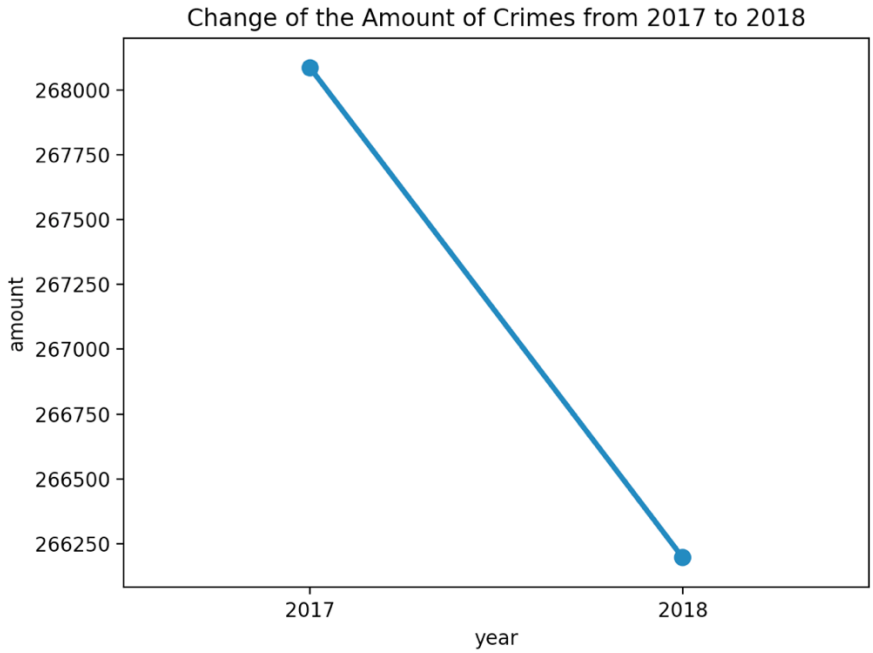
Summary Statistics for Problem 1

Xuan Bu

Problem 1: Data Acquisition and Analysis

Table 1



First, according to table 1, we can find that compared to 2017, the total number of crimes dropped 0.7 percent in 2018.

Table 2

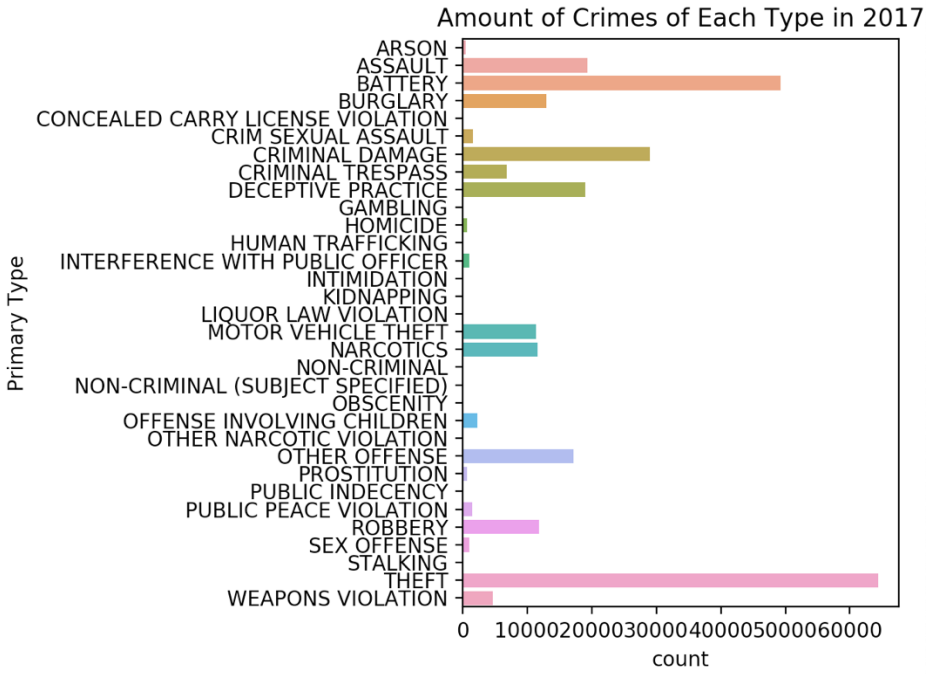
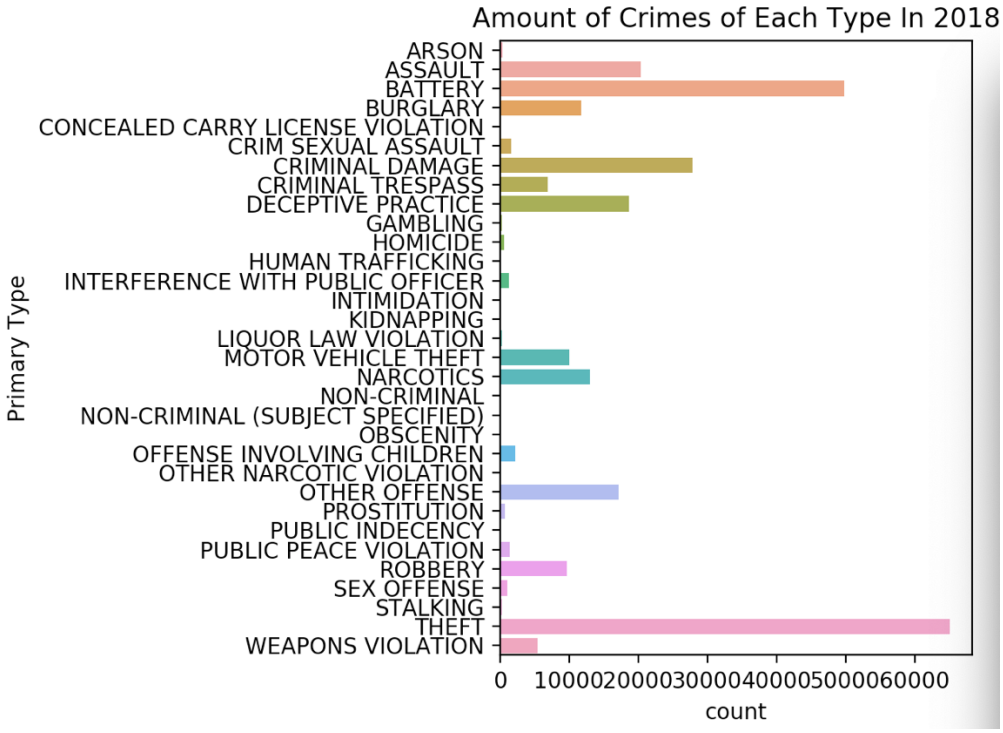


Table 3



Second, based on table 2 and table 3, we can find the distributions of the types of crimes between 2017 and 2018 are almost the same, in both years, the top 5 crimes are Theft, Battery, Criminal Damage, Assault and Deceptive Practice.

Table 4

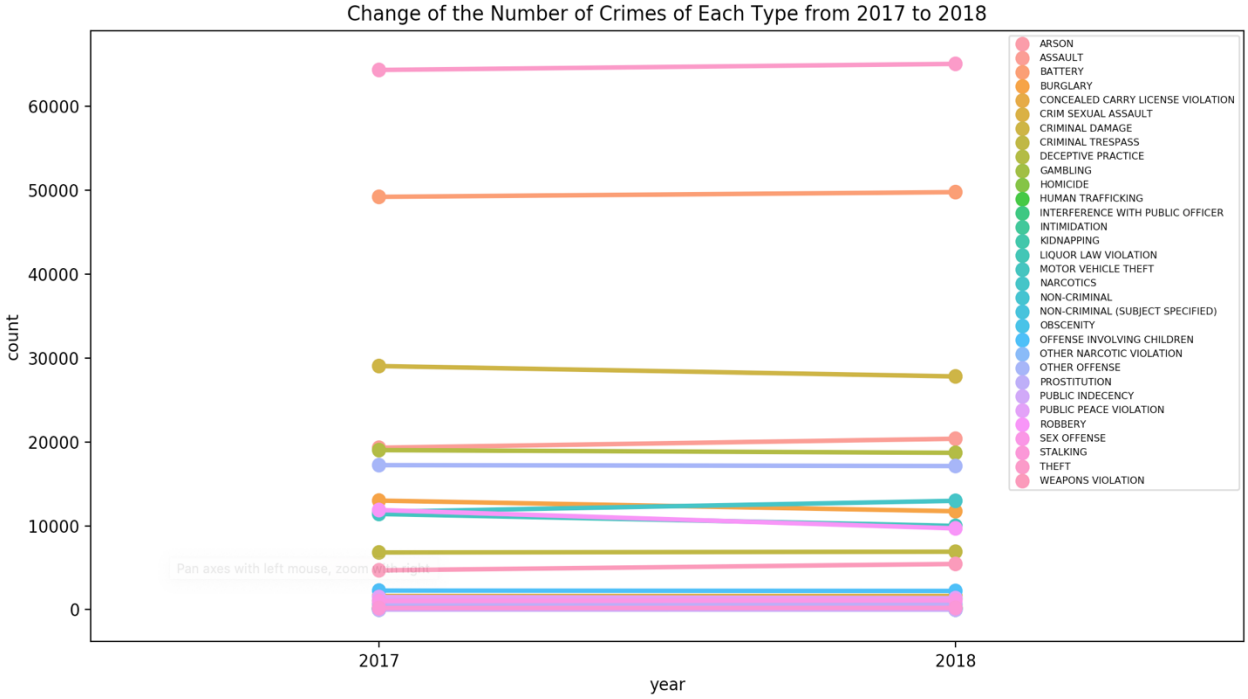
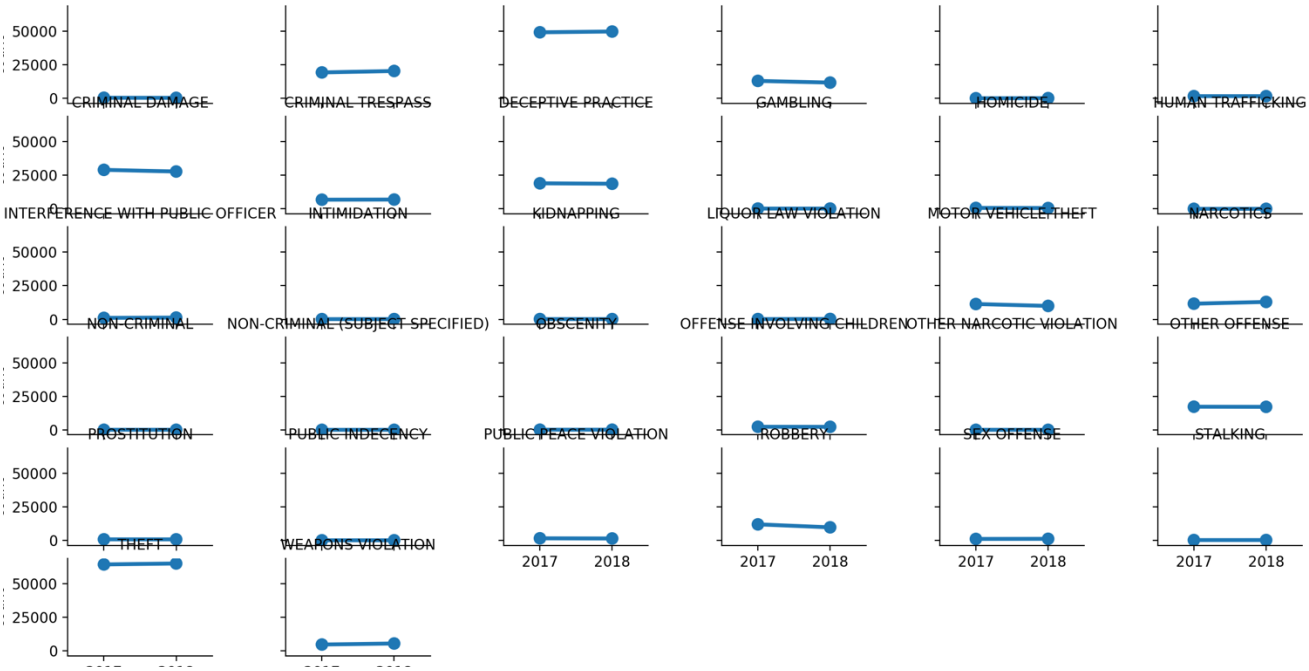


Table 5



Third, based on table 4 and table 5, we can find that the numbers of crimes of each type are roughly the same from 2017 to 2018.

Table 6

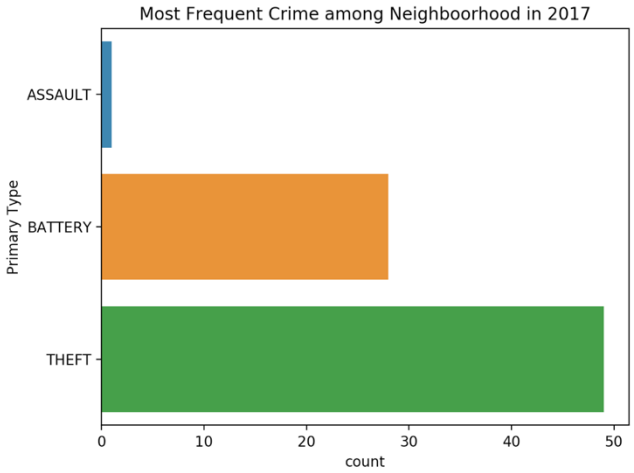


Table 7

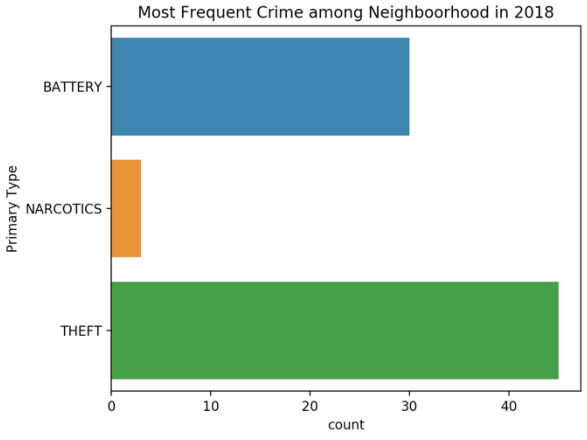
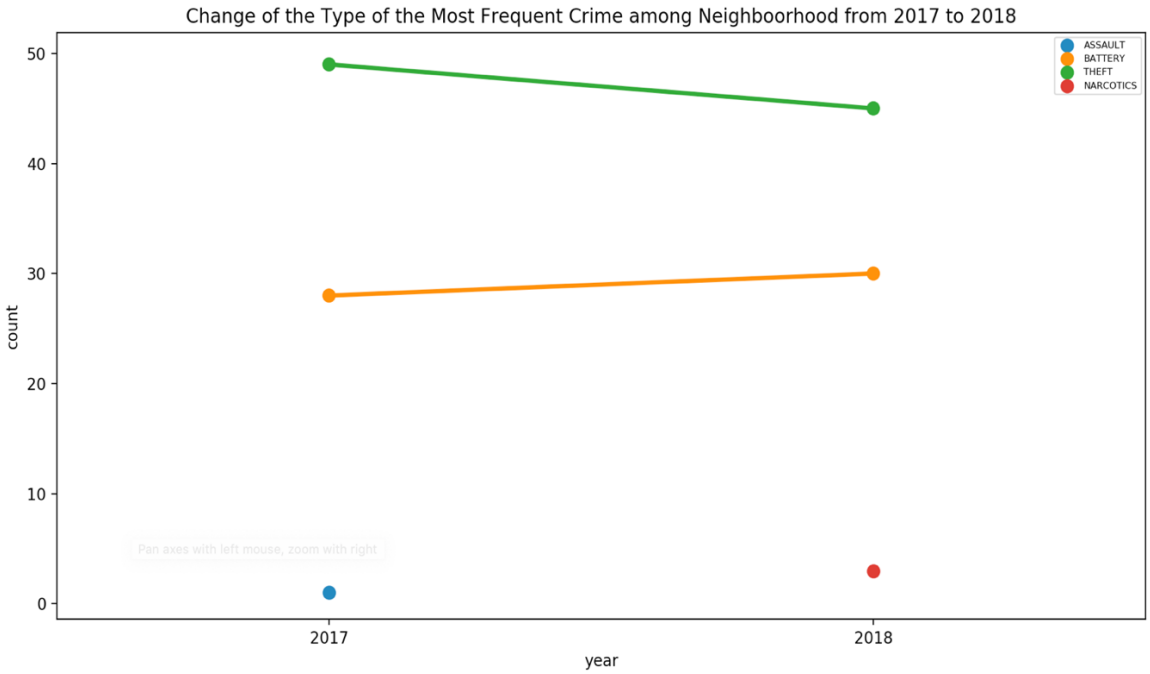
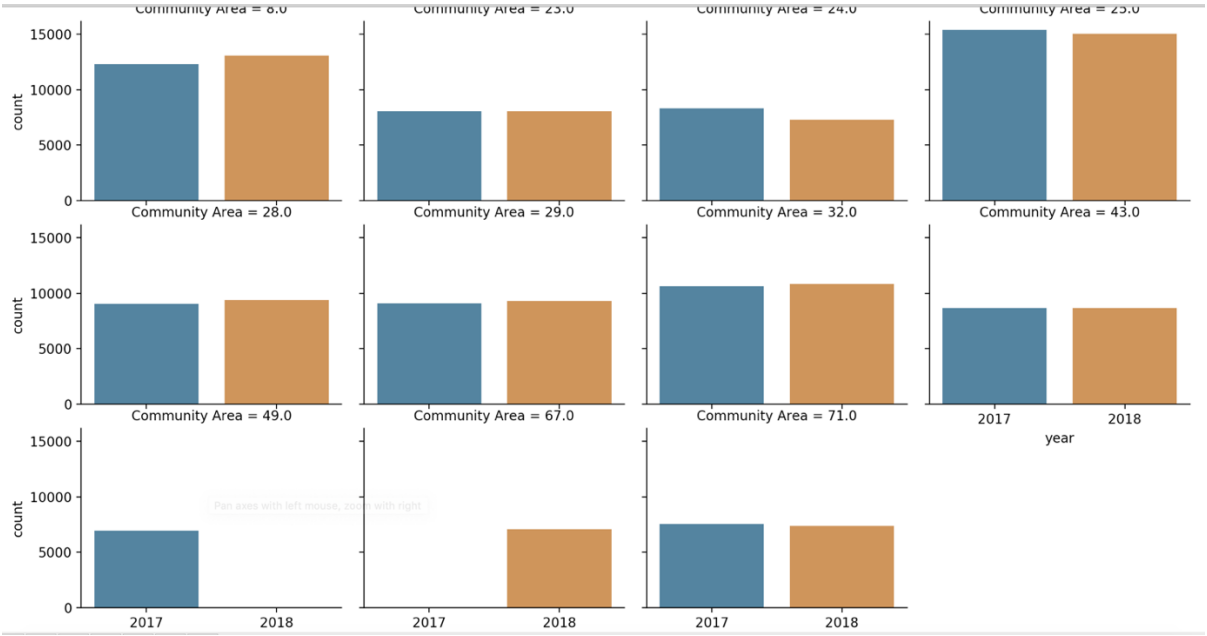


Table 8



According to table 6 to table 8, we can find that among the 78 communities, in 2017, Theft is the top 1 crime in 49 communities, Battery is the top 1 crime in 28 communities, and assault the top 1 crime in 1 community; in 2018, Theft is the top 1 crime in 45 communities, Battery is the top 1 crime in 30 communities, and Narcotics the top 1 crime in 3 community.

Table 9



According to table 9, we can find the top 10 neighborhoods with most frequent crimes in 2017 and 2018. In both years, the distributions of the top 10 neighborhoods with most frequent crimes are roughly the same except the No.49 neighborhood that is in the top 10 in 2017 but not in the top 10 in 2018, and No.67 neighborhood that is not in the top 10 in 2017 but in the top 10 in 2018.

Problem 2: Data Augmentation and APIs

To describe the features of the blocks with reports of certain type of crime, in problem 2, I select race, income and education attainment from ACS in 2017 and 2018, then merge them with the crime reports in 2017 and 2018 respectively. In the process of merging, I use the tract number and community area number from the file "2010 Tract to Community Area Equivalency File - Sheet1.csv".

1. What types of blocks have reports of "Battery"?

To describe the features of the blocks with reports of "Battery"

Table 10

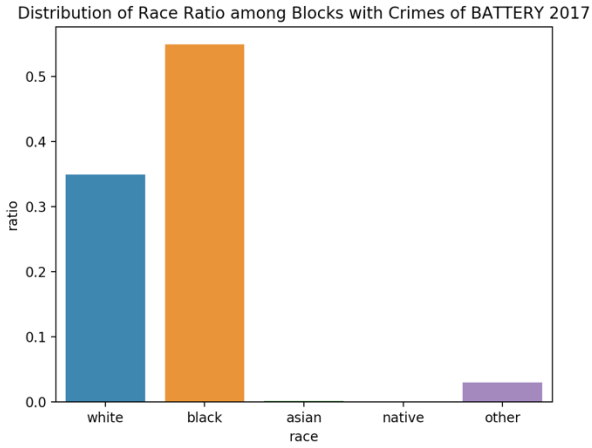


Table 11

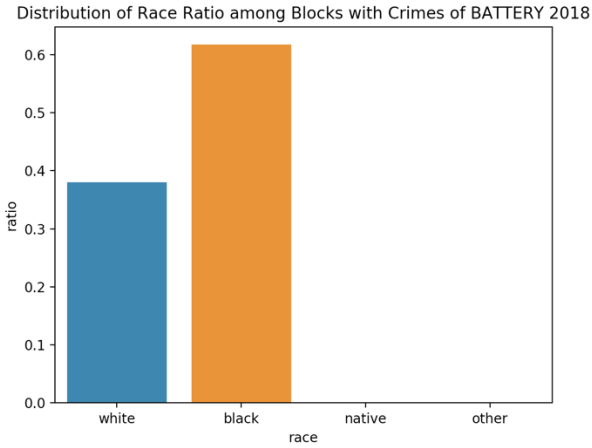


Table 12

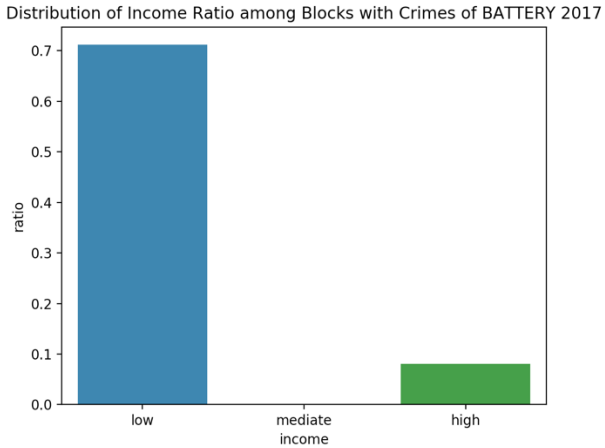
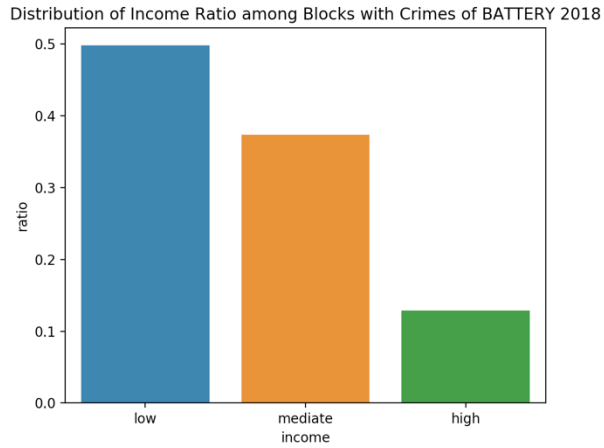


Table 13



From table 10 and table 11, we can find that the blocks that have reports of “Battery” are the ones that dominated by black people, in both years, there are over 55% residents are black.
From table 12 and table 13, we can find that the blocks with reports of “Battery” are made of low-income people.
In addition, the average ratio of educated people in this dataset in 2017 is 98.41%, in the blocks with reports of “Battery”, there are 39.05% blocks with ratio of educated people below the average. And the average ratio of people who finished high school in 2018 is 82.94%, there are 51.99% blocks with ratio of people who finished high school below the average.

2. What types of blocks get “Homicide”?

Table 14

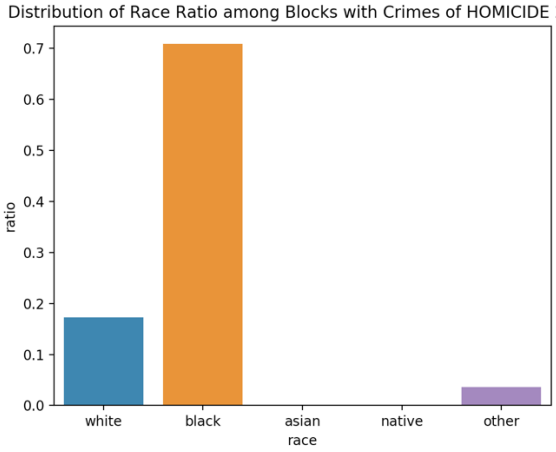


Table 15

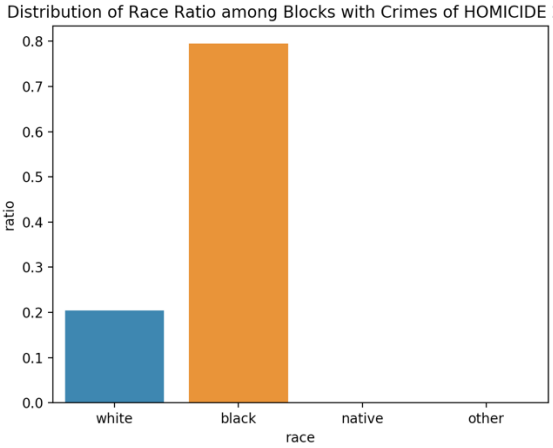


Table 16

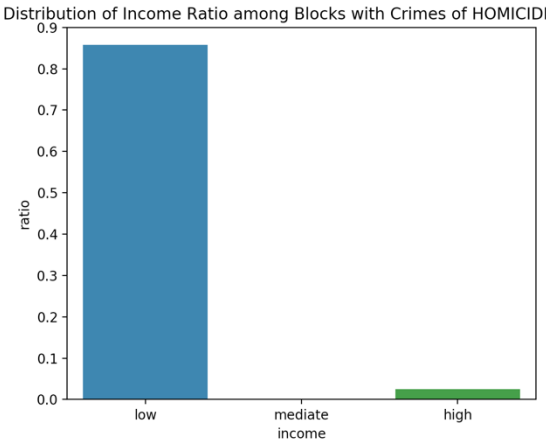
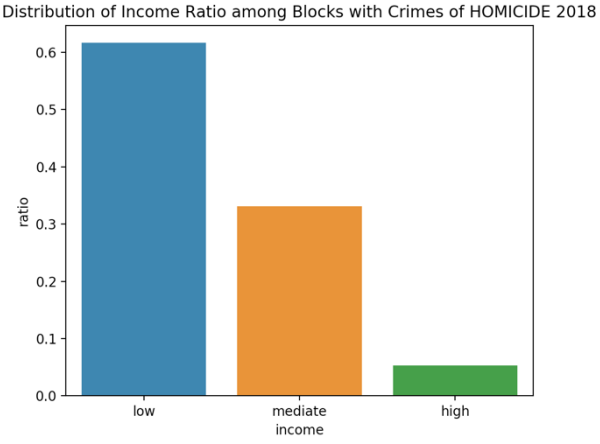


Table 17



From table 14 and table 15, we can find that the blocks that have reports of “Homicide” are dominated by black people. Compared to the blocks with reports

of “Battery”, there are more blocks with reports of “Homicide” that are mainly made by black people. In both years, there are over 65% residents are black. From table 12 and table 13, we can find that the blocks with reports of “Homicide” are made of low-income people.
In addition, the average ratio of educated people in this dataset in 2017 is 98.41%, in the blocks with reports of “Homicide”, there are 42.49% blocks with ratio of educated people below the average. And the average ratio of people who finished high school in 2018 is 82.94%, there are 62.94% blocks with ratio of people who finished high school below the average.

3. Does that change over time in the data you collected?

According to the tables 10-17, we can find that in the blocks with reports of “Homicide” and “Battery”, neither race distribution nor income distribution nor education status has changed significantly from 2017 to 2018.

4. What is the difference in blocks that get “Deceptive Practice” vs “Sex Offense”?

Table 18

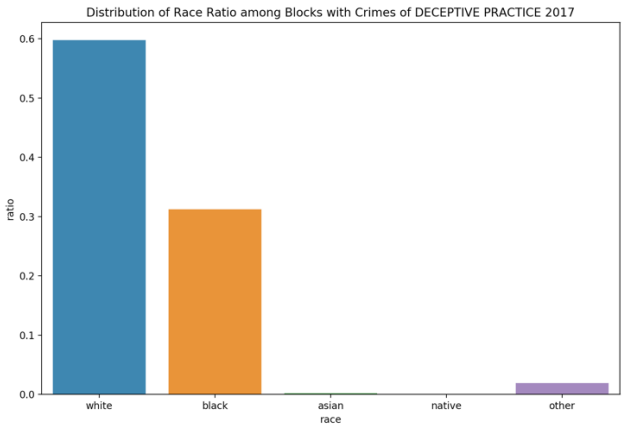


Table 19

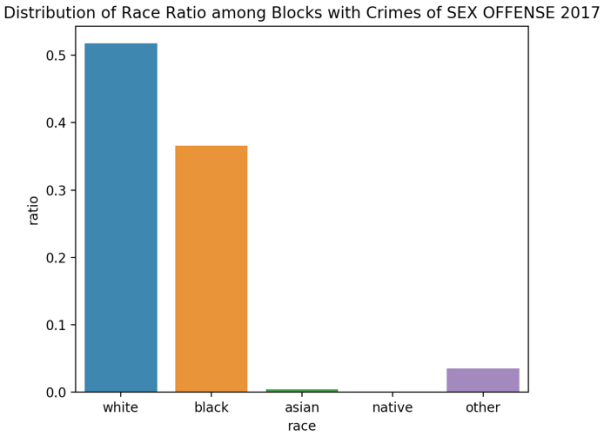


Table 20

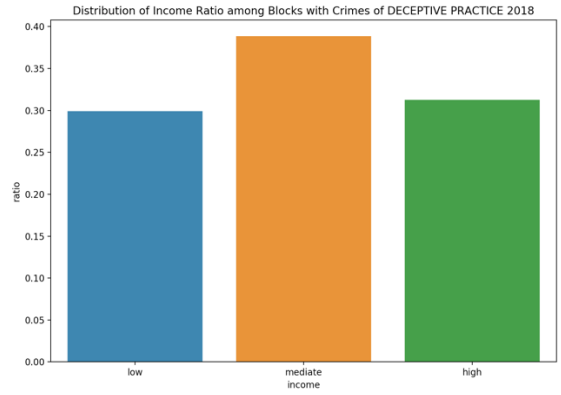
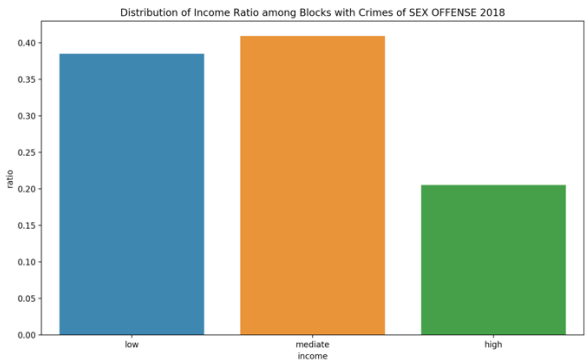


Table 21



From table 18 and 19, we can find that, although the blocks with both crime types are dominated by white people, the ratio of white people in the blocks that get "Deceptive Practice" are higher than the blocks that get "Sex Offense".
From Table 20 and 21, we can find that, although the blocks with both crime types are mainly made by mediate-income people, the ratio of high-income people in the "Deceptive Practice" are higher, which might explain partly the occurrences of "Deceptive Practice".

Problem 3: Analysis and Communication

1. Describe how crime has changed in Chicago from 2017 to 2018?

Based on table 1, we can find that from 2017 to 2018, the total number of crimes dropped 0.7 percent, that is, there are 1848 crimes less from 2017 to 2018.
In addition, from table 6 and table 7, we can find the both the ratios of battery and theft have decreased from 2017 to 2018. However, the issue of narcotics has become more serious.
In general, there is not a significant change from 2017 to 2018. From the table 5 and table 9, both the numbers of crimes of each type and the neighborhoods with most crimes are roughly the same.

2. A. Are these statistics correct?

No, it's incorrect. Based on my calculation, the change of the four types crimes (Robbery, Battery, Burglary, Motor Vehicle Theft) for the week (07/20/2018-07/26/2018), compared to the same week in 2017:

- Robberies decrease by 16.97%
- Batteries increase by 3.94%
- Burglaries decrease by 11.39%
- Motor Vehicle Thefts decrease by 12.73%

B. Could they be misleading or would you agree with the conclusions he's drawing? Why or why not?

These statistics could be misleading and I would not agree with the conclusion. First, based on my calculation, these results are incorrect for some reason, especially for battery, burglary, and motor vehicle theft, the results of these three types of crimes are far away from the correct results, which could be misleading. Second, even though these statistics are correct, they are not reliable since they depend on a short period, which cannot represent the change of a long time.

3. a. Narcotics has become a more serious issue, which should be paid more attention.
   b. Theft is still the most frequent crime whose number is far more than other types of crimes.
   c. In the wealthier blocks, deceptive practice is increasing and become a more usual crime.
   d. Lower educational attainment and poverty are two of the main reasons leading more crimes.
   e. Less diversified blocks are more likely to be the blocks with more crimes.

4. Considering the limit of dataset and statistical method, I would suggest that all of the analysis and recommendations are likely to be unreliable since they are based on the dataset from only two years with only descriptive statistics.

Problem 4

In this problem, all calculations are based on the crime report of 2018.

1. To search the 2111 S Michigan Ave in dataset, we can find its corresponding community area is 33. From the distribution of the probabilities of different crimes, the call is more likely about theft.

Primary Type	Probability
THEFT	31.877023
BATTERY	15.587918
DECEPTIVE PRACTICE	14.077670
CRIMINAL DAMAGE	10.248112
ASSAULT	6.580367
OTHER OFFENSE	5.987055
MOTOR VEHICLE THEFT	3.829558
CRIMINAL TRESPASS	2.858684
ROBBERY	2.642934
BURGLARY	1.725998
NARCOTICS	1.024811
CRIM SEXUAL ASSAULT	0.647249
WEAPONS VIOLATION	0.647249
PUBLIC PEACE VIOLATION	0.485437
SEX OFFENSE	0.485437
LIQUOR LAW VIOLATION	0.323625
OFFENSE INVOLVING CHILDREN	0.323625
HOMICIDE	0.161812
ARSON	0.107875

INTIMIDATION	0.107875
INTERFERENCE WITH PUBLIC OFFICER	0.053937
CONCEALED CARRY LICENSE VIOLATION	0.053937
OBSCENITY	0.053937
STALKING	0.053937
KIDNAPPING	0.053937

2. To search Garfield Park and Uptown in dataset, we can find Uptown corresponds to 3, and Garfield Park corresponds to 26 and 27.

Community Area	probability
3	42.518327
26	27.166882
7	30.314791

Since Garfield Park covers two community area, thus its probability is

27.17 + 30.31 = 57.48

Thus, the call is more likely from Garfield Park.

3. According to Bayes' theorem:
   Prob (Garfield) = 0.6,
   Prob (Uptown) = 0.4,
   Prob (Garfield | Battery) = $\frac{10}{26}$
   Prob (Uptown | Battery) = $\frac{16}{26}$
   Thus,
   Prob (Battery | Garfield) = $\frac{prob(Battery)*prob(Garfield \mid Battery)}{prob(Garfield)} = \frac{1}{6}$
   Prob (Battery | Uptown) = $\frac{prob(Battery)*prob(Uptown \mid Battery)}{prob(Uptown)} = \frac{2}{5}$
   $\frac{2}{5} - \frac{1}{6}$ = 0.233

   Therefore, the probability of that call is from Uptown is approximately 23.3% higher than the probability of that call is from Garfield Park.