NYPD Data Visualization

Lin Shao Chieh

2022-08-01

Subtopic #1 Borough Comparison

1. Which borough has the most shooting incidents?
2. Is the fatality rate (victim died due to the shooting incident) higher in certain boroughs?
3. Does Manhattan, the richest borough in New York, has less shooting incidents?

Conclusions

1. Brooklyn has the most shooting incidents.
2. Bronx has the highest incidents/population rate.
3. Although Manhattans' GDP is much higher than the other boroughs, its incidents/population rate is not significantly lower. (Actually, its indidents/population rate is higher than Queens and Staten Island.) However, victims in Manhattan had a better chance to survive the shooting incidents. In other words, Manhattan has the lowest fatality rate(incidents/murdered) among the 5 boroughs.

Research Methods

1. Grouping data by boroughs.
2. Generating variables: incidents per thousands, murdered victims per thousands, fatality rate.
3. Importing external data: GDP in different boroughs.
4. Evaluating GDP's affect on the generated variables.

Bias Sources

1. Although higher GDP per capita often translates to lower crime rate, GDP doesn't always reflect a boro's budget on policing or medical services. These two sectors might affect the number of shooting incidents and the shooting fatality rate.
2. GDP per capita might be misleading because it doesn't take wealth gap into consideration. A wide wealth gap may lead to more crime in the poorer neighborhoods.

Import Datasets

```
# main dataset
url <- 'https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD'
df <- read_csv(url)
```

```
## 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.
```

```
# boro stats from wikipedia (https://en.wikipedia.org/wiki/Boroughs_of_New_York_City)
BORO <- c('BRONX', 'BROOKLYN', 'MANHATTAN', 'QUEENS', 'STATEN ISLAND')
boro_pop <- c(1472654, 2736074, 1694263, 2405464, 495747)
boro_gdp <- c(36.938, 86.151, 610.386, 83.328, 14.270)
boro_stats <- data.frame(BORO, boro_pop, boro_gdp)
```

Data Preprocessing

```
boro <- df %>% count(BORO)

temp <- df %>%
filter(STATISTICAL_MURDER_FLAG==TRUE) %>%
group_by(BORO, STATISTICAL_MURDER_FLAG) %>%
count()

temp <- temp %>%
  subset(select = - STATISTICAL_MURDER_FLAG) %>%
  rename(murdered = n)

boro <- boro %>%
  left_join(temp) %>%
  left_join(boro_stats) %>%
  mutate(cases_per_thou = n/boro_pop*1000,
         murd_per_thou = murdered/boro_pop*1000,
         gdp_per_capita = boro_gdp/boro_pop*(10^8),
         fatality_rate = murdered/n)
```

```
## Joining, by = "BORO"
## Joining, by = "BORO"
```

```
## create new columns

boro
```

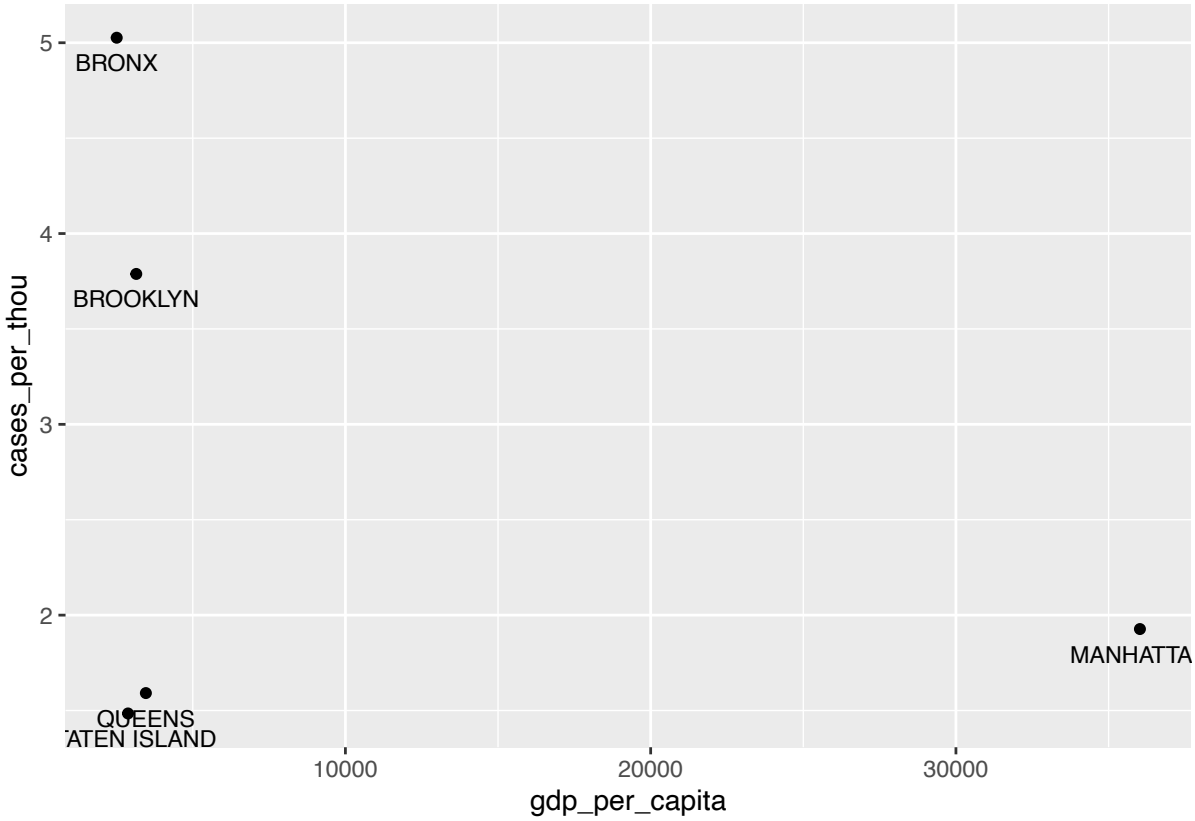
```
## # A tibble: 5 x 9
##   BORO              n murdered boro_pop boro_gdp cases~1 murd_~2 gdp_p~3 fatal~4
##   <chr>         <int>    <int>    <dbl>    <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 BRONX          7402     1417  1472654     36.9    5.03   0.962   2508.   0.191
## 2 BROOKLYN      10365     2020  2736074     86.2    3.79   0.738   3149.   0.195
## 3 MANHATTAN      3265      574  1694263    610.     1.93   0.339  36027.   0.176
```

```
## 4 QUEENS         3828      762  2405464     83.3    1.59   0.317   3464.   0.199
## 5 STATEN ISLAND   736      155   495747     14.3    1.48   0.313   2878.   0.211
## # ... with abbreviated variable names 1: cases_per_thou, 2: murd_per_thou,
## #   3: gdp_per_capita, 4: fatality_rate
```

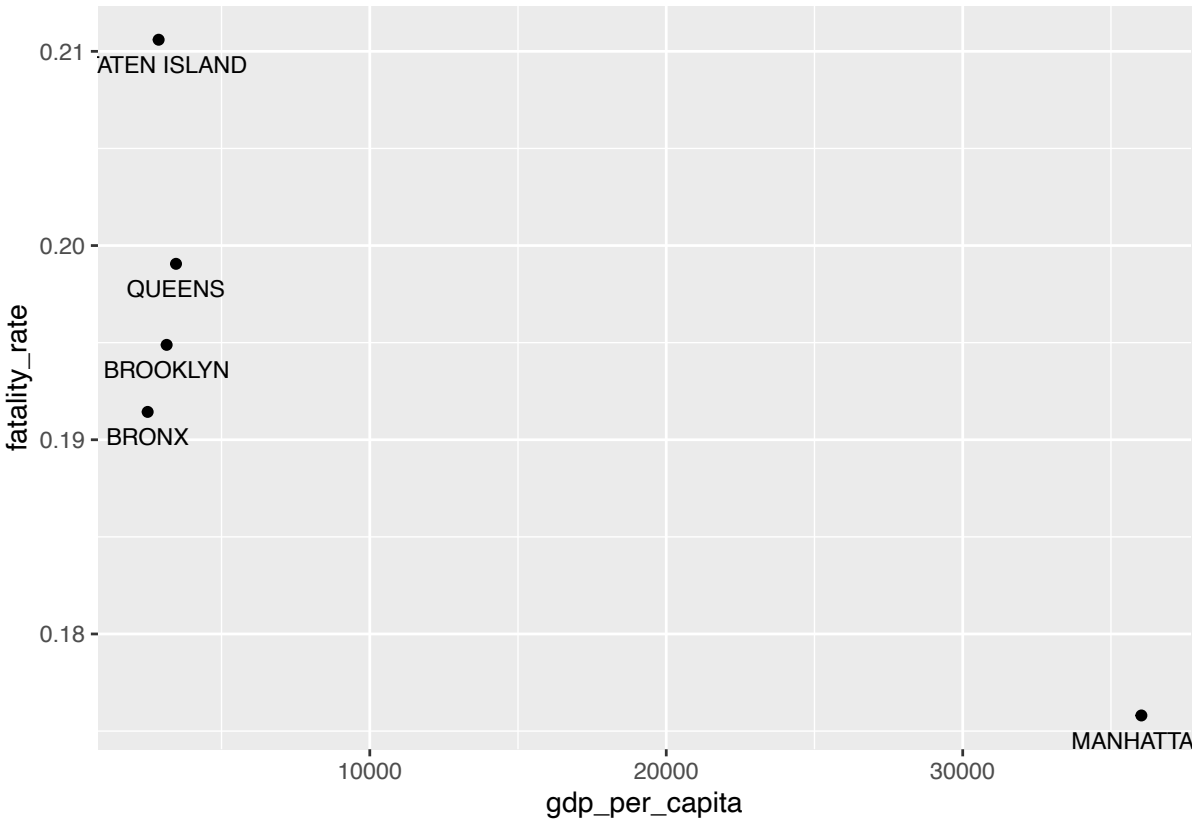
Data Visualization

```
ls_boro <- c('BRONX', 'BROOKLYN', 'MANHATTAN', 'QUEENS', 'STATEN ISLAND')

boro %>%
  ggplot(aes(y=cases_per_thou, x=gdp_per_capita, label=ls_boro)) +
  geom_point() +
  geom_text(vjust=2, size=3)
```




```
boro %>%
  ggplot(aes(x=gdp_per_capita, y=fatality_rate, label=ls_boro)) +
  geom_point() +
  geom_text(vjust=2, size=3)
```



Subtopic #2 Precinct Comparison

1. Are there similar number of shooting incidents in different precincts? Or are there 'hot spots' where most shooting incidents occurred?
2. Is the number of shooting incidents in different precincts a good predictor to gun deaths? In other words, is the gun fatality rate the same in different precincts? Or more shooting incidents lead to higher fatality rate?

Conclusions

1. The distribution of number of shooting incidents is not even. Some precincts has much more shooting incidents than the others.
2. Although the uneven distribution of shooting incidents in different precincts, the number of shooting incidents proved to be a good predictor to number of deaths due to shootings.

Research Methods

1. Grouping data by precincts
2. Plotting histogram to evaluate the distribution of shooting incidents in different precincts.
3. Generating a linear model for deaths to shooting incidents.

Bias Sources

1. Comparing to borughs, precincts are much smaller. As the policing and medical resources are very well shared among different precincts, we cannot see difference in fatality rate in different precincts.

Data Preprocessing

```
pre_n <- df %>%
group_by(PRECINCT) %>%
count() %>%
rename(n_shootings = n)

pre_deaths <- df %>%
filter(STATISTICAL_MURDER_FLAG==TRUE) %>%
group_by(PRECINCT) %>%
count() %>%
rename(deaths = n)

pre_n <- pre_n %>%
left_join(pre_deaths)
```

```
## Joining, by = "PRECINCT"
```

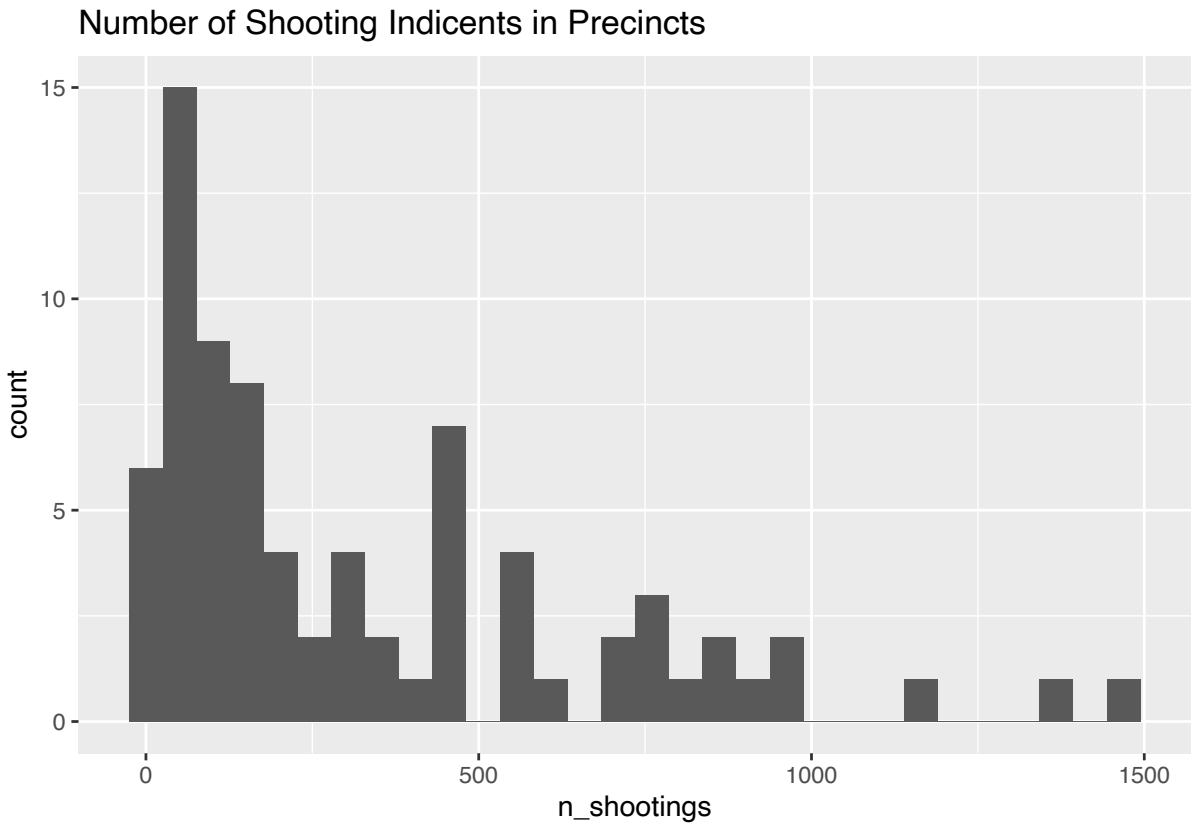
```
## fill na with 0
pre_n[is.na(pre_n)]  <- 0
```

Linear Model

```
mod <- lm(deaths ~ n_shootings, data=pre_n)
pre_n['pred'] <- predict(mod)
```

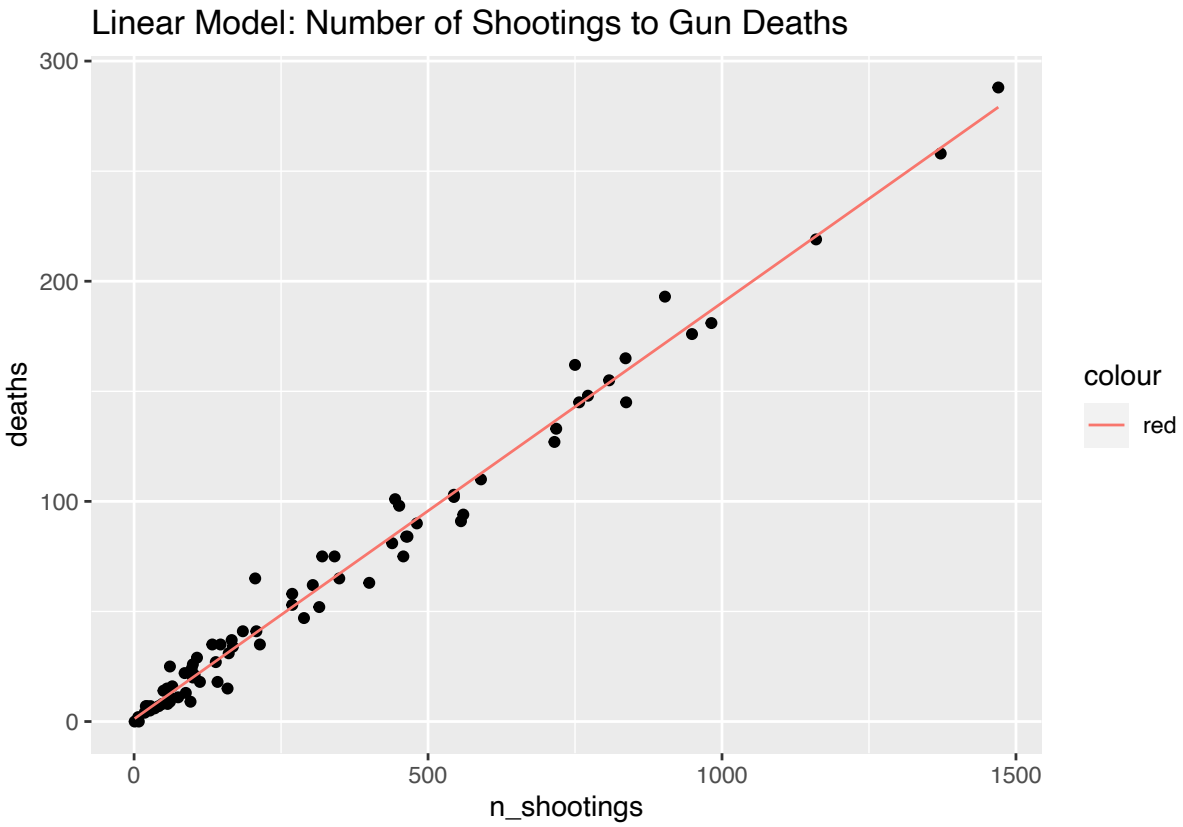
Visualization

```
pre_n %>%
ggplot(aes(n_shootings)) +
geom_histogram(bins=30) +
ggtitle('Number of Shooting Indicents in Precincts')
```




```
##

pre_n %>%
ggplot() +
geom_point(aes(x=n_shootings, y=deaths)) +
geom_line(aes(x=n_shootings, y=pred, color='red')) +
ggtitle('Linear Model: Number of Shootings to Gun Deaths')
```


Linear Model: Number of Shootings to Gun Deaths
300
200
100
0
deaths
0
500
1000
1500
n_shootings
colour
red