Data Programming

Homework 5

Lucas Lorenzo Jakin

2025-04-01

Solutions

Submit a **.html** file created using Quarto via e-classroom. A sample is attached. Display all code (packages,input) and output.

Import the database from **https://www.kaggle.com/datasets/konradb/fatal-encounters-database**

It contains the 2 dataframes shown below.

```
library(readr)
library(tidyverse)
library(ggplot2)

dataset <- read.csv("dataset.csv", sep = ";")
state_abrv_pop <- read.csv("state_abbreviations_and_populations.csv", sep = ";")
```

1

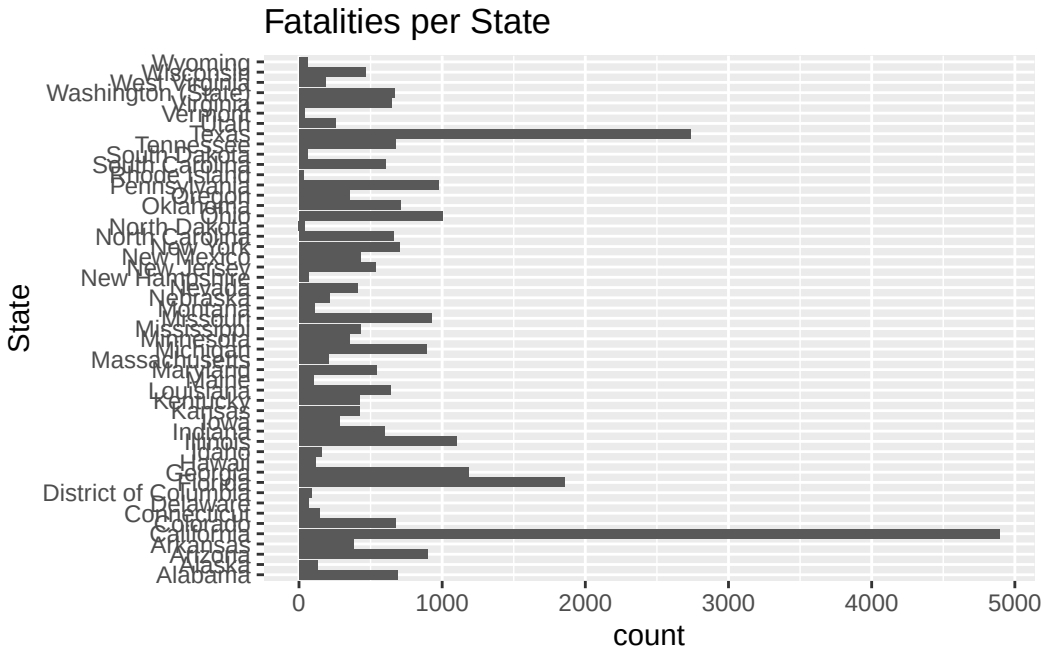
Show the number of fatalities per state (full name of state required). Furthermore, plot the results with a histogram.

```
# Solve 1 here

#a) show number of fatalities per state:
data_merged <- inner_join(state_abrv_pop, dataset, by = c("Abbreviation" = "State"))

fatalities_state <- data_merged %>% group_by(State = .$Name.x) %>%
  summarise(n = n())
```

```r
fatalities_state
```

```
# A tibble: 51 x 2
   State                   n
   <chr>               <int>
 1 Alabama               691
 2 Alaska                126
 3 Arizona               900
 4 Arkansas              384
 5 California           4892
 6 Colorado              673
 7 Connecticut           142
 8 Delaware               63
 9 District of Columbia   88
10 Florida              1851
# i 41 more rows
```

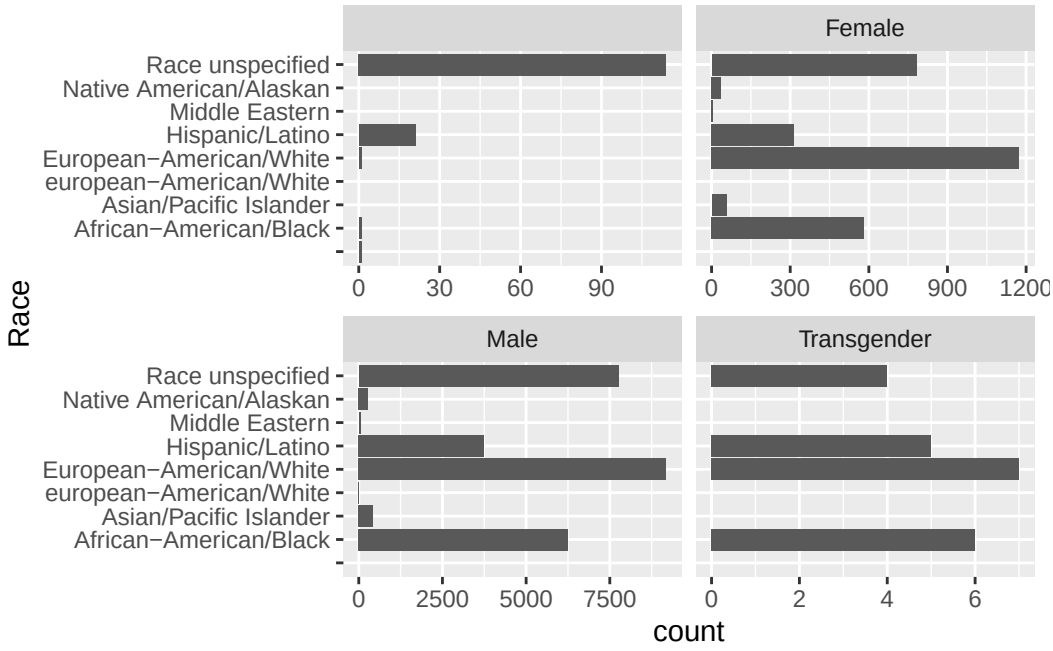
```r
#b)
ggplot(fatalities_state, aes(x = n, y = State)) +
  geom_histogram(stat = "identity") +
  labs(title = "Fatalities per State",
       x = "count",
       y = "State")
```



2

Plot death by gender across race with a histogram, entirely with R, not with markdown/quarto!

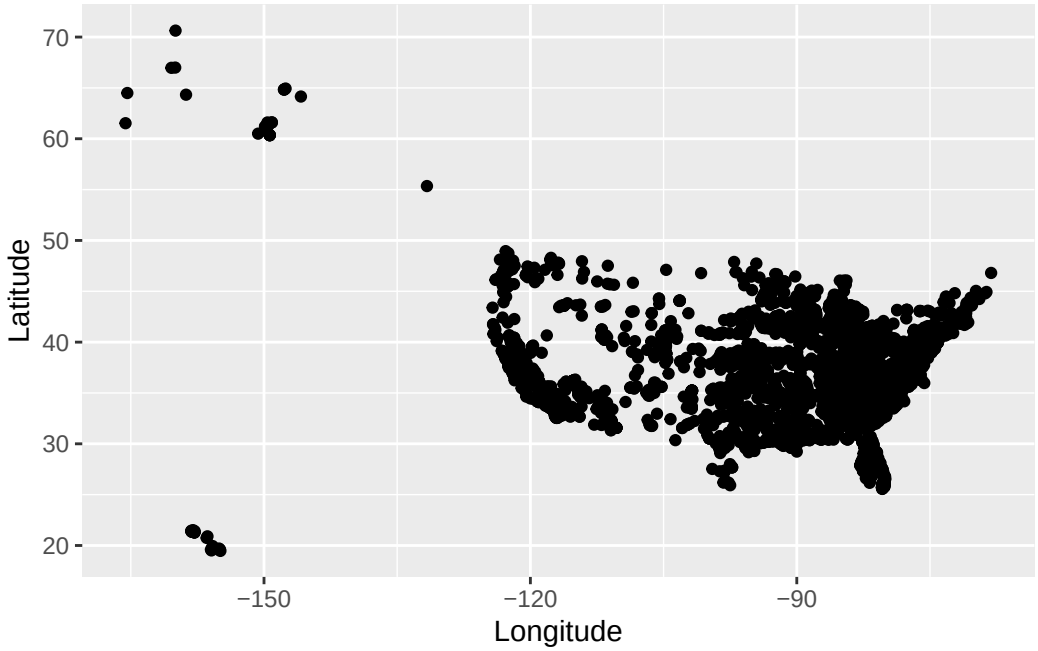
```
# Solve 2 here
gender_raceData <- dataset %>% group_by(Gender, Race) %>% summarise(n = n(), .groups = 'dr
ggplot(gender_raceData, aes(x = n, y = Race))+
  geom_col() +
  facet_wrap(~Gender, scales = "free_x")+
  labs(x = "count", y = "Race")
```



3

Plot the outline of the USA by scatterplotting the locations of deaths.

```
#Solve 3 here
dataset <- dataset %>% mutate(LongitudeNew = as.numeric(Longitude),
                              LatitudeNew = as.numeric(Latitude))
dataset_clean <- dataset %>% filter(!is.na(LongitudeNew) & !is.na(LatitudeNew))
ggplot(dataset_clean, aes(x = LongitudeNew, y = LatitudeNew)) +
  geom_point(color = "black")+
  labs(x = "Longitude",
       y = "Latitude")
```


70
60
50
40
30
20
Latitude
−150
−120
−90
Longitude