NYPD Incident Shooting Data Analysis

Eric

5/27/2021

After years of a downward trend in shootings and deaths in New York City, why was there a spike in shootings and deaths in 2020?

Description of Data:

The data set I am analyzing is the NYPD Shooting Incident Data (Historic). The data was obtained from data.gov at this url: https://catalog.data.gov/dataset and it is the home of the U.S. Governments open data.

This data set has records on any incident where there was a shooting from 01/01/2006 - 12/31/2020. It also gives us data on:

- The victims race, sex, and age range.

- The perpetrators race, sex, and age range.

- The Borough & Precinct where it occurred.

- The Date & Time the shooting occurred.

- If the Shooting resulted in a death.

- Description of the location of the shooting.

- It also has data on: X_COORD_CD, Y_COORD_CD, Latitude, Longitude, Lon_Lat, JURISDICTION_CODE, and INCIDENT_KEY, but I didn't use any of this additional data and filtered it out early on.

Load packages we will use:

```
library(tidyverse)
```

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

```
## v ggplot2 3.3.3      v purrr   0.3.4
## v tibble  3.1.2      v dplyr   1.0.6
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1
```

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

```
library(lubridate)
```

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

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

```
library(readr)
library(dplyr)
library(ggplot2)
```

Load & View Data:

I downloaded the NYPD Shooting Incident Data (Historic) report from the URL provided in the Description of Data section in the report, and saved it as a csv on my desktop and imported it from there.: ***

```
NYPD_Shooting_Incident_Data_Historic <- read_csv("/Users/ericdoci/Desktop/NYPD_Shooting_Incident_Data__
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   INCIDENT_KEY = col_double(),
##   OCCUR_DATE = col_character(),
##   OCCUR_TIME = col_time(format = ""),
##   BORO = col_character(),
##   PRECINCT = col_double(),
##   JURISDICTION_CODE = col_double(),
##   LOCATION_DESC = col_character(),
##   STATISTICAL_MURDER_FLAG = col_logical(),
##   PERP_AGE_GROUP = col_character(),
##   PERP_SEX = col_character(),
##   PERP_RACE = col_character(),
##   VIC_AGE_GROUP = col_character(),
##   VIC_SEX = col_character(),
```

```
##     VIC_RACE = col_character(),
##     X_COORD_CD = col_number(),
##     Y_COORD_CD = col_number(),
##     Latitude = col_double(),
##     Longitude = col_double(),
##     Lon_Lat = col_character()
## )
```

NYPD_Shooting_Incident_Data_Historic

```
## # A tibble: 23,568 x 19
##     INCIDENT_KEY OCCUR_DATE OCCUR_TIME BORO          PRECINCT JURISDICTION_CODE
##            <dbl> <chr>      <time>     <chr>            <dbl>             <dbl>
## 1     201575314 08/23/2019 22:10      QUEENS             103                 0
## 2     205748546 11/27/2019 15:54      BRONX               40                 0
## 3     193118596 02/02/2019 19:40      MANHATTAN           23                 0
## 4     204192600 10/24/2019 00:52      STATEN ISLAND      121                 0
## 5     201483468 08/22/2019 18:03      BRONX               46                 0
## 6     198255460 06/07/2019 17:50      BROOKLYN            73                 0
## 7     194570529 03/11/2019 16:30      BROOKLYN            81                 0
## 8     203211777 10/03/2019 01:45      BROOKLYN            67                 0
## 9     193694863 02/17/2019 03:00      QUEENS             114                 2
## 10    199582060 07/10/2019 02:56      BROOKLYN            69                 0
## # ... with 23,558 more rows, and 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>
```

Filter out data for X&Y Coordinate, Latitude & Longitude, Lon&Lat, Jurisdictiion Code,, Incident Key, Occur Time, and Location Description:

```
nypd_shootings <- NYPD_Shooting_Incident_Data_Historic %>%
  select(-c(X_COORD_CD, Y_COORD_CD, Latitude, Longitude, Lon_Lat, JURISDICTION_CODE, INCIDENT_KEY, OCCU
```

nypd_shootings

```
## # A tibble: 23,568 x 11
##    OCCUR_DATE BORO    PRECINCT LOCATION_DESC STATISTICAL_MURD~ PERP_AGE_GROUP
##    <chr>      <chr>      <dbl> <chr>         <lgl>             <chr>
## 1 08/23/2019 QUEENS       103 <NA>          FALSE             <NA>
## 2 11/27/2019 BRONX         40 <NA>          FALSE             <18
## 3 02/02/2019 MANHAT~       23 <NA>          FALSE             18-24
## 4 10/24/2019 STATEN~      121 PVT HOUSE     TRUE              25-44
## 5 08/22/2019 BRONX         46 <NA>          FALSE             25-44
## 6 06/07/2019 BROOKL~       73 <NA>          FALSE             45-64
## 7 03/11/2019 BROOKL~       81 <NA>          FALSE             18-24
```

```
##  8 10/03/2019 BROOKL~        67 MULTI DWELL - A~ TRUE               <NA>
##  9 02/17/2019 QUEENS        114 MULTI DWELL - P~ FALSE              18-24
## 10 07/10/2019 BROOKL~        69 <NA>             FALSE              25-44
## # ... with 23,558 more rows, and 5 more variables: PERP_SEX <chr>,
## #   PERP_RACE <chr>, VIC_AGE_GROUP <chr>, VIC_SEX <chr>, VIC_RACE <chr>
```

```
summary(nypd_shootings)
```

```
##   OCCUR_DATE            BORO              PRECINCT       LOCATION_DESC
## Length:23568       Length:23568       Min.   :  1.00    Length:23568
## Class :character   Class :character   1st Qu.: 44.00    Class :character
## Mode  :character   Mode  :character   Median : 69.00    Mode  :character
##                                       Mean   : 66.21
##                                       3rd Qu.: 81.00
##                                       Max.   :123.00
## STATISTICAL_MURDER_FLAG PERP_AGE_GROUP        PERP_SEX
## Mode :logical           Length:23568       Length:23568
## FALSE:19080             Class :character   Class :character
## TRUE :4488              Mode  :character   Mode  :character
##
##
##
##   PERP_RACE         VIC_AGE_GROUP        VIC_SEX            VIC_RACE
## Length:23568       Length:23568       Length:23568       Length:23568
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
```

Change date from character value to date value:

```
nypd_shootings <- nypd_shootings %>%
    mutate(OCCUR_DATE = mdy(OCCUR_DATE))
```

```
nypd_shootings
```

```
## # A tibble: 23,568 x 11
##    OCCUR_DATE BORO    PRECINCT LOCATION_DESC    STATISTICAL_MURD~ PERP_AGE_GROUP
##    <date>     <chr>      <dbl> <chr>            <lgl>             <chr>
##  1 2019-08-23 QUEENS       103 <NA>             FALSE             <NA>
##  2 2019-11-27 BRONX         40 <NA>             FALSE             <18
##  3 2019-02-02 MANHAT~       23 <NA>             FALSE             18-24
##  4 2019-10-24 STATEN~      121 PVT HOUSE        TRUE              25-44
##  5 2019-08-22 BRONX         46 <NA>             FALSE             25-44
##  6 2019-06-07 BROOKL~       73 <NA>             FALSE             45-64
##  7 2019-03-11 BROOKL~       81 <NA>             FALSE             18-24
##  8 2019-10-03 BROOKL~       67 MULTI DWELL - A~ TRUE              <NA>
##  9 2019-02-17 QUEENS       114 MULTI DWELL - P~ FALSE             18-24
```

```
## 10 2019-07-10 BROOKL~         69 <NA>              FALSE            25-44
## # ... with 23,558 more rows, and 5 more variables: PERP_SEX <chr>,
## #   PERP_RACE <chr>, VIC_AGE_GROUP <chr>, VIC_SEX <chr>, VIC_RACE <chr>
```

```
summary(nypd_shootings)
```

```
##    OCCUR_DATE              BORO              PRECINCT      LOCATION_DESC
## Min.   :2006-01-01   Length:23568       Min.   :  1.00   Length:23568
## 1st Qu.:2008-12-30   Class :character   1st Qu.: 44.00   Class :character
## Median :2012-02-26   Mode  :character   Median : 69.00   Mode  :character
## Mean   :2012-10-03                      Mean   : 66.21
## 3rd Qu.:2016-02-28                      3rd Qu.: 81.00
## Max.   :2020-12-31                      Max.   :123.00
## STATISTICAL_MURDER_FLAG PERP_AGE_GROUP       PERP_SEX
## Mode :logical           Length:23568       Length:23568
## FALSE:19080             Class :character   Class :character
## TRUE :4488              Mode  :character   Mode  :character
##
##
##
##   PERP_RACE          VIC_AGE_GROUP        VIC_SEX            VIC_RACE
## Length:23568       Length:23568       Length:23568       Length:23568
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
```

Separate date so that Month, Date, and Year each have their own column:

```
nypd_shootings <- nypd_shootings %>% separate(OCCUR_DATE, c("Year", "Month", "Day"))
```

```
nypd_shootings
```

```
## # A tibble: 23,568 x 13
##    Year  Month Day   BORO       PRECINCT LOCATION_DESC        STATISTICAL_MURDER~
##    <chr> <chr> <chr> <chr>         <dbl> <chr>                <lgl>
##  1 2019  08    23    QUEENS          103 <NA>                 FALSE
##  2 2019  11    27    BRONX            40 <NA>                 FALSE
##  3 2019  02    02    MANHATTAN        23 <NA>                 FALSE
##  4 2019  10    24    STATEN IS~      121 PVT HOUSE            TRUE
##  5 2019  08    22    BRONX            46 <NA>                 FALSE
##  6 2019  06    07    BROOKLYN         73 <NA>                 FALSE
##  7 2019  03    11    BROOKLYN         81 <NA>                 FALSE
##  8 2019  10    03    BROOKLYN         67 MULTI DWELL - APT ~  TRUE
##  9 2019  02    17    QUEENS          114 MULTI DWELL - PUBL~  FALSE
## 10 2019  07    10    BROOKLYN         69 <NA>                 FALSE
```

```
## # ... with 23,558 more rows, and 6 more variables: PERP_AGE_GROUP <chr>,
## #   PERP_SEX <chr>, PERP_RACE <chr>, VIC_AGE_GROUP <chr>, VIC_SEX <chr>,
## #   VIC_RACE <chr>
```

```
summary(nypd_shootings)
```

```
##       Year              Month              Day              BORO
##  Length:23568      Length:23568      Length:23568      Length:23568
##  Class :character  Class :character  Class :character  Class :character
##  Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
##     PRECINCT       LOCATION_DESC     STATISTICAL_MURDER_FLAG PERP_AGE_GROUP
##  Min.   :  1.00   Length:23568      Mode :logical           Length:23568
##  1st Qu.: 44.00   Class :character  FALSE:19080             Class :character
##  Median : 69.00   Mode  :character  TRUE :4488              Mode  :character
##  Mean   : 66.21
##  3rd Qu.: 81.00
##  Max.   :123.00
##    PERP_SEX          PERP_RACE         VIC_AGE_GROUP        VIC_SEX
##  Length:23568      Length:23568      Length:23568      Length:23568
##  Class :character  Class :character  Class :character  Class :character
##  Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
##    VIC_RACE
##  Length:23568
##  Class :character
##  Mode  :character
##
##
##
```

View shootings per year in NYC:

```
shootings_year_ny <- nypd_shootings %>%
  group_by(Year) %>%
  summarise_at(vars(STATISTICAL_MURDER_FLAG), list(Shootings = length))
```

```
shootings_year_ny
```

```
## # A tibble: 15 x 2
##    Year  Shootings
##    <chr>     <int>
## 1 2006       2055
## 2 2007       1887
## 3 2008       1958
```

```
##  4 2009        1828
##  5 2010        1910
##  6 2011        1939
##  7 2012        1717
##  8 2013        1339
##  9 2014        1464
## 10 2015        1434
## 11 2016        1208
## 12 2017         969
## 13 2018         951
## 14 2019         967
## 15 2020        1942
```

View deaths per Year in NYC:

```
deaths_year_ny <- nypd_shootings %>%
  group_by(Year) %>%
  summarise(Deaths = sum(STATISTICAL_MURDER_FLAG == "TRUE"))

deaths_year_ny
```

```
## # A tibble: 15 x 2
##     Year  Deaths
##     <chr>  <int>
##  1 2006      445
##  2 2007      373
##  3 2008      362
##  4 2009      348
##  5 2010      403
##  6 2011      373
##  7 2012      287
##  8 2013      223
##  9 2014      248
## 10 2015      283
## 11 2016      223
## 12 2017      174
## 13 2018      202
## 14 2019      183
## 15 2020      361
```

Merge Shootings and Deaths per year:

```
shooting_death_year_ny <- shootings_year_ny %>%
  full_join(deaths_year_ny)
```

```
## Joining, by = "Year"
```

```
shooting_death_year_ny
```

```
## # A tibble: 15 x 3
##     Year  Shootings Deaths
##     <chr>     <int>  <int>
##  1 2006       2055    445
##  2 2007       1887    373
##  3 2008       1958    362
##  4 2009       1828    348
##  5 2010       1910    403
##  6 2011       1939    373
##  7 2012       1717    287
##  8 2013       1339    223
##  9 2014       1464    248
## 10 2015       1434    283
## 11 2016       1208    223
## 12 2017        969    174
## 13 2018        951    202
## 14 2019        967    183
## 15 2020       1942    361
```

Graph Shootings & Deaths Per Year in NYC:

Set x = as.numeric(Year) so that the line can be added be added to the plots. ***

```
shooting_death_year_ny %>%
  ggplot(aes(x = as.numeric(Year), y = Shootings)) +
  geom_point(aes(color = "Shootings")) +
  geom_line(aes(color = "Shootings")) +
  geom_point(aes(color = "Shootings")) +
  geom_line(aes(y = Deaths, color = "Deaths")) +
  geom_point(aes(y = Deaths, color = "Deaths")) +
  theme(legend.position="bottom", axis.text.x = element_text(angle = 90)) +
  labs(title = "Shootings/Year & Deaths/Year in NYC", y=NULL)
```

Shootings/Year & Deaths/Year in NYC
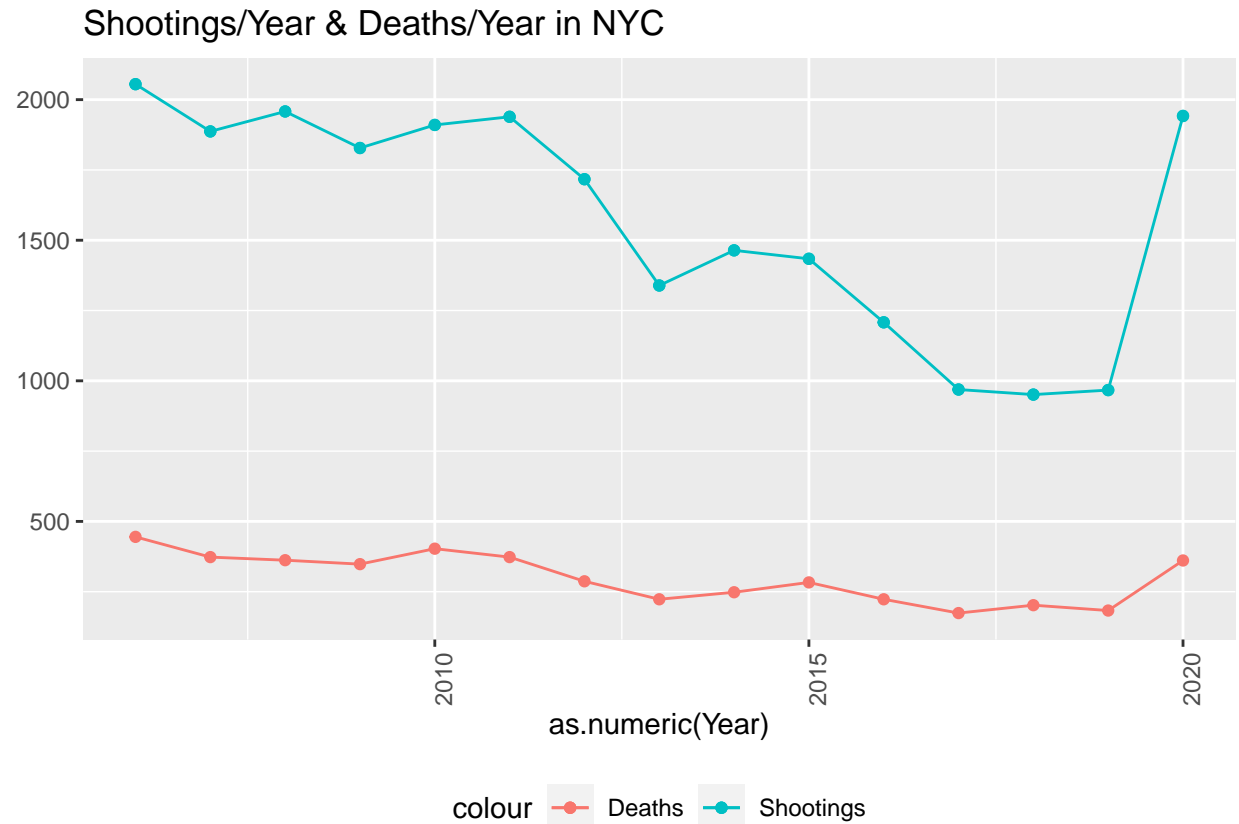
Analysis of Shootings & Deaths Per Year in NYC:

When looking at the graph of Shootings/Year and Deaths/Year in New York City, one can notice a rather consistent downward trend in the number of shooting incidents and deaths from 2006 to 2019. The most surprising thing we notice on the graph is the drastic increase (~100%) in shootings and deaths in the year 2020 compared to the previous years. A couple of other notable observations are the relatively significant decreases in shootings and deaths in 2013 compared to 2012, i.e. ~22% and ~17% respectively.

View shootings per year by BORO ("BRONX", "BROOKLYN", "STATEN ISLAND", "MANHATTAN", "QUEENS"):

```
shootings_year_boro <- nypd_shootings %>%
  group_by(BORO, Year) %>%
  summarise_at(vars(STATISTICAL_MURDER_FLAG), list(Shootings = length))

view(shootings_year_boro)
```

View deaths PER Year by BORO ("BRONX", "BROOKLYN", "STATEN ISLAND", "MANHATTAN", "QUEENS"):

```
deaths_year_boro <- nypd_shootings %>%
  group_by(BORO, Year) %>%
  summarise(Deaths = sum(STATISTICAL_MURDER_FLAG == "TRUE"))
```

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

```
View(deaths_year_boro)
```

Merge Shootings and Death per year by Boro ("BRONX", "BROOKLYN", "STATEN ISLAND", "MANHATTAN", "QUEENS"):

```
shooting_death_year_boro <- shootings_year_boro %>%
  full_join(deaths_year_boro)
```

```
## Joining, by = c("BORO", "Year")
```

```
view(shooting_death_year_boro)
```

```
summary(shooting_death_year_boro)
```

```
##      BORO              Year              Shootings          Deaths
## Length:75          Length:75          Min.   : 25.0   Min.   :   4.00
## Class :character   Class :character   1st Qu.:143.0   1st Qu.: 23.00
## Mode  :character   Mode  :character   Median :266.0   Median : 48.00
##                                       Mean   :314.2   Mean   : 59.84
##                                       3rd Qu.:502.0   3rd Qu.: 85.50
##                                       Max.   :850.0   Max.   :182.00
```

Graph Shootings & Deaths Per Year by BORO ("BRONX", "BROOKLYN", "STATEN ISLAND", "MANHATTAN", "QUEENS"):

Graph Shootings/Year & Deaths/Year in the Bronx:

10

```
boro <- "BRONX"
shooting_death_year_boro %>%
  filter(BORO == boro) %>%
  ggplot(aes(x = as.numeric(Year), y = Shootings)) +
  geom_line(aes(color = "Shootings")) +
  geom_point(aes(color = "Shootings")) +
  geom_line(aes(y = Deaths, color = "Deaths")) +
  geom_point(aes(y = Deaths, color = "Deaths")) +
  theme(legend.position="bottom", axis.text.x = element_text(angle = 90)) +
  labs(title = "Shootings/Year & Deaths/Year in Bronx", y=NULL)
```
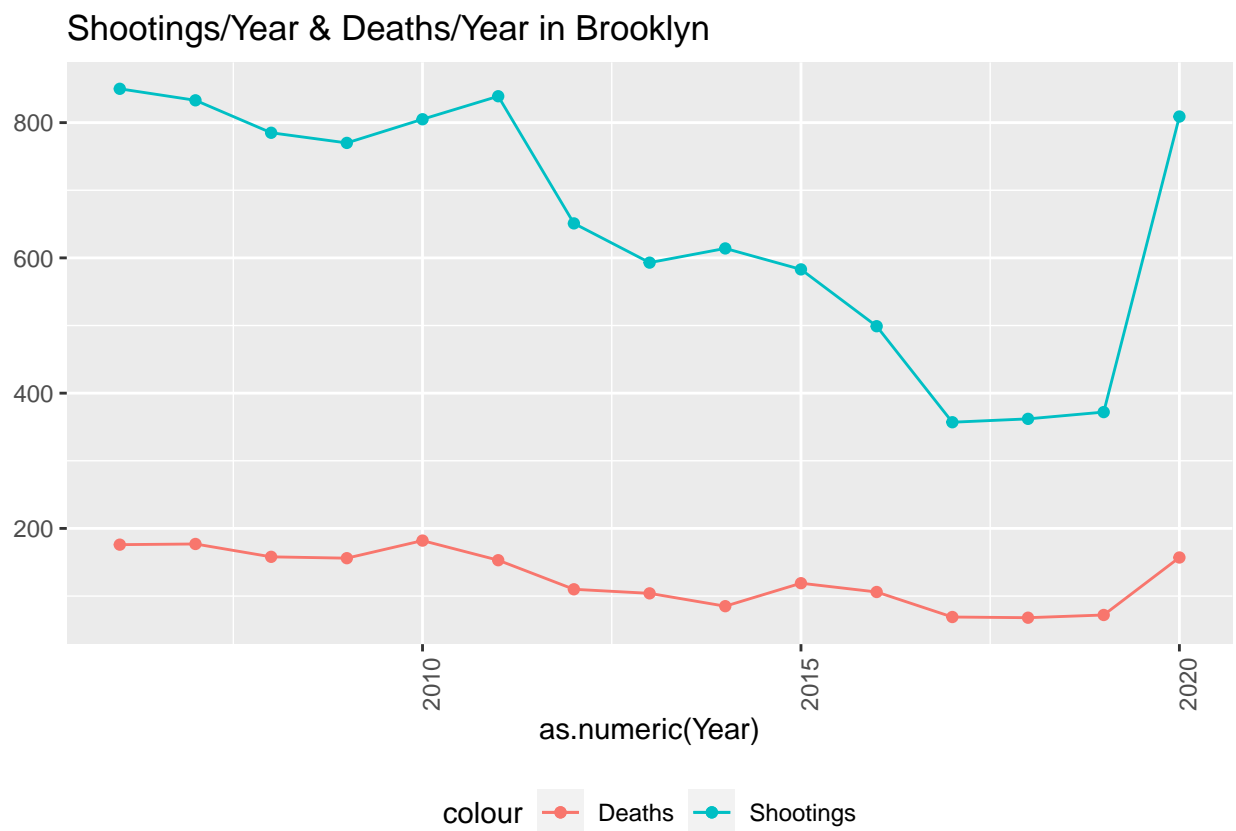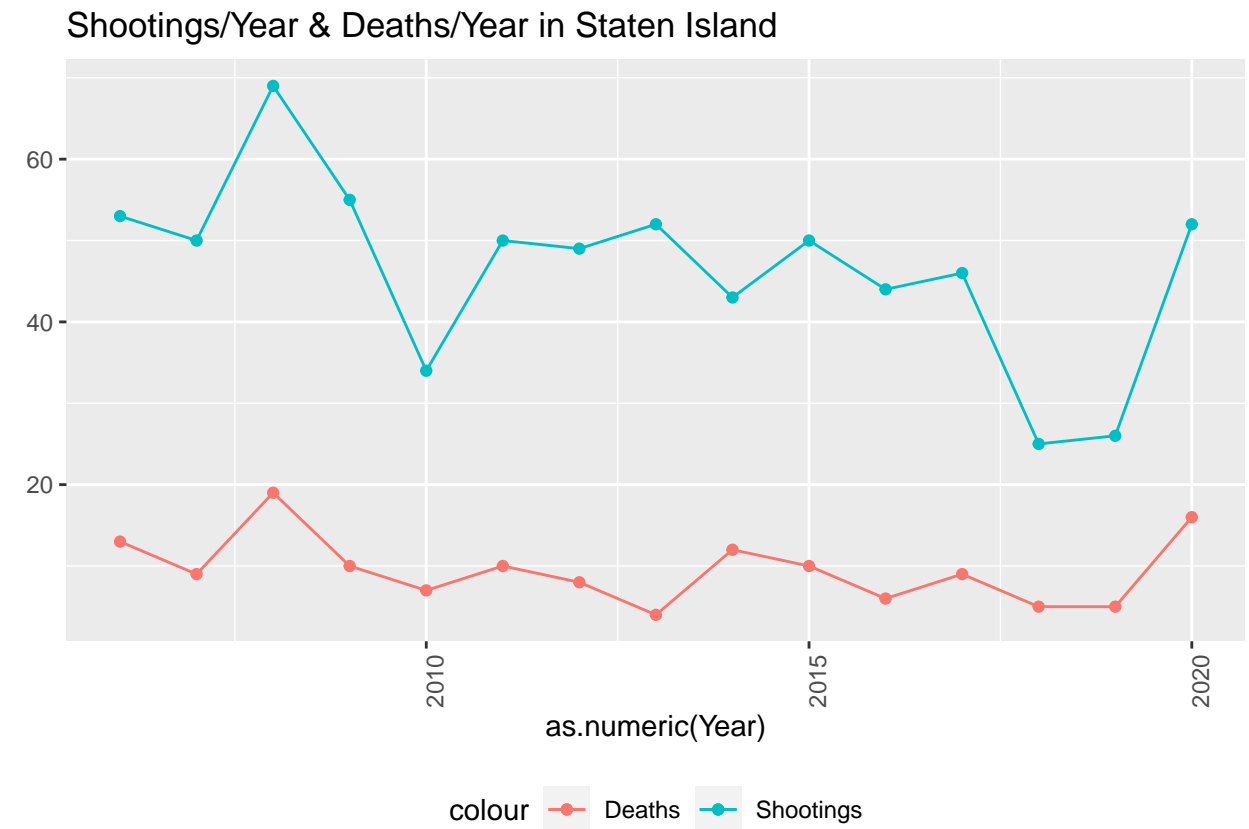

Shootings/Year & Deaths/Year in Bronx
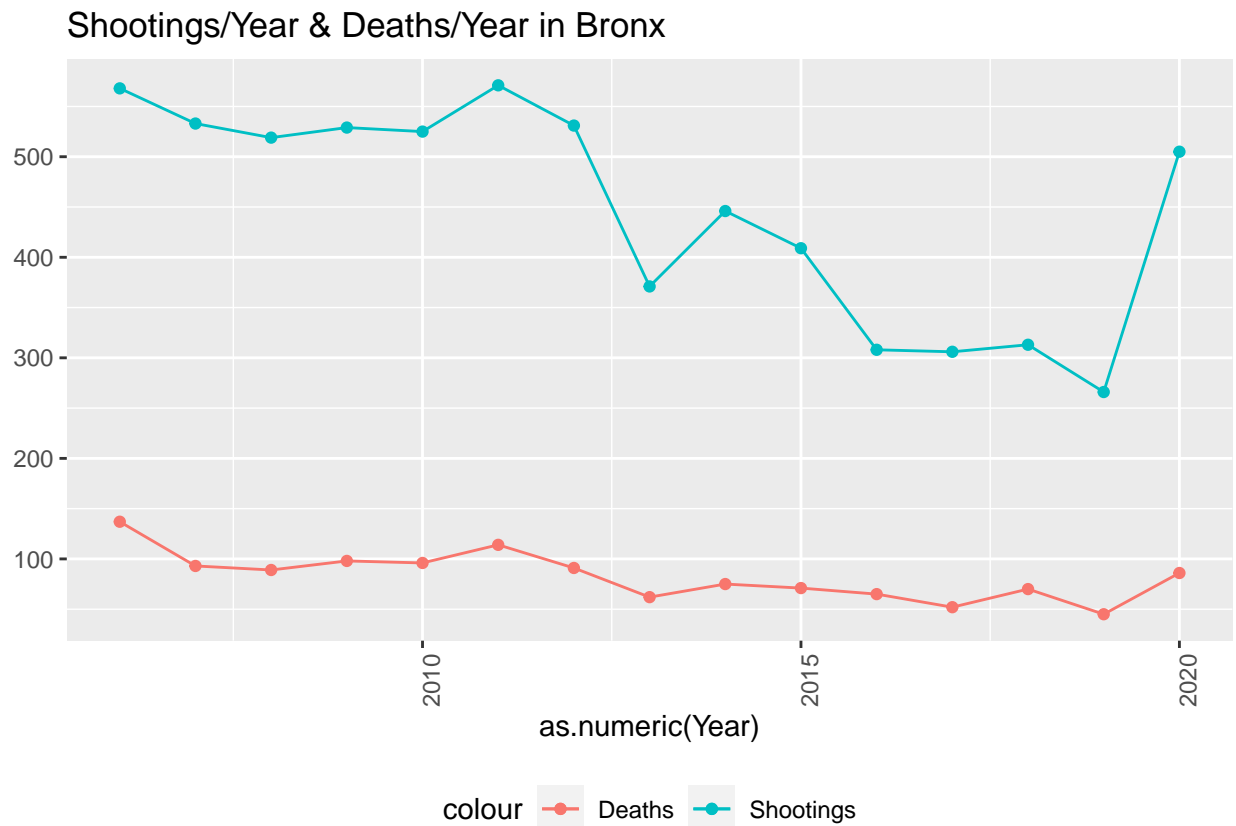
Analysis of Shootings/Year & Deaths/Year in the Bronx:

The graph of Shootings/Year and Deaths/Year for the Bronx shows a rather consistent downward trend in shootings and deaths per year very similar to what we saw in the graph for all of NYC. Also, similar to what we see in the NYC graph, we also notice a major rise in 2020 and a relative drop in 2013 for the Bronx.

Graph Shootings/Year & Deaths/Year in Brooklyn:

```
boro <- "BROOKLYN"
shooting_death_year_boro %>%
  filter(BORO == boro) %>%
  ggplot(aes(x = as.numeric(Year), y = Shootings)) +
  geom_line(aes(color = "Shootings")) +
  geom_point(aes(color = "Shootings")) +
  geom_line(aes(y = Deaths, color = "Deaths")) +
  geom_point(aes(y = Deaths, color = "Deaths")) +
  theme(legend.position="bottom", axis.text.x = element_text(angle = 90)) +
  labs(title = "Shootings/Year & Deaths/Year in Brooklyn", y=NULL)
```


Shootings/Year & Deaths/Year in Brooklyn

Analysis of Shootings/Year & Deaths/Year in Brooklyn:

Brooklyn has the most shootings and deaths per year compared to all the other boroughs. The Brooklyn graph of Shootings/Year and Deaths/Year shows a rather consistent downward trend in shootings and deaths a year like in the NYC and Bronx graphs. The same jump in Shootings and Deaths in 2020 can again be noticed, but for Brooklyn, we see a large drop in shootings and deaths in 2012 instead of 2013, which was the year that showed a notable drop for NYC and Bronx.

Graph Shootings/Year & Deaths/Year in Staten Island:

```r
boro <- "STATEN ISLAND"
shooting_death_year_boro %>%
  filter(BORO == boro) %>%
  ggplot(aes(x = as.numeric(Year), y = Shootings)) +
  geom_line(aes(color = "Shootings")) +
  geom_point(aes(color = "Shootings")) +
  geom_line(aes(y = Deaths, color = "Deaths")) +
  geom_point(aes(y = Deaths, color = "Deaths")) +
  theme(legend.position="bottom", axis.text.x = element_text(angle = 90)) +
  labs(title = "Shootings/Year & Deaths/Year in Staten Island", y=NULL)
```


Analysis of Shootings/Year & Deaths/Year in Staten Island:

Staten Island has the least amount of shootings and deaths per year compared to all the other boroughs. We notice a very slight downward trend in shootings and deaths per year in Staten Island, but the data is

rather sporadic and inconsistent. For Staten Island, the shootings and deaths in 2020 seem rather consistent with the general overall trend, but we notice a relatively large increase in shootings from 2007 to 2008 and decreases from 2009 to 2010 and 2017 to 2018.

Graph Shootings/Year & Deaths/Year in Manhattan:

```
boro <- "MANHATTAN"
shooting_death_year_boro %>%
  filter(BORO == boro) %>%
  ggplot(aes(x = as.numeric(Year), y = Shootings)) +
  geom_line(aes(color = "Shootings")) +
  geom_point(aes(color = "Shootings")) +
  geom_line(aes(y = Deaths, color = "Deaths")) +
  geom_point(aes(y = Deaths, color = "Deaths")) +
  theme(legend.position="bottom", axis.text.x = element_text(angle = 90)) +
  labs(title = "Shootings/Year & Deaths/Year in Manhattan", y=NULL)
```
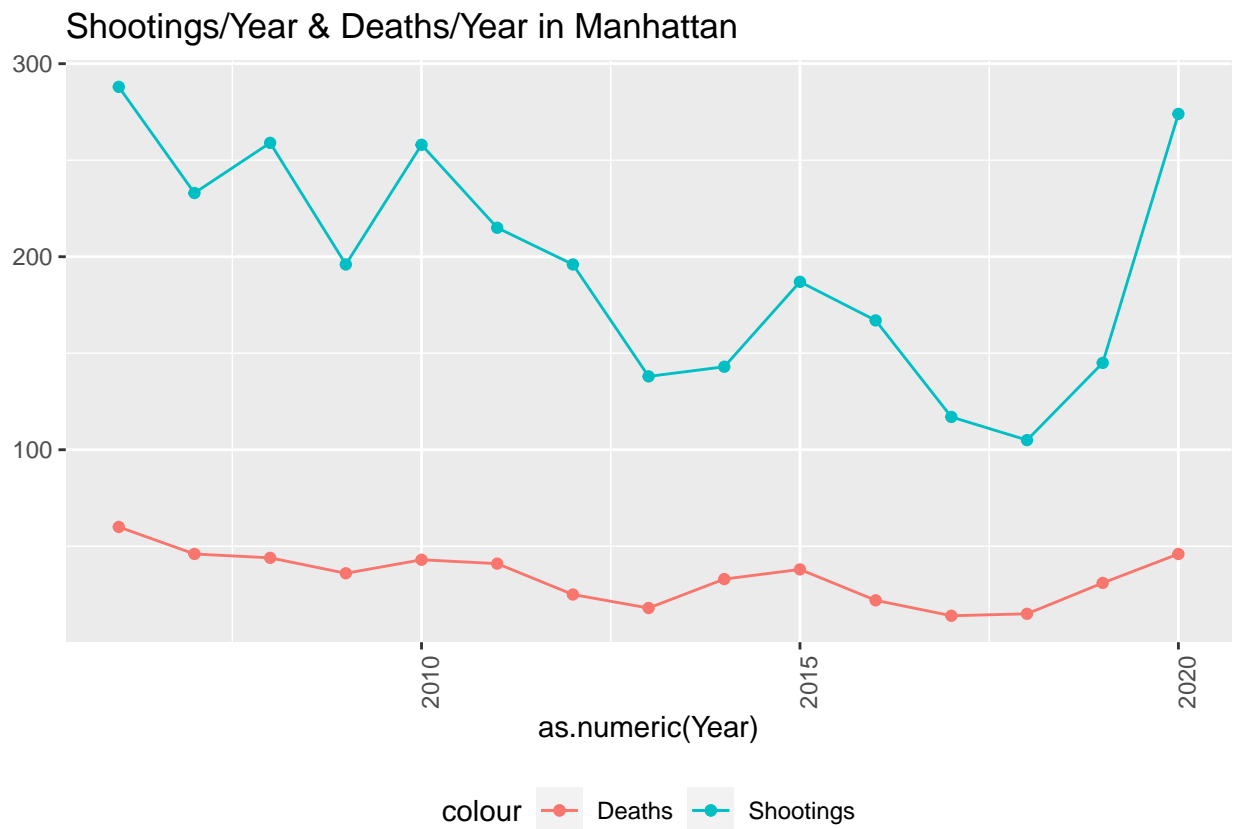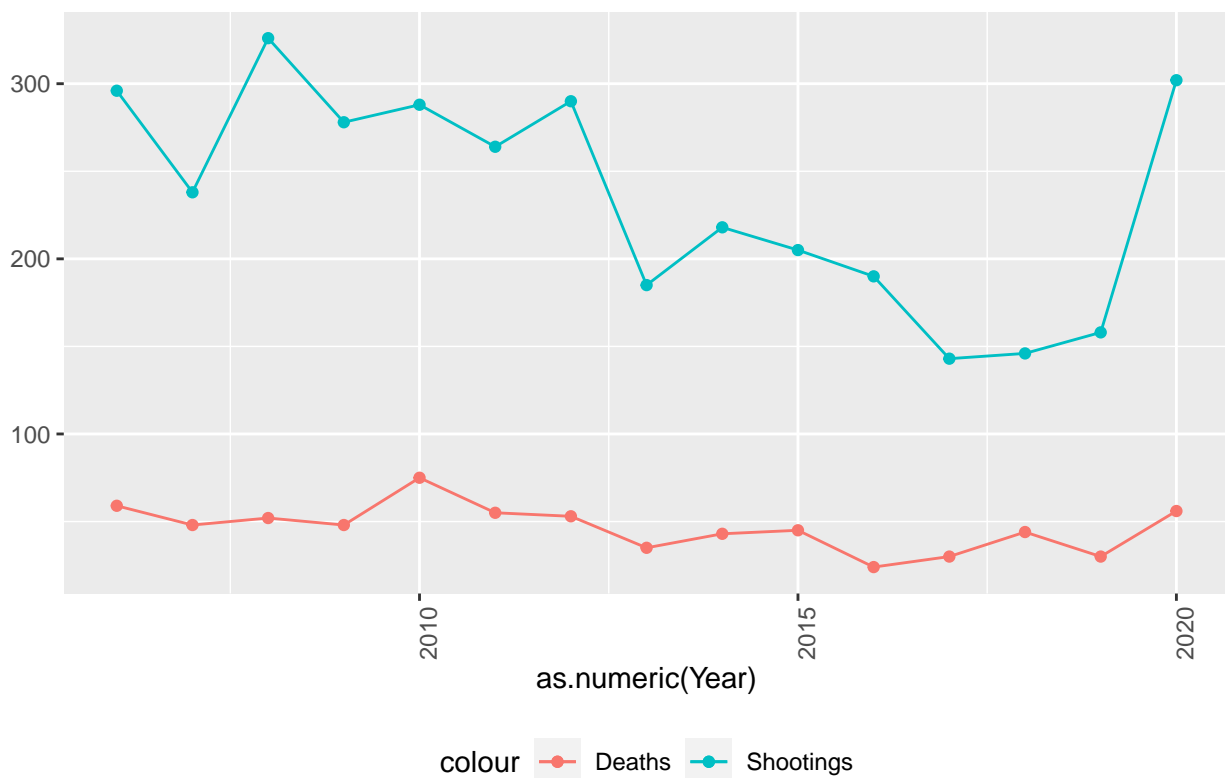
Analysis of Shootings/Year & Deaths/Year in Manhattan:

The graph of shootings and deaths per year in Manhattan is a little more sporadic compared to the graphs for The Bronx and Brooklyn, but we can still notice the same type of consistent downward trend here. For Manhattan, we notice a substantial increase in shootings from 2018 to 2019 and then making an even bigger jump from 2019 to 2020. We can also notice rather substantial drops from 2012 to 2013/14 and 2016 to 2017/18.

Graph Shootings/Year & Deaths/Year in Queens:

```
boro <- "QUEENS"
shooting_death_year_boro %>%
  filter(BORO == boro) %>%
  ggplot(aes(x = as.numeric(Year), y = Shootings)) +
  geom_line(aes(color = "Shootings")) +
  geom_point(aes(color = "Shootings")) +
  geom_line(aes(y = Deaths, color = "Deaths")) +
  geom_point(aes(y = Deaths, color = "Deaths")) +
  theme(legend.position="bottom", axis.text.x = element_text(angle = 90)) +
  labs(title = "Shootings/Year & Deaths/Year in Queens", y=NULL)
```


Shootings/Year & Deaths/Year in Queens

Analysis of Shootings/Year & Deaths/Year in Queens:

Lastly, we have the graph for Queens. This graph also shows a consistent downward trend like the NYC and the other boroughs, except for Staten Island. Like in the other graphs, we again notice the large increase in shootings from 2019 to 2020 as well as a notable drop from 2012 to 2013. Of interest are also the rather large drop from 2006 to 2007 and the the jump from 2007 to 2008, which surpasses the number of shootings in 2006.

Create Linear Model for Deaths as a function of Shootings

```
mod <- lm(Deaths ~ Shootings, data = shooting_death_year_ny)

summary(mod)
```

```
##
## Call:
## lm(formula = Deaths ~ Shootings, data = shooting_death_year_ny)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -42.364 -16.069  -1.058  10.493  45.708
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -25.86201   27.32092  -0.947    0.361
## Shootings     0.20689    0.01688  12.260 1.62e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.52 on 13 degrees of freedom
## Multiple R-squared:  0.9204, Adjusted R-squared:  0.9143
## F-statistic: 150.3 on 1 and 13 DF,  p-value: 1.616e-08
```

View Shootings Min & Max

```
shooting_death_year_ny %>% slice_min(Shootings)
```

```
## # A tibble: 1 x 3
##   Year  Shootings Deaths
##   <chr>     <int>  <int>
## 1 2018       951    202
```

```
shooting_death_year_ny %>% slice_max(Shootings)
```

```
## # A tibble: 1 x 3
##    Year  Shootings Deaths
##    <chr>     <int>  <int>
## 1 2006       2055    445
```

Create a grid from 900 (a little below our shootings min) and 2100 (a little above our shootings max)

```
x_grid <- seq(900, 2100)

new_df <- tibble(Shootings = x_grid)
```

# Add predictive linear model to shootings and deaths per year in NYC

```
shooting_death_year_ny_w_pred <- shooting_death_year_ny %>% mutate(pred = predict(mod))

view(shooting_death_year_ny_w_pred)
```

Plot predictive linear model and shootings vs deaths per year in NYC

```
shooting_death_year_ny_w_pred %>%
  ggplot() +
  geom_point(aes(x = Shootings, y = Deaths), color = "blue") +
  geom_point(aes(x = Shootings, y = pred), color = "red") +
  geom_line(aes(x = Shootings, y = pred), color = "red")
```
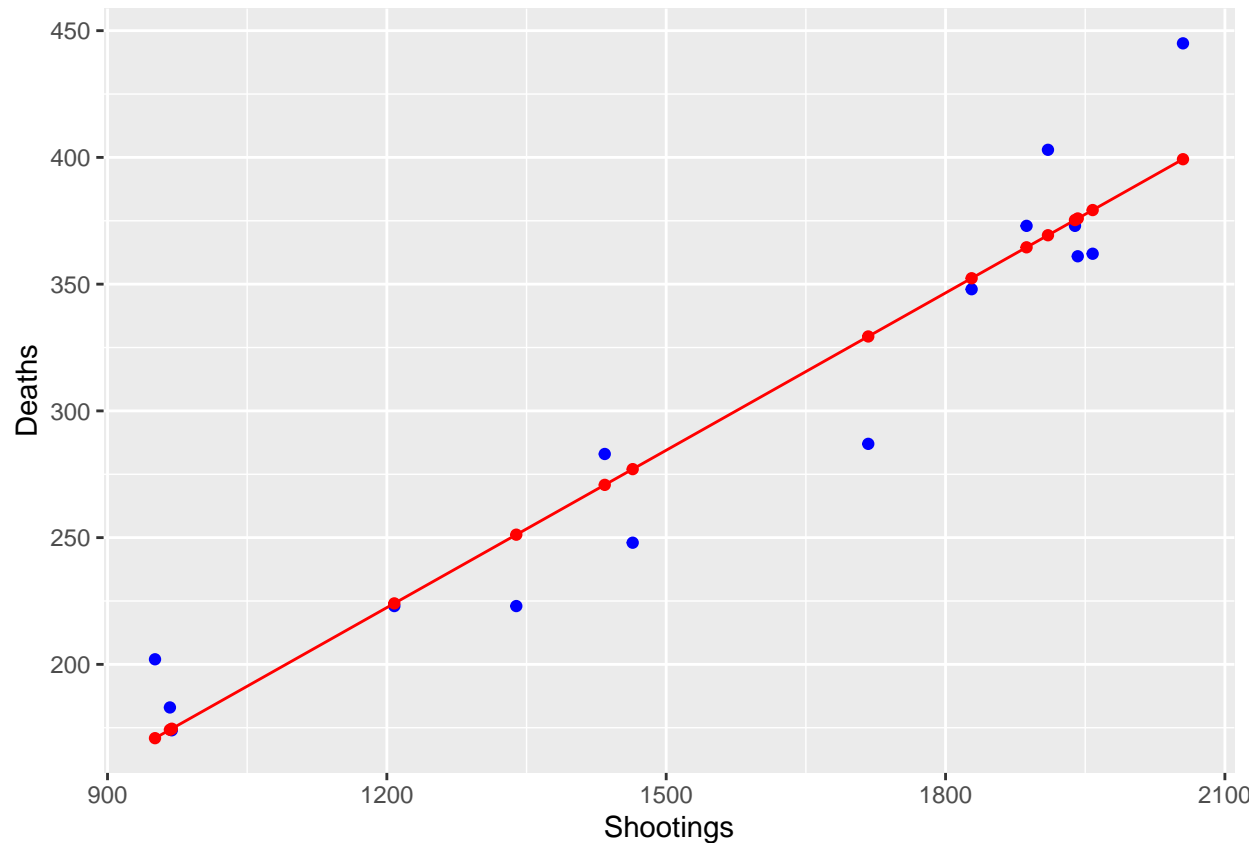
Shootings vs Deaths with pred Analysis:

When looking at the graphs for Shootings vs Deaths, there is a pretty steady linear correlation between the two. We can notice this by just looking at the graph with the predictive model. When we look at R-squared (0.9204) and Adjusted R-squared (0.9143), as well as a P-value of (1.616e-08), they reinforce the direct correlation we see between shootings in a year and deaths in a year.

Compare Shootings per month in 2019 and 2020 to see if there are insights for the 2020 spike

```
shootings_month_ny_19_20 <- nypd_shootings %>%
  filter(Year > 2018) %>%
  group_by(Month, Year) %>%
  summarise_at(vars(STATISTICAL_MURDER_FLAG), list(Shootings = length))

shootings_month_ny_19_20
```

```
## # A tibble: 24 x 3
## # Groups:   Month [12]
##    Month Year  Shootings
##    <chr> <chr>     <int>
##  1 01    2019         63
##  2 01    2020         94
##  3 02    2019         59
##  4 02    2020         51
##  5 03    2019         65
##  6 03    2020         71
##  7 04    2019         67
##  8 04    2020         71
##  9 05    2019         81
## 10 05    2020        138
## # ... with 14 more rows
```

Deaths per month 2019 and 2020 to see if there are insights for the 2020 spike

---

```
deaths_month_ny_19_20 <- nypd_shootings %>%
  filter(Year > 2018) %>%
  group_by(Month, Year) %>%
  summarise(Deaths = sum(STATISTICAL_MURDER_FLAG == "TRUE"))
```

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

```
View(deaths_month_ny_19_20)
```

Merge Shootings & Deaths per month in 2019 and 2020

```
shooting_death_month_ny_19_20 <- shootings_month_ny_19_20 %>%
  full_join(deaths_month_ny_19_20)
```

```
## Joining, by = c("Month", "Year")
```

```
view(shooting_death_month_ny_19_20)
```
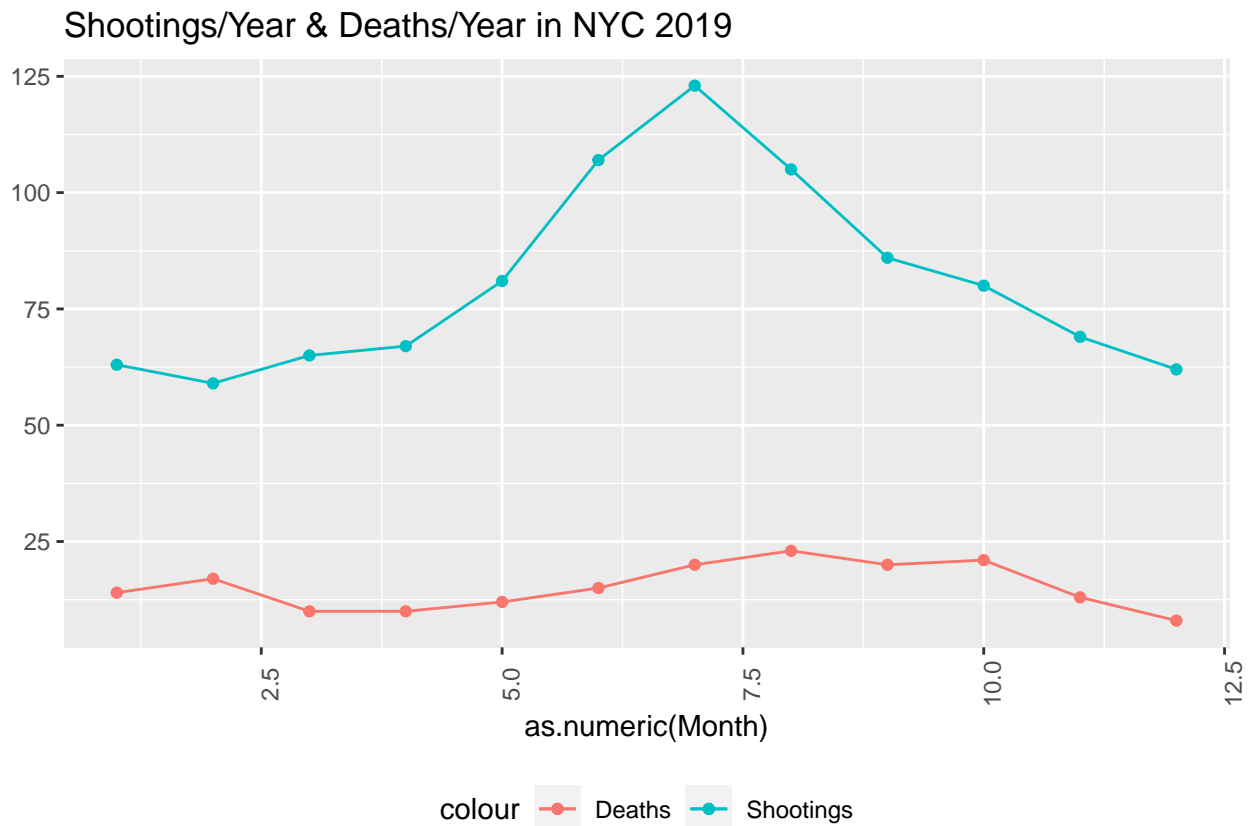
Plot shootings/deaths per month in NYC in 2019

```
year <- "2019"
shooting_death_month_ny_19_20 %>%
  filter(Year == year) %>%
  ggplot(aes(x = as.numeric(Month), y = Shootings)) +
  geom_line(aes(color = "Shootings")) +
  geom_point(aes(color = "Shootings")) +
  geom_line(aes(y = Deaths, color = "Deaths")) +
  geom_point(aes(y = Deaths, color = "Deaths")) +
  theme(legend.position="bottom", axis.text.x = element_text(angle = 90)) +
  labs(title = "Shootings/Year & Deaths/Year in NYC 2019", y=NULL)
```


Plot shootings/deaths per month in NYC in 2020

```
year <- "2020"
shooting_death_month_ny_19_20 %>%
  filter(Year == year) %>%
  ggplot(aes(x = as.numeric(Month), y = Shootings)) +
  geom_line(aes(color = "Shootings")) +
  geom_point(aes(color = "Shootings")) +
  geom_line(aes(y = Deaths, color = "Deaths")) +
  geom_point(aes(y = Deaths, color = "Deaths")) +
```

```
theme(legend.position="bottom", axis.text.x = element_text(angle = 90)) +
labs(title = "Shootings/Year & Deaths/Year in NYC 2020", y=NULL)
```

Shootings/Year & Deaths/Year in NYC 2020
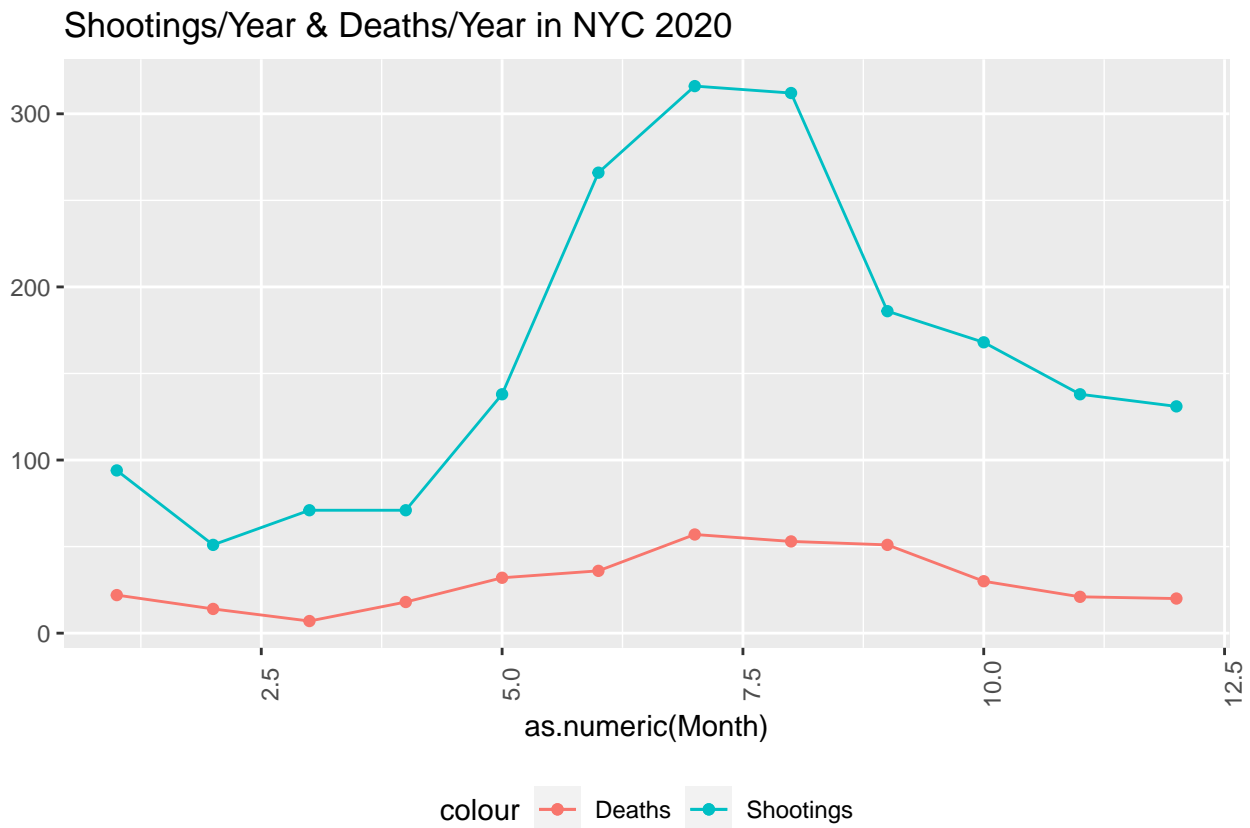
Analysis of Shootings/Month & Deaths/Month for 2019/2020

When looking at the monthly shooting data for 2019 and 2020, we will see that shootigns per month in January-April of 2020 are consistent with the shootings per month in the previous 4-month period, i.e. Sep-Dec 2019. We also notice that January-April of 2019 is also consistent with the shooting numbers in 2020. The spike in shootings & deaths starts in May of 2020 (just about a month after the US began its Covid lock-down) and it continues pretty much throughout the remainder of the year. ***

Conclusion:

After creating and analyzing all of the visualizations and models, it is clear that all of the New York City boroughs, except for Staten Island, are following the same general downward trend with a large spike in shootings from from 2019 to 2020, specifically from 967 shootings in 2019 to 1942 shootings in 2020. Similarly, all boroughs except Staten Island, experienced a rather significant drop in shootings from 2012 (1717) to 2013 (1339).

It was surprising to see that, after 14 years of shootings and deaths trending downward in New York City, the spike in 2020 was so significant that it was about twice higher than 2019 and it appears to have wiped out years of continuous decerease in shootings & deaths. I think there could be many factors to attribute to this increase in shootings, but the two most relevant ones to me are Covid and the BLM protests that occurred last year.

I think Covid played a huge role in the rise in shootings from 2019 to 2020. Due to Covid, millions of Americans lost their jobs. When looking at the monthly shooting data for 2019 and 2020, we will see that shootigns per month in January-April of 2020 are consistent with the shootings per month in the previous 4-month period, i.e. Sep-Dec 2019. The spike in shootings & deaths starts in May of 2020 (just about a month after the US began its lockdown) and it continues pretty much throughout the remainder of the year. Around this same time, people started losing their jobs and were panicking about how to provide for their families. Also, the people most affected by the lockdown were lower-income families, because of the shutdown of restaurants, stores, and other similar businesses that employed lower income individuals.

I also believe that another factor that explains the increase in shootings in 2020 is the unrest associated with the Black Lives Matter protests. After the death of George Floyd last year, the nation responded by organizing protests all across the US and some of these were turned violent and may have also lead to some of the increase in shootings.

In regards to the significant dip in 2013, I do not have any insights as to what may have caused these results, but the random drop from form 2012 to 2013 and then jump back from 2013 to 2014, but it does raise some questions.

In conclusion, I believe that New York will see a substantial decline in shootings in the year 2021. 2020 was such a crazy year, millions of people lost their jobs, mental health was a bigger issue than ever, and no one had ever experienced anything like that before. 2020 was a result of insane circumstances and I am very interested to see what the data for 2021 will show. Also, there is some missing information in the data, which if available, could have helped with a better analysis and insights, e.g. is there a possible one-to-many relationship in the data for shootings vs. deaths and how are mass shootings counted?

Possible Sources of bias:

A couple of possible sources of bias for me personally are that I am more liberal and I am in favor of gun control legislation and behind that lies my bias and belief that relaxed gun laws contribute to shootings and deaths. When it comes to my political standing, I am more Liberal and that could be the reason I mentioned the BLM protests as a cause for the increase in shootings from 2019-2020 in NYC. I do see how that could be a biased claim, but I made that claim solely because of the timing of the protests and how they correlated directly to the summer when the shooting numbers for 2020 showed an increase between 140% and 200% compared to 2019.

Also, the fact that I am against gun violence is another thing that can make me have certain opinions and affect how I view the data. When I was reviewing the data, I did my best to leave out any personal feelings and solely make predictions based on the data and knowledge I have of the past year. Lastly, attributing the increase in shooting in 2020 to Covid could also be a source of bias, but I believe that the data truly shows this correlation and my predictions are based on sound observations.