Global Country Information

Vishal Maurya

INTRODUCTION

In this visualization project, we set out to analyze and visualize various key indicators related to global demographics, land area, life expectancy, and environmental impact. The main objectives included understanding the distribution of population, population density, land area, and life expectancy at a global scale, as well as exploring their distribution across continents. Additionally, we sought to investigate the relationships between different attributes and portray them visually.

DATA DESCRIPTION

This comprehensive dataset provides a wealth of information about all countries worldwide, covering a wide range of indicators and attributes. It encompasses demographic statistics, economic indicators, environmental factors, healthcare metrics, education statistics, and much more. With every country represented, this dataset offers a complete global perspective on various aspects of nations, enabling in-depth analyses and cross-country comparisons.

Country: Name of the country.

Density (P/Km2): Population density measured in persons per square kilometer.

Abbreviation: Abbreviation or code representing the country.

Agricultural Land (%): Percentage of land area used for agricultural purposes.

Land Area (Km2): Total land area of the country in square kilometers.

Armed Forces Size: Size of the armed forces in the country.

Birth Rate: Number of births per 1,000 population per year.

Calling Code: International calling code for the country.

Capital/Major City: Name of the capital or major city.

CO2 Emissions: Carbon dioxide emissions in tons.

CPI: Consumer Price Index, a measure of inflation and purchasing power.

CPI Change (%): Percentage change in the Consumer Price Index compared to the previous year.

Currency_Code: Currency code used in the country.

Fertility Rate: Average number of children born to a woman during her lifetime.

Forested Area (%): Percentage of land area covered by forests.

Gasoline_Price: Price of gasoline per liter in local currency.

GDP: Gross Domestic Product, the total value of goods and services produced in the country.

Gross Primary Education Enrollment (%): Gross enrollment ratio for primary education.

Gross Tertiary Education Enrollment (%): Gross enrollment ratio for tertiary education.

Infant Mortality: Number of deaths per 1,000 live births before reaching one year of age.

Largest City: Name of the country's largest city.

Life Expectancy: Average number of years a newborn is expected to live.

Maternal Mortality Ratio: Number of maternal deaths per 100,000 live births.

Minimum Wage: Minimum wage level in local currency.

Official Language: Official language(s) spoken in the country.

Out of Pocket Health Expenditure (%): Percentage of total health expenditure paid out-of-pocket by individuals.

Physicians per Thousand: Number of physicians per thousand people.

Population: Total population of the country.

Population: Labor Force Participation (%): Percentage of the population that is part of the labor force.

Tax Revenue (%): Tax revenue as a percentage of GDP.

Total Tax Rate: Overall tax burden as a percentage of commercial profits.

Unemployment Rate: Percentage of the labor force that is unemployed.

Urban Population: Percentage of the population living in urban areas.

Latitude: Latitude coordinate of the country's location.

Longitude: Longitude coordinate of the country's location.

DATA SOURCE: https://www.kaggle.com/datasets/nelgiriyewithana/countries-of-the-world-2023

OBJECTIVES

Using the above dataset, we are trying to visualize and intrepret the following:

- Distribution of Population, Population density, Land Area and Life Expectancy among the Countries

- Distribution of Countries, Population and Land Area among the Continents

- Distribution of Population Density, Life Expectancy and Co2 Emission within the Continents

- Finding out significant linear relationships between different attributes and visualise them

The above objectives can be achieved using histograms, bar plots, piechart and box plots for univariate analysis, scatter plots and line plots for bivariate analysis, and gradient scatter plots for multivariate analysis.

EXPLORATORY DATA ANALYSIS

Distribution of Population, Population density, Land Area and Life Expactency among the Countries
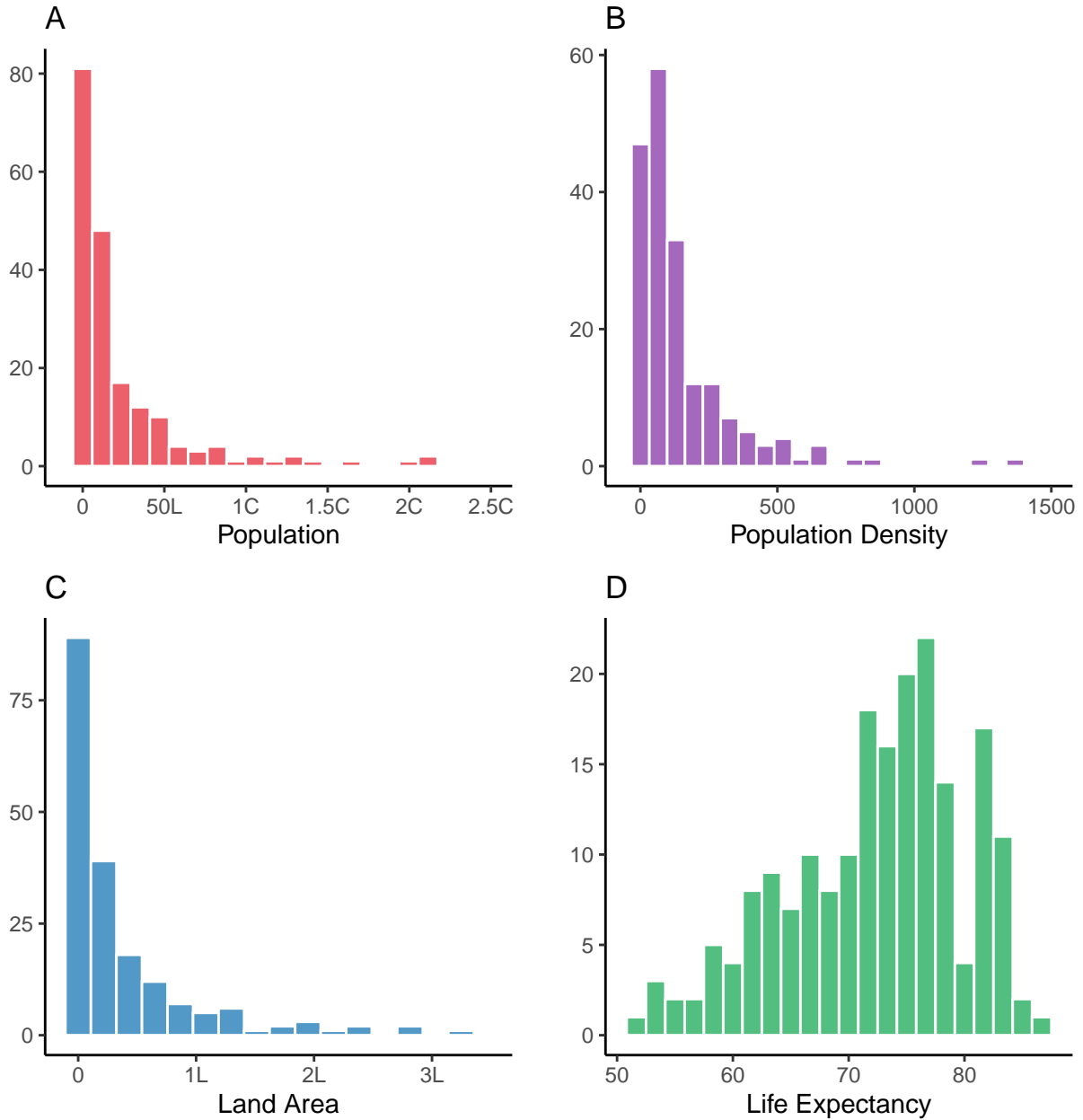

A. Most of the Population is very condensed to the inital ranges and drop down exponentially.So,it is very rare to found a country having population more than 2 crores.
B. Just like population, population density is also condensed in initial ranges and drop down exponentially.So,it is very rare to found population density more than 700 people per km square area.
C. Similarly for Land Area,it is very rare to find a country having Land Area more than 2 Lakhs km square.
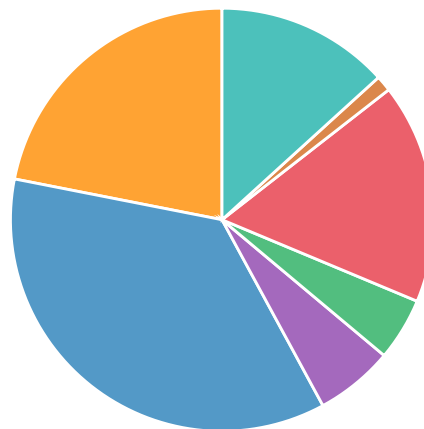**D.** Unlike other attributes,Life Expectancy is negatively distributed over the age range of 55 to 85.

Distribution of Countries, Population and Land Area among the Continents

A
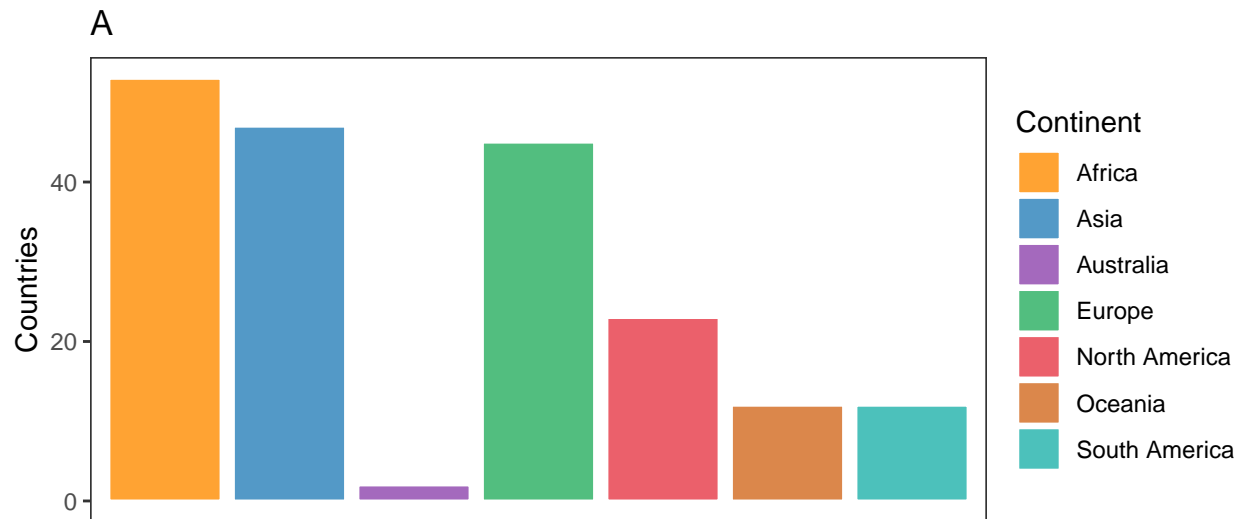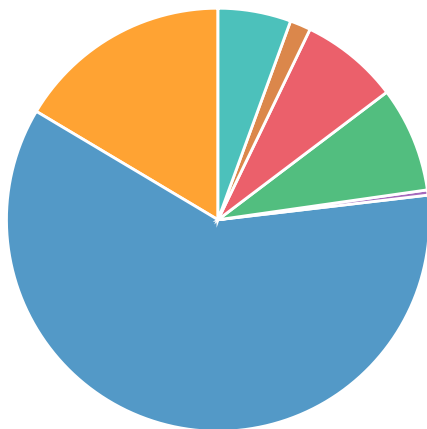

B. Population Distribution   C. Area Distribution

A. Most of the countries belongs to Asia, Africa and Europe.
Note: There are many countries which are part of more than one continent.Like Russia,s 3/4th land is in Asia and remaining in Europe.
B. Asia and Africa together have the approximately 3/4th of the total population.Continent Australia and geographical region Oceania have least population.
C. Most of the land area is in under Africa, Asia, America(Both) and Asia-Europe region.

Distribution of Population Density,Life Expectancy,Co2 Emission within the Continents
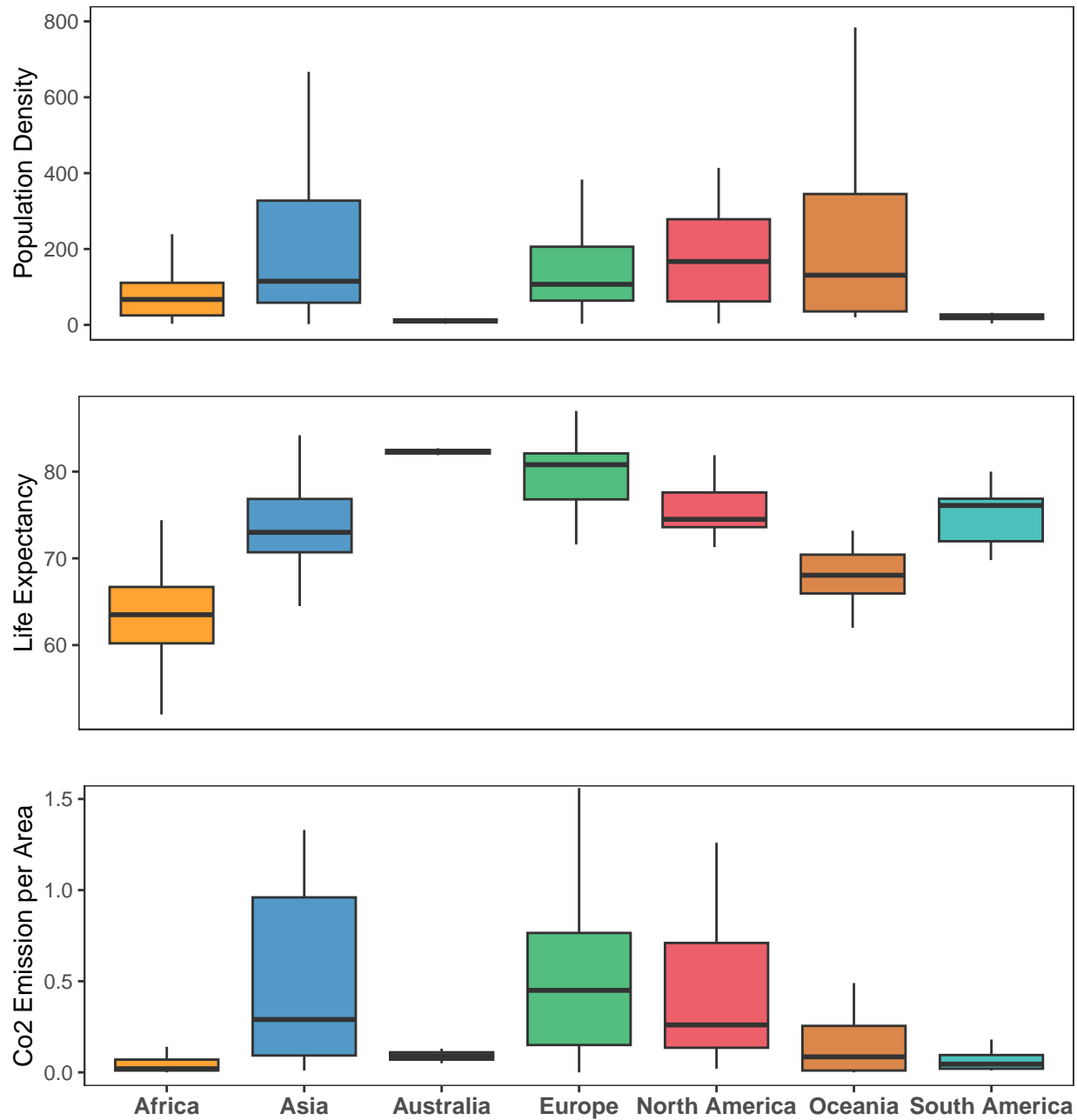
A. Population Density of Australia and South America is very condensed and very less.That of North America,Oceania and Asia is very much spread and higher than others.

**B.** Life Expectancy in Africa is very low and in Australia and Europe is high.

C. Co2 Emission is comparatively very less in Africa,Australia and South Africa and highest in North America. Note: We have taken Co2 Emission per km square area, for comparison

A and C : Australia and South America have least Co2 Emission contributed in their highest life expectancy but there can be others reason too.

Mutivariate Ananlysis

Analyze linear relationships between different attributes and visualise them

- We observed that most of the variables are showing a stronger linear relationship if variables are transformed by log10.

- It is very obvious that Land area will have a significant correlation with Agricultural Land.So, we have considered Agricultural Land Percentage with respect to total Land Area to have a meaningful comparison and correlation.
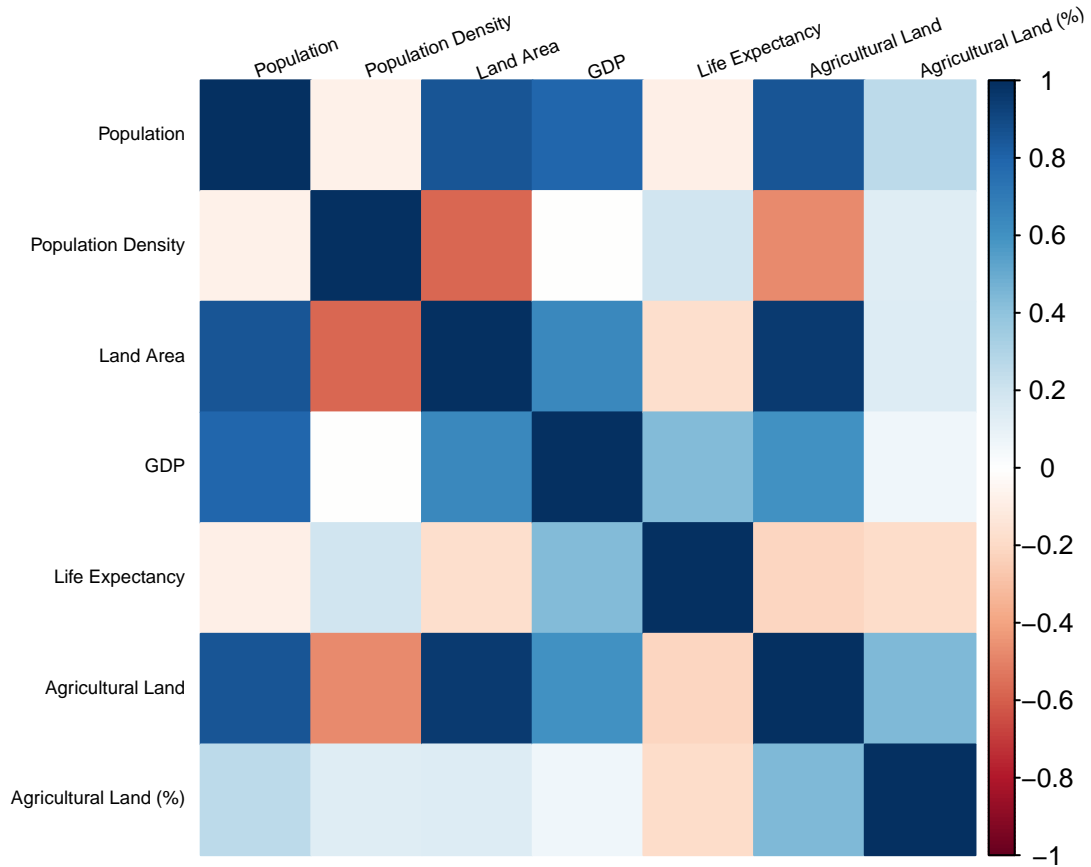

- One strong inference is that GDP is positively correlated with Agriculture Land significantly. This shows a significant contribution of Agricultural production in GDP.

- Land Area is positively correlated with Agricultural Land significantly. Which is very obvious. So, Land Area, GDP and Agricultural Land are positively correlated with each other.

- Agricultural Land is highly correlated with the population. This is may be because higher the population,higher is the food consumption.

A : Population versus Land Area with Life expectancy and Continents
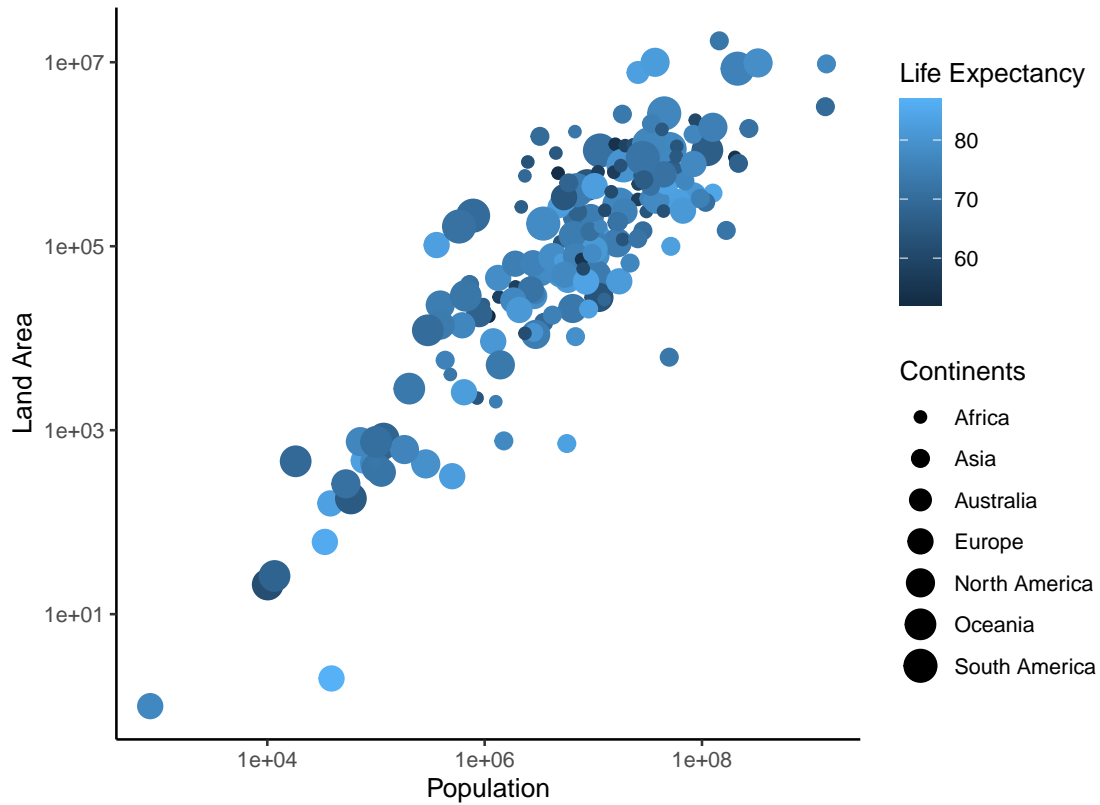
B : Agriculture Land versus GDP with Population and Continents
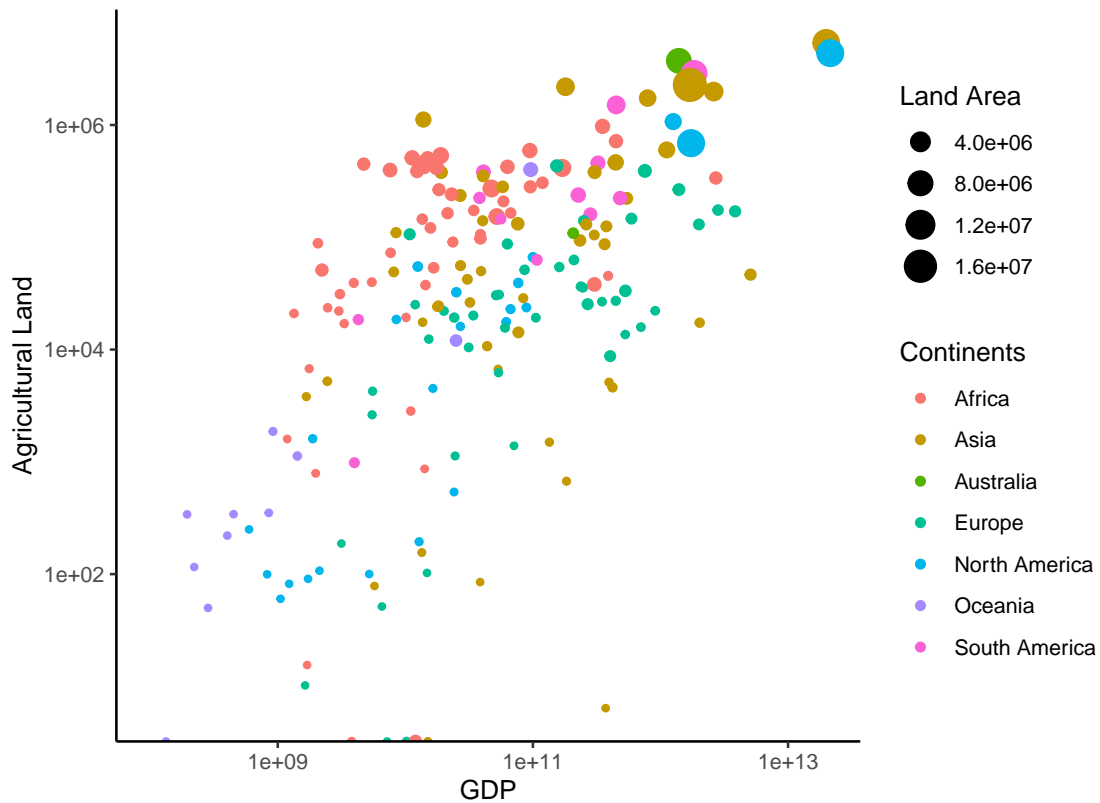
For A - The scatter plot is condensed along a line with slope -1. So,Land area and Population are significantly positively correlated when both are transformed with log10. - All the small circles are condensed to higher Population and higher Land Area.So, African and asian countries are lying in this region.

For B - Initially graph was very condensed to lower GDP and lower Agriculture land(This happens in most of the cases when both the variables are condensed initially, and drop down exponentially). So, we had change the scale of X and Y by log10 transformation to visaulise it in a better way. - This scatter plot clearly shows that there is a significant positive correlation among Land Area, GDP and Agricultural Land.

CONCLUSION

Through these visualizations and analyses, we have gained a comprehensive understanding of the global landscape in terms of population distribution, land area, life expectancy, and their interplay with environmental factors. These insights are invaluable for informed decision-making and policy formulation, aiding in addressing global challenges and fostering sustainable development.