

dlypr: Tödliche Polizeischüsse

2022-11-25

Intro

Zuerst laden wir die Libraries und die Daten

```
# Falls sie noch nicht installiert sind entferne den # damit der Code ausgeführt wird  
# install.packages("dplyr")  
# install.packages("readr")  
  
# Falls du die Pakete schon heruntergeladen hast  
library(dplyr)  
library(readr)
```

Eventuell hast du einen anderen Pfad! Falls das der Fall ist kann du den Datensatz über click-and-point (File -> Import Dataset -> auswählen) installieren oder indem du den richtigen Pfad eingibst.

```
# deadly_cops <- read_csv("data/deadly_cops/polizeischuesse.csv")  
  
# tödliche Polizeischüsse in Deutschland seit 1976  
# Eine erste Übersicht verschaffen wir uns mit glimpse()  
# glimpse(deadly_cops)
```

Aufgaben

Basics

1. In welcher Stadt wurden die meisten Menschen von der Polizei erschossen?
2. Wie viele Menschen wurden seit Beginn der Aufzeichnung von der Polizei pro Bundesland erschossen?
3. In welcher Stadt wurden die meisten Menschen von der Polizei in Bayern erschossen?

Neue Variablen erstellen I

4. Wie viele Menschen wurden pro Jahr in Deutschland von der Polizei erschossen?
 - Erstelle dafür eine neue Variable "Jahr". Nutze dazu `format()` in Verbund mit `as.Date()`.
5. In welchem Bundesland gab es die meisten tödlichen Polizeischüsse in den 1990er Jahren und in welchem in den 2000ern?
6. Wie viele minderjährige Menschen wurden von der Polizei erschossen?
 - Erstelle dafür eine neue Variable mit `mutate()` und `ifelse()`

Neue Variablen erstellen II

1. In welchem Bundesland gab es die meisten Hinweise auf eine psychische Ausnahmesituation relativ zu allen Fällen?
 - Ersetze dafür zuerst die NA mit “Nein”. Dafür kannst du die Funktion `replace_na()` aus der Bibliothek `tidyr` nutzen. Falls du die noch nicht heruntergeladen hast mach das mit `install_packages()`
 - Gruppiere die Daten. Wende dann `pivot_wider()` an.
 - Berechne dann die Prozentzahl. Dafür musst du eventuell wieder NAs erstzen..
2. Wie haben sich die Hinweise auf psychische Ausnahmesituationen über die Zeit verändert? Erstelle eine neue Variable `Datum`, welche die Daten in Jahrzehnte gruppiert. Berechne dann die Hinweise auf psychische Ausnahmesituationen over time in Prozent
3. Kreiere eine neue Variable “Quelle”, welche den Wert “Ja” annimmt, falls es eine Berichterstattung in der Presse gab und “Nein”, falls das nicht der Fall ist.
 - Bei welchen Bundesländern gab es die meiste, bei welchen die wenigste Presseberichterstattung?
 - Bei manchen Vorfällen ist ein Wikipedia Artikel verlinkt. Erstelle eine neue Variable, die angibt, ob es sich bei der Quelle um einen Wikipediaartikel handelt.
 - Tip: Nutze die Funktion `str_detect()` aus dem package `tidyr` um die Präsenz von einem string in einer Variablen zu finden.

Joining Data

1. In welchem Bundesland gab es die meisten tödlichen Polizeischüsse, relativ zur Einwohner*innenanzahl?
 - Erstelle einen Vektor mit `c()`, der die Namen aller Bundesländer hat und einen zweiten Vektor, der die Einwohner:innenanzahl beinhaltet. Auch wenn etwas analytische Tiefe verloren geht nutzen wir die gleiche Einwohner:innenanzahl für alle Jahre.
 - Verbinde die vektoren zu einem Dataframe mit der Funktion `data.frame()`
 - Für welchen Zeitraum solltest du die Daten zu tödlichen Polizeischüssen filtern, damit ein Vergleich nach Bundesland Sinn macht?
 - Verbinde nun beide Datensätze mit der Funktion `df_join`. Unterstützung findest du, wenn du `?left_join()` eingibst.
2. Werden mehr Menschen von der Polizei in den alten oder neuen Bundesländern erschossen, wenn wir für die Einwohner:innenanzahl kontrollieren? Beginne mit unseren Ausgangsdatsatz.
 - Erstelle auch hierfür einen Vektor mit neuen oder alten Bundesländern. Erstelle eine neue Variable `neue_BLs`, eine Aussage trifft, ob es sich um ein neues BL handelt.
 - Verbinde dann die Daten mit unserem Vektor zu Einwohner:innen
 - Und berechne die Fälle pro neues/altes Bundesland, relativ zu Einwohner:innenanzahl