a not-so-short answer to the question...

...is

suitable enough for biostatisticians

involved in Clinical Research & Evidence-Based Medicine?

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

Adrian Olszewski
r.clin.res@gmail.com

WeRcome!

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

Agenda

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 ½ :) Books
V R is supported by the business
V ½ :) 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 ½ :) 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

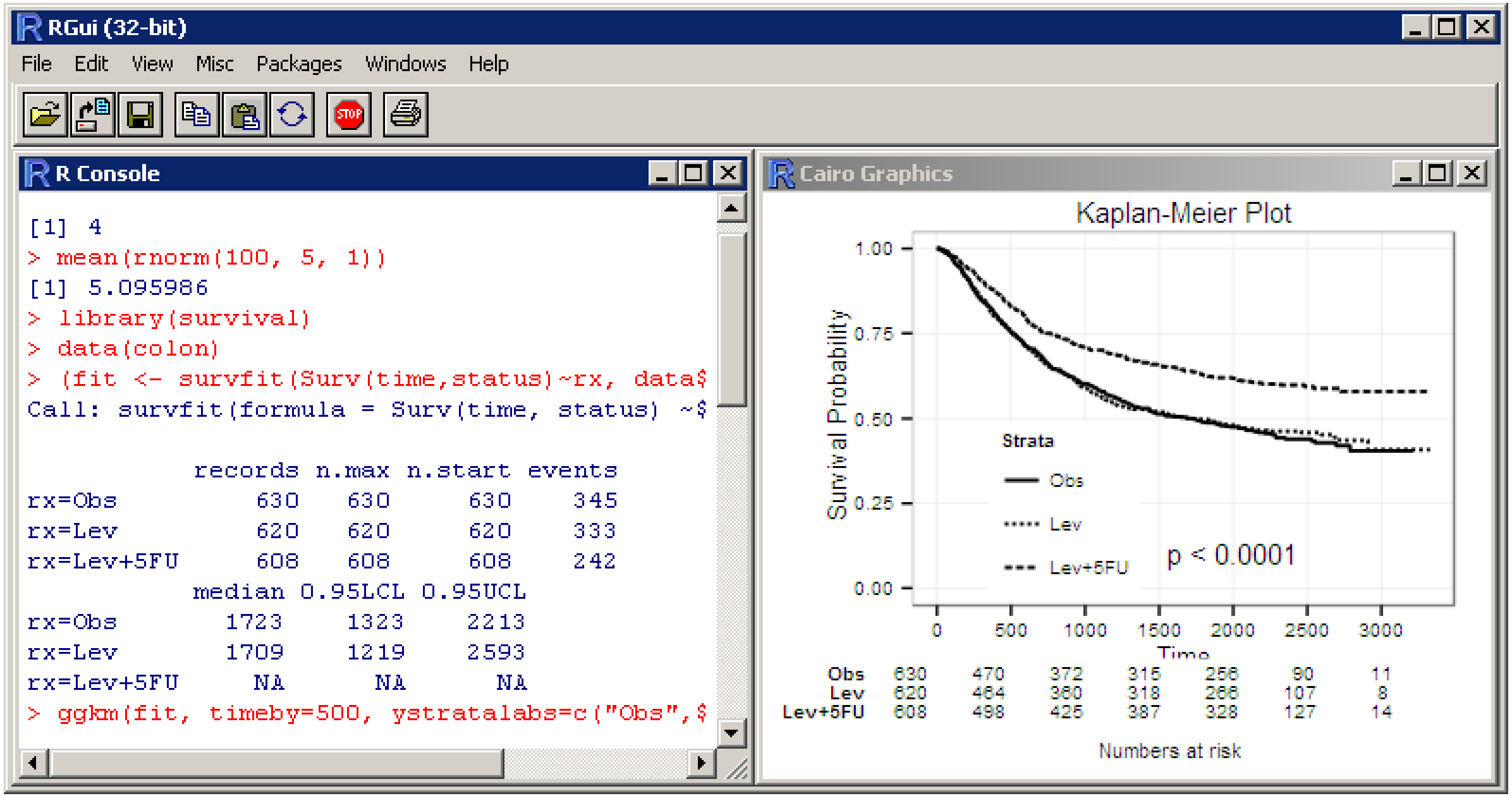
I Handling metadata
II Handling ODM and CDA metadata
III Issue with multilingual data
IV Differences between SAS and R
V Implementation of useful SAS functions

IntRoduction

...so, what is "R"?

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.



```
[1] 4
> mean(rnorm(100, 5, 1))
[1] 5.095986
> library(survival)
> data(colon)
> (fit <- survfit(Surv(time,status)~rx, data$
Call: survfit(formula = Surv(time, status) ~$

            records n.max n.start events
rx=Obs          630   630     630    345
rx=Lev          620   620     620    333
rx=Lev+5FU      608   608     608    242
            median 0.95LCL 0.95UCL
rx=Obs        1723    1323    2213
rx=Lev        1709    1219    2593
rx=Lev+5FU      NA      NA      NA
> ggkm(fit, timeby=500, ystratalabs=c("Obs",$
```

is it a popular softwaRe ?

[...] 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

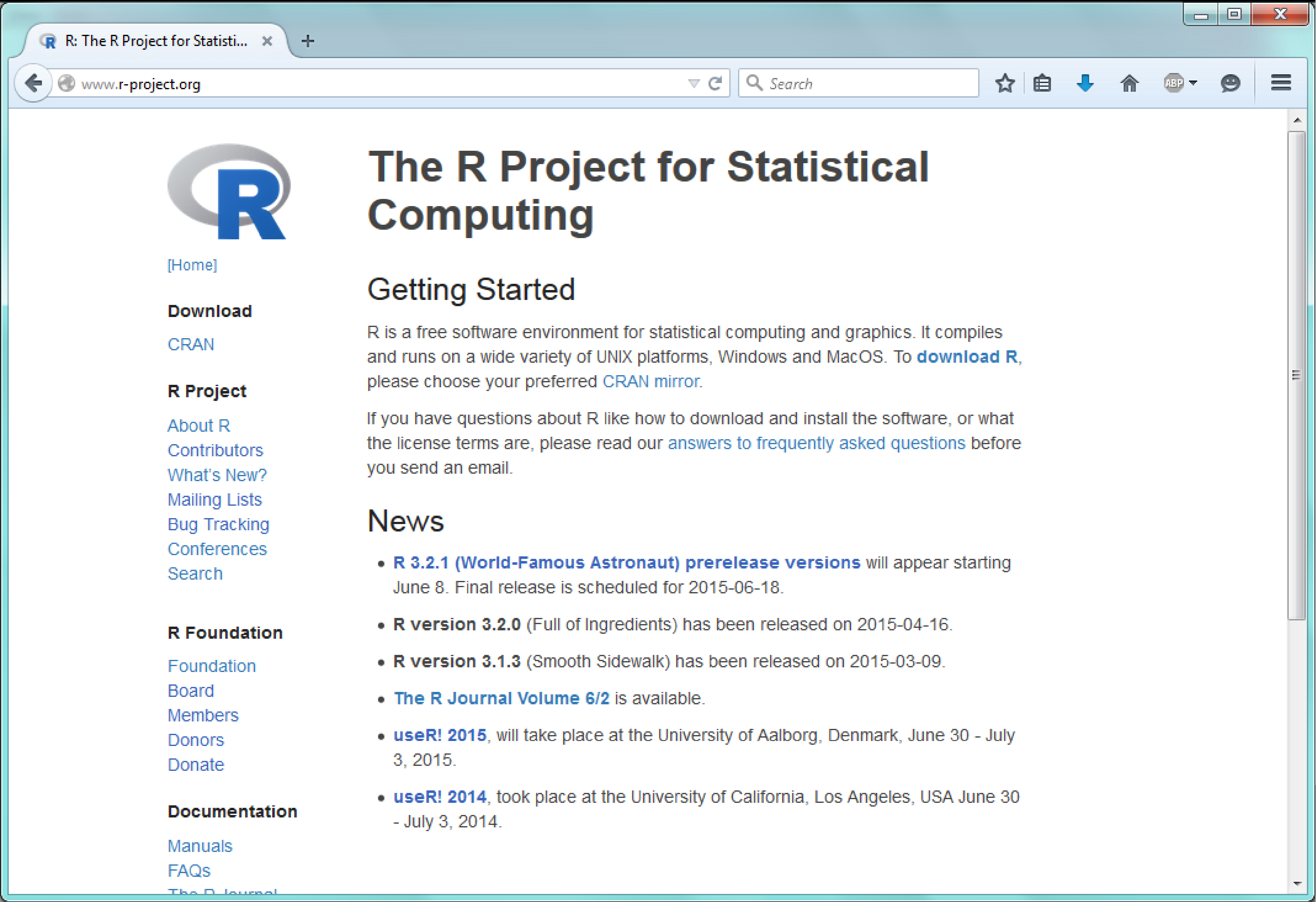
…so, what is “R”?

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)

http://www.r-project.org

…so, what is "R"?



[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

The R Project for Statistical Computing

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.

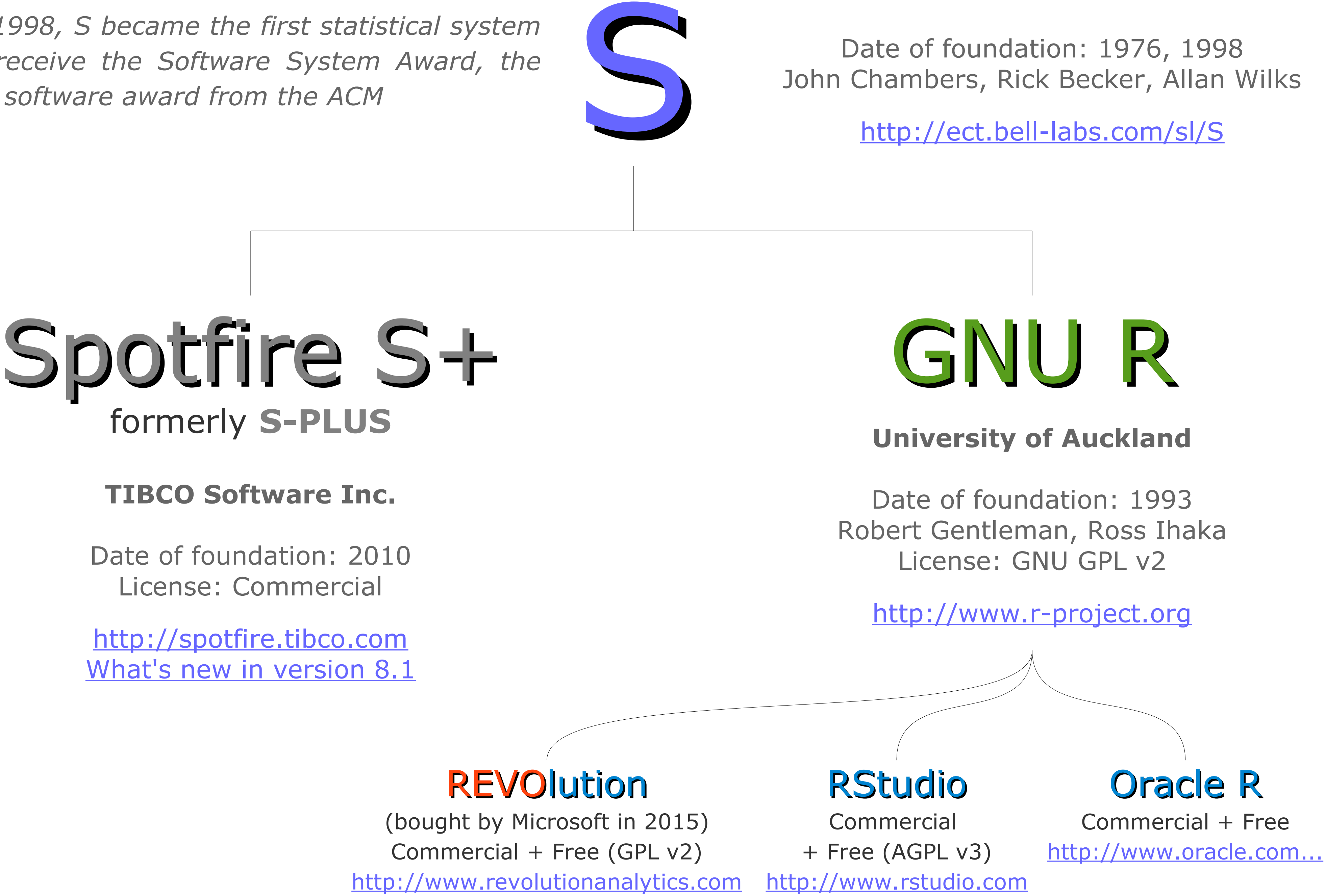
If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

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.

The R family

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



S

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...

did you know, that...

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

http://blog.revolutionanalytics.com/2014/04/seven-quick-facts-about-r.html

List of R users

They use(d) R

facebook

ebay

Bank of America

ANZ

Canada
Health Canada

mozilla
FOUNDATION

Google

MERCK

NOVARTIS

AstraZeneca

Microsoft

AMGEN

UBER

KCR

Pfizer

Allianz

DEPARTMENT OF HEALTH & HUMAN SERVICES • USA

FDA

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

BAYER

MOLINA HEALTHCARE

BIOSITE
NEW DIMENSIONS IN DIAGNOSIS

CARDIODX

Ford

Roche

benetech
TECHNOLOGY SERVING HUMANITY

European Environment Agency

KICKSTARTER

Pacific Northwest
NATIONAL LABORATORY

NewScientist

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.

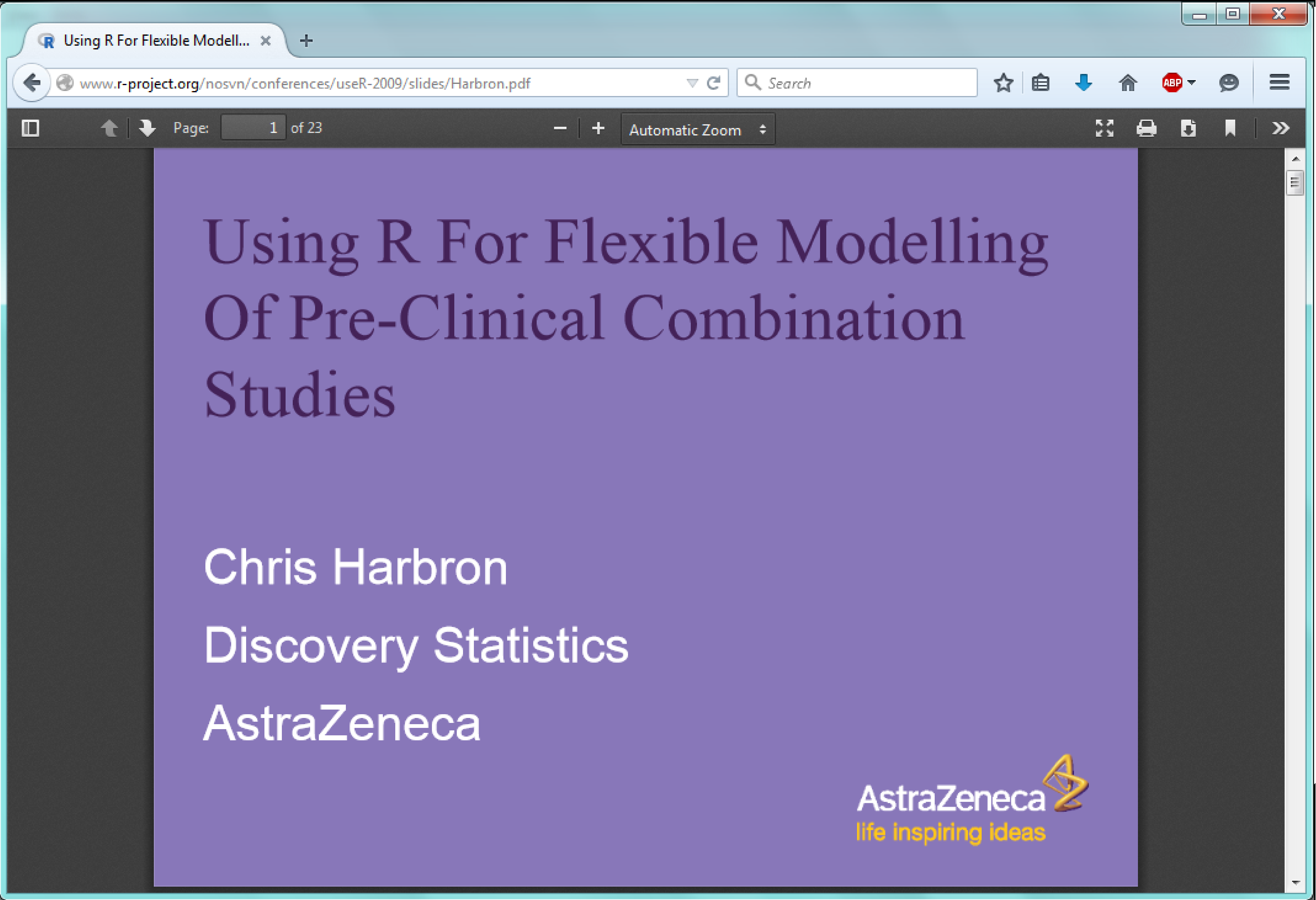
They use R

“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



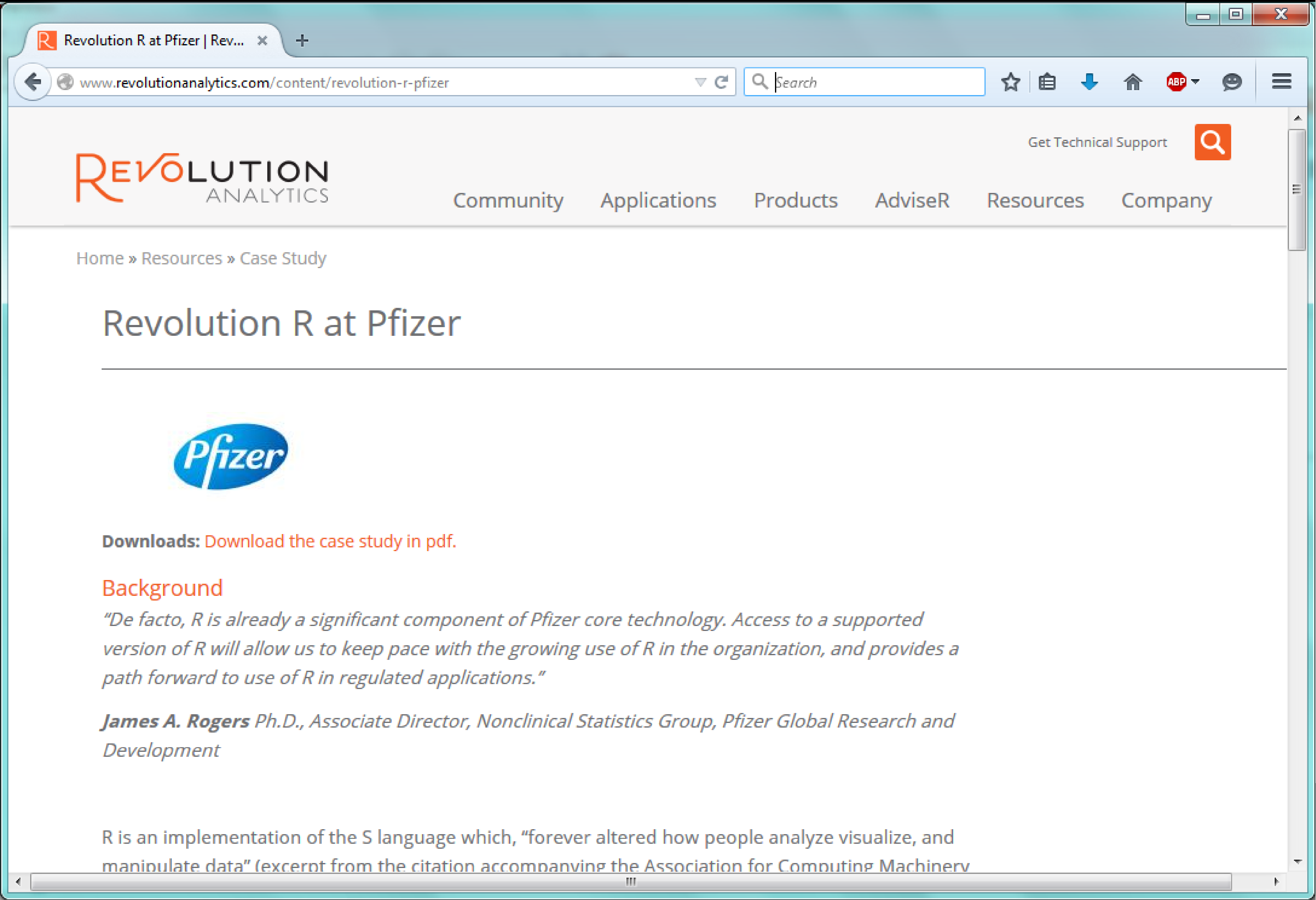
Using R For Flexible Modelling Of Pre-Clinical Combination Studies

Chris Harbron

Discovery Statistics

AstraZeneca

They use R



Revolution R at Pfizer

Downloads: Download the case study in pdf.

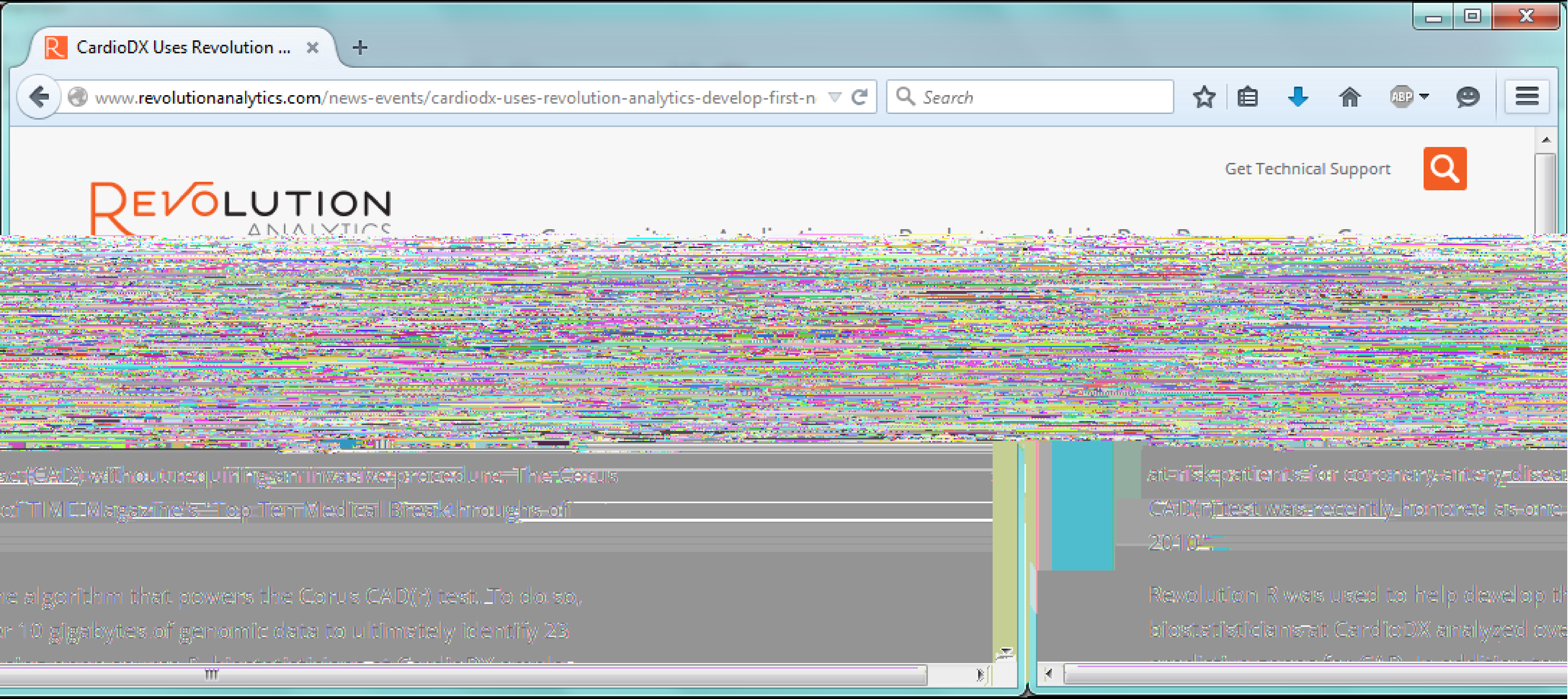
Background

"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



ase (CAD) without requiring an invasive procedure. The Corus
of TIME Magazine's "Top Ten Medical Breakthroughs of

e algorithm that powers the Corus CAD(r) test. To do so,
r 10 gigabytes of genomic data to ultimately identify 23

at-risk patients for coronary artery disea
CAD(r) test was recently honored as one
2010".

Revolution R was used to help develop th
biostatisticians at CardioDX analyzed ove

Advanced PK modeling with mixed effects

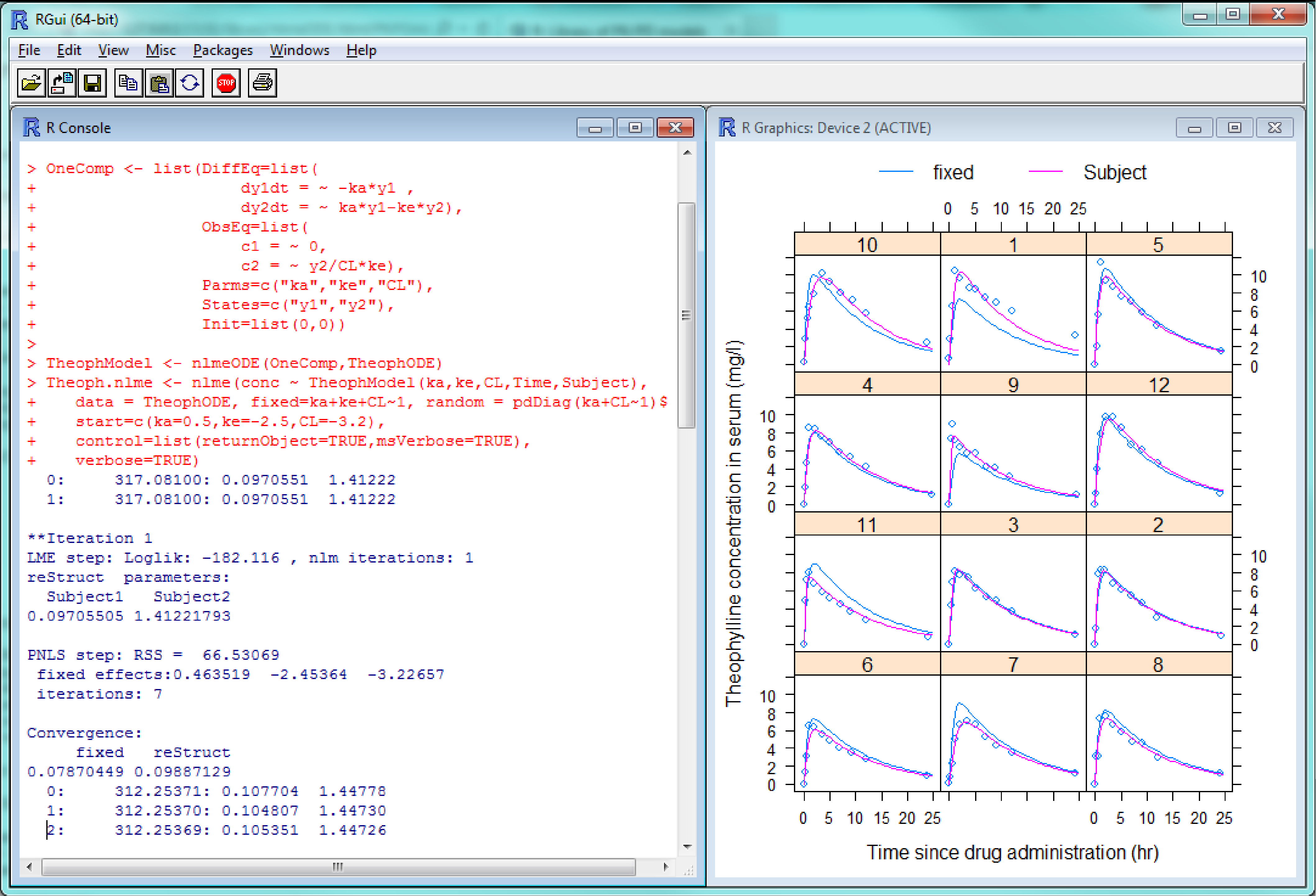
nlmeODE is a powerful combination of the famous nlme package with deSolve (an ordinary differential equations solver) for non-linear, mixed-effects modelling, designed especially for pharmaceutical research.

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



```
> OneComp <- list(DiffEq=list(
+                     dy1dt = ~ -ka*y1 ,
+                     dy2dt = ~ ka*y1-ke*y2),
+                 ObsEq=list(
+                     c1 = ~ 0,
+                     c2 = ~ y2/CL*ke),
+                 Parms=c("ka","ke","CL"),
+                 States=c("y1","y2"),
+                 Init=list(0,0))
>
> TheophModel <- nlmeODE(OneComp,TheophODE)
> Theoph.nlme <- nlme(conc ~ TheophModel(ka,ke,CL,Time,Subject),
+    data = TheophODE, fixed=ka+ke+CL~1, random = pdDiag(ka+CL~1)$
+    start=c(ka=0.5,ke=-2.5,CL=-3.2),
+    control=list(returnObject=TRUE,msVerbose=TRUE),
+    verbose=TRUE)
  0:     317.08100: 0.0970551  1.41222
  1:     317.08100: 0.0970551  1.41222

**Iteration 1
LME step: Loglik: -182.116 , nlm iterations: 1
reStruct  parameters:
  Subject1   Subject2
0.09705505 1.41221793

PNLS step: RSS =  66.53069
 fixed effects:0.463519  -2.45364  -3.22657
 iterations: 7

Convergence:
     fixed   reStruct
0.07870449 0.09887129
  0:     312.25371: 0.107704   1.44778
  1:     312.25370: 0.104807   1.44730
  2:     312.25369: 0.105351   1.44726
```

R + ADMB (automatic diffeRentiation!)

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

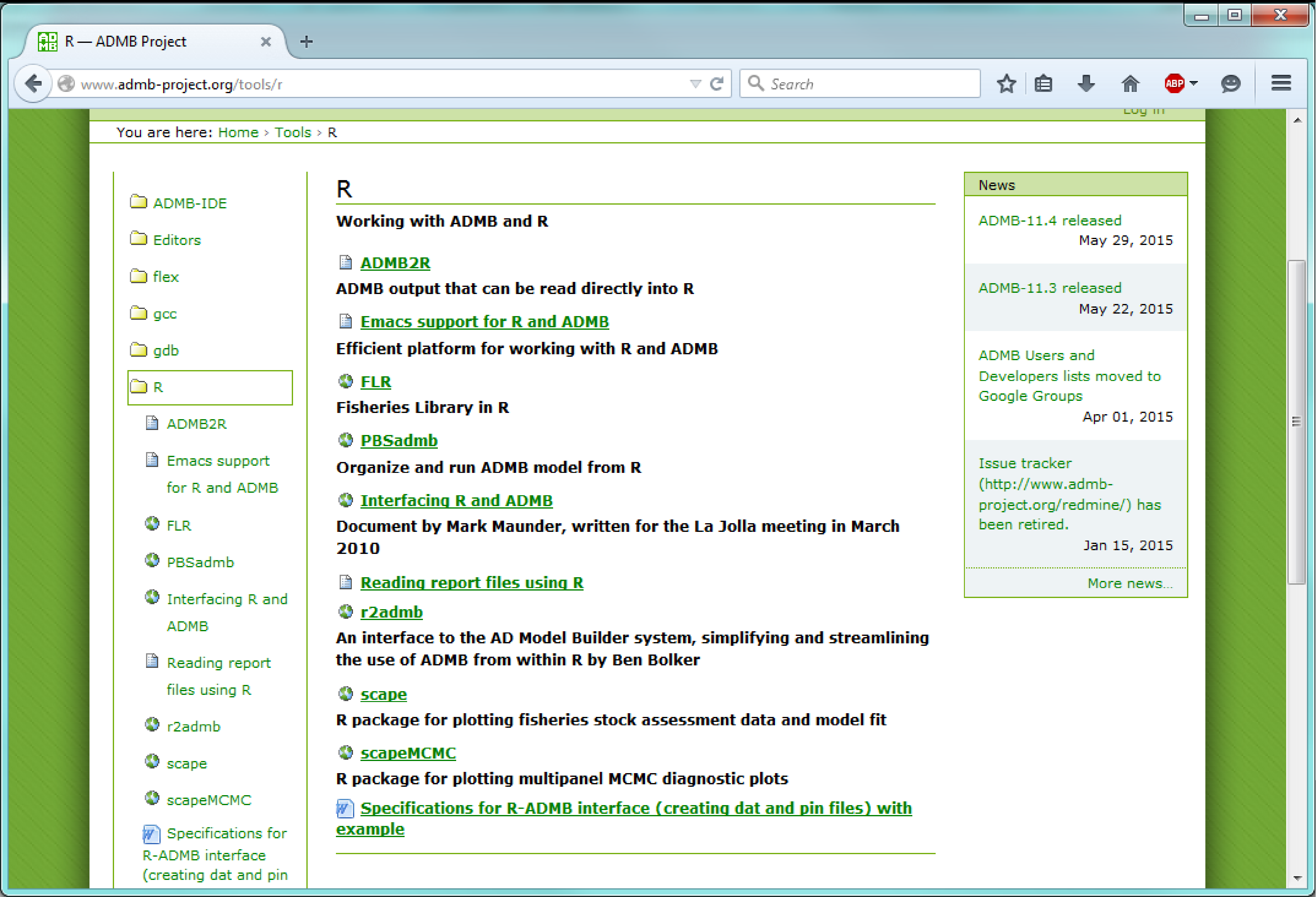
ADMB-11.3 released
May 22, 2015

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!)



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

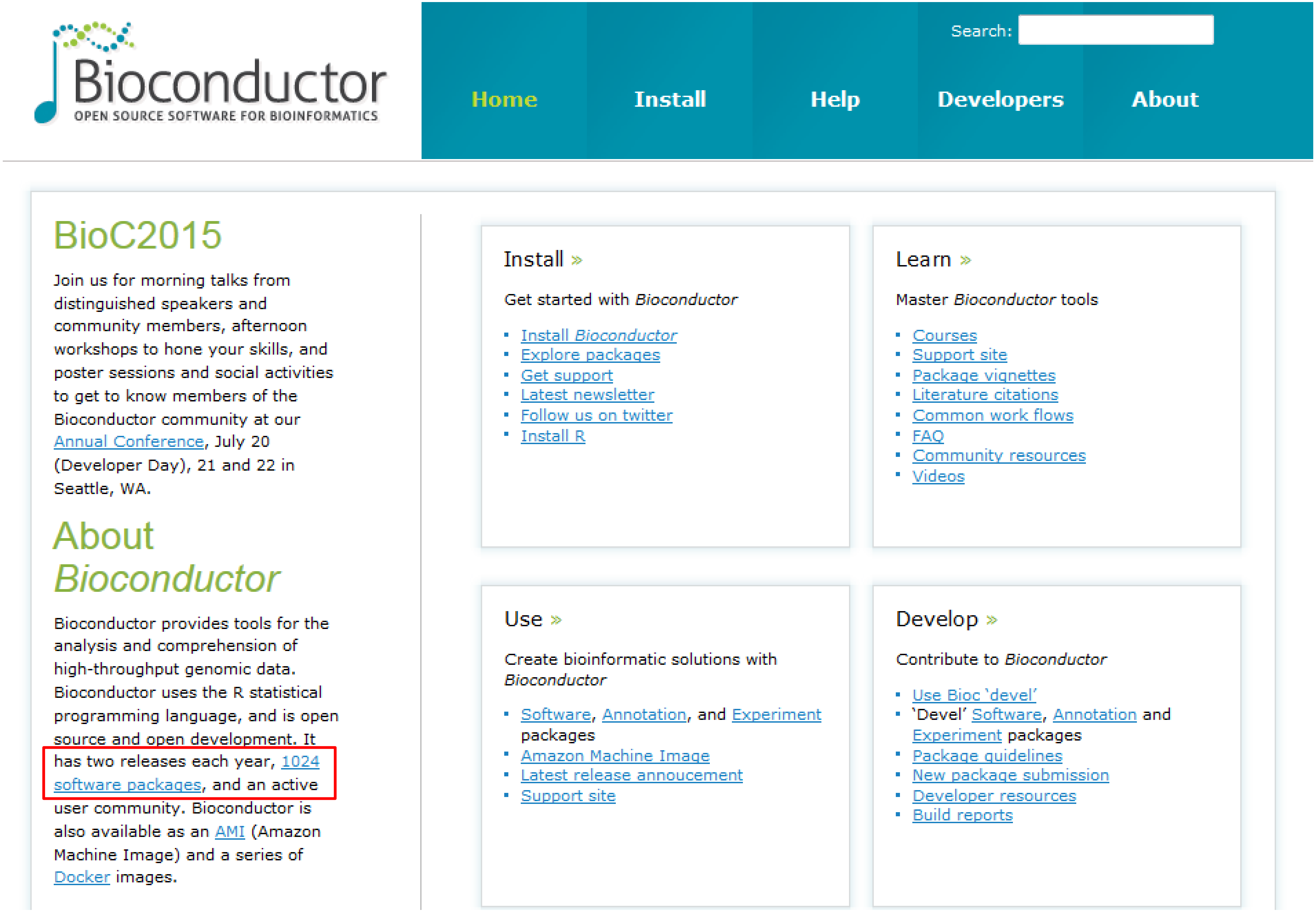
ADMB-11.3 released
May 22, 2015

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...

Genomic data analysis



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 annoucement
- Support site

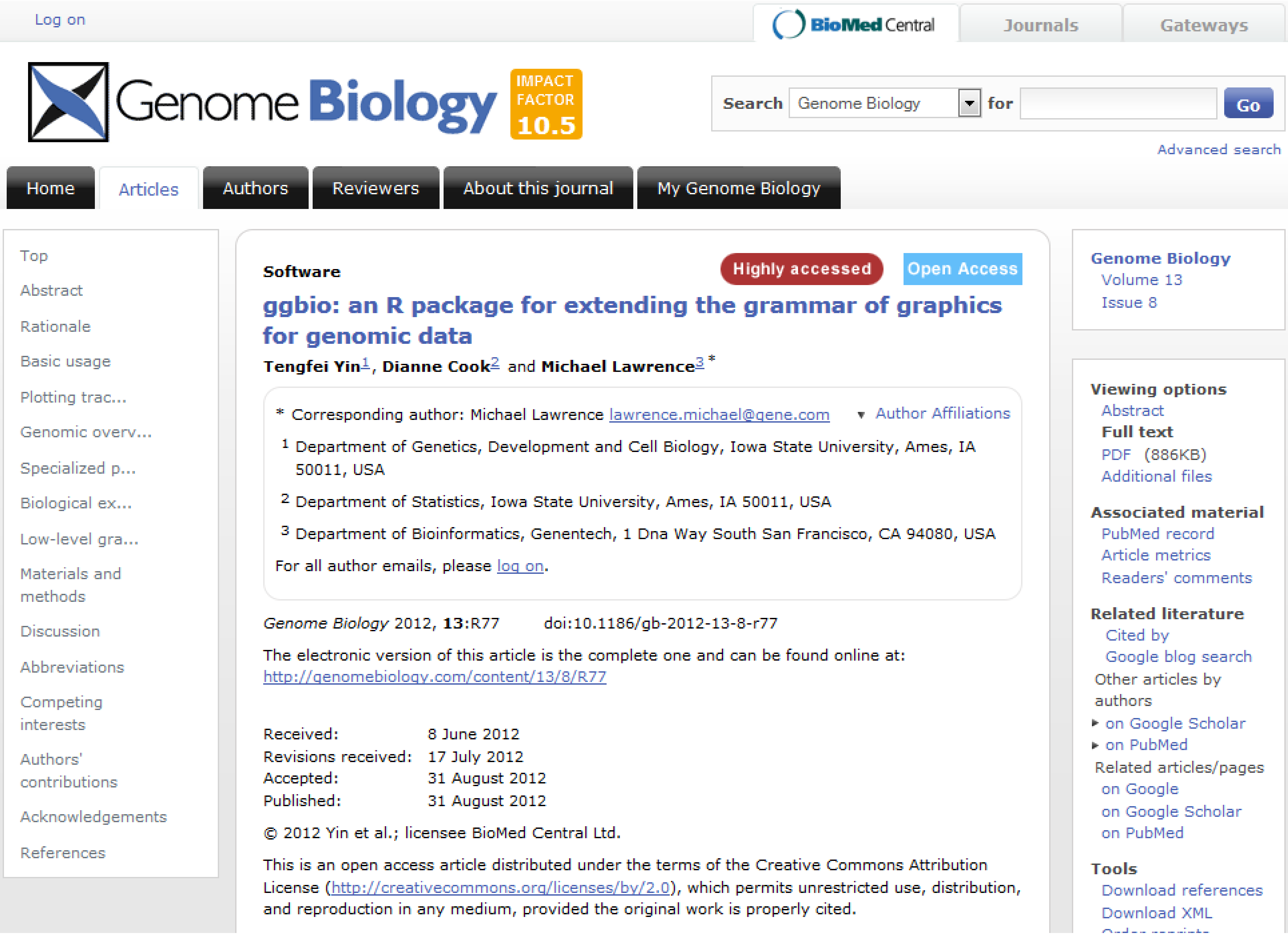
Develop »

Contribute to *Bioconductor*

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

http://tinyurl.com/bioconductor-packages

Genomic data analysis



Software

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

Tengfei Yin[1], **Dianne Cook**[2] and **Michael Lawrence**[3]*

* 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

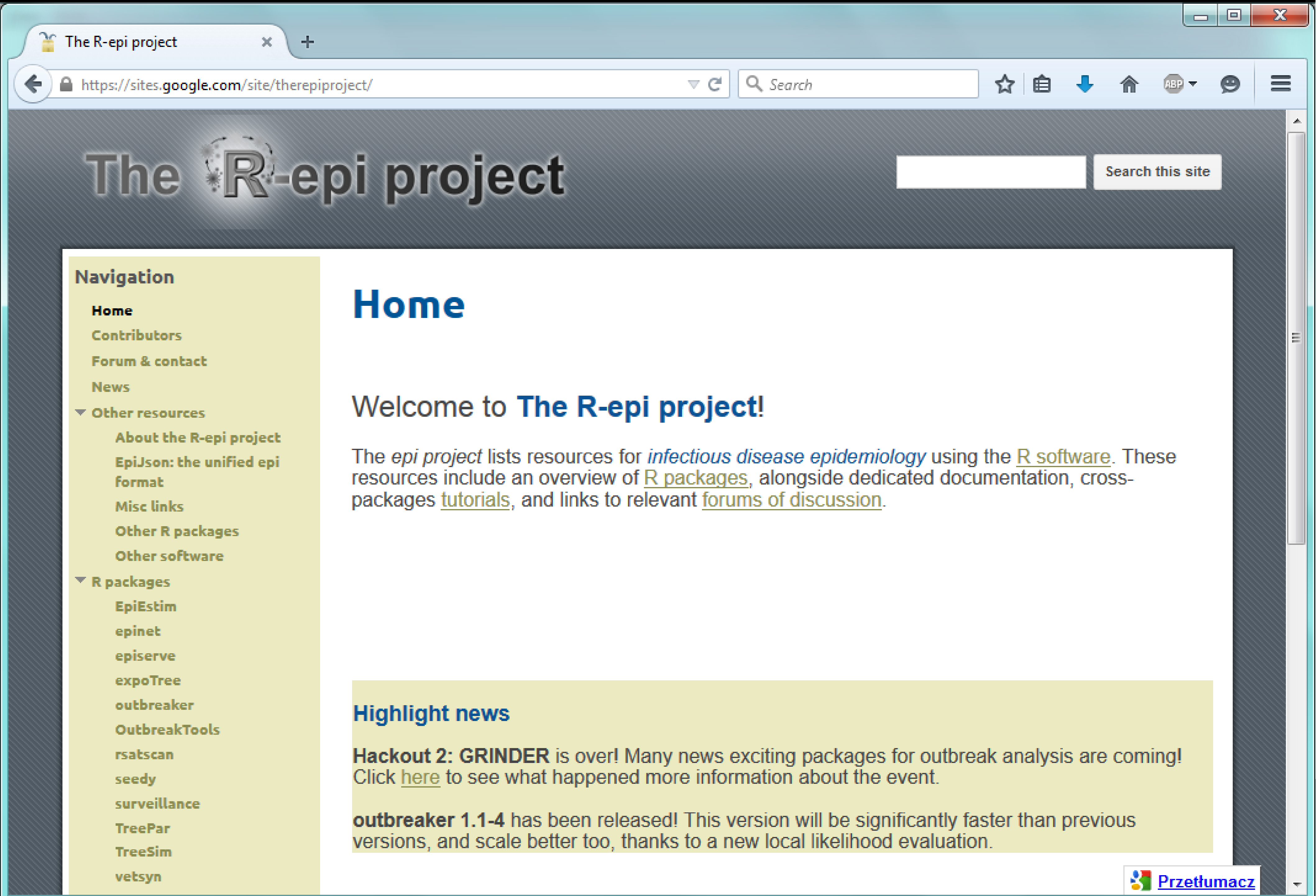
For all author emails, please log on.

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

R for Epidemiology



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!**

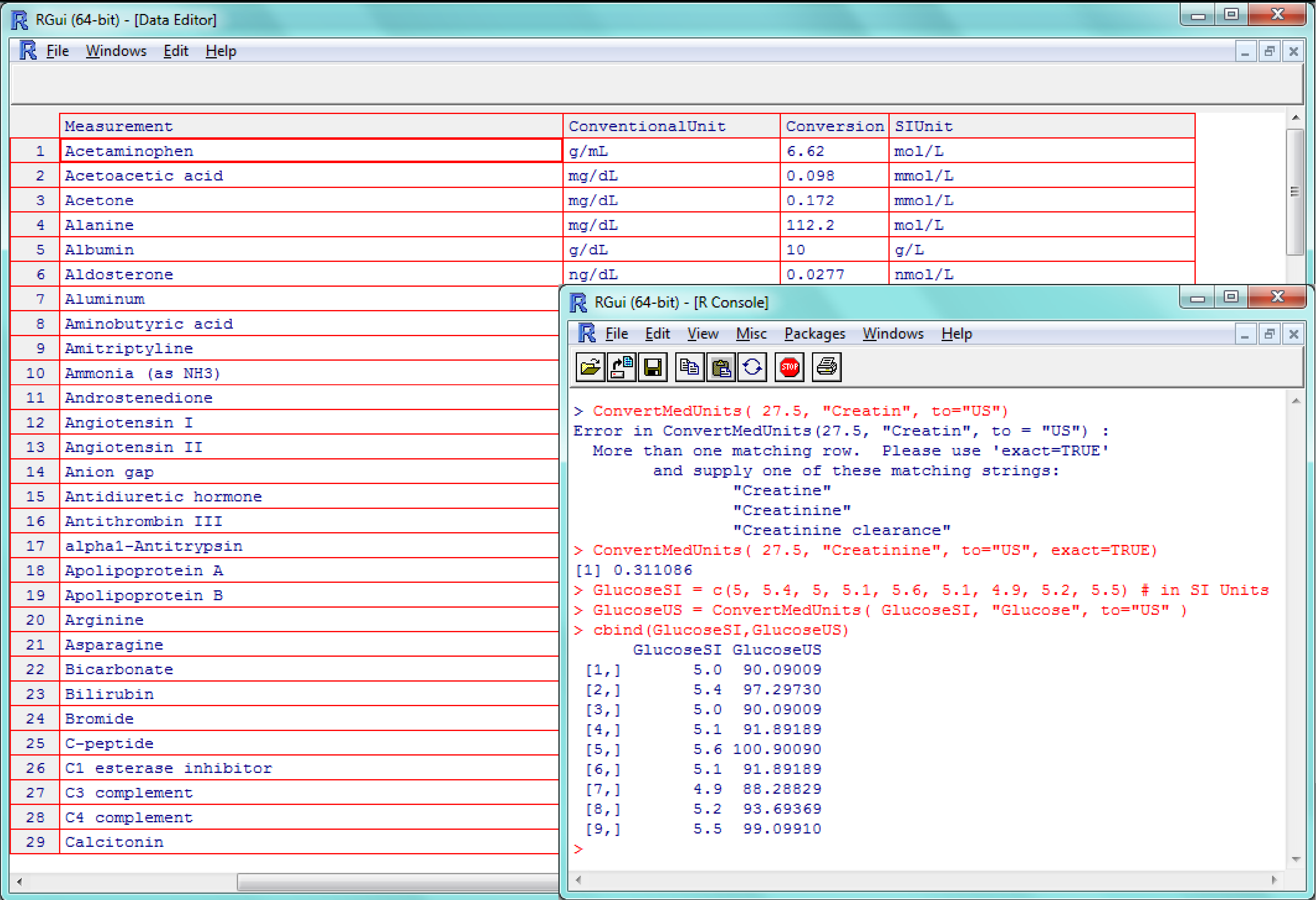
The *epi project* lists resources for *infectious disease epidemiology* using the R software. These resources include an overview of R packages, alongside dedicated documentation, cross-packages tutorials, and links to relevant forums of discussion.

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.

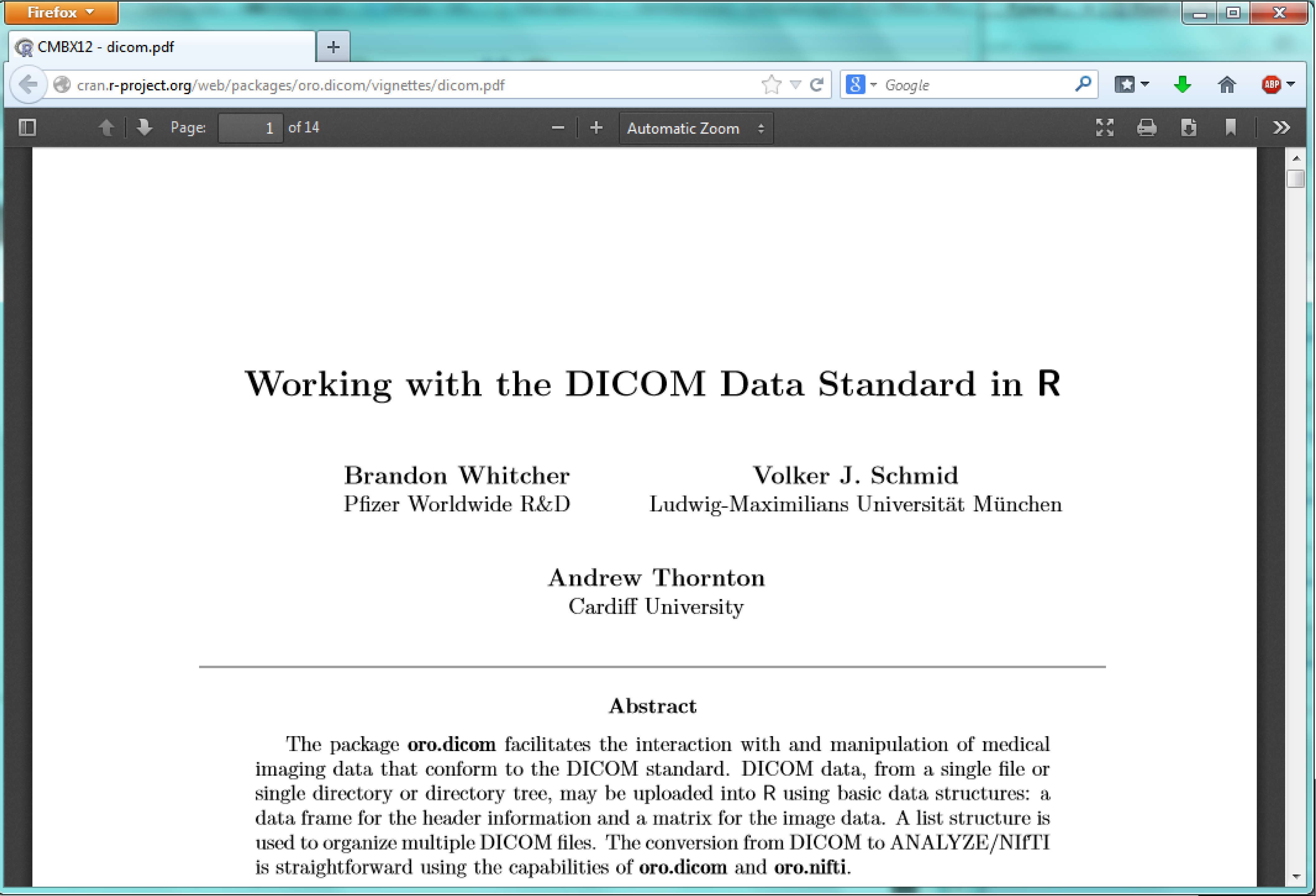
R for laboratory diagnostics



	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	alpha1-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			

```
> 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



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**

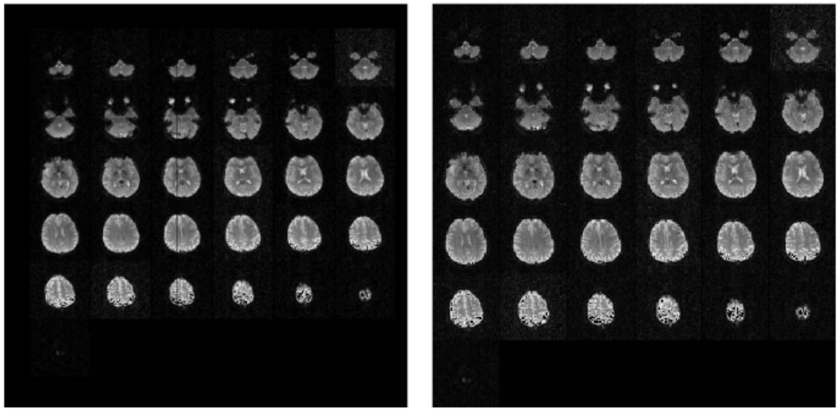

(a) (b)

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

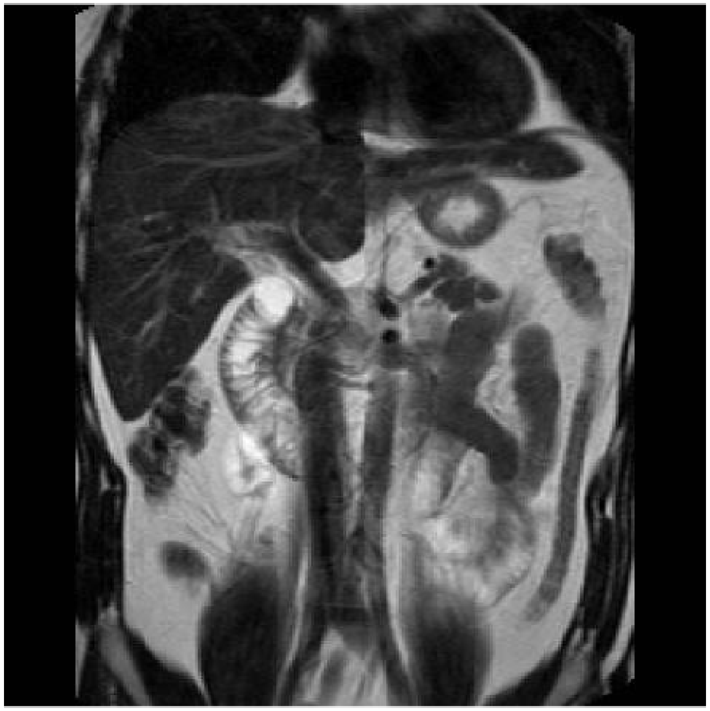
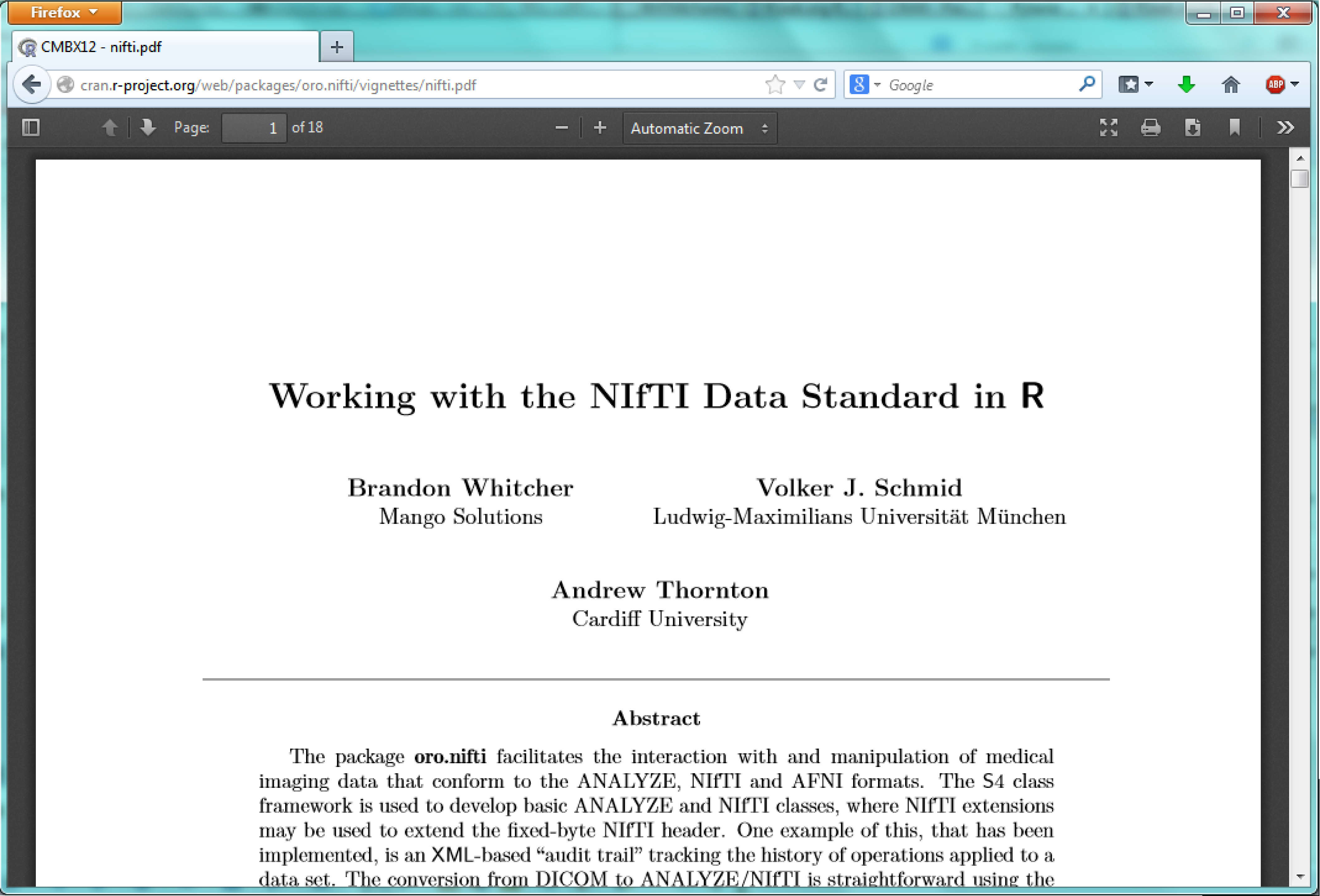


Figure 2: Coronal slice of the abdomen viewed in *neurological* convention (left is right and right is left).

```
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
```

http://cran.r-project.org/web/packages/oro.dicom/index.html

Process NIfTI data with oro.nifti



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

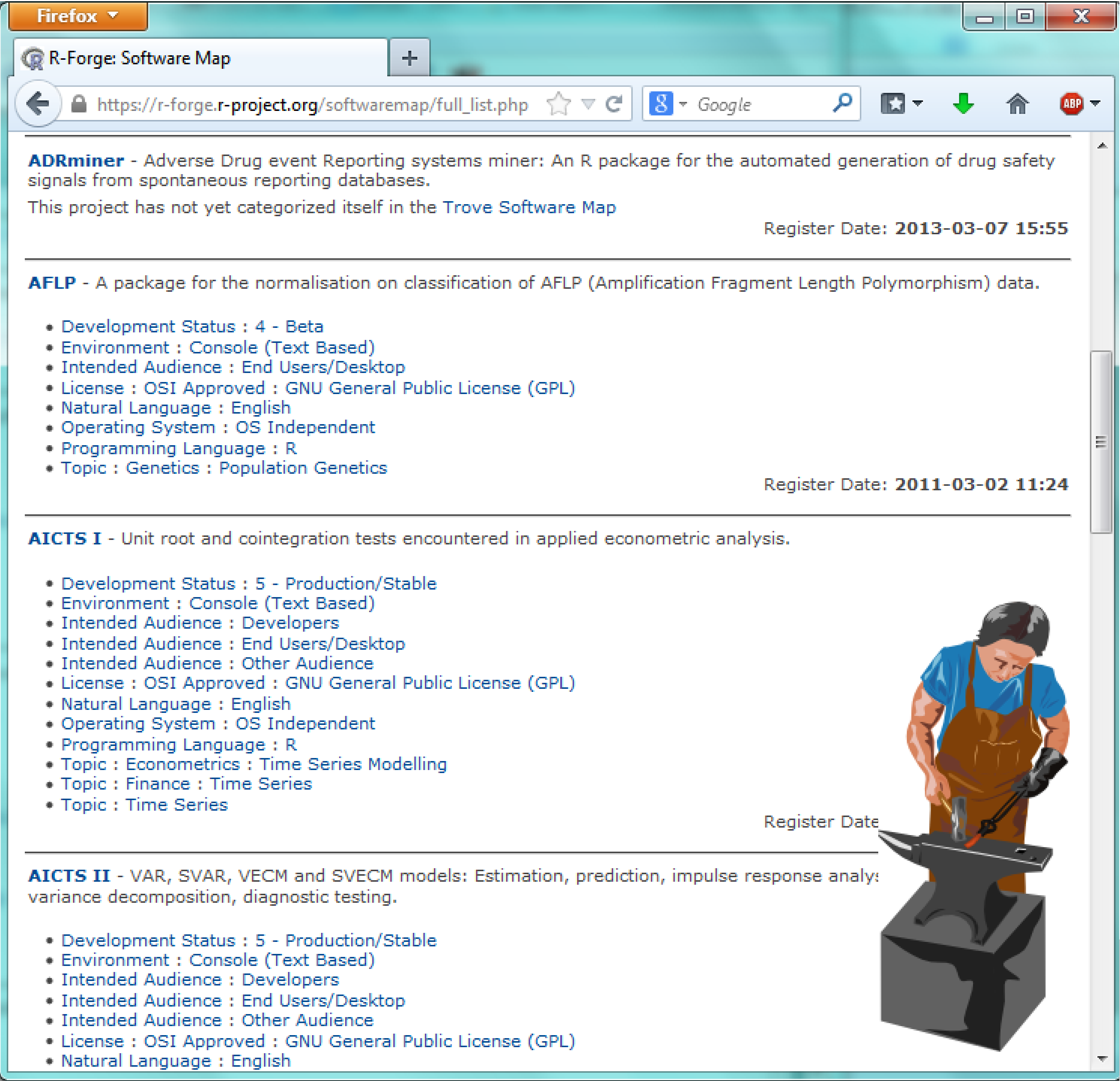
Software Map

Tag cloud | Project Tree | Project List

Project tree

Topic

- Bayesian Statistics *(53 projects)*
- Bioinformatics *(148 projects)*
- Biostatistics & Medical Statistics *(68 projects)*
- Chemoinformatics *(17 projects)*
- Cluster Analysis *(41 projects)*
- Computational Physics *(12 projects)*
- Connectivity *(20 projects)*
- Database *(27 projects)*
- Datasets *(28 projects)*
- Design of Experiments & Analysis of Experimen
- 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)*



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.

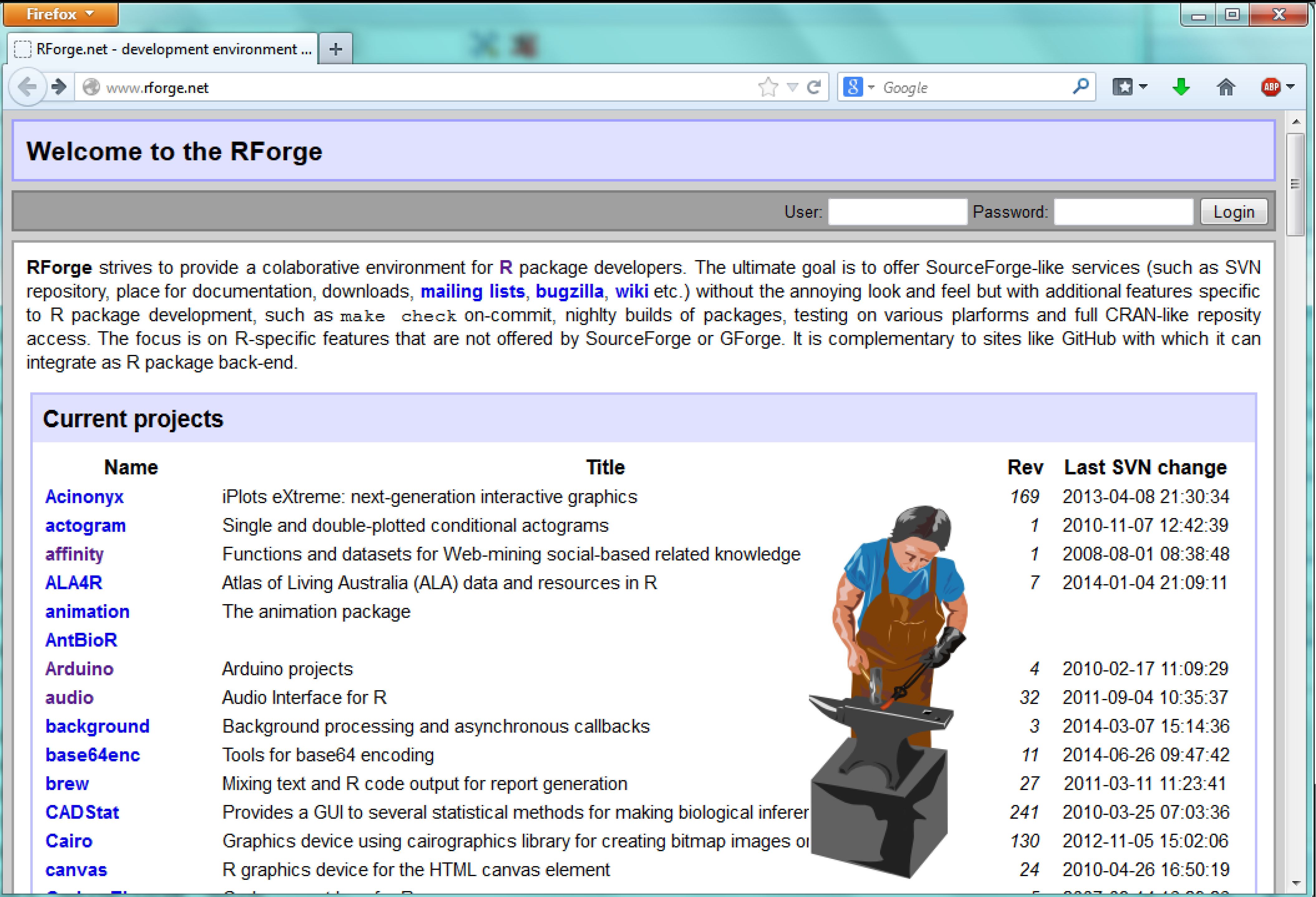
- Development Status : 5 - Production/Stable
- Environment : Console (Text Based)
- Intended Audience : Developers
- Intended Audience : End Users/Desktop
- Intended Audience : Other Audience
- License : OSI Approved : GNU General Public License (GPL)
- Natural Language : English
- Operating System :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



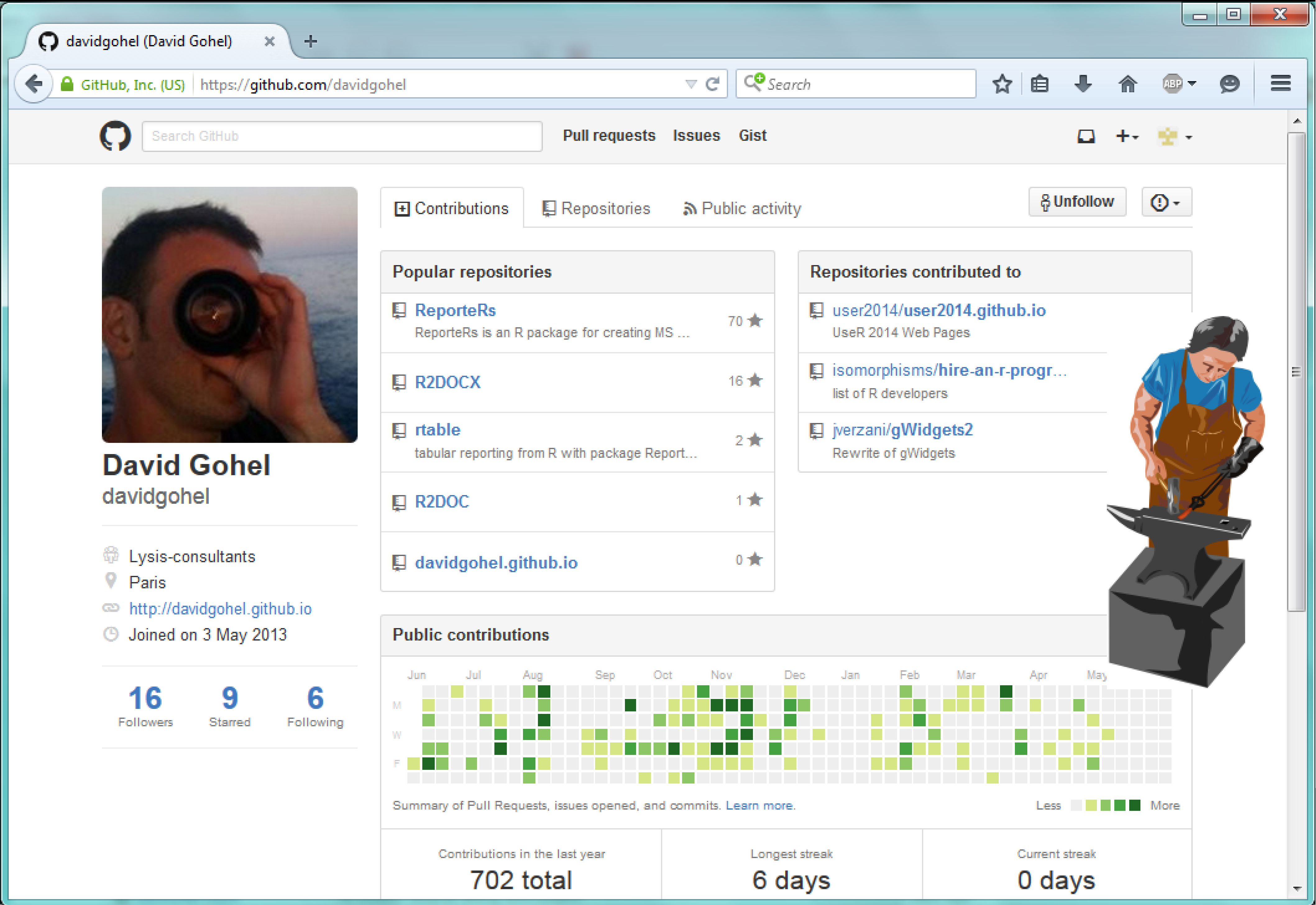
Welcome to the RForge

RForge strives to provide a colaborative 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, nighlty builds of packages, testing on various plarforms and full CRAN-like reposity 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 inferer	*241*	2010-03-25 07:03:36
Cairo	Graphics device using cairographics library for creating bitmap images or	*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



David Gohel
davidgohel

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

16	9	6
Followers	Starred	Following

Popular repositories

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

Repositories contributed to

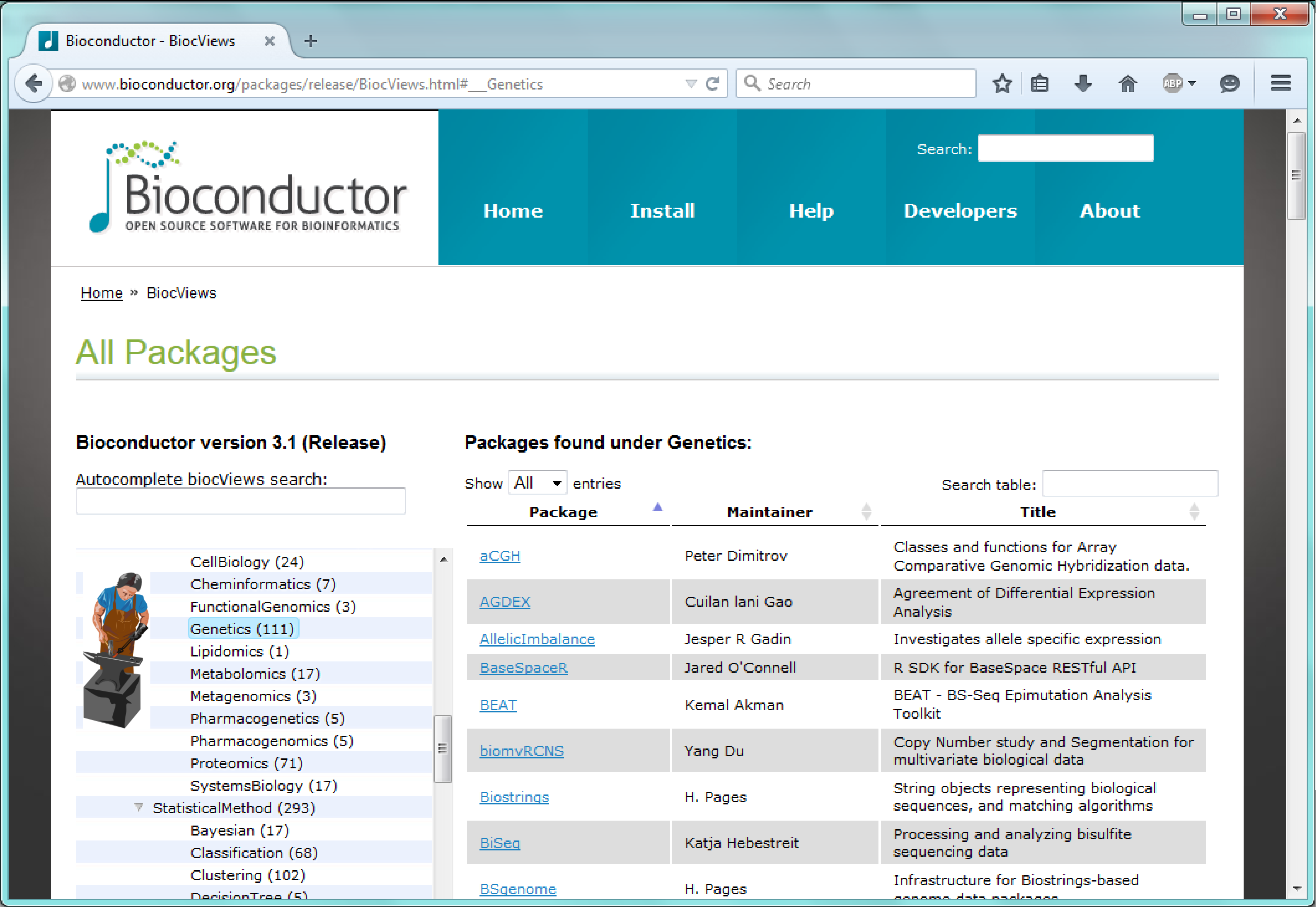
- user2014/user2014.github.io
  UseR 2014 Web Pages
- isomorphisms/hire-an-r-progr…
  list of R developers
- jverzani/gWidgets2
  Rewrite of gWidgets

Public contributions

Summary of Pull Requests, issues opened, and commits. Learn more.

Contributions in the last year	Longest streak	Current streak
702 total	6 days	0 days

The Origin Of Things - Bioconductor



Home » BiocViews

All Packages

Bioconductor version 3.1 (Release)

Autocomplete biocViews search:

- 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)

Packages found under Genetics:

Show All entries

Search table:

Package	Maintainer	Title
aCGH	Peter Dimitrov	Classes and functions for Array Comparative Genomic Hybridization data.
AGDEX	Cuilan lani 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

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 $^{\text{MD5}}$ and install them.

Packages in a binary form can be installed from local .zip archives

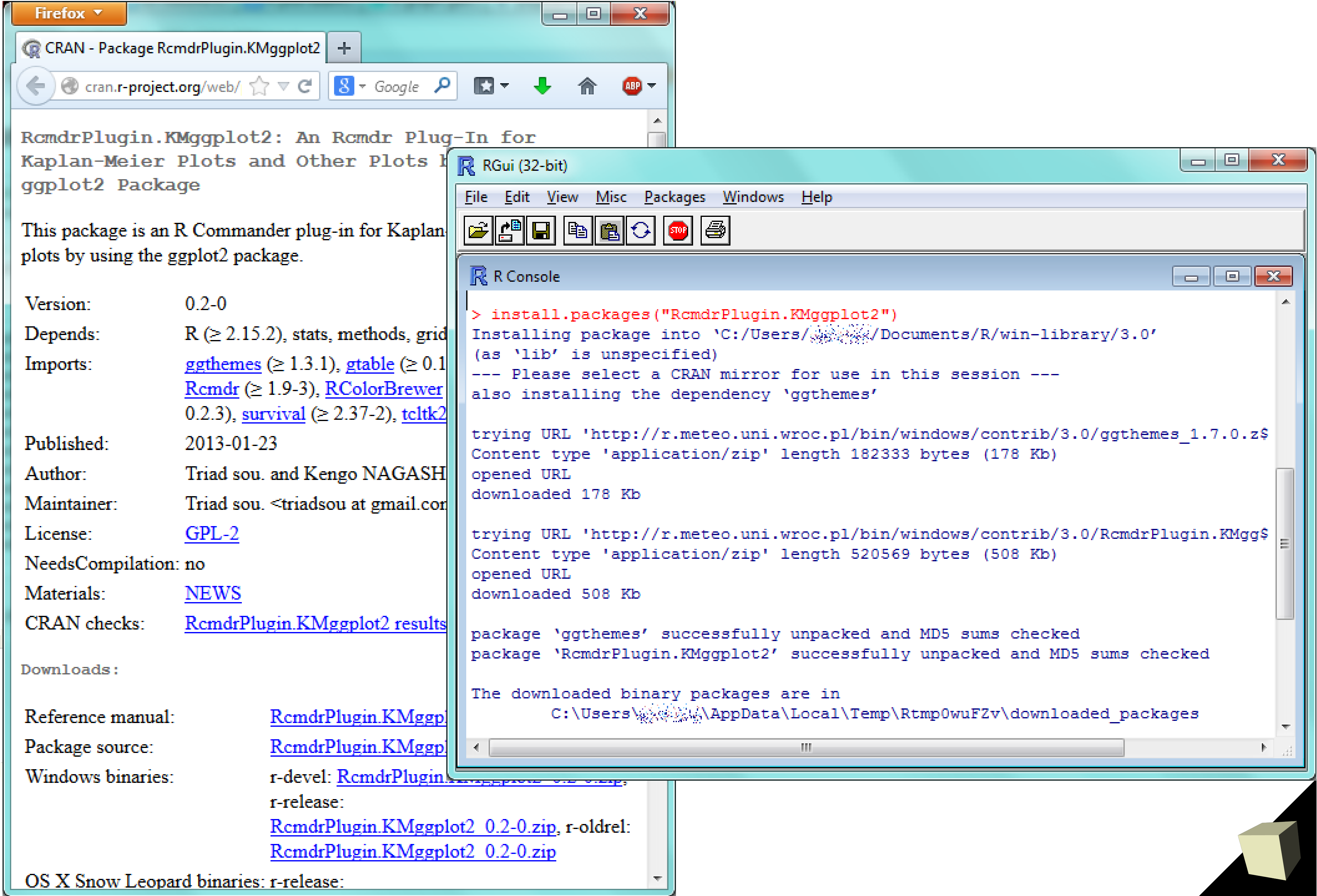
Packages can be also built from sources (common approach on Linux)

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!



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

This package is an R Commander plug-in for Kaplan
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 Rcmdr (≥ 1.9-3), RColorBrewer 0.2.3), survival (≥ 2.37-2), tcltk2
Published:	2013-01-23
Author:	Triad sou. and Kengo NAGASH
Maintainer:	Triad sou. <triadsou at gmail.cor
License:	GPL-2
NeedsCompilation:	no
Materials:	NEWS
CRAN checks:	RcmdrPlugin.KMggplot2 results

Downloads:

Reference manual:	RcmdrPlugin.KMggp
Package source:	RcmdrPlugin.KMggp
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:

```
> install.packages("RcmdrPlugin.KMggplot2")
Installing package into 'C:/Users/[redacted]/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.z$
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.KMgg$
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\[redacted]\AppData\Local\Temp\Rtmp0wuFZv\downloaded_packages
```

CRAN D.I.Y.

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:

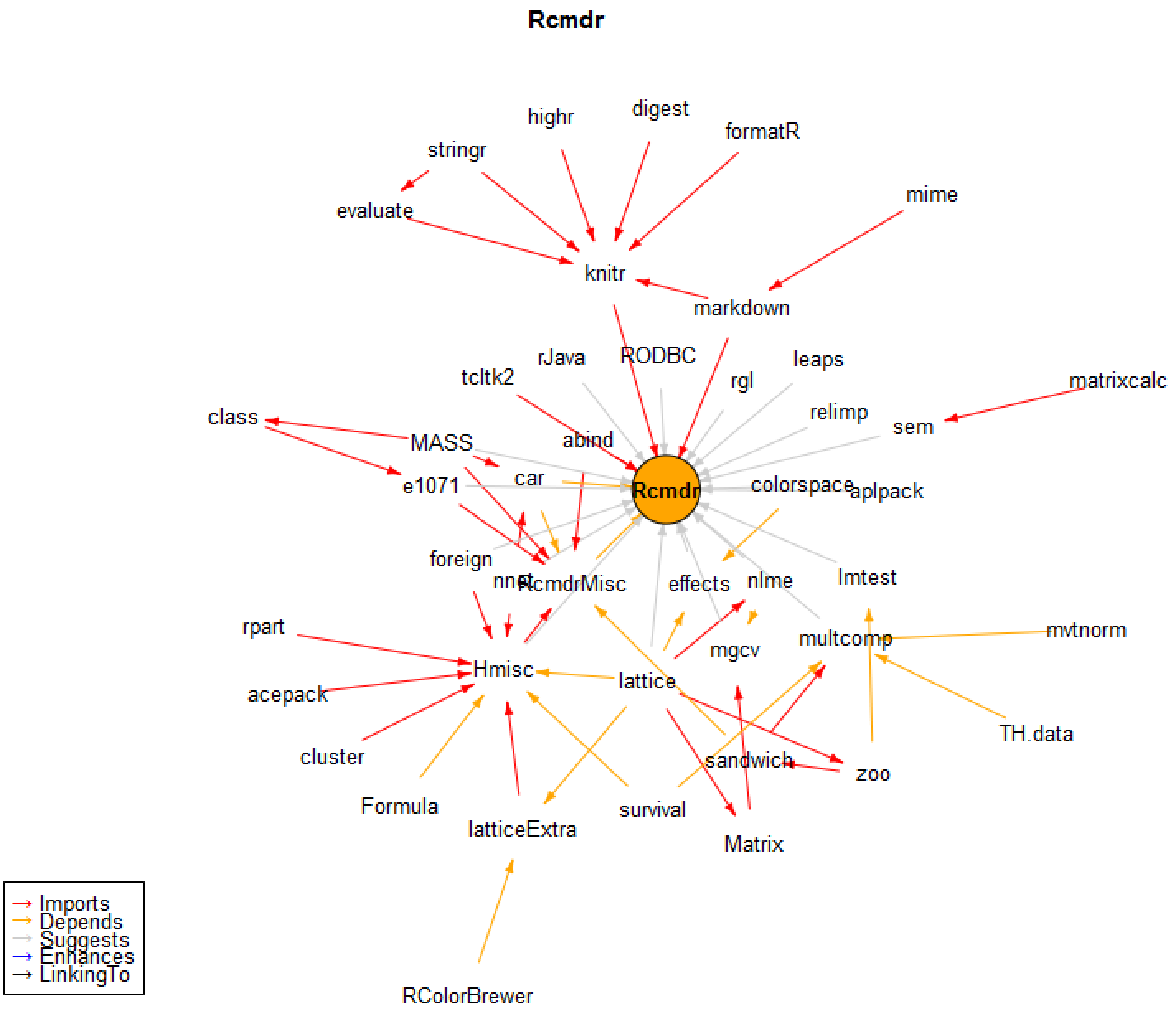
> *You may wish to mirror only a subset of CRAN, for security, legal compliance or any other in-house reason*
>
> *You may wish to restrict internal package use to a subset of public packages, to minimize package duplication, or other reasons of coding standards*
>
> *You may wish to make packages available 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/

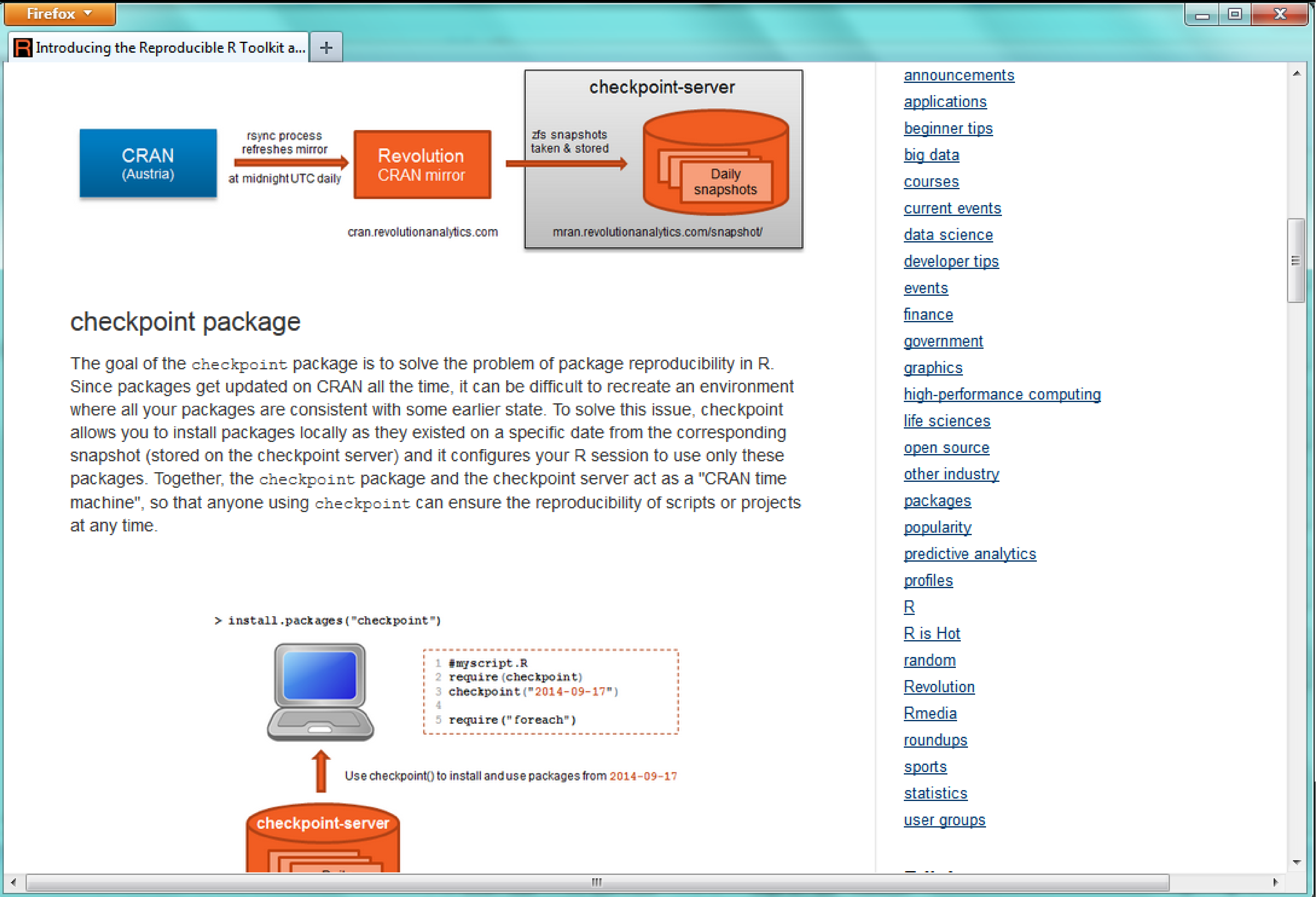
Do not forget to visit the quick introduction to miniCRAN.

CRAN D.I.Y.

MiniCRAN also helps you to track dependencies between packages.



Versioned CRAN – meet Revolution::*checkpoint*



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")
```

```
#myscript.R
require(checkpoint)
checkpoint("2014-09-17")

require("foreach")
```

- 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

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

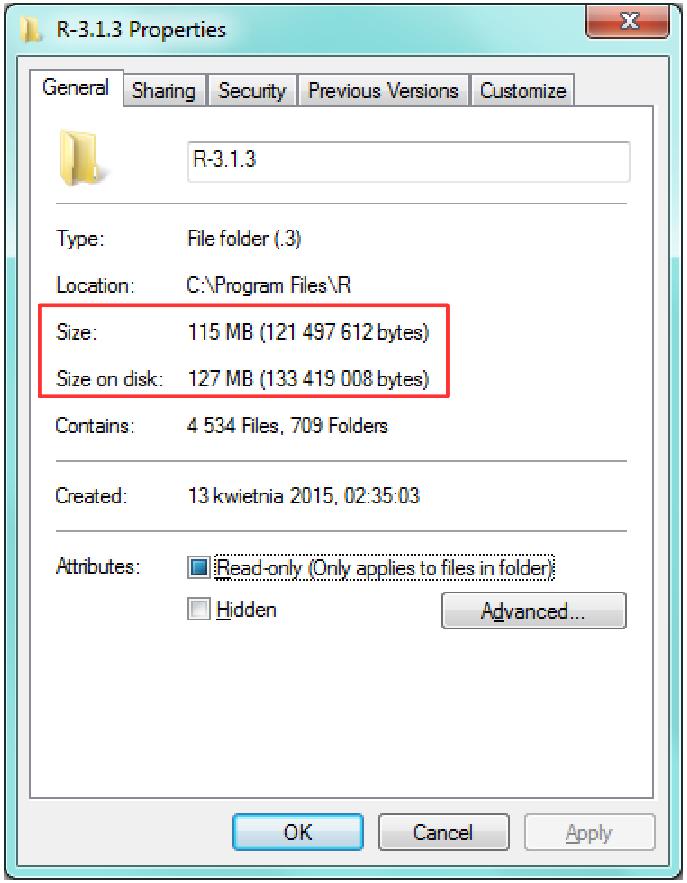
X There are many options to optimize the code

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

R-3.1.3

Type: File folder (.3)

Location: C:\Program Files\R

Size: 115 MB (121 497 612 bytes)

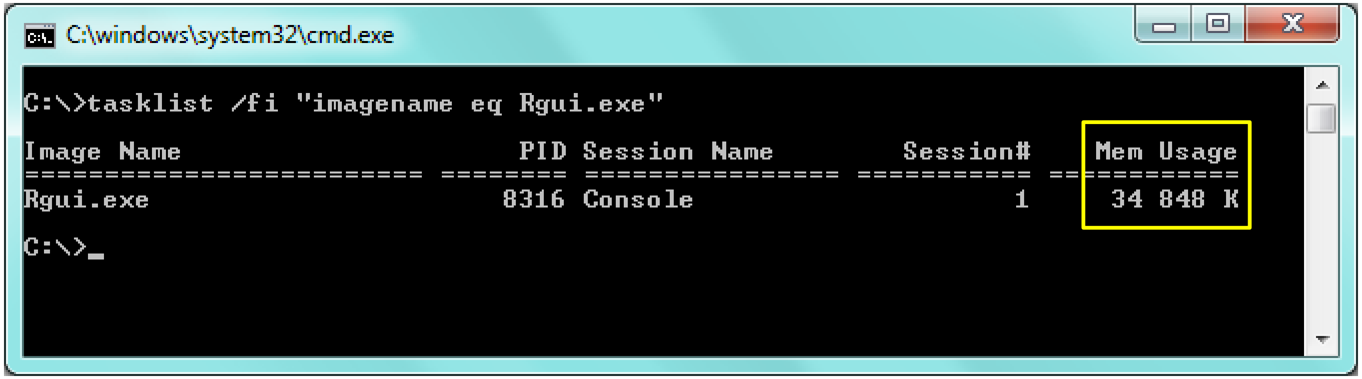
Size on disk: 127 MB (133 419 008 bytes)

Contains: 4 534 Files, 709 Folders

Created: 13 kwietnia 2015, 02:35:03

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

☐ Hidden



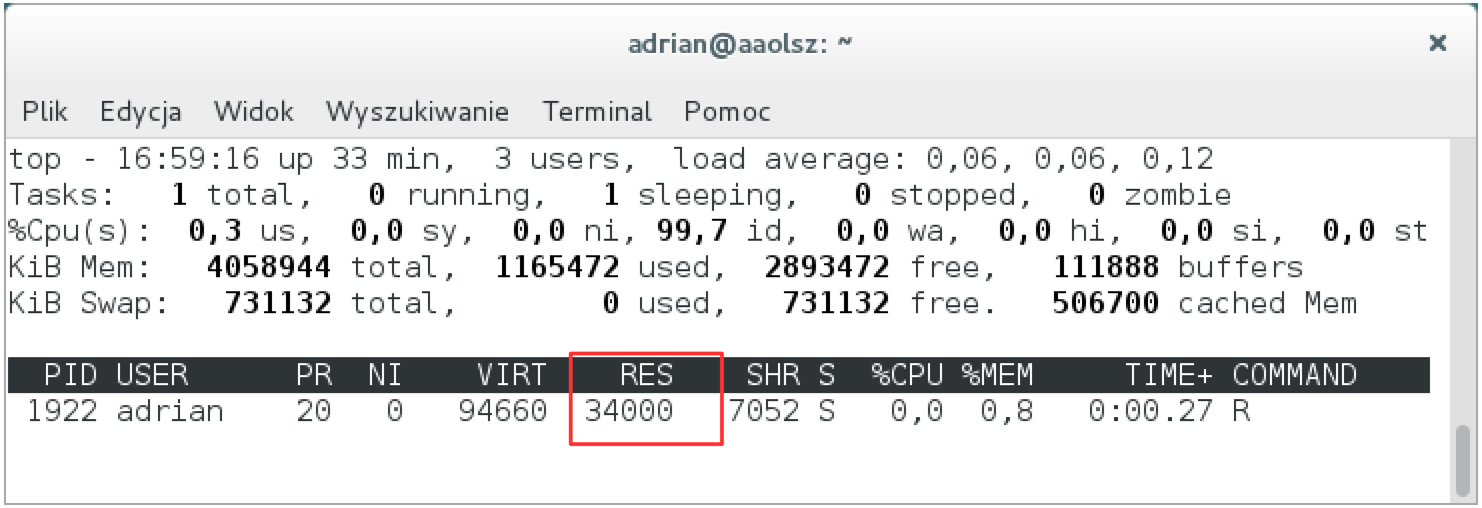
C:\windows\system32\cmd.exe

```
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

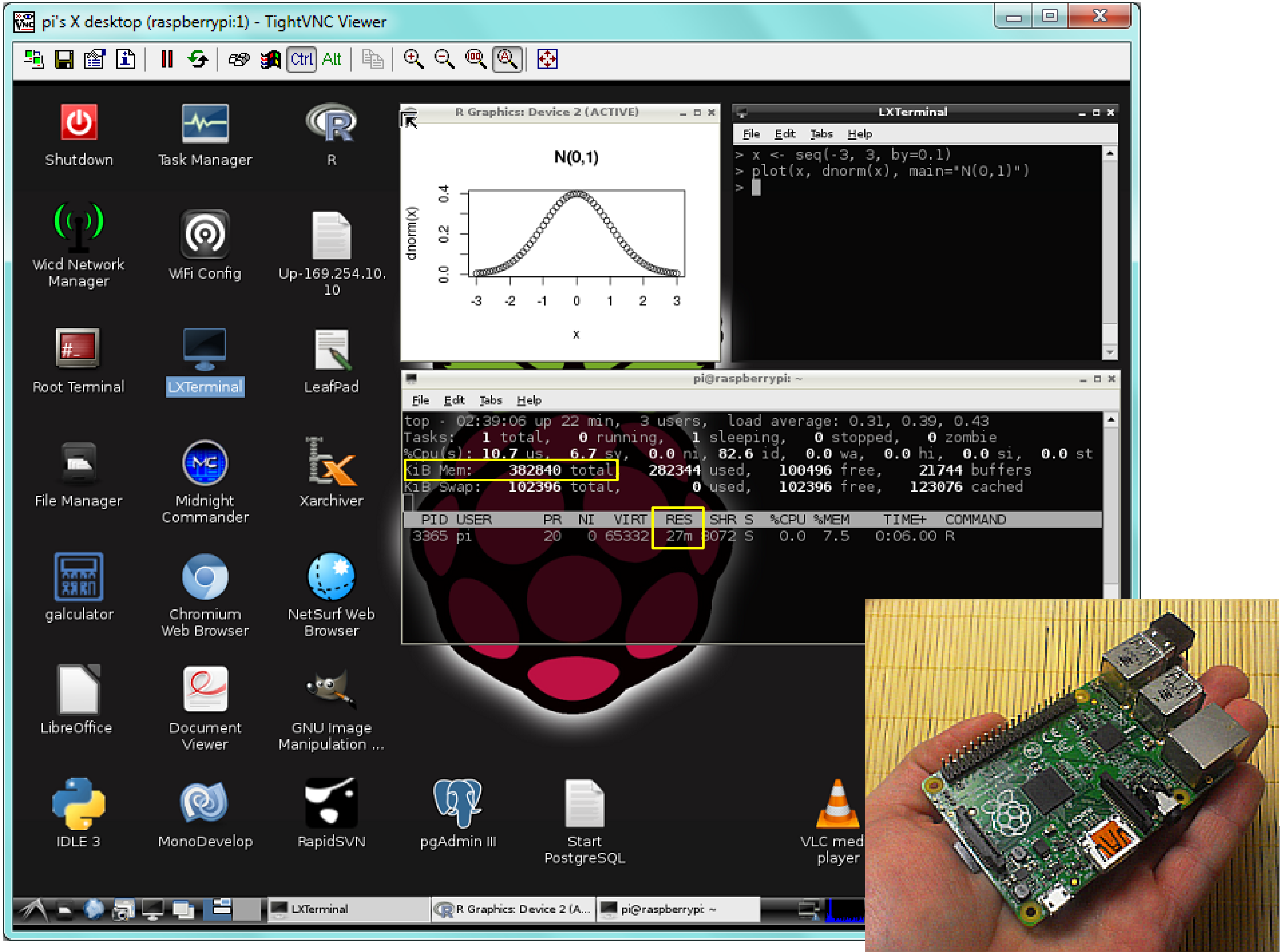
RAM: ~ 35MiB



adrian@aaolsz: ~

```
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!



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

```
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 3072 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

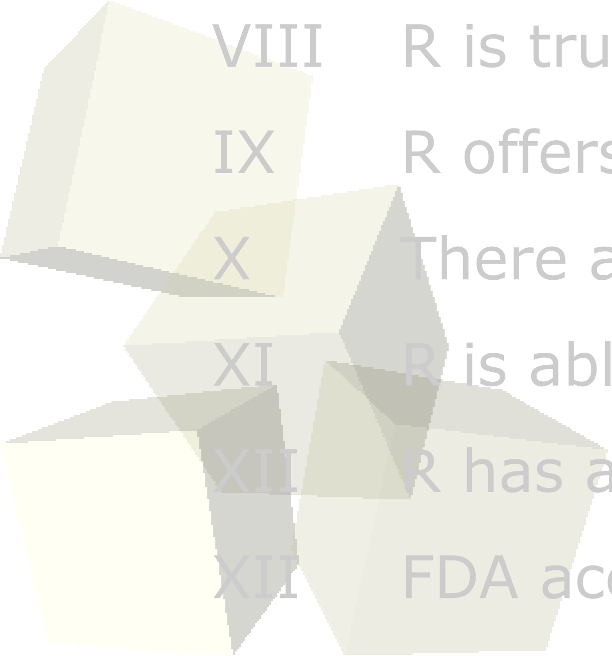
IX R offers numerous ways of presenting data

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

XII FDA accepted using R for drug trials!



…supported by the **science**

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 Pricenton, 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

ISSN: 2073-4859

About The R Journal

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

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.

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.

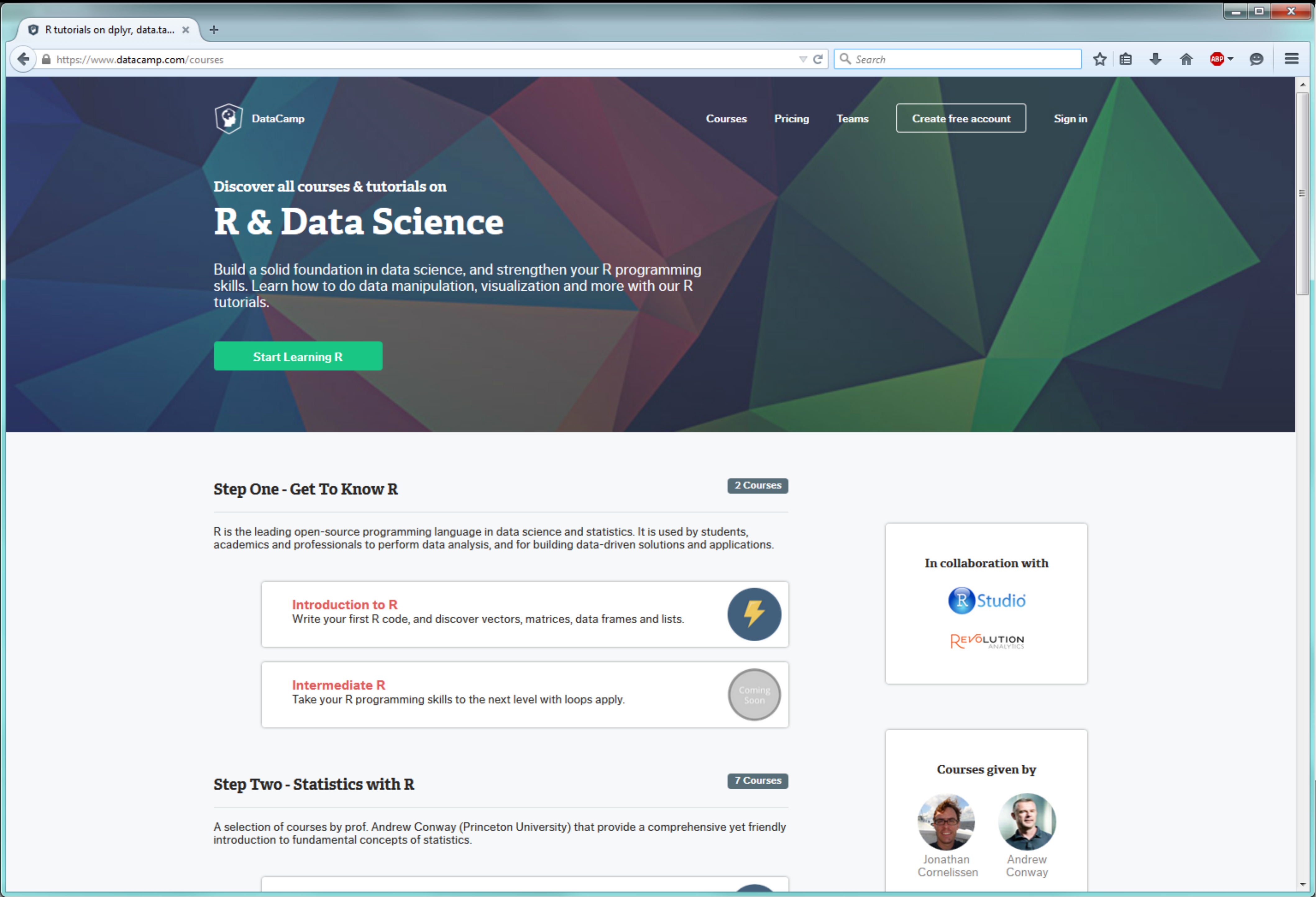
where to leaRn?

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 couRse!



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.

Step One - Get To Know R

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.

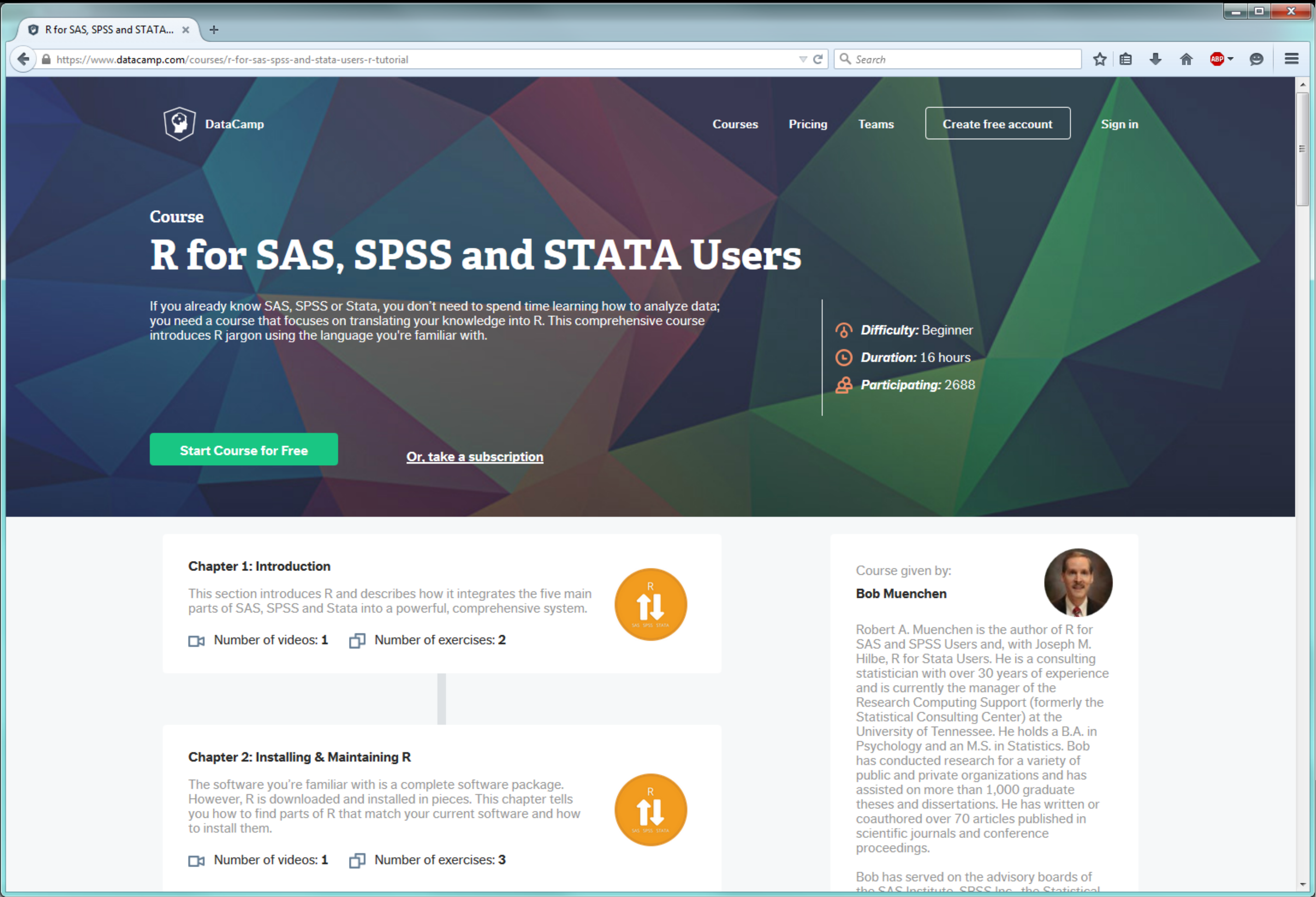
Step Two - Statistics with R

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
RStudio
Revolution Analytics

Courses given by
Jonathan Cornelissen
Andrew Conway

Take a couRse!



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

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**

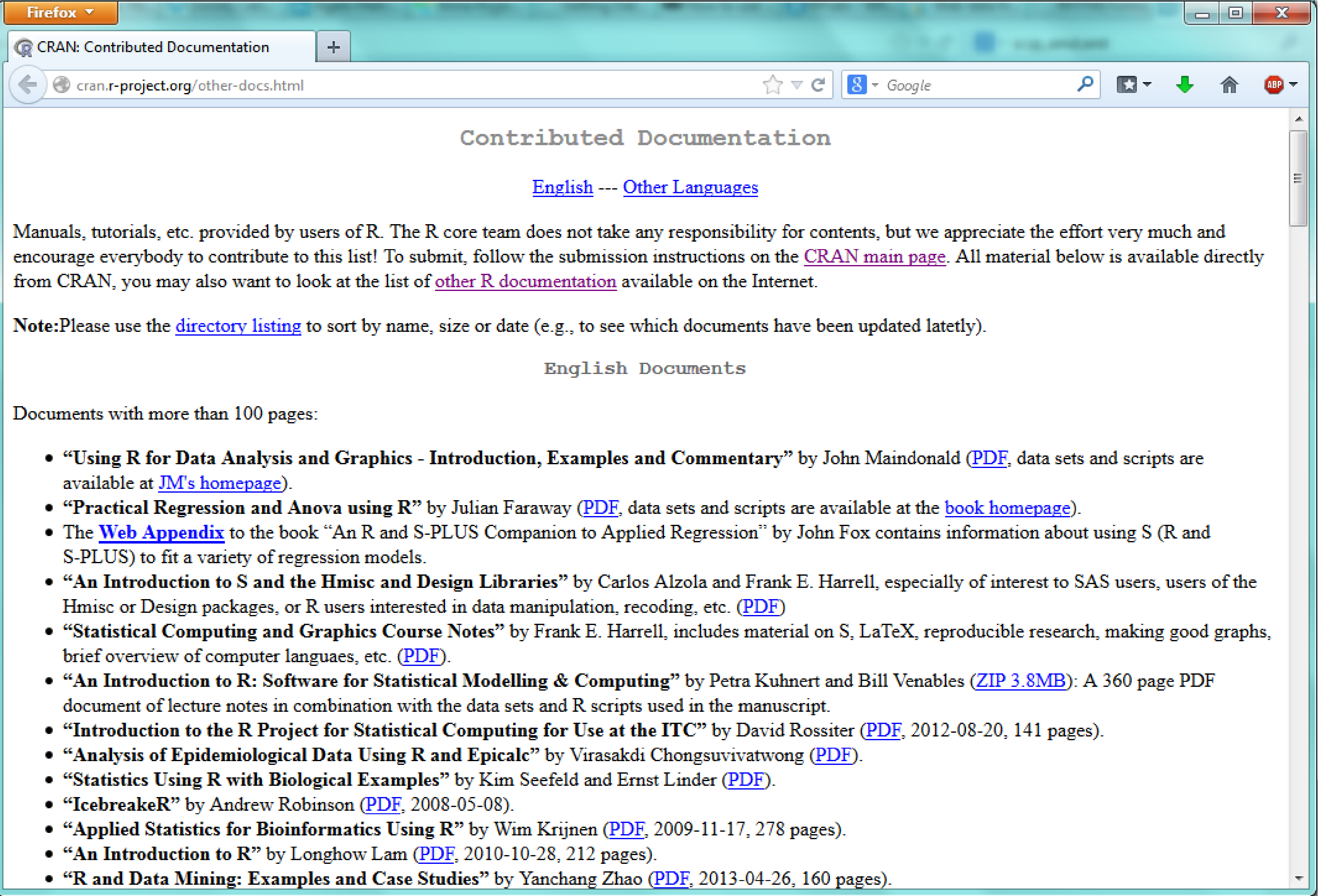
Course given by:

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

Look! This library contains pRiceless resources!



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.

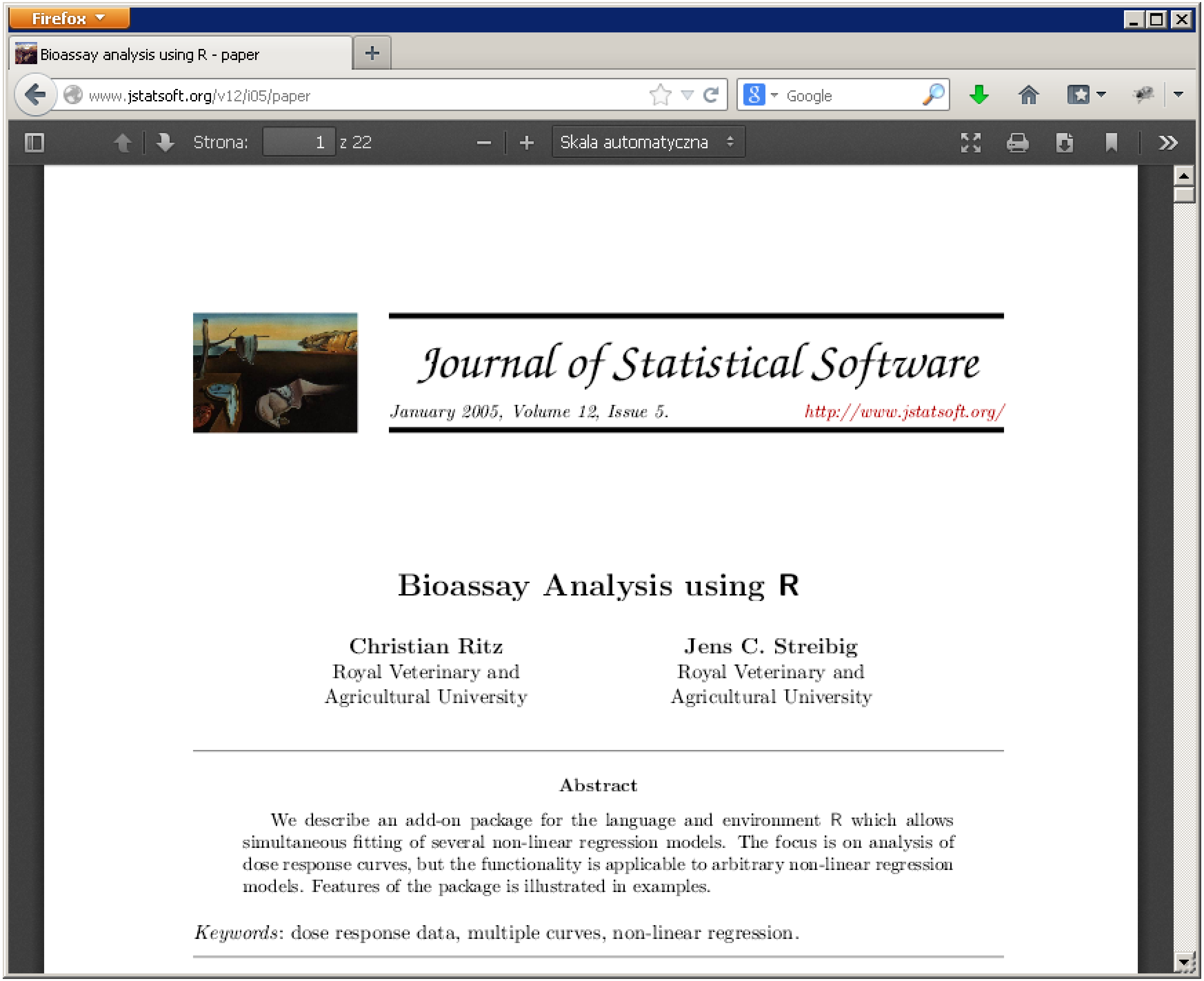
Note: Please use the directory listing to sort by name, size or date (e.g., to see which documents have been updated latetly).

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 languaes, 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 papeRs at our fingertips



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

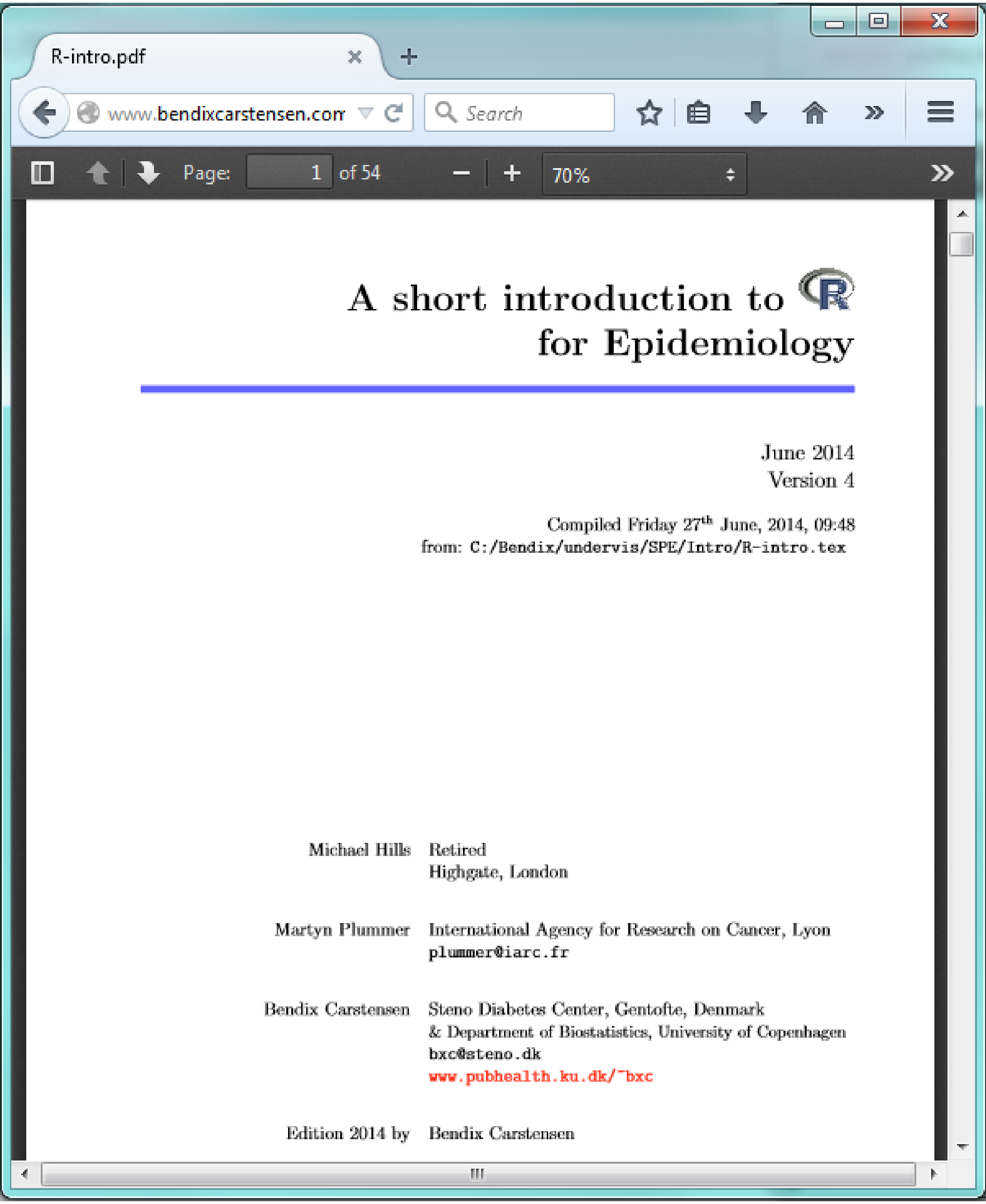
Jens C. Streibig
Royal Veterinary and
Agricultural University

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 papeRs at our fingertips



A short introduction to R for Epidemiology

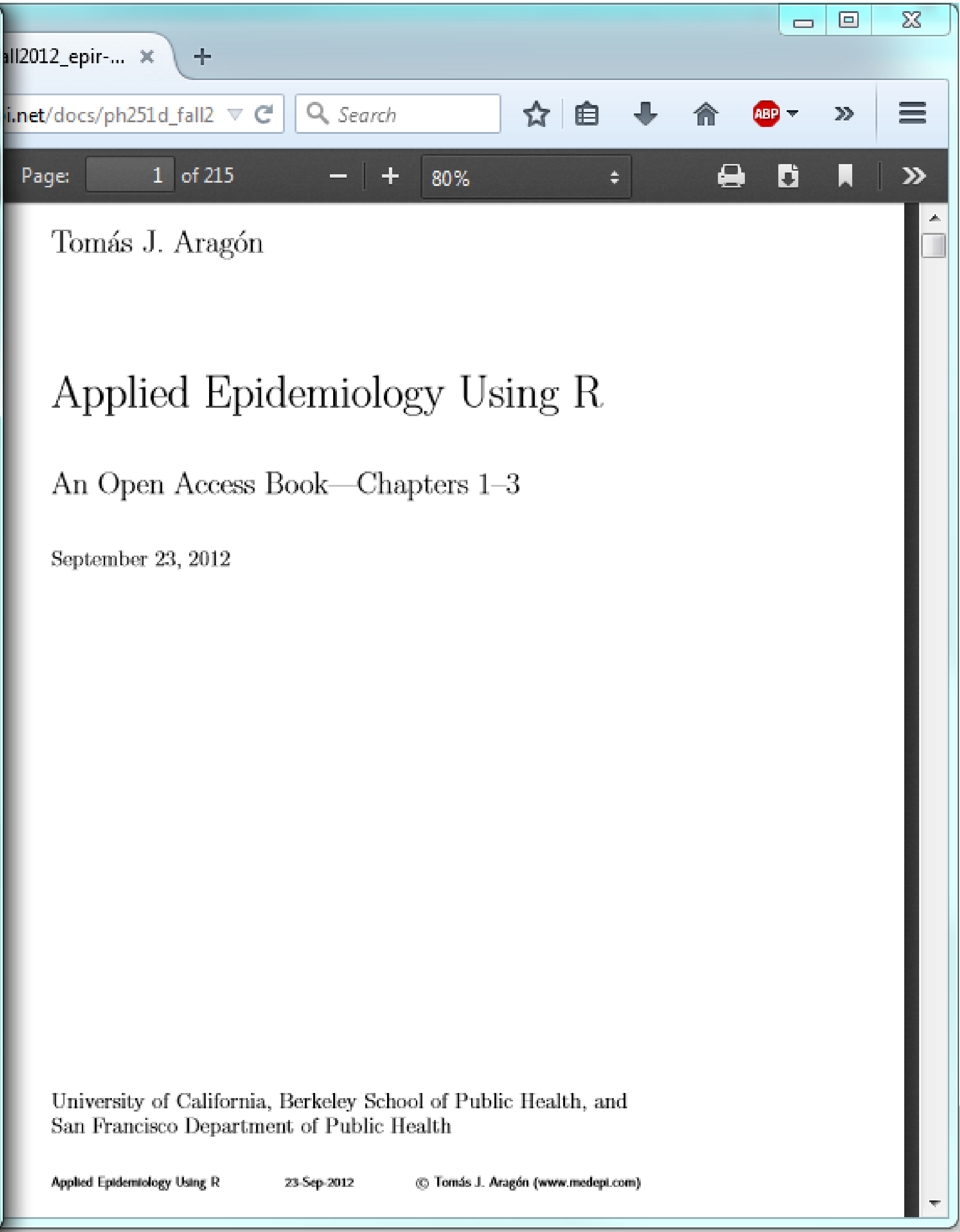
June 2014
Version 4

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



Tomás J. Aragón

Applied Epidemiology Using R

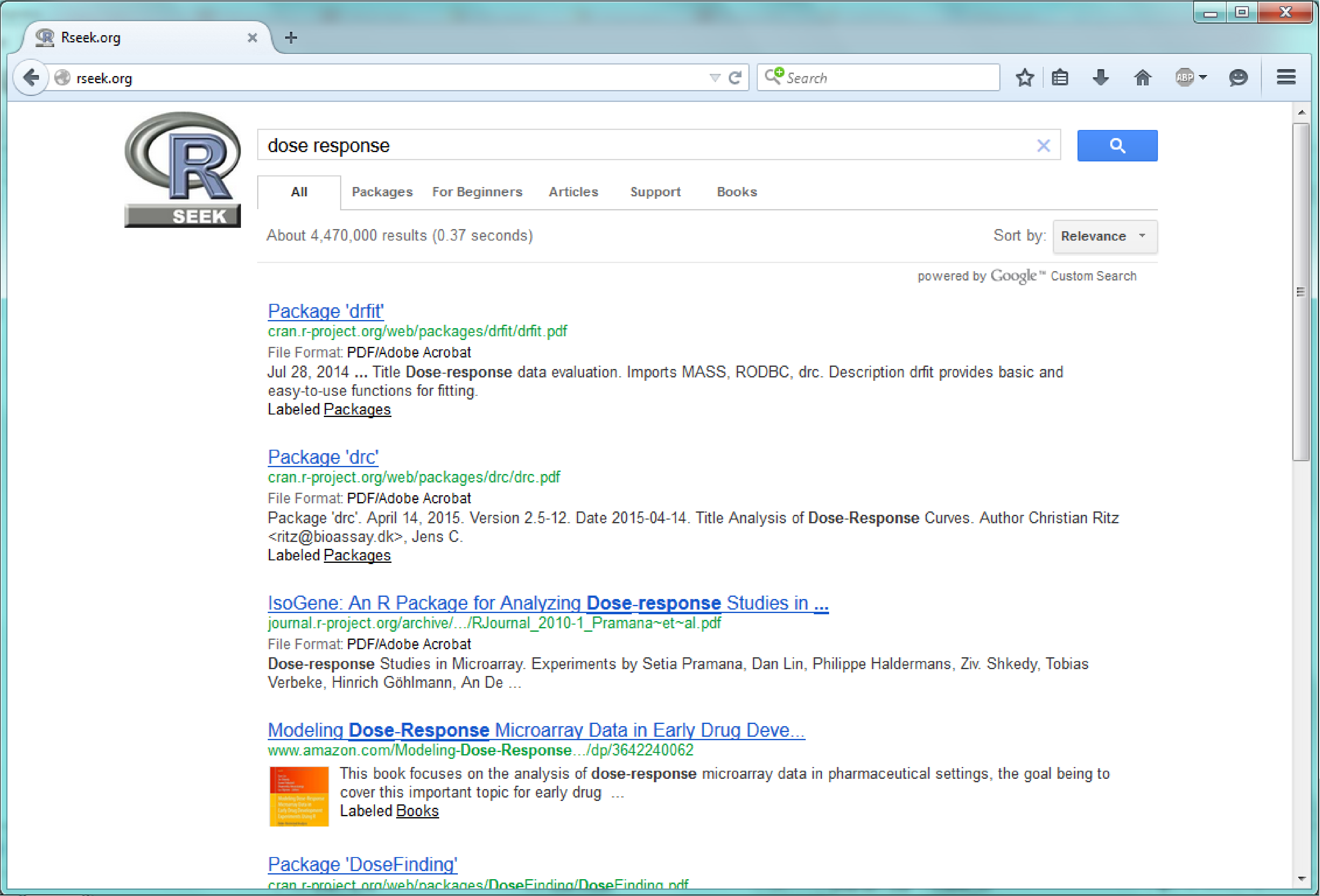
An Open Access Book—Chapters 1–3

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



About 4,470,000 results (0.37 seconds)

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

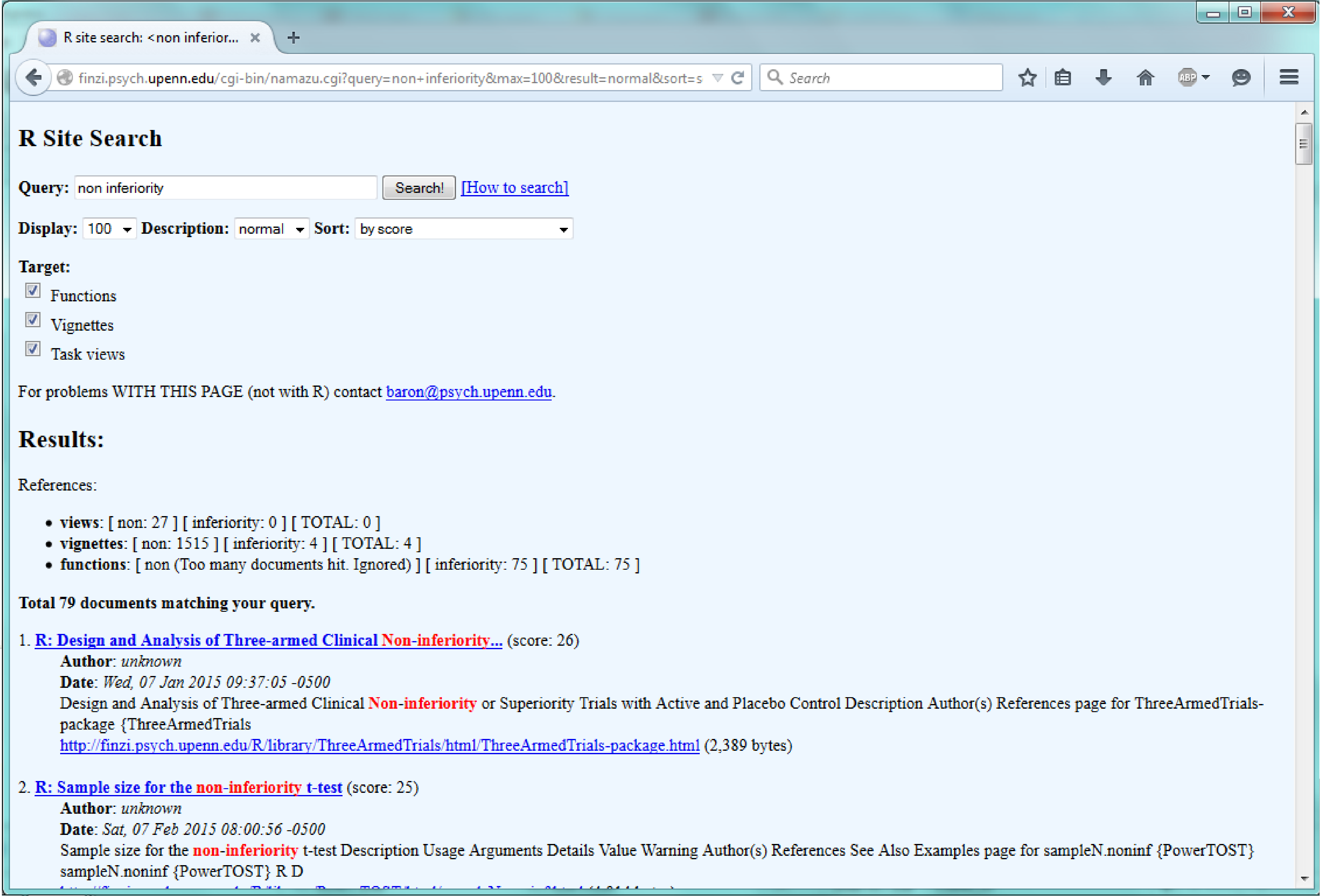
Package 'drc'
cran.r-project.org/web/packages/drc/drc.pdf
File Format: PDF/Adobe Acrobat
Package 'drc'. April 14, 2015. Version 2.5-12. Date 2015-04-14. Title Analysis of **Dose-Response** Curves. Author Christian Ritz <ritz@bioassay.dk>, Jens C.
Labeled Packages

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

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:

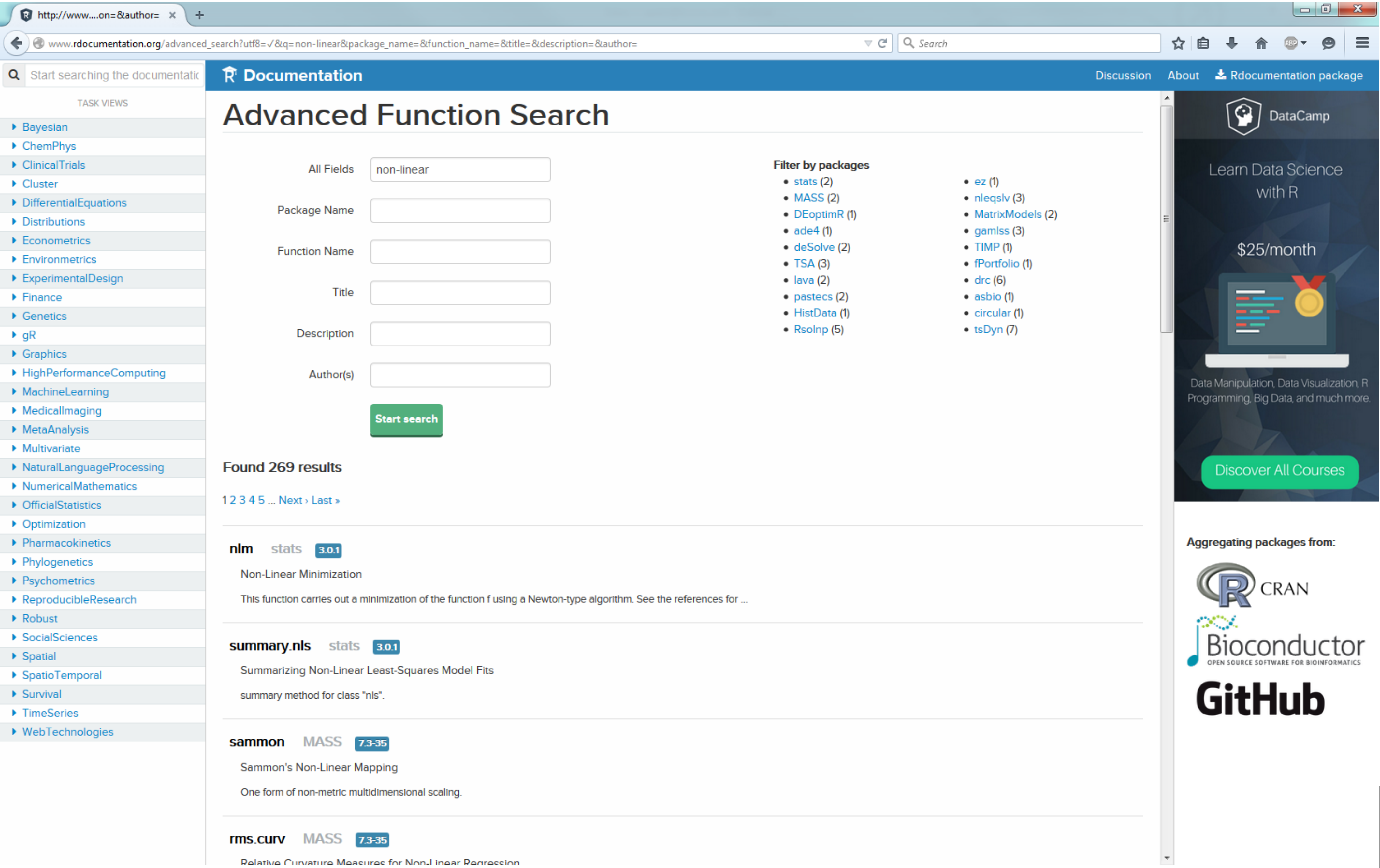
- **views**: [non: 27] [inferiority: 0] [TOTAL: 0]
- **vignettes**: [non: 1515] [inferiority: 4] [TOTAL: 4]
- **functions**: [non (Too many documents hit. Ignored)] [inferiority: 75] [TOTAL: 75]

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



Advanced Function Search

All Fields: non-linear
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)
- fPortfolio (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

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.

...supported by the community

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.

http://www.r-project.org/doc/R-FDA.pdf

...supported by the community

PAZUR

Poznański Akademicki Zlot Użytkowników R

Organizatorzy:

Uniwersytet Przyrodniczy w Poznaniu

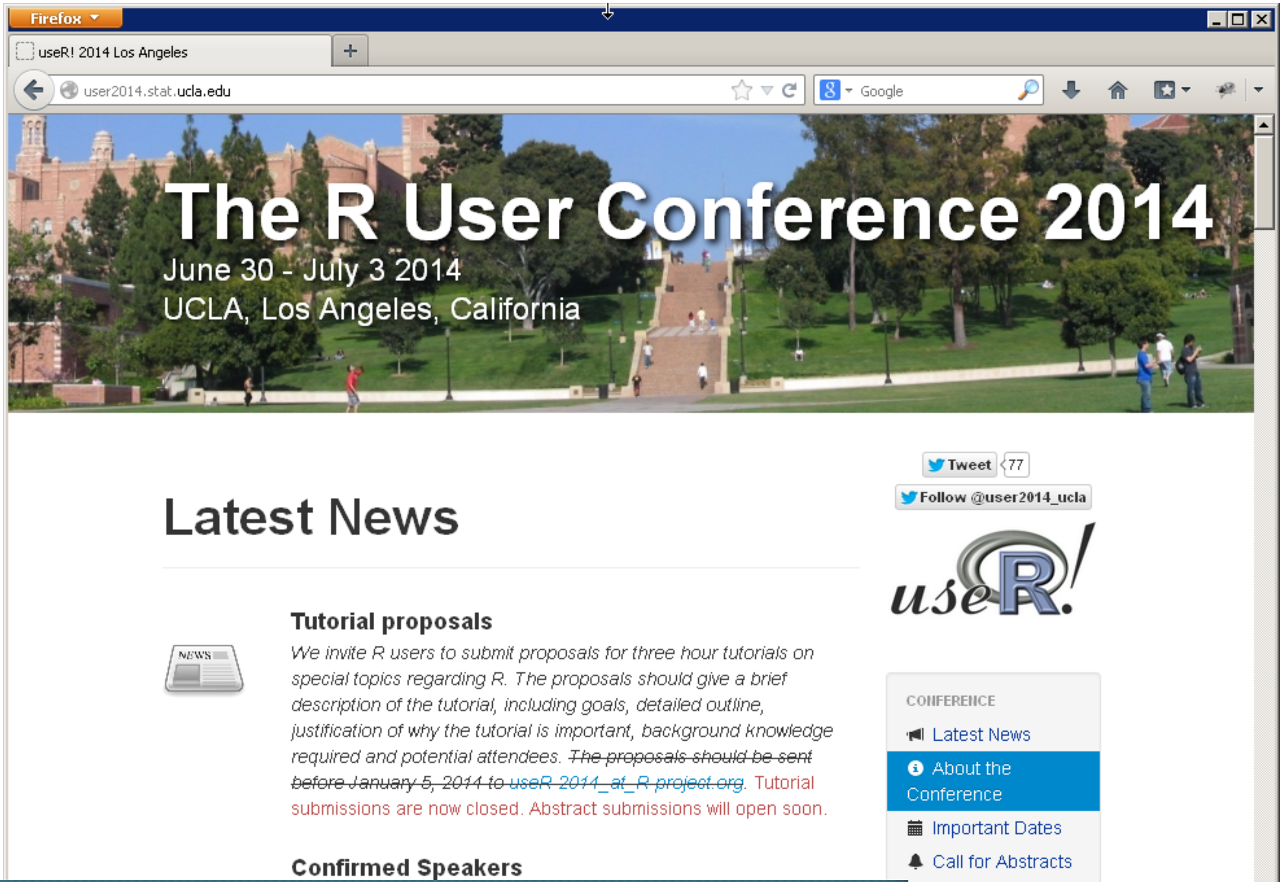
ESTYMATOR
STUDENCKIE KOŁO NAUKOWE
Uniwersytet Ekonomiczny
w Poznaniu

Sponsor:

REVOLUTION ANALYTICS

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

ZOO COWORKING SPACE POZNAN



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 useR-2014_at_R-project.org.~~ Tutorial submissions are now closed. Abstract submissions will open soon.

Confirmed Speakers

EARL 2015

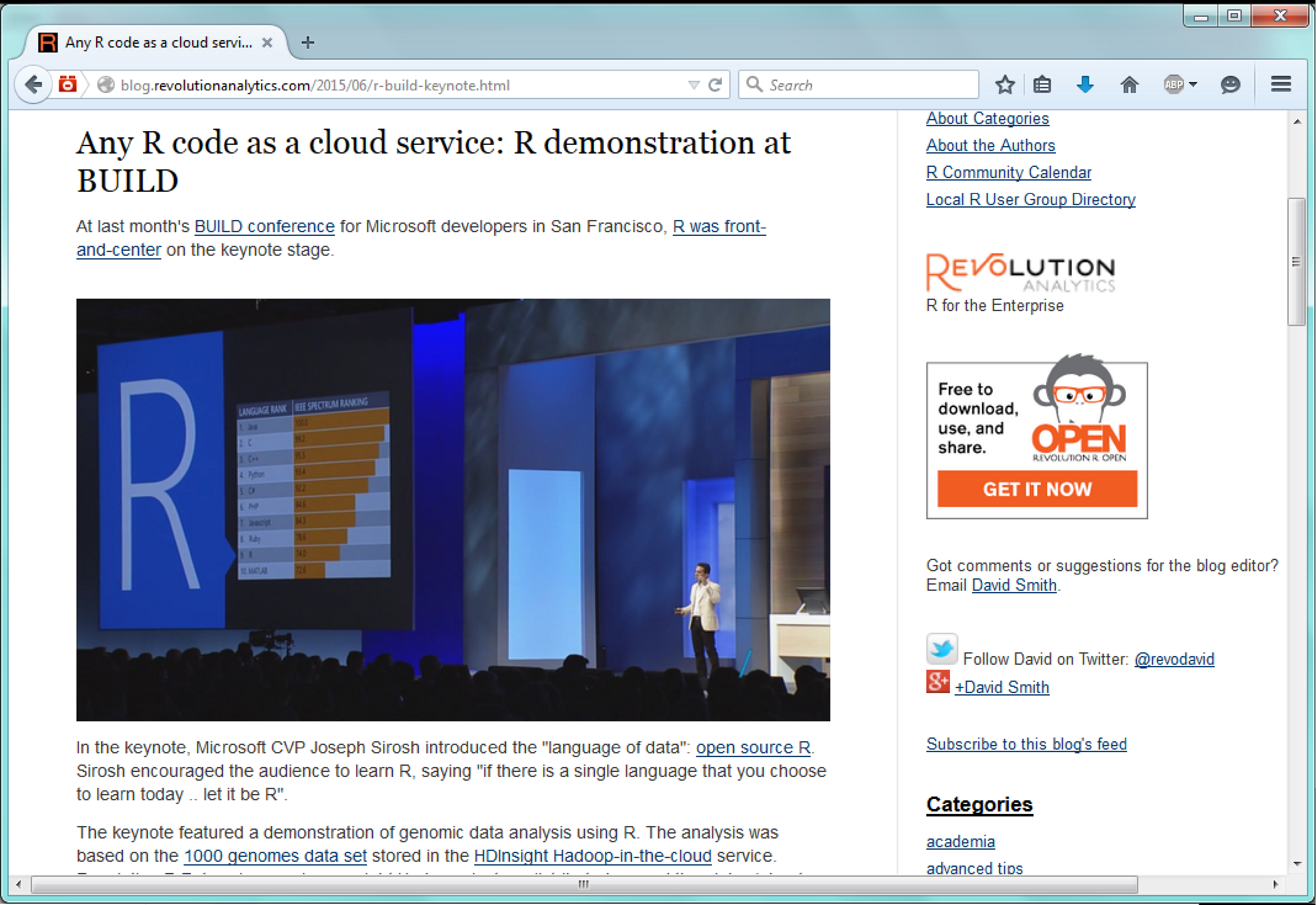
EFFECTIVE APPLICATIONS OF THE R LANGUAGE

LONDON 14 - 16 SEPTEMBER

@earlconf

R at Microsoft's BUILD 2015 conference

say “wow!”



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

REVOLUTION ANALYTICS
R for the Enterprise

Free to download, use, and share.
OPEN
REVOLUTION R OPEN
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-bloggers

R news and tutorials contributed by (552) R bloggers

WELCOME!

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

And get updates to your Facebook:

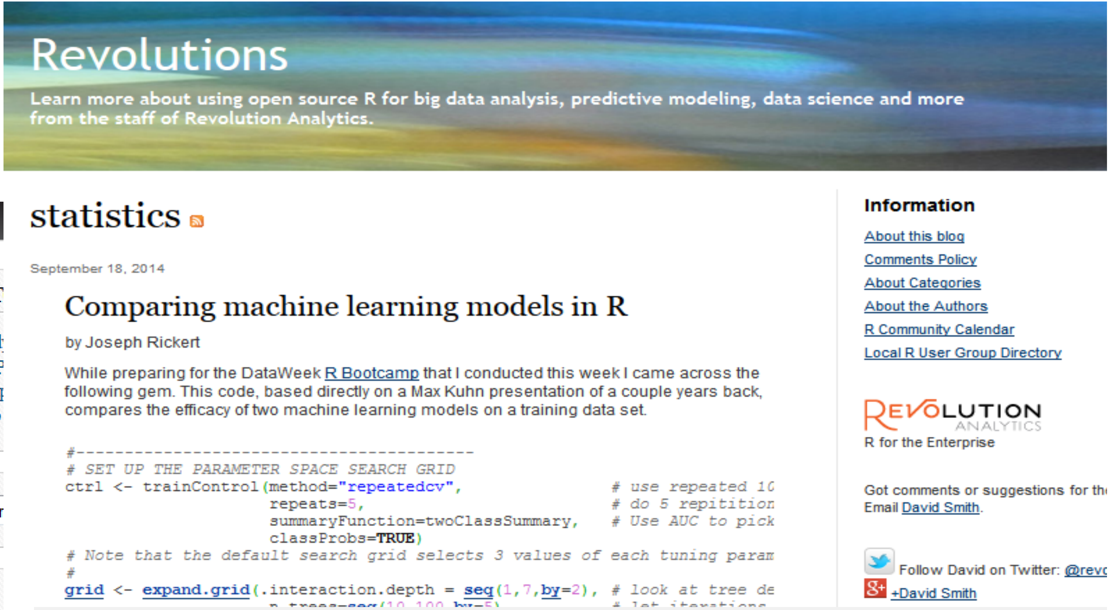
Fun with the Raspberry Pi

January 21, 2014

By Markus Gesmann

(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.



Revolutions

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

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.

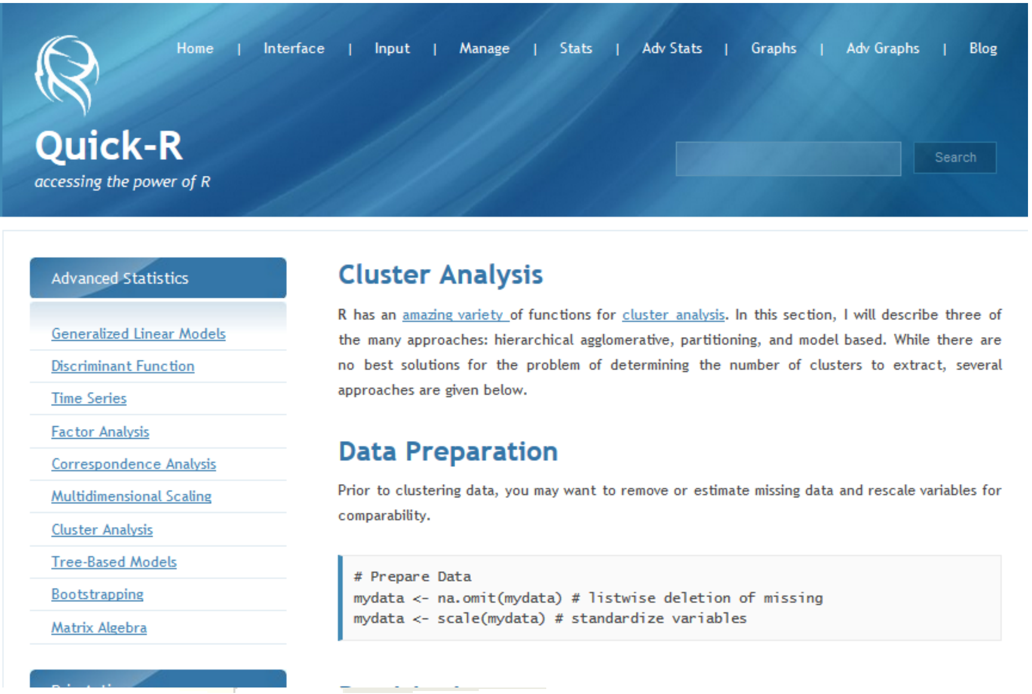
```
#-----------------------------------------
# SET UP THE PARAMETER SPACE SEARCH GRID
ctrl <- trainControl(method="repeatedcv",           # use repeated 10
                     repeats=5,                      # do 5 repitition
                     summaryFunction=twoClassSummary,   # Use AUC to pick
                     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
```

Information

- About this blog
- Comments Policy
- About Categories
- About the Authors
- R Community Calendar
- Local R User Group Directory

Got comments or suggestions for the
Email David Smith.

Follow David on Twitter: @revo
+David Smith



Advanced Statistics

- Generalized Linear Models
- Discriminant Function
- Time Series
- Factor Analysis
- Correspondence Analysis
- Multidimensional Scaling
- Cluster Analysis
- Tree-Based Models
- Bootstrapping
- Matrix Algebra

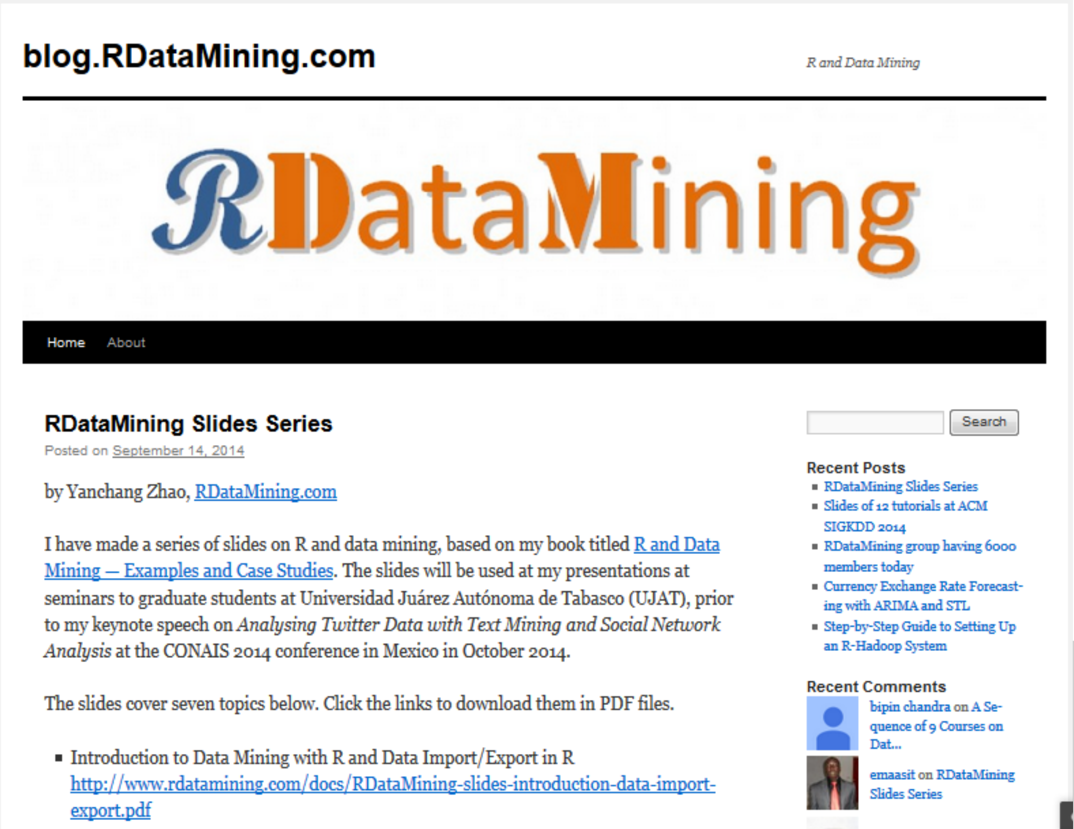
Cluster Analysis

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
```



blog.RDataMining.com

R and Data Mining

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

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

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

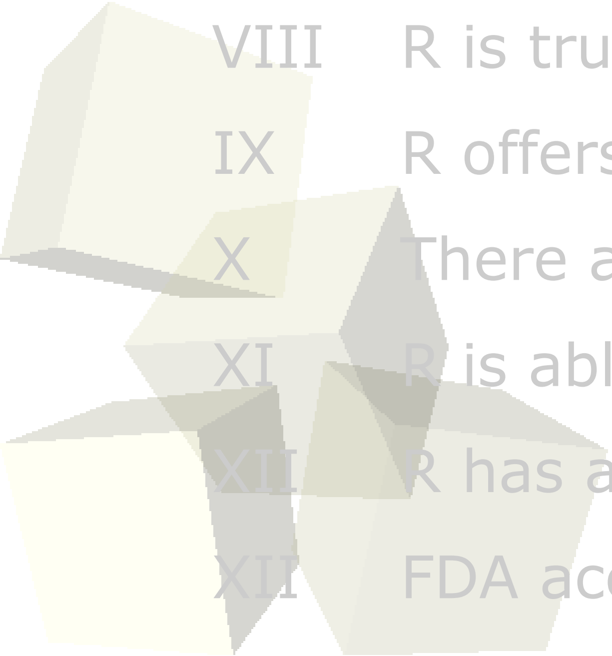
IX R offers numerous ways of presenting data

X There are many options to optimize the code

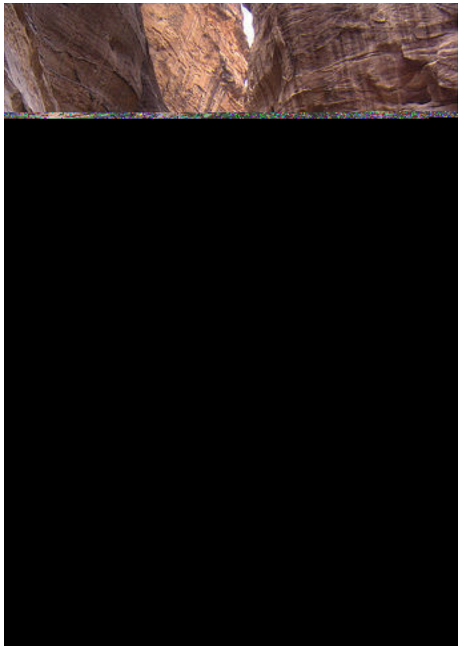
XI R is able to handle large amount of data

XII R has a set of fancy tools and IDEs

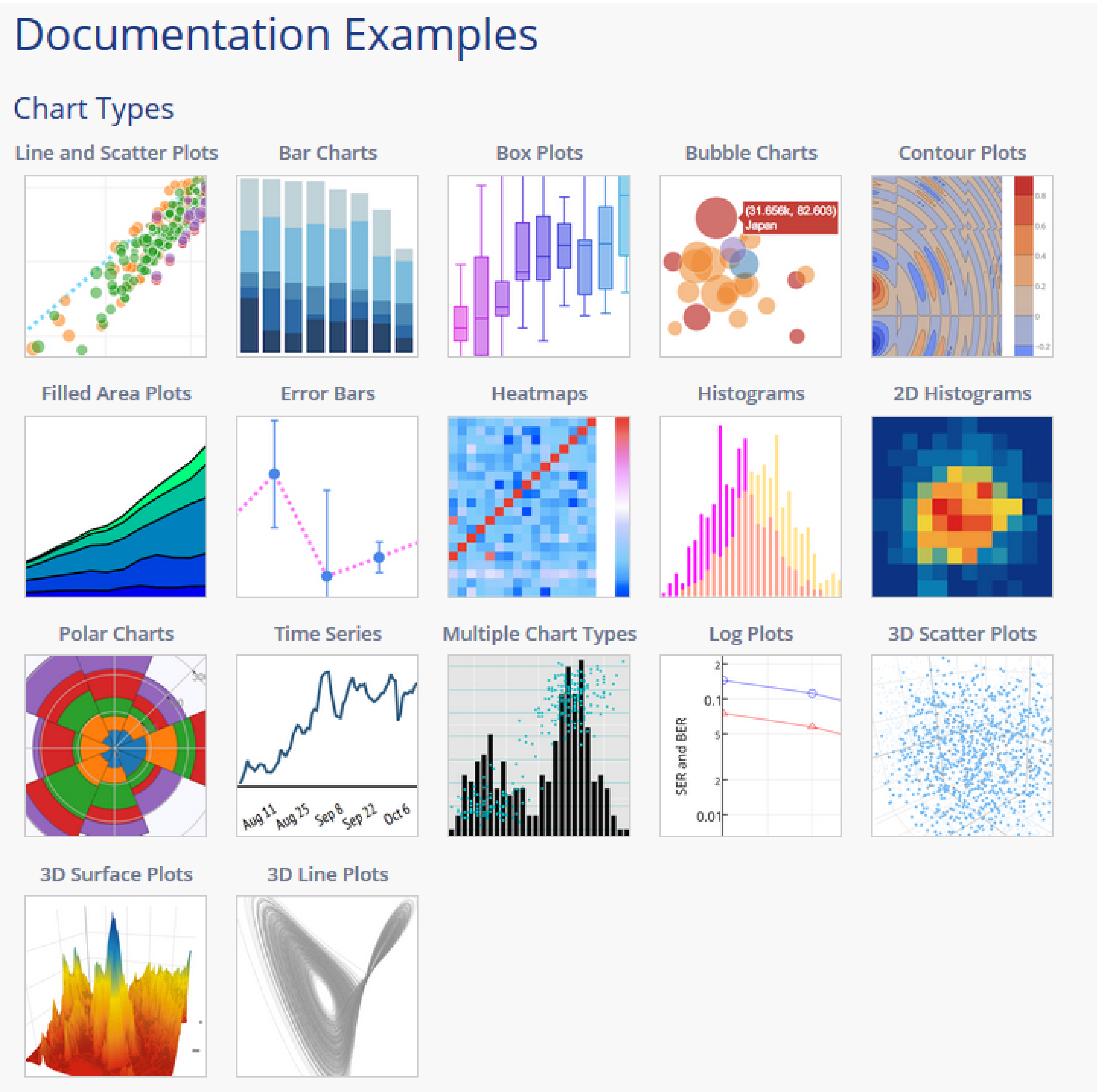
XII FDA accepted using R for drug trials!



...some polish books



...and interactive (JScript)

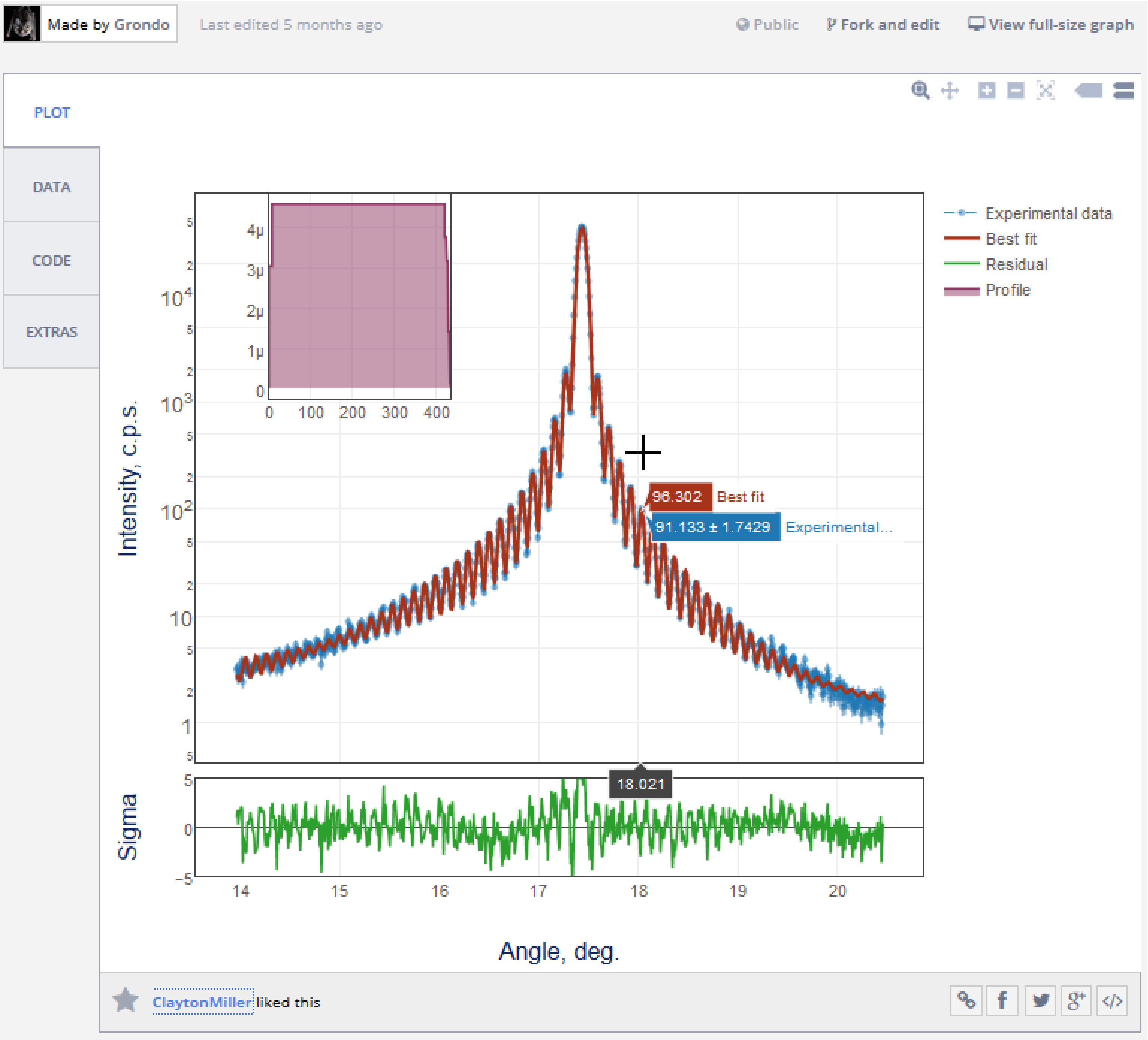


Documentation Examples

Chart Types

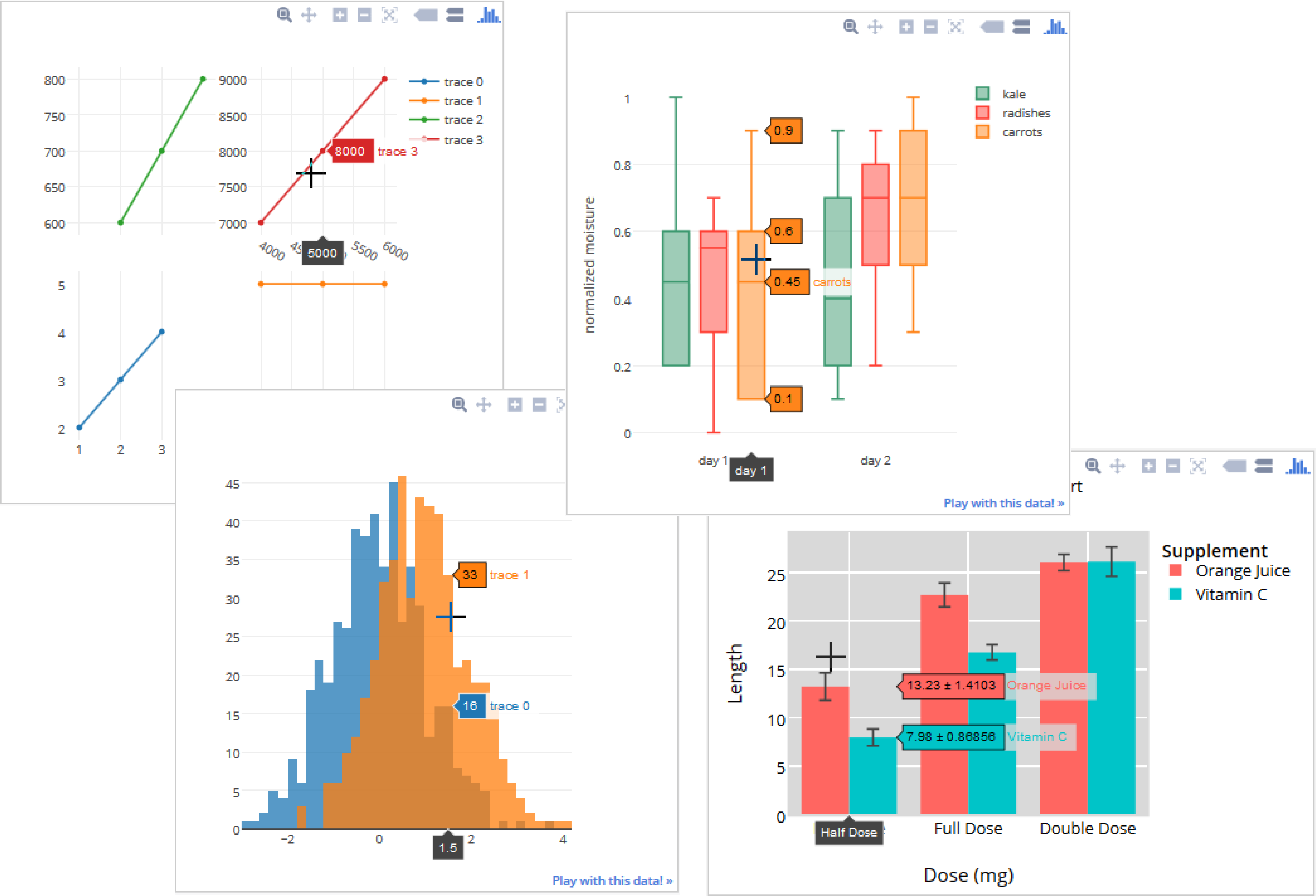
Line and Scatter Plots
Bar Charts
Box Plots
Bubble Charts
Contour Plots
Filled Area Plots
Error Bars
Heatmaps
Histograms
2D Histograms
Polar Charts
Time Series
Multiple Chart Types
Log Plots
3D Scatter Plots
3D Surface Plots
3D Line Plots

...and interactive (JScript)



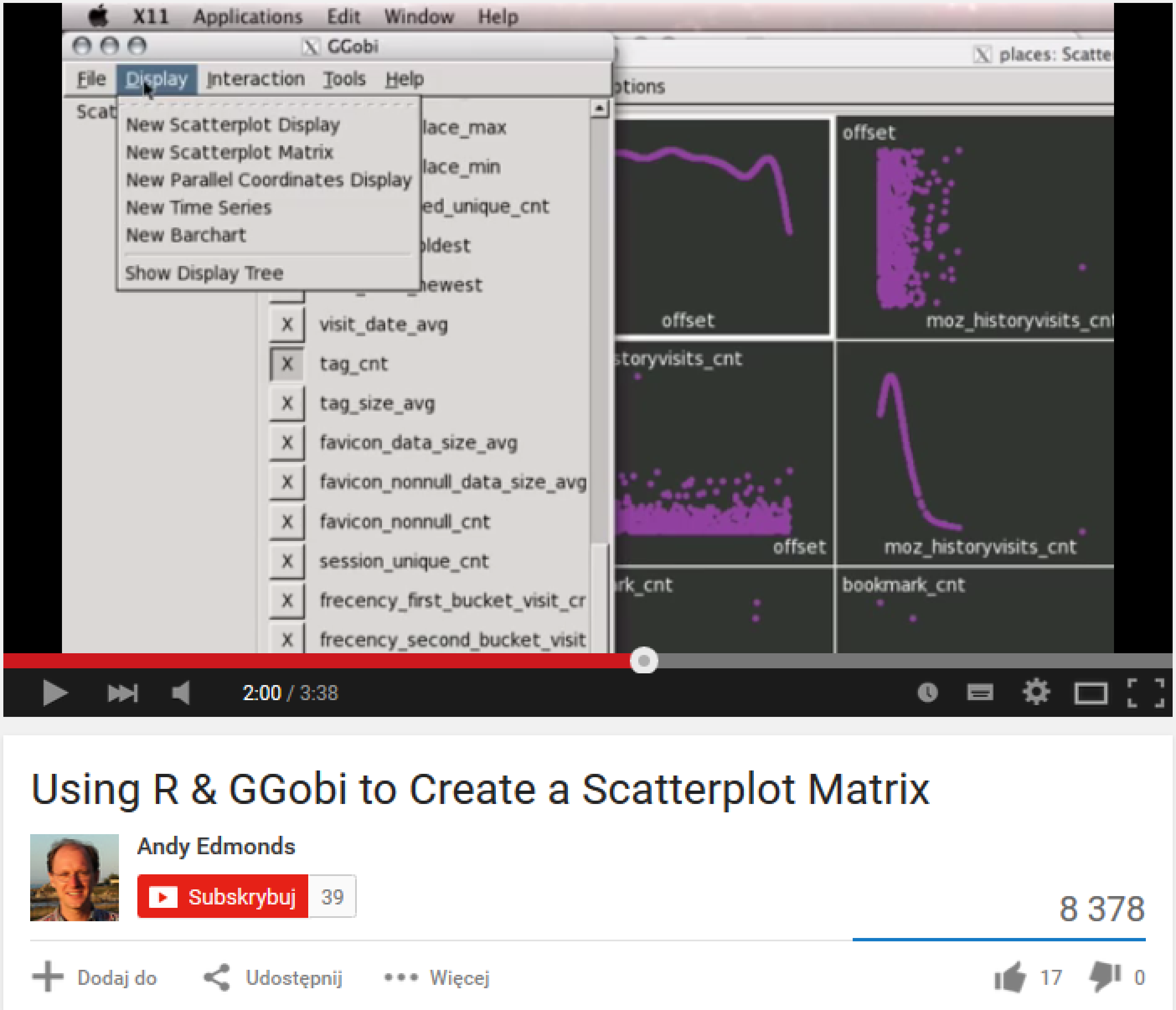
ClaytonMiller liked this

...and interactive (JScript)



R &GGobi in action

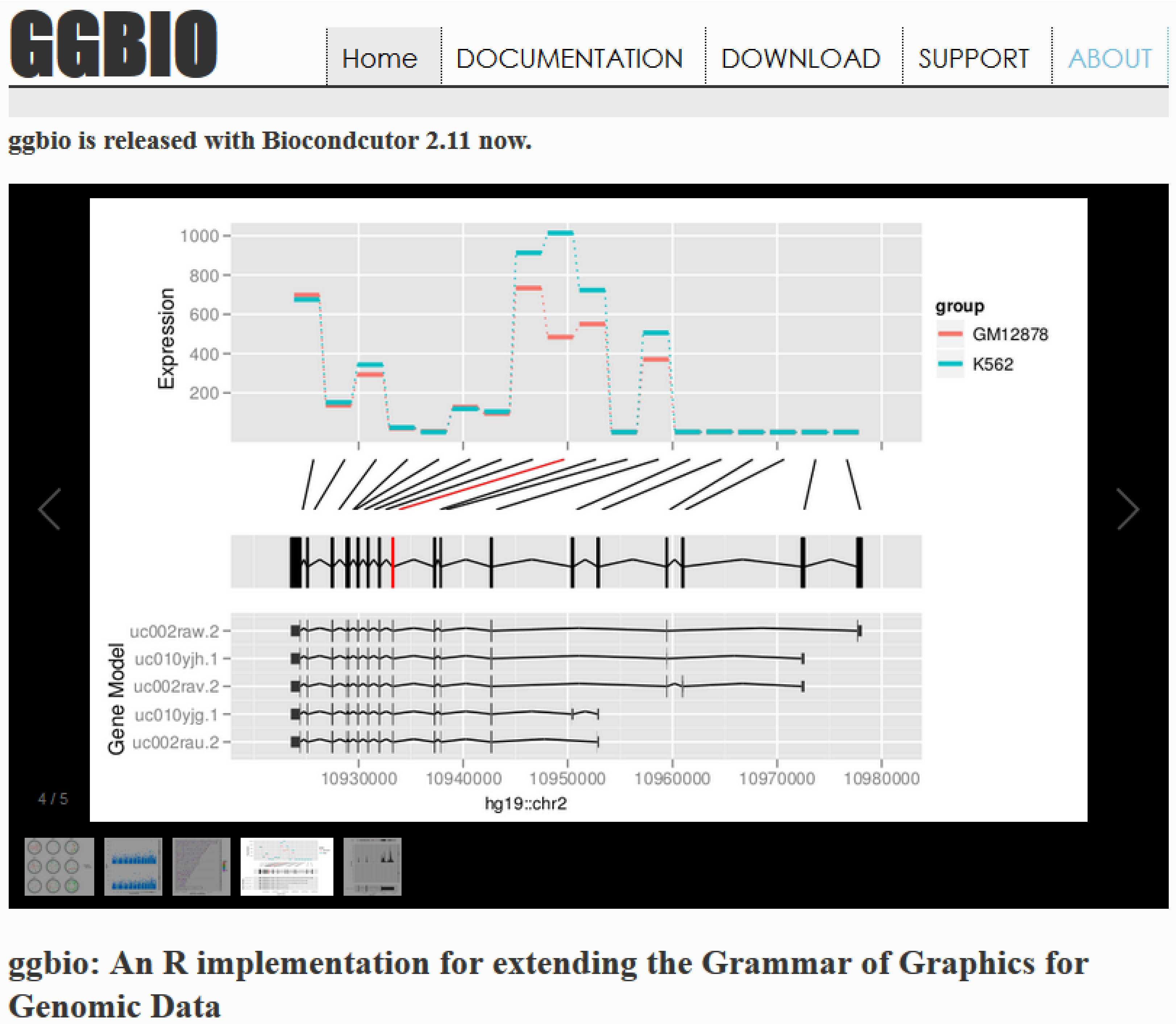
[YouTube: Using R & GGobi to Create a Scatterplot Matrix]



Using R & GGobi to Create a Scatterplot Matrix

Andy Edmonds

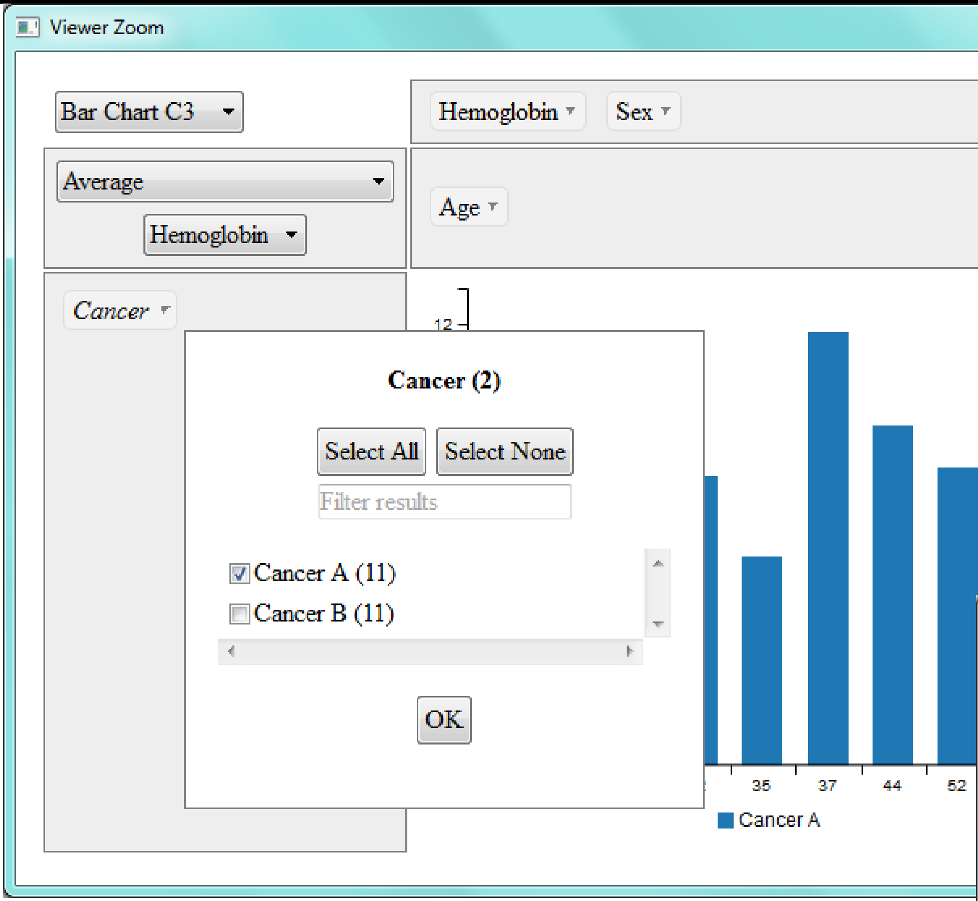
GGBio for geneticists

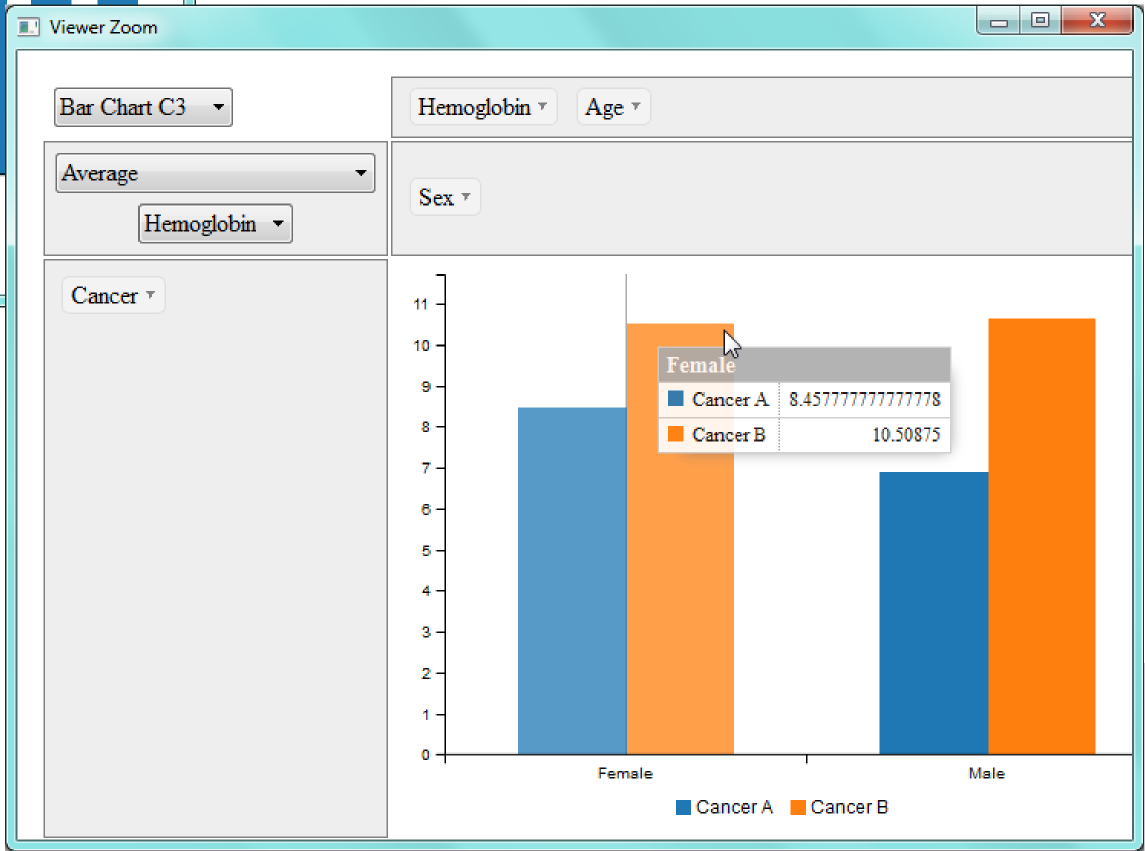


ggbio is released with Biocondcutor 2.11 now.

ggbio: An R implementation for extending the Grammar of Graphics for Genomic Data

rpivotTable – inteRactive pivot tables



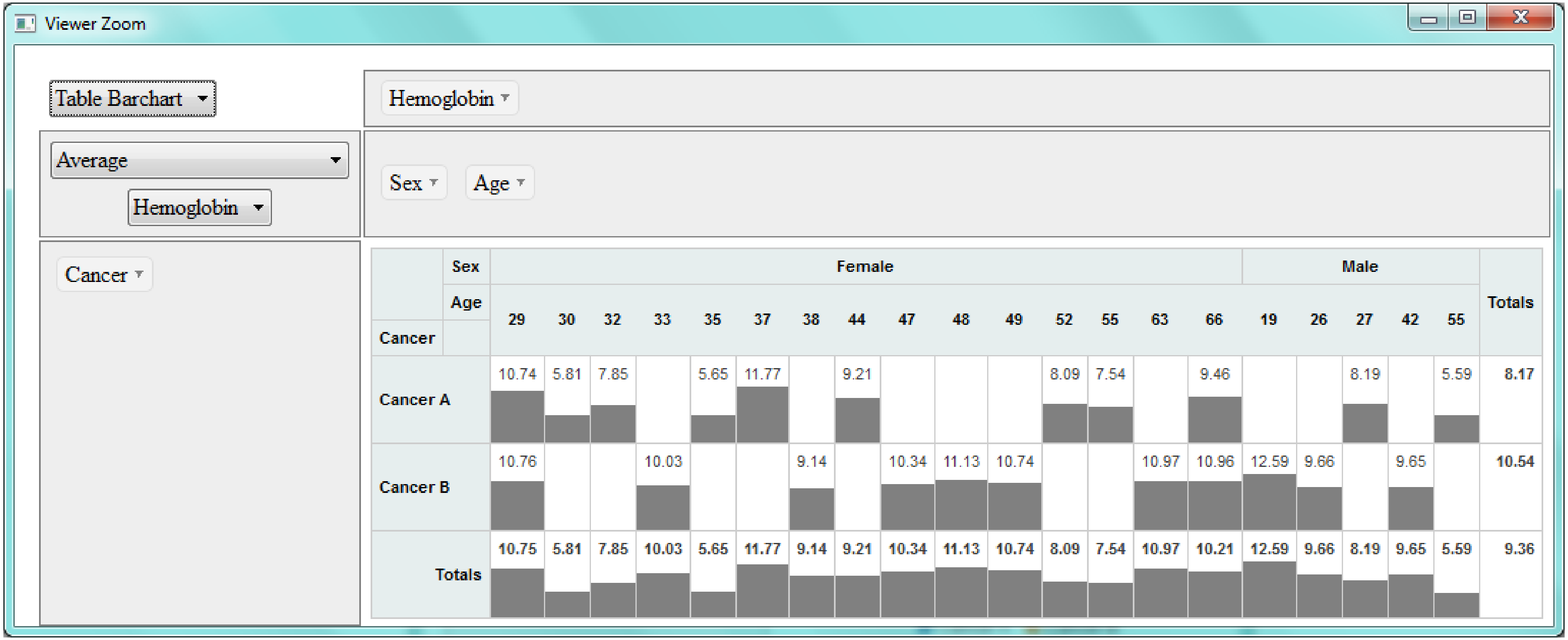


Based on PivotTable.js by
Nicolas Kruchten

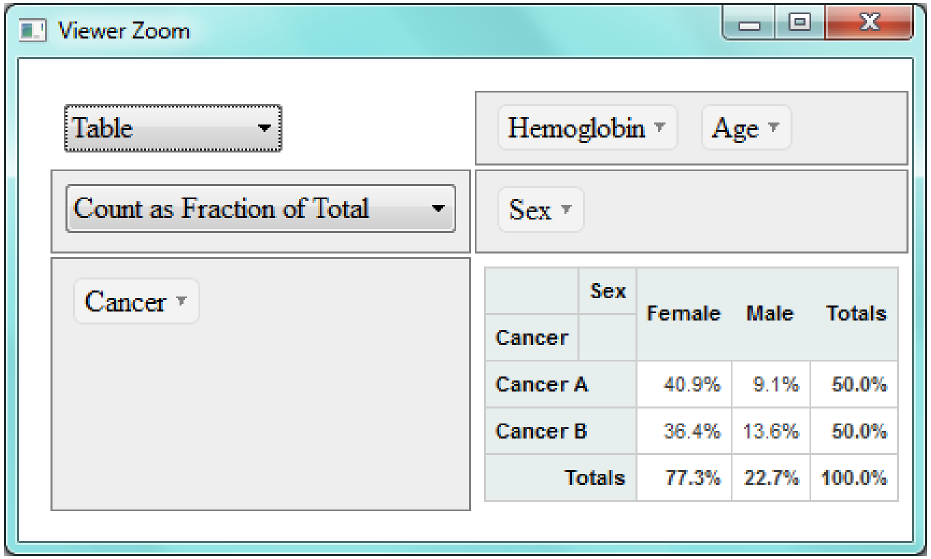
http://nicolas.kruchten.com/pivottable

http://www.magesblog.com/2015...

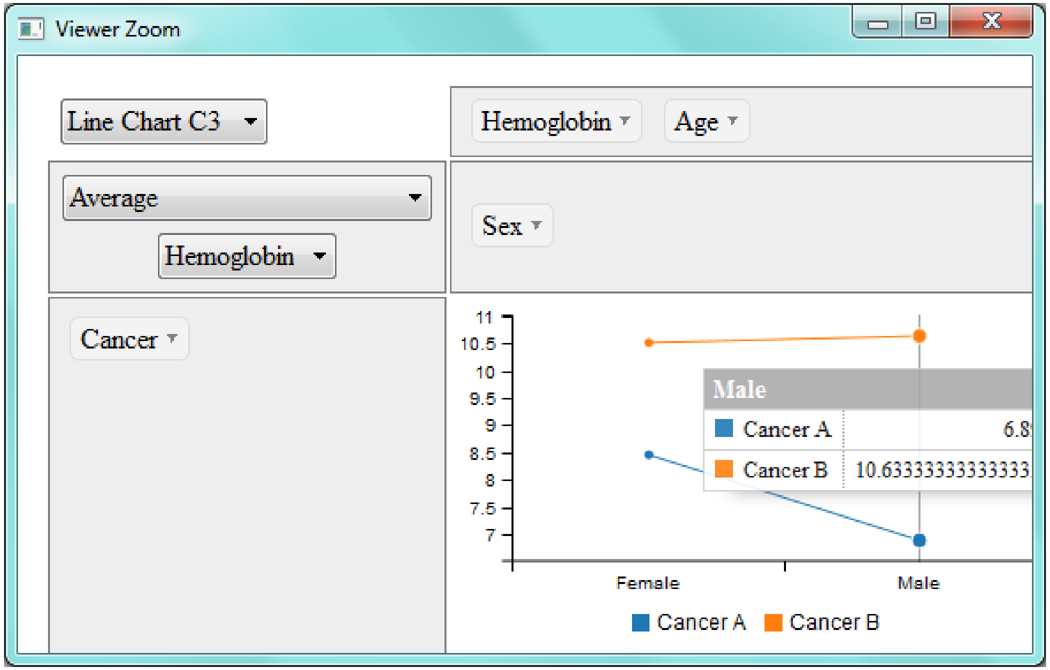
rpivotTable



Sex	Female															Male					Totals
Age	29	30	32	33	35	37	38	44	47	48	49	52	55	63	66	19	26	27	42	55	
Cancer																					
Cancer A	10.74	5.81	7.85		5.65	11.77		9.21				8.09	7.54		9.46			8.19		5.59	8.17
Cancer B	10.76			10.03			9.14		10.34	11.13	10.74			10.97	10.96	12.59	9.66		9.65		10.54
Totals	10.75	5.81	7.85	10.03	5.65	11.77	9.14	9.21	10.34	11.13	10.74	8.09	7.54	10.97	10.21	12.59	9.66	8.19	9.65	5.59	9.36



Sex	Female	Male	Totals
Cancer			
Cancer A	40.9%	9.1%	50.0%
Cancer B	36.4%	13.6%	50.0%
Totals	77.3%	22.7%	100.0%



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

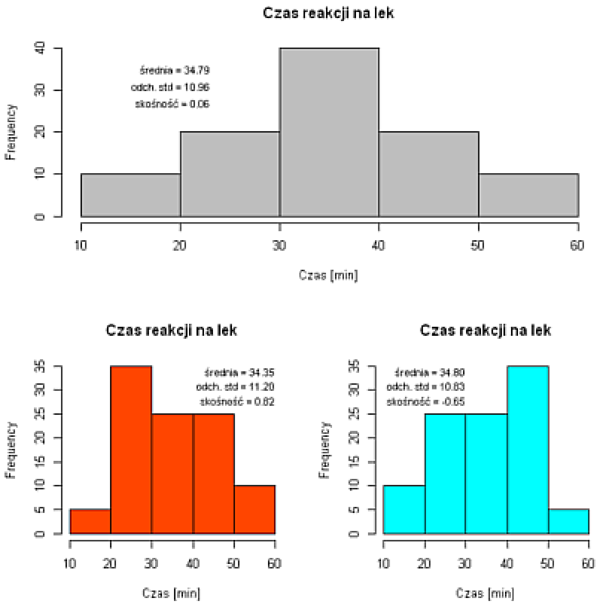
by Deepayan Sarka
University of Wisconsin

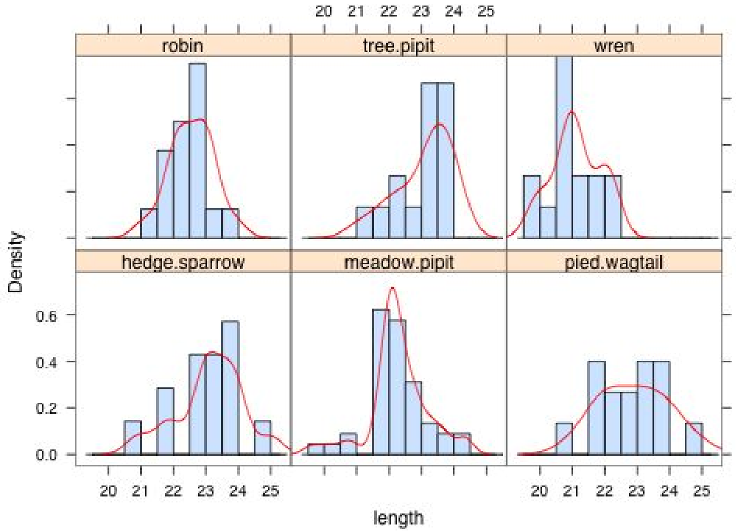
designed for viewing multivariable datasets (grid of conditioned plots)
well readable
closed set of available diagrams: Barplot, Dotplot, Box and Whiskers, Histogram, Density, QQ, Scatterplot
incompatible with other systems

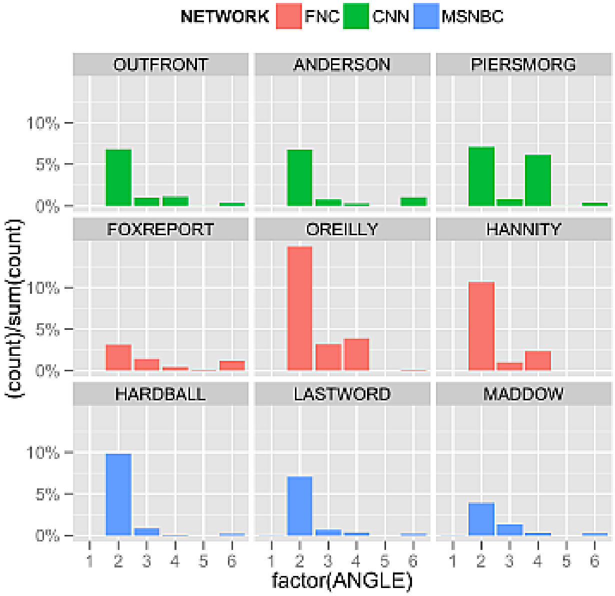
ggplot2

by Hadley Wickham
Rice University

high-level and well organized implementation of *Grammar of Graphics*
powerful, higly 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

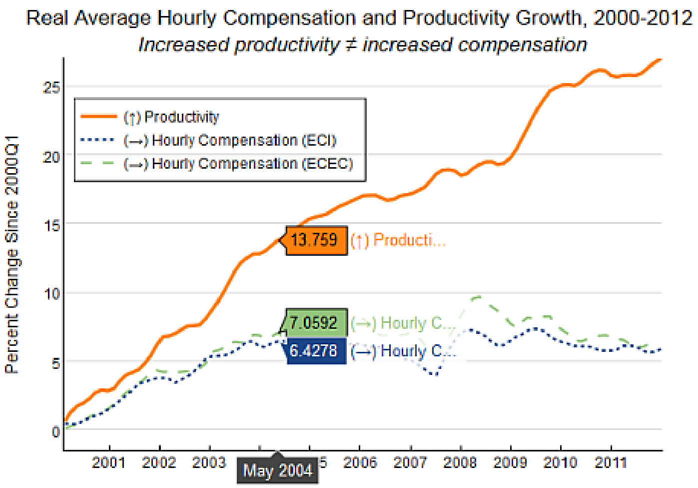
based on ggplot2
interactive (JScript)
WWW enabled (HTML/JSON/JS)
breathtaking output –must see!
well readable, clean
rich library of examples
incompatible with other systems

ggvis
by RStudio team

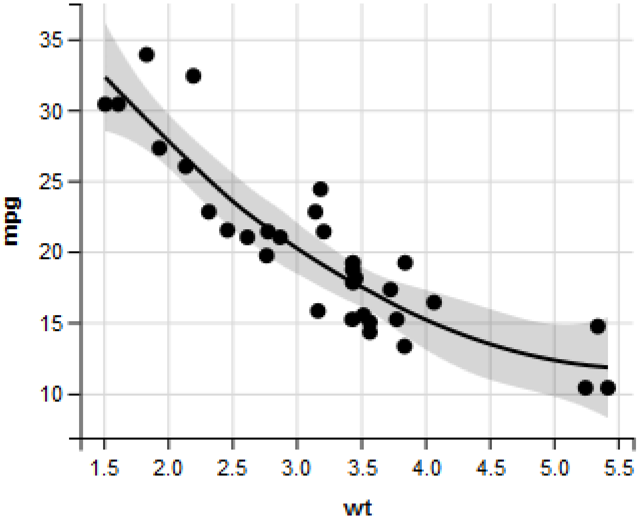
another implementation of *Grammar of graphics*, similar in spirit to ggplot2
Interactive (JScript)
WWW enabled (HTML, JS)
well readable
designed for Shiny

txtplot
by Bjoern Bornkamp

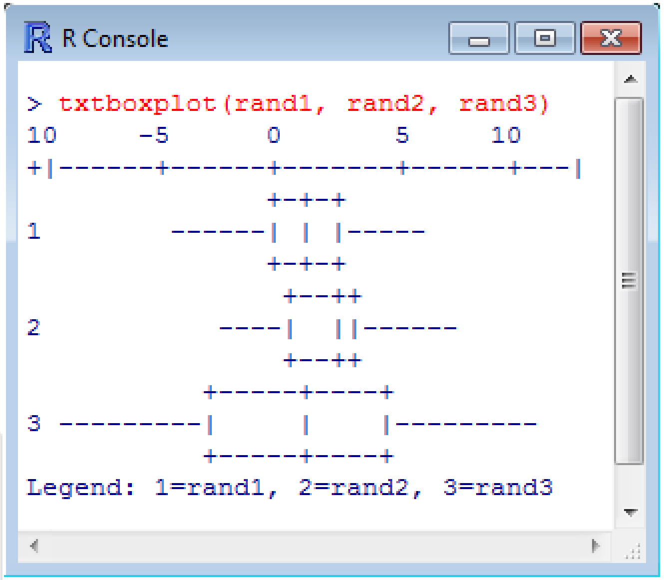
produces graphs in pure ASCII
rudimentary output
closed set of plots (boxplot, lineplot, barplot, density, ACF)
really useful when resources are limited (mobile devices, simple LCD displays, etc.) or output must be textual
incompatible with others



Source: Economic Policy Institute



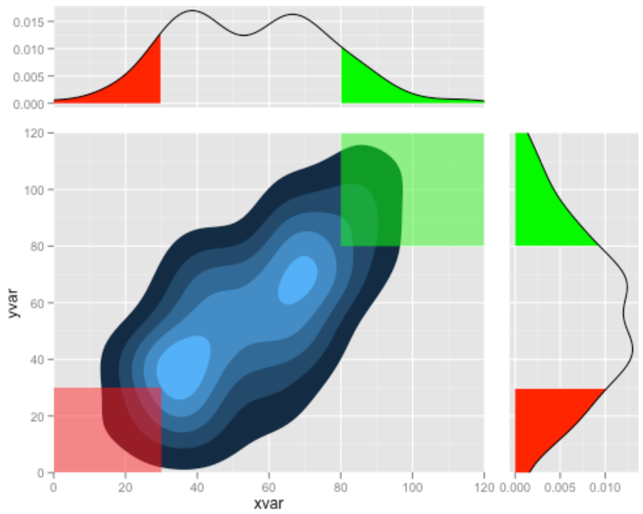

Smoothing span
0.2
0.85

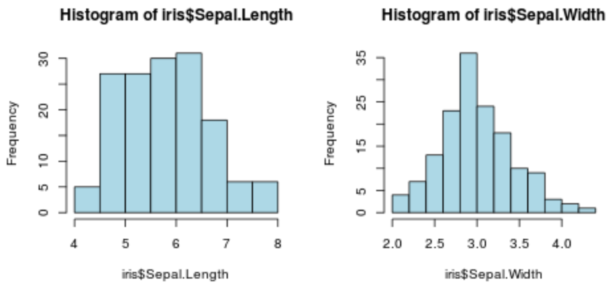


```
> txtboxplot(rand1, rand2, rand3)
10     -5      0       5     10
+|------+------+-------+------+---|
               +-+-+
1        ------| | |-----
               +-+-+
                +--++
2           ----|  ||------
                +--++
           +-----+----+
3 ---------|     |    |---------
           +-----+----+
Legend: 1=rand1, 2=rand2, 3=rand3
```

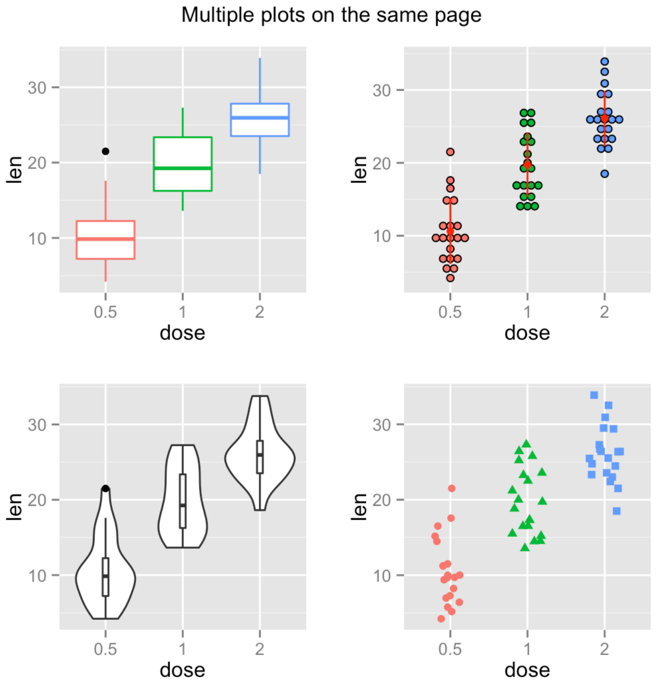
Additional helpers for ggplot2: gridExtra

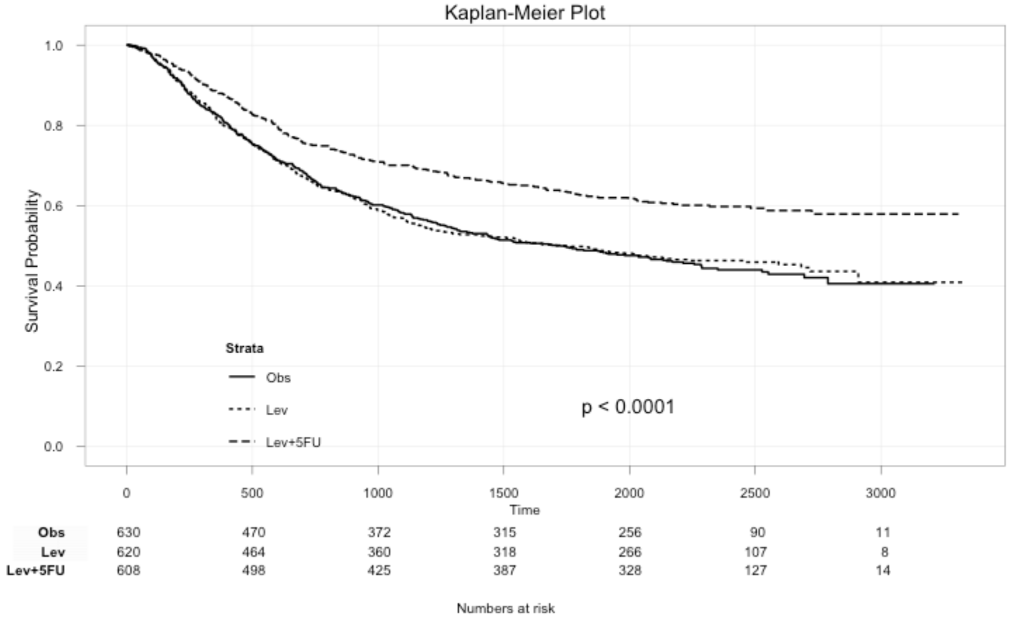
Helps to arrange multiple ggplot2 objects on the same page





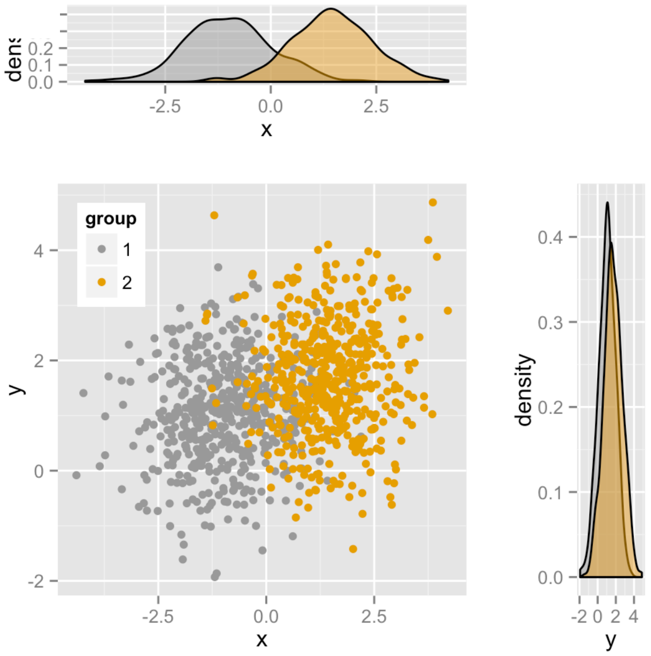
	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa





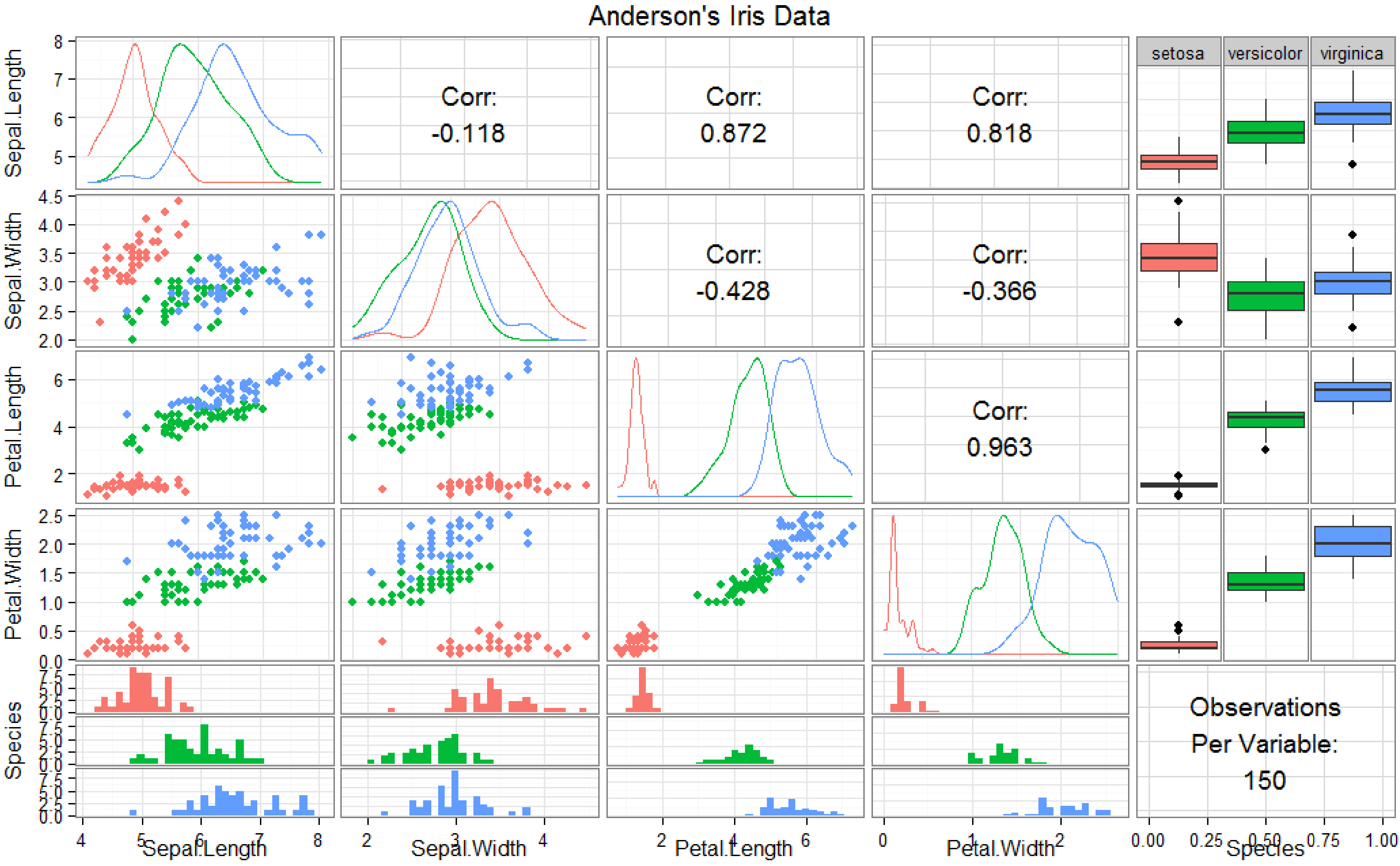
	0	500	1000	1500	2000	2500	3000
Obs	630	470	372	315	256	90	11
Lev	620	464	360	318	266	107	8
Lev+5FU	608	498	425	387	328	127	14

Numbers at risk



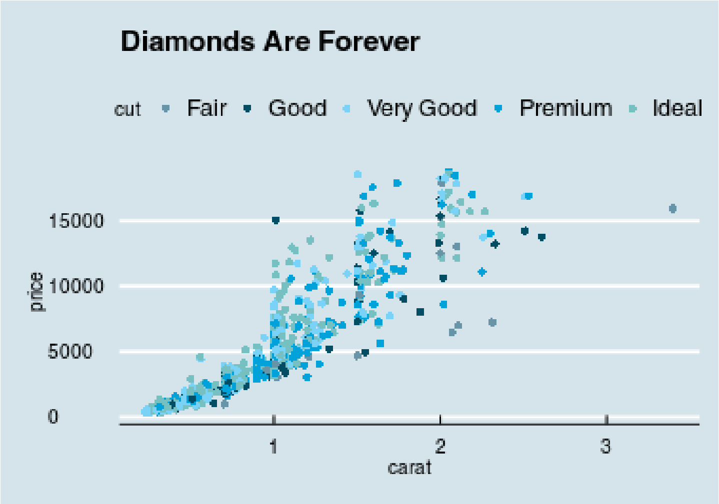
Additional helpers for ggplot2: GGally

Creates a matrix of ggplot2 graphs for data exploration purposes.

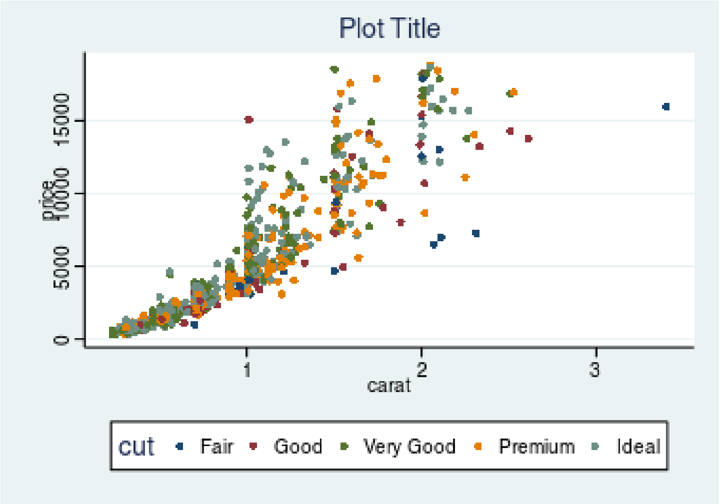


Additional helpers for ggplot2: ggthemes

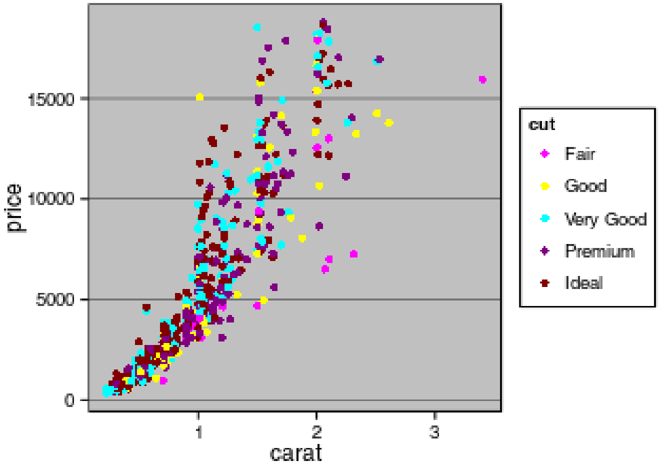
A set of themes (>15) for ggplot2: http://github.com/jrnold/ggthemes



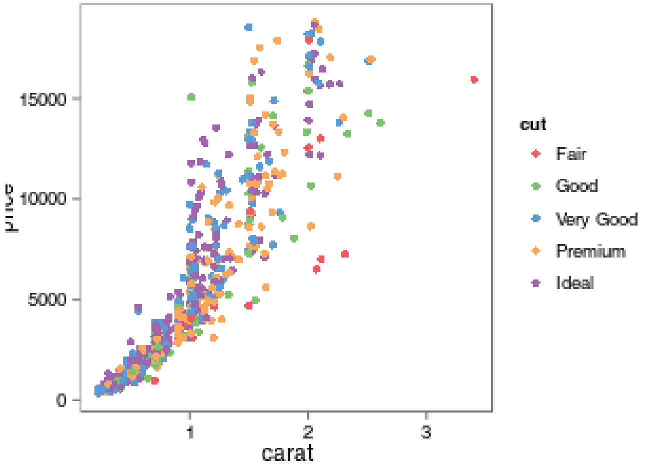
The Economist magazine theme



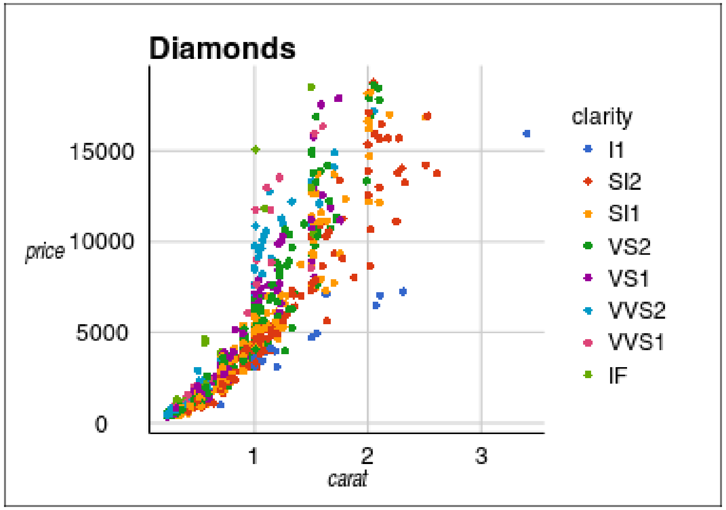
Stata theme



Excel theme

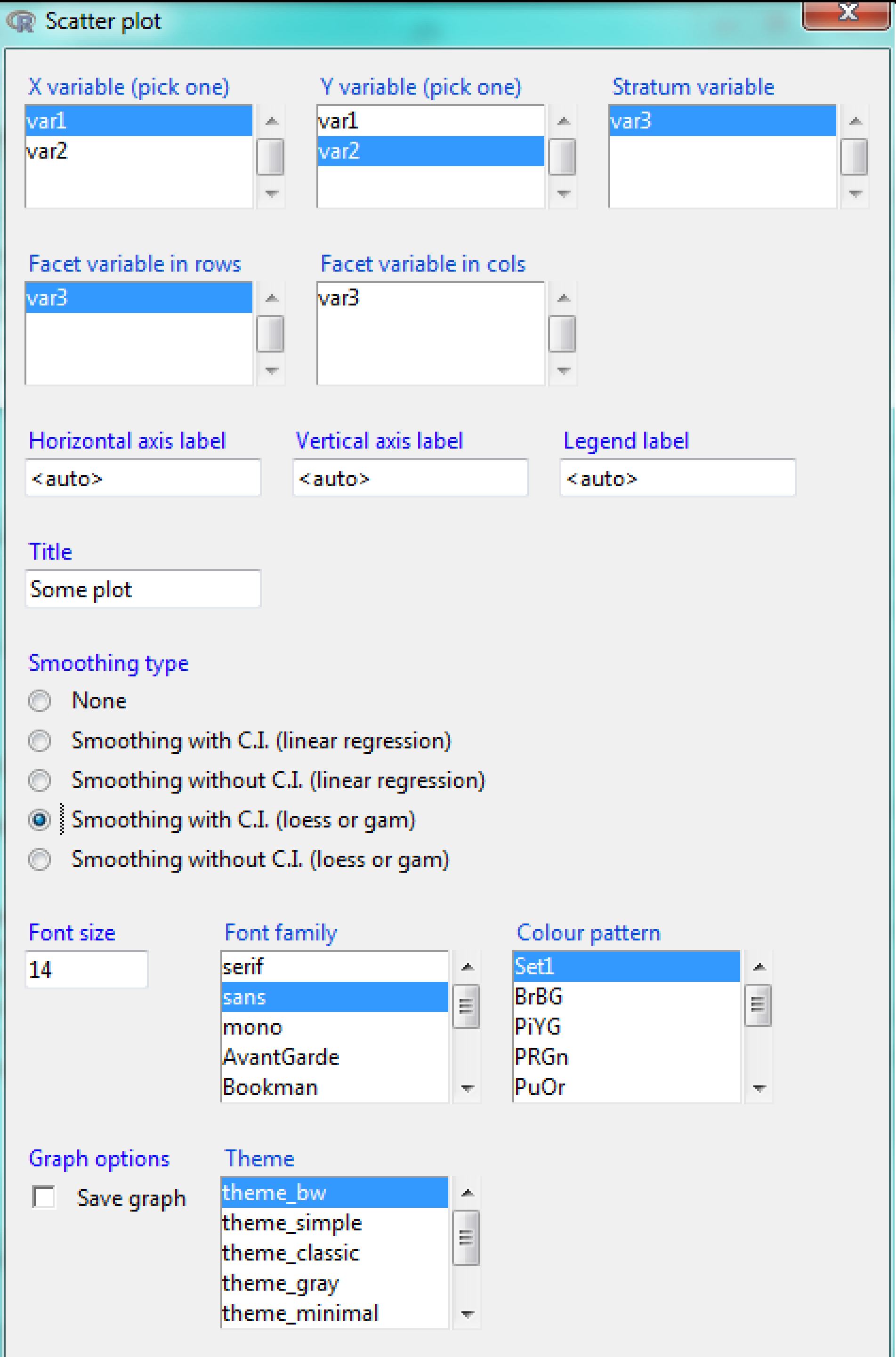


[Stephen Few's rules]()



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)

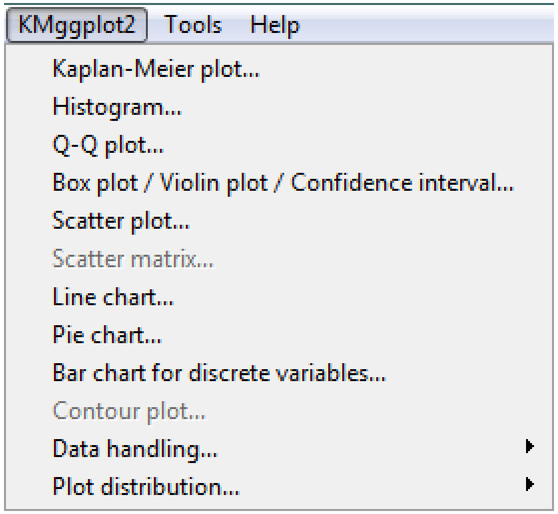
Font size: 14

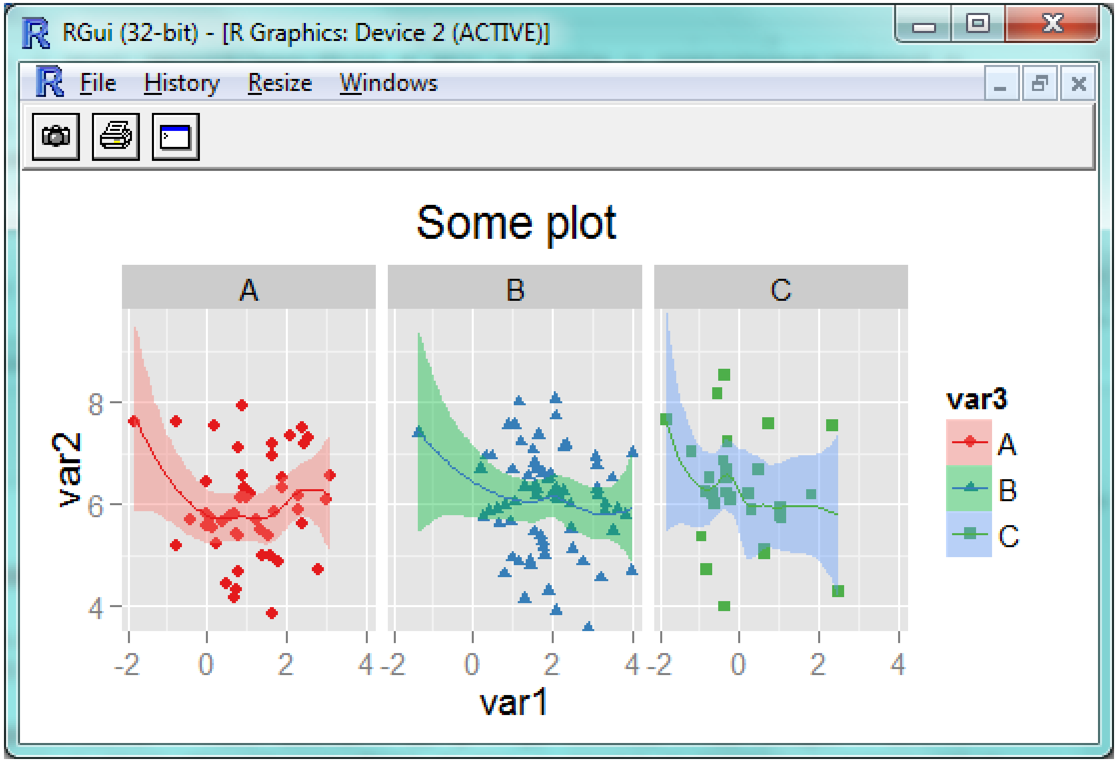
Font family
- serif
- **sans**
- mono
- AvantGarde
- Bookman

Colour pattern
- **Set1**
- BrBG
- PiYG
- PRGn
- PuOr

Graph options
- ☐ Save graph

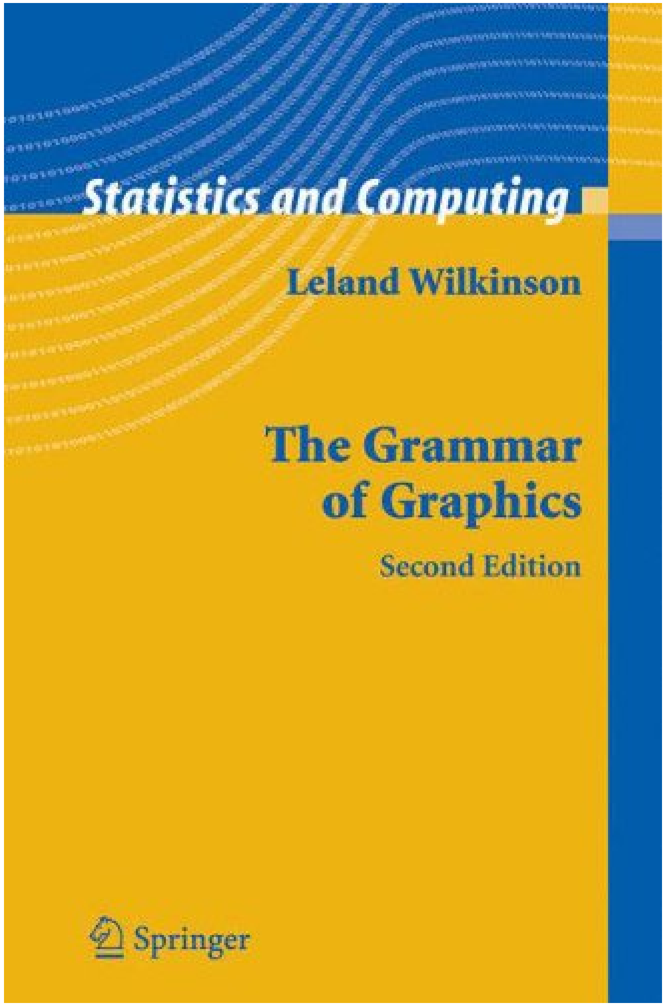
Theme
- **theme_bw**
- theme_simple
- theme_classic
- theme_gray
- theme_minimal





Deducer

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”.

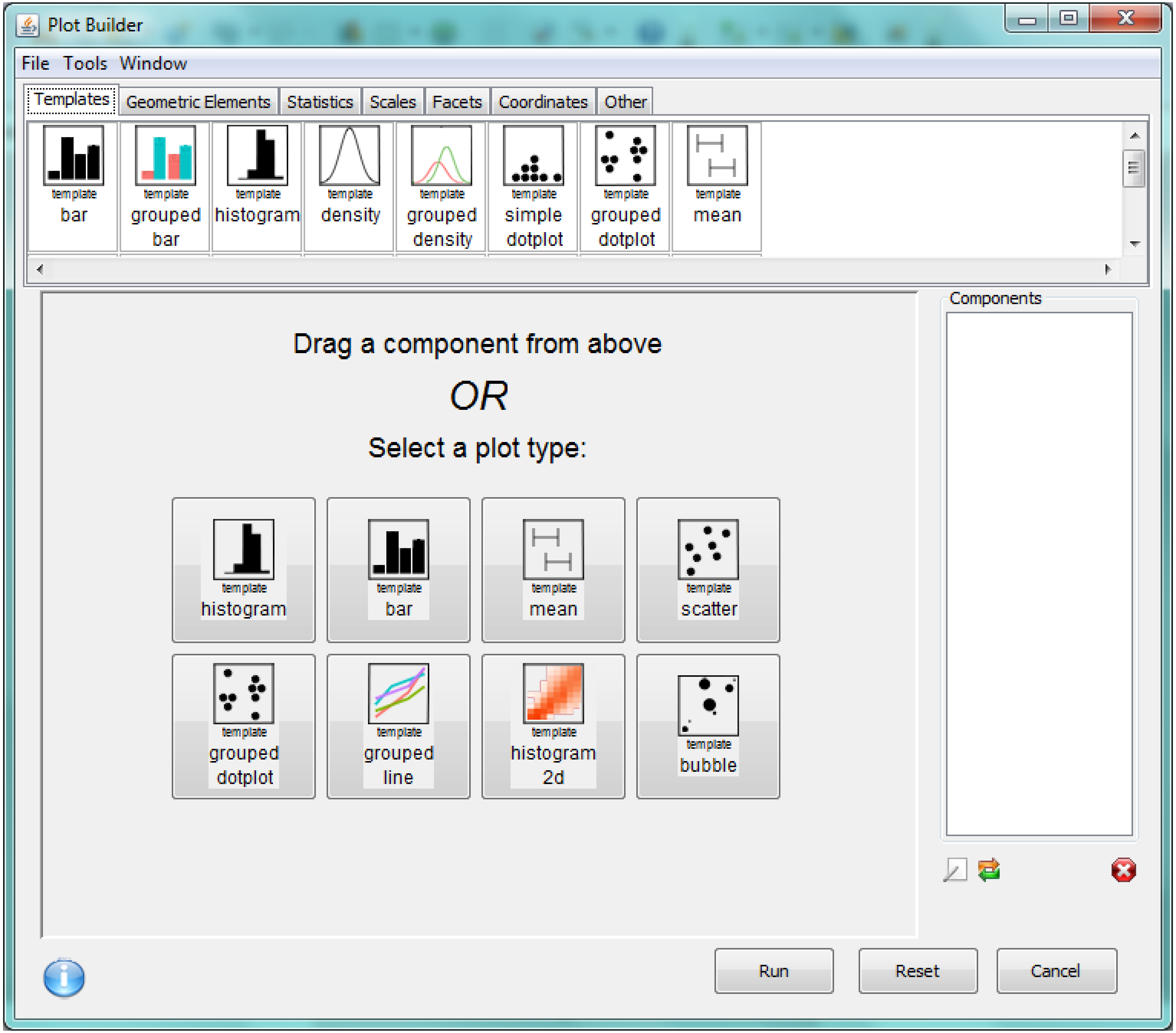


The Deducer package is an attempt to achieve this goal drawing from the power of ggplot2.

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

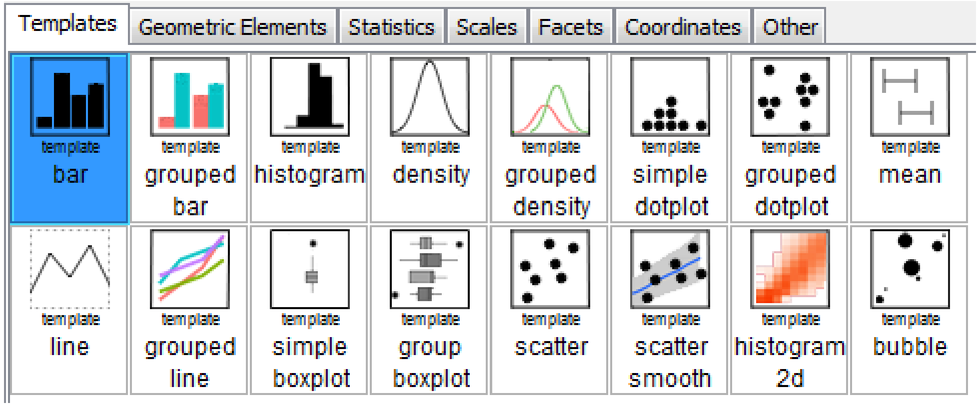


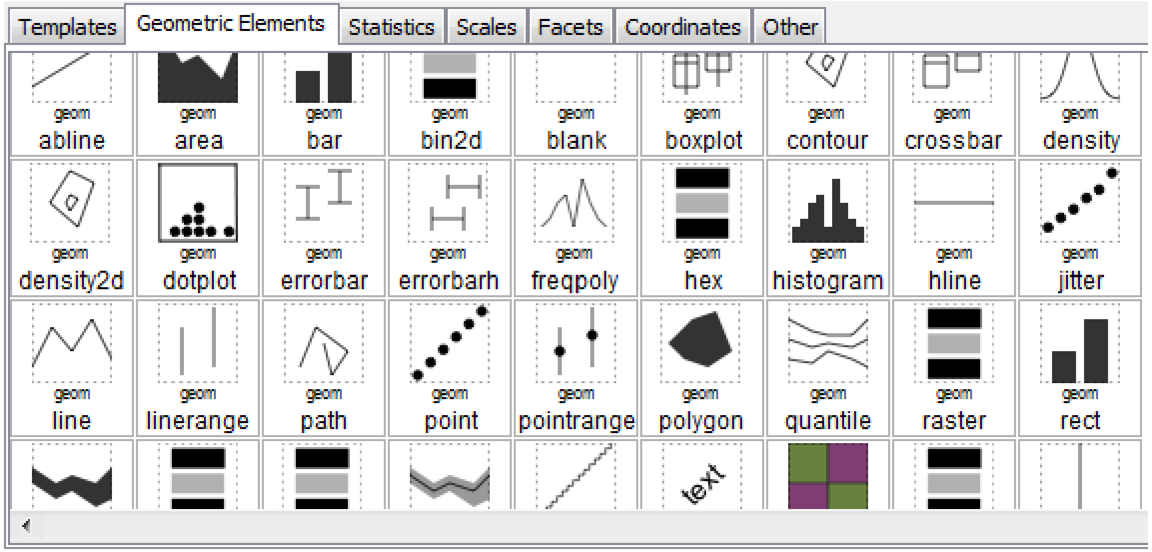
Drag a component from above

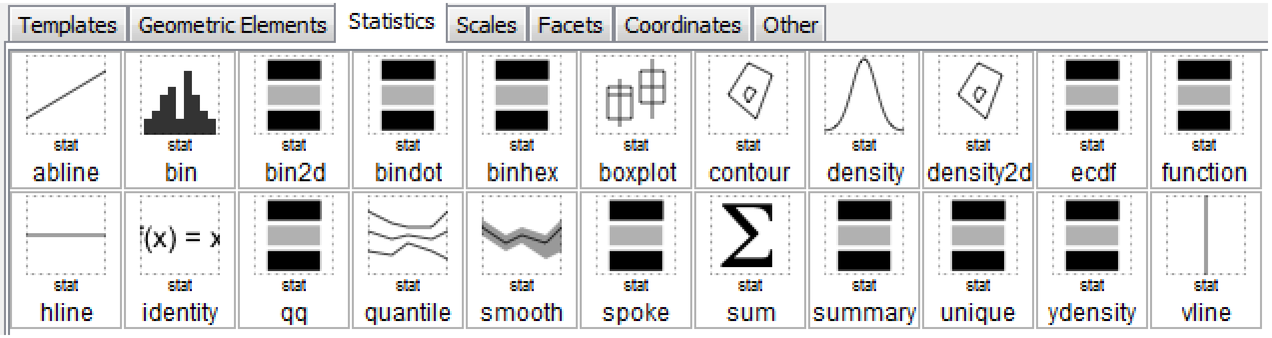
OR

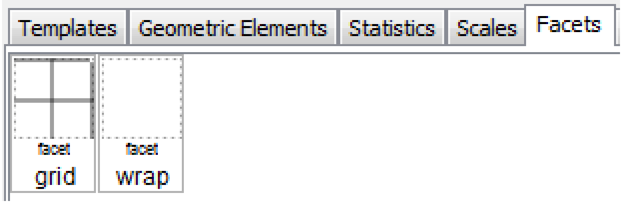
Select a plot type:

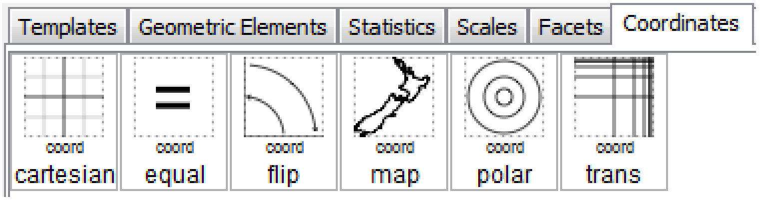
Deducer

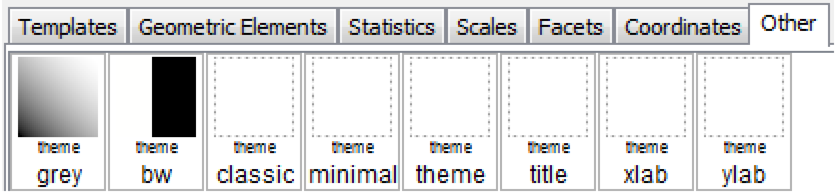

Templates
Geometric Elements
Statistics
Scales
Facets
Coordinates
Other
template bar
template grouped bar
template histogram
template density
template grouped density
template simple dotplot
template grouped dotplot
template mean
template line
template grouped line
template simple boxplot
template group boxplot
template scatter
template scatter smooth
template histogram 2d
template bubble

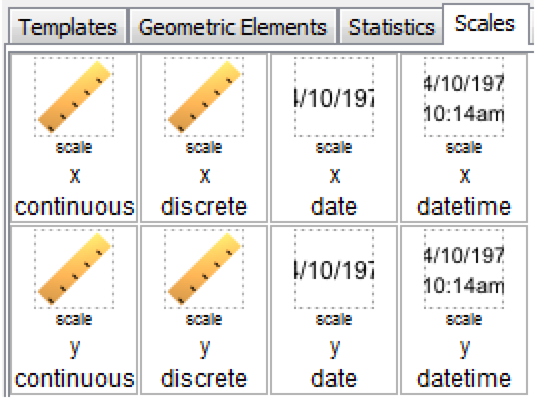

Templates
Geometric Elements
Statistics
Scales
Facets
Coordinates
Other
geom abline
geom area
geom bar
geom bin2d
geom blank
geom boxplot
geom contour
geom crossbar
geom density
geom density2d
geom dotplot
geom errorbar
geom errorbarh
geom freqpoly
geom hex
geom histogram
geom hline
geom jitter
geom line
geom linerange
geom path
geom point
geom pointrange
geom polygon
geom quantile
geom raster
geom rect
text


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
f(x) = x
stat identity
stat qq
stat quantile
stat smooth
stat spoke
stat sum
stat summary
stat unique
stat ydensity
stat vline

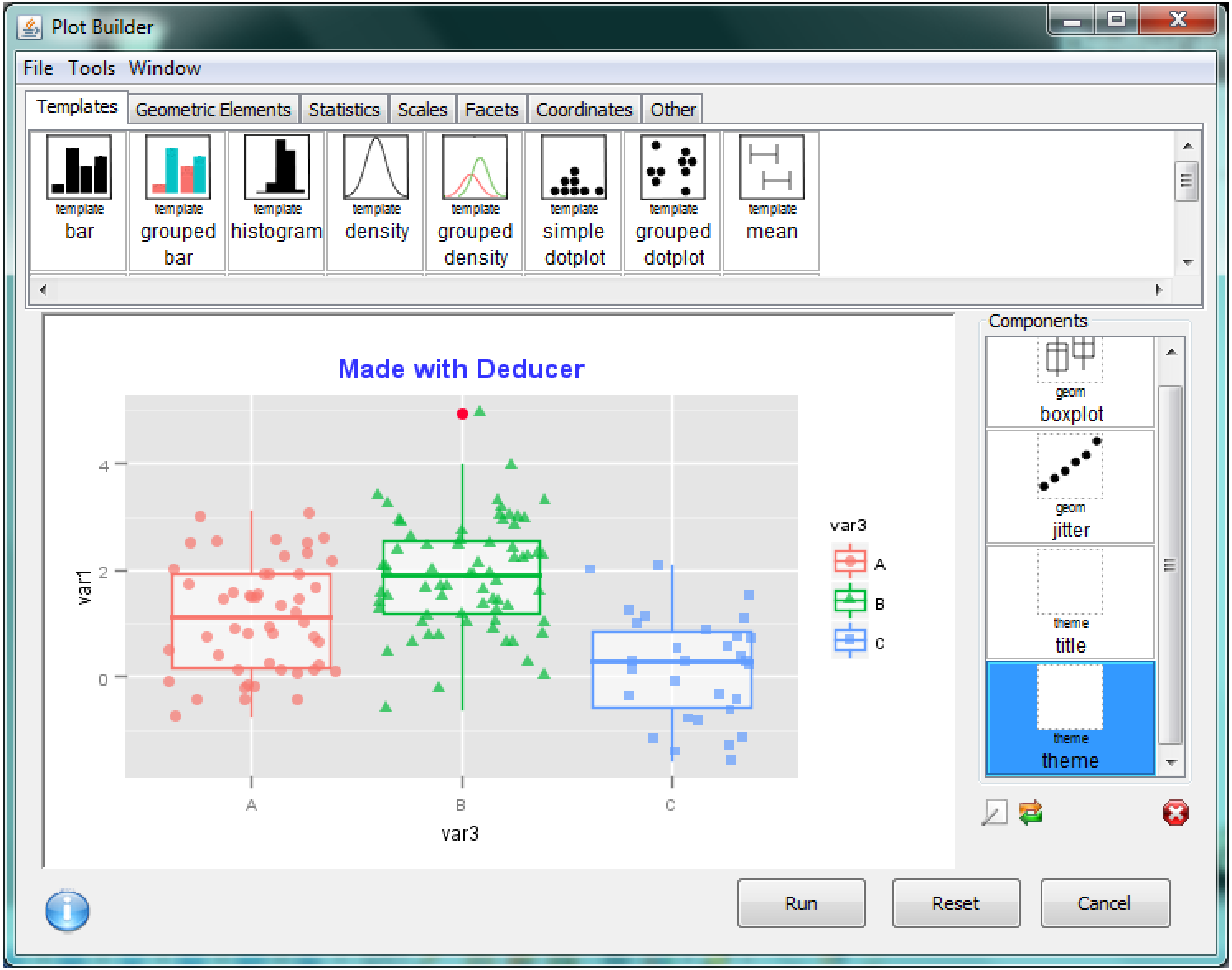

Templates
Geometric Elements
Statistics
Scales
Facets
facet grid
facet wrap


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
theme grey
theme bw
theme classic
theme minimal
theme theme
theme title
theme xlab
theme ylab

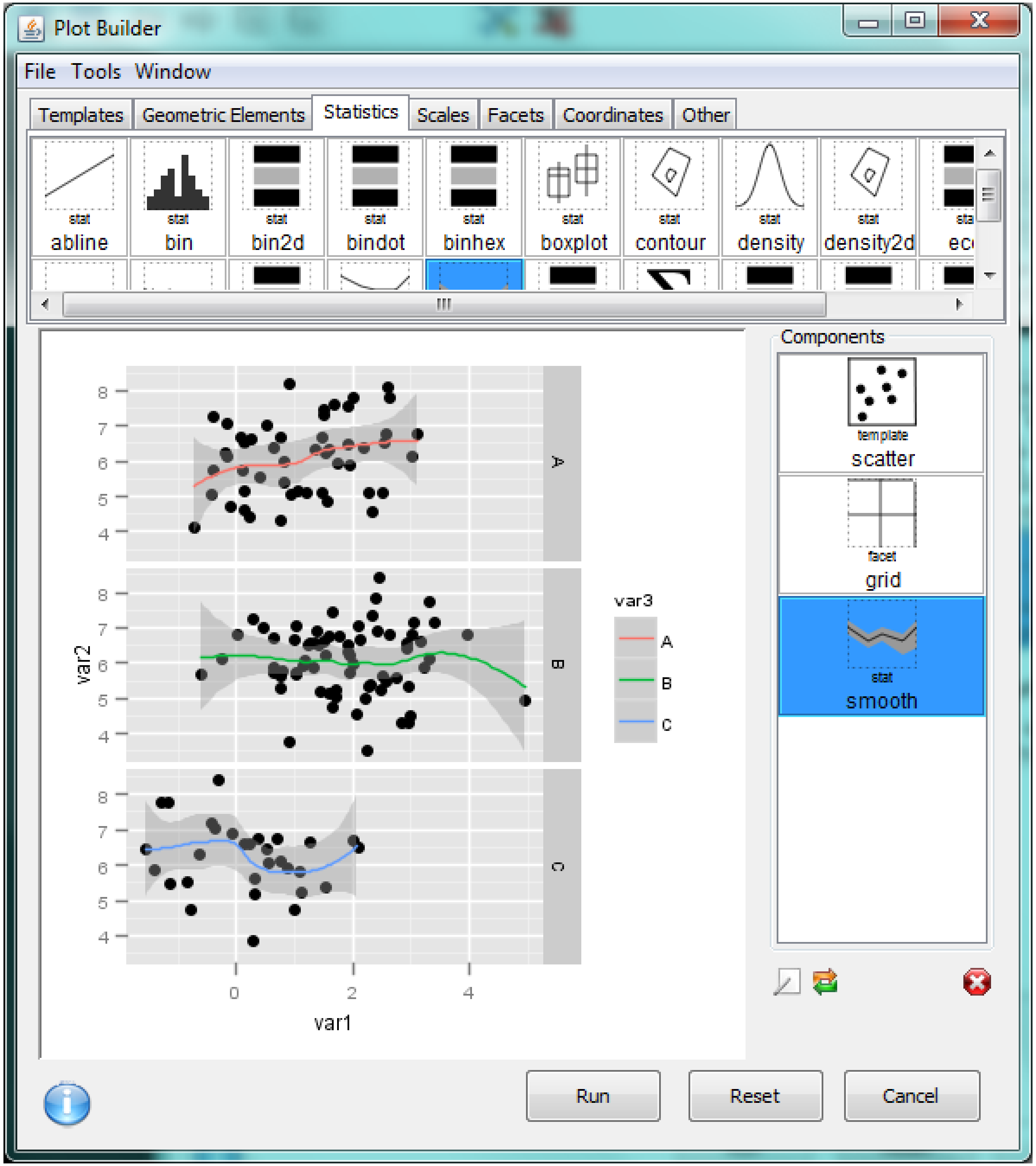

Templates
Geometric Elements
Statistics
Scales
scale x continuous
scale x discrete
1/10/197
scale x date
4/10/197
10:14am
scale x datetime
scale y continuous
scale y discrete
1/10/197
scale y date
4/10/197
10:14am
scale y datetime

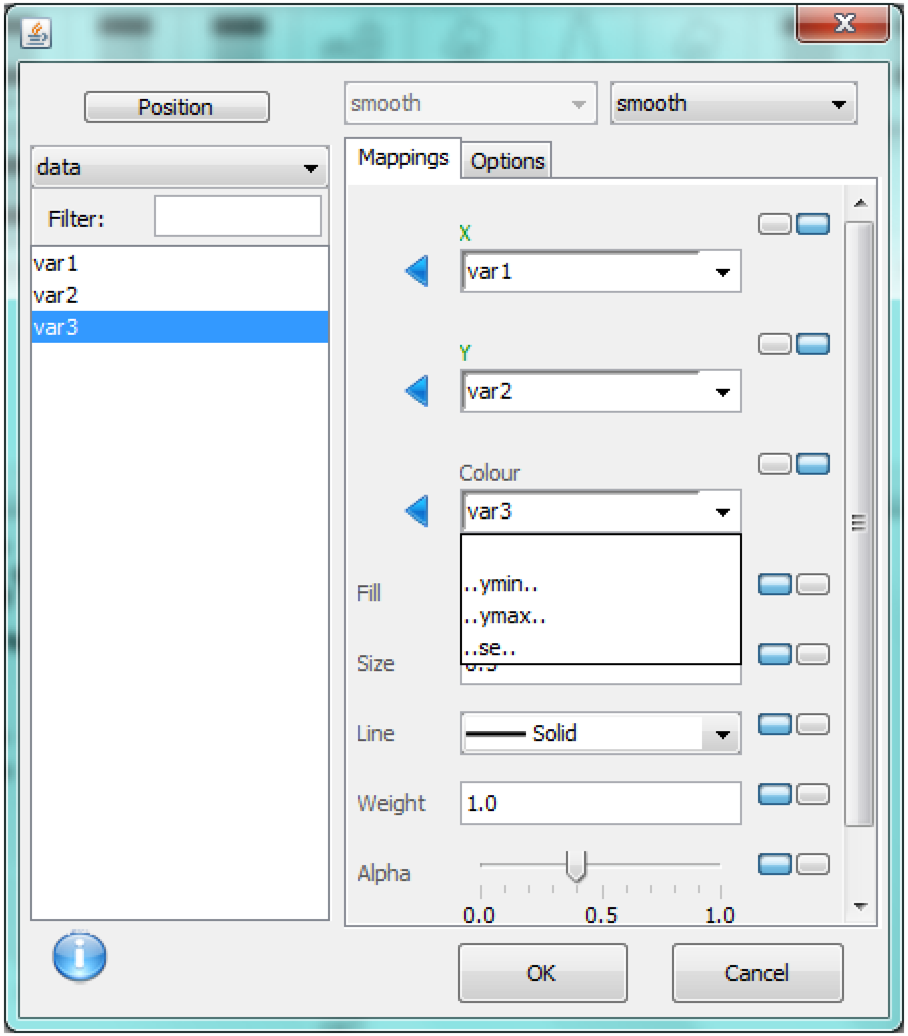
Deducer



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





smooth

X: var1

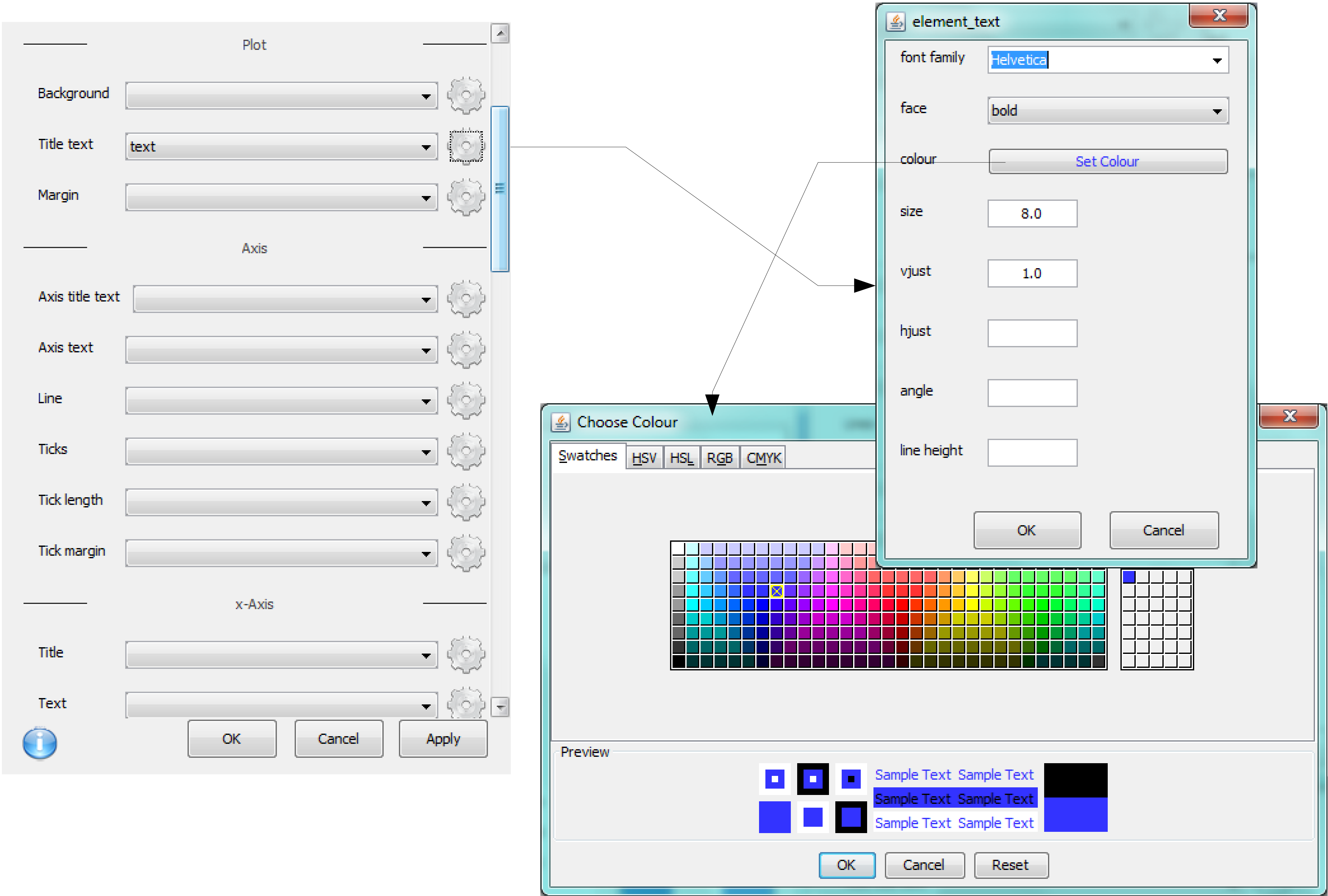
Y: var2

Colour: var3

Line: Solid

Weight: 1.0

Deducer – control dozens of graph properties



Plot

Background:

Title text: text

Margin:

Axis

Axis title text:

Axis text:

Line:

Ticks:

Tick length:

Tick margin:

x-Axis

Title:

Text:

element_text

font family: Helvetica

face: bold

colour: Set Colour

size: 8.0

vjust: 1.0

hjust:

angle:

line height:

Choose Colour

Swatches | HSV | HSL | RGB | CMYK

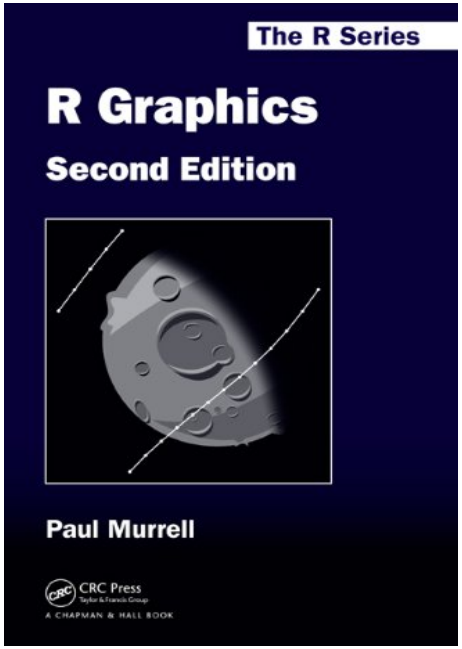
Preview

Sample Text Sample Text

Sample Text Sample Text

Sample Text Sample Text

Some books about data visualization in R


The R Series

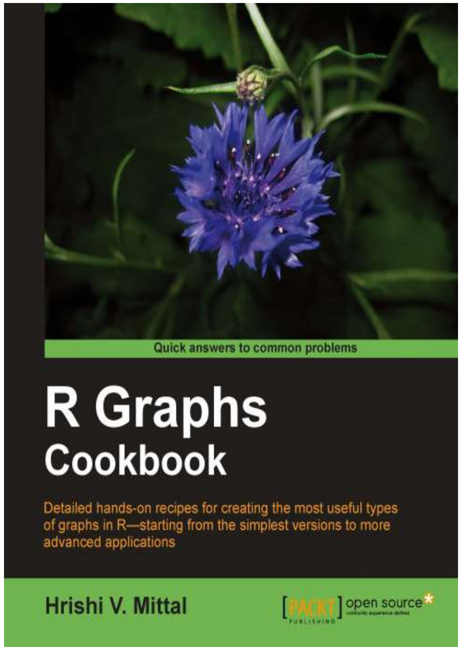
R Graphics

Second Edition

Paul Murrell

CRC Press

A CHAPMAN & HALL BOOK

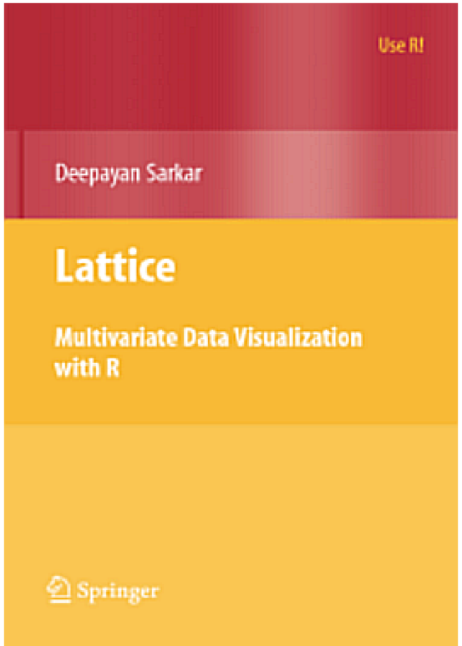

Quick answers to common problems

R Graphs Cookbook

Detailed hands-on recipes for creating the most useful types of graphs in R—starting from the simplest versions to more advanced applications

Hrishi V. Mittal

[PACKT] open source

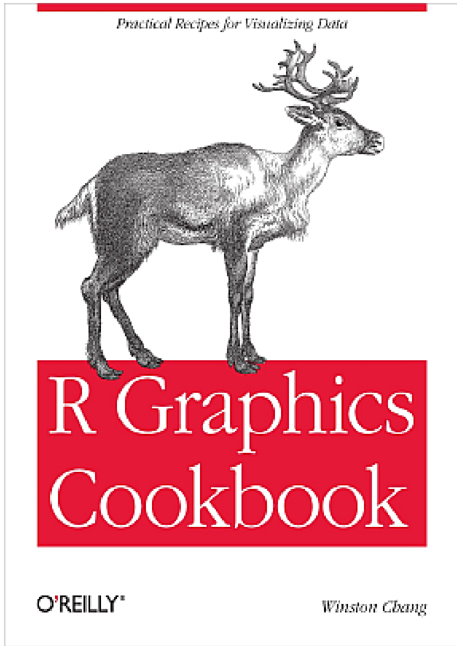

Use R!

Deepayan Sarkar

Lattice

Multivariate Data Visualization with R

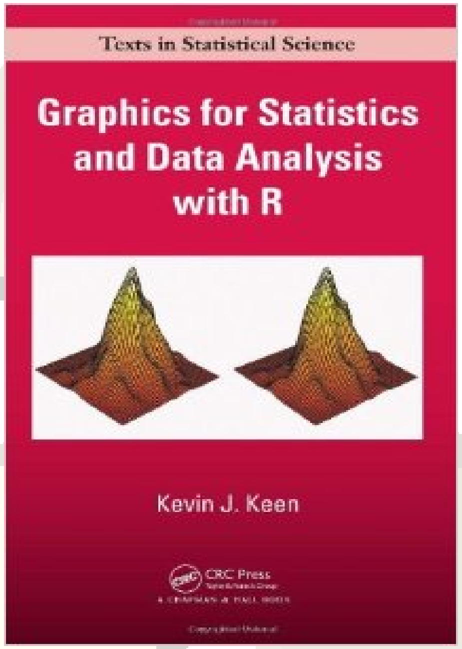
Springer


Practical Recipes for Visualizing Data

R Graphics Cookbook

O'REILLY

Winston Chang

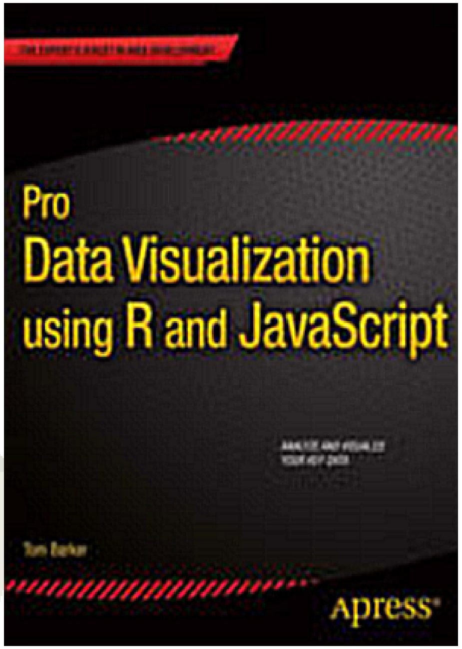

Texts in Statistical Science

Graphics for Statistics and Data Analysis with R

Kevin J. Keen

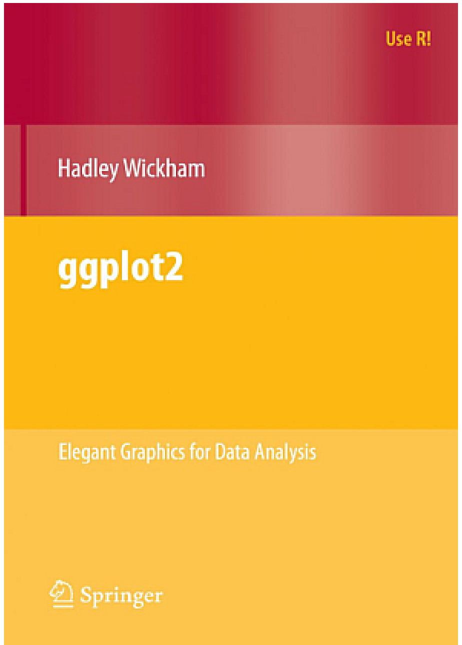
CRC Press

A CHAPMAN & HALL BOOK


Pro Data Visualization using R and JavaScript

Tom Barker

Apress

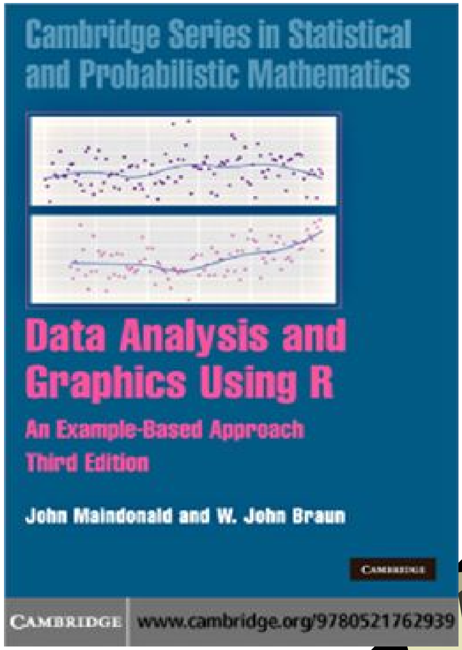

Use R!

Hadley Wickham

ggplot2

Elegant Graphics for Data Analysis

Springer


Cambridge Series in Statistical and Probabilistic Mathematics

Data Analysis and Graphics Using R

An Example-Based Approach

Third Edition

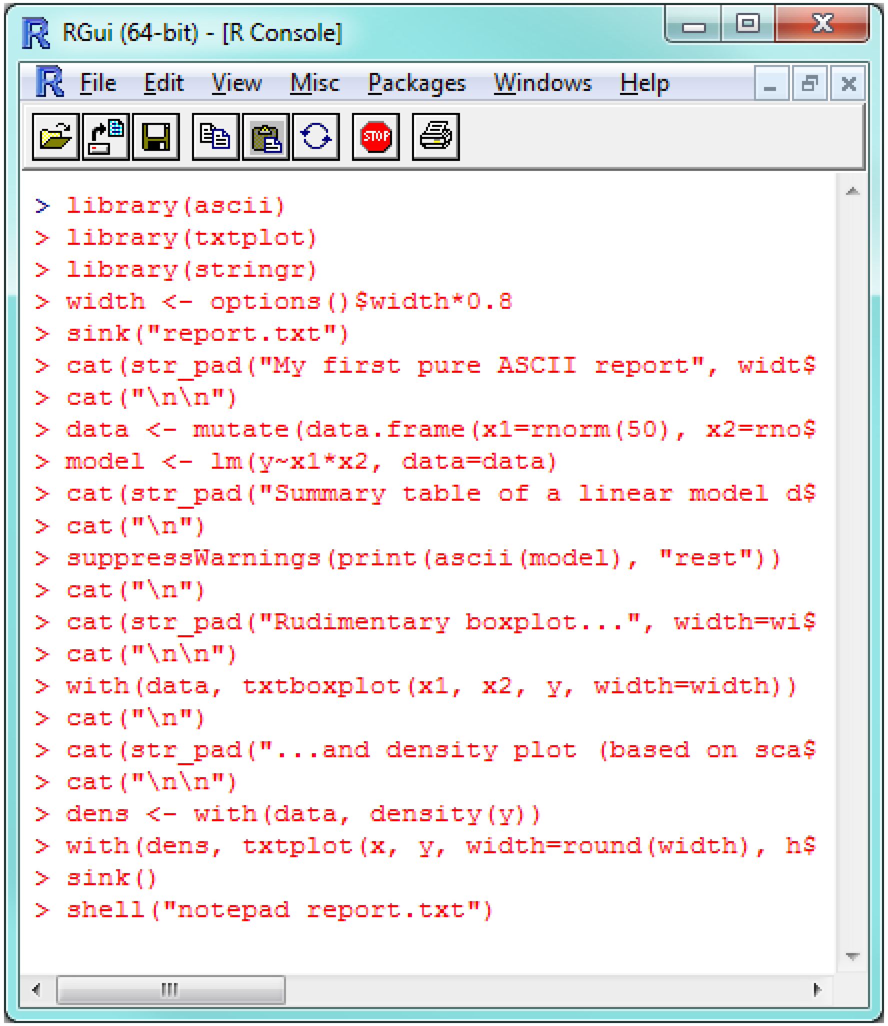
John Maindonald and W. John Braun

CAMBRIDGE

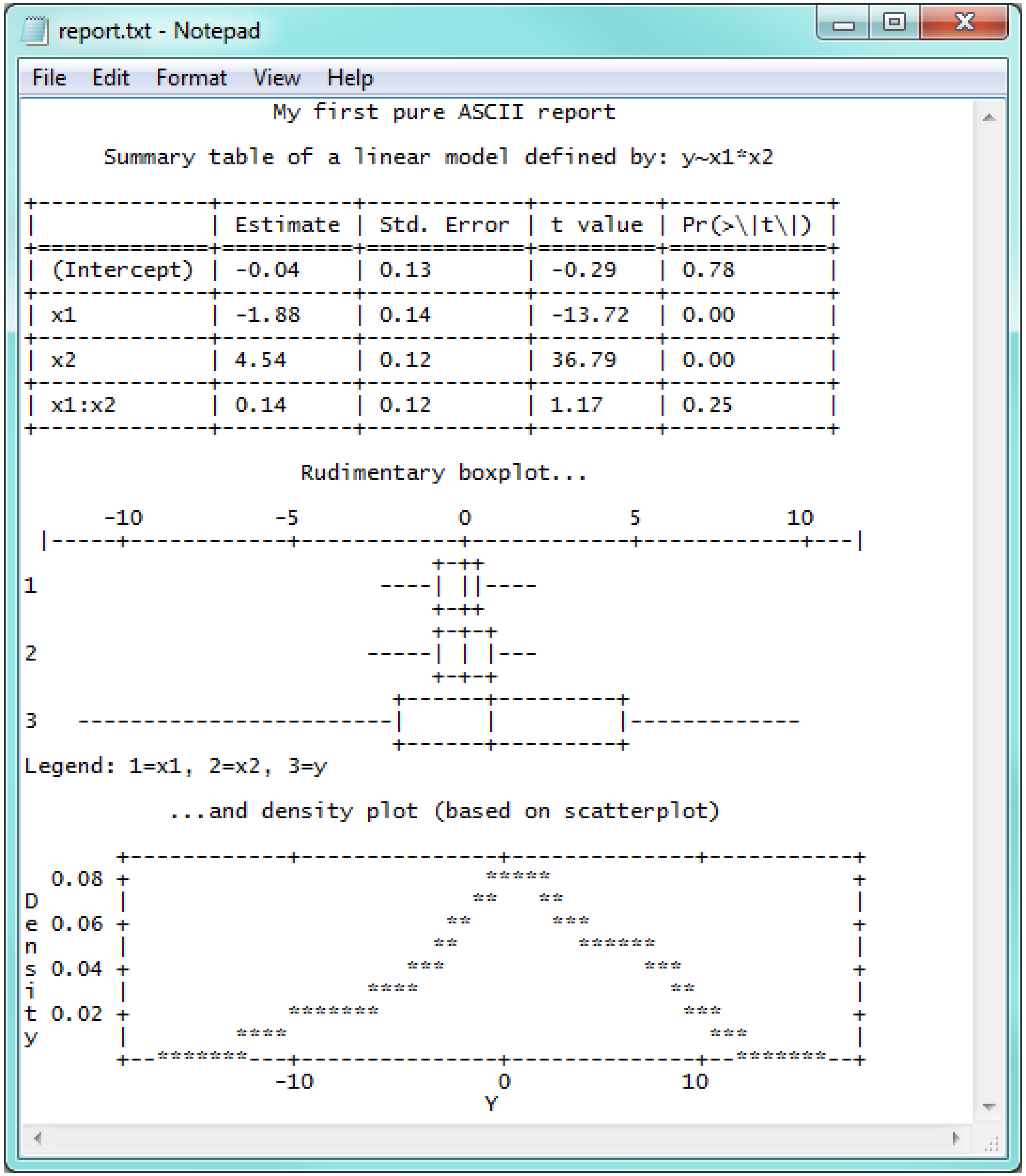
CAMBRIDGE www.cambridge.org/9780521762939

Plain ASCII output

With ascii and txtplot packages one can create text logs (listings) like SAS does.



```
> library(ascii)
> library(txtplot)
> library(stringr)
> width <- options()$width*0.8
> sink("report.txt")
> cat(str_pad("My first pure ASCII report", widt$
> 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")
```



```
                   My first pure ASCII report

      Summary table of a linear model defined by: y~x1*x2

+-------------+----------+------------+---------+------------+
|             | Estimate | Std. Error | t value | Pr(>\|t\|) |
+=============+==========+============+=========+============+
| (Intercept) | -0.04    | 0.13       | -0.29   | 0.78       |
+-------------+----------+------------+---------+------------+
| x1          | -1.88    | 0.14       | -13.72  | 0.00       |
+-------------+----------+------------+---------+------------+
| x2          | 4.54     | 0.12       | 36.79   | 0.00       |
+-------------+----------+------------+---------+------------+
| x1:x2       | 0.14     | 0.12       | 1.17    | 0.25       |
+-------------+----------+------------+---------+------------+

                     Rudimentary boxplot...

      -10          -5            0            5           10
 |-----+------------+------------+------------+------------+---|
                               +-++
1                          ----| ||----
                               +-++
                               +-+-+
2                         -----| | |---
                               +-+-+
                            +------+---------+
3   ------------------------|      |         |-------------
                            +------+---------+
Legend: 1=x1, 2=x2, 3=y

           ...and density plot (based on scatterplot)

       +------------+---------------+--------------+-----------+
  0.08 +                           *****                       +
D      |                          **   **                      |
e 0.06 +                        **      ***                    +
n      |                       **         ******               |
s 0.04 +                     ***               ***             +
i      |                  ****                   **            |
t 0.02 +            *******                       ***          +
y      |        ****                                ***        |
       +--*******---+---------------+--------------+--*******--+
                   -10              0             10
                                   Y
```

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

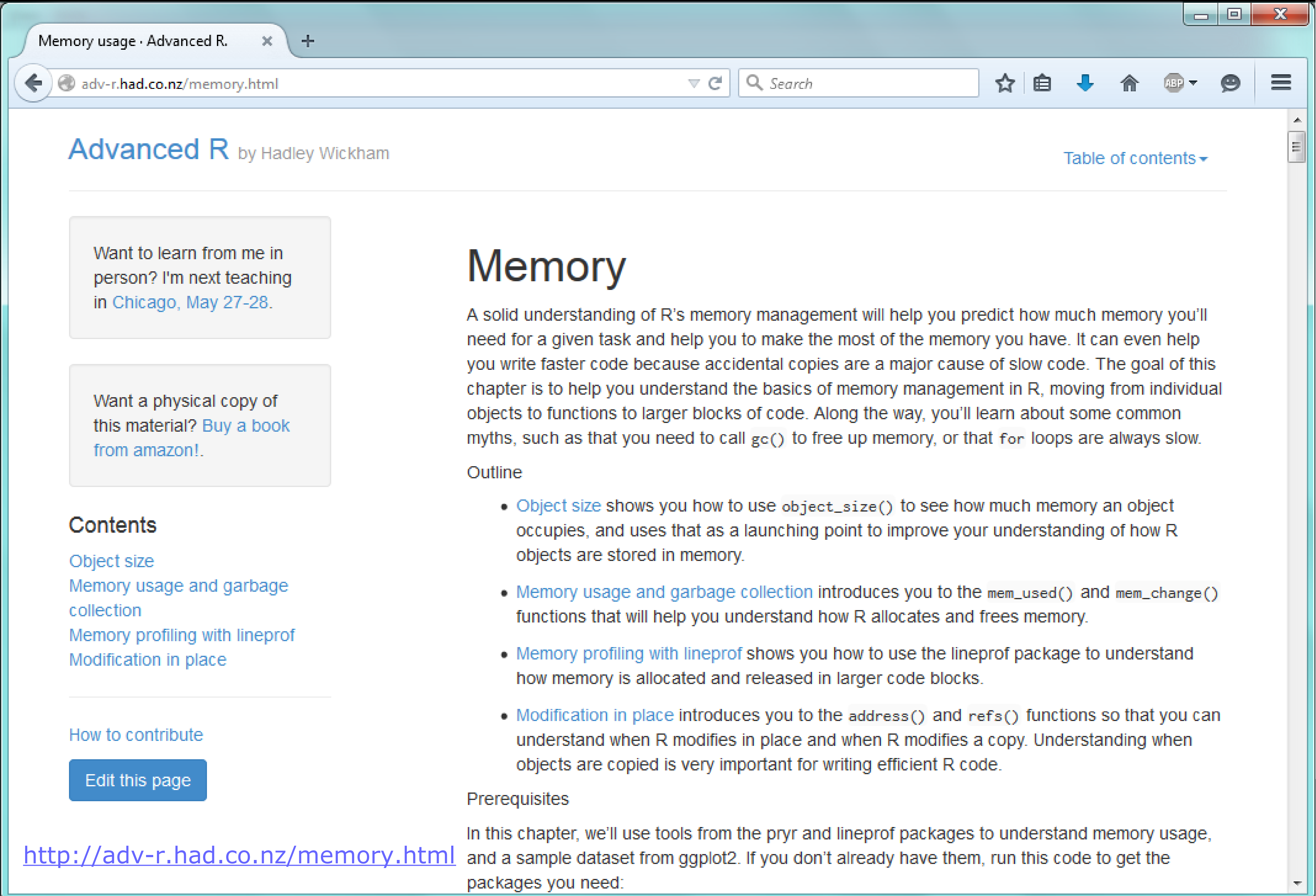
I R is (extremely) cheap. In fact - it's free :)
II R has (extremely) wide range of capabilities
III R is (widely) supported by the world of science
IV R is supported by the community
V R is (increasingly) supported by the business
VI R is able to read data in many formats
VII Interoperability is easy 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!

Tune it up!

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



Advanced R by Hadley Wickham

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

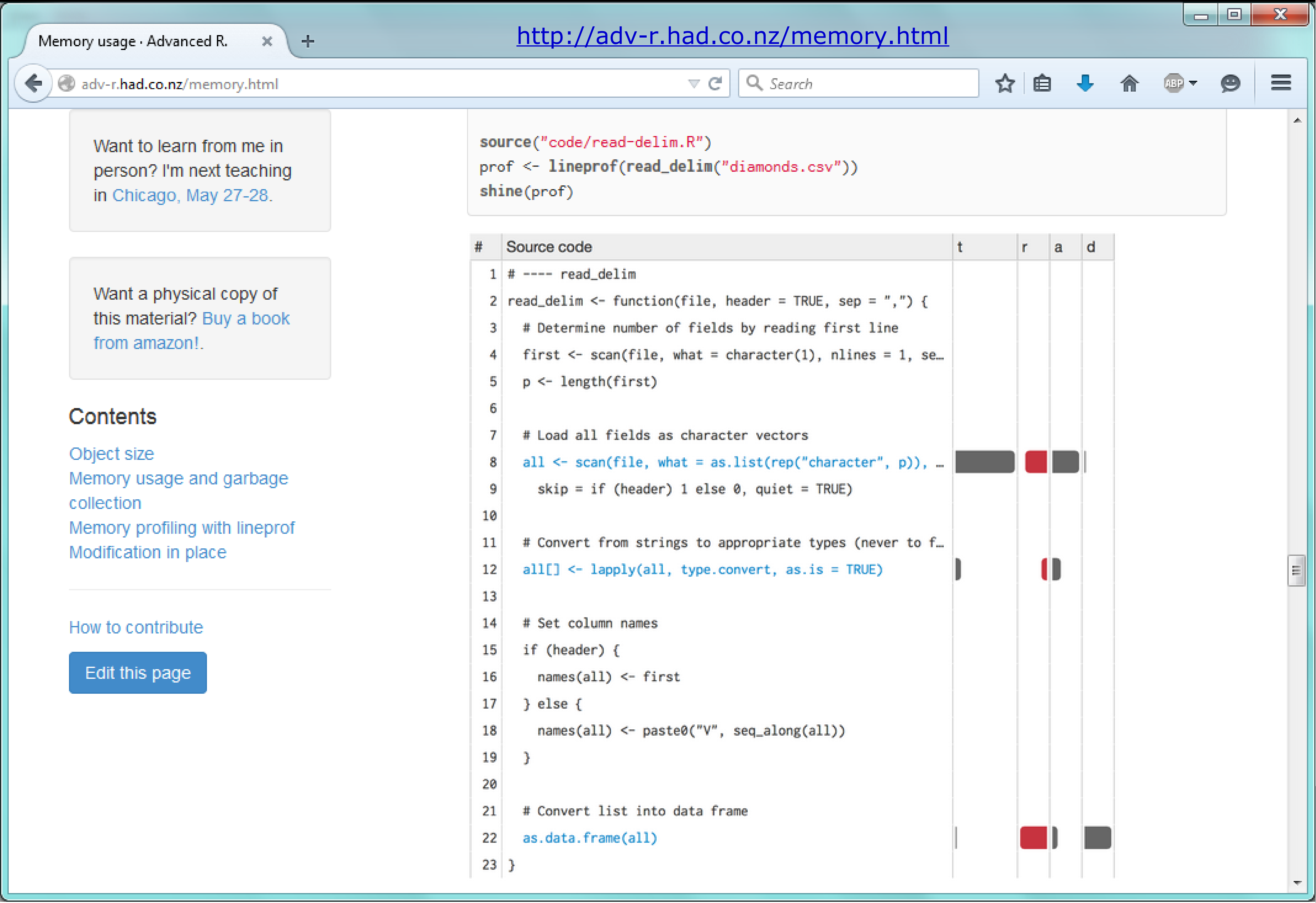
- Object size shows you how to use `object_size()` to see how much memory an object occupies, and uses that as a launching point to improve your understanding of how R objects are stored in memory.
- Memory usage and garbage collection introduces you to the `mem_used()` and `mem_change()` functions that 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



http://adv-r.had.co.nz/memory.html

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

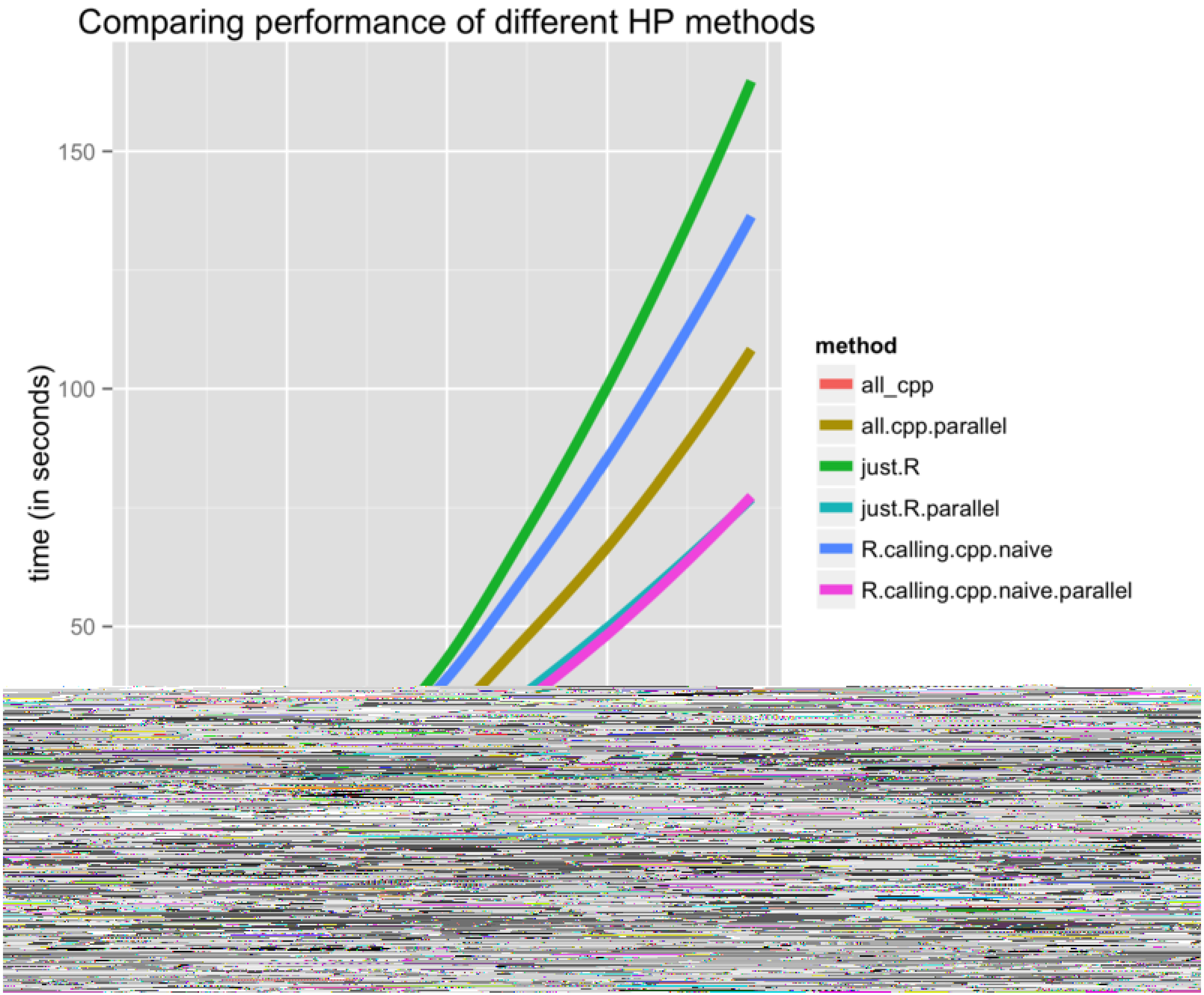
Edit this page

```
source("code/read-delim.R")
prof <- lineprof(read_delim("diamonds.csv"))
shine(prof)
```

#	Source code	t	r	a	d

```
# ---- read_delim
read_delim <- function(file, header = TRUE, sep = ",") {
  # Determine number of fields by reading first line
  first <- scan(file, what = character(1), nlines = 1, se…
  p <- length(first)

  # Load all fields as character vectors
  all <- scan(file, what = as.list(rep("character", p)), …
    skip = if (header) 1 else 0, quiet = TRUE)

  # Convert from strings to appropriate types (never to f…
  all[] <- lapply(all, type.convert, as.is = TRUE)

  # Set column names
  if (header) {
    names(all) <- first
  } else {
    names(all) <- paste0("V", seq_along(all))
  }

  # Convert list into data frame
  as.data.frame(all)
}
```

Tune it up!



http://blog.revolutionanalytics....comparison-of-high-performance-computing-techniques-in-r.html