Ryan Long
DSC640
4.2 Exercises: Charts

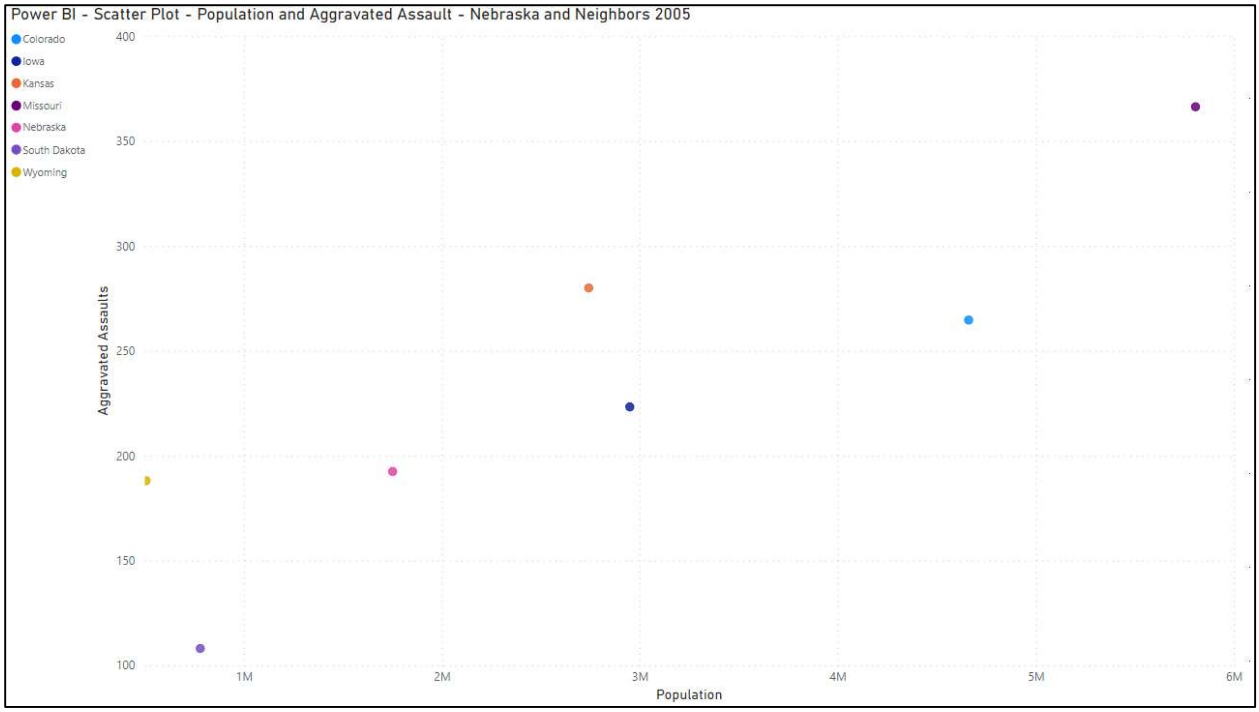
Dataset used:
crimerates-by-state-2005.csv

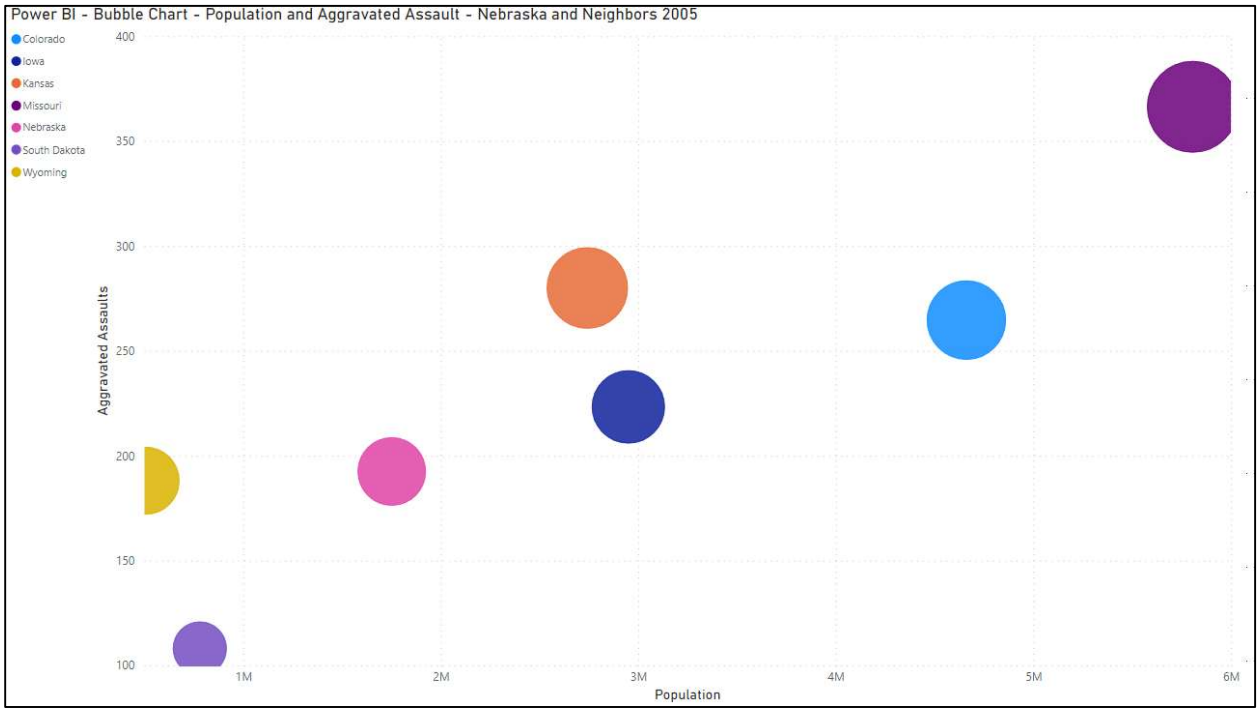
Summary
I tried two different approaches with the Scatter and Density charts. Within Power BI, I subset the data for Nebraska and neighboring states. I did this because I was interested in seeing how they all compared, and I couldn't do all 50 with a legible legend. With this approach, the reader is given more detail for each state, but lacks the broader perspective in my Python and R charts.

With the Python and R scatter and bubble charts, I used all 50 states + DC, but they are not individually labeled. The information here conveys only population and number of crimes in each. Either is a viable option, depending on what the messaging if focused on.

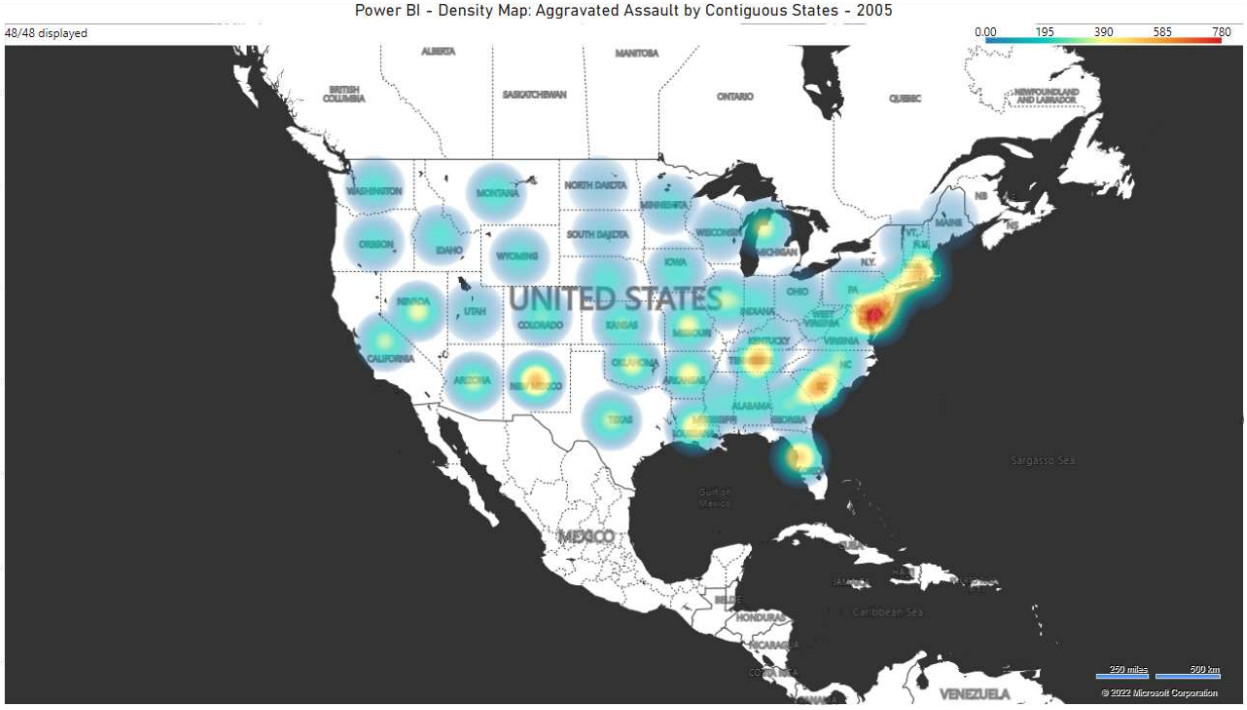
With the density charts, both Python and R were similar in nature showing a slightly right-skew distribution. For Power BI, I took the density map literally and overlaid the data on a map of the contiguous US. Visually, it is more interesting than the simple density chart. I think using them in conjunction with each other would provide the reader the best perspective on the data.

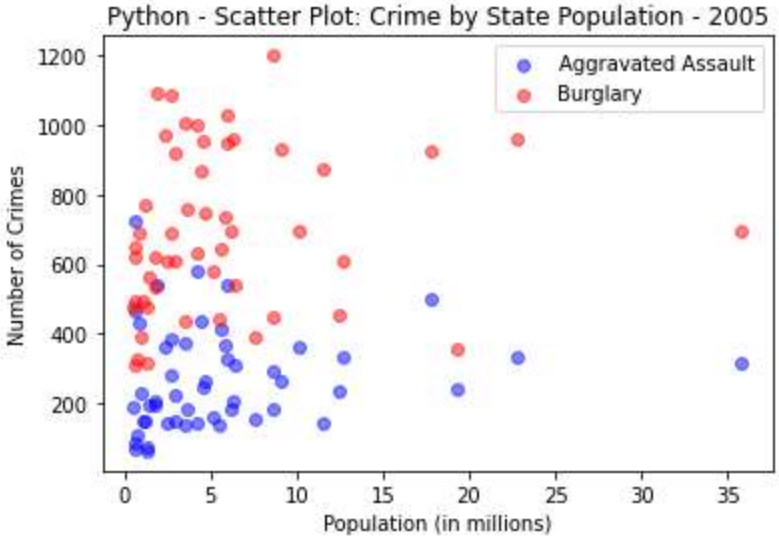
The following pages contain:

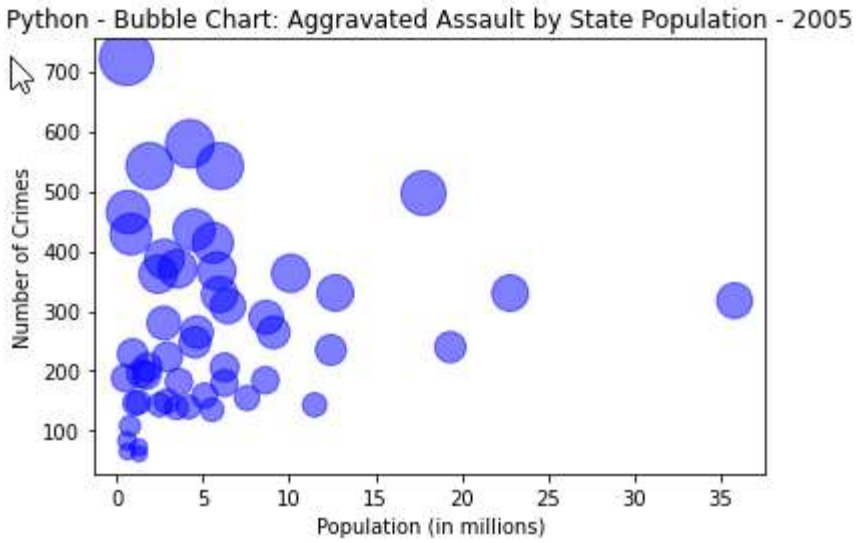

Power BI - Scatter Plot - Population and Aggravated Assault - Nebraska and Neighbors 2005
Colorado
Iowa
Kansas
Missouri
Nebraska
South Dakota
Wyoming
Aggravated Assaults
400
350
300
250
200
150
100
1M
2M
3M
4M
5M
6M
Population

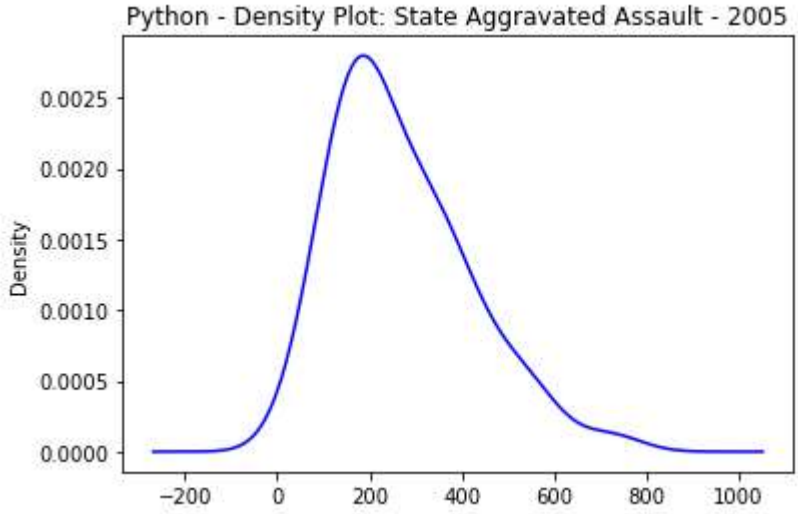

Power BI - Bubble Chart - Population and Aggravated Assault - Nebraska and Neighbors 2005
Colorado
Iowa
Kansas
Missouri
Nebraska
South Dakota
Wyoming
Aggravated Assaults
400
350
300
250
200
150
100
1M
2M
3M
4M
5M
6M
Population

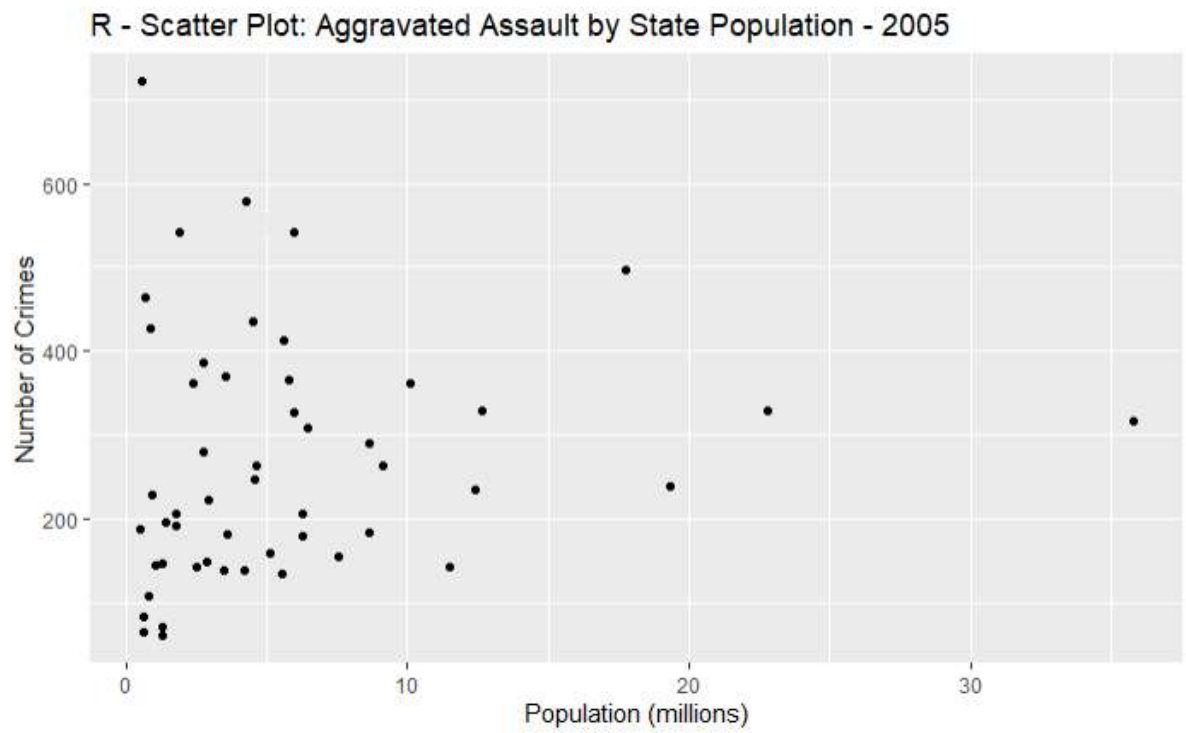
Power BI - Density Map: Aggravated Assault by Contiguous States - 2005

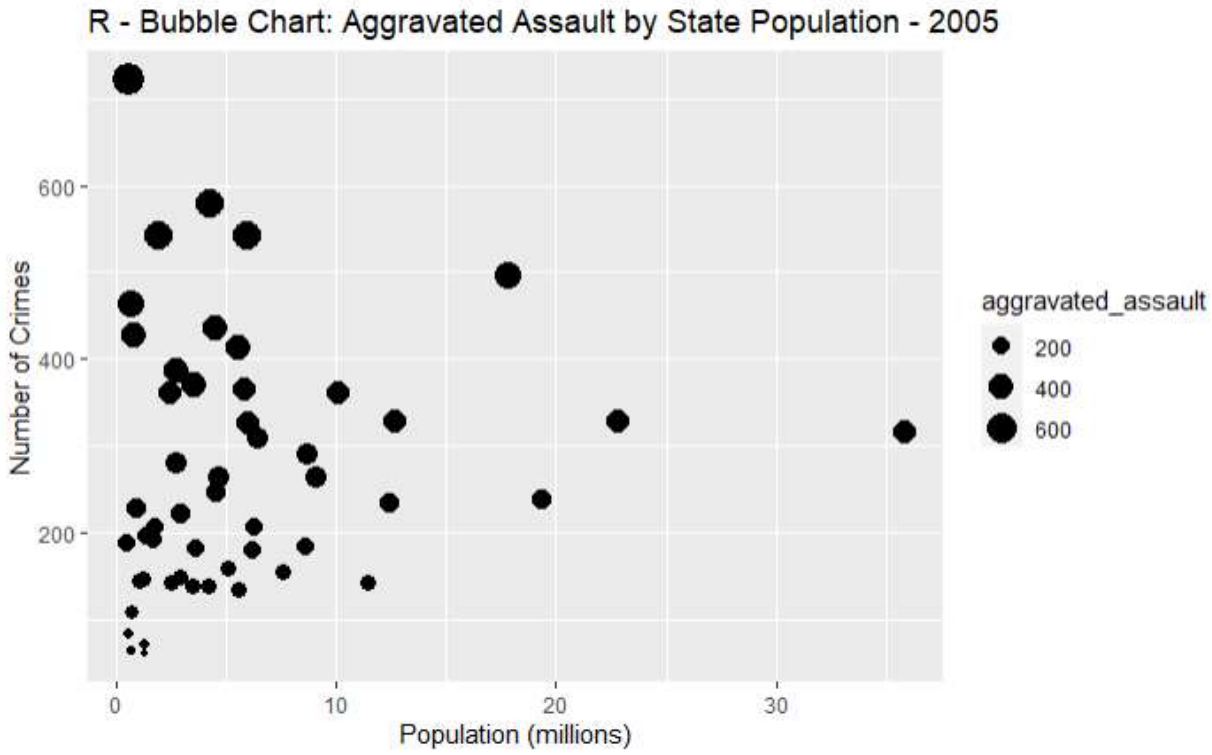


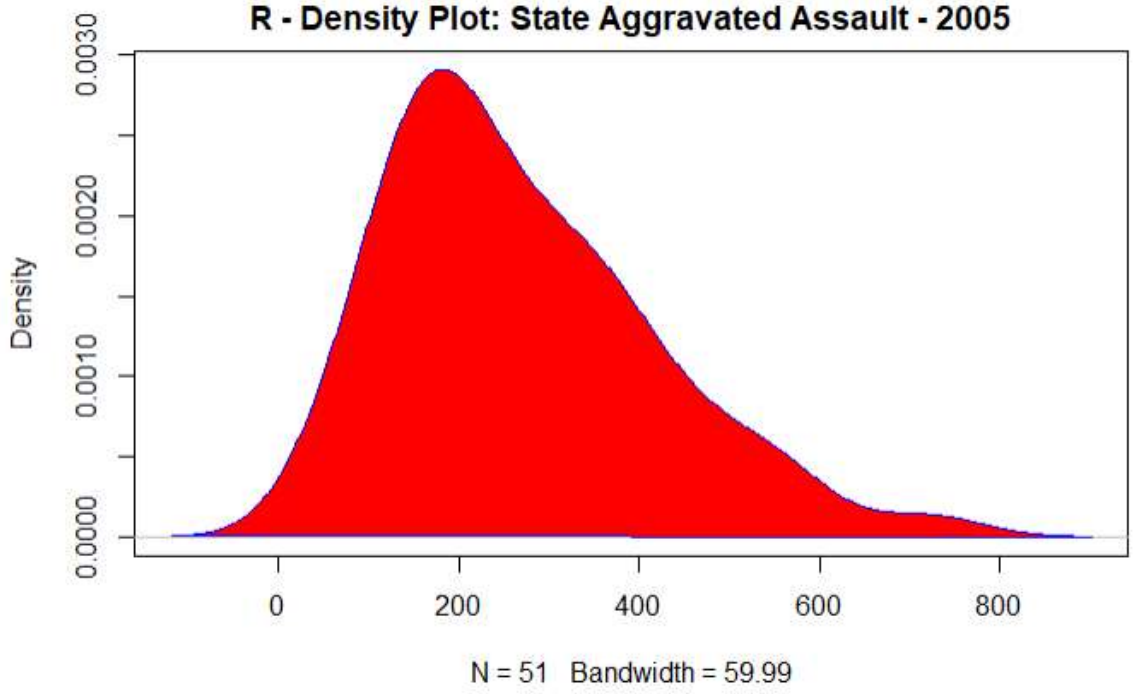



Python - Bubble Chart: Aggravated Assault by State Population - 2005
Number of Crimes
700
600
500
400
300
200
100
0
5
10
15
20
25
30
35
Population (in millions)


Python - Density Plot: State Aggravated Assault - 2005
Density
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
−200
0
200
400
600
800
1000


R - Scatter Plot: Aggravated Assault by State Population - 2005
Number of Crimes
600
400
200
0
10
20
30
Population (millions)


R - Bubble Chart: Aggravated Assault by State Population - 2005
Number of Crimes
600
400
200
0
10
20
30
Population (millions)
aggravated_assault
200
400
600


R - Density Plot: State Aggravated Assault - 2005
Density
0.0000
0.0010
0.0020
0.0030
0
200
400
600
800
N = 51 Bandwidth = 59.99

APPENDIX

In [1]:
```
#Load libraries
import pandas as pd
import seaborn as sns

import numpy as np
import matplotlib.pyplot as plt

#import squarify
#import matplotlib.ticker as plticker # for plot ticks
```

In [2]:
```
#import data as dataframe

data1 = pd.read_csv('crimerates-by-state-2005.csv')
tv = pd.read_csv('tv_sizes.txt',names=['Year','Size'], skiprows=1, sep='\t') #tab seper
#data2 = pd.read_csv('expenditures.txt',names=['Year','Category','Expenditure','Sex'],
```

In [3]:
```
data1 = data1.iloc[1: , :] # drops US row
data1['populationM'] = data1['population'].div(1000000)
```

In [4]:
```
neighbors = ['Nebraska','Iowa','Wyoming','Colorado','Kansas','South Dakota','Missouri',
```

In [5]:
```
neighborsdf = data1[data1['state'].isin(neighbors)]
```

In [6]:
```
neighborsdf
```

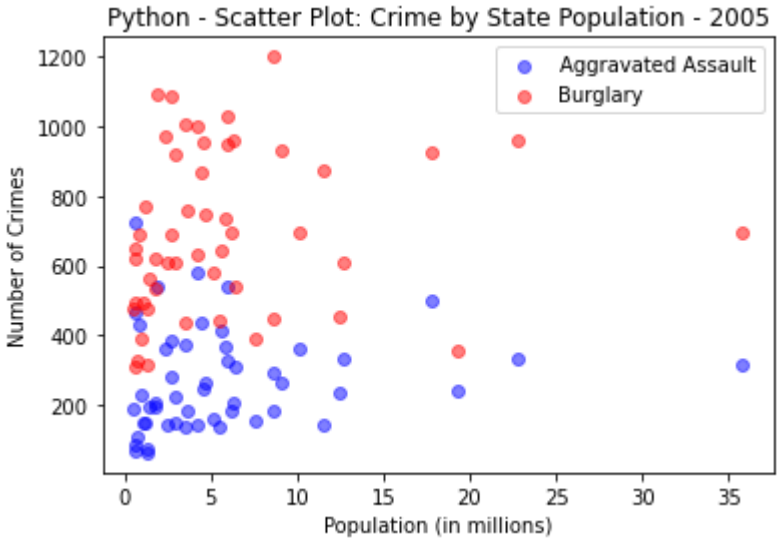
Out[6]:

	state	murder	forcible_rape	robbery	aggravated_assault	burglary	larceny_theft	motor_vehicl
6	Colorado	3.7	43.4	84.6	264.7	744.8	2735.2	
16	Iowa	1.3	27.9	38.9	223.3	606.4	2042.7	
17	Kansas	3.7	38.4	65.3	280.0	689.2	2758.1	
26	Missouri	6.9	28.0	124.1	366.4	738.3	2746.2	
28	Nebraska	2.5	32.9	59.1	192.5	532.4	2574.3	
42	South Dakota	2.3	46.7	18.6	108.1	324.4	1343.7	
51	Wyoming	2.7	24.0	15.3	188.1	476.3	2533.9	

Scatter Plot

In [7]:
```
plt.scatter(data1["populationM"], data1["aggravated_assault"], color='blue',alpha=0.5,l
plt.scatter(data1["populationM"], data1["burglary"], color='red',alpha=0.5,label='Burgl
plt.xlabel("Population (in millions)")  # X-axis label
plt.ylabel("Number of Crimes")  # Y-axis label
plt.title("Python - Scatter Plot: Crime by State Population - 2005")  # title
```

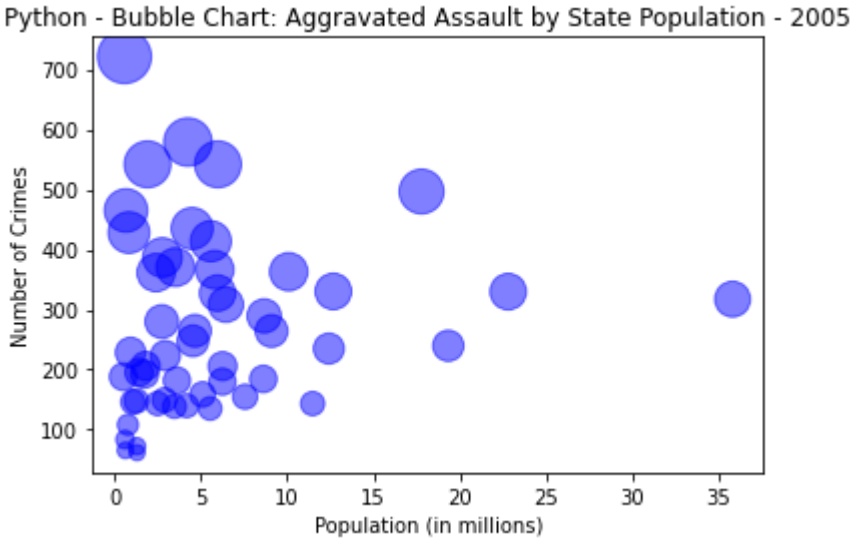
```
plt.legend(loc='upper right')
plt.show()
```



Bubble Chart

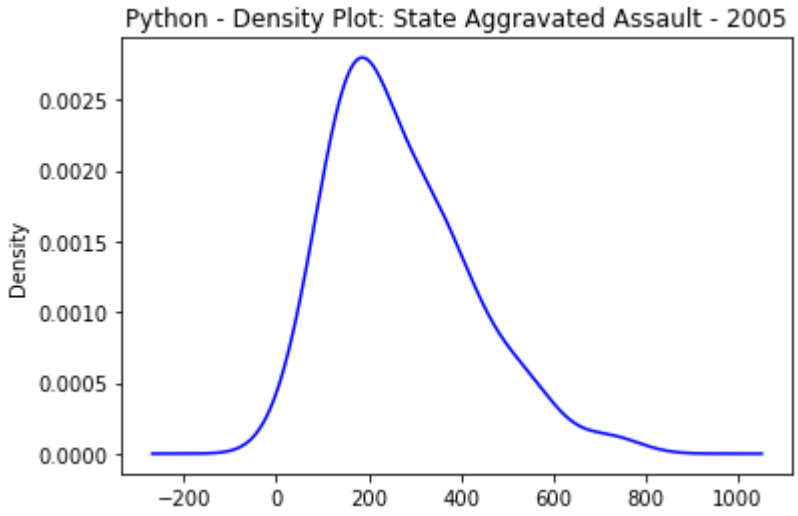
In [8]:
```
plt.scatter(data1["populationM"], data1["aggravated_assault"], color='blue',alpha=0.5,s
plt.xlabel("Population (in millions)")  # X-axis label
plt.ylabel("Number of Crimes")  # Y-axis label
plt.title("Python - Bubble Chart: Aggravated Assault by State Population - 2005")  # ti

plt.show()
```



Density Plot

In [9]:
```
data1.aggravated_assault.plot.density(color='blue')
plt.title('Python - Density Plot: State Aggravated Assault - 2005')
plt.show()
```


Python - Density Plot: State Aggravated Assault - 2005
Density
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
−200
0
200
400
600
800
1000

In [ ]:

Week 7 & 8

Code

Hide

```
#load libraries
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
```

Hide

```
library(tidyr)
library(hrbrthemes)
```

```
Registered S3 methods overwritten by 'htmltools':
  method               from
  print.html           tools:rstudio
  print.shiny.tag      tools:rstudio
  print.shiny.tag.list tools:rstudio
Registering Windows fonts with R
NOTE: Either Arial Narrow or Roboto Condensed fonts are required to use these themes.
      Please use hrbrthemes::import_roboto_condensed() to install Roboto Condensed and
      if Arial Narrow is not on your system, please see https://bit.ly/arialnarrow
```

Hide

```
library(pivottabler)
```

```
Registered S3 method overwritten by 'htmlwidgets':
  method           from
  print.htmlwidget tools:rstudio
Registered S3 method overwritten by 'data.table':
  method           from
  print.data.table
```

Hide

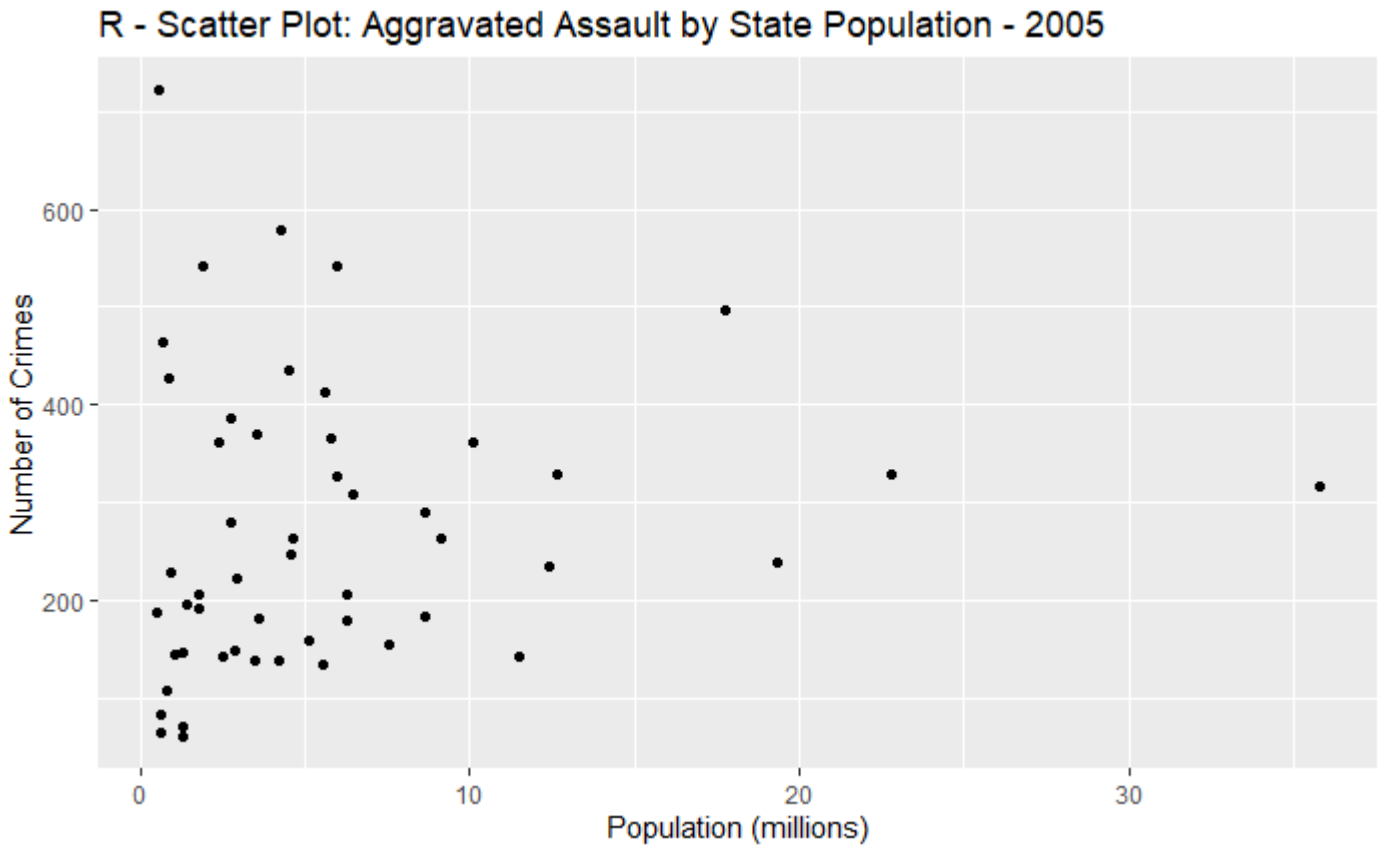
```
library(areaplot)
```

Hide

```
#import data
data = read.csv("C:\\Users\\longr\\Documents\\DSC 640\\Week 7 & 8\\4.2 Exercises\\crimerates-by-state-2005.csv")
data <- data[-c(1), ]
data$PopinM <- round(data$population/1000000,2)
```

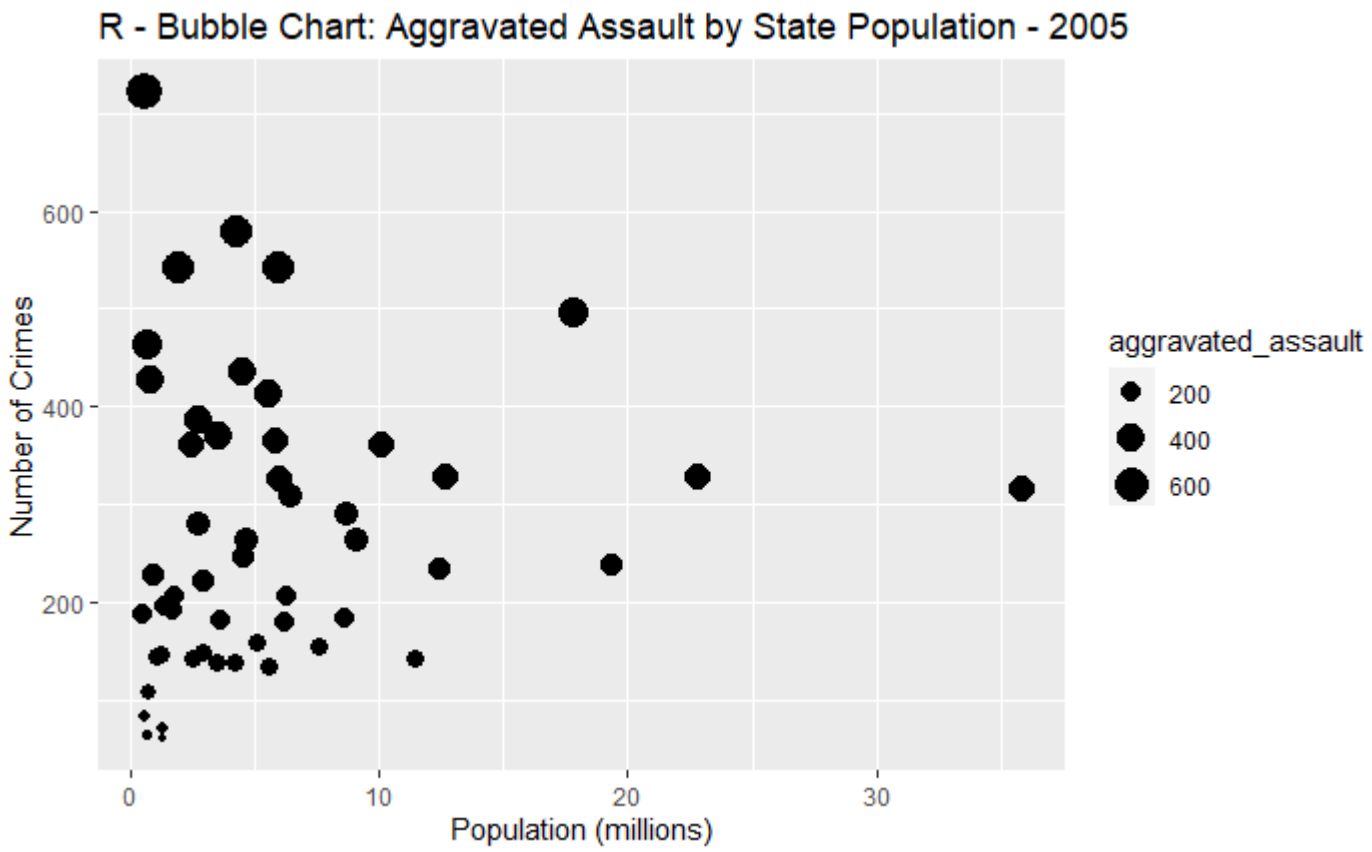
Hide

```
ggplot(data, aes(x=PopinM, y=aggravated_assault)) + geom_point()+
    xlab("Population (millions)")+ylab("Number of Crimes")+
    ggtitle("R - Scatter Plot: Aggravated Assault by State Population - 2005")
```



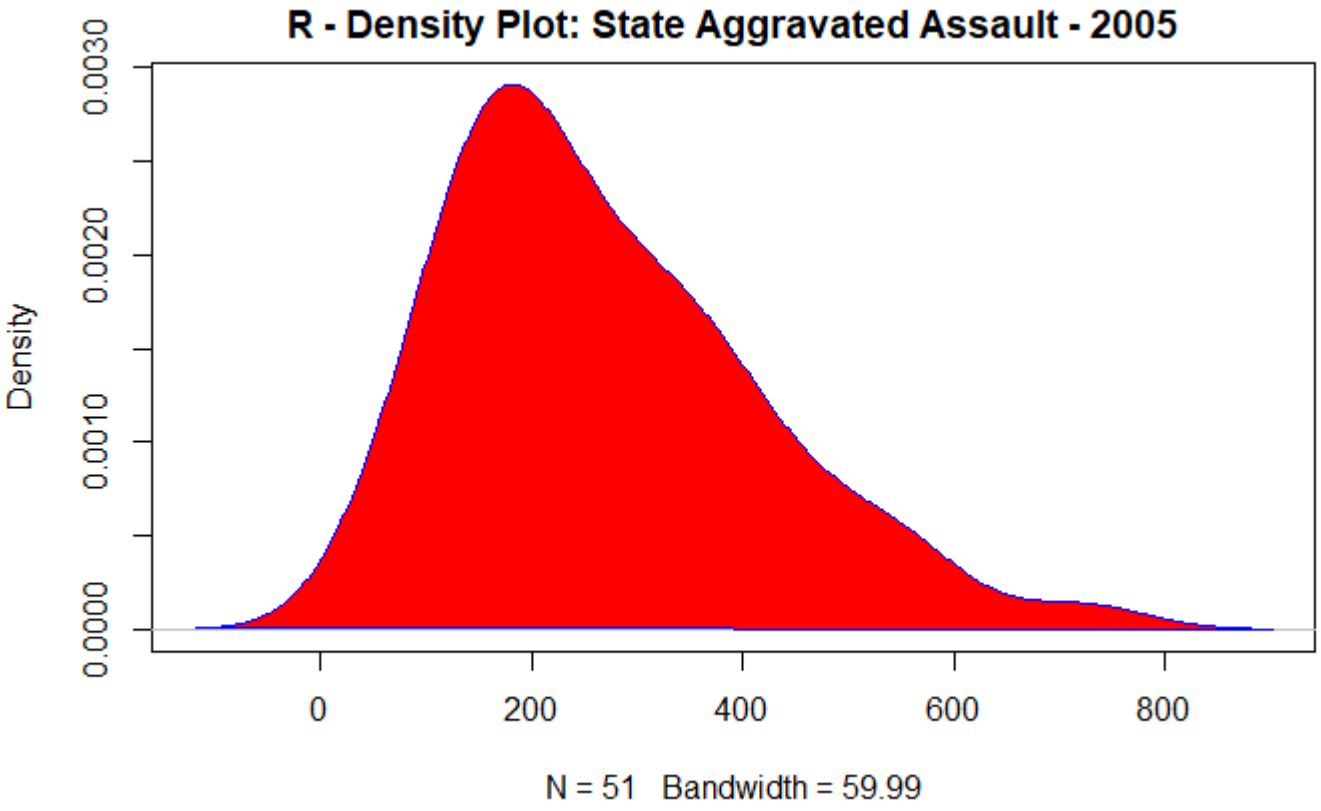
Hide

```
ggplot(data, aes(x=PopinM, y=aggravated_assault)) +
  geom_point(aes(size=aggravated_assault))+
    xlab("Population (millions)")+ylab("Number of Crimes")+
    ggtitle("R - Bubble Chart: Aggravated Assault by State Population - 2005")
```



Hide

```
d <- density(data$aggravated_assault)
plot(d, main="R - Density Plot: State Aggravated Assault - 2005")
polygon(d, col="red", border="blue")
```


R - Density Plot: State Aggravated Assault - 2005
Density
0.0000
0.0010
0.0020
0.0030
0
200
400
600
800
N = 51 Bandwidth = 59.99