NYPD Shooting Report

J. Achalapong

2025-06-01

Introduction

Project Purpose

This project is an assignment for the course DTSA 5301: Data Science as a Field. We are demonstrating our ability to complete all steps in the data science process by producing a report on the NYPD Shooting Incident data.

Question of Interest

How can data from past shooting incidents help us decide where and when to deploy police resources most effectively to reduce gun violence in New York City?

Project Step 1: Describe and Import the Dataset

Data Description

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 NYPD Shooting Incident Data (Historic) - CKAN for additional information about this dataset.

Row Description

- Each row in this dataset is a shooting incident.

Column Description

- INCIDENT_KEY: Randomly generated persistent ID for each arrest
- OCCUR_DATE: Exact date of shooting incident
- OCCUR_TIME: Exact time of the shooting incident
- BORO: Borough where the shooting incident occurred

- STATISTICAL_MURDER_FLAG: Shooting resulted in the victim's death which would be counted as a murder
- PERP_AGE_GROUP: Perpetrator's age within a category
- PERP_SEX: Perpetrator's sex description
- PERP_RACE: Perpetrator's race description
- VIC_AGE_GROUP: Victim's age within a category
- VIC_SEX: Victim's sex description
- VIC_RACE: Victim's race description

Import Libraries

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

Import Dataset

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

```
## Rows: 29,744
## Columns: 21
## $ INCIDENT_KEY            <int> 231974218, 177934247, 255028563, 25384540, 726~
## $ OCCUR_DATE              <chr> "08/09/2021", "04/07/2018", "12/02/2022", "11/~
## $ OCCUR_TIME              <chr> "01:06:00", "19:48:00", "22:57:00", "01:50:00"~
## $ BORO                    <chr> "BRONX", "BROOKLYN", "BRONX", "BROOKLYN", "BRO~
## $ LOC_OF_OCCUR_DESC       <chr> "", "", "OUTSIDE", "", "", "", "", "", "", "",~
## $ PRECINCT                <int> 40, 79, 47, 66, 46, 42, 71, 69, 75, 69, 40, 42~
## $ JURISDICTION_CODE       <int> 0, 0, 0, 0, 0, 2, 0, 2, 0, 0, 0, 2, 0, 0, 2, 0~
## $ LOC_CLASSFCTN_DESC      <chr> "", "", "STREET", "", "", "", "", "", "", "", ~
## $ LOCATION_DESC           <chr> "", "", "GROCERY/BODEGA", "PVT HOUSE", "MULTI ~
## $ STATISTICAL_MURDER_FLAG <chr> "false", "true", "false", "true", "true", "fal~
## $ PERP_AGE_GROUP          <chr> "", "25-44", "(null)", "UNKNOWN", "25-44", "18~
## $ PERP_SEX                <chr> "", "M", "(null)", "U", "M", "M", "", "", "M",~
## $ PERP_RACE               <chr> "", "WHITE HISPANIC", "(null)", "UNKNOWN", "BL~
## $ VIC_AGE_GROUP           <chr> "18-24", "25-44", "25-44", "18-24", "<18", "18~
## $ VIC_SEX                 <chr> "M", "M", "M", "M", "F", "M", "M", "M", "M", "~
## $ VIC_RACE                <chr> "BLACK", "BLACK", "BLACK", "BLACK", "BLACK", "~
## $ X_COORD_CD              <chr> "1006343", "1000082.937500000000000", "1020691~
## $ Y_COORD_CD              <chr> "234270", "189064.671875000000000", "257125", ~
## $ Latitude                <dbl> 40.80967, 40.68561, 40.87235, 40.64249, 40.845~
## $ Longitude               <dbl> -73.92019, -73.94291, -73.86823, -73.99691, -7~
## $ Lon_Lat                 <chr> "POINT (-73.92019278899994 40.80967347200004)"~
```

Step 2: Tidy and Transform Data

Remove Unnecessary Columns

The following columns are not needed for this assignment:
PRECINCT, JURISDICTION_CODE, LOCATION_DESC, X_COORD_CD, Y_COORD_CD, Lon_Lat

```
nypd_shooting <- nypd_shooting %>%
  select(-c(PRECINCT, JURISDICTION_CODE, LOCATION_DESC, X_COORD_CD, Y_COORD_CD, Lon_Lat)) %>%
  mutate(OCCUR_DATE = mdy(OCCUR_DATE),
         OCCUR_TIME = hms(OCCUR_TIME),
         Shootings = 1,
         OCCUR_YEAR = year(OCCUR_DATE),
         OCCUR_MONTH = month(OCCUR_DATE, label = TRUE, abbr = TRUE),
         OCCUR_WDAY = weekdays(OCCUR_DATE),
         OCCUR_HOUR = hour(hms(OCCUR_TIME)))
```

Replace missing and Remove extreme values in data

```
nypd_shooting = nypd_shooting %>%
  replace_na(list(PERP_AGE_GROUP = "Unknown", PERP_SEX = "Unknown", PERP_RACE = "Unknown"))

nypd_shooting = subset(nypd_shooting,
                       PERP_AGE_GROUP!="224"
                       & PERP_AGE_GROUP!="940"
                       & PERP_AGE_GROUP!="1020"
                       & PERP_AGE_GROUP!="1028"
                       & PERP_AGE_GROUP!="2021")
```

Convert Data Types

```
nypd_shooting$PERP_AGE_GROUP = recode(nypd_shooting$PERP_AGE_GROUP, UNKNOWN = "Unknown")
nypd_shooting$PERP_SEX = recode(nypd_shooting$PERP_SEX, U = "Unknown")
nypd_shooting$PERP_RACE = recode(nypd_shooting$PERP_RACE, UNKNOWN = "Unknown")
nypd_shooting$VIC_SEX   = recode(nypd_shooting$VIC_SEX, U = "Unknown")
nypd_shooting$VIC_RACE   = recode(nypd_shooting$VIC_RACE, UNKNOWN = "Unknown")
nypd_shooting$INCIDENT_KEY = as.character(nypd_shooting$INCIDENT_KEY)
nypd_shooting$BORO = as.factor(nypd_shooting$BORO)
nypd_shooting$PERP_AGE_GROUP = as.factor(nypd_shooting$PERP_AGE_GROUP)
nypd_shooting$PERP_SEX = as.factor(nypd_shooting$PERP_SEX)
nypd_shooting$PERP_RACE = as.factor(nypd_shooting$PERP_RACE)
nypd_shooting$VIC_AGE_GROUP = as.factor(nypd_shooting$VIC_AGE_GROUP)
nypd_shooting$VIC_SEX = as.factor(nypd_shooting$VIC_SEX)
nypd_shooting$VIC_RACE = as.factor(nypd_shooting$VIC_RACE)
nypd_shooting$STATISTICAL_MURDER_FLAG <- as.factor(nypd_shooting$STATISTICAL_MURDER_FLAG)
```

Summary of Data (Descriptive Statistics)

```r
# Descriptive statistics.
summary(nypd_shooting)
```

```
##  INCIDENT_KEY          OCCUR_DATE            OCCUR_TIME
##  Length:29739       Min.   :2006-01-01   Min.    :0S
##  Class :character   1st Qu.:2009-10-28   1st Qu.:3H 30M 30S
##  Mode  :character   Median :2014-03-25   Median :15H 15M 0S
##                     Mean   :2014-10-31   Mean   :12H 46M 10.9896096035518S
##                     3rd Qu.:2020-06-29   3rd Qu.:20H 44M 0S
##                     Max.   :2024-12-31   Max.   :23H 59M 0S
##
##            BORO        LOC_OF_OCCUR_DESC  LOC_CLASSFCTN_DESC
##  BRONX        : 8832   Length:29739       Length:29739
##  BROOKLYN     :11683   Class :character   Class :character
##  MANHATTAN    : 3977   Mode  :character   Mode  :character
##  QUEENS       : 4426
##  STATEN ISLAND:  821
##
##
##  STATISTICAL_MURDER_FLAG PERP_AGE_GROUP    PERP_SEX              PERP_RACE
##  false:23974                    :9344           : 9310   BLACK          :12320
##  true : 5765             18-24  :6630   (null) : 1628                   : 9310
##                          25-44  :6342   F      :  461   WHITE HISPANIC: 2665
##                          Unknown:3148   M      :16840   Unknown       : 1838
##                          <18    :1805   Unknown: 1500   (null)        : 1628
##                          (null) :1628                   BLACK HISPANIC: 1487
##                          (Other): 842                   (Other)       :  491
##  VIC_AGE_GROUP     VIC_SEX                            VIC_RACE
##  <18    : 3081   F      : 2891   AMERICAN INDIAN/ALASKAN NATIVE:   13
##  1022   :    1   M      :26836   ASIAN / PACIFIC ISLANDER      :  478
##  18-24  :10675   Unknown:   12   BLACK                         :20996
##  25-44  :13560                   BLACK HISPANIC                : 2930
##  45-64  : 2118                   Unknown                       :   72
##  65+    :  236                   WHITE                         :  741
##  UNKNOWN:   68                   WHITE HISPANIC                : 4509
##     Latitude       Longitude        Shootings    OCCUR_YEAR     OCCUR_MONTH
##  Min.   :40.51   Min.   :-74.25   Min.   :1   Min.   :2006   Jul    : 3513
##  1st Qu.:40.67   1st Qu.:-73.94   1st Qu.:1   1st Qu.:2009   Aug    : 3352
##  Median :40.70   Median :-73.91   Median :1   Median :2014   Jun    : 3091
##  Mean   :40.74   Mean   :-73.91   Mean   :1   Mean   :2014   Sep    : 2808
##  3rd Qu.:40.83   3rd Qu.:-73.88   3rd Qu.:1   3rd Qu.:2020   May    : 2794
##  Max.   :40.91   Max.   :-73.70   Max.   :1   Max.   :2024   Oct    : 2483
##  NA's   :96      NA's   :96                                  (Other):11698
##   OCCUR_WDAY          OCCUR_HOUR
##  Length:29739       Min.   : 1.00
##  Class :character   1st Qu.: 4.00
##  Mode  :character   Median :16.00
##                     Mean   :13.35
##                     3rd Qu.:21.00
##                     Max.   :23.00
##                     NA's   :2337
```

Missing Data

```r
# Identify columns with missing data and display the number of missing values per column.
colSums(is.na(nypd_shooting))
```

```
##            INCIDENT_KEY             OCCUR_DATE             OCCUR_TIME
##                       0                      0                      0
##                    BORO      LOC_OF_OCCUR_DESC      LOC_CLASSFCTN_DESC
##                       0                      0                      0
## STATISTICAL_MURDER_FLAG         PERP_AGE_GROUP               PERP_SEX
##                       0                      0                      0
##               PERP_RACE          VIC_AGE_GROUP                VIC_SEX
##                       0                      0                      0
##                VIC_RACE               Latitude              Longitude
##                       0                     96                     96
##               Shootings             OCCUR_YEAR            OCCUR_MONTH
##                       0                      0                      0
##              OCCUR_WDAY             OCCUR_HOUR
##                       0                   2337
```

```r
# Total number of missing values.
sum(is.na(nypd_shooting))
```

```
## [1] 2529
```

```r
# Percentage of missing values.
mean(is.na(nypd_shooting))
```

```
## [1] 0.004251992
```

Step 3: Add Visualizations and Analysis

Research Question

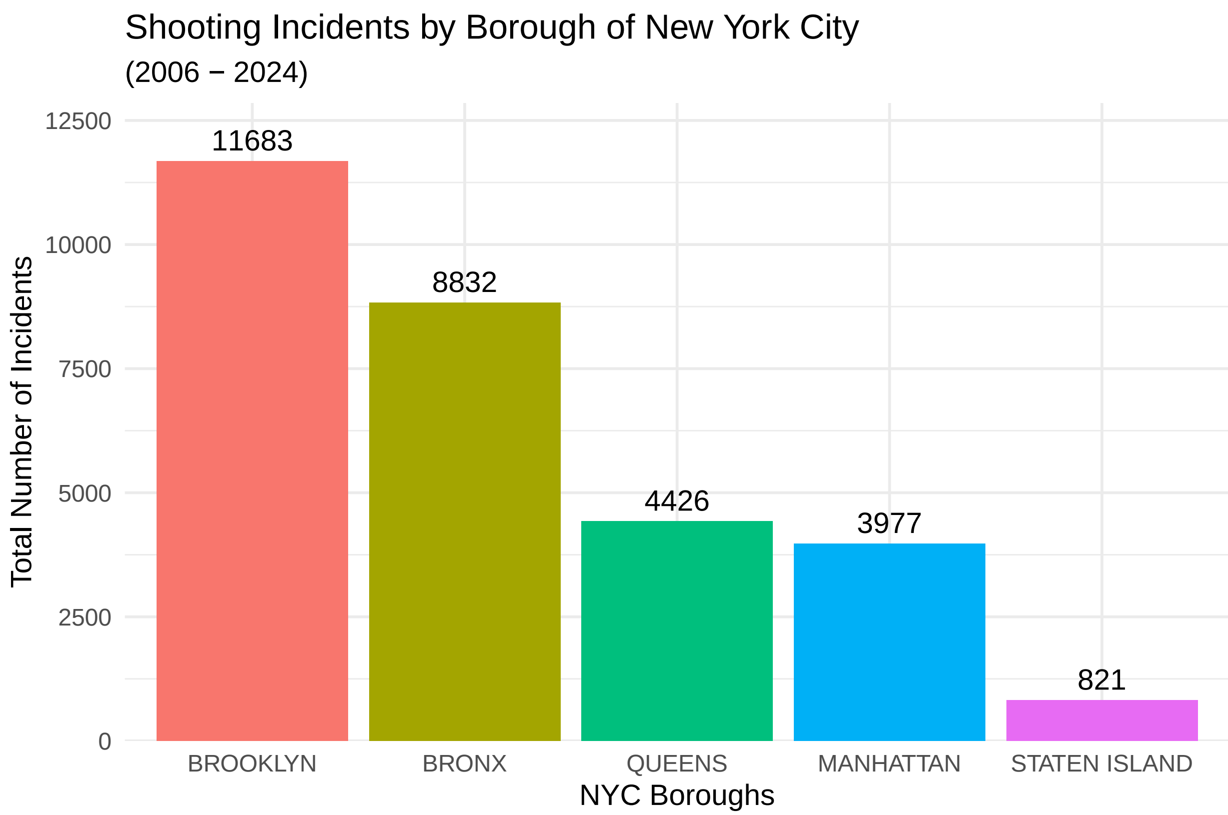
1. Which part of New York has the highest number of incidents? How many of those were murder cases?

- Brooklyn recorded the highest number of shooting incidents, followed by the Bronx and Queens. The pattern is similar when looking specifically at murder cases, with Brooklyn again leading, followed by the Bronx and Queens in the same order.

```r
boro_counts <- nypd_shooting %>%
  count(BORO) %>%
  mutate(BORO = fct_reorder(BORO, n, .desc = TRUE))

ggplot(boro_counts, aes(x = BORO, y = n, fill = BORO)) +
  geom_col() +
  geom_text(aes(label = n), vjust = -0.5, size = 4) + # Add count labels
  labs(title = "Shooting Incidents by Borough of New York City",
       subtitle = "(2006 - 2024)",
```

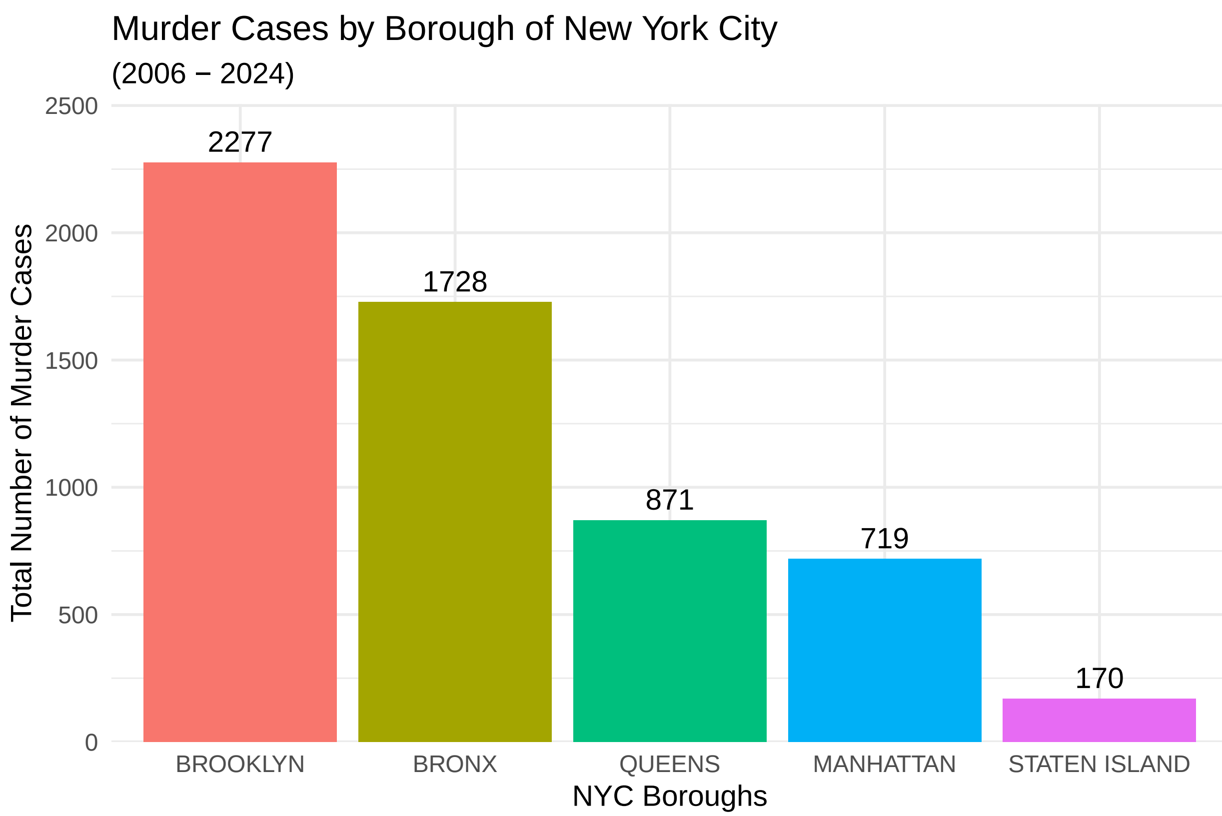
```
       x = "NYC Boroughs",
       y = "Total Number of Incidents",
       caption = "(Figure - 1)") +
  scale_y_continuous(expand = expansion(mult = c(0, 0.1))) +
  theme_minimal() +
  theme(legend.position = "none")
```



(Figure – 1)

```
boro_murders <- nypd_shooting %>%
  filter(STATISTICAL_MURDER_FLAG == "true") %>%
  count(BORO) %>%
  mutate(BORO = fct_reorder(BORO, n, .desc = TRUE))

ggplot(boro_murders, aes(x = BORO, y = n, fill = BORO)) +
  geom_col() +
  geom_text(aes(label = n), vjust = -0.5, size = 4) + # Add count labels
  labs(title = "Murder Cases by Borough of New York City",
       subtitle = "(2006 - 2024)",
       x = "NYC Boroughs",
       y = "Total Number of Murder Cases",
       caption = "(Figure - 2)") +
  scale_y_continuous(expand = expansion(mult = c(0, 0.1))) +
  theme_minimal() +
  theme(legend.position = "none")
```



(Figure − 2)

2. When should people in New York be most cautious about becoming victims of crime?

- **Summer months**—particularly **June**, **July**, and **August**—see the highest number of incidents.
- **Weekends** tend to have more criminal activity, so extra caution is advised.
- **Evenings and nighttime** are the riskiest hours. Unless it's necessary, staying indoors during these times is strongly recommended.

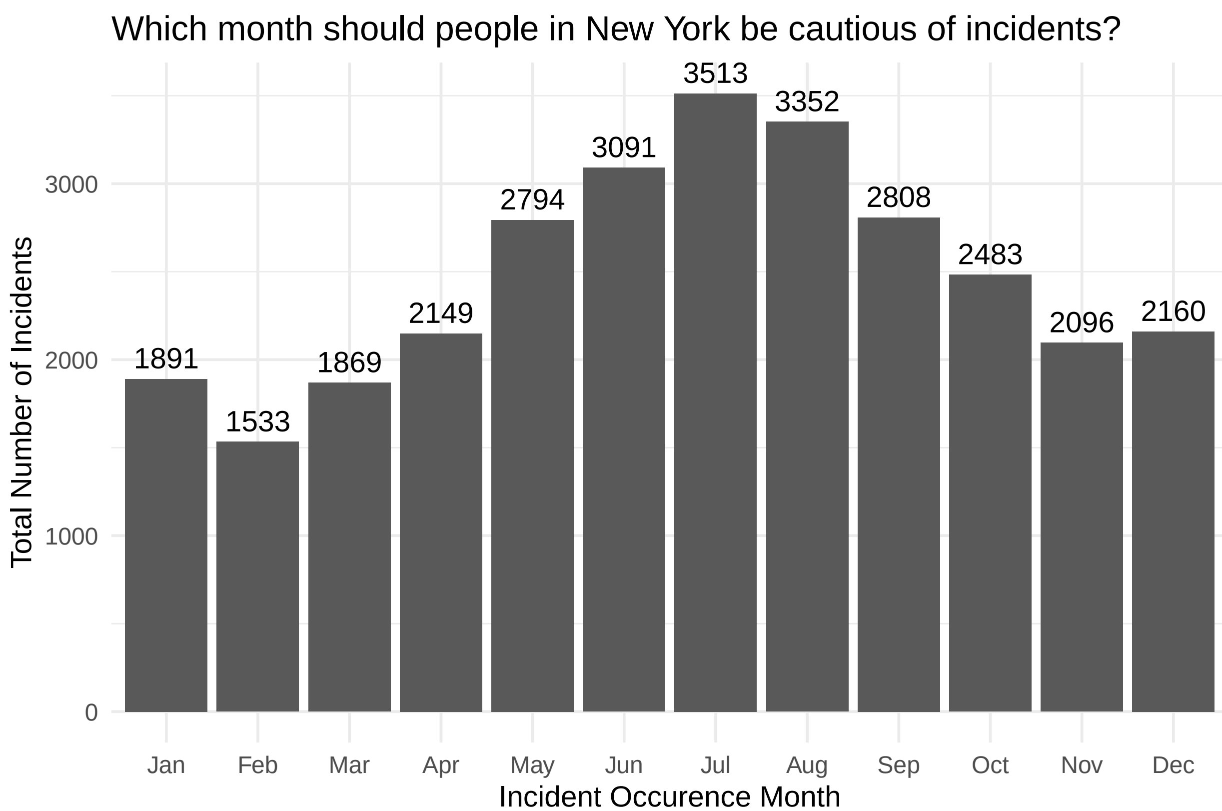
```
incident_by_month <- nypd_shooting %>%
  group_by(OCCUR_MONTH) %>%
  count()

incident_by_wday <- nypd_shooting %>%
  group_by(OCCUR_WDAY) %>%
  count()

incident_by_hour <- nypd_shooting %>%
  group_by(OCCUR_HOUR) %>%
  count()
```

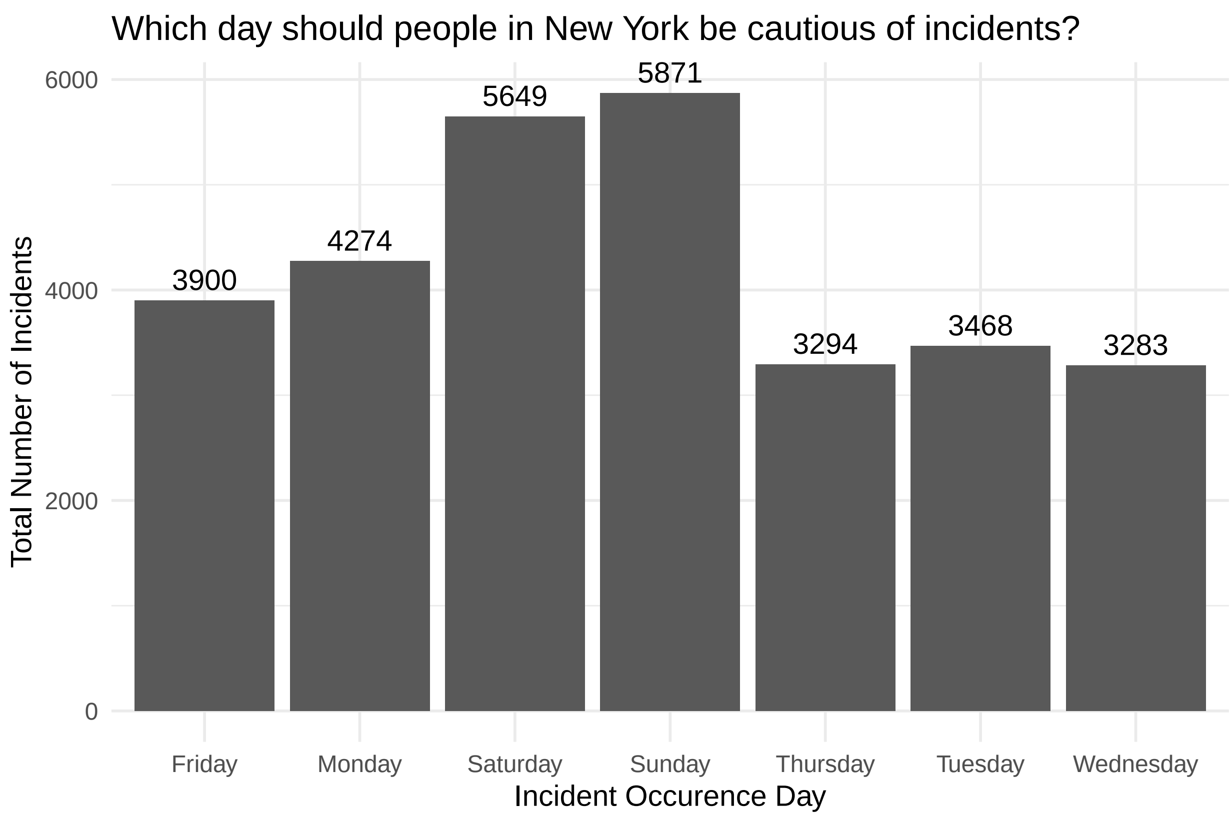
```
g <- ggplot(incident_by_month, aes(x = OCCUR_MONTH, y = n)) +
  geom_col() +
  geom_text(aes(label = n), vjust = -0.5, size = 4) +
  labs(title = "Which month should people in New York be cautious of incidents?",
```

```
      x = "Incident Occurence Month",
      y = "Total Number of Incidents",
      caption = "(Figure - 3)") +
  theme_minimal()
g
```



(Figure − 3)

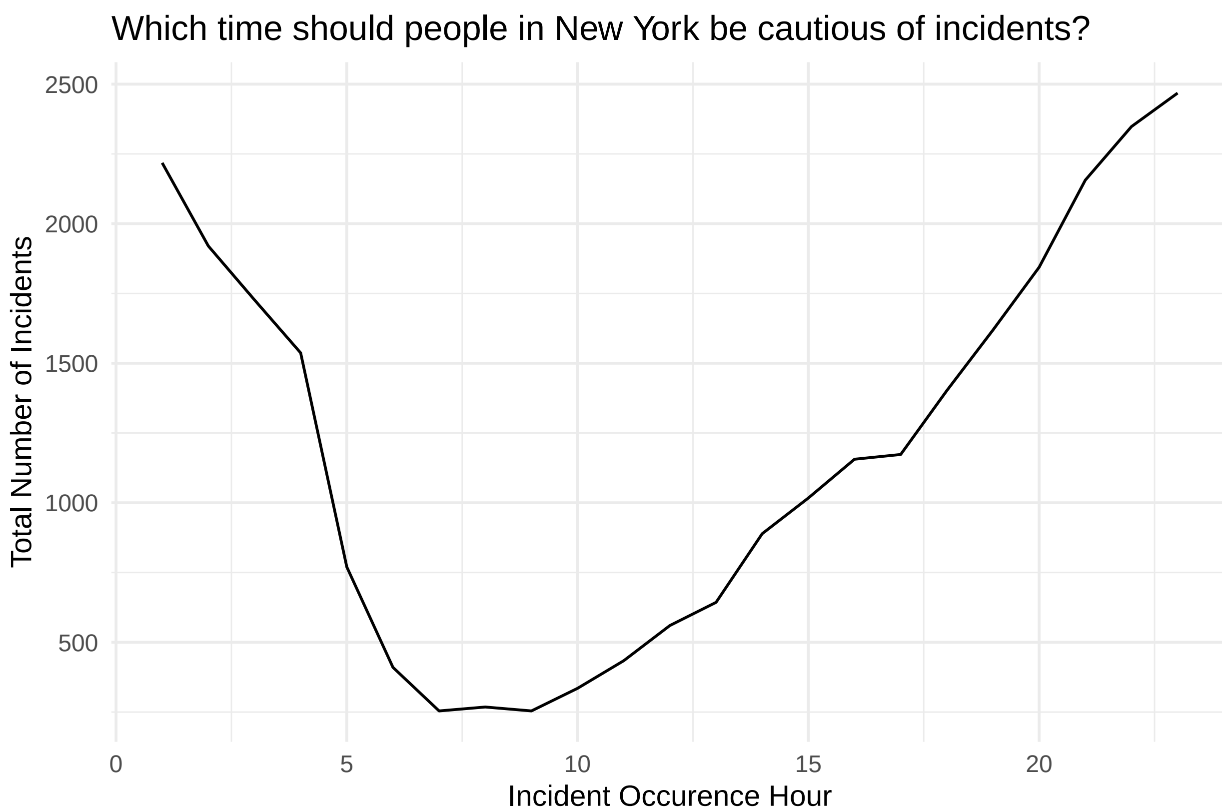
```
g <- ggplot(incident_by_wday, aes(x = OCCUR_WDAY, y = n)) +
  geom_col() +
  geom_text(aes(label = n), vjust = -0.5, size = 4) +
  labs(title = "Which day should people in New York be cautious of incidents?",
      x = "Incident Occurence Day",
      y = "Total Number of Incidents",
      caption = "(Figure - 4)") +
  theme_minimal()
g
```



(Figure – 4)

```r
g <- ggplot(incident_by_hour, aes(x = OCCUR_HOUR, y = n)) +
  geom_line() +
  labs(title = "Which time should people in New York be cautious of incidents?",
       x = "Incident Occurence Hour",
       y = "Total Number of Incidents",
       caption = "(Figure - 5)") +
  theme_minimal()
g
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_line()`).
```



(Figure − 5)

3. The Profile of Perpetrators and Victims

- A significant number of incidents involve individuals aged **18–24** and **25–44**, making these the most affected age groups.
- **Black** and **White Hispanic** individuals appear most frequently in incident records across New York City boroughs.
- The vast majority of incidents involve **male** individuals—far more than female—highlighting a clear gender disparity.

```
table(nypd_shooting$PERP_AGE_GROUP, nypd_shooting$VIC_AGE_GROUP)
```

```
##
##            <18 1022 18-24 25-44 45-64  65+ UNKNOWN
##            812    0  3568  4342   573   44       5
##   (null)   156    0   457   859   135   21       0
##   <18      566    0   669   455    90   23       2
##   18-24    825    1  2903  2483   355   49      14
##   25-44    284    0  1622  3773   571   52      40
##   45-64     22    0    90   419   221   18       5
##   65+        0    0     2    27    25   13       0
##   Unknown  416    0  1364  1202   148   16       2
```

```r
table(nypd_shooting$PERP_SEX, nypd_shooting$VIC_SEX)
```

```
##
##                F     M Unknown
##              693  8614       3
##   (null)     176  1452       0
##   F           80   380       1
##   M         1830 15003       7
##   Unknown    112  1387       1
```

```r
table(nypd_shooting$PERP_RACE, nypd_shooting$VIC_RACE)
```

```
##
##                                  AMERICAN INDIAN/ALASKAN NATIVE
##                                                               2
##   (null)                                                      2
##   AMERICAN INDIAN/ALASKAN NATIVE                              0
##   ASIAN / PACIFIC ISLANDER                                    0
##   BLACK                                                       5
##   BLACK HISPANIC                                              0
##   Unknown                                                     3
##   WHITE                                                       0
##   WHITE HISPANIC                                              1
##
##                                  ASIAN / PACIFIC ISLANDER BLACK BLACK HISPANIC
##                                                        97  7164            844
##   (null)                                               38  1147            160
##   AMERICAN INDIAN/ALASKAN NATIVE                        0     2              0
##   ASIAN / PACIFIC ISLANDER                             69    61             14
##   BLACK                                               174  9728            863
##   BLACK HISPANIC                                       24   587            399
##   Unknown                                              16  1360            155
##   WHITE                                                13    45             23
##   WHITE HISPANIC                                       47   902            472
##
##                                  Unknown WHITE WHITE HISPANIC
##                                       18   145           1040
##   (null)                               1    21            259
##   AMERICAN INDIAN/ALASKAN NATIVE       0     0              0
##   ASIAN / PACIFIC ISLANDER             0    12             28
##   BLACK                               26   211           1313
##   BLACK HISPANIC                       7    37            433
##   Unknown                              7    42            255
##   WHITE                                1   168             55
##   WHITE HISPANIC                      12   105           1126
```

4. Building a Logistic Regression Model to Predict Whether an Incident Is a Murder Case

Logistic regression is a classification technique used to predict a categorical outcome based on input variables. In this case, I use logistic regression to estimate the probability that a shooting incident results in a murder, based on factors such as the **demographic profile** of those involved, the **location** of the incident, and the **date and time** it occurred.

```r
# Logistics Regression
glm_model <- glm(STATISTICAL_MURDER_FLAG ~ PERP_RACE + PERP_SEX + PERP_AGE_GROUP + VIC_RACE + VIC_SEX +
summary(glm_model)
```

```
##
## Call:
## glm(formula = STATISTICAL_MURDER_FLAG ~ PERP_RACE + PERP_SEX +
##     PERP_AGE_GROUP + VIC_RACE + VIC_SEX + VIC_AGE_GROUP + OCCUR_HOUR +
##     OCCUR_WDAY + OCCUR_MONTH + Latitude + Longitude + BORO, family = binomial,
##     data = nypd_shooting)
##
## Coefficients: (3 not defined because of singularities)
##                                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)                        67.926327 144.547278   0.470 0.638409
## PERP_RACE(null)                    -0.297178   0.079897  -3.720 0.000200
## PERP_RACEAMERICAN INDIAN/ALASKAN NATIVE -10.836592 378.426625  -0.029 0.977155
## PERP_RACEASIAN / PACIFIC ISLANDER   2.335721   0.467274   4.999 5.77e-07
## PERP_RACEBLACK                      1.987536   0.434193   4.578 4.70e-06
## PERP_RACEBLACK HISPANIC             2.002410   0.438728   4.564 5.02e-06
## PERP_RACEUnknown                    1.373405   0.370421   3.708 0.000209
## PERP_RACEWHITE                      2.493991   0.454930   5.482 4.20e-08
## PERP_RACEWHITE HISPANIC             2.153178   0.436642   4.931 8.17e-07
## PERP_SEX(null)                            NA         NA      NA       NA
## PERP_SEXF                          -1.510590   0.290158  -5.206 1.93e-07
## PERP_SEXM                          -1.663387   0.268035  -6.206 5.44e-10
## PERP_SEXUnknown                           NA         NA      NA       NA
## PERP_AGE_GROUP(null)                      NA         NA      NA       NA
## PERP_AGE_GROUP<18                  -0.222629   0.423838  -0.525 0.599396
## PERP_AGE_GROUP18-24                -0.120684   0.420063  -0.287 0.773883
## PERP_AGE_GROUP25-44                 0.074097   0.420159   0.176 0.860016
## PERP_AGE_GROUP45-64                 0.367747   0.426952   0.861 0.389055
## PERP_AGE_GROUP65+                   0.438533   0.495961   0.884 0.376584
## PERP_AGE_GROUPUnknown              -2.694238   0.387735  -6.949 3.69e-12
## VIC_RACEASIAN / PACIFIC ISLANDER   11.900194 139.189985   0.085 0.931867
## VIC_RACEBLACK                      11.738120 139.189936   0.084 0.932793
## VIC_RACEBLACK HISPANIC             11.530947 139.189944   0.083 0.933976
## VIC_RACEUnknown                    10.982371 139.190564   0.079 0.937111
## VIC_RACEWHITE                      11.803491 139.189970   0.085 0.932419
## VIC_RACEWHITE HISPANIC             11.774534 139.189940   0.085 0.932585
## VIC_SEXM                            0.017286   0.052457   0.330 0.741756
## VIC_SEXUnknown                    -11.670140 165.721114  -0.070 0.943859
## VIC_AGE_GROUP1022                 -11.611771 535.411178  -0.022 0.982697
## VIC_AGE_GROUP18-24                  0.209116   0.063242   3.307 0.000944
## VIC_AGE_GROUP25-44                  0.487608   0.061962   7.869 3.56e-15
## VIC_AGE_GROUP45-64                  0.579500   0.079068   7.329 2.32e-13
## VIC_AGE_GROUP65+                    0.803813   0.159617   5.036 4.76e-07
## VIC_AGE_GROUPUNKNOWN                0.496835   0.329443   1.508 0.131528
## OCCUR_HOUR                         -0.001062   0.002066  -0.514 0.607113
## OCCUR_WDAYMonday                    0.027367   0.058746   0.466 0.641325
## OCCUR_WDAYSaturday                 -0.072751   0.056626  -1.285 0.198874
## OCCUR_WDAYSunday                    0.049256   0.055948   0.880 0.378657
## OCCUR_WDAYThursday                 -0.056058   0.063102  -0.888 0.374335
## OCCUR_WDAYTuesday                  -0.078677   0.062967  -1.249 0.211489
```

```
## OCCUR_WDAYWednesday                     0.039515   0.062357   0.634 0.526288
## OCCUR_MONTH.L                           0.036067   0.058567   0.616 0.538003
## OCCUR_MONTH.Q                           0.220660   0.055432   3.981 6.87e-05
## OCCUR_MONTH.C                           0.055127   0.055596   0.992 0.321412
## OCCUR_MONTH^4                          -0.099833   0.056534  -1.766 0.077413
## OCCUR_MONTH^5                           0.033818   0.056838   0.595 0.551852
## OCCUR_MONTH^6                           0.057490   0.056913   1.010 0.312428
## OCCUR_MONTH^7                           0.163090   0.056181   2.903 0.003697
## OCCUR_MONTH^8                          -0.052693   0.055197  -0.955 0.339766
## OCCUR_MONTH^9                           0.024608   0.053746   0.458 0.647051
## OCCUR_MONTH^10                         -0.069888   0.051723  -1.351 0.176633
## OCCUR_MONTH^11                         -0.116280   0.049751  -2.337 0.019426
## Latitude                               -0.169554   0.484025  -0.350 0.726113
## Longitude                               1.008971   0.456486   2.210 0.027084
## BOROBROOKLYN                            0.028330   0.096795   0.293 0.769767
## BOROMANHATTAN                          -0.116413   0.063439  -1.835 0.066497
## BOROQUEENS                             -0.124683   0.095480  -1.306 0.191604
## BOROSTATEN ISLAND                       0.081862   0.176500   0.464 0.642785
##
## (Intercept)
## PERP_RACE(null)                        ***
## PERP_RACEAMERICAN INDIAN/ALASKAN NATIVE
## PERP_RACEASIAN / PACIFIC ISLANDER      ***
## PERP_RACEBLACK                         ***
## PERP_RACEBLACK HISPANIC                ***
## PERP_RACEUnknown                       ***
## PERP_RACEWHITE                         ***
## PERP_RACEWHITE HISPANIC                ***
## PERP_SEX(null)
## PERP_SEXF                              ***
## PERP_SEXM                              ***
## PERP_SEXUnknown
## PERP_AGE_GROUP(null)
## PERP_AGE_GROUP<18
## PERP_AGE_GROUP18-24
## PERP_AGE_GROUP25-44
## PERP_AGE_GROUP45-64
## PERP_AGE_GROUP65+
## PERP_AGE_GROUPUnknown                  ***
## VIC_RACEASIAN / PACIFIC ISLANDER
## VIC_RACEBLACK
## VIC_RACEBLACK HISPANIC
## VIC_RACEUnknown
## VIC_RACEWHITE
## VIC_RACEWHITE HISPANIC
## VIC_SEXM
## VIC_SEXUnknown
## VIC_AGE_GROUP1022
## VIC_AGE_GROUP18-24                     ***
## VIC_AGE_GROUP25-44                     ***
## VIC_AGE_GROUP45-64                     ***
## VIC_AGE_GROUP65+                       ***
## VIC_AGE_GROUPUNKNOWN
## OCCUR_HOUR
```

```
## OCCUR_WDAYMonday
## OCCUR_WDAYSaturday
## OCCUR_WDAYSunday
## OCCUR_WDAYThursday
## OCCUR_WDAYTuesday
## OCCUR_WDAYWednesday
## OCCUR_MONTH.L
## OCCUR_MONTH.Q                         ***
## OCCUR_MONTH.C
## OCCUR_MONTH^4                         .
## OCCUR_MONTH^5
## OCCUR_MONTH^6
## OCCUR_MONTH^7                         **
## OCCUR_MONTH^8
## OCCUR_MONTH^9
## OCCUR_MONTH^10
## OCCUR_MONTH^11                        *
## Latitude
## Longitude                             *
## 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: 26992  on 27312  degrees of freedom
## Residual deviance: 25679  on 27258  degrees of freedom
##   (2426 observations deleted due to missingness)
## AIC: 25789
##
## Number of Fisher Scoring iterations: 12
```

Step 4: Report Conclusion and Sources of Bias

Conclusion

I aimed to explore whether certain factors—such as the victim's demographics (age, sex, race), the perpetrator's background, or the location of the incident—could help predict whether a shooting would be fatal. Using logistic regression, I found that the **victim's age group** and the **perpetrator's race** were statistically significant predictors of whether the victim survived. These results suggest that both individual characteristics and contextual factors play a role in the outcome of shooting incidents.

Sources of Bias

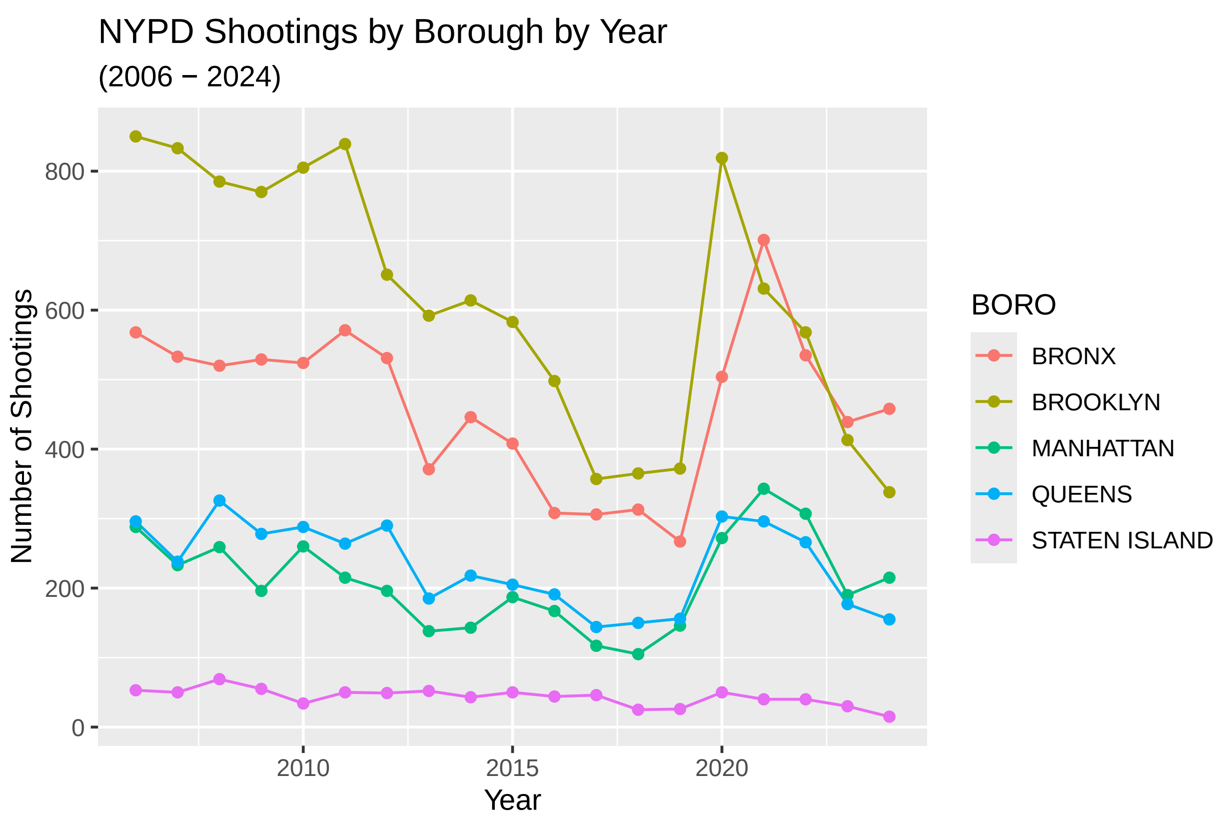
When analyzing this topic, it's important to acknowledge the possibility of unconscious bias and stereotypes, particularly when we rely on assumptions formed by media exposure rather than personal experience. As someone who has never lived in New York City, my perceptions have been largely shaped by news reports, movies, and social media. For example, I initially assumed that the Bronx would have the highest number of incidents, simply because it is often portrayed negatively in the media. I also believed that women might be more likely to be victims of such incidents.

However, when looking at the actual data, I was surprised to find that Brooklyn ranks highest in terms of total incidents, followed by the Bronx and Queens. This trend is also reflected in murder cases. Interestingly, the data shows that men are involved in significantly more incidents than women.

This highlights the importance of validating our assumptions with data. Relying solely on impressions or second-hand information can lead to flawed conclusions and reinforce stereotypes. Data-driven analysis helps us avoid these pitfalls and make more accurate, informed judgments. My findings align with a CNN report titled "Hate crimes, shooting incidents in New York City have surged since last year", which notes that shooting incidents in NYC increased by 73% in May 2021 compared to May 2020.

```
NYPD_boro_year <- nypd_shooting %>%
  group_by(BORO, OCCUR_YEAR, Shootings) %>%
  summarize(Shootings = sum(Shootings),
            .groups = 'drop') %>%
  select(BORO, OCCUR_YEAR, Shootings) %>%
  ungroup()

NYPD_boro_year %>%
  ggplot(aes(x = OCCUR_YEAR, y = Shootings ,color = BORO)) +
  geom_line() +
  geom_point() +
  labs(title = "NYPD Shootings by Borough by Year",
       subtitle = "(2006 - 2024)",
       x = "Year",
       y = "Number of Shootings",
       caption = "(Figure - 6)")
```



(Figure − 6)

Additional Resources

- NYPD Shooting Incident Data (Historic) - CKAN
- NYC, Chicago see another wave of weekend gun violence
- Hate crimes, shooting incidents in New York City have surged since last year, NYPD data show - CNN