HW1

April 9, 2019

1 Homework 1. Diagnostic

1.1 Problem 1: Data Acquisition and Analysis [20 pts]

```
In [3]: %load_ext autoreload
        %autoreload 2
        import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import util
```

1. Download reported crime data from the Chicago open data portal for 2017 and 2018.

```
In [4]: crime_df = util.load_crime_data(1000000, 2017, 2018)
```

```
WARNING:root:Requests made without an app_token will be subject to strict throttling limits.
```

2. Generate summary statistics for the crime reports data including but not limited to number of crimes of each type, how they change over time, and how they are different by neighborhood. Please use a combination of tables and graphs to present these summary stats.

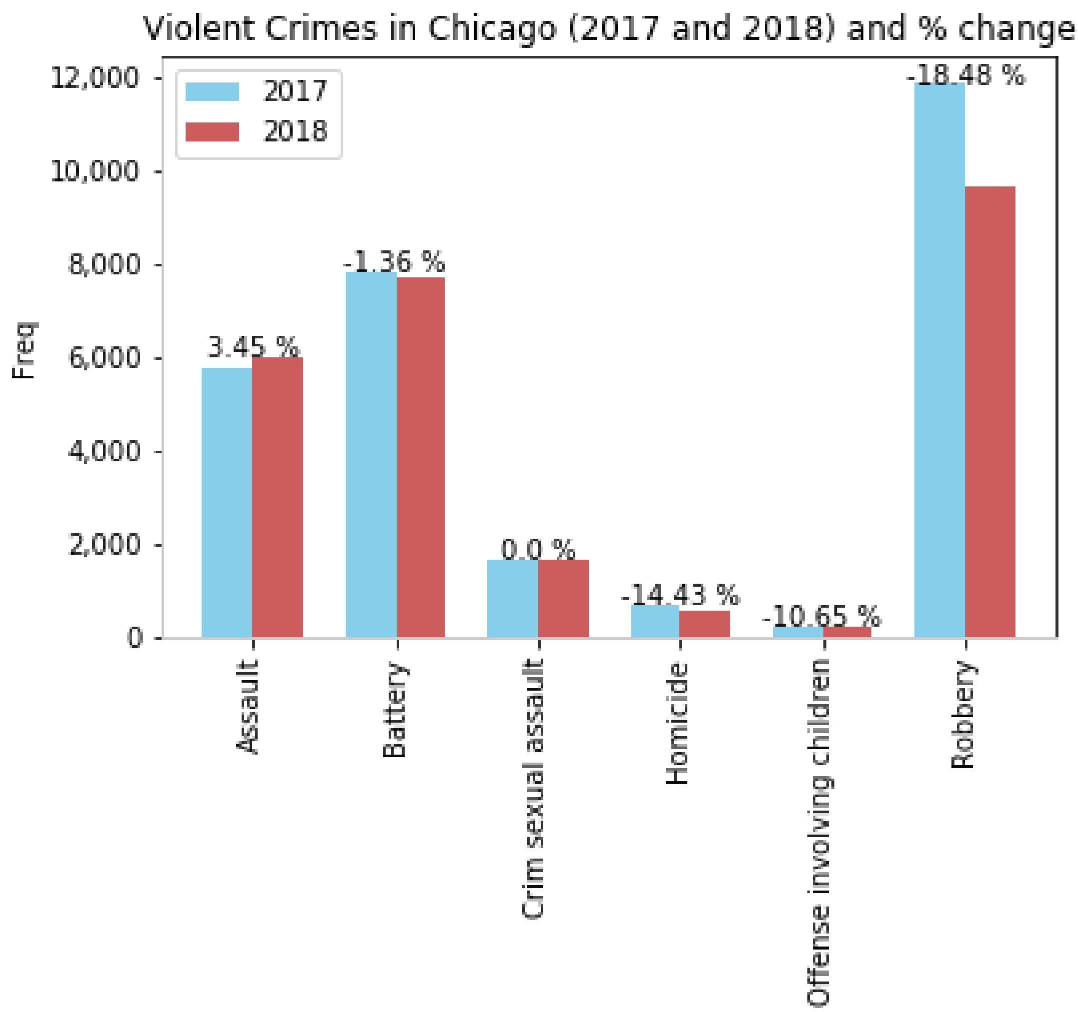
```
In [5]: util.make_cross_var_year(crime_df, 'crime_class')
```

```
Out[5]:                       Total    2017    2018  Perc Change
        CRIME CLASS
        Less serious offences  304170  150871  153299         1.61
        Property crime         176368   89196   87172        -2.27
        Violent crime           53836   28029   25807        -7.93
        Total                  534374  268096  266278        -0.68
```

In total in Chicago there were 534 thousand crimes reported in 2017 and 2018. In 2017 there were 268 thousand and in 2018 266 thousand, resulting in a very small percentual reduction of 0.68%. Between the three main types of crimes, the less serious offences were the most common, acocunting for 60% of all the crimes, followed by Property Crimes with 34% and then followed by Violent crimes with 10%. From 2017 to 2018, while the less seriouc offences slightly increased by 1.6%, positive results were observed in property crimes, which decreased by 2.2%, in and violent crimes, which decreased by 7.9%.

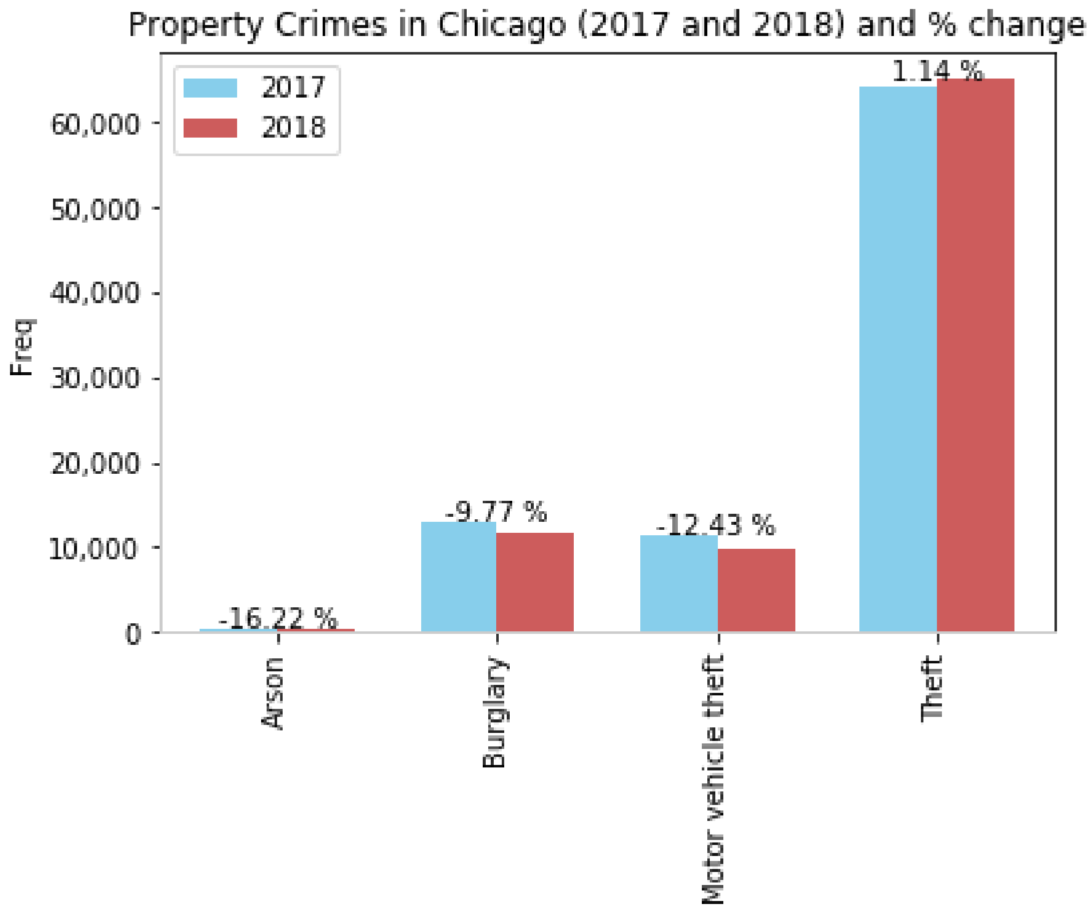
```
In [6]: %matplotlib inline

In [7]: util.graph_crimes_year(crime_df, 'Violent Crime')
```



Among the violent crimes, the most common one was Robbery, which decreased 18.5% from 2017 to 2018. Battery, the second most common violent crime, remained in similar levels, with only a reduction of 1.36%. As it may be noted, Homicide showed a substantial decrease of 14.4%.

```
In [8]: util.graph_crimes_year(crime_df, 'Property Crime')
```



Among the property crimes, theft was the most common and slightly increased from 2017 to 2018 by 1.1%. Burgalary and Motor Vehicle theft decreased by 9.8% and 12.4% respectivelly.

Communities with the highest number of crimes.

```
In [9]: comm_df = util.load_community_area_data()
```

```
WARNING:root:Requests made without an app_token will be subject to strict throttling limits.
```

```
In [10]: comm_total_crime = util.make_cross_comm_crime_year(comm_df, crime_df)
         comm_total_crime.sort_values('Total', ascending = False)[['community', 2017, 2018,'To
```

Out[10]:

	community	2017	2018	Total	Perc Change
25	AUSTIN	15386	15042	30428	-2.235799
36	NEAR NORTH SIDE	12311	13056	25367	6.051499
37	LOOP	10624	10856	21480	2.183735
28	NEAR WEST SIDE	9049	9399	18448	3.867831
29	NORTH LAWNDALE	9065	9311	18376	2.713734
38	SOUTH SHORE	8677	8648	17325	-0.334217

```
23    HUMBOLDT PARK   8056   8039  16095    -0.211023
24        WEST TOWN   8324   7301  15625   -12.289765
69   AUBURN GRESHAM   7527   7370  14897    -2.085824
64   WEST ENGLEWOOD   6950   7067  14017     1.683453
```

When considering all crimes, Austin, the Near North Side, the Loop, the Near West Side and North Lawndale were the five neighborhoods with the highers number of total crimes between 2017 and 2018. Of there, Austin was the only one that experience a reduction, of only 2.23%. However, this result is highly drawn by the minor offenses.

Communities with the highest number of violent crimes.

```
In [11]: comm_violent_crime = util.make_cross_comm_crime_year(comm_df, crime_df, 'Violent Crim
         comm_violent_crime.sort_values('Total', ascending = False)[['community', 2017, 2018,''
```

```
Out[11]:                community  2017  2018  Total  Perc Change
         25                AUSTIN  2070  1986   4056    -4.057971
         29        NORTH LAWNDALE  1165  1136   2301    -2.489270
         38           SOUTH SHORE  1196  1002   2198   -16.220736
         23         HUMBOLDT PARK  1050   917   1967   -12.666667
         69        AUBURN GRESHAM   968   900   1868    -7.024793
         66  GREATER GRAND CROSSING   970   837   1807   -13.711340
         64        WEST ENGLEWOOD   919   885   1804    -3.699674
         46              ROSELAND   837   777   1614    -7.168459
         28        NEAR WEST SIDE   862   749   1611   -13.109049
         65             ENGLEWOOD   781   753   1534    -3.585147
```

In this table we can observe that Austin is also the community with the highest number of total crimes, but in this case, is followed by North Lawndale, South Shore, Humboldt Park and Auburn Gresham. It can be observed as well, that all the 10 communities with the highest number of violent crimes experiences a decrease from 2017 to 2018, that ranges from 2.48% for North Lawndale, to 16.22% for South Shore.

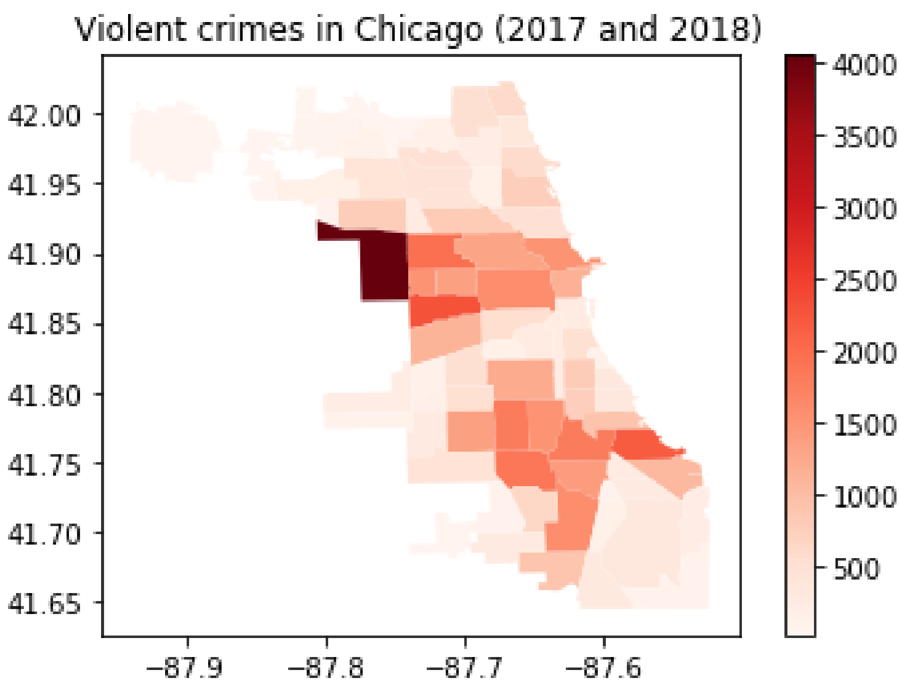
Communities with the highest increase of violent crimes.

```
In [12]: comm_violent_crime.sort_values('Perc Change', ascending = False)[['community', 2017,
```

```
Out[12]:              community  2017  2018  Total  Perc Change
         48             PULLMAN    73    88    161    20.547945
         42            BURNSIDE    46    55    101    19.565217
         13         ALBANY PARK   233   271    504    16.309013
         7           HYDE PARK   166   193    359    16.265060
         53           HEGEWISCH    45    52     97    15.555556
         19         WEST RIDGE   251   290    541    15.537849
         52           RIVERDALE   145   165    310    13.793103
         12          NORTH PARK    60    68    128    13.333333
         27  EAST GARFIELD PARK   652   724   1376    11.042945
         67        LINCOLN PARK   241   266    507    10.373444
```

Somme communities experienced a substantial increase in Violent Crimes, of which Albany Park, Hyde Park, West Ridge, Riverdale and particullarly East Garfiel Park are alarming because all had more than 100 Violent Crimes per year and had an increase above 10%.

```
In [13]: comm_area_crime = util.merge_comm_crime(comm_df, crime_df, 'crime_class')
         geo_comm_area_crime = util.convert_to_geopandas(comm_area_crime)
         util.map_comm_crime(geo_comm_area_crime, 'Violent Crime')
```

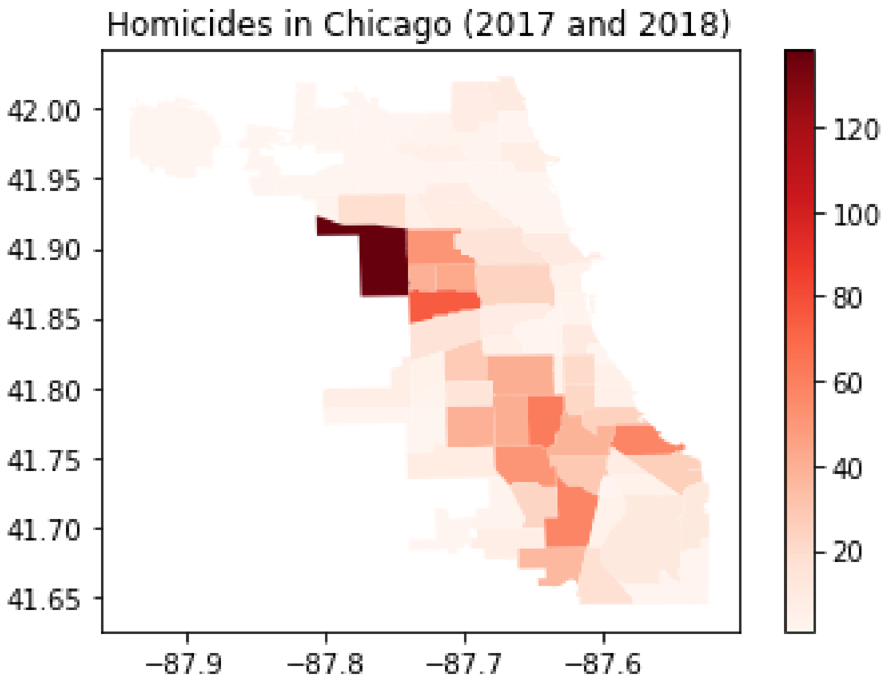



```
<Figure size 432x288 with 0 Axes>
```

As expected, there was a concentration of Violent Crimes on the southern and western communities of Chicago, and less concentration on the north and central part of the City.

```
In [14]: comm_area_crime = util.merge_comm_crime(comm_df, crime_df, 'primary_type')
         geo_comm_area_crime = util.convert_to_geopandas(comm_area_crime)
         util.map_comm_crime(geo_comm_area_crime, 'HOMICIDE')

/anaconda3/lib/python3.6/site-packages/matplotlib/colors.py:512: RuntimeWarning: invalid value
  xa[xa < 0] = -1
```



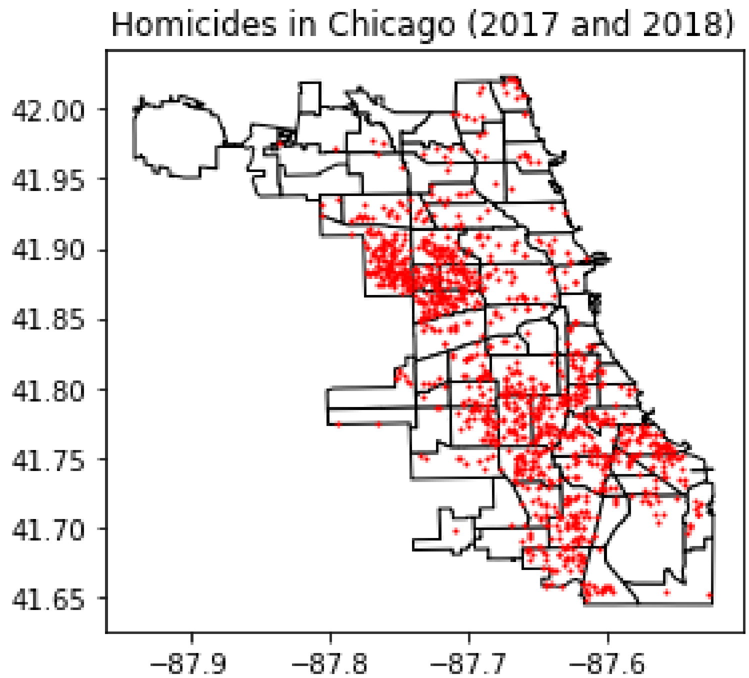

```
<Figure size 432x288 with 0 Axes>
```

Similarly, in terms of Homicides, the same pattern can be observed. A concentration of homicides on the south and west part of the city.

```
In [15]: geo_comm = util.convert_to_geopandas(comm_df)
         geo_crime = util.convert_to_geopandas(crime_df)
```

```
In [16]: util.map_crimes_loc(geo_comm, geo_crime, 'primary_type', 'HOMICIDE')
```

```
<Figure size 432x288 with 0 Axes>
```




```
<Figure size 432x288 with 0 Axes>
```

This map shows the exact location of the homicides in Chicago, reflecting the same information as the previous map.

1.2 Problem 2: Data Augmentation and APIs [40 pts]

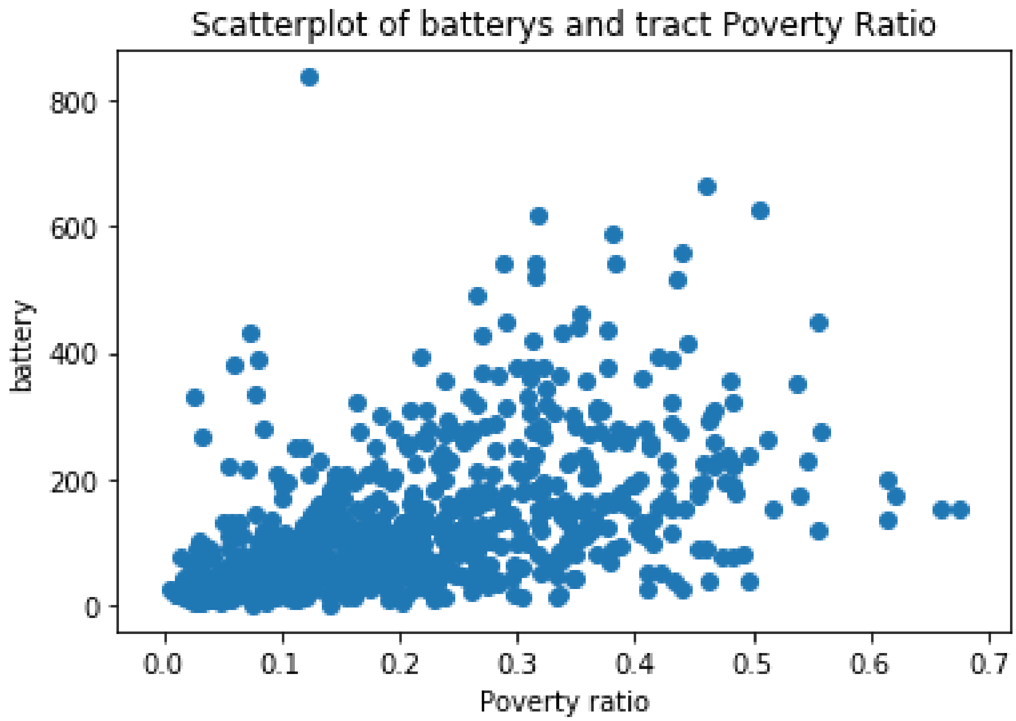
```
In [17]: acs_df = util.load_census_data(util.ACS_TABLES_KEYS)
         tract_area_df = util.load_tract_shapefile()
         tract_acs = util.merge_tract_census(tract_area_df, acs_df)
         geo_tract = util.convert_to_geopandas(tract_acs)
         tract_census_crime = util.join_tract_census_crime(geo_tract, geo_crime)
```

```
WARNING:root:Requests made without an app_token will be subject to strict throttling limits.
```

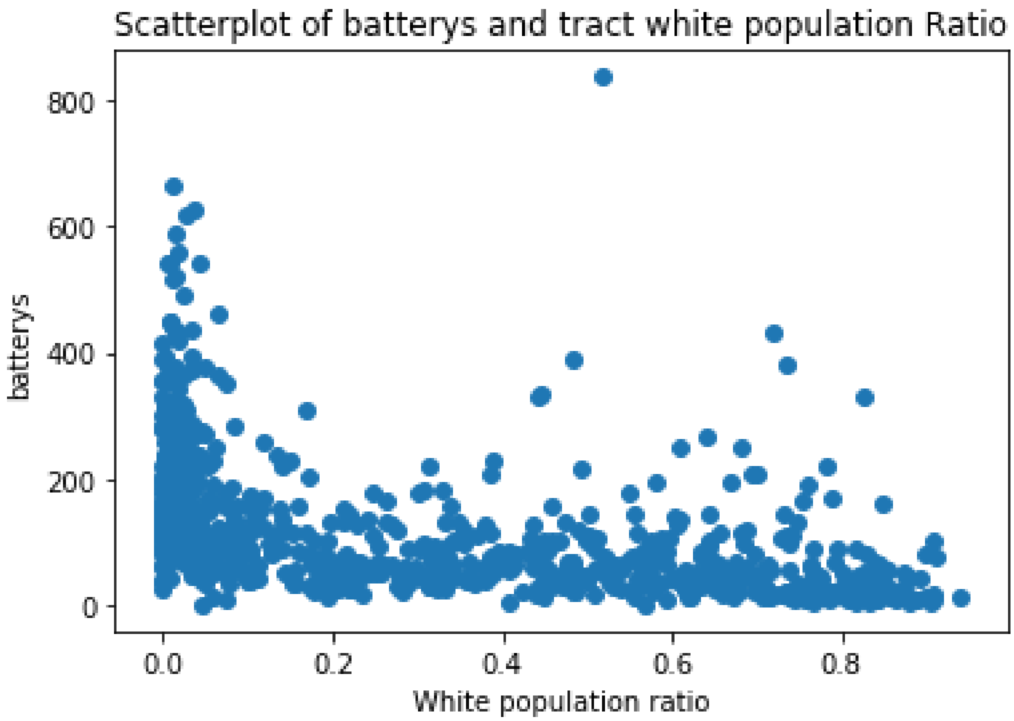
1.2.1 Tracts with Battery

```
In [18]: per_tract_battery = tract_census_crime[tract_census_crime['primary_type'] == 'BATTERY
         per_tract_battery.rename(columns={'arrest':'count'}, inplace= True)
```

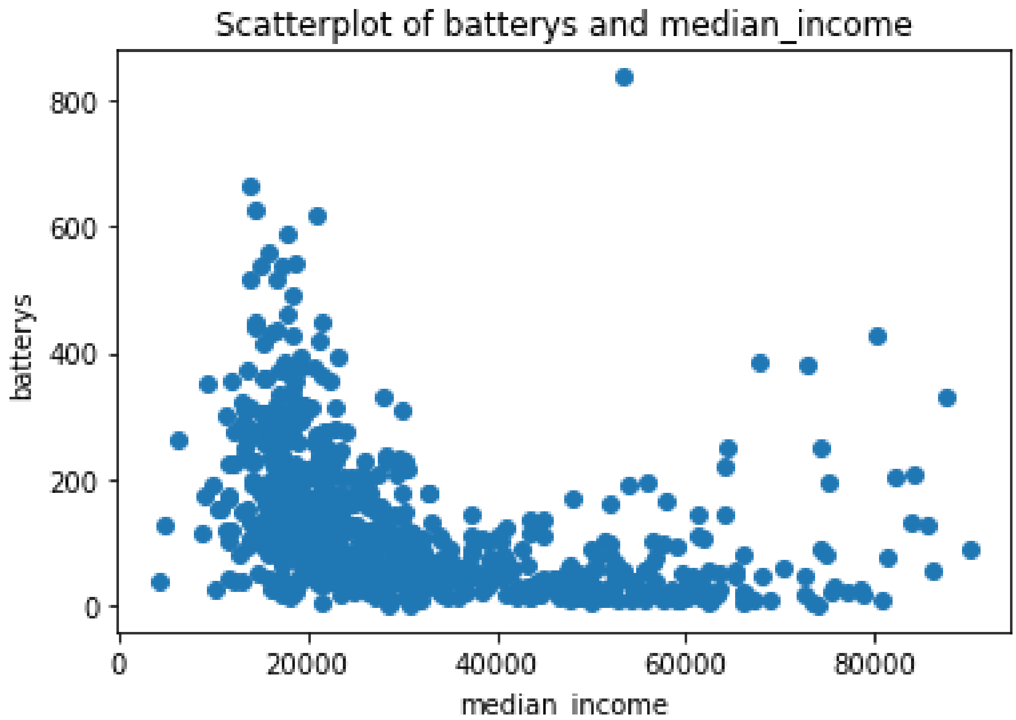
```
In [19]: plt.scatter(per_tract_battery['below_poverty_ratio'], per_tract_battery['count'])
         plt.title('Scatterplot of batterys and tract Poverty Ratio')
         plt.xlabel("Poverty ratio")
         plt.ylabel("battery")
         plt.show()
```




```
In [20]: plt.scatter(per_tract_battery['white_population_ratio'], per_tract_battery['count'])
         plt.title('Scatterplot of batterys and tract white population Ratio')
         plt.xlabel("White population ratio")
         plt.ylabel("batterys")
         plt.show()
```




```
In [21]: plt.scatter(per_tract_battery['median_income'], per_tract_battery['count'])
         plt.title('Scatterplot of batterys and median_income')
         plt.xlabel('median_income')
         plt.ylabel("batterys")
         plt.show()
```

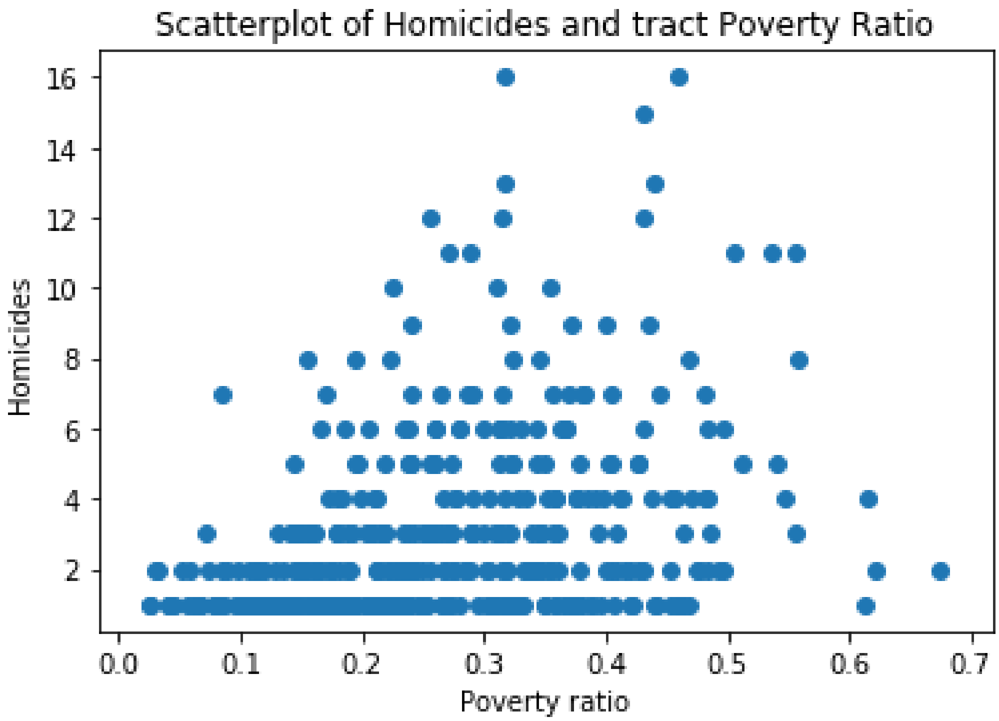


As it can be seen in these three scatterplots, Battery crimes are more present on tracts with higher poverty rate, with lower ratio of white population and with lower median income.

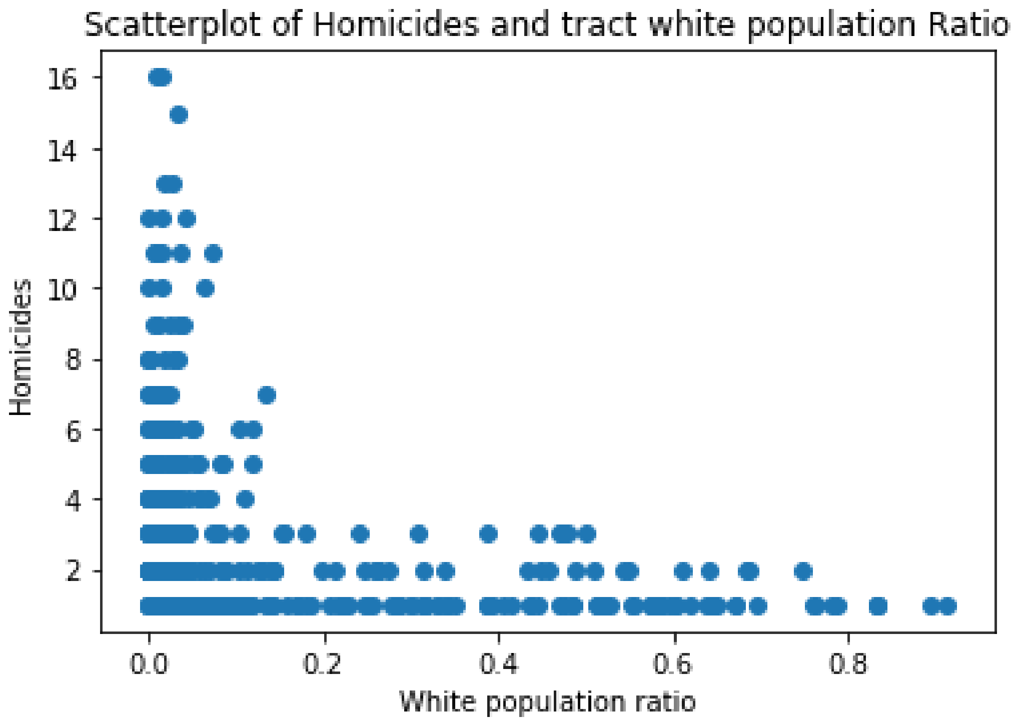
1.2.2 Tracts with Homicide

```
In [22]: per_tract_homicide = tract_census_crime[tract_census_crime['primary_type'] == 'HOMICII
         per_tract_homicide.rename(columns={'arrest':'count'}, inplace= True)
```

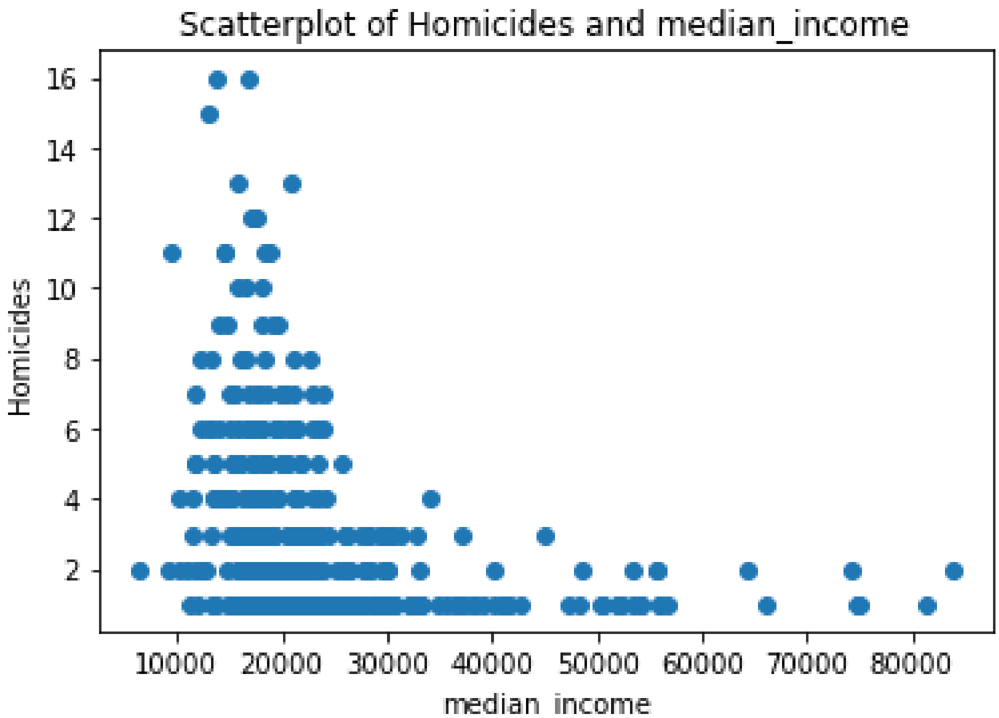
```
In [23]: plt.scatter(per_tract_homicide['below_poverty_ratio'], per_tract_homicide['count'])
         plt.title('Scatterplot of Homicides and tract Poverty Ratio')
         plt.xlabel("Poverty ratio")
         plt.ylabel("Homicides")
         plt.show()
```



In [24]:
```
plt.scatter(per_tract_homicide['white_population_ratio'], per_tract_homicide['count'])
plt.title('Scatterplot of Homicides and tract white population Ratio')
plt.xlabel("White population ratio")
plt.ylabel("Homicides")
plt.show()
```



In [25]:
```
plt.scatter(per_tract_homicide['median_income'], per_tract_homicide['count'])
plt.title('Scatterplot of Homicides and median_income')
plt.xlabel('median_income')
plt.ylabel("Homicides")
plt.show()
```



As it can be seen in these three scatterplots, Homicides are more present on tracts with higher poverty rate, with lower ratio of white population and with lower median income.

1.2.3 Comparison over time

```
In [48]: tract_batt_year = tract_census_crime[tract_census_crime['primary_type'] == 'BATTERY']
         tract_batt_year.rename(columns={'arrest':'count'}, inplace= True)
         tract_batt_year[tract_batt_year['count']>0].groupby('year').mean()
```

```
Out[48]:       below_poverty_ratio  median_income  white_population_ratio  \
         year
         2017             0.206915   30554.804786                0.306353
         2018             0.206832   30554.983648                0.306026

               foreign_born_ratio      count
         year
         2017            0.017332  61.346683
         2018            0.017310  62.011250
```

We do not observe any significant difference between the average characteristics of the trects that hac battery between 2017 and 2017.

Homicide trect mean characs

```
In [42]: tract_hom_year = tract_census_crime[tract_census_crime['primary_type'] == 'HOMICIDE']
         tract_hom_year.rename(columns={'arrest':'count'}, inplace= True)
         tract_hom_year[tract_hom_year['count']>0].groupby('year').mean()
```

```
Out[42]:       below_poverty_ratio  median_income  white_population_ratio  \
         year
         2017             0.276018   22302.349515                0.109697
         2018             0.278739   22683.433898                0.121915

               foreign_born_ratio     count
         year
         2017            0.013138  2.187702
         2018            0.012652  1.986441
```

We only observe a slight increase in the average white population ratio of the trects that had one homicide.

1.2.4 Deceptive practice vs Sex Offence

```
In [85]: tract_decep_sex = tract_census_crime[(tract_census_crime['primary_type'] == 'DECEPTIVE
         tract_decep_sex.rename(columns={'arrest':'count'}, inplace= True)
         tract_decep_sex[tract_decep_sex['count'] > 0].groupby('primary_type').mean()
```

```
Out[85]:                     below_poverty_ratio  median_income  \
         primary_type
         DECEPTIVE PRACTICE             0.207017   30551.537783
         SEX OFFENSE                    0.213702   29404.889231

                             white_population_ratio  foreign_born_ratio      count
         primary_type
         DECEPTIVE PRACTICE                0.306055            0.017350  44.463079
         SEX OFFENSE                       0.281011            0.017679   3.059724
```

We found trects very simmilar in terms of poverty ratio, median income, white population ratio and foreign born ratio.

1.3 Problem 3: Analysis and Communication [25 pts]

Overall crime in Chicago did not change between 2017 and 2018. However, violent crimes were reduced by 7.003% and Homicide crimes by 14%. The communities with the highest number of violent crimes saw a reduction that ranges from 2.48% for North Lawndale, to 16.22% for South Shore. Property crimes decresed by 2.27% in the same period. Median income, white population ratio and poverty ratio of the trects where homicide and battery occured did not change, and the trend of observing the mayority of these crimes on less favored communities did not change.

The information that the Alderman presents in misleading. Meassuring 2017 to 2018, Robberies debreased 18%, aggravated batteries decreased 1.36%, burglaries decreased 9.77% and Motor Vehicle Theft decreased by 12.43%.

1.4 Problem 4

Of the types of crimes you have data for, which crime type is the most likely given the call came from 2111 S Michigan Ave? What are the probabilities for each type of request?

```
In [120]: s_mich = crime_df[crime_df['block'] == '021XX S MICHIGAN AVE'].groupby('primary_type
          s_mich = s_mich.reset_index()
          s_mich['prob'] = s_mich[0] / s_mich[0].sum()
          s_mich
```

```
Out[120]:             primary_type   0      prob
          0                BATTERY  16  0.266667
          1          OTHER OFFENSE  13  0.216667
          2                  THEFT   6  0.100000
          3     DECEPTIVE PRACTICE   6  0.100000
          4        CRIMINAL DAMAGE   6  0.100000
          5                ASSAULT   6  0.100000
          6                ROBBERY   2  0.033333
          7    MOTOR VEHICLE THEFT   2  0.033333
          8  PUBLIC PEACE VIOLATION   1  0.016667
          9      CRIMINAL TRESPASS   1  0.016667
          10              BURGLARY   1  0.016667
```

The most probable crime is Battery with 27%.