

# Nächste Schritte

Explorative Datenanalyse mit R

The R Bootcamp @ CSS



Dezember 2019

# Hallo Tidyverse Experten

Nach zwei Tagen kennt ihr nun die Grundlagen von R und des Tidyverse!

|      | Mi. 11 Dez.   | Do. 12 Dez.  |
|------|---|--|
| 0900 | Willkommen  | Wiederholung   |
| 0930 | <b>Einführung</b><br>in R(Studio)<br>+Interactive           | <b>Wrangling II</b><br>+Einführung<br>+Übung<br>+Interactive |
| 1200 | <i>Mittagessen</i>  | <i>Mittagessen</i>   |
| 1300 | <b>Daten</b><br>+Einführung<br>+Übung<br>+Interactive       | <b>Plotting I</b><br>+Einführung<br>+Übung<br>+Interactive   |
| 1500 | <i>Pause</i>  | <i>Pause</i>   |
| 1515 | <b>Wrangling I</b><br>+Einführung<br>+Übung<br>+Interactive | <b>Plotting II</b><br>+Einführung<br>+Übung                  |
| 1730 |   | Nächste Schritte   |

# "Tidyverse" kann noch mehr



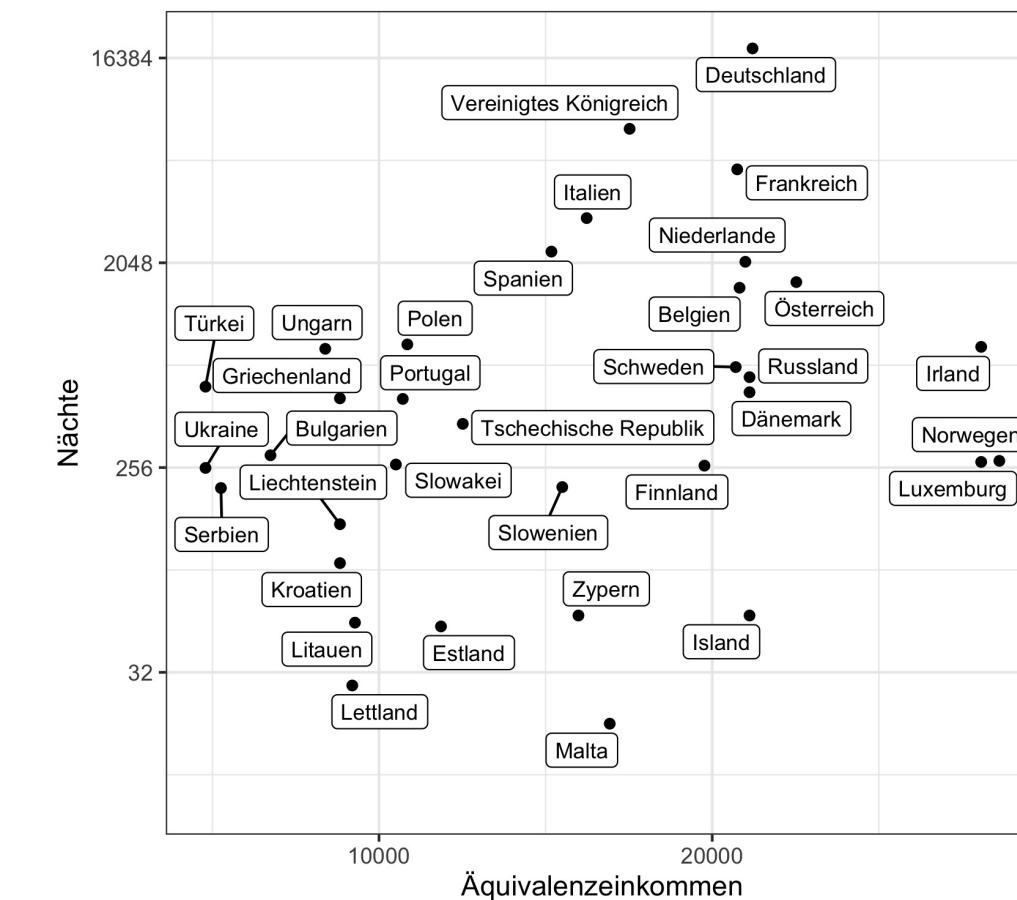
from [mitchelloharawild.com](http://mitchelloharawild.com)

# ggrepel

```
library(ggrepel)

# Lade Tourismus Daten
tour <- read_csv('1_Data/Tourismus.csv')
europa <- read_csv('1_Data/Europa.csv')

# Verbinde Nächte mit Äquivalenzeinkommen
tour %>%
  mutate(Nächte = Besucher * Dauer) %>%
  left_join(europa) %>%
  ggplot(aes(x = Äquivalenzeinkommen,
             y = Nächte,
             label = Land)) +
  scale_y_continuous(trans = 'log2') +
  geom_point() +
  geom_label_repel(size = 2) +
  theme_bw()
```



# Textanalyse

```
library(tidytext)
library(wordcloud)

# Wörter zählen
counts <-
  read_html("../R_(programming_language)") %>%
  html_text() %>%
  tibble() %>%
  unnest_tokens(w, ".") %>%
  filter(!str_detect(w, '[[:digit:]]')) %>%
  anti_join(stop_words) %>%
  count(word, sort = TRUE) %>%
  top_n(100)

# Wortwolke
wordcloud(counts$word,
           counts$n)
```



# Statistik

```
library(tidyverse)

# Lade Tourismus Daten
tour <- read_csv('1_Data/Tourismus.csv')
länd <- read_csv('1_Data/Länder.csv')

# verbinde Daten
data <- tour %>%
  inner_join(länd) %>%
  mutate(Nächte = Besucher * Dauer)

# Regressionsanalyse
model = lm(Nächte ~ Bevölkerung +
            Dichte + BIP,
            data = data)
```

```
# Bestimmtheitsmass
summary(model)$r.squared
## [1] 0.05719

# Koeffizienten
summary(model)$coef[,-2]
##                 Estimate   t value Pr(>|t|)
## (Intercept) 1.213e+01  0.01294  0.9897
## Bevölkerung 2.379e-06  1.36594  0.1775
## Dichte      -3.381e-01 -0.82598  0.4124
## BIP          2.988e+01  1.50779  0.1373
```

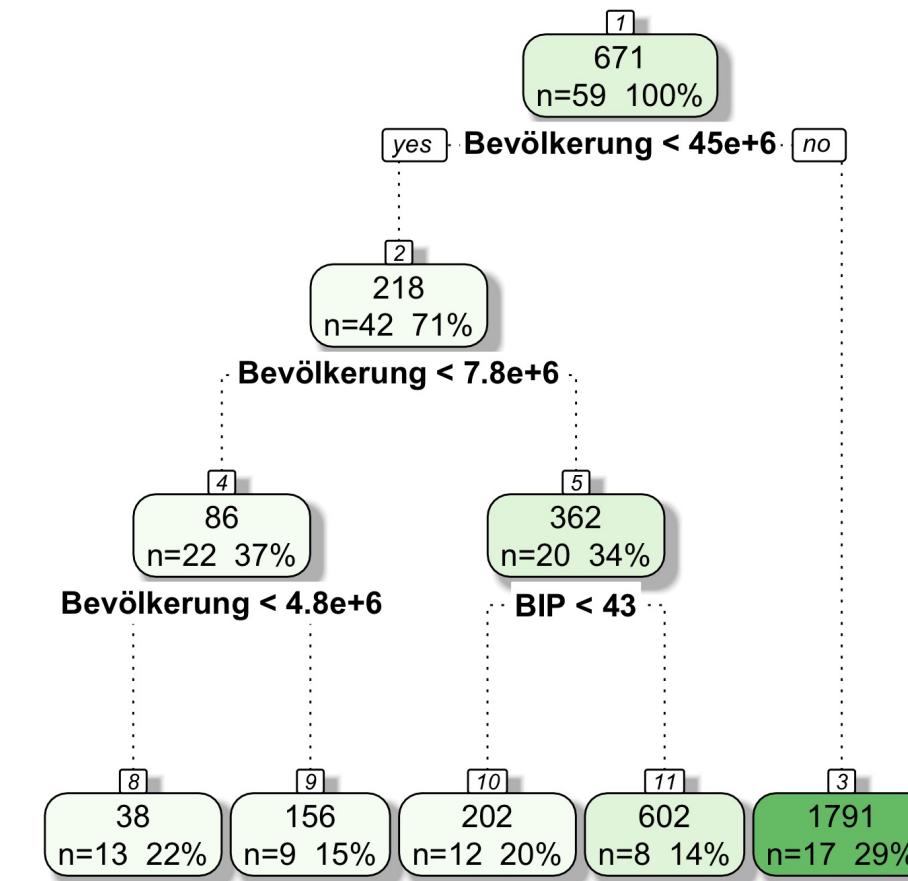
# Maschinelles lernen

```
library(tidyverse) ; library(rpart)

# Lade Tourismus Daten
tour <- read_csv('1_Data/Tourismus.csv')
länd <- read_csv('1_Data/Länder.csv')

# verbinde Daten
data <- tour %>%
  inner_join(länd) %>%
  mutate(Nächte = Besucher * Dauer)

# Regressionsanalyse
rpart(
  formula = Besucher ~ Bevölkerung +
    Dichte + BIP,
  data = data)
```



# rmarkdown

```
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = FALSE)
```

```
# Lese Daten ein
library(tidyverse)
tour <- read_csv('1_Data/Tourismus.csv')
```
```

```
## Baseler Tourismus
```

Insgesamt kamen nach Basel `$`r sum(tour$Besucher)`$` Besucher.

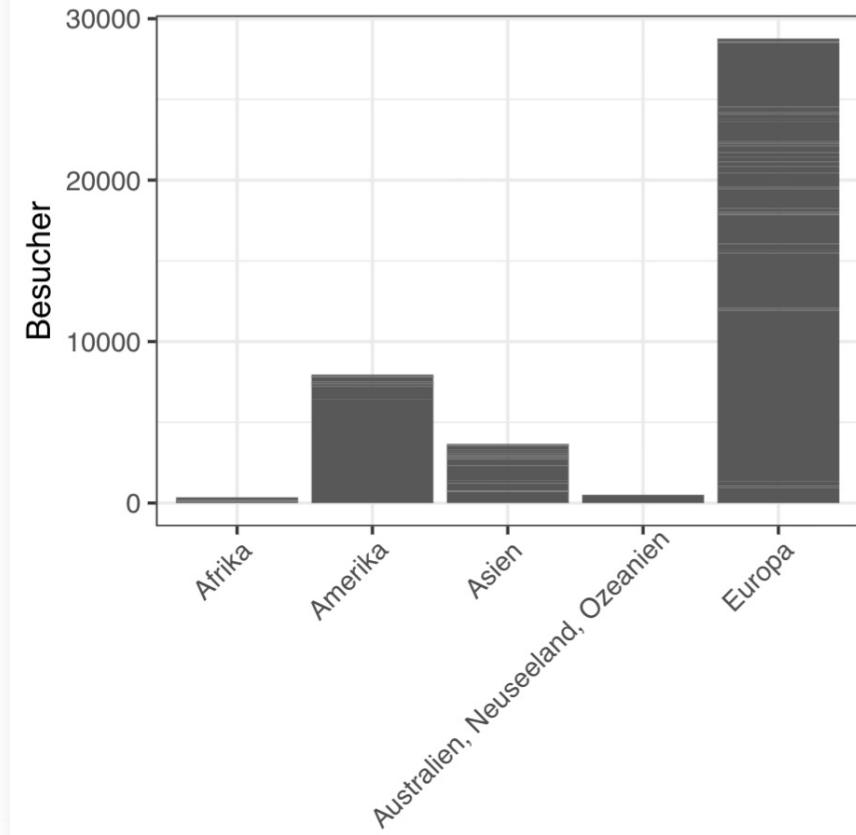
Hier sind Besucher aufgeschlüsselt nach Region.

```
```{r, message=FALSE, fig.height=4, fig.width=4}
# Plotte nach Region
tour %>%
  ggplot(aes(x = Region,
             y = Besucher)) +
  geom_bar(stat="identity") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45,
                                    hjust=0.95),
        axis.title.x=element_blank())
```
```

## Baseler Tourismus

Insgesamt kamen nach Basel  $4.1237 \times 10^4$  Besucher.

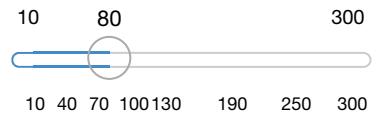
Hier sind Besucher aufgeschlüsselt nach Region.



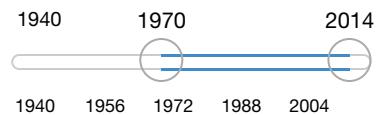
## Movie explorer

### Filter

Minimum number of reviews  
on Rotten Tomatoes



### Year released



Minimum number of Oscar  
wins (all categories)



Dollars at Box Office  
(millions)



Number of movies selected:

# Nächste Schritte

1. *Anwenden*
2. *Bücher*
3. *Webseiten*
4. *Hilfe & Consulting*
5. *Weiterführende Kurse*



# Bücher

Hier ist eine unvollständige Liste guter Bücher über R lose geordnet nach vorausgesetzter Erfahrung.



# Webseiten

Das Netz ist vll. der beste Ort für Informationen über R.

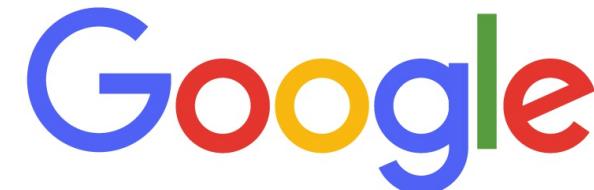
*Beginnt eure Suche mit:*

**Google**. Stellt sicher, dass ihr *R* oder *Rproject* mit in eurer Suche habt.

*Meistens werdet ihr dann weitergeleitet zu:*

**R-bloggers** informiert euch über die neusten R Entwicklungen. Meldet euch für den Newsletter an.

**Stackoverflow** ist eine Seite für R Probleme und Lösungen. Versucht selbst mal ein Problem zu posten. Oft bekommt ihr erstaunlich schnell Antwort.



# Support



**Dr. Dirk Wulff**

[dirkwulff.org](http://dirkwulff.org)  
[github.com/dwulff](https://github.com/dwulff)  
packages: `cstab`,  
`mousetraps`, `memnet` `choicepp`



**Markus Steiner**

[github.com/mdsteiner](https://github.com/mdsteiner)  
packages: `ShinyPsych`,  
`EFAdiff`

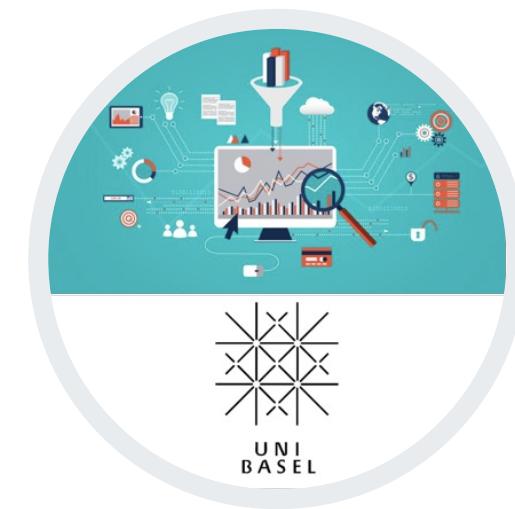


**Dr. Michael Schulte-  
Mecklenbeck**

[schulte-mecklenbeck.com](http://schulte-mecklenbeck.com)  
[github.com/schultem](https://github.com/schultem)

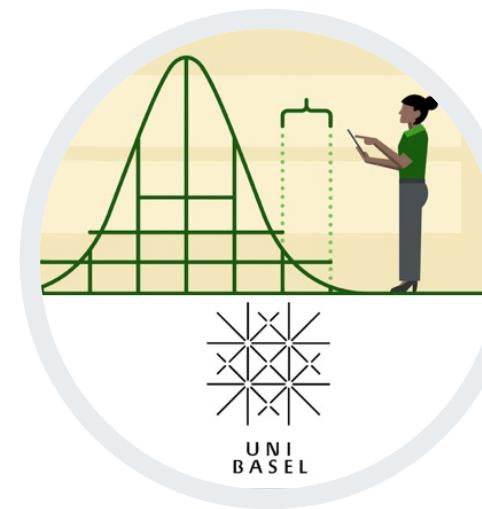
# Weiterführende Kurse

Explorative Datenanalyse mit  
R



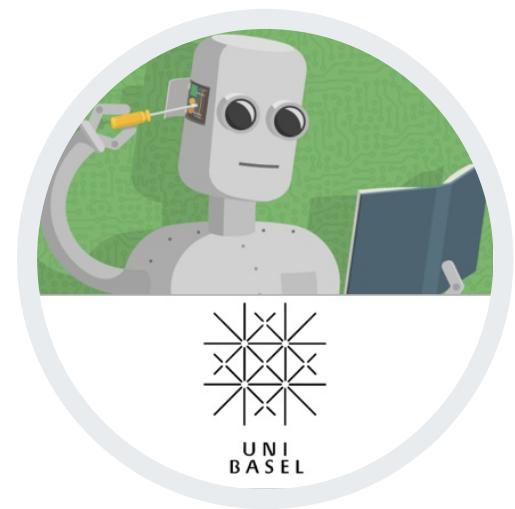
Februar 14/15, 2020

Statistik mit R



March 20/21, 2019

Maschinelles Lernen mit  
R



April 3/4, 2019

Reporting mit R



Mai 15/16, 2019

# Bitte gebt uns Feedback



from [medium.com](https://medium.com)

## Follow-up Fragebogen Explorative Datenanalyse mit R @ CSS Dezember 2019

Bitte nimm dir ein paar Minuten Zeit, um uns Feedback zum Kurs Explorative Datenanalyse mit R zu geben. Dein Feedback ist anonym.

\* Erforderlich

Wie zufrieden warst du mit dem Kurs? \*

1 2 3 4 5

Sehr unzufrieden      Sehr zufrieden

Die Präsentationen waren \*

1 2 3 4 5

Zu kurz      Zu lang

15 / 16

# Herzlichen Dank! und noch eine Sache...

16 / 16



[www.therbootcamp.com](http://www.therbootcamp.com)

Explorative Datenanalyse mit R @ CSS | Dezember 2019