

Übungsklausur Empirisches Arbeiten

Teil Explorative Datenanalyse mit R

Prüfer	Prof. Dr. Nicolas Meseth
Semester	WS 24/25
Max. Punktzahl	20
Erlaubte Hilfsmittel	keine

Hinweise zu diesem Klausurteil

- Bitte nutzt die Datei `nachname_vorname_lösungen.R` für die Beantwortung der Fragen und fügt euren R-Code jeweils unter der Frage ein. Bitte entfernt am Ende alle Codereste, die nicht zur Antwort gehören.
- Denkt daran, eure Matrikelnummer und Namen vor der Bearbeitung in die ersten beiden Zeilen einzutragen.
- Ersetzt vor der Abgabe eure Vor- und Nachnamen im Dateinamen. Als Beispiel: `mustermann_max_lösungen.R`
- Ladet die Datei über den Abgabeordner im ILIAS-Lernraum der Veranstaltung hoch! Die Abgabe muss vor dem offiziellen Ende der Bearbeitungszeit erfolgen!
- Es sind keine Hilfsmittel erlaubt!

Datensatz “Wahlpräferenzen”

In dieser Übungsklausur könnt ihr insgesamt **20 Punkte** erreichen. Bei 10 oder mehr Punkten habt ihr bestanden und ihr bekommt die Punkte > 10 als Bonus gutgeschrieben.

Die Daten für diese Übungsklausur liegen im Ordner `data/voting`. Es handelt sich um drei einfache Textdateien, in denen in jeder Zeile eine Antwort aus einer Befragung zum Wahlverhalten einer Person steht. In der Datei `age.txt` wurde das Alter festgehalten, in der `gender.txt` das Geschlecht, und in der `party.txt` die Partei, die diese Person zum Zeitpunkt der Umfrage wählen würde.

Aufgabe 1: Vektoren, Data Frames und Tibbles

a) Ladet alle drei Dateien in separate Vektoren in R! (3 Punkte)

Erinnert euch daran, dass die `scan()`-Funktion standardmäßig numerische Werte erwartet. Nutzt für nicht-numerische Werte das Argument `what = "character"`.

Um die Wahlpräferenz korrekt zu laden und den Umlaut im Wort "Grüne" korrekt darzustellen, müsst ihr der `scan()`-Funktion zusätzlich das Argument `fileEncoding = "UTF-8"` mitgeben.

```
# Fügt eure Lösung bitte in die .R-Datei unter dieser Frage ein
party <- scan("data/voting/party.txt",
             what = "character",
             fileEncoding = "UTF-8"
            )
```

b) Ermittelt auf Basis der Vektoren das Durchschnittsalter der Grünen-Wählerinnen und Wähler! (4 Punkte)

```
# Fügt eure Lösung bitte in die .R-Datei unter dieser Frage ein
```

c) Erstellt aus den drei Vektoren einen Data Frame mit dem Namen `voting` und den Spalten `gender`, `age` und `party`! (2 Punkte)

```
# Fügt eure Lösung bitte in die .R-Datei unter dieser Frage ein
```

d) Erzeugt jetzt einen Tibble aus eurem Data Frame und nennt ihn `voting_tbl`! (1 Punkt)

```
# Fügt eure Lösung bitte in die .R-Datei unter dieser Frage ein
```

Aufgabe 2: Datenvisualisierung

In dieser Aufgabe verwendet ihr den Tibble `voting_tbl` aus der Aufgabe 1d. Solltet ihr an der Erstellung des Tibbles gescheitert sein, so ladet ihn mit der folgenden Zeile:

```
voting_tbl <- read_csv("data/voting/voting.csv")
```

Findet eine passende Visualisierungsform für die folgenden Fragen und erstellt diese mit R und ggplot2!

a) Wie viele Stimmen erhält jede Partei auf Basis des vorliegenden Datensatzes? (3 Punkte)

```
# Fügt eure Lösung bitte in die .R-Datei unter dieser Frage ein
```

b) Wie ist Altersverteilung der Wählerschaft für jede Partei? Erstellt eine Visualisierung, die einen schnellen Vergleich ermöglicht! (4 Punkte)

```
# Fügt eure Lösung bitte in die .R-Datei unter dieser Frage ein
```

c) Wie ist die Verteilung des Alters im gesamten Datensatz? (3 Punkte)

```
# Fügt eure Lösung bitte in die .R-Datei unter dieser Frage ein
```

Viel Erfolg!