NYPD shooting

Ernesto Medina

2022-06-06

Import libraries

NYPD shooting incident data

- What neighborhood has more incidents?
- Who is more affected?
- Are shootings slowing down?

###Data source:

https://catalog.data.gov/dataset/nypd-shooting-incident-data-historic.

Get url for data:

```
url_in <- "https://data.cityofnewyork.us/api/views/833y-fsy8/"
file_name <- c("rows.csv?accessType=DOWNLOAD")
urls <- str_c(url_in, file_name)
urls
```

```
## [1] "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
```

Import data and print

```
incidents <- read_csv(urls[1])
incidents
```

```
## # A tibble: 23,585 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
##  7     73057167 06/05/2010 21:16      BROOKLYN       77                  0
##  8    211362213 03/20/2020 21:27      BROOKLYN       81                  0
##  9    137564752 07/04/2014 00:25      QUEENS        101                  0
```

```
## 10    147024011 10/18/2015 01:33      QUEENS        106                 0
## # ... with 23,575 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>
```

Data Description

In this document we are presenting data from NYPD data source. We will go through a process of visual analysis and modeling. As we can see in the summary:

- Data collected daily 2006-2020.
- The data consists of 23585 rows in the dataset.
- Column names are: INCIDENT_KEY, OCCUR_DATE, OCCUR_TIME, BORO, PRECINCT, JURISDICTION_CODE, LOCATION_DESC, STATISTICAL_MURDER_FLAG, PERP_AGE_GROUP, PERP_SEX, PERP_RACE, VIC_AGE_GROUP, VIC_SEX, VIC_RACE, X_COORD_CD, Y_COORD_CD, Latitude, Longitude, Lon_Lat
- STATISTICAL_MURDER_FLAG and INCIDENT_KEY will be used to differentciate between murders/non murder incidents (TRUE/FALSE flag) and individual/multiple victims (incident key).

```
summary(incidents)
```

```
##   INCIDENT_KEY        OCCUR_DATE         OCCUR_TIME           BORO
##  Min.   :  9953245   Length:23585       Length:23585      Length:23585
##  1st Qu.: 55322804   Class :character   Class1:hms        Class :character
##  Median : 83435362   Mode  :character   Class2:difftime   Mode  :character
##  Mean   :102280741                      Mode  :numeric
##  3rd Qu.:150911774
##  Max.   :230611229
##
##     PRECINCT      JURISDICTION_CODE LOCATION_DESC      STATISTICAL_MURDER_FLAG
##  Min.   :  1.00   Min.   :0.000     Length:23585       Mode :logical
##  1st Qu.: 44.00   1st Qu.:0.000     Class :character   FALSE:19085
##  Median : 69.00   Median :0.000     Mode  :character   TRUE :4500
##  Mean   : 66.21   Mean   :0.333
##  3rd Qu.: 81.00   3rd Qu.:0.000
##  Max.   :123.00   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           X_COORD_CD        Y_COORD_CD
##  Length:23585       Length:23585       Min.   : 914928   Min.   :125757
##  Class :character   Class :character   1st Qu.: 999925   1st Qu.:182539
```

```
##  Mode  :character   Mode  :character   Median :1007654   Median :193470
##                                        Mean   :1009379   Mean   :207300
##                                        3rd Qu.:1016782   3rd Qu.:239163
##                                        Max.   :1066815   Max.   :271128
##
##     Latitude       Longitude        Lon_Lat
##  Min.   :40.51   Min.   :-74.25   Length:23585
##  1st Qu.:40.67   1st Qu.:-73.94   Class :character
##  Median :40.70   Median :-73.92   Mode  :character
##  Mean   :40.74   Mean   :-73.91
##  3rd Qu.:40.82   3rd Qu.:-73.88
##  Max.   :40.91   Max.   :-73.70
##
```

Transform Data

- Transform data to see relation between incidents and deaths

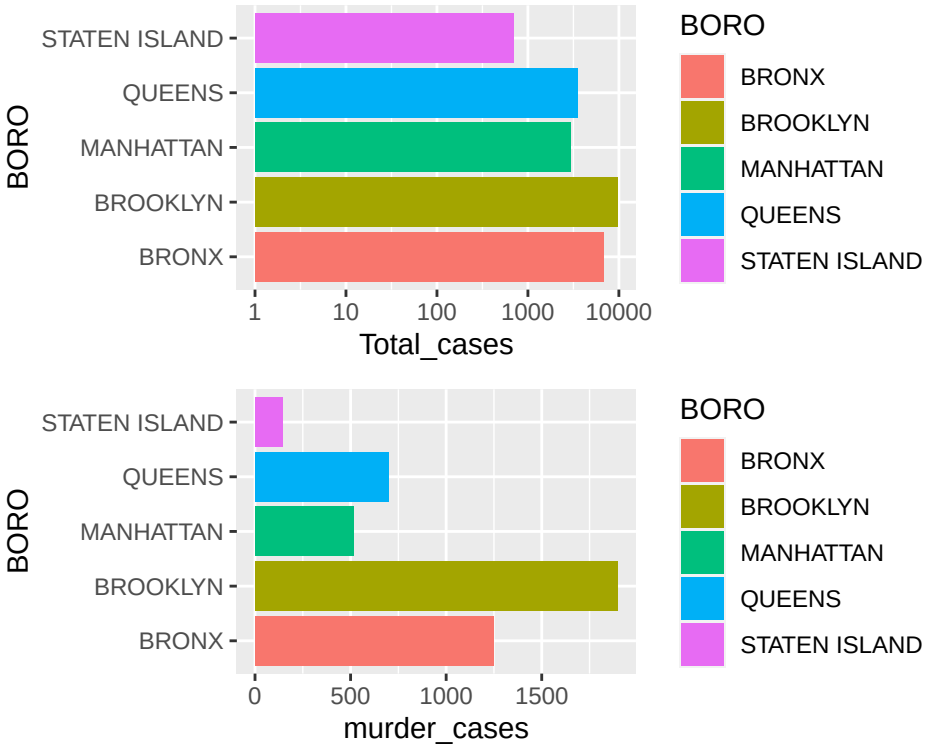
```
##    OCCUR_DATE          OCCUR_TIME       BORO_PRECINCT          BORO
##  Min.   :2006-01-01   Length:23585      Length:23585       Length:23585
##  1st Qu.:2008-12-31   Class1:hms        Class :character   Class :character
##  Median :2012-02-27   Class2:difftime   Mode  :character   Mode  :character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

- Columns we keep in dataset: OCCUR_DATE, OCCUR_TIME, BORO_PRECINCT, BORO, PRECINCT, STATISTICAL_MURDER_FLAG, PERP_AGE_GROUP, PERP_SEX, PERP_RACE, VIC_AGE_SEX, VIC_AGE_GROUP, VIC_SEX, VIC_RACE

Analyzing data

Boros deaths,cases and murder rate:

BORO	cases	murder_cases	murder_rate
BRONX	3084	1247	0.1860916
BROOKLYN	3845	1898	0.1949866
MANHATTAN	1785	515	0.1762491
QUEENS	2075	697	0.1973386
STATEN ISLAND	521	143	0.2054598

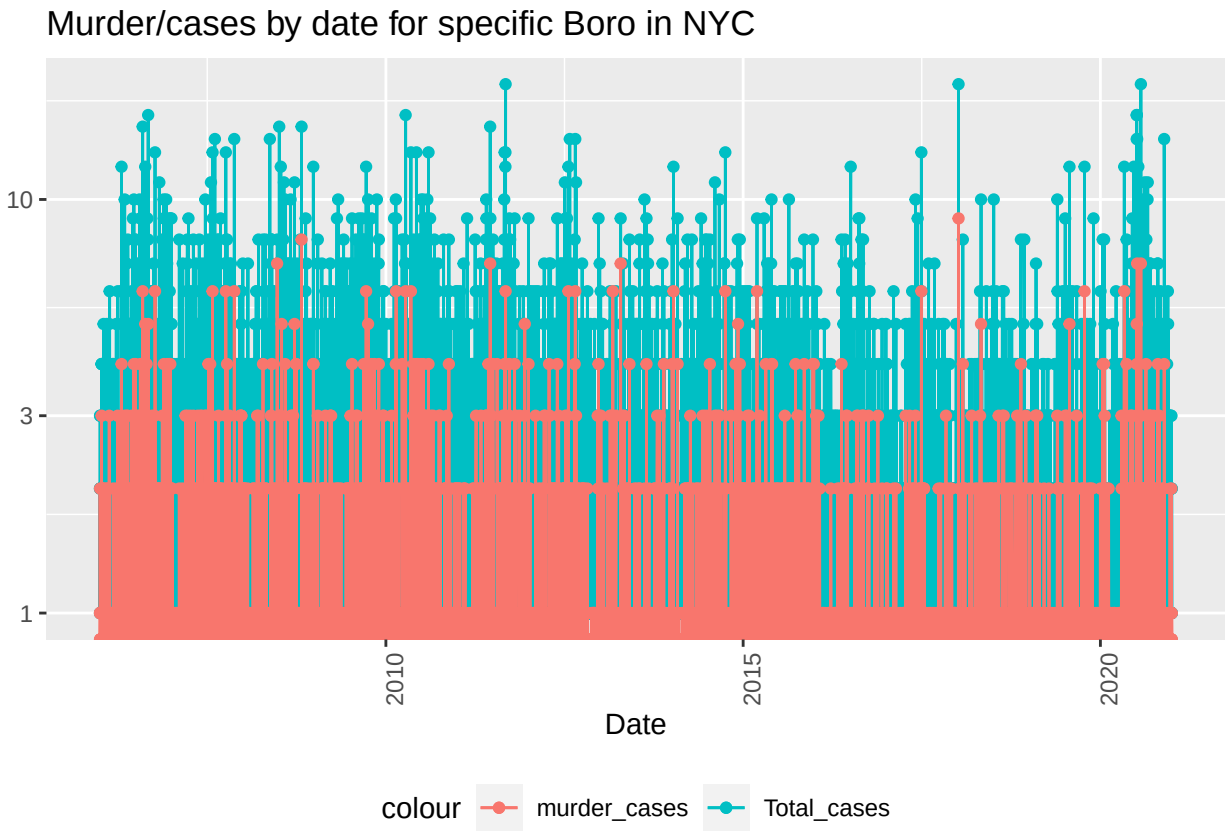


Visualizing and analyzing results

Cases Murders comparison

- Are murder and incidents related?

```
## Warning: Transformation introduced infinite values in continuous y-axis
## Transformation introduced infinite values in continuous y-axis
```

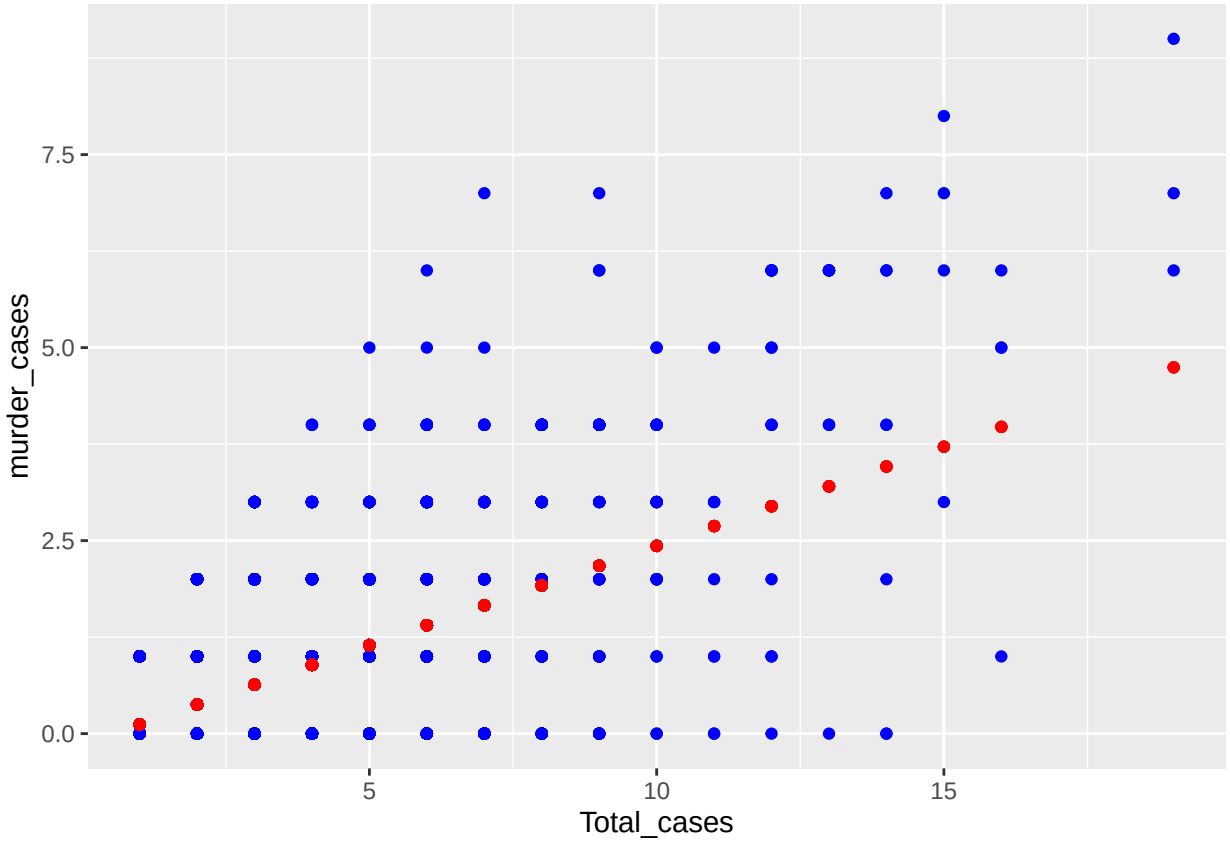


Data Modeling

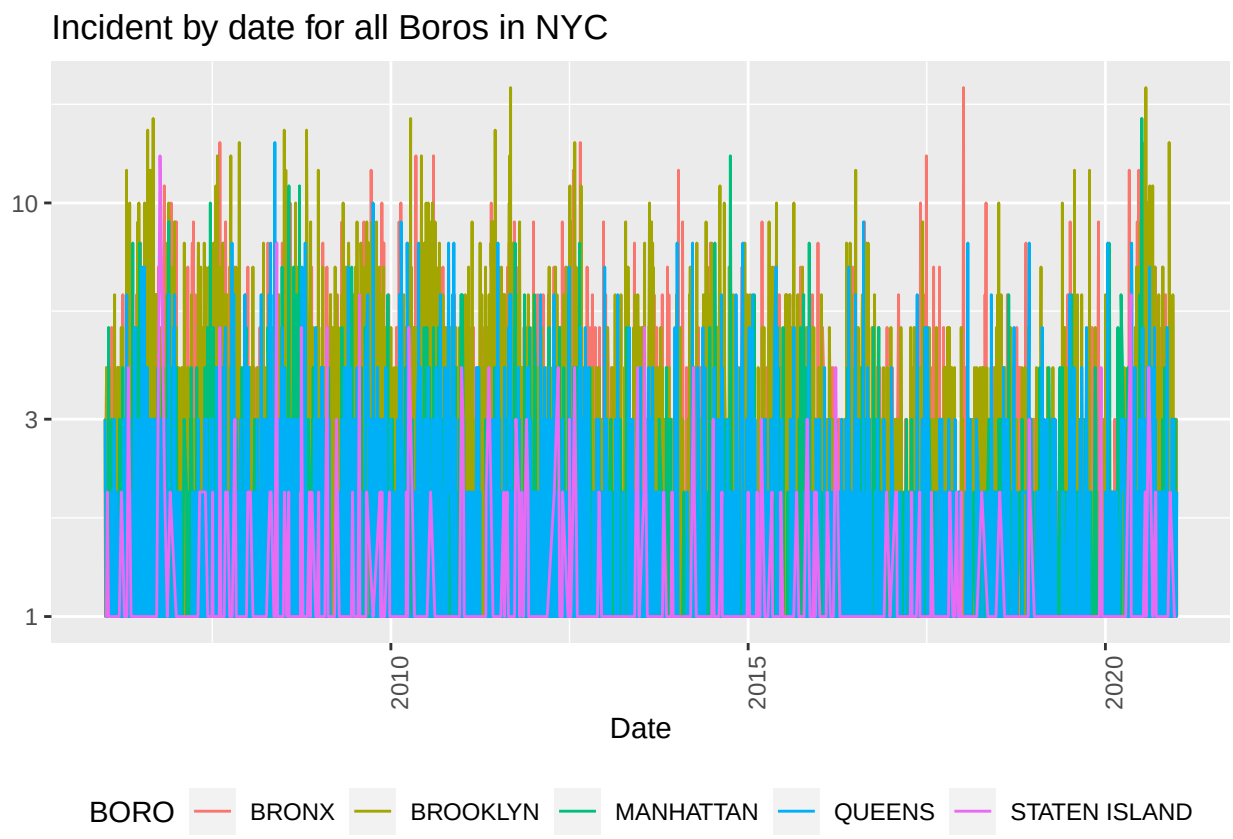
- To see if there is a relation between cases and deaths we'll use a linear model:

```
##
## Call:
## lm(formula = murder_cases ~ Total_cases, data = cases_by_boro)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4594 -0.3760 -0.1190  0.1101  5.3393
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.137959   0.009545  -14.45   <2e-16 ***
## Total_cases  0.256956   0.003535   72.68   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6448 on 11308 degrees of freedom
## Multiple R-squared:  0.3184, Adjusted R-squared:  0.3183
## F-statistic:  5283 on 1 and 11308 DF,  p-value: < 2.2e-16
```

Model plot

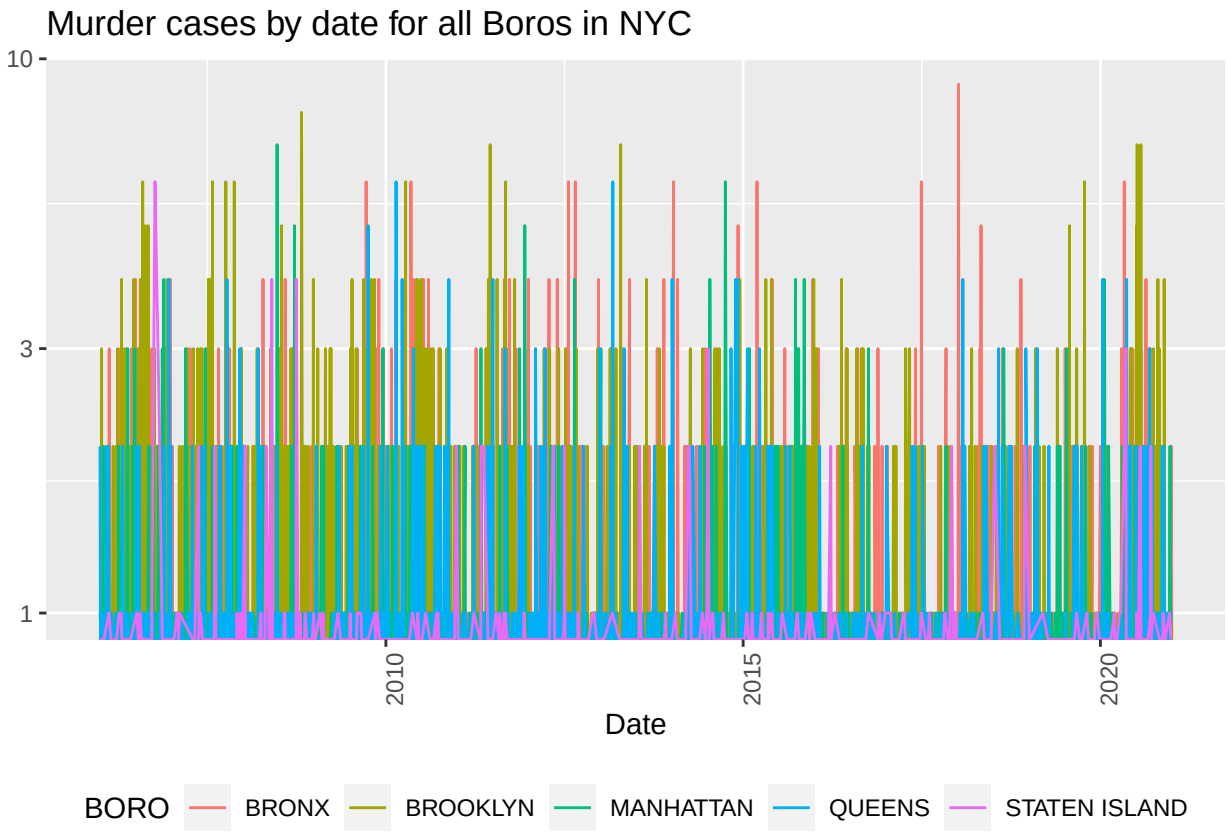


Cases comparison

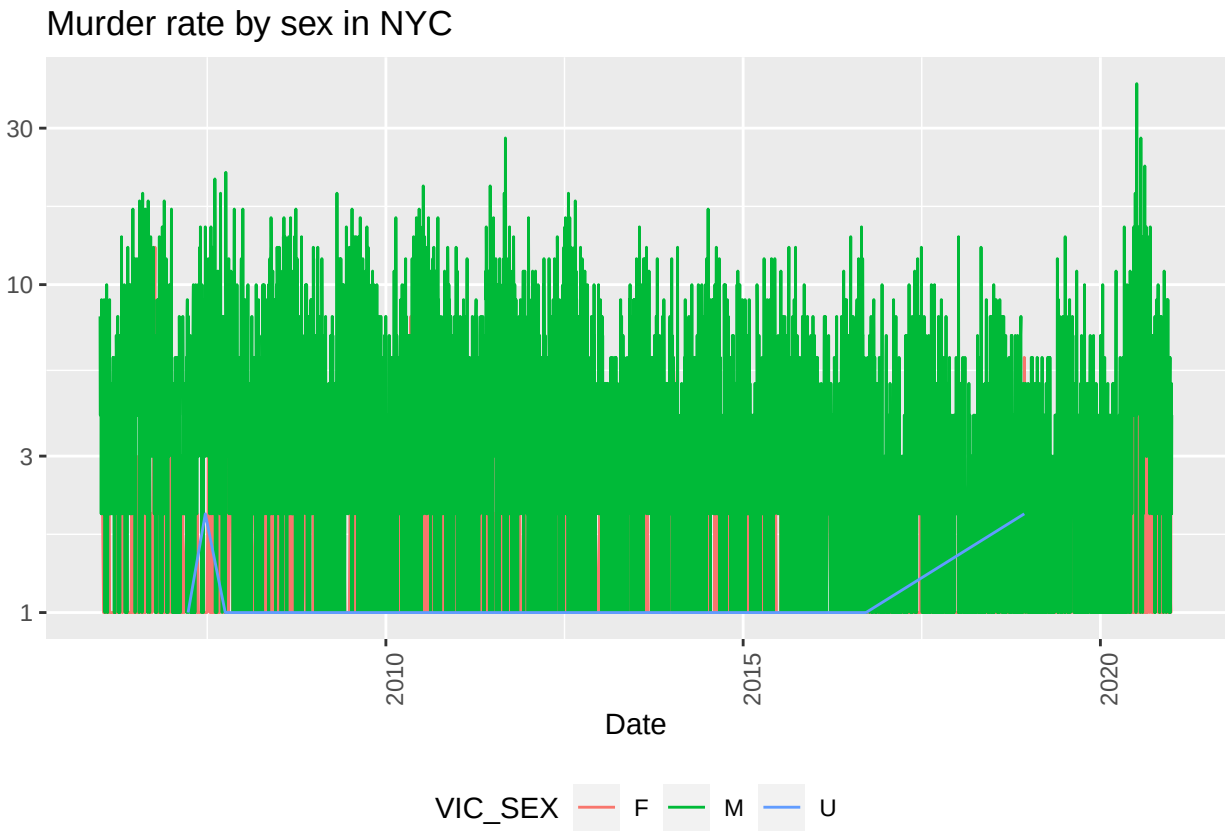


Murders comparison

```
## Warning: Transformation introduced infinite values in continuous y-axis
```


Murder cases by date for all Boros in NYC
10
3
1
2010
2015
2020
Date
BORO
BRONX
BROOKLYN
MANHATTAN
QUEENS
STATEN ISLAND

Murders Rate by victim's sex



Conclusion

Brooklyn is the Neighborhood with more incidents and murders in NYC with a murder rate of 0.195. The boro with highest murder rate is Staten Island with 0.205 although it can be moticated by a lower population. Male victims are the most affected in shootings in NYC by a factor of 10. Finally the incidents number has had an increase in the last year after some declines, probably directly related to covid-19.

As potential bias my inclination against violence and guns probably affects the analysis I've given in this report. I will favor showing bad outputs that shows how bad is gun violence.