

# Wissenschaftliches Arbeiten und Forschungsmethoden

**Einheit 9: Auswertung von Studien - Teil 2: Ergebnisse,  
Datenvizualisierung und Rückmeldung**

27.06.2024 | Dr. Caroline Zygar-Hoffmann

# Heutige Themen

Hypothesen und Hypothesentests

Datenvisualisierung

Rückmeldung

# Hypothesen und Hypothesentests

## Generell

- Die Auswertung der Daten erfolgt nach den Vorgaben des Analyseplans in der Präregistrierung
- Im Mittelpunkt der Auswertung hypothesenprüfender Untersuchungen stehen statistische Signifikanztests
- Ausgang der Tests ist die Entscheidungsgrundlage dafür, ob die forschungsleitende Hypothese als bestätigt geltend oder abgelehnt werden soll
- Die inhaltliche Interpretation der Ergebnisse nimmt auf die Theorie Bezug, aus der die Hypothese abgeleitet wurde

# Hypothesen und Hypothesentests

## Prinzip

1) Austellen von Nullhypothese und Alternativhypothese (Hypothesenpaar)

2) Bestimmung einer zugrundeliegenden Verteilung

3) Festlegung des Annahme- und Ablehnungsbereichs der Nullhypothese (kritischer Wert)

→ Schritte 1-3 haben Sie bereits in der Präregistrierung abgehandelt, durch die Aufstellung einer (Alternativ-)Hypothese und der Wahl der statistischen Analyse und Inferenzkriterien (z.B. Alpha-Niveau) zur Überprüfung dieser Hypothese

→ Zwischen Schritt 3 und 4 könnte noch die Prüfung der Voraussetzungen für die Berechnung einer Teststatistik mit einer bestimmten statistischen Analyse eingefügt werden (selbst wenn sie nur diskutiert und keine Konsequenz daraus abgeleitet wird)

4) Berechnung Teststatistik (Beobachtungswert auf Wahrscheinlichkeitsverteilung abbilden): z.B. z-Wert, t-Wert, F-Wert,  $\chi^2$ -Wert, ...

5) Vergleich kritischer Wert und Teststatistik

6) Entscheidung: Test signifikant oder nicht signifikant

→ Schritte 4-6 werden durch die Analyse in R und Interpretation des Ergebnisses abgehandelt

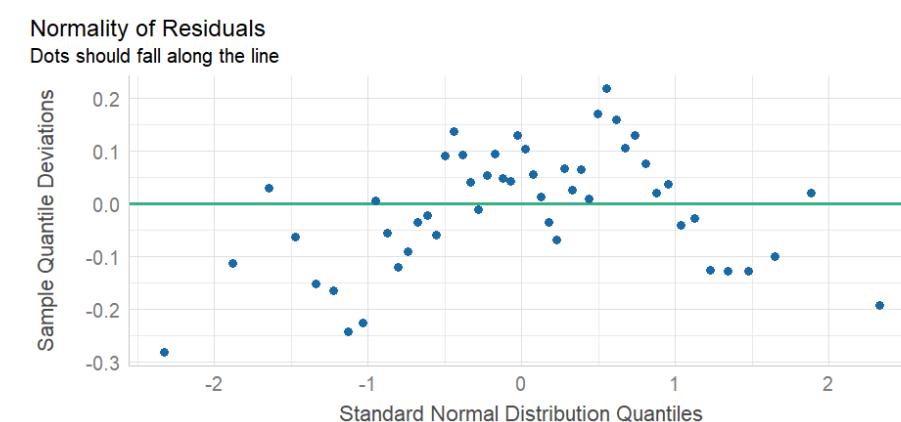
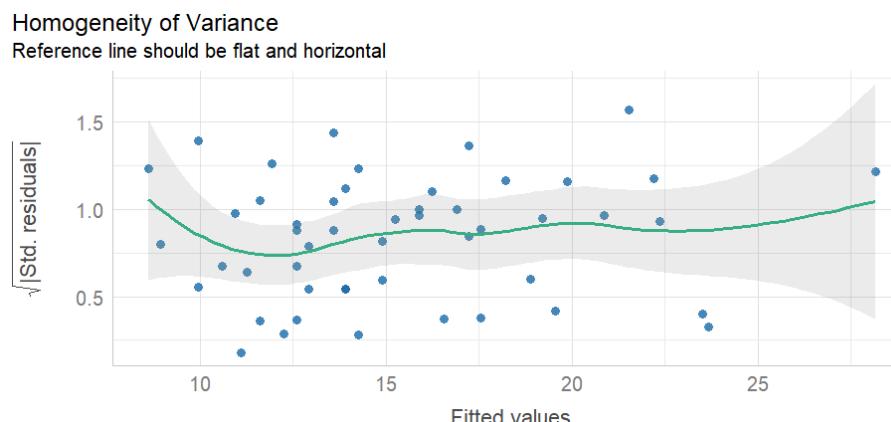
# Hypothesen und Hypothesentests

## Prüfung von Voraussetzungen

- Beim t-test z.B. durch `leveneTest()` aus dem package `car`
- Bei der Regression und ANOVA z.B. durch `check_model()` aus dem package `easystats` (oder mit dem R-Code aus der Vorlesung)

Beispiel:

```
library(easystats)
check_model(lm(y ~ x, data = data))
```



# Hypothesen und Hypothesentests

## Interpretation

- Vorsicht: Immer auch prüfen, ob der Effekt in die richtige Richtung geht, falls die Richtung der Hypothese nicht bereits im Test selbst hinterlegt (Beispiel: Bei Regression muss Richtung überprüft werden, bei t-test wird Richtung im Analysebefehl festgelegt durch das Argument **alternative**)
- Signifikante Ergebnisse (in die richtige Richtung) bestätigen (vorläufig) die Hypothese
- Nicht-signifikante Ergebnisse schränken ihren Geltungsbereich ein oder können auf geringe statistische Power zurückgeführt werden
- Advanced: Neben dem frequentistischen Nullhypotesentesten, gibt es sogenannte "bayesianische" Analysen mit denen die Evidenz für die Nullhypothese quantifiziert werden kann (durch sogenannte "Bayes Faktoren")

→ Wir werden uns in Sitzung 10 wie genau Analyse-Ergebnisse in der schriftlichen Arbeit im Fließtext berichtet werden

# Hypothesen und Hypothesentests

## Konfirmation vs. Exploration

- Die Ergebnisse von Auswertungen, die über die eigentliche Hypothesenprüfung hinausgehen, sind explorativ  
→ müssen auch in dieser Weise dargestellt werden (in einem extra Abschnitt im Ergebnisteil)
- Prinzipiell kann in der Exploration jede Analyse gerechnet werden, die einen interessiert
- Vorsicht: Je mehr Analysen man rechnet, desto wahrscheinlicher wird es ein Falsch-Positives Ergebnis zu erwischen → p-hacking
- **Illustration p-hacking App:** <https://shinyapps.org/apps/p-hacker/> → Wenn man genug ausprobiert, schafft man es ein signifikantes Ergebnis zu "generieren", ...
  - ... welches dann aber erstmal nicht vertrauenswürdig ist und in einer separaten Studie konfirmatorisch repliziert werden sollte
  - ... man explorativ aber trotzdem berichten kann
- Man kann und sollte aber auch explorative Analysen berichten, bei denen kein signifikantes Ergebnis rauskam, wenn diese Analysen naheliegend sind (man muss aber auch nicht jede explorative Analyse berichten)

# Datenvisualisierung

## Ausgewählte Ziele der Datenvisualisierung:

- Ergebnisse veranschaulichen
- Ausreißer erkennen
- Rückmeldungen erstellen

## Bewertungsschema:

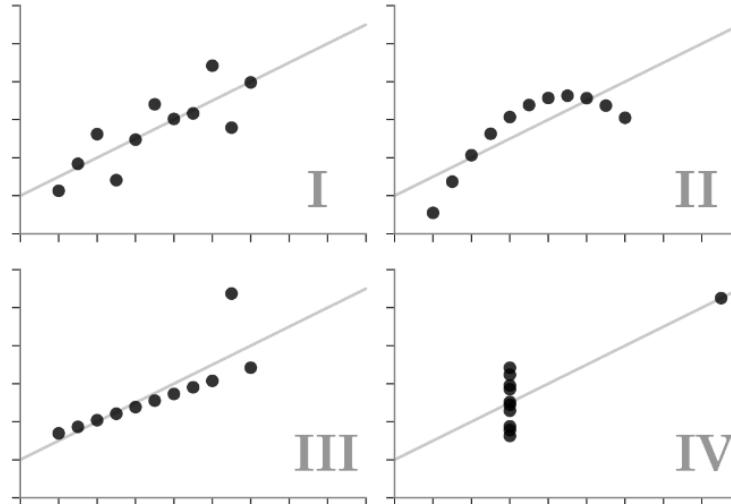
<b>(Formale) Darstellung der Ergebnisse (nach APA-Stil)</b>	<b>0</b>	Ergebnisse sind teilweise unvollständig oder mit formalen Mängeln dargestellt (z.B. fehlende Teststatistik, fehlender p-Wert, fehlerhafte Beschreibung)
	<b>1</b>	Ergebnisse sind formal fehlerfrei, aber unübersichtlich oder nicht komplett nach APA-Stil dargestellt (z.B. falsche Rundung, fehlende Effektstärke, fehlende Freiheitsgrade)
	<b>2</b>	Ergebnisse sind formal fehlerfrei, (überwiegend) nach APA-Stil und übersichtlich dargestellt
	<u><b>Bonus +1</b></u>	<u><b>Bonus:</b></u> Ergebnisse sind hilfreich grafisch veranschaulicht

# Datenvisualisierung

Hinter den gleichen Statistiken können sich unterschiedliche Muster verbergen, z.B. bei Anscombe's Quartett:

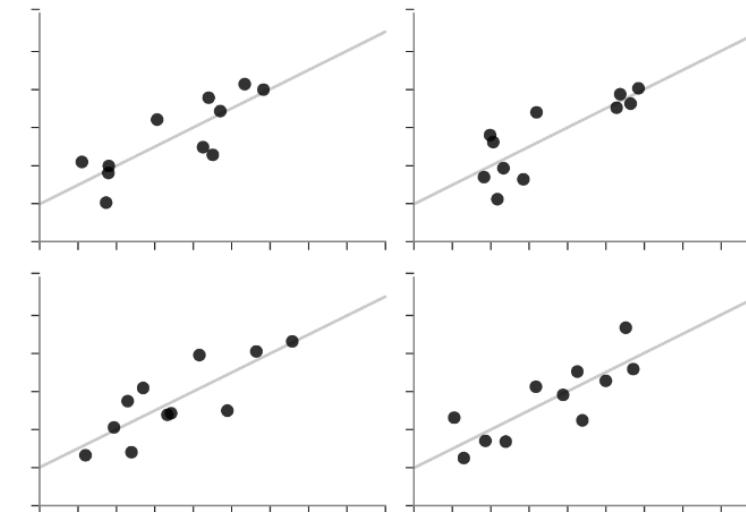
## Anscombe's Quartet

Each dataset has the same summary statistics (mean, standard deviation, correlation), and the datasets are *clearly different*, and *visually distinct*.



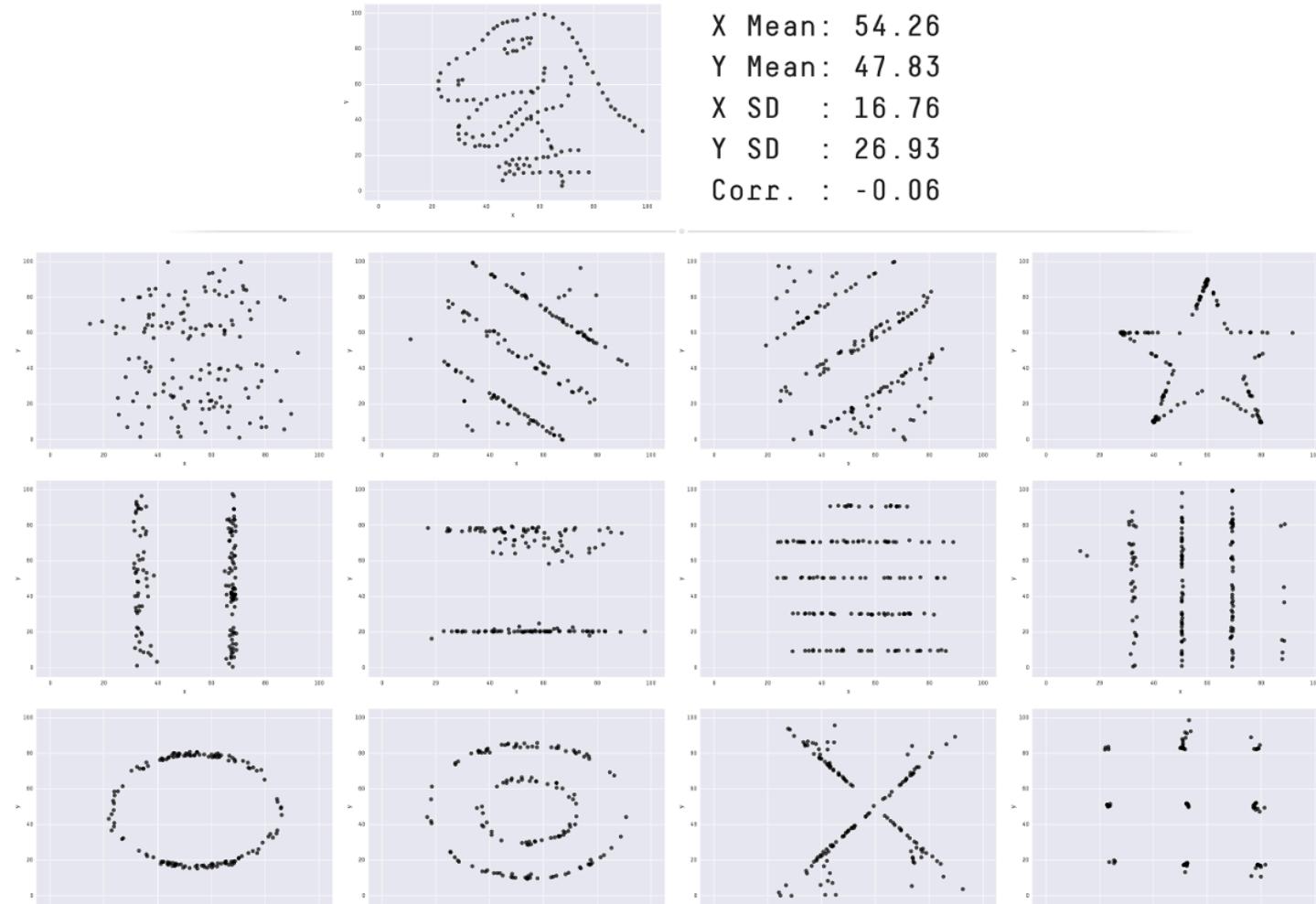
## Unstructured Quartet

Each dataset here also has the same summary statistics. However, they are not *clearly different* or *visually distinct*.



→ die Datenvisualisierung deckt auf, was für ein Muster hinter den Daten liegt

# Datenvisualisierung



# Datenvisualisierung

## Do's:

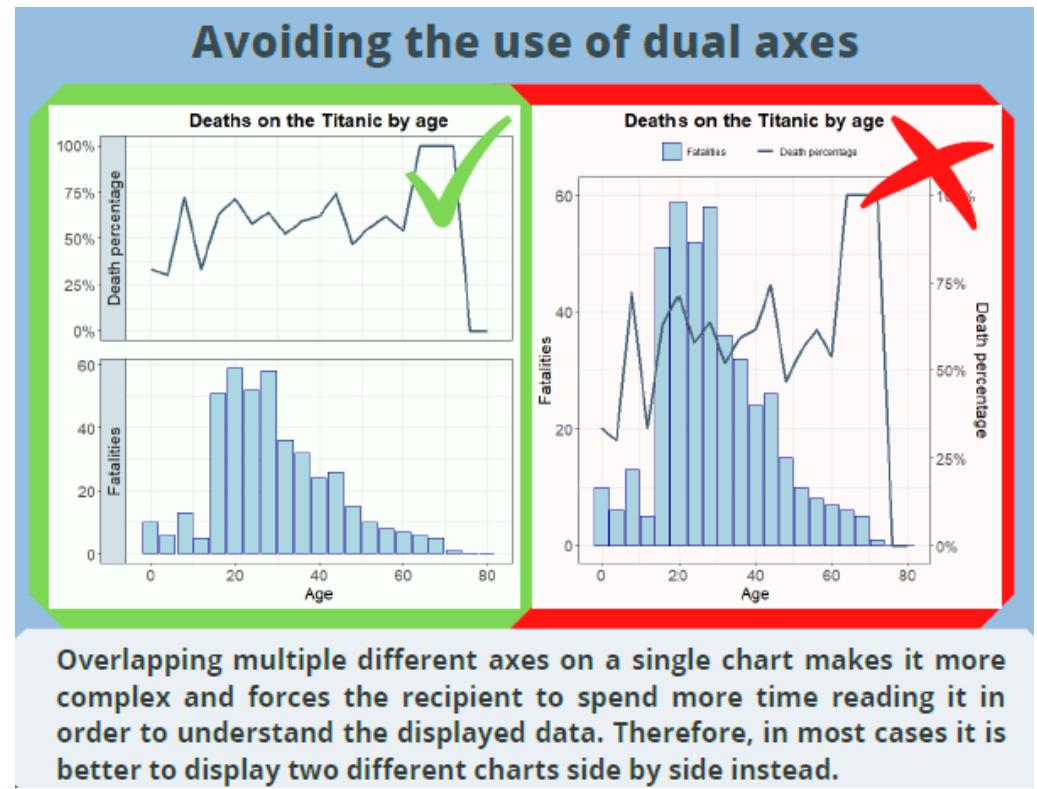
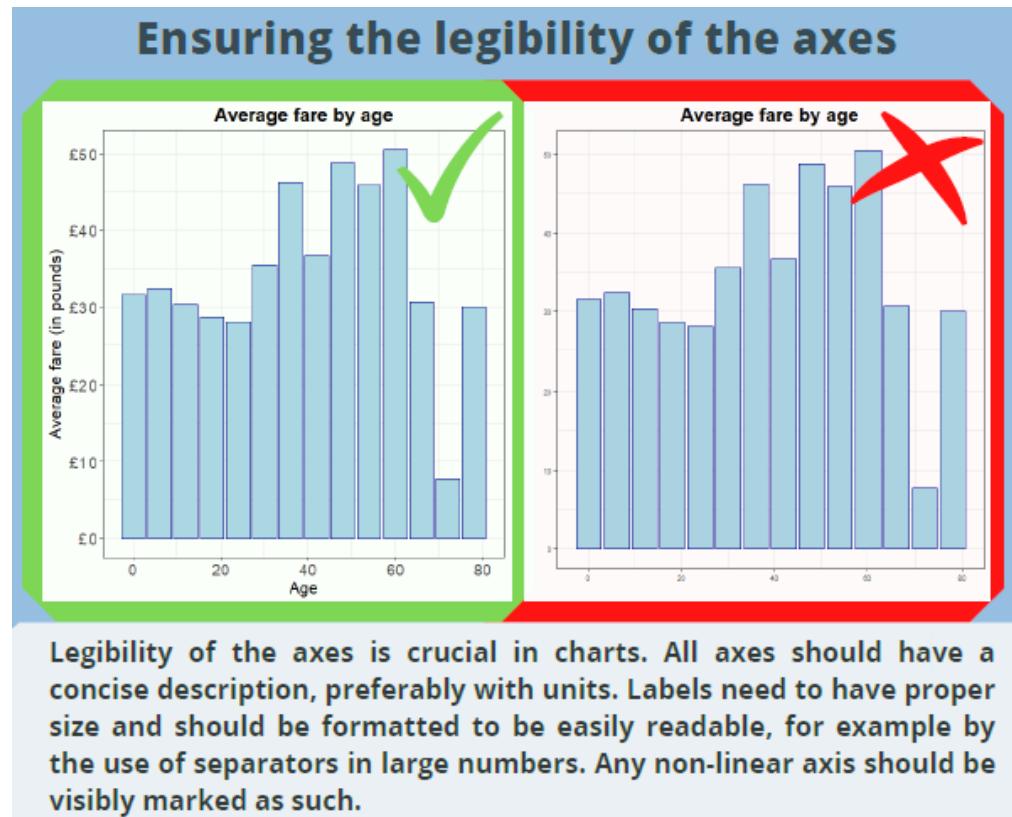
- Verständliche Achsenbeschriftung
- Gesamte Grafik sollte "stand-alone" (d.h. ohne zusätzlichen Text) verstehbar sein (Erklärungen als Anmerkungen unter die Grafik)
- schwarz-weiß kompatible Verständlichkeit (d.h. man sollte die Grafik nicht in Farbe ausdrucken müssen, damit man sie versteht)
- Rohwerte als Punkte in der Grafik aufnehmen
- Interaktionen bei verschiedenen Ausprägungen der Variablen darstellen

## Dont's:

- Wertebereich der y-Achse einschränken
- rot-grüne Farbpalette wegen Farbenblindheit
- Kreisdiagramme (Flächen werden intuitiv nicht adequat wahrgenommen)

# Datenvisualisierung

## Do: Verständliche Achsenbeschriftung



# Datenvisualisierung

## Don't: Wertebereich der y-Achse einschränken

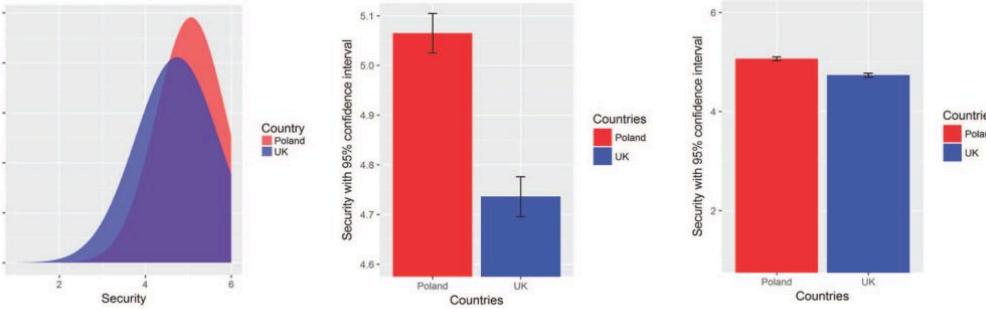
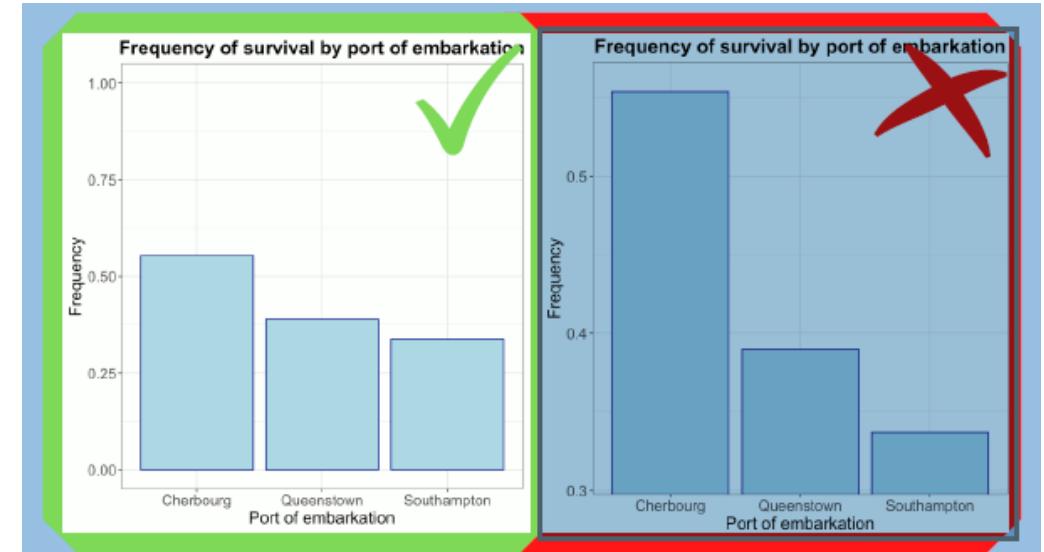
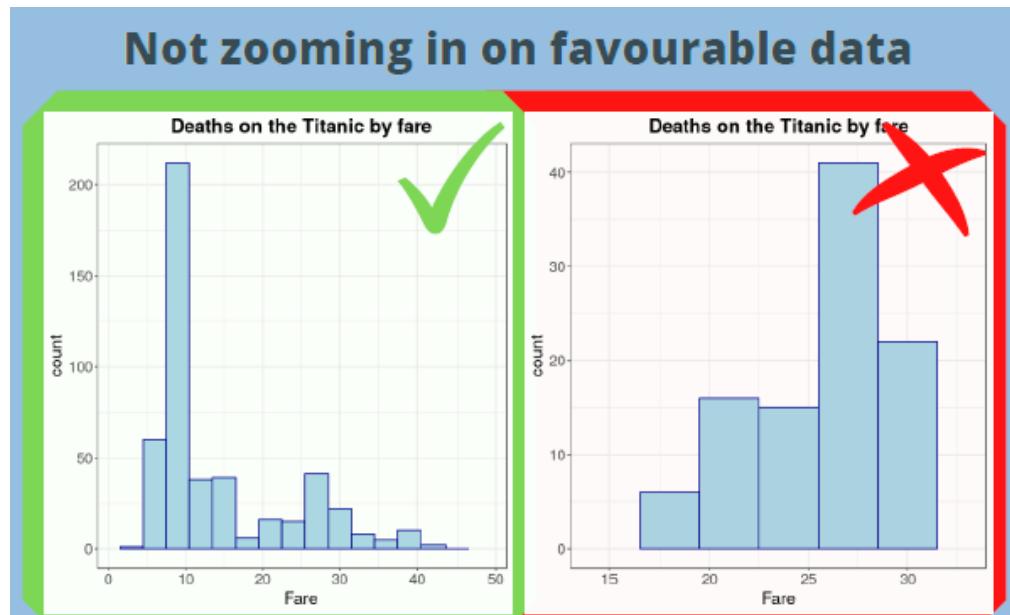


Figure 7. Three modes of depicting the same data: two superimposed normal distributions and two barplots



# Datenvisualisierung

## Weitere Hinweise zu Achsen



Not showing the entire or at least the trimmed dataset can maliciously suggest trends antithetical to the ones actually present in the data. Any zoom that leaves data out should be clearly marked as such.



With direct comparisons, using a different scale on any axis makes it difficult to easily compare data, since for any comparison the reader would be forced to read the values off the axes instead of easily comparing the different shapes.

# Datenvisualisierung

## Farben

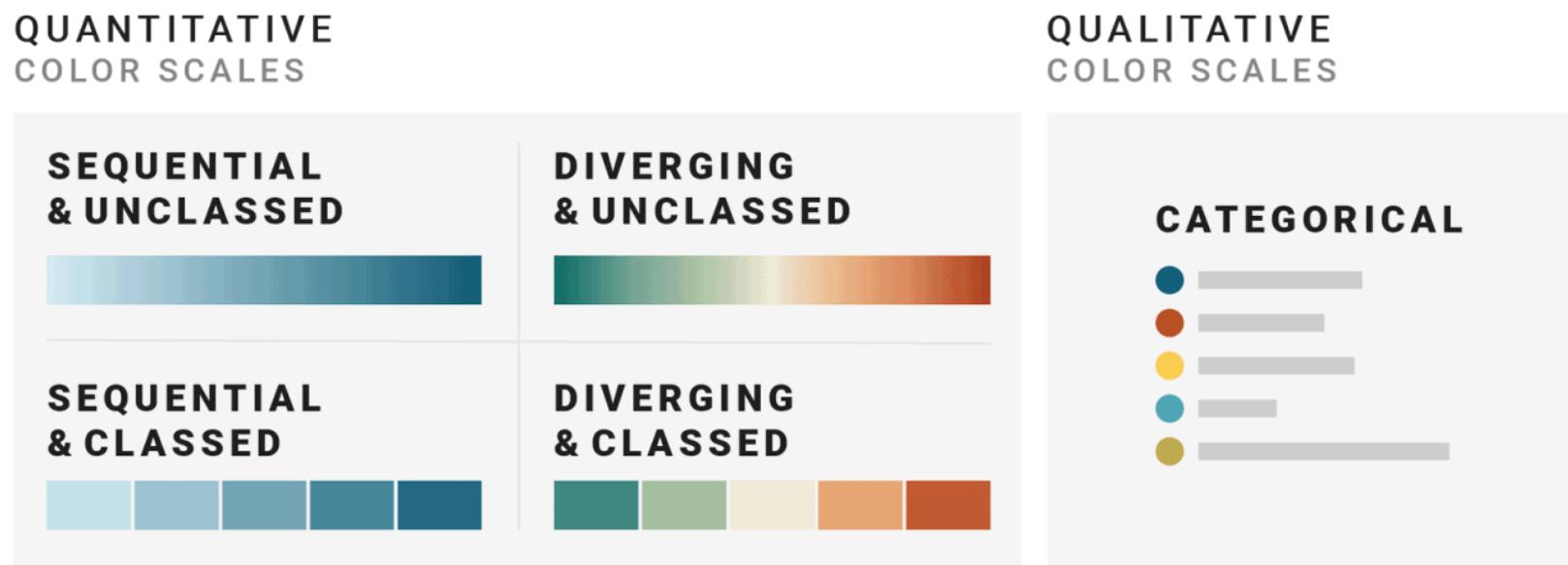


# Datenvisualisierung

## Farben

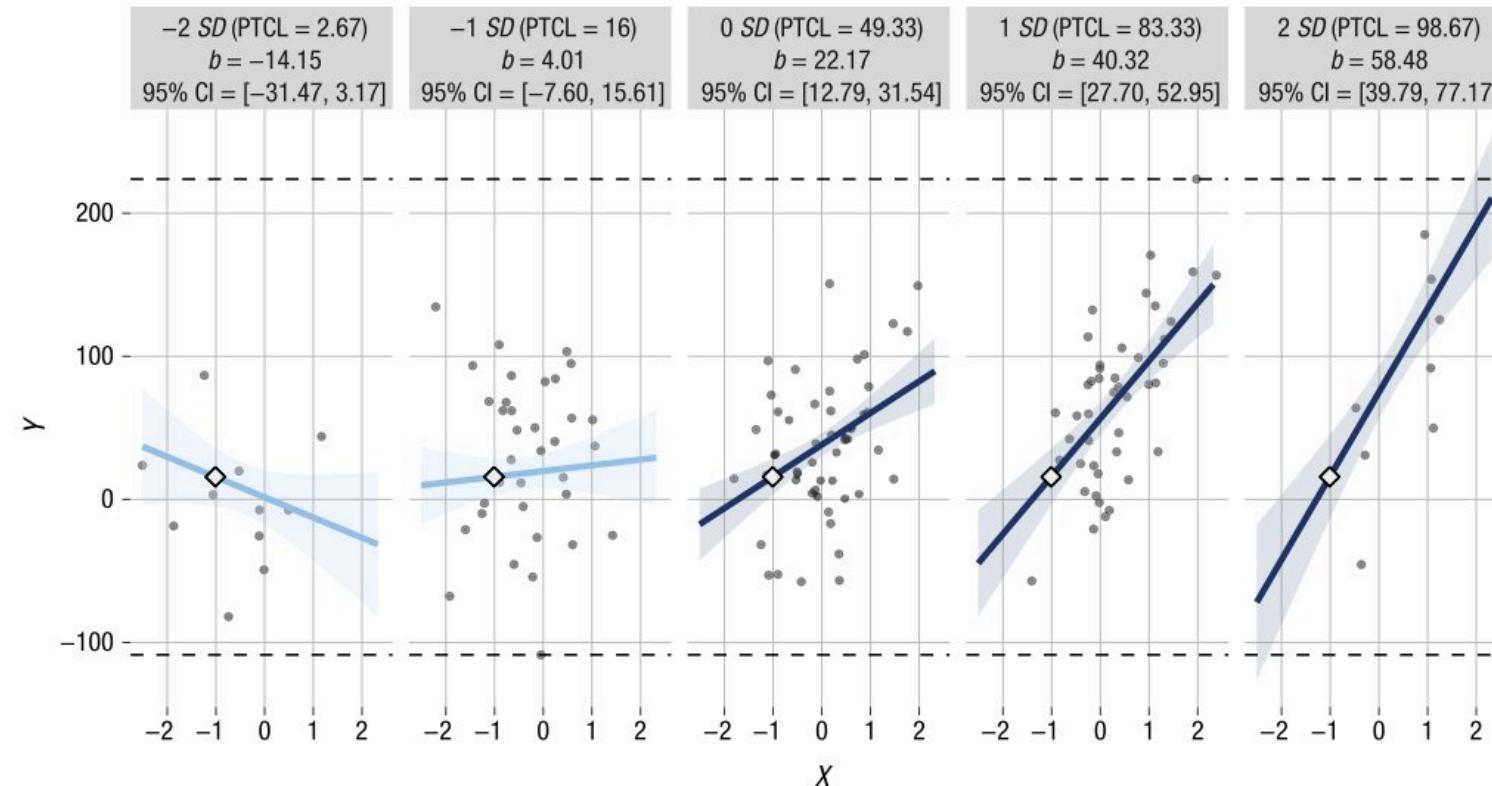
Die gleiche(n) Farbe(n) in unterschiedlicher Intenstität ("quantitative Farbskalen") oder mehrere unterschiedliche Farben ("qualitative Farbskalen")?

- <https://blog.datawrapper.de/quantitative-vs-qualitative-color-scales/>
- <https://blog.datawrapper.de/diverging-vs-sequential-color-scales/>



# Datenvisualisierung

## Do: Interaktionen bei verschiedenen Ausprägungen der Variablen darstellen

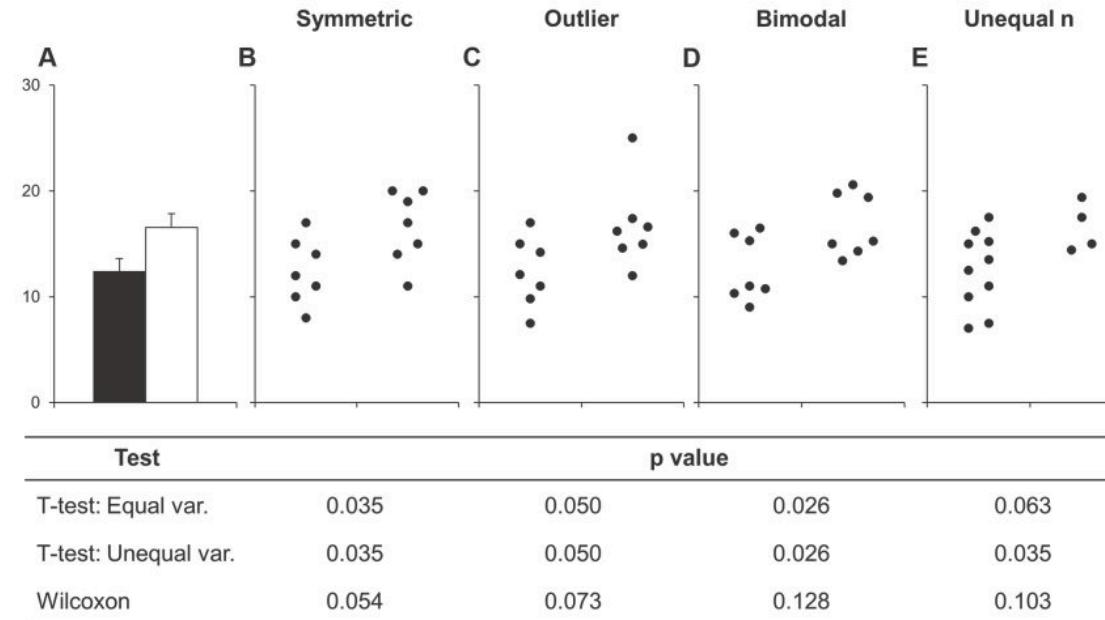


**Fig. 6.** Illustration of small multiples created by interActive for Example 1 using multivariate normal predictors. Simple slopes are provided for levels of the moderator 2 SD and 1 SD below the mean, at the mean, and 1 SD and 2 SD above the mean. Each graphic shows the computed 95% confidence region (shaded area), the observed data (gray circles), the maximum and minimum values of the outcome (dashed horizontal lines), and the crossover point (diamond). The x-axes represent the full range of the focal predictor. CI = confidence interval; PTCL = percentile.

# Datenvisualisierung

## **Suboptimal: Balkendiagramme zur Darstellung der Mittelwerte zweier oder mehrerer Gruppen**

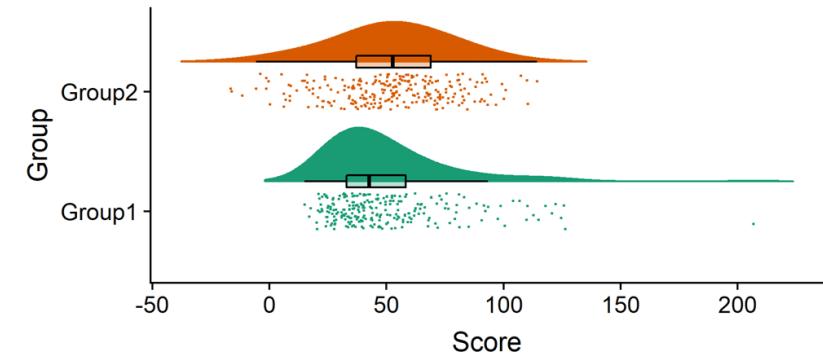
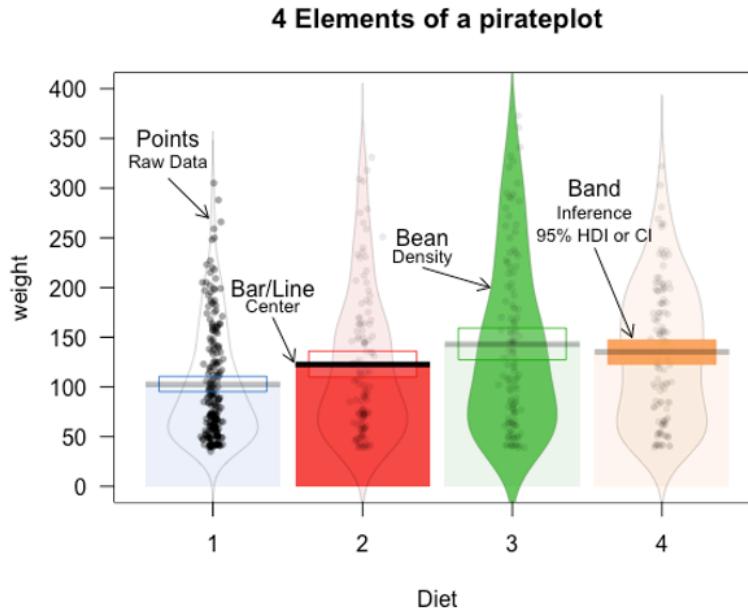
Balkendiagramme transportieren keine Information zur Verteilung der Daten innerhalb der Gruppen:



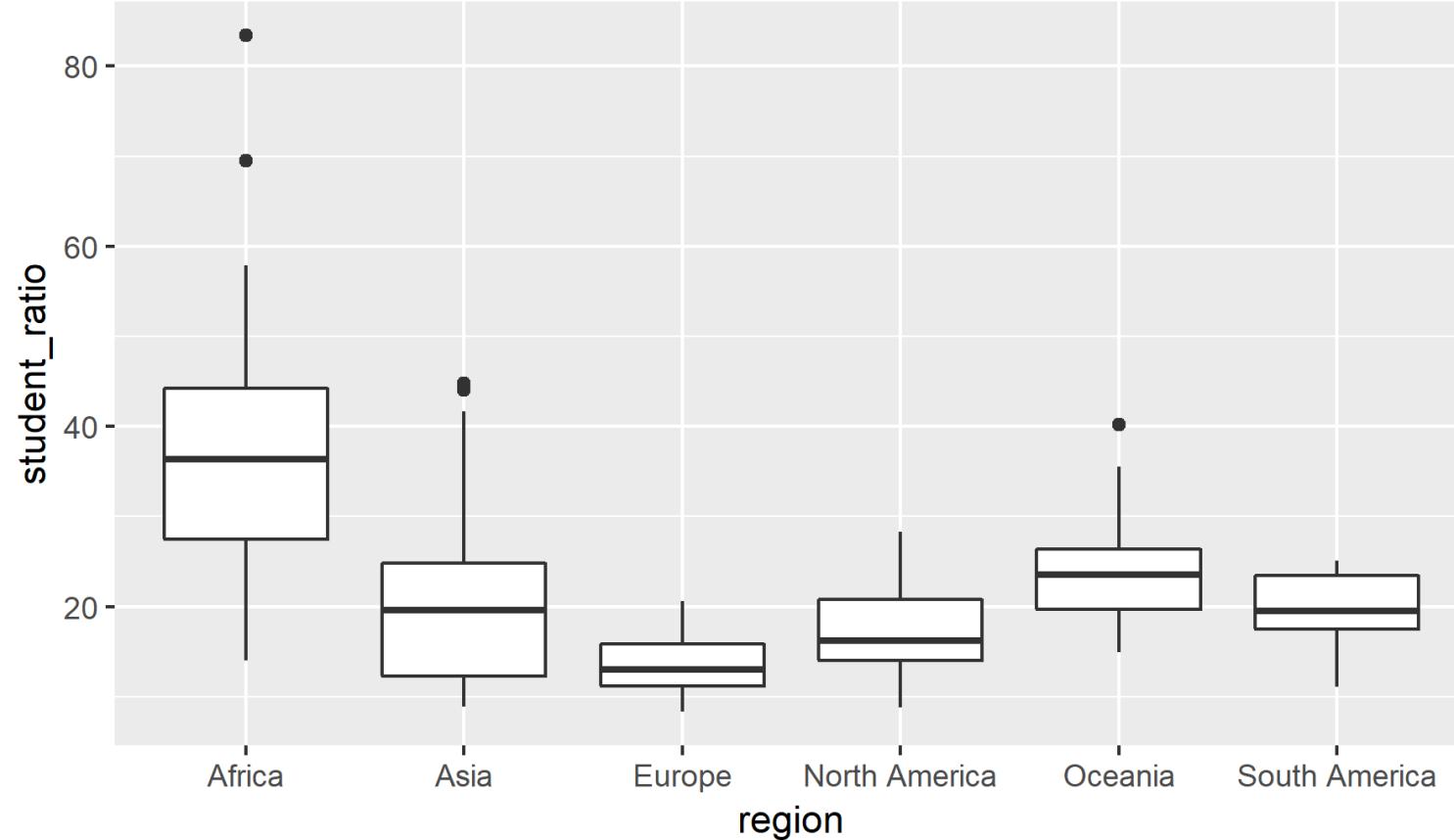
**Fig 1. Many different datasets can lead to the same bar graph.** The full data may suggest different conclusions from the summary statistics. The means and SEs for the four example datasets shown in Panels B–E are all within 0.5 units of the means and SEs shown in the bar graph (Panel A). p-values were calculated in R (version 3.0.3) using an unpaired t-test, an unpaired t-test with Welch's correction for unequal variances, or a Wilcoxon rank sum test. In Panel B, the distribution in both groups appears symmetric. Although the data suggest a small difference between groups, there is substantial overlap between groups. In Panel C, the apparent difference between groups is driven by an outlier. Panel D suggests a possible bimodal distribution. Additional data are needed to confirm that the distribution is bimodal and to determine whether this effect is explained by a covariate. In Panel E, the smaller range of values in group two may simply be due to the fact that there are only three observations. Additional data for group two would be needed to determine whether the groups are actually different.

# Datenvisualisierung

Bessere Alternative zur Darstellung der Mittelwerte zweier oder mehrerer Gruppen: Violin-Plots (auch "Pirate-Plots", linke Abbildung) oder Raincloud-Plots (rechte Abbildung)



## The Evolution of a ggplot



Data: UNESCO Institute for Statistics  
Visualization by Cédric Scherer

# Datenvisualisierung

## Hilfreiche Links für häufige Datenvisualisierungen in R:

- Histogramme, Boxplots, Dichteverteilungen, heatmaps, jitter plots: <https://medium.com/@nickmartin812/how-to-r-visualizing-distributions-49ea4141fb32#rstats>
- Pirate plots:
  - [https://dmyee.files.wordpress.com/2016/03/pirateplots\\_workshop.pdf](https://dmyee.files.wordpress.com/2016/03/pirateplots_workshop.pdf)
  - [https://indrajeetpatil.github.io/ggstatsplot/articles/web\\_only/ggbetweenstats.html](https://indrajeetpatil.github.io/ggstatsplot/articles/web_only/ggbetweenstats.html)
  - <https://cedricscherer.netlify.app/2019/05/17/the-evolution-of-a-ggplot-ep.-1/>
- Raincloud plots:
  - <https://wellcomeopenresearch.org/articles/4-63>
  - <https://www.cedricscherer.com/2021/06/06/visualizing-distributions-with-raincloud-plots-and-how-to-create-them-with-ggplot2/>
- Scatterplots für Korrelationen / einfache Regressionen:
  - [https://indrajeetpatil.github.io/ggstatsplot/articles/web\\_only/ggscatterstats.html](https://indrajeetpatil.github.io/ggstatsplot/articles/web_only/ggscatterstats.html)
  - <https://r-graphics.org/chapter-scatter> und <https://r-graphics.org/recipe-scatter-fitlines>

# Datenvisualisierung

## Hilfreiche Links für häufige Datenvisualisierungen in R:

- Interaktionen:
  - <https://cran.r-project.org/web/packages/interactions/vignettes/interactions.html>
  - [https://strengejacke.github.io/sjPlot/articles/plot\\_interactions.html](https://strengejacke.github.io/sjPlot/articles/plot_interactions.html)
  - <https://journals.sagepub.com/doi/10.1177/2515245917746792>
- Zu allen Themen:
  - <https://clauswilke.com/dataviz/index.html>
  - <https://psyteachr.github.io/msc-data-skills/ggplot.html>
  - <https://psyteachr.github.io/hack-your-data/quant-data-vis.html>

# Rückmeldung

## Elemente

- Angabe des Messinstruments mit dem gemessen wurde
- Inhaltliche Beschreibung des Maßes
- Ergebnis als verbalisiertes, klassifizierendes Intervall unter Einbeziehung der Normstichprobe

# Rückmeldung

## Angabe des Messinstruments und Beschreibung des Maßes

Beispiel Selbstwert, gemessen über Multidimensionale Selbstwertskala (MSWS, Schütz, & Rentzsch, 2020), Subskala ESW - Emotionale Selbstwertschätzung

Personen mit hohen Werten auf dieser Subskala:

- haben eine positive Einstellung zu sich selbst.
- hohe Selbstakzeptanz und Selbstzufriedenheit
- positive Einstellungen sowie Gefühle gegenüber der eigenen Person

Personen mit niedrigen Werten auf dieser Subskala:

- sind unzufrieden mit sich
- erleben Selbstzweifel
- haben eine negative Einstellung zu sich selbst
- leiden unter negativen selbstbezogenen Gefühlen.

# Rückmeldung

## Normierung

- Der einzelne Summenwert (z.B. einer Subskala) steht für starke oder weniger starke Ausprägung
- hohe Werte = hohe Ausprägung

## ABER:

- Der Wert alleine gibt uns keine Informationen über die **Einordnung** der Ausprägung
- Müssen Wert **im Vergleich** zu Werten von anderen Personen (Normen) interpretieren
- Dafür nutzen wir möglichst große, vergleichbare **Normstichproben** (Tabellen hinten im Manual eines Tests)

## Beispiel MSWS:

Empirischer Befund: Männer und Frauen unterscheiden sich signifikant in der Selbstwerteinschätzung Darum haben die Autor:innen des Fragebogens 2 getrennte Normtabellen erstellt (eine für Männer und eine für Frauen)

# Rückmeldung

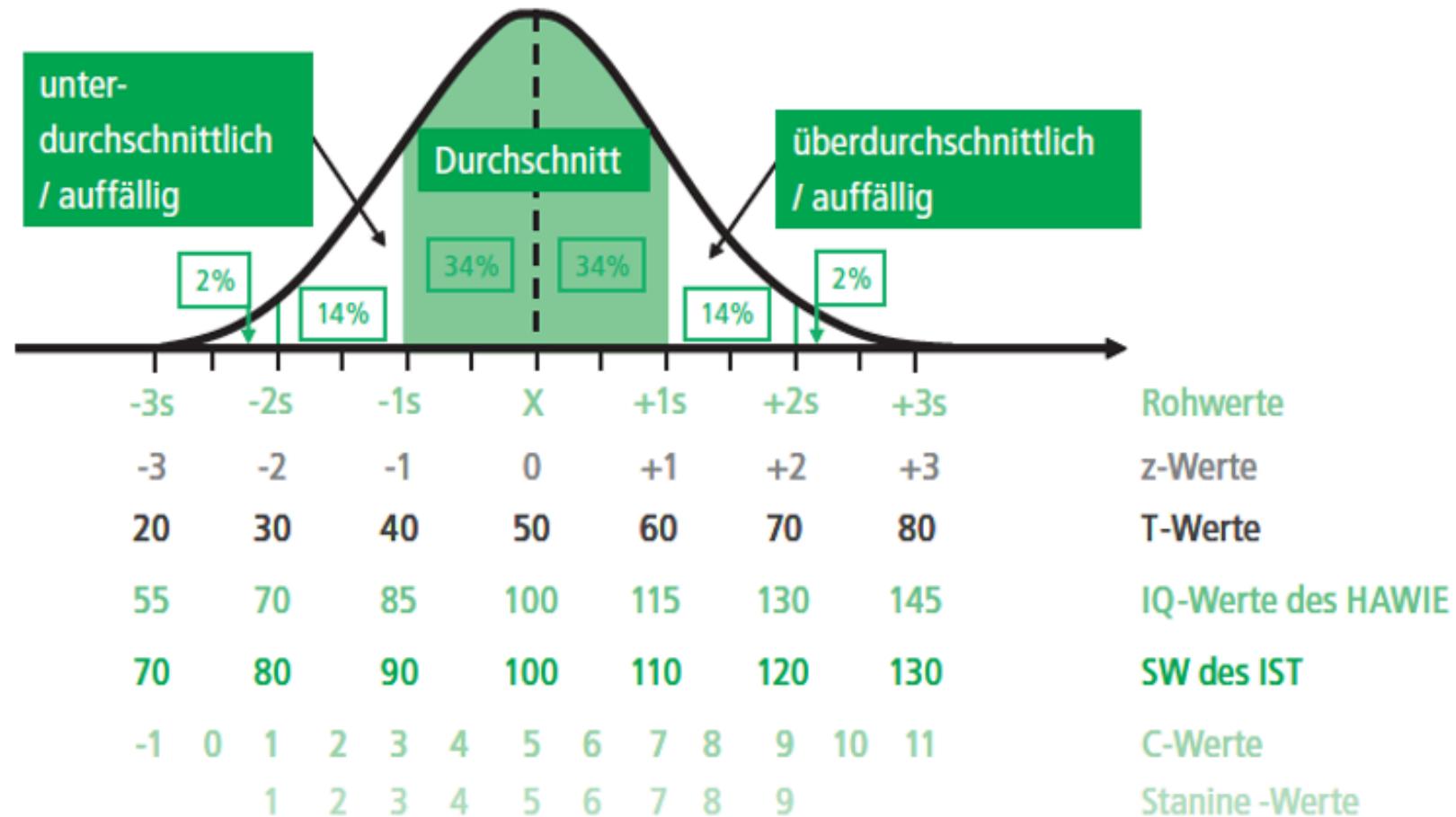
## Normierung

- Normwerte erlauben eine **Verortung des Messwertes** innerhalb der Verteilung der Werte der Normstichprobe
- z.B. Wo liegt der ESWS Wert 32 im Vergleich zu anderen Personen gleichen Geschlechts (Berücksichtigung von Mittelwert und Standardabweichung der Verteilung)
- erlaubt eine **Einteilung** in "durchschnittlich" bzw. "unter-/überdurchschnittlich" → Wichtig für eingeordnete Rückmeldung an die Testperson
- z.B. **z-Wert:** normiert auf 0 mit Standardabweichung von 1 (es gibt auch andere Normwerte mit anderem Mittelwert und Standardabweichung, z.B. bei MSWS sogenannte T-Werte mit Mittelwert 50 und Standardabweichung 10)

z.B. Eine Person mit Körpergröße 175 (angenommen dies ist der Durchschnitt der Norm) und Wert Körbergewicht 70 kg (angenommen dies ist der Durchschnitt der Norm) haben beide einen z-Wert von 0 und einen T-Wert von 50 (obwohl Variablen in unterschiedlichen Einheiten gemessen).

# Rückmeldung

## Normierung



# Rückmeldung

## Normierung

Bei uns gibt es zwei Möglichkeiten:

- 1) Entweder Sie finden für die Skala, für die Sie eine Rückmeldung anbieten, vorhandene Normen und ziehen diese heran
- 2) Oder Sie nutzen die Stichprobe der Studie als Normstichprobe, vergleichen also eine Versuchsperson mit den restlichen Versuchspersonen (aus Pragmatismus, empfehle ich Ihnen diese Variante) → Dann können Sie einfach den in der Stichprobe berechneten z-Wert zur Interpretation heranziehen (siehe Vorlesungseinheit Auswertung1)

**In der Regel sollte die Normstichprobe in der Rückmeldung detailliert beschrieben werden.** Dies ist bei Variante 2 nur eingeschränkt möglich, da wir aus Anonymitätsgründen wenige Informationen über die Stichprobe erhoben haben.

# Rückmeldung

## Interpretation

- Basis der Interpretation sind Konfidenzintervalle zur Berücksichtigung der Unsicherheit bei der Messung
- 95% Konfidenzintervall für einen einzelnen z-Wert  $z$ :  $z \pm 1.96 * \sqrt{(1 - \text{Reliabilität der Skala})}$
- Zur Reliabilität der Skala → siehe Vorlesungeinheit Auswertung2 (entweder aus vorheriger Literatur oder selbst auf eigenen Daten berechnen)
- Bei einem einzelnen Item (statt einer Skala, die aus mehreren Items besteht) kann nur eine Retest-Reliabilität berechnet werden (auf Ihren Daten nicht möglich) → Rückmeldung von Skalenwerten wünschenswert!

# Rückmeldung

## Interpretation

### Unterdurchschnittliche Ausprägung:

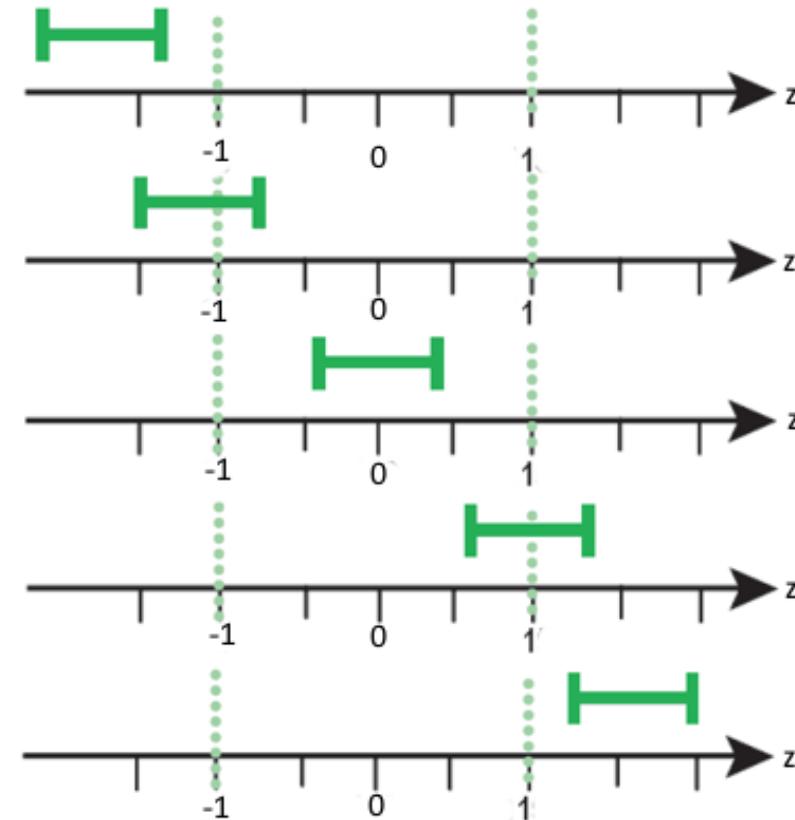
- Beide Grenzen des Konfidenzintervall kleiner als -1 Standardabweichung
- z.B. [-1.8, -1.2] oder [-3.0, -2.0]

### Unterdurchschnittliche bis durchschnittliche Ausprägung:

- Konfidenzintervall zwischen kleiner als -1 Standardabweichung und unter Mittelwert
- z.B. [-1.2, -0.2] oder [-2.0, -0.8]

### Durchschnittliche Ausprägung:

- Konfidenzintervall enthält den Mittelwert oder über/unterschreitet 1 Standardabweichung nicht
- z.B. [-0.6, 0.6] oder [-0.6, -0.2] oder [0.2, 0.6]



# Rückmeldung

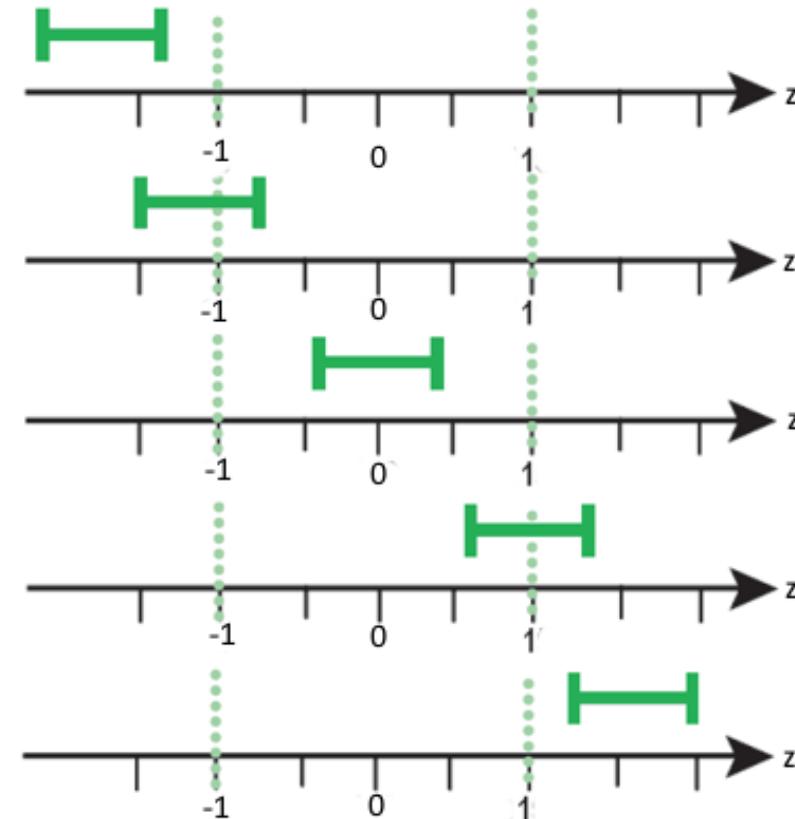
## Interpretation

### Durchschnittliche bis überdurchschnittliche Ausprägung:

- Konfidenzintervall zwischen größer als 1 Standardabweichung und über Mittelwert
- z.B. [0.2, 1.2] oder [0.8, 2.0]

### Überdurchschnittliche Ausprägung:

- Beide Grenzen des Konfidenzintervall größer als 1 Standardabweichung
- z.B. [1.2, 1.8] oder [2.0, 3.0]



## Schritt 1: Rückmeldung vorbereiten

- Rmarkdown Vorlage zur Rückmeldung "[template\\_rueckmeldung.Rmd](#)" von studynet herunterladen
- Ganz oben bei "title" den Titel Ihrer Studie einfügen
- Angabe des Messinstruments und Beschreibung des Maßes anpassen
- R-Code anpassen:
  - Zeile 7, 10 und 39: Wie heißt Ihr vorverarbeiteter Datensatz, den Sie für die Analysen einlesen werden (im Template ist es `final_data`)?
  - Bei dem Code zum Balkendiagramm, bei den z-Werten, und bei den Konfidenzintervallen: Wie heißt die Spalte mit der z-standardisierten Skala, zu der Sie eine Rückmeldung geben möchten (im Template ist es `skala_z`)?
  - Überall wo `sqrt(1-0.8)` steht (Teil der Bildung des Konfidenzintervalls): Wie ist die Reliabilität Ihrer Skala (im Template ist es 0.8)?

## Schritt 2: Rückmeldung testen

- In Zeile 10 ein Pseudonym einfügen, das auch wirklich in Ihren Daten existiert
- Rmarkdown knitten, entstandenes html prüfen

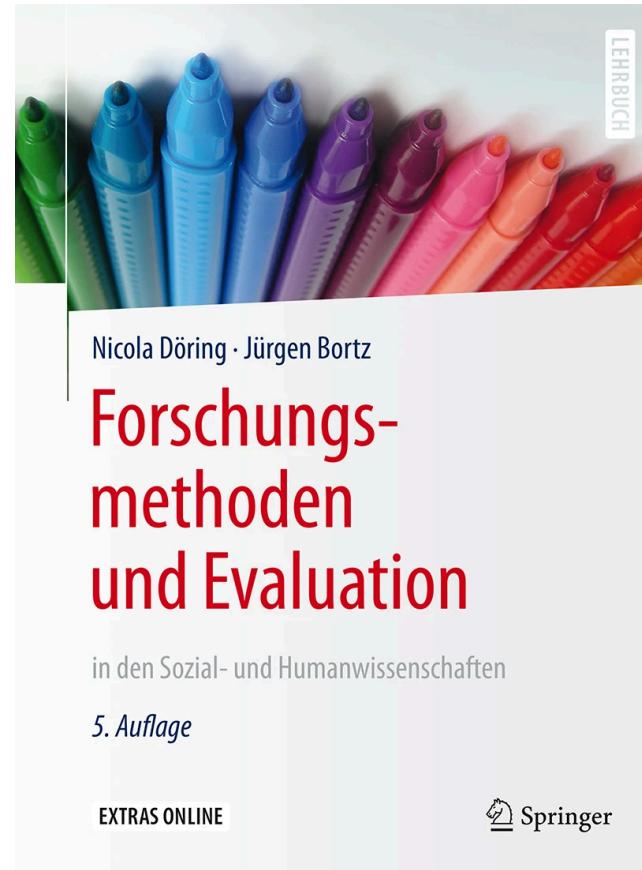
## Schritt 3: Für alle Versuchspersonen eine Rückmeldung erstellen

- Entweder manuell in Zeile 10 immer wieder: 1) das Pseudonym austauschen 2) neu knitten 3) erhaltene Datei umbenennen (sonst wird sie überschrieben -> umbenennen in das Pseudonym, z.B. `X1235ufk.html`)
- Oder das R-Skript auf studynet "`template_rueckmeldung_automate.R`" ausprobieren (da bei großen Stichproben das manuelle Vorgehen ewig dauern würde)
  - im R-Skript `template_rueckmeldung_automate.R` muss auch der Name für Ihren vorverarbeiteten Datensatz angepasst werden (`final_data`)
  - im Rmarkdown-Skript `template_rueckmeldung.Rmd` muss dann Zeile 10 auskommentiert werden (mit einem `#`), dafür Zeile 11 nicht mehr (das `#` löschen)
  - Beide Skripte müssen im selben Ordner liegen

## Schritt 4: Rückmeldungen auf studynet hochladen und Komiliton:innen informieren

- Im studynet Ordner "Rückmeldungen" einen Unterordner mit dem Titel der eigenen Studie anlegen und dort die HTML Dateien hochladen
- Komiliton:innen informieren, dass sie sich dort Ihre Rückmeldungsdatei herunterladen können (Identifizierung anhand des Dateinamens = das Pseudonym)

# Literaturempfehlung für die heutige Sitzung



Kapitel 12.2 in Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Pearson.