

NYPD Shooting Incidents Report

Sergey Ostrovsky

5/31/2021

Contents

1	Introduction	1
2	Importing	1
3	Tidying and Transforming NYPD Data	4
4	Visualizing Data	6
5	Analyzing NYPD Data	6
6	Modeling NYPD Data	7
7	Conclusion and Bias	9

1 Introduction

The data NYPD Shooting Incidents contains many exciting points to analyze the incidents based on location, region, race, or age. However, my interest in this report is to analyze the data based on political parties which are Republican or Democrats, economy, Covid-19, and presidential administration. The data contain the incident report from 2006 to 2020. Thus, based on my knowledge of the economy, presidential election, and covid-19 during these years, I would like to analyze the predominant factor which causes NYPD shooting incidents.

2 Importing

2.0.0.0.1 First, I will import the libraries to use for the report.

```
library(tidyverse)
```

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

```
## v ggplot2 3.3.3      v purrr   0.3.4
## v tibble  3.1.2      v dplyr   1.0.6
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   1.4.0      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

library(ggplot2)
options(width=60)
options(warn=-1)
```

2.0.0.0.2 Next, create a multiplot function for multiple charts.

```
multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL) {
  library(grid)

  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)

  numPlots = length(plots)

  # If layout is NULL, then use 'cols' to determine layout
  if (is.null(layout)) {
    # Make the panel
    # ncol: Number of columns of plots
    # nrow: Number of rows needed, calculated from # of cols
    layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
                      ncol = cols, nrow = ceiling(numPlots/cols))
  }

  if (numPlots==1) {
    print(plots[[1]])
  } else {
    # Set up the page
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))

    # Make each plot, in the correct location
    for (i in 1:numPlots) {
      # Get the i,j matrix positions of the regions that contain this subplot
      matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))

      print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
                                       layout.pos.col = matchidx$col))
    }
  }
}
```

2.0.0.0.3 Now I can load NYPD Data from <https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD> link.

```
url <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
nypd_shooting_incident <- read_csv(url)
```

```
##
## -- Column specification -----
## cols(
##   INCIDENT_KEY = col_double(),
##   OCCUR_DATE = col_character(),
##   OCCUR_TIME = col_time(format = ""),
##   BORO = col_character(),
##   PRECINCT = col_double(),
##   JURISDICTION_CODE = col_double(),
##   LOCATION_DESC = col_character(),
##   STATISTICAL_MURDER_FLAG = col_logical(),
##   PERP_AGE_GROUP = col_character(),
##   PERP_SEX = col_character(),
##   PERP_RACE = col_character(),
##   VIC_AGE_GROUP = col_character(),
##   VIC_SEX = col_character(),
##   VIC_RACE = col_character(),
##   X_COORD_CD = col_number(),
##   Y_COORD_CD = col_number(),
##   Latitude = col_double(),
##   Longitude = col_double(),
##   Lon_Lat = col_character()
## )
```

```
summary(nypd_shooting_incident)
```

```
##   INCIDENT_KEY      OCCUR_DATE      OCCUR_TIME
## Min.   : 9953245    Length:23568    Length:23568
## 1st Qu.: 55317014    Class :character Class1:hms
## Median : 83365370    Mode  :character Class2:difftime
## Mean   :102218616                      Mode  :numeric
## 3rd Qu.:150772442
## Max.   :222473262
##
##      BORO      PRECINCT      JURISDICTION_CODE
## Length:23568    Min.   : 1.00    Min.   :0.0000
## Class :character 1st Qu.: 44.00    1st Qu.:0.0000
## Mode  :character Median : 69.00    Median :0.0000
##                      Mean   : 66.21    Mean   :0.3323
##                      3rd Qu.: 81.00    3rd Qu.:0.0000
##                      Max.   :123.00    Max.   :2.0000
##                      NA's   :2
## LOCATION_DESC    STATISTICAL_MURDER_FLAG
## Length:23568      Mode :logical
## Class :character  FALSE:19080
## Mode  :character  TRUE :4488
##
```

```
##
##
##
## PERP_AGE_GROUP      PERP_SEX      PERP_RACE
## Length:23568      Length:23568      Length:23568
## Class :character   Class :character   Class :character
## Mode :character    Mode :character    Mode :character
##
##
##
## VIC_AGE_GROUP      VIC_SEX      VIC_RACE
## Length:23568      Length:23568      Length:23568
## Class :character   Class :character   Class :character
## Mode :character    Mode :character    Mode :character
##
##
##
## X_COORD_CD      Y_COORD_CD      Latitude
## Min. : 914928    Min. :125757    Min. :40.51
## 1st Qu.: 999900    1st Qu.:182565    1st Qu.:40.67
## Median :1007645    Median :193482    Median :40.70
## Mean :1009363      Mean :207312      Mean :40.74
## 3rd Qu.:1016807    3rd Qu.:239163    3rd Qu.:40.82
## Max. :1066815      Max. :271128      Max. :40.91
##
## Longitude      Lon_Lat
## Min. : -74.25    Length:23568
## 1st Qu.: -73.94    Class :character
## Median : -73.92    Mode :character
## Mean : -73.91
## 3rd Qu.: -73.88
## Max. : -73.70
##
##
```

2.0.0.0.4 Let's select only remarkable columns for this report.

```
nypd_shooting_incident <- nypd_shooting_incident %>%
  select(OCCUR_DATE,BORO,PERP_AGE_GROUP,PERP_RACE,
         VIC_AGE_GROUP,VIC_SEX,VIC_RACE)
```

3 Tidying and Transforming NYPD Data

3.0.0.0.1 Create YEAR and YEAR_MONTH columns for the analysis.

```
nypd_si_all <- nypd_shooting_incident %>%
  select(BORO, OCCUR_DATE) %>%
  mutate(OCCUR_DATE = mdy(OCCUR_DATE), INCIDENTS_all = 1,
         YEAR = format_ISO8601(OCCUR_DATE, precision = "y"),
         YEAR_MONTH = format_ISO8601(OCCUR_DATE, precision = "ym"))
```

3.0.0.0.2 Group and count NYPD Shooting Incidents by year.

```
nypd_si_all_global <- nypd_si_all %>%  
  select(YEAR, INCIDENTS_all) %>%  
  group_by(YEAR) %>%  
  summarise(INCIDENTS_all = sum(INCIDENTS_all), .groups = "keep") %>%  
  ungroup()
```

3.0.0.0.3 Group and count NYPD Shooting Incidents by month for 2018, 2019, and 2020.

```
nypd_si_all_global_monthly <- nypd_si_all %>%  
  select(YEAR_MONTH, INCIDENTS_all) %>%  
  group_by(YEAR_MONTH) %>%  
  summarise(INCIDENTS_all = sum(INCIDENTS_all), .groups = "keep") %>%  
  ungroup()
```

```
nypd_si_2020 <- nypd_shooting_incident %>%  
  select(BORO, OCCUR_DATE) %>%  
  mutate(OCCUR_DATE = mdy(OCCUR_DATE), INCIDENTS_2020 = 1,  
         MONTH = format(OCCUR_DATE, "%b")) %>%  
  filter(OCCUR_DATE >= as.Date("2020-01-01") & OCCUR_DATE <= as.Date("2020-12-31"))
```

```
nypd_si_2020_global <- nypd_si_2020 %>%  
  select(MONTH, INCIDENTS_2020) %>%  
  group_by(MONTH) %>%  
  summarise(INCIDENTS_2020 = sum(INCIDENTS_2020), .groups = "keep") %>%  
  ungroup()
```

```
nypd_si_2019 <- nypd_shooting_incident %>%  
  select(BORO, OCCUR_DATE) %>%  
  mutate(OCCUR_DATE = mdy(OCCUR_DATE), INCIDENTS_2019 = 1,  
         MONTH = format(OCCUR_DATE, "%b")) %>%  
  filter(OCCUR_DATE >= as.Date("2019-01-01") & OCCUR_DATE <= as.Date("2019-12-31"))
```

```
nypd_si_2019_global <- nypd_si_2019 %>%  
  select(MONTH, INCIDENTS_2019) %>%  
  group_by(MONTH) %>%  
  summarise(INCIDENTS_2019 = sum(INCIDENTS_2019), .groups = "keep") %>%  
  ungroup()
```

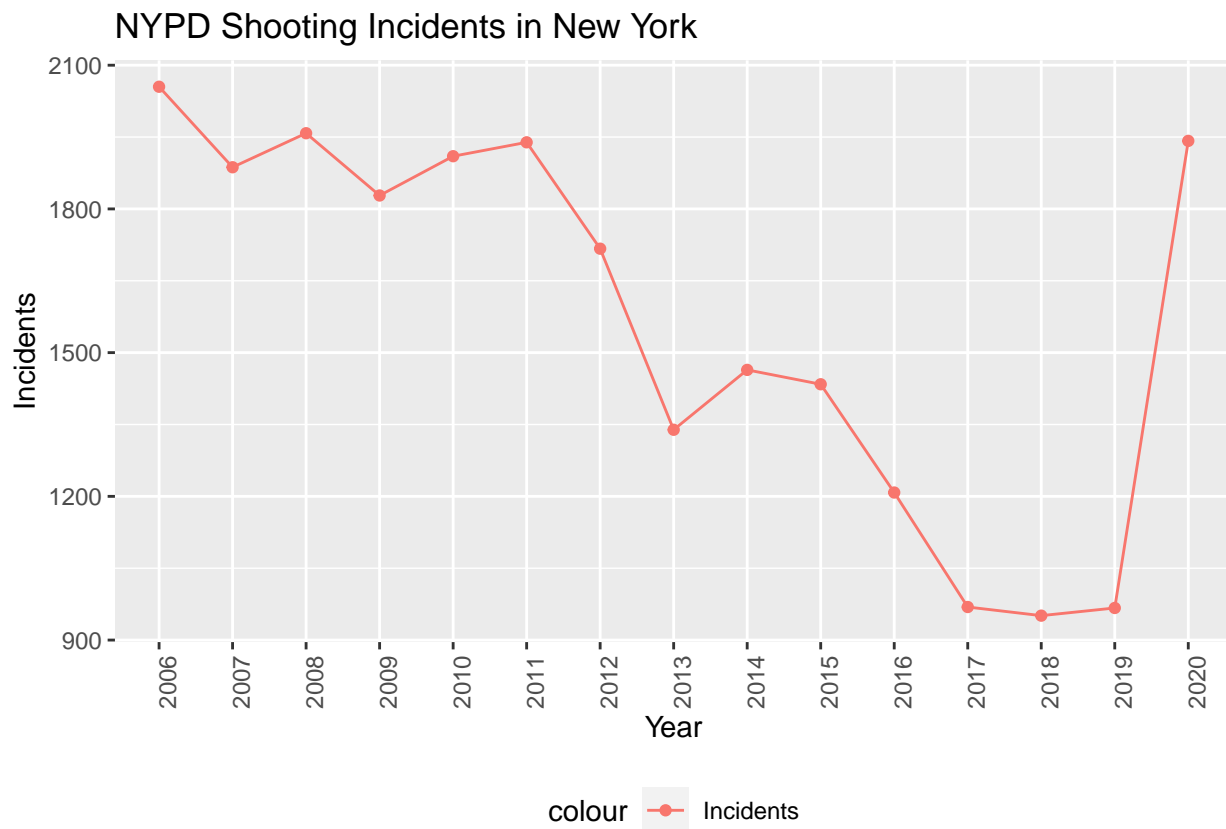
```
nypd_si_2018 <- nypd_shooting_incident %>%  
  select(BORO, OCCUR_DATE) %>%  
  mutate(OCCUR_DATE = mdy(OCCUR_DATE), INCIDENTS_2018 = 1,  
         MONTH = format(OCCUR_DATE, "%b")) %>%  
  filter(OCCUR_DATE >= as.Date("2018-01-01") & OCCUR_DATE <= as.Date("2018-12-31"))
```

```
nypd_si_2018_global <- nypd_si_2018 %>%  
  select(MONTH, INCIDENTS_2018) %>%  
  group_by(MONTH) %>%  
  summarise(INCIDENTS_2018 = sum(INCIDENTS_2018), .groups = "keep") %>%  
  ungroup()
```

4 Visualizing Data

4.0.0.0.1 Visualize NYPD Data that shows the number of incidents that occurred yearly.

```
nypd_si_all_global %>%  
  ggplot(aes(x = YEAR, y = INCIDENTS_all, group = 1)) +  
  geom_line(aes(color = "Incidents")) +  
  geom_point(aes(color = "Incidents")) +  
  theme(legend.position = "bottom",  
        axis.text.x = element_text(angle = 90)) +  
  labs(title = "NYPD Shooting Incidents in New York", y = "Incidents", x="Year")
```



5 Analyzing NYPD Data

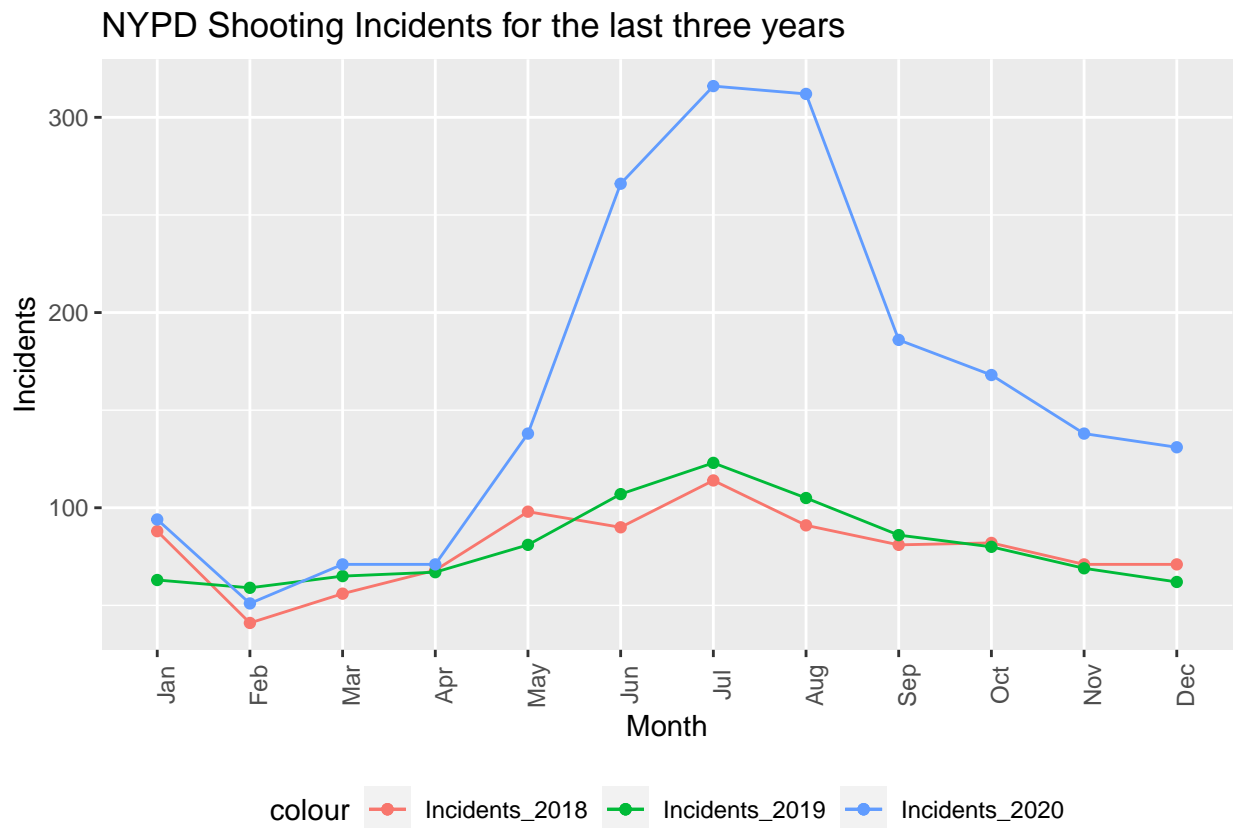
The graph above shows that the lowest NYPD Shooting Incidents were from 2016 to 2019, the Trump administration period until covid-19 came. After Covid-19, the economy became terrible, and the number of incidents became high. Thus we can see that the primary part in shooting incidents plays the status of the economy and not the political parties, which are either Republicans or Democrats.

5.0.0.0.1 For better analysis, let's see the monthly comparison for the last three years.

```
nypd_si_last_3_year <- full_join(nypd_si_2018_global,  
                                 nypd_si_2019_global, by="MONTH")
```

```
nypd_si_last_3_year <- full_join(nypd_si_last_3_year,
                                nypd_si_2020_global, by="MONTH")

level_order <- c('Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec')
nypd_si_last_3_year %>%
  ggplot(aes(x = factor(MONTH, level = level_order), y = INCIDENTS_2018, group = 1)) +
  geom_line(aes(color = "Incidents_2018")) +
  geom_point(aes(color = "Incidents_2018")) +
  geom_line(aes(y = INCIDENTS_2019, color = "Incidents_2019")) +
  geom_point(aes(y = INCIDENTS_2019, color = "Incidents_2019")) +
  geom_line(aes(y = INCIDENTS_2020, color = "Incidents_2020")) +
  geom_point(aes(y = INCIDENTS_2020, color = "Incidents_2020")) +
  theme(legend.position = "bottom",
        axis.text.x = element_text(angle = 90)) +
  labs(title = "NYPD Shooting Incidents for the last three years", y = "Incidents", x="Month")
```



The graph above shows that the number of NYPD shooting incidents started growing in May 2020. However, in September it moved lower but not up to the level of the previous two years.

6 Modeling NYPD Data

6.0.0.0.1 To see a better picture, I would like to create model which shows how the economy from 2006 to 2020 effect the shooting insidents

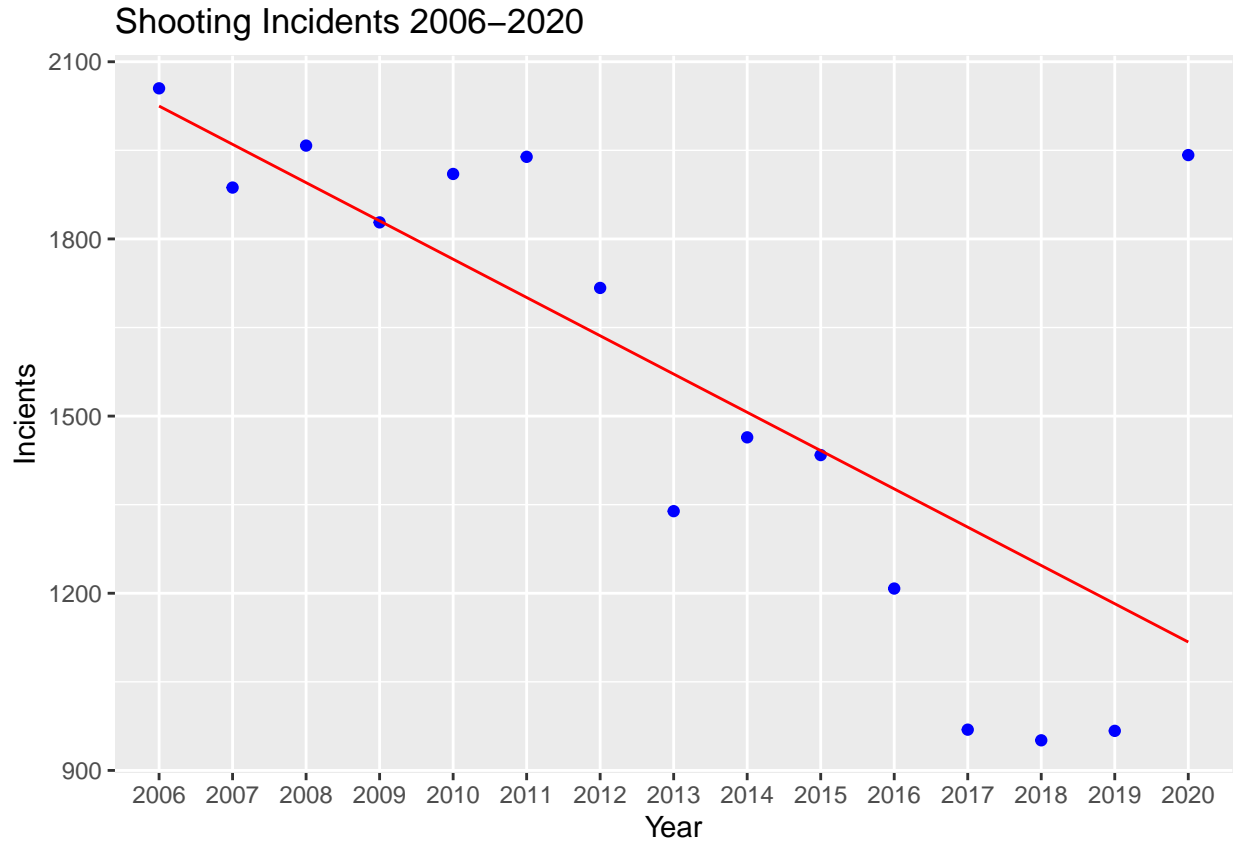
```

mod <- lm(INCIDENTS_all ~ as.integer(YEAR)-INCIDENTS_all, data = nypd_si_all_global)
summary(mod)

##
## Call:
## lm(formula = INCIDENTS_all ~ as.integer(YEAR) - INCIDENTS_all,
##     data = nypd_si_all_global)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -342.90 -191.99   -7.55    71.82   824.58
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    132063.93   35152.83     3.757  0.00240 **
## as.integer(YEAR)    -64.83     17.46    -3.712  0.00261 **
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 292.2 on 13 degrees of freedom
## Multiple R-squared:  0.5146, Adjusted R-squared:  0.4772
## F-statistic: 13.78 on 1 and 13 DF,  p-value: 0.002609

nypd_si_all_global_pred <- nypd_si_all_global %>%
  mutate(pred = predict(mod))
nypd_si_all_global_pred %>% ggplot(aes(x = YEAR, INCIDENTS_all, group = 1)) +
  geom_point(aes(y = INCIDENTS_all, x = YEAR),
             color = "blue") +
  geom_line(aes(y = pred, x = YEAR),
            color = "red") +
  labs(title = "Shooting Incidents 2006-2020", x = "Year", y="Incidents")

```

6.0.0.0.2 The graph above shows that predicted shooting incidents from 2006 to 2020 goes down because of the economy improvement.

7 Conclusion and Bias

The analysis above shows that the economy plays a primary role in shooting incidents. The other factors like President or Covid-19 can only affect the economy but not shooting incidents. The bias of this analysis can be that during the Trump administration, the economy was good before Covid-19 started, and the number of shooting incidents was low. However, when Covid-19 affected the economy, the number of shooting incidents became high again, and Trump could not handle it. So it is still unclear if Trump, who created a good economy or economy, started improving just during the second term of Obama and continue the improvement.