

DSC640 WEEKS 7 & 8

EXERCISE 4.2

Datasets - You may also download them directly from this link:

<https://content.bellevue.edu/cst/dsc/640/datasets/ex4-2.zip>
(<https://content.bellevue.edu/cst/dsc/640/datasets/ex4-2.zip>)

Exercise Goal:

You need to submit 3 scatterplots, 3 bubble charts and 3 density plot charts using Tableau or PowerBI, Python and R using the data below (or your own datasets). You can also submit using D3. You can choose which library to use in Python or R, documentation is provided to help you decide and as you start to play around in the libraries, you will decide which you prefer.

CINDY HERRERA

January 31, 2020

1. Data collection: importing data and analyzing

```
In [1]: import os
import datetime
import pprint
import pandas as pd
import numpy as np
import squarify
import seaborn as sns
from pandas import ExcelWriter
from pandas import ExcelFile
import matplotlib as mpl
import matplotlib.pyplot as plt
%matplotlib inline
```

```
In [2]: os.getcwd()
```

```
Out[2]: '/Users/Cindy/Desktop/00 data640/ex4-2'
```

Load data into dataframe (In the text file, I noticed each line had '/' so I used the sep function to clean the data being read into the dataset.)

```
In [3]: df1 = pd.read_csv('/Users/Cindy/Desktop/00 data640/ex4-2/crimerates-by-state-2005.csv')
```

```
In [4]: df1 = df1.iloc[1:,]
df1.head()
```

Out[4]:

	state	murder	forcible_rape	robbery	aggravated_assault	burglary	larceny_theft	motor
1	Alabama	8.2	34.3	141.4	247.8	953.8	2650.0	
2	Alaska	4.8	81.1	80.9	465.1	622.5	2599.1	
3	Arizona	7.5	33.8	144.4	327.4	948.4	2965.2	
4	Arkansas	6.7	42.9	91.1	386.8	1084.6	2711.2	
5	California	6.9	26.0	176.1	317.3	693.3	1916.5	

Check the dimension of table and view data Viewing the unemployment data

```
In [5]: df1.head()
```

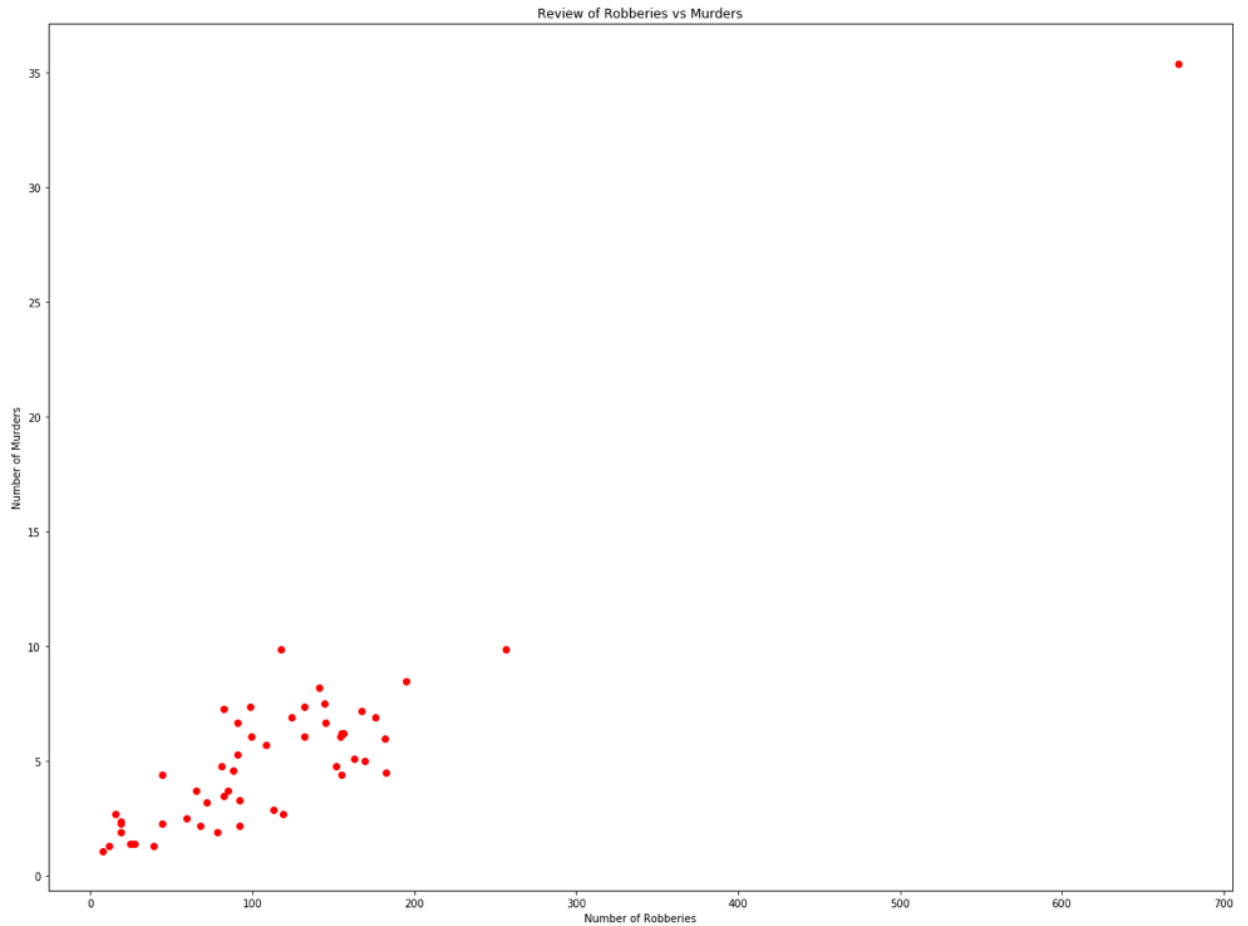
Out[5]:

	state	murder	forcible_rape	robbery	aggravated_assault	burglary	larceny_theft	motor
1	Alabama	8.2	34.3	141.4	247.8	953.8	2650.0	
2	Alaska	4.8	81.1	80.9	465.1	622.5	2599.1	
3	Arizona	7.5	33.8	144.4	327.4	948.4	2965.2	
4	Arkansas	6.7	42.9	91.1	386.8	1084.6	2711.2	
5	California	6.9	26.0	176.1	317.3	693.3	1916.5	

Scatterplot

```
In [6]: fig = plt.figure(figsize=(20,15))
plt.scatter(df1.robbery, df1.murder, color = 'r')
plt.xlabel('Number of Robberies')
plt.ylabel('Number of Murders')
plt.title('Review of Robberies vs Murders')
plt.show
```

```
Out[6]: <function matplotlib.pyplot.show(*args, **kw)>
```

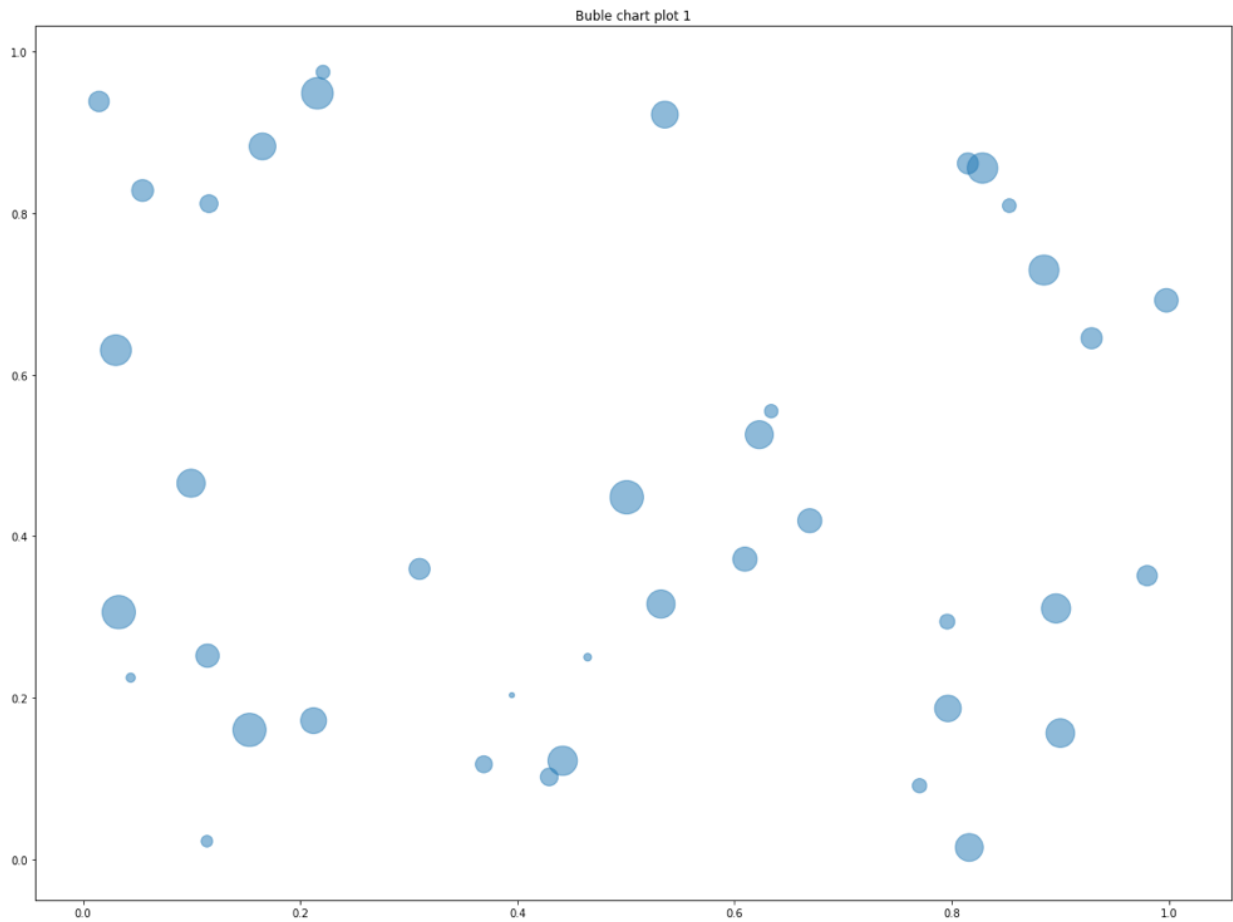


Bubble Chart

```
In [7]: fig = plt.figure(figsize=(20,15))
import numpy as np

# create data
x = np.random.rand(40)
y = np.random.rand(40)
z = np.random.rand(40)

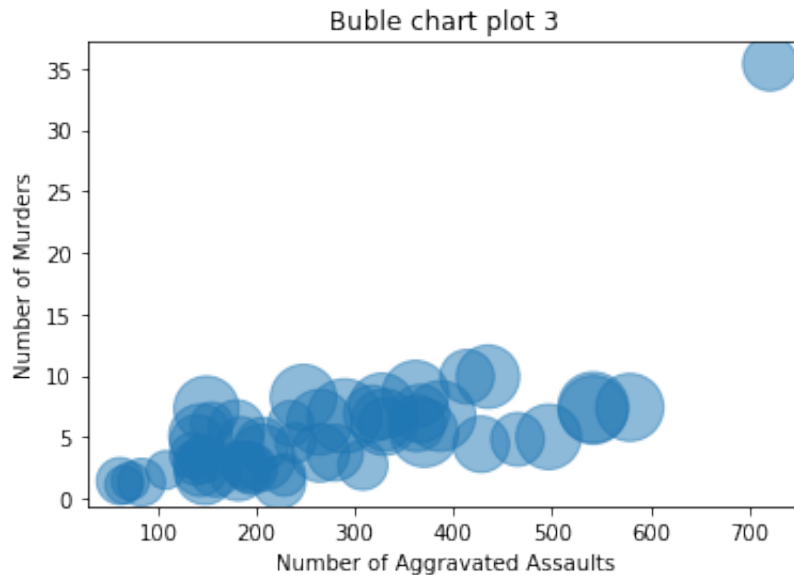
# use the scatter function
plt.scatter(x, y, s=z*1000, alpha=0.5)
plt.title('Buble chart plot 1')
plt.show()
```



Bubble Chart

```
In [9]: plt.scatter(df1.aggravated_assault, df1.murder, df1.burglary, alpha=0.5)
plt.xlabel('Number of Aggravated Assaults')
plt.ylabel('Number of Murders')
plt.title('Buble chart plot 3')
plt.show
```

```
Out[9]: <function matplotlib.pyplot.show(*args, **kw)>
```



Density Plot

```
In [12]: p1=sns.kdeplot(df1['aggravated_assault'], shade=True, color="r")
p1=sns.kdeplot(df1['burglary'], shade=True, color="b")
```

