Police Brutality in 2020 and Beyond

Zaki Ahmad

12/6/2020

Problem Statement

It is widely known that police brutality is an epidemic that plagues the United States, and this has become even more apparent in 2020, with the death of George Floyd and the propelling of Black Lives Matter into the media spotlight. What I wanted to do for this project was to analyze the data of police killings across the US, and precict how many people of each race will be killed by the police in 2020 and beyond.

Data Sets

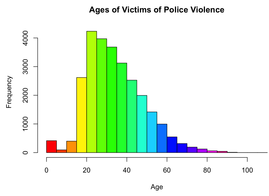
Data Cleaning

```
fatal_enc<-read.csv("~/Desktop/EAS 345/Term Project/fatal_encounters_dot_org.csv",stringsAsFactors=FALSE)
fatal_enc$URL<-NULL #not needed
fatal_enc[1]=NULL#Unique1D not needed
fatal_enc[14:15]=NULL #latitude and longitude not needed
fatal_enc[13]=NULL

#replace empty age with 0
suppressPackageStartupMessages(library(dplyr))
fatal_enc$Subject.s.age<-ifelse(fatal_enc$Subject.s.age=='',0,fatal_enc$Subject.s.age)

#replace unspecified race with imputated race, if imputation probability is >0.5
fatal_enc$Subject.s.race<-ifelse(fatal_enc$Subject.s.race=='Race unspecified' & fatal_enc$Imputation.probabilit
y > 0.5,fatal_enc$Subject.s.race.with.imputations,fatal_enc$Subject.s.race)

fatal_enc<-fatal_enc[!(fatal_enc$Subject.s.race=='Race unspecified'),] #delete rows with race still unspecifie
d, cant use that data
fatal_enc<-fatal_enc[!(is.na(fatal_enc$Subject.s.race)),]#delete any NA values, race data should be clean now
```

In the code above, I opened the csv file containing my data, and proceeded to clean it. In the column containing the different races of the victims, there were many values that were 'Race Unspecified'. I could not use these, so I removed them. However, there was a column next to this one that was the race calculated by the creator of the data set, so if the imputation probability was greater than 0.5, I chose to replace 'Race Unspecified' with the race calculated with imputations. If the race was still unspecified, then I had to delete the row, because the data is useless for this project.

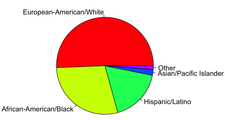
Exploratory Data Analysis

```
age<-as.numeric(fatal_enc$Subject.s.age)
hist(age, main="Ages of Victims of Police Violence",  xlab= "Age",col=rainbow(19))
```



I plotted the number of victims of police brutality by age with this histogram. An age of 0 represents a person who's age was not reported. Most of the people killed by the police from 2000 to 2020 were between the age of 20 and 40.

```
race<-fatal_enc$Subject.s.race
numOfWhite=sum(race=='European-American/White')
numOfBlack=sum(race=='African-American/Black')
numOfAsian=sum(race=='Asian/Pacific Islander')
numOfHisp=sum(race=='Hispanic/Latino')
numOfOther=sum(race!='European-American/White' & race!='African-American/Black' & race != 'Asian/Pacific Islande
r' & race != 'Hispanic/Latino')
raceTotals<-c(numOfWhite,numOfBlack,numOfHisp,numOfAsian,numOfOther)
names(raceTotals)<-c('European-American/White','African-American/Black','Hispanic/Latino','Asian/Pacific Islande
r','Other')
pie(raceTotals,col=rainbow(5))
```



This pie chart takes all of the victims from 2000 to 2020 and counts the number of each race. European-American/White Americans were killed by the police the most, which makes sense, because according to the US Census, White people make up over 60% of the population. However, the number of African-American/Black Americans killed by the police is concerning.

How Many Victims of Police Brutality Were Black?

```
numOfBlack
```

```
## [1] 7678
```

```
numOfBlack/(sum(raceTotals))*100
```

```
## [1] 28.64498
```

The above code prints out the number of Black people killed by the police in the past 20 years, and then divides that number by the number of total people killed by the police to get the percentage of the victims that were Black. Almost a third of the victims of police brutality were Black, which is alarming considering that Black people make up only 13% of the population

```
numOfHisp
```

```
## [1] 4697
```

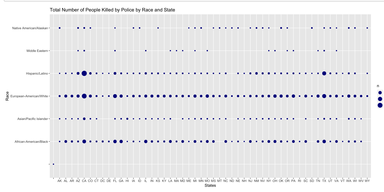
```
numOfHisp/(sum(raceTotals))*100
```

```
## [1] 17.5235
```

Hispanic victims of police brutality were only 17.5% of the total victims. This is less than the number of Black victims, despite there being more Hispanic citizens than there are Black citizens.

Total Number of People Killed by The Police by Race and by State

```
suppressPackageStartupMessages(library(ggplot2))
ggplot(fatal_enc, mapping = aes(Location.of.death..state.,Subject.s.race)) +geom_count(color="darkblue") +ggtitle
("Total Number of People Killed by Police by Race and State")+
  xlab("States") + ylab("Race")
```

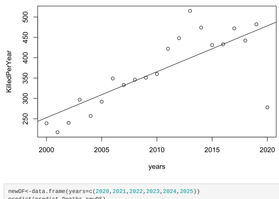


Predicting the Number of Black People that will be killed by the Police in 2020 and Beyond

```
KilledPerYear<-c()
y=2000
while(y<=2020){
  KilledPerYear<-append(KilledPerYear,sum(fatal_enc$Subject.s.race=='African-American/Black' & fatal_enc$Date..Ye
ar==y))
  y=y+1
}

years<-c(2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,2016,2017,2018,2019,2020
)
blackDF<-data.frame(KilledPerYear,years)

ggplot(blackDF,mapping=aes(years,KilledPerYear)) +geom_point(color='darkblue') +ggtitle("Number of Black People K
illed by the Police Per Year")+
  xlab("Years") + ylab("Number of Black People Killed")
```

The number of Black people being killed by the police seems to be increasing. The reason the number for 2020 is much lower than the previous years is because the data was recorded in the middle of the year. After creating this plot, I wanted to use the same data and create a linear model and finally predict the number of Black people that will be killed by the Police in the future.

```
plot(KilledPerYear ~ years, data=blackDF)
predict.Deaths = lm(KilledPerYear ~ years, data=blackDF)
abline(predict.Deaths)
```




```
newDF<-data.frame(years=c(2020,2021,2022,2023,2024,2025))
predict(predict.Deaths,newDF)
```

```
##        1        2        3        4        5        6
## 478.4502 489.7333 501.0165 512.2996 523.5827 534.8658
```

The above code creates a liner model of the number of Black People killed by the police in the last 20 years, and then predicts the number of Black people that will be killed in the future. I have taken this data and created a table below.

Predicted Number of Black People Killed by the Police

YEAR	NUMBER OF PEOPLE
2020	478
2021	489
2022	501
2023	512
2024	523
2025	534

As I had predicted before, the number of Black people that will die by the hands of the police will only increase. Unless something is done, Black people will be killed at a disproportional rate. Actions must be taken in order to protect our fellow citizens from this cruelty.

Data Product

Along with this HTML, the data I used and an R Shiny app are contained in the zip file. To run the R Shiny App, follow these steps:

- Open the app.R with Rstudio.
- Click Run App in on the top right corner
- A new window will pop up with a linear model of all the people killed by the police in the past 20 years. Click on the dropdown box to see the linear model for each race.