Exploratory Data Analysis

Levi Medina

02/21/2022

Introduction:

The Police_shootings data consists of information of fatal police shootings within 2015; such as names, dates, locations, age, gender, etc. I chose the variables gender and race in regards to the victims of this police violence.

```
library(readxl)
police_shootings <- read_excel("fatal-police-shootings-data.xlsx", sheet=1, col_names=TRUE)
police_shootings[1:10,1:5]
```

```
## # A tibble: 10 x 5
##       id name                date                manner_of_death  armed
##    <dbl> <chr>               <dttm>              <chr>            <chr>
##  1     3 Tim Elliot          2015-01-02 00:00:00 shot             gun
##  2     4 Lewis Lee Lembke    2015-01-02 00:00:00 shot             gun
##  3     5 John Paul Quintero  2015-01-03 00:00:00 shot and Tasered unarmed
##  4     8 Matthew Hoffman     2015-01-04 00:00:00 shot             toy weapon
##  5     9 Michael Rodriguez   2015-01-04 00:00:00 shot             nail gun
##  6    11 Kenneth Joe Brown   2015-01-04 00:00:00 shot             gun
##  7    13 Kenneth Arnold Buck 2015-01-05 00:00:00 shot             gun
##  8    15 Brock Nichols       2015-01-06 00:00:00 shot             gun
##  9    16 Autumn Steele       2015-01-06 00:00:00 shot             unarmed
## 10    17 Leslie Sapp III     2015-01-06 00:00:00 shot             toy weapon
```

Here we are able to read in the data and get an overall look of the data set that will be used. I chose to portray this as a chart so that i can ensure all my data was read in correctly.

Univariate Exploration:

Gender:

```
table(police_shootings$gender)
```

```
##
##    F    M
##  180 3777
```

```
library(ggplot2)
library (RColorBrewer)
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
```

```
police_shootings %>% select(gender) %>% na.omit() %>% ggplot(aes(x=gender, fill=gender)) + geom_bar() +
  ggtitle("Amount of Police Shootings Per Gender") +
  ylab("Amount of Shootings") + xlab("Gender")
```

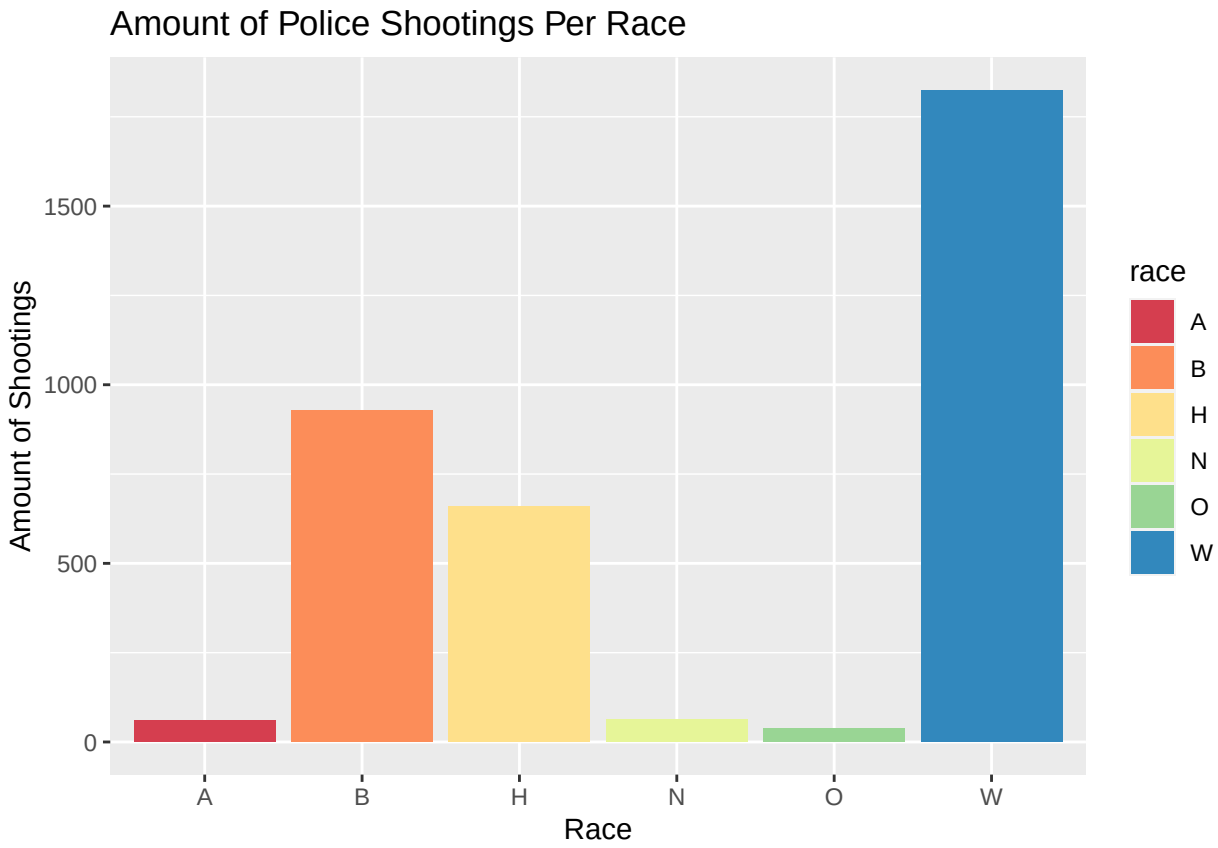
Through this bar chart we are able to directly compare both men and women in regards to the amount of shootings each faced. We are able to see the likelihood of being killed by the police based on gender.

Race:

```
table(police_shootings$race)
```

```
##
##    A    B    H    N    O    W
##   61  927  659   62   37 1825
```

```
police_shootings %>% select(race) %>% na.omit() %>% ggplot(aes(x=race, fill=race)) + geom_bar() + scale_
  ggtitle("Amount of Police Shootings Per Race") +
  ylab("Amount of Shootings") + xlab("Race")
```



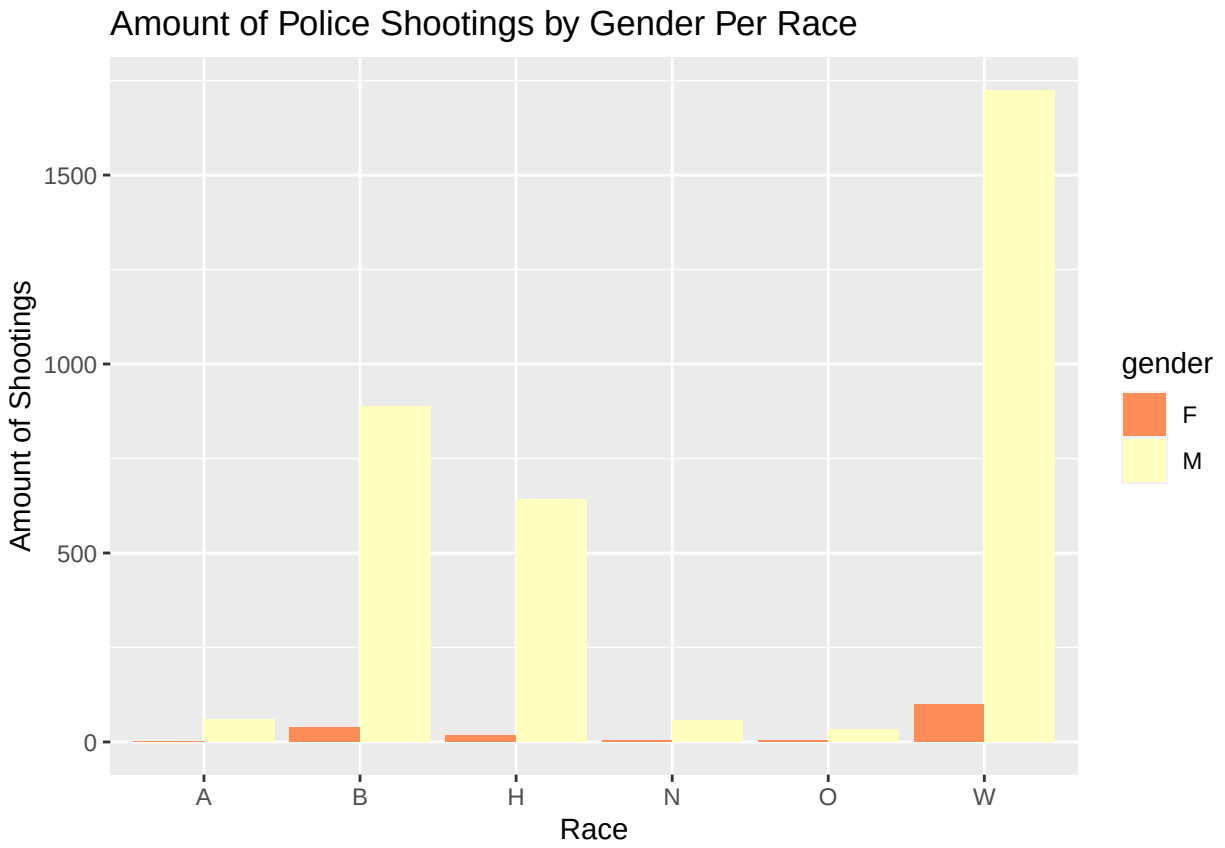
We can also use a bar chart to compare fatal police shootings based on race, similar to the way we expressed gender into R studio.In this case, A=Asian B=black, H=Hispanic N=native american O=other W=white

Bivariate Exploration

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

```
##
##        A    B    H    N    O    W
##   F    2   39   17    5    3   99
##   M   59  888  642   57   34 1725
```

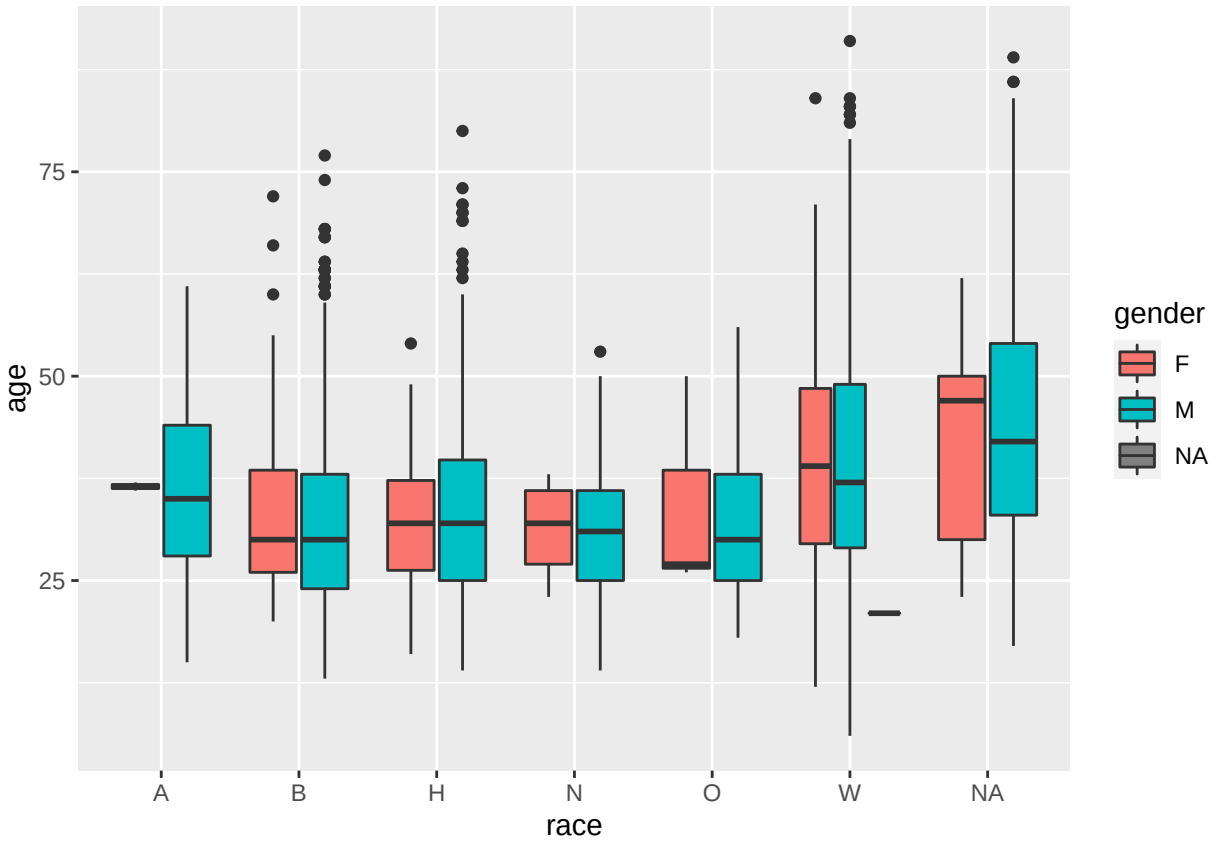
```
police_shootings %>% select(race, gender) %>% na.omit() %>% ggplot(aes(x=race, fill=gender)) + geom_bar
  ggtitle("Amount of Police Shootings by Gender Per Race") +
  ylab("Amount of Shootings") + xlab("Race")
```



Now that we have our two variables expressed, we can compare them to each other and see how race and gender play a part in the deaths in 2015 at the hand of the police. Based on the bar chart, we can see that while women were killed significantly less, this statistic was affected by race also.

```
ggplot(police_shootings, aes(y=age, x=race, fill=gender)) + geom_boxplot()
```

```
## Warning: Removed 152 rows containing non-finite values (stat_boxplot).
```



Another way we can look at data is through this box-plot; this allows us to include another variable such as age shown above. This gives us a cross of information for three variables if needed. Like this example of age, race, and gender.

Conclusion:

Overall while viewing our data, we can conclude that there is a clear relation to the number of deaths from police shootings to both race and gender. Not only is the likelihood of a fatal shooting higher for men, but also appeared to be high for those from a white or black background. Both the rates for men and women rose consistently with their listed race. By exploring the data as a chart, histogram, and box-plot; i was able to compare these variables based on each of their factors. I found the information to be most efficiently expressed as a box-plot as it gave a more direct comparison of who was most likely to be a victim to police shootings.