NYPD Shooting Incidents

2023-02-18

NYPD Shooting Incident Data (Historical)

Data Source and Summary

Sourcing data from City Of New York at: https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD The source data contains the incidents of shootings (both fatal and non-fatal) that occurred in New York City in the years from 2006 to 2021. The data includes demographics of the perpetrators and victims as well as location data for the incident.

Goal of Analysis

The goal of the analysis will be to understand the relationship of boroughs, victim's age/race/sex, and perpetrator's age/race/sex.
Analysis and visualization:
1. Determine number of shootings and murders per borough per year
2. Determine age/race/sex breakout of victims
3. Determine prevalence of young black males in incidents

Download the data and summarize

Tidy the data

Initial view

Here is a brief look at the data set:

```
## # A tibble: 10 x 19
##    INCID~1 OCCUR~2 OCCUR~3 BORO  PRECI~4 JURIS~5 LOCAT~6 STATI~7 PERP_~8 PERP_~9
##      <dbl> <chr>   <time>  <chr>   <dbl>   <dbl> <chr>   <lgl>   <chr>   <chr>
##  1  2.36e8 11/11/~ 15:04   BROO~      79       0 <NA>    FALSE   <NA>    <NA>
##  2  2.31e8 07/16/~ 22:05   BROO~      72       0 <NA>    FALSE   45-64   M
##  3  2.31e8 07/11/~ 01:09   BROO~      79       0 <NA>    FALSE   <18     M
##  4  2.38e8 12/11/~ 13:42   BROO~      81       0 <NA>    FALSE   <NA>    <NA>
##  5  2.24e8 02/16/~ 20:00   QUEE~     113       0 <NA>    FALSE   <NA>    <NA>
##  6  2.28e8 05/15/~ 04:13   QUEE~     113       0 <NA>    TRUE    <NA>    <NA>
##  7  2.27e8 04/14/~ 21:08   BRONX      42       0 COMMER~ TRUE    <NA>    <NA>
##  8  2.38e8 12/10/~ 19:30   BRONX      52       0 <NA>    FALSE   <NA>    <NA>
##  9  2.25e8 02/22/~ 00:18   MANH~      34       0 <NA>    FALSE   <NA>    <NA>
## 10  2.25e8 03/07/~ 06:15   BROO~      75       0 <NA>    TRUE    25-44   M
## # ... with 9 more variables: 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>, and abbreviated variable
## #   names 1: INCIDENT_KEY, 2: OCCUR_DATE, 3: OCCUR_TIME, 4: PRECINCT,
## #   5: JURISDICTION_CODE, 6: LOCATION_DESC, 7: STATISTICAL_MURDER_FLAG,
## #   8: PERP_AGE_GROUP, 9: PERP_SEX
```

Column reformatting and conversion

Several fields need to reformatted or converted:
- The OCCUR_DATE field needs to be reformatted to a date range - Several fields are converted to factors - A column is added for the occurrence year - Two columns are added which sum the number of shootings and murders - The Incident key, precinct, jurisdiction, and location columns are dropped

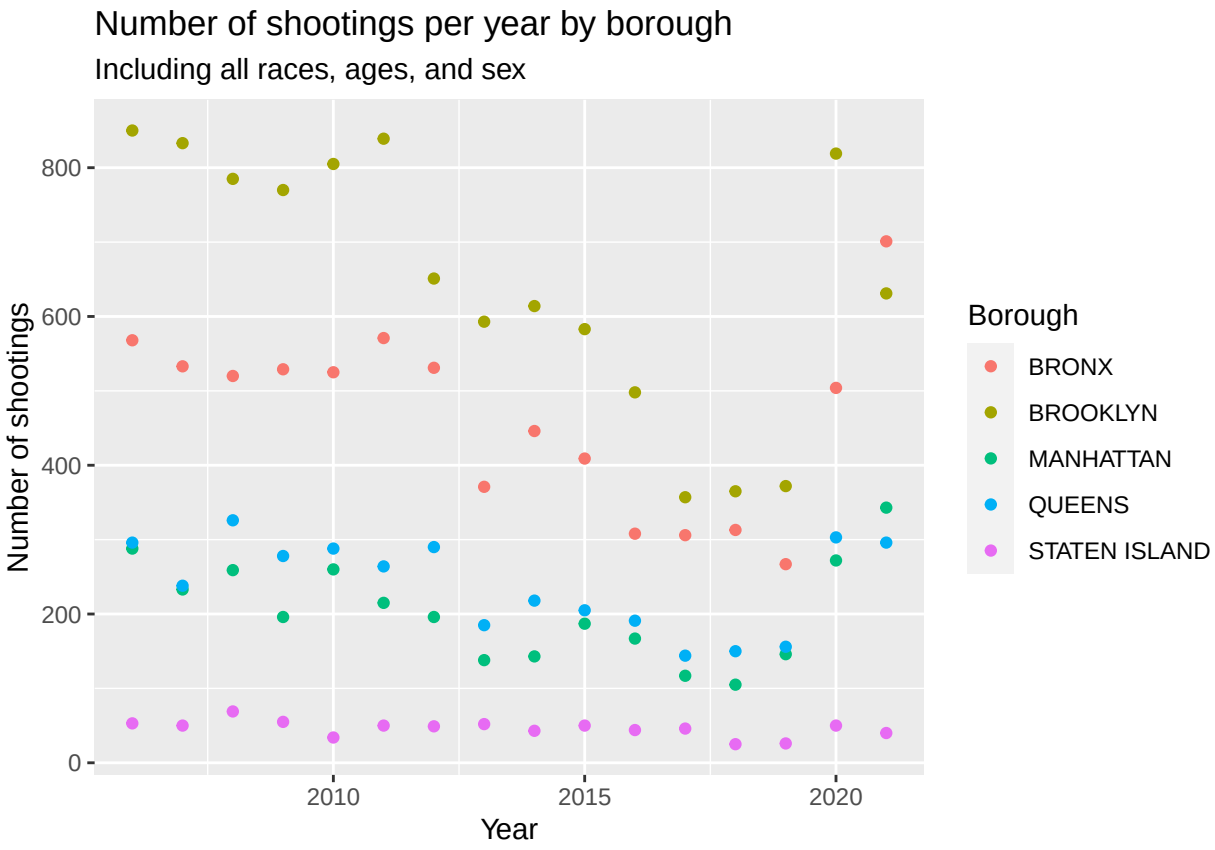
Several columns are unnecessary for the analysis to follow. These include the incident key, precinct, jurisdiction, and location data.

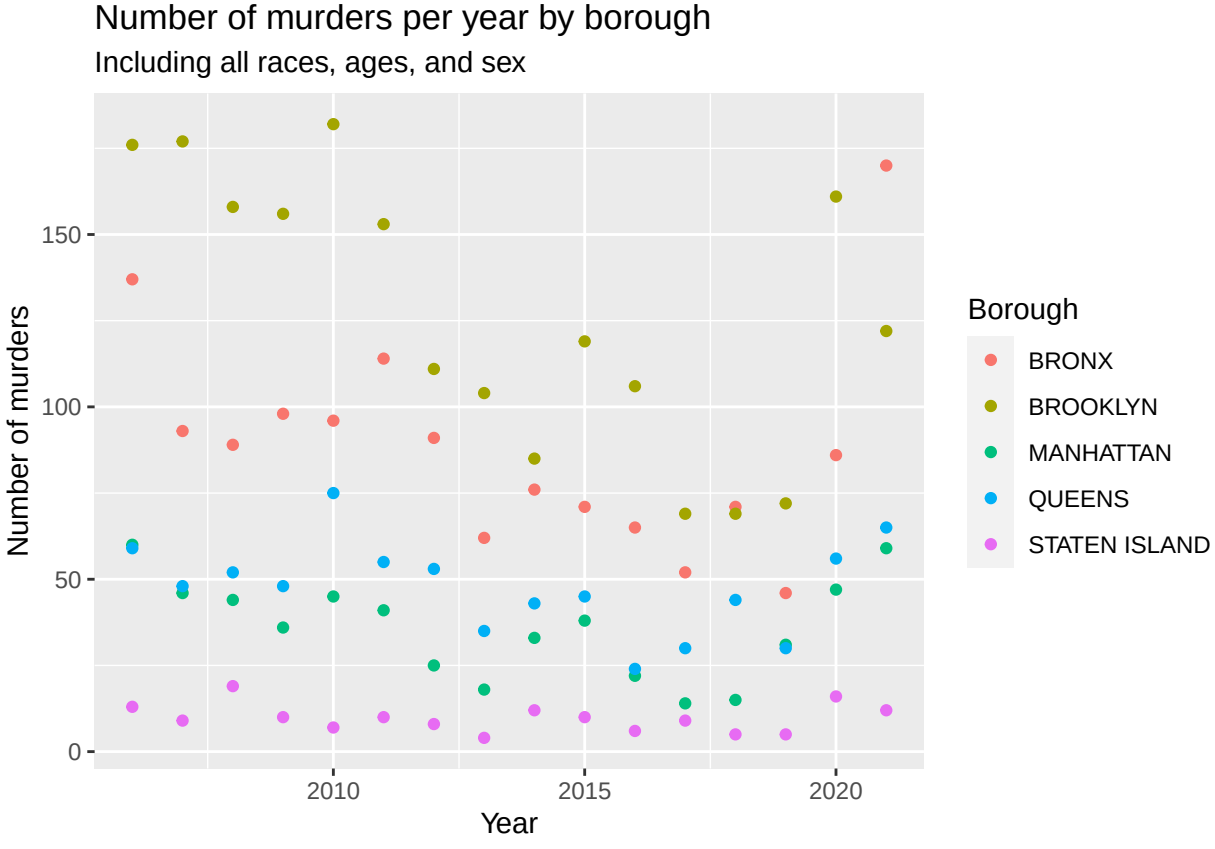
Initial Summary of Source Data

```
##    OCCUR_DATE          OCCUR_TIME                   BORO
##  Min.   :2006-01-01   Length:25596      BRONX        : 7402
##  1st Qu.:2009-05-10   Class1:hms        BROOKLYN     :10365
##  Median :2012-08-26   Class2:difftime   MANHATTAN    : 3265
##  Mean   :2013-06-13   Mode  :numeric    QUEENS       : 3828
##  3rd Qu.:2017-07-01                     STATEN ISLAND:  736
##  Max.   :2021-12-31
##
##  STATISTICAL_MURDER_FLAG PERP_AGE_GROUP PERP_SEX              PERP_RACE
##  FALSE:20668             18-24  :5844   F   :  371   BLACK         :10668
##  TRUE : 4928             25-44  :5202   M   :14416   WHITE HISPANIC: 2164
##                          UNKNOWN:3148   U   : 1499   UNKNOWN       : 1836
##                          <18    :1463   NA's: 9310   BLACK HISPANIC: 1203
##                          45-64  : 535                WHITE         :  272
##                          (Other):  60                (Other)       :  143
##                          NA's   :9344                NA's          : 9310
##  VIC_AGE_GROUP   VIC_SEX                       VIC_RACE       OCCUR_YEAR
##  <18    : 2681   F: 2403   AM INDIAN/AK NATIVE     :    9   Min.   :2006
##  18-24  : 9604   M:23182   ASIAN / PACIFIC ISLANDER:  354   1st Qu.:2009
##  25-44  :11386   U:   11   BLACK                   :18281   Median :2012
##  45-64  : 1698             BLACK HISPANIC          : 2485   Mean   :2013
##  65+    :  167             UNKNOWN                 :   65   3rd Qu.:2017
##  UNKNOWN:   60             WHITE                   :  660   Max.   :2021
##                            WHITE HISPANIC          : 3742
```

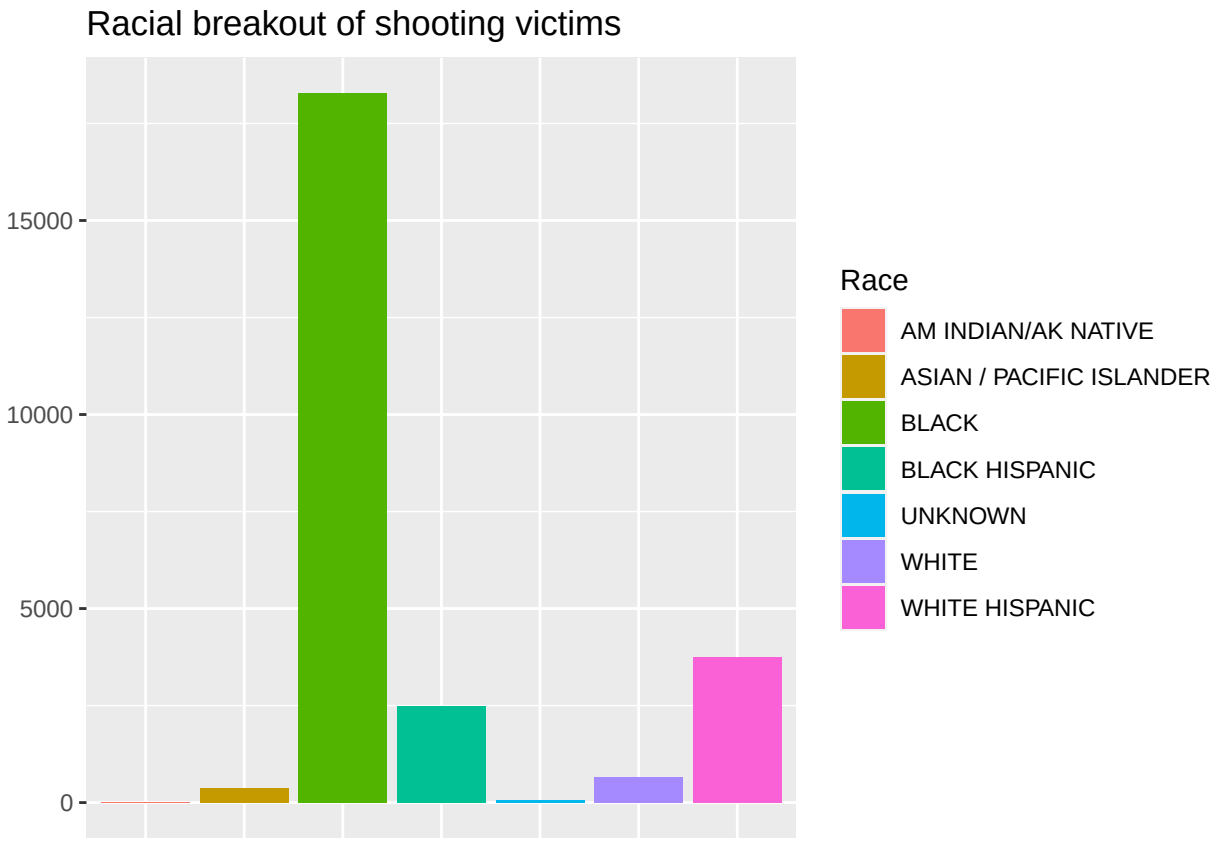
Note: Several fields are missing data or have “Unknown” as values. These will be filtered out in the following analysis.

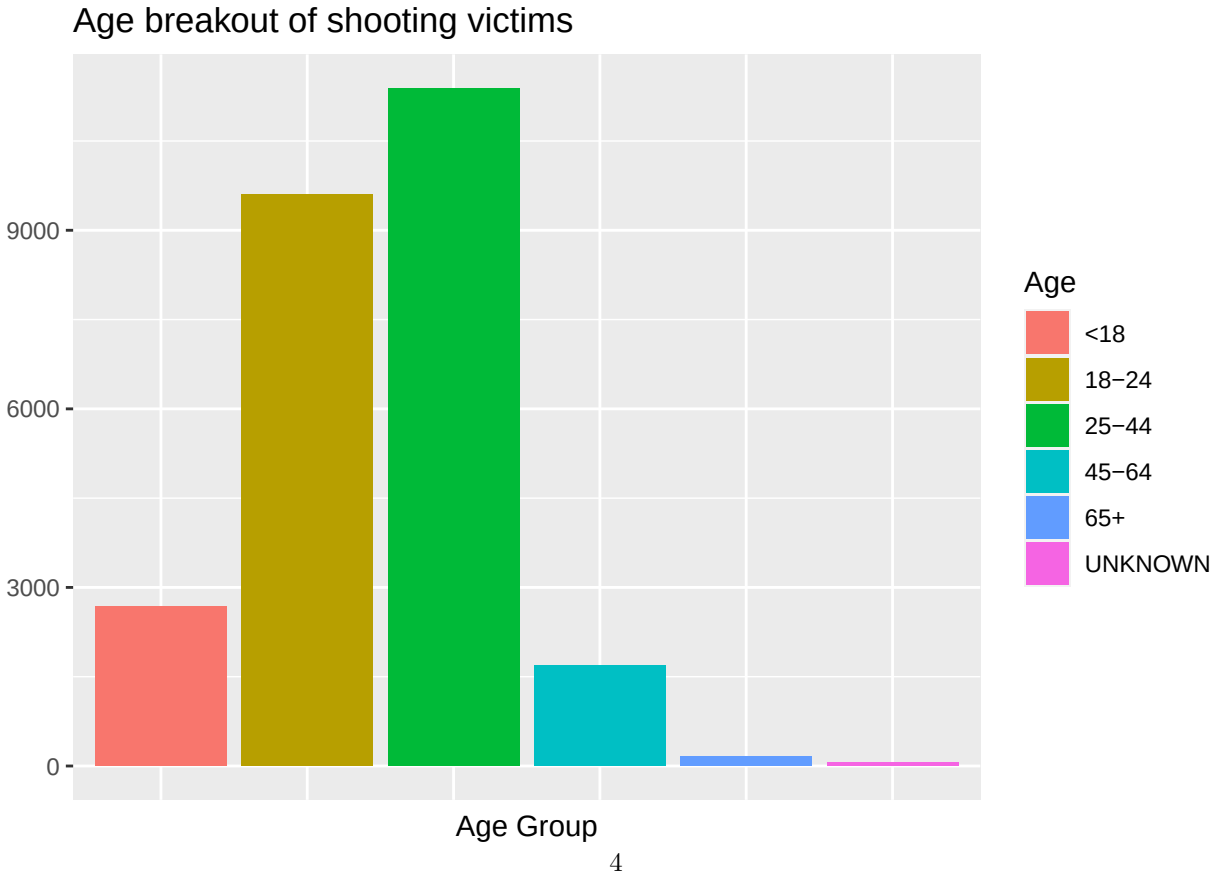
Shootings and Murders per Borough

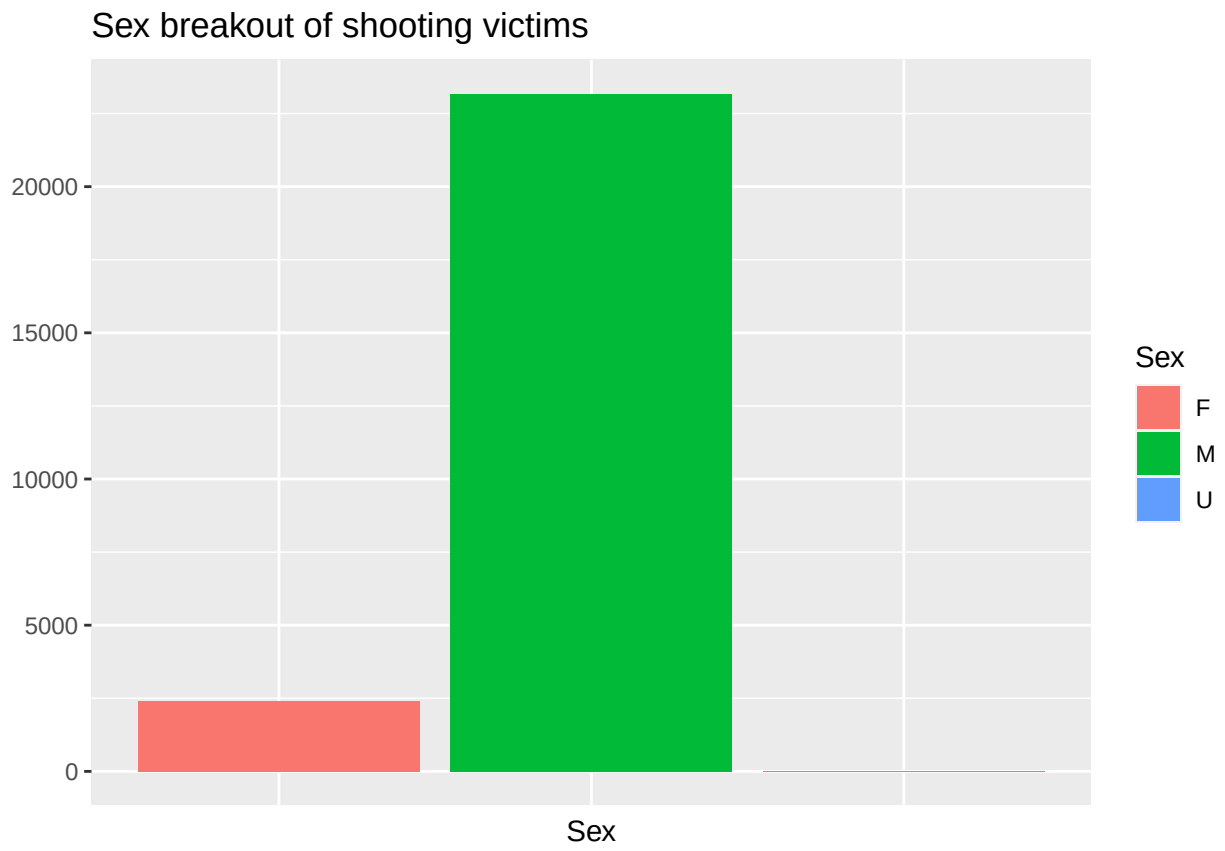




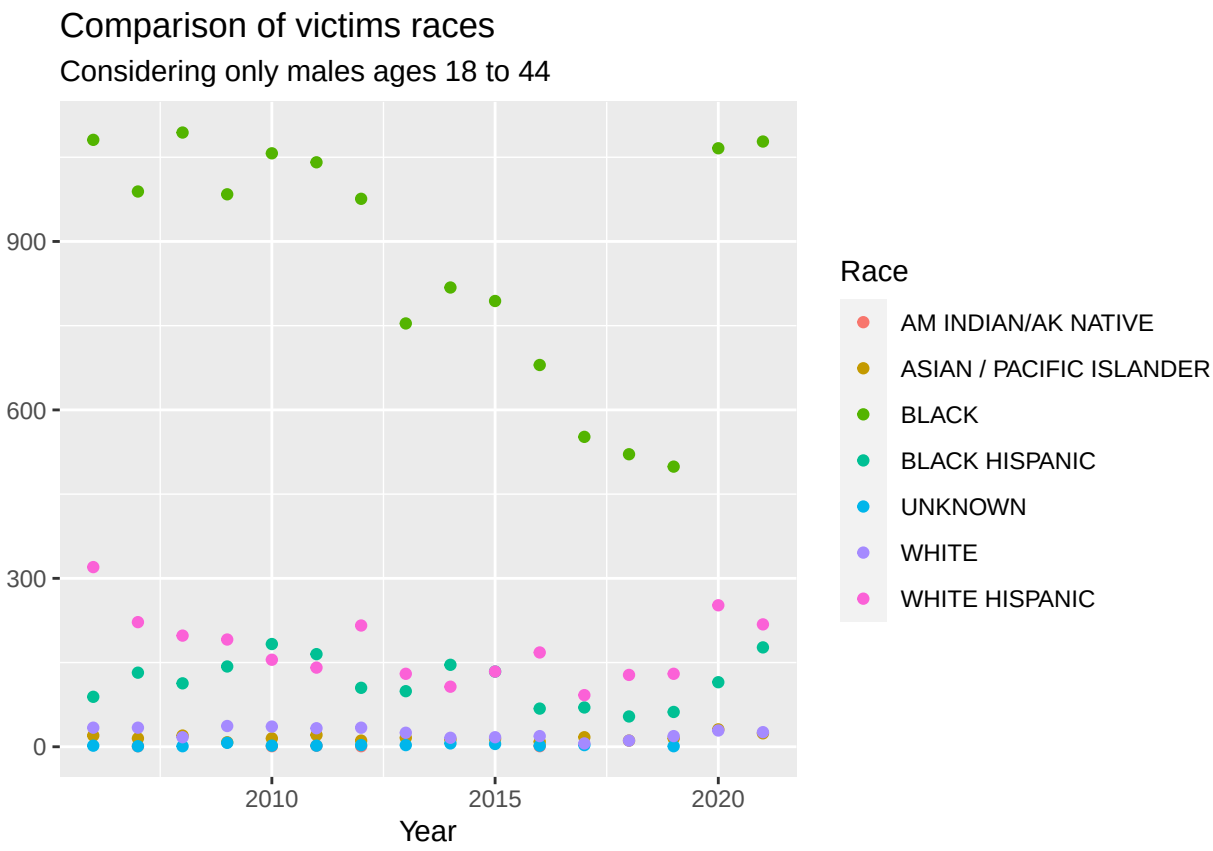
Breakout of Victims' Race, Age, and Sex



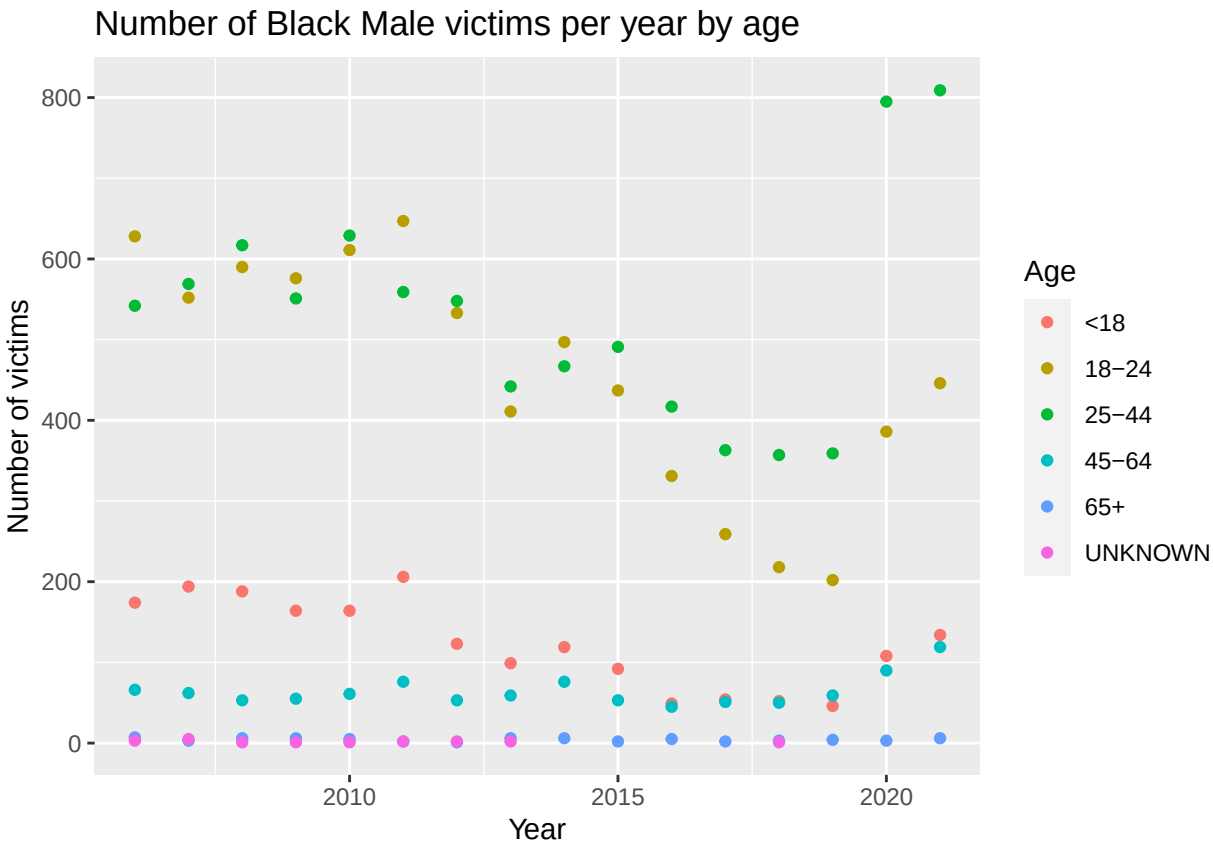


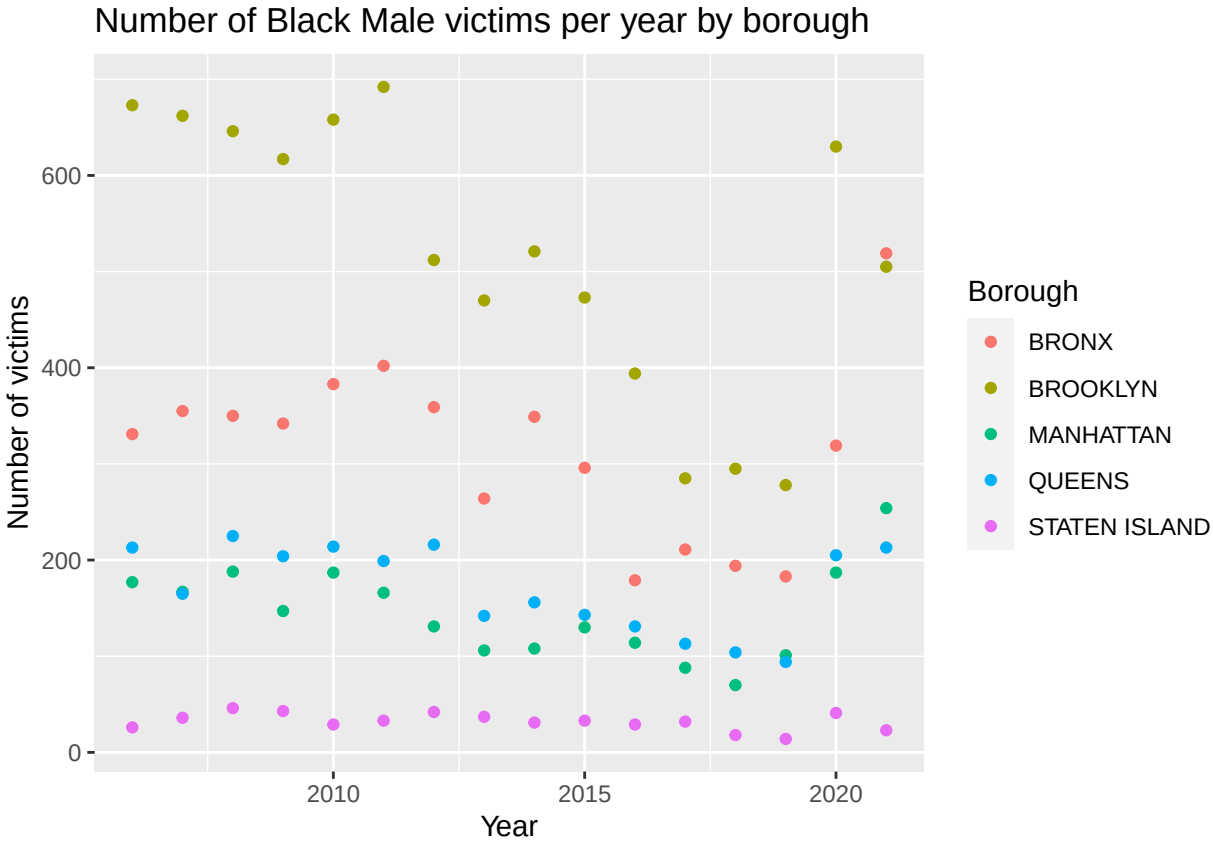


Comparing victims across races


Comparison of victims races
Considering only males ages 18 to 44
900
600
300
0
2010
2015
2020
Year
Race
AM INDIAN/AK NATIVE
ASIAN / PACIFIC ISLANDER
BLACK
BLACK HISPANIC
UNKNOWN
WHITE
WHITE HISPANIC

Focusing on Black Males





Summary and Conclusions

This analysis has shown that the highest shooting and fatality rates are in Brooklyn, followed by the Bronx, though in 2021 the Bronx moved to the highest rate. The analysis has also shown that the highest shooting victim rates are among Blacks, people ages 18 to 44, and males. Focusing on Black males, the victims are most prevalent in the 25 to 44 year age bracket, with the 18-24 year bracket as a close second. An interesting trend is the precipitous drop in shooting for 18 to 24 year-olds from 2014 to 2019.
The data and graphs raise many questions and ideas for additional analysis. For example, additional analysis could show who is committing these crimes in Brooklyn. Or, if Black males are more at risk in certain boroughs than other racial/age/sex groups.

Bias

There are many sources of bias in this data, especially in the source data itself. Capturing age, sex, and race of both the perpetrator and the victim can be inaccurate and incomplete. If the person who is supposed to be capturing the data has a certain bias against a certain race or age group, that person may be more likely or less likely to record the data or mis-record it. As I analyzed the data, I knew that it was socially and politically charged. It can even be weaponized if someone is trying to prove a point and only includes certain views or cross-sections of the data.