NYPD Shooting Incident Data Report

Lillian Brown

2023-02-22

Contents

Introduction

This report is written as part of the class, 'Data Science as a Field' from the University of Colorado Boulder Masters in Data Science program.

The data analysed in this report (https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD) is the manually extracted and NYC Office of Management Analysis and Planning reviewed breakdown of every shooting incident in NYC from 2006 through the end of the previous calendar year.

Please see NYPD Shooting Incident Data (Historic) for more information on this data set. (https://data.cityofnewyork.us/Public-Safety/NYPD-Shooting-Incident-Data-Historic-/833y-fsy8)

Import Data

The following code is used to import the NYPD shooting data from the City of New York. The necessary libraries for this report are also added in this section.

```
# Libraries
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.5
## v tibble  3.1.6      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.0
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
##
## The following objects are masked from 'package:base':
```

```
##
##     date, intersect, setdiff, union
```

```
library(ggplot2)
library(dplyr)
library(ggstream)
library(hrbrthemes)
```

```
## NOTE: Either Arial Narrow or Roboto Condensed fonts are required to use these themes.
##       Please use hrbrthemes::import_roboto_condensed() to install Roboto Condensed and
##       if Arial Narrow is not on your system, please see https://bit.ly/arialnarrow
```

```
library(wesanderson)
library(modelr)

# URL of data
url.in <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"

# Read in data
nypd.shooting.data <- read_csv(url.in)
```

```
## Rows: 25596 Columns: 19
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (10): OCCUR_DATE, BORO, LOCATION_DESC, PERP_AGE_GROUP, PERP_SEX, PERP_R...
## 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.
```

Tidy and Transform Data

From an initial review of the data columns, the following columns can be removed:

INCIDENT_KEY: This is a randomly generated persistent ID for each arrest, and is therefore not beneficial for analysis.

X_COORD_CD, Y_COORD_CD, Latitude, Longitude, Lon_Lat: Without using additional data or mapping systems, the geographical information provided by these columns is not useful for analysis. BORO, PRECINCT, and JURISTICTION_CODE provide geographical relevant information on each incident without these columns.

The following code is used to perform these steps:

```
# Remove columns
nypd.shooting.data <- nypd.shooting.data %>%
  select(-c(INCIDENT_KEY,
            X_COORD_CD,
            Y_COORD_CD,
            Latitude,
            Longitude,
            Lon_Lat,
           ))

# Return columns with the corresponding number of missing values
lapply(nypd.shooting.data, function(x) sum(is.na(x)))
```

```
## $OCCUR_DATE
## [1] 0
##
## $OCCUR_TIME
## [1] 0
##
## $BORO
## [1] 0
##
## $PRECINCT
## [1] 0
##
## $JURISDICTION_CODE
## [1] 2
##
## $LOCATION_DESC
## [1] 14977
##
## $STATISTICAL_MURDER_FLAG
## [1] 0
##
## $PERP_AGE_GROUP
## [1] 9344
##
## $PERP_SEX
## [1] 9310
##
## $PERP_RACE
## [1] 9310
##
## $VIC_AGE_GROUP
## [1] 0
##
## $VIC_SEX
## [1] 0
##
## $VIC_RACE
## [1] 0
```

Reviewing the missing values, it is apparent the LOCATION_DESC column is also a large proportion of values. According to the NYPD Shooting Incident Level Data Footnotes, "Null values appearing frequently in certain fields may be attributed to changes on official department forms where data was previously not collected. Null values may also appear in instances where information was not available or unknown at the time of the report and should be considered as either "Unknown/Not Available/Not Reported." Because almost 3/5th of the LOCATION_DESC values are missing and there is other type of location descption data available in its entirety, LOCATION_DESC will not be used in this analysis and the column can be removed.

From the information on missing values, it is also apparent that there is a significant amount of data missing on perpetrators. The perpetrator information is, according to the City of New York, on suspects. Information available on suspects does not necessarily take into account suicidal shootings, prosecution, conviction, or additional investigations. For these reasons, without additional data and analysis, the perpetrator related columns (PERP_AGE_GROUP, PER_SEX, PERP_RACE) will not be used in this analysis and can be removed.

From JURISDICTION_CODE, 2 values are missing. VIC_AGE_GROUP has 60 "UNKNOWN" values, VIC_SEX has 11 "U" values, and VIC_RACE has 65 "UNKNOWN" values. These missing values will be considered in the analysis but given the nature of the missing data type, would not be possible to complete without further information from the NYPD and are nominal enough to move forward without.

The following code is used to make the changes described:

```
# Remove columns
nypd.shooting.data <- nypd.shooting.data %>%
  select(-c(LOCATION_DESC,
            PERP_AGE_GROUP,
            PERP_SEX,
            PERP_RACE
           ))
```

The following data transformations need to be made:

The missing and unknown data should be stored uniformly, the values will all be named "UNKNOWN".

OCCUR_DATE needs to be transformed from character to date type.

BORO, PRECINCT, JURISDICTION_CODE, STATISTICAL_MURDER_FLAG, VIC_AGE_GROUP, PERP_AGE_GROUP, PERP_SEX, and PERP_RACE are all categorical data types and can be transformed to factors.

JUSRIDICTION_CODE codes are 0 (Patrol), 1 (Transit), and 2 (Housing) and can be relabeled as such.

STATISTICAL_MURDER_FLAG contains TRUE, the shooting resulted in the victim being murdered, and FALSE, the shooting did not result in the murder of the victim. These can be renamed to YES (TRUE) and NO (FALSE).

The following columns can be renamed:

DATE (from OCCUR_DATE) TIME (from OCCUR_TIME) JURISDICTION (from JURISDICTION_CODE) MURDER = (from STATISTICAL_MURDER_FLAG) VICTIM.AGE = (from VIC_AGE_GROUP) VICTIM.SEX = (from VIC_SEX) VICTIM.RACE = (from VIC_RACE)

The following code is used for the changes described:

```
# Mutate Data
nypd.shooting.data <- nypd.shooting.data %>%
  mutate(OCCUR_DATE = mdy(OCCUR_DATE))

# Rename Jurisdiction Codes
nypd.shooting.data$JURISDICTION_CODE <- as.character(nypd.shooting.data$JURISDICTION_CODE)
nypd.shooting.data$JURISDICTION_CODE <- recode(nypd.shooting.data$JURISDICTION_CODE, "0" = "PATROL", "1

#Rename Statistical Murder Flags
nypd.shooting.data$STATISTICAL_MURDER_FLAG <- as.character(nypd.shooting.data$STATISTICAL_MURDER_FLAG)
nypd.shooting.data$STATISTICAL_MURDER_FLAG <- recode(nypd.shooting.data$STATISTICAL_MURDER_FLAG, "TRUE"

# Rename Unknowns
nypd.shooting.data[is.na(nypd.shooting.data)] <- "UNKNOWN"
nypd.shooting.data$VIC_SEX <- recode(nypd.shooting.data$VIC_SEX, "U" = "UNKNOWN")

# Change data types to factors
nypd.shooting.data$BORO <- factor(nypd.shooting.data$BORO)
nypd.shooting.data$PRECINCT <- factor(nypd.shooting.data$PRECINCT)
nypd.shooting.data$JURISDICTION_CODE <- factor(nypd.shooting.data$JURISDICTION_CODE)
nypd.shooting.data$STATISTICAL_MURDER_FLAG <- factor(nypd.shooting.data$STATISTICAL_MURDER_FLAG)
```

```
nypd.shooting.data$VIC_AGE_GROUP <- factor(nypd.shooting.data$VIC_AGE_GROUP)
nypd.shooting.data$VIC_SEX <- factor(nypd.shooting.data$VIC_SEX)
nypd.shooting.data$VIC_RACE <- factor(nypd.shooting.data$VIC_RACE)

# Rename Columns
nypd.shooting.data <- nypd.shooting.data %>%
  rename(
    DATE = OCCUR_DATE,
    TIME = OCCUR_TIME,
    BOROUGH = BORO,
    JURISDICTION = JURISDICTION_CODE,
    MURDER = STATISTICAL_MURDER_FLAG,
    VICTIM.AGE = VIC_AGE_GROUP,
    VICTIM.SEX = VIC_SEX,
    VICTIM.RACE = VIC_RACE
    )
```

The summary of the data to be analyzed:

```
summary(nypd.shooting.data)
```

```
##       DATE                TIME                   BOROUGH         PRECINCT
##  Min.   :2006-01-01   Length:25596      BRONX        : 7402   75     : 1470
##  1st Qu.:2009-05-10   Class1:hms        BROOKLYN     :10365   73     : 1372
##  Median :2012-08-26   Class2:difftime   MANHATTAN    : 3265   67     : 1160
##  Mean   :2013-06-13   Mode  :numeric    QUEENS       : 3828   79     :  982
##  3rd Qu.:2017-07-01                     STATEN ISLAND:  736   44     :  949
##  Max.   :2021-12-31                                          47     :  903
##                                                              (Other):18760
##   JURISDICTION   MURDER        VICTIM.AGE      VICTIM.SEX
##  HOUSING: 4214   NO :20668   <18    : 2681   F      : 2403
##  PATROL :21321   YES: 4928   18-24  : 9604   M      :23182
##  TRANSIT:   59               25-44  :11386   UNKNOWN:   11
##  UNKNOWN:    2               45-64  : 1698
##                              65+    :  167
##                              UNKNOWN:   60
##
##                          VICTIM.RACE
##  AMERICAN INDIAN/ALASKAN NATIVE:    9
##  ASIAN / PACIFIC ISLANDER      :  354
##  BLACK                         :18281
##  BLACK HISPANIC                : 2485
##  UNKNOWN                       :   65
##  WHITE                         :  660
##  WHITE HISPANIC                : 3742
```

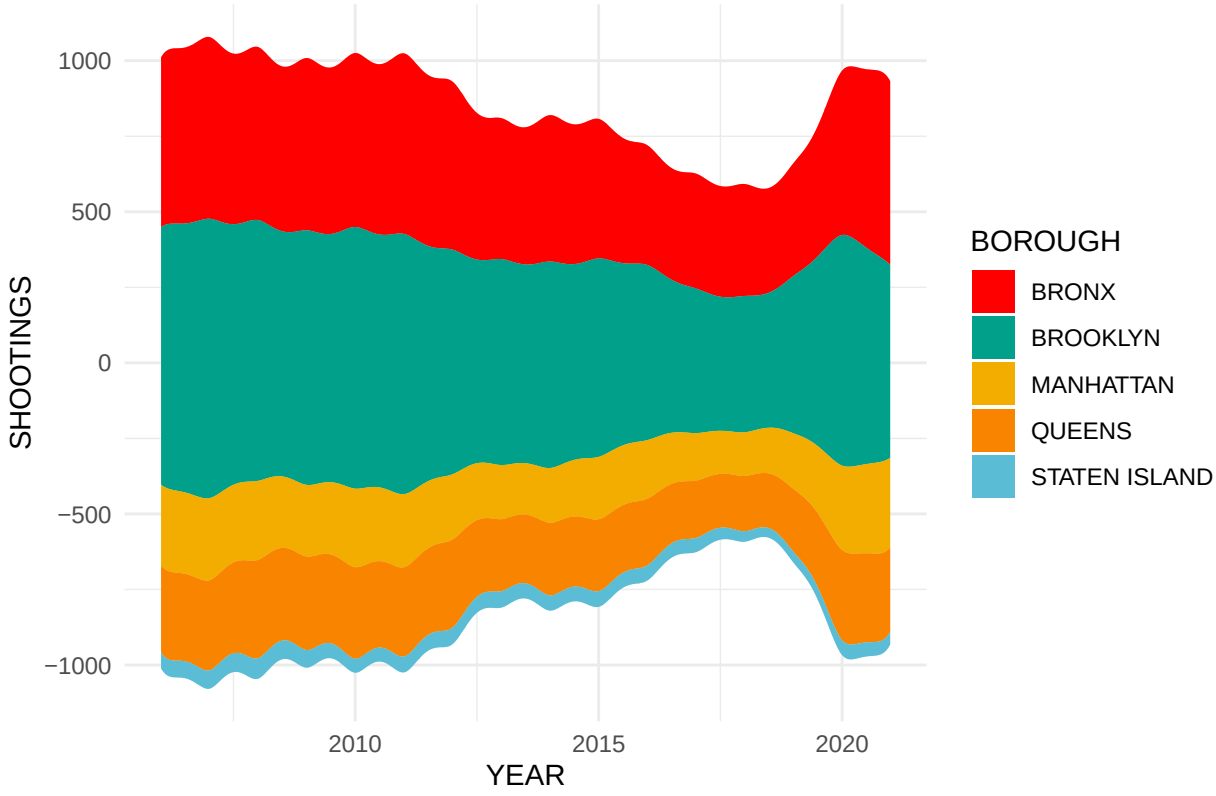
Visualizations and Analysis

From the summary of data to be analyzed, there are several general categories of data available: date/time, geography (borough, precinct, jurisdiction), murder, and victim (age, sex, race).

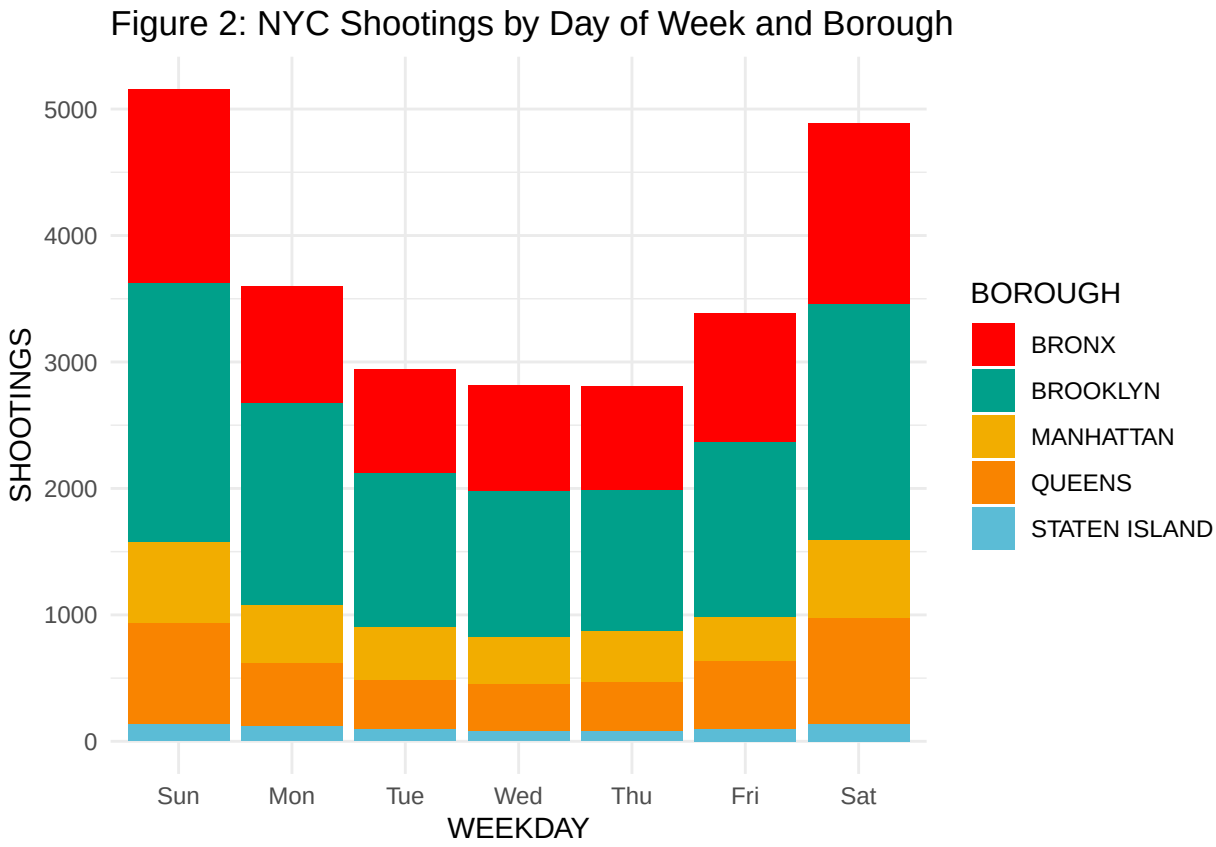
To begin analyzing these categories, the first visuals observe the number of shootings against time factors by borough:

```r
# Group shootings by year
group.byyear <- nypd.shooting.data %>%
  rename(YEAR = DATE)
group.byyear$YEAR <- format(group.byyear$YEAR, format = "%Y")
group.byyear$YEAR <- as.integer(group.byyear$YEAR)

# Tally shootings by year and borough
group.byyear <- group.byyear %>%
  group_by(YEAR) %>%
  group_by(BOROUGH, .add = TRUE) %>%
  tally()

group.byyear <- group.byyear %>%
  rename(
    SHOOTINGS = n)

# Create graph
plot.byyear <- ggplot(group.byyear, aes(x = YEAR, y = SHOOTINGS, fill = BOROUGH)) +
  geom_stream(extra_span = 0.1) +
  geom_stream(extra_span = 0.1, true_range = "both",
              alpha = 0.3) +
  theme_minimal()+
  scale_fill_manual(values = wes_palette(name="Darjeeling1", n=5))+
  ggtitle("Figure 1: NYC Shootings by Year and Borough")

# Plot graph
plot.byyear
```

Figure 1: NYC Shootings by Year and Borough



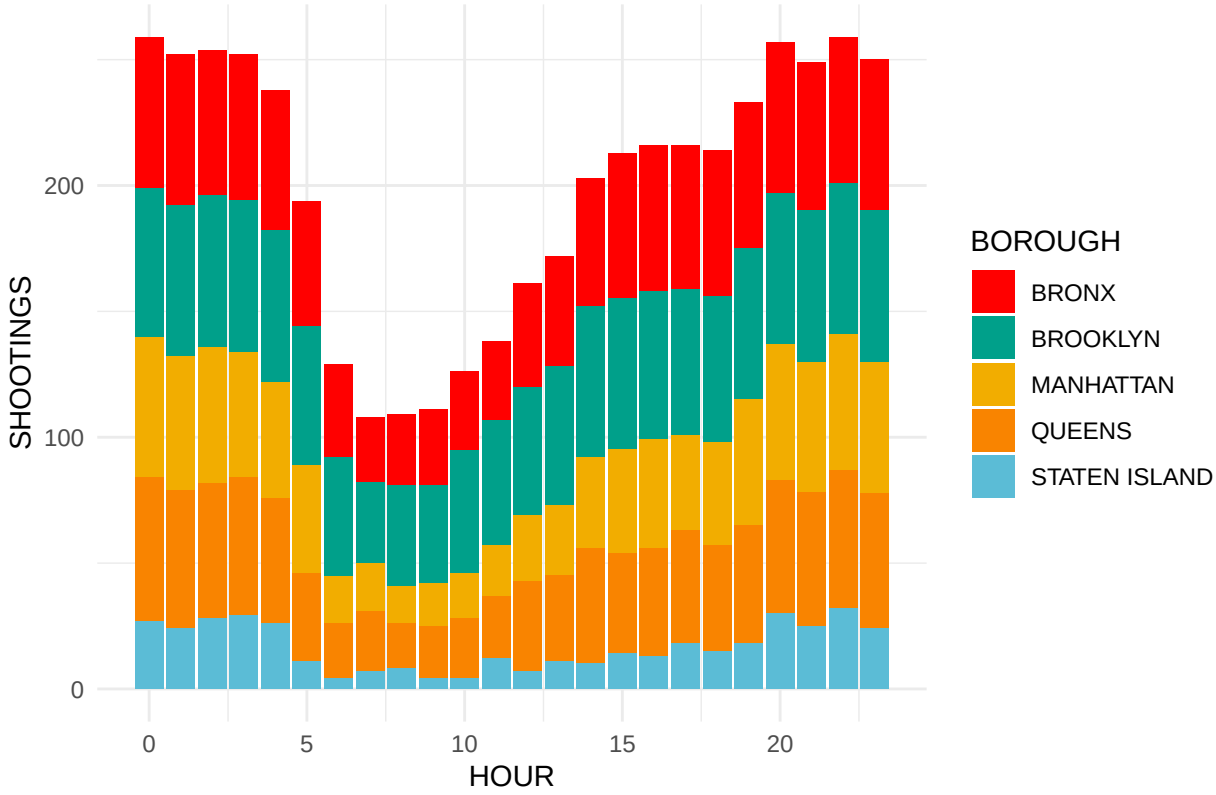
```r
# Group shootings by day of the week and borough
group.byweekday <- nypd.shooting.data %>%
  rename(WEEKDAY = DATE)

group.byweekday$WEEKDAY = wday(group.byweekday$WEEKDAY, label = TRUE)

group.byweekday <- group.byweekday %>%
  group_by(WEEKDAY) %>%
  group_by(BOROUGH, .add = TRUE) %>%
  tally()

group.byweekday <- group.byweekday %>%
  rename(
    SHOOTINGS = n)

# Plot shootings by day of the week and borough
plot.byweekday <- ggplot(group.byweekday, aes(x = WEEKDAY, y = SHOOTINGS, fill = BOROUGH)) +
  geom_bar(position="stack", stat="identity")+
  theme_minimal()+
  scale_fill_manual(values = wes_palette(name="Darjeeling1", n=5))+
  ggtitle("Figure 2: NYC Shootings by Day of Week and Borough")

plot.byweekday
```




```r
# Group shootings by time of day and borough
group.bytime <- nypd.shooting.data %>%
  group_by(TIME) %>%
```

```
  group_by(BOROUGH, .add = TRUE) %>%
  tally()

group.byhour <- group.bytime %>%
  rename(
    HOUR = TIME,
    SHOOTINGS = n)

group.byhour$HOUR <- hour(group.byhour$HOUR)

group.byhour <- group.byhour %>%
  group_by(HOUR) %>%
  group_by(BOROUGH, .add = TRUE) %>%
  tally()

group.byhour <- group.byhour %>%
  rename(
    SHOOTINGS = n)

# Plot shootings by time of day and borough
plot.byhour <- ggplot(group.byhour, aes(x = HOUR, y = SHOOTINGS, fill = BOROUGH)) +
  geom_bar(position="stack", stat="identity")+
  theme_minimal()+
  scale_fill_manual(values = wes_palette(name="Darjeeling1", n=5))+
  ggtitle("Figure 3: NYC Shootings by Time of Day (Hours) and Borough")

plot.byhour
```

Figure 3: NYC Shootings by Time of Day (Hours) and Borough



From Figure 1, it is evident that overall, in the period from 2006 to 2021, Brooklyn had the most shootings, followed by the Bronx. Manhattan and Queens had significantly less shootings while Staten Island had by far the least.

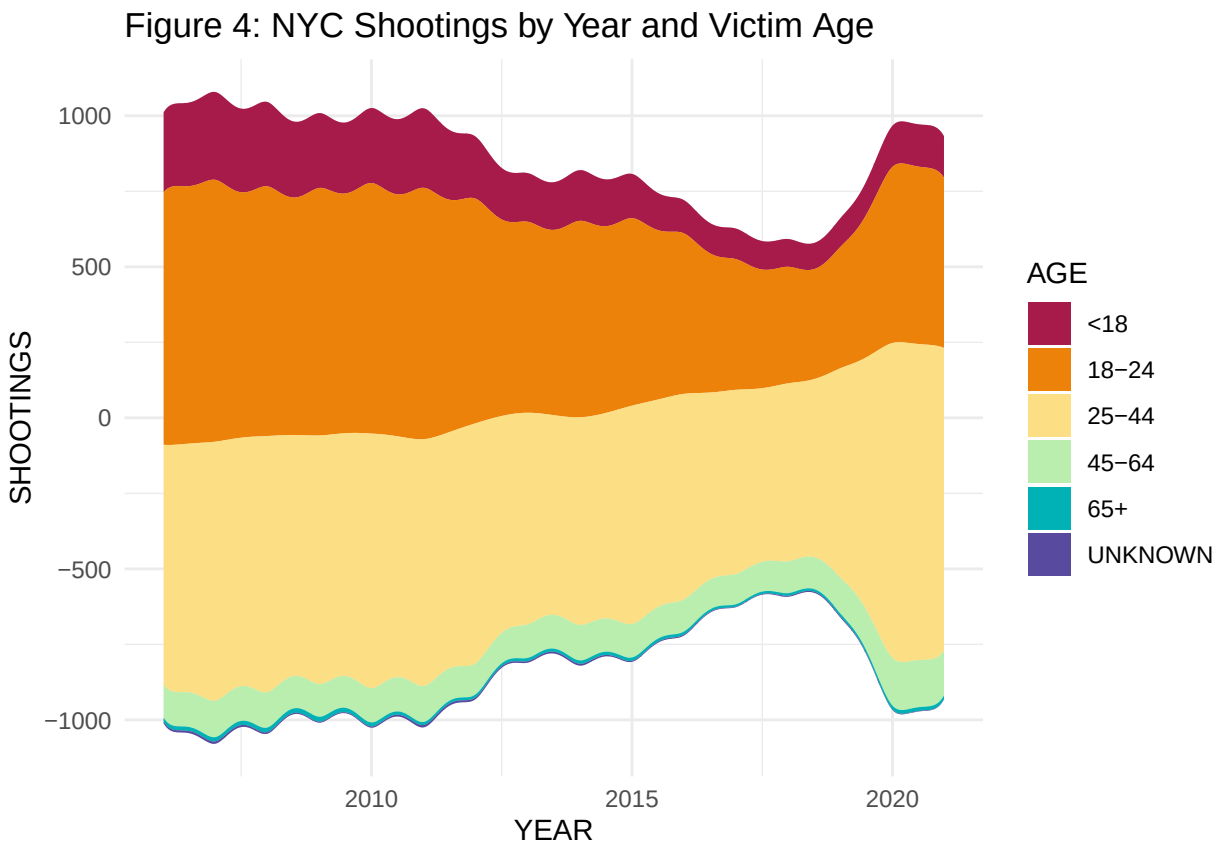
Figure 1 illustrates that from the period beginning in 2006, 2006 had the largest number of shootings. The number of shootings decreased to less than half of the 2006 levels over 2017-2019, then rapidly increased to comparable 2006 levels in 2020 and 2021.

Figures 2 shows that the highest rates of shootings tend to be on weekends while Figure 3 shows that the lowest rates of shootings tend to be in the mornings (approximately 6AM to 11AM).

Recognizing that both borough and aspects of time have greatly varying shooting records, the next visuals depict aspects of victim profiles against borough and time:

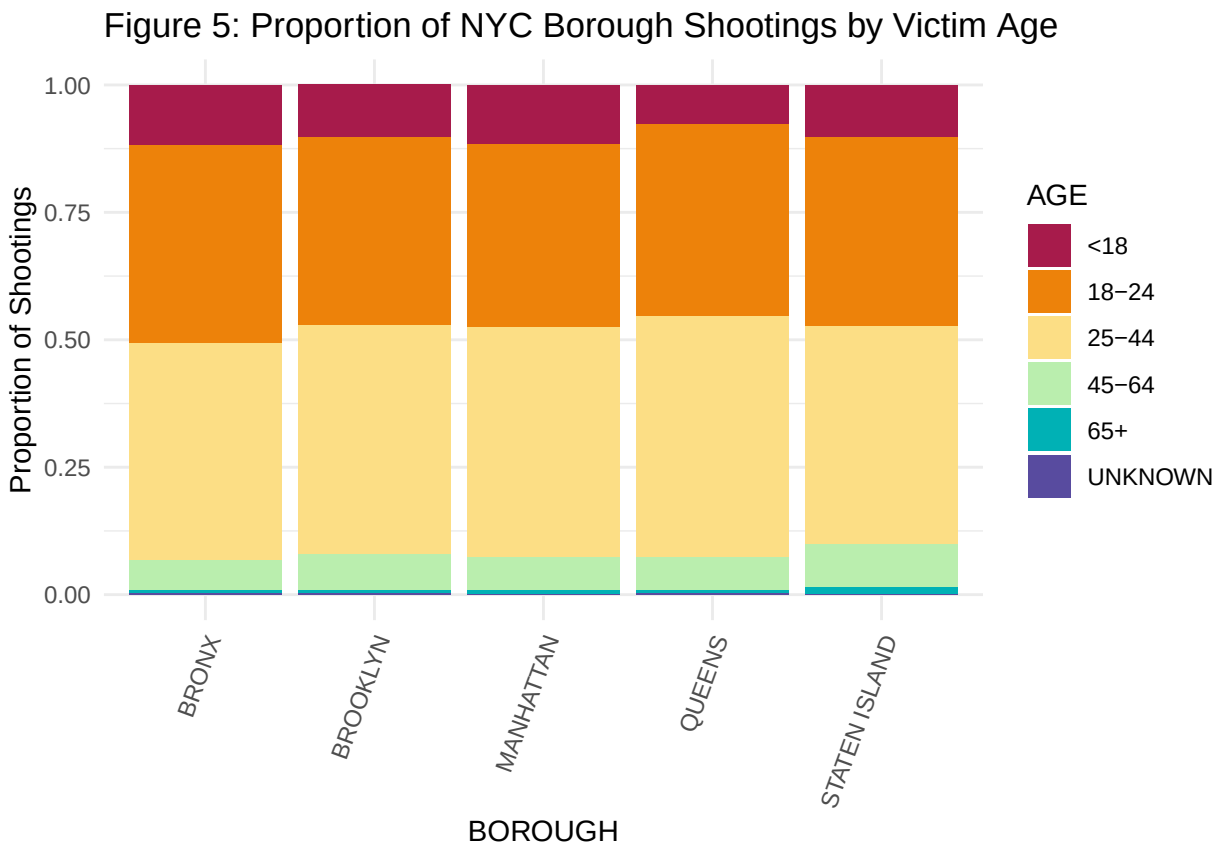
```
# Group shootings by victim age and year
group.byageyear <- nypd.shooting.data %>%
  rename(YEAR = DATE)
group.byageyear$YEAR <- format(group.byageyear$YEAR, format = "%Y")
group.byageyear$YEAR <- as.integer(group.byageyear$YEAR)

group.byageyear <- group.byageyear %>%
  group_by(YEAR) %>%
  group_by(VICTIM.AGE, .add = TRUE) %>%
  tally()

group.byageyear <- group.byageyear %>%
  rename(
    AGE = VICTIM.AGE,
    SHOOTINGS = n)
```

```
# Plot by victim age and year
plot.byageyear <- ggplot(group.byageyear, aes(x = YEAR, y = SHOOTINGS, fill = AGE)) +
  geom_stream(extra_span = 0.1) +
  geom_stream(extra_span = 0.1, true_range = "both",
              alpha = 0.3) +
  theme_minimal()+
  scale_fill_manual(values = hcl.colors(n=6, palette = "Spectral"))+
  ggtitle("Figure 4: NYC Shootings by Year and Victim Age")

plot.byageyear
```



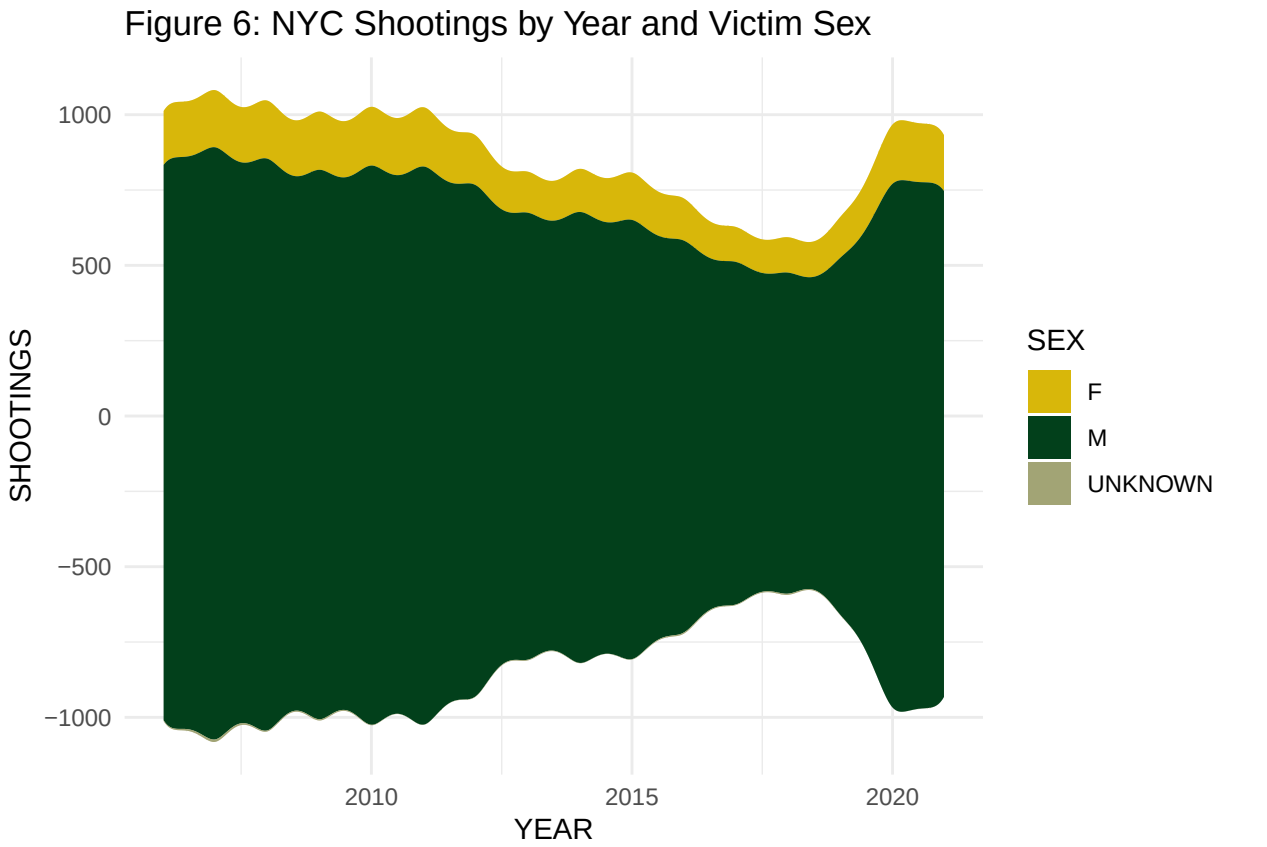

```
# Group shootings by victim age and borough
group.byageborough <- nypd.shooting.data %>%
  group_by(BOROUGH) %>%
  group_by(VICTIM.AGE, .add = TRUE) %>%
  tally()

group.byageborough <- group.byageborough %>%
  rename(
    AGE = VICTIM.AGE,
    SHOOTINGS = n)

# Plot by victim age and borough
plot.byageborough <- ggplot(group.byageborough, aes(x = BOROUGH, y = SHOOTINGS, fill = AGE)) +
  geom_bar(position="fill", stat="identity")+
  theme_minimal()+
  theme(axis.text.x = element_text(angle = 70, hjust=1))+
```

```
  labs(y="Proportion of Shootings")+
  scale_fill_manual(values = hcl.colors(n=6, palette = "Spectral"))+
  ggtitle("Figure 5: Proportion of NYC Borough Shootings by Victim Age")

plot.byageborough
```



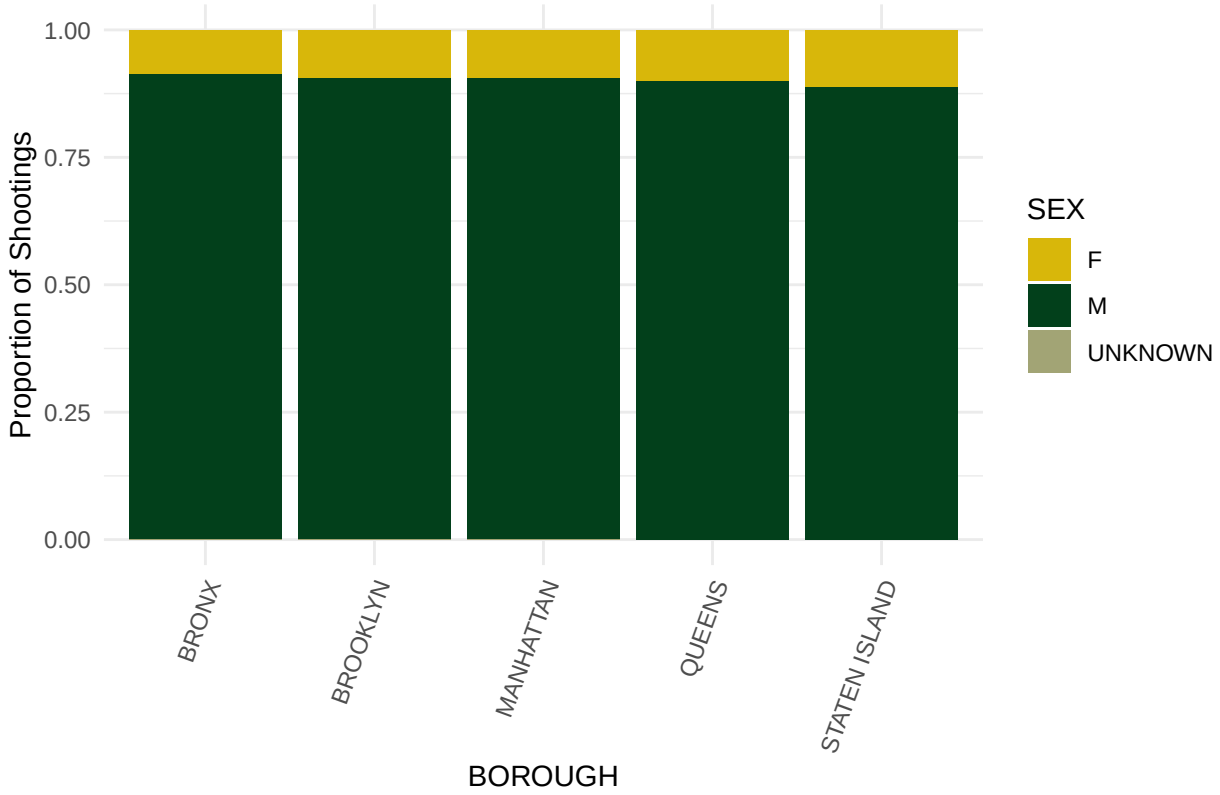

```
# Group shootings by victim sex and year
group.bysexyear <- nypd.shooting.data %>%
  rename(YEAR = DATE)
group.bysexyear$YEAR <- format(group.bysexyear$YEAR, format = "%Y")
group.bysexyear$YEAR <- as.integer(group.bysexyear$YEAR)

group.bysexyear <- group.bysexyear %>%
  group_by(YEAR) %>%
  group_by(VICTIM.SEX, .add = TRUE) %>%
  tally()

group.bysexyear <- group.bysexyear %>%
  rename(
    SEX = VICTIM.SEX,
    SHOOTINGS = n)

# Plot by victim sex and year
plot.bysexyear <- ggplot(group.bysexyear, aes(x = YEAR, y = SHOOTINGS, fill = SEX)) +
  geom_stream(extra_span = 0.1) +
  geom_stream(extra_span = 0.1, true_range = "both",
              alpha = 0.3) +
```

```
  theme_minimal()+
  scale_fill_manual(values = wes_palette(name="Cavalcanti1", n=3))+
  ggtitle("Figure 6: NYC Shootings by Year and Victim Sex")

plot.bysexyear
```



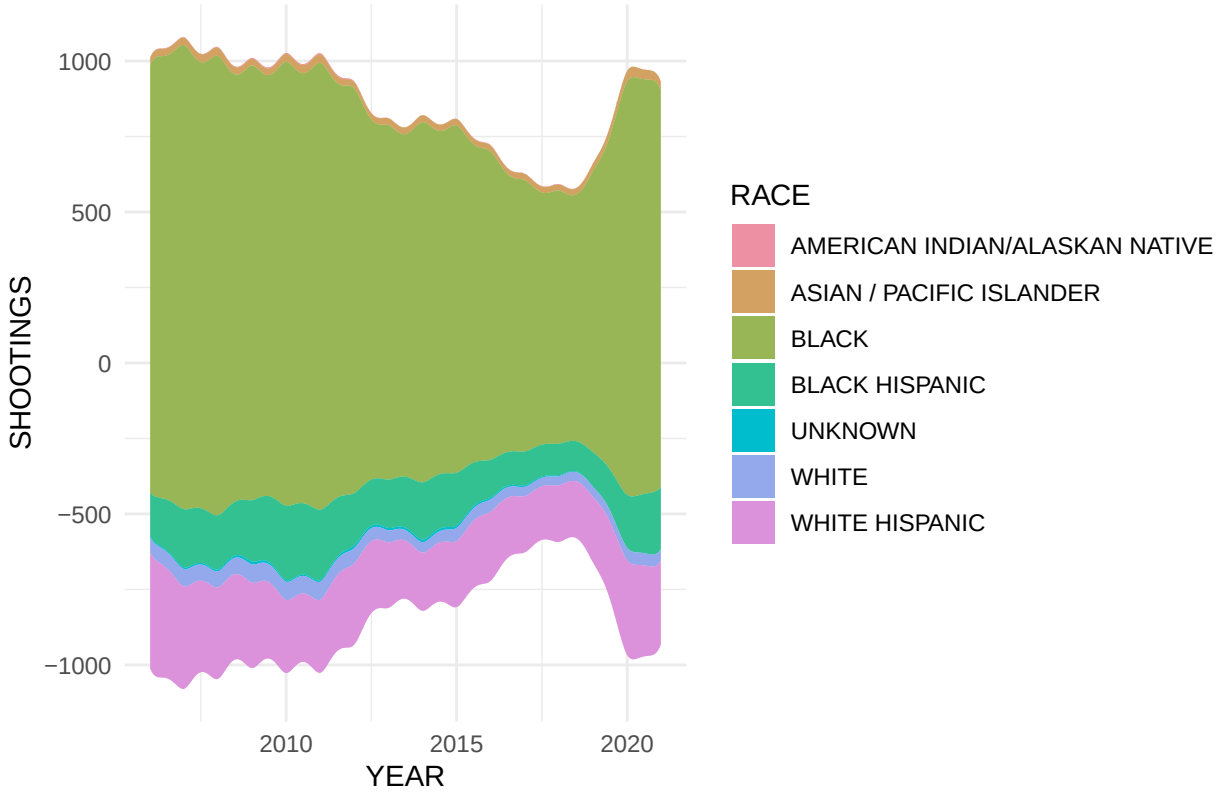

```
# Group shootings by victim sex and borough
group.bysexborough <- nypd.shooting.data %>%
  group_by(VICTIM.SEX) %>%
  group_by(BOROUGH, .add = TRUE) %>%
  tally()

group.bysexborough <- group.bysexborough %>%
  rename(
    SEX = VICTIM.SEX,
    SHOOTINGS = n)

# Plot shootings by victim sex and borough
plot.bysexborough <- ggplot(group.bysexborough, aes(x = BOROUGH, y = SHOOTINGS, fill = SEX)) +
  geom_bar(position="fill", stat="identity")+
  theme_minimal()+
  theme(axis.text.x = element_text(angle = 70, hjust=1))+
  labs(y="Proportion of Shootings")+
  scale_fill_manual(values = wes_palette(name="Cavalcanti1", n=3))+
  ggtitle("Figure 7: Proportion of NYC Borough Shootings by Victim Sex")
plot.bysexborough
```

Figure 7: Proportion of NYC Borough Shootings by Victim Sex



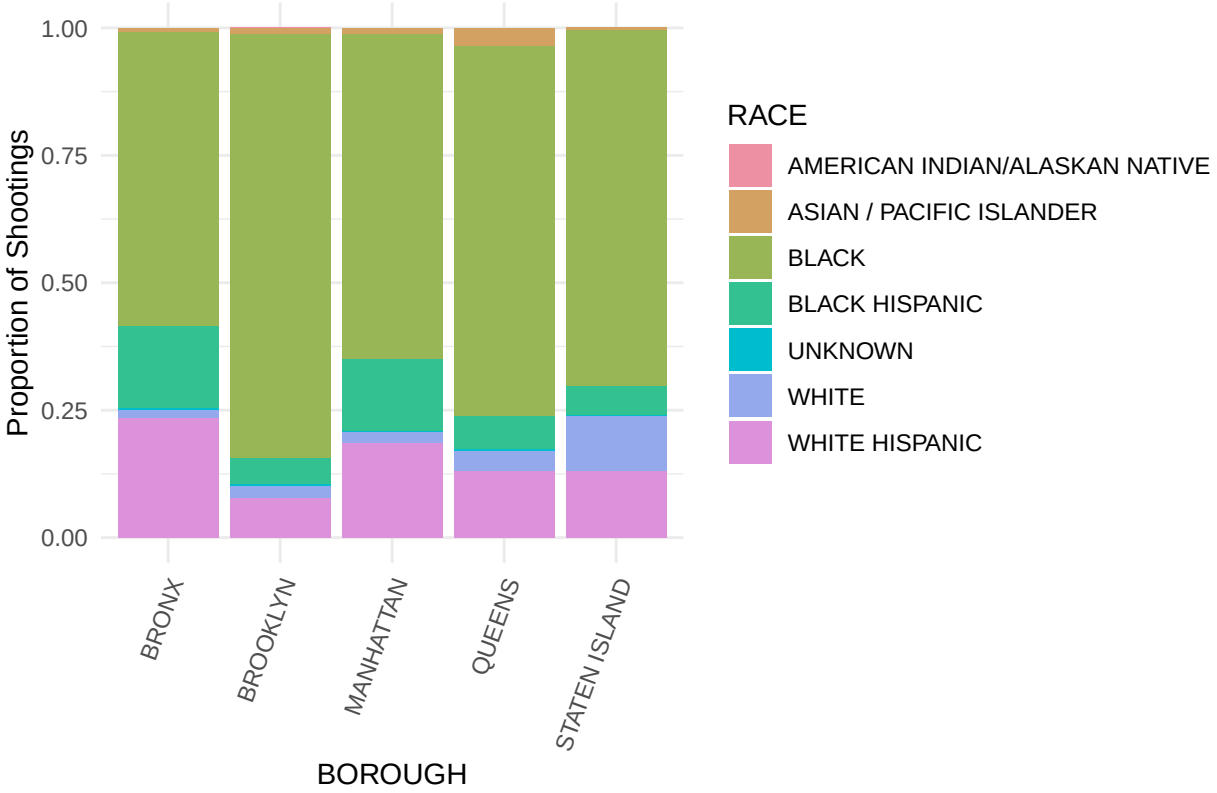

```
# Group shootings by victim race and year
group.byraceyear <- nypd.shooting.data %>%
  rename(YEAR = DATE)
group.byraceyear$YEAR <- format(group.byraceyear$YEAR, format = "%Y")
group.byraceyear$YEAR <- as.integer(group.byraceyear$YEAR)

group.byraceyear <- group.byraceyear %>%
  group_by(YEAR) %>%
  group_by(VICTIM.RACE, .add = TRUE) %>%
  tally()

group.byraceyear <- group.byraceyear %>%
  rename(
    RACE = VICTIM.RACE,
    SHOOTINGS = n)

# Plot by victim race and year
plot.byraceyear <- ggplot(group.byraceyear, aes(x = YEAR, y = SHOOTINGS, fill = RACE)) +
  geom_stream(extra_span = 0.1) +
  geom_stream(extra_span = 0.1, true_range = "both",
              alpha = 0.3) +
  theme_minimal()+
  scale_fill_manual(values = hcl.colors(n=7, palette = "Set2"))+
  ggtitle("Figure 8: NYC Shootings by Year and Victim Race")

plot.byraceyear
```

Figure 8: NYC Shootings by Year and Victim Race




```
# Group by victim race and borough
group.byraceborough <- nypd.shooting.data %>%
  group_by(VICTIM.RACE) %>%
  group_by(BOROUGH, .add = TRUE) %>%
  tally()

group.byraceborough <- group.byraceborough %>%
  rename(
    RACE = VICTIM.RACE,
    SHOOTINGS = n)

# Plot by victim race and borough
plot.byraceborough <- ggplot(group.byraceborough, aes(x = BOROUGH, y = SHOOTINGS, fill = RACE)) +
  geom_bar(position="fill", stat="identity")+
  theme_minimal()+
  theme(axis.text.x = element_text(angle = 70, hjust=1))+
  labs(y="Proportion of Shootings")+
  scale_fill_manual(values = hcl.colors(n=7, palette = "Set2"))+
  ggtitle("Figure 9: Proportion of NYC Borough Shootings by Victim Race")
plot.byraceborough
```

Figure 9: Proportion of NYC Borough Shootings by Victim Race



Figures 4 and 5 show that while the ages of shooting victims do not appear to vary much by borough, the proportion of shooting victims are overwhelmingly 18-24 and 24-45. While from 2006 the two age groups had 849 and 813 shooting victims respectively, the proportion of shooting victims 25-44 increased massively after 2016 while shooting victims 18-24 decreased and in 2021, there were 601 shooting victims 18-24 while there were 1083 25-44. It is important to keep in mind that 18-24 represents seven years of adults, while 24-45 represents twenty years of adults.

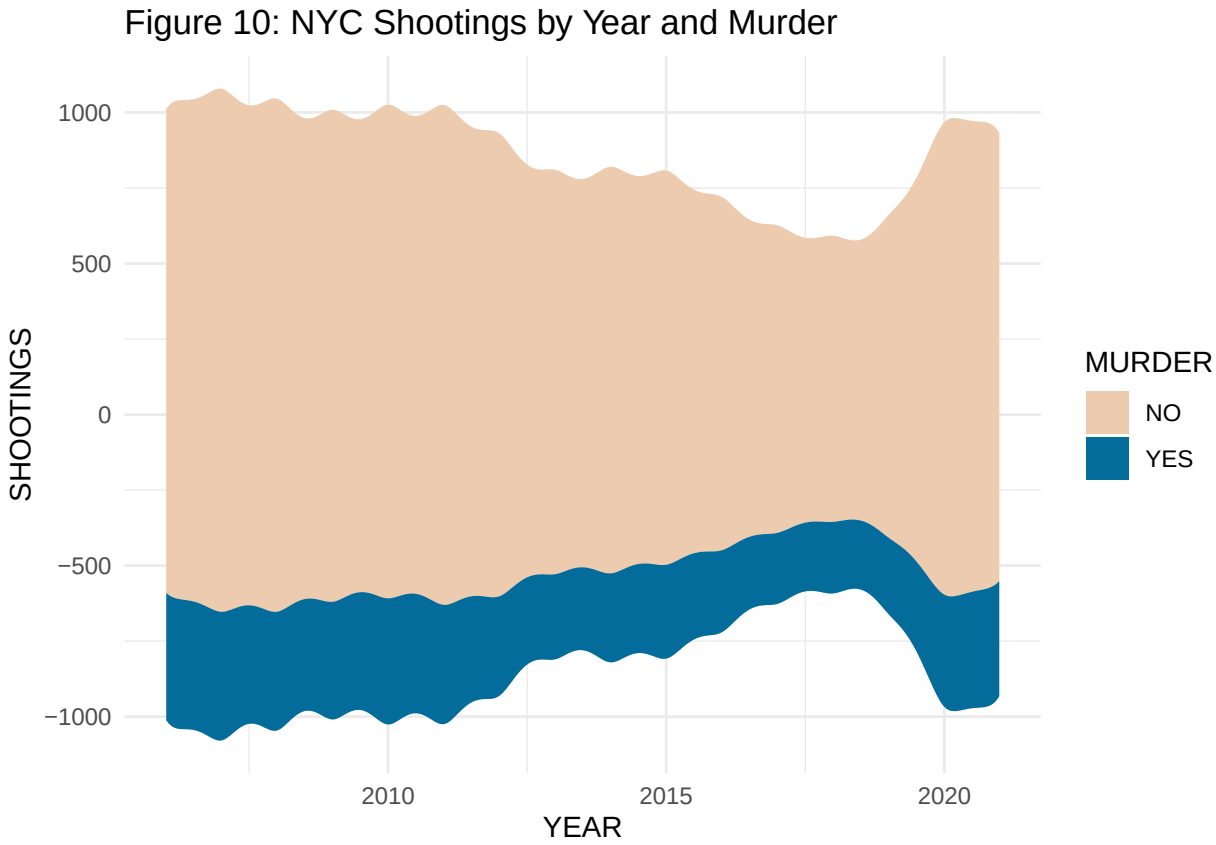
Figures 6 and 7 address victim sex. Despite there being 11 victims for whom their sex was unknown, over 90% of total shooting victims were male. The figures illustrate that in regards to year and borough, the proportion of female to male shooting victims remains largely similar.

Figures 8 and 9 display shooting victim race. The majority, approximately 71.4%, of all shooting victims are Black (not Black Hispanic). There are some notable differences by borough, there are proportionally more White Hispanic and Black Hispanic shooting victims in the Bronx than other boroughs, proportionally more Asian/Pacific Islander shooting victims in Queens than other boroughs, and more White (not White Hispanic) shooting victims on Staten Island than other boroughs, across all boroughs, Black victims make up substantially over half of all shooting victims. It should be noted there were 65 shooting victims with unknown races and American Indian/Alaskan Native victims totaled to 9.

The next visuals depict the statistical murder flag applied to annual shootings and victim information:

```
# Group shootings by victim murder and year
group.bymurderyear <- nypd.shooting.data %>%
  rename(YEAR = DATE)
group.bymurderyear$YEAR <- format(group.bymurderyear$YEAR, format = "%Y")
group.bymurderyear$YEAR <- as.integer(group.bymurderyear$YEAR)

group.bymurderyear <- group.bymurderyear %>%
  group_by(YEAR) %>%
```

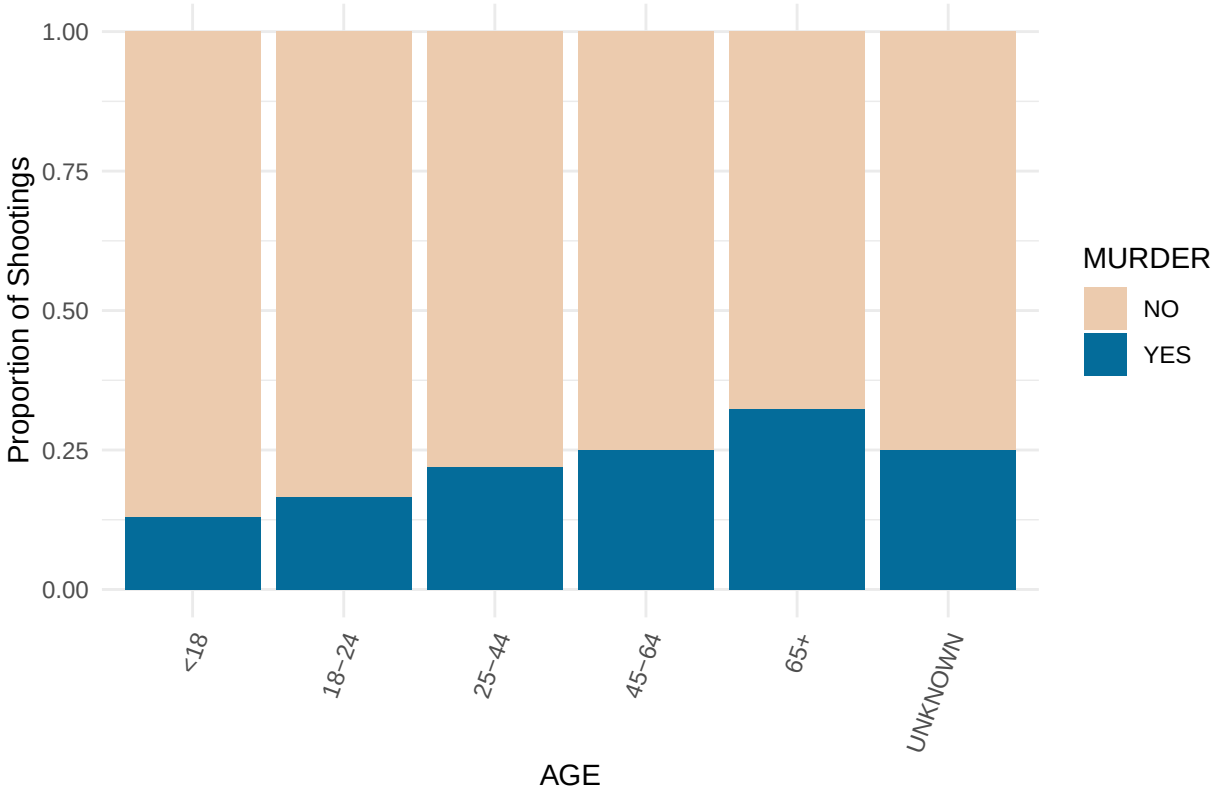
```
  group_by(MURDER, .add = TRUE) %>%
  tally()

group.bymurderyear <- group.bymurderyear %>%
  rename(
    SHOOTINGS = n)

# Plot shootings by victim murder and year
plot.bymurderyear <- ggplot(group.bymurderyear, aes(x = YEAR, y = SHOOTINGS, fill = MURDER)) +
  geom_stream(extra_span = 0.1) +
  geom_stream(extra_span = 0.1, true_range = "both",
              alpha = 0.3) +
  theme_minimal()+
  scale_fill_manual(values = wes_palette(name="Darjeeling2", n=2))+
  ggtitle("Figure 10: NYC Shootings by Year and Murder")

plot.bymurderyear
```




```
# Group shootings by victim murder and victim age
group.bymurderage <- nypd.shooting.data %>%
  group_by(VICTIM.AGE) %>%
  group_by(MURDER, .add = TRUE) %>%
  tally()

group.bymurderage <- group.bymurderage %>%
  rename(
    AGE = VICTIM.AGE,
```

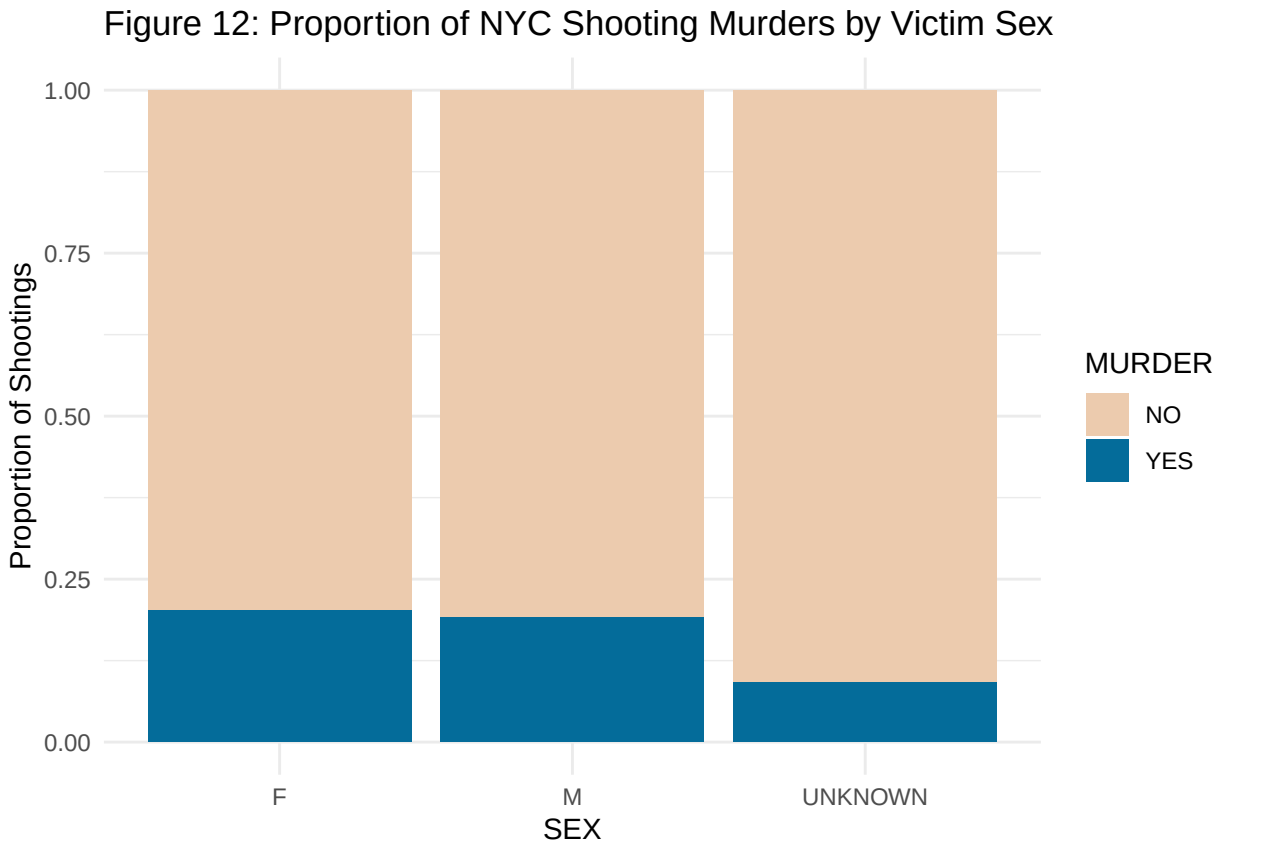
```
    SHOOTINGS = n)

# Plot shootings by victim murder and victim age
plot.bymurderage <- ggplot(group.bymurderage, aes(x = AGE, y = SHOOTINGS, fill = MURDER)) +
  geom_bar(position="fill", stat="identity")+
  theme_minimal()+
  theme(axis.text.x = element_text(angle = 70, hjust=1))+
  labs(y="Proportion of Shootings")+
  scale_fill_manual(values = wes_palette(name="Darjeeling2", n=2))+
  ggtitle("Figure 11: Proportion of NYC Shooting Murders by Victim Age")
plot.bymurderage
```

Figure 11: Proportion of NYC Shooting Murders by Victim Age




```
# Group shootings by victim murder and victim sex
group.bymurdersex <- nypd.shooting.data %>%
  group_by(VICTIM.SEX) %>%
  group_by(MURDER, .add = TRUE) %>%
  tally()

group.bymurdersex <- group.bymurdersex %>%
  rename(
    SEX = VICTIM.SEX,
    SHOOTINGS = n)

# Plot shootings by victim murder and victim sex
plot.bymurdersex <- ggplot(group.bymurdersex, aes(x = SEX, y = SHOOTINGS, fill = MURDER)) +
  geom_bar(position="fill", stat="identity")+
  theme_minimal()+
```

```
  labs(y="Proportion of Shootings")+
  scale_fill_manual(values = wes_palette(name="Darjeeling2", n=2))+
  ggtitle("Figure 12: Proportion of NYC Shooting Murders by Victim Sex")
plot.bymurdersex
```




```
# Group shootings by victim murder and victim race
group.bymurderrace <- nypd.shooting.data %>%
  group_by(VICTIM.RACE) %>%
  group_by(MURDER, .add = TRUE) %>%
  tally()

group.bymurderrace <- group.bymurderrace %>%
  rename(
    RACE = VICTIM.RACE,
    SHOOTINGS = n)

# Plot shootings by victim murder and victim race
plot.bymurderrace <- ggplot(group.bymurderrace, aes(x = RACE, y = SHOOTINGS, fill = MURDER)) +
  geom_bar(position="fill", stat="identity")+
  theme_minimal()+
  theme(axis.text.x = element_text(angle = 70, hjust=1))+
  labs(y="Proportion of Shootings")+
  scale_fill_manual(values = wes_palette(name="Darjeeling2", n=2))+
  ggtitle("Figure 13: Proportion of NYC Shooting Murders by Victim Race")
plot.bymurderrace
```

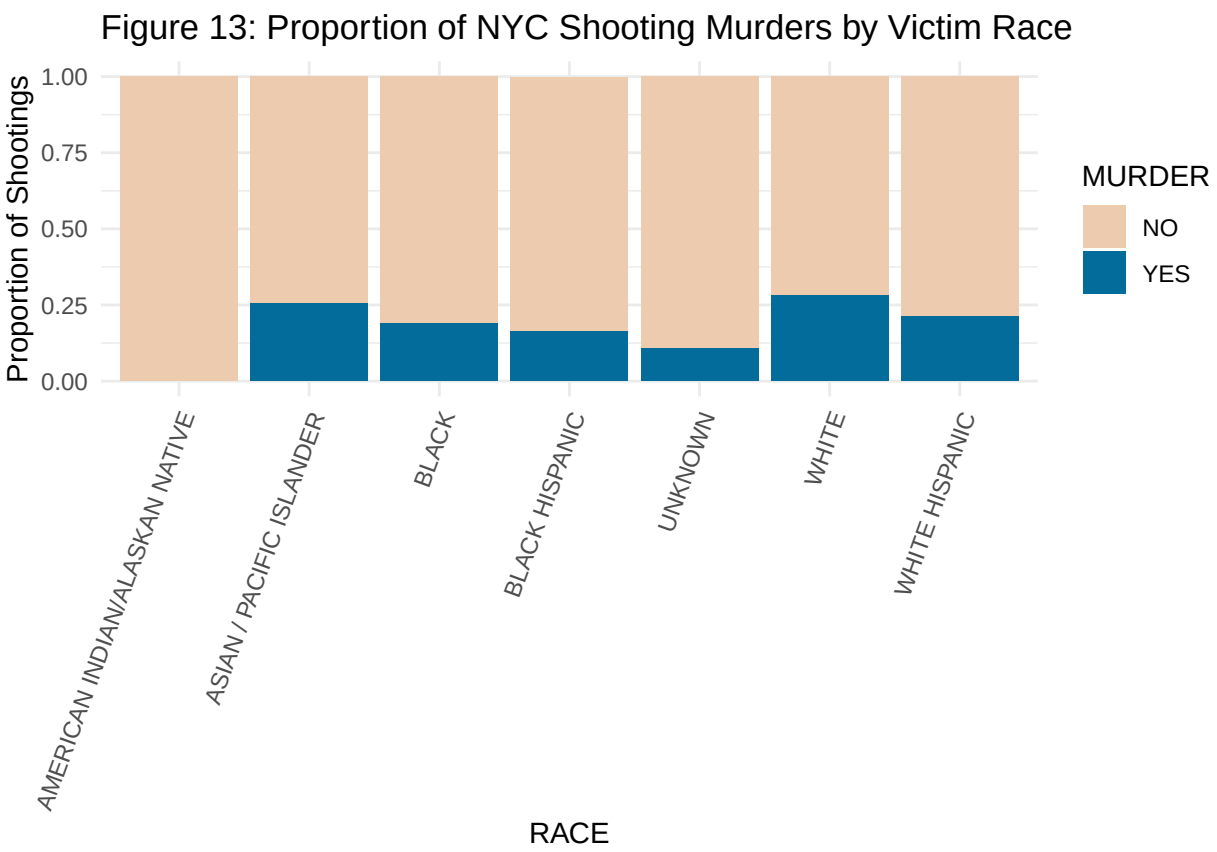


Figure 10 depicts the annual murder rate and shows the proportion of shooting murders as having little change.

Figure 11 shows that the proportion of murder shooting victims increases with age.

Figure 12 shows that proportionality of murder shooting victims, female and male, are within similar range.

Figure 13 shows that none of the 9 American Indian/Alaskan Native shooting victims were murdered, while White followed by Asian/Pacific Islander victims were proportionally most often murdered when shot.

Model

```
nypd.annualshootings <- nypd.shooting.data %>%
  mutate(DATE = format(DATE, "%Y")) %>%
  group_by(DATE) %>%
  group_by(BOROUGH, .add = TRUE) %>%
  tally()

nypd.annualshootings <- nypd.annualshootings %>%
  rename(YEAR = DATE,
         SHOOTINGS = n)

nypd.annualshootings$YEAR <- factor(nypd.annualshootings$YEAR)

nypd.annualmurders <- nypd.shooting.data %>%
  mutate(DATE = format(DATE, "%Y"))

nypd.annualmurders$MURDER = as.character(nypd.annualmurders$MURDER)
```

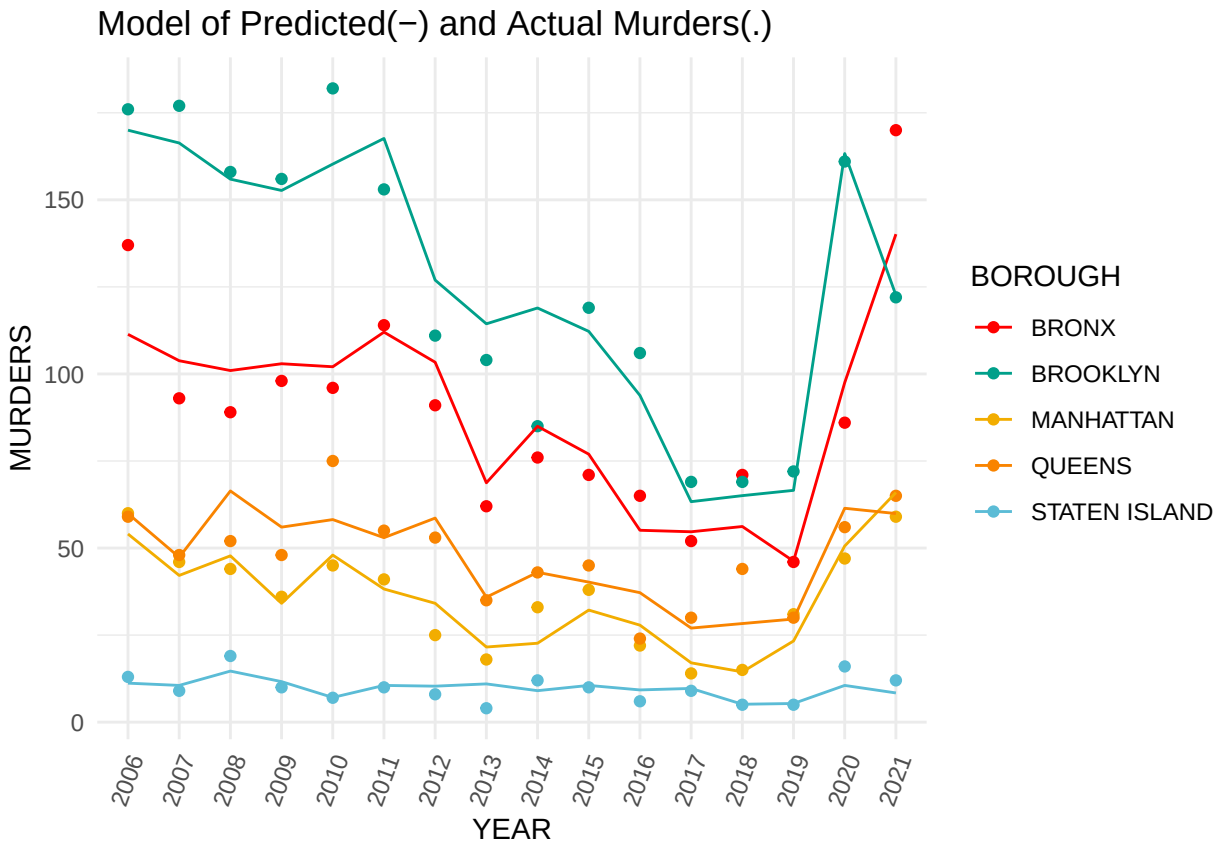
```
nypd.annualmurders$DATE = factor(nypd.annualmurders$DATE)

nypd.annualmurders <- nypd.annualmurders %>%
  filter(str_detect(MURDER, "YES")) %>%
  group_by(DATE, .add = TRUE) %>%
  group_by(BOROUGH, .add = TRUE) %>%
  tally()

nypd.annualmurders <- nypd.annualmurders %>%
  rename(MURDERS = n)

annual.shootings.murders <- nypd.annualshootings

annual.shootings.murders$MURDERS <- nypd.annualmurders$MURDERS

lm.shootingsmurders <- lm(MURDERS ~ SHOOTINGS + BOROUGH, data = annual.shootings.murders)

lm.shootingsmurders
```

```
##
## Call:
## lm(formula = MURDERS ~ SHOOTINGS + BOROUGH, data = annual.shootings.murders)
##
## Coefficients:
##          (Intercept)             SHOOTINGS       BOROUGHBROOKLYN
##              -11.5400                0.2164               -2.3832
##     BOROUGHMANHATTAN         BOROUGHQUEENS  BOROUGHSTATEN ISLAND
##                3.2601                7.3962               11.2740
```

```
predicted.murders <- data.frame(PREDICTED = predict(lm.shootingsmurders,
annual.shootings.murders), BOROUGH = annual.shootings.murders$BOROUGH, YEAR = annual.shootings.murders$

predicted.vs.actual <- annual.shootings.murders
predicted.vs.actual$PREDICTED <- predicted.murders$PREDICTED

plot.lm <-
  ggplot()+
  geom_point(data = predicted.vs.actual, aes(x=YEAR, y=MURDERS, group=BOROUGH, color=BOROUGH)) +
  scale_color_manual(values=wes_palette(name="Darjeeling1", n=5))+
  theme_minimal()+
  ggtitle("Model of Predicted(-) and Actual Murders(.)")

plot.lm <- plot.lm +
  geom_line(data = predicted.vs.actual, aes(x=YEAR, y=PREDICTED, group=BOROUGH, color=BOROUGH)) +
  scale_color_manual(values=wes_palette(name="Darjeeling1", n=5))+
  theme(axis.text.x = element_text(angle = 70, hjust=1))
```

```
## Scale for colour is already present.
## Adding another scale for colour, which will replace the existing scale.
```

```
plot.lm
```



Further Analysis

While there were some general conclusions that could be drawn in this analysis, further analysis would benefit from additional data. PEW Research Center has published that in 2020, suicides accounted for more than half of US gun deaths. Observing suicides, perpetrators with multiple victims, volume of police force, global and domestic events, and not just suspects, but convictions, prosecutions, and sentencing would help to better understand and analyze this data.

Bias Identification

Personally, I am a supporter of gun control, am a skeptic of police use of force, and a proponent of racial and gender equality.

In this assignment, I have attempted to simply take a cursory glance at the NYPD Shooting Data without analyzing any of my personal skepticism.