Exercise 4.2: Scatterplots, Bubble Charts, & Density Plots/Maps

Scott Breitbach

DSC640 - 01/27/2022

Plots Using **Python**

Load Data

In [1]:
```python
# Load lib raries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```python
# Load data
crimeDF = pd.read_csv("crimerates-by-state-2005.csv")
birthDF = pd.read_csv("birth-rate.csv")

# Remove Washington, DC due to outliers
crimeDF = crimeDF[crimeDF.state != 'District of Columbia']
# And remove United States due to averages
crimeDF = crimeDF[crimeDF.state != 'United States']

# Set color to Bellevue purple
color = "#4f3674"
```
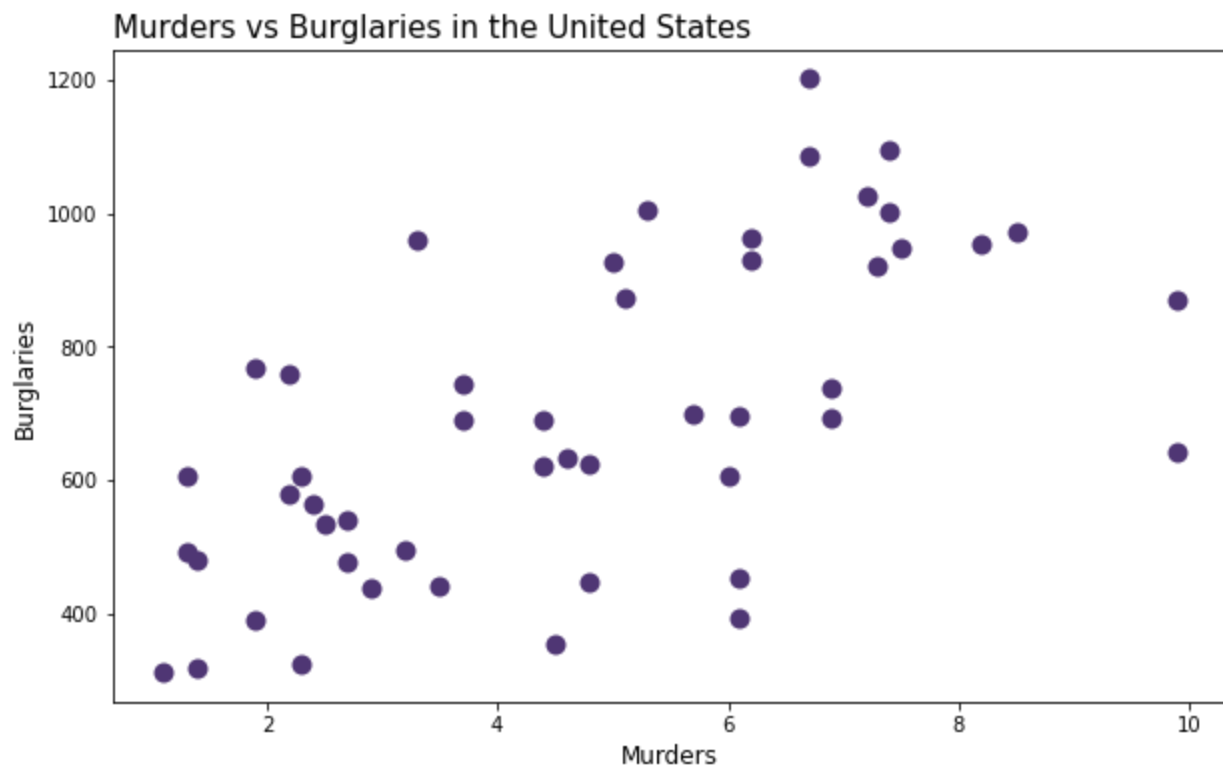
Scatterplot

In [3]:

```python
# Initialize the matplotlib figure
f, ax = plt.subplots(figsize=(10, 6))

# Create a Scatterplot
plt.plot('murder', 'burglary', data=crimeDF, linestyle='none', markersize=10,
         markerfacecolor=color, marker='o', markeredgecolor="none")

# Add chart title and labels
plt.title("Murders vs Burglaries in the United States", fontsize = 15, loc = 'left')
plt.xlabel("Murders", fontsize = 12)
plt.ylabel("Burglaries", fontsize = 12)

plt.show()
```

Bubble Chart

In [4]:

```python
# Initialize the matplotlib figure
f, ax = plt.subplots(figsize=(10, 6))

# Make a bubble plot
sc = plt.scatter('murder', 'burglary', s=crimeDF.population*.0001, data=crimeDF,
                 alpha=0.7, color=color, label='Population')

# Add chart title and labels
plt.title("Murders vs Burglaries in the United States", fontsize = 15, loc = 'left')
plt.xlabel("Murders", fontsize = 12)
plt.ylabel("Burglaries", fontsize = 12)

# Set legend for population sizes
handles, labels = sc.legend_elements(prop="sizes", alpha=0.4, num=3)
labels = ["10 MM", "20 MM", "30 MM"]
legend = ax.legend(handles, labels, loc="lower right", title="Population",
                   fontsize=11, title_fontsize=14)

plt.show()
```
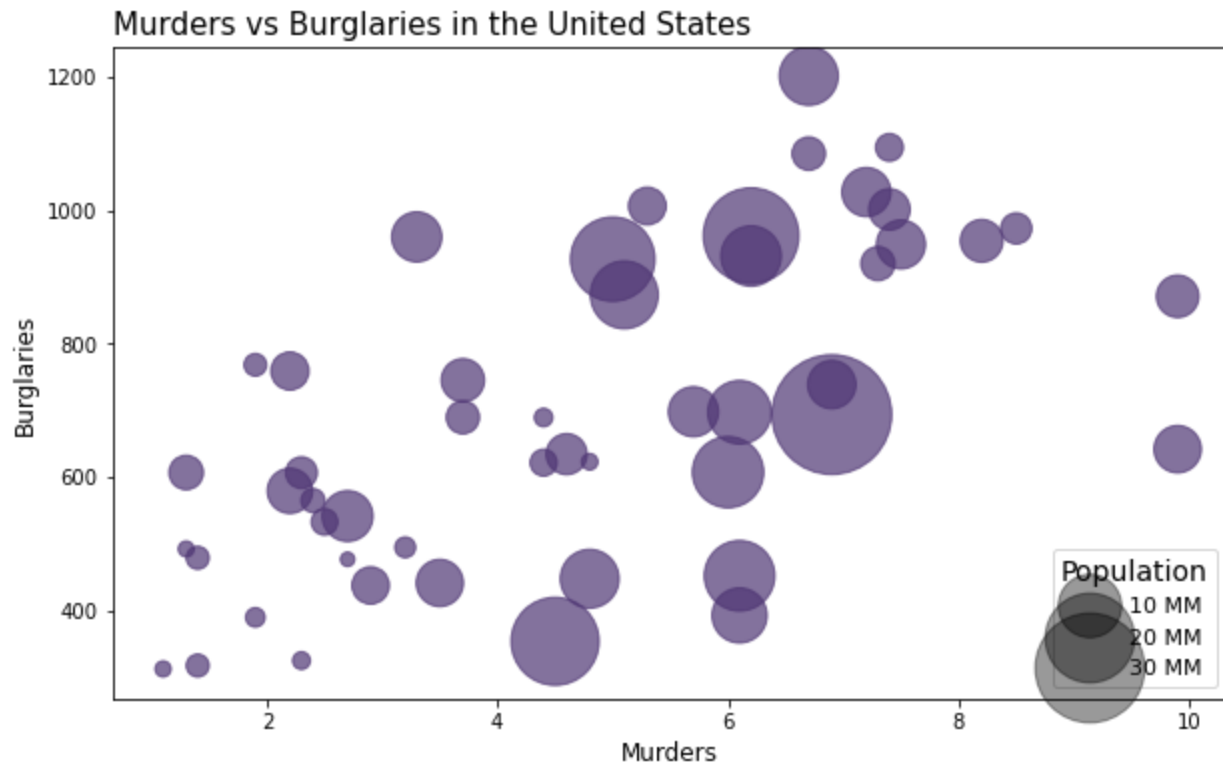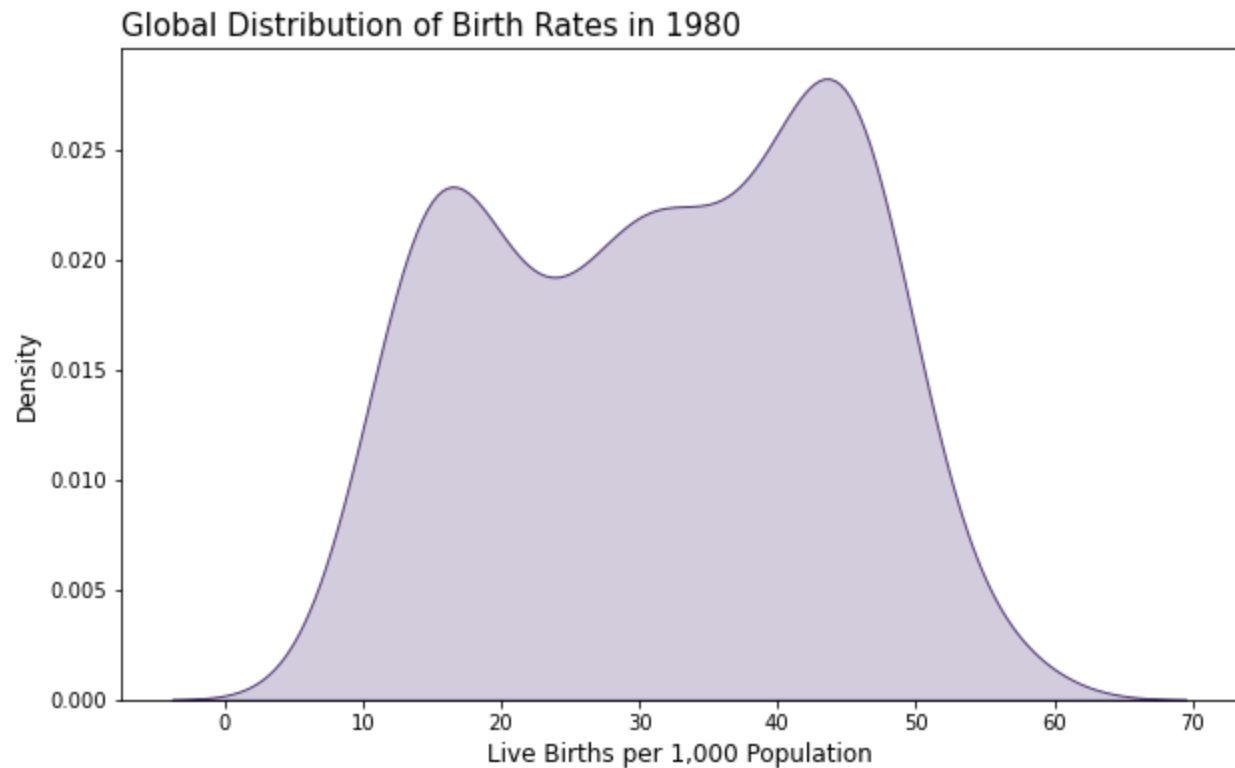
Density Plot

```
# Initialize the matplotlib figure
f, ax = plt.subplots(figsize=(10, 6))

# Make a density plot
sns.kdeplot(birthDF['1980'], color=color, fill=color)

# Add chart title and labels
plt.title("Global Distribution of Birth Rates in 1980", fontsize = 15, loc = 'left')
plt.xlabel("Live Births per 1,000 Population", fontsize = 12)
plt.ylabel("Density", fontsize = 12)
```

Text(0, 0.5, 'Density')

Exercise 4.2: Scatterplots, Bubble Charts, & Density Plots/Maps

Scott Breitbach

1/27/2022

Plots Using R

```r
knitr::opts_chunk$set(echo = TRUE, warning = FALSE)

# Set Working Directory
setwd("C:/Users/micha/OneDrive/Documents/GitHub/DSC640/Weeks7-8/")

# Load libraries
library(ggplot2)
```

Load Data

```r
# Set color to Bellevue purple
color = "#4f3674"

# Load the data
crimeDF <- read.csv("crimerates-by-state-2005.csv")
birthDF <- read.csv("birth-rate.csv")

# Remove Washington, DC due to outliers
crimeDF <- crimeDF[crimeDF$state != "District of Columbia",]
# And remove United States due to averages
crimeDF <- crimeDF[crimeDF$state != "United States",]
```
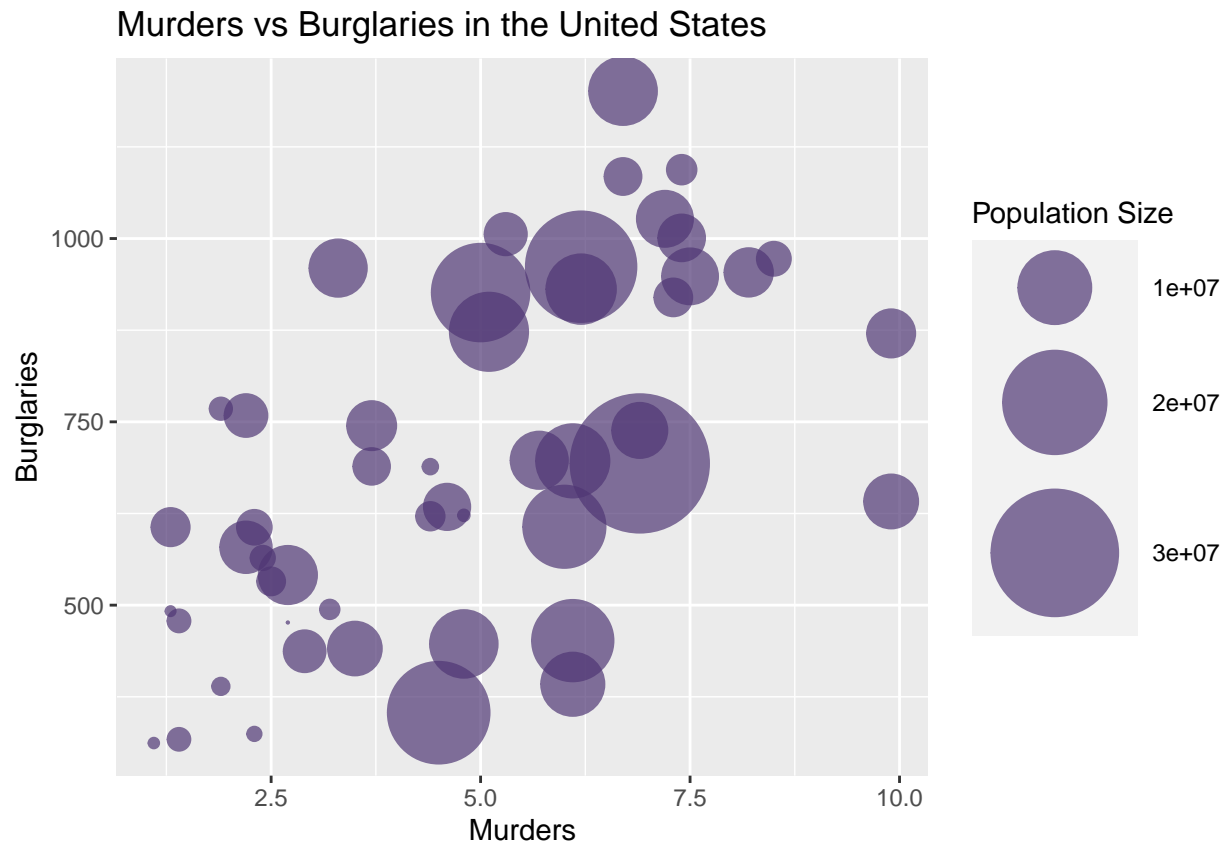
Scatter Plot

```r
# Create a Scatterplot
ggplot(crimeDF, aes(x=murder, y=burglary)) +
  geom_point(col=color) +
  theme_gray() +
  labs(x="Murders", y="Burglaries") +
  ggtitle("Murders vs Burglaries in the United States")
```
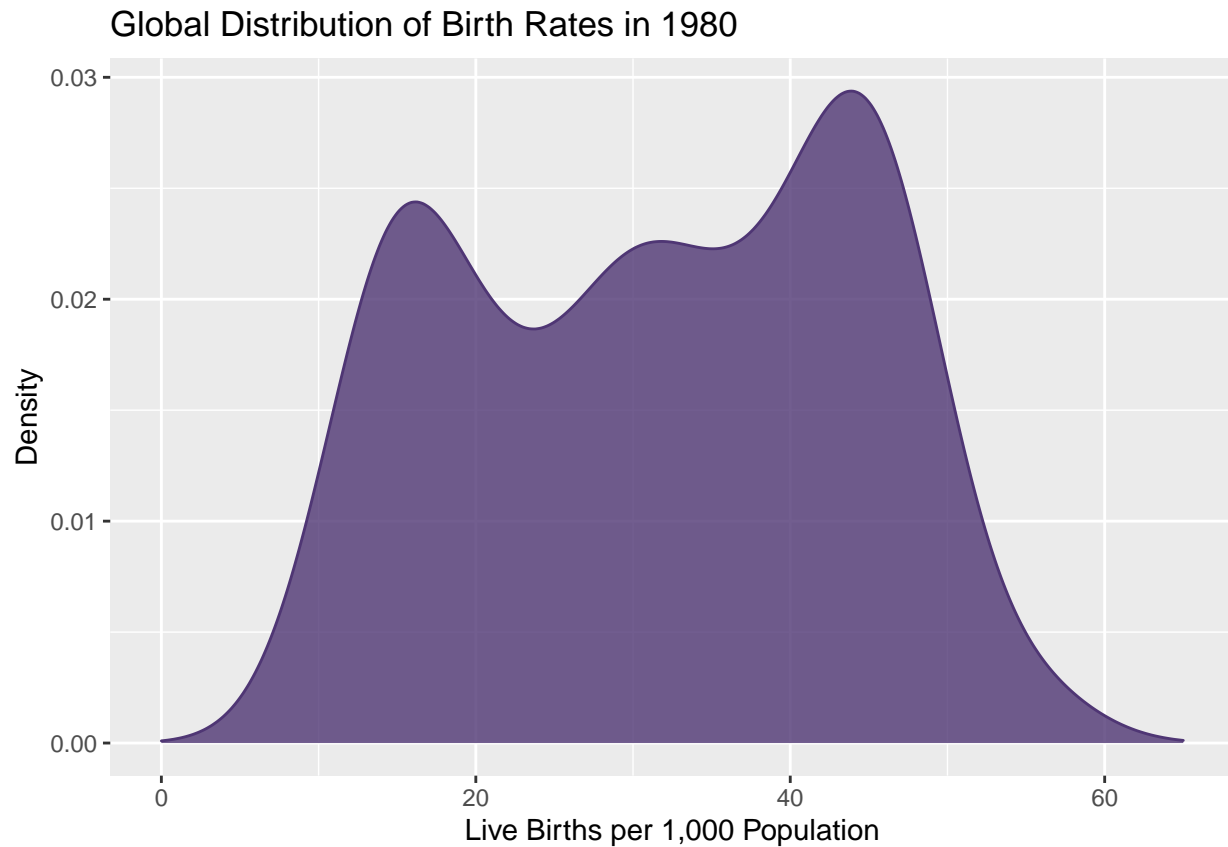

Murders vs Burglaries in the United States
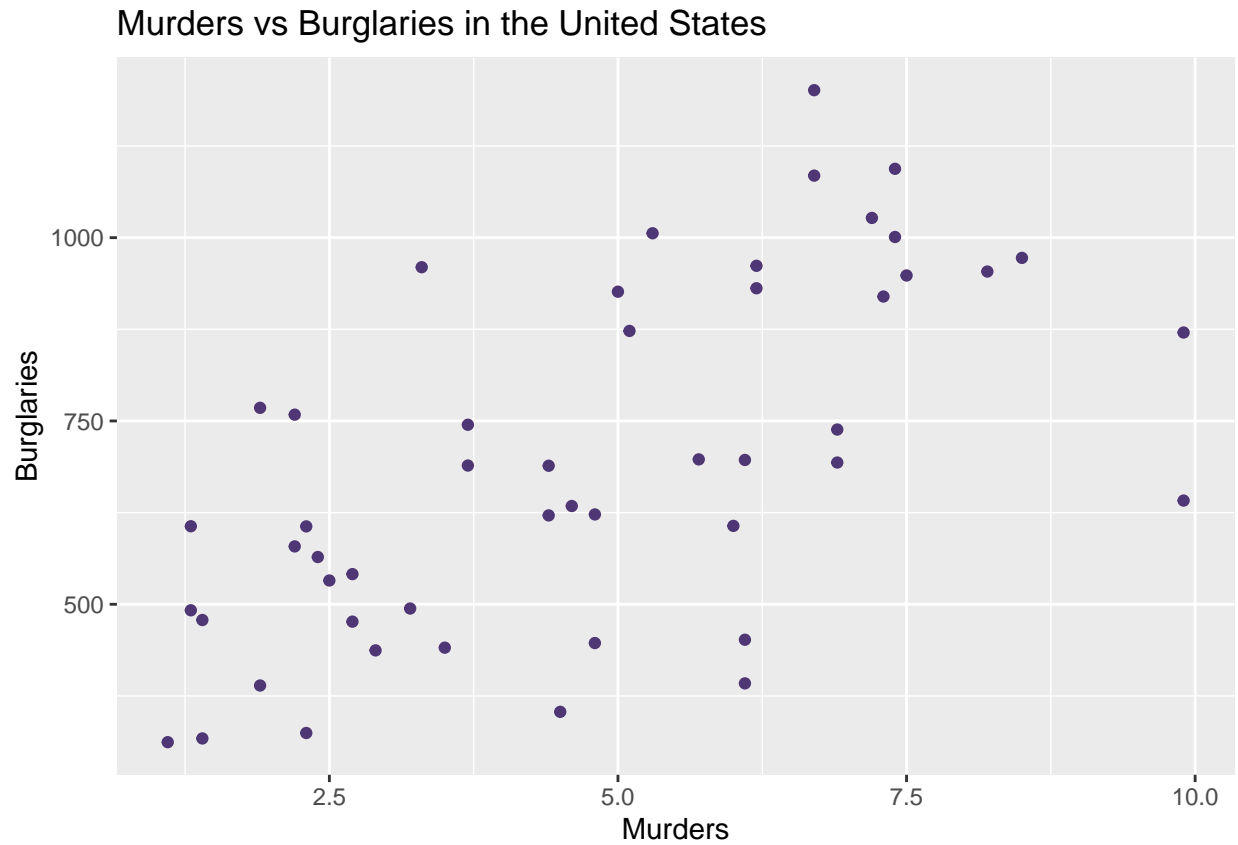
Bubble Chart

```
# Make a bubble plot
ggplot(crimeDF, aes(x=murder, y=burglary, size=population)) +
  geom_point(alpha=0.7, col=color) +
  scale_size(range=c(0.1,24), name="Population Size") +
  theme_gray() +
  labs(x="Murders", y="Burglaries") +
  ggtitle("Murders vs Burglaries in the United States")
```

Density Plot

```
# Make a density plot
ggplot(birthDF, aes(x=X1980)) +
  geom_density(fill=color, color=color, alpha=0.8) +
  xlim(0,65) +
  theme_gray() +
  labs(x="Live Births per 1,000 Population", y="Density") +
  ggtitle("Global Distribution of Birth Rates in 1980")
```
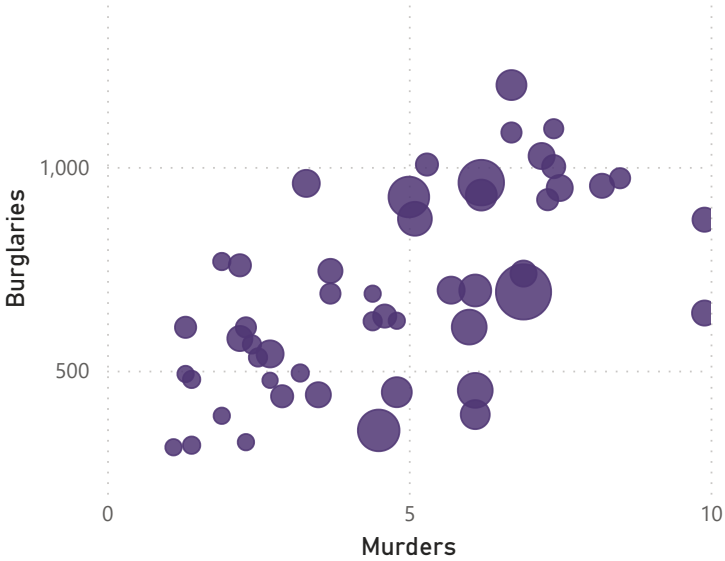
Global Distribution of Birth Rates in 1980

Power BI

Scatter Plot

Murders vs Burglaries in the United States

Bubble Plot

Murders vs Burglaries in the United States

Population represented by bubble sizes

Density Map

US Population by State