NYPD Shooting Incident Data Report

2/5/2022

Data Set Description

The data set is "a breakdown of every shooting incident that occurred in NYC going back to 2006 through the end of the previous calendar year. This data is manually extracted every quarter and reviewed by the Office of Management Analysis and Planning before being posted on the NYPD website. Each record represents a shooting incident in NYC and includes information about the event, the location and time of occurrence." https://data.cityofnewyork.us/browse?q=shooting+incidents

The data set can be found at https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType= DOWNLOAD

New York City (NYC) Shooting Incidents Are On The Rise in 2020

(Note: echo = TRUE so that R code may be seen as per course instructions)

Shooting incidents in NYC have generally been falling, most notably and consistently since 2011, however 2020 has shown a sharp increase in shooting incidents across Brooklyn, the Bronx, Manhattan, Queens, and Staten Island with the highest number of shootings and increase being in Brooklyn. The increase seems to coincide with the start of the COVID-19 pandemic though there is not enough data to confirm this hypothesis. This is shown in the following diagram:

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

shooting <- mutate(shooting, YEAR = year(mdy(OCCUR_DATE)), MONTH = month(mdy(OCCUR_DATE)), DAY = day(mdy
```

```r
total <- distinct(shooting %>% add_count(BORO, YEAR) %>% select(BORO,YEAR, n) %>% rename(SHOOTINGS = n))
xlabels = min(shooting$YEAR):max(shooting$YEAR)
```

```r
ggplot(total, aes(x = YEAR, y = SHOOTINGS)) +
  geom_line(aes(color = BORO, linetype = BORO)) +
  scale_x_continuous(labels=xlabels,breaks = xlabels) +
  theme(legend.position = "bottom") +
  labs(title = "Shootings by Year",
       x = "Year",
       y = "Shootings")
```

Shootings by Year
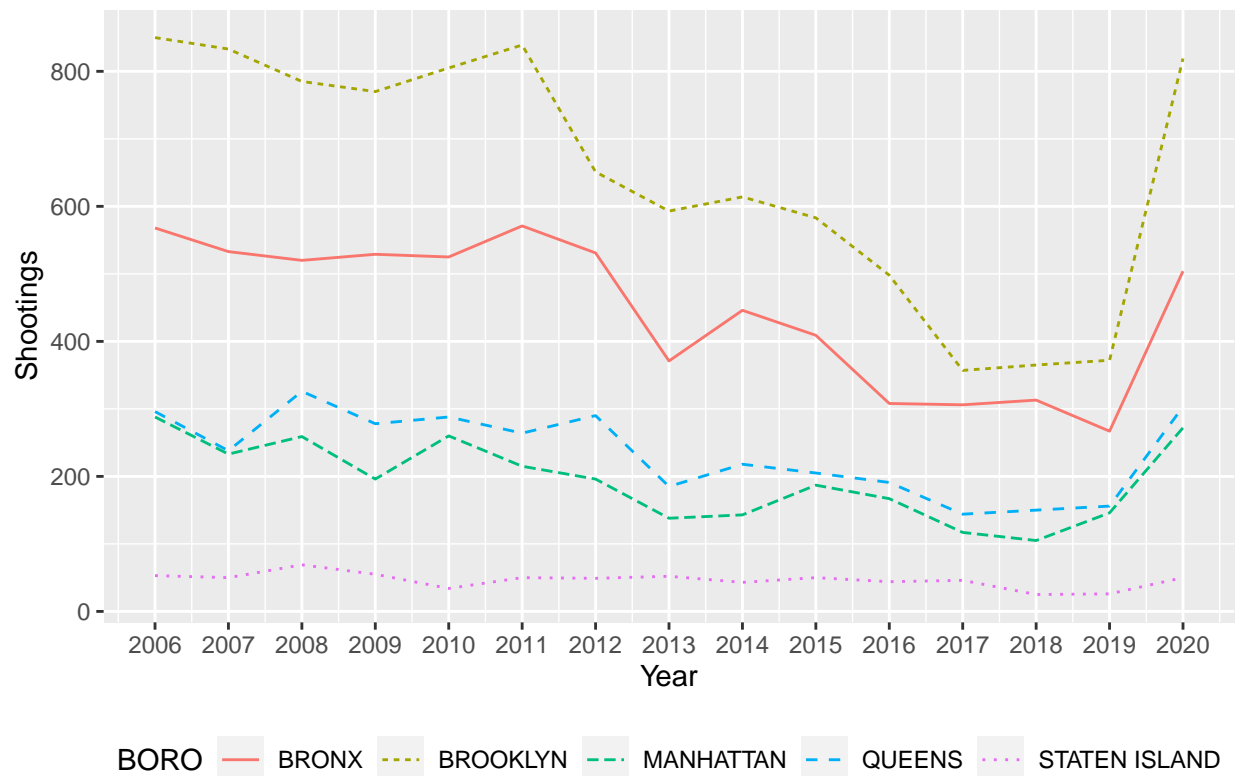
Further Analysis and Modeling

Not all shootings result in a murder. Murders are generally on the rise as well after 2011 which may indicate that as the number of shootings increase so do the numnber of murders.

```
murders <- filter(shooting, STATISTICAL_MURDER_FLAG == 'TRUE')
total_murders <- distinct(murders %>% add_count(BORO, YEAR) %>% select(BORO,YEAR, n) %>% rename(SHOOTING
xlabels = min(murders$YEAR):max(murders$YEAR)
ggplot(total_murders, aes(x = YEAR, y = SHOOTINGS)) +
  geom_line(aes(color = BORO, linetype = BORO)) +
  scale_x_continuous(labels=xlabels,breaks = xlabels) +
  theme(legend.position = "bottom") +
  labs(title = "Murders by Year",
       x = "Year",
       y = "Murders")
```

Murders by Year
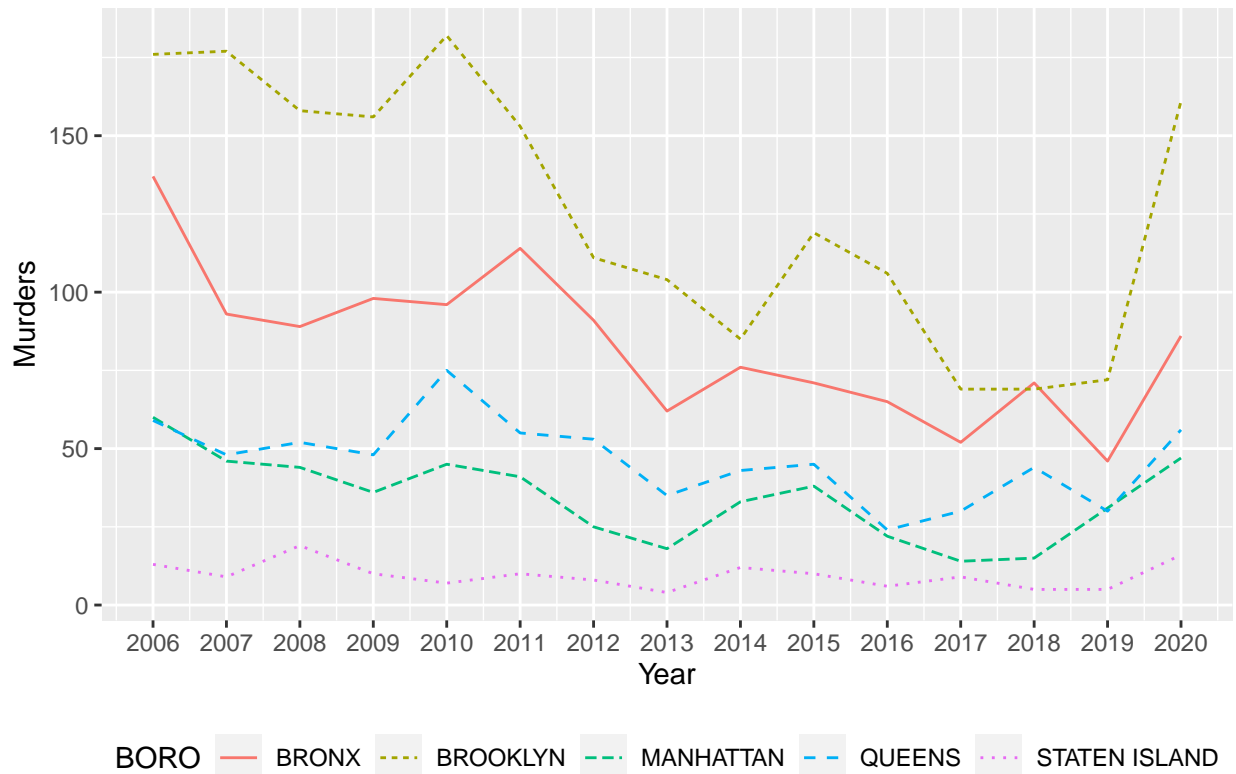
It is interesting to see if the number of shootings affect the number of murders, or if the proportion of shootings that are murders remains constant or can be predicted based on the number of shootings. Based on the the following model, the percentage of shootings that are murders does vary a fair bit but does collect around 19%.

```
total <- distinct(shooting %>% add_count(BORO, YEAR) %>% select(BORO,YEAR, n) %>% rename(SHOOTINGS = n))
total_murders <- distinct(murders %>% add_count(BORO, YEAR) %>% select(BORO,YEAR, n) %>% rename(MURDERS
tot = distinct(inner_join(total,total_murders))
tot <- mutate(tot, PERCENT_MUR=MURDERS/SHOOTINGS)
mod <- lm(PERCENT_MUR ~ SHOOTINGS, data = tot)
tot_w_pred <- tot %>% mutate(pred = predict(mod))

tot_w_pred %>% ggplot() +
  geom_point(aes(x = SHOOTINGS, y=PERCENT_MUR), color = "blue") +
  geom_point(aes(x = SHOOTINGS, y=pred), color = "red") +
  labs(title = "Graph: Model",
       x = "Shootings",
       y = "Murder Percentage")
```

Graph: Model
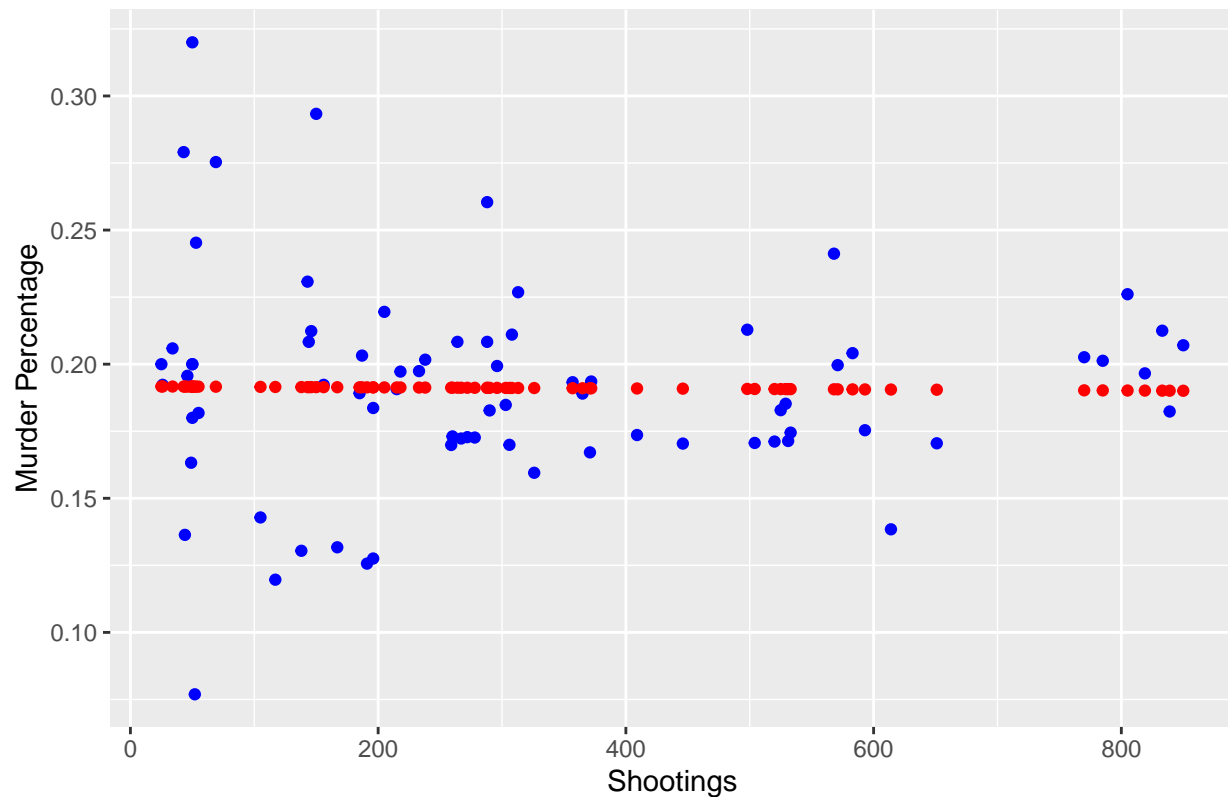
Conclusion

Shootings in NYC have risen sharply in 2020, especially in Brooklyn, which coincides with the start of the COVID-19 pandemic. This data set does not have enough data in order to confirmed that COVID-19 is a cause but this does form an interesting hypothesis that should be further investigated. The presented model does not show that the number of shootings is a strong predictor of how many of them will be murders. The only conclusion that can be drawn is that, if there are more shootings, it is likely that there are more murders. The data that is included does not look at ethnicity or pin point exact locations, so it is unlikely that any bias has been introduced.

```
## R version 4.1.0 (2021-05-18)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19042)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
```

```
##
## other attached packages:
##  [1] lubridate_1.7.10 forcats_0.5.1    stringr_1.4.0    dplyr_1.0.7
##  [5] purrr_0.3.4      readr_2.0.1      tidyr_1.1.3      tibble_3.1.4
##  [9] ggplot2_3.3.5    tidyverse_1.3.1
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.7       assertthat_0.2.1 digest_0.6.27    utf8_1.2.2
##  [5] R6_2.5.1         cellranger_1.1.0 backports_1.2.1  reprex_2.0.1
##  [9] evaluate_0.14    highr_0.9        httr_1.4.2       pillar_1.6.2
## [13] rlang_0.4.11     curl_4.3.2       readxl_1.3.1     rstudioapi_0.13
## [17] rmarkdown_2.10   labeling_0.4.2   bit_4.0.4        munsell_0.5.0
## [21] broom_0.7.9      compiler_4.1.0   modelr_0.1.8     xfun_0.25
## [25] pkgconfig_2.0.3  htmltools_0.5.2  tidyselect_1.1.1 fansi_0.5.0
## [29] crayon_1.4.1     tzdb_0.1.2       dbplyr_2.1.1     withr_2.4.2
## [33] grid_4.1.0       jsonlite_1.7.2   gtable_0.3.0     lifecycle_1.0.0
## [37] DBI_1.1.1        magrittr_2.0.1   scales_1.1.1     cli_3.0.1
## [41] stringi_1.7.4    vroom_1.5.4      farver_2.1.0     fs_1.5.0
## [45] xml2_1.3.2       ellipsis_0.3.2   generics_0.1.0   vctrs_0.3.8
## [49] tools_4.1.0      bit64_4.0.5      glue_1.4.2       hms_1.1.0
## [53] parallel_4.1.0   fastmap_1.1.0    yaml_2.2.1       colorspace_2.0-2
## [57] rvest_1.0.1      knitr_1.33       haven_2.4.3
```