NPYD Shooting Incident

Shyam Menon

5/10/2021

Dataset Maniupulation

Importing the Data - NYPD Shooting Incident Data (Historic)

List of every shooting incident that occurred in NYC going back to 2006 through the end of the previous calendar year.

This is a breakdown of every shooting incident that occurred in NYC going back to 2006 through the end of the previous calendar year. This data is manually extracted every quarter and reviewed by the Office of Management Analysis and Planning before being posted on the NYPD website. Each record represents a shooting incident in NYC and includes information about the event, the location and time of occurrence. In addition, information related to suspect and victim demographics is also included. :

```
nypd_data <- read.csv(url("https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLO
```

Cleaning and Transforming the Data

Getting rid of unneeded columns such as X_COORD_CD, Y_COORD_CD, Latitude, Longitude, and Lon_Lat. Then cleaning up the date and time columns by combining them into a singular column; then changing the data to a date object.

```
nypd <- subset(nypd_data, select = -c(X_COORD_CD, Y_COORD_CD, Latitude, Longitude, Lon_Lat))
nypd <- mutate(nypd, OCCUR_DATE_TIME = str_c(OCCUR_DATE, OCCUR_TIME, sep=" "))
nypd <- mutate(nypd, OCCUR_DATE_TIME = mdy_hms(OCCUR_DATE_TIME))
nypd <- mutate(nypd, OCCUR_DATE = mdy(OCCUR_DATE))
nypd <- mutate(nypd, OCCUR_TIME = hms(OCCUR_TIME))

nypd <- nypd %>%
filter(VIC_AGE_GROUP != "UNKNOWN")
summary(nypd)
```

```
##   INCIDENT_KEY        OCCUR_DATE           OCCUR_TIME
##  Min.   :  9953245   Min.   :2006-01-01   Min.   :0S
##  1st Qu.: 55322800   1st Qu.:2008-12-30   1st Qu.:3H 19M 0S
##  Median : 83357175   Median :2012-02-26   Median :15H 0M 0S
##  Mean   :102205248   Mean   :2012-10-03   Mean   :12H 32M 43.0719482619243S
##  3rd Qu.:150772440   3rd Qu.:2016-02-27   3rd Qu.:20H 45M 0S
##  Max.   :222473262   Max.   :2020-12-31   Max.   :23H 59M 0S
##
##      BORO              PRECINCT      JURISDICTION_CODE LOCATION_DESC
```

```
##  Length:23503      Min.  :  1.0   Min.  :0.0000    Length:23503
##  Class :character   1st Qu.: 44.0  1st Qu.:0.0000   Class :character
##  Mode  :character   Median : 69.0  Median :0.0000   Mode  :character
##                     Mean  : 66.2   Mean  :0.3325
##                     3rd Qu.: 81.0  3rd Qu.:0.0000
##                     Max.  :123.0   Max.  :2.0000
##                                    NA's  :2
##  STATISTICAL_MURDER_FLAG PERP_AGE_GROUP      PERP_SEX
##  Length:23503           Length:23503       Length:23503
##  Class :character       Class :character   Class :character
##  Mode  :character       Mode  :character   Mode  :character
##
##
##
##
##   PERP_RACE          VIC_AGE_GROUP        VIC_SEX             VIC_RACE
##  Length:23503       Length:23503       Length:23503       Length:23503
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  OCCUR_DATE_TIME
##  Min.   :2006-01-01 02:00:00
##  1st Qu.:2008-12-30 12:20:00
##  Median :2012-02-26 00:03:00
##  Mean   :2012-10-03 20:40:28
##  3rd Qu.:2016-02-27 20:55:00
##  Max.   :2020-12-31 23:45:00
##
```

Data Analysis

Total Shootings by Borough

Analysis to see the total instances of shootings in each of the Boroughs in NYC from 2006-2020.

```r
borough <- nypd %>%
  group_by(BORO) %>%
  count()

p <- ggplot(borough, aes(x = BORO, weight = n)) +
  geom_bar(width = 0.5, fill = "red") +
  labs(x = "Borough", y = "Total Shootings",
  title = "Total Shootings by Borough (2006-2020)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  geom_text(aes(x = BORO, y = n + 1, label = n))

plot(p)
```

Total Shootings by Borough (2006–2020)
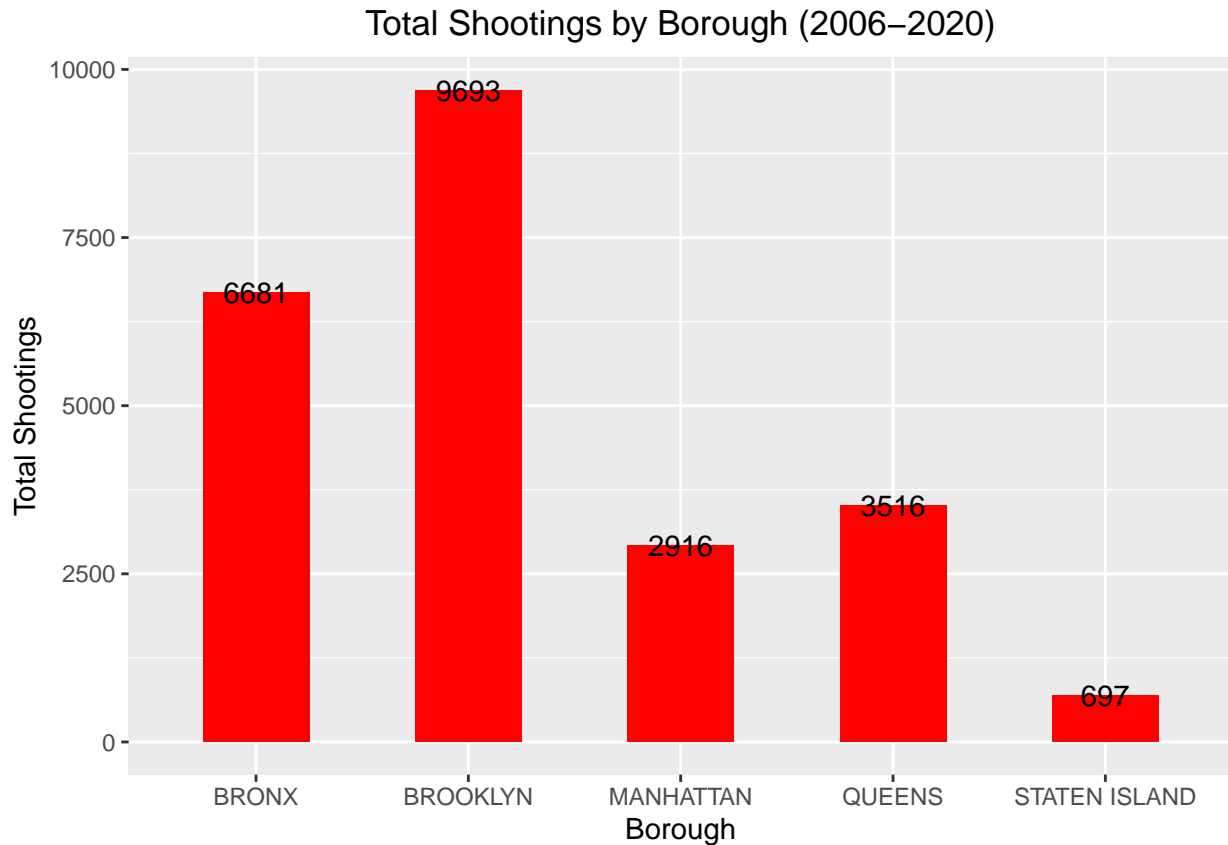
**Gender of Victims vs. Number of Shootings in Brooklyn, NYC 2020.**

Analysis to identify the genders of shooting victims in Brooklyn, NYC from 2020. F = Female M = Male

```r
brooklyn <- nypd %>%
filter(BORO == "BROOKLYN") %>%
filter(OCCUR_DATE_TIME > ymd_hms("2020-01-01 00:00:00"))


sex_brooklyn <- brooklyn %>%
filter(VIC_SEX != "U") %>%
group_by(VIC_SEX) %>%
summarise(gender_count=n())

sex_plot <- ggplot(sex_brooklyn, aes(x = VIC_SEX, weight = gender_count)) +
 geom_bar(width = 0.5, fill = "green") +
 labs(x = "Sex", y = "Number of Incidences", title = "Victim Gender from 2020 Shootings in Brooklyn") +
 theme(plot.title = element_text(hjust = 0.5)) +
 geom_text(aes(x = VIC_SEX, y = gender_count, label = gender_count))

plot(sex_plot)
```
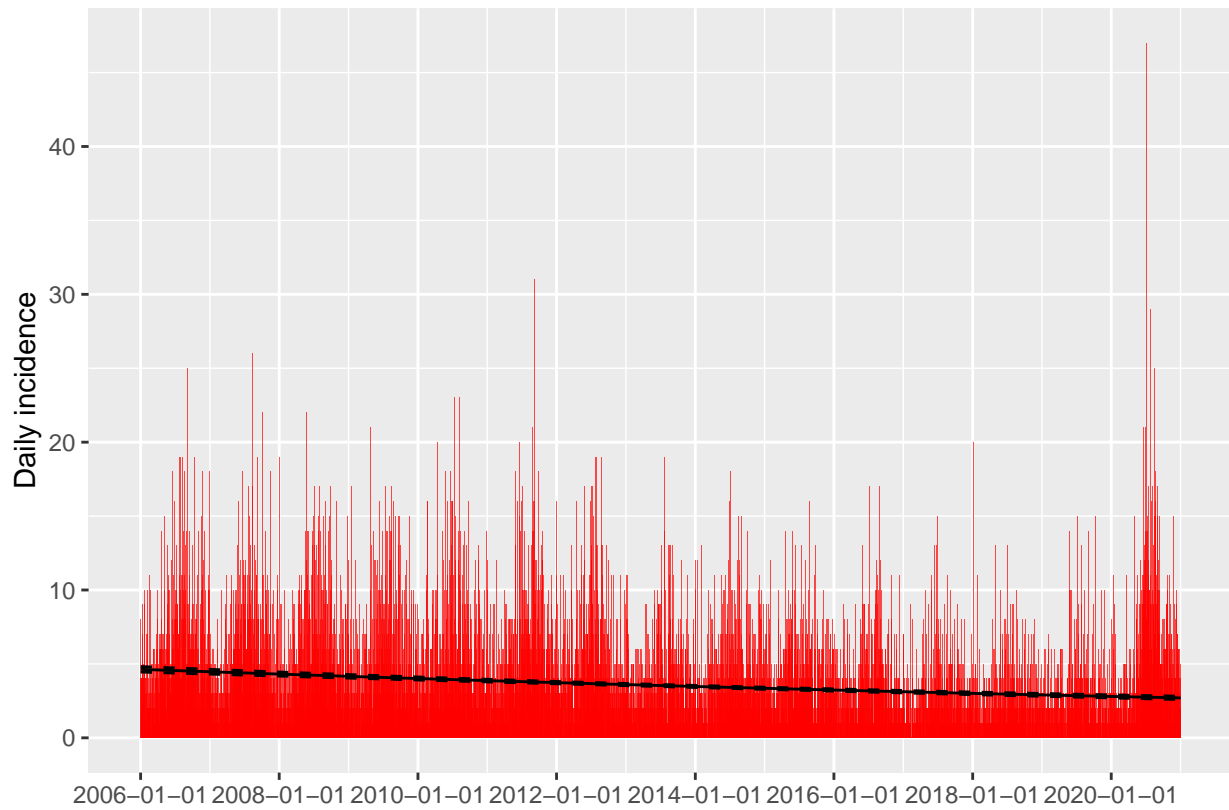
Victim Gender from 2020 Shootings in Brooklyn
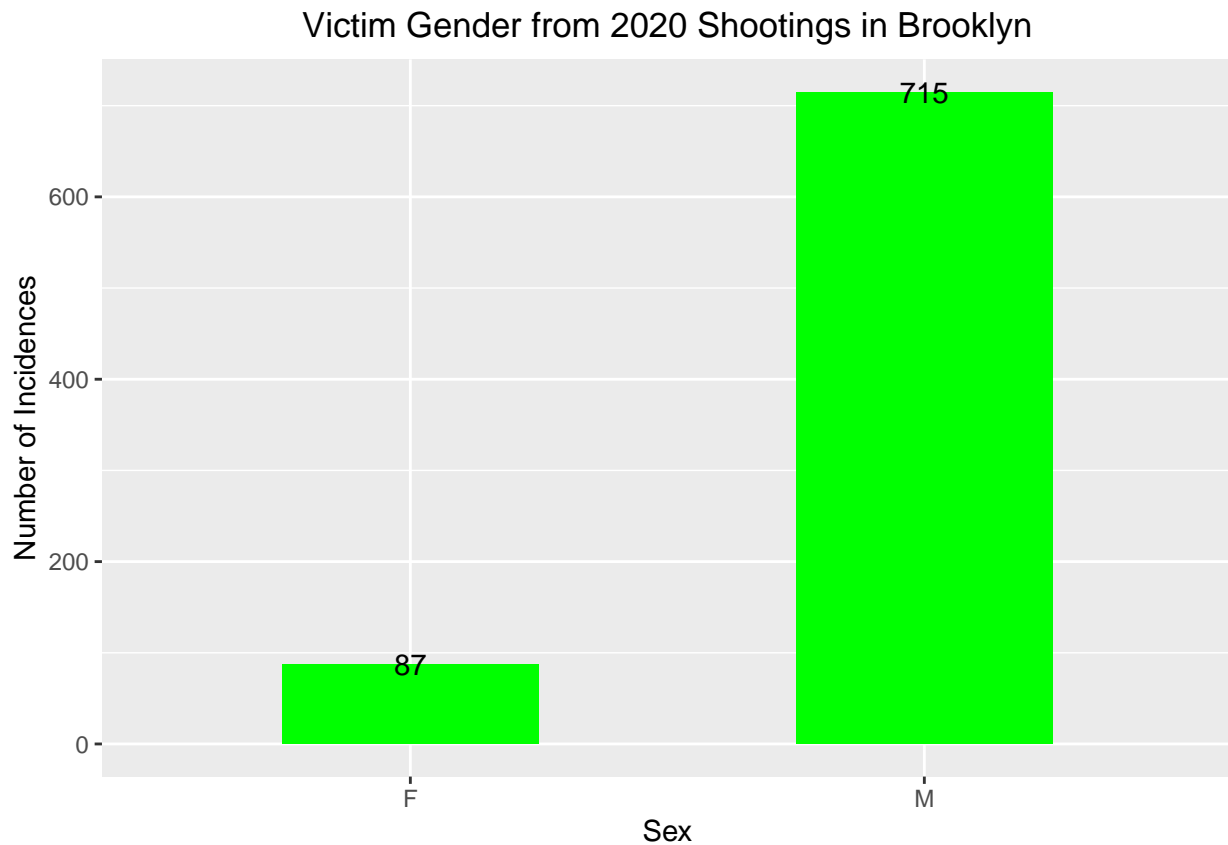
Data Modeling

Modeling the Daily Incidence of shootings in NYC

Incidence modeling of all the shootings in NYC from 2006 to 2020. A negative slope suggests an overall decline in shooting incidences in the city over time.

```
i <- incidence (nypd$OCCUR_DATE)
i_fit <- fit(i)

plot(i, fit = i_fit, color = "red")
```

Conclusion

The NYPD shooting data set was very interesting to analyze. The data on the victims of the shootings seem to be complete where as some of the data for the perpetrator was missing. This could lead the data to be bias because the lack of a complete picture. Personal bias could also be a factor if picking specific age groups, sex, or race; however, that was avoided by picking the whole data column and not excluding any data points.

While there have been many spikes of daily shooting incidences over time, the overall trend suggests that the daily incidence rates have fallen over time. Further analysis should be done at a later date in order to see if the current trend changes.