

Klausur Empirisches Arbeiten

Teil Explorative Datenanalyse mit R

Prüfer	Prof. Dr. Nicolas Meseth
Semester	WS 23/24
Max. Punktzahl	40 (oder $\frac{1}{3}$ der Gesamtpunktzahl)
Erlaubte Hilfsmittel	alle

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: `max_mustermann_lösungen.R`
- Ladet die Datei über den Abgabeordner im ILIAS-Lernraum der Veranstaltung hoch! Die Abgabe muss vor dem offiziellen Ende der Bearbeitungszeit erfolgen!

Teil 1: Datensatz “REWE-Produkte”

Im ersten von zwei Teilen könnt ihr insgesamt **20 Punkte** erreichen.

Bevor ihr mit der Bearbeitung der Aufgaben beginnt, kopiert die Datei `rewe_products.csv` in euer Arbeitsverzeichnis und ladet den Datensatz als Tibble mit dem Namen `rewe`.

```
library(tidyverse)
rewe <- read_csv("rewe_products.csv")
```

Aufgabe 1.1: Datentransformation

Beantwortet die folgenden Fragen mit R und dem Tidyverse. Das Ergebnis soll in diesem Teil als Tabelle (Tibble) und *nicht* als Visualisierung dargestellt werden.

a) Gebt alle Spaltennamen des Datensatzes aus, die bool'sche Werte enthalten! (1 Punkt)

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

b) Listet alle Produkte, denen Salz *zugesetzt* wurde! (2 Punkte)

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

c) Welche fünf Produkte bieten den besten Preis pro Gramm enthaltenem Protein? Nutzt als Startpunkt die neue Spalte grams, die das Produktgewicht in Gramm aus dem Feld grammage extrahiert! (7 Punkte)

```
# Fügt eure Lösung bitte in die .R-Datei unter dieser Frage ein
rewe |>
  mutate(grams = as.numeric(str_match(grammage, "^(\d+)g")[,2]))
```

Aufgabe 1.2: Datenvisualisierung

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

a) Enthalten Milchprodukte durchschnittlich mehr Eiweiß als andere Produkte? (5 Punkte)

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

b) Wie ist die Verteilung des Salzgehaltes für jede Unterkategorie der Kategorie "Nahrungsmittel"? Wählt eine sinnvolle Visualisierungsform, um die Verteilungen gut miteinander vergleichen zu können und zeigt nur den relevanten Bereich der Daten! (5 Punkte)

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

Teil 2: Datensatz “Energie”

Im zweiten Teil könnt ihr insgesamt **20 Punkte** erreichen!

Bevor ihr mit der Bearbeitung der Aufgaben beginnt, kopiert die Datei `owid-energy-data.csv` in euer Arbeitsverzeichnis und ladet den Datensatz als Tibble mit dem Namen `energy`. Die Daten stammen aus Ritchie, Rosado, and Roser (2023). Ein Codebuch für die Spalten findet ihr unter [diesem Link](#) (wird für diese Klausur nicht benötigt).

```
library(tidyverse)
energy <- read_csv("owid-energy-data.csv")
```

Der **Datensatz** enthält Informationen zur Energie- und Elektrizitätserzeugung und zum Verbrauch aller Länder der Welt im Zeitverlauf.

Aufgabe 2.1: Datentransformation

Beantwortet die folgenden Fragen mit R. Das Ergebnis soll in diesem Teil als Tabelle (Tibble) und *nicht* als Visualisierung ausgegeben werden.

a) **Wie viele Variablen und Beobachtungen enthält der Datensatz?** (1 Punkt)

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

b) **Wählt alle Variablen aus, die einen Wert “pro Kopf” enthalten!** (3 Punkte)

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

c) **Welche 10 Länder hatten im Jahr 2021 den höchsten Energiebedarf (energy_per_capita) pro Einwohner?** (3 Punkte)

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

d) **Welches Land hatte 2019 die prozentual größte Reduktion beim Konsum fossiler Brennstoffe im Vergleich zum Vorjahr? (fossil_cons_change_pct)? Auf welchem Platz liegt Deutschland?** (3 Punkte)

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

Aufgabe 2.2: Datenvisualisierung

a) Überprüft einen möglichen Zusammenhang zwischen der Einwohnerzahl (population) und dem Anteil an erneuerbaren Energien bei der Erzeugung von Strom (renewables_share_elec)? Betrachtet dabei das Jahr 2021 und nur Länder mit weniger als 100 Mio. Einwohner! (5 Punkte)

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

b) Visualisiert die Entwicklung des Anteils der Stromerzeugung aus Atomkraftwerken (nuclear_share_elec) für die G7-Länder über die Zeit! (5 Punkte)

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

Quellen

Ritchie, Hannah, Pablo Rosado, and Max Roser. 2023. "Energy." *Our World in Data*.