



R and Revolution R Enterprise in regulatory environments

Dr. Eike Nicklas, HMS Analytical Software
Andrie de Vries, Revolution Analytics

PhUSE 2014
SD01



HMS Analytical Software

Service and solution provider in the field of
data analysis and business intelligence



■ Profile

- 50 employees in Heidelberg and Munich, Germany
- SAS Institute Partner for 15 years
- Doing data oriented software projects for 25 years
- Focus on **life science** and finance industry



■ Technologies

- Analytics and Data Management:
SAS, JMP, **R**, Matlab, Microsoft SQL Server
- Application Development: Microsoft .NET, Java, Open Source

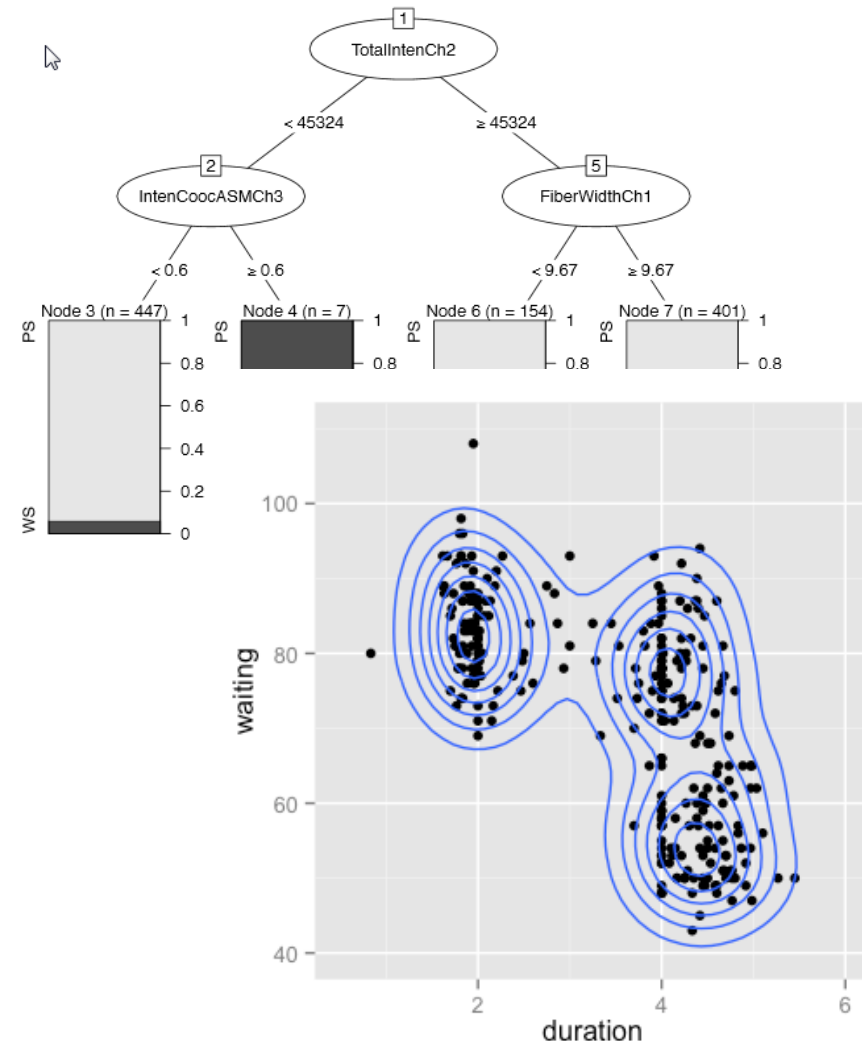


What is R?

“R is a free software environment for statistical computing and graphics” - www.r-project.org

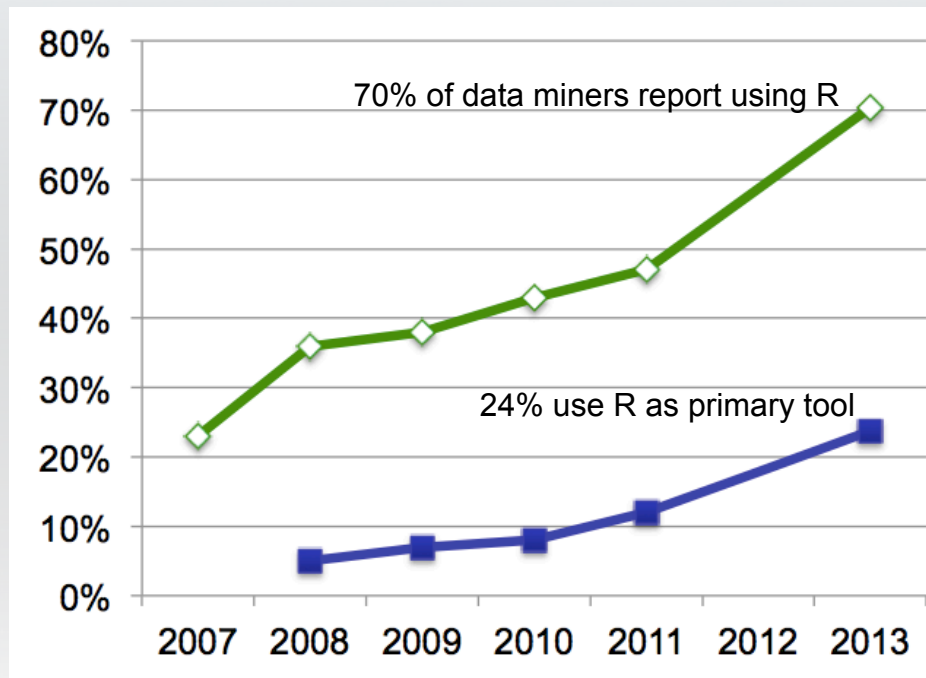


- Open Source (GPL)
- Widely used at universities
 - Talent Pool
 - Latest innovative methods available
- Large Community
- Many freely available extension packages



R usage is growing

Rexer Data Miner Survey, 2007-2013



Source: www.rexeranalytics.com

What is Validation?

Definition by the FDA

Process,
not product

Convince the
FDA

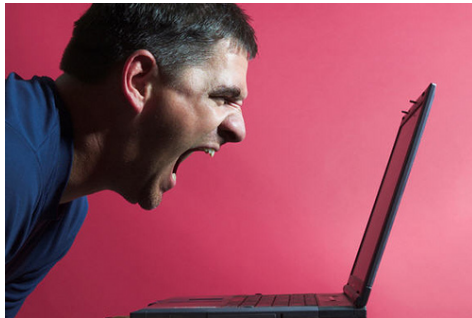
Validation - Establishing documented evidence which provides a high degree of assurance that a specific process will consistently produce a product meeting its predetermined specifications and quality attributes.

Reproducible

Correct
results

R in the regulated environment

Common concerns



„We cannot use R because it is not validated“

„The FDA does not allow the use of R?“

 Cross Validated

QUESTIONS TAGS USERS BADGES UNANSWERED ASK QUESTION

When submitting analyses to FDA how does using R affect issues of software validation?

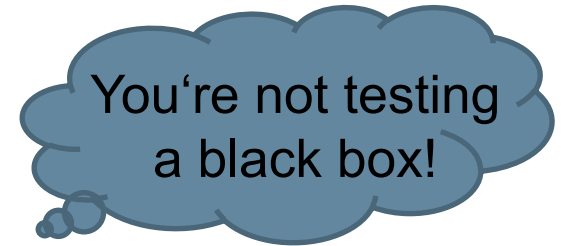
6	I have worked on clinical trials in FDA submissions for many years. I use SAS almost exclusively. Recently I discussed a consulting job I had to bootstrap a Deming regression. Bootstrapping is much more easily done in R I think and several experts have suggested ways that I can do this in R. Since	asked 1 year ago viewed 259 times active 2 months ago
---	--	---

Validation in the context of R

What are the different aspects?



- Validation of software products (R Environment or third party packages)
 - Quality statement
 - IQ / OQ
 - Availability of support and trainings
 - Source code of R and R packages is generally available
- Validation of custom scripts
 - Technical helpers: Integrated development environments, unit testing, ...
 - Methodical helpers: Pair programming, code inspection, independent re-implementation, ...



Using R in Clinical Trial Environments

Quality Statement

**R: Regulatory Compliance and Validation
A Guidance Document for the Use of R in Reg
Trial Environments**

December 28, 2012

The R Foundation for Statistical Computing
c/o Institute for Statistics and Mathematics
Wirtschaftsuniversität Wien
Augasse 2-6
1090 Vienna, Austria
Tel: (+43 1) 31336 4754
Fax: (+43 1) 31336 774
Email: R-foundation@R-project.org

**R: Regulatory Compliance and Validation Issues
A Guidance Document for the Use of R in Regulated Clinical
Trial Environments**

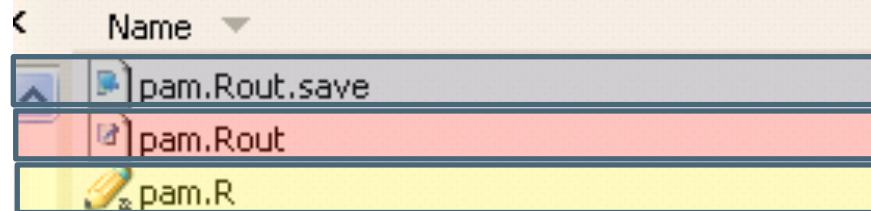
August 8, 2013

The R Foundation for Statistical Computing
c/o Institute for Statistics and Mathematics
Wirtschaftsuniversität Wien
Augasse 2-6
1090 Vienna, Austria
Tel: (+43 1) 31336 4754
Fax: (+43 1) 31336 904754
Email: R-foundation@R-project.org

Installation Qualification of R

Running the test

C:\Programme\R\R-2.10.1\test_testInstalledPackages_base\cluster-tests



2) Reference-File

3) Test-Script Output-File

Called by
`testInstalledPackages("base")`

1) Test-Script

```
.proctime00 <- proc.time()

summary(px2 <- pam(x, 2))
pdx <- pam(dist(x), 2)
all.equal(px2[nms], pdx[nms], tol = 1e-12) ## TRUE
pdxK <- pam(dist(x), 2, keep.diss = TRUE)
stopifnot(identical(pdx[nm2], pdxK[nm2]))
```

Installation Qualification of R

Evaluating the result

2) Reference-File

```
> pdx <- pam(dist(x), 2)
> all.equal(px2[nms], pdx[nms], tol = 1e-12)
[1] TRUE
> pdxK <- pam(dist(x), 2, keep.diss = TRUE)
> stopifnot(identical(pdx[nm2], pdxK[nm2]))
```



3) Test-Script Output-File

```
> pdx <- pam(dist(x), 2)
> all.equal(px2[nms], pdx[nms], tol = 1e-12)
[1] TRUE
> pdxK <- pam(dist(x), 2, keep.diss = TRUE)
> stopifnot(identical(pdx[nm2], pdxK[nm2]))
```

Operational Qualification



- IQ tests do not suffice for operational qualification
- What tests are already implemented for the packages / functions of interest?
 - What is tested?
 - Are tests missing?
 - Do reference results exist? Cross checking?
- If needed, write your own unit tests, e.g. using RUnit

What does the FDA say about R?

Mat Soukup, useR! 2007

Regulations and Guidances

Relying on R for Statistical Analysis

FDA ReviewerS?

The FutuRe

Closing Remarks

■ Regulatory Conclusions

- Regulations do NOT prevent use of R.
- However, validation, documentation, and accountability are required of sponsors relying on R.
- A public listing of validated R functions would
 - be a great benefit to FDA reviewers and sponsoring companies.
 - *exceed* current standards - **R sets the standard!**

■ FDA Reviewer/UseR Perspectives

- The Agency's primary objective is to determine **safety** and **efficacy** which should be independent of the software used to derive the results.
- For data driven conclusions there is a need for accuracy, reliability, and consistency.
- Sponsor's conclusions need to be reproducible - should be independent of software.

What does the FDA say about R?

Jae Brodsky, useR! 2012





Some Challenges of Using R in a Regulatory Environment

Jae Brodsky
Office of Biostatistics, OTS, CDER, FDA

The Big Question

How well can R be used as a statistical tool in a large federal regulatory agency such as the FDA?

Current R Usage at the FDA

- FDA offices use R on a daily basis
 - includes Office of Biostatistics
- FDA scientists have written R packages for other scientists' use (FDA or non-FDA)
- FDA does not endorse or require the use of any specific software for statistical analysis

Regulations and Guidelines for Regulatory Use

- Code of Federal Regulations Title 21: FDA
 - Part 11: Electronic records and electronic signatures
- General Principles of Software Validation: Final Guidance for Industry and FDA Staff (2002)
- Guidance Part 11: Electronic Records (Final) (2005)
- Guidance for Industry: Computerized Systems Used in Clinical Investigations (2007)
- ICH E6: Good Clinical Practice Consolidated Guideline
- ICH E9: Statistical Principles for Clinical Trials

Code of Federal Regulations

21 CFR 11 deals with electronic records and signatures that are "trustworthy, reliable and generally equivalent to paper records."

- Qualification, verification and validation
- R is only part of a larger system that is verified
- R Foundation compliance document (2008): <http://www.r-project.org/doc/R-FDA.pdf>

Regulations and Guidances: What Applies to R at the FDA

Regulations state the specifics of what is legally required and enforceable. Guidances describe the agency's current position and are recommendations. 21 CFR Part 11 regulates electronic records and their storage, not the software that is used to generate reports. There are no regulations that restrict the use of open source software (including R) at the FDA. However, there are concerns for the FDA beyond CFR Title 21.

Current FDA R Setup

Only base and recommended packages are fully supported by the R core. Code or functions outside of these packages are used at the reviewer's own risk.

- Base: *datasets*, *graphics*, *grDevices*, *grid*, *methods*, *splines*, *stats*, *statsf*, *utils*, *utils*
- All results from data should be reproducible independent of software used.
- FDA is not responsible if results have discrepancies.

Sponsors may use R in their submissions. Data must be submitted in *zpt* format. Reviewers would like to know:

- which R functions were used
- how they were accessed
- that the results are accurate, reliable, and consistent

Regulatory Agency Needs

FreeWare is good in theory for a government agency, but comes with several issues:

- No tech/customer support
- How should the FDA provide support to R users?

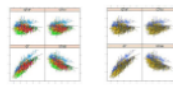
The FDA has used SAS for 30+ years. FDA groups are using SAS and R concurrently. Legacy code is generally dealt with by calling SAS from R.

- Is this an effective or efficient way to handle old SAS code?
- R users must keep up with SAS code changes, and vice versa.

The FDA does not endorse any statistical software. Can we be innovative with R without appearing as though we "endorse" R?

Further Regulatory Issues That Impact R Usage at the FDA

As a government agency, the FDA must comply with additional regulations that do not impact R. Section 508 of the Rehabilitation Act of 1973 (29 U.S.C. 794d) established guidelines for ensuring that the federal government's electronic information is accessible to people with disabilities, including blindness, deafness, colorblindness, epilepsy, and reduced motor function. Section 508 includes guidelines for luminosity or brightness, use of color, font size, text alternatives for figures or plots, and text-only alternatives. All reports released online by the FDA should be 508 compliant, including reports from the Office of Biostatistics.



Example from an FDA report: Left: a plot made with the default rainbow color palette. This figure is not 508 compliant because some information is lost to colorblind people. Right: the red-green colorblind simulation. It is not possible to determine which lines correspond to individual subjects when the color palette is reduced.

My Personal Experience as an R User at the FDA

The QT interval is a biomarker for cardiac repolarization. The QT Interdisciplinary Review Team (IRT) reviews Thorough QT (TQT) studies to investigate QT interval lengthening. The QT IRT consists of statisticians, clinical pharmacologists, and clinicians.

I am the only R user in the stats group (4 statisticians). I use:

- SAS for reviews (for consistency with other QT IRT stats reviewers)
- R for research
- R for any graphics outside reviews

The QT IRT clinical pharmacologists use R for their reviews. The R package *QT* was written specifically for the QT IRT.

- QT calls the stats group's SAS macros for areas in the clin pharm portion of the review
- other areas of the clin pharm review are done in R
- original author of *QT* no longer at the FDA

My future plans: to write an R package for QT IRT use that covers all parts of a TQT review.

Timeline: ?

In my experience, R use at the FDA is acceptable and has not caused any problems.

References

- [1] Title 21 CFR Part 11 Section 11.1 (a)
- [2] General Principles of Software Validation: Final Guidance for Industry and FDA Staff
- [3] Guidance Part 11: Electronic Records (Final)
- [4] Guidance for Industry: Computerized Systems Used in Clinical Investigations
- [5] ICH E6(R1): Good Clinical Practices
- [6] ICH E9: Statistical Principles for Clinical Trials
- [7] R: Regulatory Compliance and Validation Issues: A Guidance Document for the Use of R in Regulated Clinical Trial Environments
- [8] Section 508 of the Rehabilitation Act of 1973 (29 U.S.C. 794d)

„R use at the FDA is completely acceptable and has not caused any problems.“

Example Projects by HMS

What have we done?



- Near Munich
 - Creation of a validated R installation version for use in pharmaceutical research
- Near Heidelberg
 - Validation of custom R scripts for transferring clinical trial data from one system to another
- Even closer to Heidelberg
 - Validation of R scripts for generating and visualizing statistics regarding the handling of customer complaints of medical devices

Further Resources

Where to get more information



- [Peter Bewerunge's paper of the PhUSE 2010](#)
- Revolution Analytics Blog
 - [Using R in the pharmaceutical industry](#)
 - [Analyzing clinical trial data for FDA submissions with R](#)
 - [Clinical reporting with R](#)
 - [FDA: R OK for drug trials](#)
- [Mat Soukup's talk at UseR 2007](#)
- [Jae Brodsky's poster at UseR 2012](#)
- [Ian Francis' talk 2010](#)

Vielen Dank für Ihre Aufmerksamkeit!



Dr. Eike Nicklas

Software-Ingenieur

eike.nicklas@analytical-software.de

HMS Analytical Software GmbH
Rohrbacher Str. 26
69115 Heidelberg
www.analytical-software.de
info@analytical-software.de
+49-6221-6051-0

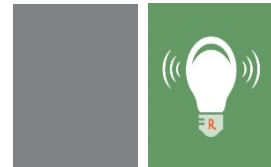


Validating R

Andrie de Vries

andrie@revolutionanalytics.com

@RevoAndrie





OUR COMPANY



MOUNTAIN VIEW ■ LONDON ■ SINGAPORE

The leading provider
of **advanced analytics
software and services**
based on open source R,
since 2007

OUR PRODUCT



REVOLUTION R: The
enterprise-grade predictive
analytics application platform
based on the R language

SOME KUDOS



Visionary
Gartner Magic Quadrant
for Advanced Analytics
Platforms, 2014

Gartner



Post-installation validation of R



Running the post install test protocol

```
library(RevobaseEnt)  
RevoIOQ()
```

RUNIT TEST PROTOCOL-- x

file:///C:/Users/Andrie%20de%20Vries/Documents/checkpoint-demo/RevoIOQ/Revo-IOQ-Report-10-08-2014-15

RUNIT TEST PROTOCOL--Wed Oct 08 15:22:33 2014

Number of test functions: 1596
Number of deactivated test functions: 47
Number of errors: 0
Number of failures: 1

1 Test suite

Name	Test functions	Deactivated	Errors	Failures
Revolution R Enterprise version 7.3.0 (2014-09-18 01:47:43 UTC)	1596	47	0	1

Failures

Test suite : test function	message
Revolution R Enterprise version 7.3.0 (2014-09-18 01:47:43 UTC): test rxMkl version same version	Error in checkEquals(rxMklVersion, revoMklVersion, errmsg) : 1 string mismatch : rxMklVersion 'Intel(R) Math Kernel Library Version 11.1.1 Product Build 20131010 for Intel(R) 64 architecture applications' is not equal to revoMklVersion 'Intel(R) Math Kernel Library Version 11.1.1 Product Build 20131010 for Intel(R) 64 architecture applications βA'



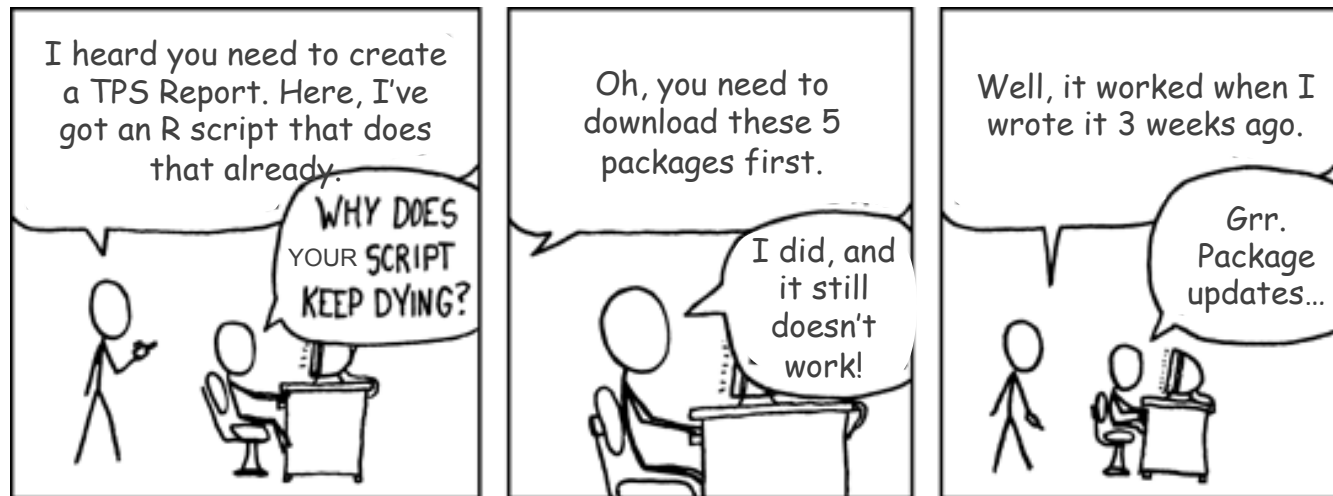
Demonstration



Package reproducibility in R



The problem of reproducibility



With apologies to xkcd (<http://xkcd.com/234/>)



Sharing a script reproducibly ... and simply

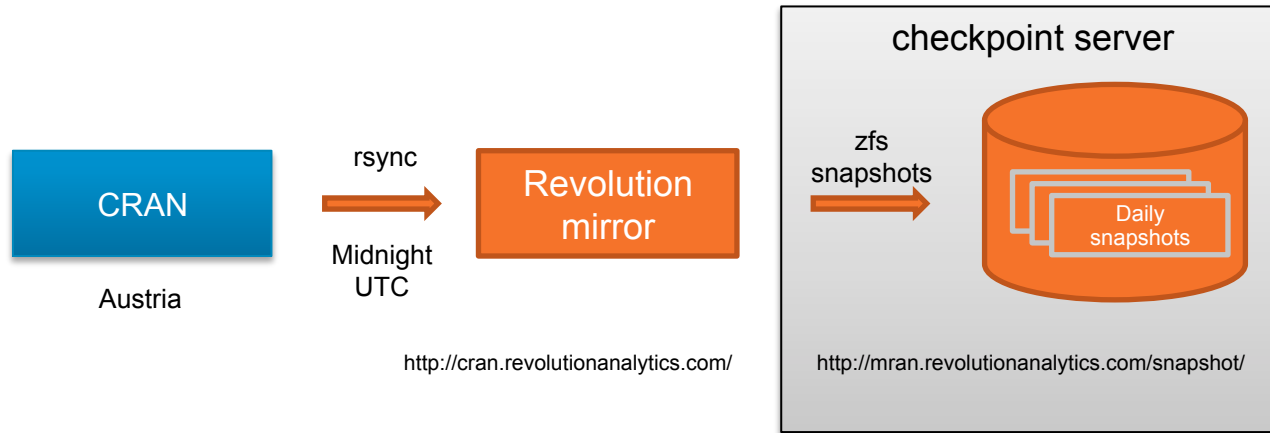
```
# Run with R 3.1.0
require(checkpoint)
checkpoint("2014-09-17")

# find packages used in this project
# install packages in checkpoint folder
# set library path to use checkpointed packages

require(ggplot2)
require(data.table)
require(knitr)
...
```



The Reproducible R Toolkit – checkpoint server





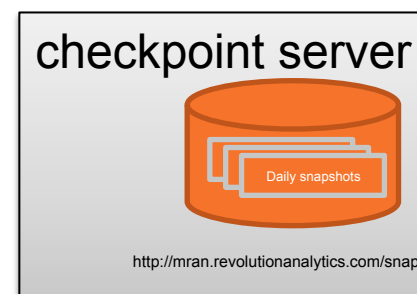
The Reproducible R Toolkit – checkpoint package



Austria



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



```
require(checkpoint)  
checkpoint("2014-09-17")  
install.packages()
```



Demonstration

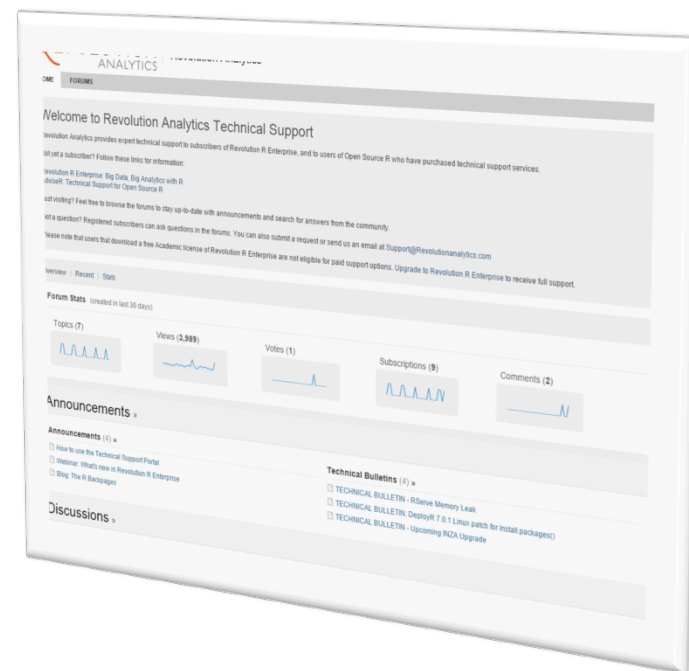


Technical support



AdviseR™ Technical Support

- Technical support for R, from the R experts.
- 10x5 email and phone support (in user's local time zone)
- Support for R, validated packages, and third-party software connections
- On-line case management and knowledgebase
- Access to technical resources, documentation and user forums
- Exclusive on-line webinars from community experts
- Guaranteed response times



Thank you.

www.revolutionanalytics.com

Twitter: @RevolutionR

