Analysis of NYDP Shooting Incident Data

Kodai.Fukuda

2023-07-24

Analyzing NYDP Shooting Incident Data

This report is about the data of gun shooting incidents in NYC, from 2006 to 2022. A csv file can be downloaded from https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD.

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

```
summary(NYPD.data)
```

```
##   INCIDENT_KEY        OCCUR_DATE         OCCUR_TIME            BORO
##  Min.   :  9953245   Length:27312       Length:27312       Length:27312
##  1st Qu.: 63860880   Class :character   Class :character   Class :character
##  Median : 90372218   Mode  :character   Mode  :character   Mode  :character
##  Mean   :120860536
##  3rd Qu.:188810230
##  Max.   :261190187
##
##  LOC_OF_OCCUR_DESC     PRECINCT      JURISDICTION_CODE LOC_CLASSFCTN_DESC
##  Length:27312       Min.   :  1.00   Min.   :0.0000    Length:27312
##  Class :character   1st Qu.: 44.00   1st Qu.:0.0000    Class :character
##  Mode  :character   Median : 68.00   Median :0.0000    Mode  :character
##                     Mean   : 65.64   Mean   :0.3269
##                     3rd Qu.: 81.00   3rd Qu.:0.0000
##                     Max.   :123.00   Max.   :2.0000
##                                      NA's   :2
##  LOCATION_DESC      STATISTICAL_MURDER_FLAG PERP_AGE_GROUP
##  Length:27312       Length:27312            Length:27312
##  Class :character   Class :character        Class :character
##  Mode  :character   Mode  :character        Mode  :character
##
##
##
##
##    PERP_SEX           PERP_RACE         VIC_AGE_GROUP        VIC_SEX
##  Length:27312       Length:27312       Length:27312       Length:27312
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
```

```
##
##    VIC_RACE           X_COORD_CD        Y_COORD_CD        Latitude
##  Length:27312       Min.   : 914928   Min.   :125757   Min.   :40.51
##  Class :character   1st Qu.:1000028   1st Qu.:182834   1st Qu.:40.67
##  Mode  :character   Median :1007731   Median :194487   Median :40.70
##                     Mean   :1009449   Mean   :208127   Mean   :40.74
##                     3rd Qu.:1016838   3rd Qu.:239518   3rd Qu.:40.82
##                     Max.   :1066815   Max.   :271128   Max.   :40.91
##                                                        NA's   :10
##    Longitude        Lon_Lat
##  Min.   :-74.25   Length:27312
##  1st Qu.:-73.94   Class :character
##  Median :-73.92   Mode  :character
##  Mean   :-73.91
##  3rd Qu.:-73.88
##  Max.   :-73.70
##  NA's   :10
```

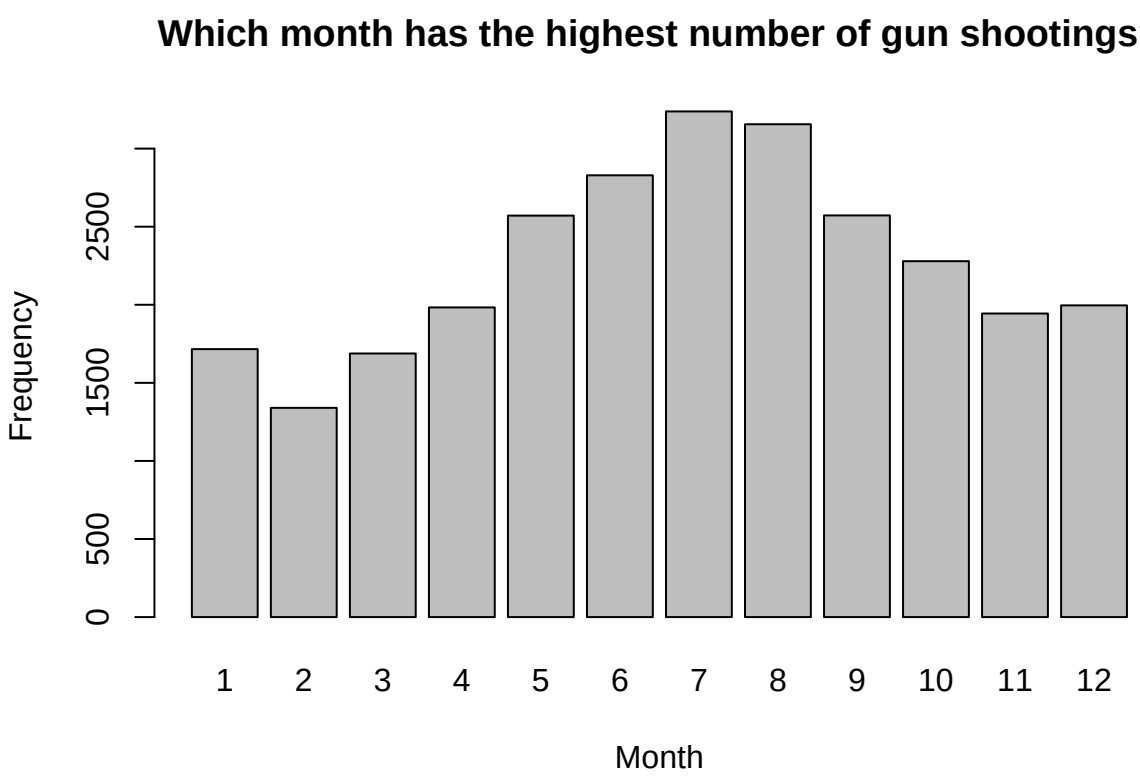
```
NYPD.datacleaning <- subset(NYPD.data, select = setdiff(colnames(NYPD.data), c('Latitude', 'Longitude',

library(lubridate)
NYPD.datacleaning$OCCUR_DATE <- mdy(NYPD.datacleaning$OCCUR_DATE)
NYPD.datacleaning$OCCUR_TIME <- hms(NYPD.datacleaning$OCCUR_TIME)
head(NYPD.datacleaning)
```

```
##   INCIDENT_KEY OCCUR_DATE OCCUR_TIME     BORO PRECINCT STATISTICAL_MURDER_FLAG
## 1    228798151 2021-05-27 21H 30M 0S   QUEENS      105                   false
## 2    137471050 2014-06-27 17H 40M 0S    BRONX       40                   false
## 3    147998800 2015-11-21  3H 56M 0S   QUEENS      108                    true
## 4    146837977 2015-10-09 18H 30M 0S    BRONX       44                   false
## 5     58921844 2009-02-19 22H 58M 0S    BRONX       47                    true
## 6    219559682 2020-10-21 21H 36M 0S BROOKLYN       81                    true
##   PERP_AGE_GROUP PERP_SEX PERP_RACE VIC_AGE_GROUP VIC_SEX       VIC_RACE
## 1                                          18-24       M          BLACK
## 2                                          18-24       M          BLACK
## 3                                          25-44       M          WHITE
## 4                                            <18       M WHITE HISPANIC
## 5          25-44        M     BLACK         45-64       M          BLACK
## 6                                          25-44       M          BLACK
```

Which month has the highest number of gun shootings?

```
occur_month <- month(NYPD.datacleaning$OCCUR_DATE)
occur_month.df <- data.frame(table(occur_month))
barplot(occur_month.df[,2],
        names.arg=c("1","2","3","4","5","6","7","8","9","10","11","12"),
     main = "Which month has the highest number of gun shootings?",
     xlab = "Month",
     ylab = "Frequency")
```

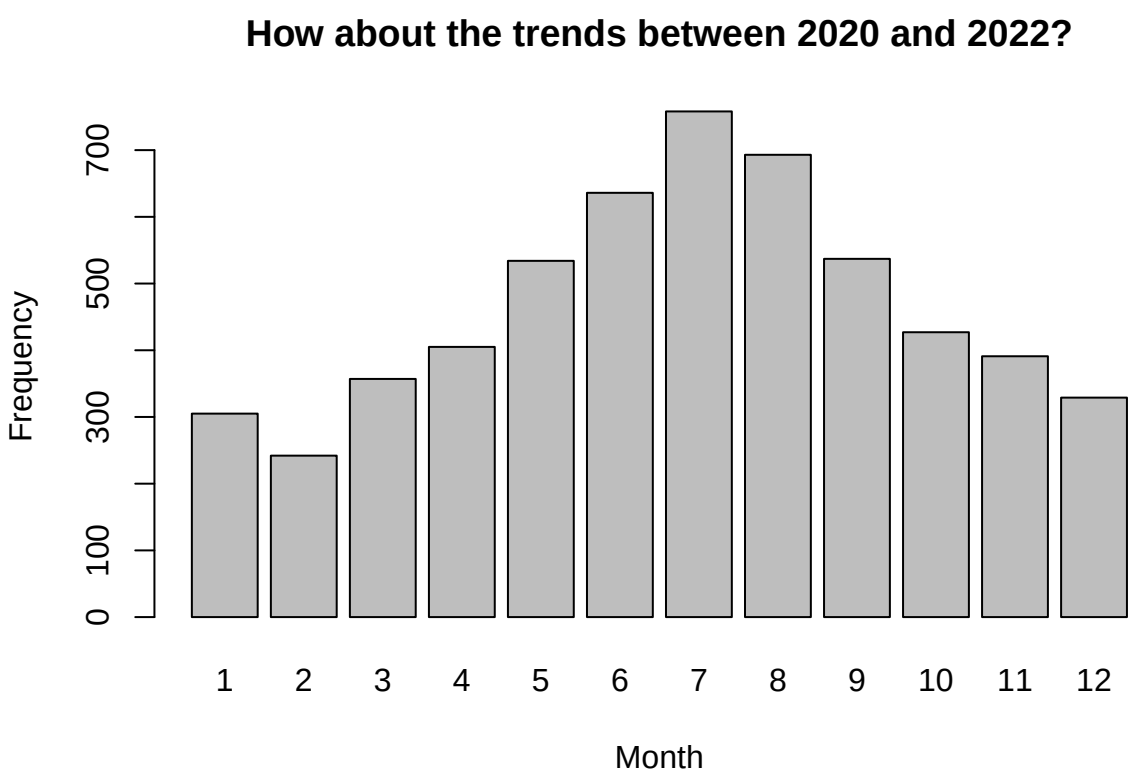


Comparing the data from 2006 to 2022 by month, it is clear that the number of gun shooting incident which occurred in NYC increases over the summer. The number reached its peak in July.

```
five_years <- NYPD.datacleaning %>%
  filter(between(
    OCCUR_DATE,
    ymd("2020-01-01"),
    ymd("2022-12-12")))

five_years.df <- data.frame(table(month(five_years$OCCUR_DATE)))

barplot(five_years.df[,2],
        names.arg=c("1","2","3","4","5","6","7","8","9","10","11","12"),
     main = "How about the trends between 2020 and 2022?",
     xlab = "Month",
     ylab = "Frequency")
```

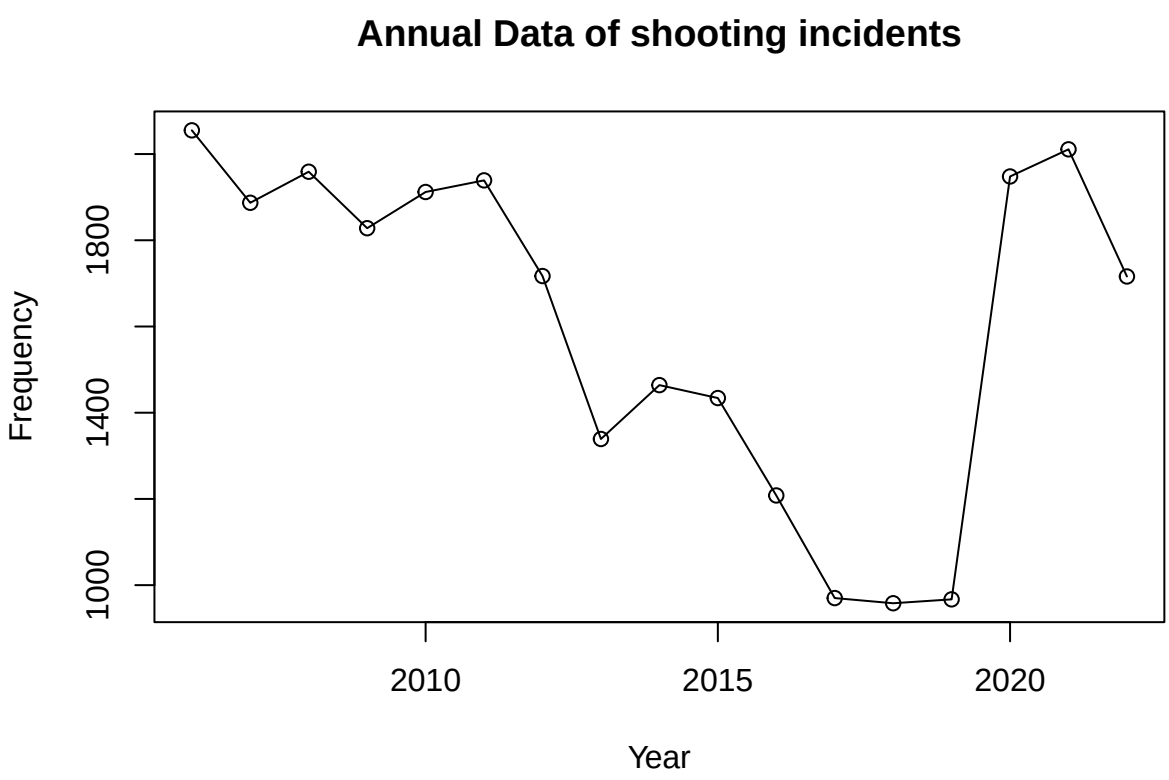


The trends have not changed in the last 3 years.

Trends in the number of gun shootings

```
occur_year <- year(NYPD.datacleaning$OCCUR_DATE)
occur_year.df = data.frame(table(occur_year))

plot(2006:2022, occur_year.df[,2], type="o",
     main = "Annual Data of shooting incidents",
     xlab = "Year",
     ylab = "Frequency")
```



Yearly data shows that the number of gun shootings had been gradually declining since 2006, but then spiked in 2020, returning to pre-2010 levels. This leads to the question of why the number suddenly jumped in 2020.

Conclusion and Possible Bias

First, the number of gun shootings has been increasing again in the last couple of years, so New York City needs to increase its budget and increase the number of police.

Second, however, it is not a dangerous city all year round. According to the data, gun shootings tend to occur in summer. It may be necessary to increase the number of police patrolling the streets from June through September.

A possible bias in this analysis was that the trend had changed over a period of about 15 years. The first analysis, which analyzed totals from 2006 to more recent years, found that gun shootings were more likely to occur in the summer. However, there was concern that the social context had changed over the past 15 years and that the results might change depending on the time period analyzed. Therefore, I focused on data from the last three years to see if the trend had changed. The results showed that the trend had not changed in the two periods, indicating that the analysis was valid.