hw1

April 8, 2019

In [3]:
```
# Problem 1: Data Acquisition and Analysis
## download the first 50000 rows of data
import requests
import pandas as pd
path = 'https://data.cityofchicago.org/resource/d62x-nvdr.json?$limit=50000'
r = requests.get(path).json()
crimes_2017 = pd.DataFrame(r[1:], columns=r[0])

path = 'https://data.cityofchicago.org/resource/3i3m-jwuy.json?$limit=50000'
j = requests.get(path).json()
crimes_2018 = pd.DataFrame(j[1:], columns=j[0])

## 1. count the frequency of each type of crime
frequency_2017 = crimes_2017['primary_type'].value_counts()
frequency_2018 = crimes_2018['primary_type'].value_counts()

print('The Three Most Common Types of Crime in 2017:')
print(frequency_2017.head(3))
print('The Three Most Common Types of Crime in 2018:')
print(frequency_2018.head(3))
```

```
The Three Most Common Types of Crime in 2017:
THEFT              12600
BATTERY             8717
CRIMINAL DAMAGE     5327
Name: primary_type, dtype: int64
The Three Most Common Types of Crime in 2018:
THEFT              12733
BATTERY             8990
CRIMINAL DAMAGE     5144
Name: primary_type, dtype: int64
```

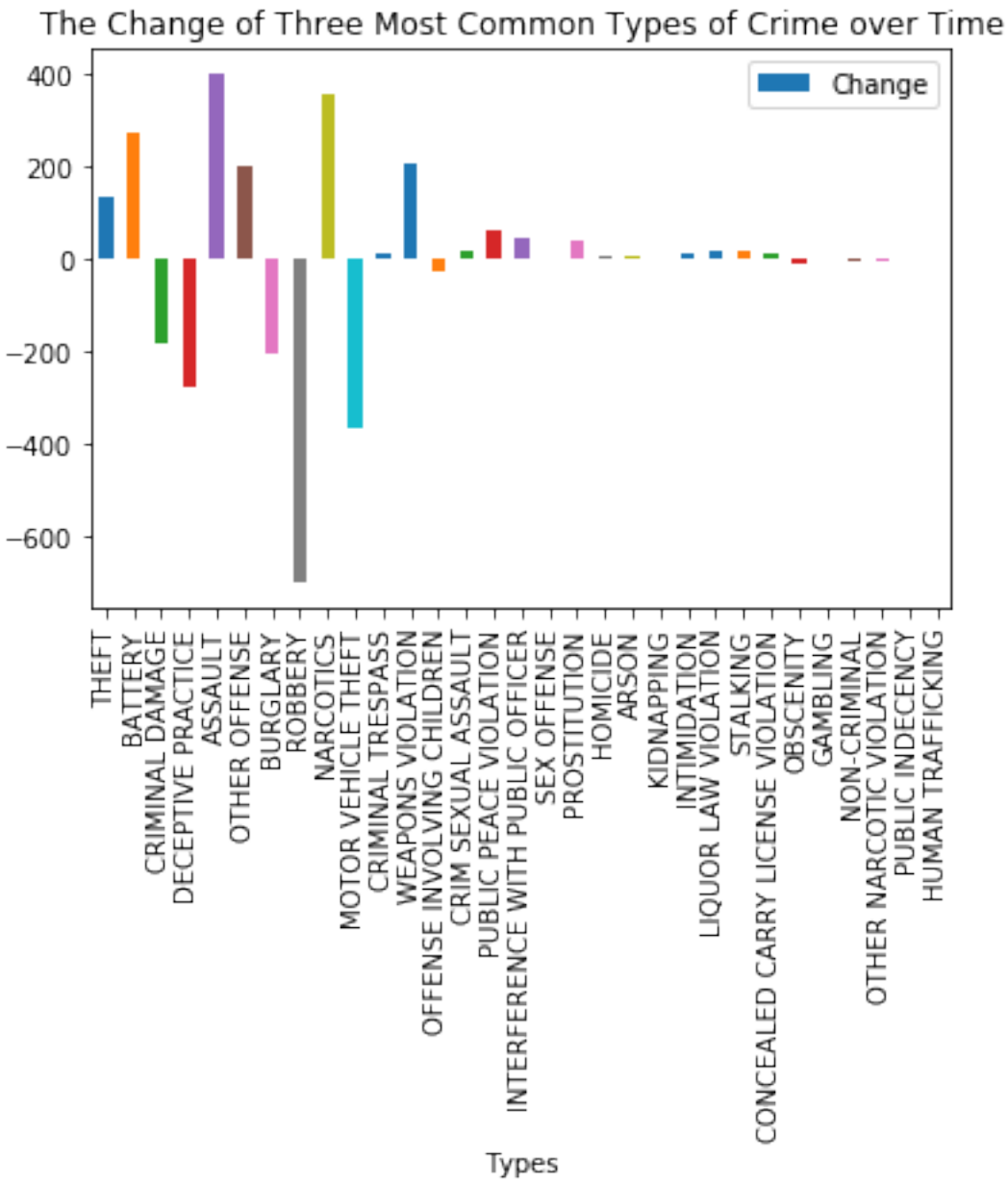
In [4]:
```
## 2. calculate the change of crimes over time
time_trend = frequency_2017.reset_index().merge(frequency_2018.reset_index(),
                                                on='index', how='outer')
time_trend.columns = ['Types', '2017', '2018']
time_trend['Sum'] = time_trend['2018'] + time_trend['2017']
```

```
        time_trend['Change'] = time_trend['2018'] - time_trend['2017']
        time_trend = time_trend.sort_values('Sum', ascending=False)
        print('The Change of Three Most Common Types of Crime over Time:')
        print(time_trend.head(3))
```

```
The Change of Three Most Common Types of Crime over Time:
              Types   2017   2018    Sum  Change
0            THEFT  12600  12733  25333     133
1          BATTERY   8717   8990  17707     273
2  CRIMINAL DAMAGE   5327   5144  10471    -183
```

In [5]:
```
        ## plot the change over time
        import matplotlib.pyplot as plt
        ax1 = time_trend.plot.bar(x='Types', y='Change')
        ax1.set_title('The Change of Three Most Common Types of Crime over Time')
        ax1
```

Out[5]: <matplotlib.axes._subplots.AxesSubplot at 0x1231e5358>



In [6]:
```
## 3. find the three community area with the most homicides
homicide_2017 = crimes_2017[crimes_2017['primary_type'] == 'HOMICIDE']
community_2017 = homicide_2017['community_area'].value_counts().reset_index()
homicide_2018 = crimes_2018[crimes_2018['primary_type'] == 'HOMICIDE']
community_2018 = homicide_2018['community_area'].value_counts().reset_index()

homicide = community_2017.merge(community_2018, on='index', how='outer').fillna(0)
homicide.columns = ['Community Area', '2017', '2018']
```

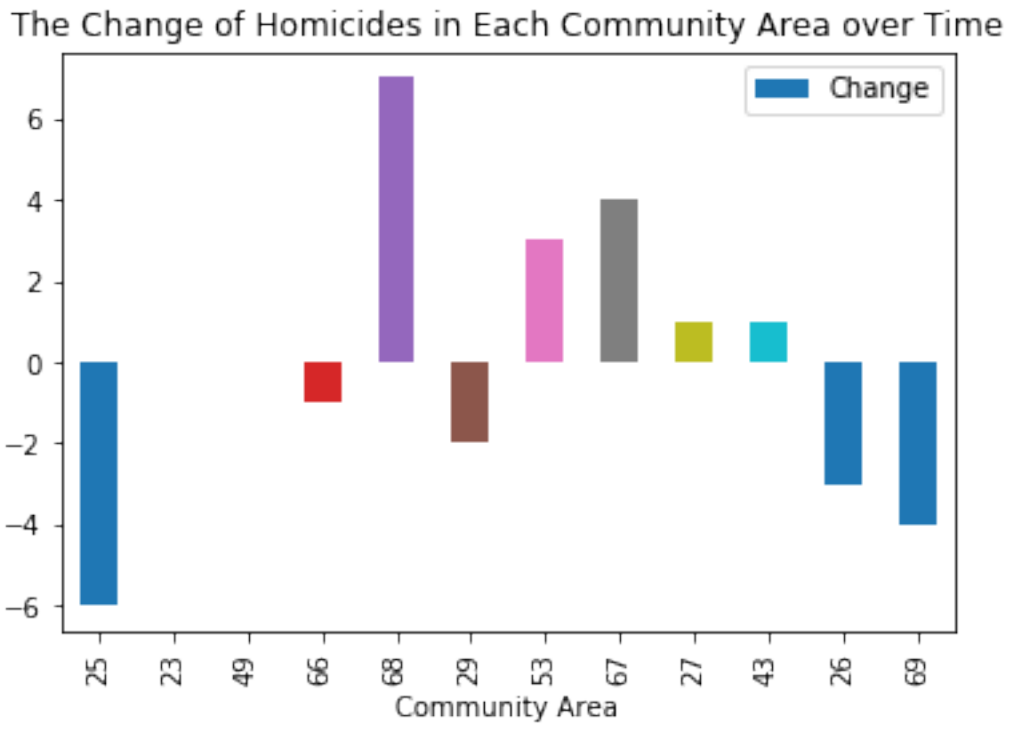
In [7]:
```
## 4. calculate the change by community area over time
```

```
homicide['Sum'] = homicide['2018'] + homicide['2017']
homicide['Change'] = homicide['2018'] - homicide['2017']
homicide = homicide.sort_values('Sum', ascending=False)
print('The Three Community Areas with Most Homicides:')
print('25 Austin, 23 Humboldt Park, 49 Roseland')
print(homicide.head(3))
## data from https://www.chicago.gov/content/dam/city/depts/doit/general
##  /GIS/Chicago_Maps/Citywide_Maps/Community_Areas_W_Numbers.pdf
```

```
The Three Community Areas with Most Homicides:
25 Austin, 23 Humboldt Park, 49 Roseland
  Community Area  2017  2018   Sum  Change
0             25  14.0   8.0  22.0    -6.0
1             23   7.0   7.0  14.0     0.0
2             49   6.0   6.0  12.0     0.0
```

In [8]:
```
## plot the change over time
ax2 = homicide[homicide['Sum']>5].plot.bar(x='Community Area', y='Change')
ax2.set_title('The Change of Homicides in Each Community Area over Time')
ax2
```

Out[8]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x11c62ea58>
```



In [9]:
```
# Problem 2: Data Augmentation and APIs
## download the census data
api = 'https://data.cityofchicago.org/resource/kn9c-c2s2.json'
j = requests.get(api).json()
census = pd.DataFrame(j[1:], columns=j[0])

## 1. What types of blocks have reports of BATTERY?
battery_2017 = crimes_2017[crimes_2017['primary_type'] == 'BATTERY']
community_2017 = battery_2017['community_area'].value_counts().reset_index()
battery_2018 = crimes_2018[crimes_2018['primary_type'] == 'BATTERY']
community_2018 = battery_2018['community_area'].value_counts().reset_index()

battery = community_2017.merge(community_2018, on='index', how='outer').fillna(0)
battery.columns = ['ca', '2017', '2018']
batteries = battery.head(3).merge(census, how='left')
batteries = batteries.append(census[census['community_area_name'] == 'CHICAGO'])
print(batteries)
print('Compare to the Average Level of Chicago,')
print('Community Areas with Most Batteries have:')
print('lower per capita income, higher unemployment rate, and lower education level.')
```

```
     2017   2018   ca community_area_name hardship_index per_capita_income_  \
0   543.0  605.0   25              Austin             73              15957
1   383.0  381.0   43         South Shore             55              19398
2   334.0  328.0   29      North Lawndale             87              12034
76    NaN    NaN  NaN             CHICAGO            NaN              28202

   percent_aged_16_unemployed percent_aged_25_without_high_school_diploma  \
0                        22.6                                        24.4
1                          20                                          14
2                        21.2                                        27.6
76                       12.9                                        19.5

   percent_aged_under_18_or_over_64 percent_households_below_poverty  \
0                              37.9                             28.6
1                              35.7                             31.1
2                              42.7                             43.1
76                             33.5                             19.7

   percent_of_housing_crowded
0                         6.3
1                         2.8
2                         7.4
76                        4.7
Compare to the Average Level of Chicago,
Community Areas with Most Batteries have:
lower per capita income, higher unemployment rate, and lower education level.
```

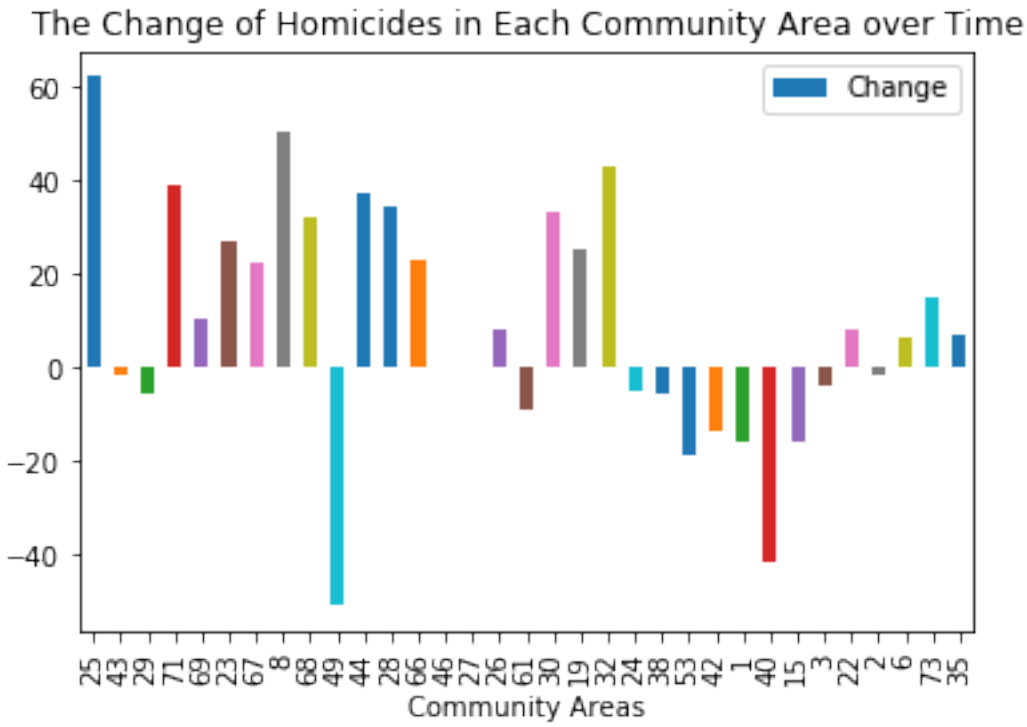
In [10]:
```
## 2. What types of blocks get Homicide?
homicides = homicide.head(3).merge(census,
                                   left_on='Community Area', right_on='ca', how='left
homicides = homicides.append(census[census['community_area_name'] == 'CHICAGO'])
print(homicides)
print('Compare to the Average Level of Chicago,')
print('Community Areas with Most Homicides have:')
print('lower per capita income, higher unemployment rate, and lower education level.')
```

```
    2017  2018  Change Community Area   Sum   ca community_area_name  \
0   14.0   8.0    -6.0             25  22.0   25              Austin
1    7.0   7.0     0.0             23  14.0   23       Humboldt park
2    6.0   6.0     0.0             49  12.0   49            Roseland
76   NaN   NaN     NaN            NaN   NaN  NaN             CHICAGO

   hardship_index per_capita_income_ percent_aged_16_unemployed  \
0              73              15957                       22.6
1              85              13781                       17.3
2              52              17949                       20.3
76            NaN              28202                       12.9

   percent_aged_25_without_high_school_diploma  \
0                                         24.4
1                                         35.4
2                                         16.9
76                                        19.5

   percent_aged_under_18_or_over_64 percent_households_below_poverty  \
0                              37.9                              28.6
1                                38                              33.9
2                              41.2                              19.8
76                             33.5                              19.7

   percent_of_housing_crowded
0                         6.3
1                        14.8
2                         2.5
76                        4.7
Compare to the Average Level of Chicago,
Community Areas with Most Homicides have:
lower per capita income, higher unemployment rate, and lower education level.
```

In [11]:
```
## 3. Does that change over time in the data you collected?
battery['Sum'] = battery['2018'] + battery['2017']
battery['Change'] = battery['2018'] - battery['2017']
battery = battery.sort_values('Sum', ascending=False)
ax3 = battery[battery['Sum']>200].plot.bar(x='ca', y='Change')
```

```
ax3.set_xlabel('Community Areas')
ax3.set_title('The Change of Homicides in Each Community Area over Time')
ax3

ax2
```

Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x11c62ea58>



In [13]:
```
## 4. What is the difference in blocks that get Deceptive Practice vs Sex Offense?
## deceptive practice
deceptive_2017 = crimes_2017[crimes_2017['primary_type'] == 'DECEPTIVE PRACTICE']
community_2017 = battery_2017['community_area'].value_counts().reset_index()
deceptive_2018 = crimes_2018[crimes_2018['primary_type'] == 'DECEPTIVE PRACTICE']
community_2018 = battery_2018['community_area'].value_counts().reset_index()

deceptive = community_2017.merge(community_2018, on='index', how='outer').fillna(0)
deceptive.columns = ['ca', '2017', '2018']
deceptive['Sum'] = deceptive['2018'] + deceptive['2017']
deceptive['Change'] = deceptive['2018'] - deceptive['2017']
deceptive = deceptive.sort_values('Sum', ascending=False)

deceptives = deceptive.head(3).merge(census, how='left')
deceptives[['per_capita_income_', 'percent_aged_16_unemployed',
            'percent_aged_25_without_high_school_diploma']]
```

```
## sex offense
sex_2017 = crimes_2017[crimes_2017['primary_type'] == 'SEX OFFENSE']
community_2017 = sex_2017['community_area'].value_counts().reset_index()
sex_2018 = crimes_2018[crimes_2018['primary_type'] == 'SEX OFFENSE']
community_2018 = sex_2018['community_area'].value_counts().reset_index()

sex = community_2017.merge(community_2018, on='index', how='outer').fillna(0)
sex.columns = ['ca', '2017', '2018']
sex['Sum'] = sex['2018'] + sex['2017']
sex['Change'] = sex['2018'] - sex['2017']
sex = sex.sort_values('Sum', ascending=False)

sex_offenses = sex.head(3).merge(census, how='left')
sex_offenses[['per_capita_income_', 'percent_aged_16_unemployed',
              'percent_aged_25_without_high_school_diploma']]

## print out results
print('Compare to Community Areas with Most Deceptive Practice,')
print('Areas with Most Sex Offense have:')
print('lower per capita income, higher unemployment rate, and lower education level.')
```

```
Compare to Community Areas with Most Deceptive Practice,
Areas with Most Sex Offense have:
lower per capita income, higher unemployment rate, and lower education level.
```