DCS 640 Data Presentation & Visualization (DSC640-T302 2231-1)

Bellevue University

4.2 Exercises: Scatterplots, Bubble Charts, and Density Plots

Author: Jake Meyer

Date: 01/28/2023

Assignment Instructions:

Submit 1 scatterplot, 1 bubble chart, and 1 density plot with Python

In [2]:
```
'''
Import the necessary libraries to complete Exercise 2.2.
'''
import numpy as np
import pandas as pd
import seaborn as sns
import scipy.stats
import matplotlib
import matplotlib.pyplot as plt
import matplotlib.patches as mpatches
import plotly.express as px
```

In [3]:
```
'''
Check the versions of the packages.
'''
print('numpy version:', np.__version__)
print('pandas version:', pd.__version__)
print('seaborn version:', sns.__version__)
print('matplotlib version:', matplotlib.__version__)
```

```
numpy version: 1.20.3
pandas version: 1.3.4
seaborn version: 0.11.2
matplotlib version: 3.4.3
```

Dataset Understanding

In [4]:
```
'''
Import the datasets.
Note: A copy of the CSV file was placed into the same directory as this notebook.
Utilize pd.read_csv() to read the file as a pandas data frame.
```

```
'''
df1 = pd.read_csv('crimerates-by-state-2005.csv')
```

In [5]:
```
'''
Use head() function to display the first 5 rows of data of df1.
'''
df1.head()
```

Out[5]:

	state	murder	forcible_rape	robbery	aggravated_assault	burglary	larceny_theft	motor_vehicle_t
0	United States	5.6	31.7	140.7	291.1	726.7	2286.3	4
1	Alabama	8.2	34.3	141.4	247.8	953.8	2650.0	2
2	Alaska	4.8	81.1	80.9	465.1	622.5	2599.1	3
3	Arizona	7.5	33.8	144.4	327.4	948.4	2965.2	9
4	Arkansas	6.7	42.9	91.1	386.8	1084.6	2711.2	2

In [6]:
```
'''
Understand the shape of the df1.
'''
print('There are {} rows and {} columns in the df1.'.format(df1.shape[0], df1.shape[1]))
```

```
There are 52 rows and 9 columns in the df1.
```

In [7]:
```
'''
Find the type of data within each df1 column initially.
'''
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 52 entries, 0 to 51
Data columns (total 9 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   state                52 non-null     object
 1   murder               52 non-null     float64
 2   forcible_rape        52 non-null     float64
 3   robbery              52 non-null     float64
 4   aggravated_assault   52 non-null     float64
 5   burglary             52 non-null     float64
 6   larceny_theft        52 non-null     float64
 7   motor_vehicle_theft  52 non-null     float64
 8   population           52 non-null     int64
dtypes: float64(7), int64(1), object(1)
memory usage: 3.8+ KB
```

In [8]:
```
'''
Understand if there are any missing values in df1.
'''
df1.isna().sum().sort_values(ascending = False)
```

Out[8]:
```
state                 0
murder                0
forcible_rape         0
robbery               0
aggravated_assault    0
burglary              0
larceny_theft         0
motor_vehicle_theft   0
population            0
dtype: int64
```

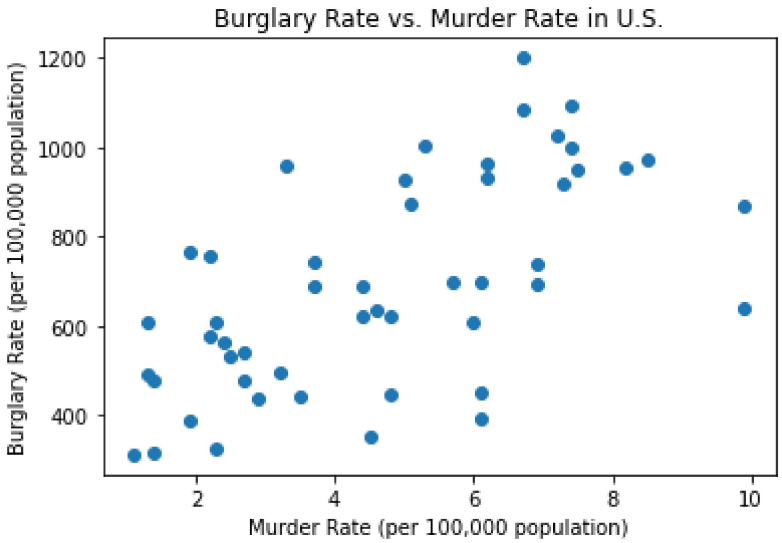
Chart Creation from the Dataset.

Scatterplot

In [13]:
```
'''
Drop the District of Columbia and United States rows within the Data Frame.
'''
df1.drop(df1[df1['state'] == "District of Columbia"].index, inplace = True)
df1.drop(df1[df1['state'] == "United States"].index, inplace = True)
```

In [20]:
```
'''
Create a scatterplot with matplotlib.
'''
plt.scatter(df1['murder'], df1['burglary'])
plt.xlabel('Murder Rate (per 100,000 population)')
plt.ylabel('Burglary Rate (per 100,000 population)')
plt.title('Burglary Rate vs. Murder Rate in U.S.')
plt.show
```

Out[20]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



Bubble Chart

In [25]:
```
'''
Create the bubble chart using plotly.
'''
```

```
fig = px.scatter(df1, x = "murder", y = "burglary", color ="state", size = 'population'
                 title = "Murder Rate vs. Burglary Rate by State Population (2005)",
                 labels={
                     "murder":"Murder Rate (per 100,000 population)",
                     "burglary": "Burglary Rate (per 100,000 population)"
                 })
fig.show()
```

Density Plot

In [30]:
```
'''
Read the life expectancy data in from the csv file.
'''
df2 = pd.read_csv('life-expectancy.csv')
```

In [36]:
```
'''
Create a density plot using seaborn and matplotlib.
'''
plt.figure(figsize = (5, 3))
sns.displot(data = df2, x='expectancy', kind = "kde")
plt.xlabel('Life Expectancy')
plt.ylabel('Density')
plt.title("World Life Expectancy (2008)")
plt.show()
```

<Figure size 360x216 with 0 Axes>