

# Intro - Erste Schritte

Jan-Philipp Kolb

08 Mai 2017

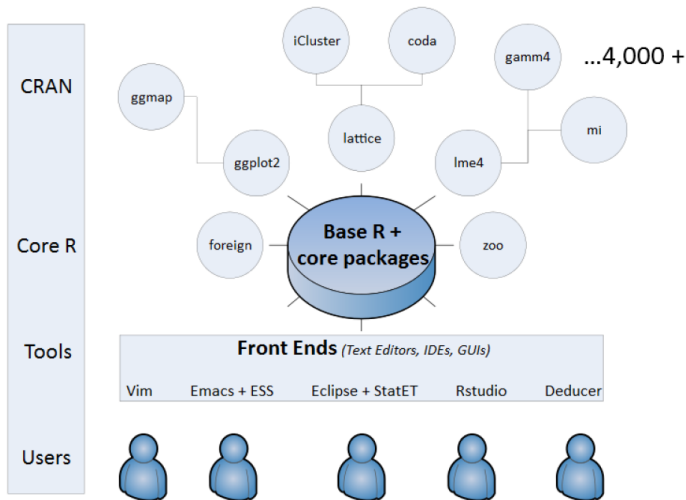
- 1 Warum R?
- 2 Warum die Nutzung von Schnittstellen?
- 3 Nutzung der Unterlagen auf GitHub

# Warum R?

# Gründe für die Nutzung 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 ;-)
- ...

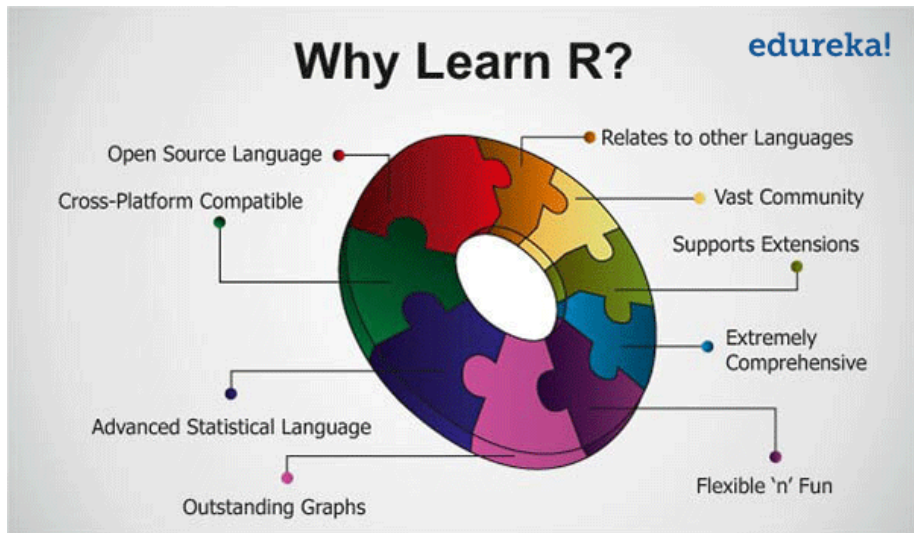
# Warum R?



# Gründe

- R ist frei verfügbar. Es kann umsonst runtergeladen werden.
- R ist eine Skriptsprache
- Gute Möglichkeiten für die Visualisierung ([Link](#) )
- R wird immer populärer
- Popularität von R

# Übersicht - warum R



# Nachteile von R

- ① Daten werden oft anderswo erfasst
- 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
- ③ Möglichkeit HTML Präsentationen zu erzeugen
- ④ Nutzung von C++
- ⑤ Nutzung von Datenbanken

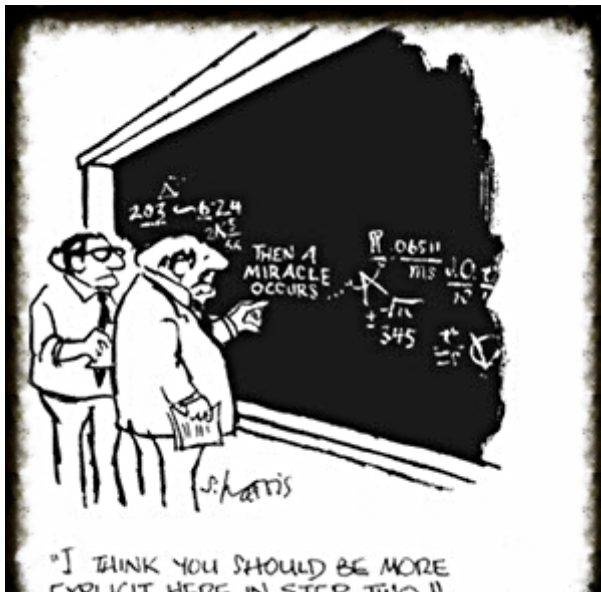
# Warum die Nutzung von Schnittstellen?

# Die Nutzung von Schnittstellen beim Import/Export

- Interaktion mit 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 ist 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

# 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



# Nutzung der Unterlagen auf GitHub

# Wie wird das Github Verzeichnis genutzt

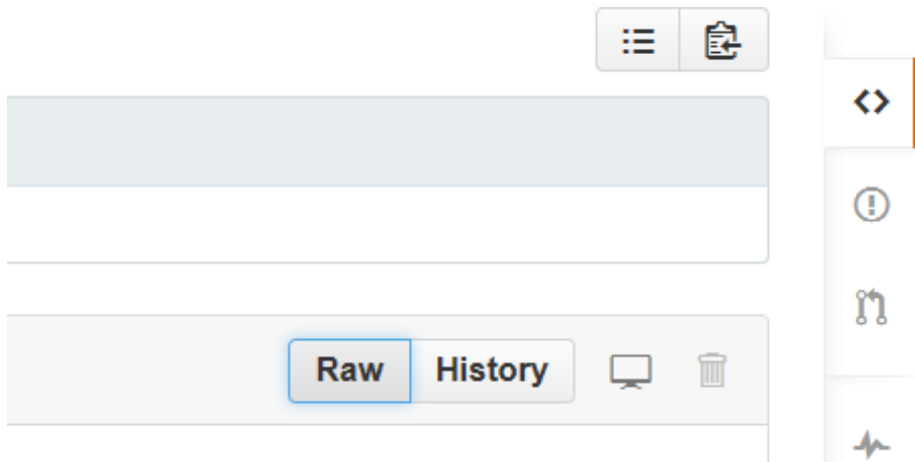
- Die folgende Seite ist die Startseite für den Kurs:

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

- Es lohnt sich immer wieder zu dieser Seite zurückzukehren, weil hier alle relevanten Dokumente verlinkt sind.
- Grundsätzlich bietet es sich an, den einzelnen Teilen der Veranstaltung mit dem File zu folgen, dass mit dem Begriff *Browser* verlinkt ist.

# 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.

# Organisatorisches

- Zusätzlich gibt es für jedes Kapitel eine Aufgabe, da man nur durch eigenes Trainieren auf der Lernkurve vorankommt.

# Links und Quellen

Wen Github näher interessiert:

- Hello World
- Understanding the GitHub