NYPD Shooting Incident Data Analysis

2024-06-01

NYPD Shooting Incident Data Analysis

Description of the Dataset

This dataset details every shooting incident in NYC from 2006 to the end of the previous year, including event specifics, location, time, suspect and victim demographics, and is reviewed quarterly by the Office of Management Analysis and Planning before being posted on the NYPD website.

Load and Clean the Data

Set data source

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

Load the data

```
shootings_data <- read_csv(url_in, show_col_types = FALSE)
```

Clean the data

```
shootings_data <- shootings_data %>% select(c("OCCUR_DATE", "OCCUR_TIME", "BORO",
                                              "PRECINCT",
                                              "STATISTICAL_MURDER_FLAG",
                                              "VIC_AGE_GROUP", "VIC_SEX",
                                              "VIC_RACE", "PERP_AGE_GROUP",
                                              "PERP_SEX", "PERP_RACE",
                                              "Latitude", "Longitude")) %>%
  mutate(OCCUR_DATE = mdy(OCCUR_DATE),
         STATISTICAL_MURDER_FLAG = as.logical(STATISTICAL_MURDER_FLAG),
         BORO = as_factor(BORO), VIC_AGE_GROUP = as_factor(VIC_AGE_GROUP),
         VIC_SEX = as_factor(VIC_SEX), VIC_RACE = as_factor(VIC_RACE),
         PERP_AGE_GROUP = as_factor(PERP_AGE_GROUP),
         PERP_SEX = as_factor(PERP_SEX),
         PERP_RACE = as_factor(PERP_RACE)) %>%
  mutate(YEAR = year(OCCUR_DATE), MONTH = month(OCCUR_DATE), SHOOTINGS_COUNT = 1)
```

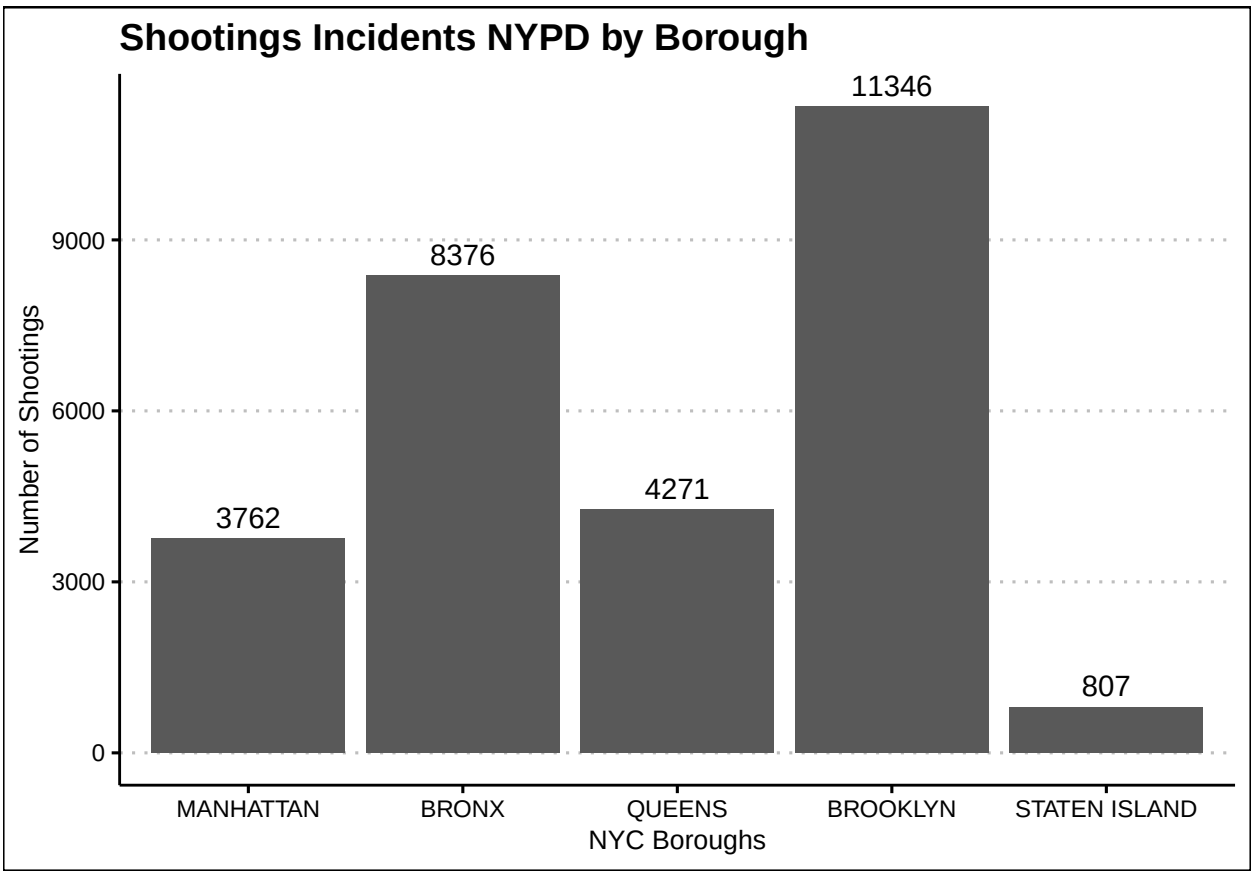
Data Visualization

Shootings Incidents NYPD by Borough

Total Count

```
shootings_by_borough <- shootings_data %>%
  ggplot(aes(x = BORO)) +
  geom_bar() +
  geom_text(stat='count', aes(label=after_stat(count)), vjust=-0.5) +
  labs(title = "Shootings Incidents NYPD by Borough",
       x = "NYC Boroughs",
       y = "Number of Shootings") +
  theme_clean() +
  scale_fill_economist()

shootings_by_borough
```



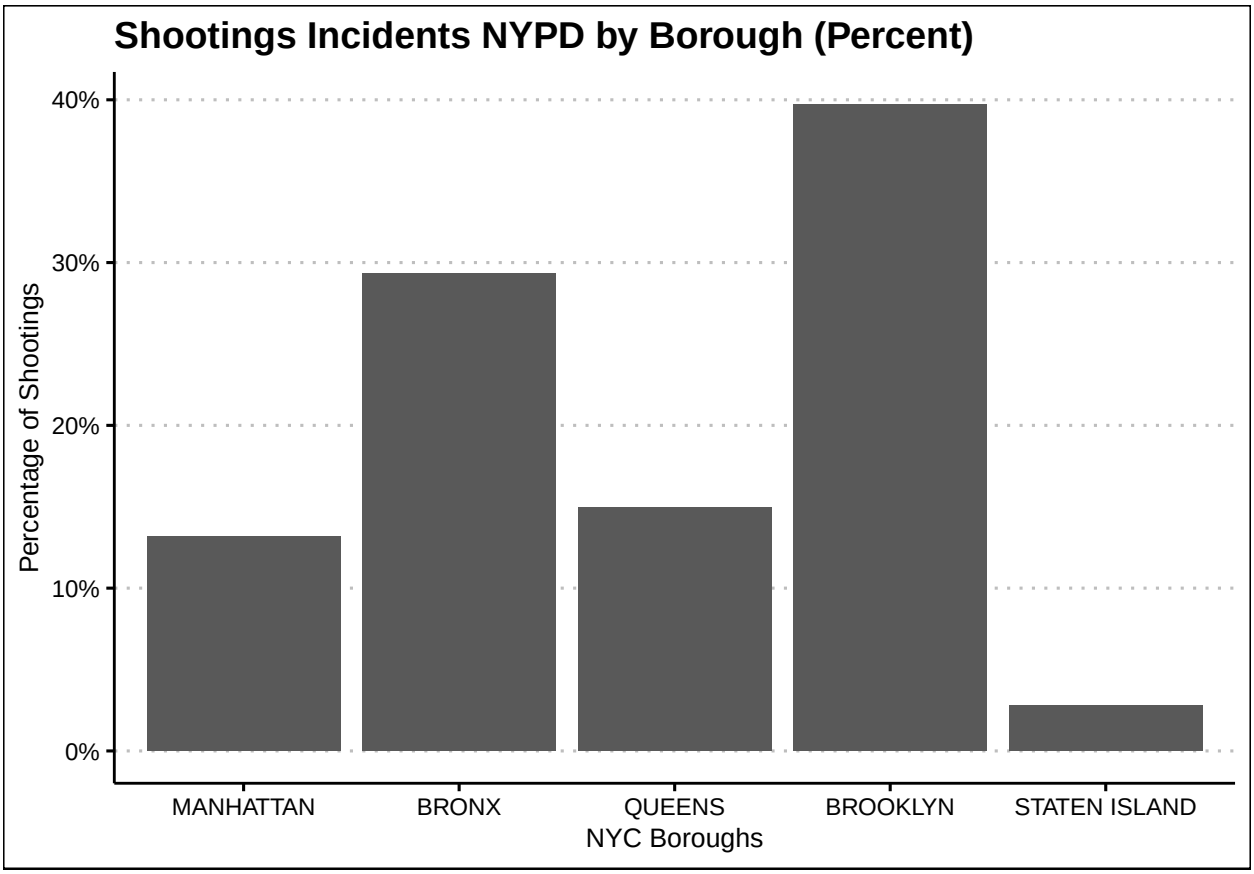
It can be observed that there are significant differences in the number of shooting incidents among the boroughs.

In Percent

```r
shootings_by_borough_perc <- shootings_data %>%
  count(BORO) %>%
  mutate(percentage = n / sum(n))

shootings_by_borough_perc <- shootings_by_borough_perc %>%
  ggplot(aes(x = BORO, y = percentage)) +
  geom_bar(stat = "identity") +
  scale_y_continuous(labels = scales::percent_format()) +
  labs(title = "Shootings Incidents NYPD by Borough (Percent)",
       x = "NYC Boroughs",
       y = "Percentage of Shootings") +
  theme_clean() +
  scale_fill_economist()

shootings_by_borough_perc
```



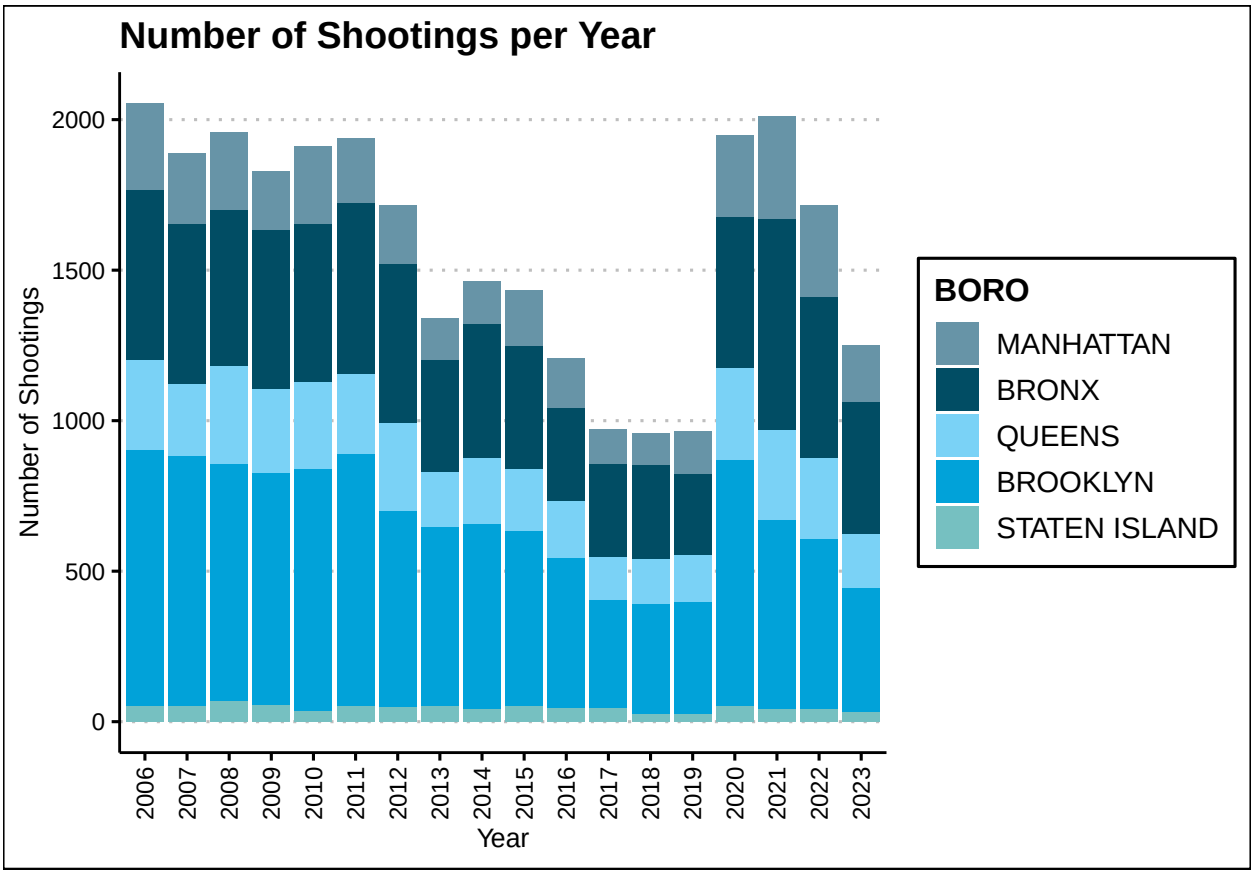
Brooklyn alone accounts for nearly 40% of all shooting incidents.

Shooting Incidents by Year and Borough

```r
yearly_data <- shootings_data %>%
  group_by(YEAR, BORO) %>%
  summarise(SHOOTINGS_COUNT = n()) %>%
  ungroup()
```

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

```
yearly_data %>%  ggplot(aes(x = factor(YEAR), y = SHOOTINGS_COUNT, fill = BORO)) +
  geom_bar(stat = "identity") +
  labs(title = "Number of Shootings per Year",
       x = "Year",
       y = "Number of Shootings") +
  theme_clean() +
  scale_fill_economist() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```



The number of shootings has substantially dropped during 2017-2019. The distribution among the boroughs remained stable over the years.

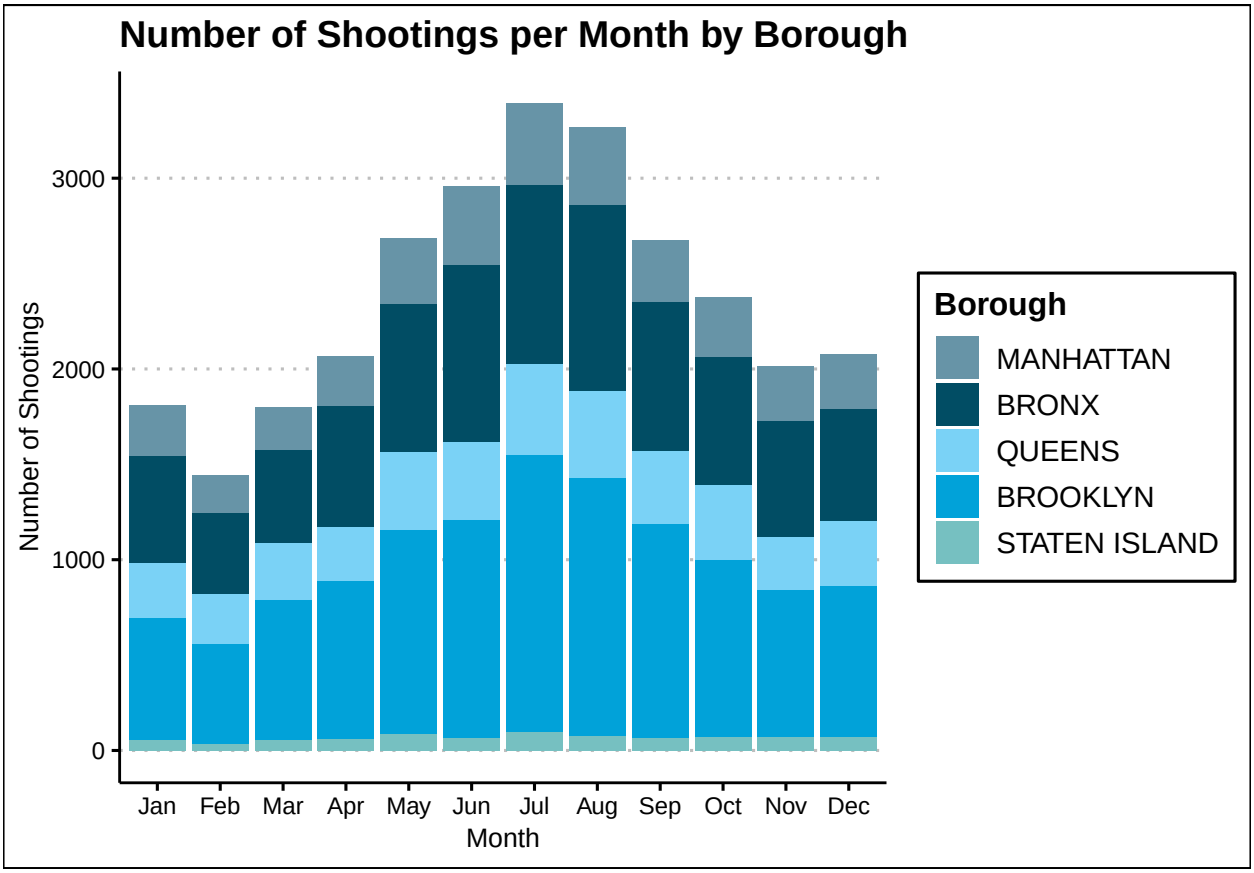
Shooting Incidents by Month and Borough

```
shootings_by_month <- shootings_data %>%
  mutate(MONTH = month.abb[MONTH])

monthly_boro_data <- shootings_by_month %>%
  group_by(MONTH, BORO) %>%
  summarise(SHOOTINGS_COUNT = n()) %>%
  ungroup()
```

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

```
ggplot(monthly_boro_data, aes(x = factor(MONTH, levels = month.abb), y = SHOOTINGS_COUNT, fill = BORO)) +
  geom_bar(stat = "identity") +
  labs(title = "Number of Shootings per Month by Borough",
       x = "Month",
       y = "Number of Shootings",
       fill = "Borough") +
  theme_clean() +
  scale_fill_economist()
```



The number of shooting incidents seems to exhibit seasonality: There are significantly more shootings in the summer months compared to the winter months.

Statistical Modeling and Results

Model: Investigating the Effects of the Borough and the Month

```
model <- lm(SHOOTINGS_COUNT ~ BORO + MONTH, data = monthly_boro_data)
summary(model)
```

```
##
```

```
## Call:
## lm(formula = SHOOTINGS_COUNT ~ BORO + MONTH, data = monthly_boro_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -234.267  -60.917   -4.225   58.821  304.533
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)         251.07      57.10   4.397 6.86e-05 ***
## BOROBRONX           384.50      45.15   8.517 7.39e-11 ***
## BOROQUEENS           42.42      45.15   0.940 0.352571
## BOROBROOKLYN        632.00      45.15  13.999  < 2e-16 ***
## BOROSTATEN ISLAND  -246.25      45.15  -5.455 2.12e-06 ***
## MONTHAug            239.20      69.94   3.420 0.001361 **
## MONTHDec              2.20      69.94   0.031 0.975048
## MONTHFeb           -124.80      69.94  -1.784 0.081251 .
## MONTHJan            -51.80      69.94  -0.741 0.462840
## MONTHJul            264.40      69.94   3.780 0.000468 ***
## MONTHJun            178.20      69.94   2.548 0.014404 *
## MONTHMar            -54.20      69.94  -0.775 0.442504
## MONTHMay            122.80      69.94   1.756 0.086077 .
## MONTHNov            -10.60      69.94  -0.152 0.880225
## MONTHOct             62.00      69.94   0.886 0.380172
## MONTHSep            121.80      69.94   1.742 0.088579 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 110.6 on 44 degrees of freedom
## Multiple R-squared:  0.9243, Adjusted R-squared:  0.8985
## F-statistic:  35.8 on 15 and 44 DF,  p-value: < 2.2e-16
```

Coefficients

- Borough: All boroughs except Queens exhibit highly significant and substantially different occurrences of shootings compared to the reference borough (Manhattan).
- Month: July demonstrates a highly significant ($p < 0.001$) higher number of shootings compared to the reference month (January). June and August indicate a significantly higher number of shootings ($p = 0.014$ and $p = 0.001$, respectively). September and May show weakly significant higher numbers of shootings ($p < 0.1$), while February shows a weakly significant lower number of shootings. The size of the effect is substantial in all cases.

Model Fit

- Multiple R-squared: 0.9243, meaning that approximately 92.43% of the variance in shootings is explained by the model.
- Adjusted R-squared: 0.8985, which adjusts the R-squared value based on the number of predictors and the sample size.
- F-statistic: 35.8 on 15 and 44 degrees of freedom, with a highly significant p-value (< 2.2e-16), indicating that the model as a whole is significant.

Interpretation

- Borough Effects: The number of shootings is significantly higher in the Bronx and Brooklyn compared to the reference borough (Manhattan). Staten Island has a significantly lower number of shootings.
- Monthly Effects: There are significant increases in shootings in July and August. Other months show mixed results, with some months showing marginally significant effects (e.g., February, May, and September).

Discussion

Model Results

Overall, the model explains a large portion of the variance in shootings, with significant contributions from certain boroughs and months. The significant p-values for many coefficients indicate strong evidence against the null hypothesis, suggesting that these factors are indeed associated with the number of shootings.

Potential Bias

Several potential biases could affect the results such as:

1. Underreporting or misclassification: Some police shooting incidents may not be accurately reported or classified, leading to underestimation or misrepresentation of the true number.
2. Demographic bias: The demographics of individuals involved in police shootings may not be representative of the overall population. Certain demographic groups may be disproportionately targeted by law enforcement, introducing bias into the data.
3. Enforcement bias: Policing practices and priorities vary between areas based on factors like crime rates, demographics, and community preferences. More heavily policed areas may employ different enforcement tactics, potentially impacting the frequency and character of police shootings.
4. Geographic Bias: The analysis did not account for differences in population numbers among the boroughs.