NYPD Shooting Analysis

This report tries to analyze and answer questions related to rate of crime from NYPD incident reports since 2005 based on the gender and race of the victims and the perpetrators.

IMPORTING THE LIBRARIES:

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.2     v dplyr   1.0.6
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- 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
```

IMPORTING THE DATA:

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

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   INCIDENT_KEY = col_double(),
##   OCCUR_DATE = col_character(),
##   OCCUR_TIME = col_time(format = ""),
##   BORO = col_character(),
##   PRECINCT = col_double(),
##   JURISDICTION_CODE = col_double(),
##   LOCATION_DESC = col_character(),
##   STATISTICAL_MURDER_FLAG = col_logical(),
##   PERP_AGE_GROUP = col_character(),
##   PERP_SEX = col_character(),
##   PERP_RACE = col_character(),
##   VIC_AGE_GROUP = col_character(),
##   VIC_SEX = col_character(),
##   VIC_RACE = col_character(),
##   X_COORD_CD = col_number(),
##   Y_COORD_CD = col_number(),
```

```
##   Latitude = col_double(),
##   Longitude = col_double(),
##   Lon_Lat = col_character()
## )
```

```
head(shooting_data)
```

```
## # A tibble: 6 x 19
##   INCIDENT_KEY OCCUR_DATE OCCUR_TIME BORO          PRECINCT JURISDICTION_CODE
##          <dbl> <chr>      <time>     <chr>            <dbl>              <dbl>
## 1    201575314 08/23/2019 22:10      QUEENS             103                  0
## 2    205748546 11/27/2019 15:54      BRONX               40                  0
## 3    193118596 02/02/2019 19:40      MANHATTAN           23                  0
## 4    204192600 10/24/2019 00:52      STATEN ISLAND      121                  0
## 5    201483468 08/22/2019 18:03      BRONX               46                  0
## 6    198255460 06/07/2019 17:50      BROOKLYN            73                  0
## # ... with 13 more variables: LOCATION_DESC <chr>,
## #   STATISTICAL_MURDER_FLAG <lgl>, PERP_AGE_GROUP <chr>, PERP_SEX <chr>,
## #   PERP_RACE <chr>, VIC_AGE_GROUP <chr>, VIC_SEX <chr>, VIC_RACE <chr>,
## #   X_COORD_CD <dbl>, Y_COORD_CD <dbl>, Latitude <dbl>, Longitude <dbl>,
## #   Lon_Lat <chr>
```

CLEANING THE DATA:

To clean the data, we first drop the columns that are not required for our analysis. In this case, we are dropping the columns X_COORD_CD, Y_COORD_CD, Latitude, Longitude, Lon_Lat.

```
shooting_data <- shooting_data %>% select(-c(X_COORD_CD,Y_COORD_CD,Latitude,Longitude,Lon_Lat))
summary(shooting_data)
```

```
##   INCIDENT_KEY        OCCUR_DATE         OCCUR_TIME           BORO
##  Min.   :  9953245   Length:23568       Length:23568      Length:23568
##  1st Qu.: 55317014   Class :character   Class1:hms        Class :character
##  Median : 83365370   Mode  :character   Class2:difftime   Mode  :character
##  Mean   :102218616                      Mode  :numeric
##  3rd Qu.:150772442
##  Max.   :222473262
##
##     PRECINCT      JURISDICTION_CODE LOCATION_DESC      STATISTICAL_MURDER_FLAG
##  Min.   :  1.00   Min.   :0.0000    Length:23568       Mode :logical
##  1st Qu.: 44.00   1st Qu.:0.0000    Class :character   FALSE:19080
##  Median : 69.00   Median :0.0000    Mode  :character   TRUE :4488
##  Mean   : 66.21   Mean   :0.3323
##  3rd Qu.: 81.00   3rd Qu.:0.0000
##  Max.   :123.00   Max.   :2.0000
##                   NA's   :2
##  PERP_AGE_GROUP       PERP_SEX          PERP_RACE         VIC_AGE_GROUP
##  Length:23568       Length:23568       Length:23568       Length:23568
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    VIC_SEX            VIC_RACE
##  Length:23568       Length:23568
```

```
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
##
```

Next, we change the datatype of OCCUR_DATE from **chr** to **date**. We accomplish this using the *lubridate* library.

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

We also look for **NULL** values and omit such rows:

```
colSums(is.na(shooting_data))
```

```
##            INCIDENT_KEY             OCCUR_DATE             OCCUR_TIME
##                       0                      0                      0
##                    BORO                PRECINCT       JURISDICTION_CODE
##                       0                      0                      2
##           LOCATION_DESC STATISTICAL_MURDER_FLAG          PERP_AGE_GROUP
##                   13581                      0                   8459
##                PERP_SEX              PERP_RACE           VIC_AGE_GROUP
##                    8425                   8425                      0
##                 VIC_SEX              VIC_RACE
##                       0                      0
```

```
na.omit(shooting_data)
```

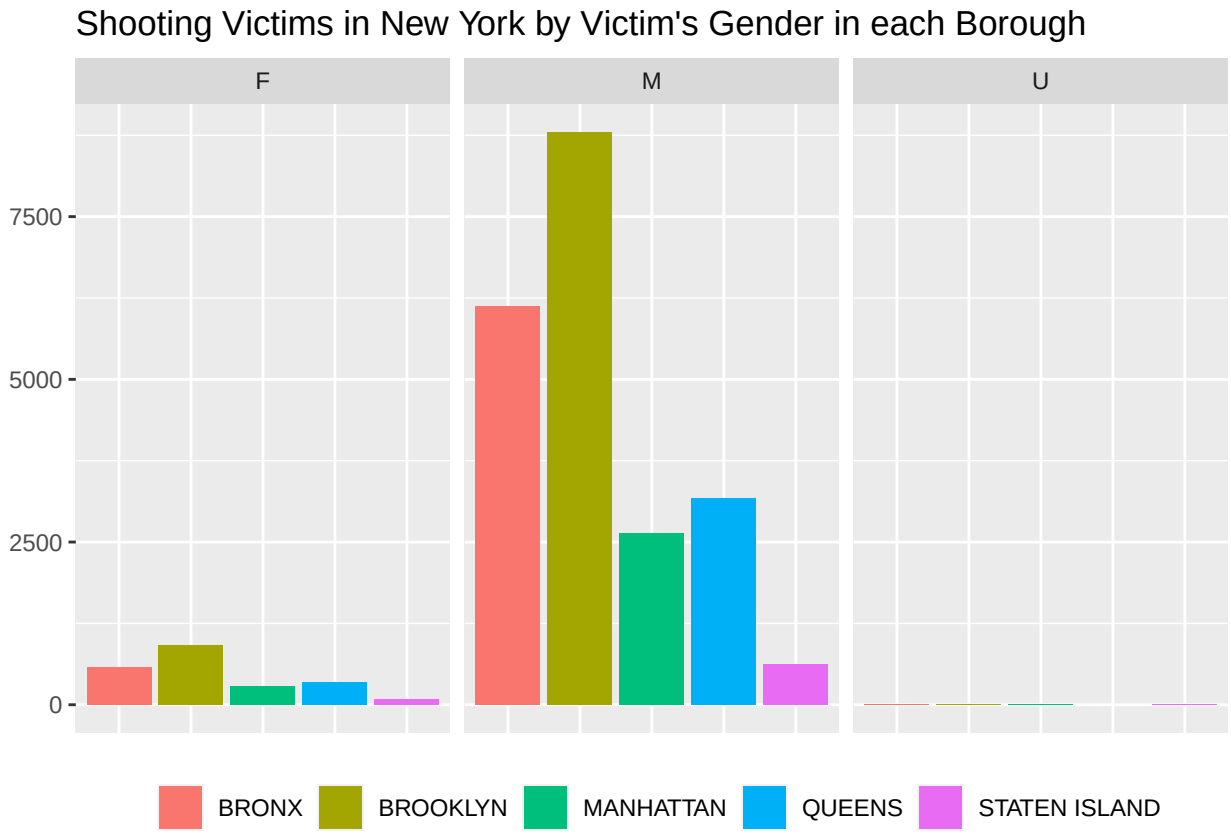
```
## # A tibble: 6,843 x 14
##    INCIDENT_KEY OCCUR_DATE OCCUR_TIME BORO          PRECINCT JURISDICTION_CODE
##           <dbl> <date>     <time>     <chr>            <dbl>              <dbl>
##  1    204192600 2019-10-24 00:52      STATEN ISLAND      121                  0
##  2    193694863 2019-02-17 03:00      QUEENS             114                  2
##  3    201436772 2019-08-21 23:34      STATEN ISLAND      120                  0
##  4    201852654 2019-08-31 07:42      BRONX               45                  0
##  5    193939359 2019-02-24 23:20      BRONX               44                  2
##  6    199247701 2019-07-03 00:04      QUEENS             114                  2
##  7    199134406 2019-06-29 05:48      BROOKLYN            69                  0
##  8    204971625 2019-11-10 14:03      BROOKLYN            63                  0
##  9    200365034 2019-07-28 14:35      MANHATTAN           30                  2
## 10    199422329 2019-07-07 10:50      BROOKLYN            60                  0
## # ... with 6,833 more rows, and 8 more variables: LOCATION_DESC <chr>,
## #   STATISTICAL_MURDER_FLAG <lgl>, PERP_AGE_GROUP <chr>, PERP_SEX <chr>,
## #   PERP_RACE <chr>, VIC_AGE_GROUP <chr>, VIC_SEX <chr>, VIC_RACE <chr>
```

ANALYSIS:

We visualize the victim data by gender in each BORO. By faceting the number of shooting incidents by victim race, we can see a clear breakdown of victims by race in each BORO.

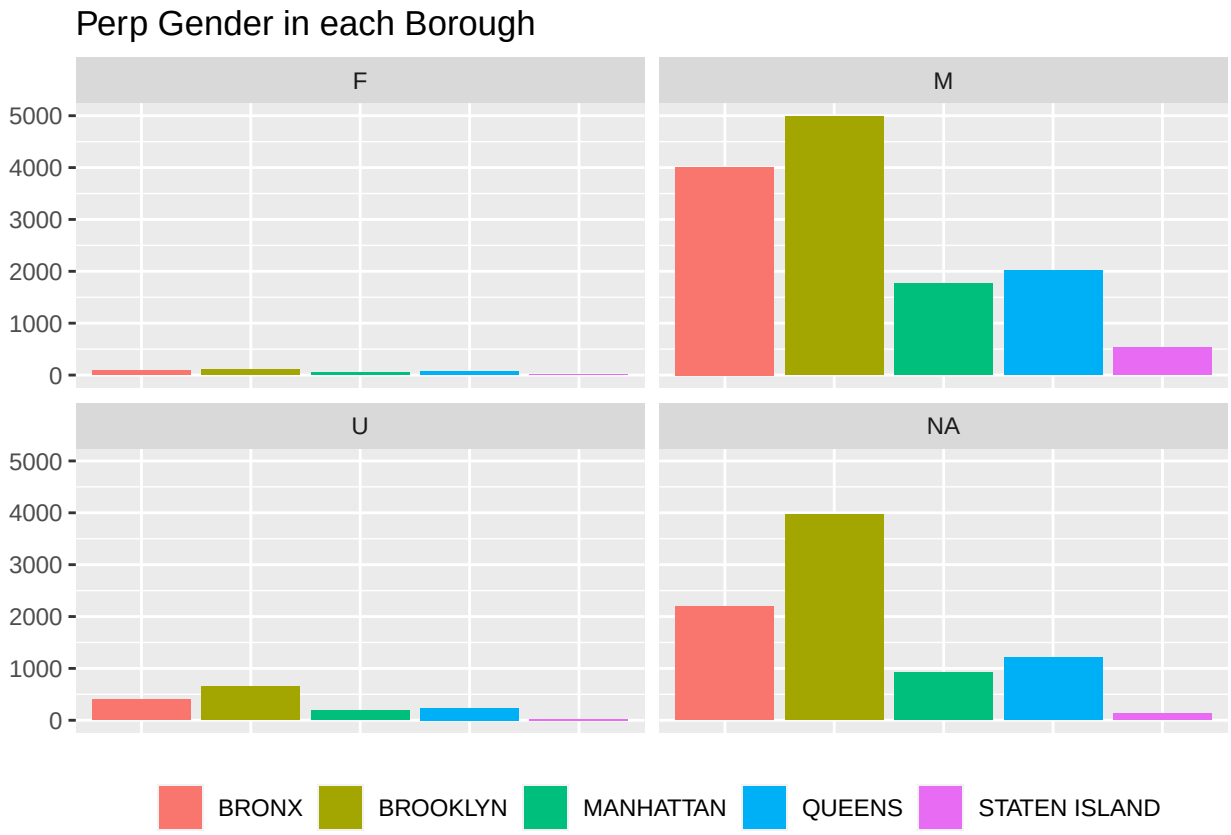
```
ggplot(shooting_data) +
  geom_bar(aes(x = BORO, fill = BORO)) +
  facet_wrap( ~ VIC_SEX) +
  theme(
    legend.position = "bottom",
    legend.title = element_blank(),
    axis.text.x = element_blank(),
```

```
    axis.ticks.x = element_blank(),
    axis.title = element_blank()
  ) +
  labs(title = "Shooting Victims in New York by Victim's Gender in each Borough", y = NULL)
```



Similarly we visualize the Perp data according to gender in each BORO.

```
ggplot(shooting_data) +
  geom_bar(aes(x = BORO, fill = BORO)) +
  facet_wrap( ~ PERP_SEX) +
  theme(
    legend.position = "bottom",
    legend.title = element_blank(),
    axis.text.x = element_blank(),
    axis.ticks.x = element_blank(),
    axis.title = element_blank()
  ) +
  labs(title = "Perp Gender in each Borough", y = NULL)
```

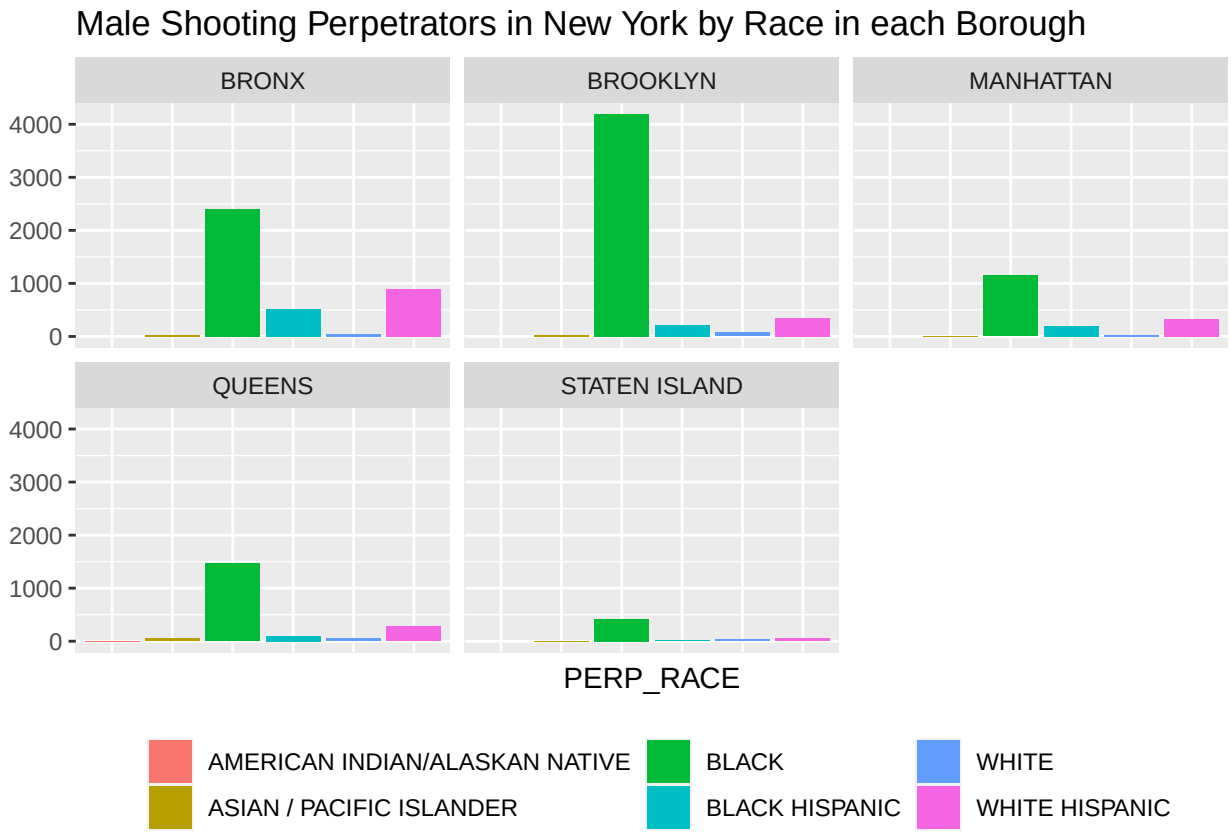


From the above plot we realize that according to the data set males are the predominant gender who are perpetrators in these incidents. Next, we analyze the number of males by race in each borough. In order to accomplish this, we first drop the non-male records from the data set(clean), then the records with unknown race:

```
shooting_data_male <- shooting_data %>% filter(PERP_SEX == "M")
shooting_data_male <- shooting_data_male %>% filter(PERP_RACE != "UNKNOWN")
```

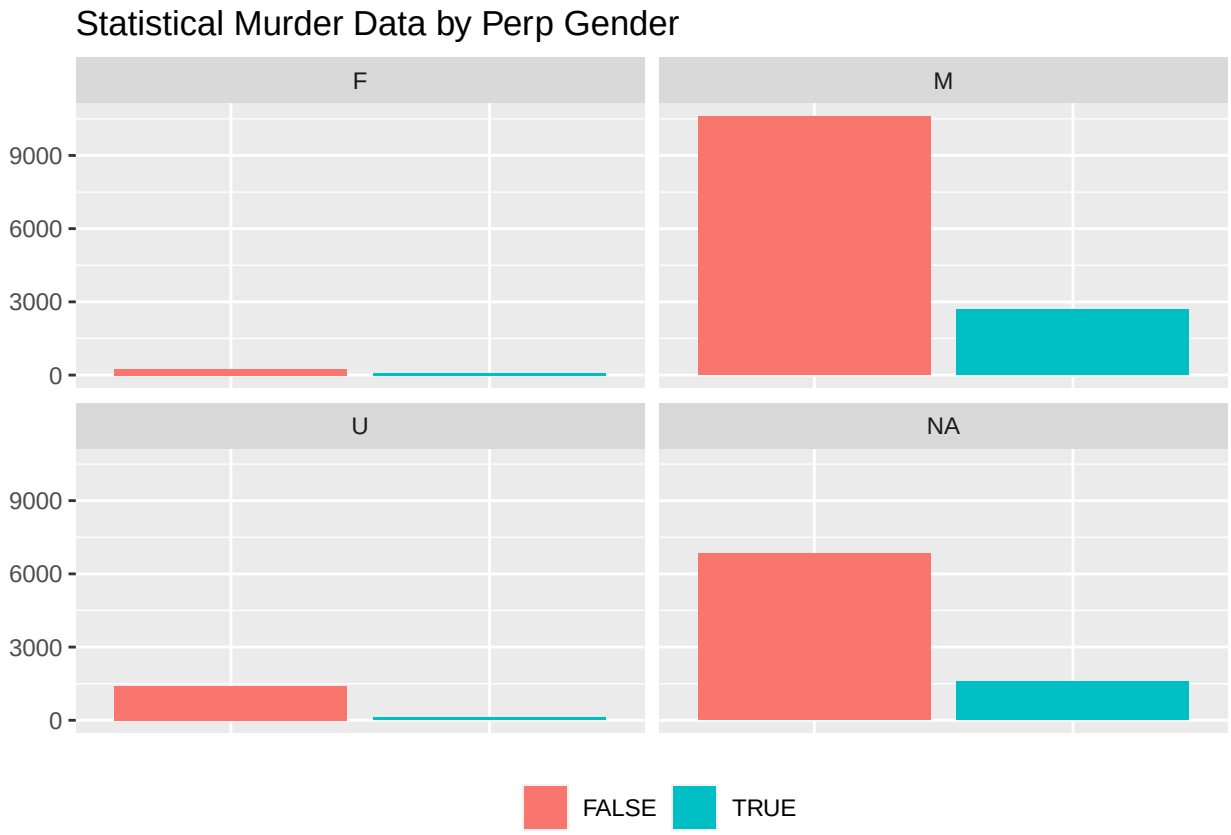
Finally, we plot the data:

```
ggplot(shooting_data_male) +
  geom_bar(aes(x = PERP_RACE, fill = PERP_RACE)) +
  facet_wrap( ~ BORO) +
  theme(
    legend.position = "bottom",
    legend.title = element_blank(),
    axis.text.x = element_blank(),
    axis.ticks.x = element_blank()
  ) +
  labs(title = "Male Shooting Perpetrators in New York by Race in each Borough", y = NULL)
```



Here, we analyze the statistical murder data by gender.

```
ggplot(shooting_data) +
  geom_bar(aes(x = STATISTICAL_MURDER_FLAG, fill = STATISTICAL_MURDER_FLAG)) +
  facet_wrap( ~ PERP_SEX) +
  theme(
    legend.position = "bottom",
    legend.title = element_blank(),
    axis.text.x = element_blank(),
    axis.ticks.x = element_blank(),
    axis.title = element_blank()
  ) +
  labs(title = "Statistical Murder Data by Perp Gender", y = NULL)
```



CONCLUSION:

By analyzing the data, we find that the most of the perpetrators are male while the victims are also overwhelmingly male. We can clearly note that males from the Bronx and Brooklyn seem to be more likely to be part of or be a victim of shooting incidents. This is a worrying trend. It could imply that the male population from these areas are not being paid attention to, in the sense that maybe they are not being offered opportunities to productively utilize their time, either at work or at school. I admit that this data set alone does not paint the entire picture and that we need to analyze the demography by the economic, social and health perspectives as well.

This analysis was done with the data collected from reported incidents. Unfortunately, a lot of incidents go unreported. Including those incidents may change the conclusion of this report. This is a potential bias. Another source of potential bias is the fact that there are a lot of `NULL` values in the race and gender columns. This could mean that either the data was unavailable or that the reporting officer failed to mention it in their incident report.