NYPD_Shooting_Incidents

2023-07-04

NYPD Shooting Incidents

This projects examines NYPD Shooting Incidents using historical data. Data is sourced from https://catalog.data.gov/dataset/nypd-shooting-incident-data-historic

This report will attempt to examine multiple areas of this data to see how it relates to shootings, including physical location, race, and time.

First we will import the data and explore the dataset.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)
url_in <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"


shooting <- read_csv(url_in)
```

```
## Rows: 27312 Columns: 21
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (12): OCCUR_DATE, BORO, LOC_OF_OCCUR_DESC, LOC_CLASSFCTN_DESC, LOCATION...
## dbl   (7): INCIDENT_KEY, PRECINCT, JURISDICTION_CODE, X_COORD_CD, Y_COORD_CD...
## lgl   (1): STATISTICAL_MURDER_FLAG
## time  (1): OCCUR_TIME
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
summary(shooting)
```

```
##   INCIDENT_KEY        OCCUR_DATE         OCCUR_TIME           BORO
##  Min.   :  9953245   Length:27312       Length:27312      Length:27312
##  1st Qu.: 63860880   Class :character   Class1:hms        Class :character
##  Median : 90372218   Mode  :character   Class2:difftime   Mode  :character
##  Mean   :120860536                      Mode  :numeric
##  3rd Qu.:188810230
```

```
##  Max.   :261190187
##
##  LOC_OF_OCCUR_DESC     PRECINCT      JURISDICTION_CODE LOC_CLASSFCTN_DESC
##  Length:27312       Min.   :  1.00   Min.   :0.0000    Length:27312
##  Class :character   1st Qu.: 44.00   1st Qu.:0.0000    Class :character
##  Mode  :character   Median : 68.00   Median :0.0000    Mode  :character
##                     Mean   : 65.64   Mean   :0.3269
##                     3rd Qu.: 81.00   3rd Qu.:0.0000
##                     Max.   :123.00   Max.   :2.0000
##                                      NA's   :2
##  LOCATION_DESC      STATISTICAL_MURDER_FLAG PERP_AGE_GROUP
##  Length:27312       Mode :logical           Length:27312
##  Class :character   FALSE:22046             Class :character
##  Mode  :character   TRUE :5266              Mode  :character
##
##
##
##
##    PERP_SEX           PERP_RACE         VIC_AGE_GROUP        VIC_SEX
##  Length:27312       Length:27312       Length:27312       Length:27312
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    VIC_RACE             X_COORD_CD        Y_COORD_CD        Latitude
##  Length:27312       Min.   : 914928   Min.   :125757   Min.   :40.51
##  Class :character   1st Qu.:1000028   1st Qu.:182834   1st Qu.:40.67
##  Mode  :character   Median :1007731   Median :194487   Median :40.70
##                     Mean   :1009449   Mean   :208127   Mean   :40.74
##                     3rd Qu.:1016838   3rd Qu.:239518   3rd Qu.:40.82
##                     Max.   :1066815   Max.   :271128   Max.   :40.91
##                                                        NA's   :10
##    Longitude         Lon_Lat
##  Min.   :-74.25   Length:27312
##  1st Qu.:-73.94   Class :character
##  Median :-73.92   Mode  :character
##  Mean   :-73.91
##  3rd Qu.:-73.88
##  Max.   :-73.70
##  NA's   :10
```

The OCCUR_DATE column appears to be a string value, so we can transform it into a date.

```
shooting <- shooting %>%
  mutate(OCCUR_DATE = mdy(OCCUR_DATE))
```

There are also a large number of columns that we will not be using for our analysis. These can be removed from the dataset.

```
shooting <- shooting %>%
  select(-OCCUR_TIME, -LOC_OF_OCCUR_DESC, -JURISDICTION_CODE,
  -LOC_CLASSFCTN_DESC, -LOCATION_DESC, -X_COORD_CD,
  -Y_COORD_CD, -Latitude, -Longitude, -Lon_Lat)
```

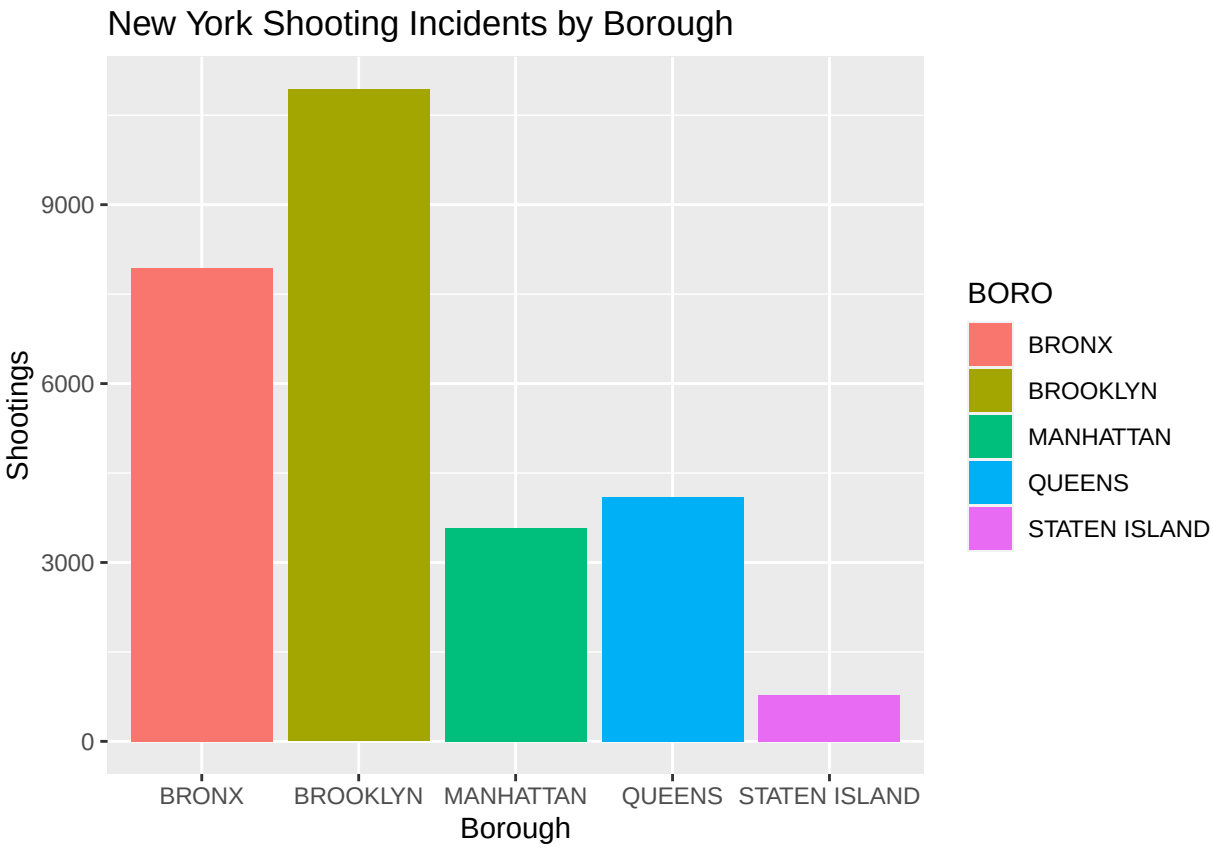
```
head(shooting)
```

```
## # A tibble: 6 x 11
##   INCIDENT_KEY OCCUR_DATE BORO    PRECINCT STATISTICAL_MURDER_F~1 PERP_AGE_GROUP
##          <dbl> <date>     <chr>      <dbl> <lgl>                  <chr>
## 1    228798151 2021-05-27 QUEENS       105 FALSE                  <NA>
## 2    137471050 2014-06-27 BRONX         40 FALSE                  <NA>
## 3    147998800 2015-11-21 QUEENS       108 TRUE                   <NA>
## 4    146837977 2015-10-09 BRONX         44 FALSE                  <NA>
## 5     58921844 2009-02-19 BRONX         47 TRUE                   25-44
## 6    219559682 2020-10-21 BROOKL~       81 TRUE                   <NA>
## # i abbreviated name: 1: STATISTICAL_MURDER_FLAG
## # i 5 more variables: PERP_SEX <chr>, PERP_RACE <chr>, VIC_AGE_GROUP <chr>,
## #   VIC_SEX <chr>, VIC_RACE <chr>
```

Shootings By Borough

One area that we can examine is the shootings by Borough.

```
boro_bar <- ggplot(shooting, aes(x=BORO)) +
  geom_bar(aes(fill = BORO)) +
  labs(title = 'New York Shooting Incidents by Borough', x = 'Borough', y = 'Shootings')

boro_bar
```



This shows that Brooklyn has had the most total shootings out of the New York Boroughs, followed by the Bronx. Brooklyn's large number of shootinga may be related to the fact that it has the largest population of New York's boroughs.

Shootings By Race

Another area that we could examine is a breakdown of shootings by race. Before we begin we need to clean up some of the race data to remove null values.

```
unique(shooting[c('PERP_RACE')])
```

```
## # A tibble: 9 x 1
##   PERP_RACE
##   <chr>
## 1 <NA>
## 2 BLACK
## 3 UNKNOWN
## 4 BLACK HISPANIC
## 5 ASIAN / PACIFIC ISLANDER
## 6 WHITE HISPANIC
## 7 WHITE
## 8 (null)
## 9 AMERICAN INDIAN/ALASKAN NATIVE
```

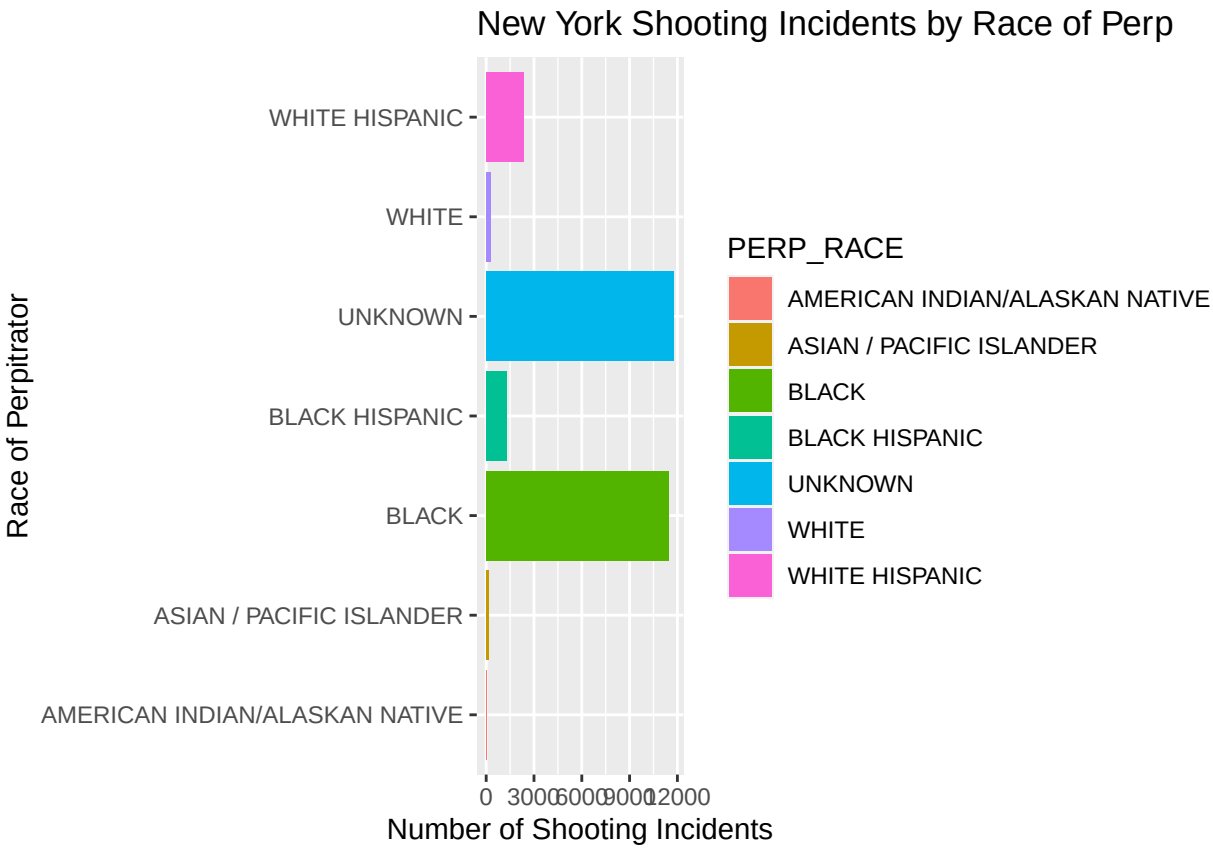
The data contains races listed as '(null)' and NA. Since there is also an 'UNKNOWN' category, we should combine all other null-type values to simply be categorized as unknown.

```
shooting$PERP_RACE[shooting$PERP_RACE == '(null)'] <- 'UNKNOWN'
shooting$PERP_RACE[is.na(shooting$PERP_RACE)] <- 'UNKNOWN'

unique(shooting[c('PERP_RACE')])
```

```
## # A tibble: 7 x 1
##   PERP_RACE
##   <chr>
## 1 UNKNOWN
## 2 BLACK
## 3 BLACK HISPANIC
## 4 ASIAN / PACIFIC ISLANDER
## 5 WHITE HISPANIC
## 6 WHITE
## 7 AMERICAN INDIAN/ALASKAN NATIVE
```

Now we can visualize the shooting incidents by the race of the perpetrator.

```
race_bar <- ggplot(shooting, aes(x=PERP_RACE)) +
  geom_bar(aes(fill = PERP_RACE)) +
  labs(title = 'New York Shooting Incidents by Race of Perp', x = 'Race of Perpitrator', y = 'Number of
  coord_flip()

race_bar
```



This shows that the race of the perpetrator is not known in a large number of the incidents.For incidents in which their is race data, black perpetrators were the most frequent. To get a better understanding of how race relates to the shooting incidents, more accurate data would be beneficial. It would also be beneficial to better understand the demographics of New York's overall population to provide context for the racial breakdown of shooting incidents.

Shootings Over Time

We can also examine the shooting incidents over time. Since the shootings are given by specific dates, we can instead view them by year by adding a year value to our dataframe. We can then group the number of shootings by year.

```
shooting$Year <- format(shooting$OCCUR_DATE, format="%Y")
unique(shooting$Year)
```

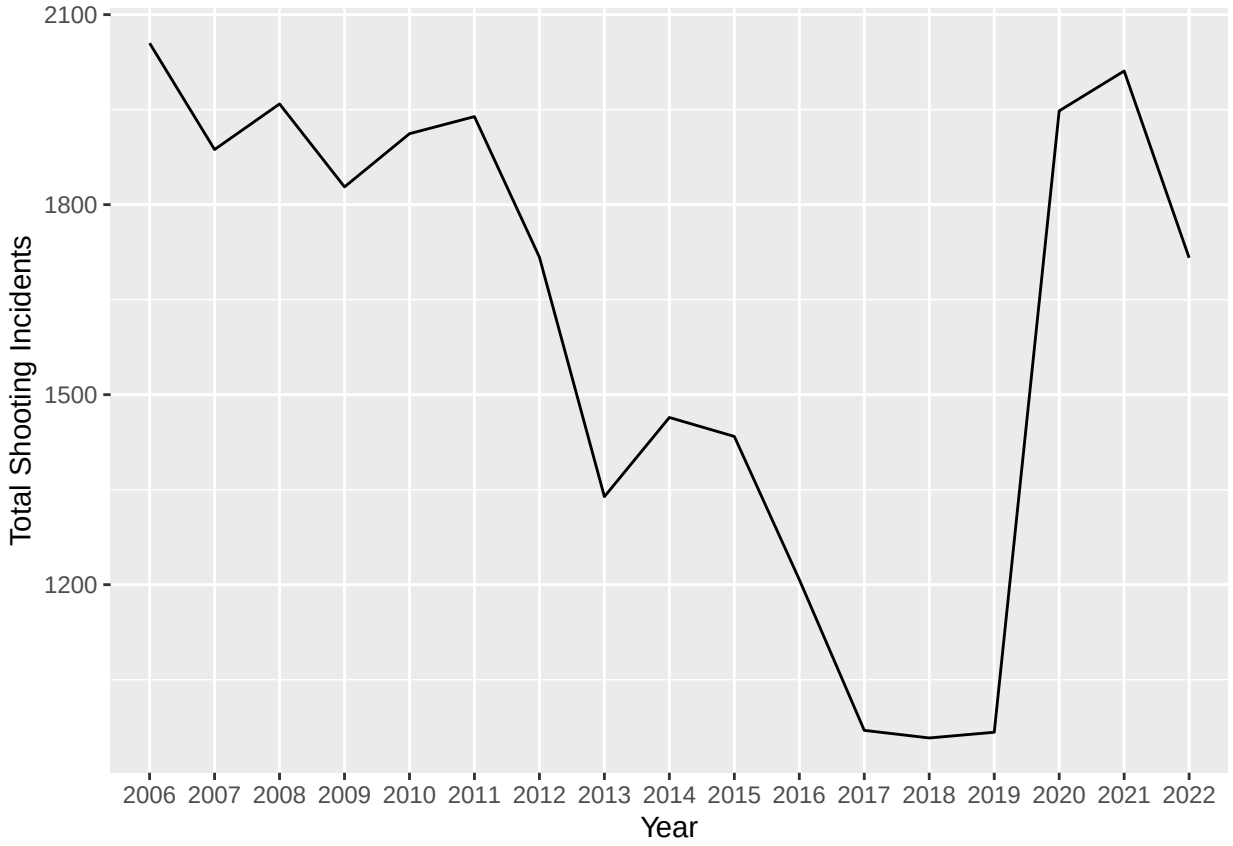
```
##  [1] "2021" "2014" "2015" "2009" "2020" "2012" "2010" "2011" "2008" "2007"
## [11] "2006" "2017" "2016" "2013" "2018" "2019" "2022"
```

```
shootings_by_year <-shooting %>%
  group_by(Year) %>%
  summarise(total_shootings = n()) %>%
  arrange(Year)

head(shootings_by_year)
```

```
## # A tibble: 6 x 2
##   Year  total_shootings
##   <chr>           <int>
## 1 2006             2055
```

```
## 2 2007              1887
## 3 2008              1959
## 4 2009              1828
## 5 2010              1912
## 6 2011              1939
```

```
shootings_year_line <- ggplot(shootings_by_year, aes(x = Year, y = total_shootings, group = 1)) +
  geom_line() +
  labs(totle = "New York Shooting Incidents by Year", x = 'Year', y = 'Total Shooting Incidents')

shootings_year_line
```



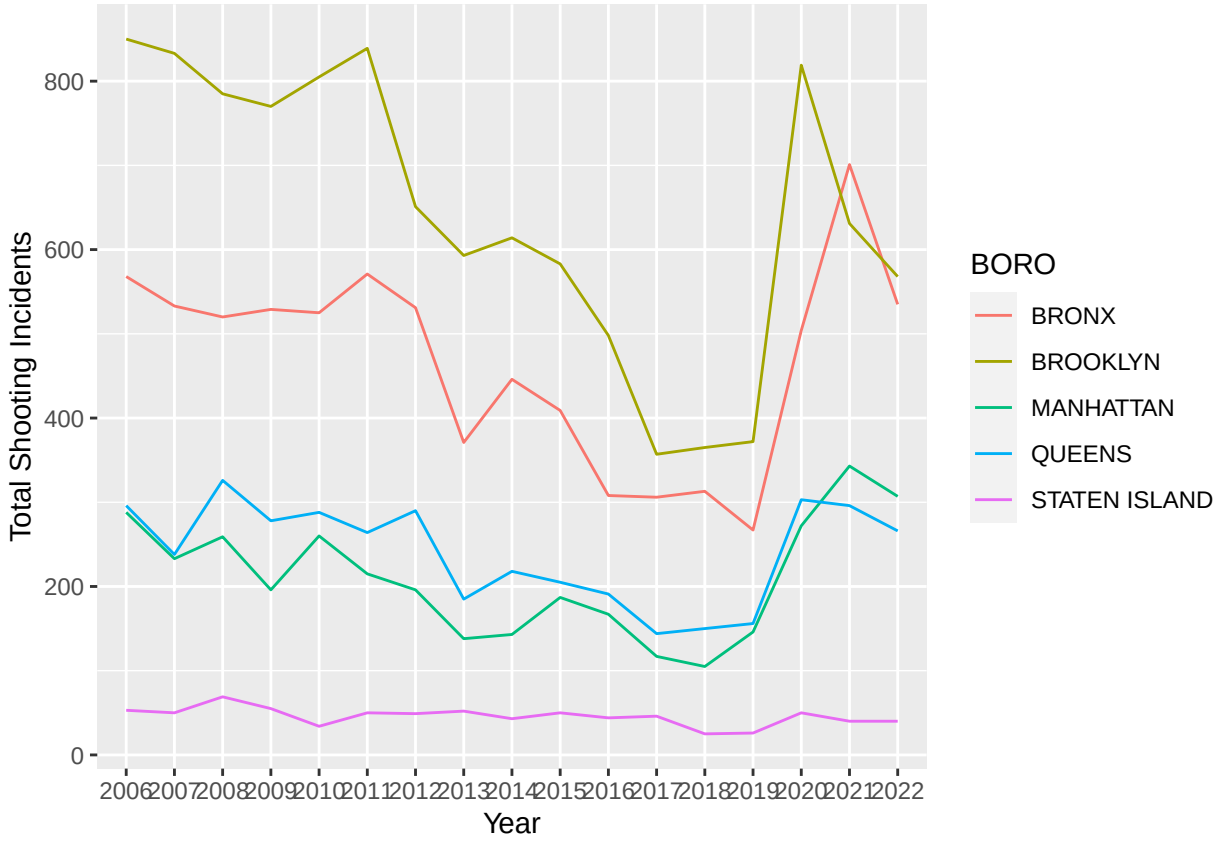
Interestingly, the graph shows a downward trend of shooting incidents from 2006 - 2019. In 2019 there is a large spike back up to around 2011 levels. One possible explanation for this spike is that it coincides with a difficult time in history with covid as well as political unrest and an election in 2020. Further examination of this time period would be beneficial for understanding this spike.

We can additionally return to our previous investigation of shootings by boroughs, now broken down over time. We can now look at the shootings by year for each borough to see how they compare to the total shootings in New York over time.

```
shootings_by_year_boro <-shooting %>%
    group_by(Year, BORO) %>%
    summarise(total_shootings = n()) %>%
    arrange(Year)
```

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

```
shootings_year_boros <- ggplot(shootings_by_year_boro, aes(x = Year, y = total_shootings, group = BORO))
  geom_line(aes(color = BORO)) +
  labs(totle = "New York Shooting Incidents by Year", x = 'Year', y = 'Total Shooting Incidents')

shootings_year_boros
```



This shows that the shootings across the boroughs follow a similar pattern to that of New York as a whole, with no significant outliers. This may indicate that the spike around 2019 was due to a wide reaching phenomenon, rather than a localized issue in only one of the boroughs. Interestingly during the spike in shootings around 2019 Bronx briefly surpasses Brooklyn and Manhattan briefly surpasses Queens.

Conclusion

This report demonstrates a broad overview of historical shooting incident data in New York city. The report examined the shooting data by location, race, and as a trend over time. With this initial analysis, some hypothesis can be made such as borough population relating to number of shootings. Additionally, there may be evidence that covid lockdowns and a tense political climate in 2019-2020 led to an increase in shooting incidents. More data and analysis would be needed to come to strong conclusions around these issues - such as demographic and economic analysis of each borough that may impact shooting rates, or additional data around the year 2019 that might explain the spike in shooting incidents. Additionally, more information would be needed for an analysis regarding race as much of the data was unkown.