

Assignment 4.2 Charts

DSC640

Taniya Adhikari

Tableau – Scatter Plot

Tableau - Scatter Plot: Countries Birth Rates vs. Life Expectancy For Year 2008

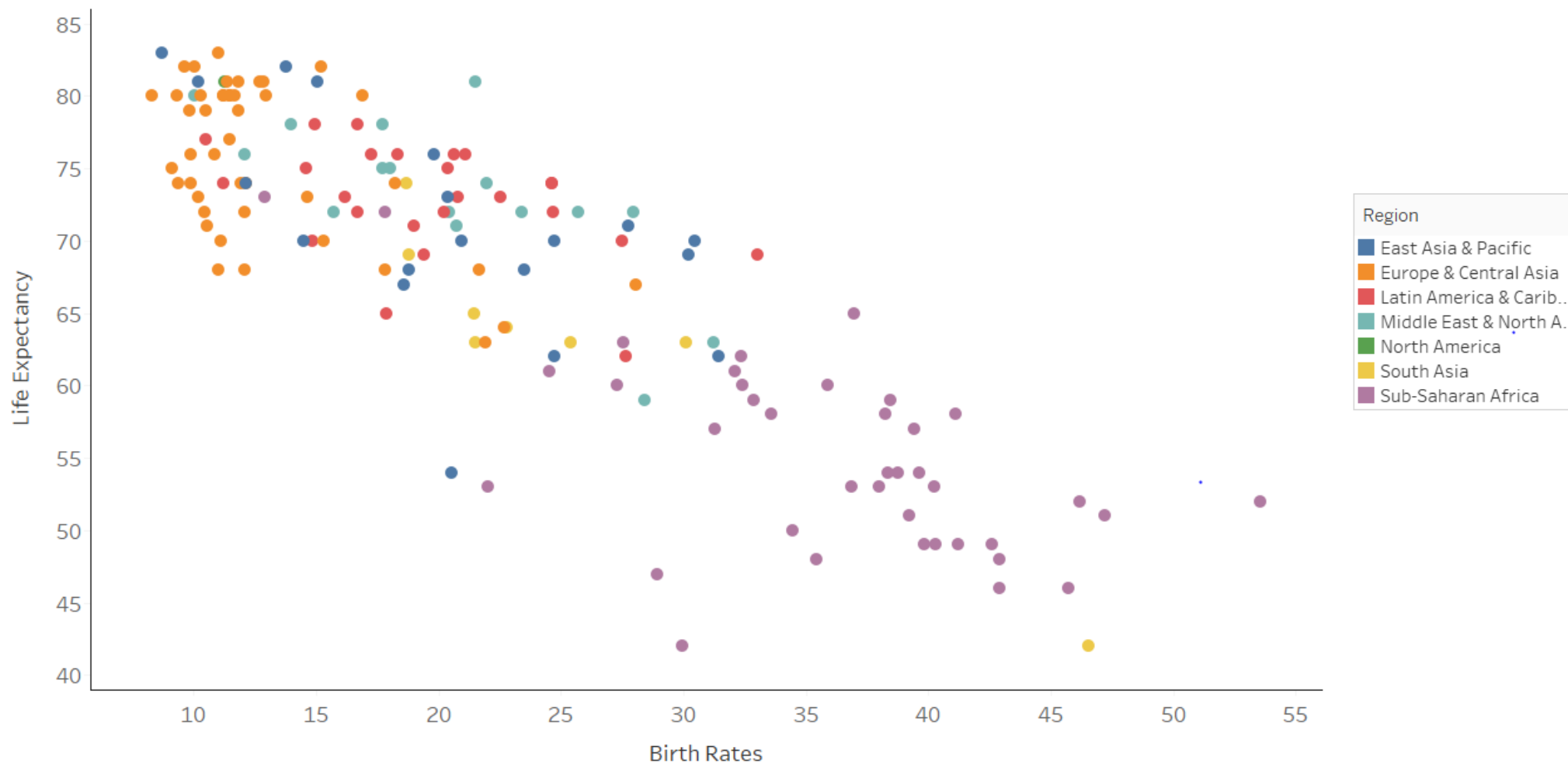


Tableau – Bubble Chart

Tableau - Bubble Chart: Countries Birth Rates vs. Life Expectancy For Year 2008

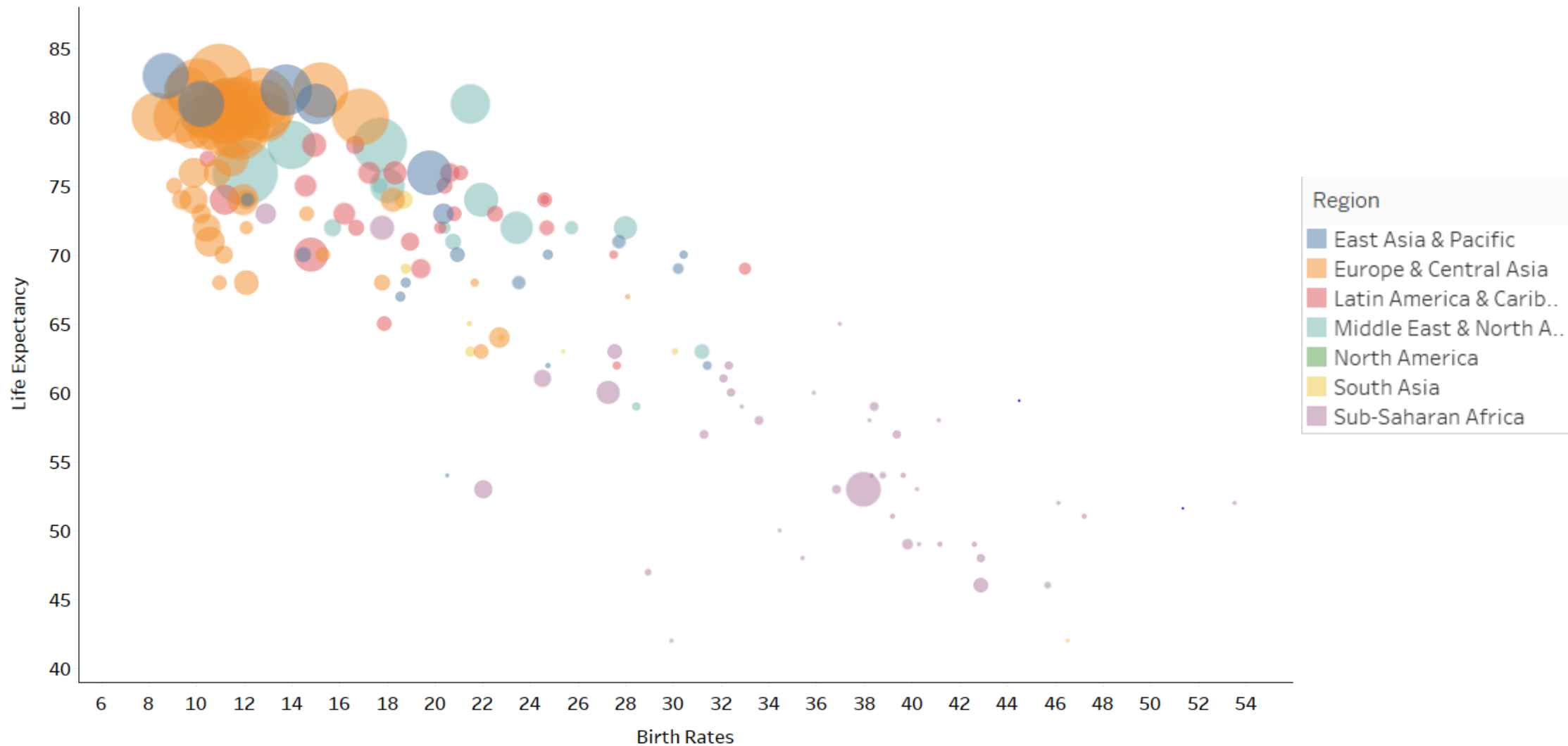
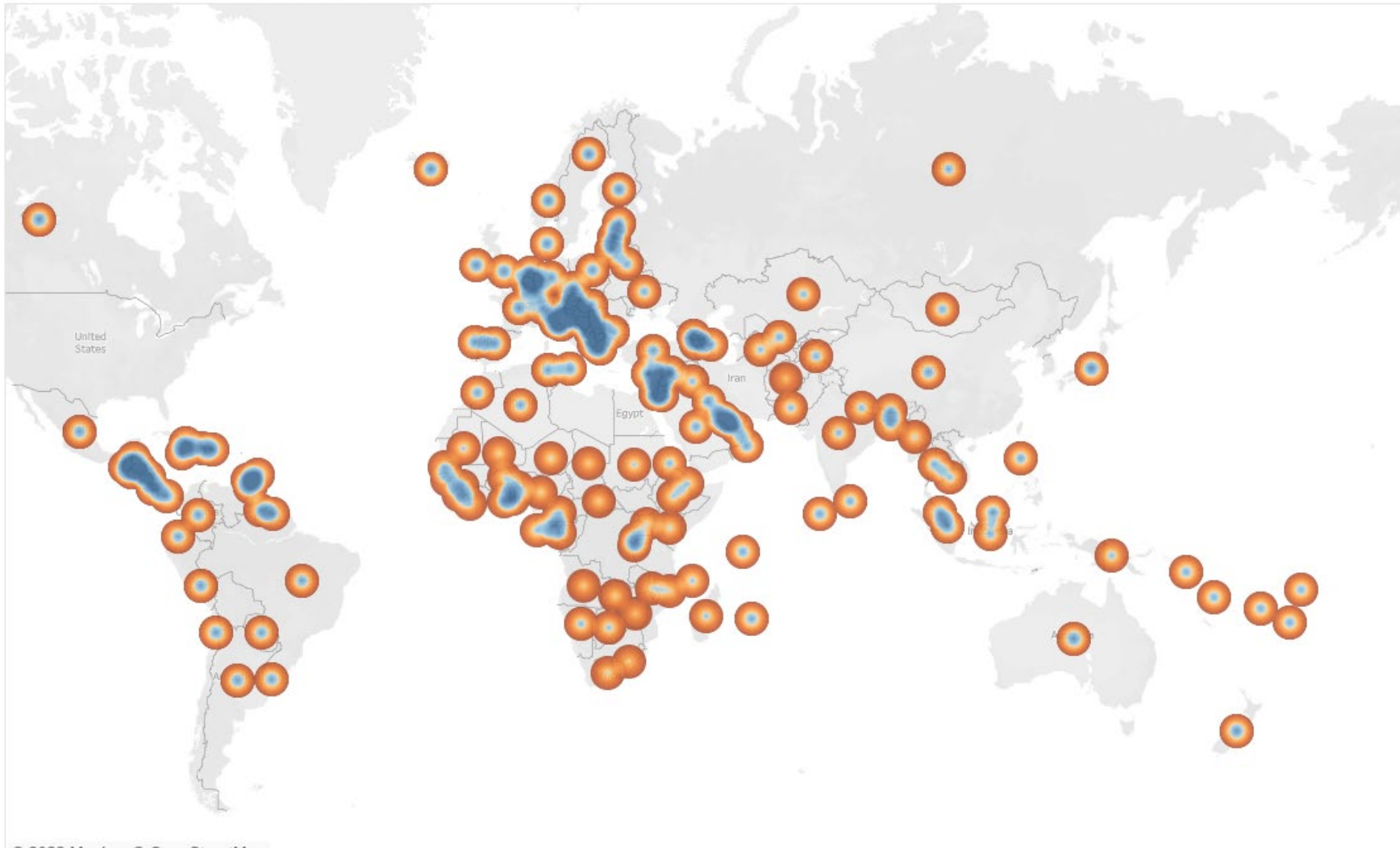


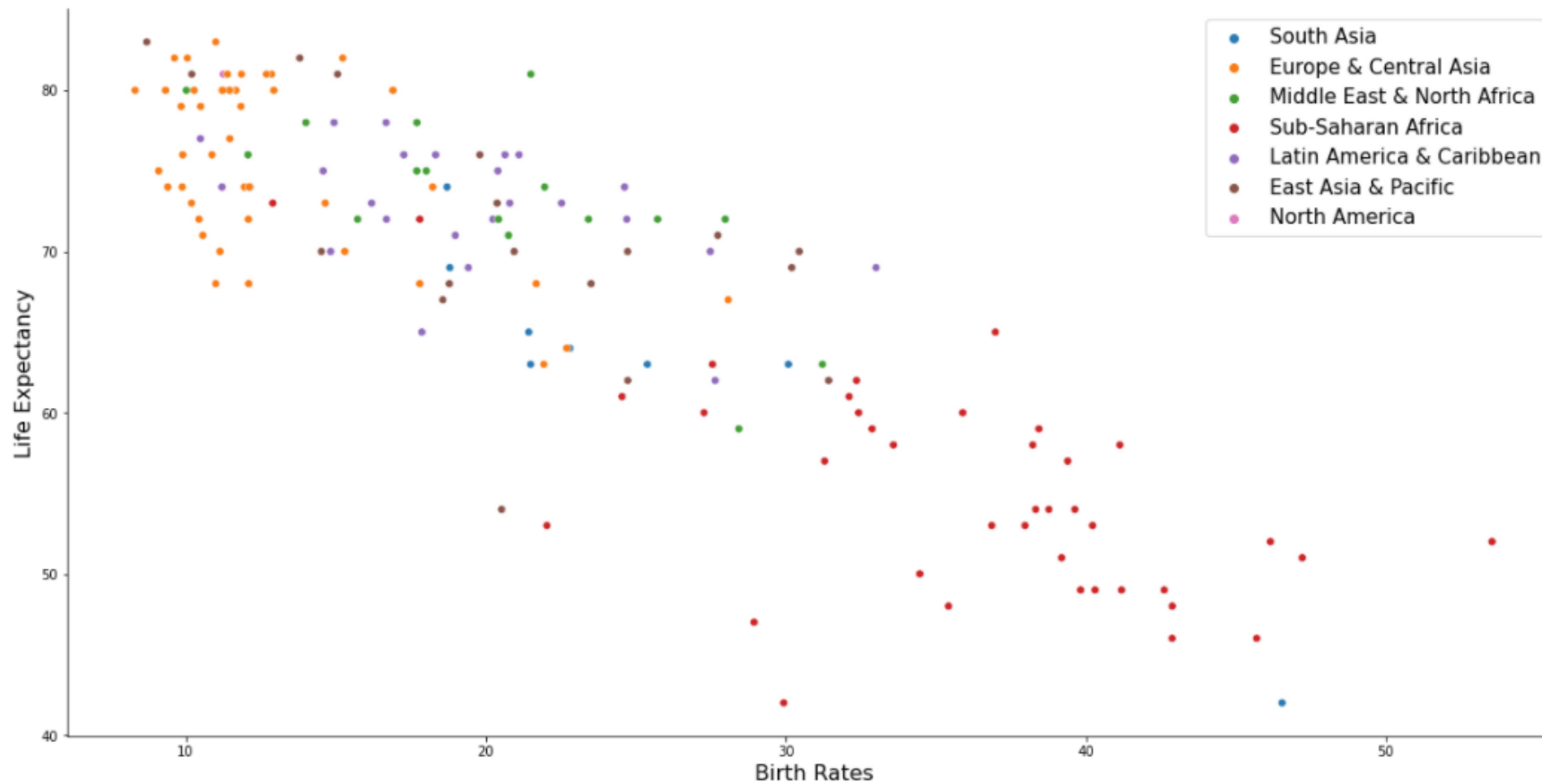
Tableau – Density Map

Tableau - Density Map: Life Expectancy For Year 2008



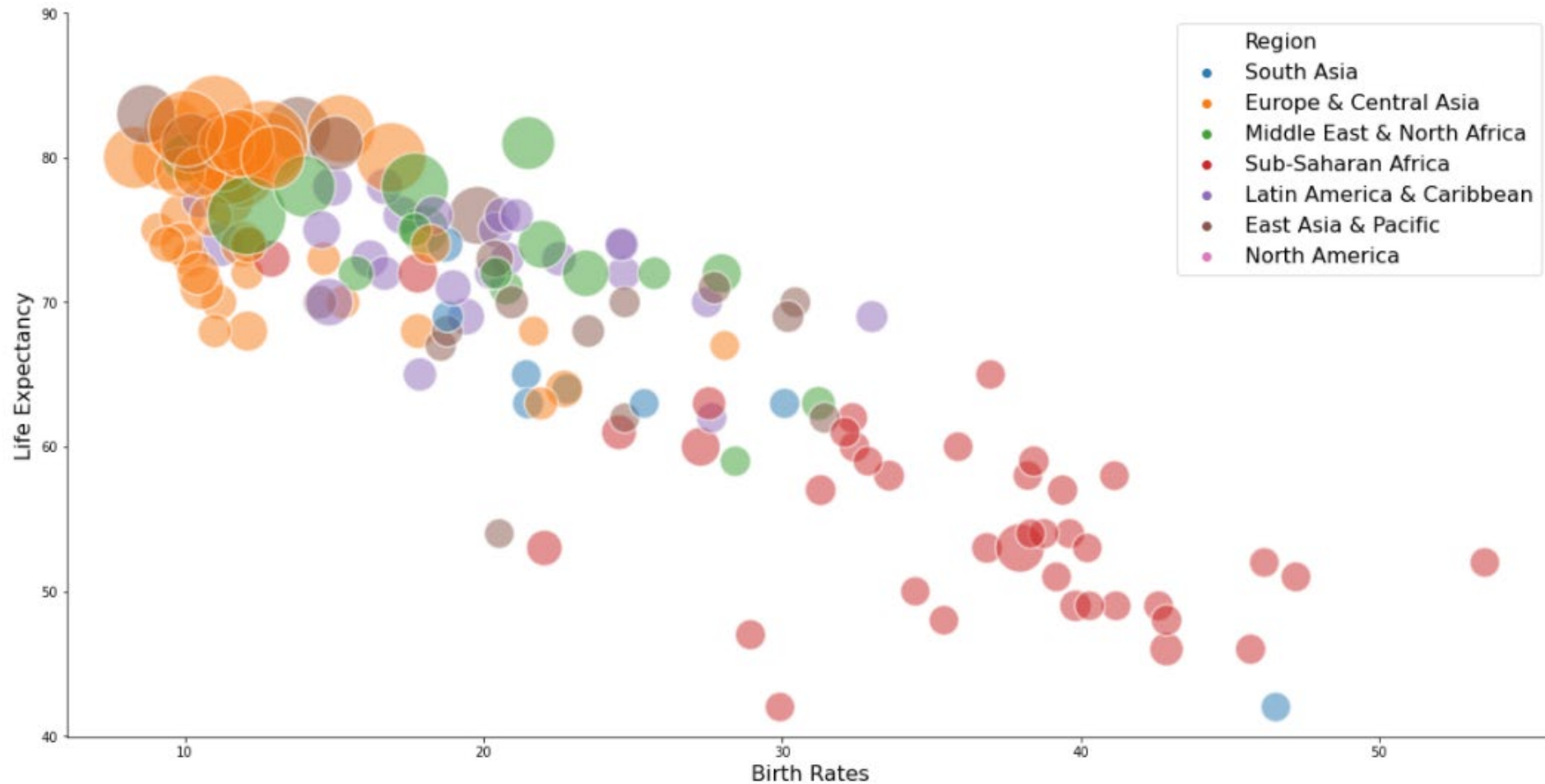
Python – Scatter Plot

Python - Scatter Plot: Countries Birth Rates vs. Life Expectancy For Year 2008



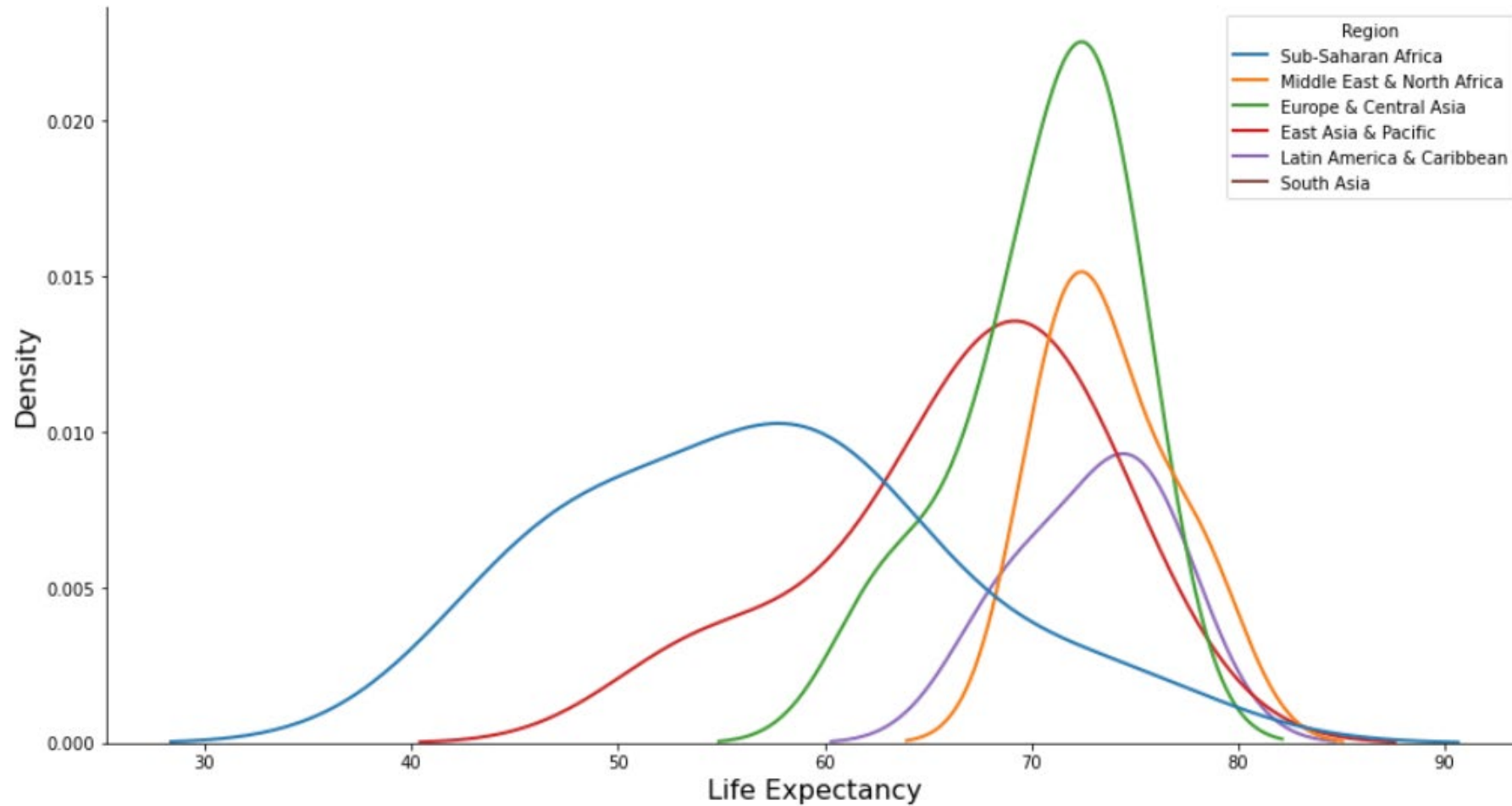
Python – Bubble Chart

Python - Bubble Chart: Countries Birth Rates vs. Life Expectancy For Year 2008

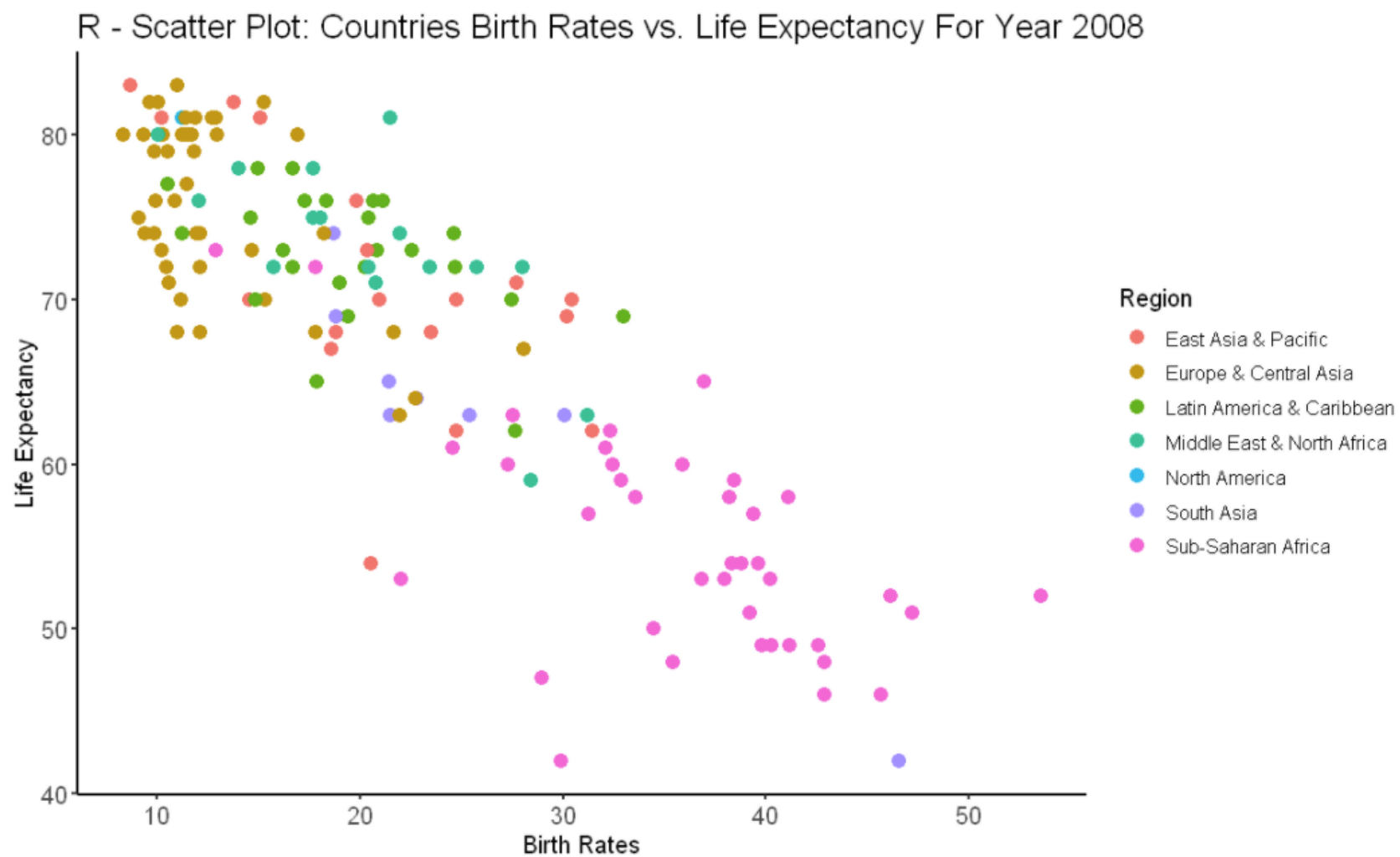


Python – Density Plot

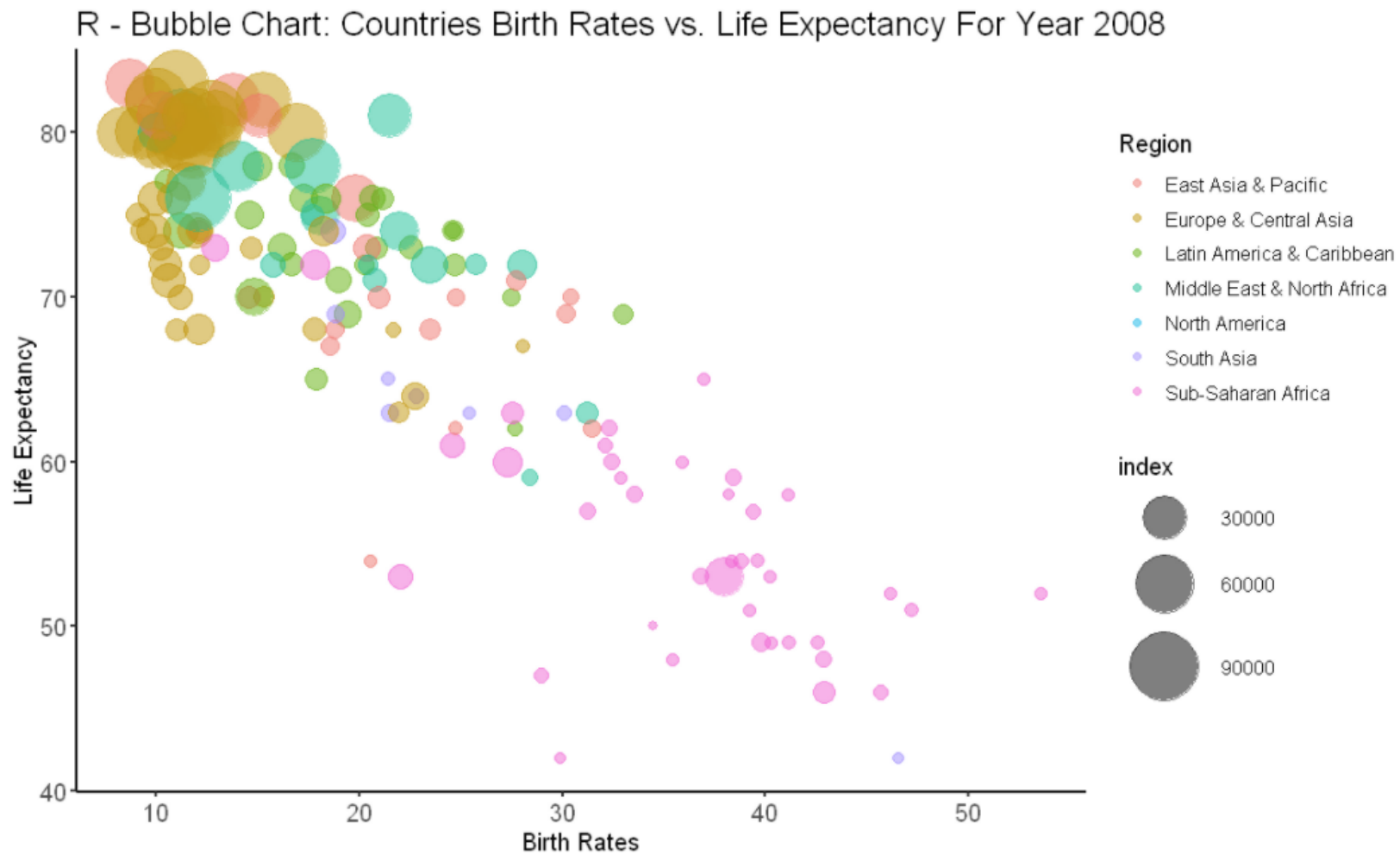
Python - Density Plot: Life Expectancy For Year 2008



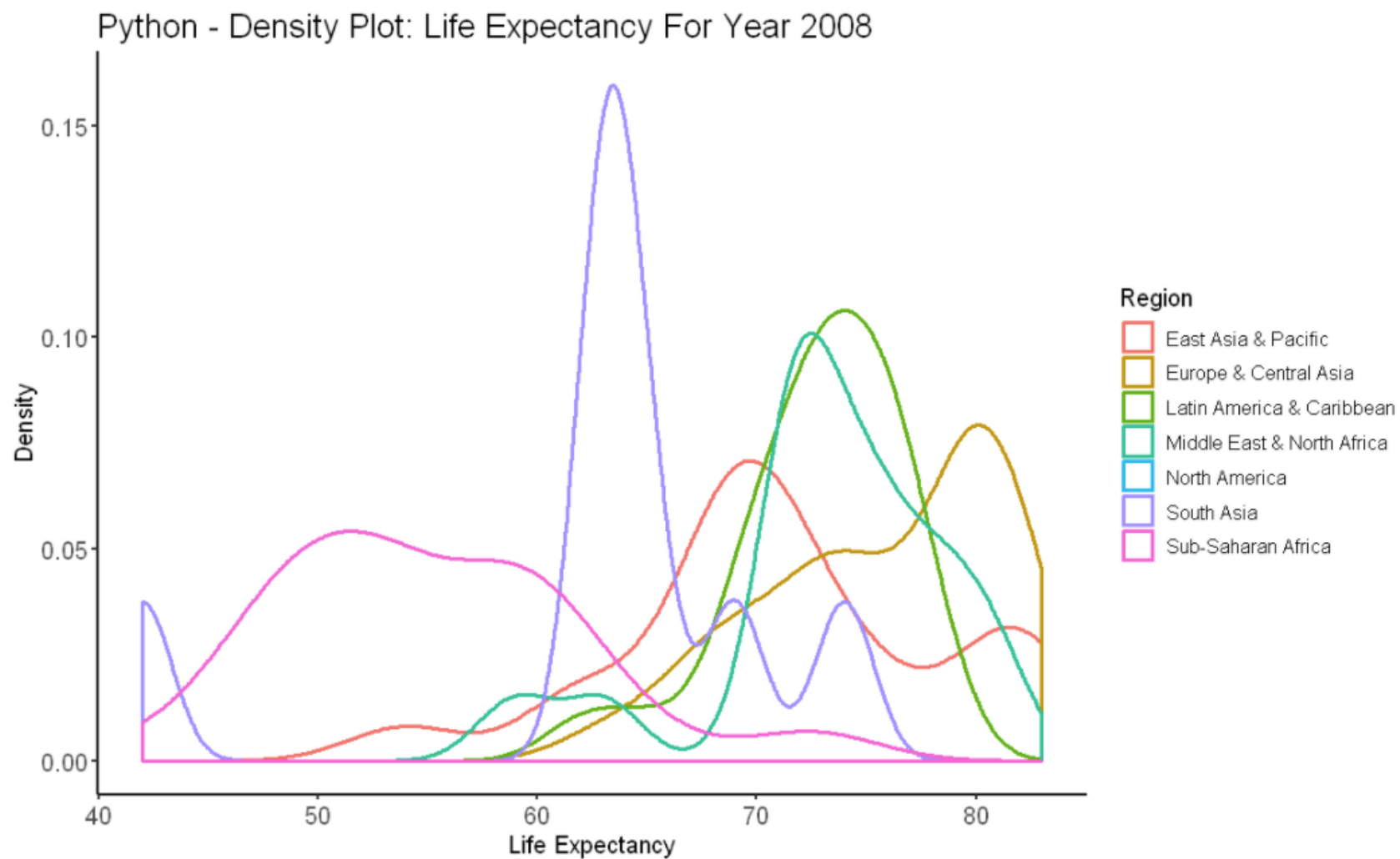
R- Scatter Plot



R – Bubble Chart



R – Density Plot



Supplemental Files

- Tableau Screenshot
- Python Code
- R Code



Format Font

A [font settings icons] Fields

Sheet Rows Columns

Default

Worksheet: [dropdown]

Pane: [dropdown]

Header: [dropdown]

Tooltip: [dropdown]

Title: [dropdown]

Total

Pane: [dropdown]

Header: [dropdown]

Grand Total

Pane: [dropdown]

Header: [dropdown]

Clear

Data Source

ScatterPlot

BubbleChart

DensityPlot

Dashboard 1

Tableau - Density Map: Life Expectancy For Year 2008

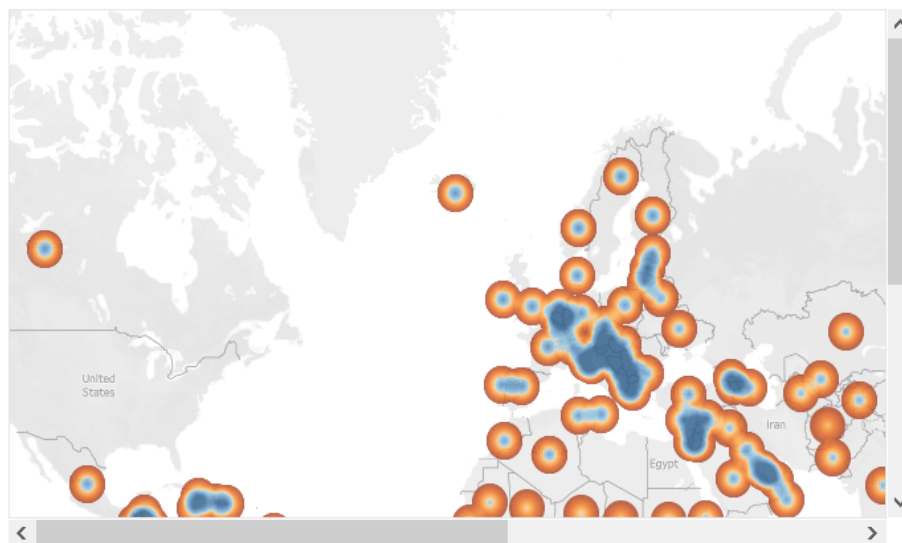


Tableau - Bubble Chart: Countries Birth Rates vs. Life Expectancy For Year 2008

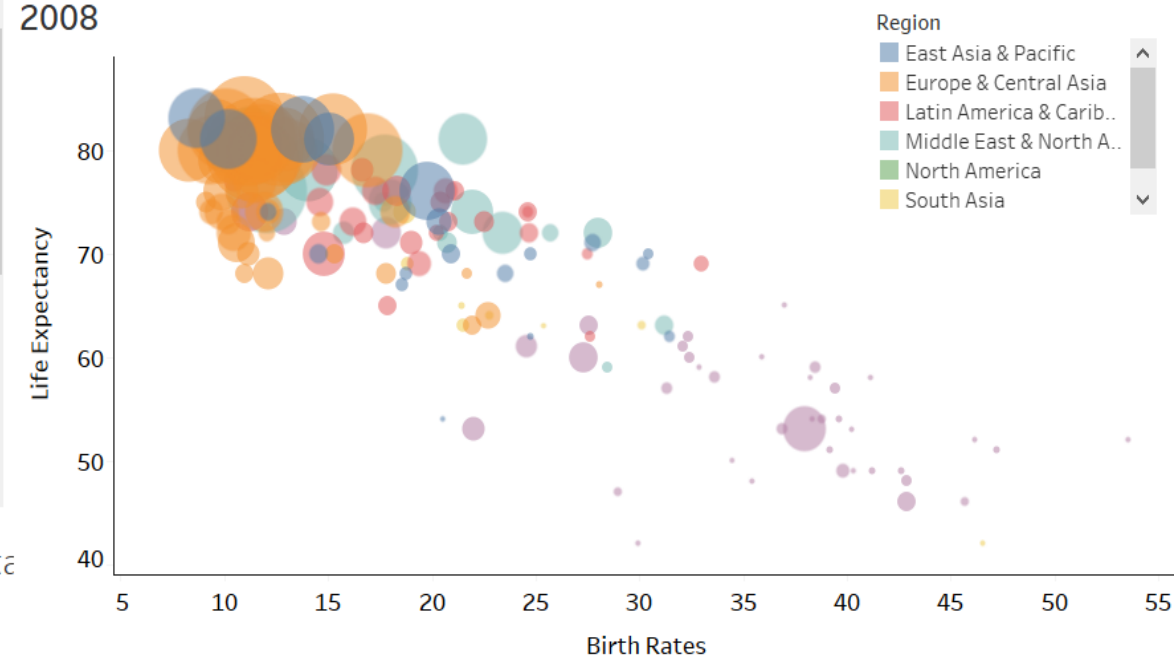
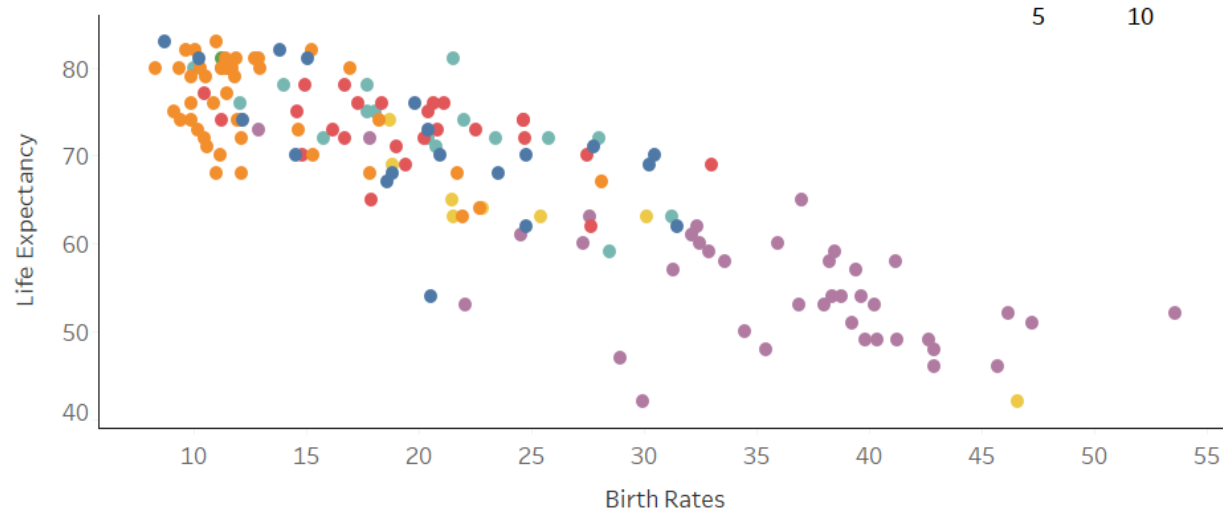


Tableau - Scatter Plot: Countries Birth Rates vs. Life Expecta



Python Script

Assignment 4.2: Scatterplots, Bubble Charts, DensityMaps/Plots

DSC640

Taniya Adhikari

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
from numerize import numerize
import matplotlib.ticker as ticker
from matplotlib.ticker import FuncFormatter
import plotly.express as px
```

```
In [2]: birthrates = pd.read_csv('birth-rate.csv')
birthrates.head(3)
```

```
Out[2]:
```

	Country	1960	1961	1962	1963	1964	1965	1966	1967	1968	...	1999	2000
0	Aruba	36.400	35.179	33.863	32.459	30.994	29.513	28.069	26.721	25.518	...	15.024	14.52
1	Afghanistan	52.201	52.206	52.208	52.204	52.192	52.168	52.130	52.076	52.006	...	51.229	50.90
2	Angola	54.432	54.394	54.317	54.199	54.040	53.836	53.585	53.296	52.984	...	48.662	48.35

3 rows × 50 columns

```
In [3]: gdp_per_capita = pd.read_csv('GDP.csv')
gdp_per_capita.head(3)
```

```
Out[3]:
```

	Country Name	Country Code	1960	1961	1962	1963	1964	1965
0	Aruba	ABW	NaN	NaN	NaN	NaN	NaN	NaN
1	Africa Eastern and Southern	AFE	147.612227	147.014904	156.189192	182.243917	162.347592	180.214908
2	Afghanistan	AFG	59.773234	59.860900	58.458009	78.706429	82.095307	101.108325

3 rows × 63 columns

```
In [4]: metadata= pd.read_csv('metadata_country.csv')
```

```
metadata = metadata[['Country Code', 'Region']]
metadata.head(3)
```

Out[4]:

	Country Code	Region
0	ABW	Latin America & Caribbean
1	AFE	NaN
2	AFG	South Asia

In [5]:

```
life_expectancy = pd.read_csv('life-expectancy.csv')
life_expectancy.head(3)
```

Out[5]:

	country	year	expectancy
0	Afghanistan	2008	42
1	Albania	2008	73
2	Algeria	2008	71

In [6]:

```
life_expectancy.shape
```

Out[6]: (187, 3)

In [7]:

```
life_expectancy['year'].describe()
```

Out[7]:

```
count      187.0
mean       2008.0
std         0.0
min        2008.0
25%        2008.0
50%        2008.0
75%        2008.0
max        2008.0
Name: year, dtype: float64
```

In [8]:

```
country = life_expectancy['country'].to_list()
```

In [9]:

```
bd = birthrates[birthrates['Country'].isin(country)]
gdp = gdp_per_capita[gdp_per_capita['Country Name'].isin(country)]
```

In [10]:

```
gdp = gdp[['Country Name', 'Country Code', '2008']]
gdp.shape
```

Out[10]: (161, 3)

In [11]:

```
gdp = pd.merge(gdp, metadata, how="inner", on=["Country Code"])
print(gdp.shape)
gdp.head(4)
```

(161, 4)

Out[11]:

	Country Name	Country Code	2008	Region
0	Afghanistan	AFG	364.663542	South Asia
1	Angola	AGO	4080.941034	Sub-Saharan Africa
2	Albania	ALB	4370.539925	Europe & Central Asia
3	United Arab Emirates	ARE	44498.940510	Middle East & North Africa

In [12]:

```
bd = bd[['Country', '2008']]
bd.shape
```

Out[12]: (161, 2)

In [13]:

```
bd.rename(columns={'2008': 'birthrates', 'Country': 'country'}, inplace=True)
gdp.rename(columns={'2008': 'gdp', 'Country Name': 'country'}, inplace=True)
```

In [14]:

```
demographics = pd.merge(life_expectancy, bd, how="inner", on=["country"])
demographics.shape
```

Out[14]: (161, 4)

In [15]:

```
demographics = pd.merge(demographics, gdp, how="inner", on=["country"])
```

In [16]:

```
demographics[demographics['birthrates'].isnull()]
```

Out[16]:

	country	year	expectancy	birthrates	Country Code	gdp	Region
4	Antigua and Barbuda	2008	74	NaN	ATG	16044.105800	Latin America & Caribbean
41	Dominica	2008	74	NaN	DMA	6468.944767	Latin America & Caribbean
77	Kiribati	2008	67	NaN	KIR	1428.134977	East Asia & Pacific
91	Marshall Islands	2008	59	NaN	MHL	2704.463554	East Asia & Pacific
95	Monaco	2008	82	NaN	MCO	185785.725900	Europe & Central Asia
111	Palau	2008	72	NaN	PLW	10598.882830	East Asia & Pacific

In [17]:

```
demographics[demographics['gdp'].isnull()]
```

Out[17]:

	country	year	expectancy	birthrates	Country Code	gdp	Region
133	Somalia	2008	48	44.105	SOM	NaN	Sub-Saharan Africa

```
In [18]: demographics.dropna(inplace=True)
demographics.shape
```

```
Out[18]: (152, 7)
```

Python Scatter Plot

```
In [73]: plt.rcParams['figure.figsize'] = [20,10]
fig, ax = plt.subplots()

sns.scatterplot(x="birthrates", y="expectancy", hue="Region", data=demographics)

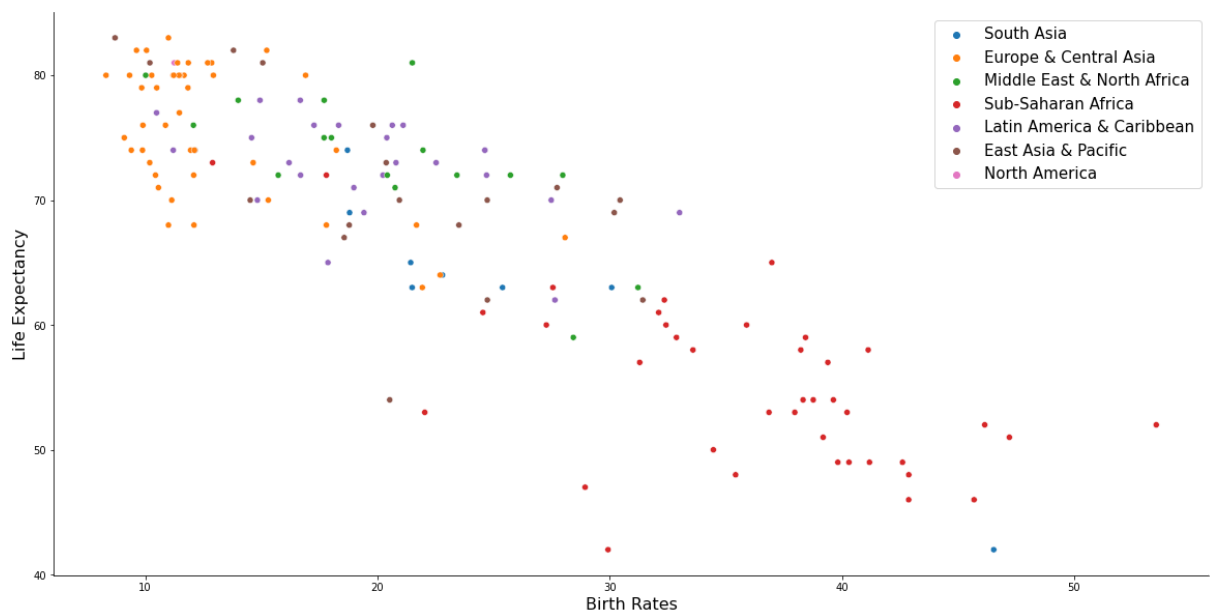
plt.suptitle("Python - Scatter Plot: Countries Birth Rates vs. Life Expectancy For Ye
             size=20, x=0.08, y=.95, horizontalalignment='left', verticalalignment='to

plt.ylabel('Life Expectancy', size=16)
plt.xlabel('Birth Rates', size=16)
plt.legend(fontsize=15)

right_side = ax.spines["right"]
right_side.set_visible(False)
top = ax.spines["top"]
top.set_visible(False)

plt.show()
```

Python - Scatter Plot: Countries Birth Rates vs. Life Expectancy For Year 2008



Python - Bubble Chart

```
In [61]: plt.rcParams['figure.figsize'] = [20,10]
fig, ax = plt.subplots()

g = sns.scatterplot(x="birthrates", y="expectancy",
                    size="gdp", sizes=(500,5000),
```



```

alpha=0.5, hue="Region", data=demographics)

h,l = g.get_legend_handles_labels()
plt.legend(h[0:8],l[0:8], fontsize=16)

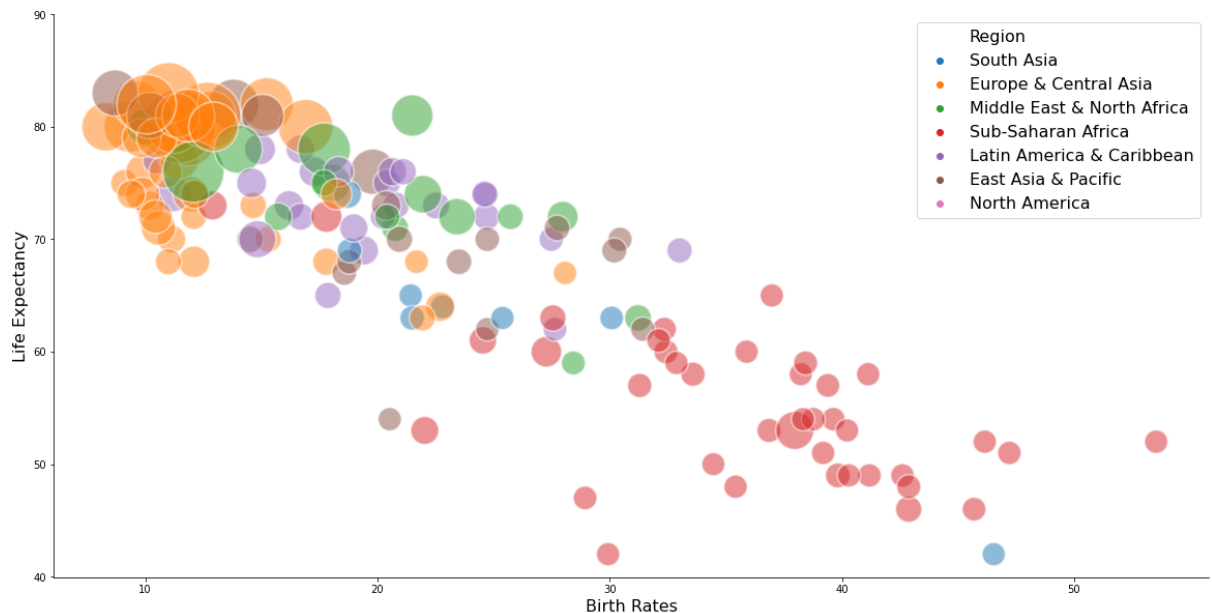
plt.suptitle("Python - Bubble Chart: Countries Birth Rates vs. Life Expectancy For Year 2008",
             size=20, x=0.08, y=.95, horizontalalignment='left', verticalalignment='top')

plt.ylim(top=90)
plt.ylabel('Life Expectancy', size=16)
plt.xlabel('Birth Rates', size=16)

sns.despine()
plt.show(g)

```

Python - Bubble Chart: Countries Birth Rates vs. Life Expectancy For Year 2008



Python - Density Plot

```

In [68]: plt.rcParams['figure.figsize'] = [15,8]
sns.kdeplot(x = sample['expectancy'],
            hue = sample['Region'],
            linewidth = 2)
plt.suptitle("Python - Density Plot: Life Expectancy For Year 2008",
             size=20, x=0.08, y=.95, horizontalalignment='left', verticalalignment='top')

plt.xlabel('Life Expectancy', size=16)
plt.ylabel("Density", size=16)

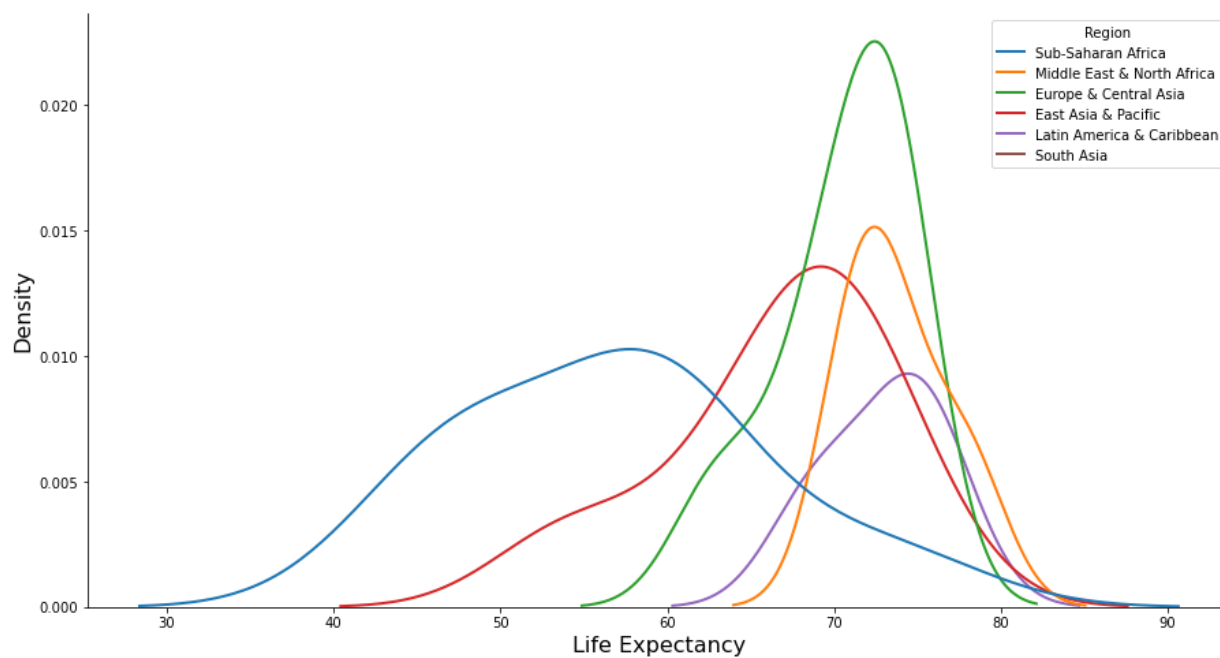
sns.despine()

plt.show()

```

C:\Users\bibek\anaconda3\envs\my_env\lib\site-packages\seaborn\distributions.py:316:
 UserWarning: Dataset has 0 variance; skipping density estimate. Pass `warn_singular=False` to disable this warning.
 warnings.warn(msg, UserWarning)

Python - Density Plot: Life Expectancy For Year 2008



```
In [69]: demographics.to_csv('sample.csv')
```

```
In [ ]:
```

R Script

Assignment 4.2: Scatterplots, Bubble Charts, DensityMaps/Plots

DSC640

Taniya Adhikari

```
In [15]: 1 library(ggplot2)
          2 library(readxl)
          3 library(scales)
          4 library(plyr)
          5 library(dplyr)
          6 library(ggrepel)
          7 library(RColorBrewer)
```

```
In [3]: 1 demographics <- read.csv("sample.csv")
          2 head(demographics)
```

X	country	year	expectancy	birthrates	Country.Code	gdp	Region
0	Afghanistan	2008	42	46.538	AFG	364.6635	South Asia
1	Albania	2008	73	14.649	ALB	4370.5399	Europe & Central Asia
2	Algeria	2008	71	20.759	DZA	4923.6316	Middle East & North Africa
3	Angola	2008	46	42.875	AGO	4080.9410	Sub-Saharan Africa
5	Argentina	2008	76	17.269	ARG	9020.8733	Latin America & Caribbean
6	Armenia	2008	70	15.299	ARM	4010.8614	Europe & Central Asia

In [4]: 1 str(demographics)

```
'data.frame':  152 obs. of  8 variables:
 $ X          : int  0 1 2 3 5 6 7 8 9 10 ...
 $ country    : Factor w/ 152 levels "Afghanistan",...: 1 2 3 4 5 6 7 8 9 10 ...
 $ year       : int  2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 ...
 $ expectancy : int  42 73 71 46 76 70 82 80 68 75 ...
 $ birthrates : num  46.5 14.6 20.8 42.9 17.3 ...
 $ Country.Code: Factor w/ 152 levels "AFG","AGO","ALB",...: 1 3 41 2 5 6 7 8 9 16 ...
 $ gdp        : num  365 4371 4924 4081 9021 ...
 $ Region     : Factor w/ 7 levels "East Asia & Pacific",...: 6 2 4 7 3 2 1 2 2 4 ...
```

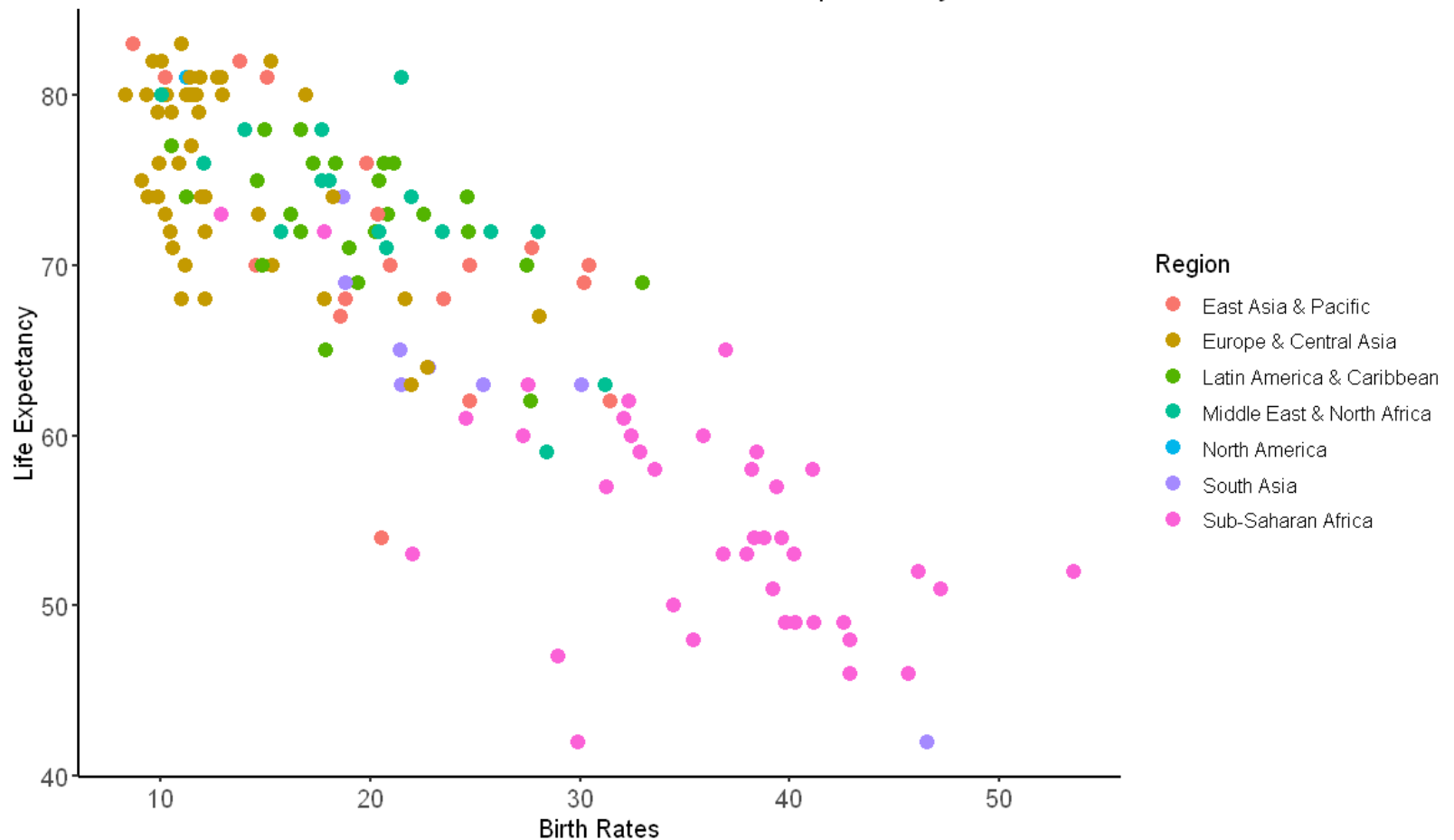
R - Scatter Plot

```

In [7]: 1 options(repr.plot.width =10, repr.plot.height =6)
2
3 ggplot(demographics, aes(x=birthrates, y=expectancy)) +
4   geom_point(size = 3, aes(color=Region)) +
5   theme_classic() +
6     theme(text = element_text(family="sans",size =12, color="black"), element_line(size = .6),
7           plot.title = element_text(size = 16), axis.text.x = element_text(size=12),
8           axis.text.y = element_text(size=12))+
9   ylab("Life Expectancy") +
10  xlab("Birth Rates") +
11  ggtitle("R - Scatter Plot: Countries Birth Rates vs. Life Expectancy For Year 2008")
12

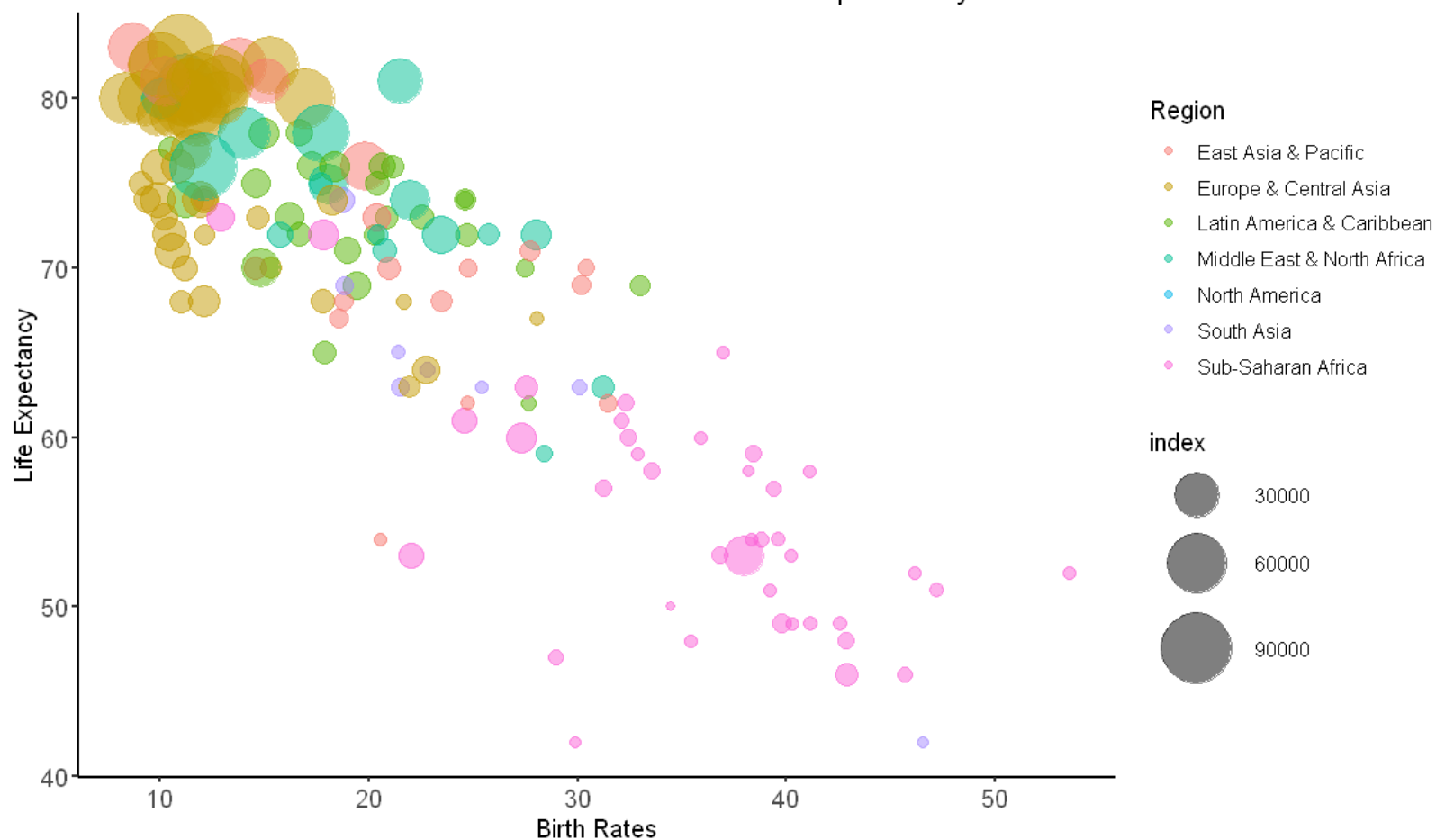
```

R - Scatter Plot: Countries Birth Rates vs. Life Expectancy For Year 2008



```
In [12]: 1 options(repr.plot.width =10, repr.plot.height =6)
2
3 ggplot(demographics, aes(x=birthrates, y=expectancy, size=gdp)) +
4   geom_point(aes(color=Region), alpha=0.5) +
5   scale_size(range = c(2,18), name="index") +
6   theme_classic() +
7   theme(text = element_text(family="sans",size =12, color="black"), element_line(size =.7),
8         plot.title = element_text(size = 16), axis.text.x = element_text(size=12),
9         axis.text.y = element_text(size=12))+
10  ylab("Life Expectancy") +
11  xlab("Birth Rates") +
12  ggtitle("R - Bubble Chart: Countries Birth Rates vs. Life Expectancy For Year 2008")
```

R - Bubble Chart: Countries Birth Rates vs. Life Expectancy For Year 2008

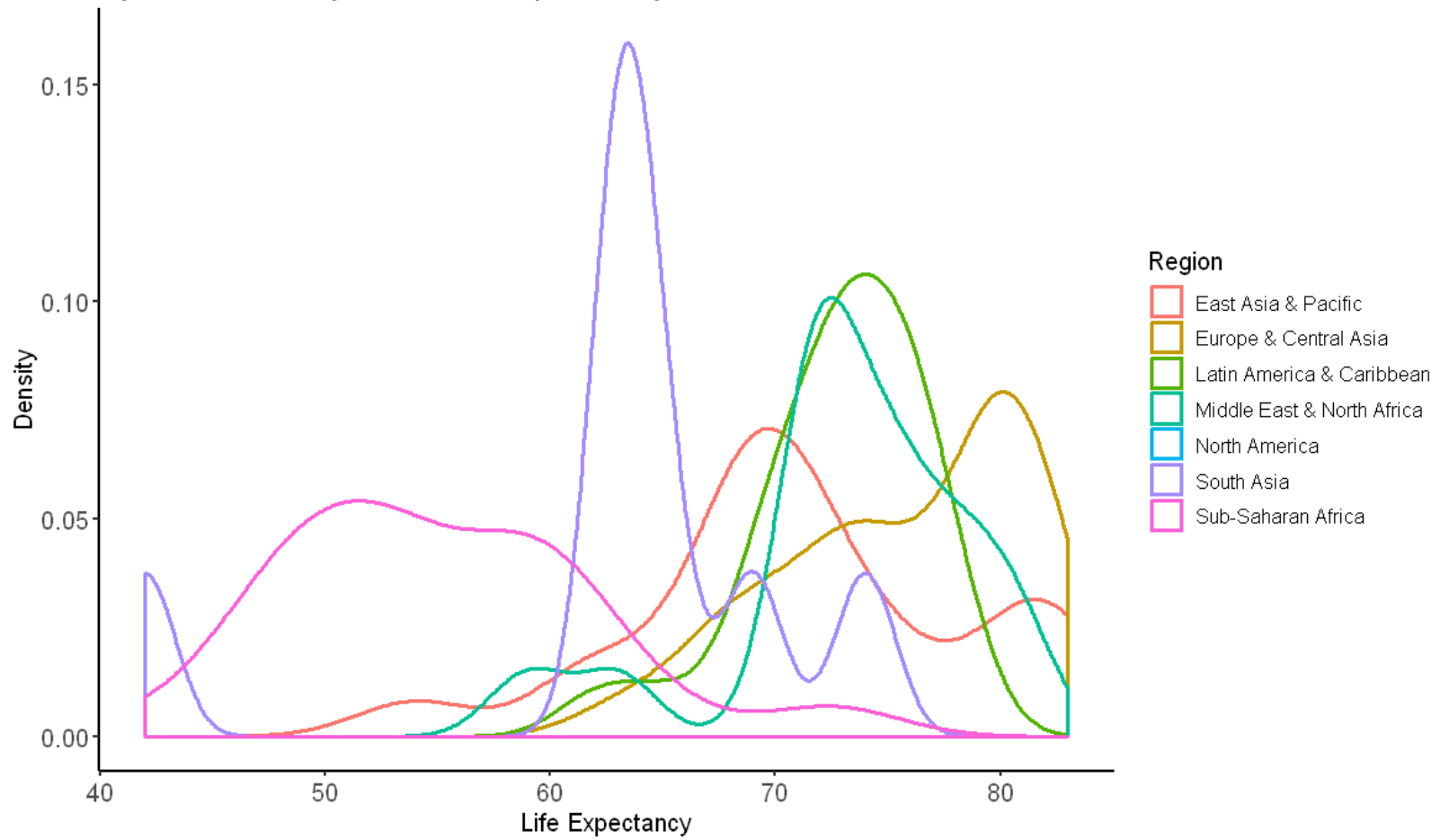


```
In [22]: 1 options(repr.plot.width =10, repr.plot.height =6)
2
3 ggplot(data = demographics, aes(x = expectancy, color =Region)) +
4 geom_density(size = 1) +
5 theme_classic() +
6   theme(text = element_text(family="sans",size =12, color="black"), element_line(size = .6),
7         plot.title = element_text(size = 16), axis.text.x = element_text(size=12),
8         axis.text.y = element_text(size=12))+
9   xlab("Life Expectancy") +
10  ylab("Density") +
11  ggtitle("Python - Density Plot: Life Expectancy For Year 2008")
12
```

Warning message:

"Groups with fewer than two data points have been dropped."

Python - Density Plot: Life Expectancy For Year 2008



In []:

1