NYPD Shooting Incident Report

2023-11-08

Objective: Explore which regions within New York City appear to have the higher instances and observe a demographic of the instances with the parties involved.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- 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(lubridate)

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

```
## 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.
```

```
# To explore the data, we can utilize the command of "view(stats)"
view(stats)

head(stats)
```

```
## # A tibble: 6 x 21
##   INCIDENT_KEY OCCUR_DATE OCCUR_TIME BORO     LOC_OF_OCCUR_DESC PRECINCT
##          <dbl> <chr>      <time>     <chr>    <chr>                <dbl>
## 1    228798151 05/27/2021 21:30      QUEENS   <NA>                   105
## 2    137471050 06/27/2014 17:40      BRONX    <NA>                    40
## 3    147998800 11/21/2015 03:56      QUEENS   <NA>                   108
```

```
## 4    146837977 10/09/2015 18:30      BRONX    <NA>                    44
## 5     58921844 02/19/2009 22:58      BRONX    <NA>                    47
## 6    219559682 10/21/2020 21:36      BROOKLYN <NA>                    81
## # i 15 more variables: JURISDICTION_CODE <dbl>, LOC_CLASSFCTN_DESC <chr>,
## #   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>
```

```
# The command "head(stats)" allows us to visualize the first six occurances within our data named "stat
str(stats)
```

```
## spc_tbl_ [27,312 x 21] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ INCIDENT_KEY           : num [1:27312] 2.29e+08 1.37e+08 1.48e+08 1.47e+08 5.89e+07 ...
##  $ OCCUR_DATE             : chr [1:27312] "05/27/2021" "06/27/2014" "11/21/2015" "10/09/2015" ...
##  $ OCCUR_TIME             : 'hms' num [1:27312] 21:30:00 17:40:00 03:56:00 18:30:00 ...
##   ..- attr(*, "units")= chr "secs"
##  $ BORO                   : chr [1:27312] "QUEENS" "BRONX" "QUEENS" "BRONX" ...
##  $ LOC_OF_OCCUR_DESC      : chr [1:27312] NA NA NA NA ...
##  $ PRECINCT               : num [1:27312] 105 40 108 44 47 81 114 81 105 101 ...
##  $ JURISDICTION_CODE      : num [1:27312] 0 0 0 0 0 0 0 0 0 0 ...
##  $ LOC_CLASSFCTN_DESC     : chr [1:27312] NA NA NA NA ...
##  $ LOCATION_DESC          : chr [1:27312] NA NA NA NA ...
##  $ STATISTICAL_MURDER_FLAG: logi [1:27312] FALSE FALSE TRUE FALSE TRUE TRUE ...
##  $ PERP_AGE_GROUP         : chr [1:27312] NA NA NA NA ...
##  $ PERP_SEX               : chr [1:27312] NA NA NA NA ...
##  $ PERP_RACE              : chr [1:27312] NA NA NA NA ...
##  $ VIC_AGE_GROUP          : chr [1:27312] "18-24" "18-24" "25-44" "<18" ...
##  $ VIC_SEX                : chr [1:27312] "M" "M" "M" "M" ...
##  $ VIC_RACE               : chr [1:27312] "BLACK" "BLACK" "WHITE" "WHITE HISPANIC" ...
##  $ X_COORD_CD             : num [1:27312] 1058925 1005028 1007668 1006537 1024922 ...
##  $ Y_COORD_CD             : num [1:27312] 180924 234516 209837 244511 262189 ...
##  $ Latitude               : num [1:27312] 40.7 40.8 40.7 40.8 40.9 ...
##  $ Longitude              : num [1:27312] -73.7 -73.9 -73.9 -73.9 -73.9 ...
##  $ Lon_Lat                : chr [1:27312] "POINT (-73.73083868899994 40.662964620000025)" "POINT (-73
##  - attr(*, "spec")=
##   .. cols(
##   ..   INCIDENT_KEY = col_double(),
##   ..   OCCUR_DATE = col_character(),
##   ..   OCCUR_TIME = col_time(format = ""),
##   ..   BORO = col_character(),
##   ..   LOC_OF_OCCUR_DESC = col_character(),
##   ..   PRECINCT = col_double(),
##   ..   JURISDICTION_CODE = col_double(),
##   ..   LOC_CLASSFCTN_DESC = col_character(),
##   ..   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_double(),
```

```
##   ..   Y_COORD_CD = col_double(),
##   ..   Latitude = col_double(),
##   ..   Longitude = col_double(),
##   ..   Lon_Lat = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
# The structure of our data indicates that we need to tidy up our information to make sure that our dat

# This will begin our tidying and transforming section from our basic data import section


library(dplyr)
newstats = na.omit(stats)

newstats2 = newstats %>% select(INCIDENT_KEY, OCCUR_DATE,OCCUR_TIME, BORO, STATISTICAL_MURDER_FLAG, PERP

sum(is.null(newstats2))
```

```
## [1] 0
```

```
sum(na.omit(NULL))
```

```
## [1] 0
```

```
glimpse(newstats2)
```

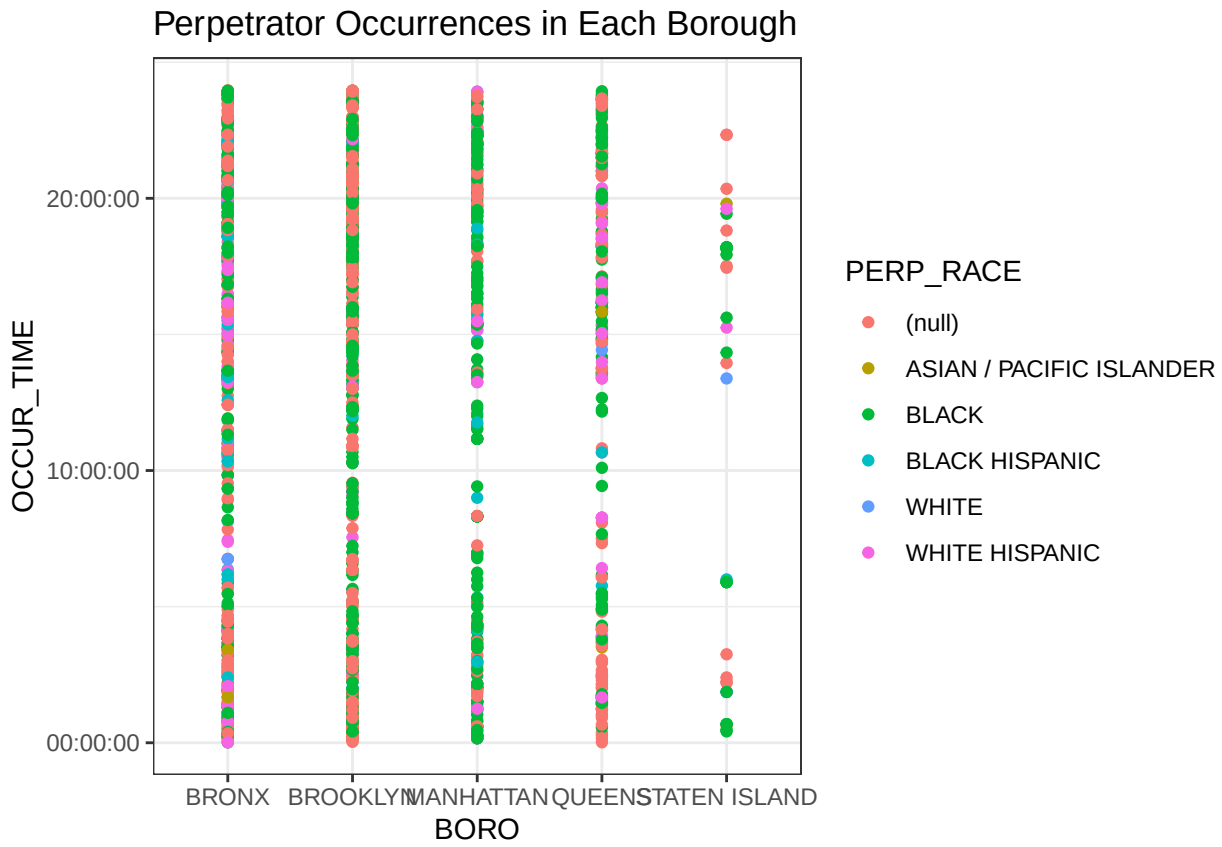
```
## Rows: 1,706
## Columns: 11
## $ INCIDENT_KEY            <dbl> 243566884, 256484816, 250216145, 239207164, 24~
## $ OCCUR_DATE              <chr> "04/12/2022", "12/17/2022", "08/27/2022", "01/~
## $ OCCUR_TIME              <time> 22:08:00, 04:08:00, 00:21:00, 19:50:00, 01:19~
## $ BORO                    <chr> "BRONX", "BRONX", "BRONX", "QUEENS", "BROOKLYN~
## $ STATISTICAL_MURDER_FLAG <lgl> TRUE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE~
## $ PERP_AGE_GROUP          <chr> "18-24", "25-44", "(null)", "(null)", "(null)"~
## $ PERP_SEX                <chr> "M", "M", "(null)", "(null)", "(null)", "(null~
## $ PERP_RACE               <chr> "BLACK HISPANIC", "WHITE HISPANIC", "(null)", ~
## $ VIC_AGE_GROUP           <chr> "18-24", "25-44", "25-44", "18-24", "25-44", "~
## $ VIC_SEX                 <chr> "F", "M", "M", "M", "F", "M", "M", "M", "F", "~
## $ VIC_RACE                <chr> "BLACK", "WHITE HISPANIC", "WHITE HISPANIC", "~
```

```
# Basic check up to determine if the coding script selected only the following columns and discarded th
```

The following code will illustrate charts and graphs to best represent the data that has been supplied.

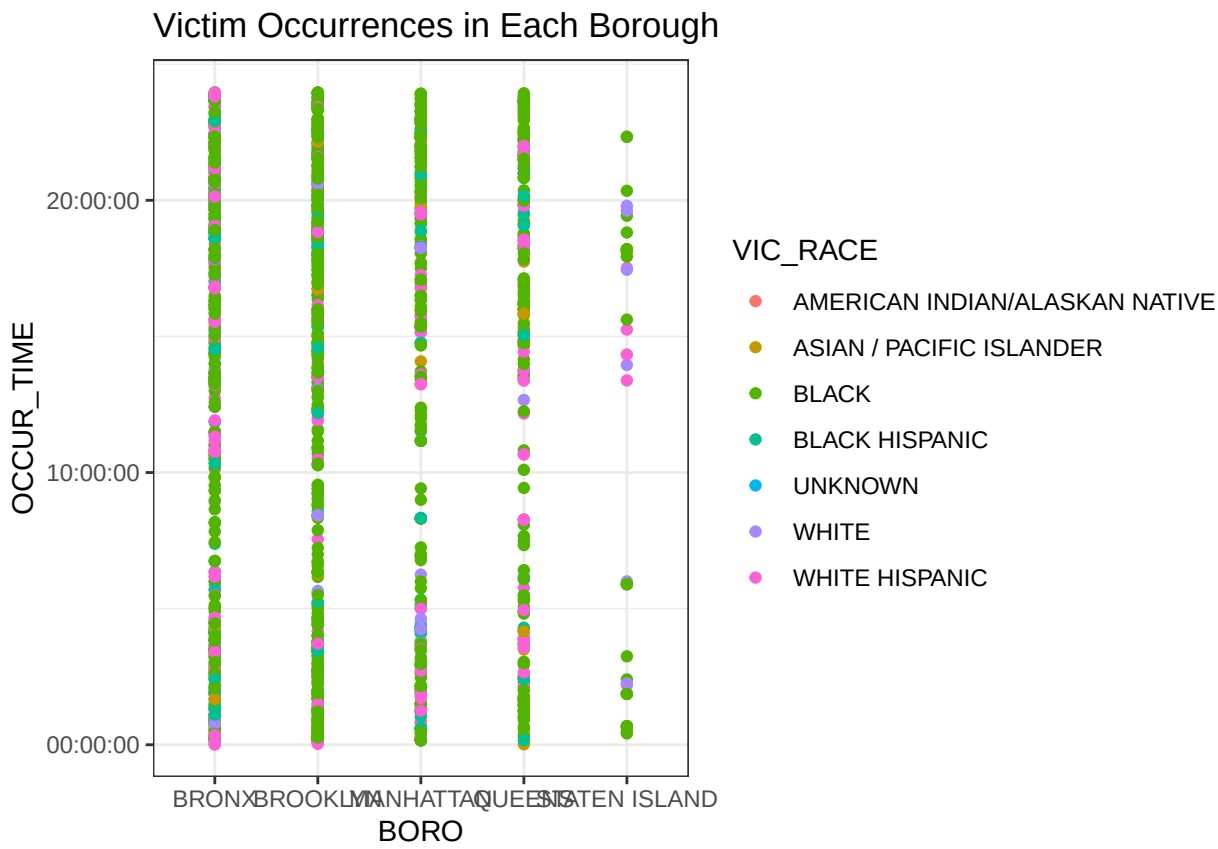
```
data()

ggplot(newstats2, aes(x = BORO,
                      y = OCCUR_TIME,))+
  geom_point(mapping = aes(color = PERP_RACE))+
  labs(title = "Perpetrator Occurrences in Each Borough")+
  theme_bw()
```



This graph indicates which areas within NYC display the most frequent altercations by all the races at the specific times.

```
ggplot(newstats2, aes(x = BORO,
                      y = OCCUR_TIME,))+
  geom_point(mapping = aes(color = VIC_RACE))+
  labs(title = "Victim Occurrences in Each Borough")+
  theme_bw()
```



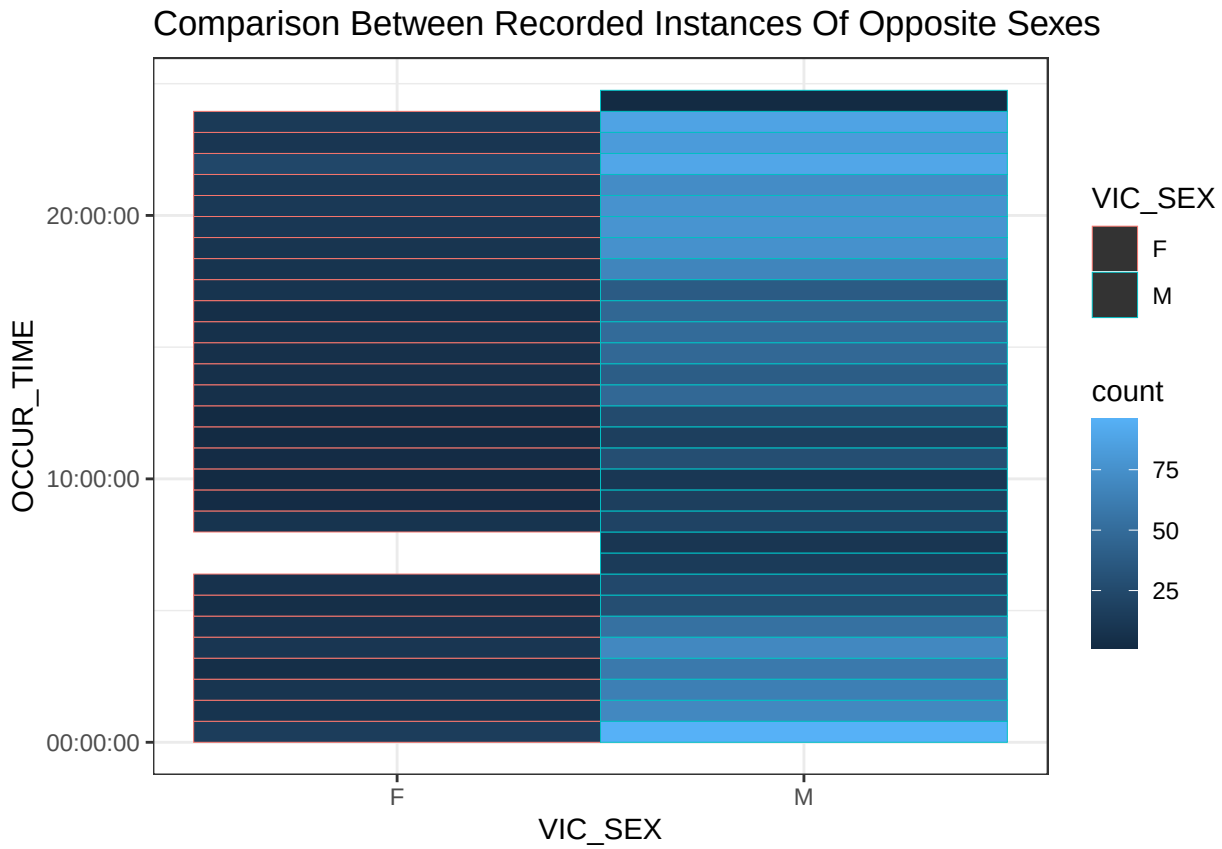
To use as a comparative measure, we will now illustrate the occurrences from the victims within each of the boroughs within NYC. One thing to note of importance is these all correlate exactly with each other and there don't appear to be any outliers. This would make sense as if there were noticeable outliers that would mean the incident was reported to the authorities but had no victim or perpetrator proving it to be illogical unless it was observed by a passerby.

Question of Interest

Are there any statistical differences that illuminate to a vast difference in women and men victim incidences/altercations?

```
view(newstats2)

newstats2 %>%
  ggplot(aes(VIC_SEX, OCCUR_TIME,
             color = VIC_SEX))+
  geom_bin2d()+
  labs(title = "Comparison Between Recorded Instances Of Opposite Sexes")+
  theme_bw()
```



Analysis

During this final analysis, I was curious to see how the recorded instances between both male and female sexes stacked in comparison to one another. In this chart we are able to understand that it appears there is a greater number of recorded incidences (victims) within the male population as a whole than the female population regardless of the boroughs of NYC. It's also very interesting to note that during this observation we are able to see there aren't any recorded instances between, what would appear to be about 6AM and 7:30AM for women; and during that same time frame, men also experience a relatively low incident rate as well. That time frame typically associated with the beginning of the work day could be a plausible reason.

Conclusion And Bias Interpretation

After a thorough investigation (statistically) it's beyond reasonable doubt that there are a vast majority of instances that differ from one another involving multiple factors such as race, what day it was, and sex. As shown in the model, there is a stark difference in the recordings as overall there are much more male victims regardless of the time. Now, sources for bias could stem from multiple angles such as policing differences within the boroughs as these data collection (model) shows no exact borough rather just a collection of data supplied by the NYPD data sheet. Additionally, this data could be strongly connected to the actions the individuals are partaking in rather than just random violence. My one alteration I would do would be to get an even more precise area involving these shootings as well as additional information pertaining to the stark difference between both genders.