NYPD Shooting Incident Project

Shruti Chandrasekaran

2025-02-04

NYPD Shooting Incident Data Report

This report consists of shooting incident that occurred in City Of NewYork from 2006 through the end of the previous calendar year. This data is obtained from **DATA.GOV**, where dataset is intended for public access and use.

Libraries

The library used in this project for analyzing and visualizing data is tidyverse and lubridate. By installing the tidyverse package and loading it using library(tidyverse), I will be able to use dplyr for data manipulation and ggplot2 for visualization. The lubridate package helps with date and time conversions.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(lubridate)
```

Importing Data

The data is import from the above mentioned site, the csv file consist of details about the shooting occured in the city of new york.

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

```
summary(nypd_incident)
```

```
##   INCIDENT_KEY        OCCUR_DATE         OCCUR_TIME            BORO
##  Min.   :  9953245   Length:28562       Length:28562       Length:28562
##  1st Qu.: 65439914   Class :character   Class :character   Class :character
##  Median : 92711254   Mode  :character   Mode  :character   Mode  :character
##  Mean   :127405824
##  3rd Qu.:203131993
##  Max.   :279758069
##
##  LOC_OF_OCCUR_DESC     PRECINCT     JURISDICTION_CODE LOC_CLASSFCTN_DESC
##  Length:28562       Min.   :  1.0   Min.   :0.0000    Length:28562
##  Class :character   1st Qu.: 44.0   1st Qu.:0.0000    Class :character
##  Mode  :character   Median : 67.0   Median :0.0000    Mode  :character
##                     Mean   : 65.5   Mean   :0.3219
##                     3rd Qu.: 81.0   3rd Qu.:0.0000
##                     Max.   :123.0   Max.   :2.0000
##                                     NA's   :2
##  LOCATION_DESC      STATISTICAL_MURDER_FLAG PERP_AGE_GROUP
##  Length:28562       Length:28562            Length:28562
##  Class :character   Class :character        Class :character
##  Mode  :character   Mode  :character        Mode  :character
##
##
##
##
##    PERP_SEX           PERP_RACE          VIC_AGE_GROUP        VIC_SEX
##  Length:28562       Length:28562       Length:28562       Length:28562
##  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:28562       Min.   : 914928   Min.   :125757   Min.   :40.51
##  Class :character   1st Qu.:1000068   1st Qu.:182912   1st Qu.:40.67
##  Mode  :character   Median :1007772   Median :194901   Median :40.70
##                     Mean   :1009424   Mean   :208380   Mean   :40.74
##                     3rd Qu.:1016807   3rd Qu.:239814   3rd Qu.:40.82
##                     Max.   :1066815   Max.   :271128   Max.   :40.91
##                                                        NA's   :59
##    Longitude         Lon_Lat
##  Min.   :-74.25   Length:28562
##  1st Qu.:-73.94   Class :character
##  Median :-73.92   Mode  :character
##  Mean   :-73.91
##  3rd Qu.:-73.88
##  Max.   :-73.70
##  NA's   :59
```

```
head(nypd_incident)
```

```
##   INCIDENT_KEY OCCUR_DATE OCCUR_TIME     BORO LOC_OF_OCCUR_DESC PRECINCT
## 1    231974218 08/09/2021   01:06:00    BRONX                        40
## 2    177934247 04/07/2018   19:48:00 BROOKLYN                        79
```

```
## 3    255028563 12/02/2022   22:57:00    BRONX           OUTSIDE       47
## 4     25384540 11/19/2006   01:50:00 BROOKLYN                         66
## 5     72616285 05/09/2010   01:58:00    BRONX                         46
## 6     85875439 07/22/2012   21:35:00    BRONX                         42
##   JURISDICTION_CODE LOC_CLASSFCTN_DESC            LOCATION_DESC
## 1                 0
## 2                 0
## 3                 0             STREET            GROCERY/BODEGA
## 4                 0                                   PVT HOUSE
## 5                 0                     MULTI DWELL - APT BUILD
## 6                 2                   MULTI DWELL - PUBLIC HOUS
##   STATISTICAL_MURDER_FLAG PERP_AGE_GROUP PERP_SEX      PERP_RACE VIC_AGE_GROUP
## 1                   false                                              18-24
## 2                    true          25-44        M WHITE HISPANIC         25-44
## 3                   false         (null)   (null)         (null)         25-44
## 4                    true        UNKNOWN        U        UNKNOWN         18-24
## 5                    true          25-44        M          BLACK           <18
## 6                   false          18-24        M          BLACK         18-24
##   VIC_SEX VIC_RACE X_COORD_CD Y_COORD_CD Latitude Longitude
## 1       M    BLACK  1006343.0   234270.0 40.80967 -73.92019
## 2       M    BLACK  1000082.9   189064.7 40.68561 -73.94291
## 3       M    BLACK  1020691.0   257125.0 40.87235 -73.86823
## 4       M    BLACK   985107.3   173349.8 40.64249 -73.99691
## 5       F    BLACK  1009853.5   247502.6 40.84598 -73.90746
## 6       M    BLACK  1011046.7   239814.2 40.82488 -73.90318
##                                     Lon_Lat
## 1  POINT (-73.92019278899994 40.80967347200004)
## 2 POINT (-73.94291302299996 40.685609672000055)
## 3                  POINT (-73.868233 40.872349)
## 4 POINT (-73.99691224999998 40.642489932000046)
## 5  POINT (-73.90746098599993 40.84598358900007)
## 6  POINT (-73.90317908399999 40.82487781900005)
```

Data Cleaning

Once the data is imported we will proceed with the data cleaning process. The summary function provides a quick overview of the key characteristics of the data which will help us with the data cleaning process.

- The column names and data types for for OCCUR_DATE and OCCUR_TIME is changed to date and time respectively.Additionally, YEAR and MONTH columns are included for further analysis.
- Unnecessary columns are removed, and some columns are converted to factors for better analysis.
- To remove duplicate records

```
nypd_incident <- nypd_incident %>%
 rename(DATE = OCCUR_DATE,
        TIME = OCCUR_TIME) %>%
 mutate(DATE = mdy(DATE),
        TIME = hms(TIME),
        YEAR = year(DATE),
        MONTHS = month(DATE, label = TRUE))
```

```
nypd_incident <- nypd_incident %>%
  select(DATE,TIME,BORO,PRECINCT,STATISTICAL_MURDER_FLAG,
           PERP_AGE_GROUP,PERP_SEX,PERP_RACE, VIC_AGE_GROUP, VIC_SEX,VIC_RACE, YEAR, MONTHS)
```

```
nypd_incident$BORO = as.factor(nypd_incident$BORO)
nypd_incident$PERP_AGE_GROUP = as.factor(nypd_incident$PERP_AGE_GROUP)
nypd_incident$PERP_SEX = as.factor(nypd_incident$PERP_SEX)
nypd_incident$PERP_RACE = as.factor(nypd_incident$PERP_RACE)
nypd_incident$VIC_AGE_GROUP = as.factor(nypd_incident$VIC_AGE_GROUP)
nypd_incident$VIC_SEX = as.factor(nypd_incident$VIC_SEX)
nypd_incident$VIC_RACE = as.factor(nypd_incident$VIC_RACE)
```

```
nypd_incident <- unique(nypd_incident)
```

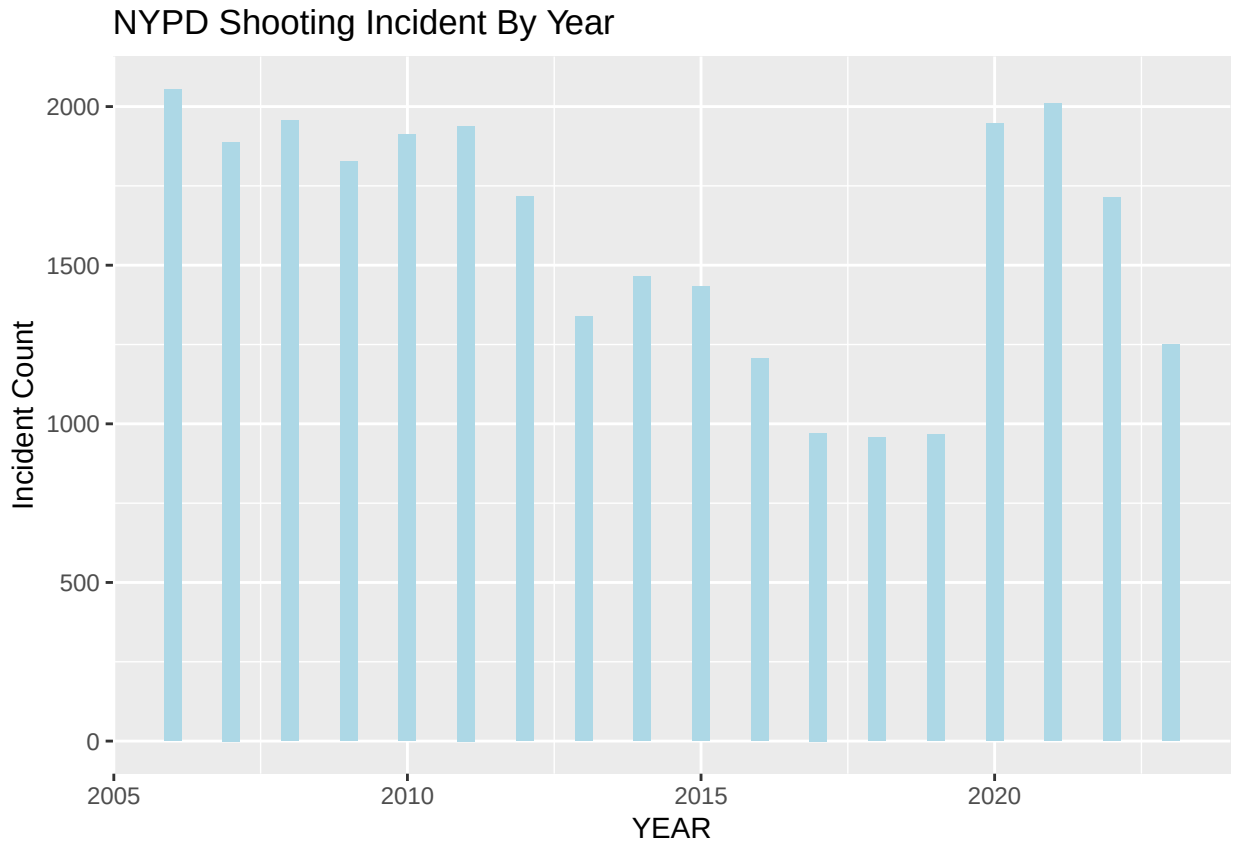
```
summary(nypd_incident)
```

```
##       DATE                 TIME                                      BORO
##  Min.   :2006-01-01   Min.   :0S                          BRONX        : 8376
##  1st Qu.:2009-09-04   1st Qu.:3H 30M 0S                   BROOKLYN     :11345
##  Median :2013-09-20   Median :15H 15M 0S                  MANHATTAN    : 3762
##  Mean   :2014-06-07   Mean   :12H 44M 16.30195021182595   QUEENS       : 4271
##  3rd Qu.:2019-09-29   3rd Qu.:20H 45M 0S                  STATEN ISLAND:  807
##  Max.   :2023-12-29   Max.   :23H 59M 0S
##
##     PRECINCT     STATISTICAL_MURDER_FLAG PERP_AGE_GROUP   PERP_SEX
##  Min.   :  1.0   Length:28561                    :9344         : 9310
##  1st Qu.: 44.0   Class :character        18-24   :6438   (null): 1141
##  Median : 67.0   Mode  :character        25-44   :6040   F     :  444
##  Mean   : 65.5                          UNKNOWN:3148   M     :16167
##  3rd Qu.: 81.0                          <18     :1682   U     : 1499
##  Max.   :123.0                          (null)  :1141
##                                         (Other): 768
##           PERP_RACE     VIC_AGE_GROUP   VIC_SEX
##  BLACK         :11902   <18     : 2954   F: 2760
##                : 9310   1022    :    1   M:25789
##  WHITE HISPANIC: 2510   18-24   :10384   U:   12
##  UNKNOWN       : 1837   25-44   :12972
##  BLACK HISPANIC: 1392   45-64   : 1981
##  (null)        : 1141   65+     :  205
##  (Other)       :  469   UNKNOWN:   64
##                              VIC_RACE          YEAR          MONTHS
##  AMERICAN INDIAN/ALASKAN NATIVE:   11   Min.   :2006   Jul    : 3390
##  ASIAN / PACIFIC ISLANDER      :  440   1st Qu.:2009   Aug    : 3264
##  BLACK                         :20234   Median :2013   Jun    : 2959
##  BLACK HISPANIC                : 2795   Mean   :2014   May    : 2682
##  UNKNOWN                       :   70   3rd Qu.:2019   Sep    : 2677
##  WHITE                         :  728   Max.   :2023   Oct    : 2378
##  WHITE HISPANIC                : 4283                  (Other):11211
```

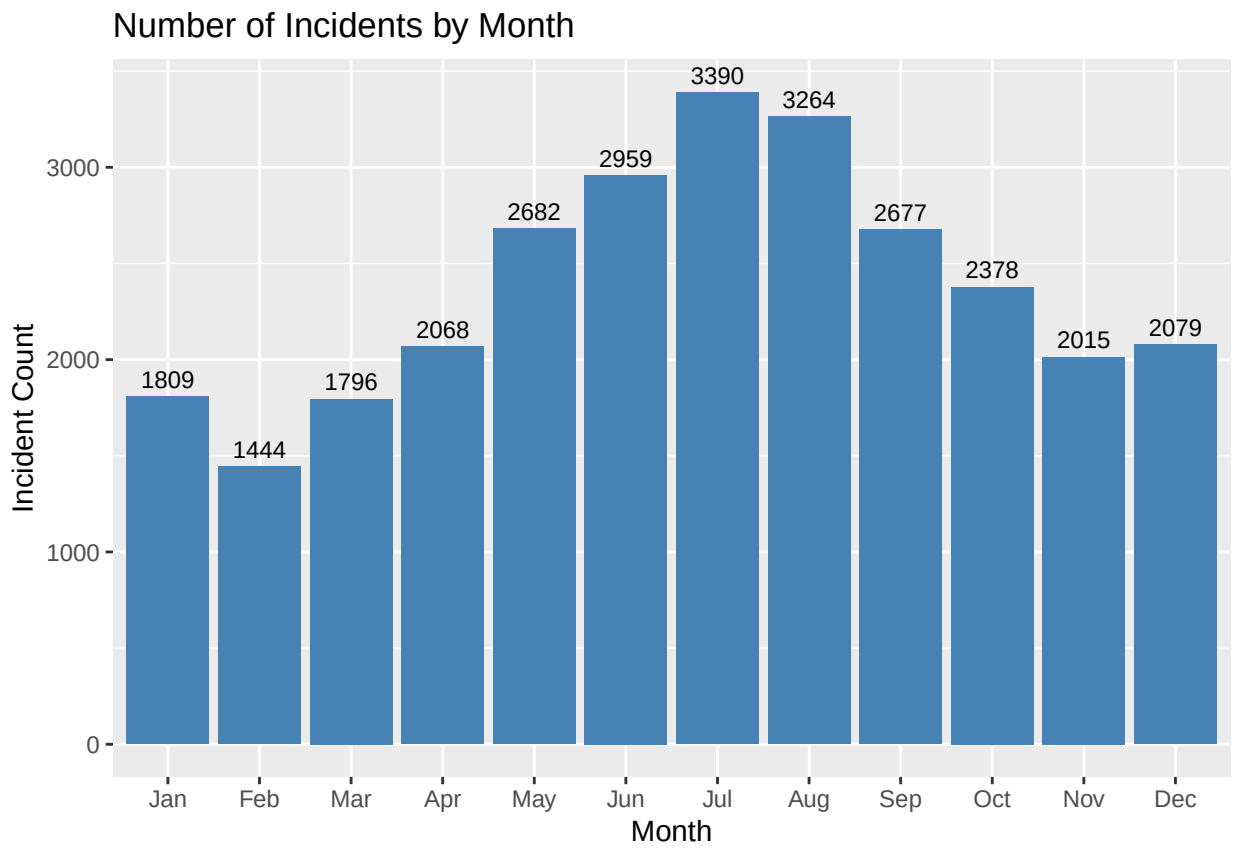
Data Analysis And Visualization

Once the data cleaning is done the next step is to proceed with data analysis and visualization of data.

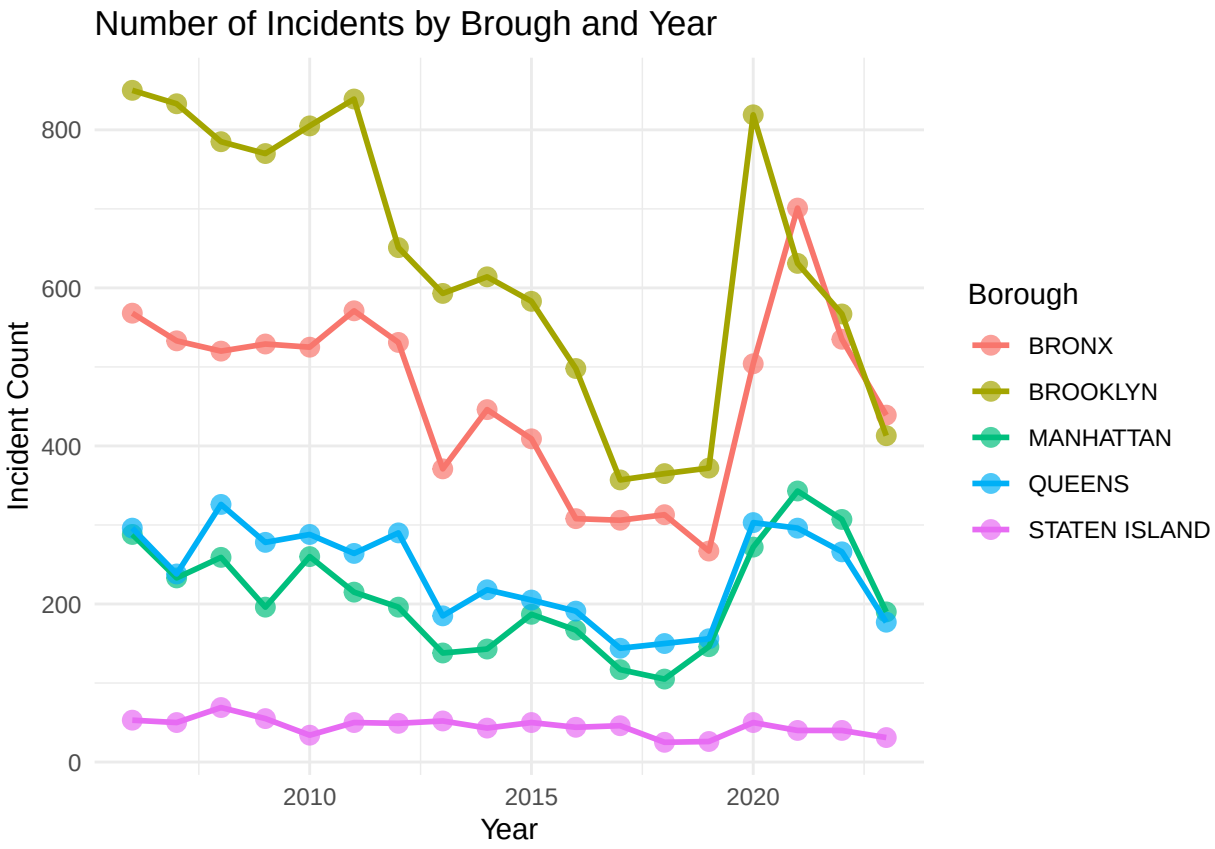
```r
nypd_boro_year <- nypd_incident %>%
  group_by(BORO, YEAR) %>%
  summarize(incident_count = n(), .groups = "drop")

ggplot(nypd_boro_year, aes(x = YEAR, y = incident_count)) +
  geom_bar(stat = "identity", fill = "lightblue", width =0.3) +
  labs(title = "NYPD Shooting Incident By Year", x = "YEAR", y = "Incident Count")
```



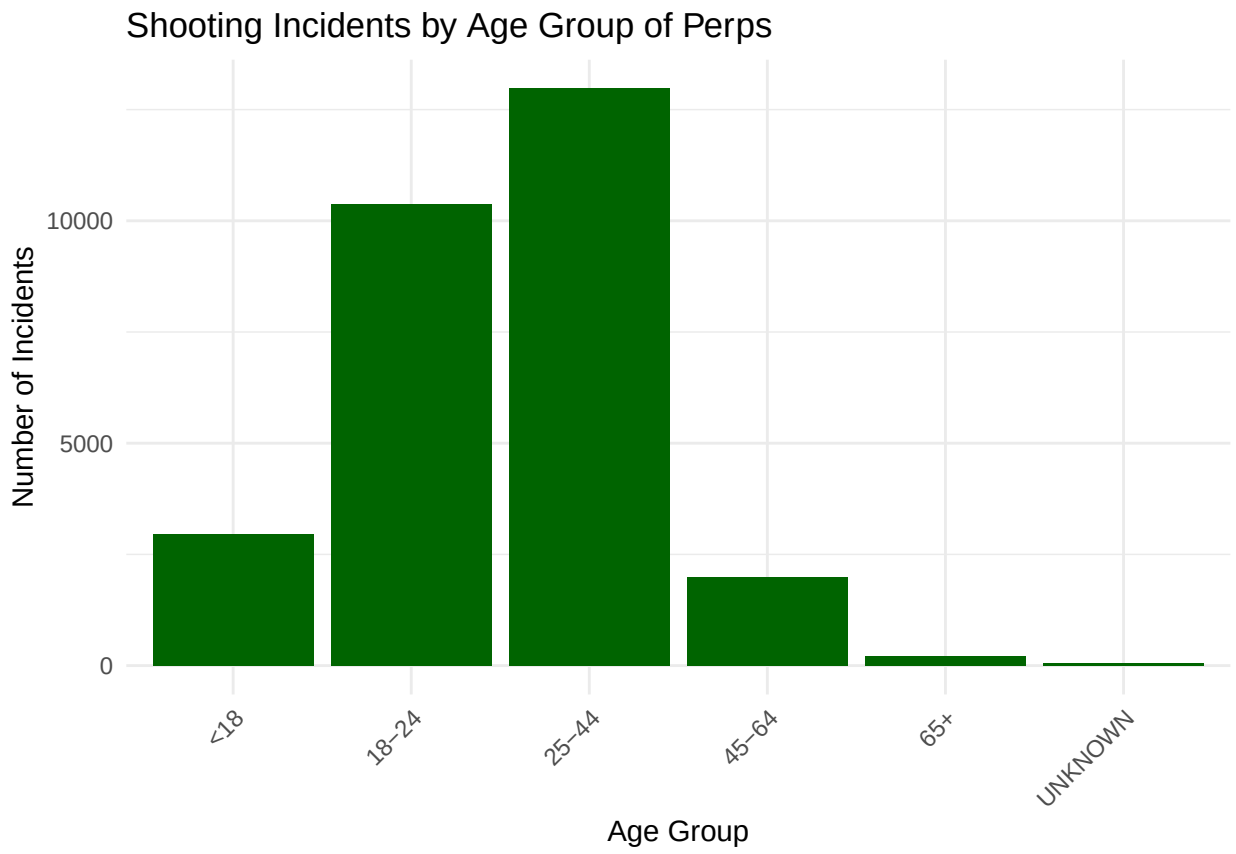

```r
incident_by_months <- nypd_incident %>%
  group_by(MONTHS) %>%
  summarise(incident_count = n()) %>%
  ungroup()

ggplot(incident_by_months, aes(x = MONTHS, y = incident_count)) +
  geom_bar(stat = "identity", fill = "steelblue") +  # Adds color for clarity
  geom_text(aes(label = incident_count),
            vjust = -0.5, color = "black", size = 3) +
  labs(title = "Number of Incidents by Month", x = "Month", y = "Incident Count")
```




```
nypd_last_5years <- nypd_incident %>%
  group_by(BORO, YEAR) %>%

  summarize(incident_count = n(), .groups = "drop")

 ggplot(nypd_last_5years, aes(x = YEAR, y = incident_count, color = BORO)) +
  geom_point(size = 3, alpha = 0.7) +
  geom_line(aes(group = BORO), linewidth = 1) +
  labs(title = "Number of Incidents by Brough and Year",
       x = "Year",
       y = "Incident Count",
       color = "Borough") +
  theme_minimal()
```



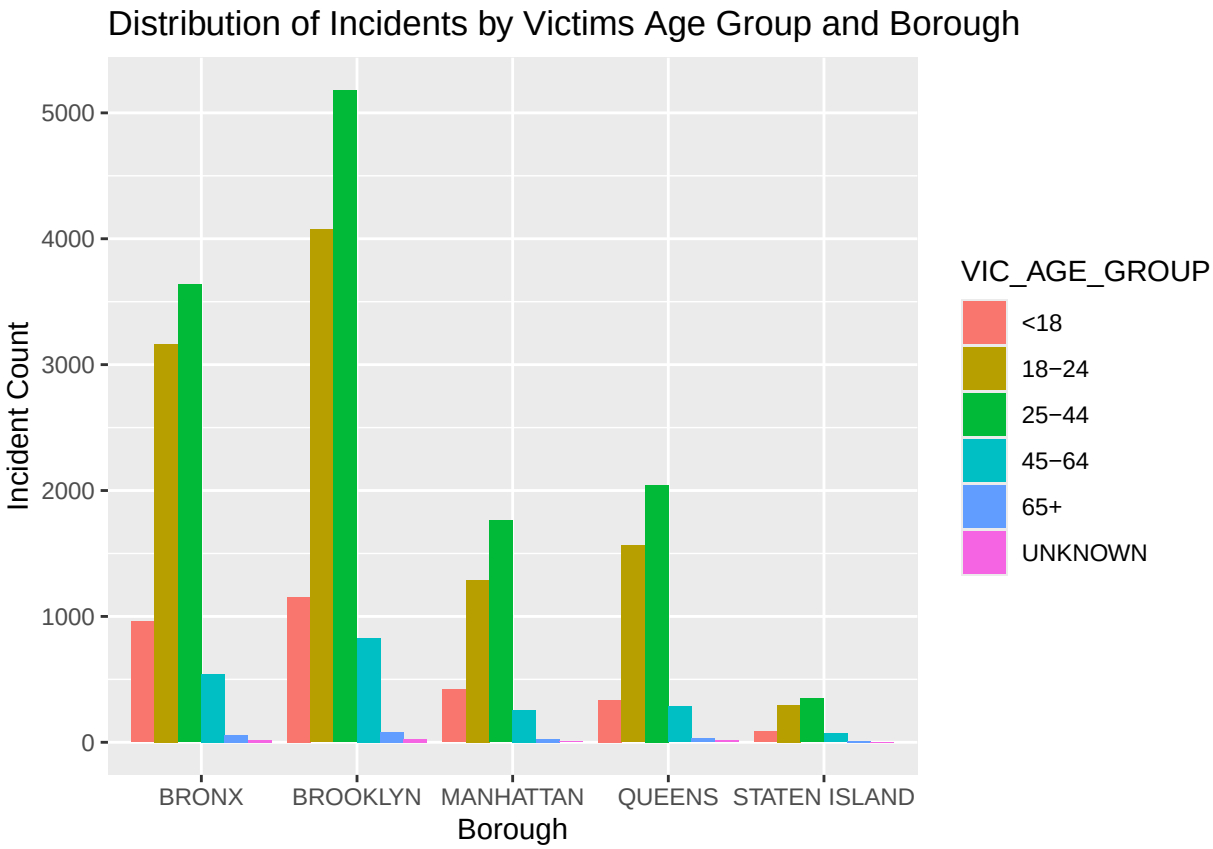

```
incident_by_age <- filter(nypd_incident, VIC_AGE_GROUP!= 1022 )

ggplot(incident_by_age, aes(x = VIC_AGE_GROUP)) +
  geom_bar(fill = "darkgreen") +
  labs(title = "Shooting Incidents by Age Group of Perps",
       x = "Age Group",
       y = "Number of Incidents") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
nypd_vic <- incident_by_age %>%
  select(BORO, VIC_AGE_GROUP)
colSums(is.na(nypd_vic))
```

```
##           BORO VIC_AGE_GROUP
##              0             0
```

```
ggplot(nypd_vic, aes(x = BORO, fill = VIC_AGE_GROUP)) +
  geom_bar(position = "dodge") +
  labs(title = "Distribution of Incidents by Victims Age Group and Borough",
       x = "Borough", y = "Incident Count")
```



Data Modelling

Once data analysis and visualization are completed, we will proceed with data modeling, which is a mathematical representation used to identify patterns, relationships, and dependencies. This enables us to make predictions or classifications based on input data. Since Brooklyn had the highest number of incidents, we focus on it for modeling.

```
brooklyn_shootings <- nypd_incident %>%
filter(BORO=='BROOKLYN')%>%
  group_by(YEAR) %>%
  summarise(number_of_crimes=n())


mod <- lm(number_of_crimes ~YEAR, brooklyn_shootings)

shooting_pred <- mutate(brooklyn_shootings, pred = predict(mod))
```

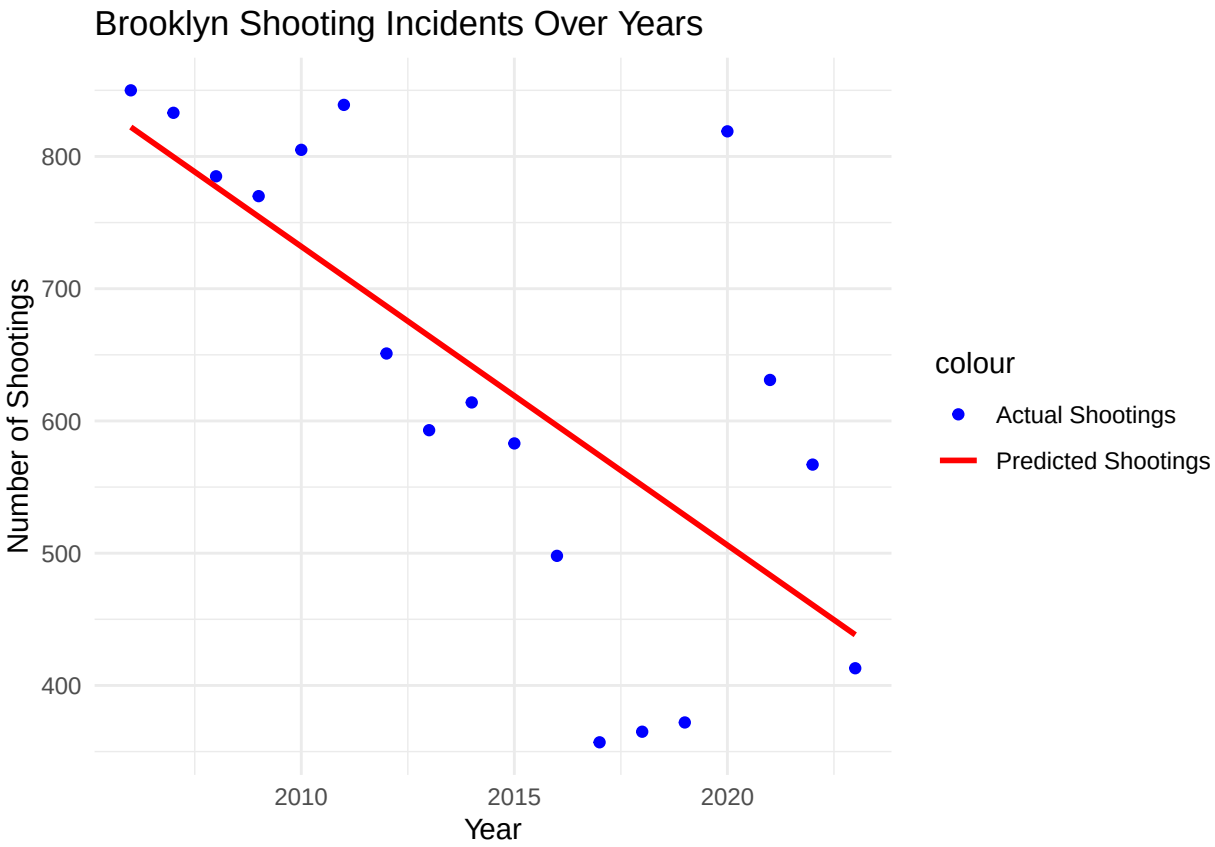
```
summary(mod)
```

```
##
## Call:
## lm(formula = number_of_crimes ~ YEAR, data = brooklyn_shootings)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -216.820  -62.361   -8.695   63.161  312.929
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 46123.883  12112.370   3.808  0.00155 **
## YEAR          -22.583      6.013  -3.756  0.00173 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 132.3 on 16 degrees of freedom
## Multiple R-squared:  0.4686, Adjusted R-squared:  0.4354
## F-statistic: 14.11 on 1 and 16 DF,  p-value: 0.001726
```

```
ggplot(shooting_pred, aes(x = YEAR)) +
  geom_point(aes(y = number_of_crimes, color = "Actual Shootings")) +  # Actual data points
  geom_line(aes(y = pred, color = "Predicted Shootings"), size = 1) +  # Regression line
  labs(title = "Brooklyn Shooting Incidents Over Years",
       x = "Year",
       y = "Number of Shootings") +
  scale_color_manual(values = c("Actual Shootings" = "blue", "Predicted Shootings" = "red")) +
  theme_minimal()
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```



Conclusion and Bias Identification

The analysis of NYPD shooting incidents from 2006 to 2024 highlights key trends and patterns in gun violence across New York City, with a particular focus on Brooklyn due to its consistently high number of reported incidents. Our linear regression model suggests a gradual decline in shooting incidents in Brooklyn, with an estimated reduction of approximately 23 incidents per year. I found that the NYPD dataset is racially biased and unfair treatment towards the minority.