

Data Story

Jeff Tsui

February 24, 2019

Analysis on Public Crime That violate Pedestrians in New York City

Introduction

Public Order Crime is an act that deviate the society's general ideas of normal social behavior and moral values. Different types of public crime can happen anywhere in the city, some borough have more crimes, other have less. By idenitifying the hotspots from the past criminal activities, it will bring caution to tourists and locals to be aware of a certain areas during certain times.

This project aim to explore and solve:

*Have the public crime activity in New York City increased or decreased at the end of 2017?

*Which borough in New York City have the most public crime activity that violate pedestrians?

*Is there any correlation between public crime activity and weather temperature?

*Which time period of the day (00:00 - 08:00, 08:01 - 16:00, or 16:01 - 24:00) has the most crime activity?

*Where are the hotspots for public crimes that violate pedestrians?

Data Acquisition

Datasets from Crime reports in NYC are availabe at NYC Opendata. I will be using the datasets of Incident Level Complaint Data - 2006 through 2017. The dataset contains: the incident level complaint from the beginning of January 2010 to the end of December 2017. The variables that I will be working on will be the boroughs, date of the complaints, time of the complaints, level of offenses, description of the offenses, description of the premises, suspect's age group, suspect's race, suspect's sex, victim's age group, victim's race, and victim's sex.

Dataset of daily average temperature from 2010 to 2017 in New York City is obtained from Weather Underground.

Important Fields and Information

According to the NY Police Department post on the The New York Job Source, the NYPD shifts are divided by three 8-hours and 35 minute shifts: 11:15 PM to 7:50 AM, 7:05 AM to 3:40 PM, and 3:00 PM tp 11:35 PM. But to simplify they call the shifts: (12 to 8), (8 to 4), and (4 to 12). I will be creating 3 time period of the day correspond to the police shifts to test out whih time period of the day has the most crime.

Data Limitations

Since there isn't a specific whole area weather temperature for the entire New York City that includes all five boroughs on the historical data on the Weather Underground website. I took the average temperature of the most centered borough (Manhattan).

There are limited data on the suspect's age, race, and sex because there might be a case where the suspect was never caught. As well as there are limited data on the victim's age, race and sex because of the protection of personal information.

None of the murder crimes have any premises description in the dataset of NYC Opendata, therefore none of them was included in this project. Since this crime is one of the most serious crime and the worst crime that can happen to a pedestrian, without the data of this crime can impact the attention that the locals and tourists would have give.

Data Cleaning and wrangling

The following packages was used for data cleaning and wrangling: `tidyverse`, `dplyr`, `lubridate`, `chron`, and `zoo`.

*Deleting useless columns by using e.g. `df[, -c(1,2,3,4)]`.

*Rearranging the columns by using e.g. `df[, c(2,1,3,4)]`.

*Renaming the columns to become more readable by using colnames.

*Used the `select()` and `filter()` function from the `dplyr` package to filter out all premises except public premises: “PARK/PLAYGROUND”, “PARKING LOT/GARAGE(PUBLIC)”, “BUS (NYC TRANSIT)”, “OPEN AREAS (OPEN LOTS)”, “BUS STOP”, “STREET”, “TRANSIT - NYC SUBWAY”, “PUBLIC BUILDING”.

*Used the `select()` and `filter()` function from the `dplyr` package to filter out all offensive except the ones that affects pedestrians: “ARSON”, “ASSAULT & RELATED OFFENSES”, “DANGEROUS WEAPONS”, “FELONY ASSAULT”, “HARRASSMENT”, “KIDNAPPING”, “MURDER & NON-NEGLECTFUL MANSLAUGHTER”, “RAPE”, “ROBBERY”, “SEX CRIMES”.

*Used the `year` function from the `lubridate` package to add a new column for the year.

*Used the `yearmon` function from the `zoo` package to add a new column for the year with month.

*Used the `chron` function from the `chron` package to convert the rows in the Complaint time column into the formate of “h:m:s”

Preliminary exploration

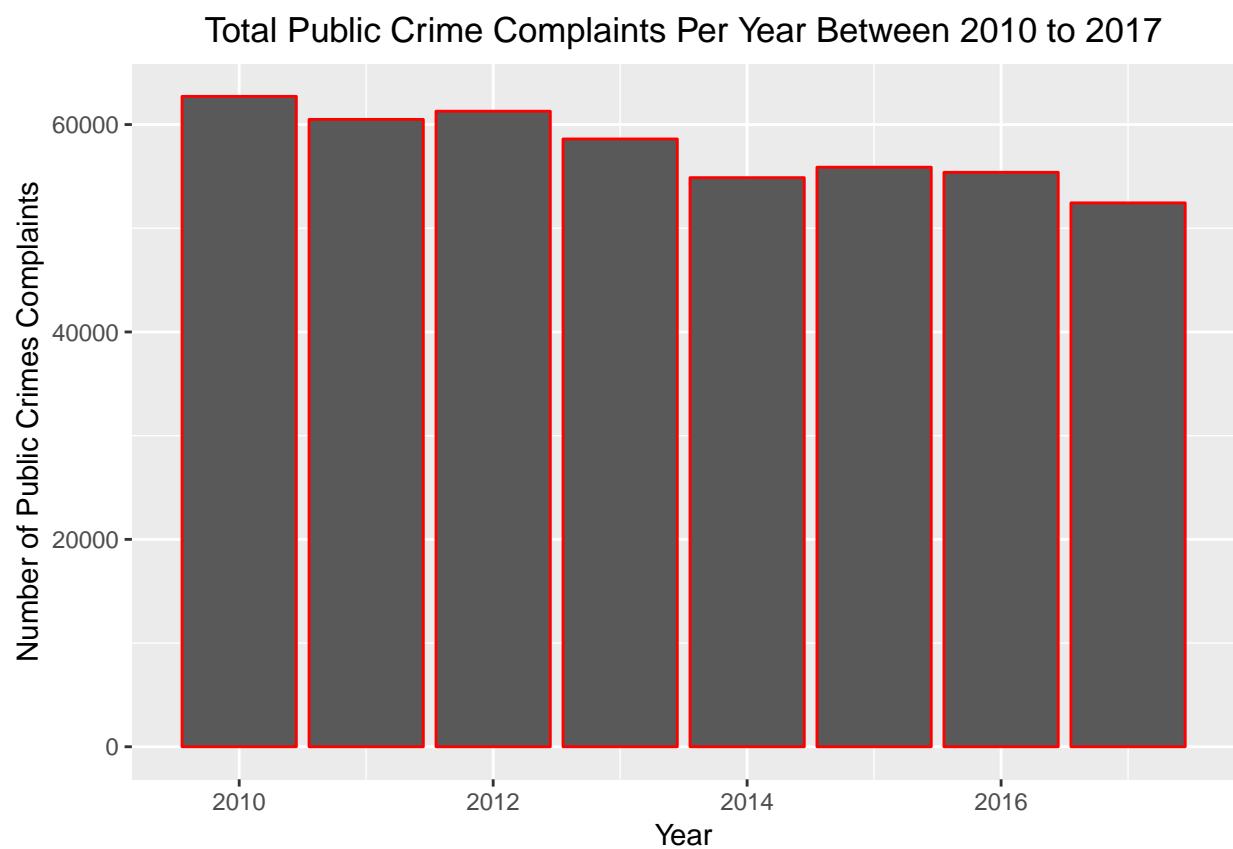
Load the library

```
library(dplyr)
library(tidyverse)
library(scales)
library(plyr)
library(lubridate)
library(ggplot2)
library(zoo)
library(gtools)
library(chron)
```

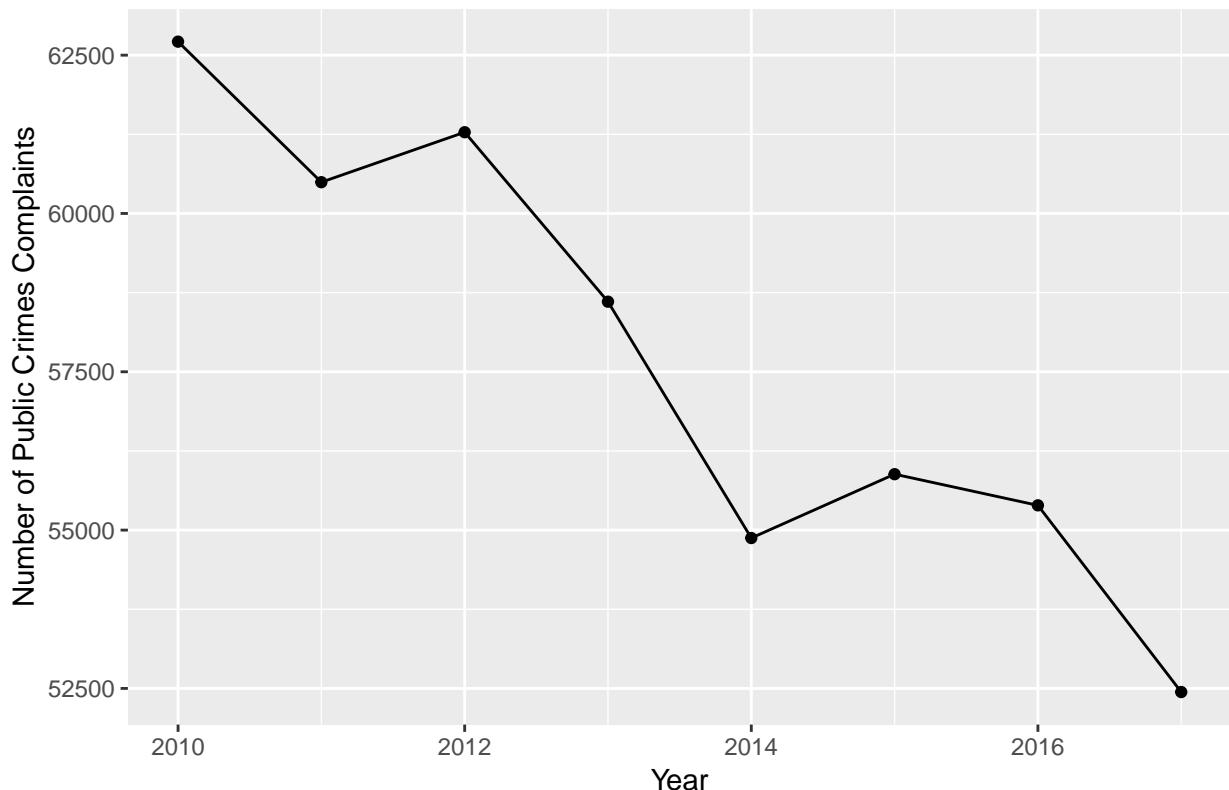
Load the dataframe

The dataframe only included the reported complaints of crimes that violates other people in the public. The public premises includes: “bus stop”, “open areas (open lots)”, “park/playground”, “public buildings”, “street”, “transit (bus)”, and “transit (subway)”. The public crimes includes: “arson”, “assault and related offenses”, “dangerous weapons”, “felony assault”, “harrassment”, “kidnapping”, “rape”, “robbery”, and “sex crimes”. The complaint reported for public crimes are 11.8% of all reported crimes between 2010 to 2017.

Analysis of Crime Complaints that happened in the Public that Violates Other People



Total Public Crime Complaints Per Year Between 2010 to 2017



```
#Number of public crime complaints in 2010  
length(which(AllComplaint$Year == 2010))
```

```
## [1] 62713
```

```
#Number of public crime complaints in 2017  
length(which(AllComplaint$Year == 2017))
```

```
## [1] 52443
```

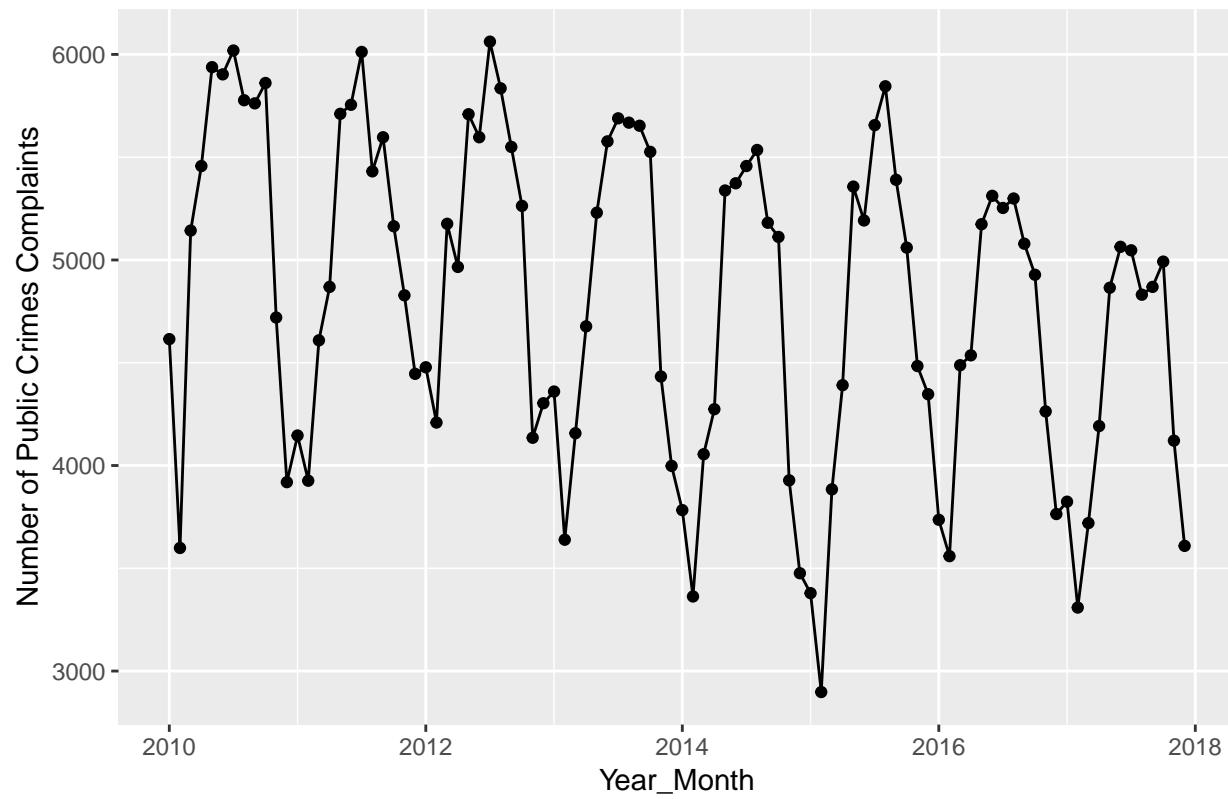
```
#The percentage change in public crime complaints from 2010 to 2017  
A <- length(which(AllComplaint$Year == 2010))  
B <- length(which(AllComplaint$Year == 2017))  
(A-B)/A
```

```
## [1] 0.1637619
```

The line graph below shows the changes more clearly, there is a trend of decreasing public crimes between 2010 to 2017.

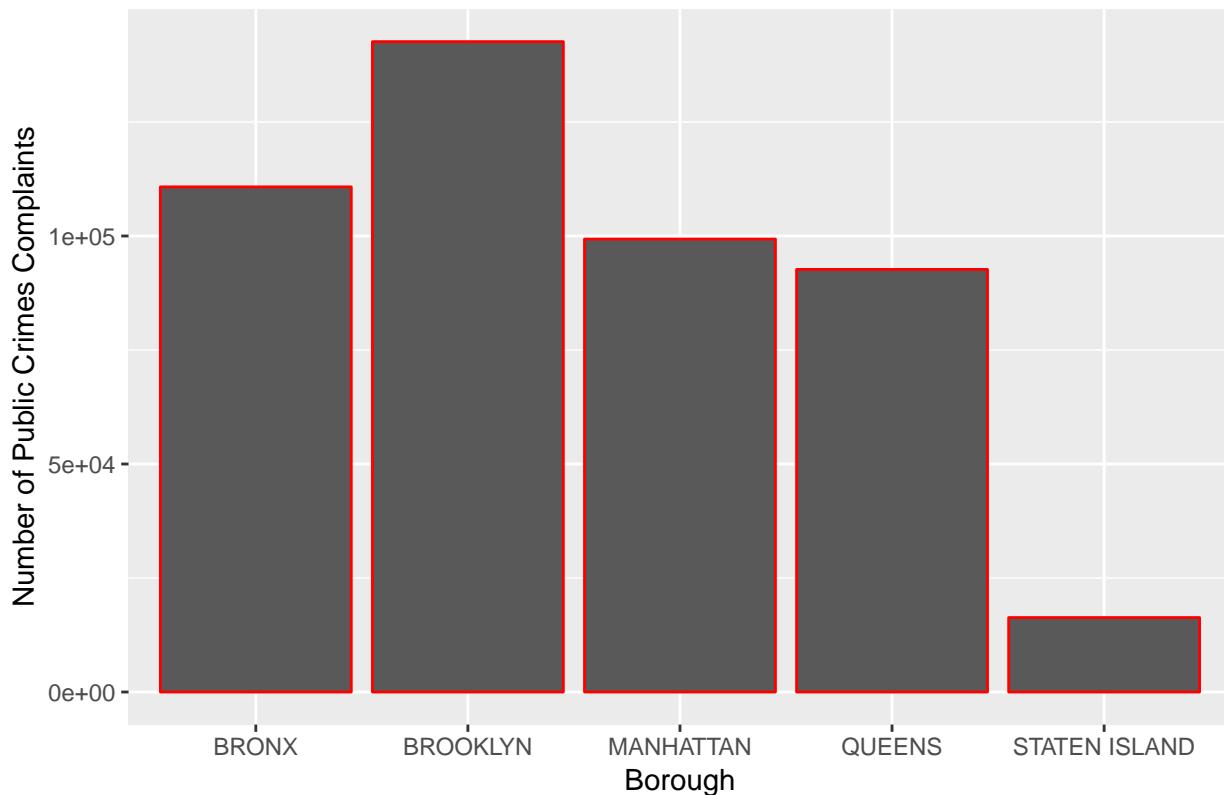
There is a 16.38% decrease of public crimes between 2010 to 2017.

Total Public Crime Complaints Per Month Between 2010 to 2017



There is a trend of decrease in public crime complaints every year between the month of November, December, and January. The crime complaints start to increase after January and reaches the peak around the month of June and July of each year.

Total Public Crime Complaints Per Borough Between 2010 to 2017



According to the bar graph above. It clearly shown that Staten Island have significantly less public crime complaints than the other four boroughs. This could be the case of the size of population. According to the city population website, Staten Island only hold 5.6% of the New York City population. But Brooklyn (31% NYC population) shows to have significantly more public crime complaints than other boroughs even when Queens (27% NYC population) have similar population size. Population size are determined by City Population

USA: New York City Boroughs

Boroughs

The population of the boroughs of New York City according to census results and latest official estimates.

Name	Status	Population	Population	Population	Population
		Census 1990-04-01	Census 2000-04-01	Census 2010-04-01	Estimate 2017-07-01
Bronx	Borough	1,203,789	1,332,244	1,384,794	1,471,160
Brooklyn (Kings County)	Borough	2,300,664	2,465,689	2,504,706	2,648,771
Manhattan (New York County)	Borough	1,487,536	1,538,096	1,586,184	1,664,727
Queens	Borough	1,951,598	2,229,394	2,230,545	2,358,582
Staten Island (Richmond County)	Borough	378,977	443,762	468,730	479,458
New York City	City	7,322,564	8,009,185	8,174,959	8,622,698

Source: US Census Bureau (web).

```
length(which(AllComplaint$Borough == "BRONX")) /1471160
```

```

## [1] 0.07528889

length(which(AllComplaint$Borough == "BROOKLYN")) /2648771

## [1] 0.05384195

length(which(AllComplaint$Borough == "MANHATTAN")) /1664727

## [1] 0.05965783

length(which(AllComplaint$Borough == "QUEENS")) /2358582

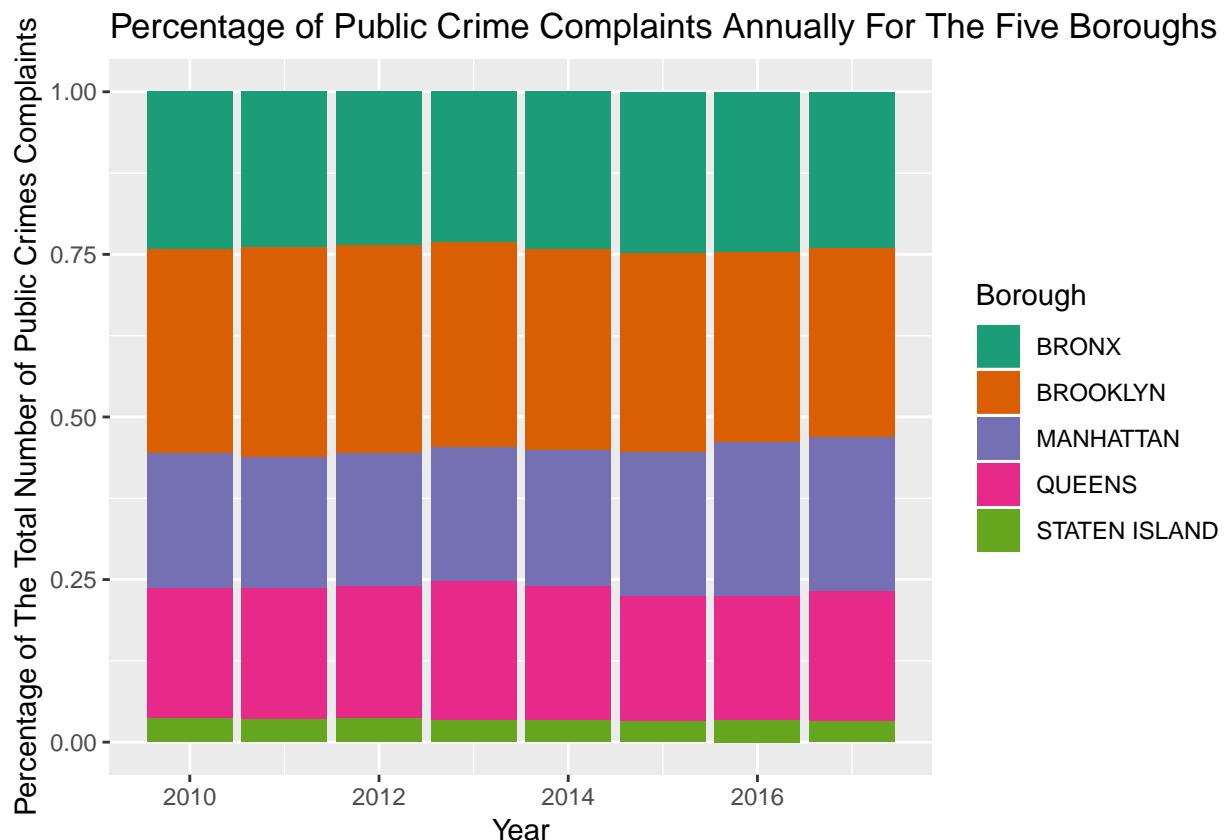
## [1] 0.03929013

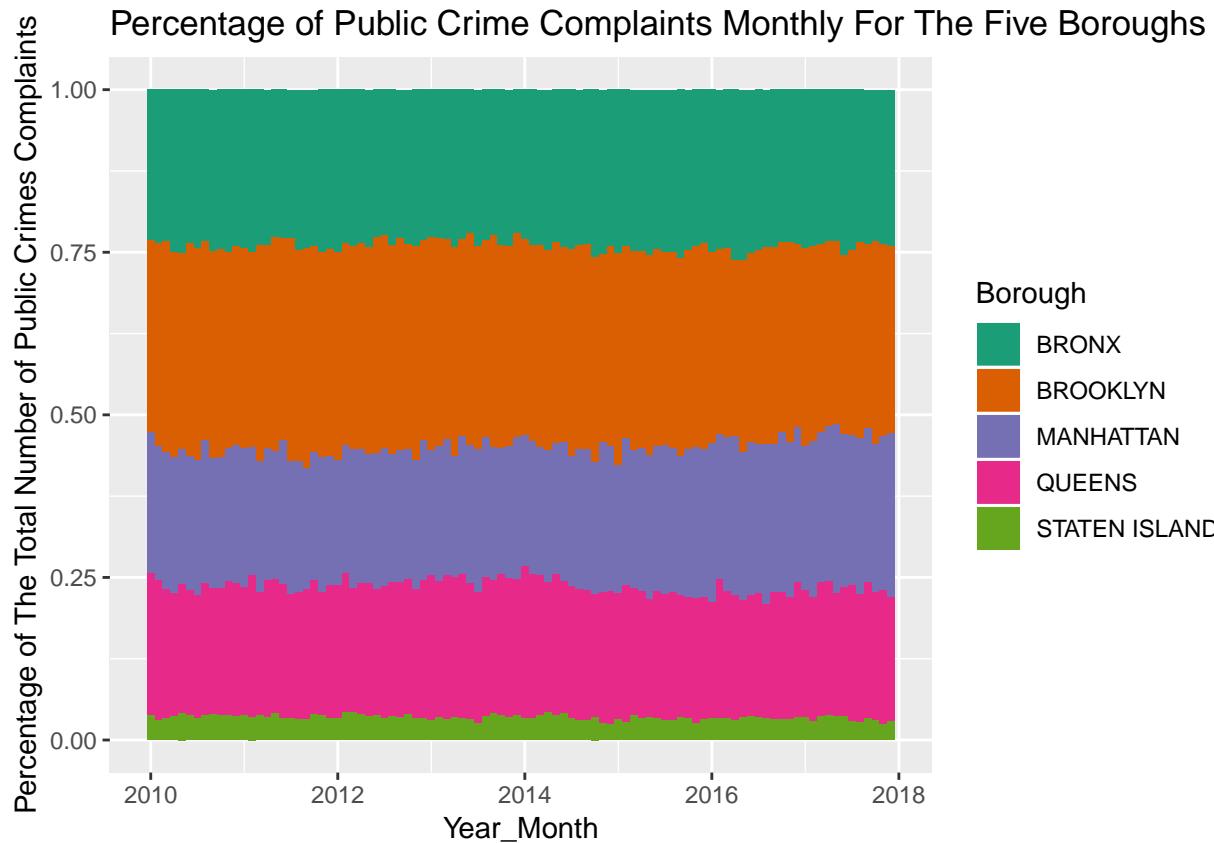
length(which(AllComplaint$Borough == "STATEN ISLAND")) /479458

## [1] 0.03405721

```

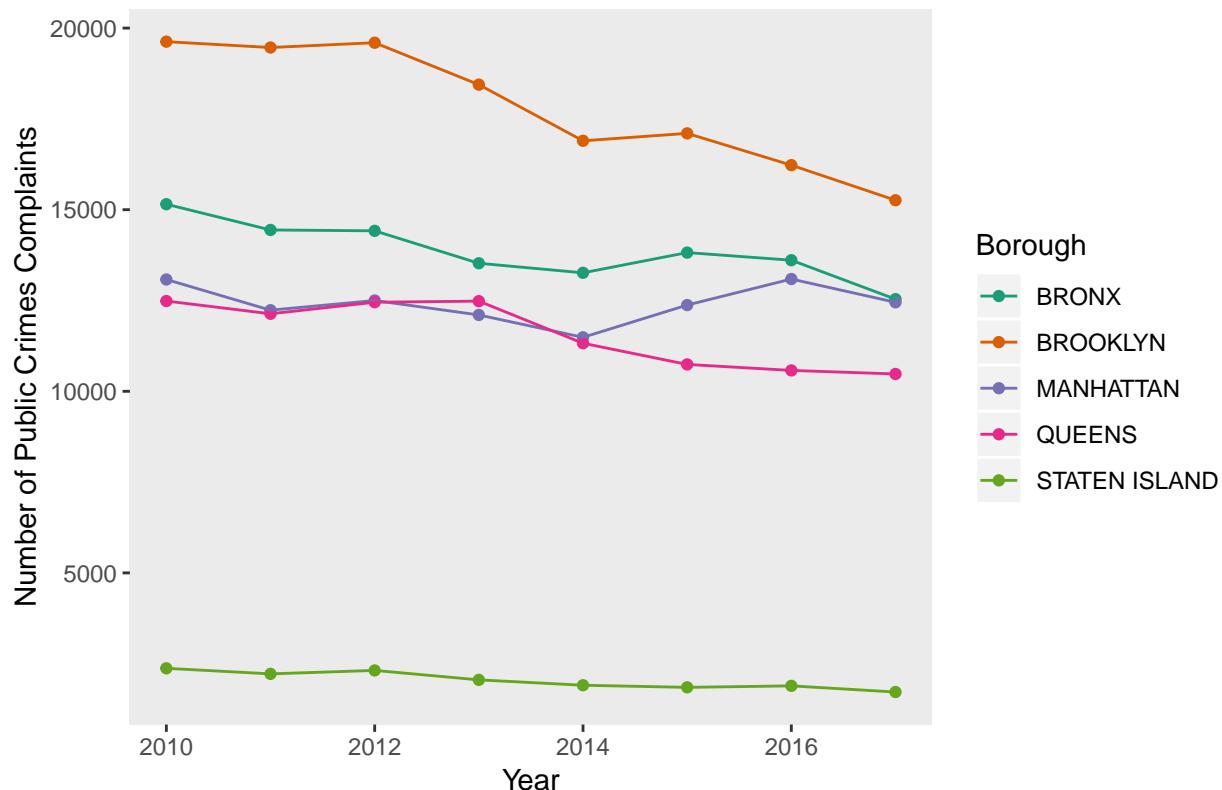
From the year of 2010 to 2017, there is a public crime complaint rate of 7.5 per 100 people in Bronx, 5.4 per 100 people in Brooklyn, 6 per 100 people in Manhattan, 3.9 per 100 people in Queens, and 3.4 per 100 people in Staten Island. Even though Brooklyn have the most public crime complaints overall, Bronx and Manhattan have more public crime complaint per 100 people.





Both of the bar graph above shown that the percentage of public crime complaints for the five boroughs did not have any huge changes yearly or monthly. Bronx's public crime complaints averaged around a little less than 25% of the total number of complaints. Brooklyn's public crime complaints averaged around 30% of the total number of complaints. Manhattan and Queens averaged around 20% of the total number of complaints. Staten Island averaged a little less than 5% of the total number of complaints.

Annual Total Public Crime Complaints for The Five Boroughs



From the annual total public crime complaint line graph above, it had shown that Brooklyn's public crime complaints have a more noticeable sight of decreasing over the years than the other boroughs.

```
#Number of public crime complaints in 2010 For Bronx
length(which(AllComplaint$Year == 2010 & AllComplaint$Borough == "BRONX"))
```

```
## [1] 15151
```

```
#Number of public crime complaints in 2017 For Bronx
length(which(AllComplaint$Year == 2017 & AllComplaint$Borough == "BRONX"))
```

```
## [1] 12536
```

```
#The percentage change in public crime complaints from 2010 to 2017 in Bronx
BX10 <- length(which(AllComplaint$Year == 2010 & AllComplaint$Borough == "BRONX"))
BX17 <- length(which(AllComplaint$Year == 2017 & AllComplaint$Borough == "BRONX"))
(BX10 - BX17)/BX10
```

```
## [1] 0.1725959
```

```
#Number of public crime complaints in 2010 For Brooklyn
length(which(AllComplaint$Year == 2010 & AllComplaint$Borough == "BROOKLYN"))
```

```
## [1] 19626
```

```

#Number of public crime complaints in 2017 For Brooklyn
length(which(AllComplaint$Year == 2017 & AllComplaint$Borough == "BROOKLYN"))

## [1] 15260

#The percentage change in public crime complaints from 2010 to 2017 in Brooklyn
BN10 <- length(which(AllComplaint$Year == 2010 & AllComplaint$Borough == "BROOKLYN"))
BN17 <- length(which(AllComplaint$Year == 2017 & AllComplaint$Borough == "BROOKLYN"))
(BN10 - BN17)/BN10

## [1] 0.22246

#Number of public crime complaints in 2010 For Manhattan
length(which(AllComplaint$Year == 2010 & AllComplaint$Borough == "MANHATTAN"))

## [1] 13078

#Number of public crime complaints in 2017 For Manhattan
length(which(AllComplaint$Year == 2017 & AllComplaint$Borough == "MANHATTAN"))

## [1] 12450

#The percentage change in public crime complaints from 2010 to 2017 in Manhattan
MN10 <- length(which(AllComplaint$Year == 2010 | AllComplaint$Borough == "MANHATTAN"))
MN17 <- length(which(AllComplaint$Year == 2017 | AllComplaint$Borough == "MANHATTAN"))
(MN10 - MN17)/MN10

## [1] 0.06473357

#Number of public crime complaints in 2010 For Queens
length(which(AllComplaint$Year == 2010 & AllComplaint$Borough == "QUEENS"))

## [1] 12485

#Number of public crime complaints in 2017 For Queens
length(which(AllComplaint$Year == 2017 & AllComplaint$Borough == "QUEENS"))

## [1] 10477

#The percentage change in public crime complaints from 2010 to 2017 in Queens
QS10 <- length(which(AllComplaint$Year == 2010 | AllComplaint$Borough == "QUEENS"))
QS17 <- length(which(AllComplaint$Year == 2017 | AllComplaint$Borough == "QUEENS"))
(QS10 - QS17)/QS10

## [1] 0.05781787

```

```
#Number of public crime complaints in 2010 For Staten Island
length(which(AllComplaint$Year == 2010 & AllComplaint$Borough == "STATEN ISLAND"))
```

```
## [1] 2373
```

```
#Number of public crime complaints in 2017 For Staten Island
length(which(AllComplaint$Year == 2017 & AllComplaint$Borough == "STATEN ISLAND"))
```

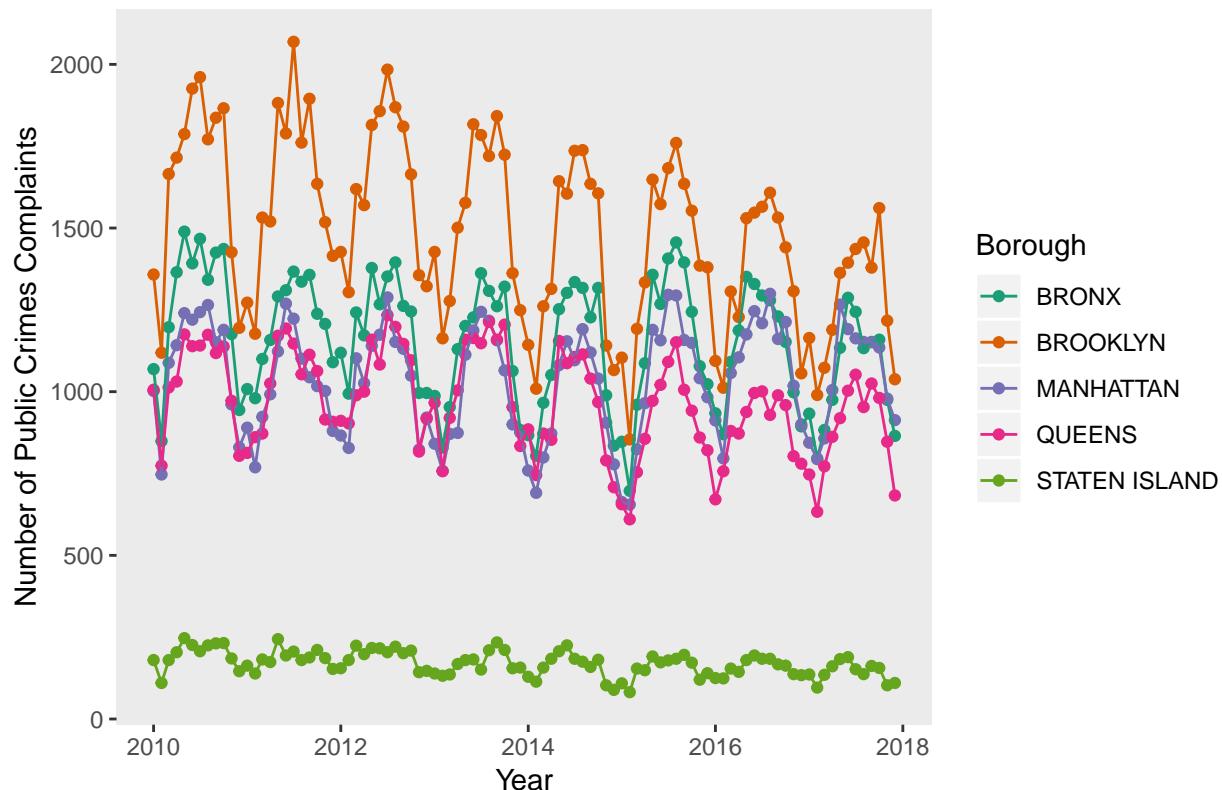
```
## [1] 1720
```

```
#The percentage change in public crime complaints from 2010 to 2017 in Staten Island
SD10 <- length(which(AllComplaint$Year == 2010 & AllComplaint$Borough == "STATEN ISLAND"))
SD17 <- length(which(AllComplaint$Year == 2017 & AllComplaint$Borough == "STATEN ISLAND"))
(SD10 - SD17)/SD10
```

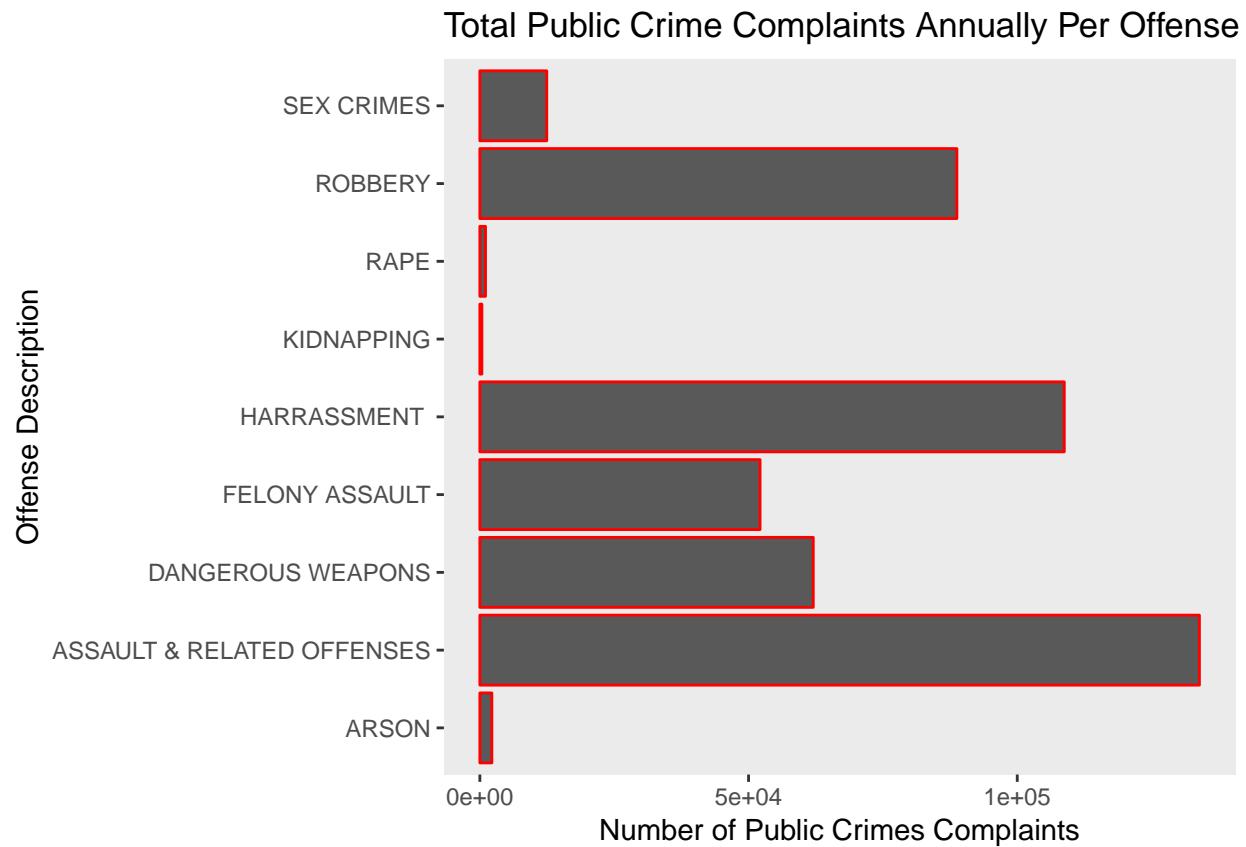
```
## [1] 0.2751791
```

The public crime complaints have decreased for all five boroughs at the end of 2017 from 2010. Bronx with 17.3% decrease, Brooklyn with 22.2% decrease, Manhattan with 6.5% decrease, Queens with 5.8% decrease, and Staten Island with 27.5% decrease. Brooklyn still have the most public crime complaints even though its public crime complaints was decreased by 22.2% over the years, and Queens and Manhattan's public crime complaints have barely changed.

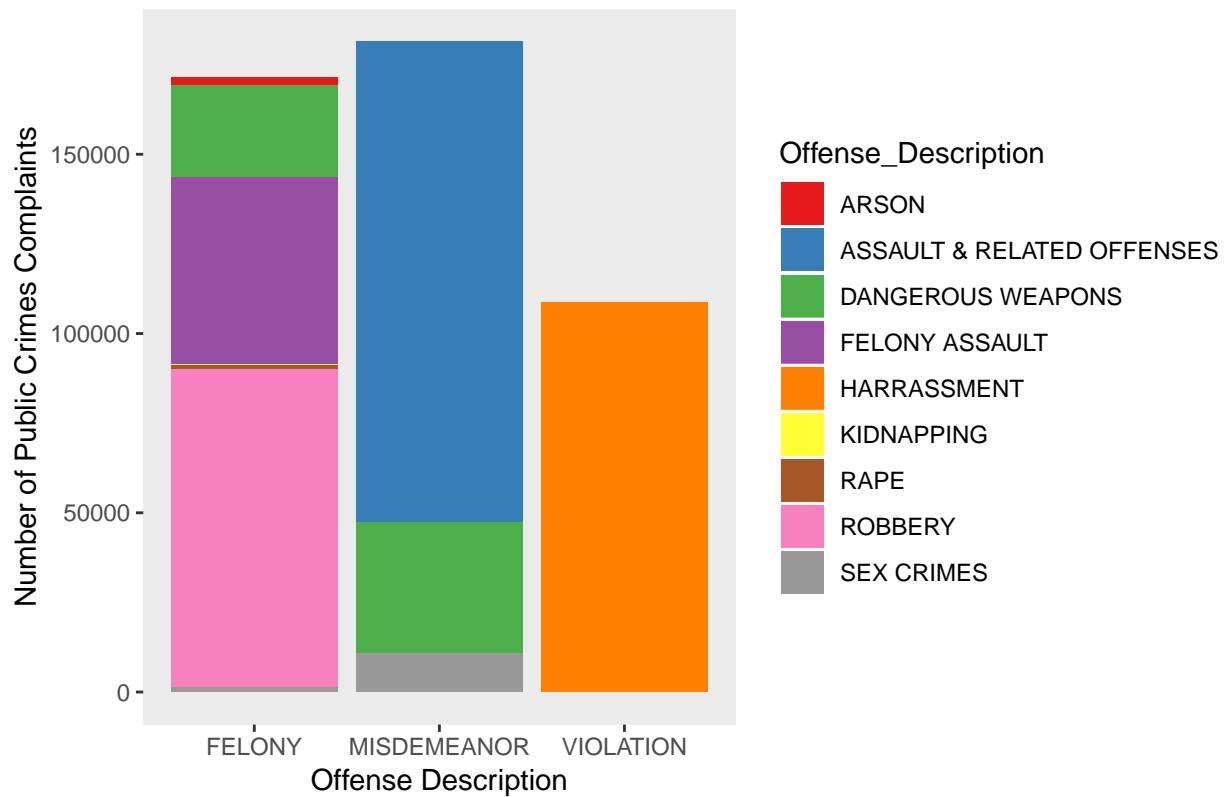
Total Public Crime Complaints Monthly for The Five Borough



The monthly complaints for the five boroughs line graph above once again shown that there is a significantly less public crime complaints during the month of November, December, and January. And there is a significantly more public crime complaints during the month of June, July, and August.



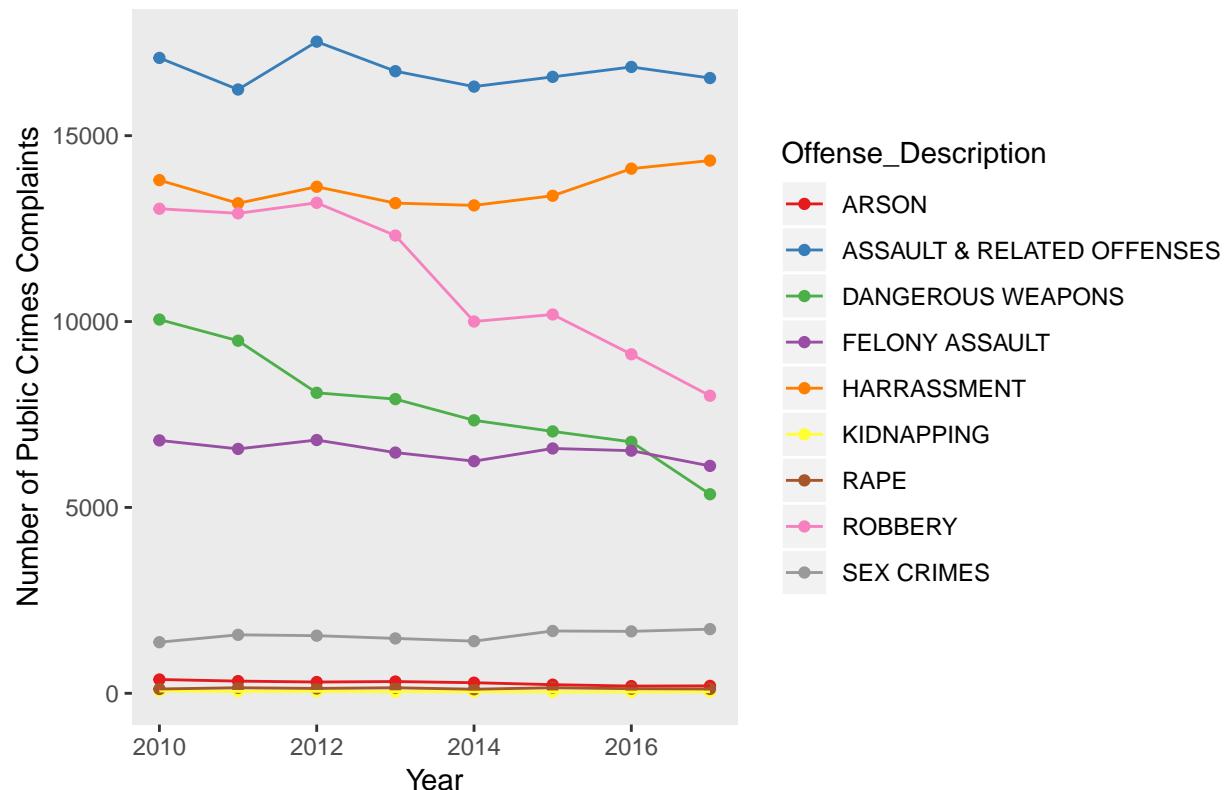
Total Public Crime Complaints Annually Per Offense



In the world of Criminal Justice, violations are considered to be the most minor of offenses. Violations are punishable by a fine and will not result in any jail or prison time. Misdemeanor offenses are more serious than violation offenses. Misdemeanor can result up to one year in jail. Felonies are the most offense out of the three. Felonies are separated by letter (Class A - Class E). Class A felonies are the most serious and class E is the least. Punishable by class (A: up to life time in prison, B: 25 years+, C: 10 - 25 years, D: 5 - 10 years, and E: 1 - 5 years). Legal Dictionary

Crimes such as arson, kidnapping, and rape are considered class A felonies, therefore those crimes are the minority among the total public crime complaints. Robbery and felony assault are considered class B felonies, which is the majority of the number of public crime complaints. Sex crime and the possession of dangerous weapons are mixture of felony and misdemeanor. Assault and related offenses are the majority of misdemeanor offenses. And harassment is only considered as violation.

Annual Total Public Crime Complaints Per Offenses



The annual total public crime complaints per offenses line graph above shown that harrassment and sex crimes had clearly increased. The possession of dangerous weapons and robbery had clearly decreased. But other offenses are hard to tell.

```
#Number of public crime complaints in 2010 For arson
length(which(AllComplaint$Year == 2010 & AllComplaint$Offense_Description == "ARSON"))
```

```
## [1] 371
```

```
#Number of public crime complaints in 2017 For arson
length(which(AllComplaint$Year == 2017 & AllComplaint$Offense_Description == "ARSON"))
```

```
## [1] 199
```

```
#The percentage change in public crime complaints from 2010 to 2017 in arson
AN10 <- length(which(AllComplaint$Year == 2010 & AllComplaint$Offense_Description == "ARSON"))
AN17 <- length(which(AllComplaint$Year == 2017 & AllComplaint$Offense_Description == "ARSON"))
(AN10 - AN17)/AN10
```

```
## [1] 0.4636119
```

```
#Number of public crime complaints in 2010 For assault & related offenses
length(which(AllComplaint$Year == 2010 & AllComplaint$Offense_Description == "ASSAULT & RELATED OFFENSES"))
```

```
## [1] 17500
```

```

## [1] 17092

#Number of public crime complaints in 2017 For assault & related offenses
length(which(AllComplaint$Year == 2017 & AllComplaint$Offense_Description == "ASSAULT & RELATED OFFENSES"))

## [1] 16551

#The percentage change in public crime complaints from 2010 to 2017 in assault & related offenses
AT10 <- length(which(AllComplaint$Year == 2010 & AllComplaint$Offense_Description == "ASSAULT & RELATED OFFENSES"))
AT17 <- length(which(AllComplaint$Year == 2017 & AllComplaint$Offense_Description == "ASSAULT & RELATED OFFENSES"))
(AT10 - AT17)/AT10

## [1] 0.03165223

#Number of public crime complaints in 2010 For dangerous weapons
length(which(AllComplaint$Year == 2010 & AllComplaint$Offense_Description == "DANGEROUS WEAPONS"))

## [1] 10054

#Number of public crime complaints in 2017 For dangerous weapons
length(which(AllComplaint$Year == 2017 & AllComplaint$Offense_Description == "DANGEROUS WEAPONS"))

## [1] 5355

#The percentage change in public crime complaints from 2010 to 2017 in dangerous weapons
DW10 <- length(which(AllComplaint$Year == 2010 & AllComplaint$Offense_Description == "DANGEROUS WEAPONS"))
DW17 <- length(which(AllComplaint$Year == 2017 & AllComplaint$Offense_Description == "DANGEROUS WEAPONS"))
(DW10 - DW17)/DW10

## [1] 0.4673762

#Number of public crime complaints in 2010 For felony assault
length(which(AllComplaint$Year == 2010 & AllComplaint$Offense_Description == "FELONY ASSAULT"))

## [1] 6803

#Number of public crime complaints in 2017 For felony assault
length(which(AllComplaint$Year == 2017 & AllComplaint$Offense_Description == "FELONY ASSAULT"))

## [1] 6117

#The percentage change in public crime complaints from 2010 to 2017 in felony assault
FA10 <- length(which(AllComplaint$Year == 2010 & AllComplaint$Offense_Description == "FELONY ASSAULT"))
FA17 <- length(which(AllComplaint$Year == 2017 & AllComplaint$Offense_Description == "FELONY ASSAULT"))
(FA10 - FA17)/FA10

## [1] 0.1008379

```

```

#Number of public crime complaints in 2010 For harrassment
length(which(AllComplaint$Year == 2010 & AllComplaint$Offense_Description == "HARRASSMENT"))

## [1] 13804

#Number of public crime complaints in 2017 For harrassment
length(which(AllComplaint$Year == 2017 & AllComplaint$Offense_Description == "HARRASSMENT"))

## [1] 14331

#The percentage change in public crime complaints from 2010 to 2017 in harrassment
HT10 <- length(which(AllComplaint$Year == 2010 & AllComplaint$Offense_Description == "HARRASSMENT"))
HT17 <- length(which(AllComplaint$Year == 2017 & AllComplaint$Offense_Description == "HARRASSMENT"))
(HT10 - HT17)/HT10

## [1] -0.03817734

#Number of public crime complaints in 2010 For kidnapping
length(which(AllComplaint$Year == 2010 & AllComplaint$Offense_Description == "KIDNAPPING"))

## [1] 64

#Number of public crime complaints in 2017 For kidnapping
length(which(AllComplaint$Year == 2017 & AllComplaint$Offense_Description == "KIDNAPPING"))

## [1] 44

#The percentage change in public crime complaints from 2010 to 2017 in kidnapping
KG10 <- length(which(AllComplaint$Year == 2010 & AllComplaint$Offense_Description == "KIDNAPPING"))
KG17 <- length(which(AllComplaint$Year == 2017 & AllComplaint$Offense_Description == "KIDNAPPING"))
(KG10 - KG17)/KG10

## [1] 0.3125

#Number of public crime complaints in 2010 For rape
length(which(AllComplaint$Year == 2010 & AllComplaint$Offense_Description == "RAPE"))

## [1] 116

#Number of public crime complaints in 2017 For rape
length(which(AllComplaint$Year == 2017 & AllComplaint$Offense_Description == "RAPE"))

## [1] 0

#The percentage change in public crime complaints from 2010 to 2017 in rape
RE10 <- length(which(AllComplaint$Year == 2010 & AllComplaint$Offense_Description == "RAPE"))
RE17 <- length(which(AllComplaint$Year == 2017 & AllComplaint$Offense_Description == "RAPE"))
(RE10 - RE17)/RE10

```

```

#Number of public crime complaints in 2010 For robbery
length(which(AllComplaint$Year == 2010 & AllComplaint$Offense_Description == "ROBBERY"))

## [1] 13035

#Number of public crime complaints in 2017 For robbery
length(which(AllComplaint$Year == 2017 & AllComplaint$Offense_Description == "ROBBERY"))

## [1] 8004

#The percentage change in public crime complaints from 2010 to 2017 in robbery
RY10 <- length(which(AllComplaint$Year == 2010 & AllComplaint$Offense_Description == "ROBBERY"))
RY17 <- length(which(AllComplaint$Year == 2017 & AllComplaint$Offense_Description == "ROBBERY"))
(RY10 - RY17)/RY10

## [1] 0.3859609

#Number of public crime complaints in 2010 For sex crimes
length(which(AllComplaint$Year == 2010 & AllComplaint$Offense_Description == "SEX CRIMES"))

## [1] 1374

#Number of public crime complaints in 2017 For sex crimes
length(which(AllComplaint$Year == 2017 & AllComplaint$Offense_Description == "SEX CRIMES"))

## [1] 1726

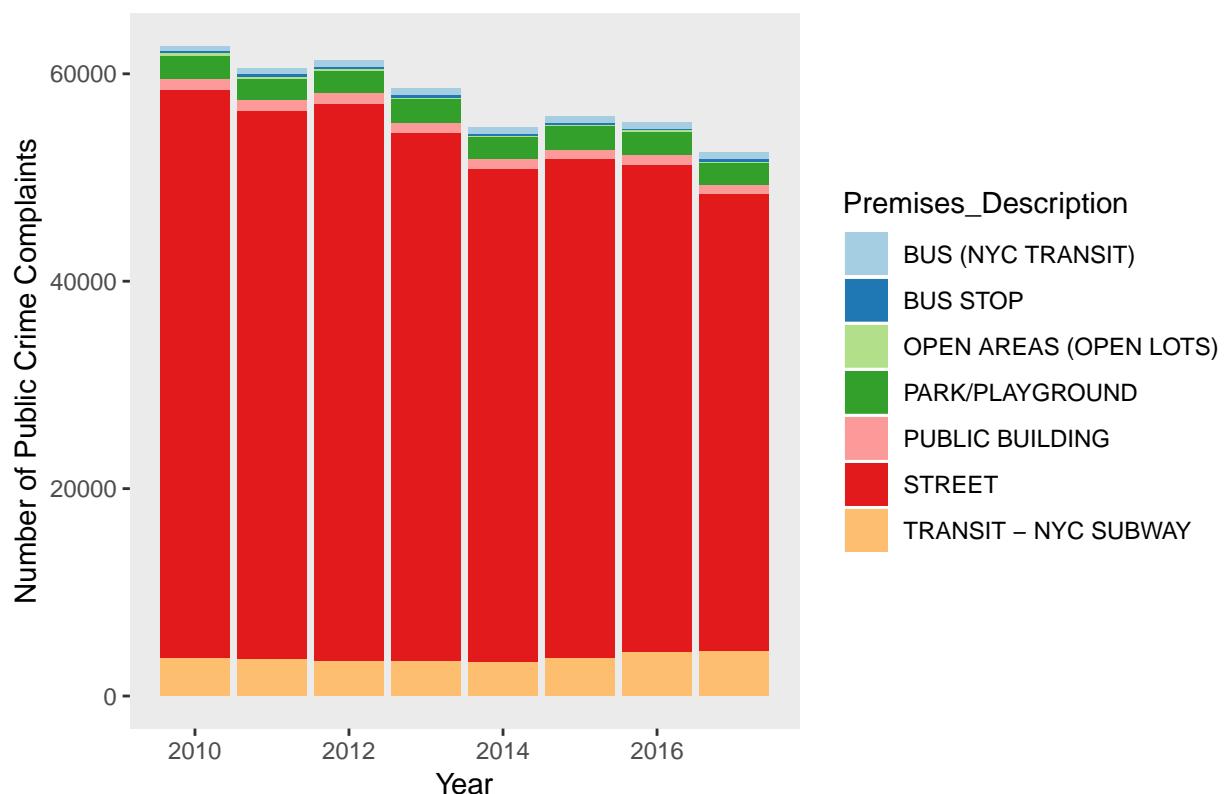
#The percentage change in public crime complaints from 2010 to 2017 in sex crimes
SC10 <- length(which(AllComplaint$Year == 2010 & AllComplaint$Offense_Description == "SEX CRIMES"))
SC17 <- length(which(AllComplaint$Year == 2017 & AllComplaint$Offense_Description == "SEX CRIMES"))
(SC10 - SC17)/SC10

## [1] -0.2561863

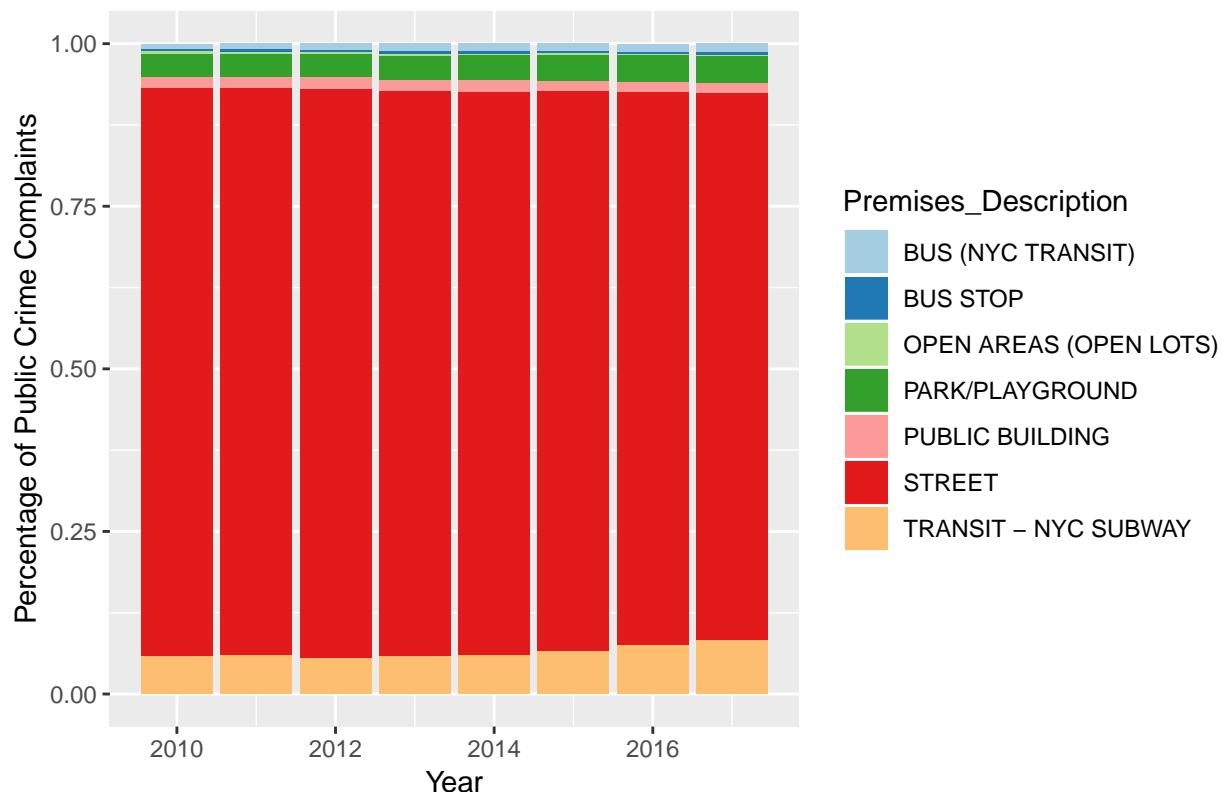
```

By using the functions and formula above, the percentage change in 2017 from 2010 can be shown, which helps to identify if the offense have decreased or increased. The percentage change in public crime complaints in 2017 from 2010: arson have decreased by 46.4%, assault & related offenses have decreased by 3.2%, possession of dangerous weapons have decreased by 46.7%, felony assault have decreased by 10.1%, harrassment have increased by 3.8%, kidnapping have decreased by 31.3%, rape is unchanged, robbery have decreased by 38.6%, and sex crimes have increased by 25.6%.

Total Public Crime Complaints Annually in Different Premises

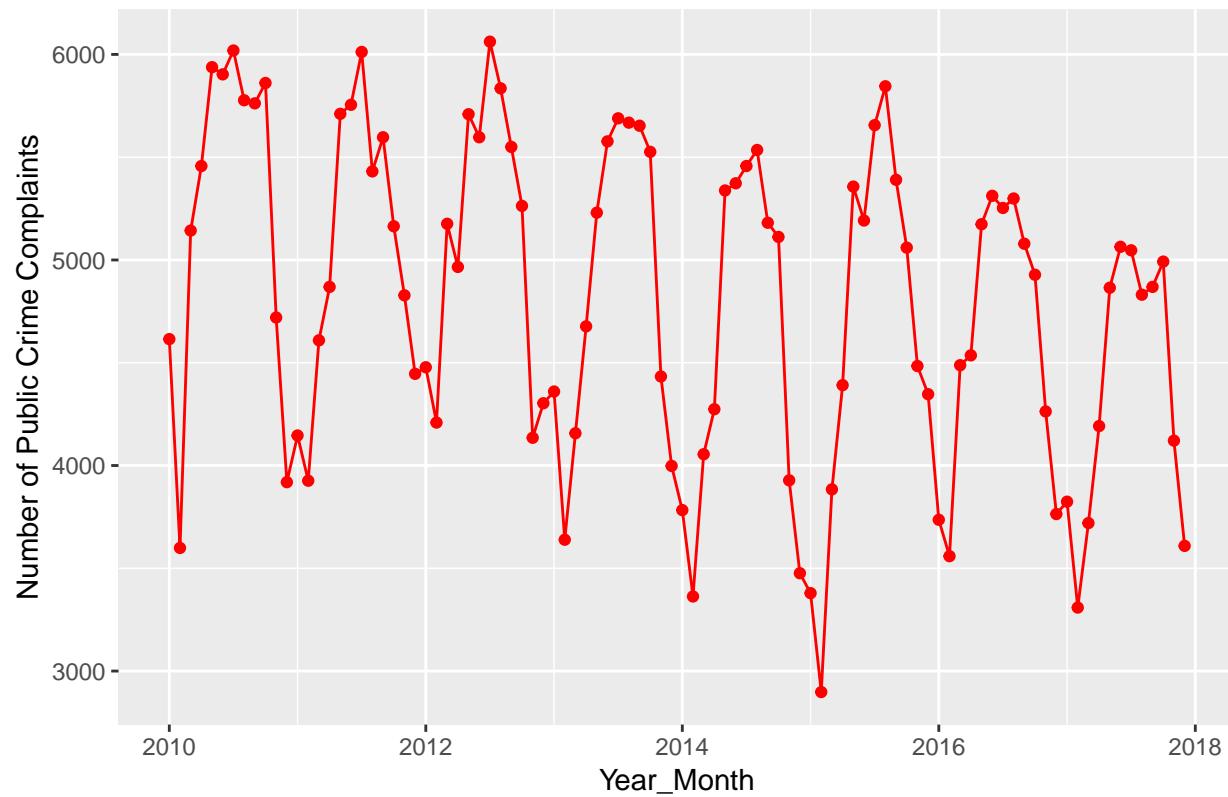


Percentage of The Total Public Crime Complaints Annually in Different Premises

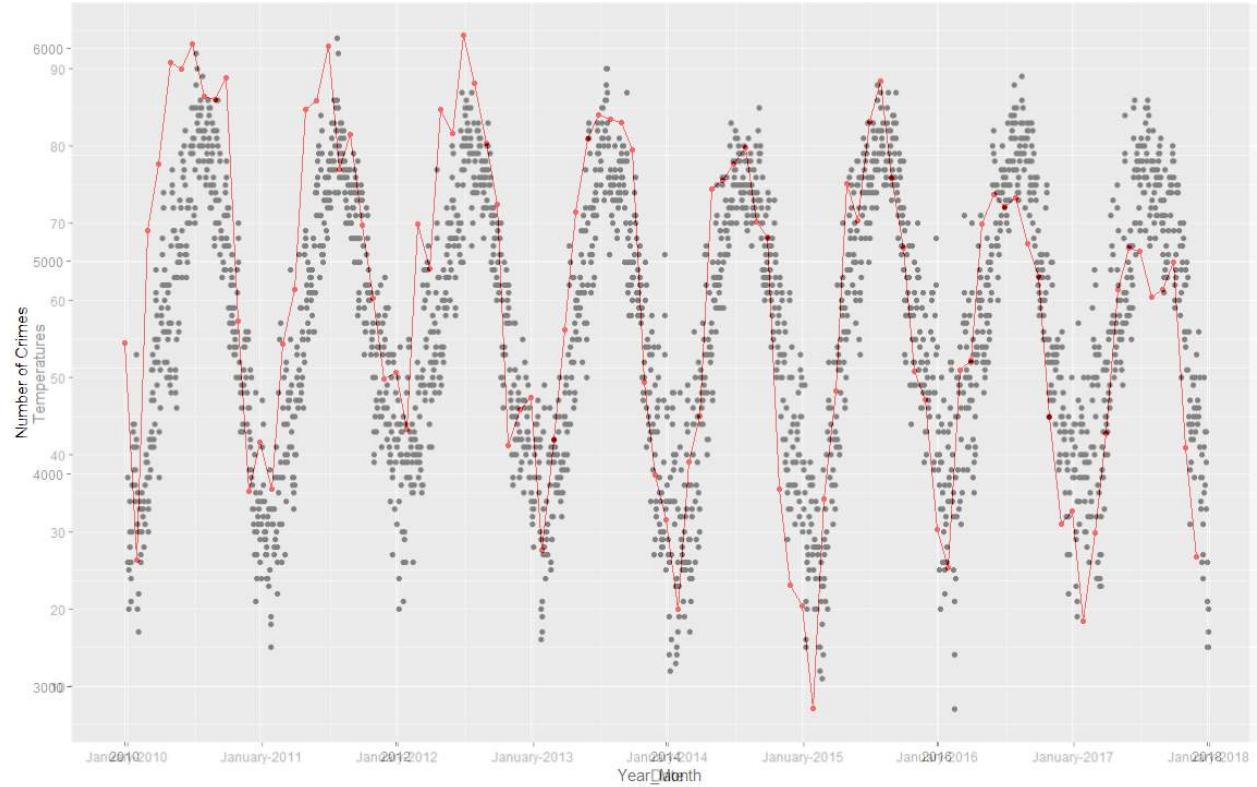
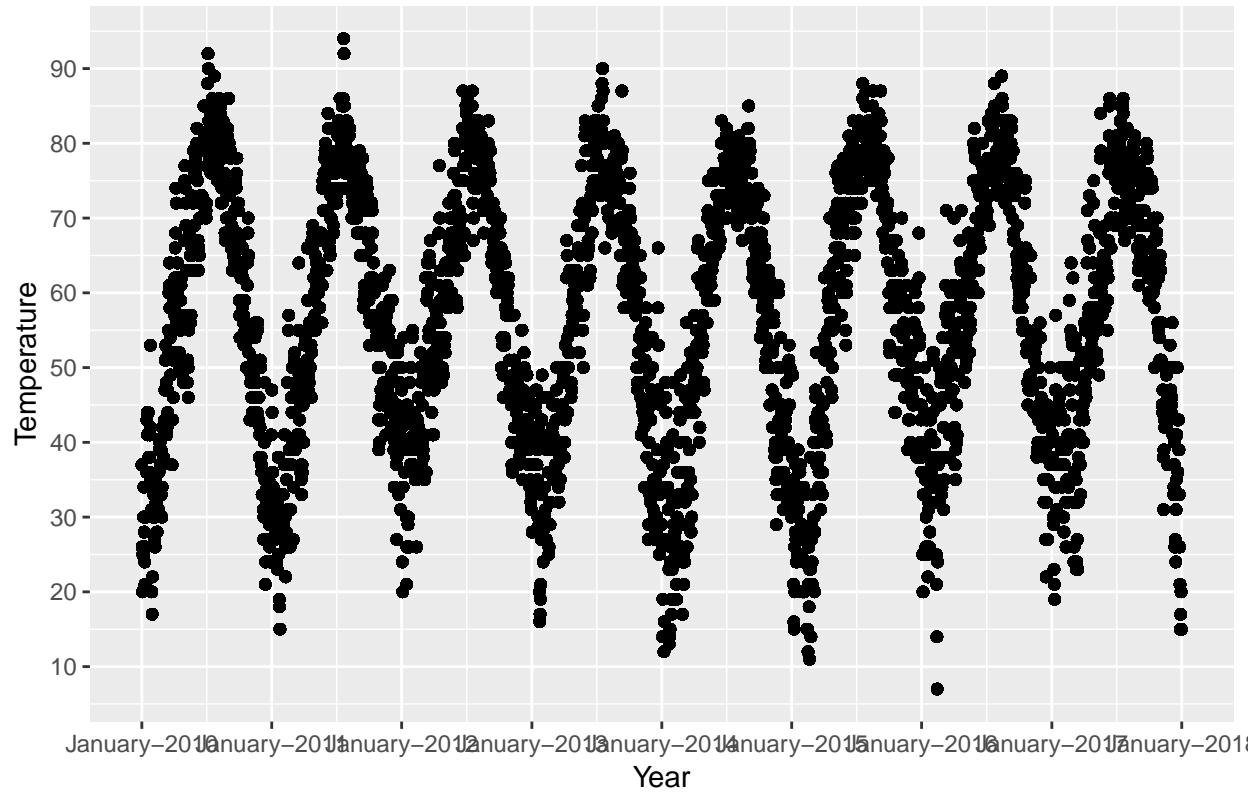


Majority of the total public crime complaints are from the street. Over the years, the percentage of street complaints have decreased while the percentage of transit - NYC subway have increased.

Total Public Crime Complaints Per Month Between 2010 to 2017



Daily Average Temperature Between 2010 to 2017



By compare the temperature with the total public crime The figure shows a strong correlation between the

daily average temperature and total crime complaint over the year of 2010 to 2017. As the temperature decrease, the public crime decreases as well and as the temperature increase, the public crime complaints increases as well.

```
#Total public crime complaints during shift 1  
nrow(Shift1)
```

```
## [1] 118733
```

```
#Total public crime complaints during shift 2  
nrow(Shift2)
```

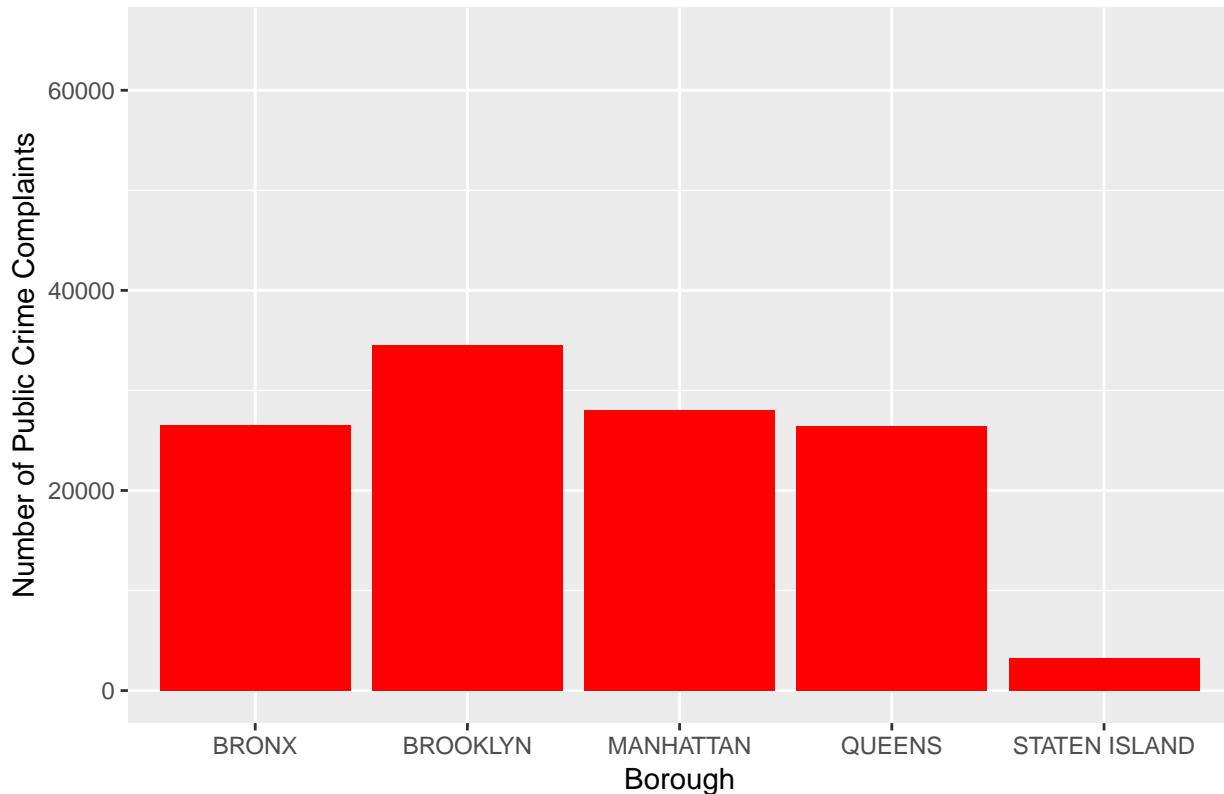
```
## [1] 145766
```

```
#Total public crime complaints during shift 3  
nrow(Shift3)
```

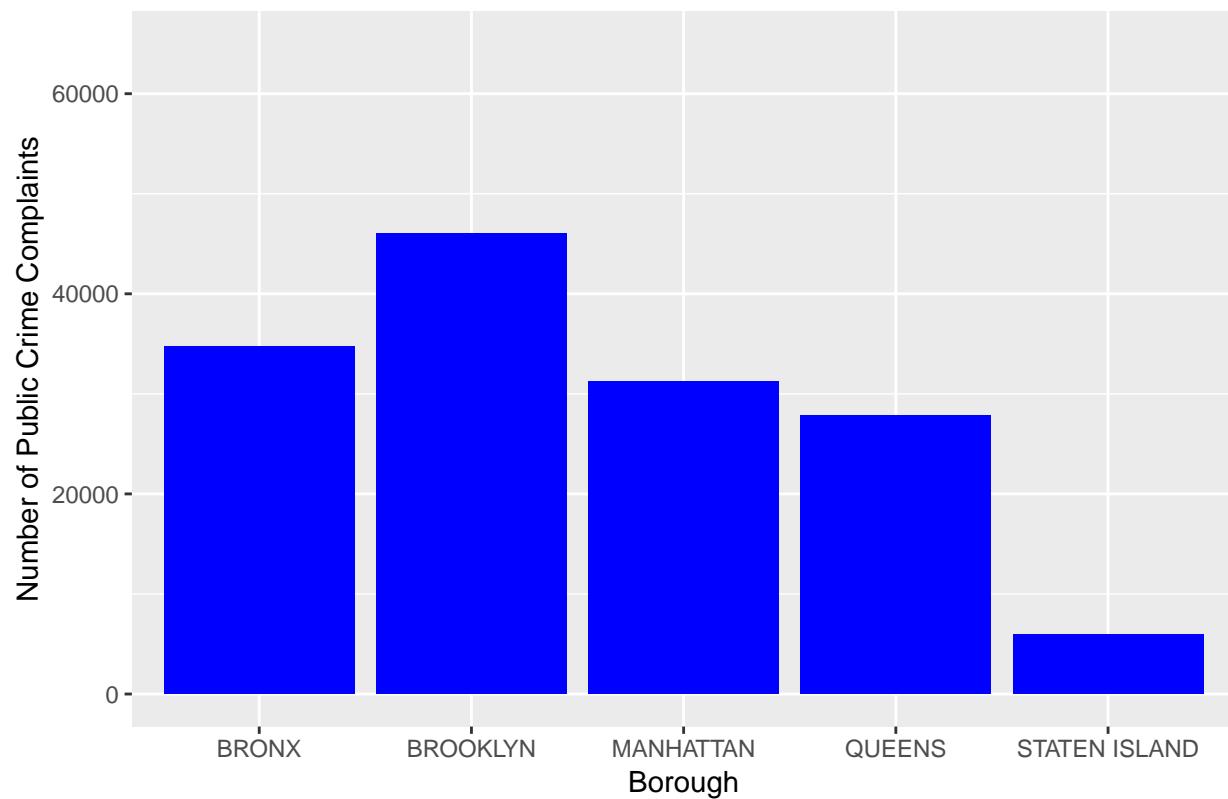
```
## [1] 197408
```

From the nrow function, Shift 3 (hour 16:00 - 24:00) was found to have the most public crime complaints. Shift 1 have 118733 complaints, shift 2 have 145766 complaints, and shift 3 have 197408 complaints.

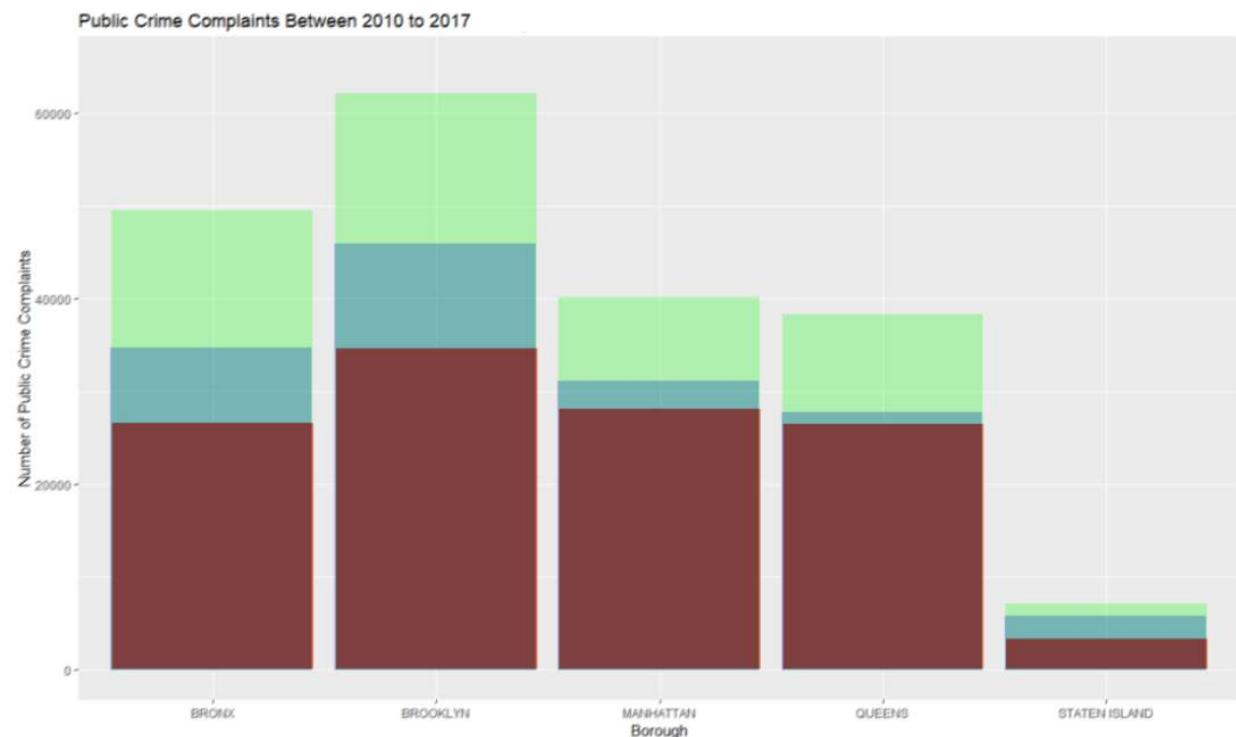
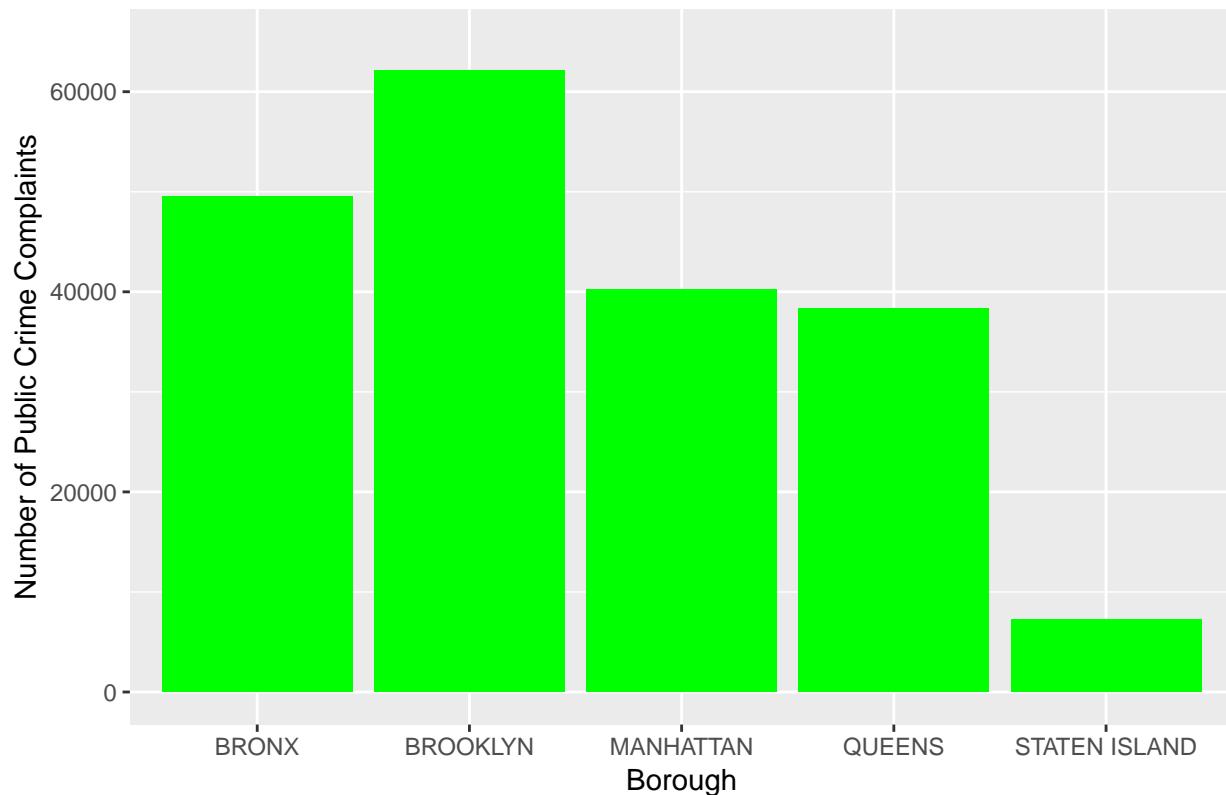
Public Crime Complaints Between 2010 to 2017 During Shift 1



Public Crime Complaints Between 2010 to 2017 During Shift 2



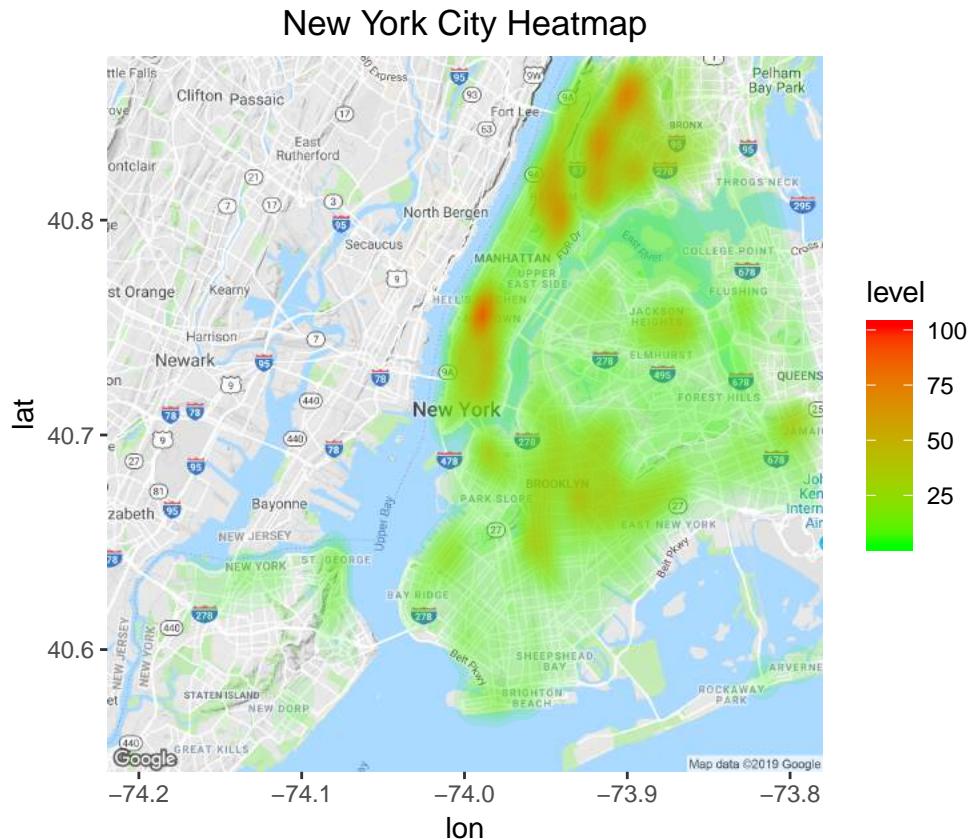
Public Crime Complaints Between 2010 to 2017 During Shift 3



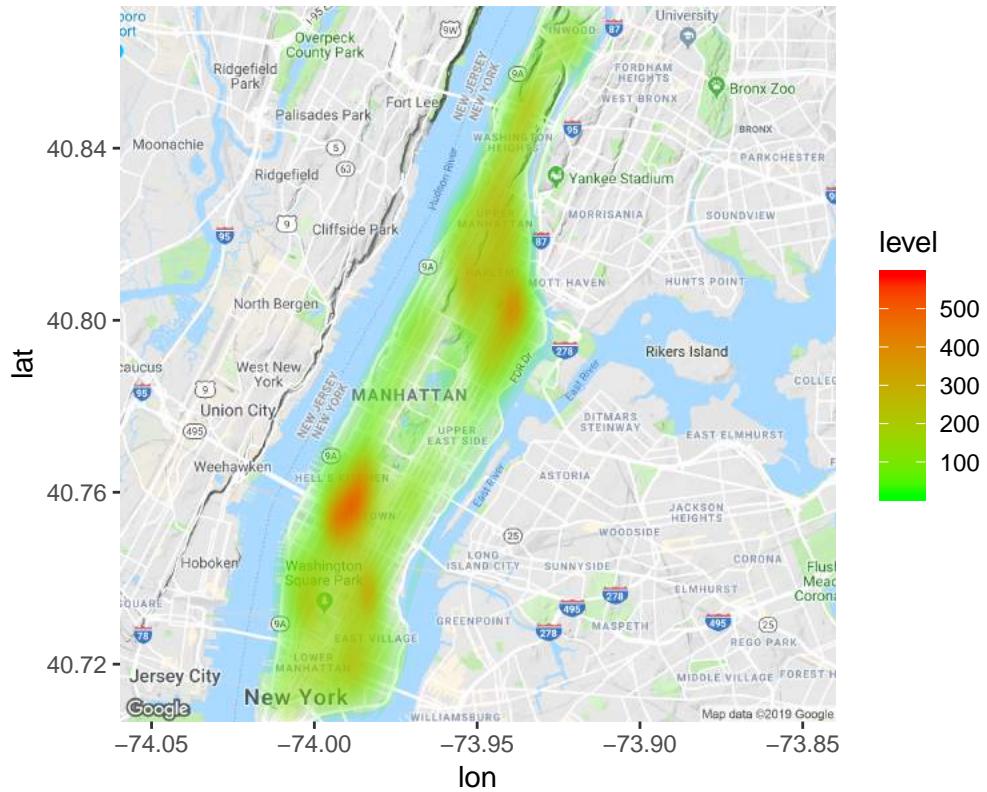
The image above shown that the number of public crime complaints in all five boroughs are in order for every shifts. Brooklyn always have the most complaints, Bronx always come second, Manhattan always third, Queens fourth, and Staten Island always last.

Load the library for plotting heatmap

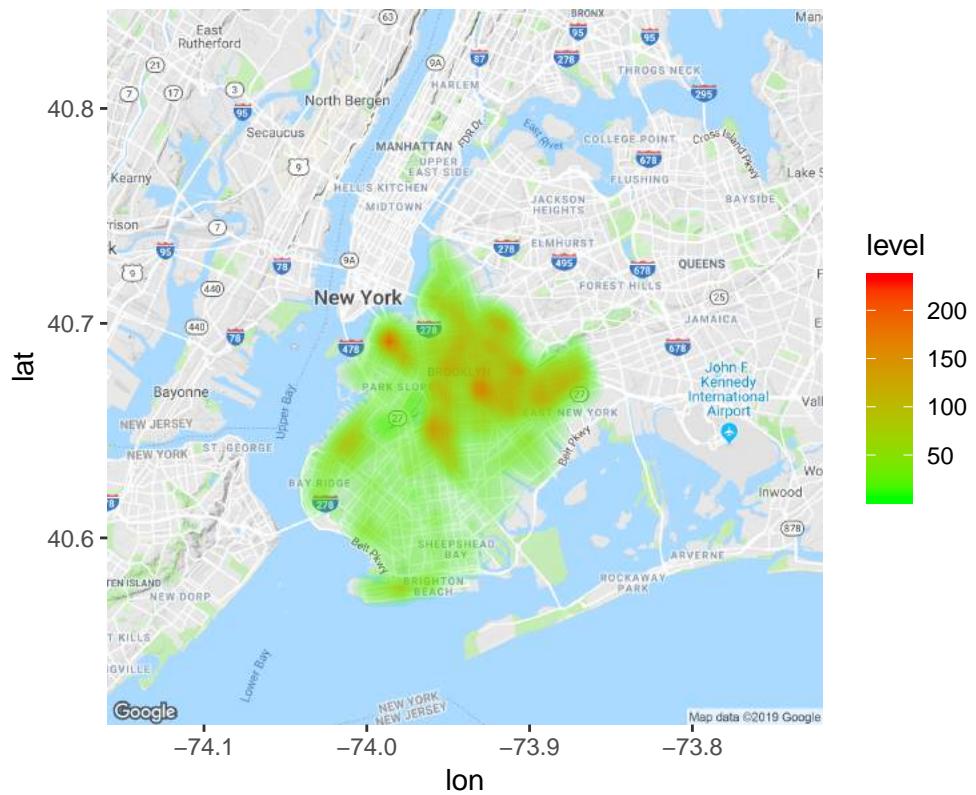
```
library(ggplot2)
library(ggthemes)
library(viridis)
library(ggmap)
library(scales)
library(grid)
library(dplyr)
library(gridExtra)
library(tigris)
library(leaflet)
library(sp)
library(maptools)
library(broom)
library(httr)
library(rgdal)
```



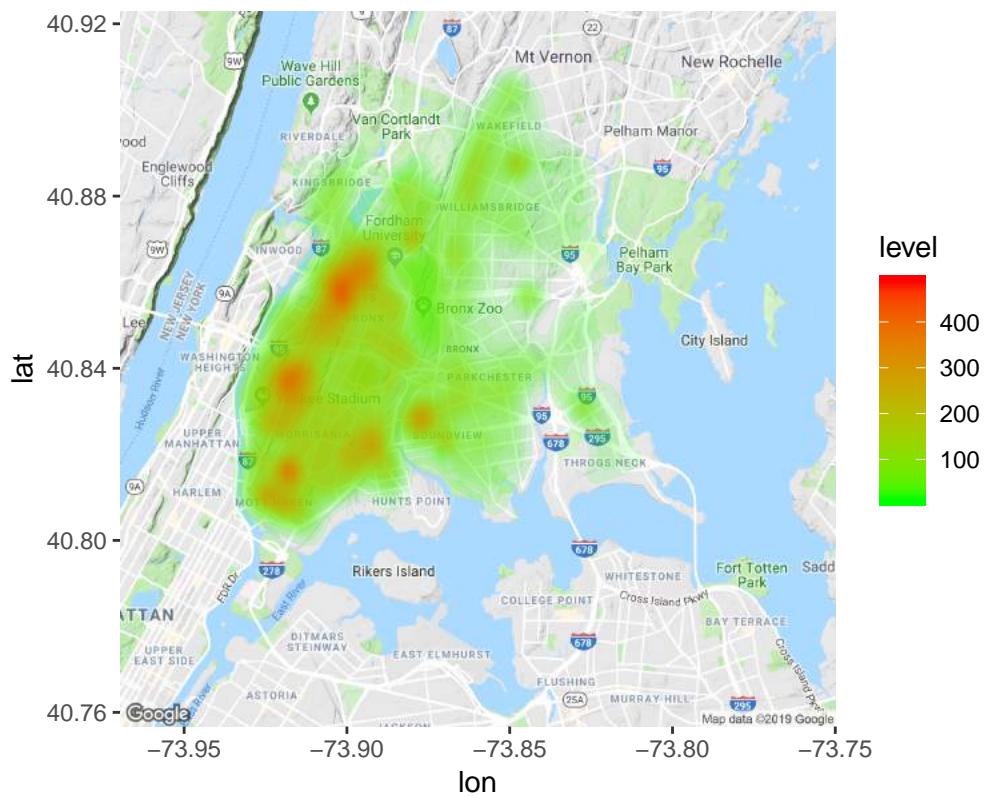
Manhattan Heatmap



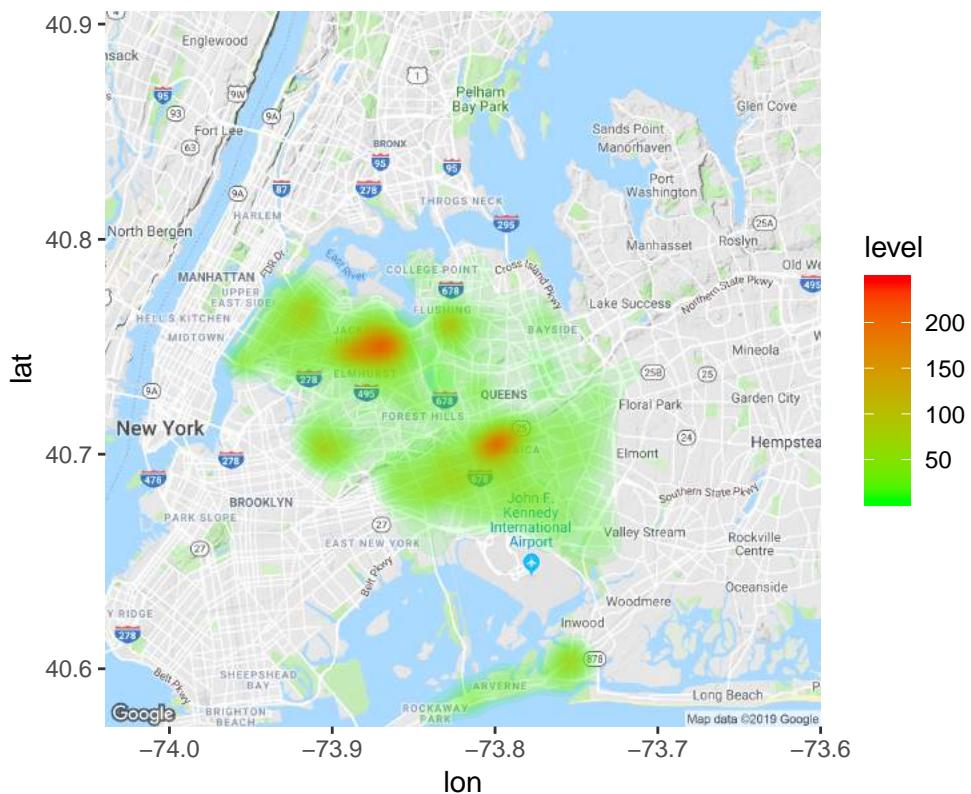
Brooklyn Heatmap



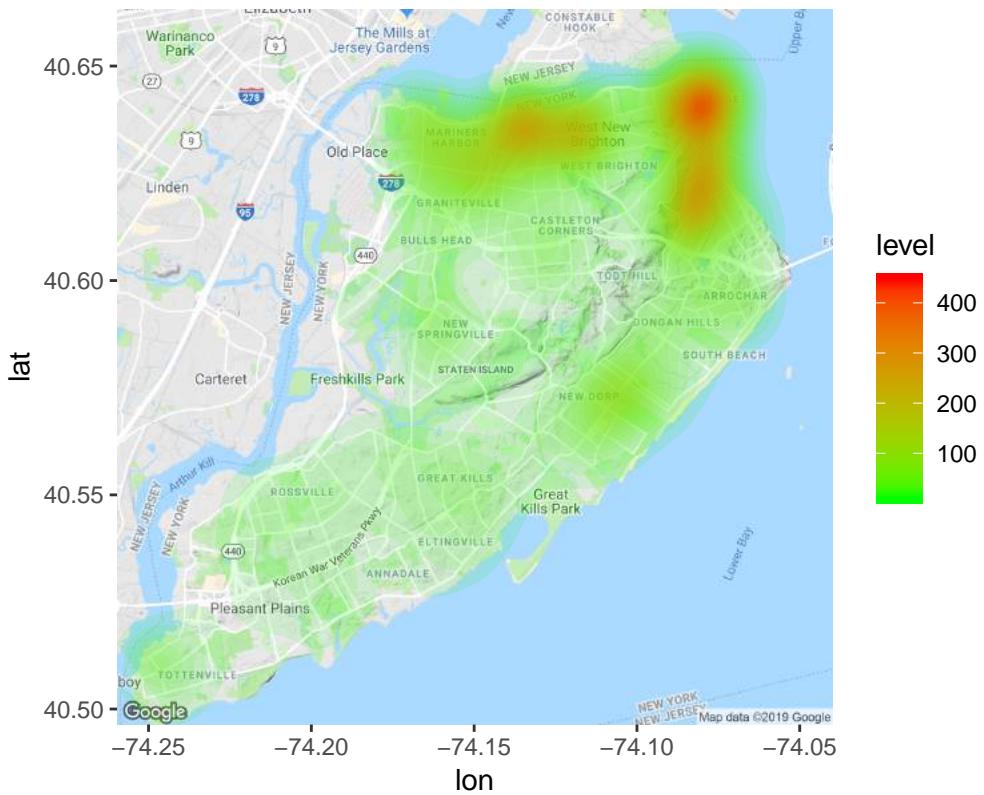
Bronx Heatmap



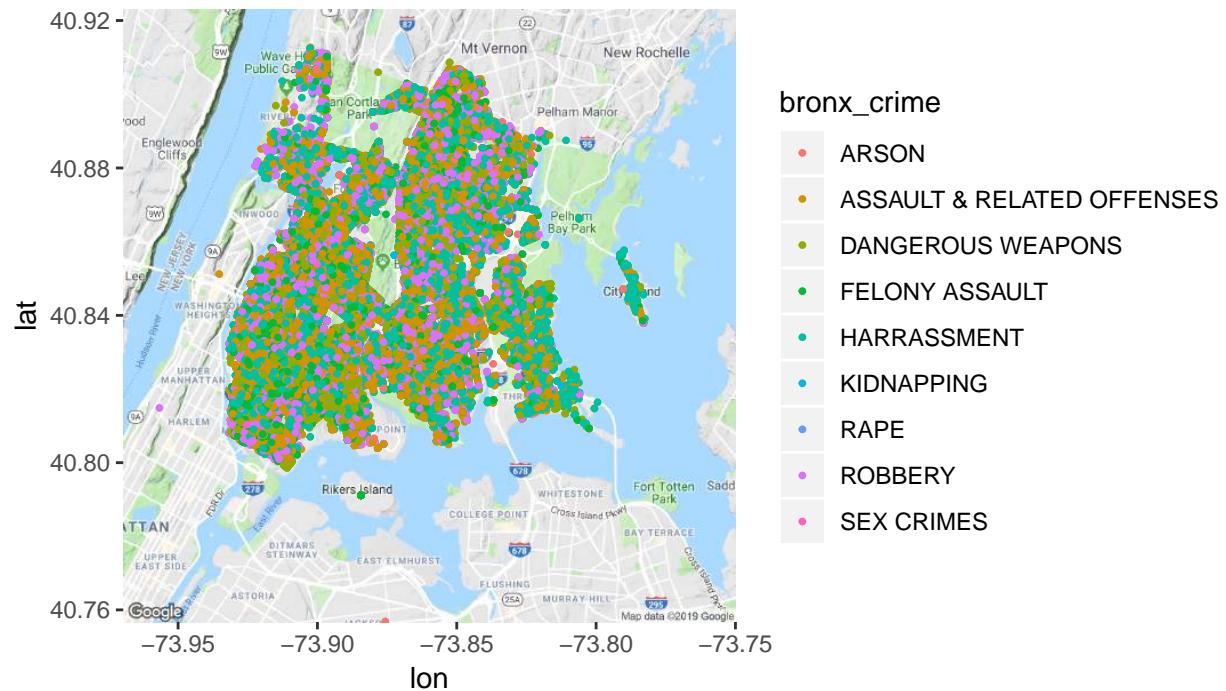
Queens Heatmap



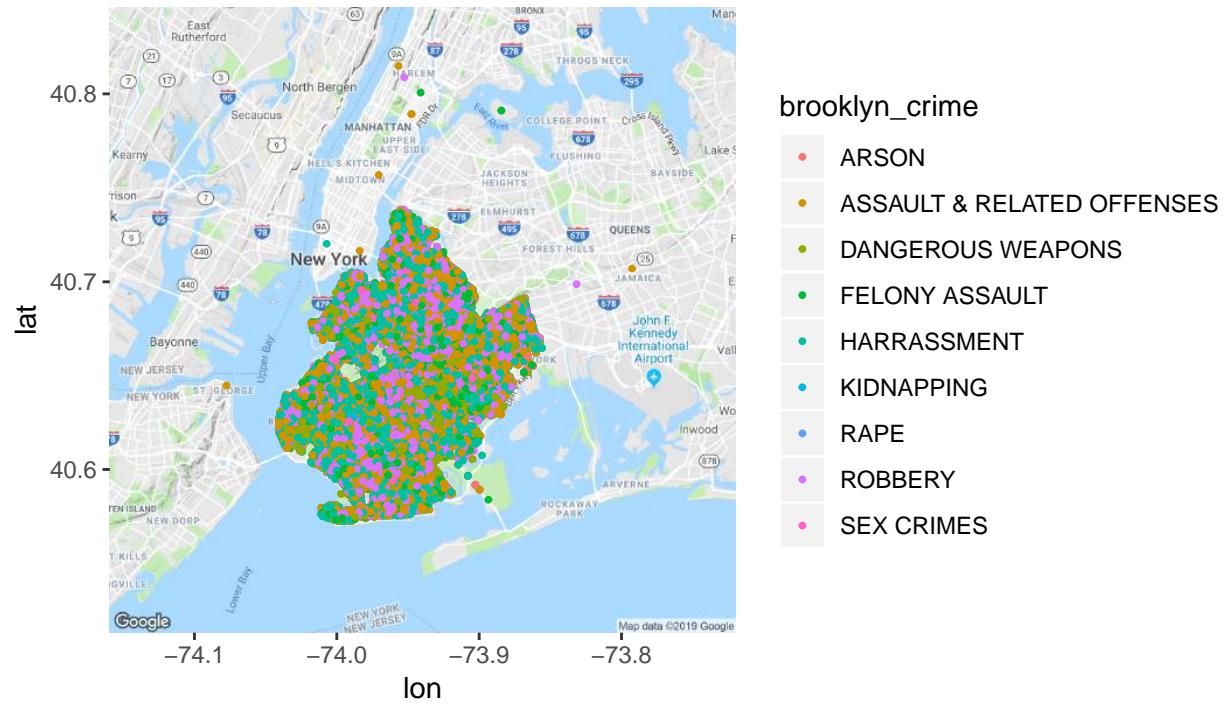
Staten Island Heatmap



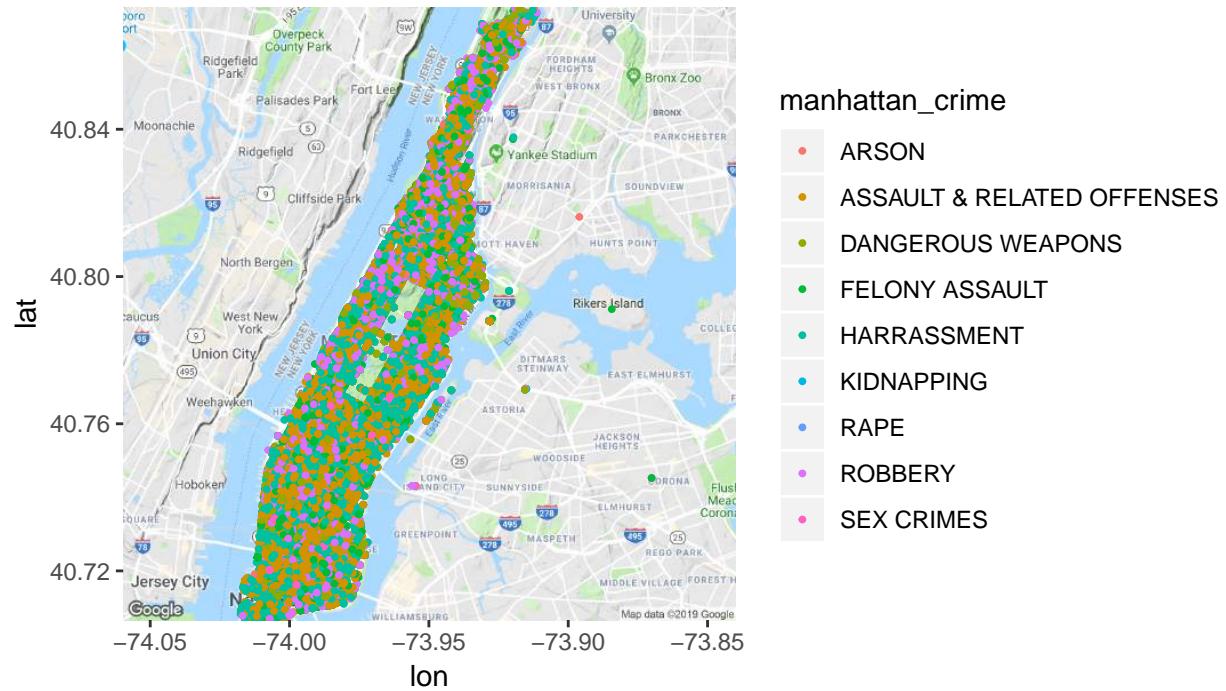
Bronx Offenses Map



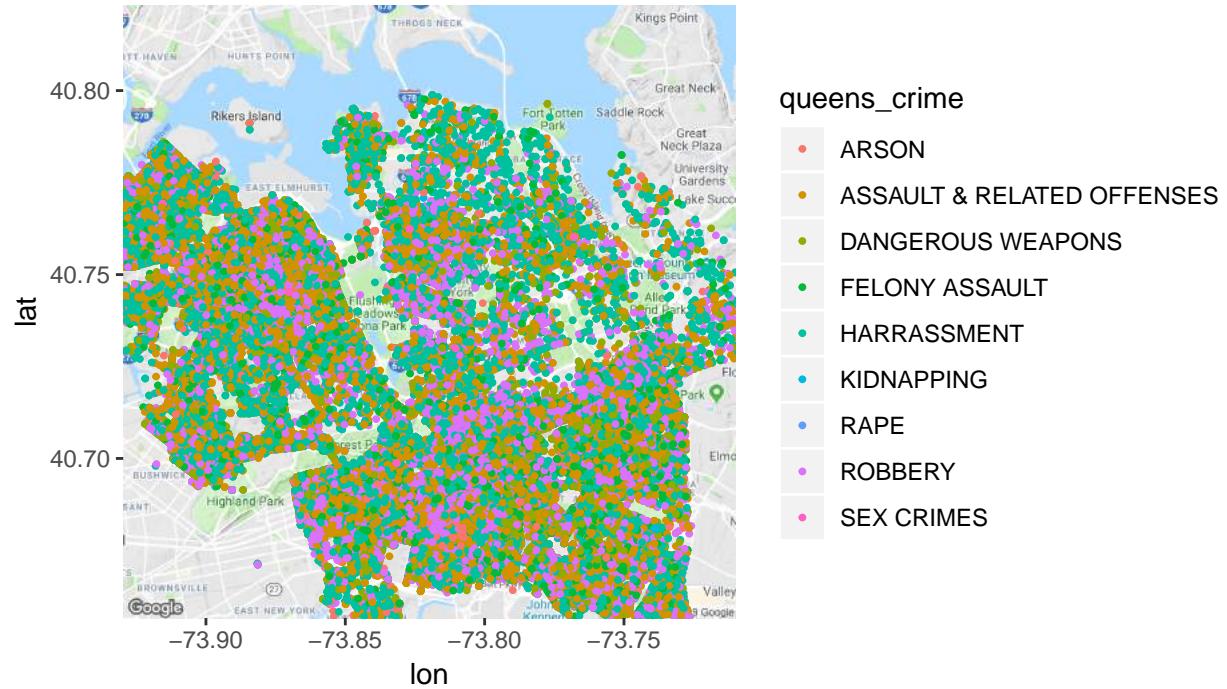
Brooklyn Offenses Map



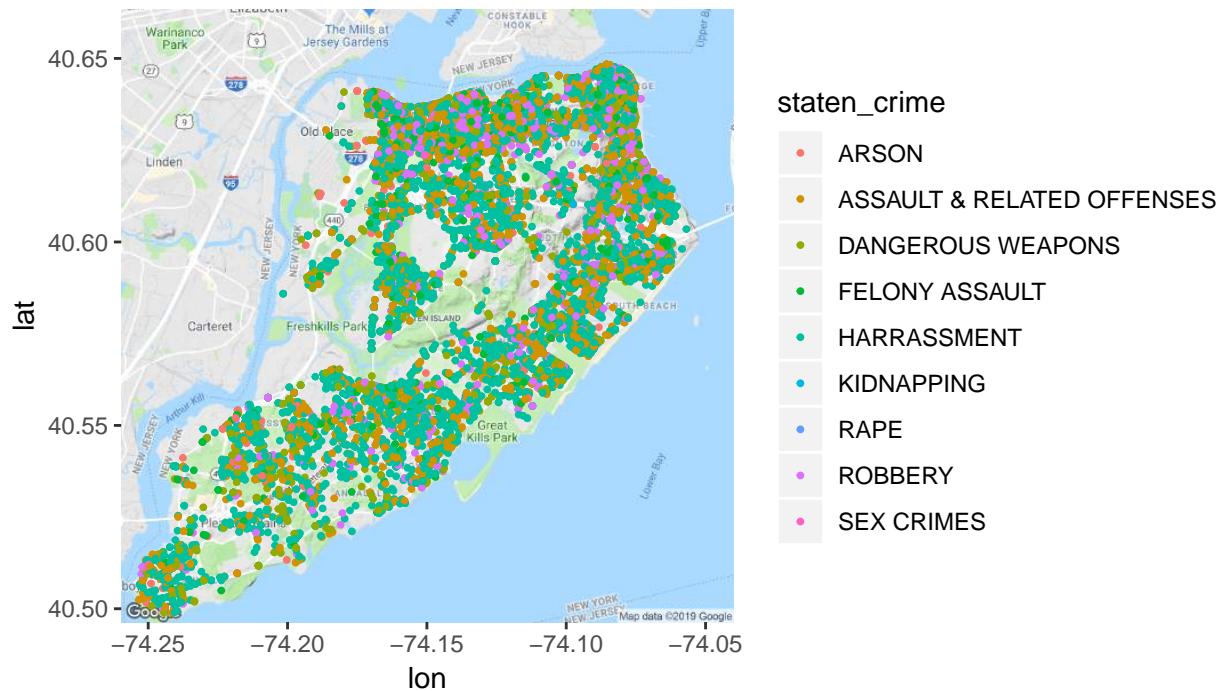
Manhattan Offenses Map



Queens Offenses Map



Staten Island Offenses Map



Conclusion

The public crime complaints in New York City have decreased 16.38% at the end of 2017. Brooklyn is the borough in New York City that have the most public crime activity that violate pedestrians. There is a strong correlation between public crime activity and weather temperature. As the temperature decreased, the number of crime activities decreased as well. Same as the temperature increased, the number of crime activities increased. The most crime activity happens at the hour 16:01 - 24:00 (shift 3). The hotspots for each borough are shown on the heatmaps above.