



ein Statistiksystem für alle!

- zumindest für alle, die empirisch arbeiten wollen –

Erste Schritte in Statistik und Computeralgebra
auf/mit Debian GNU/Linux

Bruno Hopp, SFD am 21.Sept. 2013 in Köln

Was ist



?

laut „R Development Core Team“ (2004) bietet **R** eine leistungsfähige (*performante*) Umgebung für die Durchführung von statistischen Analysen aller Art, von Computeralgebra und ebenso regelmäßig gewünschten Grafiken bzw. Visualisierungen:

„R is a language and environment for statistical computing and graphics.“

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

license()

This software is distributed under the terms of the GNU General Public License, either Version 2, June 1991 or Version 3, June 2007.

The terms of version 2 of the license are in a file called COPYING which you should have received with this software and which can be displayed by RShowDoc("COPYING").

Copies of both versions 2 and 3 of the license can be found at <http://www.R-project.org/Licenses/>.

A small number of files (the API header files listed in R_DOC_DIR/COPYRIGHTS) are distributed under the Lesser GNU General Public License version 2.1.

This can be displayed by RShowDoc("COPYING.LIB"), or obtained at the URI given.

WAS HEISST DAS FÜR UNS ANWENDER?

was ist R ... (Forts.)

- **Implementierung des S4-Standards
(kommerz.Software „S-Plus“ von www.insightful.com)**
- **Modular aufgebaut,**
- **ursprünglich auch ein „soziales Experiment“ bzgl. Open Source
Entwicklung**
- **richtet sich an alle Nutzer, die große Mengen an numerischem
Material auswerten wollen/müssen.**
- **eignet sich für sog. „High Performance Computing“ (=HPC) dank
Parallelisierung und Clusterfähigkeit.**

Geschichte, Plattformen, releases

“**R** was created by Ross Ihaka and Robert Gentleman^[4] at the University of Auckland, New Zealand[...] “

nach Wikipedia: gestartet 1992, seit 1993 öffentlich, seit 1995 unter GNU/GPL lizenziert.

- **R** ist verfügbar im Sourcecode sowie binär für Debian GNU/Linux, für RedHat/Fedora Linux, OpenSuSE, UBUNTU Linux, MacOS und für Microsoft Windows Systeme.
- Generell ist **R** je in 32-bit und 64-bit Variante je Plattform verfügbar – soweit technisch möglich.
- **R** ist in Debian (GNU/Linux) seit 1997 enthalten, damals in release 0.61.
- Nicht separat erwähnt aber leicht verfügbar ist **R** in OpenBSD (Version 5.3 oder höher) über die Ports collection (R version 2.15.2) und in FreeBSD (release 8.x und höher).
- *Zweimal pro Jahr* wird veröffentlicht, jeweils Ende April und Ende Oktober.
- Anfang Sept. 2013: **3.0.1** „Good Sport“ ; angekündigtes Release **3.0.2** („Frisbee Sailing“) ca zum 15.Sept. 2013.



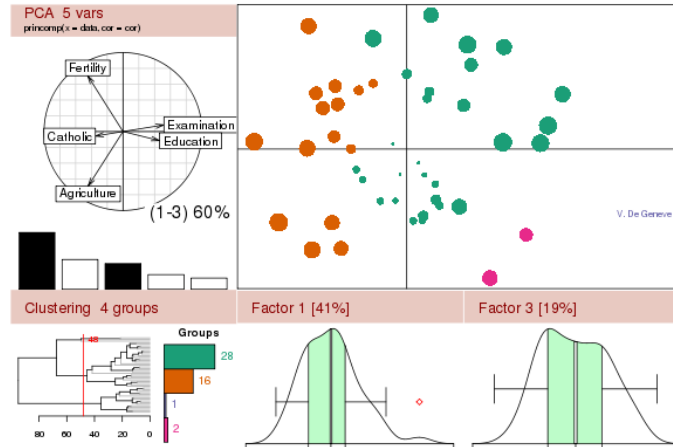
About R
[What is R?](#)
[Contributors](#)
[Screenshots](#)
[What's new?](#)

Download, Packages
[CRAN](#)

R Project
[Foundation](#)
[Members & Donors](#)
[Mailing Lists](#)
[Bug Tracking](#)
[Developer Page](#)
[Conferences](#)
[Search!](#)

Documentation
[Manuals](#)
[FAQs](#)
[The R Journal](#)
[Wiki](#)
[Books](#)
[Certification](#)
[Other](#)

Misc
[Bioconductor](#)
[Related Projects](#)
[User Groups](#)
[Links](#)



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 email.

News:

- [R 3.0.2 \(Frisbee Sailing\) prerelease versions](#) will appear starting September 15. Final release is scheduled for September 25, 2013. Thanks to Erin Hodgens for the idea leading to the name.
- [The R Journal Vol.5/1](#) is available.
- [R version 3.0.1](#) (Good Sport) has been released on 2013-05-16.
- [R version 2.15.3](#) (Security Blanket) has been released on 2013-03-01.
- [useR! 2013](#), will take place at the University of Castilla-La Mancha, Albacete, Spain, July 10-12 2013. .

CRAN mirrors weltweit

- Wichtigste Sammlung von **R**-weiterungen findet sich im **CRAN**, einem Netz von ftp- und Webservern (**C**omprehensive **R** **A**rchive **N**etwork):

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

- CRAN master node an der TU Wien <http://cran.at.r-project.org/>

- **4833 Pakete (Sept.2013)** source code plus: Debian Linux, Ubuntu, Fedora Linux, MacOS, win32.... ist sowas nicht redundant?

- Neben dem allgemeinen CRAN-Netzwerk gibt es spezialisierte Repositories, z.B. für die Biowissenschaften (gestartet 2001, ca 675 Pakete im Sept.2013):

<http://www.bioconductor.org/>

Quelle: <http://www.bioconductor.org/packages/release/bioc/>

R Installation für/in Debian Linux (+ Ubuntu)

- Sektion „GNU R“ unter <http://packages.debian.org/stable/> würde ausreichen
- Wir hätten aber gerne das noch aktuellere release 3.0.1:
<http://cran.at.r-project.org/bin/linux/debian>
- **deb** <http://<favorite-cran-mirror>/bin/linux/debian> **wheezy-cran3/**
und ersetze *favorite-cran-mirror* durch einen Mirror deiner Präferenz! Vorschlag etwa:
- **deb** <http://ftp5.gwdg.de/pub/misc/cran/bin/linux/debian> **wheezy-cran3/**
- Ubuntu: <http://cran.at.r-project.org/bin/linux/ubuntu/> **raring/**

R: Debian Linux Installation (Forts.)

Falls apt bzw. synaptic noch Kommentare auswerfen, liegt es i.d.R. am fehlenden *Key*, der das repository identifiziert (s.o.). Dazu in Shell bitte eingeben (wie [hier](#) erklärt):

```
gpg --keyserver subkeys.pgp.net --recv-key 381BA480  
gpg -a --export 381BA480 > jranke_cran.asc  
apt-key add jranke_cran.asc
```

Der key ist signiert von <jranke@uni-bremen.de> Johannes Ranke.

R unter Debian Linux starten

ein installiertes R unter Linux/Unix starten von Konsole starten mit:

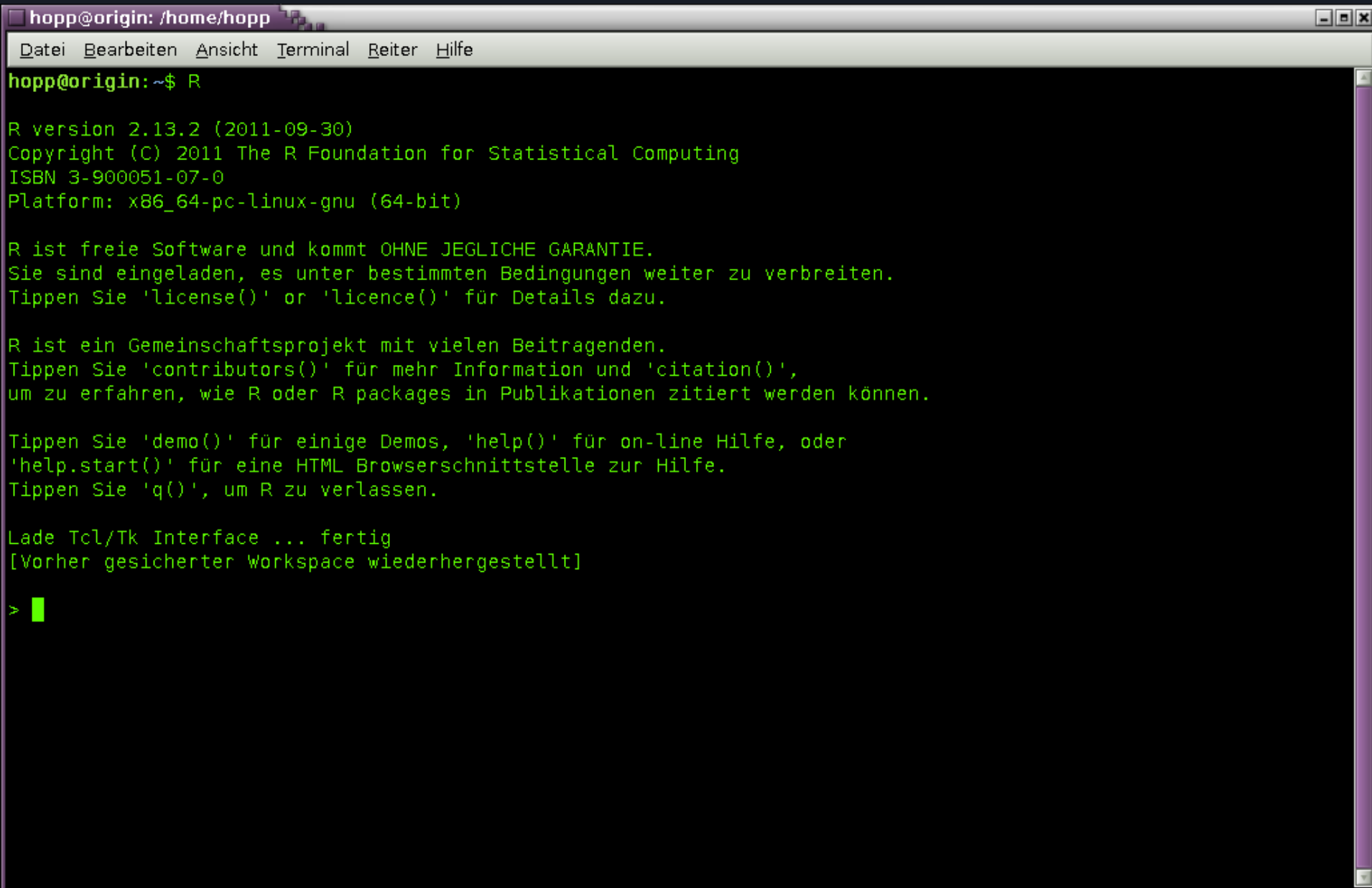
\$R (nächste Folie)

bitte beachten – case-sensitiv ($R \neq r$)

Um ein R interface vergleichbar bequem wie unter MS Windows inkl. einer Menüleiste zu erhalten, starten wir R unter Debian in der Konsole mit dem Tk/Gtk interface:

\$ R -g Tk & (auf übernächster Folie)

Benutzeroberfläche: Debian CLI mit „R“



```
hopp@origin: /home/hopp
Datei Bearbeiten Ansicht Terminal Reiter Hilfe
hopp@origin:~$ R
R version 2.13.2 (2011-09-30)
Copyright (C) 2011 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: x86_64-pc-linux-gnu (64-bit)

R ist freie Software und kommt OHNE JEGLICHE GARANTIE.
Sie sind eingeladen, es unter bestimmten Bedingungen weiter zu verbreiten.
Tippen Sie 'license()' or 'licence()' für Details dazu.

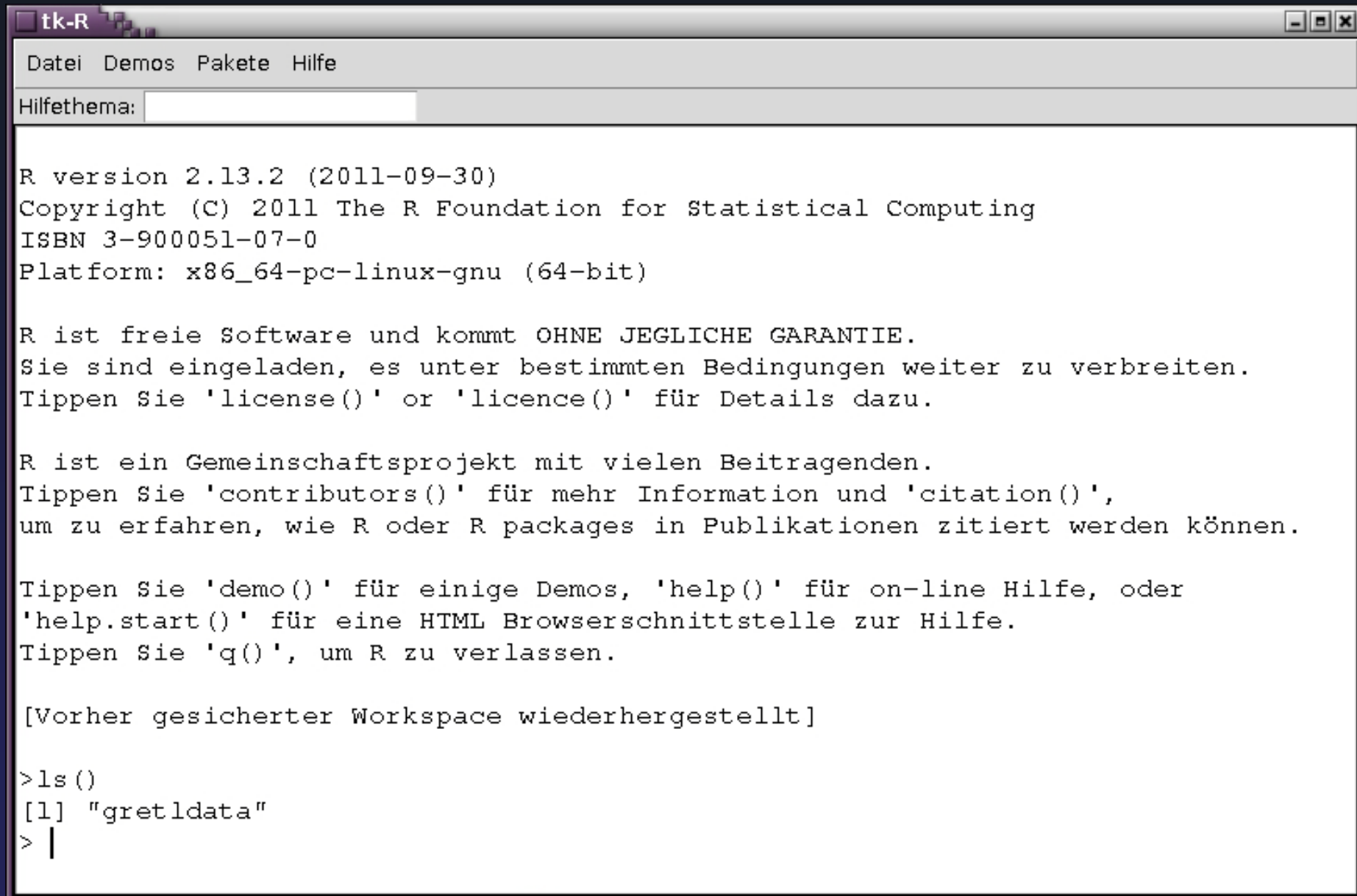
R ist ein Gemeinschaftsprojekt mit vielen Beitragenden.
Tippen Sie 'contributors()' für mehr Information und 'citation()',
um zu erfahren, wie R oder R packages in Publikationen zitiert werden können.

Tippen Sie 'demo()' für einige Demos, 'help()' für on-line Hilfe, oder
'help.start()' für eine HTML Browserschnittstelle zur Hilfe.
Tippen Sie 'q()', um R zu verlassen.

Lade Tcl/Tk Interface ... fertig
[Vorher gesicherter Workspace wiederhergestellt]

> █
```

Benutzeroberfläche Debian: „R –g Tk“



```
tk-R
Datei Demos Pakete Hilfe
Hilfethema:
R version 2.13.2 (2011-09-30)
Copyright (C) 2011 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: x86_64-pc-linux-gnu (64-bit)

R ist freie Software und kommt OHNE JEGLICHE GARANTIE.
Sie sind eingeladen, es unter bestimmten Bedingungen weiter zu verbreiten.
Tippen Sie 'license()' or 'licence()' für Details dazu.

R ist ein Gemeinschaftsprojekt mit vielen Beitragenden.
Tippen Sie 'contributors()' für mehr Information und 'citation()',
um zu erfahren, wie R oder R packages in Publikationen zitiert werden können.

Tippen Sie 'demo()' für einige Demos, 'help()' für on-line Hilfe, oder
'help.start()' für eine HTML Browserschnittstelle zur Hilfe.
Tippen Sie 'q()', um R zu verlassen.

[Vorher gesicherter Workspace wiederhergestellt]

>ls()
[1] "gretldata"
> |
```

GUI zum Vergleich - R interface unter win32 (*hüstel, würg*)

```
RGui (32-bit)
File Edit View Misc Packages Windows Help

R Console

R version 3.0.1 (2013-05-16) -- "Good Sport"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> ls()
 [1] "a"                "airquality"      "Animals"         "area"            "austpop"         "b"               "b0"              "b1"              "bols"
[10] "brunotest"       "c0"              "c1"              "caesar.shoe"    "catheter"       "ChickEgg"       "Credit"          "Credit.ols"     "df"
[19] "draw.plotmath.cell" "draw.title.cell" "drivers"         "elastic.lm"     "errors"          "eurodist"       "Example"         "fbeta"          "fbeta.tmp"
[28] "g"               "get.c"          "get.r"          "growth.rates"  "i"              "intake"         "iris"           "ironslag.lm"   "j"
[37] "k"               "K"              "K1"             "Kerstin"        "Kk"             "l"              "L"              "last.warning"  "lik.mean"
[46] "lik.regress"     "lik.regressR"   "LR"             "make.table"    "Mediadaten"     "model"          "modelmu"        "modelR"        "mu"
[55] "n"               "nc"             "nr"             "oldpar"        "OLS.model"      "ones"           "op"             "opar"          "open.account"
[64] "Ostrom"         "p"              "p0.true"       "p1.true"       "pie.sales"     "pin"            "pp"             "presidents"    "q"
[73] "quakes"         "r"              "r0"             "r1"             "regress.model"  "robert"        "robust.se"      "ross"          "rr"
[82] "S2"             "scale"         "se"             "sebastianstest" "Senate"         "showSmooth"    "Sig"            "Sigunb"        "sm"
[91] "sm.3RS"         "sm2"           "sm3"           "SSE"           "toy cars"       "trees"          "tstat"          "USDistLag"     "uspop"
[100] "usr"            "u"             "u1"            "volcano"       "wmeanR"         "x"              "x1"             "x3R"           "xadd"
[109] "xdelta"         "xscale"        "xx"            "y"             "y0"             "y1"            "yadd"           "ydelta"        "yhat"
[118] "yscale"         "yy"            "z"             "Z0"            "Z00"           "Z05599"        "zstat"

> |
```

R incl GUI: Rcmdr mit Code, warnings, session logging, Tab.editor, Grafik

RGui

Datei Windows Bearbeiten Hilfe

R Console

Dateneditor

	Y	X1	X2	X3	NEWVAR	X1X2	X3sq	residuals	fitted	var10
1	1	-0.545	0.16	0.033	0.648	-0.0872	0.001089	0.1793127	0.8206873	
[1]	2	0	0.892	0.125	1.476	2.493	0.1115	2.178576	-0.5963473	0.5963473
[12]	3	2	1.647	0.619	-0.262	4.004	1.019493	0.068644	0.4133659	1.586634
[23]	4	2	1.749	-1.446	0.31	2.613	-2.529054	0.0961	-0.5714932	2.571493
[34]	5	2	0.362	-0.589	-1.404	0.369	-0.213218	1.971216	1.236769	0.763231
> shi	6	0	0.531	-0.606	0.777	0.702	-0.321786	0.603729	-1.413514	1.413514
func	7	2	0.003	-0.8	-0.897	0.306	-0.0024	0.804609	0.8252701	1.17473
s	8	0	0.26	0.597	-0.64	0.217	0.15522	0.4096	-0.8227212	0.8227212
wsi	9	3	1.502	-0.309	0.112	4.305	-0.464118	0.012544	1.030575	1.969425
t	10	0	0.613	0.273	-0.845	0.041	0.167349	0.714025	-1.005694	1.005694
df	11	0	-1.028	-0.307	-1.17	-2.505	0.315596	1.3689	-0.2440337	0.2440337
p	12	2	0.155	-0.262	-0.534	1.359	-0.04061	0.285156	0.7844338	1.215566
re	13	1	-1.795	-2.051	-0.398	-3.244	3.681545	0.158404	-0.1673938	1.167394
re	14	0	-1.007	1.974	0.189	1.156	-1.987818	0.035721	0.2190535	-0.2190535
> rol	15	1	0.596	-0.493	-1.369	-0.266	-0.293828	1.074161	0.1324172	0.8675828
func	16									
s	17									
wsi	18									
t	19									
df										
p										
re										
dii										
re										
results										

RGL device 1 [Focus]

Rcmdr Version 1.7

Attache Paket: 'R'

The following obj

tclvalue

Warnmeldungen:

1: Paket 'Rcmdr'

2: Paket 'surviva'

>

R version 2.13.1 (2011-07-08)

R Commander

Datei Bearbeiten Datenmanagement Statistik Grafiken Modelle Verteilungen Extras Hilfe

Datenmatrix: [Beispiel](#) [Datenmatrix bearbeiten](#) [Datenmatrix betrachten](#) Modell: [<Kein aktuelles Modell>](#)

Skriptfenster

```
library(rgl, pos=4)
library(mgcv, pos=4)
scatter3d(Example$residuals, Example$fitted, Example$X1, fit="linear", residuals=TRUE,
          bg="white", axis.scales=TRUE, grid=TRUE, ellipsoid=FALSE, xlab="residuals", ylab="fitted",
          zlab="X1")
```

Ausgabefenster

Befehl ausführen

```
> library(rgl, pos=4)
> library(mgcv, pos=4)
> scatter3d(Example$residuals, Example$fitted, Example$X1, fit="linear", residuals=TRUE,
+          bg="white", axis.scales=TRUE, grid=TRUE, ellipsoid=FALSE, xlab="residuals", ylab="fitted",
+          zlab="X1")
```

Meldungen

with the single-document interface (SDI); see ?Commander.

[3] HINWEIS: Die Datenmatrix 'Example' hat 15 Zeilen und 9 Spalten.

Benutzeroberfläche — der eingebaute „Dateneditor“ von R

	Y	C	var3	var4	var5
1	791.8	733.2			
2	819	748.7			
3	844.3	771.4			
4	880	802.5			
5	894	822.7			
6	944.5	873.8			
7	989.4	899.8			
8	1012.1	919.7			
9	1028.8	932.9			
10	1067.2	979.4			
11	1091.1	1005.1			
12	1123.2	1025.2			
13	1170.2	1069			
14	1207.3	1108.4			
15	1291	1170.6			
16	1365.7	1236.4			
17	1431.3	1298.9			
18	1493.2	1337.7			
19	1551.3	1405.9			
20	1599.8	1456.7			
21	1688.1	1492			
22	1728.4	1538.8			
23	1797.4	1621.9			
24	1916.3	1689.6			
25	1896.6	1674			

**das hier ist
kein Formeleditor!**

Rstudio 0.97.336: Codefenster, Objektliste, Editor, Grafik bzw. Hilfefenster

The screenshot displays the RStudio 0.97.336 interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Project, Build, Tools, and Help. The toolbar contains icons for file operations and a 'Go to file/function' search bar. The main workspace shows a data table with 22 observations of 3 variables (YEAR, US, USSR) and a list of objects in the environment, such as 'Animals', 'Credit', 'Example', 'Ostrom', 'S2', 'SSE', 'Senate', 'airquality', 'austpop', 'bols', 'caesar.shoe', 'catheter', 'errors', 'growth.rates', 'intake', and 'isic'.

	YEAR	US	USSR
1	1967	67.357	40.000
2	1968	77.265	40.000
3	1969	77.785	47.000
4	1970	77.070	52.000
5	1971	74.472	55.000
6	1972	75.076	60.000
7	1973	73.223	65.000
8	1974	77.550	74.000
9	1975	84.900	81.000
10	1976	87.891	86.000
11	1977	95.557	103.000
12	1978	103.042	119.000
13	1979	115.013	127.000
14	1980	132.840	139.788
15	1981	156.096	153.600
16	1982	182.850	181.400
17	1983	210.484	207.400
18	1984	227.413	236.700

The console window shows the R version (3.0.1) and copyright information. It also displays the R license text and instructions for using R and its packages. The package manager window is open, showing a list of available packages with their descriptions and versions.

```
R version 3.0.1 (2013-05-16) -- "Good Sport"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from H:/Bruno/Rwork/.RData]

> View(Ostrom)
There were 50 or more warnings (use warnings() to see the first 50)
> View(Ostrom)
>
```

Package	Description	Version
<input type="checkbox"/> abind	Combine multi-dimensional arrays	1.4-0
<input type="checkbox"/> abn	Data Modelling with Additive Bayesian Networks	0.83
<input type="checkbox"/> acepack	ace() and avas() for selecting regression transformations	1.3-3.3
<input type="checkbox"/> ade4	Analysis of Ecological Data : Exploratory and Euclidean methods in Environmental sciences	1.5-2
<input type="checkbox"/> AER	Applied Econometrics with R	1.1-9
<input type="checkbox"/> alr3	Data to accompany Applied Linear Regression 3rd edition	2.0.5
<input type="checkbox"/> amap	Another Multidimensional Analysis Package	0.8-7
<input type="checkbox"/> anacor	Simple and Canonical Correspondence Analysis.	1.0-2
<input type="checkbox"/> animation	A gallery of animations in statistics and utilities to create animations	2.2
<input type="checkbox"/> aod	Analysis of Overdispersed Data	1.3
<input type="checkbox"/> ape	Analyses of Phylogenetics and Evolution	3.0-10
<input type="checkbox"/> applot	Another Plot PACKAGE: stem, leaf, bagplot, faces, spin3R, and some slider functions	1.2.7
<input type="checkbox"/> apt	Asymmetric Price Transmission (apt)	1.3
<input type="checkbox"/> ash	David Scott's ASH routines	1.0-14
<input type="checkbox"/> AutoSEARCH	General-to-Specific (GETS) Model Selection	1.2
<input type="checkbox"/> bayesm	Bayesian Inference for Marketing/Micro-econometrics	2.2-5
<input type="checkbox"/> bayesmix	Bayesian Mixture Models with JAGS	0.7-2
<input type="checkbox"/> bdsmatrix	Routines for Block Diagonal Symmetric matrices	1.3-1
<input type="checkbox"/> betareq	Beta Regression	3.0-3
<input type="checkbox"/> biclust	BiCluster Algorithms	1.0.2

Wichtige Eigenschaften von R

- Kommandozeileninterface (nur Text)
- Optionale GUIs wie „**R Cmdr**“ und andere für Mausschubser setzen auf R auf, kein integraler Teil der Codebasis.
- interne Repräsentation: *strikte Objektorientierung*
- Dokumentation jedes Paketes ist in LaTeX geschrieben!
- Grafiken universell – sowohl auf Monitor, als file oder Druck (als eps/xy-pic direkt in LaTeX bzw. PicTeX)

Vergleich andere Software, Web

- **R** = *viele/beliebige* Objekte/Tabellen gleichzeitig,
- SPSS, STATA und viele viele andere = **immer eine Datenmatrix!**
(sind die doof oder was?)

- **R** kann als „Serverprozess“ im Hintergrund hinter Apache/nginx fungieren, um bei Klick auf ein Tabellenobjekt (in Webseite) irgendwelche Berechnungen oder Grafiken an den Nutzer auszugeben (z.B. via cgi oder Hadoop):
<http://www.math.montana.edu/Rweb/>

- Speziell für Webseiten lassen sich statistische Analysen in **R** gut skripten:
Perl (Statistics::R), **Python** (RPy), **Ruby** (rsruby)

Fremdformate mit R: Daten Import & Export

- Datenaustausch: foreign-Paket für MS EXCEL-, SPSS-, STATA-Minitab, Octave, S, SAS, Systat und veraltetes dbase-Format.
Frei verfügbare Formate wie **CSV** werden unterstützt. Bitte hierzu das freie „R Import/Export Manual“ unter:
<http://cran.at.r-project.org/doc/manuals/r-release/R-data.html> lesen!
- **Datenbanken:** Oracle, MySQL, PostgreSQL, SQLite als Datenbanken, **ODBC** generell unterstützt! Allgemein zu Fremdformaten:
<http://cran.at.r-project.org/doc/manuals/R-data.html>
- Spezielle **R**-Pakete für MS Excel Formate einlesen/schreiben:
xlsReadWrite, xlsx
- Austausch mit Apple „Xgrid“: Paket „xgrid“

weitere Eigenschaften von R...

generell kann der gesamte Arbeitsspeicher genutzt werden oder nur Teile davon („`--max-mem-size=150M`“ als Option)

lässt sich per „nice“ in der Priorität gg.über anderen Prozessen einstellen,

für Entwickler: eigene Profiling-werkzeuge...

lokale Hilfeseiten

Statistical Data Analysis



Manuals

[An Introduction to R](#)
[Writing R Extensions](#)
[R Data Import/Export](#)

[The R Language Definition](#)
[R Installation and Administration](#)
[R Internals](#)

Reference

[Packages](#)

[Search Engine & Keywords](#)

Miscellaneous Material

[About R](#)
[License](#)

[Authors](#)
[Frequently Asked Questions](#)
[NEWS](#)

[Resources](#)
[Thanks](#)

Material specific to the Windows port

[CHANGES](#)

[Windows FAQ](#)

Zusatzpakete/Namen von Modulen

- **R** besitzt in der Basisinstallation nur *wenige* Funktionen
- jede ‚scientific domain‘ braucht eigene spezielle Funktionen
- jede neue Funktion wird aus kleineren, einfachen Funktionen zusammengesetzt!
- **„lmtest“** = Linear Modeling + Testing Linear Regression Models
- **„nlme“** = Linear and Nonlinear Mixed Effects Models
- **„nnDiag“** = k-Nearest Neighbor Diagnostic Tools
- **„pegas“** = Population and Evolutionary Genetics Analysis System
- **„tgp“** = Bayesian treed Gaussian process models
- **„tseries“** = Time Series (= Zeitreihen)

Hinzufügen gewünschter Funktionalität (Module)

```
library(lmtest,Paketname...)  
# und nun mach etwas mit den verfügbaren Routinen...  
meinObjekt <- FunktionausPaket(Variable-X, Option1.....)
```

R Erweiterungen systematisch: „TASKVIEW“

- 4833 verschiedene Pakete („*packages*“) im Sept. 2013
<http://cran.at.r-project.org/web/packages/>
- 31 TaskViews, vergleichbar „Metapaket“ unter Debian Linux
- „CRAN taskviewer“ (ctv) zuerst installieren:

```
install.packages("ctv")  
library("ctv")
```
- interessantes TaskView wählen <http://cran.at.r-project.org/web/views/>

```
install.views(„ChemPhys“)
```
- Wenige Sekunden warten, während ctv nötige CRAN Pakete einspielt, fertig!

R Erweiterungen systematisch – TASKVIEW (Forts.)

eine individuelle Auswahl möglicher TaskViews:

„Experimental Design“, „MedicalImage“,
„Clinical Trials“ für Mediziner / Epidemiologen,
„HighPerformance Computing“ für die Informatik,
„Genetics“ und „Phylogenetics“ für Biowissenschaften,
„Optimization“ für Mathematik,
„Multivariate“ für Multivariate Analysen,
„Psychometrics“ für Sozialwissenschaftler, Psychologen und Marketingexperten,
„SocialSciences“ für Sozialwissenschaftler mit John Fox als Betreuer,
„Environmetrics“ für die Umweltwissenschaften

Auch die spezialisierten Repos bieten „Views“:

<http://www.bioconductor.org/packages/devel/BiocViews.html>

Zum Einstieg

Welche **Eigenschaften** hat mein (Daten-) Objekt?

Welche **Variablen** stecken in diesen Daten?

Habe ich eine wirklich *klare Frage* an den „Zahlenberg“?

deutschsprachige Einführung von Andreas Handl:

<http://www.wiwi.uni-bielefeld.de/~frohn/Mitarbeiter/Handl/statskript.pdf>

FAQ: <http://cran.at.r-project.org/doc/FAQ/R-FAQ.html>

Wiki: <http://rwiki.sciviews.org/doku.php> (hier auch kommentierte Datensätze, die innerhalb der Module mitgeliefert werden).

Suche in R-spezifischer Doku: <http://www.r-project.org/other-docs.html>

Konventionen zum Scripting in R

R kennt *mehrere* Betriebsmodi.

Ausgangspunkt ist der gute alte *batch-Modus*!

- ein R Skript besteht aus einer Folge von R-Befehlen, um eine Reihe logischer Arbeitsschritte hintereinander auszuführen.
- R Skripte enden auf *.R im Dateinamen.
- R Skripte lassen sich mit *jedem* Editor bearbeiten (d.h. vi- oder Emacs)
- R Skripte sind portabel
- R ist prinzipiell SMP-/Clusterfähig (HPC): „library(parallel)“
„library(snow)“
- # am Zeilenanfang deklariert den Rest der Zeile als Kommentar!

zur Arbeitsweise in R ... (und Skript anlegen)

```
> pi/2  
[1] 1.570796
```

```
> 3+5  
[1] 8
```

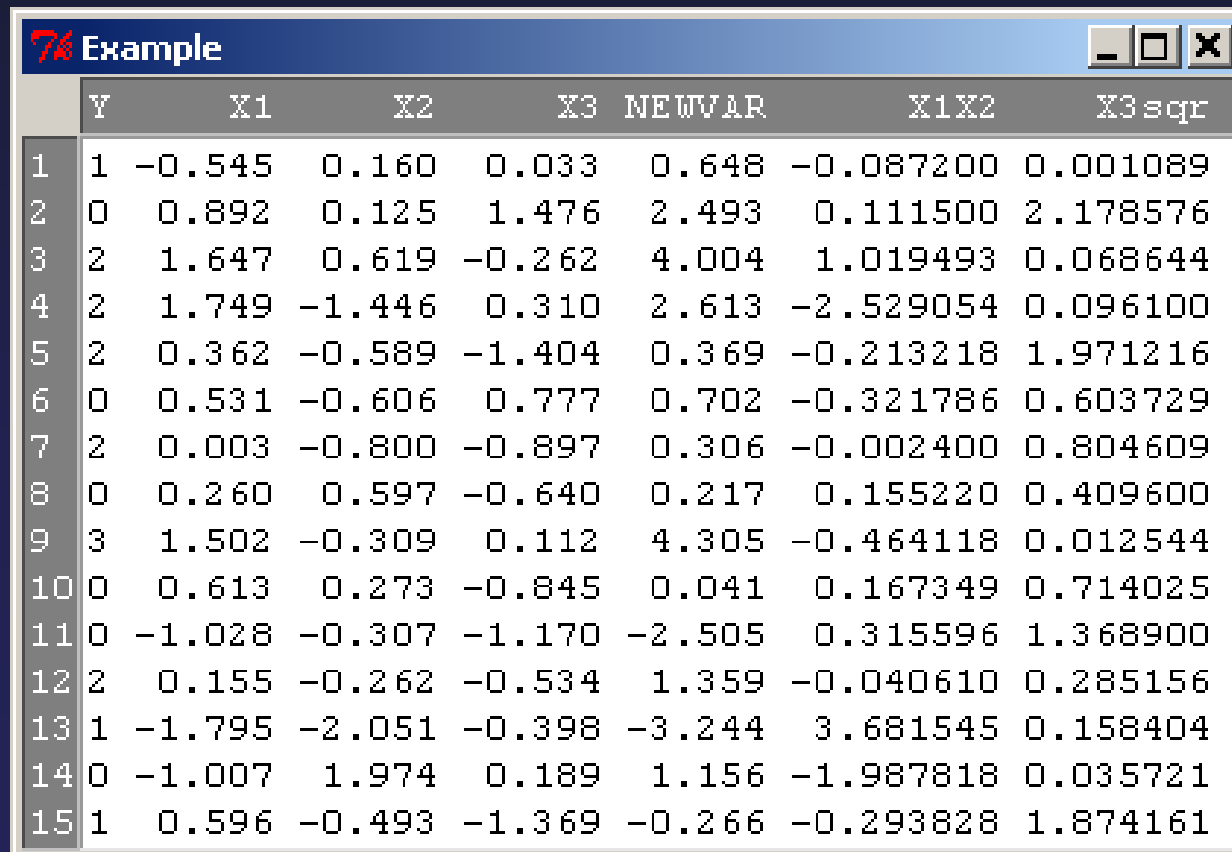
```
> testvar <- c(2,4,7,17,38,64)  
> show(testvar)  
[1] 2 4 7 17 38 64  
> sum(testvar)  
[1] 132
```

Beispiel: Daten einlesen ASCII Format via R Skript

hier wird ein KLARTEXT Rohdatenfile eingelesen,
die 1.Zeile enthält Var-labels.

```
Example <- read.table("EXAMPLE.TXT", header=TRUE)  
attach(Example)
```

Example



	Y	X1	X2	X3	NEWVAR	X1X2	X3sqr
1	1	-0.545	0.160	0.033	0.648	-0.087200	0.001089
2	0	0.892	0.125	1.476	2.493	0.111500	2.178576
3	2	1.647	0.619	-0.262	4.004	1.019493	0.068644
4	2	1.749	-1.446	0.310	2.613	-2.529054	0.096100
5	2	0.362	-0.589	-1.404	0.369	-0.213218	1.971216
6	0	0.531	-0.606	0.777	0.702	-0.321786	0.603729
7	2	0.003	-0.800	-0.897	0.306	-0.002400	0.804609
8	0	0.260	0.597	-0.640	0.217	0.155220	0.409600
9	3	1.502	-0.309	0.112	4.305	-0.464118	0.012544
10	0	0.613	0.273	-0.845	0.041	0.167349	0.714025
11	0	-1.028	-0.307	-1.170	-2.505	0.315596	1.368900
12	2	0.155	-0.262	-0.534	1.359	-0.040610	0.285156
13	1	-1.795	-2.051	-0.398	-3.244	3.681545	0.158404
14	0	-1.007	1.974	0.189	1.156	-1.987818	0.035721
15	1	0.596	-0.493	-1.369	-0.266	-0.293828	1.874161

wir wollen ganz einfache statistische Maße z.B.
Mittelwerte haben ... vom Beispieldatensatz „Example“

```
> summary(Example)
```

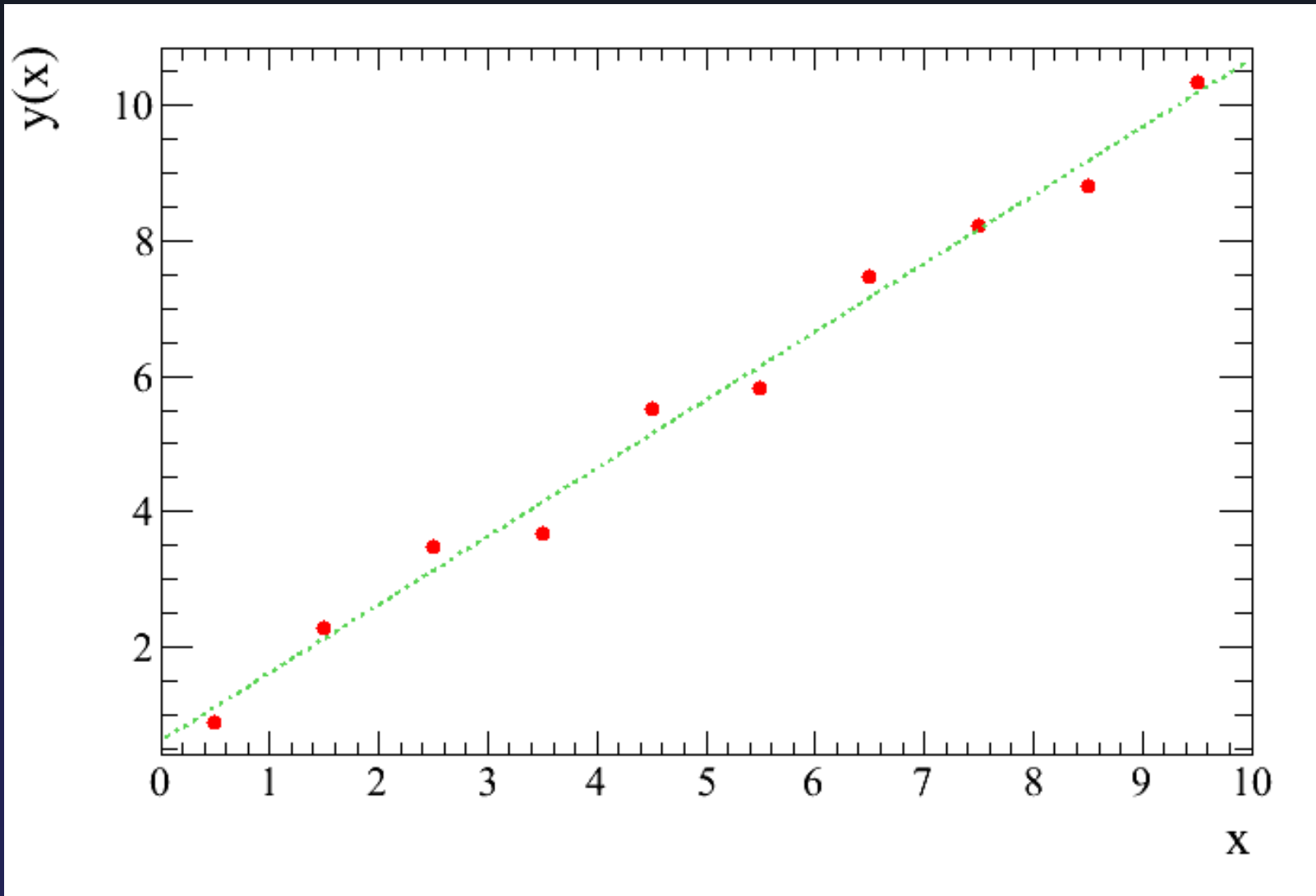
Y	X1	X2	X3
Min. :0.000	Min. :-1.7950	Min. :-2.0510	Min. :-1.4040
1st Qu.:0.000	1st Qu.: -0.2710	1st Qu.: -0.5975	1st Qu.: -0.8710
Median :1.000	Median : 0.3620	Median : -0.3070	Median : -0.3980
Mean :1.067	Mean : 0.2623	Mean : -0.2077	Mean : -0.3081
3rd Qu.:2.000	3rd Qu.: 0.7525	3rd Qu.: 0.2165	3rd Qu.: 0.1505
Max. :3.000	Max. : 1.7490	Max. : 1.9740	Max. : 1.4760

Ein erstes eigenes Skript in R

starte einen beliebigen Editor deiner Wahl (möglichst einen mit Syntax-highlighting)
Der Emacs beherrscht den sog. ESS („*Emacs speaks Statistics*“).

```
# das wird ein erstes R skript
# ein Objekt "Example" aus eingelesener Datei erzeugen.
Example <- read.table("EXAMPLE.TXT", header=TRUE)
attach(Example)
Example
# gib eine erste Übersicht über alle Variablen und
# deren wichtigste Eigenschaften aus "Example"
summary(Example)
```

Beispiel: wie formuliere ich einen Zusammenhang? (etwa in einer Regression)



Quelle: <http://de.wikipedia.org/wiki/Ausgleichsrechnung>

Matrixoperationen, ATLAS und LAPACK für Lineare Algebra

zu Begriffen/nützlichen Libraries in der Linearen Algebra

ATLAS = Automatically Tuned Linear Algebra Software

(<http://math-atlas.sourceforge.net/>)

LAPACK = Linear Algebra Packages (<http://www.netlib.org/lapack/>)

BLAS = Basic Linear Algebra Subprograms (<http://netlib.org/blas/>)

#

importiere R Pakete

library(Matrix, matrixStats, Matrixmodels, matrixcalc, maxLinear, maxLik)

library(maxstat)

und nun der restliche Kram

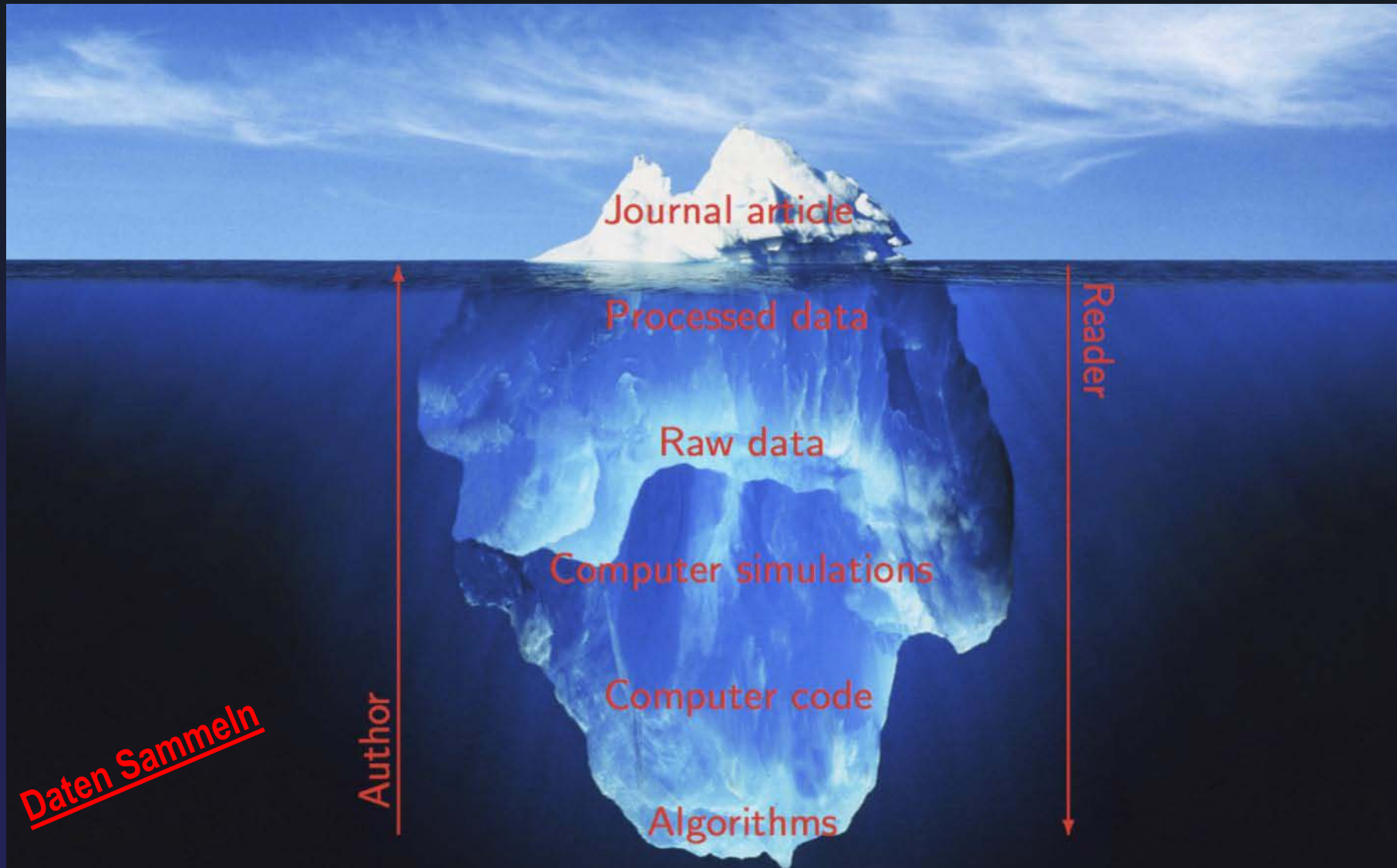
Methodenausblick: was geht ...

- für:
Mehrebenenanalysen, Regressionen, MLE, Cluster-/Faktor-/Korrespondenzanalysen, MDS, Zeitreihen, Bayes'sche Analysen und MCMC, Numerik, dynamische Modelle (Wirtschaftssimulationen, Epidemiologie...),
jede Visualisierung incl. mehrdimensionaler Modelle (!)
- Grafiker sollten den „hwriter“ (HTML writer) für Webseitengestaltung prüfen.
hwriter erzeugt aus R-output direkt HTML

Zum Überblick gehe bitte auf

http://cran.at.r-project.org/web/packages/available_packages_by_name.html

Welche Schritte erfolgen in (fast) jeder Datenanalyse?



<http://www.r-bloggers.com/cycles-in-finite-populations-a-reproducible-seminar-in-three-acts/>

ab in die Zukunft – R-forge

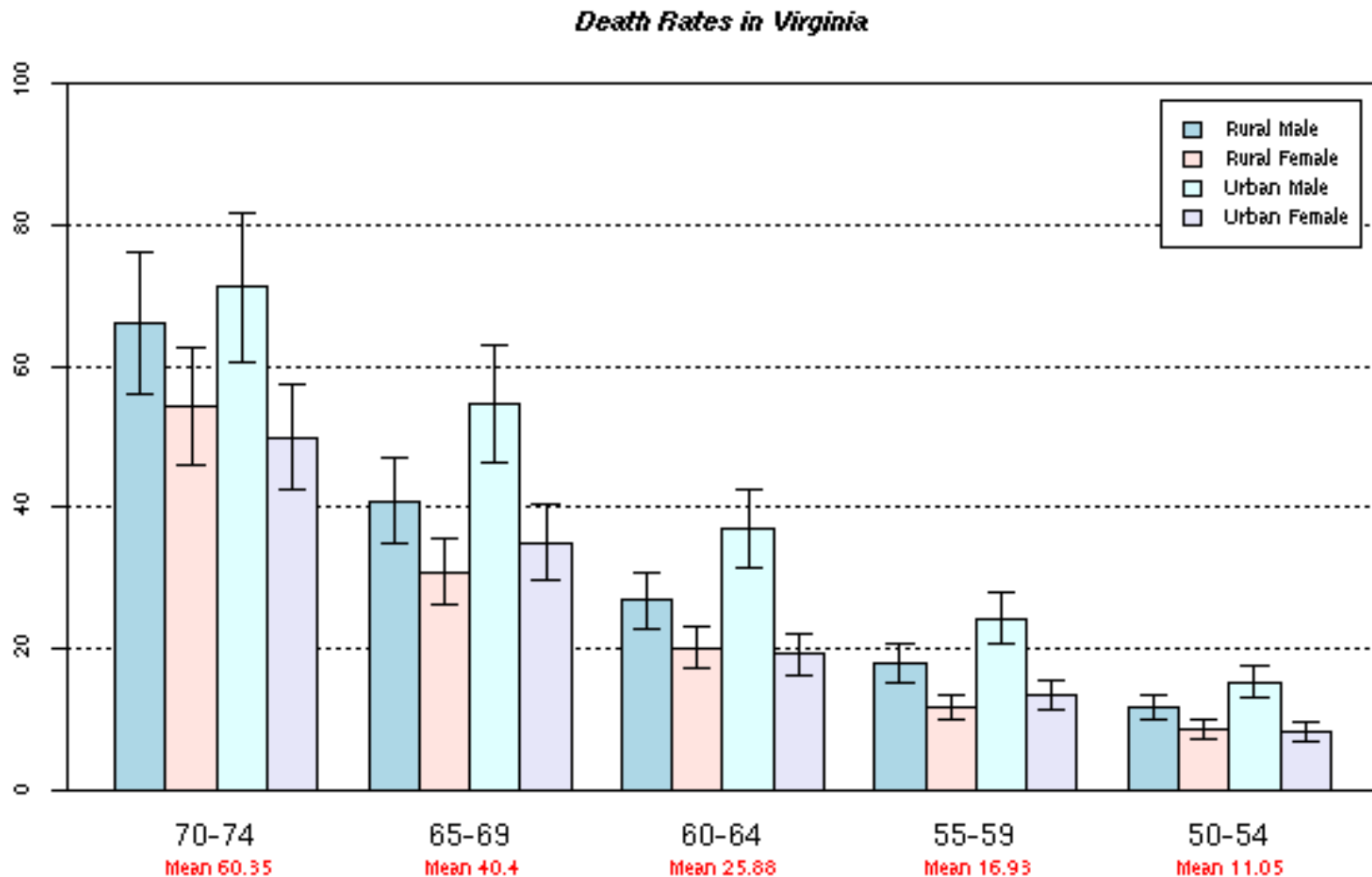
<http://r-forge.r-project.org/>

„**R-Forge** offers a central platform for the development of R packages, R-related software and further projects. It is based on FusionForge offering easy access to the best in SVN, daily built and checked packages, mailing lists, bug tracking, message boards/forums, site hosting, permanent file archival, full backups, and total web-based administration.“

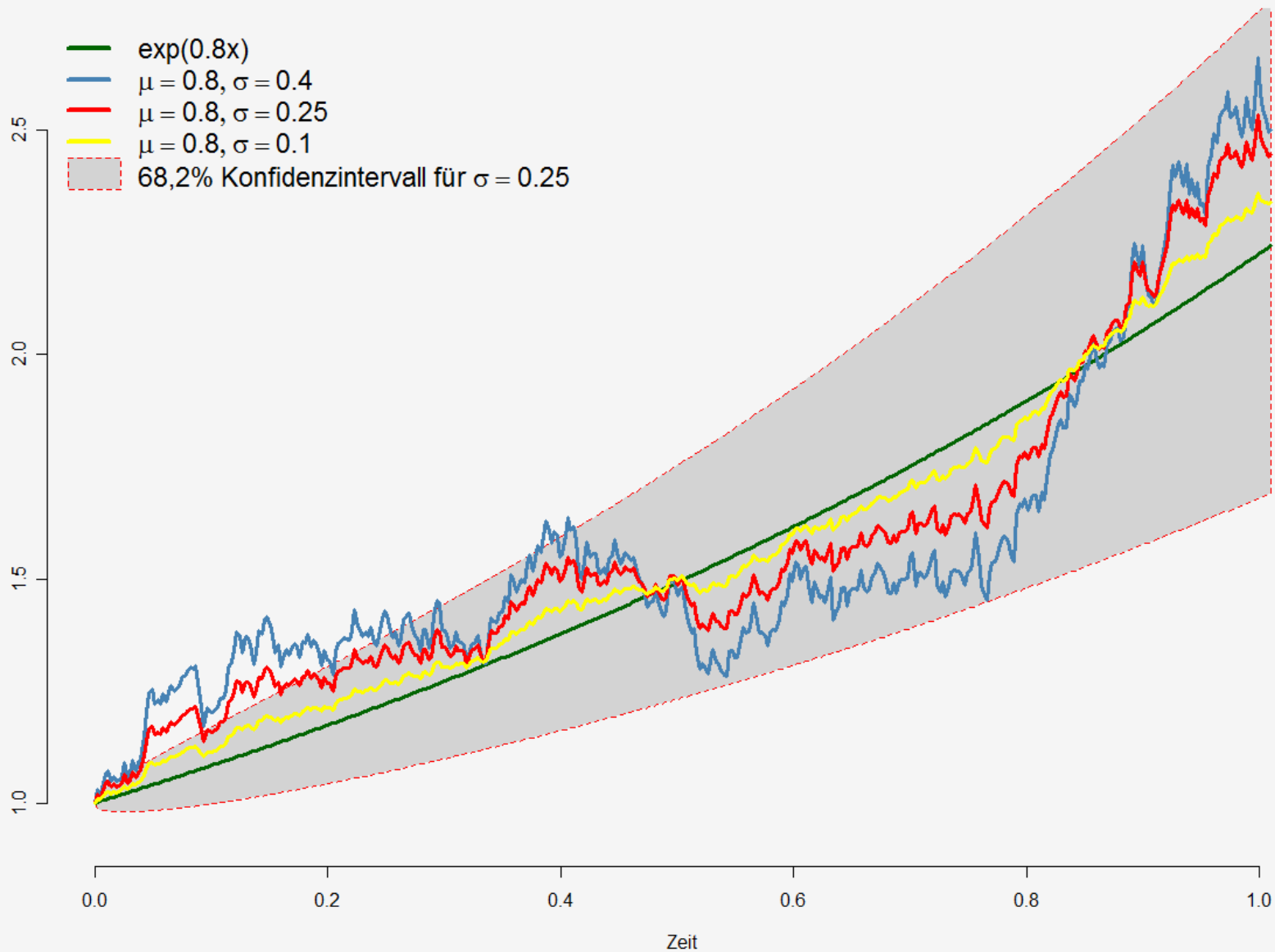
- **Intro:** http://download.r-forge.r-project.org/RJournal_2009-1_Theussl+Zeileis.pdf
- **User Manual:** http://r-forge.r-project.org/R-Forge_Manual.pdf

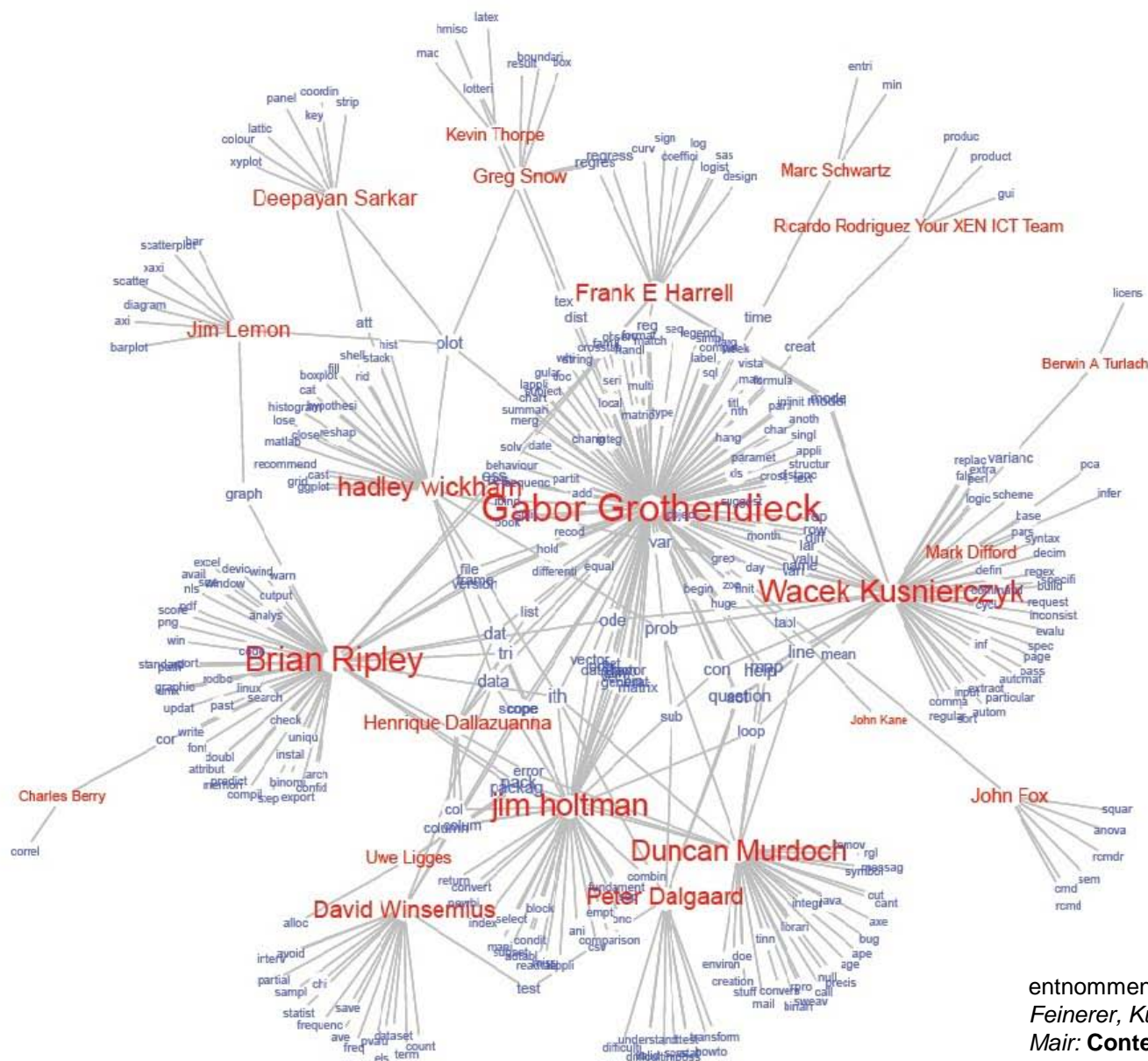
jetzt folgen grafische Beispiele

Beispiel: Sterbefälle nach Altersgruppen (in Virginia/USA): *M/W* und *Stadt/Land*



Geometrische brownische Bewegung





entnommen: Angela Bohn, Ingo Feinerer, Kurt Hornik und Patrick Mair: Content-Based Social Network Analysis of Mailing Lists , in: R Journal 2011, Heft 1, Seite 15

ausgewählte kleine Hilfsmittel für R

R Suchmaschine: <http://www.rseek.org/>

das R Journal: <http://journal.r-project.org/>

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

Data Mining z.B.: <http://www.rdatamining.com/>

Vier Mailinglisten: <http://www.r-project.org/mail.html>

(R-announce, R-packages, R-help, R-devel)

Veröffentlichungen zu R: <http://www.r-project.org/doc/bib/R-books.html>

(TeX-Anwender: *bibtex*-files zum download)

Mehrsprachige Doku unter: <http://www.r-project.org/other-docs.html> (EN, DE, F, ES, JP, PT, RU ...)

Neueste Entwicklungen und Meinungen: <http://www.r-bloggers.com/>

Konferenzen wie use!R <http://www.r-project.org/conferences.html> (teilnehmen!)

Falls jemand R als nützlich erlebt, bitte über eigenes Engagement in der „R-foundation“ d.h. der Stiftung zum Wohle der Community nachdenken:

<http://www.r-project.org/foundation/main.html>

Ausgewählte vertiefende Literatur

- **Adler J.** (2010): R in a Nutshell. Köln: O'Reilly (deutsche Ausgabe, prima Nachschlagewerk)
- **Dalgaard P.** (2002): Introductory Statistics with R. New York: Springer (Einsteiger)
- **Leisch F.** (2002): Sweave: dynamic generation of statistical reports using literate data analysis. In: *Compstat 2002 - Proceedings in Computational Statistics*, Hrsg.: Härdle W, Rönz B.; Heidelberg: Physika Verlag, 575-580.
- **Maindonald J./W.J. Braun** (2010): Data Analysis and Graphics using R. Cambridge: Cambridge University Press (3.Auflage) english, Grundriß+Nachschlagewerk.
- **Venables W.M./D. Smith/R core Project** (2013): An Introduction to R. <http://cran.r-project.org/doc/manuals/R-intro.pdf> (ca 100 Seiten, wichtigste Leseempfehlung)

Fragen, Kommentare?

Danke!

and now something totally different:

Mittagessen beim Mexikaner

(Cantina Mexicana, Weyertal 38, ca 250 m)

Dank an unseren Sponsor!