

NYPD Shooting Incident - Historic Data Analysis

This report analyzes the data relating to shooting incidents which have occurred in NYC boroughs between 2006 to 2020. This data is extracted and reviewed every quarter by the Office of Management Analysis and Planning, and will be used in an overall analysis of shootings in each borough.

Importing libraries

First, the libraries that are needed for this analysis are loaded.

```
library(lubridate)
library(tidyverse)
library(knitr)
library(incidence)
```

Importing Data

Data is read in from the .csv file from the DATA.GOV website

```
url <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv"
nyshootings <- read.csv(url)
colnames(nyshootings)
```

```
## [1] "INCIDENT_KEY"      "OCCUR_DATE"
## [3] "OCCUR_TIME"        "BORO"
## [5] "PRECINCT"          "JURISDICTION_CODE"
## [7] "LOCATION_DESC"       "STATISTICAL_MURDER_FLAG"
## [9] "PERP_AGE_GROUP"    "PERP_SEX"
## [11] "PERP_RACE"          "VIC_AGE_GROUP"
## [13] "VIC_SEX"            "VIC_RACE"
## [15] "X_COORD_CD"         "Y_COORD_CD"
## [17] "Latitude"           "Longitude"
## [19] "Lon_Lat"
```

Data Wrangling

Now that our data has been read in, it needs to be cleaned up and any data that are not pertinent to this analysis can be removed. Because I'm only interested in shootings in the boroughs, I am removing data such as Longitude and Latitude.

```
nyshootings <- nyshootings %>% select(-c(Latitude, Longitude, Lon_Lat,
                                          INCIDENT_KEY, JURISDICTION_CODE))
nyshootings <- mutate(nyshootings, OCCUR_DATE = mdy(OCCUR_DATE))
summary(nyshootings)
```

```
## OCCUR_DATE          OCCUR_TIME          BORO          PRECINCT
## Min. :2006-01-01    Length:23568    Length:23568    Min. : 1.00
## 1st Qu.:2008-12-30    Class :character    Class :character    1st Qu.: 44.00
## Median :2012-02-26    Mode :character    Mode :character    Median : 69.00
## Mean :2012-10-03                                     Mean : 66.21
## 3rd Qu.:2016-02-28                                     3rd Qu.: 81.00
## Max. :2020-12-31                                     Max. :123.00
## LOCATION_DESC      STATISTICAL_MURDER_FLAG PERP_AGE_GROUP
## Length:23568        Length:23568        Length:23568
## Class :character    Class :character    Class :character
## Mode :character     Mode :character     Mode :character
##
##
##
## PERP_SEX           PERP_RACE           VIC_AGE_GROUP       VIC_SEX
## Length:23568        Length:23568        Length:23568        Length:23568
## Class :character    Class :character    Class :character    Class :character
## Mode :character     Mode :character     Mode :character     Mode :character
##
##
##
## VIC_RACE           X_COORD_CD          Y_COORD_CD
## Length:23568        Length:23568        Length:23568
## Class :character    Class :character    Class :character
## Mode :character     Mode :character     Mode :character
##
##
##
```

Data Analysis

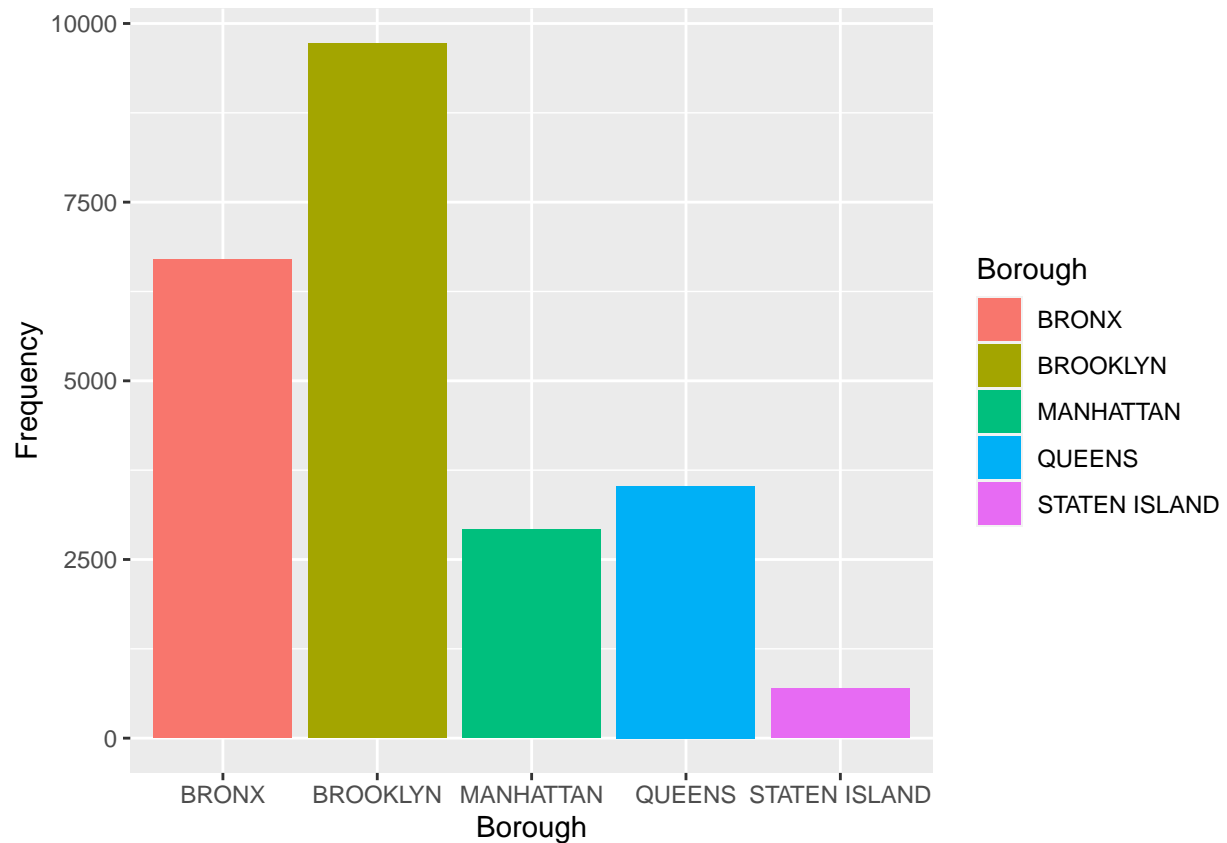
After cleaning up the data we can analyse the data. The first analysis will be the total number of shootings per borough.

```
borough_totals <- table(nyshootings$BORO)
borough_totals <- as.data.frame(borough_totals) %>%
  rename(Borough = Var1, Frequency = Freq)
borough_totals$Percent <- round((borough_totals$Frequency / sum(borough_totals$Frequency)*100),2)
kable(borough_totals)
```

Borough	Frequency	Percent
BRONX	6700	28.43
BROOKLYN	9722	41.25
MANHATTAN	2921	12.39
QUEENS	3527	14.97
STATEN ISLAND	698	2.96

This summary can be visualized in a histogram

```
ggplot(borough_totals, aes(x=Borough, y=Frequency, fill=Borough)) + geom_bar(stat="identity")
```



The histogram shows that the frequency of shootings in Brooklyn is the highest, with Staten Island having the lowest incidence. But this is the total shootings from 2006 to 2020, what has this data looked over time? We can analyze the data further by taking shooting incidence from Brooklyn and Staten Island and plotting against the year.

Brooklyn

```
BR <-subset(nyshootings, BORO=='BROOKLYN', select=c(BORO, OCCUR_DATE))
n <- 5

BR$YEAR <- substr(BR$OCCUR_DATE, nchar(BR$OCCUR_DATE) - n + 1, nchar(BR$OCCUR_DATE))

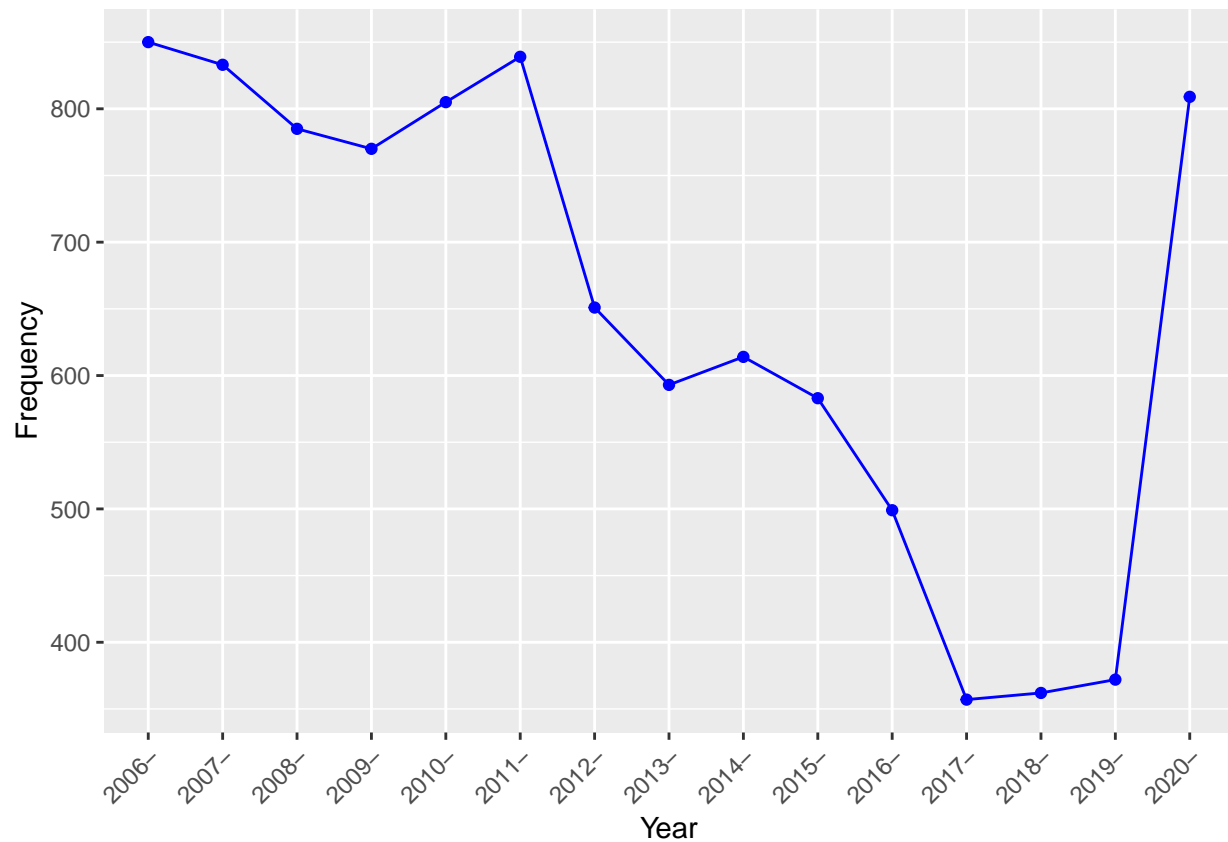
BR <- subset(BR, select = -c(OCCUR_DATE))

BROOKLYN <- table(BR$YEAR)
BROOKLYN <- as.data.frame(BROOKLYN)

names(BROOKLYN)[names(BROOKLYN) == "Var1"] <- "Year"
names(BROOKLYN)[names(BROOKLYN) == "Freq"] <- "Frequency"

ggplot(data=BROOKLYN, aes(x=Year, y=Frequency, group=1)) +
  geom_line(colour = "blue")+
```

```
geom_point(colour = "blue")+
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Staten Island

```
SI <- subset(nyshootings, BORO=='STATEN ISLAND', select=c(BORO, OCCUR_DATE))
n <- 5

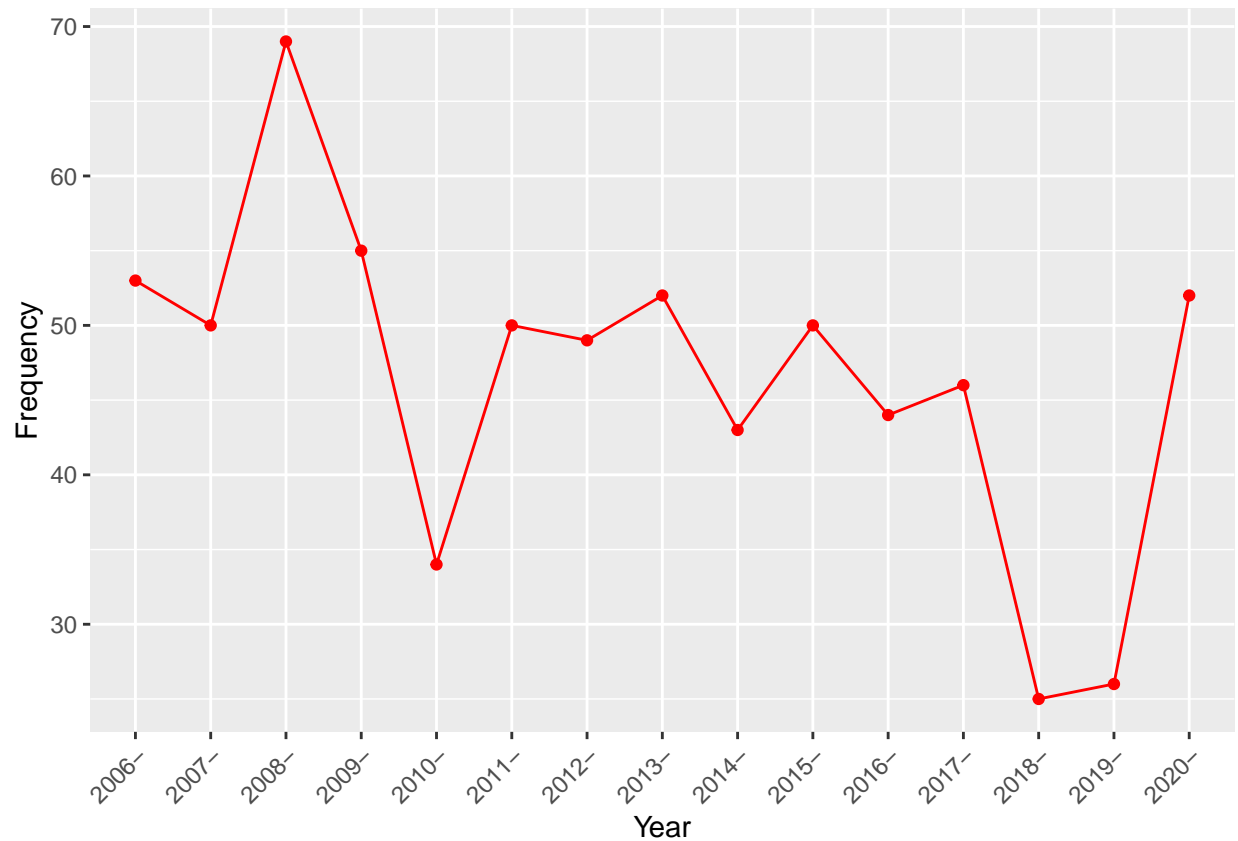
SI$YEAR <- substr(SI$OCCUR_DATE, nchar(SI$OCCUR_DATE) - n + 1, nchar(SI$OCCUR_DATE))

SI <- subset(SI, select = -c(OCCUR_DATE))

STATEN_ISLAND <- table(SI$YEAR)
STATEN_ISLAND <- as.data.frame(STATEN_ISLAND)

names(STATEN_ISLAND)[names(STATEN_ISLAND) == "Var1"] <- "Year"
names(STATEN_ISLAND)[names(STATEN_ISLAND) == "Freq"] <- "Frequency"

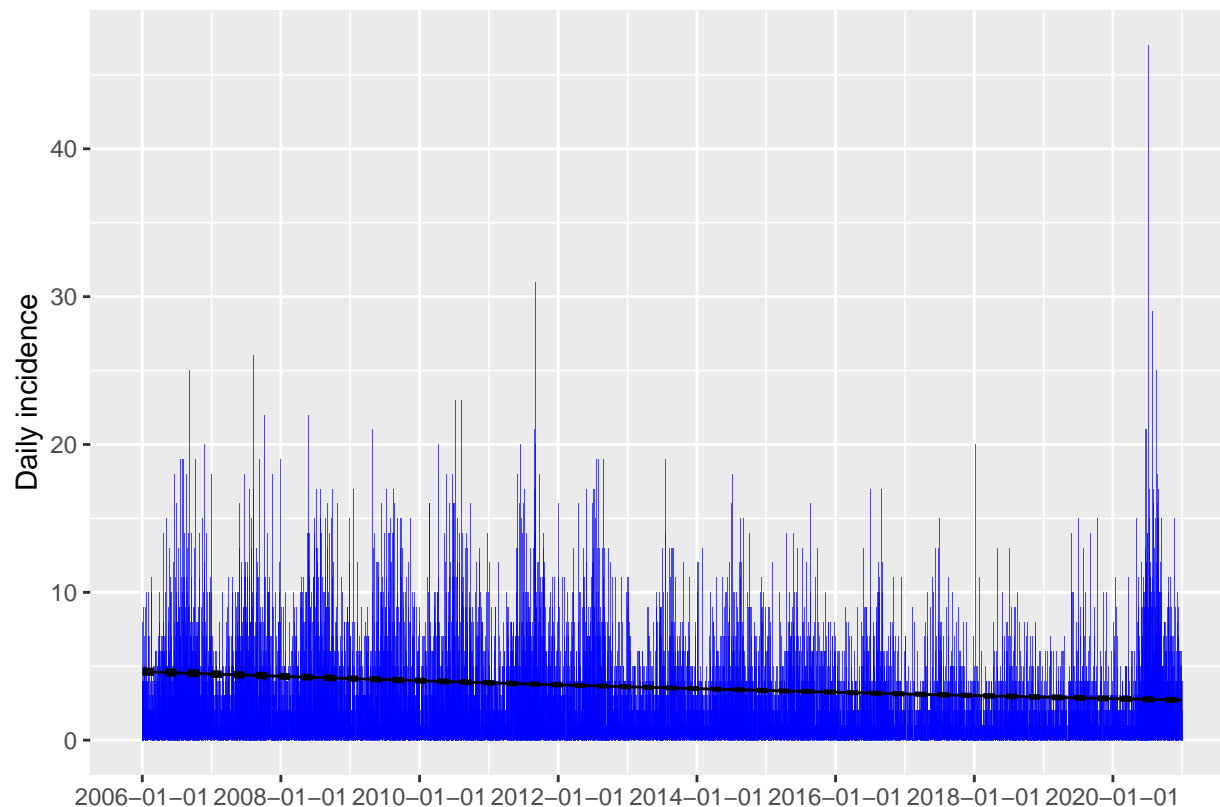
ggplot(data=STATEN_ISLAND, aes(x=Year, y=Frequency, group=1)) +
  geom_line(colour = "red")+
  geom_point(colour = "red")+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Shootings Overall

The plots from Brooklyn and Staten Island show an overall decline with a sharp increase in shootings in the year 2020. Since this is only in 2 of the boroughs, we can use a model to determine whether shootings in NYC as a whole have increased or decreased over time.

```
sh <- incidence (nyshootings$OCCUR_DATE)
sh_fit <- fit(sh)
plot(sh, fit = sh_fit, color = "blue")
```



Conclusion

This data set contained a lot of data which could be used for analysis, from victim and perpetrator demographics, daily/yearly shootings, and even the latitude and longitude of each shooting. This particular analysis focused on the overall totals for each of the boroughs listed as well as breaking down the totals for the boroughs with the most and least shootings: Brooklyn and Staten Island.

Overall, in 2020, there has been an increase in the frequency of shootings in both Brooklyn and Staten Island as seen in the spike of each graph. However using the model, we can see that the overall trend of shootings in NYC as a whole trended downward. This analysis doesn't take in to consideration any outside influences on these numbers, nor does it deal with the demographics of the victims and perpetrators, or whether the size or geography of each borough played a role. This could be potential sources of bias in the analysis. On a personal level, I have no connection to NYC and the boroughs listed, and chose my analysis based on the shooting data. Being an immigrant to America, my knowledge of these places is based on TV shows and news stories, which create a favorable vs. unfavourable image that could be a source of bias for future analysis.