DSC640 Assignment # 4.2 - Week_07_08

Ayachit Madhukar

Jan 24 2022

```r
# setting current working diirectory
setwd("/Users/madhukarayachit/DSC640")
```

```r
##load required libraries
library(ggplot2)      ## to draw plots
library(ggrepel)      ## to add labels to bubble chart
```

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.

```r
df1 <- read.csv('Data/ex4-2/crimerates-by-state-2005.csv')
df2 <- read.csv('Data/ex4-2/life-expectancy.csv')
head(df1, 10)
```

```
##                   state murder forcible_rape robbery aggravated_assault
## 1        United States    5.6          31.7   140.7              291.1
## 2              Alabama    8.2          34.3   141.4              247.8
## 3               Alaska    4.8          81.1    80.9              465.1
## 4              Arizona    7.5          33.8   144.4              327.4
## 5             Arkansas    6.7          42.9    91.1              386.8
## 6           California    6.9          26.0   176.1              317.3
## 7             Colorado    3.7          43.4    84.6              264.7
## 8          Connecticut    2.9          20.0   113.0              138.6
## 9             Delaware    4.4          44.7   154.8              428.2
## 10 District of Columbia   35.4          30.2   672.1              721.3
##     burglary larceny_theft motor_vehicle_theft population
## 1      726.7        2286.3               416.7  295753151
## 2      953.8        2650.0               288.3    4545049
## 3      622.5        2599.1               391.0     669488
## 4      948.4        2965.2               924.4    5974834
## 5     1084.6        2711.2               262.1    2776221
## 6      693.3        1916.5               712.8   35795255
## 7      744.8        2735.2               559.5    4660780
## 8      437.1        1824.1               296.8    3477416
## 9      688.9        2144.0               278.5     839906
## 10     649.7        2694.9              1402.3     582049
```

```r
head(df2)
```

```
##              country year expectancy
## 1        Afghanistan 2008         42
```

```
## 2               Albania 2008        73
## 3               Algeria 2008        71
## 4                Angola 2008        46
## 5 Antigua and Barbuda 2008        74
## 6             Argentina 2008        76
```

```r
## from df1, we will filter out records related to "District of Columbia" and
## entire country which has state name as "United States"
df1_temp <- df1[((df1$state != "District of Columbia") & (df1$state != "United States")), ]
## display initial 10 records to ensure filtering has worked as expected
head(df1_temp, 10)
```

```
##          state murder forcible_rape robbery aggravated_assault burglary
## 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
## 7      Colorado    3.7          43.4    84.6              264.7    744.8
## 8   Connecticut    2.9          20.0   113.0              138.6    437.1
## 9      Delaware    4.4          44.7   154.8              428.2    688.9
## 11      Florida    5.0          37.1   169.4              496.6    926.3
## 12      Georgia    6.2          23.6   154.8              264.3    931.0
##    larceny_theft motor_vehicle_theft population
## 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
## 7         2735.2               559.5    4660780
## 8         1824.1               296.8    3477416
## 9         2144.0               278.5     839906
## 11        2658.3               423.3   17783868
## 12        2751.1               490.2    9097428
```

Scatter Plots

```r
## plot Scatter plot using ggplot
ggplot(df1_temp, aes(x = murder, y = burglary)) +
    geom_point() +
    ggtitle('US state wise murder vs burglary') +
    ggrepel::geom_text_repel(aes(label=state),
                        max.overlaps = Inf,
                        box.padding = 0.3) +
    theme_bw()
```

US state wise murder vs burglary
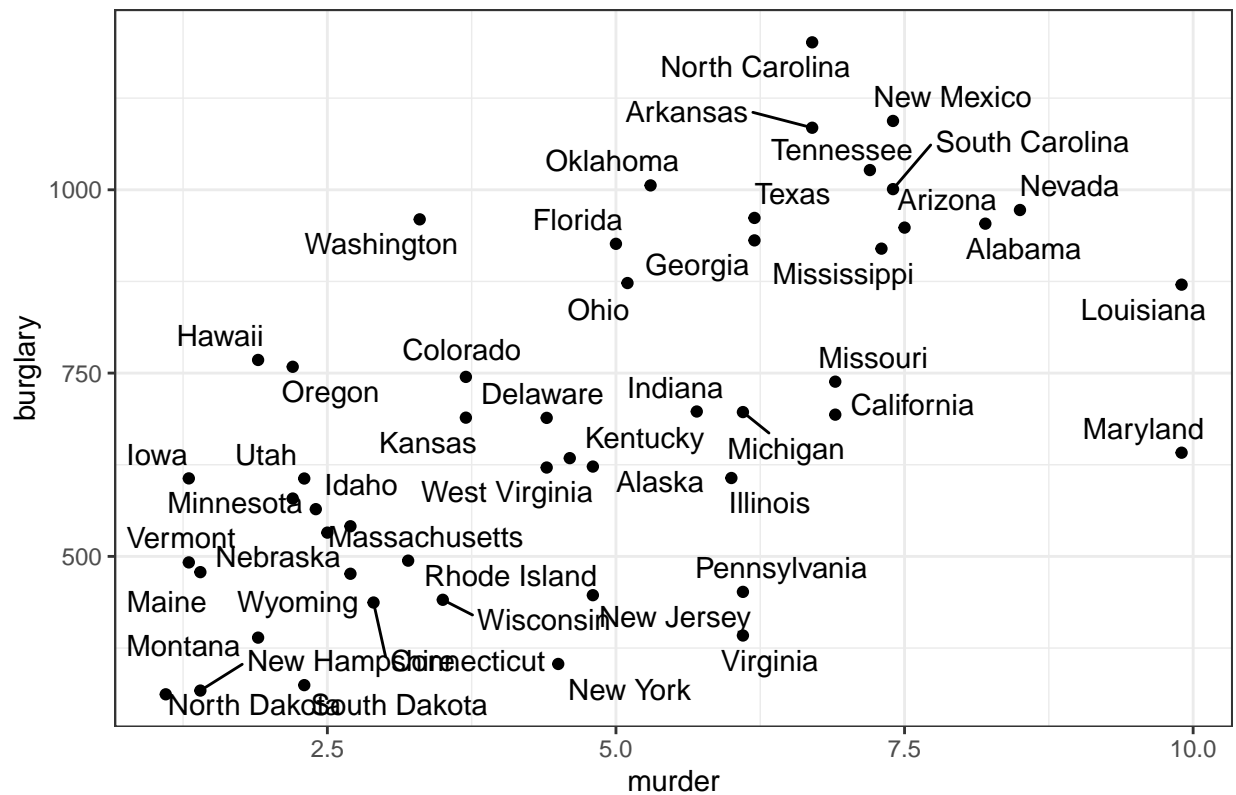
Bubble Plots

```
## plot Bubble chart using ggplot
ggplot(df1_temp, aes(x = murder, y = burglary, size = population)) +
    geom_point(alpha = 0.8) +
    scale_size(name = "Population") +
    ggrepel::geom_text_repel(aes(murder, burglary, label=state))
```
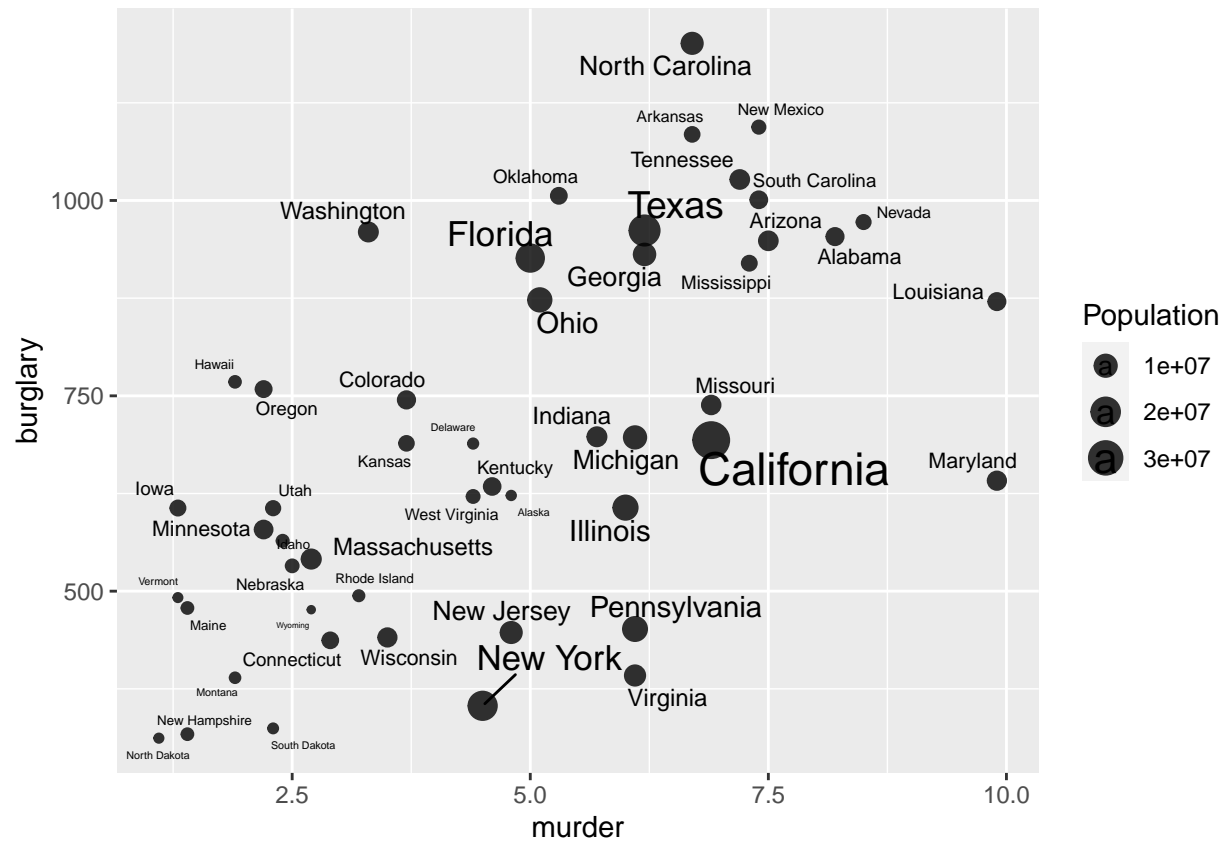
```
    ggtitle('US state wise murder vs burglary') +
    theme_bw()
```
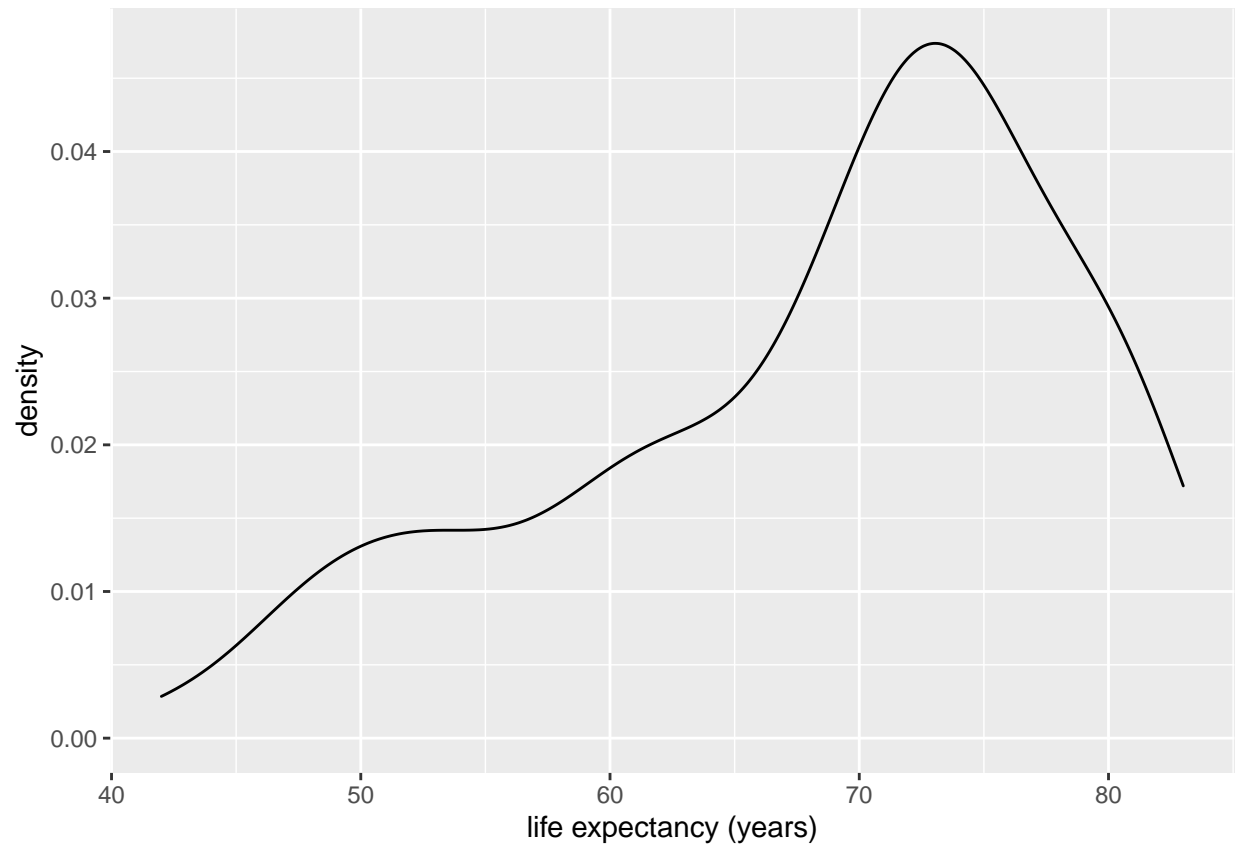
NULL

Density Plot

```
## Density plot using plot() method for life expectancy
plot(density(df2$expectancy, bw = 0.5),
     xlab = "life expectancy (years)")
```

density.default(x = df2$expectancy, bw = 0.5)

life expectancy (years)

```
## Density plot using ggplot for life expectancy
ggplot(df2, aes(x = expectancy)) +
  geom_density() +
  xlab("life expectancy (years)")
```

```
ggtitle("World wide life expectancy density map") +
theme_bw()
```

```
## NULL
```

Week 7&8 - Exercise 4.2

Name: Madhukar Ayachit

Date: 25 Jan 2022

Class: DSC-640

Assignment: scatterplots, bubble charts, and density plots/maps

```
In [1]:   ## import required libraries
          import pandas as pd
          import matplotlib.pyplot as plt
          import matplotlib
          import plotly.graph_objects as go
          import seaborn as sns
          from scipy.stats import stats
          ### pip install plotly
```

```
In [3]:   ##  load dataset from crimerate csv

          df = pd.read_csv('Data/ex4-2/crimerates-by-state-2005.csv')
          # removing "united state"
          df_crime = df[(df.state != "United States") ]
```

```
In [4]:   ## load dataset life expectancy csv

          df_lifeexp = pd.read_csv('Data/ex4-2/life-expectancy.csv')
```

```
In [5]:   df_crime.head()
```

Out[5]:

	state	murder	forcible_rape	robbery	aggravated_assault	burglary	larceny_theft	motor_
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	
5	California	6.9	26.0	176.1	317.3	693.3	1916.5	

```
In [6]:   x=df_crime.state
          y=df_crime.robbery
          plt.figure(figsize=(18, 16))
          plt.bar(x,y)
          plt.xticks(rotation=90)
          plt.show()
```
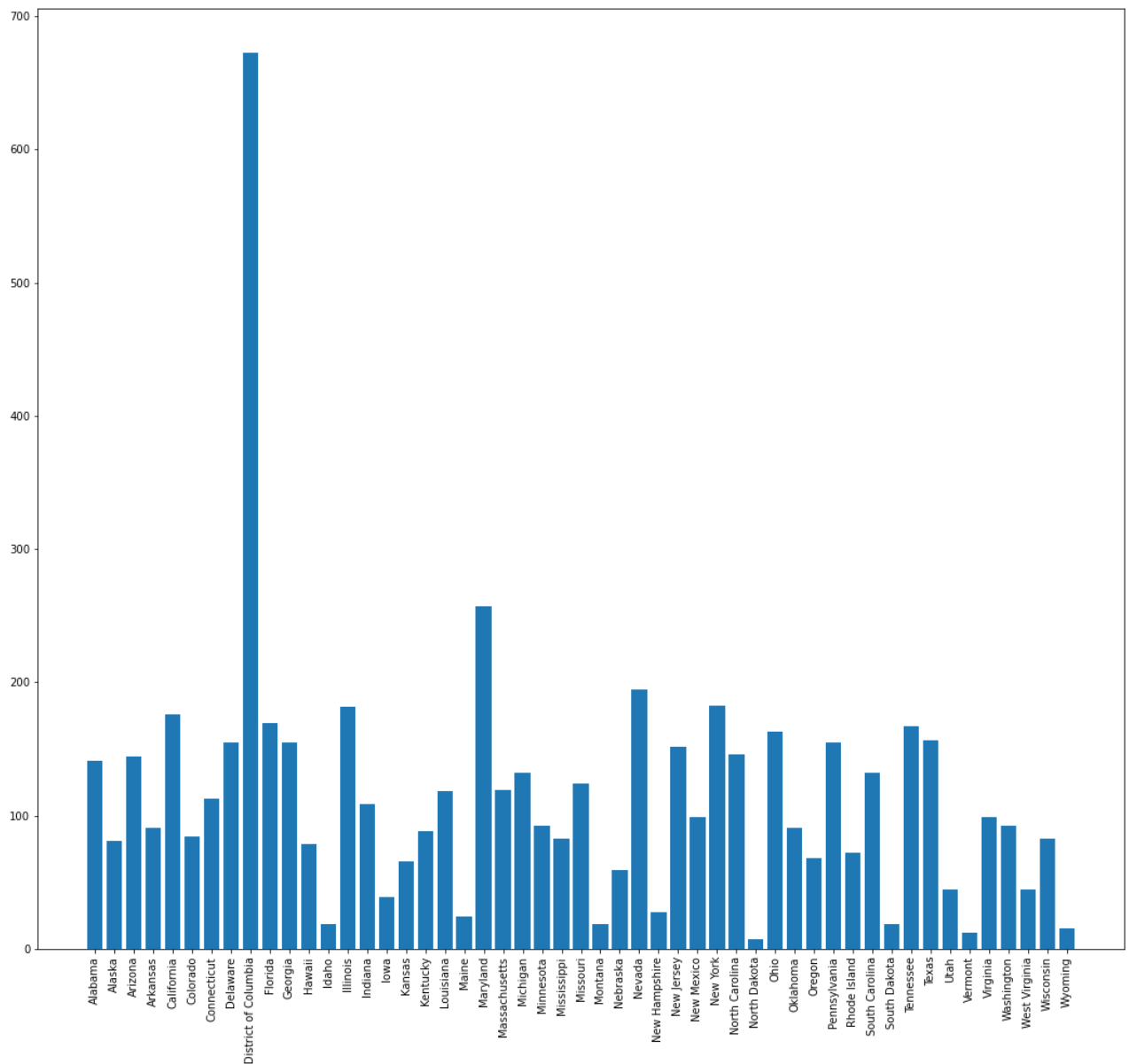
In [7]: `df_lifeexp.head()`

Out[7]:

|	country	year	expectancy
0	Afghanistan	2008	42
1	Albania	2008	73
2	Algeria	2008	71
3	Angola	2008	46
4	Antigua and Barbuda	2008	74

Scatterplot

In [8]:
```
## Scatter Plot

plt.figure(figsize=(18, 16))
plt.scatter(df_crime.murder, df_crime.burglary)
for state, murder, burglary in zip(df_crime.state, df_crime.murder, df_crime.bur
```

```
        plt.text(x = murder - 0.3, y = burglary - 10, s = state)
plt.grid()
plt.title("US States wise murders vs burglary")
plt.xlabel("murder rate")
plt.ylabel("burglary")
plt.show()
```
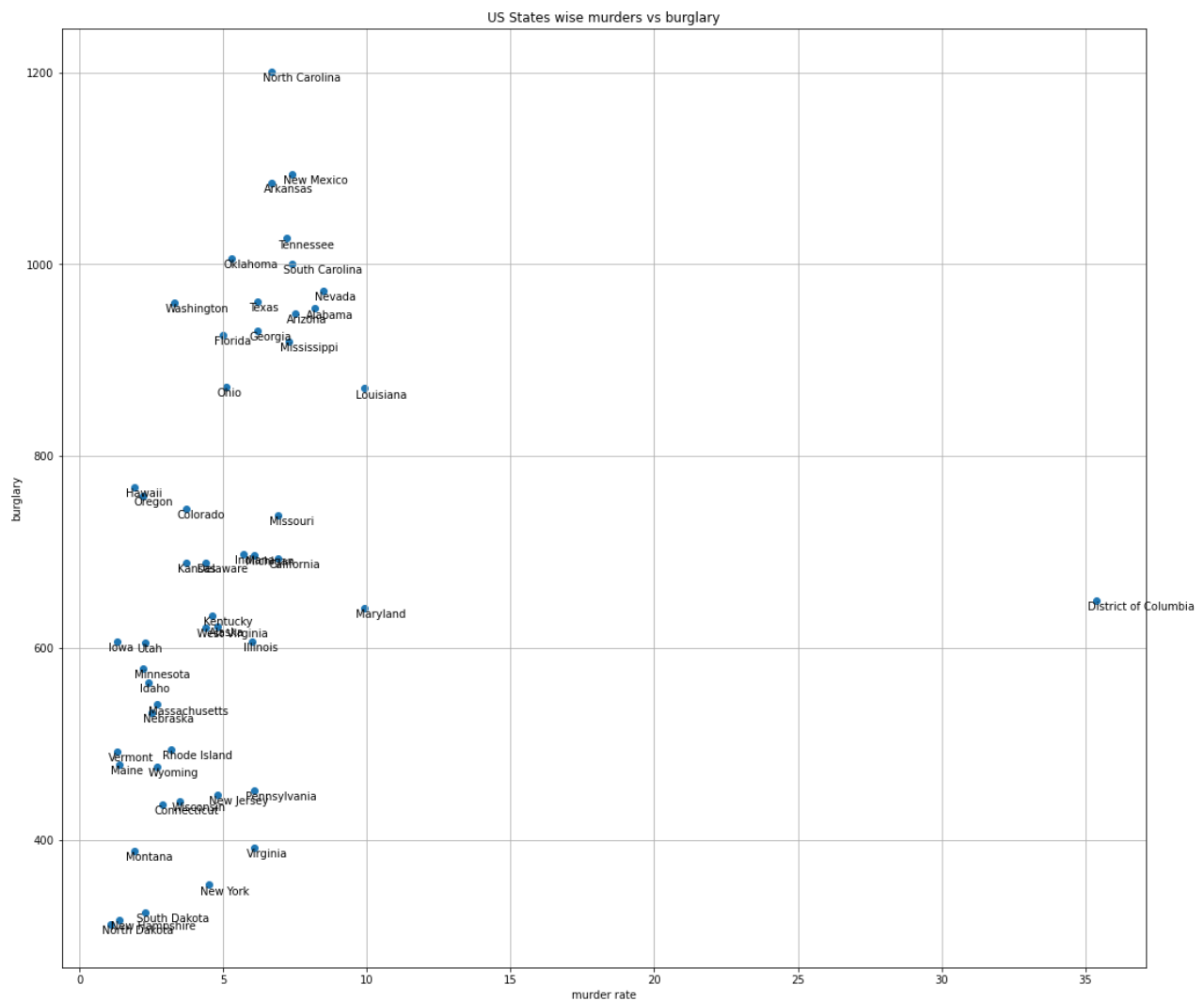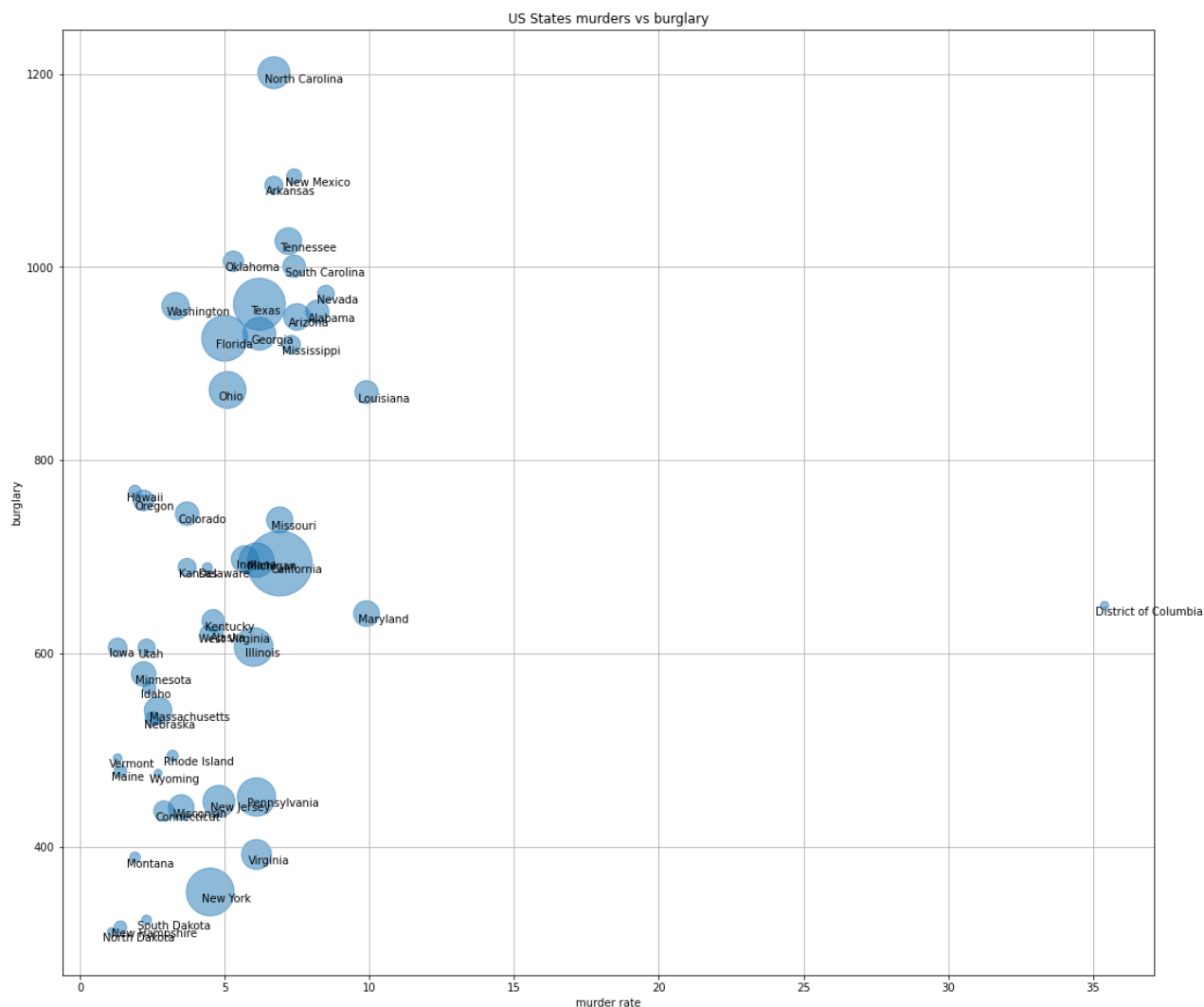

Bubble chart

Matplotlib

```
In [9]:   ## Bubble Plot

          plt.figure(figsize=(18, 16))
          plt.scatter(x=df_crime.murder, y=df_crime.burglary, alpha = 0.5, s = df_crime.po
          for state, murder, burglary in zip(df_crime.state, df_crime.murder, df_crime.bur
              plt.text(x = murder -0.3, y = burglary - 10, s = state)
          plt.grid()
          plt.title("US States murders vs burglary")
          plt.xlabel("murder rate")
          plt.ylabel("burglary")

          plt.show()
```

US States murders vs burglary

Plotly

In [10]:
```python
## Bubble charts using plotly with Hover Text
## https://plotly.com/python/bubble-charts/#scaling-the-size-of-bubble-charts


fig = go.Figure(data=[go.Scatter(
    x = df_crime.murder, y = df_crime.burglary,
    text = df_crime.state,
    mode = 'markers',
    marker = dict(size=df_crime.population/250000)
)],
    layout=go.Layout(height=800, width=1000)
        )


fig.show()
```
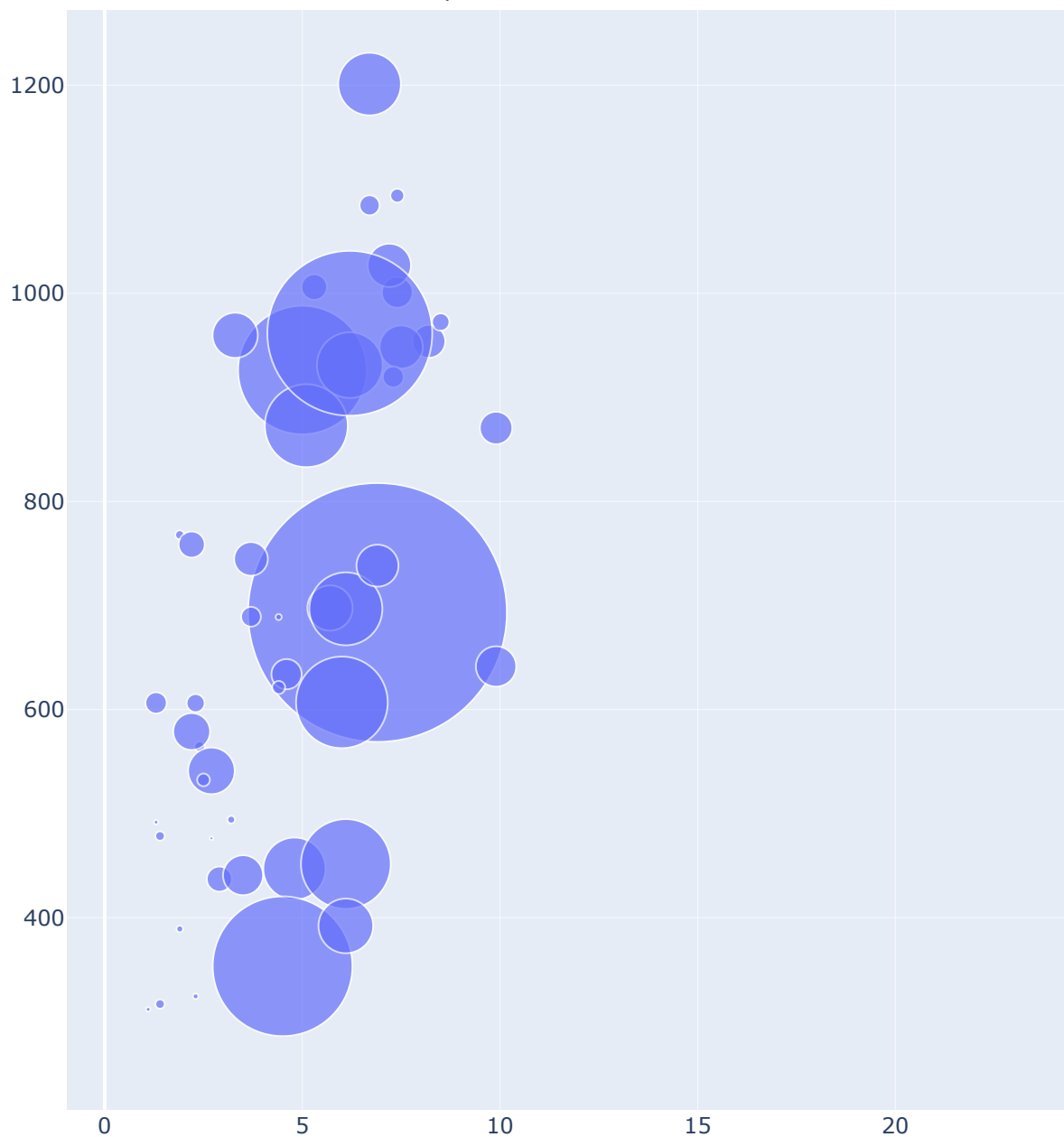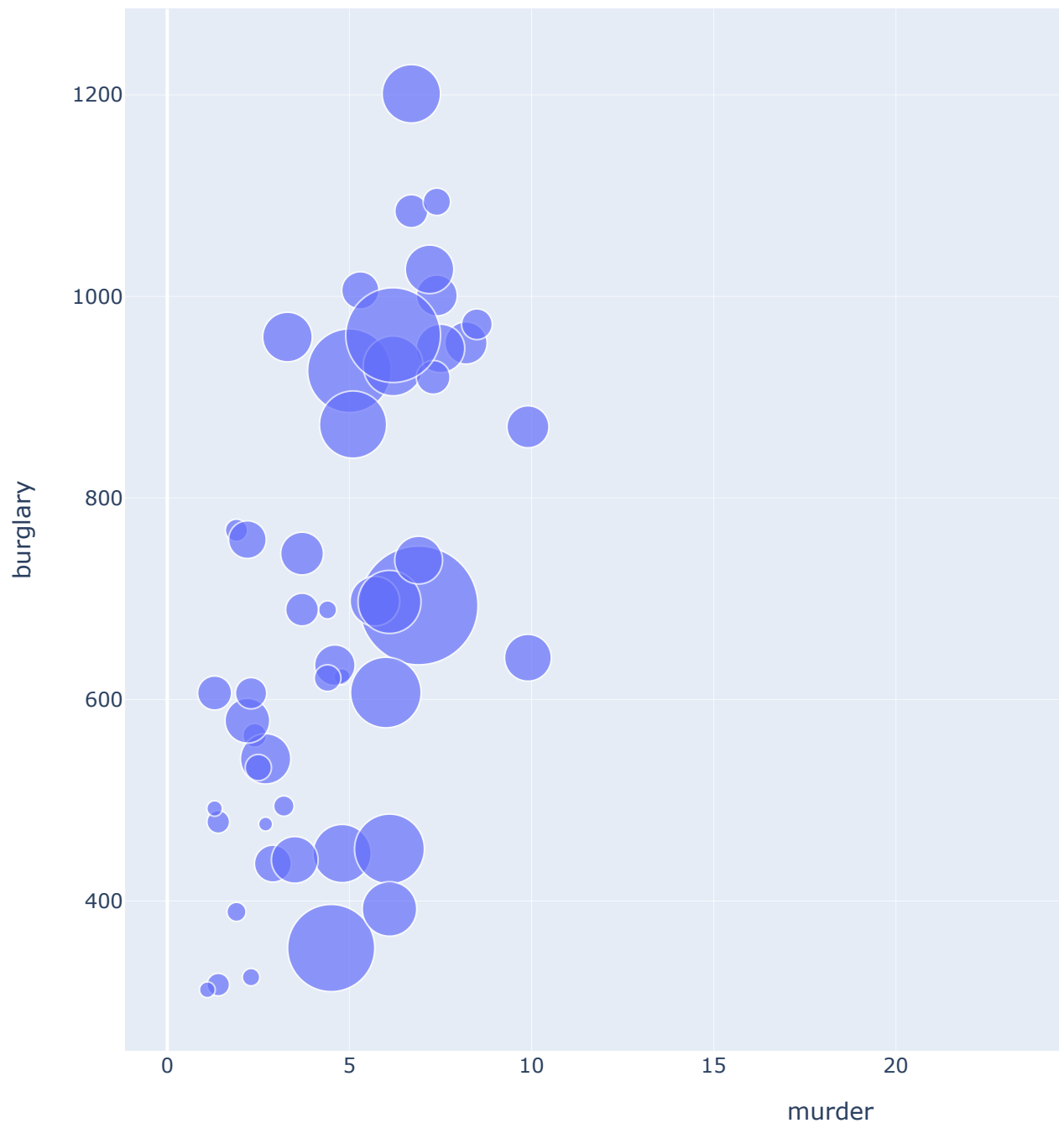
```
In [11]:   import plotly.express as px

           fig = px.scatter(df_crime, x = "murder", y = "burglary", size = "population",
                            hover_name="state", log_x=False, size_max = 50, title = "US sta

           fig.update_layout(
               autosize=False,
               width=1000,
               height=800)

           fig.show()
```

US statewise murder vs. burglary
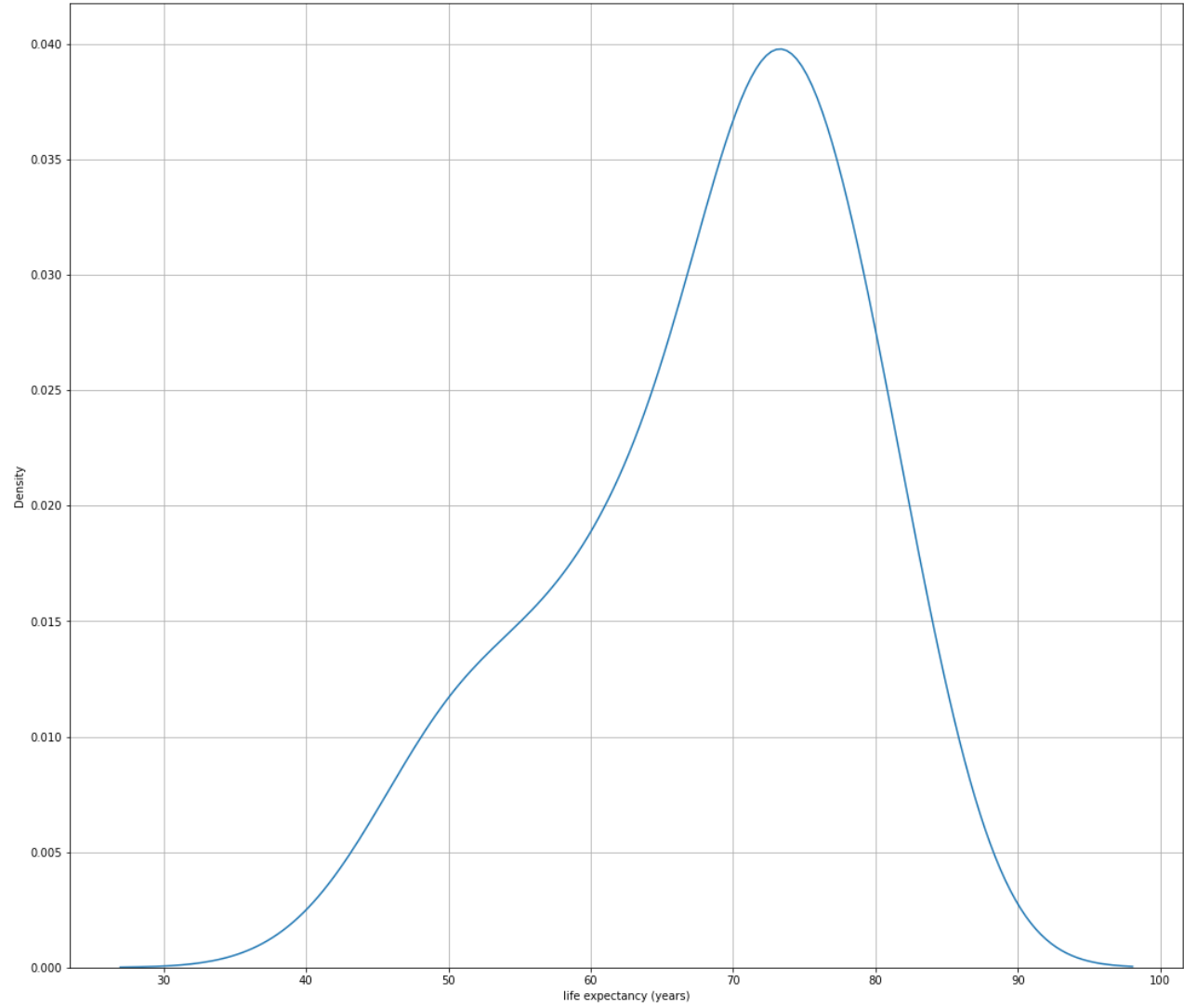
Density chart

```
In [12]:   ## plot kde density chart


           plt.figure(figsize=(18, 16))

           sns.kdeplot(df_lifeexp.expectancy, bw_method=0.5)
           plt.grid()
           plt.xlabel("life expectancy (years)")
           plt.show()
```
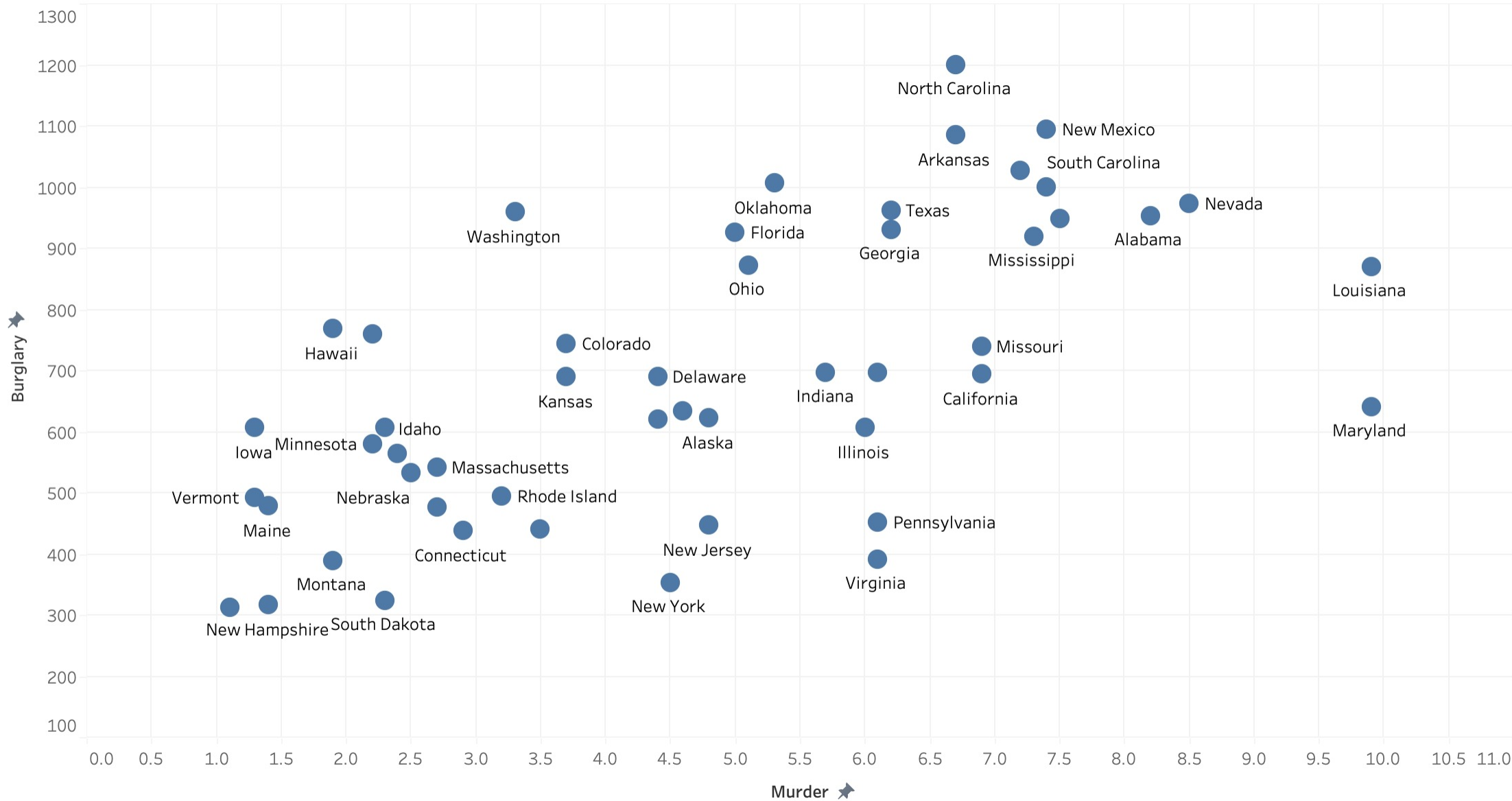
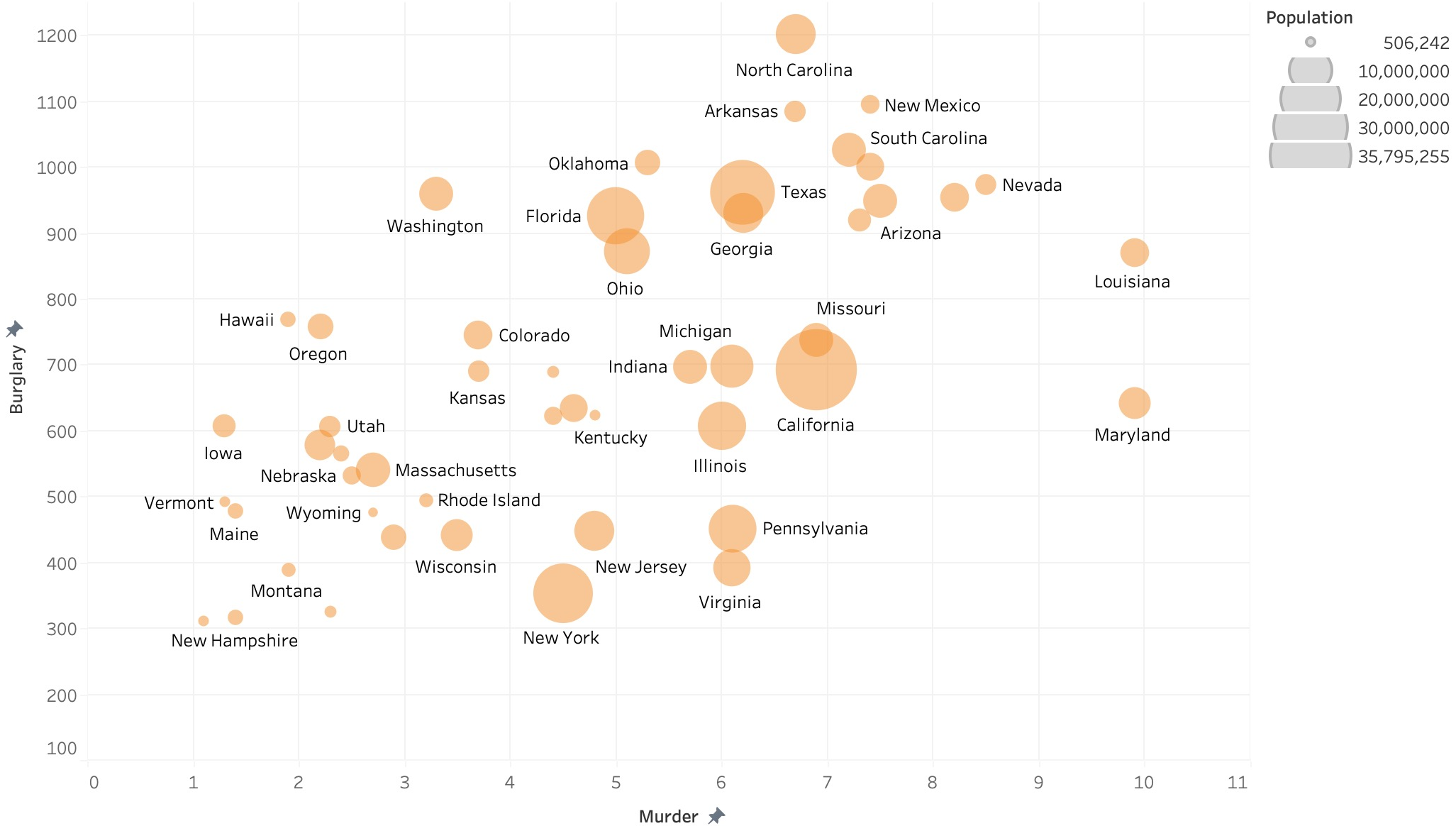In [ ]:

Madhukar_ayachit_Week_07_08

File created on: 2/3/22 4:00:16 PM EST

US States - Murder vs. Burglary

Murder vs. Burglary. The marks are labeled by State. The view is filtered on State, which excludes District of Columbia and United States.

US States - Murder vs. Burglary

Sum of Murder vs. sum of Burglary. Size shows sum of Population. The marks are labeled by State. The view is filtered on State, which excludes District of Columbia and United States.

Life Expectancy country

Expectancy
- 42
- 50
- 60
- 70
- 83

© 2022 Mapbox © OpenStreetMap

Map based on Longitude (generated) and Latitude (generated). Size shows sum of Expectancy. Details are shown for Country.

Motor Vehicle Theft

Motor Vehicle Theft

- 102
- 500
- 1,000
- 1,402
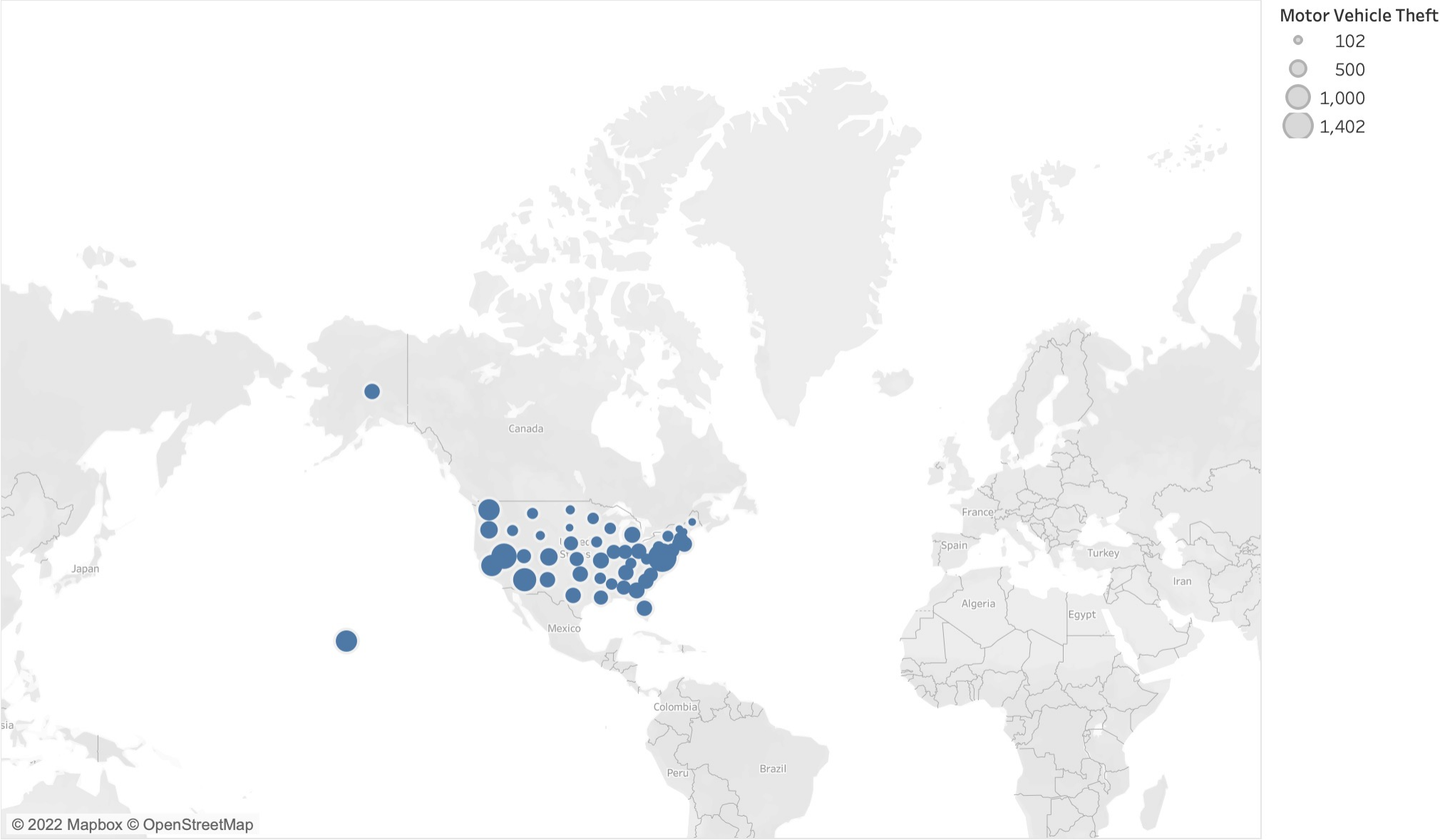
Map based on Longitude (generated) and Latitude (generated). Size shows sum of Motor Vehicle Theft. Details are shown for State.

© 2022 Mapbox © OpenStreetMap