

# Nichtlineare Effekte in der linearen Regression

Die wichtigsten Werkzeuge für den Kurs

Jan-Philipp Kolb

Freitag, 20.06.2014



# Inhalt

R kam, sah und blieb

Wie bekommt man Hilfe?

Modularer Aufbau von R

Literatur

# Gliederung

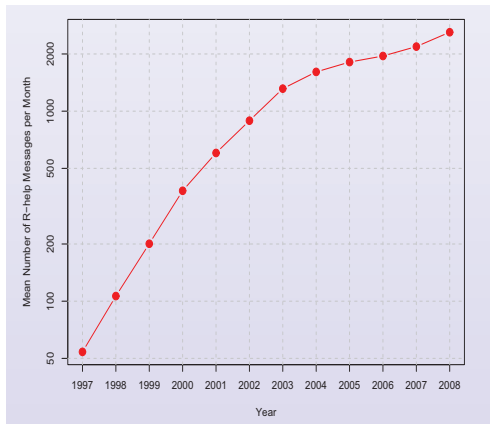
R kam, sah und blieb

Wie bekommt man Hilfe?

Modularer Aufbau von R

Literatur

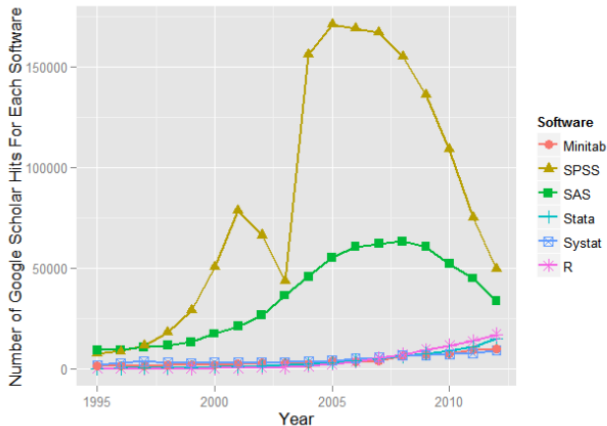
## Anzahl der Fragen in Hilfeforen zu R



## Warum R nutzen?

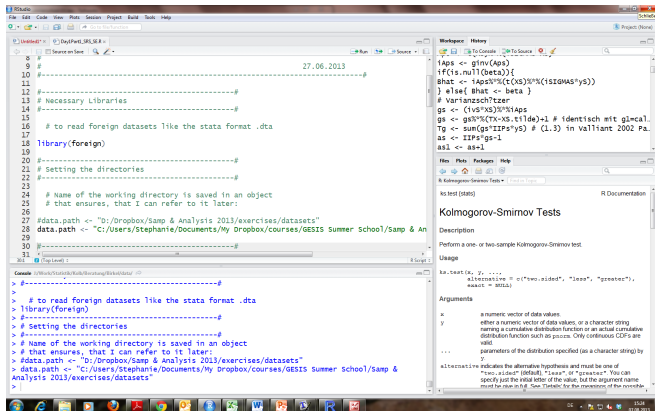
- ▶ Als Weg kreativ zu sein ...
- ▶ Graphiken, Graphiken, Graphiken
- ▶ In Kombination mit anderen Programmen nutzbar
- ▶ Zur Verbindung von Datenstrukturen
- ▶ Zum Automatisieren
- ▶ Um die Intelligenz anderer Leute zu nutzen ;-)
- ▶ ...

## Popularität von R



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

# Ich nutze Rstudio!



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

# Ein Einführungskurs in R

`https://github.com/Japhilko/IntroR\_Mannheim2014.git`

The screenshot shows the GitHub interface for the repository 'IntroR\_Mannheim2014' by user 'Japhilko'. At the top, there's a navigation bar with the GitHub logo, a search bar, and links for 'Explore', 'Gist', 'Blog', and 'Help'. Below this, the repository name is displayed with a 'PUBLIC' label. To the right of the repository name are buttons for 'Unwatch' (with a dropdown arrow), 'Star' (with a count of 0), and 'Fork' (with a count of 0). Below the repository name, there's a breadcrumb trail: 'branch: master' > 'IntroR\_Mannheim2014' > 'slides' > 'GESIS\_R\_Kurs\_2014\_Teil1.pdf'. The main content area shows a commit by 'Japhilko' from '19 minutes ago' with the message 'Upload Slides'. Below the commit message, it says '1 contributor'. At the bottom, there's a file entry for 'file' with a size of '9935.143 kb'. To the right of the file entry are buttons for 'Open', 'Raw', 'History', and 'Delete'. On the far right, there's a vertical sidebar with icons for code, issues, pull requests, and a file explorer.



# Die Unterlagen für diesen Kurs

<https://github.com/Japhilko/npRegression>

The screenshot shows the GitHub repository page for **Japhilko / npRegression**. The repository is public and has 5 commits, 1 branch (master), 0 releases, and 1 contributor. The main content area displays a list of files and commits:

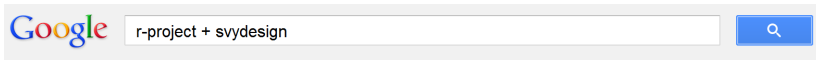
File	Commit Message	Time Ago
data	upload data	13 minutes ago
rcode	Upload Files	31 minutes ago
slides	Upload Files	31 minutes ago
.gitattributes	Revert "Revert "Start""	37 minutes ago
.gitignore	Revert "Revert "Start""	37 minutes ago

Below the file list, there is a blue banner recommending to add a README to the repository, with a button labeled "Add a README".

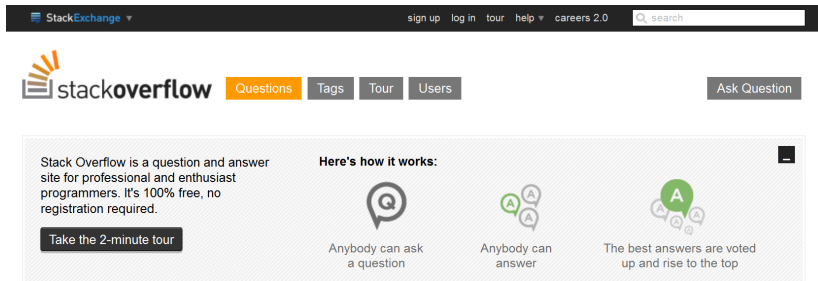
On the right side, there is a sidebar with links to various repository features: Code, Issues (0), Pull Requests (0), Wiki, Pulse, Graphs, Network, and Settings. At the bottom of the sidebar, there is a section for cloning the repository via HTTPS, with the URL <https://github.com/Japhilko/npRegression>.

## 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!



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

## 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](#).

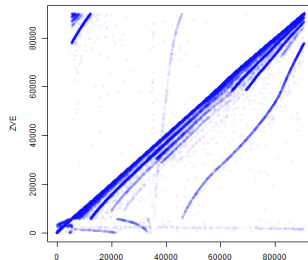
- [1] Victor A. Bloomfield. *Using R for Numerical Analysis in Science and Engineering*. Chapman & Hall/CRC, 2014. ISBN 978-1439884485. [ [bib](#) | <http://www.crcpress.com/product/isbn/9781439884485> ]

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:

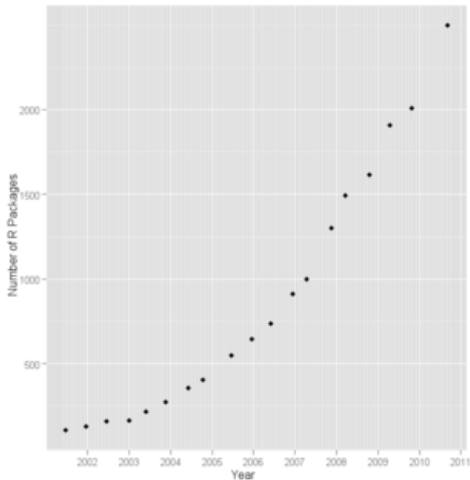
<http://www.r-project.org/doc/bib/R-books.html>

## Warum nicht **ein** Datensatz - mit dem die ganze Zeit gearbeitet wird?

- ▶ Viele Datensätze um zu zeigen, was mit R alles möglich ist
- ▶ Jeder Datensatz ist anders und es erfordert Zeit sich einzuarbeiten
- ▶ Für Scientific-Use-Files muss extra ein Antrag gestellt werden
- ▶ Bei Public-Use-Files gehen oft sinnvolle Zusammenhänge verloren



## Modularer Aufbau von R



<http://blog.revolutionanalytics.com/2010/09/what-can-other-languages-learn-from-r.html>





## Wichtige Bibliotheken

Bibliothek	Thema
foreign	Functions for reading and writing data stored by statistical packages
sampling	Functions for drawing and calibrating samples.
survey	Analysis of complex survey samples
MASS	Functions and Datasets for Venables and Ripley's Modern Applied Statistics with S'

## Weitere nützliche Bibliotheken

Bibliothek	Thema
xtable	Coerce data to LaTeX and HTML tables
dummies	Expands factors, characters and other eligible classes into dummy/indicator variables.
mvtnorm	Multivariate Normal and t Distributions
maptools	Tools for reading and handling spatial objects

## Aufgabe A1 - Interessante Zusatzpakete

Gehen Sie auf `cran.r-project.org` und suchen Sie in dem Bereich, wo die Pakete vorgestellt werden, nach Paketen,...

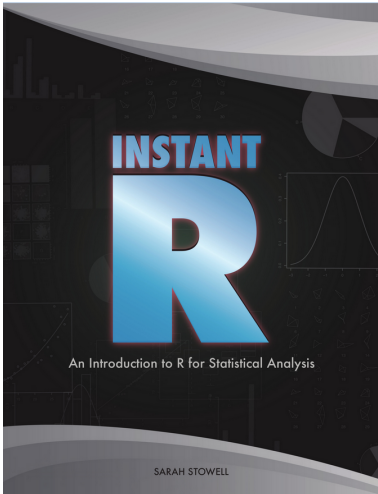
1. die für die deskriptive Datenanalyse geeignet sind.
2. um fremde Datensätze einzulesen (z.B. SPSS-Daten)
3. für Survival-Analysen geeignet sind
4. Splines und andere Glättungsfunktionen berechnen
5. das Lasso und Elastic Net berechnen

## Literatur



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

## Literatur



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

## Aufgabe A2 - Zuweisungen und Funktionen

Erzeugen Sie einen Vektor  $b$  mit den Zahlen von 1 bis 5 und berechnen Sie...

1. den Mittelwert
2. die Varianz
3. die Standardabweichung
4. die quadratische Wurzel aus dem Mittelwert