NYPD Shooting Incident Data Report

Eric Yu

2024-06-21

Introduction

This report is an analysis of the NYPD Shooting Incident Data (Historic), a data-set recording shooting incidents that occurred from 2006 to the present day in New York City. The goal of this project is to explore trends in the data and uncover potential insights that may help to understand the shooting incidents.

Research Questions

- Where and when are shootings most likely to occur?
- What are the demographics of perpetrators and victims?
- Which factors are correlated with incidents that resulted in murders?

Importing Libraries and Data-set

The analysis will be performed using R libraries. These tools will be used to import the data-set of interest, clean/transform the data, and generate results/visualizations.

Load R libraries

```
library(tidyverse)
library(ggplot2)
library(tidyr)
```

Load data-set from csv file

```
url <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv"
nypd_data <- read_csv(url)
```

Tidy and clean data

For the purposes of this report, some columns include extra information that are unnecessary for answering the research questions. I am interested in victim/perpetrator demographics information, and general descriptions of the locations of the incidents.

```r
nypd_data_cleaned <- nypd_data %>%
  select(-c(LOC_OF_OCCUR_DESC,
            PRECINCT,
            JURISDICTION_CODE,
            LOC_CLASSFCTN_DESC,
            Latitude,
            Longitude,
            Lon_Lat,
            X_COORD_CD,Y_COORD_CD))
```

Remove unnecessary columns The dates and times at which the shootings occurred are given in `MM/DD/YYYY` and `H:M:S` formats, respectively. For the purposes of this report, I will consider only the months and hours of the incidents for analysis.

```r
# Retain only the month of the incident
nypd_data_cleaned$OCCUR_DATE <-
  as.integer(substr(nypd_data_cleaned$OCCUR_DATE,
                    start=1,
                    stop=2))

# Convert month number to name of month
nypd_data_cleaned$OCCUR_DATE <-
  month.name[nypd_data_cleaned$OCCUR_DATE]

# Shorten months to first three letters
nypd_data_cleaned$OCCUR_DATE <-
  substr(nypd_data_cleaned$OCCUR_DATE, start=1, stop=3)

# Retain only the hour of the incident
nypd_data_cleaned$OCCUR_TIME <-
  as.integer(substr(nypd_data_cleaned$OCCUR_TIME,
                    start=1,
                    stop=2))

# Convert to AM/PM format
am_pm_mapping <- c(`0`="12:00 AM", `1`="1:00 AM", `2`="2:00 AM",
                   `3`="3:00 AM", `4`="4:00 AM", `5`="5:00 AM",
                   `6`="6:00 AM", `7`="7:00 AM", `8`="8:00 AM",
                   `9`="9:00 AM", `10`="10:00 AM", `11`="11:00 AM",
                   `12`="12:00 PM", `13`="1:00 PM", `14`="2:00 PM",
                   `15`="3:00 PM", `16`="4:00 PM", `17`="5:00 PM",
                   `18`="6:00 PM", `19`="7:00 PM", `20`="8:00 PM",
                   `21`="9:00 PM", `22`="10:00 PM", `23`="11:00 PM")
nypd_data_cleaned$OCCUR_TIME <-
  am_pm_mapping[as.character(nypd_data_cleaned$OCCUR_TIME)]
```

Remove entries with unknown values Some columns contain null values or values that do not make sense in the context of the other values in the column. I will drop rows that have these values since I cannot accurately interpret them.

- **`PERP_AGE_GROUP`** and **`VIC_AGE_GROUP`** contain `"(null)"` values (distinct from `"UNKNOWN"`) and values that do not follow a standard format.

- **PERP_RACE** contains "(null)" values (distinct from "UNKNOWN")
- **LOCATION_DESC** contains "(null)" values (distinct from "NONE")

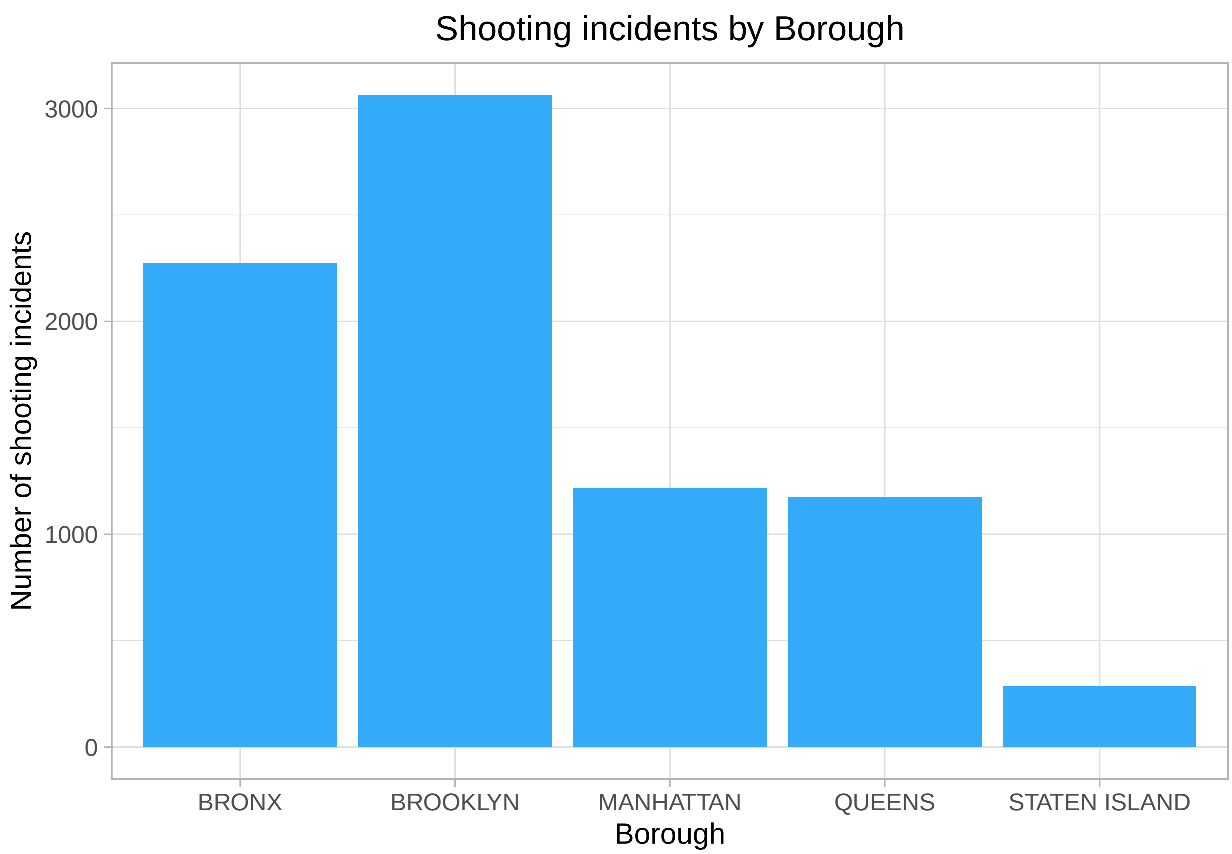
```
excluded_age <- c("(null)", "1020", "1022", "1028", "224", "940")
excluded_race <- c("(null)")
excluded_loc <- c("(null)")

nypd_data_cleaned <- nypd_data_cleaned %>%
  filter(
    !is.na(nypd_data_cleaned$PERP_AGE_GROUP),
    !(nypd_data_cleaned$PERP_AGE_GROUP %in% excluded_age),
    !is.na(nypd_data_cleaned$VIC_AGE_GROUP),
    !(nypd_data_cleaned$VIC_AGE_GROUP %in% excluded_age),
    !is.na(nypd_data_cleaned$PERP_RACE),
    !(nypd_data_cleaned$PERP_RACE %in% excluded_race),
    !is.na(nypd_data_cleaned$LOCATION_DESC),
    !(nypd_data_cleaned$LOCATION_DESC %in% excluded_loc))

glimpse(nypd_data_cleaned)
```

```
## Rows: 8,017
## Columns: 12
## $ INCIDENT_KEY            <dbl> 244608249, 33445716, 11118232, 10137418, 15675~
## $ OCCUR_DATE              <chr> "May", "Jul", "Apr", "Jan", "Sep", "Jun", "May~
## $ OCCUR_TIME              <chr> "12:00 AM", "4:00 AM", "6:00 PM", "7:00 PM", "~
## $ BORO                    <chr> "MANHATTAN", "QUEENS", "QUEENS", "BROOKLYN", "~
## $ LOCATION_DESC           <chr> "VIDEO STORE", "BAR/NIGHT CLUB", "PVT HOUSE", ~
## $ STATISTICAL_MURDER_FLAG <lgl> TRUE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE~
## $ PERP_AGE_GROUP          <chr> "25-44", "UNKNOWN", "45-64", "25-44", "25-44",~
## $ PERP_SEX                <chr> "M", "M", "M", "M", "M", "M", "M", "M", "M", "~
## $ PERP_RACE               <chr> "BLACK", "BLACK", "BLACK", "BLACK", "BLACK", "~
## $ VIC_AGE_GROUP           <chr> "25-44", "25-44", "45-64", "25-44", "18-24", "~
## $ VIC_SEX                 <chr> "M", "M", "M", "M", "M", "M", "M", "M", "M", "~
## $ VIC_RACE                <chr> "BLACK", "BLACK", "BLACK", "BLACK", "WHITE HIS~
```

Exploring the Data

Shooting incidents by location (borough)

```
nypd_data_cleaned %>%
  ggplot(aes(x=BORO)) +
  theme_light() +
  theme(plot.title = element_text(hjust=0.5)) +
  geom_bar(fill="#33ABF9") +
  labs(x="Borough",
       y="Number of shooting incidents",
       title="Shooting incidents by Borough")
```




```r
# Table of counts to extract exact values
table(nypd_data_cleaned$BORO)
```

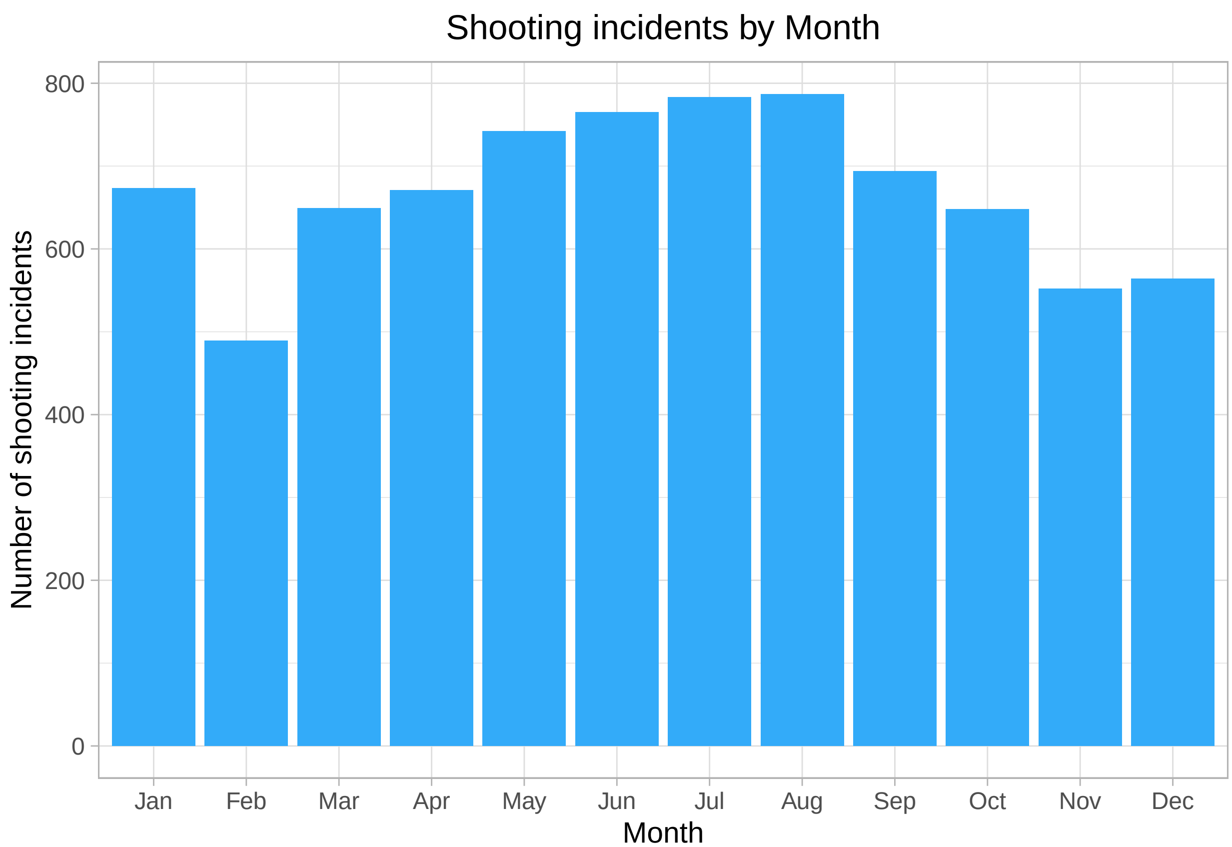
```
##
##         BRONX      BROOKLYN     MANHATTAN        QUEENS STATEN ISLAND
##          2272          3063          1218          1176           288
```

The number of shootings that occurred in Staten Island appear to be significantly less than in the other four boroughs. This observation makes logical sense because the population of New York City is largely concentrated outside of Staten Island (roughly 490,000 in Staten Island, 7,800,000 in the other four boroughs in 2022).

Shooting incidents by month

```r
month_order <- c("Jan", "Feb", "Mar", "Apr",
                 "May", "Jun", "Jul", "Aug",
                 "Sep", "Oct", "Nov", "Dec")

# Order by month in plot, default order is alphabetical
nypd_data_cleaned$OCCUR_DATE <- factor(nypd_data_cleaned$OCCUR_DATE, levels = month_order)

nypd_data_cleaned %>%
  ggplot(aes(x=OCCUR_DATE)) +
```

```
  theme_light() +
  theme(plot.title = element_text(hjust=0.5)) +
  geom_bar(fill="#33ABF9") +
  labs(x="Month",
       y="Number of shooting incidents",
       title="Shooting incidents by Month")
```



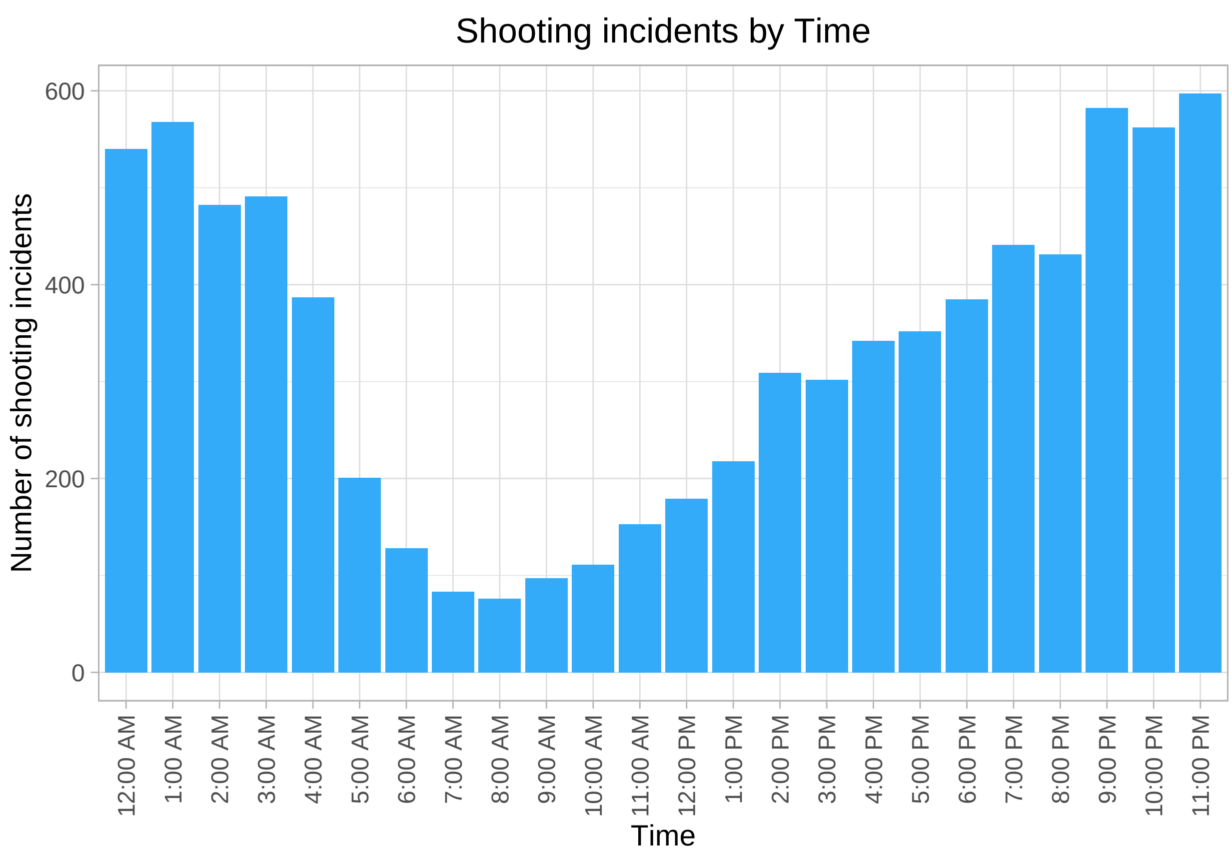

```
# Table of counts to extract exact values
table(nypd_data_cleaned$OCCUR_DATE)
```

```
##
## Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
## 673 489 649 671 742 765 783 787 694 648 552 564
```

The number of shootings appear to occur more frequently between June-August, which are summer months. There are several factors associated with the summer-time, which may be related to higher rates of shootings, such as heat, more frequent traveling (vacations), and more people going outdoors. However, the difference does not appear to be extremely significant.

Shooting incidents by time

```r
time_order <- c("12:00 AM","1:00 AM","2:00 AM","3:00 AM",
                "4:00 AM","5:00 AM","6:00 AM","7:00 AM",
                "8:00 AM","9:00 AM","10:00 AM","11:00 AM",
                "12:00 PM","1:00 PM","2:00 PM","3:00 PM",
                "4:00 PM","5:00 PM","6:00 PM","7:00 PM",
                "8:00 PM","9:00 PM","10:00 PM","11:00 PM")

# Order by time in plot
nypd_data_cleaned$OCCUR_TIME <- factor(nypd_data_cleaned$OCCUR_TIME, levels = time_order)

nypd_data_cleaned %>%
  ggplot(aes(x=OCCUR_TIME)) +
  theme_light() +
  theme(
    axis.text.x = element_text(angle=90, vjust=0.5, hjust=1),
    plot.title = element_text(hjust=0.5)) +
  geom_bar(fill="#33ABF9") +
  labs(x="Time",
       y="Number of shooting incidents",
       title="Shooting incidents by Time")
```




```r
# Table of counts to extract exact values
table(nypd_data_cleaned$OCCUR_TIME)
```

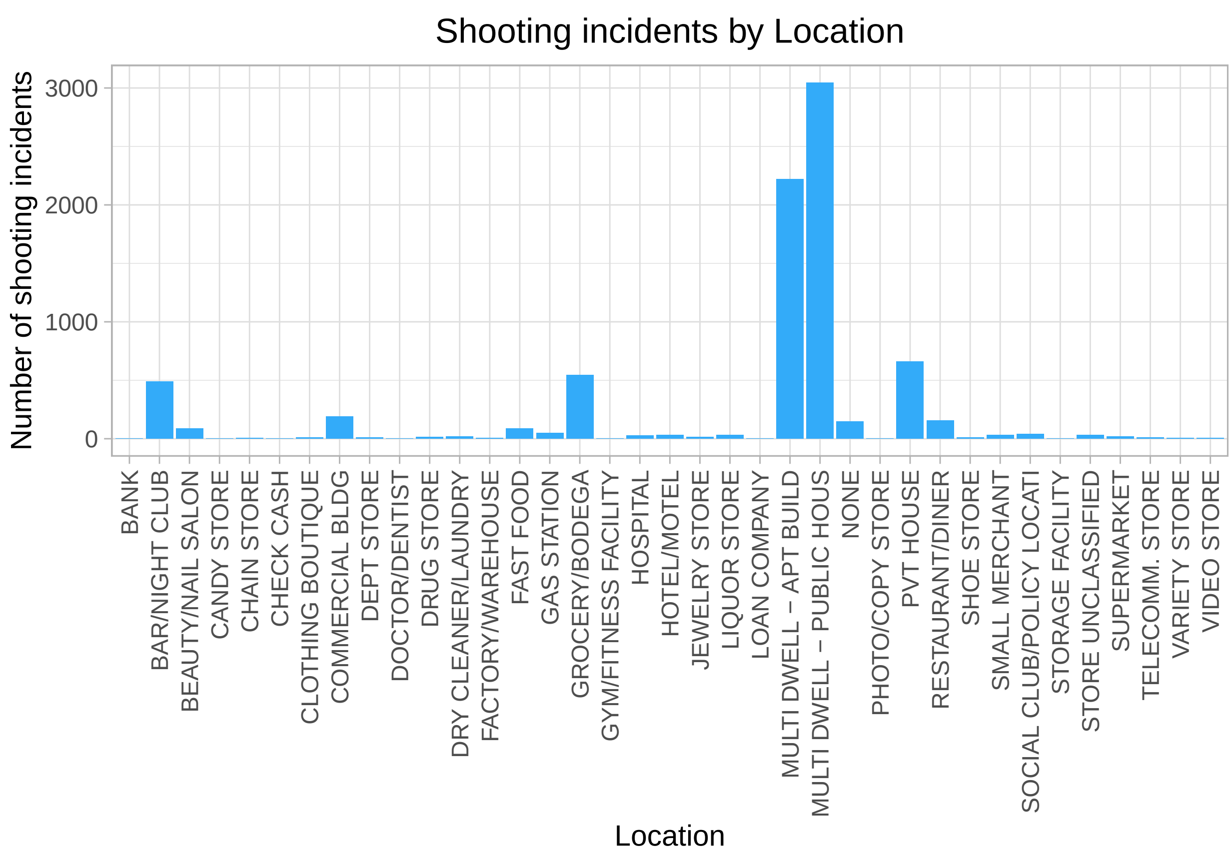
```
##
```

```
## 12:00 AM  1:00 AM  2:00 AM  3:00 AM  4:00 AM  5:00 AM  6:00 AM  7:00 AM
##      540      568      482      491      387      201      128       83
##  8:00 AM  9:00 AM 10:00 AM 11:00 AM 12:00 PM  1:00 PM  2:00 PM  3:00 PM
##       76       97      111      153      179      218      309      302
##  4:00 PM  5:00 PM  6:00 PM  7:00 PM  8:00 PM  9:00 PM 10:00 PM 11:00 PM
##      342      352      385      441      431      582      562      597
```

Shooting incidents appear to be most frequent during the evening to the early morning (past midnight but before sunrise). A possible explanation is that people tend to be asleep during these times and police activity may be lower at night, allowing perpetrators more opportunities to commit crimes.

Locations of shooting incidents

```
nypd_data_cleaned %>%
  ggplot(aes(x=LOCATION_DESC)) +
  theme_light() +
  theme(
    axis.text.x = element_text(angle=90, vjust=0.5, hjust=1),
    plot.title = element_text(hjust=0.5)) +
  geom_bar(fill="#33ABF9") +
  labs(x="Location",
       y="Number of shooting incidents",
       title="Shooting incidents by Location")
```



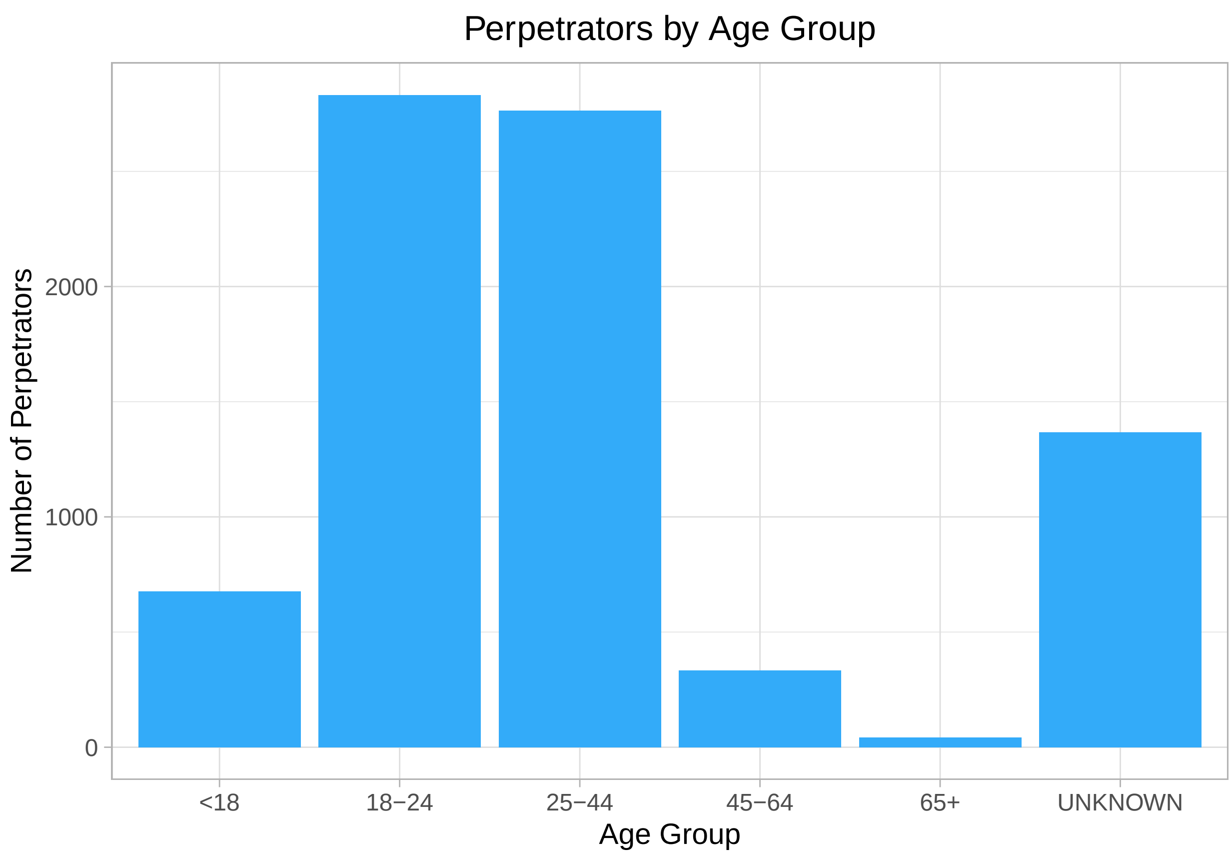
```r
# Table of counts to extract exact values
table(nypd_data_cleaned$LOCATION_DESC)
```

```
##
##                      BANK             BAR/NIGHT CLUB         BEAUTY/NAIL SALON
##                         2                        490                        87
##               CANDY STORE                CHAIN STORE                CHECK CASH
##                         3                          5                         1
##          CLOTHING BOUTIQUE            COMMERCIAL BLDG                DEPT STORE
##                         9                        189                         9
##            DOCTOR/DENTIST                 DRUG STORE      DRY CLEANER/LAUNDRY
##                         1                         13                        20
##         FACTORY/WAREHOUSE                  FAST FOOD               GAS STATION
##                         7                         89                        48
##            GROCERY/BODEGA       GYM/FITNESS FACILITY                  HOSPITAL
##                       546                          3                        26
##               HOTEL/MOTEL              JEWELRY STORE              LIQUOR STORE
##                        33                         13                        33
##              LOAN COMPANY   MULTI DWELL - APT BUILD MULTI DWELL - PUBLIC HOUS
##                         1                       2221                      3045
##                      NONE           PHOTO/COPY STORE                 PVT HOUSE
##                       147                          1                       662
##          RESTAURANT/DINER                 SHOE STORE            SMALL MERCHANT
##                       154                          9                        32
## SOCIAL CLUB/POLICY LOCATI          STORAGE FACILITY         STORE UNCLASSIFIED
##                        41                          1                        31
##               SUPERMARKET             TELECOMM. STORE             VARIETY STORE
##                        19                         11                         8
##               VIDEO STORE
##                         7
```

Shooting incidents appear to occur mostly in apartment buildings and public housing. A possible explanation is that people tend to keep money and valuable possessions at their homes, which may be motivations for perpetrators. The privacy of homes also allow criminals to conceal their crimes, which can explain why shootings occur more frequently in homes than in public places.

Perpetrator and victim demographics

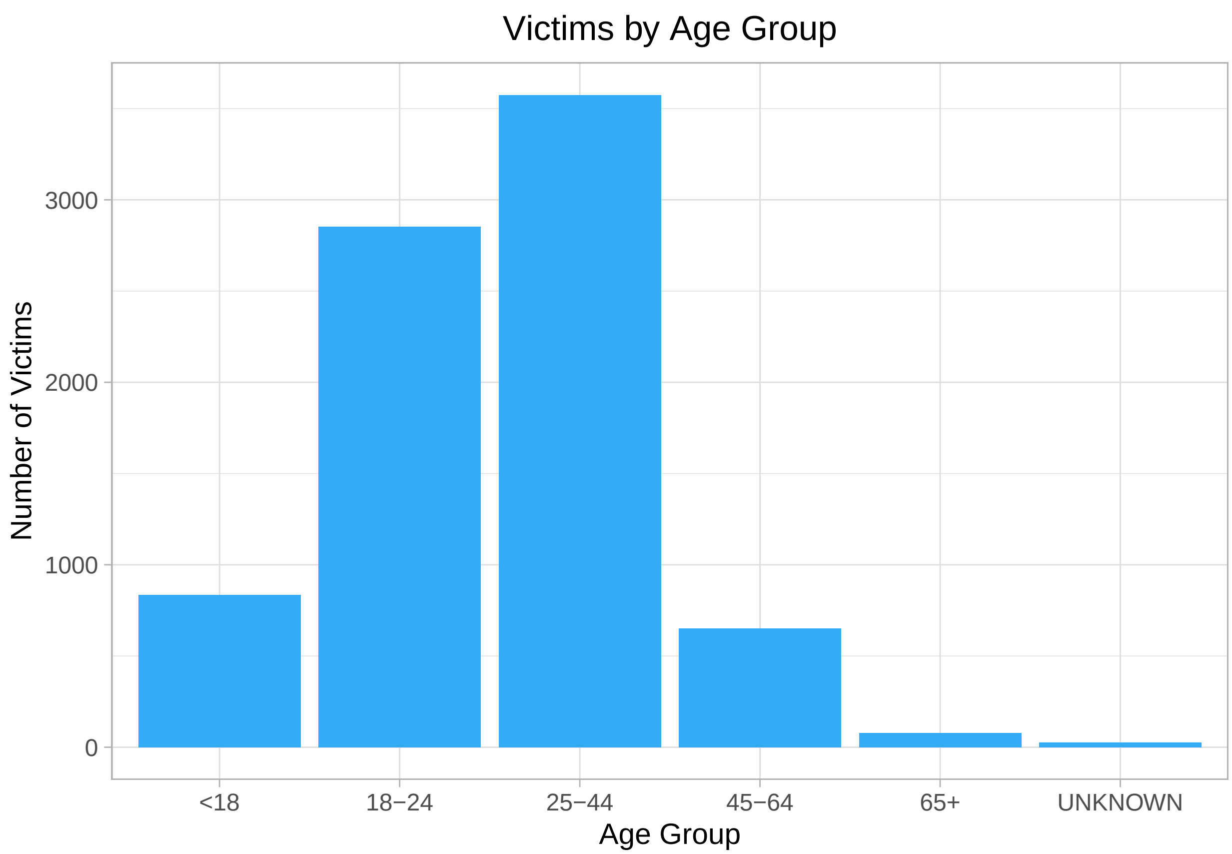
```r
# Age groups
nypd_data_cleaned %>%
  ggplot(aes(x=PERP_AGE_GROUP)) +
  theme_light() +
  theme(plot.title = element_text(hjust=0.5)) +
  geom_bar(fill="#33ABF9") +
  labs(x="Age Group",
       y="Number of Perpetrators",
       title="Perpetrators by Age Group")
```




```r
# Table of counts to extract exact values
table(nypd_data_cleaned$PERP_AGE_GROUP)
```

```
##
##     <18   18-24   25-44   45-64     65+ UNKNOWN
##     677    2832    2764     334      42    1368
```

```r
nypd_data_cleaned %>%
  ggplot(aes(x=VIC_AGE_GROUP)) +
  theme_light() +
  theme(plot.title = element_text(hjust=0.5)) +
  geom_bar(fill="#33ABF9") +
  labs(x="Age Group",
       y="Number of Victims",
       title="Victims by Age Group")
```

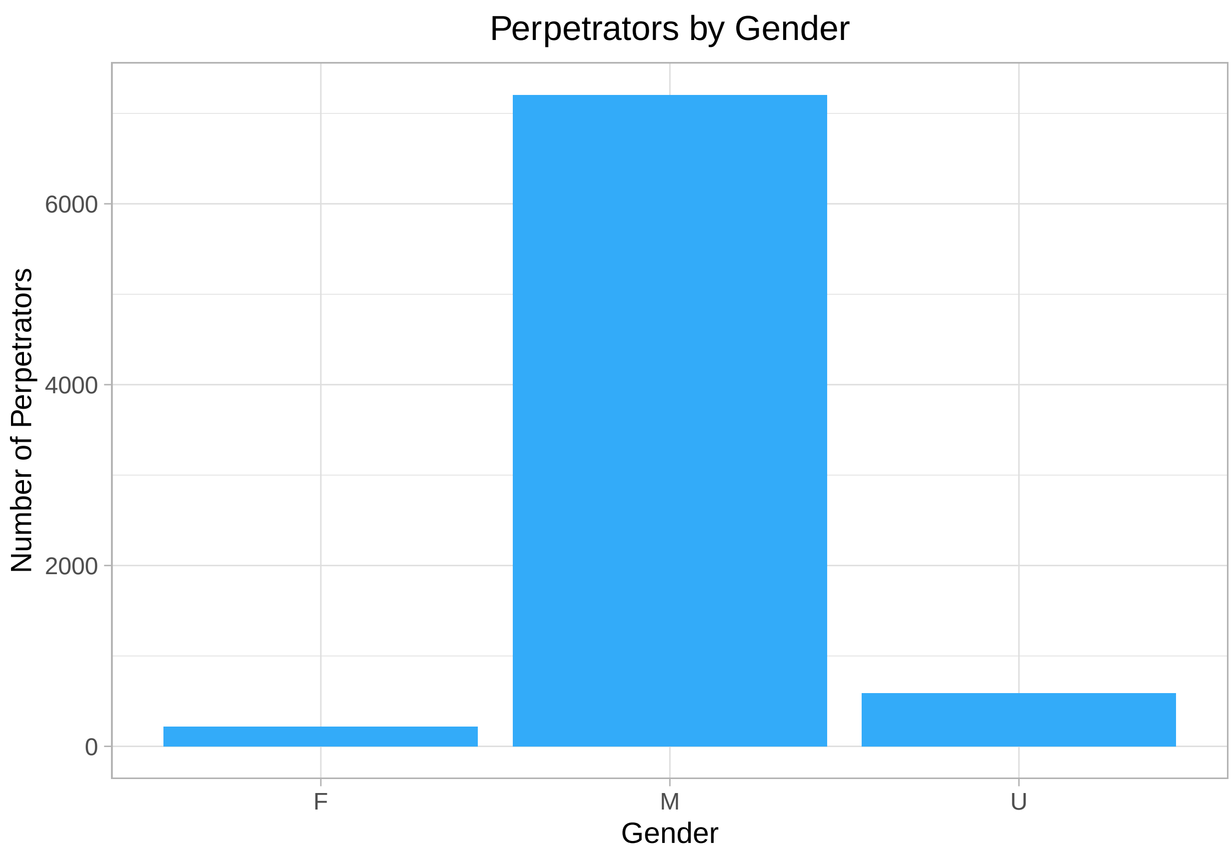



```
# Table of counts to extract exact values
table(nypd_data_cleaned$VIC_AGE_GROUP)
```

```
##
##     <18   18-24   25-44   45-64     65+ UNKNOWN
##     835    2852    3575     650      78      27
```

In both groups (perpetrators and victims), shootings appear to mostly involve those between the ages of 18 and 44. A possible explanation for why most perpetrators are in those age groups could be that the ages 18-44 are when people are most physically active and have access to firearms. The age groups of the victims are similar, but are more weighted towards the 25-44 age group. A possible explanation is that there may be more motivation to commit a crime against those in that age group (have more possessions, money, etc.).

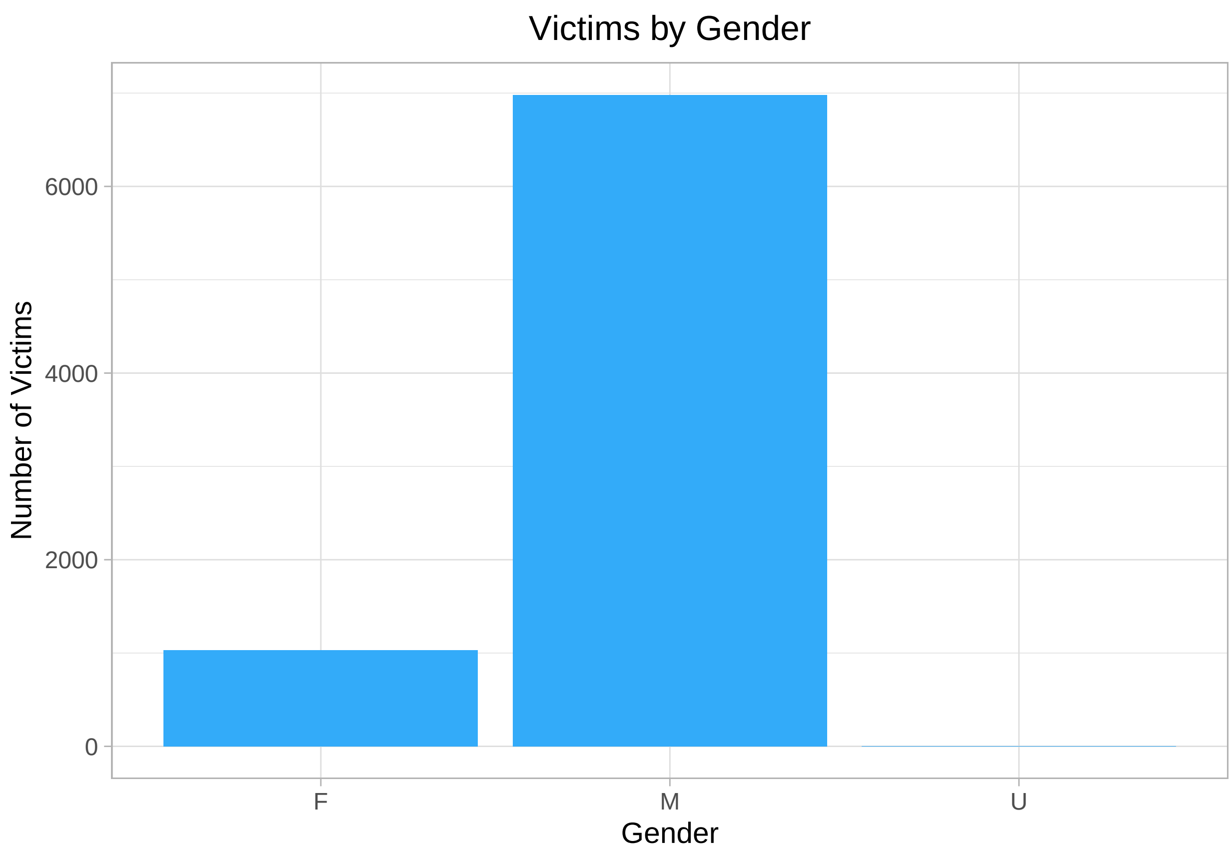
```
# Gender
nypd_data_cleaned %>%
  ggplot(aes(x=PERP_SEX)) +
  theme_light() +
  theme(plot.title = element_text(hjust=0.5)) +
  geom_bar(fill="#33ABF9") +
  labs(x="Gender",
       y="Number of Perpetrators",
       title="Perpetrators by Gender")
```




```r
# Table of counts to extract exact values
table(nypd_data_cleaned$PERP_SEX)
```

```
##
##    F    M    U
##  220 7206  591
```

```r
nypd_data_cleaned %>%
  ggplot(aes(x=VIC_SEX)) +
  theme_light() +
  theme(plot.title = element_text(hjust=0.5)) +
  geom_bar(fill="#33ABF9") +
  labs(x="Gender",
       y="Number of Victims",
       title="Victims by Gender")
```

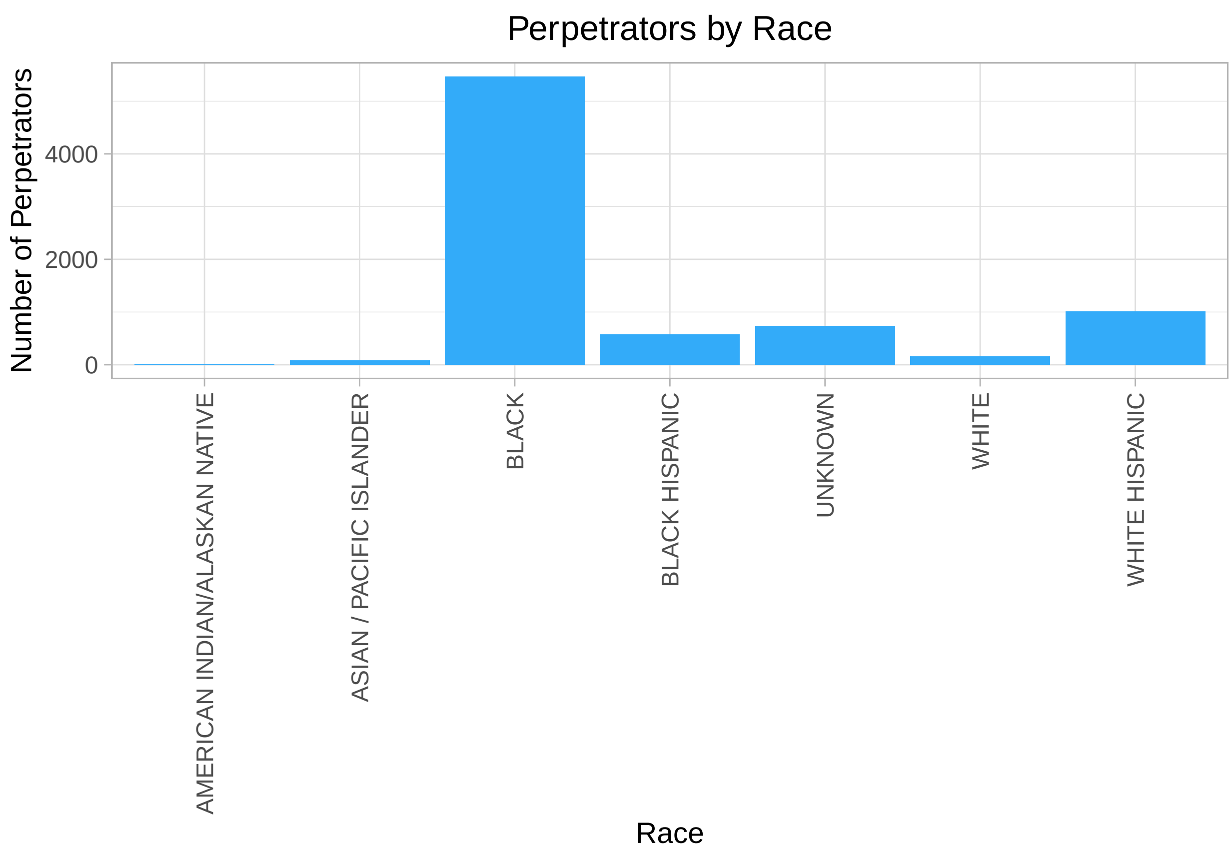



```r
# Table of counts to extract exact values
table(nypd_data_cleaned$VIC_SEX)
```

```
##
##    F    M    U
## 1033 6982    2
```

In both groups, males tend to be more involved in shootings than females or unknown gender. According to the data, males account for 90% and 87% of perpetrators and victims, respectively.

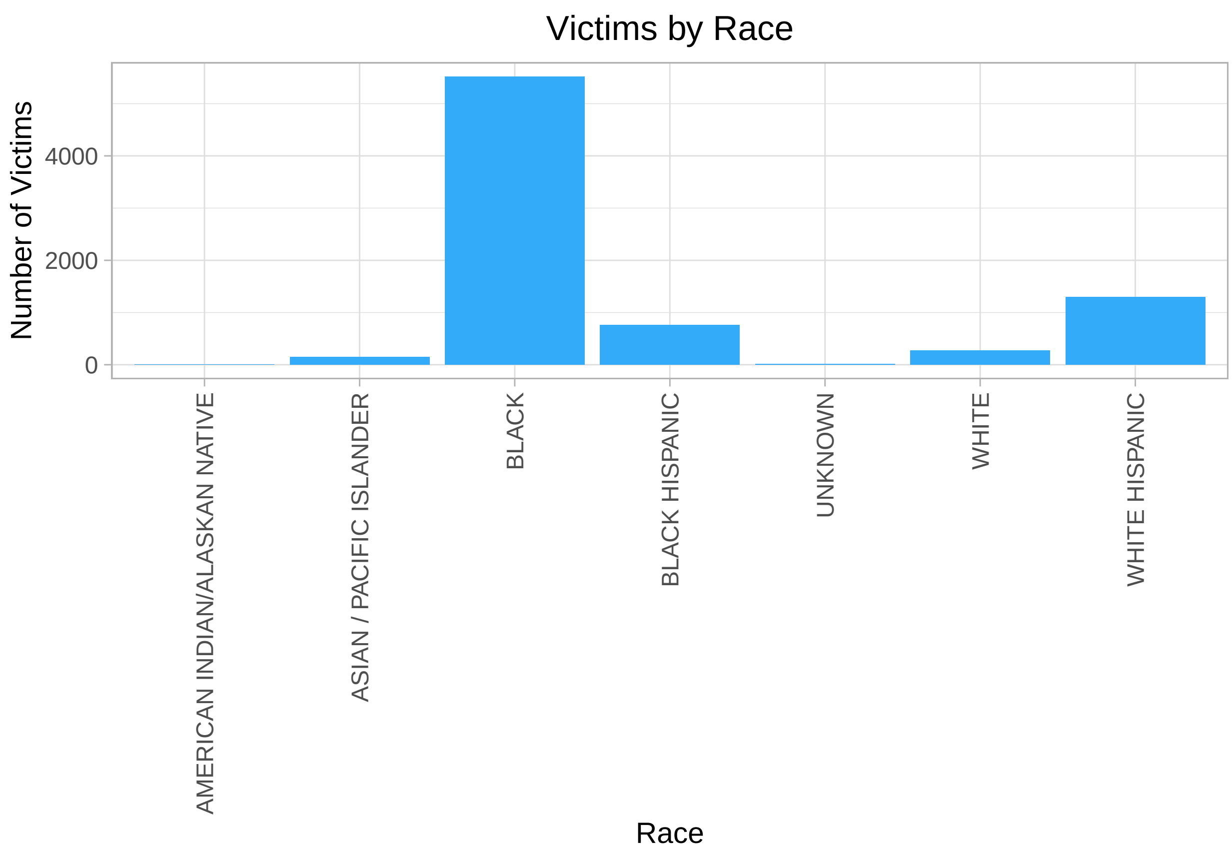
```r
# Race
nypd_data_cleaned %>%
  ggplot(aes(x=PERP_RACE)) +
  theme_light() +
  theme(
    axis.text.x = element_text(angle=90, vjust=0.5, hjust=1),
    plot.title = element_text(hjust=0.5)) +
  geom_bar(fill="#33ABF9") +
  labs(x="Race",
       y="Number of Perpetrators",
       title="Perpetrators by Race")
```




```
# Table of counts to extract exact values
table(nypd_data_cleaned$PERP_RACE)
```

```
##
## AMERICAN INDIAN/ALASKAN NATIVE       ASIAN / PACIFIC ISLANDER
##                              1                             77
##                          BLACK                 BLACK HISPANIC
##                           5466                            573
##                        UNKNOWN                          WHITE
##                            739                            155
##                 WHITE HISPANIC
##                           1006
```

```
nypd_data_cleaned %>%
  ggplot(aes(x=VIC_RACE)) +
  theme_light() +
  theme(
    axis.text.x = element_text(angle=90, vjust=0.5, hjust=1),
    plot.title = element_text(hjust=0.5)) +
  geom_bar(fill="#33ABF9") +
  labs(x="Race",
       y="Number of Victims",
       title="Victims by Race")
```




```
# Table of counts to extract exact values
table(nypd_data_cleaned$VIC_RACE)
```

```
##
## AMERICAN INDIAN/ALASKAN NATIVE       ASIAN / PACIFIC ISLANDER
##                              4                            145
##                          BLACK                 BLACK HISPANIC
##                           5518                            765
##                        UNKNOWN                          WHITE
##                             18                            270
##                 WHITE HISPANIC
##                           1297
```

In most of the shootings, the perpetrators and victims were black. According to the data, black perpetrators account for 68% of all perpetrators, and black victims account for 69% of all victims.

Analysis

Logistic Regression

I will use a logistic regression to identify factors that are significant to whether or not a shooting incident involved a murder.

```r
# Model parameters

# Incident info
occur_date <- nypd_data_cleaned$OCCUR_DATE
occur_time <- nypd_data_cleaned$OCCUR_TIME

# Perpetrator/victim demographics
age_grp_perp <- nypd_data_cleaned$PERP_AGE_GROUP
age_grp_vic <- nypd_data_cleaned$VIC_AGE_GROUP
gender_perp <- nypd_data_cleaned$PERP_SEX
gender_vic <- nypd_data_cleaned$VIC_SEX
race_perp <- nypd_data_cleaned$PERP_RACE
race_vic <- nypd_data_cleaned$VIC_RACE

# Murder (binary)
murder <- nypd_data_cleaned$STATISTICAL_MURDER_FLAG

# Logistic regression model
log_reg <- glm(murder ~
                 occur_date + occur_time + age_grp_perp +                          age_grp_vic + gende
                 race_perp + race_vic,
               data = nypd_data_cleaned,
               family = "binomial")

summary(log_reg)
```

```
##
## Call:
## glm(formula = murder ~ occur_date + occur_time + age_grp_perp +
##     age_grp_vic + gender_perp + gender_vic + race_perp + race_vic,
##     family = "binomial", data = nypd_data_cleaned)
##
## Coefficients:
##                                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)                      -25.87512  586.39369  -0.044 0.964804
## occur_dateFeb                      0.03903    0.14274   0.273 0.784507
## occur_dateMar                     -0.22579    0.13643  -1.655 0.097918 .
## occur_dateApr                     -0.11213    0.13379  -0.838 0.401954
## occur_dateMay                      0.02099    0.12892   0.163 0.870664
## occur_dateJun                     -0.25613    0.13178  -1.944 0.051933 .
## occur_dateJul                     -0.12154    0.13108  -0.927 0.353807
## occur_dateAug                     -0.13274    0.12984  -1.022 0.306628
## occur_dateSep                      0.04299    0.13102   0.328 0.742829
## occur_dateOct                     -0.10112    0.13505  -0.749 0.453989
## occur_dateNov                     -0.15578    0.14295  -1.090 0.275828
## occur_dateDec                      0.08310    0.13729   0.605 0.544996
## occur_time1:00 AM                  0.05765    0.15054   0.383 0.701746
## occur_time2:00 AM                 -0.01410    0.15905  -0.089 0.929349
## occur_time3:00 AM                 -0.09323    0.16328  -0.571 0.568025
## occur_time4:00 AM                  0.14686    0.16862   0.871 0.383789
## occur_time5:00 AM                  0.27306    0.19796   1.379 0.167785
## occur_time6:00 AM                  0.19441    0.23383   0.831 0.405735
## occur_time7:00 AM                  0.49848    0.27015   1.845 0.065015 .
```

```
## occur_time8:00 AM                  0.24503    0.28636   0.856 0.392171
## occur_time9:00 AM                  0.05805    0.26514   0.219 0.826704
## occur_time10:00 AM                 0.05516    0.24696   0.223 0.823255
## occur_time11:00 AM                 0.11070    0.22093   0.501 0.616317
## occur_time12:00 PM                 0.06422    0.20811   0.309 0.757635
## occur_time1:00 PM                  0.14949    0.19026   0.786 0.432043
## occur_time2:00 PM                  0.06316    0.17369   0.364 0.716145
## occur_time3:00 PM                 -0.01884    0.18227  -0.103 0.917664
## occur_time4:00 PM                 -0.26546    0.18131  -1.464 0.143160
## occur_time5:00 PM                  0.17722    0.16633   1.065 0.286677
## occur_time6:00 PM                  0.10643    0.16295   0.653 0.513680
## occur_time7:00 PM                  0.13869    0.15747   0.881 0.378434
## occur_time8:00 PM                 -0.13023    0.16543  -0.787 0.431140
## occur_time9:00 PM                  0.04196    0.14980   0.280 0.779408
## occur_time10:00 PM                 0.16736    0.14821   1.129 0.258797
## occur_time11:00 PM                -0.03698    0.15101  -0.245 0.806547
## age_grp_perp18-24                  0.20419    0.10945   1.866 0.062096 .
## age_grp_perp25-44                  0.44618    0.11093   4.022 5.77e-05 ***
## age_grp_perp45-64                  0.92730    0.15667   5.919 3.24e-09 ***
## age_grp_perp65+                    0.59248    0.35559   1.666 0.095669 .
## age_grp_perpUNKNOWN               -2.61056    0.27720  -9.418  < 2e-16 ***
## age_grp_vic18-24                   0.27645    0.11012   2.511 0.012056 *
## age_grp_vic25-44                   0.41469    0.10866   3.816 0.000135 ***
## age_grp_vic45-64                   0.37164    0.13928   2.668 0.007623 **
## age_grp_vic65+                     1.29420    0.26518   4.880 1.06e-06 ***
## age_grp_vicUNKNOWN                 0.46300    0.48574   0.953 0.340496
## gender_perpM                      -0.19780    0.15266  -1.296 0.195086
## gender_perpU                       1.67694    0.43863   3.823 0.000132 ***
## gender_vicM                       -0.06588    0.08073  -0.816 0.414471
## gender_vicU                      -10.26496  355.01447  -0.029 0.976933
## race_perpASIAN / PACIFIC ISLANDER  13.20737  535.41128   0.025 0.980320
## race_perpBLACK                    12.89405  535.41122   0.024 0.980787
## race_perpBLACK HISPANIC           12.85652  535.41122   0.024 0.980843
## race_perpUNKNOWN                  12.40943  535.41132   0.023 0.981509
## race_perpWHITE                    13.43711  535.41125   0.025 0.979978
## race_perpWHITE HISPANIC           13.06264  535.41122   0.024 0.980536
## race_vicASIAN / PACIFIC ISLANDER   11.72181  239.14936   0.049 0.960908
## race_vicBLACK                     11.57731  239.14927   0.048 0.961389
## race_vicBLACK HISPANIC            11.38430  239.14928   0.048 0.962032
## race_vicUNKNOWN                    9.71267  239.15160   0.041 0.967604
## race_vicWHITE                     11.63844  239.14932   0.049 0.961185
## race_vicWHITE HISPANIC            11.57833  239.14928   0.048 0.961386
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 8695.0  on 8016  degrees of freedom
## Residual deviance: 8006.4  on 7956  degrees of freedom
## AIC: 8128.4
##
## Number of Fisher Scoring iterations: 12
```

Interpreting Results

According to the model summary, murders appear to be strongly associated with age group, for both perpetrators and victims. The results suggest that perpetrators in the age groups `45-64` (**p-value = 3.24e-09 and z-value = 5.919**) have a strong positive correlation with a murder occurring. Victims in the age groups `65+` (**p-value = 1.06e-06 and z-value = 4.880**) have a strong positive correlation with murder. My interpretation of these results is that murders are most common when the victims are older and when perpetrators are roughly middle-aged.

In my initial data exploration, I concluded that most of the shooting incidents involved black, male perpetrators and victims. However, the logistic regression model does not appear to suggest any significant correlation between race/gender and murder. The dates and times of shootings also do not appear to be factors strongly associated with murder.

Bias

My source of bias is how I define murder. In legal practice, there are distinctions between murder and manslaughter, namely intent and the circumstances surrounding them. To avoid introducing my biases in the analysis, I treated murder as death of victim in a shooting. There is no additional context provided (intent of perpetrator, retaliation from victim), so the most logical choice to me was to define murder as victim death.

Conclusion

Referring back to my research questions:

- Where and when are shootings most likely to occur?
- What are the demographics of perpetrators and victims?
- Which factors are correlated with incidents that resulted in murders?

Shootings appear to occur mostly in public housing and apartments, likely due to lack of obstruction by police and incentives to commit associated crimes like robbery and burglary. Evenings and early morning immediately after midnight are the most common times for shootings to occur, possibly due to lower activity (both police and victim). The most common demographic among both perpetrators and victims is black male between the ages 18-44.

Based on these observations, a suggestion I could make is to allocate more resources for home security. Multi-family homes and apartment complexes may benefit from extra security guards and monitoring systems, since these locations are where shootings occur the most. Police departments may want to consider allocating resources to monitor public activity during late hours. Additionally, dispatch and medical services may want to consider prioritizing victims above the age of 65, since this age group is the most likely to be involved in murders.

Most factors related to the number of shooting incidents are not very strong predictors of whether or not a murder was in involved. Only perpetrator and victim age were suggested to be strongly correlated with murder.

Bibliography

- https://catalog.data.gov/dataset/nypd-shooting-incident-data-historic
- https://www.tidyverse.org/packages/
- https://www.citypopulation.de/en/usa/newyorkcity/