Exploratory Data Analysis

Brianna Sandoval

September 20,2021

Selected data set: Police Shootings

Introduction

The data set that I chose to analyze for my exploratory data project was the Police Shooting Data set. This data set tells us many characteristics(variables) of those that were killed by the police in 2015 but I will be looking into the race and gender of those affected.

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(ggplot2)
library(readxl)

police_shootings  <-
read_excel("C:/Users/Sbria/OneDrive/Desktop/math_130/data/fatal-police-
shootings-data.xlsx", sheet=1, col_names=TRUE)
```

Univariate Analysis of Variables

Race

```
police_shootings$race[police_shootings$race %in% c('A')] <- "Asian"
police_shootings$race[police_shootings$race %in% c('B')] <- "Black"
police_shootings$race[police_shootings$race %in% c('H')] <- "Hispanic"
police_shootings$race[police_shootings$race %in% c('W')] <- "White"
police_shootings$race[police_shootings$race %in% c('N')] <- "Native American"
police_shootings$race[police_shootings$race %in% c('O')] <- "Other"


table(police_shootings$race)

##
##           Asian           Black         Hispanic Native American
```

```
Other
61   927   659   62
37
White
1825
```

With the information provided to us from the table above we can see the race of those who were attacked by the police along with the amount of people from that race.The first thing that we can see is that the majority of the victims are White, Hispanic, and Black The table above shows,61 were Asian, 927 were black, 659 were Hispanic, 62 were Native American, 1825 were white, and 37 were classified as other, and 369 were unclassified

```
ggplot(police_shootings,aes(x=race,fill=race))+geom_bar()
```
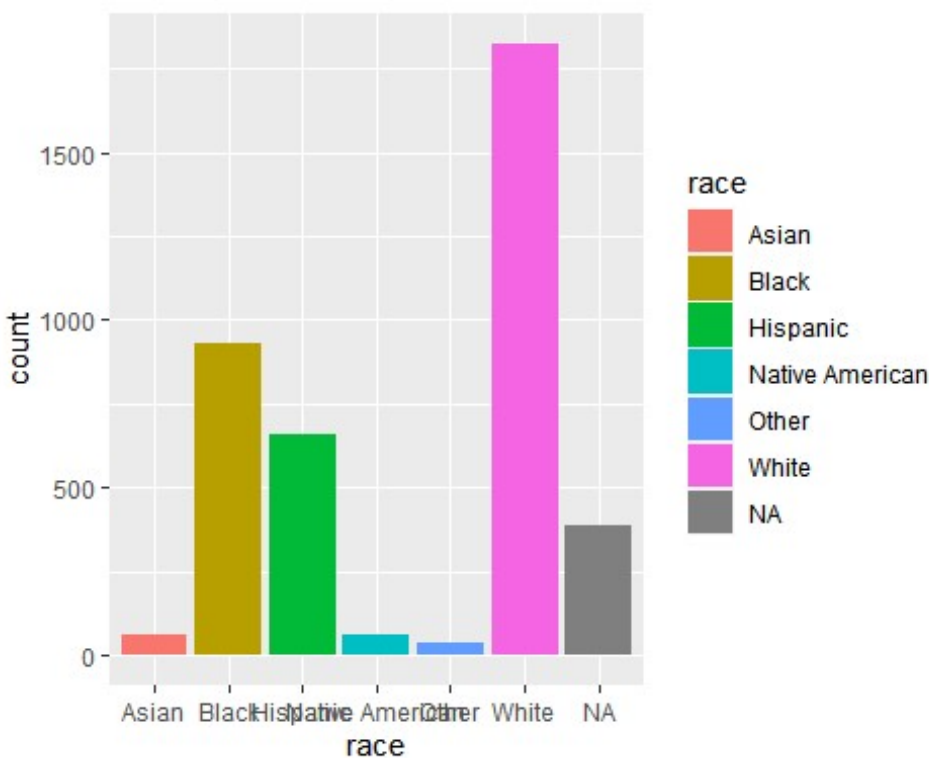

This graph gives a visual representation of how many people within the races were victims to police shootings and as expected based on the chart above we can see that the tallest bar are Black, Hispanic, and White.

Gender

```
police_shootings$gender[police_shootings$gender %in% c('M')] <- "Male"
police_shootings$gender[police_shootings$gender %in% c('F')] <- "Female"

table(police_shootings$gender)

##
## Female    Male
##    180    3777
```
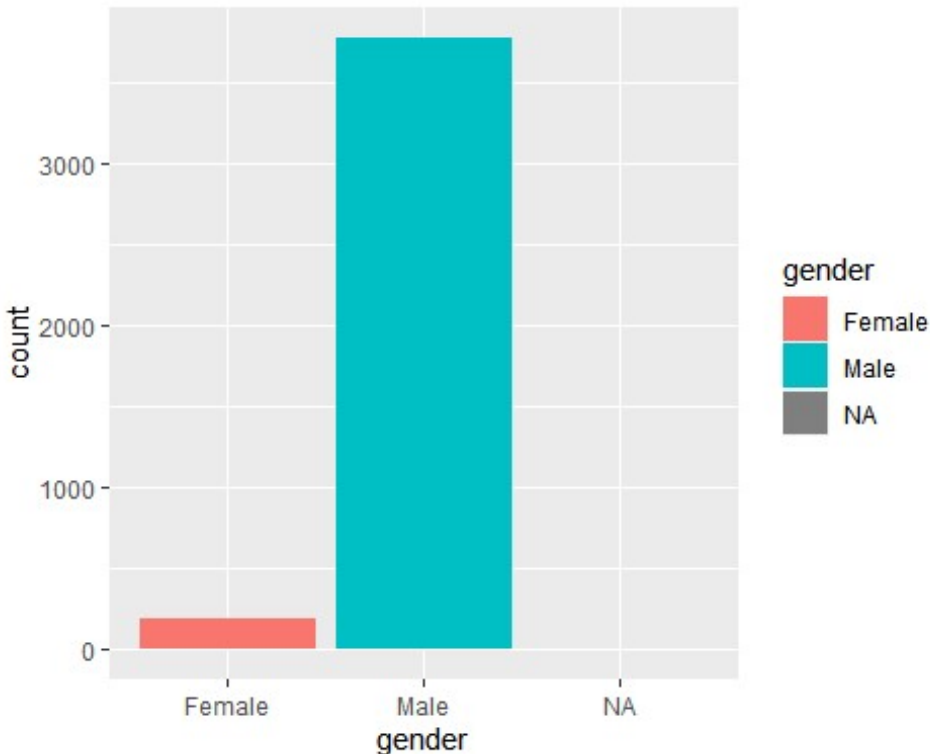
With the information provided to us from the table above we can see that the majority of the victims to police shootings were male with a total of 3777.

```
ggplot(police_shootings,aes(x=gender,fill=gender))+geom_bar()
```


As expected based on the information provided to us from the chart above we can see that the male bar is extremely higher compared to the female bar.

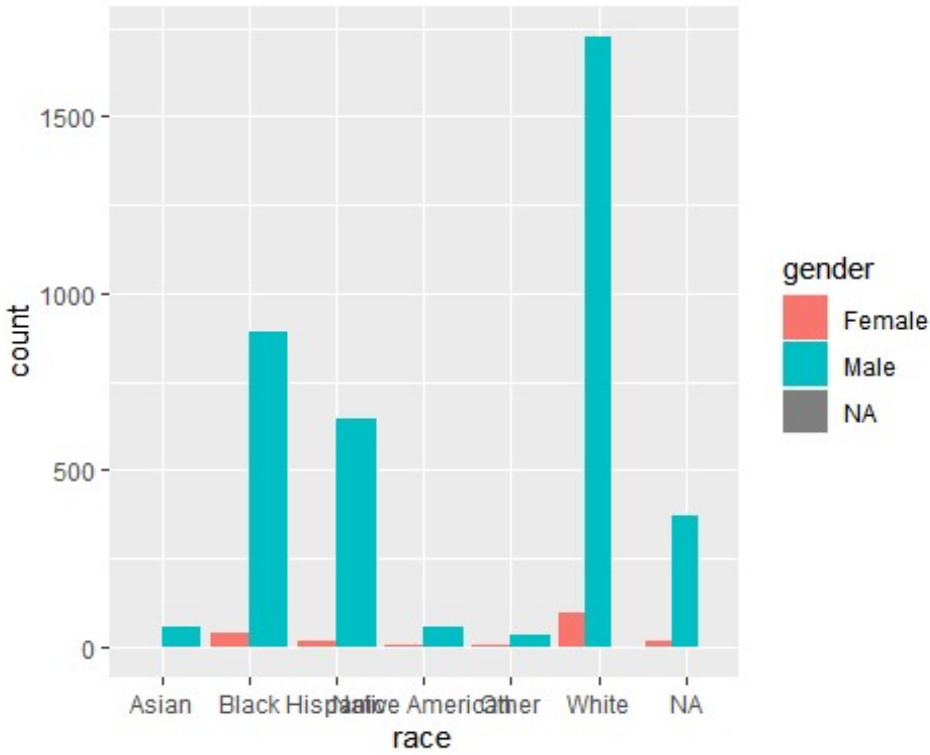Bivariate Analysis

```
table(police_shootings$race, police_shootings$gender)
```

```
##
##                    Female Male
##    Asian                2   59
##    Black               39  888
##    Hispanic            17  642
##    Native American      5   57
##    Other                3   34
##    White               99 1725
```

In this table we can see how many victims within each race or male or female.

```
ggplot(police_shootings, aes(x=race, fill=gender)) + geom_bar(position =
"dodge")
```

This graph gives us a visual representation of how many males and females within each race were affected by the police shootings. This graph helps us see that a majority of those affected were white,Hispanic and black males.