Final Project 1 NYPD Shooting Incident Report

jz

2024-06-24

About NYPD Shooting Incident data

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. This data can be used by the public to explore the nature of shooting/criminal activity. Please refer to the attached data footnotes for additional information about this dataset. Also see https://data.cityofnewyork.us/Public-Safety/NYPD-Shooting-Incident-Data-Historic-/833y-fsy8/about_data

This report is mainly focus on yearly trend and quarterly total incidents happened in NYC. It contains a model to find out a likelihood of different factors affecting shooting result. In order to achieve reproducibility, this report shows each step including how to import, tidy and analyze the data.

Step 0: Import Library

```
# install.packages("tidyverse")
library(tidyverse)
library(lubridate)
```

Step 1: Importing Data

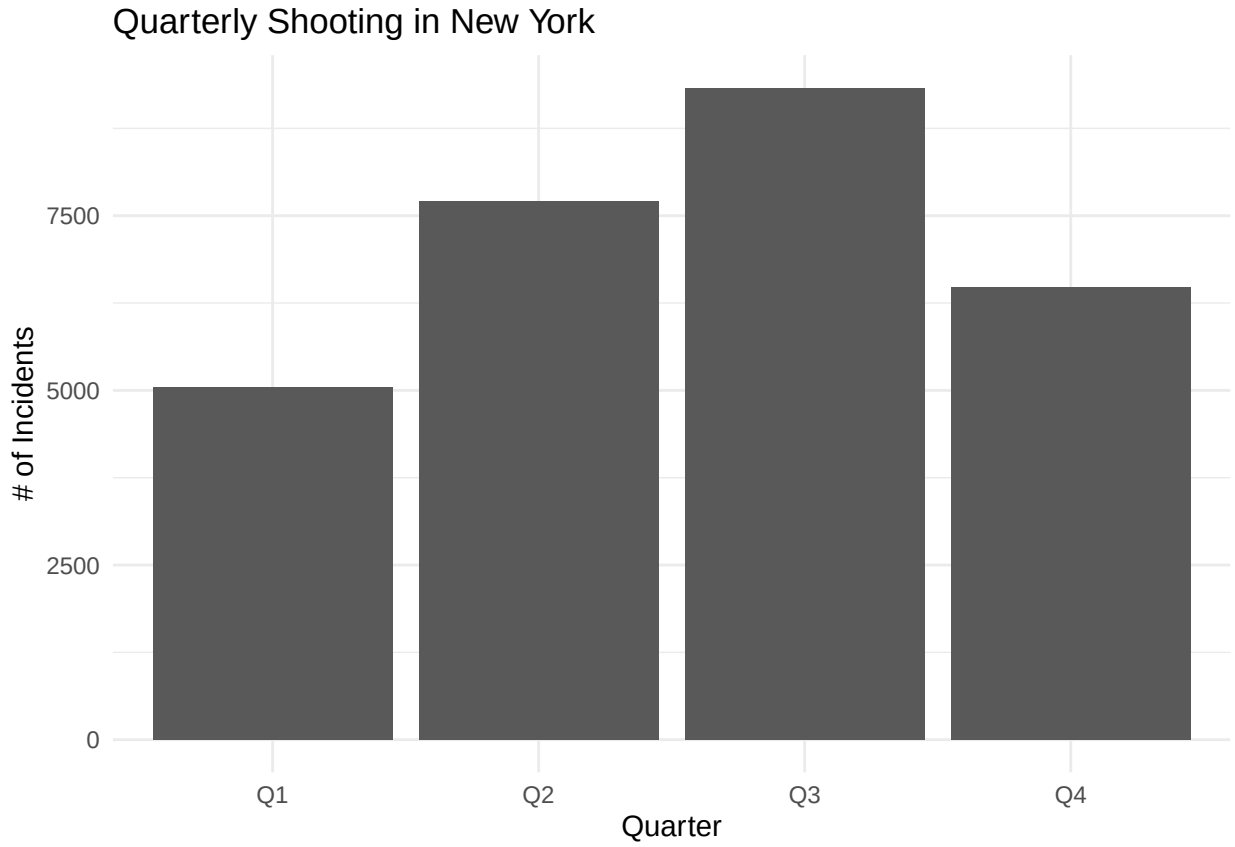
```
## read_csv() reads comma delimited files, message=FALSE, warning=FALSE}
url_in <-"https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
NYPDdf      <-read_csv(url_in)
```

```
## Rows: 28562 Columns: 21
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (12): OCCUR_DATE, BORO, LOC_OF_OCCUR_DESC, LOC_CLASSFCTN_DESC, LOCATION...
## dbl   (7): INCIDENT_KEY, PRECINCT, JURISDICTION_CODE, X_COORD_CD, Y_COORD_CD...
## lgl   (1): STATISTICAL_MURDER_FLAG
## time  (1): OCCUR_TIME
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Step 2: Transform and visualize Data

Step 2.1: Transform and vdata to quarterly

```
NYPDdf$OCCUR_DAY = mdy(NYPDdf$OCCUR_DATE)
NYPDdf$OCCUR_Quarter = quarters(NYPDdf$OCCUR_DAY)

NYPD_by_quarter <-NYPDdf %>%
      group_by ( OCCUR_Quarter ) %>%
      count()

g <- ggplot(NYPD_by_quarter , aes(x=OCCUR_Quarter,y=n)) +
   geom_col() +
  labs(title = "Quarterly Shooting in New York",
       x = "Quarter",
       y = "# of Incidents") +
  theme_minimal()
g
```

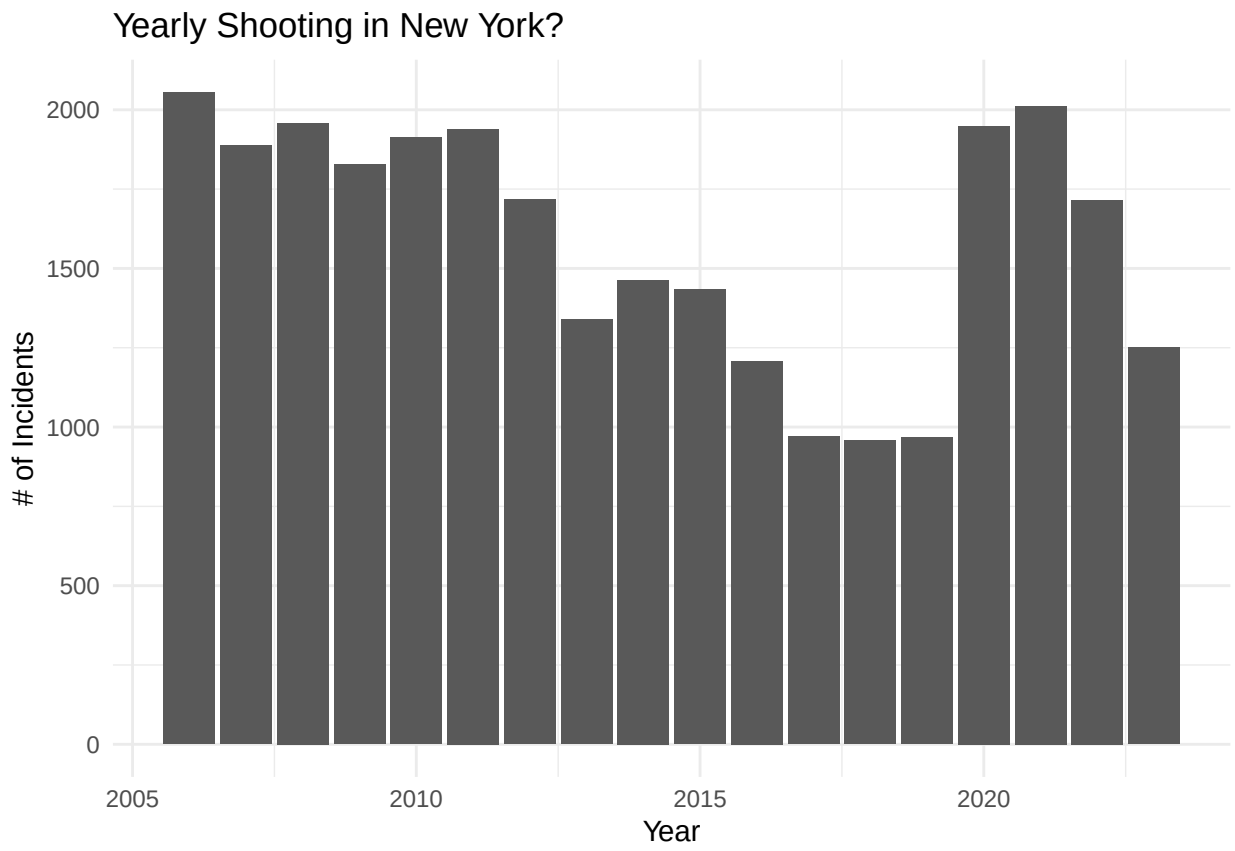


Step 2.2: Visualize yearly data

```r
NYPDdf$OCCUR_DAY = mdy(NYPDdf$OCCUR_DATE)
NYPDdf$OCCUR_Year = year(NYPDdf$OCCUR_DAY)

NYPD_by_year <-NYPDdf %>%
      group_by ( OCCUR_Year) %>%
      count()


tail(NYPDdf)
```

```
## # A tibble: 6 x 24
##   INCIDENT_KEY OCCUR_DATE OCCUR_TIME BORO      LOC_OF_OCCUR_DESC PRECINCT
##          <dbl> <chr>      <time>     <chr>     <chr>                <dbl>
## 1    270719378 07/02/2023 21:40      BRONX     OUTSIDE                 46
## 2    265354835 03/19/2023 23:48      BRONX     INSIDE                  47
## 3    272968931 08/16/2023 02:46      BRONX     OUTSIDE                 41
## 4    270489846 06/27/2023 12:27      BRONX     INSIDE                  41
## 5    271021661 07/08/2023 11:27      QUEENS    OUTSIDE                102
## 6    271818283 07/24/2023 23:38      MANHATTAN OUTSIDE                 28
## # i 18 more variables: JURISDICTION_CODE <dbl>, LOC_CLASSFCTN_DESC <chr>,
## #   LOCATION_DESC <chr>, STATISTICAL_MURDER_FLAG <lgl>, PERP_AGE_GROUP <chr>,
## #   PERP_SEX <chr>, PERP_RACE <chr>, VIC_AGE_GROUP <chr>, VIC_SEX <chr>,
## #   VIC_RACE <chr>, X_COORD_CD <dbl>, Y_COORD_CD <dbl>, Latitude <dbl>,
## #   Longitude <dbl>, Lon_Lat <chr>, OCCUR_DAY <date>, OCCUR_Quarter <chr>,
## #   OCCUR_Year <dbl>
```

```r
g <- ggplot(NYPD_by_year , aes(x=OCCUR_Year,y=n)) +
   geom_col() +
  labs(title = "Yearly Shooting in New York?",
       x = "Year",
       y = "# of Incidents") +
  theme_minimal()

g
```



Step 3: Modling NYPD shooting data

Utilizing logistic regression model, the independent variables will be VIC_AGE_GROUP, VIC_SEX, VIC_RACE and BORO and dependent variable will be STATISTICAL_MURDER_FLAG. This can get an estimate for how much the age, gender, and race and boro of the victims will affect the shooting result.

```
mod <- glm(STATISTICAL_MURDER_FLAG ~ PERP_RACE + PERP_SEX + PERP_AGE_GROUP + BORO, data = NYPDdf, family
summary(mod)
```

```
##
## Call:
## glm(formula = STATISTICAL_MURDER_FLAG ~ PERP_RACE + PERP_SEX +
##     PERP_AGE_GROUP + BORO, family = binomial, data = NYPDdf)
##
## Coefficients: (2 not defined because of singularities)
##                                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)                             -1.67807    0.08830 -19.005  < 2e-16
## PERP_RACEAMERICAN INDIAN/ALASKAN NATIVE -11.92165  229.62616  -0.052  0.95859
## PERP_RACEASIAN / PACIFIC ISLANDER        -0.03129    0.30669  -0.102  0.91874
## PERP_RACEBLACK                           -0.44169    0.25501  -1.732  0.08327
## PERP_RACEBLACK HISPANIC                  -0.55671    0.26294  -2.117  0.03424
## PERP_RACEUNKNOWN                         -1.10382    0.14077  -7.841 4.47e-15
## PERP_RACEWHITE                            0.10654    0.28345   0.376  0.70702
## PERP_RACEWHITE HISPANIC                  -0.32938    0.25879  -1.273  0.20310
```

```
## PERP_SEXF                                -1.53727    0.28434  -5.406 6.43e-08
## PERP_SEXM                                -1.71400    0.26209  -6.540 6.16e-11
## PERP_SEXU                                      NA         NA      NA       NA
## PERP_AGE_GROUP<18                         2.37968    0.17915  13.283  < 2e-16
## PERP_AGE_GROUP1020                       -8.73230  324.74374  -0.027  0.97855
## PERP_AGE_GROUP1028                       -8.57449  324.74375  -0.026  0.97894
## PERP_AGE_GROUP18-24                       2.57397    0.17019  15.124  < 2e-16
## PERP_AGE_GROUP224                        -8.84460  324.74374  -0.027  0.97827
## PERP_AGE_GROUP25-44                       2.88148    0.17018  16.932  < 2e-16
## PERP_AGE_GROUP45-64                       3.24297    0.18602  17.433  < 2e-16
## PERP_AGE_GROUP65+                         3.30755    0.30908  10.701  < 2e-16
## PERP_AGE_GROUP940                        -8.73791  324.74374  -0.027  0.97853
## PERP_AGE_GROUPUNKNOWN                          NA         NA      NA       NA
## BOROBROOKLYN                             -0.10669    0.04643  -2.298  0.02158
## BOROMANHATTAN                            -0.15367    0.05950  -2.583  0.00981
## BOROQUEENS                               -0.13681    0.05886  -2.324  0.02010
## BOROSTATEN ISLAND                        -0.15780    0.10277  -1.536  0.12466
##
## (Intercept)                             ***
## PERP_RACEAMERICAN INDIAN/ALASKAN NATIVE
## PERP_RACEASIAN / PACIFIC ISLANDER
## PERP_RACEBLACK                          .
## PERP_RACEBLACK HISPANIC                 *
## PERP_RACEUNKNOWN                        ***
## PERP_RACEWHITE
## PERP_RACEWHITE HISPANIC
## PERP_SEXF                               ***
## PERP_SEXM                               ***
## PERP_SEXU
## PERP_AGE_GROUP<18                       ***
## PERP_AGE_GROUP1020
## PERP_AGE_GROUP1028
## PERP_AGE_GROUP18-24                     ***
## PERP_AGE_GROUP224
## PERP_AGE_GROUP25-44                     ***
## PERP_AGE_GROUP45-64                     ***
## PERP_AGE_GROUP65+                       ***
## PERP_AGE_GROUP940
## PERP_AGE_GROUPUNKNOWN
## BOROBROOKLYN                            *
## BOROMANHATTAN                           **
## BOROQUEENS                              *
## BOROSTATEN ISLAND
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 19168  on 19217  degrees of freedom
## Residual deviance: 18020  on 19195  degrees of freedom
##   (9344 observations deleted due to missingness)
## AIC: 18066
##
## Number of Fisher Scoring iterations: 11
```

Step 4: Identify Bias

I would like to address two areas of biases. One is the data. Another is interpretation of the data. From NY City Controller site https://comptroller.nyc.gov/newsroom/nypds-shotspotter-gunshot-detection-system-overwhelmingly-sends-officers-to-locations-where-no-confirmed-shooting-occurred-new-audit-uncovers/, Comptroller Brad Lander stated. "The evidence shows that NYPD is wasting precious time and money on this technology and needs to do a better job managing its resources. Chasing down car backfires and construction noise does not make us safer." which helps to understand the nature of criminal activity and shooting and limitation of ShotSpotter and much more stated in the article.

Another is the analysis focus of this report general trend yearly and quarterly. The analysis is totally based on the data.
I have also viewed different articles about the same subject and tried to eliminate anything that is from my personal experience.

Additional Resources

- Comparing different models to forecast the number of mass shootings in the United States: An application of forecasting rare event time series data
- Most alerts from the NYPD's gunfire detection system are unconfirmed shootings, city audit finds