NYCPD Report

B. Patel

4/18/2022'

Introduction

We will be looking at NYPD Shooting Incident Data (Historic). The site used is https://catalog.data.gov/ dataset. From this data set, I hope to understand who is targeted, where they are targeted, and how often

Intial Preparation

We will use tidyverse and lubridate libraries.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.6      v dplyr   1.0.8
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

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

Import Data

At this stage, we need to import the actual data so that we can analyze in the latter steps. We first store the url and then store the actual file. The final url for the nycpd report will be the combination od the url and the file.

```
url_in <- "https://data.cityofnewyork.us/api/views/833y-fsy8/"
file_name <- c("rows.csv")
url <- str_c(url_in, file_name)
nycpd <- read.csv(url)
```

Understanding the Data

Since, the data is imported. It is great to check the summary in order to examine what portions of the data
set is relevant for the report and analysis.This step can be skipped.

```
summary(nycpd)
```

```
##    INCIDENT_KEY         OCCUR_DATE          OCCUR_TIME           BORO
##  Min.   :  9953245   Length:23585       Length:23585       Length:23585
##  1st Qu.: 55322804   Class :character   Class :character   Class :character
##  Median : 83435362   Mode  :character   Mode  :character   Mode  :character
##  Mean   :102280741
##  3rd Qu.:150911774
##  Max.   :230611229
##
##     PRECINCT      JURISDICTION_CODE LOCATION_DESC      STATISTICAL_MURDER_FLAG
##  Min.   :  1.00   Min.   :0.000     Length:23585       Length:23585
##  1st Qu.: 44.00   1st Qu.:0.000     Class :character   Class :character
##  Median : 69.00   Median :0.000     Mode  :character   Mode  :character
##  Mean   : 66.21   Mean   :0.333
##  3rd Qu.: 81.00   3rd Qu.:0.000
##  Max.   :123.00   Max.   :2.000
##                   NA's   :2
##  PERP_AGE_GROUP      PERP_SEX           PERP_RACE          VIC_AGE_GROUP
##  Length:23585       Length:23585       Length:23585       Length:23585
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##     VIC_SEX            VIC_RACE           X_COORD_CD         Y_COORD_CD
##  Length:23585       Length:23585       Min.   : 914928    Min.   :125757
##  Class :character   Class :character   1st Qu.: 999925    1st Qu.:182539
##  Mode  :character   Mode  :character   Median :1007654    Median :193470
##                                        Mean   :1009379    Mean   :207300
##                                        3rd Qu.:1016782    3rd Qu.:239163
##                                        Max.   :1066815    Max.   :271128
##
##     Latitude        Longitude         Lon_Lat
##  Min.   :40.51   Min.   :-74.25    Length:23585
##  1st Qu.:40.67   1st Qu.:-73.94    Class :character
##  Median :40.70   Median :-73.92    Mode  :character
##  Mean   :40.74   Mean   :-73.91
##  3rd Qu.:40.82   3rd Qu.:-73.88
##  Max.   :40.91   Max.   :-73.70
##
```

Visualization and Analysis

Determing the safest district in New York City

I think the first thing I want to explore is which is the safest place and the most dangerous locations of New York City. Ideally, the locations that have more incidents should be avoided.

To check where the greatest number of incidents took place, a map needs to be generated with the latitude and longitude data.

```
ggplot(data = nycpd, aes(x = Longitude, y = Latitude)) + geom_point()
```
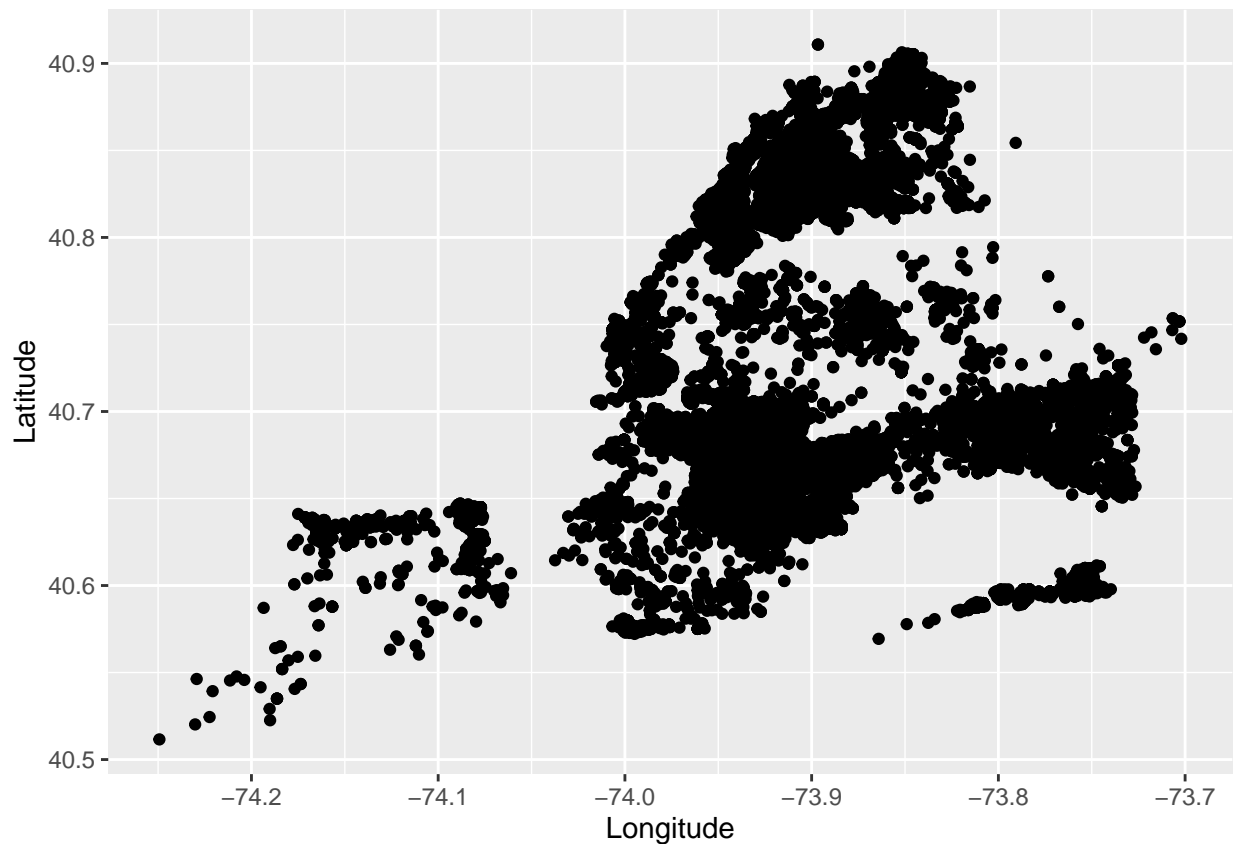

After seeing the map, there is not a clear indicator of which location has more or less incidents. Therefore, a table can be generated to determine the most safe and unsafe districts of New York City. To do this, the frequency of incidents per district needs to be calculated. Then, the frquencies can be ordered so that the first row of the table is the safest location and the last row of the table is the most unsafe location.

```
district <- as.data.frame(table(nycpd["BORO"]))
district[order(district$Freq),]
```

```
##              BORO Freq
## 5 STATEN ISLAND  696
## 3     MANHATTAN 2922
## 4        QUEENS 3532
## 1         BRONX 6701
## 2      BROOKLYN 9734
```

It appears that majority of incidents occur in Brooklyn, while the least amount of incidents occur in Staten Island.

Determing the profile of an average victim of New York City crimes

Race is one aspect of victim profiling. We can begin by generating a bar graph according to victim's race.

```r
victim_race <- table(nycpd["VIC_RACE"])
par(las=2)
rownames(victim_race) <- c('Indian/Native','Asian/Islander','Black', 'Black Hispanic',
                           'Unknown', 'White', 'White Hispanic')
barplot(victim_race, cex.names=.50, horiz= TRUE)
```
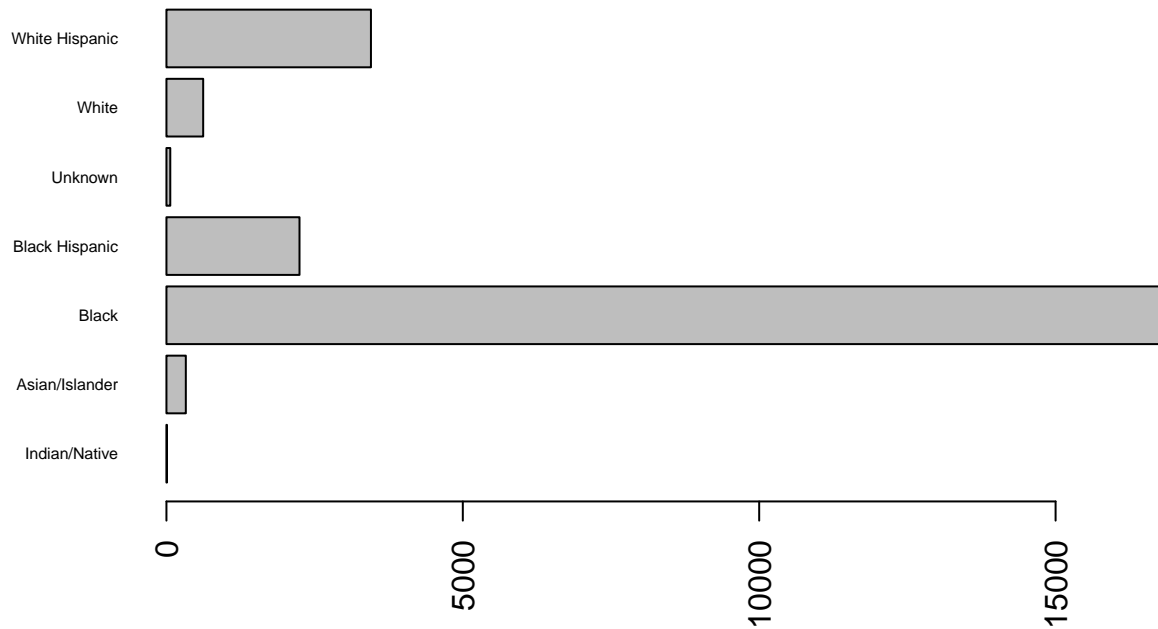

According to the graph, the average victim is more likely to be Black.

Gender is another aspect of victim profiling. We can examine the gender next.

```r
victim_sex <- table(nycpd["VIC_SEX"])
par(las=2)
barplot(victim_sex, cex.names=.50, horiz= TRUE)
```
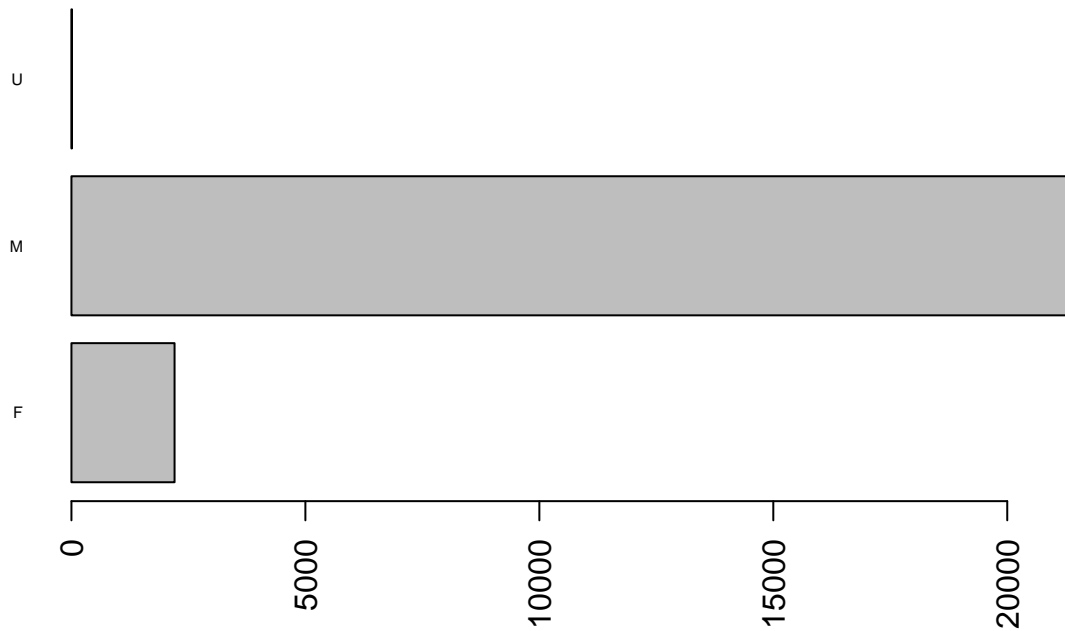
As displayed in the graph, the average victim is more likely to identify as a male.

Lastly, a victim's profile includes age.

```r
victim_age <- table(nycpd["VIC_AGE_GROUP"])
par(las=2)
barplot(victim_age, cex.names=.50, horiz= TRUE)
```

The graph indicates that the most common age of a victim is between 25 and 44.

Bias

There are quite a lot of biases possible that may have affected this data. For example, the number of incidents can be much higher than assumed by this data set. Quite often, some incidents do not get recorded because victim's are not found. Another category that needs to be analyzed is the household incomes of the victims. It also must be noted that Staten Island is the smallest district while Brooklyn is the largest district. To consider the crime rates for each district, a percent must be taken for evaluating the district's safety level.

Conclusion

Overall, I learned that when visiting New York City, Staten Island is a safer visiting spot than Brooklyn. Also, the average victim to die in a New York City crime incident is a 25-44 year old Black man.