NYPD Shooting Incident Data

Eric Schaffer

2024-09-08

Abstract

List of every shooting incident that occurred in NYC going back to 2006 through the end of the previous calendar year.

This 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. In addition, information related to suspect and victim demographics is also included. This data can be used by the public to explore the nature of shooting/criminal activity.

Import NYPD Shooting Incident Data

```
# import the data and conver the missing values to NA
raw_data <- read.csv("https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD", r
```

Inspect Data

Intial view of the raw data to determine data shape and quality for the tidying step

```
head(raw_data)
```

```
##   INCIDENT_KEY OCCUR_DATE OCCUR_TIME      BORO LOC_OF_OCCUR_DESC PRECINCT
## 1    244608249 05/05/2022   00:10:00 MANHATTAN            INSIDE       14
## 2    247542571 07/04/2022   22:20:00     BRONX           OUTSIDE       48
## 3     84967535 05/27/2012   19:35:00    QUEENS              <NA>      103
## 4    202853370 09/24/2019   21:00:00     BRONX              <NA>       42
## 5     27078636 02/25/2007   21:00:00  BROOKLYN              <NA>       83
## 6    230311078 07/01/2021   23:07:00 MANHATTAN              <NA>       23
##   JURISDICTION_CODE LOC_CLASSFCTN_DESC             LOCATION_DESC
## 1                 0         COMMERCIAL               VIDEO STORE
## 2                 0             STREET                      <NA>
## 3                 0               <NA>                      <NA>
## 4                 0               <NA>                      <NA>
## 5                 0               <NA>                      <NA>
## 6                 2               <NA> MULTI DWELL - PUBLIC HOUS
##   STATISTICAL_MURDER_FLAG PERP_AGE_GROUP PERP_SEX PERP_RACE VIC_AGE_GROUP
```

```
## 1                   true          25-44        M     BLACK         25-44
## 2                   true           <NA>     <NA>      <NA>         18-24
## 3                  false           <NA>     <NA>      <NA>         18-24
## 4                  false          25-44        M      <NA>         25-44
## 5                  false          25-44        M     BLACK         25-44
## 6                  false           <NA>     <NA>      <NA>         25-44
##   VIC_SEX VIC_RACE X_COORD_CD Y_COORD_CD Latitude Longitude
## 1       M    BLACK     986050   214231.0 40.75469 -73.99350
## 2       M    BLACK    1016802   250581.0 40.85440 -73.88233
## 3       M    BLACK    1048632   198262.0 40.71063 -73.76777
## 4       M    BLACK    1014493   242565.0 40.83242 -73.89071
## 5       M    BLACK    1009149   190104.7 40.68844 -73.91022
## 6       M    BLACK     999061   229912.0 40.79773 -73.94651
##                                     Lon_Lat
## 1                    POINT (-73.9935 40.754692)
## 2                   POINT (-73.88233 40.854402)
## 3  POINT (-73.76777349199995 40.71063412500007)
## 4 POINT (-73.89071440599997 40.832416753000075)
## 5  POINT (-73.91021857399994 40.68844345900004)
## 6  POINT (-73.94650786199998 40.79772716600007)
```

```
summary(raw_data)
```

```
##   INCIDENT_KEY        OCCUR_DATE         OCCUR_TIME            BORO
##  Min.   :  9953245   Length:28562       Length:28562       Length:28562
##  1st Qu.: 65439914   Class :character   Class :character   Class :character
##  Median : 92711254   Mode  :character   Mode  :character   Mode  :character
##  Mean   :127405824
##  3rd Qu.:203131993
##  Max.   :279758069
##
##  LOC_OF_OCCUR_DESC     PRECINCT     JURISDICTION_CODE LOC_CLASSFCTN_DESC
##  Length:28562       Min.   :  1.0   Min.   :0.0000    Length:28562
##  Class :character   1st Qu.: 44.0   1st Qu.:0.0000    Class :character
##  Mode  :character   Median : 67.0   Median :0.0000    Mode  :character
##                     Mean   : 65.5   Mean   :0.3219
##                     3rd Qu.: 81.0   3rd Qu.:0.0000
##                     Max.   :123.0   Max.   :2.0000
##                                     NA's   :2
##  LOCATION_DESC      STATISTICAL_MURDER_FLAG PERP_AGE_GROUP
##  Length:28562       Length:28562            Length:28562
##  Class :character   Class :character        Class :character
##  Mode  :character   Mode  :character        Mode  :character
##
##
##
##
##    PERP_SEX           PERP_RACE         VIC_AGE_GROUP        VIC_SEX
##  Length:28562       Length:28562       Length:28562       Length:28562
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
```

```
##
##    VIC_RACE           X_COORD_CD        Y_COORD_CD        Latitude
##  Length:28562       Min.   : 914928   Min.   :125757   Min.   :40.51
##  Class :character   1st Qu.:1000068   1st Qu.:182912   1st Qu.:40.67
##  Mode  :character   Median :1007772   Median :194901   Median :40.70
##                     Mean   :1009424   Mean   :208380   Mean   :40.74
##                     3rd Qu.:1016807   3rd Qu.:239814   3rd Qu.:40.82
##                     Max.   :1066815   Max.   :271128   Max.   :40.91
##                                                        NA's   :59
##    Longitude         Lon_Lat
##  Min.   :-74.25   Length:28562
##  1st Qu.:-73.94   Class :character
##  Median :-73.92   Mode  :character
##  Mean   :-73.91
##  3rd Qu.:-73.88
##  Max.   :-73.70
##  NA's   :59
```

Tidy Data

To tidy the NYPD Shooting data set I first converted the date and time columns into a date time column, converted the murder flag column into a logical boolean column, and created factors for the categorical columns. I removed the columns that had non-categorical or numerical data that would be difficult to analyze or was missing a lot of data like the location description and classification columns. Finally, I removed the small number (~60) rows that were missing either location data or a jurisdiction code. As part of this tidying I notices that the classifcation data for the perpetrator has a lot of missing data compared to the victims. I attribute this to the fact that the perpetrator may not have been aprehended or present at the scene when the report was captured. Any analysis on the perpetrators will need to highlight that the majority of the data is missing.

```
NYPD_Data <- raw_data %>%
  # convert the date and time columns into a single datetime column
  mutate(OCCUR_DATE = mdy(OCCUR_DATE)) %>%
  unite(OCCUR_DATETIME, sep = " ", OCCUR_DATE:OCCUR_TIME) %>%
  mutate(OCCUR_DATETIME = as_datetime(OCCUR_DATETIME)) %>%
  # convert the murder flag to a boolean
  mutate(STATISTICAL_MURDER_FLAG = as.logical(STATISTICAL_MURDER_FLAG))%>%
  # remove columns that have data non-categorical or numerical data
  subset(select = -c(LOC_OF_OCCUR_DESC, LOCATION_DESC, LOC_CLASSFCTN_DESC, X_COORD_CD, Y_COORD_CD)) %>%
  # remove rows that don't have location data or a jurisdiction code (there are only a small number of
  filter(!is.na(Latitude)) %>%
  filter(!is.na(JURISDICTION_CODE)) %>%
  # Create factors for categorical data
  mutate(BORO = as.factor(BORO)) %>%
  mutate(PRECINCT = as.factor(PRECINCT)) %>%
  mutate(PERP_AGE_GROUP = factor(PERP_AGE_GROUP, levels = c("<18", "18-24", "25-44", "45-64", "65+"))) %
  mutate(VIC_AGE_GROUP =  factor(VIC_AGE_GROUP, levels = c("<18", "18-24", "25-44", "45-64", "65+"))) %>
  mutate(PERP_SEX = as.factor(PERP_SEX)) %>%
  mutate(PERP_RACE = as.factor(PERP_RACE)) %>%
  mutate(VIC_SEX = as.factor(VIC_SEX)) %>%
  mutate(VIC_RACE = as.factor(VIC_RACE)) %>%
  mutate(JURISDICTION_CODE = as.factor(JURISDICTION_CODE))
```

```
head(NYPD_Data)
```

```
##   INCIDENT_KEY      OCCUR_DATETIME      BORO PRECINCT JURISDICTION_CODE
## 1    244608249 2022-05-05 00:10:00 MANHATTAN       14                 0
## 2    247542571 2022-07-04 22:20:00     BRONX       48                 0
## 3     84967535 2012-05-27 19:35:00    QUEENS      103                 0
## 4    202853370 2019-09-24 21:00:00     BRONX       42                 0
## 5     27078636 2007-02-25 21:00:00  BROOKLYN       83                 0
## 6    230311078 2021-07-01 23:07:00 MANHATTAN       23                 2
##   STATISTICAL_MURDER_FLAG PERP_AGE_GROUP PERP_SEX PERP_RACE VIC_AGE_GROUP
## 1                    TRUE          25-44        M     BLACK         25-44
## 2                    TRUE           <NA>     <NA>      <NA>         18-24
## 3                   FALSE           <NA>     <NA>      <NA>         18-24
## 4                   FALSE          25-44        M      <NA>         25-44
## 5                   FALSE          25-44        M     BLACK         25-44
## 6                   FALSE           <NA>     <NA>      <NA>         25-44
##   VIC_SEX VIC_RACE Latitude Longitude
## 1       M    BLACK 40.75469 -73.99350
## 2       M    BLACK 40.85440 -73.88233
## 3       M    BLACK 40.71063 -73.76777
## 4       M    BLACK 40.83242 -73.89071
## 5       M    BLACK 40.68844 -73.91022
## 6       M    BLACK 40.79773 -73.94651
##                                      Lon_Lat
## 1                    POINT (-73.9935 40.754692)
## 2                   POINT (-73.88233 40.854402)
## 3  POINT (-73.76777349199995 40.71063412500007)
## 4 POINT (-73.89071440599997 40.832416753000075)
## 5  POINT (-73.91021857399994 40.68844345900004)
## 6  POINT (-73.94650786199998 40.79772716600007)
```

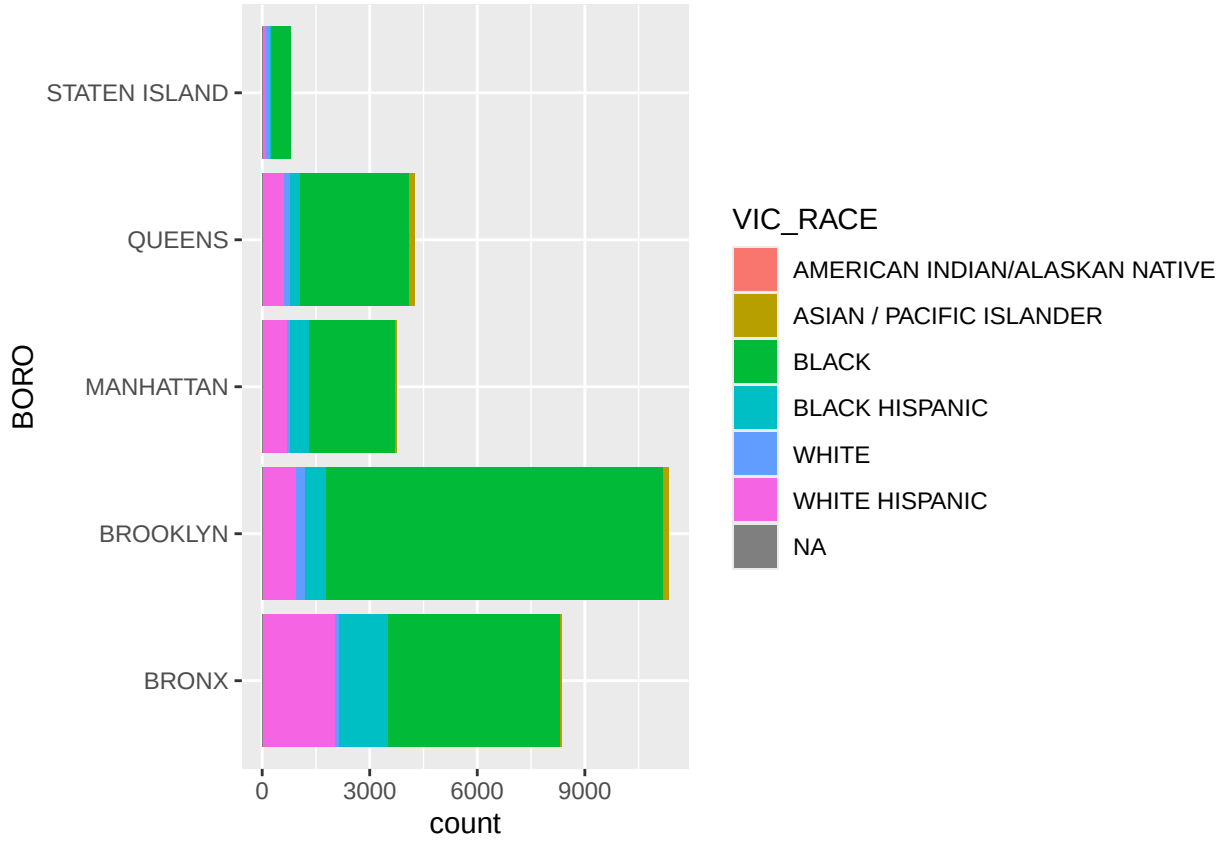
```
summary(NYPD_Data)
```

```
##   INCIDENT_KEY       OCCUR_DATETIME                                BORO
##  Min.   :  9953245   Min.   :2006-01-01 02:00:00.00   BRONX        : 8363
##  1st Qu.: 65276038   1st Qu.:2009-08-31 00:50:00.00   BROOKLYN     :11331
##  Median : 92550741   Median :2013-09-09 05:26:00.00   MANHATTAN    : 3745
##  Mean   :127118170   Mean   :2014-06-01 02:56:39.49   QUEENS       : 4262
##  3rd Qu.:202504685   3rd Qu.:2019-09-16 00:20:00.00   STATEN ISLAND:  800
##  Max.   :279758069   Max.   :2023-12-29 21:22:00.00
##
##     PRECINCT     JURISDICTION_CODE STATISTICAL_MURDER_FLAG PERP_AGE_GROUP
##  75     : 1626   0:23865           Mode :logical           <18  : 1673
##  73     : 1498   1:   80           FALSE:22979             18-24: 6425
##  67     : 1259   2: 4556           TRUE :5522              25-44: 6032
##  44     : 1076                                             45-64:  697
##  79     : 1045                                             65+  :   65
##  47     : 1006                                             NA's :13609
##  (Other):20991
##  PERP_SEX                              PERP_RACE     VIC_AGE_GROUP VIC_SEX
##  F   :  443   AMERICAN INDIAN/ALASKAN NATIVE:    2   <18  : 2946   F: 2753
##  M   :16134   ASIAN / PACIFIC ISLANDER      :  169   18-24:10362   M:25736
```

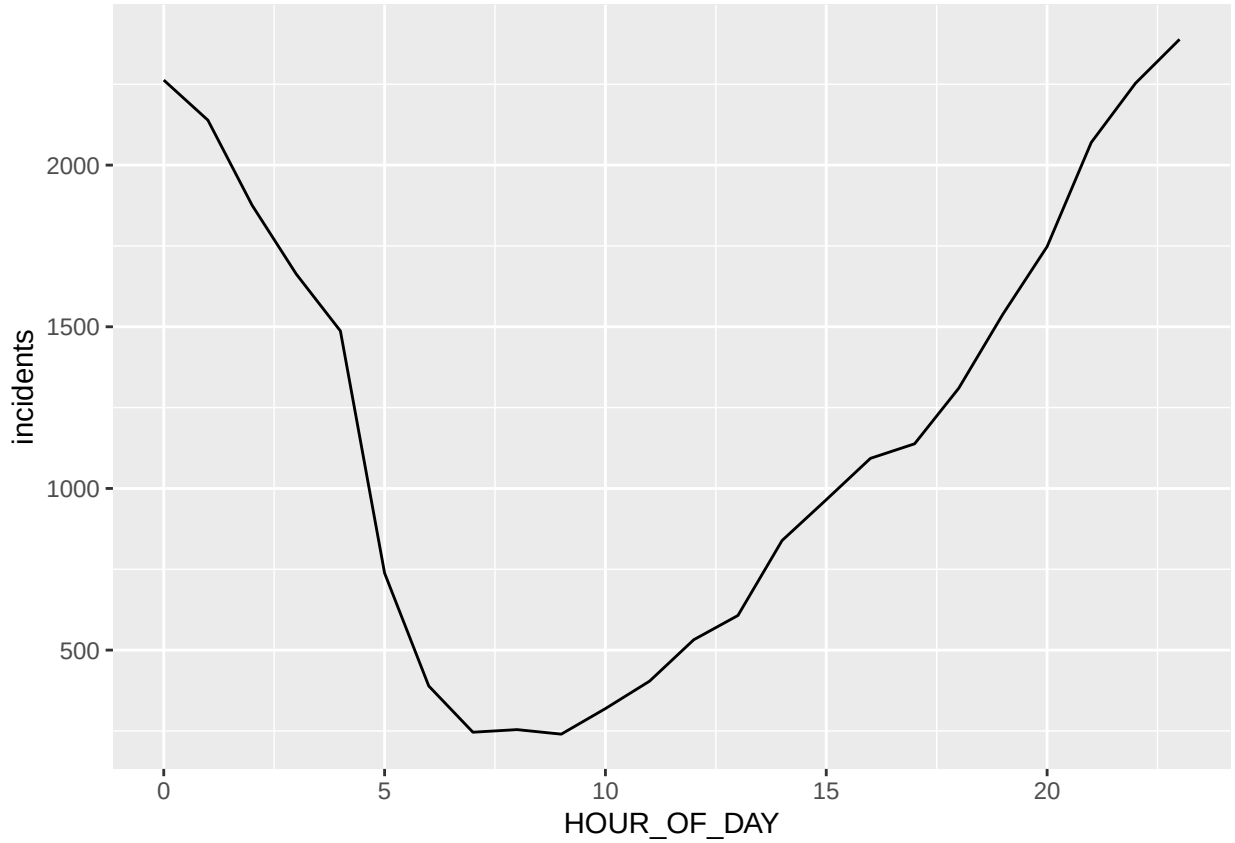
```
##  U   : 1499   BLACK                          :11880   25-44:12945   U:   12
##  NA's:10425   BLACK HISPANIC                 : 1388   45-64: 1978
##               WHITE                          :  298   65+  :  205
##               WHITE HISPANIC                 : 2502   NA's :   65
##               NA's                           :12262
##                              VIC_RACE        Latitude       Longitude
##  AMERICAN INDIAN/ALASKAN NATIVE:   11   Min.   :40.51   Min.   :-74.25
##  ASIAN / PACIFIC ISLANDER      :  440   1st Qu.:40.67   1st Qu.:-73.94
##  BLACK                         :20201   Median :40.70   Median :-73.92
##  BLACK HISPANIC                : 2787   Mean   :40.74   Mean   :-73.91
##  WHITE                         :  728   3rd Qu.:40.82   3rd Qu.:-73.88
##  WHITE HISPANIC                : 4264   Max.   :40.91   Max.   :-73.70
##  NA's                          :   70
##    Lon_Lat
##  Length:28501
##  Class :character
##  Mode  :character
##
##
##
##
```

Analysis

```
# Graph of the sum of shootings by borough and by victim race

ggplot(NYPD_Data, aes(y = BORO)) +
  geom_bar(aes(fill = VIC_RACE))
```




```
# Graph of the shootings by time of day

NYPD_Data_by_time <- NYPD_Data %>%
  mutate(HOUR_OF_DAY = hour(OCCUR_DATETIME)) %>%
  group_by(HOUR_OF_DAY) %>%
  summarise(incidents = n())

ggplot(NYPD_Data_by_time, aes(x = HOUR_OF_DAY, y = incidents)) +
  geom_line()
```



Appendix

```
sessionInfo()
```

```
## R version 4.3.2 (2023-10-31)
## Platform: aarch64-apple-darwin20 (64-bit)
## Running under: macOS Sonoma 14.2.1
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRlapack.dylib;  LAPACK v
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: America/Los_Angeles
## tzcode source: internal
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] ggpubr_0.6.0    usmap_0.7.1     lubridate_1.9.3 forcats_1.0.0
##  [5] stringr_1.5.1   dplyr_1.1.4     purrr_1.0.2     readr_2.1.5
```

```
##  [9] tidyr_1.3.1     tibble_3.2.1    ggplot2_3.5.0   tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
##  [1] utf8_1.2.4        generics_0.1.3    rstatix_0.7.2     stringi_1.8.3
##  [5] hms_1.1.3         digest_0.6.34     magrittr_2.0.3    evaluate_0.23
##  [9] grid_4.3.2        timechange_0.3.0  fastmap_1.1.1     backports_1.4.1
## [13] fansi_1.0.6       scales_1.3.0      abind_1.4-5       cli_3.6.2
## [17] rlang_1.1.3       munsell_0.5.0     withr_3.0.0       yaml_2.3.8
## [21] tools_4.3.2       tzdb_0.4.0        ggsignif_0.6.4    colorspace_2.1-0
## [25] broom_1.0.5       vctrs_0.6.5       R6_2.5.1          lifecycle_1.0.4
## [29] car_3.1-2         pkgconfig_2.0.3   pillar_1.9.0      gtable_0.3.4
## [33] glue_1.7.0        highr_0.10        xfun_0.42         tidyselect_1.2.0
## [37] rstudioapi_0.15.0 knitr_1.45        farver_2.1.1      htmltools_0.5.7
## [41] rmarkdown_2.25    carData_3.0-5     labeling_0.4.3    compiler_4.3.2
```