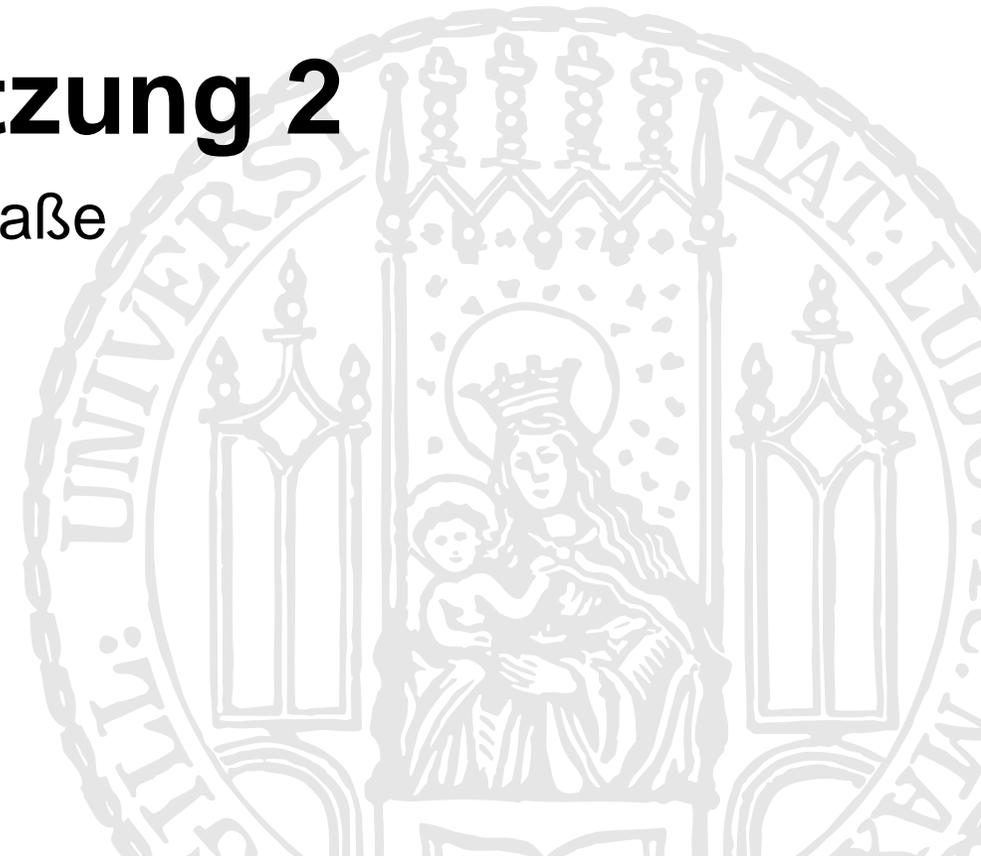


Datenanalyse – Sitzung 2

Häufigkeiten und Lagemaße

Institut für Kommunikationswissenschaft und Medienforschung
Ludwig-Maximilians-Universität München



Ablauf der Sitzung

1. Warum Daten bereinigen?
2. Übung 1: Datenbereinigung
3. Warum die Berechnung von Häufigkeiten und Lagemaßen erlernen?
4. Kurze Wiederholung: Häufigkeiten und Lagemaße
5. Übung 2: Häufigkeiten und Lagemasse manuell berechnen
6. Häufigkeiten und Lagemaße in R berechnen
7. Übung 3: Übung 2 in R umsetzen
8. Übung 4: Häufigkeiten und Lagemaße in R berechnen
9. Visualisierung von Häufigkeiten in R

1. WARUM DATEN BEREINIGEN?

Ziele der Datenbereinigung

- **Datenqualität sichern:** die Bereinigung von Daten ist für deren Qualität entscheidend
- **Verbesserung der Analysegenauigkeit:** durch das Entfernen oder Korrigieren von fehlerhaften Daten werden statistische Analysen präziser
- **Effiziente Datenanalyse:** bereinigte Daten vereinfachen den Analyseprozess, indem sie die Komplexität reduzieren

Wann und warum Datenbereinigung?

Rohdaten aus einer Befragung weisen oft **Fehler und Probleme** auf, z.B.:

- „**Abbrecher**“: Personen, die den Fragebogen an einer Stelle abgebrochen haben
- „**Durchklicker**“: Personen, die den Fragebogen zwar zu Ende, aber unsauber ausgefüllt haben (passiert vor allem dann, wenn mit Incentives gelockt wurde, bspw. einem Gewinnspiel am Ende) – einige Indikatoren dafür sind:
 - Viele fehlende Werte
 - Sehr schnelles Ausfüllen (Fragebogensoftware weist oft die Gesamtdauer oder die Verweildauer auf einzelnen Seiten in eigener Variable aus)
 - Muster im Antwortverhalten (tritt vor allem bei vielen gleichförmigen Skalen auf)
- **Unplausible Angaben** (z. B. Alter 233 Jahre, die sich ggf. korrigieren lassen)
- **Fehler bei der Dateneingabe** (z. B. nicht vergebene Werte)

Wie geht man mit solchen Fällen um?

Einzelfallentscheidung

- Betrifft das Problem nur einen kleinen Teil des Fragebogens, der für die Analysen eher nebensächlich ist?
 - Könnte das extreme Antwortverhalten auch die tatsächlichen Werte widerspiegeln?
 - Unplausible Antworten vielleicht nur durch Tippfehler entstanden (Geburtsjahr 2980 statt 1980) oder durch „unernstes“ Ausfüllen? In ersten Fall kann man den Fall behalten und den Wert korrigieren oder nur diesen einen Wert entfernen.
- Fälle löschen?
 - Filtervariable erstellen, um Fälle für bestimmte Analysen auszuschließen?
 - Falsche Angaben als fehlende Werte definieren?
 - Angabe korrigieren?

2. ÜBUNG 1: DATENBEREINIGUNG

Übung 1: Datenbereinigung

Sie haben 8 Studierende bzw. Dozierende auf einer 5-stufigen Likert-Skala (1=sehr schlecht, 5=sehr gut, **NA=keine Angabe**) zu ihrer Meinung zum Produkt „Super-Duper-R-Handbuch“ und dem Kaufinteresse (1=ja, 2=vielleicht, 3=nein) gefragt. Folgende Daten haben Sie dafür in R eingetragen:

Befragte:r	Rolle	Alter	Meinung	Kaufinteresse	Dauer des Ausfüllens des Fragebogens (in Sek.)
1	Studierender	24	4	1	31
2	Studierender	29	2	3	24
3	Studierender	21	5	3	4
4	Dozierender	31	5	3	44
5	Studierender	1998	NA	1	53
6	Dozierender	35	2	2	19
7	Studierendner	19	3	2	29
8	Dozierender	33	4	4	411

- 1) Wo haben sich Fehler oder Probleme eingeschlichen?
- 2) Wie würden Sie damit umgehen?

Lösung für Übung 1: Datenbereinigung

Sie haben 8 Studierende bzw. Dozierende auf einer 5-stufigen Likert-Skala (1=sehr schlecht, 5=sehr gut, NA=keine Angabe) zu ihrer Meinung zum Produkt „Super-Duper-R-Handbuch“ und dem Kaufinteresse (1=ja, 2=vielleicht, 3=nein) gefragt. Folgende Daten haben Sie dafür in R eingetragen:

Befragte:r	Rolle	Alter	Meinung	Kaufinteresse	Dauer des Ausfüllens des Fragebogens (in Sek.)
1	Studierender	24	4	1	31
2	Studierender	29	2	3	24
3	Studierender	21	5	3	4
4	Dozierender	31	5	3	44
5	Studierender	1998	NA	1	53
6	Dozierender	35	2	2	19
7	Studierender	19	3	2	29
8	Dozierender	33	4	4	411

- 1) Wo haben sich Fehler oder Probleme eingeschlichen?
- 2) Wie würden Sie damit umgehen?

3. WARUM DIE BERECHNUNG VON HÄUFIGKEITEN UND LAGEMASSEN ERLERNEN?

Häufigkeiten und Lagemaße...

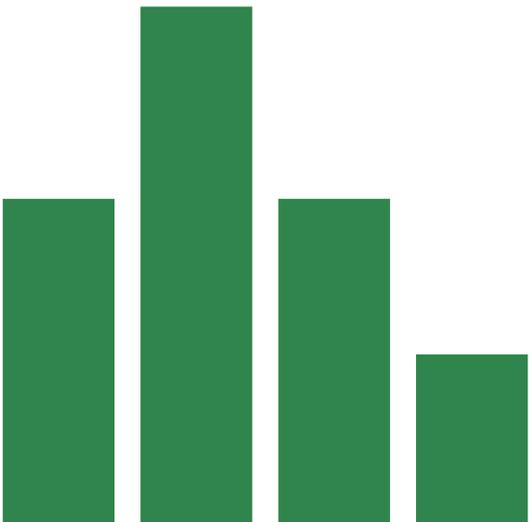
- ...sind Grundbausteine der **deskriptiven Statistik**, d.h. der erste Schritt zur Datenanalyse und -interpretation
- ...erlauben die schnelle **Einschätzung der Datenverteilung** sowie Identifizierung von Mustern, Trends und Anomalien
- ...helfen somit bei der **Aufdeckung von Fehlern** (z.B. fehlende Daten, Duplikate, oder unplausible Antworten) **und Ausreißern** in den Daten
- ...erleichtern nicht zuletzt die Entscheidungsfindung für passende Datenbereinigungsmethoden und **sichern dadurch die Datenqualität**

4. KURZE WIEDERHOLUNG: HÄUFIGKEITEN UND LAGEMASSE

Was sind Lagemaße?

Lagemaße sind statistische Kennzahlen, die beschreiben, **wo sich die Daten konzentrieren oder „liegen“**. Sie versuchen, den **typischen oder zentralen Wert eines Datensatzes** zu identifizieren.

Dazu zählen: **Modus, Median** und **arithmetisches Mittel**.



Lagemaße

- **Modus** (Modalwert)
 - Merkmalswert mit der höchsten absoluten Häufigkeit
 - für alle Daten (nominales Datenniveau oder höher)

- **Median**
 - Teilt die Daten in zwei gleiche Hälften
 - für Daten mit mindestens ordinalem Datenniveau

- **Arithmetisches Mittel** (ggf. mit Standardabweichung)
 - für intervallskalierte Daten, die näherungsweise normalverteilt sind
 - nicht geeignet für zweigipflige Verteilungen

5. ÜBUNG 2:

HÄUFIGKEITEN UND LAGEMASSE MANUELL BERECHNEN

Übung 2a

Um sich einen Überblick über die Lesefreudigkeit der Studierenden zu machen, wurde von den KW-Professor*innen eine Befragung durchgeführt. Die Studierenden mussten angeben, an wie vielen Tagen der Woche sie das Institut zum Literaturstudium betreten. Die Stichprobe lieferte folgende Ergebnisse:

3	2	5	3	1	0	3	0	2	2
3	0	2	3	0	4	2	4	5	3
4	3	2	4	4	3	3	1	3	1

Ermitteln Sie ...

- a) die absoluten Häufigkeiten
- b) Modus
- c) Median
- d) arithmetisches Mittel

Tipp: Schneller ohne Taschenrechner !

Lösung für Übung 2a

Die Stichprobe lieferte folgende Ergebnisse:

3 2 5 3 1 0 3 0 2 2
 3 0 2 3 0 4 2 4 5 3
 4 3 2 4 4 3 3 1 3 1

Ermitteln Sie ...

a) die absoluten Häufigkeiten (mittels Strichliste)

0 IIII
 1 III
 2 IIII
 3 III III
 4 III
 5 II

x	n
0	4
1	3
2	6
3	10
4	5
5	2
N	30

Lösung für Übung 2b

Ermitteln Sie ...

b) Modus

x	n
0	4
1	3
2	6
3	10
4	5
5	2
N	30

Modus: **3** (n=10)

Lösung für Übung 2c

Ermitteln Sie ...

c) Median

x	n
0	4
1	3
2	6
3	10
4	5
5	2
N	30

$$N = 30 \rightarrow \text{gerade, daher } \bar{x}_{med} = \frac{1}{2} \left(x_{\left(\frac{N}{2}\right)} + x_{\left(\frac{N}{2}+1\right)} \right)$$

$$\bar{x}_{med} = \frac{1}{2} \left(x_{\left(\frac{30}{2}\right)} + x_{\left(\frac{30}{2}+1\right)} \right)$$

$$\bar{x}_{med} = \frac{1}{2} \left(x_{(15)} + x_{(16)} \right)$$

$$\bar{x}_{med} = \frac{1}{2} (3 + 3)$$

$$\bar{x}_{med} = 3$$

Lösung für Übung 2c

Falls N gerade ist: In der Mitte teilen

$$MD = \bar{x}_{med} = \frac{1}{2} [x_{(\frac{N}{2})} + x_{(\frac{N}{2}+1)}] =$$

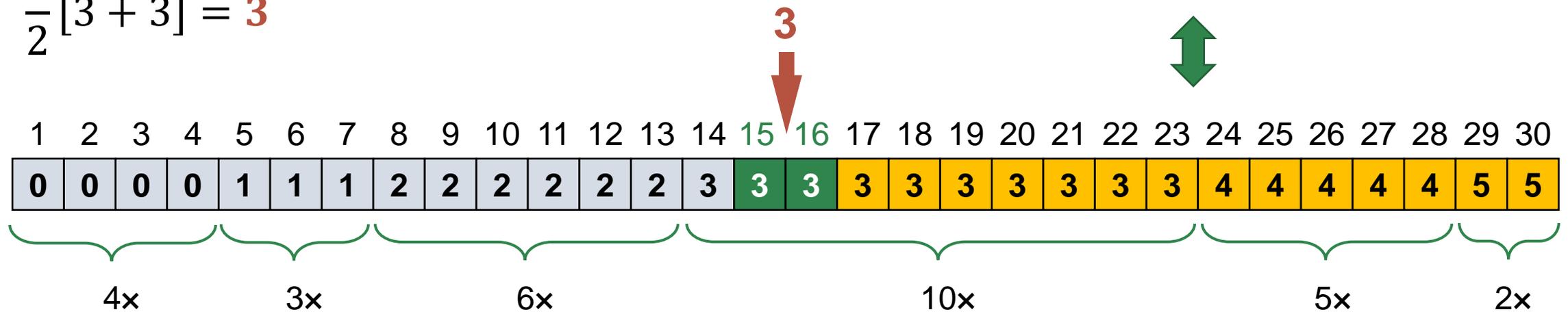
$$\frac{1}{2} [x_{(\frac{30}{2})} + x_{(\frac{30}{2}+1)}] =$$

$$\frac{1}{2} [x_{15} + x_{16}] =$$

$$\frac{1}{2} [3 + 3] = 3$$

x	n	F
0	4	4
1	3	7
2	6	13
3	10	23
4	5	28
5	2	30
N	30	

Da ist die 15 und 16 drin



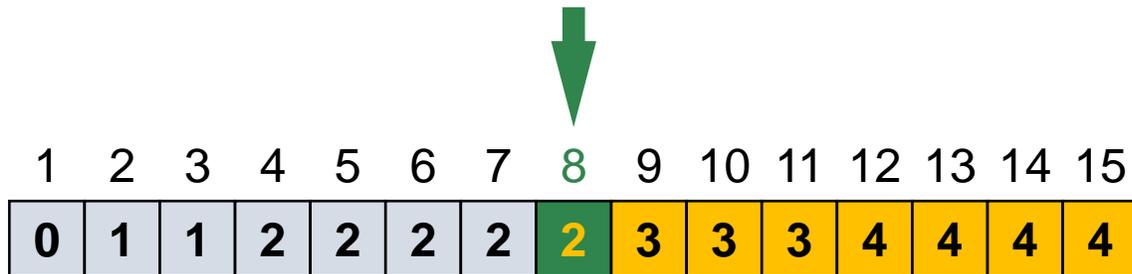
Lösung für Übung 2c

(Exkurs) Falls N ungerade ist, ist es einfach der Wert in der Mitte

$$MD = \bar{x}_{med} = x_{\left(\frac{N+1}{2}\right)} = x_{\left(\frac{15+1}{2}\right)} = x_8 = 2$$

x	n	F
0	1	1
1	2	3
2	5	8
3	3	11
4	4	15
5	0	15
N	15	

Da ist die 8 drin



Lösung für Übung 2d

Ermitteln Sie ...

d) Arithmetisches Mittel

x	n	n · x
0	4	0
1	3	3
2	6	12
3	10	30
4	5	20
5	2	10
N	30	75

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{x_1 + x_2 + \dots + x_N}{N}$$

$$\bar{x} = \frac{0 + 0 + 0 + 0 + 1 + 1 + 1 + 2 + \dots + 4 + 5 + 5}{30}$$

$$\bar{x} = \frac{4 \cdot 0 + 3 \cdot 1 + 6 \cdot 2 + 10 \cdot 3 + 5 \cdot 4 + 2 \cdot 5}{30} = \frac{75}{30}$$

$$\bar{x} = 2,5$$

6. HÄUFIGKEITEN UND LAGEMASSE IN R BERECHNEN

Ein wichtiges Wort vorweg

Der nachfolgende Code mag Ihnen zuerst einmal schwer überschaubar erscheinen, auch wenn es sich in der Praxis nur um ein paar wenige Funktionen handelt.

Bis Sie diese Funktionen verinnerlicht haben, bedarf es Übung und noch ein paar empirische Forschungsseminare. **Niemand erwartet von Ihnen, dass Sie all diese Funktionen auswendig lernen** und perfekt aus dem Kopf nieder schreiben können.

Daten mit R auswerten zu können, heißt in der Praxis **zu wissen, wo man den passenden Code nachschlagen kann**, diesen zu übernehmen **und an die eigenen Variablen und Datensätze anzupassen**.

Lagemaße berechnen mittels **tidyverse** (**dplyr**)

Zusätzlich zu den **dplyr**-Funktionen, die Sie aus der letzten Sitzung kennen, kommen noch zwei neue dazu. Mit diesen können Sie Lagemaße berechnen:

1. **summarise()** / **summarize()**
2. [ggf. in Kombination mit **group_by()**].



Erlaubt die Ausgabe der Lagemaße für jede Gruppe getrennt, zum Beispiel den Mittelwert für jedes Geschlecht separat.

Struktur:

```
data %>%
#   group_by(group_variable) %>%
  summarize(
    n = n(),
    Mode = names(which.max(table(variable))),
    Mode_N = max(table(variable)),
    Mdn = median(variable),
    M = mean(variable, na.rm = TRUE),
  )
```

Lagemaße berechnen mittels **tidycomm**

Sie können die Lagemaße alternativ mit dem **tidycomm**-Package berechnen. Dafür benötigen Sie je nach Skalenniveau der Sie interessierenden Variablen dann folgende Funktionen:

1. **tab_frequencies()** * [Häufigkeitentabelle]
2. **describe_cat()** ** [Lagemaße für nominal- und ordinalskalierte Variablen]
3. **describe()** [Lagemaße für metrische Variablen]

* Eselsbrücke: tab_frequencies steht für „tabulate frequencies“

** Eselsbrücke: describe_cat steht für „describe categorical variable“

Struktur:

```
data %>%  
  tab_frequencies(variable)
```

```
data %>%  
  describe_cat(variable)
```

```
data %>%  
  describe(variable)
```

7. ÜBUNG 3: ÜBUNG 2 IN R UMSETZEN

Übung 2 in R umsetzen

Erinnerung: 30 Studierende mussten angeben, an wie vielen Tagen der Woche sie das Institut zum Literaturstudium betreten. Die Stichprobe lieferte folgende Ergebnisse:

3	2	5	3	1	0	3	0	2	2
3	0	2	3	0	4	2	4	5	3
4	3	2	4	4	3	3	1	3	1

Wir ermitteln nun mittels R (`tidycomm`):

- die absoluten Häufigkeiten,
- Modus,
- Median,
- arithmetisches Mittel

Übung 2 in R umsetzen

Legen wir zunächst die 30 Werte samt der ID der Studierenden als Vektoren und Dataframe an:

```
4  ## Anlegen des dataframes
5  bib_besuche <- c(3, 2, 5, 3, 1, 0, 3, 0, 2, 2, 3, 0, 2, 3, 0, 4, 2, 4, 5, 3, 4
6                    3, 2, 4, 4, 3, 3, 1, 3, 1)
7  student_id <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,
8                 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30)
9  data <- cbind(student_id, bib_besuche)
10 data <- as_tibble(data)
```

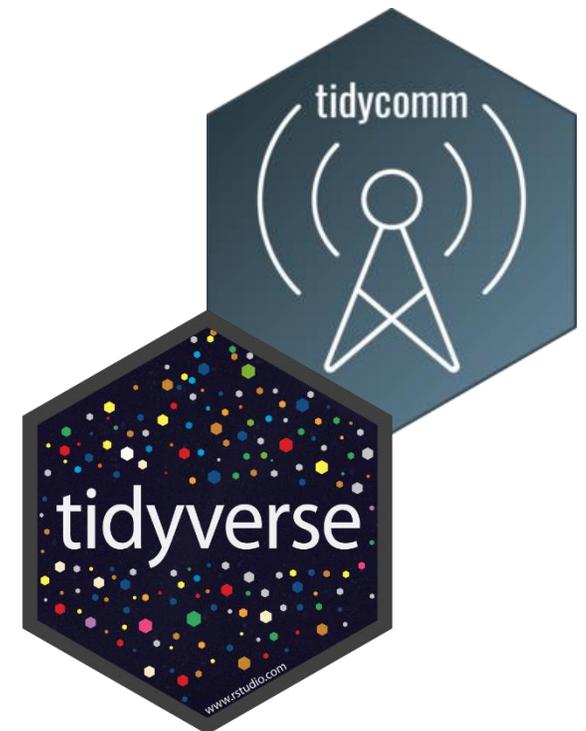
Übung 2 in R umsetzen

Als nächstes aktivieren wir die Packages, die die Zusatzfunktionen enthalten, die wir gleich zur Berechnung der Lagemaße nutzen wollen. Dazu benutzen wir die Funktion `library()` und aktivieren zwei Packages:

```
library(tidyverse)
```

```
library(tidycomm)
```

→ `tidycomm` ist unser IfKW-eigenes Package zur Berechnung vieler KW-typischer Kennwerte und Testverfahren. Es folgt der „tidy“ Logik des tidyverse.



Lösung für Übung 2 in R umsetzen mittels tidycomm

Aufgabe 2a): Ausgeben der absoluten Häufigkeiten.

Sie können `tab_frequencies()` aus dem `tidycomm`-Package anwenden:

Der Code: `14 data %>% tab_frequencies(bib_besuche)`

Das Ergebnis:

```
# A tibble: 6 × 5
  bib_besuche  n percent cum_n cum_percent
*   <dbl> <int> <dbl> <int> <dbl>
1         0     4 0.133     4  0.133
2         1     3  0.1       7  0.233
3         2     6  0.2       13 0.433
4         3    10 0.333     23 0.767
5         4     5 0.167     28 0.933
6         5     2 0.0667    30  1
```

Zum Vergleich hier
noch einmal Ihre
manuell ausgezählten
Häufigkeiten:

x	n
0	4
1	3
2	6
3	10
4	5
5	2
N	30

Lösung für Übung 2 in R umsetzen mittels tidycomm

Aufgabe 2b): Ausgeben des Modus.

Die Funktion `describe_cat()` aus dem `tidycomm`-Package anwenden:

Der Code: `17 data %>% describe_cat(bib_besuche)`

Das Ergebnis:

```
# A tibble: 1 × 6
  Variable N Missing Unique Mode Mode_N
* <chr>   <int> <int> <dbl> <chr> <int>
1 bib_besuche 30 0 6 3 10
```

Anzahl der Befragten

Modus

Häufigkeit, mit der die Ausprägung des Modus in der Stichprobe vertreten ist.

Lösung für Übung 2 in R umsetzen mittels tidycomm

Aufgabe 2c) + 2d): Ausgeben des Medians und Mittelwerts.

Die Funktion `describe()` aus dem `tidycomm`-Package anwenden:

Der Code: `20 data %>% describe(bib_besuche)`

Das Ergebnis:

```
# A tibble: 1 × 15
  Variable      N Missing      M      SD   Min   Q25   Mdn   Q75   Max Range CI_95_LL CI_95_UL Skewness Kurtosis
* <chr>      <int> <int> <dbl> <dbl>
1 bib_besuche 30      0    2.5  1.43    0     2     3     3     5     5    1.97    3.03  -0.286  2.34
```

Anzahl der Befragten

Mittelwert der Befragten

Median der Befragten

8. ÜBUNG 4

HÄUFIGKEITEN UND LAGEMASSE IN R BERECHNEN

Übung 4

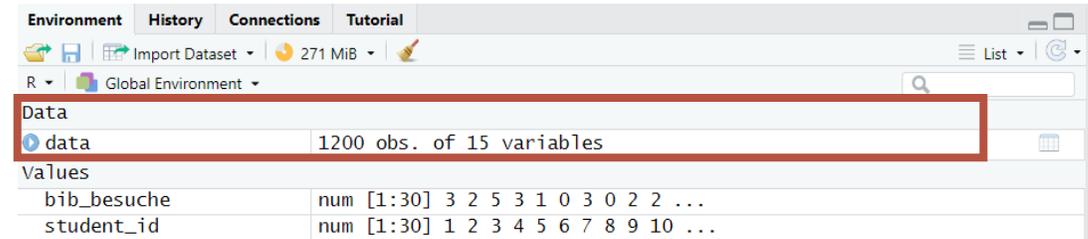
Laden Sie nun den Datensatz „WoJ“ vom tidycomm-Package in Ihre RStudio-Environment. WoJ ist ein Datensatz, der bei der Installation des tidycomm-Package gleich „mitgeliefert“ wird. Dazu führen Sie diesen Code aus, nachdem Sie das tidycomm-Package bereits aktiviert haben (was Sie bereits haben):

```
24 data <- WoJ
```

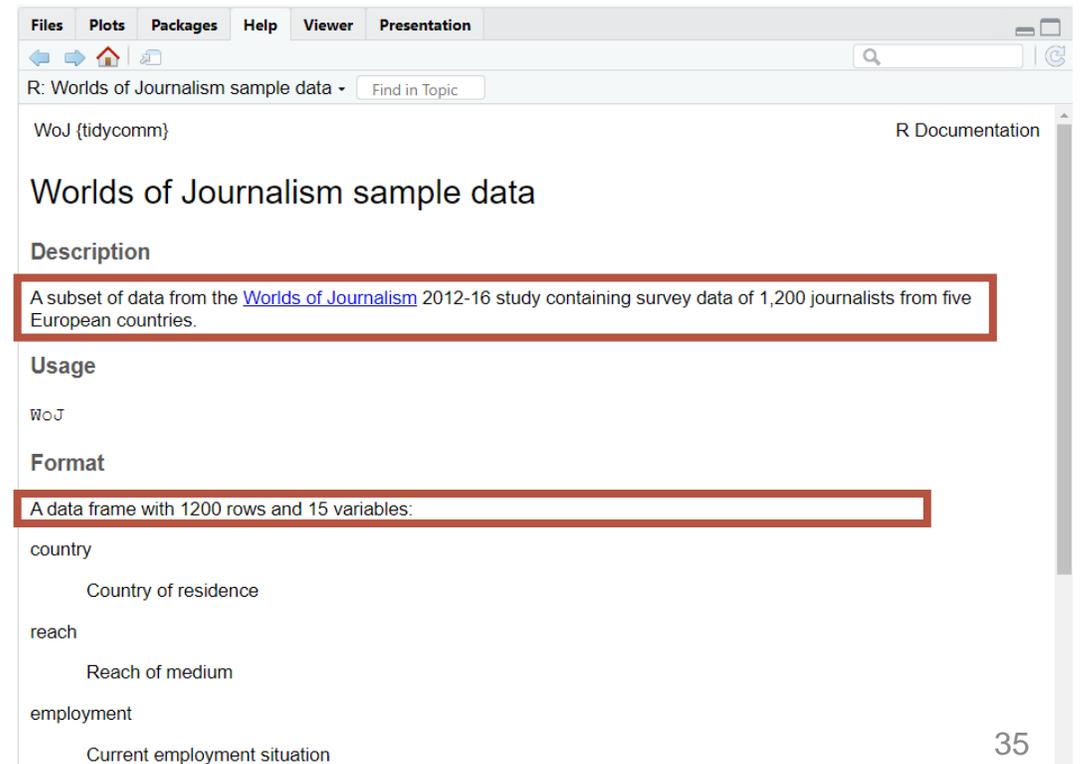
Über diesen Datensatz können wir mehr mittels der Hilfe-Funktion ? erfahren:

```
25 ?WoJ
```

Die Environment nach dem Laden der Daten:



Hilfeseite zum WoJ-Datensatz:



Übung 4

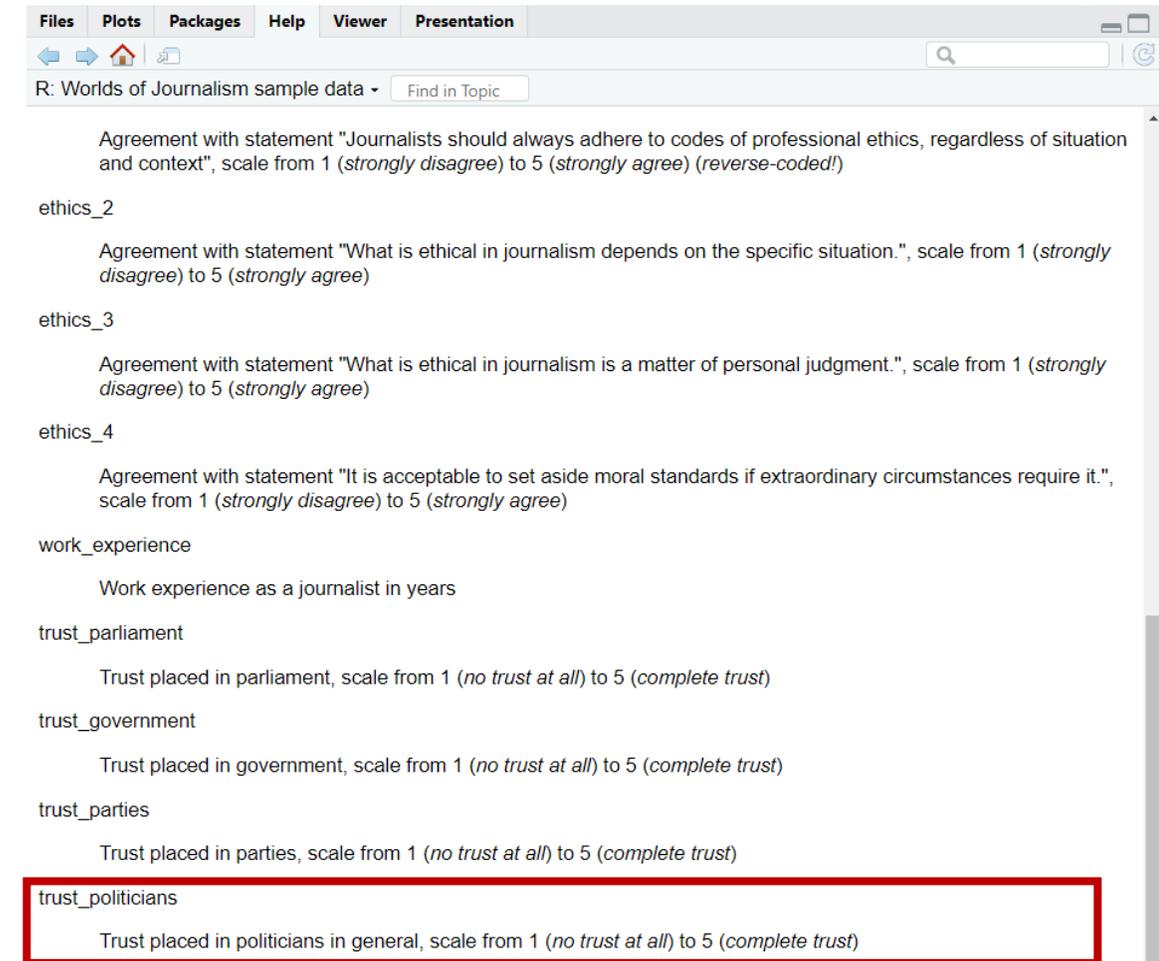
Durch die Hilfeseite wissen wir jetzt, dass der Datensatz 1200 Befragte enthält. Scrollt man die Hilfeseite weiter runter, erfährt man auch, dass der Datensatz eine Variable namens „**trust_politicians**“ enthält, die von 1 (gar kein Vertrauen) bis 5 (volles Vertrauen) skaliert ist.

Berechnen Sie für «trust_politicians» in R mittels **tidycmm...**

- a) die absoluten Häufigkeiten
- b) Modus
- c) Median
- d) arithmetisches Mittel

Zusatzaufgabe: Wer besonders schnell ist, kann die Übungen auch noch einmal mit **dplyr** zu lösen versuchen (Tipp: b-d) sind leichter als a).

Hilfeseite zum WoJ-Datensatz:



Files Plots Packages Help Viewer Presentation

R: Worlds of Journalism sample data - Find in Topic

Agreement with statement "Journalists should always adhere to codes of professional ethics, regardless of situation and context", scale from 1 (*strongly disagree*) to 5 (*strongly agree*) (*reverse-coded!*)

ethics_2

Agreement with statement "What is ethical in journalism depends on the specific situation.", scale from 1 (*strongly disagree*) to 5 (*strongly agree*)

ethics_3

Agreement with statement "What is ethical in journalism is a matter of personal judgment.", scale from 1 (*strongly disagree*) to 5 (*strongly agree*)

ethics_4

Agreement with statement "It is acceptable to set aside moral standards if extraordinary circumstances require it.", scale from 1 (*strongly disagree*) to 5 (*strongly agree*)

work_experience

Work experience as a journalist in years

trust_parliament

Trust placed in parliament, scale from 1 (*no trust at all*) to 5 (*complete trust*)

trust_government

Trust placed in government, scale from 1 (*no trust at all*) to 5 (*complete trust*)

trust_parties

Trust placed in parties, scale from 1 (*no trust at all*) to 5 (*complete trust*)

trust_politicians

Trust placed in politicians in general, scale from 1 (*no trust at all*) to 5 (*complete trust*)

Source

'<https://worldsofjournalism.org/data/data-and-key-tables-2012-2016>'

Lösung für Übung 4

Da der WoJ-Datensatz aus dem tidycomm-Package bereits als Tibble vorliegt, können direkt `tab_frequencies()`, `describe_cat()` und `describe()` zum Einsatz kommen:

Input im Source-Fenster:

```
27 # Auswertung 4a)
28 data %>% tab_frequencies(trust_politicians)
29
30 # Auswertung 4b)
31 data %>% describe_cat(trust_politicians)
32
33 # Auswertung 4c) + 4d)
34 data %>% describe(trust_politicians)
```

Output in der Console:

```
> # Auswertung 4a)
> data %>% tab_frequencies(trust_politicians)
# A tibble: 4 x 5
  trust_politicians n percent cum_n cum_percent
  <dbl> <int> <dbl> <int> <dbl>
1 1 95 0.0792 95 0.0792
2 2 445 0.371 540 0.45
3 3 601 0.501 1141 0.951
4 4 59 0.0492 1200 1

> # Auswertung 4b)
> data %>% describe_cat(trust_politicians)
# A tibble: 1 x 6
  Variable N Missing Unique Mode Mode_N
  <chr> <int> <int> <dbl> <chr> <int>
1 trust_politicians 1200 0 4 3 601

> # Auswertung 4c) + 4d)
> data %>% describe(trust_politicians)
# A tibble: 1 x 15
  Variable N Missing M SD Min Q25 Mdn Q75 Max Range CI_95_LL CI_95_UL Skewness Kurtosis
  <chr> <int> <int> <dbl> <dbl>
1 trust_pol... 1200 0 2.52 0.712 1 2 3 3 4 3 2.48 2.56 -0.320 2.78
```

Modus

Mittelwert

Median

Lösung für Übung 4 (Zusatzaufgabe)

Input im Source-Fenster:

```

36 # Zusatzaufgabe 4a-d) mit dplyr:
37 data %>%
38   group_by(trust_politicians) %>%
39   summarize(n = n())
40
41 data %>%
42   summarize(
43     n = n(),
44     Mode = names(which.max(table(trust_politicians))),
45     Mode_N = max(table(trust_politicians)),
46     M = mean(trust_politicians, na.rm = TRUE),
47     Mdn = median(trust_politicians)
48   )

```

Output in der Console:

```

# A tibble: 4 × 2
  trust_politicians     n
      <dbl> <int>
1             1         95
2             2        445
3             3        601
4             4         59

# A tibble: 1 × 5
   n Mode  Mode_N     M   Mdn
  <int> <chr> <int> <dbl> <dbl>
1  1200 3      601  2.52    3

```

9. VISUALISIERUNG VON HÄUFIGKEITEN IN R

Visualisierungen anfordern in R

Das **tidycomm**-Package erlaubt es, alle Lagemaße mittels passender Grafiken zu visualisieren. Dazu muss nur `%>% visualize()` an die Funktion angehängt werden, die das Lagemaß berechnet.

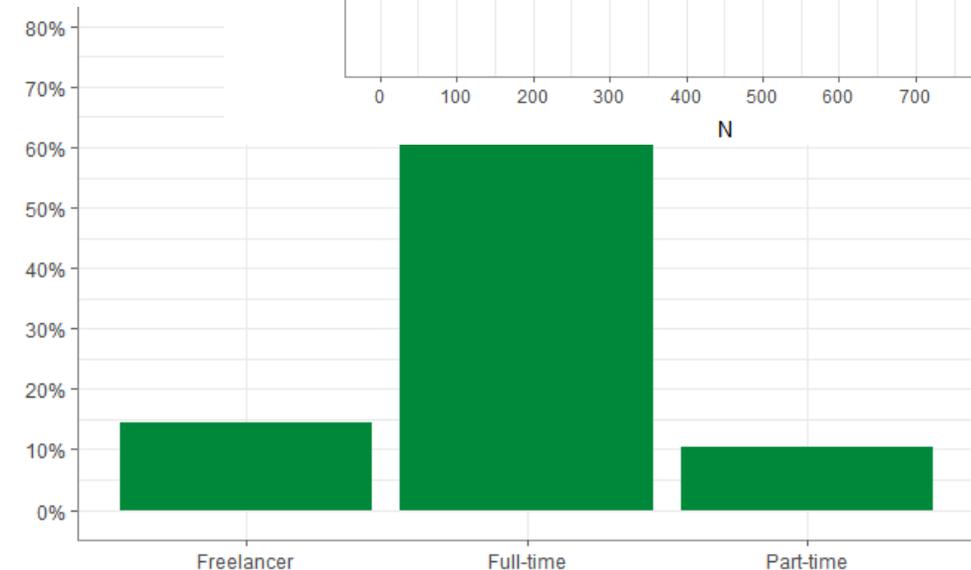
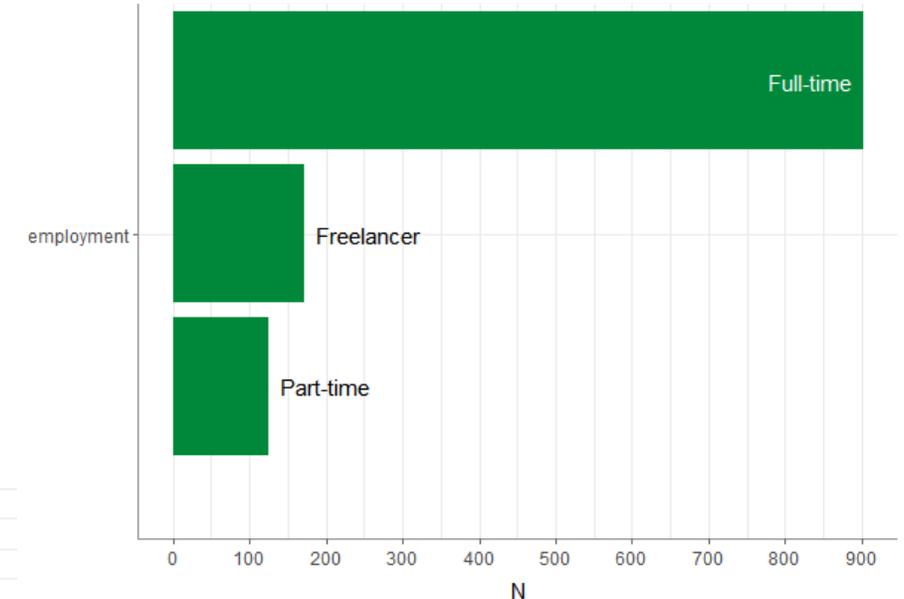
Skalenniveau	Diagrammtyp	Beschreibung	tidycomm-Code
Nominal oder Ordinal	Balkendiagramm	Zeigt die Häufigkeit / den Anteil jeder Kategorie.	absolute Häuf.: <code>describe_cat() %>% visualize()</code> oder relative Häuf.: <code>tab_frequencies() %>% visualize()</code>
	Histogramm	Zeigt die Verteilung einer metrischen Variable, indem die Häufigkeit der Werte in Intervallen dargestellt wird.	<code>tab_frequencies() %>% visualize()</code>
Metrisch (Intervall oder Verhältnis)	Boxplot	Fasst die Verteilung einer metrischen Variable zusammen und zeigt Median und Quartile.	<code>describe() %>% visualize()</code>

Visualisierungen anfordern in R: Das Balkendiagramm

Nominalskalierte Variablen wie „employment“ im WoJ-Datensatz werden automatisch als Balkendiagramm visualisiert.

`describe_cat()` + `visualize()` liefert absolute Häufigkeiten.

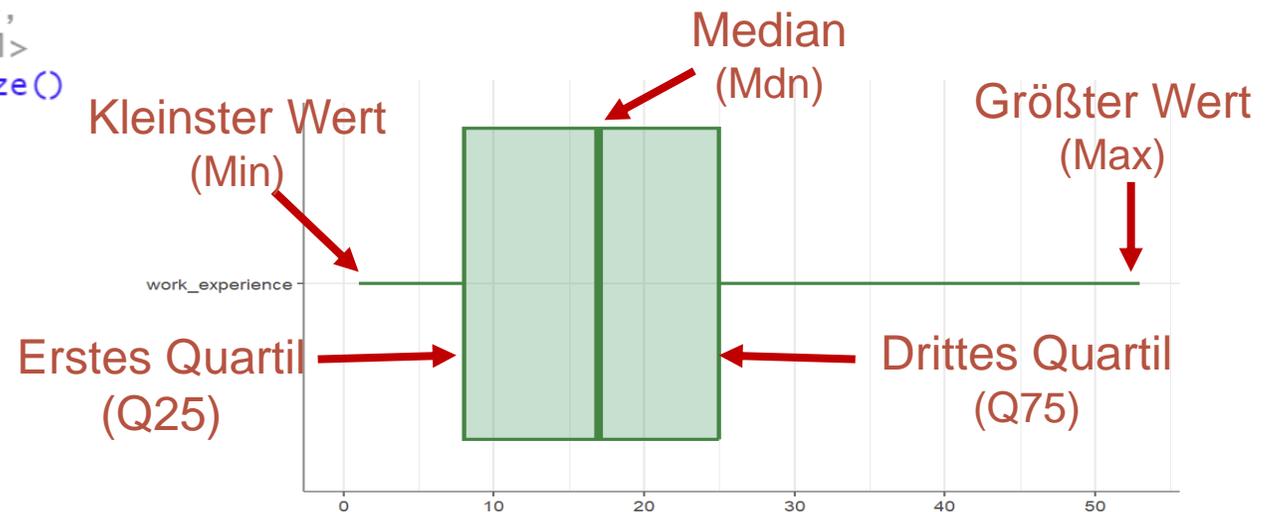
`tab_frequencies()` + `visualize()` liefert relative Häufigkeiten.



Visualisierungen anfordern in R: Der **Boxplot**

Metrische Variablen (z.B. „work_experience“) können als Boxplot visualisiert werden, indem `%>% visualize()` an die `describe()`-Funktion angefügt wird:

```
> data %>% describe(work_experience)
# A tibble: 1 × 15
  Variable      N Missing      M     SD   Min   Q25   Mdn   Q75   Max
* <chr>      <int>  <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 work_ex...  1187    13  17.8  10.9    1     8    17    25    53
# i 5 more variables: Range <dbl>, CI_95_LL <dbl>,
# CI_95_UL <dbl>, skewness <dbl>, kurtosis <dbl>
> data %>% describe(work_experience) %>% visualize()
```



Visualisierungen anfordern in R: Das **Histogramm**

Häufigkeitstabellen für metrisch skalierte Variablen wie „work_experience“ im WoJ-Datensatz werden mittels **visualize()** zu einem Histogramm:

```
> data %>% tab_frequencies(work_experience)
# A tibble: 51 × 5
  work_experience      n percent cum_n cum_percent
*   <dbl> <int> <dbl> <int> <dbl>
1         1     21 0.0175     21 0.0175
2         2     33 0.0275     54 0.045
3         3     39 0.0325     93 0.0775
4         4     46 0.0383    139 0.116
5         5     56 0.0467    195 0.162
6         6     19 0.0158    214 0.178
7         7     37 0.0308    251 0.209
8         8     47 0.0392    298 0.248
9         9     23 0.0192    321 0.268
10        10     74 0.0617    395 0.329
# i 41 more rows
# i Use `print(n = ...)` to see more rows
```

```
> data %>% tab_frequencies(work_experience) %>% visualize()
```

