

07 Visualisierungen II

Einführung in die quantitativen Forschungsmethoden

Heute

- Wiederholung
- Weitere Arten von Plots
- Inspirationen
- Übungen

Wiederholung

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

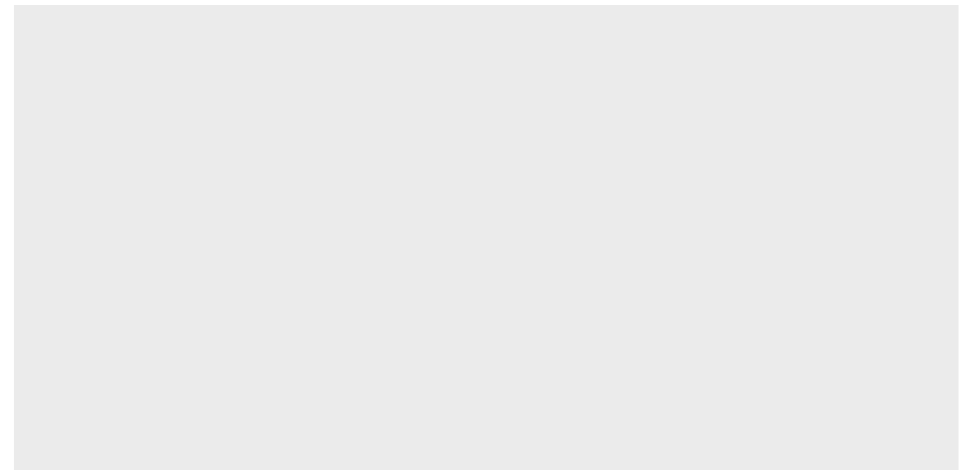
Beispiel

```
1 library(tidyverse)
2 library(haven)
3 ess8 <- read_dta("../data/ESS8e02_2.dta")
4
5 # gruppieren
6 ess8 <- group_by(ess8, cntry)
7
8 # zusammenfassen
9 democracy <- summarise(ess8,
10                       mean_stfgov=mean(stfgov, na.rm=T),
11                       mean_stfdem=mean(stfdem, na.rm=T))
12 # Demokratie Variable
13 democracy$cee <- ifelse(democracy$cntry %in%
14                         c("CZ", "EE", "HU", "LT", "PL", "SI"),
15                         "CEE", "other")
```

Beispiel

- `ggplot()` erzeugt Plot
- `data`: Datensatz

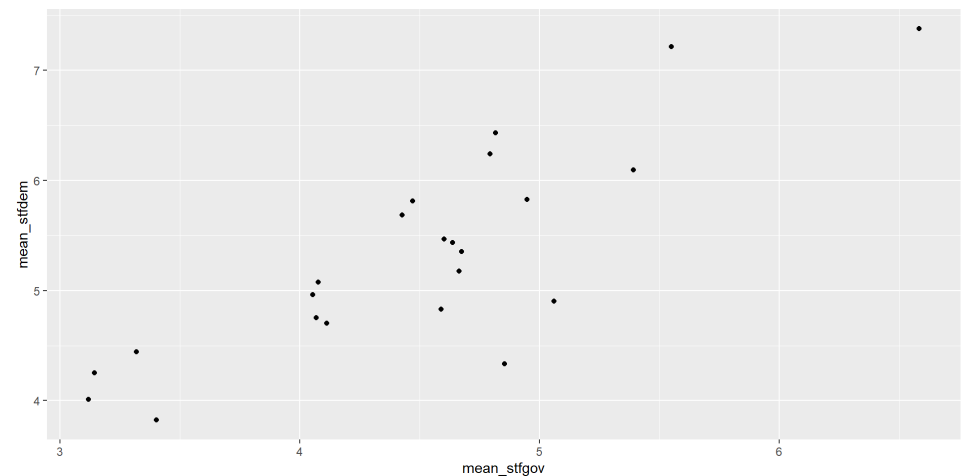
```
1 ggplot(data=democracy)
```



Beispiel

- geoms definieren die Art des Plots
- **aes** definiert x- und y-Werte

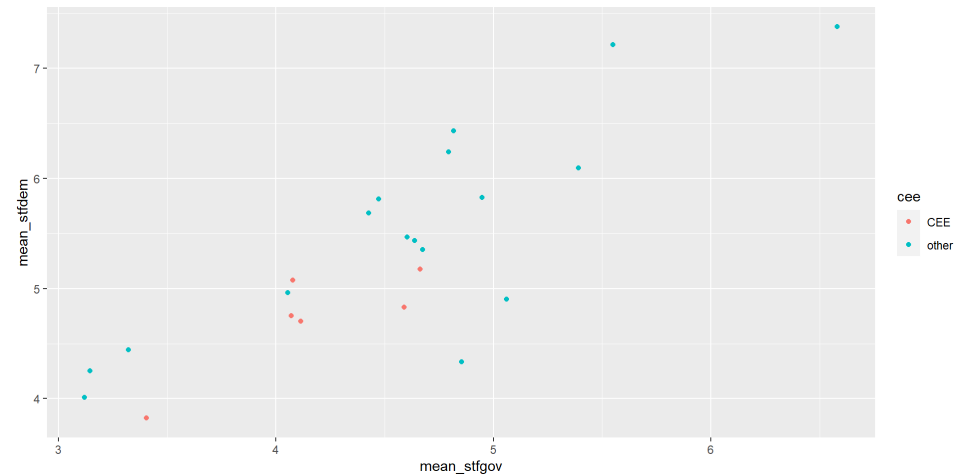
```
1 ggplot(data=democracy) +  
2   geom_point(aes(x=mean_stfgov,  
3                 y=mean_stfdem))
```



Beispiel

- `aes` definiert auch weitere Variablen

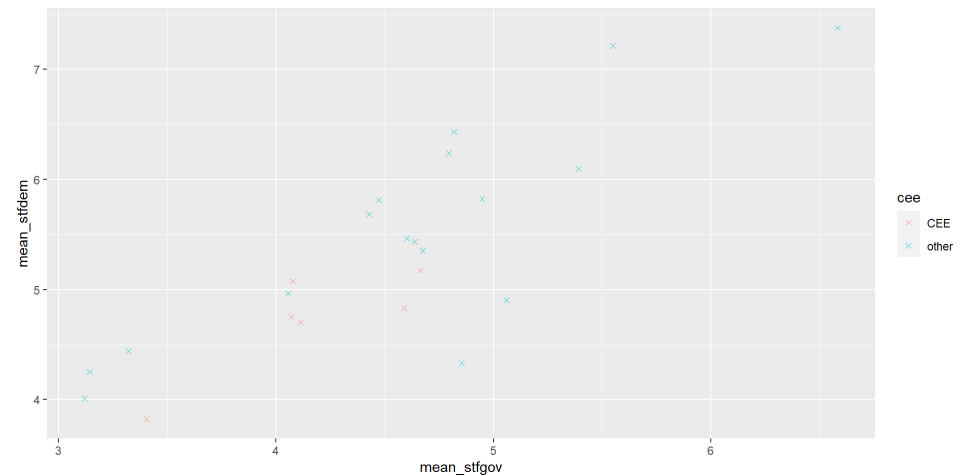
```
1 ggplot(data=democracy) +  
2   geom_point(aes(x=mean_stfgov,  
3                 y=mean_stfdem,  
4                 color=cee))
```



Heute neu

- Attribute verändern Eigenschaften unabhängig von Variablen - außerhalb von `aes`

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



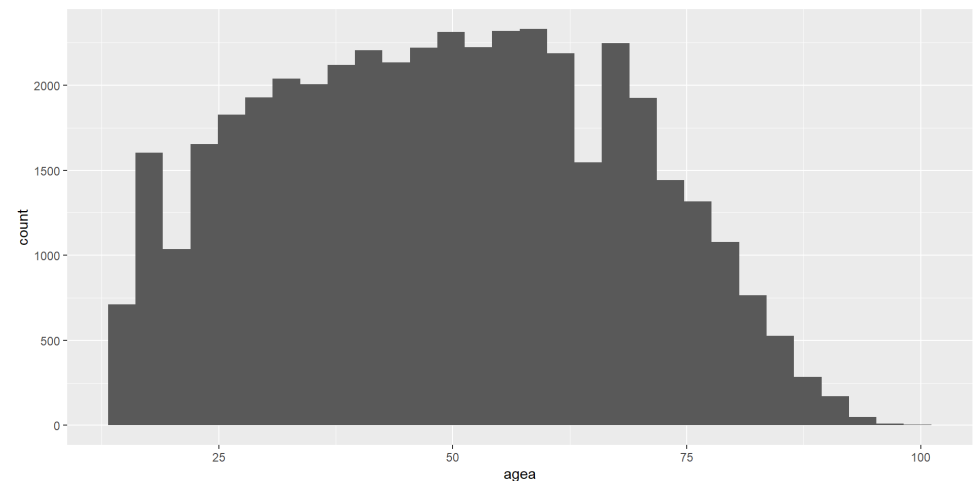
→ zum Beispiel für Farben

Weitere Arten von Plots

Häufigkeitsverteilungen

- als Histogramm mit `geom_histogram()`
 - zählt Kategorien aus (nur x-Wert)
 - bei sehr vielen Kategorien werden Kategorien zusammengefasst
 - `geom_bar()` funktioniert sehr ähnlich

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

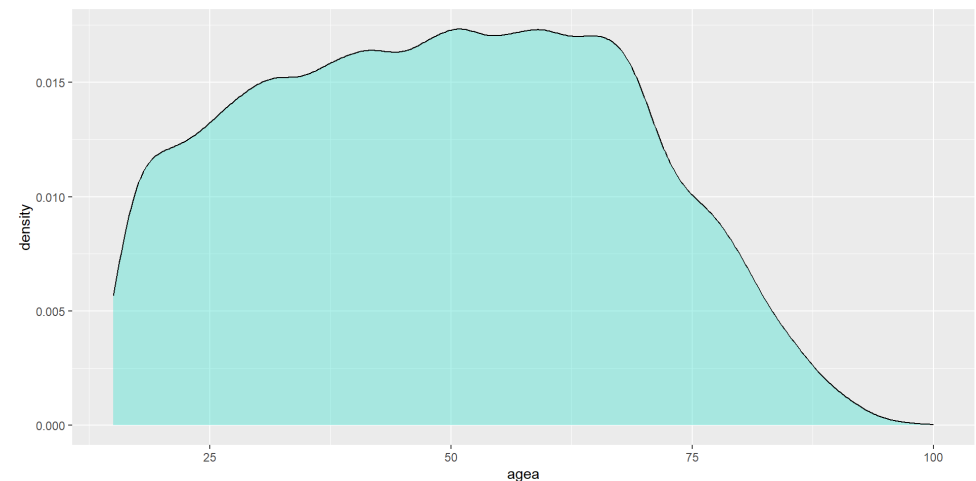


→ nützlich v.a. für wenige Kategorien

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)
```

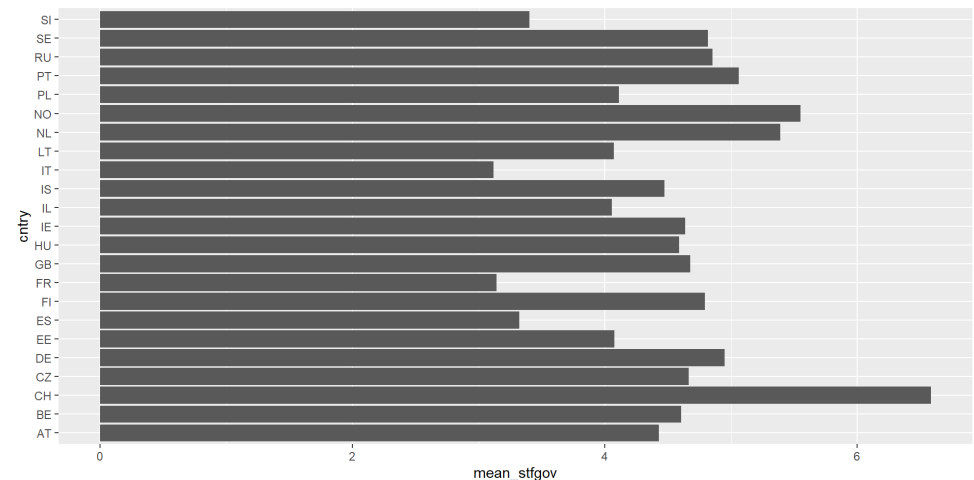


→ geeigneter für viele Kategorien

Bar chart

- Bar charts (Balkendiagramme) 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

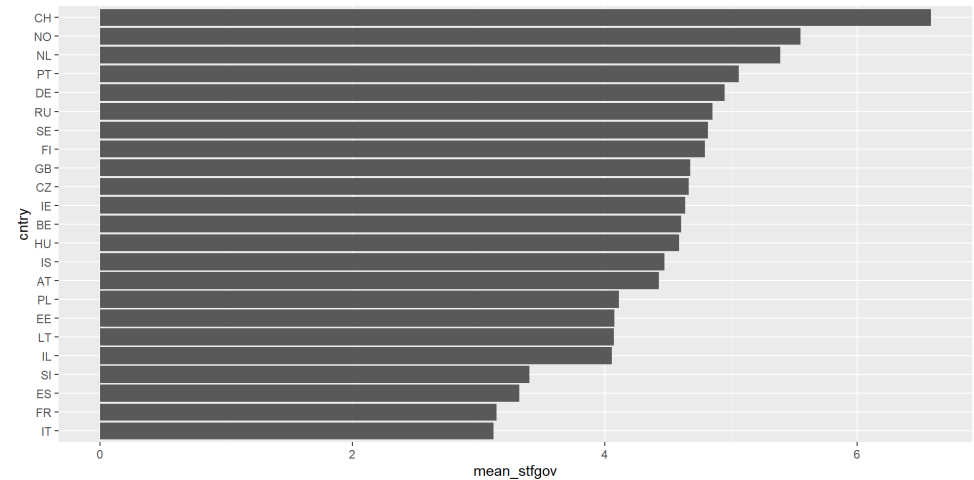
```
1 ggplot(data=democracy,  
2         aes(x=mean_stfgov,  
3             y=cntry)) +  
4   geom_col()
```



Bar chart

- **Ergänzung:** Wir können die Variable sortieren, damit die Grafik übersichtlicher wird

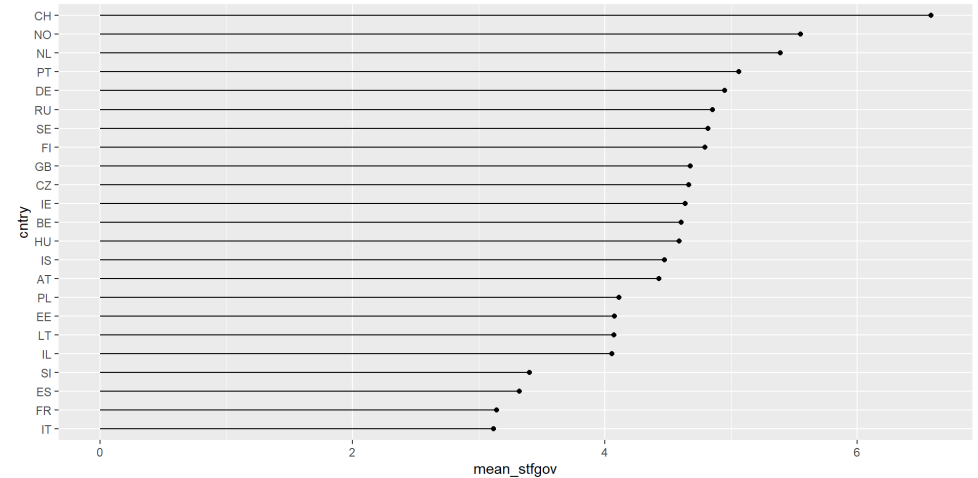
```
1 # sortieren der Werte
2 democracy <- arrange(democracy,mean_stfgov)
3 # festschreiben der Ordnung durch
4 # Faktorisierung der Variable cntry
5 democracy$cntry <- factor(democracy$cntry,
6                           democracy$cntry)
7
8 # Grafik
9 ggplot(data=democracy,
10        aes(x=mean_stfgov,
11            y=cntry))+
12 geom_col()
```



Lollipop chart

- Alternative: Lollipop chart mit `geom_segment()`
 - `xend` und `yend` → Linie dazwischen

```
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()
```

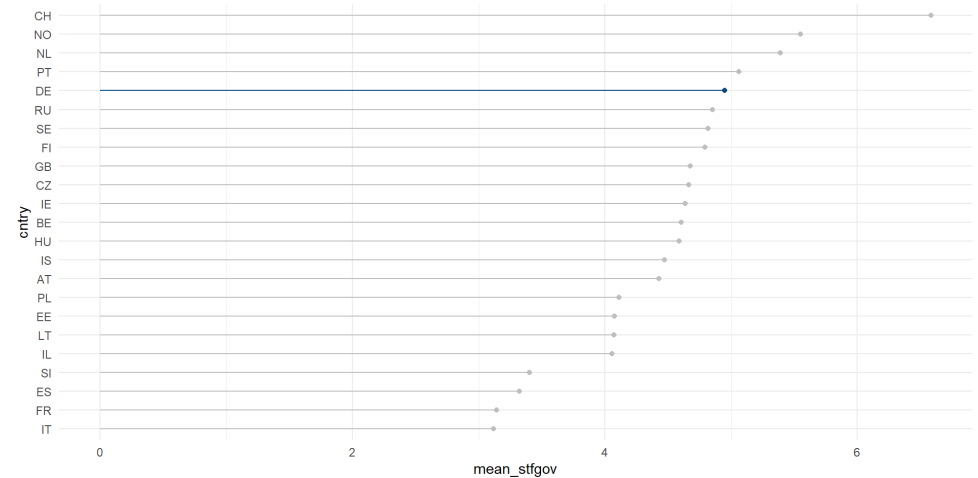


→ Achtung: `geom_segment()` braucht auch `xend` und `yend` als Spezifikation, weil es eine Fläche zwischen diesen Punkten markiert!

Lollipop chart

- Bsp.: Änderung der Farben über Variable

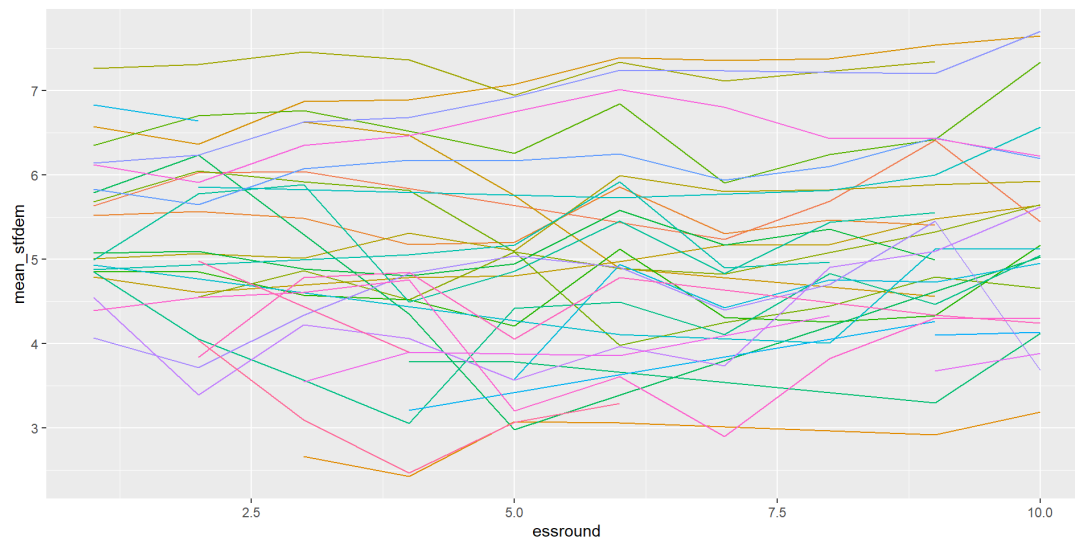
```
1 democracy$col <- ifelse(democracy$centry=="DE",  
2                           "highlight", "else")  
3 ggplot(data=democracy,  
4         aes(x=mean_stfgov, xend=0,  
5             y=centry, yend=centry,  
6             # Definition Farbe  
7             color=col))+  
8   geom_segment()+  
9 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



Linien

- vor allem für über-Zeit Vergleiche
 - Linie impliziert Kontinuität → keine Vergleiche über Gruppen hinweg

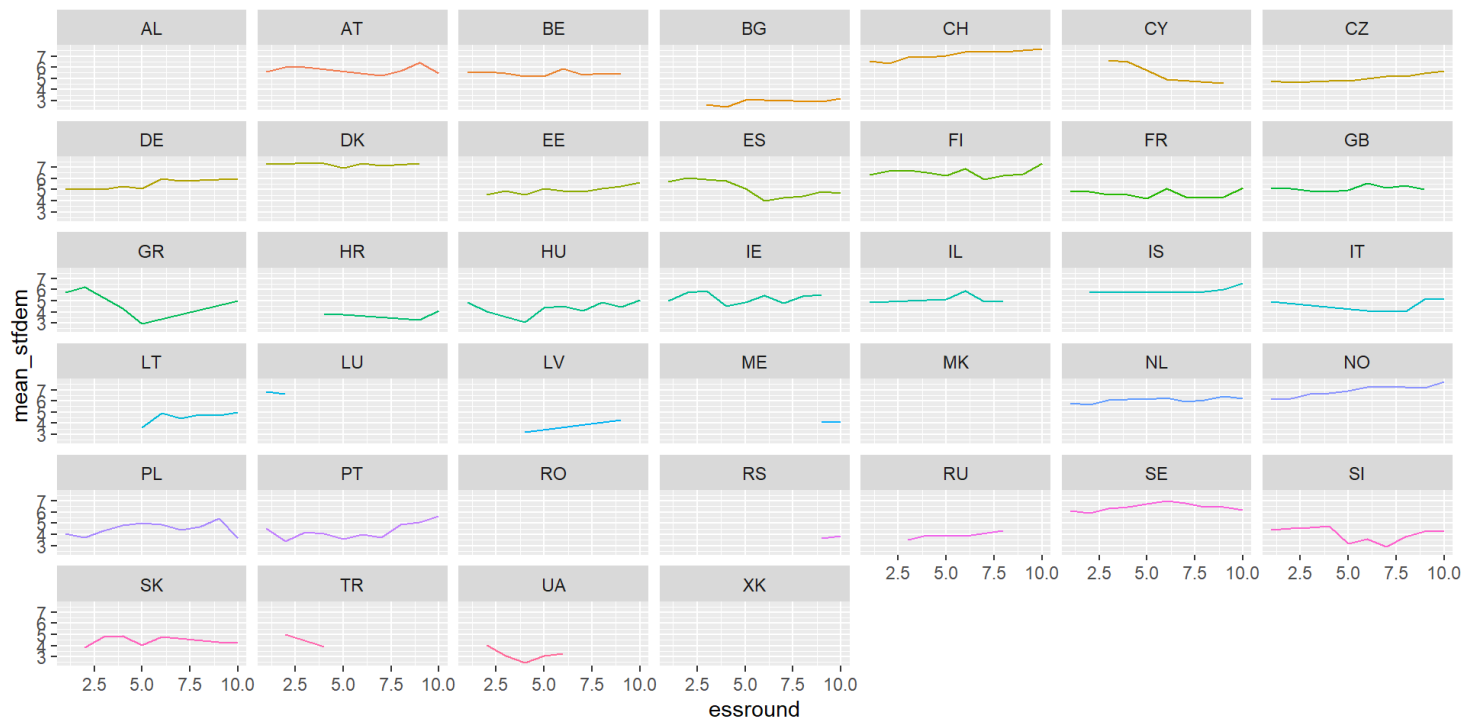
```
1 overtime <- haven::read_dta("../data/ESS-Data-Wizard-subset-2023-01-11.dta")
2
3 overtime <- group_by(overtime, cntry, essround)
4 overtime_stfdem <- summarize(overtime, mean_stfdem = mean(stfdem, na.rm=T))
5
6 ggplot(data=overtime_stfdem) +
7   geom_line(aes(x=essround, y=mean_stfdem, color=cntry)) +
8   # Legende unterdrückt
9   theme(legend.position="none")
```



Linien

- `facet_wrap()` stellt Gruppen deutlicher hervor!

```
1 ggplot(data=overtime_stfdem) +  
2   geom_line(aes(x=essround, y=mean_stfdem, color=cntry)) +  
3   facet_wrap(~cntry) + theme(legend.position="none")
```



Karten

Es gibt sehr viele weitere Möglichkeiten: z.B. ESS Karte ist auch mit R erzeugt



Figure 1: Länder nach ESS-Runde

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

Inspirationen

Herangehensweise

- **Was** will ich zeigen?
 - z.B. Verteilung, Beziehung, Unterschiede oder Gemeinsamkeiten...
- **Wie** zeige ich das?
 - Scatterplot, Bar chart, Linien,
 - Inspirationen auf Cheatsheet & Links (→ nächste Folie)

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](#)

Übungen

Aufgabe

Bearbeiten Sie das Übungsfile

Erstellen Sie eine Visualisierung für Ihre Forschungsfrage

Nächste Woche: Stichproben und Wahrscheinlichkeit

Llaudet/Imai (2023) S. 177-191

