

# Explorative Datenanalyse: Wahlen in Berlin

2023-01-30

## 1. Intro / Set up

- Erstelle ein **New Project...** mit jeweils einem Folder für Daten und Code.
- Erstelle ein Markdown Dokument und definiere die Working Directory.
- Kopiere die Daten in den Data Folder und lade sie mit einem relativen Pfad in dein Projekt

```
# setting the working directory with your relative path  
# knitr::opts_knit$set(root.dir = "yourPath")
```

Anwendung: Bundestagswahl in Berlin 2021. Wahldaten findest du hier.

## 2. Was passiert in den Daten?

Nutze dplyr, ggplot2 und tmap um die Daten zu verstehen und visualisieren. Suche dir eine oder mehrere Bezirke und Partei(en) heraus. Gruppiere die Daten nach Kategorien mit dplyr.

- In welchem Bezirk hat Partei X die meisten Stimmen bekommen?
  - Speichere die Daten in einem neuen dataframe.
  - Extra: Schreibe die Analyse als Funktion mit dem Argument “Partei”. So kannst du die Analyse schnell für verschiedene Parteien wiederholen. Hier findest du Unterstützung
- In welcher Adresse gab es die meisten ungültigen Stimmen?
  - Speichere die Analyse in einem neuen dataframe
- Hat die Linkspartei mehr Stimmen im Osten oder Westen der Stadt bekommen?
  - Beantworte die Frage rechnerisch.
  - Und visuell mit ggplot2. Nutze geom\_histogram und geom\_density. Ändere das alpha und die “position”. Mache eine passende Beschriftung der Achsen und einen Titel.
  - Bei Fragen zu ggplot2 schaue in dein Skript oder nutze das Internet.
- Download und speichere georeferenzierte Daten.
  - Geometrien der Berliner Wahllokale.
  - Geometrien der Wahlbezirke für die Wahlen zum Deutschen Bundestag in Berlin und zum Abgeordnetenhaus von Berlin 2021
  - Lade die Shapefiles mit read\_sf. Welche Bibliothek musst du dafür laden?
- Nutze tmap um die Datensätze erst einzeln und dann gemeinsam zu plotten.
  - Füge einen Titel hinzu, ändere die Farbe und Größe der Punkte.
  - Eine gute Dokumentation von tmap findest du hier.
- Wir möchten plotten wie Partei X in Berlin abgeschnitten hat.
  - Verbinde die Wahldaten mit Geometrien. Nutze dafür die Funktion left\_join().
  - Wie musst du die Daten vorher mit dplyr bearbeiten? Hinweis: Um Datensätze miteinander verbinden zu können brauchen wir einen individuellen key
  - Schreibe die Lösung als Funktion, um schnell die Ergebnisse von anderen Parteien plotten zu können.

### 3. Kontextualisierung

Lade weitere Daten herunter und verbinde sie mit unseren Ursprungsdaten. Wir nutzen Strukturdaten für die Wahl zum AGH.

- Wir brauchen einen einzigartigen Key zum mergen. Kannst du ihn finden oder musst du ihn erst kreieren?
- Wie können ein Objekt der Klasse `sf` und `df` gemerged werden?

### 4. Zusammenhänge in Daten finden

Wiederhole Schritt zwei mit dem neuen Datensatz.

- Welche sozio-ökonomischen Unterschiede finden sich zwischen Ost- und Westberlin?
- Wie kannst du sie visualisieren?
- Wie korrelieren sie mit Wahlergebnissen?