

Datenvisualisierung mit R

Armin Schäfer

Wintersemester 2024/25

E-Mail: schaefer@politik.uni-mainz.de
Raum: GFG 01-512

Uhrzeit: 8.00-10.00 Uhr
Büro: GFG 04-436

Beschreibung

Inhalt: Der kompetente Umgang mit Daten und deren grafisch Darstellung werden immer wichtiger. Wissenschaftliche Ergebnisse werden immer häufiger grafisch dargestellt und nicht nur im Text beschrieben. Auch im Journalismus gibt es zunehmend spezialisierte Teams, die sich ausschließlich mit Datenvisualisierung beschäftigen. Im Zentrum dieses Forschungsseminars steht deshalb die Arbeit mit dem kostenfreien Statistikprogramm “R”. Alle Teilnehmer:innen werden lernen, wie sich Daten mit diesem Programm bearbeiten und auswerten lassen, um Analysen durchführen und grafische Darstellungen erstellen zu können. Dabei gehen wir streng *anwendungsorientiert* vor, denn dieses Forschungsseminar wird keine Einführung in die Statistik ersetzen. Ziel ist es, verschiedene Varianten der Datenvisualisierung zu erlernen, die Sie im weiteren Studium nutzen können. Um teilnehmen zu können, sind keine Vorkenntnisse notwendig.

Aktive Mitarbeit

Um den Lernerfolg sicherzustellen, müssen Sie eine Übungsaufgabe bearbeiten und den Lösungsweg in Moodle hochladen. Diese Abgabe ist *verpflichtend*.

Ablauf

17. April: Einführung

24. April: Grundlagen

Vorab lesen: Kapitel 1 aus [Data Visualization. A practical introduction](#).

- a) R, RStudio und Quarto
- b) Projekte anlegen & mit Skripten arbeiten
- c) Bibliotheken installieren und aufrufen
- d) Erstes Beispiel: Ein simulierter Datensatz
- e) Datensätze inspizieren
- f) Die erste Grafik!

8. Mai: Balkendiagramme

Vorab lesen: Kapitel 6 aus [Fundamentals of Data Visualization](#).

- a) Excel-Daten laden: csv und xlsx
- b) Balkendiagramm nach Parteizugehörigkeit
- c) Den Parteien Farben zuordnen
- d) Langes Datenformat
- e) Balkendiagramme nach Partei und Geschlecht

15. Mai: Gestaltungsarten – die “Geome”

Beispiel: Abgeordnete im Bundestag

- a) Punktediagramm: Durchschnittliches Alter
- b) Alle Datenpunkte zeigen: Jitter
- c) Verteilungen anzeigen: Histogramm und Density plot
- d) Alpha effektiv einsetzen
- e) Das Erscheinungsbild ändern

22. Mai: Übungswoche

Bitte bearbeiten Sie bis zum **2. Juni (18:00 Uhr)** die in Moodle beschriebene Übungsaufgabe und laden Sie Ihr Skript als html-Dokument in den dafür vorgesehenen Ordner hoch. Nach Ablauf der Frist lade ich eine (einfache) **Musterlösung** hoch, die Woche danach lernen wir weitere Schritte, um das Skript weniger fehleranfällig zu machen.

5. Juni: Daten umformen und weitere Darstellungsarten

- a) Variablen auswählen, filtern und umbenennen
- b) Variablen verändern: `mutate()`, `case_when()` und `if_else()`
- c) Missings
- d) Mit Gruppen arbeiten

12. Juni: Funktionen I: Unnötige Wiederholungen vermeiden

Never type the same code twice.

- a) Funktionen erstellen
- b) Funktionen auf mehrere Spalten anwenden
- c) Eine ggplot-Funktion
- d) Ein eigenes “theme” erstellen

26. Juni: Funktionen II: Grafiken automatisiert erstellen

- a) Mehr zu `mutate(across())`
- b) Eine Funktion für Grafiken
- c) Grafiken kombinieren
- d) Ein “theme” selbst erstellen

3. Juli: Trends im Zeitverlauf

Beispiel: Autokratisierung

- a) Trends in wenigen Ländern
- b) Trends in vielen Ländern pro Jahrzehnt
- c) Dumbbell plots
- d) Faceting
- e) Highlighting Beispiel: Wahlbeteiligung und Parteiergebnisse in den Wahlkreisen

10. Juli: Lineare Regressionsmodelle und Interaktionen grafisch darstellen

Beispiel: Wo schneiden Parteien gut oder schlecht ab?

- a) Das unersetzliche Streudiagramm
- b) Trendlinien hinzufügen
- c) Koeffizienten als Abbildung
- d) Interaktionen einfügen
- e) Interaktionen darstellen

17. Juli: Logistische Regression und vorhergesagte Wahrscheinlichkeiten

Beispiel: Wer wählt die Grünen?

- a) Logistische Regressionen
- b) Koeffizientenplots
- c) Vorhergesagte Wahrscheinlichkeiten
- d) Marginale Effekte