

New York Shooting Data by Boro

2023-02-18

IMPORTANT

PLEASE NOTE: The mapview package has an output that does *NOT* support knit to PDF output. Please Knit to html only.

IMPORT DATA

First we are going to go through and import the data. We are going to read in the data from <https://data.cityofnewyork.us> (<https://data.cityofnewyork.us>).

We are also going to use the tidyverse, lubridate, ggplot2, and the mapview library.

We are also going to read in the CSV version of the data.

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

TIDY DATA

The Types of Data we are interested in is the following:

- Where the incident occurred.
- Which NY City borough
- Which Year an incident occurred
- if the victim was murdered.

TRANSFORM and VISUALIZE

Lets go ahead and summarize the victims and murders.

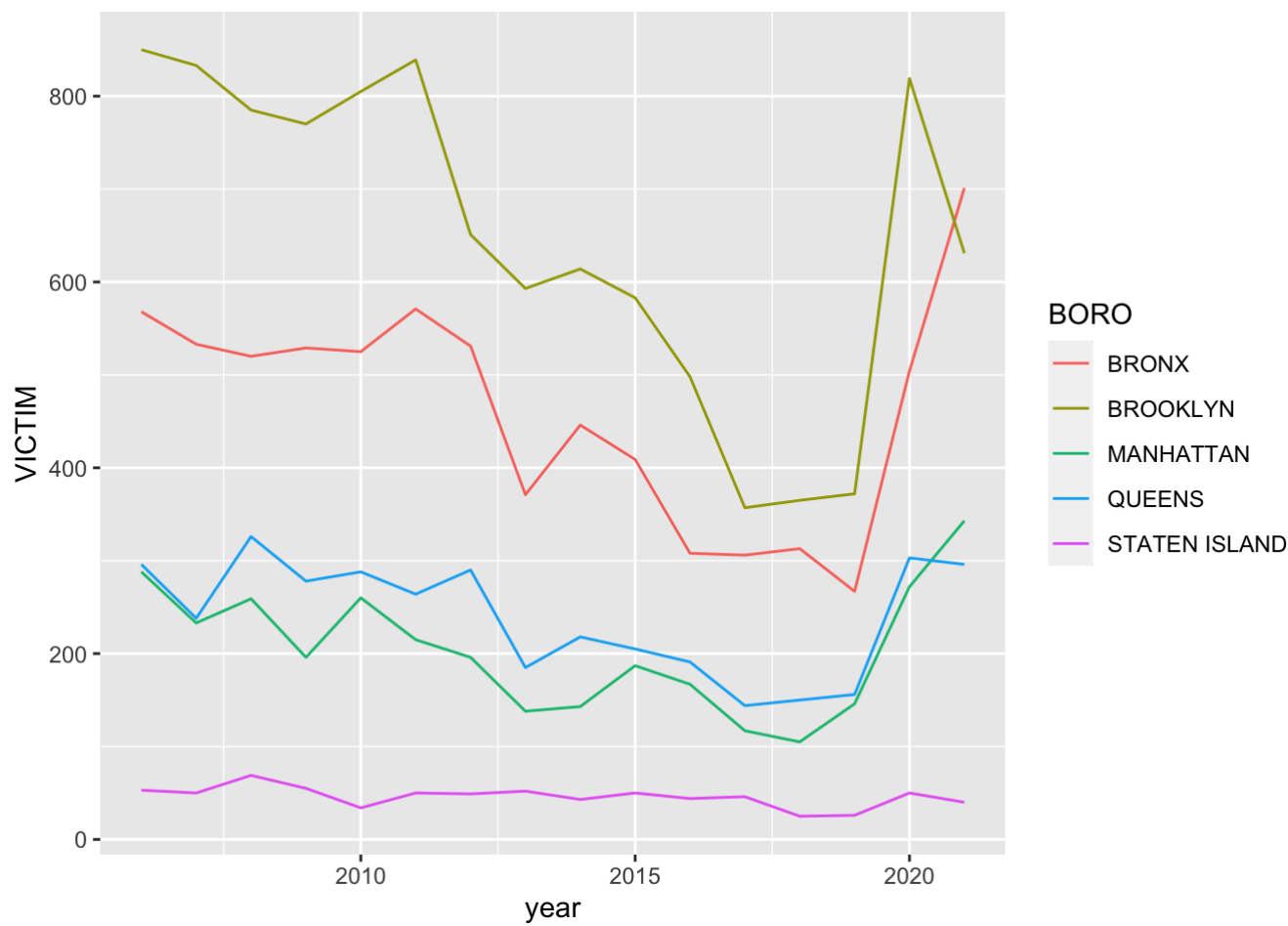
```
# NYBorobyYear <- NYBoro %>% group_by(BORO, year) %>% summarise(VICTIM = sum(VICTIM), murder = sum(murder)) %>% ungroup()
```

```
# Percent murders
```

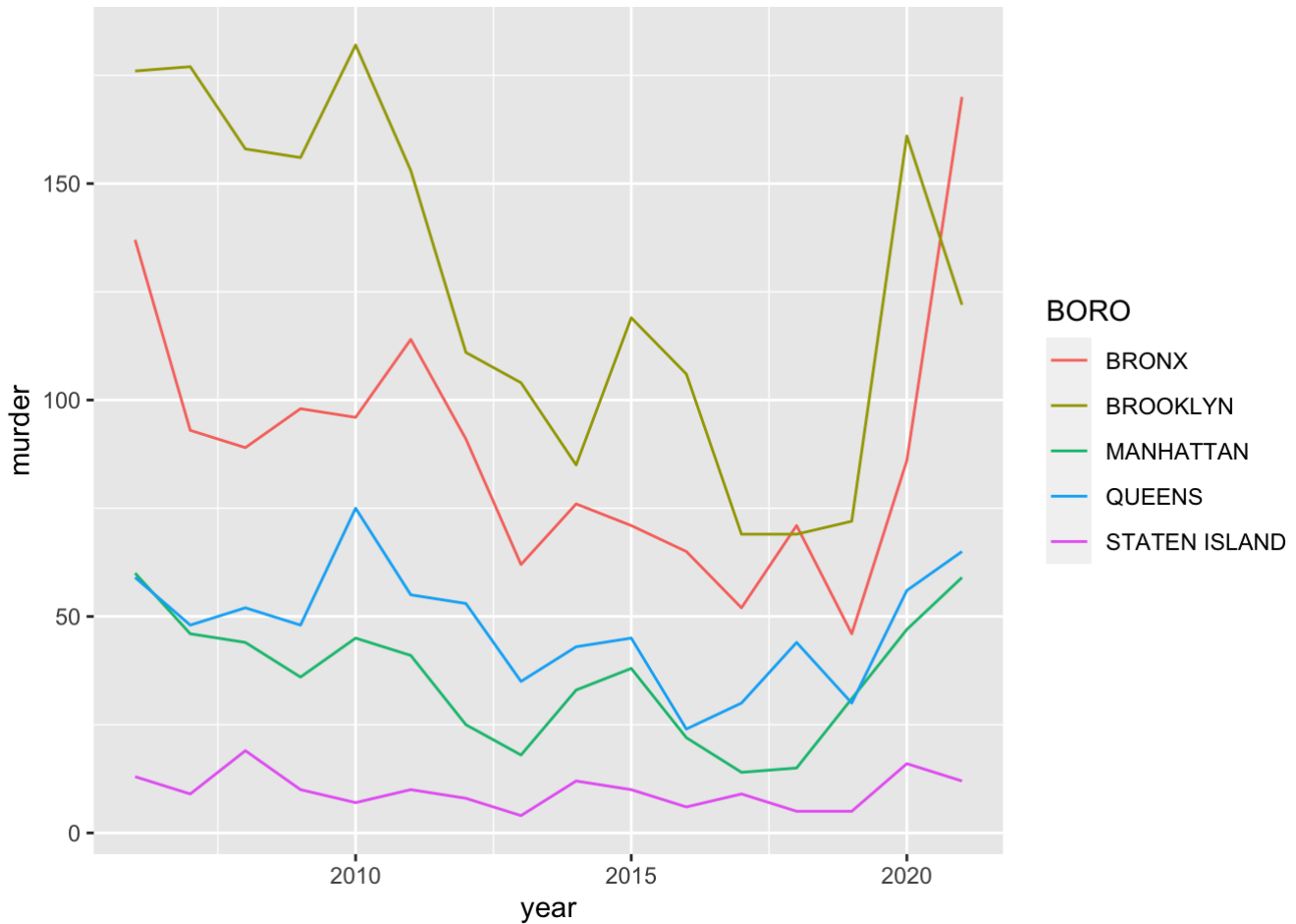
```
# NYBorobyYear <- NYBorobyYear %>% mutate(perMurder = (murder/VICTIM)*100)
```

Lets graph the number of Victims per year in each

BORO



Lets graph the number of Murders per year



ANALYSIS

Here we can see that the numbers of Victims and the number of murders. The headlines recently had stated that Gun incidents are up recently.

This we can clearly see in the above graphs.

VICTIMS vs. MURDERS

So I was curious if the number of murders vs. the number of victims had Changes over time.



ANALYSIS

I really thought that we would see more people surviving gun violence. But that was not the case. It's around 20%.

We will also model this later as well.

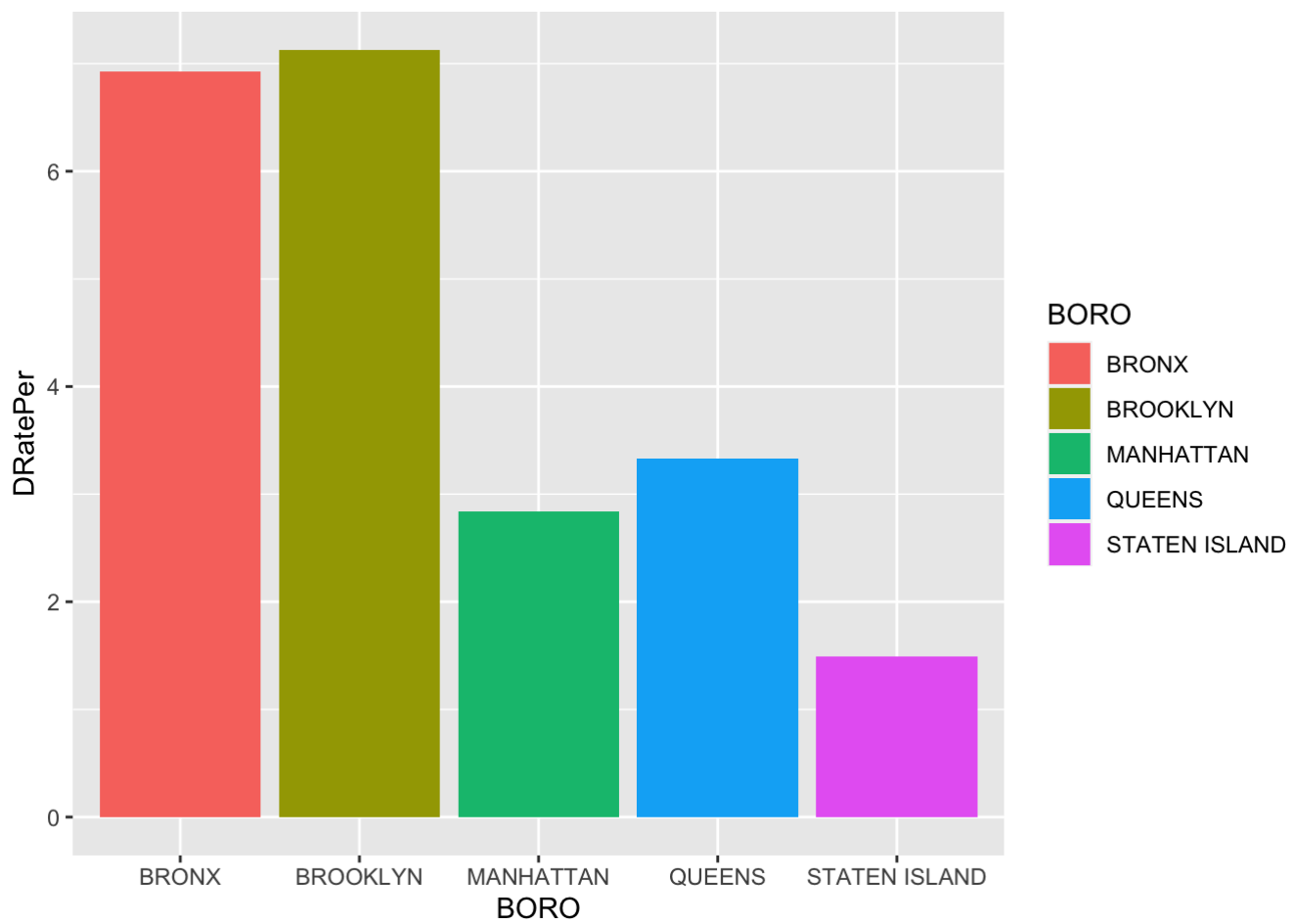
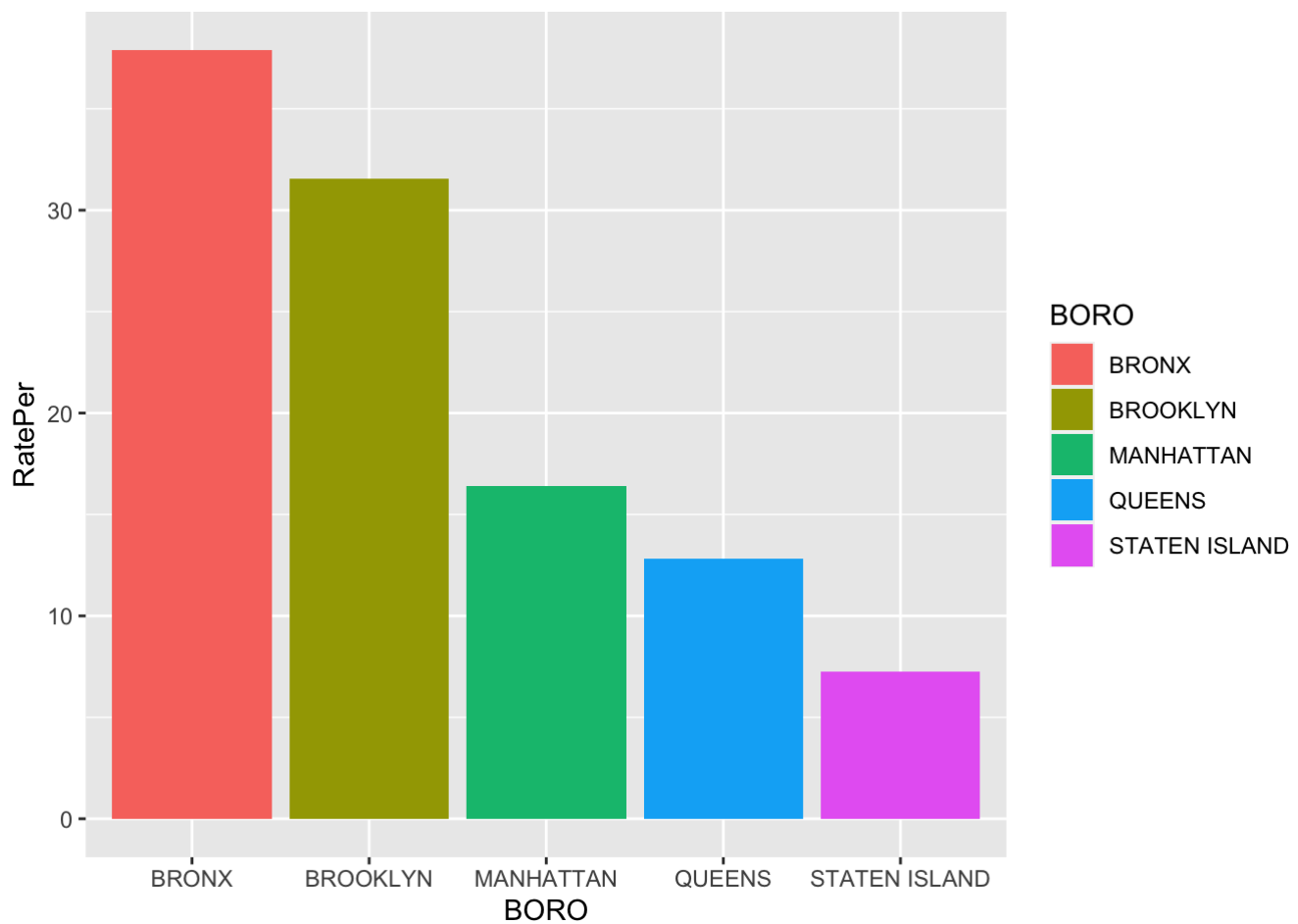
Note that Staten Island is a little bit of an outlier. It's because of so few murders occurring in the Borough.

MORE VISUALS AND MURDERS

I wanted to see one borough had higher rate of incidents vs. the population in that borough. In this case I grabbed some population data for each borough.

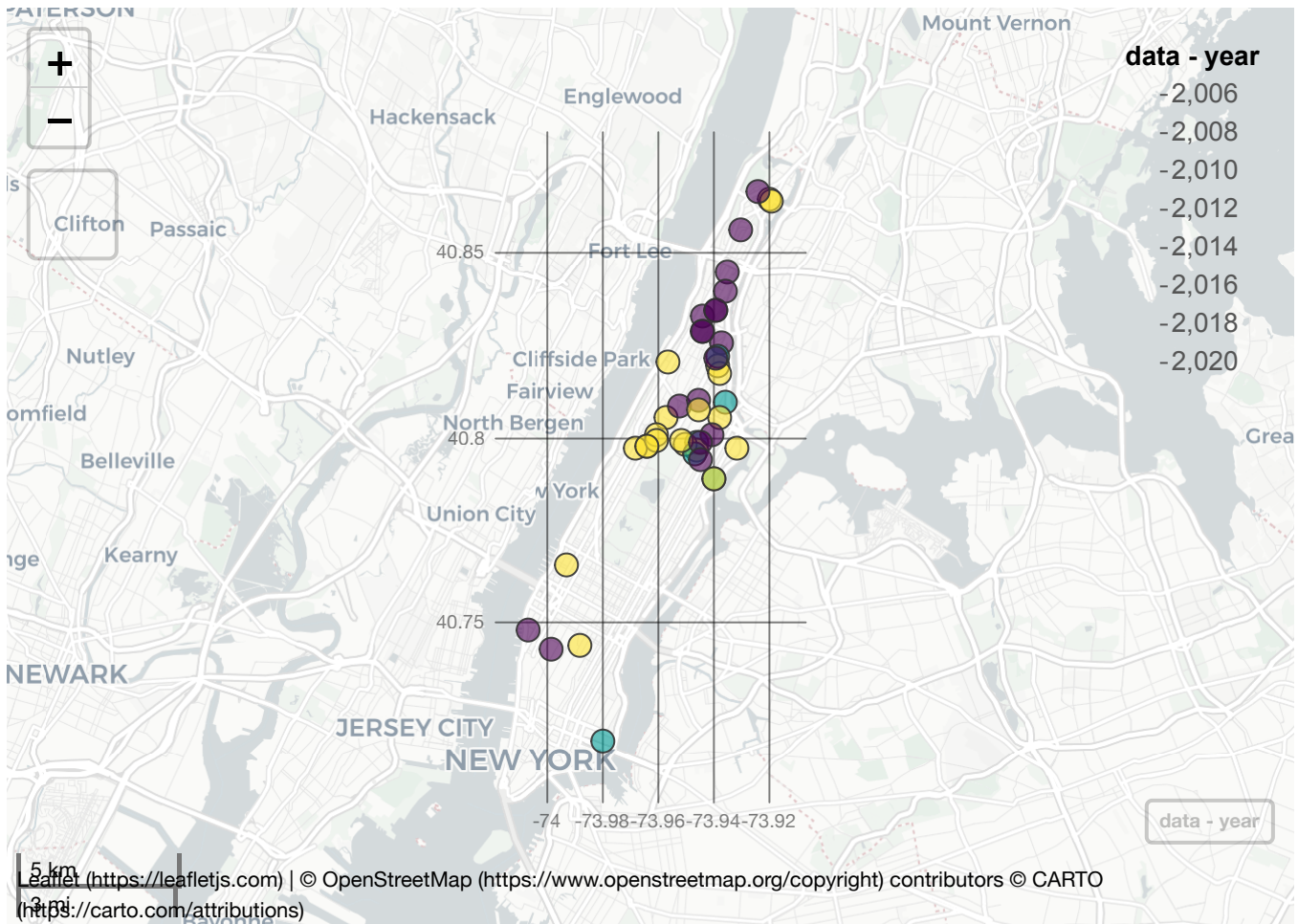
In this case I went through the whole process of Tidy, transforming, Visualize, and Model the data.

I took the population data and joined it with the incident data. Created a few new columns for Victims per 100000 population, and murders per 100000 population.



Lets map out the location of incidents on a Map. Just to see if gun crime is localized to certain areas.

```
## Rows: 25596 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.
```



BIAS

- Personal Bias
 - I am familiar with Manhattan.
 - Confirmation bias
- Other Bias
 - Not sure how the data is collected.
 - Limited data availability. I would have loved to start from the 1960's.