3"]] <- sqlQuery(otherCon, "SELECT * FROM Table") # source 3
...
myCon <- odbcConnectAccess("c:/tmp/my.mdb")

for(table in names(tables)){
  sqlSave(myCon, tables[[table]], tablename = table)      # that's all!
}
OdbcClose(myCon)
```

Isn't this just beautiful?



13 reasons why **you will** love GNU R

- I R is (extremely) cheap. In fact - it's free :)
- II R has (extremely) wide range of capabilities
- III R is (widely) supported by the world of science
- IV R is supported by the community
- V R is (increasingly) supported by the business

VI 2/3 :) Advanced data manipulation

- VII Interoperability is easy to achieve
- VIII R is truly cross-platform
- IX R offers numerous ways of presenting data
- X There are many options to optimize the code
- XI R is able to handle large amount of data
- XII R has a set of fancy tools and IDEs
- XII FDA accepted using R for drug trials!



Manipulating data

SQL is extremely useful for querying data but it is not the only option the user can choose. [ddplyr](#), [reshape2](#), [tidyr](#) and [data.table](#) libraries provide a rich set of functions for advanced querying and manipulating data structures, i.e.:

- [Filtering](#) (sub-setting) rows
- [Selecting](#) variables (columns)
- [Adding](#) new variables, e.g. computed (mutating and transmutating)
- [Ordering](#) results (sorting)
- [Aggregating](#) (grouping and summarizing)
- [Reshaping](#) data from wide to long format and vice-versa
- [Combining \(piping\) the above operations](#) by the *chain operator* `%>%`

Unquestionably, these packages are one of the top-most useful, “must-have” packages in the analyst's toolbox.



Manipulating data

The chain operator `%>%` is one of the most useful operators in R.

It remarkably facilitates common data processing tasks placing them in a flow.

```
dataset %>%  
  filter(column1 == 2 & column2 %in% c("A", "B", "C")) %>%  
  select(column1, column2, column3) %>%  
  group_by(column1, column2, column3) %>%  
    summarize(Avg = mean(column3), SD = sd(column3)) %>%  
      mutate(CV = SD/Avg) %>%  
        arrange(desc(CV)) %>%  
          merge(., dictionary, by.x="column1", by.y="ID") %>%  
            head(., 10)
```



Querying data with dplyr compared to SQL

```
RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help
[Icons]
> head(diamonds, 2)
  carat  cut color clarity depth table price   x   y   z
1 0.23  Ideal  E    SI2   61.5   55   326 3.95 3.98 2.43
2 0.21 Premium  E    SI1   59.8   61   326 3.89 3.84 2.31
>
> diamonds %>%
+   filter(cut=="Ideal") %>%
+   select(carat, cut, color, price, clarity, depth) %>%
+   mutate(price_per_carat = price/carat) %>%
+   group_by(carat, clarity, depth) %>%
+   summarise(average_price_per_carat=mean(price_per_carat)) %>%
+   data.frame %>%
+   arrange(desc(average_price_per_carat)) %>%
+   head(., 5)
  carat clarity depth average_price_per_carat
1  1.03      IF  62.0             17077.67
2  1.00      IF  60.7             16469.00
3  1.07      IF  60.9             15927.10
4  1.02      IF  63.0             15269.61
5  1.06      IF  61.2             14917.92
>
> sqldf("SELECT carat, clarity, depth, avg(average_price_per_carat)
+       FROM
+       (
+         SELECT t.carat, t.clarity, t.depth, t.price/t.carat AS average_price_per_carat
+         FROM diamonds t
+         WHERE t.cut = 'Ideal'
+       )
+       GROUP BY carat, clarity, depth
+       ORDER BY 4 DESC
+       LIMIT 5")
  carat clarity depth avg(average_price_per_carat)
1  1.03      IF  62.0             17077.67
2  1.00      IF  60.7             16469.00
3  1.07      IF  60.9             15927.10
4  1.02      IF  63.0             15269.61
5  1.06      IF  61.2             14917.92
> |
```

dplyr

sqldf

With dplyr “group-by summaries” are easy!

```
RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help
[Icons]
> head(melanoma)
  time      status sex age year thickness  ulcer
1   10    Died-other M  76 1972     6.76 Present
2   30    Died-other M  56 1968     0.65  Absent
3   35      Alive M  41 1977     1.34  Absent
4   99    Died-other F  71 1968     2.90  Absent
5  185 Died-melanoma M  52 1965    12.08 Present
6  204 Died-melanoma M  28 1971     4.84 Present
> melanoma %>%
+   group_by( status, ulcer, sex ) %>%
+   summarize( n = n(),
+             Missing = sum( is.na( thickness ) ),
+             " "="|",
+             Mean = round( mean( thickness, na.rm = T ), 3 ),
+             SD = round( sd( thickness, na.rm = T ), 3 ),
+             Median = median( thickness, na.rm = T ),
+             Min = min( thickness, na.rm = T ),
+             Max = max( thickness, na.rm = T )
+           )
Source: local data frame [12 x 11]
Groups: status, ulcer

   status  ulcer sex  n Missing  Mean   SD Median  Min  Max
1   Alive  Absent  F  68      0 | 1.693 2.004  1.290 0.10 12.88
2   Alive  Absent  M  24      0 | 1.468 1.719  0.970 0.16  7.09
3   Alive Present  F  23      0 | 2.972 2.593  1.940 0.32 12.24
4   Alive Present  M  19      0 | 4.319 2.423  3.870 0.81  8.38
5 Died-melanoma Absent  F   8      0 | 2.139 1.184  2.020 0.32  3.56
6 Died-melanoma Absent  M   8      0 | 3.266 4.681  1.780 0.81 14.66
7 Died-melanoma Present  F  20      0 | 4.724 4.128  3.540 0.97 17.42
8 Died-melanoma Present  M  21      0 | 5.143 2.862  4.830 1.62 12.88
9   Died-other Absent  F   3      0 | 1.667 1.141  1.450 0.65  2.90
10  Died-other Absent  M   4      0 | 2.420 2.499  1.455 0.65  6.12
11  Died-other Present  F   4      0 | 3.302 3.713  2.255 0.16  8.54
12  Died-other Present  M   3      0 | 8.053 4.019  6.760 4.84 12.56
> |
```




Reshaping long ↔ wide format with reshape2

RGui (64-bit) - [R Console]

File Edit View Misc Packages Windows Help

```
> head(data, 10)
  SiteID SubjID LabTest TestResult
1      1      1     ALAT         42.0
2      1      1     AspAT         55.0
3      1      1     GGTP         18.0
4      1      1     RBC           5.3
5      1      1     WBC           6.2
6      1      2     ALAT         54.0
7      1      2     AspAT         47.0
8      1      2     GGTP         22.0
9      1      2     RBC           4.9
10     1      2     WBC           5.8
```

← source (long)

```
> (wide <- dcast(data, formula = SiteID + SubjID ~ LabTest, value.var="TestResult"))
  SiteID SubjID ALAT AspAT GGTP RBC WBC
1      1      1    42   55  18 5.3 6.2
2      1      2    54   47  22 4.9 5.8
3      1      3    51  NA  21 4.5 NA
4      2      1    59   47  27 4.9 5.1
5      2      2    30  NA  16 5.7 4.2
6      2      3    79   75  31 NA 9.3
```

← long → wide

```
> melt(wide, id.vars=c("SiteID", "SubjID"), variable.name="LabTest", value.name="TestResult") %>% arrange(SiteID, SubjID) %>% head(10)
  SiteID SubjID LabTest TestResult
1      1      1     ALAT         42.0
2      1      1     AspAT         55.0
3      1      1     GGTP         18.0
4      1      1     RBC           5.3
5      1      1     WBC           6.2
6      1      2     ALAT         54.0
7      1      2     AspAT         47.0
8      1      2     GGTP         22.0
9      1      2     RBC           4.9
10     1      2     WBC           5.8
```

← wide → long

Manipulating data

http://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf

www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf

Data Wrangling with dplyr and tidyr

Cheat Sheet



Syntax - Helpful conventions for wrangling

dplyr::tbl_df(iris)

Converts data to tbl class. tbl's are easier to examine than data frames. R displays only the data that fits onscreen:

```
Source: local data frame [100 x 5]
  Sepal.Length Sepal.Width Petal.Length
1             5.1           3.5           1.4
2             4.9           3.0           1.4
3             4.7           3.2           1.3
4             4.6           3.1           1.5
5             5.0           3.0           1.4
Variables not shown: Petal.Width (dbl),
Species (factor)
```

dplyr::glimpse(iris)

Information dense summary of tbl data.

utils::View(iris)

View data set in spreadsheet-like display (note capital V).

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.0	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa

dplyr::%>%

Passes object on left hand side as first argument (or argument) of function on righthand side.

```
x %>% f(y) is the same as f(x, y)
y %>% f(x, .. z) is the same as f(x, y, z)
```

"Piping" with %>% makes code more readable, eg.

```
iris %>%
  group_by(Species) %>%
  summarise(avg = mean(Sepal.Width)) %>%
  arrange(avg)
```

Tidy Data - A foundation for wrangling in R

In a tidy data set:

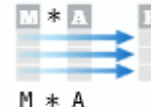


Each variable is saved in its own column



Each observation is saved in its own row

Tidy data complements R's **vectorized operations**. R will automatically preserve observations as you manipulate variables. No other format works as intuitively with R.



Reshaping Data - change the layout of a data set



tidyr::gather(cases, "year", "n", 2:4)
Gather columns into rows.



tidyr::spread(pollution, size, amount)
Spread rows into columns.



tidyr::separate(storms, date, c("y", "m", "d"))
Separate one column into several.



tidyr::unite(data, col, ..., sep)
Unite several columns into one.

dplyr::data_frame(a = 1:3, b = 4:6)
Combine vectors into data frame (optimized).

dplyr::arrange(mtcars, mpg)
Order rows by values of a column (low to high).

dplyr::arrange(mtcars, desc(mpg))
Order rows by values of a column (high to low).

dplyr::rename(tb, y = year)
Rename the columns of a data frame.

Subset Observations (Rows)



dplyr::filter(iris, Sepal.Length > 7)

Extract rows that meet logical criteria.

dplyr::distinct(iris)

Remove duplicate rows.

dplyr::sample_frac(iris, 0.5, replace = TRUE)

Randomly select fraction of rows.

dplyr::sample_n(iris, 10, replace = TRUE)

Randomly select n rows.

dplyr::slice(iris, 10:15)

Select rows by position.

dplyr::top_n(storms, 2, date)

Select and order top n entries (by group if grouped data).

Subset Variables (Columns)



dplyr::select(iris, Sepal.Width, Petal.Length, Species)

Select columns by name or helper function.

Helper functions for select - ?select

select(iris, contains(" "))

Select columns whose name contains a character string.

select(iris, ends_with("Length"))

Select columns whose name ends with a character string.

select(iris, everything())

Select every column.

select(iris, matches(".*"))

Select columns whose name matches a regular expression.

select(iris, num_range("x", 1:5))

Select columns named x1, x2, x3, x4, x5.

select(iris, one_of(c("Species", "Genus")))

Select columns whose names are in a group of names.

select(iris, starts_with("Sepal"))

Select columns whose name starts with a character string.

select(iris, Sepal.Length:Petal.Width)

Select all columns between Sepal.Length and Petal.Width (inclusive).

select(iris, -Species)

Select all columns except Species.

Logic in R - ?Comparison, ?base::Logic

<	Less than	!=	Not equal to
>	Greater than	%in%	Group membership
==	Equal to	is.na	Is NA
<=	Less than or equal to	!is.na	Is not NA
>=	Greater than or equal to	%, , !, xor, any, all	Boolean operators

Manipulating data

http://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf

www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf

Page: 2 of 2

Page Fit

Summarise Data



- dplyr::summarise(iris, avg = mean(Sepal.Length))**
Summarise data into single row of values.
- dplyr::summarise_each(iris, funs(mean))**
Apply summary function to each column.
- dplyr::count(iris, Species, wt = Sepal.Length)**
Count number of rows with each unique value of variable (with or without weights).



Summarise uses **summary functions**, functions that take a vector of values and return a single value, such as:

- | | |
|---|--|
| dplyr::first
First value of a vector. | min
Minimum value in a vector. |
| dplyr::last
Last value of a vector. | max
Maximum value in a vector. |
| dplyr::nth
Nth value of a vector. | mean
Mean value of a vector. |
| dplyr::n
of values in a vector. | median
Median value of a vector. |
| dplyr::n_distinct
of distinct values in a vector. | var
Variance of a vector. |
| IQR
IQR of a vector. | sd
Standard deviation of a vector. |

Group Data

- dplyr::group_by(iris, Species)**
Group data into rows with the same value of Species.
- dplyr::ungroup(iris)**
Remove grouping information from data frame.
- iris %>% group_by(Species) %>% summarise(...)**
Compute separate summary row for each group.



Make New Variables



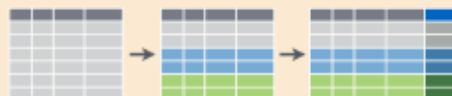
- dplyr::mutate(iris, sepal = Sepal.Length + Sepal.Width)**
Compute and append one or more new columns.
- dplyr::mutate_each(iris, funs(min_rank))**
Apply window function to each column.
- dplyr::transmute(iris, sepal = Sepal.Length + Sepal.Width)**
Compute one or more new columns. Drop original columns.



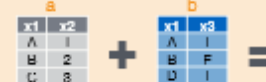
Mutate uses **window functions**, functions that take a vector of values and return another vector of values, such as:

- | | |
|---|--|
| dplyr::lead
Copy with values shifted by 1. | dplyr::cumall
Cumulative all |
| dplyr::lag
Copy with values lagged by 1. | dplyr::cumany
Cumulative any |
| dplyr::dense_rank
Ranks with no gaps. | dplyr::cummean
Cumulative mean |
| dplyr::min_rank
Ranks. Ties get min rank. | cumsum
Cumulative sum |
| dplyr::percent_rank
Ranks rescaled to [0, 1]. | cummax
Cumulative max |
| dplyr::row_number
Ranks. Ties got to first value. | cummin
Cumulative min |
| dplyr::ntile
Bin vector into n buckets. | cumprod
Cumulative prod |
| dplyr::between
Are values between a and b? | pmax
Element-wise max |
| dplyr::cume_dist
Cumulative distribution. | pmin
Element-wise min |

- iris %>% group_by(Species) %>% mutate(...)**
Compute new variables by group.



Combine Data Sets



Mutating Joins

x1	x2	x3
A	1	1
B	2	2
C	3	NA

- dplyr::left_join(a, b, by = "x1")**
Join matching rows from b to a.

- dplyr::right_join(a, b, by = "x1")**
Join matching rows from a to b.

- dplyr::inner_join(a, b, by = "x1")**
Join data. Retain only rows in both sets.

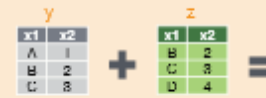
- dplyr::full_join(a, b, by = "x1")**
Join data. Retain all values, all rows.

Filtering Joins

x1	x2
A	1
B	2
C	3

- dplyr::semi_join(a, b, by = "x1")**
All rows in a that have a match in b.

- dplyr::anti_join(a, b, by = "x1")**
All rows in a that do not have a match in b.



Set Operations

y	z
A	1
B	2
C	3
D	4

- dplyr::intersect(y, z)**
Rows that appear in both y and z.

- dplyr::union(y, z)**
Rows that appear in either or both y and z.

- dplyr::setdiff(y, z)**
Rows that appear in y but not z.

Binding

x1	x2
A	1
B	2
C	3
D	4

- dplyr::bind_rows(y, z)**
Append z to y as new rows.

- dplyr::bind_cols(y, z)**
Append z to y as new columns.
Caution: matches rows by position.



data.frame → data.table

[data.table](#) is a library providing a significant enhancement of the regular R data.frame. It allows the user to perform **really fast** (indexing) in-memory processing of **huge** volumes of data (>100 GB) with a relatively easy syntax.

It covers the following tasks:

- **filtering** rows and **selecting** columns
- **adding and deleting** (mutating) columns using no intermediate copies at all
- **joining** tables by indexes
- **aggregating** (calculations in sub-groups)
- **reading** huge CSV files – this is the fastest method currently available in R
- **indexing** selected content (set of columns) of a data.table

Data.tables are compatible with data.frames with some exceptions ([FAQ 2.17](#)).

With data.table summaries are easy too!

```
RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help
> head(data, 20)
  SiteID SubjID TestID TestResult SiteName TestName
1:     1     1     1      42.0   Site 1     ALAT
2:     1     1     2      55.0   Site 1     AspAT
3:     1     1     3      18.0   Site 1     GGTP
4:     1     1     4       5.3   Site 1     RBC
5:     1     1     5       6.2   Site 1     WBC
6:     1     2     1      54.0   Site 1     ALAT
7:     1     2     2      47.0   Site 1     AspAT
8:     1     2     3      22.0   Site 1     GGTP
9:     1     2     4       4.9   Site 1     RBC
10:    1     2     5       5.8   Site 1     WBC
11:    1     3     1      51.0   Site 1     ALAT
12:    1     3     2       NA    Site 1     AspAT
13:    1     3     3      21.0   Site 1     GGTP
14:    1     3     4       4.5   Site 1     RBC
15:    1     3     5       NA    Site 1     WBC
16:    2     1     1      59.0   Site 2     ALAT
17:    2     1     2      47.0   Site 2     AspAT
18:    2     1     3      27.0   Site 2     GGTP
19:    2     1     4       4.9   Site 2     RBC
20:    2     1     5       5.1   Site 2     WBC
> data[, .( .N,
+         Min = min(TestResult, na.rm=TRUE),
+         Avg = round(mean(TestResult, na.rm=TRUE), 1),
+         Med=median(TestResult, na.rm=TRUE),
+         Max=max(TestResult, na.rm=TRUE)),
+       by=.(SiteName, TestName)][, SomeFlag := Med > Avg,
+                                by=.(SiteName, TestName)][SomeFlag==TRUE]
  SiteName TestName N  Min  Avg  Med  Max  SomeFlag
1: Site 1     ALAT 3 42.0 49.0 51.0 54.0     TRUE
2: Site 1     GGTP 3 18.0 20.3 21.0 22.0     TRUE
3: Site 2     ALAT 3 30.0 56.0 59.0 79.0     TRUE
4: Site 2     GGTP 3 16.0 24.7 27.0 31.0     TRUE
5: Site 2     RBC  3  4.9  5.3  5.3  5.7     TRUE
>
```

grouping,
mutating,
filtering



data.table feat. dplyr

```

RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help
[Icons: Home, Copy, Paste, Undo, Redo, Stop, Print]

> head(data, 3)
  SiteID SubjID TestID TestResult
1:     1      1      1          42
2:     1      2      1          54
3:     1      3      1          51

> head(sites)
  SiteID SiteName
1:     1   Site 1
2:     2   Site 2

> head(tests)
  TestID TestName
1:     1    ALAT
2:     2   AspAT
3:     3   GGTP
4:     4    RBC
5:     5    WBC

> class(data)
[1] "data.table" "data.frame"

> setkey(data, SiteID); setkey(tests, TestID); setkey(sites, SiteID);
> tables()
  NAME  NROW NCOL MB COLS          KEY
[1,] data   30   4  1 SiteID,SubjID,TestID,TestResult SiteID
[2,] sites   2   2  1 SiteID,SiteName           SiteID
[3,] tests   5   2  1 TestID,TestName           TestID
Total: 3MB

> data.table(data[sites], key="TestID") %>% .[tests] %>% arrange(SiteID, SubjID, TestID) %>%
+ (function(x, y){ rbind(head(x, 5), tail(y, 5)) })(x=., y=.)
  SiteID SubjID TestID TestResult SiteName TestName
1:     1      1      1      42.0   Site 1    ALAT
2:     1      1      2      55.0   Site 1    AspAT
3:     1      1      3      18.0   Site 1    GGTP
4:     1      1      4       5.3   Site 1    RBC
5:     1      1      5       6.2   Site 1    WBC
6:     2      3      1      79.0   Site 2    ALAT
7:     2      3      2      75.0   Site 2    AspAT
8:     2      3      3      31.0   Site 2    GGTP
9:     2      3      4       NA     Site 2    RBC
10:    2      3      5       9.3   Site 2    WBC

```

3 x indexed data.table

content of a data.table storage

merging, chaining, anonym. fn.



Manipulating data

Valuable resources:

- <http://seananderson.ca/2013/10/19/reshape.html>
- <http://seananderson.ca/2014/09/13/dplyr-intro.html>
- <http://www.sharpsightlabs.com/dplyr-intro-data-manipulation-with-r>
- <http://cran.rstudio.com/web/packages/dplyr/vignettes/introduction.html>
- <http://www.statsblogs.com/2014/02/10/how-dplyr-replaced-my-most...>
- http://www.cookbookr.com/Manipulating_data/Converting_data_between...
- <http://datatable.r-forge.r-project.org/datatable-faq.pdf>
- <http://s3.amazonaws.com/assets.datacamp.com/img/blog/data+table...>
- <http://github.com/Rdatatable/data.table/wiki/Benchmarks-%3A-Grouping>



13 reasons why **you will** love GNU R

- I R is (extremely) cheap. In fact - it's free :)
- II R has (extremely) wide range of capabilities
- III R is (widely) supported by the world of science
- IV R is supported by the community
- V R is (increasingly) supported by the business
- VI R is able to read data in many formats
- VII Interoperability is easy to achieve**
- VIII R is truly cross-platform
- IX R offers numerous ways of presenting data
- X There are many possibilities to optimize the code
- XI R is able to handle large amount of data
- XII R has a set of fancy tools and IDEs
- XIII FDA accepted using R for drug trials!



“R” stands for inteRoperability

R is not only a great statistical package. It is often used as a part of more complex systems as a [computing engine](#). It may also function as a standalone computational server through the Web.

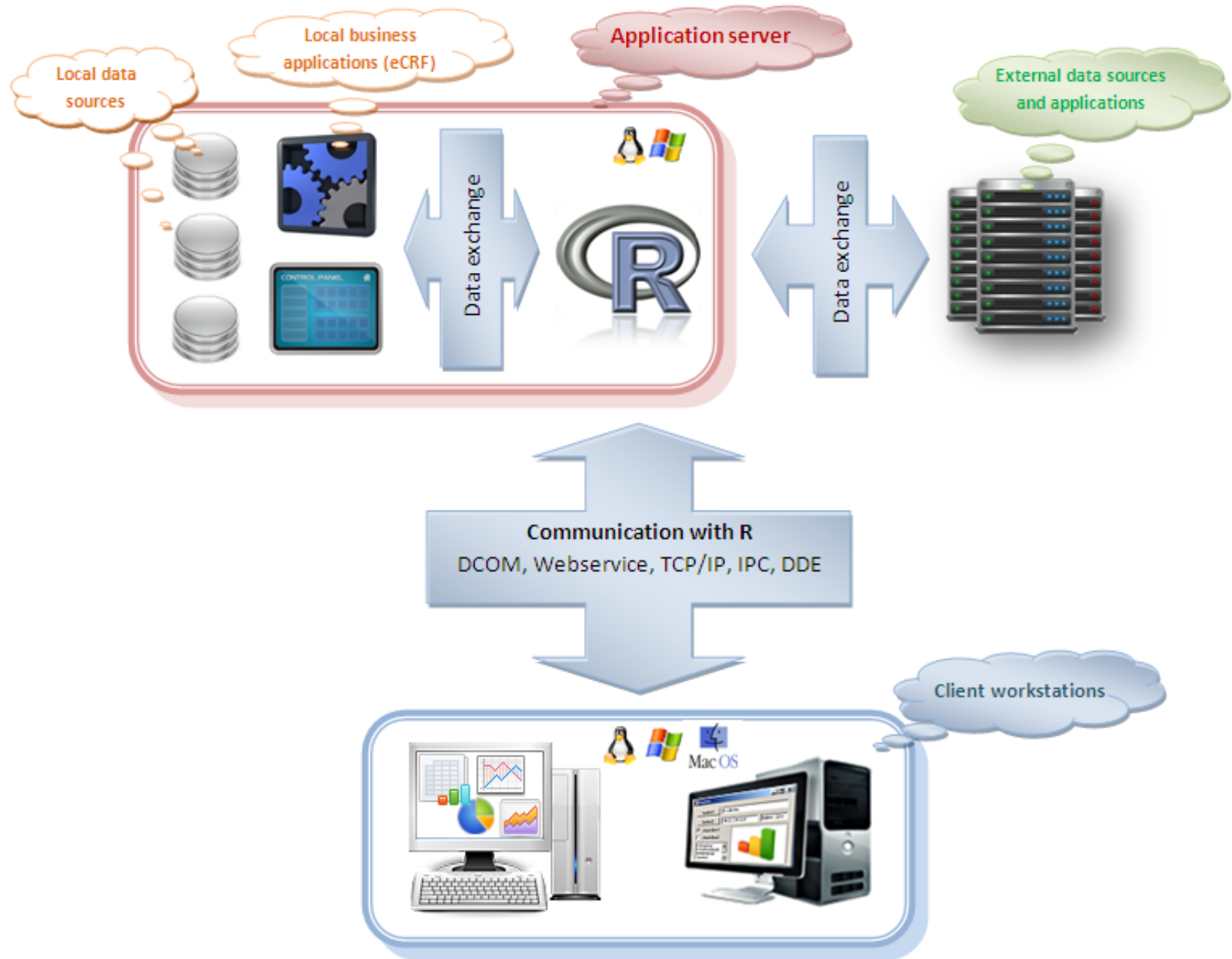
There are number of libraries that allow to communicate with R in many ways ([COM](#), [TCP/IP](#), [WebServices](#), [DDE](#), [direct linking](#)) from many programming languages and platforms ([C++](#), [Java](#), [.NET/Mono](#), [Perl](#), [Python](#), [Scala](#), ...) or external applications able to act as a COM, DDE or Webservice client.

It's worth noting that R is able to call [C/C++](#), Java, .NET, Perl and Python code.

In addition, R allows to create embedded [graphical user interfaces](#) (GUI) for easier entering data and performing analyzes by people not familiar with R. One can make use of the [GTK](#), [Tcl/Tk](#), [wxWidgets](#) and [gWidgets](#) toolkits or the [R GUI Generator](#).

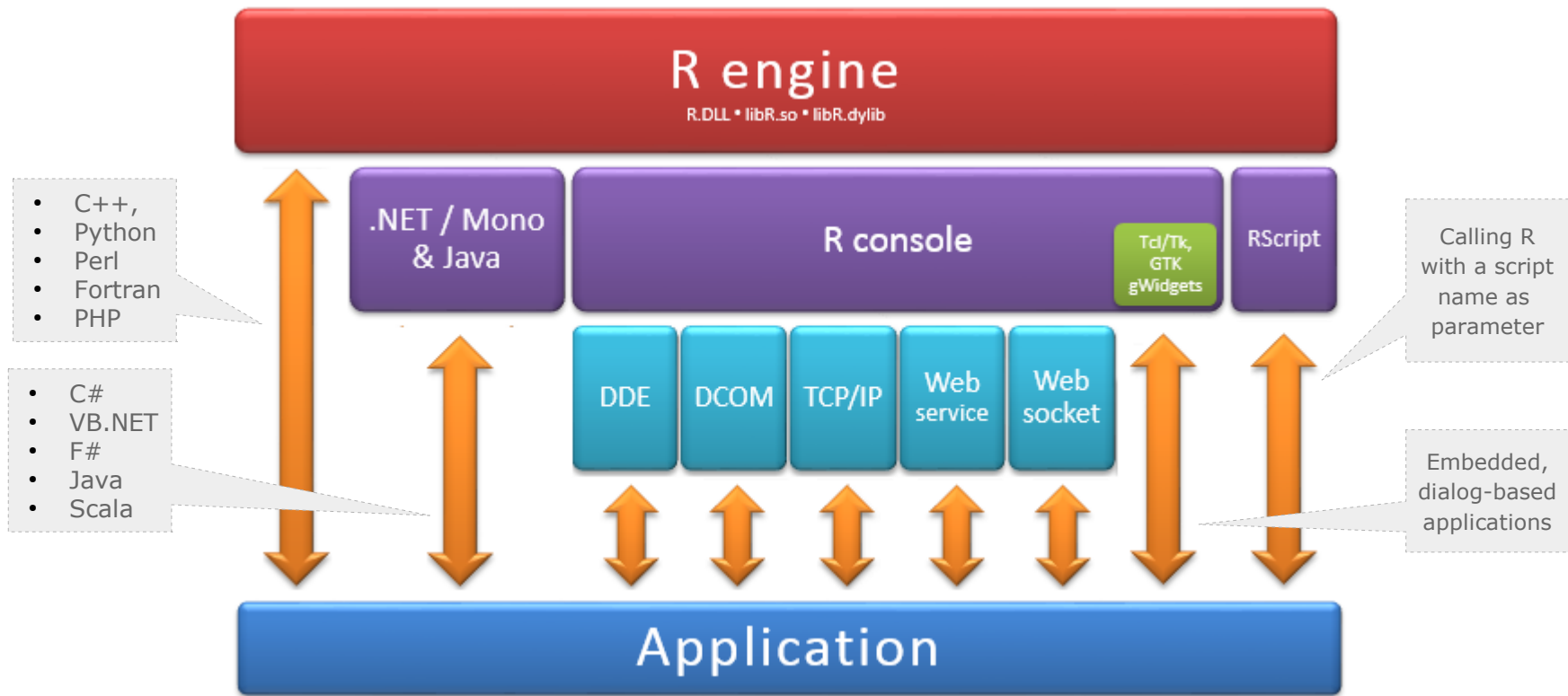


“R” stands for inteRoperability





...”there are so many ways of speaking”



<http://www.rforge.net/rJava>

<http://rpy.sourceforge.net>

<http://rclr.codeplex.com>

<http://rservecli.codeplex.com>

<http://rcom.univie.ac.at>

<http://ndde.codeplex.com>

<http://www.rforge.net/Rserve>

<http://www.rcpp.org>

<http://rdotnet.codeplex.com>

<http://dahl.byu.edu/software/jvmr>

<http://www.rstudio.com/shiny>

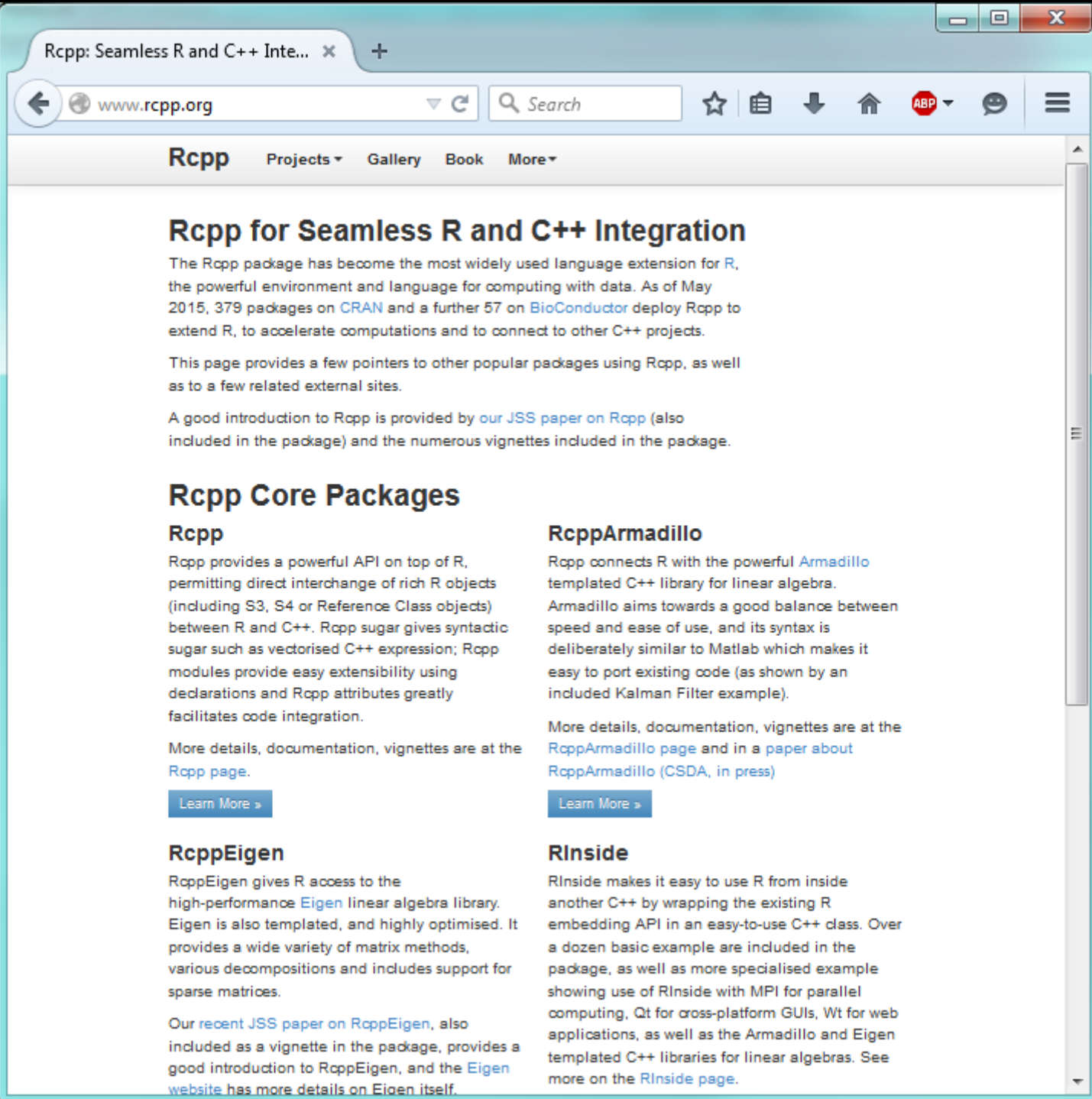
<http://rpython.r-forge.r-project.org>

<http://www.omegahat.org/RSPerl>

<http://www.goldenline.pl/forum/2478242/nowy-interfejs-r-net/> ^{PL}



Easy R – C++ integration



The screenshot shows a web browser window with the URL www.rcpp.org. The page title is "Rcpp: Seamless R and C++ Integration". The main heading is "Rcpp for Seamless R and C++ Integration". The text describes the Rcpp package as the most widely used language extension for R, with 379 packages on CRAN and 57 on BioConductor as of May 2015. It provides pointers to other packages and external sites. Below this, there are sections for "Rcpp Core Packages" including Rcpp, RcppArmadillo, RcppEigen, and RInside, each with a brief description and a "Learn More" button.

Rcpp for Seamless R and C++ Integration

The Rcpp package has become the most widely used language extension for R, the powerful environment and language for computing with data. As of May 2015, 379 packages on [CRAN](#) and a further 57 on [BioConductor](#) deploy Rcpp to extend R, to accelerate computations and to connect to other C++ projects.

This page provides a few pointers to other popular packages using Rcpp, as well as to a few related external sites.

A good introduction to Rcpp is provided by our [JSS paper on Rcpp](#) (also included in the package) and the numerous vignettes included in the package.

Rcpp Core Packages

Rcpp

Rcpp provides a powerful API on top of R, permitting direct interchange of rich R objects (including S3, S4 or Reference Class objects) between R and C++. Rcpp sugar gives syntactic sugar such as vectorised C++ expression; Rcpp modules provide easy extensibility using declarations and Rcpp attributes greatly facilitates code integration.

More details, documentation, vignettes are at the [Rcpp page](#).

[Learn More >](#)

RcppArmadillo

Rcpp connects R with the powerful [Armadillo](#) templated C++ library for linear algebra. Armadillo aims towards a good balance between speed and ease of use, and its syntax is deliberately similar to Matlab which makes it easy to port existing code (as shown by an included Kalman Filter example).

More details, documentation, vignettes are at the [RcppArmadillo page](#) and in a [paper about RcppArmadillo](#) (CSDA, in press)

[Learn More >](#)

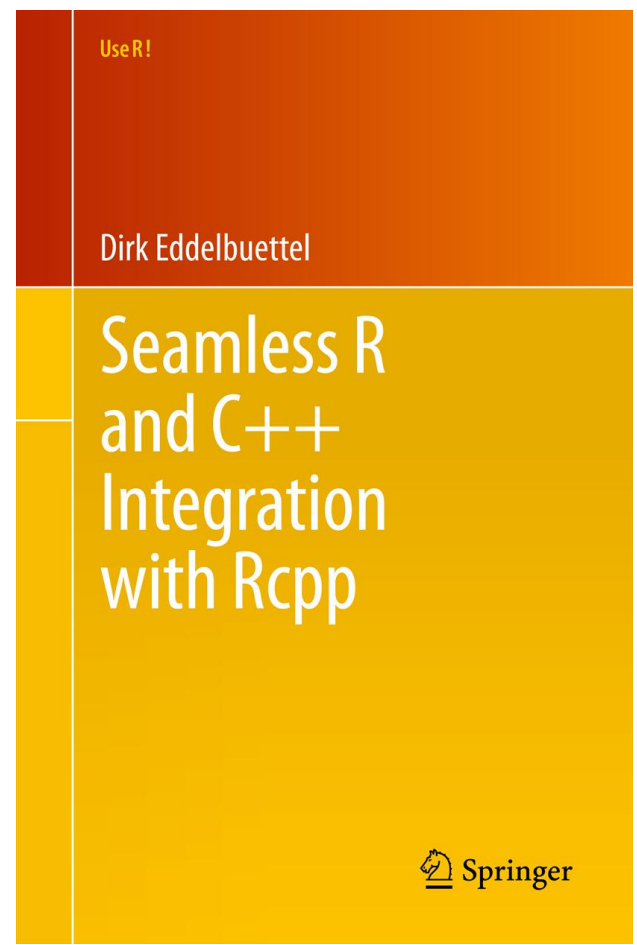
RcppEigen

RcppEigen gives R access to the high-performance [Eigen](#) linear algebra library. Eigen is also templated, and highly optimised. It provides a wide variety of matrix methods, various decompositions and includes support for sparse matrices.

Our [recent JSS paper on RcppEigen](#), also included as a vignette in the package, provides a good introduction to RcppEigen, and the [Eigen website](#) has more details on Eigen itself.

RInside

RInside makes it easy to use R from inside another C++ by wrapping the existing R embedding API in an easy-to-use C++ class. Over a dozen basic example are included in the package, as well as more specialised example showing use of RInside with MPI for parallel computing, Qt for cross-platform GUIs, Wt for web applications, as well as the Armadillo and Eigen templated C++ libraries for linear algebras. See more on the [RInside page](#).




The book cover features a red top section with the text "Use R!" and the author's name "Dirk Eddelbuettel". The main title "Seamless R and C++ Integration with Rcpp" is written in white on a yellow background. The Springer logo is in the bottom right corner.

Use R!

Dirk Eddelbuettel

Seamless R and C++ Integration with Rcpp

 Springer



Various R bindings

R.NET

CodePlex Project Hosting for Open Source Software

Register | Sign In | Search all projects

HOME | SOURCE CODE | DOWNLOADS | DOCUMENTATION | DISCUSSIONS | ISSUES | PEOPLE | LICENSE

Page Info | Change History (all pages) | Follow (252) | Subscribe

Project Description: R.NET enables the .NET Framework to interoperate with the R statistical language in the same process. R.NET requires .NET Framework 4 and the native R DLLs installed with the R environment. R.NET works on Windows, Linux and MacOS. Enjoy statistics and programming in your special language with R.

You should head for the [documentation](#) to get started with R.NET.

News

2014-12-18 Release of the R package rClr 0.7-2 at [rClr](#). This is a significant release due to the fact that all unit tests that can pass using Mono indeed pass.

2014-11-16 R.NET source code reference repository is now hosted on [GitHub](#). This has been requested by several contributors amongst other things to facilitate pull requests.

2014-06-17 Thanks to [Evelina Gabasova](#) for documenting [Setting up R.NET on Mac](#)

2014-05-23 Version 1.5.13 is available on [nuget.org](#). Note that the package identifier is R.NET.Community

NuGet Packages

Search Wiki & Documentation

CURRENT R.NET 1.5.19 (NuGet)
DATE Fri Dec 5, 2014 at 9:00 AM
STATUS Beta
DOWNLOADS 0
RATING ★★★★★ 6 ratings
[Review this release](#)

MOST HELPFUL REVIEWS

★★★★★ Great library! I'm using it often and since its first releases. A real life saver!

★★★★☆ Very useful. A bit of a learning curve, but not too steep.

[View all reviews](#)

RserveCLI

CodePlex Project Hosting for Open Source Software

Register | Sign In | Search all projects

HOME | SOURCE CODE | DOWNLOADS | DOCUMENTATION | DISCUSSIONS | ISSUES | PEOPLE | LICENSE

Page Info | Change History (all pages) | Follow (13) | Subscribe

We have moved!

The original developer was fortunate enough no longer to work in a Windows-environment, reducing the need for this client. Suraj Gupta has been maintaining and improving the code, and his fork is at <https://github.com/SurajGupta/RserveCLI>. I recommend that you use his version for any new projects.

Project Description
RserveCLI is a .NET/CLI client for Rserve. It allows .NET/CLI client to access R on the same machine or across the network.

Download the source code and or look at the slides on the [Documentation](#) page.

Have fun!

Summary for the search engines:

Rserve (<http://www.forge.net/Rserve/>) is a server application that allows users to access an R (<http://www.r-project.org/>) session remotely across the network. RserveCLI is a client library that allows one to access an Rserve server from a .NET/CLI environment using languages such as C#, Visual Basic, Boo, Python, etc.
Last edited: Sep 20, 2013 at 6:44 PM by [nhamold_version.8](#)

Search Wiki & Documentation

download

CURRENT RserveCLI 1.0.0.0
DATE Thu Apr 28, 2011 at 9:00 AM
STATUS Stable
DOWNLOADS 754
RATING ★★★★★ 2 ratings
[Review this release](#)

MOST HELPFUL REVIEWS

★★★★★ RserveCLI was my savior, when I needed to integrate R-forecasting into ASP.NET application. It works perfectly.

rJava - Low-level R to Java interface

RForge.net

About rJava

What is rJava?
rJava is a simple R-to-Java interface. It is comparable to the `.C/.Call` C interface. rJava provides a low-level bridge between R and Java (via JNI). It allows to create objects, call methods and access fields of Java objects from R.

rJava release versions can be obtained from CRAN - usually `install.packages("rJava")` in R will do the trick. The current development version can be downloaded from the [files section](#).

In a sense the inverse of rJava is [JRI](#) (Java/R Interface) which provides the opposite direction - calling R from Java. [JRI](#) is now shipped as a part of the [rJava](#) package, although it still can be used as a separate entity (especially for development). Currently rJava is used as a part of [JGR](#), [iPlots](#) and [JavaGD](#) software/packages.

Please report any bugs or wishes related to rJava or JRI using [Issues on GitHub](#).

What's new?
rJava source repository is now on [GitHub](#) and that is also the place to report bugs. The main page and builds are still on RForge.net.


2012/12/23 - rJava 0.9-6 released. Fixes Java parameter issue introduced in 0.9-5 on systems with headless mode (e.g. OS X).

2011/06/22 - rJava 0.9-0 released. This is a major upgrade that changes behavior of array references in low-level calls back to early 0.8 state as intended by the original design. It should be more consistent now. We have had rJava 0.9-0 in RC state for a long time so hopefully all developers relying on rJava have checked their packages. For the full list of fixes and changes see [NEWS](#).

2009/10/27 - rJava 0.8-0 released. Many new features mostly thanks to Romain Francois -- check the [NEWS](#) file for details.

2009/08/22 - rJava 0.7-0 released. Recommended update (includes bugfixes to fields handling, new support for with(), methodField

rpy2 - R in Python



[Get it on Pypi](#)

About
rpy2 is a redesign and rewrite of [rpy](#). It is providing a low-level interface to **R** from **Python**, a proposed high-level interface, including wrappers to graphical libraries, as well as R-like structures and functions.

Source and installation

Questions and issues
Questions should preferably be asked on the [rpy mailing-list](#) on SourceForge, or on [StackOverflow](#). Bugs, or wishes, can also go to the [bitbucket page](#) for the project.

Time travel
If looking for mv or the old website... this



R under Mono on Debian via RserveCLI

The screenshot shows the MonoDevelop IDE interface. The main window displays the output of an R linear model fit. The console output includes the following information:

```
plik Edycja Widok Terminal Pomoc
y = 10,5 x + -36,5. R^2 = 94,35%
[ 273, 243, 244, 102, 455, 235, 205, 160 ]
BŁĄD: nie znaleziono obiektu 'alaniemakota'

Call:
lm(formula = y ~ x1 * x2)

Residuals:
    Min       1Q   Median       3Q      Max
-2.58259 -0.61681  0.03649  0.74187  2.22947

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.06182    0.10387  -0.595    0.553
x1           2.91777    0.10734  27.181 <2e-16 ***
x2          -5.99019    0.11487 -52.146 <2e-16 ***
x1:x2       -0.05920    0.11467  -0.516    0.607
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.035 on 96 degrees of freedom
Multiple R-squared:  0.9728,    Adjusted R-squared:  0.972
F-statistic: 1145 on 3 and 96 DF,  p-value: < 2.2e-16
```

Below the console output, a snippet of C# code is visible, showing error handling logic:

```
76 if(e.Source.Equals("RserveCli"))
77 {
78     if(e.Message.Equals("R threw an error. "))
79     {
80         Console.WriteLine(s.Eval("geterrmessage()").AsString);
81     }

```

The IDE interface includes a Solution Explorer on the left showing the project structure, a Class Explorer below it, and a status bar at the bottom with the text "Gotowe" and "86 : 40 WST".



R.NET interface

Sample .NET applications

```
Edytor skryptów R
( year <- c(2000, 2001, 2002, 2003, 2004) )
( rate <- c(9.34, 8.50, 7.62, 6.93, 6.60) )

summary(lm(rate ~ year))
```

Zapisz wyniki do pliku Wyczyść okno wyników Informuj, który skrypt wykonano Wykonaj skrypt

2011-10-16 19:39:19 - Wykonanie skryptu użytkownika

```
> ( year <- c(2000, 2001, 2002, 2003, 2004) )
[1] 2000 2001 2002 2003 2004

> ( rate <- c(9.34, 8.50, 7.62, 6.93, 6.60) )
[1] 9.34 8.50 7.62 6.93 6.60

> summary(lm(rate ~ year))

Call:
lm(formula = rate ~ year)

Residuals:
    1     2     3     4     5 
0.132 -0.003 -0.178 -0.163  0.212

Coefficients:
(Intercept) 1419.20800 126.94957 11
year          -0.70500  0.06341 -11
---
Signif. codes:  0 '***' 0.001 '**' 0.0

Residual standard error: 0.2005 on 3 d
Multiple R-squared:  0.9763, Adjusted
F-statistic: 123.6 on 1 and 3 DF, p-v
```

Panel badacza

Okres od: 1 czerwca 2012 do: 31 sierpnia 2012 **Wykonaj analizę** [Zakończ żądanie nowej analizy](#)

Porównanie skuteczności leków Efekt działania leku w miesiącach terapii Porównanie skuteczności drogi podania leku

Wybór leków W grupach rozpoznai

- Lek 1
- Lek 2
- Lek 3
- Lek 4

- ICD xxx
- ICD yyy
- ICD zzz

Dodatkowe parametry

Parametr 3 = wartość

[Dodaj do listy z warunkiem:](#) ORAZ

- Parametr 1 = "Polska"
- ORAZ Parametr 2 = "Per Protocol"
- ORAZ Parametr 5 >= 11,5
 - ORAZ Parametr 5 <= 15,0
- ORAZ Parametr 9 = "Aktywny"

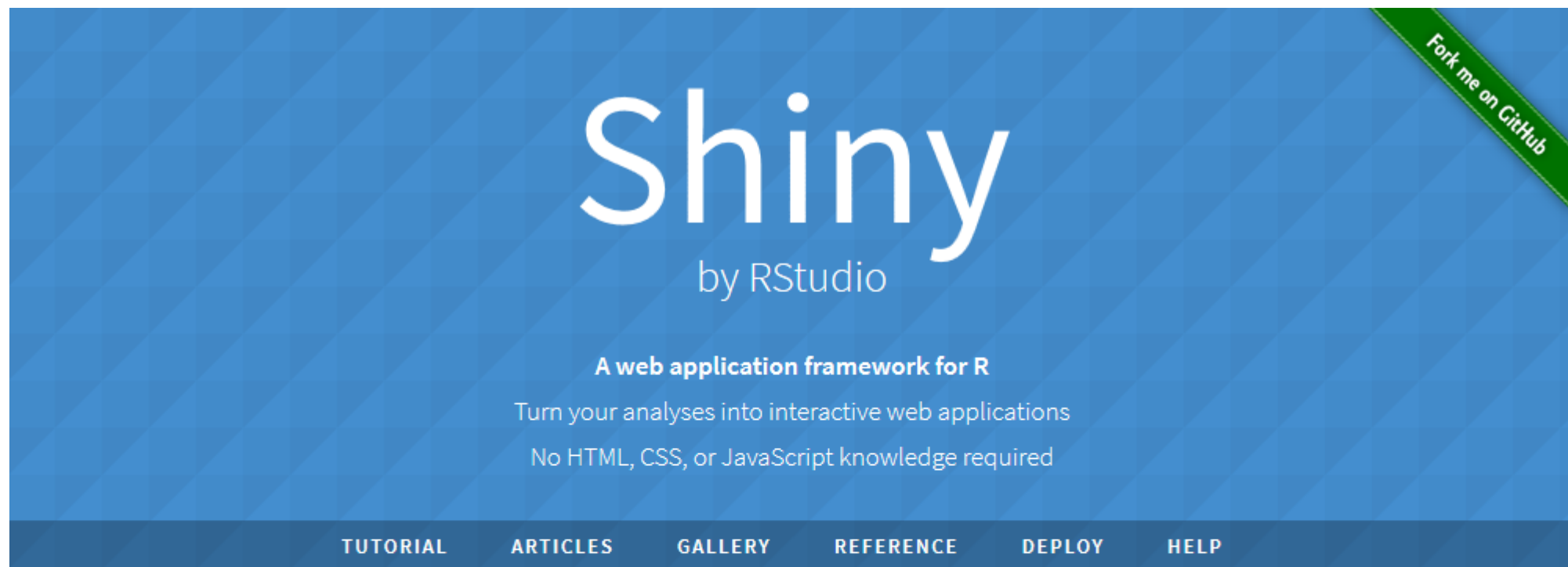
Wykres średnich ANOVA (treatment) Trend liniowy

Średnie stężenia w funkcji leku

Group	n	Mean
Lek A	13	12.77
Lek B	14	13.50
Lek C	13	11.01
Placebo	10	10.06



shining web applications with R and Shiny



Shiny
by RStudio

A web application framework for R
Turn your analyses into interactive web applications
No HTML, CSS, or JavaScript knowledge required

[TUTORIAL](#) [ARTICLES](#) [GALLERY](#) [REFERENCE](#) [DEPLOY](#) [HELP](#)

Fork me on GitHub



Get inspired
(gallery)



Get started
(tutorial)



Go deeper
(articles)



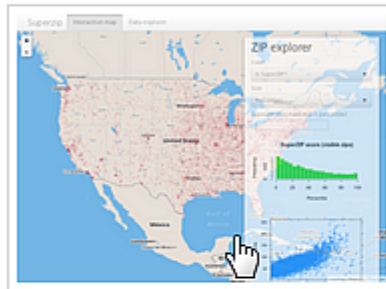
shining web applications with R and Shiny

Gallery

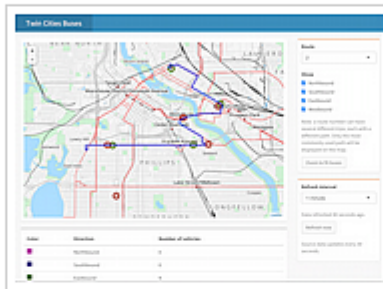
This gallery contains useful examples to learn from. Visit the [Shiny User Showcase](#) to see an inspiring set of sophisticated apps.

Interactive visualizations

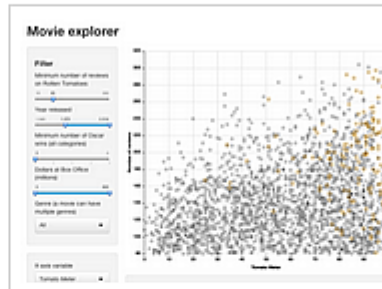
Shiny is designed for fully interactive visualization, using JavaScript libraries like [d3](#), [Leaflet](#), and [Google Charts](#).



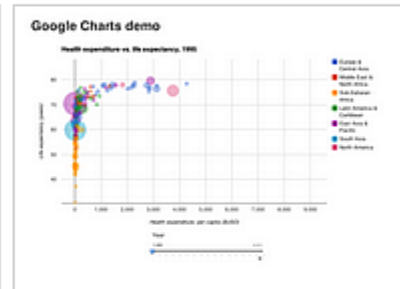
SuperZip example



Bus dashboard



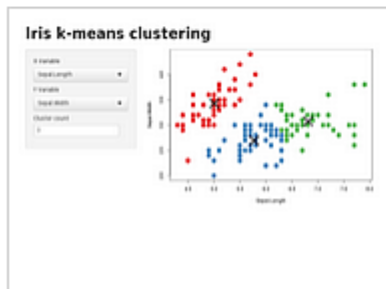
Movie explorer



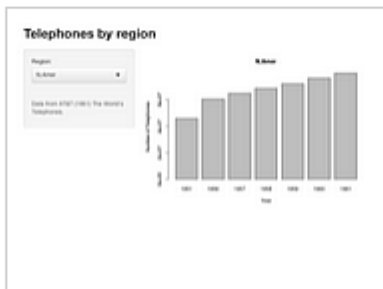
Google Charts

Start simple

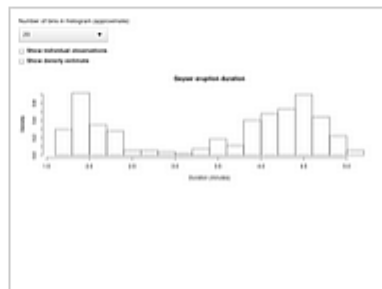
If you're new to Shiny, these simple but complete applications are designed for you to study.



Kmeans example



Telephones by region



Faithful



Word cloud



shining web applications with R and Shiny

Widgets

Get to know many of the input and output widgets that are available in Shiny with these examples.

Shiny Widgets Gallery

The right widget below the Current Values widget displays the value that the widget below it displays. Below that is the widget that you can interact with the widget.

Action button
Action

Single checkbox
Check A
Check B
Check C

Checkbox group
Check A
Check B
Check C

Date input
Current Date: 02/19/2017

Date range
Current Dates: 02/17/2017

File input
No file chosen

Widget Gallery

Sliders

Slider: 50

Name	Value
# Slider	50
# Range	0.5
# Range	100.00
# Range	0
# Range	0
# Range	0

Sliders

More Widgets

Choose a dataset: iris

Summary

Species	sepal	petal	other	petal
Setosa	5.00	3.50	1.50	0.38
versicolour	5.76	4.35	1.51	0.76
virginica	5.90	4.91	1.61	1.01

Observations

data	petal	other	petal
1	5.10	3.50	1.40
2	4.90	3.60	1.40
3	5.00	3.40	1.50
4	4.40	3.70	1.40
5	4.80	3.40	1.50

Widgets

Basic DataTable

Manufacturer	Model	Weight	Price	Capacity
Canon	EOS 100D	4.9	549	18
Canon	EOS 1100D	5.5	449	18
Canon	EOS 1200D	5.5	499	18
Canon	EOS 1300D	5.5	549	18
Canon	EOS 1500D	5.5	649	18

Basic DataTable

DataTables Demo

Name	Age	Salary	Sex
John	28	52000	Male
Jane	32	68000	Female
Bob	45	85000	Male
Alice	58	120000	Female

DataTables Demo

DataTables Options

Species	Sepal.Length	Petal.Length	Sepals.Width
Setosa	5.1	1.6	0.2
Setosa	4.9	1.4	0.2
Setosa	5.4	1.7	0.2
Setosa	4.6	1.2	0.2
Setosa	5.0	1.5	0.2

DataTables Options

Date and date range

Date input: 2014-03-01

Date range: 2014-03-01 to 2014-03-02

Date range with dropdown: 2014-03-01 to 2014-03-02

Date and date range

Dynamic Clustering in Shiny

Scatter plot showing data points clustered into groups. Legend: Group 1 (red), Group 2 (blue), Group 3 (green).

Dynamic Clustering

File Upload

Choose file to upload

Upload

File Upload

File Download

Choose a dataset: iris

Download

File Download

Selectize Examples

Examples of selectize widgets: single selection, multiple selection, search, and disabled options.

Selectize Examples

Selectize vs. Select

Comparison of selectize and select widgets. Selectize offers search and multiple selection, while select is a simple dropdown.

Selectize vs. Select

Custom input example

URL: http://www.r-project.org/

Custom input example

URL: http://www.r-project.org/

Custom input example

URL: http://www.r-project.org/

Custom input example

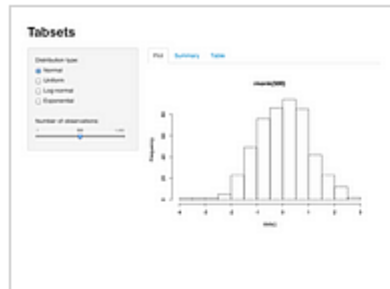
URL: http://www.r-project.org/



shining web applications with R and Shiny

Application layout

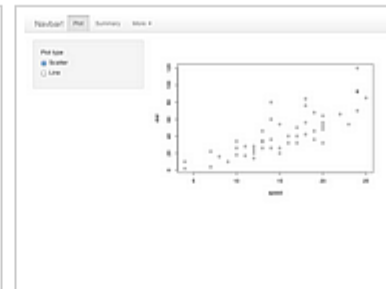
Each example in this category demonstrates one or more of the functions you can use to organize app UI.



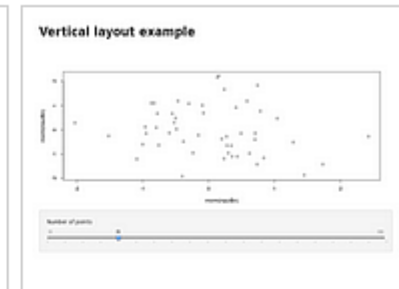
Tabsets



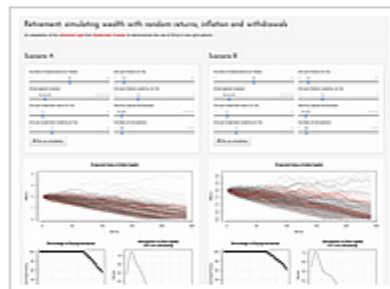
Plot plus three columns



Navbar Example



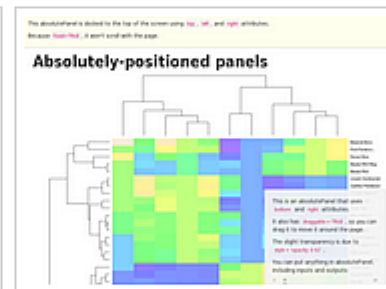
Vertical Layout



Retirement simulation



navlistPanel example



Absolutely-positioned panels



Including HTML, text, and Markdown files



Inline Output

easy (but shining) web applications with R and Shiny

Firefox

Height and weight of schoolchildren

glimmer.rstudio.com/winston/heightweight/

X variable: Age

Limit range: 11 to 21

Y variable: Height (inches)

Limit range: 50 to 72

Separate male/female

Model predictions

Linear (dot-dash)

Quadratic (dashed)

Locally weighted LOESS (solid)

Smoothing (alpha): 0.15 to 1

Quadratic model

```
Call:
lm(formula = "heightIn ~ I(ageYear^2) + ageYear * sex", data = hw_sub)

Residuals:
    Min       1Q   Median       3Q      Max
-8.1284 -1.6284  0.0967  1.6099  7.6396

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.49769    10.22530  -0.342   0.733
I(ageYear^2) -0.24747     0.05187  -4.771 3.24e-06 ***
ageYear * sex
```



some GUI hosted by R

R Latticist: CO2 v. 0.9-41

uptake vs conc by Type and Treatm

Treatment: nonchilled (blue), chilled (pink)

Regions: Quebec, Mississippi

Y-axis: uptake (40)

X-axis: conc (200, 400, 600)

Kernel: gaussian

Density Estimation kernel: Gaussian, Epanechnikov, Rectangular, Triangular, Cosine

Density estimation bandwidth (scaled by 100): 100

R command for random data creation: `c(rnorm(100,0,1), rnorm(50,5,1))`

Enter Two-Way Table

Table: Statistics

Number of Rows: 3

Number of Columns: 4

Enter counts:

	2	3	4
1			
2			
3			

Plot Means

Data Options

Error Bars: Standard errors (selected), Standard deviations, Confidence intervals (Level of confidence: 0.95), No error bars

Plot Labels: x-axis label < auto >, y-axis label < auto >, Graph title < auto >

Base Settings

Name of new design: Design.1

Size and randomization:

- Number of runs: 8 (Specify nruns checked)
- Number of factors: 4
- Number of center points: 0
- Number of blocks: 1 (blocks may be aliased with 2fis unchecked)
- Replications: 1 (Repeat only unchecked)
- Seed for randomization: 3777 (Randomization checked)

Design properties:

- Minimum resolution: III (NOTE: affects design generation for MaxC2 choice OR unspecified number of runs only)
- MA (Maximum resolution and minimum aberration) (selected)
- MaxC2 (Maximum number of clear 2fis) (unselected)

Show available designs

Show best 10 designs for 4 factors in 8 runs
The menu remains open, fetch it back after looking at designs

Buttons: Help, OK, Cancel, Apply, Reset



13 reasons why **you will** love GNU R

- I R is (extremely) cheap. In fact - it's free :)
- II R has (extremely) wide range of capabilities
- III R is (widely) supported by the world of science
- IV R is supported by the community
- V R is (increasingly) supported by the business
- VI R is able to read data in many formats
- VII Interoperability is easy to achieve

VIII R is truly cross-platform

- IX R offers numerous ways of presenting data
- X There are many possibilities to optimize the code
- XI R is able to handle large amount of data
- XII R has a set of fancy tools and IDEs
- XIII FDA accepted using R for drug trials!



R is truly cRoss-platform



R can be run:

- on **many operating systems**: Windows (95-8.1), Unix (Solaris, AIX, HP-UX), Unix-like (GNU/Linux, FreeBSD), OSX and mobile: [iOS](#), [Android](#), Maemo
- on **minicomputers** like [RaspberryPi](#) ([Raspbian OS](#))
- **without installation**, from any storage (USB device, CD/DVD, SD card)^{portable}
– just copy/paste the directory with R. You can have many versions of it.

Those properties make R an ideal base for creating:

- **Handy, self-contained tools** used by CRO's *Data Management* and *Biometrics* departments for easy monitoring the quality of collected data.
- **Cheap, easy to deliver and backup** (copy/paste) computing environments
/ free Linux + R + RStudio + PostgreSQL/MySQL + SVN + TrueCrypt + OpenVPN + VNC + rsync + Apache /
- **Independent** computing units based on simplified minicomputers.



GNU/Debian "Wheezy"

The screenshot displays the GNU/Debian Wheezy desktop environment. The top panel shows the system tray with the date and time: "nie 23 wrz, 17:20". The desktop background is dark blue with a star pattern. A window titled "Komputer" is visible in the background. The foreground features two windows:

- R Commander:** A window with a menu bar (File, Edit, Data, Statistics, Graphs, Models, Distributions, Tools, Help) and a toolbar. The "Data set" and "Model" fields are both set to "<No active dataset>". The "Script Window" contains the following R code:

```
x <- rnorm(100)
y <- 2*x + rnorm(100)
summary(lm(y~x))
```

The "Output Window" displays the results of the linear model fit:

```
Call:
lm(formula = y ~ x)

Residuals:
    Min       1Q   Median       3Q      Max
-2.76833 -0.65465  0.02382  0.79758  2.29783

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.07161    0.10429   0.687   0.494
x            2.06915    0.10204  20.278 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.038 on 98 degrees of freedom
Multiple R-squared:  0.8075, Adjusted R-squared:  0.8056
F-statistic: 411.2 on 1 and 98 DF, p-value: < 2.2e-16
```
- Terminal (mono@debian: ~):** A terminal window showing the R version (2.15.1) and the installation of the Rcmdr package using the command `> library(Rcmdr)`. The output shows the installation of several dependencies: tcltk, MASS, and nnet.



R and friends on Debian "Stretch"

Podgląd RStudio pią, 02:38 pl

R Commander

Plik Edycja Dane Statystyki Wykresy Modele Rozkłady Narzędzia Pomoc

Zbiór danych: data Edytuj zbiór danych Zobacz zb

Skrypt R R Markdown

```
font=getRcmdr('logFont'), maxwidth=80, maxheight=30,
suppress.X11.warnings=FALSE)
densityPlot(y~x, data=data, bw="SJ", adjust=1,
kernel="gaussian", main="Density in both groups")
```

Wyjście Wykonaj

```
> data <- data.frame(y=rnorm(100), x=c(rep("A", 50), rep("B", 50)))
> library(relimp, pos=14)
```

R Graphics: Device 2 (ACTIVE)

Density in both groups

data

y	x
1	-0.633457893 A
2	-0.856902216 A
3	1.264061611 A
4	-0.469223534 A
5	-0.749086866 A
6	-0.247264247 A
7	-0.427223345 A
8	1.214380229 A
9	-1.085049608 A
10	-1.136368578 A
11	0.866431842 A
12	-0.652096474 A
13	0.213989189 A
14	-1.476227920 A
15	0.198889763 A
16	0.140341900 A

~/KCR/Projects/Biostat-Projects - RStudio

File Edit Code View Plots Session Build Debug Tools Help

Go to file/function SVN

Biostat-Projects

setup.r * XMLDataLoader.r * test.r *

```
classes <- sapply(getNodeSet(thisDoc, "/xsd
13
14
15
16 node <- node
17 type <- xmlGetAttr(node, "type")
18 name <- xmlGetAttr(node, "name")
19 if(is.null(type)) {
20 tmpDoc <- xmlDoc(node)
21 type <- str_replace(unlist(xmlApp
22 free(tmpDoc)
23 }
24 tmp <- c(ifelse(type %in% c("integer",
25 names(tmp) <- name
26 tmp
27 })
28 free(thisDoc)
29 rm(thisDoc)
30 return(classes)
```

21:26 FUN(node) R Script

Console ~/KCR/Projects/Biostat-Projects/

```
> x<-seq(0, 2*pi, 0.1)
> plo
> lin
> gri
> Vie
> ses
> ses
[1] "
> ses
[1] "
> ses
[1] "
> ses$
```

platform arch os system status major minor year

Environment History SVN

Import Dataset Grid

Global Environment

Name	Type	Len...	Size	Value
df	data.f...	2	2.3 ...	6 obs. of 2 va...
doc	docx	6	6.8 ...	List of 6
ft	FlexTa...	14	31.8...	List of 14
ses	list	14	2.5 ...	List of 14
x	numeric	63	544 B num	[1:63] 0 0.1...

Files Plots Packages Help Viewer

Zoom Export

sin(x)

sin(x)

x



MacOS / OS X

R File Edit Format Workspace Packages & Data Misc Window Help

100% (Charged) Tue 2:14 PM stefano iacus

R Console

```

rgl.sr> ylen <- ylim[2] - ylim[1] + 1
rgl.sr> colorlut <- terrain.colors(ylen)
rgl.sr> col <- colorlut[y - ylim[1] + 1]
rgl.sr> rgl.clear()
rgl.sr> rgl.surface(x, z, y, color = col)

```

R Data Editor

height	weight
58	115
59	117
60	120
61	123
62	126
63	129
64	132
65	135
66	139
67	142
68	146
69	150
70	154
71	159
72	164

Quartz (2) - Active

Given : depth

R Workspace Browser

Object	Type	Structure
dati	data.frame	dim: 20 4
g	factor	levels: 10
l	numeric	length: 12
n	numeric	length: 1
opar	list	length: 2
pie.sales	numeric	length: 6
pin	numeric	length: 2
scale	numeric	length: 1
usr	numeric	length: 4
women	data.frame	dim: 15 2
height	numeric	length: 15
weight	numeric	length: 15
x	numeric	length: 87

Refresh List

```

BoxDens=function(data, npts = 200., x = c(0.,
add = TRUE, col = 11., border=FALSE,collin
{
dens <- density(data, n = npts)
dx <- dens$x
dy <- dens$y
if(add == FALSE)
plot(0., 0., axes = F, main = "", xlim = x, ylim = y,
ylab = "")
if(orientation == "paysage") {
dx2 <- (dx - min(dx))/(max(dx) - min(dx)) * (x[2.] - x
x[1.]
dy2 <- (dy - min(dy))/(max(dy) - min(dy)) * (y[2.] - y
y[1.]
seqbelow <- rep(y[1.], length(dx))
if(Fill == T)
confshade(dx2, seqbelow, dy2, col = col)
if (border==TRUE) points(dx2, dy2, type = "l", col = c
}
else {
dy2 <- (dx - min(dx))/(max(dx) - min(dx)) * (y[2.] - y
v[1.]

```

RGL device 1 (active)

R Package Manager

Refresh List

status	Package	Description
<input checked="" type="checkbox"/> loaded	graphics	The R Graphics Package
<input type="checkbox"/> not loaded	grid	The Grid Graphics Package
<input type="checkbox"/> not loaded	lattice	Lattice Graphics
<input checked="" type="checkbox"/> loaded	methods	Formal Methods and Classes
<input type="checkbox"/> not loaded	mgcv	CAMs with CCV smoothness estimation

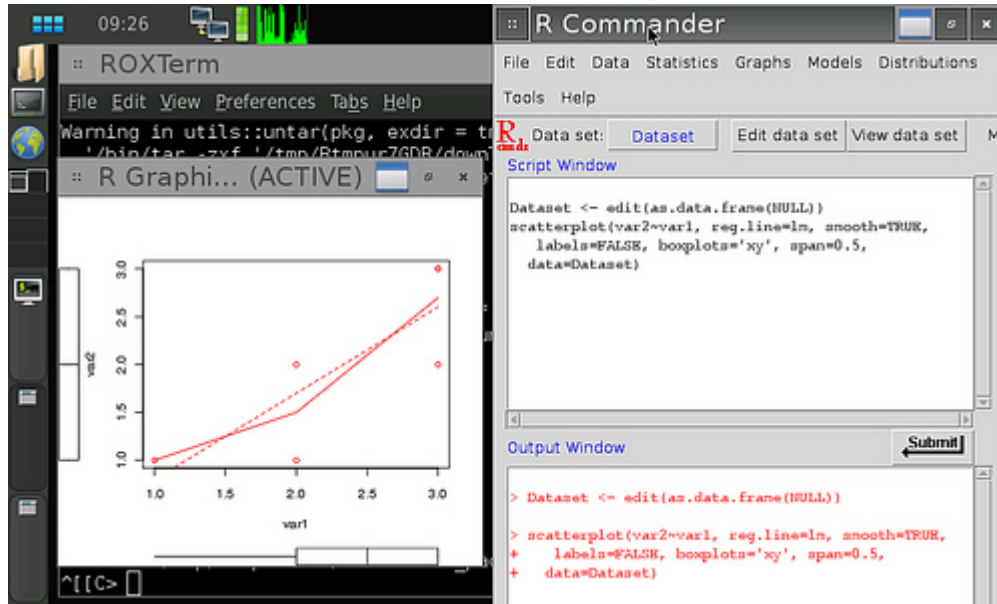
The R Graphics Package

Documentation for package `graphics' version 2.0.0

Help Pages

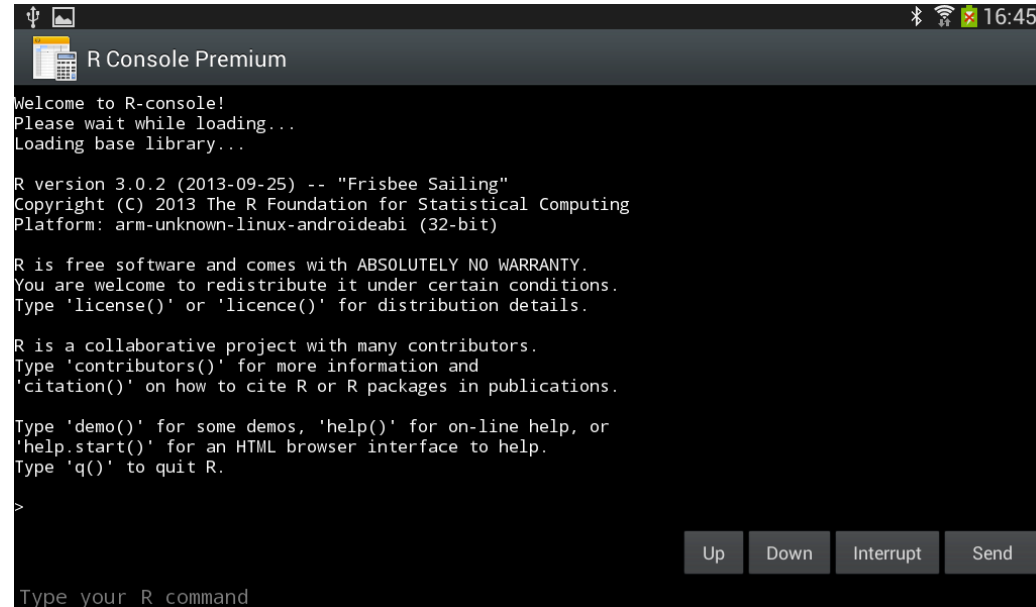
[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [L](#) [M](#) [N](#) [P](#) [R](#) [S](#) [T](#) [X](#)

R on mobile platforms



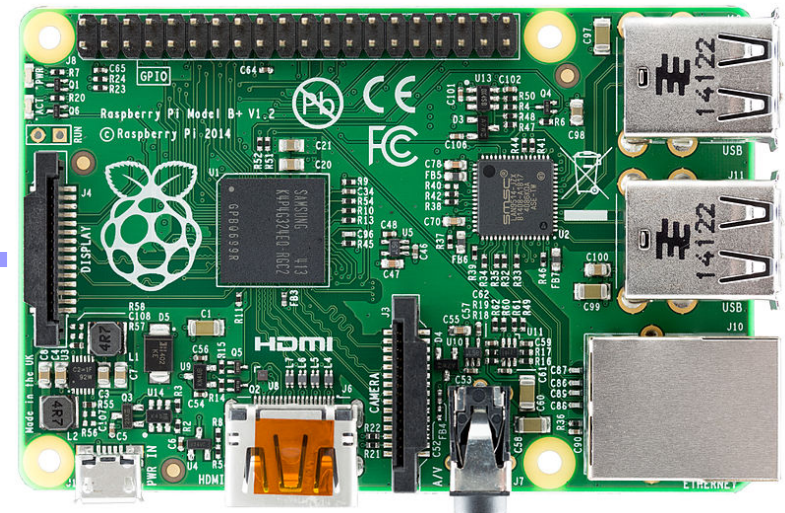
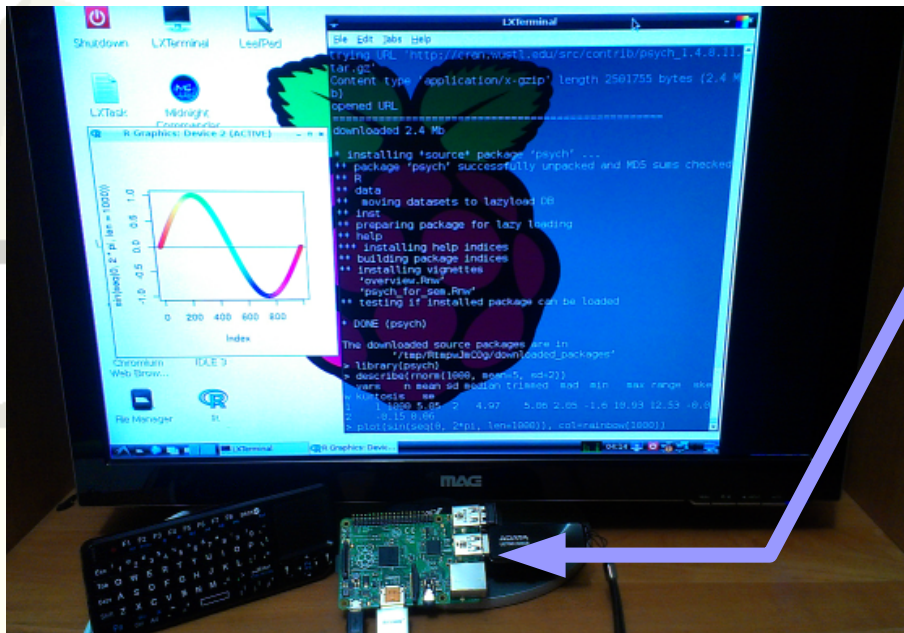
Maemo (Nokia N900)

<https://www.flickr.com/photos/n900user/>



Android

[R Console Premium \(Google Play\)](#)

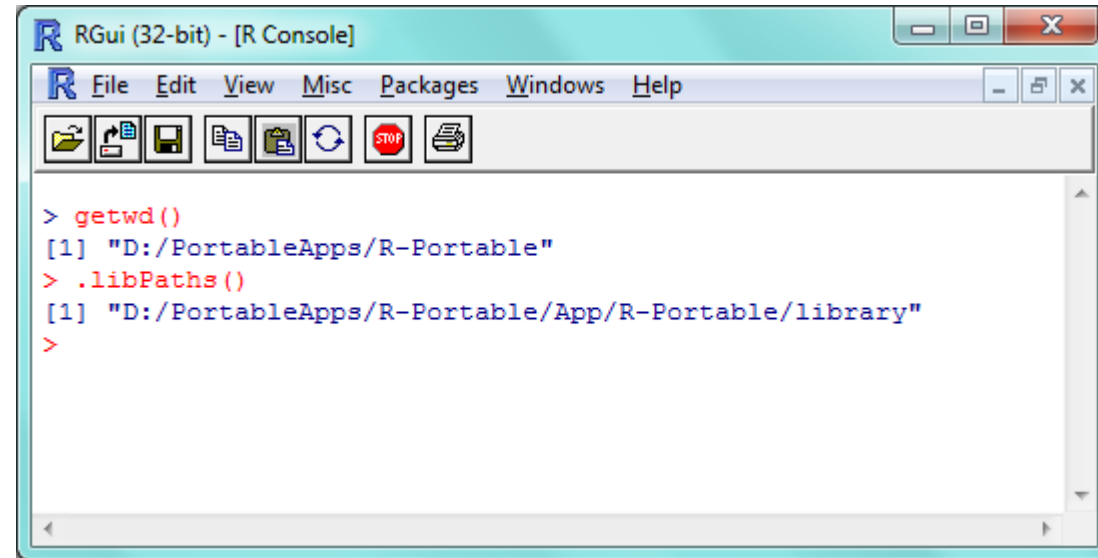
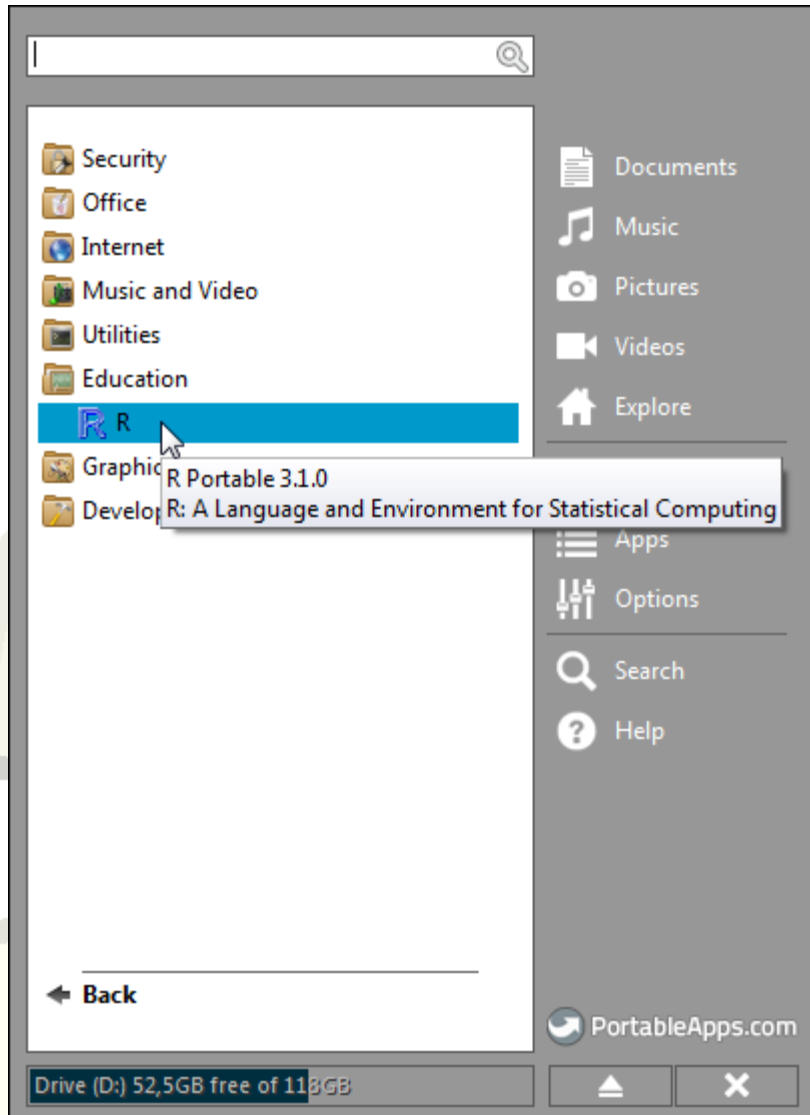


Raspberry Pi B+



<http://sourceforge.net/projects/rportable/>

(also older, 2.11 – 2.15 versions are available!)





PoRtable in use – Simple Data Inspector



Instructions

- * Pick an Access database file (.mdb only), select desired analyses, choose a viewer application and push the [Generate] button.
- * After all the selected application(s) will be opened showing the report. For tabular result CSV files are available as well.
- * After the report is generated, it is cached and can be opened many times in a desired viewer (use the "Open last" button).
- * The database file should be placed locally rather than on a network share.
- * The more analyses are selected, the longer it takes.

Pick a database: Open in:

The report is stored in:

Web browser continuous with TOC

Excel with multiple sheets

Excel continuous with TOC

Word continuous with TOC

OpenOffice Writer continuous with TOC

List of available analyzes.

You selected 4 analyzes. Go on :)

Database summary | **Retreatment** | Simple Study Tracker | Lab results reconciliation

Descriptive statistics of scores **Parameters:** Shortened

Graphical inspection of raw data (boxplots, histograms, densities, scatterplots) **Parameters:** acr. sites (N) acr. countr (N) don't split (N)

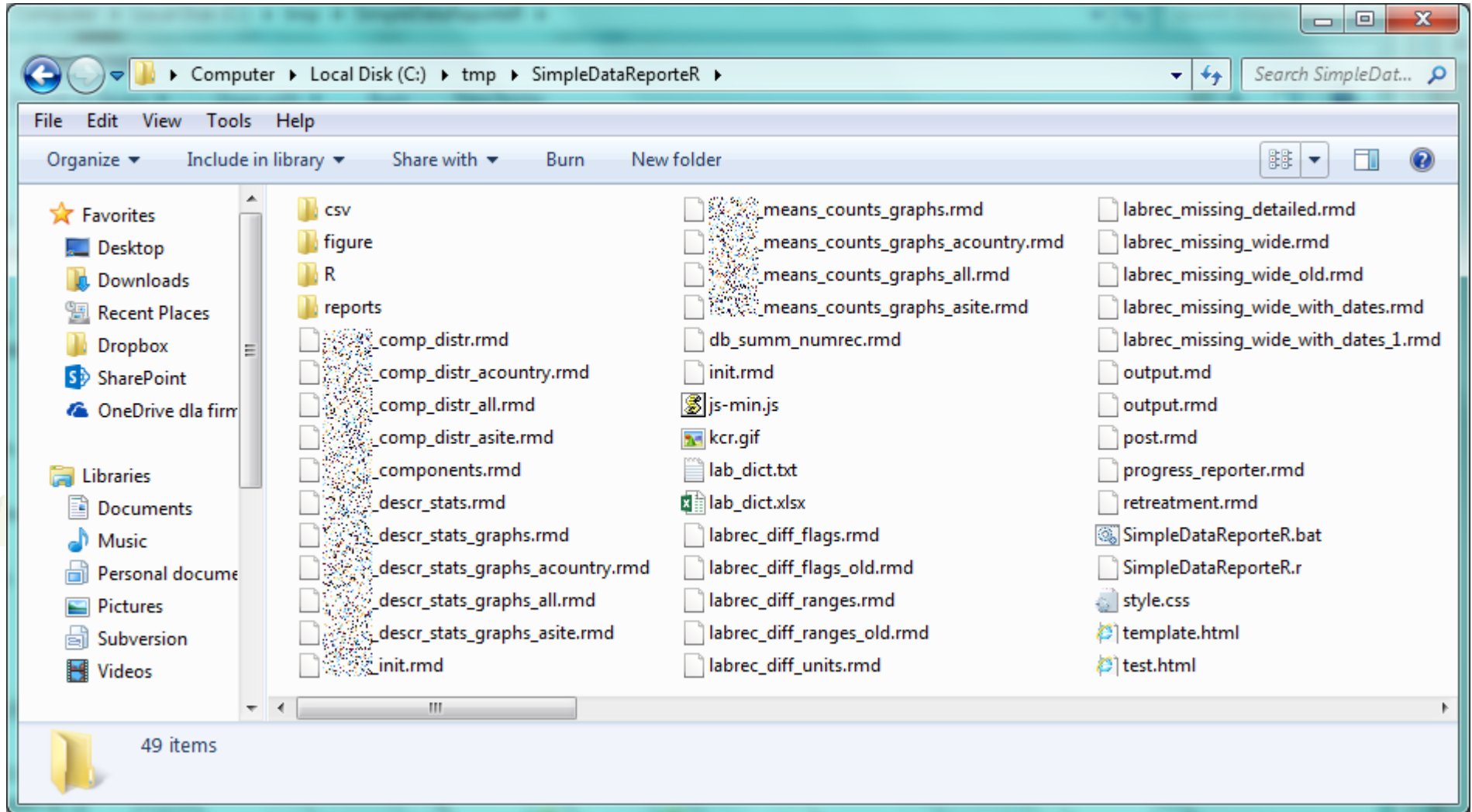
Mean + median + count barchart **Parameters:** acr. sites (N) acr. countr (N) don't split (N)

Comparison of distributions from D1, W16, 20, W24 **Parameters:** acr. sites (N) acr. countr (N) don't split (N)

View on and its components across selected sites **Parameters:** Period: site 1: Site 2: site 3:
site 4: Site 5: site 6:



♪ ...all you need is **R** ♪



Complete application: ~240MB raw,
~140MB after compression



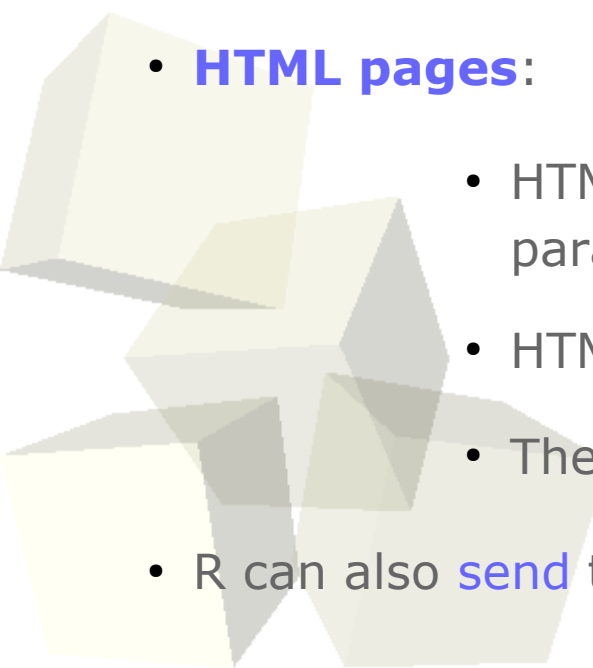
13 reasons why **you will** love GNU R

- I R is (extremely) cheap. In fact - it's free :)
- II R has (extremely) wide range of capabilities
- III R is (widely) supported by the world of science
- IV R is supported by the community
- V R is (increasingly) supported by the business
- VI R is able to read data in many formats
- VII Interoperability is easy to achieve
- VIII R is truly cross-platform
- IX R offers numerous ways of presenting data**
- X There are many possibilities to optimize the code
- XI R is able to handle large amount of data
- XII R has a set of fancy tools and IDEs
- XIII FDA accepted using R for drug trials!



R can produce output in a rich set of formats, such as:

- **pictures:** JPG, PNG, BMP, GIF, TIFF, SVG, EMF
- **animations:** GIF directly or via ImageMagic, AVI via Ffmpeg
- **plain text files:** CSV, TSV, XML, XPT (**SAS transport files**), ASCII
- **rich format docs:** RTF, PS, PDF, MS Word (*.doc; *.docx OpenXML), Open/Libre Office Writer (*.odt OpenDocument)
- **spreadsheets:** MS Excel (*.xls; *.xlsx), OpenOffice Calc (*.ods)
- **HTML pages:** simple pages, complex reports or presentations
 - HTML files can be easily read by **Word** and **Excel**. Tables, paragraphs, and styles are preserved
 - HTML is convertible to many other formats via **pandoc**
 - There is [a way to obtain multi-sheet Excel workbooks](#)
- R can also **send** the results to an external application via Webservice JSON





Definitions:

- *The goal of reproducible research is **to tie specific instructions to data analysis and experimental data so that scholarship can be recreated, better understood and verified.***

<http://cran.r-project.org/web/views/ReproducibleResearch.html>

- *The term reproducible research refers to the idea that the ultimate product of academic research is the **paper along with the full computational environment used to produce the results in the paper such as the code, data, etc. that can be used to reproduce the results and create new work based on the research.***

http://en.wikipedia.org/wiki/Reproducibility#Reproducible_research



In simple words, documents are created from templates containing **presentation code** (Markdown, HTML, LaTeX) or **formatted paragraphs** (MS Word, Open Document, RTF) mixed with **chunks of R code**.

After the template is processed, chunks of R code are replaced with the result they produce.

With this approach one can forget about continuous copying and pasting objects (charts and tables) from the statistical package into a word processor.

Such documents are **self-contained** in that, they are capable to perform alone all the tasks required to complete rendering:

- setup the environment (install required packages, apply configuration),
- read data from the indicated sources or generate artificial datasets
- analyze the retrieved data and produce output objects (tables, graphics)



Sample analysis

Ver. 1.2, date: 2011-02-01

1. Paragraph A

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

```
```{r}
tab <- sqlQuery(„SELECT * FROM Table”)
attach(tab)
model <- lm(Response ~ Treatment)
summary(model)
```
```

2. Paragraph B



- Knitr + pandoc
- Sweave
- OdfWeave
- ReporteRs

Sample analysis

Ver. 1.2, date: 2011-02-01

1. Paragraph A

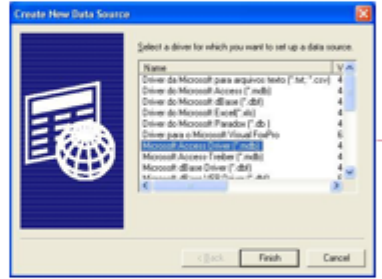
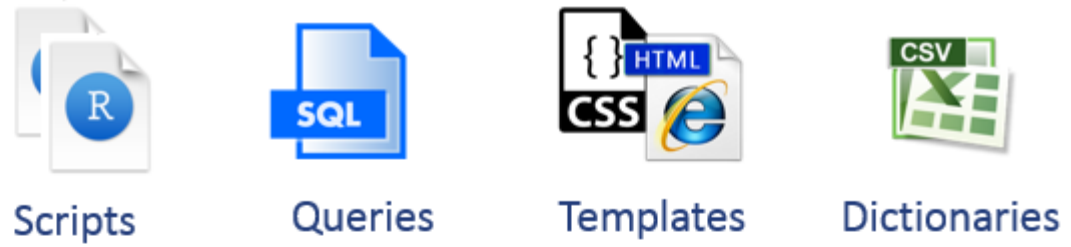
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

| | Estimate | Std. Error | t value | Pr(> t) | |
|-----------|----------|------------|---------|-----------|-----|
| Intercept | 10.7902 | 0.242 | 44.498 | <2e-16 | *** |
| Trt.B | 5.3418 | 0.331 | 16.111 | <2e-16 | *** |
| Trt.C | -0.4563 | 0.342 | -1.331 | 0.188 | |

2. Paragraph B



Reproducible Research



| SiteID | SubjID | Lab Test | Screening | Day 1 | Day 2 |
|--------|--------|----------|-----------|---------|---------|
| 1 | 3 | RBC | OK | OK | OK |
| 1 | 3 | WBC | OK | MISSING | N/A |
| 1 | 4 | ESR | OK | MISSING | MISSING |
| 1 | 5 | Hb | OK | OK | N/A |
| 2 | 5 | B-HCG | N/A | OK | MISSING |





R for Clinical Trial Reporting

Vanderbilt Biostatistics

Software Quality and Validation

Quality and Error Sources
What is Called "Validation" and What Should it Be?

Example of a Comprehensive Analysis Validation

High-Level Tools for Reproducible Analysis and Reporting

Background Tools
Mock Tables
Statistical Methods
Example

R for Clinical Trial Reporting: Reproducible Research, Quality and Validation

Frank E Harrell Jr

Department of Biostatistics, Vanderbilt University School of Medicine

useR! 2007 Conference

10 Aug 2007

Slides and Code at <http://biostat.mc.vanderbilt.edu/Rreport>



The R Series

Reproducible Research with R and RStudio



Christopher Gandrud

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

The R Series

Dynamic Documents with R and knitr



Yihui Xie

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

<http://biostat.mc.vanderbilt.edu/wiki/pub/Main/FHHandouts/dmcreport.pdf>

<http://handsondatascience.com/KnitRO.pdf>

Reproducible Research

Firefox

CRAN Task View: Reproducible Research

cran.r-project.org/web/views/ReproducibleResearch.html

Google

CRAN Task View: Reproducible Research

Maintainer: Max Kuhn

Contact: max.kuhn at pfizer.com

Version: 2014-09-23

The goal of reproducible research is to tie specific instructions to data analysis and experimental data so that scholarship can be recreated, better understood and verified.

R largely facilitates reproducible research using literate programming; a document that is a combination of content and data analysis code. The `Sweave` function (in the base R utils package) and the [knitr](#) package can be used to blend the subject matter and R code so that a single document defines the content and the algorithms.

Basic packages can be structured into the following groups:

- *LaTeX Markup* : The [Hmisc](#), [xtable](#) and [tables](#) packages contain functions to write R objects into LaTeX representations. [Hmisc](#) also includes methods for translating strings to proper LaTeX markup (e.g., " \geq " to " \geq "). Animations can be inserted into LaTeX documents being converted to PDF via the [animation](#) package. The `picTeX` function in the base `grDevices` package is a `PicTeX` graphics driver and the [tikzDevice](#) can convert R graphics to [TikZ](#) markup. The [tth](#) package can convert TeX to HTML.
- *HTML Markup* : The [R2HTML](#) package has drivers that allow `Sweave` to process HTML documents via `Sweave`. Packages [R2HTML](#), [hwriter](#) and [ReporteRs](#) can be used to build HTML pages sequentially. [R2HTML](#), [xtable](#) and [hwriter](#) can also convert some R objects into HTML representations. [knitr](#) also has facilities to weave R code with HTML as well as convert markdown to HTML.
- *ODF Markup* : The [odfWeave](#) package extends `Sweave` to the [Open Document Format](#). Word processing tools, such as OpenOffice.org, can then be used to blend content and programs. Many word processors can be used to translate the ODF document to other formats (e.g., Word, PDF, HTML, etc.)
- *Microsoft Formats* : The [R2wd](#) and [R2PPT](#) packages for Windows can be used to communicate between R and Word or PowerPoint via the COM interface. Document elements (e.g. sections, text, images, etc) that are created in R can be inserted into the document from R. The [rtf](#) can also be used to create RTF format documents directly from R. Commercial R products that work with RTF and/or Word are [RTFGen](#), [Inference for R](#) and [SWord](#). The output from other packages ([odfWeave](#) and [R2HTML](#)) can also be opened by Word. [ReporteRs](#) can be used to create Word and PowerPoint documents. [RExcel](#) can integrate code with Microsoft Excel. Additionally, the [table1xls](#) can convert summary tables to Excel files.
- *Plain Text Formats* : R code and output in `Sweave` files can be converted into [AsciiDoc](#) and other structured text formats using the [ascii](#) package. The [markdown](#) and [knitr](#) packages have tools for [markdown](#) format.
- *Syntax Highlighting* : The [SweaveListingUtils](#) package can also provide enhanced control over how R code chunks and their output are rendered in LaTeX.
- *Caching of R Objects* : The [weaver](#) package allows caching of specific code chunks. The [R.cache](#) package can also be used but is not integrated with `Sweave`. [knitr](#) also has the ability to cache the results of code chunks.
- *Others* : The [brew](#) and [R.rsp](#) packages contain alternative approaches to embedding R code into various markups. [knitr](#) is a comprehensive package derived from `Sweave` that includes code formatting, highlighting, caching, fine control of graphics, conditional evaluation, multiple markup formats and other features. The [pander](#) package can write R objects into [Pandoc's markdown](#) and also to convert those or complex reports to PDF/HTML/docx/ODT. The [rapport](#) package builds on [pander](#) and provides a way to create reproducible statistical report templates with graphs, tables and annotations to be applied to any R data frame and export the results in different formats. The [installr](#) package for Windows can download and install MikTeX, pandoc (and other software), as well as quickly update R itself.



Reproducible Research – knitr and Sweave

With [HTML](#), [Latex](#), [R Markdown](#), [xtable](#) and [CSS](#) processed by [Sweave](#) or [knitr](#) and (optionally) [pandoc](#), one can produce complex, professionally looking and reproducible reports in many formats, e.g. DOCX, ODT, HTML, RTF and PDF.

This R-scripts-driven approach might seem a bit complicated, but it gives the user maximum flexibility and control over how the final product will look like. HTML and CSS WYSIWYG editors are useful here.

Chunks of R code may be saved into named sub-reports for later reuse in many places in the template. They can be turned on/off depending on the result of a conditional expression.

Produced HTML file can be displayed in a web browser, opened directly in a word processor or spreadsheet or converted to another format (DOCX, PDF).

This is a perfect tool for [automated generation of reports](#) or [record the flow of analyses](#) (= R code along with results).



knitr + R Markdown + HTML + CSS

```
*C:\tmp\SimpleDataReporteR\template.html - Notepad++
File Edit Search View Encoding Language Settings Macro Run Plugins Window ?
template.htm
1 <!DOCTYPE html>
2 <html>
3 <head>
4 <title>#!title#</title>
5 <link rel="Stylesheet" type="text/css" href="style.css" />
6 <script src='js-min.js' type='text/javascript'></script>
7 </script>
8 </head>
9 <div class="container">
10 #!html_output#
11 </div>
12 <script type="text/javascript">
13 $(function() {
14     $('#td:contains("NA")').addClass('RedBackWhiteCol');
15 });
16 </script>
17 </div>
18 </html>
```

```
*C:\tmp\SimpleDataReporteR\style.css - Notepad++
File Edit Search View Encoding Language Settings Macro Run
Plugins Window ?
style.css
1 body {
2 font-family: Arial, Verdana, Tahoma;
3 line-height:130%;
4 }
5 #TOC {
6 margin-left:30px;
7 }
8 p, ul, ol {
9 margin: 0px;
10 font-size:13px;
11 text-align:justify;
12 }
13 table {
```

```
*c:\tmp\SimpleDataReporteR\init.rmd - Notepad++
File Edit Search View Encoding Language Settings Macro Run Plugins
init.rmc
15 suppressMessages(library(plyr))
16 con <- odbcConnectAccess(database)
17 ...
18
19 <!-- REPORT SECTION TabTitle="Header"-->
20 <a name="top"></a>
21 Report of data inspection for XXXX
22 =====
23
24 ```{r Report-time, echo=FALSE, results='asis'}
25 dbname <- gsub(pattern = "(.*)/(.*)($)", replacement = "\
26 cat(paste("Database:<span class='ReportDetails'>", dbname
27 ...
28 <hr>
29 <p>Lorem ipsum dolor sit amet, consectetur adipiscing eli
30 <hr>
31
32 ### Table of contents:
33 ```{r TOC, echo=FALSE, results='asis'}
34 ToDo <- subset(analyzes, run==TRUE)
35 ToDoByParams <- subset(parameters, fkid %in% ToDo$id & me
36
37 cat("<ol id='TOC'>");
38 for(anID in ToDo$id){
```

```
*C:\tmp\SimpleDataReporteR\comp_distr_all.rmd - Notepad++
File Edit Search View Encoding Language Settings Macro Run Plugins Window ?
comp_distr_all.rmc
1 <!-- REPORT SECTION TabTitle="Distr. of XXX"-->
2
3 ```{r XXX_comp_distr_all_anchor, echo=FALSE, results='asis'}
4 param <- subset(XXX_comp_distr_params, code.1 == "all")
5 cat(paste("<h3><a name='\
6 param$descr_label, "\
7 param$analysis), " - ", param$descr_label, "<a href='\
8 class='\
9 </a>>Top</a>", sep="")
10 ...
11
12 #### Disitributions of XXX- - overall
13 ```{r XXX_comp_distr_all, echo=FALSE, warnings = FALSE,
14 messages=FALSE, comment = NA, fig.width=12, fig.height=8}
15
16 XXX_dens_distr <- XXX[! XXX$visit_name %in% c("Week 48", "Early
17 Termination", "Week 4", "Week 8", "Week 12"),]
18
19 ggplot(XXX_dens_distr, aes(XXX, fill=visit_name)) +
20 geom_density(alpha=.5) +
21 theme(legend.position=c(1,1), legend.justification=c(1,1)) +
22 labs(title="Distributions of XXX\n", colour="Visits") +
23 theme(plot.title = element_text(size = rel(1.3), face = "bold"))
24 ...
```




HTML report opened in Word and Excel

report_aaolsz_d

PLIK NAR WST PRO UKŁ OD

Report of data inspection

Database: 17sep.mdb | Created on: 2014-10-03 05:20:48

Some description...

Table of contents:

- [Descriptive statistics of xxx scores](#)

Descriptive statistics of xxx scores

| visit_name | N | mean | SD |
|-------------------|-----|------|------|
| Day 1 | 372 | 6.84 | 0.75 |
| Week 4 | 314 | 5.43 | 1.21 |
| Week 8 | 288 | 4.78 | 1.23 |
| Week 12 | 268 | 4.36 | 1.27 |
| Week 16 | 246 | 3.75 | 1.38 |
| Week 20 | 226 | 3.64 | 0.92 |
| Week 24 | 192 | 4.00 | 1.17 |
| Week 48 | 73 | 3.51 | 1.49 |
| Early Termination | 7 | 5.00 | 2.12 |

Report generated in: 00:00:07

STRONA 1 Z 1 WYRAZY: 153

report_aaolsz_db_17sep_... 2014-10-03_...

PLIK NARZĘDZI WSTAWIA UKŁAD ST FORMUŁY DANE RECENZJA WIDOK Olszewski...

B19 : 288

Report of data inspection for XXXX

Database: 17sep.mdb | Created on: 2014-10-03 05:20:48

Some description...

Table of contents:

- [Descriptive statistics of xxx scores](#)

Descriptive statistics of xxx scores Top

| visit_name | N | mean | SD | MAD | QD | min | Q1 | Me | Q3 | max |
|-------------------|-----|------|------|------|------|------|------|------|------|------|
| Day 1 | 372 | 6.84 | 0.75 | 0.81 | 0.53 | 4.51 | 6.34 | 6.80 | 7.40 | 8.83 |
| Week 4 | 314 | 5.43 | 1.21 | 1.22 | 0.79 | 2.15 | 4.74 | 5.49 | 6.33 | 7.85 |
| Week 8 | 288 | 4.78 | 1.23 | 1.28 | 0.86 | 0.56 | 3.97 | 4.88 | 5.69 | 7.80 |
| Week 12 | 268 | 4.36 | 1.27 | 1.38 | 0.92 | 1.26 | 3.41 | 4.46 | 5.25 | 8.32 |
| Week 16 | 246 | 3.75 | 1.21 | 1.08 | 0.74 | 0.56 | 2.93 | 3.66 | 4.41 | 7.85 |
| Week 20 | 226 | 3.64 | 1.17 | 0.72 | 0.66 | 1.09 | 2.90 | 3.20 | 4.22 | 7.09 |
| Week 24 | 192 | 4.00 | 1.11 | 0.65 | 0.42 | 0.80 | 3.62 | 3.94 | 4.46 | 7.36 |
| Week 48 | 73 | 3.51 | 1.49 | 1.42 | 0.96 | 1.00 | 2.34 | 3.30 | 4.26 | 7.22 |
| Early Termination | 7 | 5.00 | 2.12 | 1.93 | 1.68 | 1.74 | 3.42 | 5.80 | 6.78 | 7.10 |

Report generated in: 00:00:07

report_aaolsz_db_17sep_...

GOTOWY 100%



pandoc – convert to real DOC / OXML / PDF / ODF

Firefox

Pandoc - About pandoc

Pandoc a universal document converter

Donate

Flattr

424

About

Installing

Getting started

Demos

Try pandoc online

Examples

Documentation

User's Guide

FAQ

API documentation

Contributing

Mailing lists

Scripting

Making an ebook

Extras

Releases

About pandoc

If you need to convert files from one markup format into another, pandoc is your swiss-army knife. Pandoc can convert documents in [markdown](#), [reStructuredText](#), [textile](#), [HTML](#), [DocBook](#), [LaTeX](#), [MediaWiki markup](#), [OPML](#), Emacs [Org-Mode](#), [Txt2Tags](#), Microsoft Word [docx](#), [EPUB](#), or [Haddock markup](#) to

- HTML formats: XHTML, HTML5, and HTML slide shows using [Slidy](#), [reveal.js](#), [Slideous](#), [S5](#), or [DZSlides](#).
- Word processor formats: Microsoft Word [docx](#), OpenOffice/LibreOffice [ODT](#), [OpenDocument XML](#)
- Ebooks: [EPUB](#) version 2 or 3, [FictionBook2](#)
- Documentation formats: [DocBook](#), [GNU TexInfo](#), [Groff man](#) pages, [Haddock markup](#)
- Page layout formats: [InDesign ICML](#)
- Outline formats: [OPML](#)
- TeX formats: [LaTeX](#), [ConTeXt](#), LaTeX Beamer slides
- [PDF](#) via LaTeX
- Lightweight markup formats: [Markdown](#), [reStructuredText](#), [AsciiDoc](#), [MediaWiki markup](#), [DokuWiki markup](#), Emacs [Org-Mode](#), [Textile](#)
- Custom formats: custom writers can be written in [lua](#).

Pandoc understands a number of useful markdown syntax extensions, including document metadata (title, author, date); footnotes; tables; definition lists; superscript and subscript; strikethrough; enhanced ordered lists (start number and numbering style are significant); running example lists; delimited code blocks with syntax highlighting; smart quotes, dashes, and ellipses; markdown inside HTML blocks; and inline LaTeX. If strict markdown compatibility is desired, all of these extensions can be turned off.

LaTeX math (and even macros) can be used in markdown documents. Several different methods of rendering math in HTML are provided, including MathJax and translation to MathML. LaTeX



Reproducible Research with RMarkdown RStudio

X:/projects/Test - RStudio

File Edit Code View Plots Session Build Debug Tools Help

Go to file/function

ormats.r * Untitled5* * rmdtest2.Rmd* * test.css * st >>

Knit HTML Chunks

```

1 The Analysis (own CSS applied)
2
3
4 This is an R Markdown document. Markdown is a
  simple formatting syntax for authoring web pages.
5
6 An unordered list:
7 - item
8 - item
9
10 when you click the Knit HTML button a web page
   will be generated that includes both content as
   well as the output of any embedded R code chunks
   within the document. You can embed an R code chunk
   like this:
11
12 {r, results='asis', echo=FALSE}
13 names <- c("John", "Marry", "Amana", "Adrian",
14            "James", "Peter", "Salad")
15 surnames <- c("Kowalski", "Doe", "Papanouk",
16              "Bond", "Fingers", "Lomax")
17 data <- data.frame(Name=names[1:nrow(data)],
18                   Surname=surnames[1:nrow(data)],
19                   , 1, 6))
20
21 You can also embed plots, for example:
22 {r fig.width=7, fig.height=6}

```

2:40 (Top Level)

Console X:/projects/Test/

```

+ function(inputFile, outputFile) {
+   require(markdown)
+   markdownToHTML(inputFile, outputFile)
+ }
Loading required package: markdown
>

```

RStudio: Preview HTML

Preview: X:/projects/Test/rmdtest2.html Log Save As Publish Find

The Analysis (own CSS applied)

This is an R Markdown document. Markdown is a simple formatting syntax for authoring web pages.

An unordered list:

- item
- item

When you click the **Knit HTML** button a web page will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

| ID | Name | Surname |
|----|--------|----------|
| 1 | Adrian | Bond |
| 2 | Amana | Doe |
| 3 | Marry | Bond |
| 4 | Marry | Papanouk |
| 5 | Adrian | Doe |

...s, for example:

Sample plot

Listner - [X:\projects\Test\test.css]

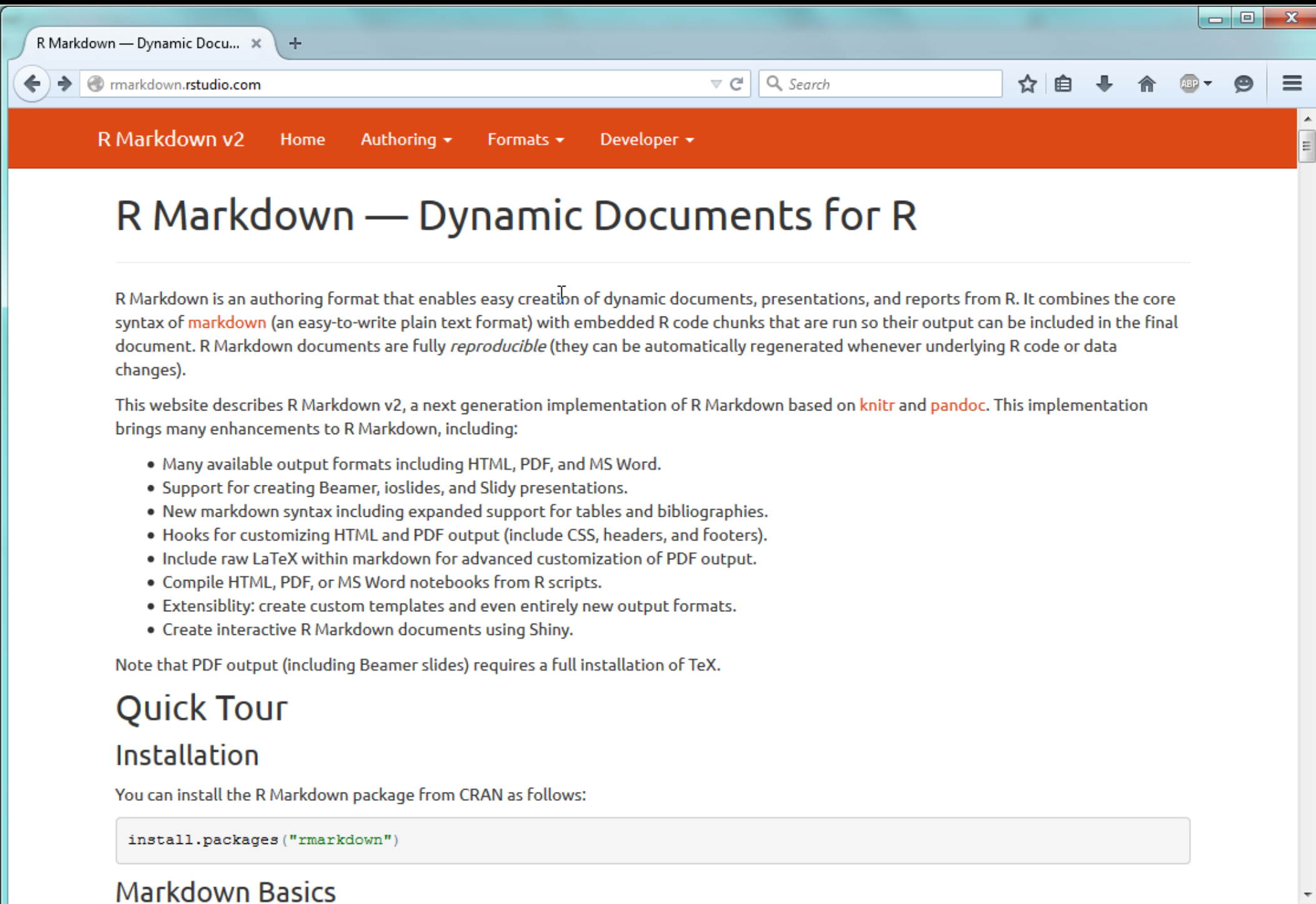
File Edit Options Encoding Help 100 %

```

td {
  background-color: #dcdcdc;
  text-align:right;
}
th:nth-child(1){
  width:50px;
}
th:nth-child(2){
  width:150px;
}
th:nth-child(3){
  width:150px;
}

```

Take a moment and discover R Markdown!



The image shows a browser window displaying the R Markdown website. The browser's address bar shows the URL `rmarkdown.rstudio.com`. The website has a navigation bar with links for "R Markdown v2", "Home", "Authoring", "Formats", and "Developer". The main heading is "R Markdown — Dynamic Documents for R". The text describes R Markdown as an authoring format for dynamic documents, presentations, and reports, combining markdown with embedded R code. It lists several enhancements in v2, such as support for Beamer presentations, new markdown syntax for tables and bibliographies, and the ability to create interactive documents. A note mentions that PDF output requires TeX. The page also includes sections for "Quick Tour", "Installation", and "Markdown Basics".

R Markdown — Dynamic Documents for R

R Markdown is an authoring format that enables easy creation of dynamic documents, presentations, and reports from R. It combines the core syntax of `markdown` (an easy-to-write plain text format) with embedded R code chunks that are run so their output can be included in the final document. R Markdown documents are fully *reproducible* (they can be automatically regenerated whenever underlying R code or data changes).

This website describes R Markdown v2, a next generation implementation of R Markdown based on `knitr` and `pandoc`. This implementation brings many enhancements to R Markdown, including:

- Many available output formats including HTML, PDF, and MS Word.
- Support for creating Beamer, ioslides, and Slidy presentations.
- New markdown syntax including expanded support for tables and bibliographies.
- Hooks for customizing HTML and PDF output (include CSS, headers, and footers).
- Include raw LaTeX within markdown for advanced customization of PDF output.
- Compile HTML, PDF, or MS Word notebooks from R scripts.
- Extensibility: create custom templates and even entirely new output formats.
- Create interactive R Markdown documents using Shiny.

Note that PDF output (including Beamer slides) requires a full installation of TeX.

Quick Tour

Installation

You can install the R Markdown package from CRAN as follows:

```
install.packages("rmarkdown")
```

Markdown Basics



Reproducible Research with ODF – **odfWeave**

odfWeave is an adaptation of Sweave for **OpenDocument** files. Such files serve as a template of the report with R code embedded in. MS Office, Open/Libre Office and Calligra Suite (formerly KOffice) work with the OpenDocument format smoothly.

odfWeave is flexible and simple to use. It allows to create templates directly in a word processor. This is **the most convenient way** to accomplish the Reproducible Research goals. Styling can be applied (in a limited extent) from the R code and, mostly – in the editor.

The R code may be embedded into a template document or stored in a separate file, which is then referenced in the template. Both approaches have their pros and cons. The second approach is particularly useful, when the same code is shared between (reused in) many documents.



Reproducible Research with ODF – **odfWeave**

Unfortunately, some serious caveats and limitations must be noted:

- MS Word incorrectly saves a definition of autonumbered captions which causes all captions begin from: "Table: 999"
- MS Windows does not honor UTF-8 locale. Therefore odfWeave cannot handle mixed characters from different alphabets (óęléüš яфы βθΣΩ), especially ASCII (English) and non-ASCII (Russian, Greek). There are four options to solve this issue:
 - process document under Linux or MacOS, which are Unicode-enabled
 - use old good non-Unicode fonts, one for each language, e.g. Symbol for Greek characters and math symbols.
 - replace all non-ASCII characters with their ASCII equivalents in the fetched data by using **iconv(text, "UTF-8", "ASCII//TRANSLIT")**
 - Ä, Á, Å --> A | Ë, É, Ě --> E | Õ, Ó --> O (U)
 - Ń, Ň --> N | Ü, Ó --> U | Š, Ş, Ś --> S etc.

From an ODF template...

Very Important Analysis Template.odt - OpenOffice Writer

Plik Edytuj Widok Wstaw Format Tabela Narzędzia Okno Pomoc

Normalny Calibri 11 G K P

The analysis

a few words of description

| Sample header | |
|-------------------|-----------------------|
| Date of creation: | Under request number: |
| Created by: | Validated by: |
| Accepted by: | Submitted by: |

1. Introduction

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nulla mattis, turpis non tincidunt dictum, leo turpis porttitor felis, nec vulputate sem turpis vitae purus.

Sed varius scelerisque rhoncus. Integer pellentesque tristique luctus. Fusce nec odio lorem. Aliquam finibus sapien risus, sit amet venenatis arcu pharetra vitae. Aliquam accumsan eu odio id finibus.

2. Description of the data

Praesent pellentesque condimentum vestibulum. Curabitur rutrum, eros sit amet varius imperdiet, risus ex efficitur felis, a scelerisque expurus id lectus. Vestibulum ac fermentum tortor. Maecenas ut tempus lacus.

```
<<load_libs, results=hide, echo=FALSE>>=
library(RODBC)
library(ggplot2)
@
<<set_styles, results=hide, echo=FALSE>>=
currentDefs <- getStyleDefs()
currentDefs$TableCellValue <- currentDefs$ArialNormal
currentDefs$TableCellValue$fontSize <- "10pt"
currentDefs$TableHeaderValue <- currentDefs$ArialNormal
currentDefs$TableHeaderValue$fontType <- "bold"
currentDefs$TableHeaderValue$fontSize <- "10pt"
currentDefs$TableHeaderValue$marginBottom <- "5px"
currentDefs$TableHeaderValue$borderBottom <- "1px solid black"
@
```

```
<<get_prepare_data, results=hide, echo=FALSE>>=
con <- odbcConnect("test1")
query <- "SELECT s.ID,
          s.Name + ' ' + s.Surname AS Subject,
          rr.Time,
          rr.RR_SYS, rr.RR_DIA, rr.BPM
FROM Subject s
INNER JOIN RR rr ON s.ID = rr.SubjID"
result <- sqlQuery(con, query)
odbcClose(con)
result$Time <- as.POSIXct(strptime(result$Time, "%Y-%m-%d %H:%M:%S"))
@
```

2.1 Summary

The dataset contains $\backslash\text{Sexpr}\{ \text{nrow}(\text{result}) \}$ cases and $\backslash\text{Sexpr}\{ \text{length}(\text{result}\$Subject) \}$ subjects.

List of subjects in the data set:

```
<<subj_list, results=xml, echo=FALSE>>=
odfTableCaption("The content of the dataset")
odfTable(result, useRowNames=F)
@
```

2.1.1 Special case of Adrian Olszewski

Aliquam arcu ligula, egestas a feugiat quis, pulvinar bibendum lectus. Fusce lacina ante est, sit amet malesuada nisi dui, sit amet mattis dolor malesuada pretium.

```
<<olszewski_data, results=xml, echo=FALSE>>=
source("name_of_file_with_R_code.r")
@
```

The graphical representation is as follows:

```
<<olszewski_data_plot, results=xml, echo=FALSE, fig=TRUE>>=
par(mar=c(6, 4, 4, 2) + 0.5)
```

much better option:
don't embed a code into the template,
put it in a separate file and refer to it:
source("an019_32.r")

Normalny Notatki Numeracja 1 Numeracja 2 Numeracja 3

Właściwości

Tekst

Calibri

G K P

Akapit

Odstęp:

0,00 "

0,11 "

Strona

Strona 1 / 3 MPO Angielski (Wielka Brytania) WSTAW STD * 75 %



...via the odfWeave engine...

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for the `odfweave` function, which packages a file into a zip archive.
- Environment:** Shows the global environment with variables like `currentDef...`, `currentsty...`, `query`, `result`, `soleszewski`, and `tck`.
- Console:** Shows the execution output, including file removal, packaging with 7z, and compression of various XML and PNG files.
- Plots:** A line plot titled "Blood pressure of: Olszewski Adrian, #1" showing SYS and DIA values over time.

```
1 # Sys.setlocale(category = "LC_ALL", locale = "Polish_Poland.1250")
2
3 odfweave("c:/tmp/Very Important Analysis Template.odt",
4         "c:/tmp/Very Important Analysis.odt",
5         control = odfweaveControl(zipCmd = c("7z a -tzip $$file$$ .",
6                                             "7z x -tzip $$file$$"))))
7
8
```

Console Output:

```
Removing extra files

Packaging file using 7z a -tzip "Very Important Analysis Template.odt" .
7-Zip 9.20 Copyright (c) 1999-2010 Igor Pavlov 2010-11-18
Scanning

Creating archive Very Important Analysis Template.odt

Compressing  Configurations2\accelerator\current.xml
Compressing  content.xml
Compressing  layout-cache
Compressing  manifest.rdf
Compressing  META-INF\manifest.xml
Compressing  meta.xml
Compressing  mimetype
Compressing  Pictures\content_1-olszewski_data_plot.png
Compressing  settings.xml
Compressing  styles.xml
Compressing  Thumbnails\thumbnail.png

Everything is Ok
Copying Very Important Analysis Template.odt
Resetting wd
Removing C:\Users\...\AppData\Local\Temp\RtmpEDJzji\odfweave040534483
41

Done
>
```

Plot Data:

| Time | SYS | DIA |
|--------------|-----|-----|
| Jan-02 10:10 | 135 | 75 |
| Jan-02 15:30 | 140 | 80 |
| Jan-02 18:25 | 135 | 75 |
| Jan-03 10:45 | 140 | 85 |
| Jan-03 15:30 | 135 | 75 |
| Jan-04 11:30 | 145 | 90 |

...to the final document

Very Important Analysis.odt - OpenOffice Writer

Plik Edytuj Widok Wstaw Format Tabela Narzędzia Okno Pomoc

Domyślnie Calibri 11 G K P

2 1 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

| Sample header | |
|-------------------|-----------------------|
| Date of creation: | Under request number: |
| Created by: | Validated by: |
| Accepted by: | Submitted by: |

1. Introduction

..... STRIPPED

2. Description of the data

..... STRIPPED

2.1 Summary

The dataset contains 9 cases and 6 variables.

List of subjects in the data set:

- Adrian Olszewski
- Sheryll Cwiąkalska-Zółć
- John Doe

Table 1: The content of the dataset

| ID | Subject | Time | RR_SYS | RR_DIA | BPM |
|----|-------------------------|---------------------|--------|--------|-----|
| 1 | Adrian Olszewski | 2014-01-02 10:10:00 | 138 | 78 | 76 |
| 1 | Adrian Olszewski | 2014-01-02 15:30:00 | 142 | 82 | 99 |
| 1 | Adrian Olszewski | 2014-01-02 18:25:00 | 137 | 78 | 82 |
| 1 | Adrian Olszewski | 2014-01-03 10:45:00 | 142 | 88 | 87 |
| 1 | Adrian Olszewski | 2014-01-03 15:30:00 | 138 | 74 | 76 |
| 1 | Adrian Olszewski | 2014-01-04 11:30:00 | 146 | 91 | 91 |
| 2 | Sheryll Cwiąkalska-Zółć | 2014-01-02 10:20:00 | 127 | 68 | 72 |
| 2 | Sheryll Cwiąkalska-Zółć | 2014-01-02 15:30:00 | 131 | 78 | 76 |
| 3 | John Doe | 2014-01-03 10:20:00 | 127 | 69 | 67 |

Table 2: Blood pressure of Adrian Olszewski

| ID | Subject | Time | RR_SYS | RR_DIA | BPM |
|----|------------------|---------------------|--------|--------|-----|
| 1 | Adrian Olszewski | 2014-01-02 10:10:00 | 138 | 78 | 76 |
| 1 | Adrian Olszewski | 2014-01-02 15:30:00 | 142 | 82 | 99 |
| 1 | Adrian Olszewski | 2014-01-02 18:25:00 | 137 | 78 | 82 |
| 1 | Adrian Olszewski | 2014-01-03 10:45:00 | 142 | 88 | 87 |
| 1 | Adrian Olszewski | 2014-01-03 15:30:00 | 138 | 74 | 76 |
| 1 | Adrian Olszewski | 2014-01-04 11:30:00 | 146 | 91 | 91 |

The graphical representation is as follows:

Blood pressure of: Olszewski Adrian, #1

| Time | SYS | DIA |
|--------------|-----|-----|
| Jan-02 10:10 | 138 | 78 |
| Jan-02 15:30 | 142 | 82 |
| Jan-02 18:25 | 137 | 78 |
| Jan-03 10:45 | 142 | 88 |
| Jan-03 15:30 | 138 | 74 |
| Jan-04 11:30 | 146 | 91 |

Illustration 1: Blood pressure of Adrian Olszewski

Strona 2 / 2 | MP0 | Angielski (Wielka Brytania) | WSTAW | STD * | 70 %

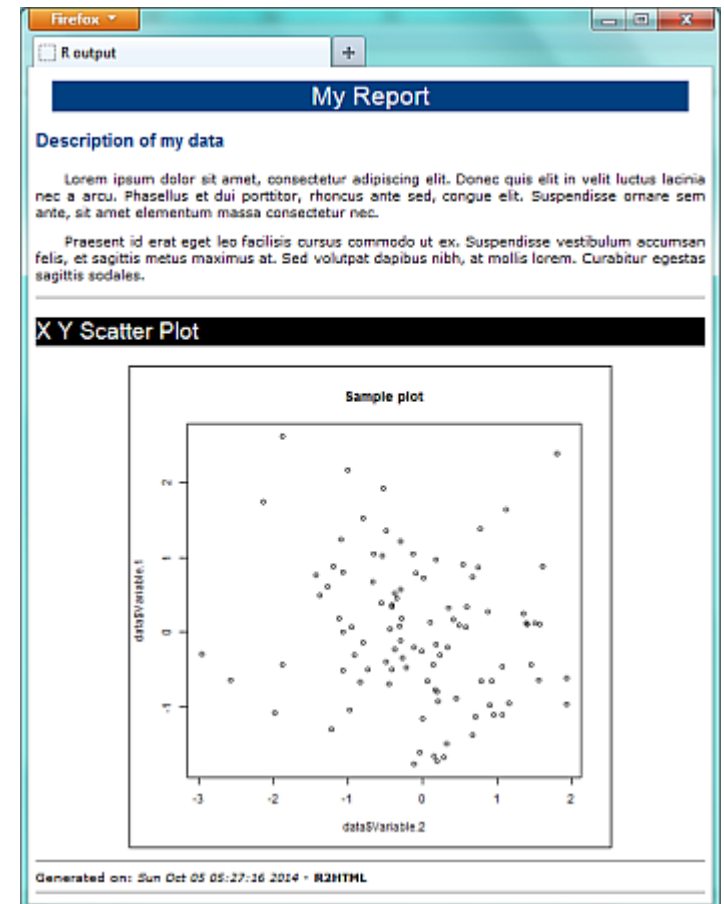


Let's do the same other **R** way

R allows the user to create documents directly from code. [rtf](#), [R2HTML](#), [R2PPT](#) and [ReporteRs](#) packages do the job. This is similar to how it is done in SAS.

The philosophy of document creation here is different from the approach taken in [odfWeave](#) or [knitr](#). There is **no mixing** formatted paragraphs with R code. Every single object (literal, paragraph, table, graphics) is created by a specific function, **directly in R**.

```
data <- data.frame("Var 1" = rnorm(100),
                  "Var 2" = rnorm(100))
HTMLStart(outdir="c:/mydir", file="myreport",
          extension="html", echo=FALSE...)
HTML.title("My Report", HR=1)
HTML.title("Description of my data", HR=3)
HTML("Lorem ipsum dolor sit amet,...")
HTML("Praesent id erat eget leo facilisis...")
summary(data)
HTMLhr()
HTML.title("X Y Scatter Plot", HR=2)
plot(data$Var.1~data$Var.2, main="Sample plot")
HTMLplot()
HTMLStop()
```



R2HTML is a well known yet a bit outdated tool...

Firefox

Rnews_2003-3.pdf

Page: 33 of 45 Automatic Zoom

limitations

The R2HTML Package

Introduction to HTML and the R2HTML package

Generating HTML output on the fly

Creating personalized reports

A simple analysis

A template scheme to complete the report

Going one step further with CSS

Summary

R Help Desk

Preface

Documentation on package management

Libraries

Source vs. binary packages

Installing and updating source packages

Package management on the Macintosh

Package management on Windows

Programmer's Niche

References

Recent Events

R at the Statistics Canada Symposium

Upcoming Events

useR! 2004

Vol. 3/3, December 2003 33

The R2HTML Package

Formatting HTML output on the fly or by using a template scheme

By Eric Lecoutre

Statistics is not only theory and methodology, but also computing and communication. Applied statisticians are aware that they have to pay particular attention to the last step of an analysis: the report. A very elegant way to handle the final report with R is to use the wonderful Sweave system (Leisch, 2002a) in package **tools**: not only does it allow professional quality reports by using \LaTeX , but it also stores the code that is used for the report within the document, which is very useful when returning to the analysis later on. This solution, however, is not always applicable, as the user may not be familiar with \LaTeX or may need another format to communicate with a client, who, in many cases, will expect a report that he can edit or to which he can add some details. RTF format is ideal for this type of communication, as it can be opened on many platforms and allows some formatting enhancements (bold, tables, ...). Nevertheless, it is not easy to produce and does not enable the user to embed graphs. Another universal format which can achieve our goal is HTML: it is lightweight, readable on almost all platforms, editable, and allows graphs. Moreover, it can easily be exported to other formats.

Consequently, the only required knowledge in order to write basic HTML documents is the list of existing tags and their functionality. By way of illustration, here is the structure of a (rather basic) HTML document:

```
<html>
<h1>My first HTML page </h1>
<p>This is some basic text with a
  <b>bold</b> word.</p>
<p>It uses h1, and p tags which allows to create
  a title and to define a paragraph</p>
</html>
```

Now, we have a very easy way to create our first webpage from R: simply using the `cat` function to write text into an external file. In the following example, consider how we call the `cat` function several times having set the `append` argument to `TRUE` in order to add information to the page.

```
> htmlfile = file.path(tempdir(),
+ "page1.html")
> cat("<html><h1>My first HTML page from R</h1>",
+ file = htmlfile)
> cat("\n<br>Hello Web World!",
+ append = TRUE, file = htmlfile)
> cat("\n</html>", append = TRUE,
+ file = htmlfile)
```

The package **R2HTML** contains a collection of wrapper functions that call the `cat` function to write

...as well as **rtf** package...

```
RGui (32-bit) - [R Console]
File Edit View Misc Packages Windows Help

> as.data.frame(round(coef(s2), 2))
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.07      0.24   0.28   0.78
x1           0.92      0.36   2.52   0.01
x2           0.89      0.19   4.57   0.00

> outtab
      Model 1      Model 2
Intercept "0.61 (0.23)" "0.07 (0.24)"
x1        "0.79 (0.4)"  "0.92 (0.36)"
x2        ""           "0.89 (0.19)"
sigma     "1.06"       "0.97"
Adj. R-Squared "0.03"    "0.19"
n         "100"       "100"

> library(rtf)
> rtffile <- RTF("test.rtf")
> addParagraph(rtffile, "This is the output of a regression coeff")
> addTable(rtffile, as.data.frame(round(coef(s2), 2)))
> addParagraph(rtffile, "\n\nThis is the nicer looking table we m")
> addTable(rtffile, cbind(rownames(outtab), outtab))
> done(rtffile)
```

test.rtf [Tryb zgodności] - Word

PLIK NARZ WSTA PROJE UKŁA ODW KORE RECE WIDO Olszewski...

This is the output of a regression coefficients:

Estimate	Std. Error	t value	Pr(> t)
0.07	0.24	0.28	0.78
0.92	0.36	2.52	0.01
0.89	0.19	4.57	0

This is the nicer looking table we made above:

V1	Model 1	Model 2
Intercept	0.61 (0.23)	0.07 (0.24)
x1	0.79 (0.4)	0.92 (0.36)
x2		0.89 (0.19)
sigma	1.06	0.97
Adj. R-Squared	0.03	0.19
n	100	100



...but the new wave comes! The **ReporteRs**

ReporteRs Getting Started Formatting content Word PowerPoint HTML Main functions

ReporteRs

An R package to generate Microsoft Word, Microsoft PowerPoint and HTML reports.



PowerPoint documents

Create Powerpoint presentations from R with your corporate template. Produce nice outputs for graphical, tabular and textual reporting.

[More](#)



Word documents

Create your own Word documents from R. Use your corporate template. Replace content in existing Word documents. Produce nice outputs for graphical, tabular and textual reporting.

[More](#)



HTML documents

Create your own HTML documents from R. Document will be styled with bootstrap. Produce nice outputs for graphical, tabular and textual reporting.

[More](#)



FlexTable objects

A set of tools to customize tables, their formats and their contents

[More](#)



pot objects

Create and format any text from R with pot objects.

[More](#)



Markdown

Use markdown for text output in documents.

[More](#)



Editable vector graphics

Add R plots as **editable** vector graphics into documents.

[More](#)



Package rtable

rtable will let you transform **xtable** objects into FlexTable

[More](#)

<http://davidgohel.github.io/ReporteRs/index.html>



[ReporteRs](#) is just outstanding package which allows to create [Word](#), [PowerPoint](#) (OpenXML) and [HTML5 presentations](#) without need to have MS Office installed.

It is hard to enumerate all capabilities offered by the package:

- [Styles](#) – use the power of styles to easy control format of each object, and
 - add auto-numbered captions to tables and graphics. These captions are recognizable by TOC generator, so one can easily add [a list of tables](#).
 - add nested, numbered or bulleted lists (also from code)
- [Templates](#) – use predefined styles, prepare overall layout of a document and fill it with a content, replacing “bookmarks” with tables and graphics
- [Editable, scalable graphics](#) (requires MS Office > 2007) – must see this!
- [Flexible tables](#), with cells spanning multiple rows and columns, advanced, conditional formatting, “zebra style”. FlexTables are [compatible with knitr](#).
- [Multilevel ordered lists](#) with advanced, level-dependent, numbering
- [Footnotes](#), “pot objects” and many more!



```

RGui (32-bit) - [R Console]
File Edit View Misc Packages Windows Help installr
> options("ReporteRs-default-font"="Arial")
> doc = docx( title = 'My document' )
> doc = addTitle( doc , 'First 5 lines of iris', level = 1)
> doc = addTable( doc , iris[1:5, ], layout.properties= get$
> doc = addParagraph( doc, "My table", stylename = "rTableL$
> doc = addTitle( doc , 'ggplot2 example', level = 1)
> myggplot = qplot(Sepal.Length, Petal.Length, data = iris,$
> doc = addPlot( doc = doc , fun = print, x = myggplot )
> writeDoc( doc, 'c:/tmp/my_first_doc.docx' )

```

my_first_doc.docx - Word

PLIK NARZ WSTA PROJE UKŁA ODW KORE RECE WIDO Olszewski...

1 FIRST 5 LINES OF IRIS

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa

Table 1: MY TABLE

2 GGLOT2 EXAMPLE

Main functions list

All these functions are documented in the package.

The following functions can be used whatever the output format is (docx, pptx, html):

- **addTitle**: Add a title
- **addFlexTable**: Add a FlexTable
- **addPlot**: Add plots
- **addImage**: Add external images
- **addRScript**: Add syntax highlighted R code
- **addParagraph**: Add paragraphs of text
- **addTable**: Add a table
- **writeDoc**: Write the document into a file

The following functions can only be used when the output format is docx.

- **styles**: Get available styles
- **addTOC**: Add a table of contents
- **addPageBreak**: Add a page break
- **dim**: Get page dimensions

The following functions can only be used when the output format is pptx.

- **slide.layouts**: Get available layout names
- **addSlide**: Add a slide
- **addDate**: Add a date
- **addPageNumber**: Add a page number
- **addFooter**: Add a comment in the footer
- **addSubtitle**: Add a sub title
- **dim**: Get shape dimensions

The following functions can only be used when the output format is html.

- **addPage**: Add a page



ReporteRs - tables

	cyl	gear	carb	disp	mpg	wt
4	3	1		120.1	21.50	2.465
	4	1		84.2	29.10	2.072
	4	2		121.0	24.75	2.684
	5	2		107.7	28.20	1.827
6	3	1		241.5	19.75	3.337
	4	4		163.8	19.75	3.094
	5	6		145.0	19.70	2.770
8	3	2		345.5	17.15	3.560
	3	3		275.8	16.30	3.860
	3	4		416.4	12.62	4.686
	5	4		351.0	15.80	3.170
	5	8		301.0	15.00	3.570

	Estimate	Std. Error	t value	Pr(> t)	Signif
(Intercept)	33.991	1.888	18.006	< 0.001	***
wt	-3.206	0.754	-4.252	< 0.001	***
cyl6	-4.256	1.386	-3.070	0.00472	**
cyl8	-6.071	1.652	-3.674	< 0.001	***

	Status	Ulceration	n	Mean	SD	Median	Min	Max	Missing
Alive	Alive	Absent	92	1.635	1.927	1.13	0.10	12.88	0
	Alive	Present	42	3.581	2.578	3.06	0.32	12.24	0
Melanoma	Melanoma	Absent	16	2.703	3.350	1.94	0.32	14.66	0
	Melanoma	Present	41	4.939	3.498	4.04	0.97	17.42	0
Non-melanoma	Non-melanoma	Absent	7	2.097	1.928	1.45	0.65	6.12	0
Non-melanoma	Non-melanoma	Present	7	5.339	4.328	4.84	0.16	12.56	0

nitro	Golden Rain		Marvellous		Victory	
	est	df	est	df	est	df
0	4.35 (0.08)	11.8	4.41 (0.08)	11.8	4.28 (0.08)	11.8
0.2	4.58 (0.07)	10.3	4.64 (0.07)	10.3	4.50 (0.07)	10.3
0.4	4.73 (0.07)	10.3	4.79 (0.07)	10.3	4.65 (0.07)	10.3
0.6	4.81 (0.08)	11.8	4.86 (0.08)	11.8	4.73 (0.08)	11.8

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
sex	1	75.37	75.37	0.38	0.5417
ethnicity	3	2572.15	857.38	4.27	0.0072
grade	1	36.31	36.31	0.18	0.6717
disadv	1	59.30	59.30	0.30	0.5882
Residuals	93	18682.87	200.89		

Summary statistics for thickness

Status	Gender	Ulceration	Count	Mean (\pm sd) *
Alive	Female	Absent	68	1.693 (\pm 2.004)
		Present	23	2.972 (\pm 2.593)
	Male	Absent	24	1.468 (\pm 1.719)
		Present	19	4.319 (\pm 2.423)
Melanoma	Female	Absent	8	2.139 (\pm 1.184)
		Present	20	4.724 (\pm 4.128)
	Male	Absent	8	3.266 (\pm 4.681)
		Present	21	5.143 (\pm 2.862)
Non-melanoma	Female	Absent	3	1.667 (\pm 1.141)
		Present	4	3.302 (\pm 3.713)
	Male	Absent	4	2.420 (\pm 2.499)
		Present	3	8.053 (\pm 4.019)

* This is a footer note

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
mpg	1.000	-0.852	-0.848	-0.776	0.681	-0.868	0.419	0.664	0.600	0.480	-0.551
cyl	-0.852	1.000	0.902	0.832	-0.700	0.782	-0.591	-0.811	-0.523	-0.493	0.527
disp	-0.848	0.902	1.000	0.791	-0.710	0.888	-0.434	-0.710	-0.591	-0.556	0.395
hp	-0.776	0.832	0.791	1.000	-0.449	0.659	-0.708	-0.723	-0.243	-0.126	0.750
drat	0.681	-0.700	-0.710	-0.449	1.000	-0.712	0.091	0.440	0.713	0.700	-0.091
wt	-0.868	0.782	0.888	0.659	-0.712	1.000	-0.175	-0.555	-0.692	-0.583	0.428
qsec	0.419	-0.591	-0.434	-0.708	0.091	-0.175	1.000	0.745	-0.230	-0.213	-0.656
vs	0.664	-0.811	-0.710	-0.723	0.440	-0.555	0.745	1.000	0.168	0.206	-0.570
am	0.600	-0.523	-0.591	-0.243	0.713	-0.692	-0.230	0.168	1.000	0.794	0.058
gear	0.480	-0.493	-0.556	-0.126	0.700	-0.583	-0.213	0.206	0.794	1.000	0.274
carb	-0.551	0.527	0.395	0.750	-0.091	0.428	-0.656	-0.570	0.058	0.274	1.000



[rtable](#) is a set of functions to make easy tabular reporting from R with the ReporteRs package. It easily [integrates](#) with knitr and shiny.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
sex	1	75.37	75.37	0.38	0.5417
ethnicity	3	2572.15	857.38	4.27	0.0072
grade	1	36.31	36.31	0.18	0.6717
disadv	1	59.30	59.30	0.30	0.5882
Residuals	93	18682.87	200.89		

[xtable](#) → FlexTable

		Cylinders	4	4	6	6	8	8
		Transmission	0	1	0	1	0	1
V/S	Gears							
0	3	0	0	0	0	12	0	
0	4	0	0	0	2	0	0	
0	5	0	1	0	1	0	2	
1	3	1	0	2	0	0	0	
1	4	2	6	2	0	0	0	
1	5	0	1	0	0	0	0	

[ftable](#) → FlexTable



		High	Low	Middle	Sum
current	#	51	43	22	116
	%	14.33	12.08	6.18	32.58
	row %	43.97	37.07	18.97	
	col %	24.17	46.24	42.31	
former	#	92	28	21	141
	%	25.84	7.87	5.90	39.61
	row %	65.25	19.86	14.89	
	col %	43.60	30.11	40.38	
never	#	68	22	9	99
	%	19.10	6.18	2.53	27.81
	row %	68.69	22.22	9.09	
	col %	32.23	23.66	17.31	
Sum	#	211	93	52	356
	%	59.27	26.12	14.61	100.00

freqtable

nitro	Golden Rain		Marvellous		Victory	
	est	df	est	df	est	df
0	4.35 (0.08)	11.8	4.41 (0.08)	11.8	4.28 (0.08)	11.8
0.2	4.58 (0.07)	10.3	4.64 (0.07)	10.3	4.50 (0.07)	10.3
0.4	4.73 (0.07)	10.3	4.79 (0.07)	10.3	4.65 (0.07)	10.3
0.6	4.81 (0.08)	11.8	4.86 (0.08)	11.8	4.73 (0.08)	11.8

pivot table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
sex	1	75.37	75.37	0.38	0.5417
ethnicity	3	2572.15	857.38	4.27	0.0072
grade	1	36.31	36.31	0.18	0.6717
disadv	1	59.30	59.30	0.30	0.5882
Residuals	93	18682.87	200.89		

custom forms



ReporteRs – editable graphics

word_simple_example.docx - Word

PLIK NARZĘDZIA GŁ WSTAWIANIE PROJEKTOWA UKŁAD STRON ODWOŁANIA KORESPONDEN RECENZJA WIDOK **FORMATOWANIE** Olszewski...

Kształty

Wstawianie kształt... Style kształtów Style WordArt Tekst Rozmieszczanie

2 1 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

Species

- setosa
- versicolor
- virginica

Petal.Width

- 0.5
- 1.0
- 1.5
- 2.0
- 2.5

Kolory motywu

Kolory standardowe

- Brak konturu
- Więcej kolorów konturów...
- Grubość
- Kreski

Styl Wypełnienie Kontur

STRONA 1 Z 1 WYRAZY: 19 ANGIELSKI (ZJEDNOCZONE KRÓLESTWO) 70%

ReporteRs – replacing content in a template

word_bookmark_template.docx - Word

PLIK NARZE WSTA PROJE UKŁAD ODWO KORES RECEN WIDOK Olszewski...

Authors AUTHOR Reviewers REVIEWER

PLOT

Illud tamen clausos vehementer angebat quod captis navigiis, qua flumen, Isauri quidem alimentorum copiis adfluebant, ipsi vero so consumendo inediae propinquantis aerumnas exitialis horrebant.

Illud tamen clausos vehementer angebat quod captis navigiis, qua flumen, Isauri quidem alimentorum copiis adfluebant, ipsi vero so consumendo inediae propinquantis aerumnas exitialis horrebant.

Below a table

Illud tamen clausos vehementer angebat quod captis navigiis, qua flumen, Isauri quidem alimentorum copiis adfluebant, ipsi vero so consumendo inediae propinquantis aerumnas exitialis horrebant.

STRONA 2 Z 2 WYRAZY: 96

word_replacement.docx - Word

PLIK NARZE WSTA PROJE UKŁAD ODWO KORES RECEN WIDOK Olszewski...

Petal.Length

Sepal.Length

Petal.Width

- 0.5
- 1.0
- 1.5
- 2.0
- 2.5

Illud tamen clausos vehementer angebat quod captis navigiis, quae frumenta vehebant per flumen, Isauri quidem alimentorum copiis adfluebant, ipsi vero solitarum rerum cibos iam consumendo inediae propinquantis aerumnas exitialis horrebant.

Illud tamen clausos vehementer angebat quod captis navigiis, quae frumenta vehebant per flumen, Isauri quidem alimentorum copiis adfluebant, ipsi vero solitarum rerum cibos iam consumendo inediae propinquantis aerumnas exitialis horrebant.

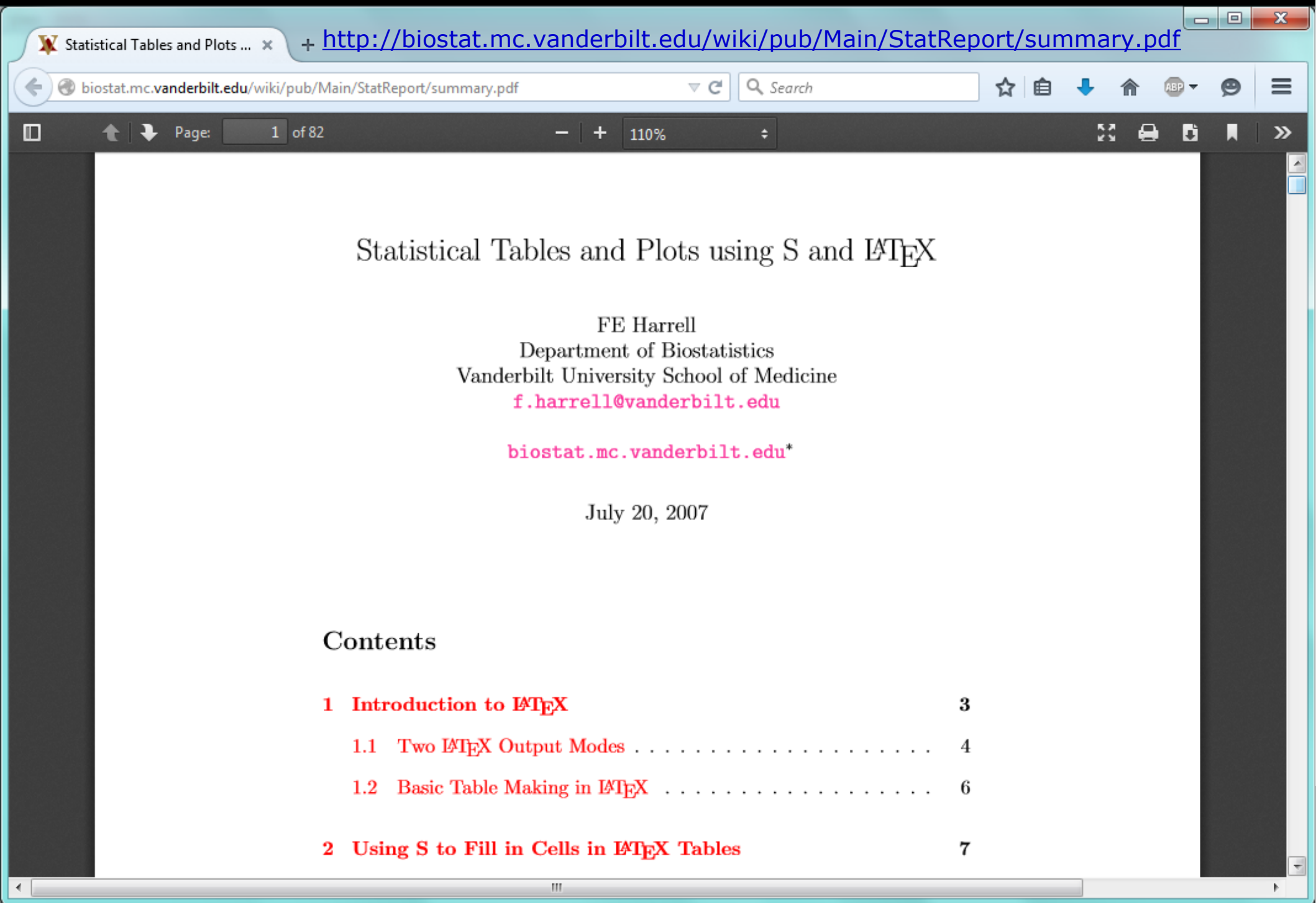
Below a table

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet											

STRONA 2 Z 3 WYRAZY: 237

70%

What about those fancy, publication-Ready, tables?



Statistical Tables and Plots using S and LaTeX

FE Harrell
Department of Biostatistics
Vanderbilt University School of Medicine
f.harrell@vanderbilt.edu

biostat.mc.vanderbilt.edu*

July 20, 2007

Contents

1 Introduction to LaTeX	3
1.1 Two LaTeX Output Modes	4
1.2 Basic Table Making in LaTeX	6
2 Using S to Fill in Cells in LaTeX Tables	7



What about those fancy, publication-Ready, tables?

stargazer

beautiful LATEX, HTML and
ASCII tables from R
statistical output

Marek Hlavac
Harvard University

Table 7: Correlation Matrix

	rating	complaints	privileges
rating	1	0.825	0.426
complaints	0.825	1	0.558
privileges	0.426	0.558	1

Table 1:

Statistic	N	Mean	St. Dev.	Min	Max
rating	30	64.633	12.173	40	85
complaints	30	66.600	13.315	37	90
privileges	30	53.133	12.235	30	83
learning	30	56.367	11.737	34	75
raises	30	64.633	10.397	43	88
critical	30	74.767	9.895	49	92
advance	30	42.933	10.289	25	72
high.rating	30	0.333	0.479	0	1

Table 3: Results

	<i>Dependent variable:</i>		
	rating <i>OLS</i>		high.rating <i>probit</i>
	(1)	(2)	(3)
complaints	0.692*** (0.149)	0.682*** (0.129)	
privileges	-0.104 (0.135)	-0.103 (0.129)	
learning	0.249 (0.160)	0.238* (0.139)	0.164*** (0.053)
raises	-0.033 (0.202)		
critical	0.015 (0.147)		-0.001 (0.044)
advance			-0.062 (0.042)
Constant	11.011 (11.704)	11.258 (7.318)	-7.476** (3.570)
Observations	30	30	30
R ²	0.715	0.715	
Adjusted R ²	0.656	0.682	
Log likelihood			-9.087
Akaike Inf. Crit.			26.175
Residual Std. Error	7.139(df = 24)	6.863(df = 26)	
F statistic	12.063***(df = 5; 24)	21.743***(df = 3; 26)	

Note:

*p<0.1; **p<0.05; ***p<0.01

What about those fancy, publication-ready, tables?

reporttools

R Functions to Generate
LATEX Tables of
Descriptive Statistics

Kaspar Rubach
University of Zurich

Variable	Levels	n	Min	\tilde{x}	\bar{x}_{trim}	Max	IQR	c_v	s	#NA
Age	no	34	8.8	46.3	44.0	59.1	12.5	0.3	11.4	0
	yes	69	19.6	48.0	46.4	64.4	9.5	0.2	8.9	0
$p = 0.39$	all	103	8.8	47.8	45.8	64.4	10.9	0.2	9.8	0
Follow up time	no	34	0.0	20.0	57.9	1400.0	41.5	2.6	250.5	0
	yes	69	4.0	206.0	374.7	1799.0	548.0	1.1	458.9	0
$p = 1.1e-08$	all	103	0.0	89.0	259.0	1799.0	379.0	1.4	428.3	0
Mismatch score	no	NA	NA	NA	NA	NA	NA	NA	NA	NA
	yes	65	0.0	1.1	1.1	3.0	0.8	0.5	0.6	4
	all	65	0.0	1.1	1.1	3.0	0.8	0.5	0.6	38

Variable	Levels	n_{no}	$\%_{no}$	$\sum \%_{no}$	n_{yes}	$\%_{yes}$	$\sum \%_{yes}$	n_{all}	$\%_{all}$	$\sum \%_{all}$
Surgery	no	21	91.3	91.3	35	76.1	76.1	56	81.2	81.2
	yes	2	8.7	100.0	11	23.9	100.0	13	18.8	100.0
$p = 0.19$	all	23	100.0		46	100.0		69	100.0	
Survival status	alive	4	17.4	17.4	21	45.6	45.6	25	36.2	36.2
	dead	19	82.6	100.0	25	54.4	100.0	44	63.8	100.0
$p = 0.033$	all	23	100.0		46	100.0		69	100.0	
HLA A2 score	0	0	0.0	0.0	35	76.1	76.1	35	50.7	50.7
	1	0	0.0	0.0	8	17.4	93.5	8	11.6	62.3
	missing	23	100.0	100.0	3	6.5	100.0	26	37.7	100.0
$p = 5e-04$	all	23	100.0		46	100.0		69	100.0	



What about those fancy, publication-ready, tables?

tables

Computes and displays complex tables of summary statistics

Duncan Murdoch
University of Western
Ontario

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
<i>Iris setosa</i>				
mean	5.0060	3.4280	1.4620	0.2460
sd	0.3525	0.3791	0.1737	0.1054
<i>Iris versicolor</i>				
mean	5.9360	2.7700	4.2600	1.3260
sd	0.5162	0.3138	0.4699	0.1978
<i>Iris virginica</i>				
mean	6.5880	2.9740	5.5520	2.0260
sd	0.6359	0.3225	0.5519	0.2747

		X	
	n	mean	sd
X > 0	5	0.43369	0.3496
X < 0	5	-0.46960	0.2761
All	10	-0.01796	0.5611

Species	n	Sepal.Length		Sepal.Width	
		mean	sd	mean	sd
setosa	50	5.01	0.35	3.43	0.38
versicolor	50	5.94	0.52	2.77	0.31
virginica	50	6.59	0.64	2.97	0.32

Overall, we see the following:

All	150	5.84	0.83	3.06	0.44
-----	-----	------	------	------	------

Species	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
setosa	5.01 ± 0.05	3.43 ± 0.05	1.46 ± 0.02	0.25 ± 0.01
versicolor	5.94 ± 0.07	2.77 ± 0.04	4.26 ± 0.07	1.33 ± 0.03
virginica	6.59 ± 0.09	2.97 ± 0.05	5.55 ± 0.08	2.03 ± 0.04
All	5.84 ± 0.07	3.06 ± 0.04	3.76 ± 0.14	1.20 ± 0.06

What about those fancy, publication-ready, tables?

xtables

Coerce data to LaTeX and HTML tables

David B. Dahl
Brigham Young University

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1954							1	0	-1	-1	-1	0
1955	2	4	6	7	8	9	10	12	13	13	12	15
1956	15	16	18	20	21	20	21	21	21	24	26	27
1957	30	32	32	34	36	36	37	37	40	39	40	39
1958	40	39	40	42	42	45	45	46	49	51	53	56
1959	57	60	61	60	62	64	64	66	69	68	70	70
1960	71	72	75	75	75	76	78	79	79	80	80	81
1961	82	83	84	86	86	87	87	87	88	89	92	91
1962	92	93	92	92	92	92	94	98	99	99		

	PC1	PC2	PC3	PC4
Standard deviation	83.7324	14.2124	6.4894	2.4828
Proportion of Variance	0.9655	0.0278	0.0058	0.0008
Cumulative Proportion	0.9655	0.9933	0.9991	1.0000

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.0445	0.1709	17.81	0.0000
outcome2	-0.4543	0.2022	-2.25	0.0246
outcome3	-0.2930	0.1927	-1.52	0.1285
treatment2	0.0000	0.2000	0.00	1.0000
treatment3	0.0000	0.2000	0.00	1.0000

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
sex	1	75.37	75.37	0.38	0.5417
ethnicity	3	2572.15	857.38	4.27	0.0072
grade	1	36.31	36.31	0.18	0.6717
disadv	1	59.30	59.30	0.30	0.5882
Residuals	93	18682.87	200.89		

	R^2	R^2	F-stat	S.E.E	DW
y_{t-1}	0.90	0.89	200.00	0.04	2.00



13 reasons why **you will** love GNU R

- I R is (extremely) cheap. In fact - it's free :)
- II R has (extremely) wide range of capabilities
- III R is (widely) supported by the world of science
- IV R is supported by the community
- V R is (increasingly) supported by the business
- VI R is able to read data in many formats
- VII Interoperability is easy to achieve
- VIII R is truly cross-platform
- IX ½ :)** **Graphics** – all have waited for this moment :)
- X There are many possibilities to optimize the code
- XI R is able to handle large amount of data
- XII R has a set of fancy tools and IDEs
- XII FDA accepted using R for drug trials!



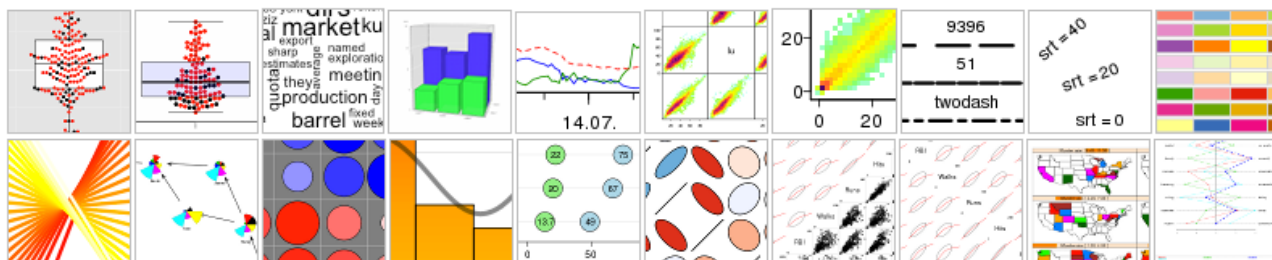
a picture is woRth a thousand words

It is really hard to describe all the potential of R in creating graphics. Internet is full of of examples and tutorials of how to create complex and sophisticated charts in R. There are thousands of examples.

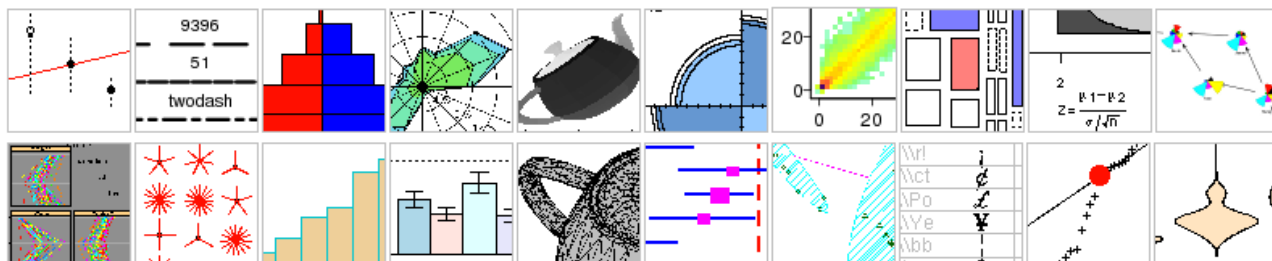
The user has at his disposal a number of libraries allowing creation of almost every chart or graphics he needs. R is able to produce clean, ascetic charts suitable for journals and books, and super fancy graphics, perfect for presentations.

**Imagination
is the only limit here**

» Last entries ...



» Random entries



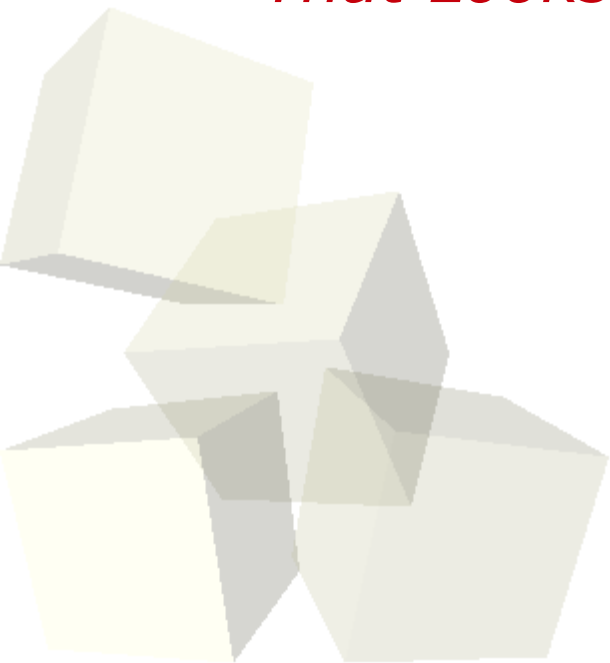
<http://rgraphgallery.blogspot.com/>



If anything, there should be a Law:

*Thou Shalt **Not Even Think** Of Producing A Graph
That Looks Like **Anything From A Spreadsheet***

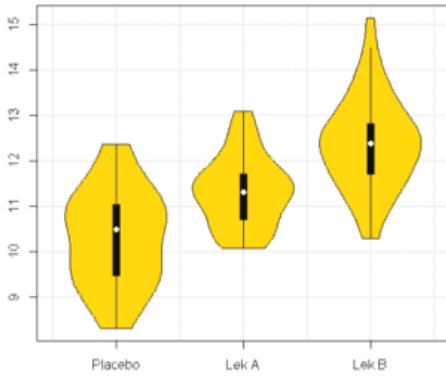
Ted Harding



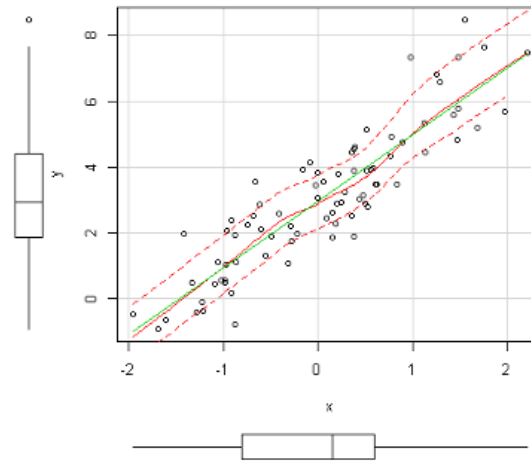


R has great charting capabilities

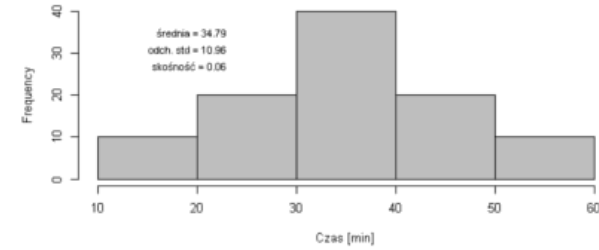
Średnie stężenia Hb w funkcji leku



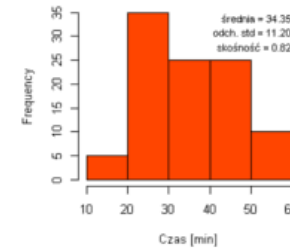
Kombinowany wykres X-Y



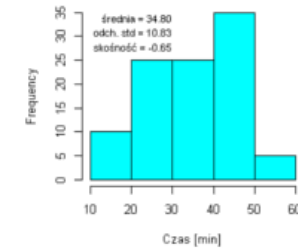
Czas reakcji na lek



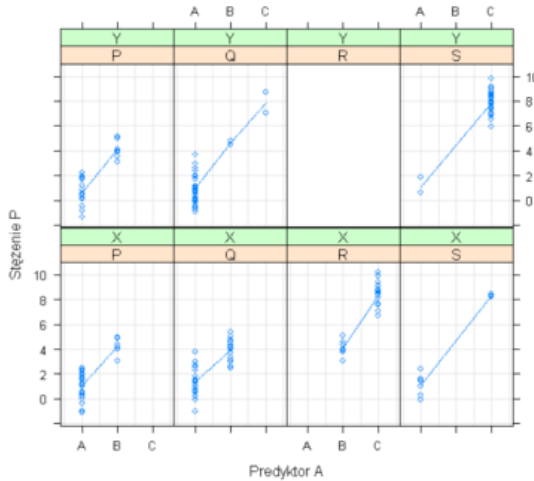
Czas reakcji na lek



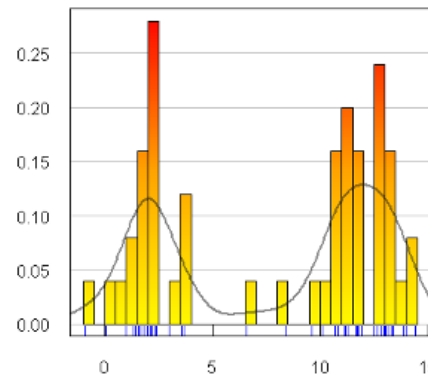
Czas reakcji na lek



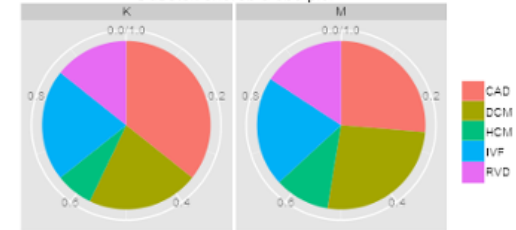
Parametr P w funkcji predyktorów jakościowych



Mieszanka dwóch rozkładów



Odsetek chorób u obu płci



Wpływ leku na stężenie Hb

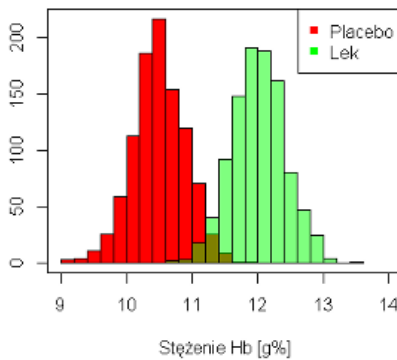


Diagram mozaikowy

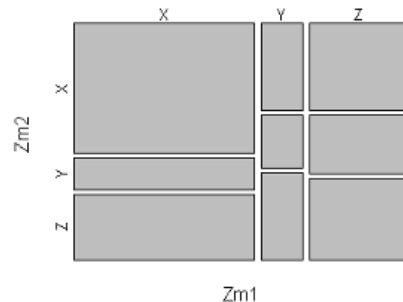
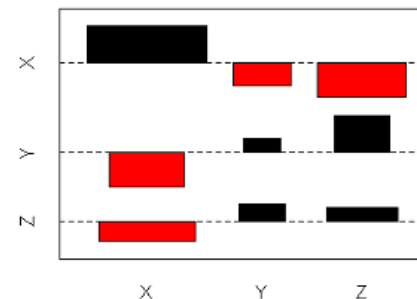
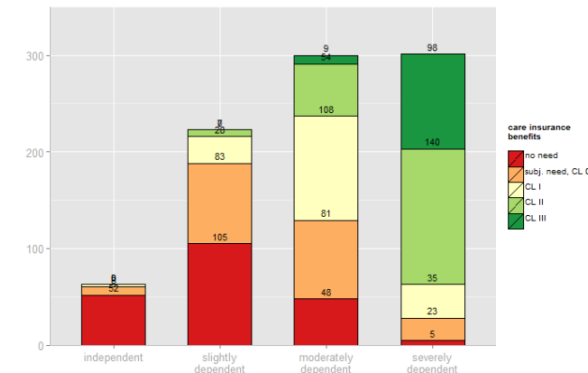


Diagram asocjacji

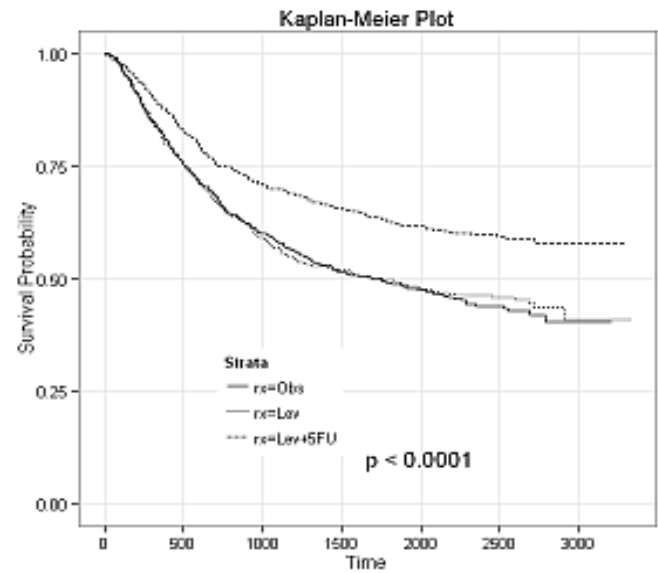
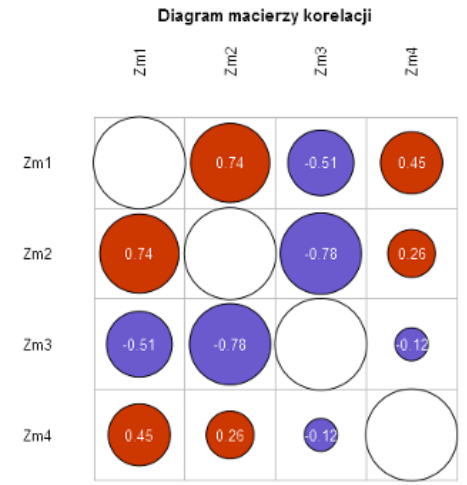
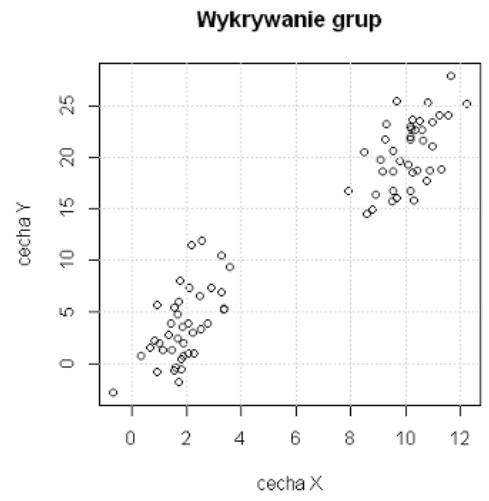
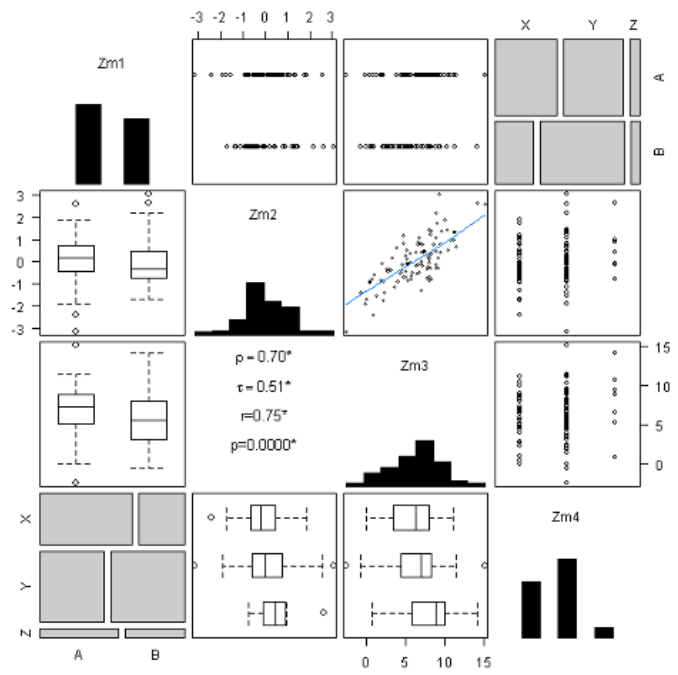


(subjective) dependency



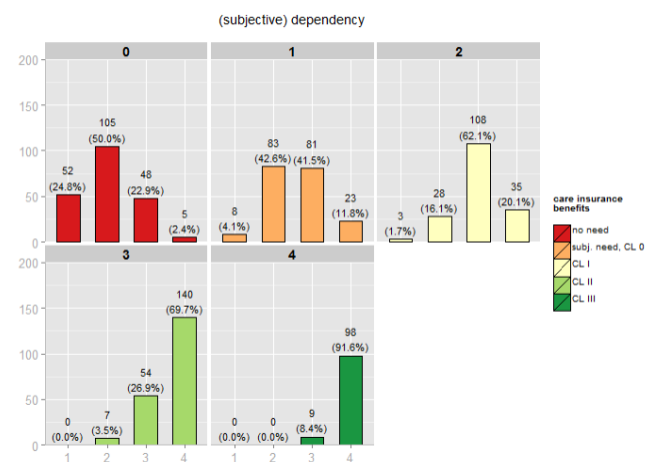
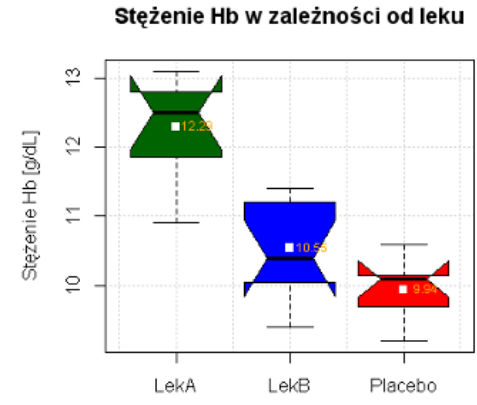
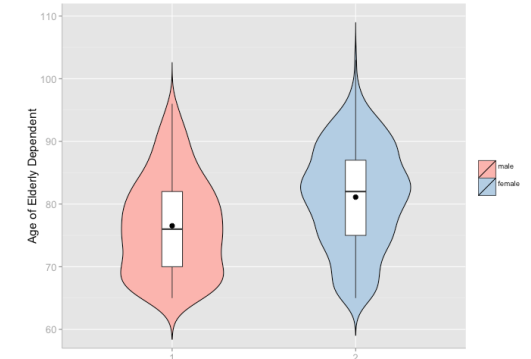
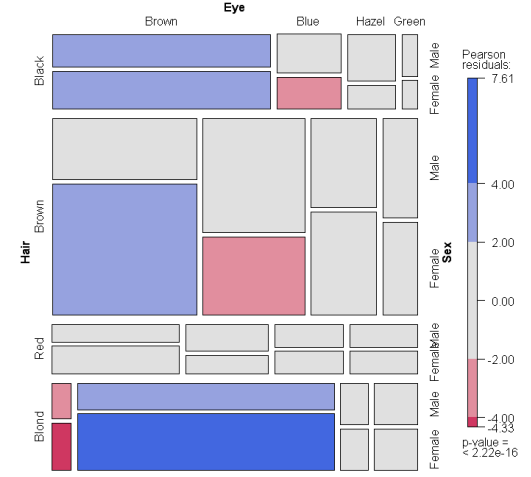


...of any kind



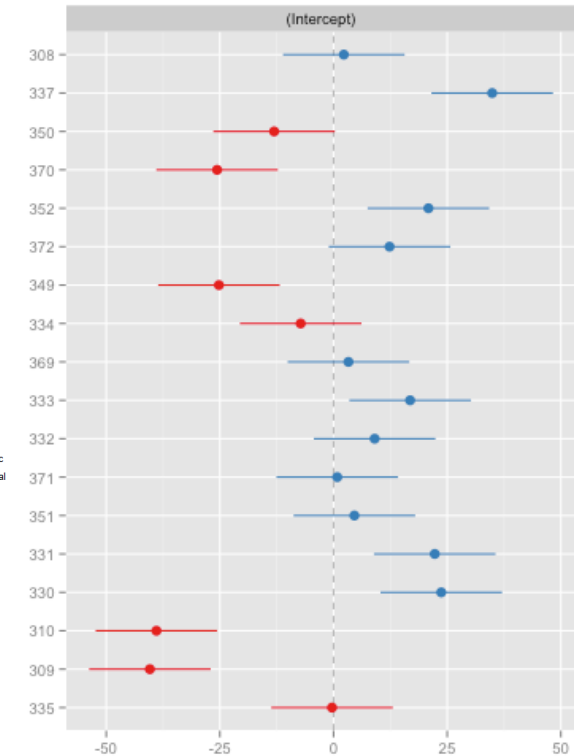
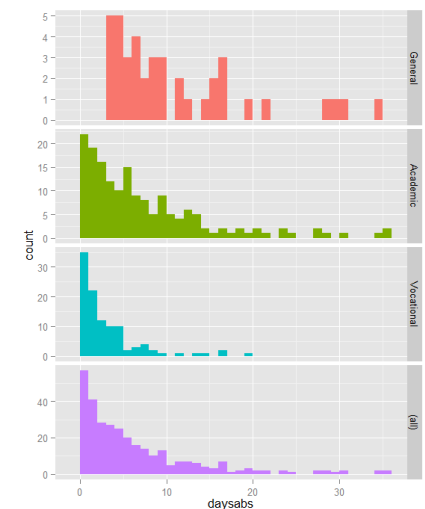
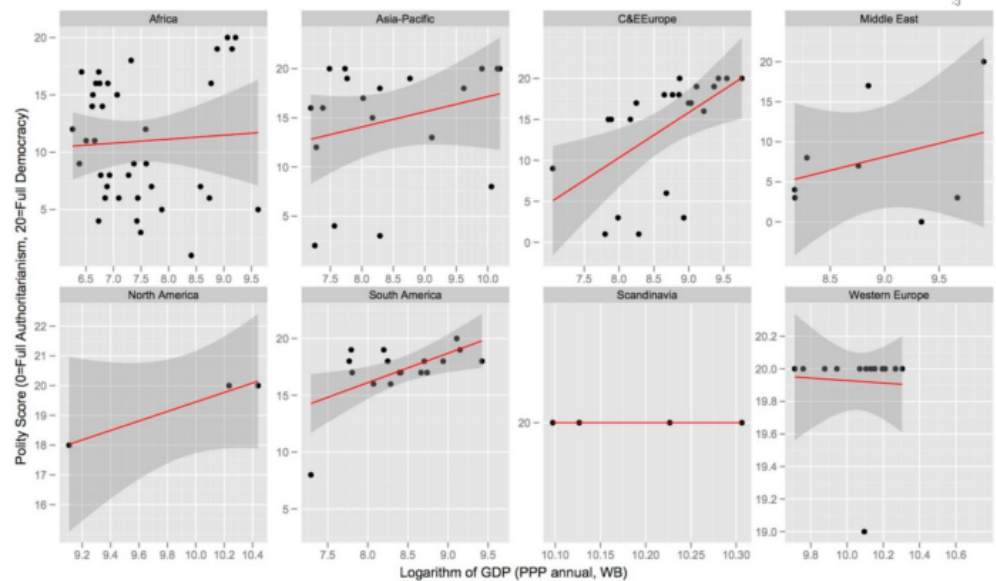
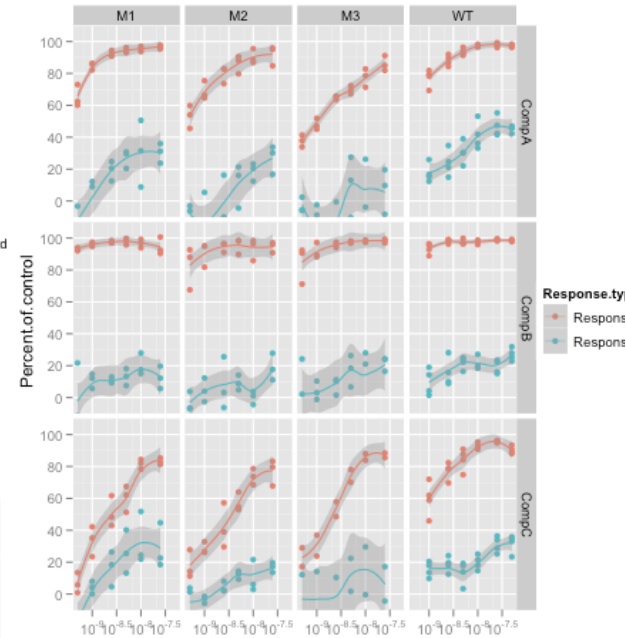
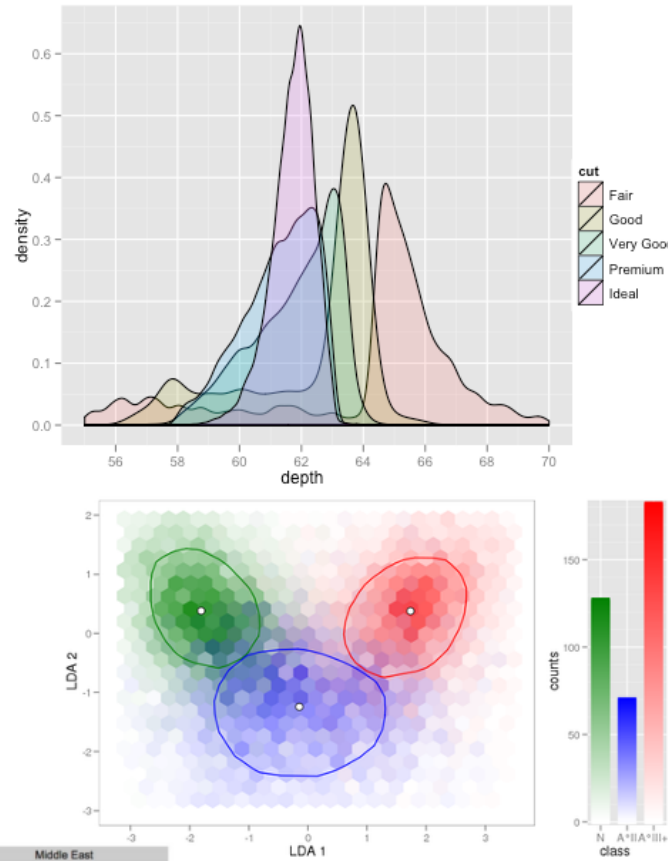
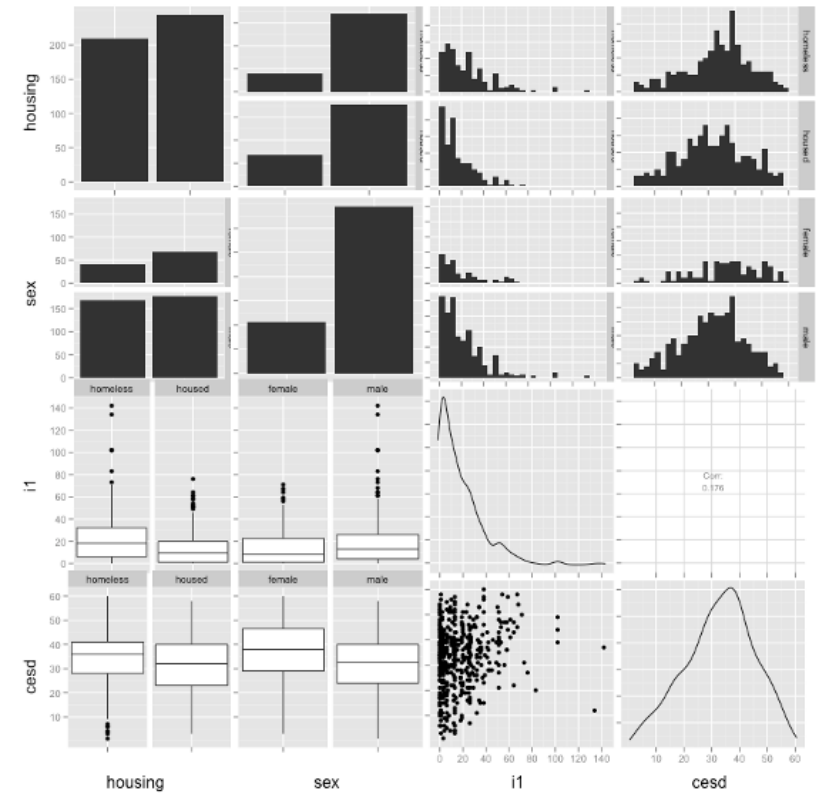
	0	500	1000	1500	2000	2500	3000
rx=0bs	630	470	372	315	256	90	11
rx=Lev	630	464	360	318	266	107	8
rx=Lev+5FU	608	498	425	387	328	127	14

Numbers at risk



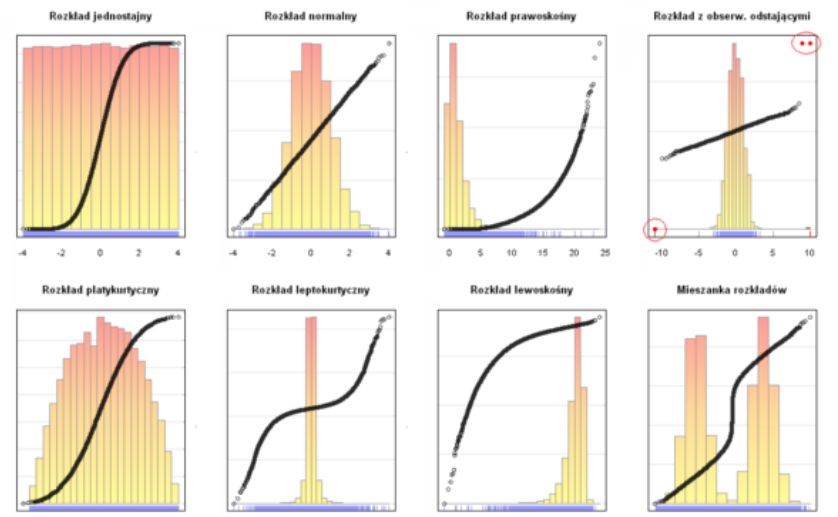
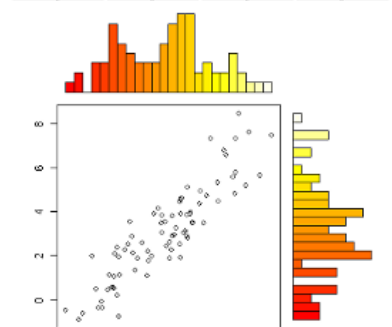
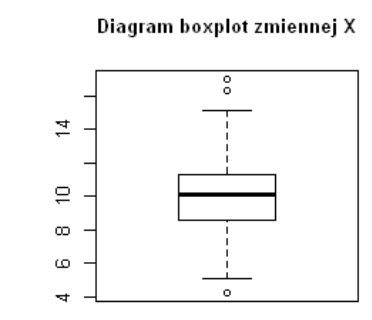
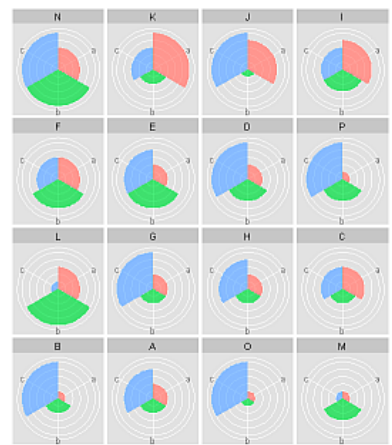
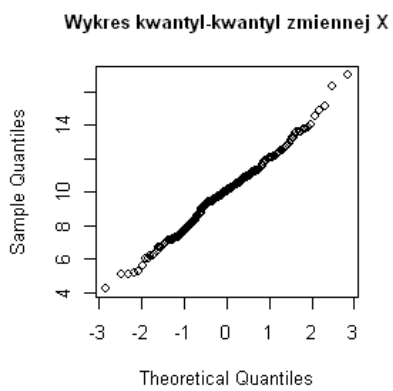
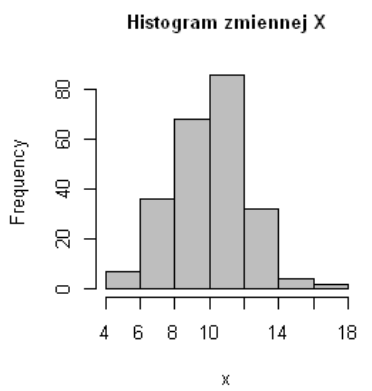
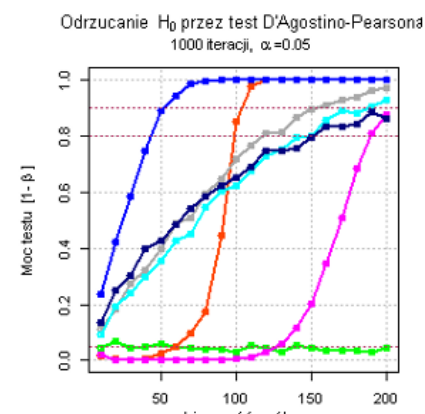
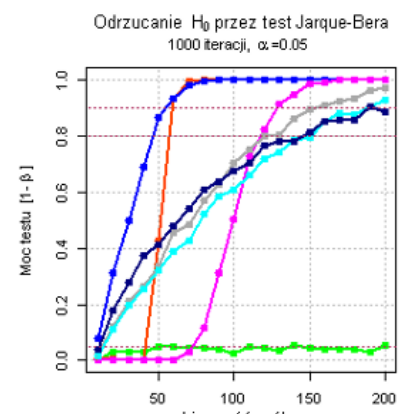
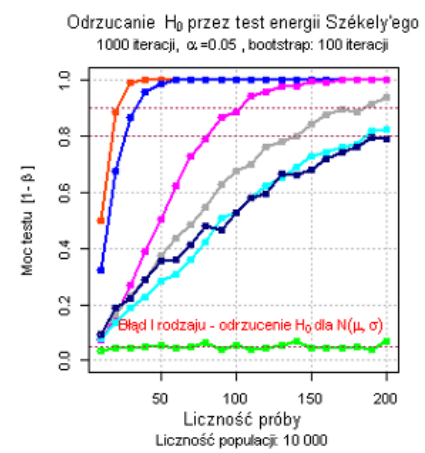
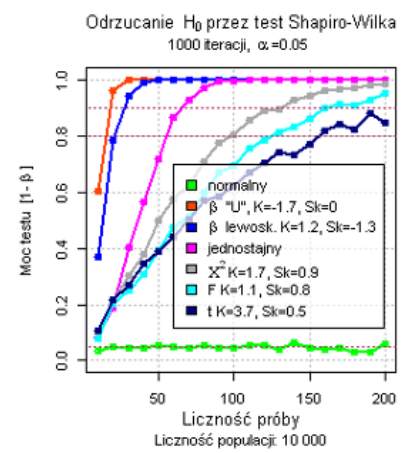
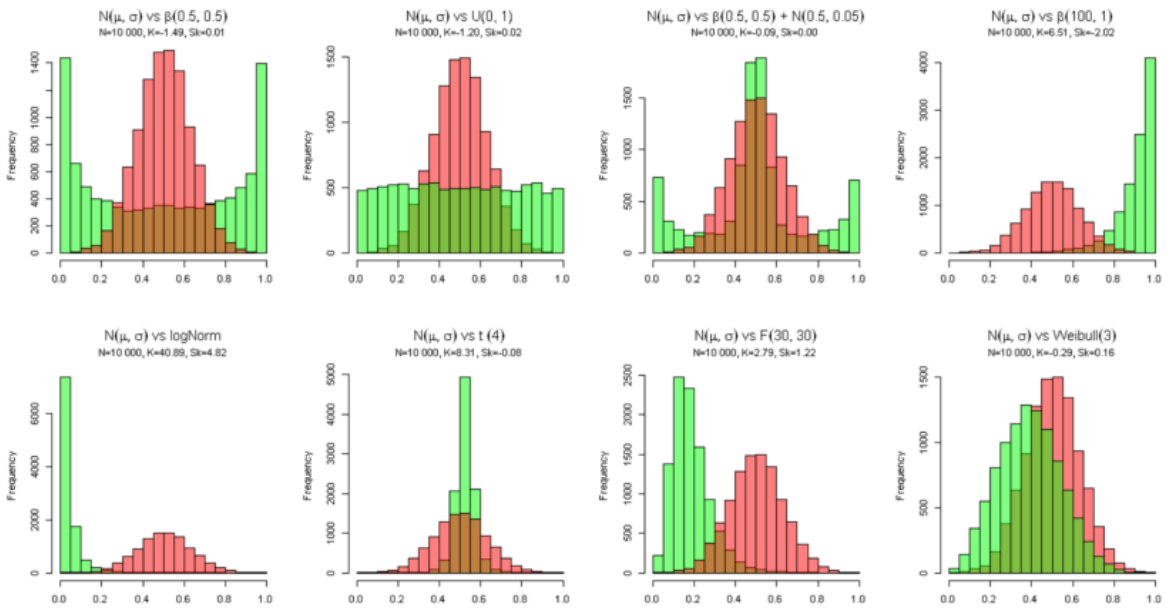


...and yes, these charts can be very elegant :)



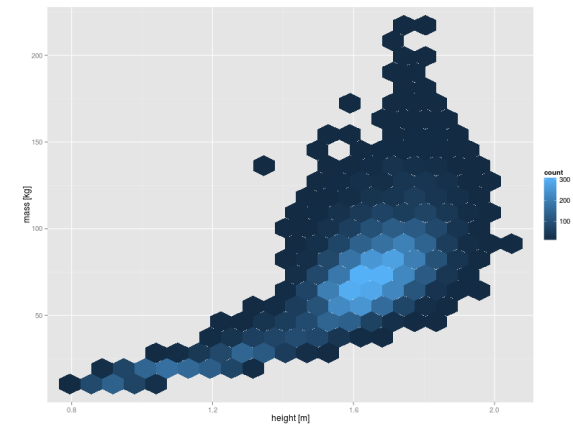
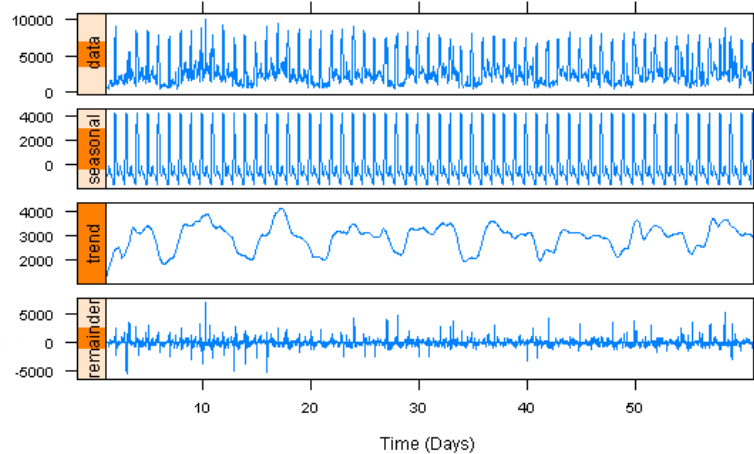
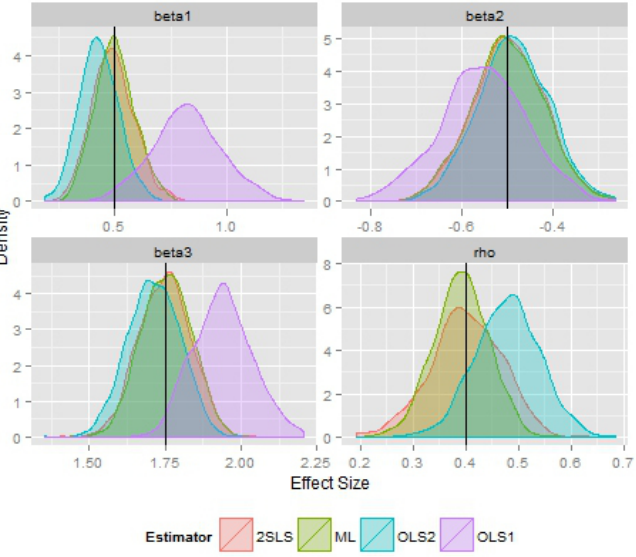
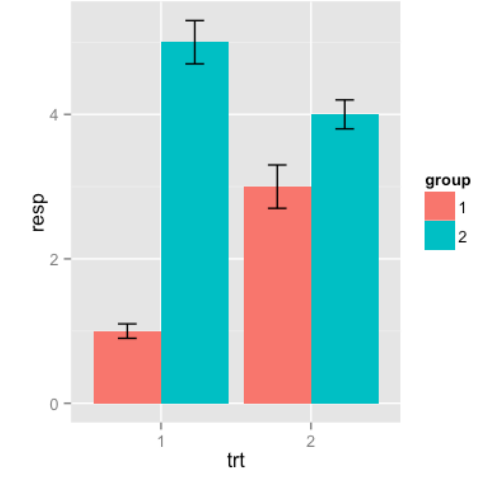
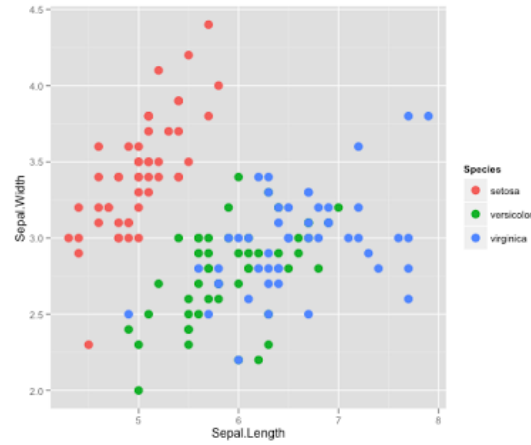
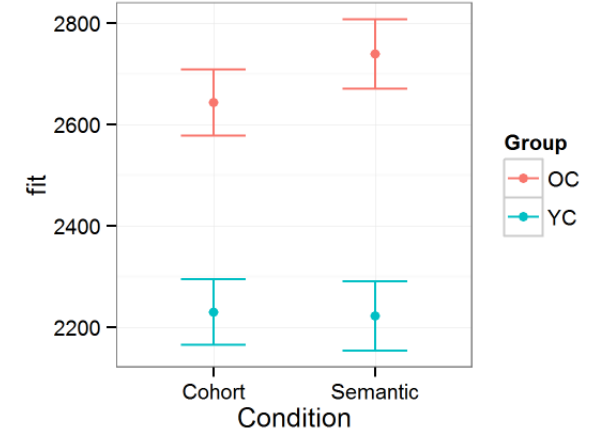
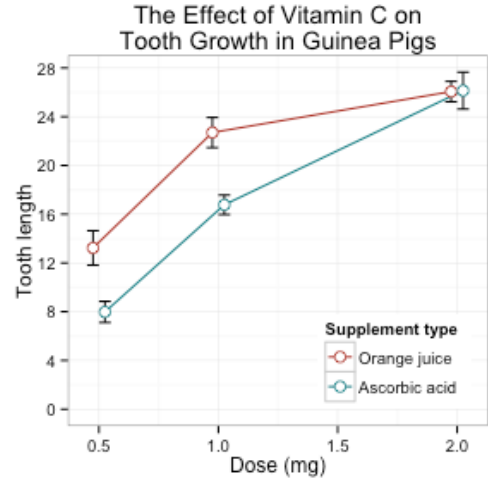
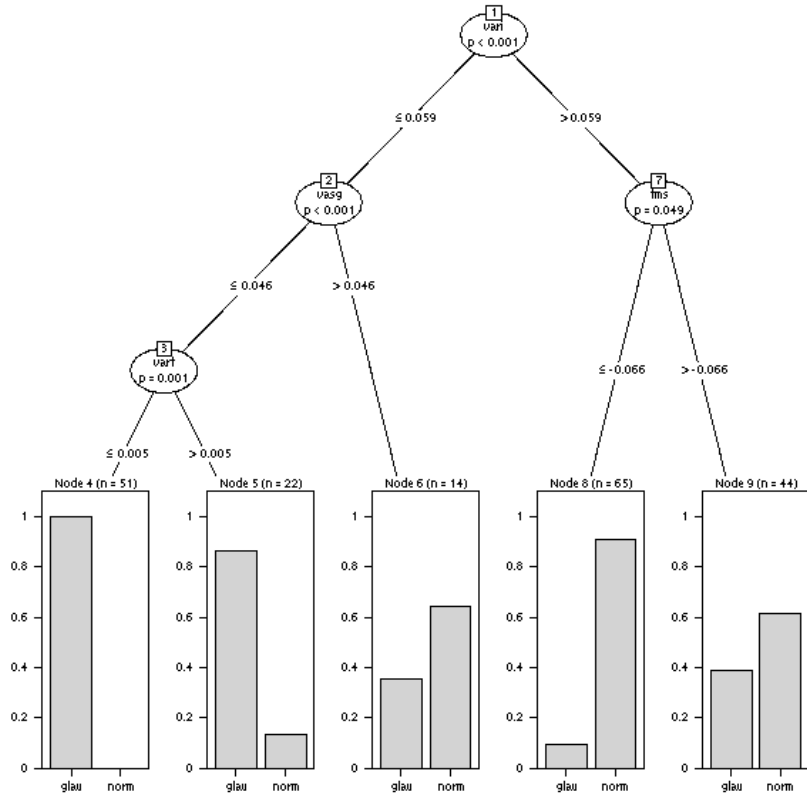


...and varied...



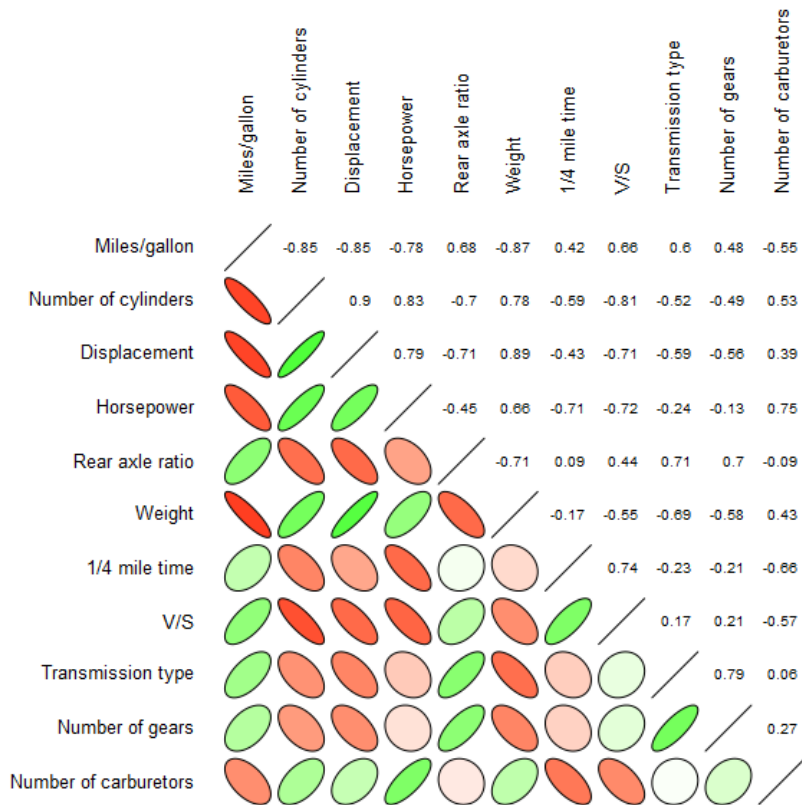
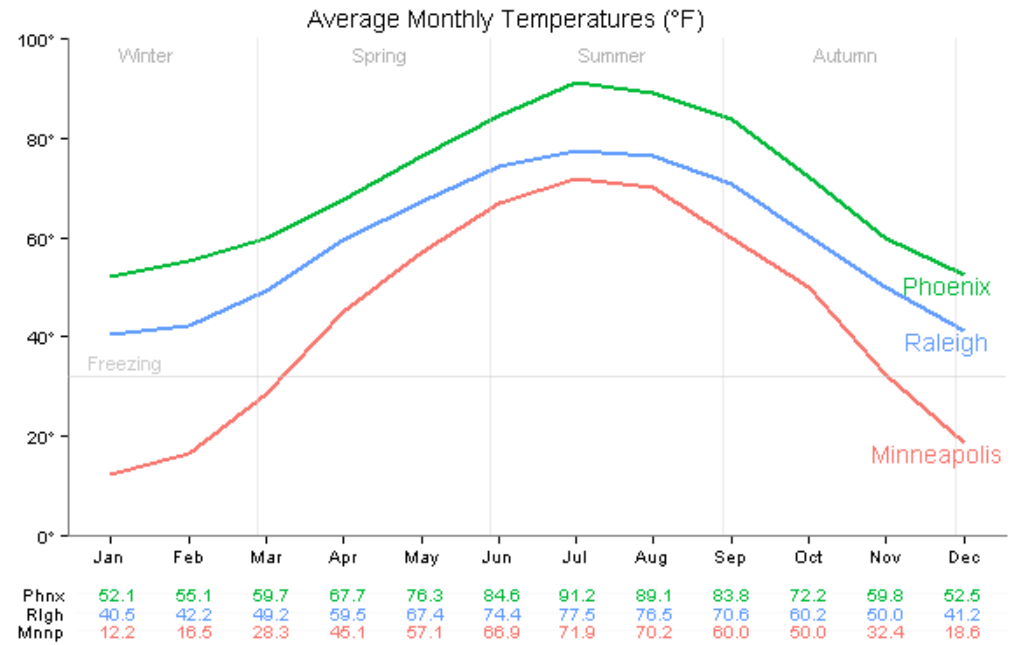
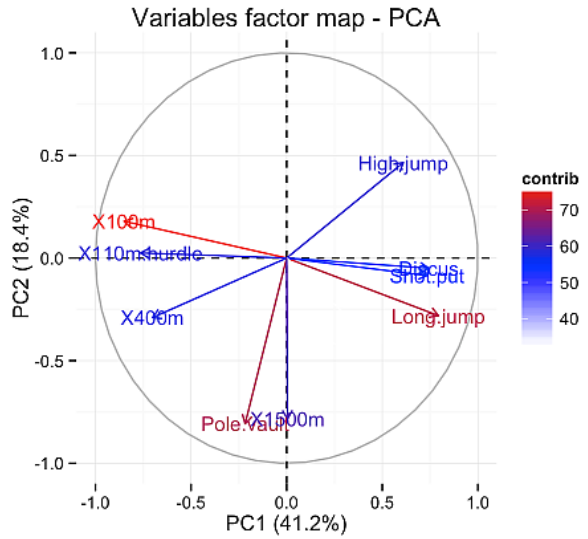


...and varied...

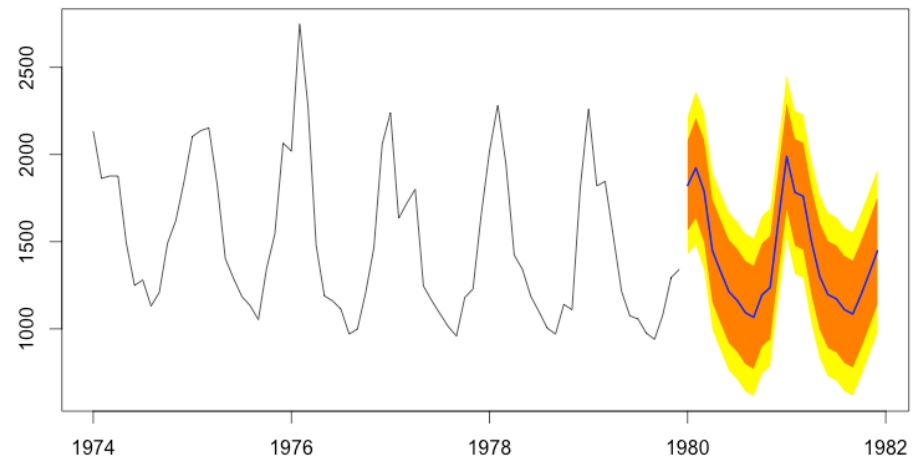




...and varied...



Forecast of Lung Disease Related Deaths in the UK

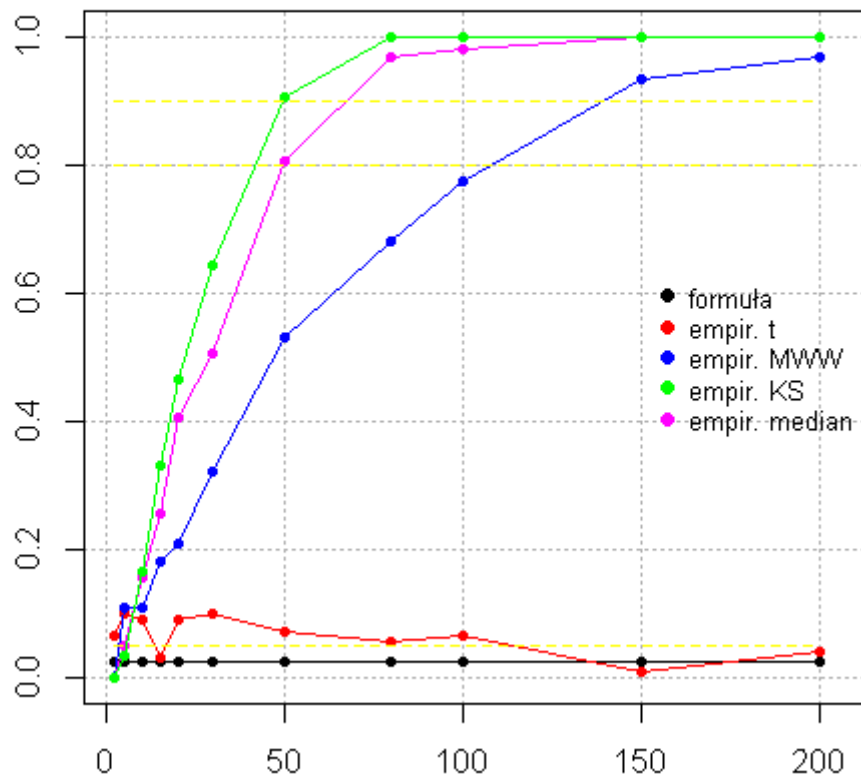




Power of tests

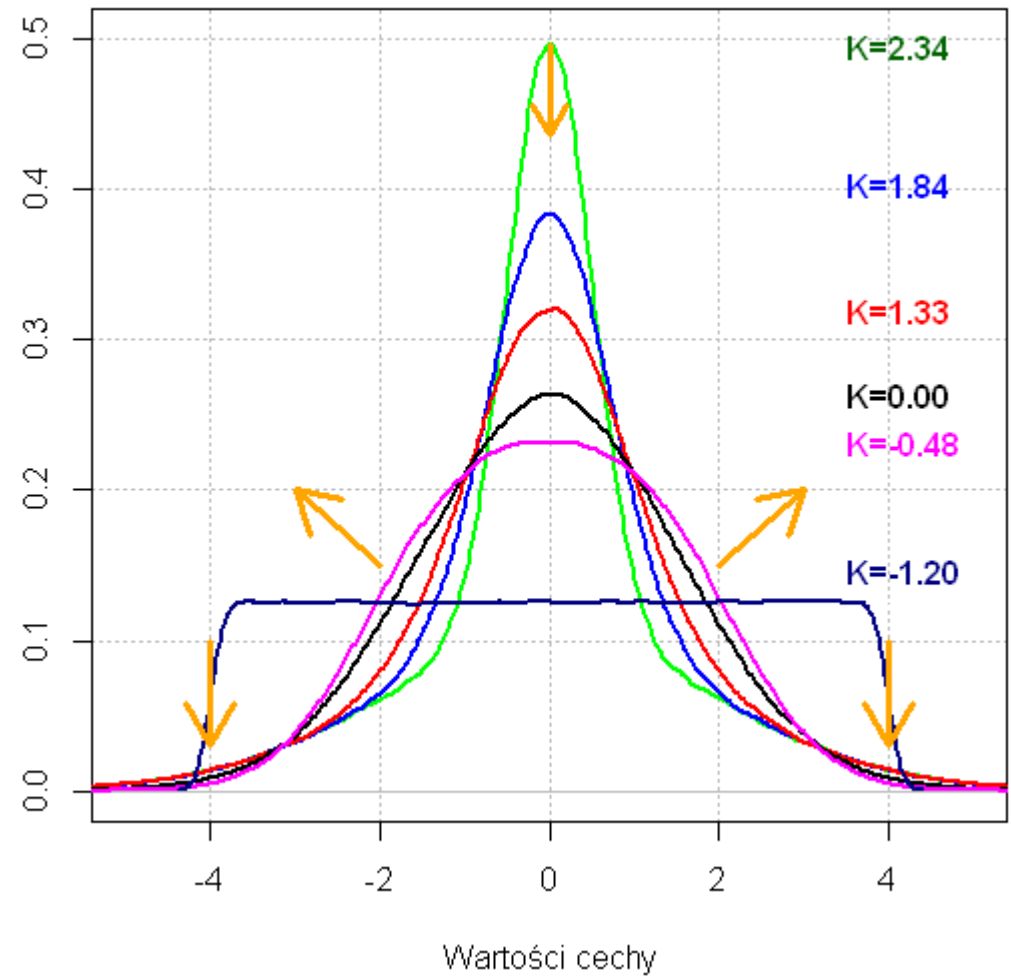
Moc testów t, MWW, KS i median

$\beta(100, 1)$ vs. logNorm | $\delta = 0.001$



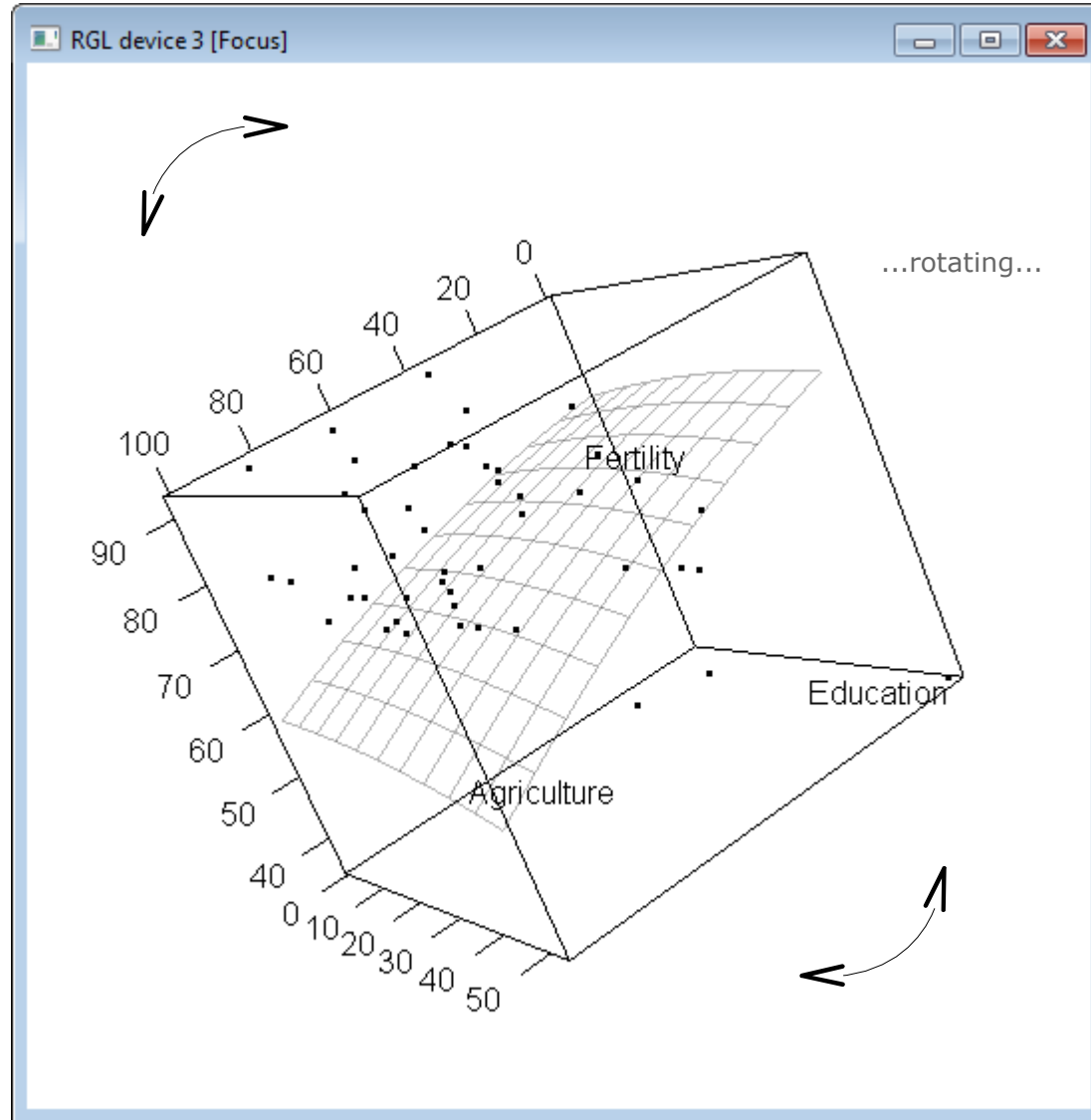
Decreasing kurtosis

Zmniejszająca się kurtoza



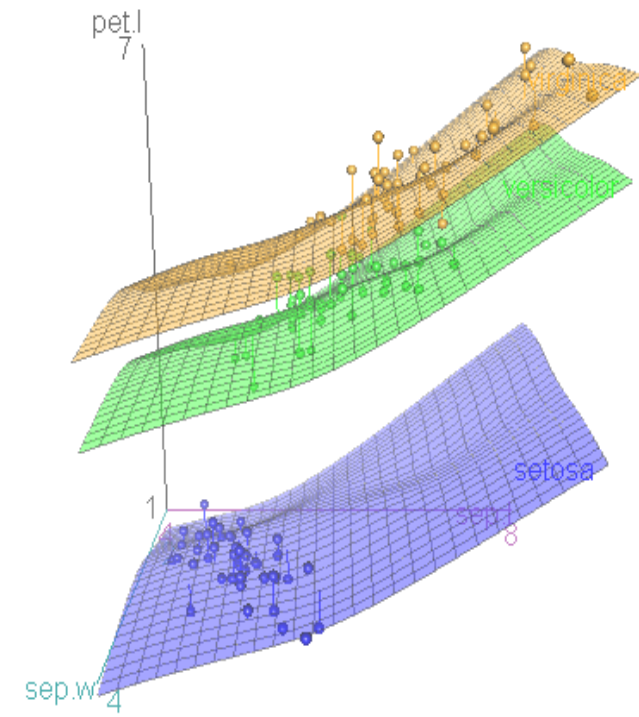
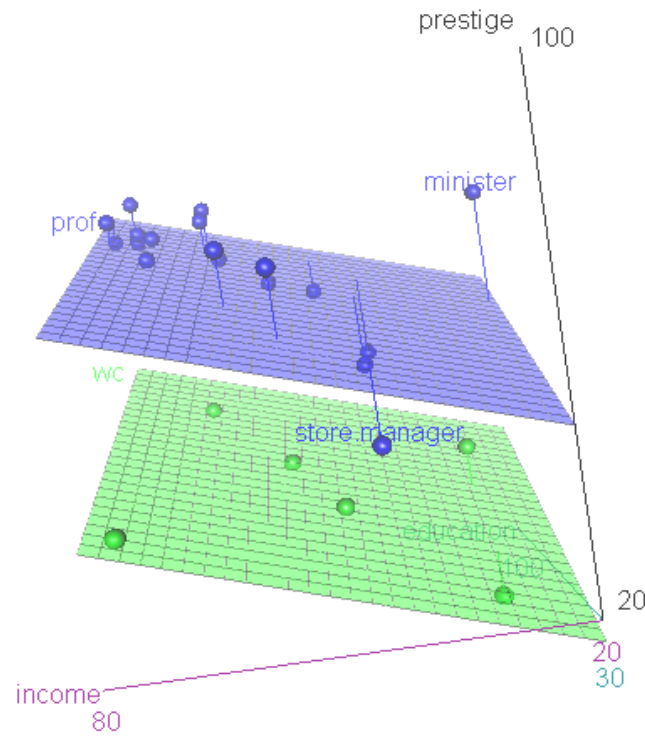
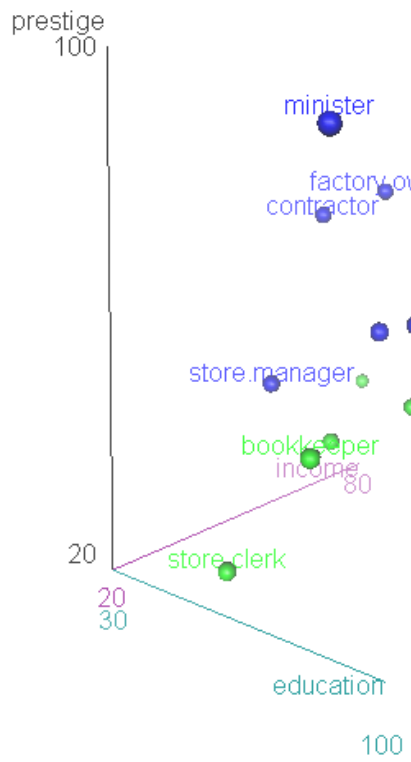


...and 3D, interactive and animated (rgl)



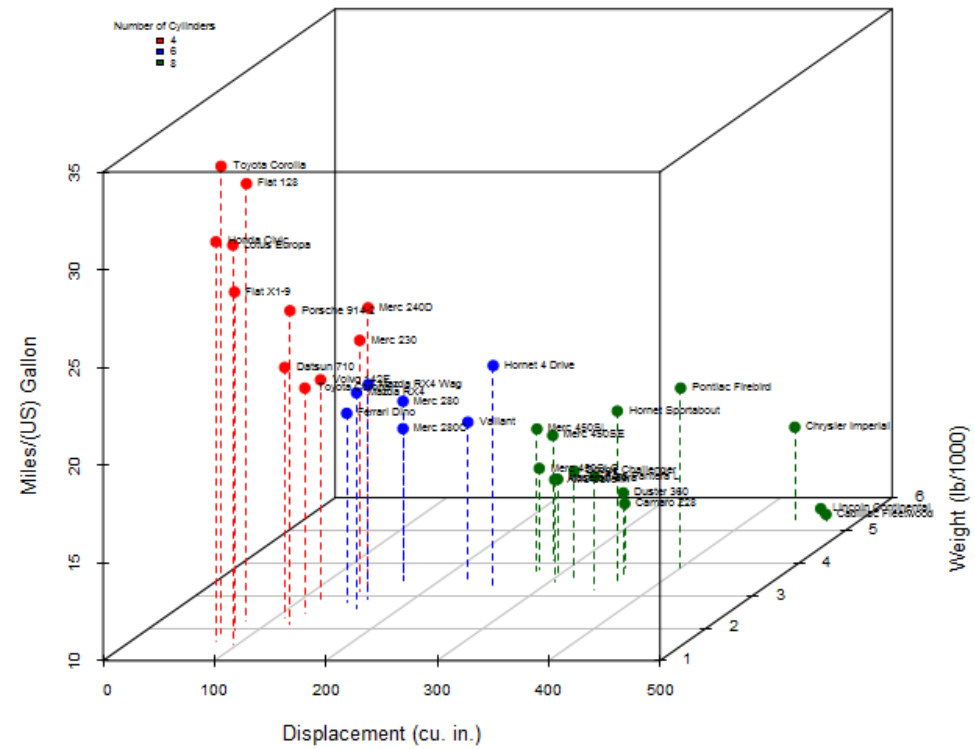
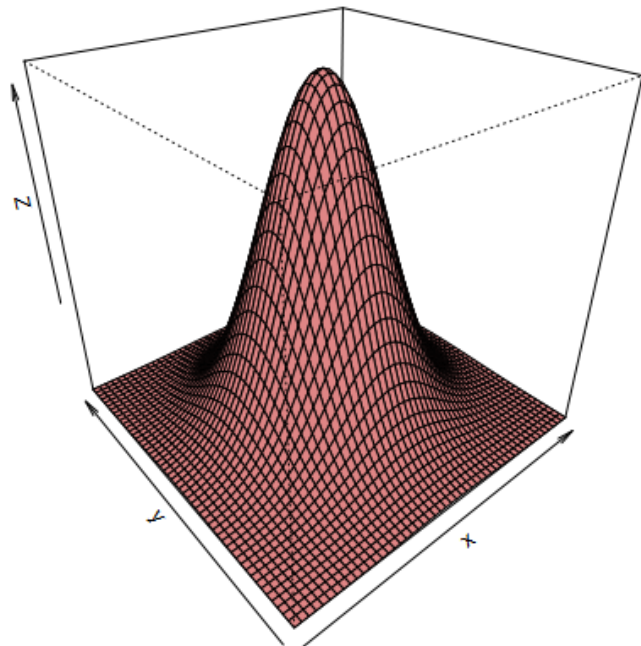
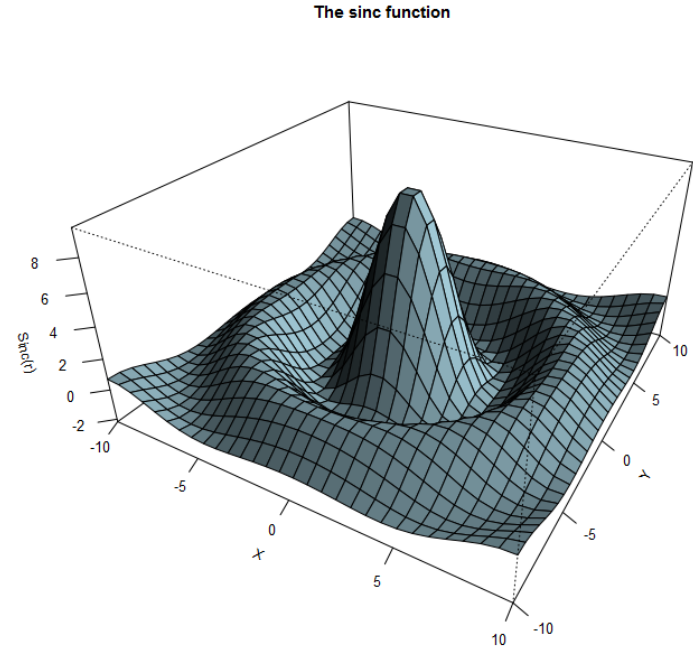
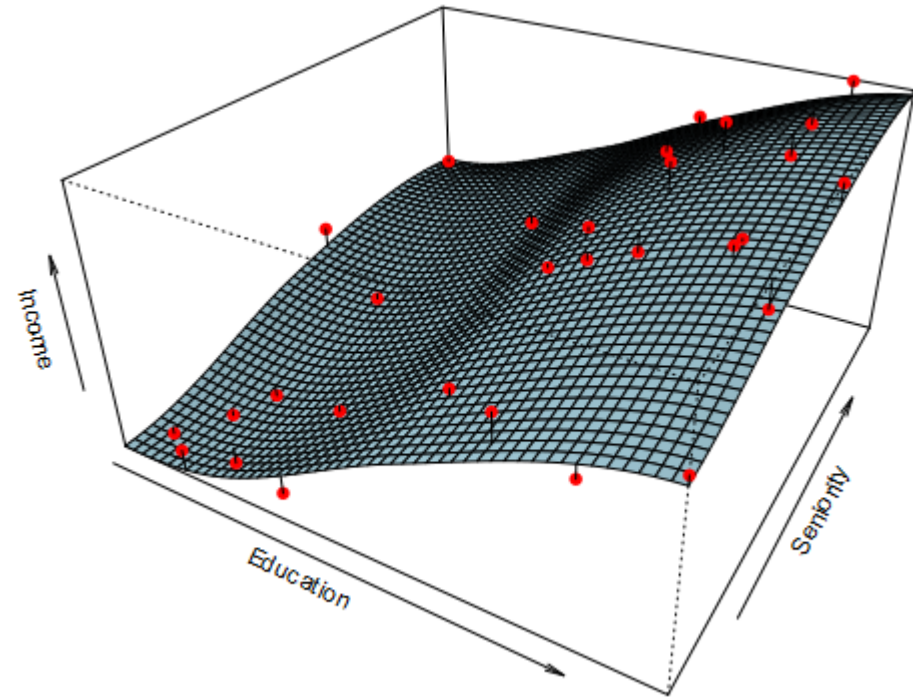


...and 3D, interactive and animated (rgl)





...and 3D interactive (scatterplot3d)



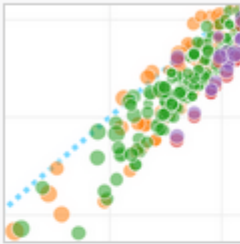


...and interactive (JScript)

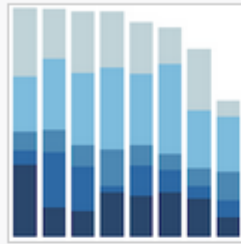
Documentation Examples

Chart Types

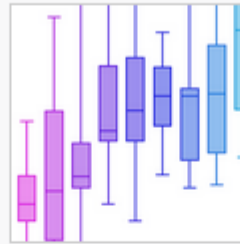
Line and Scatter Plots



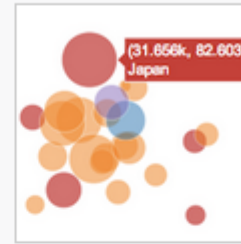
Bar Charts



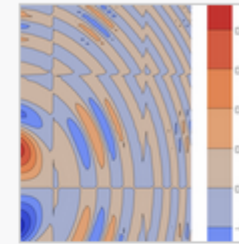
Box Plots



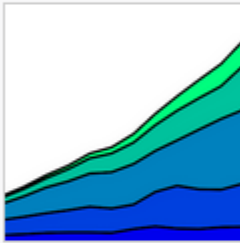
Bubble Charts



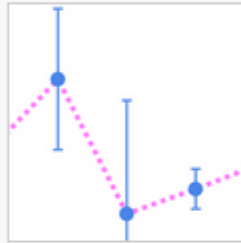
Contour Plots



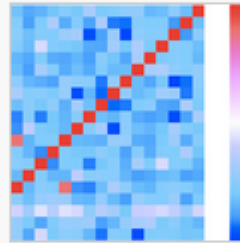
Filled Area Plots



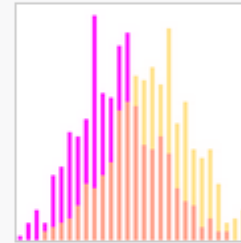
Error Bars



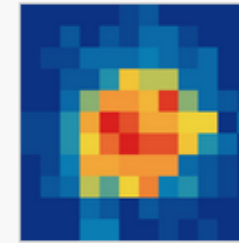
Heatmaps



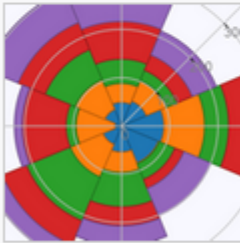
Histograms



2D Histograms



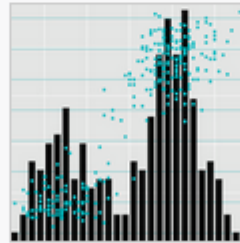
Polar Charts



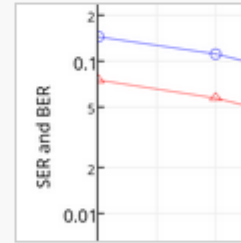
Time Series



Multiple Chart Types



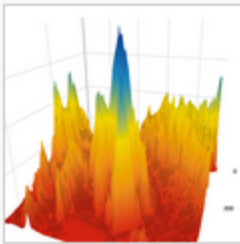
Log Plots



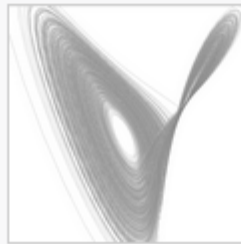
3D Scatter Plots



3D Surface Plots

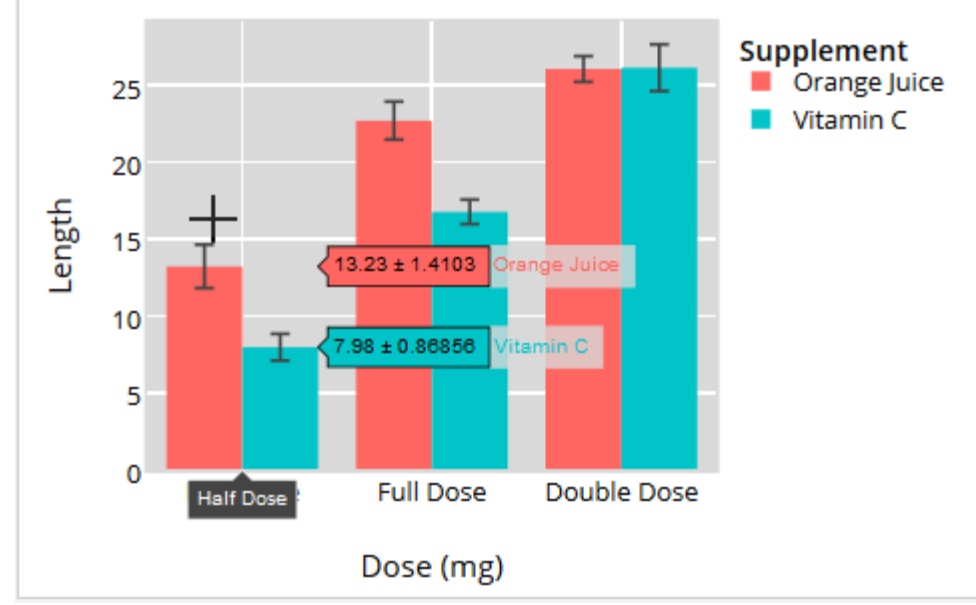
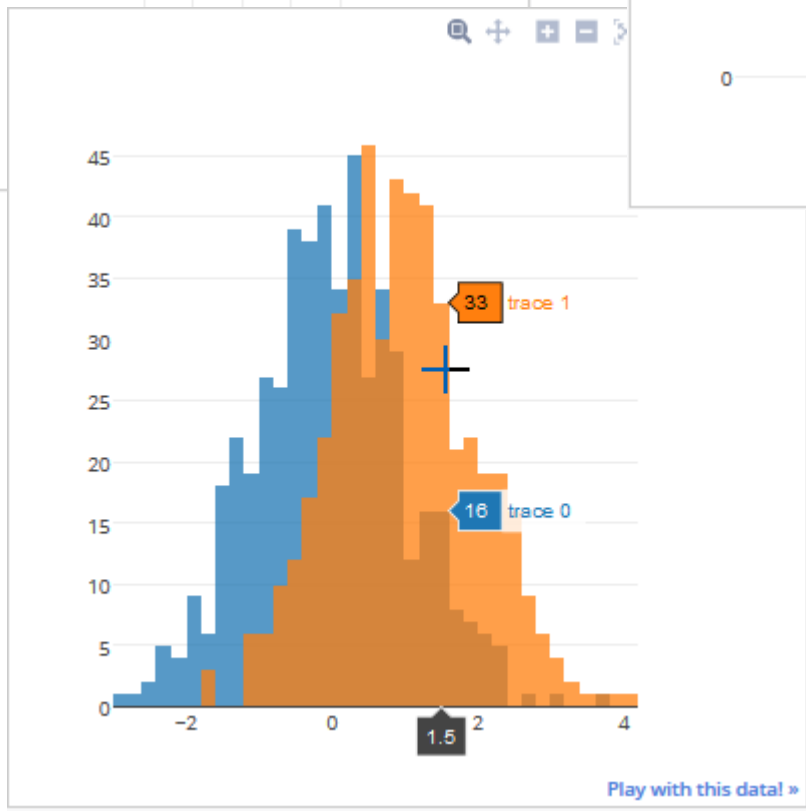
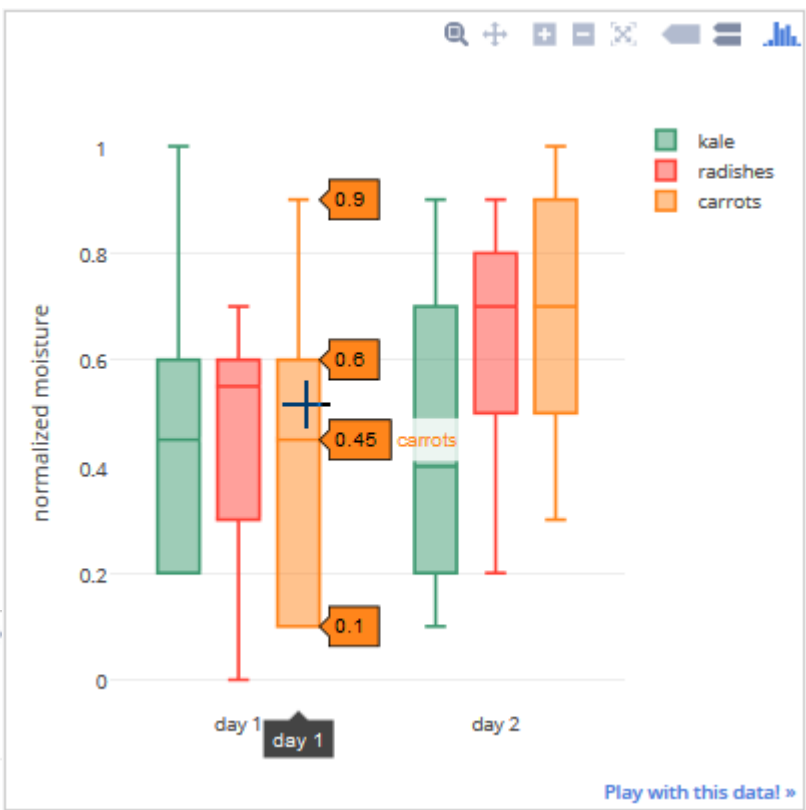
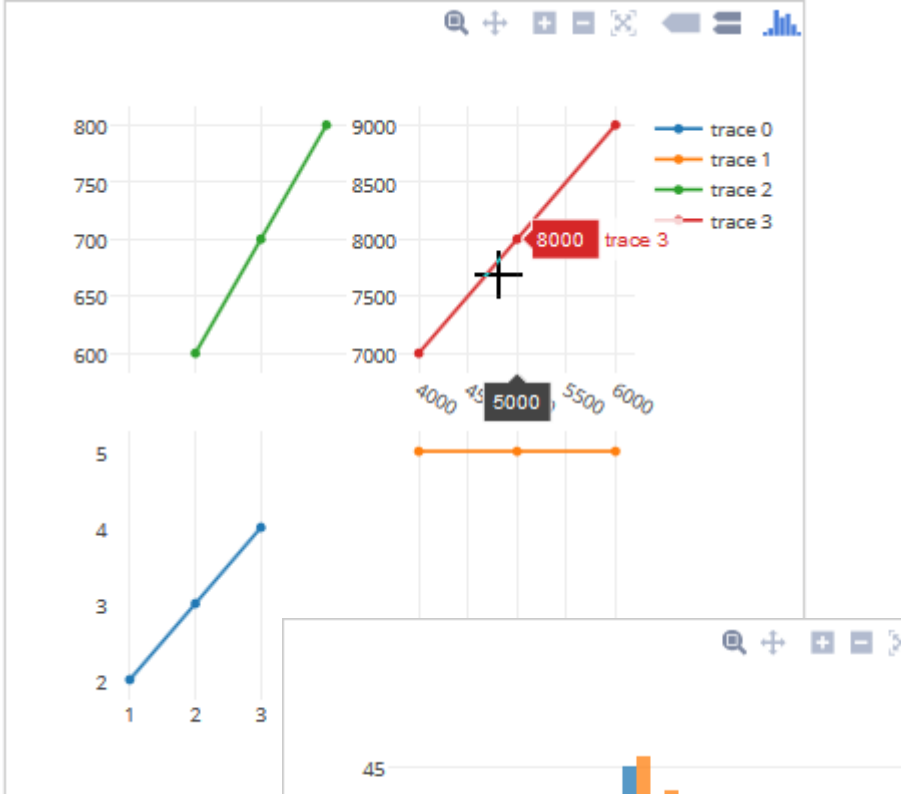


3D Line Plots



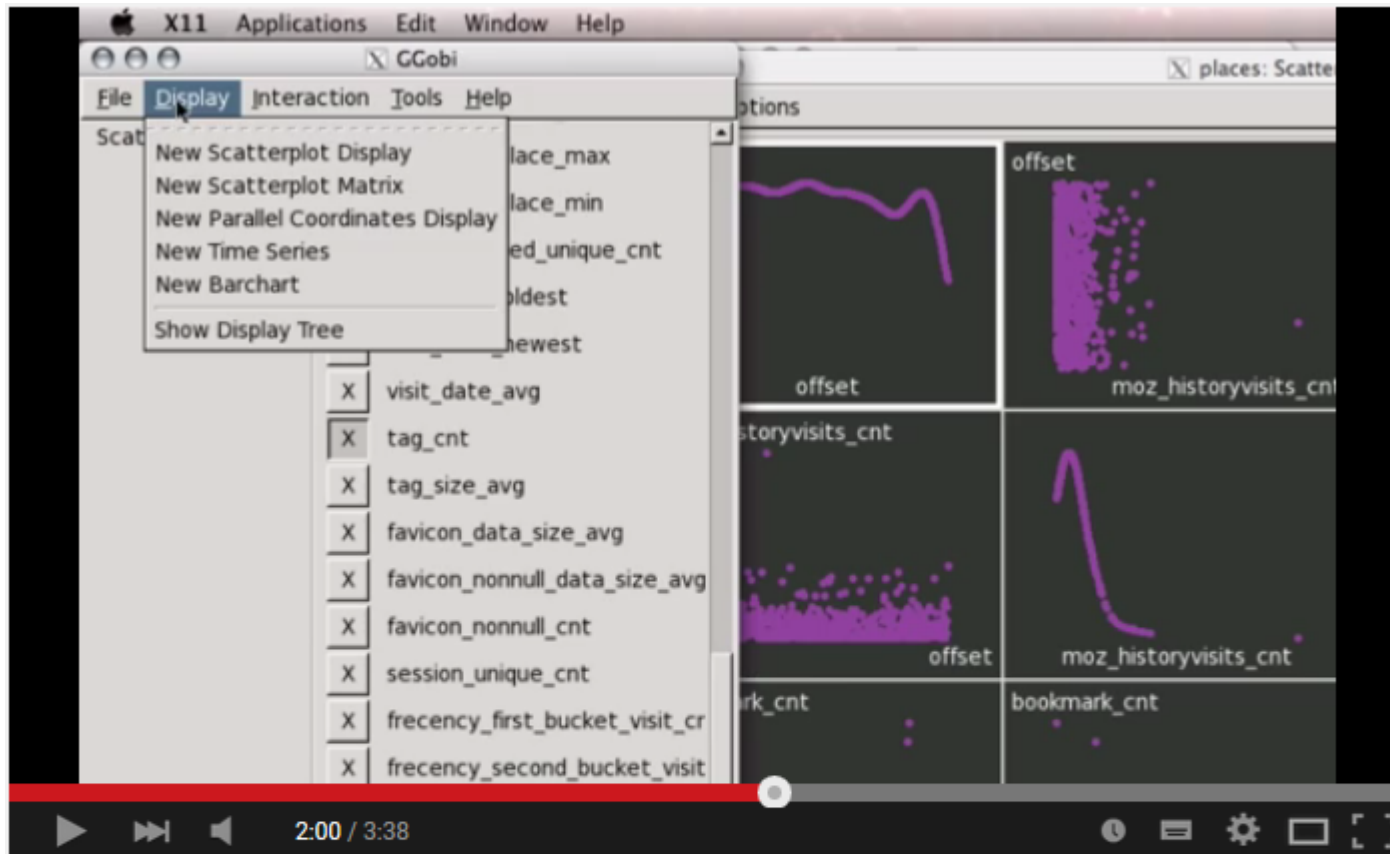


...and interactive (JScript)





[YouTube: Using R & GGobi to Create a Scatterplot Matrix](#)



Using R & GGobi to Create a Scatterplot Matrix



Andy Edmonds

Subskrybuj 39

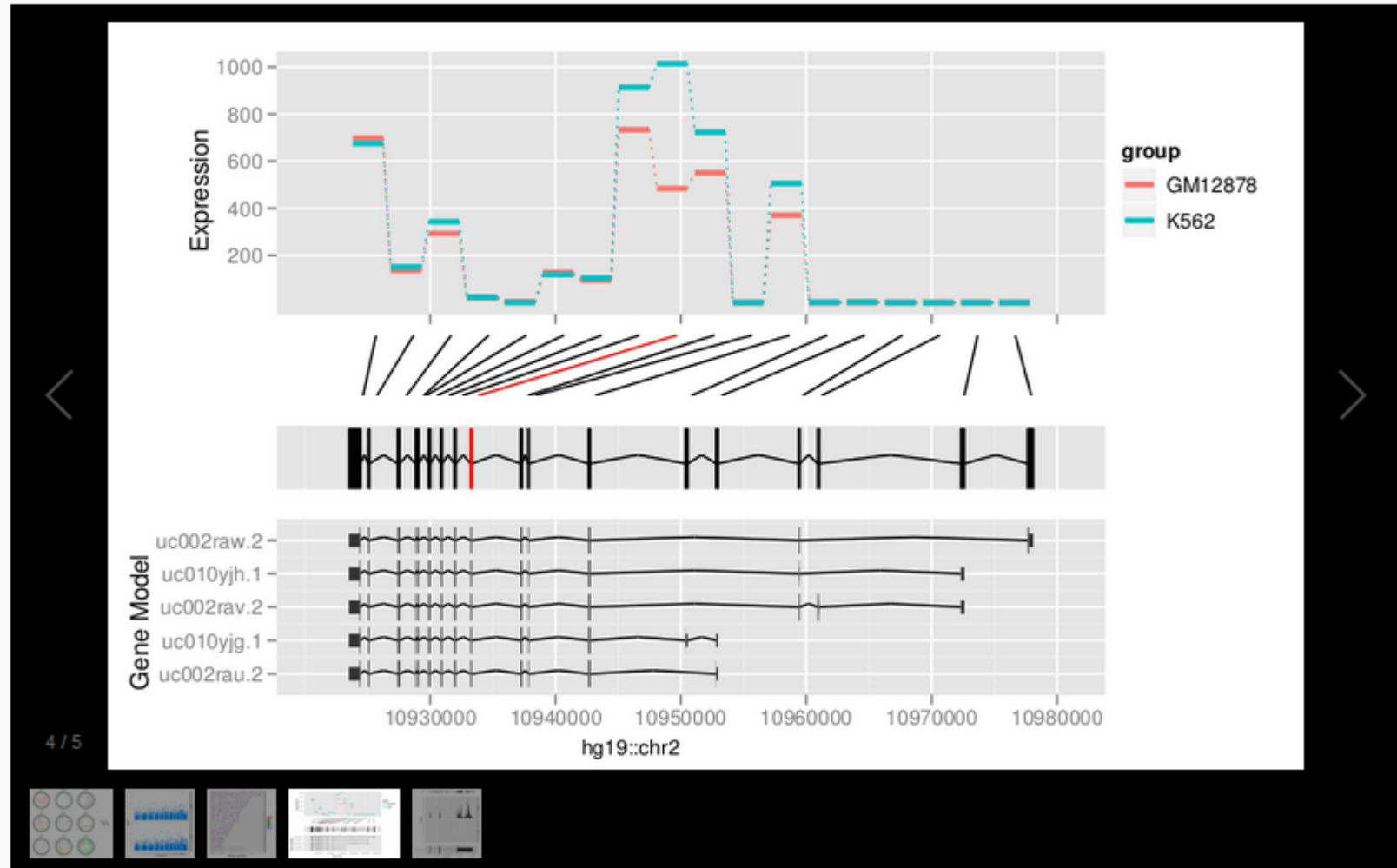
8 378

+ Dodaj do Udostępnij Więcej

17 0



ggbio is released with Bioconductor 2.11 now.



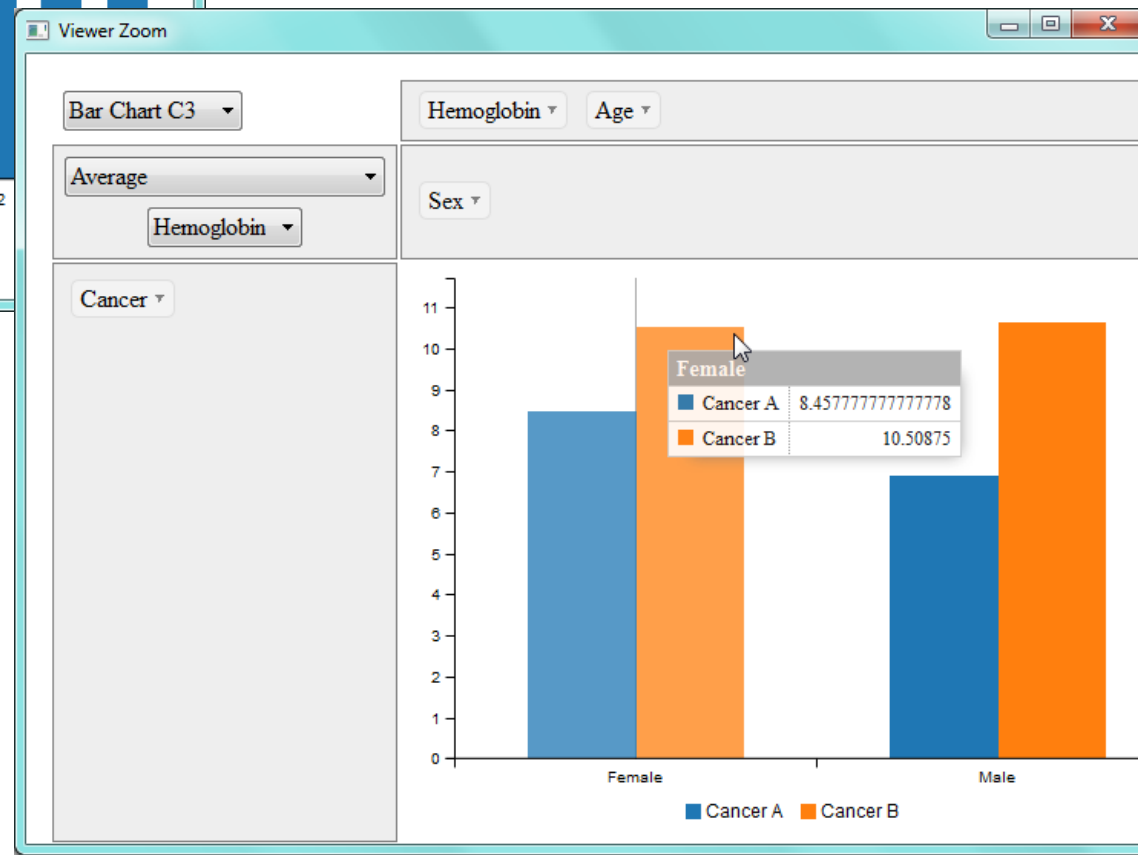
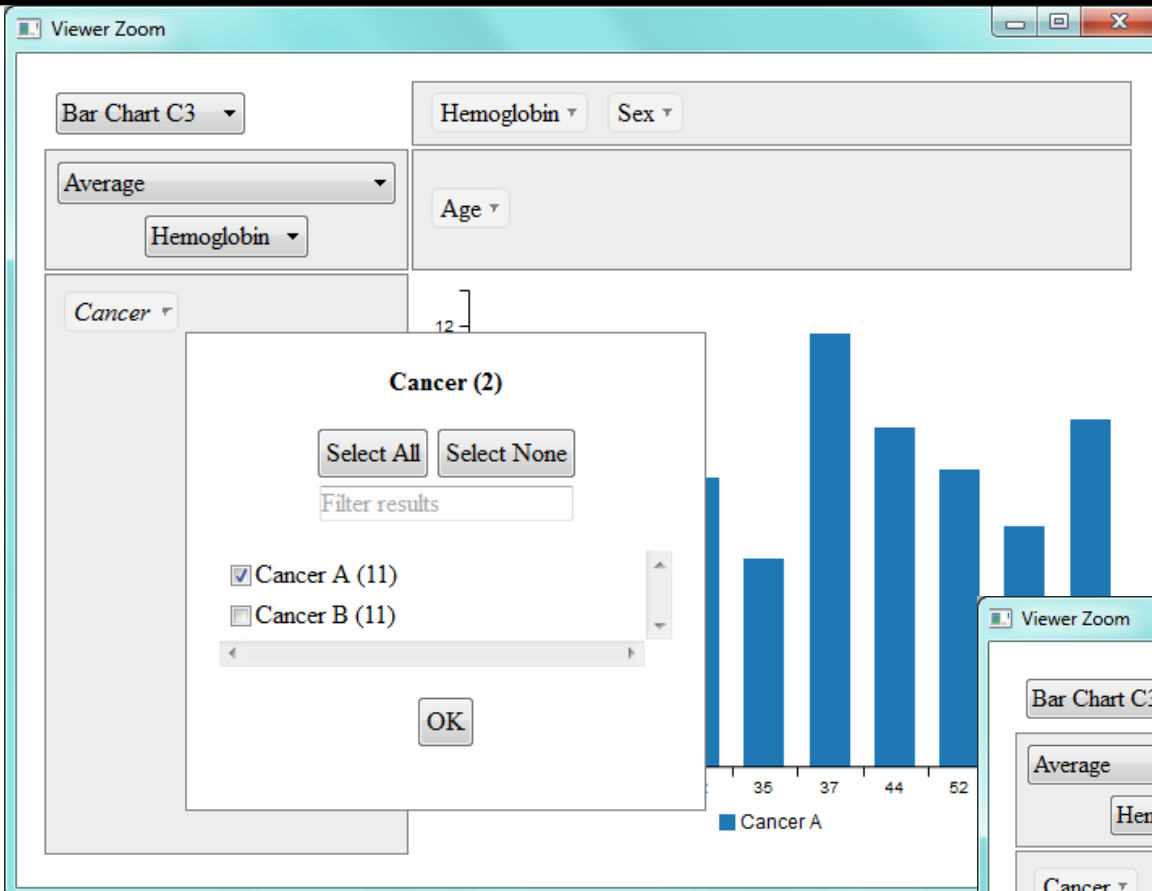
ggbio: An R implementation for extending the Grammar of Graphics for Genomic Data



rpivotTable – interactive pivot tables

Based on PivotTable.js by
[Nicolas Kruchten](http://nicolas.kruchten.com/pivottable)

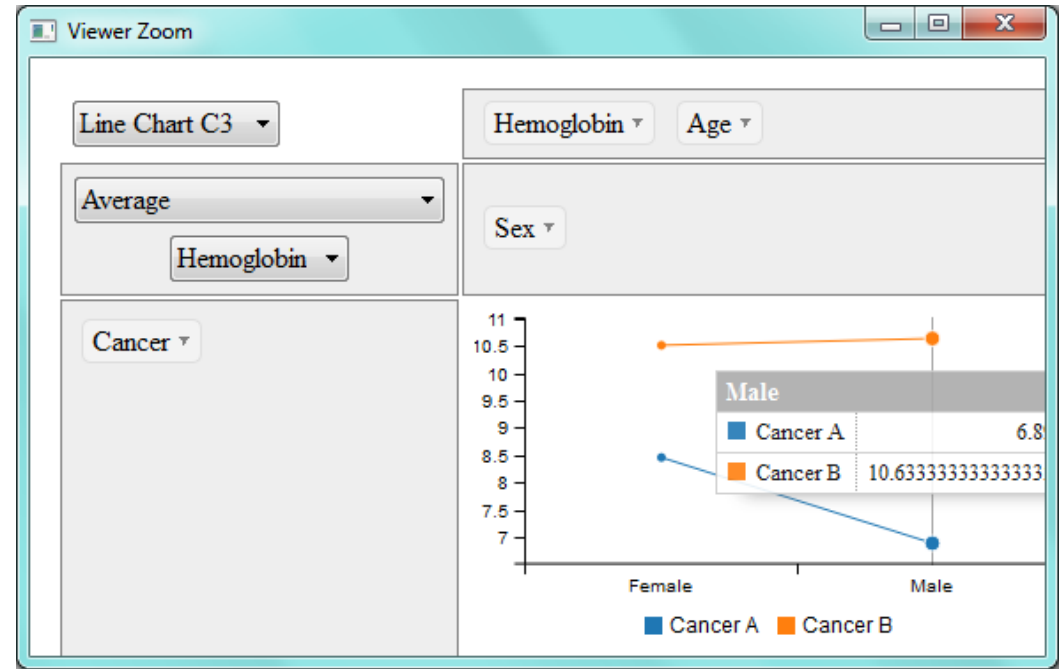
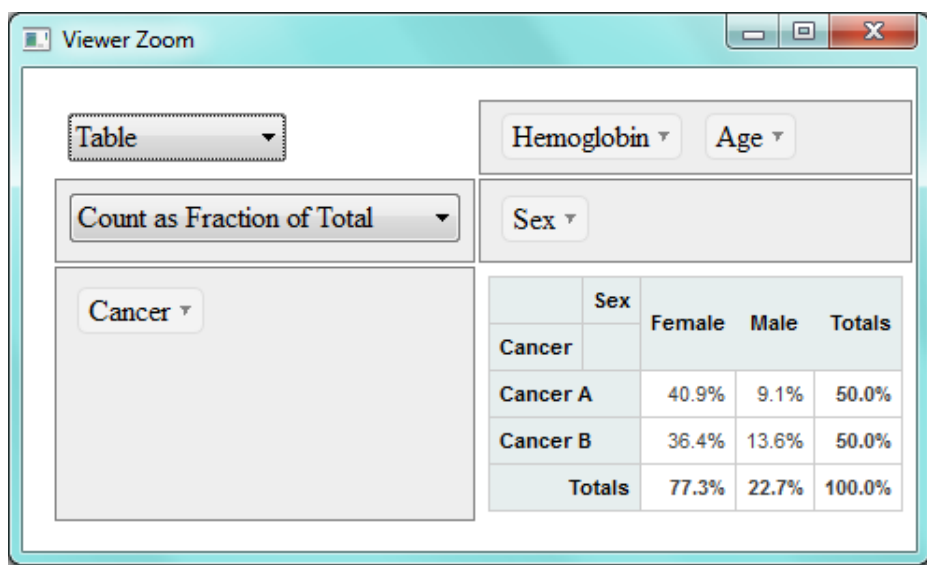
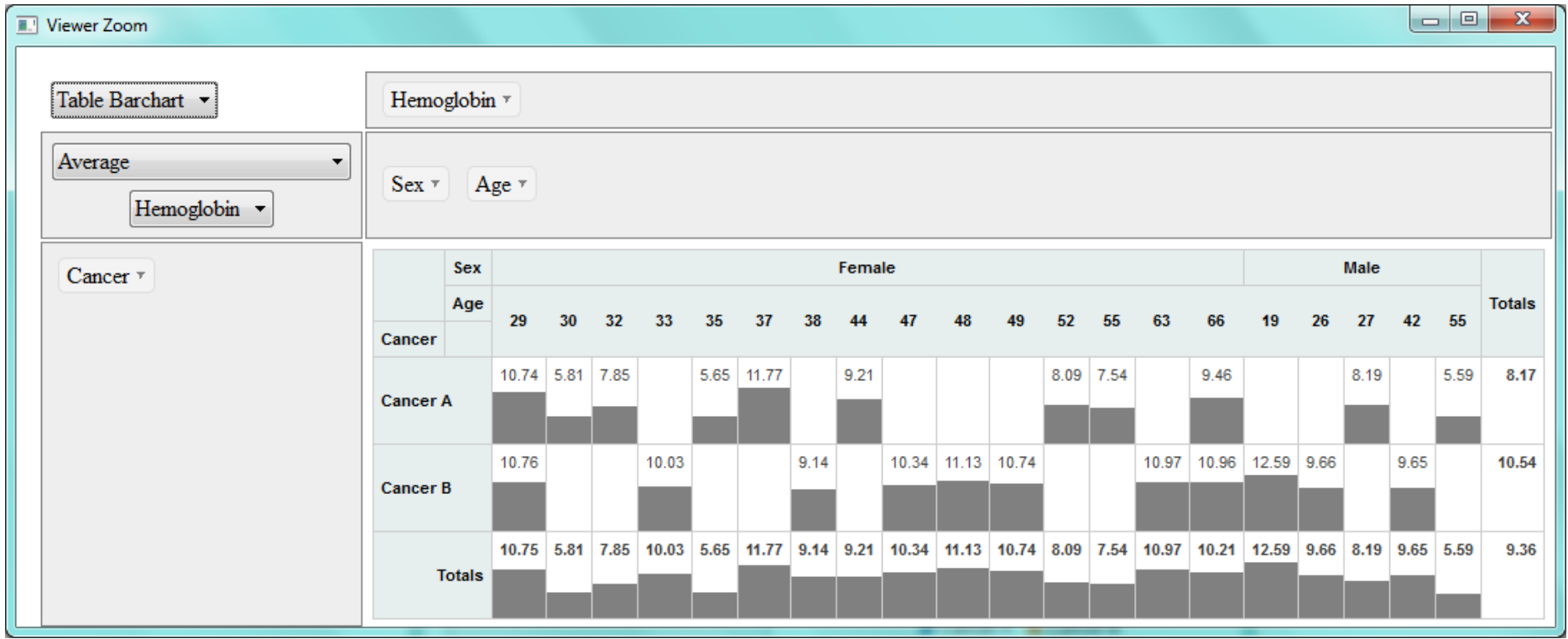
<http://nicolas.kruchten.com/pivottable>



<http://www.magesblog.com/2015...>



rpivotTable



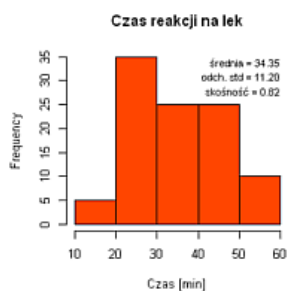
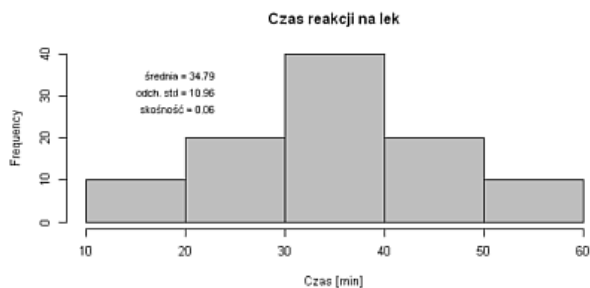


Charting subsystems

Base, default library

by Ross Ihaka
University of Auckland

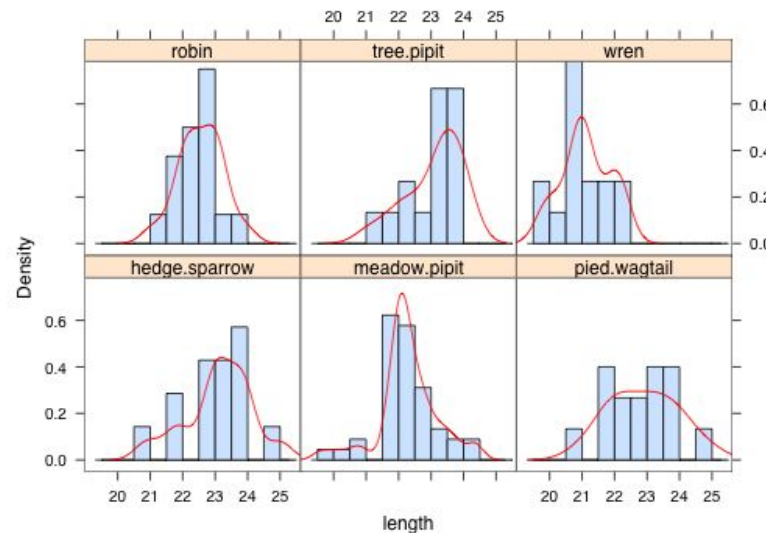
- low-level – graph. primitives
- easy to learn*
- most powerful – no limits
- well readable, ascetic
- may involve a lot of coding to get fancy results
- supports multiple plots
- interactive locator of points
- No anti-aliasing but it can draw on Cairo devices



Trellis

by Deepayan Sarka
University of Wisconsin

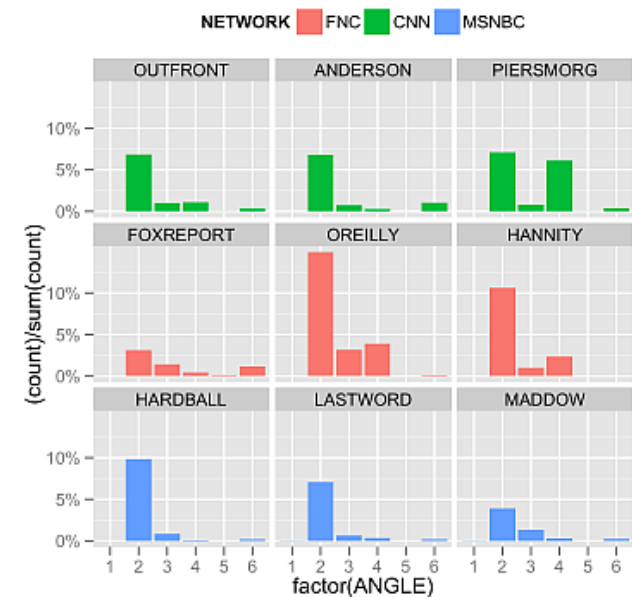
- designed for viewing multivariable datasets (grid of conditioned plots)
- well readable
- closed set of available diagrams: Barplot, Dotplot, Box and Whiskers, Histogram, Density, QQ, Scatterplot
- incompatible with other systems



ggplot2

by Hadley Wickham
Rice University

- high-level and well organized
- implementation of *Grammar of Graphics*
- powerful, highly customizable
- well readable, polished output
- anti-aliased by design
- closed set of diagrams but easily expandable
- supports multiple plots (grid)
- incompatible with others

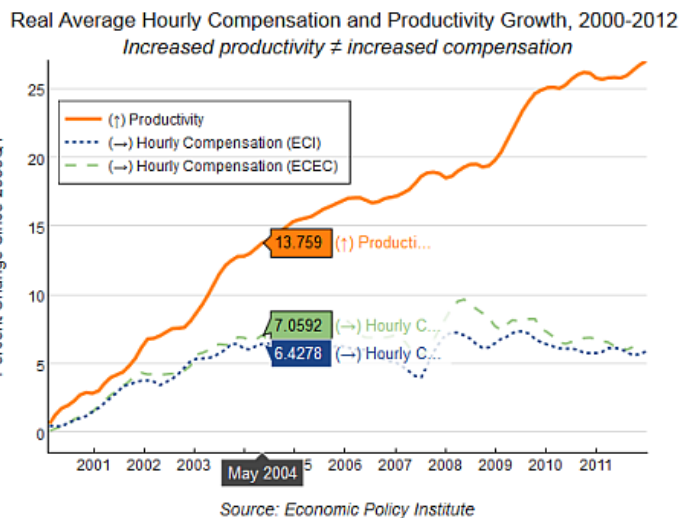




Charting subsystems

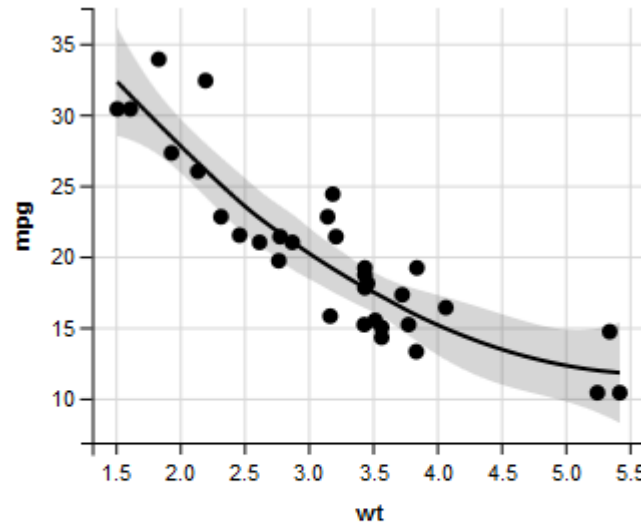
plotly [Teamwork](#)

- based on ggplot2
- interactive (JScript)
- WWW enabled (HTML/JSON/JS)
- breathtaking output –must see!
- well readable, clean
- rich library of examples
- incompatible with other systems



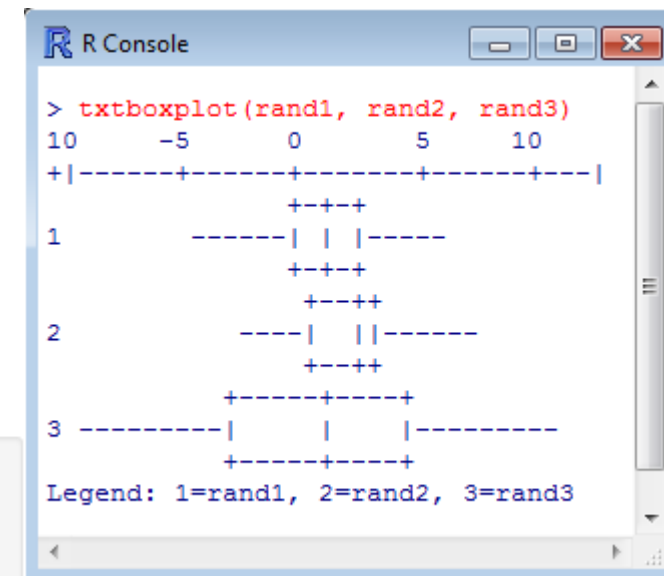
ggvis by RStudio team

- another implementation of *Grammar of graphics*, similar in spirit to ggplot2
- Interactive (JScript)
- WWW enabled (HTML, JS)
- well readable
- designed for [Shiny](#)



txtplot by Bjoern Bornkamp

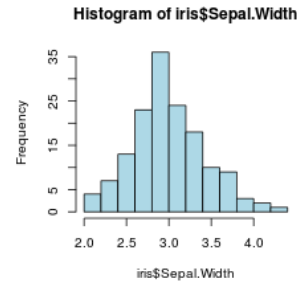
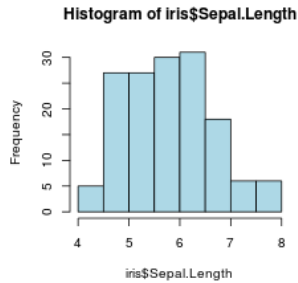
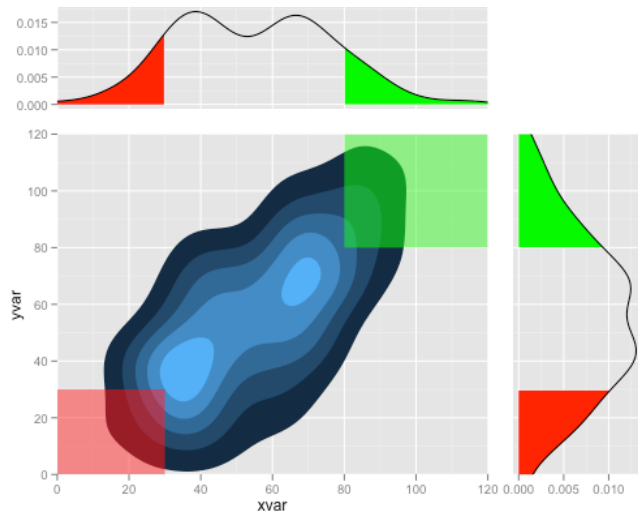
- produces graphs in pure ASCII
- rudimentary output
- closed set of plots (boxplot, lineplot, barplot, density, ACF)
- really useful when resources are limited (mobile devices, simple LCD displays, etc.) or output must be textual
- incompatible with others





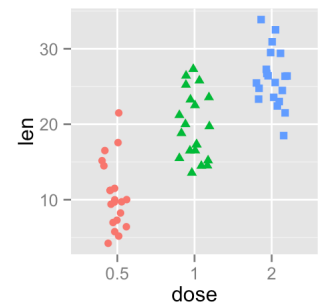
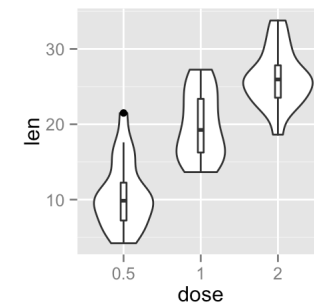
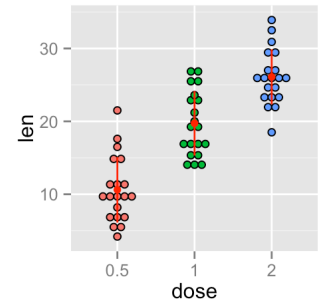
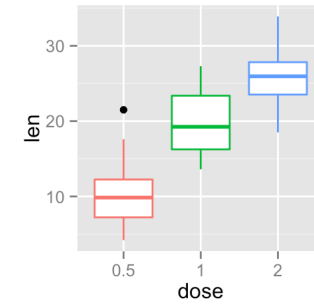
Additional helpers for ggplot2: **gridExtra**

Helps to arrange multiple ggplot2 objects on the same page

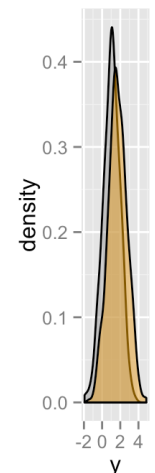
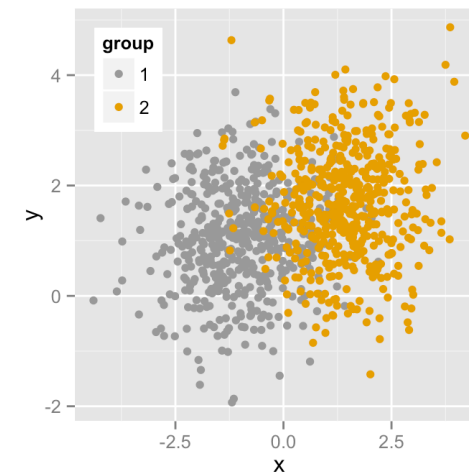
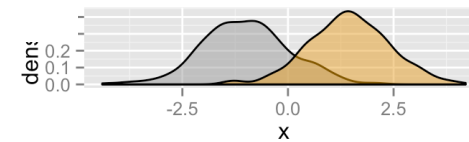
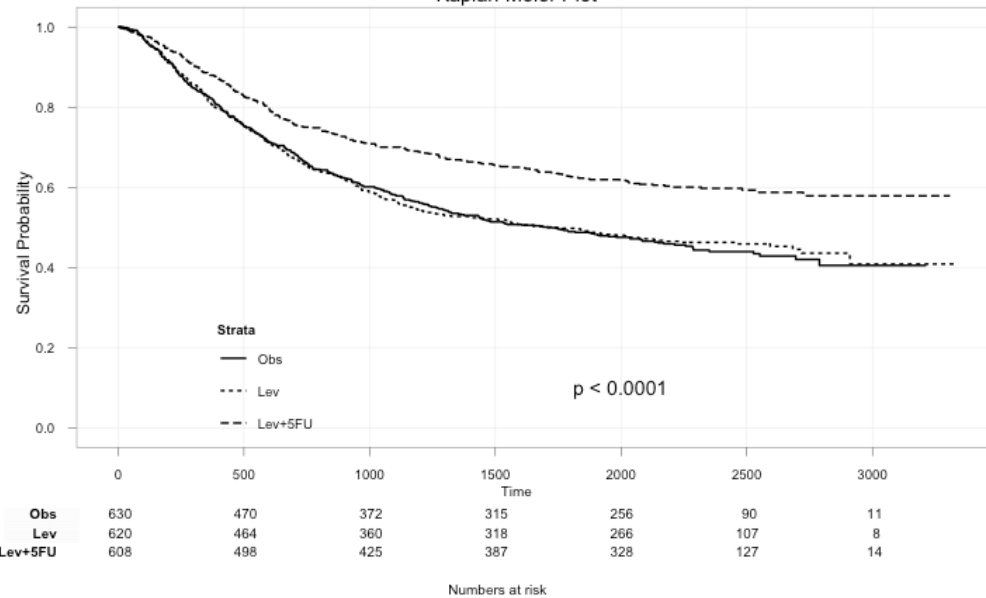


	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa

Multiple plots on the same page



Kaplan-Meier Plot

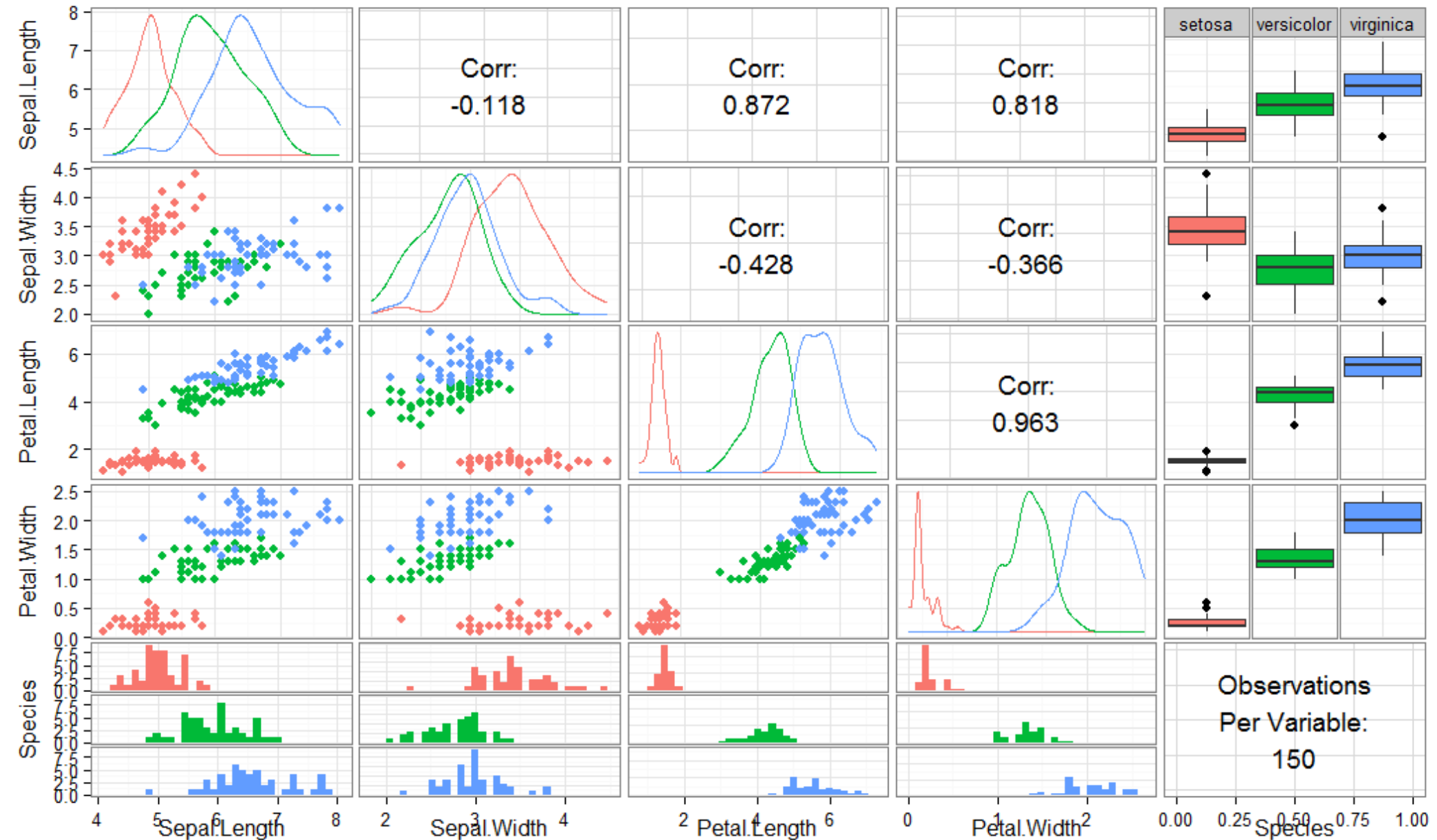




Additional helpers for ggplot2: **GGally**

Creates a matrix of ggplot2 graphs for data exploration purposes.

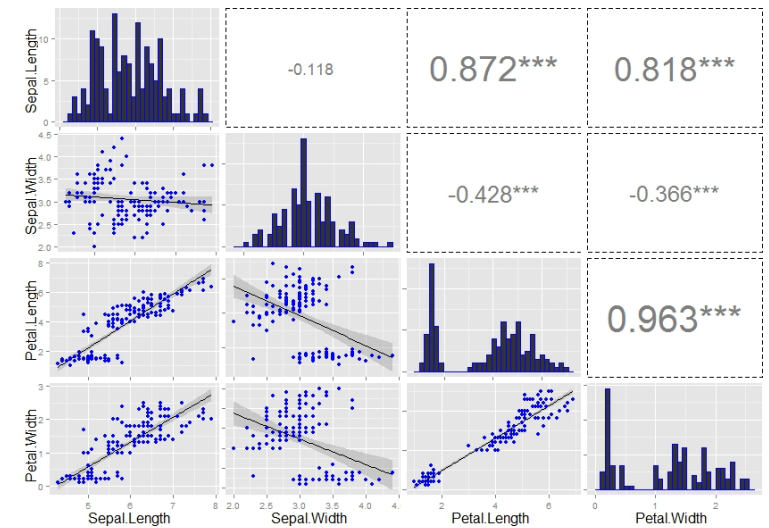
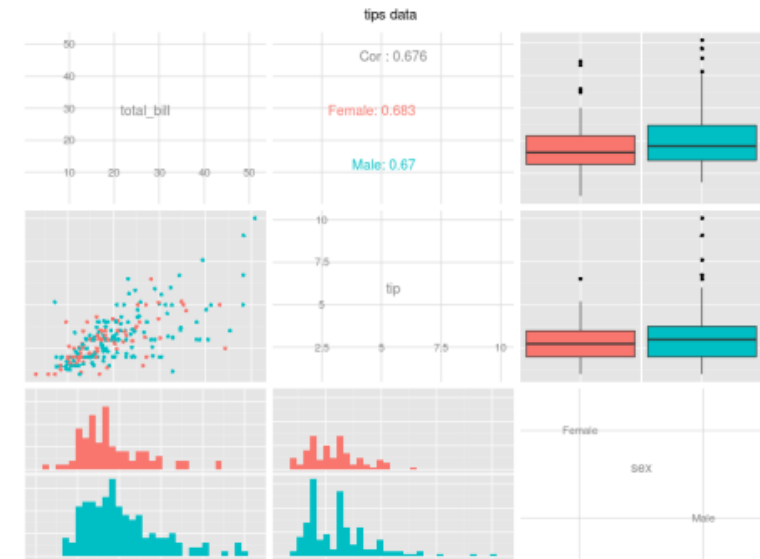
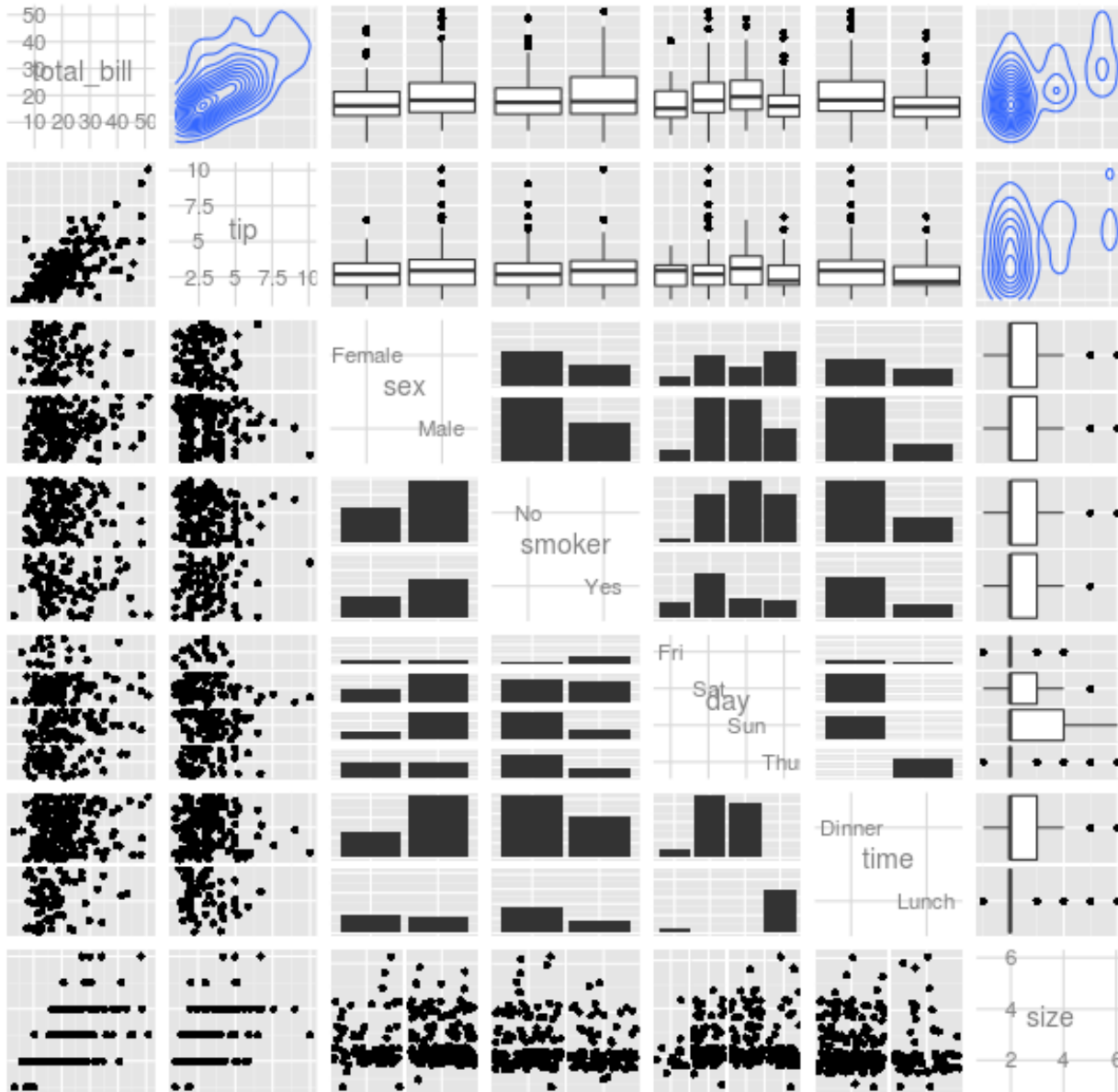
Anderson's Iris Data





Additional helpers for ggplot2: **GGally**

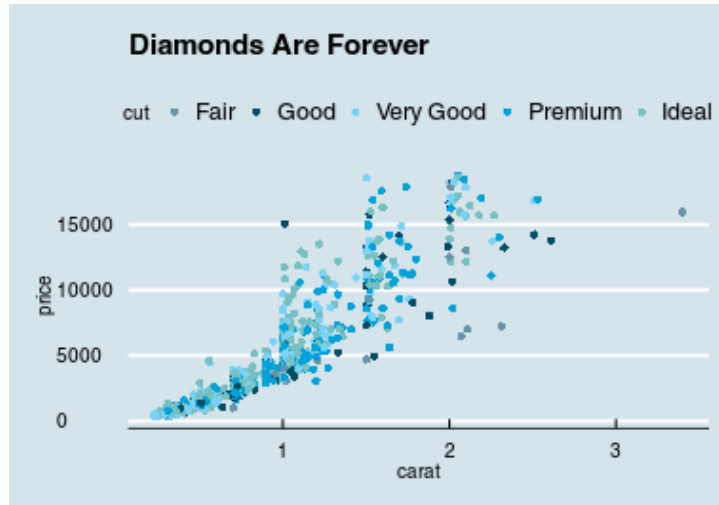
Creates a matrix of ggplot2 graphs for data exploration purposes.





Additional helpers for ggplot2: **ggthemes**

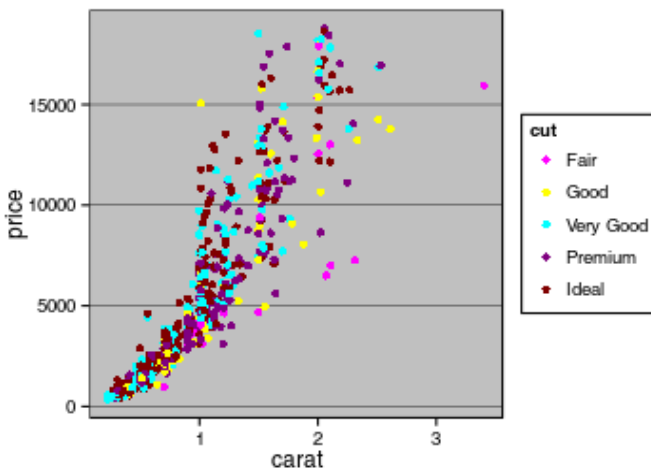
A set of themes (>15) for ggplot2: <http://github.com/jrnold/ggthemes>



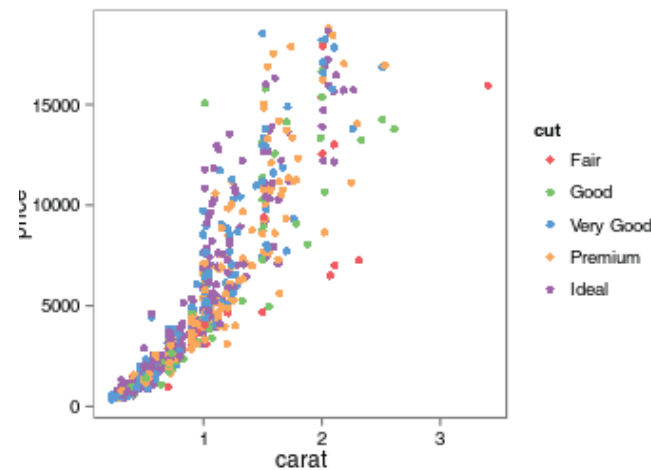
The Economist magazine theme



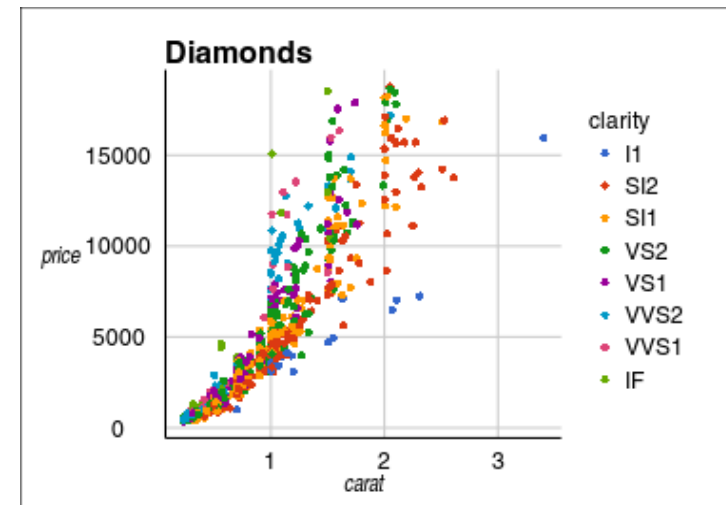
Stata theme



Excel theme



[Stephen Few's rules](#)



Google Docs theme



KMggplot2 plugin for RCommander

Scatter plot

X variable (pick one)
var1
var2

Y variable (pick one)
var1
var2

Stratum variable
var3

Facet variable in rows
var3

Facet variable in cols
var3

Horizontal axis label
<auto>

Vertical axis label
<auto>

Legend label
<auto>

Title
Some plot

Smoothing type

- None
- Smoothing with C.I. (linear regression)
- Smoothing without C.I. (linear regression)
- Smoothing with C.I. (loess or gam)
- Smoothing without C.I. (loess or gam)

Font size
14

Font family
serif
sans
mono
AvantGarde
Bookman

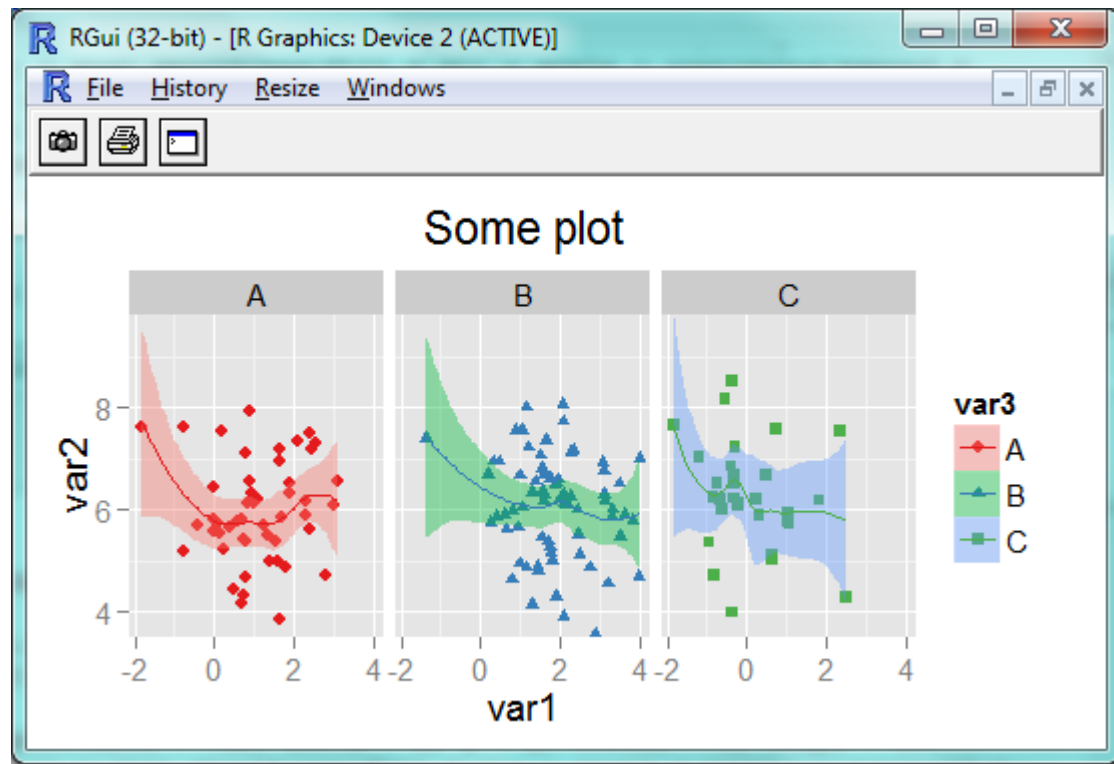
Colour pattern
Set1
BrBG
PiYG
PRGn
PuOr

Graph options
 Save graph

Theme
theme_bw
theme_simple
theme_classic
theme_gray
theme_minimal

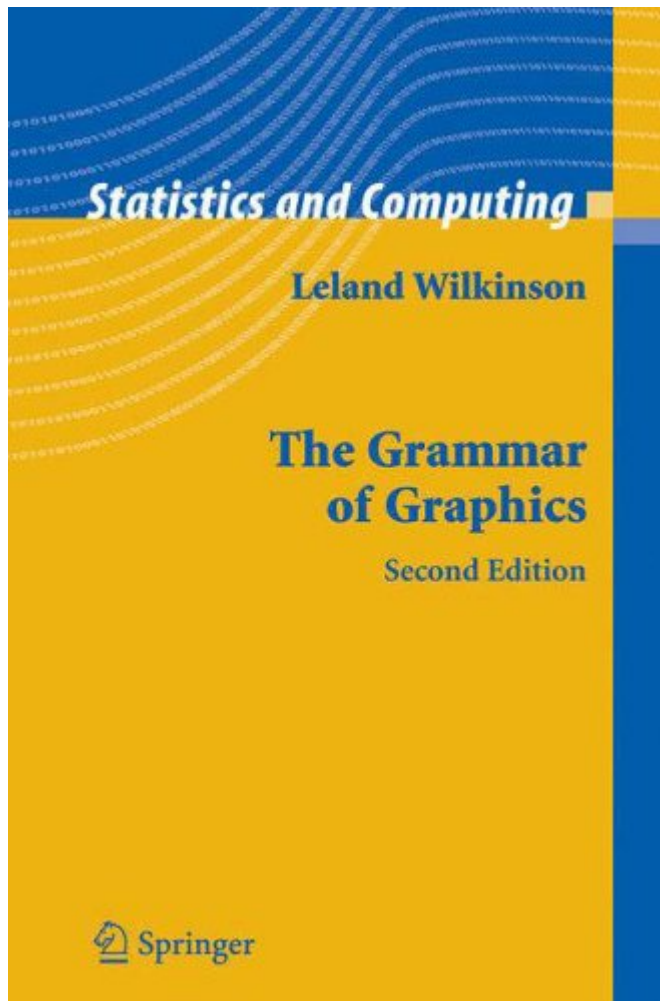
KMggplot2 Tools Help

- Kaplan-Meier plot...
- Histogram...
- Q-Q plot...
- Box plot / Violin plot / Confidence interval...
- Scatter plot...
- Scatter matrix...
- Line chart...
- Pie chart...
- Bar chart for discrete variables...
- Contour plot...
- Data handling... ▶
- Plot distribution... ▶





Since ggplot2 is an implementation of [Grammar of Graphics](#), which defines any graphics as a set of objects and layers and properties, it is possible to create graphical editor working on the principle “point and click”.



The [Deducer](#) package is an attempt to achieve this goal drawing from the power of [ggplot2](#).

It allows the user to define complex panel of charts using only mouse.

This is, in my opinion, one of the most advanced, graphical, free chart creator available in the Internet.



Deducer

Plot Builder

File Tools Window

Templates Geometric Elements Statistics Scales Facets Coordinates Other

template bar
template grouped bar
template histogram
template density
template grouped density
template simple dotplot
template grouped dotplot
template mean

Drag a component from above
OR
Select a plot type:

template histogram
template bar
template mean
template scatter
template grouped dotplot
template grouped line
template histogram 2d
template bubble

Components

Run Reset Cancel



Deducer

Templates	Geometric Elements	Statistics	Scales	Facets	Coordinates	Other
template bar	template grouped bar	template histogram	template density	template grouped density	template simple dotplot	template grouped dotplot
template line	template grouped line	template simple boxplot	template group boxplot	template scatter	template scatter smooth	template histogram 2d
						template bubble

Templates	Geometric Elements	Statistics	Scales	Facets
facet grid	facet wrap			

Templates	Geometric Elements	Statistics	Scales	Facets	Coordinates
coord cartesian	coord equal	coord flip	coord map	coord polar	coord trans

Templates	Geometric Elements	Statistics	Scales	Facets	Coordinates	Other
geom abline	geom area	geom bar	geom bin2d	geom blank	geom boxplot	geom contour
geom crossbar	geom density	geom density2d	geom dotplot	geom errorbar	geom errorbarh	geom freqpoly
geom hex	geom histogram	geom hline	geom jitter	geom line	geom linerrange	geom path
geom point	geom pointrange	geom polygon	geom quantile	geom raster	geom rect	geom text

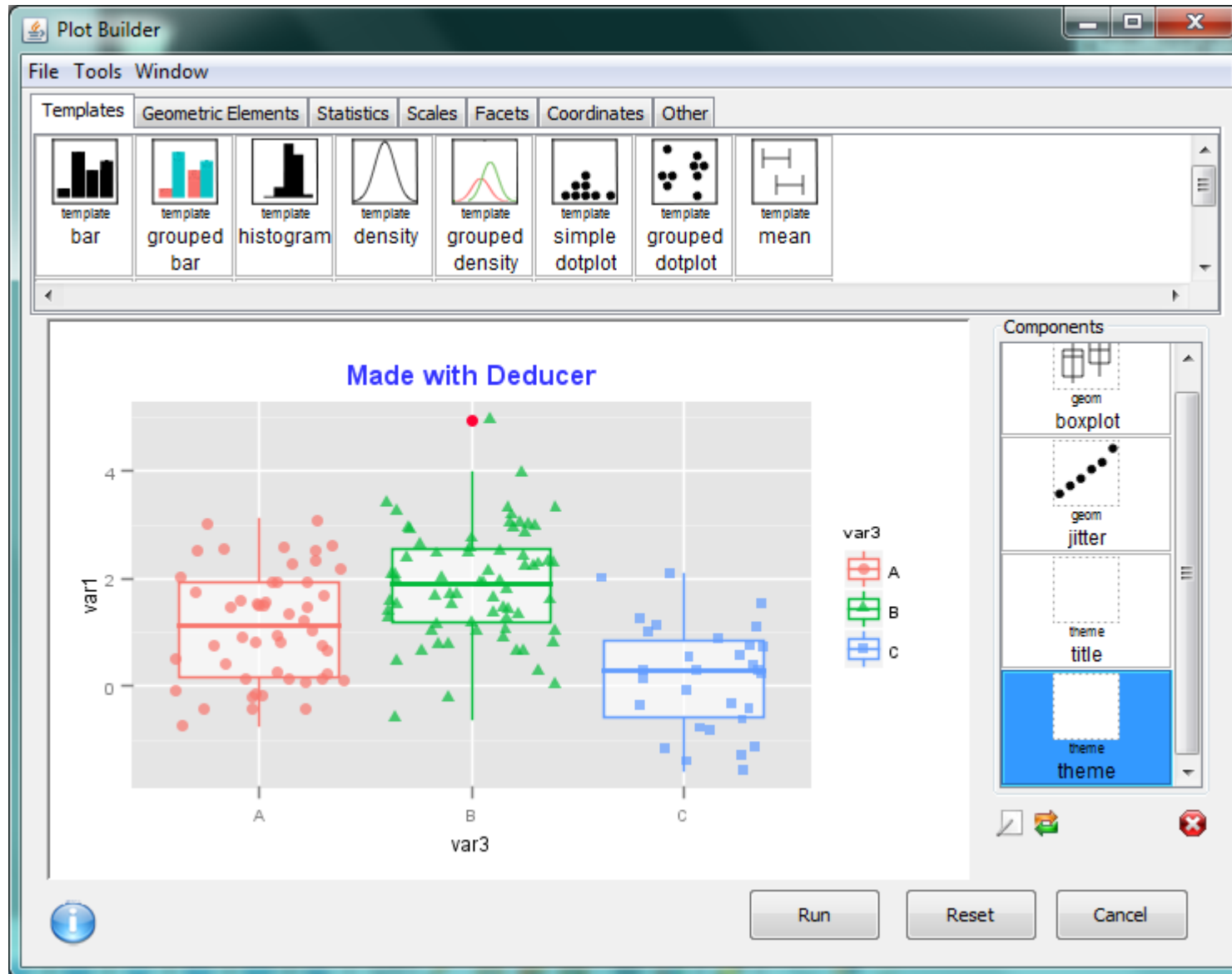
Templates	Geometric Elements	Statistics	Scales	Facets	Coordinates	Other
theme grey	theme bw	theme classic	theme minimal	theme theme	theme title	theme xlab
						theme ylab

Templates	Geometric Elements	Statistics	Scales
scale X continuous	scale X discrete	scale Y date	scale Y datetime
scale Y continuous	scale Y discrete	scale Y date	scale Y datetime

Templates	Geometric Elements	Statistics	Scales	Facets	Coordinates	Other
stat abline	stat bin	stat bin2d	stat bindot	stat binhex	stat boxplot	stat contour
stat density	stat density2d	stat ecdf	stat function	stat hline	stat identity	stat qq
stat quantile	stat smooth	stat spoke	stat sum	stat summary	stat unique	stat ydensity
						stat vline



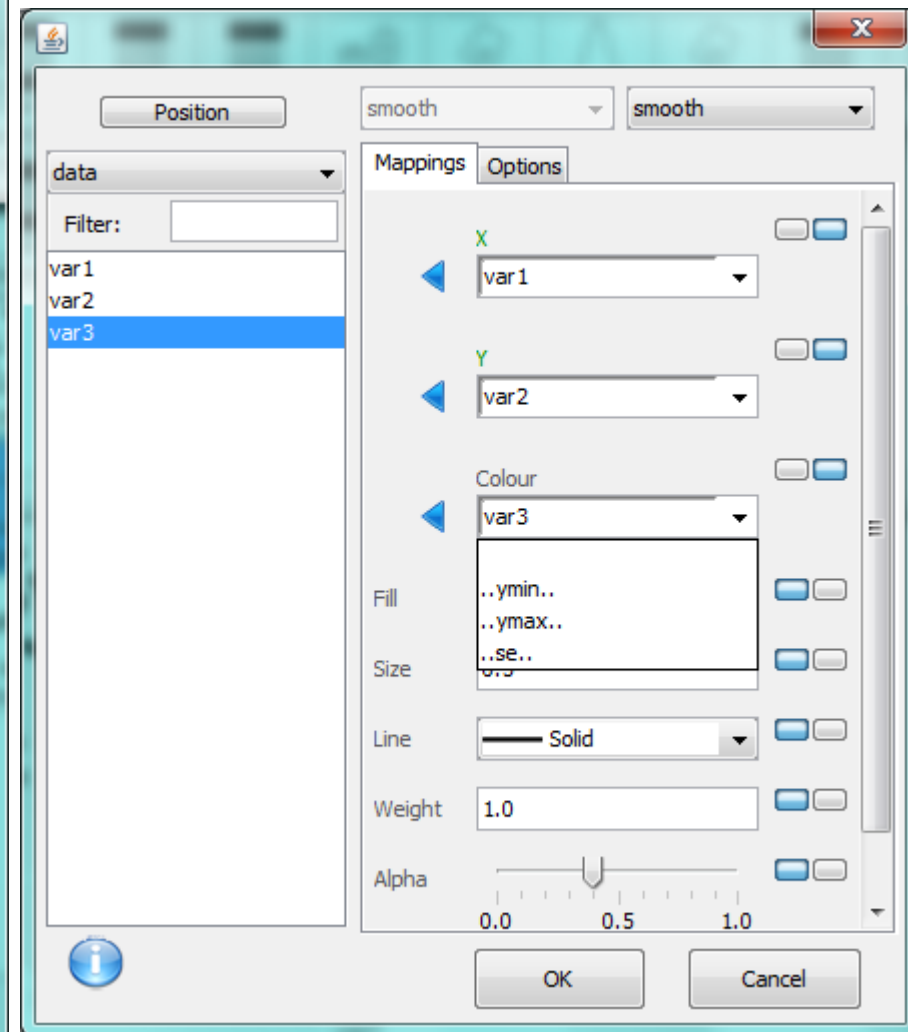
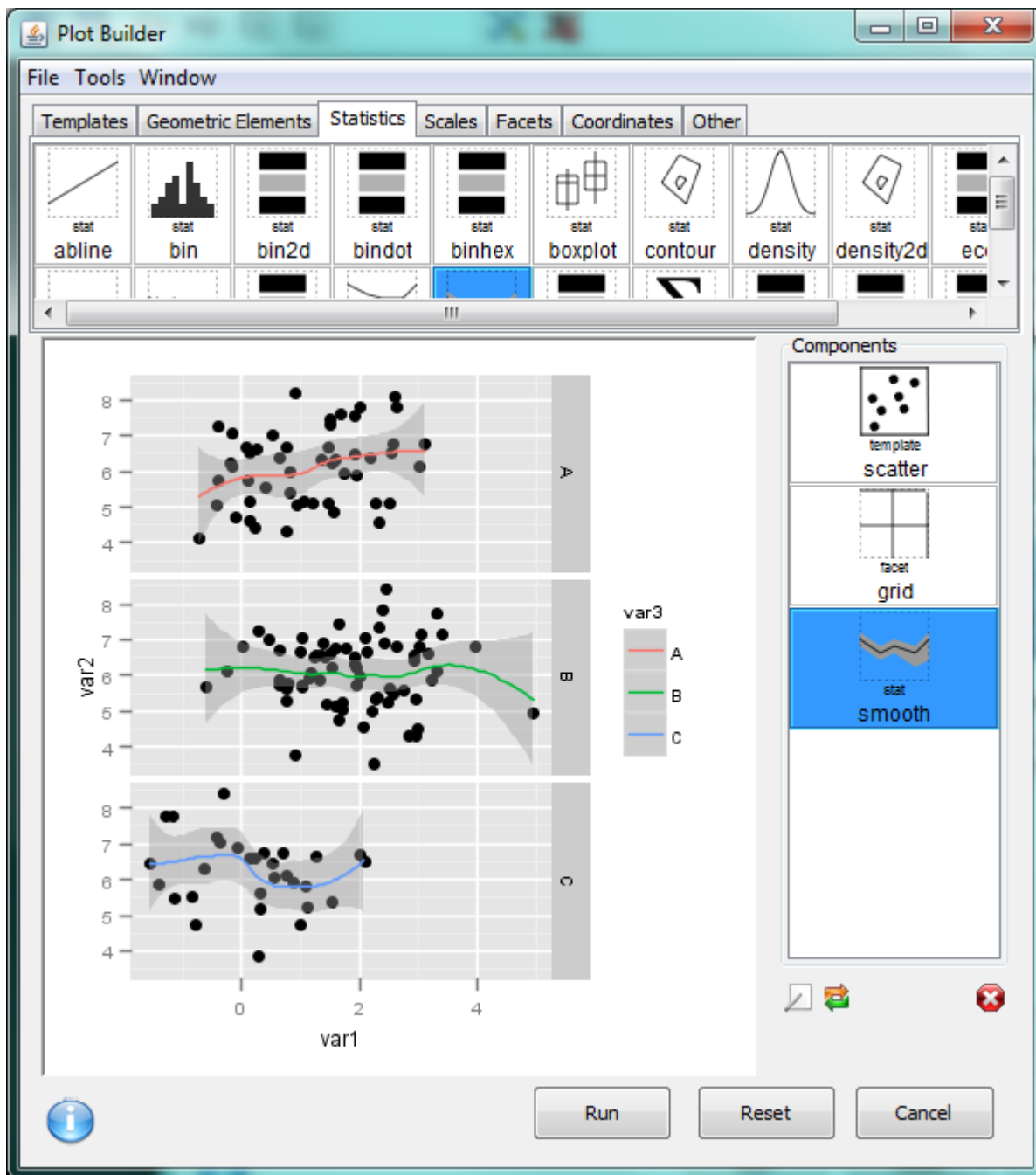
Deducer



```
ggplot() + geom_boxplot(aes(y = var1,x = var3,colour = var3),data=data,alpha = 0.6,outlier.colour = '#ff0033') +  
  geom_jitter(aes(x = var3,y = var1,shape = var3,colour = var3),data=data,alpha = 0.7) +  
  ggtitle(label = 'Made with Deducer') +  
  theme(plot.title = element_text(family = 'Helvetica',face = 'bold',colour = '#3333ff',size = 17.0,vjust =  
1.0),panel.border = element_line())
```




Deducer





Deducer – control dozens of graph properties

Plot

Background

Title text

Margin

Axis

Axis title text

Axis text

Line

Ticks

Tick length

Tick margin

x-Axis

Title

Text

OK Cancel Apply

element_text

font family

face

colour

size

vjust

hjust

angle

line height

OK Cancel

Choose Colour

Swatches

Preview

Sample Text Sample Text

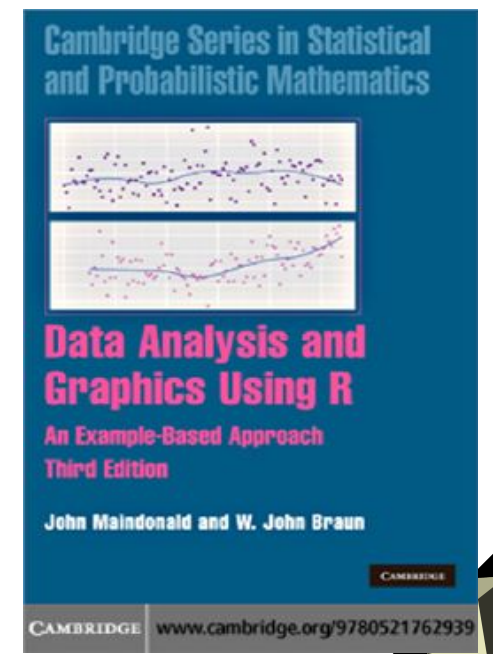
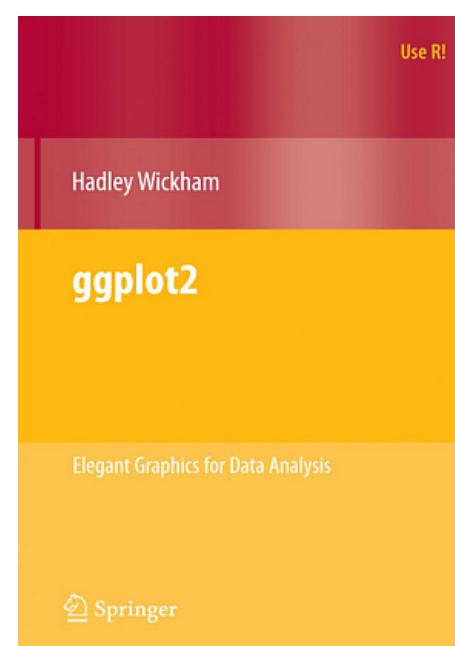
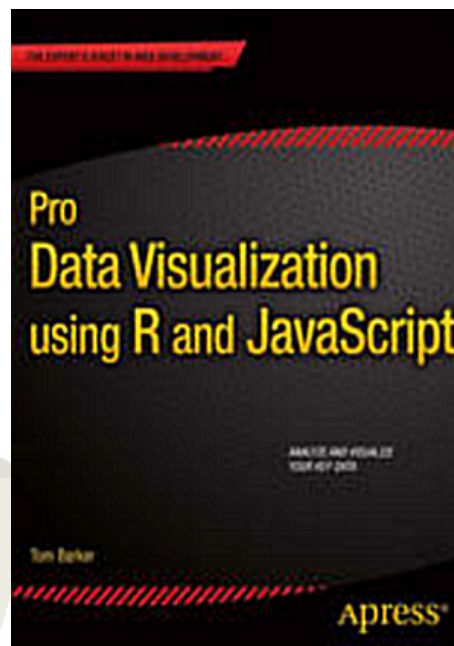
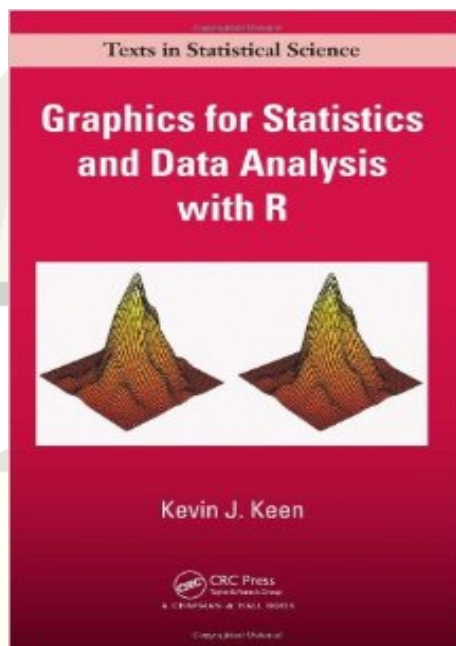
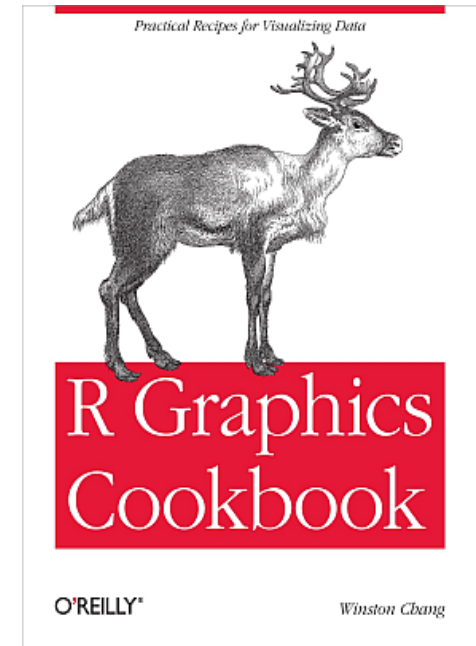
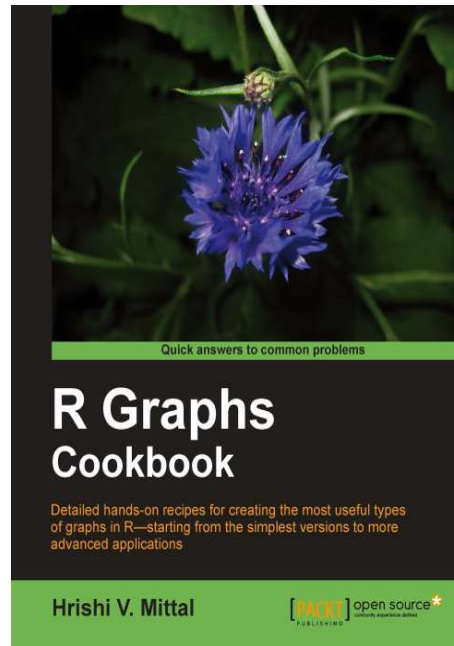
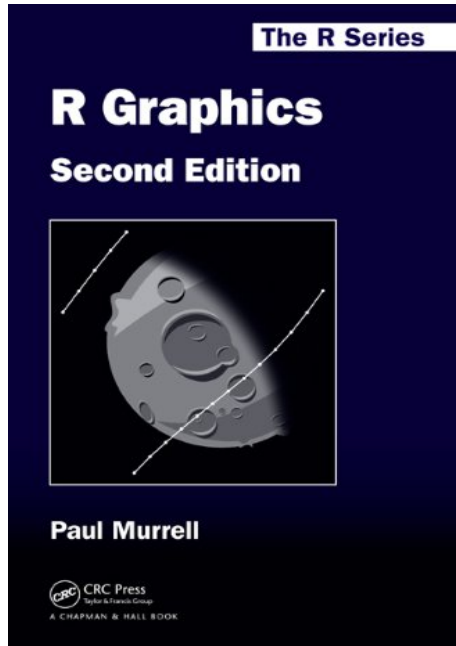
Sample Text Sample Text

Sample Text Sample Text

OK Cancel Reset



Some books about data visualization in R

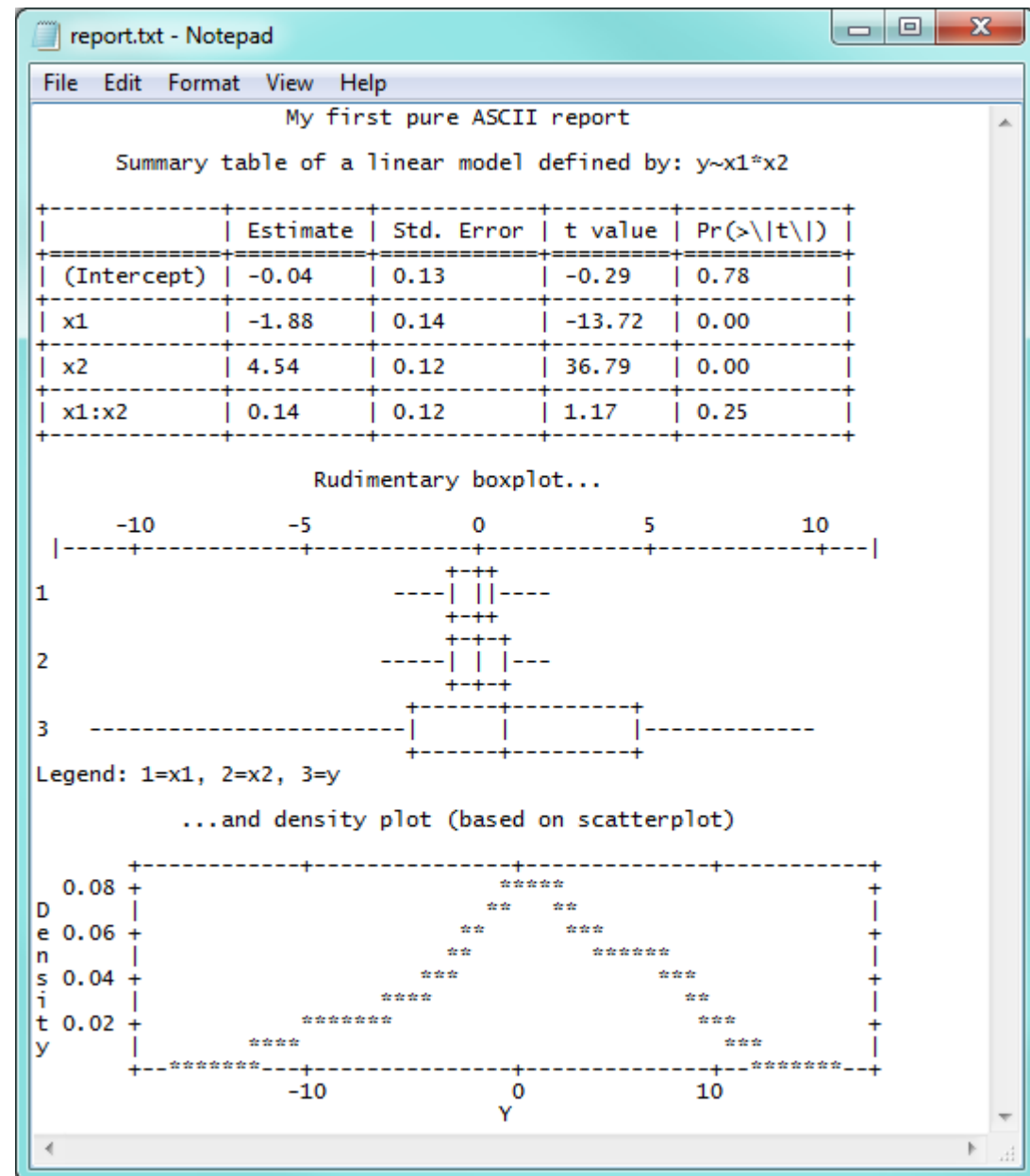




Plain ASCII output

With [ascii](#) and [txtplot](#) packages one can create text logs (listings) like SAS does.

```
RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help
> library(ascii)
> library(txtplot)
> library(stringr)
> width <- options()$width*0.8
> sink("report.txt")
> cat(str_pad("My first pure ASCII report", width)
> cat("\n\n")
> data <- mutate(data.frame(x1=rnorm(50), x2=rno$
> model <- lm(y~x1*x2, data=data)
> cat(str_pad("Summary table of a linear model d$
> cat("\n")
> suppressWarnings(print(ascii(model), "rest"))
> cat("\n")
> cat(str_pad("Rudimentary boxplot...", width=wi$
> cat("\n\n")
> with(data, txtboxplot(x1, x2, y, width=width))
> cat("\n")
> cat(str_pad("...and density plot (based on sca$
> cat("\n\n")
> dens <- with(data, density(y))
> with(dens, txtplot(x, y, width=round(width), h$
> sink()
> shell("notepad report.txt")
```





13 reasons why **you will** love GNU R

- I R is (extremely) cheap. In fact - it's free :)
- II R has (extremely) wide range of capabilities
- III R is (widely) supported by the world of science
- IV R is supported by the community
- V R is (increasingly) supported by the business
- VI R is able to read data in many formats
- VII Interoperability is easy to achieve
- VIII R is truly cross-platform
- IX R offers numerous ways of presenting data
- X There are many options to optimize the code**
- XI R is able to handle large amount of data
- XII R has a set of fancy tools and IDEs
- XII FDA accepted using R for drug trials!



R offers a bunch of methods that can remarkably **increase the performance** of algorithms in the case of performing complex calculations:

- **Profiling** the code, which enables us to determine the computationally intensive portions of the program
- Turning on the “**Just In Time**” compilation of the code
- **Vectorizing** calculations – which means avoiding explicit loops. It can speed up computations 5-10 times
- Performing all **algebraic** computations with the use of libraries **tuned** for our hardware (e.g. BLAS)
- Become familiar with methods of **algorithmic differentiation** (ADMB)
- Executing **parallel** computations in a **cluster** environment
- Using the power of a **graphic card processor** (CUDA, OpenCL)
- **Implement key parts of algorithm in C++** and call them using RCPP



Profiling the memoRy usage

Memory usage · Advanced R. x +

adv-r.had.co.nz/memory.html

Search

Advanced R by Hadley Wickham

[Table of contents](#)

Want to learn from me in person? I'm next teaching in [Chicago, May 27-28](#).

Want a physical copy of this material? [Buy a book from amazon!](#)

Contents

[Object size](#)

[Memory usage and garbage collection](#)

[Memory profiling with lineprof](#)
[Modification in place](#)

[How to contribute](#)

[Edit this page](#)

Memory

A solid understanding of R's memory management will help you predict how much memory you'll need for a given task and help you to make the most of the memory you have. It can even help you write faster code because accidental copies are a major cause of slow code. The goal of this chapter is to help you understand the basics of memory management in R, moving from individual objects to functions to larger blocks of code. Along the way, you'll learn about some common myths, such as that you need to call `gc()` to free up memory, or that `for` loops are always slow.

Outline

- [Object size](#) shows you how to use `object_size()` to see how much memory an object occupies, and uses that as a launching point to improve your understanding of how R objects are stored in memory.
- [Memory usage and garbage collection](#) introduces you to the `mem_used()` and `mem_change()` functions that will help you understand how R allocates and frees memory.
- [Memory profiling with lineprof](#) shows you how to use the `lineprof` package to understand how memory is allocated and released in larger code blocks.
- [Modification in place](#) introduces you to the `address()` and `refs()` functions so that you can understand when R modifies in place and when R modifies a copy. Understanding when objects are copied is very important for writing efficient R code.

Prerequisites

In this chapter, we'll use tools from the `pryr` and `lineprof` packages to understand memory usage, and a sample dataset from `ggplot2`. If you don't already have them, run this code to get the packages you need:

<http://adv-r.had.co.nz/memory.html>



Profiling the memoRy usage

Memory usage · Advanced R. x +

<http://adv-r.had.co.nz/memory.html>

adv-r.had.co.nz/memory.html

Search

Want to learn from me in person? I'm next teaching in Chicago, May 27-28.

Want a physical copy of this material? [Buy a book from amazon!](#)

Contents

[Object size](#)

[Memory usage and garbage collection](#)

[Memory profiling with lineprof](#)

[Modification in place](#)

[How to contribute](#)

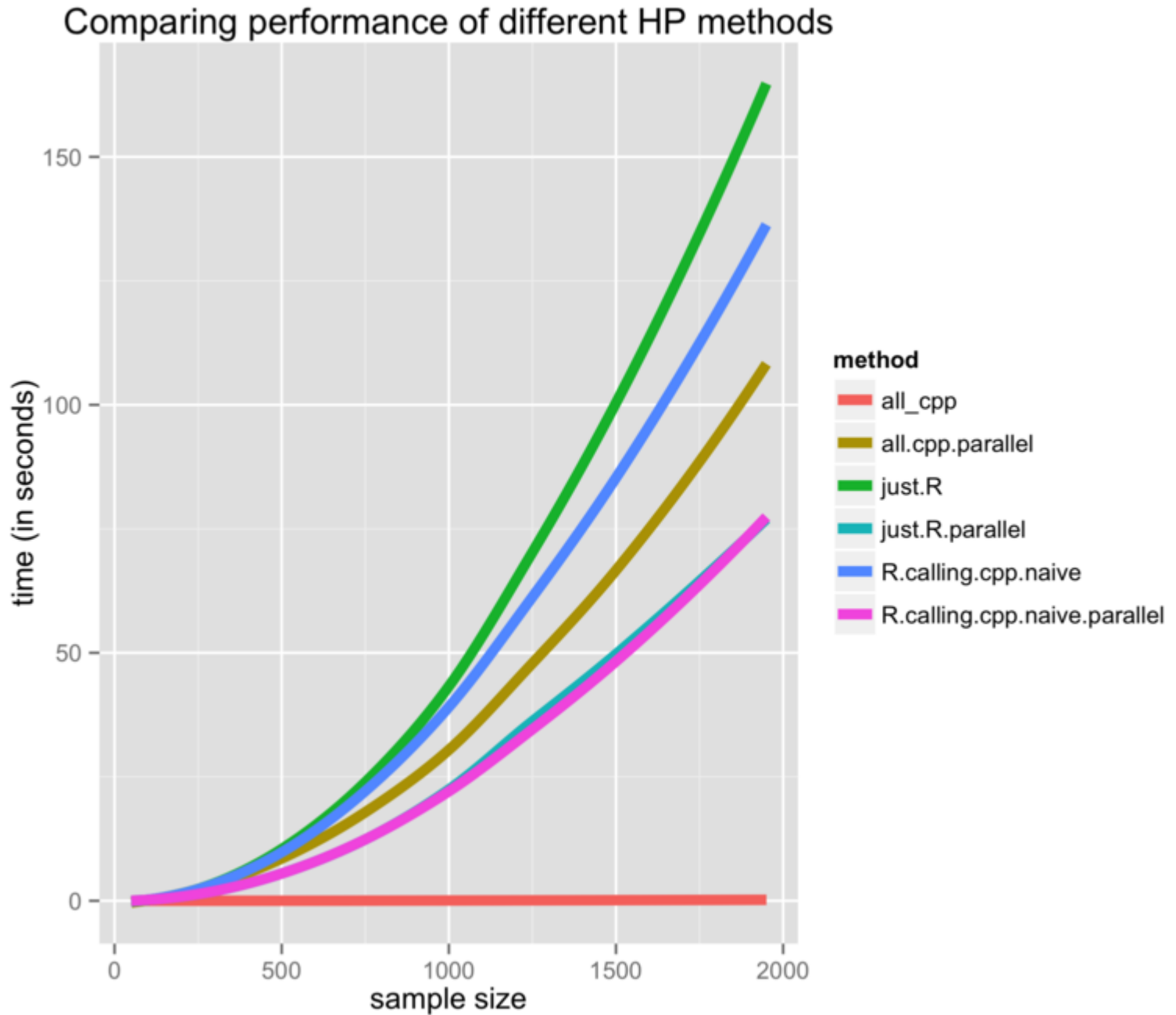
[Edit this page](#)

```
source("code/read-delim.R")
prof <- lineprof(read_delim("diamonds.csv"))
shine(prof)
```

#	Source code	t	r	a	d
1	# ---- read_delim				
2	read_delim <- function(file, header = TRUE, sep = ",") {				
3	# Determine number of fields by reading first line				
4	first <- scan(file, what = character(1), nlines = 1, se...				
5	p <- length(first)				
6					
7	# Load all fields as character vectors				
8	all <- scan(file, what = as.list(rep("character", p)), ...	█	█	█	
9	skip = if (header) 1 else 0, quiet = TRUE)				
10					
11	# Convert from strings to appropriate types (never to f...				
12	all[] <- lapply(all, type.convert, as.is = TRUE)	█	█	█	
13					
14	# Set column names				
15	if (header) {				
16	names(all) <- first				
17	} else {				
18	names(all) <- paste0("V", seq_along(all))				
19	}				
20					
21	# Convert list into data frame				
22	as.data.frame(all)	█	█	█	
23	}				



Tune it up!





Debug your code!

R-debug-tools.pdf

www.biostat.jhsph.edu/~rpeng/docs/R-debug-tools.pdf

Search

Page: 1 of 15

100%

An Introduction to the Interactive Debugging Tools in R

Roger D. Peng

UCLA Department of Statistics

August 28, 2002

1 Introduction

The purpose of this document is to provide a brief introduction to the built-in program debugging tools in the R statistical computing environment. The five functions that will mainly be covered are `traceback`, `debug`, `browser`, `trace`, and `recover`. Throughout, the `typewriter` font will be used to indicate R code.

It is important to note that debugging is a practice which gets considerably easier as one's familiarity with the language increases. In some ways it can be more "art" than "science". For example, knowing where to look in a 500 line program after it has just halted execution is sometimes just a "feeling" one develops after much previous suffering.

2 Trouble with your Droids?

R has a very rich language with which users can write very useful but potentially complex functions. As with programs written in any other language, functions written in R can contain unforeseen problems which lead to failure. The purpose of the debugging tools is to help the programmer find these problems quickly and efficiently. R is a generally friendly language and the problems which arise are typically unlike those found in a language such as



Debug your code!

```
RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help
[Icons: File Explorer, Print, Save, Copy, Paste, Refresh, Stop, Print]

> fn <- function(x, y) {
+   z <- x * y
+   browser()
+   return(sum(z > 6))
+ }
> fn(x = 1:5, y = 2:6)
Called from: fn(x = 1:5, y = 2:6)
Browse[1]> n
debug at #4: return(sum(z > 6))
Browse[2]> x
[1] 1 2 3 4 5
Browse[2]> y
[1] 2 3 4 5 6
Browse[2]> n
[1] 3
>
> debug(fn)
> fn(x = 1:5, y = 2:6)
debugging in: fn(x = 1:5, y = 2:6)
debug at #1: {
  z <- x * y
  browser()
  return(sum(z > 6))
}
Browse[2]> n
debug at #2: z <- x * y
Browse[2]> n
debug at #3: browser()
Browse[2]> z
[1] 2 6 12 20 30
Browse[2]> n
Browse[2]> n
debug at #4: return(sum(z > 6))
Browse[2]> n
exiting from: fn(x = 1:5, y = 2:6)
[1] 3
> |
```



Debug your code!

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains the following R code:

```
1  
2 f1 <- function(p) {  
3   x <- 1  
4   y <- 2  
5  
6   if(p > y) {  
7     x <- 5  
8   }  
9 }  
10  
11 f1(6)
```
- Console:** Shows the execution of the code with debug output:

```
> debugSource('X:/projects/Test/debug.r', echo=TRUE)  
f1 <- function(p) {  
  x <- 1  
  y <- 2  
  
  if(p > y) {  
    x <- 5  
  }  
}  
  
f1(6)  
Called from: eval(expr, envir, enclos)  
Browse[1]> n  
debug at X:/projects/Test/debug.r#3: x <- 1  
Browse[2]> n  
debug at X:/projects/Test/debug.r#4: y <- 2  
Browse[2]> |
```
- Environment Pane:** Shows the current environment with variables:

Name	Type	Length	Size	Value
p	promise	0	0 B	6
x	numeric	1	32 B	1
- Traceback Pane:** Shows the call stack:

```
f1(6) at debug.r:4  
eval(expr, envir, enclos)  
eval(expr, globalenv())  
withVisible(eval(expr, envir = globalenv()))  
doTryCatch(return(expr), name, parentenv, handler...  
tryCatchone(expr, names, parentenv, handlers[[1L]]...
```



13 reasons why **you will** love GNU R

- I R is (extremely) cheap. In fact - it's free :)
- II R has (extremely) wide range of capabilities
- III R is (widely) supported by the world of science
- IV R is supported by the community
- V R is (increasingly) supported by the business
- VI R is able to read data in many formats
- VII Interoperability is easy to achieve
- VIII R is truly cross-platform
- IX R offers numerous ways of presenting data
- X There are many options to optimize the code
- XI R is able to handle large amount of data**
- XII R has a set of fancy tools and IDEs
- XII FDA accepted using R for drug trials!



“Big data” doesn't mean a “big trouble”

Sometimes there is a need for processing large datasets that don't fit in the RAM memory. Size of the data cannot exceed 2GB per process in a 32-bit machine and 8GB per process in a 64-bit machine.

In such cases R gives a [couple of methods](#) for handling massive data:

- Building models on chunks of the data. Only linear models are supported
- Mapping physical memory on a set of files (disk virtual memory)
- Storing parts of datasets outside of the current process memory



pbodR
Programming with Big Data in R

<http://www.r-pbd.org/>



“Big data” doesn't mean a “big trouble”

Get Technical Support

Community Applications Products Services Resources Company

Big Data Big Analytics Software

Innovation, performance and value for today's data architecture.

- WHAT IS R?
Open Source Data Science
- R IN BUSINESS
Your Competitive Edge
- RRE SOFTWARE
Big Data Big Analytics
- RRE IN THE CLOUD
Big Data R on AWS
- ADVISER
Support for Open Source R



Community Experience Distilled

Big Data Analytics with R and Hadoop

Set up an integrated infrastructure of R and Hadoop to turn your data analytics into Big Data analytics

Vignesh Prajapati

PACKT open source*
PUBLISHING ordinary experience. unusual.

High-Performance and Parallel Computing with R:

- Parallel computing:
 - Explicit and implicit parallelism
 - Grid computing
 - Hadoop
 - Random numbers
 - Resource managers and batch schedulers
 - Applications
 - GPUs
- Large memory and out-of-memory data
- Easier interfaces for Compiled code
- Profiling tools



“Big data” doesn't mean a “big trouble”

A screenshot of a PDF viewer window. The browser's address bar shows the file path: file:///C:/tmp/rre-in-teradata.pdf. The viewer interface includes a search bar, navigation icons, and a page indicator showing 'Page: 1 of 13' at 80% zoom. The PDF content features a header with a row of icons (orange square, grey square, dark grey square, document icon, handshake icon, lightbulb icon, antenna icon) and the 'REVOLUTION ANALYTICS' logo. Below the header, it is labeled 'WHITE PAPER'. The main title is 'Maximizing the Value of Big Data', followed by the subtitle 'Introducing RRE for the Teradata Database'. The author is listed as 'Bill Jacobs, Director, Product Marketing, Revolution Analytics', and the date is 'April 2014'.

rre-in-teradata.pdf

file:///C:/tmp/rre-in-teradata.pdf

Page: 1 of 13 80%

REVOLUTION ANALYTICS

WHITE PAPER

Maximizing the Value of Big Data

Introducing RRE for the Teradata Database

Bill Jacobs
Director, Product Marketing, Revolution Analytics

April 2014



“Big data” doesn't mean a “big trouble”

Step-by-Step Guide to Setti... x +

www.rdatamining.com/big-data/r-hadoop-setup-guide

RDM RDataMining.com: R and Data Mining

Search this site

- Home
- News
- Course & Seminar News
- Job News
- Resource News
- Training
- Documents
 - Introduction to Data Mining with R
 - R Reference Card for Data Mining
 - R and Data Mining: Examples and Case Studies
 - Introduction to Data Mining with R and Data Import/Export in R
 - Data Exploration and Visualization with R
 - Regression and Classification with R
 - Data Clustering with R
 - Association Rule Mining with R
 - Text Mining with R: Twitter Data Analysis
 - Time Series Analysis and Mining with R
- Examples
 - Data Exploration
 - Decision Trees
 - Random Forest

[Big Data >](#)

Step-by-Step Guide to Setting Up an R-Hadoop System

30 May 2014

This is a step-by-step guide to setting up an R-Hadoop system. I have tested it both on a single computer and on a cluster of computers. Note that this process is for Mac OS X and some steps or settings might be different for Windows or Ubuntu.

To install Hadoop on Windows, you can find detailed instructions at

- [Build and Install Hadoop 2.2 or newer on Windows](#)

Below is a list of software used for this setup.

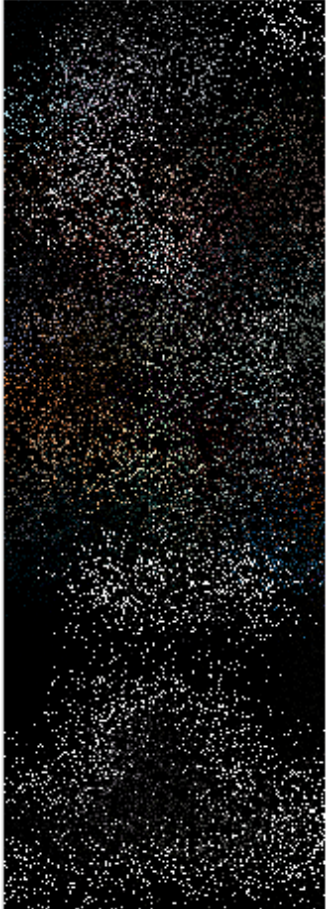
- OS and other tools:
 - Mac OS X 10.6.8, Java 1.6.0_65, Homebrew, thrift 0.9.0
- Hadoop and HBase:
 - Hadoop 1.1.2, HBase 0.94.17
- R and RHadoop packages:
 - R 3.1.0, rhdfs 1.0.8, rmr2 3.1.0, plymr 0.2.0, rhbase 1.2.0

This process should work with Hadoop 2.2 or above and newer versions of HBase as well, but I haven't tested it yet.

Homebrew is a missing package manager for Mac OS X, and it is needed for install *git*, *pkg-config* and *thrift*. For other operating systems, the equivalents to Homebrew are *apt-get* on Ubuntu and *yum* on CentOS.

By the way, two painful steps in this process are setting up HBase on Hadoop in cluster mode and installing *rhbase*. If you want to have a quick start or are not going to use HBase, you donot need to intall *thrift*, HBase or *rhbase*, and therefore can skip

- step 3 - Install HBase,





13 reasons why **you will** love GNU R

- I R is (extremely) cheap. In fact - it's free :)
- II R has (extremely) wide range of capabilities
- III R is (widely) supported by the world of science
- IV R is supported by the community
- V R is (increasingly) supported by the business
- VI R is able to read data in many formats
- VII Interoperability is easy to achieve
- VIII R is truly cross-platform
- IX R offers numerous ways of presenting data
- X There are many options to optimize the code
- XI R is able to handle large amount of data
- XII R has a set of fancy tools and IDEs**
- XII FDA accepted using R for drug trials!



Integrated Development Environments



1. RStudio

<http://www.rstudio.com/>

Welcome to RStudio - Open source and enterprise-ready professional software for R

Download RStudio

Discover Shiny



Powerful IDE for R

RStudio IDE is a powerful and productive user interface for R. It's free and open source, and works great on Windows, Mac, and Linux.

[Learn More >](#)



R Packages

Our developers and expert trainers are the authors of several popular R packages, including ggplot2, plyr, lubridate, and others.

[Learn More >](#)



Bring R to the web

Shiny is an elegant and powerful web framework for building interactive reports and visualizations using R — with or without web development skills.

[Learn More >](#)



Take control of your R code

RStudio is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management. [Click here to see more RStudio features.](#)

RStudio is available in open source and commercial editions and runs on the desktop (Windows, Mac, and Linux) or in a browser connected to RStudio Server or RStudio Server Pro (Debian/Ubuntu, RedHat/CentOS, and SUSE Linux).



Desktop

Run RStudio on your desktop

[RStudio Desktop >](#)



Server

Centralize access and computation

[RStudio Server >](#)



[CLICK HERE TO SEE ADDITIONAL FEATURES](#)



Firefox

Home - RStudio

www.rstudio.com

Google



[Products](#) [Resources](#) [Pricing](#) [About Us](#) [Blog](#)

A few of our professional fans.

RStudio is an active member of the R community. We believe free and open source data analysis software is a foundation for innovative and important work in science, education, and industry. The many customers who value our professional software capabilities help us contribute to this community.



STITCH FIX™



Bayer

ebay™

AMGEN



UC San Diego
SCHOOL OF MEDICINE



250 Northern Ave, Boston, MA 02210
Phone: 844-448-1212
Email: info@rstudio.com

[DMCA](#)
[Trademark](#)
[Support](#)
[ECCN](#)





RStudio

X:/projects/Test - RStudio

File Edit Code View Plots Session Build Debug Tools Help

test.rmd x test.pres.Rpres x Untitled1 x data x Untitled2* x

```
1 library(RODBC)
2 library(ggplot2)
3
4 con <- odbcConnectAccess("c:/tmp/11_aug.mdb")
5
6 param <- sqlQuery(con, "
7 SELECT q1.visitid,
8        (SELECT LEFT(dict.CLLABEL, 255)
9         FROM CODELST dict
10        WHERE dict.CODELST='cl_visit'
11              AND dict.CLCODE=CStr(q1.VISITID)) AS visit_name,
12        q1.siteid,
13        LEFT(q1.siteid, 3) AS country,
14        q1.subjid,
15        q1.XXXX AS param
16        FROM QS1 q1
17        WHERE q1.XXXX IS NOT NULL
18        ORDER BY VISITID, SITEID, SUBJID");
19
20 param$visit_name <- factor(param$visit_name, levels=c("Day 1",
21
22 p1 <- ggplot(param, aes(x=1:length(param), colour=visit_name,
23                        theme_bw() + geom_point(shape=1, aes(colour=visit_name))
24                        theme(axis.ticks = element_blank(), axis.text.x = elemen
25                        xlab("") + labs(title="some param raw data points") +
26                        scale_y_continuous(breaks=seq(0, 10, by=1)) + ylab("Some
27                        geom_boxplot(alpha=1/3, colour=I("#3366FF"), show_guide=
28                        facet_wrap(~visit_name, scales="free_x") + labs(colour="
29
30
```

Environment History SVN Presentation x

Name	Type	Length	Size	Value
con	RODBC	1	1.6 KB	Class 'RODBC' atomic [...]
p1	gg	9	71.1 KB	List of 9
param	data.frame	6	49 KB	1682 obs. of 6 vari...
test	function	1	5.9 KB	function (x)

Files Plots Packages Help Viewer

Zoom Export Save Plot as Image... Save Plot as PDF... Copy Plot to Clipboard... Clear All

Visit

- Day 1
- Week 4
- Week 8
- Week 12
- Week 16
- Week 20
- Week 24
- Week 48
- Early Termination

Console X:/projects/Test/

```
+ geom_boxplot(alpha=1/3, colour=I("#3366FF"), show_guide=FALSE, outlier.size=0) +
+ facet_wrap(~visit_name, scales="free_x") + labs(colour="visit")
> plot(p1)
>
```



RStudio – debugging

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Shows the R script `debug.r` with the following code:

```
1
2 f1 <- function(p) {
3   x <- 1
4   y <- 2
5
6   if(p > y) {
7     x <- 5
8   }
9 }
10
11 f1(6)
```
- Console:** Shows the execution of `debugSource` and the function call with debug output:

```
> debugSource('X:/projects/Test/debug.r', echo=TRUE)
f1 <- function(p) {
  x <- 1
  y <- 2

  if(p > y) {
    x <- 5
  }
}

f1(6)
called from: eval(expr, envir, enclos)
Browse[1]> n
debug at X:/projects/Test/debug.r#3: x <- 1
Browse[2]> n
debug at X:/projects/Test/debug.r#4: y <- 2
Browse[2]> |
```
- Environment Pane:** Shows the current environment with variables:

Name	Type	Length	Size	Value
p	promise	0	0 B	6
x	numeric	1	32 B	1
- Traceback Pane:** Shows the call stack:

```
f1(6) at debug.r:4
eval(expr, envir, enclos)
eval(expr, globalenv())
withVisible(eval(expr, envir = globalenv()))
doTryCatch(return(expr), name, parentenv, handler...)
tryCatchone(expr, names, parentenv, handlers[[1L]]...
```



RStudio – version control

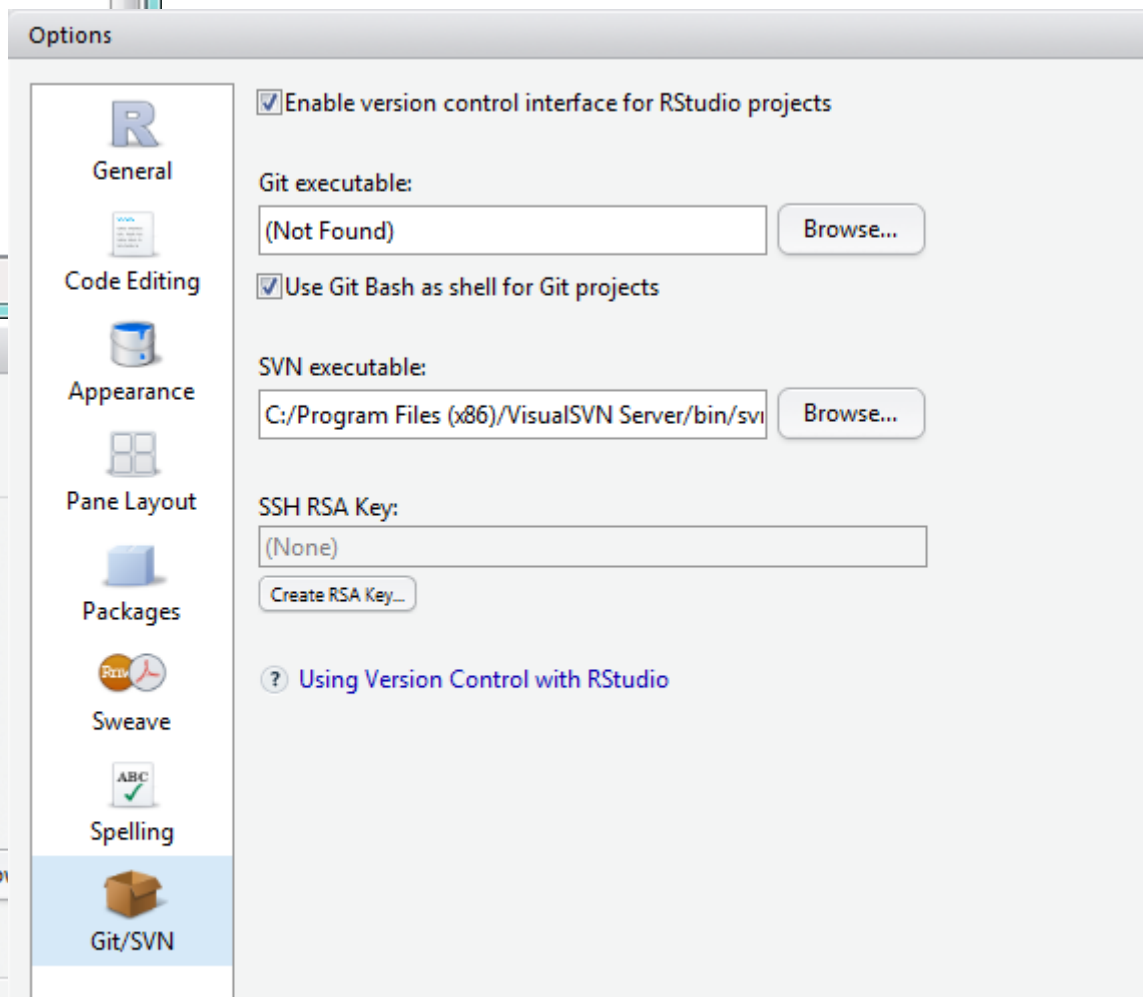
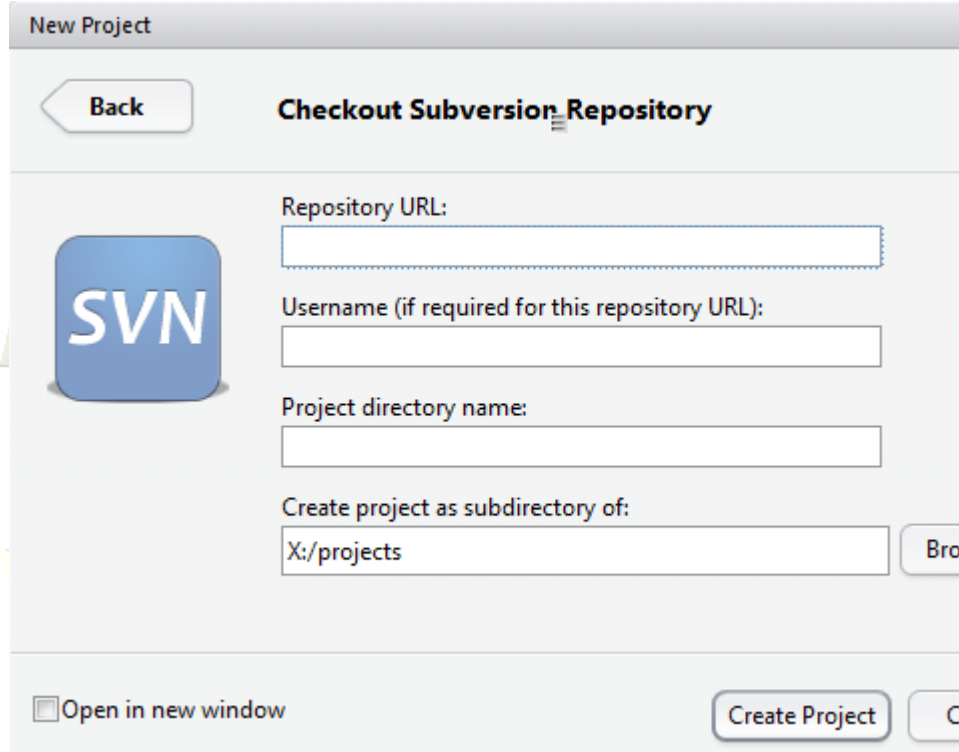
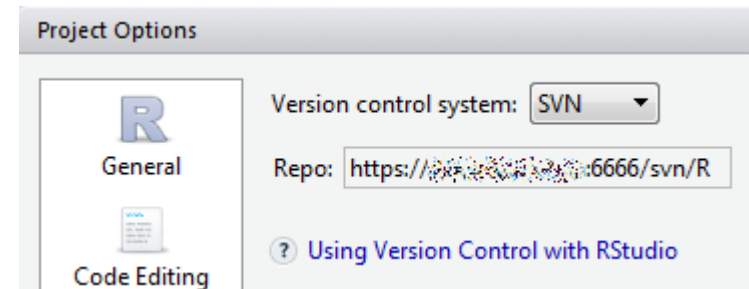
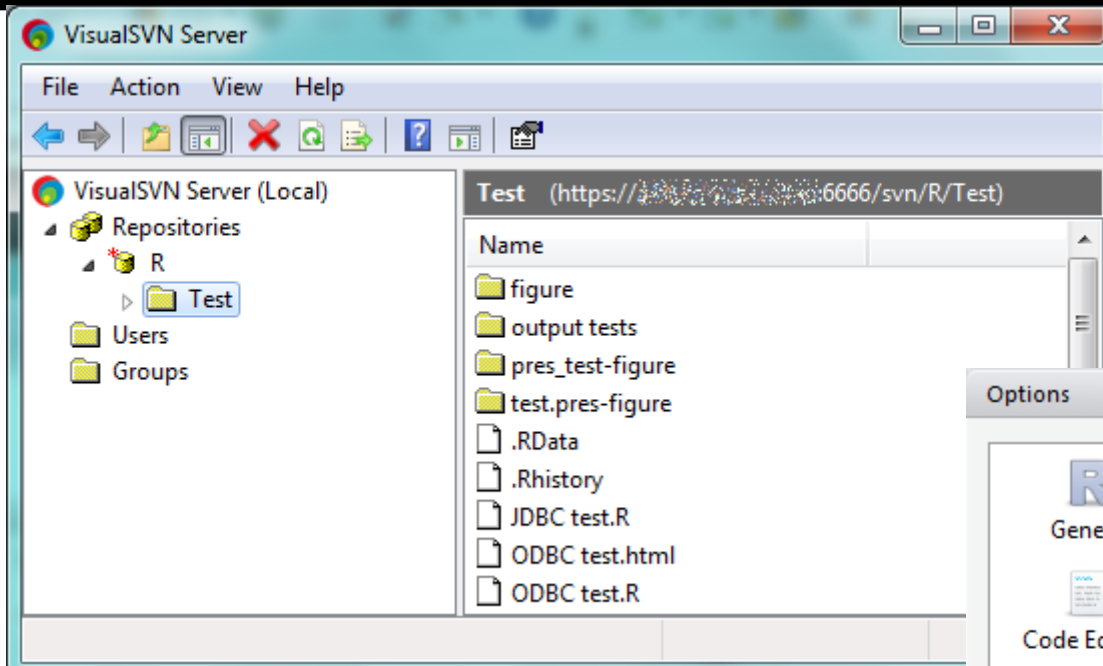
RStudio integrates seamlessly with version control systems like Git or Subversion (SVN). **Version control** is one of the **major requirements** for documentation management in clinical research.

The screenshot displays the RStudio interface with the following components:

- File Explorer:** Shows the project directory structure for 'Test'. The file list includes:
 - .Rproj.user
 - figure
 - output tests
 - pres_test-figure
 - test.pres-figure
 - .RData
 - .Rhistory
 - JDBC test.R
 - ODBC test.html
 - ODBC test.R
 - pres_test.md
 - pres_test.Rpres
 - test.html.html
 - test text document.txt
 - test.html
 - test.md
 - test.pres.md
 - test.pres.Rpres
 - test.rmd
 - Test.Rproj
 - testhtml2.html
 - testhtml2.Rhtml
- SVN Panel:** Shows the status of files in the working copy:
 - .RData (Modified)
 - .Rhistory (Modified)
 - ODBC test.R (Modified)
 - ODBC test.html (Unknown)
- Context Menu:** A 'More' menu is open over the SVN panel, showing options: Revert..., Ignore..., Resolve..., Update, Cleanup (highlighted), History, and Shell... A tooltip for 'Cleanup' reads: 'Recursively clean up the working copy (removing locks, etc)'.
- Console:** Shows the output of the 'view(param)' and 'summary(param)' functions.



RStudio – version control





RStudio – some taste...

Edit R Markdown Document Options

Output Format: **HTML**

Recommended format for authoring (you can switch to PDF or Word output anytime).

General Figures Advanced

Include table of contents

Syntax highlighting: default

Apply theme: space tango

Apply CSS file: X:/Biostat-Projects/...s.css

Number section headings

Choose R Installation

RStudio requires an existing installation of R in order to work. Please select the version of R to use.

Use your machine's default version of R64 (64-bit)

Use your machine's default version of R (32-bit)

Choose a specific version of R:

- [64-bit] C:\Program Files\R\R-3.0.3
- [32-bit] C:\Program Files\R\R-3.0.3
- [64-bit] C:\PROGRA~1\R\R-30~1.3
- [32-bit] C:\PROGRA~1\R\R-30~1.3

OK Cancel

File Edit Code View Plots Session Build Debug Tools Help

data x Untitled2* x param x testhtml2.Rhtml* x

```
1 library(RODBC)
2 library(ggplot2)
3
4 con <- odbcConnectAcc
5
6 param <- sqlQuery(con,
7 SELECT q1.visitid,
8 (SELECT LEFT(di
9 FROM CODELST
10 WHERE dict.COD
11 AND dict
12 q1.siteid,
13 LEFT(q1.siteid
```

Code Completion Tab

Go To Help

Go To Function Definition

Extract Function

Reindent Lines Ctrl+I

Reflow Comment Ctrl+Shift+/
Comment/Uncomment Lines Ctrl+Shift+C

Show code completions at the current cursor location

Run Chunks

Insert Chunk Ctrl+Alt+I

Jump To... Shift+Alt+J

Run Previous Code Ctrl+Alt+P

Run Current Chunk Ctrl+Alt+C

Run Next Chunk Ctrl+Alt+N

Run the current code chunk Ctrl+Alt+R

Go to File/Function

xml|

- xml (UnitsReporter.r)
- XMLDatasetLoader.r
- LoadXMLDataset (XMLDatasetLoader.r)

Untitled2* x param x testhtml2.Rhtml x pres_test.Rpres x

pres_test

```
1 pres_test
2
3 author:
4 date:
```

test.rmd

test.pres.Rpres

testhtml2.Rhtml



RStudio – syntax and validation

```

17 table {
18   font-family: "Lucida Console", "Courier New", Courier, monospace;
19   margin-top: 25px;
20   max-width: 95%;
21   border:0px none;
22   border-top:2px solid black;
23   border-bottom:2px solid black;
24   /*width:70% !important;*/
25   width:auto !important;
26   border-collapse: collapse;

```

21 border:0px none;
 22 border-top:2px solid black;
 Values of 0 shouldn't have units specified.

```

1 select * from table
2
3

```

table	local
from	local
select	local
s*	snippet
seq	snippet
addcol	snippet
colcom	snippet
tblcom	snippet

select * from \${1:table}

26 border-collapse: collapse
 27
 28 margin-left:auto !important;
 29 margin-right:auto !important;
 Expected RBRACE at line 28, col 15.

TeX
Markdown
XML
YAML
DCF
Shell
HTML
CSS
JavaScript
C/C++
Python
SQL
R Markdown

R Script
R Markdown
R Sweave
R HTML
R Presentation
Rd File
Text File
TeX
Markdown
XML
YAML
DCF
R Markdown

```
border-collapse: collapse;
```

bxz:bb	snippet
bxz	snippet
bxsh:n	snippet
bxsh	snippet
bxsh+	snippet
b:a	snippet
b	snippet
bd:n	snippet

bxsh+
 box-shadow: \${1:0} \${2:0} \${3:0} #\${4:000};

```

77 # Level 1 header
78 ## Level 2 header
79 ```{r}
80 result <- with(ds,
81   sqldf("SELECT Test, Name, Unit, OrigUnitN, COUNT(*)
82     FROM labs
83     GROUP BY Test, Name, Unit, OrigUnitN")
84   ...

```

expected ',' after expression

Knit HTML

```

...{r}
result <- with
  sqldf("S
  ...
Say hellow
...
### List of 1
...{r}
knitr::kable(
  with(ds,
    FROM 1
    INNER
    WHERE
  ...

```

Check Spelling

Not in dictionary: hellow

Ignore Ignore All Add

Change to: hello

hello
 hellos
 hallow
 hollow
 mellow

Change Change All

Cancel



RStudio - autocompletion

Console R Markdown x
X:/Biostat-Projects/UnitsReporter/

```
> knitr::|
```

- all_labels {knitr}
- all_patterns {knitr}
- asis_output {knitr}
- clean_cache {knitr}
- current_input {knitr}
- dep_auto {knitr}
- dep_prev {knitr}
- eclipse_theme {knitr}

clean_cache(clean = FALSE, path =
opts_chunk\$get("cache.path"))

If you remove or rename some cached code chunks, their original cache files will not be automatically cleaned. You can use this function to identify these possible files, and clean them if you are sure they are no longer needed.

Press F1 for additional help

Console R Markdown x
X:/Biostat-Projects/UnitsReporter/

```
> lists|
```

- vector
- dataFrame
- list
- fuction {list}

```
```{r Named chunk, results=}
knitr::kable(
 with(ds, sqldf("SELEC
 FROM labs l
 INNER JOIN CODELST
 WHERE l.OrigUnitN =
 ...
Comment:
```

- outDir
- dbName
- dbPath
- scientific\_10
- 'markup'
- 'asis'
- 'hold'
- 'hide'

```
It seems that **q/l** play
Chunk 6: Named chunk
```

Console R | read.csv(file, header = TRUE, sep = ",", quote = "\"", dec = ".",  
fill = TRUE, comment.char = "", ...)

```
> read.csv(fi|
```

- file =
- fill =
- fileEncoding =
- fifo {read.csv}
- file {read.csv}
- file.access {read.csv}
- file.append {read.csv}
- file.choose {read.csv}

**file**

the name of the file which the data are to be read from. Each row of the table appears as one line of the file. If it does not contain an *absolute* path, the file name is *relative* to the current working directory, `getwd()`. Tilde-expansion is performed where supported. This can be a compressed file (see `file`).

Alternatively, `file` can be a readable text-mode connection (which will be opened for reading if necessary, and if so `closed` (and hence

Press F1 for additional help



# RStudio – support for Rcpp

```
gibbs.cpp *
Source on Save
1 #include <Rcpp.h>
2 using namespace Rcpp;
3
4 // [[Rcpp::export]]
5 NumericMatrix gibbs(int N, int thin) {
6 NumericMatrix mat(N, 2);
7 double x = 0, y = 0;
8
9
10 for(int i = 0; i < N; i++) {
11 for(int j = 0; j < thin; j++) {
12 x = R::rgamma(3.0, 1.0 / (y * y + 4));
13 y = R::rnorm(1.0 / (x + 1), 1.0 / sqrt(2 * x + 2));
14 }
15 mat(i, 0) = x;
16 mat(i, 1) = y;
17 }
18
19 return(mat);
20 }
```

19:15 gibbs(int N, int thin): NumericMatrix C/C++

```
fibonacci.cpp *
Source on Save
1 #include <Rcpp.h>
2
3 // [[Rcpp::export]]
4 int fibonacci(const int x) {
5 if (x < 2)
6 return x;
7 else
8 return (fibonacci(x - 1)) + fibonacci(x - 2);
9 }
10
11 /** R
12 # Call the fibonacci function defined in C++
13 fibonacci(10)
14 */
```

```
convolve.cpp *
Source on Save
1 #include <Rcpp.h>
2 using namespace Rcpp;
3
4 // [[Rcpp::export]]
5 NumericVector convolveCpp(NumericVector a, NumericVector b) {
6 int na = a.size(), nb = b.size();
7 int nab = na + nb - 1;
8 NumericVector xab(nab);
9 for (int i = 0; i < na; i++)
10 for (int j = 0; j < nb; j++)
11 xab[i + j] += a[i] * b[j];
12 return xab;
13 }
14
```

7:3 C/C++

Console Source Cpp \*  
~/convolve.cpp Output Issues

~/convolve.cpp

- Line 7 'n' was not declared in this scope
- Line 13 expected ';' before '}' token

```
9
10 for (int i=0; i<n; i++)
11 y[i] = R::pnorm(x[i], 0.0, 1.0, 1);
12
13 too few arguments to function call, expected 5, have 4
14 }
```

```
11
12 for (int i=0; i<n; i++)
13 y[i] = R::pe[
14
15 pentagamma
16 pexp
17]
```

double pentagamma(double x)



# RStudio - menu

**File** Edit Code View Plots Session Build Debug Tools Help

- New File
  - R Script **Ctrl+Shift+N**
  - R Markdown...
  - Text File
  - C++ File
  - R Sweave
  - R HTML
  - R Presentation
  - R Documentation
- New Project...
- Open File... **Ctrl+O**
- Reopen with Encoding...
- Recent Files
- Open Project...
- Open Project in New Window...
- Recent Projects
- Save **Ctrl+S**
- Save As...
- Save with Encoding...
- Save All **Ctrl+Alt+S**
- Knit **Ctrl+Shift+K**
- Compile Notebook...
- Print...
- Close **Ctrl+W**
- Close All **Ctrl+Shift+W**
- Close Project
- Quit RStudio... **Ctrl+Q**

Edit Code View Plots Session Build

- Undo
- Redo
- Cut
- Copy
- Paste
- Folding
- Go to Line... **Alt+Shift+G**
- Find... **Ctrl+F**
- Find Next **F3**
- Find Previous **Shift+F3**
- Use Selection for Find **Ctrl+F3**
- Replace and Find **Ctrl+Shift+J**
- Find in Files... **Ctrl+Shift+F**
- Clear Console **Ctrl+L**

Tools Help

- Import Dataset
- Install Packages...
- Check for Package Updates...
- Version Control
- Shell...
- Project Options...
- Global Options...

Debug Tools Help

- Toggle Breakpoint **Shift+F9**
- Clear All Breakpoints...
- Execute Next Line **F10**
- Step Into Function **Shift+F4**
- Finish Function/Loop **Shift+F6**
- Continue **Shift+F5**
- Stop Debugging **Shift+F8**
- On Error
- Debugging Help

Session Build Debug Tools Help

- Interrupt R
- Restart R **Ctrl+Shift+F10**
- Terminate R...
- Set Working Directory
- Load Workspace...
- Save Workspace As...
- Clear Workspace...

- To Project Directory
- To Source File Location
- To Files Pane Location
- Choose Directory... **Ctrl+Shift+H**

- Message Only
- Error Inspector
- Break in Code



# RStudio - menu

Code	View	Plots	Session	Build	Debug	Tools
Back						Ctrl+F9
Forward						Ctrl+F10
Insert Section...						Ctrl+Shift+R
Jump To...						Alt+Shift+J
Go To File/Function...						Ctrl+.
Go To Help						
Go To Function Definition						
Extract Function						Ctrl+Alt+X
Extract Variable						Ctrl+Alt+V
Reflow Comment						Ctrl+Shift+/ /
Comment/Uncomment Lines						Ctrl+Shift+C
Insert Roxygen Skeleton						Ctrl+Alt+Shift+R
Reindent Lines						Ctrl+I
Reformat Code						Ctrl+Shift+A
Show Diagnostics						
Show Diagnostics (Project)						Ctrl+Alt+Shift+P
Run Line(s)						Ctrl+Enter
Re-Run Previous						Ctrl+Shift+P
Run Region						
Source						Ctrl+Shift+S
Source with Echo						Ctrl+Shift+Enter
Source File...						Ctrl+Shift+O

View	Plots	Session	Build	Debug	Tools
Hide Toolbar					
Zoom In					Ctrl++
Zoom Out					Ctrl+-
Switch to Tab...					Ctrl+Shift+.
Next Tab					Ctrl+F12
Previous Tab					Ctrl+F11
First Tab					Ctrl+Shift+F11
Last Tab					Ctrl+Shift+F12
Move Focus to Source					Ctrl+1
Move Focus to Console					Ctrl+2
Move Focus to Help					Ctrl+3
Show History					Ctrl+4
Show Files					Ctrl+5
Show Plots					Ctrl+6
Show Packages					Ctrl+7
Show Environment					Ctrl+8

Plots	Session	Build	Debug	Tools
Next Plot				Ctrl+Alt+F12
Previous Plot				Ctrl+Alt+F11
Zoom Plot...				
Save as Image...				
Save as PDF...				
Copy to Clipboard...				
Remove Plot...				
Clear All...				

Build	Debug	Tools	Help
Load All			Ctrl+Shift+L
Build and Reload			Ctrl+Shift+B
Clean and Rebuild			
Clean All			
Test Package			Ctrl+Shift+T
Check Package			Ctrl+Shift+E
Build Source Package			
Build Binary Package			
Document			Ctrl+Shift+D
Stop Build			
Configure Build Tools...			

Run From Beginning To Line	Ctrl+Alt+B
Run From Line to End	Ctrl+Alt+E
Run Function Definition	Ctrl+Alt+F
Run Code Section	Ctrl+Alt+T
Run All	Ctrl+Alt+R



# RStudio – the notebook (RMarkdown)

```
RStudio: Preview HTML
Preview: X:/projects/Test/test_notebook.html
s.Name + " " + s.Surname AS Subject,
rr.Time,
rr.RR_SYS, rr.RR_DIA, rr.BPM
FROM Subject s
INNER JOIN RR rr ON s.ID = rr.SubjID"

result <- sqlQuery(con, query)
odbcClose(con)

result$Time <- as.POSIXct(strptime(result$Time, "%Y-%m-%d %H:%M"))
(sOlszewski <- subset(result, ID==1))

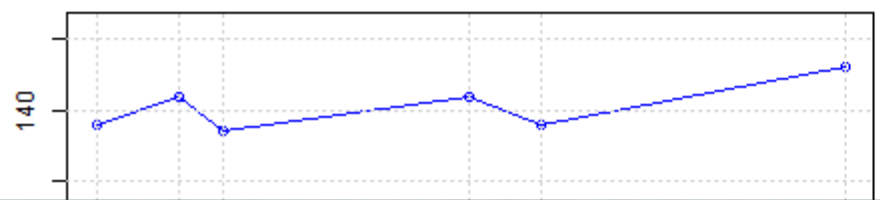
ID Subject Time RR_SYS RR_DIA BPM
1 1 Adrian Olszewski 2014-01-02 10:10:00 138 78 76
2 1 Adrian Olszewski 2014-01-02 15:30:00 142 82 99
3 1 Adrian Olszewski 2014-01-02 18:25:00 137 78 82
4 1 Adrian Olszewski 2014-01-03 10:45:00 142 88 87
5 1 Adrian Olszewski 2014-01-03 15:30:00 138 74 76
6 1 Adrian Olszewski 2014-01-04 11:30:00 146 91 91

par(mar=c(6, 4, 4, 2) + 0.5)
with(sOlszewski, plot(Time, RR_SYS, col="blue", type="o", ylim=c(60, 150), xaxt="n", main="Blood pressure of:
Olszewski Adrian, #1", ylab="SYS / DIA", xlab=""))
with(sOlszewski, lines(Time, RR_DIA, col="magenta", type="o"))
title(xlab='Time', line=5)
grid(nx=NA, ny=NA)
Sys.setlocale(category = "LC_TIME", locale="C")

[1] "C"
```

```
tck <- axis(1, at=sOlszewski$Time, labels=F)
text(tck, par("usr")[3], labels=format(sOlszewski$Time, "%b-%d %H:%M"), srt=50, xpd=TRUE, adj=c(1,2),
cex=0.8)
abline(v=tck, col = "lightgray", lty = "dotted", lwd = par("lwd"))
abline(h=axis(2, at=seq(60, 150, 10), labels=F), col = "lightgray", lty = "dotted", lwd = par("lwd"))
legend("bottomright", c("SYS", "DIA"), col=c("blue", "magenta"), lty=c(1,1))
```

Blood pressure of: Olszewski Adrian, #1



Compile Notebook from R Script

A Notebook is a standalone HTML file that contains the code and output from your R script.

Title (optional):  
test\_notebook.R

Author (optional):  
aaolsz

Notebook type: ?  
knitr::spin

Compile Cancel





# RStudio Server

## RStudio Server

	Open Source Edition	Professional Edition
Overview	<ul style="list-style-type: none"><li>• Access via a web browser</li><li>• Move computation closer to the data</li><li>• Scale compute and RAM centrally</li></ul>	<ul style="list-style-type: none"><li>• All of the features of open source; plus:</li><li>• Administrative Tools</li><li>• Enhanced Security and Authentication</li><li>• Metrics and Monitoring</li><li>• Advanced Resource Management</li></ul>
Documentation	<a href="#">Getting Started with RStudio Server</a>	<a href="#">RStudio Server Professional Admin Guide</a>
Support	Community forums only	<ul style="list-style-type: none"><li>• Priority Email Support</li><li>• 8 hour response during business hours (ET)</li></ul>
License	AGPL v3	<a href="#">RStudio License Agreement</a>
Pricing	Free	\$9,995/server/year <a href="#">Academic</a> and <a href="#">Small Business</a> discounts available
	<a href="#">DOWNLOAD RSTUDIO SERVER</a>	<a href="#">DOWNLOAD FREE RSTUDIO SERVER PRO EVAL</a>

[Purchase Now](#) [Contact Sales](#) [Learn More](#)



# RStudio Server

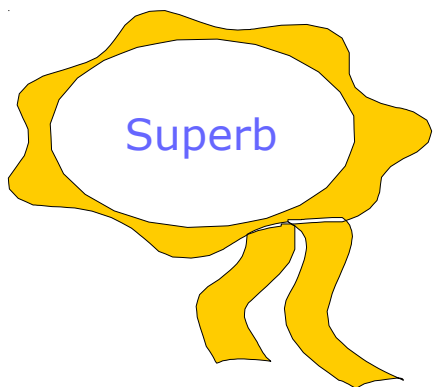
On my... pocket HTC :)

The screenshot displays the RStudio Server interface on a mobile device. The browser address bar shows the URL `ec2-54-213-155-143.us-west-2.compute.amazona...`. The RStudio interface includes a menu bar (File, Edit, Code, View, Plots, Session, Project, Build, Tools, Help), a console with the command `> hist(rbeta(100, 10, 1))`, and a plot window showing a histogram titled "Histogram of rbeta(100, 10, 1)". The histogram shows a distribution of values between 0.6 and 1.0, with a peak frequency of approximately 40.

Value Range	Frequency
0.60 - 0.65	1
0.65 - 0.70	5
0.70 - 0.75	4
0.75 - 0.80	6
0.80 - 0.85	5
0.85 - 0.90	14
0.90 - 0.95	24
0.95 - 1.00	40



# R is somewhat ascetic, isn't it?



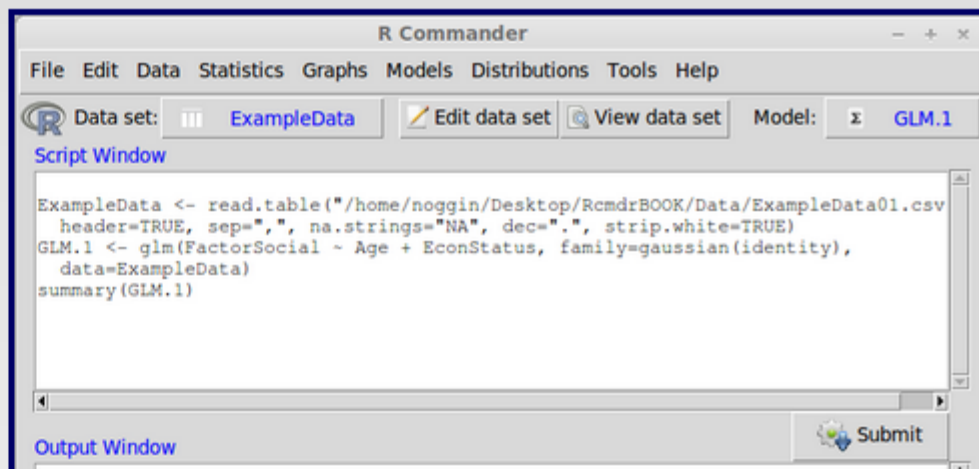
## 2. Rcommander

<http://www.rcommander.com>

## Rcommander a graphical interface for R

### R commander (Rcmdr)

R provides a powerful and comprehensive system for analysing data and when used in conjunction with the R-commander (a graphical user interface, commonly known as Rcmdr) it also provides one that is easy and intuitive to use. Basically, R provides the engine that carries out the analyses and Rcmdr provides a convenient way for users to input commands. The Rcmdr program enables analysts to access a selection of commonly-used R commands using a simple interface that should be familiar to most computer users. It also serves the important role of helping users to implement R commands and develop their knowledge and expertise in using the command line --- an important skill for those wishing to exploit the full power of the program.





# RCommander

R Commander

File Edit Data Statistics Graphs Models Distributions Tools Help

Data set: **data** Edit data set View data set Model: **GLM.1**

R Script R Markdown

```
GLM.1 <- glm(Y ~ X1 + X2 + bs(X3, df=5), family=gaussian(identity), data=data)
summary(GLM.1)
```

data

	Y	X1	X2	X3	X4
45	11.9081663	1.75384672	4.429748	13.13053	A
46	5.6032840	2.93563879	6.070810	11.19974	A
47	3.1406395	2.04441250	6.300311	11.58215	A
48	8.4691229	2.88742110	4.317648	11.17634	A
49	2.3783664	-0.05447750	5.158012	12.10283	A
50	5.1963735	0.97917484	5.051698	12.78101	A
51	7.2573921	2.46548105	5.455997	13.41225	B
52	7.7366457	3.28434256	5.589160	12.51206	B

Generalized Linear Model

Enter name for model: GLM.3

Variables (double-click to formula)

X1  
X2  
X3  
X4 [factor]  
Y

Model Formula

Operators (click to formula): + \* / %in% - ^ ( )

Splines/Polynomials: (select variable and click) B-spline natural spline orthogonal polynomial raw polynomial df for splines: 5 deg. for polynomials: 2

Y ~ X1 + X2 + bs(X3, df = 5)

Subset expression

<all valid cases>

Family (double-click to select) Link function

gaussian identity  
binomial inverse  
poisson log  
Gamma  
inverse.gaussian  
quasibinomial

Help Reset OK Cancel Apply

Output

```
> summary(GLM.1)
```

Call:

```
glm(formula = Y ~ X1 + X2 + bs(X3, df = 5), family =
 data = data)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-2.17524	-0.78511	0.05095	0.63205	2.40555

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	16.10279	0.93855	17.157	< 2e-16
X1	2.02949	0.07525	26.971	< 2e-16
X2	-3.08902	0.09956	-31.025	< 2e-16
bs(X3, df = 5)1	-0.02892	1.36739	-0.021	0.9832
bs(X3, df = 5)2	2.11650	0.88694	2.386	0.0191

Messages

```
[9] NOTE: The dataset data has 100 rows and 5 columns.
[10] WARNING: There is only one model in memory.
```



# R Commander

### Enter Two-Way Table

Table Statistics

Number of Rows:  3

Number of Columns:  4

Enter counts:

	1	2	3	4
1				
2				
3				

Help Reset OK

### Enter Two-Way Table

Table Statistics

Compute Percentages

- Row percentages
- Column percentages
- Percentages of total
- No percentages

Hypothesis Test

- Chi-square test of independence
- Components of chi-square statistic
- Print expected frequencies
- Fisher's exact test

Help Reset OK

### Scatterplot Matrix

Data Options

Graph title:

On Diagonal

- Density plots
- Histograms
- Boxplots
- One-dimensional scatterplots
- Normal QQ plots
- Nothing (empty)

Other Options

- Least-squares lines
- Smooth lines
- Show spread

Number of points to identify in each panel and group:

Span for smooth:

Help Reset OK Cancel Apply

### Plot Means

Data Options

Error Bars

- Standard errors
- Standard deviations
- Confidence intervals Level of confidence:
- No error bars

Plot Labels

x-axis label:

y-axis label:

Graph title:

Help Reset OK Cancel Apply

### Levene's Test

Factors (pick one or more):

Response Variable (pick one):

Center

- median
- mean

Help Reset OK Cancel Apply

Survival Regression Model

Enter name for model:

Time or start/end times (select one or two) | Event indicator (select one or none)

age	age
inst	inst
meal.cal	meal.cal
pat.karno	pat.karno
ph.ecog	ph.ecog
ph.karno	ph.karno

Strata (select zero or more) | Clusters (optional)

Distribution:  Weibull,  Exponential,  Gaussian,  Logistic,  Log-normal,  Log-logistic

Robust Standard Errors:  Default,  Yes,  No

Type of Censoring:  Default,  Right,  Left,  Interval,  Counting,  Interval type

Variables (double-click to formula)

age
inst
meal.cal
pat.karno
ph.ecog
ph.karno

~ sex + age

Subset expression: <all valid cases>

Create regular 2-level design ...

Base Settings | Factor Details | Export

Name of new design:  Tab Help

Size and randomization

Number of runs	<input type="text" value="8"/>	<input checked="" type="checkbox"/> Specify nruns
Number of factors	<input type="text" value="4"/>	
Number of center points	<input type="text" value="0"/>	
Number of blocks	<input type="text" value="1"/>	<input type="checkbox"/> blocks may be aliased with 2fis
Replications	<input type="text" value="1"/>	<input type="checkbox"/> Repeat only

You normally do not need to change randomization settings

Seed for randomization:   Randomization

Design properties

Minimum resolution:  Show available designs

NOTE: affects design generation for MaxC2 choice OR unspecified number of runs only

MA (Maximum resolution and minimum aberration)

MaxC2 (Maximum number of clear 2fis)

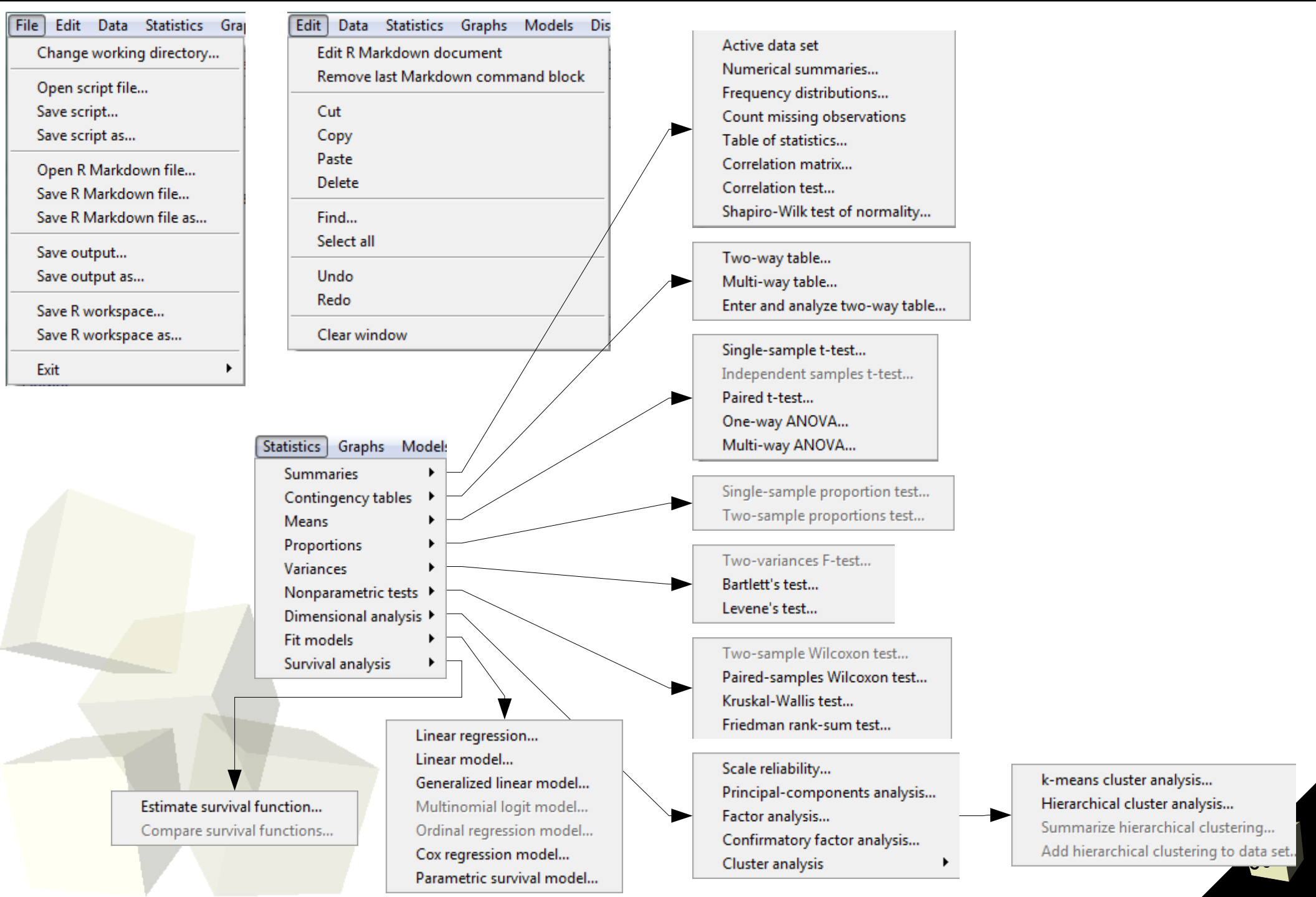
Show best 10 designs for 4 factors in 8 runs  
The menu remains open, fetch it back after looking at designs

Help OK Cancel  Activate Special Choices

Store form Load form Reset form



# R Commander





# RCommander

Models Distributions Tools Help

- Select active model...
- Summarize model
- Add observation statistics to data...

---

- Akaike Information Criterion (AIC)
- Bayesian Information Criterion (BIC)
- Stepwise model selection...
- Subset model selection...

---

- Confidence intervals...
- Hypothesis tests ▶

---

- Numerical diagnostics ▶
- Graphs ▶

ANOVA table...

- Compare two models...
- Linear hypothesis...

- Variance-inflation factors
- Breusch-Pagan test for heteroscedasticity...
- Durbin-Watson test for autocorrelation...
- RESET test for nonlinearity...
- Bonferroni outlier test

- Basic diagnostic plots
- Residual quantile-comparison plot...
- Component+residual plots
- Added-variable plots
- Influence plot
- Effect plots

- Cox-model survival function...
- Plot terms in Cox model
- Plot survival-regression dfbetas
- Plot survival-regression dfbeta
- Plot null Martingale residuals
- Cox-model partial-residual plots

Graphs Models Distributions

- Color palette...

---

- Index plot...
- Histogram...
- Density estimate...
- Stem-and-leaf display...
- Boxplot...
- Quantile-comparison plot...

---

- Scatterplot...
- Scatterplot matrix...
- Line graph...
- XY conditioning plot...
- Plot of means...
- Strip chart...

---

- Bar graph...
- Pie chart...

---

- 3D graph ▶
- Save graph to file ▶

3D scatterplot...

- Identify observations with mouse
- Save graph to file

- as bitmap...
- as PDF/Postscript/EPS...
- 3D RGL graph...

Distributions Tools Help

- Continuous distributions ▶
- Discrete distributions ▶

Normal distribution ▶	Normal quantiles...
t distribution ▶	Normal probabilities...
Chi-squared distribution ▶	Plot normal distribution...
F distribution ▶	Sample from normal distribution...
Exponential distribution ▶	
Uniform distribution ▶	
Beta distribution ▶	
Cauchy distribution ▶	
Logistic distribution ▶	
Lognormal distribution ▶	
Gamma distribution ▶	
Weibull distribution ▶	
Gumbel distribution ▶	

Distributions Tools Help

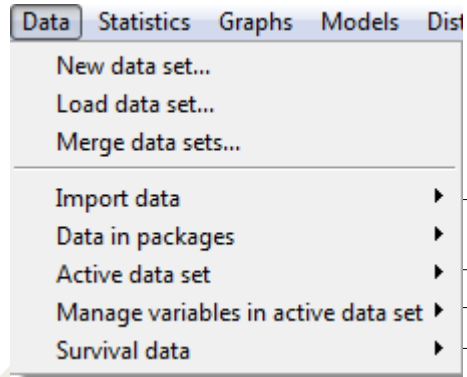
- Continuous distributions ▶
- Discrete distributions ▶

Binomial distribution ▶	Binomial quantiles...
Poisson distribution ▶	Binomial tail probabilities...
Geometric distribution ▶	Binomial probabilities...
Hypergeometric distribution ▶	Plot binomial distribution...
Negative binomial distribution ▶	Sample from binomial distribution...





# RCommander



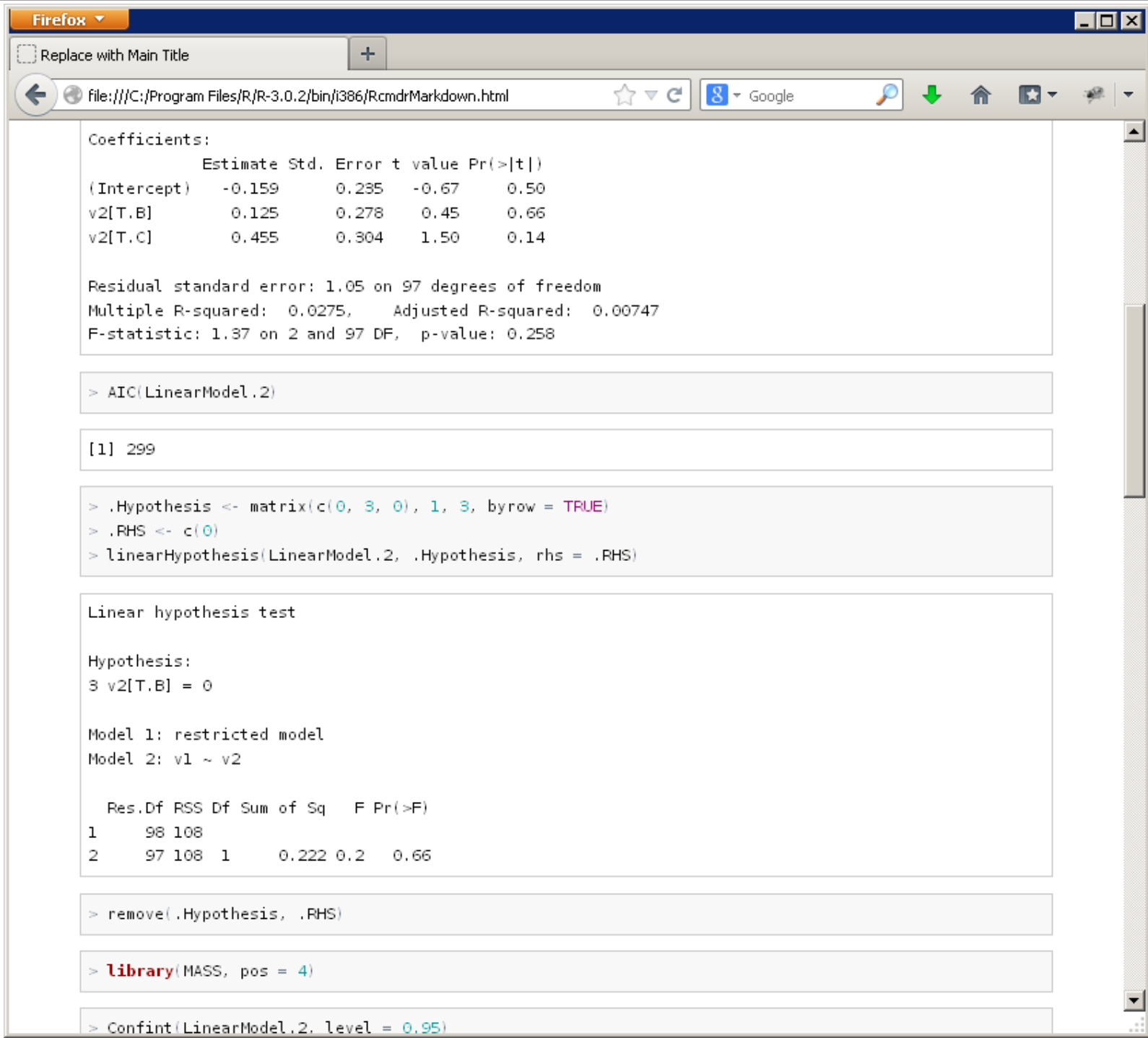
from text file, clipboard, or URL...  
from SPSS data set...  
from SAS xport file...  
from Minitab data set...  
from STATA data set...  
from Excel, Access or dBase data set...

Select active data set...  
Refresh active data set  
Help on active data set (if available)  
Variables in active data set  
Set case names...  
Subset active data set...  
Aggregate variables in active data set...  
Remove row(s) from active data set...  
Stack variables in active data set...  
Remove cases with missing data...  
Save active data set...  
Export active data set...

Recode variables...  
Compute new variable...  
Add observation numbers to data set  
Standardize variables...  
Convert numeric variables to factors...  
Bin numeric variable...  
Reorder factor levels...  
Define contrasts for a factor...  
Rename variables...  
Delete variables from data set ...

Survival data definition...  
Convert wide to long data...  
Convert variable to date...

# RCommander – the notebook (RMarkdown)



Firefox

Replace with Main Title

file:///C:/Program Files/R/R-3.0.2/bin/i386/RcmdrMarkdown.html

Google

```
Coefficients:
 Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.159 0.235 -0.67 0.50
v2[T.B] 0.125 0.278 0.45 0.66
v2[T.C] 0.455 0.304 1.50 0.14

Residual standard error: 1.05 on 97 degrees of freedom
Multiple R-squared: 0.0275, Adjusted R-squared: 0.00747
F-statistic: 1.37 on 2 and 97 DF, p-value: 0.258

> AIC(LinearModel.2)

[1] 299

> .Hypothesis <- matrix(c(0, 3, 0), 1, 3, byrow = TRUE)
> .RHS <- c(0)
> linearHypothesis(LinearModel.2, .Hypothesis, rhs = .RHS)

Linear hypothesis test

Hypothesis:
3 v2[T.B] = 0

Model 1: restricted model
Model 2: v1 ~ v2

 Res.Df RSS Df Sum of Sq F Pr(>F)
1 98 108
2 97 108 1 0.222 0.2 0.66

> remove(.Hypothesis, .RHS)

> library(MASS, pos = 4)

> Confint(LinearModel.2, level = 0.95)
```



# RCommander - plugins

## R commander Plugins (RcmdrPlugin)

A number of plugins are available that provide direct access to R packages through the Rcmdr interface. These plugins are installed in the same way as for other R packages (for information about installation see [www.UsingRcmdr.com](http://www.UsingRcmdr.com)) and can be loaded via the R-console or by using the Rcmdr menus **Tools, Load Rcmdr plugin(s)...**. There are currently 29 Plugins that provide support for specific analyses, graphics, books and teaching. Full information about the Plugins can be obtained by following the links provided in the table...

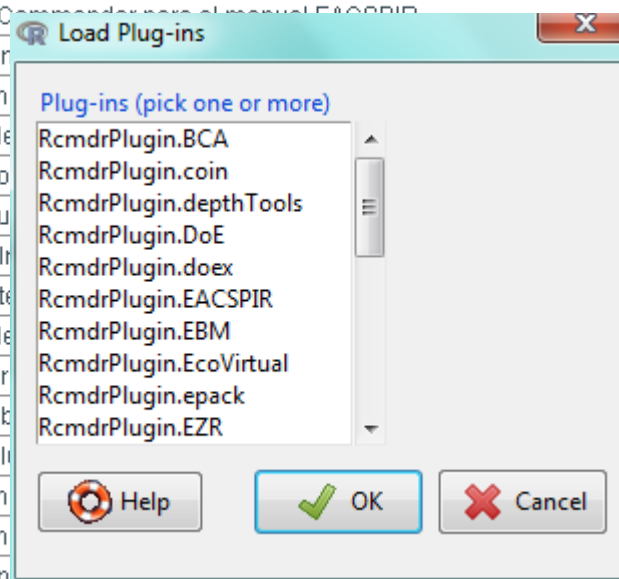
## Using the Rcmdr in conjunction with the Rstudio.

Rcmdr integrates with Rstudio

## Further information and resources.

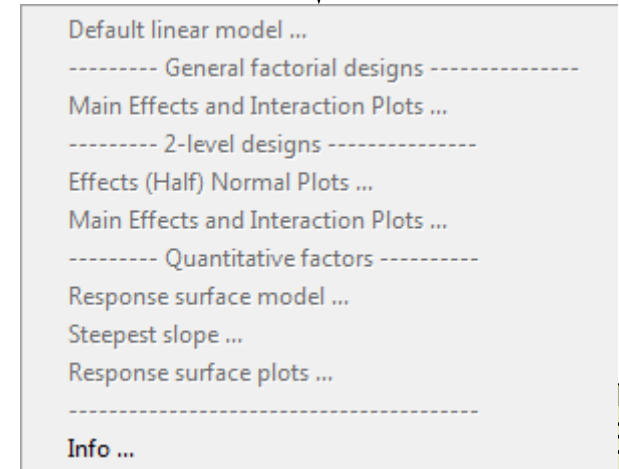
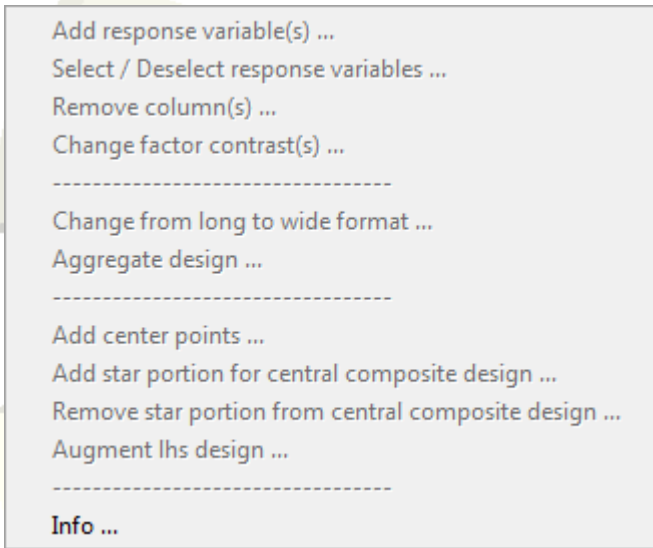
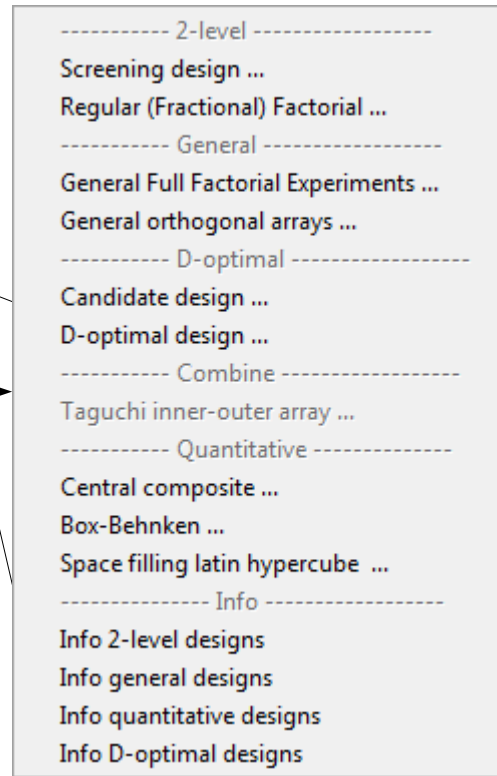
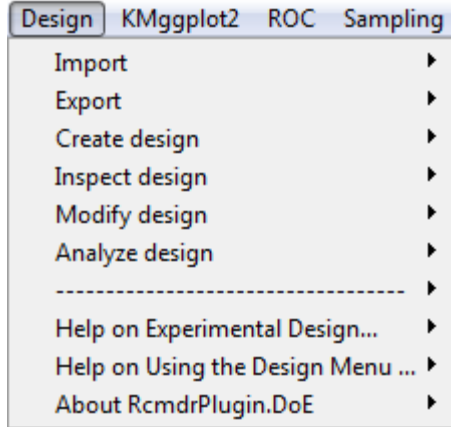
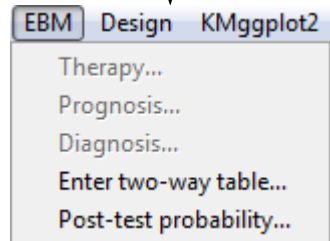
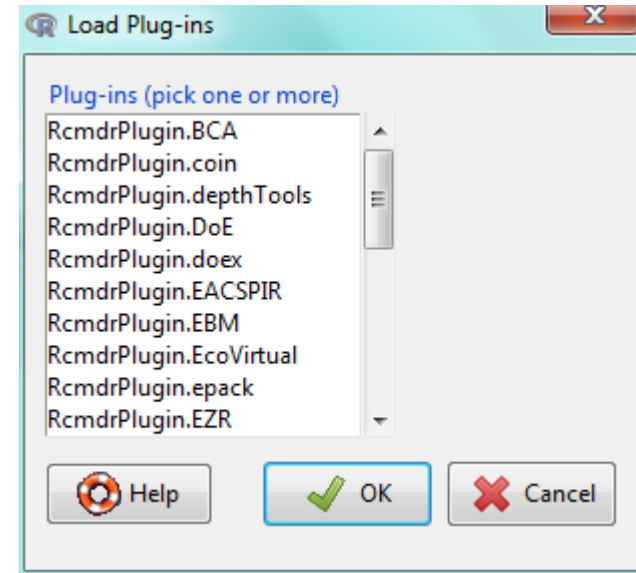
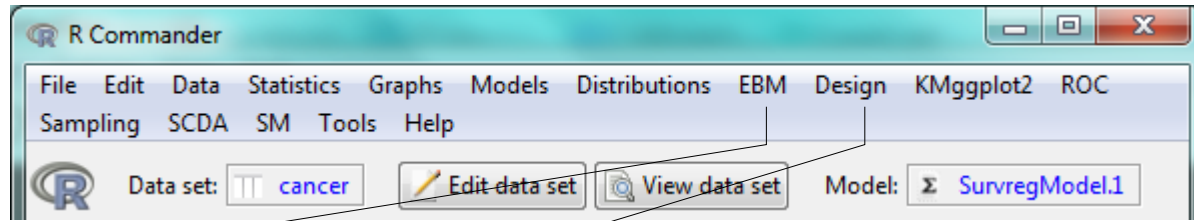
**Forthcoming book** (due out in 2013): **Data Analysis using R and the R commander**. Hutcheson, G. D. Sage Publications.

<a href="#">RcmdrPlugin.BCA</a>	Rcmdr Plug-In for Business and Customer Analytics
<a href="#">RcmdrPlugin.coin</a>	Rcmdr Coin Plug-In
<a href="#">RcmdrPlugin.depthTools</a>	R commander Depth Tools Plug-In
<a href="#">RcmdrPlugin.doBy</a>	Rcmdr doBy Plug-In
<a href="#">RcmdrPlugin.DoE</a>	R Commander Plugin for (industrial) Design of Experiments
<a href="#">RcmdrPlugin.doex</a>	Rcmdr plugin for Stat 4309 course
<a href="#">RcmdrPlugin.EACSPIR</a>	Plugin de R-Commander pour le manuel EACSPIR
<a href="#">RcmdrPlugin.EBM</a>	Rcmdr Eviden
<a href="#">RcmdrPlugin.epack</a>	Rcmdr plugin
<a href="#">RcmdrPlugin.EZR</a>	R Commande
<a href="#">RcmdrPlugin.HH</a>	Rcmdr suppo
<a href="#">RcmdrPlugin.IPSUR</a>	An IPSUR Plu
<a href="#">RcmdrPlugin.KMggplot2</a>	Rcmdr Plug-In
<a href="#">RcmdrPlugin.mosaic</a>	Adds menu ite
<a href="#">RcmdrPlugin.MPASTats</a>	R Commande
<a href="#">RcmdrPlugin.orloca</a>	orloca Rcmdr
<a href="#">RcmdrPlugin.plotByGroup</a>	Rcmdr plots b
<a href="#">RcmdrPlugin.qcc</a>	Rcmdr qcc Pl
<a href="#">RcmdrPlugin.qual</a>	Rcmdr plugin
<a href="#">RcmdrPlugin.SCDA</a>	Rcmdr plugin
<a href="#">RcmdrPlugin.seeg</a>	Rcmdr Plugin
<a href="#">RcmdrPlugin.SLC</a>	SLC Rcmdr Plug-in
<a href="#">RcmdrPlugin.SM</a>	Rcmdr Sport Management Plug-In
<a href="#">RcmdrPlugin.StatisticalURV</a>	Statistical URV Rcmdr Plug-In
<a href="#">RcmdrPlugin.steeppness</a>	Steeppness Rcmdr Plug-in
<a href="#">RcmdrPlugin.survival</a>	R Commander Plug-in for the survival Package
<a href="#">RcmdrPlugin.TeachingDemos</a>	Rcmdr Teaching Demos Plug-In
<a href="#">RcmdrPlugin.temis</a>	Graphical user interface providing an integrated text mining solution
<a href="#">RcmdrPlugin.UCA</a>	UCA Rcmdr Plug-in





# bio RCommander





# bio RCommander

The screenshot displays the bio RCommander software interface. At the top, the main menu bar includes File, Edit, Data, Statistics, Graphs, Models, Distributions, EBM, Design, KMggplot2, ROC, Sampling, SCDA, SM, Tools, and Help. Below the menu bar, the current data set is 'cancer' and the model is 'SurvregModel.1'. A dropdown menu is open under 'Design', showing options: SCVA, SCRT, and SCMA. SCRT is expanded to show 'Design your experiment' and 'Analyze your data'. Below this, a dialog box titled 'Number of possible assignments' is open, showing options for design type (AB Phase Design, ABA Phase Design, ABAB Phase Design, Completely Randomized Design) and test statistics (A-B, B-A, [A-B], PA-PB). The 'Randomization distribution' dialog box is also open, showing options for design type, test statistic, randomization distribution (Systematic or Monte Carlo), and data file selection. The 'Randomization distribution' dialog box has the following fields:

- Select the design type:** AB Phase Design, ABA Phase Design, ABAB Phase Design, Completely Randomized Design
- Select the test statistic:** A-B, B-A, [A-B], PA-PB (only for ABA and ABAB)
- For phase designs:** Minimum number of observations per phase: [input field]
- For alternating treatments designs:** Maximum number of consecutive administrations of the same condition: [input field]
- Select the randomization distribution:**  Systematic randomization distribution,  Monte Carlo randomization distribution
- Save the distribution?:**  NO,  YES
- Number of randomizations:** [input field]
- Selected location:** [input field]
- Select the data file:**  Use the active data set,  Use data from a text file [Select File]
- Selected file:** [input field]
- Select the start points file (only for MBD):** Selected file: [input field]
- Selected file:** [input field]

At the bottom of both dialog boxes are buttons for Help, OK, and Cancel.



# R is somewhat ascetic, isn't it?

## 3. RExcel

<http://www.statconn.com/products.html>

### Products

- [RExcel](#)
- [SWord for 32 Bit R and for 64 Bit R](#)
- [statconnDCOM for 32 Bit R and for 64 Bit R](#)
- [statconn.NET for 32 Bit R and for 64 Bit R](#)
- [statconnWS](#)
- [Pricing](#)

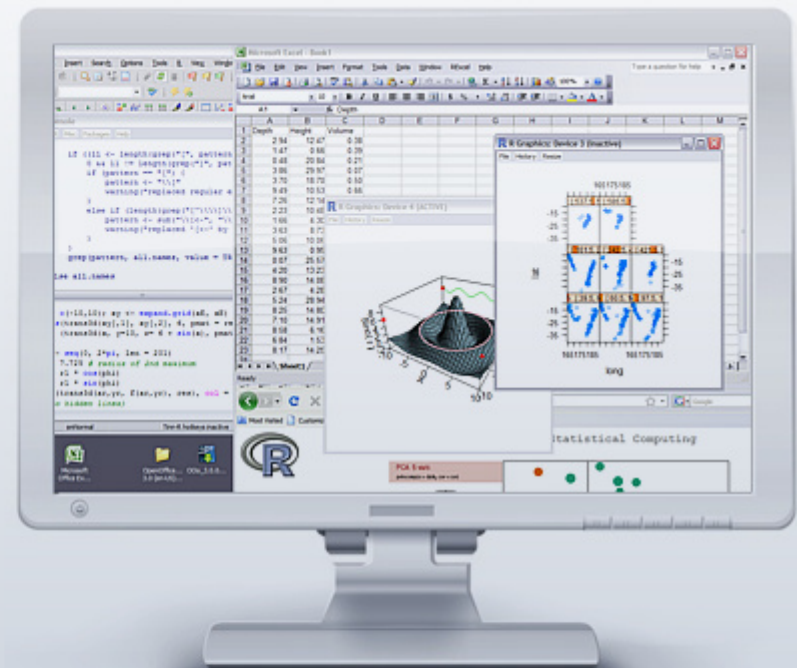
### R in Microsoft Excel: RExcel

*RExcel* integrates R seamlessly into Microsoft Excel. RExcel currently supports Microsoft Excel versions 2003, 2007 and 2010.

RExcel allows to use R from within Excel in different ways

- Transferring data between R and Excel interactively from within Excel, using Excel as data editor for data to be analyzed with R
- Using Excel as container (and simple code editor) for R commands, and running R commands from Excel
- Running R commands from Excel VBA macros allowing to create Excel based applications and Excel add-ins using R for other users
- Using R functions in Excel worksheets functions, integrating R functionality into the automatic recalculation engine of Excel

RExcel is used for introductory statistics courses at universities and colleges in many countries (USA, Austria, Germany, Japan, New Zealand ...). It is also used for custom Excel applications in finance, pharmaceutical industry, and chemical production industry.





# Excel + RExcel + RCommander

Zeszyt1 - Microsoft Excel

Narzędzia główne Wstawianie Układ strony Formuły Dane Recenzja Widok Dodatki

Interfejs Bluetooth R Excel


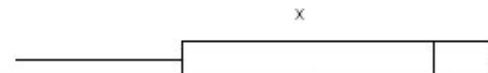
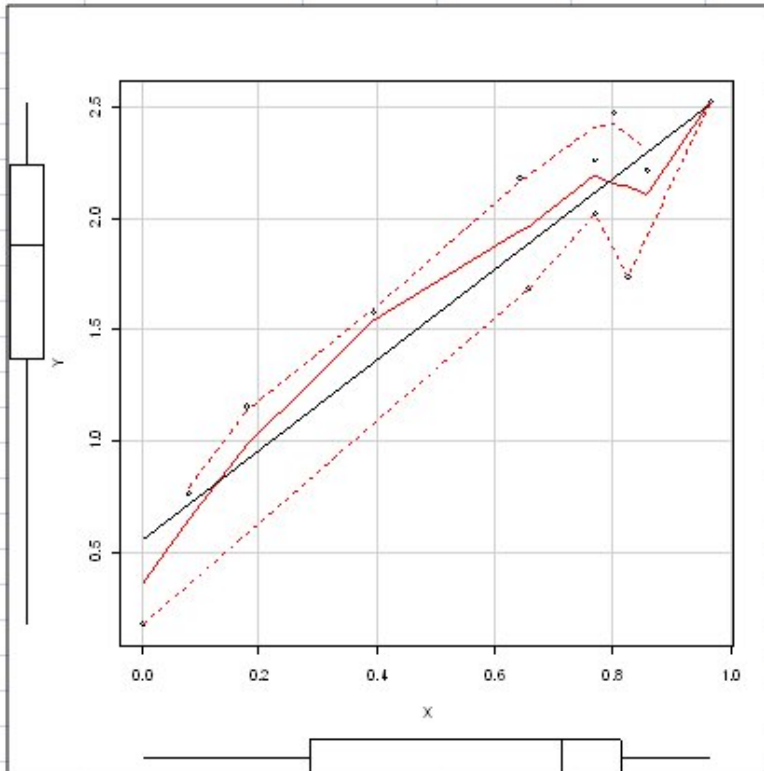
File Data Statistics Graphs Models Distributions Tools Help

Dataset: Regresja Model: RegModel.1

Polecenia menu Niestandardowe paski narzędzi

I25

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2		Dane do regresji		REval("RegModel.1[coefficients]")									
3		X	Y	<b>c(0.554263716743928, 2.03433984170854)</b>					> RegModel.1 <- lm(Y~X, data=Regresja)				
4		0,178147	1,154119										
5		0,769566	2,023968	REval("summary(lm(Y~X, data=Regresja))\$sigma")					> summary(RegModel.1)				
6		0,82564	1,738452	<b>0,276806354</b>					Call:				
7		0,857405	2,217537						lm(formula = Y ~ X, data = Regresja)				
8		0,768989	2,26635						Residuals:				
9		0,001861	0,172507						Min 1Q Median 3Q Max				
10		0,393118	1,580379						-0.49544 -0.12359 0.02505 0.22915 0.31711				
11		0,802951	2,473617						Coefficients:				
12		0,080386	0,760383						Estimate Std. Error t value Pr(> t )				
13		0,966064	2,527085						(Intercept) 0.5543 0.1668 3.324 0.0077 **				
14		0,64321	2,179882						X 2.0343 0.2529 8.044 1.12e-05 ***				
15		0,658135	1,686334						---				
16									Signif. codes: 0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1				
17									Residual standard error: 0.2768 on 10 degrees of freedom				
18									Multiple R-squared: 0.8661, Adjusted R-squared: 0.8528				
19									F-statistic: 64.71 on 1 and 10 DF, p-value: 1.122e-05				
20													
21													
22													
23													
24													
25													
26													
27													
28													
29													
30													



# Excel + RExcel + RCommander

FactoMineR menu options:

- Principal Component Analysis (PCA)
- Correspondence Analysis (CA)
- Multiple Correspondence Analysis (MCA)
- Multiple Factor Analysis (MFA)
- Hierarchical Multiple Factor Analysis (HMFA)
- Dual Multiple Factor Analysis (DMFA)
- Factor Analysis of Mixed Data (FAMD)
- General Procrustes Analysis (GPA)
- Scatter plot with additional variables
- Description of categories
- Hierarchical Clustering on Principal Components (HCPC)

	A	B	C	D	E	F
1						
2	Dane do regresji			REval("RegModel.1[coefficients"		
3	X	Y		c(0.554263716743928, 2.03433984		
4	0,178147	1,154119				
5	0,769566	2,023968		REval("summary(lm(Y~X, data=Regresja		
6	0,82564	1,738452		0,276806354		
7	0,857405	2,217537				

Put dataframe in R dialog box:

- Dataframe name in R: Regresja
- Get from R
- Get from Cell
- with rownames
- colorize source range
- make active in RCommander

	A	B	C	D	E	F
1						
2	Dane do regresji					
3	X	Y				
4	0,178147	1,154119				
5	0,769566	2,023968				
6	0,82564	1,738452				
7	0,857405	2,217537				
8	0,768989	2,26635				
9	0,001861	0,172507				
10	0,393118	1,580379				
11	0,802951	2,473617				
12	0,080386	0,760383				
13	0,966064	2,527085				
14	0,64321	2,179882				
15	0,658135	1,686334				
16						

Context menu options:

- Run code in Rcmdr
- Run code
- Get R Value
- Put R Var
- Get R DataFrame
- Put R DataFrame
- Rcmdr Get
- Get R Output
- Insert Current R Plot
- Name Range
- Prettyformat Numbers

	A	B	C	D	E	F
1						
2	Dane do					
3	X	Y				
4	0,178147					
5	0,769566					
6	0,82564					
7	0,857405					
8	0,768989					
9	0,001861					
10	0,393118					
11	0,802951					
12	0,080386					
13	0,966064					
14	0,64321					
15	0,658135					
16						





# R is somewhat ascetic, isn't it?

## 4. R Analytic Flow

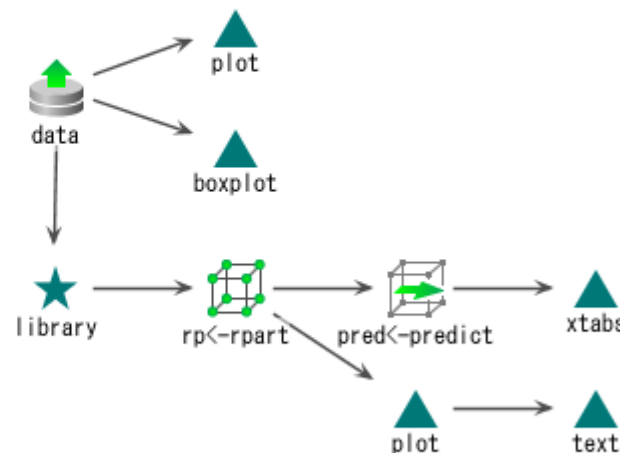
[http://www.ef-prime.com/products/ranalyticflow\\_en/](http://www.ef-prime.com/products/ranalyticflow_en/)

### Features



#### [Analysis Flow]

Processes of data analysis are described as a flowchart, and then conducted according to it. So once a flow is written, anyone can reproduce the result easily and precisely.





# R Analytic Flow

R AnalyticFlow - BostonAnalysis.rflow

File Edit Node Run Preferences Window Help

write.table(file = "prdata.csv", x = dat.pr, sep = ",", row.names = FALSE)

SampleAnalysis  
- .raf  
- script  
- BostonAnalysis.rflow  
- IrisAnalysis.rflow

library → data → attach → plot  
data.entry → Transform → Sampling → mod<-lm → mod.stp<-step → pred<-predict → write.table  
dat.pr<-data.frame → plot → hist

R Edit - write.table

Setting Property

Code

```
write.table(file = "prdata.csv", x = dat.pr, sep = ",", row.names = FALSE)
```

Assignment option

Evaluate only  
 Assign to: \_\_\_\_\_

Function

Parameter	Value
file	"prdata.csv"
x	dat.pr
sep	","
row.names	FALSE
append	FALSE
create	TRUE

OK Cancel Apply

R Graphics: Device 2 (ACTIVE)

pred

log(dat[test, "medv"])

R console

```
Command
- rm 1 0.4093 8.8595 -828.03
- nox 1 0.5709 9.0211 -823.46
- rad 1 0.6837 9.1339 -820.31
- ptratio 1 0.8154 9.2657 -816.69
- log.dis 1 1.2055 9.6557 -806.26
- crim 1 1.4940 9.9443 -798.81
- log.lstat 1 5.0428 13.4930 -721.60
> pred <- predict(mod.stp, newdata = dat[test,])
> plot(pred ~ log(dat[test, "medv"]))
>
```



# R Analytic Flow

R AnalyticFlow - \*New Analysis Flow

File Edit Node Run Preferences Window Help

File Function Hierarchy Object

RFlow

- .raf
- SampleAnalysis
- Tutorial

Zmienna <- 3

x<-runif

Skrypt od Jarka  
węzeł skryptu R do wykonania

if else

plot

parametry węzła prostego

R Edit - plot

Setting Property

Code

```
plot(1:Zmienna)
```

Assignment option

Evaluate only

Assign to:

Function

plot src ?

Parameter	Value
	1:Zmienna
x	
y	
...	

OK Cancel Apply

if

Condition: Zmienna == 1

IF ELSE

między IN i OUT można umieścić dowolny ciąg węzłów prostych i/lub złożonych, w tym kolejne warunki, pętle, itd.

in for out

for

Condition: i in 1:Zmienna

między IN i OUT można umieścić dowolny ciąg węzłów prostych i/lub złożonych, w tym kolejne warunki, pętle, itd.

in print out

R Graphics: Device 2 (ACTIVE)

File History Resize

R console

Command

```
[1] "Iteracja nr 6"
[1] "Iteracja nr 7"
[1] "Iteracja nr 8"
[1] "Iteracja nr 9"
[1] "Iteracja nr 10"
[1] "Iteracja nr 11"
[1] "Iteracja nr 12"
[1] "Iteracja nr 13"
[1] "Iteracja nr 14"
[1] "Iteracja nr 15"
[1] "Iteracja nr 16"
[1] "Iteracja nr 17"
[1] "Iteracja nr 18"
[1] "Iteracja nr 19"
[1] "Iteracja nr 20"
> plot(1:Zmienna)
>
```



# R Analytic Flow



R AnalyticFlow - Search\_and\_Replace.rflow

File Edit View Node Run Preferences Help

File Object Search Breakpoint Task

rp : rpart [14]

- frame : data.frame [5,9]
  - where : integer [150]
  - call : call [3]
  - terms : terms [3]
  - cptable : matrix [3,5]
  - method : character "class"
- parms : list [3]
  - prior : array [3]
  - loss : matrix [3,3]
  - split : numeric 1
- control : list [9]
- functions : list [3]
- numresp : integer 5L
- enlite : matrix [14 5]

R Console

```
> .clearObjects()
> data(iris)
> boxplot(Petal.Length ~ Species, data = iris, col
= 3, main = "Petal.Length")
> data(iris)
> Petal <- iris[, c("Petal.Length", "Petal.Width")]
> plot(Petal)
> data(iris)
> library(rpart)
> rp <- rpart(Species ~ ., iris)
> plot(rp, margin = 0.2, branch = 0.3)
> text(rp, fancy = T, all = T, use.n = T)
>
```

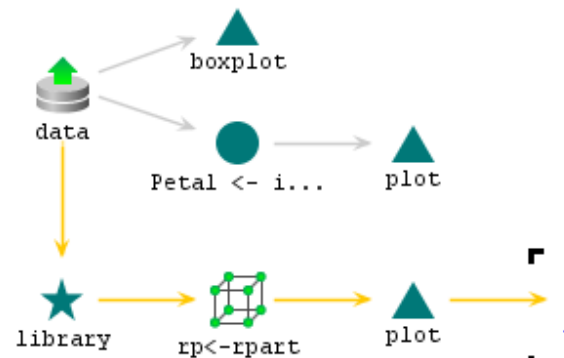
Top

Graphics:2 x Petal <- i... x data x rp<-rpart x text x

▲ text(rp, fancy = T, all = T, use.n = T)

You can search/replace the character strings in R scripts. Click on the "Search" tab at the upper-left side, or press Ctrl + F (Command + F on Mac OS X).

1. Find "iris" from the following flow.
2. Replace "Petal" to "Sepal".
3. Find "Sepal". Enable "Whole Word" and compare the results.



- Clear and Run Ctrl+E
- Run Ctrl+R
- Run Selected Nodes
- Add to Task
- Cache
- New Comment
- Set Breakpoint
- Box
- Copy R Script
- Export R Script...
- Open Enter
- Open in New Tab
- Cut Ctrl+X
- Copy Ctrl+C
- Remove Delete
- Properties



# R is somewhat ascetic, isn't it?

## 5. Rattle

<http://rattle.togaware.com/>

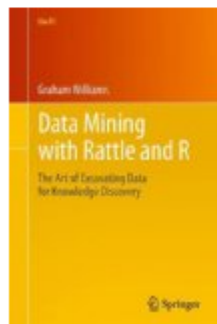
### Rattle: A Graphical User Interface for Data Mining using R

Version 2.6.27 release 148 dated 2013-12-23.

```
> install.packages("rattle", repos="http://rattle.togaware.com", type="source")
```

```
$ wget http://togaware.com.au/access/rattle_2.6.27.tar.gz
```

**Rattle** (the R Analytical Tool To Learn Easily) presents statistical and visual summaries of data, transforms data into forms that can be readily modelled, builds both unsupervised and supervised models from the data, presents the performance of models graphically, and scores new datasets.



[Errata](#)



[Brochure](#)



# Rattle your data

R Data Miner - [Rattle (audit.csv)] Rattle Version 2.6.6 [togaware.com](http://togaware.com)

Project Tools Settings Help

Execute New Open Save Report Export Stop Quit

Data Explore Test Transform Cluster Associate Model Evaluate Log

Source:  Spreadsheet  ARFF  ODBC  R Dataset  RData File  Library  Corpus  Script

Filename:  Separator:  Decimal:   Header

Partition  Seed:

Input  Ignore Weight Calculator:

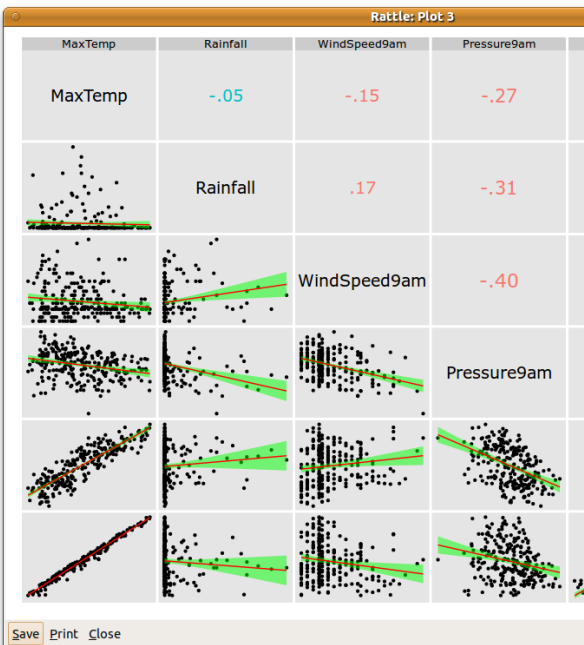
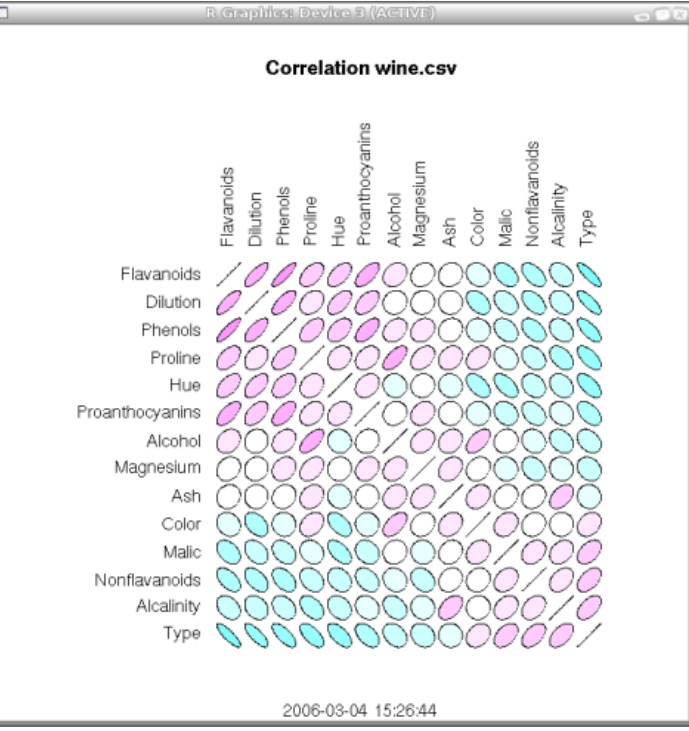
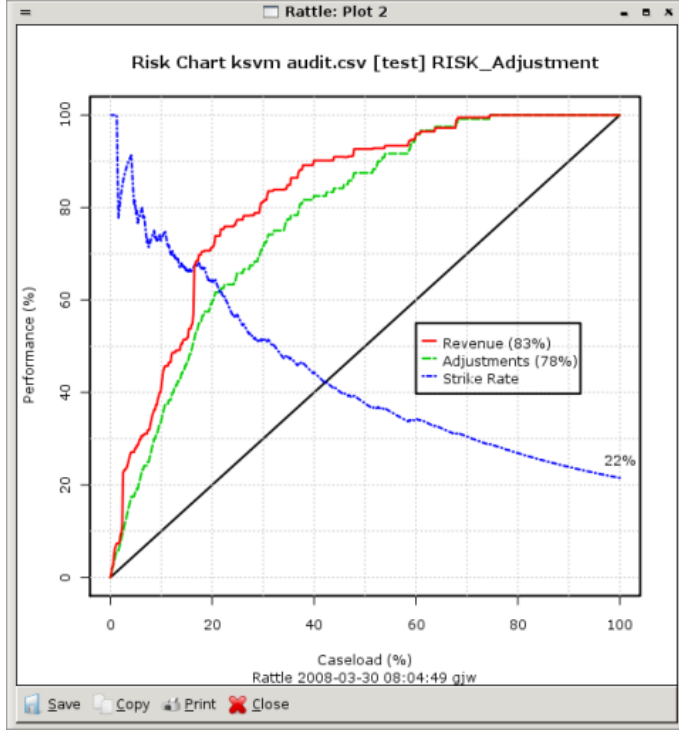
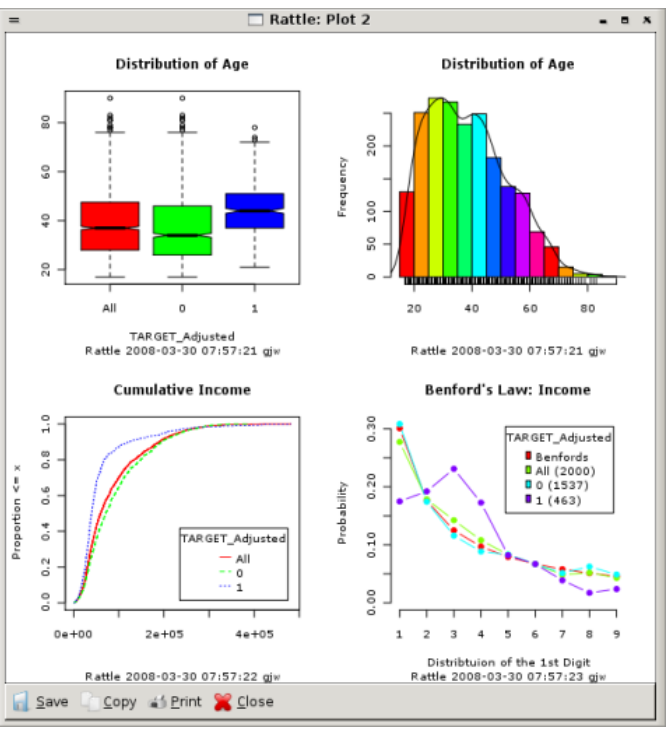
Target Data Type  
 Auto  Categorical  Numeric  Survival

No.	Variable	Data Type	Input	Target	Risk	Ident	Ignore	Weight	Comment
1	ID	Numeric	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 2000
2	Age	Numeric	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 67
3	Employment	Categorical	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 8 Missing: 100
4	Education	Categorical	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 16
5	Marital	Categorical	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 6
6	Occupation	Categorical	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 14 Missing: 101
7	Income	Numeric	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 2000
8	Gender	Categorical	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 2
9	Deductions	Numeric	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 41
10	Hours	Numeric	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 68
11	IGNORE_Accounts	Categorical	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Unique: 33 Missing: 43
12	RISK_Adjustment	Numeric	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 310
13	TARGET_Adjusted	Numeric	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 2

Roles noted. 2000 observations and 9 input variables. The target is TARGET\_Adjusted. Categorical 2. Classification models enabled.



# Rattle your data



**R Data Miner - [Rattle]**

Rattle Version 2.6.6 togaware.com

Project Tools Settings Help

Execute New Open Save Report Export Stop Quit

Data Explore Test Transform Cluster Associate Model Evaluate Log

Source:  Spreadsheet  ARFF  ODBC  RDataset  RData File  Library  Corpus  Script

Filename: (None) Separator: Decimal:  Header

Partition 70/15/15 Seed: 42 View Edit

Input  Ignore Weight Calculator: Target Data Type:  Auto  Categorical  Numeric  Survival

Welcome to Rattle (rattle.togaware.com).

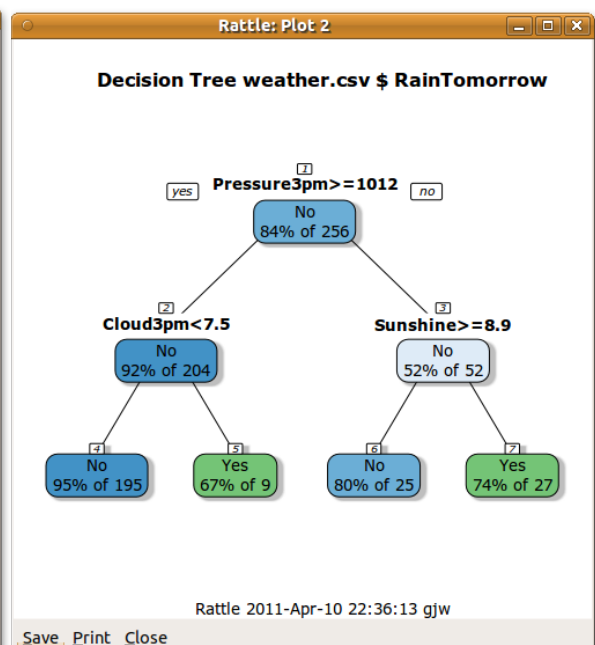
Rattle is a free graphical user interface for Data Mining, developed using R. R is a free software environment for statistical computing and graphics. Together they provide a sophisticated environments for data mining, statistical analyses, and data visualisation.

See the Help menu for extensive support in using Rattle. The Togaware Desktop Data Mining Survival Guide includes Rattle documentation and is available from datamining.togaware.com

Rattle is licensed under the GNU General Public License, Version 2. Rattle comes with ABSOLUTELY NO WARRANTY. See Help -> About for details.

Rattle Version 2.6.6. Copyright 2006-2011 Togaware Pty Ltd  
Rattle is a registered trademark of Togaware Pty Ltd

To Begin: Choose the data source, specify the details, then click the Execute button.



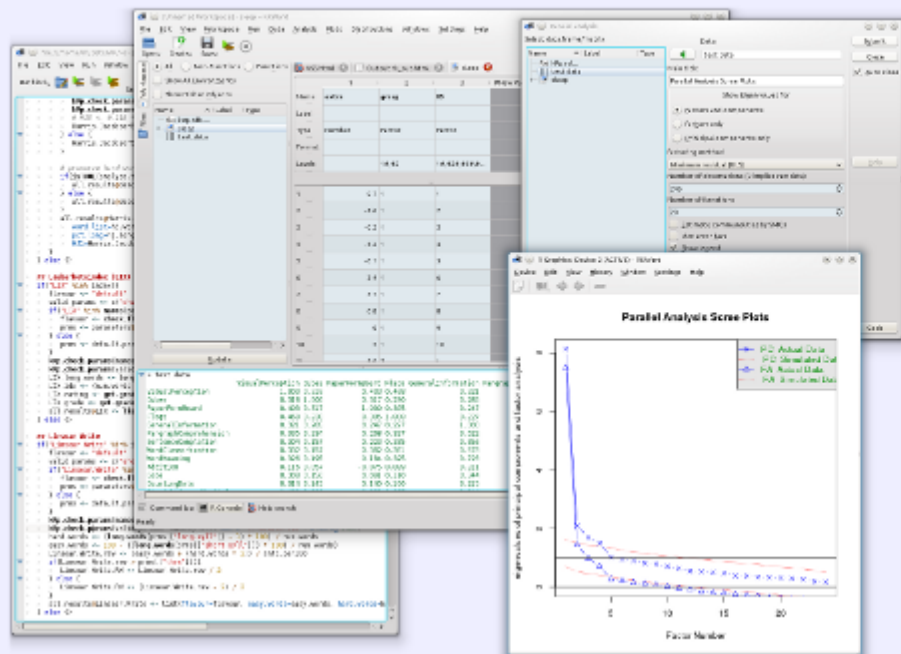


# R is somewhat ascetic, isn't it?

## 6. RKWard (Linux only)

<http://sourceforge.net/projects/rkward/>

### Welcome to RKWard



**RKWard is an easy to use and easily extensible IDE/GUI for R.** It aims to combine the power of the R-language with the ease of use of commercial statistics tools.

RKWard's features include:

- Spreadsheet-like data editor
- Syntax highlighting, code folding and code completion
- Data import (e.g. SPSS, Stata and CSV)
- Plot preview and browsable history
- R package management
- Workspace browser
- GUI dialogs for all kinds of statistics and plots

Its features can be extended by plugins, and it's all free software.





# RKward

[[ Unnamed Workspace ] - file:///home/seraph/software/rkward\_wiki/codes4wiki.R - RKward

File Edit View Workspace Run Analysis Plots Distributions Windows Settings Help

Dataset Script File Open R Script File Open Recent Previous Window Next Window Run selection Interrupt running command

Workspace: All Non-Functions Functions Show All Environments Show Hidden Objects

Name	Class	Type	Label
package:base			
Autoloads			
package:met...			
package:data...			
package:utils			
package:grDe...			
package:grap...			
package:stats			
package:rkward			
package:MASS			
package:lattice			
.GlobalEnv			
lcrabs.pc	matrix		
crabs	data.frame		

```

Multiple Graphics Windows:

library(MASS) # crabs
library(lattice)

levels(crabs$sex) <- c("Female", "Male")
levels(crabs$sp) <- c("Blue", "Orange")
lcrabs.pc <- predict(princomp(log(crabs[,4:8])))
splom(~lcrabs.pc[, 1:3] | sp, data=crabs, pscals=0, cex=0.5, layout=c(1,2))
splom(~lcrabs.pc[, 1:3] | sex, data=crabs, pscals=0, cex=0.5, layout=c(1,2))
splom(~lcrabs.pc[, 1:3] | sp*sex, data=crabs, pscals=0, cex=0.5)

```

/home/seraph/software/rkward\_wiki/

Name	Size	Date
RKwardApplicationData...	209.9 KiB	05/08/09 01:
RKwardApplicationCom...	179.5 KiB	05/08/09 01:
MultiDevices.png	182.1 KiB	05/08/09 01:
LoadPackages.png	42.4 KiB	05/08/09 12:
GraphicsExport2.png	27.5 KiB	05/08/09 01:
GraphicsExport1.png	32.4 KiB	05/08/09 01:
ConfigureRKward3.png	38.5 KiB	05/08/09 01:
ConfigureRKward2.png	38.6 KiB	05/08/09 01:
ConfigureRKward1.png	40.3 KiB	05/08/09 01:
codes4wiki.R	425 B	05/08/09 01:

traffix.Rdata - ppscales - RKward

File Edit View Workspace Run Analysis Plots Distributions Windows Settings Help

Dataset Script File Open R Script File Open Recent Previous Window Next Window Cut Copy Paste Paste inside selection

Workspace: All Non-Functions Functions scaleweights ppscales print.html Show All Environments Show Hidden Objects

Name	Label	Type	Format	Levels
pn6	Personal ...	Number		
an1	Awarenes...	Number		
pn4	Personal ...	Number		
ac13	Awarenes...	Number		
ac8	Awarenes...	Number		
ac11	Awarenes...	Number		

```

> library(MASS) # crabs
> library(lattice)
> levels(crabs$sex) <- c("Female", "Male")
> levels(crabs$sp) <- c("Blue", "Orange")
> lcrabs.pc <- predict(princomp(log(crabs[,4:8])))
> splom(~lcrabs.pc[, 1:3] | sp, data=crabs, pscals=0, cex=0.5, layout=c(1,2))
> splom(~lcrabs.pc[, 1:3] | sex, data=crabs, pscals=0, cex=0.5, layout=c(1,2))
> splom(~lcrabs.pc[, 1:3] | sp*sex, data=crabs, pscals=0, cex=0.5)
>

```

	pn6	an1	pn4	ac13	ac8	ac11
1	3	5	2	4	3	1
2	3	6	2	1	6	6
3	7	7	1	2	7	1
4	5	7	5	2	7	5
5	3	4	4	4	1	6
6	4	7	3	invalid	4	2
7	2	6	2	2	5	2
8	3	5	2	6	3	6
9	5	5	4	6	5	7
10	1	3	7	6	1	3
11	6	7	7	2	6	5
12	7	5	7	3	5	1

Command log Pending Jobs R Console Help search

Ready.

/home/thomas R engine idle



# R is somewhat ascetic, isn't it?

## 7. Deducer

<http://www.deducer.org>

View Edit History Print

### Deducer Manual

(redirected from *Main.HomePage*)

#### An R Graphical User Interface (GUI) for Everyone

Deducer is designed to be a free easy to use alternative to proprietary data analysis software such as SPSS, JMP, and Minitab. It has a menu system to do common data manipulation and analysis tasks, and an excel-like spreadsheet in which to view and edit data frames. The goal of the project is two fold.

1. Provide an intuitive graphical user interface (GUI) for R, encouraging non-technical users to learn and perform analyses without programming getting in their way.
2. Increase the efficiency of expert R users when performing common tasks by replacing hundreds of keystrokes with a few mouse clicks. Also, as much as possible the GUI should not get in their way if they just want to do some programming.

Deducer is designed to be used with the Java based R console JGR, though it supports a number of other R environments (e.g. Windows RGUI and RTerm).

[Download and Install](#)



#### Deducer Manual:

- [Home](#)
- [Download](#)
- [Viewer](#)
- [Teaching](#)
- [Plug-in](#)
- [Development](#)

#### Poker AI Research:

- [Home](#)
- [AAAI 2007](#)
- [AAAI 2008](#)



# Deducer

The screenshot shows the R console window with the 'Analysis' menu open. The menu items are: Frequencies, Descriptives, Contingency Tables, One Sample Test, Two Sample Test, K-Sample Test, Paired Test, Correlation, Linear Model, Logistic Model, and Generalized Linear Model. The 'Plots' sub-menu is also open, showing: Plot Builder, Import Template, Open Plot, Quick, and Interactive. The 'Quick' sub-menu is further open, listing various plot types: bar, grouped bar, histogram, density, grouped density, simple dotplot, grouped dotplot, mean, line, grouped line, simple boxplot, grouped boxplot, scatter, scatter smooth, histogram 2d, and bubble.

```
R version 3.0.3 (2014-11-14)
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/x86_64-w64-mingw32

R is free software and you are welcome to redistribute it
under certain conditions. See the file LICENSE for details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

>
Loading required package: JGR
Loading required package: rJava

Loading required package: JavaGD
Loading required package: iplots
starting httpd help server ... done
> JGR::package.manager()
```

Three overlapping menu panels are shown on the right side of the image. The top panel, titled 'Analysis', lists: Frequencies, Descriptives, Contingency Tables, One Sample Test, Two Sample Test, K-Sample Test, Paired Test, Correlation, Linear Model, Logistic Model, and Generalized Linear Model. The middle panel, titled 'Extras', lists: Distribution quantiles, Distribution function values, Cumulative distribution function, Load data from package, Data summary, Single proportion, Single proportion: Exact, k-sample proportion, k-sample variance test, t-test power, k-means cluster, Apply k-means to data, Hierarchical cluster, Multi-dimensional scaling, Ranking analysis, and Open Vignette. The bottom panel, titled 'Text Packages & Data Window', lists: Distribution quantiles, Distribution function values, Cumulative distribution function, Load data from package, Data summary, Single proportion, Single proportion: Exact, k-sample proportion, k-sample variance test, t-test power, k-means cluster, Apply k-means to data, Hierarchical cluster, Multi-dimensional scaling, Ranking analysis, and Open Vignette.



# Deducer

Generalized Linear Model Explorer

$N$   $f(x)$   $\mathcal{F}$  var1 ~ var3 + var2

General Diagnostics Terms Added Variable

Preview

```
>.gui.working.env$model.glm <- glm(formula=var1 ~ var3 + var2,family=gaussian(),data=.gui.working.env$SomeData,na.action=na.omit)

>Anova(.gui.working.env$model.glm,type='II',test.statistic='Wald')
```

Analysis of Deviance Table (Type II tests)

Response: var1

	Df	Chisq	Pr(>Chisq)
var3	1	0.1006	0.7511
var2	2	1.2681	0.5304

Residuals 96

```
>summary(.gui.working.env$model.glm)
```

Call:

```
glm(formula = var1 ~ var3 + var2, family = gaussian(), data = .gui.working.env$SomeData, na.action = na.omit)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-----	----	--------	----	-----

Options

Post Hoc

Tests

Plots

Means

Export

Update Model

Run Reset Cancel

Options

Tables

Anova Table

Type:  II  III

Test:  LR  Wald  F

Summary Table

Parameter Correlations

Model Diagnostics

Variance Inflation Factors

Influence Summary

OK Cancel



# Deducer

Data Viewer

File Edit Help

Data Set (df) S...

Data View Variable View

	var1	var2	var3	var5
1	-0.361194470	A	0.139364628	Y
2	0.491281568	A	0.145281157	R
3	-2.263512565	A	0.960004409	T
4	0.324549515	A	0.363441332	Q
5	1.164971145	A	0.985196047	L
6	-0.999952689	A	0.970113546	U
7	0.796392712	A	0.344870559	U
8	0.680357541	A	0.884673960	G
9	0.411478560	A		
10	-0.030439523	A		
11	-0.882940467	A		
12	0.189137456	A		
13	1.312458194	A		
14	-0.866649047	A		

Data Viewer

File Edit Help

Data Set (df) SomeData

Data View Variable View

	Variable	Type	Factor Levels
1	var1	Double	
2	var2	Factor	(1) A; (2) B; (3) C;
3	var3	Double	
4	var5	Factor	(1) A; (2) B; (3) C; (4) D; (5) E; (6) F; (7) G; (8) H; (9) I; (1...

Object Browser

- data
  - var1
  - var2
  - var3
- dd
  - Site
  - normDataWithin
  - summarySE
  - summarySEwithin
  - levnum
  - Ymax
- aovmodel

- Data
- Analysis
- Plots
- Extra
- Edit Factor
- Recode Variables
- Transform
- Reset Row Names
- Sort
- Transpose
- Merge Data
- Subset

Data Subset

SomeData

Filter:

var1  
var2  
var3  
var5

Subset Name: <auto>

Subset Expression

```
is.logical(var3) & var1 >= 34
```

Recent:

Logical Functions

- is.character
- is.integer
- is.logical
- is.complex
- is.double
- is.numeric
- is.vector
- is.factor
- is.finite

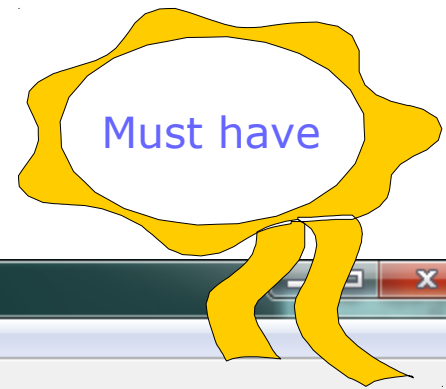
Logical Operators

- >=
- <=
- >
- <
- =
- !=
- And
- Or
- Not

OK Reset Cancel



# Deducer



Plot Builder

File Tools Window

Templates Geometric Elements Statistics Scales Facets Coordinates Other

bar grouped bar histogram density grouped density

bar grouped bar histogram density grouped density

Drag a component  
OR  
Select a plot

histogram bar grouped dotplot grouped line

Plot Builder

File Tools Window

Templates Geometric Elements Statistics Scales Facets Coordinates Other

bar grouped bar histogram density grouped density simple dotplot grouped dotplot mean

Made with Deducer

var1

var3

A B C

var3

- A
- B
- C

Components

- geom boxplot
- geom jitter
- theme title
- theme theme




Run Reset Cancel



# R is somewhat ascetic, isn't it?


## 8. TinnR

<http://sourceforge.net/projects/tinn-r>

Enterprise / Text Processing / Tinn-R Share    

 **Tinn-R**

★★★★★ (31) [Read Reviews](#) | Last Updated 2014-02-28

 **Download** 799 Downloads (This Week)

Windows

[Reviews](#) | [Files](#) | [Tools](#) ▼



Tinn-R - [D:\dados\tex\jcfaria\exercicios\res\_lista01.rnw]

File Project Edit Format Marks Insert Search Options Encoding Tools R View Window Web Help

Tinn-R\_example of script.r Tinn-R\_recognized words.r res\_lista01.rnw

Tools Misc Markup Results Spell Database R

Explorer Card Mirrors

0-Cloud  
 Argentina  
 Australia  
 Austria  
 Belgium  
 Brazil  
 Canada  
 Chile  
 China  
 Colombia  
 Denmark  
 Ecuador  
 France  
 Germany  
 Greece  
 Hungary  
 India  
 Indonesia

Name	City
Brazil (BA)	Ilheus
Brazil (PR)	Curitiba
Brazil (RJ)	Rio de Janeiro
Brazil (SP 1)	Sao Paulo
Brazil (SP 2)	Piracicaba

Center for Comp. Biol. at Universidade Estadual de Santa Cruz  
<http://nbcgib.uesc.br/mirrors/cran/>  
 br  
 Default: <http://nbcgib.uesc.br/mirrors/cran/>

```

\subsection{Decomposição}
\subsubsection{Decomposição espectral}
<<>>=
dA = eigen(A); dA
@

\subsubsection{Somatório da decompos}
<<>>=
sdA = 1/dA$values[1]*dA$vectors[,1]%
 1/dA$values[2]*dA$vectors[,2]%
@

2 \subsubsection{Inversa de A}
<<>>=
iA = solve(A); iA
@
a=
b=
tol=Machine$double.eps
LINPACK=FALSE
...

\subsubsection{...}
<<>>=
all(round(iA, 5) == round(sdA, 5))
@

\pagebreak
\section{Questão} %3
\subsection{$(A^{-1})^{-1}=(A^{-1})^{-1}$}
<<>>=
A = matrix(c(1,2,3,4), nr=2); A
B = matrix(c(5,6,7,8), nr=2); B

```

Rterm

```

> A = matrix(c(3,1,1,2), nr=2); A
 [,1] [,2]
[1,] 3 1
[2,] 1 2

> dA = eigen(A); dA
$values
[1] 3.618034 1.381966

$vectors
 [,1] [,2]
[1,] -0.8506508 0.5257311
[2,] -0.5257311 -0.8506508

> iA = solve(A); iA
 [,1] [,2]
[1,] 0.4 -0.2
[2,] -0.2 0.6

```

ANSI WIN 149/748: 12 Editing Normal Size: 18.26 KB R hotkeys active R -> TCP/IP [base] <solve.default>





**Tinn-R**

File Project Edit Format Marks Insert Search Options Encoding Tools R View Window Web Help

Tinn-R\_example of script.r Tinn-R\_recognized words.r pandoc.markdown

Tools Misc Markup Results Spell Database R

Explorer Card Mirrors

package:datasets

All

- ability.cov
- airquality
- anscombe
- attenu
- attitude
- austres
- beaver1
- beaver2
- Bjsales
- Bjsales.lead
- BOD
- cars
- ChickWeight
- chickwts
- co2
- CO2
- crimtab
- discoveries
- DNase
- esoph
- euro
- euro.cross
- eurodist
- EuStockMarkets
- faithful
- fdeaths
- Formaldehyde
- freeny
- freeny.x
- freeny.y
- HairEyeColor
- Harman23.cor
- Harman74.cor
- Indometh
- infert
- InsectSprays
- iris
- iris3

Total: 103 Index: Sel:

C:\Tinn-R\doc\Tinn-R\_recogni... C:\Tinn-R\sample\pandoc.mar... C:\Tinn-R\sample\Tinn-R\_exa...

```
=====
Mail: <<< joseclaudio.far
More: <<< http://zoonek2.
=====

help.start() start the htm
your machine). You should
Iconify the help window an

1 x <- rnorm(50)
x

2 y <- rnorm(50)
y
Generate two pseudo-random

3 x11(w=4,
 h=4)

bringToTop(s=T)

plot(x,
 y)

Plot the points in the pla
```

Rterm

IO Log

```
> plot
function (x, y, ...)
UseMethod("plot")
<bytecode: 0x0000000012dbeb48>
<environment: file=
header=FALSE
sep="
quote="\ "
dec='.'
row.names=
col.names=
as.is=!stringsAsFactors
na.strings='NA'
colClasses=NA
nrows=-1
skip=0
check.names=TRUE
fill=blank.lines.skip
strip.white=FALSE
blank.lines.skip=TRUE
comment.char='#'
allowEscapes=FALSE
flush=FALSE
stringsAsFactors=default.stringsAsFac
fileEncoding="
encoding='unknown'
text=

> 1:1e2
[1] 1 2
[9] 9 10
[17] 17 18
[25] 25 26
[33] 33 34
[41] 41 42
[49] 49 50
[57] 57 58
[65] 65 66
[73] 73 74
[81] 81 82
[89] 89 90
[97] 97 98

>read.table(
```

ANSI WIN 17/100: 2 Editing Normal Size: 2.45 KB R hotkeys active R -> TCP/IP [utils] <read.table>



**Pandoc (document converter)**

From

<input type="radio"/> docbook (DocBook XML)	<input checked="" type="radio"/> markdown (markdown)
<input type="radio"/> html (HTML)	<input type="radio"/> native (native Haskell)
<input type="radio"/> json (JSON version of native AST)	<input type="radio"/> rst (reStructuredText)
<input type="radio"/> latex (LaTeX)	<input type="radio"/> textile (Textile)

To

<input type="radio"/> asciidoc (AsciiDoc)	<input type="radio"/> mediawiki (MediaWiki markup)
<input type="radio"/> beamer (LaTeX beamer slide show)	<input type="radio"/> native (native Haskell)
<input type="radio"/> context (ConTeXt)	<input type="radio"/> odt (OpenOffice text document)
<input type="radio"/> docbook (DocBook XML)	<input type="radio"/> opendocument (OpenDocument XML)
<input type="radio"/> docx (Word docx)	<input type="radio"/> org (Emacs Org-Mode)
<input type="radio"/> dzslides (HTML5 + javascript slide show)	<input type="radio"/> plain (plain text)
<input type="radio"/> epub (EPUB book)	<input type="radio"/> rst (reStructuredText)
<input checked="" type="radio"/> html (XHTML 1)	<input type="radio"/> rtf (rich text format)
<input type="radio"/> html5 (HTML 5)	<input type="radio"/> s5 (S5 HTML and javascript slide show)
<input type="radio"/> json (JSON version of native AST)	<input type="radio"/> slideous (Slideous HTML and javascript slide show)
<input type="radio"/> latex (LaTeX)	<input type="radio"/> slidy (Slidy HTML and javascript slide show)
<input type="radio"/> man (groff man)	<input type="radio"/> texinfo (GNU Texinfo)
<input type="radio"/> markdown (markdown)	<input type="radio"/> textile (Textile)

Instruction (will be send to pandoc.exe)

```
pandoc -f markdown -t html %from -o %to
```

%from: C:\tmp\pandoc.md  
%to: C:\tmp\pandoc.html

Instruction (history)

%from:  
%to:

Restore default    OK    Cancel

**Tools**

Misc   Markup   Results   Spell   Database   R

Windows expl.   Work expl.   Project

- tinn-r\_user\_guide.tps
  - 00. user\_guide
    - user\_guide.html
    - user\_guide.t2t
  - 01. chapter\_1
    - 0\_index\_overview.t2t
    - acknowledgments.t2t
    - feedback.t2t
    - have\_to\_pay.t2t
    - motivation.t2t
    - phrase.t2t
    - quick\_start.t2t
    - tools\_used.t2t
    - what\_is\_tinn-r.t2t
    - what\_you\_get.t2t
    - why\_tinn-r.t2t
  - 02. chapter\_2
  - 03. chapter\_3
  - 04. chapter\_4
  - 05. chapter\_5
  - 05. chapter\_6

Total: 226    Index: 1    Sel: tinn-r\_user\_guide.tps



# 13 reasons why **you will** love GNU R

- I R is (extremely) cheap. In fact - it's free :)
- II R has (extremely) wide range of capabilities
- III R is (widely) supported by the world of science
- IV R is supported by the community
- V R is (increasingly) supported by the business
- VI R is able to read data in many formats
- VII Interoperability is easy to achieve
- VIII R is truly cross-platform
- IX R offers numerous ways of presenting data
- X There are many options to optimize the code
- XI R is able to handle large amount of data
- XII R has a set of fancy tools and IDEs
- XII FDA accepted using R for drug trials!**



# FDA: R is **OK** for drug trials!

In a poster presented at the UseR 2012 conference, FDA biostatistician Jae Brodsky reiterated the FDA policy regarding software used to prepare submissions for drug approvals with clinical trials: **Sponsors may use R in their submissions.** [...] SAS **is not required** to be used for clinical trials.



The FDA **does not endorse or require** any particular software to be used for clinical trial submissions, and there are **no regulations that restrict the use** of open source software (including R) at the FDA.

Nonetheless, any software (R included) used to prepare data analysis from clinical trials **must comply** with the various FDA regulations and guidances (e.g. [21 CFR Part 11](#)). Even MS Excel can be made "21 CFR 11 Compliant".



The R Foundation helpfully **provides a guidance document** for the use of R in regulated clinical trial environments, which provides details of the specific FDA regulations and how R complies with them.



# the (veRy important) guidance

Firefox

R-FDA.pdf

www.r-project.org/doc/R-FDA.pdf

Page: 1 of 25 90%

**R: Regulatory Compliance and Validation Issues**  
**A Guidance Document for the Use of R in Regulated Clinical Trial Environments**

*August 8, 2013*

The R Foundation for Statistical Computing  
c/o Institute for Statistics and Mathematics  
Wirtschaftsuniversität Wien  
Augasse 2-6  
1090 Vienna, Austria  
Tel: (+43 1) 31336 4754  
Fax: (+43 1) 31336 904754  
Email: R-foundation@R-project.org

<http://www.r-project.org/doc/R-FDA.pdf>



# the (veRy important) poster



## Some Challenges of Using *R* in a Regulatory Environment

Jae Brodsky

Office of Biostatistics, OTS, CDER, FDA

### The Big Question

How well can *R* be used as a statistical tool in a large federal regulatory agency such as the FDA?

### Current *R* Usage at the FDA

- FDA offices use *R* on a daily basis
  - includes Office of Biostatistics
- FDA scientists have written *R* packages for other scientists' use (FDA or non-FDA)
- FDA does not endorse or require the use of any specific software for statistical analysis

### Regulations and Guidelines for Regulatory Use

- Code of Federal Regulations Title 21: FDA
  - Part 11: Electronic records and electronic signatures
- General Principles of Software Validation: Final Guidance for Industry and FDA Staff (2002)
- Guidance Part 11: Electronic Records (Final) (2003)
- Guidance for Industry: Computerized Systems Used in Clinical Investigations (2007)
- ICH E6: Good Clinical Practice Consolidated Guideline
- ICH E9: Statistical Principles for Clinical Trials

### Code of Federal Regulations

21 CFR 11 deals with electronic records and signatures that are "trustworthy, reliable and generally equivalent to paper records."

- Qualification, verification and validation
- *R* is only part of a larger system that is verified
- *R* Foundation compliance document (2008): <http://www.r-project.org/doc/R-FDA.pdf>

This poster reflects the views of the author and should not be construed to represent FDA's views or policies.

### Regulations and Guidances: What Applies to *R* at the FDA

Regulations state the specifics of what is legally required and enforceable. Guidances describe the agency's current position and are recommendations. 21 CFR Part 11 regulates electronic records and their storage, not the software that is used to generate reports. There are no regulations that restrict the use of open source software (including *R*) at the FDA. However, there are concerns for the FDA beyond CFR Title 21.

### Current FDA *R* Setup

Only base and recommended packages are fully supported by the *R* core. Code or functions outside of these packages are used at the reviewer's own risk.

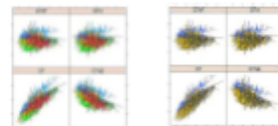
- Base: *datasets*, *graphics*, *grDevices*, *grid*, *methods*, *splines*, *stats*, *stats4*, *tecltk*, *tools*, *utils*
- All results from data should be reproducible independent of software used.
- FDA is not responsible if results have discrepancies.

Sponsors may use *R* in their submissions. Data must be submitted in *xpt* format. Reviewers would like to know:

- which *R* functions were used
- how they were accessed
- that the results are accurate, reliable, and consistent

### Further Regulatory Issues That Impact *R* Usage at the FDA

As a government agency, the FDA must comply with additional regulations that do not impact *R*. Section 508 of the Rehabilitation Act of 1973 (29 U.S.C. 794d) established guidelines for ensuring that the federal government's electronic information is accessible to people with disabilities, including blindness, deafness, colorblindness, epilepsy, and reduced motor function. Section 508 includes guidelines for luminosity or brightness, use of color, font size, text alternatives for figures or plots, and text-only alternatives. All reports released online by the FDA should be 508 compliant, including reports from the Office of Biostatistics.



Example from an FDA report: Left: a plot made with the default rainbow color palette. This figure is not 508 compliant because some information is lost to colorblind people. Right: the red-green colorblind simulation. It is not possible to determine which lines correspond to individual subjects when the color palette is reduced.

Would the FDA need to modify or write its own version of popular *R* packages to meet additional guidelines? Section 508 compliance is not mandated for *R*, and the FDA cannot ask the *R* community to automatically make packages 508 compliant.

### My Personal Experience as an *R* User at the FDA

The QT interval is a biomarker for cardiac repolarization. The QT Interdisciplinary Review Team (IRT) reviews Thorough QT (TQT) studies to investigate QT interval lengthening. The QT IRT consists of statisticians, clinical pharmacologists, and clinicians.

I am the only *R* user in the stats group (4 statisticians). I use:

- *SAS* for reviews (for consistency with other QT IRT stats reviewers)
- *R* for research
- *R* for any graphics outside reviews

The QT IRT clinical pharmacologists use *R* for their reviews. The *R* package *QT* was written specifically for the QT IRT.

- *QT* calls the stats group's *SAS* macros for areas in the clin pharm portion of the review
- other areas of the clin pharm review are done in *R*
- original author of *QT* no longer at the FDA

My future plans: to write an *R* package for QT IRT use that covers all parts of a TQT review.

Timeline: ?

In my experience, *R* use at the FDA is completely acceptable and has not caused any problems.

### References

- [1] Title 21 CFR Part 11 Section 11.1 (a)
- [2] General Principles of Software Validation; Final Guidance for Industry and FDA Staff
- [3] Guidance Part 11: Electronic Records (Final)
- [4] Guidance for Industry: Computerized Systems Used in Clinical Investigations
- [5] ICH E6(R1): Good Clinical Practices
- [6] ICH E9: Statistical Principles for Clinical Trials
- [7] *R*: Regulatory Compliance and Validation Issues: A Guidance Document for the Use of *R* in Regulated Clinical Trial Environments
- [8] Section 508 of the Rehabilitation Act of 1973 (29 U.S.C. 794d):



# some important information

SoukupR@FDA.dvi - soukup.pdf x +

user2007.org/program/presentations/soukup.pdf

Page: 1 of 22 Page Fit

Regulations and Guidances Relying on R for Statistical Analysis FDA Reviewers? The Future

## USING R: PERSPECTIVES OF A FDA STATISTICAL REVIEWER

Mat Soukup, Ph.D.  
Mat.Soukup@fda.hhs.gov

FDA/CDER/OTS/OB/DB3

<http://user2007.org/program/presentations/soukup.pdf>

Perspectives of a FDA Statistical Reviewer SLIDE: 1



FurtheR impoRtant issues





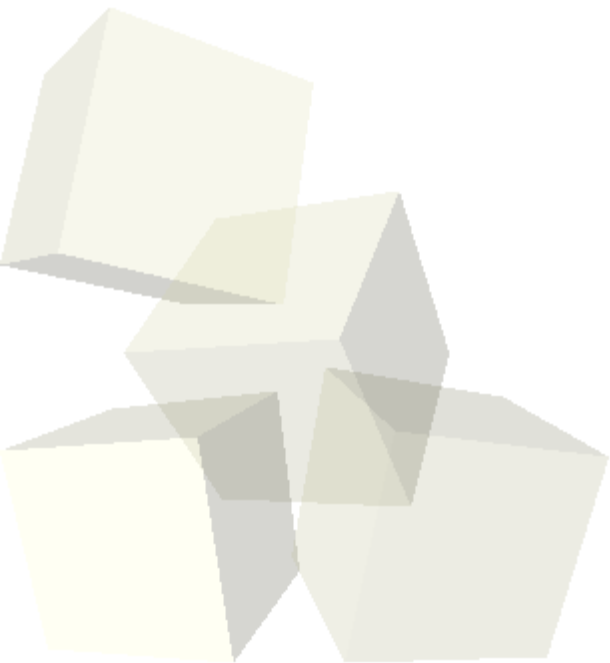
## **I** Handling metadata

II Handling ODM and CDA metadata

III Issue with multilingual data

IV Differences between SAS and R

V Implementation of useful SAS functions





# Handling metadata

Briefly, *metadata is data that describes other data.*

Metadata can be used for:

- Commenting and labeling (titling) objects (datasets, functions, graphs, tables)
- Adding instructions of how the content of a variable should be rendered, e.g.:
  - number of decimal digits, format of dates and times, length of texts
  - format of outputted text: font name, color, size, style
  - dictionary of which values should appear instead of raw data (translation)
- Searching purposes (keywords, tags)

Note: *metadata doesn't change the underlying, raw value*, it is just additional information.



# Handling metadata

SAS uses metadata extensively to control how data is displayed. There are dozens of [formats](#) and ["informats"](#) specific to literals, numbers, currencies, dates and times.

On the contrary, **R doesn't use "formats"**. There is no something like *"numeric(10,2)"* or *"char(20)"*. Without additional efforts data is displayed **"as is"**.

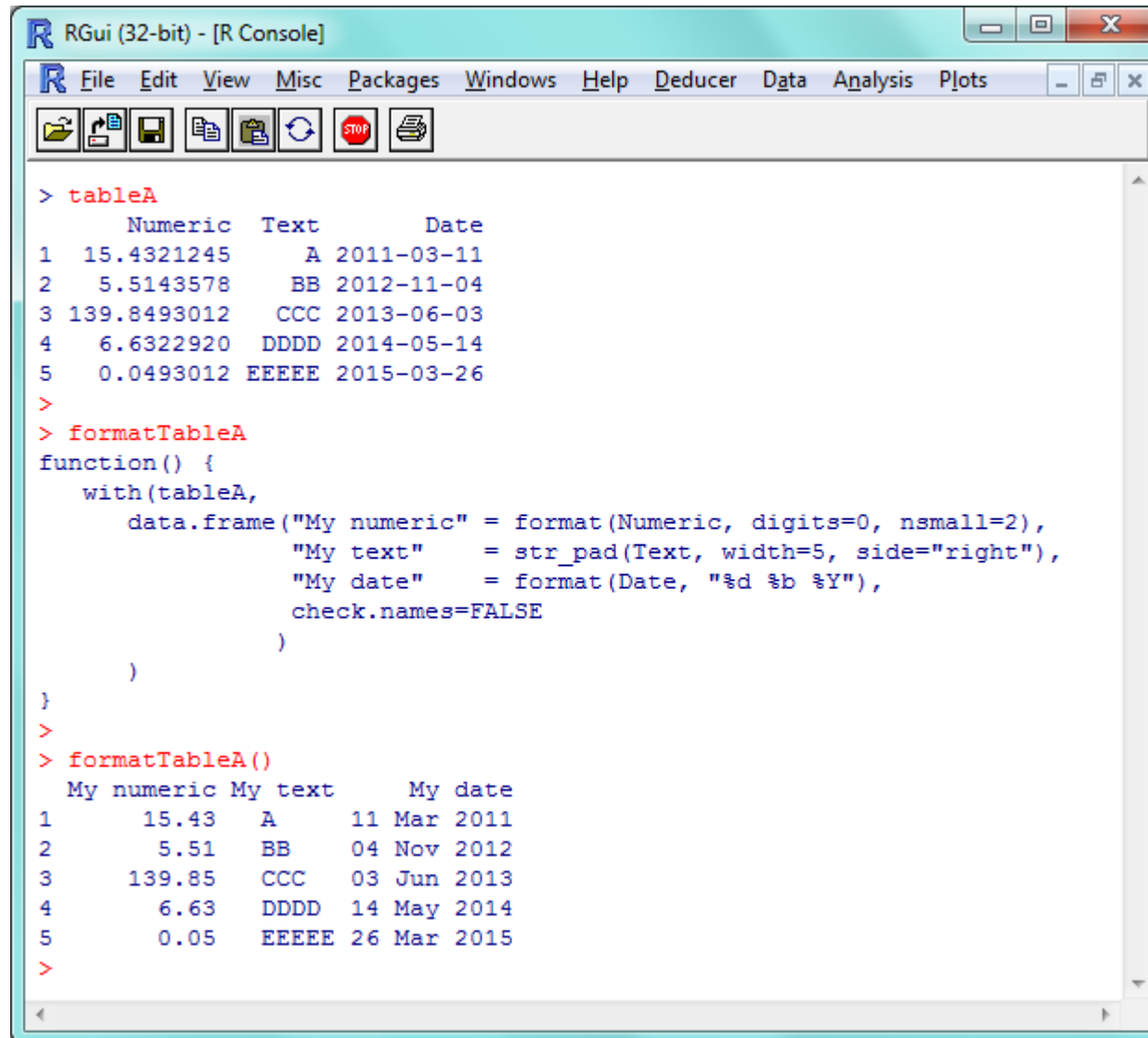
There are only some options to control **globally** the number of decimal digits, formatting dates or setting timezone.

```
> options("digits")$digit
[1] 7
> (x <- 12.3456789)
[1] 12.34568
> options(digits = 4)
> x
[1] 12.35
> x == 12.3456789
[1] TRUE
```



# Handling metadata

To control formats on the column or cell level one must **write additional formatting code every time it is required** (usually a function is created for this purpose).



```
RGui (32-bit) - [R Console]
File Edit View Misc Packages Windows Help Deducer Data Analysis Plots
> tableA
 Numeric Text Date
1 15.4321245 A 2011-03-11
2 5.5143578 BB 2012-11-04
3 139.8493012 CCC 2013-06-03
4 6.6322920 DDDD 2014-05-14
5 0.0493012 EEEEE 2015-03-26
>
> formatTableA
function() {
 with(tableA,
 data.frame("My numeric" = format(Numeric, digits=0, nsmall=2),
 "My text" = str_pad(Text, width=5, side="right"),
 "My date" = format(Date, "%d %b %Y"),
 check.names=FALSE
)
)
}
>
> formatTableA()
My numeric My text My date
1 15.43 A 11 Mar 2011
2 5.51 BB 04 Nov 2012
3 139.85 CCC 03 Jun 2013
4 6.63 DDDD 14 May 2014
5 0.05 EEEEE 26 Mar 2015
>
```



# Handling metadata

This **separates metadata from data** and is an anti-pattern leading to:

- **Creation of dedicated formatting code specific to a data object.** Amount of the code **grows quickly** as new variables are created (*formatTableA, formatListB, etc.*)
- **Errors**, when someone mistakes one metadata for other. This happens easily for large sets of variables.

The **golden rule** is that the ***data should be "self-describing"***.

This indicates the need to introduce a separate, independent, intermediate metadata layer between the data and the output rendering code, which consists of:

- **Set of attributes** describing data objects
- **Functions assigning** these attributes to data objects
- **Functions retrieving** values of the attributes **and applying** them to the raw data



# Metadata in use

```
RGui (32-bit) - [R Console]
File Edit View Misc Packages Windows Help Deducer Data Analysis Plots

> SetVarAttributes(tableA$Numeric, decimPlaces=2, label="My numeric")
> SetVarAttributes(tableA$Text, textWidth=5, label="My text")
> SetVarAttributes(tableA$Date, dateFormat="%d %b %Y", label="My date")
>
> tableA
 Numeric Text Date
1 15.4321245 A 2011-03-11
2 5.5143578 BB 2012-11-04
3 139.8493012 CCC 2013-06-03
4 6.6322920 DDDD 2014-05-14
5 0.0493012 EEEEE 2015-03-26
>
> attributes(tableA$Date)
$class
[1] "Date"

$comment
[1] "My date"

$user.date.format
[1] "%d %b %Y"

> fTableA <- ApplyAttributesToDataFrame(tableA, sourceOfMetadata=tableA)
> knitr::kable(fTableA$dataFrame, align=fTableA$aligns)

| My numeric|My text | My date|
|-----:|:-----:|-----:|
| 15.43|A | 11 Mar 2011|
| 5.51|BB | 04 Nov 2012|
| 139.85|CCC | 03 Jun 2013|
| 6.63|DDDD | 14 May 2014|
| 0.05|EEEEE | 26 Mar 2015|
>
```

Assigning attributes (metadata) to the original table

Original data stays unchanged

Metadata bound to a column of the original table

Applying metadata to the table

Displaying formatted result



Fortunately, it is possible to create one's own [metadata layer](#) by using R [attributes](#).

Author of this presentation wrote for his own use a set of functions for:

- Managing and using dictionaries (*code lists*)
- Binding metadata to objects
- Controlling the way data is displayed (applying metadata to datasets)
- Displaying the data with metadata overlapped
- Copying metadata between objects (when attributes were wiped out)
- Describing datasets in detail
- Reading labels of variables from sasXmlMap file
- Creating SAS Transport Files (XPT) using the attached metadata

They are supposed to be released as an R package in future.



# Own metadata layer - example

Definition of an exemplary dataset:

```
> x <- c(5L,NA,6L,7L,1L)

> data <- data.frame(
 unlabeled = c(NA, "A", "B", "C", "D"),
 vlt = c("This is a long string", "Another long string", "Much longer string than previous",
 "A short string", NA),
 dv = c(NA, rnorm(4)),
 pct = x/sum(na.omit(x)),
 int = x,
 dt = as.POSIXct(c("2011-02-01 08:00", "2010-04-21 12:30", NA, "1999-11-02 14:15", "2000-10-15 07:30")),
 dat = as.POSIXct(c("2011-02-01 08:00", "2010-04-21 12:30", NA, "1999-11-02 14:15", "2000-10-15 07:30")),
 tim = as.POSIXct(c("2011-02-01 08:00", "2010-04-21 12:30", NA, "1999-11-02 14:15", "2000-10-15 07:30")),
 trt = c(1,NA,2,4,3))
```

```
> # raw, unformatted data
```

```
> data
 unlabeled vlt dv pct int dt dat tim trt
1 <NA> This is a long string NA 0.26315789 5 2011-02-01 08:00:00 2011-02-01 08:00:00 2011-02-01 08:00:00 1
2 A Another long string -1.4286644 NA NA 2010-04-21 12:30:00 2010-04-21 12:30:00 2010-04-21 12:30:00 NA
3 B Much longer than previous 1.5035051 0.31578947 6 <NA> <NA> <NA> 2
4 C A short string -1.3464709 0.36842105 7 1999-11-02 14:15:00 1999-11-02 14:15:00 1999-11-02 14:15:00 4
5 D <NA> -0.7571589 0.05263158 1 2000-10-15 07:30:00 2000-10-15 07:30:00 2000-10-15 07:30:00 3
```





# Metadata - example

Definition of exemplary dictionaries:

```
Look how a definition in SAS format can be used directly
```

```
> AddUserDictFromDef(definitions = 'value MyDict
 1 = "Treatment A"
 2 = "Treatment B"
 3 = "Placebo";', SASFormat=TRUE)
```

```
> AddUserDict("sex", codes = c("F", "M"),
 labels = c("Female", "Male"))
```

```
> UCODELST
```

	code1st	clcode	cllabel
1	MyDict	1	Treatment A
2	MyDict	2	Treatment B
3	MyDict	3	Placebo
4	sex	F	Female
5	sex	M	Male



# Metadata - example

Add some metadata:

```
> SetVarAttributes(data$vl, label = "Very long text", textWidth = 15, sasFormat = "$15.")
> SetVarAttributes(data$dv, label = "Double value", decimPlaces = 2, sasFormat = "8.2")
> SetVarAttributes(data$pct, label = "Percentage", decimPlaces = 1, asPercent = TRUE)
> SetSVarAttributes("data$int", label = "Integer") # variable name passed as a literal
> SetVarAttributes(data$dat, label = "Date", dateFormat = "%d%b%Y", sasFormat = "yymmdd10.")
> SetVarAttributes(data$dt, label = "Date & Time (ISO 8601)", dateFormat = "%Y-%m-%d %H:%M:%S",
 sasFormat = "e8601dt19.")
> SetVarAttributes(data$tim, label = "Time", dateFormat = "%H:%M")
> SetVarAttributes(data$str, label = "Translated", translDict = "MyDict")
> SetVarAttributes(data, label = "My data set")
```

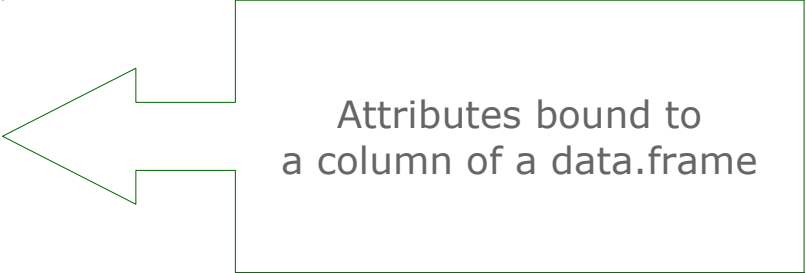
```
> attributes(data$dt)
$class
[1] "POSIXct" "POSIXt"

$tzone
[1] ""

$comment
[1] "Date & Time (ISO 8601)"

$user.date.format
[1] "%Y-%m-%d %H:%M:%S"

$user.sas.format
[1] "e8601dt19."
```



Attributes bound to  
a column of a data.frame



# Metadata - example

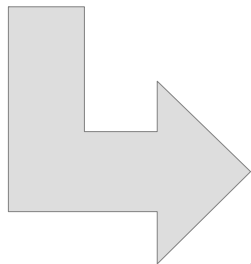
Display the data with metadata applied:

```
Raw data
> data
 unlabeled vlt dv pct int dt dat tim trt
1 <NA> This is a long string NA 0.26315789 5 2011-02-01 08:00:00 2011-02-01 08:00:00 2011-02-01 08:00:00 1
2 A Another long string -1.4286644 NA NA 2010-04-21 12:30:00 2010-04-21 12:30:00 2010-04-21 12:30:00 NA
3 B Much longer than previous 1.5035051 0.31578947 6 <NA> <NA> <NA> 2
4 C A short string -1.3464709 0.36842105 7 1999-11-02 14:15:00 1999-11-02 14:15:00 1999-11-02 14:15:00 4
5 D <NA> -0.7571589 0.05263158 1 2000-10-15 07:30:00 2000-10-15 07:30:00 2000-10-15 07:30:00 3
```

# Data with metadata overlapped (output is in RMarkdown format)

```
> PrintDataFrame(data, sourceOfMetadata = data)
```

```
| Row|unlabeled |Very long text | Double value| Percentage| Integer| Date & Time (ISO 8601)| Date| Time|Translated |
|---:|:-----|:-----|:-----|:-----|:-----|:-----|:-----|:-----|:-----|
| 1| |This is a long | | 26.3%| 5| 2011-02-01 08:00:00| 01Feb2011| 08:00|Treatment A |
| 2|A|Another long st | -1.43| | | 2010-04-21 12:30:00| 21Apr2010| 12:30| |
| 3|B|Much longer tha | 1.50| 31.6%| 6| | | |Treatment B |
| 4|C|A short string | -1.35| 36.8%| 7| 1999-11-02 14:15:00| 02Nov1999| 14:15|4 |
| 5|D| | -0.76| 5.3%| 1| 2000-10-15 07:30:00| 15Oct2000| 07:30|Placebo |
```



Row	unlabeled	Very long text	Double value	Percentage	Integer	Date & Time (ISO 8601)	Date	Time	Translated
1		This is a long		26.3%	5	2011-02-01 08:00:00	01Feb2011	08:00	Treatment A
2	A	Another long st	2.24			2010-04-21 12:30:00	21Apr2010	12:30	
3	B	Much longer tha	-0.23	31.6%	6				Treatment B
4	C	A short string	0.22	36.8%	7	1999-11-02 14:15:00	02Nov1999	14:15	4
5	D		-0.65	5.3%	1	2000-10-15 07:30:00	15Oct2000	07:30	Placebo



# Metadata - example

Describe dataset with metadata attached:

```
> DescribeTable(data) # simple, default form
```

My data set

---

```
Rows: 5

dat* DAT maxL: 19 W: D: T: N: 4 NA: 1 U: 5 | 01Feb2011, 21Apr2010, NA, 02Nov1999, 15Oct2000
dt* DAT maxL: 19 W: D: T: N: 4 NA: 1 U: 5 | 2011-02-01 08:00:00, 2010-04-21 12:30:00, NA, 1999-11-02 14:15:00, 2000-10-15 07:30:00
dv* dbl maxL: 18 W: D:2 T: N: 4 NA: 1 U: 5 | Min -1.43, Q1 -1.39, Me -1.05, Q3 0.37, Max 1.5 | NA, -1.42866436243726, 1.50350508808 ...
int* int maxL: 2 W: D:0 T: N: 4 NA: 1 U: 5 | Min 1, Q1 3, Me 5.5, Q3 6.5, Max 7 | 5, NA, 6, 7, 1
pct* dbl maxL: 18 W: D:1 T: N: 4 NA: 1 U: 5 | Min 0.05, Q1 0.16, Me 0.29, Q3 0.34, Max 0.37 | 0.263157894736842, NA, 0.3157894736842 ...
tim* DAT maxL: 19 W: D: T: N: 4 NA: 1 U: 5 | 08:00, 12:30, NA, 14:15, 07:30
trt* dbl maxL: 2 W: D:2 T:Y N: 4 NA: 1 U: 5 | Min 1, Q1 1.5, Me 2.5, Q3 3.5, Max 4 | (MyDict) Treatment A, NA, Treatment B, 4, Placebo
unlabeled txt maxL: 2 W: D: T: N: 4 NA: 1 U: 5 | NA, A, B, C, D
vlt* txt maxL: 25 W: 15 D: T: N: 4 NA: 1 U: 5 | This is a long string, Another long string, Much longer than previous, A short string, NA
```

```
> DescribeTable(data, displayLabels=TRUE, displaySASFormats = TRUE) # more detailed
```

My data set

---

```
Rows: 5

dat [Date] DAT maxL: 19 W: D: SASf: yymmdd10. T: N: 4 NA: 1 U: 5 | 01Feb2011, 21Apr2010, NA, 02Nov1999, 15Oct2000
dt [Date & Time (ISO 8601)]DAT maxL: 19 W: D: SASf:e8601dt19. T: N: 4 NA: 1 U: 5 | 2011-02-01 08:00:00, 2010-04-21 12:30:00, NA, 1 ...
dv [Double value] dbl maxL: 18 W: D:2 SASf: 8.2 T: N: 4 NA: 1 U: 5 | Min -1.43, Q1 -1.39, Me -1.05, Q3 0.37, Max 1.5 ...
int [Integer] int maxL: 2 W: D:0 SASf: T: N: 4 NA: 1 U: 5 | Min 1, Q1 3, Me 5.5, Q3 6.5, Max 7 | 5, NA, 6, ...
pct [Percentage] dbl maxL: 18 W: D:1 SASf: T: N: 4 NA: 1 U: 5 | Min 0.05, Q1 0.16, Me 0.29, Q3 0.34, Max 0.37 | ...
tim [Time] DAT maxL: 19 W: D: SASf: T: N: 4 NA: 1 U: 5 | 08:00, 12:30, NA, 14:15, 07:30
trt [Translated] dbl maxL: 2 W: D:2 SASf: T:Y N: 4 NA: 1 U: 5 | Min 1, Q1 1.5, Me 2.5, Q3 3.5, Max 4 | (MyDict) ...
unlabeled txt maxL: 2 W: D: SASf: T: N: 4 NA: 1 U: 5 | NA, A, B, C, D
vlt [Very long text] txt maxL: 25 W: 15 D: SASf: $15. T: N: 4 NA: 1 U: 5 | This is a long string, Another long string, Muc ...
```



# Metadata - example

Transfer attributes between two datasets.

```
> newData <- sqldf("SELECT * FROM data") # sqldf wipes out attributes
```

```
> DescribeTable(newData, displayLabels=TRUE, displaySASFormats = TRUE)
```

```
Rows: 5
```

dat	DAT	maxL: 19	W:	D:	SASf:	T:	N:	4	NA:	1	U:	5		2011-02-01 08:00:00, 2010-04-21 12:30:00, NA, 1999-11-02 14:15:00, 2000-10-15 07 ...
dt	DAT	maxL: 19	W:	D:	SASf:	T:	N:	4	NA:	1	U:	5		2011-02-01 08:00:00, 2010-04-21 12:30:00, NA, 1999-11-02 14:15:00, 2000-10-15 07 ...
dv	dbl	maxL: 18	W:	D:2	SASf:	T:	N:	4	NA:	1	U:	5		Min -0.65, Q1 -0.44, Me 0, Q3 1.23, Max 2.24   NA, 2.24243108561811, -0.22569800 ...
int	int	maxL: 2	W:	D:0	SASf:	T:	N:	4	NA:	1	U:	5		Min 1, Q1 3, Me 5.5, Q3 6.5, Max 7   5, NA, 6, 7, 1
pct	dbl	maxL: 18	W:	D:2	SASf:	T:	N:	4	NA:	1	U:	5		Min 0.05, Q1 0.16, Me 0.29, Q3 0.34, Max 0.37   0.263157894736842, NA, 0.3157894 ...
tim	DAT	maxL: 19	W:	D:	SASf:	T:	N:	4	NA:	1	U:	5		2011-02-01 08:00:00, 2010-04-21 12:30:00, NA, 1999-11-02 14:15:00, 2000-10-15 07 ...
trt	dbl	maxL: 2	W:	D:2	SASf:	T:	N:	4	NA:	1	U:	5		Min 1, Q1 1.5, Me 2.5, Q3 3.5, Max 4   1, NA, 2, 4, 3
unlabeled	txt	maxL: 2	W:	D:	SASf:	T:	N:	4	NA:	1	U:	5		NA, A, B, C, D
vlt	txt	maxL: 25	W:	D:	SASf:	T:	N:	4	NA:	1	U:	5		This is a long string, Another long string, Much longer than previous, A short s ...

```
> CopyUserAttributes(srcDS = data, destDS = newData, env=environment())
```

```
Copied table attribute: [newData]@comment = My data set
Copied column attribute: [vlt]@comment = Very long text
Copied column attribute: [vlt]@user.str.len = 15
Copied column attribute: [vlt]@user.sas.format = $15.
...
Copied column attribute: [trt]@comment = Translated
Copied column attribute: [trt]@user.transl.dict = MyDict
```

```
> DescribeTable(newData, displayLabels=TRUE, displaySASFormats = TRUE)
```

My data set

```
Rows: 5
```

dat	[Date]	DAT	maxL: 19	W:	D:	SASf:	yymmdd10.	T:	N:	4	NA:	1	U:	5		01Feb2011, 21Apr2010, NA, 02Nov1999, 15Oct2000
dt	[Date & Time (ISO 8601)]	DAT	maxL: 19	W:	D:	SASf:	e8601dt19.	T:	N:	4	NA:	1	U:	5		2011-02-01 08:00:00, 2010-04-21 12:30:00, NA, 1 ...
dv	[Double value]	dbl	maxL: 18	W:	D:2	SASf:	8.2	T:	N:	4	NA:	1	U:	5		Min -0.65, Q1 -0.44, Me 0, Q3 1.23, Max 2.24   ...
int	[Integer]	int	maxL: 2	W:	D:0	SASf:		T:	N:	4	NA:	1	U:	5		Min 1, Q1 3, Me 5.5, Q3 6.5, Max 7   5, NA, 6, ...
pct	[Percentage]	dbl	maxL: 18	W:	D:1	SASf:		T:	N:	4	NA:	1	U:	5		Min 0.05, Q1 0.16, Me 0.29, Q3 0.34, Max 0.37   ...
tim	[Time]	DAT	maxL: 19	W:	D:	SASf:		T:	N:	4	NA:	1	U:	5		08:00, 12:30, NA, 14:15, 07:30
trt	[Translated]	dbl	maxL: 2	W:	D:2	SASf:		T:Y	N:	4	NA:	1	U:	5		Min 1, Q1 1.5, Me 2.5, Q3 3.5, Max 4   (MyDict) ...
unlabeled		txt	maxL: 2	W:	D:	SASf:		T:	N:	4	NA:	1	U:	5		NA, A, B, C, D
vlt	[Very long text]	txt	maxL: 25	W: 15	D:	SASf:	\$15.	T:	N:	4	NA:	1	U:	5		This is a long string, Another long string, Muc ...



# Metadata - example

Transfer attributes when datasets are copied via <- or =

```
keep.attr <- function(x)
{
 a <- attributes(x)
 a[c('names', 'row.names', 'class', 'dim', 'dimnames')] <- NULL
 a
}

keep <- function(.Data, ..., .Attr=NULL)
{
 cl <- union('keep', class(.Data))
 do.call('structure', c(list(.Data, class=cl, ...), .Attr))
}

' [.keep' <- function(.Data, ...)
 keep(NextMethod(), .Attr=keep.attr(.Data))

' [<-.keep' <- function(.Data, ...)
 keep(NextMethod(), .Attr=keep.attr(.Data))
```

```
> newData <- data # dataset "data" has attributes which will be copied
```

```
> DescribeTable(newData, displayLabels=TRUE, displaySASFormats = TRUE)
```

My data set

```
Rows: 5
```

dat	[Date]	DAT	maxL: 19 W:	D:	SASf: yymmdd10.	T:	N: 4 NA: 1 U:	5   01Feb2011, 21Apr2010, NA, 02Nov1999, 15Oct2000
dt	[Date & Time (ISO 8601)]	DAT	maxL: 19 W:	D:	SASf: e8601dt19.	T:	N: 4 NA: 1 U:	5   2011-02-01 08:00:00, 2010-04-21 12:30:00, NA, 1 ...
dv	[Double value]	dbl	maxL: 18 W:	D:2	SASf:	8.2 T:	N: 4 NA: 1 U:	5   Min -0.65, Q1 -0.44, Me 0, Q3 1.23, Max 2.24   ...
int	[Integer]	int	maxL: 2 W:	D:0	SASf:	T:	N: 4 NA: 1 U:	5   Min 1, Q1 3, Me 5.5, Q3 6.5, Max 7   5, NA, 6, ...
pct	[Percentage]	dbl	maxL: 18 W:	D:1	SASf:	T:	N: 4 NA: 1 U:	5   Min 0.05, Q1 0.16, Me 0.29, Q3 0.34, Max 0.37   ...
tim	[Time]	DAT	maxL: 19 W:	D:	SASf:	T:	N: 4 NA: 1 U:	5   08:00, 12:30, NA, 14:15, 07:30
trt	[Translated]	dbl	maxL: 2 W:	D:2	SASf:	T:Y	N: 4 NA: 1 U:	5   Min 1, Q1 1.5, Me 2.5, Q3 3.5, Max 4   (MyDict) ...
unlabeled		txt	maxL: 2 W:	D:	SASf:	T:	N: 4 NA: 1 U:	5   NA, A, B, C, D
vlt	[Very long text]	txt	maxL: 25 W: 15	D:	SASf:	\$15.	T:	N: 4 NA: 1 U: 5   This is a long string, Another long string, Muc ...



# Metadata – reading \*.sasXmlMap

With R one can easily parse additional XML files containing definition of metadata.

*# Quick and dirty reading labels from the \*.sasXmlMap file. This can be easily extended.*

```
ApplyLabelsFromSasXmlMap <- function(collectionOfTables, pathToSasXmlMap) {

 nameOfTablesCollection <- deparse(substitute(collectionOfTables))

 if(file.exists(pathToSasXmlMap)) {

 DOM <- xmlParse(pathToSasXmlMap)

 for(tab in names(collectionOfTables)) {

 for(col in toupper(colnames(collectionOfTables[[tab]]))) {

 labelList <- xpathApply(DOM, paste0("//TABLE[@name = '", tab, "']/COLUMN[@name = '", col,
 "']/DESCRIPTION"), xmlValue)

 if(length(labelList) > 0) {

 cat(paste("Table:", tab, " Column:", col, "=", labelList[[1]], "\n"))

 SetSVarAttributes(varName = paste0(nameOfTablesCollection, "$", tab, "$",
 tolower(col)), label = labelList[[1]])

 }

 }

 }

 }

}
```







# Metadata – creating SAS Transport File

Having metadata attached to a data.frame and using [SASxport](#) package SAS Transport files can be created easily.

```
> MakeSASTransportFile(data, "c:/tmp/test.xpt")
```

```
Processing: test.xpt
```

```
*** Setting metadata... ***
```

```
Processing column: unlabeled of class: character
```

```
Processing column: vlt of class: character | Specified width: 15 | Assigned SAS format: $15.
```

```
Processing column: dv of class: numeric | Specified decimal places: 2 | Assigned SAS format: 8.2
```

```
Processing column: pct of class: numeric | Specified decimal places: 1 | Assigned DEFAULT NUMERIC SAS format: Best12.
```

```
Processing column: int of class: integer | Assigned DEFAULT NUMERIC SAS format: Best12.
```

```
Processing column: dt of class: POSIXct | Specified format: %Y-%m-%d %H:%M:%S | Assigned SAS format: e8601dt19.
```

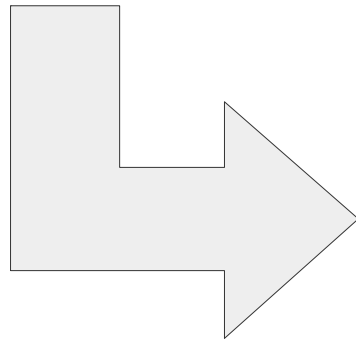
```
Processing column: dat of class: POSIXct | Specified format: %d%b%Y | Assigned SAS format: yymmdd10.
```

```
Processing column: tim of class: POSIXct | Specified format: %H:%M
```

```
Processing column: trt of class: numeric | Assigned DE
```

```
*** Exporting the dataset into SAS transport file..
```

```
*** Done ***
```



VIEWTABLE: Work.Data

	UNLABLE	Very long text	Double value	Percentage	Integer	Date & Time (ISO 8601)
1		This is a long	.	0.2631578947	5	2011-02-01T08:00:00
2	A	Another long st	2.24	.	.	2010-04-21T12:30:00
3	B	Much longer tha	-0.23	0.3157894737	6	.
4	C	A short string	0.22	0.3684210526	7	1999-11-02T14:15:00
5	D		-0.65	0.0526315789	1	2000-10-15T07:30:00

Column Attributes

General | Colors | Fonts | Close

Name: DT

Label: Date & Time (ISO 8601)

Length: 8

Format: E8601DT19. ...

Informat: DATETIME19. ...

Type  
 Character  
 Numeric

Apply Help



- I Handling metadata
- II Handling ODM and CDA metadata**
- III Issue with multilingual data
- IV Differences between SAS and R
- V Implementation of useful SAS functions





# Handling ODM and CDA metadata

By using the [ODMconverter](#) package one can work with [ODM](#) and CDA formats in R. Both formats can be reciprocally translated into each other. There are also functions for creating R data.frames decorated with ODM metadata as well as creating ODM definitions based on metadata bound to an existing R data.frame.



# Handling ODM and CDA metadata

ODM-Test.xlsx - Excel

PLIK NARZĘDZI WSTAWIA UKŁAD STR FORMUŁY DANE RECENZJA WIDOK DEWLOPE POWERPIV TEAM

A1 : X ✓ fx StudyOID

	A	B	C	D
1	StudyOID	S.0000		
2	Sponsor	Testsponsor		
3	Condition	Testcondition		
4	StudyName	ODM Test Study		
5	StudyDescription	Test of ODM tools		
6	Form	ODM-Test		
7	FirstName	Test		
8	LastName	Testname		
9	Organization	Test organization		
10				
11	Type	Name		
12	itemgroup	General Info		
		Willingness to participate		
		clinical trials		
13	boolean			
14	integer	Age		
15	date	Date of Birth		
16	integer	Gender		
17	codelistitem	1		
18	codelistitem	2		
19	string	Diagnosis text		
20	string	Diagnosis code		
21	float	Creatinine		
22	time	Time of lab value		
23				

Sheet1

GOTOWY

XLSx → R

RGui (64-bit) - [R Console]

```

> attributes(odmdata)
$names
[1] "I.1001" "I.1002" "I.1003" "I.1004"

$row.names
integer(0)

$StudyOID
[1] "S.0000"

$Sponsor
[1] "Testsponsor"

$Condition
[1] "Testcondition"

$StudyName
[1] "ODM Test Study"

$StudyDescription
[1] "Test of ODM tools"

$Form
[1] "ODM-Test"

```

Lister - [c:\Users\aaolsz\Documents\R\win-l...

```

File Edit Options Encoding Help 52 %
odmdata <- data.frame(
 I.1001=logical(),
 I.1002=integer(),
 I.1003=character(),
 I.1004=integer(),
 I.1005=character(),
 I.1006=character(),
 I.1007=numeric(),
 I.1008=character(),
 stringsAsFactors = F)

attr(odmdata, "StudyOID") <- "S.0000"
attr(odmdata, "Sponsor") <- "Testsponsor"
attr(odmdata, "Condition") <- "Testcondition"
attr(odmdata, "StudyName") <- "ODM Test Study"
attr(odmdata, "StudyDescription") <- "Test of ODM tools"
attr(odmdata, "Form") <- "ODM-Test"
attr(odmdata, "FormID") <- "F.1"

```

R → ODM

Lister - [c:\Users\aaolsz\Documents\R\win-l...

```

File Edit Options Encoding Help 17 %
<?xml version="1.0" encoding="UTF-8" ?>
<ODM xmlns="http://www.cdisc.org/ns/odm/v1.3"
Description="ODM-Test S.0000 Testcondition"
ODMVersion="1.3.1" CreationDateTime="2013-05-27T21:11:11"
FileOID="ODM-Test S.0000.xml"
FileType="Snapshot">
<Study OID="S.0000">
<GlobalVariables>
<StudyName>ODM Test Study</StudyName>
<StudyDescription>Test of ODM tools</StudyDescription>
<ProtocolName>Testsponsor</ProtocolName>
</GlobalVariables>
<BasicDefinitions>
<MeasurementUnit OID="MU.1" Name="unit">
<Symbol><TranslatedText xml:lang="en">Unit</TranslatedText>
</MeasurementUnit>
</BasicDefinitions>
<MetaDataVersion OID="MD.1" Name="Metadataversion">
<Protocol><StudyEventRef StudyEventOID="SE.1" OrderNumber="1">
<StudyEventDef OID="SE.1" Name="Testcondition" Repeatable="Y">
<FormRef FormOID="F.1" OrderNumber="1" Mandatory="Y">
</StudyEventDef>

```



- I Handling metadata
- II Handling ODM and CDA metadata
- III Issue with multilingual data**
- IV Differences between SAS and R
- V Implementation of useful SAS functions





# Issue with multilingual data

If multilingual data has to be processed, the best choice is to do it under Linux or MacOS. Both are unicode-enabled operating systems in contrast to Windows.

## Windows

```

RGui (32-bit) - [R Console]
File Edit View Misc Packages Windows Help
> sessionInfo()Rplatform
[1] "i386-w64-mingw32"
> data.frame(LANG=c("PL", "DE", "FR", "DU", "TUR", "RU$
+ Text=c("ąęóąśłźźćń",
+ "äöüß",
+ "àââûüæçÿ€éèèëïîô",
+ "éëïóöü€",
+ "çğıİöşü",
+ "чяюф"))
 LANG Text
1 PL ąęóąśłźźćń
2 DE äöüß
3 FR àââûüæçÿ€éèèëïîô
4 DU éëïóöü€
5 TUR çğıİöşü
6 RU <U+0447><U+044F><U+044E><U+0444>
> Sys.getlocale("LC_ALL")
[1] "LC_COLLATE=Polish_Poland.1250;LC_CTYPE=Polish_Pol$
> Sys.setlocale("LC_ALL", 'pl_PL.UTF-8')
[1] ""
Warning message:
In Sys.setlocale("LC_ALL", "pl_PL.UTF-8") :
 OS reports request to set locale to "pl_PL.UTF-8" ca$

```



## Linux

```

Terminal
Plik Edycja Widok Wyszukiwanie Terminal Pomoc
> sessionInfo()Rplatform
[1] "x86_64-pc-linux-gnu"
> data.frame(LANG=c("PL", "DE", "FR", "DU", "TUR", "RU"),
+ Text=c("ąęóąśłźźćń",
+ "äöüß",
+ "àââûüæçÿ€éèèëïîô",
+ "éëïóöü€",
+ "çğıİöşü",
+ "зчяющылф"))
 LANG Text
1 PL ąęóąśłźźćń
2 DE äöüß
3 FR àââûüæçÿ€éèèëïîô
4 DU éëïóöü€
5 TUR çğıİöşü
6 RU зчяющылф
> Sys.getlocale("LC_ALL")
[1] "LC_CTYPE=pl_PL.utf8;LC_NUMERIC=C;LC_TIME=pl_PL.utf8;LC_CO
LLATE=pl_PL.utf8;LC_MONETARY=pl_PL.utf8;LC_MESSAGES=pl_PL.utf8
;LC_PAPER=pl_PL.utf8;LC_NAME=C;LC_ADDRESS=C;LC_TELEPHONE=C;LC_
MEASUREMENT=pl_PL.utf8;LC_IDENTIFICATION=C"
> █

```





- I Handling metadata
- II Handling ODM and CDA metadata
- III Issue with multilingual data
- IV Differences between SAS and R**
- V Implementation of useful SAS functions





# Differences between SAS and R

It is worth noting that:

- R and SAS use different types of sum of square, type I and III respectively
  - R can be set to calculate SS of the same type (in fact – of any type)
- R uses different default contrast coding than SAS
  - R can be set to use the same contrasts (contr.SAS)
- Dates of origin of POSIX date/time differ across both packages:
  - in R: 1970-01-01 00:00:00 UTC
  - in SAS: 1960-01-01 00:00:00 GMT
- R uses different algorithm for calculating quantiles. Luckily, SAS-compliant algorithm is implemented (type 3)





# further **R** impo**R**tant issues

- I Handling metadata
- II Handling ODM and CDA metadata
- III Issue with multilingual data
- IV Differences between SAS and R
- V Implementation of useful SAS functions**





# R implementation of useful SAS functions #1

```
FirstRowsBy <- function(dataFrame, byColumns) {
 DT <- data.table(dataFrame, key=byColumns)
 as.data.frame(DT[unique(DT[, key(DT), with = FALSE]), mult = 'first'])
}

LastRowsBy <- function(dataFrame, byColumns) {
 DT <- data.table(dataFrame, key=byColumns)
 as.data.frame(DT[unique(DT[, key(DT), with = FALSE]), mult = 'last'])
}

MarkFirstLastBy <- function(dataFrame, byColumns) {
 idName <- paste0("ID", paste0(sample(c(LETTERS, letters, 0:9), size=5), collapse=""))
 dataFrame[, idName] <- seq_len(nrow(dataFrame))

 firstIDs <- FirstRowsBy(dataFrame = dataFrame, byColumns = byColumns)[, idName]
 lastIDs <- LastRowsBy(dataFrame = dataFrame, byColumns = byColumns)[, idName]

 dataFrame$FIRST <- dataFrame[, idName] %in% firstIDs
 dataFrame$LAST <- dataFrame[, idName] %in% lastIDs

 dataFrame[, idName] <- NULL
 return(dataFrame)
}
```

# R implementation of useful SAS functions #1

```
RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help
[Icons: Home, Copy, Paste, Undo, Redo, Stop, Print]

> data
 ID Country Sex Age
1 1 PL F 43
2 3 PL F 37
3 2 PL M 25
4 5 RU F 33
5 6 RU F 27
6 4 RU M 63
7 9 EN F 48
8 7 EN M 46
9 8 EN M 52

> FirstRowsBy(data, c("Country", "Sex"))
 ID Country Sex Age
1 1 PL F 43
2 2 PL M 25
3 5 RU F 33
4 4 RU M 63
5 9 EN F 48
6 7 EN M 46

> LastRowsBy(data, c("Country", "Sex"))
 ID Country Sex Age
1 3 PL F 37
2 2 PL M 25
3 6 RU F 27
4 4 RU M 63
5 9 EN F 48
6 8 EN M 52

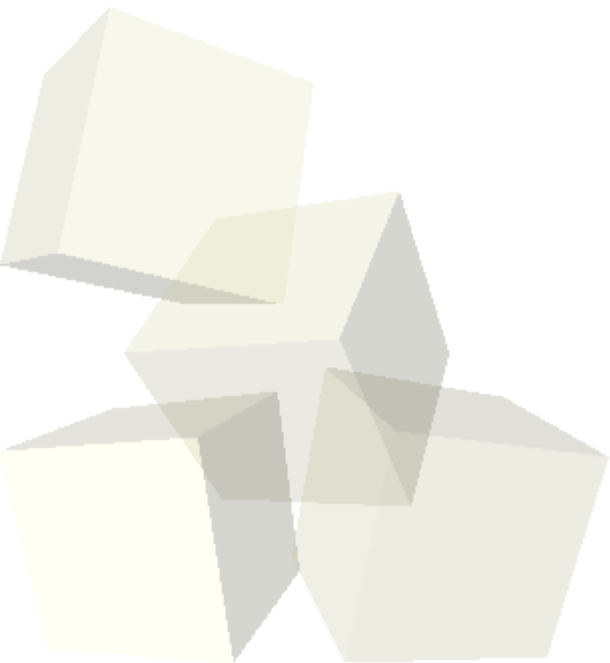
> MarkFirstLastBy(data, c("Country", "Sex"))
 ID Country Sex Age FIRST LAST
1 1 PL F 43 TRUE FALSE
2 3 PL F 37 FALSE TRUE
3 2 PL M 25 TRUE TRUE
4 5 RU F 33 TRUE FALSE
5 6 RU F 27 FALSE TRUE
6 4 RU M 63 TRUE TRUE
7 9 EN F 48 TRUE TRUE
8 7 EN M 46 TRUE FALSE
9 8 EN M 52 FALSE TRUE
```



Undoubtedly, R **does meet** all the requirements involved in biostatistician's work.

It is **perfectly fitted** for this purpose.

R can entirely **replace** SAS as well as **cooperate** with.





Thank you for your attention!

Please find the latest version of this document at:

<http://www.scribd.com/AdrianOlszewski>

or, in HTML version at:

<http://www.r-clinical-research.com>

This presentation was published first time on Scribd on April 1<sup>st</sup>, 2014

Last update: 13<sup>th</sup> June, 2015

HTML version was created with [PDF2HTML](#) provided by Lu Wang