R Programming

Week 5-6: Prepare - Scatterplots, Bubble Charts, and Density Plots

Shani Kumar

## Week 5-6: Exercises: Charts

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.

**Data source** We are using dataset from [Data Source URL](https://content.bellevue.edu/cst/dsc/640/datasets/ex4-2.zip) file.

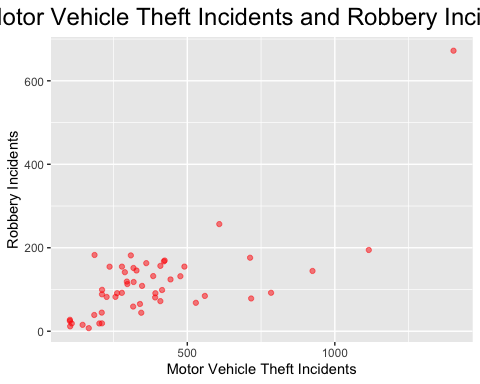
## state murder forcible\_rape robbery aggravated\_assault burglary  
## 1 Alabama 8.2 34.3 141.4 247.8 953.8  
## 2 Alaska 4.8 81.1 80.9 465.1 622.5  
## 3 Arizona 7.5 33.8 144.4 327.4 948.4  
## 4 Arkansas 6.7 42.9 91.1 386.8 1084.6  
## 5 California 6.9 26.0 176.1 317.3 693.3  
## 6 Colorado 3.7 43.4 84.6 264.7 744.8  
## larceny\_theft motor\_vehicle\_theft population  
## 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  
## 6 2735.2 559.5 4660780

### Data structure:

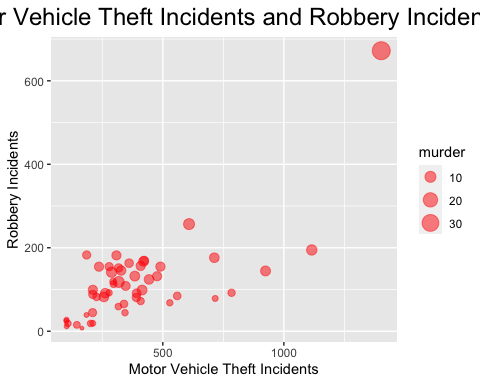
## 'data.frame': 51 obs. of 9 variables:  
## $ state : Factor w/ 52 levels "Alabama","Alaska",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ murder : num 8.2 4.8 7.5 6.7 6.9 3.7 2.9 4.4 35.4 5 ...  
## $ forcible\_rape : num 34.3 81.1 33.8 42.9 26 43.4 20 44.7 30.2 37.1 ...  
## $ robbery : num 141.4 80.9 144.4 91.1 176.1 ...  
## $ aggravated\_assault : num 248 465 327 387 317 ...  
## $ burglary : num 954 622 948 1085 693 ...  
## $ larceny\_theft : num 2650 2599 2965 2711 1916 ...  
## $ motor\_vehicle\_theft: num 288 391 924 262 713 ...  
## $ population : int 4545049 669488 5974834 2776221 35795255 4660780 3477416 839906 582049 17783868 ...

### Construct Charts:

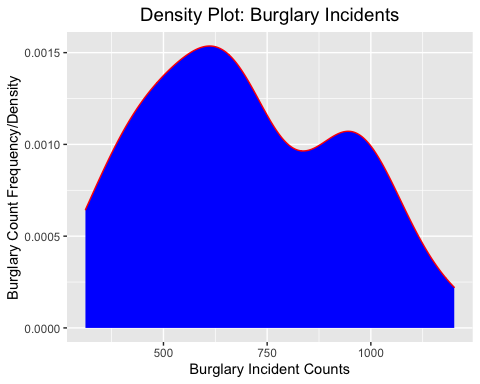
**Scatterplot**

Plot: Motor Vehicle Theft Incidents vs Robbery Incidents 

**Bubble Chart**

Plot: Motor Vehicle Theft Incidents vs Robbery Incidents (Bubble Size - Murder) 

**Density Plot**

Density Plot: For Burglary Incidents 

Python – Jupyter Notebook

# Week 7-8 - Assignment

**Prepare - Scatterplots, Bubble Chart & Density Plot**

**By**

**Shani Kumar**

### Introduction: Assignment Details

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.

### Source Data

<https://content.bellevue.edu/cst/dsc/640/datasets/ex4-2.zip>

In [1]:

*# Impprt required libraries/packages*

**import** **numpy** **as** **np**

**import** **pandas** **as** **pd**

**import** **squarify**

**import** **matplotlib.pyplot** **as** **plt**

**import** **seaborn** **as** **sns**

*# configure display of graph*

%**matplotlib** inline

### Load data into a dataframe

In [2]:

*# load the csv file as a data frame*

crime\_rate\_raw = pd.read\_csv('crimerates-by-state-2005.csv')

crime\_rate = crime\_rate\_raw[(crime\_rate\_raw['state']!='United States')]

*# summarize the shape of the dataset*

print("Crime Rate by State Data:**\n**")

print("Dataset Shape: ",crime\_rate.shape)

*# see the sample of the data*

print("Sample Data: ")

crime\_rate.head()

Crime Rate by State Data:

Dataset Shape: (51, 9)

Sample Data:

Out[2]:

|  | **state** | **murder** | **forcible\_rape** | **robbery** | **aggravated\_assault** | **burglary** | **larceny\_theft** | **motor\_vehicle\_theft** | **population** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | Alabama | 8.2 | 34.3 | 141.4 | 247.8 | 953.8 | 2650.0 | 288.3 | 4545049 |
| **2** | Alaska | 4.8 | 81.1 | 80.9 | 465.1 | 622.5 | 2599.1 | 391.0 | 669488 |
| **3** | Arizona | 7.5 | 33.8 | 144.4 | 327.4 | 948.4 | 2965.2 | 924.4 | 5974834 |
| **4** | Arkansas | 6.7 | 42.9 | 91.1 | 386.8 | 1084.6 | 2711.2 | 262.1 | 2776221 |
| **5** | California | 6.9 | 26.0 | 176.1 | 317.3 | 693.3 | 1916.5 | 712.8 | 35795255 |

## Scatterplot

In [3]:

*# Create a scatter plot showing correlation between robbery and motor\_vehicle\_theft*

plt.scatter(crime\_rate['robbery'], crime\_rate['motor\_vehicle\_theft'],alpha=0.5)

plt.title('Correlation between robbery and motor\_vehicle\_theft')

plt.ylabel('Motor Vehicle Theft')

plt.xlabel('Robbery')

plt.show()

Chart, scatter chart

Description automatically generated

## Bubble Chart

In [4]:

*# Create a bubble plot showing correlation between motor\_vehicle\_theft and robbery,*

*# and using murder for the size of the bubble*

plt.scatter(crime\_rate['motor\_vehicle\_theft'], crime\_rate['robbery'], s=crime\_rate['murder']\*10,alpha=0.5)

plt.title('Correlation between motor\_vehicle\_theft and robbery (Murder - Bubble Size)')

plt.xlabel('Motor Vehicle Theft Incidents')

plt.ylabel('Robbery Incidents')

plt.show()

Chart, scatter chart

Description automatically generated

## Density Plot

In [5]:

*# Plotting distribution of burglary incidents*

sns.distplot(crime\_rate['burglary'], hist = **False**, kde = **True**,

kde\_kws = {'shade': **True**, 'linewidth': 3})

plt.title('Density Plot for Burglary Incidents')

plt.xlabel('Burglary Incident counts')

plt.ylabel('Frequency')

plt.show()

Icon

Description automatically generated

Tableau





