

Statistische Analysen mit R

Liars, Outliers and Statistics

Bruno Hopp, zum SOFTWARE FREEDOM DAY 2012

15. Sept. 2012

erst mal Generalfrage “wat soll dä Quatsch?”

“ R is a language and environment for statistical computing
and graphics” (www.r-project.org/about.html)

erst mal Generalfrage “wat soll dä Quatsch?”

“ R is a language and environment for statistical computing and graphics” (www.r-project.org/about.html)

”is an integrated suite of software facilities for data manipulation, calculation and graphical display.” (dto).

erst mal Generalfrage “wat soll dä Quatsch?”

“ R is a language and environment for statistical computing and graphics” (www.r-project.org/about.html)

”is an integrated suite of software facilities for data manipulation, calculation and graphical display.” (dto).

Lizenz

Sowohl GNU/GPL 2 als auch GNU/GPL 3

erst mal Generalfrage “wat soll dä Quatsch?”

“ R is a language and environment for statistical computing and graphics” (www.r-project.org/about.html)

”is an integrated suite of software facilities for data manipulation, calculation and graphical display.” (dto).

Lizenz

Sowohl GNU/GPL 2 als auch GNU/GPL 3

Homepage: www.r-project.org/

weitere Eigenschaften ...

modular, mehr als 4000 Module via **CRAN**,
mehr als 554 in BioConductor (Aug.2012).

Alle relevanten Architekturen: Linux i386+AMD64+ARM,
Microsoft Windows, SPARC, BSD unices: OpenBSD, FreeBSD
etc.

FREIER SOURCE CODE! Kann mit den meisten RDBMS
(MySQL/PostgreSQL, ORACLE...) kooperieren. I/O in
Fremdformaten kein Problem.

In the beginning . . .

Geschichte

“R was created by Ross IHAKA and Robert GENTLEMAN at the University of Auckland, New Zealand...”

[http://en.wikipedia.org/wiki/R_\(programming_language\)](http://en.wikipedia.org/wiki/R_(programming_language))
Start 1992, seit 1993 öffentlich, seit 1995 unter GNU/GPL.

In the beginning . . .

Geschichte

“R was created by Ross IHAKA and Robert GENTLEMAN at the University of Auckland, New Zealand...”

[http://en.wikipedia.org/wiki/R_\(programming_language\)](http://en.wikipedia.org/wiki/R_(programming_language))
Start 1992, seit 1993 öffentlich, seit 1995 unter GNU/GPL.

Releases

Zweimal im Jahr: Ende April und Ende Oktober. Zur Zeit gilt: Version 2.15.1, aka ROASTED MARSHMALLOWS, zwei bis max. drei Sub-Releases (2.15.1 oder 2.15.2).

Mirrors weltweit: **CRAN**

Das “Comprehensive R Archive Network” dient schneller Verbreitung und leichtem Zugriff:

<http://cran.r-project.org/mirrors.html>

CRAN dient für alle Zusatzpakete in **R**.

Daneben gibt es spezialisierte Sammlungen von Zusatzpaketen, etwa BIOCONDUCTOR für die Biologischen Wissenschaften.

CRAN und VIEWS

Mirrors weltweit: **CRAN**

Das “Comprehensive R Archive Network” dient schneller Verbreitung und leichtem Zugriff:

<http://cran.r-project.org/mirrors.html>

CRAN dient für alle Zusatzpakete in **R**.

Daneben gibt es spezialisierte Sammlungen von Zusatzpaketen, etwa BIOCONDUCTOR für die Biologischen Wissenschaften.

TaskView

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

bei der Bewältigung aktueller Aufgaben: z.B. Lineare Modellierung oder Bayesian Statistik verlangen *mehrere* Pakete. Um hier schnell ans gewünschte Ziel zu kommen, installiere die jeweiligen **TaskViews**.

Haben will. . .

Installation

gehe zu Sektion "GNU R" in
<http://packages.debian.org/stable> und nutze eine leicht
veraltete Version (meist vorletzte).

Haben will...

Installation

gehe zu Sektion "GNU R" in
<http://packages.debian.org/stable> und nutze eine leicht
veraltete Version (meist vorletzte).

Aktuelles...

<http://cran.at.r-project.org/bin/linux/debian/>
wenn du in Debian das *aktuellste* Release willst.
Fedora/RH Nutzer gehen bitte zu [.../bin/linux/redhat/](http://cran.at.r-project.org/bin/linux/redhat/).
Sourcen unter cran.at.r-project.org/sources.html.

Haben will...

Installation

gehe zu Sektion "GNU R" in <http://packages.debian.org/stable> und nutze eine leicht veraltete Version (meist vorletzte).

Aktuelles...

<http://cran.at.r-project.org/bin/linux/debian/>
wenn du in Debian das *aktuellste* Release willst.
Fedora/RH Nutzer gehen bitte zu [.../bin/linux/redhat/](http://cran.at.r-project.org/bin/linux/redhat/).
Sourcen unter cran.at.r-project.org/sources.html.

Zusätzlich

Pakete **r-base**, **r-recommended**, **lmtest** (Lin.Modeling+Testing) und **Rcommander** einspielen. Dann ein passendes **TaskView** –Namen beachten!

Erster Start

aus der Shell (Bash) heraus mit: “R” (case- sensitiv!). Leichte Frustration. . .

R -g TK bietet nur etwas mehr an Luxus.

Erster Start

aus der Shell (Bash) heraus mit: “R” (case- sensitiv!). Leichte Frustration. . .

R -g TK bietet nur etwas mehr an Luxus.

tippe in der **R** Umgebung bitte bei installiertem
“R-commander”: `library(Rcmdr)`

es startet nach einem Moment die grafische Oberfläche des **R**
Commander. Es gibt weitere GUIs!! (mindestens zehn, eher
mehr)

mehr Eigenschaften von R

R ist

grundsätzlich SMP-/Clusterfähig, die Skripte bestehen aus Klartext, Skripte besitzen ein *.R als Endung. Skripte sind systemunabhängig! Grafiken sind exportierbar z.B. für eine Publikation. Ein `#` am Zeilenanfang im Skript macht aus der ganzen Zeile einen Kommentar.

was ist mit BIG data?

sehr große Datensätze wie über Google API, Messreihen mit mehr als tausend oder zehntausend Datenpunkten, komplexe Datensätze aus Datenbanken... lassen sich hervorragend mit **R** analysieren. Dank Import-/Exportfilter sind Fremdformate kein Problem (CSV, SPSS, STATA, SAS, Excel, etc): (Paket **foreign**)

Webseite

Statistische Analysen mit R

Bruno Hopp,
zum
SOFTWARE
FREEDOM
DAY 2012

The R Project for Statistical Computing

PCA 5 variables
generated by plot.prcomp

Clustering 4 groups

Density

Factor 1 [41%]

Factor 3 [19%]

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 version 2.15.1** (Roasted Marshmallows) has been released on 2012-06-22.
- [The R Journal Vol. 4/1](#) is available.
- [useR! 2012](#), took place at Vanderbilt University, Nashville Tennessee, USA, June 12-15, 2012.
- [useR! 2013](#), will take place at the University of Castilla-La Mancha, Albacete, Spain, July 10-12 2013. .

This server is hosted by the [Institute for Statistics and Mathematics](#) of the [WU Wien](#).

welche (Analyse-) Methoden??

Regression, Faktoranalyse, Zeitreihen, Bayesian Statistik, GLM, SEM und zahlreiche andere werden von **R** unterstützt. Missing Values (NA) werden gesondert behandelt werden soweit gewünscht.

Grafiken + Visualisierungen

Statistische
Analysen mit
R

Bruno Hopp,
zum
SOFTWARE
FREEDOM
DAY 2012

sind eine der enormen Stärken von **R**.

Grafiken + Visualisierungen

Statistische
Analysen mit
R

Bruno Hopp,
zum
SOFTWARE
FREEDOM
DAY 2012

sind eine der enormen Stärken von **R**.

Scatterplots, Kreisdiagramme, Balken, Linien, Heatmaps,
Lattices etc.

Grafiken + Visualisierungen

sind eine der enormen Stärken von **R**.

Scatterplots, Kreisdiagramme, Balken, Linien, Heatmaps, Lattices etc.

sind vorsichtig zu nutzen! Gefahr von Fehlinterpretationen. . . Sehgewohnheiten beachten!

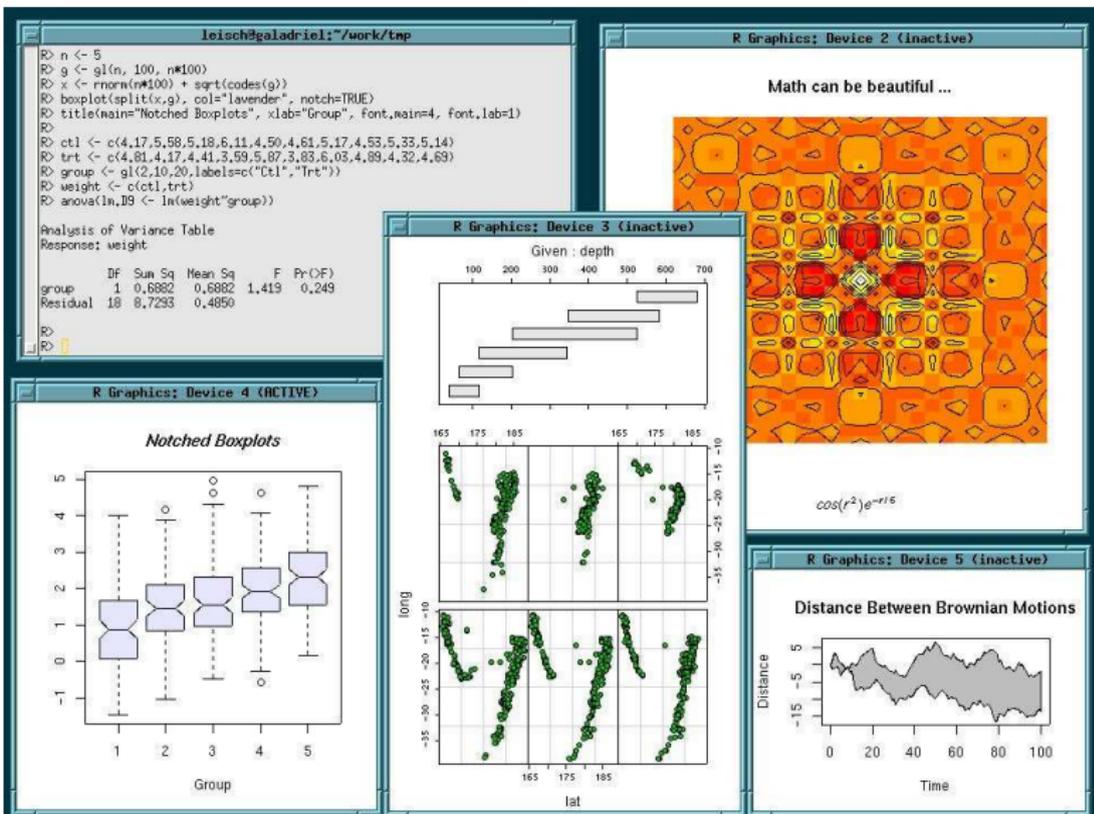
Datenbanken, Tabellen: **foreign**

werden voll unterstützt. MySQL und PostGreSQL dürften die bekanntesten OS-Lösungen sein. **R** kann mit proprietären RDBMS (Oracle) zusammenarbeiten (R/W).
Import/Export nach MS Excel funktioniert.
Import/Export nach OpenOffice/LibreOffice/Gnumeric funktioniert.
Import/Export als CSV/clear text funktioniert.

Arbeitsumgebung

Statistische Analysen mit R

Bruno Hopp,
zum
SOFTWARE FREEDOM
DAY 2012



Lesen

- Joseph ADLER (2010): R in a nutshell. Köln: O'Reilly.
- W.M. VENABLES/D. SMITH/R Core Project(2012): An Introduction to R.
(cran.at.r-project.org/doc/manuals/R-intro.pdf),
22.Juni 2012.
- <http://www.r-project.org/doc/bib/R-books.html>

Lesen

- Joseph ADLER (2010): R in a nutshell. Köln: O'Reilly.
- W.M. VENABLES/D. SMITH/R Core Project(2012): An Introduction to R.
(cran.at.r-project.org/doc/manuals/R-intro.pdf),
22.Juni 2012.
- <http://www.r-project.org/doc/bib/R-books.html>

in Köln existiert schon eine **R** Usergroup!

das Ende naht ...

Statistische
Analysen mit
R

Bruno Hopp,
zum
SOFTWARE
FREEDOM
DAY 2012

Fragen, Kommentare?
Danke!