

# Intro - Erste Schritte

Jan-Philipp Kolb

08 Mai 2017

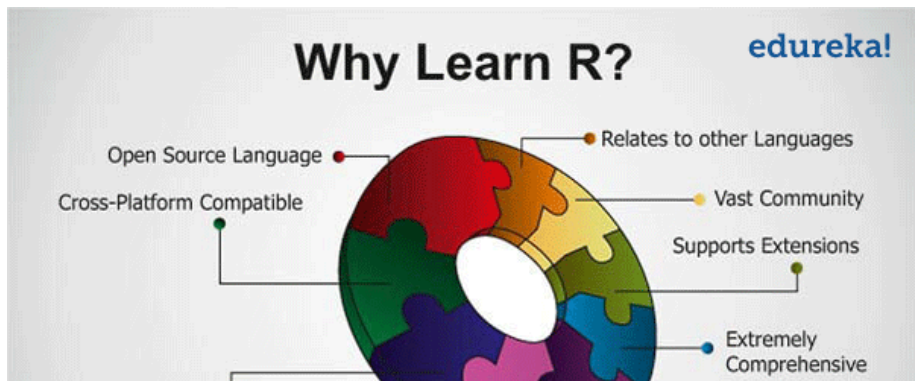


# Pluspunkte von R

- 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 ;-)
- ...

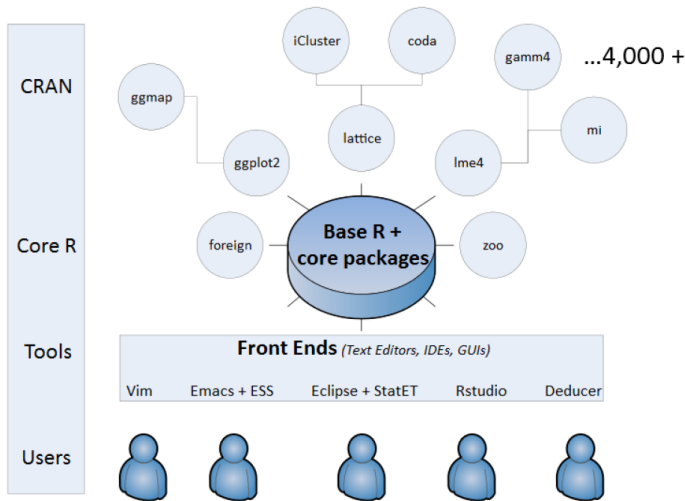
# Gründe

- R ist frei verfügbar. Es kann umsonst heruntergeladen werden.
- R ist eine Skriptsprache
- Gute Möglichkeiten für die Visualisierung
- R wird immer populärer
- Popularität von R ist in vielen Bereichen sehr hoch.

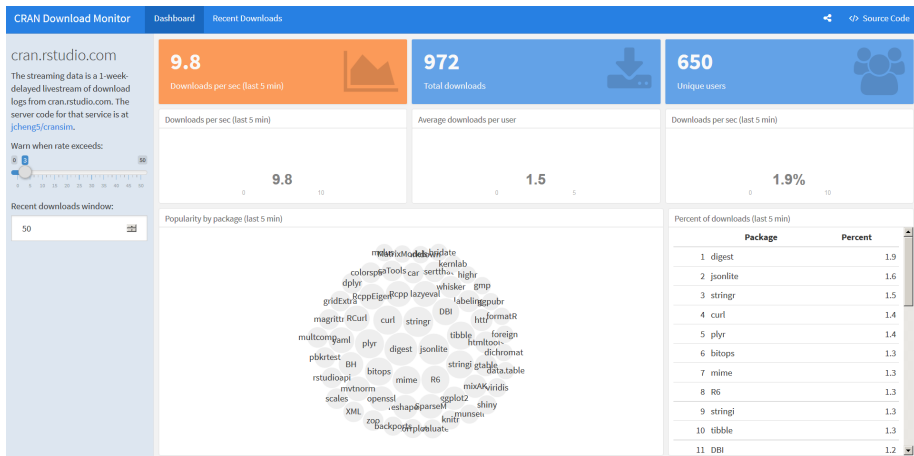


edureka!

# Warum R?



## Die am meisten heruntergeladenen Pakete



# CRAN Task Views

## 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
<a href="#">ExperimentalDesign</a>	Design of Experiments (DoE) & Analysis of Experimental Data
<a href="#">ExtremeValue</a>	Extreme Value Analysis
<a href="#">Finance</a>	Empirical Finance

# Motivation - Nachteile von R

- ① Daten werden oft anderswo erfasst/eingegeben (oft Excel, SPSS etc.)
- ② Nicht jeder ist bereit mit R zu arbeiten
- ③ Nicht auf jedem Rechner ist R installiert
- ④ R ist manchmal zu langsam
- ⑤ Schwierigkeiten bei der Arbeit mit großen Datenmengen



# Was folgt daraus

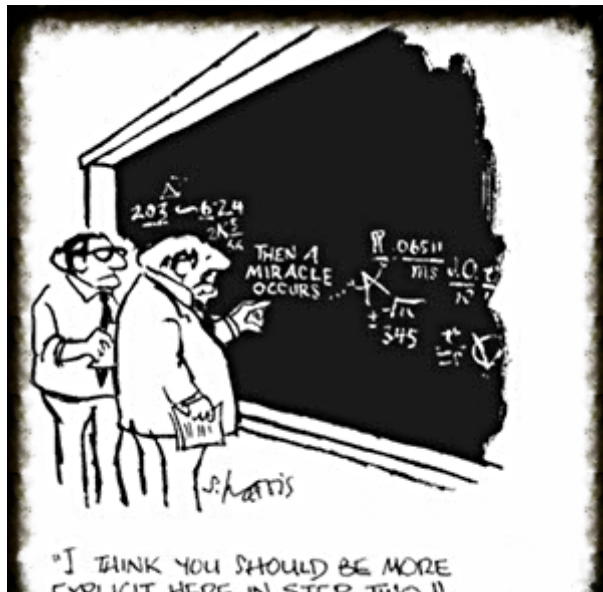
- ① Schnittstelle zu SPSS/Stata/Excel zum Import von Daten
- ② Schnittstelle zu Word/LaTeX
- ③ Möglichkeit HTML Präsentationen zu erzeugen
- ④ Nutzung von C++
- ⑤ Nutzung von Datenbanken

# Die Nutzung von Schnittstellen beim Import/Export

- Interaktion mit Excel, SPSS, Stata, ...



# Reproducible Research



# Was wird bei Wikipedia unter Reproducibility verstanden?



# Darstellung von Ergebnissen

- Mit der Schnittstelle zu Javascript lassen sich interaktive Graphiken erzeugen
- Diese kann man auf Websites, in Präsentationen oder in Dashboards verwenden



basic usage



random nodes



images



shapes



social network



groups



selections



mobile friendly



sizing



multiline text



custom style



scalable images



# Warum die Schnittstelle zu C++?

- Wenn Schnelligkeit wichtig ist, bietet sich C++ an.
- Dies kann bspw. der Fall sein, wenn sich Schleifen nicht vermeiden lassen.
- Man wird bei der Programmierung durch RStudio unterstützt.
- Es gibt eine Rcpp Galerie, wo man sich Anregungen holen kann.
- Allerdings sollte man zunächst versuchen den Rcode so schnell wie möglich zu gestalten.

# Die Nutzung von Datenbanken

- Wenn große Datenmengen anfallen, kann die zusätzliche Nutzung von Datenbanken sinnvoll sein.
- In R bestehen Schnittstellen sowohl zu relationalen als auch nicht relationalen Datenbanken.
- Datenbanken sollten allerdings erst genutzt werden, wenn alle Möglichkeiten in R ausgeschöpft sind.

# Nutzung der Unterlagen auf GitHub

- Die folgende Seite ist die Startseite für den Kurs:

<https://japhilko.github.io/Interfaces4R/>

The logo for 'Interfaces4R' features the text 'Interfaces4R' in a white, handwritten-style font. It is set against a blue rectangular background that has a subtle grid pattern and small white star-like specks, reminiscent of a night sky or a technical grid.

// Schnittstellen für den Austausch von Daten

- Einführung und Motivation (pdf)
- Datenimport (pdf, code)



# Wo sind die Sourcecodes?

Wie wird das Github Verzeichnis genutzt?

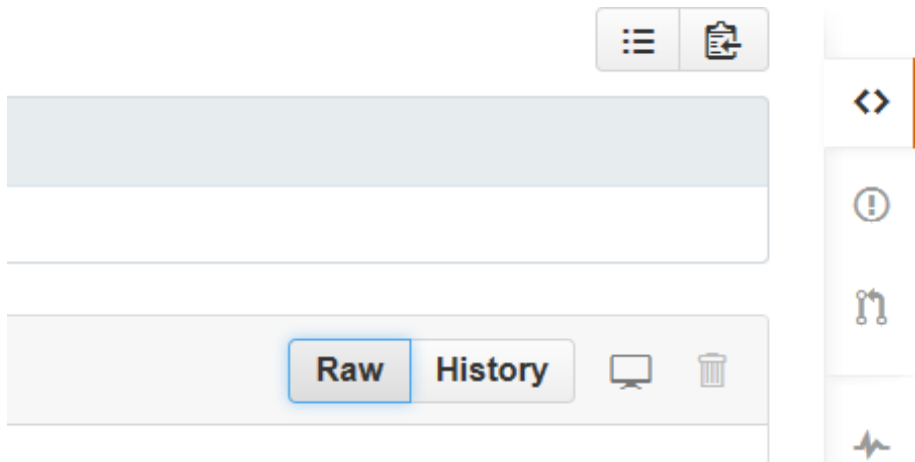
- Auf der folgenden Seite sind alle Sourcecodes enthalten:

`https://github.com/Japhilko/RInterfaces`

- Es lohnt sich immer wieder zu dieser Seite zurückzukehren, weil auch hier alle relevanten Dokumente verlinkt sind.
- Grundsätzlich kann man der Veranstaltung am Besten mit den kompletten File oder der kompletten Browserversion eines Kapitels (sind unter den Kapitelüberschriften verlinkt) folgen. Wenn Teile heruntergeladen werden sollen, bietet es sich an, das entsprechende pdf herunterzuladen.
- Falls Links ins Leere führen - bitte Bescheid sagen.

## Informationen ausdrucken

- Zum Ausdrucken eignen sich die pdf-Dateien besser.
- Diese können mit dem Raw Button heruntergeladen werden.



## Weitere Dateien herunterladen

- Begleitend zu den Folien wird meistens auch ein R-File angeboten.
- Hier können Sie entweder das gesamte R-File herunterladen und in R ausführen oder einzelne Befehle per Copy/Paste übernehmen.
- Vereinzelt sind auch Datensätze vorhanden.
- .csv Dateien können direkt von R eingelesen werden (wie das geht werde ich noch zeigen).
- Wenn die .csv Dateien heruntergeladen werden sollen auch den Raw Button verwenden.
- Alle anderen Dateien (bspw. .RData) auch mittels Raw Button herunterladen.

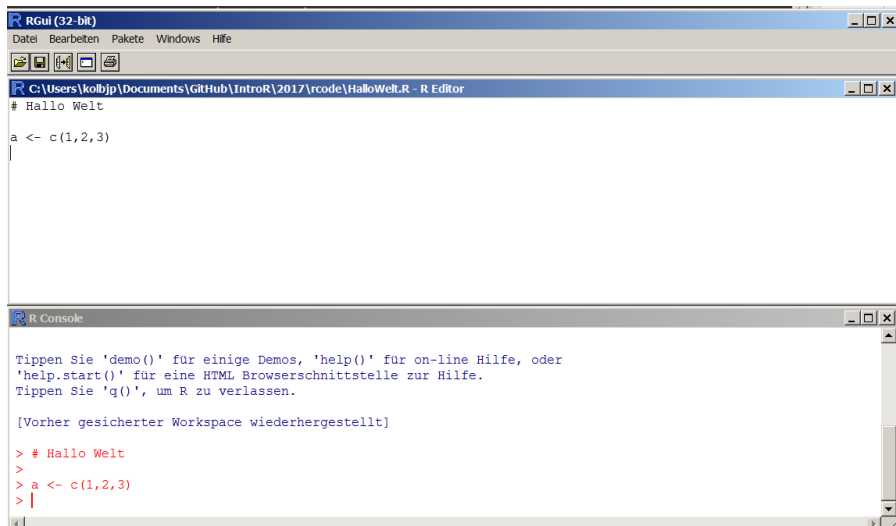
- Zusätzlich gibt es in jedem Kapitel eine oder mehrere Aufgabe(n), da man nur durch eigenes Trainieren auf der Lernkurve vorankommt.
- Die Quellen für die Punkte auf den Folien sind als Link meist in der Überschrift hinterlegt.
- Die Links sind nur im HTML Dokument zu sehen aber auch in der pdf vorhanden.

Wen Github näher interessiert:

- Hello World
- Understanding the GitHub flow

# Basis R ...

- Wenn man nur R herunterlädt und installiert, sieht das so aus:



The screenshot displays the RGui (32-bit) application window. The top menu bar includes 'Datei', 'Bearbeiten', 'Pakete', 'Windows', and 'Hilfe'. Below the menu is a toolbar with icons for file operations. The main window is titled 'C:\Users\kolbjp\Documents\GitHub\IntroR\2017\rcode\HalloWelt.R - R Editor' and contains the following R code:

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

Below the editor is the 'R Console' window, which shows the execution of the code and some help 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)  
> |
```

## ... und Rstudio

- Rstudio bietet Heute sehr viel Unterstützung:

