

Klausur Empirisches Arbeiten im WS 21/22 - Teil R

Hinweise zu diesem Klausurteil

- Ihr könnt in diesem Teil der Klausur maximal **40 Punkte** erreichen.
- Der Teil wird zu einem Drittel in der Gesamtnote gewichtet.
- Bitte nutzt das Script `lösungen.R` für die Beantwortung der Fragen und fügt euren R-Code jeweils unter die 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.
- Benennt vor der Abgabe die Datei nach folgendem Muster: `nachname_vorname.R` und verwendet die Abgabefunktion in der OSCA-Veranstaltung des Moduls für die endgültige Abgabe!

Teil 1: Datensatz Environmental Impacts of Food Production

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

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

```
library(tidyverse)

fp <- read_csv("food_production.csv")
```

Aufgabe 1.1: Datenverarbeitung

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

a) Gebt alle Spaltennamen des Datensatzes aus! (1 Punkt)

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

b) Welche 3 Produkte verbrauchen am meisten Wasser pro 1000 kcal? (2 Punkte)

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

c) Welche 5 Produkte benötigen die kleinste Anbaufläche pro Kg? (2 Punkte)

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

d) Erstellt zwei neue Spalten. Die Erste soll den durchschnittlichen Wert der Gesamtemissionen über alle Produkte enthalten. Die Zweite soll TRUE sein, wenn das Produkt den durchschnittlichen Wert der Gesamtemissionen über alle Produkte überschreitet. Vergibt sprechende Namen für die neuen Spalten und behaltet nur sie im Ergebnis! (5 Punkte)

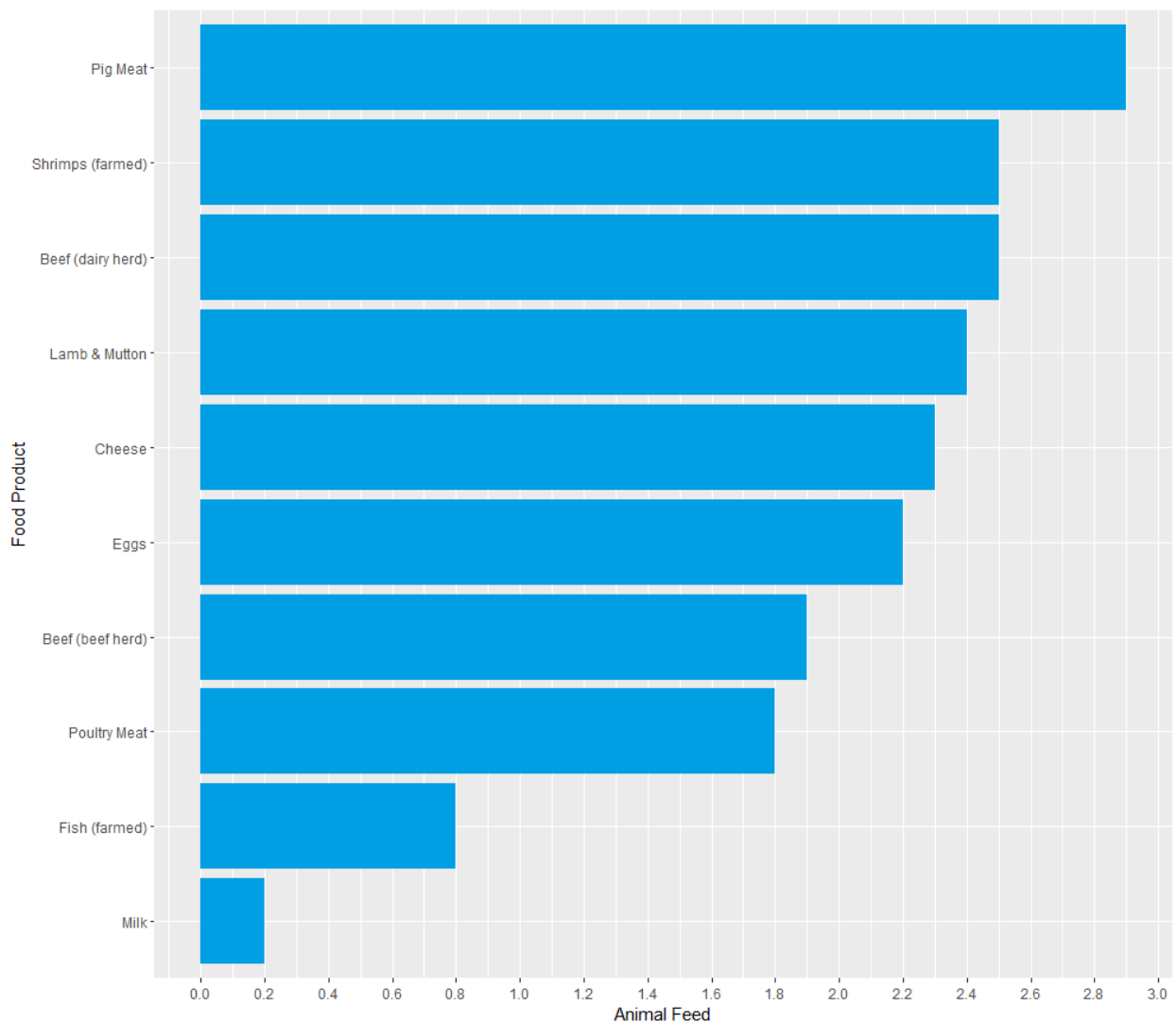
Fügt eure Lösung bitte in die Datei lösungen.R unter dieser Frage ein

Aufgabe 1.2: Datenvisualisierung

a) Erstellt den R-Code für das Balkendiagramm unten und versucht dabei, das Diagramm möglichst exakt nachzubilden! (5 Punkte)

Hinweis: Der Farbcode der Balken ist #009ee3.

Fügt eure Lösung bitte in die Datei lösungen.R unter dieser Frage ein



b) Emittieren Produkte, die eine große Fläche pro Kilogramm beanspruchen, auch mehr CO₂? Erstellt eine geeignete Visualisierung, um diesen Zusammenhang zu überprüfen! (5 Punkte)

Fügt eure Lösung bitte in die Datei lösungen.R unter dieser Frage ein

Teil 2: Datensatz Orangenlimonade

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

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

```
limo <- read_csv("limonade.csv")
```

Aufgabe 2.1: Datenverarbeitung

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

a) **Erstellt einen Tibble, der nur die numerischen Spalten enthält!** (1 Punkt)

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

b) **Welches sind die häufigsten Postleitzahlen unter den Teilnehmer_innen_?** (1 Punkt)

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

c) **Wie viel % der befragten Personen sind weiblich?** (4 Punkte)

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

d) **Wie alt ist die älteste Person, die kürzlich Limonade gekauft hat (f1)?** (4 Punkte)

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

Aufgabe 2.2: Datenvisualisierung

a) **Erstellt Boxplots für die Verteilung des Preises, den die Probanden als passend für eine 0,3 l Flasche Fanta angegeben haben. Stellt die Antworten der weiblichen und männlichen Teilnehmer:innen in einem Diagramm aber jeweils als eigene Box dar!** (6 Punkte)

Hinweise:

- Die Antworten findet ihr in der Spalte `f18_preisoffen_A_fanta`.
- Die Spalte `f53_geschlecht` müsst ihr in einen Faktor umwandeln, z. B.: `mutate(f53_geschlecht = as.factor(f53_geschlecht))`

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

b) **Gibt es Unterschiede bei der Preisbereitschaft für 0,3 l Fanta (`f18_preisoffen_A_fanta`) und dem Schulabschluss (`f40_schulabschluss`)? Wählt eine geeignete Visualisierung, um das zu prüfen!** (4 Punkte)

Hinweis: Es gelten die folgenden Kodierungen für die Werte der Spalte `f40_schulabschluss`:

- 0 = keine Angabe
- 1 = Erwerbslos

- 2 = Hausfrau/Hausmann
- 3 = Rentner/Pensionär
- 4 = In Ausbildung/Studium
- 5 = Teilzeiterwerbstätig
- 6 = Vollerwerbstätig

Fügt eure Lösung bitte in die Datei lösungen.R unter dieser Frage ein