Assignment 4.2

Veera Koppula

05/04/2022

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.

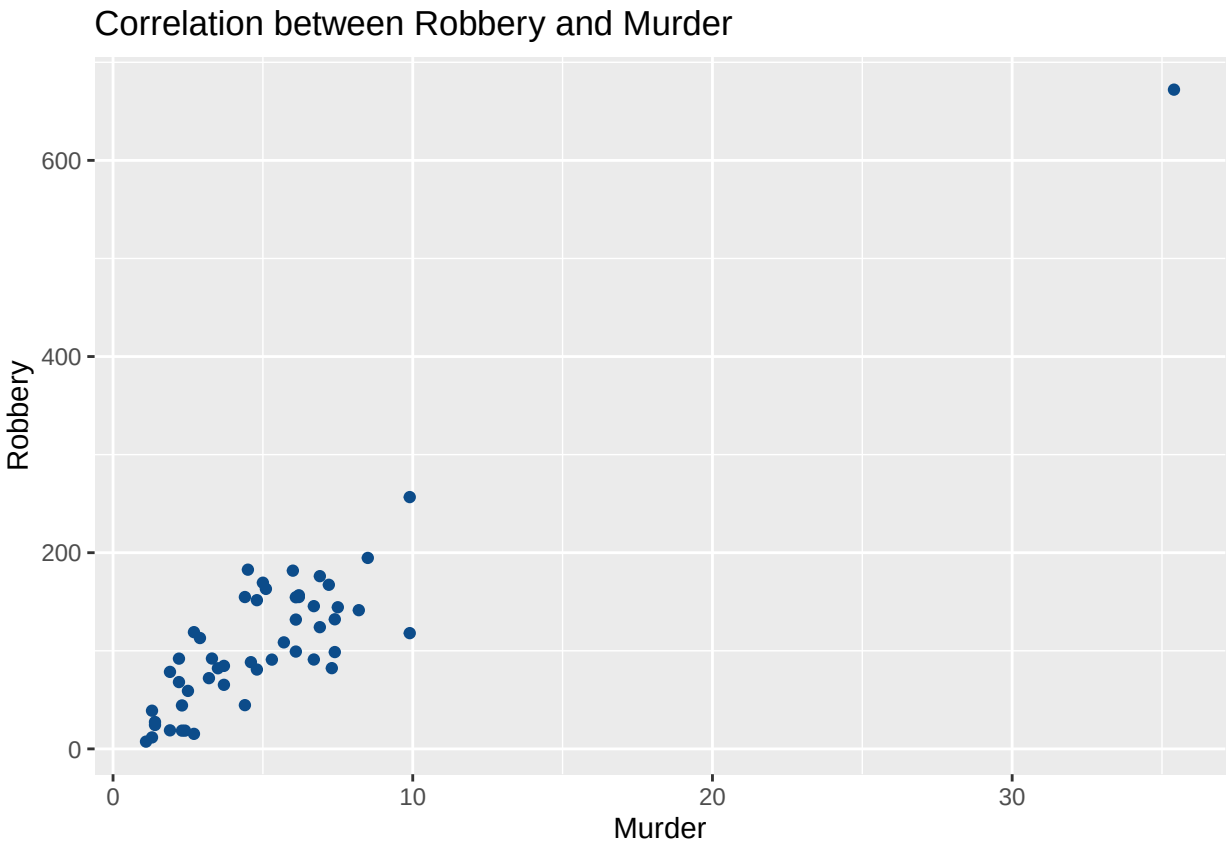
```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
##        state murder forcible_rape robbery aggravated_assault burglary
## 1    Alabama    8.2          34.3   141.4              247.8    953.8
## 2     Alaska    4.8          81.1    80.9              465.1    622.5
## 3    Arizona    7.5          33.8   144.4              327.4    948.4
## 4   Arkansas    6.7          42.9    91.1              386.8   1084.6
## 5 California    6.9          26.0   176.1              317.3    693.3
## 6   Colorado    3.7          43.4    84.6              264.7    744.8
##   larceny_theft motor_vehicle_theft population total_crime state_cont
## 1        2650.0               288.3    4545049      4323.8          1
## 2        2599.1               391.0     669488      4244.5          2
## 3        2965.2               924.4    5974834      5351.1          3
## 4        2711.2               262.1    2776221      4585.4          4
## 5        1916.5               712.8   35795255      3848.9          5
## 6        2735.2               559.5    4660780      4435.9          6
##    name_caps abbr      region midwest
## 1    ALABAMA   AL Rest of USA       0
## 2     ALASKA   AK Rest of USA       0
## 3    ARIZONA   AZ Rest of USA       0
## 4   ARKANSAS   AR Rest of USA       0
## 5 CALIFORNIA   CA Rest of USA       0
## 6   COLORADO   CO Rest of USA       0
```

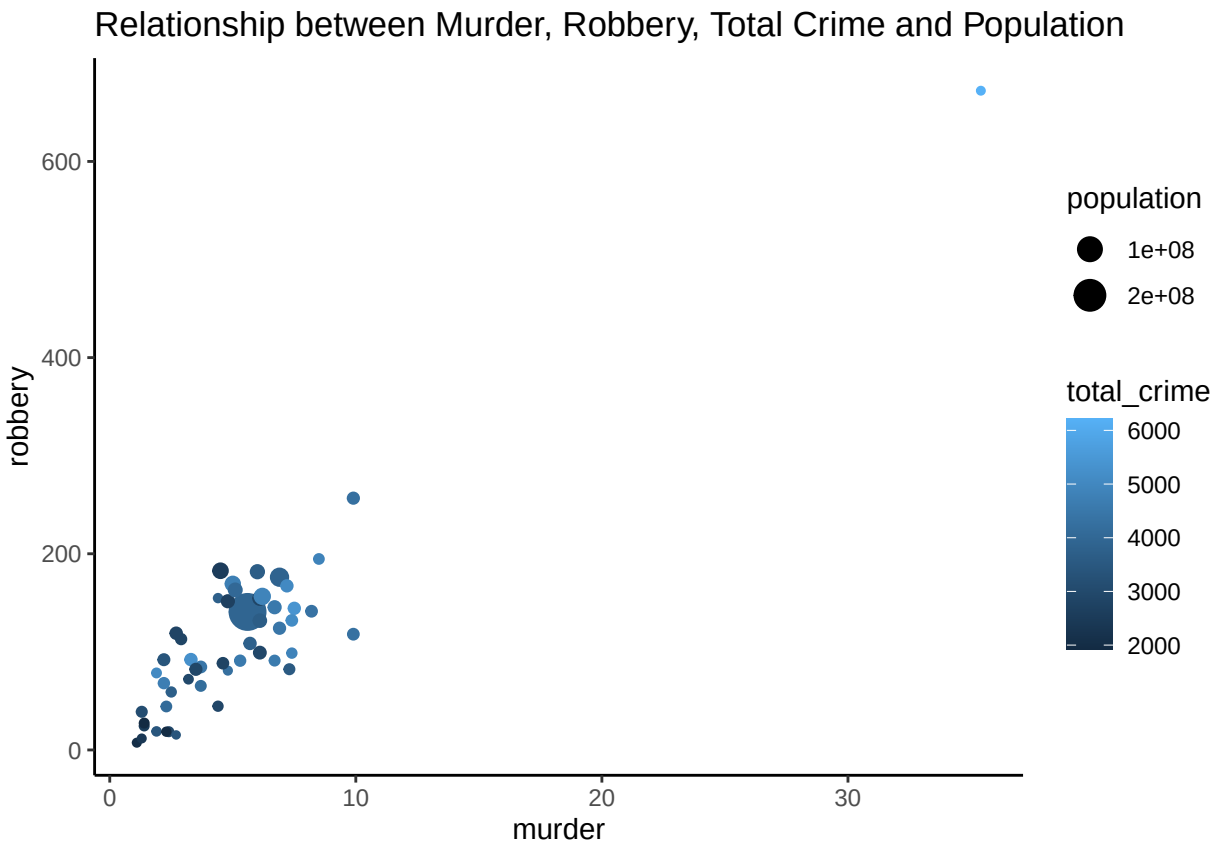
```
##               country year expectancy
## 1         Afghanistan 2008         42
## 2             Albania 2008         73
## 3             Algeria 2008         71
## 4              Angola 2008         46
```

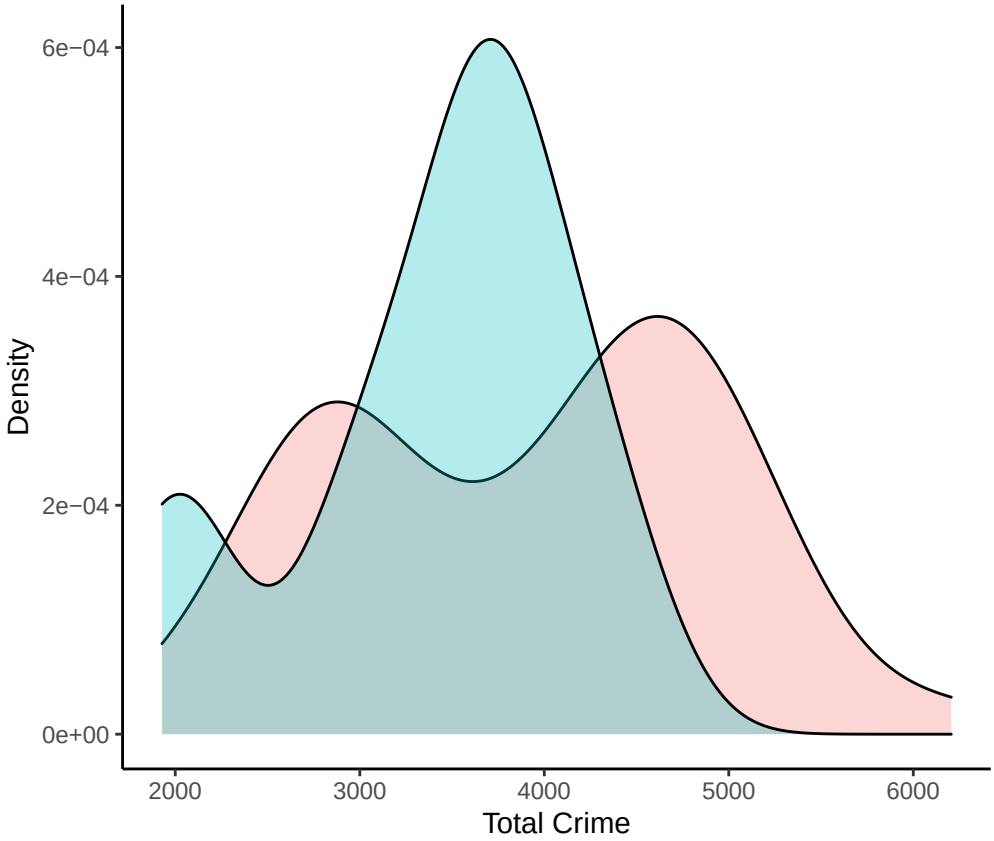
```
## 5 Antigua and Barbuda 2008         74
## 6           Argentina 2008         76
```

Plot1: Scatter Plot



Plot2: Bubble Plot





Plot3: Density Plot

You need to submit 3 tree maps, 3 area charts and 3 stacked area 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.

Python

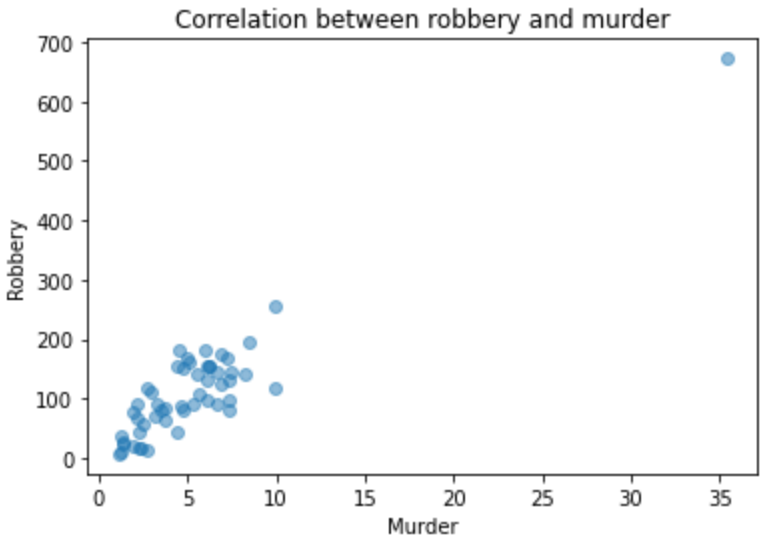
In [1]:
```
# Import libraries
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import chart_studio.plotly as py
import cufflinks as cf
import seaborn as sns
```

In [2]:
```
# Read world population data
dirData = 'ex4-2/'
f_crime = 'crimerates-by-state-2005.csv'

dir_crime = dirData+f_crime
crime = pd.read_csv(dir_crime)
crime_state = crime[crime['state']!='United States']

print(crime_state.head())
```

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

   larceny_theft  motor_vehicle_theft  population
1         2650.0                288.3     4545049
2         2599.1                391.0      669488
3         2965.2                924.4     5974834
4         2711.2                262.1     2776221
5         1916.5                712.8    35795255
```

1. Scatter plot

In [3]:
```
# Create a scatter plot showing correlation between murder and robbery

plt.scatter(x=crime['murder'], y=crime['robbery'],alpha=0.5)
plt.title('Correlation between robbery and murder')
plt.xlabel('Murder')
plt.ylabel('Robbery')
plt.show()
```

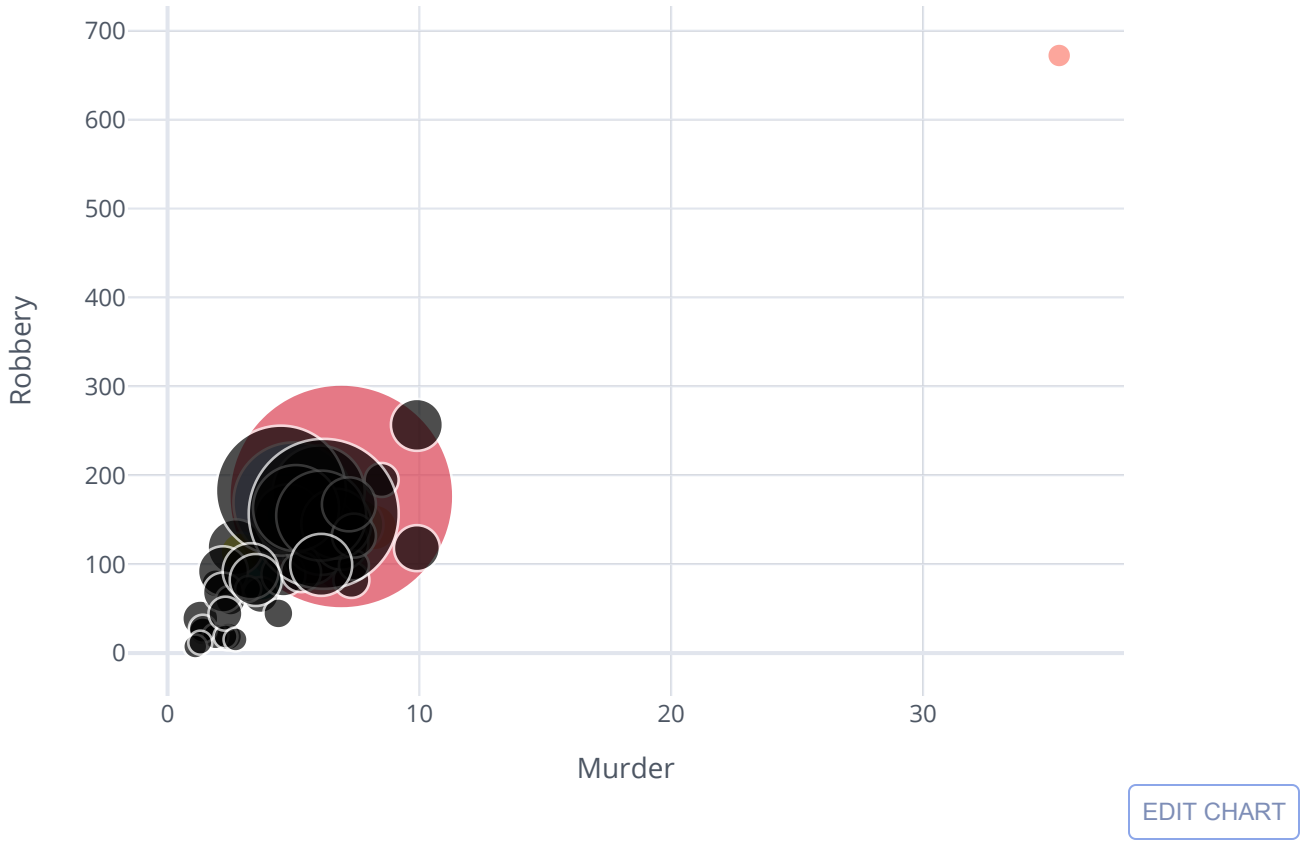


2. Bubble Chart

In [4]:

```
# For this excercise I used the API service from plotly (https://plot.ly/)

py.plotly.tools.set_credentials_file(username='vkoppul', api_key='HZkq8JtTPj1Q4RbEm6eV')
cf.set_config_file(offline=False, world_readable=True, theme='pearl')

crime_state.iplot(kind='bubble', x='murder', y='robbery', size='population', text='state',
             xTitle='Murder', yTitle='Robbery',
             filename='Murder vs Robbery by Population')
```

Out[4]:

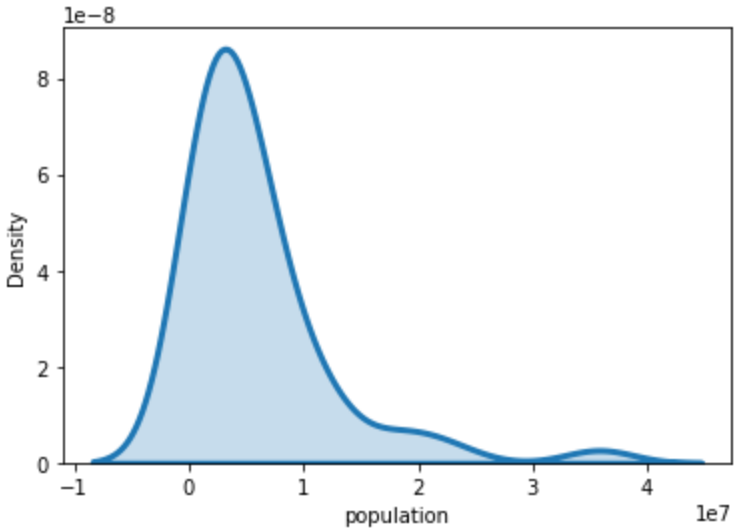


3. Density plot

In [5]:
```
# Plotting distribution of population

sns.distplot(crime_state['population'], hist = False, kde = True,
                 kde_kws = {'shade': True, 'linewidth': 3})
```

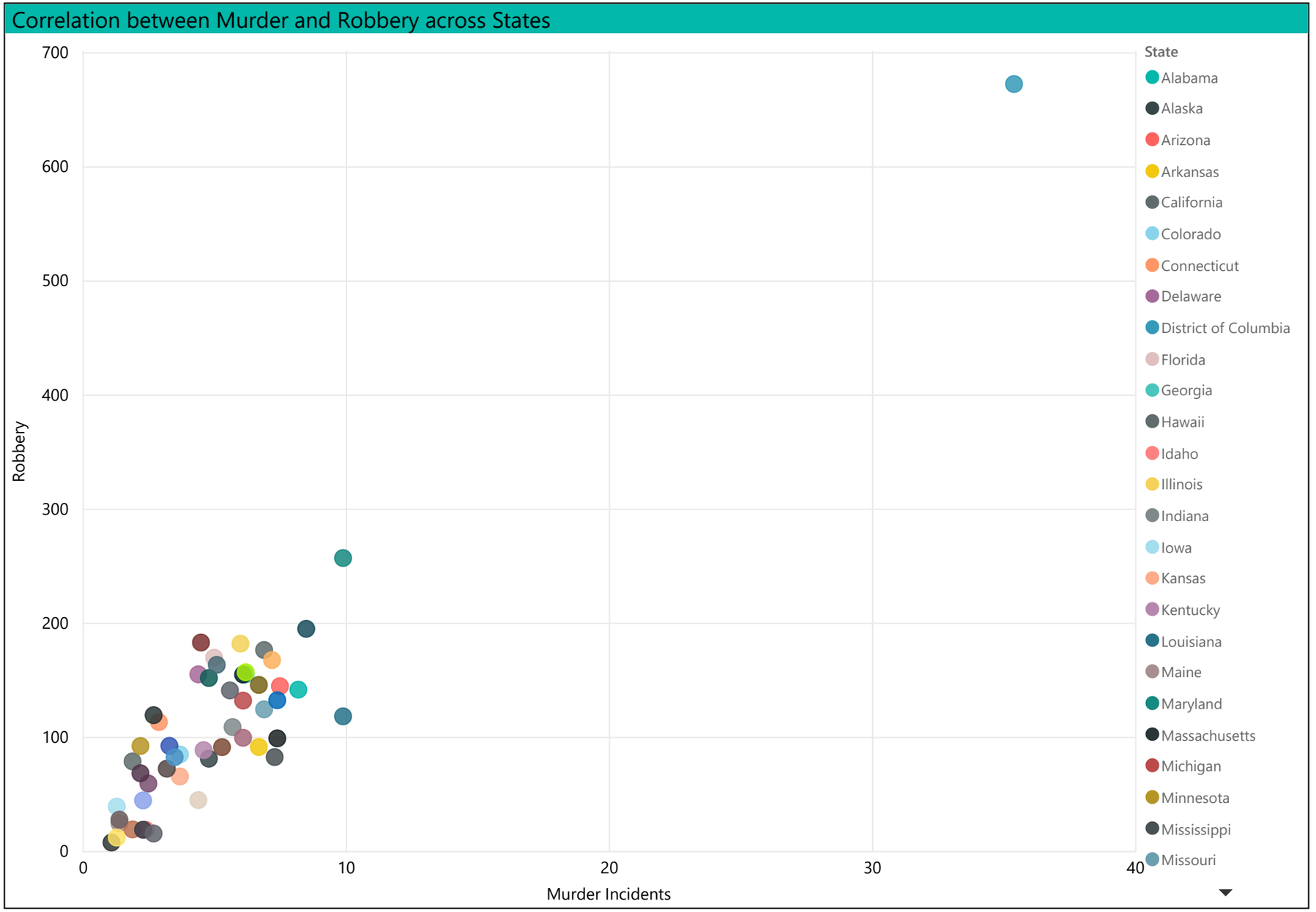
```
/Users/veerareddykoppula/opt/anaconda3/lib/python3.9/site-packages/seaborn/distributions.py:2619: FutureWarning:

`distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `kdeplot` (an axes-level function for kernel density plots).
```

Out[5]:
```
<AxesSubplot:xlabel='population', ylabel='Density'>
```

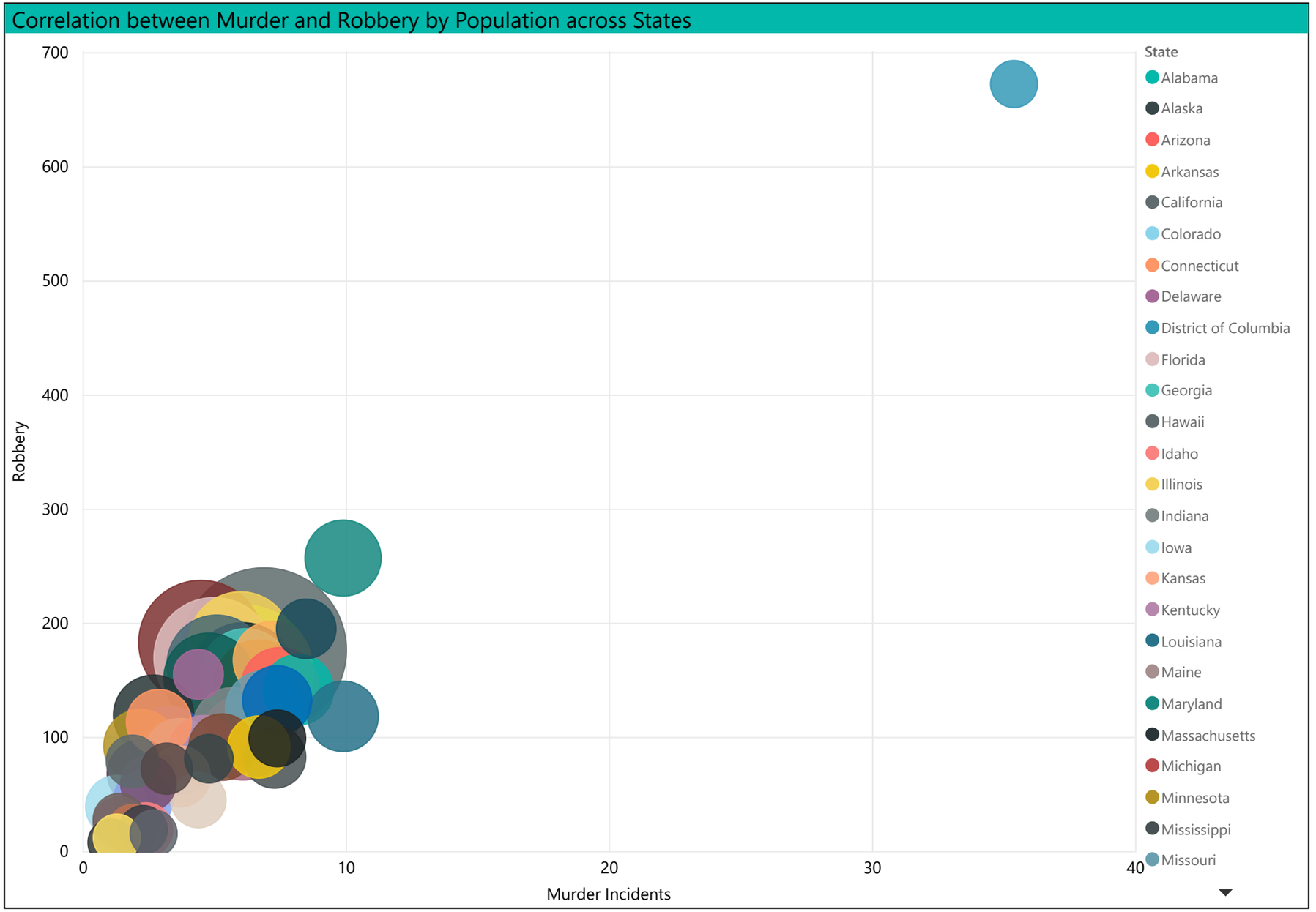


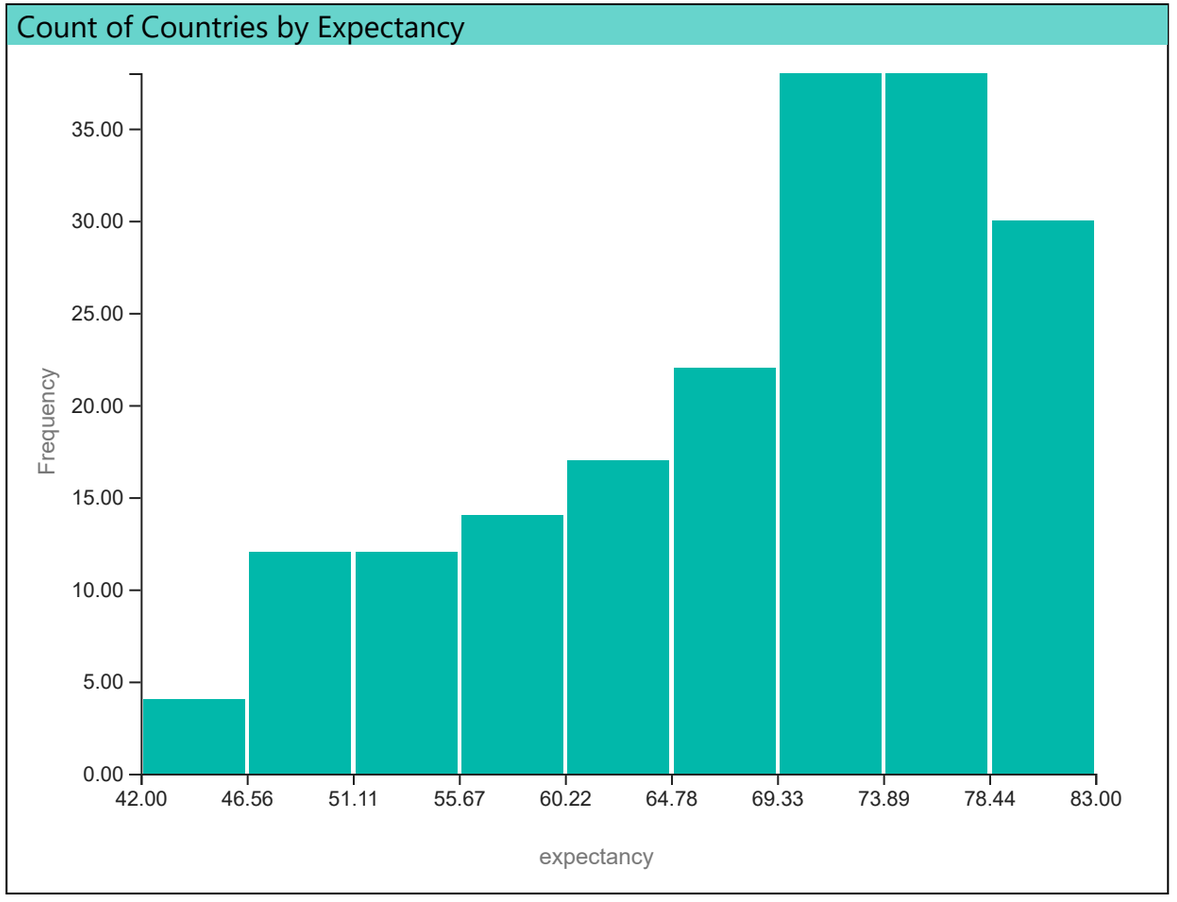
End of code

Find Whether there is any correlation between murder as a result of robbery across states



Find Whether there is any correlation between murder as a result of robbery across states and population




Count of Countries by Expectancy
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
Frequency
42.00
46.56
51.11
55.67
60.22
64.78
69.33
73.89
78.44
83.00
expectancy

External data source is used for State name and state Code mapping