Sowndeep SivaRama Tellagadda - 10509470

Vinay Reddy Mallugari - 10388072

Data visualization Assignment-2

Interactive Visual Analysis

# **Assignment 2: Interactive Visual Analysis**

## Step 1: Data Selection

The dataset contains the information about all the countries in the world. It contains Information on population, region, area size, infant mortality and more.

* **Data is compiled from US Government 1970-2017.**
* **Dimension of the data set:** 227 rows and 20 columns
* The dataset is taken from the Kaggle repository and can be found in the below link:
* <https://www.kaggle.com/fernandol/countries-of-the-world>
* The dataset contains one csv file and we used that.

Step 2: Interactive Visual Analysis

We have performed exploratory analysis on the selected dataset using a visualization tool such as Python, matplotlib, RShiny, D3, Seaborn and Bokeh etc. I have used the visualization tool named python, seaborn, matplotlib to Visualize my selected dataset.

**The dataset contains the following variables:**

1. Country: country name
2. Region: Region Name
3. Population: count of population
4. Area(sq.mi) : Displayed in sq.mi
5. Population.Density(sq.mi): Displayed per sq.mi
6. Coastline (coast/area ratio): Shows coastal area
7. Net migration: Shows people migrate to other countries
8. Infant mortality (per 1000 births): Displays Infant mortality per every 1000 births
9. GDP($ per capita): Gdp per capita
10. Literacy(%): Literacy Rate
11. Phones (per1000): Phones for every 1000 people.
12. Arable(%): Shows percentage of agricultural land.
13. Crops: Crops percentage
14. Other: Others percentage
15. Climate: Climate readings
16. Birthrate: Birth count
17. Deathrate: Death count
18. Agriculture: agriculture
19. Industry: industry
20. Service: Service

The questions raised by exploring dataset are mentioned under each visualization

**Initial Analysis Questions**

## How do you visualize the top 100 birth and death rates of different countries in the world?

## How do you visualize the last 100 birth and death rates of different countries in the world?

## What is (Service, industry and agriculture) sectors percentage of world countries?

## What are the agriculture and service features of top 7 countries?

## What are the industry and service features of top 7 countries?

## What is the GDP rate of top 7 countries?

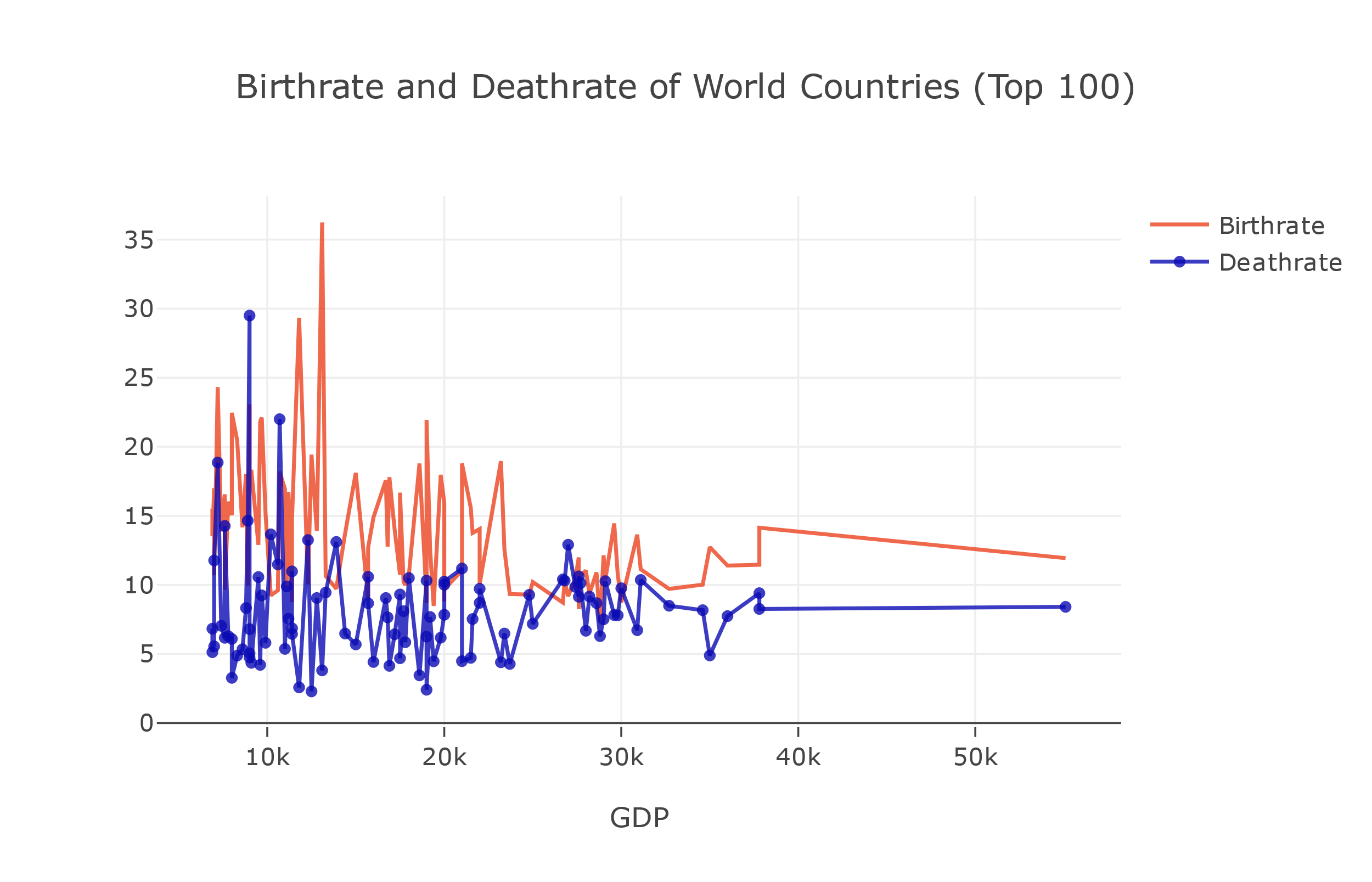
* What are the service and literacy rates of countries?
* What is the population per each country?
* Feature selections or correlation analysis?
* What is the relationship between variables?

**VISUALIZATIONS**

Visualization 1:

## **Question1:**

## How do you visualize the top 100 birth and death rates of different countries in the world?

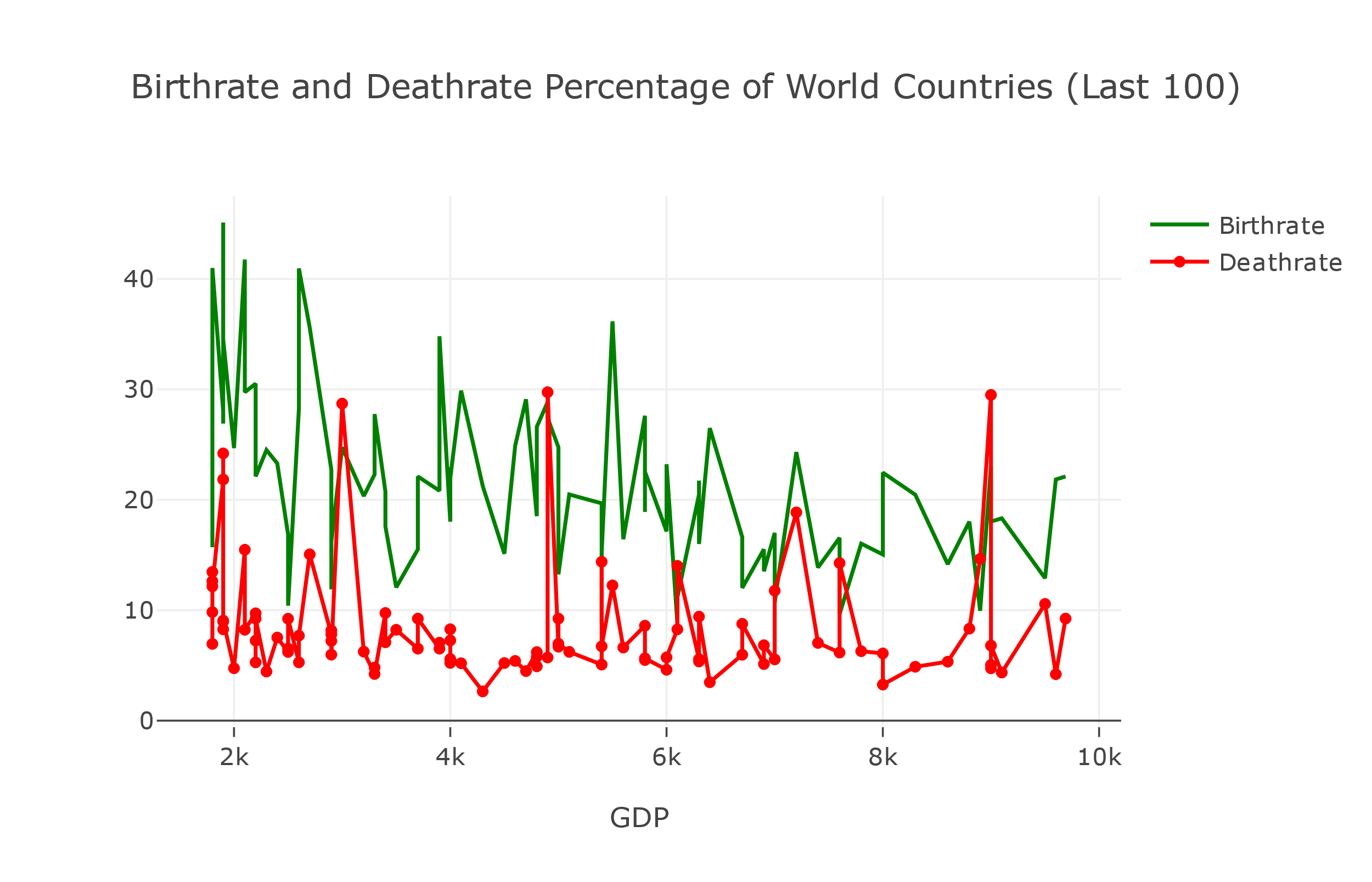


* Above graph is plotted by taking top 100 countries with high gdp.
* From the above visualization, it is clearly observed that as birth and death rates increases then Gdp rate decreases.
* The birth rates are higher than death rates almost at every point.

Visualization 2:

**Question 2:**

## How do you visualize the last 100 birth and death rates of different countries in the world?

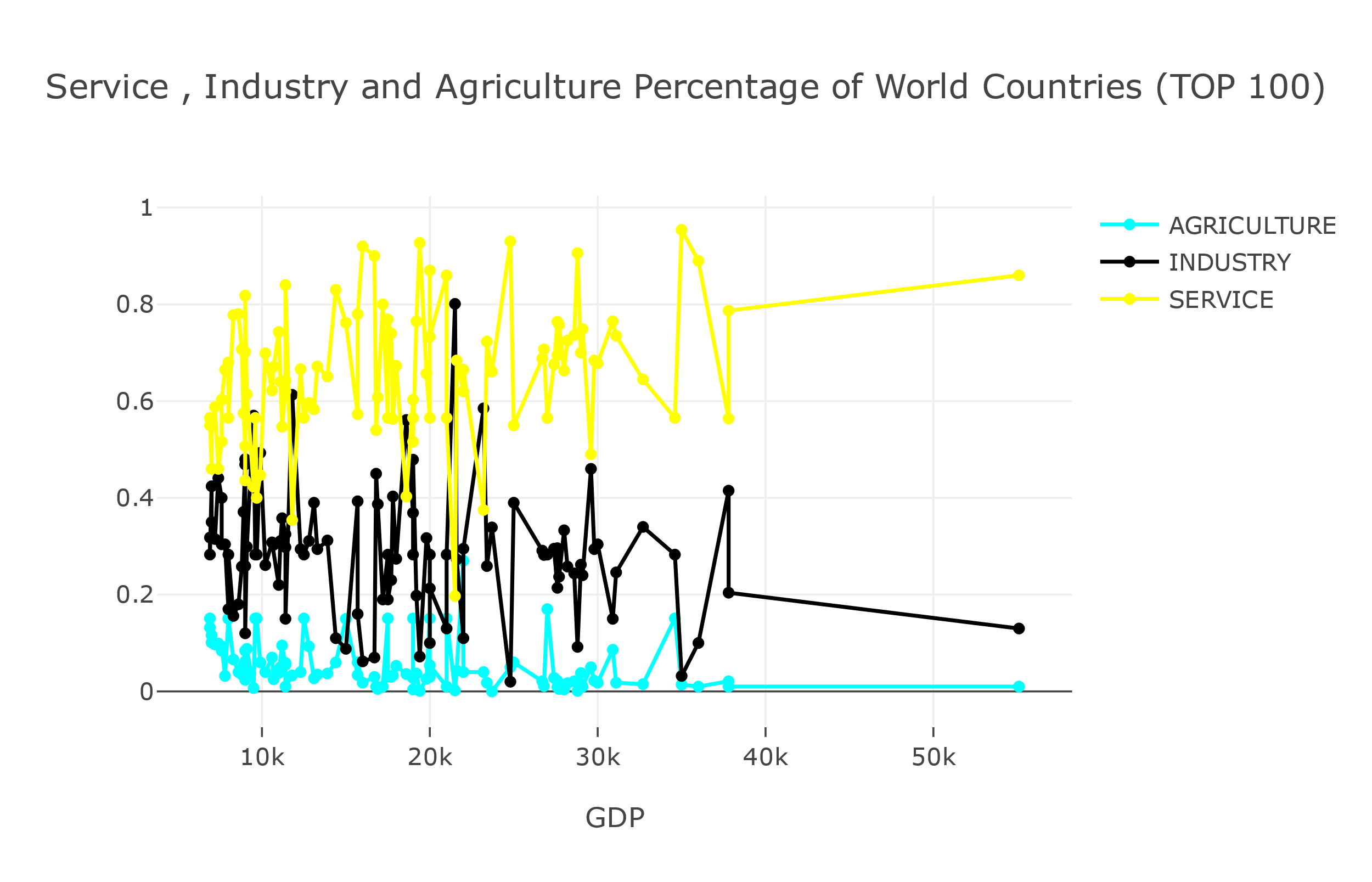


* Above graph is plotted by considering 100 countries with least gdp rates.
* From the above visualization we can observe that as the Death rate increases the gdp rate also increases gradually.
* As the birth rate increases gdp rate decreases unlike deathrate.

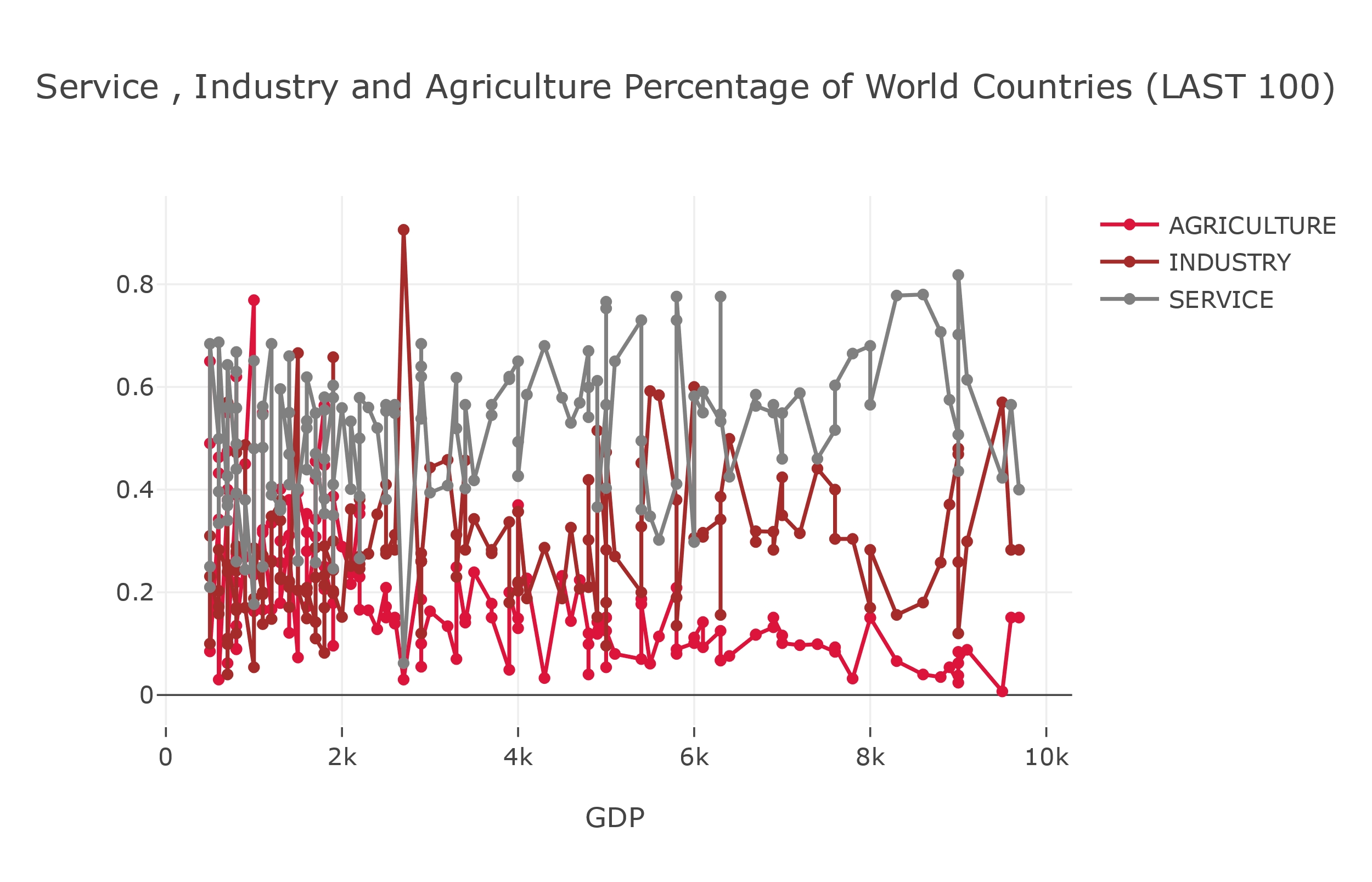
Visualization 3:

**Question 3:**

What is (Service, industry and agriculture) sectors percentage of world countries?



* Graph is plotted by taking top 100 countries with higher gdp percentage.
* From the above visualization we can observe that as the service rate increases the gdp rate also increases.
* In most of the countries as both industry and Agriculture rate decreases gdp rate is elevated.

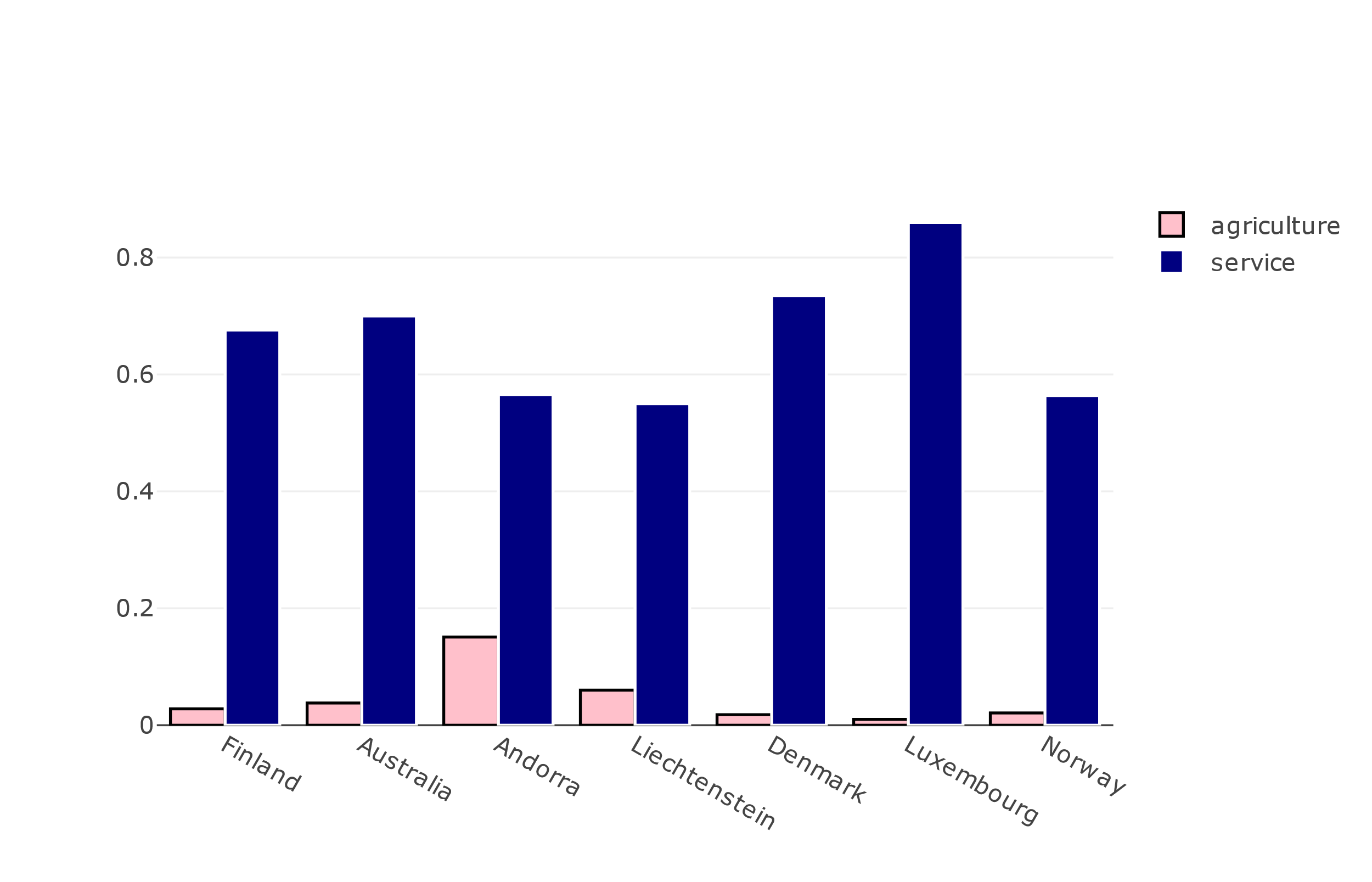


* In most of the countries as the agriculture rate decreases gdp rate increases.
* As service rate increases the gdp rate also increases.
* Same with the industry, Industry rate increases when gdp rate is also high.

Visualization 4:

## **Question4:**

## What are the agriculture and service features of top 7 countries?

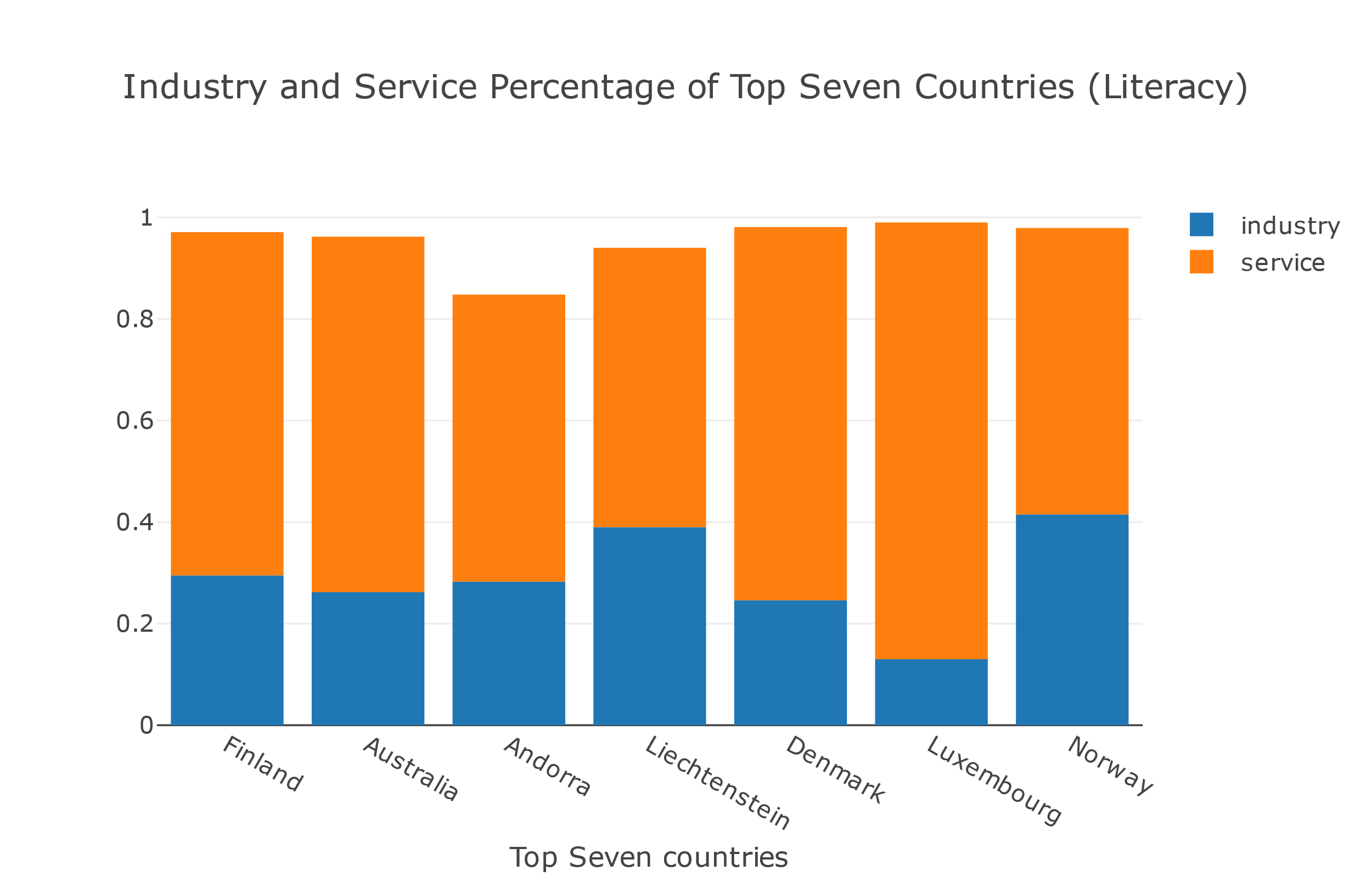


* From the above plot we can clearly see that the service sector have higher percentage compared to agriculture sector for every country.
* Luxembourg have higher service percentage compared to all other countries. In contrast Luxembourg stood at least position than other countries in agriculture sector.

Visualization 5:

## **Question 5:**

## What are the industry and service features of top 7 countries?

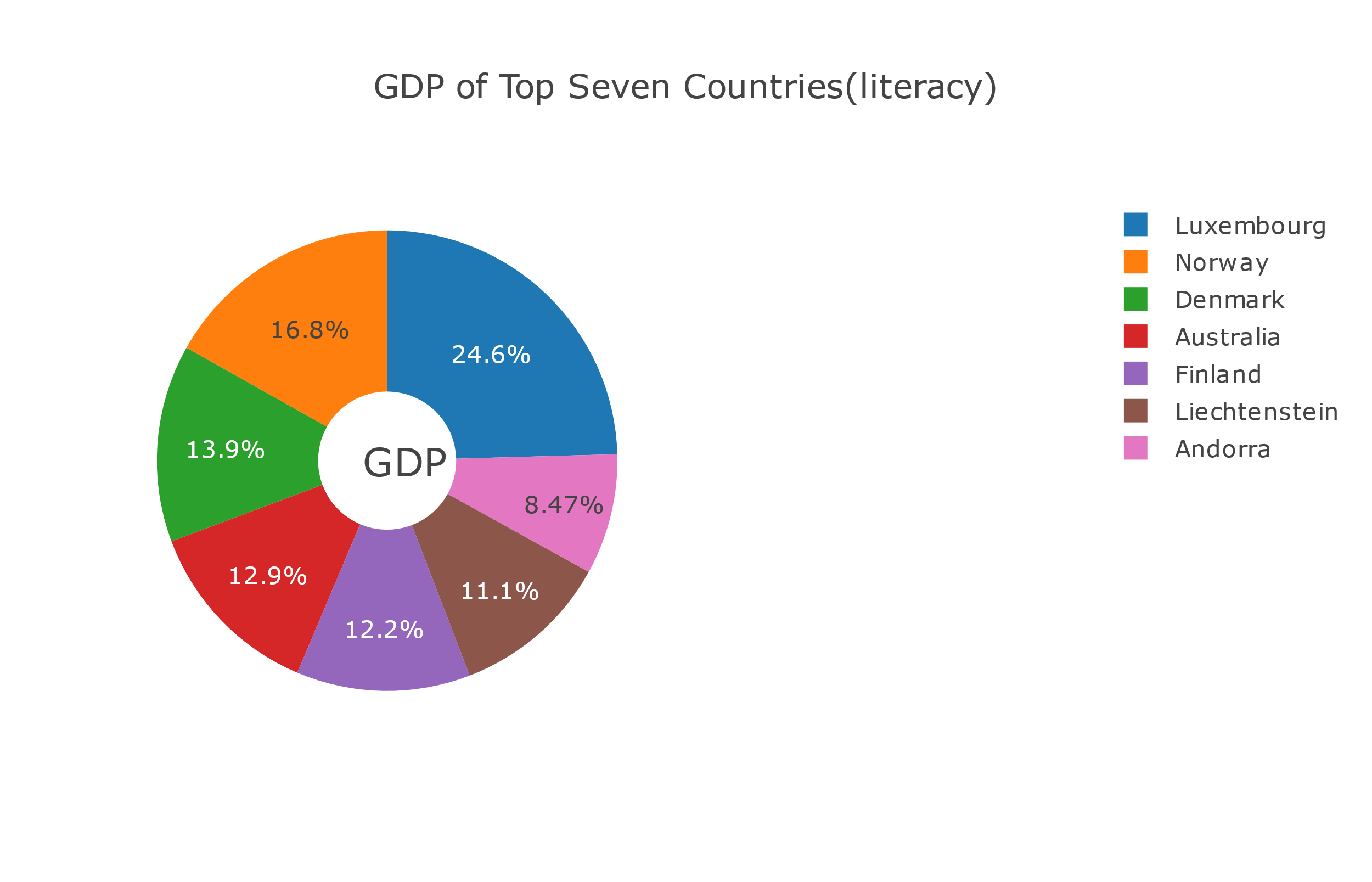


* The above graph is plotted by taking top seven literacy rate countries.
* We can clearly see that service sector have more percentage than industry sector for every country.
* Luxembourg have highest service percentage and Norway have higher industry percentage.

Visualization 6:

## **Question 6:**

## What is the GDP rate of top 7 countries?



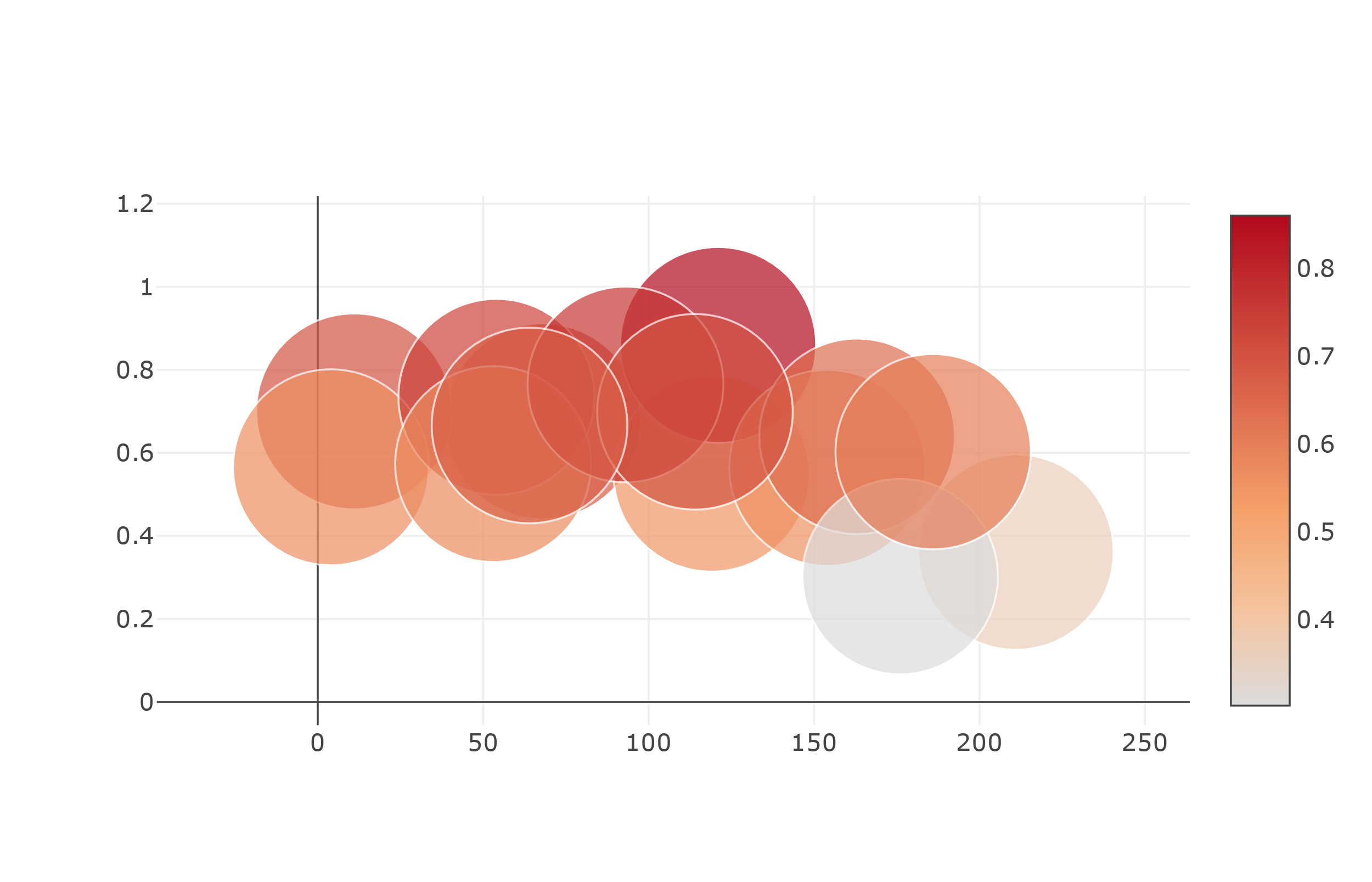
* Above graph is plotted by taking top seven literacy rate countries.
* Luxembourg stood in first place with 24.6% in gdp whereas, Andorra is at last position with 8.47% gdp rate.

Visualization 7:

## **Question 7:**

What are the service and literacy rates of countries?

Bubble chart with service and literacy of countries

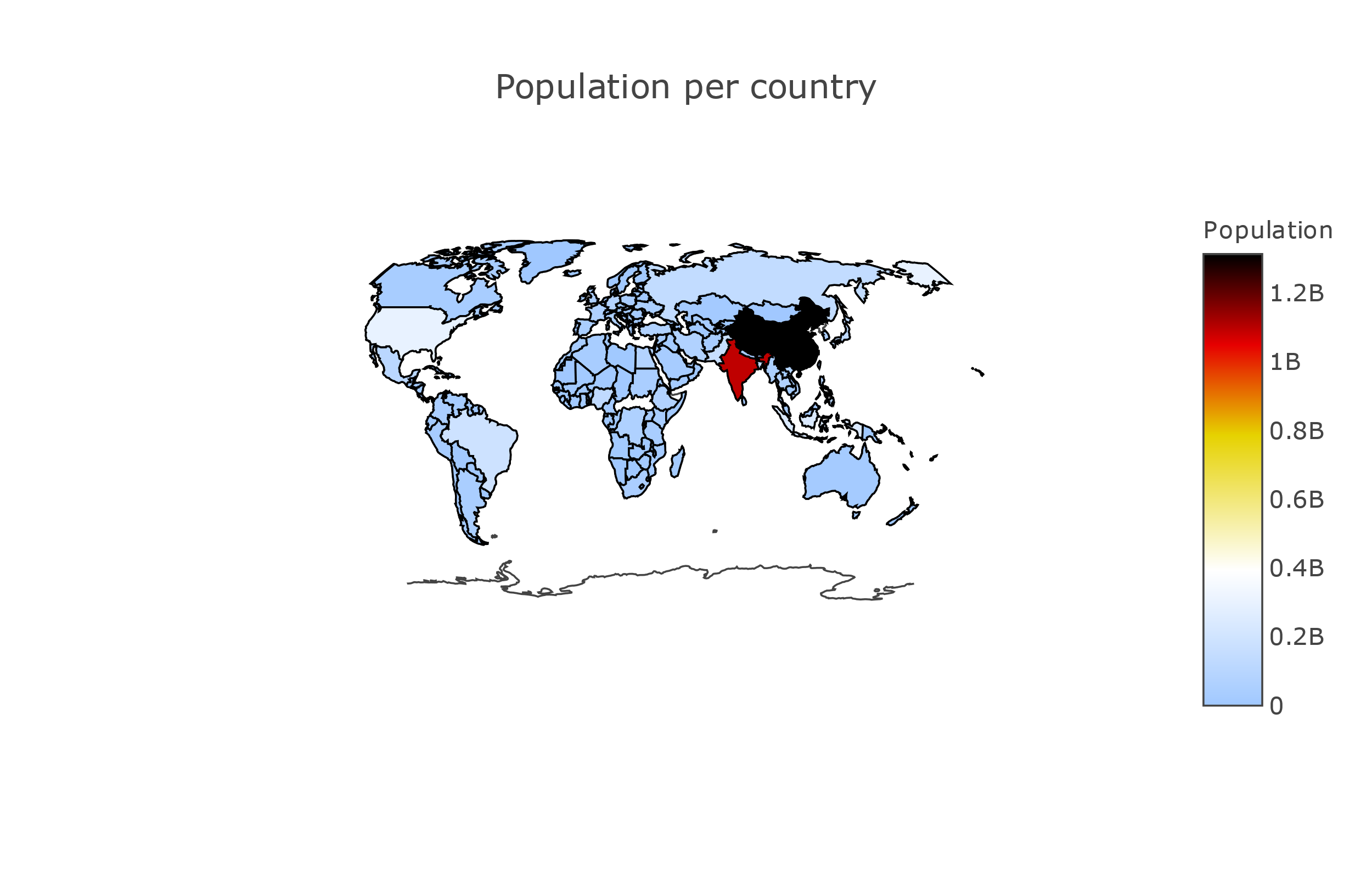


* The above bubble chart is plotted by considering service and literacy sectors.
* The countries with high services percentage contains moderate literacy rate.
* The countries with average service percentage contains low literacy rate and countries with low service rate have high literacy percentage.

Visualization 8:

**Question 8:**

What is the population per each country?

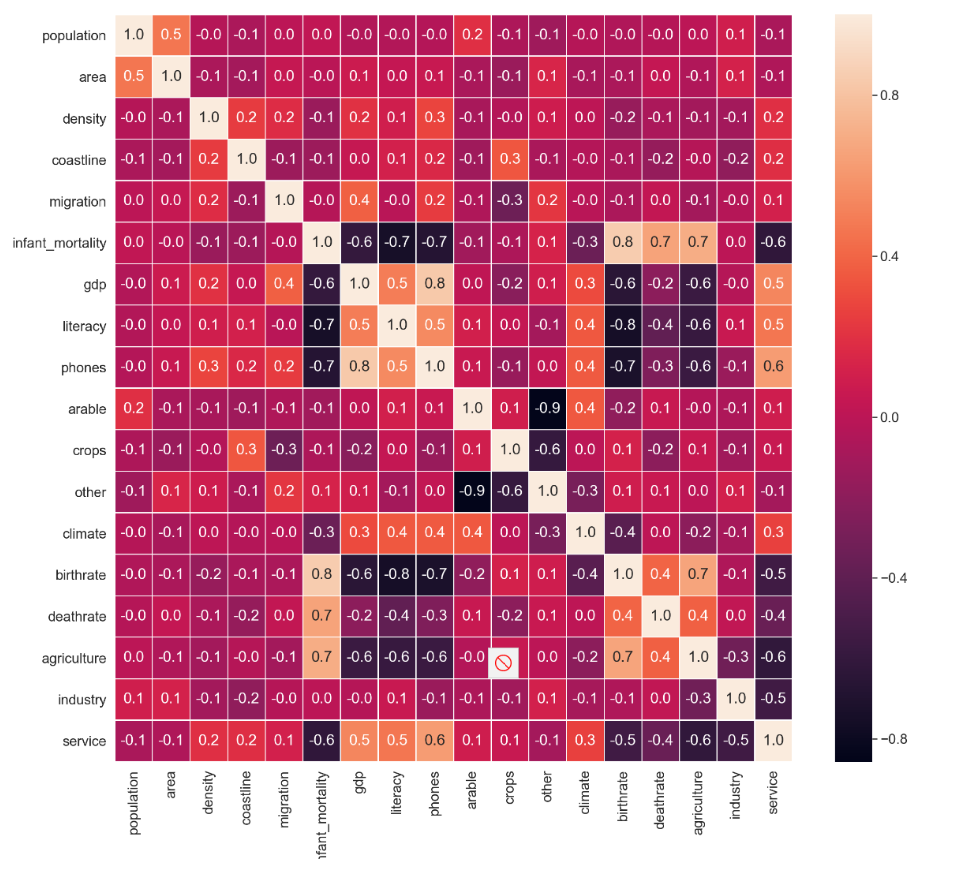
****

* The above plot depicts population count for each country.
* China takes the first position with around 1.4 billion, following by india with 1.2 billion.

Visualization 9:

**Question 9:**

Feature selections or correlation analysis?

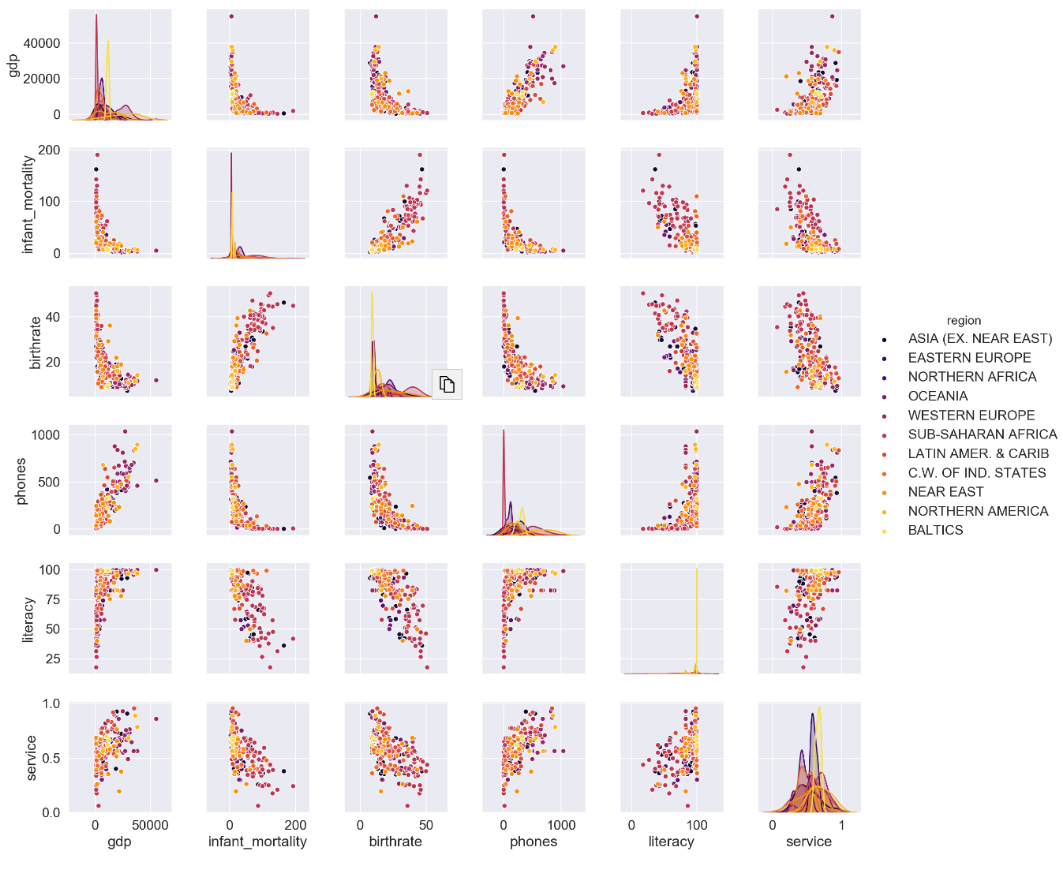


* Heat maps shows the correlation between features which are related.
* The primary purpose of heat maps is to better visualize the volume of locations/events with in a dataset.

Visualization 10:

**Question 10:**

What is the relationship between variables?



* It is used to explain the relationship between two variables or to form the most separated clusters.
* It also helps to form some simple classification models by drawing some simple lines.

**Conclusion:**

* Overall, Luxembourg stood at first place in service sector.
* After taking top 7 countries with high literacy rates, Again Luxembourg stood first in gdp rating and Andorra is at last position with 8.47 percentage.
* China and India is at first and second positions in population comparison.
* China also stood in first place in land area.