Week 5 & 6 Assignment - Python

Name: Kesav Adithya Venkidusamy
Course: DSC640 - Data Presentation and Visualization
Instructor: Catherine Williams

These two weeks we are going to be focused on scatterplots, bubble charts, and density plots and using various tools to create these visualizations. You must consolidate all the charts into ONE document with each chart labeled with the type of chart and technology - for example: Python - Bar Chart. Failure to label and consolidate the charts will resort in points being taken off or a 0 for the assignment.

Sample Datasets (click on the Downloads tab.)

You may also download them directly from this link: Exercise 4.2 Datasets (click the link to download a folder containing the datasets.)

You need to submit:
1 scatterplot, 1 bubble chart and 1 density map using Tableau or PowerBI
1 scatterplot, 1 bubble chart and 1 density plot chart using Python
1 scatterplot, 1 bubble chart and 1 density plot chart using R

1 scatterplot, 1 bubble chart and 1 density plot chart using Python

In [49]:
```
## Importing libraries required for this exercise
import pandas as pd
import numpy as np
import squarify
import matplotlib.pyplot as plt
%matplotlib inline
import plotly.express as px
import seaborn as sns
import plotly.io as pio
pio.renderers.default='notebook'
```

1. Python - Scatterplot Chart

Plotting Scatterplot chart for Crime rates by State for 2005

In [12]:
```
## Load the dataset into dataframe
crime_df = pd.read_csv("crimerates-by-state-2005.csv")
crime_df.head()
```

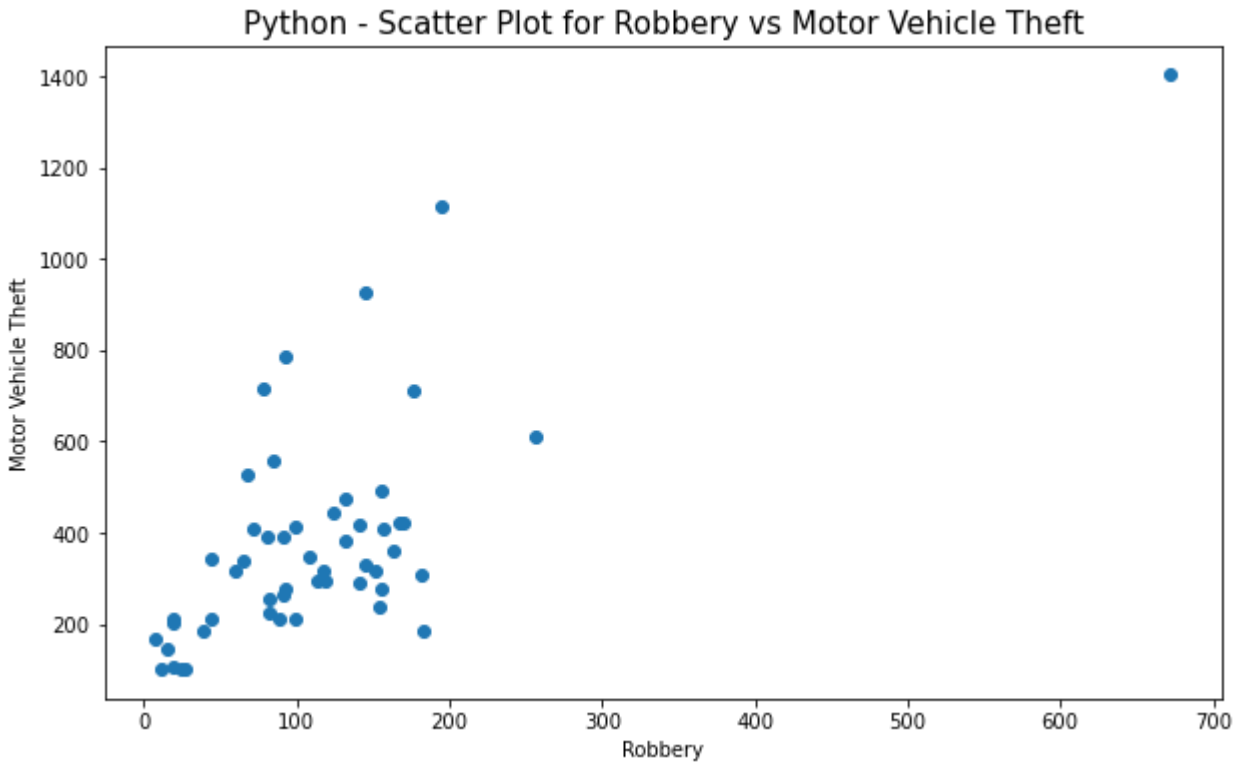
Out[12]:

	state	murder	forcible_rape	robbery	aggravated_assault	burglary	larceny_theft	motor_vehicle_
0	United States	5.6	31.7	140.7	291.1	726.7	2286.3	
1	Alabama	8.2	34.3	141.4	247.8	953.8	2650.0	

	state	murder	forcible_rape	robbery	aggravated_assault	burglary	larceny_theft	motor_vehicle_
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	

We will find correlation between Robbery and Motor Vehicle theft

In [19]:
```
plt.figure(figsize=(10,6))
plt.scatter(crime_df.robbery ,crime_df.motor_vehicle_theft, marker ="o")
plt.title("Python - Scatter Plot for Robbery vs Motor Vehicle Theft", fontsize = 15)
plt.xlabel('Robbery')
plt.ylabel('Motor Vehicle Theft')
plt.show()
```



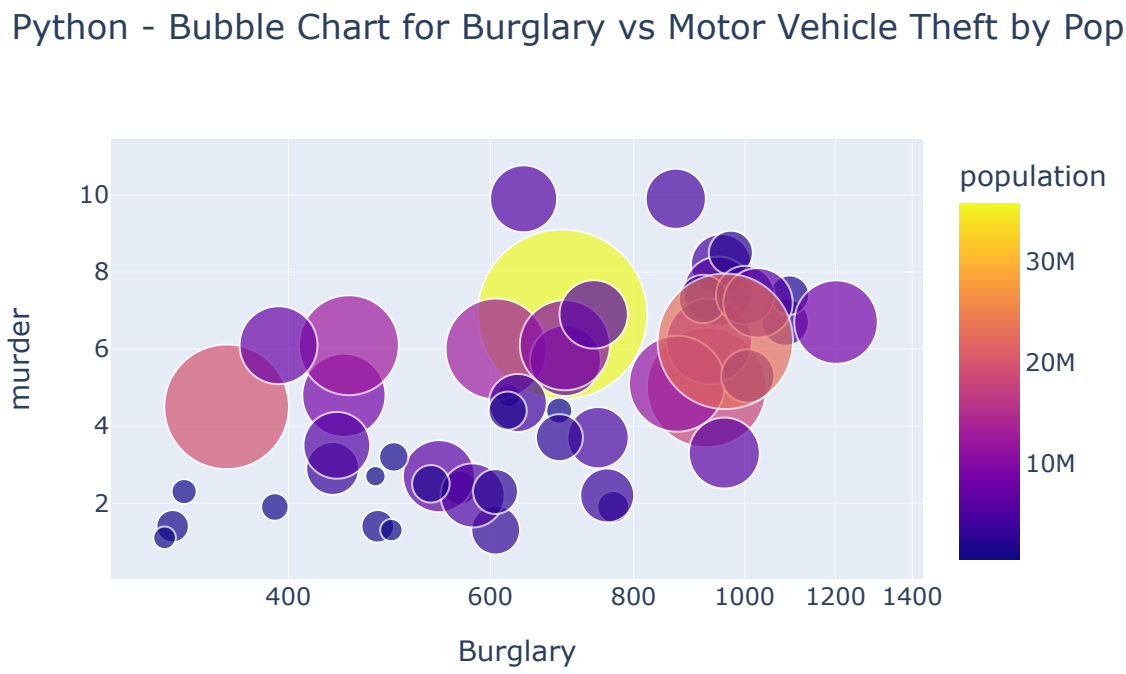
2. Python - Bubbel Plot

Plotting bubble plot for Motor Vehicle Theft by Population

In [31]:
```
## Remving duplicate and outliers from the dataset; So, removing "United States" and "D
crime_state_df = crime_df[crime_df['state'] != 'United States']
crime_state_df = crime_state_df[crime_state_df['state'] != 'District of Columbia']
```

In [50]:
```
## Bubble Plot
fig = px.scatter(crime_state_df, x = 'burglary', y = 'murder', size = 'population', col
                 hover_name = 'state', log_x = True, size_max = 60,
                 labels = {'burglary':'Burglary','motor_vehicle_theft':'Motor Vehicle Th
                 title = 'Python - Bubble Chart for Burglary vs Motor Vehicle Theft by P
```

```
                  width=600, height=400)
fig.show()
```



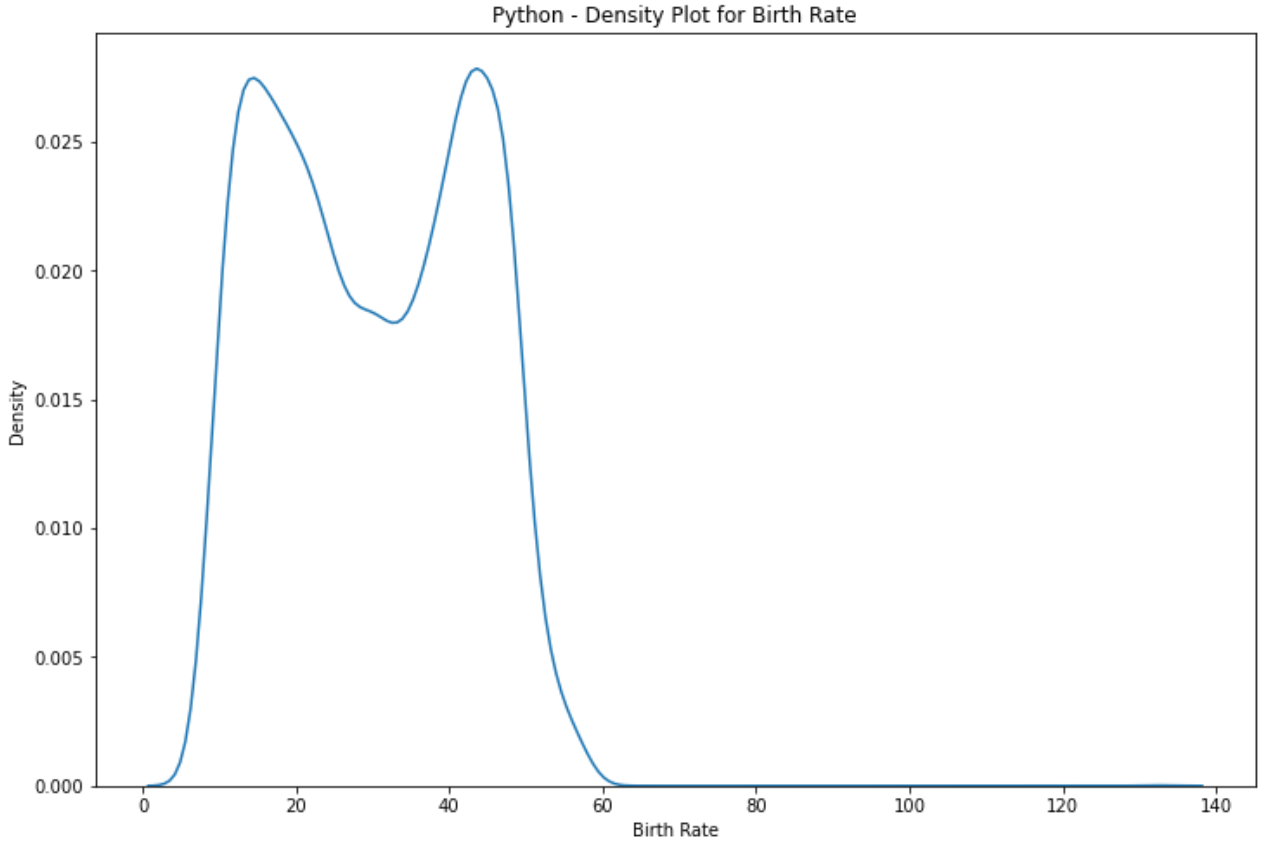
3. Python - Density Plot

In [33]:
```
## Creating the dataframe for birth rate yearly dataset
birth_df = pd.read_csv("birth-rates-yearly.csv")
birth_df.head()
```

Out[33]:

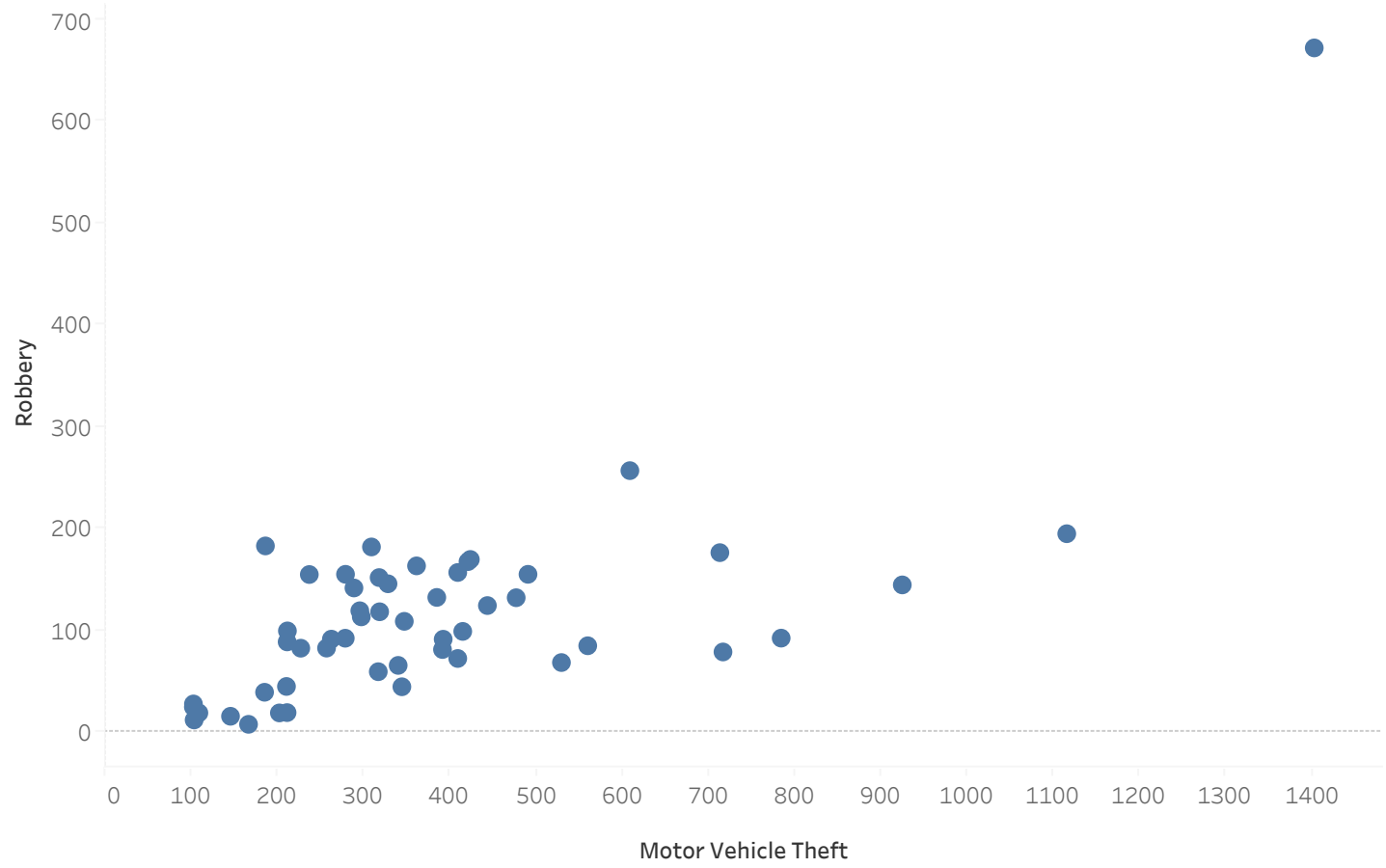
	year	rate
0	1960	36.400
1	1961	35.179
2	1962	33.863
3	1963	32.459
4	1964	30.994

In [35]:
```
## Creating density chart
plt.figure(figsize=(12,8))
sns.kdeplot(birth_df['rate'])
plt.xlabel('Birth Rate')
plt.title("Python - Density Plot for Birth Rate")
plt.show()
```


Python - Density Plot for Birth Rate
Density
0.025
0.020
0.015
0.010
0.005
0.000
0
20
40
60
80
100
120
140
Birth Rate

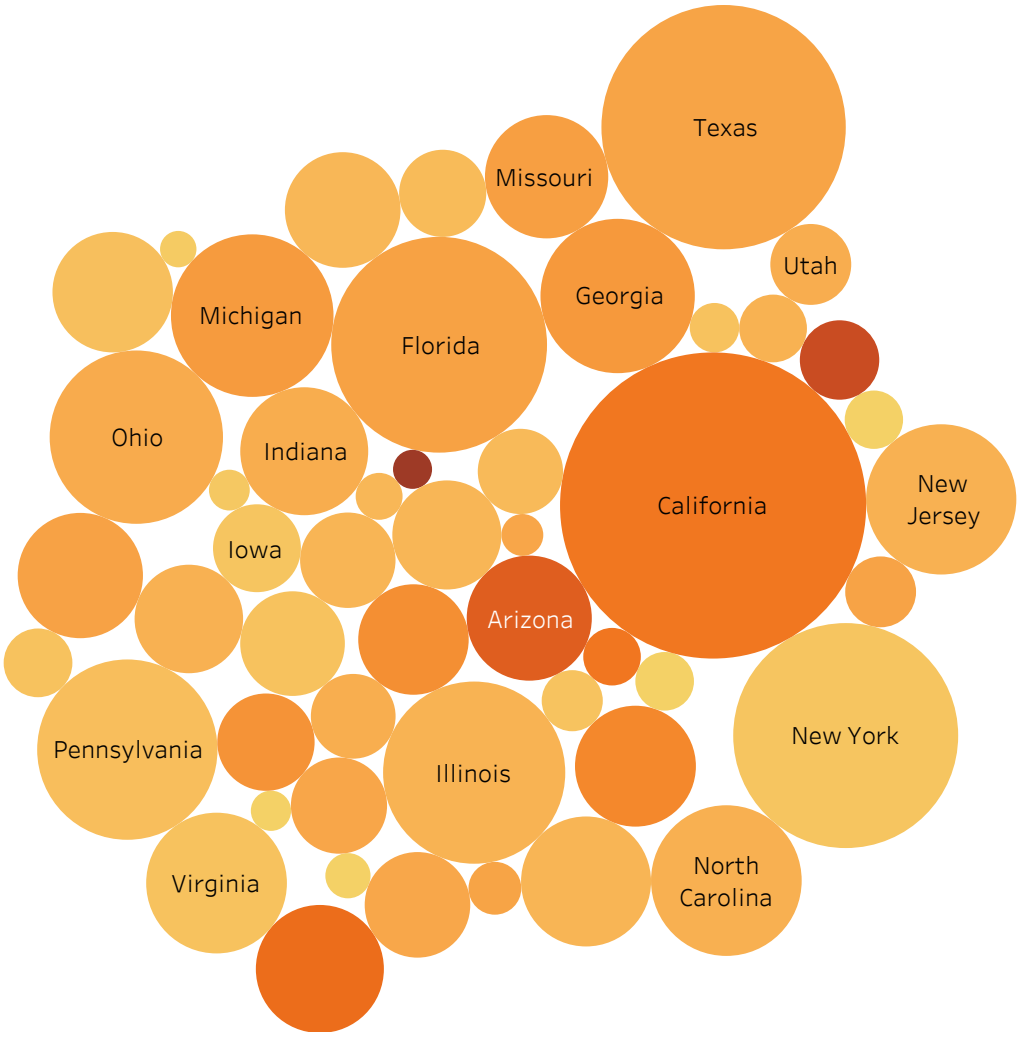
In [ ]:

Tableau - Scatter Chart for Robbery vs Motor Vehicle Theft by State



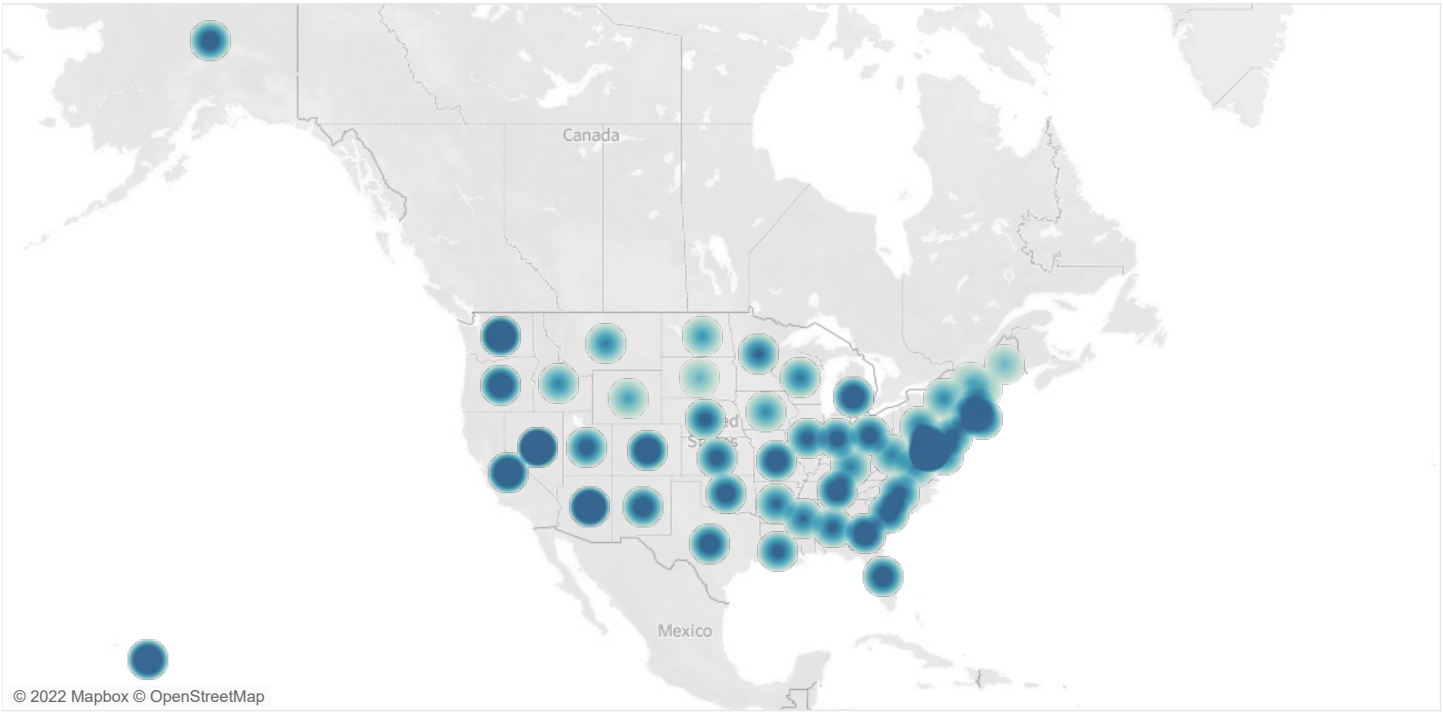
Sum of Motor Vehicle Theft vs. sum of Robbery. Details are shown for State. The view is filtered on State, which excludes United States.

Tableau - Bubble Chart for Motor Vehicle Theft by Population



State. Color shows sum of Motor Vehicle Theft. Size shows sum of Population. The marks are labeled by State. The view is filtered on State, which excludes United States.

Tableau - Density Plot for Motor Vehicle Theft by State



Map based on Longitude (generated) and Latitude (generated). Color shows sum of Motor Vehicle Theft. Details are shown for State. The view is filtered on State, which excludes United States.

Assignment_Week_5&6_Venkidusamy_KesavAdithya

Kesav Adithya Venkidusamy

2022/07/05

```
knitr::opts_chunk$set(echo = TRUE)

library(readxl)
library(ggplot2)
library(dplyr)
```

```
##
## 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
```

Data Loading

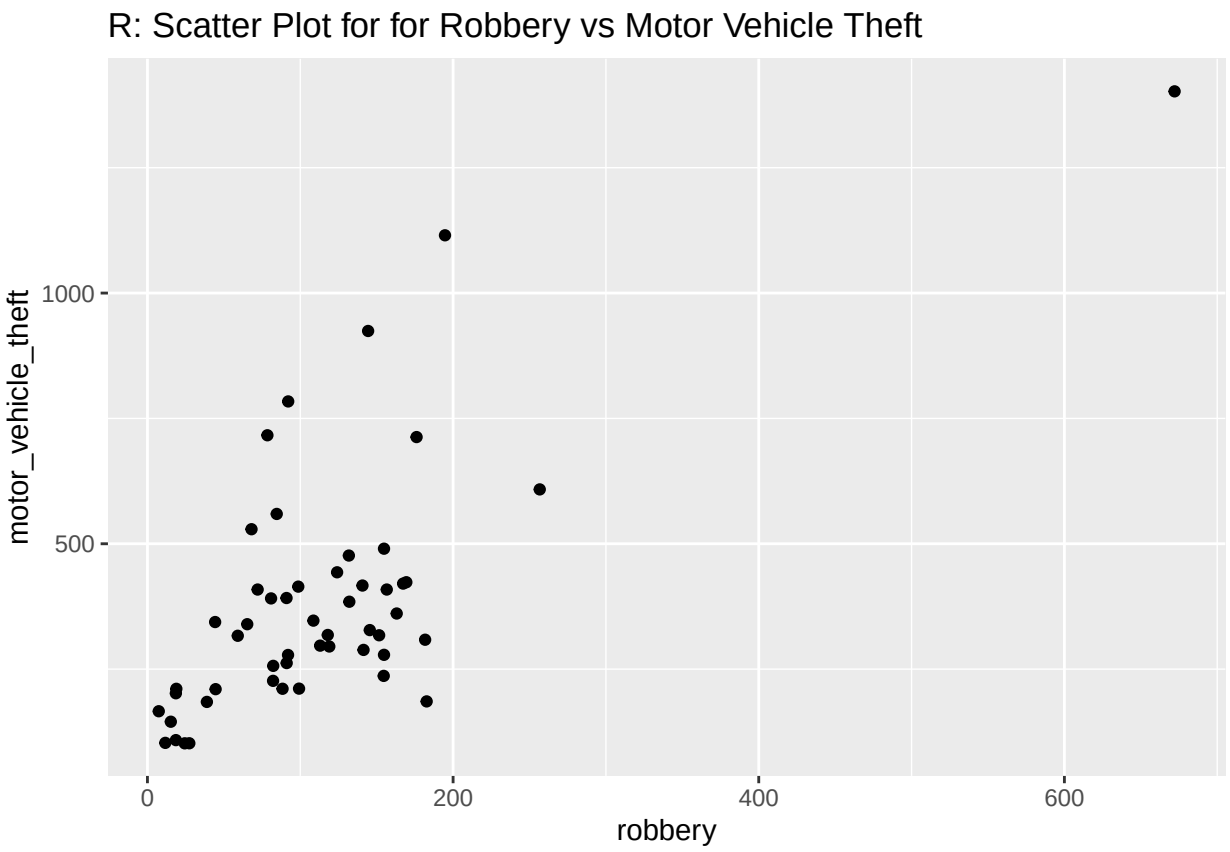
```
# Creating dataframe
crime_df <- read.csv("E:/Personal/Bellevue University/Course/github/dsc640/Week 5&6/crimerates-by-state
head(crime_df)
```

```
##           state murder forcible_rape robbery aggravated_assault burglary
## 1 United States    5.6          31.7   140.7              291.1    726.7
## 2       Alabama    8.2          34.3   141.4              247.8    953.8
## 3        Alaska    4.8          81.1    80.9              465.1    622.5
## 4       Arizona    7.5          33.8   144.4              327.4    948.4
## 5      Arkansas    6.7          42.9    91.1              386.8   1084.6
## 6    California    6.9          26.0   176.1              317.3    693.3
##   larceny_theft motor_vehicle_theft population
## 1        2286.3               416.7  295753151
## 2        2650.0               288.3    4545049
## 3        2599.1               391.0     669488
## 4        2965.2               924.4    5974834
## 5        2711.2               262.1    2776221
## 6        1916.5               712.8   35795255
```

```
# Total number of records present in the data set
nrow(crime_df)
```

```
## [1] 52
```

```
# Scatter Plot
ggplot(crime_df, aes(x=robbery, y=motor_vehicle_theft)) + geom_point() + ggtitle("R: Scatter Plot for f
```




```
## Creating dataframe
crime_df <- read.csv("E:/Personal/Bellevue University/Course/github/dsc640/Week 5&6/crimerates-by-state
head(crime_df)
```

```
##           state murder forcible_rape robbery aggravated_assault burglary
## 1 United States    5.6          31.7   140.7              291.1    726.7
## 2       Alabama    8.2          34.3   141.4              247.8    953.8
## 3        Alaska    4.8          81.1    80.9              465.1    622.5
## 4       Arizona    7.5          33.8   144.4              327.4    948.4
## 5      Arkansas    6.7          42.9    91.1              386.8   1084.6
## 6    California    6.9          26.0   176.1              317.3    693.3
##   larceny_theft motor_vehicle_theft population
## 1        2286.3               416.7  295753151
## 2        2650.0               288.3    4545049
## 3        2599.1               391.0     669488
## 4        2965.2               924.4    5974834
```

```
## 5        2711.2                262.1    2776221
## 6        1916.5                712.8   35795255
```

```
# Total number of records present in the data set
nrow(crime_df)
```

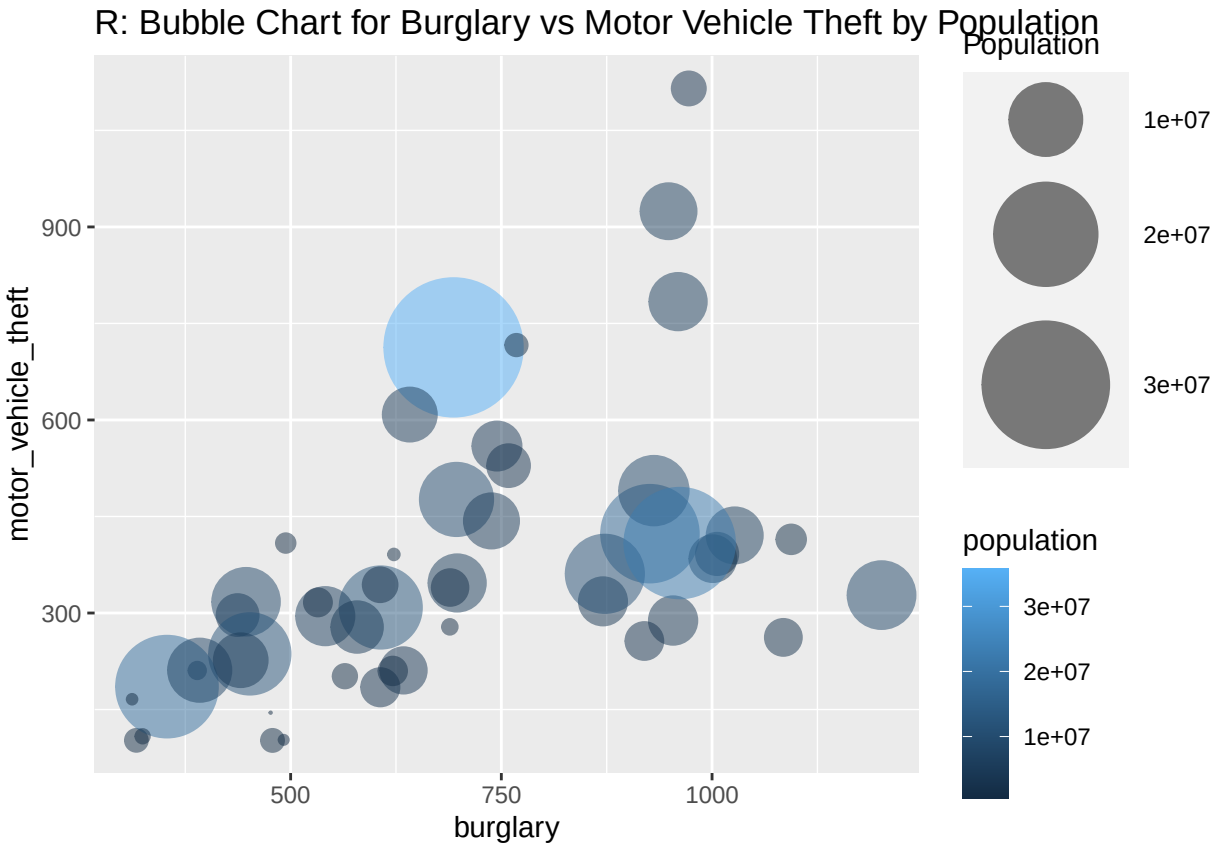
```
## [1] 52
```

```
# Filter outliers
crime_us_df <- filter(crime_df, state != 'United States')
crime_us_df <- filter(crime_us_df, state != 'District of Columbia')

# Total number of records present in the data set
nrow(crime_df)
```

```
## [1] 52
```

```
## Create Bubble Chart
ggplot(crime_us_df, aes(x=burglary, y=motor_vehicle_theft, size=population, color = population))+geom_p
```




```
birth_df <- read.csv("E:/Personal/Bellevue University/Course/github/dsc640/Week 5&6/birth-rates-yearly.
head(birth_df)
```

```
##   year   rate
## 1 1960 36.400
## 2 1961 35.179
## 3 1962 33.863
## 4 1963 32.459
## 5 1964 30.994
## 6 1965 29.513
```

```
# Total number of records present in the data set
nrow(birth_df)
```

```
## [1] 9870
```

```
## Create Stacked Area Chart

ggplot(birth_df, aes(x=rate)) +
    geom_density(color = 'darkblue', fill = 'lightblue', alpha = 0.8) + ggtitle("R: Density Chart for Bi
```