NYPD Shooting Incident Data Analysis

Cristian

2025-03-04

Introduction

This report analyzes the NYPD Shooting Incident Data Historic dataset. It includes data on shooting incidents across various boroughs over time. The analysis involves data cleaning, visualizations, and a logistic regression model to predict shooting incidents.

Load Libraries

```r
library(tidyverse)
library(lubridate)
library(ggplot2)
library(broom)
```
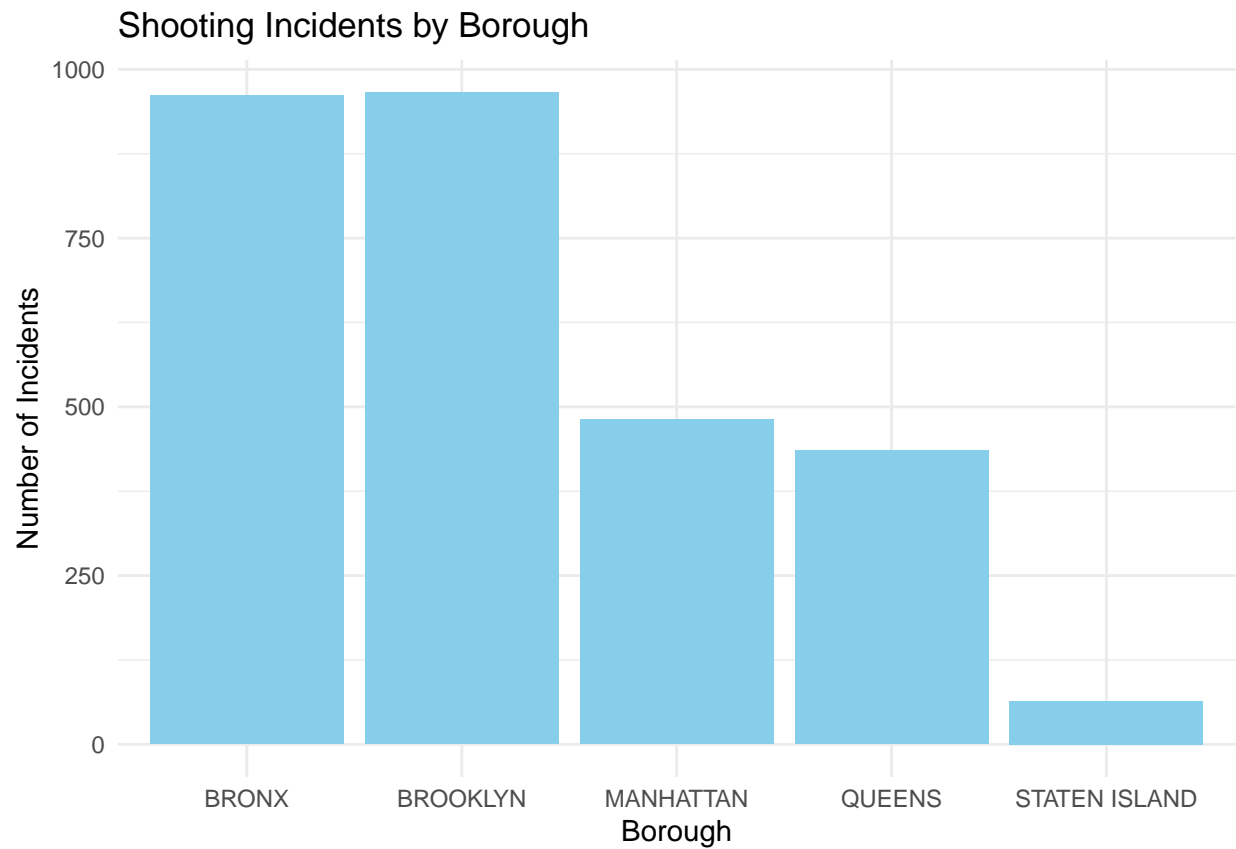
Load and Clean Data

```r
nypd_data <- read_csv("NYPD_Shooting_Incident_Data__Historic_ (1).csv")

nypd_data_clean <- nypd_data %>%
  drop_na() %>%
  distinct() %>%
  mutate(
    OCCUR_DATE = mdy(OCCUR_DATE),
    BORO = factor(BORO),
    STATISTICAL_MURDER_FLAG = as.factor(STATISTICAL_MURDER_FLAG)
  )
```

## Visualization 1: Shooting Incidents by Borough

```r
ggplot(nypd_data_clean, aes(x = BORO)) +
  geom_bar(fill = "skyblue") +
  labs(title = "Shooting Incidents by Borough", x = "Borough", y = "Number of Incidents") +
  theme_minimal()
```

Shooting Incidents by Borough

Number of Incidents vs Borough — bar chart showing:
- BRONX: ~960
- BROOKLYN: ~965
- MANHATTAN: ~480
- QUEENS: ~435
- STATEN ISLAND: ~65

Visualization 2: Shooting Incidents Over Time

```r
ggplot(nypd_data_clean, aes(x = OCCUR_DATE)) +
  geom_histogram(binwidth = 30, fill = "lightgreen") +
  labs(title = "Shooting Incidents Over Time", x = "Date", y = "Number of Incidents") +
  theme_minimal() +
  scale_x_date(date_breaks = "1 year", date_labels = "%Y")
```

Shooting Incidents Over Time
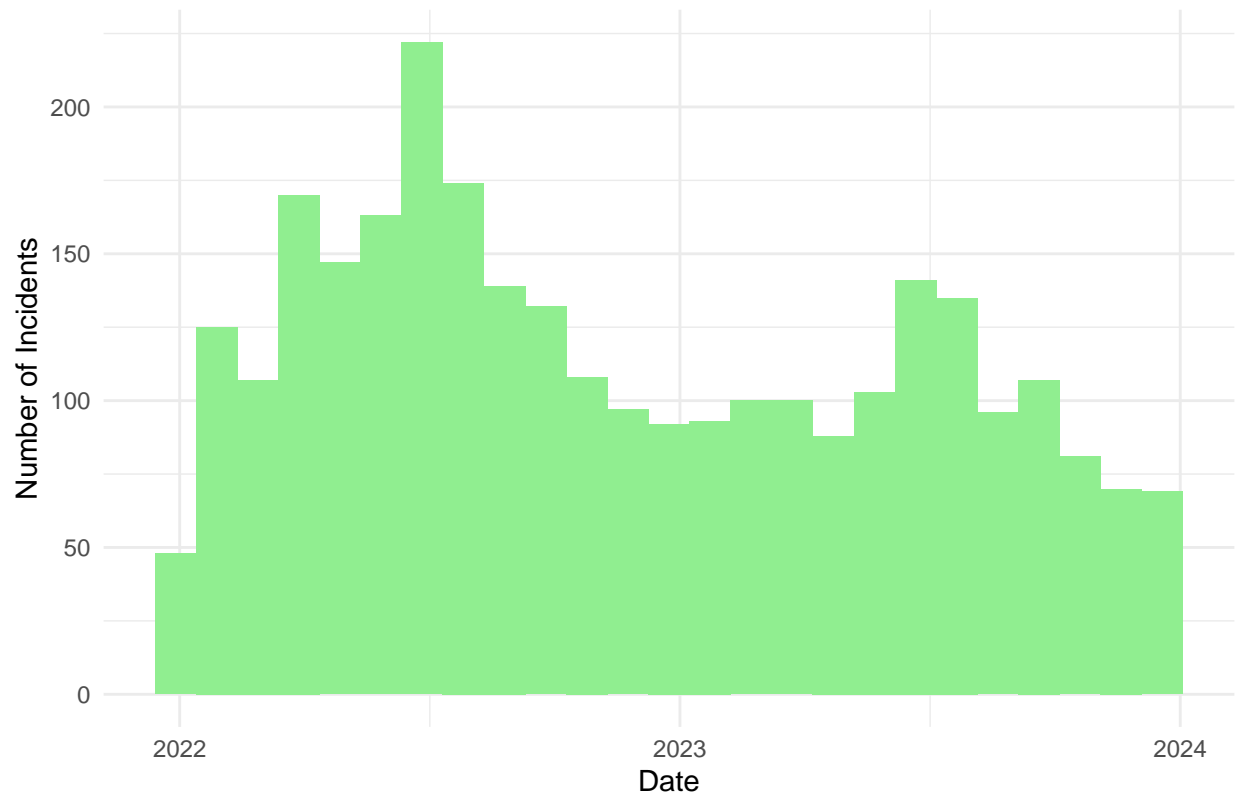
Predictive Model: Logistic Regression

```r
model <- glm(STATISTICAL_MURDER_FLAG ~ BORO + OCCUR_DATE, data = nypd_data_clean, family = binomial())
summary(model)
```

```
##
## Call:
## glm(formula = STATISTICAL_MURDER_FLAG ~ BORO + OCCUR_DATE, family = binomial(),
##     data = nypd_data_clean)
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -1.334e-01  4.354e+00  -0.031   0.9756
## BOROBROOKLYN     -1.947e-01  1.122e-01  -1.736   0.0826 .
## BOROMANHATTAN    -1.325e-01  1.371e-01  -0.966   0.3338
## BOROQUEENS       -2.894e-01  1.469e-01  -1.970   0.0488 *
## BOROSTATEN ISLAND -4.176e-02 3.121e-01  -0.134   0.8935
## OCCUR_DATE       -5.685e-05  2.253e-04  -0.252   0.8008
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##     Null deviance: 2943.9  on 2906  degrees of freedom
## Residual deviance: 2938.7  on 2901  degrees of freedom
## AIC: 2950.7
##
## Number of Fisher Scoring iterations: 4
```

Bias Considerations

- **Temporal Bias**: Shooting incidents may be more prevalent in certain years or months.
- **Demographic Bias**: Certain racial groups may be overrepresented due to systemic factors.
- **Reporting Bias**: Data may not capture unreported incidents.

Conclusion

This report explored trends in NYPD shooting incidents using visualizations and logistic regression modeling. Further research could incorporate additional predictive factors such as location and time of day.