```
---
title: "NYPD Shooting Report"

date: '2024-06-19'
output:
  html_document: default
  pdf_document: default
---





## Step 1: Load Dataset

### Description

__NYPD Shooting Incident Data (Historic)__


__Source__: [NYPD Shooting Incident Data (Historic)](https://catalog.data.gov/dataset/nypd-shooting-inc


### Column Description

- __INCIDENT_KEY__: Randomly generated persistent ID for each incident.
- __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__: Indicates whether the shooting resulted in the victim's death (TRUE if fa
- __PERP_AGE_GROUP__: Perpetrator's age group.
- __PERP_SEX__: Perpetrator's sex.
- __PERP_RACE__: Perpetrator's race.
- __VIC_AGE_GROUP__: Victim's age group.
- __VIC_SEX__: Victim's sex.
- __VIC_RACE__: Victim's race.

### Import Dataset


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

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

shootings <- read_csv(url)

glimpse(shootings)  # View structure of the dataset
```

```
## Rows: 28,562
## Columns: 21
## $ INCIDENT_KEY            <dbl> 244608249, 247542571, 84967535, 202853370, 270~
```
```

```
## $ OCCUR_DATE              <chr> "05/05/2022", "07/04/2022", "05/27/2012", "09/~
## $ OCCUR_TIME              <time> 00:10:00, 22:20:00, 19:35:00, 21:00:00, 21:00~
## $ BORO                    <chr> "MANHATTAN", "BRONX", "QUEENS", "BRONX", "BROO~
## $ LOC_OF_OCCUR_DESC       <chr> "INSIDE", "OUTSIDE", NA, NA, NA, NA, NA, NA, N~
## $ PRECINCT                <dbl> 14, 48, 103, 42, 83, 23, 113, 77, 48, 49, 73, ~
## $ JURISDICTION_CODE       <dbl> 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ LOC_CLASSFCTN_DESC      <chr> "COMMERCIAL", "STREET", NA, NA, NA, NA, NA, NA~
## $ LOCATION_DESC           <chr> "VIDEO STORE", "(null)", NA, NA, NA, "MULTI DW~
## $ STATISTICAL_MURDER_FLAG <lgl> TRUE, TRUE, FALSE, FALSE, FALSE, FALSE, TRUE, ~
## $ PERP_AGE_GROUP          <chr> "25-44", "(null)", NA, "25-44", "25-44", NA, N~
## $ PERP_SEX                <chr> "M", "(null)", NA, "M", "M", NA, NA, NA, NA, "~
## $ PERP_RACE               <chr> "BLACK", "(null)", NA, "UNKNOWN", "BLACK", NA,~
## $ VIC_AGE_GROUP           <chr> "25-44", "18-24", "18-24", "25-44", "25-44", "~
## $ VIC_SEX                 <chr> "M", "M", "M", "M", "M", "M", "M", "M", "M", "~
## $ VIC_RACE                <chr> "BLACK", "BLACK", "BLACK", "BLACK", "BLACK", "~
## $ X_COORD_CD              <dbl> 986050, 1016802, 1048632, 1014493, 1009149, 99~
## $ Y_COORD_CD              <dbl> 214231.0, 250581.0, 198262.0, 242565.0, 190104~
## $ Latitude                <dbl> 40.75469, 40.85440, 40.71063, 40.83242, 40.688~
## $ Longitude               <dbl> -73.99350, -73.88233, -73.76777, -73.89071, -7~
## $ Lon_Lat                 <chr> "POINT (-73.9935 40.754692)", "POINT (-73.8823~
```

Step 2: Tidy and Transform Data

```r
shootings <- shootings %>%
  select(-c(PRECINCT, JURISDICTION_CODE, LOCATION_DESC, X_COORD_CD, Y_COORD_CD, Lon_Lat))
```

Convert Data Types

Convert OCCUR_DATE to date object.

```r
shootings <- shootings %>%
  mutate(OCCUR_DATE = mdy(OCCUR_DATE))
```

Convert Variables to Factors

```r
shootings <- shootings %>%
  mutate(
    BORO = factor(BORO),
    PERP_AGE_GROUP = factor(PERP_AGE_GROUP),
    PERP_SEX = factor(PERP_SEX),
    PERP_RACE = factor(PERP_RACE),
    VIC_AGE_GROUP = factor(VIC_AGE_GROUP),
    VIC_SEX = factor(VIC_SEX),
    VIC_RACE = factor(VIC_RACE),
    STATISTICAL_MURDER_FLAG = factor(STATISTICAL_MURDER_FLAG)
  )
```

Step 3: Add Visualizations and Analysis

```r
table(shootings$STATISTICAL_MURDER_FLAG)
```

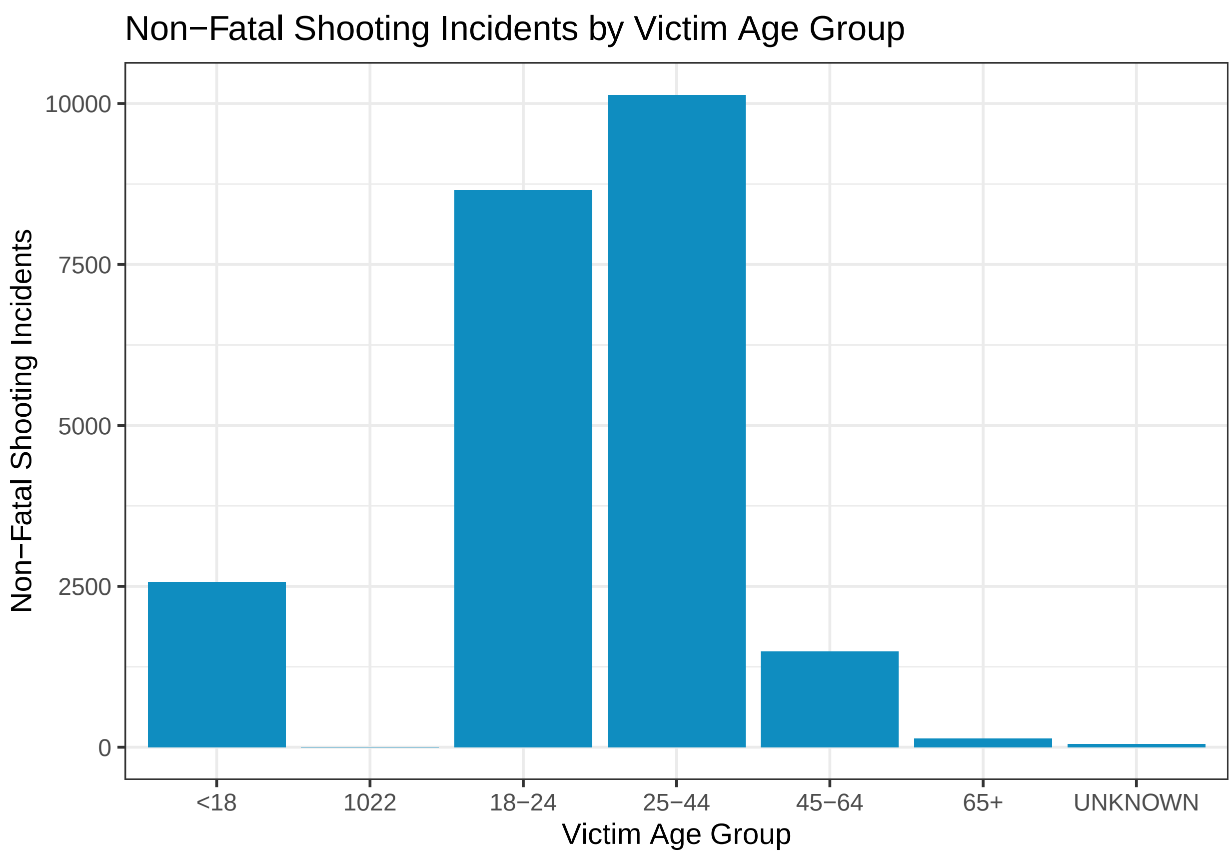
```
##
## FALSE  TRUE
## 23036  5526
```

Victim Age

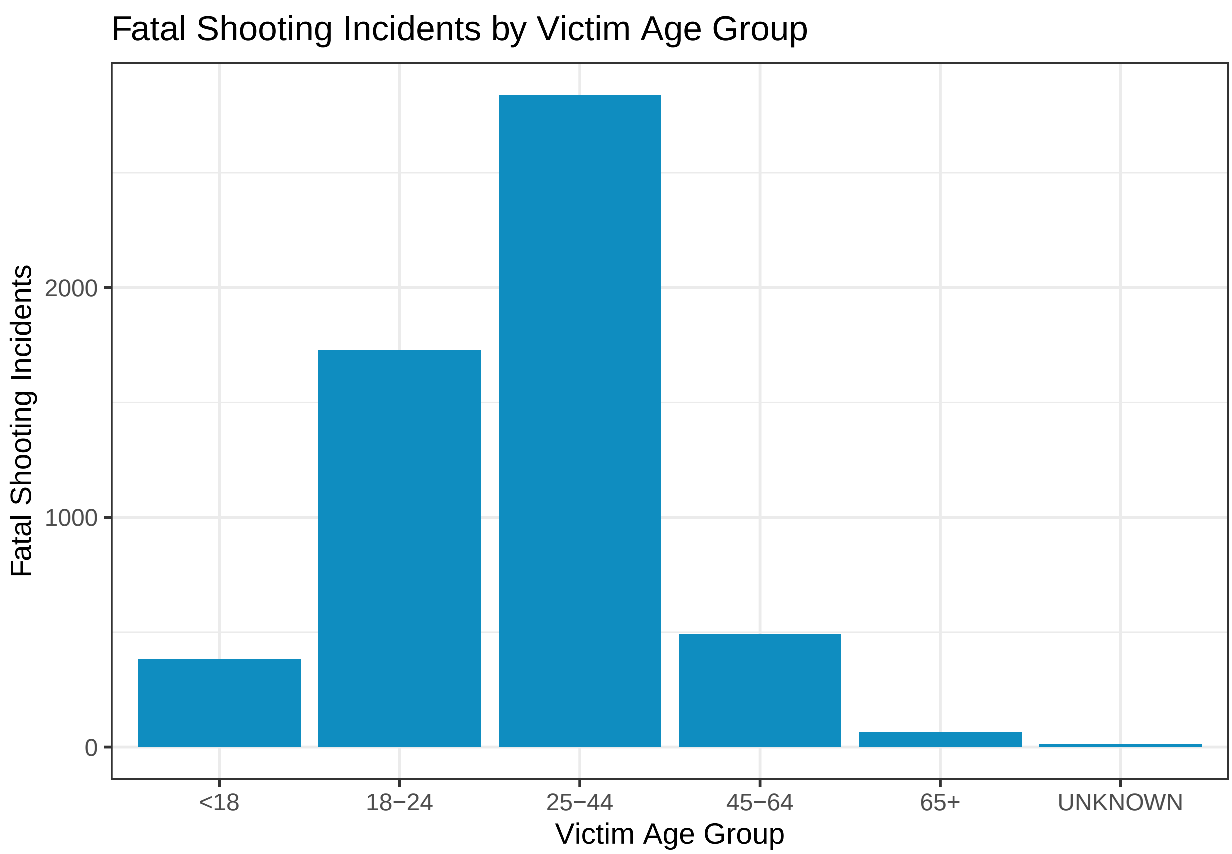
```r
table(shootings$STATISTICAL_MURDER_FLAG, shootings$VIC_AGE_GROUP)
```

```
##
##           <18  1022 18-24 25-44 45-64   65+ UNKNOWN
##   FALSE  2569     1  8654 10135  1489   139      49
##   TRUE    385     0  1730  2838   492    66      15
```

```r
shootings %>%
  filter(STATISTICAL_MURDER_FLAG == FALSE) %>%
  ggplot(aes(x = VIC_AGE_GROUP)) +
  geom_bar(fill = "#0F8DC0") +
  theme_bw() +
  labs(x = "Victim Age Group",
       y = "Non-Fatal Shooting Incidents",
       title = "Non-Fatal Shooting Incidents by Victim Age Group")
```




```r
shootings %>%
  filter(STATISTICAL_MURDER_FLAG == TRUE) %>%
  ggplot(aes(x = VIC_AGE_GROUP)) +
  geom_bar(fill = "#0F8DC0") +
  theme_bw() +
  labs(x = "Victim Age Group",
       y = "Fatal Shooting Incidents",
       title = "Fatal Shooting Incidents by Victim Age Group")
```

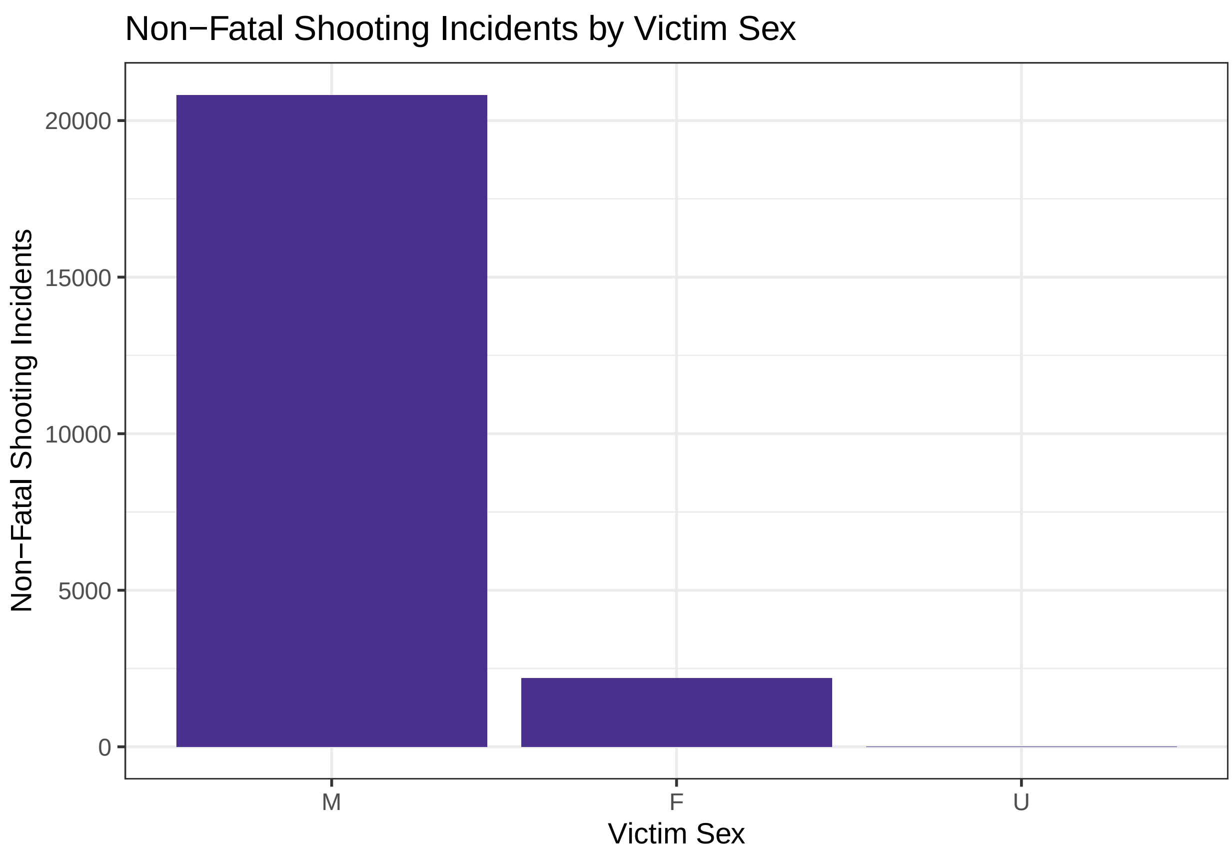


Victim Sex

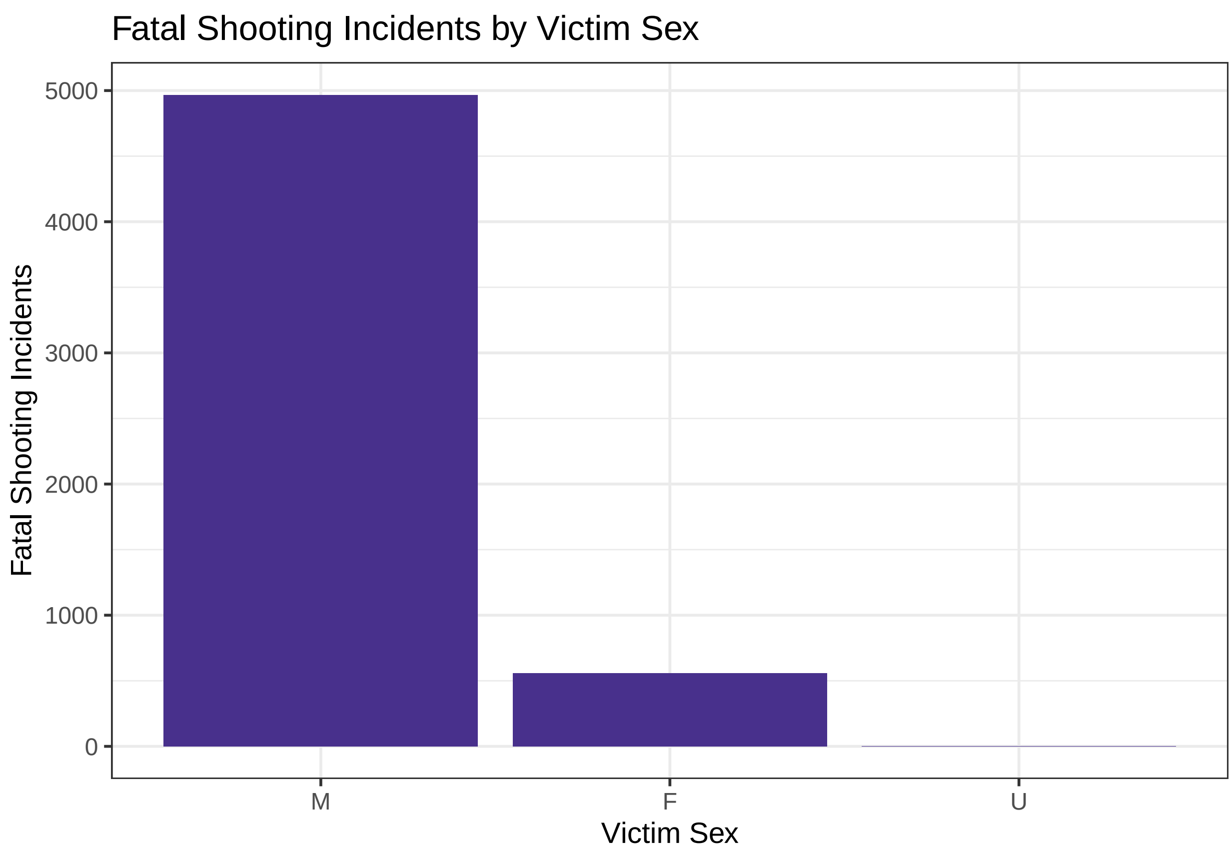
```
table(shootings$STATISTICAL_MURDER_FLAG, shootings$VIC_SEX)
```

```
##
##              F     M     U
##   FALSE   2203 20822    11
##   TRUE     557  4968     1
```

```
shootings %>%
  filter(STATISTICAL_MURDER_FLAG == FALSE) %>%
  ggplot(aes(x = fct_infreq(VIC_SEX))) +
  geom_bar(stat = 'count') +
  geom_bar(fill = "#48308C") +
  theme_bw() +
  labs(x = "Victim Sex",
       y = "Non-Fatal Shooting Incidents",
       title = "Non-Fatal Shooting Incidents by Victim Sex")
```




```
shootings %>%
  filter(STATISTICAL_MURDER_FLAG == TRUE) %>%
  ggplot(aes(x = fct_infreq(VIC_SEX))) +
  geom_bar(stat = 'count') +
  geom_bar(fill = "#48308C") +
  theme_bw() +
  labs(x = "Victim Sex",
       y = "Fatal Shooting Incidents",
       title = "Fatal Shooting Incidents by Victim Sex")
```

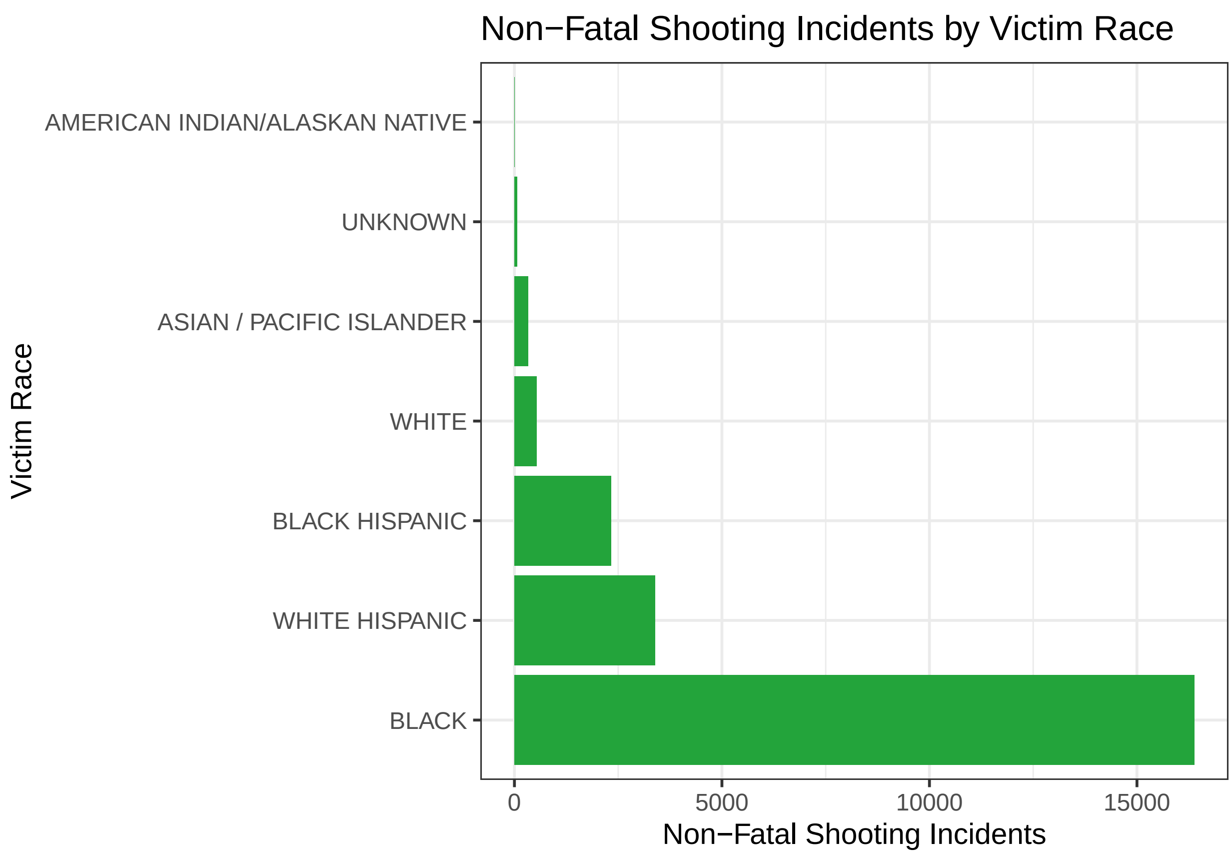


Victim Race

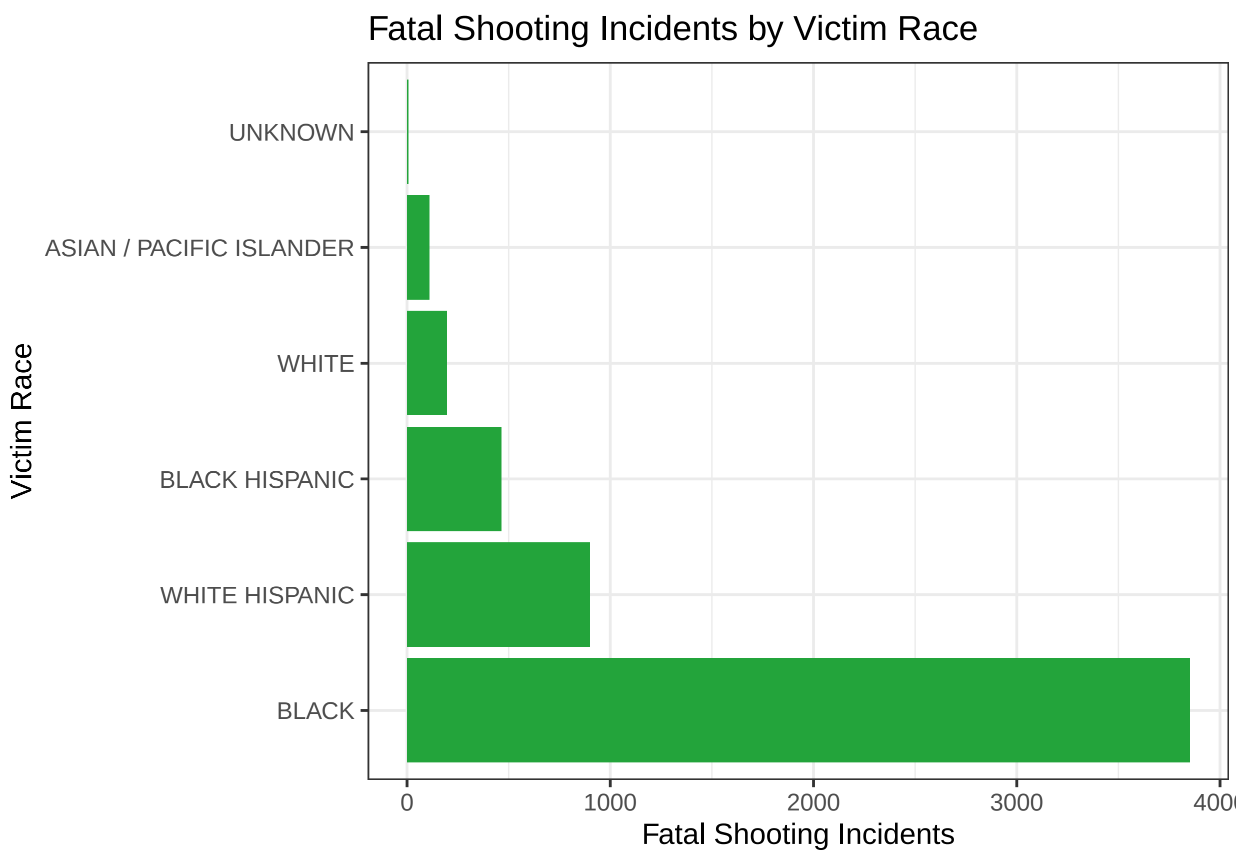
```
table(shootings$STATISTICAL_MURDER_FLAG, shootings$VIC_RACE)
```

```
##
##         AMERICAN INDIAN/ALASKAN NATIVE ASIAN / PACIFIC ISLANDER BLACK
##   FALSE                             11                      330 16384
##   TRUE                               0                      110  3851
##
##         BLACK HISPANIC UNKNOWN WHITE WHITE HISPANIC
##   FALSE           2332      63   532           3384
##   TRUE             463       7   196            899
```

```
shootings %>%
  filter(STATISTICAL_MURDER_FLAG == FALSE) %>%
  ggplot(aes(x = fct_infreq(VIC_RACE))) +
  geom_bar(stat = 'count') +
  geom_bar(fill = "#23A43B") +
  coord_flip() +
  theme_bw() +
  labs(x = "Victim Race",
       y = "Non-Fatal Shooting Incidents",
       title = "Non-Fatal Shooting Incidents by Victim Race")
```




```r
shootings %>%
  filter(STATISTICAL_MURDER_FLAG == TRUE) %>%
  ggplot(aes(x = fct_infreq(VIC_RACE))) +
  geom_bar(stat = 'count') +
  geom_bar(fill = "#23A43B") +
  coord_flip() +
  theme_bw() +
  labs(x = "Victim Race",
       y = "Fatal Shooting Incidents",
       title = "Fatal Shooting Incidents by Victim Race")
```



Multivariable Logistic Regression Model

```
glm_model <- glm(STATISTICAL_MURDER_FLAG ~ VIC_AGE_GROUP + VIC_SEX + VIC_RACE, data = shootings, family

summary(glm_model)
```

```
##
## Call:
## glm(formula = STATISTICAL_MURDER_FLAG ~ VIC_AGE_GROUP + VIC_SEX +
##     VIC_RACE, family = "binomial", data = shootings)
##
## Coefficients:
##                                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)                       -12.90411   97.39229  -0.132   0.8946
## VIC_AGE_GROUP1022                 -10.66158  324.74370  -0.033   0.9738
## VIC_AGE_GROUP18-24                  0.28445    0.06081   4.678 2.90e-06 ***
## VIC_AGE_GROUP25-44                  0.61495    0.05880  10.458  < 2e-16 ***
## VIC_AGE_GROUP45-64                  0.75645    0.07591   9.965  < 2e-16 ***
## VIC_AGE_GROUP65+                    1.08054    0.16030   6.741 1.58e-11 ***
## VIC_AGE_GROUPUNKNOWN                0.86260    0.31637   2.727   0.0064 **
## VIC_SEXM                           -0.03247    0.05080  -0.639   0.5227
## VIC_SEXU                           -0.61990    1.08054  -0.574   0.5662
## VIC_RACEASIAN / PACIFIC ISLANDER   11.29529   97.39233   0.116   0.9077
## VIC_RACEBLACK                      11.03210   97.39227   0.113   0.9098
```

```
## VIC_RACEBLACK HISPANIC           10.86233   97.39228   0.112   0.9112
## VIC_RACEUNKNOWN                  10.21286   97.39317   0.105   0.9165
## VIC_RACEWHITE                    11.34597   97.39230   0.116   0.9073
## VIC_RACEWHITE HISPANIC           11.13984   97.39227   0.114   0.9089
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 28061  on 28561  degrees of freedom
## Residual deviance: 27772  on 28547  degrees of freedom
## AIC: 27802
##
## Number of Fisher Scoring iterations: 11
```

Conclusion

Through the analysis of NYPD shooting incident data, this report aimed to determine whether demographic factors such as age, sex, or race of the victim could predict the likelihood of a shooting being fatal. The findings indicate that victim age group emerges as a significant predictor of the outcome of shooting incidents. Specifically, younger victims, particularly those in the age groups under 18 and 18-24, exhibit higher survival rates compared to older age groups. Conversely, shootings involving victims aged 65 and above are more likely to result in fatalities.

The analysis also explored the roles of victim sex and race in shooting outcomes. While there are disparities in the distribution of incidents across these demographics, particularly with a higher incidence among males and certain racial groups, these factors alone did not show a clear predictive relationship with fatality outcomes once controlled for other variables.

The multivariable logistic regression model reinforced the significance of victim age group in predicting fatality, suggesting that age-related physiological factors and potentially different circumstances surrounding incidents involving different age groups may influence survival rates.

Recommendations for Further Investigation

To further enhance predictive models and deepen understanding of shooting incidents, future investigations could consider integrating additional variables such as location-specific factors (e.g., neighborhood demographics, economic conditions, policing practices) and situational variables (e.g., time of day, presence of firearms). Such factors could provide a more nuanced view of the complex dynamics influencing shooting outcomes in urban environments.

Sources of Bias

While efforts were made to maintain objectivity throughout the analysis, it's important to acknowledge potential biases inherent in the data and analysis process. These biases could stem from limitations in data collection methodologies, inherent societal biases reflected in crime reporting and policing practices, as well as the researcher's own perspectives and interpretations.

Implications for Policy and Community Action

The insights from this study underscore the importance of targeted interventions aimed at reducing gun violence and improving emergency response strategies, particularly for vulnerable age groups identified in the analysis. Policies focusing on youth engagement, community policing, and firearm regulations tailored to high-risk areas and demographic groups could potentially mitigate the impact of gun violence and improve overall public safety.

In conclusion, while victim demographics, particularly age, provide valuable insights into the outcomes of shooting incidents, a comprehensive understanding requires consideration of a broader range of factors. Continued research and data-driven approaches are essential for developing effective strategies to prevent gun violence and ensure safer communities for all residents. “‘