Evaluating NYPD Historical Shooting Incidents

DTSA 5301, University of Colorado

March 6th, 2022

OBJECTIVE

The objective of this analysis is to determine the demographics of when and where historical shootings in the City of New York using data from the New York Police Department, evaluated over a 15 year period from January 1st, 2006 to December 29th, 2020. Throughout the analysis any sources of bias such as age or race will be identified and objectively evaluated to ensure a fair, balanced, and inclusive summary has been generated.

IMPORTING & TIDYING DATA

DATA IMPORT:

The first step will be to import the shootings data from the City of New York website. A summary is run to determine the initial size of the imported data frame and the type of data in each column. This is done to determine if it is required to change the data types of any columns in the data frame after it has been imported. The data summary indicates that the data frame has 23,585 rows and 19 columns, and that a few of the columns are of the data class "character".

```
#Import historical shootings data from the City of New York website
url <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv"
data_import <- read_csv(url)
```

```
#Determine the initial shape and data types of the data frame
glimpse(data_import)
```

```
## Rows: 23,585
## Columns: 19
## $ INCIDENT_KEY            <dbl> 24050482, 77673979, 203350417, 80584527, 90843~
## $ OCCUR_DATE              <chr> "08/27/2006", "03/11/2011", "10/06/2019", "09/~
## $ OCCUR_TIME              <time> 05:35:00, 12:03:00, 01:09:00, 03:35:00, 21:16~
## $ BORO                    <chr> "BRONX", "QUEENS", "BROOKLYN", "BRONX", "QUEEN~
## $ PRECINCT                <dbl> 52, 106, 77, 40, 100, 67, 77, 81, 101, 106, 71~
## $ JURISDICTION_CODE       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ LOCATION_DESC           <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
## $ STATISTICAL_MURDER_FLAG <lgl> TRUE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE~
## $ PERP_AGE_GROUP          <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
## $ PERP_SEX                <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
## $ PERP_RACE               <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
```

```
## $ VIC_AGE_GROUP           <chr> "25-44", "65+", "18-24", "<18", "18-24", "<18"~
## $ VIC_SEX                 <chr> "F", "M", "F", "M", "M", "M", "M", "M", "M", "~
## $ VIC_RACE                <chr> "BLACK HISPANIC", "WHITE", "BLACK", "BLACK", "~
## $ X_COORD_CD              <dbl> 1017542, 1027543, 995325, 1007453, 1041267, 10~
## $ Y_COORD_CD              <dbl> 255918.9, 186095.0, 185155.0, 233952.0, 157133~
## $ Latitude                <dbl> 40.86906, 40.67737, 40.67489, 40.80880, 40.597~
## $ Longitude               <dbl> -73.87963, -73.84392, -73.96008, -73.91618, -7~
## $ Lon_Lat                 <chr> "POINT (-73.87963173099996 40.86905819000003)"~
```

TIDYING DATA:

The data can now be organized in a tidy fashion where each variable is in it's own column, and each observation is in it's own row. The columns which are of the data type "character" will be changed to the type "factor" to allow for further analysis by grouping the data into common groups. After tidying has been completed the resulting "shootings" data frame has 5,340 rows and 18 columns.

```
#Define column names to be changed into factor data type
column_names <- c('BORO','PRECINCT','JURISDICTION_CODE',
                  'LOCATION_DESC','STATISTICAL_MURDER_FLAG',
                  'PERP_AGE_GROUP','PERP_SEX','PERP_RACE',
                  'VIC_AGE_GROUP','VIC_SEX','VIC_RACE')

#Create "shootings" data frame by changing column data types and removing errant values
shootings <- data_import %>%
  na_if('UNKNOWN') %>% #Remove 'UNKNOWN' values from data frame
  na_if('NONE') %>% #Remove 'NONE' values from data frame
  drop_na() %>% #Remove NA values from data frame
  filter(!PERP_AGE_GROUP %in% c(224, 940, 1020), #Remove errant values in PERP_AGE_GROUP
         !PERP_SEX %in% 'U') %>% #Remove errant values in PERP_SEX column
  mutate(OCCUR_DATE = as_date(OCCUR_DATE, format = "%m/%d/%Y"), #Format date
         across(column_names, as_factor)) %>% #Change these cols to factors
  select(-Lon_Lat) #Remove Lon_Lat column
```

```
#Determine shape and data types of the formatted "shootings" data frame
glimpse(shootings)
```

```
## Rows: 5,340
## Columns: 18
## $ INCIDENT_KEY            <dbl> 16814011, 144732382, 68964080, 85875439, 17577~
## $ OCCUR_DATE              <date> 2006-06-13, 2015-07-23, 2009-12-17, 2012-07-2~
## $ OCCUR_TIME              <time> 20:56:00, 00:22:00, 13:45:00, 21:35:00, 03:40~
## $ BORO                    <fct> BROOKLYN, BROOKLYN, MANHATTAN, BRONX, QUEENS, ~
## $ PRECINCT                <fct> 70, 90, 20, 42, 102, 73, 49, 73, 50, 40, 48, 4~
## $ JURISDICTION_CODE       <fct> 0, 2, 0, 2, 0, 0, 0, 0, 0, 2, 0, 2, 0, 2, 2, 2~
## $ LOCATION_DESC           <fct> MULTI DWELL - APT BUILD, MULTI DWELL - PUBLIC ~
## $ STATISTICAL_MURDER_FLAG <fct> TRUE, FALSE, TRUE, FALSE, FALSE, FALSE, TRUE, ~
## $ PERP_AGE_GROUP          <fct> 25-44, 25-44, 25-44, <18, 25-44, <18, 18-24, 1~
## $ PERP_SEX                <fct> M, M, M, M, M, M, M, M, M, M, M, M, F, M, M, M~
## $ PERP_RACE               <fct> BLACK, WHITE HISPANIC, WHITE HISPANIC, BLACK, ~
## $ VIC_AGE_GROUP           <fct> 25-44, 25-44, 45-64, 25-44, 25-44, <18, 18-24,~
## $ VIC_SEX                 <fct> M, M, F, M, F, F, M, M, M, M, M, M, M, M, M, M~
## $ VIC_RACE                <fct> BLACK, BLACK, WHITE HISPANIC, BLACK, AMERICAN ~
## $ X_COORD_CD              <dbl> 995793.0, 1001019.9, 990838.6, 1011046.7, 1031~
```

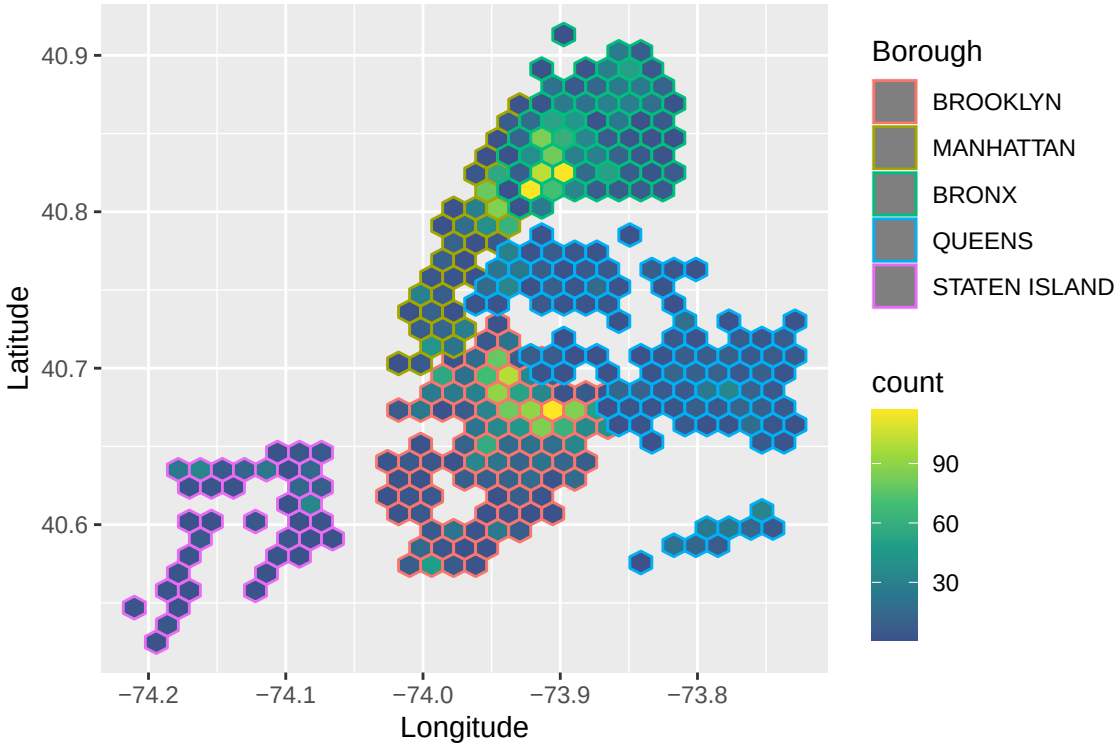
```
## $ Y_COORD_CD              <dbl> 173152.0, 196399.9, 225657.7, 239814.2, 190474~
## $ Latitude                <dbl> 40.64194, 40.70574, 40.78606, 40.82488, 40.689~
## $ Longitude               <dbl> -73.95841, -73.93952, -73.97621, -73.90318, -7~
```

EXPLORATORY DATA ANALYSIS

DENSITY OF SHOOTINGS:

The next step is to determine which areas of New York city have had the most shootings over the 15 year period. A map indicating the density of shootings is prepared below using the location that has been reported for each shooting. To further the analysis the area's with higher densities of shootings will be isolated and the demographics of the shootings in these area's evaluated. The map indicates that the boroughs of the Bronx, Brooklyn, and Manhattan have had the majority of the shootings in New York City during this time frame.

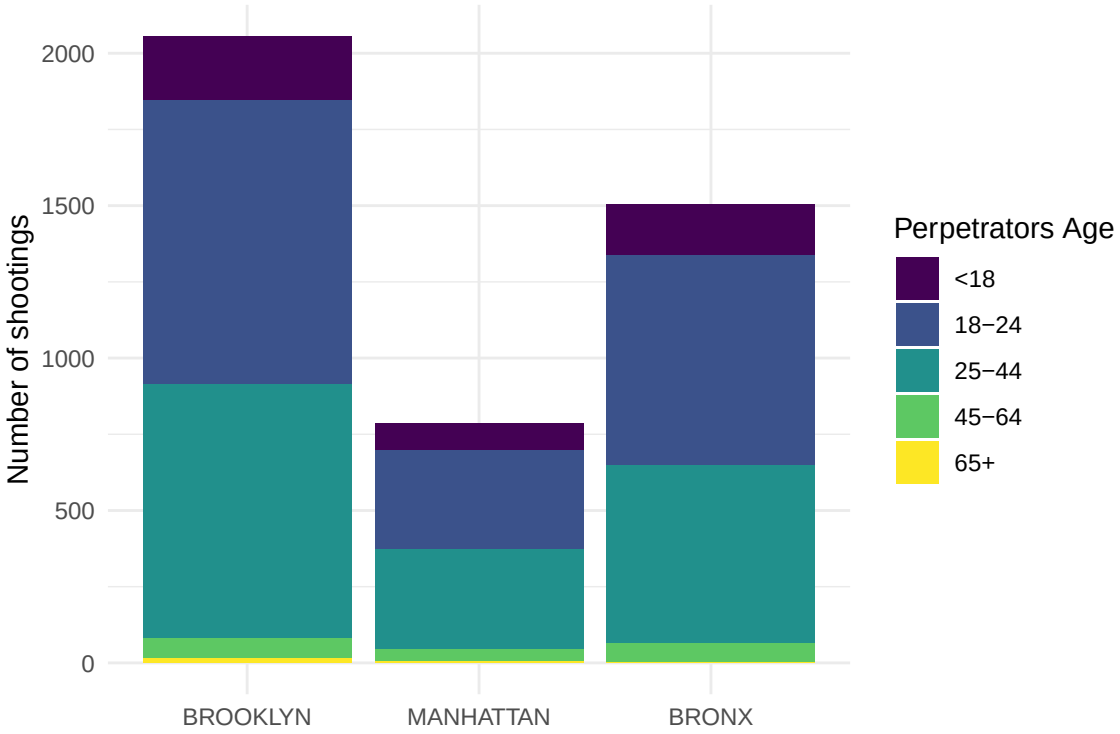
```
shootings %>%
  ggplot() +
  geom_hex(aes(x=Longitude, y=Latitude, color=BORO)) +
  scale_fill_viridis_c(begin = 0.25, end=1) +
  guides(color=guide_legend("Borough"))
```



SHOOTINGS BY AGE GROUPS:

A possible bias would be to assume that only young people (aged under 18) commit crimes. The plot below indicates that the majority of perpetrators are actually below the ages 45.

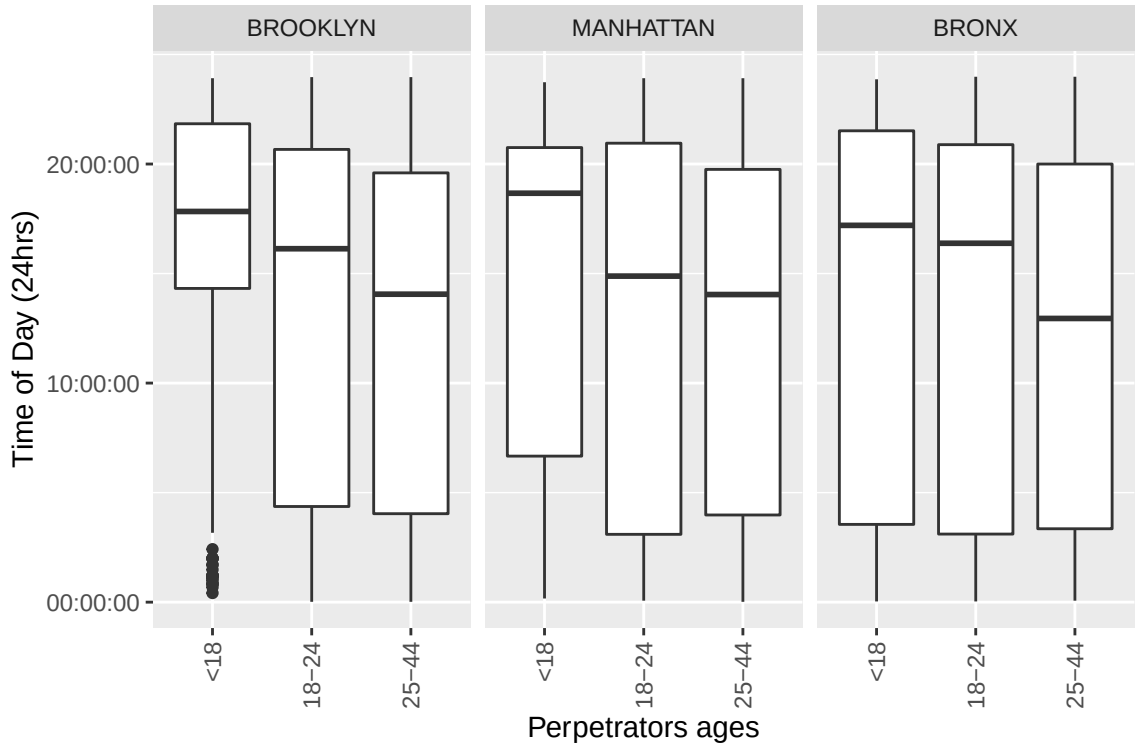
```
shootings %>%
  filter(BORO %in% c('BRONX','MANHATTAN','BROOKLYN')) %>%
  mutate(PERP_AGE_GROUP=factor(PERP_AGE_GROUP,
                               levels=c('<18','18-24','25-44','45-64','65+'))) %>%
  ggplot(aes(x = BORO, fill = PERP_AGE_GROUP)) +
  geom_bar() +
  scale_fill_viridis_d(option = "viridis", direction = 1) +
  labs(x=element_blank(), y="Number of shootings") +
  guides(fill=guide_legend("Perpetrators Age")) +
  theme_minimal()
```



DISTRIBUTION OF WHEN SHOOTINGS OCCUR:

It would be interesting to understand the distribution of when shootings occurred in the Bronx, Brooklyn, and Manhattan for those perpetrators under the age of 45. The plot below reveals an interesting trend such that the older a perpetrator is, the more likely they are to commit a shooting earlier in the day.

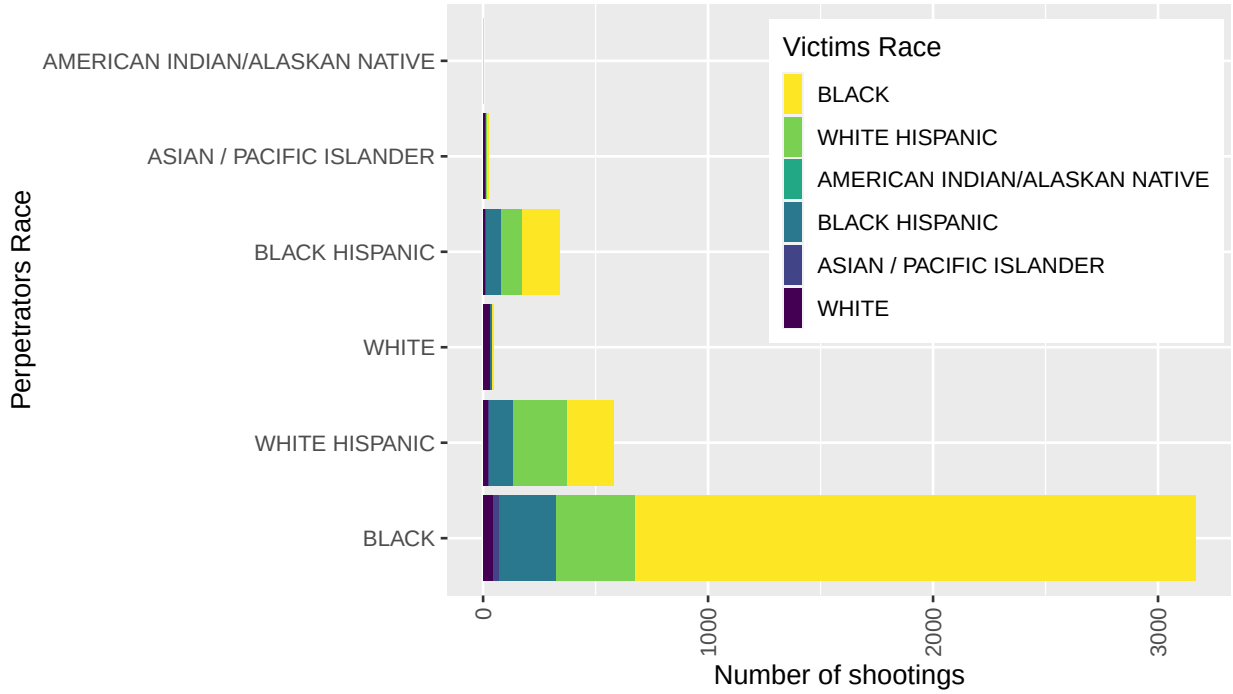
```
shootings %>%
  filter(BORO %in% c('BRONX','MANHATTAN','BROOKLYN'),
         PERP_AGE_GROUP %in% c('<18','18-24','25-44')) %>%
  ggplot() +
  geom_boxplot(aes(x=factor(PERP_AGE_GROUP,
                            levels=c('<18','18-24','25-44')),
                            y=OCCUR_TIME)) +
  labs(x = 'Perpetrators ages', y = 'Time of Day (24hrs)') +
  facet_grid(cols=vars(BORO)) +
  guides(x = guide_axis(angle = 90))
```



DYNAMICS OF RACE IN SHOOTINGS:

Another bias would be to only evaluate the data based on one particular race. The data below compares the race of perpetrators under the age of 45 to the race of their victims in the Bronx, Brooklyn, and Manhattan. Although the majority of shootings occur between Black people, the data shows that other races can also commit shootings.

```
shootings %>%
  filter(BORO %in% c('BRONX','MANHATTAN','BROOKLYN'),
         PERP_AGE_GROUP %in% c('<18','18-24','25-44')) %>%
  ggplot() +
  aes(y = PERP_RACE, fill = VIC_RACE) +
  geom_bar() +
  scale_fill_viridis_d(option = "viridis", direction = -1) +
  guides(x = guide_axis(angle = 90), fill=guide_legend("Victims Race")) +
  labs(x="Number of shootings", y="Perpetrators Race") +
  theme(legend.position=c(0.7,0.7), legend.key.width = unit(0.3,"cm"))
```



CONCLUSION

In summary it has been shown that the majority of shootings have occured between Black people who are under the age or 45 in the boroughs of the Bronx, Brooklyn, and Manhattan. Further analysis is required to understand the dynamics between perpetrators and their victims. In particular it would be important to understand why there is such a high rate of Black on Black shootings and how the location and time of day are factors in these crimes.

R SESSION INFORMATION:

```
##  package   * version date (UTC) lib source
##  dplyr     * 1.0.8   2022-02-08 [1] CRAN (R 4.1.2)
##  forcats   * 0.5.1   2021-01-27 [1] CRAN (R 4.1.1)
##  ggplot2   * 3.3.5   2021-06-25 [1] CRAN (R 4.1.0)
##  knitr     * 1.37    2021-12-16 [1] CRAN (R 4.1.2)
##  lubridate * 1.8.0   2021-10-07 [1] CRAN (R 4.1.2)
##  purrr     * 0.3.4   2020-04-17 [1] CRAN (R 4.1.1)
##  readr     * 2.1.2   2022-01-30 [1] CRAN (R 4.1.2)
##  stringr   * 1.4.0   2019-02-10 [1] CRAN (R 4.1.1)
##  tibble    * 3.1.6   2021-11-07 [1] CRAN (R 4.1.2)
##  tidyr     * 1.2.0   2022-02-01 [1] CRAN (R 4.1.2)
##  tidyverse * 1.3.1   2021-04-15 [1] CRAN (R 4.1.1)
##
##  [1] C:/Users/ryand/Documents/R/win-library/4.1
##  [2] C:/Program Files/R/R-4.1.1/library
```