NYPD_Data

2024-10-15

NYPD Historical Shooting Data Project

For this project, I chose to analyze NYPD historic shooting data to understand more about how reported shootings and murders have changed over time in each of NYC's boroughs. The dataset includes data from 2006 - 2023. I chose to primarily look at data in what I would consider "public" places, so I have excluded data from unknown or unrecorded locations, as well as private residences.

Bias Identification

It is important to note for this project that there may be sources of bias both in the data collection and aggregation, and in my analysis itself.

1. The data included only includes reported incidents. One can assume the actual numbers are higher than the NYPD database would reflect.
2. I have personally excluded data that occurs at a private residence in order to examine the likelihood of the average person being shot or murdered in each of the boroughs over time. I was interested in examining if NYC has gotten more or less dangerous over time, but I have excluded the private location data under the assumption that these shootings are between parties that know eachother, rather than a random act of violence. This is again, an assumption that reflects my personal bias.

Setting Up

Some graphics in this R Markdown document require XQuartz for MacOS to run properly.

```r
# Set Up Environment
library(tidyverse)
library(lubridate)
library(forecast)
library(mgcv)
library(ggplot2)

# Read the CSV data from the URL
nyc_data <- read.csv('https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD')
```

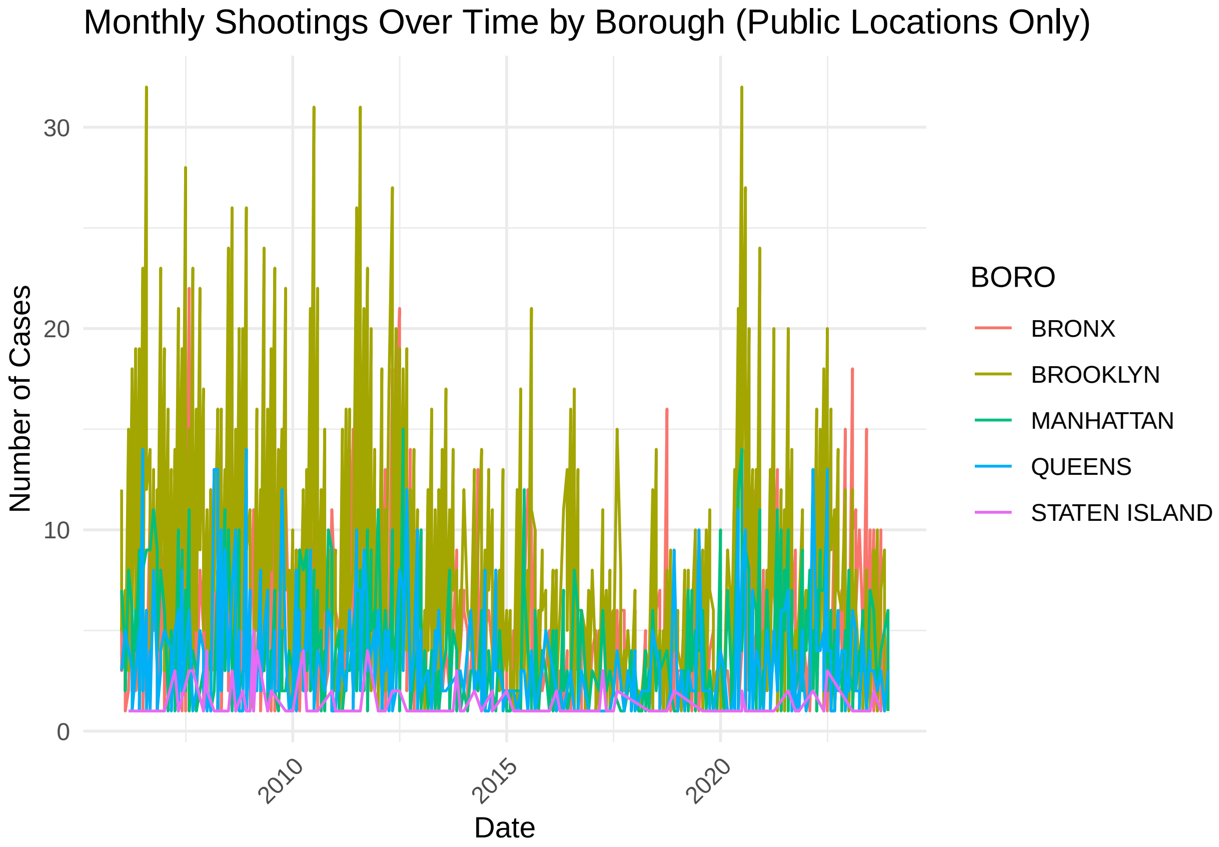
```r
# Remove columns 1, 5-8, and 17-21
data_reduced <- nyc_data %>% select(-c(1, 5:8, 17:21))

# Cleaning up empty or null entries
data_reduced <- data_reduced %>% mutate(across(everything(), ~replace_na(.x, 'UNKNOWN'))) %>% mutate(ac
data_reduced <- data_reduced %>% mutate(across(everything(), ~ifelse(. == "(null)", "UNKNOWN", .)))
```

```r
# Condensing Entries for Location to Public, Private or Unknown
locations = unique(data_reduced$LOCATION_DESC)
private <- c('MULTI DWELL - APT BUILD','PVT HOUSE')
unknown <- c('NONE','UNKNOWN')
tmp = setdiff(locations, private)
public = setdiff(tmp, unknown)

data_reduced$LOCATION_DESC <- ifelse(data_reduced$LOCATION_DESC %in% public, 'PUBLIC', data_reduced$LOCA
data_reduced$LOCATION_DESC <- ifelse(data_reduced$LOCATION_DESC %in% private, 'PRIVATE', data_reduced$L
data_reduced$LOCATION_DESC <- ifelse(data_reduced$LOCATION_DESC %in% unknown, 'UNKNOWN', data_reduced$L
```

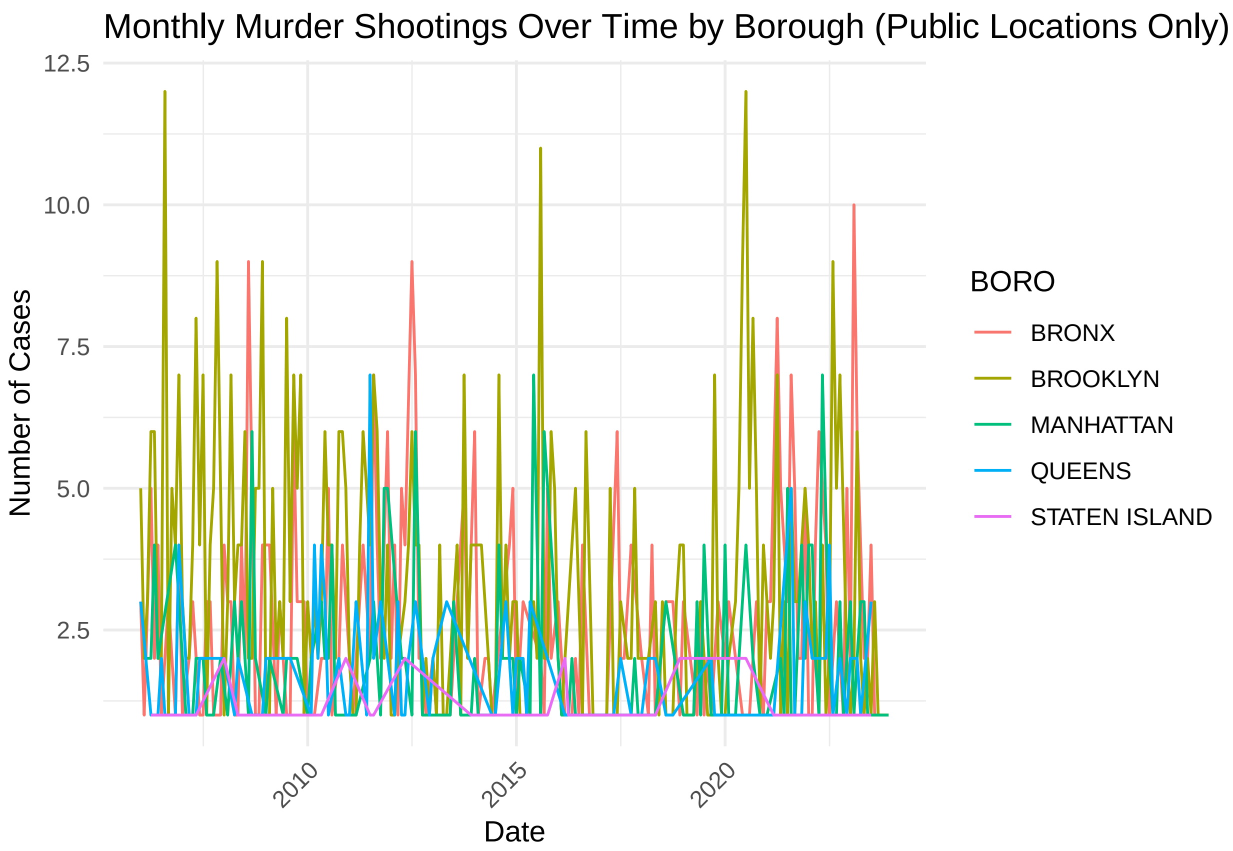
```r
#Original Location Data
view(locations)
#Condensed Location Data
new_locations = unique(data_reduced$LOCATION_DESC)
view(new_locations)
```

##Visualizing Public Shootings by Borough The below plots investigate the number of shootings in each borough, and then examine a subset of those shootings that were classified as murders. I chose to examine only the shootings and murders that occurred in public places, which was included in the dataset as the location description.

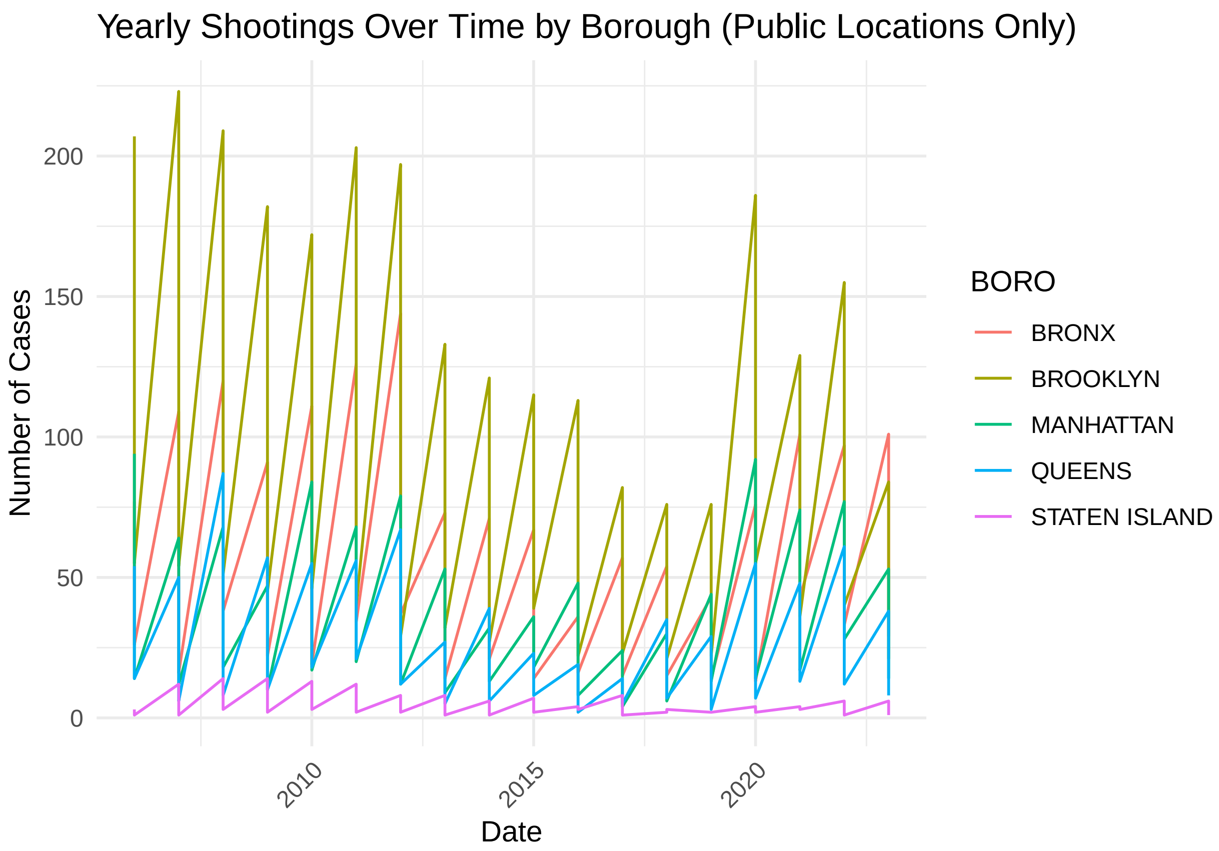
```r
#Questions to Answer
# 1. Let's examine public shootings only, by borough over time
data_reduced <- data_reduced %>% mutate(OCCUR_DATE = mdy(OCCUR_DATE))
boro_monthly_data <- data_reduced %>% mutate(year_month = floor_date(OCCUR_DATE, "month")) %>% group_by
boro_monthly_data <- boro_monthly_data %>% rename(OCCUR_DATE = year_month)

boro_public_plot <- boro_monthly_data %>% filter(LOCATION_DESC == "PUBLIC") %>% ggplot(aes(x = OCCUR_DA
print(boro_public_plot)
```



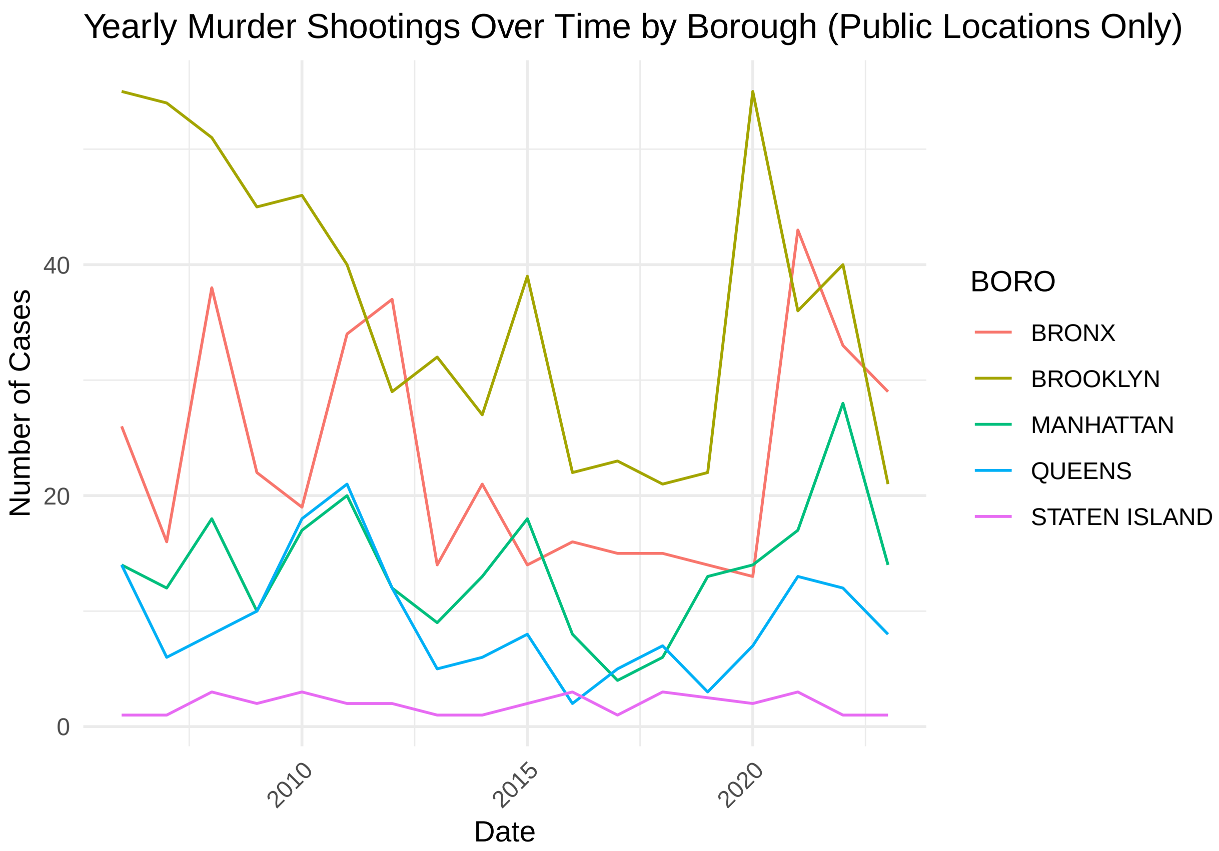
```r
# 2. Add murder flag to be true filter
boro_murder_plot <- boro_monthly_data %>% filter(LOCATION_DESC == "PUBLIC", STATISTICAL_MURDER_FLAG ==
print(boro_murder_plot)
```



```
# 3. Reduce to yearly data
yearly_data <- boro_monthly_data %>% mutate(YEAR = year(OCCUR_DATE))
boro_yearly_data <- yearly_data %>% group_by(YEAR, BORO, LOCATION_DESC, STATISTICAL_MURDER_FLAG) %>% sur

#Plot Yearly Data
boro_yearly_public_plot <- boro_yearly_data %>% filter(LOCATION_DESC == "PUBLIC") %>% ggplot(aes(x = YEA
print(boro_yearly_public_plot)
```



```
# Add murder flag
boro_yearly_murder_plot <- boro_yearly_data %>% filter(LOCATION_DESC == "PUBLIC", STATISTICAL_MURDER_FLA
print(boro_yearly_murder_plot)
```



```
#Examine Safest and Least Safe Months
filtered_data <- boro_monthly_data %>% filter(LOCATION_DESC == "PUBLIC") %>% group_by(OCCUR_DATE, BORO)
lowest_cases_per_boro <- filtered_data %>% group_by(BORO) %>% slice_min(TOTAL_CASES, n = 1) %>% ungroup
highest_cases_per_boro <- filtered_data %>% group_by(BORO) %>% slice_max(TOTAL_CASES, n = 1) %>% ungrou

year_filtered_data <- boro_yearly_data %>% filter(LOCATION_DESC == "PUBLIC") %>% group_by(YEAR, BORO) %
lowest_year_per_boro <- year_filtered_data %>% group_by(BORO) %>% slice_min(TOTAL_CASES, n = 1) %>% ung
highest_year_per_boro <- year_filtered_data %>% group_by(BORO) %>% slice_max(TOTAL_CASES, n = 1) %>% un
```

Examining High and Low Records

The tables below show the highest and lowest shooting recorded for each borough.

```
print(lowest_year_per_boro)
```

```
## # A tibble: 5 x 3
##    YEAR BORO          TOTAL_CASES
##   <dbl> <chr>               <int>
## 1  2016 BRONX                  52
## 2  2018 BROOKLYN               97
## 3  2017 MANHATTAN              28
## 4  2017 QUEENS                 19
## 5  2019 STATEN ISLAND           2
```

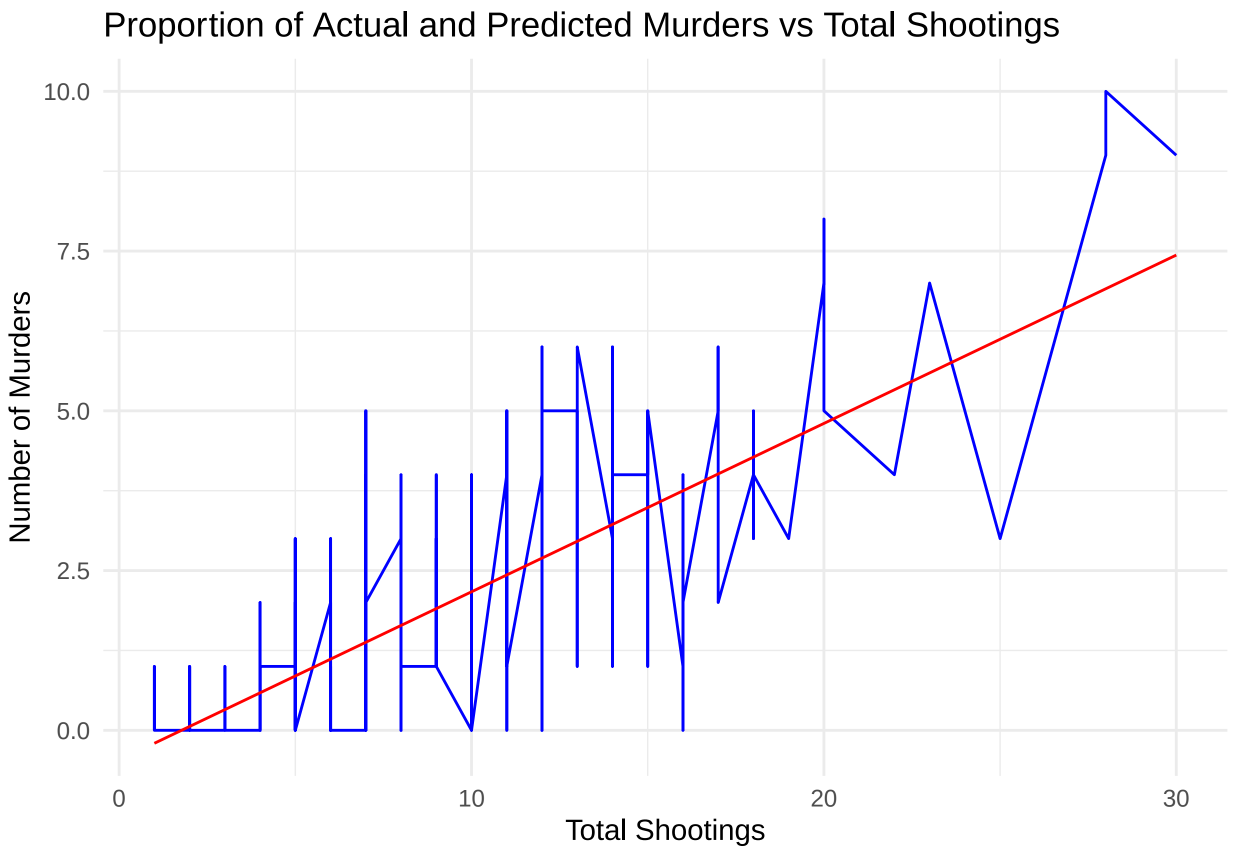
```
print(highest_year_per_boro)
```

```
## # A tibble: 5 x 3
##    YEAR BORO          TOTAL_CASES
##   <dbl> <chr>               <int>
## 1  2012 BRONX                 181
## 2  2007 BROOKLYN              277
## 3  2006 MANHATTAN             108
## 4  2008 QUEENS                 95
## 5  2008 STATEN ISLAND          17
```

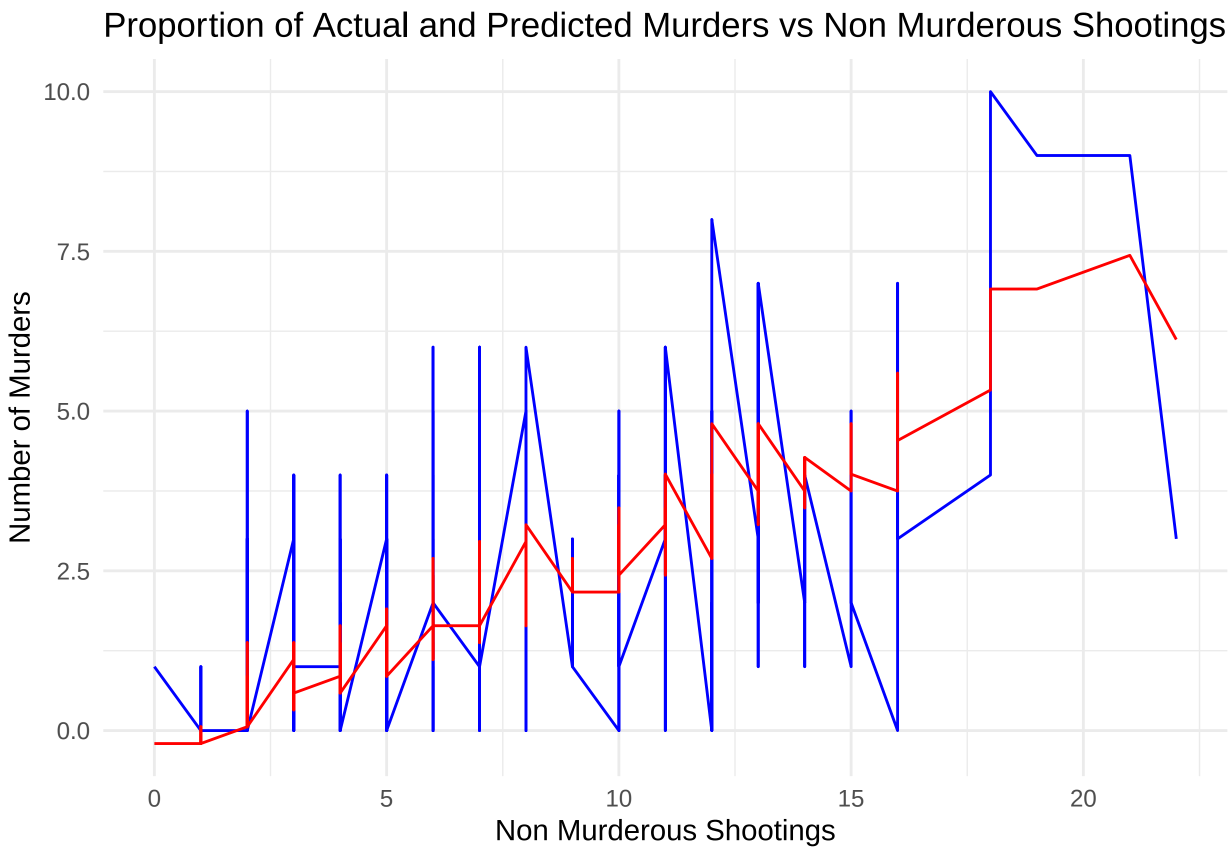
Predicting Murders as a Function of Total Shootings

Here I use a linear model to examine the number of murders compared to the total shootings. The plots below use the same model, but plot against two different metrics, 1. The Overall Shootings and 2. The Non Murderous Shootings.

```
#Add Linear Model to either public or private shootings
boro_murders_shootings <- boro_monthly_data %>% filter(LOCATION_DESC == "PUBLIC") %>% group_by(BORO, OC
bronx_murders_shootings <- boro_murders_shootings %>% filter(BORO == 'BRONX')
mod = lm(MURDERS ~ TOTAL, data = bronx_murders_shootings)
bronx_pred <- bronx_murders_shootings %>% mutate(pred = predict(mod))


bronx_pred %>% ggplot() + geom_line(aes(x = TOTAL, y = MURDERS), color = "blue") + geom_line(aes(x = TO
```



```
bronx_pred %>% ggplot() + geom_line(aes(x = SHOOTINGS, y = MURDERS), color = "blue") + geom_line(aes(x =
```


Proportion of Actual and Predicted Murders vs Non Murderous Shootings
Number of Murders
Non Murderous Shootings