

# A1 Erste Schritte mit R

Jan-Philipp Kolb

15 Oktober 2018

# Disclaimer/ Informationen vorab

Normalerweise gibt es große Unterschiede in den Kenntnissen und Fähigkeiten der Teilnehmer - bitte gebt Bescheid, wenn es zu schnell oder zu langsam geht oder etwas unklar geblieben ist.

- Wenn es Fragen gibt - immer fragen
- In diesem Kurs gibt es viele **Übungen**, denn das Programmieren lernt man am Ende nur allein.
- Ich habe viele **Beispiele** - probiert sie aus
- R macht mehr Spaß zusammen - arbeitet zusammen!

# Gründe R zu nutzen...

- ... R ist eine **quelloffene Sprache**
- ... hervorragende **Grafiken, Grafiken, Grafiken**
- ... **R kann in Kombination mit anderen Programmen verwendet werden** - z.B. zur **Verknüpfung von Daten**
- ... R kann **zur Automatisierung** verwendet werden
- ... Breite und aktive Community - **Man kann die Intelligenz anderer Leute nutzen ;-)**

# R kann in Kombination mit anderen Programmen genutzt werden...

Use R!

Richard M. Heiberger  
Erich Neuwirth

**R Through Excel**



SASmixed



rPython R package

Statistics and Computing

Robert A. Muenchen · Joseph M. Hilbe

**R for  
Stata Users**

[IBM SPSS Statistics Essentials for R: Project Web Hosting - Open Source Software](#)

**IBM SPSS Statistics Essentials for R**

**Users**

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

[Donate money](#)

[Project detail and discuss](#)

[Get support](#)

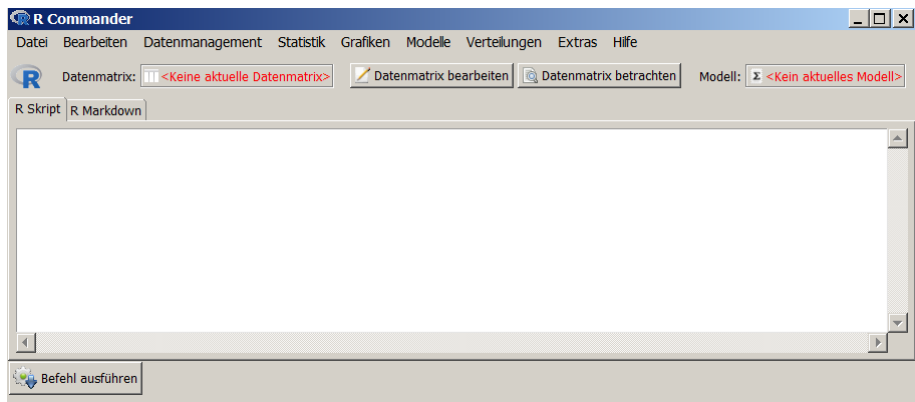
Not what you're looking for?

Open Source Software

# R für SPSS Nutzer

## Bob Muenchen - R for SPSS and SAS Users

- R commander (Rcmdr)



# R sollte genutzt werden, weil andere Programme Fehler provozieren:

## FAQ: Reinhart, Rogoff, and the Excel Error That Changed History

By Peter Coy



PHOTOGRAPH BY GREGOR SCHUSTER

# Die Beliebtheit von R-Paketen

The screenshot displays the CRAN.rstudio.com dashboard with the following components:

- Navigation Bar:** Includes links for "Dashboard", "Recent Downloads", and "Source Code".
- Main Metrics Section:**
  - Downloads per sec (last 5 min):** 9.8 (represented by an orange box with a bar chart icon).
  - Total downloads:** 972 (represented by a blue box with a download arrow icon).
  - Unique users:** 650 (represented by a blue box with a group of people icon).
- Warning Section:** A slider control labeled "Warn when rate exceeds:" with a value set at 9.8.
- Recent downloads window:** A section showing a count of 50 recent downloads.
- Popularity by package (last 5 min):** A bubble chart where the size of each bubble corresponds to the number of downloads for a specific package. The most prominent packages are digest, jsonlite, stringr, and curl.
- Percent of downloads (last 5 min):** A table listing the top packages by percentage of total downloads.

Package	Percent
1 digest	1.9
2 jsonlite	1.6
3 stringr	1.5
4 curl	1.4
5 plyr	1.4
6 bitops	1.3
7 mime	1.3
8 R6	1.3
9 stringi	1.3
10 tibble	1.3
11 DBI	1.2

# Download R:

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



*CRAN*  
[Mirrors](#)  
[What's new?](#)  
[Task Views](#)  
[Search](#)

*About R*  
[R Homepage](#)  
[The R Journal](#)

*Software*  
[R Sources](#)  
[R Binaries](#)  
[Packages](#)  
[Other](#)

## The Comprehensive R Archive Network

### Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

### Source Code for all Platforms

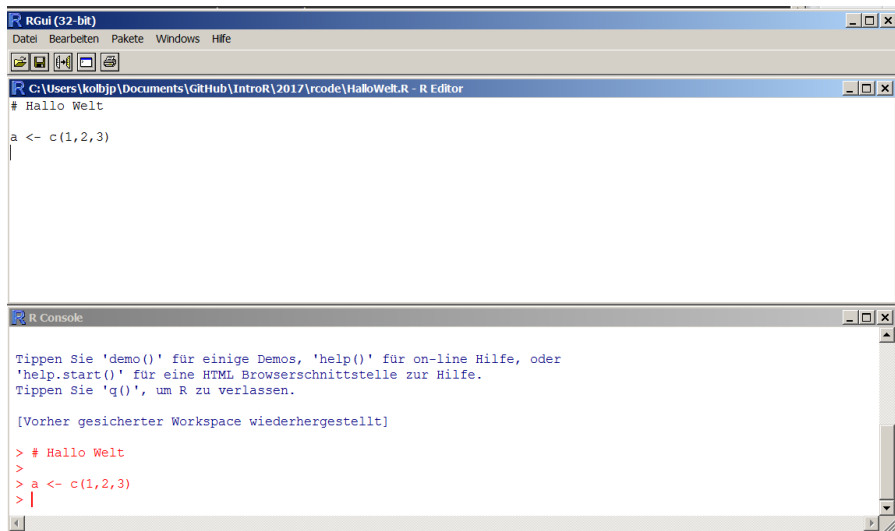
Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (Friday 2017-04-21, You Stupid Darkness)  
[R-3.4.0.tar.gz](#), read [what's new](#) in the latest version.



# Open Source Programm R

## Das ist das Basis-R:



The screenshot displays the R GUI (32-bit) interface. The top window, titled "R GUI (32-bit)", contains a menu bar with "Datei", "Bearbeiten", "Pakete", "Windows", and "Hilfe", and a toolbar with icons for file operations. Below this is the "R Editor" window, which shows the file path "C:\Users\kolbjp\Documents\GitHub\IntroR\2017\rcode\HalloWelt.R - R Editor" and the following R code:

```
# Hallo Welt  
  
a <- c(1,2,3)
```

The bottom window, titled "R Console", displays the following text:

```
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]  
  
> # Hallo Welt  
>  
> a <- c(1,2,3)  
> |
```

# Graphical user interface

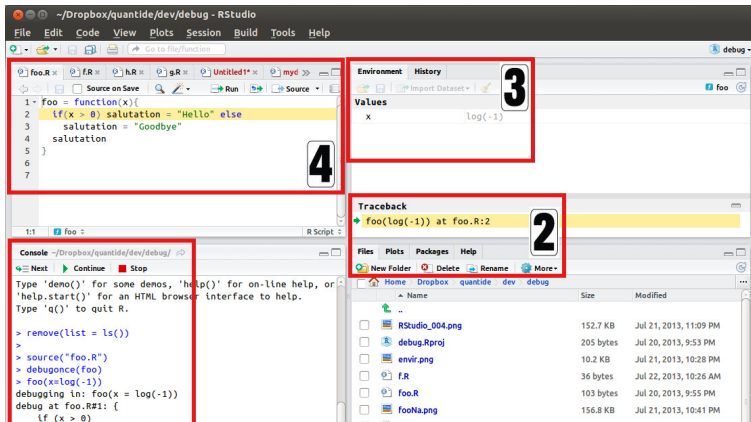
Viele Leute nutzen ein **Graphical User Interface** (GUI) oder ein **Integrated Development Interface** (IDE).

Aus den folgenden Gründen:

- Syntax-Hervorhebung
- Auto-Vervollständigung
- Bessere Übersicht über Graphiken, Pakete, Dateien, ...

# Various text editors / IDEs

- **Gedit** with R-specific Add-ons for Linux
- **Emacs** and **ESS** (Emacs speaks statistics)- An extensible, customizable, free/libre text editor — and more.
- I use **Rstudio**!



# RStudio

The screenshot displays the RStudio IDE interface. The main editor window shows an R script with the following content:

```
42 c("A1 Getting started", "A2 How to get help", "A3 Data import", "A4 The GESIS panel data", "A5 data export")
43
44 c("B1 Basic data analysis", "B2 The use of the survey package", "B3 Graphics", "B4 linear regression", "B5 logistic regression")
45
46
47 c("C1 Understanding error messages", "C2 Hierarchical/Multilevel models")
48
49
50 ## {r,echo=F}
51 library(knitr)
52 sched <- xlsx::read.xlsx2("../orga/schedule.xlsx", 1)
53 data <- sched[1:7, c(1, 4, 5)]
54 data <- data[which(data$Part == "Break"), ]
55
56 kable(data, row.names = F)
57
58
59
60 ## Overview - advantages of R
61
```

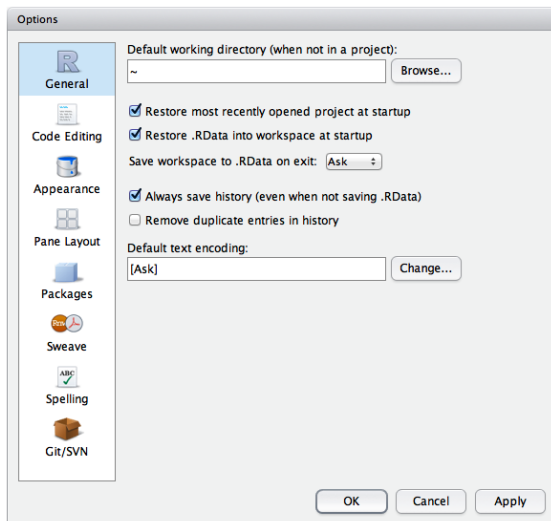
The Environment pane on the right shows the loaded packages: `misc/addin.general.Rmd`, `misc/psw_pres.Rmd`, `misc/psw_pres.pdf`, `slides/A1_GettingStarted.Rmd`, `slides/A1_GettingStarted.Rmd.pdf`, `slides/A1_GettingStarted.Rmd`, `slides/A2_How2gethelp.pdf`, and `slides/A2_How2gethelp.pdf`.

The File Explorer pane on the right shows the file structure of the project, including folders like `slides` and files like `A1_GettingStarted.html`, `A1_GettingStarted.Rmd`, `A1_GettingStarted.pdf`, `A1_GettingStarted.tex`, `A2_How2gethelp.html`, `A2_How2gethelp.Rmd`, `A2_How2gethelp.pdf`, `A2_How2gethelp.tex`, `A3_GESISPanel.pdf`, `A3_GESISPanel.Rmd`, `A3_GESISPanel.tex`, `A3_GESISPanel.cache`, `A3_GESISPanel_files`, `A4_DataExport.pdf`, `A4_DataExport.Rmd`, `A4_GESISPanel.cache`, `A4_GESISPanel_files`, `A5_DataExport.pdf`, `A5_DataExport.Rmd`, `B1_DataProcessing.html`, and `B1_DataProcessing.Rmd`.

The Console pane at the bottom shows the output of the script, including the knitr chunk output and the final output file created: `A2_How2gethelp.pdf`.

# Customizing RStudio

## Six reasons to use Rstudio.



# A1A Exercise - Preparation

- Check if R is installed on your computer.
- If not, download **R** and install it.
- Check if Rstudio is installed.
- If not - **install** Rstudio.
- Start RStudio. Go to the console (lower left window) and write

```
3+2
```

- If there is not already an editor open in the upper left window, then go to the file menu and open a new script. Check the date with `date()` and the R version with `sessionInfo()`.

```
date()
```

```
sessionInfo()
```

# R ist eine objektorientierte Sprache.

## Vektoren und Zuweisungen

- <- ist der Zuweisungsoperator

```
b <- c(1,2) # create an object with the numbers 1 and 2
```

- Auf dieses Objekt kann eine Funktion angewendet werden:

```
mean(b) # computes the mean
```

```
## [1] 1.5
```

Mit den folgenden Funktionen können wir etwas über die Eigenschaften des Objekts erfahren:

```
length(b) # b has the length 2
```

# Funktionen in base-Paket

Function	Meaning	Example
<code>str()</code>	Object structure	<code>str(b)</code>
<code>max()</code>	Maximum	<code>max(b)</code>
<code>min()</code>	Minimum	<code>min(b)</code>
<code>sd()</code>	Standard deviation	<code>sd(b)</code>
<code>var()</code>	Variance	<code>var(b)</code>
<code>mean()</code>	Mean	<code>mean(b)</code>
<code>median()</code>	Median	<code>median(b)</code>

Diese Funktionen benötigen nur ein Argument.



# Funktionen mit mehr Argumenten

Andere Funktionen benötigen mehr Argumente:

Argument	Bedeutung	Beispiel
quantile()	90 % Quantile	quantile(b,.9)
sample()	Draw a sample	sample(b,1)

```
quantile(b,.9)
```

```
## 90%
```

```
## 1.9
```

```
sample(b,1)
```

```
## [1] 1
```

# Überblick Funktionen

<http://cran.r-project.org/doc/manuals/R-intro.html>

## An Introduction to R

### Table of Contents

#### [Preface](#)

#### [1 Introduction and preliminaries](#)

##### [1.1 The R environment](#)

##### [1.2 Related software and documentation](#)

##### [1.3 R and statistics](#)

##### [1.4 R and the window system](#)

##### [1.5 Using R interactively](#)

##### [1.6 An introductory session](#)

##### [1.7 Getting help with functions and features](#)

##### [1.8 R commands, case sensitivity, etc.](#)

##### [1.9 Recall and correction of previous commands](#)

##### [1.10 Executing commands from or diverting output to a file](#)

##### [1.11 Data permanency and removing objects](#)

# A1B Übung - Zuweisungen und Funktionen

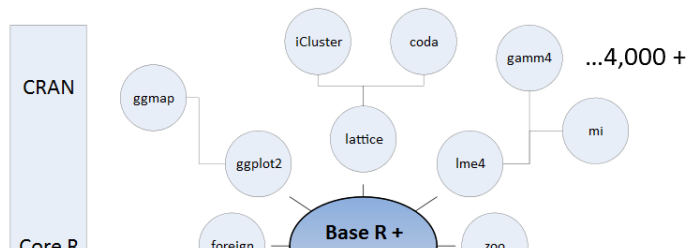
Erstellen Sie einen Vektor `b` mit den Zahlen von 1 bis 5 und berechnen Sie. . . .

- 1 den Mittelwert
- 2 die Varianz
- 3 die Standardabweichung
- 4 die Quadratwurzel aus dem Mittelwert

# Wo man Routinen findet

- Viele Funktionen sind in Basis-R enthalten.
- Viele spezifische Funktionen sind in zusätzliche Bibliotheken integriert.
- R kann modular durch sogenannte Pakete oder Bibliotheken erweitert werden.
- Die wichtigsten Pakete, die auf CRAN gehostet werden (13087 at Do Okt 04)
- Weitere Pakete finden Sie z.B. unter **Bioleiter**

## Übersicht R-Pakete



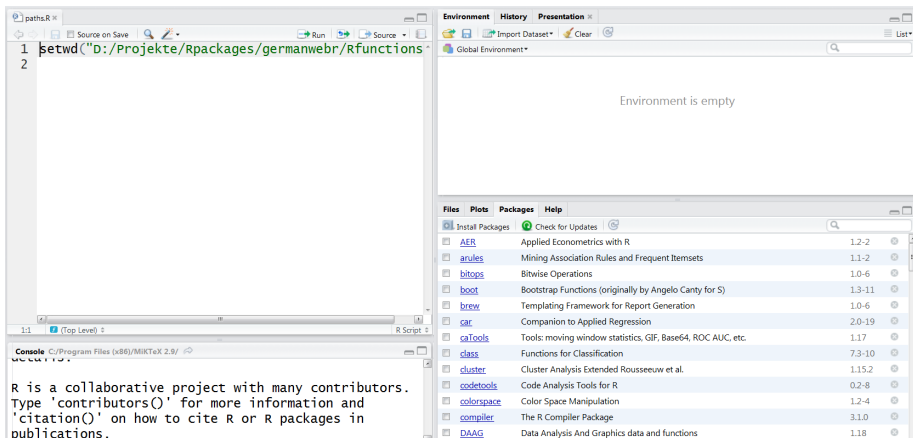
# Installation von Paketen

- Die Anführungszeichen um den Paketnamen herum sind für den Befehl `install.packages` notwendig.
- Sie sind optional für den Befehl `library`.
- Man kann auch `require` anstelle von `library` verwenden.

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

```
library(lme4)
```

# Installation von Paketen mit RStudio



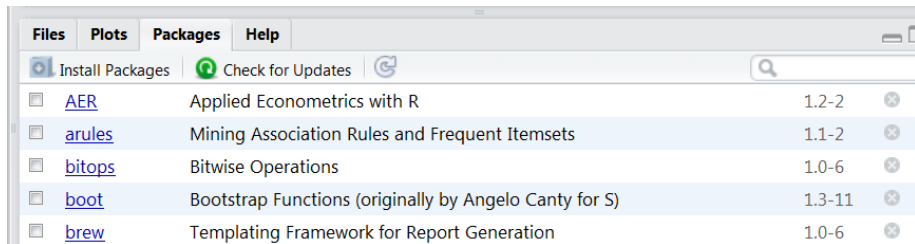
The screenshot shows the RStudio interface with the following components:

- Script Editor:** Contains the R code `setwd("D:/Projekte/Rpackages/germanwebr/Rfunctions")`.
- Environment Pane:** Displays "Global Environment\*" and "Environment is empty".
- Packages Pane:** Shows a list of installed and available packages. The installed packages are highlighted in blue.

Package	Description	Version
AER	Applied Econometrics with R	1.2-2
arules	Mining Association Rules and Frequent Itemsets	1.1-2
bitops	Bitwise Operations	1.0-6
boot	Bootstrap Functions (originally by Angelo Canty for S)	1.3-11
brew	Templating Framework for Report Generation	1.0-6
car	Companion to Applied Regression	2.0-19
caTools	Tools: moving window statistics, GIF, Base64, ROC AUC, etc.	1.17
class	Functions for Classification	7.3-10
cluster	Cluster Analysis Extended Rousseeuw et al.	1.15.2
codetools	Code Analysis Tools for R	0.2-8
colorspace	Color Space Manipulation	1.2-4
compiler	The R Compiler Package	3.1.0
DAAG	Data Analysis And Graphics data and functions	1.18

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.

# Bestehende Pakete und Installation



# Übersicht über viele nützliche Pakete:

- Luhmann - **Table with many useful packages**

## Weitere interessante Pakete:

- Paket für Import/Export - `foreign`
- **sampling-Paket für die Stichprobenziehung**
- `xtable` Paket zur Integration von LaTeX in R (**xtable Galerie**)
- `dummies` - **Paket zur Erstellung von Dummies**
- **Paket `mvtnorm` um eine multivariate Normalverteilung zu erhalten.**
- **Paket `maptools` um Karten zu erzeugen**



# Pakete aus verschiedenen Quellen installieren

## Pakete vom CRAN Server installieren

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

## Pakete vom Bioconductor Server installieren

```
source("https://bioconductor.org/biocLite.R")  
biocLite(c("GenomicFeatures", "AnnotationDbi"))
```

## Pakete von Github installieren

```
install.packages("devtools")  
library(devtools)  
  
install_github("hadley/ggplot2")
```

# Wie bekomme ich einen Überblick?

- Entdecke Pakete, die kürzlich auf den **CRAN** Server hochgeladen wurden
- Nutze eine Shiny Web-App, die **Pakete anzeigt, die kürzlich von CRAN** heruntergeladen wurden.
- Werfe einen Blick auf eine **Quick-Liste nützlicher Pakete**
- . . . . , oder auf eine Liste mit den **besten Paketen für die Datenverarbeitung und -analyse**, . . . . .
- . . . . , oder schaue unter **die 50 meistgenutzten Pakete**

# CRAN Task Views

- Bezüglich mancher Themen gibt es einen Überblick über alle wichtigen Pakete - (**CRAN Task Views**)
- Momentan gibt es 35 Task Views.
- Alle Pakete einer Task-View können mit folgendem Befehl installiert werden: **command:**

```
install.packages("ctv")  
library("ctv")  
install.views("Bayesian")
```

## CRAN Task Views

<a href="#">Bayesian</a>	Bayesian Inference
<a href="#">ChemPhys</a>	Chemometrics and Computational Physics
<a href="#">ClinicalTrials</a>	Clinical Trial Design, Monitoring, and Analysis
<a href="#">Cluster</a>	Cluster Analysis & Finite Mixture Models
<a href="#">DifferentialEquations</a>	Differential Equations
<a href="#">Distributions</a>	Probability Distributions
<a href="#">Econometrics</a>	Econometrics
<a href="#">Environmetrics</a>	Analysis of Ecological and Environmental Data

# A1C Übung - zusätzliche Pakete

Geht auf <https://cran.r-project.org/> und sucht nach Paketen...

- die sich für die deskriptive Datenanalyse eignen.
- mit denen man fremde Datensätze einlesen kann (z.B. SPSS data)
- mit denen man Lasso Regressionen rechnen kann
- mit denen man große Datenmengen bearbeiten kann

## Links zum Weiterlesen:

- **Warum man R zuerst lernen sollte wenn man Data Science machen möchte**
- RStudio hat den **Infoworld 2015 Technology of the Year Award**. bekommen
- **Warum R gut für Unternehmen ist**
- Schaut auf **R-bloggers**
- Vergleich zwischen **python** und **R**
- R und Stata - **Side-by-side**
- **AWESOME R**
- **1000 R tutorials/Links**
- **Zwei Minuten Videos auf Github**

# Shiny App - Einführung in R

<http://www.intro-stats.com/>

