Data 607 Final Project Proposal

Zaneta Paulusova & Inna Yedzinovich

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(tidyr)
library(dplyr)
library(ggplot2)
library(usmap)
```

Introduction:

We will examine the relationship between the number of drivers and the incidence of fatal accidents in the USA. Our focus will be on fatality data from the National Highway Traffic Safety Administration and driving population statistics from the Federal Highway Administration for the years 2018, 2020, and 2022. Our analysis will investigate whether the number of fatalities increases as the driving population grows. I believe that an increase in the number of drivers is likely to lead to an increase in the number of fatalities as well.

Data Source

Data of fatal accidents per state - https://www-fars.nhtsa.dot.gov/states/statesfatalitiesfatalityrates.aspx

Population of drivers per state: - 2022 - https://www.fhwa.dot.gov/policyinformation/statistics/2022/dl1c.cfm - 2020 - https://www.fhwa.dot.gov/policyinformation/statistics/2020/dl1c.cfm - 2018 - https://www.fhwa.dot.gov/policyinformation/statistics/2018/dl1c.cfm

Fatality Data

```
Fatality_data <- read.csv("https://raw.githubusercontent.com/ZanetaP02/DAta607-Final-Project/refs/heads/
Fatality_data <- Fatality_data[, c("State", "X2018", "X2020", "X2022")]
names(Fatality_data) <- c("State", "Y18", "Y20", "Y22")
```

```r
Fatality_data <- Fatality_data[-c(52), ]
head(Fatality_data)
```

```
##        State  Y18  Y20  Y22
## 1    Alabama  953  934  988
## 2     Alaska   80   64   82
## 3    Arizona 1011 1053 1302
## 4   Arkansas  520  651  643
## 5 California 3798 3980 4428
## 6   Colorado  632  622  764
```

Driving Population for Year 2022

```r
data_2022 <- read.csv("https://raw.githubusercontent.com/ZanetaP02/DAta607-Final-Project/refs/heads/mai
names(data_2022) <- c("State", "Male_Drivers_Y22", "Male_Driver_%_Y22", "Female_Driver_Y22", "Female_Dr
data_2022 <- data_2022[-c(1,2,3,4,5,57,58,59,60,61,62,63,64), ]
head(data_2022)
```

```
##          State Male_Drivers_Y22 Male_Driver_%_Y22 Female_Driver_Y22
## 6      Alabama        1,979,339             48.42         2,108,546
## 7    Alaska(2)          271,585             52.11           249,635
## 8      Arizona        2,954,582             50.53         2,893,079
## 9  Arkansas(4)        1,133,111             49.12         1,173,810
## 10  California       14,034,707             50.79        13,597,396
## 11 Colorado(6)        2,161,648             48.28         2,315,799
##    Female_Driver_%_Y22 Total_Drivers_Y22
## 6                51.58         4,087,885
## 7                47.89           521,220
## 8                49.47         5,847,661
## 9                50.88         2,306,921
## 10               49.21        27,632,103
## 11               51.72         4,477,447
##    Commercial_Motor_Vehicles_Registered_Y22 Population_Resident_Y22
## 6                                     0.75               5,074,296
## 7                                     0.79                 733,583
## 8                                     0.98               7,359,197
## 9                                     0.66               3,045,637
## 10                                    0.91              39,029,342
## 11                                    0.89               5,839,926
##    Population_Male_Y22 Population_Female_Y22 Population_Total_Y22
## 6            1,959,932             2,133,808            4,093,740
## 7              305,057               270,877              575,934
## 8            2,966,110             2,995,623            5,961,733
## 9            1,190,700             1,239,900            2,430,600
## 10          15,715,745            15,836,963           31,552,708
## 11           2,419,496             2,356,821            4,776,317
##    Drivers_Per_1K_Total_Resident_Polulation_Y22
## 6                                           806
## 7                                           711
## 8                                           795
## 9                                           757
```

```
## 10                                        708
## 11                                        767
##    Drivers_Per_1K_Age_Population_Y22
## 6                                999
## 7                                905
## 8                                981
## 9                                949
## 10                               876
## 11                               937
```

Driving Population for Year 2020

```
data_2020 <- read.csv("https://raw.githubusercontent.com/ZanetaP02/DAta607-Final-Project/refs/heads/mai
names(data_2020) <- c("State", "Male_Drivers_Y20", "Male_Driver_%_Y20", "Female_Driver_Y20", "Female_Dr
data_2020 <- data_2020[-c(1,2,3,4,5,6,58,59,60,61,62,63,64), ]
head(data_2020)
```

```
##          State Male_Drivers_Y20 Male_Driver_%_Y20 Female_Driver_Y20
## 7      Alabama        1,956,800              48.4         2,086,100
## 8       Alaska          271,451             52.32           247,421
## 9      Arizona        2,877,305             50.64         2,804,190
## 10 Arkansas(5)        1,057,402             49.09         1,096,527
## 11  California       13,730,114             50.84        13,275,188
## 12    Colorado        2,099,231             48.83         2,200,216
##    Female_Driver_%_Y20 Total_Drivers_Y20
## 7                 51.6         4,042,900
## 8                47.68           518,872
## 9                49.36         5,681,495
## 10               50.91         2,153,929
## 11               49.16        27,005,302
## 12               51.17         4,299,447
##    Commercial_Motor_Vehicles_Registered_Y20 Population_Resident_Y20
## 7                                     0.78               4,921,532
## 8                                     0.67                 731,158
## 9                                     0.96               7,421,401
## 10                                    0.75               3,030,522
## 11                                    0.88              39,368,078
## 12                                     0.8               5,807,719
##    Population_Male_Y20 Population_Female_Y20 Population_Total_Y20
## 7            1,828,314             2,005,935            3,917,625
## 8              263,418               289,009              553,317
## 9            2,753,727             3,021,251            5,662,328
## 10           1,111,417             1,219,391            2,322,502
## 11          14,580,188            15,996,656           30,465,205
## 12           2,173,275             2,384,409            4,568,613
##    Drivers_Per_1K_Total_Resident_Polulation_Y20
## 7                                           821
## 8                                           710
## 9                                           766
## 10                                          711
## 11                                          686
## 12                                          740
```

```
##    Drivers_Per_1K_Age_Population_Y20
## 7                             1,032
## 8                               938
## 9                             1,003
## 10                              927
## 11                              886
## 12                              941
```

Driving Population for Year 2018

```
data_2018 <- read.csv("https://raw.githubusercontent.com/ZanetaP02/DAta607-Final-Project/refs/heads/mai
names(data_2018) <- c("State", "Male_Drivers_Y18", "Male_Driver_%_Y18", "Female_Driver_Y18", "Female_Dr
data_2018 <- data_2018[-c(1,2,3,4,5,6,58,59,60), ]
head(data_2018)
```

```
##         State Male_Drivers_Y18 Male_Driver_%_Y18 Female_Driver_Y18
## 7     Alabama        1,939,120             48.49         2,059,937
## 8    Alaska2/          281,297             52.48           254,736
## 9     Arizona        2,645,777             50.06         2,639,193
## 10   Arkansas        1,052,671             49.07         1,092,663
## 11 California       13,755,501             50.87        13,283,899
## 12   Colorado        2,194,476              51.7         2,050,237
##    Female_Driver_%_Y18 Total_Drivers_Y18
## 7                51.51         3,999,057
## 8                47.52           536,033
## 9                49.94         5,284,970
## 10               50.93         2,145,334
## 11               49.13        27,039,400
## 12                48.3         4,244,713
##    Commercial_Motor_Vehicles_Registered_Y18 Population_Resident_Y18
## 7                                     0.77               4,887,871
## 8                                     0.69                 737,438
## 9                                     0.92               7,171,646
## 10                                    0.77               3,013,825
## 11                                    0.89              39,557,045
## 12                                     0.8               5,695,564
##    Population_Male_Y18 Population_Female_Y18 Population_Total_Y18
## 7            1,873,206             2,051,459            3,924,665
## 8              300,173               272,699              572,872
## 9            2,822,894             2,893,459            5,716,353
## 10           1,160,540             1,229,433            2,389,973
## 11          15,584,687            15,992,312           31,576,999
## 12           2,293,825             2,278,929            4,572,754
##    Drivers_Per_1K_Total_Resident_Polulation_Y18
## 7                                           818
## 8                                           727
## 9                                           737
## 10                                          712
## 11                                          684
## 12                                          745
##    Drivers_Per_1K_Age_Population_Y18
## 7                             1,019
```

```
## 8                                936
## 9                                925
## 10                               898
## 11                               856
## 12                               928
```

Merging and Cleaning Data

```
df22 <- data_2022[, c("State", "Male_Drivers_Y22", "Female_Driver_Y22", "Total_Drivers_Y22", "Population
df20 <- data_2020[, c("State", "Male_Drivers_Y20", "Female_Driver_Y20", "Total_Drivers_Y20", "Population
df18 <- data_2018[, c("State", "Male_Drivers_Y18", "Female_Driver_Y18", "Total_Drivers_Y18", "Population

pop_d <- merge(df22, df20, by = "State", all = TRUE)
pop_drive <- merge(pop_d, df18, by = "State", all = TRUE)

pop_driver <- pop_drive %>%
  pivot_longer(cols = c('Total_Drivers_Y22', 'Total_Drivers_Y20', 'Total_Drivers_Y18'), names_to = "Tota

drivers_pop <- pop_driver[, c("State", "Total_Drivers_Per_Years", "Drivers_Population")]
dp <- na.omit(drivers_pop)
head(dp)
```

```
## # A tibble: 6 x 3
##   State     Total_Drivers_Per_Years Drivers_Population
##   <chr>     <chr>                   <chr>
## 1 Alabama   Total_Drivers_Y22       4,087,885
## 2 Alabama   Total_Drivers_Y20       4,042,900
## 3 Alabama   Total_Drivers_Y18       3,999,057
## 4 Alaska    Total_Drivers_Y20       518,872
## 5 Alaska(2) Total_Drivers_Y22       521,220
## 6 Alaska2/  Total_Drivers_Y18       536,033
```

Descriptive Statistics

```
summary(dp)
```

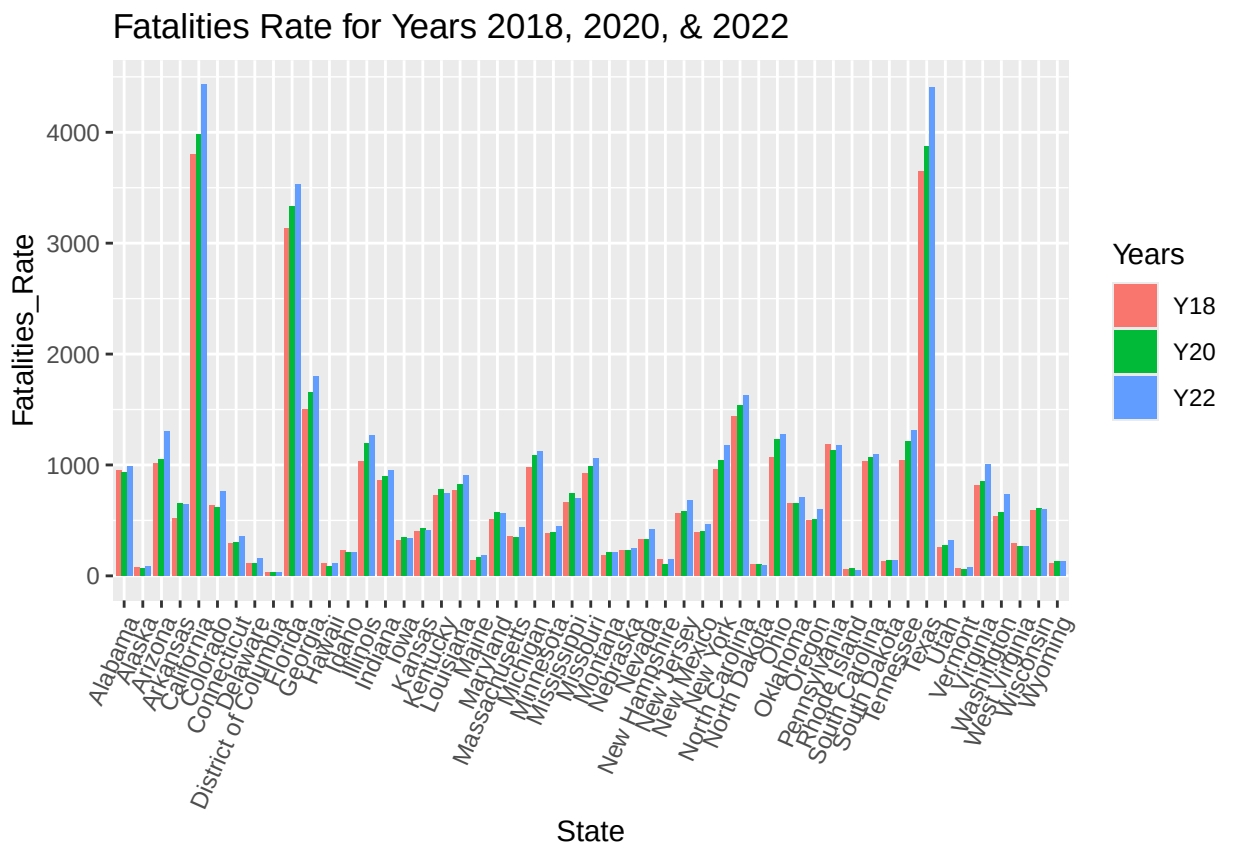
```
##     State           Total_Drivers_Per_Years Drivers_Population
##  Length:153         Length:153              Length:153
##  Class :character   Class :character        Class :character
##  Mode  :character   Mode  :character        Mode  :character
```

Graph of Fatalities

```
fd <- Fatality_data %>%
  pivot_longer(cols = c('Y18', 'Y20', 'Y22'), names_to = "Years", values_to = "Fatalities_Rate")

ggplot((data = fd), aes(x = State, y = Fatalities_Rate, fill = Years)) +
```

```
  geom_col(position = position_dodge()) +
  ggtitle("Fatalities Rate for Years 2018, 2020, & 2022") +
  theme(axis.text.x = element_text(angle = 66, hjust = 1))
```



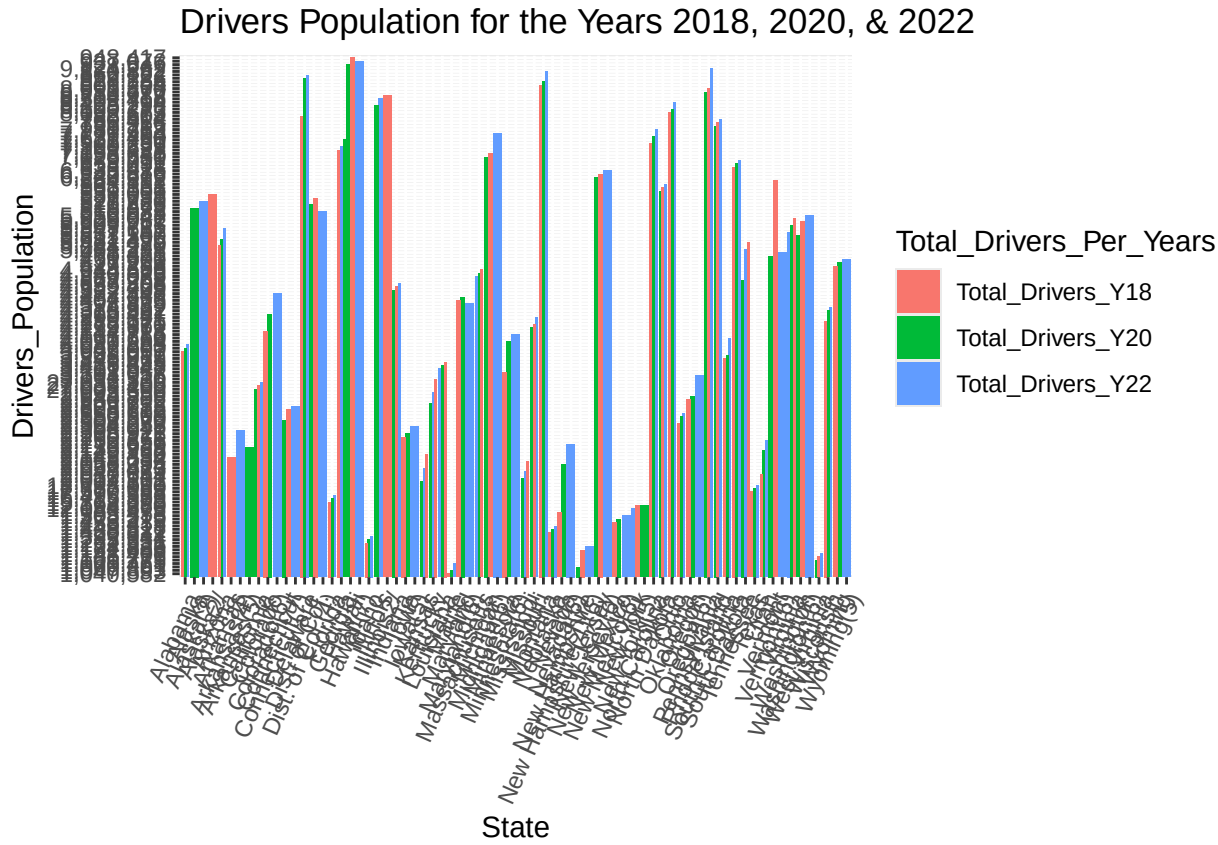
This graph provides a clear visual representation of fatalities rates across states and years, allowing for comparisons and analysis of trends.

The graph effectively visualizes traffic fatalities across US states for 2018, 2020, and 2022, highlighting significant variations between populous states like California, Texas, and Florida compared to smaller states.

Key Observations: - High Fatalities States: California, Florida, and Texas consistently have the highest fatalities rates. - Low Fatalities States: Vermont, Wyoming, and Rhode Island have the lowest rates. - Yearly Changes: While some states show slight increases or decreases, the overall trend varies by state.

Graph of Drivers Population

```
ggplot((data = dp), aes(x = State, y = Drivers_Population, fill = Total_Drivers_Per_Years)) +
  geom_col(position = position_dodge()) +
  ggtitle("Drivers Population for the Years 2018, 2020, & 2022") +
  theme(axis.text.x = element_text(angle = 66, hjust = 1))
```


Drivers Population for the Years 2018, 2020, & 2022
Drivers_Population
State
Total_Drivers_Per_Years
Total_Drivers_Y18
Total_Drivers_Y20
Total_Drivers_Y22