Introduction

This report provides a data analysis of shooting incidents that occurred in New York City, starting from 2006 through the end of the previous calendar year. It break downs data based on **perpetrators' age group** and **boroughs**. Data was obtained through https://catalog.data.gov/dataset/nypd-shooting-incident-data-historic.

Perpetrators' Age Group

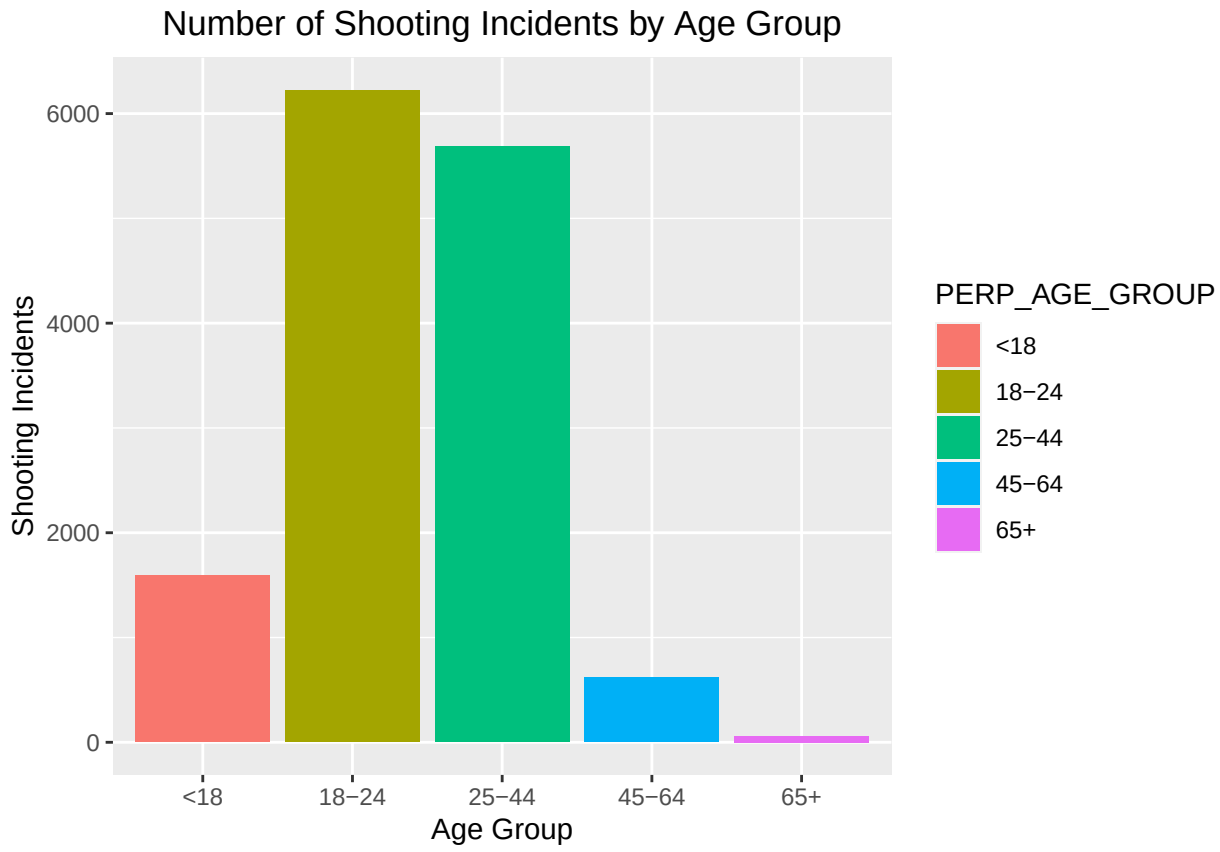
The perpetrators' data for shooting incidents were provided in the following groups:

- <18
- 18-24
  - *Most number of shooting incidents*
- 25-44
- 45-64
- 65
  - *Least number of shooting incidents*

In order to visualize the shooting incidents based on age groups, all columns besides PERP_AGE_GROUP were removed. There were vast amounts of data that were not viable to use for visualization. Those data were blank, (null), UNKNOWN and others that did not go into the provided age groups. Besides UNKNOWN, other categories appear to be either incorrectly inputted or not inputted at all.

```
PERP_AGE_GROUP <- data %>% select(PERP_AGE_GROUP)
PERP_AGE_GROUP_data <- PERP_AGE_GROUP %>%
  filter(PERP_AGE_GROUP != "",
         PERP_AGE_GROUP !="UNKNOWN",
         PERP_AGE_GROUP != "1020",
         PERP_AGE_GROUP != "940",
         PERP_AGE_GROUP != "224",
         PERP_AGE_GROUP != "(null)") %>%
  group_by(PERP_AGE_GROUP) %>%
  summarise(INCIDENTS = n()) %>%
  arrange(PERP_AGE_GROUP)

ggplot(PERP_AGE_GROUP_data, aes(PERP_AGE_GROUP, INCIDENTS, fill=PERP_AGE_GROUP))+
  geom_bar(stat = "identity")+
  labs(title = "Number of Shooting Incidents by Age Group",
       x = "Age Group",
       y = "Shooting Incidents")+
  theme(plot.title = element_text(hjust = 0.5))
```



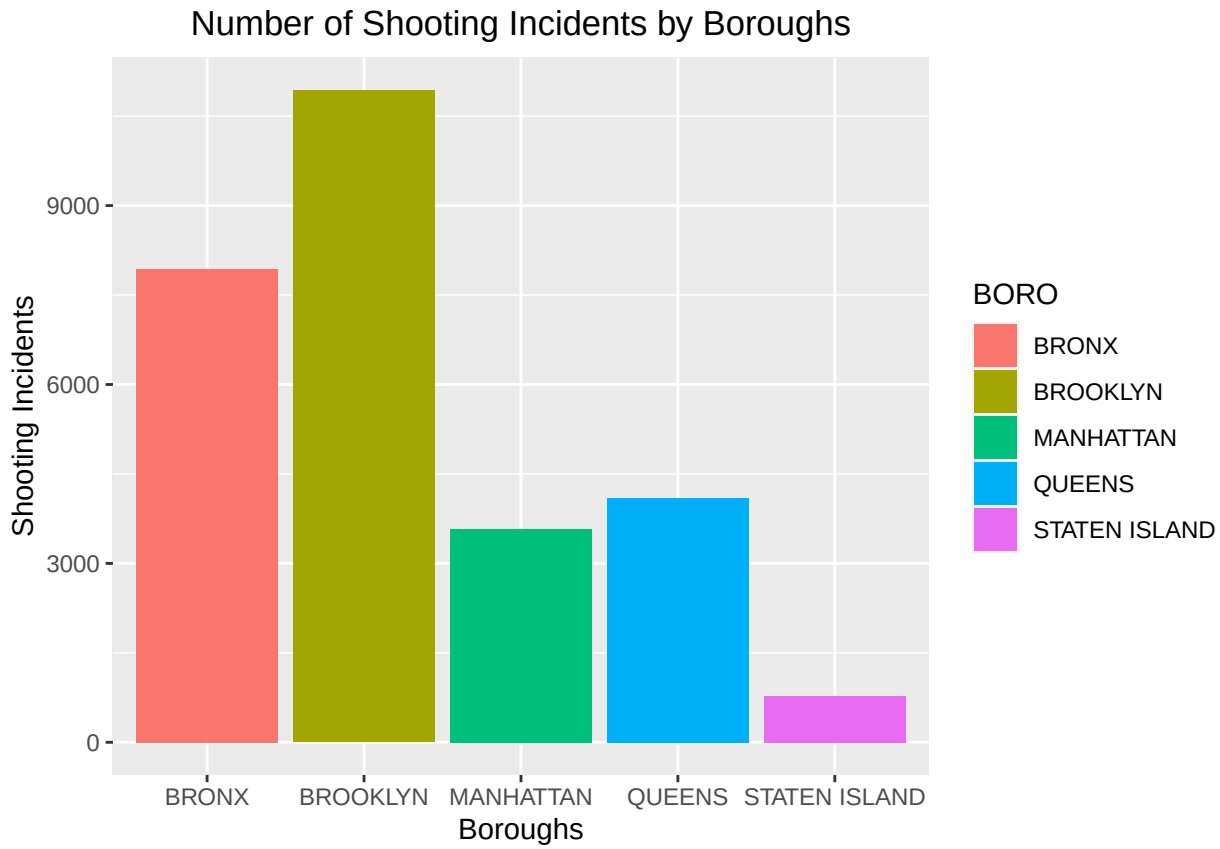

```
##  PERP_AGE_GROUP       INCIDENTS
##  Length:5           Min.   :  60
##  Class :character   1st Qu.: 617
##  Mode  :character   Median :1591
##                     Mean   :2835
##                     3rd Qu.:5687
##                     Max.   :6222
```

Majority of shooting incidents occurred between the ages 18-44. There is a sharp decline after the age of 45. Based on the data, it appears that most incidents occur in the person's most active stage in life. I would investigate the age group 45-64 because the decline is significant. Is there a decline because the perpetrators before the age of 45 are incarcerated and therefore are unable to commit any additional shootings? Or is it that as people approaches the age 50, there are other environmental factors that influence their decisions?

Boroughs

There are five boroughs in New York City. The following lists the boroughs from the highest shooting incidents to the lowest.

1. Brooklyn
2. Bronx
3. Queens
4. Manhattan
5. Staten Island




```
##      BORO             INCIDENTS
##  Length:5           Min.   :  776
##  Class :character   1st Qu.: 3572
##  Mode  :character   Median : 4094
##                     Mean   : 5462
##                     3rd Qu.: 7937
##                     Max.   :10933
```

Based on the table above, I would investigate why Brooklyn has the highest shooting incidents. Is it because of the social economics factors? If so, what are those factors? Bronx has the 2nd highest incident. With Manhattan and Queens between Brooklyn and Bronx, I do not think the reasoning is because of their influence on each other. I would dive into more data to see the similaries between Brooklyn and Bronx as well as the differences between them from Manhattan and Queens.

Conclusion

The highest shooting incidents in New York City occurred in Brooklyn and with the perpetrators between the ages 18-24. Below is a table summarizing the highest to the lowest incidents by categories discussed.

```
##    Boroughs_NYC Age_Groups
## 1      Brooklyn      18-24
## 2         Bronx      25-44
## 3        Queens        <18
## 4     Manhattan      45-64
## 5 Staten Island        65+
```

A possible source of bias are outliers. There were data in the NYPD shooting incident document that does not correlate to any of the groups assigned. My personal bias is to ignore these outliers as I assume that they are either incorrect or an error. To mitigate my bias, I looked at each individual outliers to determine if they are essential to my analysis.

```
## [1] "Data Science as a Field, Summer 2023"
```