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

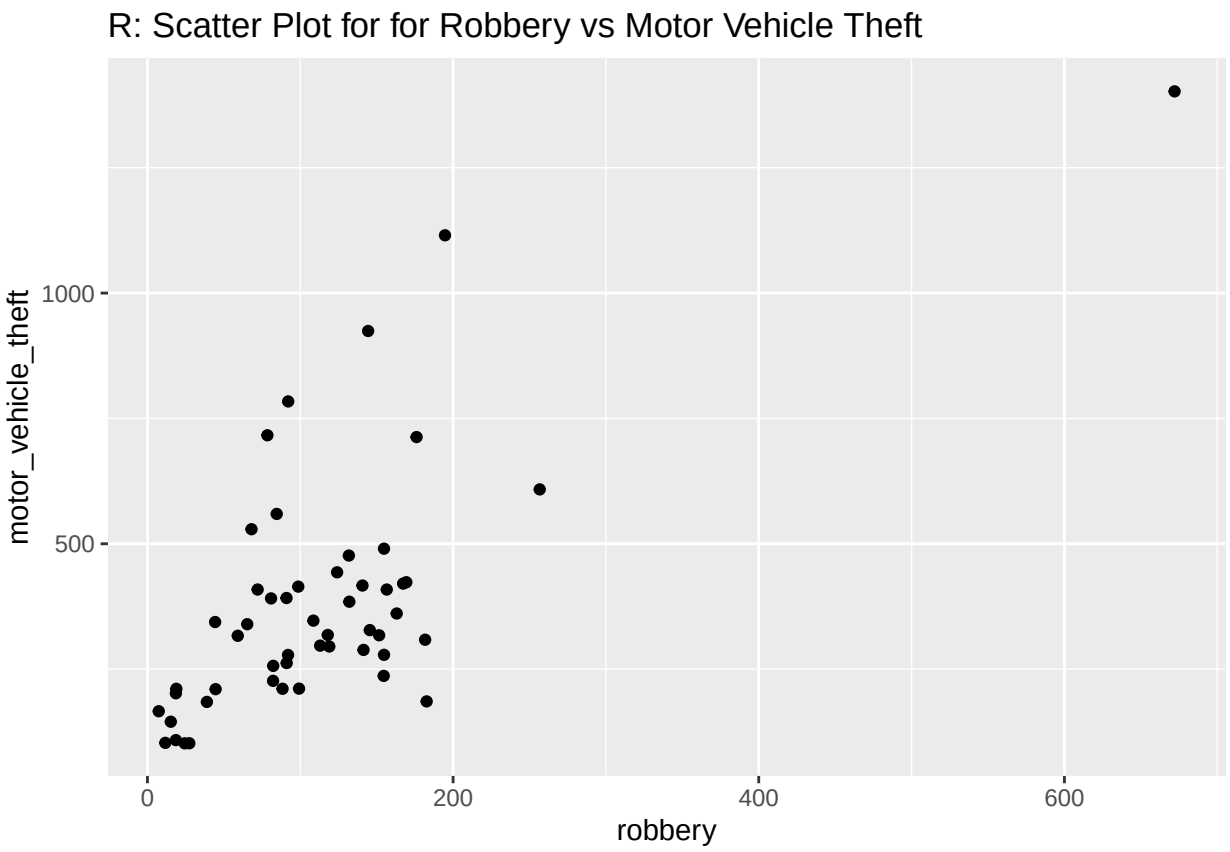
```
# Creating dataframe
crime_df <- read.csv("E:/Personal/Bellevue University/Course/github/dsc640/Week 5&6/crimerates-by-state
head(crime_df)
```

```
##           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)
```

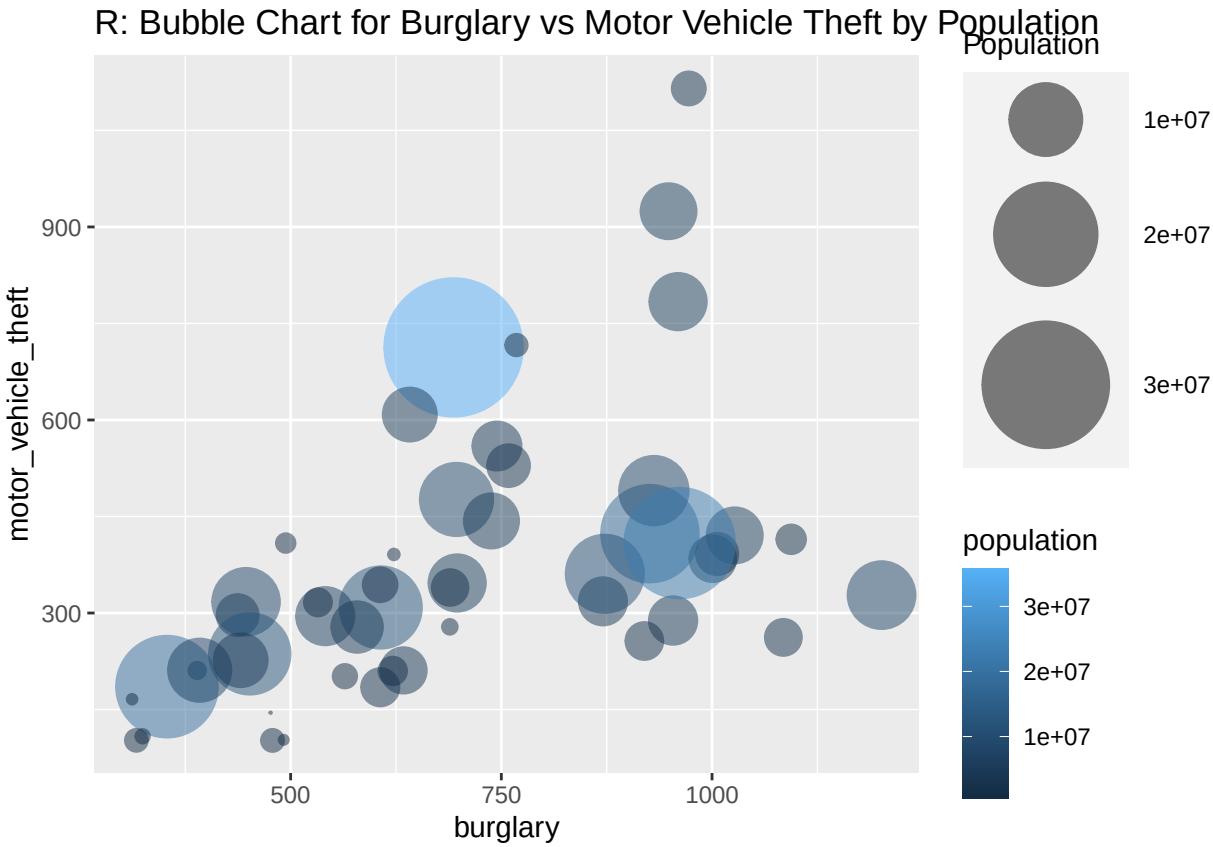
```
## [1] 52
```

```
# 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
```