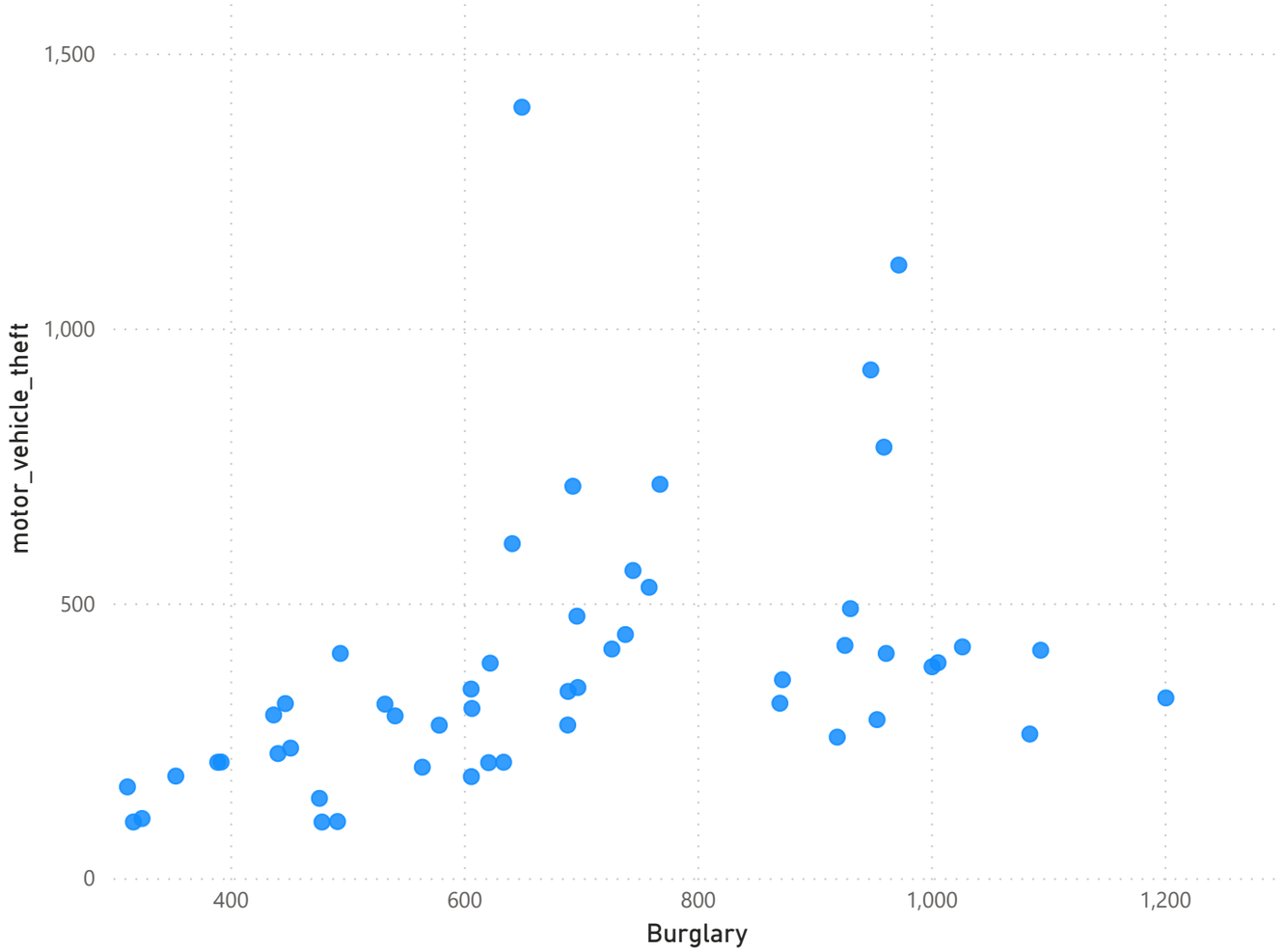
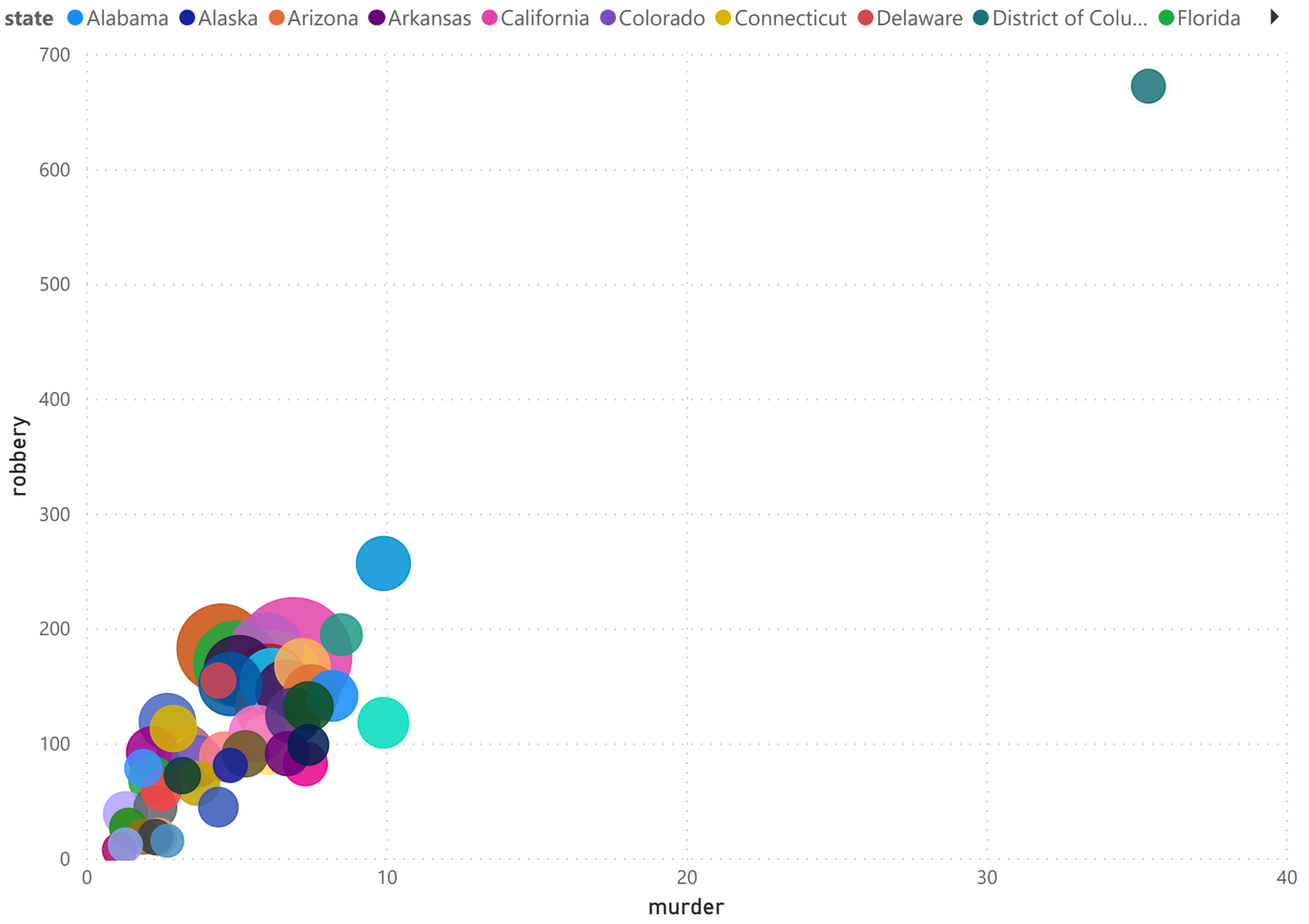
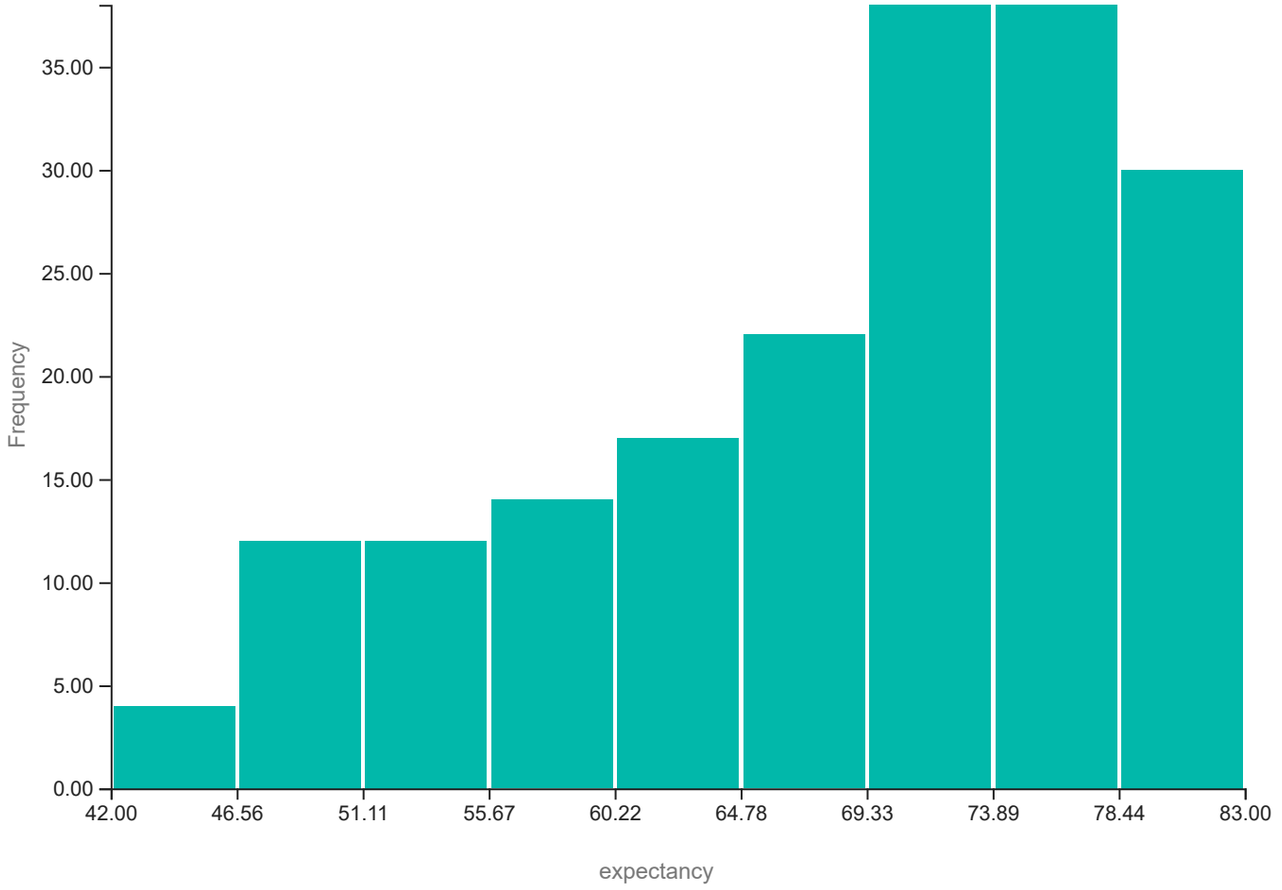
PowerBI-ScatterPlot



PowerBI-Bubble chart



PowerBI-Density Plot



Python plots

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import chart_studio.plotly as py
import cufflinks as cf
import seaborn as sns
```

In [2]:
```
df = pd.read_csv("crimerates-by-state-2005.csv")
```

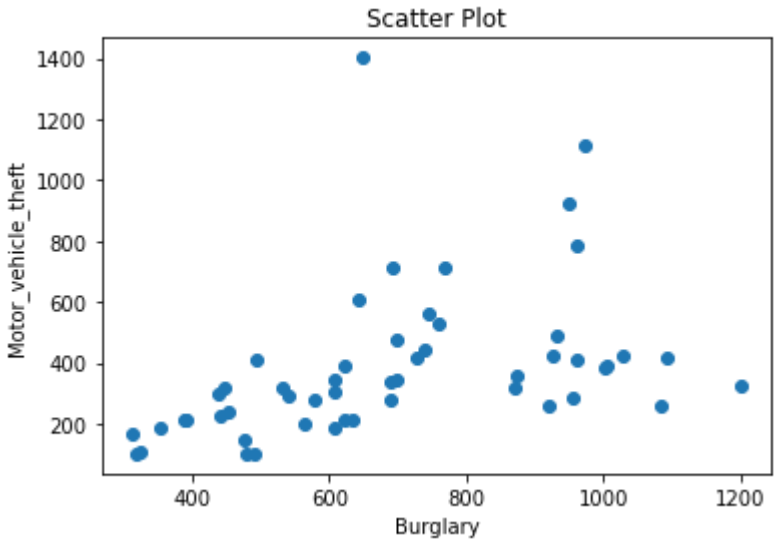
In [3]:
```
df.head()
```

Out[3]:

	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	
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	

Python - Scatter plot

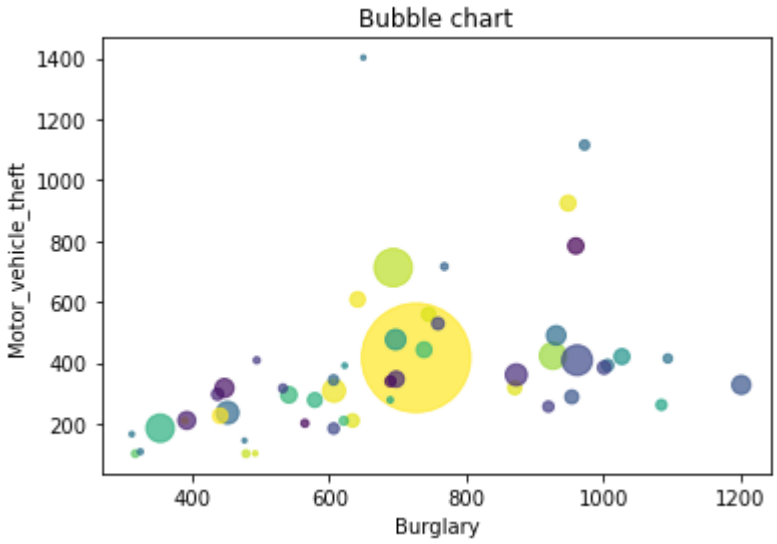
In [4]:
```
plt.scatter(x=df['burglary'], y=df['motor_vehicle_theft'])
plt.title('Scatter Plot')
plt.xlabel('Burglary')
plt.ylabel('Motor_vehicle_theft')
plt.show()
```



Python - bubble chart

In [5]:
```
x = df['burglary']
y = df['motor_vehicle_theft']
z = df['population']/100000
colors = np.random.rand(52)
plt.scatter(x=x, y=y, s=z, c=colors, alpha=0.7)
plt.xlabel("Burglary")
plt.ylabel("Motor_vehicle_theft")
plt.title("Bubble chart")
```

Out[5]:
```
Text(0.5, 1.0, 'Bubble chart')
```



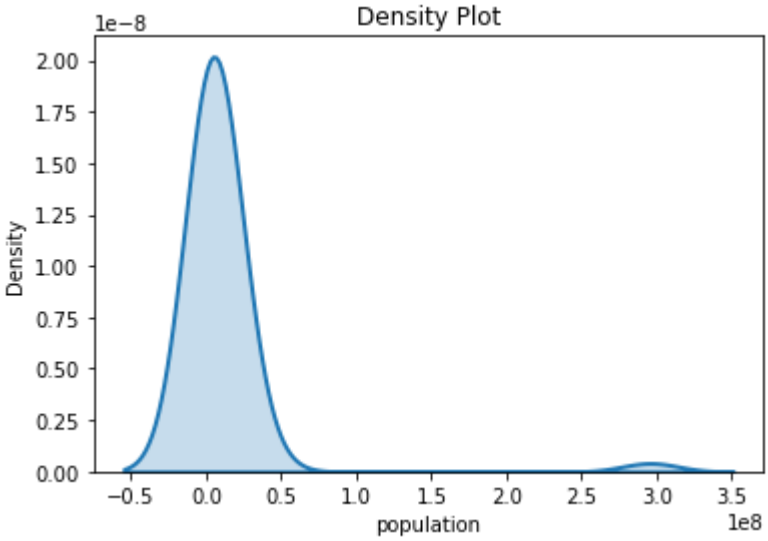
Python - Density Map

In [6]:
```
sns.distplot(df['population'], hist = False, kde = True,
                 kde_kws = {'shade': True, 'linewidth': 2})
plt.title("Density Plot")
```

```
C:\Users\meena\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning:
```

```
`distplot` is a deprecated function and will be removed in a future version. Please adap
t your code to use either `displot` (a figure-level function with similar flexibility) o
r `kdeplot` (an axes-level function for kernel density plots).
```

Out[6]:
```
Text(0.5, 1.0, 'Density Plot')
```



R - Plots

In [1]:
```
library('magrittr')
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
```

In [2]:
```
crimerate <- read.csv2(paste('crimerates-by-state-2005.csv',sep=''), header=TRUE, sep='
  dplyr::mutate(murder = as.numeric(murder),
                forcible_rape = as.numeric(forcible_rape),
                robbery = as.numeric(robbery),
                aggravated_assault = as.numeric(aggravated_assault),
                burglary = as.numeric(burglary),
                larceny_theft = as.numeric(larceny_theft),
                motor_vehicle_theft = as.numeric(motor_vehicle_theft),
                population = as.integer(population),
                total_crime = murder+
                  forcible_rape+
                  robbery+
                  aggravated_assault+
                  burglary+
                  larceny_theft+
                  motor_vehicle_theft,
                state_cont = rank(state, ties.method = 'first'))

head(crimerate)
```

A data.frame: 6 × 11

	state	murder	forcible_rape	robbery	aggravated_assault	burglary	larceny_theft	motor_vehicle
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	
1	United States	5.6	31.7	140.7	291.1	726.7	2286.3	
2	Alabama	8.2	34.3	141.4	247.8	953.8	2650.0	
3	Alaska	4.8	81.1	80.9	465.1	622.5	2599.1	
4	Arizona	7.5	33.8	144.4	327.4	948.4	2965.2	
5	Arkansas	6.7	42.9	91.1	386.8	1084.6	2711.2	
6	California	6.9	26.0	176.1	317.3	693.3	1916.5	

In [3]:
```
state_detail <- xlsx::read.xlsx2(paste("states_detail.xlsx",sep=''),sheetIndex = 1, str

state_detail = rename(state_detail, "state"="full_name")
state_detail
```

```
java.home option:

JAVA_HOME environment variable: C:\Users\meena\anaconda3\Library\lib\jvm

Warning message in fun(libname, pkgname):
"Java home setting is INVALID, it will be ignored.
Please do NOT set it unless you want to override system settings."
```

A data.frame: 50 × 4

name_caps	state	abbr	region
<chr>	<chr>	<chr>	<chr>
ALABAMA	Alabama	AL	Rest of USA
ALASKA	Alaska	AK	Rest of USA
ARIZONA	Arizona	AZ	Rest of USA
ARKANSAS	Arkansas	AR	Rest of USA
CALIFORNIA	California	CA	Rest of USA
COLORADO	Colorado	CO	Rest of USA
CONNECTICUT	Connecticut	CT	Rest of USA
DELAWARE	Delaware	DE	Rest of USA
FLORIDA	Florida	FL	Rest of USA
GEORGIA	Georgia	GA	Rest of USA
HAWAII	Hawaii	HI	Rest of USA
IDAHO	Idaho	ID	Rest of USA
ILLINOIS	Illinois	IL	MidWest
INDIANA	Indiana	IN	MidWest
IOWA	Iowa	IA	MidWest
KANSAS	Kansas	KS	MidWest
KENTUCKY	Kentucky	KY	Rest of USA
LOUISIANA	Louisiana	LA	Rest of USA
MAINE	Maine	ME	Rest of USA
MARYLAND	Maryland	MD	Rest of USA
MASSACHUSETTS	Massachusetts	MA	Rest of USA
MICHIGAN	Michigan	MI	MidWest
MINNESOTA	Minnesota	MN	MidWest

name_caps	state	abbr	region
<chr>	<chr>	<chr>	<chr>
MISSISSIPPI	Mississippi	MS	Rest of USA
MISSOURI	Missouri	MO	MidWest
MONTANA	Montana	MT	Rest of USA
NEBRASKA	Nebraska	NE	MidWest
NEVADA	Nevada	NV	Rest of USA
NEW HAMPSHIRE	New Hampshire	NH	Rest of USA
NEW JERSEY	New Jersey	NJ	Rest of USA
NEW MEXICO	New Mexico	NM	Rest of USA
NEW YORK	New York	NY	Rest of USA
NORTH CAROLINA	North Carolina	NC	Rest of USA
NORTH DAKOTA	North Dakota	ND	MidWest
OHIO	Ohio	OH	MidWest
OKLAHOMA	Oklahoma	OK	Rest of USA
OREGON	Oregon	OR	Rest of USA
PENNSYLVANIA	Pennsylvania	PA	Rest of USA
RHODE ISLAND	Rhode Island	RI	Rest of USA
SOUTH CAROLINA	South Carolina	SC	Rest of USA
SOUTH DAKOTA	South Dakota	SD	MidWest
TENNESSEE	Tennessee	TN	Rest of USA
TEXAS	Texas	TX	Rest of USA
UTAH	Utah	UT	Rest of USA
VERMONT	Vermont	VT	Rest of USA
VIRGINIA	Virginia	VA	Rest of USA
WASHINGTON	Washington	WA	Rest of USA
WEST VIRGINIA	West Virginia	WV	Rest of USA
WISCONSIN	Wisconsin	WI	MidWest
WYOMING	Wyoming	WY	Rest of USA

In [4]:

```
state_detail <- mutate(state_detail, midwest = case_when(state_detail['region'] == 'Mid
state_detail
```

A data.frame: 50 × 5

name_caps	state	abbr	region	midwest
<chr>	<chr>	<chr>	<chr>	<chr>

name_caps	state	abbr	region	midwest
<chr>	<chr>	<chr>	<chr>	<chr>
ALABAMA	Alabama	AL	Rest of USA	0
ALASKA	Alaska	AK	Rest of USA	0
ARIZONA	Arizona	AZ	Rest of USA	0
ARKANSAS	Arkansas	AR	Rest of USA	0
CALIFORNIA	California	CA	Rest of USA	0
COLORADO	Colorado	CO	Rest of USA	0
CONNECTICUT	Connecticut	CT	Rest of USA	0
DELAWARE	Delaware	DE	Rest of USA	0
FLORIDA	Florida	FL	Rest of USA	0
GEORGIA	Georgia	GA	Rest of USA	0
HAWAII	Hawaii	HI	Rest of USA	0
IDAHO	Idaho	ID	Rest of USA	0
ILLINOIS	Illinois	IL	MidWest	1
INDIANA	Indiana	IN	MidWest	1
IOWA	Iowa	IA	MidWest	1
KANSAS	Kansas	KS	MidWest	1
KENTUCKY	Kentucky	KY	Rest of USA	0
LOUISIANA	Louisiana	LA	Rest of USA	0
MAINE	Maine	ME	Rest of USA	0
MARYLAND	Maryland	MD	Rest of USA	0
MASSACHUSETTS	Massachusetts	MA	Rest of USA	0
MICHIGAN	Michigan	MI	MidWest	1
MINNESOTA	Minnesota	MN	MidWest	1
MISSISSIPPI	Mississippi	MS	Rest of USA	0
MISSOURI	Missouri	MO	MidWest	1
MONTANA	Montana	MT	Rest of USA	0
NEBRASKA	Nebraska	NE	MidWest	1
NEVADA	Nevada	NV	Rest of USA	0
NEW HAMPSHIRE	New Hampshire	NH	Rest of USA	0
NEW JERSEY	New Jersey	NJ	Rest of USA	0
NEW MEXICO	New Mexico	NM	Rest of USA	0
NEW YORK	New York	NY	Rest of USA	0

name_caps	state	abbr	region	midwest
<chr>	<chr>	<chr>	<chr>	<chr>
NORTH CAROLINA	North Carolina	NC	Rest of USA	0
NORTH DAKOTA	North Dakota	ND	MidWest	1
OHIO	Ohio	OH	MidWest	1
OKLAHOMA	Oklahoma	OK	Rest of USA	0
OREGON	Oregon	OR	Rest of USA	0
PENNSYLVANIA	Pennsylvania	PA	Rest of USA	0
RHODE ISLAND	Rhode Island	RI	Rest of USA	0
SOUTH CAROLINA	South Carolina	SC	Rest of USA	0
SOUTH DAKOTA	South Dakota	SD	MidWest	1
TENNESSEE	Tennessee	TN	Rest of USA	0
TEXAS	Texas	TX	Rest of USA	0
UTAH	Utah	UT	Rest of USA	0
VERMONT	Vermont	VT	Rest of USA	0
VIRGINIA	Virginia	VA	Rest of USA	0
WASHINGTON	Washington	WA	Rest of USA	0
WEST VIRGINIA	West Virginia	WV	Rest of USA	0
WISCONSIN	Wisconsin	WI	MidWest	1
WYOMING	Wyoming	WY	Rest of USA	0

In [5]:

```
# Remove USA as a state and add region
crimerate_states <- crimerate %>%
  dplyr::left_join(state_detail, by='state') %>%
  dplyr::filter(state !='United States') %>%
  dplyr::mutate(midwest = as.integer(midwest))


crimerate_states[is.na(crimerate_states)] <- 0


head(crimerate_states)
```

A data.frame: 6 × 15

	state	murder	forcible_rape	robbery	aggravated_assault	burglary	larceny_theft	motor_vehicle
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	
1	Alabama	8.2	34.3	141.4	247.8	953.8	2650.0	
2	Alaska	4.8	81.1	80.9	465.1	622.5	2599.1	

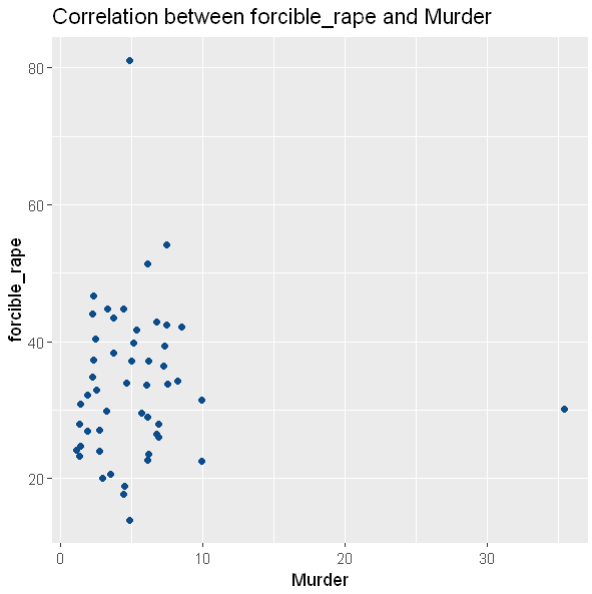
	state	murder	forcible_rape	robbery	aggravated_assault	burglary	larceny_theft	motor_vehicle
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	
3	Arizona	7.5	33.8	144.4	327.4	948.4	2965.2	
4	Arkansas	6.7	42.9	91.1	386.8	1084.6	2711.2	
5	California	6.9	26.0	176.1	317.3	693.3	1916.5	
6	Colorado	3.7	43.4	84.6	264.7	744.8	2735.2	

R - Scatter Plot

In [6]:

```
options(repr.plot.width = 5, repr.plot.height = 5)

# Plot
ggplot2::ggplot(data = crimerate_states) +
  ggplot2::aes(x = murder, y = forcible_rape) +
  ggplot2::geom_point(color = "#0c4c8a") +
  ggplot2::labs(title = "Correlation between forcible_rape and Murder",
       x = "Murder",
       y = "forcible_rape") +
  ggplot2::theme_grey()
```



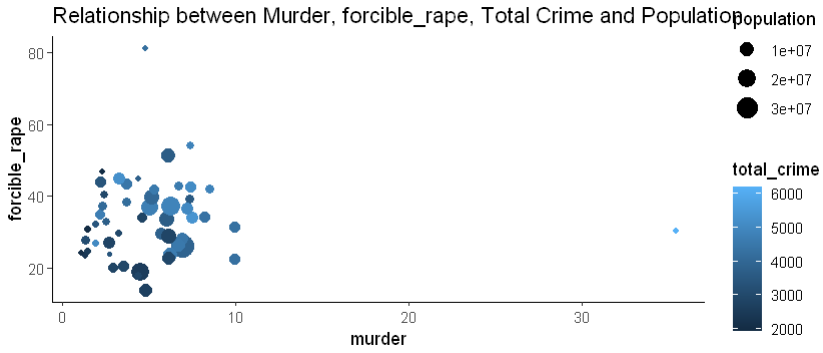
R - Bubble chart

In [7]:

```
# Format graph size
options(repr.plot.width = 7, repr.plot.height = 3)

# Plot
ggplot2::ggplot(data = crimerate_states) +
```

```
  ggplot2::aes(x = murder, y = forcible_rape, color = total_crime, size = population) +
  ggplot2::geom_point() +
  ggplot2::labs(title = "Relationship between Murder, forcible_rape, Total Crime and Po
  ggplot2::theme_classic()
```



R - Density Plot

In [8]:

```
# For this I want to compare total crime rate of mid west states to the rest of the sta
crimerate_compare <- crimerate_states[,-1]
rownames(crimerate_compare) <- crimerate_states[,1]

ggplot2::ggplot(crimerate_compare) +
  ggplot2::aes(total_crime, fill=as.factor(crimerate_compare$midwest)) +
  ggplot2::geom_density(alpha = 0.3) +
  ggplot2::labs(x='Total Crime', y='Density') +
  ggplot2::theme_classic() +
  ggplot2::guides(fill=ggplot2::guide_legend(title="Mid West flag"))
```

```
Warning message:
"Use of `crimerate_compare$midwest` is discouraged. Use `midwest` instead."
```