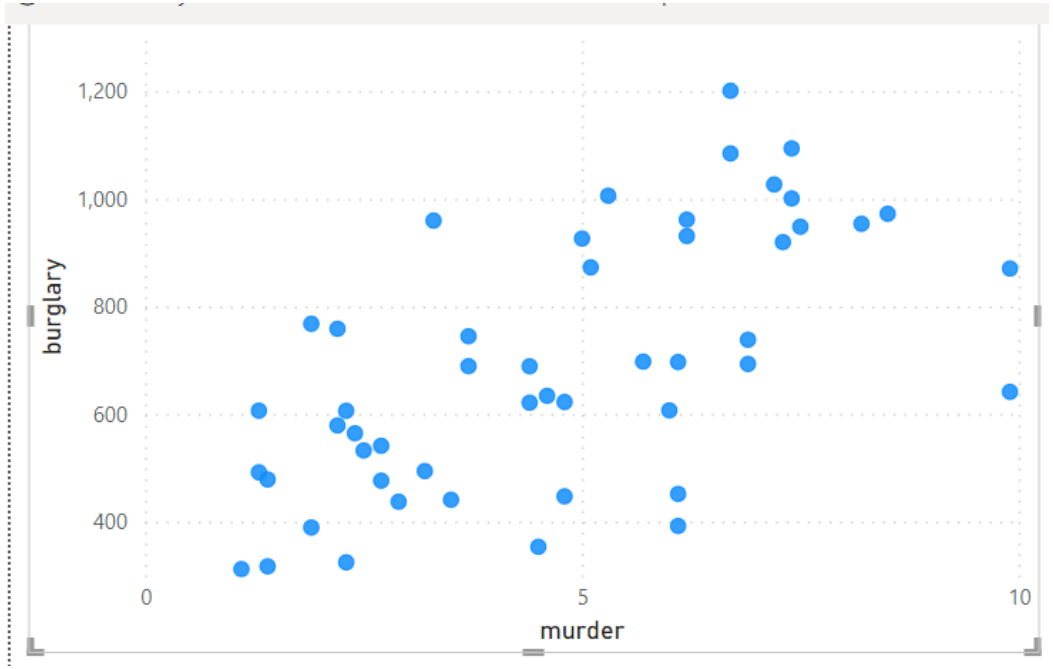
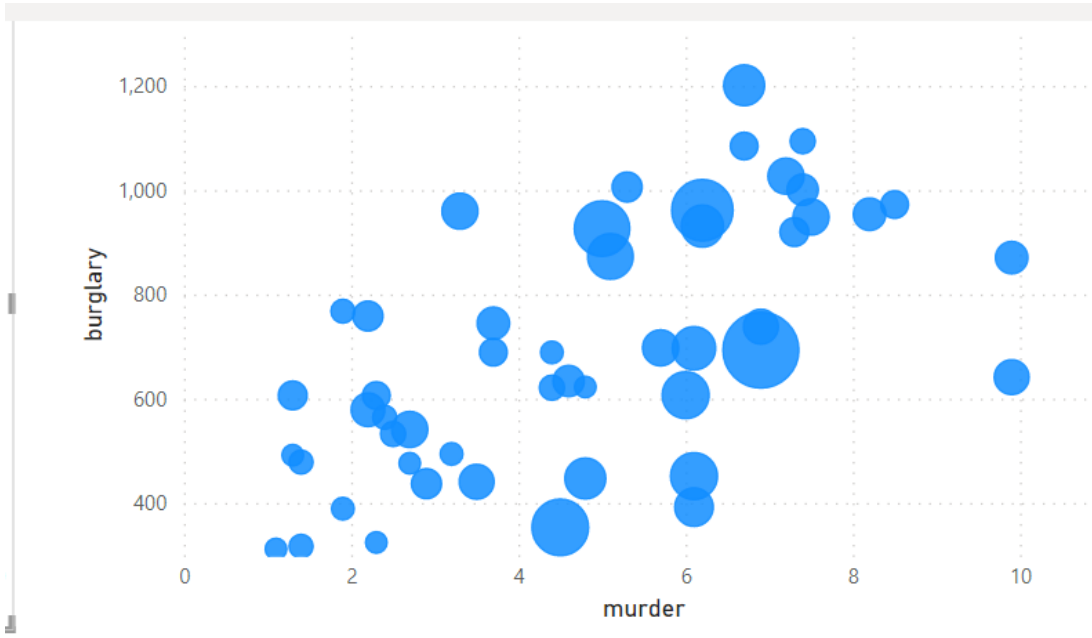
DSC 640 – Week 7 & 8
Michael Ersevim

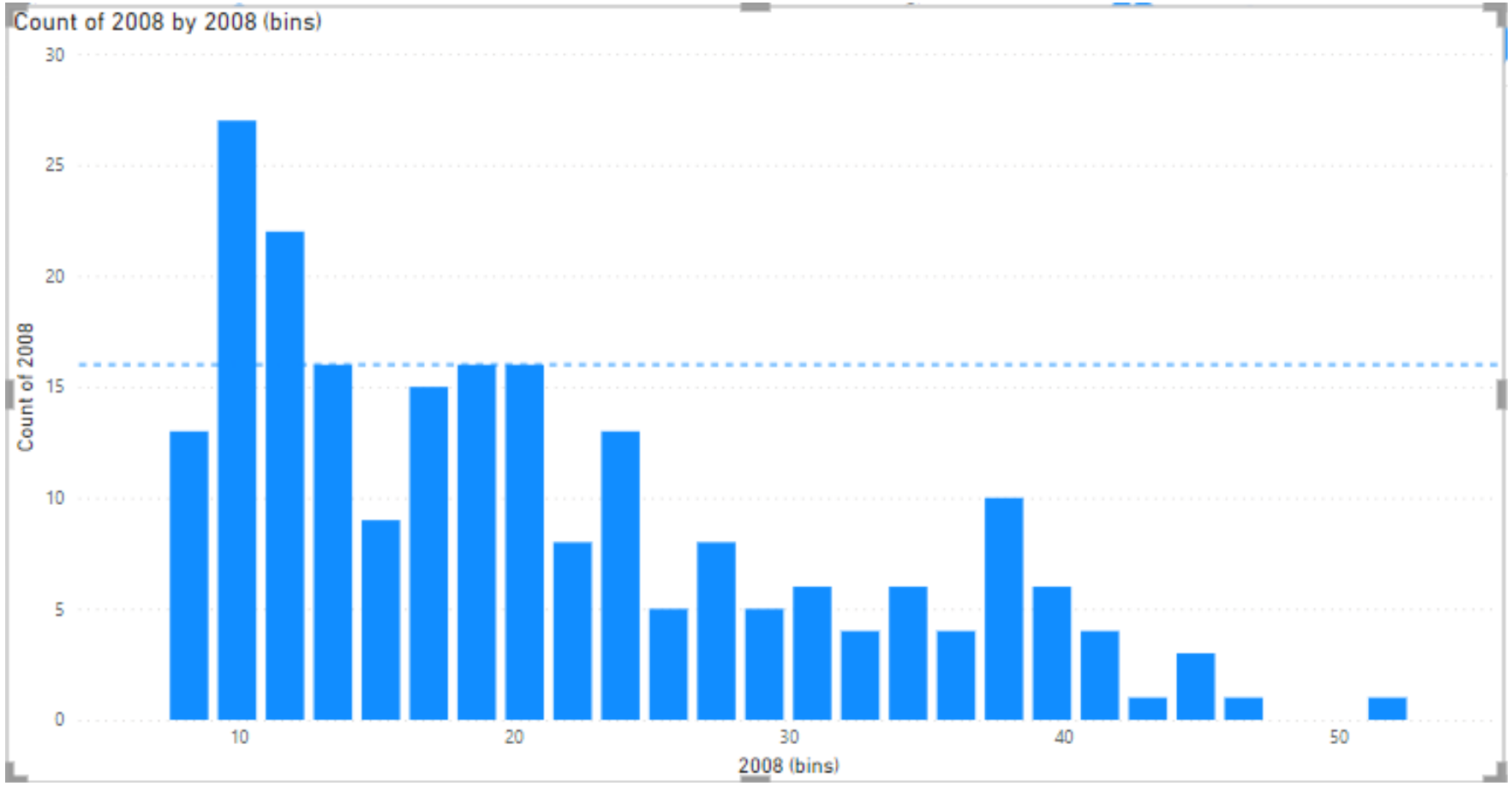
Power BI: Scatter graph and bubble graph



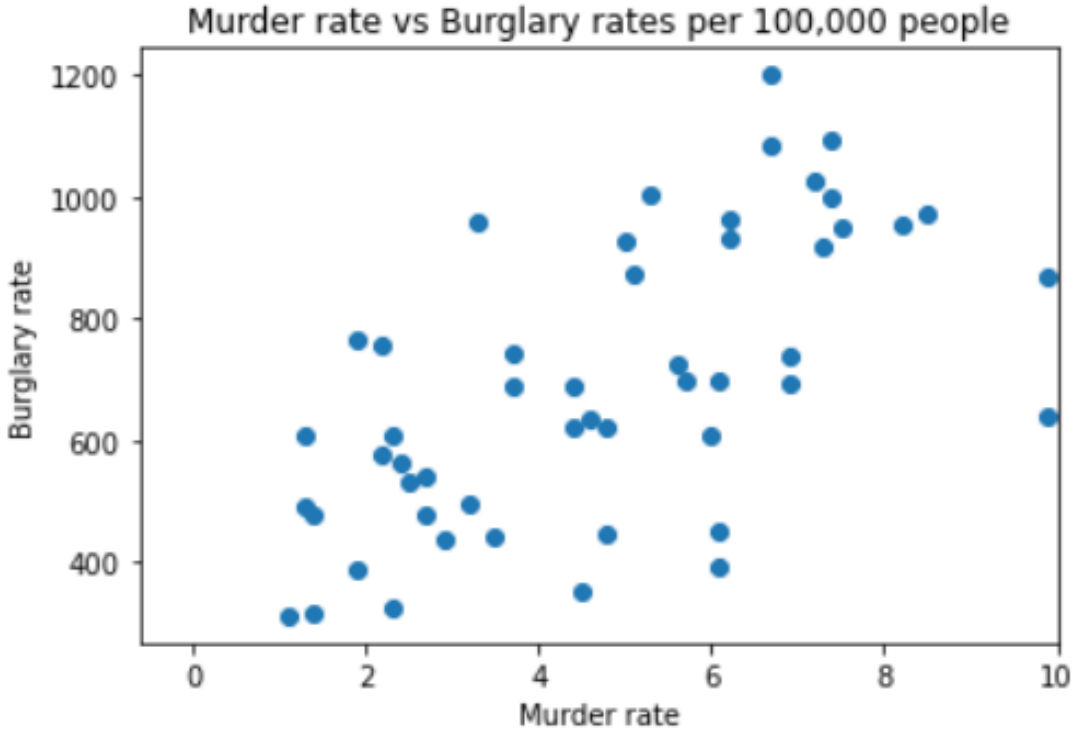


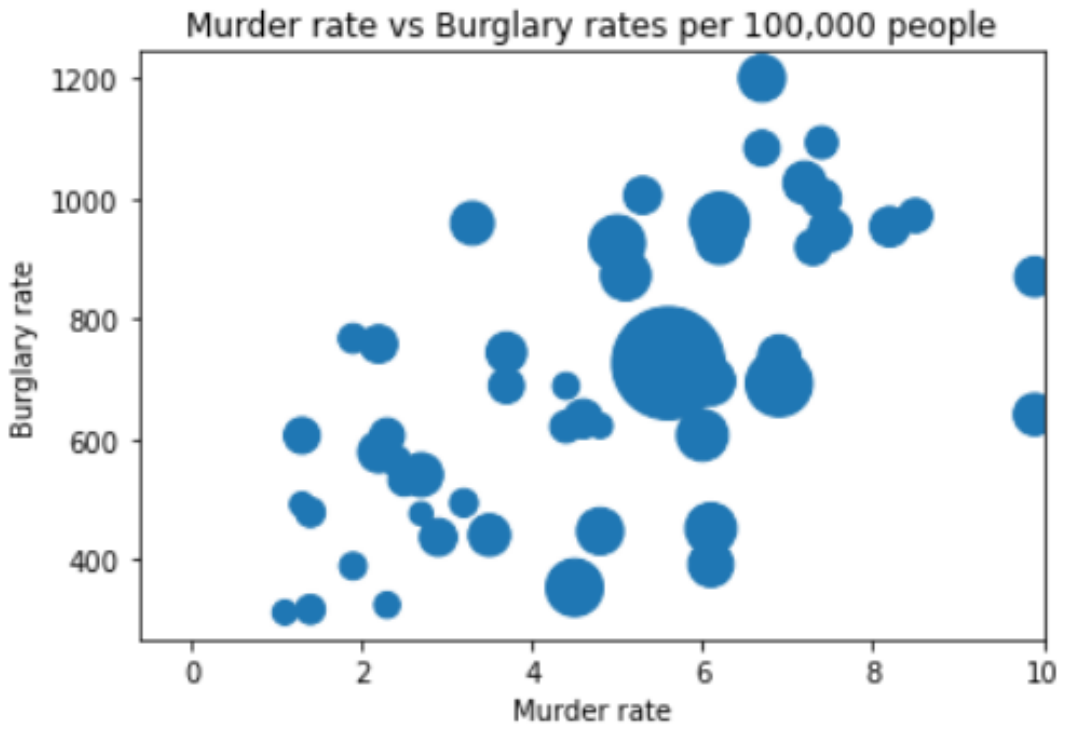
Power BI: Density graph - Bar Graph *proxy*

I know this isn't a PDF however, I couldn't figure it out in Power BI so a 25 bin histogram is a proxy to the shape of the PDF or PMF graph.

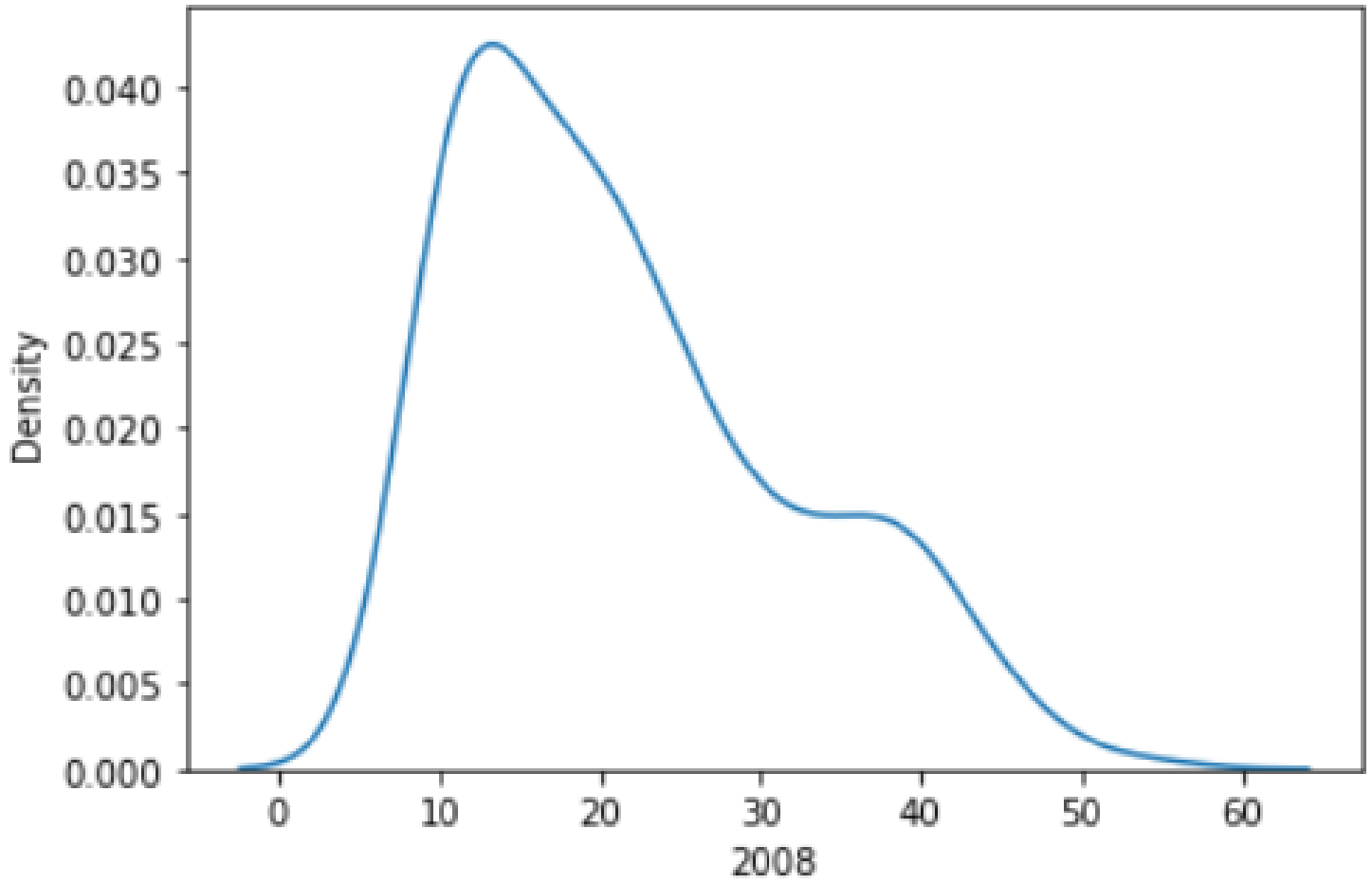


Python: Scatter and Bubble graphs





Python: PDF graph of birth rates from 2008: Countries around the world



Python: CODE for generating graphs

DSCC640 - Michael Ersevim - Week 7&8 assignment

In [20]:
```
# Call in libraries
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from statsmodels.distributions.empirical_distribution import ECDF
```

In [3]:
```
# read in birth rate ddata
dfbr = pd.read_excel('C:\\Users\\Kate\\Documents\\Bellevue DS classes\\DSC640\\birth-rate.xlsx')
dfbr.head()
```

Out[3]:

	Country	1960	1961	1962	1963	1964	1965	1966	1967	1968	...	1999	2000	2001	2002	2003	2004	2005	2006	2007
0	Aruba	36.400	35.179	33.863	32.459	30.994	29.513	28.069	26.721	25.518	...	15.024	14.528	14.041	13.579	13.153	12.772	12.441	12.159	11.919
1	Afghanistan	52.201	52.206	52.208	52.204	52.192	52.168	52.130	52.076	52.006	...	51.229	50.903	50.486	49.984	49.416	48.803	48.177	47.575	47.023
2	Angola	54.432	54.394	54.317	54.199	54.040	53.836	53.585	53.296	52.984	...	48.662	48.355	48.005	47.545	46.936	46.184	45.330	44.444	43.607
3	Albania	40.886	40.312	39.604	38.792	37.913	37.008	36.112	35.245	34.421	...	17.713	16.850	16.081	15.444	14.962	14.644	14.485	14.464	14.534
4	Netherlands Antilles	32.321	30.987	29.618	28.229	26.849	25.518	24.280	23.173	22.230	...	15.809	15.412	15.096	14.824	14.565	14.309	14.051	13.790	13.532

5 rows × 50 columns

Python: CODE for generating graphs

In [4]:
```
# read in crime rate file
dfcr = pd.read_excel('C:\\Users\\Kate\\Documents\\Bellevue DS classes\\DSC640\\crimeRatesByState2005.xlsx')
dfcr.head()
```

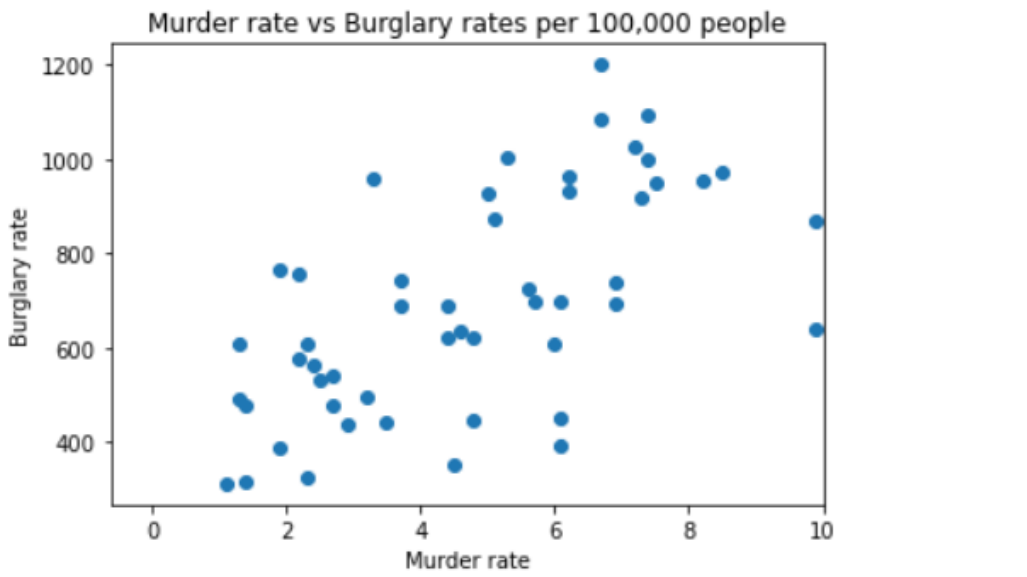
Out[4]:

	state	murder	forcible_rape	robbery	aggravated_assault	burglary	larceny_theft	motor_vehicle_theft	population
0	United States	5.6	31.7	140.7	291.1	726.7	2286.3	416.7	295753151
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

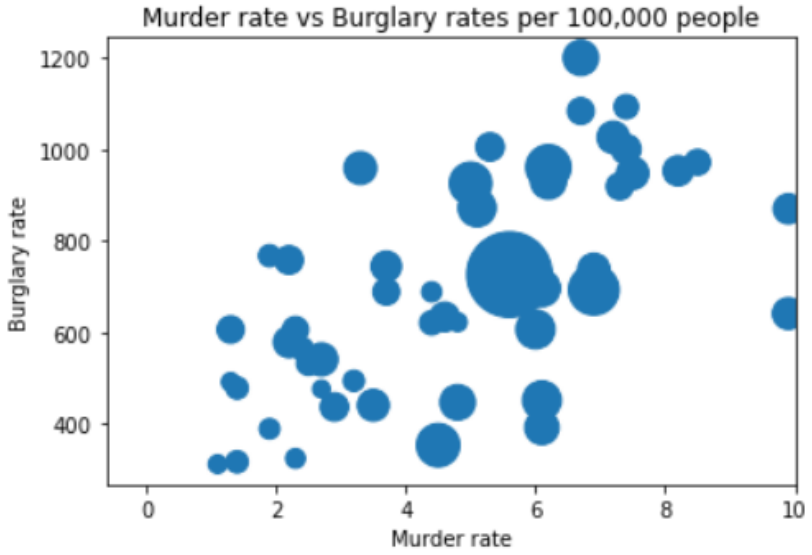
In [18]:
```
# define x, y and bubble size for bubblegraph
x = dfcr['murder']
y = dfcr['burglary']
size = dfcr['population']**0.5*.1 # Square root tule AND scale factor for reasonable size
```

In [16]:
```
# Make scatter plot
plt.scatter(x, y)
plt.title('Murder rate vs Burglary rates per 100,000 people')
plt.xlabel('Murder rate')
plt.ylabel('Burglary rate')
plt.xlim(xmax=10)
plt.show()
```

Python: CODE for generating graphs




```
In [19]: # Make bubblegraph plot
         plt.scatter(x, y, s=size)
         plt.title('Murder rate vs Burglary rates per 100,000 people')
         plt.xlabel('Murder rate')
         plt.ylabel('Burglary rate')
         plt.xlim(xmax=10)
         plt.show()
```

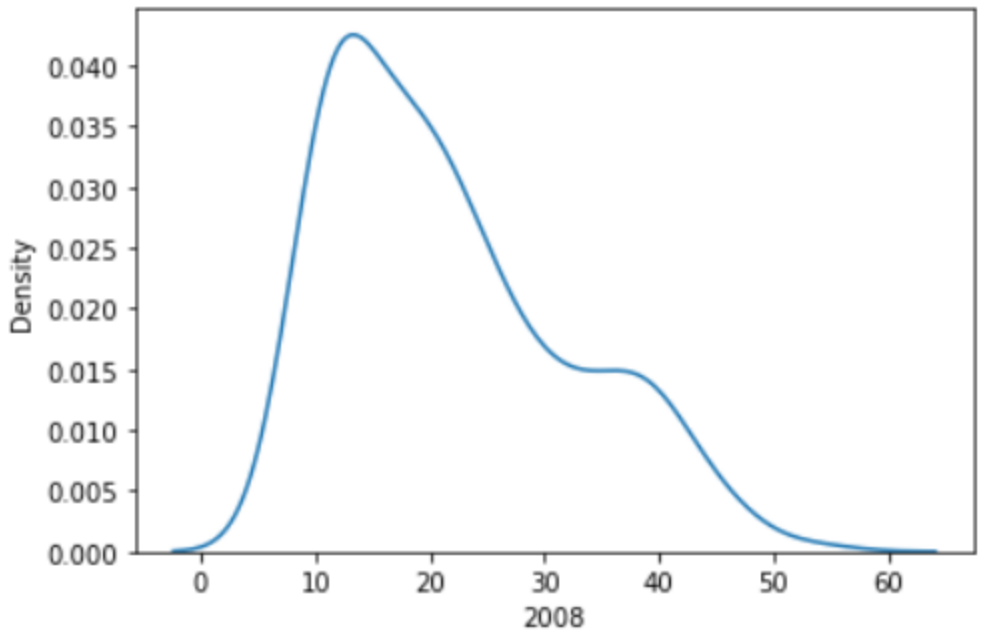



```
In [23]: # define points for PDF

         ecdf = ECDF(dfbr[2008])

         # plot the cdf
         plt.plot(ecdf.x, ecdf.y)
         plt.show()
         #Got a CDF...next, use KDE to get the PMF
```

Python: CODE for generating graphs

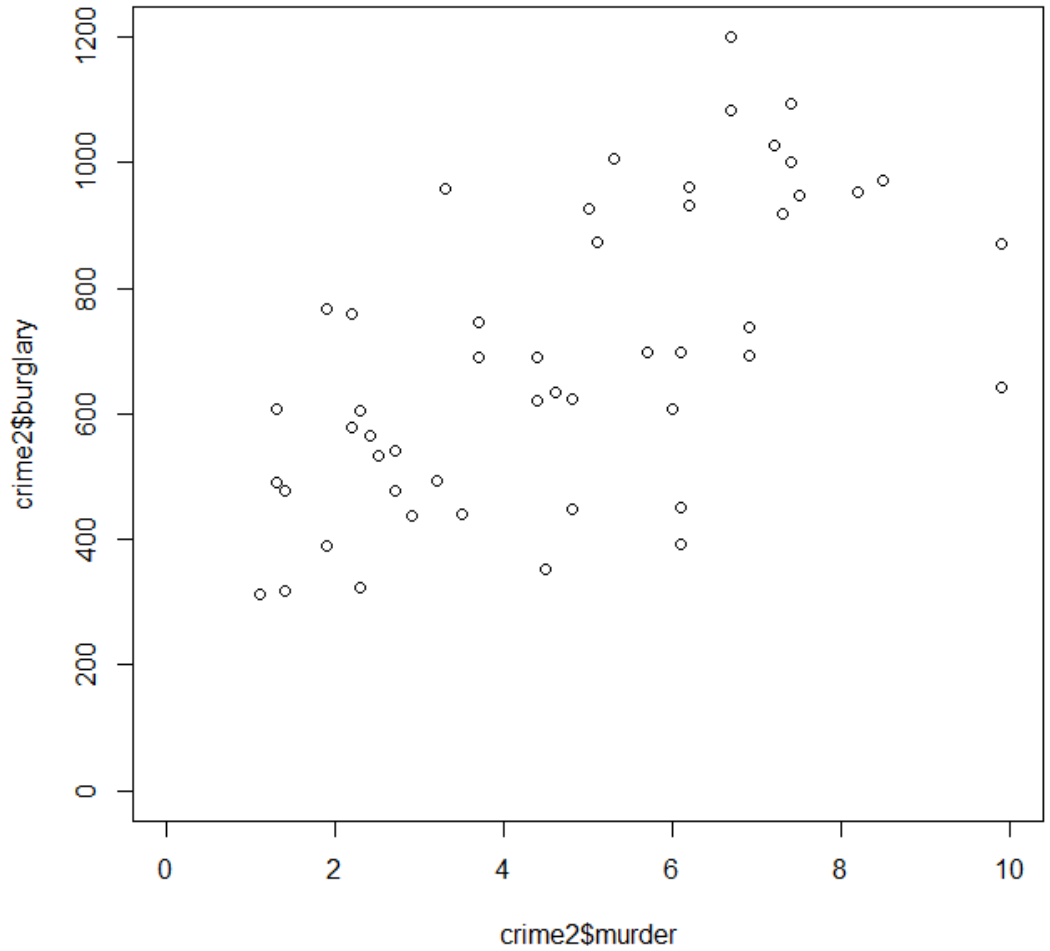
In [24]:
```
import seaborn as sns
# Plotting the KDE Plot
sns.kdeplot(dfbr[2008]) # Uses the values from 2008 to create the PDF
```

Out[24]:
```
<AxesSubplot:xlabel='2008', ylabel='Density'>
```

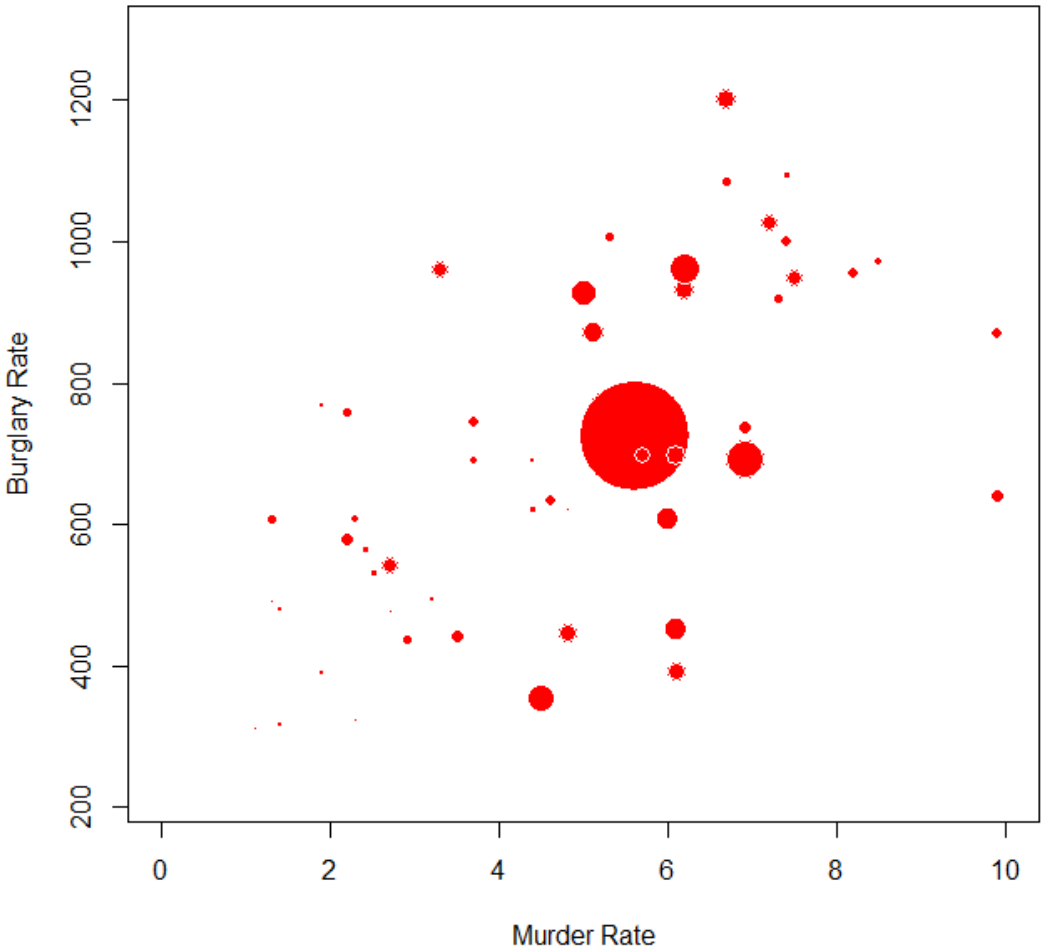


R-graphs

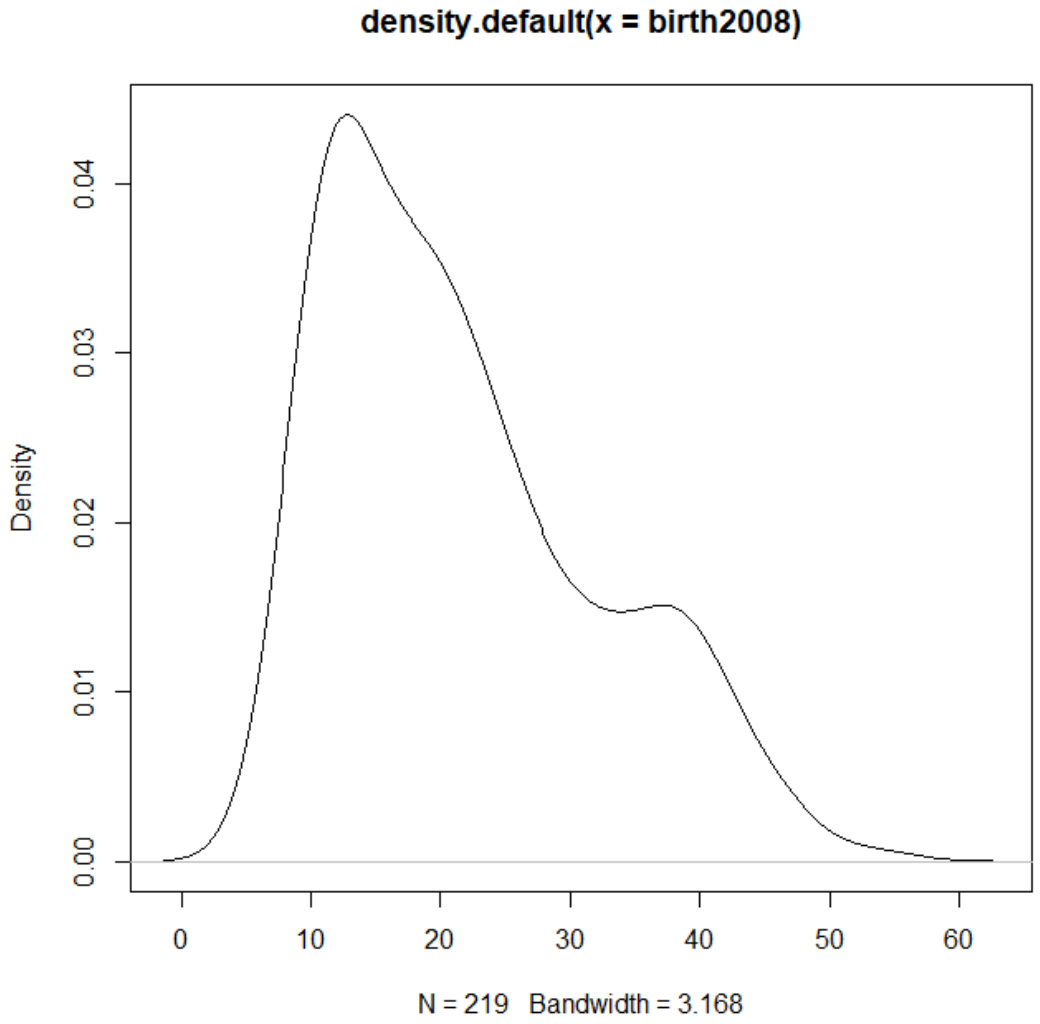
Scattergraph of the murder rate plotted against burglary rate to check for correlation (both per 100,000 people)



Bubble graph of murder vs burglary rate. Area of bubble should be proportional to population of state, but sqrt of radius function seems not to be working...



R: Probability Density graph of birth rates (per 1,000 people) across countries from around the world



R: Code for prior graphs
Based upon code supplied with 'Visualize This (N. Yau)', with a few additions as directed by the text in the book

```
setwd("C:/Users/Kate/Documents/Bellevue DS classes/DSC640")
# Load the data
crime <- read.csv('crimeRatesByState2005.csv', sep=",", header=TRUE)

# Remove US total and DC
crime2 <- crime[crime$state != "District of Columbia",]
crime2 <- crime2[crime2$state != "United States",]

# Scatterplot for murder and burglary
plot(crime$murder, crime$burglary)
plot(crime2$murder, crime2$burglary)
plot(crime2$murder, crime2$burglary, xlim=c(0,10), ylim=c(0, 1200))

# make bubble chart
radius <- sqrt( crime$population/ pi )
symbols(crime$murder, crime$burglary, circles=radius, inches=0.35,
      fg='white', bg='red', xlab='Murder Rate', ylab='Burglary Rate', xlim = c(0,10))

# Scatterplot matrix
plot(crime2[,2:9])
```