

# 0 - Organisatorisches und Einführung

Sozialwissenschaftliche Forschungsmethoden für Fortgeschrittene

**Dr. André Calero Valdez**

Prof. Dr. Martina Ziefle

Human-Computer Interaction Center

Wintersemester 19/20



# Organisatorisches

LV.Nummer: 79.41688

Termine: MO, 08:30-10:00 Uhr im Raum H08 (1385|219)

Nr	Datum
1	14.10.2019
2	21.10.2019
3	28.10.2019
4	04.11.2019
5	11.11.2019
6	18.11.2019
7	25.11.2019

Nr	Datum
8	02.12.2019
9	09.12.2019
10	16.12.2019
11	06.01.2020
12	13.01.2020
13	20.01.2020
14	27.01.2020

# Folien

<https://goo.gl/RifCHp>



# Anmeldung und Klausur

## Anmeldung

- 2 Anmeldungen: Zur Vorlesung **und** Klausur (benotete Prüfungsleistung)
- Plus 2 Anmeldungen im Seminar "Forschungsmethoden".
  - Seminar dient der Übung der Inhalte der Vorlesung.

## Klausur

- Die Veranstaltung hat einen **Selbststudiumsanteil** von 60h ~ **4h/Woche**.
- Das Seminar hat ebenfalls einen Selbststudiumsanteil von 60h ~ 4h/Woche.
- Die Klausur wird 50% Multiple-Choice- und 50% Freitext-Aufgaben sein.

# 2 Teile und 2 Ziele

1. Theorie und Mathematik

2. Praxis und R

# Vorlesung und Seminar

## Inhalt der Vorlesung

- Theorie und Grundlagen
- Einführung in die Praxis mit R

## Inhalt des Seminars

- Anwendung mit echter Forschungsfrage
- Übungen für den Umgang mit R

**Beides** ist klausurrelevant.



# Unterschied R und SPSS

## Datenanalyse mit GUI vs. Datenanalyse als Programmiersprache

- GUI<sup>1</sup> Datenanalyse-Tools: **SPSS**, STATA, RapidMiner, etc.
- Programmiersprachen für die Datenanalyse: Python, Julia, **R**, etc.



# Unterschied R und SPSS

## Datenanalyse mit GUI vs. Datenanalyse als Programmiersprache

- GUI<sup>1</sup> Datenanalyse-Tools: **SPSS**, STATA, RapidMiner, etc.
- Programmiersprachen für die Datenanalyse: Python, Julia, R, etc.

## SPSS

- Kostenpflichtig
- "Einfach" zu lernen
- Schnelle Analyse durch Menüführung
- Wiederholung von Analysen nur bei Anwendung von Syntax
- Schlechte Visualisierung

# Unterschied R und SPSS

## Datenanalyse mit GUI vs. Datenanalyse als Programmiersprache

- GUI<sup>1</sup> Datenanalyse-Tools: **SPSS**, STATA, RapidMiner, etc.
- Programmiersprachen für die Datenanalyse: Python, Julia, **R**, etc.

### SPSS

- Kostenpflichtig
- "Einfach" zu lernen
- Schnelle Analyse durch Menüführung
- Wiederholung von Analysen nur bei Anwendung von Syntax
- Schlechte Visualisierung

### R

- Kostenfrei
- Cross-Plattform kompatibel
- Höherer Lernaufwand
- Analyse erfordert das Schreiben eines Programms
- Wiederholung eingebaut
- Reproduzierbarkeit
- Druckreife Visualisierung, Gute Integration in Publishing Prozesse

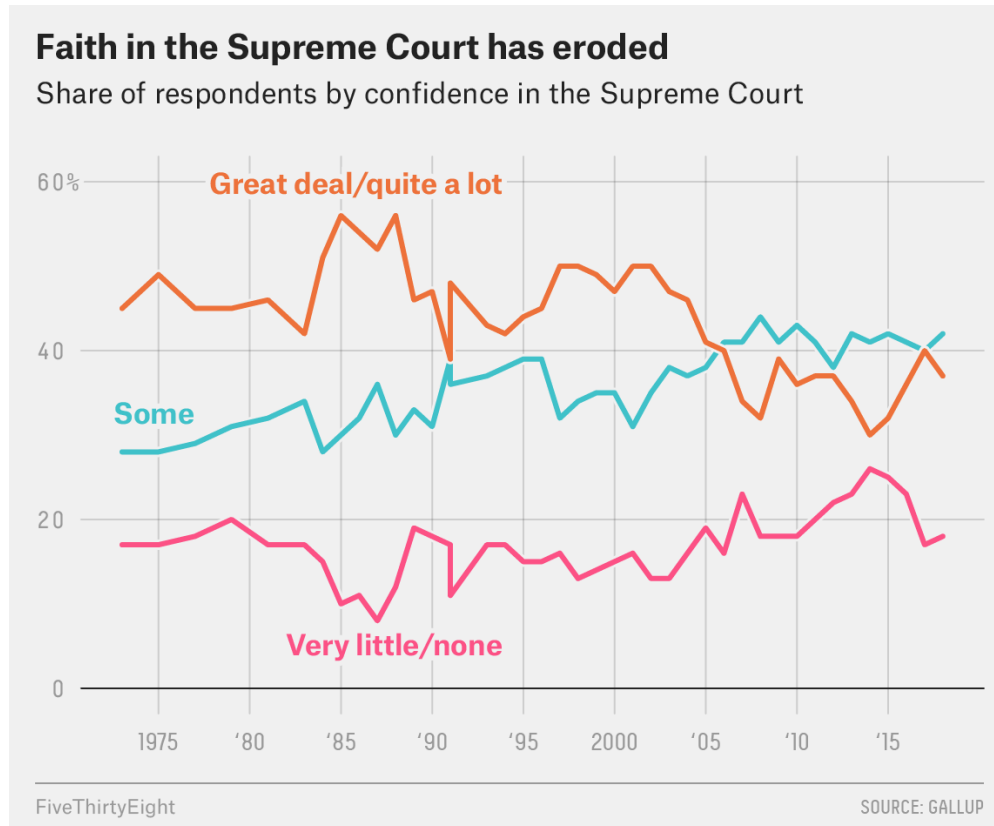
[1] Graphical User Interface. Programme mit Maussteuerung.

# Vorteile von R im Detail (exemplarisch)

## Reproduzierbarkeit (Reproducibility)

- Im Sinne von Open-Science, nachvollziehbare Analysen
- Im Sinne der Arbeitserleichterung (z.B. neue, gleiche Studie)

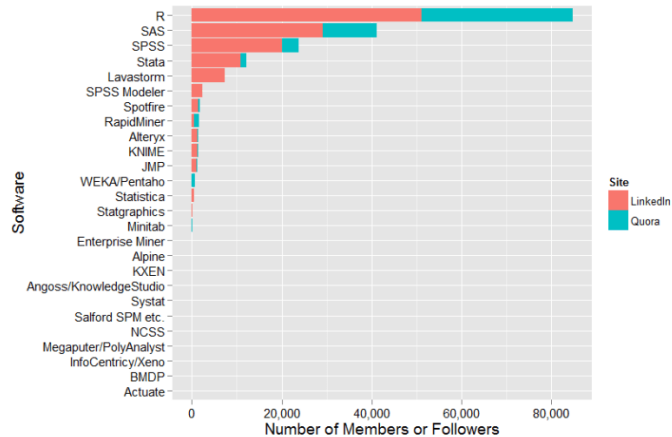
# Einsatz im Berufsleben



Quelle: <http://fivethirtyeight.com>

# Ist R wichtiger auf dem Arbeitsmarkt?

Größere und aktivere Community



**Abb. 1:** Anzahl LinkedIn Jobs und Quora Member 2015

# Ist R wichtiger auf dem Arbeitsmarkt?

## Größere und aktivere Community

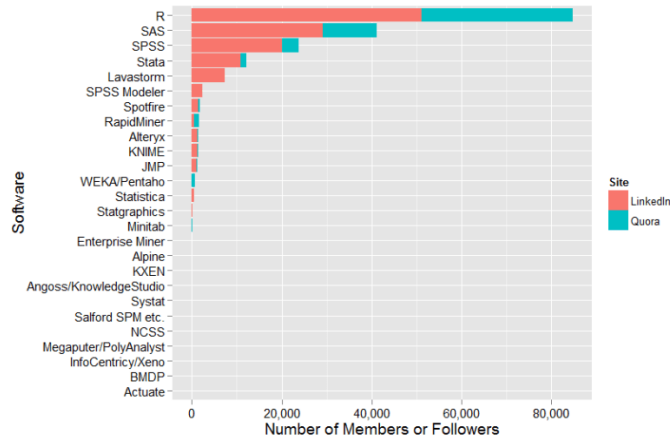


Abb. 1: Anzahl LinkedIn Jobs und Quora Member 2015

## Gewinnt wissenschaftlich Bedeutung

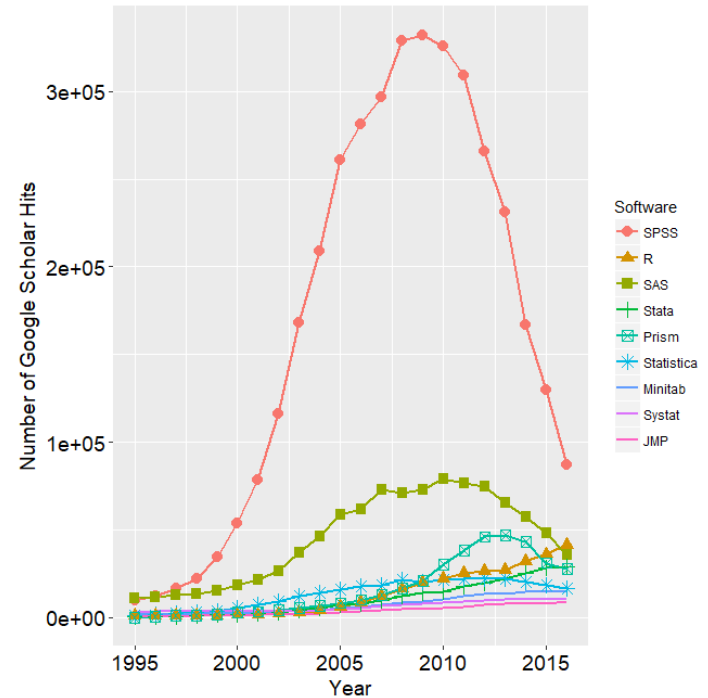
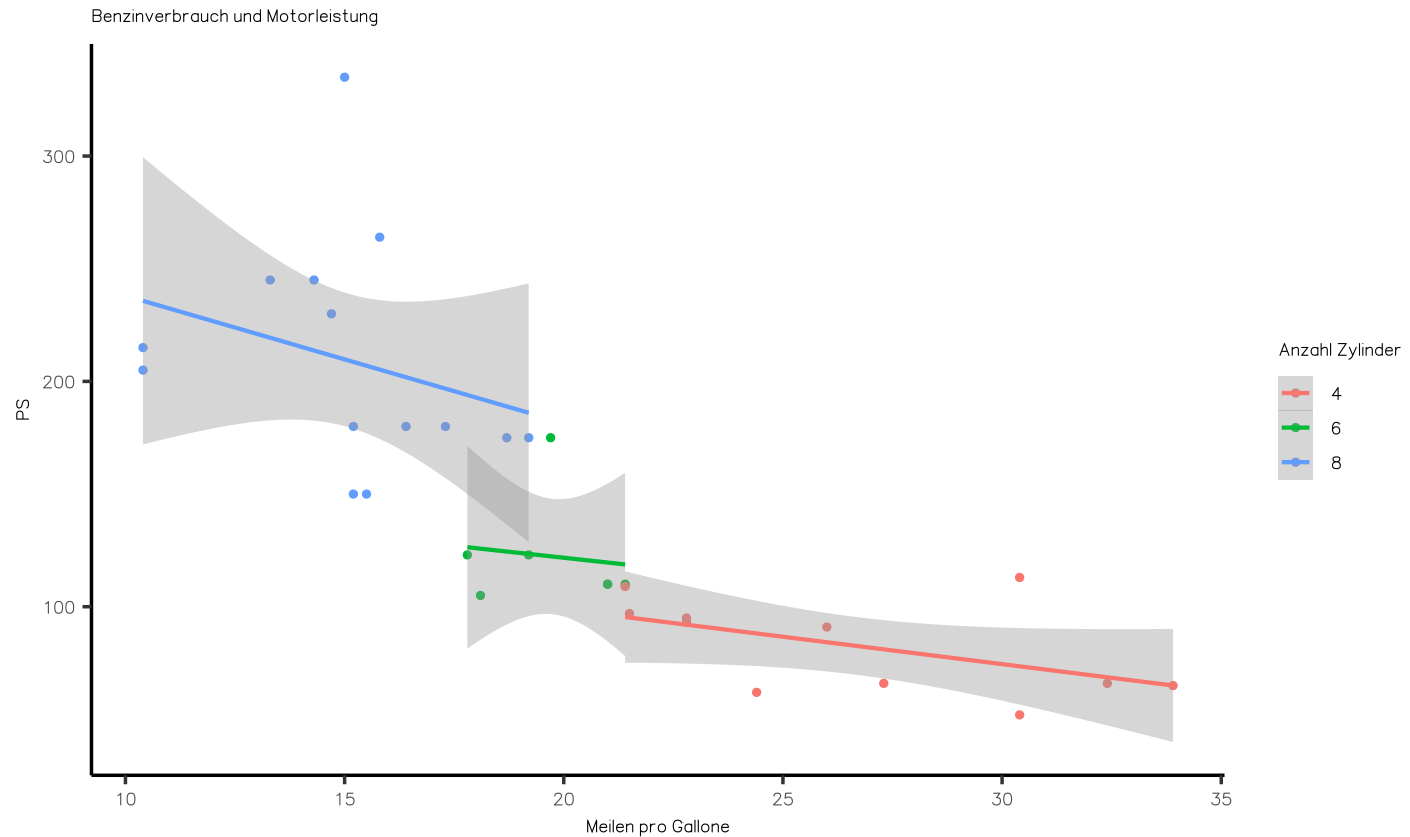


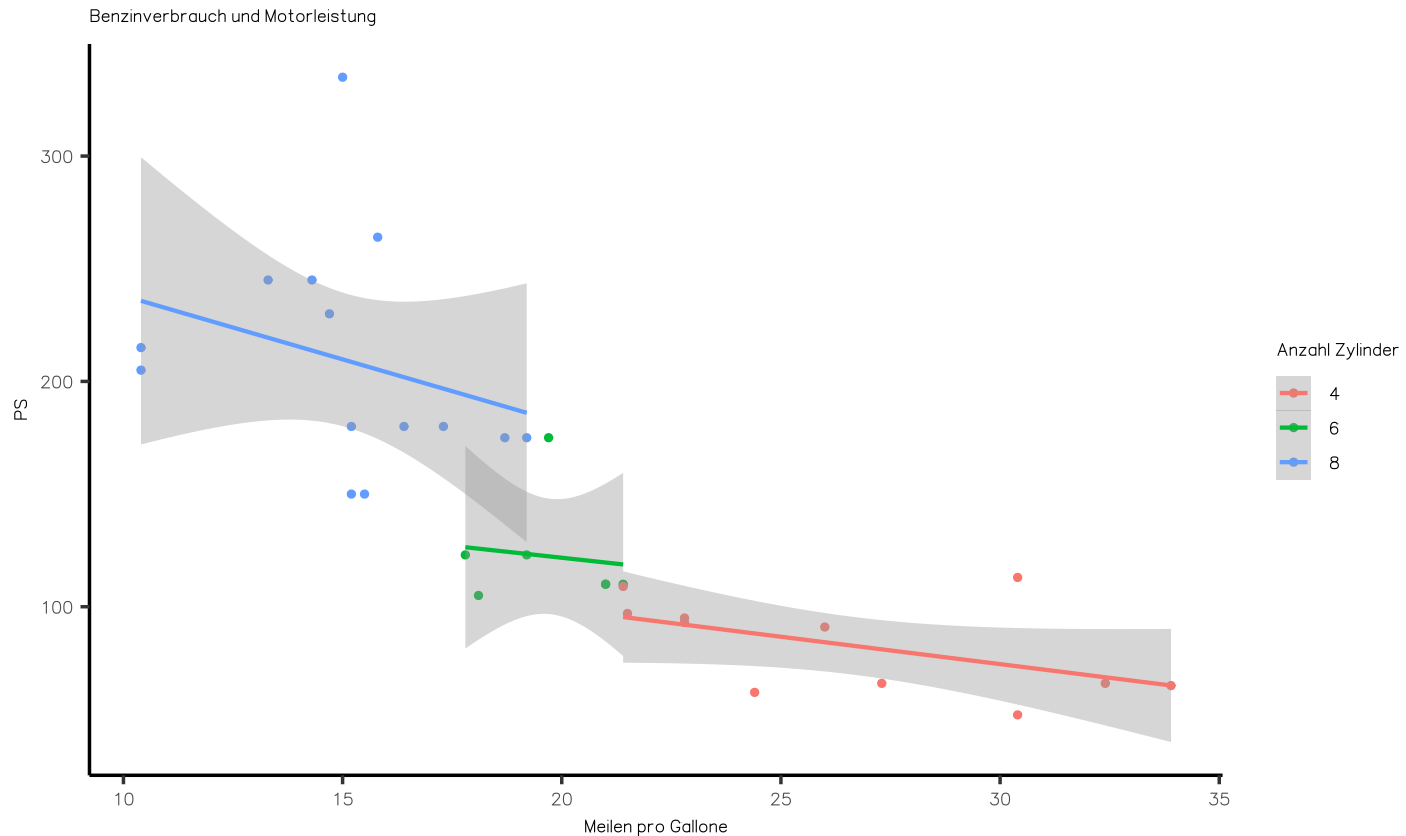
Abb. 2: Anzahl wissenschaftlicher Veröffentlichungen auf Google Scholar

Quelle: <http://r4stats.com/articles/popularity/>

# Datenvisualisierung mit R



# Datenvisualisierung mit R



Eine (lange) Zeile in R:

```
ggplot(mtcars, aes(x=mpg, y=hp, color=factor(cyl))) +  
  geom_point() + geom_smooth(method="lm", formula=y~x) +  
  labs(title="Benzinverbrauch und Motorleistung", x="Meilen pro Gallone",
```





# Lernmaterialien

## e-Learning der RWTH Aachen und Partner

- Alle Kursmaterialien: <https://www.calerovaldez.com/course/methoden/>
- Im L2P werden sämtliche Inhalte zur Verfügung gestellt.
- DataCamp Übungsraum (online Trainingstool)
- Videoaufzeichnungen der Veranstaltungen + Tutorials

## Literatur zur Veranstaltung

- Döring/Bortz - Forschungsmethoden und Evaluation
- Online-Bücher:
  - Computational Communication Science, Link: <https://bookdown.org/andreacalerovaldez/ccs/>
  - R for Data Science, Link: <http://r4ds.had.co.nz>
  - R for Social Science, Link: <http://socviz.co/>
  - Modern Dive: <https://moderndive.com/index.html>

# Zusätzliches Material

Es gibt hunderte Webseiten, Bücher, Foren zur R:

- <http://compcogscisydney.org/learning-statistics-with-r/>
- <http://compcogscisydney.org/psyr/>
- <http://www.sthda.com/english/>
- <https://www.statmethods.net/index.html>
- <https://www.r-bloggers.com/>
- Cheat-Sheets: <https://www.rstudio.com/resources/cheatsheets/>

## Twitter-Hashtags und User:

Twitter-Account anlegen und nach Hashtag suchen und Usern folgen.

Hashtags:

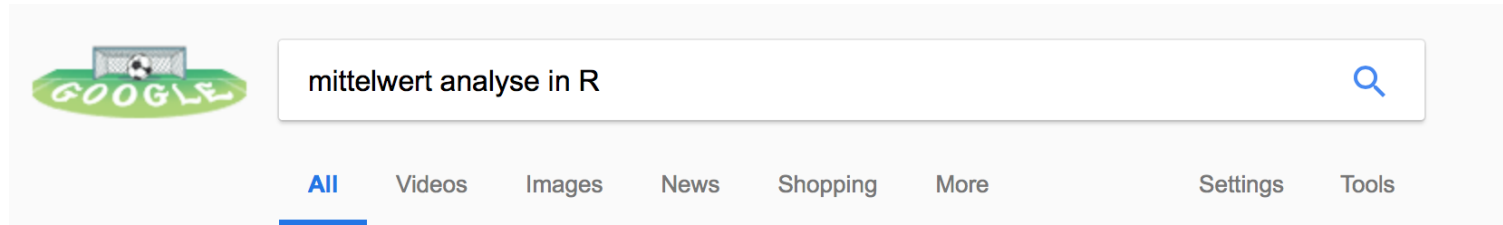
- #rlang, #rstats, #r4ds, #rladies

User:

- @RLangTip, @RBloggers, @rOpenSci, @RLadiesGlobal, @hadleywickham, @juliasilge

# Was tun wenn etwas nicht funktioniert?

Google!



The image shows a Google search interface. At the top left is the Google logo. To its right is a search bar containing the text "mittelwert analyse in R". Below the search bar are tabs for "All", "Videos", "Images", "News", "Shopping", "More", "Settings", and "Tools". Below the tabs, it says "About 824,000 results (0.58 seconds)". The first search result is titled "[PDF] 3 Deskriptive Statistik in R (univariat) - TU Chemnitz" with a green link to "https://www.tu-chemnitz.de/hsw/.../mb/03\_desc\_in\_R\_univar.pdf" and a "Translate this page" link. The snippet below the link reads: "diese Art der **Analyse** sinnvoll ist, hängt natürlich von der Struktur Ihrer Daten ab. ... erkennen, dass der **Mittelwert** der Männer vor allem durch die 3 Männer ...". The second result is titled "Deskriptive Statistik mit R - Datenanalyse mit R, STATA & SPSS" with a green link to "https://www.beratung-statistik.de/statistik.../r.../deskriptive-statistik-r..." and a "Translate this page" link. The snippet reads: "Sehen Sie sich den Datensatz zunächst an, indem Sie in die R-Konsole ... Der **Mittelwert** und der Median werden in R mit folgenden Befehlen berechnet:". The third result is titled "GNU R: Deskriptive Statistik – Wikibooks, Sammlung freier Lehr-, Sach ..." with a green link to "https://de.wikibooks.org/wiki/GNU\_R:\_Deskriptive\_Statistik" and a "Translate this page" link. The snippet reads: "Jump to **Mittelwerte** - **Mittelwerte**[Bearbeiten] ... R> tore <- Bundesliga\$Tore. ... Zentralwert (Median) als ebenfalls gebräuchlicher **Mittelwert** kann auch für ... Mittelwert, Extremwerte ... · Streuung · Häufigkeitstabellen und ...". The fourth result is titled "Deskriptive Kennwerte - R-Stutorials" with a green link to "www.r-stutorials.de/eda/deskriptive-kennwerte" and a "Translate this page" link. The snippet reads: "In diesem Kapitel geben wir Ihnen eine Einführung zur deskriptiven **Analyse in R**. Sie ... Der getrimmte **Mittelwert** ohne die oberen und unteren 10% der Daten."

About 824,000 results (0.58 seconds)

[\[PDF\] 3 Deskriptive Statistik in R \(univariat\) - TU Chemnitz](#)  
[https://www.tu-chemnitz.de/hsw/.../mb/03\\_desc\\_in\\_R\\_univar.pdf](https://www.tu-chemnitz.de/hsw/.../mb/03_desc_in_R_univar.pdf) ▾ [Translate this page](#)  
diese Art der **Analyse** sinnvoll ist, hängt natürlich von der Struktur Ihrer Daten ab. ... erkennen, dass der **Mittelwert** der Männer vor allem durch die 3 Männer ...

[Deskriptive Statistik mit R - Datenanalyse mit R, STATA & SPSS](#)  
<https://www.beratung-statistik.de/statistik.../r.../deskriptive-statistik-r...> ▾ [Translate this page](#)  
Sehen Sie sich den Datensatz zunächst an, indem Sie in die R-Konsole ... Der **Mittelwert** und der Median werden in R mit folgenden Befehlen berechnet:.

[GNU R: Deskriptive Statistik – Wikibooks, Sammlung freier Lehr-, Sach ...](#)  
[https://de.wikibooks.org/wiki/GNU\\_R:\\_Deskriptive\\_Statistik](https://de.wikibooks.org/wiki/GNU_R:_Deskriptive_Statistik) ▾ [Translate this page](#)  
Jump to **Mittelwerte** - **Mittelwerte**[Bearbeiten] ... R> tore <- Bundesliga\$Tore. ... Zentralwert (Median) als ebenfalls gebräuchlicher **Mittelwert** kann auch für ...  
[Mittelwert, Extremwerte ...](#) · [Streuung](#) · [Häufigkeitstabellen und ...](#)

[Deskriptive Kennwerte - R-Stutorials](#)  
[www.r-stutorials.de/eda/deskriptive-kennwerte](http://www.r-stutorials.de/eda/deskriptive-kennwerte) ▾ [Translate this page](#)  
In diesem Kapitel geben wir Ihnen eine Einführung zur deskriptiven **Analyse in R**. Sie ... Der getrimmte **Mittelwert** ohne die oberen und unteren 10% der Daten.

# Übersicht

1. Was sind Methoden? => [Link](#)
2. Qualitative und Quantitative Daten? Forschungsfrage wählen => [Link](#)
3. Wissenschaftstheorie, Empirie und Theorie => [Link](#)
4. Forschungsinstrument entwickeln, Messtheorie, Skalenniveaus => [Link](#)
5. Deskriptive Statistik, zentrale Tendenz und Streuung => [Link](#)
6. Verteilungen, Stichproben und Wahrscheinlichkeit => [Link](#)
7. Inferenz, Hypothesen, Fehler 1. und 2. Art, t-Test => [Link](#)
8. Alpha-Fehler Kummulierung, ANOVA, MANOVA, Bonferroni => [Link](#)
9. Zusammenhänge, Korrelation, lineare Regression => [Link](#)
10. Skalen, Likert-Skalen, Reliabilität und Faktoren-Analyse => [Link](#)
11. Explorative Statistik, parametrische und nicht-parametrische Verfahren => [Link](#)
12. Conjoint-Verfahren, Cluster-Analyse => [Link](#)
13. Effekt-Stärken und Poweranalyse => [Link](#)