NYPD_Shooting_Incident

M. Li

2024-12-23

Project Step 1: Start an Rmd Document

In this analysis, we explore the NYPD shooting incidents dataset.

```
knitr::opts_chunk$set(echo = TRUE)
# Load necessary libraries
library(magrittr)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x tidyr::extract()   masks magrittr::extract()
## x dplyr::filter()    masks stats::filter()
## x dplyr::lag()       masks stats::lag()
## x purrr::set_names() masks magrittr::set_names()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(lubridate)
library(ggplot2)
```

```
# Access the file and load data
url_NYPD <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
NYPD <- read.csv(url_NYPD)
```

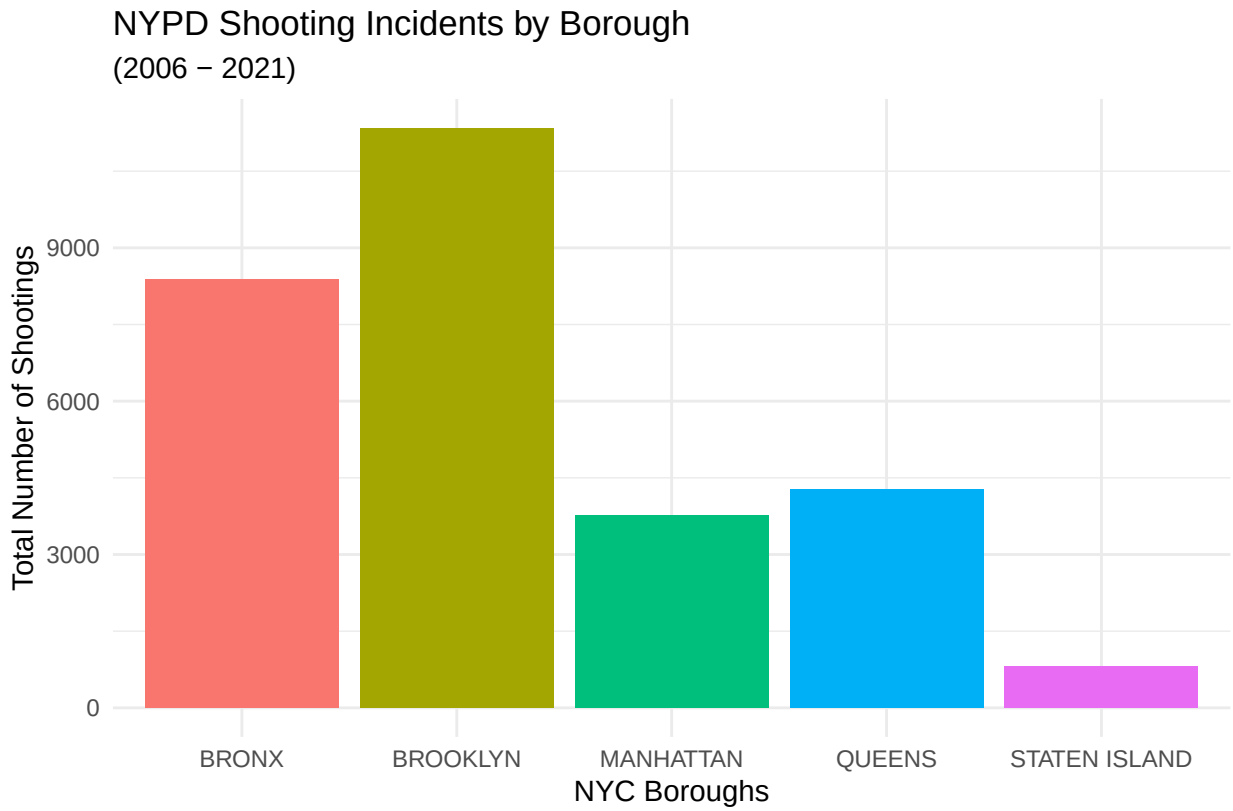
Project Step 2: Tidy and Transform Data

```
NYPD_clean <- NYPD %>%
  select(c("OCCUR_DATE", "OCCUR_TIME", "BORO", "PRECINCT",
           "STATISTICAL_MURDER_FLAG", "VIC_AGE_GROUP", "VIC_SEX", "VIC_RACE")) %>%
  mutate(OCCUR_DATE = mdy(OCCUR_DATE),
         OCCUR_TIME = hms(OCCUR_TIME),
         STATISTICAL_MURDER_FLAG = as.logical(STATISTICAL_MURDER_FLAG),
         Shootings = 1,
         Year = year(OCCUR_DATE))
```

Project Step 3: Add Visualizations and Analysis

This visualization will show the distribution of shooting incidents across New York City boroughs. The bar lengths represent the total number of incidents in each borough.

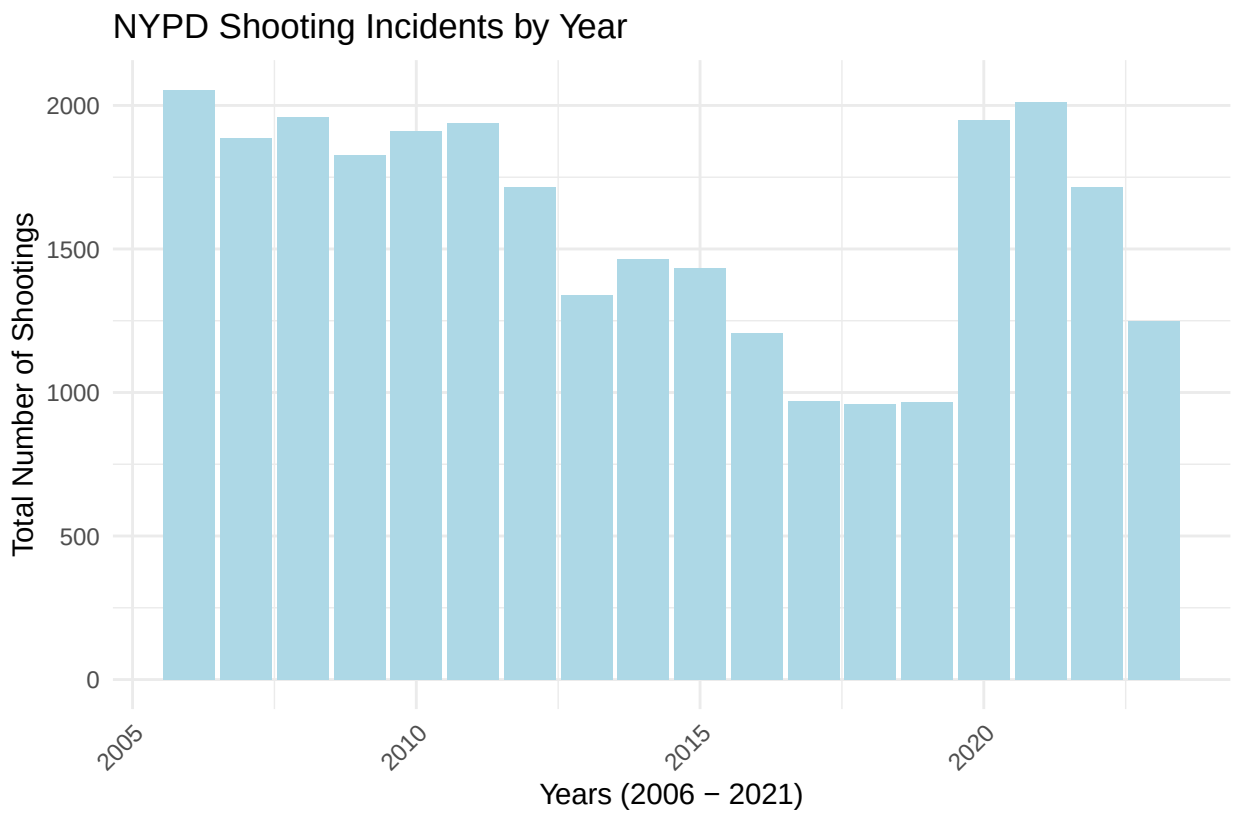
```
# Bar plot of shooting incidents by borough
NYPD_clean %>%
  ggplot(aes(x = BORO, fill = BORO)) +
  geom_bar() +
  labs(
    title = "NYPD Shooting Incidents by Borough",
    subtitle = "(2006 - 2021)",
    x = "NYC Boroughs",
    y = "Total Number of Shootings",
    caption = "(Figure - 1)"
  ) +
  theme_minimal() +  # Optional: gives a clean, minimal theme
  theme(legend.position = "none")
```



This will show the trend in shooting incidents over time (from 2006 to 2021). The trends over time might reveal if certain years had higher or lower shooting incidents.

```
# Bar plot of shooting incidents by year
NYPD_clean %>%
  ggplot(aes(x = Year)) +
  geom_bar(fill = "lightblue", show.legend = FALSE) +
```

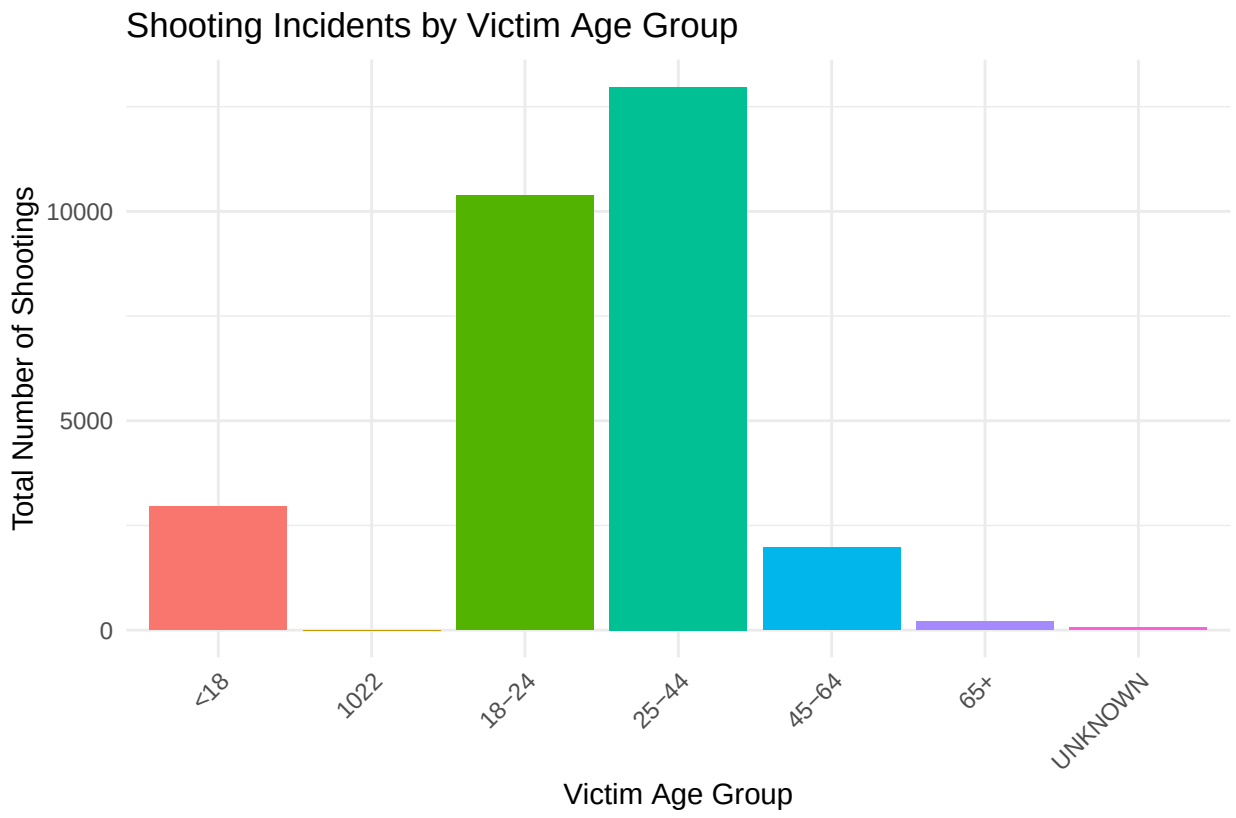
```
  labs(
    title = "NYPD Shooting Incidents by Year",
    x = "Years (2006 - 2021)",
    y = "Total Number of Shootings",
    caption = "(Figure - 2)"
  ) +
  theme_minimal() +  # Optional: gives a clean, minimal theme
  theme(axis.text.x = element_text(angle = 45, hjust = 1))  # Rotate x-axis labels for better readabili
```



(Figure − 2)

This visualization will show how shooting incidents are distributed across different age groups of victims. By visualizing the incidents by age group, we can identify if younger or older age groups are more likely to be involved in shooting incidents. For example, it might reveal that younger age groups (e.g., 18-24) are disproportionately involved.

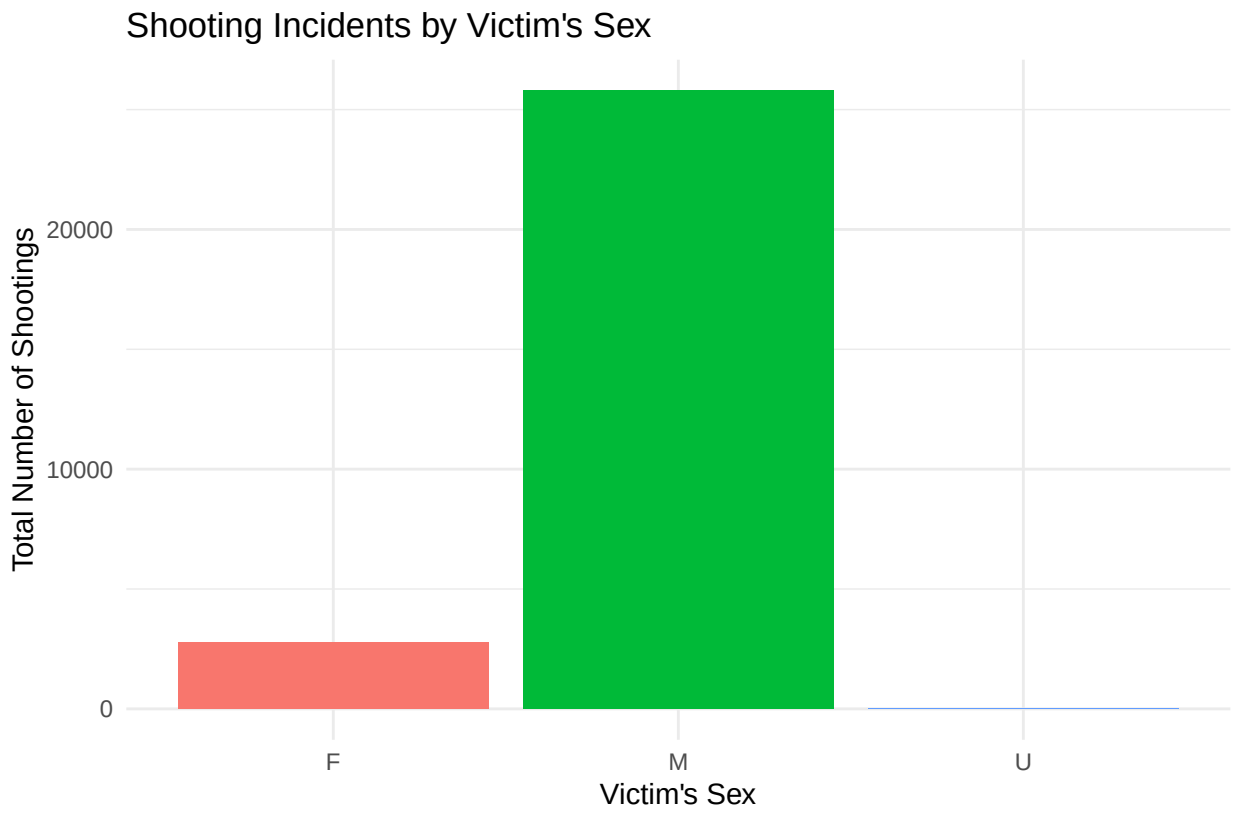
```
# Shooting Incidents by Victim Age Group
NYPD_clean %>%
  ggplot(aes(x = VIC_AGE_GROUP, fill = VIC_AGE_GROUP)) +
  geom_bar() +
  labs(
    title = "Shooting Incidents by Victim Age Group",
    x = "Victim Age Group",
    y = "Total Number of Shootings",
    caption = "(Figure - 3)"
  ) +
  theme_minimal() +
  theme(legend.position = "none", axis.text.x = element_text(angle = 45, hjust = 1)) # Rotate x labels
```



(Figure – 3)

This plot will show how shooting incidents are distributed by the gender of the victims. The incidents by victim sex can help reveal whether one gender (typically male, based on crime statistics) is more likely to be involved in shooting incidents. If the data suggests an imbalance, it may warrant further investigation into the reasons behind this pattern.

```
# Shooting Incidents by Victim's Sex
NYPD_clean %>%
  ggplot(aes(x = VIC_SEX, fill = VIC_SEX)) +
  geom_bar() +
  labs(
    title = "Shooting Incidents by Victim's Sex",
    x = "Victim's Sex",
    y = "Total Number of Shootings",
    caption = "(Figure - 4)"
  ) +
  theme_minimal() +
  theme(legend.position = "none")
```



(Figure – 4)

Project Step 4: Add Bias Identification

1. Geographical Bias:

- Issue: Different boroughs have different population sizes, densities, and socio-economic conditions. High-density areas (e.g., Manhattan, Brooklyn) will naturally have a higher number of shootings simply due to the larger population. This might lead to misinterpretation if we don't account for population density.
- Impact: If population size or density is not accounted for, you might incorrectly assume that boroughs like Brooklyn or the Bronx have a "higher" crime rate, even though the incidents might just be more frequent due to population size.

2. Seasonality:

- Issue: Shooting incidents may show seasonal patterns (e.g., spikes in warmer months due to increased outdoor activities or social gatherings). This might skew the data depending on which months are included in each year.
- Impact: If these patterns are not considered, year-over-year comparisons might overlook seasonal variations, leading to false conclusions about trends.

3. Bias in Crime Categorization:

- Issue: Young people might be involved in certain types of crime (e.g., gang-related shootings) that are more likely to be categorized differently, which could affect the way incidents are documented or reported.

- Impact: This could lead to overrepresentation of certain age groups in the dataset, particularly those in more visible and heavily policed areas, while underreporting incidents in other, less-affected demographics.

4. Gender Bias in Policing and Crime Reporting:

- Issue: Gender bias within law enforcement or data collection practices could skew reporting. For instance, a higher priority may be given to documenting incidents involving male victims in certain contexts (e.g., gang-related shootings), while incidents involving women might be underreported or misclassified.
- Impact: This could result in a skewed representation of gender in the shooting data, potentially underrepresenting female victims.