NYPD Shooting Incident Data Report

7/7/2021

Contents

Abstract

This report studies and analyzes the shooting incidents reported by the New York Police department. The question of interest is to investigate the geospatial correlation of gun violence.

Data and Sources

- Data for the NYPD shooting data was retrieved from https://catalog.data.gov/dataset with the following link https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD
- Shapefile for NYPD shooting data was retried from < link > with the following link https://data.cityofnewyork.us/Public-Safety/Police-Precincts/78dh-3ptz

Below is a the NYPD Precinct Map colored with Shooting incidents.
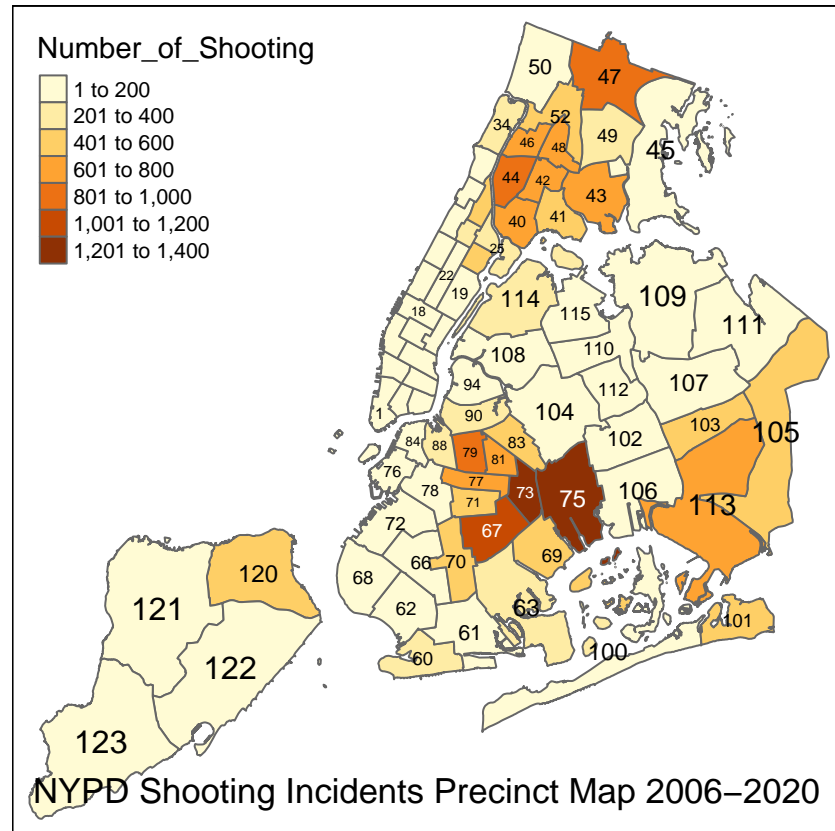
Process data and clean up values

Looking at the initial summary of the data, categorical data such as BORO, JURISDICTION_CODE, provide no useful information for analysis and were thus removed. Lon_Lat provides the same information as Longitude and Latitude in a more compact form, it is removed in this case.

Categorical information such as perpetrator race and age are often missing in the reported values.They are categorized as NA in the data set and there is no need to modify them. Missing numerical data are filled in with numerical averages or mean data, in this case, there is no numerical data missing. Perpetrator & victim's race, sex, age group, and occur time information were also removed as they are not of interest in this report.

```
summary(NYPD_shooting_processed)
```

```
##    OCCUR_DATE           PRECINCT       INCIDENT_KEY
## Min.   :2006-01-01   75     : 1367   Min.   :  9953245
## 1st Qu.:2008-12-30   73     : 1282   1st Qu.: 55317014
## Median :2012-02-26   67     : 1102   Median : 83365370
## Mean   :2012-10-03   79     :  920   Mean   :102218616
## 3rd Qu.:2016-02-28   44     :  842   3rd Qu.:150772442
## Max.   :2020-12-31   47     :  815   Max.   :222473262
##                      (Other):17240
##                   LOCATION_DESC   STATISTICAL_MURDER_FLAG   X_COORD_CD
```

```
##  MULTI DWELL - PUBLIC HOUS: 4230    Mode :logical       Min.   : 914928
##  MULTI DWELL - APT BUILD  : 2551    FALSE:19080          1st Qu.: 999900
##  PVT HOUSE                :  858    TRUE :4488           Median :1007645
##  GROCERY/BODEGA           :  572                         Mean   :1009363
##  BAR/NIGHT CLUB           :  558                         3rd Qu.:1016807
##  (Other)                  : 1218                         Max.   :1066815
##  NA's                     :13581
##    Y_COORD_CD         Latitude        Longitude
##  Min.   :125757    Min.   :40.51    Min.   :-74.25
##  1st Qu.:182565    1st Qu.:40.67    1st Qu.:-73.94
##  Median :193482    Median :40.70    Median :-73.92
##  Mean   :207312    Mean   :40.74    Mean   :-73.91
##  3rd Qu.:239163    3rd Qu.:40.82    3rd Qu.:-73.88
##  Max.   :271128    Max.   :40.91    Max.   :-73.70
##
```

Initial Investigation

Looking at the data from PRECINT, it appears that precinct 75, 73, and 79 have the highest shooting incidents, then precinct 44 and 47.

By plotting a histogram of shooting incidents based on precinct number, it is observed that there are about 2 group of precincts that has very high shooting incidents, namely, precinct 40s and precinct 70s. These group of precincts are geographically connected to each other numerically. Interestingly, because these are two peaks with precinct number very far apart, without reviewing the NYPD precinct map, one might draw the conclusion that either precinct 40s and precinct 70s are somehow neighboring districts, or they are two separate locations quite far apart.

Histogram of Shooting by Precinct
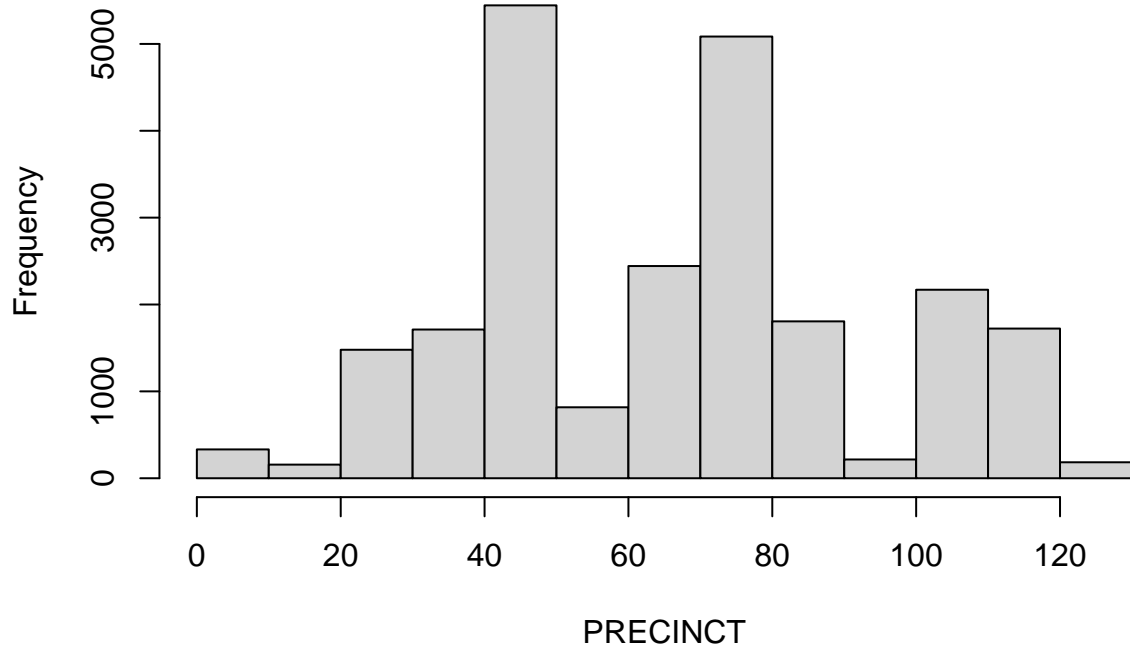
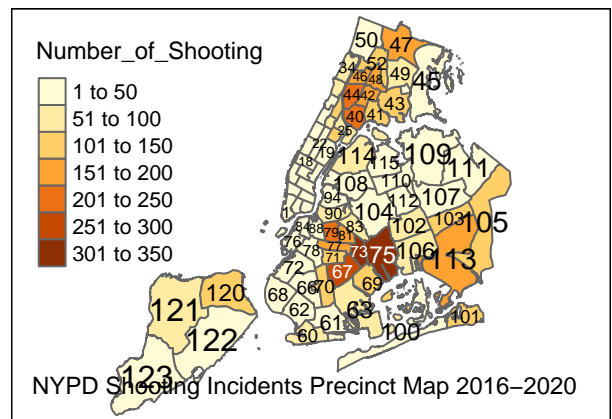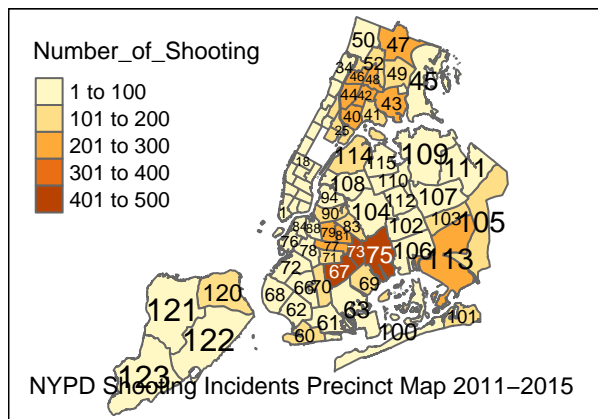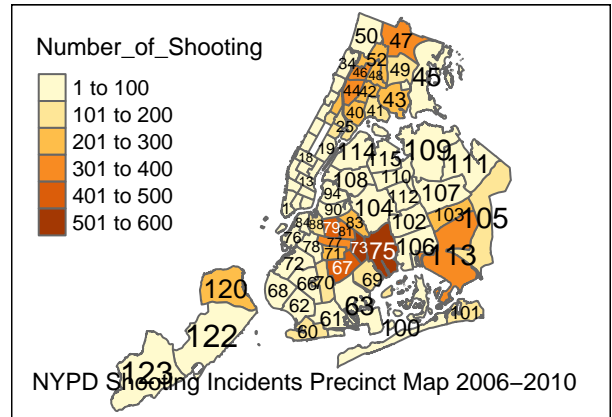## Sorting Shooting Incidents from highest to lowest

A quick sorting of the precincts and rank them from the highest occurrence to the lowest shows some striking features. The logarithmic appearance implies that shooting incidents can be modeled as a exponentially decaying function where the center of the peak values have high shooting activities or crime rate.

Shooting by Precinct sorted from highest to lowest occurance
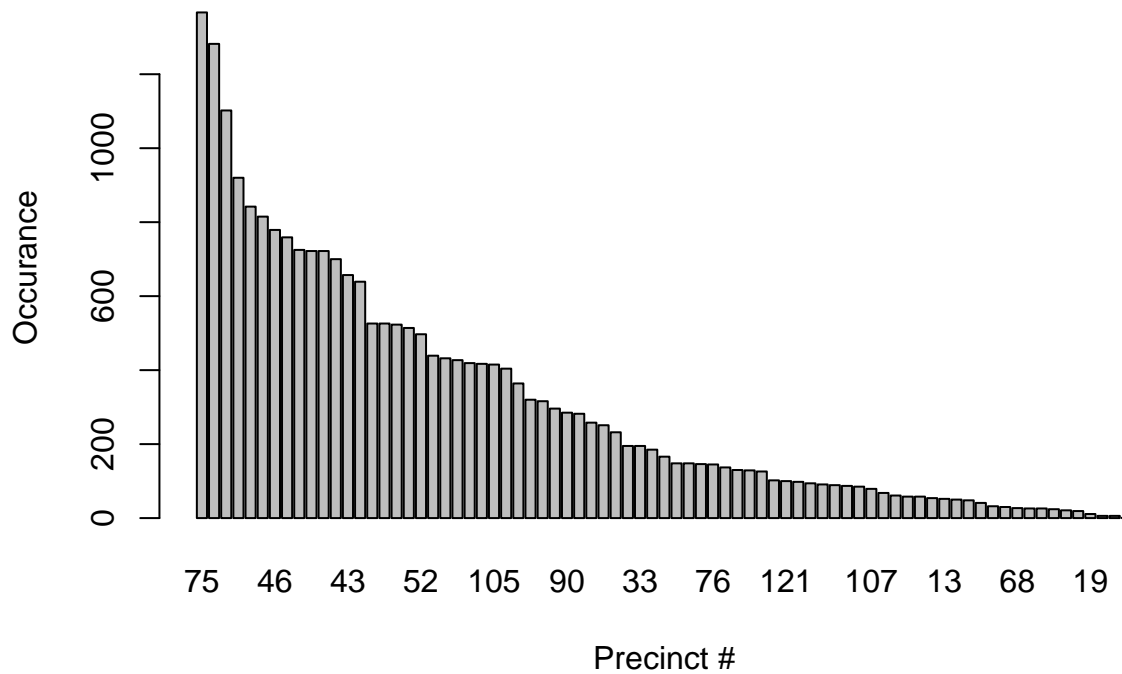
### Plotting Incident ontop of Precinct Map

By plotting all the incidents onto the NYPD Precinct Map, it basically covers the entire space.however, it is noticeable that there are less shooting on Staton Island, which is less populated than Manhattan.

Number_of_Shooting

1 to 200
201 to 400
401 to 600
601 to 800
801 to 1,000
1,001 to 1,200
1,201 to 1,400

NYPD Shooting Incidents Precinct Map 2006–2020

Number_of_Shooting

1 to 100
101 to 200
201 to 300
301 to 400
401 to 500
501 to 600

NYPD Shooting Incidents Precinct Map 2006–2010

Number_of_Shooting

1 to 100
101 to 200
201 to 300
301 to 400
401 to 500

NYPD Shooting Incidents Precinct Map 2011–2015

Number_of_Shooting

1 to 50
51 to 100
101 to 150
151 to 200
201 to 250
251 to 300
301 to 350

NYPD Shooting Incidents Precinct Map 2016–2020
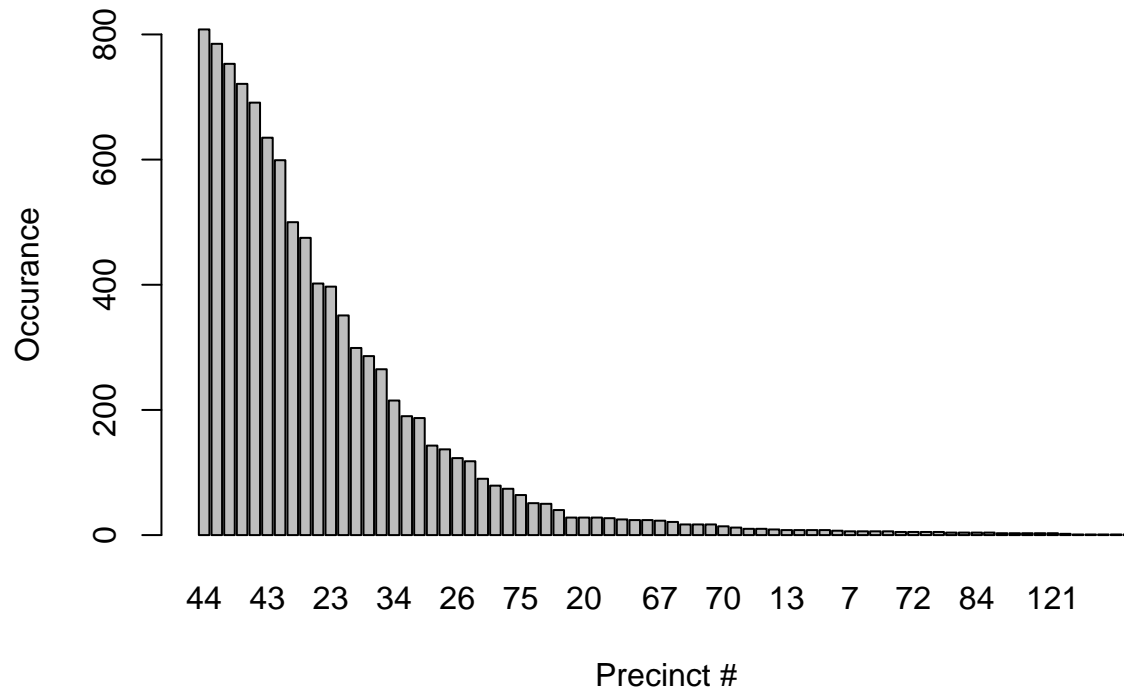
Plotting Contour Map of the Shooting Incidents

To futher explore the data set, in this case, by plotting the incidents and create a contour map of the shooting incidents, it is obvious that the contour map has 2 peaks.

Contour Plot of NYPD Shooting Incidents
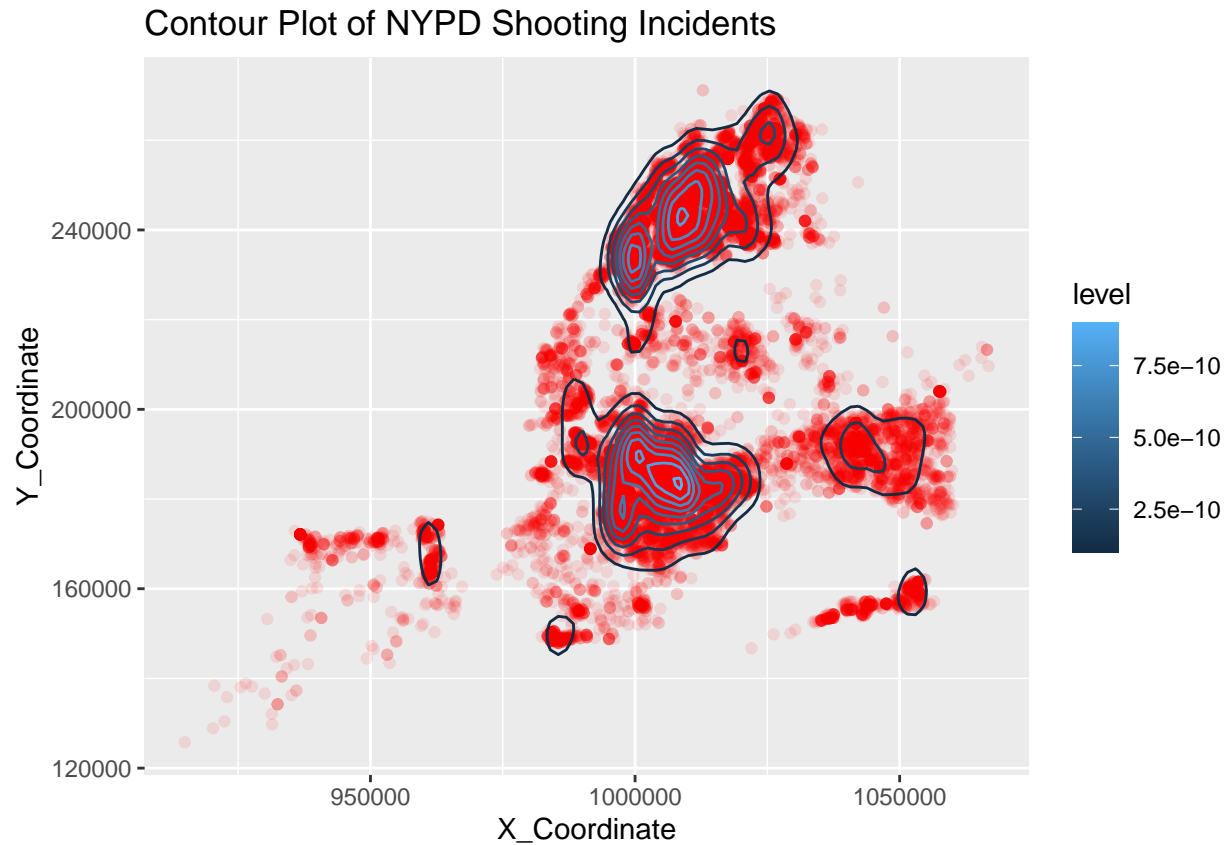
Futher Categorizing data

Given there are two peaks shown in the contour graph separated by Longtidue (Y_COORD_CD) around 210000, sorting them into two bins and re-plot the barplot gives the following results. The two plots still resembles the same exponential decaying function from the center of the peak values.

Manhattan Area Shooting by Precinct

Occurance (y-axis): 0, 200, 400, 600, 800

Precinct # (x-axis): 44, 43, 23, 34, 26, 75, 20, 67, 70, 13, 7, 72, 84, 121

Brooklyn Area Shooting by Precinct

Occurance (y-axis): 0, 200, 600, 1000

Precinct # (x-axis): 75, 77, 70, 90, 76, 9, 94, 5, 42, 46, 43, 34, 49, 26, 18
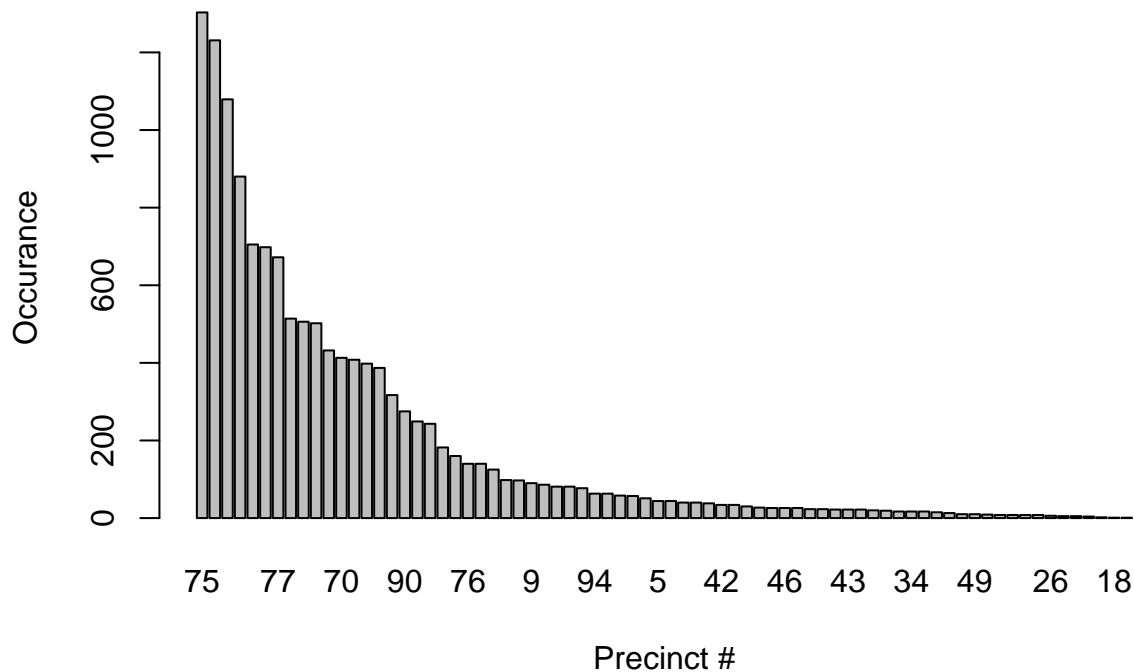
Comparing Population Density Map versus Shooting Incident Map 2016-2020

Population Density Map of New York City Plotted over Neighborhood Tabulation Areas (NTA), which is slightly different from police precinct map. It gives a general idea of population distribution over the area.

Sources:

- Poppulation data for New York City by Neighborhood Tabulation https://data.cityofnewyork.us/City-Government/New-York-City-Population-By-Neighborhood-Tabulatio/swpk-hqdp
- Shapefile for New York City Population by Neighborhood Tabulation https://data.cityofnewyork.us/City-Government/NTA-map/d3qk-pfyz

```
## Joining, by = "ntacode"
```

```
## Some legend labels were too wide. These labels have been resized to 0.65, 0.65. Increase legend.widt
```

Population Density NTA Map New York City 2010

NYPD Shooting Incidents Precinct Map 2010
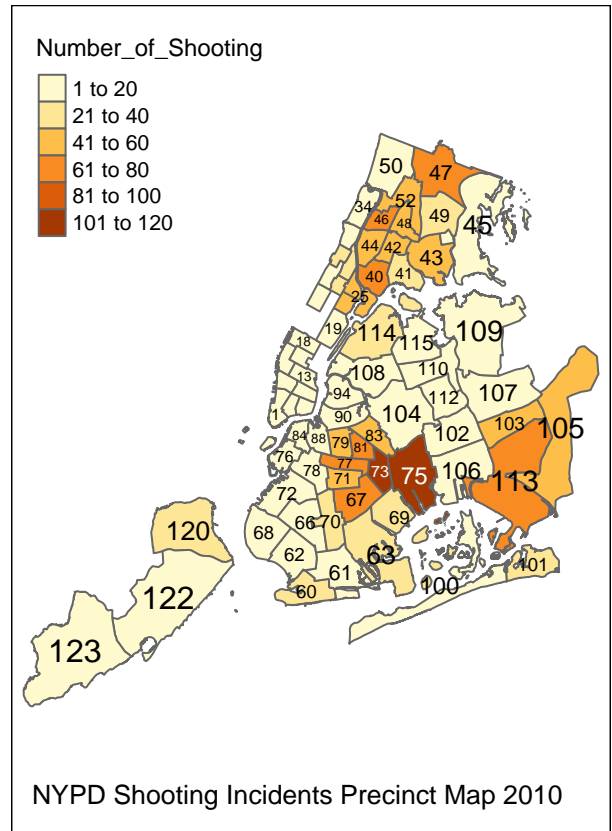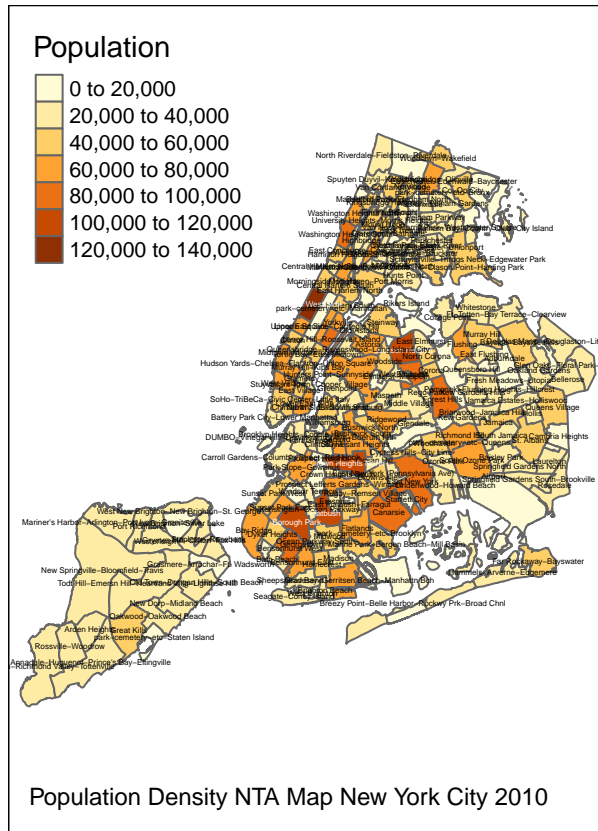
Analysis

There is basically two peak shown in the Contour Plot of the shooting incidents, each centered at Brooklyn and Manhattan, which are both heavily populated areas.

In a sense, this is another method of looking at population density of New York City where heavily populated area have higher concentration of gun violence.

For city planning purposes, the city planner would want to avoid constructing heavily populated buildings such as public housing and apartments. The fact that the summary report in Location_Desc field shows Multistory Dwelling, public housing and apartment building have the highest rate of shooting raises the question whether there is a positive correlation with the high density housing project and the occurrence of shooting. Then again, most of the Location Description field is categorized with NA or Other, might have other implications.

By comparing the population density map, one can see the shooting incidents tend to happen more frequent in the more densely populated area, with several exceptions. There are several spots in the population density map that were heavily populated in the range of 120,000 ~ 150,000, but has 1-50 shooting. And yes, due to the slight difference in the mapping between NTA and police district, the data would need to combined an remapped to give better representation. Based on difference between population density map and shooting incident map, one would see that shooting doesn't necessarily occurs in the more populated area, some area could be business district wherer more human activites happens.

Statistically murder flag shows that the majority of the incidents were non-murder related shootings, implies that gun violence does not automatically implies murder. Gun is only a choice of weapon when committing murder. However, the prevalence of shooting incidents is a sign of violence, and as policy makers, gun control

in the region needs to be better regulated. For shootings other than murder, whether they are domestic violence or gang related shootings are not categorized.

From a mathematic point of view, the center of the gun violence in those two regions have statistical significance. As gun violence drop exponentially away from the center, such behavior are usually represented as triggering event or an impulse function. In another words, eliminated the triggering event, or the center of the gun violence, would reduce gun violence significantly. My suspicion is gang activity, as this exponential drop in behavior models closely to human activity as well (i.e. sphere of influence).

Identification of Bias

Perpetrator race and victim race information were excluded from the study, as police statically identifying a person based on skin color could a source a bias. Violence, and especially gun violence is a symptom of socioeconomic condition and government policy on gun control. Personal bias against racial categorization and violence in general might affect my study on such topic.

Conclusion

This is an initial analysis of the data gathered and two distinct center of interest were identified as center of gun violence in the study. Further analysis is required to explore correlation of gun violence map in correlation to socio-economic condition, government policy (gun access), and population density. A remap of the NTA data into Police Precinct data is required to provide more accurate insight into the correlation between population density and gun violence.