

# 04

# Sozialwissenschaftliche Daten

Einführung in die quantitativen Forschungsmethoden

# Heute

- Quiz zu letzter Sitzung
- Arten sozialwissenschaftlicher Daten
- Besonderheiten von Umfragedaten
- Arten von Fragen
- Gütekriterien
- Die European Social Survey

# Quiz zu letzter Sitzung

Quiz Link

Quiz Link2

# Arten sozialwissenschaftlicher Daten



# Arten sozialwissenschaftlicher Daten

- amtliche Daten
  - z.B. Bevölkerungsstatistik, Wirtschaftsdaten, ...
- Umfragedaten
  - Befragungen zu verschiedenen Themen durch Umfrageinstitute
- Rohdaten
  - z.B. Pressemitteilungen, social media posts, Wahlplakate, ...
- Tracking Daten
  - z.B. Konsumdaten, digital trace data, social media Daten

# Besonderheiten von Umfragedaten

# Umfragedaten: Grundlagen

- [Pew Video](#)
- Befragung einer **repräsentativen Stichprobe** - d.h. einer Teilgruppe, die in ihren Charakteristiken der Bevölkerung entspricht
  - **Bevölkerung / Population** ↔ **Stichprobe**
  - häufig durch Zufallsstichprobe (randomisiert) oder mithilfe von Quoten
  - Gewichtung von Antworten
- Erhebung durch **Befragung**
  - z.B. Umfragen per Telefon, Fußgängerzone, Hausbesuch, self-completion Fragebogen, ...
  - der **Umfragemodus** hat einen Einfluss auf die Repräsentativität

# Umfragedaten

**Welche möglichen Besonderheiten von Umfragedaten als Befragungsdaten fallen Ihnen ein?**

Im Vergleich z.B. zu amtlichen Statistiken, (physikalischen) Messungen, ...

# Subjektivität und mentale Abkürzungen

- Menschen haben häufig Schwierigkeiten Fragen zu Fakten oder Sachverhalte richtig zu beantworten
  - Vergessen (z.B. mit zeitlichem Abstand)
  - z.B. Platz in der Einkommensverteilung
  - z.B. unterschiedliche Bewertungen eines Zustands
  - z.B. Präferenz für 'runde' Antworten

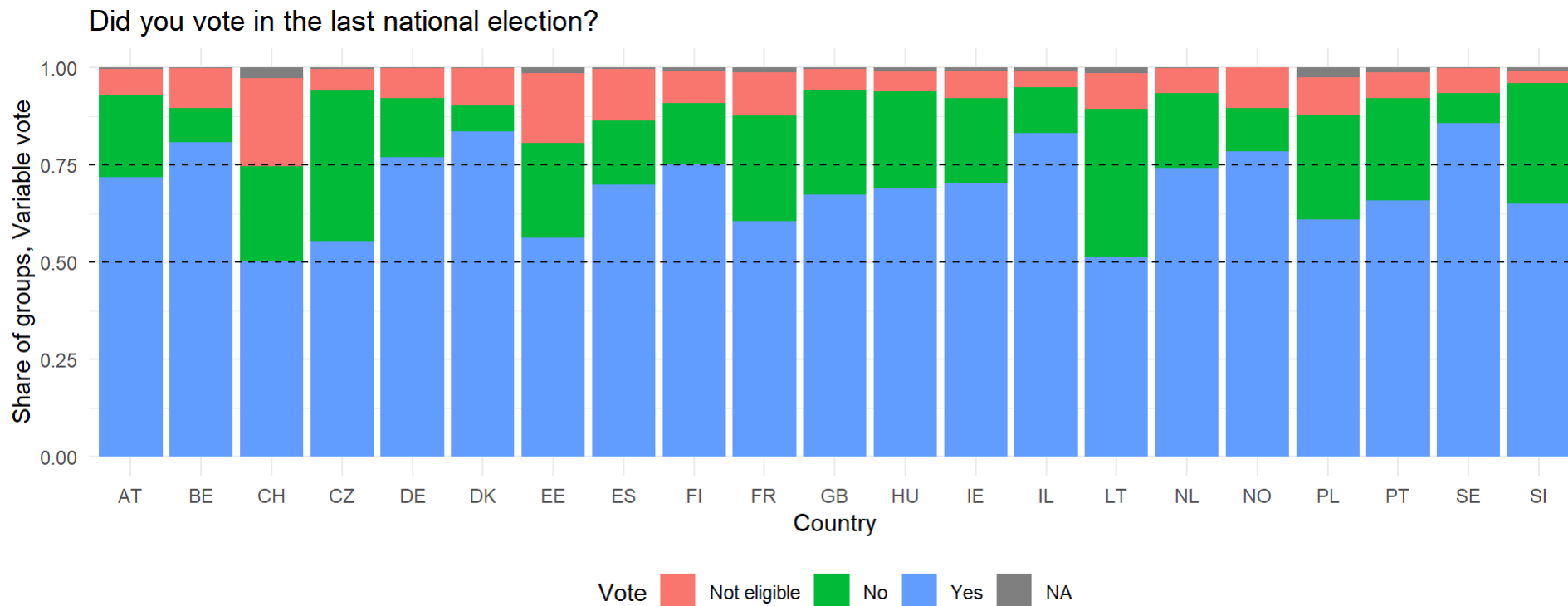
# Subjektivität und mentale Abkürzungen

Beispiel: Schätzungen von Bevölkerungsanteilen

# Subjektivität und mentale Abkürzungen

# Soziale Situation der Befragung

- mögliche Messfehler durch **soziale Erwünschtheit**
  - z.B. Wahlbeteiligung und Wahlverhalten





# Soziale Situation der Befragung

- Situation der Befragung (& Umfragemodus) machen einen Unterschied für soziale Erwünschtheit! (z.B. [Valentim 2021](#))
  - Fragestellung
  - Verhalten des Interviewers
  - self-completion vs. Interview
  - Fragetechniken

# Zeitgebundenheit

- Abbild des **Zeitpunkts der Erhebung**
- und nicht rückwirkend erhebbar

z.B.: Steiner et al. (n.d.) zeigen, dass Studierende nach der russischen Invasion in der Ukraine

- pro-europäischer
- interessierter an europäischer Politik und
- überzeugter von den Vorteilen der EU für ihr Land

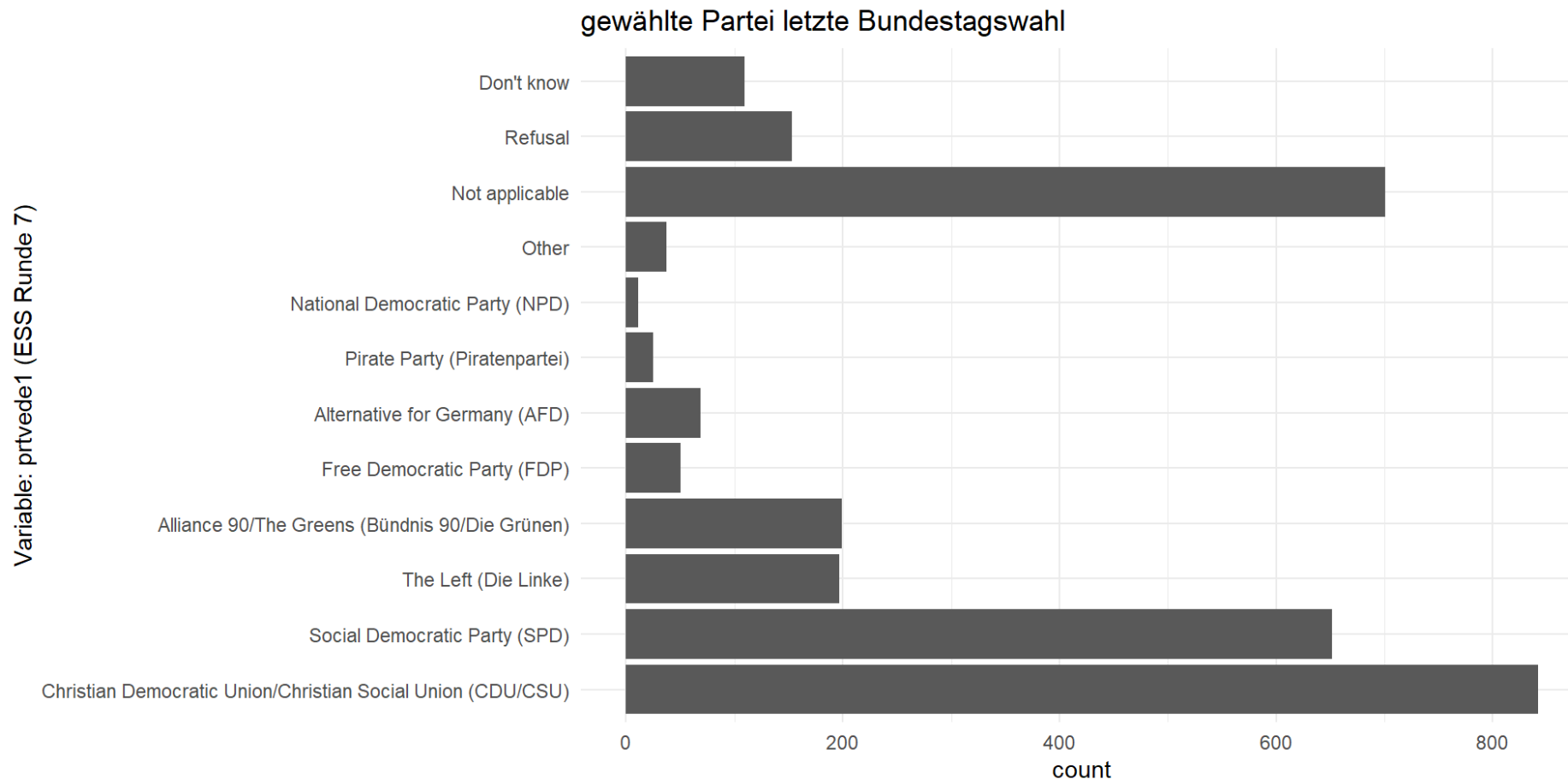
sind.

# Unvollständigkeit

- Umfragen enthalten oft (viele) fehlende Observationen
  - z.B. weil Fragen unverständlich sind
  - z.B. weil Themen unangenehm sind
- welche Werte fehlen ist häufig kein Zufall
  - z.B. soziale Erwünschtheit für bestimmte Antwortkategorien (Wahlentscheidung, sexuelle Orientierung, ...)

# Unvollständigkeit

z.B.: Wer gibt die Parteipräferenz nicht an?



→ Wir werden dem Problem des Umgangs mit fehlenden Daten immer wieder begegnen

# Schwächen von Umfragedaten

Umfragedaten - als Selbstauskunft in einer bestimmten sozialen Situation - haben **Schwächen, die wir kennen sollten**

- Subjektivität
- Aufmerksamkeit / mentale Abkürzungen
- Soziale Erwünschtheit
- Zeitgebundenheit
- Unvollständigkeit

# Aber: Gestaltungsmöglichkeiten

Umfragen bieten thematisch viele Möglichkeiten:

- für Nutzer:innen: umfassendes Spektrum an potenziellen Themen im Vergleich zu z.B. Zensus-Daten
  - ...wenn jemand die Daten bereits erhoben hat
  - ...solange es ein valides Instrument gibt
- für Umfrageforscher:innen: Forschung zur Umfragegestaltung
  - z.B. Frageformulierung mit Pre-Tests, Eyetracking, ...
  - z.B. Umfragemodi

# Arten von Fragen

# Offene und geschlossene Fragen

- **Offene Fragen:** Fragen ohne vorgegebene Antwortkategorien → Aufzeichnung der Antwort

## Vorteile

- mehr Information
- Interpretation der befragten Person
- Gegenstück: **geschlossene & halboffene Fragen**
  - Vorgabe aller oder einiger Antwortkategorien

## Nachteile

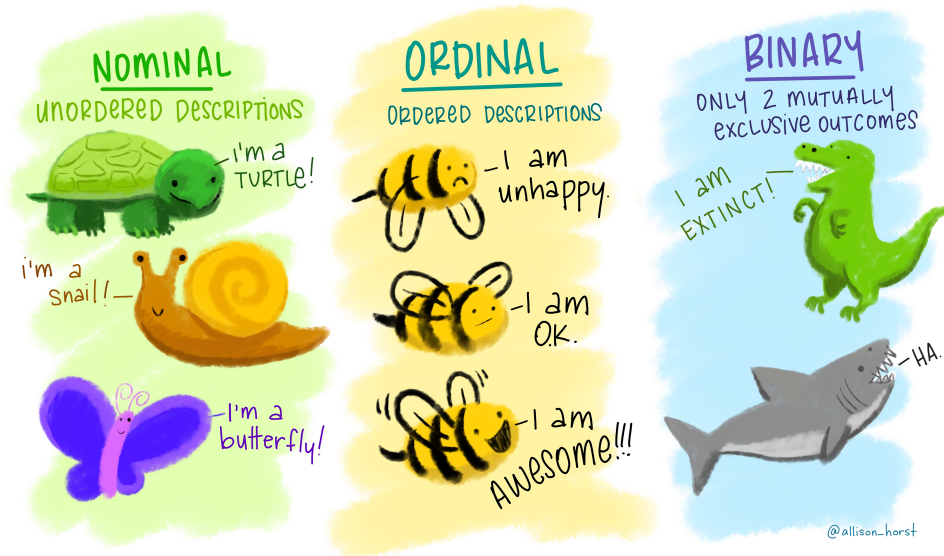
- schwierige Analyse mit standardisierten Methoden
- Antworten geben nicht immer die gewünschte Information



# Variablen

- Fragen können sich aber auch in der Art der Information unterscheiden
  - z.B. *Wie alt sind Sie?* ↔ *Zu welcher Generation gehören Sie?*
- Die Art der Frage bestimmt die *Skala* der Variablen, die wir anschließend messen
  - die *Skala* bestimmt welche Berechnungen wir anstellen können

# Variablen



Artwork by @allison\_horst

→ siehe auch Tausendpfund (2018), 119-124

- Nominalskala: keine Reihenfolge
- Ordinalskala: Reihenfolge, ohne klare Abstände
- (binär): 2 Zustände, exklusiv
- metrische Skala: Reihenfolge & klare Abstände
  - auch Intervall und Ratioskalen (mit echtem Nullpunkt) genannt

# Variablen

## CONTINUOUS

MEASURED DATA, can have  $\infty$  values within possible range.



I AM 3.1" TALL  
I WEIGH 34.16 grams

## DISCRETE

OBSERVATIONS can only exist at LIMITED VALUES, OFTEN COUNTS.



I HAVE 8 LEGS  
and  
4 SPOTS!

@allison\_horst

- diskret: nur ganze Zahlen (z.B. Anzahl)
- kontinuierlich: auch Dezimalzahlen

Artwork by @allison\_horst

# Variablen

Auf welchem Niveau werden die folgenden Variablen gemessen?

- Herkunftsland
- Größe des Zimmers in  $m^2$
- Zahl der Personen, die man wöchentlich sieht
- Bildungsgrad
- Nettoeinkommen
- Art des Jobs
- subjektiver sozialer Status

# Gütekriterien

# Was macht eine Umfrage eigentlich gut?

Zwei wissenschaftliche Standards: Reliabilität & Validität  
(Gehring and Weins ([2010](#)), Kap. 3)

# Reliabilität

- **Reliabilität als Zuverlässigkeit einer Messung**
  - Gibt es Messfehler?
  - Ist das Messinstrument konsistent (über Befragte und Zeit)?

z.B. standardisierte Fragestellung, hochqualitative Übersetzung, ...

# Validität

- **Validität als Gültigkeit** einer Messung
    - ist das gemessene das, was wir messen wollten?
  - z.B. Parteiidentifikation vs. Wahl einer Partei
    - langfristige Zugehörigkeit vs. einmalige Wahlentscheidung
- Die Frage nach der Wahl einer Partei kann das Konzept Parteiidentifikation nicht valide messen



# Die European Social Survey

# Fakten

- alle zwei Jahre durchgeführte repräsentative Umfrage
  - seit 2002
- repräsentativ und mit genug Befragten, um auch Aussagen über einzelne Länder zu treffen
- breites Themenspektrum (sozial & politisch)

# Länder

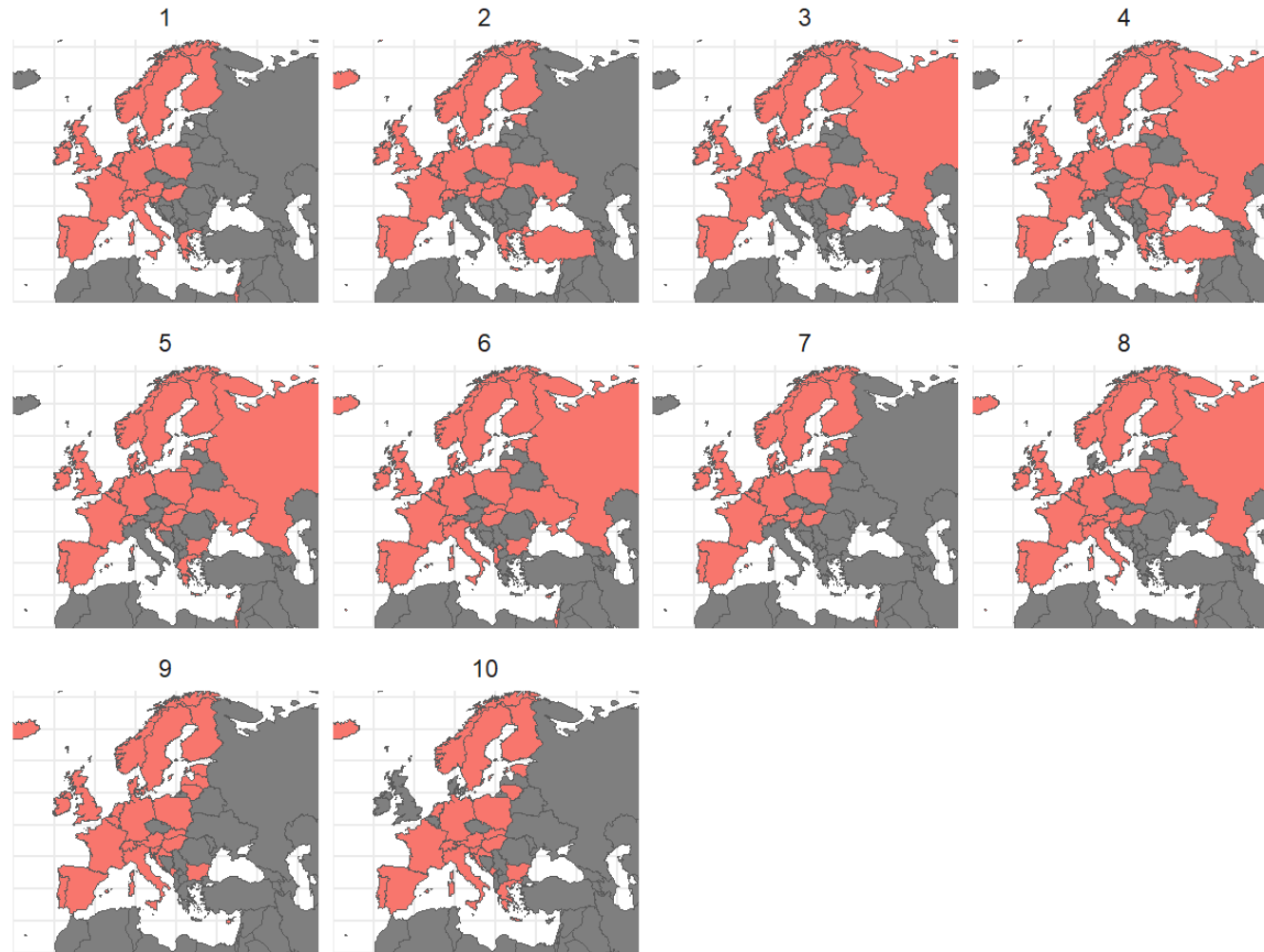


Figure 1: Länder nach ESS-Runde

# Themen: Aufgabe

Machen Sie sich mit dem [Fragebogen](#) der aktuellen European Social Survey vertraut: Welche Themen gibt es darin? Was könnte Sie interessieren? ([deutsche Version](#))

Sie können auch nach Themen suchen, die nur in einem speziellen Jahr gefragt wurden ([Rotationsmodule](#), wie z.B. unser Klimathema).

→ bis zur sechsten Sitzung entwerfen wir daraus eine Forschungsfrage.

# Aufgabe: Dateien öffnen und die Survey ansehen

Wir arbeiten im Kurs v.a. mit der achten Runde (2016/2017), in der es u.a. um Einstellungen zum Klima geht.

- speichern Sie das File der Survey (s. Moodle) im Kurs-Ordner
- öffnen mit dem `read_dta()` Befehls aus dem `haven` Paket
  - das wichtigste ist der richtige Dateipfad: finden Sie mit `getwd()` ihren aktuellen Ordner und geben Sie den Pfad im Verhältnis dazu an ([mehr Infos zu Dateipfaden](#))

# Aufgabe: Dateien öffnen und die Survey ansehen

Die Daten sind eigentlich für Stata (ein anderes Statistikprogramm) geschrieben → wir verwenden das Paket **haven**, um die Daten zu importieren

```
1 getwd()  
2 library(tidyverse)  
3 library(haven)  
4 ess8 <- read_dta("../data/ESS8e02_2.dta")
```

# Aufgabe: Dateien öffnen und die Survey ansehen

Probieren Sie einige Befehle aus der R-Einführungssitzung:

→ Siehe Aufgabenblatt

# Nächste Woche: Daten zusammenfassen: Mittelwerte und Streuungen

**Thema:** Wie können wir Daten zusammenfassen und Zusammenhänge zwischen Variablen erkennen?

- Llaudet and Imai ([2023](#)), S. 51-89
- optional: Gehring and Weins ([2010](#)), S. 120-139 zum nachlesen über Konzepte auf deutsch
- optional mehr Zusammenhangsmaße: Gehring and Weins ([2010](#)), 141-174



# References

- Gehring, Uwe W., and Cornelia Weins. 2010. *Grundkurs Statistik für Politologen und Soziologen*. Wiesbaden: VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-91879-2>.
- Llaudet, Elena, and Kosuke Imai. 2023. *Data Analysis for Social Science: A Friendly and Practical Introduction*. Princeton: Princeton University Press.
- Schnell, Rainer. 2019. *Survey-Interviews: Methoden standardisierter Befragungen*. Studienskripten zur Soziologie. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-531-19901-6>.
- Steiner, Nils D., Ruxanda Berlinschi, Etienne Farvaque, Jan Fidrmuc, Philipp Harms, Alexander Mihailov, Michael Neugart, and Piotr Stanek. n.d. “Rallying Around the EU Flag: Russia’s Invasion of Ukraine and Attitudes Toward European Integration.” *JCMS: Journal of Common Market Studies* n/a (n/a). Accessed December 21, 2022. <https://doi.org/10.1111/jcms.13449>.
- Tausendpfund, Markus. 2018. *Quantitative Methoden in der Politikwissenschaft*. Wiesbaden: Springer Fachmedien. <https://doi.org/10.1007/978-3-658-20698-7>.
- Valentim, Vicente. 2021. “Parliamentary Representation and the Normalization of Radical Right Support.” *Comparative Political Studies*, March, 0010414021997159. <https://doi.org/10.1177/0010414021997159>.

