

NYPD Shooting Incidents

Carlos Barron

2022-01-16

Libraries

First I will import the libraries I will use for my report.

```
library(tidyverse)
library(knitr)
library(lubridate)
library(ggplot2)
```

Data Source

I will include the NYPD shooting incidents data that can be found here: (<https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD>)

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

Variables

Then I will create some variables for manipulation and categorization of data.

```
time_shifts = hour(hm("00:00", "6:00", "12:00", "18:00", "23:59"))
time_of_day = c("Night", "Morning", "Afternoon", "Evening")
q = setNames( rep(c('Winter', 'Spring', 'Summer', 'Fall'), each=3), c(12, 1:11))
```

Dataset cleanup

I will not use Lon_Lat, X_COORD_CD, Y_COORD_CD, VIC_RACE, PERP_RACE, LOCATION_DESC, JURISDICTION_CODE, PRECINCT so I will omit from my data set. I will create the new columns MONTH_NUM, change the date as type mdy, and set the time of day and season.

```
nypd_data = nypd_data %>%
  mutate(
    MONTH_NUM = as.numeric(format(as.Date(OCCUR_DATE, '%m/%d/%Y'), '%m')),
    HOUR = as.numeric(hour(OCCUR_TIME)),
    OCCUR_DATE = mdy(OCCUR_DATE),
    TIME_OF_DAY = cut(x=hour(OCCUR_TIME), breaks = time_shifts, labels = time_of_day, include.lowest=TRUE),
    SEASON = unname(q[as.character(MONTH_NUM)]),
  ) %>%
```

```

rename(
  DATE = 'OCCUR_DATE',
  TIME = 'OCCUR_TIME',
  ID = 'INCIDENT_KEY') %>%
select(-c(Lon_Lat, X_COORD_CD, Y_COORD_CD, VIC_RACE,
          PERP_RACE, LOCATION_DESC, JURISDICTION_CODE, PRECINCT))

```

Then I set my model on the TIME_SQ column as a quadratic function.

```

crimes_by_time_of_day = nypd_data %>% group_by(TIME_OF_DAY, TIME, HOUR, BORO, SEASON) %>% summarize(CASES = sum(CASES))
quadr = lm(CASES ~ TIME_SQ, data = crimes_by_time_of_day)
crimes_by_time_of_day = crimes_by_time_of_day %>% mutate(PREDICTION = predict(quadr))
crimes_by_season = nypd_data %>% group_by(SEASON, BORO) %>% summarize(CASES = sum(Latitude/Latitude), MURDERS = sum(MURDERS))

```

Data Summary

Here is a summary of the data by time of day and season:

```
summary(crimes_by_time_of_day)
```

```

##      TIME_OF_DAY      TIME      BORO      CASES
## Night      :9679  Length:25596  Length:25596  Min.   : 1.000
## Morning    :1707  Class1:hms     Class :character 1st Qu.: 2.000
## Afternoon  :5202  Class2:difftime  Mode  :character Median : 3.000
## Evening    :9008  Mode :numeric    Mean  : 4.805
##                                     3rd Qu.: 6.000
##                                     Max.   :34.000
##      MURDERS      TIME_SQ      HOUR      SEASON
## Min.   : 0.0000  Min.   : 2.00  Min.   : 0.00  Length:25596
## 1st Qu.: 0.0000  1st Qu.: 6.50  1st Qu.: 3.00  Class :character
## Median : 0.0000  Median :12.12  Median :15.00  Mode  :character
## Mean   : 0.9556  Mean   :11.24  Mean   :12.19
## 3rd Qu.: 1.0000  3rd Qu.:14.50  3rd Qu.:20.00
## Max.   :10.0000  Max.   :20.00  Max.   :23.00
##      PREDICTION
## Min.   :2.648
## 1st Qu.:3.699
## Median :5.012
## Mean   :4.805
## 3rd Qu.:5.567
## Max.   :6.851

```

```
summary(crimes_by_season)
```

```

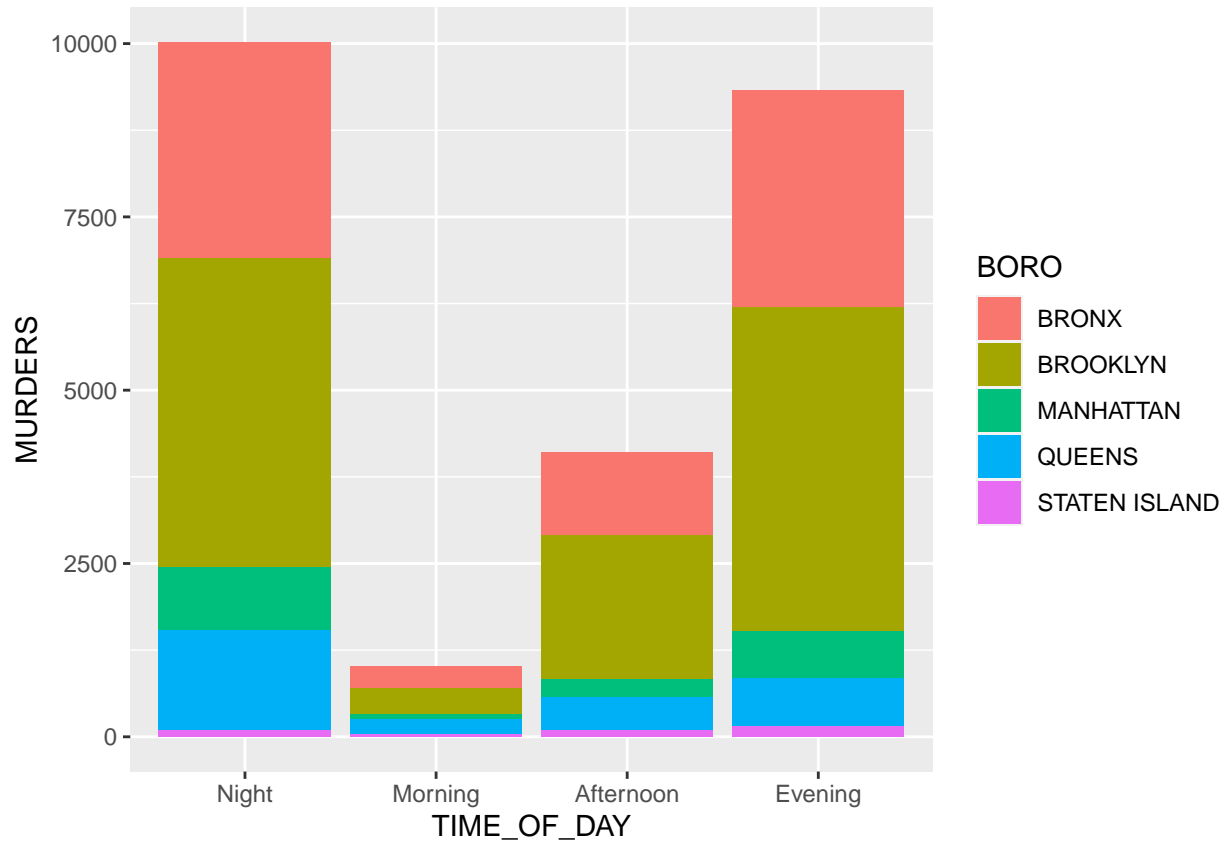
##      SEASON      BORO      CASES      MURDERS
## Length:20      Length:20  Min.   : 157  Min.   : 34.0
## Class :character  Class :character 1st Qu.: 686  1st Qu.:117.8
## Mode  :character  Mode  :character Median :1016  Median :190.0
##                                     Mean  :1280  Mean  :246.4
##                                     3rd Qu.:1800 3rd Qu.:351.5
##                                     Max.   :3642  Max.   :671.0

```

Visualizations and Analysis

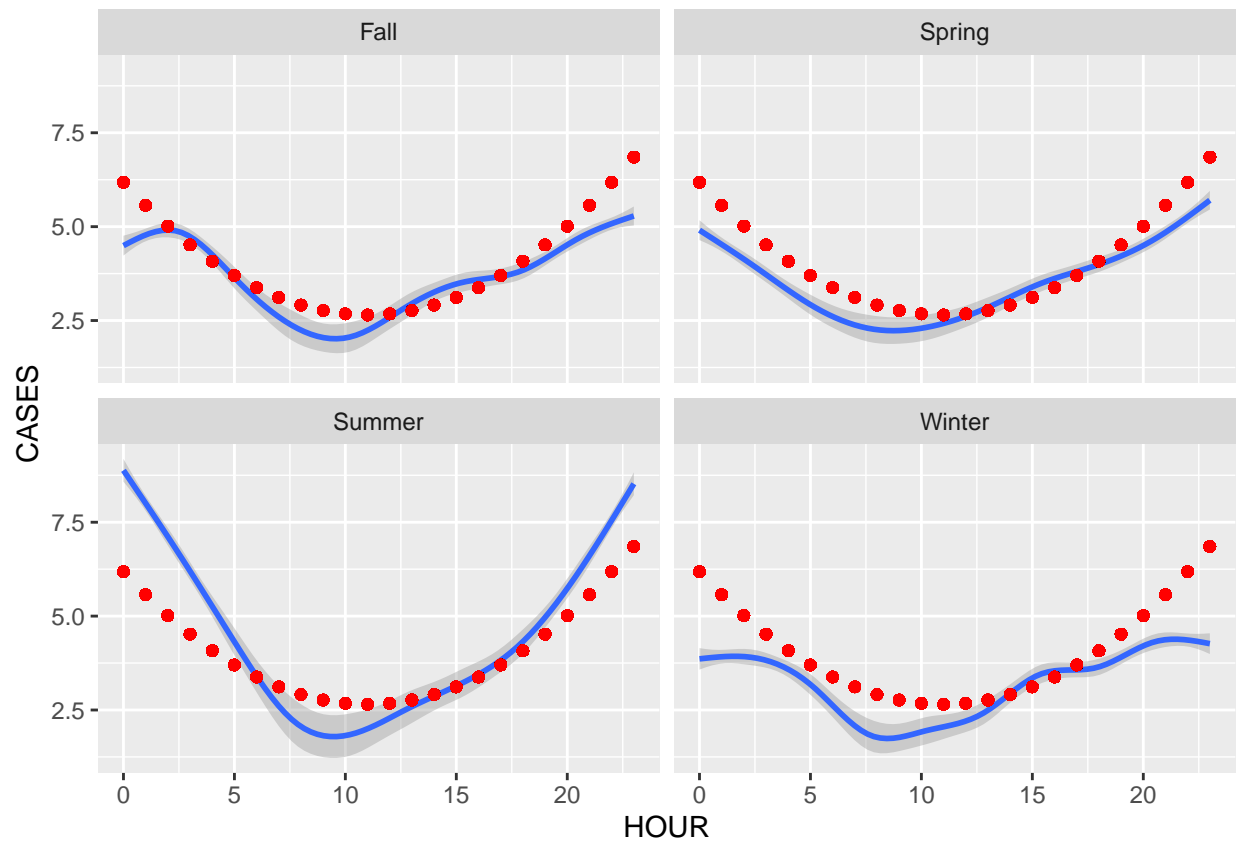
In the graphs it can be seen how there are more crimes during the evening and at night. During some seasons, there are less hours of sunlight, this could be affecting how my model is predicting incidents. It seems like during the summer more incidents occur.

```
ggplot2::ggplot(data = crimes_by_time_of_day) + geom_bar(mapping = aes(x=TIME_OF_DAY, y=MURDERS, fill=BORO))
```

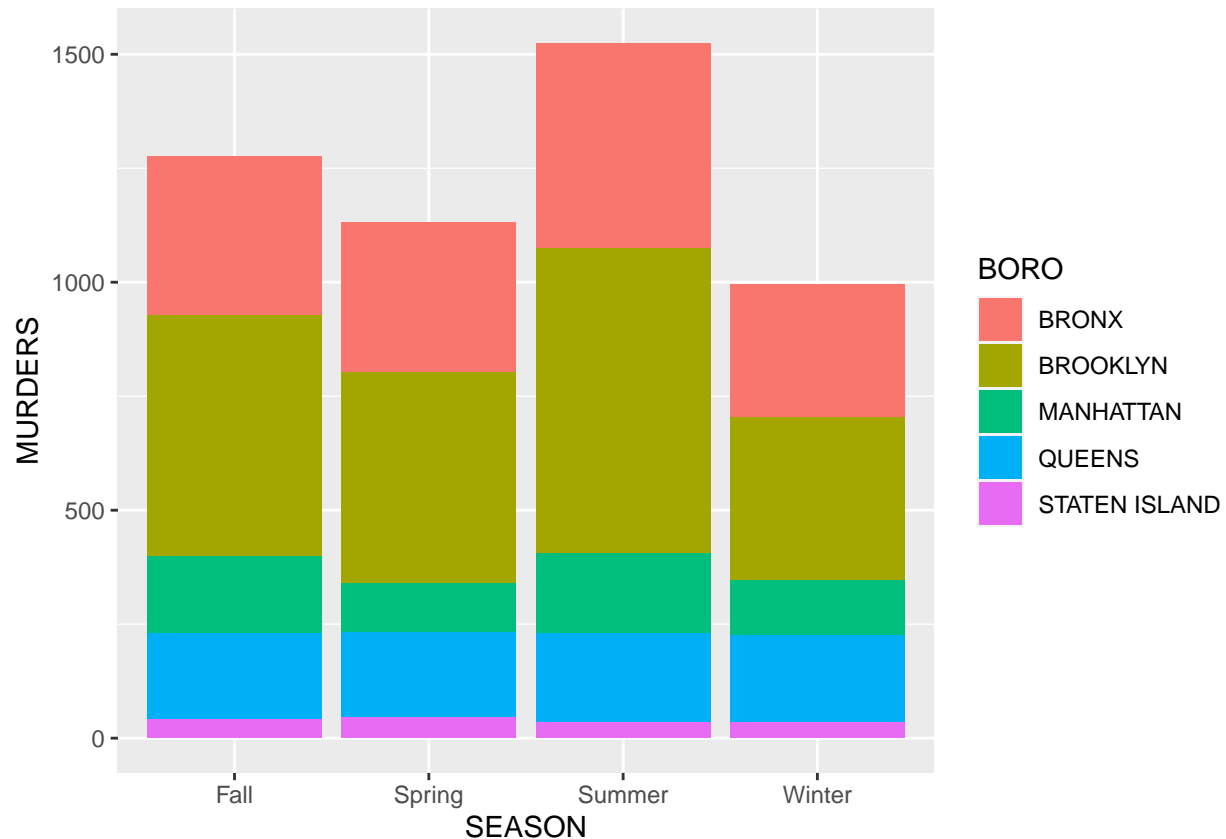


```
# `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'  
ggplot2::ggplot(data = crimes_by_time_of_day) + geom_smooth(mapping = aes(x=HOUR, y=CASES)) + geom_point
```

```
## 'geom_smooth()' using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```



```
ggplot2::ggplot(data = crimes_by_season) + geom_bar(mapping = aes(x=SEASON, y=MURDERS, fill=BORO), stat=
```



BIAS

In the data set there are columns for victim and perpetrator race, I believe these are common sources of bias, and have not included in the data for study.

```
sessionInfo()
```

```
## R version 4.2.1 (2022-06-23)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur ... 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
## [1] lubridate_1.9.0  timechange_0.1.1 knitr_1.40      forcats_0.5.2
## [5] stringr_1.4.1    dplyr_1.0.10     purrr_0.3.5     readr_2.1.3
## [9] tidyr_1.2.1      tibble_3.1.8     ggplot2_3.4.0   tidyverse_1.3.2
```

```
##
## loaded via a namespace (and not attached):
## [1] lattice_0.20-45      assertthat_0.2.1    digest_0.6.30
## [4] utf8_1.2.2           R6_2.5.1            cellranger_1.1.0
## [7] backports_1.4.1      reprex_2.0.2        evaluate_0.18
## [10] highr_0.9            httr_1.4.4          pillar_1.8.1
## [13] rlang_1.0.6          googlesheets4_1.0.1 curl_4.3.3
## [16] readxl_1.4.1         rstudioapi_0.14     Matrix_1.4-1
## [19] rmarkdown_2.17       splines_4.2.1       labeling_0.4.2
## [22] googledrive_2.0.0    bit_4.0.4           munsell_0.5.0
## [25] broom_1.0.1          compiler_4.2.1      modelr_0.1.9
## [28] xfun_0.34            pkgconfig_2.0.3     mgcv_1.8-40
## [31] htmltools_0.5.3      tidyselect_1.2.0    fansi_1.0.3
## [34] crayon_1.5.2         tzdb_0.3.0          dbplyr_2.2.1
## [37] withr_2.5.0          grid_4.2.1          nlme_3.1-157
## [40] jsonlite_1.8.3       gtable_0.3.1        lifecycle_1.0.3
## [43] DBI_1.1.3            magrittr_2.0.3      scales_1.2.1
## [46] cli_3.4.1            stringi_1.7.8       vroom_1.6.0
## [49] farver_2.1.1         fs_1.5.2            xml2_1.3.3
## [52] ellipsis_0.3.2       generics_0.1.3      vctrs_0.5.0
## [55] tools_4.2.1          bit64_4.0.5         glue_1.6.2
## [58] hms_1.1.2            parallel_4.2.1      fastmap_1.1.0
## [61] yaml_2.3.6           colorspace_2.0-3    gargle_1.2.1
## [64] rvest_1.0.3          haven_2.5.1
```