

GESIS-Workshop
"Datenanalyse mit R"
DLR Berlin

Jan-Philipp Kolb

Donnerstag, 06. November, 2014





Gliederung

Warum R nutzen

Modularer Aufbau von R

Anwendungsbeispiele

Regression/Zeitreihen

Graphiken

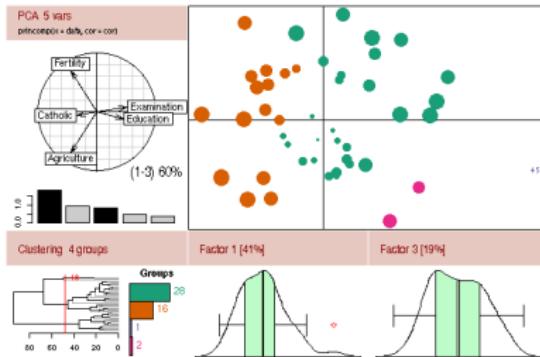
Reproducible Research

Schnittstellen

Quellen/Literatur

Open Source Programm R

- ▶ R ist eine freie, nicht-kommerzielle Implementierung der Programmiersprache S (von AT&T Bell Laboratories entwickelt)
- ▶ Freie Beteiligung ⇒ modularer Aufbau (immer mehr Erweiterungspakete)



www.r-project.org

Gliederung

Warum R nutzen

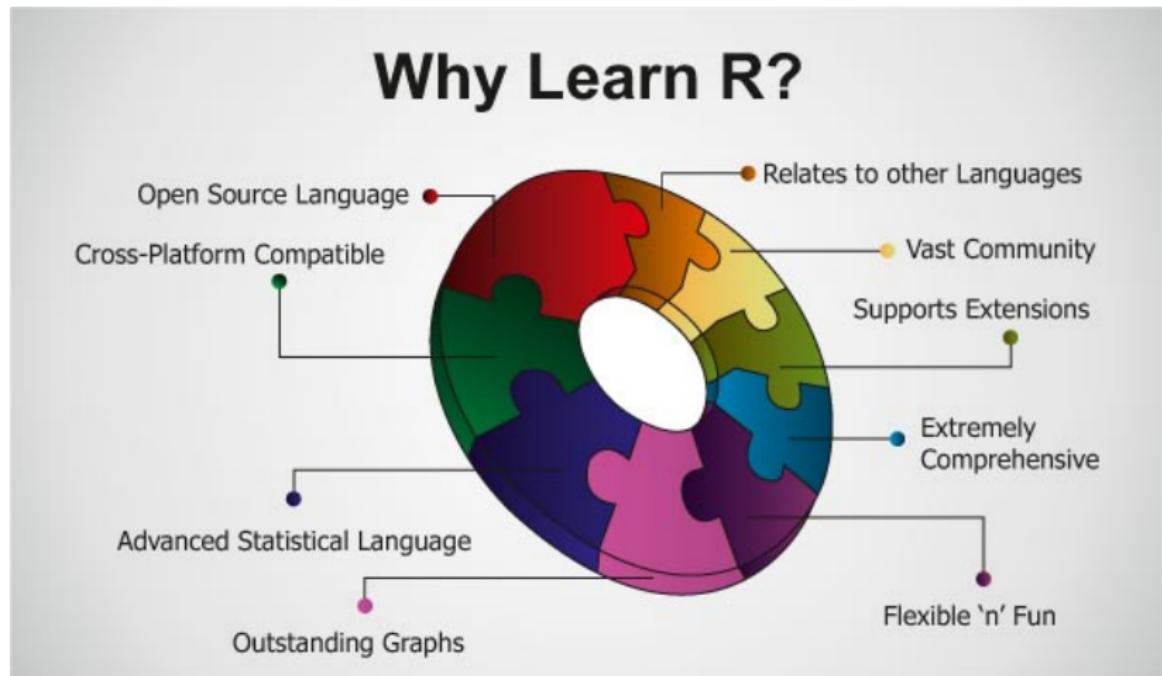
Modularer Aufbau von R

Anwendungsbeispiele

Schnittstellen

Quellen/Literatur

Warum R nutzen?



<http://www.edureka.co/blog/why-learn-r/>

Warum R nutzen?

- ▶ Als Weg kreativ zu sein ...
- ▶ Graphiken, Graphiken, Graphiken
- ▶ In Kombination mit anderen Programmen nutzbar
- ▶ Zur Verbindung von Datenstrukturen
- ▶ Zum Automatisieren
- ▶ Um die Intelligenz anderer Leute zu nutzen ;-)
- ▶ ...

Vorteile von R

- ▶ Open Source
- ▶ Sehr aktive Nutzer
 - ▶ Hilfeforen
 - ▶ Viele spezialisierte Erweiterungen
- ▶ Visualisierung
- ▶ Nutzung von Schnittstellen
- ▶ Reproducible research

Why R? - its free!

R ist ein Open-Source Projekt, man kann R ohne Kosten nutzen.
Man muss sich keine Gedanken machen über:

- ▶ Lizenzgebühren,
- ▶ Lizenzmanager,
- ▶ Beschränkungen bei der Zahl der Nutzer.

Aber mindestens genauso wichtig ist:

R ist offen:

- ▶ Man kann den code untersuchen und daran herumbasteln wie man will
- ▶ Tausende von Experten haben das getan und von ihren Erfahrungen profitieren heute Millionen von Nutzern.
- ▶ Viele Analyse-Methoden werden zuerst in R umgesetzt

<http://www.inside-r.org/why-use-r>

R-Nutzer rund um die Welt



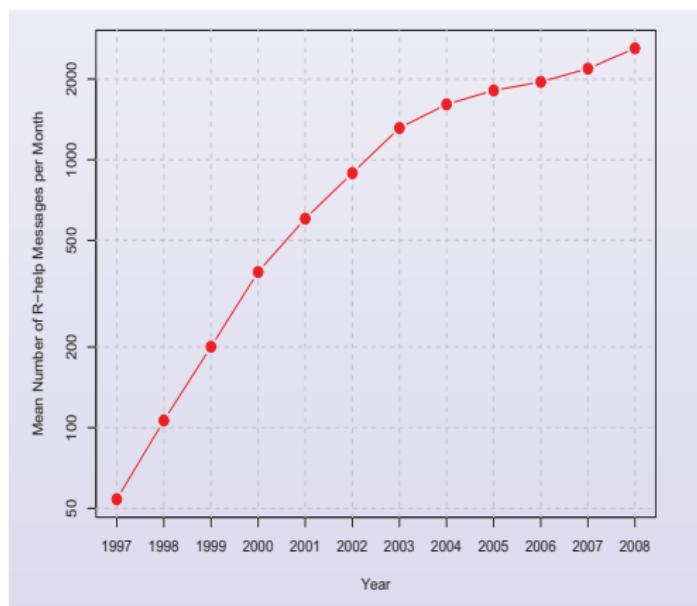
R-Nutzer rund um die Welt

R Activity Around the World

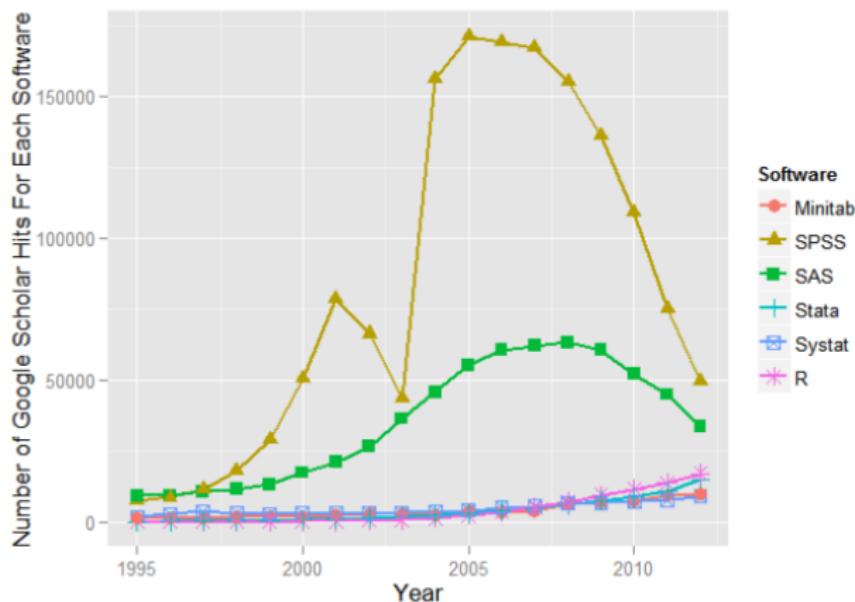


http://spatial.ly/2013/06/r_activity/

Sehr Aktive R-Nutzer - Hilfeforen/Bloggs



Popularität von R



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

github.com



Ein Einführungskurs in R

https://github.com/Japhilko/IntroR_Mannheim2014.git

The screenshot shows a GitHub repository page for 'IntroR_Mannheim2014'. At the top, there's a navigation bar with links for 'Explore', 'Gist', 'Blog', and 'Help'. On the right, there's a profile icon for 'Japhilko' and options to 'Unwatch', 'Star', and 'Fork'. Below the header, the repository name 'Japhilko / IntroR_Mannheim2014' is displayed, along with a 'branch: master' dropdown and a file list. A recent commit from 'Japhilko' is shown, indicating they uploaded slides 19 minutes ago. The main content area displays a single file: 'GESIS_R_Kurs_2014_Teil1.pdf', which is 9935.143 kb in size. There are buttons for 'Open', 'Raw', 'History', and 'Delete'.

Drei Prinzipien



3 Key Principles

Object

Everything that exists is an object.

Function

Everything that happens is a function call.

Interface

R is built on interfaces to many “algorithms”.

<http://user2014.stat.ucla.edu/files/chambers.pdf>

Wie wir denken...

- ▶ Die Programmiersprache sollte die Denkweise der Menschen wiederspiegeln.
- ▶ Ein einfaches Beispiel: Wir denken Gewicht ist eine Funktion von (abhängig von) Größe und Geburstag.

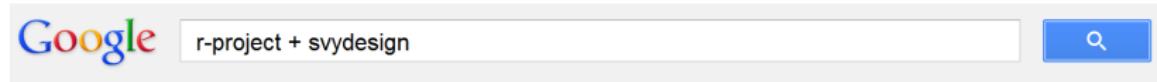
Die R Formel um dies auszudrücken ist:

```
lm(Gewicht ~ Groesse + Geburstag)
```

<http://www.burns-stat.com/documents/tutorials/why-use-the-r-language/>

Wie bekommt man Hilfe?

- ▶ Um generell Hilfe zu bekommen: `help.start()`
- ▶ Online Dokumentation für die meisten Funktionen:
`help(name)`
- ▶ Nutze `?` um Hilfe zu bekommen.
Beispiel: `?mean`
- ▶ `example(lm)` gibt ein Beispiel für die lineare Regression



- ▶ Ich nutze meistens google
- ▶ Tippe:
R-project + Was ich schon immer wissen wollte
- ▶ Das funktioniert natürlich mit jeder Suchmaschine!

└ Warum R nutzen

The screenshot shows the top navigation bar of Stack Overflow with links for sign up, log in, tour, help, careers 2.0, and search. Below the header is the Stack Overflow logo and navigation links for Questions, Tags, Tour, and Users. A prominent 'Ask Question' button is on the right.

Stack Overflow is a question and answer site for professional and enthusiast programmers. It's 100% free, no registration required.

Here's how it works:

- Anybody can ask a question
- Anybody can answer
- The best answers are voted up and rise to the top

- ▶ <http://stackoverflow.com/>
- ▶ Für Fragen zum Programmieren
- ▶ Ist nicht auf R fokussiert
- ▶ Sehr detaillierte Diskussionen

Vignette Paket `zoo` - Paket für Zeitreihen

Vignetten sind kurze Paper mit Beschreibung/Beispielen zu einem Paket.

`zoo` Quick Reference

Ajay Shah

National Institute of Public
Finance and Policy, India

Achim Zeileis

Universität Innsbruck

Gabor Grothendieck

GKX Associates Inc.

Abstract

This vignette gives a brief overview of (some of) the functionality contained in `zoo` including several nifty code snippets when dealing with (daily) financial data. For a more complete overview of the package's functionality and extensibility see [Zeileis and Grothendieck \(2005\)](#) (contained as vignette "zoo" in the package), the manual pages and the reference card.

Keywords: irregular time series, daily data, weekly data, returns.

Gliederung

Warum R nutzen

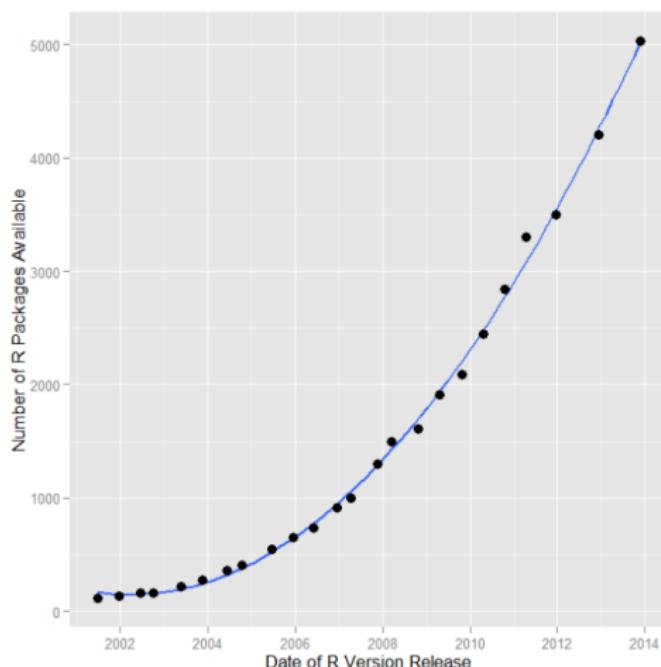
Modularer Aufbau von R

Anwendungsbeispiele

Schnittstellen

Quellen/Literatur

Modularer Aufbau von R



<http://www.statsblogs.com/2014/04/07/r-continues-its-rapid-growth/>

Modularer Aufbau

- ▶ Viele Funktionen sind im Basis-R enthalten
- ▶ Viele spezifische Funktionen sind in zusätzlichen Bibliotheken integriert
- ▶ R kann modular erweitert werden durch sog. **packages** bzw. **libraries**
- ▶ Auf **CRAN** werden die wichtigsten packages gehostet (im Moment 4567)
- ▶ Weitergehende Pakete finden sich z.B. bei www.bioconductor.org

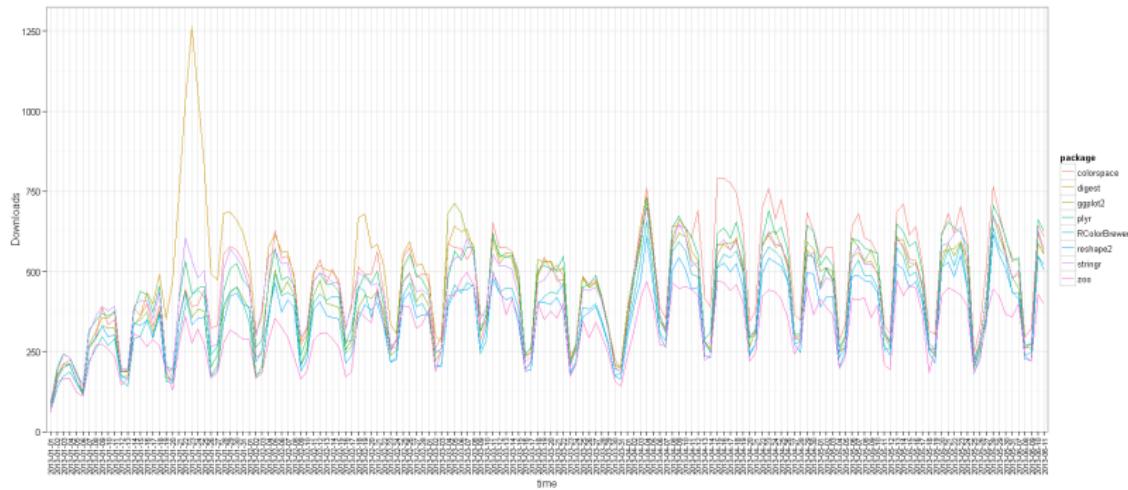
```
install.packages("lme4")
```

```
library(lme4)
```

Die am meisten genutzten Pakete



Die wichtigsten Pakete:



<http://www.r-statistics.com/2013/06/top-100-r-packages-for-2013-jan-may/>

Die 10 wichtigsten Pakete:



[http://blog.yhathq.com/posts/
10-R-packages-I-wish-I-knew-about-earlier.html](http://blog.yhathq.com/posts/10-R-packages-I-wish-I-knew-about-earlier.html)

Gliederung

Warum R nutzen

Modularer Aufbau von R

Anwendungsbeispiele

Regression/Zeitreihen

Graphiken

Reproducible Research

Schnittstellen

Quellen/Literatur

Lineare Regression in R - Beispieldatensatz

Lawn Roller Data

This data frame has 10 rows and 2 columns. Different weights of roller were rolled over
Description and the depression was recorded.

The roller data frame has 10 rows and 2 columns. Different weights of roller were rolled over different parts of a lawn, and the depression was recorded.

Usage

`roller`

This data frame contains the following columns:

Format

This data frame contains the following columns:

`:weight`

`:depth` a numeric vector consisting of the roller weights

`:depression`

`:the depth of the depression made in the grass under the roller area.`

```
library(DAAG)
data(roller)
?roller
```

Das lineare Regressionsmodell in R

Schätzen eines Regressionsmodells:

```
roller.lm <- lm(depression ~ weight, data = roller)
```

So bekommt man die Schätzwerte:

```
summary(roller.lm)
```

Falls das Modell ohne Intercept geschätzt werden soll:

```
lm(depression ~ -1 + weight, data = roller)
```

Summary des Modells

```
summary(roller.lm)
```

```
call:  
lm(formula = depression ~ weight, data = roller)  
  
Residuals:  
    Min      1Q  Median      3Q     Max  
-8.180 -5.580 -1.346  5.920  8.020  
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) -2.0871     4.7543  -0.439  0.67227  
weight        2.6667     0.7002   3.808  0.00518 **  
---  
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 6.735 on 8 degrees of freedom  
Multiple R-squared:  0.6445, Adjusted R-squared:  0.6001  
F-statistic: 14.5 on 1 and 8 DF,  p-value: 0.005175
```

R arbeitet mit Objekten

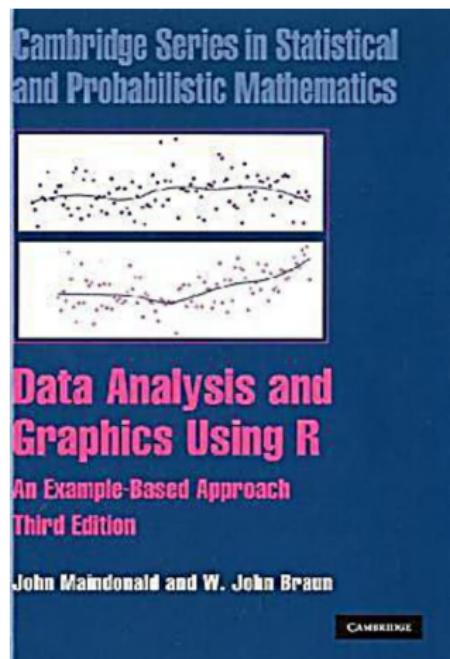
- ▶ `roller.lm` ist nun ein spezielles Regressions-Objekt
- ▶ Auf dieses Objekt können nun verschiedene Funktionen angewendet werden

```
predict(roller.lm) # Vorhersage
resid(roller.lm) # Residuen
```

Linkliste - lineare Regression

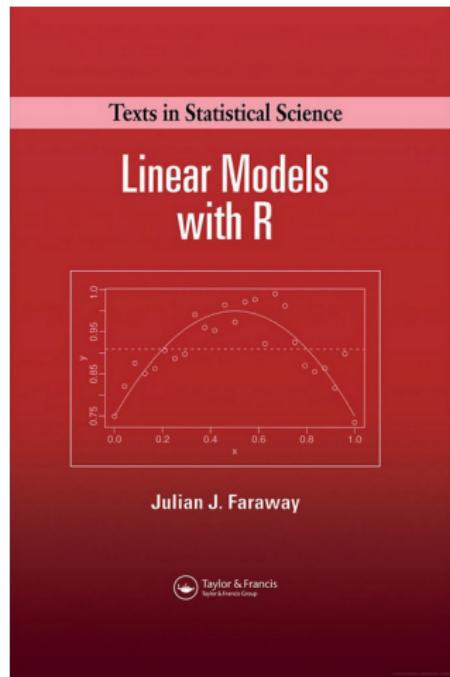
- ▶ Auf dem Kurs an der Uni Leipzig von Verena Zuber basieren auch viele der Aufgaben in diesem Workshop:
<http://www.uni-leipzig.de/~zuber/teaching/ws09/r-kurs/theorie/Kurs9.pdf>
- ▶ Eine der vielen interessanten Blogs auf r-bloggers:
<http://www.r-bloggers.com/r-tutorial-series-simple-linear-regression/>
- ▶ Komplettes Buch von Faraway (sehr intuitiv geschrieben):
<http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf>
- ▶ Gute Einführung auf Quick-R:
<http://www.statmethods.net/stats/regression.html>

Literatur Regression



1. Einführung in R
2. Datenanalyse
3. Statistische Modelle
4. Inferenzkonzepte
5. Regression mit einem Prädiktor
6. Multiple lineare Regression
7. Ausweitung des linearen Modells
8. ...

Literatur



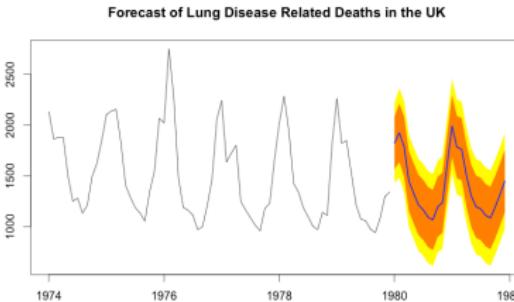
- ▶ Lineare Regression gut erklärt
- ▶ Beispiele mit R-code

Pakete - Regression

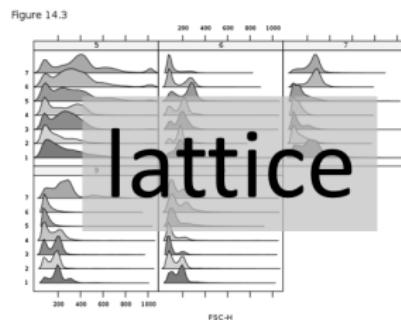
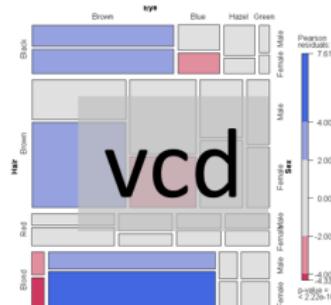
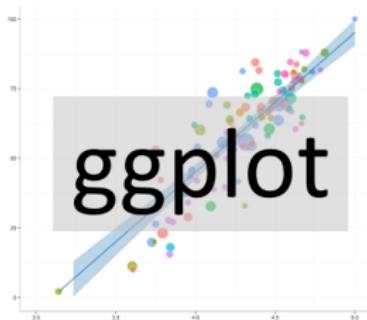
Paket	Für was?
base{lm}	Einfache lineare Regression
base{glm}	Generalisierte Lineare Modelle
tsDyn	Autoregressive Modelle (Zeitreihen)
robustbase	Robuste Regressionen
crs	Nichtparametrische Regression
glmnet	Lasso Verfahren

Zeitreihen

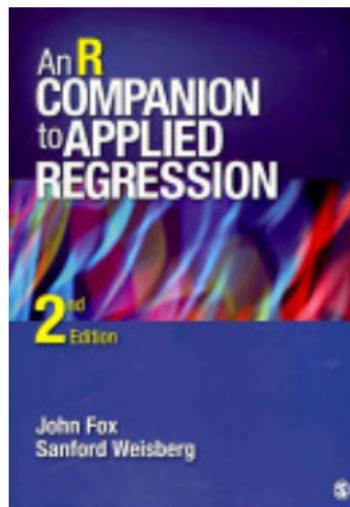
```
install.packages("forecast")
library(forecast)
fit <- auto.arima(mdeaths)
forecast(fit, level=c(80, 95, 99), h=3)
plot(forecast(fit), shadecols="oldstyle")
```



Pakete - deskriptive Datenanalyse



Der Prestige Datensatz



- ▶ Benötigte library: car
- ▶ Als Datenmaterial werden die *Prestige* Daten verwendet.
- ▶ Zugrunde liegen 102 Beobachtungen.
- ▶ Variablen des Datensatzes:
 - ▶ **income**
 - ▶ **education**
 - ▶ **women**
 - ▶ **prestige**
 - ▶ **census**
 - ▶ **type**

Der Prestige Datensatz

```
library(car)
data(Prestige)
attach(Prestige)
```

Prestige {car}

R Documentation

Prestige of Canadian Occupations

Description

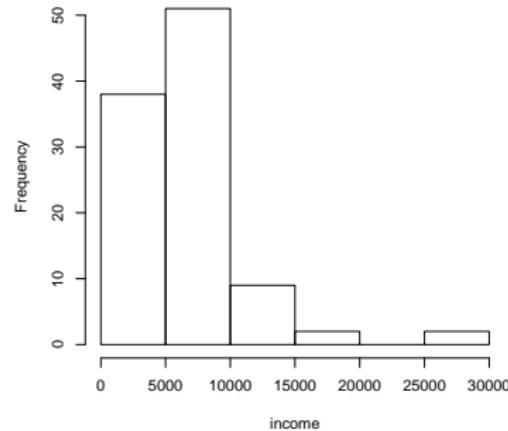
The `Prestige` data frame has 102 rows and 6 columns. The observations are occupations.

Univariate Datenanalyse

Histogramm der Variable *income*

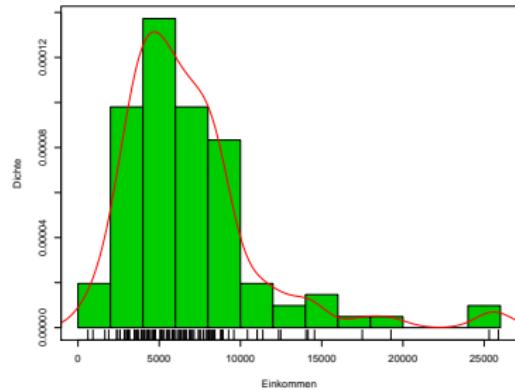
```
hist(income)
```

Histogram of income



```
hist(income, probability=T)  
points(density(income),  
type="l", col="red")
```

Histogramm: Data Prestige



Barplot

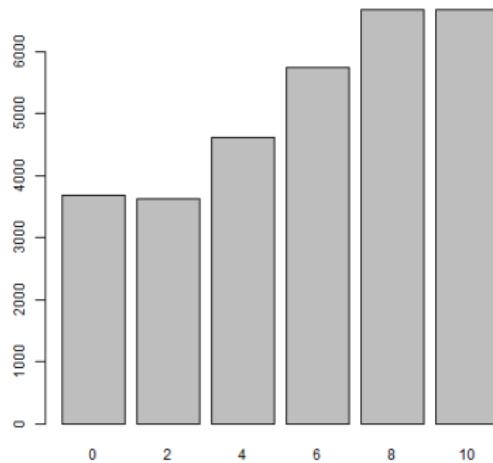
- ▶ Die Funktion `barplot()` erzeugt aus einer Häufigkeitstabelle einen Barplot
- ▶ Ist das übergebene Tabellen-Objekt zweidimensional wird ein bedingter Barplot erstellt

```
tabScore <- table(Chem97$score)
barplot(tabScore)
```

Barplots und barcharts

Mehr Farben!

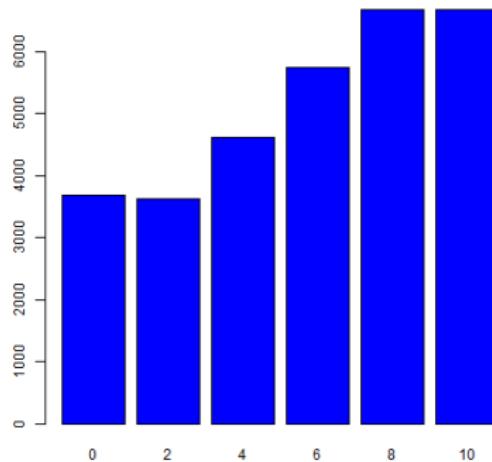
`barplot(tabScore)`



Barplots und barcharts

Mehr Farben!

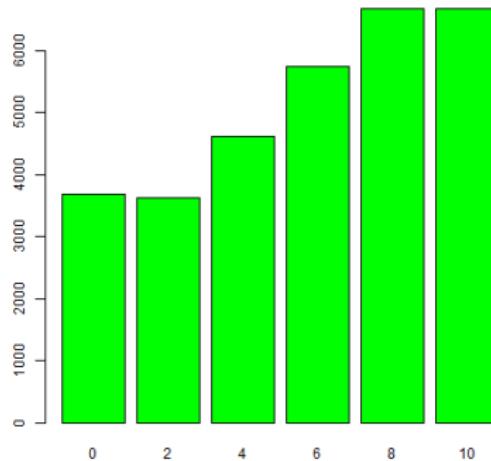
```
barplot(tabScore,col=rgb(0,0,1))
```



Barplots und barcharts

Mehr Farben!

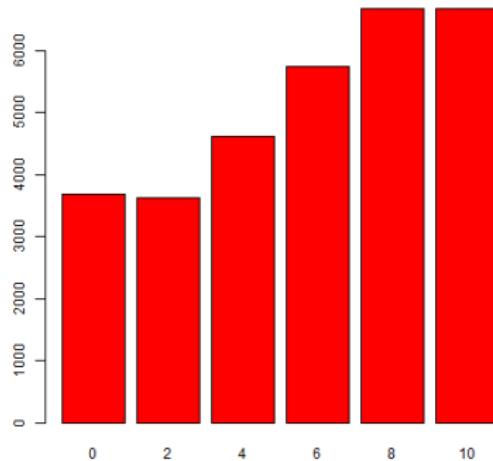
```
barplot(tabScore,col=rgb(0,1,0))
```



Barplots und barcharts

Mehr Farben!

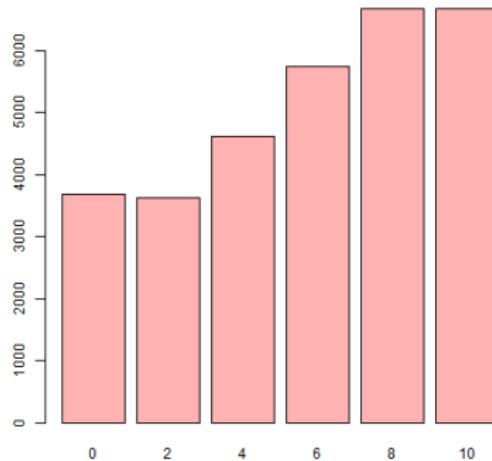
```
barplot(tabScore,col=rgb(1,0,0))
```



Barplots und barcharts

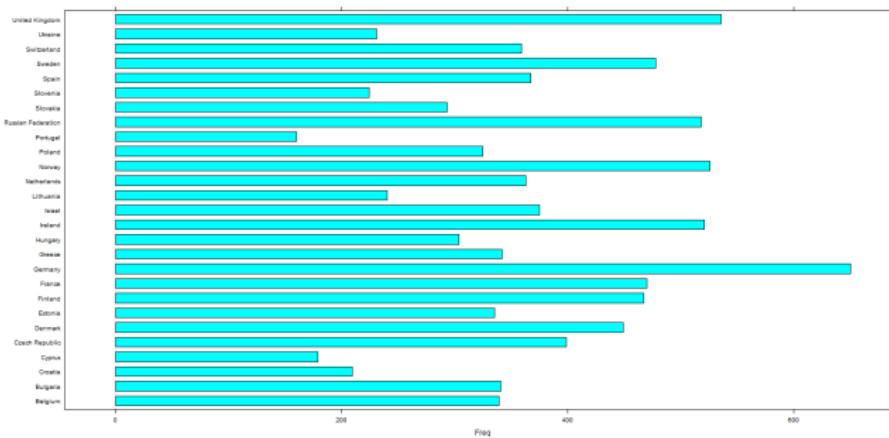
Mehr Farben!

```
barplot(tabScore,col=rgb(1,0,0,.3))
```



Barplots und barcharts

Eine erste lattice-Graphik barchart(tabScore)

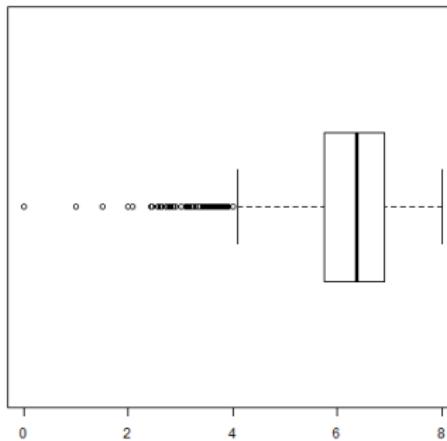


Boxplot

- ▶ Einen einfachen Boxplot erstellt man mit `boxplot()`

```
?boxplot
```

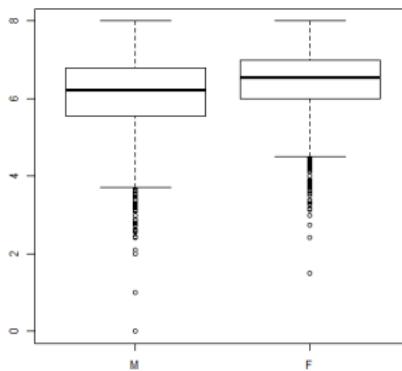
```
boxplot(Chem97$gcsescore ,  
horizontal=TRUE)
```



Gruppierte Boxplots

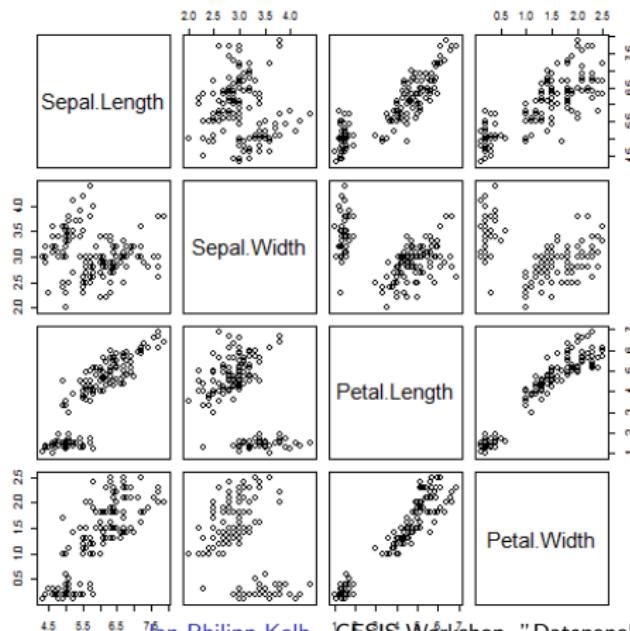
Die Funktion boxplot()

```
boxplot(Chem97$gcsescore ~ Chem97$gender)
```



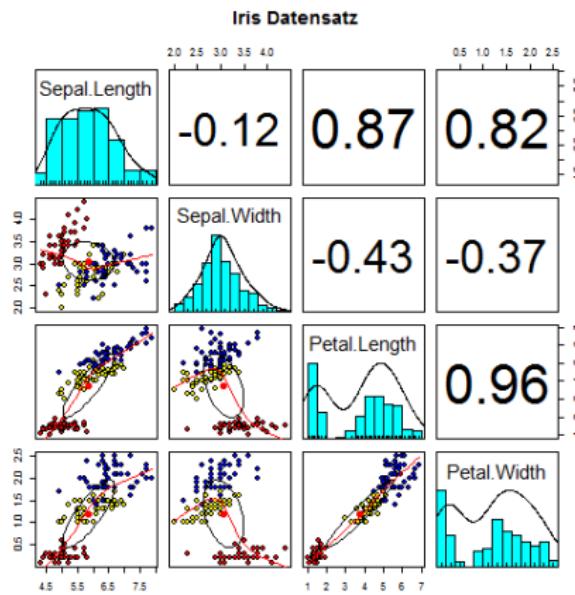
Zusammenhang zwischen mehreren Variablen

```
pairs(iris[, 1:4])
```



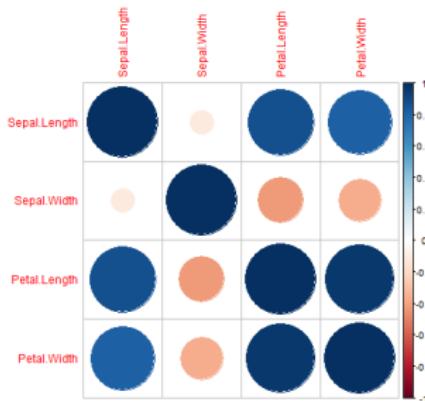
Zusammenhang zwischen mehreren Variablen

```
pairs.panels(iris[1:4], bg=c("red", "yellow", "blue")  
[iris$Species], pch=21, main="Iris Datensatz")
```

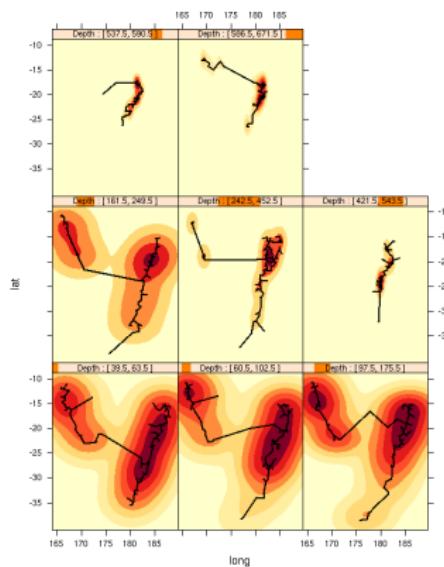


Zusammenhang - corrplot

```
library(corrplot)
M <- cor(iris[,1:4])
corrplot(M)
```



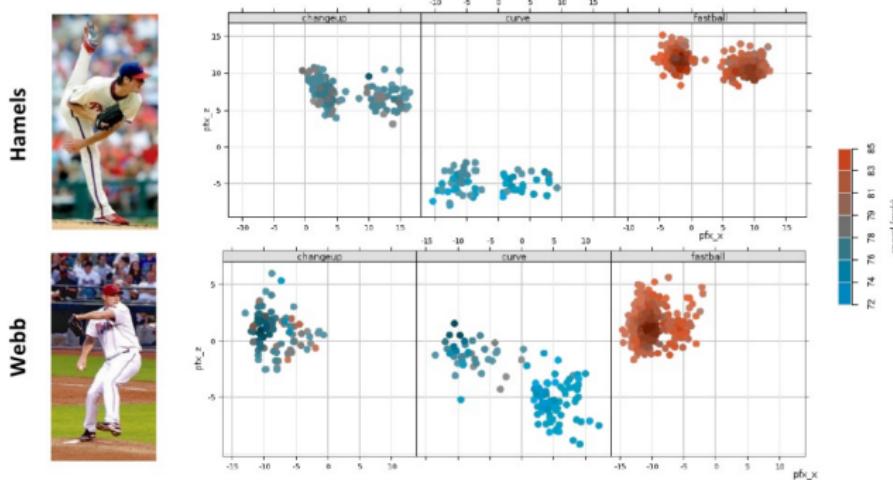
Graphiken mit lattice



http://zonnek.free.fr/blosxom/R/2006-08-10_R_Graphics.html

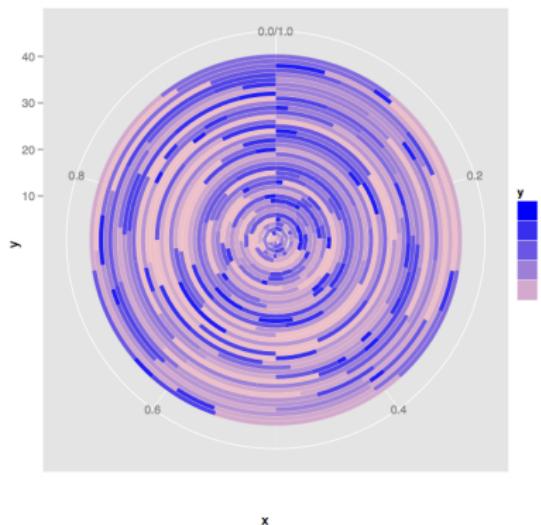
Graphiken mit lattice

A Story of Two Pitchers



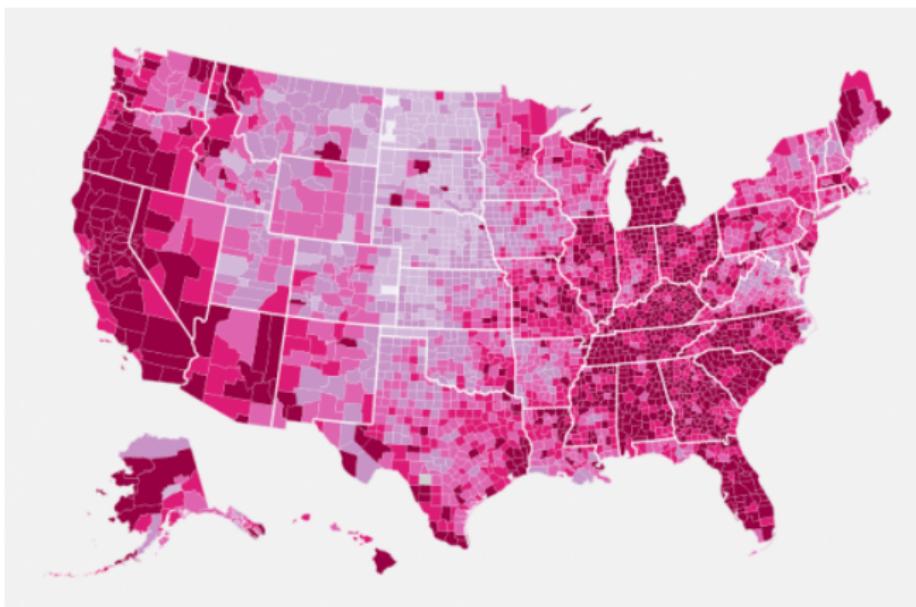
<http://de.slideshare.net/dataspora/a-survey-of-r-graphics>

Graphiken mit ggplot



<http://stackoverflow.com/questions/9483033/increase-polygonal-resolution-of-ggplot-polar-plots>

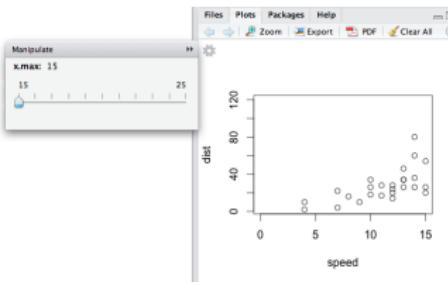
Graphiken mit ggplot



<http://www.thisisthegreenroom.com/2009/choropleths-in-r/>

Interaktive Graphiken mit manipulate

```
library(manipulate)
manipulate(plot(1:x), x = slider(1, 100))
```



<https://support.rstudio.com/hc/en-us/articles/200551906-Interactive-Plotting-with-Manipulate>

Interaktive Graphiken mit shiny

[OVERVIEW](#)[TUTORIAL](#)[ARTICLES](#)[GALLERY](#)[REFERENCE](#)[DEPLOY](#)[HELP](#)

Gallery

Interactive visualizations

Shiny is designed for fully interactive visualization, using JavaScript libraries like d3, Leaflet, and Google Charts.



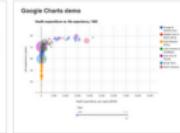
SuperZip example



Movie explorer



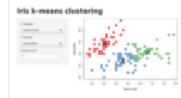
NVD3 line chart output



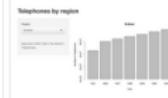
Google Charts

Start simple

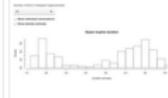
If you're new to Shiny, these simple but complete applications are designed for you to study.



Kmeans example



Telephones by region



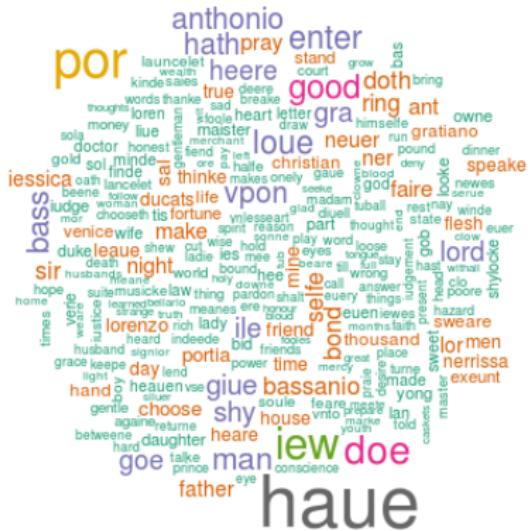
Faithful



Word cloud

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

Beispiel: wordcloud



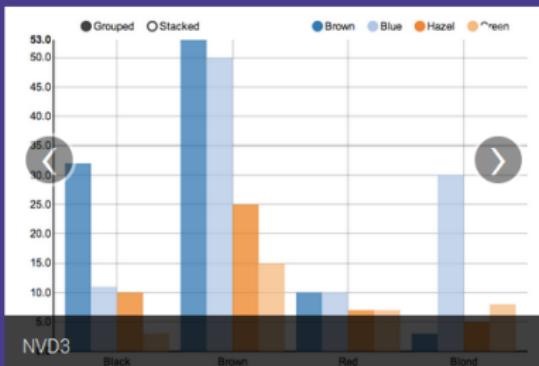
<http://shiny.rstudio.com/gallery/word-cloud.html>

Das Paket rCharts



is an R package to create, customize and publish interactive javascript visualizations from R using a familiar lattice style plotting interface.

[Learn more »](#)

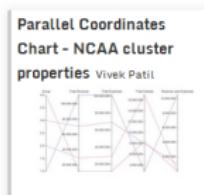
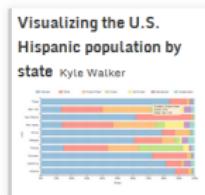
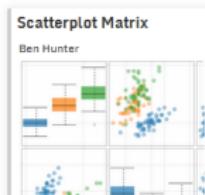
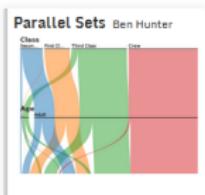
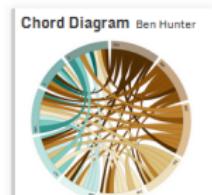
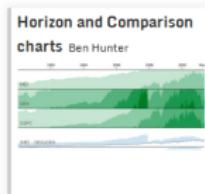
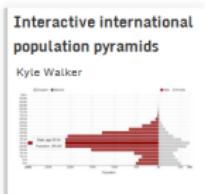
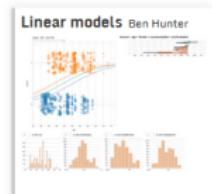


<http://rcharts.io/>

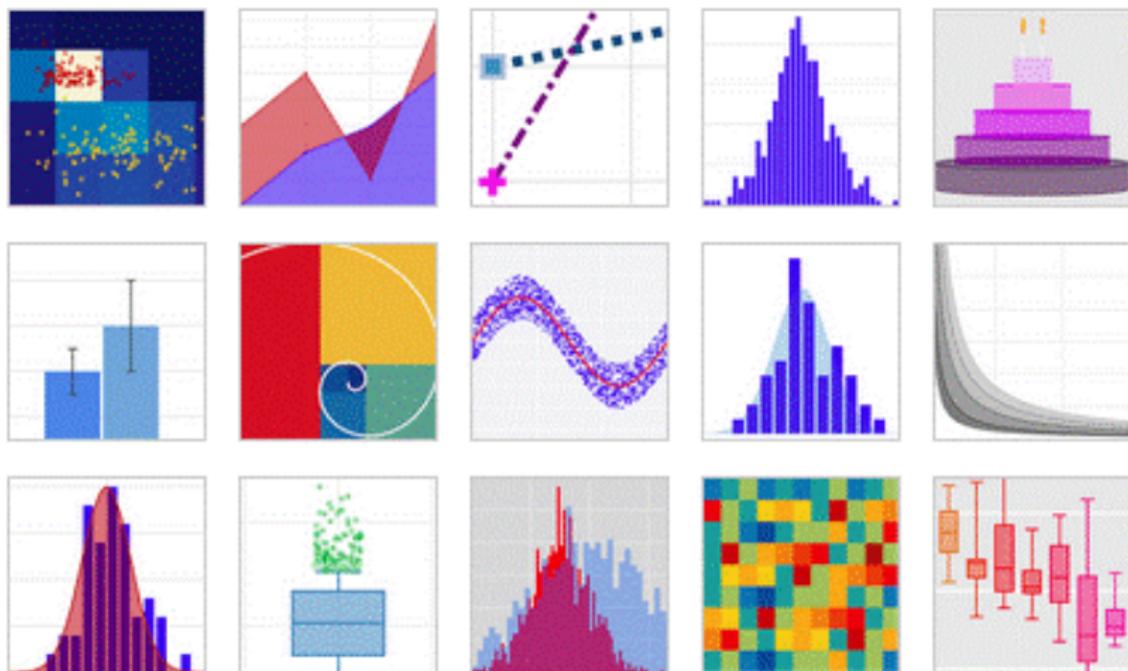
└ Anwendungsbeispiele

└ Graphiken

<http://rcharts.io/>



Plotly Beta: Collaborative Plotting with R

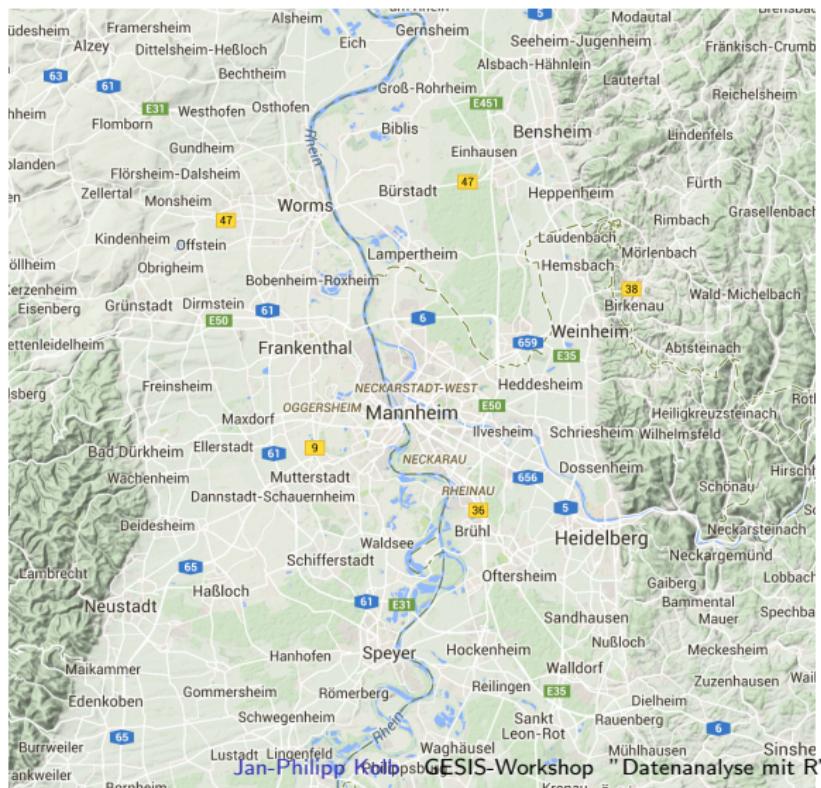


<http://www.r-statistics.com/on/r-and-the-web/>

R-Paket ggmap

```
library(ggmap)
qmap("Mannheim")
```

R-Paket ggmap - Mannheim



R-Paket ggmap - zoom

```
qmap(location = 'Mannheim', zoom = 12)
```



R-Paket ggmap - zoom

```
qmap(location = 'Mannheim', zoom = 13)
```



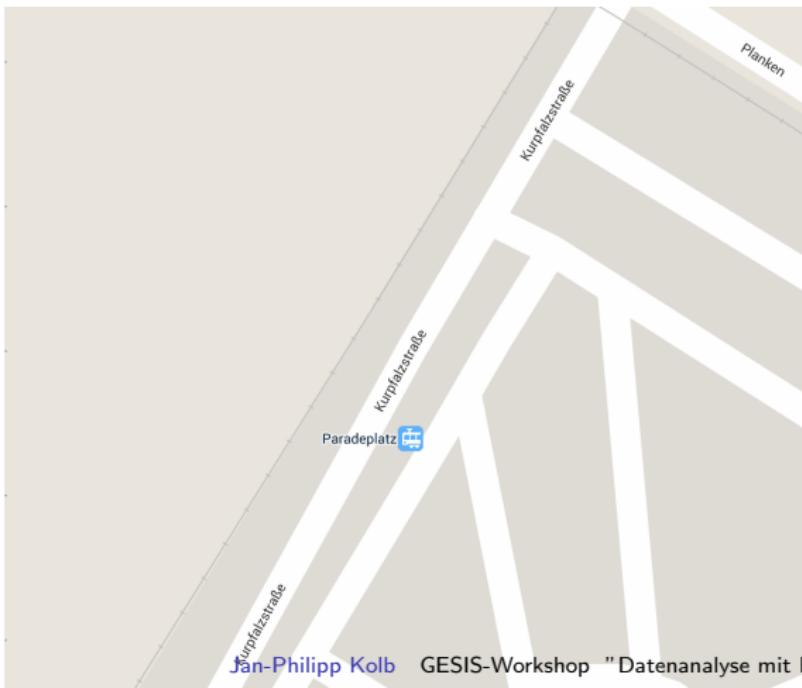
R-Paket ggmap - zoom

```
qmap(location = 'Mannheim', zoom = 14)
```



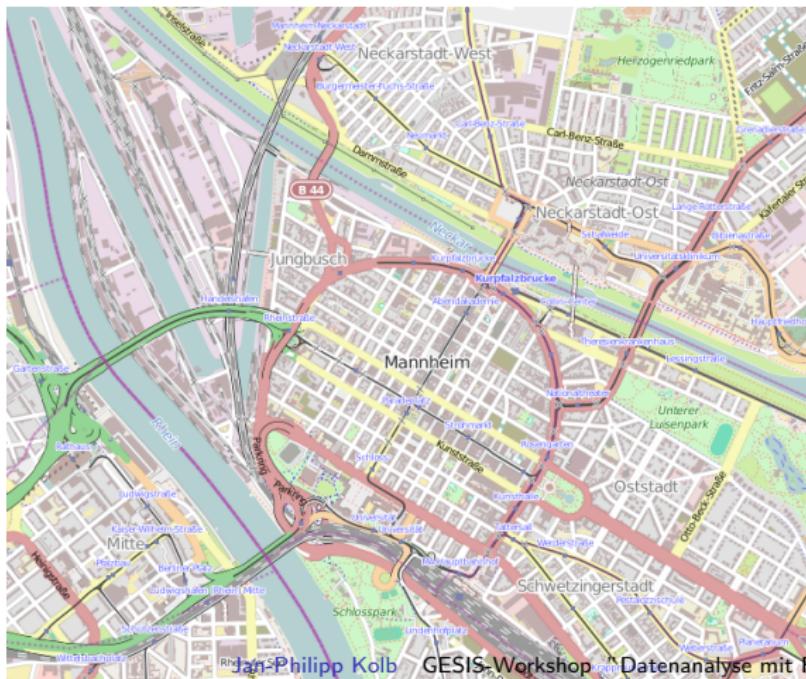
R-Paket ggmap - zoom

```
qmap(location = 'Mannheim', zoom = 20)
```



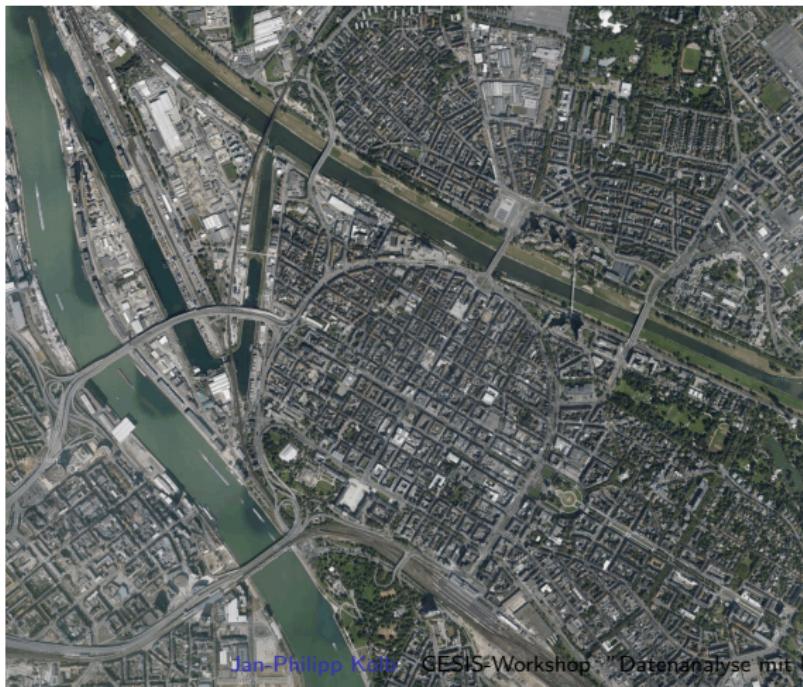
R-Paket ggmap - source

```
qmap(location = 'Mannheim', zoom = 14, source="osm")
```



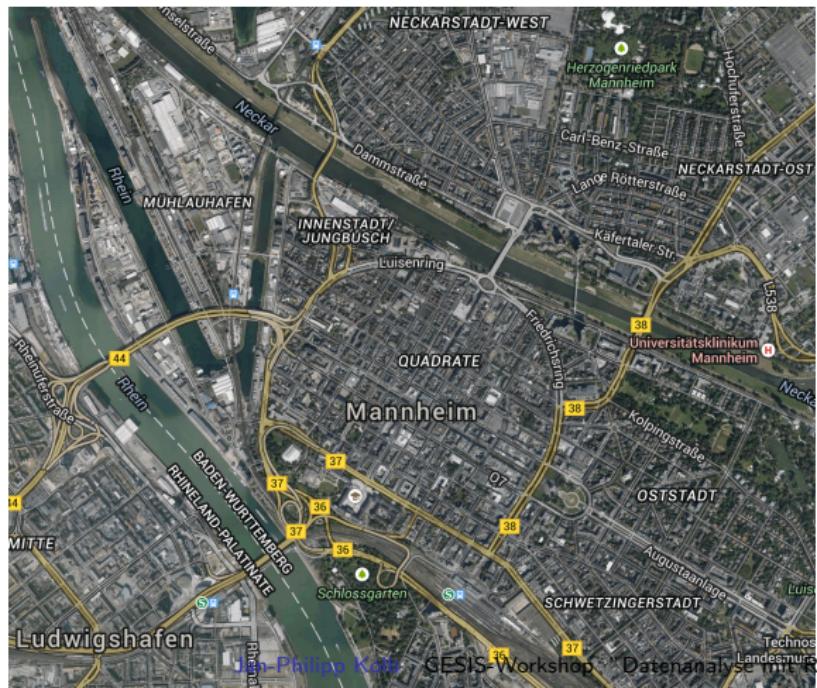
R-Paket ggmap - maptype

```
qmap(location = 'Mannheim', zoom = 14, maptype="satellite")
```



R-Paket ggmap - maptype

```
qmap(location = 'Mannheim', zoom = 14, maptype="hybrid")
```



R-Paket ggmap - maptype

```
qmap(location = 'Mannheim', zoom = 14,  
      maptype="toner",source="stamen")
```



R und LaTeX

Table 1: The Importance of Clustering Standard Errors

	M1	M2	M3
(Intercept)	4.35*** (1.32)	4.35*** (0.76)	4.35*** (1.32)
x	2.26* (1.27)	2.26*** (0.73)	2.26* (1.32)
R ²	0.00	0.00	0.00
Adj. R ²	0.00	0.00	0.00
Num. obs.	1000	3000	3000

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Quelle: <http://diffuseprior.wordpress.com/tag/stargazer/>

Das R-Paket xtable

- ▶ R bietet tolle Möglichkeiten mit LaTeX zu interagieren
- ▶ Zusammen mit word funktioniert das leider weniger gut
(höchstens Paket R2wd)

```
N <- 1000

a <- sample(1:5,N,replace=T)
b <- sample(1:3,N,replace=T)

tabAB <- table(a,b)

xtable(tabAB)
```

Das R-Paket stargazer

```
stargazer(attitude)
```

Tabelle :

Statistic	N	Mean	St. Dev.	Min	Max
rating	30	64.633	12.173	40	85
complaints	30	66.600	13.315	37	90
privileges	30	53.133	12.235	30	83
learning	30	56.367	11.737	34	75
raises	30	64.633	10.397	43	88
critical	30	74.767	9.895	49	92
advance	30	42.933	10.289	25	72

Mehr zu reproducible research mit R

- ▶ Das R-Paket `library(texreg)`
- ▶ Das R-Paket `library(R2wd)`
- ▶ <http://knutur.at/wsmt/slides/long-slides.pdf>
- ▶ R Task View zu reproducible research:

CRAN Task View: Reproducible Research

Maintainer: Max Kuhn

Contact: max.kuhn at pfizer.com

Version: 2014-01-20

Gliederung

Warum R nutzen

Modularer Aufbau von R

Anwendungsbeispiele

Schnittstellen

Quellen/Literatur

R lässt sich kombinieren...

Use R!

Richard M. Heiberger
Erich Neuwirth

R Through Excel

IBM SPSS Statistics Essentials for R: Project Web Hosting - Open Source Software

IBM SPSS Statistics Essentials for R

Open Source Software

Users

[Download IBM SPSS Statistics Essentials for R files](#)

[Donate money](#)

[Project detail and discuss](#)

[Get support](#)

Not what you're looking for?

R-Forge

rPython R package

SASmixed

Statistics and Computing

Robert A. Muenchen · Joseph M. Hilbe

R for Stata Users

Datenimport



SPSS-Dateien einlesen

```
midcar <- read.spss("MiD2008_PUF_Auto.sav",
                      to.data.frame = T,
                      use.value.labels =T)
```

```
midcar2 <- read.spss("MiD2008_PUF_Auto.sav",
                       to.data.frame = T,
                       use.value.labels =F)
```

Datenimport - Pakete

Daten von Datenbanken einlesen

RODBC, RMySQL, RPostgreSQL, RSQLite

Microsoft Excel Daten einlesen

XLConnect, xlsx

Man kann die Daten auch als .csv abspeichern und einlesen.

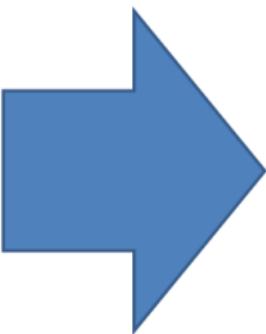
Sonstige Daten einlesen

foreign

SAS-Daten, SPSS-Daten und weitere Datenformate.

<https://support.rstudio.com/hc/en-us/articles/201057987-Quick-list-of-useful-R-packages>

Datenexport



SPSS®

STATA®

dBASE™

Minitab ▶



MySQL™

Paket - fremde Datensätze (foreign)

read.spss
write.dta
read.dta write.dbf
write.arff
read.epiinfo
read.ssd
read.mtp
write.foreign
read.octave
read.arff
read.xport

Überblick Daten Import/Export

R Data Import/Export

Version 3.1.0 (2014-04-10)

<http://cran.r-project.org/doc/manuals/r-release/R-data.pdf>

Datenmanagement wie in SPSS oder Stata

```
install.packages("Rz")
library(Rz)
```

Rz - dataset1

File Preferences Help

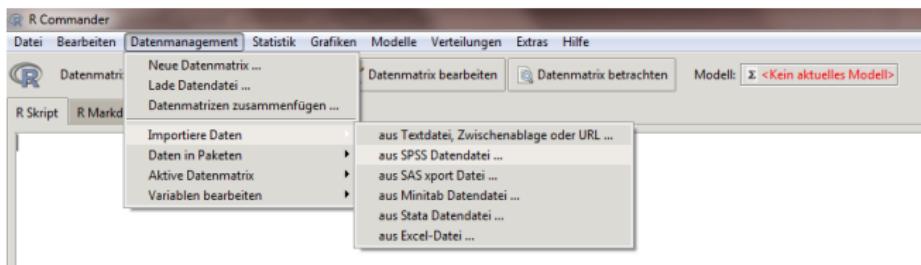
dataset1 (spss_cf_sohi98.sav)

Variables Descriptive statistics Plot

Meas: Names Labels Value Labels Missing

	Names	Labels	Value Labels	Missing
1	gebiet	Gebietsstand	1 "Früh. Bundes"	
2	traeger	Art des Trägers	1 "örtlich", 2 "üb."	
3	lfd_p_nr	Nummer der Person		
4	stellung	Stellung zum Haushaltsvorstand	1 "Haushaltsvorstand"	
5	geschl	Geschlecht	1 "männlich", 2 "	
6	geb_jahr	Geburtsjahr		
7	staat	Personengruppe	1 "Deutsche(r)",	
8	zuschl1	Mehrbedarfszuschlag, 1. Art	1 "an Personen",	
9	zuschl2	Mehrbedarfszuschlag, 2. Art	1 "an Personen",	
10	zuschl3	Mehrbedarfszuschlag, 3. Art	1 "an Personen",	
11	zuschl4	Mehrbedarfszuschlag, 4. Art	1 "an Personen",	
12	erv_stat	Erwerbsstatus	1 "vollzeiterwerb",	
13	schule	höchster allgemeinbildender Schulabschluss	1 "in schulischer",	
14	beruf	höchster Berufsausbildungsabschluss	1 "kein Ausbildung",	
15	dauer_e	bisherige Dauer der Arbeitslosigkeit		
16	lfd_hhnr	Lfd. Nr. des Haushalts		
17	personen	Anzahl der Personen in der Bedarfsgem.		
18	einricht	Lfd. HLU wird gewährt	1 "ausserhalb vo",	
19	bedarf	Bruttobedarf der Bedarfsgem. in DM/Monat		
20	miete	anerkannte Bruttokaltmiete in DM/Monat		
21	anspruc	Anspruch der Bedarfsgem. Jp. DM/Monat (W)		
22	soz_sit1	besondere soziale Situation, 1. Möglichkeit, 1 Tod eines Par		

Graphical User Interface (GUI) - vorhanden



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

Weitere Schnittstellen



Paket rJava



Paket RPy



Paket Rcpp

0xdata H2O

Apache Hadoop ist ein freies, in Java geschriebenes Framework für skalierbare, verteilt arbeitende Software.

http://de.wikipedia.org/wiki/Apache_Hadoop



Run H2O From Within R

With the REST API, it's simple to run H2O operations from within R using similar syntax to all your favorite R functions. In this post, we'll walk through a simple demo of its capabilities. First, get H2O installed and running by following the tutorial here. Once you have the R package loaded, you can take a look at the included demos by typing `demo(package="h2o")`, and run one of them by typing e.g., `demo(h2o.glm)`. We'll be stepping through a few of the basic statistical functions in this tutorial.

Starting up H2O in R

```
library(h2o)
localH2O = new("H2OClient", ip = "127.0.0.1", port = 54321)
h2o.checkClient(localH2O)
```

<http://0xdata.com/blog/2013/08/run-h2o-from-within-r/>



Written by
[Anqi Fu](#)



Aug 13 2013 [2 Comments](#)

Categories

[Uncategorized](#)

Tags

[0xdata Posts](#)

> [Next Post](#)

< [Previous Post](#)

[Back to all posts](#)

Gliederung

Warum R nutzen

Modularer Aufbau von R

Anwendungsbeispiele

Schnittstellen

Quellen/Literatur

Quellen für den Start

Schauen Sie auf:

A (very) short introduction to R

[http://cran.r-project.org/doc/contrib/Torfs+
Brauer-Short-R-Intro.pdf](http://cran.r-project.org/doc/contrib/Torfs+
Brauer-Short-R-Intro.pdf)

Das Paket mosaic



<http://mosaic-web.org/>

Datenmanipulation

`plyr` - Essential shortcuts for subsetting, summarizing, rearranging, and joining together data sets. `plyr` is the go to package for doing “groupwise” operations with your data.

`reshape2` - Tools for changing the layout of your data sets. Use the `melt` function to convert your data to long format, the layout R likes best.

`stringr` - Easy to learn tools for regular expressions and character strings.

`lubridate` - Tools that make working with dates and times easier.

<https://support.rstudio.com/hc/en-us/articles/201057987-Quick-list-of-useful-R-packages>

Wer benutzt R

- ▶ Who uses the R programming language and how do they use it?

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

Who-uses-the-R-programming-language-and-how-do-they-use-it

- ▶ Companies Using R

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

- ▶ How Google uses R to make online advertising more effective

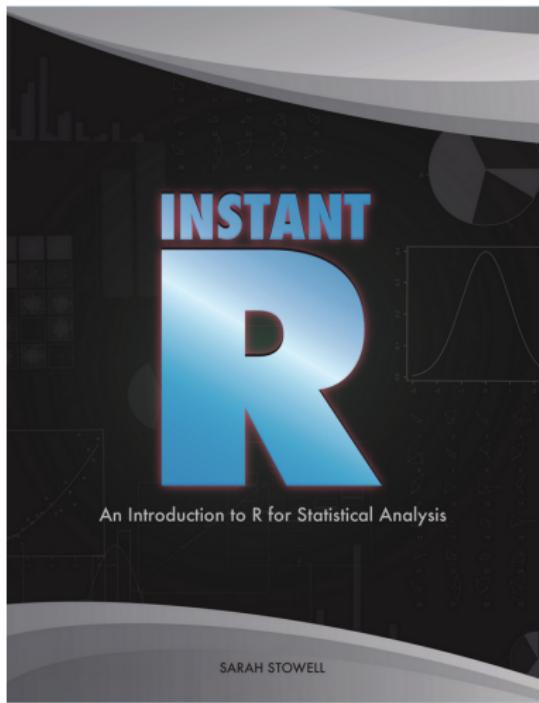
<http://blog.revolutionanalytics.com/2011/08/google-r-effective-ads.html>

Literatur



- ▶ Ligges, U. (2008):
Programmieren mit R.
Springer.
- ▶ Gut für Anfänger

Literatur



- ▶ Import und Export von Daten
- ▶ Daten editieren
- ▶ Graphiken

Literatur zu R

Books related to R

This page gives a partially annotated list of books that are related to S or R and may be useful to the R user community. See also the list of [other publications](#) related to R. An alternative searchable listing of both sets together is available [here](#).

- [1] Victor A. Bloomfield. *Using R for Numerical Analysis in Science and Engineering*. Chapman & Hall/CRC, 2014. ISBN 978-1439884485. [[bib](#) | <http://www.crcpress.com/product/isbn/9781439884485>]

Instead of presenting the standard theoretical treatments that underlie the various numerical methods used by scientists and engineers, Using R for Numerical Analysis in Science and Engineering shows how to use R and its add-on packages to obtain numerical solutions to the complex mathematical problems commonly faced by scientists and engineers. This practical guide to the capabilities of R demonstrates Monte Carlo, stochastic, deterministic, and other numerical methods through an abundance of worked examples and code, covering the solution of systems of linear algebraic equations and nonlinear equations as well as ordinary differential equations and partial differential equations. It not only shows how to use R's powerful graphic tools to construct the types of plots most useful in scientific and engineering work, but also:

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



- ▶ R project tutorial: reading Excel spreadsheets
<http://www.youtube.com/watch?v=Hq0JmSnBX8I>

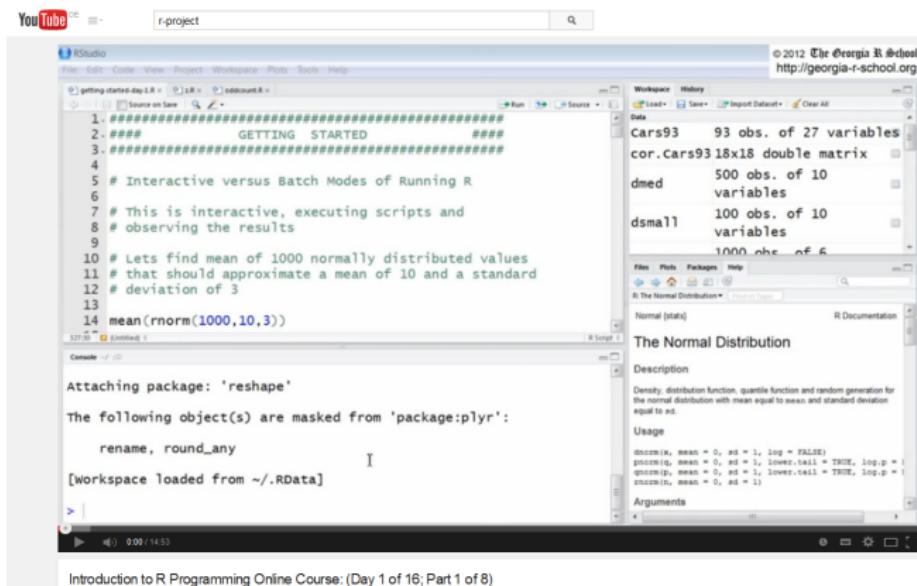
- ▶ Loading Data Into R Software -
(read.table, Data/CSV Import Tutorial)
<http://www.youtube.com/watch?v=VLtazaiYo-c>

Vielen Dank für die
Aufmerksamkeit

Jan-Philippe.Kolb@gesis.org

Anhang

Youtube-Videos



Introduction to R Programming Online Course: (Day 1 of 16; Part 1 of 8)

<http://www.youtube.com/watch?v=qHfSTRNg6jE>

Youtube Video - Graphiken

- ▶ R project tutorial: how to create and interpret a matrix scatter plot

<http://www.youtube.com/watch?v=kkhdB4dNg0>

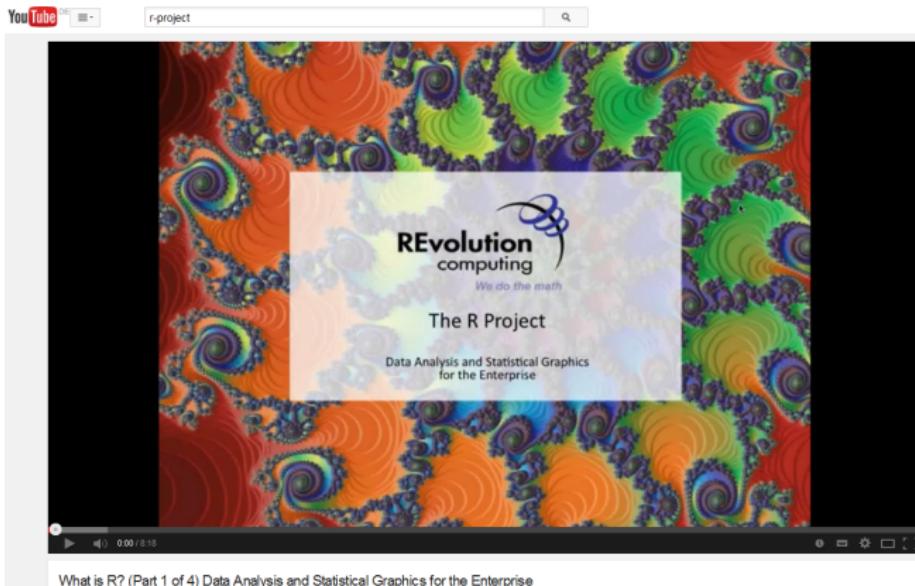
Youtube-videos

- ▶ Statistics with R (part 0: Download and Install R)
<http://www.youtube.com/watch?v=NfH5peM1RtI&list=PL9FE4E32D5F2AA0BA&index=3>
- ▶ R Tutorial 1 - Download, Installation, Setup - Statistical Programming Language R
<http://www.youtube.com/watch?v=ZoPJGmpYJzw>
- ▶ An Introduction to R - A Brief Tutorial for R Software for Statistical Analysis
<http://www.youtube.com/watch?v=LjuXiBjxryQ>
- ▶ R für Anfänger (4): Daten einlesen, analysieren, plotten für Eilige (German)
<http://www.youtube.com/watch?v=0ADUACud99k>
- ▶ R project tutorial: simple linear regression
<http://www.youtube.com/watch?v=ZGTB0hbahmY>

Youtube-videos

- ▶ R Tutorial 3 - More about Vectors - Statistical Programming Language R
<http://www.youtube.com/watch?v=hLozrMhpWuY>
- ▶ Twitter Sentiment Analysis
http://www.youtube.com/watch?v=adIvt_lu01o

Youtube-videos



What is R? (Part 1 of 4) Data Analysis and Statistical Graphics for the Enterprise

http://www.youtube.com/watch?v=M2u7kbcXI_k