NYPDShootingIncidentData

Rupa

2023-07-03

NYPD Shooting Incident

Introduction

This report is related to the New York Shooting Incidents from 2006 to 2022. This report helps understand the pattern and trends in the shootings and the Boroughs that the shootings are most seen.

Import and Read Data

```r
# Get the NYPD Shooting Incident Data:


ny_data <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv"


ny_cases <- read_csv (ny_data)
```

```
## Rows: 27312 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.
```

Tidy and Transform the data

```r
# Clean the data keeping only columns needed for the analysis

ny_cases <- ny_cases %>% select (-c(JURISDICTION_CODE, OCCUR_TIME, LOC_OF_OCCUR_DESC,
                                    LOC_CLASSFCTN_DESC, LOCATION_DESC, PERP_AGE_GROUP,
                                    PERP_SEX, PERP_RACE, X_COORD_CD, Y_COORD_CD,
                                    Latitude, Longitude, Lon_Lat))

# Extract Unique Elements

ny_cases <- unique (ny_cases)

# Extract the year from the DATE

ny_cases$Year <- format(as.Date(ny_cases$OCCUR_DATE, format="%m/%d/%Y"),"%Y")

# Arrange data in Descending Order of Most Shootings

cases_by_boro <- ny_cases %>% group_by(BORO, Year) %>%  summarize (Cases = n())
```

```
## `summarise()` has grouped output by 'BORO'. You can override using the
## `.groups` argument.
# Get the updated list of the cases after cleaning up the data
# and keeping only columns needed for the analysis

cases_by_boro
```

```
## # A tibble: 85 x 3
## # Groups:   BORO [5]
##    BORO  Year  Cases
##    <chr> <chr> <int>
##  1 BRONX 2006    493
##  2 BRONX 2007    451
##  3 BRONX 2008    472
##  4 BRONX 2009    484
##  5 BRONX 2010    487
##  6 BRONX 2011    525
##  7 BRONX 2012    495
##  8 BRONX 2013    341
##  9 BRONX 2014    399
## 10 BRONX 2015    361
## # i 75 more rows
```

Plot the cases by Borough by Year

```
# Plot the number of cases by Year by Borough

cases_by_boro %>% ggplot(aes (x = Year, y = Cases, fill = BORO)) + geom_bar (stat = "identity", widt
```
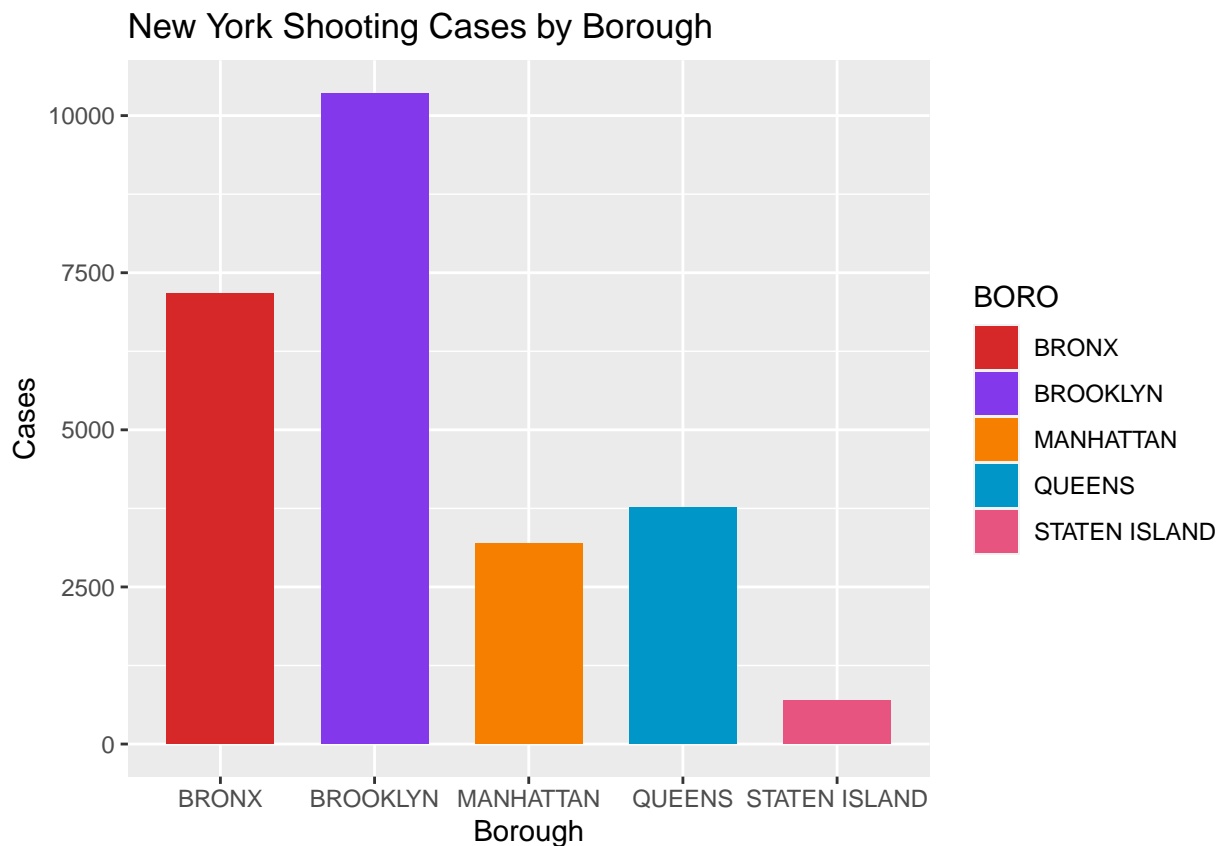

New York Shooting Cases by Borough by Year
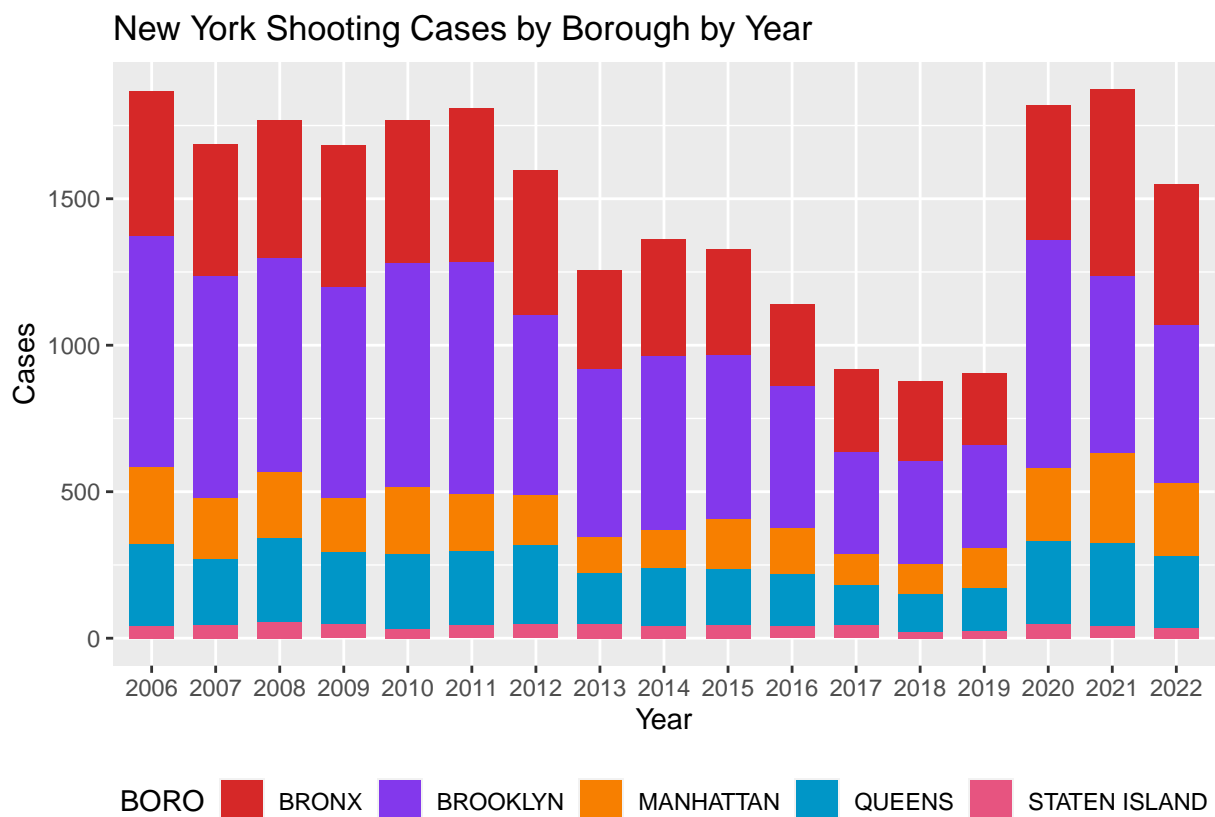
Plot the cases by Borough

```
# Plot the number of cases by Borough

cases_by_boro %>% ggplot(aes (x = BORO, y = Cases, fill = BORO)) + geom_bar (stat = "identity", widt
```

New York Shooting Cases by Borough

Plot the cases by Victim Age Group

```
# Plot the cases by Victim Age Group

cases_by_vic_age_group <- ny_cases %>% filter (VIC_AGE_GROUP != 1022)

cases_by_vic_age_group <- cases_by_vic_age_group %>% filter (VIC_AGE_GROUP != "UNKNOWN")

cases_by_vic_age_group_by_year <- cases_by_vic_age_group %>% group_by(VIC_AGE_GROUP, Year) %>%  summ
```
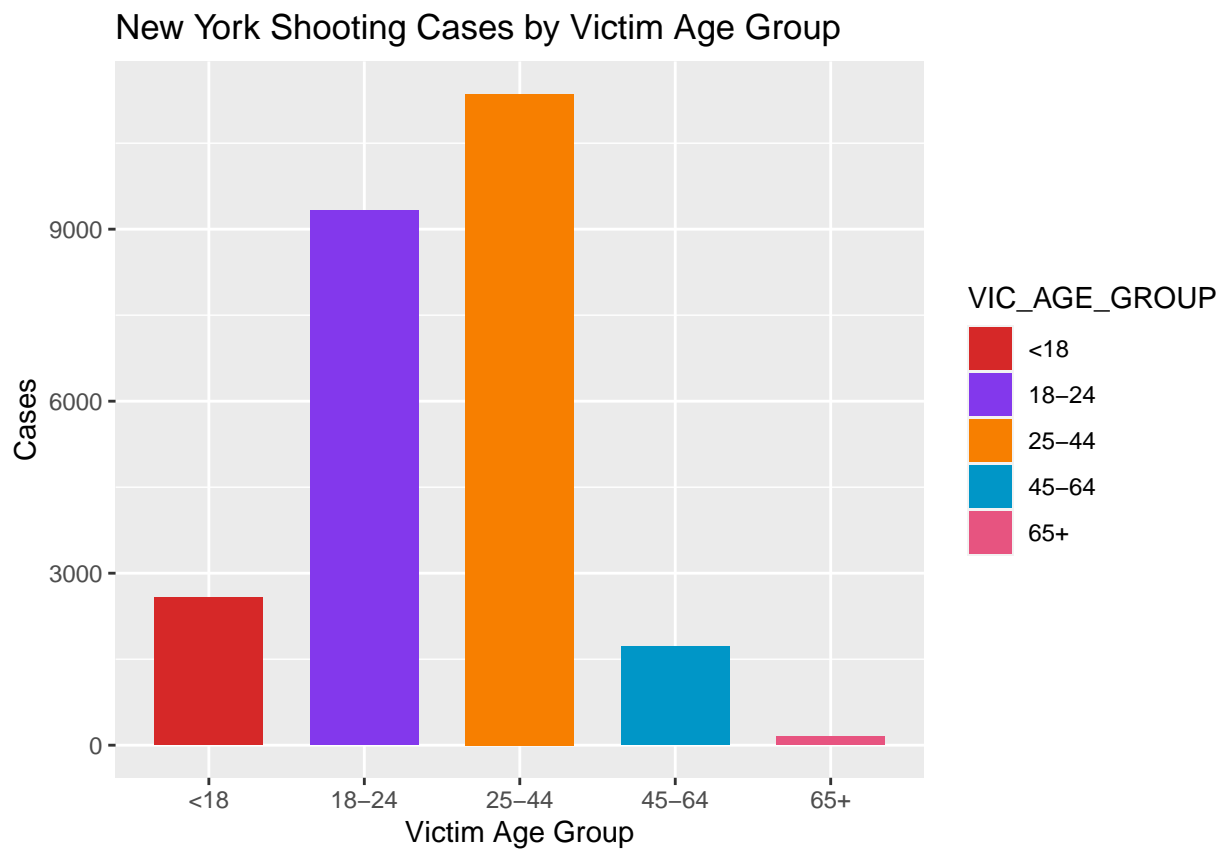
```
## `summarise()` has grouped output by 'VIC_AGE_GROUP'. You can override using the
## `.groups` argument.
```

```
cases_by_vic_age_group_by_year[is.na(cases_by_vic_age_group_by_year) | cases_by_vic_age_group_by_yea

cases_by_vic_age_group_by_year %>% ggplot(aes (x = VIC_AGE_GROUP, y = Cases, fill = VIC_AGE_GROUP))
```
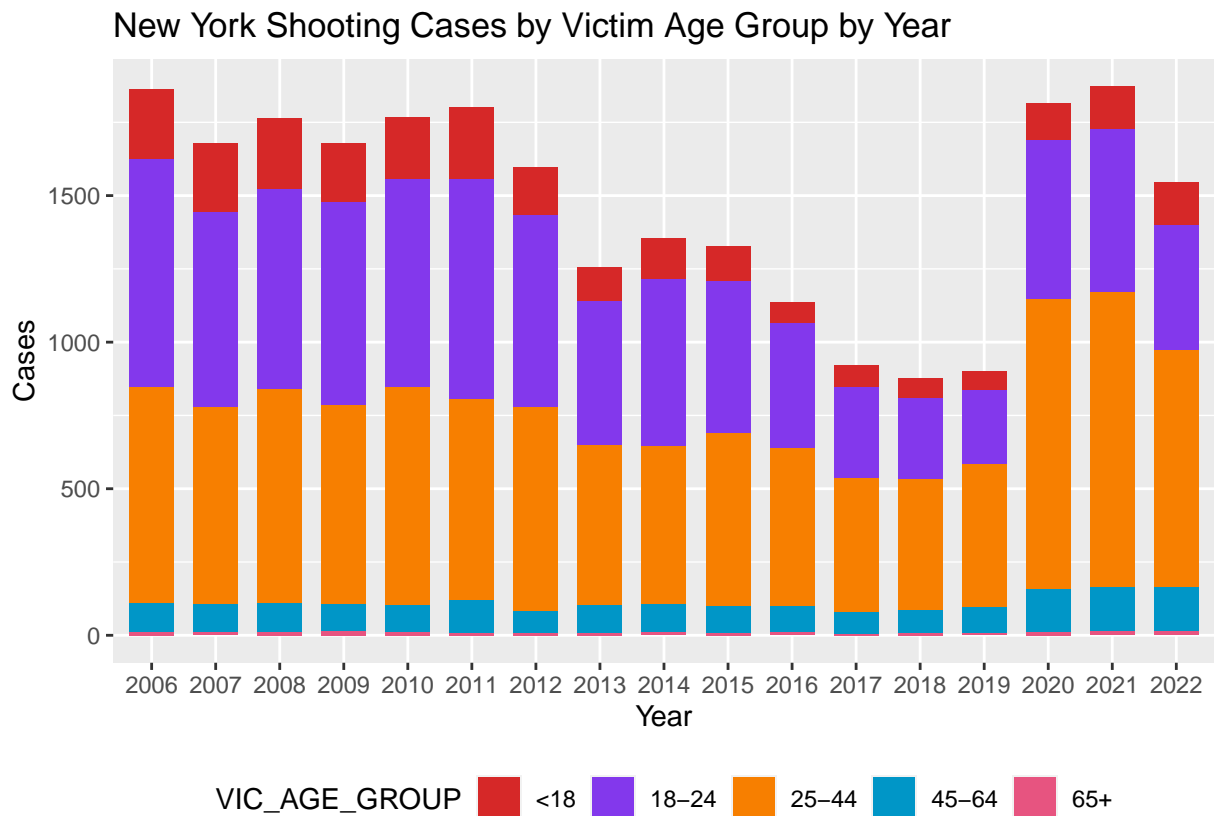
New York Shooting Cases by Victim Age Group

Plot the cases by Victim Age Group by Year

```r
#Plot the cases by Victim Age Group by Year

cases_by_vic_age_group_by_year %>% ggplot(aes (x = Year, y = Cases, fill = VIC_AGE_GROUP)) + geom_ba
```

New York Shooting Cases by Victim Age Group by Year

Plot the cases by Victim Age Group by Borough

```
#Plot the cases by Victim Age Group by Borough

cases_by_vic_by_boro <- cases_by_vic_age_group %>% group_by(Year, VIC_AGE_GROUP, BORO) %>% summarize
```

```
## `summarise()` has grouped output by 'Year', 'VIC_AGE_GROUP'. You can override
## using the `.groups` argument.
```

```
cases_by_vic_by_boro
```

```
## # A tibble: 399 x 4
## # Groups:   Year, VIC_AGE_GROUP [85]
##    Year  VIC_AGE_GROUP BORO          Cases
##    <chr> <chr>         <chr>         <int>
##  1 2006  18-24         BRONX           205
##  2 2006  18-24         BROOKLYN        328
##  3 2006  18-24         MANHATTAN       107
##  4 2006  18-24         QUEENS          122
##  5 2006  18-24         STATEN ISLAND    18
##  6 2006  25-44         BRONX           194
##  7 2006  25-44         BROOKLYN        315
##  8 2006  25-44         MANHATTAN        98
##  9 2006  25-44         QUEENS          113
## 10 2006  25-44         STATEN ISLAND    15
## # i 389 more rows
```

```
cases_by_vic_by_boro[is.na(cases_by_vic_by_boro) | cases_by_vic_by_boro=="Inf"] = NA

cases_by_vic_by_boro %>% ggplot(aes (x = VIC_AGE_GROUP, y = Cases, fill = BORO, label = Cases)) + ge
```

New York Shooting Cases by Victim Age Group by Borough
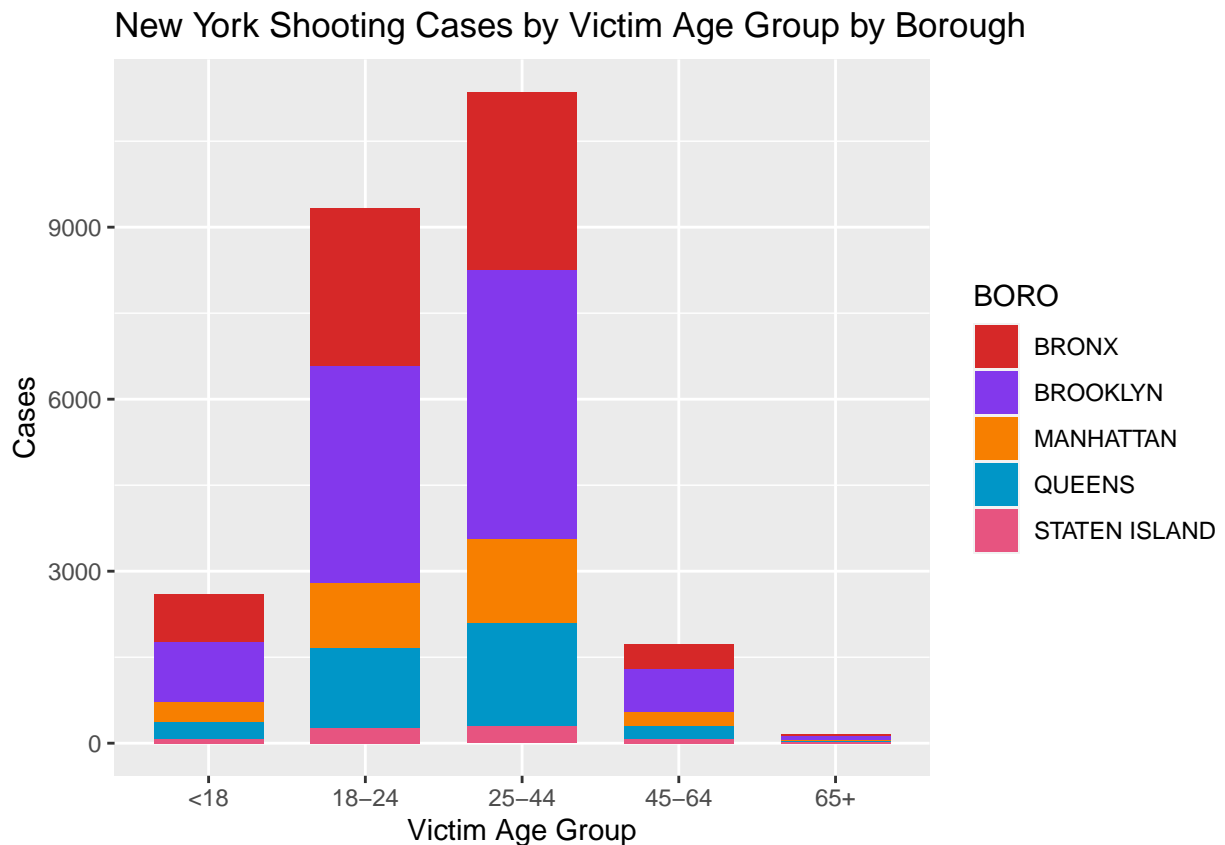
Linear Modeling the Cases by Victim's Age Group and Cases by Borough

```
#Update the Cases by Borough to include Statistical Murder Analysis by Borough

total_cases <- ny_cases %>% summarize (boro_cases = cases_by_boro$Cases, vic_age_cases = cases_by_vi
```

```
## Warning: Returning more (or less) than 1 row per `summarise()` group was deprecated in
## dplyr 1.1.0.
## i Please use `reframe()` instead.
## i When switching from `summarise()` to `reframe()`, remember that `reframe()`
##   always returns an ungrouped data frame and adjust accordingly.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
total_cases
```

```
## # A tibble: 85 x 2
##    boro_cases vic_age_cases
##         <int>         <int>
## 1         493           780
## 2         451           668
## 3         472           684
## 4         484           693
## 5         487           711
## 6         525           752
## 7         495           657
## 8         341           490
## 9         399           573
## 10        361           518
## # i 75 more rows
```

```
lin_mod <- lm (formula = boro_cases ~ vic_age_cases, data = total_cases)

summary (lin_mod)

##
## Call:
## lm(formula = boro_cases ~ vic_age_cases, data = total_cases)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -219.084 -84.980   2.833  91.379 244.014
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  105.80767   18.92869    5.59 2.83e-07 ***
## vic_age_cases  0.64457    0.04628   13.93  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 120.5 on 83 degrees of freedom
## Multiple R-squared:  0.7004, Adjusted R-squared:  0.6968
## F-statistic:   194 on 1 and 83 DF,  p-value: < 2.2e-16

total_cases %>% slice_min (vic_age_cases)

## # A tibble: 1 x 2
##   boro_cases vic_age_cases
##        <int>         <int>
## 1        137             4

total_cases %>% slice_max (vic_age_cases)

## # A tibble: 1 x 2
##   boro_cases vic_age_cases
##        <int>         <int>
## 1        605          1009

total_cases

## # A tibble: 85 x 2
##    boro_cases vic_age_cases
##         <int>         <int>
## 1         493           780
## 2         451           668
## 3         472           684
## 4         484           693
## 5         487           711
## 6         525           752
## 7         495           657
## 8         341           490
## 9         399           573
## 10        361           518
## # i 75 more rows

x_grid <- seq (-250, 250)
#lin_mod

lm_cases_with_pred <- total_cases %>% mutate (pred = predict (lin_mod))

lm_cases_with_pred %>% ggplot () + geom_point (aes (x = boro_cases, y = vic_age_cases), color = "blu
```
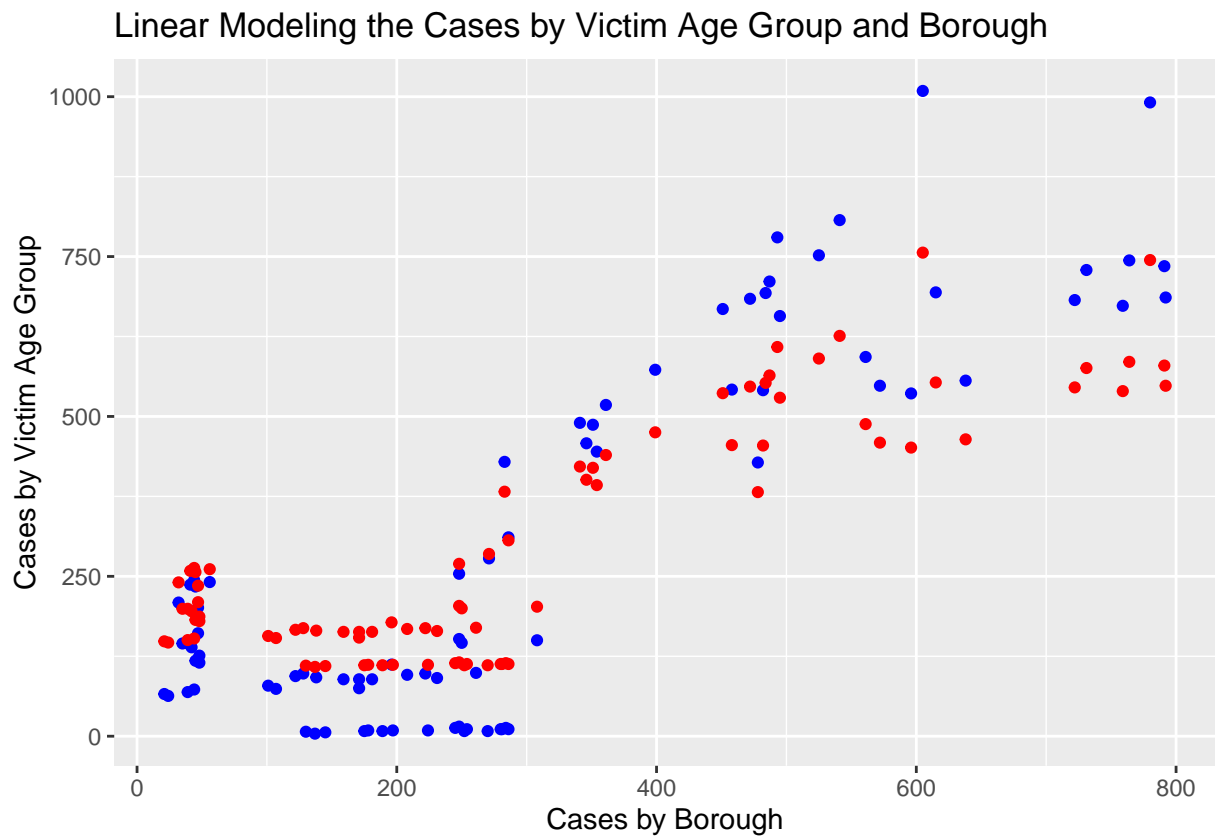
Linear Modeling the Cases by Victim Age Group and Borough

Conclusion

The analysis shows that most of the cases are from Brooklyn and the lowest cases are from Staten Island. The rates had gone down between 2017-2019, but seems to have risen back in 2020. The reason could be because of the lower population of Staten Island compared to Brooklyn. However, not having strict gun laws and seeing the increase in shooting incidents is very concerning.