NYPD shooting incidents by time and age

Kai Gui

2023-10-11

Importing Police Shooting datset

https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD

The NYPD Shooting Dataset is imported from DATA.gov, which provided a dataset from cityofnewyork.us. The scope of this dataset is reported shootings within New York City from 2006 - 2022.

This report aims to better understand the questions: When do shootings occur, and what age groups are involved?

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

```
## 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.
```

```r
shooting_dataset
```

```
## # A tibble: 27,312 x 21
##    INCIDENT_KEY OCCUR_DATE OCCUR_TIME BORO     LOC_OF_OCCUR_DESC PRECINCT
##           <dbl> <chr>      <time>     <chr>    <chr>                <dbl>
##  1    228798151 05/27/2021 21:30      QUEENS   <NA>                   105
##  2    137471050 06/27/2014 17:40      BRONX    <NA>                    40
##  3    147998800 11/21/2015 03:56      QUEENS   <NA>                   108
##  4    146837977 10/09/2015 18:30      BRONX    <NA>                    44
##  5     58921844 02/19/2009 22:58      BRONX    <NA>                    47
##  6    219559682 10/21/2020 21:36      BROOKLYN <NA>                    81
##  7     85295722 06/17/2012 22:47      QUEENS   <NA>                   114
##  8     71662474 03/08/2010 19:41      BROOKLYN <NA>                    81
##  9     83002139 02/05/2012 05:45      QUEENS   <NA>                   105
## 10     86437261 08/26/2012 01:10      QUEENS   <NA>                   101
## # i 27,302 more rows
## # i 15 more variables: JURISDICTION_CODE <dbl>, LOC_CLASSFCTN_DESC <chr>,
```

```
## #   LOCATION_DESC <chr>, STATISTICAL_MURDER_FLAG <lgl>, PERP_AGE_GROUP <chr>,
## #   PERP_SEX <chr>, PERP_RACE <chr>, VIC_AGE_GROUP <chr>, VIC_SEX <chr>,
## #   VIC_RACE <chr>, X_COORD_CD <dbl>, Y_COORD_CD <dbl>, Latitude <dbl>,
## #   Longitude <dbl>, Lon_Lat <chr>
```

Tidying Data: Casting date from string to date object

OCCUR_DATE was imported as a string, so the column was cast as a date. Columns that were not needed for the planned analysis are filtered out and saved to new data frames prior to plotting, and additional cleaning is included with the visualization code (see sections below). Missing data that could be used for additional analysis was identified, but it was decided to not bring external data into this analysis to avoid over complicating this project and expanding the scope beyond what is expected. (See sections on visualizations: age distribution, borough population and police funding)

```
df <- shooting_dataset %>%
  mutate(OCCUR_DATE = mdy(OCCUR_DATE))
summary(df)
```

```
##   INCIDENT_KEY          OCCUR_DATE          OCCUR_TIME           BORO
##  Min.   :  9953245   Min.   :2006-01-01   Length:27312        Length:27312
##  1st Qu.: 63860880   1st Qu.:2009-07-18   Class1:hms          Class :character
##  Median : 90372218   Median :2013-04-29   Class2:difftime     Mode  :character
##  Mean   :120860536   Mean   :2014-01-06   Mode  :numeric
##  3rd Qu.:188810230   3rd Qu.:2018-10-15
##  Max.   :261190187   Max.   :2022-12-31
##
##  LOC_OF_OCCUR_DESC     PRECINCT      JURISDICTION_CODE LOC_CLASSFCTN_DESC
##  Length:27312        Min.   :  1.00   Min.   :0.0000    Length:27312
##  Class :character    1st Qu.: 44.00   1st Qu.:0.0000    Class :character
##  Mode  :character    Median : 68.00   Median :0.0000    Mode  :character
##                      Mean   : 65.64   Mean   :0.3269
##                      3rd Qu.: 81.00   3rd Qu.:0.0000
##                      Max.   :123.00   Max.   :2.0000
##                                       NA's   :2
##  LOCATION_DESC       STATISTICAL_MURDER_FLAG PERP_AGE_GROUP
##  Length:27312        Mode :logical           Length:27312
##  Class :character    FALSE:22046             Class :character
##  Mode  :character    TRUE :5266              Mode  :character
##
##
##
##
##    PERP_SEX            PERP_RACE          VIC_AGE_GROUP        VIC_SEX
##  Length:27312        Length:27312        Length:27312        Length:27312
##  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:27312        Min.   : 914928    Min.   :125757     Min.   :40.51
##  Class :character    1st Qu.:1000029    1st Qu.:182834     1st Qu.:40.67
```

```
##   Mode  :character    Median :1007731   Median :194487   Median :40.70
##                        Mean    :1009449   Mean    :208127   Mean    :40.74
##                        3rd Qu.:1016838   3rd Qu.:239518   3rd Qu.:40.82
##                        Max.    :1066815   Max.    :271128   Max.    :40.91
##                                                             NA's    :10
##      Longitude          Lon_Lat
##   Min.    :-74.25    Length:27312
##   1st Qu.:-73.94    Class :character
##   Median :-73.92    Mode  :character
##   Mean    :-73.91
##   3rd Qu.:-73.88
##   Max.    :-73.70
##   NA's    :10
```

Plotting Shooting occurances over time

To investigate when shootings in New York occur, the number of shootings per hour were obtained and plotted. It is apparent that shootings occur more frequently overnight between 10PM - 4AM. This observation likely affects staffing planning for emergency services in New York, where sufficient personnel must be available to respond to shootings at times outside normal working hours.

By comparing the time-shooting profiles between different boroughs, we observe that, shootings are reported more frequently in the Bronx and Brooklyn. For future work, it would be interesting to bring in external data like population and police department funding per borough. We would then be able to better understand which boroughs have higher frequencies of shootings when controlling for other variables.

```r
df %>%
  mutate(hour = hour(OCCUR_TIME)) %>%
  group_by(hour, BORO) %>%
  summarise(freq = n()) -> hour_freqs
```

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

```r
ggplot(hour_freqs, aes(x = hour, y = freq, fill = BORO)) +
  geom_bar(stat = 'identity') +
  labs(
    x = "Hour of the Day",
    y = "Frequency",
    title = "NYPD Shooting Incidents by Hour (Stacked)",
    fill = "BORO"
  )
```

NYPD Shooting Incidents by Hour (Stacked)
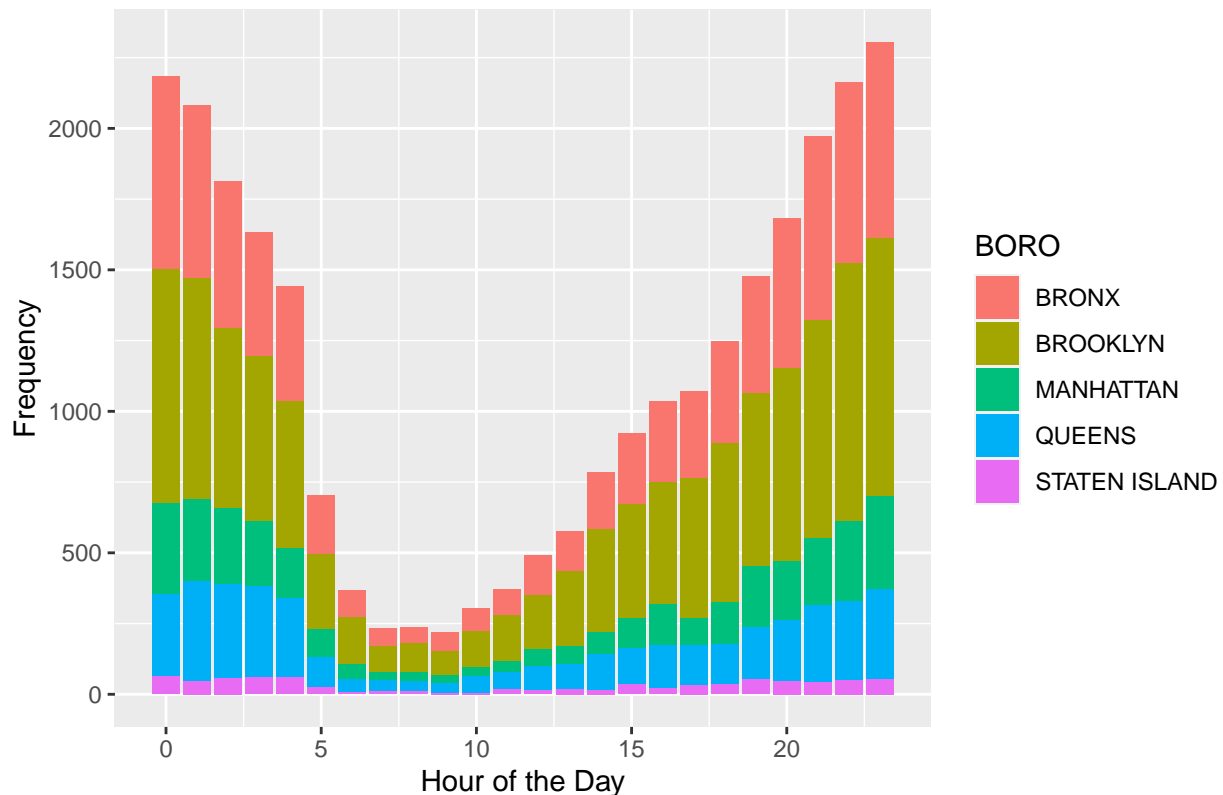
Plotting Shooting occurances vs age groups.

Now that we have established when shootings occur, we now turn our attention to who perpetrators and victims of shootings are. Since shootings occur overnight, we may expect victims and perpetrators to bo be more common with age groups that are more active, in general, over night. When plotting the interaction of perpetrator and victim age groups in a heat map, we do see that the age groups of 18-24 and 25-44 are most likely to be involved with reported shootings. Additionally we do see that the perpetrators and victims age groups are more frequently the same. We can hypothesize that this occurs because there are higher number of interactions for people of similar ages. For future work, we could bring in external data on the typical number of interactions between age groups, and the age-distribution within New York. This would give us a better idea if shootings are occurring as a byproduct of social interactions, or perhaps the age profile of victims and perpetrators simply follow the population distribution of the area.

```r
# Combine messy age groups to 'NA' category
df_age_clean <- df %>%
  mutate(
    PERP_AGE_GROUP = case_when(
      is.na(PERP_AGE_GROUP) ~ 'NA',
      PERP_AGE_GROUP %in% c('(null)', '1020', '224', '940', 'UNKNOWN') ~ 'NA',
      TRUE ~ PERP_AGE_GROUP
    ),
    VIC_AGE_GROUP = case_when(
      VIC_AGE_GROUP == '1022' ~ 'UNKNOWN',
      VIC_AGE_GROUP == 'NA' ~ 'NA',
      TRUE ~ VIC_AGE_GROUP
```

```
    )
  )

cross_tab <- df_age_clean %>%
  group_by(VIC_AGE_GROUP, PERP_AGE_GROUP) %>%
  summarize(count = n()) %>%
  ungroup() %>%
  complete(VIC_AGE_GROUP, PERP_AGE_GROUP, fill = list(count = 0)) %>%
  pivot_wider(names_from = VIC_AGE_GROUP, values_from = count, values_fill = 0)
```

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

```
cross_tab_long <- cross_tab %>%
  pivot_longer(cols = -PERP_AGE_GROUP, names_to = "VIC_AGE_GROUP", values_to = "count")

# Create the heatmap using ggplot2
ggplot(cross_tab_long, aes(x = PERP_AGE_GROUP, y = VIC_AGE_GROUP, fill = count)) +
  geom_tile() +
  scale_fill_gradient(low = "white", high = "blue") +
  labs(
    x = "Perpetrator Age Group",
    y = "Victim Age Group",
    title = "Interaction Between Victim and Perpetrator Age Groups"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Interaction Between Victim and Perpetrator Age Groups
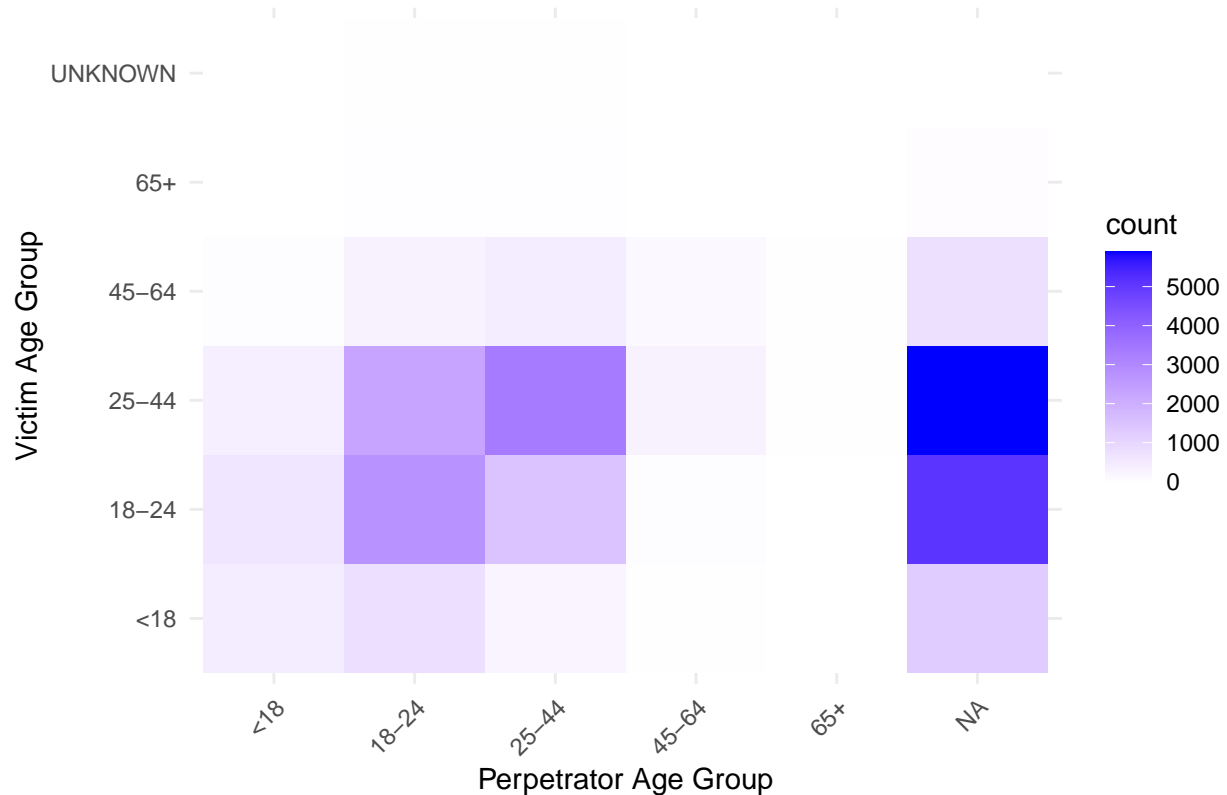
Conclusion

On the questions posed by this report, when do shootings occur and what ages groups are involved, we observed through visualizations that shootings occur most frequently between 10PM-4AM and affect age groups between 18-44 years. We propose additional avenues of investigation bringing in external data, such as population data and the typical number of interactions between different age groups, to further explore questions of which boroughs are experiencing higher than expected (based on population) numbers of shootings, and if the occurrence of shootings is correlated to typical social interactions.

On the topic of bias, we do note that this analysis is completed on only reported shootings. When considering if shootings are higher than expected per borough, for example, we need to be aware that the true number of shootings may differ from what is reported. Certain neighborhoods may have increased police presence or technology in place to monitor shootings. If this is the case, a borough with lower reported shootings may actually be experiencing worse amounts of gun violence. The consequence of this possibility is that if this analysis was used to determine allocation of police funds, the amount of resources may be allocated ineffectively. It would be important to holistically evaluate the state of gun violence before taking action, either by collaborating this report with additional datasets, or by working closely with local stakeholders who are more knowledgeable about the situation on the ground.