NYPD Historic Shooting Data

–

2022-05-29

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

NYPD Historic Shooting Data

This project explores the NYPD Historic Shooting Data, as part of the Data Science as a Field course of the University of Colorado Boulder MSDS program.

```
## Get the historic data on NYPD shootings that is available at data.gov.
url_nypd <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
data_nypd <- read_csv(url_nypd)
```

```
## Rows: 23585 Columns: 19
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (10): OCCUR_DATE, BORO, LOCATION_DESC, PERP_AGE_GROUP, PERP_SEX, PERP_R...
## 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.
```

```
head(data_nypd)
```

```
## # A tibble: 6 x 19
##   INCIDENT_KEY OCCUR_DATE OCCUR_TIME BORO     PRECINCT JURISDICTION_CODE
##          <dbl> <chr>      <time>     <chr>       <dbl>              <dbl>
## 1     24050482 08/27/2006 05:35      BRONX          52                  0
## 2     77673979 03/11/2011 12:03      QUEENS        106                  0
## 3    203350417 10/06/2019 01:09      BROOKLYN       77                  0
## 4     80584527 09/04/2011 03:35      BRONX          40                  0
## 5     90843766 05/27/2013 21:16      QUEENS        100                  0
## 6     92393427 09/01/2013 04:17      BROOKLYN       67                  0
## # ... with 13 more variables: 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>
```

Data cleaning

The columns of INCIDENT_KEY, X_COORD_CD, Y_COORD_CD, Latitude, Longitude, Lon_Lat are removed from **data_nypd**, and the column for OCCUR_DATE is changed to *'date'* type. The variable year, derived from OCCUR_DATE, is added to support future group_by functions.

```
data_nypd <- data_nypd %>%
  select(-c(INCIDENT_KEY, X_COORD_CD, Y_COORD_CD,
         Latitude, Longitude, Lon_Lat)) %>%
  mutate(OCCUR_DATE = mdy(OCCUR_DATE))

data_nypd$year <- year(data_nypd$OCCUR_DATE)

head(data_nypd)
```

```
## # A tibble: 6 x 14
##   OCCUR_DATE OCCUR_TIME BORO     PRECINCT JURISDICTION_CODE LOCATION_DESC
##   <date>     <time>     <chr>       <dbl>             <dbl> <chr>
## 1 2006-08-27 05:35      BRONX          52                 0 <NA>
## 2 2011-03-11 12:03      QUEENS        106                 0 <NA>
## 3 2019-10-06 01:09      BROOKLYN       77                 0 <NA>
## 4 2011-09-04 03:35      BRONX          40                 0 <NA>
## 5 2013-05-27 21:16      QUEENS        100                 0 <NA>
## 6 2013-09-01 04:17      BROOKLYN       67                 0 <NA>
## # ... with 8 more variables: STATISTICAL_MURDER_FLAG <lgl>,
## #   PERP_AGE_GROUP <chr>, PERP_SEX <chr>, PERP_RACE <chr>, VIC_AGE_GROUP <chr>,
## #   VIC_SEX <chr>, VIC_RACE <chr>, year <dbl>
```

```
summary(data_nypd)
```

```
##    OCCUR_DATE          OCCUR_TIME           BORO              PRECINCT
##  Min.   :2006-01-01   Length:23585      Length:23585       Min.   :  1.00
##  1st Qu.:2008-12-31   Class1:hms        Class :character   1st Qu.: 44.00
##  Median :2012-02-27   Class2:difftime   Mode  :character   Median : 69.00
##  Mean   :2012-10-05   Mode  :numeric                       Mean   : 66.21
##  3rd Qu.:2016-03-02                                        3rd Qu.: 81.00
##  Max.   :2020-12-31                                        Max.   :123.00
##
##  JURISDICTION_CODE LOCATION_DESC      STATISTICAL_MURDER_FLAG
##  Min.   :0.000     Length:23585       Mode :logical
##  1st Qu.:0.000     Class :character   FALSE:19085
##  Median :0.000     Mode  :character   TRUE :4500
##  Mean   :0.333
##  3rd Qu.:0.000
##  Max.   :2.000
##  NA's   :2
##  PERP_AGE_GROUP       PERP_SEX          PERP_RACE         VIC_AGE_GROUP
##  Length:23585       Length:23585       Length:23585       Length:23585
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    VIC_SEX            VIC_RACE             year
##  Length:23585       Length:23585       Min.   :2006
##  Class :character   Class :character   1st Qu.:2008
##  Mode  :character   Mode  :character   Median :2012
##                                        Mean   :2012
##                                        3rd Qu.:2016
##                                        Max.   :2020
##
```

The summary of **data_nypd** shows two *NA*'s in JURISDICTION_CODE, but looking at the two rows and counts of the JURISDICTION_CODE, it appears that this variable does not include additional relevant information, since there is also the BORO and PRECINCT. Therefore, the JURISDICTION_CODE is removed.

```
data_nypd %>% filter(is.na(JURISDICTION_CODE))
```

```
## # A tibble: 2 x 14
##   OCCUR_DATE OCCUR_TIME BORO      PRECINCT JURISDICTION_CODE LOCATION_DESC
##   <date>     <time>     <chr>        <dbl>             <dbl> <chr>
## 1 2007-07-13 01:10      QUEENS         104                NA SOCIAL CLUB/POLICY~
## 2 2019-03-09 02:41      MANHATTAN       25                NA <NA>
## # ... with 8 more variables: STATISTICAL_MURDER_FLAG <lgl>,
## #   PERP_AGE_GROUP <chr>, PERP_SEX <chr>, PERP_RACE <chr>, VIC_AGE_GROUP <chr>,
## #   VIC_SEX <chr>, VIC_RACE <chr>, year <dbl>
```

```
data_nypd %>% count(JURISDICTION_CODE)
```

```
## # A tibble: 4 x 2
##   JURISDICTION_CODE     n
##               <dbl> <int>
## 1                 0 19629
## 2                 1    54
## 3                 2  3900
## 4                NA     2
```

```
data_nypd <- data_nypd %>% select(-JURISDICTION_CODE)
```

There are no other *NA* in the **data__nypd** summary, however there are 7 *character* data-type columns. The count function is used to inspect the data.

```
data_nypd %>% count(BORO)
```

```
## # A tibble: 5 x 2
##   BORO              n
##   <chr>         <int>
## 1 BRONX          6701
## 2 BROOKLYN       9734
## 3 MANHATTAN      2922
## 4 QUEENS         3532
## 5 STATEN ISLAND   696
```

```
data_nypd %>% count(LOCATION_DESC)
```

```
## # A tibble: 40 x 2
##    LOCATION_DESC         n
##    <chr>             <int>
##  1 ATM                   1
##  2 BANK                  1
##  3 BAR/NIGHT CLUB      562
##  4 BEAUTY/NAIL SALON   100
##  5 CANDY STORE           6
##  6 CHAIN STORE           5
##  7 CHECK CASH            1
##  8 CLOTHING BOUTIQUE    14
##  9 COMMERCIAL BLDG     234
## 10 DEPT STORE            5
## # ... with 30 more rows
```

```
data_nypd %>% count(PERP_AGE_GROUP)
```

```
## # A tibble: 10 x 2
##    PERP_AGE_GROUP     n
##    <chr>          <int>
##  1 <18             1368
##  2 1020               1
```

```
##  3 18-24           5508
##  4 224                1
##  5 25-44           4714
##  6 45-64            495
##  7 65+               54
##  8 940                1
##  9 UNKNOWN         3148
## 10 <NA>            8295
```

```
data_nypd %>% count(PERP_SEX)
```

```
## # A tibble: 4 x 2
##   PERP_SEX     n
##   <chr>    <int>
## 1 F          335
## 2 M        13490
## 3 U         1499
## 4 <NA>      8261
```

```
data_nypd %>% count(PERP_RACE)
```

```
## # A tibble: 8 x 2
##   PERP_RACE                          n
##   <chr>                          <int>
## 1 AMERICAN INDIAN/ALASKAN NATIVE     2
## 2 ASIAN / PACIFIC ISLANDER         122
## 3 BLACK                          10025
## 4 BLACK HISPANIC                  1096
## 5 UNKNOWN                         1836
## 6 WHITE                            255
## 7 WHITE HISPANIC                  1988
## 8 <NA>                            8261
```

```
data_nypd %>% count(VIC_AGE_GROUP)
```

```
## # A tibble: 6 x 2
##   VIC_AGE_GROUP     n
##   <chr>         <int>
## 1 <18            2525
## 2 18-24          9003
## 3 25-44         10303
## 4 45-64          1541
## 5 65+             154
## 6 UNKNOWN          59
```

```
data_nypd %>% count(VIC_SEX)
```

```
## # A tibble: 3 x 2
##   VIC_SEX     n
##   <chr>   <int>
## 1 F        2204
## 2 M       21370
## 3 U          11
```

```
data_nypd %>% count(VIC_RACE)
```

```
## # A tibble: 7 x 2
##   VIC_RACE                          n
##   <chr>                         <int>
## 1 AMERICAN INDIAN/ALASKAN NATIVE    9
## 2 ASIAN / PACIFIC ISLANDER        327
## 3 BLACK                         16869
## 4 BLACK HISPANIC                 2245
## 5 UNKNOWN                          65
## 6 WHITE                           620
## 7 WHITE HISPANIC                 3450
```

The columns BORO, VIC_AGE, VIC_SEX and VIC_RACE have no *NA* values. VIC_AGE and VIC_RACE have a small number of *"UNKNOWN"* values, but this is less than 1% of the total so it is unlikely to be significant.

The column LOCATION_DESC has the value *NA* 13581 times (58%) and *"None"* 175 times (0.7%). With over half the values missing, LOCATION_DESC will be removed from this data.

```
data_nypd %>%
  filter(is.na(LOCATION_DESC) | LOCATION_DESC == 'NONE') %>%
  count(LOCATION_DESC)
```

```
## # A tibble: 2 x 2
##   LOCATION_DESC     n
##   <chr>         <int>
## 1 NONE            175
## 2 <NA>          13581
```

```
data_nypd <- data_nypd %>% select(-LOCATION_DESC)
```

The columns with identifying information about the perpetrator contain a large number of *NA* and *"UNKNOWN"* values, with 35% of the values as *NA*. Since this is significant and probably a result of the police not being able to identify the perpetrators, *NA* with be replaced with *"UNKNOWN"*. In addition, there are 1499 entries for PERP_SEX as *"U"*, which will be replaced with *"UNKNOWN"*, to combine the *"U"* and *NA* values. The single entries for PERP_AGE_GROUP of *224*, *940* and *1020* will be removed.

```
data_nypd <- data_nypd %>%
  replace_na(list(PERP_AGE_GROUP = 'UNKNOWN',
                  PERP_SEX = 'UNKNOWN',
                  PERP_RACE = 'UNKNOWN')) %>%
  filter(PERP_AGE_GROUP != 940, PERP_AGE_GROUP != 224, PERP_AGE_GROUP != 1020)

data_nypd["PERP_SEX"][data_nypd["PERP_SEX"] == "U"] <- "UNKNOWN"
```

Data analysis part 1

The analysis starts with an overall question about the shootings reported to the NYPD between 2006 and 2020:
How have the number of shootings and murders changed?

To explore this question, the number of shootings and murders per day were calculated, and joined using a full_join to the data frame **daily_total**.

```
daily_total <- data_nypd %>%
  count(OCCUR_DATE) %>%
  rename(no_shooting = n)

daily_murders <- data_nypd %>%
  group_by(OCCUR_DATE) %>%
  summarize(murders = sum(STATISTICAL_MURDER_FLAG))

daily_total <- daily_total %>%
  full_join(daily_murders)
```

```
## Joining, by = "OCCUR_DATE"
```

```
head(daily_total)
```

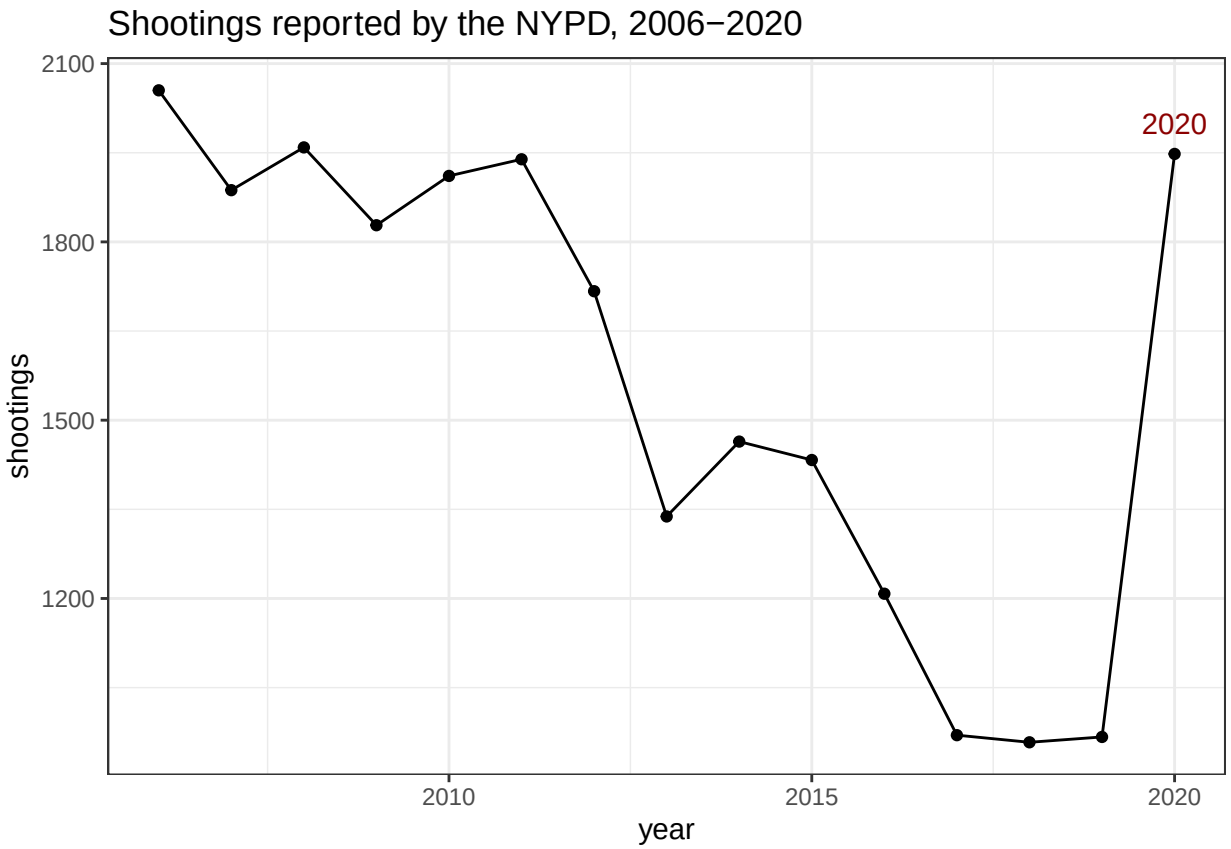
```
## # A tibble: 6 x 3
##   OCCUR_DATE no_shooting murders
##   <date>           <int>   <int>
## 1 2006-01-01           8       4
## 2 2006-01-02           4       1
## 3 2006-01-03           4       1
## 4 2006-01-04           4       0
## 5 2006-01-05           4       0
## 6 2006-01-06           4       0
```

For the purpose of this exploration, we will consider the yearly, not daily, statistics. A new data frame is created, **daily_year**, grouping the shooting and murder statistics.

```
daily_total$year <- year(daily_total$OCCUR_DATE)

yearly_total <- daily_total %>%
  group_by(year) %>%
  summarize(shootings = sum(no_shooting), murders = sum(murders))

head(yearly_total)
```

```
## # A tibble: 6 x 3
##    year shootings murders
##   <dbl>     <int>   <int>
## 1  2006      2055     445
## 2  2007      1887     373
## 3  2008      1959     362
## 4  2009      1828     348
## 5  2010      1911     405
## 6  2011      1939     373
```

An initial visualization shows a dramatic increase in shootings in 2020, after 14 years of declining.



Two linear regression models were considered with *'year'* as the predictor variable and *'number of shootings'* as the response: the whole data set from 2006-2020, and a reduced data set excluding 2020 to consider how the trend prior to 2020. This process was repeated for number of reported shootings that resulted in murders.

```
shoot_mod = lm(shootings ~ year, yearly_total)
summary(shoot_mod)
```

```
##
## Call:
## lm(formula = shootings ~ year, data = yearly_total)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -343.85 -194.06   -9.99   72.15  827.87
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 131554.42   35247.96   3.732  0.00251 **
## year           -64.57      17.51  -3.688  0.00273 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 293 on 13 degrees of freedom
## Multiple R-squared:  0.5113, Adjusted R-squared:  0.4737
## F-statistic:  13.6 on 1 and 13 DF,  p-value: 0.002734
```

```r
murder_mod = lm(murders~ year, yearly_total)
summary(murder_mod)
```

```
##
## Call:
## lm(formula = murders ~ year, data = yearly_total)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -77.00 -33.21 -11.39  28.00 164.45
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 28611.407   7554.616   3.787  0.00226 **
## year          -14.064      3.753  -3.748  0.00244 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 62.8 on 13 degrees of freedom
## Multiple R-squared:  0.5193, Adjusted R-squared:  0.4823
## F-statistic: 14.04 on 1 and 13 DF,  p-value: 0.002439
```

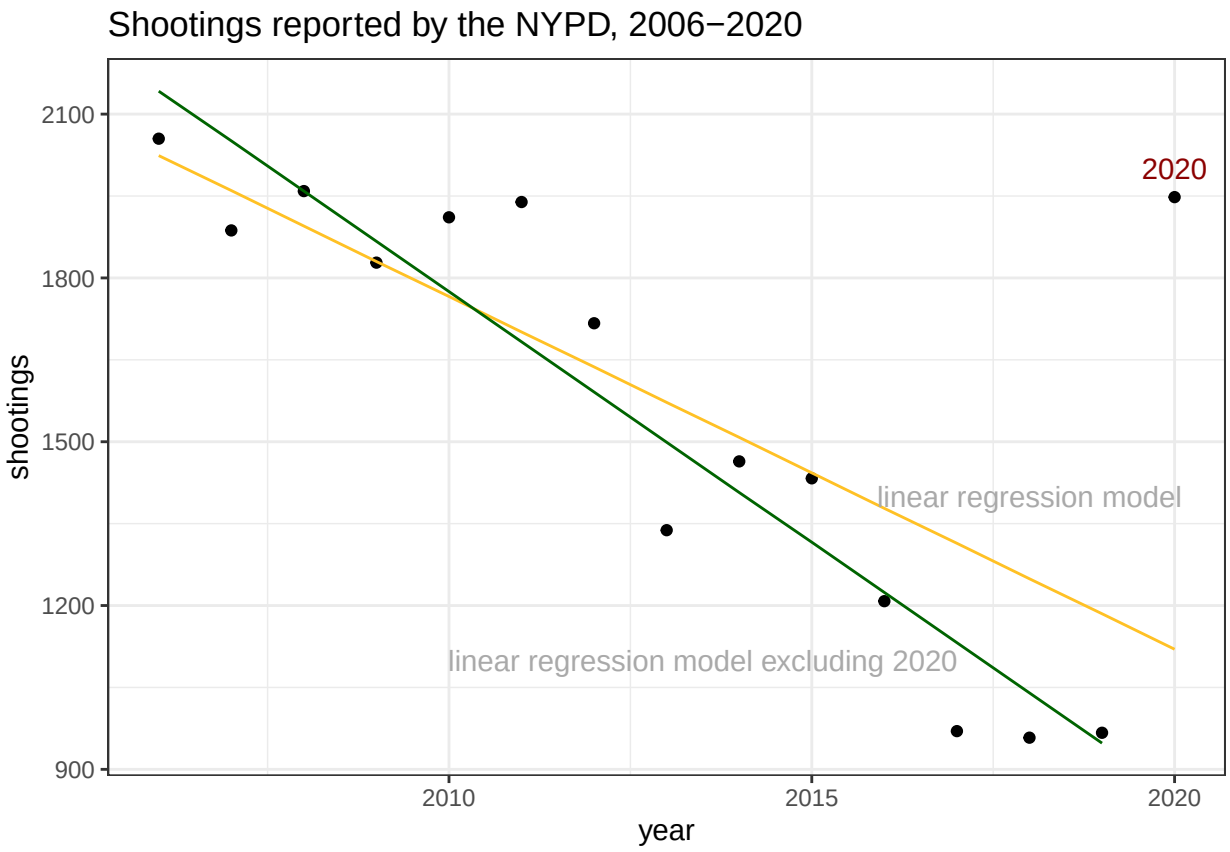
```r
yearly_total_reduced <- yearly_total %>% filter(year < 2020)

shoot_mod_rd = lm(shootings ~ year, yearly_total_reduced)
summary(shoot_mod_rd)
```

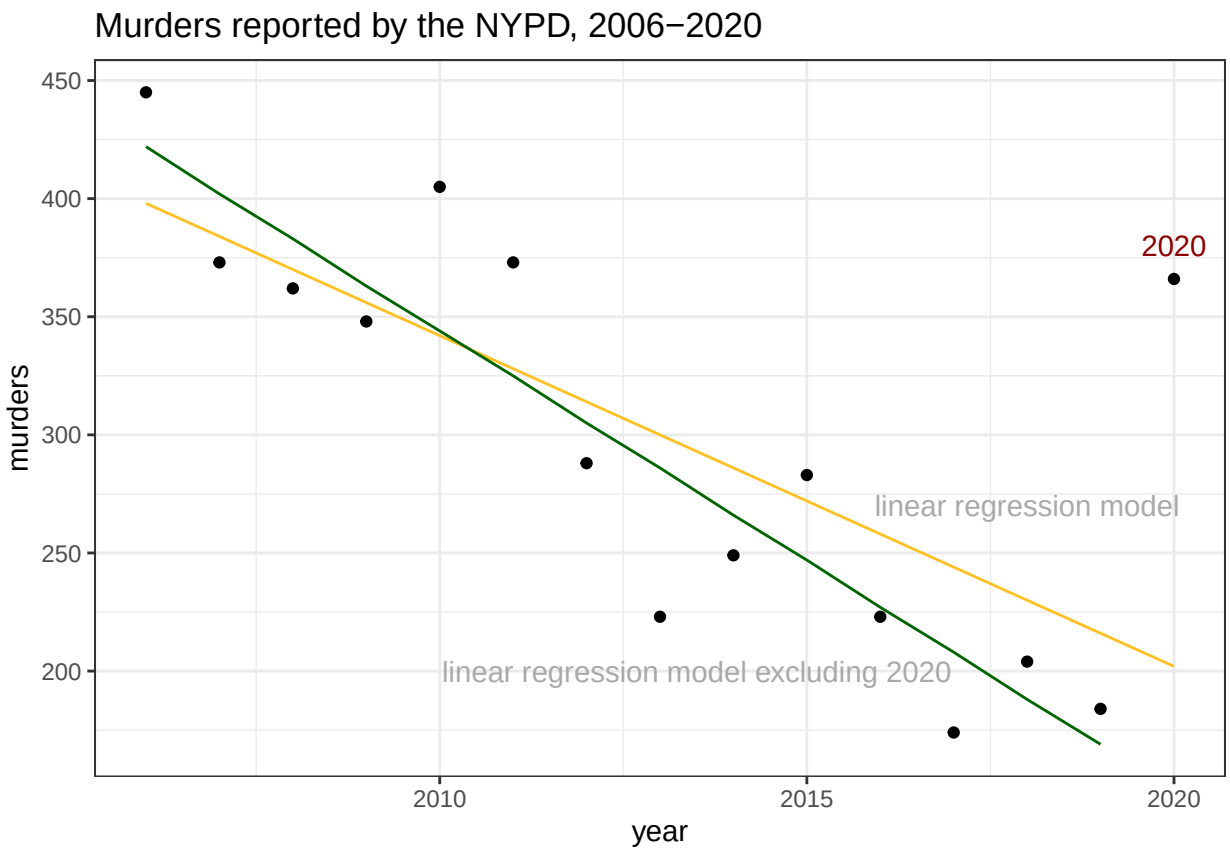
```
##
## Call:
## lm(formula = shootings ~ year, data = yearly_total_reduced)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -163.536  -86.059   -7.718  102.158  255.919
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 186421.055  17745.746   10.51  2.1e-07 ***
## year           -91.864      8.818  -10.42  2.3e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 133 on 12 degrees of freedom
## Multiple R-squared:  0.9004, Adjusted R-squared:  0.8921
## F-statistic: 108.5 on 1 and 12 DF,  p-value: 2.297e-07
```

```r
murder_mod_rd = lm(murders ~ year, yearly_total_reduced)
summary(murder_mod_rd)
```

```
##
## Call:
```

```
## lm(formula = murders ~ year, data = yearly_total_reduced)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -62.543 -19.993  -9.786  21.264  61.000
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 39510.286   4811.527   8.212 2.88e-06 ***
## year          -19.486      2.391  -8.150 3.11e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 36.06 on 12 degrees of freedom
## Multiple R-squared:  0.847,  Adjusted R-squared:  0.8342
## F-statistic: 66.43 on 1 and 12 DF,  p-value: 3.106e-06
```





While the p-values and t-tests on all four models was below alpha = 0.05, they were much smaller on the reduced data sets. This result is also evident in the visualizations, which suggests that 2020 is a significant outlier in the trend of decreasing shootings and murders from shootings in NY.
These findings are consistent with news reporting of an overall increase in gun violence in the US in 2020 https://www.npr.org/transcripts/1040904770.

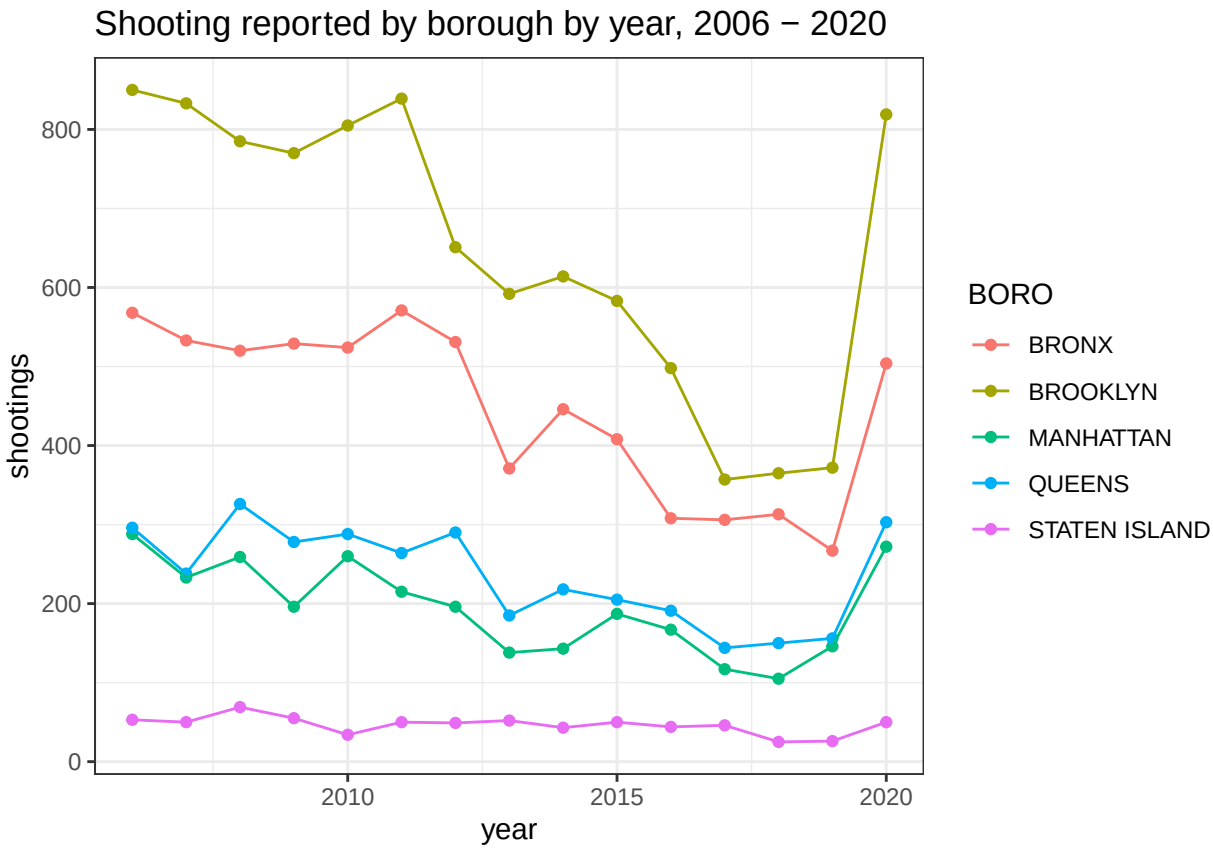
Data analysis, part 2

The follow-up question to these initial findings is whether the dramatic increase in shootings was consistent across different areas, victim characteristics and perpetrator characteristics. **1. Which boroughs of NY had the greatest increase in shootings in 2020?**
2. Which victims experienced the greatest increase in shootings in 2020?
3. What were the characteristics of the perpetrators associated with this increase in shootings in 2020?

NY Boroughs

```
boro_count <- data_nypd %>%
  group_by(year) %>%
  count(BORO) %>%
  rename(shootings = n)
```

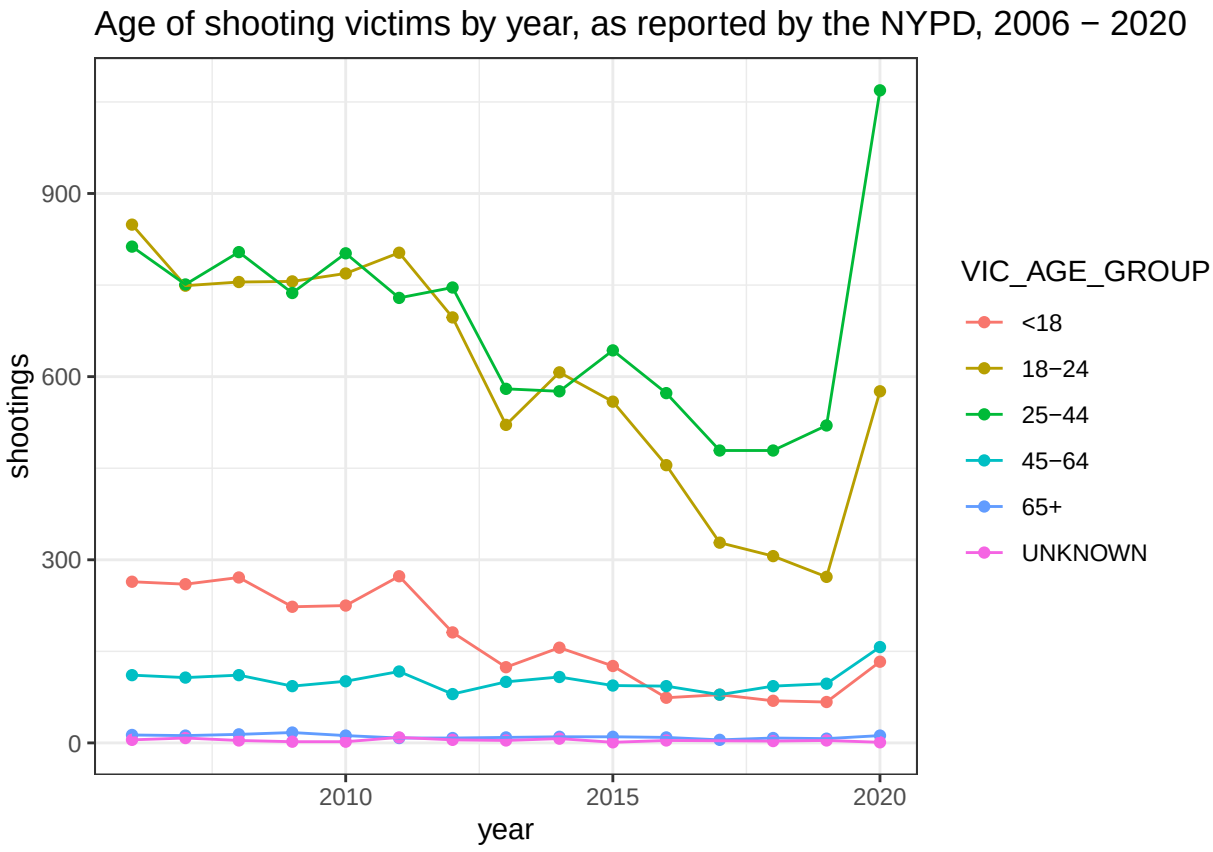
```
ggplot(boro_count, aes(x = year, y = shootings, color = BORO)) +
  geom_point() +
  geom_line() +
  labs(title = "Shooting reported by borough by year, 2006 - 2020") +
  theme_bw()
```



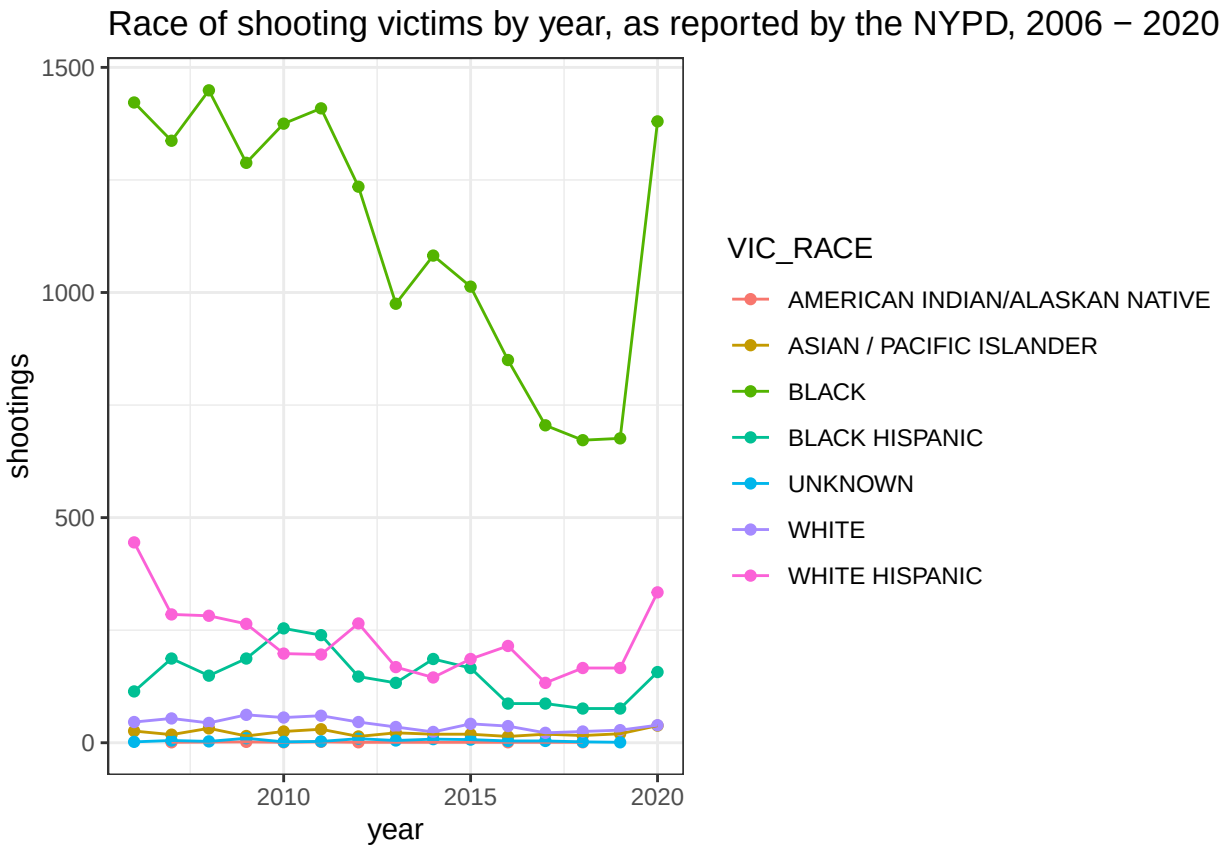
The most greatest increase in shootings happened in Brooklyn, followed by the Bronx. It is notable that in the four boroughs of Brooklyn, Bronx, Queens and Manhattan, the total number of shootings had an overall decreasing trend from 2006 to 2019, but in 2020 they rose to approximately their 2006 levels. Staten Island had the lowest number of shootings and the least amount of variability in 2020.

Victim characteristics

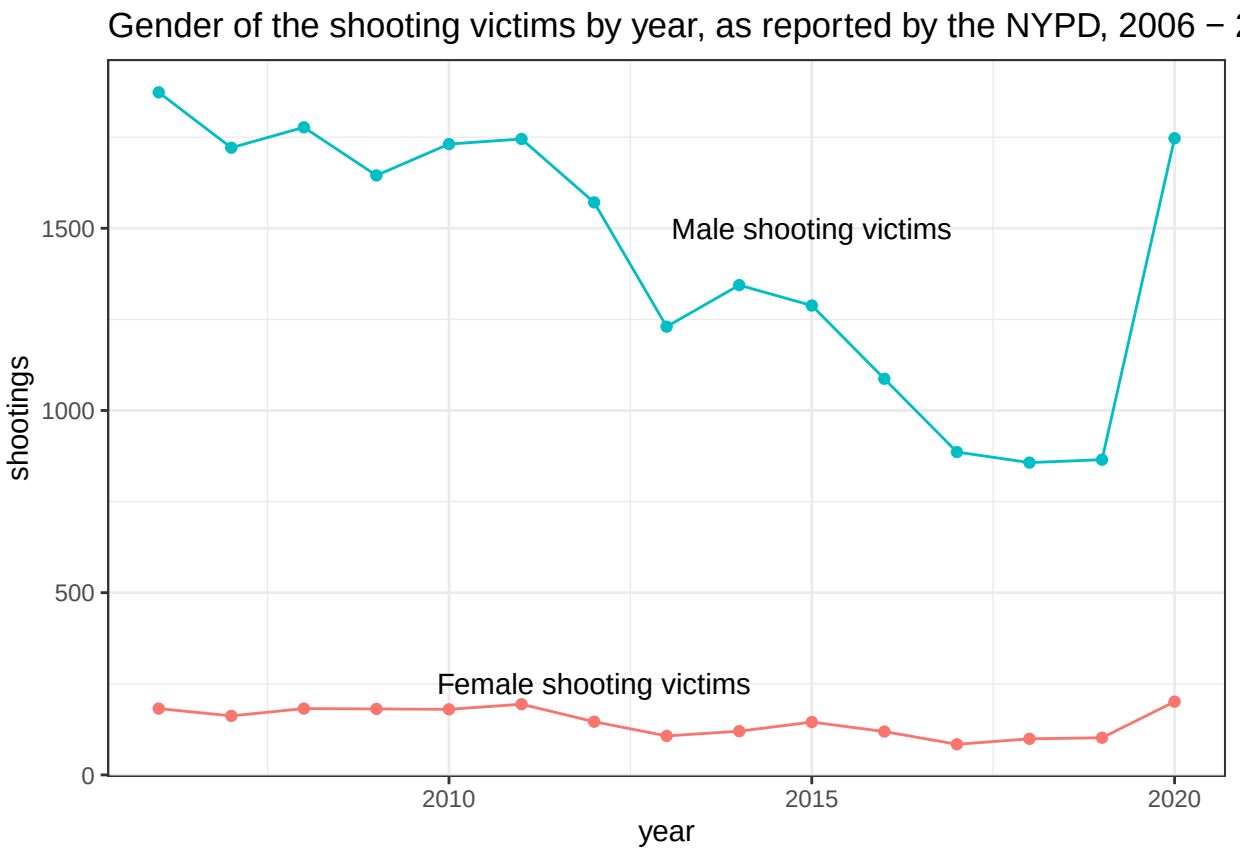
```
vic_age_count <- data_nypd %>%
  group_by(year) %>%
  count(VIC_AGE_GROUP) %>%
  rename(shootings = n)

ggplot(vic_age_count, aes(x = year, y = shootings, color = VIC_AGE_GROUP)) +
  geom_point() +
  geom_line() +
  labs(title = "Age of shooting victims by year, as reported by the NYPD, 2006 - 2020") +
  theme_bw()
```




```
vic_race_count <- data_nypd %>%
  group_by(year) %>%
  count(VIC_RACE) %>%
  rename(shootings = n)

ggplot(vic_race_count, aes(x = year, y = shootings, color = VIC_RACE)) +
  geom_point() +
  geom_line() +
  labs(title = "Race of shooting victims by year, as reported by the NYPD, 2006 - 2020") +
  theme_bw()
```



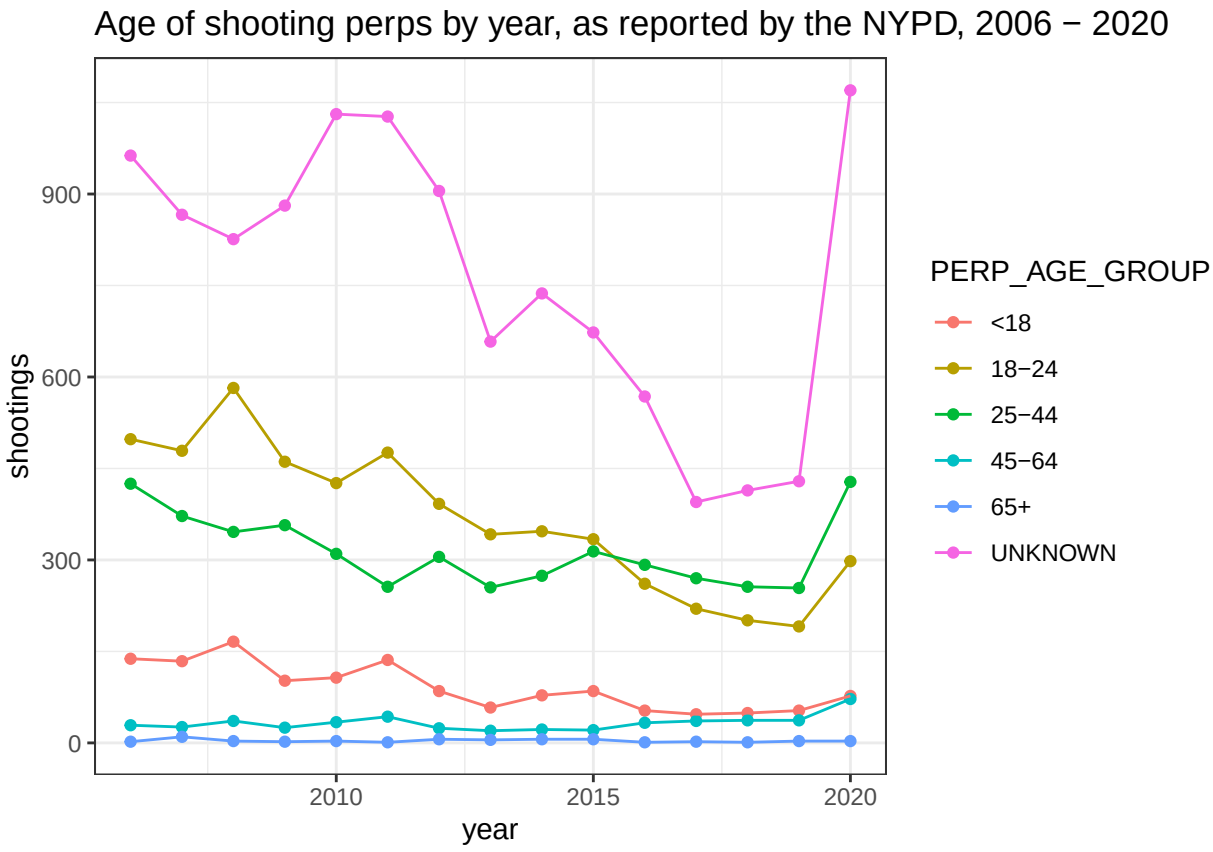

```
vic_sex_count <- data_nypd %>%
  group_by(year) %>%
  count(VIC_SEX) %>%
  filter(VIC_SEX != "U") %>%
  rename(shootings = n)

ggplot(vic_sex_count, aes(x = year, y = shootings, color = VIC_SEX)) +
  geom_point() +
  geom_line() +
  labs(title = "Gender of the shooting victims by year, as reported by the NYPD, 2006 - 2020") +
  annotate("text", x = 2012, y = 250, label = "Female shooting victims") +
  annotate("text", x = 2015, y = 1500, label = "Male shooting victims") +
  theme_bw() +
  theme(legend.position="none")
```



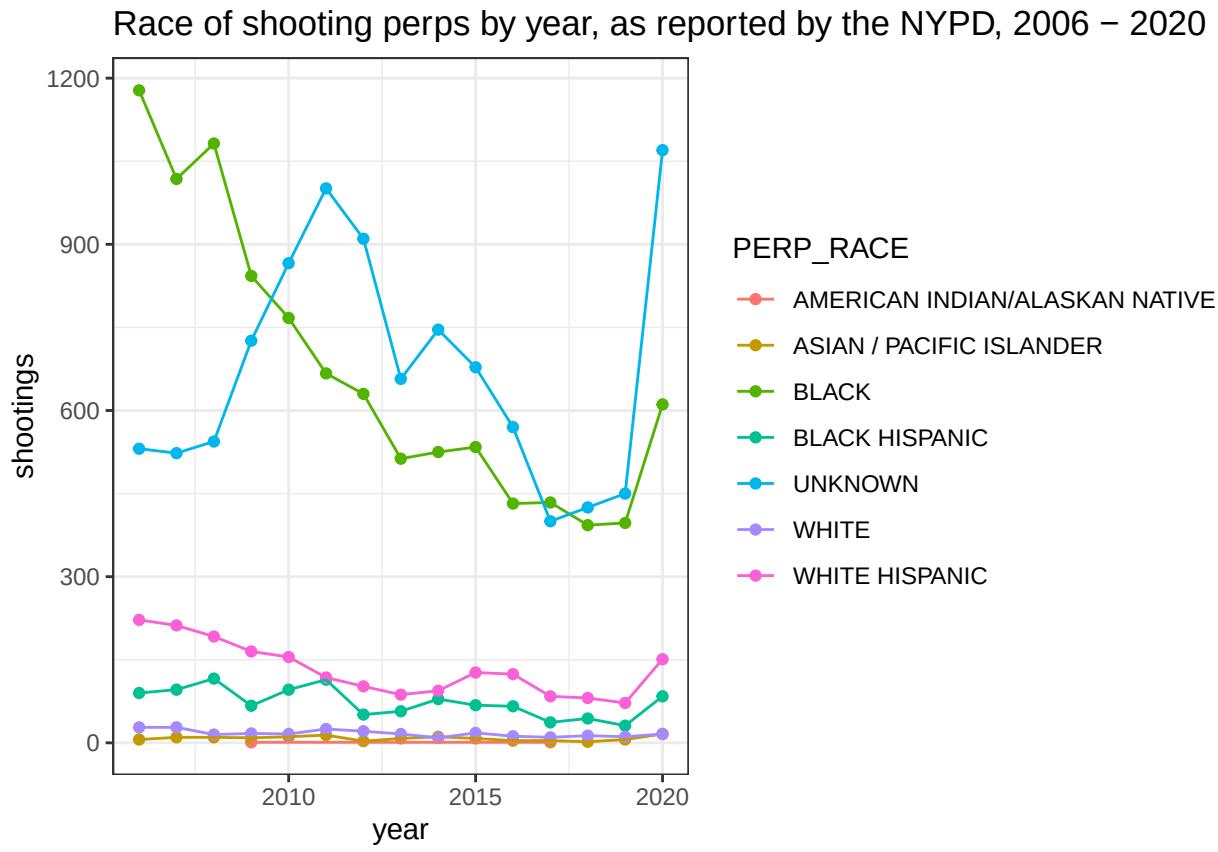
Just as in the case of the total number of shootings by borough, the characteristics of the victims (race, age and gender) that generally had the highest count of shootings, also had the greatest increase in 2020, reversing the downward trends from 2006 to 2019.
The highest counts of shooting victims have the following characteristics:
- gender: male
- ages: between 25-44 years old , with notable increases for 18-24 years old
- race: Black, with notable increases also for White Hispanic and Black Hispanic
In general, the 2020 increases returned shooting counts to approximately 2006 levels, except for the age group of 25-44.

Perpetrator characteristics

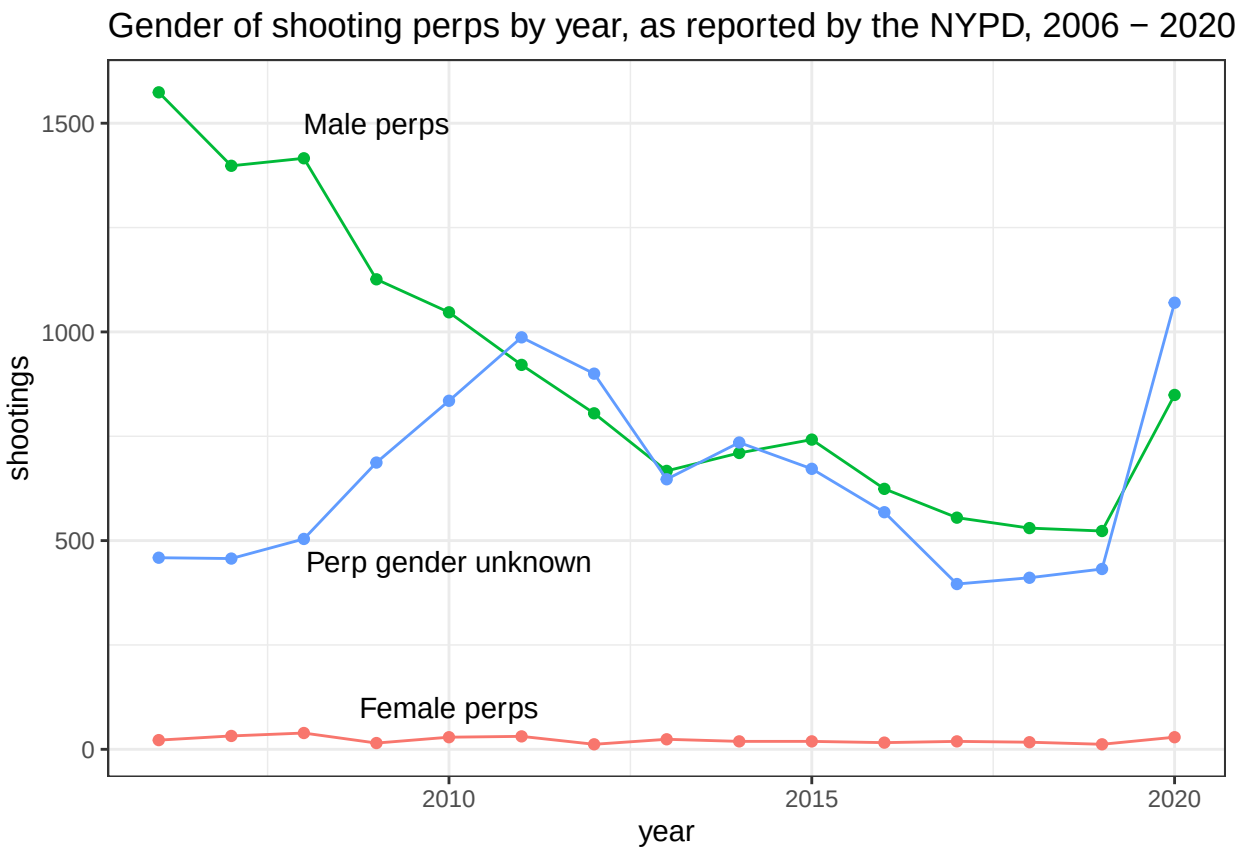
```
perp_age_count <- data_nypd %>%
  group_by(year) %>%
  count(PERP_AGE_GROUP) %>%
  rename(shootings = n)

ggplot(perp_age_count, aes(x = year, y = shootings, color = PERP_AGE_GROUP)) +
  geom_point() +
  geom_line() +
  labs(title = "Age of shooting perps by year, as reported by the NYPD, 2006 - 2020") +
  theme_bw()
```




```
view(perp_age_count)

perp_race_count <- data_nypd %>%
  group_by(year) %>%
  count(PERP_RACE) %>%
  rename(shootings = n)

ggplot(perp_race_count, aes(x = year, y = shootings, color = PERP_RACE)) +
  geom_point() +
  geom_line() +
  labs(title = "Race of shooting perps by year, as reported by the NYPD, 2006 - 2020") +
  theme_bw()
```




```
perp_sex_count <- data_nypd %>%
  group_by(year) %>%
  count(PERP_SEX) %>%
  rename(shootings = n)

ggplot(perp_sex_count, aes(x = year, y = shootings, color = PERP_SEX)) +
  geom_point() +
  geom_line() +
  annotate("text", x = 2010, y = 100, label = "Female perps") +
  annotate("text", x = 2009, y = 1500, label = "Male perps") +
  annotate("text", x = 2010, y = 450, label = "Perp gender unknown") +
  labs(title = "Gender of shooting perps by year, as reported by the NYPD, 2006 - 2020") +
  theme_bw() +
  theme(legend.position="None")
```



The characteristics of the shooting perpetrators shared many similarities to those of the victim population: 2020 saw an increase in the number of male perpetrators and perpetrators between the the ages of 25-44 and 18-24. The number of perpetrators identified as Black also increased, but only to the levels of 2011-2012, still roughly half as many as recorded in 2006. One notable feature of the increase in in shootings reported in 2020 is the sudden increase in the number count of characteristics that are *"UNKNOWN"*, especially for the race and age group. There could be many reasons for this, including a police force that was unprepared for a surge in the number of shootings, and other events, such as COVID and BLM movement, might have left the police force overwhelmed and/or understaffed. Moreover, the early months of COVID caused dramatic changes in social interactions, and the emergence of BLM movement influenced attitudes towards policing and trust in the police force.

Potential biases

In looking at these crime statistics, we need to consider whether there are biases present in who is reporting shooting crimes in NYC and how the police are treating these reports. Some groups may be reluctant to report crimes, and it is possible that not all reported crimes are given the same value. Secondly, there is usually pressure on police departments to reduce crime, which is recorded in this data (except for in 2020), but it is not clear what measures were taken to reduce shooting crime and if this affected the data collection. Finally, there was a dramatic change in the relationship between the police and the general public with the BLM movement in 2020, and it could be that changed behaviours in how crimes are reported by the general public.