Coursework

Applied Statistics and Data Visualisation

(Principles of Data Science)

MSc Data Science

THE ANALYSIS OF THE SOCIOECONOMIC STATUS OF THE G7 COUNTRIES, BRAZIL, CHINA, AND NIGERIA ACROSS TEN YEARS

Oluwaniyi Grace Oluwaseun

@00600719

December 2022

Table of Contents

Part One: Interactive Dashboard Design

1. Introduction

The Social and Economic development of a country is measured over different indicators in different sector which ranges from the health sector to the educational sector to the import and export sector, even the self-development sector also known as the Human Development Sector and so much more.

In this study, the analysis was done based on the measurement of the Economic conditions and the measurement of the Human Development Index and Component of the selected countries. We also looked at how the countries progressed over time in the urban and rural areas. The data collected from the World Bank Databank contains indicators such as Population, Gross Domestic Product (GDP), Gross National Income (GNI), Life expectancy, Final Consumption and so on.

For this report, a total of 10 countries were selected over a period of 10 years (2011-2020). Seven of these countries are the G7 (Group of seven) countries, also known as the most industrialised countries with advanced economies in the world. They are Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States. Nigeria, the country with the largest GDP in Africa was also selected. It is also the most populous country in Africa. China was also selected because of its rank as the country with the second largest economy in the world. It is also the most populous country in Asia. The last country that was selected for the research purpose is Brazil. A country with the largest economy in South America and is also the most populous in its continent. The selection of countries cuts across 5 out of the 7 continents of the world which are Europe, North America, South America, Asia, and Africa which is a good basis and selection for the study of the Social and Economic development of countries in the world.

At the end of this study, we would have seen how these countries with high GDPs and high number of citizens have grown over 10 years in both sectors. We would also see the relationship between these indicators and cross examine the living situations in these countries.

2. Background Research

It is now easier than ever to visualise data thanks to contemporary technology. A large table of raw data can be virtually instantly transformed into a beautiful and simple to understand diagram with a few mouse clicks. Prior to the 17th century, data visualisation was mostly used in maps, but better visualisations were required as the requirement for more precise mapping and physical measurement increased. A Flemish astronomer named Michael Florent Van Langren is said to be the first to develop the first visual depiction of statistical data in 1644. The one-dimensional line graph displayed the twelve estimations of the longitude difference between Toledo and Rome at the time (Insightsoftware, 2022). Before Tableau and Power Bi, most people would use PowerPoint, Microsoft Excel, and such tools to create charts to visually explain a variable or tell a story (Allen Hillery, 2020).

A dashboard is a visual depiction of the most crucial data that must be gathered and organised on a single screen to allow for quick access to information to accomplish one or more goals. The best technologies for designing a dashboard now are, Power Bi and Tableau. A dashboard's interactive features, such as filtering and drill-down, should be simple to use. It should be usable by everyone and not just executives (Arimetrics, s.d.).

An interactive dashboard is a tool used in the management of data that tracks and monitors key business and economic metrics in a consolidated format. With the availability of different interactive features, users can explore a selected dataset on a deeper level and make well-informed, data-driven business or economic decisions. A dashboard can be operational, analytic, and strategic depending on its service (Ratibor Sekirov, 2021).

For this study, the Tableau dashboard was selected for its data visualisation prowess. It is one of the fast-growing data visualisation tools in the Data Science world. With a very easy, interactive, simple, and fast design, Tableau is extremely popular. It accepts very large data sets and allows the user present results and dashboards in understandable and eyes-pleasing displays (Neelam Tyagi, 2019).

The Economic state of a country can be measured based on its macroeconomic performance (gross domestic product [GDP], consumption, investment, and international trade) (The World Bank, s.d.). The Human Development Index (HDI) is an index that measures key major aspects of human development. The three key aspects are long and healthy life which is measured by life expectancy, Access to education which is measured by expected years of schooling and a good standard of living which is measured by Gross National Income per capita (Max Roser, 2014).

The G7 countries will be compared with the 3 largest countries both in Economy and Population in Africa, Asia, and South America to study the social and economic growths or decrease in the space of 10 years. With the dashboard, the aim is to see the changes that has happened over time using the measurement tools i.e., indicators that have been selected. This dashboard is to be easily understood, it is to be interactive and is to portray the increase or decrease recorded over time, also considering the rural and urban population statistics of the countries.

The interactive dashboard will contain charts such as Line charts, Bar charts, Dual Line plots, OpenStreet Maps. These will be used because they are easily readable by the audience, and they vividly show the progression of these indicators over time.

3. Exploration of Data Set

The data downloaded from the World Bank website was downloaded in the CSV format. The data was first imported into R for pre-processing. Columns like ‘Country Code’ and ‘Time Code’ were completely removed as they did not contribute to the study. The missing values were inputted using the mean of the values for the available years. This was done according to the different countries, indicators and the time frame which is 10 years. A basic descriptive analysis was also carried out to see the different statistical values such as the mean, mode, and quantiles. Also, to see if there were any outliers that were going to influence the study.

The cleaned data was then imported into Tableau through the connect page. The dataset contains 101 rows and 16 columns. Twelve indicators were selected for this study. They are Population, Urban Population, Rural Population, Gross Domestic Product (GDP), Per Capita Income (GDP), Final Consumption Expenditure (% of the GDP), Trade (Domestic and Foreign trade), Life Expectancy, Gross National Income (GNI), GNI Per Capita, Labor Force, and Unemployment (% of the Labor Force). The first three for the population analysis, next four for measuring and comparing the economic conditions and the last five for measuring the Human Development Index. For HDI, Expected Years of Schooling and the Percentage of Adult education would have been selected as one of the indicators for this measurement, but these indicators and the one relative to these only had inputs for the year 2018 on the World bank and UN sites, and as a result, this indicator was left out. Labor Force and Unemployment (% of Labor Force) were then used instead for the measurement of HDI.

The study was divided into three broad areas following the selected indicators. The areas are Population, Economic conditions, and Human Development Index.

Here is a brief explanation of the indicators chosen for this study.

Population – It measures the total number of citizens of a country.

GDP – It measures the monetary value of final goods and services. It is the average productivity of each citizen.

GDP per Capita – it is GDP divided by midyear population.

GNI – It is total amount earned by the citizens of a country through jobs, businesses, etc.

Final Consumption Expenditure (% of the GDP) – Consists of the total expenditure spent by a household or group on goods and services used to satisfy needs and basic wants and the total government expenditure. It is measured as a share of GDP.

Trade – It is the transfer of goods and services between two or more people. The unit for exchange is usually of monetary value. For us, Trade includes both foreign (exchange between two or more people of different countries) and domestic (trade done between two people of same nationality) trade. It is measured as a share of GDP.

Life Expectancy – It is defined as the number of years, on average, that an individual is expected to live for.

Labour Force – Is the number of people, age 15 and older, that supplies labour to produce goods and services. This includes employed citizens and unemployed citizens looking for jobs.

Unemployment (% of Labour Force) – Refers to the percentage of people in the Labour Force currently unemployed but seeking for jobs.

We would be comparing the relationship between GDP, GDP Per Capita Income, Final Consumption Expenditure and Trade, also the relationship between GNI, GNI per Capita, Life Expectancy, Labor Force and Unemployment.

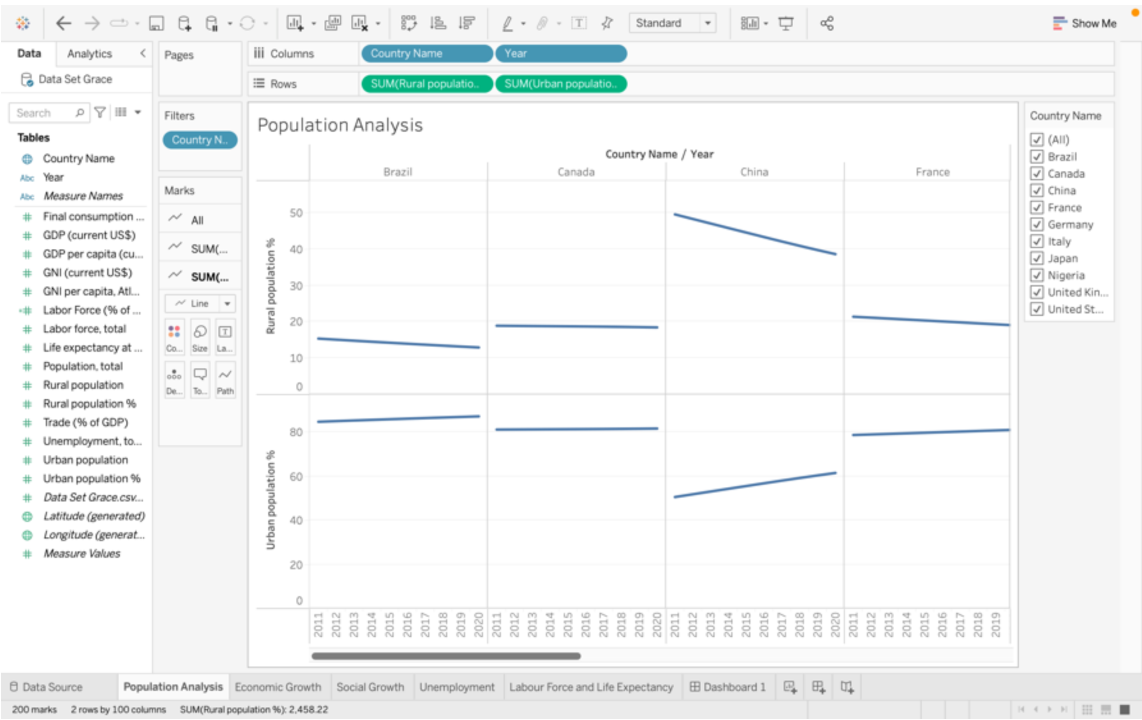
4. Investigation of Data Workflows & Proposal for Design of Dashboard

To build a dashboard, the first step to be carried out is to create different worksheets which will then be added to a single-screen dashboard. This comes after connecting the cleaned dataset to Tableau. The icon for this is located at the bottom left of the page. Clicking on this brings us to the blank worksheet where visualisations are created based on the objectives of the dashboard, i.e., what we really want the audience to see. In this area, variables are dragged and dropped on the row or column area to create the visualisations. The variables from the dataset are found on the left side of the worksheet. From the icons on the left of the variable names, we can know the type of the variable, whether it is a string, numeric or even countries. There are other tools such as the filter section, pages section, mark section which allows the user streamline and add filters to aid the visualisations, or even use colours or texts to aid the audience understanding. As we know, the use of colour (not too much) and annotations are important to help the audience the visualisation, particularly if they are part of a data story.

For this report, we aim to see the situation and progression of the economic and human development of the G7 countries, Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States, the country with the highest GDP in Africa - Nigeria, the country with the highest GDP in Asia - China and the country with the highest GDP in South America - Brazil over the period of ten years, 2011 – 2020. We would be comparing the varying situations in these countries. Secondly, we would be studying the relationship between the selected indicators for measuring economic growth, and the selected indicators for measuring human development index for each country over time. Thirdly, we hope to see the association between Life Expectancy and Labor Force.

The variable type for the variable 'Year' was changed to string type from number type because this allowed for a more suitable visualisation. A new measure called Labor Force (% of Population) was created using Analysis under Calculated Field on Tableau. This was done to know the percentage of the Labor Force that formed the total population for every year. This new variable was used in the comparison of Life Expectancy versus Labor Force.

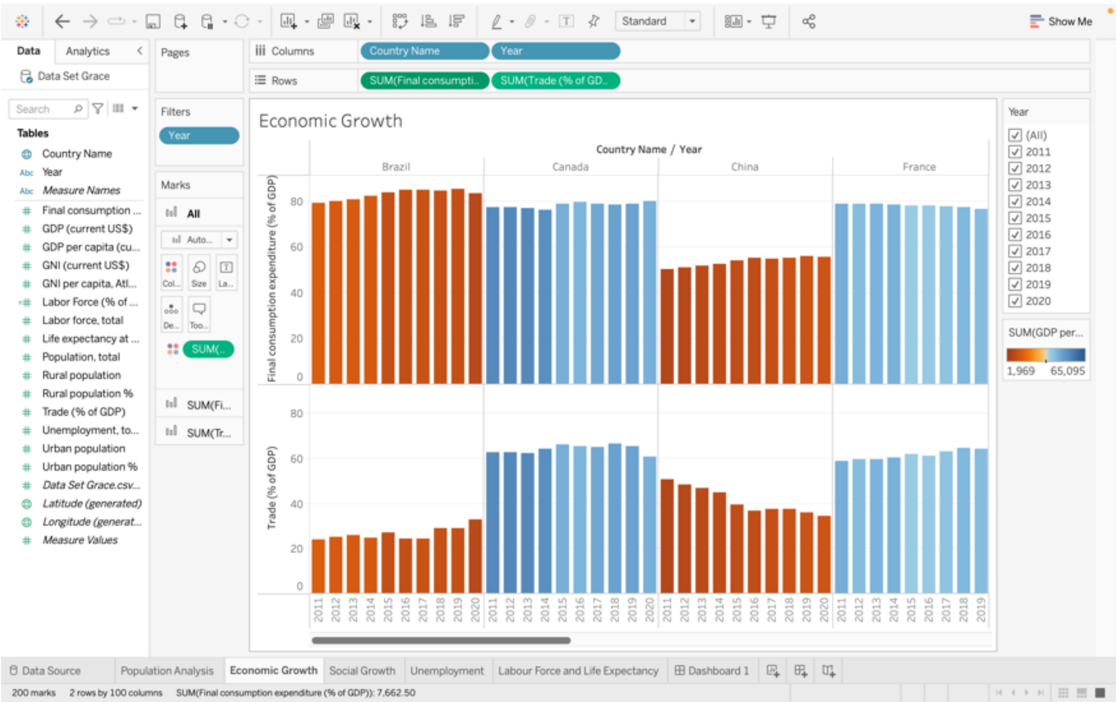
Comparison of Urban and Rural Population



A dual line chart was chosen for this visualisation because it clearly shows the growth for each year for both rural and urban population. From the line plots for each country, we can see that there are countries with little or no change, countries with remarkable difference and countries with very little difference. For Canada, Germany, and Japan, over the years, there was little or no record of a variation in the proportions of both urban and rural population. For China and Nigeria, we noticed a remarkable change in the proportion of urban to rural population. There is a remarkable drop in the percentage of rural population as there is an increase in that of urban population. For Brazil, France, Italy, UK and USA, the change in proportion is not as much as the previous countries. We can see that over the period of 10 years, some countries were able to reduce the percentage of rural population significantly, and others, little or none.

Economic situation over the years

Here, we used a dual bar chart to show the variation of Trade and Final Consumption Expenditure. We used colours (orange to blue diverging) to show the Gross Domestic Product for each year for each country. With the lowest GDP being 1,969 and highest being 65,095. We see the variation of Final Consumption and Trade for each year in units of % of GDP.

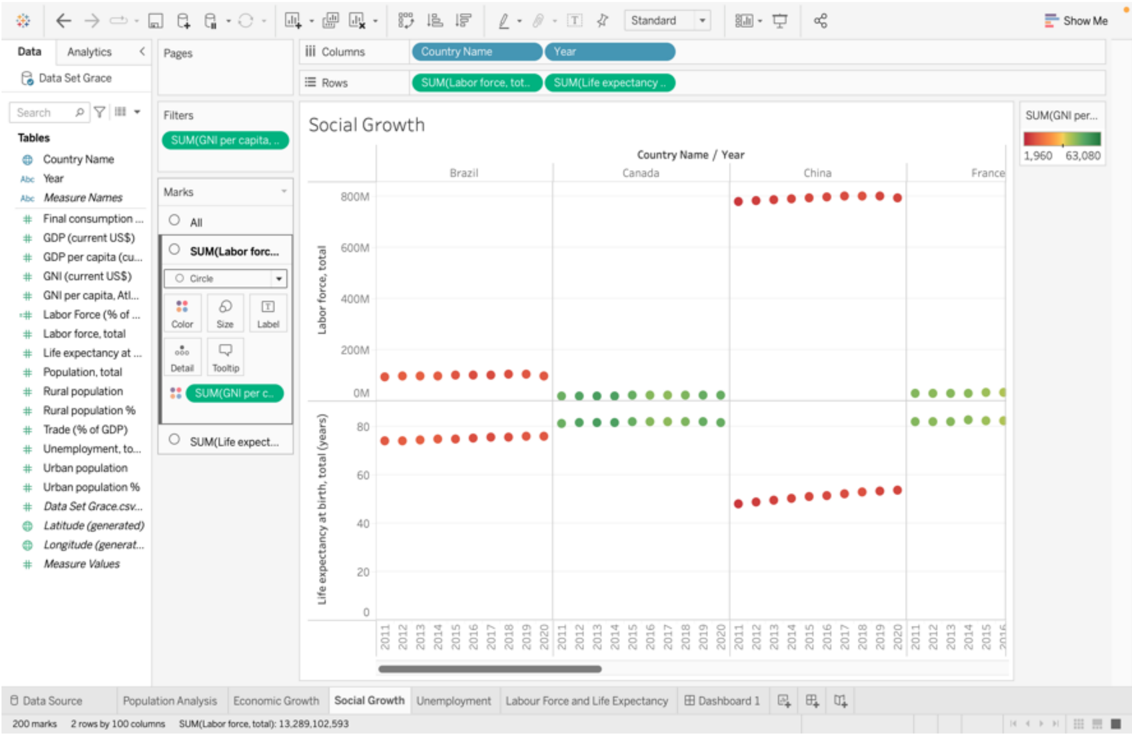


If GDP falls, it means there's a negative growth in the economy. An increase means that the economy of a country is doing well. A rise in Final Consumption implies a rise in GDP as this goes in same direction. This is expected as, the more citizens earn, the more they spend which brings about growth in the economy. It is said to be the highest GDP component (Valentino Piana, 2011). The same pattern goes for Trade indicator. The growth between GDP and Trade is in the same direction.

For most countries of our countries, we see that they either remain on the side with high GDP (e.g., United States) or the side with low GDP (e.g., Nigeria). There are of course, fluctuations over the years but it remains in the same range for most countries except Italy. We see very serious changes in their GDP over the years.

Social (Human Development Metrics) situation over the years

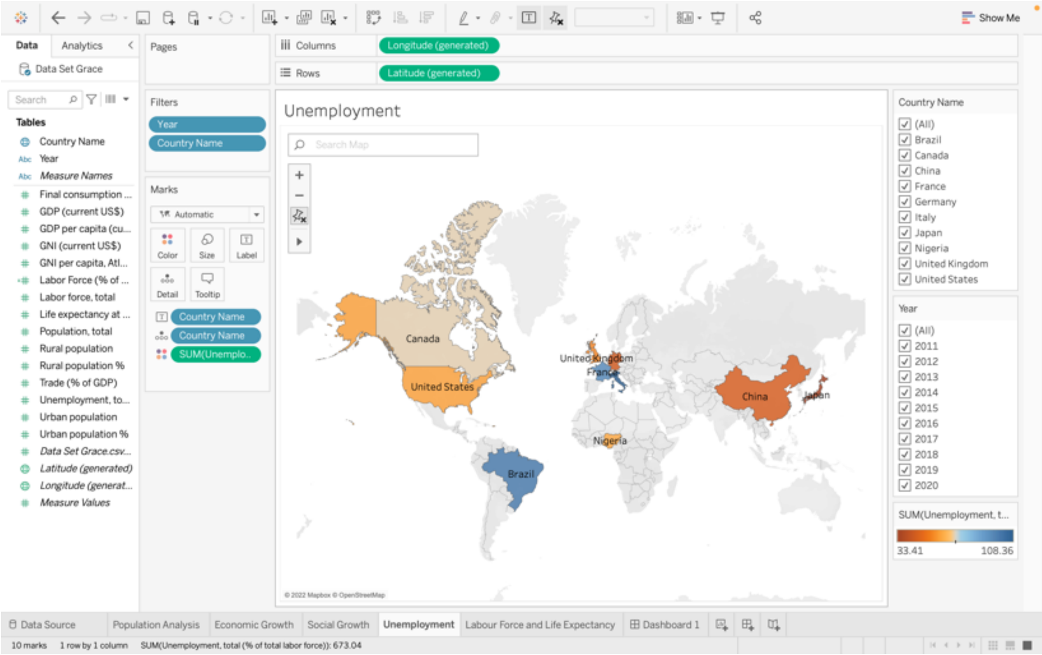
Here we are looking Life Expectancy and Labour Force using a dual chart with circles as the indicator for the points. This choice works well as we can visually see the relationship between the indicators, whether it is parallel or not. We would look at how these factors have changed over the years. And how GNI has progressed over the years, using the colour filter.



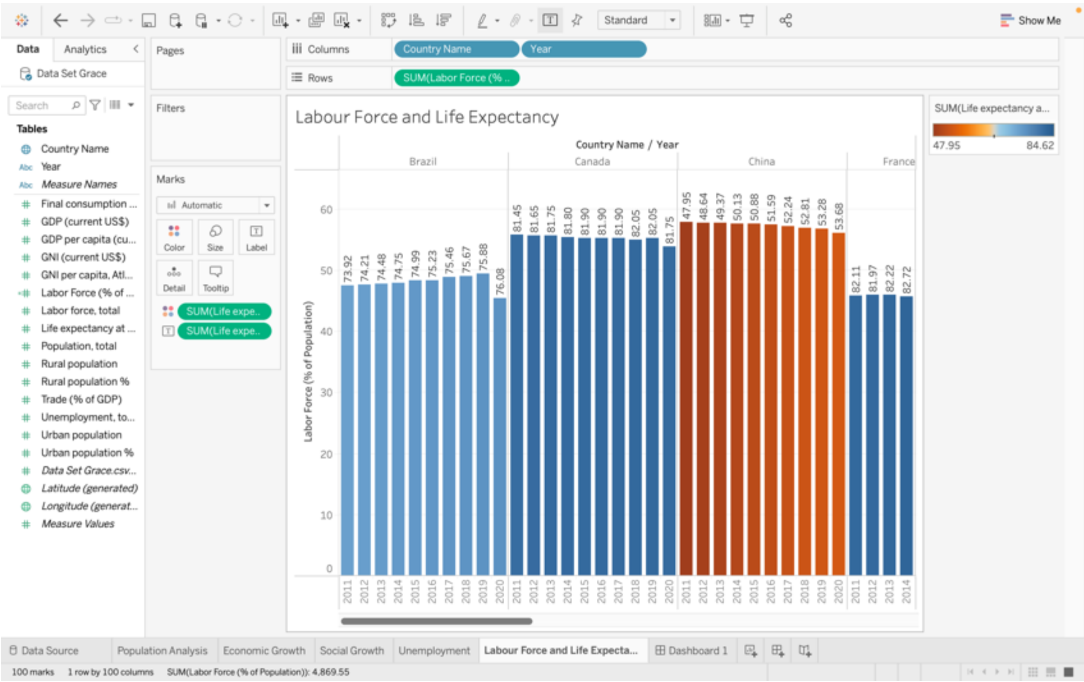
We see that Life Expectancy and Labour Force has almost remained same for most countries over the year as the points are also in a straight line. Italy recorded a stealth growth in the numbers for Life Expectancy over the years. Nigeria also recorded growths but in a like manner as Italy. United States recorded a visible reduction in Life Expectancy in 2020. Nigeria had the lowest GNI recorded over the years with 1,960 in 2018. United States had the highest GNI recorded over the years with 63,080 in 2018. Once again, we see a bit of fluctuation with the GNI of Italy.

Unemployment rate

Using the latitude and longitude coordinates, we created a geographic visual on OpenStreet Map as we look at the unemployment rate for the 10 countries. We used a colour filter to view this on the map. This is the number of people without a job amongst the population of people in the labour force. We see that Japan has the lowest Unemployment rate within the 10 selected countries. Italy has the highest rate of unemployment.



Life Expectancy and Labour Force

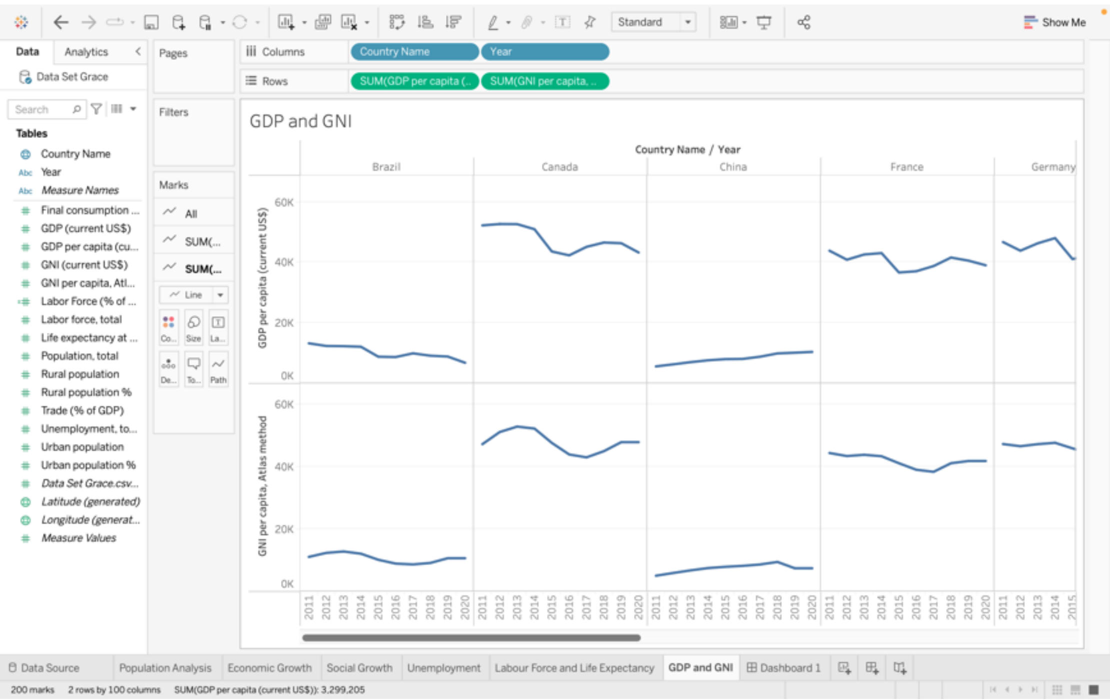


A new measure called Labour Force (% of Population) was created as earlier stated. We want to see the progression of the number of people registered to be on the Labour force for every year and how they make up the calculated percentage of the population using horizontal bar charts. We also look at the progression of Life Expectancy over the years for each country. This is highlighted using the orange-blue diverging colour filter. The label on each bar also speaks to Life Expectancy.

An increase in Life Expectancy suggests a growth in the social development of a country. Amongst the ten countries, we can say five out of them had a constant social growth over the years because of the increase in years of Life Expectancy. The other countries, France, Germany, Italy, United Kingdom, and United States had varying growth over the years. There were years where the age for Life Expectancy reduced.

We read that the Labour Force is an estimate of a country's economy workforce (Adam Hayes, 2022). With this indicator and unemployment percentage, we can see some perspective of the country's economic and social (in our case, human development) state. Labour force rate fluctuates over time based on social, economic, and demographic trends. We see countries like Canada, China, and France with decrease over the years, countries like Brazil, Germany (with a reduction in 2015), Italy (decline in 2013 and 2015) and Japan, countries like Nigeria, United Kingdom and United States with increases and decreases over the years. For every country, there is a very visible drop in the Labour force rate in the year 2020. We would assume this is because of the Covid-19 pandemic.

Gross Domestic Product (GDP) and Gross National Income (GNI)



The chart chosen for this visualisation is a dual line chart as it is suitable to show how the countries fared economically and socially over the years. Here, we just want to see how GDP, our economic growth metric and GNI, our Human Development metric progressed over the years and if there is a relationship between both indicators.

We can see there's a close relationship from the maps. They progress in an almost same manner.

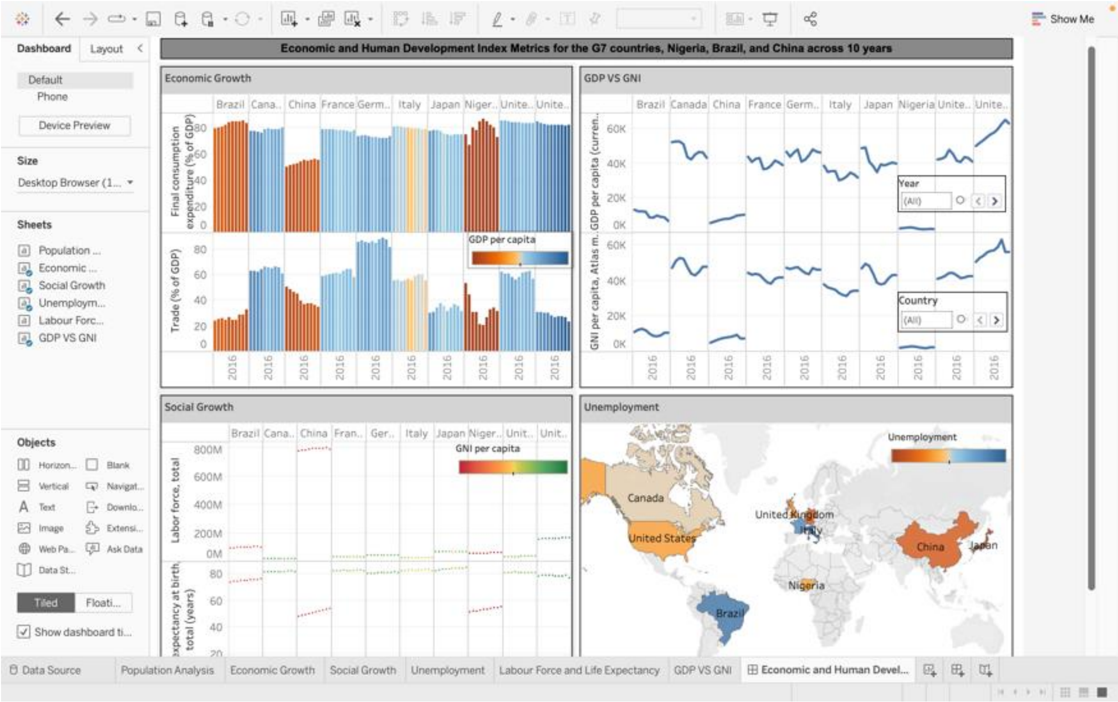
5. Discussion

I wanted my dashboard to portray the objectives of my analysis in a very clear and concise way. Not too noisy with too many worksheets and colours so that the users and readers are able to understand the results of the analysis. I also researched past works and publications on dashboard to see the best way and formats such as fonts, to use. Some examples of the dashboards seen are the NHS e-Referral Service (e-RS) dashboard (NHS Digital, s.d.) and the UK dairy trade dashboard by AHDB (AHDB, s.d.). I went through my worksheets to also select the ones that were going to best describe my objectives.

It is normal for most humans to read from left to right, top to bottom, so I knew I had to place my sheets following hierarchy, starting with the most important sheet. I knew that using a lot of colours would be very distracting for my users, so I used colours only where necessary. I also wanted my dashboard to be interactive and not just a picture that could be done using tools like Microsoft PowerPoint, so I made sure to add that to my dashboard (Sandra Durcevic, 2022). I made sure I used simple fonts that would be readable by everyone including those with some form of sight deficiency. I also got rid of some headers that weren't important to declutter my dashboard and make it as simple but functional as possible. I changed the settings of my filters to floating to enable me to move them to anywhere on the dashboard. I made sure my sheets were sized to fit to enable the users see all the data. I also changed the filter for the years and countries to the slider format.

My dashboard was titled 'Economic and Human Development Index Metrics for the G7 countries, Nigeria, Brazil, and China across 10 years'. This was suitable and self-explanatory as it describes the dashboard just like the charts and plots. As the dashboard wasn't created just for the Non-Government Organisation (NGO) where I work as a Data Scientist but also for the wider audience, I knew that every detail that could play as a tool for communicating my findings to the public that includes people not in the field, had to be utilised.

For the dashboard, I selected the sheets that analysed the economic and human development metrics. I also used the sheet that contains the OpenStreet map to show the unemployment rates in the selected countries. I also used the sheet that shows the progression of GNI VS GDP across the ten years.



From the dashboard, we can see the changes that occurred over the ten years duration for each of the indicators. We can see how the countries fared economically (using GDP) vs socially (using GNI). We see the relationships between the indicators and how for some of the indicators, there was constant growth for certain countries, and in some cases, it was a decline for certain countries. For some indicators, it was a case of fluctuations, where there was either an increase or a decrease, over the years.

6. Conclusions

From our study, we can conclude that for our selected countries, Brazil, Canada, China, France, Germany, Italy, Japan, Nigeria, United Kingdom, and United States, there will need to be a review of the structures of measures and infrastructures put together to grow these countries both economically and socially. Yes, we see that for all countries, there was an increase in the proportion of urban areas to that of rural areas which could be translated to urbanisation of the nations, but we need to also reiterate the fact that these countries haven't recorded a forward growth over the years.

There were a lot of decline in numbers and percentages over the years. Take for example, Italy, for the GDP and GNI indicators which are our main metrics for the measurement of economic and social growth, we see serious fluctuations in these metrics over the ten years.

We also noticed how the G7 countries, have much more larger GDP rates and GNI rates to the other countries. Looking at Nigeria, we see how the GDP and the GNI is very far apart to that of Japan. We can therefore say that indeed, the G7 countries are the countries with the best economies in the world.

Part Two: Statistical Analysis

1. Introduction

The aim of this study is to statistically analyse the social and economic growth of 10 countries over a period of ten years (2011 to 2020). The dataset used for the interactive dashboard was also used for this study.

The countries were divided into two groups following the size of population. The countries in the first group are Canada, France, Germany, Italy, and United Kingdom and second group countries are Brazil, China, Japan, Nigeria, and United States of America. The indicators cover both the economic and social sector. They are Population, Urban Population, Rural Population, Gross Domestic Product (GDP), Per Capita Income (GDP), Final Consumption Expenditure (% of the GDP), Trade (Domestic and Foreign trade), Life Expectancy, Gross National Income (GNI), GNI Per Capita, Labor Force, and Unemployment (% of the Labor Force).

We want to understand the progression of indicators for each country to be able to see how these countries have fared over the years. We will be carrying out a statistical descriptive analysis to get an understanding of this. We also want to see how the indicators are related. If they are strongly correlated or not, positively, or not. We also examine the variability between variables. Compare the living situations in the countries that are densely populated to those that aren't to proffer solutions and discuss the causes of decline or growth over the selected period and to see the if the percentage of life expectancy in the densely populated countries is different to that in the other countries.

2. Background Research

"The basic objective of development is to create an enabling environment for people to enjoy long, healthy, and creative lives. But it is often forgotten in the immediate concern with the accumulation of commodities and financial wealth" (World Bank, 2001). GDP per capita is not enough to measure the social growth of a country which is what is widely believed. Although, in recent times there has been lots of critics and debates over this topic. GDP per capita is not enough to measure the social aspect of the development of the citizens of a country, which is where the Human Development Index comes in. This cuts across the social growth (GNI), how long a citizen is expected to live for (Life Expectancy), and the rate of literacy (Adult Literacy Rate). For economic development indicators, the macroeconomy indicators includes GDP, Final Consumption Expenditure and Trade. Most world bank economic indicators are used for tracking Sustainable Development Goals (SDG) goal 8 which aims to promote full and productive employment, inclusive and sustainable economic growth, and decent work for all (The World Bank, s.d.).

We see an example of a research study that investigates the existence of a bidirectional causal relationship between the economic and non-economic factors that influence country performance using a simultaneous equation model (SEM), as well as the strength of this reciprocal causality. The conclusion of the research showed that GDP has a bidirectional causal link with life expectancy, but only a unidirectional one with literacy rate (Maria Francesca Cracolici et al, 2009).

For this study, we selected 10 countries that are considered the countries with the best economies in the world and in their continents. They are the G7 countries which are Canada, France, Germany, Italy, Japan, United Kingdom, and the United States of America, Brazil, the country with the largest GDP in South America, China, the country with the largest GDP in Asia and Nigeria, the country with the largest GDP in Nigeria. We got the median of the total population of our dataset, and we used this to divide the countries into two groups. The countries with larger population rates are Brazil, China, Japan, Nigeria and USA, and the countries with lower population rates are Canada, France, Germany, Italy, and United Kingdom.

It is quite interesting to study these two groups of countries to see if there is an effect on the social and economic growth based on the population of people living in the country. Will a country be more productive in terms of Labor force because of a larger population rate, or would it be the opposite? Is there decent work for all considering the unemployment rates in this country? Has life expectancy increased over the years to signify increasingly good conditions of living? These and more are the questions we would be answering using statistical methods.

Using the research and statistical analysis methods in some of the papers reviewed, we get an overview of the analysis to do. Our analysis will entail a general and comprehensive descriptive analysis, correlation analysis using the Pearson’s correlation, analysis of variances using ANOVA, simple and multiple linear regression, one sample and two sample t-tests, using ARMA and MA models for time series analysis.

To understand the growth for each country, we perform a descriptive analysis to get a general overview of the factors (variables) chosen for this study. Statistics such as the mean, medium, standard deviation etc, will be used for the key indicators for the selected countries.

Pearson's correlation coefficient (Pearson's r) is widely used for measuring linear relationship. It measures the statistical association between two continuous variables. It ranges between -1 and 1. -1 indicates an absolute negative relation, 0 indications no relationship, 1 indicates an absolute positive relationship. This was used to test for positive or negative linear relationship between the key indicators in our study. For example, what would the linear relationship between Final consumption expenditure and Trade look like. One way ANOVA was used to evaluate the difference among the means of groups. What is GDP, GNI, Life expectancy like in our country groups of large and smaller population sizes? Are the population mean significantly different?

With the help of the statistical model called simple linear regression, one may predict the linear connection between a response variable (dependent variable) and an explanatory variable (independent variable). For example, we can use the population size as an explanatory value to the life expectancy indicator. Coefficient of determination (R^2) is used to measure how much of the variation of the dependent variable is explained by the variation of the independent variable. It takes values between 0 and 1. For values greater than 0.5, we can say that the independent variable is a good explanatory variable for the dependent variable. Multiple linear regression is same as SLR except that in this case, we have two or more explanatory variables

The objective of a time series analysis is to provide a model of a numeric data that is ordered over time which can then help with predicting future values of that variable. The time can be in intervals of annual, quarter, daily or hourly type. Over the 10 years of our stipulated study time, we can create a model for the GDP or GNI indicator which can then predict the future values for the selected countries.

3. Exploration of Data Set

The dataset used for this study was downloaded from the World Bank website. It contains 15 variables and 101 rows. The row contains the header and the 10 selected countries over a period of 10 years. The time frame for this dataset is ten years, 2011 to 2020.

The missing values were inputted using the mean of the values for the available years per country and indicator. The variables are defined in the table below.

Variable Name	Description	Source
GDP per capita	GDP divided by midyear population	World Bank national accounts data, and OECD National Accounts data files.

GNI per capita	GNI divided by the midyear population	World Bank national accounts data, and OECD National Accounts data files
GDP (current US$)	It measures the monetary value of final goods and services. It is the average productivity of each citizen.	World Bank national accounts data, and OECD National Accounts data files.
GNI (current US$)	It is total amount earned by the citizens of a country through jobs, businesses, etc.	World Bank national accounts data, and OECD National Accounts data files.
Population, total	It measures the total number of citizens of a country	World Bank national accounts data, and OECD National Accounts data files.
Rural population	Rural population refers to the total number of people living in rural areas.	World Bank national accounts data, and OECD National Accounts data files.
Urban population	Urban population refers to the total number of people living in rural areas.	World Bank national accounts data, and OECD National Accounts data files.
Final consumption expenditure (% of GDP)	Consists of the total expenditure spent by a household or group on goods and services used to satisfy needs and basic wants and the total government expenditure	World Bank national accounts data, and OECD National Accounts data files.
Trade (% of GDP)	It is the transfer of goods and services between two or more people	World Bank national accounts data, and OECD National Accounts data files.
Life expectancy at birth, total (years)	It is defined as the number of years, on average, that an individual is expected to live for.	United Nations Population Division. World Population Prospects

Labor force, total	It is the number of people, age 15 and older, that supplies labour to produce goods and services.	Derived using data from International Labour Organization, ILOSTAT database
Unemployment, total (% of total Labor force)	Refers to the percentage of people in the Labour Force currently unemployed but seeking for jobs.	International Labour Organization, ILOSTAT database
Rural population %	Percentage of people living in rural areas	Derived
Urban population %	Percentage of people living in urban areas	Derived
Group	Defined by the size of the population. Greater or less than 1.047e+08	Derived

4. Analysis

We will be implementing different statistical analysis on this dataset to answer the questions raised in the introduction and research background of this report. The aim is to fully understand each country's overview, see how they've increased or declined over the years in the social and economic sectors, compare the countries in group A and B to see if there's any effect of the population size on some of the indicators, see the relationship between indicators, create time series models that can predict future values, and conduct tests of different hypothesis. Like we earlier mentioned, we divided our countries into two groups. The reason for this was to be able to perform one way ANOVA on our dataset. R will be used for all parts of this analysis.

4.1 Descriptive Statistical Analysis

We used the function (stat.desc) to get the comprehensive statistics for each indicator and for every country. This includes the measures of central dependency (mean, mode, median), disparity (range), skewness, kurtosis, and other values.

```
CountryName: Brazil
           CountryName          Year  GNIPerCapita PopulationTotal RuralPopulation RuralPopulation_. UrbanPopulation UrbanPopulation_.           GDP           GNI
nbr.val             NA  1.000000e+01  1.000000e+01    1.000000e+01    1.000000e+01       10.00000000    1.000000e+01       10.000000000  1.000000e+01  1.000000e+01
nbr.null            NA  0.000000e+00  0.000000e+00    0.000000e+00    0.000000e+00        0.00000000    0.000000e+00        0.000000000  0.000000e+00  0.000000e+00
nbr.na              NA  0.000000e+00  0.000000e+00    0.000000e+00    0.000000e+00        0.00000000    0.000000e+00        0.000000000  0.000000e+00  0.000000e+00
min                 NA  2.011000e+03  8.670000e+03    1.975145e+08    2.747756e+07       12.92700000    1.671585e+08       84.631000000  1.450000e+12  1.410000e+12
max                 NA  2.020000e+03  1.281000e+04    2.125594e+08    3.035601e+07       15.36900000    1.850819e+08       87.073000000  2.620000e+12  2.550000e+12
range               NA  9.000000e+00  4.140000e+03    1.504487e+07    2.878455e+06        2.44200000    1.792332e+07        2.442000000  1.170000e+12  1.140000e+12
sum                 NA  2.015500e+04  1.065620e+05    2.052148e+09    2.893011e+08      141.15800000    1.762847e+09      858.842000000  2.092000e+13  2.042000e+13
median              NA  2.015500e+03  1.065600e+04    2.053174e+08    2.893628e+07       14.09400000    1.763811e+08       85.906000000  1.990000e+12  1.940000e+12
mean                NA  2.015500e+03  1.065620e+04    2.052148e+08    2.893011e+07       14.11580000    1.762847e+08       85.884200000  2.092000e+12  2.042000e+12
SE.mean             NA  9.574271e-01  4.627909e+02    1.606580e+06    3.063351e+05        0.26008438    1.912803e+06        0.260084379  1.230429e+11  1.214340e+11
CI.mean             NA  2.165851e+00  1.046906e+03    3.634337e+06    6.929782e+05        0.58835174    4.327061e+06        0.588351742  2.783424e+11  2.747028e+11
var                 NA  9.166667e+00  2.141754e+06    2.581099e+13    9.384120e+11        0.67643884    3.658815e+13        0.676438844  1.513956e+23  1.474622e+23
std.dev             NA  3.027650e+00  1.463473e+03    5.080452e+06    9.687167e+05        0.82245902    6.048814e+06        0.822459023  3.890958e+11  3.840081e+11
coef.var            NA  1.502183e-03  1.373354e-01    2.475675e-02    3.348472e-02        0.05826514    3.431275e-02        0.009576372  1.859923e-01  1.880549e-01
skewness            NA  0.000000e+00  4.581634e-02   -4.752841e-02   -1.778304e-02        0.05644269   -3.713809e-02       -0.056442693 -4.313796e-02 -4.746709e-02
skew.2SE            NA  0.000000e+00  3.334314e-02   -3.458911e-02   -1.294173e-02        0.04107654   -2.702749e-02       -0.041076541 -3.139393e-02 -3.454449e-02
kurtosis            NA -1.561636e+00 -1.602367e+00   -1.568613e+00   -1.562811e+00       -1.56148164   -1.568271e+00       -1.561481637 -1.583737e+00 -1.606057e+00
kurt.2SE            NA -5.852118e-01 -6.004753e-01   -5.878261e-01   -5.856522e-01       -0.58515386   -5.876980e-01       -0.585153860 -5.934940e-01 -6.018581e-01
normtest.W          NA  9.701646e-01  9.374915e-01    9.684423e-01    9.699364e-01        0.96923178    9.687950e-01        0.969231777  9.096856e-01  9.083820e-01
normtest.p          NA  8.923673e-01  5.254533e-01    8.760383e-01    8.902551e-01        0.88363241    8.794535e-01        0.883632414  2.788382e-01  2.700609e-01
           LifeExpectancy  GDPPerCapita FinalConsumptionExpenditure        Trade    LaborForce Unemployment group
nbr.val       10.000000000  1.000000e+01                 10.00000000  10.00000000  1.000000e+01  10.00000000    NA
nbr.null       0.000000000  0.000000e+00                  0.00000000   0.00000000  0.000000e+00   0.00000000    NA
nbr.na         0.000000000  0.000000e+00                  0.00000000   0.00000000  0.000000e+00   0.00000000    NA
min           73.921000000  6.814876e+03                 78.94310870  23.93440518  9.379402e+07   6.65999985    NA
max           76.084000000  1.324539e+04                 85.12762867  32.89045210  1.043770e+08  13.68999958    NA
range          2.163000000  6.430512e+03                  6.18451997   8.95604692  1.058294e+07   7.02999973    NA
sum          750.675000000  1.023239e+05                827.64300718 265.98399767  9.854198e+08  98.55000018    NA
median        75.112000000  9.540029e+03                 83.55565520  25.45011547  9.795736e+07  10.01500034    NA
mean          75.067500000  1.023239e+04                 82.76430072  26.59839977  9.854198e+07   9.85500002    NA
SE.mean        0.229500581  6.714029e+02                  0.70445773   0.90204454  1.097619e+06   0.90182190    NA
CI.mean        0.519166383  1.518819e+03                  1.59359409   2.04056653  2.482986e+06   2.04006287    NA
var            0.526705167  4.507818e+06                  4.96260689   8.13684359  1.204767e+13   8.13282738    NA
std.dev        0.725744560  2.123162e+03                  2.22769093   2.85251531  3.470975e+06   2.85181125    NA
coef.var       0.009667893  2.074943e-01                  0.02691608   0.10724387  3.522331e-02   0.28937709    NA
skewness      -0.131035205  2.647844e-02                 -0.50313675   0.95929787  2.638854e-01   0.03619451    NA
skew.2SE      -0.095361732  1.926986e-02                 -0.36616108   0.69813533  1.920443e-01   0.02634079    NA
kurtosis      -1.536075683 -1.599049e+00                 -1.51805300  -0.37786938 -1.451232e+00  -2.00183347    NA
```

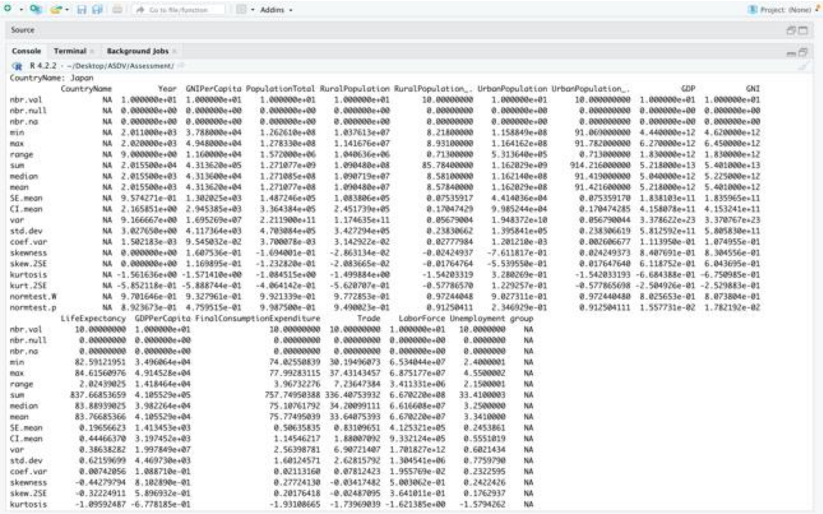
```
CountryName: Canada
           CountryName          Year  GNIPerCapita PopulationTotal RuralPopulation RuralPopulation_. UrbanPopulation UrbanPopulation_.           GDP           GNI
nbr.val             NA  1.000000e+01  1.000000e+01    1.000000e+01    1.000000e+01       10.00000000    1.000000e+01       10.00000000  1.000000e+01  1.000000e+01
nbr.null            NA  0.000000e+00  0.000000e+00    0.000000e+00    0.000000e+00        0.00000000    0.000000e+00        0.00000000  0.000000e+00  0.000000e+00
nbr.na              NA  0.000000e+00  0.000000e+00    0.000000e+00    0.000000e+00        0.00000000    0.000000e+00        0.00000000  0.000000e+00  0.000000e+00
min                 NA  2.011000e+03  4.296000e+04    3.433933e+07    6.491507e+06       18.43800000    2.784782e+07       81.09600000  1.530000e+12  1.510000e+12
max                 NA  2.020000e+03  5.281000e+04    3.803720e+07    7.013300e+06       18.90400000    3.102390e+07       81.56200000  1.850000e+12  1.820000e+12
range               NA  9.000000e+00  9.850000e+03    3.697876e+06    5.217930e+05        0.46600000    3.176083e+06        0.46600000  3.200000e+11  3.100000e+11
sum                 NA  2.015500e+04  4.783000e+05    3.606351e+08    6.742439e+07      187.00700000    2.932107e+08      812.99300000  1.714000e+13  1.688000e+13
median              NA  2.015500e+03  4.771000e+04    3.590620e+07    6.721778e+06       18.72050000    2.918442e+07       81.27950000  1.735000e+12  1.710000e+12
mean                NA  2.015500e+03  4.783000e+04    3.606351e+07    6.742439e+06       18.70070000    2.932107e+07       81.29930000  1.714000e+12  1.688000e+12
SE.mean             NA  9.574271e-01  1.067246e+03    3.912125e+05    5.572416e+04        0.048080731   3.355123e+05        0.048080731  3.553402e+10  3.466667e+10
CI.mean             NA  2.165851e+00  2.414277e+03    8.849824e+05    1.260568e+05        0.108766169   7.589815e+05        0.108766169  8.038354e+10  7.842145e+10
var                 NA  9.166667e+00  1.139013e+07    1.530472e+12    3.105182e+10        0.023117567   1.125685e+12        0.023117567  1.262667e+22  1.201778e+22
std.dev             NA  3.027650e+00  3.374927e+03    1.237123e+06    1.762153e+05        0.152044621   1.060983e+06        0.152044621  1.123684e+11  1.096256e+11
coef.var            NA  1.502183e-03  7.056089e-02    3.430400e-02    2.613524e-02        0.008130424   3.618500e-02        0.001870184  6.555918e-02  6.494409e-02
skewness            NA  0.000000e+00  1.019000e-01    1.914018e-01    1.388441e-01       -0.304028547   2.003692e-01        0.304028547 -3.512957e-01 -3.858162e-01
skew.2SE            NA  0.000000e+00  7.415842e-02    1.392939e-01    1.010447e-01       -0.221258774   1.458200e-01        0.221258774 -2.556578e-01 -2.807803e-01
kurtosis            NA -1.561636e+00 -1.486263e+00   -1.502647e+00   -1.539757e+00       -1.371335473  -1.495603e+00       -1.371335473 -1.504562e+00 -1.469649e+00
kurt.2SE            NA -5.852118e-01 -5.569661e-01   -5.631059e-01   -5.770126e-01       -0.513897972  -5.604665e-01       -0.513897972 -5.638237e-01 -5.507401e-01
normtest.W          NA  9.701646e-01  9.347982e-01    9.663870e-01    9.652434e-01        0.967416231   9.664392e-01        0.967416231  9.254595e-01  9.208329e-01
normtest.p          NA  8.923673e-01  4.967072e-01    8.554756e-01    8.435854e-01        0.865910136   8.560117e-01        0.865910136  4.047436e-01  3.639541e-01
           LifeExpectancy  GDPPerCapita FinalConsumptionExpenditure        Trade    LaborForce Unemployment group
nbr.val       10.000000000  1.000000e+01                  10.0000000   10.0000000  1.000000e+01   10.0000000    NA
nbr.null       0.000000000  0.000000e+00                   0.0000000    0.0000000  0.000000e+00    0.0000000    NA
nbr.na         0.000000000  0.000000e+00                   0.0000000    0.0000000  0.000000e+00    0.0000000    NA
min           81.448780490  4.231560e+04                  76.06287442  60.78253044  1.914740e+07   5.6599998    NA
max           82.048780490  5.266909e+04                  79.80694809  66.44771684  2.079260e+07   9.4600000    NA
range          0.600000000  1.035349e+04                   3.74407367   5.66518640  1.645204e+06   3.8000002    NA
sum          818.192682940  4.756607e+05                 780.64887621 640.97628157  1.992664e+08  69.9800000    NA
median        81.850000000  4.643866e+04                  78.54418427  64.73986150  1.985152e+07   6.9549999    NA
mean          81.819268294  4.756607e+04                  78.06488762  64.09762816  1.992664e+07   6.9980000    NA
SE.mean        0.057883039  1.313647e+03                   0.39984107   0.61010384  1.710623e+05   0.3336159    NA
CI.mean        0.130940530  2.971677e+03                   0.90450333   1.38015077  3.869699e+05   0.7546916    NA
var            0.033504462  1.725669e+07                   1.59872879   3.72226697  2.926233e+11   1.1129957    NA
std.dev        0.183042240  4.154118e+03                   1.26440847   1.92931775  5.409466e+05   1.0549861    NA
coef.var       0.002237153  8.733363e-02                   0.01619689   0.03009967  2.714691e-02   0.1507554    NA
skewness      -0.501547830  1.373658e-01                  -0.17972139  -0.32394098  1.239533e-01   0.9204391    NA
skew.2SE      -0.365004730  9.996888e-02                  -0.13079342  -0.23575017  9.020781e-02   0.6698556    NA
kurtosis      -0.736230883 -1.880952e+00                  -1.58501395  -1.55733369 -1.522169e+00   0.3962031    NA
```

```
CountryName: China
           CountryName          Year  GNIPerCapita PopulationTotal RuralPopulation RuralPopulation_. UrbanPopulation UrbanPopulation_.           GDP           GNI
nbr.val             NA  1.000000e+01  1.000000e+01    1.000000e+01    1.000000e+01       10.00000000    1.000000e+01       10.00000000  1.000000e+01  1.000000e+01
nbr.null            NA  0.000000e+00  0.000000e+00    0.000000e+00    0.000000e+00        0.00000000    0.000000e+00        0.00000000  0.000000e+00  0.000000e+00
nbr.na              NA  0.000000e+00  0.000000e+00    0.000000e+00    0.000000e+00        0.00000000    0.000000e+00        0.00000000  0.000000e+00  0.000000e+00
min                 NA  2.011000e+03  5.050000e+03    1.345035e+09    5.442895e+08       38.57200000    6.793906e+08       50.51100000  1.691544e+09  1.706709e+09
max                 NA  2.020000e+03  9.460000e+03    1.411100e+09    6.656444e+08       49.48900000    8.668105e+08       61.42800000  2.510127e+09  2.521348e+09
range               NA  9.000000e+00  4.410000e+03    6.606500e+07    1.213549e+08       10.91700000    1.874199e+08       10.91700000  8.185825e+08  8.146391e+08
sum                 NA  2.015500e+04  7.435000e+04    1.381980e+10    6.064357e+09      439.36800000    7.755438e+09      560.63200000  2.089666e+10  2.156244e+10
median              NA  2.015500e+03  7.467500e+03    1.383825e+09    6.072256e+08       43.88200000    7.765994e+08       56.11800000  2.146664e+09  2.281980e+09
mean                NA  2.015500e+03  7.435000e+03    1.381980e+09    6.064357e+08       43.93680000    7.755438e+08       56.06320000  2.089666e+09  2.156244e+09
SE.mean             NA  9.574271e-01  4.055559e+02    7.284171e+06    1.292078e+07        1.16699091    2.016629e+07        1.16699091  9.491486e+07  1.034775e+08
CI.mean             NA  2.165851e+00  9.174311e+02    1.647794e+07    2.922882e+07        2.63991686    4.561931e+07        2.63991686  2.147123e+08  2.340824e+08
var                 NA  9.166667e+00  1.644756e+06    5.305914e+14    1.669464e+15       13.61867796    4.066791e+15       13.61867796  9.008831e+16  1.070760e+17
std.dev             NA  3.027650e+00  1.282480e+03    2.303457e+07    4.085908e+07        3.69034930    6.377140e+07        3.69034930  3.001471e+08  3.272247e+08
coef.var            NA  1.502183e-03  1.724923e-01    1.666781e-02    6.737578e-02        0.08399222    8.222798e-02        0.06582481  1.436340e-01  1.517568e-01
skewness            NA  0.000000e+00 -3.128451e-01   -2.099415e-01   -4.995859e-02        0.03483462   -4.878489e-02       -0.03483462 -5.796172e-02 -2.687552e-01
skew.2SE            NA  0.000000e+00 -2.276751e-01   -1.527863e-01   -3.635769e-02        0.02535112   -3.550352e-02       -0.02535112 -4.218202e-02 -1.955883e-01
kurtosis            NA -1.561636e+00 -8.897159e-01   -1.585696e+00   -1.561231e+00       -1.57331603   -1.587938e+00       -1.57331603 -1.703070e+00 -1.872027e+00
kurt.2SE            NA -5.852118e-01 -3.334145e-01   -5.942279e-01   -5.850599e-01       -0.58958871   -5.950683e-01       -0.58958871 -6.382131e-01 -7.015283e-01
normtest.W          NA  9.701646e-01  9.726562e-01    9.484781e-01    9.694440e-01        0.96818664    9.655504e-01        0.96818664  9.278881e-01  8.512038e-01
normtest.p          NA  8.923673e-01  9.143206e-01    6.505294e-01    8.856425e-01        0.87354169    8.468070e-01        0.87354169  4.274441e-01  6.003598e-02
           LifeExpectancy  GDPPerCapita FinalConsumptionExpenditure        Trade    LaborForce Unemployment group
nbr.val        10.00000000  1.000000e+01                  10.0000000   10.0000000  1.000000e+01  10.00000000    NA
nbr.null        0.00000000  0.000000e+00                   0.0000000    0.0000000  0.000000e+00   0.00000000    NA
nbr.na          0.00000000  0.000000e+00                   0.0000000    0.0000000  0.000000e+00   0.00000000    NA
min            47.95000000  5.614352e+03                  50.1592933   34.5892495  7.789777e+08   4.28000021    NA
max            53.67900000  1.040867e+04                  56.0185730   50.7409046  8.000210e+08   5.00000000    NA
range           5.72900000  4.794318e+03                   5.8592798   16.1516551  2.104324e+07   0.71999979    NA
sum           510.56900000  8.195705e+04                 535.8479368  412.6941457  7.923933e+09  45.72000122    NA
median         51.23700000  8.055397e+03                  54.4292494   38.5482913  7.938743e+08   4.56000018    NA
mean           51.05690000  8.195705e+03                  53.5847937   41.2694146  7.923933e+08   4.57200012    NA
SE.mean         0.63080209  5.163232e+02                   0.6444132    1.8402377  2.345253e+06   0.05734495    NA
CI.mean         1.42697347  1.168004e+03                   1.4577640    4.1629070  5.305331e+06   0.12972328    NA
var             3.97911277  2.665897e+06                   4.1526840   33.8647497  5.500212e+13   0.03288443    NA
std.dev         1.99477136  1.632757e+03                   2.0378135    5.8193427  7.416342e+06   0.18134065    NA
coef.var        0.03906957  1.992211e-01                   0.0380297    0.1410086  9.359420e-03   0.03966331    NA
skewness       -0.17043502 -5.386280e-02                  -0.4051274    0.3868053 -5.359155e-01   0.85919292    NA
skew.2SE       -0.12403521 -3.919900e-02                  -0.2948341    0.2815001 -3.900160e-01   0.62528330    NA
kurtosis       -1.61405475 -1.503100e+00                  -1.5910493   -1.6843453 -1.328450e+00   0.80567643    NA
```

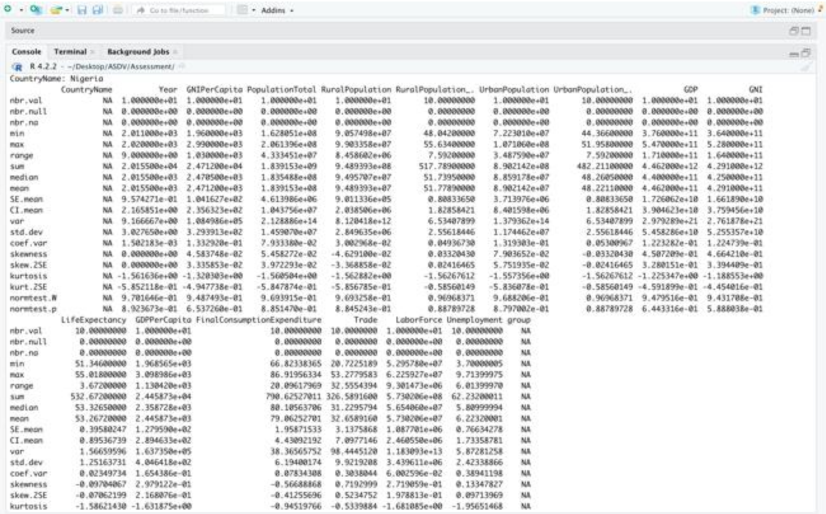
```
R 4.2.2 · ~/Desktop/ASDV/Assessment/
CountryName: France
           CountryName          Year   GNIPerCapita PopulationTotal RuralPopulation RuralPopulation_. UrbanPopulation UrbanPopulation_.           GDP           GNI
nbr.val             NA  1.000000e+01  1.000000e+01    1.000000e+01    1.000000e+01       10.00000000    1.000000e+01       10.00000000  1.000000e+01  1.000000e+01
nbr.null            NA  0.000000e+00  0.000000e+00    0.000000e+00    0.000000e+00        0.00000000    0.000000e+00        0.00000000  0.000000e+00  0.000000e+00
nbr.na              NA  0.000000e+00  0.000000e+00    0.000000e+00    0.000000e+00        0.00000000    0.000000e+00        0.00000000  0.000000e+00  0.000000e+00
min                 NA  2.011000e+03  3.833000e+04    6.534279e+07    1.281903e+07       19.02500000    5.137381e+07       78.62200000  2.440000e+12  2.490000e+12
max                 NA  2.020000e+03  4.435000e+04    6.737991e+07    1.396898e+07       21.37800000    5.456088e+07       80.97500000  2.870000e+12  2.940000e+12
range               NA  9.000000e+00  6.020000e+03    2.037119e+06    1.149954e+06        2.35300000    3.187073e+06        2.35300000  4.300000e+11  4.500000e+11
sum                 NA  2.015500e+04  4.178860e+05    6.652345e+08    1.343896e+08      202.09100000    5.308449e+08      797.90900000  2.688000e+13  2.746000e+13
median              NA  2.015500e+03  4.178800e+04    6.663619e+07    1.346972e+07       20.21400000    5.316646e+07       79.78600000  2.705000e+12  2.765000e+12
mean                NA  2.015500e+03  4.178860e+04    6.652345e+07    1.343896e+07       20.20910000    5.308449e+07       79.79090000  2.688000e+12  2.746000e+12
SE.mean             NA  9.574271e-01  6.342317e+02    2.173304e+05    1.233974e+05        0.25043866    3.384927e+05        0.25043866  4.829999e+10  5.004887e+10
CI.mean             NA  2.165851e+00  1.434732e+03    4.916356e+05    2.791444e+05        0.56653160    7.657236e+05        0.56653160  1.092622e+11  1.132184e+11
var                 NA  9.166667e+00  4.022499e+06    4.723252e+11    1.522693e+11        0.62719521    1.145773e+12        0.62719521  2.332889e+22  2.504889e+22
std.dev             NA  3.027650e+00  2.005617e+03    6.872592e+05    3.902170e+05        0.79195657    1.070408e+06        0.79195657  1.527380e+11  1.582684e+11
coef.var            NA  1.502183e-03  4.799435e-02    1.033108e-02    2.903625e-02        0.03918812    2.016423e-02        0.00992540  5.682216e-02  5.763598e-02
skewness            NA  0.000000e+00 -3.926629e-01   -3.606675e-01   -1.673895e-01       -0.01378586   -1.676305e-01        0.01378586 -3.486230e-01 -3.061228e-01
skew.2SE            NA  0.000000e+00 -2.857630e-01   -2.624781e-01   -1.218188e-01       -0.01003275   -1.219942e-01        0.01003275 -2.537127e-01 -2.227829e-01
kurtosis            NA -1.561636e+00 -1.296601e+00   -1.417605e+00   -1.561077e+00       -1.56292328   -1.518163e+00       -1.56292328 -1.453273e+00 -1.512668e+00
kurt.2SE            NA -5.852118e-01 -4.858918e-01   -5.312371e-01   -5.850024e-01       -0.58569411   -5.689205e-01       -0.58569411 -5.446033e-01 -5.668613e-01
normtest.W          NA  9.701646e-01  9.284435e-01    9.498646e-01    9.619416e-01        0.96994537    9.668736e-01        0.96994537  9.290505e-01  9.343752e-01
normtest.p          NA  8.923673e-01  4.327585e-01    6.668991e-01    8.077659e-01        0.89033824    8.604418e-01        0.89033824  4.386180e-01  4.922770e-01
           LifeExpectancy  GDPPerCapita FinalConsumptionExpenditure        Trade   LaborForce Unemployment group
nbr.val       10.000000000  1.000000e+01                 10.00000000  10.00000000  1.000000e+01  10.00000000    NA
nbr.null       0.000000000  0.000000e+00                  0.00000000   0.00000000  0.000000e+00   0.00000000    NA
nbr.na         0.000000000  0.000000e+00                  0.00000000   0.00000000  0.000000e+00   0.00000000    NA
min           81.968292680  3.665292e+04                 76.59272204  57.76742833  2.996538e+07   8.01000023    NA
max           82.826829270  4.384811e+04                 78.74738627  64.43795260  3.062036e+07  10.35000038    NA
range          0.858536590  7.195184e+03                  2.15466423   6.67052427  6.549810e+05   2.34000015    NA
sum          824.170731730  4.040991e+05                780.01194130 610.89602490  3.037124e+08  93.67000103    NA
median        82.447560975  4.072550e+04                 78.10197336  60.78946945  3.040168e+07   9.40499973    NA
mean          82.417073173  4.040991e+04                 78.00119413  61.08960249  3.037124e+07   9.36700010    NA
SE.mean        0.092794114  7.832081e+02                  0.22669024   0.70590319  5.953601e+04   0.25345194    NA
CI.mean        0.209914870  1.771740e+03                  0.51280896   1.59686396  1.346798e+05   0.57334813    NA
var            0.086107476  6.134149e+06                  0.51388467   4.98299315  3.544536e+10   0.64237887    NA
std.dev        0.293440754  2.476721e+03                  0.71685749   2.23226189  1.882694e+05   0.80148542    NA
coef.var       0.003560436  6.128994e-02                  0.00919034   0.03654078  6.198936e-03   0.08556479    NA
skewness      -0.075864394 -1.689901e-01                 -0.61661162   0.16875773 -7.173500e-01  -0.26529120    NA
skew.2SE      -0.055210811 -1.229837e-01                 -0.44874316   0.12281454 -5.220562e-01  -0.19306741    NA
kurtosis      -1.702742840 -1.537518e+00                 -0.99441172  -1.46313621 -3.772450e-01  -1.47059484    NA
```

```
R 4.2.2 · ~/Desktop/ASDV/Assessment/
CountryName: Germany
           CountryName          Year   GNIPerCapita PopulationTotal RuralPopulation RuralPopulation_. UrbanPopulation UrbanPopulation_.           GDP           GNI
nbr.val             NA  1.000000e+01  1.000000e+01    1.000000e+01    1.000000e+01      10.000000000    1.000000e+01      10.000000000  1.000000e+01  1.000000e+01
nbr.null            NA  0.000000e+00  0.000000e+00    0.000000e+00    0.000000e+00       0.000000000    0.000000e+00       0.000000000  0.000000e+00  0.000000e+00
nbr.na              NA  0.000000e+00  0.000000e+00    0.000000e+00    0.000000e+00       0.000000000    0.000000e+00       0.000000000  0.000000e+00  0.000000e+00
min                 NA  2.011000e+03  4.364000e+04    8.027498e+07    1.833481e+07      22.547000000    6.194018e+07      77.160000000  3.360000e+12  3.430000e+12
max                 NA  2.020000e+03  4.762000e+04    8.316087e+07    1.880966e+07      22.840000000    6.441059e+07      77.453000000  3.980000e+12  4.110000e+12
range               NA  9.000000e+00  3.980000e+03    2.885888e+06    4.748580e+05       0.293000000    2.470412e+06       0.293000000  6.200000e+11  6.800000e+11
sum                 NA  2.015500e+04  4.617860e+05    8.181808e+08    1.861060e+08     227.474000000    6.320748e+08     772.526000000  3.714000e+13  3.809000e+13
median              NA  2.015500e+03  4.635400e+04    8.201764e+07    1.868741e+07      22.788000000    6.332750e+07      77.212000000  3.740000e+12  3.835000e+12
mean                NA  2.015500e+03  4.617860e+04    8.181808e+07    1.861060e+07      22.747400000    6.320748e+07      77.252600000  3.714000e+12  3.809000e+12
SE.mean             NA  9.574271e-01  4.117508e+02    3.652476e+05    6.239762e+04       0.031119554    3.045048e+05       0.031119554  6.429446e+10  6.869336e+10
CI.mean             NA  2.165851e+00  9.314451e+02    8.262476e+05    1.411532e+05       0.070397323    6.888377e+05       0.070397323  1.454442e+11  1.553952e+11
var                 NA  9.166667e+00  1.695388e+06    1.334058e+12    3.893463e+10       0.009684267    9.272317e+11       0.009684267  4.133778e+22  4.718778e+22
std.dev             NA  3.027650e+00  1.302071e+03    1.155014e+06    1.973186e+05       0.098408672    9.629287e+05       0.098408672  2.033169e+11  2.172275e+11
coef.var            NA  1.502183e-03  2.819640e-02    1.411686e-02    1.060248e-02       0.004326150    1.523441e-02       0.001273856  5.474339e-02  5.703005e-02
skewness            NA  0.000000e+00 -7.779152e-01   -1.330845e-01   -2.783683e-01      -0.837643464   -7.907722e-02       0.837643464 -3.907485e-01 -3.448123e-01
skew.2SE            NA  0.000000e+00 -5.661329e-01   -9.685311e-02   -2.025844e-01      -0.609600537   -5.754897e-02       0.609600537 -2.843698e-01 -2.509394e-01
kurtosis            NA -1.561636e+00 -8.600441e-01   -1.894661e+00   -1.899268e+00      -0.836851245   -1.870895e+00      -0.836851245 -1.381069e+00 -1.340288e+00
kurt.2SE            NA -5.852118e-01 -3.222953e-01   -7.100103e-01   -7.117368e-01      -0.313603904   -7.011041e-01      -0.313603904 -5.175457e-01 -5.022631e-01
normtest.W          NA  9.701646e-01  8.781948e-01    8.795708e-01    8.332054e-01       0.860372140    8.965516e-01       0.860372140  9.403663e-01  9.566635e-01
normtest.p          NA  8.923673e-01  1.244019e-01    1.290207e-01    3.656914e-02       0.077083620    2.007188e-01       0.077083620  5.570961e-01  7.472631e-01
           LifeExpectancy  GDPPerCapita FinalConsumptionExpenditure        Trade   LaborForce Unemployment group
nbr.val       10.000000000  1.000000e+01                 10.00000000  10.00000000  1.000000e+01  10.00000000    NA
nbr.null       0.000000000  0.000000e+00                  0.00000000   0.00000000  0.000000e+00   0.00000000    NA
nbr.na         0.000000000  0.000000e+00                  0.00000000   0.00000000  0.000000e+00   0.00000000    NA
min           80.436585370  4.110326e+04                 71.92766326  81.10854743  4.172073e+07   3.14000010    NA
max           81.292682930  4.802387e+04                 74.18777456  88.43440642  4.435116e+07   5.82000017    NA
range          0.856097560  6.920614e+03                  2.26011130   7.32585899  2.630437e+06   2.68000007    NA
sum          808.307317070  4.537977e+05                729.36929064 856.81483965  4.304742e+08  44.23000026    NA
median        80.917073170  4.627581e+04                 72.83654037  85.72617294  4.310289e+07   4.36999989    NA
mean          80.830731707  4.537977e+04                 72.93692906  85.68148396  4.304742e+07   4.42300003    NA
SE.mean        0.090720342  7.509122e+02                  0.26052657   0.65106088  3.126224e+05   0.28910226    NA
CI.mean        0.205223672  1.698682e+03                  0.58935204   1.47280204  7.072010e+05   0.65399474    NA
var            0.082301805  5.638692e+06                  0.67874092   4.23880273  9.773277e+11   0.83580115    NA
std.dev        0.286882911  2.374593e+03                  0.82385734   2.05883529  9.885989e+05   0.91422161    NA
coef.var       0.003549181  5.232713e-02                  0.01129548   0.02402894  2.296534e-02   0.20669717    NA
skewness      -0.026885511 -5.505582e-01                  0.29031007  -0.75013977 -6.876164e-02   0.06534872    NA
skew.2SE      -0.019566108 -4.006724e-01                  0.21127506  -0.54591915 -5.004174e-02   0.04755796    NA
kurtosis      -1.558025278 -1.259521e+00                 -1.43247984  -0.12336727 -1.855730e+00  -1.65191305    NA
```

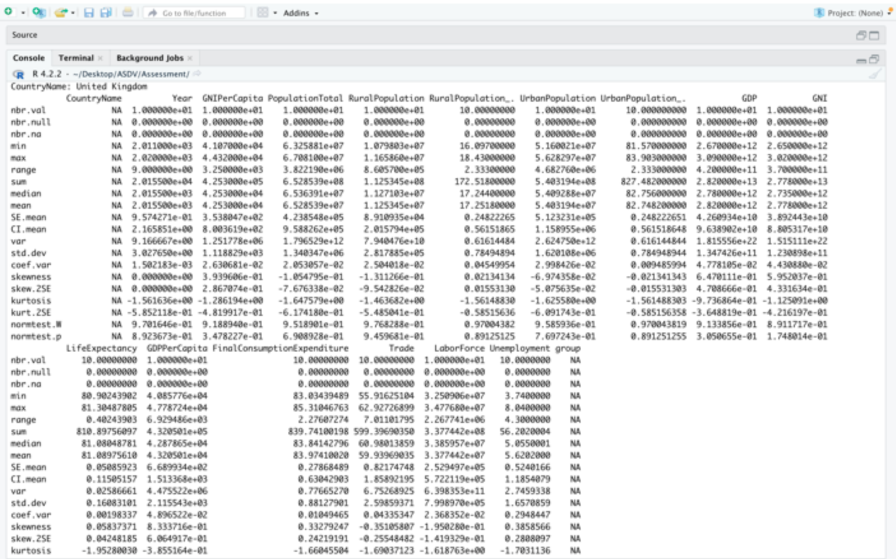
```
R 4.2.2 · ~/Desktop/ASDV/Assessment/
CountryName: Italy
           CountryName          Year   GNIPerCapita PopulationTotal RuralPopulation RuralPopulation_. UrbanPopulation UrbanPopulation_.           GDP           GNI
nbr.val             NA  1.000000e+01  1.000000e+01    1.000000e+01    1.000000e+01       10.00000000    1.000000e+01       10.00000000  1.000000e+01  1.000000e+01
nbr.null            NA  0.000000e+00  0.000000e+00    0.000000e+00    0.000000e+00        0.00000000    0.000000e+00        0.00000000  0.000000e+00  0.000000e+00
nbr.na              NA  0.000000e+00  0.000000e+00    0.000000e+00    0.000000e+00        0.00000000    0.000000e+00        0.00000000  0.000000e+00  0.000000e+00
min                 NA  2.011000e+03  3.134000e+04    5.937945e+07    1.721718e+07       28.96100000    4.064167e+07       68.44400000  1.840000e+12  1.820000e+12
max                 NA  2.020000e+03  3.790000e+04    6.078914e+07    1.873778e+07       31.55600000    4.255988e+07       71.03900000  2.290000e+12  2.290000e+12
range               NA  9.000000e+00  6.560000e+03    1.409691e+06    1.520601e+06        2.59500000    1.918209e+06        2.59500000  4.500000e+11  4.700000e+11
sum                 NA  2.015500e+04  3.431620e+05    6.014374e+08    1.821392e+08      302.84400000    4.192982e+08      697.15600000  2.035000e+13  2.041000e+13
median              NA  2.015500e+03  3.431600e+04    6.032785e+07    1.837976e+07       30.29000000    4.223979e+07       69.71000000  2.050000e+12  2.055000e+12
mean                NA  2.015500e+03  3.431620e+04    6.014374e+07    1.821392e+07       30.28440000    4.192982e+07       69.71560000  2.035000e+12  2.041000e+12
SE.mean             NA  9.574271e-01  6.225127e+02    1.775657e+05    1.716828e+05        0.27810131    2.124691e+05        0.27810131  4.558874e+10  4.573960e+10
CI.mean             NA  2.165851e+00  1.408222e+03    4.016815e+05    3.883735e+05        0.62910887    4.806386e+05        0.62910887  1.031289e+11  1.034702e+11
var                 NA  9.166667e+00  3.875221e+06    3.152958e+11    2.947499e+11        0.77340338    4.514313e+11        0.77340338  2.078333e+22  2.092111e+22
std.dev             NA  3.027650e+00  1.968558e+03    5.615120e+05    5.429088e+05        0.87943356    6.718864e+05        0.87943356  1.441643e+11  1.446413e+11
coef.var            NA  1.502183e-03  5.736527e-02    9.336168e-03    2.980735e-02        0.02903916    1.602407e-02        0.01261459  7.084239e-02  7.086787e-02
skewness            NA  0.000000e+00  1.627389e-01   -2.291233e-01   -6.149372e-01       -0.02855726   -8.995758e-01        0.02855726  1.802275e-01  5.115964e-02
skew.2SE            NA  0.000000e+00  1.184343e-01   -1.667460e-01   -4.475246e-01       -0.02078273   -6.546722e-01        0.02078273  1.311617e-01  3.723177e-02
kurtosis            NA -1.561636e+00 -1.044763e+00   -1.880495e+00   -1.266743e+00       -1.57250218   -9.270684e-01       -1.57250218 -1.354097e+00 -1.291243e+00
kurt.2SE            NA -5.852118e-01 -3.915174e-01   -7.047018e-01   -4.747029e-01       -0.58928373   -3.474121e-01       -0.58928373 -5.074381e-01 -4.838839e-01
normtest.W          NA  9.701646e-01  9.840190e-01    8.686955e-01    8.742177e-01        0.96764645    8.031088e-01        0.96764645  9.567664e-01  9.789715e-01
normtest.p          NA  8.923673e-01  9.830213e-01    9.652298e-02    1.119118e-01        0.86820723    1.581591e-02        0.86820723  7.484670e-01  9.594269e-01
           LifeExpectancy  GDPPerCapita FinalConsumptionExpenditure        Trade   LaborForce Unemployment group
nbr.val        10.00000000  1.000000e+01                 10.00000000  10.000000000  1.000000e+01   10.0000000    NA
nbr.null        0.00000000  0.000000e+00                  0.00000000   0.000000000  0.000000e+00    0.0000000    NA
nbr.na          0.00000000  0.000000e+00                  0.00000000   0.000000000  0.000000e+00    0.0000000    NA
min            82.18780488  3.024239e+04                 78.43418620  54.867579400  2.456474e+07    8.3599997    NA
max            83.49756098  3.864964e+04                 81.10672230  60.303551240  2.614214e+07   12.6800003    NA
range           1.30975610  8.407253e+03                  2.67253610   5.435971840  1.577403e+06    4.3200007    NA
sum           828.12926829  3.385674e+05                797.14284580 566.965555880  2.554670e+08  108.3599978    NA
median         82.81829268  3.414782e+04                 79.51487173  55.528681280  2.567821e+07   10.9299998    NA
mean           82.81292683  3.385674e+04                 79.71428458  56.696555588  2.554670e+07   10.8359998    NA
SE.mean         0.15133663  8.060974e+02                  0.30845495   0.658479441  1.616020e+05    0.4338259    NA
CI.mean         0.34234725  1.823519e+03                  0.69777357   1.489583984  3.655690e+05    0.9813824    NA
var             0.22902776  6.497929e+06                  0.95144456   4.335951744  2.611520e+11    1.8820491    NA
std.dev         0.47856845  2.549104e+03                  0.97542020   2.082294826  5.110303e+05    1.3718779    NA
coef.var        0.00577891  7.529087e-02                  0.01223645   0.036727008  2.000377e-02    0.1266037    NA
skewness        0.01699506  2.376567e-01                  0.16355679   0.753201955 -4.974736e-01   -0.3943531    NA
skew.2SE        0.01236827  1.729562e-01                  0.11902953   0.548147673 -3.620397e-01   -0.2869930    NA
kurtosis       -1.74854294 -1.083603e+00                 -1.73462874  -1.336202596 -1.161417e+00   -1.2363365    NA
```



```
CountryName: Japan
           CountryName          Year  GNIPerCapita PopulationTotal RuralPopulation RuralPopulation_. UrbanPopulation UrbanPopulation_.          GDP          GNI
nbr.val             NA  1.000000e+01  1.000000e+01    1.000000e+01    1.000000e+01        10.0000000    1.000000e+01       10.00000000 1.000000e+01 1.000000e+01
nbr.null            NA  0.000000e+00  0.000000e+00    0.000000e+00    0.000000e+00         0.0000000    0.000000e+00        0.00000000 0.000000e+00 0.000000e+00
nbr.na              NA  0.000000e+00  0.000000e+00    0.000000e+00    0.000000e+00         0.0000000    0.000000e+00        0.00000000 0.000000e+00 0.000000e+00
min                 NA  2.011000e+03  3.788000e+04    1.262610e+08    1.037613e+07         8.2180000    1.158849e+08       91.06900000 4.440000e+12 4.620000e+12
max                 NA  2.020000e+03  4.948000e+04    1.278330e+08    1.141676e+07         8.9310000    1.164162e+08       91.78200000 6.270000e+12 6.450000e+12
range               NA  9.000000e+00  1.160000e+04    1.572000e+06    1.040636e+06         0.7130000    5.313640e+05        0.71300000 1.830000e+12 1.830000e+12
sum                 NA  2.015500e+04  4.313620e+05    1.271077e+09    1.090480e+08        85.7840000    1.162029e+09      914.21600000 5.218000e+13 5.401000e+13
median              NA  2.015500e+03  4.313600e+04    1.271085e+08    1.090719e+07         8.5810000    1.162140e+08       91.41900000 5.040000e+12 5.225000e+12
mean                NA  2.015500e+03  4.313620e+04    1.271077e+08    1.090480e+07         8.5784000    1.162029e+08       91.42160000 5.218000e+12 5.401000e+12
SE.mean             NA  9.574271e-01  1.302025e+03    1.487246e+05    1.083806e+05        0.07535917    4.414036e+04       0.075359170 1.838103e+11 1.835965e+11
CI.mean             NA  2.165851e+00  2.945385e+03    3.364384e+05    2.451739e+05        0.17047429    9.985244e+04       0.170474285 4.158078e+11 4.153241e+11
var                 NA  9.166667e+00  1.695269e+07    2.211900e+11    1.174635e+11        0.05679004    1.948372e+10       0.056790044 3.378622e+23 3.370767e+23
std.dev             NA  3.027650e+00  4.117364e+03    4.703084e+05    3.427294e+05        0.23830662    1.395841e+05       0.238306619 5.812592e+11 5.805830e+11
coef.var            NA  1.502183e-03  9.545032e-02    3.700078e-03    3.142922e-02        0.02777984    1.201210e-03       0.002606677 1.113950e-01 1.074955e-01
skewness            NA  0.000000e+00  1.607536e-01   -1.694001e-01   -2.863134e-02       -0.02424937   -7.611817e-01       0.024249373 8.407691e-01 8.304556e-01
skew.2SE            NA  0.000000e+00  1.169895e-01   -1.232820e-01   -2.083665e-02       -0.01764764   -5.539550e-01       0.017647640 6.118752e-01 6.043695e-01
kurtosis            NA -1.561636e+00 -1.571410e+00   -1.084515e+00   -1.499884e+00       -1.54203319    3.280269e-01      -1.542033193 -6.684388e-01 -6.750985e-01
kurt.2SE            NA -5.852118e-01 -5.888744e-01   -4.064142e-01   -5.620707e-01       -0.57786570    1.229257e-01      -0.577865698 -2.504926e-01 -2.529883e-01
normtest.W          NA  9.701646e-01  9.327961e-01    9.921339e-01    9.772853e-01        0.97244048    9.027311e-01       0.972440480 8.025653e-01 8.073804e-01
normtest.p          NA  8.923673e-01  4.759515e-01    9.987500e-01    9.490023e-01        0.91250411    2.346929e-01       0.912504111 1.557731e-02 1.782192e-02
           LifeExpectancy  GDPPerCapita FinalConsumptionExpenditure        Trade   LaborForce Unemployment group
nbr.val       10.00000000  1.000000e+01                 10.00000000  10.00000000 1.000000e+01   10.0000000    NA
nbr.null       0.00000000  0.000000e+00                  0.00000000   0.00000000 0.000000e+00    0.0000000    NA
nbr.na         0.00000000  0.000000e+00                  0.00000000   0.00000000 0.000000e+00    0.0000000    NA
min           82.59121951  3.496064e+04                 74.02550839  30.19496073 6.534044e+07    2.4000001    NA
max           84.61560976  4.914528e+04                 77.99283115  37.43143457 6.875177e+07    4.5500002    NA
range          2.02439025  1.418464e+04                  3.96732276   7.23647384 3.411331e+06    2.1500001    NA
sum          837.66853659  4.105529e+05                757.74950388 336.40753932 6.670220e+08   33.4100003    NA
median        83.88939025  3.982264e+04                 75.10761792  34.20099111 6.616608e+07    3.2500000    NA
mean          83.76685366  4.105529e+04                 75.77495039  33.64075393 6.670220e+07    3.3410000    NA
SE.mean        0.19656623  1.413453e+03                  0.50635835   0.83109651 4.125321e+05    0.2453861    NA
CI.mean        0.44466370  3.197452e+03                  1.14546217   1.88007092 9.332124e+05    0.5551019    NA
var            0.38638282  1.997849e+07                  2.56398781   6.90721407 1.701827e+12    0.6021434    NA
std.dev        0.62159699  4.469730e+03                  1.60124571   2.62815792 1.304541e+06    0.7759790    NA
coef.var       0.00742056  1.088710e-01                  0.02113160   0.07812423 1.955769e-02    0.2322595    NA
skewness      -0.44279794  8.102890e-01                  0.27724130  -0.03417482 5.003062e-01    0.2422426    NA
skew.2SE      -0.32224911  5.896932e-01                  0.20176418  -0.02487095 3.641011e-01    0.1762937    NA
kurtosis      -1.09592487 -6.778185e-01                 -1.93108665  -1.73969039 -1.621385e+00   -1.5794262    NA
```



```
CountryName: Nigeria
           CountryName          Year  GNIPerCapita PopulationTotal RuralPopulation RuralPopulation_. UrbanPopulation UrbanPopulation_.          GDP          GNI
nbr.val             NA  1.000000e+01  1.000000e+01    1.000000e+01    1.000000e+01        10.0000000    1.000000e+01        10.0000000 1.000000e+01 1.000000e+01
nbr.null            NA  0.000000e+00  0.000000e+00    0.000000e+00    0.000000e+00         0.0000000    0.000000e+00         0.0000000 0.000000e+00 0.000000e+00
nbr.na              NA  0.000000e+00  0.000000e+00    0.000000e+00    0.000000e+00         0.0000000    0.000000e+00         0.0000000 0.000000e+00 0.000000e+00
min                 NA  2.011000e+03  1.960000e+03    1.628051e+08    9.057498e+07        48.0420000    7.223010e+07        44.3660000 3.760000e+11 3.640000e+11
max                 NA  2.020000e+03  2.990000e+03    2.061396e+08    9.903358e+07        55.6340000    1.071060e+08        51.9580000 5.470000e+11 5.280000e+11
range               NA  9.000000e+00  1.030000e+03    4.333451e+07    8.458602e+06         7.5920000    3.487590e+07         7.5920000 1.710000e+11 1.640000e+11
sum                 NA  2.015500e+04  2.471200e+04    1.839153e+09    9.489393e+08       517.7890000    8.902142e+08       482.2110000 4.462000e+12 4.291000e+12
median              NA  2.015500e+03  2.470500e+03    1.835488e+08    9.495707e+07        51.7395000    8.859178e+07        48.2605000 4.400000e+11 4.250000e+11
mean                NA  2.015500e+03  2.471200e+03    1.839153e+08    9.489393e+07        51.7789000    8.902142e+07        48.2211000 4.462000e+11 4.291000e+11
SE.mean             NA  9.574271e-01  1.041627e+02    4.613986e+06    9.011336e+05         0.8083365    3.713976e+06         0.8083365 1.726062e+10 1.661890e+10
CI.mean             NA  2.165851e+00  2.356323e+02    1.043756e+07    2.038506e+06         1.8285842    8.401598e+06         1.8285842 3.904623e+10 3.759456e+10
var                 NA  9.166667e+00  1.084986e+05    2.128886e+14    8.120418e+12         6.5340790    1.379362e+14         6.5340790 2.979289e+21 2.761878e+21
std.dev             NA  3.027650e+00  3.293913e+02    1.459070e+07    2.849635e+06         2.5561845    1.174462e+07         2.5561845 5.458286e+10 5.255357e+10
coef.var            NA  1.502183e-03  1.332920e-01    7.933380e-02    3.002968e-02        0.04936730    1.319303e-01        0.05300967 1.223282e-01 1.224739e-01
skewness            NA  0.000000e+00  4.583748e-02    5.458272e-02   -4.629100e-02        0.03320430    7.903652e-02       -0.03320430 4.507209e-01 4.664210e-01
skew.2SE            NA  0.000000e+00  3.335853e-02    3.972293e-02   -3.368858e-02        0.02416465    5.751935e-02       -0.02416465 3.280151e-01 3.394409e-01
kurtosis            NA -1.561636e+00 -1.320303e+00   -1.560504e+00   -1.562882e+00       -1.56267612   -1.557356e+00       -1.56267612 -1.225347e+00 -1.188553e+00
kurt.2SE            NA -5.852118e-01 -4.947738e-01   -5.847874e-01   -5.856785e-01       -0.58560149   -5.836078e-01       -0.58560149 -4.591899e-01 -4.454016e-01
normtest.W          NA  9.701646e-01  9.487493e-01    9.693915e-01    9.693258e-01        0.96968371    9.688206e-01        0.96968371 9.479516e-01 9.431708e-01
normtest.p          NA  8.923673e-01  6.537260e-01    8.851470e-01    8.845243e-01        0.88789728    8.797002e-01        0.88789728 6.443316e-01 5.888038e-01
           LifeExpectancy  GDPPerCapita FinalConsumptionExpenditure       Trade   LaborForce Unemployment group
nbr.val       10.00000000  1.000000e+01                 10.00000000  10.0000000 1.000000e+01  10.00000000    NA
nbr.null       0.00000000  0.000000e+00                  0.00000000   0.0000000 0.000000e+00   0.00000000    NA
nbr.na         0.00000000  0.000000e+00                  0.00000000   0.0000000 0.000000e+00   0.00000000    NA
min           51.34600000  1.968565e+03                 66.82338365  20.7225189 5.295780e+07   3.70000005    NA
max           55.01800000  3.098986e+03                 86.91956334  53.2779583 6.225927e+07   9.71399975    NA
range          3.67200000  1.130420e+03                 20.09617969  32.5554394 9.301473e+06   6.01399970    NA
sum          532.67200000  2.445873e+04                790.62527011 326.5891600 5.730206e+08  62.23200011    NA
median        53.32650000  2.358728e+03                 80.10563706  31.2295794 5.654060e+07   5.80999994    NA
mean          53.26720000  2.445873e+03                 79.06252701  32.6589160 5.730206e+07   6.22320001    NA
SE.mean        0.39580247  1.279590e+02                  1.95871533   3.1375868 1.087701e+06   0.76634278    NA
CI.mean        0.89536739  2.894633e+02                  4.43092192   7.0977146 2.460550e+06   1.73358781    NA
var            1.56659596  1.637350e+05                 38.36565752  98.4445120 1.183093e+13   5.87281258    NA
std.dev        1.25163731  4.046418e+02                  6.19400174   9.9219208 3.439611e+06   2.42338866    NA
coef.var       0.02349734  1.654386e-01                  0.07834308   0.3038044 6.002596e-02   0.38941198    NA
skewness      -0.09704067  2.979122e-01                 -0.56688868   0.7192999 2.719059e-01   0.13347827    NA
skew.2SE      -0.07062199  2.168076e-01                 -0.41255696   0.5234752 1.978813e-01   0.09713969    NA
kurtosis      -1.58621430 -1.631875e+00                 -0.94519766  -0.5339884 -1.681085e+00  -1.95651468    NA
```



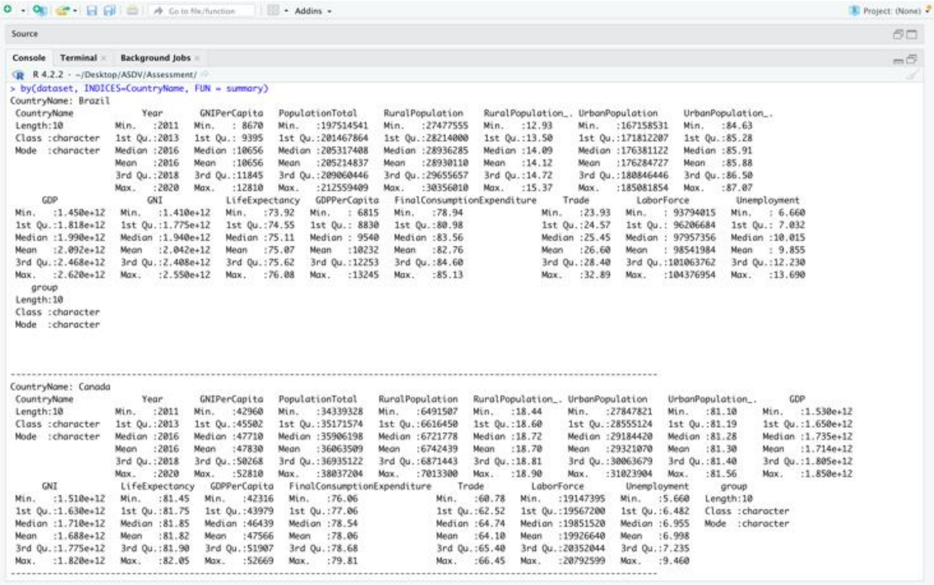
```
CountryName: United Kingdom
           CountryName          Year  GNIPerCapita PopulationTotal RuralPopulation RuralPopulation_. UrbanPopulation UrbanPopulation_.           GDP           GNI
nbr.val             NA  1.000000e+01  1.000000e+01    1.000000e+01    1.000000e+01       10.00000000    1.000000e+01       10.000000000  1.000000e+01  1.000000e+01
nbr.null            NA  0.000000e+00  0.000000e+00    0.000000e+00    0.000000e+00        0.00000000    0.000000e+00        0.000000000  0.000000e+00  0.000000e+00
nbr.na              NA  0.000000e+00  0.000000e+00    0.000000e+00    0.000000e+00        0.00000000    0.000000e+00        0.000000000  0.000000e+00  0.000000e+00
min                 NA  2.011000e+03  4.107000e+04    6.325881e+07    1.079803e+07       16.09700000    5.160021e+07       81.570000000  2.670000e+12  2.650000e+12
max                 NA  2.020000e+03  4.432000e+04    6.708100e+07    1.165860e+07       18.43000000    5.628297e+07       83.903000000  3.090000e+12  3.020000e+12
range               NA  9.000000e+00  3.250000e+03    3.822190e+06    8.605700e+05        2.33300000    4.682760e+06        2.333000000  4.200000e+11  3.700000e+11
sum                 NA  2.015500e+04  4.253000e+05    6.528539e+08    1.125345e+08      172.51800000    5.403194e+08      827.482000000  2.820000e+13  2.778000e+13
median              NA  2.015500e+03  4.253000e+04    6.536391e+07    1.127103e+07       17.24400000    5.409288e+07       82.756000000  2.780000e+12  2.735000e+12
mean                NA  2.015500e+03  4.253000e+04    6.528539e+07    1.125345e+07       17.25180000    5.403194e+07       82.748200000  2.820000e+12  2.778000e+12
SE.mean             NA  9.574271e-01  3.538047e+02    4.238548e+05    8.910935e+04        0.24822265    5.123231e+05        0.248222651  4.260934e+10  3.892443e+10
CI.mean             NA  2.165851e+00  8.003619e+02    9.588262e+05    2.015794e+05        0.56151865    1.158955e+06        0.561518648  9.638902e+10  8.805317e+10
var                 NA  9.166667e+00  1.251778e+06    1.796529e+12    7.940476e+10        0.61614484    2.624750e+12        0.616144844  1.815556e+22  1.515111e+22
std.dev             NA  3.027650e+00  1.118829e+03    1.340347e+06    2.817885e+05        0.78494894    1.620108e+06        0.784948944  1.347426e+11  1.230898e+11
coef.var            NA  1.502183e-03  2.630681e-02    2.053057e-02    2.504018e-02        0.04549954    2.998426e-02        0.009485994  4.778105e-02  4.430880e-02
skewness            NA  0.000000e+00  3.939606e-01   -1.054795e-01   -1.311266e-01        0.02134134   -6.974358e-02       -0.021341343  6.470111e-01  5.952037e-01
skew.2SE            NA  0.000000e+00  2.867074e-01   -7.676338e-02   -9.542826e-02        0.01553130   -5.075635e-02       -0.015531303  4.708666e-01  4.331634e-01
kurtosis            NA -1.561636e+00 -1.286194e+00   -1.647579e+00   -1.463682e+00       -1.56148830   -1.625580e+00       -1.561488303 -9.736864e-01 -1.125091e+00
kurt.2SE            NA -5.852118e-01 -4.819917e-01   -6.174180e-01   -5.485041e-01       -0.58515636   -6.091743e-01       -0.585156358 -3.648819e-01 -4.216197e-01
normtest.W          NA  9.701646e-01  9.188940e-01    9.518901e-01    9.768288e-01        0.97004382    9.585936e-01        0.970043819  9.133856e-01  8.911717e-01
normtest.p          NA  8.923673e-01  3.478227e-01    6.908928e-01    9.459681e-01        0.89125125    7.697243e-01        0.891251255  3.050655e-01  1.748014e-01
           LifeExpectancy  GDPPerCapita FinalConsumptionExpenditure        Trade   LaborForce Unemployment group
nbr.val       10.00000000  1.000000e+01                 10.00000000  10.00000000 1.000000e+01   10.0000000    NA
nbr.null       0.00000000  0.000000e+00                  0.00000000   0.00000000 0.000000e+00    0.0000000    NA
nbr.na         0.00000000  0.000000e+00                  0.00000000   0.00000000 0.000000e+00    0.0000000    NA
min           80.90243902  4.085776e+04                 83.03439489  55.91625104 3.250906e+07    3.7400000    NA
max           81.30487805  4.778724e+04                 85.31046763  62.92726899 3.477680e+07    8.0400000    NA
range          0.40243903  6.929486e+03                  2.27607274   7.01101795 2.267741e+06    4.3000000    NA
sum          810.89756097  4.320501e+05                839.74100198 599.39690350 3.377442e+08   56.2020004    NA
median        81.08048781  4.287865e+04                 83.84142796  60.98013859 3.385957e+07    5.0550001    NA
mean          81.08975610  4.320501e+04                 83.97410020  59.93969035 3.377442e+07    5.6202000    NA
SE.mean        0.05085923  6.689934e+02                  0.27868489   0.82174748 2.529497e+05    0.5240166    NA
CI.mean        0.11505157  1.513368e+03                  0.63042903   1.85892195 5.722119e+05    1.1854079    NA
var            0.02586661  4.475522e+06                  0.77665270   6.75268925 6.398353e+11    2.7459338    NA
std.dev        0.16083101  2.115543e+03                  0.88127901   2.59859371 7.998970e+05    1.6570859    NA
coef.var       0.00198337  4.896522e-02                  0.01049465   0.04335347 2.368352e-02    0.2948447    NA
skewness       0.05837371  8.333716e-01                  0.33279247  -0.35105807 -1.950280e-01    0.3858566    NA
skew.2SE       0.04248185  6.064917e-01                  0.24219191  -0.25548482 -1.419329e-01    0.2808097    NA
kurtosis      -1.95280030 -3.855164e-01                 -1.66045504  -1.69037123 -1.618763e+00   -1.7031136    NA
```

```
CountryName: United States
           CountryName           Year  GNIPerCapita PopulationTotal RuralPopulation RuralPopulation_.. UrbanPopulation UrbanPopulation_..           GDP           GNI
nbr.val             NA  1.000000e+01  1.000000e+01    1.000000e+01    1.000000e+01        10.00000000    1.000000e+01        10.000000000  1.000000e+01  1.000000e+01
nbr.null            NA  0.000000e+00  0.000000e+00    0.000000e+00    0.000000e+00         0.00000000    0.000000e+00         0.000000000  0.000000e+00  0.000000e+00
nbr.na              NA  0.000000e+00  0.000000e+00    0.000000e+00    0.000000e+00         0.00000000    0.000000e+00         0.000000000  0.000000e+00  0.000000e+00
min                 NA  2.011000e+03  5.057000e+04    3.115835e+08    5.746903e+07        17.33600000    2.522081e+08        80.944000000  1.560000e+13  1.590000e+13
max                 NA  2.020000e+03  6.308000e+04    3.315011e+08    5.937535e+07        19.05600000    2.740321e+08        82.664000000  2.140000e+13  2.170000e+13
range               NA  9.000000e+00  1.251000e+04    1.991760e+07    1.906321e+06         1.72000000    2.182392e+07         1.720000000  5.800000e+12  5.800000e+12
sum                 NA  2.015500e+04  5.613620e+05    3.215510e+09    5.854765e+08       182.18500000    2.630033e+09       817.815000000  1.855000e+14  1.893000e+14
median              NA  2.015500e+03  5.613600e+04    3.219054e+08    5.869350e+07        18.23350000    2.632119e+08        81.766500000  1.845000e+13  1.885000e+13
mean                NA  2.015500e+03  5.613620e+04    3.215510e+08    5.854765e+07        18.21850000    2.630033e+08        81.781500000  1.855000e+13  1.893000e+13
SE.mean             NA  9.574271e-01  1.081551e+03    2.073741e+06    2.151097e+05         0.18319166    2.283845e+06         0.183191657  6.344289e+11  6.340435e+11
CI.mean             NA  2.165851e+00  2.446639e+03    4.691128e+06    4.866118e+05         0.41440832    5.166417e+06         0.414408320  1.435178e+12  1.434306e+12
var                 NA  9.166667e+00  1.169753e+07    4.300401e+13    4.627216e+11         0.33559183    5.215949e+13         0.335591833  4.025000e+24  4.020111e+24
std.dev             NA  3.027650e+00  3.420165e+03    6.557745e+06    6.802365e+05         0.57930289    7.222153e+06         0.579302886  2.006240e+12  2.005021e+12
coef.var            NA  1.502183e-03  6.092620e-02    2.039411e-02    1.161851e-02         0.03179751    2.746031e-02         0.007083544  1.081531e-01  1.059177e-01
skewness            NA  0.000000e+00  3.468062e-01   -5.189672e-02   -3.546394e-01        -0.05570173   -1.480194e-02         0.055701731  1.337442e-02 -2.205297e-02
skew.2SE            NA  0.000000e+00  2.523905e-01   -3.776818e-02   -2.580911e-01        -0.04053730   -1.077221e-02         0.040537301  9.733321e-03 -1.604920e-02
kurtosis            NA -1.561636e+00 -4.353404e-01   -1.528644e+00   -1.532353e+00        -1.56093507   -1.537705e+00        -1.560935066 -1.621548e+00 -1.607187e+00
kurt.2SE            NA -5.852118e-01 -1.631406e-01   -5.728483e-01   -5.742383e-01        -0.58494904   -5.762437e-01        -0.584949037 -6.076632e-01 -6.022815e-01
normtest.W          NA  9.701646e-01  9.573487e-01    9.736500e-01    9.278847e-01         0.96927611    9.735553e-01         0.969276107  9.567884e-01  9.553445e-01
normtest.p          NA  8.923673e-01  7.552709e-01    9.224704e-01    4.274114e-01         0.88405345    9.217099e-01         0.884053449  7.487245e-01  7.317663e-01
           LifeExpectancy  GDPPerCapita FinalConsumptionExpenditure       Trade   LaborForce Unemployment group
nbr.val      1.000000e+01  1.000000e+01                 10.00000000  10.0000000 1.000000e+01   10.0000000    NA
nbr.null     0.000000e+00  0.000000e+00                  0.00000000   0.0000000 0.000000e+00    0.0000000    NA
nbr.na       0.000000e+00  0.000000e+00                  0.00000000   0.0000000 0.000000e+00    0.0000000    NA
min          7.728049e+01  5.006597e+04                 81.42489189  23.4424574 1.571632e+08    3.6700001    NA
max          7.884146e+01  6.509480e+04                 84.68203994  30.8424862 1.663556e+08    8.9499998    NA
range        1.560976e+00  1.502883e+04                  3.25714805   7.4000288 9.192466e+06    5.2799997    NA
sum          7.854415e+02  5.757373e+05                823.71862467 280.5703222 1.613043e+09   60.6900001    NA
median       7.866585e+01  5.731474e+04                 82.05492423  27.7005459 1.608744e+08    5.7250001    NA
mean         7.854415e+01  5.757373e+04                 82.37186247  28.0570322 1.613043e+08    6.0690000    NA
SE.mean      1.438494e-01  1.612900e+03                  0.31034927   0.7411827 1.049794e+06    0.6089325    NA
CI.mean      3.254098e-01  3.648634e+03                  0.70205881   1.6766718 2.374800e+06    1.3775009    NA
var          2.069264e-01  2.601447e+07                  0.96316667   5.4935182 1.102068e+13    3.7079874    NA
std.dev      4.548916e-01  5.100438e+03                  0.98141055   2.3438255 3.319741e+06    1.9256135    NA
coef.var     5.791540e-03  8.858967e-02                  0.01191439   0.0835379 2.058061e-02    0.3172868    NA
skewness    -2.078896e+00  2.464404e-02                  1.27067256  -0.4079916 1.594492e-01    0.1496441    NA
skew.2SE    -1.512930e+00  1.793486e-02                  0.92474031  -0.2969186 1.160402e-01    0.1089045    NA
kurtosis     2.990100e+00 -1.590768e+00                  0.34349805  -1.0022334 -1.770096e+00  -1.7677067    NA
```

Looking at the measures of central dependency (mean and median) of the GDP and GNI values for each country, we can see that USA has the highest values of both measures while China has the lowest values. Meanwhile, China has the highest values of population. We also noticed that the countries in Group A (lower population) have the mean and median of the Life Expectancy higher than 80 but for group B, the values are lower with Japan being the only country with values higher than 80. We'll analyse further in this study to understand the effect of the size of population on Life expectancy.

Using range for detection of outliers may not be very efficient. Kurtosis on the other hand is a metric that indicates how heavy-tailed, or light tailed the data are in comparison to a normal distribution. Data with outliers tend to have heavy tails which relates to high kurtosis and vice versa. There are three categories; Mesokurtic (kurtosis = 3) like a normal distribution, Leptokurtic (kurtosis > 3) it shows as a curve with long tails which signifies outliers, and Platykurtic (kurtosis < 3) short tails meaning fewer outliers (Will Kenton, 2022). Looking at the kurtosis for our dataset, the values are less than 3 which signifies no or few outliers.

Skewness is a metric for the asymmetry or distortion of symmetrical distribution in a set of data. When the skew is negative, the distribution's tail is on the left side, when it is positive, the distribution's tail is on the right side and when it is zero it means there is no skewness and the distribution is symmetric (Zach, 2020).



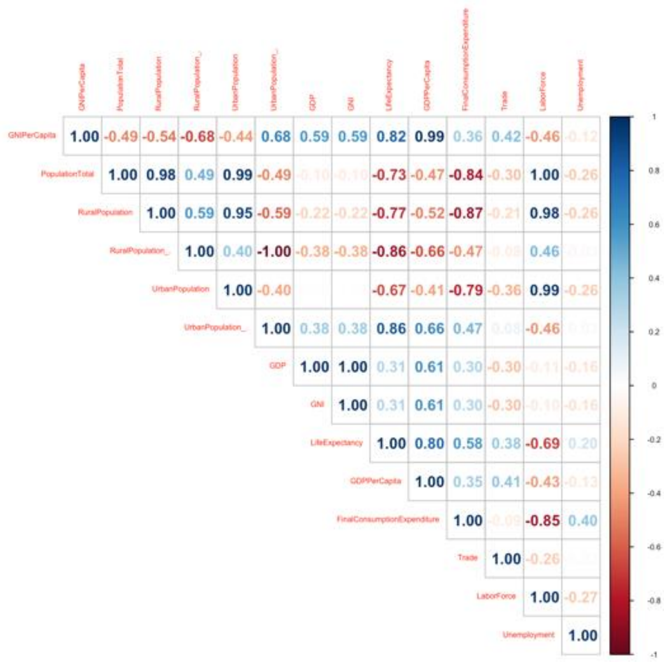
```
> by(dataset, INDICES=CountryName, FUN = summary)
CountryName: Brazil
 CountryName             Year       GNIPerCapita    PopulationTotal       RuralPopulation      RuralPopulation_.  UrbanPopulation        UrbanPopulation_.
 Length:10          Min.   :2011   Min.   : 8670   Min.   :197514541   Min.   :27477555   Min.   :12.93   Min.   :167158531   Min.   :84.63
 Class :character   1st Qu.:2013   1st Qu.: 9395   1st Qu.:201467864   1st Qu.:28214000   1st Qu.:13.50   1st Qu.:171812207   1st Qu.:85.28
 Mode  :character   Median :2016   Median :10656   Median :205317408   Median :28936285   Median :14.09   Median :176381122   Median :85.91
                    Mean   :2016   Mean   :10656   Mean   :205214837   Mean   :28930110   Mean   :14.12   Mean   :176284727   Mean   :85.88
                    3rd Qu.:2018   3rd Qu.:11845   3rd Qu.:209060446   3rd Qu.:29655657   3rd Qu.:14.72   3rd Qu.:180846446   3rd Qu.:86.50
                    Max.   :2020   Max.   :12810   Max.   :212559409   Max.   :30356010   Max.   :15.37   Max.   :185081854   Max.   :87.07
      GDP                 GNI             LifeExpectancy   GDPPerCapita    FinalConsumptionExpenditure     Trade          LaborForce          Unemployment
 Min.   :1.450e+12   Min.   :1.410e+12   Min.   :73.92   Min.   : 6815   Min.   :78.94                 Min.   :23.93   Min.   : 93794015   Min.   : 6.660
 1st Qu.:1.818e+12   1st Qu.:1.775e+12   1st Qu.:74.55   1st Qu.: 8830   1st Qu.:80.98                 1st Qu.:24.57   1st Qu.: 96206684   1st Qu.: 7.032
 Median :1.990e+12   Median :1.940e+12   Median :75.11   Median : 9540   Median :83.56                 Median :25.45   Median : 97957356   Median :10.015
 Mean   :2.092e+12   Mean   :2.042e+12   Mean   :75.07   Mean   :10232   Mean   :82.76                 Mean   :26.60   Mean   : 98541984   Mean   : 9.855
 3rd Qu.:2.468e+12   3rd Qu.:2.408e+12   3rd Qu.:75.62   3rd Qu.:12253   3rd Qu.:84.60                 3rd Qu.:28.40   3rd Qu.:101063762   3rd Qu.:12.230
 Max.   :2.620e+12   Max.   :2.550e+12   Max.   :76.08   Max.   :13245   Max.   :85.13                 Max.   :32.89   Max.   :104376954   Max.   :13.690
    group
 Length:10
 Class :character
 Mode  :character

------------------------------------------------------------------------------------------------------------------------
CountryName: Canada
 CountryName             Year       GNIPerCapita    PopulationTotal      RuralPopulation     RuralPopulation_.  UrbanPopulation       UrbanPopulation_.      GDP
 Length:10          Min.   :2011   Min.   :42960   Min.   :34339328   Min.   :6491507   Min.   :18.44   Min.   :27847821   Min.   :81.10   Min.   :1.530e+12
 Class :character   1st Qu.:2013   1st Qu.:45502   1st Qu.:35171574   1st Qu.:6616450   1st Qu.:18.60   1st Qu.:28555124   1st Qu.:81.19   1st Qu.:1.650e+12
 Mode  :character   Median :2016   Median :47710   Median :35906198   Median :6721778   Median :18.72   Median :29184420   Median :81.28   Median :1.735e+12
                    Mean   :2016   Mean   :47830   Mean   :36063509   Mean   :6742439   Mean   :18.70   Mean   :29321070   Mean   :81.30   Mean   :1.714e+12
                    3rd Qu.:2018   3rd Qu.:50268   3rd Qu.:36935122   3rd Qu.:6871443   3rd Qu.:18.81   3rd Qu.:30063679   3rd Qu.:81.40   3rd Qu.:1.805e+12
                    Max.   :2020   Max.   :52810   Max.   :38037204   Max.   :7013300   Max.   :18.90   Max.   :31023904   Max.   :81.56   Max.   :1.850e+12
      GNI              LifeExpectancy   GDPPerCapita   FinalConsumptionExpenditure     Trade          LaborForce         Unemployment        group
 Min.   :1.510e+12   Min.   :81.45   Min.   :42316   Min.   :76.06                 Min.   :60.78   Min.   :19147395   Min.   :5.660   Length:10
 1st Qu.:1.630e+12   1st Qu.:81.75   1st Qu.:43979   1st Qu.:77.06                 1st Qu.:62.52   1st Qu.:19567200   1st Qu.:6.482   Class :character
 Median :1.710e+12   Median :81.85   Median :46439   Median :78.54                 Median :64.74   Median :19851520   Median :6.955   Mode  :character
 Mean   :1.688e+12   Mean   :81.82   Mean   :47566   Mean   :78.06                 Mean   :64.10   Mean   :19926640   Mean   :6.998
 3rd Qu.:1.775e+12   3rd Qu.:81.90   3rd Qu.:51907   3rd Qu.:78.68                 3rd Qu.:65.40   3rd Qu.:20352044   3rd Qu.:7.235
 Max.   :1.820e+12   Max.   :82.05   Max.   :52669   Max.   :79.81                 Max.   :66.45   Max.   :20792599   Max.   :9.460
------------------------------------------------------------------------------------------------------------------------
```

We also looked at the quartile levels of all the indicators for all the countries using the ‘summary’ function in R. Examining the ranges of values with the minimum and maximum values, we see that values are close in proximity which mean no extreme outliers.

4.2 Correlation Analysis

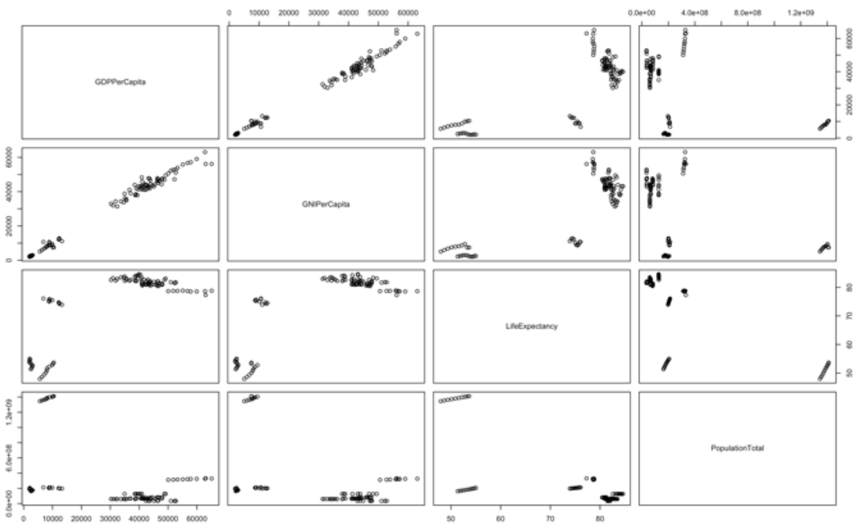
It is important to know what type of relationship exists between the key indicators in this study. If there is a linear relationship between them and if yes, is it positive or negative. Also, what is the effect of the population size on some of these indicators.

We used the Pearson’s correlation (cor () in R) function to compute the correlation value for each of our continuous variables. We first removed the ‘Country name’, ‘Year’ and ‘Group’ columns as they’re not continuous variables.



For the general correlation plot, we see the indicators that are strongly correlated either positively or not, the indicators that weakly correlated and those that aren't corelated at all. If there is a negative correlation, it means that the two variables are changing in the opposite ways; for example, if one variable rises, the other falls, and vice versa. For positive correlation, there go in the same direction. We can clearly see a strong positive correlation between GDP and GNI. This is not surprising as both factors strongly contribute to the socio-economic state of a country.

The correlation matrix maybe not be so readable, so we will be drawing a scatterplot to visualize the correlation between key indicators. The indicators are GDP per capita, GNI per capita, Population, and Life expectancy.



We can see negative correlation between Population with GDP per capita, GNI per capita and Life expectancy but positive correlations between GDP per capita, GNI per capita and Life expectancy. This speaks further to what we saw in the Descriptive analysis part. The size of population does matter. We will carry out a one-way ANOVA test to determine the effect of population size on some of these indicators.

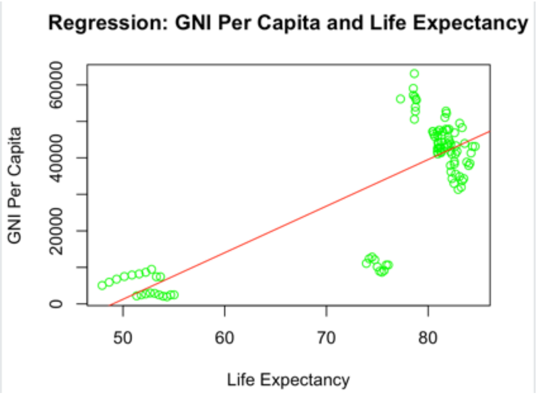
4.3 Regression Analysis

A strong statistical technique that enables one to investigate the relationship between two or more relevant variables is regression analysis. We can confidently establish which elements are most important, which ones can be ignored, and how these factors interact when we do a regression. To perform regression analysis, we need a dependent variable (Y) and one or more independent variables (X_i). Y is the variable we want to study or predict, while X_i are the variables we assume will influence Y. When we have one independent variable, it called a Simple Linear Regression (SLR). The model is defined by $Y = \beta_0 + \beta_1 X_1 + \varepsilon$. When we have more than one independent variable, it is called a Multiple Linear Regression (MLR). The model is defined by $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$. Where ε is the random error term. For SLR and MLR, the independent variables and dependent variable must be numerical (Brain Beers, 2022).

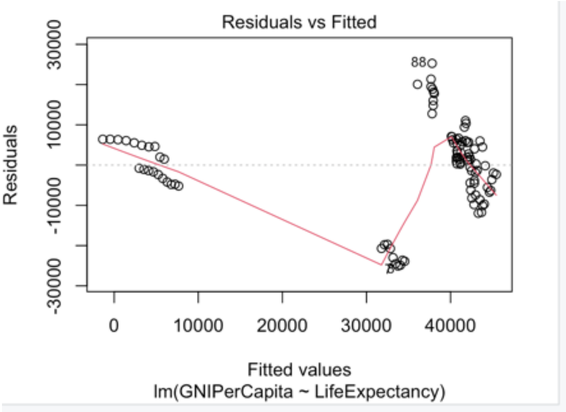
We want to examine the possible linear relation between GNI per capita and Life expectancy. Coefficient of determination (R^2) (R square greater than 0.5 is good) as explained in the background research will explain how much of the variation of the GNI is explained by the variation of the Life expectancy.

```
> model1 <- lm(GNIPerCapita ~ LifeExpectancy, dat)
> summary(model1)

Call:
lm(formula = GNIPerCapita ~ LifeExpectancy, data = dat)

Residuals:
   Min     1Q Median     3Q    Max 
-25074  -4567   1365   5704  25274 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)    -62550.38    6850.82   -9.13 9.34e-15 ***
LifeExpectancy   1276.17      90.17   14.15  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10570 on 98 degrees of freedom
Multiple R-squared:  0.6715,	Adjusted R-squared:  0.6681 
F-statistic: 200.3 on 1 and 98 DF,  p-value: < 2.2e-16
```

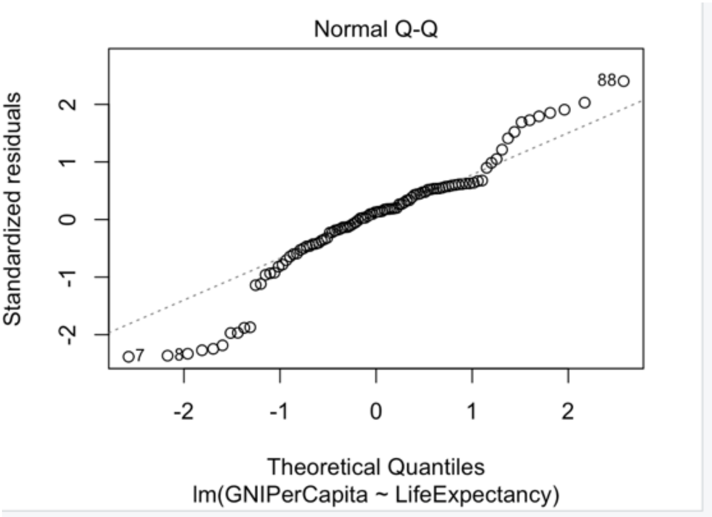
We see that 66% of the variation of the GNI per capita is explained by the variation of the Life expectancy. Now we will carry out the assumption tests. Assumption 1 which speaks about linearity between X and Y is verified in the image below.



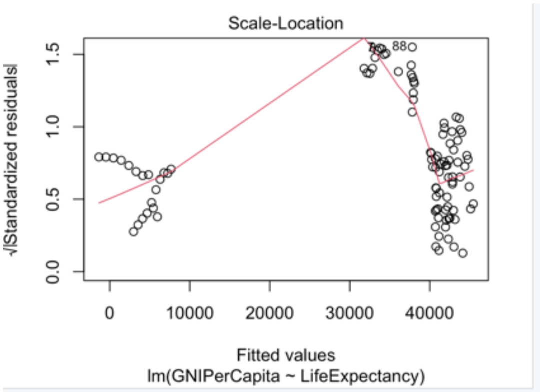
Assumption 2 which is Residuals Independence is not verified as shown in the plot below. We can see some sort of relationships.



Assumption 3 which is Normality Independence is verified as shown in the plot below. It looks like a normal distribution plot.

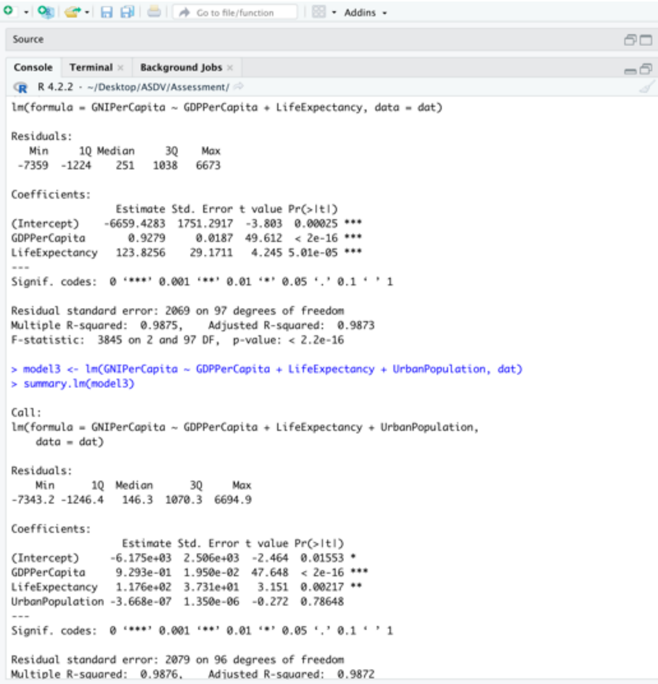


Assumption 4 which is equal variance is not verified as shown in the plot below. We can see some sort of relationships.



Now you can see that we cannot use this equation to predict GNI per capita based on Life Expectancy for a new data point in any of the countries (a new sample from the population) which is expected.

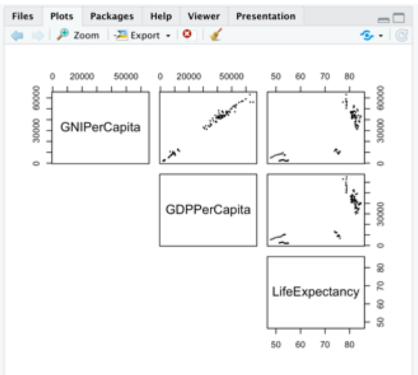
Now we will perform a MLR using the forward stepwise method (which function in a manner that until a predetermined stopping rule is reached or all the variables being taken into consideration are included in the model, the most important variables are added one after the other). We will add the indicators by the degree of correlation with GNI per capita using the correlation matrix in (4.2). We see that GNI per capita has a strong correlation with GDP per capita, Life expectancy and Urban Population in this order. We will be using these indicators for the method.



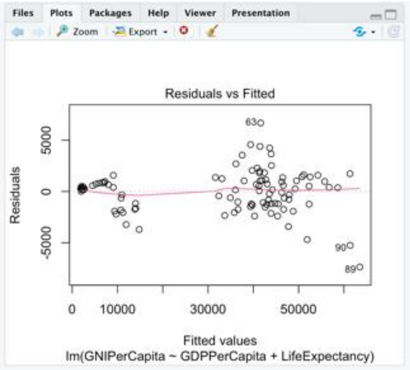
```
lm(formula = GNIPerCapita ~ GDPPerCapita + LifeExpectancy, data = dat)

Residuals:
   Min     1Q Median     3Q    Max 
 -7359  -1224    251   1038   6673 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)   -6659.4283  1751.2917  -3.803  0.00025 ***
GDPPerCapita      0.9279     0.0187  49.612  < 2e-16 ***
LifeExpectancy  123.8256    29.1711   4.245 5.01e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2069 on 97 degrees of freedom
Multiple R-squared:  0.9875,	Adjusted R-squared:  0.9873 
F-statistic:  3845 on 2 and 97 DF,  p-value: < 2.2e-16

> model3 <- lm(GNIPerCapita ~ GDPPerCapita + LifeExpectancy + UrbanPopulation, dat)
> summary.lm(model3)

Call:
lm(formula = GNIPerCapita ~ GDPPerCapita + LifeExpectancy + UrbanPopulation, 
    data = dat)

Residuals:
    Min      1Q  Median      3Q     Max 
-7343.2 -1246.4   146.3  1070.3  6694.9 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -6.175e+03  2.506e+03  -2.464  0.01553 *  
GDPPerCapita     9.293e-01  1.950e-02  47.648  < 2e-16 ***
LifeExpectancy   1.176e+02  3.731e+01   3.151  0.00217 ** 
UrbanPopulation -3.668e-07  1.350e-06  -0.272  0.78648    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2079 on 96 degrees of freedom
Multiple R-squared:  0.9876,	Adjusted R-squared:  0.9872 
```

We see that model2 has a slightly greater R squared value to model3. We will carry out the tests for the assumptions to see if model2 is a good model.

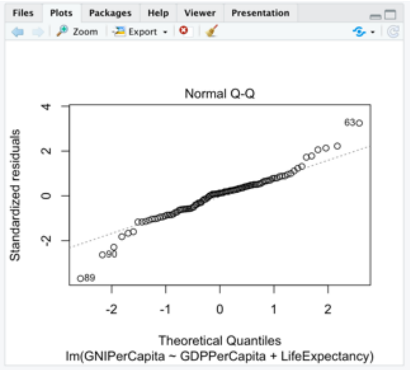
Assumption 1 which is linearity between the Xs and Y is verified looking at the figure below.



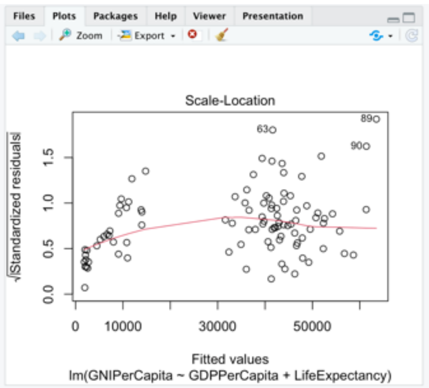
Assumption 2 which demands that the Residuals be independent is also verified as seen in the plot below. There is no visible relationship.



Assumption 3 which is normality of residuals is verified because we can see in the plot below that the curve looks like it follows that of a normal distribution.



Assumption 4 which demands that all residuals must have equal variance (homoscedasticity) is verified in the plot below as we see no relationships.



Assumption 5 which demands that there is multicollinearity is verified using the variance inflation factor (VIF) metric which evaluates the degree to which the predictor variables in a regression model are correlated (Ronald N. Forthofer, Eun Sul Lee and Mike Hernandez, 2007). VIF value starts at 1. A value between 1 and 5 shows weak corelation which is not severe enough but a value greater than 5 shows serious correlation. In the image below, we see that our VIFs values are <5 which means the assumption is verified.

```
> vif(model2) #VIF less than 5
GDPPerCapita LifeExpectancy
     2.73224        2.73224
> 
```

All assumptions are verified so our final model cane be written as

$$GNI\ Per\ Capita = -6659.4283 + 0.9279 * GDP\ Per\ Capita + 123.8256 * Life\ Expectancy$$

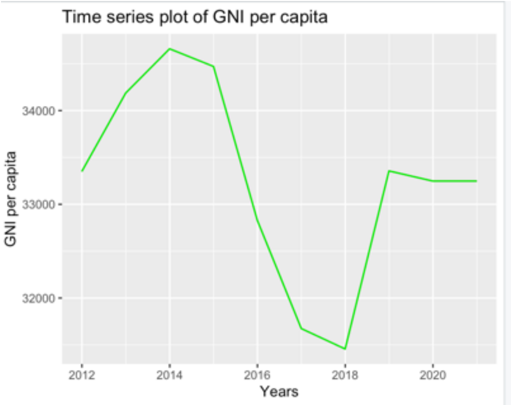
In real world, this would be expected as GDP is also an important factor in determining socio-economic state of a country like GNI.

4.4 Time Series Analysis

A time series is a group of data points arranged chronologically. Time series analysis is the process of analysing a sequence of data points collected over an interval of time which can be annually, quarterly, daily, etc, in a specific manner. It aides with the understanding of trends, patterns, and their causes over time. One can also predict the likelihood of an occurrence using time series forecasting. Some popular areas where time series is used are weather data, stock, interest rates, etc (Aryan Gupta, 2022).

A time series has four components namely Trend component, Seasonal component, Cyclical component, and Irregular component. There are two major types of time series data type: stationary and non-stationary. A stationary time series is one with constant mean, variance, and covariance over time, without having the trend and seasonality components. To model a time series, we need to make sure it is stationary. A non-stationary time series can be made stationary by Detrending (removing the trend component from the data), Differencing (stabilize the time series' mean and get rid of the series' dependence on time) and Transforming (the use of Power Transform, Square Root, and Log Transfer methods) (Shanthababu Pandian, 2021).

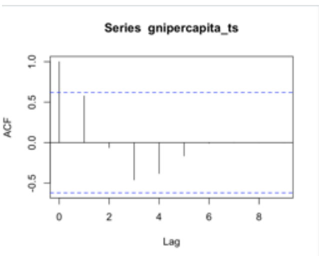
For this section, GNI per capita was selected for analysis and forecasting using time series analysis. We grouped the data by year and used the mean values of the GNI per capita for each year to build a time series data. The plot of the time series is seen in the image below.

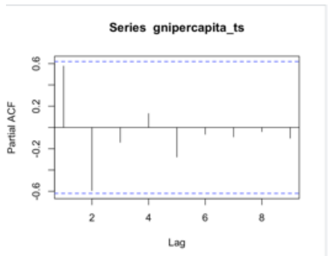


We can see our time series is non seasonal and decomposing it using simple moving average does not add much information to what we already can see from the plot of the time series because there aren't many fluctuations. Now we will be building an ARIMA model because we do not have a seasonal component (Jason Brownlee , 2018).

ARIMA (autoregressive integrated moving average). The AR stands for Autoregression (based on past values, forecasts future values), I stand for integrated (determines how static data values differ from earlier values) and MA stands for moving average (dependence between an observed value and a residual error from previous observations using a moving average model). An ARIMA model order is shown as (p,d,q), where the values represent the order or frequency of the function during model execution. We differenced the time series, but it did not make the graph more stationary, so we have d= 0.

To plot the correlogram and partial parallelogram to determine the values of p and q, we used the acf() and pacf() functions.



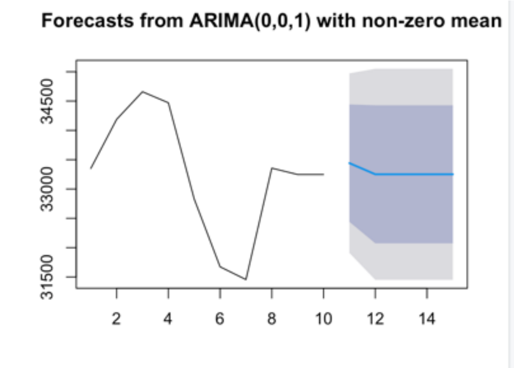


From the pacf plot, we see that the correlogram tails off to zero after lag one so our p=0 and our q=1. Now we will fit an ARIMA(0,0,1) model.

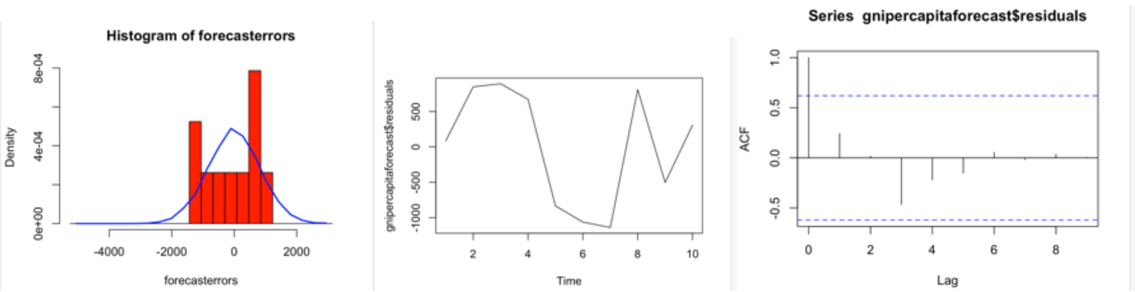
```
> gnipercapitaarima

Call:
arima(x = gnipercapita_ts, order = c(0, 0, 1))

Coefficients:
         ma1   intercept
      0.6186  33251.2571
s.e.  0.2144    385.3676

sigma^2 estimated as 609845:  log likelihood = -81.04,  aic = 168.07
```

We can now use the ARIMA model to make predictions using the “forecast” function. The image below shows the future mean value of the GNI per capita indicator for our 10 countries for the next 5 years. The plot is also included below.

```
> gnipercapitaforecast <- forecast(gnipercapitaarima, h=5)
> gnipercapitaforecast
   Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
11       33441.73 32440.93 34442.54 31911.13 34972.33
12       33251.26 32074.47 34428.04 31451.52 35050.99
13       33251.26 32074.47 34428.04 31451.52 35050.99
14       33251.26 32074.47 34428.04 31451.52 35050.99
15       33251.26 32074.47 34428.04 31451.52 35050.99
>
```



The ARIMA(0,0,1) seem to efficiently forecast the mean GNI per capita for our 10 countries because from the plots below, we see on the histogram that the mean of the forecast errors is close to zero and the errors are normally distributed, the forecast error plot shows constant variance and none of the lags exceeds the significance bound on the correlogram.



4.5 Tests

As a researcher, it is necessary to answer the important questions following the research topic. It is paramount that the right methods are implemented to satisfy the budding curiosity. In our case, we want to understand the growth of a country using different indicators. Does the size of a population affect the socio-economic growth of a country? We will be using ANOVA and T-tests to compare GNI per capita, GDP per capita and life expectancy recorded for the different groups of population sizes, i.e., the densely populated and the less. We use ANOVA because it is suitable to test the impact on a dependent variable of two or more class of one or two independent, and we use T-tests to evaluate whether there is a significant difference between the means of two groups and their relationships.

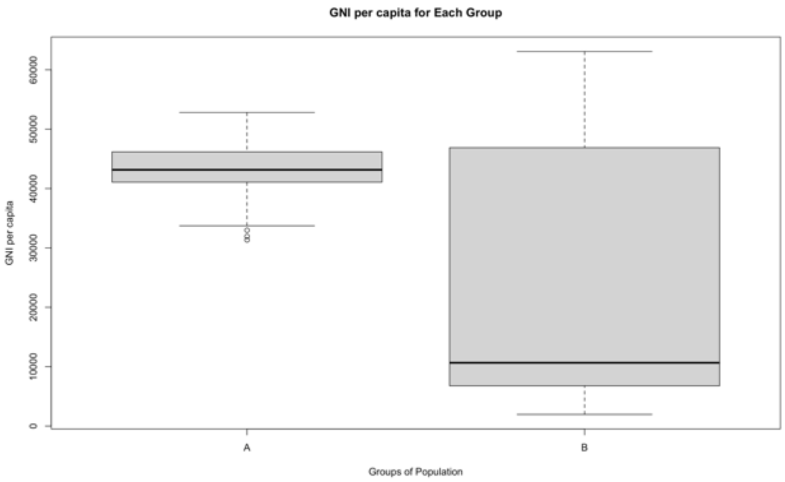
ANOVA

We want to compare the GNI per Capita for the groups of countries (Large and smaller population). We will be using a way one ANOVA analysis for this. In this case, the Null Hypothesis is that the different population sizes have similar effects and none of them have a different effect on the GNI value.

$$H0{:}\mu A = \mu B$$

There are 6 assumptions that needs to be treated before we can carry out an ANOVA. We will be checking for those assumptions.

Firstly, our dependent variable will be GNI per capita, and our independent variable will be Population grouped into two. Assumption 1 and 2 which says that the dependent variable should be continuous, and that the independent variable should be categorical with two or more groups is verified. Assumption 3 which refers to zero relationship between or in the groups is also verified. Assumption 4 asks that there should not be any significant outliers. From the image below, we see that there are only 3 outliers, but they do not seem to be influential.



Assumption 5 requests that the dependent variable should be normally distributed for each group. We will be using the Shapiro wilk test for this with the Null Hypothesis (H null) as the variable follows a normal distribution. The diagram below shows that we do not reject the Null hypothesis.

```
> #A5 assumption to check for normality
> byf.shapiro(GNIPerCapita ~ group, data=dataset)

        Shapiro-Wilk normality tests

data:  GNIPerCapita by group

            W   p-value
GroupA 0.9564   0.06259 .
GroupB 0.7989 8.271e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

>
```

This means that we can’t perform one-way ANOVA. We will be using the log, sqrt and exp transformation to transform our dataset into a normal distribution and then we confirm assumption 5 and 6.

```
> #Bartlett homogeneity test
> var<- bartlett.test(GNIPerCapita ~ group, data= dataset)
> var

        Bartlett test of homogeneity of variances

data:  GNIPerCapita by group
Bartlett's K-squared = 78.468, df = 1, p-value < 2.2e-16
```

From the image above, we see that p<0.05 which indicates that there is a considerable difference in the variance levels of the two groups (assumption 6), so we draw the conclusion that we can’t use ANOVA. We will now be using two sample t-tests.

TWO SAMPLE T-TEST

Two sample t-Test is a technique used to determine if the unknown population means of two groups are equal or not. We’re going to investigate whether the GNI per Capita and Life Expectancy is different for the two groups of population. In this case, our Null Hypothesis is:

$$H_0: \mu_{GroupA} = \mu_{GroupB}$$

And our Alternative Hypothesis is:

$$H_1: \mu_{GroupA} \neq \mu_{GroupB}$$

```
> #TWO SAMPLE T-TEST
> t.test(GNIPerCapita ~ group, dataset)

        Welch Two Sample t-test

data:  GNIPerCapita by group
t = 5.8417, df = 54.395, p-value = 2.991e-07
alternative hypothesis: true difference in means between group GroupA and group GroupB is not equal to
 0
95 percent confidence interval:
 12192.34 24931.10
sample estimates:
mean in group GroupA mean in group GroupB
            42528.68             23966.96

>
```

We see that our p value is less than 0.05 so we reject the null hypothesis. We will be carrying out same test for life expectancy like we mentioned earlier in the report.

```
> t.test(LifeExpectancy ~ group, dataset) #for Life expectancy

	Welch Two Sample t-test

data:  LifeExpectancy by group
t = 6.9331, df = 49.348, p-value = 8.138e-09
alternative hypothesis: true difference in means between group GroupA and group GroupB is not equal to
 0
95 percent confidence interval:
  9.554608 17.352254
sample estimates:
mean in group GroupA mean in group GroupB
            81.79395             68.34052

> |
```

We see that our p value is less than 0.05 so we reject the null hypothesis.

5. Discussion

To study the socio-economic states of these countries, it was paramount we carry out a comprehensive statistical analysis. Firstly, getting the general overview of the indicators helped to form the important questions needed to fully understand the objective and implementation of this analysis. Using the mean, median, max, min, mode, standard deviation, skewness, and kurtosis values gave a general sense of each of the indicators per country. We noticed a pattern with the Life Expectancy of each country following their population size. We divided our dataset into two using the median of population and we noticed that the countries with lower population size had the mean of the value of Life Expectancy greater than 80 and it was lower than 80 for the countries with larger population except Japan.

The Pearson's correlation analysis was used to analyse the relationships between indicators. This method helped us to visually see how negatively or positively correlated these indicators were. It also helped build curiosity on an already existing important research question which was the effect of population size on the indicators as we saw negative correlation between Population and GNI per Capita, GDP per Capita and Life Expectancy.

Using the forward stepwise method in multiple linear regression, we were able to efficiently build a model that showed the linear relationship between GNI per capita and the indicators that were strongly correlated with this indicator. We were able to build different models and pick the one which best predicted a future value for GNI per capita. It efficiently showed the weights needed for each indicator in the final equation to be able to predict future values of the GNI per capita.

We extracted a time series data from our dataset using the year and the GNI per capita columns. We did this to see the variation of this indicator over time and to build a model that can predict future values of this indicator. We did not pick a country for this analysis but instead, we used the mean of the GNI per capita for each year to see a general overview of the indicator for all the countries combined. From the time series plot, we noticed an exponential decline in the rate of GNI per capita from 2015 – 2017. This could be because of different factors in the social-economic sectors of our countries.

We were unable to use ANOVA for our dataset because we could not verify all the assumptions needed. We were able to use the two-sample t-test to test for the effects of population size on our key indicators.

For this study, we had a few limitations. First, some important indicators like Rate of Adult literacy and Expected years of schooling lacked substantial data so we were unable to use these indicators. We also could not get a real variation of points for our time series because we had limited number of years.

6. Conclusions

From the descriptive analysis we could see from the statistics summary that China has the lowest GNI per capita, and GDP per capita values and it is the most populous country amongst our 10 selected countries. Also, the countries with their mean population lower than 1.047e+08 have higher values of Life expectancy compared to countries with a higher population mean. The correlation analysis made us see that Population is negatively correlated with GDP per capita, GNI per capita and Life expectancy which means they go in opposite directions. From the regression analysis, we see that we can predict the values of GNI per capita using GDP per capita and Life Expectancy. Time series analysis showed a massive decline in GNI per capita rates from 2015 – 2017. The tests confirmed that indeed the population size of a country affects its socio-economic growth. Countries with lower population seem to fare better compared to those with a larger population size.

Part Three: References

Will Kenton. (2022). *Kurtosis Definition, Types, and Importance*. https://www.investopedia.com/terms/k/kurtosis.asp

Zach. (2020, October 23). *How to Calculate Skewness & Kurtosis in R*. https://www.statology.org/skewness-kurtosis-in-r/

Brain Beers. (2022). *What is Regression? Definition, Calculation, and Example*. https://www.investopedia.com/terms/r/regression.asp

Ronald N. Forthofer, Eun Sul Lee and Mike Hernandez. (2007). Linear Regression. In E. S. Ronald N. Forthofer, *Biostatistics.*

Aryan Gupta. (2022, September). *A Complete Guide To Get A Grasp Of Time Series Analysis*. https://www.simplilearn.com/tutorials/statistics-tutorial/what-is-time-series-analysis

Shanthababu Pandian. (2021, October). *A Comprehensive Guide to Time Series Analysis*. https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-to-time-series-analysis/

Jason Brownlee . (2018, August 17). *A Gentle Introduction to SARIMA for Time Series Forecasting in Python*. https://machinelearningmastery.com/sarima-for-time-series-forecasting-in-python/

The World Bank. (n.d.). *Economy*. https://datatopics.worldbank.org/world-development-indicators/themes/economy.html

Max Roser. (2014). *Human Development Index (HDI)*. https://ourworldindata.org/human-development-index

Valentino Piana. (2011). *Consumption*. http://www.economicswebinstitute.org/glossary/cons.html

Adam Hayes. (2022, October 8). *Labor Force Participation Rate: Purpose, Formula, and Trends*. https://www.investopedia.com/terms/p/participationrate.asp

Sandra Durcevic. (2022, October 14). *Move Beyond Excel, PowerPoint & Static Business Reporting with Powerful Interactive Dashboards*. https://www.datapine.com/blog/interactive-dashboard-features/

World Bank. (2001). *World development indicators 2007*. https://scholar.google.com/scholar_lookup?title=World+development+indicators&publication_year=2001&

Maria Francesca Cracolici et al. (2009, April 9). *The Measurement of Economic, Social and Environmental Performance of Countries: A Novel Approach*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2785896/

Insightsoftware. (2022). *A Brief History of Data Visualization*. https://insightsoftware.com/blog/a-brief-history-of-data-visualization/

Allen Hillery. (2020, October 19). *The Evolution of Data Visualization*. https://chartio.com/blog/the-evolution-of-data-visualization/

Arimetrics. (n.d.). *What is Dashboard*. https://www.arimetrics.com/en/digital-glossary/dashboard

Ratibor Sekirov. (2021, October 1). *Dashboard Design Overview: Trends, Examples and Best Practices*. https://aspirity.com/blog/dashboard-design-overview#popup:close

Neelam Tyagi. (2019, August 21). *What is Tableau? Working and Key Features*. https://www.analyticssteps.com/blogs/why-is-tableau-used-for-read-the-working-and-key-features-of-tableau

NHS Digital. (n.d.). *NHS e-Referral Service (e-RS) open data dashboard*. https://digital.nhs.uk/dashboards/ers-open-data

AHDB. (n.d.). *UK dairy trade dashboard*. https://ahdb.org.uk/dairy/trade-dashboard