Project 1 - NYPD Analysis

Ian Thulin

2023-03-01

After importing the data, we will begin the data cleaning or "tidying" stage of our project. Initially we will explore the data set to understand what information we will use in our analysis and what we can remove. In my case I have decided that the Date and Time records have not been recorded in a method that will allow further study, so I had to go through the process of formatting each column of the data set into the proper date/time type. I also decided to remove the columns related to the incident, key, jurisdiction code, location description, x coordinate, and y coordinates.

```r
# Adjust the formatting for the OCCUR_DATE column and remove unneccesary columns of data
NYPD_data$OCCUR_DATE <- mdy(NYPD_data$OCCUR_DATE)
NYPD_data$OCCUR_DATE <- format(as.Date(NYPD_data$OCCUR_DATE), "%Y")
NYPD_data <- subset(NYPD_data, select = -c(INCIDENT_KEY, JURISDICTION_CODE, OCCUR_TIME, LOCATION_DESC, )
NYPD_data$VIC_SEX[NYPD_data$VIC_SEX == "M"] <- 1
NYPD_data$VIC_SEX[NYPD_data$VIC_SEX == "F"] <- 0
NYPD_data$VIC_SEX <- as.integer(NYPD_data$VIC_SEX)
NYPD_data$num <- 1:nrow(NYPD_data)
```


In order to more effectively map the different boroughs we need to separate our data frame into subframes associated with each Borough of New York.

```r
# Create a new data frame for each Boro to allow for graphing in seperate colors
NYPD_data_Bronx <- NYPD_data[NYPD_data$BORO == 'BRONX',]
NYPD_data_Brooklyn <- NYPD_data[NYPD_data$BORO == 'BROOKLYN',]
NYPD_data_Manhattan <- NYPD_data[NYPD_data$BORO == 'MANHATTAN',]
NYPD_data_Queens <- NYPD_data[NYPD_data$BORO == 'QUEENS',]
NYPD_data_Staten_Island <- NYPD_data[NYPD_data$BORO == 'STATEN ISLAND',]
```

In this next code block we will set the boundary for the mapping region, then plot each Borough in it's own color as an overlay filter for the Google Map we previously Obtained.

Shooting Incidents as reported to NYPD

6/1/2006 – 11/11/2021, https://data.cityofnewyork.us/api/views/833y–f:
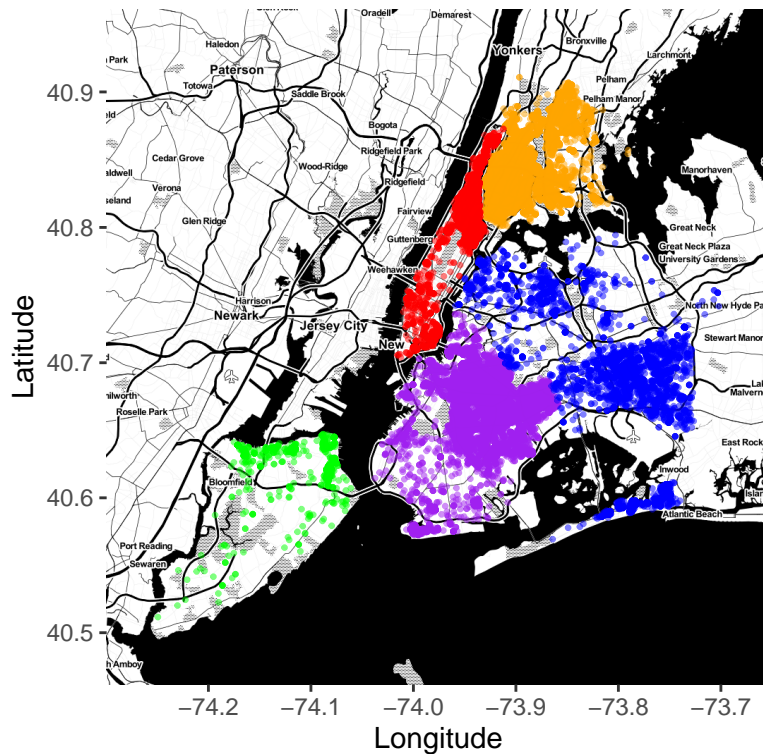
Finally we are left with the above plot. The location of each incident included in our data set has been plotted on the google map overlay by utilizing the latitude and longitude of the data point. AS we can see there are several area on the map that have a higher density of incidents than their surrounding neighborhoods. This may be reflective of the overall crime rate throughout different areas of the city, but ultimately is probably quite muddled by the numerous confounding variables that our data most likely has.

One of the most pressing confounding variables is the fact that all of the data points in this data set come from reported crimes that occured after a victim had already been discovered. It is possible that the victims of these incidents had been moved from the scene of the crime in order to hide the actual location of the crime. Additionally I would expect some areas of the city to have a higher rate of patrol activity which may act to further deter the activities of many criminal elements.

Correlatation between male and female victims in different Boroughs
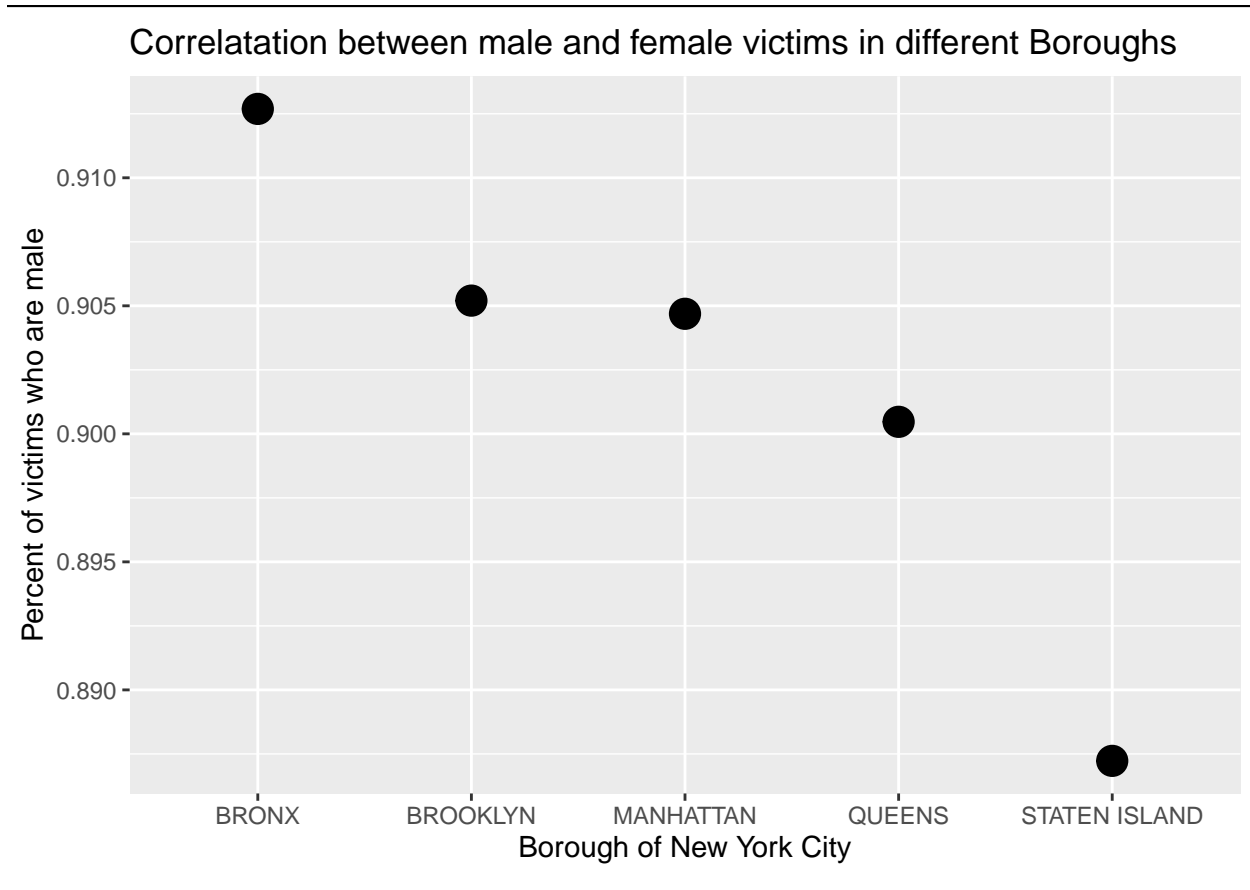
A second point of interest for this data set was to look at how skewed the proportion of male to female victims is. I believe most people would not bbe surprised to learn that men are significantly more likely to die to aggressive circumstances. In particular I was aware that the vast majority of gun violence victims were male, regardless of if the gun was used on oneself or on a different individual. For this analysis I looked at two seperate charte. On the 1st chart we graphed the percentage of male victims against the Borough in which the victim was killed. I was curious if different districs in the city hade a significant skew towards victims of a particular sex. Overall the vast majority, more than 88% in all boroughs, of victims were male with the most victims being found in the Bronx. While the overall percentage was only 2% higher, the Bronx data point was noticeably higher than the other Boroughs. I would be curious to gather some socio-economic data about the residents of each bough in order to determine if the difference in percent male victims is tied to lower income areas of the city.

The second plot sought was created in order to examine the changed in the percent of male victims over subsequent years. Overall the mean percentage of male victims is consistent with our previous graph, showing a difference of roughly 2% between the highlest and lowest points of our data set. It does apear that while the pattern is somewhat cyclical there has been an overall downward trend in male victims as time goes on. This could be an interesting point to explore the relationship between changing levels of gang-related crime incidents and the proportion of male victims of gun related crimes.

Correlatation between male and female victims in different Boroughs

D. Kahle and H. Wickham. ggmap: Spatial Visualization with ggplot2. The R Journal, 5(1), 144-161. URL http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf

```
## R version 4.2.1 (2022-06-23 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19044)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.utf8
## [2] LC_CTYPE=English_United States.utf8
## [3] LC_MONETARY=English_United States.utf8
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.utf8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] ggmap_3.0.1     ggplot2_3.4.1   lubridate_1.9.2
##
```

```
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.10        highr_0.10         plyr_1.8.8
##  [4] pillar_1.8.1       compiler_4.2.1     bitops_1.0-7
##  [7] tools_4.2.1        digest_0.6.31      evaluate_0.20
## [10] lifecycle_1.0.3    tibble_3.1.8       gtable_0.3.1
## [13] timechange_0.2.0   lattice_0.20-45    pkgconfig_2.0.3
## [16] png_0.1-8          rlang_1.0.6        cli_3.6.0
## [19] rstudioapi_0.14    curl_5.0.0         yaml_2.3.7
## [22] xfun_0.37          fastmap_1.1.1      withr_2.5.0
## [25] stringr_1.5.0      dplyr_1.1.0        httr_1.4.5
## [28] knitr_1.42         generics_0.1.3     RgoogleMaps_1.4.5.3
## [31] vctrs_0.5.2        grid_4.2.1         tidyselect_1.2.0
## [34] glue_1.6.2         R6_2.5.1           jpeg_0.1-10
## [37] fansi_1.0.4        rmarkdown_2.20     sp_1.6-0
## [40] farver_2.1.1       tidyr_1.3.0        purrr_1.0.1
## [43] magrittr_2.0.3     scales_1.2.1       htmltools_0.5.4
## [46] colorspace_2.1-0   labeling_0.4.2     utf8_1.2.3
## [49] stringi_1.7.12     munsell_0.5.0
```