NYPD Shooting Incidents

By

Agenda

- Goals
- Data Examined.
- Methodology.
- Data Cleanup Steps
- Question 1 results.
- Question 2 results.

What questions are we trying to solve?

- Do the number of gun incidents verses the number of deaths changed over time.
- Which NY Borough had the least number of gun incidents vs. the population of that Borough.

Data Examined

Examined the following data sets:

- City of New York shooting incidents:

  https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD

- City of New York population Data:

  https://data.cityofnewyork.us/api/views/xywu-7bv9/rows.csv?accessType=DOWNLOAD

Cleanup Steps (Gun incident data)

- Remove some unnecessary fields
  - NYCity_data <- NYCity_data %>% select(-c(INCIDENT_KEY, LOCATION_DESC, JURISDICTION_CODE, Lon_Lat,X_COORD_CD, Y_COORD_CD,OCCUR_TIME))
  - NYBoro <- NYCity_data %>% select(OCCUR_DATE,BORO,STATISTICAL_MURDER_FLAG)

- Rename some fields and add a field.

NYBoro <- NYBoro %>% rename(date = OCCUR_DATE) %>% mutate(date = mdy(date)) %>% rename(murder = STATISTICAL_MURDER_FLAG) %>% mutate(VICTIM = 1)

- Fix the date field into months and year.
  - NYBoro <- NYBoro %>% mutate(year=year(date), month=month(date))

-

More Cleanup Steps

- 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
```

Cleanup Population Data

- Select the population data.

  NYPopData <-NyPop_data %>% select(Borough,`2010`,`2020`)

- Pivot the data

  NYPopData <- NYPopData %>% pivot_longer(cols = -c(`Borough`), names_to = "year", values_to = "population"

- Convert Year to Double.

  NYPopData$year <- as.double(NYPopData$year)

- Delete first few rows and rename

  NYPopData <- NYPopData[-1,]
  NYPopData <- NYPopData[-1,]
  NYPopData <- NYPopData %>% rename("BORO"="Borough")
  NYPopData <- NYPopData %>% mutate(BORO = toupper(BORO))

- Filter By year
  - NYBoroPopulation <- NYBorobyYear %>% filter(year==2010 | year==2020)

Cleanup Population Data

- Join the data...

```
NYBoroPopulation <- NYBoroPopulation %>%
left_join(NYPopData, by = c("year","BORO"))
```

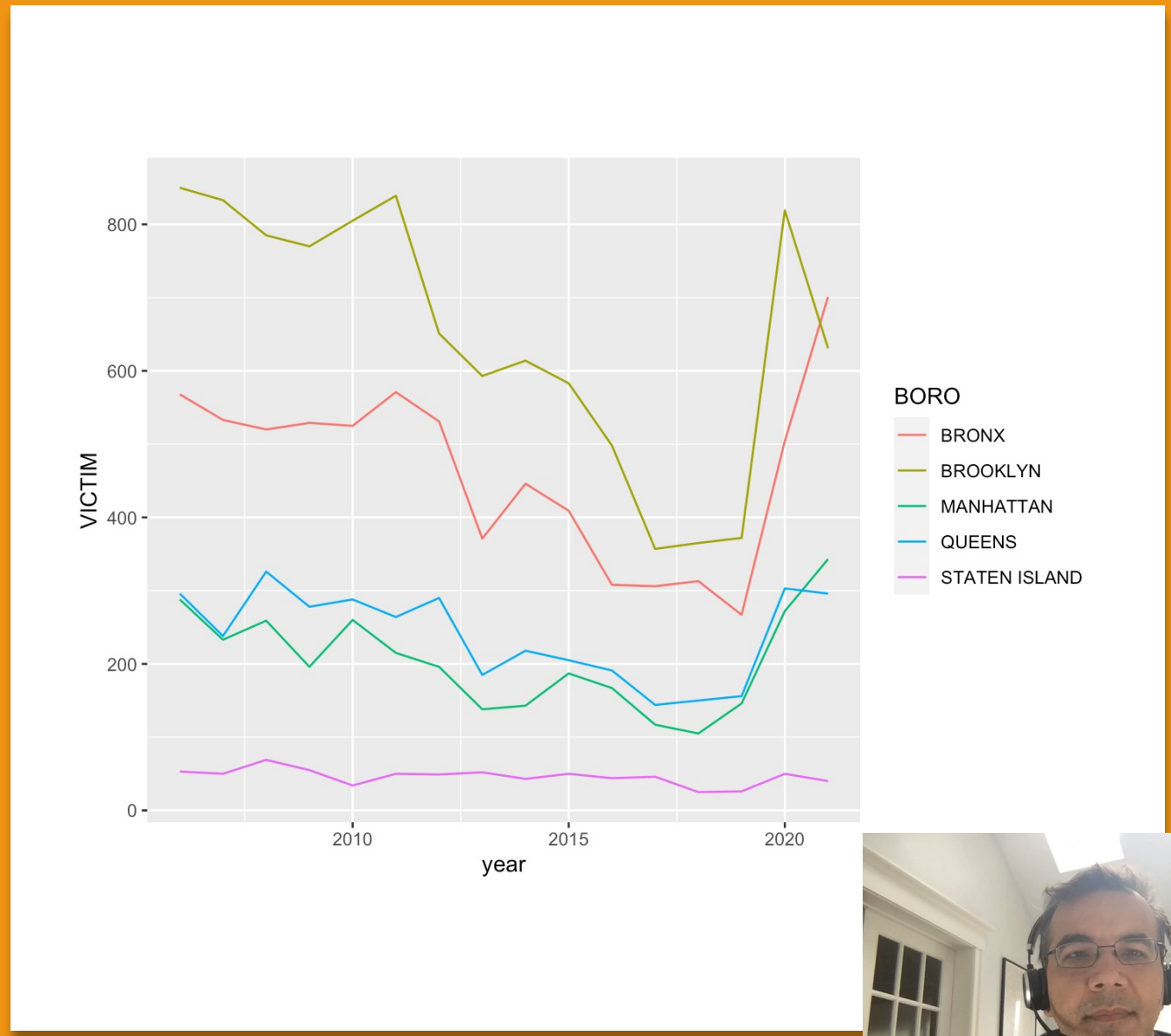
- Now let's calculate the rate per 100000

```
NYBoroPopulation <- NYBoroPopulation %>% mutate(RatePer =
(VICTIM/population) * 100000)
```

```
NYBoroPopulation <- NYBoroPopulation %>% mutate(DRatePer
= (murder/population) * 100000)
```

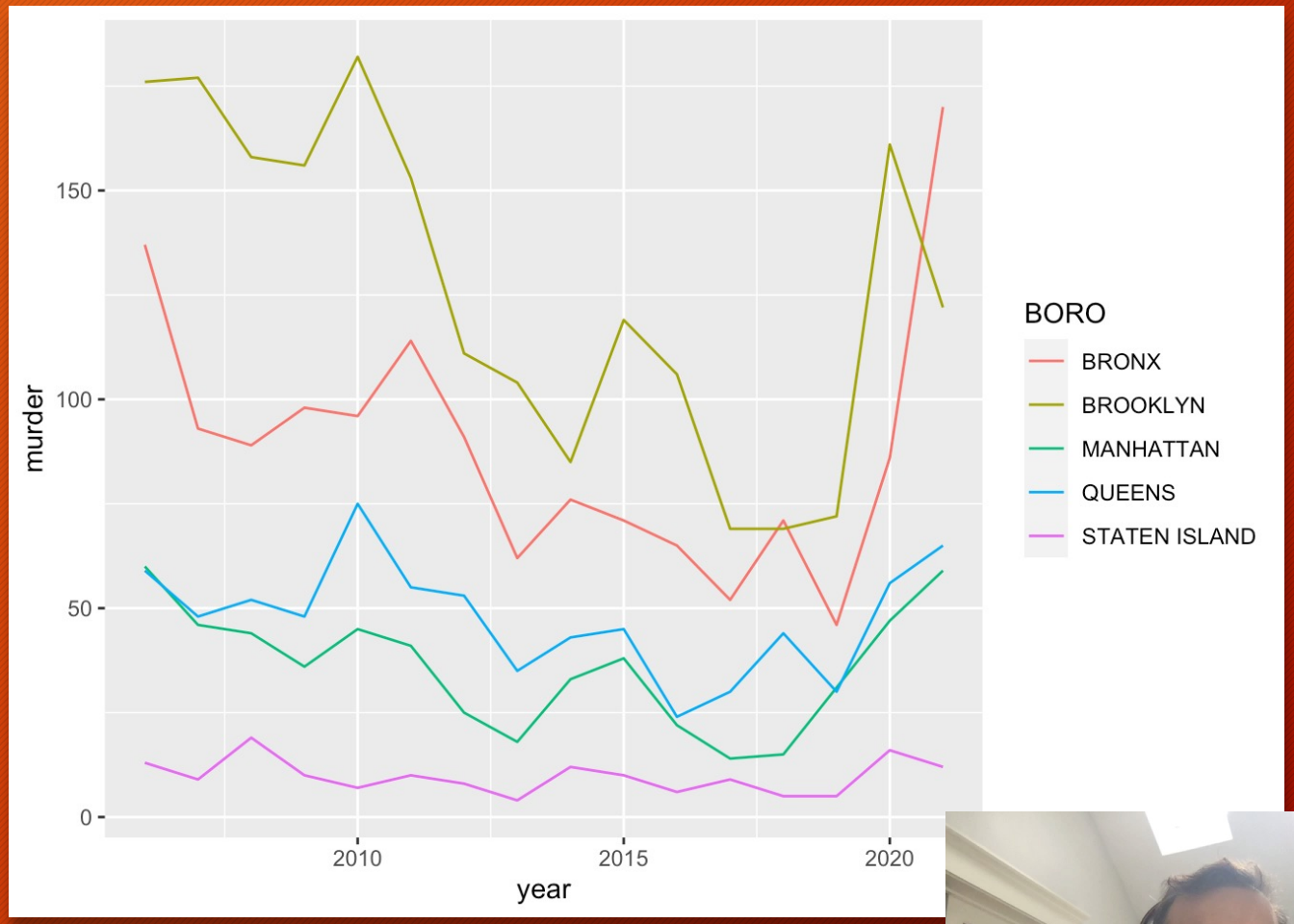
Results for Question 1.

- This shows the number of Victims per new York City Borough
- Its worth pointing out that gun crime seems to be up the last few years.



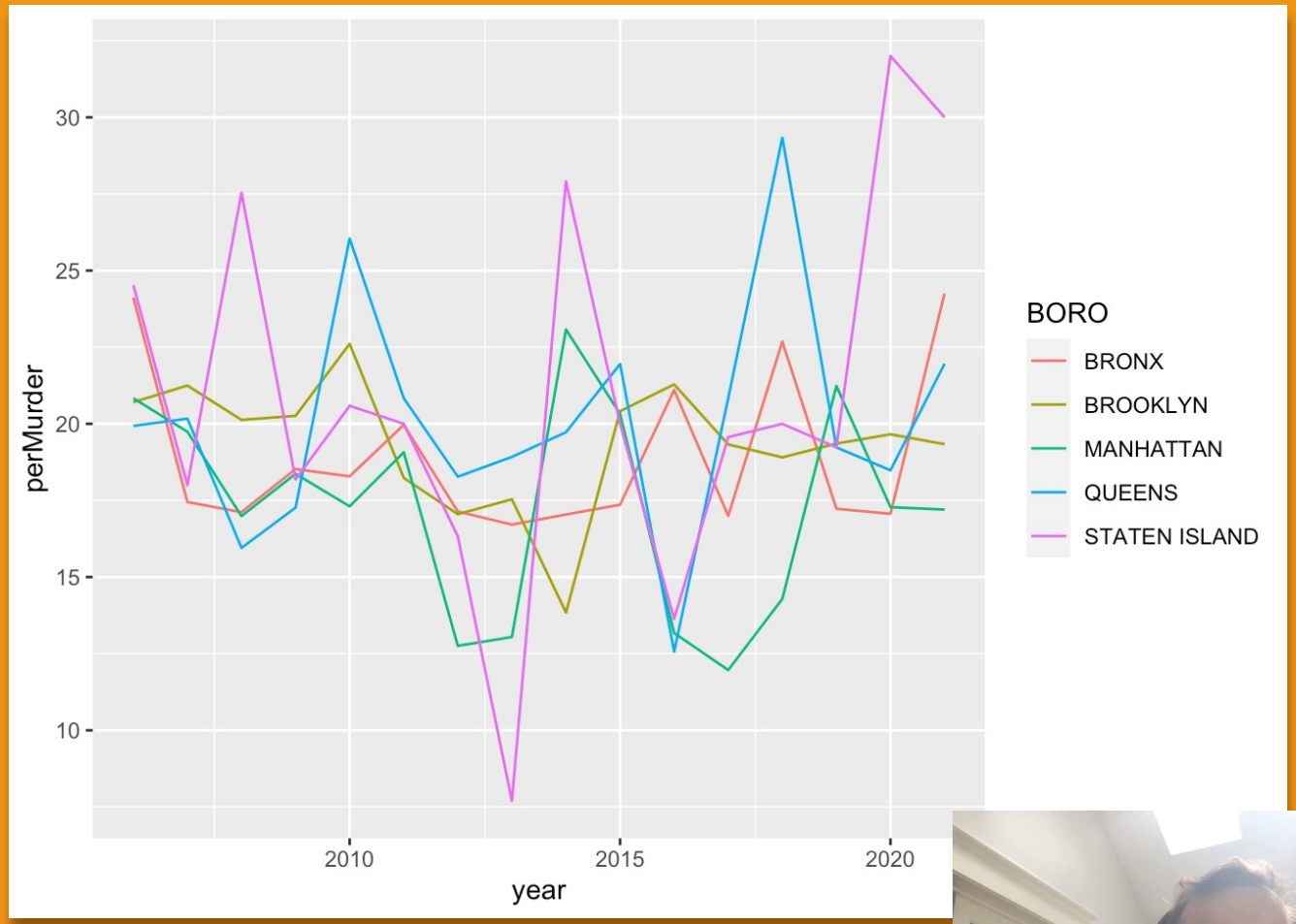
Results for Question 1.

- This graph shows the number of Murder victims per Borough.



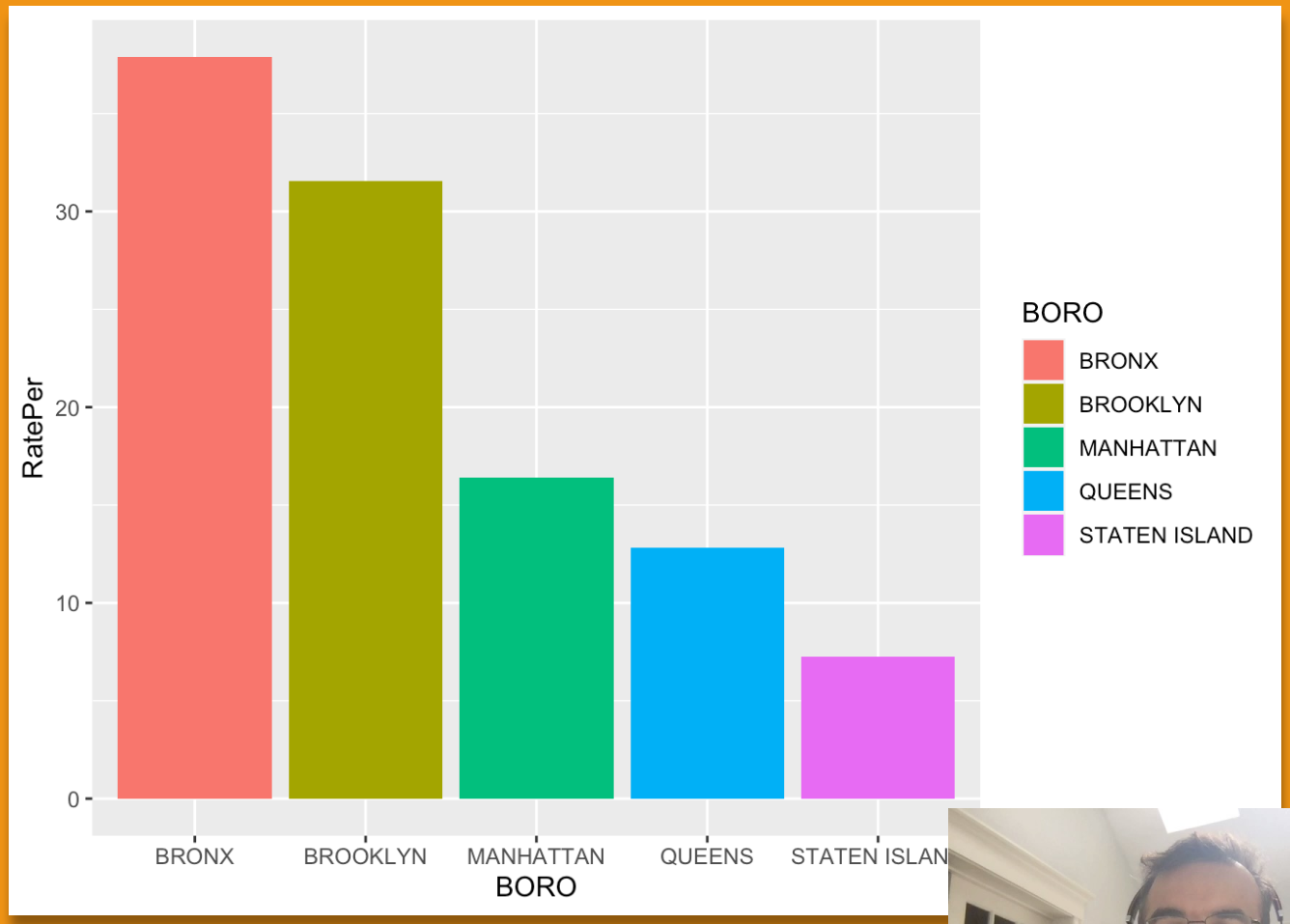
Results from Question 1

- This shows the percent of people killed each year. (deaths divided by the number of Victims.
- As we can the data does Vary year to year… but overall the tread is no change.



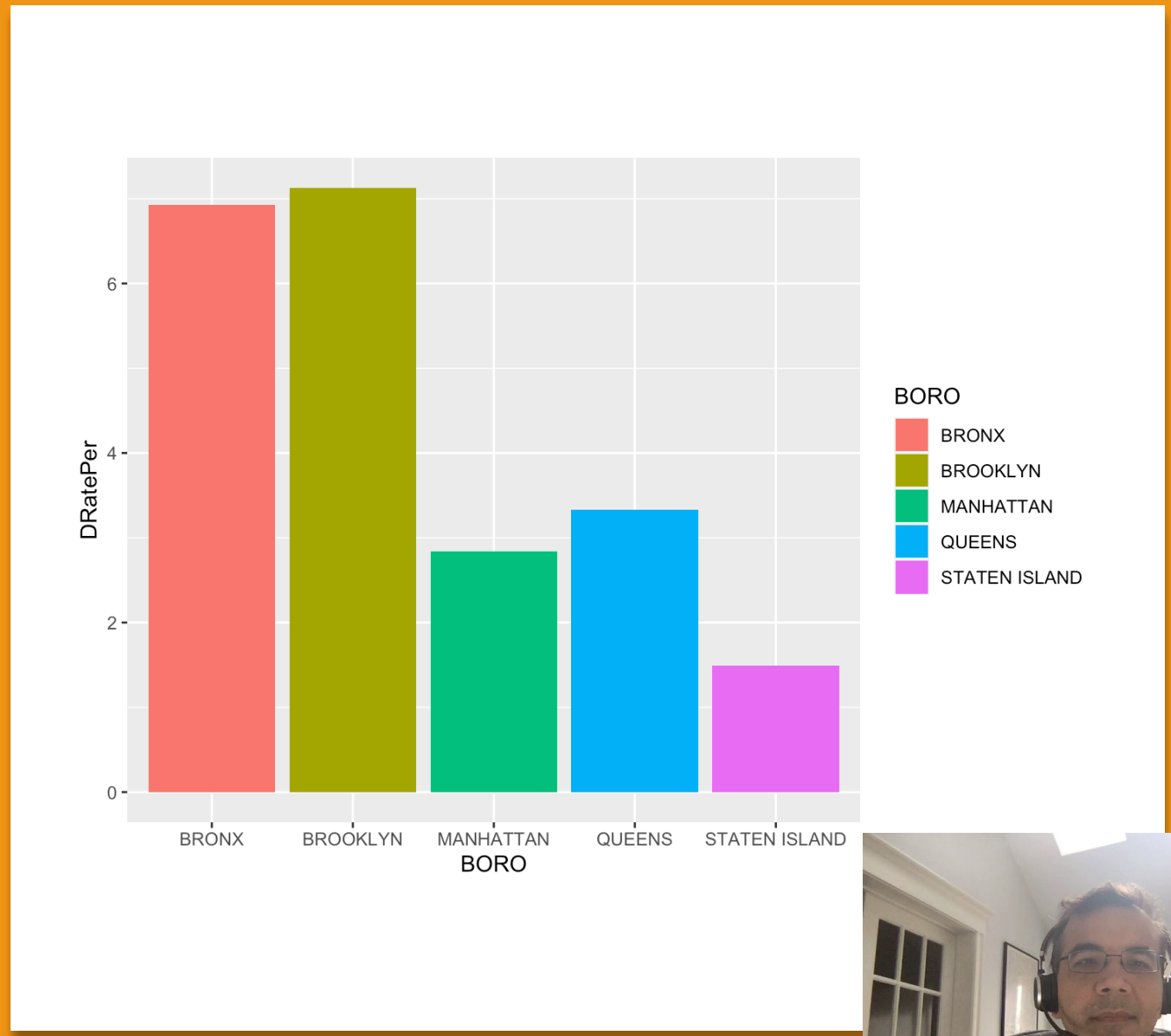
Results for Question 2.

- This graph shows the number of incidents/population * 100000.
- This number is the rate per 100000



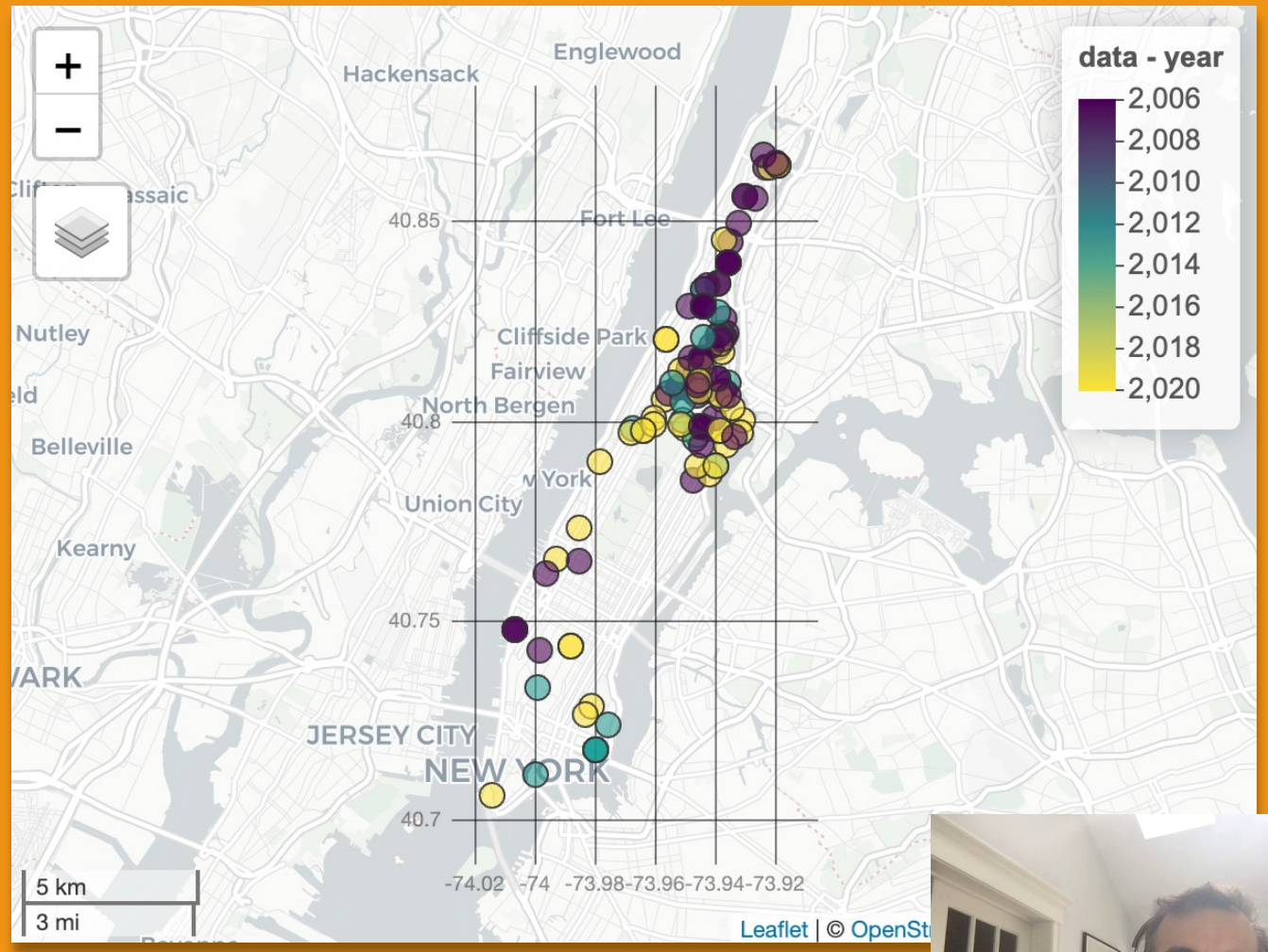
Results for Question 2.

- This shows the same as previous graph but with **DEATHS**.
- Number of Deaths/population * 100000



Something a little extra (Manhattan crime map)

- This is using the mapview package.
- I was just curios to see if crime was moving around in a Borough
- As you can see Manhattan's crime problem is mostly on the North side of town
- It seems to be around "Harlem" and Washington Heights neighborhood.
-



BIAS

- It could be how the data is collected.
- My personal biases in ignoring the outliers for question 1 could be my personal bias.