Project

2022-09-21

[Introduction]

This report will analyze killings made by police. It will compare race of the victim, the date of their murder, and the state it took place in. This data was collected by the Washington Post starting in 2015 through culling local news report, law enforcement websites, and social media. Using this data I would like to see if this type of violence is a background constant, or if there are specific times that there are spikes. The

```
library(ggplot2)
library(forcats)
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
```

```
FPS <- read.csv("../data/fatal-police-shootings-data - Copy.csv")
dim(FPS)
```

```
## [1] 7768   18
```

[Univariate Exploration]

Because of the low amount of some data points across different races, the Asian, Native American, and other categories were combine. These categories were also renamed for better readability. With the below code it can be seen the the majority of victims to fatal police shootings were either white, Black, or Hispanic. There is also a significant portion of fatal police shootings where the race of the victim was no reported, which could be for various reasons, but because of the amount of data it will be left in for the analysis.

```
FPS$race[FPS$race=="A"]<-"O"
FPS$race[FPS$race=="N"]<-"O"
FPS$race <- factor(FPS$race, levels=c("W","B","","H","O"), labels=c("White","Black","Unknown","Hispanic","Other"))
ggplot(FPS, aes(x=race)) + geom_bar() + ggtitle("Number of Victims of Police Fatal Shootings by Race") + xlab("Race") + ylab("")
```


Number of Victims of Police Fatal Shootings by Race
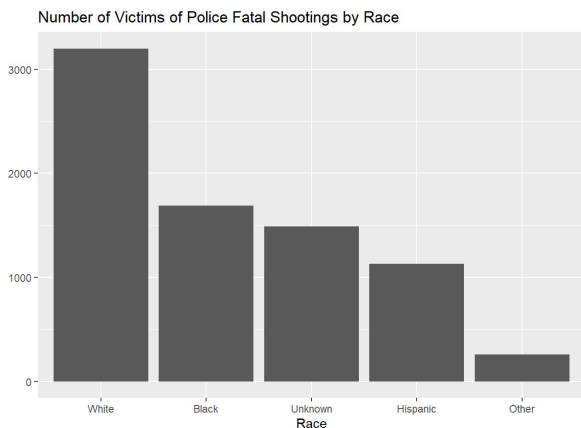
Next is a look at the Gender of the victims. Here it can be seen that the vast majority of fatal police shootings had male victims, however there is still a large amount of female victims as well. There were only 21 cases where the victim's gender was not reported. To simplify the data visualization these small amount of points will be ignored in the full analysis.

```
FPS$gender[FPS$gender=="F"]<-"Female"
FPS$gender[FPS$gender=="M"]<-"Male"
FPS$gender[FPS$gender==""]<- NA
table(FPS$gender, useNA="always")
```

```
##
## Female   Male   <NA>
##    350   7397     21
```

```
ggplot(FPS, aes(x=gender)) + geom_bar() + ggtitle("Number of Victims of Police Fatal Shootings by Gender") + xlab("Gender")
+ ylab("")
```


Number of Victims of Police Fatal Shootings by Gender
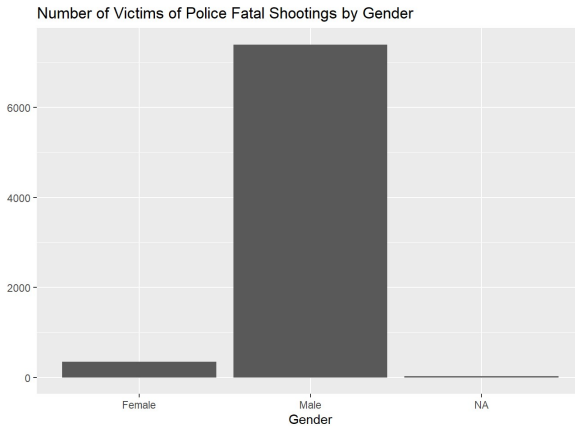
[Bivariate Exploration]

To find the amount of shootings over time separated by race and gender there are 2 histogram plots of shootings over time, One of male victims, and one of female victims. Here we can see that on average the number of shootings has stayed relatively constant. It also seems that there has been a large decrease in reporting the race of the victim, which can be seen in the spike in the "Unknown" race slice. There was also a spike in Hispanic women victims in

```
FPS <- filter(FPS, gender=="Male" | gender=="Female")
ggplot(FPS, aes(x=days, fill=race)) + geom_density(alpha=.25) + facet_wrap(~gender, ncol = 1) + scale_fill_brewer(palette="S
et3") + ggtitle("Number of Victims Per Day Since January First, 2015") + xlab("Days Since January First 2015") + ylab("")
```


Number of Victims Per Day Since January First, 2015
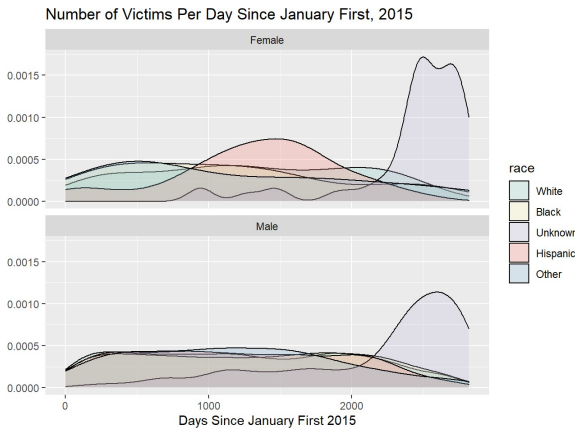
[Conclusions]

This analysis has shown that there is a pretty large gap in fatal police shootings by gender. The racial gap is a little harder to parse without larger population data. It also seems that the general rate of fatal police shootings have increased since 2015, however there have been a recent increase in not reporting the race of the victim of the murder.