NYC Shooting Incident Analysis Report

Sriram Nyshadham

2/20/2022

Historical NYPD Shooting Incident data

The New York Police Department (NYPD) provides a dataset of shooting incidents in New York City from 2006 to the previous calendar year. This data only includes shooting incidents resulting in an injured victim. Thus, it does not represent every shooting incident in New York City.

The data and data dictionary are available at https://data.cityofnewyork.us/Public-Safety/NYPD-Shooting-Incident-Data-Historic-/833y-fsy8. Footnotes accompanying the data are available at https://data.cityofnewyork.us/api/views/833y-fsy8/files/e4e3d86c-348f-4a16-a17f-19480c089429?download=true&filename=NYPD_Shootings_Incident_Level_Data_Footnotes.pdf.

Raw Shooting Data Structure

```
url_in <-"https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
nydata <- read.csv(url_in)
# check out the data
str(nydata)
```

```
## 'data.frame':    23585 obs. of  19 variables:
##  $ INCIDENT_KEY           : int  24050482 77673979 203350417 80584527 90843766 92393427 73057167 2113
##  $ OCCUR_DATE             : chr  "08/27/2006" "03/11/2011" "10/06/2019" "09/04/2011" ...
##  $ OCCUR_TIME             : chr  "05:35:00" "12:03:00" "01:09:00" "03:35:00" ...
##  $ BORO                   : chr  "BRONX" "QUEENS" "BROOKLYN" "BRONX" ...
##  $ PRECINCT               : int  52 106 77 40 100 67 77 81 101 106 ...
##  $ JURISDICTION_CODE      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ LOCATION_DESC          : chr  "" "" "" "" ...
##  $ STATISTICAL_MURDER_FLAG: chr  "true" "false" "false" "false" ...
##  $ PERP_AGE_GROUP         : chr  "" "" "" "" ...
##  $ PERP_SEX               : chr  "" "" "" "" ...
##  $ PERP_RACE              : chr  "" "" "" "" ...
##  $ VIC_AGE_GROUP          : chr  "25-44" "65+" "18-24" "<18" ...
##  $ VIC_SEX                : chr  "F" "M" "F" "M" ...
##  $ VIC_RACE               : chr  "BLACK HISPANIC" "WHITE" "BLACK" "BLACK" ...
##  $ X_COORD_CD             : num  1017542 1027543 995325 1007453 1041267 ...
##  $ Y_COORD_CD             : num  255919 186095 185155 233952 157134 ...
##  $ Latitude               : num  40.9 40.7 40.7 40.8 40.6 ...
##  $ Longitude              : num  -73.9 -73.8 -74 -73.9 -73.8 ...
##  $ Lon_Lat                : chr  "POINT (-73.87963173099996 40.86905819000003)" "POINT (-73.843920197
```

Raw Shooting Data Summary

```
summary(nydata)
```

```
##    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
##
```

Recoded Shooting data with variables of interest.

Age group, sex, and race contain inconsistent labels for both victim and perpetrator. These six variables are recoded so that all unknown values are missing. Records with missing values will be excluded from analyses unless noted otherwise.

```r
# relabel the empty strings as missing on the factor variables
#levels(nydata$PERP_AGE_GROUP)
which.one <- which( levels(nydata$PERP_AGE_GROUP) == "" )
#which.one
levels(nydata$PERP_AGE_GROUP)[which.one] <- "UNKNOWN"
#levels(nydata$PERP_AGE_GROUP)

#levels(nydata$PERP_SEX)
```

```r
which.one <- which( levels(nydata$PERP_SEX) == "" )
#which.one
levels(nydata$PERP_SEX)[which.one] <- "U"
#levels(nydata$PERP_SEX)

#levels(nydata$PERP_RACE)
which.one <- which( levels(nydata$PERP_RACE) == "" )
#which.one
levels(nydata$PERP_RACE)[which.one] <- "UNKNOWN"
#levels(nydata$PERP_RACE)

# keep variables of interest
ny <- nydata %>%
  select(INCIDENT_KEY, OCCUR_DATE, BORO, STATISTICAL_MURDER_FLAG,
         PERP_AGE_GROUP, VIC_AGE_GROUP, VIC_SEX, PERP_SEX, PERP_RACE, VIC_RACE, Latitude, Longitude) %>%
  mutate(OCCUR_DATE = mdy(OCCUR_DATE)) %>%
  mutate(YEAR = year(OCCUR_DATE)) %>%
  mutate(PERP_AGE_GROUP = recode_factor(PERP_AGE_GROUP, "<18" = "<18", "18-24" = "18-24",
                                        "25-44"="25-44", "45-64"="45-64", "65+" = "65+", "224" = NULL,
                                        "940" = NULL, "1020" = NULL, "UNKNOWN" = NULL, .missing= NULL)) %>%
  mutate(VIC_AGE_GROUP = recode_factor(VIC_AGE_GROUP, "<18" = "<18", "18-24" = "18-24",
                                       "25-44"="25-44", "45-64"="45-64", "65+" = "65+", "UNKNOWN" = NULL, .m
  mutate(PERP_SEX = recode_factor(PERP_SEX, "F" = "FEMALE", "M" = "MALE", "U" = NULL, .missing= NULL))
  mutate(VIC_SEX = recode_factor(VIC_SEX, "F" = "FEMALE", "M" = "MALE", "U" = NULL, .missing= NULL)) %>%
  mutate(PERP_RACE = recode_factor(PERP_RACE, "WHITE" = "WHITE", "WHITE HISPANIC" = "WHITE",
                                   "BLACK" = "BLACK", "BLACK HISPANIC" = "BLACK", "ASIAN / PACIFIC ISLA
                                   "AMERICAN INDIAN/ALASKAN NATIVE" = "AAPI/AIAN", "UNKNOWN" = NULL, .m
  mutate(VIC_RACE = recode_factor(VIC_RACE, "WHITE" = "WHITE", "WHITE HISPANIC" = "WHITE",
                                  "BLACK" = "BLACK", "BLACK HISPANIC" = "BLACK", "ASIAN / PACIFIC ISLA
                                  "AMERICAN INDIAN/ALASKAN NATIVE" = "AAPI/AIAN", "UNKNOWN" = NULL, .m

# summarize the prepared data
summary(ny)
```

```
##   INCIDENT_KEY          OCCUR_DATE              BORO
## Min.   :  9953245   Min.   :2006-01-01   Length:23585
## 1st Qu.: 55322804   1st Qu.:2008-12-31   Class :character
## Median : 83435362   Median :2012-02-27   Mode  :character
## Mean   :102280741   Mean   :2012-10-05
## 3rd Qu.:150911774   3rd Qu.:2016-03-02
## Max.   :230611229   Max.   :2020-12-31
##
## STATISTICAL_MURDER_FLAG PERP_AGE_GROUP VIC_AGE_GROUP   VIC_SEX
## Length:23585            <18  :1368     <18  : 2525   FEMALE: 2204
## Class :character        18-24:5508     18-24: 9003   MALE  :21370
## Mode  :character        25-44:4714     25-44:10303   NA's  :   11
##                         45-64: 495     45-64: 1541
##                         65+  :  54     65+  :  154
##                              :8295     NA's :   59
##                         NA's :3151
##    PERP_SEX          PERP_RACE           VIC_RACE         Latitude
## FEMALE:  335   WHITE   : 2243   WHITE   : 4070   Min.   :40.51
## MALE  :13490   BLACK   :11121   BLACK   :19114   1st Qu.:40.67
##       : 8261   AAPI/AIAN:  124   AAPI/AIAN:  336   Median :40.70
```

```
##   NA's  : 1499                : 8261   NA's      :   65   Mean   :40.74
##                 NA's    : 1836                           3rd Qu.:40.82
##                                                          Max.   :40.91
##
##    Longitude          YEAR
##  Min.   :-74.25   Min.   :2006
##  1st Qu.:-73.94   1st Qu.:2008
##  Median :-73.92   Median :2012
##  Mean   :-73.91   Mean   :2012
##  3rd Qu.:-73.88   3rd Qu.:2016
##  Max.   :-73.70   Max.   :2020
##
```

Basic statistics of New York City Shooting Incidents

Brooklyn has the highest total number of murdered shooting incident victims, but Staten Island has the
highest percentage of shooting victims that resulted in murder.

```
table(ny$BORO, ny$STATISTICAL_MURDER_FLAG, useNA="ifany")
```

```
##
##                 false true
##    BRONX         5454 1247
##    BROOKLYN      7836 1898
##    MANHATTAN     2407  515
##    QUEENS        2835  697
##    STATEN ISLAND  553  143
```

```
# row percentages
prop.table(table(ny$BORO, ny$STATISTICAL_MURDER_FLAG, useNA="ifany"), 1)
```

```
##
##                     false      true
##    BRONX         0.8139084 0.1860916
##    BROOKLYN      0.8050134 0.1949866
##    MANHATTAN     0.8237509 0.1762491
##    QUEENS        0.8026614 0.1973386
##    STATEN ISLAND 0.7945402 0.2054598
```

Victim demographics

The majority of shooting victims are black males aged 25-44.

```
table(ny$VIC_AGE_GROUP, ny$VIC_SEX, ny$VIC_RACE, useNA="ifany")
```

```
## , ,  = WHITE
##
##
##        FEMALE MALE <NA>
##   <18       67  312    0
##   18-24    111 1341    0
##   25-44    178 1610    0
##   45-64     93  280    0
##   65+       16   41    0
##   <NA>       1   20    0
##
## , ,  = BLACK
```

4

```
##
##
##        FEMALE MALE <NA>
##   <18      295 1832    0
##   18-24    553 6878    4
##   25-44    622 7706    0
##   45-64    201  909    0
##   65+       31   61    0
##   <NA>       4   18    0
##
## , ,  = AAPI/AIAN
##
##
##        FEMALE MALE <NA>
##   <18        3   12    0
##   18-24      4   92    0
##   25-44     16  150    0
##   45-64      6   47    0
##   65+        2    3    0
##   <NA>       0    1    0
##
## , ,  = NA
##
##
##        FEMALE MALE <NA>
##   <18        0    4    0
##   18-24      0   20    0
##   25-44      1   18    2
##   45-64      0    5    0
##   65+        0    0    0
##   <NA>       0   10    5
```

```
#murder <- filter(ny, STATISTICAL_MURDER_FLAG=="true")
#table(murder$VIC_AGE_GROUP, murder$VIC_SEX, murder$VIC_RACE, useNA="ifany")
```

Perpetrator demographics

A large proportion of perpetrator demographics are missing in this data. Race is missing for over half of the incidents. One should use caution making any statements about the demographics of perpetrators based on the non-missing data. It should also be noted that this data is presented as one row per injured victim. There is no identifier key for the perpetrator, so it is not possible to make conclusions about the perpetrators (e.g., one shooting incident may involve multiple perpetrators).

```
table(ny$PERP_AGE_GROUP, ny$PERP_SEX, ny$PERP_RACE, useNA="ifany")
```

```
## , ,  = WHITE
##
##
##        FEMALE MALE      <NA>
##   <18        9  206    0    0
##   18-24     38  882    0    0
##   25-44     30  782    0    0
##   45-64      4  129    0    0
##   65+        0   23    0    0
##            0    0    0    0
```

```
##    <NA>       2  138    0    0
## 
## , ,  = BLACK
## 
## 
##          FEMALE MALE      <NA>
##    <18       25 1099    0    1
##    18-24     85 4395    0    9
##    25-44    109 3671    0    3
##    45-64     13  338    0    0
##    65+        1   29    0    0
##               0    0    0    0
##    <NA>      12 1324    0    7
## 
## , ,  = AAPI/AIAN
## 
## 
##          FEMALE MALE      <NA>
##    <18        0   12    0    0
##    18-24      2   33    0    0
##    25-44      0   60    0    0
##    45-64      0    6    0    0
##    65+        0    0    0    0
##               0    0    0    0
##    <NA>       0   11    0    0
## 
## , ,  =
## 
## 
##          FEMALE MALE      <NA>
##    <18        0    0    0    0
##    18-24      0    0    0    0
##    25-44      0    0    0    0
##    45-64      0    0    0    0
##    65+        0    0    0    0
##               0    0 8261    0
##    <NA>       0    0    0    0
## 
## , ,  = NA
## 
## 
##          FEMALE MALE      <NA>
##    <18        0   14    0    2
##    18-24      1   56    0    7
##    25-44      0   55    0    4
##    45-64      0    5    0    0
##    65+        0    1    0    0
##               0    0    0   34
##    <NA>       4  221    0 1432
```

Monthly shooting incidents in New York City

Shooting incidents peak each year during summer months. The years 2017-2019 have lower peaks than other years. However, 2020 has a higher peak than all previous years in the data.
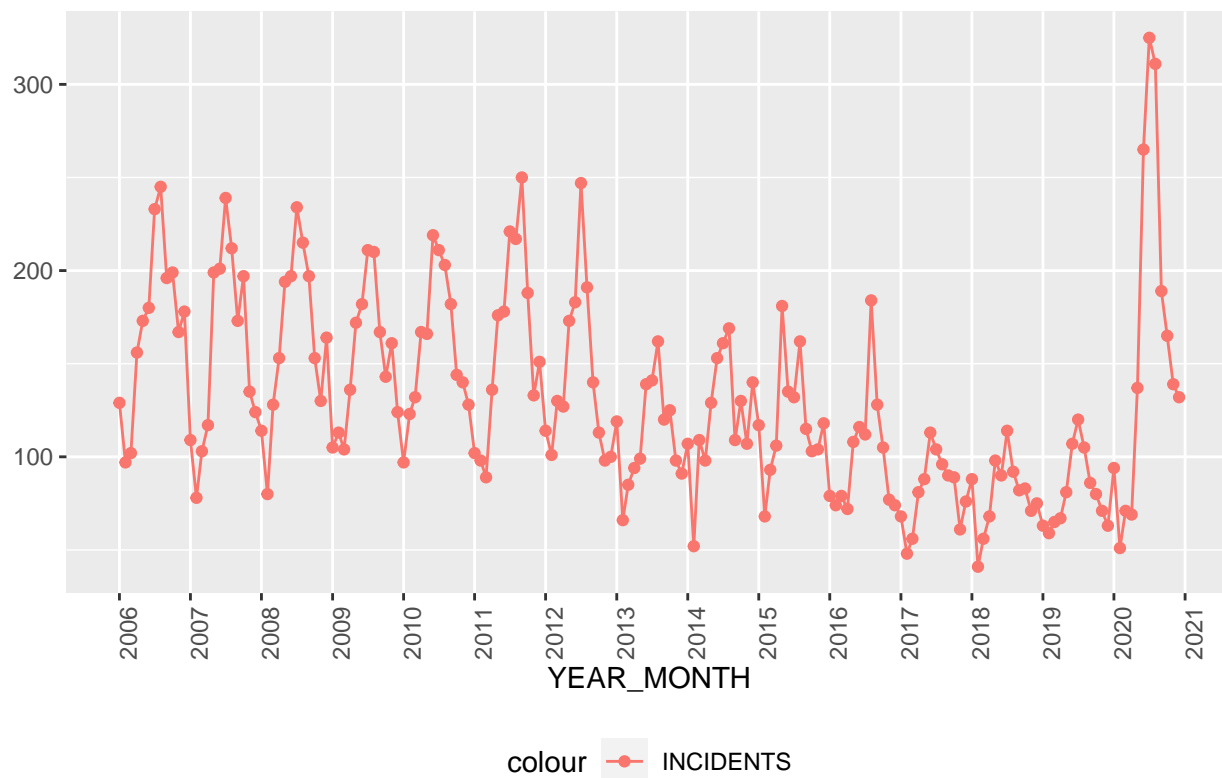
```
NY_Totals <- ny %>%
  group_by(YEAR_MONTH = floor_date(OCCUR_DATE, "month")) %>%
  summarize(INCIDENTS = n()) %>%
  ungroup()
# NY_Totals

NY_Totals %>%
  ggplot(aes(x=YEAR_MONTH, y=INCIDENTS)) +
  geom_line(aes(color="INCIDENTS")) +
  geom_point(aes(color="INCIDENTS")) +
  theme(legend.position = "bottom",
        axis.text.x = element_text(angle=90)) +
  labs(title="Number of Shooting Incidents in NYC by Month", y=NULL) +
  scale_x_date(date_breaks="year", date_labels="%Y", minor_breaks=NULL)
```


Shooting incidents in New York City by Borough

More shooting incidents occur in Brooklyn and Bronx than the other boroughs. Staten Island has the fewest shooting incidents.

```
Boro_Totals <- ny %>%
  group_by(YEAR_MONTH = floor_date(OCCUR_DATE, "month"), BORO) %>%
  summarize(INCIDENTS = n(), .groups='drop') %>%
  ungroup()

#Boro_Totals
```

```
Boro_Totals %>%
  ggplot(aes(x=YEAR_MONTH, y=INCIDENTS, group=BORO, color=BORO)) +
  geom_line() +
  geom_point() +
  theme(legend.position = "bottom",
        axis.text.x = element_text(angle=90)) +
  labs(title="Number of Shooting Incidents by Borough", y=NULL) +
  scale_x_date(date_breaks="year", date_labels="%Y", minor_breaks=NULL)
```

Number of Shooting Incidents by Borough
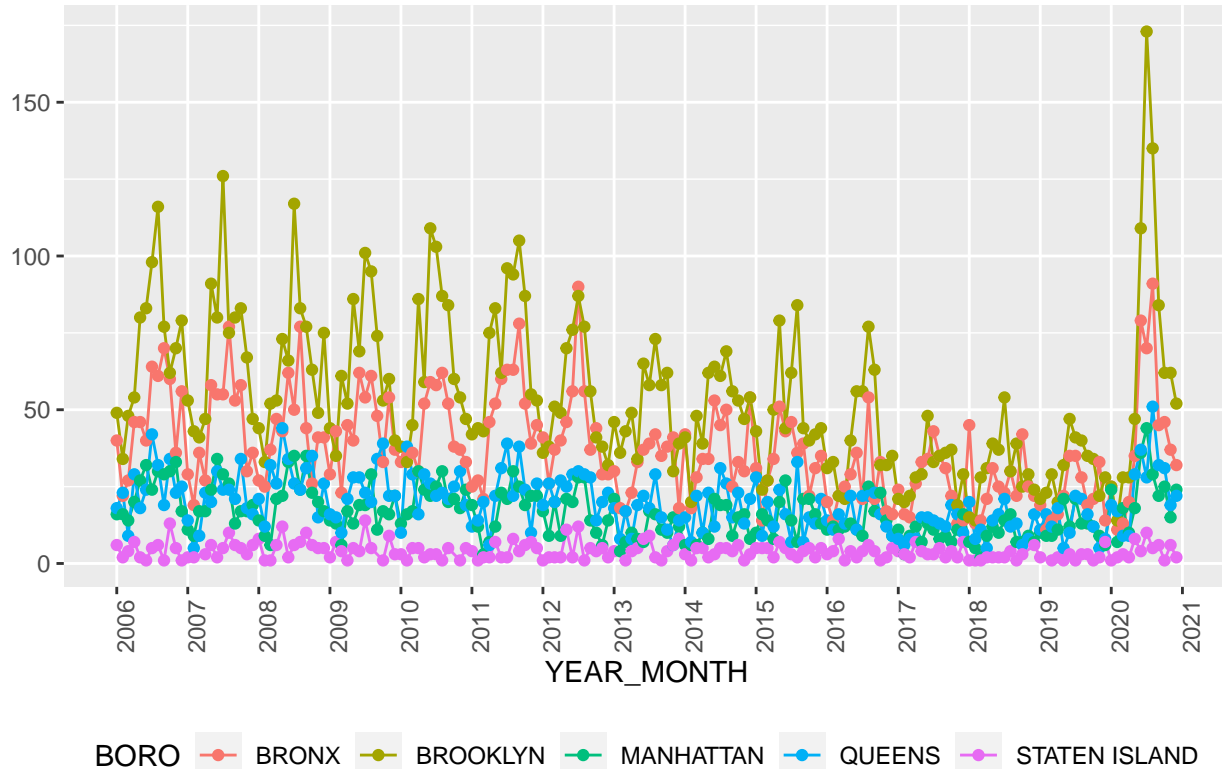

```
# Create a map of New York City based on data latitude and longitude
# Using Stamen Maps at http://maps.stamen.com/ to avoid Google Maps charges and security risk

#summary(nydata$Latitude)
#summary(nydata$Longitude)

# create map borders based on the data
height <- max(nydata$Latitude) - min(nydata$Latitude)
width <- max(nydata$Longitude) - min(nydata$Longitude)
borders <- c(bottom  = min(nydata$Latitude)  - 0.1 * height,
             top     = max(nydata$Latitude)  + 0.1 * height,
             left    = min(nydata$Longitude) - 0.1 * width,
             right   = max(nydata$Longitude) + 0.1 * width)

# get map data within borders
nyc_map <- get_stamenmap(borders, zoom = 11, maptype = "toner-lite")
# ggmap(nyc_map)
```
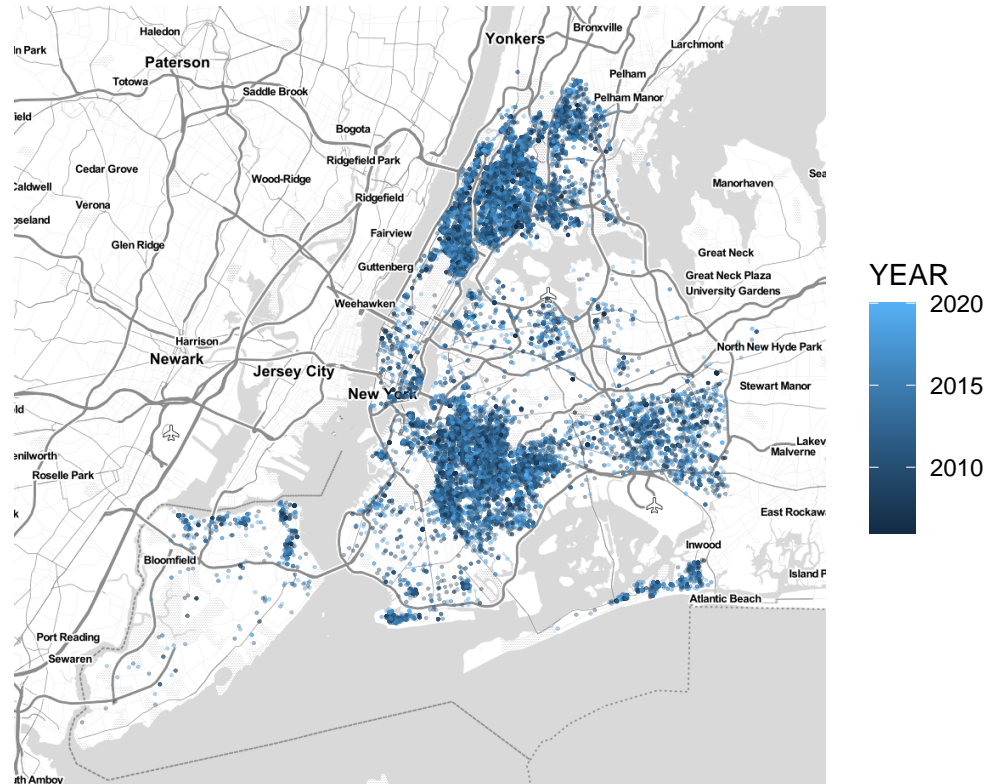
Map of shooting incidents in New York City, 2006-2020

```
ggmap(nyc_map) +
  geom_point(data=ny, aes(x=Longitude, y=Latitude, color=YEAR),
               size=0.1, alpha=0.4)  +
    ggtitle("Shooting Incidents in New York City from 2006 to 2020") +
  theme(axis.ticks = element_blank(),
          axis.text = element_blank(),
          axis.title = element_blank())
```

Shooting Incidents in New York City from 2006 to 2020

Map of shooting incidents in New York City, 2020

```
year=2020

ny_year <- ny %>%
  filter(YEAR==year)

# summary(ny_year)

ggmap(nyc_map) +
  geom_point(data=ny_year, aes(x=Longitude, y=Latitude, color=YEAR),
               size=0.1, alpha=0.4)  +
    ggtitle("Shooting Incidents in New York City in 2020") +
  theme(axis.ticks = element_blank(),
          axis.text = element_blank(),
          axis.title = element_blank(),
        legend.position="none")
```

Shooting Incidents in New York City in 2020
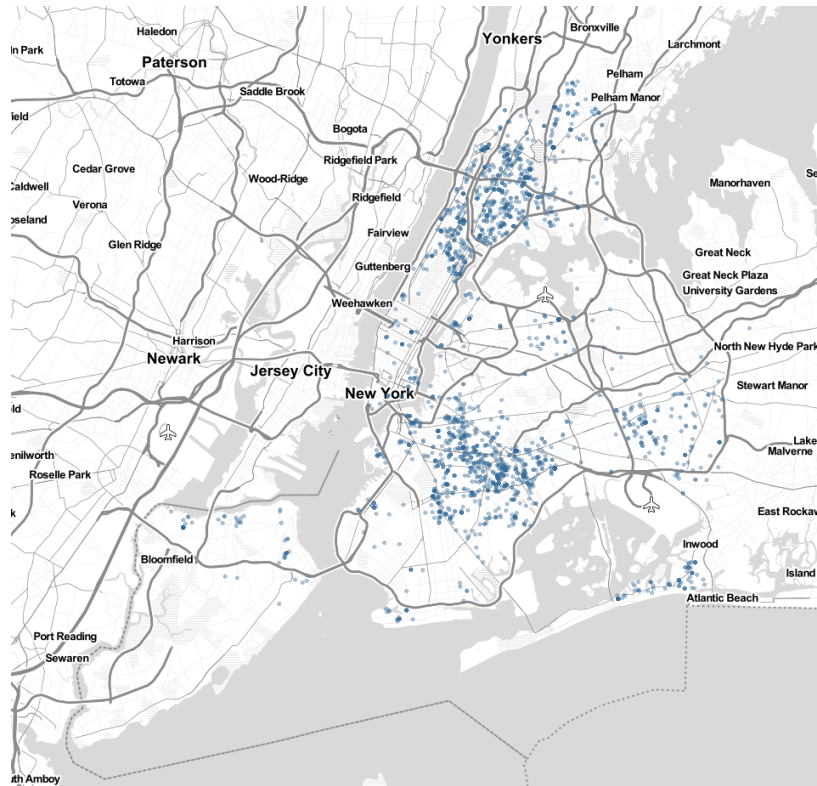
New York City Unemployment Data

The COVID-19 pandemic of 2020 led to massive unemployment across the USA. This next section examines if there is an association between monthly unemployment rates and shooting incidents.

Historical New York City monthly unemployment data are available at https://statistics.labor.ny.gov/lausCSV.asp?PASS=1&geog=21093561

Raw unemployment data structure

```r
url_in <-"https://statistics.labor.ny.gov/lausCSV.asp?PASS=1&geog=21093561"
empdata <- read.csv(url_in, skip=6, nrows=46)

# check out the data
str(empdata)
```

```
## 'data.frame':    46 obs. of  15 variables:
##  $ Year   : int  2021 2020 2019 2018 2017 2016 2015 2014 2013 2012 ...
##  $ Jan    : chr  "13.1%" "3.8%" "4.8%" "4.6%" ...
##  $ Feb    : chr  "13.2%" "3.8%" "4.4%" "4.5%" ...
##  $ Mar    : chr  "11.2%" "4.2%" "4.1%" "4.3%" ...
##  $ Apr    : chr  "10.8%" "15.5%" "3.6%" "3.9%" ...
##  $ May    : chr  "9.9%" "20.2%" "3.8%" "3.6%" ...
##  $ Jun    : chr  "10.1%" "18.7%" "3.8%" "4.2%" ...
##  $ Jul    : chr  "10.2%" "18.8%" "4.1%" "4.3%" ...
##  $ Aug    : chr  "9.8%" "14.9%" "4%" "4.3%" ...
##  $ Sep    : chr  "8.9%" "14.7%" "3.5%" "3.9%" ...
```

```
## $ Oct    : chr  "8.4%" "11.7%" "3.7%" "4%" ...
## $ Nov    : chr  "8%" "11.7%" "3.4%" "3.8%" ...
## $ Dec    : chr  "7.9%" "11.6%" "3%" "4%" ...
## $ Ann.Avg: chr  "" "12.3%" "3.9%" "4.1%" ...
## $ X      : logi  NA NA NA NA NA NA ...
```

Transform and Append Unemployment Data to Shooting Data

```
unemp <- empdata %>%
  select(Year, Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec) %>%
  mutate(Year=as.numeric(as.character(Year))) %>%
  pivot_longer(!Year, names_to = "Month",
               values_to = "unemp_rate") %>%
  mutate(MonNum=match(Month,month.abb)) %>%
  mutate(YEAR_MONTH = make_date(year=Year, month=MonNum, day="01")) %>%
  mutate(unemp_rate=as.numeric(sub("%", "", unemp_rate))) %>%
  select(YEAR_MONTH, unemp_rate)
#str(unemp)
#summary(unemp)
#head(unemp)
#tail(unemp)
#str(NY_Totals)
#summary(NY_Totals)


Monthly_Shooting_Unemployment <-  inner_join(unemp, NY_Totals, by = c("YEAR_MONTH" = "YEAR_MONTH"))
str(Monthly_Shooting_Unemployment)
```

```
## tibble [180 x 3] (S3: tbl_df/tbl/data.frame)
##  $ YEAR_MONTH: Date[1:180], format: "2020-01-01" "2020-02-01" ...
##  $ unemp_rate: num [1:180] 3.8 3.8 4.2 15.5 20.2 18.7 18.8 14.9 14.7 11.7 ...
##  $ INCIDENTS : int [1:180] 94 51 71 69 137 265 325 311 189 165 ...
```

```
summary(Monthly_Shooting_Unemployment)
```

```
##    YEAR_MONTH              unemp_rate        INCIDENTS
## Min.   :2006-01-01   Min.   : 3.000   Min.   : 41.00
## 1st Qu.:2009-09-23   1st Qu.: 4.775   1st Qu.: 92.75
## Median :2013-06-16   Median : 5.700   Median :119.50
## Mean   :2013-06-16   Mean   : 6.952   Mean   :131.03
## 3rd Qu.:2017-03-08   3rd Qu.: 9.200   3rd Qu.:167.00
## Max.   :2020-12-01   Max.   :20.200   Max.   :325.00
```

## Examine association between monthly unemployment rate and monthly shooting incidents.

```
#linear model
mod <-lm(INCIDENTS ~ unemp_rate, data=Monthly_Shooting_Unemployment)
summary(mod)
```

```
##
## Call:
## lm(formula = INCIDENTS ~ unemp_rate, data = Monthly_Shooting_Unemployment)
##
## Residuals:
##     Min      1Q   Median      3Q      Max
```

```
## -128.846   -34.419    -9.102    29.017   130.009
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    76.690      9.089   8.438 1.09e-14 ***
## unemp_rate      7.817      1.203   6.497 7.92e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 47.75 on 178 degrees of freedom
## Multiple R-squared:  0.1917, Adjusted R-squared:  0.1872
## F-statistic: 42.22 on 1 and 178 DF,  p-value: 7.921e-10
```

```r
# add the predicted value to the data
#Monthly_Shooting_Unemployment %>% mutate (pred=predict(mod))
Monthly_w_pred <- Monthly_Shooting_Unemployment %>% mutate(pred=predict(mod))

# plot model
Monthly_w_pred %>% ggplot() +
  geom_point(aes(x=unemp_rate, y=INCIDENTS),
             color="blue") +
  geom_point(aes(x=unemp_rate, y=pred), color="red")
```
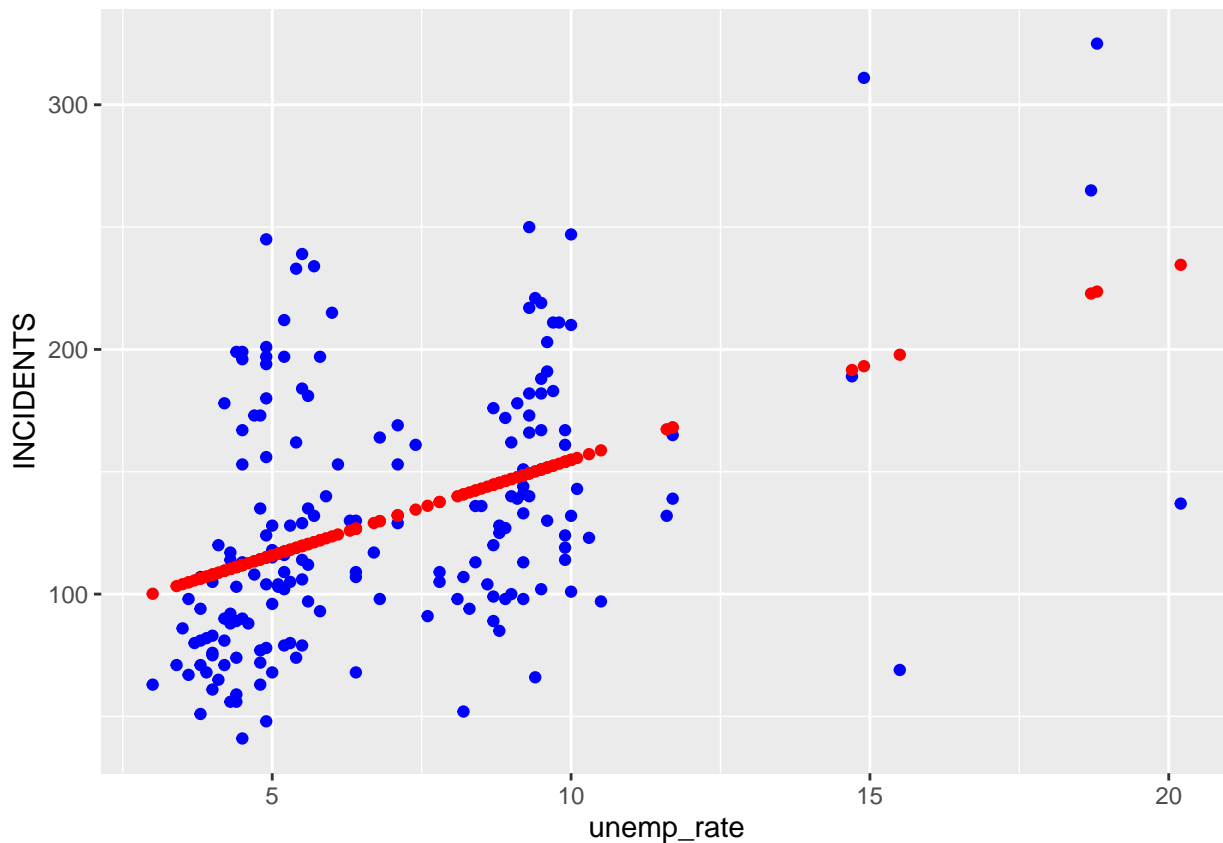


```r
#plot residuals
resid <-resid(mod)
plot(Monthly_Shooting_Unemployment$unemp_rate, resid,
     ylab="Residuals", xlab="Unemployment Rate",
     main="Shooting Incidents") +
```

```
abline(0, 0)
```

Shooting Incidents
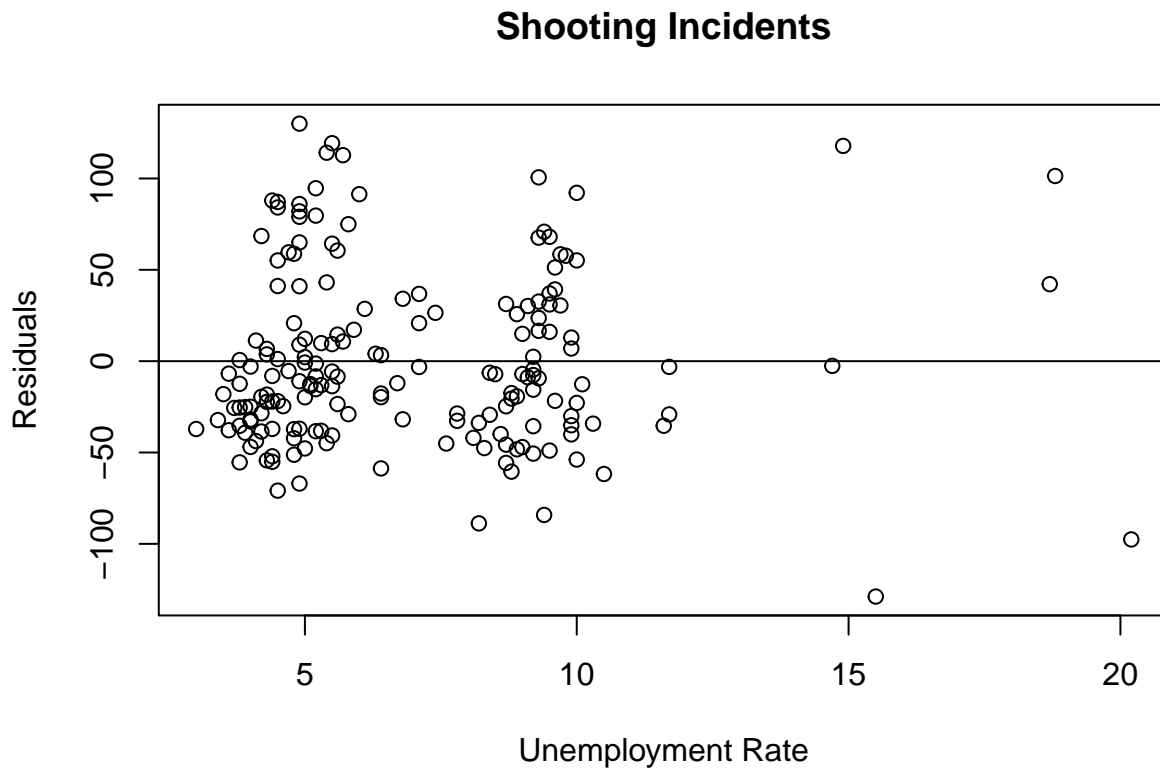

```
## integer(0)
```

Conclusion

New York City shooting incidents occur more often in Brooklyn and Bronx than other boroughs. More shooting incidents occur in summer months. The number of shooting incidents was lower between 2013-2019 compared to 2006-2012. There was an increase in shooting incidents in 2020. Shooting incidents are slightly associated with unemployment, but this does not explain most of the variation.

Other potential social and environmental factors associated with the COVID-19 pandemic should be considered, such as the school closures, reduction in availability of social services, and the effects of social isolation.

Missing Data and Bias

As mentioned at the start, this is not complete data on all shooting incidents in New York City. It excludes shooting incidents without injured victims.

The data are laid out so that each row represents one injured victim. There is an incident identifier (one incident may involve multiple injured victims). However, there is not an perpetrator identifier. Thus it is not possible to examine individual perpetrators.

There is also a high proportion of missing information on perpetrator age, sex, and race. It is possible the missing data are biased, as the collection method may affect the coding (e.g., self-report versus assumptions made by investigators). However, information on how the demographic data were collected is not immediately available.

My personal bias may be an inclination to distrust the veracity of shooting incident data provided by NYC. There is not much documentation on how the data are collected and verified. I tried to avoid this bias affecting my analysis by not making any statements that were not supported by the data.

R Session Info

```
sessionInfo()
```

```
## R version 4.1.2 (2021-11-01)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 21.10
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/blas/libblas.so.3.9.0
## LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.9.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] forcats_0.5.1   stringr_1.4.0   dplyr_1.0.8     purrr_0.3.4
##  [5] readr_2.1.2     tidyr_1.2.0     tibble_3.1.6    tidyverse_1.3.1
##  [9] readxl_1.3.1    lubridate_1.8.0 knitr_1.37      ggmap_3.0.0
## [13] ggplot2_3.3.3
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.6          lattice_0.20-45   png_0.1-7
##  [4] assertthat_0.2.1    digest_0.6.27     utf8_1.1.4
##  [7] R6_2.5.0            cellranger_1.1.0  plyr_1.8.6
## [10] backports_1.2.1     reprex_2.0.1      evaluate_0.14
## [13] highr_0.9           httr_1.4.2        pillar_1.7.0
## [16] RgoogleMaps_1.4.5.3 rlang_1.0.1       curl_4.3
## [19] rstudioapi_0.13     rmarkdown_2.11    labeling_0.4.2
## [22] munsell_0.5.0       broom_0.7.12      compiler_4.1.2
## [25] modelr_0.1.8        xfun_0.29         pkgconfig_2.0.3
## [28] htmltools_0.5.2     tidyselect_1.1.1  fansi_0.4.2
## [31] crayon_1.5.0        tzdb_0.2.0        dbplyr_2.1.1
## [34] withr_2.4.1         bitops_1.0-7      grid_4.1.2
## [37] jsonlite_1.7.2      gtable_0.3.0      lifecycle_1.0.1
## [40] DBI_1.1.2           magrittr_2.0.1    scales_1.1.1
## [43] cli_3.2.0           stringi_1.7.6     farver_2.0.3
## [46] fs_1.5.2            sp_1.4-6          xml2_1.3.3
## [49] ellipsis_0.3.2      generics_0.1.2    vctrs_0.3.8
## [52] rjson_0.2.21        tools_4.1.2       glue_1.6.1
## [55] hms_1.1.1           jpeg_0.1-9        fastmap_1.1.0
## [58] yaml_2.3.4          colorspace_2.0-0  rvest_1.0.2
## [61] haven_2.4.3
```