

# 06 Visualisierungen

Einführung in die quantitativen Forschungsmethoden

# Heute

- Quiz zur letzten Sitzung
- Grundprinzipien der Visualisierung
- Visualisierungen und Forschung
- Visualisierungen erstellen
- Weitere Arten von Plots
- Inspirationen

# Quiz zur letzten Sitzung

Quiz Link

# Grundprinzipien der Visualisierung

# Grundprinzipien

Welche Visualisierungen finden Sie hilfreich, welche weniger?

# Weniger ist mehr

- Visualisierungen zeigen oft unnötige Informationen
  - ‘chart junk’ (more examples)
- Edward Tufte: “Above all else show the data”
  - data-ink-ratio as proportion of ink devoted to non-redundant information

$$\text{data-ink-ratio} = \frac{\text{data-ink}}{\text{total ink used to print figure}}$$

# Visuelle Merkmale

Bestimmte visuelle Merkmale (pre-attentive attributes) werden von unserem Auge vorgefiltert

- **Farben**
- **Größe**
- einfache Formen
- Länge von Linien

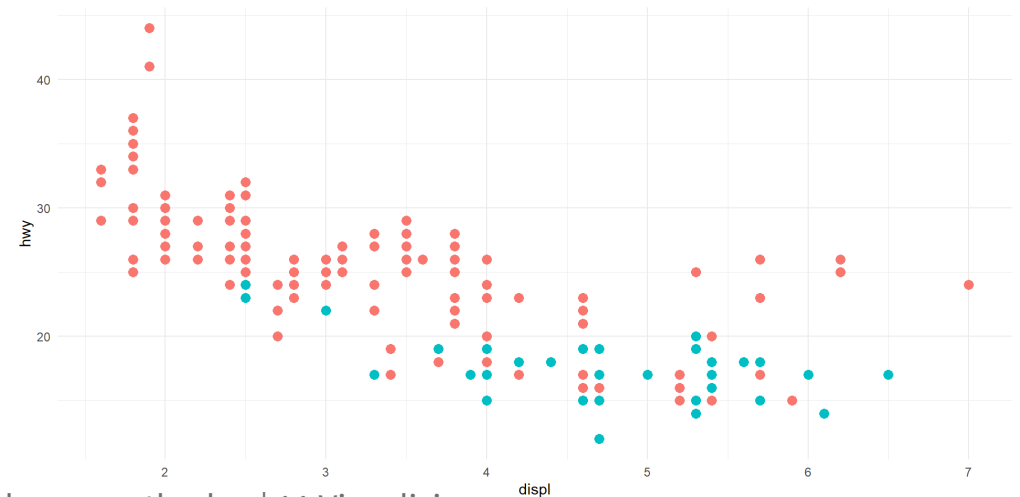
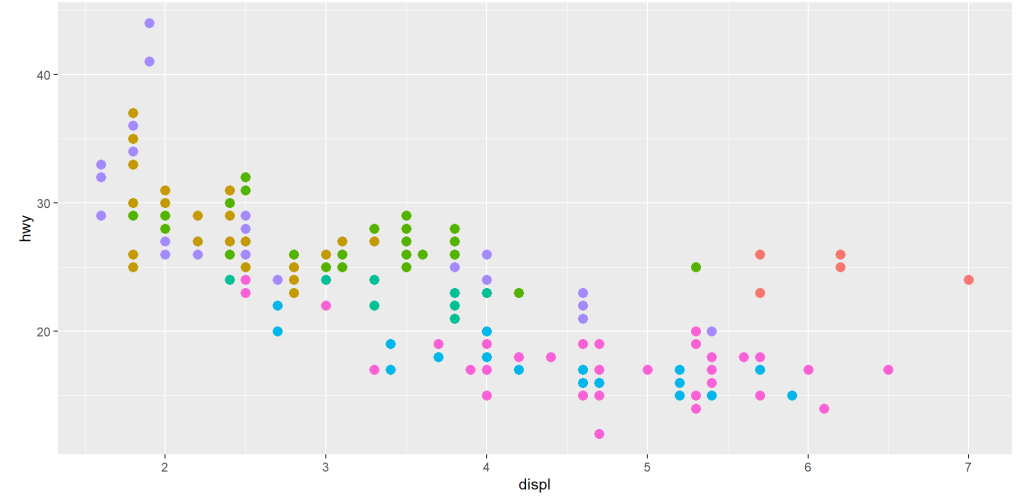
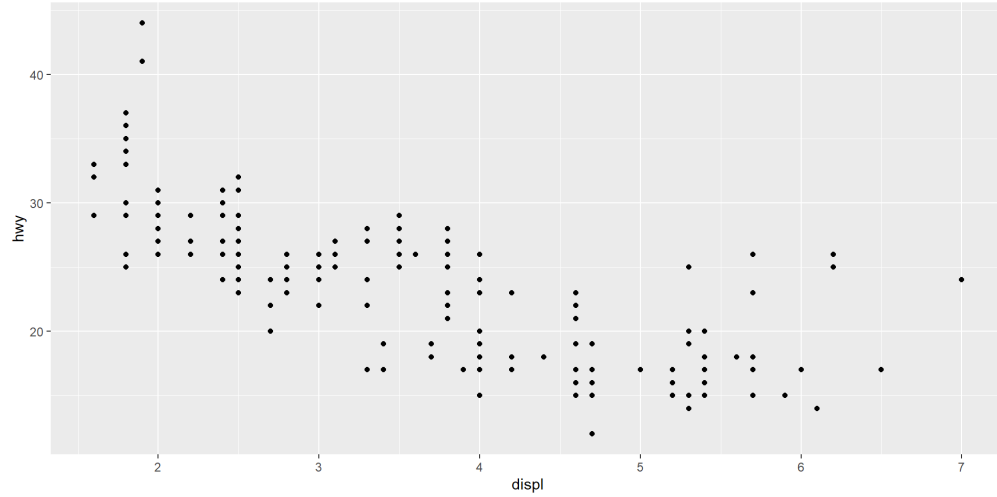
**Andere Merkmale brauchen unser Bewusstsein**

- z.B. erkennen von Buchstaben & Zahlen

→ **visuelle Merkmale können helfen Information prominenter zu machen**

# Visuelle Merkmale

Die hervorhebende Wirkung von visuellen Merkmalen hängt auch vom Kontext ab

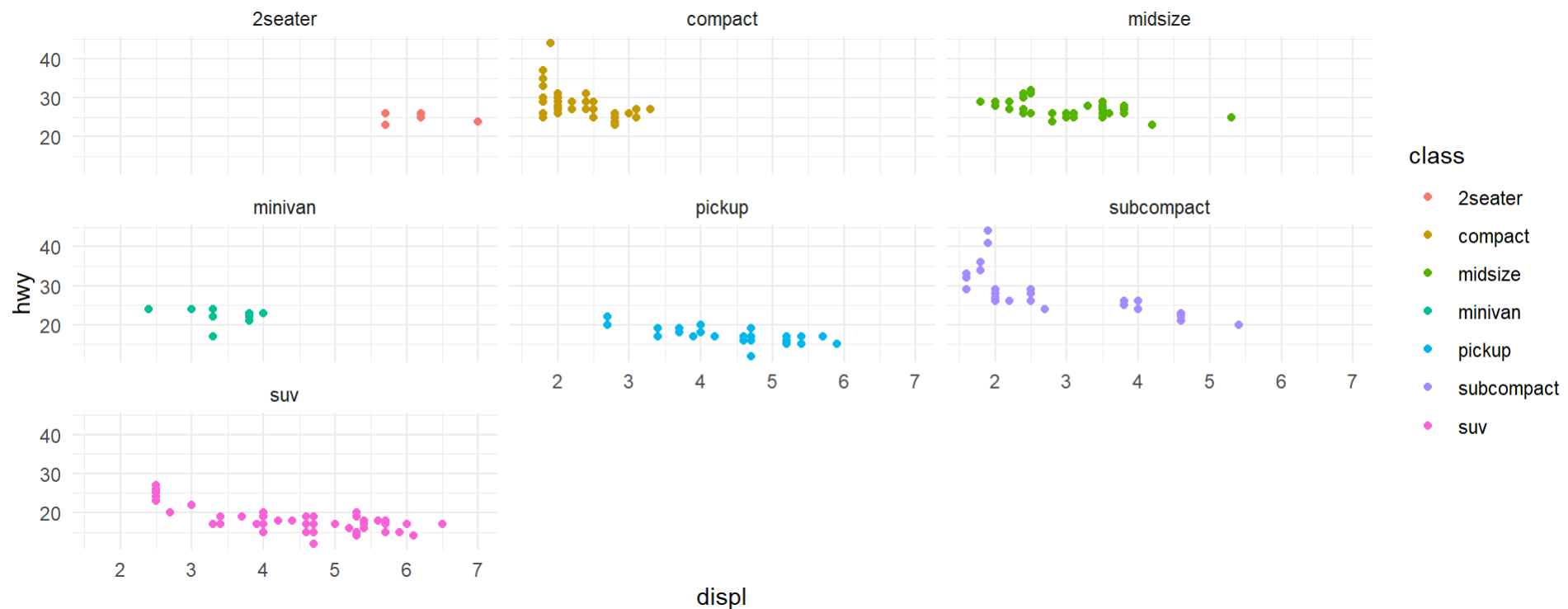






# Aufteilen von Grafiken

- häufig ist Information besser auf mehrere, kleinere Grafiken verteilt ('small multiples')
  - → zerlegen in verschiedene daten-basierte Schichten
- → Vergleiche zwischen Gruppen bieten sich visuell an



# Mehr zu Wahrnehmung

- z.B. in Healy ([2018](#))

# Visualisierungen und Forschung

# Visualisierungen und Forschung

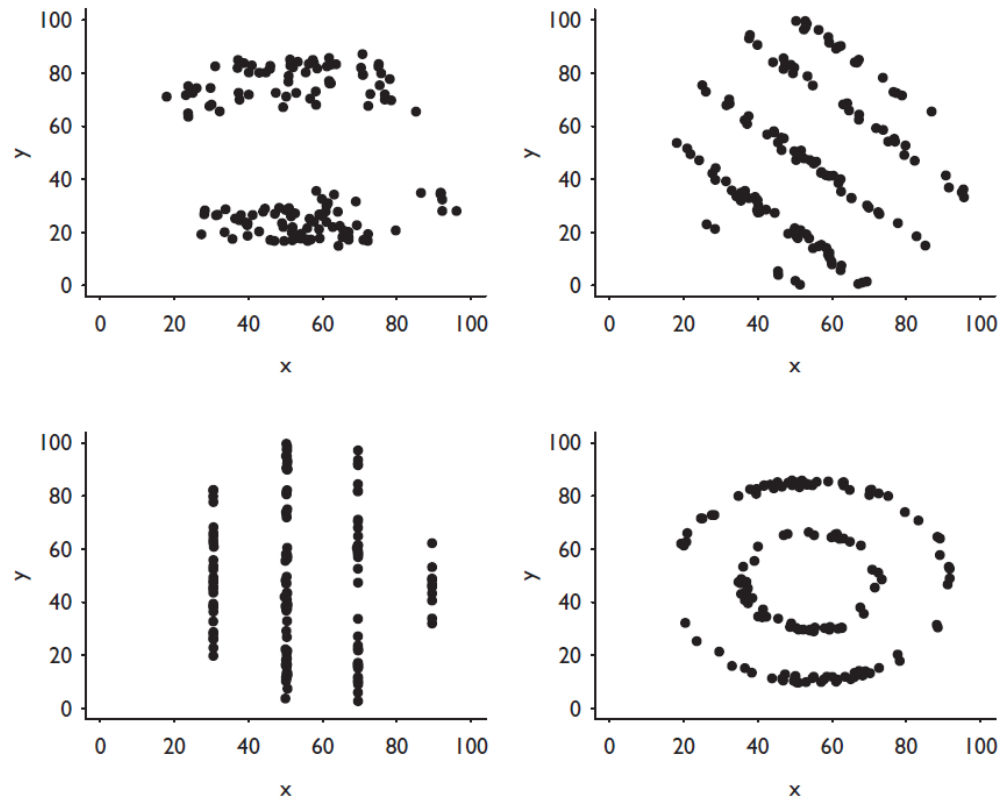


Figure 25.1 Four scatter plots of four data sets that show wildly different patterns - although summary statistics are identical

aus Traunmüller (2020), S. 437

# Visualisierungen und Forschung

Die vier Grafiken zeigen

- gleiche Mittelwerte für X & Y
- gleiche Standardabweichungen für X & Y
- gleiche Korrelationen zwischen X & Y

→ Visualisierungen helfen uns Zusammenhänge zu erkennen

# Visualisierungen und Forschung

“Data visualization can be understood as a translation tool that assigns abstract numerical values to physical properties such as spatial position along a scale, the length of a bar or the geometric shape and color of a plotting symbol.

However, data visualization involves more than simply mapping numbers to visual stimuli. Ideally, data visualization is a method that helps us and our audience understand the political world by assisting analytical thinking.”

Trautmüller ([2020](#)), S. 437

→ **Visualisierung als Form der Präsentation und Kommunikation**

→ **Visualisierung als Form der (explorativen) Analyse**

# Visualisierungen erstellen



# Visualisierungen erstellen

In R hilft das `ggplot2` Paket (Teil von `tidyverse`) beim Herstellen komplexerer Grafiken. Es ist ebenfalls in `tidyverse` enthalten.

```
1 library(tidyverse)
```

# Beispiel: Zusammenhang Regierungs- & Demokratiezufriedenheit

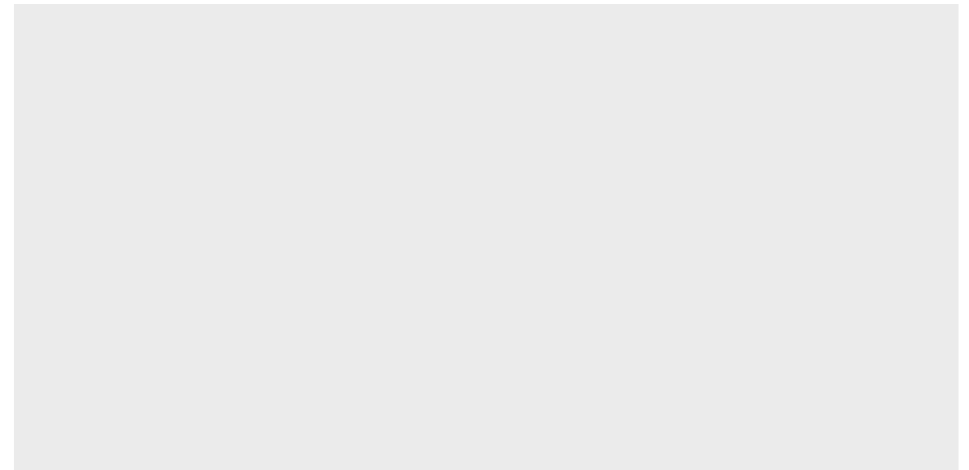
Wir berechnen den Durchschnitt der Regierungs- (`stfgov`) und Demokratiezufriedenheit (`stfdem`)

```
1 library(haven)
2 ess8 <- read_dta("../data/ESS8e02_2.dta")
3
4 # gruppieren
5 ess8 <- group_by(ess8, cntry)
6
7 # zusammenfassen
8 democracy <- summarise(ess8,
9                       mean_stfgov=mean(stfgov, na.rm=T),
10                      mean_stfdem=mean(stfdem, na.rm=T))
```

# Visualisierungen erstellen

`ggplot()` erstellt den Plot

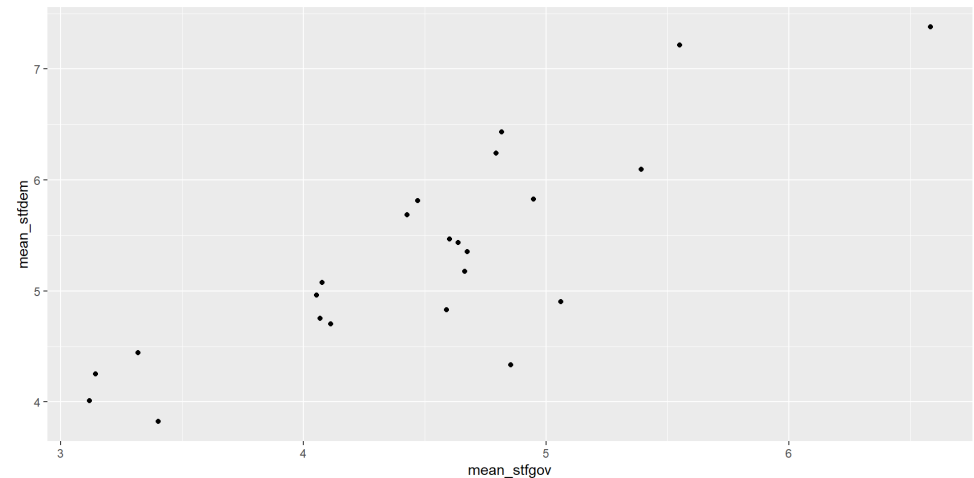
```
1 ggplot(data=democracy)
```



# Visualisierungen erstellen

- Definition von Variablen als *aesthetics* mit `aes()`
  - x-Variable, y-Variable
- `geom_point()` erzeugt einen Scatterplot basierend auf Werten
  - **Achtung:** Hinzufügen mit `+`

```
1 ggplot(data=democracy) +  
2   geom_point(aes(x=mean_stfgov,  
3                 y=mean_stfdem))
```

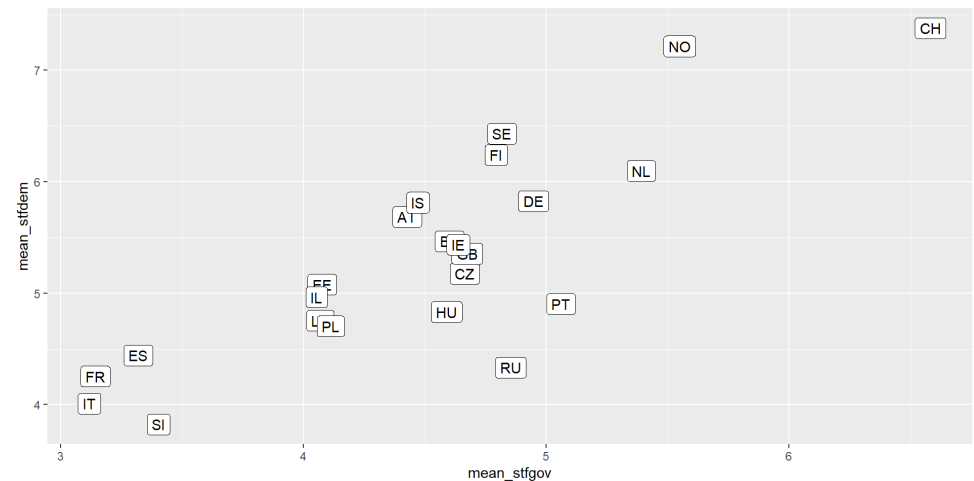


# Geoms

- `geom_point()` ist nur ein sogenanntes 'geom' → viele weitere in R
  - verschiedene geoms brauchen oft unterschiedliche aesthetics

Beispiel: `geom_label()` braucht eine Variable mit Label-Text → R-Dokumentation hilft

```
1 ggplot(data=democracy) +  
2   geom_label(aes(x=mean_stfgov,  
3                 y=mean_stfdem,  
4                 label=cntry))
```

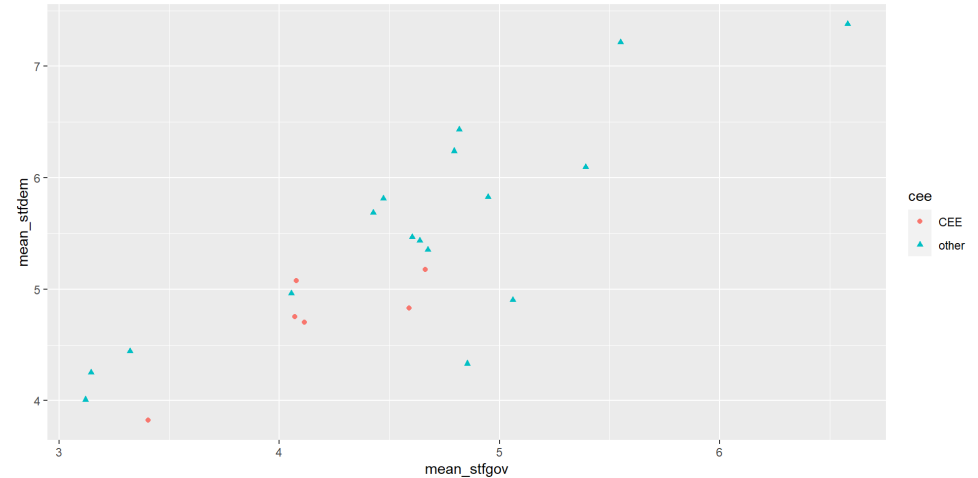


# Weitere visuelle Merkmale

- Kontrolle weiterer Merkmale über `aes`
  - z.B. `fill`, `color`, `shape`, ...
  - Verknüpfung mit Variable

```
1 democracy$cee <- ifelse(democracy$cntry %in%  
2     c("CZ", "EE", "HU", "LT", "PL", "SI"),  
3     "CEE", "other")
```

```
1 ggplot(data=democracy) +  
2   geom_point(aes(x=mean_stfgov,  
3                 y=mean_stfdem,  
4                 color=cee,  
5                 shape=cee))
```

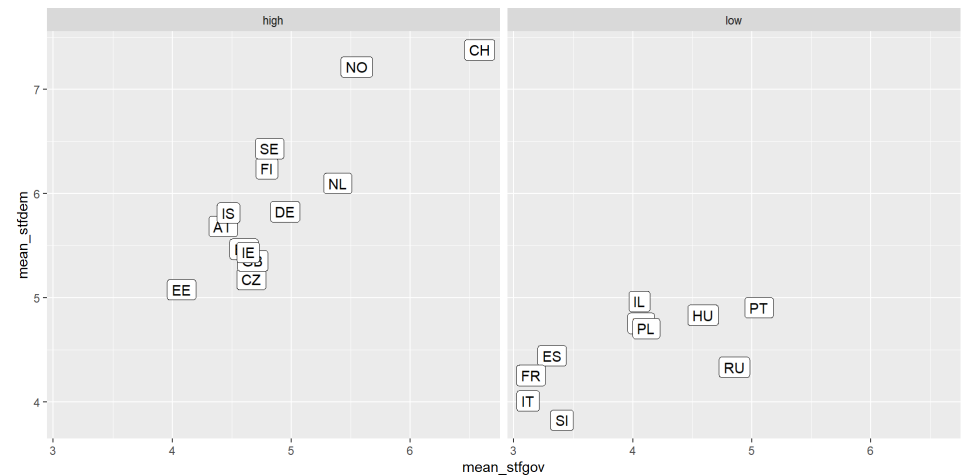


# Faceting

- mit `facet_wrap()` kann man small multiples erzeugen und den Plot entlang einer Variable 'schichten'

```
1 democracy$lowhigh <- ifelse(democracy$mean_stfdem>5,  
2                             "high", "low")
```

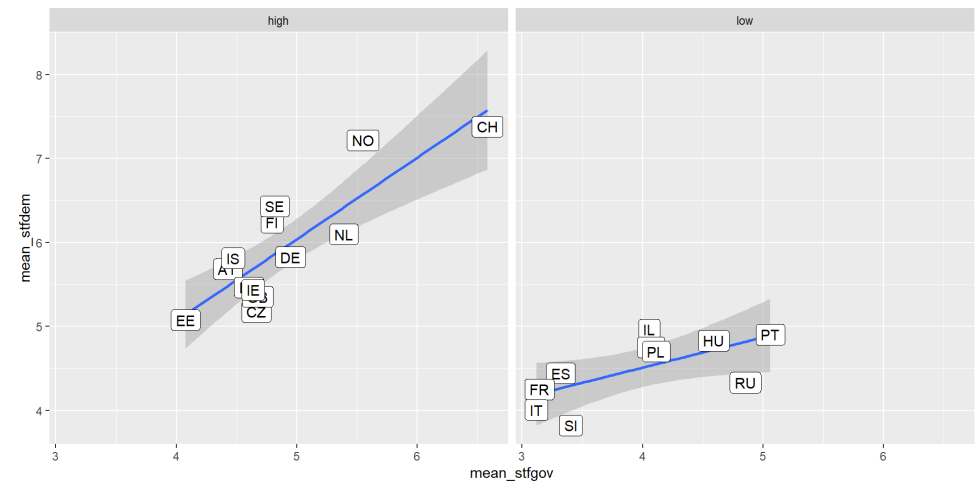
```
1 ggplot(data=democracy,  
2       aes(x=mean_stfgov,  
3           y=mean_stfdem,  
4           label=cntry))+  
5   geom_label()+  
6   facet_wrap(~lowhigh)
```



# Faceting

- besonders hilfreich, wenn wir Statistiken für jede Facet berechnen
  - z.B. zeichnet `geom_smooth()` eine Regressionslinie (`lm`: lineares Modell)

```
1 ggplot(data=democracy,  
2       aes(x=mean_stfgov,  
3           y=mean_stfdem,  
4           label=cntry))+  
5 geom_smooth(method="lm")+  
6 geom_label()+  
7 facet_wrap(~lowhigh)
```



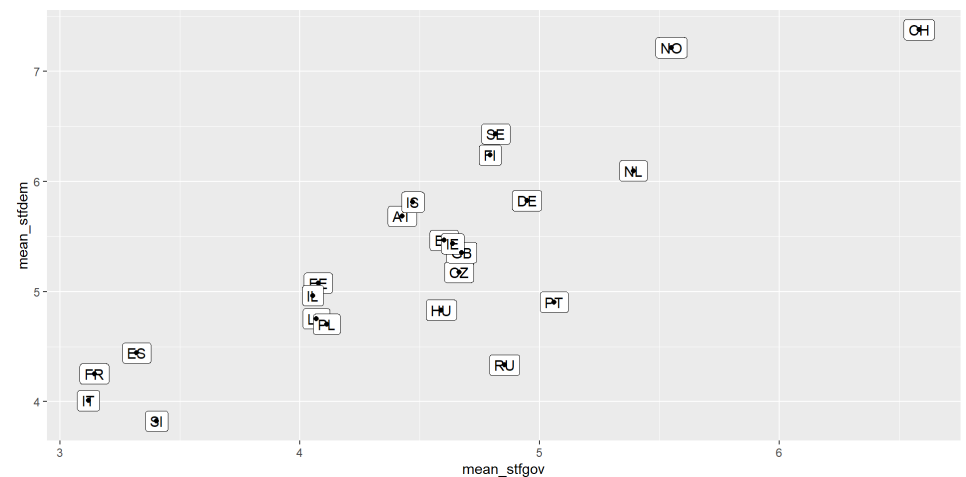
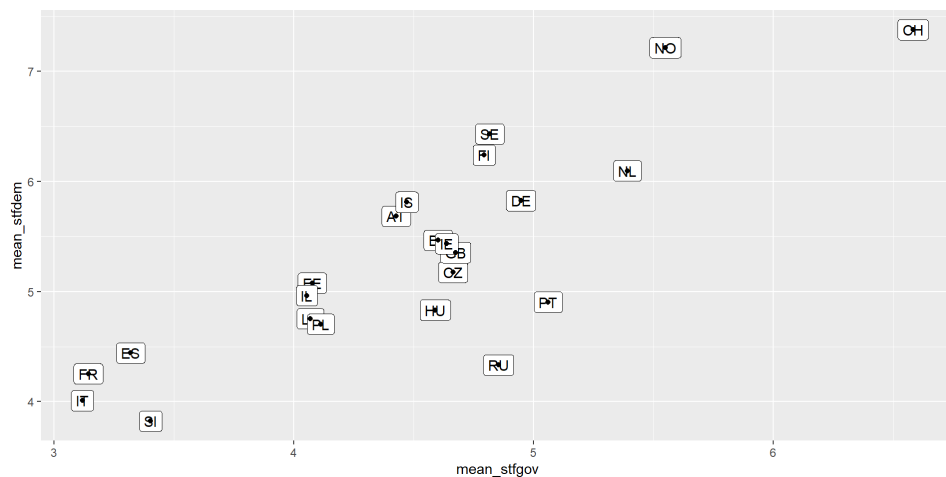


# Geoms

- Möglichkeit `aes()` für alle geoms zu definieren durch vorziehen der `aes()`-Definition in `ggplot()`

```
1 ggplot(data=democracy)+
2   geom_label(aes(x=mean_stfgov,
3                 y=mean_stfdem,
4                 label=cntry))+
5   geom_point(aes(x=mean_stfgov,
6                 y=mean_stfdem))
```

```
1 ggplot(data=democracy,
2         aes(x=mean_stfgov,
3             y=mean_stfdem,
4             label=cntry))+
5   geom_label()+
6   geom_point()
```



# Themes

Themes sind Möglichkeiten das Aussehen des Plots zu verändern

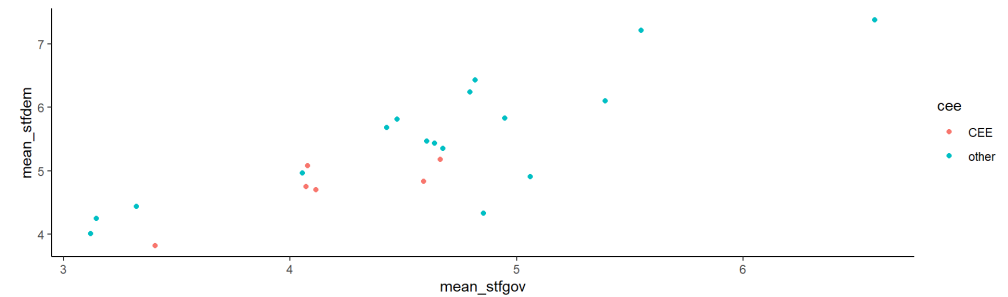
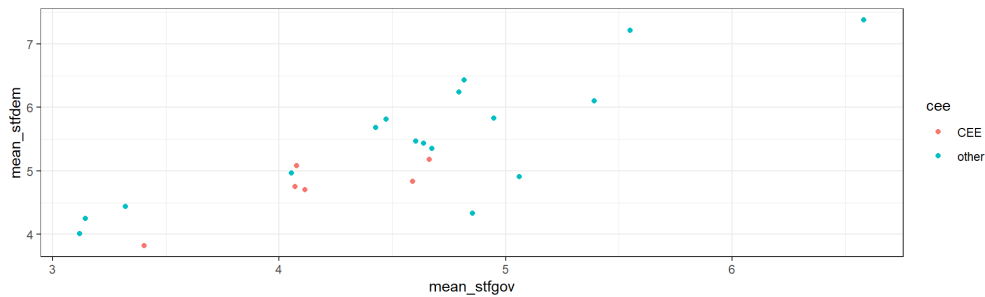
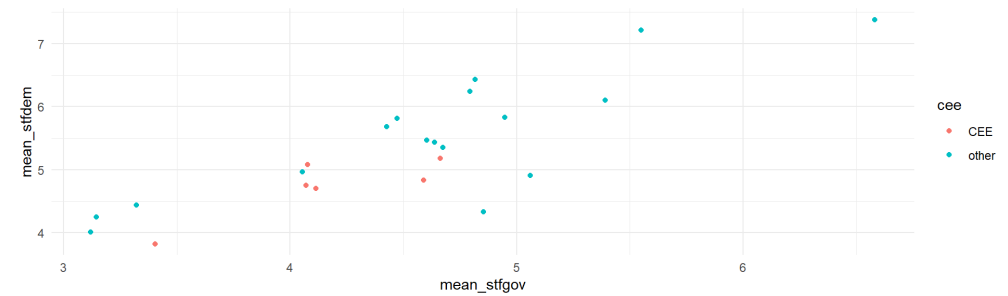
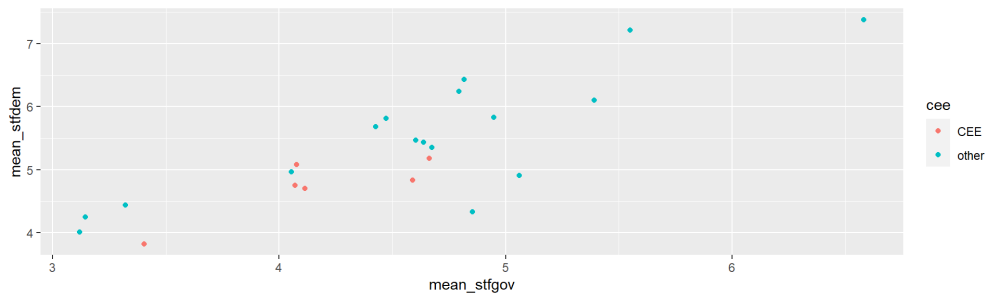
- vorgefertigte komplette themes
- Generierung eigener themes

```
1 plot <- ggplot(data=democracy) +  
2   geom_point(aes(x=mean_stfgov,  
3                 y=mean_stfdem,  
4                 color=cee))
```

# Themes

```
1 plot
```

```
1 plot+theme_minimal()  
2 plot+theme_bw()  
3 plot+theme_classic()
```

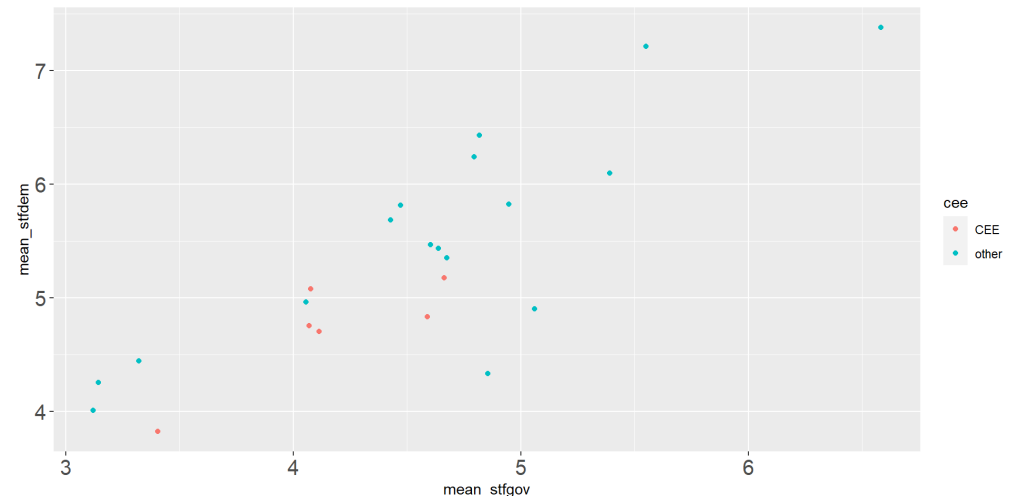
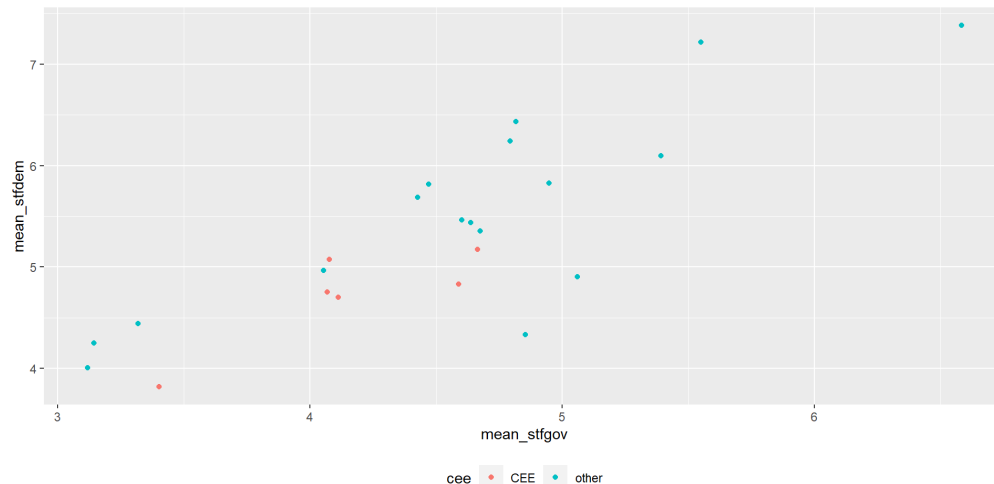


# Themes

Einige typische eigene Modifikationen

- `theme(legend.position="bottom")` / `theme(legend.position="none")`
  - platziert Legende anders / verbirgt sie
- `theme(axis.text=element_text(size=4))`
  - verändert Größe der Achsen-Annotation

```
1 plot+theme(legend.position="bottom")
2 plot+theme(axis.text=element_text(size=15))
```



# Übersicht

- Geoms: Art der Visualisierung
  - `geom_point()`
  - `geom_label()`
  - `geom_smooth()`
- aesthetics: Variable Eigenschaften, in geom oder `ggplot` definiert
  - `x, y`
  - `color, shape, fill` etc.
- Facets: `facet_wrap()`
- Themes
  - `theme_minimal(), theme_classic(), theme_bw()`
- CheatSheet

# Übungen 1

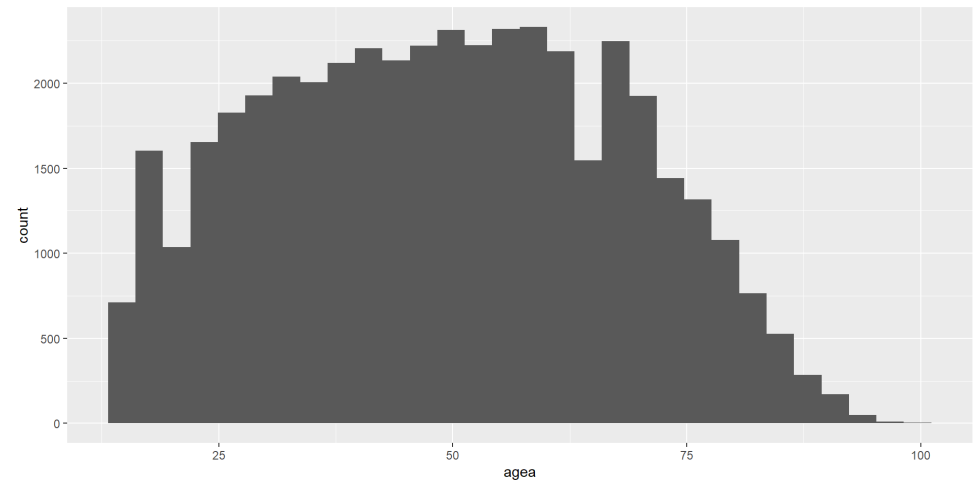
**Bearbeiten Sie den ersten Teil des Übungsfiles**

# Weitere Arten von Plots

# Häufigkeitsverteilungen

- als Histogramm mit `geom_histogram()`
  - zählt Kategorien aus

```
1 ggplot(data=ess8, aes(x=agea)) +  
2   geom_histogram()
```



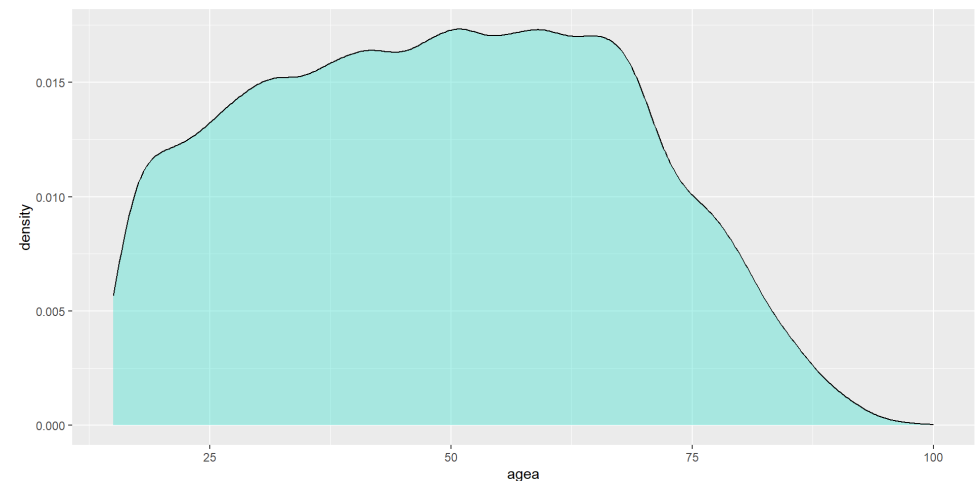
→ nützlich v.a. für wenige Kategorien



# Häufigkeitsverteilungen

- als weichgezeichnete Verteilung mit `geom_density()`
  - Häufigkeitsverteilung, keine direkten Werte → sinnvoll für viele Kategorien
- hier wird `fill` als 'attribute' verwendet (siehe [cheat sheet](#)), d.h. es ist nicht mit einer Variable verknüpft und steht deswegen nicht in `aes`, sondern im `geom`

```
1 ggplot(data=ess8, aes(x=agea)) +  
2   geom_density(fill="turquoise", alpha=0.4)
```

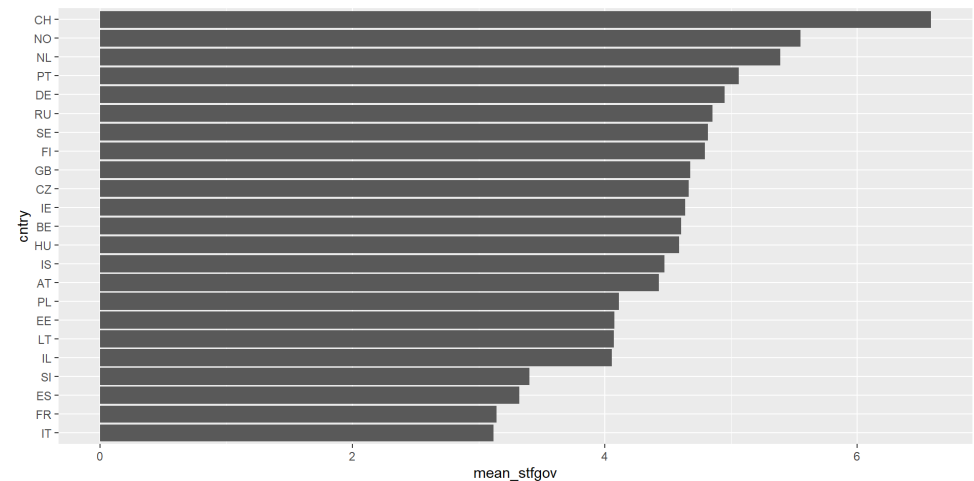


→ geeigneter für viele Kategorien

# Bar chart

- Bar charts mit `geom_col()`
  - auch `geom_bar(stat="identity")`
  - `geom_bar()` zählt Häufigkeit, `geom_col()` und `geom_bar(stat="identity")` funktionieren, wenn wir die Häufigkeit schon berechnet haben

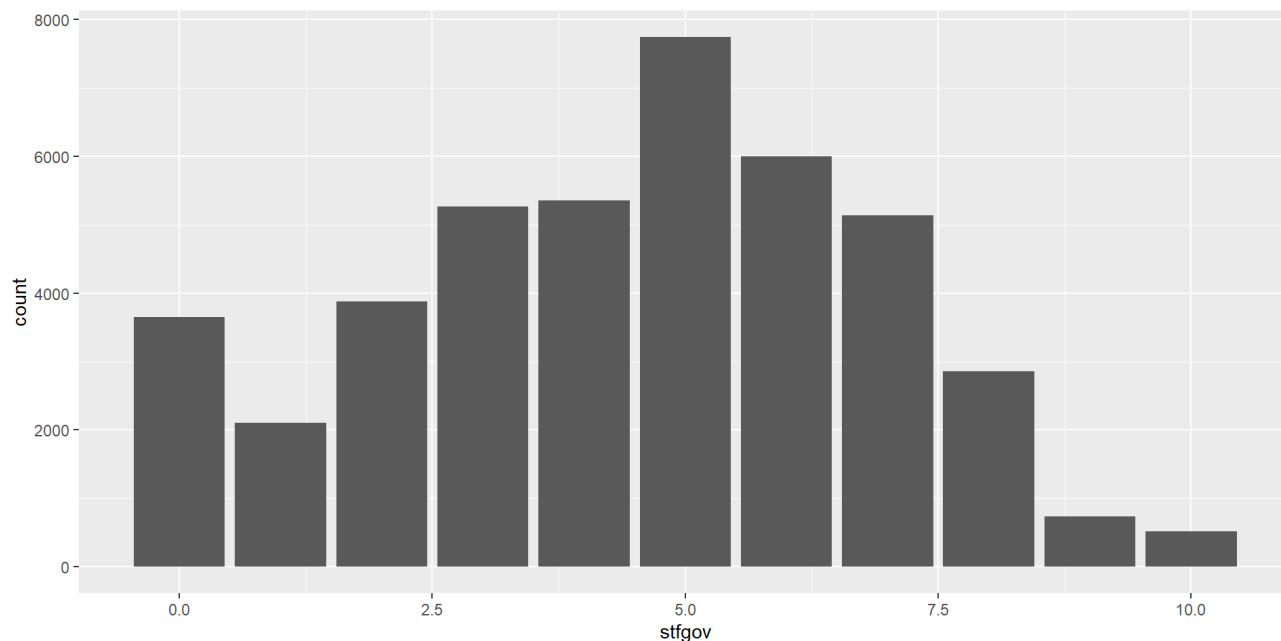
```
1 # sortieren der Werte
2 democracy <- democracy %>%
3   arrange(mean_stfgov) %>%
4   mutate(cntry=factor(cntry,cntry))
5
6 # Grafik
7 ggplot(data=democracy,
8         aes(x=mean_stfgov,xend=0,
9             y=cntry, yend=cntry))+
10  geom_col()
```



# Bar chart 2

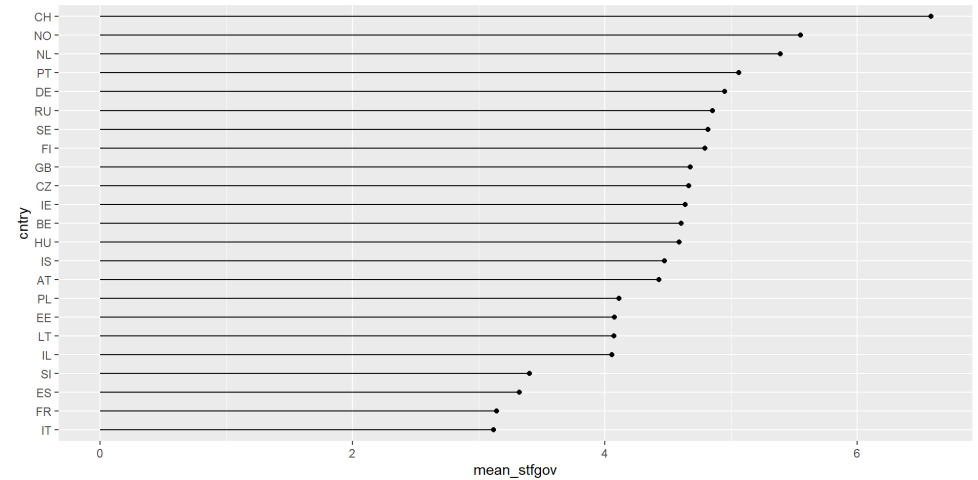
- Bar charts können auch die Häufigkeit von Werten anzeigen
  - das ist die Standardeinstellung bei `geom_bar()` !

```
1 ggplot(data=ess8,  
2       aes(x=stfgov)) + geom_bar()
```



# Lollipop chart

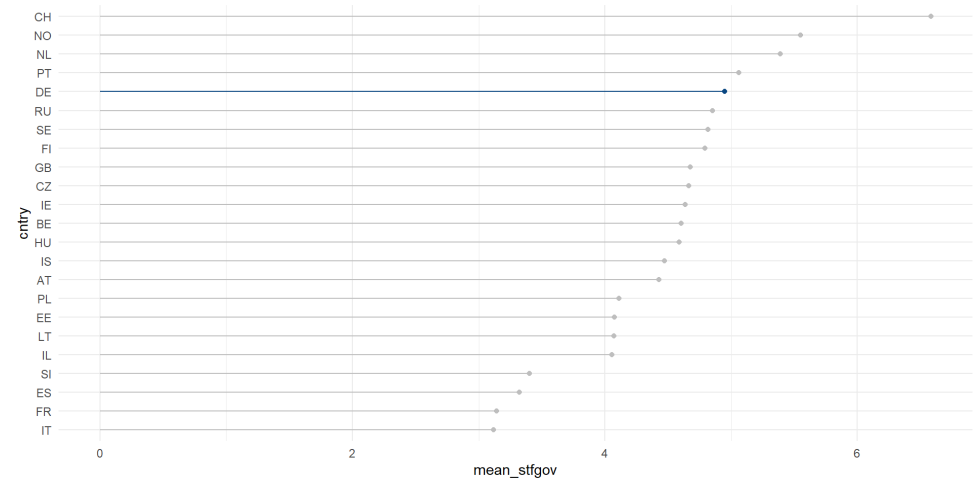
```
1 # sortieren
2 democracy <- democracy %>%
3   arrange(mean_stfgov) %>%
4   mutate(cntry=factor(cntry,cntry))
5
6 # lollipop chart
7 ggplot(data=democracy,
8         aes(x=mean_stfgov,xend=0,
9             y=cntry, yend=cntry))+
10  geom_segment()+
11  geom_point()
```



→ Achtung: `geom_segment()` braucht auch `xend` und `yend` als Spezifikation, weil es eine Fläche zwischen diesen Punkten markiert!

# Lollipop chart

```
1 democracy$col <- ifelse(democracy$cntry=="DE", "high", "low")
2 ggplot(data=democracy,
3       aes(x=mean_stfgov, xend=0,
4           y=cntry, yend=cntry,
5           # Definition Farbe
6           color=col))+
7   geom_segment()+
8   geom_point()+
9   # Legende unterdrücken
10  theme_minimal() + theme(legend.position="none")+
11  # zuweisen von Farben
12  scale_color_manual(values=c("gray", "#094782"))
```



# Karten

z.B. ESS Karte ist auch mit R erzeugt

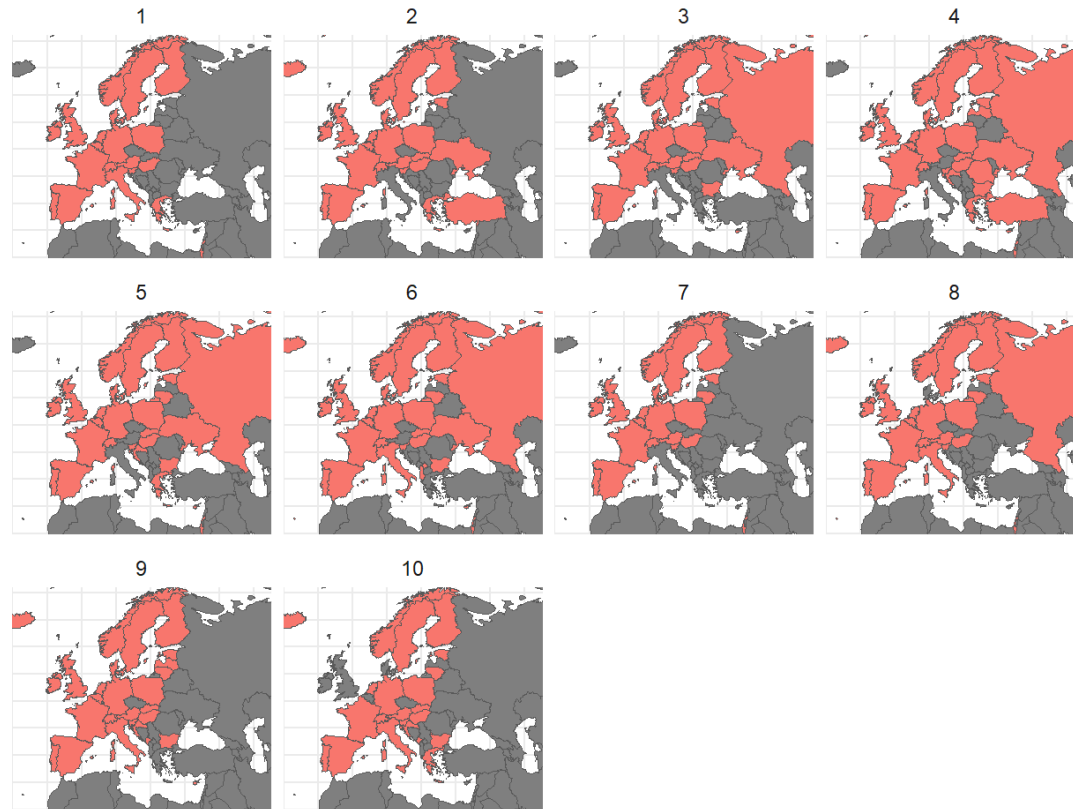


Figure 1: Länder nach ESS-Runde

→ ...aber etwas komplizierter → nur für Extra-motivierte

# Inspirationen

# Inspirationen

- Grafiken und Code in R: [The R Graph Gallery](#)
  - [Fokus auf ggplot2](#)
- Inspiration: [Data Visualisation Catalogue](#)
- Gute (journalistische) Infografiken: [The Economist](#)
- Gute (journalistische) Infografiken: [Flowing Data](#)
- Schlechte Beispiele: [Junk Charts](#)



# Aufgabe

**Bearbeiten Sie das Übungsfile**

**Erstellen Sie eine Visualisierung für Ihre Forschungsfrage**

# Nächste Woche: Visualisierungen - Teil 2

# Referenzen

Healy, Kieran. 2018. *Data Visualization: A Practical Introduction*. Princeton, NJ: Princeton University Press.

Traunmüller, Richard. 2020. “Visualizing Data in Political Science.” In *The SAGE Handbook of Research Methods in Political Science and International Relations*, 436–60. 1 Oliver’s Yard, 55 City Road London EC1Y 1SP: SAGE Publications Ltd.  
<https://doi.org/10.4135/9781526486387.n28>.

