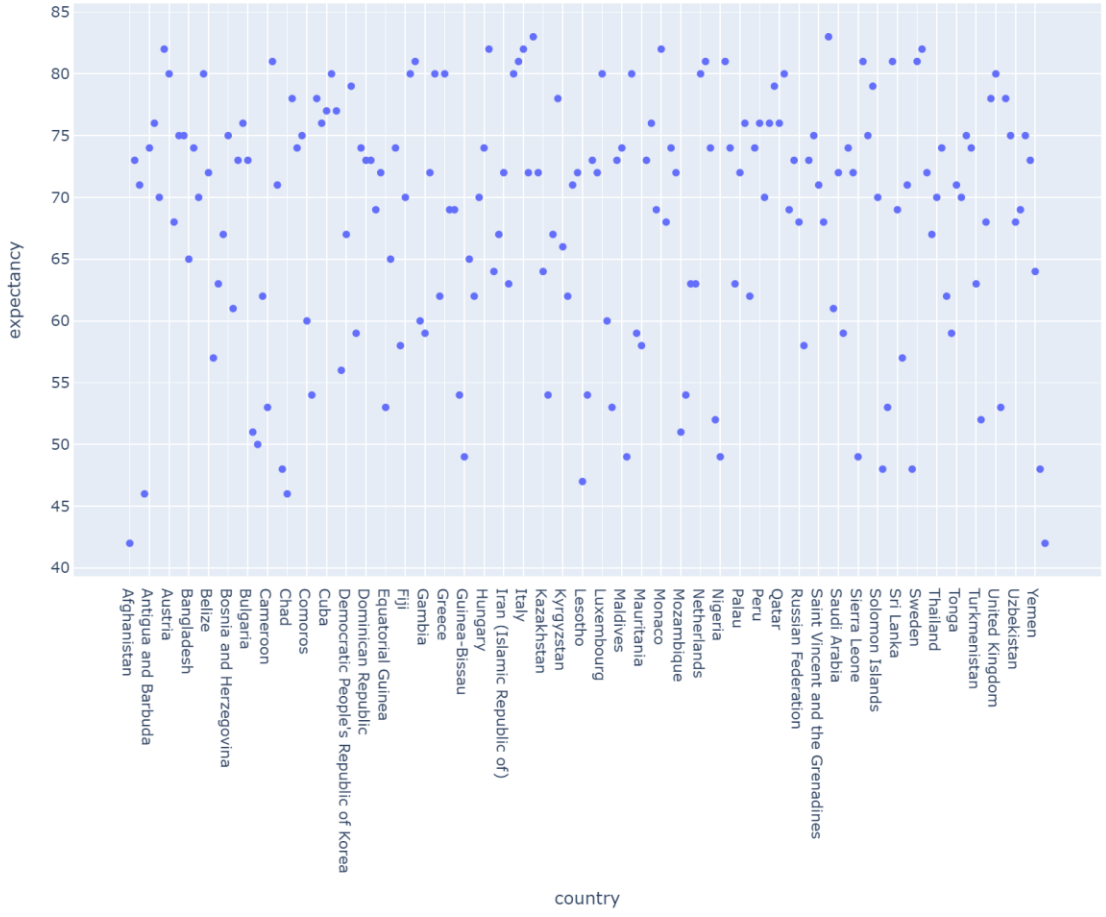
Scatterplot – Python

```
# draw a scatterplot of life expectancy for each country
fig = px.scatter(life, x='country', y='expectancy',
                 width=990, height=800)
fig.show()
```



Background Information:

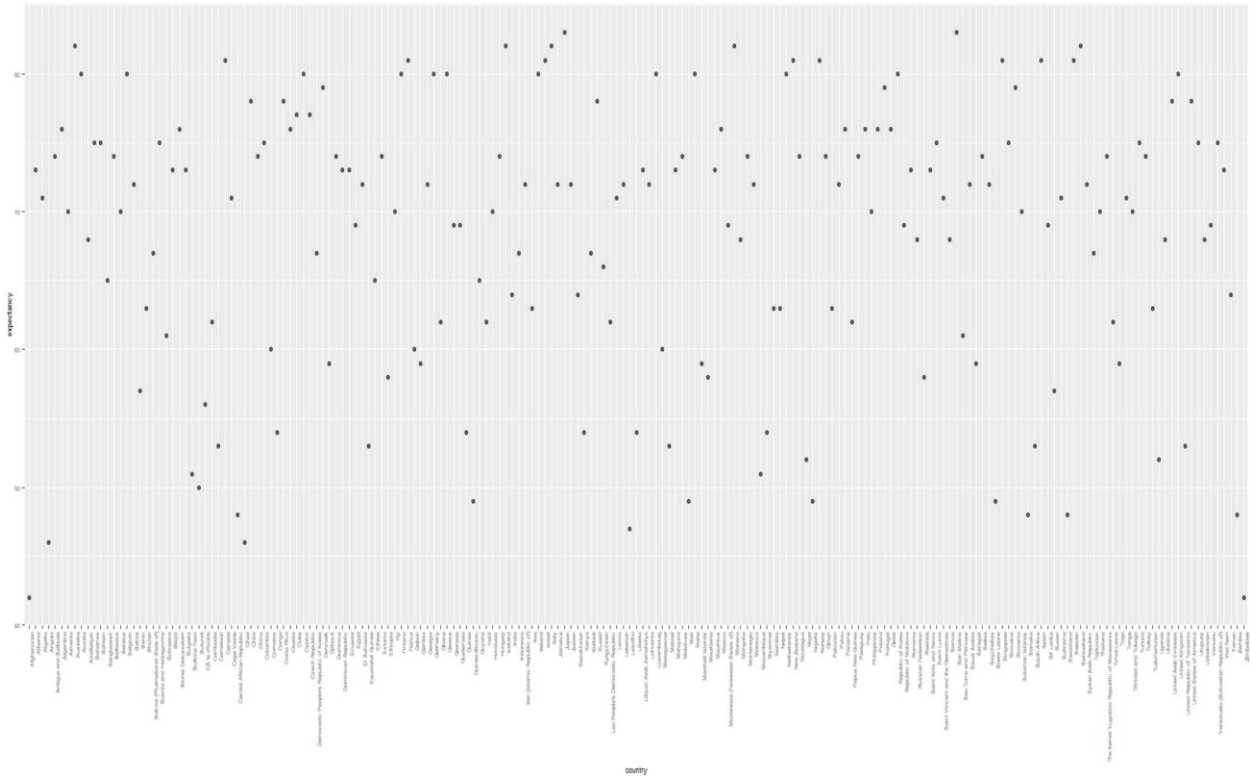
A scatterplot is created by using Python – Plotly. The data set is the life-expectancy.scv provided by the instructor. The countries are plotted on the x-axis. The life expectancy for each country is plotted on the y-axis. From this graph, Japan and San Mario show the highest life expectancy at 83. Afghanistan and Zimbabwe show the lowest life expectancy at 42. Plotly provides an interactive map, which allows the users to hover their mouse to see the value of each dot.

Method:

- Load the libraries and load the dataset into a data frame
- Plot the country as x-variable and the values of life expectancy as y-variable
- Use the scatter() function in the plotly package to graph the scatterplot

Scatter Plot – R

```
4  library(tidyverse)
5
6  ggplot(life, aes(x=country, y=expectancy)) +
7    geom_point(alpha=0.6) +
8    theme(axis.text.x = element_text(angle = 90, hjust = 1), text = element_text(size = 7))
9
```



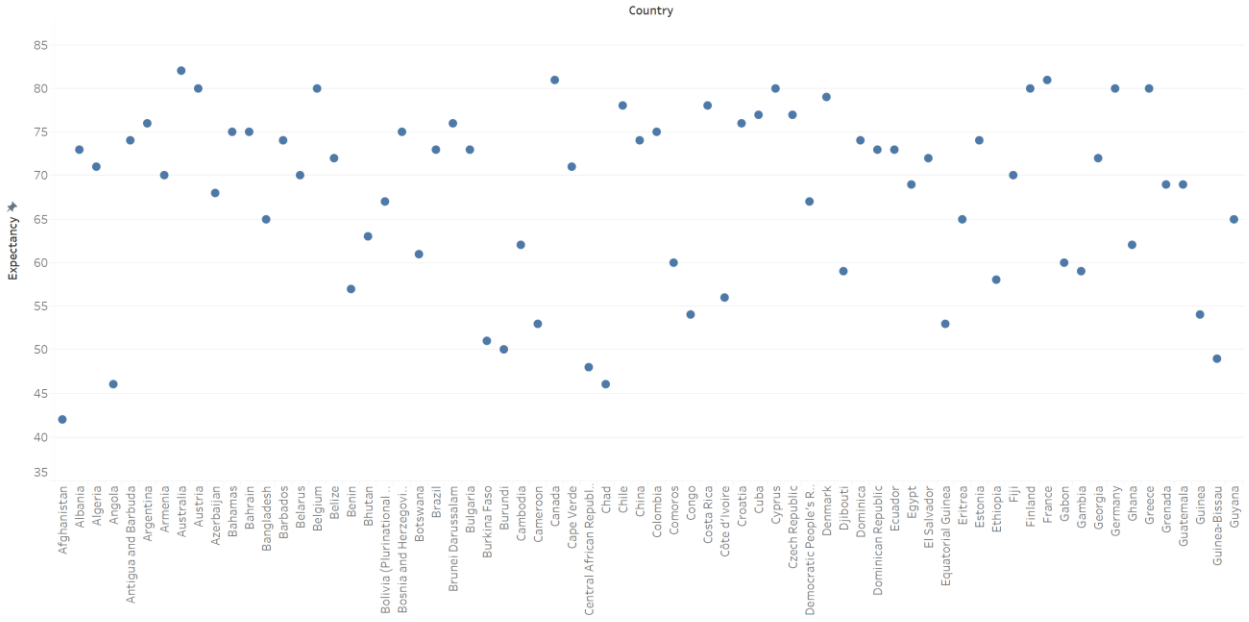
Background Information:

A scatterplot is created by using R – Tidyverse. The data set is the life-expectancy.scv provided by the instructor. The countries are plotted on the x-axis. The life expectancy for each country is plotted on the y-axis. From this graph, Japan and San Mario show the highest life expectancy at 83. Afghanistan and Zimbabwe show the lowest life expectancy at 42.

Method:

- Load the library (Tidyverse) and load the dataset into a data frame
- Plot the country as x-variable and the values of life expectancy as y-variable
- Use the ggplot() & geom_point() function in to graph the scatterplot

Scatter Plot – Tableau



Background Information:

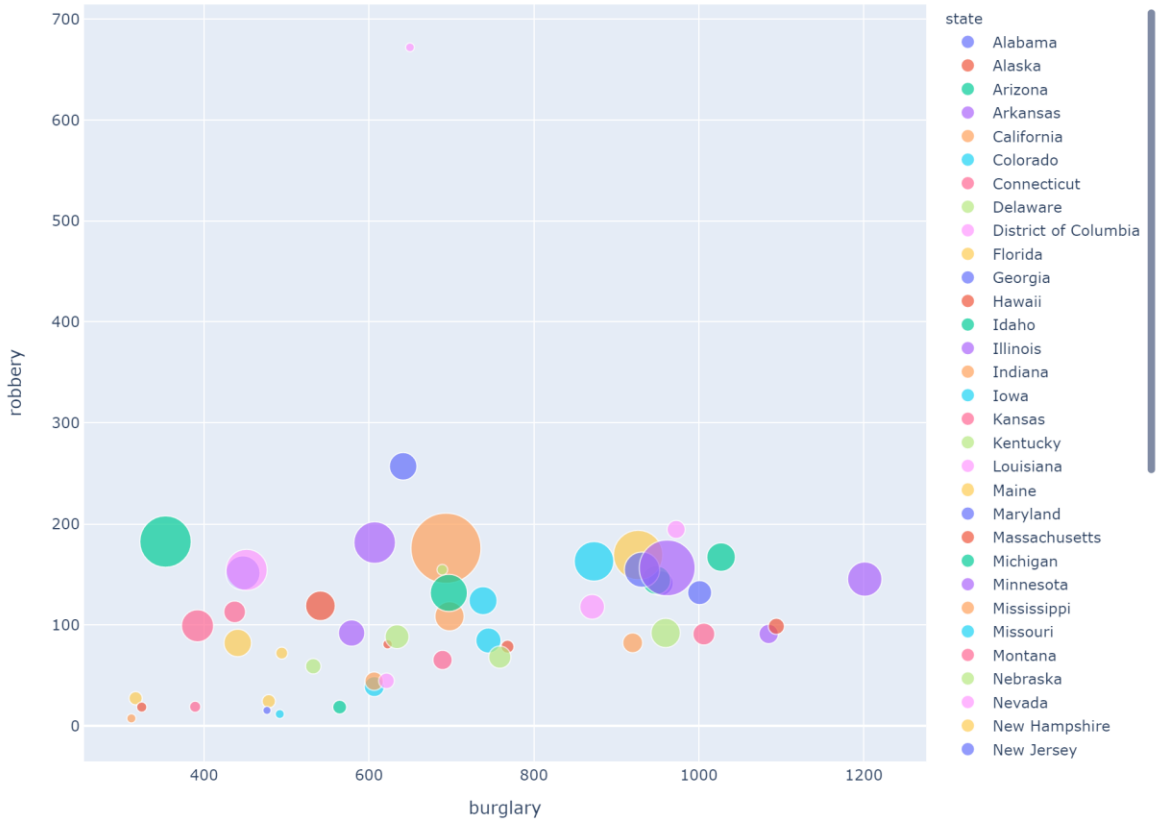
A scatterplot is created by using Tableau. The data set is the life-expectancy.scv provided by the instructor. The countries are plotted on the x-axis. The life expectancy for each country is plotted on the y-axis. From this graph, Japan and San Mario show the highest life expectancy at 83. Afghanistan and Zimbabwe show the lowest life expectancy at 42.

Method:

- Load the data from the datasource
- Place the Country from the Data Table into the Column of Sheet 1
- Place the Sum(Expectancy) from the Data Table into the Row of Sheet 1
- On the Marks box, select Circle

Bubble Chart – Python

```
# plot a bubble chart for 50 US states, which includes burglary, robbery and population
fig = px.scatter(crime, x='burglary', y='robbery', color='state', size='population',
                 size_max=40, width=950, height=750)
fig.show()
```



Background Information:

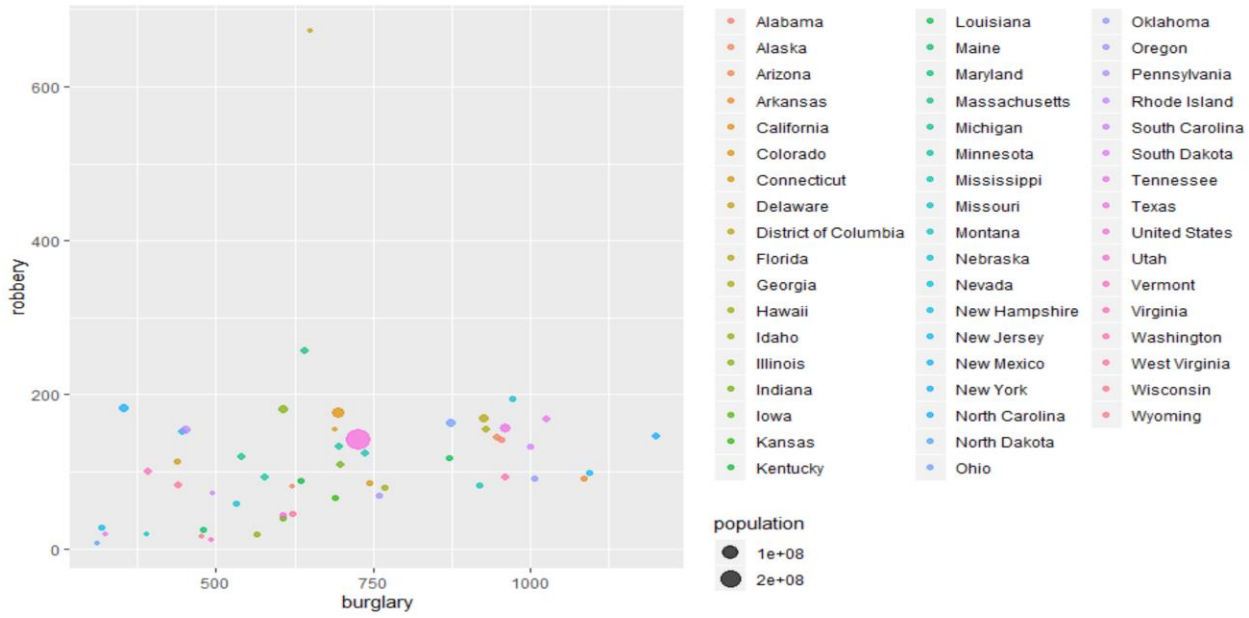
A bubble chart is created by using Python – Plotly. The data set is the crimerates-by-state-2005.csv provided by the instructor. The burglary for each state is plotted on the x-axis. The robbery for each state is plotted on the y-axis. The color of the bubble represents the state. The size of the bubble represents the population of the state. As shown on the chart, there is a small pink bubble at the top, which represents the highest robbery at District of Columbia. The state with highest number of burglary is North Carolina.

Method:

- Load the libraries and load the dataset into a data frame
- Plot the burglary as x-variable, the robbery as y-variable, the state as the color and the population as the size of the bubble
- Use the scatter() function in the plotly package to graph the bubble chart

Bubble Chart – R

```
12  ggplot(crime, aes(x=burglary, y=robbery, size=population, col=state)) +
13    geom_point(alpha=0.7)
14
```



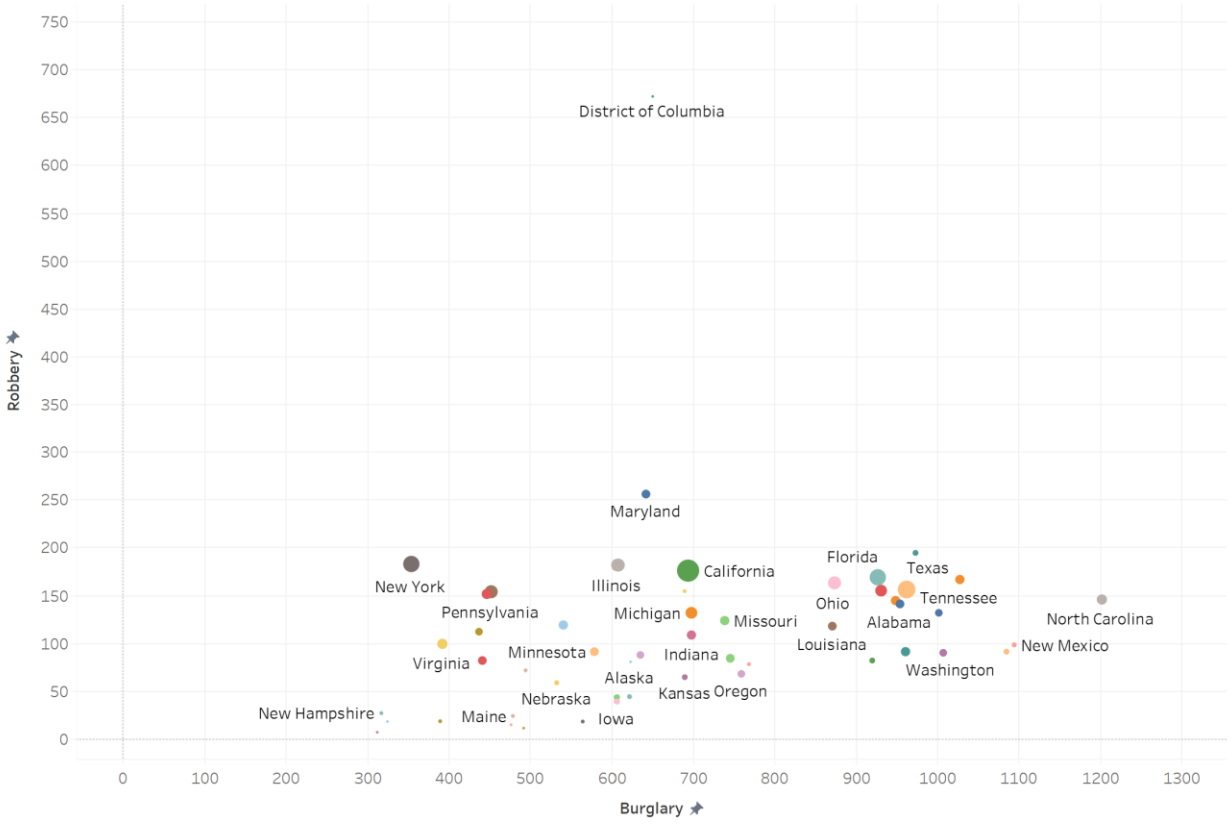
Background Information:

A bubble chart is created by using R – Tidyverse. The data set is the crimerates-by-state-2005.csv provided by the instructor. The burglary for each state is plotted on the x-axis. The robbery for each state is plotted on the y-axis. The color of the bubble represents the state. The size of the bubble represents the population of the state. As shown on the chart, there is a small pink bubble at the top, which represents the highest robbery at District of Columbia. The state with highest number of burglary is North Carolina.

Method:

- Load the libraries and load the dataset into a data frame
- Plot the burglary as x-variable, the robbery as y-variable, the state as the color and the population as the size of the bubble
- Use the ggplot() + geom_point() function in the ggplot2 package to graph the bubble chart

Bubble Chart – Tableau



Background Information:

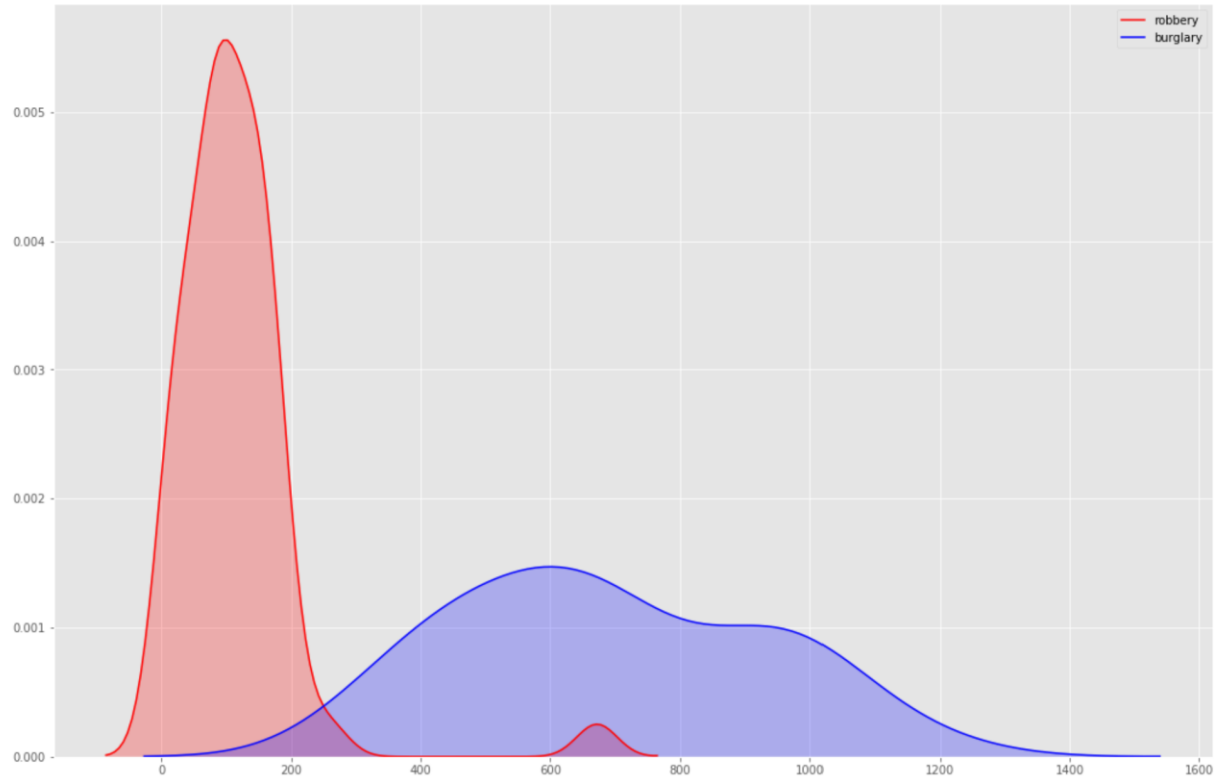
A bubble chart is created by using Tableau. The data set is the crimerates-by-state-2005.csv provided by the instructor. The burglary for each state is plotted on the x-axis. The robbery for each state is plotted on the y-axis. The color of the bubble represents the state. The size of the bubble represents the population of the state. As shown on the chart, there is a small pink bubble at the top, which represents the highest robbery at District of Columbia. The state with highest number of burglary is North Carolina.

Method:

- Load the data from the datasource
- Place the Burglary from the Data Table into the Column of Sheet 2
- Place the Robbery from the Data Table into the Row of Sheet 2
- Place the State from the Data Table to the Color on the Marks
- Place the Population from the Data Table to the Size on the Marks
- Place the State from the Data Table to the Label on the Marks

Density Plot – Python

```
#  plot 2 variables: murder and burglary
plt.style.use('ggplot')
plt.figure(figsize=(18,12))
p1 = sns.kdeplot(crime['robbery'], shade=True, color='r')
p2 = sns.kdeplot(crime['burglary'], shade=True, color='b')
plt.show()
```



Background Information:

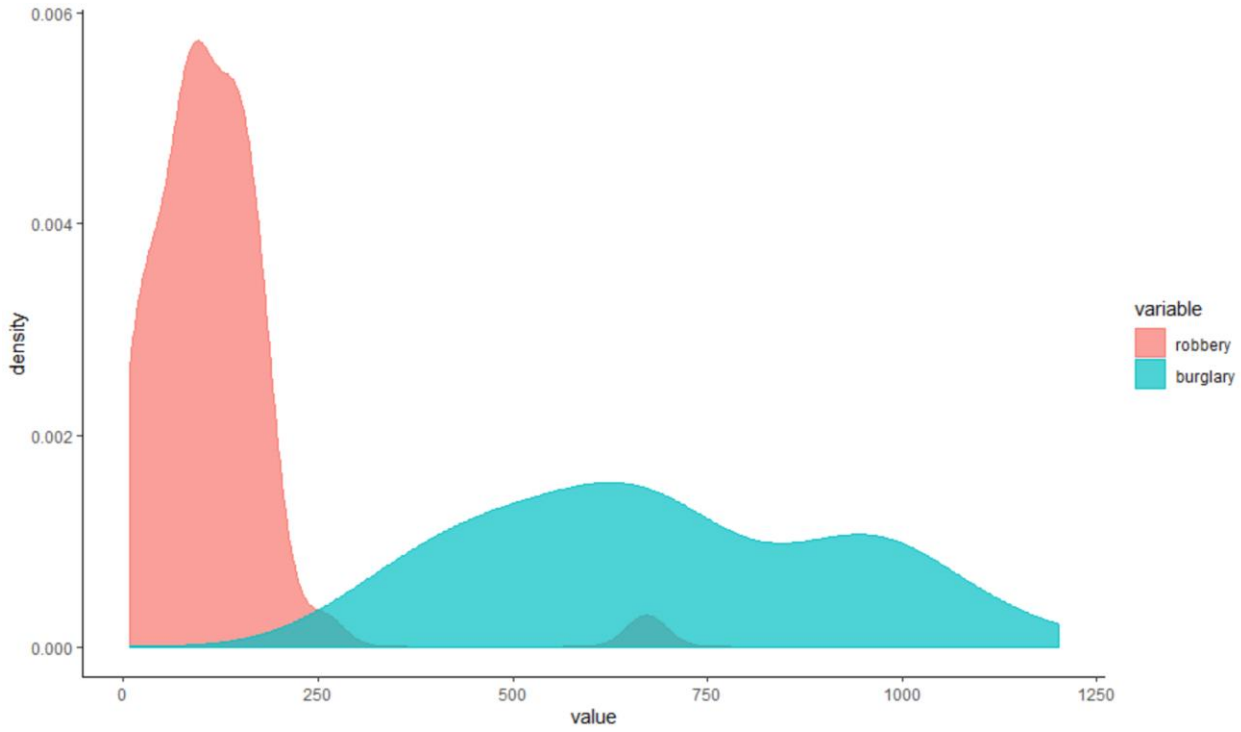
A density plot is created by using Python – Seaborn. The data set is the crimerates-by-state-2005.csv provided by the instructor. This density plot includes two variables: robbery (red) and burglary (blue). The density plot shows the probability distribution function of robbery and burglary that occurs in 50 US states in 2005. The x-axis represents the number of cases. The y-axis represents the probability. As shown in the graph, the area and the peak of the robbery focuses on a smaller number of cases among different states. The number of burglary cases is more spread-out for different states.

Method:

- Load the libraries and load the dataset into a data frame
- Use the kdeplot() function of the seaborn package
- Draw the density plot for robbery with shade = True and set the color to red
- Draw the density plot for burglary with shade = True and set the color to blue

Density Plot – R

```
17 robbery <- crime$robbery
18 burglary <- crime$burglary
19
20 df <- data.frame(robbery, burglary)
21
22 # plot densities
23 library(reshape2)
24
25 # use melt() to break down the components
26 df.plot <- melt(df)
27
28 # plot
29 p <- ggplot(aes(x=value, colour=variable, fill=variable), data=df.plot)
30 p + geom_density(alpha=0.7)
31
```



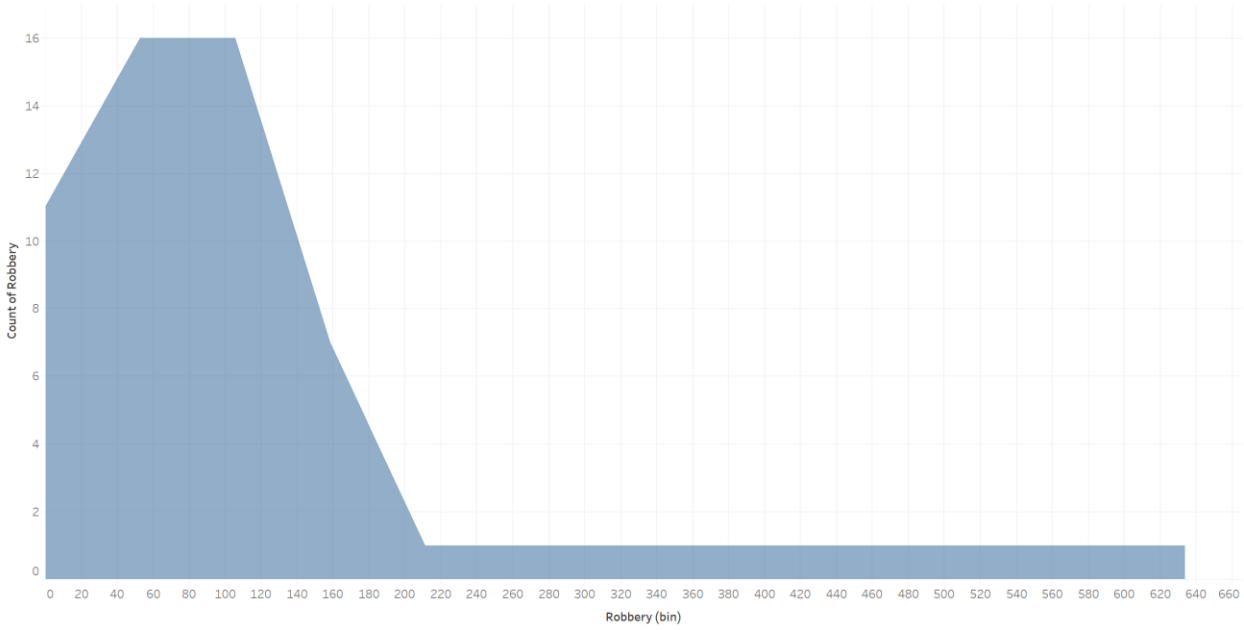
Background Information:

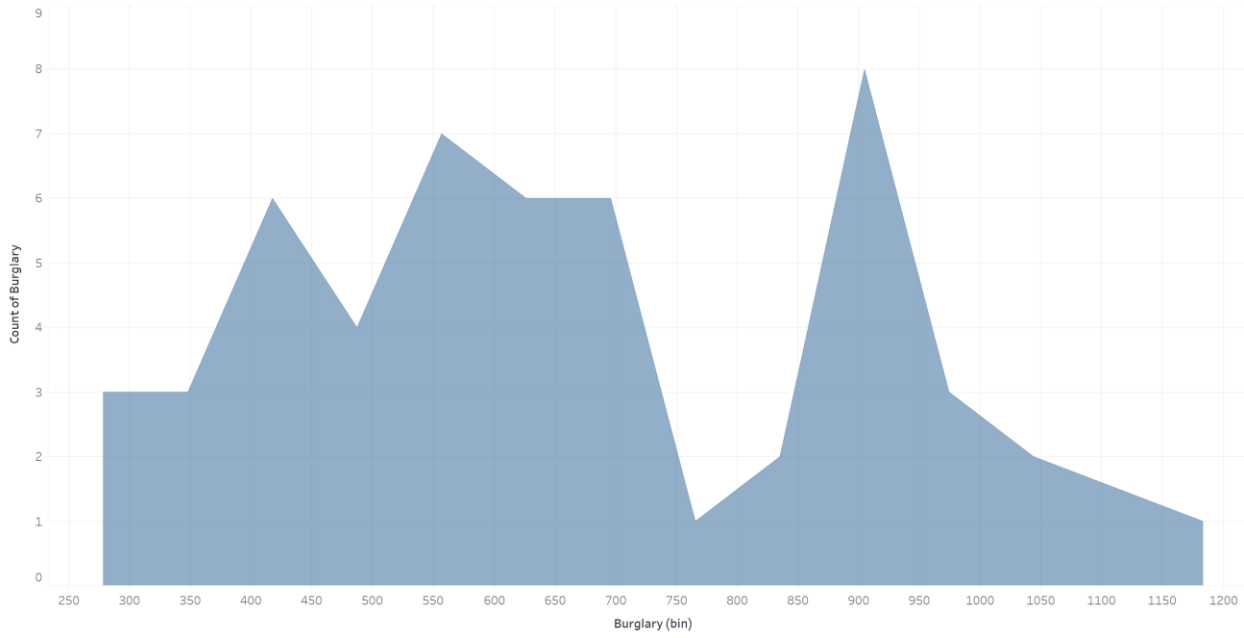
A density plot is created by using R – ggplot. The data set is the crimerates-by-state-2005.csv provided by the instructor. This density plot includes two variables: robbery (red) and burglary (teal). The density plot shows the probability distribution function of robbery and burglary that occurs in 50 US states in 2005. The x-axis represents the number of cases. The y-axis represents the probability. As shown in the graph, the area and the peak of the robbery focuses on a smaller number of cases among different states. The number of burglary cases is more spread-out for different states.

Method:

- Load the libraries and load the dataset into a data frame
- Select the interested variables from the dataframe to form a new dataframe
- Use reshape2 package and the melt() function to break down the dataframe into values and varaibles
- Use the geom_density() function of the ggplot2 package to draw the density plot

Density Plot – Tableau





Background Information:

Because both Tableau and Power BI lacks the ability to directly draw a kernel density plot, so I have to use Histogram to draw each of them in Tableau. Kernel density plot is basically a variation of histogram, which uses Kernel smoothing to plot values and smooths out the noise. The data set is the crimerates-by-state-2005.csv provided by the instructor. The burglary and the robbery for each state is plotted on the x-axis. The count is plotted on the y-axis, for kernel density plot, the count is converted to the probability. Method:

- Load the data from the datasource
- Place the Burglary from the Data Table into the Column of Sheet 2
- On the Show Me tab, choose Histogram
- On the Marks tab, choose Area.
- Place the Robbery from the Data Table into the Column of Sheet 3
- On the Show Me tab, choose Histogram
- On the Marks tab, choose Area.