

Einführung in die Datenanalyse mit R

Jan-Philipp Kolb

9 Februar 2016

Warum R nutzen?

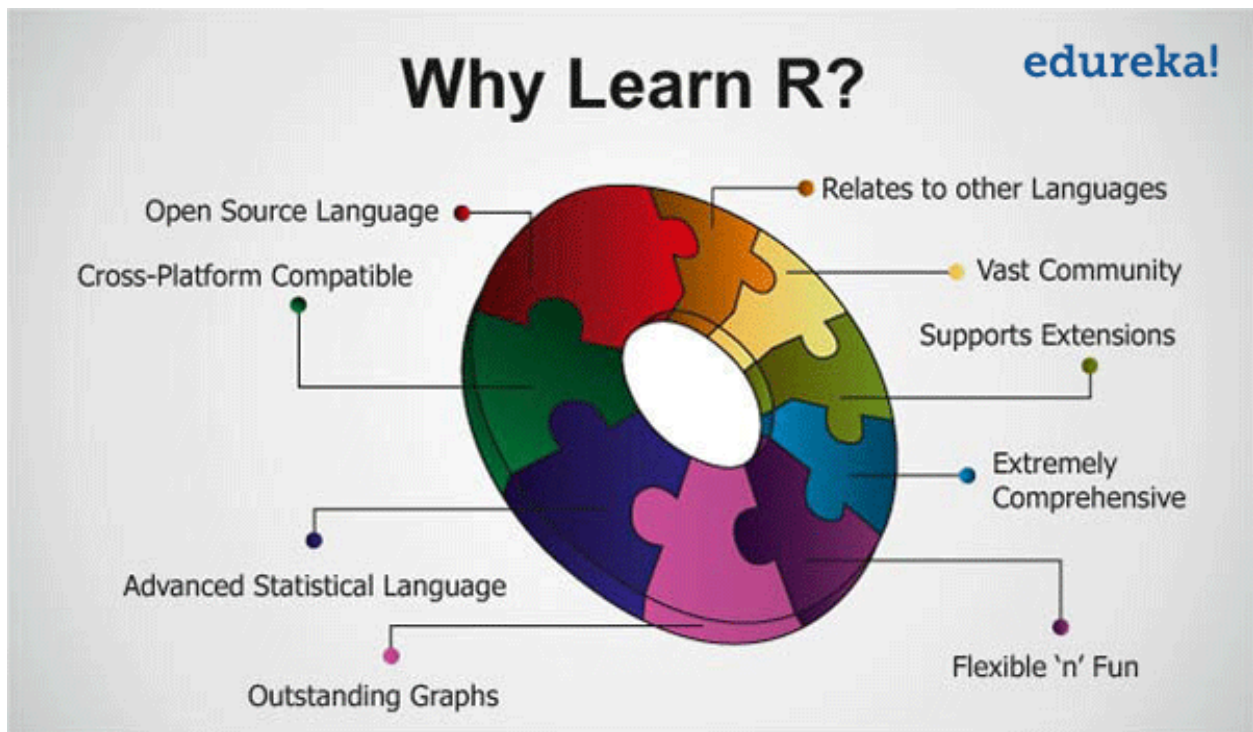
Gründe für die Nutzung von R

- 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 ;-)
- ...

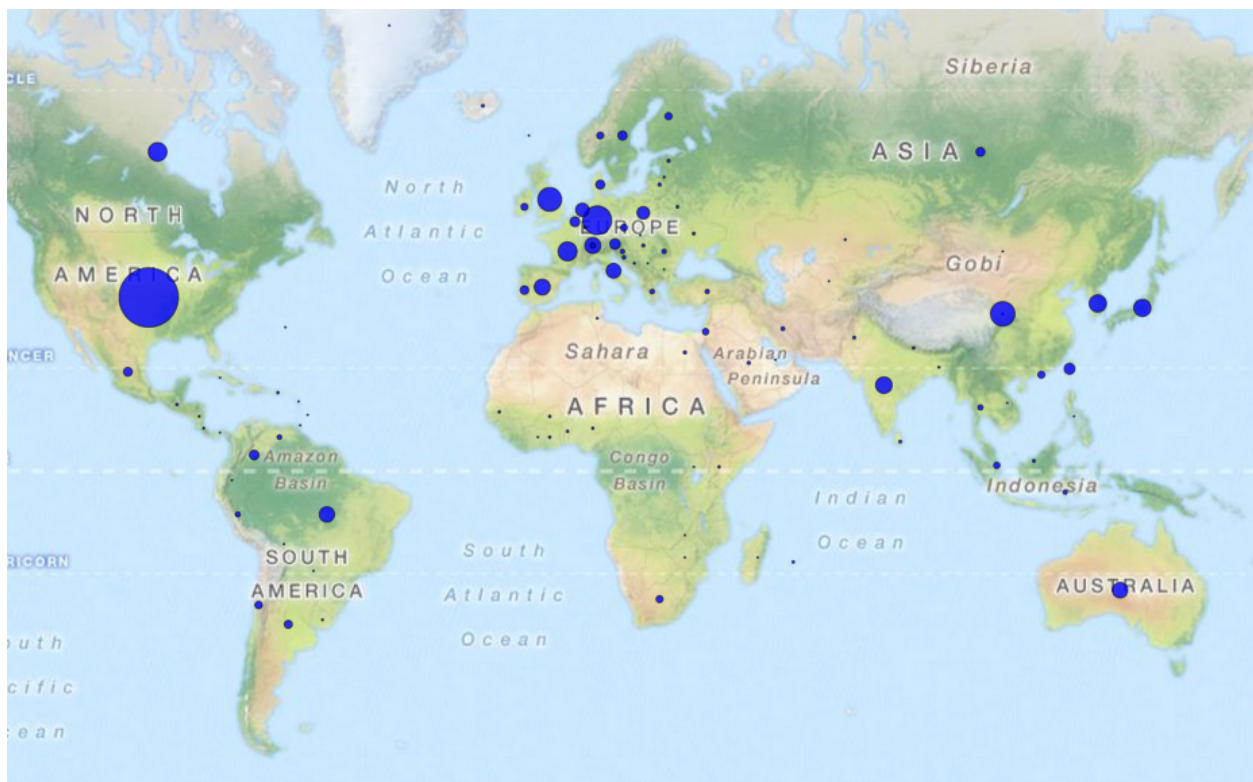
Gründe

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

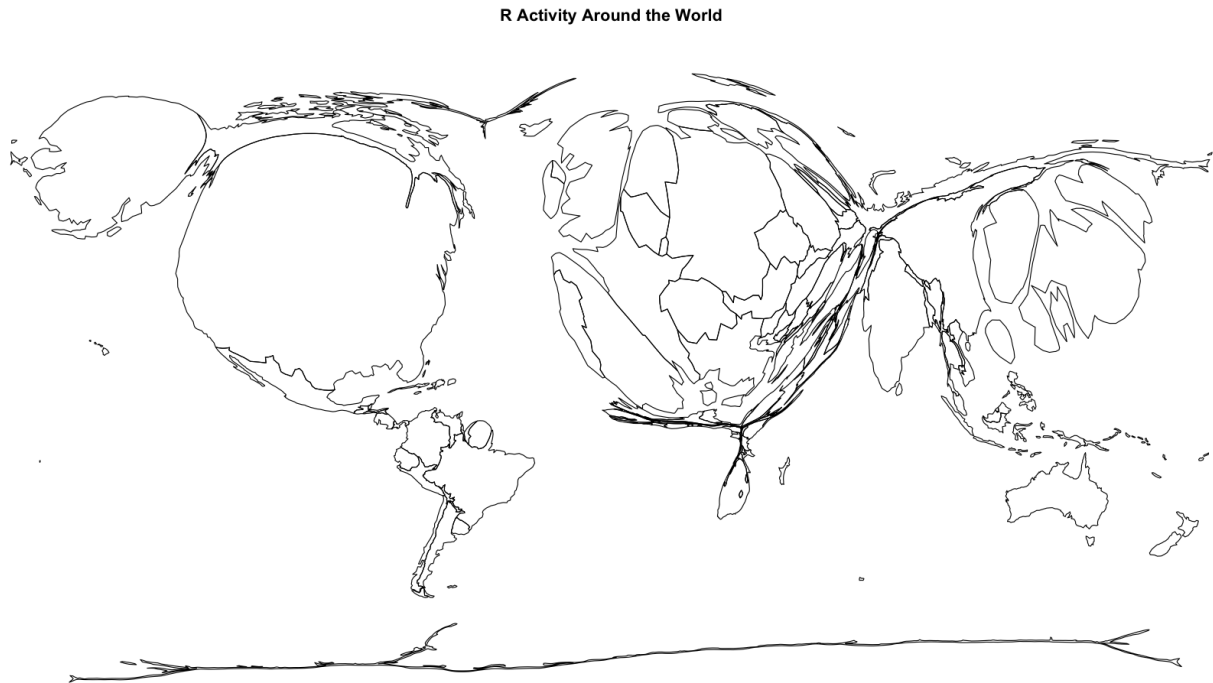
Übersicht - warum R



R Nutzer rund um die Welt



Wo sind die aktivsten Nutzer?



R für SPSS Nutzer

```
library("Rcmdr")
```

Bob Munich - [R for SPSS and SAS Users](#)

Links

- [Vergleich python und R](#)
- [Warum man R für Data Science lernen sollte](#)
- [R Technologie des Jahres](#)
- [Why R is Good for Business](#)
- [Warum R auf r-bloggers](#)
- [Intro R](#)
- [Intro R II](#)

Probleme mit Excel

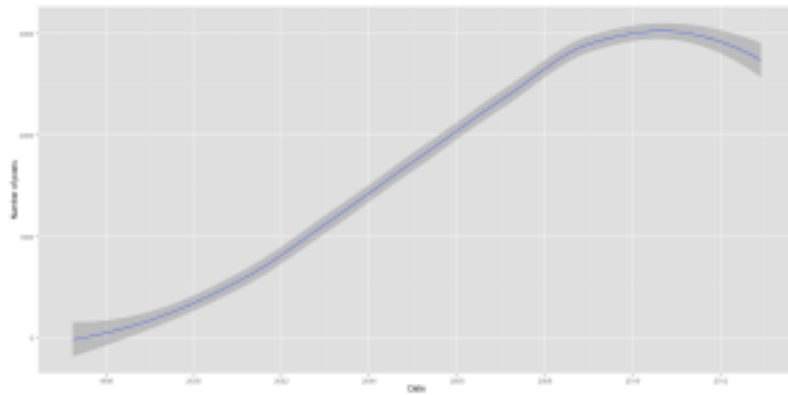
Weil andere Programme große Fehler haben:

- [Excel bug](#)

- [Datum in Excel](#)
- [Probleme mit Excel](#)

Dein Freund das Graphical User Interface (GUI)

Anzahl der Fragen in Hilfeforen zu R



[Source](#)

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)
- Der Download ist auf dieser Seite möglich:

www.r-project.org

Grpahisches User Interface

Aber die meisten Menschen nutzen einen Editor oder ein graphical user interface (GUI).

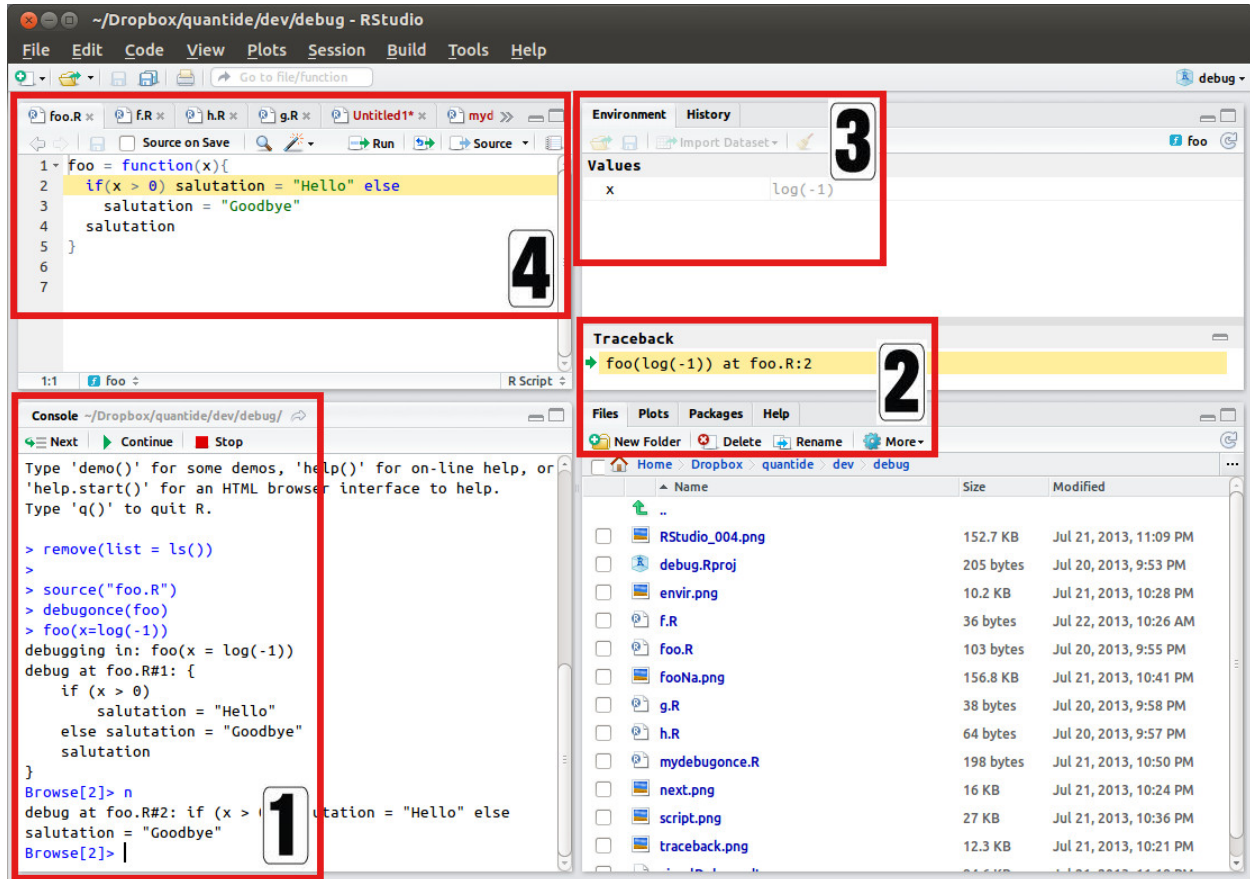
Aus den folgenden Gründen:

- Syntax highlighting
- Auto-Vervollständigung
- Bessere Übersicht über Graphiken, Bibliotheken

Verschiedene GUIs

- [Gedit](#) mit R-spezifischen Add-ons für Linux
- [Emacs](#)
- [TinnR](#)

- Ich nutze [Rstudio](#)!



Download der Unterlagen

Auf [github](#) sind alle Unterlagen für diesen Kurs zu finden.

Wie nutzt man github?

Rstudio

Sechs [Gründe](#) Rstudio zu nutzen.

Wie man Rstudio [nutzen kann](#).

Grundlagen in R

R ist eine Objekt-orientierte Sprache

Vektoren und Zuweisungen

- R ist eine Objekt-orientierte Sprache
- `<-` ist der Zuweisungsoperator

```
b <- c(1,2) # erzeugt ein Objekt mit den Zahlen 1 und 2
```

- Eine Funktion kann auf dieses Objekt angewendet werden:

```
mean(b) # berechnet den Mittelwert
```

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

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

```
length(b) # b hat die Länge 2
```

```
## [1] 2
```

Objektstruktur

```
str(b) # b ist ein numerischer Vektor
```

```
##  num [1:2] 1 2
```

Funktionen im base-Paket

Funktion	Bedeutung	Beispiel
length()	Länge	length(b)
max()	Maximum	max(b)
min()	Minimum	min(b)
sd()	Standardabweichung	sd(b)
var()	Varianz	var(b)
mean()	Mittelwert	mean(b)
median()	Median	median(b)

Diese Funktionen brauchen nur ein Argument.

Andere Funktionen brauchen mehr:

Argument	Bedeutung	Beispiel
quantile()	90 % Quantile	quantile(b,.9)
sample()	Stichprobe ziehen	sample(b,1)

Funktionen mit einem Argument

```
max(b)
```

```
## [1] 2
```

```
min(b)
```

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

```
sd(b)
```

```
## [1] 0.7071068
```

```
var(b)
```

```
## [1] 0.5
```

```
mean(b)
```

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

```
median(b)
```

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

Funktionen mit mehr Argumenten

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

```
## 90%
```

```
## 1.9
```

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

```
## [1] 2
```

Übersicht Befehle

Eine einführende Übersicht findet man [hier](#)