

NYPD Shooting Data

JamesYildiz

June 8, 2021

This report analyzes the shooting incidents data from years 2006 through 2020 in New York City boroughs.

Importing Data

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

```
## [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"
```

```
summary(nypd_data)
```

```
##  INCIDENT_KEY      OCCUR_DATE      OCCUR_TIME      BORO
##  Min.   : 9953245   Length:23568   Length:23568   Length:23568
##  1st Qu.: 55317014  Class :character  Class :character  Class :character
##  Median : 83365370  Mode  :character  Mode  :character  Mode  :character
##  Mean    :102218616
##  3rd Qu.:150772442
##  Max.    :222473262
##
##  PRECINCT      JURISDICTION_CODE  LOCATION_DESC      STATISTICAL_MURDER_FLAG
##  Min.   : 1.00   Min.   :0.0000   Length:23568   Length:23568
##  1st Qu.: 44.00  1st Qu.:0.0000   Class :character  Class :character
##  Median : 69.00  Median :0.0000   Mode  :character  Mode  :character
##  Mean    : 66.21  Mean    :0.3323
##  3rd Qu.: 81.00  3rd Qu.:0.0000
##  Max.    :123.00  Max.    :2.0000
##
##  NA's      :2
##  PERP_AGE_GROUP      PERP_SEX      PERP_RACE      VIC_AGE_GROUP
```

```
## 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_SEX          VIC_RACE          X_COORD_CD        Y_COORD_CD
## Length:23568      Length:23568      Length:23568      Length:23568
## Class :character  Class :character  Class :character  Class :character
## Mode :character  Mode :character  Mode :character  Mode :character
##
##
##
##
## Latitude         Longitude         Lon_Lat
## Min. :40.51      Min. : -74.25      Length:23568
## 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
##
##
```

Importing Libraries

The following libraries will be needed for the cleaning and analysis of the data.

```
library(lubridate)
library(tidyverse)
library(ggplot2)
library(tidyr)
```

Data cleaning

I will remove the columns I won't need in the analysis.

```
#Removing columns
```

```
nypd_data <- nypd_data %>% select(-c(PRECINCT, JURISDICTION_CODE, X_COORD_CD, Y_COORD_CD, Latitude, Longitude))
colnames(nypd_data)
```

```
## [1] "OCCUR_DATE"          "BORO"
## [3] "STATISTICAL_MURDER_FLAG"
```

```
#Convert date column to proper date format
```

```
nypd_data <- mutate(nypd_data, OCCUR_DATE = mdy(OCCUR_DATE))
```

Total number of shootings per borough,

```
boro_totals <- table(nypd_data$BORO)
boro_totals <- as.data.frame(boro_totals) %>% rename(Borough = Var1, Frequency = Freq)
boro_totals
```

```
##      Borough Frequency
## 1      BRONX      6700
## 2    BROOKLYN     9722
## 3    MANHATTAN     2921
## 4      QUEENS     3527
## 5 STATEN ISLAND      698
```

The total population of each borough and shootings per 1,000 people

```
borough_population = c(1418000, 2560000, 1629000, 2254000, 476000)

boro_totals <- boro_totals %>%
  mutate(Population = borough_population)

Per_kcapita = round(((boro_totals$Frequency/boro_totals$Population) * 1000),2)

boro_totals <- boro_totals %>%
  mutate(Per_kcapita = Per_kcapita)

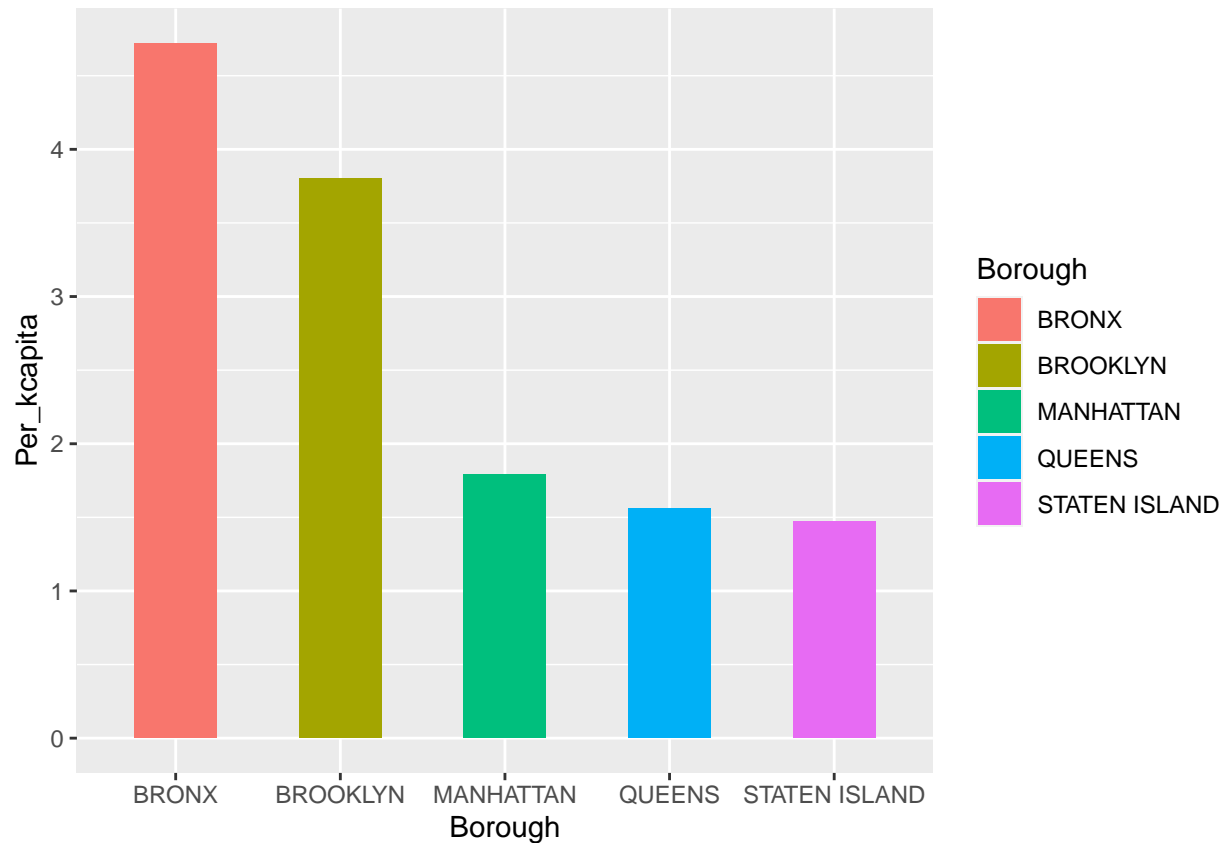
boro_totals
```

```
##      Borough Frequency Population Per_kcapita
## 1      BRONX      6700    1418000         4.72
## 2    BROOKLYN     9722    2560000         3.80
## 3    MANHATTAN     2921    1629000         1.79
## 4      QUEENS     3527    2254000         1.56
## 5 STATEN ISLAND      698     476000         1.47
```

>The data shows that percapita shooting is the highest in Bronx. Also, while Queens seems to have about five times higher number of shootings than Staten Island, the per capita shooting is about the same for both boroughs.

The histogram of the shootings per 1000 people by borough can be visualized as the following.

```
ggplot(boro_totals, aes(x=Borough, y=Per_kcapita, fill = Borough)) + geom_col(width=0.5)
```



Then, find the shootings in Bronx (borough with highest per capita shooting) by each month of the year.

```
BR <- subset(nypd_data, BORO=='BRONX', select=c(BORO, OCCUR_DATE))

BR$month <- substr(BR$OCCUR_DATE,6,7)

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

BRONX <- table(BR$month)
BRONX <- as.data.frame(BRONX)

colnames(BRONX)[colnames(BRONX) == "Var1"] <- "Months"
colnames(BRONX)[colnames(BRONX) == "Freq"] <- "Incidents"

ggplot(data=BRONX, aes(x=Months, y=Incidents, group = 1)) +
  geom_line(colour = "Blue") +
  geom_point(colour = "red")
```



Let's plot the number of incidents per month for Brooklyn.

```
BRLYN <- subset(nypd_data, BORO=='BROOKLYN', select=c(BORO, OCCUR_DATE))

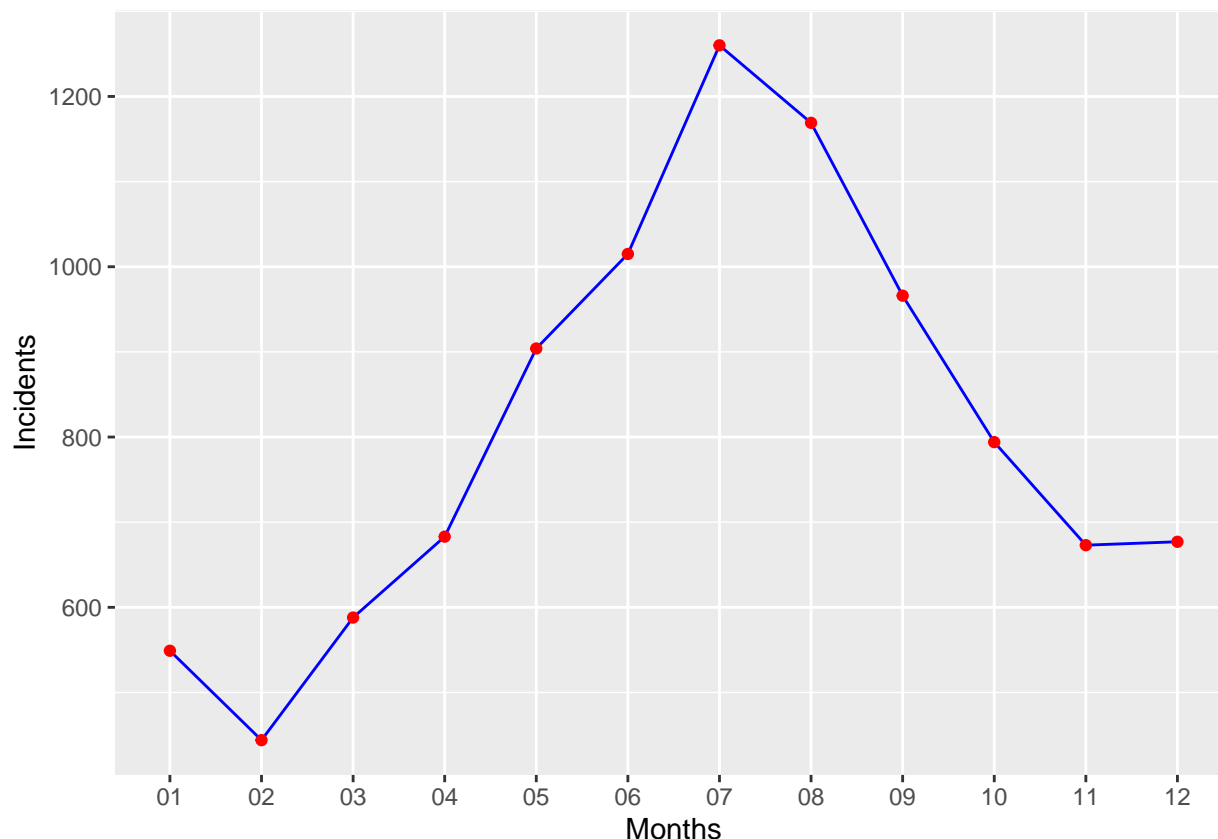
BRLYN$month <- substr(BRLYN$OCCUR_DATE,6,7)

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

BROOKLYN <- table(BRLYN$month)
BROOKLYN <- as.data.frame(BROOKLYN)

colnames(BROOKLYN)[colnames(BROOKLYN) == "Var1"] <- "Months"
colnames(BROOKLYN)[colnames(BROOKLYN) == "Freq"] <- "Incidents"

ggplot(data=BROOKLYN, aes(x=Months, y=Incidents, group = 1)) +
  geom_line(colour = "Blue") +
  geom_point(colour = "red")
```



Bias Analysis

In boroughs with heavy gang activities, many shooting activities may not get reported to the police. This may add bias to the number from these boroughs.

On a personal level, I do not see any sources of bias on this data set. However, the selection of cities that I plan to extend this study may be affected by my personal connection to the cities. The cities that I know well may produce a more detailed report which will add bias to NYC data as it will not be as detailed and accurate on a comparative basis.

Conclusion and the next steps.

As it can be seen on the plots, There is a significant increase in the shooting cases during the summer months. This may be due to increased social interactions among people during these months. I would like to analyze similar data sets from Los Angeles CA, Houston TX, and Orlando FL where the seasonal effects on social interactions are considerably less than that of NYC. Also, based on these two boroughs there is a predictive ratio between summer peak number and the months of November and December. The summer peak to holiday-season incidents ratio is around 1.7:1.0.

Another addition to this project would be to analyze smaller zones in each borough using the map coordinates and investigate the effect of the shooting incidents on property prices and school districts