NYPD Shooting Incident Data Report

Michael Grybko

2024-02-05

Introduction

This project will analyze the NYPD Shooting Incident Data (Historic). This is a record of the shootings that have occurred in New York City from 2006 through April 27, 2023. This data includes information about the victim, the perpetrator if known, and incident location. The focus of this research will be to examine the variables associated with shootings that resulted in murder. The data set used in this project is available through DATA.GOV at https://catalog.data.gov/dataset/nypd-shooting-incident-data-historic.

Libraries

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr
              1.1.2
                        v readr
                                    2.1.4
## v forcats
              1.0.0
                                    1.5.0
                        v stringr
## v ggplot2
             3.4.2
                        v tibble
                                    3.2.1
## v lubridate 1.9.2
                        v tidyr
                                    1.3.0
## v purrr
              1.0.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()
                    masks stats::lag()
## i Use the conflicted package (<a href="http://conflicted.r-lib.org/">http://conflicted.r-lib.org/</a>) to force all conflicts to become error
## Warning: package 'writexl' was built under R version 4.2.3
```

Read in data

```
## 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.
```

Clean Data

Both date and time need to be correctly formatted, and unnecessary variables related to location and specific incident identification will be removed. The variable STATISTICAL_MURDER_FLAG, which indicates

which shootings resulted in the victim's death and are counted as murder, will be converted to a binary variable. 1 will indicate flagged for murder and 0 will indicate not flagged for murder.

```
shooting_data <- shooting_data %>%
  mutate(OCCUR DATE = mdy(OCCUR DATE)) %>%
  mutate(OCCUR_TIME = as.POSIXct(OCCUR_TIME,
                                           format = "%H:%M:%S")) %>%
  select(-c(X_COORD_CD, Y_COORD_CD, Latitude, Longitude, Lon_Lat, LOC_OF_OCCUR_DESC, LOCATION_DESC,LOC_
#transform variable STATISTICAL_MURDER_FLAG to binary 0 and 1
shooting_data$STATISTICAL_MURDER_FLAG <- ifelse(shooting_data$STATISTICAL_MURDER_FLAG == TRUE, 1, 0)
# change variables to factors
shooting_data$STATISTICAL_MURDER_FLAG <- factor(shooting_data$STATISTICAL_MURDER_FLAG)
shooting_data$VIC_RACE <- factor(shooting_data$VIC_RACE)</pre>
shooting_data$VIC_AGE_GROUP <- factor(shooting_data$VIC_AGE_GROUP)</pre>
summary(shooting_data)
      OCCUR DATE
                            OCCUR_TIME
                                                                  BORO
##
##
   \mathtt{Min}.
           :2006-01-01
                         \mathtt{Min}.
                                 :1970-01-01 00:00:00.0000
                                                              Length: 27312
##
    1st Qu.:2009-07-18
                         1st Qu.:1970-01-01 03:27:00.0000
                                                              Class : character
##
  Median :2013-04-29
                         Median :1970-01-01 15:11:00.0000
                                                              Mode :character
   Mean
           :2014-01-06
                         Mean
                                 :1970-01-01 12:41:31.7091
                         3rd Qu.:1970-01-01 20:45:00.0000
##
    3rd Qu.:2018-10-15
##
    Max.
           :2022-12-31
                         Max.
                                 :1970-01-01 23:59:00.0000
##
##
       PRECINCT
                     STATISTICAL_MURDER_FLAG PERP_AGE_GROUP
                                                                    PERP_SEX
##
    Min. : 1.00
                     0:22046
                                              Length: 27312
                                                                  Length: 27312
##
    1st Qu.: 44.00
                     1: 5266
                                                                  Class :character
                                              Class : character
   Median : 68.00
                                              Mode :character
                                                                  Mode :character
          : 65.64
##
   Mean
##
    3rd Qu.: 81.00
##
    Max. :123.00
##
                        VIC_AGE_GROUP
##
     PERP_RACE
                                          VIC_SEX
    Length: 27312
                               : 2839
                                        Length: 27312
##
                        <18
##
    Class : character
                        1022
                                        Class : character
                                    1
    Mode :character
                        18-24 :10086
                                        Mode : character
##
                       25-44 :12281
##
                        45-64 : 1863
##
                        65+
                               : 181
##
                       UNKNOWN:
                                   61
                               VIC RACE
##
##
    AMERICAN INDIAN/ALASKAN NATIVE:
##
  ASIAN / PACIFIC ISLANDER
                                      404
## BLACK
                                   :19439
## BLACK HISPANIC
                                   : 2646
## UNKNOWN
                                       66
## WHITE
                                      698
## WHITE HISPANIC
                                   : 4049
```

Missing Data

There are some missing data entries that need to be addressed. Some of these entries related to perpetrator are not filled in at all and some contain the UNKOWN or U.

```
## Missing values in PERP_AGE_GROUP: NA
## Missing values in PERP_SEX: NA
## Missing values in PERP_RACE: NA
## PERP_AGE_GROUP labeled UNKOWN: NA
## PERP_SEX labeled U: NA
## PERP_RACE labeled UNKOWN: NA
```

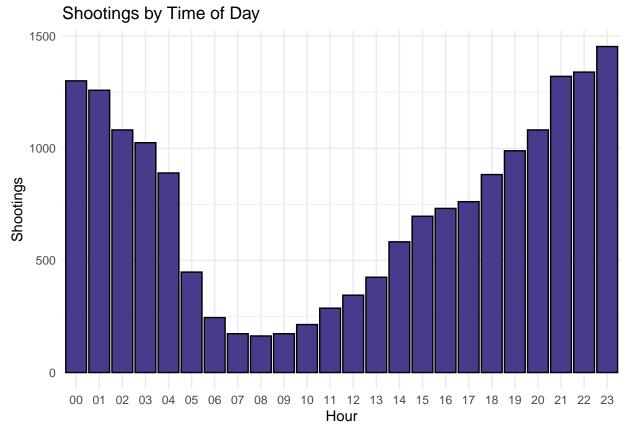
Handeling Missing Data

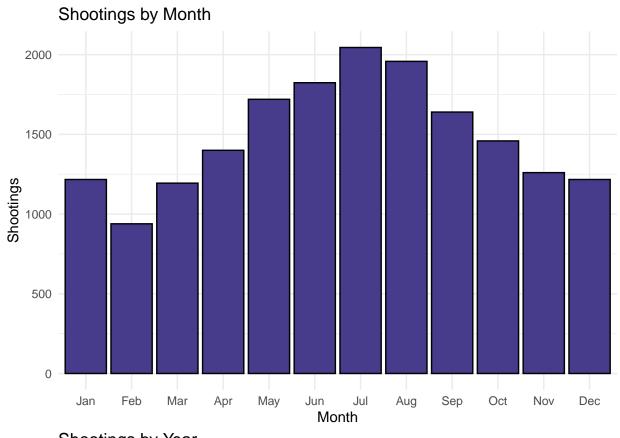
Empty data cells or cells containing(null), for the variables related to perpetrator will be filled in with UNKNOWN, and for consistency the entries of U for the variable PERP_SEX will be converted to UNKNOWN. Entries of UNKNOWN will not be imputed for the variables related to perpetrator because it is unclear what they should be imputed to. For example, some of these may indicate the perpetrator was never seen and/or caught, which may be of statistical importance as is, while other shootings may be self-inflected. Since there is only one instance each of "1020", "940", and "224" in the variable PER_AGE_GROUP they will be dropped. Shooting incidents that include entries of UNKNOWN for VIC_RACE and VIC_AGE_GROUP will be dropped. There are relatively few of these entries (66 and 61) and dropping them should not impact statistical analysis and modeling. VIC_AGE_GROUP also contains one entry of 1022 that will be removed. There are only a few instances with the gender of the victim is "U", so they will be dropped.

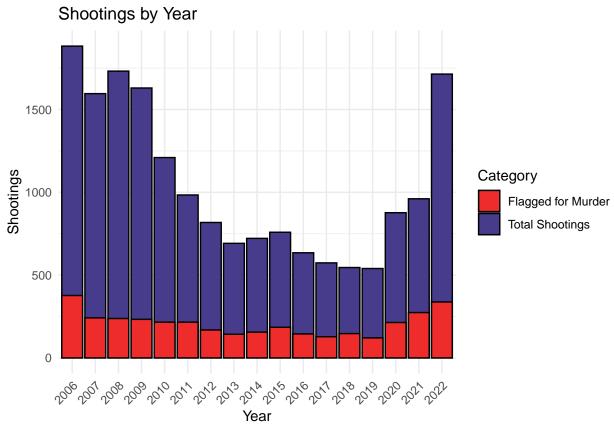
```
shooting_data <- shooting_data %>%
  mutate(PERP_AGE_GROUP = ifelse(PERP_AGE_GROUP == "", "UNKNOWN", PERP_AGE_GROUP),
         PERP_AGE_GROUP = ifelse(PERP_AGE_GROUP == "(null)", "UNKNOWN", PERP_AGE_GROUP),
         PERP RACE = ifelse(PERP_RACE == "", "UNKNOWN", PERP_RACE),
         PERP_RACE = ifelse(PERP_RACE == "(null)", "UNKNOWN", PERP_RACE),
         PERP_SEX = ifelse(PERP_SEX == "U", "UNKNOWN", ifelse(PERP_SEX == "", "UNKNOWN", PERP_SEX)),
         PERP_SEX = ifelse(PERP_SEX == "(null)", "UNKNOWN", PERP_SEX)) %>%
  filter(VIC RACE != "UNKNOWN" & VIC AGE GROUP != "UNKNOWN" & VIC AGE GROUP != "1022" & PERP AGE GROUP
         & PERP AGE GROUP != "224" & PERP AGE GROUP != "940"
         & VIC SEX != "U")
shooting data$VIC RACE <- factor(shooting data$VIC RACE, exclude = "UNKNOWN")
shooting_data$VIC_AGE_GROUP <- factor(shooting_data$VIC_AGE_GROUP, exclude = "UNKNOWN")</pre>
shooting_data$VIC_AGE_GROUP <- factor(shooting_data$VIC_AGE_GROUP, exclude = "1022")</pre>
shooting_data$VIC_SEX <- factor(shooting_data$VIC_SEX, exclude = "U")</pre>
shooting_data$PERP_AGE_GROUP <- factor(shooting_data$PERP_AGE_GROUP, exclude = c("1020", "224", "940"))
summary_data <- summary(shooting_data)</pre>
summary_data <- as.data.frame(summary_data)</pre>
excel file <- "/Users/michaelgrybko/Desktop/MSDS/DataScienceAsaField/Week5/summary shooting data.xlsx"
write_xlsx(list(summary_data = summary_data), excel_file)
```

Data Exploration

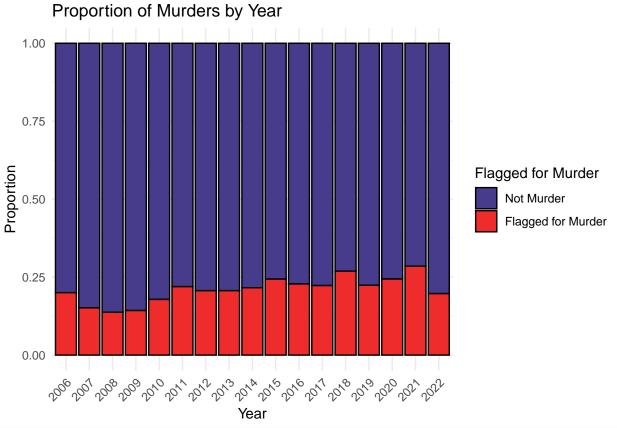
Graphs Showing Relationship of Shootings and Time

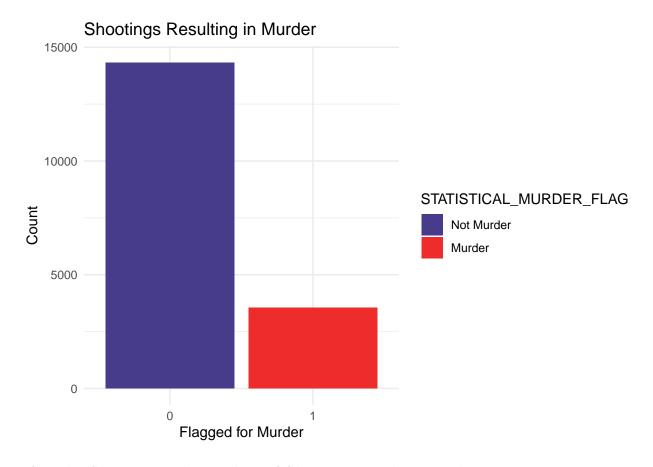






`summarise()` has grouped output by 'year'. You can override using the
`.groups` argument.

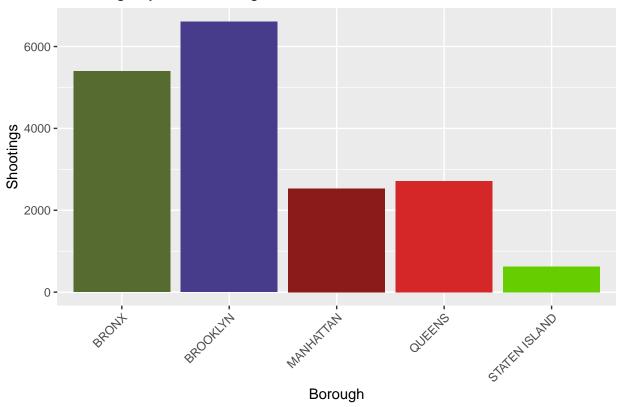




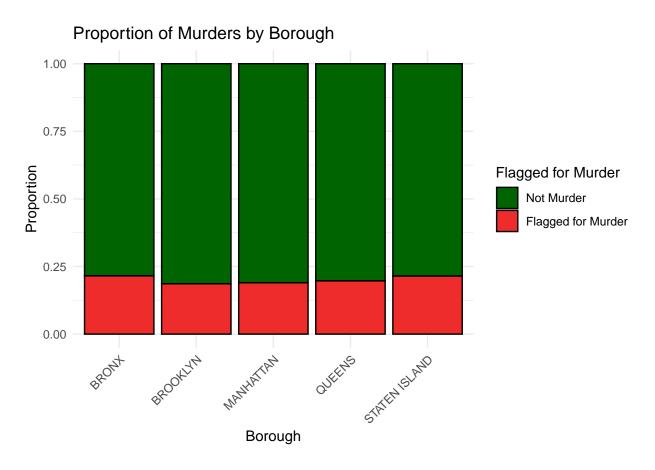
Graphs Showing Relationship of Shootings and Boroughs

- ## Warning: The `<scale>` argument of `guides()` cannot be `FALSE`. Use "none" instead as
 ## of ggplot2 3.3.4.
- ## This warning is displayed once every 8 hours.
- ## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
- ## generated.

Shootings by NYC Borough



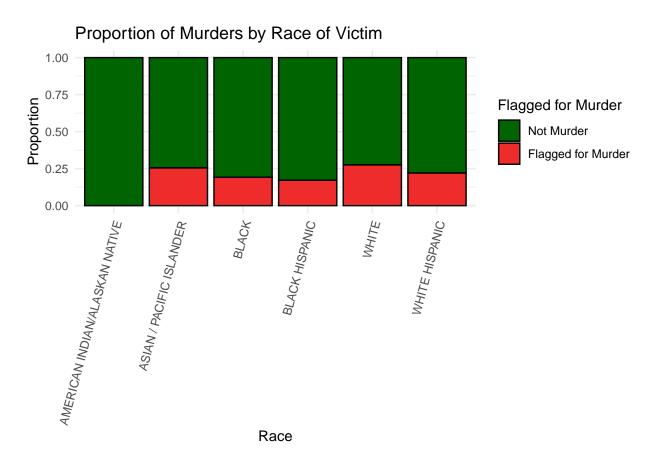
 $\mbox{\tt \#\# `summarise()` has grouped output by 'BORO'. You can override using the $\mbox{\tt \#\# `.groups` argument.}$}$



Relationship of Shootings and Attributes of Victims

These plots use data normalized showing the murder rate as a percentage of total shootings.

 $\mbox{\tt ## `summarise()` has grouped output by 'VIC_RACE'. You can override using the <math display="inline">\mbox{\tt ## `.groups` argument.}$



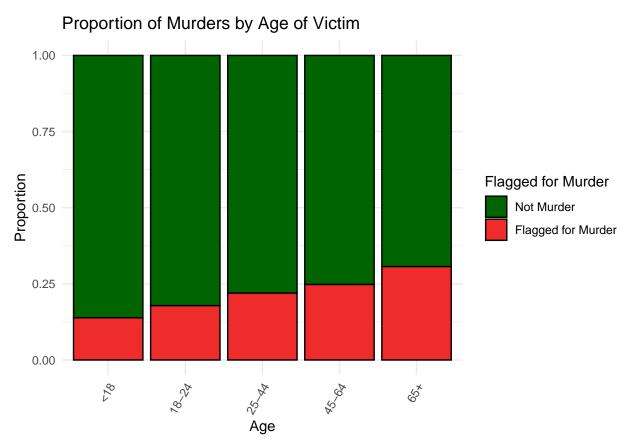
 $\mbox{\tt \#\# `summarise()` has grouped output by 'VIC_SEX'. You can override using the <math display="inline">\mbox{\tt \#\# `.groups` argument.}$



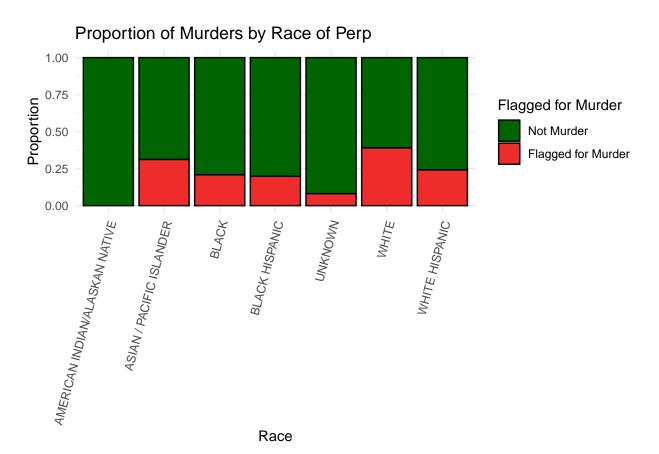
Relationship of Shootings and Attributes of Perpetrators

These plots use data normalized showing the murder rate as a percentage of total shootings.

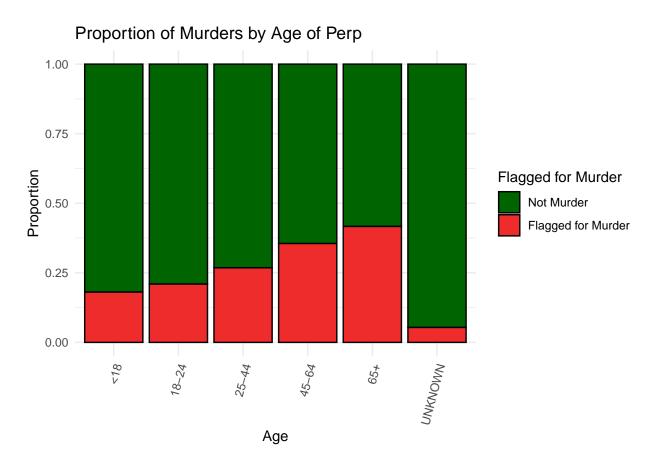
`summarise()` has grouped output by 'VIC_AGE_GROUP'. You can override using the ## `.groups` argument.



`summarise()` has grouped output by 'PERP_RACE'. You can override using the
`.groups` argument.



 $\mbox{\tt ## `summarise()` has grouped output by 'PERP_AGE_GROUP'. You can override using <math display="inline">\mbox{\tt ## the `.groups` argument.}$



Statistical Analysis and Models

Full Logistical Regression Model

Here the response variable is STATISTICAL_MURDER_FLAG, which has been converted to a binary factor with 1 indicating a shooting that has been flagged for murder and 0 a shooting that did not result in murder. The first model is a full logistical regression model including all predictors.

```
log_reg_mod_all <- glm(STATISTICAL_MURDER_FLAG ~ ., data = shooting_data, family = "binomial")</pre>
summary(log_reg_mod_all)
##
## Call:
  glm(formula = STATISTICAL_MURDER_FLAG ~ ., family = "binomial",
##
       data = shooting_data)
##
  Deviance Residuals:
##
                      Median
##
       Min
                 1Q
                                   3Q
                                           Max
            -0.7322 -0.6034 -0.1938
                                         3.0007
## -1.6871
##
## Coefficients:
                                       Estimate Std. Error z value Pr(>|z|)
##
## (Intercept)
                                      4.537e+00 2.543e+02
                                                              0.018 0.985765
## OCCUR_DATE
                                     -2.092e-03 2.190e-03 -0.955 0.339428
## OCCUR TIME
                                      5.861e-05 1.865e-05
                                                              3.143 0.001675 **
## BOROBROOKLYN
                                     -1.158e-01 1.086e-01 -1.067 0.286125
## BOROMANHATTAN
                                     -2.255e-01 9.157e-02 -2.463 0.013771 *
```

```
## BOROQUEENS
                                    -7.674e-02 2.126e-01 -0.361 0.718098
                                    -1.291e-01 2.688e-01 -0.480 0.631037
## BOROSTATEN ISLAND
## PRECINCT
                                   -1.196e-03 3.245e-03 -0.369 0.712468
## PERP_AGE_GROUP18-24
                                    1.244e-01 7.422e-02
                                                           1.676 0.093747
## PERP_AGE_GROUP25-44
                                    3.614e-01
                                               7.574e-02
                                                           4.771 1.83e-06 ***
## PERP AGE GROUP45-64
                                               1.130e-01
                                                           6.163 7.12e-10 ***
                                   6.967e-01
## PERP AGE GROUP65+
                                   8.756e-01 2.859e-01
                                                           3.062 0.002198 **
## PERP AGE GROUPUNKNOWN
                                  -2.494e+00 1.797e-01 -13.882 < 2e-16 ***
## PERP SEXM
                                   -1.606e-01
                                               1.152e-01 -1.394 0.163170
## PERP_SEXUNKNOWN
                                    1.873e+00 2.785e-01
                                                           6.725 1.76e-11 ***
## PERP_RACEASIAN / PACIFIC ISLANDER 1.205e+01 2.294e+02
                                                           0.053 0.958097
## PERP_RACEBLACK
                                     1.168e+01
                                               2.294e+02
                                                           0.051 0.959379
## PERP_RACEBLACK HISPANIC
                                     1.156e+01 2.294e+02
                                                           0.050 0.959815
                                               2.294e+02
## PERP_RACEUNKNOWN
                                    1.115e+01
                                                           0.049 0.961240
## PERP_RACEWHITE
                                    1.225e+01
                                               2.294e+02
                                                           0.053 0.957425
## PERP_RACEWHITE HISPANIC
                                    1.176e+01
                                               2.294e+02
                                                           0.051 0.959118
## VIC_AGE_GROUP18-24
                                     2.522e-01
                                               7.508e-02
                                                           3.359 0.000782 ***
## VIC AGE GROUP25-44
                                               7.478e-02
                                                           4.867 1.13e-06 ***
                                   3.640e-01
## VIC_AGE_GROUP45-64
                                     3.892e-01 9.777e-02
                                                           3.981 6.86e-05 ***
## VIC AGE GROUP65+
                                     6.716e-01
                                               2.096e-01
                                                           3.204 0.001354 **
## VIC_SEXM
                                    -7.263e-02 6.133e-02 -1.184 0.236286
## VIC RACEASIAN / PACIFIC ISLANDER 1.043e+01 1.058e+02
                                                          0.099 0.921432
                                                           0.097 0.922656
## VIC_RACEBLACK
                                     1.027e+01 1.058e+02
## VIC RACEBLACK HISPANIC
                                    1.005e+01
                                               1.058e+02
                                                           0.095 0.924309
## VIC RACEWHITE
                                   1.029e+01 1.058e+02
                                                           0.097 0.922501
## VIC RACEWHITE HISPANIC
                                   1.033e+01 1.058e+02
                                                           0.098 0.922223
## hour01
                                    -2.959e-01
                                               1.251e-01 -2.364 0.018056 *
## hour02
                                    -5.984e-01
                                               1.745e-01 -3.429 0.000605 ***
## hour03
                                    -6.498e-01 2.327e-01 -2.792 0.005239 **
## hour04
                                    -6.770e-01 2.871e-01 -2.358 0.018372 *
## hour05
                                    -7.245e-01
                                               3.573e-01 -2.027 0.042614 *
## hour06
                                    -1.054e+00 4.365e-01 -2.416 0.015690 *
## hour07
                                    -1.096e+00 5.074e-01 -2.160 0.030805 *
## hour08
                                    -1.645e+00 5.772e-01
                                                          -2.850 0.004373 **
## hour09
                                    -1.871e+00
                                               6.363e-01
                                                          -2.941 0.003276 **
## hour10
                                    -2.097e+00 6.975e-01 -3.007 0.002638 **
## hour11
                                    -2.251e+00 7.591e-01 -2.965 0.003026 **
## hour12
                                    -2.629e+00 8.252e-01 -3.186 0.001443 **
## hour13
                                    -2.688e+00 8.853e-01 -3.036 0.002397 **
                                    -2.987e+00 9.516e-01 -3.139 0.001697 **
## hour14
## hour15
                                    -3.325e+00 1.015e+00 -3.275 0.001055 **
## hour16
                                    -3.741e+00 1.083e+00 -3.455 0.000550 ***
## hour17
                                    -3.592e+00
                                               1.150e+00 -3.122 0.001793 **
## hour18
                                    -3.665e+00 1.213e+00 -3.022 0.002509 **
## hour19
                                    -4.112e+00 1.283e+00 -3.205 0.001352 **
                                               1.347e+00 -3.249 0.001160 **
## hour20
                                    -4.375e+00
## hour21
                                    -4.357e+00
                                               1.413e+00 -3.083 0.002047 **
## hour22
                                    -4.509e+00
                                               1.481e+00 -3.044 0.002335 **
## hour23
                                    -5.009e+00
                                               1.549e+00
                                                          -3.233 0.001225 **
## month02
                                    5.482e-02
                                               1.311e-01
                                                           0.418 0.675814
## month03
                                   -2.314e-02 1.705e-01 -0.136 0.892065
## month04
                                    1.950e-01 2.235e-01
                                                          0.872 0.382937
## month05
                                    3.706e-01 2.841e-01
                                                           1.305 0.192046
## month06
                                    1.883e-01 3.476e-01
                                                           0.542 0.587980
```

```
## month07
                                      2.979e-01 4.091e-01
                                                             0.728 0.466464
## month08
                                      3.486e-01 4.772e-01
                                                             0.731 0.465046
## month09
                                      5.847e-01 5.420e-01
                                                             1.079 0.280696
## month10
                                      5.463e-01 6.077e-01
                                                             0.899 0.368705
## month11
                                      6.168e-01
                                                 6.754e-01
                                                             0.913 0.361138
## month12
                                      8.987e-01
                                                7.419e-01
                                                             1.211 0.225761
## year2007
                                      3.289e-01 8.057e-01
                                                             0.408 0.683111
## year2008
                                      9.439e-01
                                                1.604e+00
                                                             0.588 0.556318
## year2009
                                      1.799e+00
                                                 2.403e+00
                                                             0.749 0.454147
## year2010
                                      2.732e+00
                                                 3.202e+00
                                                             0.853 0.393532
## year2011
                                      3.510e+00
                                                 3.999e+00
                                                             0.878 0.380138
## year2012
                                      4.088e+00
                                                 4.801e+00
                                                             0.851 0.394572
## year2013
                                      4.858e+00
                                                 5.600e+00
                                                             0.867 0.385735
## year2014
                                      5.679e+00
                                                 6.398e+00
                                                             0.888 0.374743
## year2015
                                      6.553e+00
                                                 7.200e+00
                                                             0.910 0.362691
## year2016
                                      7.217e+00
                                                 8.000e+00
                                                             0.902 0.367004
## year2017
                                      7.921e+00
                                                 8.800e+00
                                                             0.900 0.368082
## year2018
                                      8.921e+00
                                                 9.598e+00
                                                             0.929 0.352672
## year2019
                                      9.441e+00
                                                1.040e+01
                                                             0.908 0.363933
## year2020
                                      1.032e+01
                                                 1.120e+01
                                                             0.921 0.356957
## year2021
                                      1.133e+01
                                                1.200e+01
                                                             0.944 0.344949
## year2022
                                      1.195e+01 1.280e+01
                                                             0.934 0.350415
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
##
##
  (Dispersion parameter for binomial family taken to be 1)
##
##
       Null deviance: 17797
                             on 17872 degrees of freedom
## Residual deviance: 16502
                             on 17792 degrees of freedom
## AIC: 16664
## Number of Fisher Scoring iterations: 11
```

Reduced Logistical Regression Model

Coefficients:

VIC_AGE_GROUP18-24

(Intercept)

##

The second model is a reduced logistical regression model. STATISTICAL_MURDER_FLAG is still the response variable, however only the predictor variables with significant p-values in the full model were included.

```
log_reg_mod_red <- glm(STATISTICAL_MURDER_FLAG ~ VIC_AGE_GROUP + PERP_AGE_GROUP + PERP_SEX + hour, data
summary(log_reg_mod_red)
##
## Call:
## glm(formula = STATISTICAL_MURDER_FLAG ~ VIC_AGE_GROUP + PERP_AGE_GROUP +
##
       PERP SEX + hour, family = "binomial", data = shooting data)
##
## Deviance Residuals:
##
       Min
                 1Q
                      Median
                                    30
                                            Max
##
  -1.4863
           -0.7430 -0.6379
                              -0.1977
                                         2.9654
##
```

0.157633 -9.964 < 2e-16 ***

3.406 0.000660 ***

Estimate Std. Error z value Pr(>|z|)

0.074372

-1.570656

0.253300

```
## VIC AGE GROUP25-44
                           0.369161
                                       0.073788
                                                  5.003 5.64e-07 ***
## VIC_AGE_GROUP45-64
                           0.416951
                                       0.096526
                                                  4.320 1.56e-05 ***
## VIC AGE GROUP65+
                           0.721360
                                       0.206137
                                                  3.499 0.000466 ***
## PERP_AGE_GROUP18-24
                           0.115007
                                       0.073689
                                                  1.561 0.118590
## PERP AGE GROUP25-44
                           0.378682
                                       0.075019
                                                  5.048 4.47e-07 ***
## PERP AGE GROUP45-64
                           0.751212
                                       0.111111
                                                  6.761 1.37e-11 ***
## PERP AGE GROUP65+
                           0.911892
                                       0.276265
                                                  3.301 0.000964 ***
## PERP AGE GROUPUNKNOWN -2.509343
                                       0.175641 -14.287 < 2e-16 ***
## PERP SEXM
                          -0.175115
                                       0.113984
                                                 -1.536 0.124461
## PERP_SEXUNKNOWN
                           1.363711
                                       0.207707
                                                  6.566 5.18e-11 ***
## hour01
                          -0.064753
                                       0.105568
                                                 -0.613 0.539623
## hour02
                          -0.178820
                                       0.112492
                                                 -1.590 0.111919
## hour03
                           0.003698
                                       0.111334
                                                  0.033 0.973500
## hour04
                           0.161068
                                       0.112360
                                                  1.434 0.151714
## hour05
                           0.335389
                                       0.133524
                                                  2.512 0.012011 *
## hour06
                           0.222264
                                       0.168568
                                                  1.319 0.187323
## hour07
                                       0.190292
                                                  2.087 0.036858 *
                           0.397202
## hour08
                           0.141650
                                       0.199167
                                                  0.711 0.476953
## hour09
                           0.087267
                                       0.197399
                                                  0.442 0.658427
## hour10
                           0.067611
                                       0.181816
                                                  0.372 0.709993
## hour11
                           0.155104
                                       0.159917
                                                  0.970 0.332096
                          -0.041999
## hour12
                                       0.153220
                                                 -0.274 0.784000
## hour13
                           0.085625
                                       0.139818
                                                  0.612 0.540271
## hour14
                          -0.005432
                                       0.127529
                                                 -0.043 0.966025
## hour15
                          -0.096190
                                       0.124482
                                                 -0.773 0.439684
## hour16
                          -0.338707
                                       0.127604
                                                 -2.654 0.007946
                           0.076127
## hour17
                                       0.116734
                                                  0.652 0.514311
## hour18
                           0.168385
                                       0.110734
                                                  1.521 0.128353
## hour19
                          -0.033973
                                       0.110450
                                                 -0.308 0.758398
## hour20
                          -0.100795
                                       0.109559
                                                 -0.920 0.357572
## hour21
                           0.117112
                                       0.101884
                                                  1.149 0.250364
## hour22
                           0.164338
                                       0.100365
                                                  1.637 0.101548
## hour23
                          -0.129423
                                       0.102768
                                                 -1.259 0.207895
##
                   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
  Signif. codes:
##
##
   (Dispersion parameter for binomial family taken to be 1)
##
       Null deviance: 17797
##
                              on 17872
                                         degrees of freedom
  Residual deviance: 16683
                              on 17838
                                         degrees of freedom
  AIC: 16753
##
## Number of Fisher Scoring iterations: 6
```

Exploration of Bias and Conclusion

It is very important to be cognizant of the potential for bias to impact the analysis of data such as this. There can be bias in the collection and formatting of the data before it is analyzed, and the cleaning and modeling of the data can be impacted by personal bias. One reason for this is because this dataset is documenting violent crime, which can be an emotional topic. Also, in the United States the topic of gun ownership is very contentious. Some believe gun ownership is a fundamental right of an American citizen, while others believe stricter limits on gun ownership need to be put into place to help prevent violent crime.

The variable STATISTICAL_MURDER_FLAG could be an indication of bias in this dataset. This factor

only documents shootings that resulted in murder, however there was no information about other causes of shootings. For instance, were any of these shootings defensive, accidental, self-inflicted or police related? By only flagging murders and not including other reasons for shootings we are inclined to interpret this data in a certain way, introducing a bias. This is part of my reasoning behind not imputing or removing shooting instances when there was missing data for the perpetrator. I do not know how to interpret an "UNKNOWN" perpetrator. I am generally in favor of stricter gun laws, and I could have introduced bias into the data by trying to impute missing data in attributes associated with the perpetrator.

I found it interesting that in both logistical regression models the age of both the victim and the perpetrator were statistically significant when examining which shootings were flagged as murders. There were not very many other impactful attributes. Certain hours of the day had some significance. The only other factors that stood out as statistically significant were PERP_SEX "UNKNOWN" and PERP_AGE_GROUP "UNKNOWN", with PERP_SEX "UNKNOWN" having a positive correlation and PERP_AGE_GROUP "UNKNOWN" having a negative correlation. This could be an interesting avenue for further exploration.

Session Information

sessionInfo()

```
## R version 4.2.2 (2022-10-31)
## Platform: aarch64-apple-darwin20 (64-bit)
## Running under: macOS 14.2.1
##
## Matrix products: default
           /Library/Frameworks/R.framework/Versions/4.2-arm64/Resources/lib/libRblas.0.dylib
## BLAS:
## LAPACK: /Library/Frameworks/R.framework/Versions/4.2-arm64/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] writexl 1.5.0
                        lubridate 1.9.2 forcats 1.0.0
                                                         stringr 1.5.0
##
    [5] dplyr_1.1.2
                                         readr_2.1.4
                                                         tidyr_1.3.0
##
                        purrr_1.0.1
   [9] tibble_3.2.1
                        ggplot2_3.4.2
                                         tidyverse_2.0.0
##
##
## loaded via a namespace (and not attached):
   [1] highr_0.10
                         pillar_1.9.0
##
                                           compiler_4.2.2
                                                            tools_4.2.2
    [5] bit_4.0.5
                         digest_0.6.31
                                           timechange_0.2.0 evaluate_0.20
##
##
   [9] lifecycle_1.0.3
                         gtable_0.3.3
                                           pkgconfig_2.0.3
                                                            rlang_1.1.1
## [13] cli_3.6.1
                         rstudioapi_0.14
                                           curl_5.0.0
                                                            parallel_4.2.2
  [17] yaml_2.3.7
                                           fastmap_1.1.1
                                                            withr_2.5.0
                         xfun_0.39
  [21] knitr_1.42
                         generics_0.1.3
                                           vctrs_0.6.2
                                                            hms_1.1.3
##
  [25] bit64_4.0.5
                         grid_4.2.2
                                           tidyselect_1.2.0 glue_1.6.2
## [29] R6_2.5.1
                         fansi_1.0.4
                                           vroom_1.6.3
                                                            rmarkdown_2.21
                                           magrittr_2.0.3
                                                            scales_1.2.1
  [33] farver_2.1.1
                         tzdb_0.4.0
  [37] htmltools_0.5.5
                         colorspace_2.1-0 labeling_0.4.2
                                                            utf8_1.2.3
## [41] stringi 1.7.12
                         munsell 0.5.0
                                           crayon 1.5.2
```