NYPD Shooting Incidents

2025-03-04

Introduction

This report looks at every shooting incident in NYC that occurred from 2006 to the end of 2023, and, specifically, aims to analyze the demographics of the victims to determine if certain communities are more vulnerable.

Import Data

```
NYPD_url <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
```

```
NYPD_data <- read_csv(NYPD_url)
```

Tidy Data

Eliminate the columns unnecessary for the current analysis.

```
NYPD_data <- NYPD_data %>%
  select(STATISTICAL_MURDER_FLAG, PERP_AGE_GROUP, PERP_SEX, PERP_RACE, VIC_AGE_GROUP, VIC_SEX, VIC_RACE
summarize(NYPD_data)
```
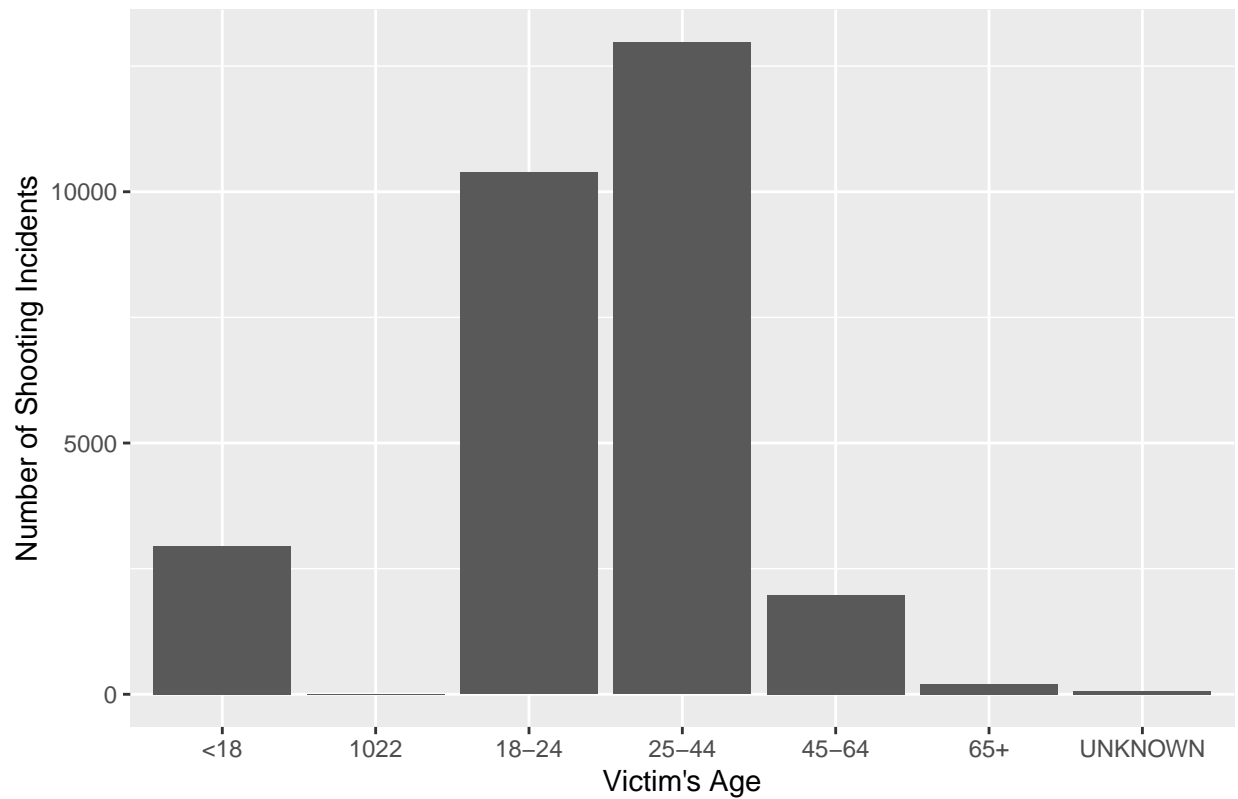
```
## # A tibble: 1 x 0
```

## Visualize and Analyze the Data

Victim's Age

Plot the data based on the victim's age group.

```
NYPD_data %>%
  ggplot(aes(x = VIC_AGE_GROUP)) +
  geom_bar() +
  labs(title = "Shooting Incidents by Victim's Age Group",
      x = "Victim's Age",
      y = "Number of Shooting Incidents")
```

Shooting Incidents by Victim's Age Group


```r
NYPD_data %>%
  filter(VIC_AGE_GROUP == '25-44') %>%
  ggplot(aes(x = PERP_AGE_GROUP)) +
  geom_bar() +
  labs(title = "Shooting Incidents where victim is 25-44",
       x = "Perp's Age",
       y = "Number of Shooting Incidents") +
  guides(x = guide_axis(angle = 45))
```
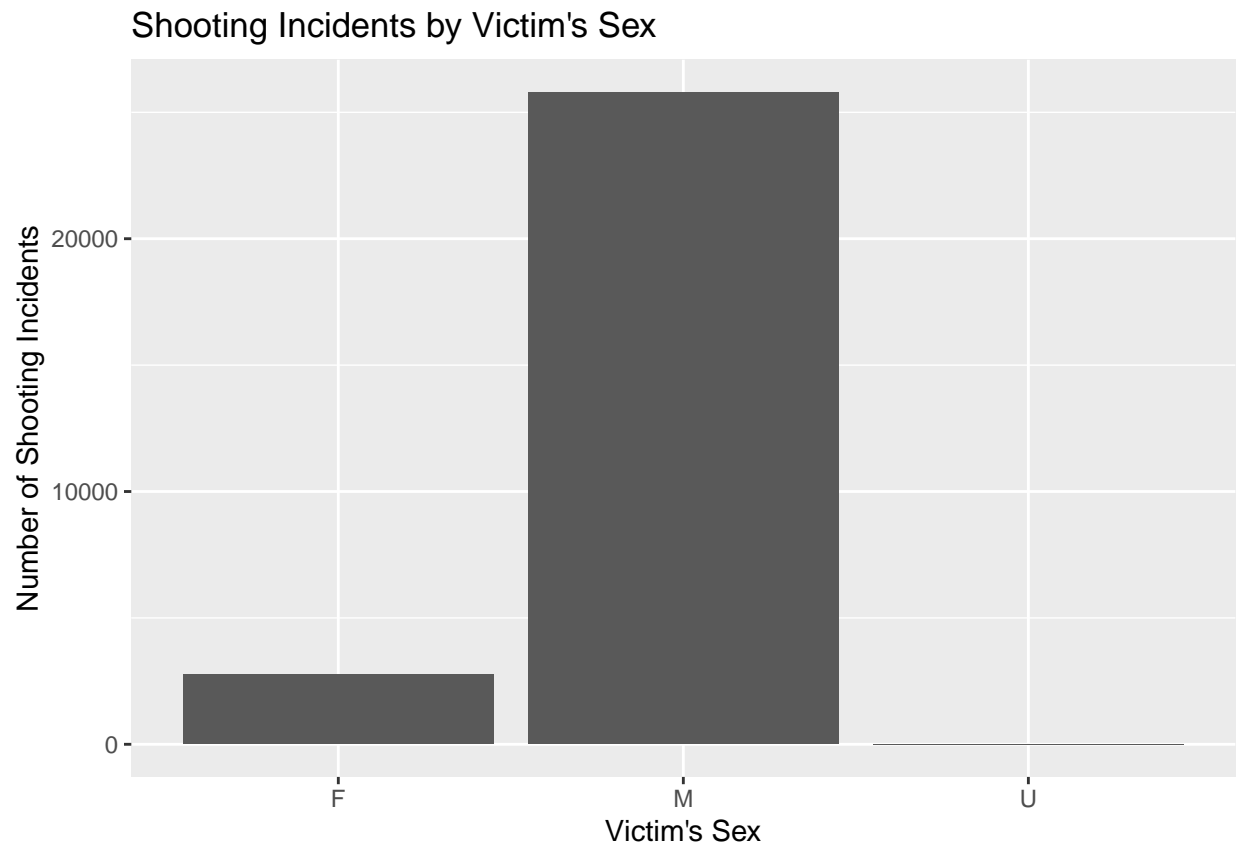
Shooting Incidents where victim is 25−44
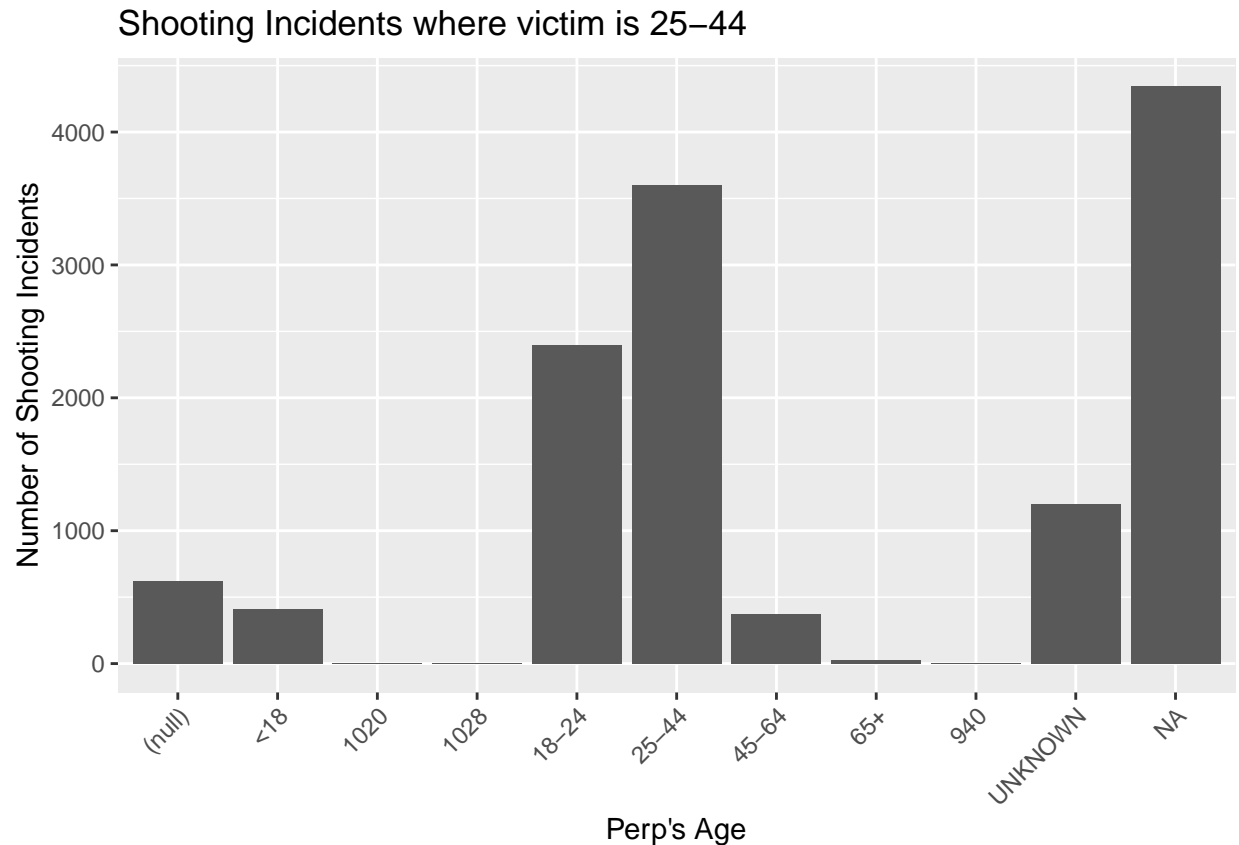
The number of incidents is highest within the age range 25-44, followed by 18-24. When filtering out for the perpetrators of of the highest victim age group, it seems they are targeted the most by people within their own age range.
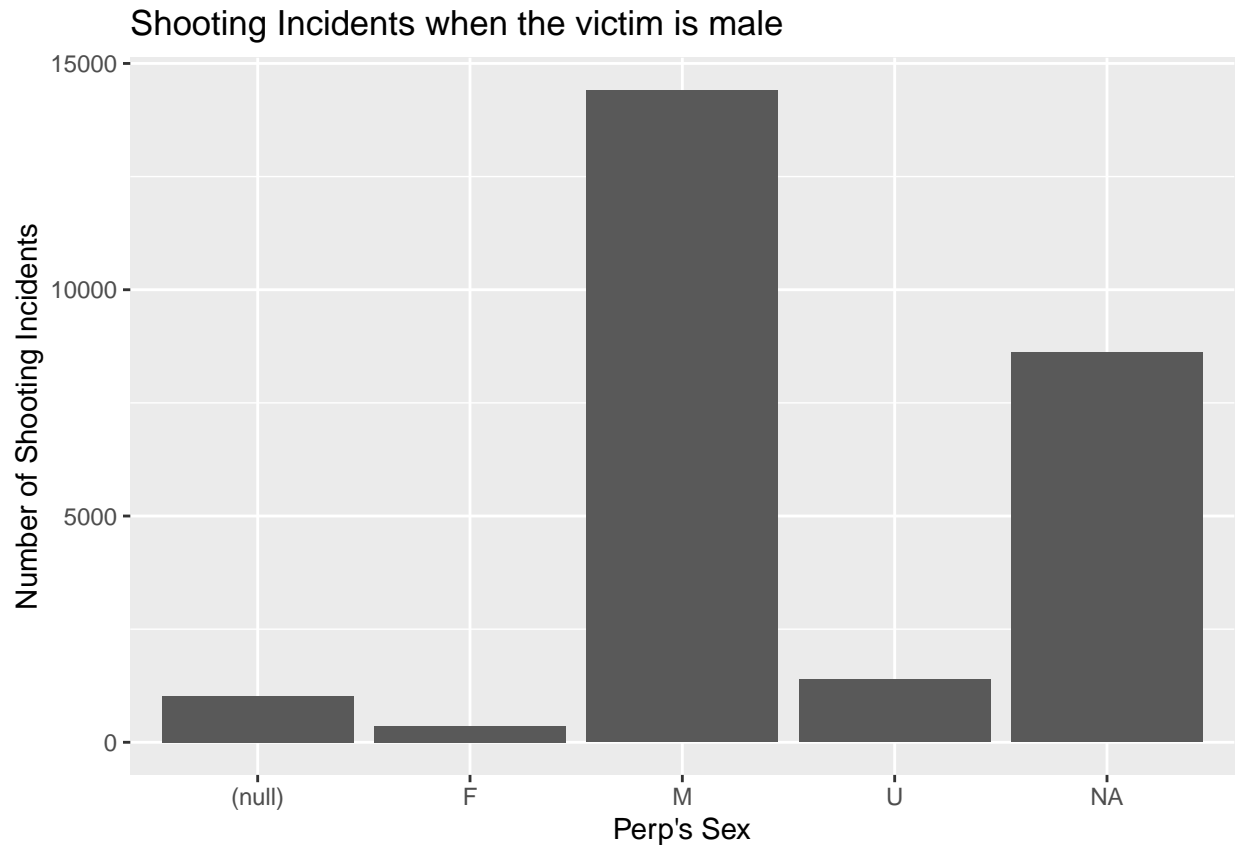
Victim's Sex

Plot the data based on the victim's sex.

```
NYPD_data %>%
  ggplot(aes(x = VIC_SEX)) +
  geom_bar() +
  labs(title = "Shooting Incidents by Victim's Sex",
       x = "Victim's Sex",
       y = "Number of Shooting Incidents")
```

Shooting Incidents by Victim's Sex


```
NYPD_data %>%
  filter(VIC_SEX == 'M') %>%
  ggplot(aes(x = PERP_SEX)) +
  geom_bar() +
  labs(title = "Shooting Incidents when the victim is male",
       x = "Perp's Sex",
       y = "Number of Shooting Incidents")
```
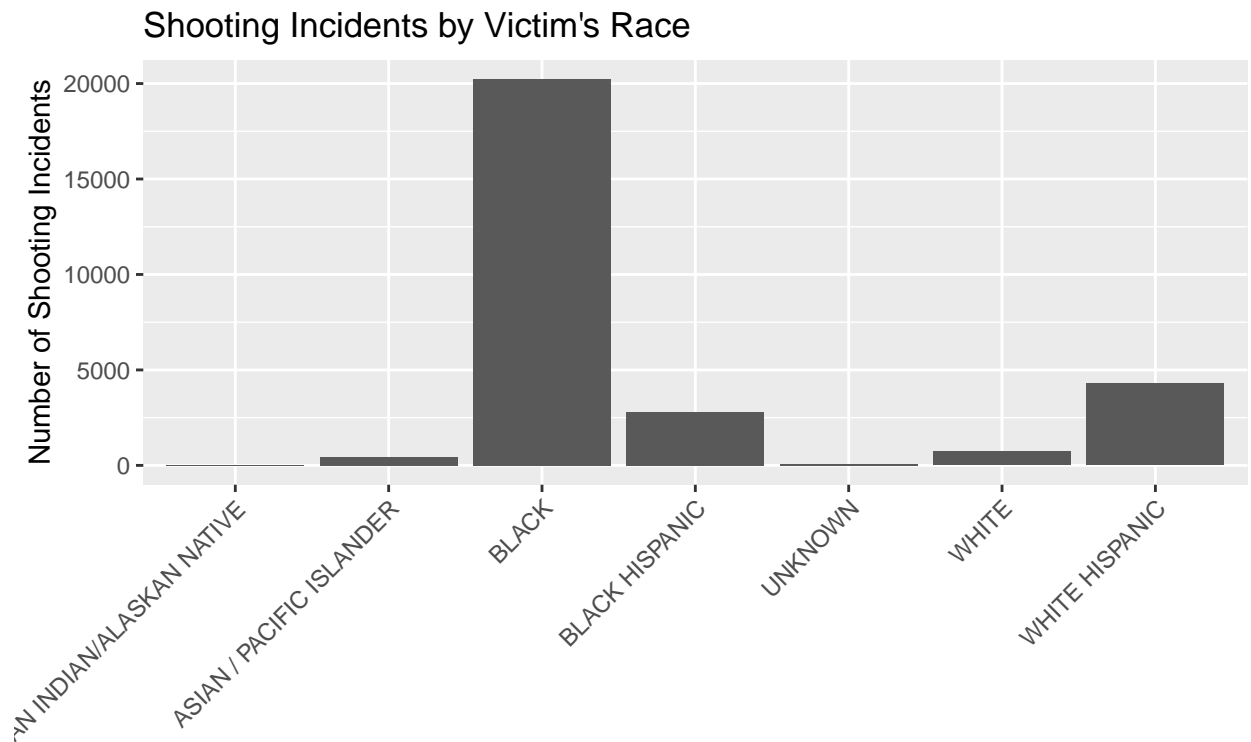
Shooting Incidents when the victim is male

The amount of male victims far outnumber the amount of female victims, by a ratio of almost 10:1. Like the previous section, it seems the victims are targeted the most by others of their own sex, i.e. male.
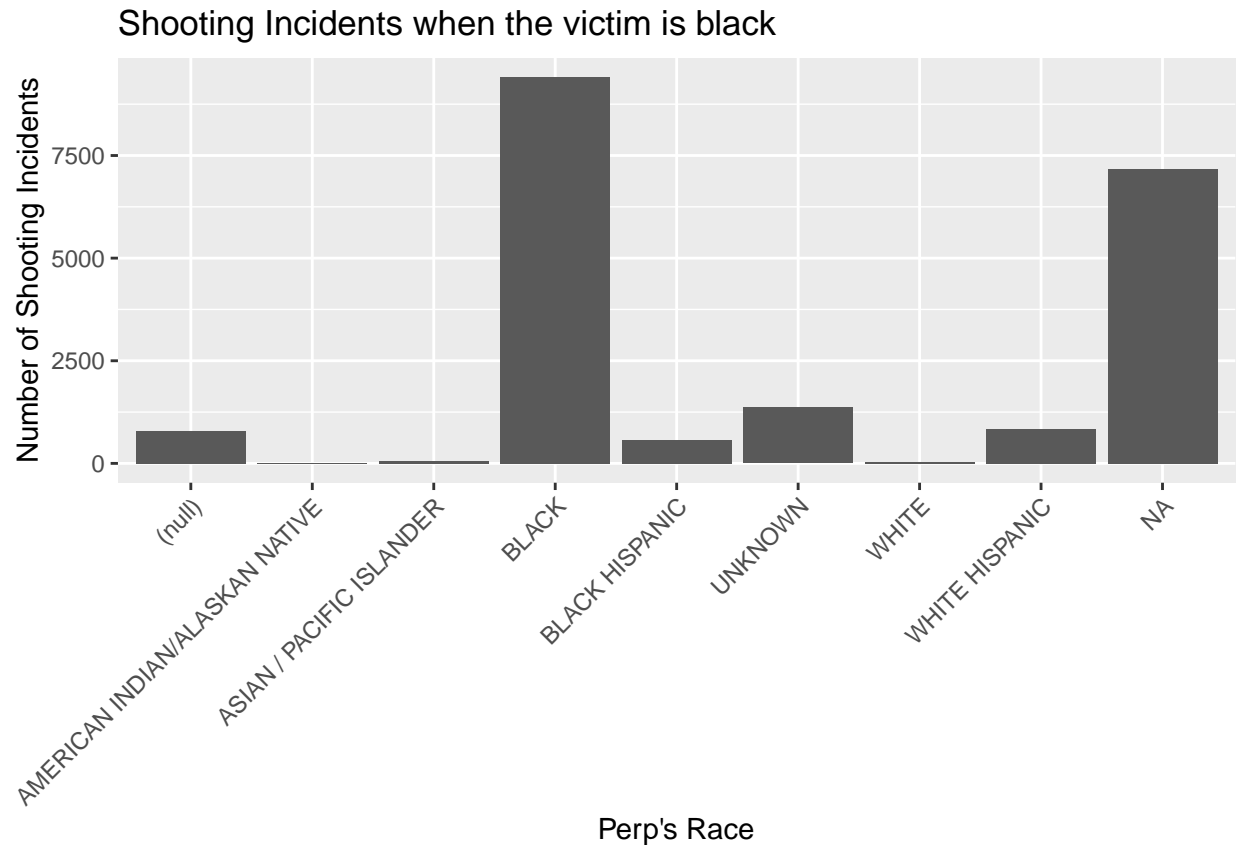
Victim's Race

Plot the date based on the victim's race.

```
NYPD_data %>%
  ggplot(aes(x = VIC_RACE)) +
  geom_bar() +
  labs(title = "Shooting Incidents by Victim's Race",
       x = "Victim's Race",
       y = "Number of Shooting Incidents") +
  guides(x = guide_axis(angle = 45))
```

Shooting Incidents by Victim's Race

Victim's Race

```
NYPD_data %>%
  filter(VIC_RACE == 'BLACK') %>%
  ggplot(aes(x = PERP_RACE)) +
  geom_bar() +
  labs(title = "Shooting Incidents when the victim is black",
      x = "Perp's Race",
      y = "Number of Shooting Incidents") +
  guides(x = guide_axis(angle = 45))
```

Shooting Incidents when the victim is black

The Black community has the highest number of victims, followed by White Hispanic and Black Hispanic. Similarly to the previous, the perpetrators are of the same race.

Linear Regression Model

```
NYPD_model <- lm(STATISTICAL_MURDER_FLAG ~ VIC_AGE_GROUP + VIC_SEX + VIC_RACE, data = NYPD_data)
summary(NYPD_model)
```

```
##
## Call:
## lm(formula = STATISTICAL_MURDER_FLAG ~ VIC_AGE_GROUP + VIC_SEX +
##     VIC_RACE, data = NYPD_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3701 -0.2160 -0.1650 -0.1294  0.9727
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -0.043196   0.118902  -0.363   0.7164
## VIC_AGE_GROUP1022        -0.129366   0.393223  -0.329   0.7422
## VIC_AGE_GROUP18-24        0.035626   0.008217   4.336 1.46e-05 ***
## VIC_AGE_GROUP25-44        0.086663   0.008039  10.780  < 2e-16 ***
## VIC_AGE_GROUP45-64        0.112239   0.011481   9.776  < 2e-16 ***
## VIC_AGE_GROUP65+          0.179189   0.028518   6.283 3.36e-10 ***
## VIC_AGE_GROUPUNKNOWN      0.127628   0.052455   2.433   0.0150 *
```

```
## VIC_SEXM                          -0.004963   0.007940  -0.625    0.5319
## VIC_SEXU                          -0.073066   0.119688  -0.610    0.5416
## VIC_RACEASIAN / PACIFIC ISLANDER   0.223102   0.120028   1.859    0.0631 .
## VIC_RACEBLACK                       0.177524   0.118576   1.497    0.1344
## VIC_RACEBLACK HISPANIC              0.152914   0.118776   1.287    0.1980
## VIC_RACEUNKNOWN                     0.075431   0.128800   0.586    0.5581
## VIC_RACEWHITE                       0.234062   0.119467   1.959    0.0501 .
## VIC_RACEWHITE HISPANIC              0.194689   0.118697   1.640    0.1010
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3932 on 28547 degrees of freedom
## Multiple R-squared:  0.009961,   Adjusted R-squared:  0.009476
## F-statistic: 20.52 on 14 and 28547 DF,  p-value: < 2.2e-16
```
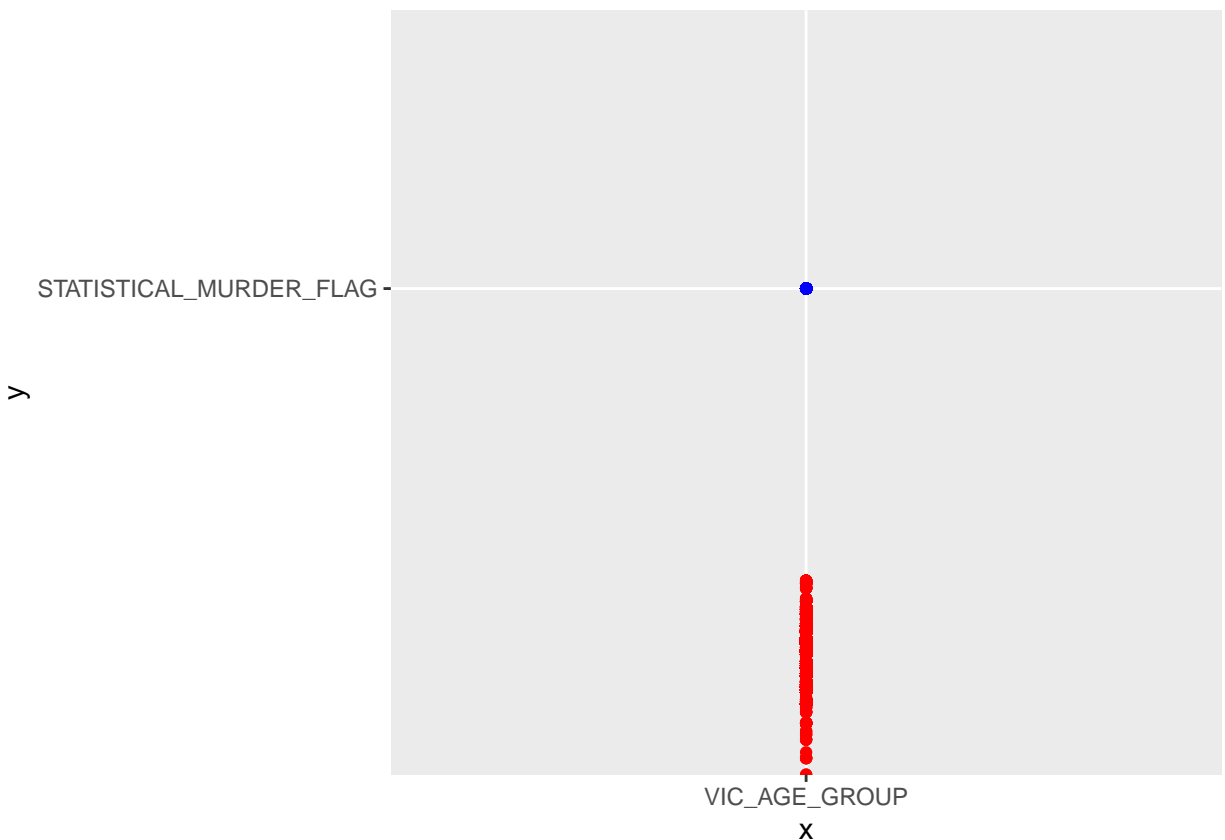
```r
NYPD_w_pred <- NYPD_data %>% mutate(pred = predict(NYPD_model))

NYPD_w_pred %>%
  ggplot() +
  geom_point(aes(x = 'VIC_AGE_GROUP', y = 'STATISTICAL_MURDER_FLAG'), color = "blue") +
  geom_point(aes(x = 'VIC_AGE_GROUP', y = pred), color = "red")
```

```
## Warning in geom_point(aes(x = "VIC_AGE_GROUP", y = "STATISTICAL_MURDER_FLAG"), : All aesthetics have
## i Please consider using `annotate()` or provide this layer with data containing
##   a single row.
```


```

## Sources of Bias

As for sources of bias, as someone who has never lived in a large city, the large number of shooting incidents was a surprise to me. I believe I was able to eliminate any personal biases due to focusing on the numbers on their own.

## Conclusion

The targets of shooting incidents disproportionately skew towards the black and Hispanic communities. In addition, mot victims are men and young adults. Additionally, this raises additional questions of whether the intersection of these various demographics will result in different conclusions (i.e. comparing the number of male to female communities within the age range of 25-44 and seeing if the ratio is different).