Visualizations and statistical analysis of comparisons between the healthcare systems (by extension, Life expectancies) of African countries and other developed countries around the world

Chike Ayogu

Table of Contents

TASK 1: Visualisation

1. Introduction

A measure of how developed a country is can be reflected almost perfectly in the state of that country's healthcare system. The vast majority of African countries still fall very low in the rankings on the world development scale as compared to the more developed countries across the world.

According to the World Health Organization (WHO), Health Systems can be defined as "all the activities whose primary purpose is to promote, restore, or maintain health" (WHO, 2000). This definition is broad and can be further described to include all health-related activities and policies, which are in some way, influenced by the government of the country.

There are also health-related activities carried out by the private sector of a country, but this report focuses on only the activities that are partially or completely carried out by the government.

The statistics generated by these activities are called indicators, and a study of these indicators is what gives a representation of the state of the countries' healthcare systems.

The problem this dashboard seeks to illuminate and thereafter, attempt to rectify, is the difference in performance between the healthcare systems of the selected African countries and other developed countries.

The objective of the visualisation is to show the disparity between the healthcare systems of these two sets of countries, explore the relationships between some of the chosen indicators, show how the health indicators of these sets of countries have performed over the years, and then give a forecast to how they'll perform in coming years.

The basic strategy employed in tackling the problem is intertwined with the objectives, as answering the objectives' questions will lead to the identification of a proposed solution to the problem as stated above.

By researching, as shown in the next section, how the various indicators depict the state of the healthcare system, the trend that these indicators have undergone over the years, and then the relationship between key trends, an overview can be visualised and this makes up the general research direction.

2. Background Research

Telling a story using interactive dashboards has since gone beyond being a marginal feature in making decisions and extracting information from data. It has now evolved into being essential in the presentation and/or analysis of data. This is especially true when dealing with extensive data.

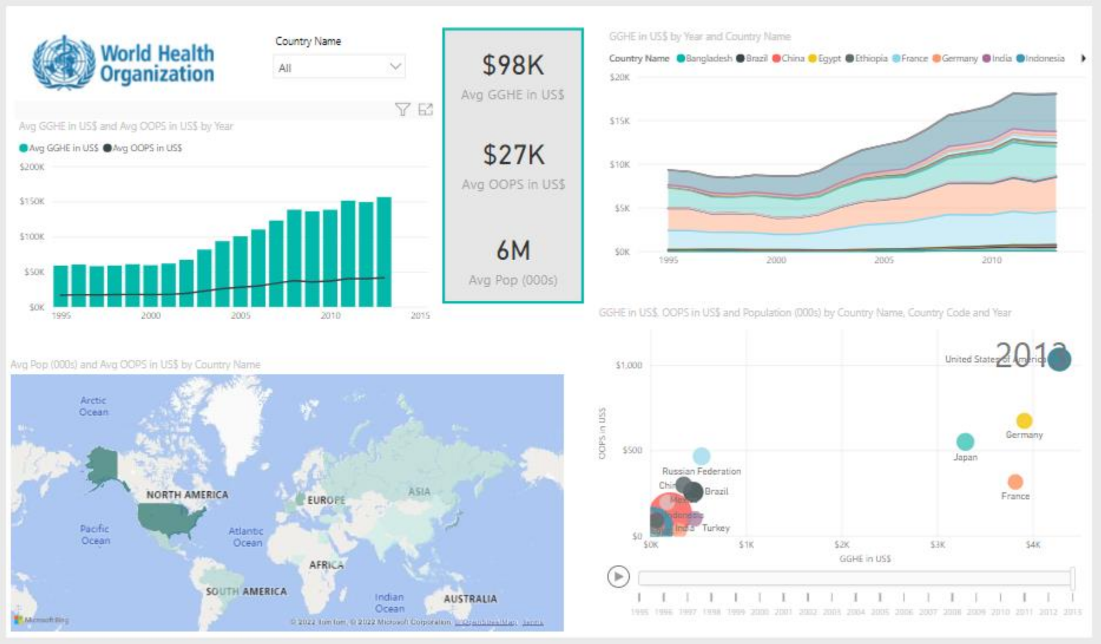
This report emphasized the GESTALT principles of visualisation.

According to research, to build great-looking and effective dashboards, some features need to be considered and incorporated into their design. Some of these features are outlined below.

2.1 Moving graphs

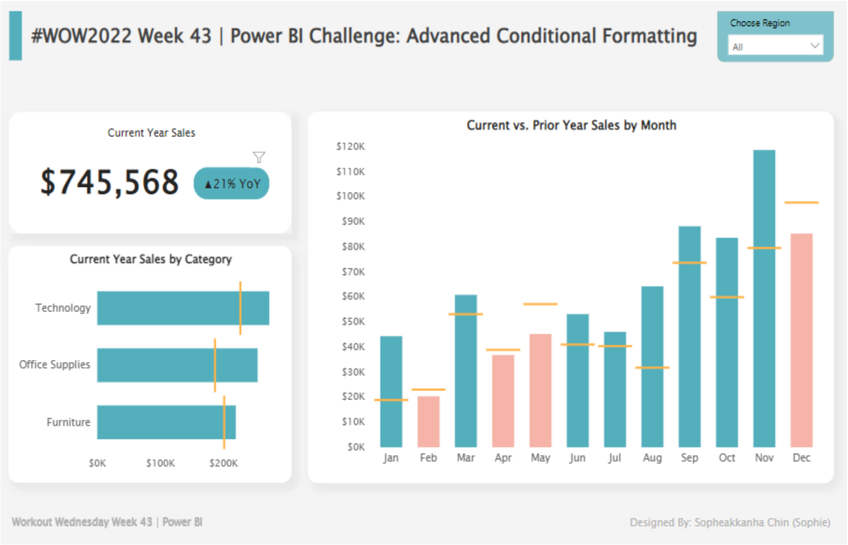
While researching similar topic dashboards with state-of-the-art designs, a feature of the dashboard below stood out; Moving graphs.

This can effectively capture a visual representation of the changes in data trends over time. As can be seen from this WHO dashboard of analysing healthcare with open data using power-BI, the graph on the bottom right corner of the dashboard has a 'play' button which makes the graph play like a video and shows the movement of data time. (Analyzing healthcare open data with power BI, 2017)



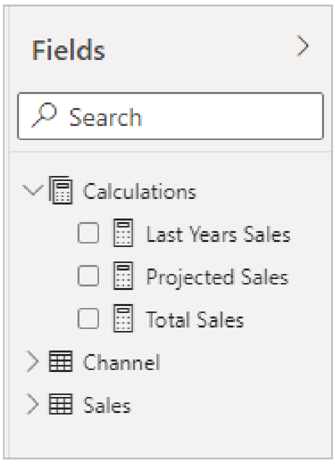
2.2 Conditional Formatting

The use of conditional formatting is a current trend that cannot be overlooked. This makes the dashboard more interactive by setting rules to a chart for the appearance of some of its features (like text, color, and size) to be changed if certain conditions are met: (WOW2022 Week43 | Power BI: Advanced conditional formatting, 2022)


#WOW2022 Week 43 | Power BI Challenge: Advanced Conditional Formatting
Choose Region
All
Current Year Sales
$745,568
▲21% YoY
Current Year Sales by Category
Technology
Office Supplies
Furniture
$0K
$100K
$200K
Current vs. Prior Year Sales by Month
$120K
$110K
$100K
$90K
$80K
$70K
$60K
$50K
$40K
$30K
$20K
$10K
$0K
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Workout Wednesday Week 43 | Power BI
Designed By: Sopheakkanha Chin (Sophie)

2.3 Creating a separate table for measures

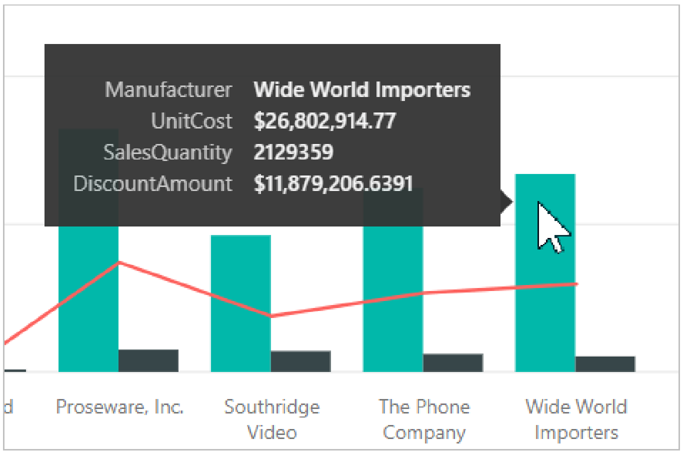
This might not directly impact the appearance of the dashboard, but it is good practice to have all measures kept in a separate table. It keeps the work organized and if the measures cut across (taking arguments from) different tables, the conundrum as to where the result should be placed is solved with the measures table. (Winter, 2021)



Notice how the ‘Calculations’ table has an icon different from that of the ‘Channel’ and ‘Sales’ tables. (Davidiseminger, n.d.)

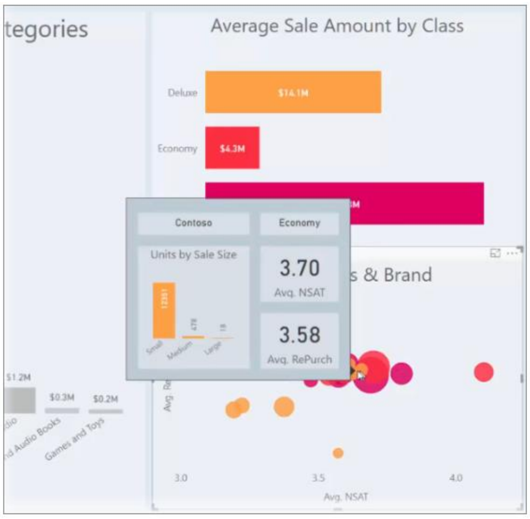
2.4 Adding to tooltips

Tooltips are the information that is shown in Power BI whenever you hover over any points of data in the visual. By default, there’s usually some information displayed, but often, this is not enough. In more recent state-of-the-art dashboards, these tooltips are important tools to convey additional information in a visual, and their potential can be maximized when used appropriately. (Compton, 2022)



Manufacturer Wide World Importers
UnitCost $26,802,914.77
SalesQuantity 2129359
DiscountAmount $11,879,206.6391

In addition to just adding more text to tooltips, extra visuals can also be added to tooltips such that whenever any data point is hovered over, the visual is displayed instead of just text. (MaggiesMSFT, n.d.)



3. Exploration of Dataset

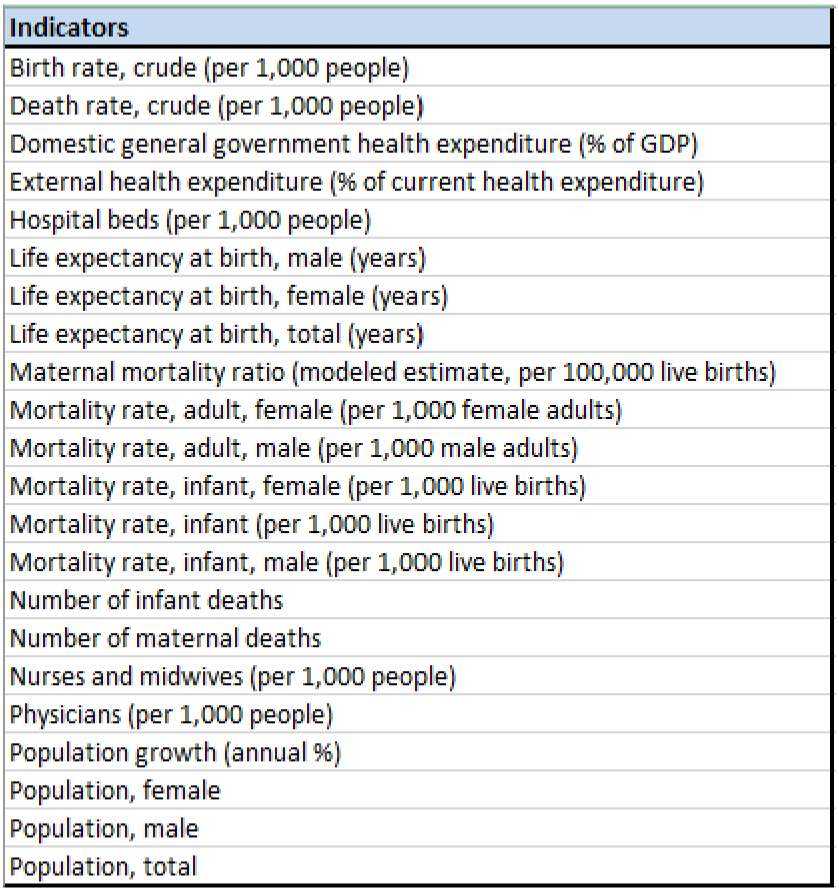
The dataset used in this task was downloaded from the World Bank databank, and consists of 12 countries - 6 countries from Africa, and 6 from developed countries in other parts of the world. The data collected spanned a period of 16 years, from 2002 – 2017, and comprises 22 economic indicators.

3.1 Countries and Indicators

The 12 countries selected are:

1. Australia	4. Canada	7. Japan	10. Sudan
2. Burkina Faso	5. Denmark	8. Kenya	11. Uganda
3. Cameroon	6. Ethiopia	9. Norway	12. United Kingdom

The 22 indicators are:



Indicators
Birth rate, crude (per 1,000 people)
Death rate, crude (per 1,000 people)
Domestic general government health expenditure (% of GDP)
External health expenditure (% of current health expenditure)
Hospital beds (per 1,000 people)
Life expectancy at birth, male (years)
Life expectancy at birth, female (years)
Life expectancy at birth, total (years)
Maternal mortality ratio (modeled estimate, per 100,000 live births)
Mortality rate, adult, female (per 1,000 female adults)
Mortality rate, adult, male (per 1,000 male adults)
Mortality rate, infant, female (per 1,000 live births)
Mortality rate, infant (per 1,000 live births)
Mortality rate, infant, male (per 1,000 live births)
Number of infant deaths
Number of maternal deaths
Nurses and midwives (per 1,000 people)
Physicians (per 1,000 people)
Population growth (annual %)
Population, female
Population, male
Population, total

3.2 Data pre-processing

Data Type Check

After uploading the dataset, and opening the Power Query via 'Transform', the 'Detect Data Type' function was used to automatically register the data types, and a check was done to ensure accuracy.

The column 'Year' came in the whole number data type and was first converted to text before finally being converted into a date.

Renaming Columns

Next, Some of the columns were renamed to more appropriate names like Country Name to Country, Time to Year, and Time Code to Year Code.

Grouping, Hierarchies, or Binning

The original dataset was duplicated, the 'Country' and 'Country Code' columns were selected and the duplicate rows were removed. Then, a conditional column to group all the countries into African or Developed countries was added.

DAX codes created

A calendar table was created from scratch, and some new measures from existing columns using a DAX code. This is to have a very varied diversity of filter and customisation functions. Then a one-to-many relationship between the created calendar table and the main facts data table. A relationship was established by dragging the date (from the calendar) and placing it on the year column in the main data table.

3.3 Relational nature of attributes

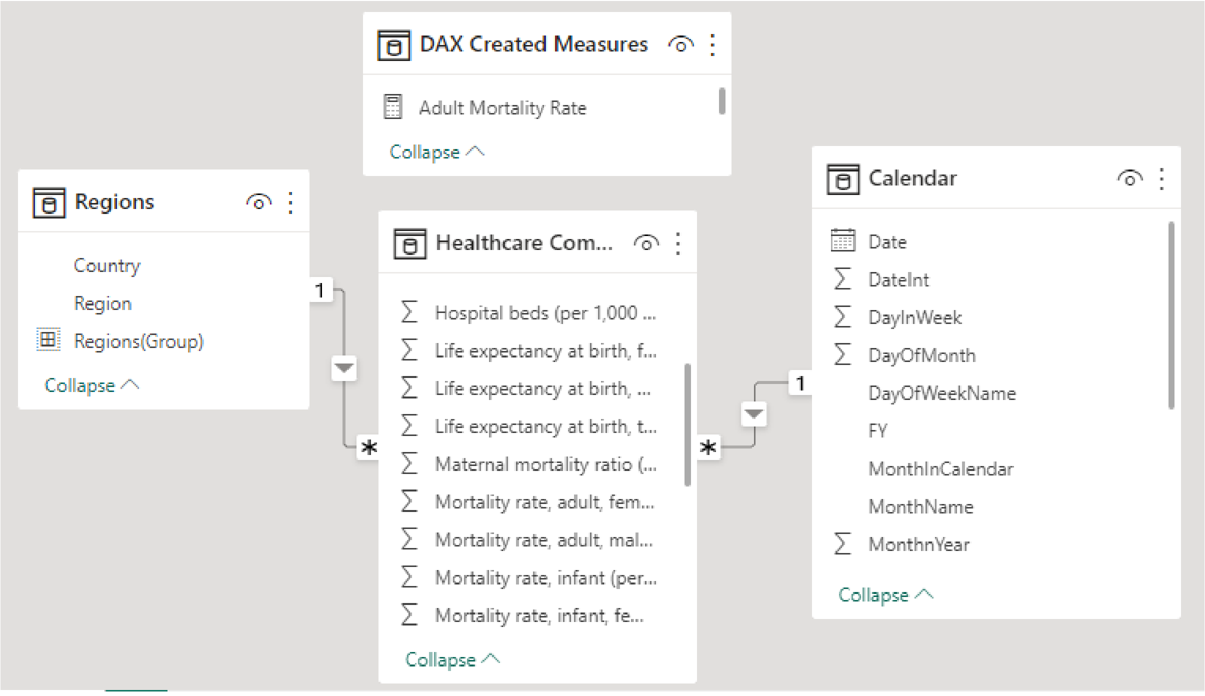
The attributes were selected such that there are different groups of attributes in the dataset. The groups include birth & death rate, government spending, infrastructure, life expectancy, mortality, personnel, and population. These are all connected using the country column.

A 'calendar' and 'Regions' table were also created using DAX codes and were connected to the main table using Date-to-Year one-to-many connection and a country-to-country one-to-many connection respectively.

3.4 Usage of attributes in dashboard

This report utilizes the use of 4 tables, 3 of which were created. The created tables were for the DAX-created measures, Regions, and the calendar tables. The fourth table was the main data table imported.

They were all connected using the star topology, with the data table as the centre, and other tables connected:


DAX Created Measures
Adult Mortality Rate
Collapse
Regions
Country
Region
Regions(Group)
Collapse
1
*
Healthcare Com...
Hospital beds (per 1,000 ...
Life expectancy at birth, f...
Life expectancy at birth, ...
Life expectancy at birth, t...
Maternal mortality ratio (...
Mortality rate, adult, fem...
Mortality rate, adult, mal...
Mortality rate, infant (per...
Mortality rate, infant, fe...
Collapse
*
1
Calendar
Date
DateInt
DayInWeek
DayOfMonth
DayOfWeekName
FY
MonthInCalendar
MonthName
MonthnYear
Collapse

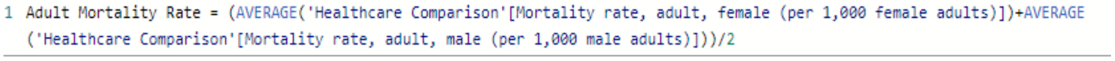
4. Investigation Of Data Workflow & Proposal For Design Of Dashboard

The first thing that was done after the initial data was loaded and preprocessed was the use of DAX codes to create extra measures both from scratch and from already existing attributes.

4.1 DAX created measures/Table

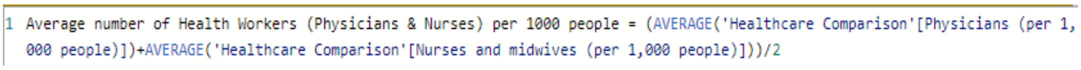
The following columns were created from already existing columns and put in a table called ‘DAX Created Measures’:

- The ‘adult mortality rate’ was calculated as the average between the ‘mortality rate, adult, female (per 1000 female adults)’ and the ‘mortality rate, adult, male (per 1000 male adults)’.



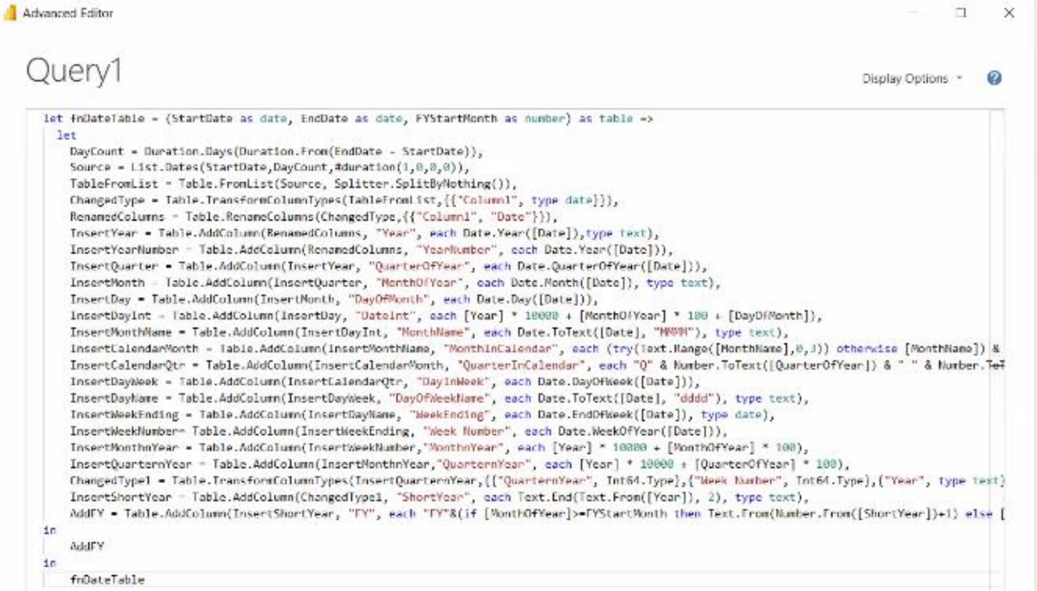
```
Adult Mortality Rate = (AVERAGE('Healthcare Comparison'[Mortality rate, adult, female (per 1,000 female adults)])+AVERAGE
('Healthcare Comparison'[Mortality rate, adult, male (per 1,000 male adults)]))/2
```

- The ‘Average number of Health Workers (Physicians & Nurses) per 1000 people’ was also calculated as the average between the ‘Physicians (per 1000 people) and the ‘Nurses and midwives (per 1000 people)’.



```
Average number of Health Workers (Physicians & Nurses) per 1000 people = (AVERAGE('Healthcare Comparison'[Physicians (per 1,
000 people)])+AVERAGE('Healthcare Comparison'[Nurses and midwives (per 1,000 people)]))/2
```

Next, the calendar table was created from scratch using the code below, and was made as a table:



Query1

```
let fnDateTable = (StartDate as date, EndDate as date, FYStartMonth as number) as table =>
  let
    DayCount = Duration.Days(Duration.From(EndDate - StartDate)),
    Source = List.Dates(StartDate,DayCount,#duration(1,0,0,0)),
    TableFromList = Table.FromList(Source, Splitter.SplitByNothing()),
    ChangedType = Table.TransformColumnTypes(TableFromList,{{"Column1", type date}}),
    RenamedColumns = Table.RenameColumns(ChangedType,{{"Column1", "Date"}}),
    InsertYear = Table.AddColumn(RenamedColumns, "Year", each Date.Year([Date]),type text),
    InsertYearNumber = Table.AddColumn(RenamedColumns, "YearNumber", each Date.Year([Date])),
    InsertQuarter = Table.AddColumn(InsertYear, "QuarterOfYear", each Date.QuarterOfYear([Date])),
    InsertMonth = Table.AddColumn(InsertQuarter, "MonthOfYear", each Date.Month([Date]), type text),
    InsertDay = Table.AddColumn(InsertMonth, "DayOfMonth", each Date.Day([Date])),
    InsertDayInt = Table.AddColumn(InsertDay, "DateInt", each [Year] * 10000 + [MonthOfYear] * 100 + [DayOfMonth]),
    InsertMonthName = Table.AddColumn(InsertDayInt, "MonthName", each Date.ToText([Date], "MMMM"), type text),
    InsertCalendarMonth = Table.AddColumn(InsertMonthName, "MonthInCalendar", each (try(Text.Range([MonthName],0,3)) otherwise [MonthName]) &
    InsertCalendarQtr = Table.AddColumn(InsertCalendarMonth, "QuarterInCalendar", each "Q" & Number.ToText([QuarterOfYear]) & " " & Number.ToT
    InsertDayWeek = Table.AddColumn(InsertCalendarQtr, "DayInWeek", each Date.DayOfWeek([Date])),
    InsertDayName = Table.AddColumn(InsertDayWeek, "DayOfWeekName", each Date.ToText([Date], "dddd"), type text),
    InsertWeekEnding = Table.AddColumn(InsertDayName, "WeekEnding", each Date.EndOfWeek([Date]), type date),
    InsertWeekNumber= Table.AddColumn(InsertWeekEnding, "Week Number", each Date.WeekOfYear([Date])),
    InsertMonthnYear = Table.AddColumn(InsertWeekNumber,"MonthnYear", each [Year] * 10000 + [MonthOfYear] * 100),
    InsertQuarternYear = Table.AddColumn(InsertMonthnYear,"QuarternYear", each [Year] * 10000 + [QuarterOfYear] * 100),
    ChangedType1 = Table.TransformColumnTypes(InsertQuarternYear,{{"QuarternYear", Int64.Type},{"Week Number", Int64.Type},{"Year", type text}
    InsertShortYear = Table.AddColumn(ChangedType1, "ShortYear", each Text.End(Text.From([Year]), 2), type text),
    AddFY = Table.AddColumn(InsertShortYear, "FY", each "FY"&(if [MonthOfYear]>=FYStartMonth then Text.From(Number.From([ShortYear])+1) else [
in
    AddFY
in
    fnDateTable
```

Then, all the tables were connected in a star-topology and the relationship model was built as shown in the previous section.

4.2 Building the dashboard

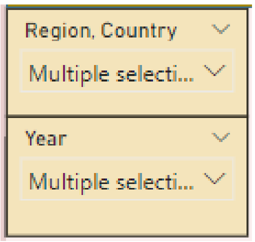
The main dashboard seeks to provide insights into the disparity between the healthcare systems of African and Developed countries. This is to be achieved systematically by carrying out a visual analysis of individual pairings (or groups) of the economic indicators that make up the healthcare system of the country.

The dashboard consists of 2 rows of visuals, and these would be picked out and explained individually from left to right, starting from the top row and then to the bottom row in the following paragraphs.

Visual 1: Slicer

This is the visual that selects the indicators to be displayed across all the other visuals in the dashboard. The countries were divided into 2 groups and this is reflected in the slicer visual as well. The slicer comprises 2 sections, one for the regions (group of countries) and another for the years. These were set in such a way that multiple countries and years can be selected with just a single click.

This was done to enable easier selection of data based on countries or periods to focus on.



Region, Country

Multiple selecti...

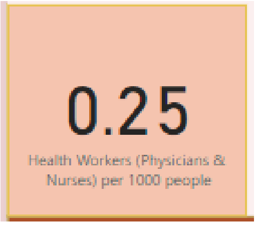
Year

Multiple selecti...

Visual 2: Card

This card shows the number of health workers per 1000 people, aggregated by mean. This indicator gives an idea of the healthcare personnel count of a country, which is an indicator that contributes to assessing the strength of that county's healthcare system. This is a result of the fact that the higher the number of physicians & nurses in a country, the more access there is to healthcare provision, and the better the overall healthcare system is. (Aday & Andersen, 2021).

The card was selected as the visual of choice in this instance because its simplicity helps to keep the dashboard organized, and more importantly because the data to be visualised is a single value number.



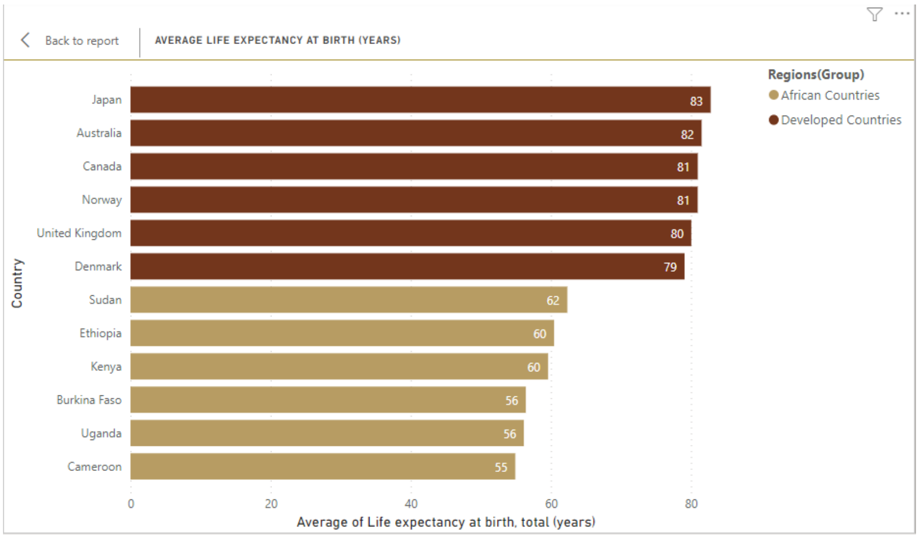
0.25

Health Workers (Physicians & Nurses) per 1000 people

Visual 3: Average life expectancy at birth (Years)

This visual is a bar chart (aligned horizontally) to show a comparison between the life expectancies of each country (or group of countries). One of the most contributory economic indicators of the state of the healthcare system of a country is life expectancy (Stiefel et al., 2010). Therefore, the horizontal bar chart was employed as the ideal visual to compare the life expectancies of 2 countries, any combination of countries, or the groups of countries.

This was done because the horizontal bar chart offers a clear representation of a graph of a categorical variable and a corresponding numerical variable. (Croxton & Stein, 1932)



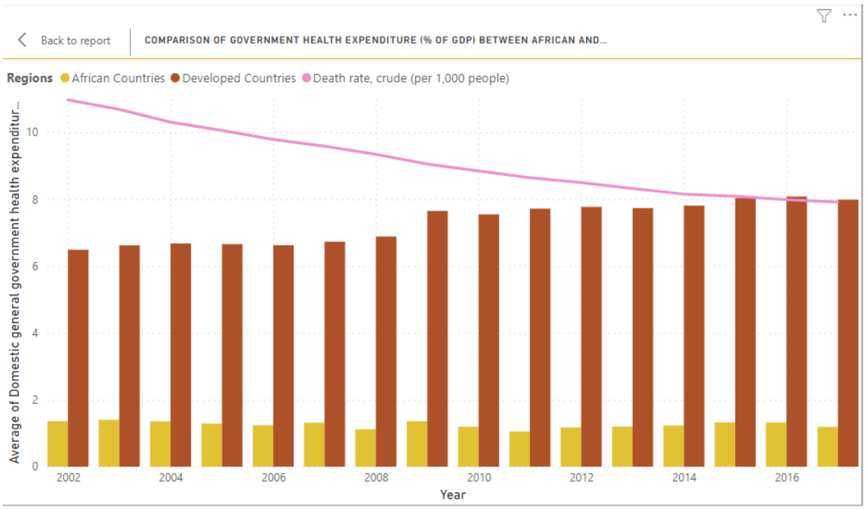
Visual 4: Comparison of government health expenditure (% of GDP) between African and Developed countries showing the death rate per 1000 people.

This visual is a line and clustered column chart that visually portrays the following:

- A comparison between the domestic government health expenditures of the African and Developed countries, and how they changed over the years in focus.
- A trendline showing the death rate per 1000 people in the period in focus.

This is important to the entire scope of this research because previous research (Berger & Messer, 2002) has proven that there exists a relationship between government health expenditure and the death rate in a country.

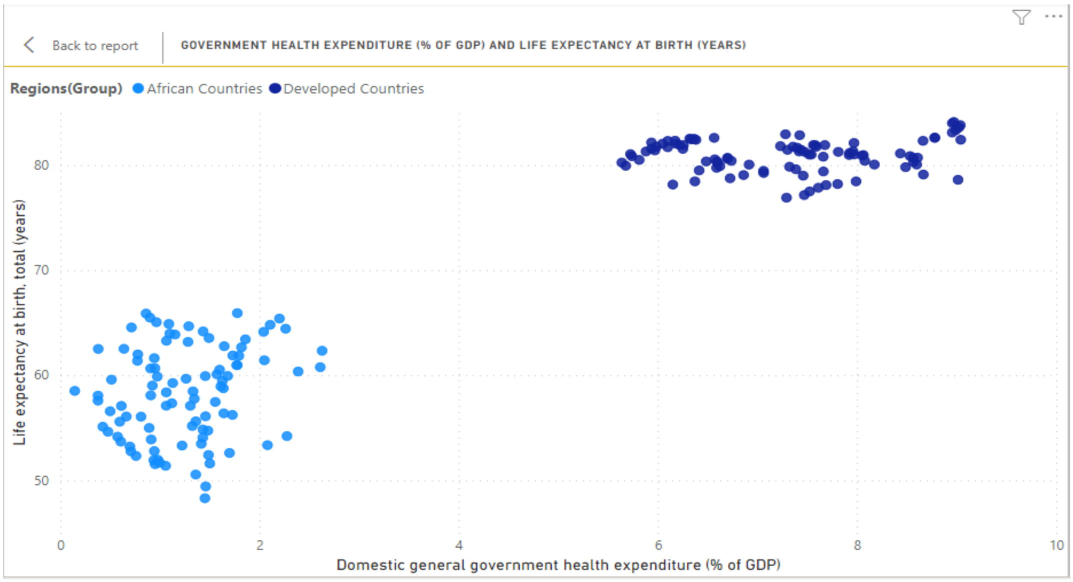
This visual is ideal for this analysis because it helps to explore these economic indicators separately and then compare them in the same graph.



Visual 5: Government health expenditure (% of GDP) and life expectancy at birth (years)

This visual was included to analyse and portray the relationship between the government health expenditure and the expected life expectancy at birth. It is important in the data workflow because a critical part of the research is an investigation into which indicators contribute to the disparity (seen in visual 3 above) between the 2 groups of countries, and the life expectancy (which in previous research works has been proven to be a suitable measure for assessing a country's healthcare system (Stiefel et al., 2010))

Government health expenditure has been proven to have a strong impact on the healthcare system of a country and by extension, its life expectancy (Deshpande et al., 2014). Knowledge of the relationship between these 2 variables can help to effect change in the system.

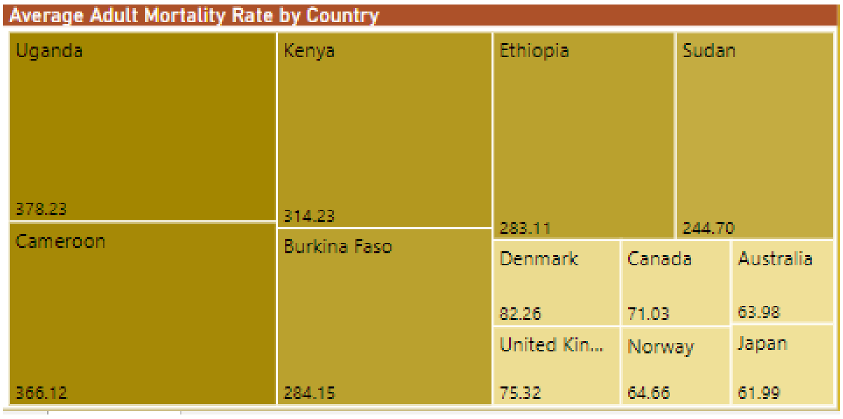


Visual 6: Average adult mortality rate by country

This is a treemap showing the distribution of the adult mortality rates of the countries selected. The treemap was chosen for this visual because it is the ideal choice for arranging and displaying the rankings and the distribution of the economic indicator to be focused on.

The adult mortality rate is an important indicator in this study because it is a good measure of the lack of adequate access to good health care which in turn is a means to assess the healthcare system of a country.

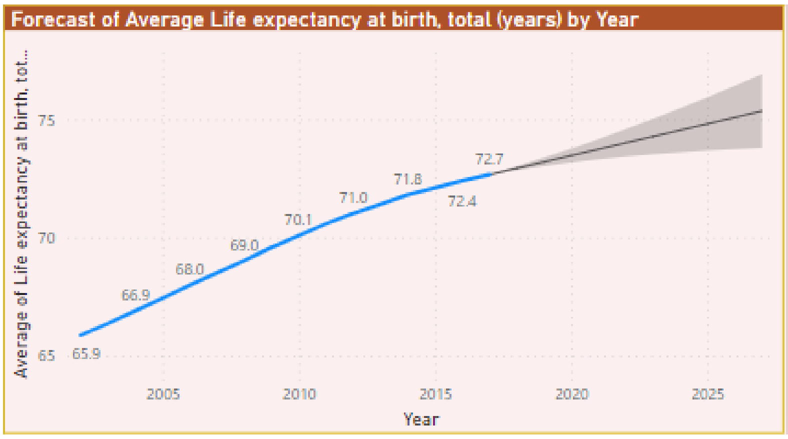
This simply means that a country with a below-par healthcare system is likely to have a higher adult mortality rate than countries with adequate healthcare systems. (Owusu et al., 2021)



Visual 7: Forecast of Average life expectancy at birth by Year

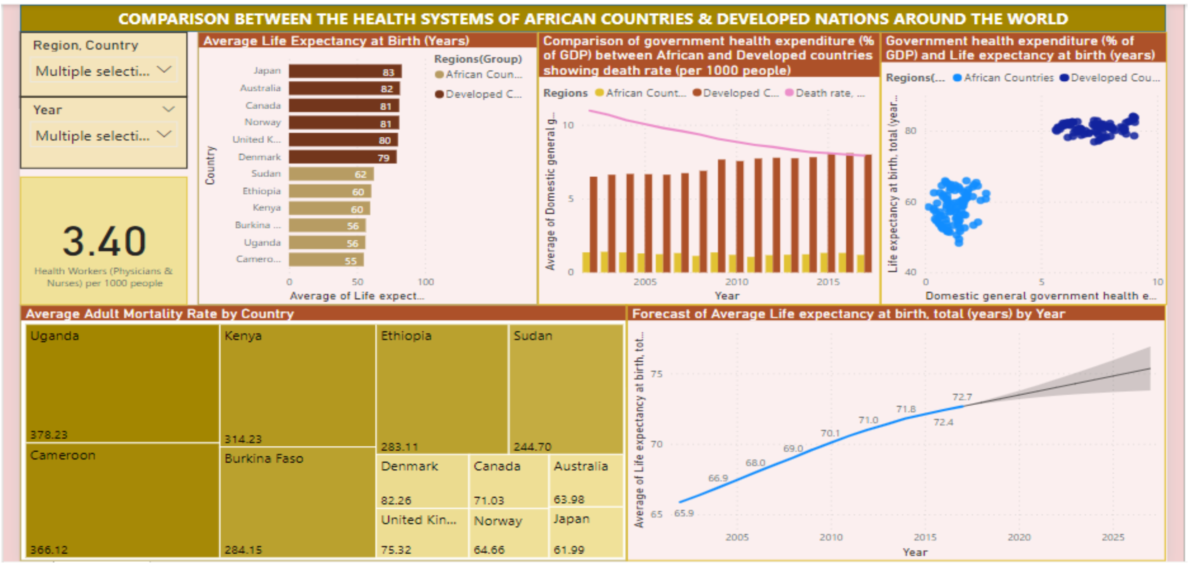
No visual is more appropriate in the visualization of trends over a stated period than the line chart. This is why it was chosen as the chart for the display of the trend that the life expectancy variable has taken over the chosen number of years in focus. It aggregates this economic indicator by getting a mean of the values of the countries selected using the slicer tool (visual 1 as explained earlier).

In this study, apart from showing the trend of life expectancy over the years, it was also used to plot a forecast for the next 10 years. Armed with this projection, other factors that have been shown to contribute to this indicator can be tweaked to further improve the actual values of life expectancies (Lee & Carter, 1992).



These visuals collectively make up the single-screen dashboard and show how the indicators and economic variables selected have been used in this study in collaboration with past research referenced to give insights into the healthcare systems of the individual and/or group of countries specified.

A complete representation of the single-screen dashboard in its entirety is shown in the image below:


COMPARISON BETWEEN THE HEALTH SYSTEMS OF AFRICAN COUNTRIES & DEVELOPED NATIONS AROUND THE WORLD
Region, Country
Multiple selecti...
Year
Multiple selecti...
3.40
Health Workers (Physicians & Nurses) per 1000 people
Average Life Expectancy at Birth (Years)
Regions(Group)
African Coun...
Developed C...
Country
Japan 83
Australia 82
Canada 81
Norway 81
United K... 80
Denmark 79
Sudan 62
Ethiopia 60
Kenya 60
Burkina ... 56
Uganda 56
Camero... 55
0
50
100
Average of Life expect...
Comparison of government health expenditure (% of GDP) between African and Developed countries showing death rate (per 1000 people)
Regions
African Count...
Developed C...
Death rate, ...
Average of Domestic general g...
10
5
0
2005
2010
2015
Year
Government health expenditure (% of GDP) and Life expectancy at birth (years)
Regions(...
African Countries
Developed Cou...
Life expectancy at birth, total (year...
80
60
40
0
5
10
Domestic general government health e...
Average Adult Mortality Rate by Country
Uganda 378.23
Kenya 314.23
Ethiopia 283.11
Sudan 244.70
Cameroon 366.12
Burkina Faso 284.15
Denmark 82.26
Canada 71.03
Australia 63.98
United Kin... 75.32
Norway 64.66
Japan 61.99
Forecast of Average Life expectancy at birth, total (years) by Year
Average of Life expectancy at birth, tot...
75
70
65
65.9
66.9
68.0
69.0
70.1
71.0
71.8
72.4
72.7
2005
2010
2015
2020
2025
Year

5. Discussion

This study sought to explore the disparity between the healthcare systems of 12 countries split into two groups, explore the relationships between some of the chosen indicators, show how the health indicators of these sets of countries have performed over the years, and then give a forecast of how they'll perform in coming years.

The card (visual 2) shows a clear number of the healthcare personnel in the country or group of countries chosen in the slicer (visual 1). This number gives an immediate idea of how stretched the health personnel is.

The result of visual 3 shows that the developed countries over the years have a higher life expectancy at birth than the African countries with Japan having the overall highest life expectancy at 83 years, while Cameroon was the lowest at 55 years. These numbers show a big difference as the average of both groups of countries has a numerical difference of 22.83 years. Visual 4 exposes the increasing trend in the government health expenditure of the developed group as opposed to the much lower yet slightly decreasing amounts spent by the African countries on the healthcare systems, which could be a contributing factor to the low life expectancy in the African cohorts.

The government health expenditure is then shown in visual 5 to have a strong positive relationship with life expectancy indicating that as the government health expenditure of a group is increasing over the years, the life expectancy also increases. The visual shows the African group displaying a slower rate of increase than their developed cohorts.

An exploration into the adult mortality rate in visual 6 shows that the African countries have the higher mortality rates with Uganda leading the pack with 378.23 per 1000 people and the lowest being Burkina Faso at 284.15 per 1000 people. In contrast, Denmark has the highest mortality rate among the developed countries at 82.26 per 1000 people, with Japan being the least at 61.99 per 1000 people.

These relatively much higher mortality rates in the African group can be, on some level, attributed to the government's lack of sufficient allocation of GDP percentage to the healthcare system of the individual countries, and even though there is not enough evidence to conclude that low GDP allocation is the reason for the low life expectancy and high mortality rates of the African countries, the visuals show that low GDP healthcare percentages are major contributors.

The last visual (visual 7) shows the life expectancy of any selected country or the average life expectancy in any combination of countries chosen. It also gives a 10-year forecast and the greyed-out parts are the confidence levels of the said forecast.

In a simplified summary of the results, it can be stated that the developed countries have a much better healthcare system than their African cohorts, clearly reflected in the life expectancy of their citizens. This life expectancy also shows a positive relationship with government health expenditure. The low government spending by African countries reflects macabrely in the high mortality rates of its people.

6. Conclusion

This study aimed to compare the healthcare systems of the African group of countries to that of their developed counterparts and also proffer solutions to bridge the gap between the groups.

Results from this study conducted through the dashboard show that there is a relationship between government health expenditure and the state of the healthcare of a country and by extension, the life expectancy.

The final solution proposed to reduce the just-discovered gulf between the healthcare systems of both groups of countries is for the government of the African countries to increase the percentage of the country’s GDP being allocated to healthcare services. This would serve not only to increase the life expectancy in the country but also to reduce the mortality rates in the country as well.

With this solution, it can be said that the objectives of this study have been very successfully met.

With the use of this dashboard designed, the reports and results have been written in such a way that the wider public who might not have in-depth analytic knowledge can use the dashboard and come up with more data and information regarding the healthcare systems of the selected country or group of countries.

TASK 2: Statistical Analysis

1. Introduction

Healthcare systems are designed by countries to provide the optimum health services to their citizens, a lot of advancement has been made, from the ease of access to emergency services to the minimization of the cost of care for the vulnerable population. Some countries have made major strides in optimizing their healthcare systems while others are still some ways behind. (Burgers et al., 2014)

This part of the report seeks to use different statistical methods to analyse the same dataset used in task 1 above. It explores the economic indicators which tell the overall state of the healthcare systems of the 12 countries over 16 years.

The main objective of this research is to compare the healthcare systems of the 2 groups of countries, the 'African' and 'Developed' groups, which are made up of the 12 selected countries, by getting the relationship between other indicators and life expectancy (which is regarded as an ideal indicator that gives a measure of the state of the health system).

In various previous research (Asiskovitch, 2010), life expectancy has been said to be a prime tool in the estimation, both locally and internationally, of the health system and general population health of a country.

The secondary objective is to show that the birth and death rates of the countries are not equal.

The strategy to be followed in the course of this research is to carry out descriptive statistical analysis, correlation, regression, and time-series analysis on some select indicators in the dataset. This would be done to explore the data, infer & examine the relationships between indicators, and understand and predict the behavior of the indicators.

2. Background Research

2.1 Brief Description of Theoretical Backgrounds

2.1.1 Correlation Analysis

This is a method in statistics that explores the relationship between 2 datasets or variables.

Statistical correlation can be classified using three major coefficients used, these include:

i. Pearson's Coefficient: Regarded as the most popular method, this method assesses how weak/strong the linear relationship between variables is. It is used when the data is quantitative and parametric. The formula is

$$r = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma(X-\overline{X})^2}\sqrt{(Y-\overline{Y})^2}}$$

Where, $\overline{X}$ = mean of X variable

$\overline{Y}$ = mean of Y variable

ii. Spearman's Rank Correlation Coefficient: This is used to assess the significance of the relationship between variables/datasets. It is utilised when the data is of ordinal type and shows ranking, and is qualitative and non-parametric. The formula is

$$r_s = 1 - \frac{6\sum D^2}{n\left(n^2-1\right)}$$

Where D = difference between ranks

n = number of observations

iii. Kendall's Rank Correlation Coefficient: This is used to measure how dependent 2 non-parametric variables are between themselves. The formula is:

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}$$

Where n_c = number of concordant

And n_d = number of discordant

n = sample size

2.1.2 Regression Analysis

This is a deeper measurement and examination of the relationship between more than 1 variable (2 or more) in a dataset. It calculates the significance of various combinations of independent variables and their impact on the dependent variable.

To perform this analysis, the dependent and independent variables need to be defined before the relationships between them can be plotted.

This analysis further computes the formula (to a certain degree of error) with which a combination of the independent variables can be used to achieve the dependent variable. It is majorly divided into linear and non-linear regression.

i. Linear regression: This works on the principle that the dependent variable can be derived from a linear combination of the independent variables.
ii. Non-Linear regression: This is a model in which the function is not a linear combination of the independent variables. It utilises an iterative process to reduce the errors generated.

2.1.3 Time Series Analysis

This is a method used to analyse data points that have been collected over time and usually at consistent intervals. The results from time series can be of different types which include descriptive analysis, segmentation, forecasting, intervention analysis, etc.

The time series has 4 components:

i. Cyclical: This is the up-and-down movement pattern in a time cycle.
ii. Trend: This refers to a general outlook over the entire period to know if there is an increasing, decreasing, or consistent movement of the variable over time.
iii. Seasonality: These are rhythmic patterns that occur repeatedly over the same period throughout the data.
iv. Irregularity: These are variations that cannot be controlled and are generally unavoidable.

Time series data can be stationary and non-stationary.

There are three major classes of models used in time series analysis. These include:

a. Autoregressive (AR): This method works on the principle that the output variable can be derived linearly from its previous values
b. Moving Average (MA): This model is generally seen as a linear regression of the series' current values plotted against past and presently occurring white noise errors.
c. Integrated model (I): This mostly refers to a consistent repetition of a particular attribute of a model to improve accuracy.

A combination of the above classes yields further models such as ARMA, and ARIMA.

2.2 Literature Reviews

2.2.1 Literature review 1

In a conference article, (Pacáková et al., 2016) take a look at the comparisons of healthcare results in public health systems of European countries. The work focuses on presenting the results obtained from the use of statistical methods such as correlation analysis, factor analysis, cluster analysis, and regression analysis to examine the differences in the public health systems of the various selected countries.

The data consists of indicators that were selected based on factors ranging from the country’s health expenditure to the available healthcare resources.

Statistical theoretical background investigation

- *Correlation*

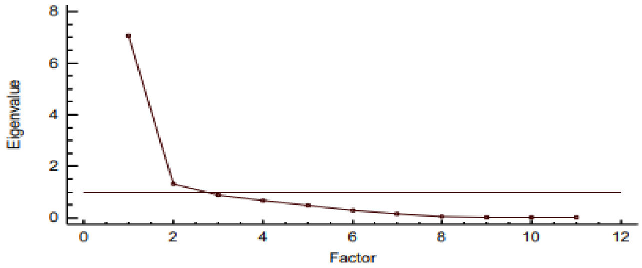
In this work, during the investigation of the relationship between the selected indicators, since the work is tilted more towards ranking, the Spearman coefficient was chosen over the Pearson correlations because the data value ranks are used in the calculations as opposed to the use of the actual values in the Pearson method. (Sokolowski, 1999)

- *Multidimensional Comparative Analysis*

Finally, a multidimensional comparative analysis was implemented and an aggregate scoring of the performance of healthcare systems in each country was gotten. The score was sorted in an order such that the highest-performing country was top and the lowest-performing country was at the bottom.

The methods used were effective in showing the results visually, like the factor analysis that helped to reduce the number of causal factors from 11 to 2 using the Scree Plot.

Figure 1 Scree Plot



Source: Own calculation, output from Statgraphics Centurion XV

The results also determined that some of the health system variables had very little impact on the health status of some of the European countries and therefore suggest a subpar performance of the health systems.

2.2.2 Literature review 2

This journal article (Kaya Samut & Cafri, 2016), analyzes the hospital efficiency of 29 OECD (Organization for Economic Cooperation and Development) countries between 2000 and 2010. 5 input variables were selected ranging from available equipment to the number of health personnel (nurses and doctors).

Statistical theoretical background investigation

- *Regression Analysis*

The analysis in this work was carried out in several parts, but the part focused on is the Tobit Panel Model which is also called the Censored Regression Model.

Censored Regression Model is the type of analysis carried out when there is some restriction or censoring in the variance of the dependent variable. It estimates linear relationships amongst variables when this type of censoring occurs. (Long, 1997)

The regression model was able to establish a strong negative relationship between government spending on the health system and its efficiency.

2.2.3 Literature review 3

This work by (Jaba et al., 2014) sought to establish the relationship between the independent and dependent variables in the dataset of economic indicators that defines a healthcare system. The input variable is defined as the government expenditure while the life expectancy at birth was selected as the output.

Statistical theoretical background investigation

- *Time Series Analysis*

The fixed effects regression model was used in this research as it explains the variation between life expectancy and total health expenditure. All the countries showed significant positive relationships between the input and output variables, and the null hypothesis, which believes that the coefficient of the country binary variables equals 0, was rejected.

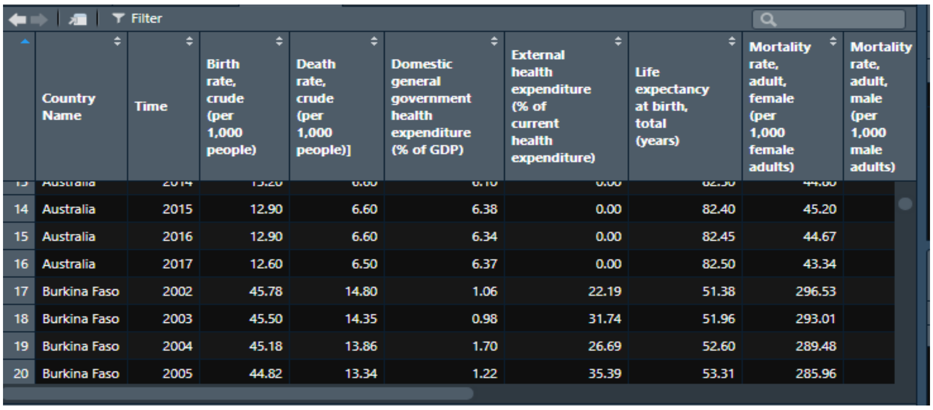
3. Exploration of Dataset

The dataset used was sourced from the World Bank Development Indicators Databank https://databank.worldbank.org/source/world-development-indicators

12 countries were selected, with 6 African countries and 6 developed countries. The data spans 16 years, 2002 – 2017, and contains 12 indicators for the healthcare system comparison of the 2 sets of countries.

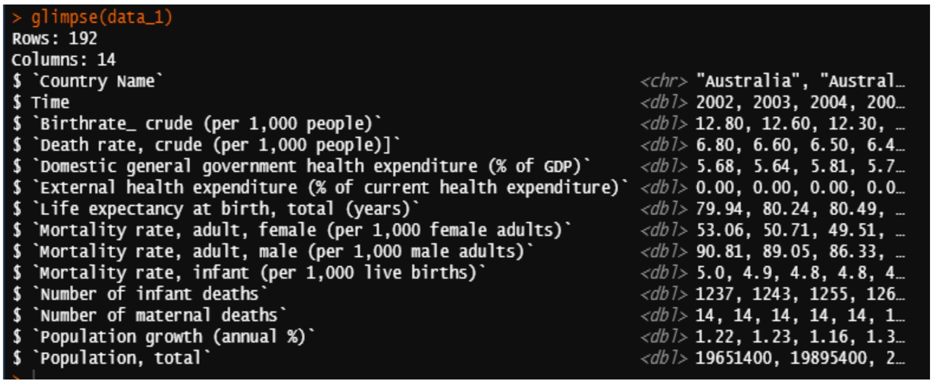
3.1 Pre-processing steps taken

First, the dataset was loaded into a variable called 'data_1' using the Tidyverse library, and a preview was done:



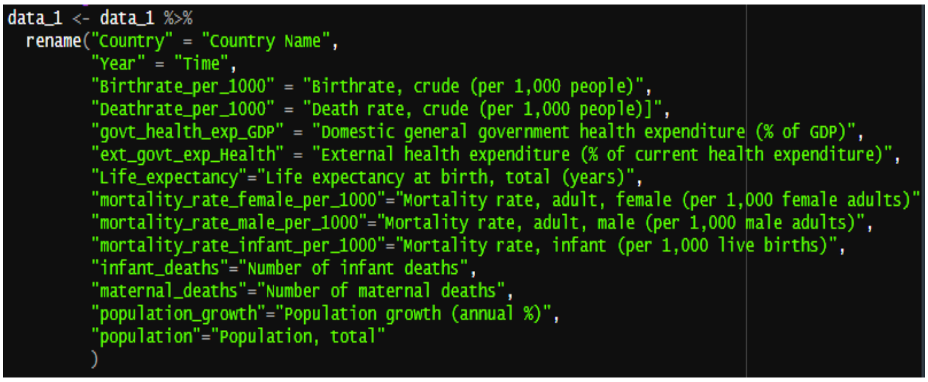
	Country Name	Time	Birth rate, crude (per 1,000 people)	Death rate, crude (per 1,000 people)]	Domestic general government health expenditure (% of GDP)	External health expenditure (% of current health expenditure)	Life expectancy at birth, total (years)	Mortality rate, adult, female (per 1,000 female adults)	Mortality rate, adult, male (per 1,000 male adults)
13	Australia	2014	13.20	6.60	6.10	0.00	82.30	44.00	
14	Australia	2015	12.90	6.60	6.38	0.00	82.40	45.20	
15	Australia	2016	12.90	6.60	6.34	0.00	82.45	44.67	
16	Australia	2017	12.60	6.50	6.37	0.00	82.50	43.34	
17	Burkina Faso	2002	45.78	14.80	1.06	22.19	51.38	296.53	
18	Burkina Faso	2003	45.50	14.35	0.98	31.74	51.96	293.01	
19	Burkina Faso	2004	45.18	13.86	1.70	26.69	52.60	289.48	
20	Burkina Faso	2005	44.82	13.34	1.22	35.39	53.31	285.96	

Then, a 'glimpse' function was used to show the indicators selected. See below:

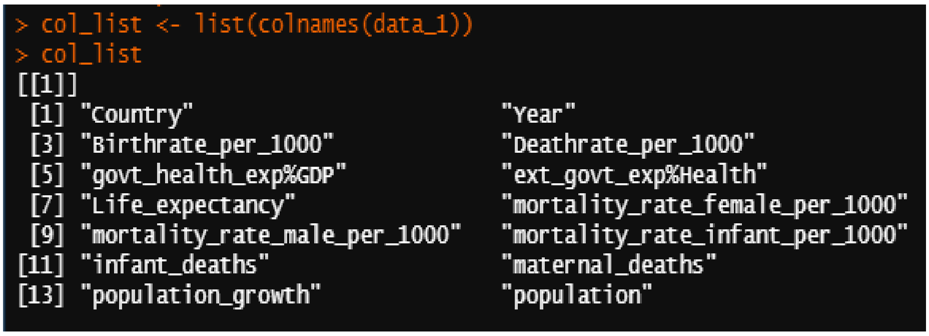


```
> glimpse(data_1)
Rows: 192
Columns: 14
$ `Country Name`                                              <chr> "Australia", "Austral…
$ Time                                                        <dbl> 2002, 2003, 2004, 200…
$ `Birthrate_ crude (per 1,000 people)`                       <dbl> 12.80, 12.60, 12.30, …
$ `Death rate, crude (per 1,000 people)]`                     <dbl> 6.80, 6.60, 6.50, 6.4…
$ `Domestic general government health expenditure (% of GDP)` <dbl> 5.68, 5.64, 5.81, 5.7…
$ `External health expenditure (% of current health expenditure)` <dbl> 0.00, 0.00, 0.00, 0.0…
$ `Life expectancy at birth, total (years)`                   <dbl> 79.94, 80.24, 80.49, …
$ `Mortality rate, adult, female (per 1,000 female adults)`   <dbl> 53.06, 50.71, 49.51, …
$ `Mortality rate, adult, male (per 1,000 male adults)`       <dbl> 90.81, 89.05, 86.33, …
$ `Mortality rate, infant (per 1,000 live births)`            <dbl> 5.0, 4.9, 4.8, 4.8, 4…
$ `Number of infant deaths`                                   <dbl> 1237, 1243, 1255, 126…
$ `Number of maternal deaths`                                 <dbl> 14, 14, 14, 14, 14, 1…
$ `Population growth (annual %)`                              <dbl> 1.22, 1.23, 1.16, 1.3…
$ `Population, total`                                         <dbl> 19651400, 19895400, 2…
```

The columns were renamed to make processing easier. Codes & results are below:

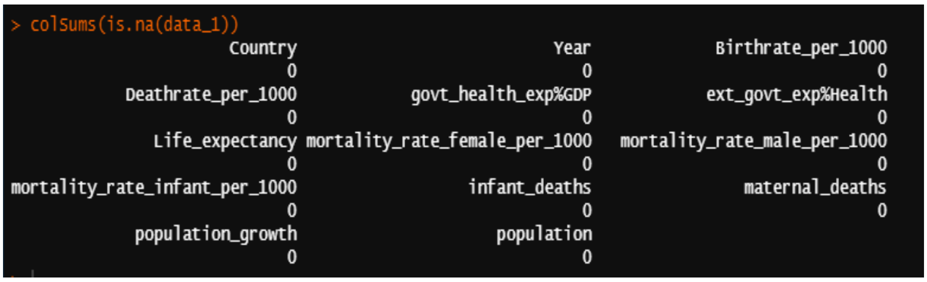


```
data_1 <- data_1 %>%
  rename("Country" = "Country Name",
         "Year" = "Time",
         "Birthrate_per_1000" = "Birthrate, crude (per 1,000 people)",
         "Deathrate_per_1000" = "Death rate, crude (per 1,000 people)]",
         "govt_health_exp_GDP" = "Domestic general government health expenditure (% of GDP)",
         "ext_govt_exp_Health" = "External health expenditure (% of current health expenditure)",
         "Life_expectancy"="Life expectancy at birth, total (years)",
         "mortality_rate_female_per_1000"="Mortality rate, adult, female (per 1,000 female adults)"
         "mortality_rate_male_per_1000"="Mortality rate, adult, male (per 1,000 male adults)",
         "mortality_rate_infant_per_1000"="Mortality rate, infant (per 1,000 live births)",
         "infant_deaths"="Number of infant deaths",
         "maternal_deaths"="Number of maternal deaths",
         "population_growth"="Population growth (annual %)",
         "population"="Population, total"
         )
```



```
> col_list <- list(colnames(data_1))
> col_list
[[1]]
 [1] "Country"                      "Year"
 [3] "Birthrate_per_1000"           "Deathrate_per_1000"
 [5] "govt_health_exp%GDP"          "ext_govt_exp%Health"
 [7] "Life_expectancy"              "mortality_rate_female_per_1000"
 [9] "mortality_rate_male_per_1000" "mortality_rate_infant_per_1000"
[11] "infant_deaths"                "maternal_deaths"
[13] "population_growth"            "population"
```

Next, the dataset was checked for missing values:



```
> colSums(is.na(data_1))
                       Country                           Year             Birthrate_per_1000
                             0                              0                              0
            Deathrate_per_1000            govt_health_exp%GDP            ext_govt_exp%Health
                             0                              0                              0
               Life_expectancy mortality_rate_female_per_1000   mortality_rate_male_per_1000
                             0                              0                              0
mortality_rate_infant_per_1000                  infant_deaths                maternal_deaths
                             0                              0                              0
             population_growth                     population
                             0                              0
```

3.2 Tables showing list of countries:

Countries	Group Assigned
Australia	Developed
Burkina Faso	African
Cameroon	African
Canada	Developed
Denmark	Developed
Ethiopia	African

Countries	Group Assigned
Japan	Developed
Kenya	African
Norway	Developed
Sudan	African
Uganda	African
United Kingdom	Developed

3.3 Table showing list of indicators:

Indicators	Shortened versions	Short descriptions
Birthrate, crude (per 1,000 people)	Birthrate_per_1000	Crude birth rate indicates the number of live births per 1,000 midyear population.
Death rate, crude (per 1,000 people)	Deathrate_per_1000	Crude death rate indicates the number of deaths per 1,000 midyear population.
Domestic general government health expenditure (% of GDP)	govt_health_exp%GDP	Public expenditure on health from domestic sources as measured by GDP.
External health expenditure (% of current health expenditure)	ext_govt_exp%Health	Share of current health expenditures funded from external sources
Life expectancy at birth, total (years)	Life_expectancy	Number of years a newborn infant would live if prevailing

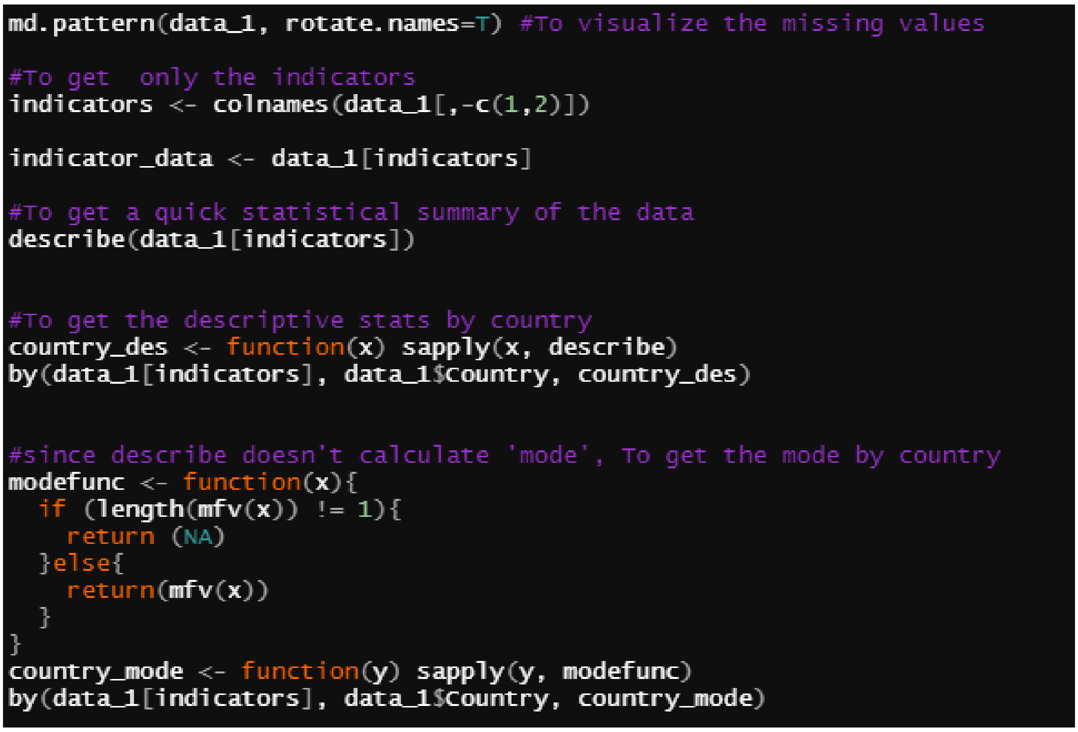
		patterns of mortality at birth were to stay the same throughout
Mortality rate, adult, female (per 1,000 female adults)	mortality_rate_female_per_1000	The probability of females dying between the ages of 15 and 60
Mortality rate, adult, male (per 1,000 male adults)	mortality_rate_male_per_1000	The probability of males dying between the ages of 15 and 60
Mortality rate, infant (per 1,000 live births)	mortality_rate_infant_per_1000	Number of infants dying before reaching one year of age, per 1,000 live births in a given year.
Number of infant deaths	infant_deaths	Number of infants dying before reaching one year of age
Number of maternal deaths	maternal_deaths	Number of deaths of women who are pregnant or within 42 days of termination of pregnancy
Population growth (annual %)	population_growth	Change in population expressed as a percentage
Population, total	population	Total number of people

4. Analysis

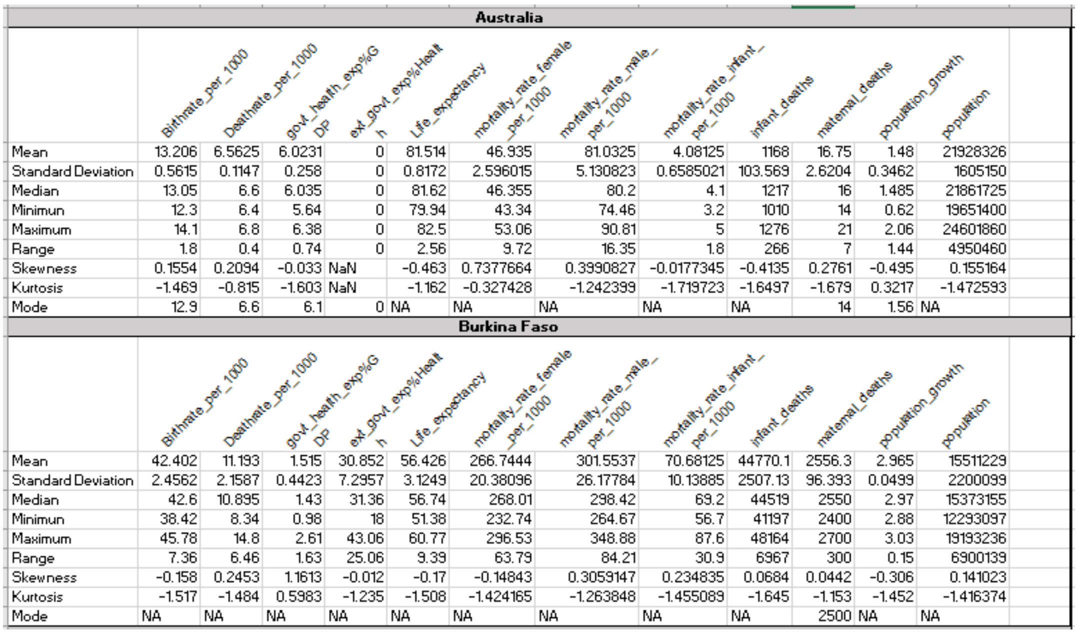
4.1 Descriptive Statistics

For the sake of clarity, the results of the descriptive statistics are presented in a tabular format.

The ‘mode’ isn’t displayed with the ‘describe’ function, so a separate function (using modeest) was written to compute the mode. The codes and tabulated output respectively are shown below:



```
md.pattern(data_1, rotate.names=T) #To visualize the missing values

#To get  only the indicators
indicators <- colnames(data_1[,-c(1,2)])

indicator_data <- data_1[indicators]

#To get a quick statistical summary of the data
describe(data_1[indicators])


#To get the descriptive stats by country
country_des <- function(x) sapply(x, describe)
by(data_1[indicators], data_1$Country, country_des)


#since describe doesn't calculate 'mode', To get the mode by country
modefunc <- function(x){
  if (length(mfv(x)) != 1){
    return (NA)
  }else{
    return(mfv(x))
  }
}
country_mode <- function(y) sapply(y, modefunc)
by(data_1[indicators], data_1$Country, country_mode)
```

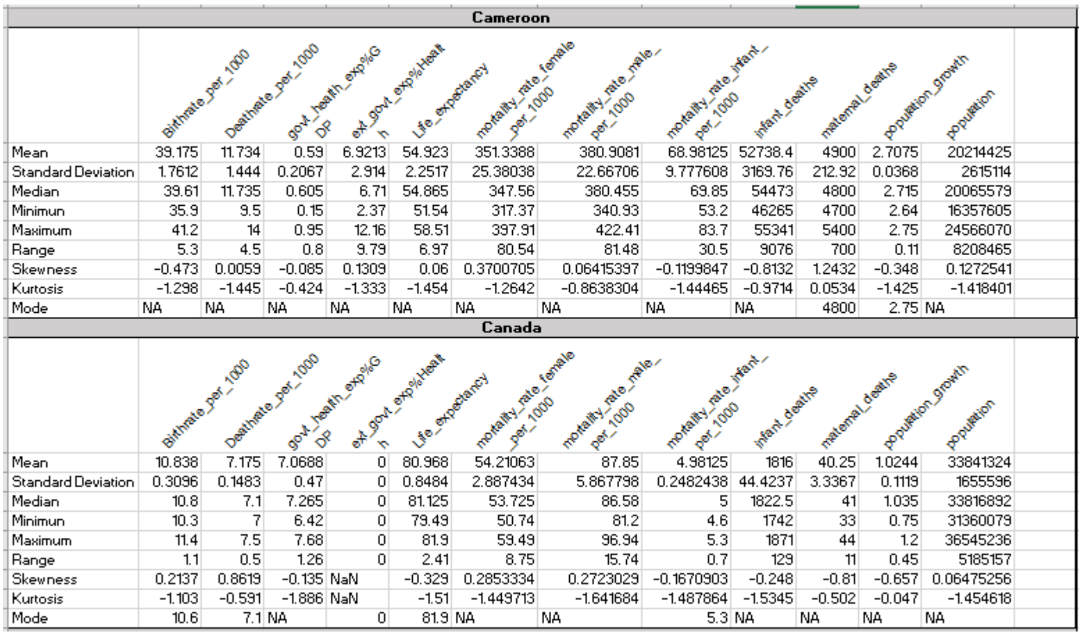


Australia

	Birthrate_per_1000	Deathrate_per_1000	govt_health_exp%GDP	ext_govt_exp%Health	Life_expectancy	mortality_rate_female_per_1000	mortality_rate_male_per_1000	mortality_rate_infant_per_1000	infant_deaths	maternal_deaths	population_growth	population
Mean	13.206	6.5625	6.0231	0	81.514	46.935	81.0325	4.08125	1168	16.75	1.48	21928326
Standard Deviation	0.5615	0.1147	0.258	0	0.8172	2.596015	5.130823	0.6585021	103.569	2.6204	0.3462	1605150
Median	13.05	6.6	6.035	0	81.62	46.355	80.2	4.1	1217	16	1.485	21861725
Minimun	12.3	6.4	5.64	0	79.94	43.34	74.46	3.2	1010	14	0.62	19651400
Maximum	14.1	6.8	6.38	0	82.5	53.06	90.81	5	1276	21	2.06	24601860
Range	1.8	0.4	0.74	0	2.56	9.72	16.35	1.8	266	7	1.44	4950460
Skewness	0.1554	0.2094	-0.033	NaN	-0.463	0.7377664	0.3990827	-0.0177345	-0.4135	0.2761	-0.495	0.155164
Kurtosis	-1.469	-0.815	-1.603	NaN	-1.162	-0.327428	-1.242399	-1.719723	-1.6497	-1.679	0.3217	-1.472593
Mode	12.9	6.6	6.1	0	NA	NA	NA	NA	NA	14	1.56	NA

Burkina Faso

	Birthrate_per_1000	Deathrate_per_1000	govt_health_exp%GDP	ext_govt_exp%Health	Life_expectancy	mortality_rate_female_per_1000	mortality_rate_male_per_1000	mortality_rate_infant_per_1000	infant_deaths	maternal_deaths	population_growth	population
Mean	42.402	11.193	1.515	30.852	56.426	266.7444	301.5537	70.68125	44770.1	2556.3	2.965	15511229
Standard Deviation	2.4562	2.1587	0.4423	7.2957	3.1249	20.38096	26.17784	10.13885	2507.13	96.393	0.0499	2200099
Median	42.6	10.895	1.43	31.36	56.74	268.01	298.42	69.2	44519	2550	2.97	15373155
Minimun	38.42	8.34	0.98	18	51.38	232.74	264.67	56.7	41197	2400	2.88	12293097
Maximum	45.78	14.8	2.61	43.06	60.77	296.53	348.88	87.6	48164	2700	3.03	19193236
Range	7.36	6.46	1.63	25.06	9.39	63.79	84.21	30.9	6967	300	0.15	6900139
Skewness	-0.158	0.2453	1.1613	-0.012	-0.17	-0.14843	0.3059147	0.234835	0.0684	0.0442	-0.306	0.141023
Kurtosis	-1.517	-1.484	0.5983	-1.235	-1.508	-1.424165	-1.263848	-1.455089	-1.645	-1.153	-1.452	-1.416374
Mode	NA	NA	NA	NA	NA	NA	NA	NA	NA	2500	NA	NA

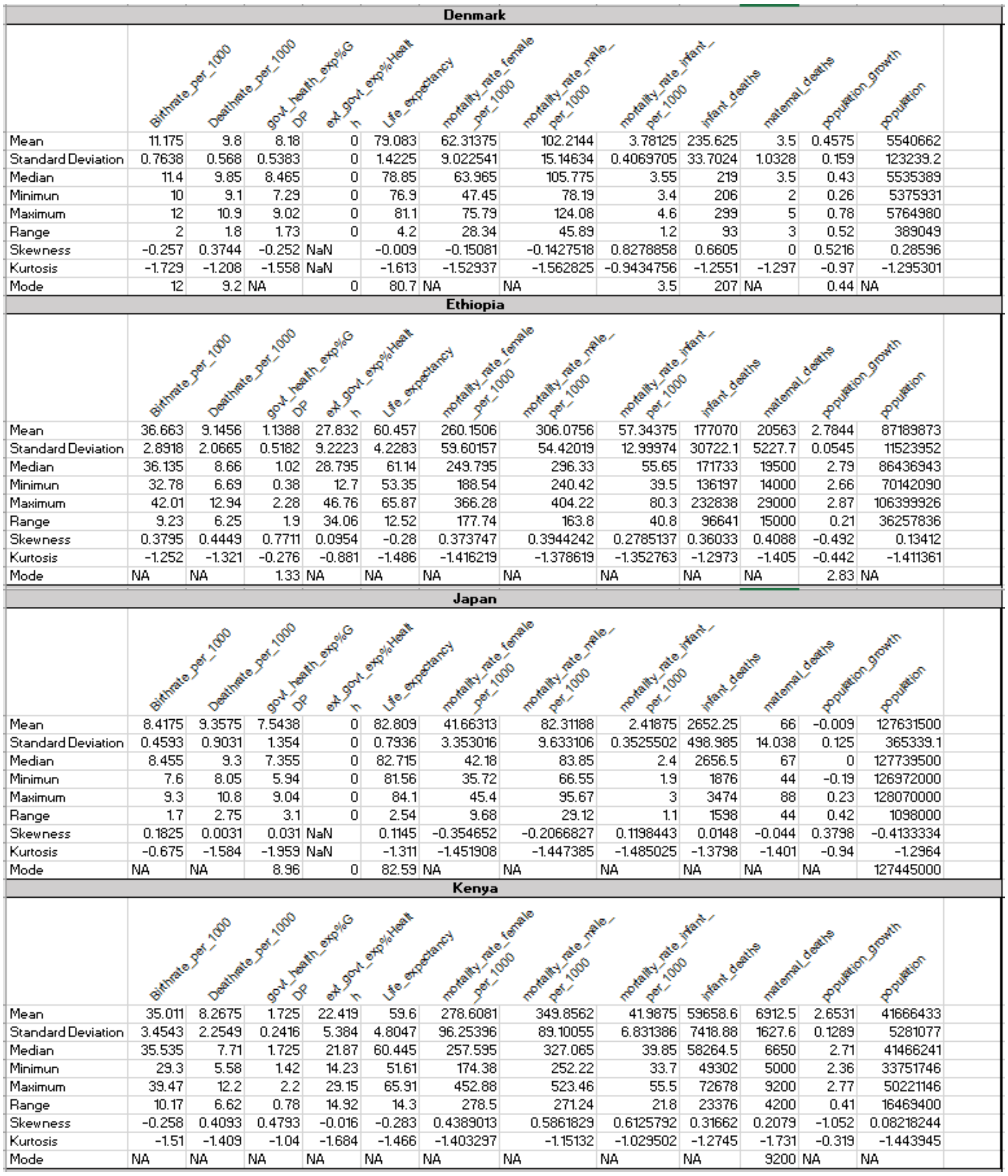


Cameroon

	Birthrate_per_1000	Deathrate_per_1000	govt_health_exp%GDP	ext_govt_exp%Health	Life_expectancy	mortality_rate_female_per_1000	mortality_rate_male_per_1000	mortality_rate_infant_per_1000	infant_deaths	maternal_deaths	population_growth	population
Mean	39.175	11.734	0.59	6.9213	54.923	351.3388	380.9081	68.98125	52738.4	4900	2.7075	20214425
Standard Deviation	1.7612	1.444	0.2067	2.914	2.2517	25.38038	22.66706	9.777608	3169.76	212.92	0.0368	2615114
Median	39.61	11.735	0.605	6.71	54.865	347.56	380.455	69.85	54473	4800	2.715	20065579
Minimun	35.9	9.5	0.15	2.37	51.54	317.37	340.93	53.2	46265	4700	2.64	16357605
Maximum	41.2	14	0.95	12.16	58.51	397.91	422.41	83.7	55341	5400	2.75	24566070
Range	5.3	4.5	0.8	9.79	6.97	80.54	81.48	30.5	9076	700	0.11	8208465
Skewness	-0.473	0.0059	-0.085	0.1309	0.06	0.3700705	0.06415397	-0.1199847	-0.8132	1.2432	-0.348	0.1272541
Kurtosis	-1.298	-1.445	-0.424	-1.333	-1.454	-1.2642	-0.8638304	-1.44465	-0.9714	0.0534	-1.425	-1.418401
Mode	NA	NA	NA	NA	NA	NA	NA	NA	NA	4800	2.75	NA

Canada

	Birthrate_per_1000	Deathrate_per_1000	govt_health_exp%GDP	ext_govt_exp%Health	Life_expectancy	mortality_rate_female_per_1000	mortality_rate_male_per_1000	mortality_rate_infant_per_1000	infant_deaths	maternal_deaths	population_growth	population
Mean	10.838	7.175	7.0688	0	80.968	54.21063	87.85	4.98125	1816	40.25	1.0244	33841324
Standard Deviation	0.3096	0.1483	0.47	0	0.8484	2.887434	5.867798	0.2482438	44.4237	3.3367	0.1119	1655596
Median	10.8	7.1	7.265	0	81.125	53.725	86.58	5	1822.5	41	1.035	33816892
Minimun	10.3	7	6.42	0	79.49	50.74	81.2	4.6	1742	33	0.75	31360079
Maximum	11.4	7.5	7.68	0	81.9	59.49	96.94	5.3	1871	44	1.2	36545236
Range	1.1	0.5	1.26	0	2.41	8.75	15.74	0.7	129	11	0.45	5185157
Skewness	0.2137	0.8619	-0.135	NaN	-0.329	0.2853334	0.2723029	-0.1670903	-0.248	-0.81	-0.657	0.06475256
Kurtosis	-1.103	-0.591	-1.886	NaN	-1.51	-1.449713	-1.641684	-1.487864	-1.5345	-0.502	-0.047	-1.454618
Mode	10.6	7.1	NA	0	81.9	NA	NA	5.3	NA	NA	NA	NA



Denmark

	Birthrate_per_1000	Deathrate_per_1000	govt_health_exp%GDP	ext_govt_exp%Health	Life_expectancy	mortality_rate_female_per_1000	mortality_rate_male_per_1000	mortality_rate_infant_per_1000	infant_deaths	maternal_deaths	population_growth	population
Mean	11.175	9.8	8.18	0	79.083	62.31375	102.2144	3.78125	235.625	3.5	0.4575	5540662
Standard Deviation	0.7638	0.568	0.5383	0	1.4225	9.022541	15.14634	0.4069705	33.7024	1.0328	0.159	123239.2
Median	11.4	9.85	8.465	0	78.85	63.965	105.775	3.55	219	3.5	0.43	5535389
Minimun	10	9.1	7.29	0	76.9	47.45	78.19	3.4	206	2	0.26	5375931
Maximum	12	10.9	9.02	0	81.1	75.79	124.08	4.6	299	5	0.78	5764980
Range	2	1.8	1.73	0	4.2	28.34	45.89	1.2	93	3	0.52	389049
Skewness	-0.257	0.3744	-0.252	NaN	-0.009	-0.15081	-0.1427518	0.8278858	0.6605	0	0.5216	0.28596
Kurtosis	-1.729	-1.208	-1.558	NaN	-1.613	-1.52937	-1.562825	-0.9434756	-1.2551	-1.297	-0.97	-1.295301
Mode	12	9.2	NA	0	80.7	NA	NA	3.5	207	NA	0.44	NA

Ethiopia

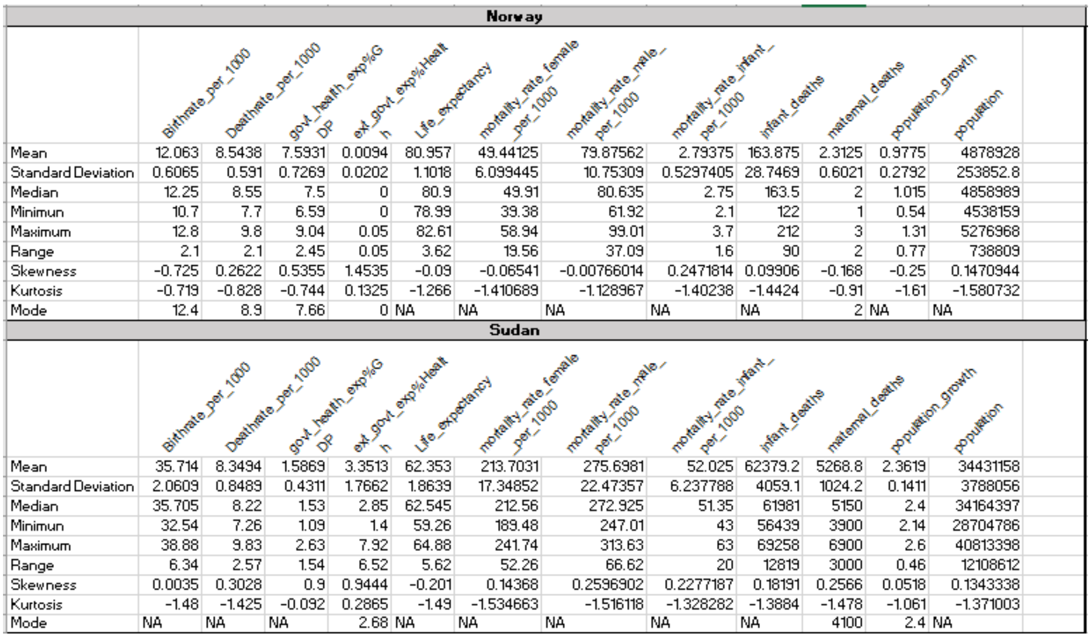
	Birthrate_per_1000	Deathrate_per_1000	govt_health_exp%GDP	ext_govt_exp%Health	Life_expectancy	mortality_rate_female_per_1000	mortality_rate_male_per_1000	mortality_rate_infant_per_1000	infant_deaths	maternal_deaths	population_growth	population
Mean	36.663	9.1456	1.1388	27.832	60.457	260.1506	306.0756	57.34375	177070	20563	2.7844	87189873
Standard Deviation	2.8918	2.0665	0.5182	9.2223	4.2283	59.60157	54.42019	12.99974	30722.1	5227.7	0.0545	11523952
Median	36.135	8.66	1.02	28.795	61.14	249.795	296.33	55.65	171733	19500	2.79	86436943
Minimun	32.78	6.69	0.38	12.7	53.35	188.54	240.42	39.5	136197	14000	2.66	70142090
Maximum	42.01	12.94	2.28	46.76	65.87	366.28	404.22	80.3	232838	29000	2.87	106399926
Range	9.23	6.25	1.9	34.06	12.52	177.74	163.8	40.8	96641	15000	0.21	36257836
Skewness	0.3795	0.4449	0.7711	0.0954	-0.28	0.373747	0.3944242	0.2785137	0.36033	0.4088	-0.492	0.13412
Kurtosis	-1.252	-1.321	-0.276	-0.881	-1.486	-1.416219	-1.378619	-1.352763	-1.2973	-1.405	-0.442	-1.411361
Mode	NA	NA	1.33	NA	NA	NA	NA	NA	NA	NA	2.83	NA

Japan

	Birthrate_per_1000	Deathrate_per_1000	govt_health_exp%GDP	ext_govt_exp%Health	Life_expectancy	mortality_rate_female_per_1000	mortality_rate_male_per_1000	mortality_rate_infant_per_1000	infant_deaths	maternal_deaths	population_growth	population
Mean	8.4175	9.3575	7.5438	0	82.809	41.66313	82.31188	2.41875	2652.25	66	-0.009	127631500
Standard Deviation	0.4593	0.9031	1.354	0	0.7936	3.353016	9.633106	0.3525502	498.985	14.038	0.125	365339.1
Median	8.455	9.3	7.355	0	82.715	42.18	83.85	2.4	2656.5	67	0	127739500
Minimun	7.6	8.05	5.94	0	81.56	35.72	66.55	1.9	1876	44	-0.19	126972000
Maximum	9.3	10.8	9.04	0	84.1	45.4	95.67	3	3474	88	0.23	128070000
Range	1.7	2.75	3.1	0	2.54	9.68	29.12	1.1	1598	44	0.42	1098000
Skewness	0.1825	0.0031	0.031	NaN	0.1145	-0.354652	-0.2066827	0.1198443	0.0148	-0.044	0.3798	-0.4133334
Kurtosis	-0.675	-1.584	-1.959	NaN	-1.311	-1.451908	-1.447385	-1.485025	-1.3798	-1.401	-0.94	-1.2964
Mode	NA	NA	8.96	0	82.59	NA	NA	NA	NA	NA	NA	127445000

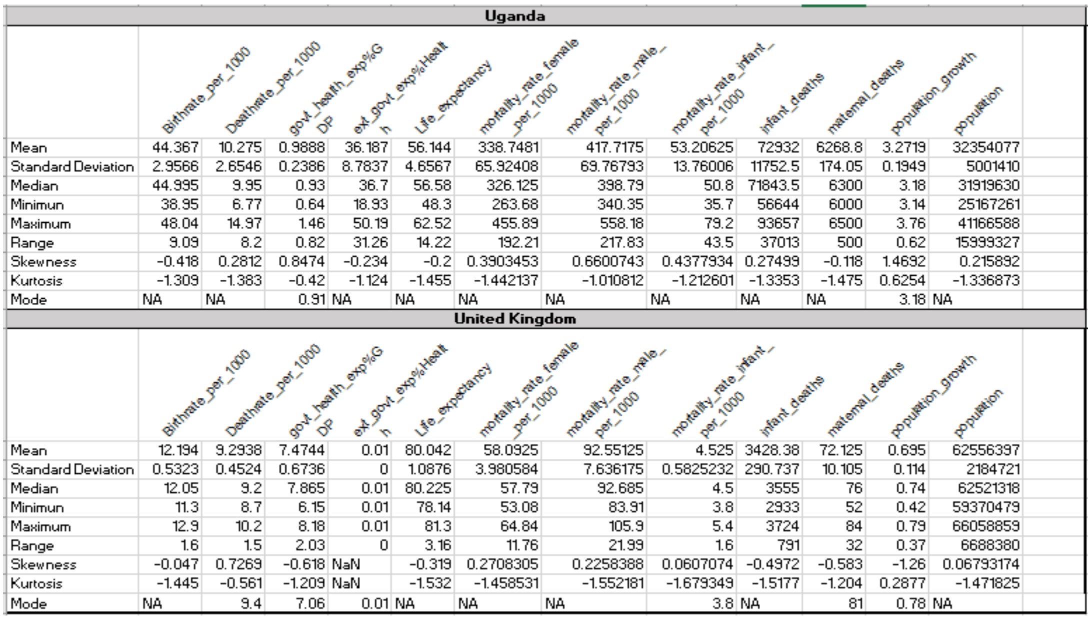
Kenya

	Birthrate_per_1000	Deathrate_per_1000	govt_health_exp%GDP	ext_govt_exp%Health	Life_expectancy	mortality_rate_female_per_1000	mortality_rate_male_per_1000	mortality_rate_infant_per_1000	infant_deaths	maternal_deaths	population_growth	population
Mean	35.011	8.2675	1.725	22.419	59.6	278.6081	349.8562	41.9875	59658.6	6912.5	2.6531	41666433
Standard Deviation	3.4543	2.2549	0.2416	5.384	4.8047	96.25396	89.10055	6.831386	7418.88	1627.6	0.1289	5281077
Median	35.535	7.71	1.725	21.87	60.445	257.595	327.065	39.85	58264.5	6650	2.71	41466241
Minimun	29.3	5.58	1.42	14.23	51.61	174.38	252.22	33.7	49302	5000	2.36	33751746
Maximum	39.47	12.2	2.2	29.15	65.91	452.88	523.46	55.5	72678	9200	2.77	50221146
Range	10.17	6.62	0.78	14.92	14.3	278.5	271.24	21.8	23376	4200	0.41	16469400
Skewness	-0.258	0.4093	0.4793	-0.016	-0.283	0.4389013	0.5861829	0.6125792	0.31662	0.2079	-1.052	0.08218244
Kurtosis	-1.51	-1.409	-1.04	-1.684	-1.466	-1.403297	-1.15132	-1.029502	-1.2745	-1.731	-0.319	-1.443945
Mode	NA	NA	NA	NA	NA	NA	NA	NA	NA	9200	NA	NA



Norway	Birthrate_per_1000	Deathrate_per_1000	govt_health_exp%GDP	ext_govt_exp%Health	Life_expectancy	mortality_rate_female_per_1000	mortality_rate_male_per_1000	mortality_rate_infant_per_1000	infant_deaths	maternal_deaths	population_growth	population
Mean	12.063	8.5438	7.5931	0.0094	80.957	49.44125	79.87562	2.79375	163.875	2.3125	0.9775	4878928
Standard Deviation	0.6065	0.591	0.7269	0.0202	1.1018	6.099445	10.75309	0.5297405	28.7469	0.6021	0.2792	253852.8
Median	12.25	8.55	7.5	0	80.9	49.91	80.635	2.75	163.5	2	1.015	4858989
Minimun	10.7	7.7	6.59	0	78.99	39.38	61.92	2.1	122	1	0.54	4538159
Maximum	12.8	9.8	9.04	0.05	82.61	58.94	99.01	3.7	212	3	1.31	5276968
Range	2.1	2.1	2.45	0.05	3.62	19.56	37.09	1.6	90	2	0.77	738809
Skewness	-0.725	0.2622	0.5355	1.4535	-0.09	-0.06541	-0.00766014	0.2471814	0.09906	-0.168	-0.25	0.1470944
Kurtosis	-0.719	-0.828	-0.744	0.1325	-1.266	-1.410689	-1.128967	-1.40238	-1.4424	-0.91	-1.61	-1.580732
Mode	12.4	8.9	7.66	0	NA	NA	NA	NA	NA	2	NA	NA

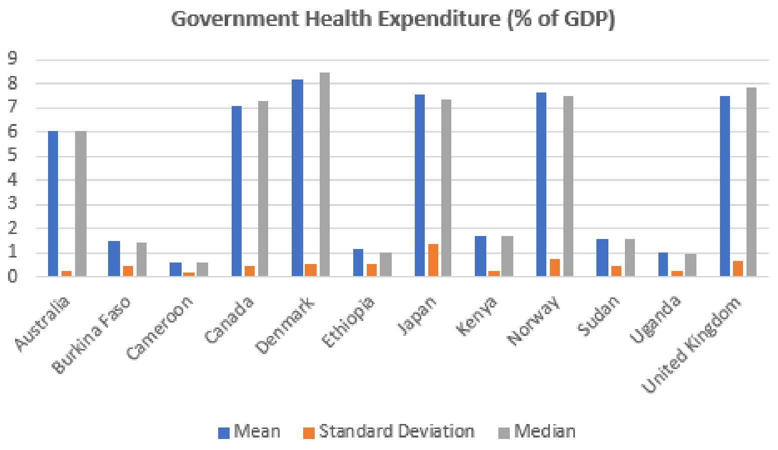
Sudan	Birthrate_per_1000	Deathrate_per_1000	govt_health_exp%GDP	ext_govt_exp%Health	Life_expectancy	mortality_rate_female_per_1000	mortality_rate_male_per_1000	mortality_rate_infant_per_1000	infant_deaths	maternal_deaths	population_growth	population
Mean	35.714	8.3494	1.5869	3.3513	62.353	213.7031	275.6981	52.025	62379.2	5268.8	2.3619	34431158
Standard Deviation	2.0609	0.8489	0.4311	1.7662	1.8639	17.34852	22.47357	6.237788	4059.1	1024.2	0.1411	3788056
Median	35.705	8.22	1.53	2.85	62.545	212.56	272.925	51.35	61981	5150	2.4	34164397
Minimun	32.54	7.26	1.09	1.4	59.26	189.48	247.01	43	56439	3900	2.14	28704786
Maximum	38.88	9.83	2.63	7.92	64.88	241.74	313.63	63	69258	6900	2.6	40813398
Range	6.34	2.57	1.54	6.52	5.62	52.26	66.62	20	12819	3000	0.46	12108612
Skewness	0.0035	0.3028	0.9	0.9444	-0.201	0.14368	0.2596902	0.2277187	0.18191	0.2566	0.0518	0.1343338
Kurtosis	-1.48	-1.425	-0.092	0.2865	-1.49	-1.534663	-1.516118	-1.328282	-1.3884	-1.478	-1.061	-1.371003
Mode	NA	NA	NA	2.68	NA	NA	NA	NA	NA	4100	2.4	NA



Uganda	Birthrate_per_1000	Deathrate_per_1000	govt_health_exp%GDP	ext_govt_exp%Health	Life_expectancy	mortality_rate_female_per_1000	mortality_rate_male_per_1000	mortality_rate_infant_per_1000	infant_deaths	maternal_deaths	population_growth	population
Mean	44.367	10.275	0.9888	36.187	56.144	338.7481	417.7175	53.20625	72932	6268.8	3.2719	32354077
Standard Deviation	2.9566	2.6546	0.2386	8.7837	4.6567	65.92408	69.76793	13.76006	11752.5	174.05	0.1949	5001410
Median	44.995	9.95	0.93	36.7	56.58	326.125	398.79	50.8	71843.5	6300	3.18	31919630
Minimun	38.95	6.77	0.64	18.93	48.3	263.68	340.35	35.7	56644	6000	3.14	25167261
Maximum	48.04	14.97	1.46	50.19	62.52	455.89	558.18	79.2	93657	6500	3.76	41166588
Range	9.09	8.2	0.82	31.26	14.22	192.21	217.83	43.5	37013	500	0.62	15999327
Skewness	-0.418	0.2812	0.8474	-0.234	-0.2	0.3903453	0.6600743	0.4377934	0.27499	-0.118	1.4692	0.215892
Kurtosis	-1.309	-1.383	-0.42	-1.124	-1.455	-1.442137	-1.010812	-1.212601	-1.3353	-1.475	0.6254	-1.336873
Mode	NA	NA	0.91	NA	NA	NA	NA	NA	NA	NA	3.18	NA

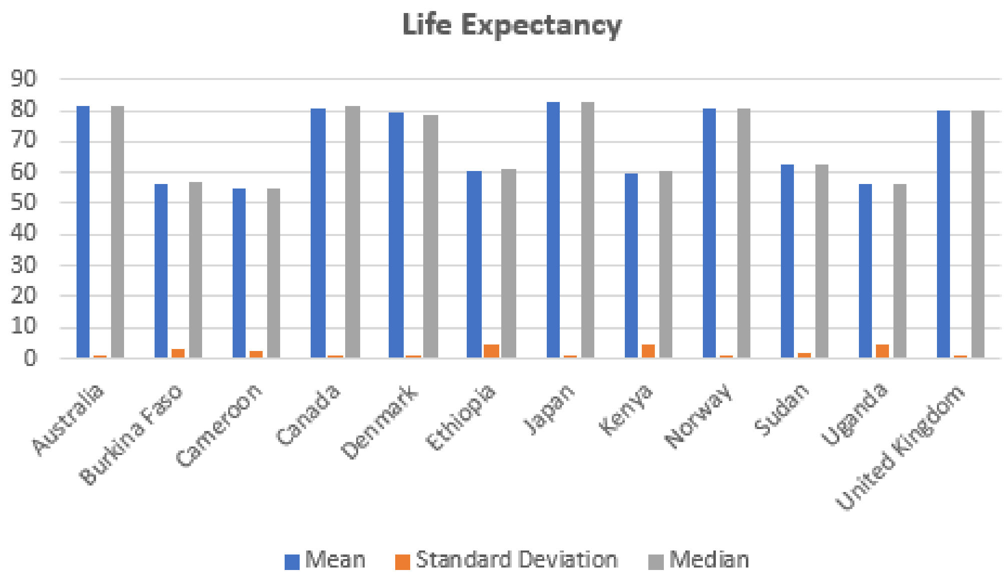
United Kingdom	Birthrate_per_1000	Deathrate_per_1000	govt_health_exp%GDP	ext_govt_exp%Health	Life_expectancy	mortality_rate_female_per_1000	mortality_rate_male_per_1000	mortality_rate_infant_per_1000	infant_deaths	maternal_deaths	population_growth	population
Mean	12.194	9.2938	7.4744	0.01	80.042	58.0925	92.55125	4.525	3428.38	72.125	0.695	62556397
Standard Deviation	0.5323	0.4524	0.6736	0	1.0876	3.980584	7.636175	0.5825232	290.737	10.105	0.114	2184721
Median	12.05	9.2	7.865	0.01	80.225	57.79	92.685	4.5	3555	76	0.74	62521318
Minimun	11.3	8.7	6.15	0.01	78.14	53.08	83.91	3.8	2933	52	0.42	59370479
Maximum	12.9	10.2	8.18	0.01	81.3	64.84	105.9	5.4	3724	84	0.79	66058859
Range	1.6	1.5	2.03	0	3.16	11.76	21.99	1.6	791	32	0.37	6688380
Skewness	-0.047	0.7269	-0.618	NaN	-0.319	0.2708305	0.2258388	0.0607074	-0.4972	-0.583	-1.26	0.06793174
Kurtosis	-1.445	-0.561	-1.209	NaN	-1.532	-1.458531	-1.552181	-1.679349	-1.5177	-1.204	0.2877	-1.471825
Mode	NA	9.4	7.06	0.01	NA	NA	NA	3.8	NA	81	0.78	NA

The image below shows a graph of the mean, median, and standard deviation of the percentage of the GDP allocated to the health systems of each country:



As seen from the image above, there is a remarkable difference between how much the countries from the developed countries budget for healthcare as compared to their African counterparts.

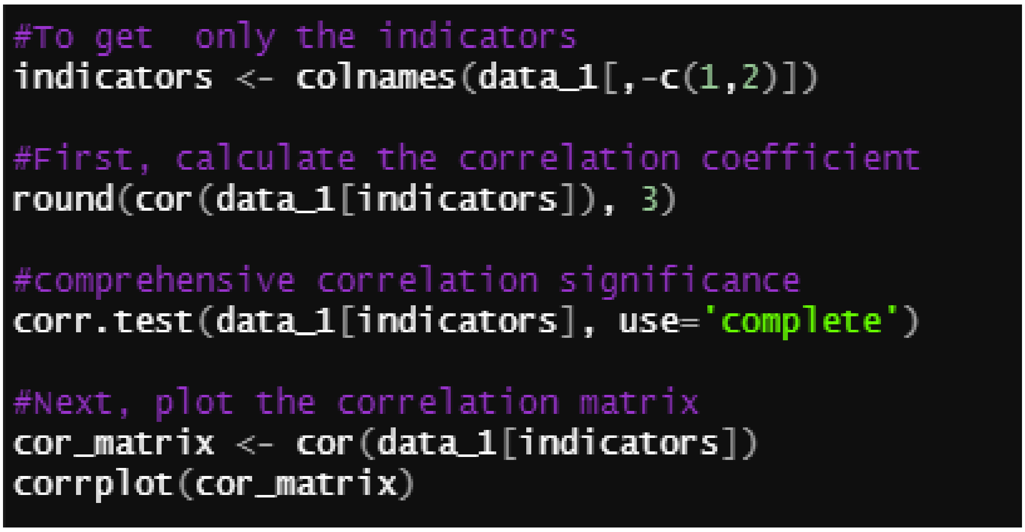
Further analysis of the results also shows a similar trend, as the chart below shows the life expectancy in the developed countries in a region higher than that of the African countries.



4.2 Correlation Analysis

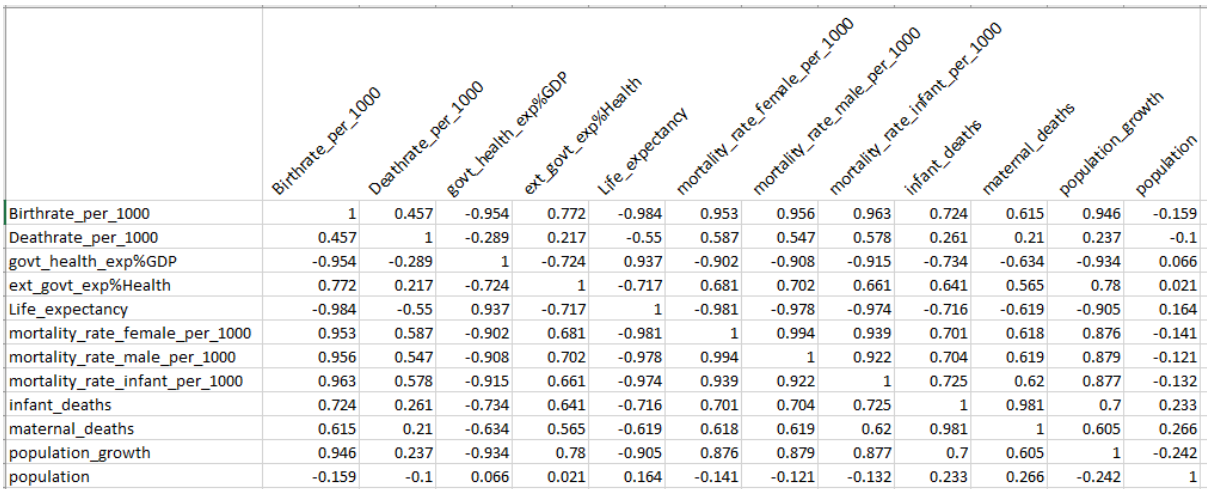
Correlation is a measure of the presence or strength of a relationship between variables.

This report used the Pearson correlation coefficient, which is ideal because it measures the linear dependence between 2 variables. Codes are shown below:



```
#To get  only the indicators
indicators <- colnames(data_1[,-c(1,2)])

#First, calculate the correlation coefficient
round(cor(data_1[indicators]), 3)

#comprehensive correlation significance
corr.test(data_1[indicators], use='complete')

#Next, plot the correlation matrix
cor_matrix <- cor(data_1[indicators])
corrplot(cor_matrix)
```

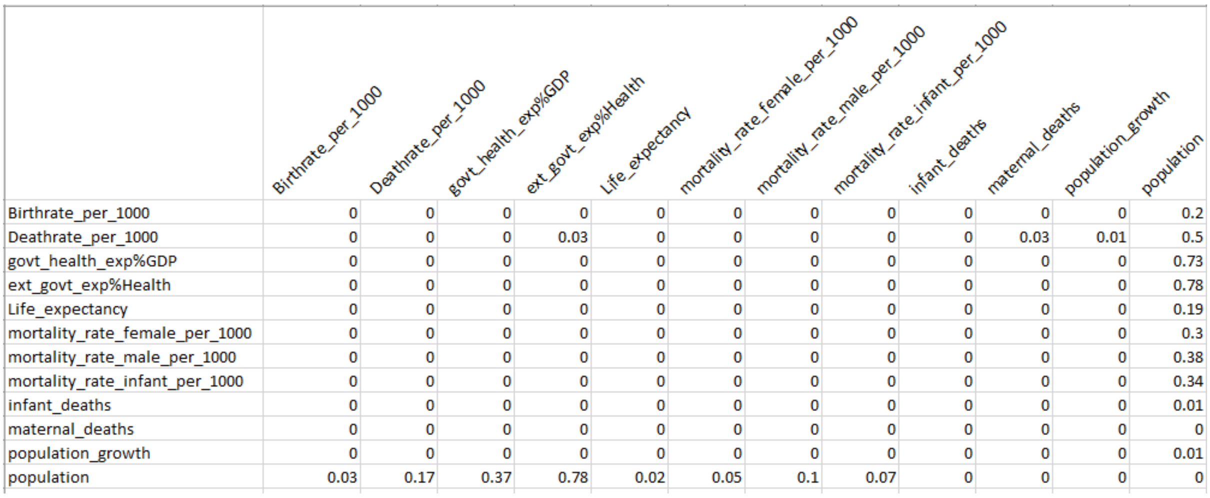
First, we calculate the correlation between the variables:



	Birthrate_per_1000	Deathrate_per_1000	govt_health_exp%GDP	ext_govt_exp%Health	Life_expectancy	mortality_rate_female_per_1000	mortality_rate_male_per_1000	mortality_rate_infant_per_1000	infant_deaths	maternal_deaths	population_growth	population
Birthrate_per_1000	1	0.457	-0.954	0.772	-0.984	0.953	0.956	0.963	0.724	0.615	0.946	-0.159
Deathrate_per_1000	0.457	1	-0.289	0.217	-0.55	0.587	0.547	0.578	0.261	0.21	0.237	-0.1
govt_health_exp%GDP	-0.954	-0.289	1	-0.724	0.937	-0.902	-0.908	-0.915	-0.734	-0.634	-0.934	0.066
ext_govt_exp%Health	0.772	0.217	-0.724	1	-0.717	0.681	0.702	0.661	0.641	0.565	0.78	0.021
Life_expectancy	-0.984	-0.55	0.937	-0.717	1	-0.981	-0.978	-0.974	-0.716	-0.619	-0.905	0.164
mortality_rate_female_per_1000	0.953	0.587	-0.902	0.681	-0.981	1	0.994	0.939	0.701	0.618	0.876	-0.141
mortality_rate_male_per_1000	0.956	0.547	-0.908	0.702	-0.978	0.994	1	0.922	0.704	0.619	0.879	-0.121
mortality_rate_infant_per_1000	0.963	0.578	-0.915	0.661	-0.974	0.939	0.922	1	0.725	0.62	0.877	-0.132
infant_deaths	0.724	0.261	-0.734	0.641	-0.716	0.701	0.704	0.725	1	0.981	0.7	0.233
maternal_deaths	0.615	0.21	-0.634	0.565	-0.619	0.618	0.619	0.62	0.981	1	0.605	0.266
population_growth	0.946	0.237	-0.934	0.78	-0.905	0.876	0.879	0.877	0.7	0.605	1	-0.242
population	-0.159	-0.1	0.066	0.021	0.164	-0.141	-0.121	-0.132	0.233	0.266	-0.242	1

One thing that stands out from the image above is the main diagonal. It is all 1, and that is because a variable has the maximum positive correlation with itself.

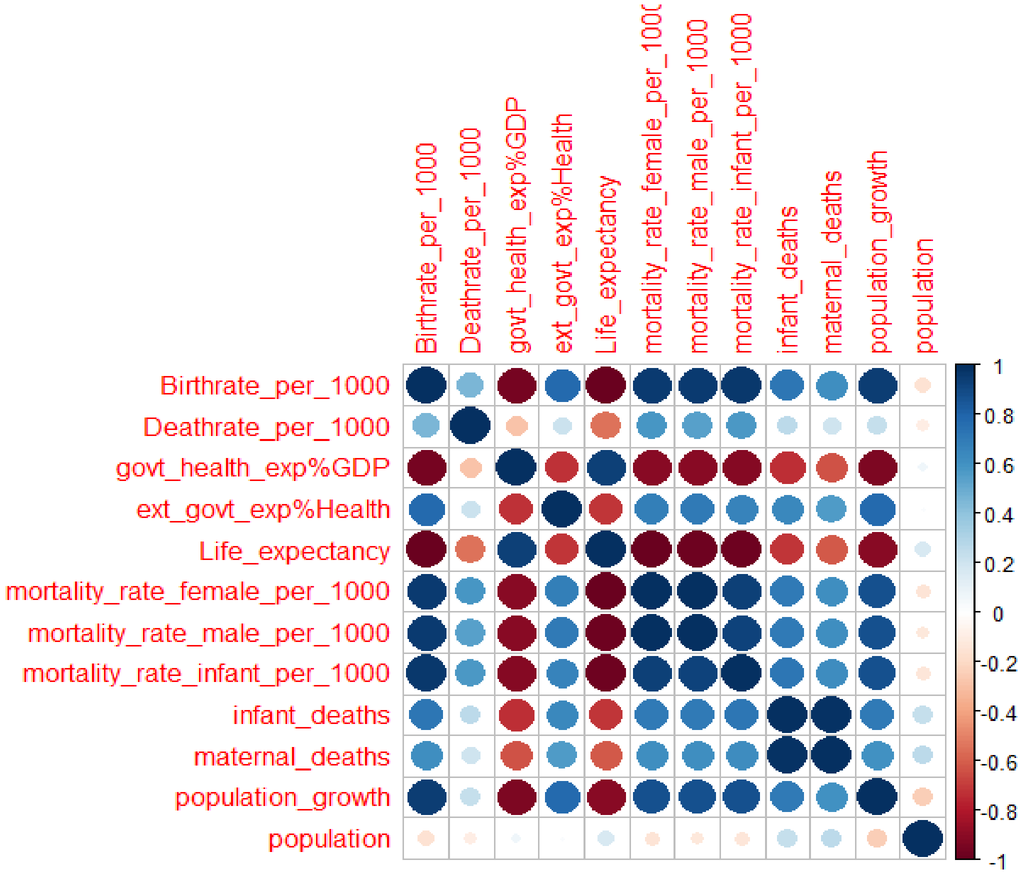
To do a comprehensive test to get the significance of the correlations, the 'psych' package has a 'test' function. The results are shown below:



	Birthrate_per_1000	Deathrate_per_1000	govt_health_exp%GDP	ext_govt_exp%Health	Life_expectancy	mortality_rate_female_per_1000	mortality_rate_male_per_1000	mortality_rate_infant_per_1000	infant_deaths	maternal_deaths	population_growth	population
Birthrate_per_1000	0	0	0	0	0	0	0	0	0	0	0	0.2
Deathrate_per_1000	0	0	0	0.03	0	0	0	0	0	0.03	0.01	0.5
govt_health_exp%GDP	0	0	0	0	0	0	0	0	0	0	0	0.73
ext_govt_exp%Health	0	0	0	0	0	0	0	0	0	0	0	0.78
Life_expectancy	0	0	0	0	0	0	0	0	0	0	0	0.19
mortality_rate_female_per_1000	0	0	0	0	0	0	0	0	0	0	0	0.3
mortality_rate_male_per_1000	0	0	0	0	0	0	0	0	0	0	0	0.38
mortality_rate_infant_per_1000	0	0	0	0	0	0	0	0	0	0	0	0.34
infant_deaths	0	0	0	0	0	0	0	0	0	0	0	0.01
maternal_deaths	0	0	0	0	0	0	0	0	0	0	0	0
population_growth	0	0	0	0	0	0	0	0	0	0	0	0.01
population	0.03	0.17	0.37	0.78	0.02	0.05	0.1	0.07	0	0	0	0

In the image above, all values less than 0.05 can be said to have a significant correlation, as opposed to being a random occurrence. This means that majority of the indicators are correlated.

To see the relationship between variables at a glance, we plot the correlation matrix using the 'corrplot' library.



From the image above, the variables with strong relationships with each other can easily be identified. For example, government health expenditure and life expectancy have a strong positive correlation, while all the mortality rates and life expectancy have a correlation coefficient of -0.981, -0.978, and -0.974, which can be interpreted as a strong negative relationship. This posits that as the mortality rates are decreasing in this dataset, life expectancy is increasing.

Some indicator pairs show no relationship with each other like the government health funds from external sources and the population.

Rules:

Correlation coefficient (r)	Positive
> 0.7	Very High
0.5 < r < 0.7	High
0.3 < r <0.5	Medium
< 0.3	Low

Correlation coefficient (r)	Negative
< - 0.7	Very High
- 0.7 < r < - 0.5	High
- 0.5 < r < - 0.3	Medium
> - 0.3	Low

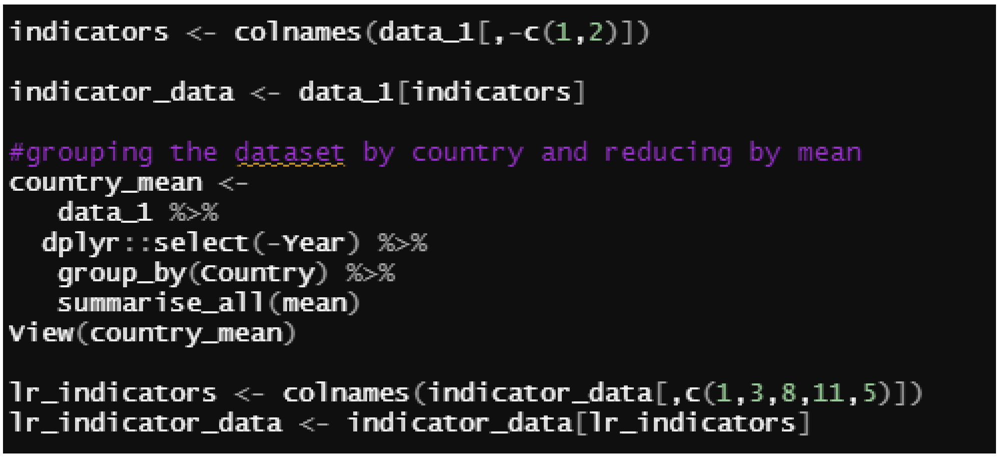
The correlation between the percentage of the country's GDP dedicated to healthcare and the life expectancy in the country is 0.937. This is regarded as a very high positive correlation, so It can be said that from the analysis, the GDP increases as Life expectancy increases.

4.3 Regression Analysis

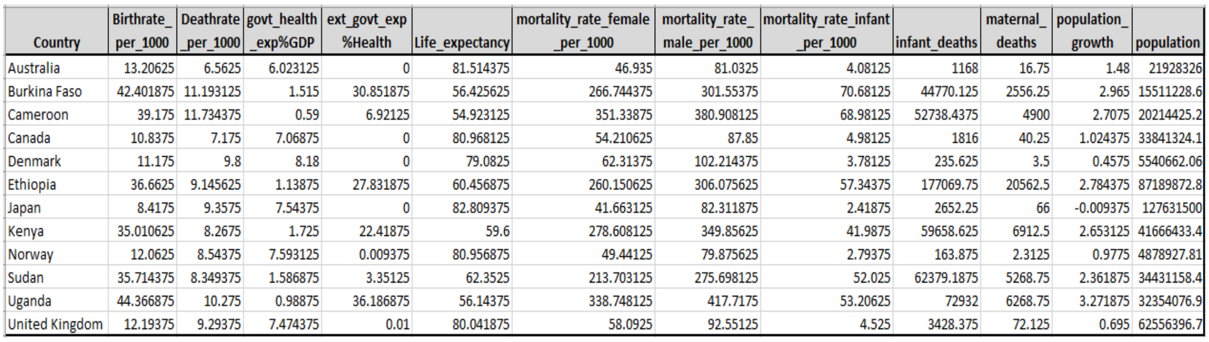
This task seeks to further assess the possible relationship between the life expectancy of each country and some indicators selected based on their correlation with the life expectancy indicator.

A reference to the correlation matrix plotted in the previous section shows that Birth rate, government health expenditure (percentage of GDP), infant mortality rate, and population growth show significant impacts on life expectancy.

The above columns are selected, and since the performance of countries is what's being compared, data is grouped by country and reduced by the mean, and afterward, the 'year' column is dropped.



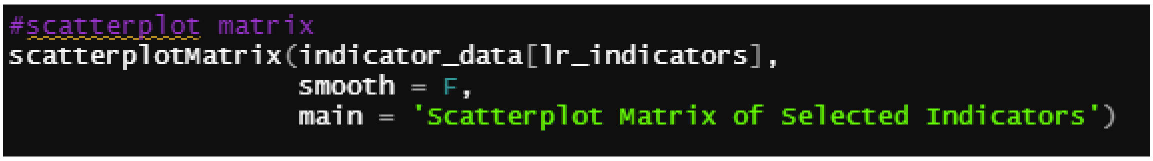
```
indicators <- colnames(data_1[,-c(1,2)])

indicator_data <- data_1[indicators]

#grouping the dataset by country and reducing by mean
country_mean <-
   data_1 %>%
  dplyr::select(-Year) %>%
   group_by(Country) %>%
   summarise_all(mean)
view(country_mean)

lr_indicators <- colnames(indicator_data[,c(1,3,8,11,5)])
lr_indicator_data <- indicator_data[lr_indicators]
```



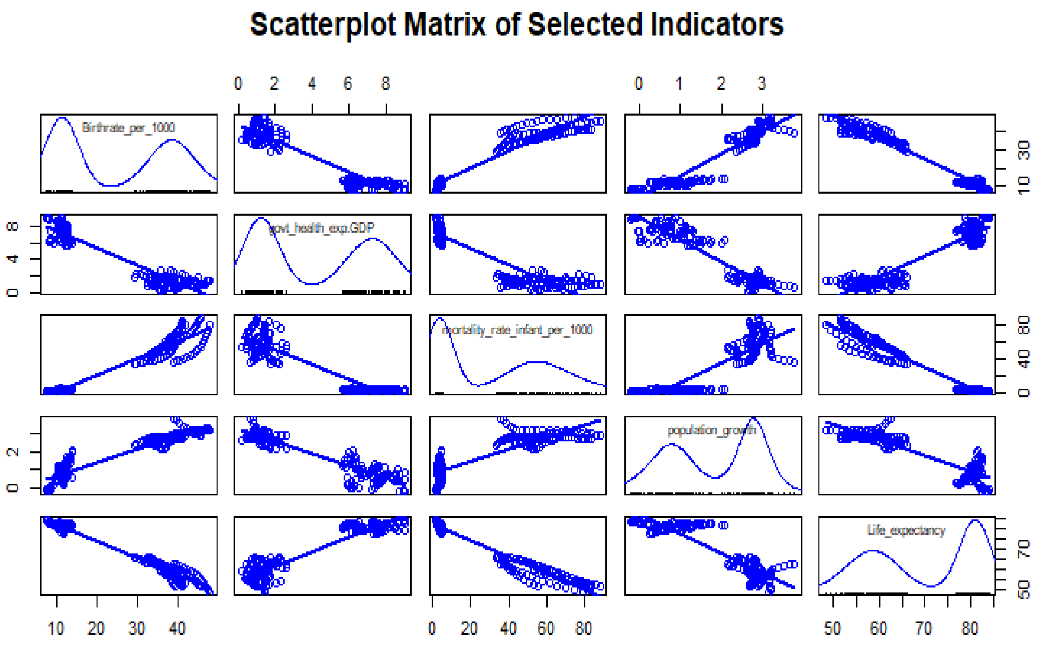
	Birthrate_	Deathrate	govt_health	ext_govt_exp		mortality_rate_female	mortality_rate_	mortality_rate_infant		maternal_	population_	
Country	per_1000	_per_1000	_exp%GDP	%Health	Life_expectancy	_per_1000	male_per_1000	_per_1000	infant_deaths	deaths	growth	population
Australia	13.20625	6.5625	6.023125	0	81.514375	46.935	81.0325	4.08125	1168	16.75	1.48	21928326
Burkina Faso	42.401875	11.193125	1.515	30.851875	56.425625	266.744375	301.55375	70.68125	44770.125	2556.25	2.965	15511228.6
Cameroon	39.175	11.734375	0.59	6.92125	54.923125	351.33875	380.908125	68.98125	52738.4375	4900	2.7075	20214425.2
Canada	10.8375	7.175	7.06875	0	80.968125	54.210625	87.85	4.98125	1816	40.25	1.024375	33841324.1
Denmark	11.175	9.8	8.18	0	79.0825	62.31375	102.214375	3.78125	235.625	3.5	0.4575	5540662.06
Ethiopia	36.6625	9.145625	1.13875	27.831875	60.456875	260.150625	306.075625	57.34375	177069.75	20562.5	2.784375	87189872.8
Japan	8.4175	9.3575	7.54375	0	82.809375	41.663125	82.311875	2.41875	2652.25	66	-0.009375	127631500
Kenya	35.010625	8.2675	1.725	22.41875	59.6	278.608125	349.85625	41.9875	59658.625	6912.5	2.653125	41666433.4
Norway	12.0625	8.54375	7.593125	0.009375	80.956875	49.44125	79.875625	2.79375	163.875	2.3125	0.9775	4878927.81
Sudan	35.714375	8.349375	1.586875	3.35125	62.3525	213.703125	275.698125	52.025	62379.1875	5268.75	2.361875	34431158.4
Uganda	44.366875	10.275	0.98875	36.186875	56.14375	338.748125	417.7175	53.20625	72932	6268.75	3.271875	32354076.9
United Kingdom	12.19375	9.29375	7.474375	0.01	80.041875	58.0925	92.55125	4.525	3428.375	72.125	0.695	62556396.7

Since multiple indicators are being tested, the Multiple Linear Regression is appropriate, and the statistical model will be the Ordinary Least Squares (OLS) model seeing that the predicting and target indicators are of continuous data types.

A scatterplot matrix of the selected indicators was plotted:



```
#scatterplot matrix
scatterplotMatrix(indicator_data[lr_indicators],
                  smooth = F,
                  main = 'Scatterplot Matrix of Selected Indicators')
```



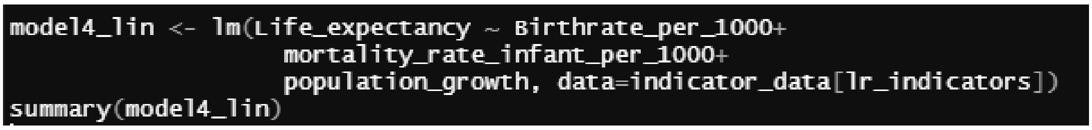
Afterward, the linear regression models can now be built.

4.3.1 *Advanced linear regression models*

Ordinary Least Square (OLS) Regression

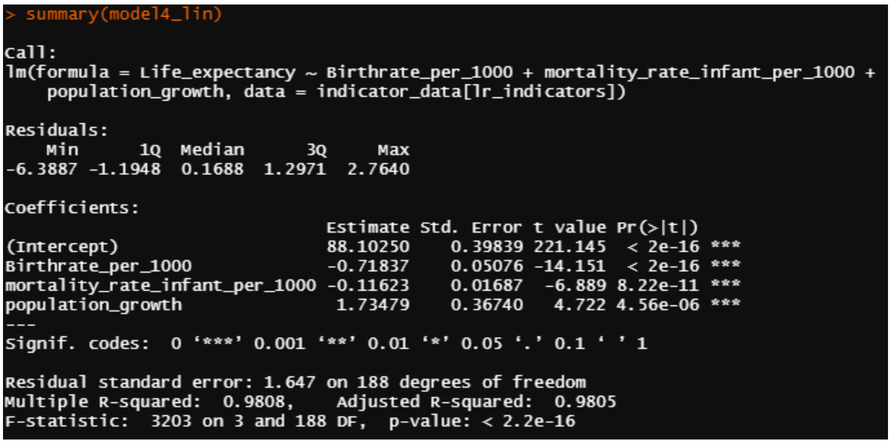
The OLS linear regression was carried out on the independent variables using the 'lm' function. Several combinations of variables were tested, and the ideal variables were picked based on if their correlation is significant, if adjusted R^2 values are higher, and if their Residual Standard Error is lower.

The variables used in building the model are as shown in the code below:



```
model4_lin <- lm(Life_expectancy ~ Birthrate_per_1000+
                    mortality_rate_infant_per_1000+
                    population_growth, data=indicator_data[lr_indicators])
summary(model4_lin)
```

The result:



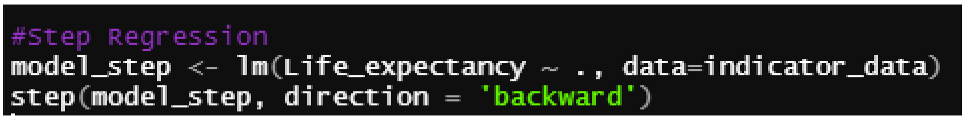
```
> summary(model4_lin)

Call:
lm(formula = Life_expectancy ~ Birthrate_per_1000 + mortality_rate_infant_per_1000 +
    population_growth, data = indicator_data[lr_indicators])

Residuals:
    Min      1Q  Median      3Q     Max 
-6.3887 -1.1948  0.1688  1.2971  2.7640 

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)    
(Intercept)                    88.10250    0.39839 221.145  < 2e-16 ***
Birthrate_per_1000             -0.71837    0.05076 -14.151  < 2e-16 ***
mortality_rate_infant_per_1000 -0.11623    0.01687  -6.889 8.22e-11 ***
population_growth               1.73479    0.36740   4.722 4.56e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.647 on 188 degrees of freedom
Multiple R-squared:  0.9808,	Adjusted R-squared:  0.9805 
F-statistic:  3203 on 3 and 188 DF,  p-value: < 2.2e-16
```

From the results shown above,

- The ‘estimate’ in the coefficients show the relationship between the dependent and independent variable.
- The (***) beside each variable shows the level of significance.
- Since the p-value shows less than 0.05, the null hypothesis, which states that our regression model is not significant, is rejected. This means that at least one variable is significantly associated with life expectancy.

Backward Stepwise Regression

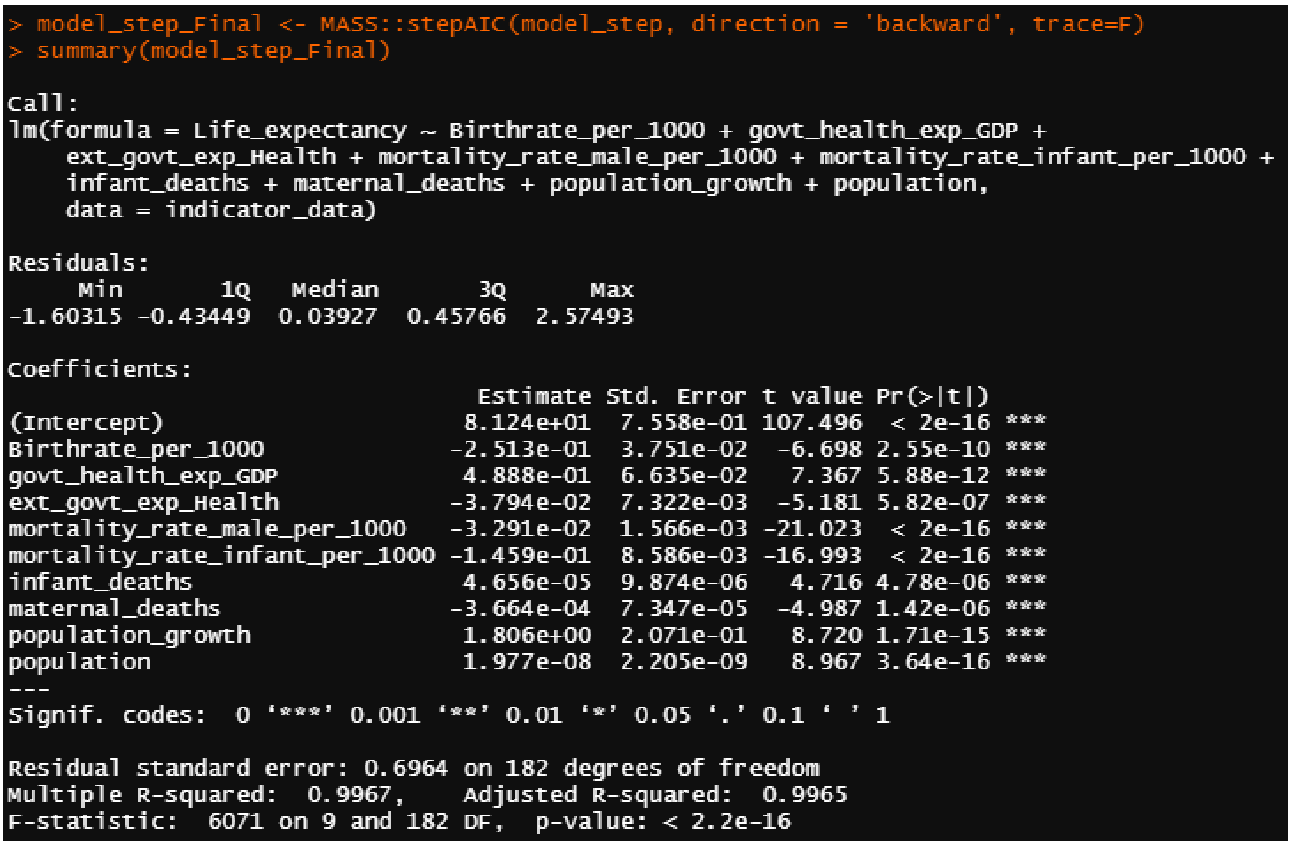
This means the algorithm starts by using all the indicator values, compares different models, then removes an economic indicator at a time while recording their Akaike information criterion (AIC). It then returns the model with the least AIC value:



```
#Step Regression
model_step <- lm(Life_expectancy ~ ., data=indicator_data)
step(model_step, direction = 'backward')
```

Even though this type of regression is great and saves time while looking for the best model, the drawback is that it does not evaluate every single combination of indicators.

To overcome this, the StepAIC function was employed:

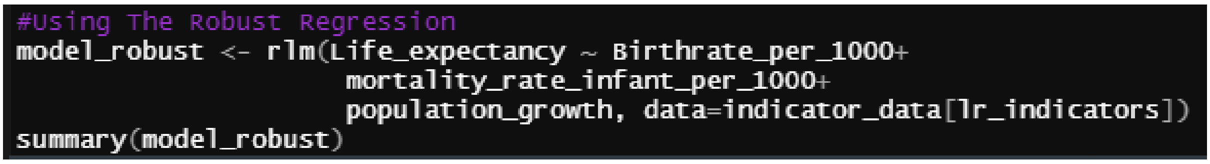


```
> model_step_Final <- MASS::stepAIC(model_step, direction = 'backward', trace=F)
> summary(model_step_Final)

Call:
lm(formula = Life_expectancy ~ Birthrate_per_1000 + govt_health_exp_GDP + 
    ext_govt_exp_Health + mortality_rate_male_per_1000 + mortality_rate_infant_per_1000 + 
    infant_deaths + maternal_deaths + population_growth + population, 
    data = indicator_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.60315 -0.43449  0.03927  0.45766  2.57493 

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)    
(Intercept)                    8.124e+01  7.558e-01 107.496  < 2e-16 ***
Birthrate_per_1000            -2.513e-01  3.751e-02  -6.698 2.55e-10 ***
govt_health_exp_GDP            4.888e-01  6.635e-02   7.367 5.88e-12 ***
ext_govt_exp_Health           -3.794e-02  7.322e-03  -5.181 5.82e-07 ***
mortality_rate_male_per_1000  -3.291e-02  1.566e-03 -21.023  < 2e-16 ***
mortality_rate_infant_per_1000 -1.459e-01  8.586e-03 -16.993  < 2e-16 ***
infant_deaths                  4.656e-05  9.874e-06   4.716 4.78e-06 ***
maternal_deaths               -3.664e-04  7.347e-05  -4.987 1.42e-06 ***
population_growth              1.806e+00  2.071e-01   8.720 1.71e-15 ***
population                     1.977e-08  2.205e-09   8.967 3.64e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6964 on 182 degrees of freedom
Multiple R-squared:  0.9967,	Adjusted R-squared:  0.9965 
F-statistic:  6071 on 9 and 182 DF,  p-value: < 2.2e-16
```

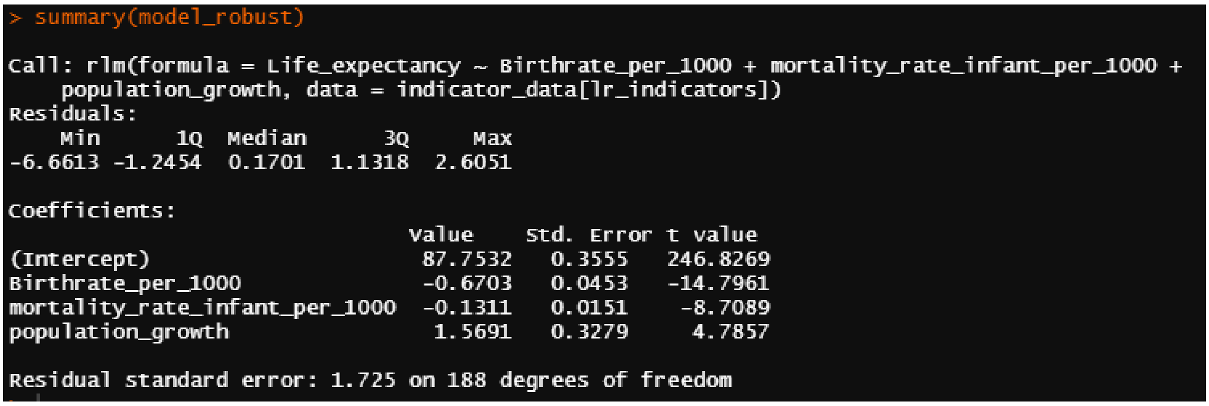
The robust method

This is an advanced form of regression that takes outliers and other influencing observations into consideration.

It is done using the ‘rlm’ function in the ‘Mass’ package:



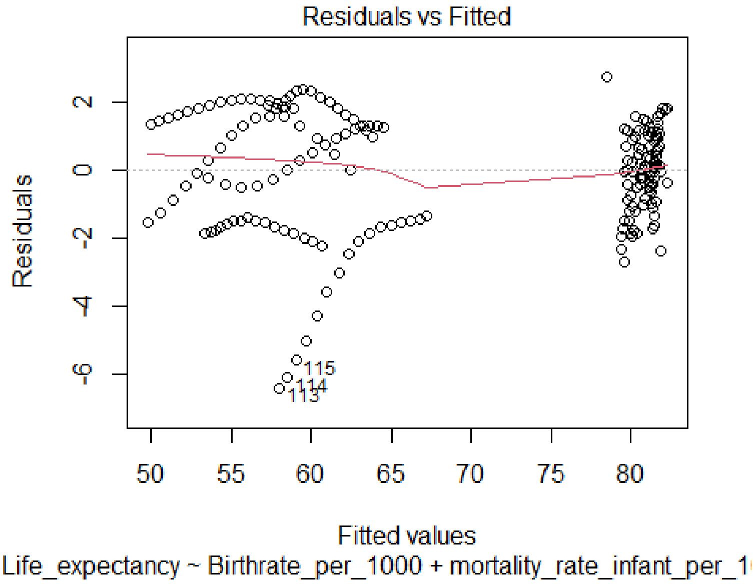
```
#Using The Robust Regression
model_robust <- rlm(Life_expectancy ~ Birthrate_per_1000+
                          mortality_rate_infant_per_1000+
                          population_growth, data=indicator_data[lr_indicators])
summary(model_robust)
```



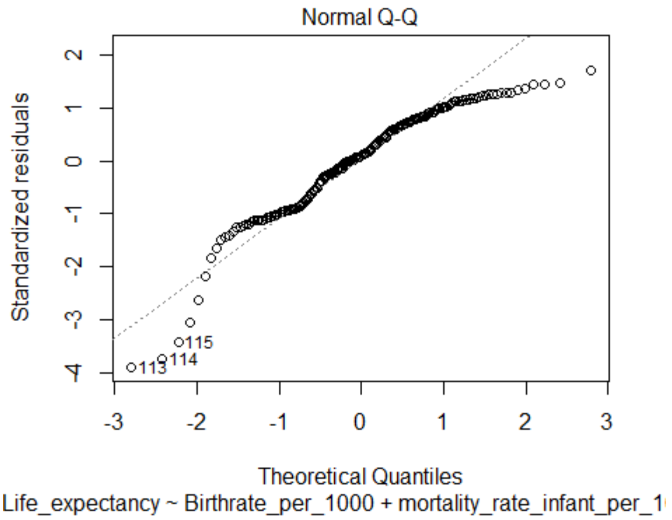
```
> summary(model_robust)

Call: rlm(formula = Life_expectancy ~ Birthrate_per_1000 + mortality_rate_infant_per_1000 +
    population_growth, data = indicator_data[lr_indicators])
Residuals:
    Min      1Q  Median      3Q     Max 
-6.6613 -1.2454  0.1701  1.1318  2.6051 

Coefficients:
                               Value    Std. Error t value 
(Intercept)                     87.7532   0.3555   246.8269
Birthrate_per_1000              -0.6703   0.0453   -14.7961
mortality_rate_infant_per_1000  -0.1311   0.0151    -8.7089
population_growth                1.5691   0.3279     4.7857

Residual standard error: 1.725 on 188 degrees of freedom
```

4.3.2 *Meeting of assumptions*

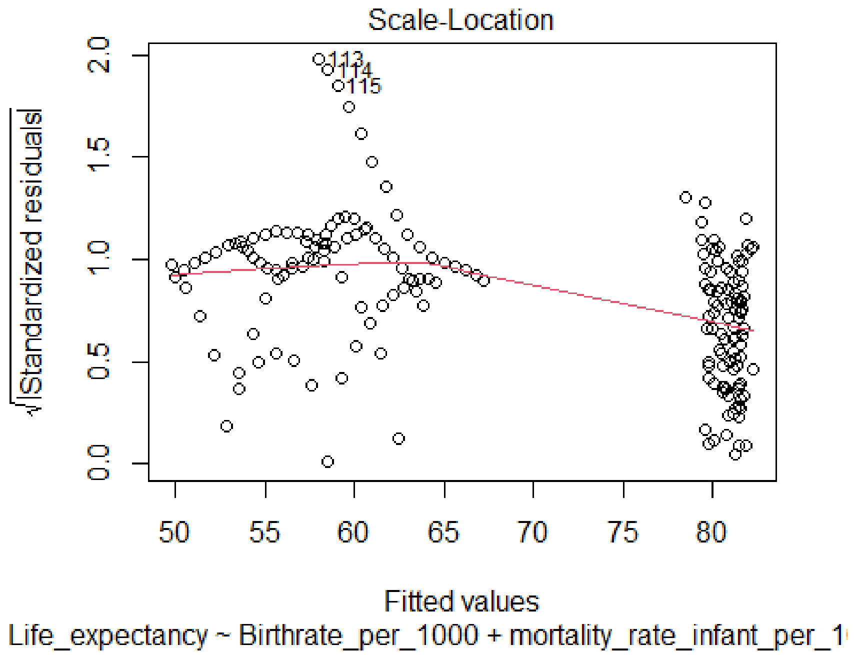
1. Linearity: From the scatterplot shown earlier, the variables show a linear relationship
2. Residual’s Independence: The graph below shows the residuals as independent as the plot line is generally horizontal.



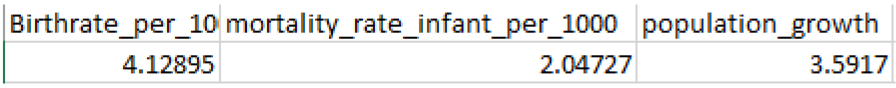
3. Normality of residuals: The graph below shows reasonably normally distributed residuals



4. **Homoscedasticity:** This shows that the residual variance is randomly scattered, meaning approximately constant.



5. **Colinearity:** A quick check using the VIF measures in the ‘car’ package shows no collinearity.



Birthrate_per_10	mortality_rate_infant_per_1000	population_growth
4.12895	2.04727	3.5917

Since all 5 assumptions were approved, the fitted regression line is as follows:

Regression model

Life expectancy = 88.10250 + (-0.71837) x Birthrate_per_1000 + (-0.11623) x mortality_rate_infant_per_1000 + 1.73479 x population_growth

4.3.3 Similar research:

OLS Method

This work by (Liu et al., 2021) seeks to analyse the factors that influence the change in life expectancy in Chinese cities. It uses the Ordinary least squares regression method to explore the factors of the change in this economic indicator. The OLS method in this instance is carried out based on the economic development levels. The results show that even though the OLS model solves spatial autocorrelation problems, the results can be improved upon using the Geographically Weighted Regression (GWR) model.

Stepwise regression Method

This work, (Samadder et al., 2014), used stepwise regression to predict how arsenic pollution affects life expectancy. The algorithm selected 4 out of the 8 independent variables as the ideal prediction model.

Robust Regression Method

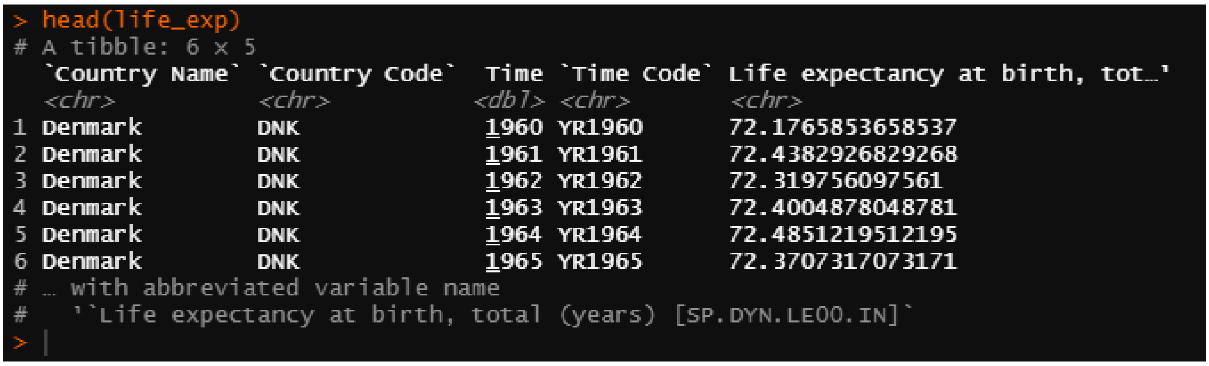
The journal article by (Hakim et al., 2020) applies a type of robust regression called the Spatial Durbin (SD) Robust regression model to analyse the relationship between independent and dependent variables with spatial influence. The results show that the SD model produces greater adjusted R^2 values than the other regression models tested.

4.4 Time series analysis

Time series analysis requires that the data be long enough to produce accurate and reasonable predictions for the phenomena of interest. (Hassouna & Al-Sahili, 2020)

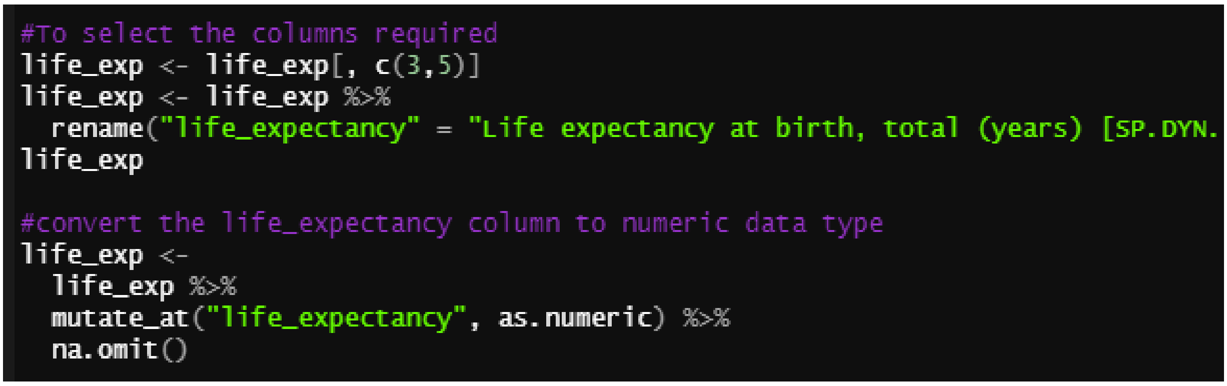
As such, for this task, a developed country (Denmark), was selected, and the period downloaded was extended to 60 years.

First, a quick look is done to have a feel of the data:

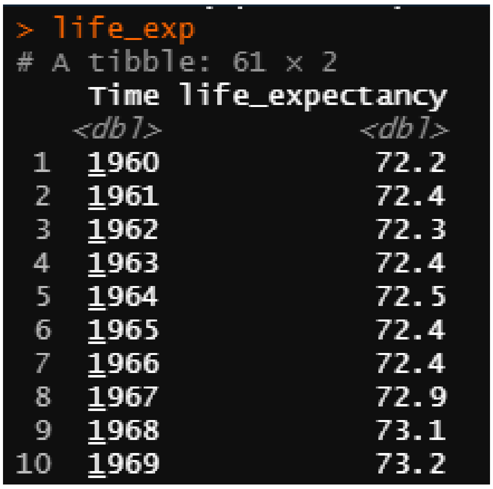


```
> head(life_exp)
# A tibble: 6 × 5
  `Country Name` `Country Code`  Time `Time Code` Life expectancy at birth, tot…¹
  <chr>          <chr>          <dbl> <chr>       <chr>
1 Denmark        DNK             1960 YR1960      72.1765853658537
2 Denmark        DNK             1961 YR1961      72.4382926829268
3 Denmark        DNK             1962 YR1962      72.319756097561
4 Denmark        DNK             1963 YR1963      72.4004878048781
5 Denmark        DNK             1964 YR1964      72.4851219512195
6 Denmark        DNK             1965 YR1965      72.3707317073171
# … with abbreviated variable name
#   ¹`Life expectancy at birth, total (years) [SP.DYN.LE00.IN]`
> |
```

Next, the columns required are selected, and as the image above shows that the life_expectancy column is a character column, it was changed into a numeric column. 'NA's were introduced by R because there were some characters that couldn't be converted to numbers. These were eventually dropped to get the data to be used:



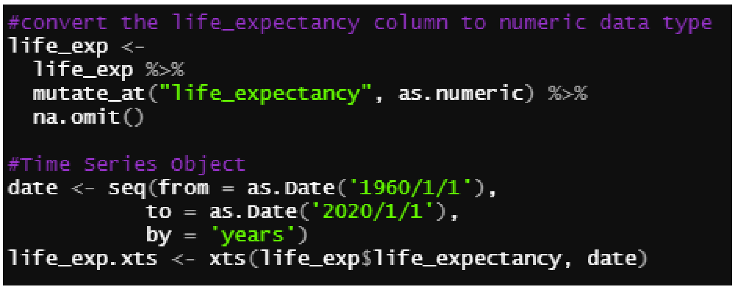
```
#To select the columns required
life_exp <- life_exp[, c(3,5)]
life_exp <- life_exp %>%
  rename("life_expectancy" = "Life expectancy at birth, total (years) [SP.DYN.
life_exp

#convert the life_expectancy column to numeric data type
life_exp <-
  life_exp %>%
  mutate_at("life_expectancy", as.numeric) %>%
  na.omit()
```



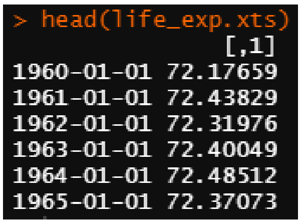
```
> life_exp
# A tibble: 61 × 2
    Time life_expectancy
   <dbl>           <dbl>
 1  1960            72.2
 2  1961            72.4
 3  1962            72.3
 4  1963            72.4
 5  1964            72.5
 6  1965            72.4
 7  1966            72.4
 8  1967            72.9
 9  1968            73.1
10  1969            73.2
```

4.4.1 Time series object

Next, a time series object containing both the observations and corresponding date specifications was created from the time series data to allow for easier manipulation:

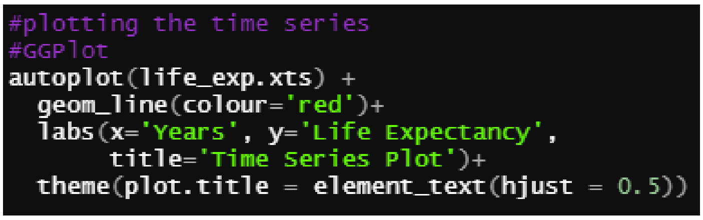


```
#convert the life_expectancy column to numeric data type
life_exp <-
  life_exp %>%
  mutate_at("life_expectancy", as.numeric) %>%
  na.omit()

#Time Series Object
date <- seq(from = as.Date('1960/1/1'),
            to = as.Date('2020/1/1'),
            by = 'years')
life_exp.xts <- xts(life_exp$life_expectancy, date)
```

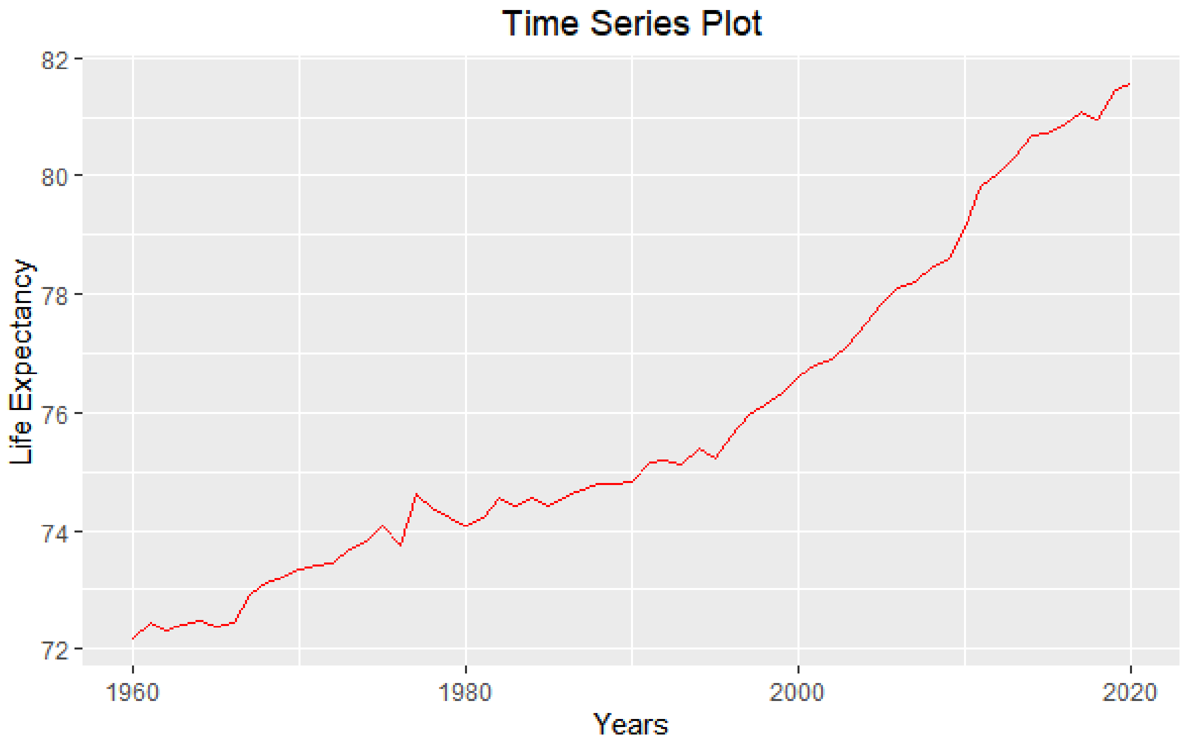


```
> head(life_exp.xts)
               [,1]
1960-01-01 72.17659
1961-01-01 72.43829
1962-01-01 72.31976
1963-01-01 72.40049
1964-01-01 72.48512
1965-01-01 72.37073
```

Next, the time series of the life expectancy in Denmark from 1960 to 2020 was plotted as shown below:



```
#plotting the time series
#GGPlot
autoplot(life_exp.xts) +
  geom_line(colour='red')+
  labs(x='Years', y='Life Expectancy',
       title='Time Series Plot')+
  theme(plot.title = element_text(hjust = 0.5))
```



4.4.2 *Decomposing the time series*

This step takes a look at 4 components of the time series:

- Trend: this is the general overall outlook
- Seasonality: this is the pattern that repeats at a particular season
- Observed Cyclic behavior: a bit like seasonality, it shows patterns that repeat over the years
- Random errors: are random components that have no explanation

In this report, since the data is non-seasonal, decomposition doesn't need to be carried out.

4.4.3 *Using the AutoRegressive 'Integrated' Moving Average (ARMA/ARIMA) model*

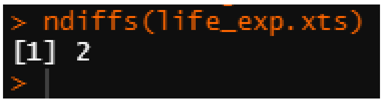
Here, the predicted values are derived as a linear function of both the recent actual values and the recent residuals (prediction errors)

Making the data stationary

For the ARIMA model to be used effectively, the data has to be stationary. Being stationary in this case indicates that the mean, variance, and auto-covariance have to be stable over time.

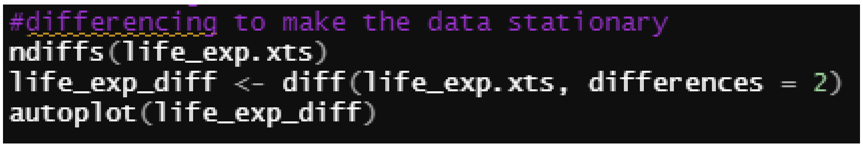
Since the data is not stationary (shows a clear upward trend over time). There is a need to 'difference' the dataset to make it stationary.

This repeated 'differencing' is the 'I' component that makes the model 'ARIMA' instead of 'ARMA'. The 'ndiffs' function tells us the data needs to be differenced twice.

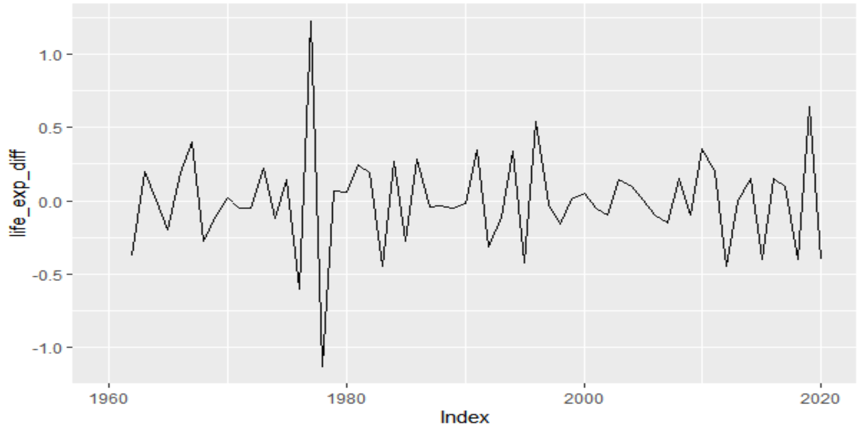


```
> ndiffs(life_exp.xts)
[1] 2
>
```

Then, the differencing was done, and the time-series data were plotted

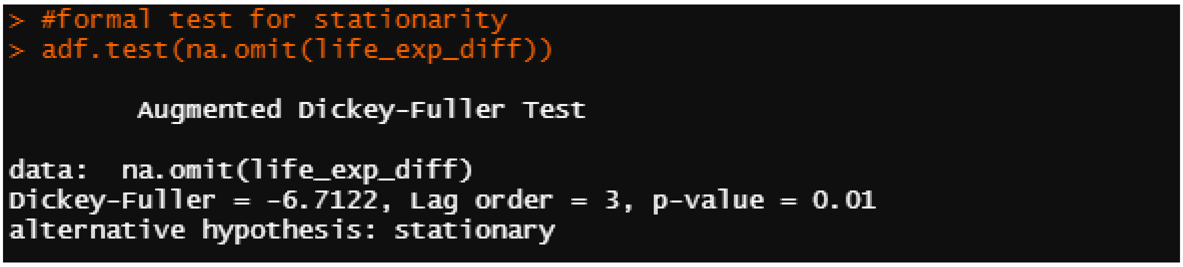


```
#differencing to make the data stationary
ndiffs(life_exp.xts)
life_exp_diff <- diff(life_exp.xts, differences = 2)
autoplot(life_exp_diff)
```



Test for Stationarity

Next, a test was carried out on the time-series data for stationarity called the Augmented Dickey-Fuller (ADF) test



```
> #formal test for stationarity
> adf.test(na.omit(life_exp_diff))

        Augmented Dickey-Fuller Test

data:  na.omit(life_exp_diff)
Dickey-Fuller = -6.7122, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

In the image above, the null hypothesis is that the image is not stationary, with the p-value (0.01) being less than 0.05, the null hypothesis was rejected, meaning the data is now stationary.

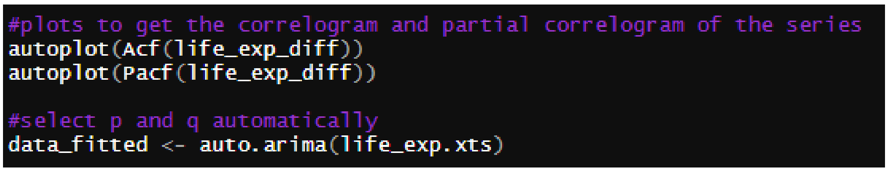
Fitting the ARIMA model (Correlogram & Partial correlogram)

The ARIMA model uses the function ARIMA(p,d,q) where:

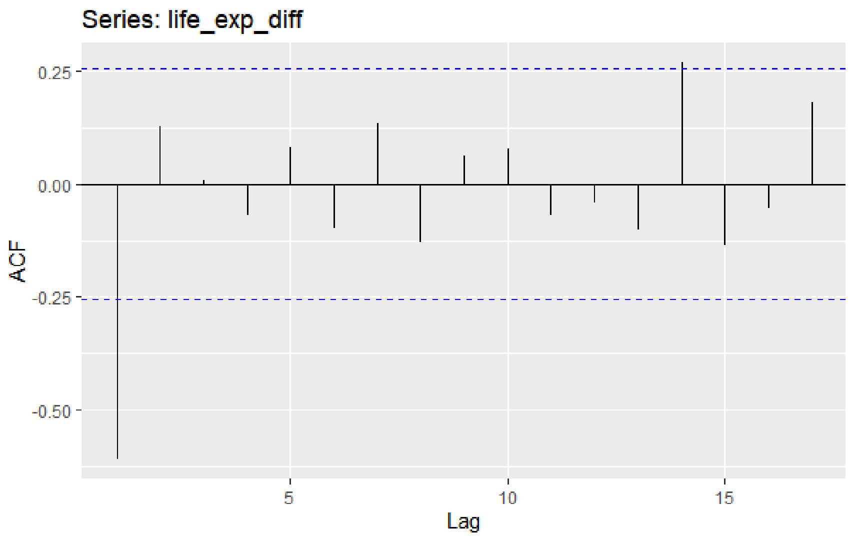
- p = number of autoregressive terms
- d = differences needed for data stationarity
- q = number of lagged forecast errors

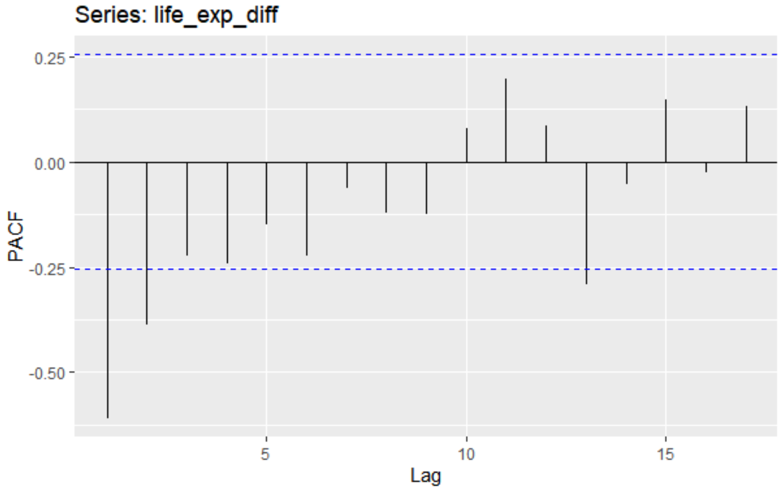
d is already known, but p and q need to be derived.

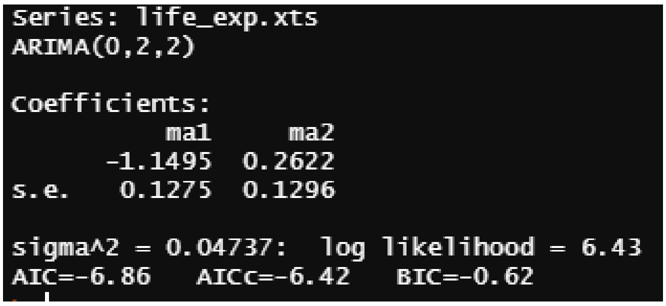
To do this, the correlogram and partial correlogram of the series are plotted, and the ‘auto arima’ function is utilized:



```
#plots to get the correlogram and partial correlogram of the series
autoplot(Acf(life_exp_diff))
autoplot(Pacf(life_exp_diff))

#select p and q automatically
data_fitted <- auto.arima(life_exp.xts)
```





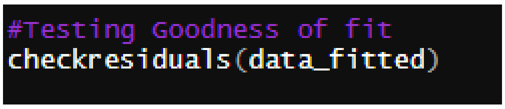


```
Series: life_exp.xts
ARIMA(0,2,2)

Coefficients:
          ma1     ma2
      -1.1495  0.2622
s.e.   0.1275  0.1296

sigma^2 = 0.04737:  log likelihood = 6.43
AIC=-6.86   AICc=-6.42   BIC=-0.62
```

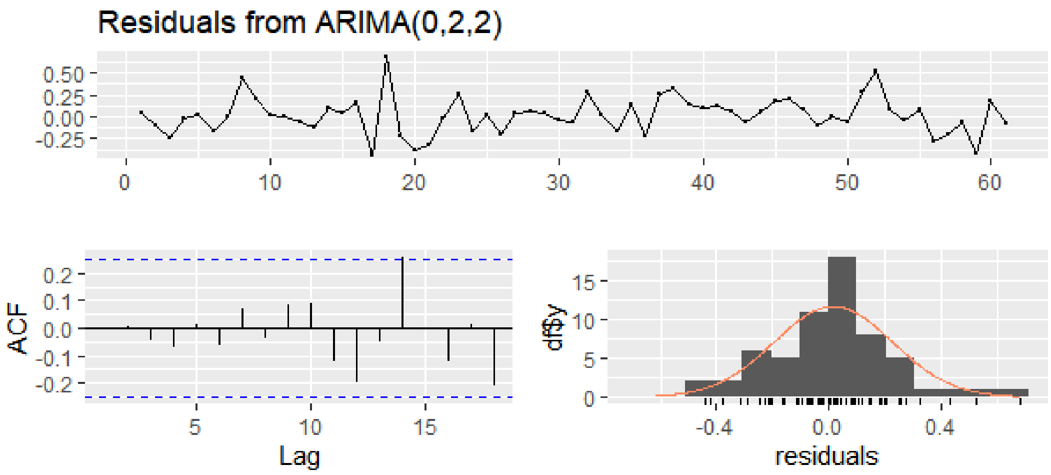
The image above shows that the regression or lag order is not needed, the differencing should be done to the order of 2, and an MA residual order of 2.

Goodness of fit test

This test evaluates how much the sample data fits a distribution gotten from a normally distributed population. The model needs to have normally distributed residuals, and the autocorrelation should be identical to white noise (zero lags). View code & result:



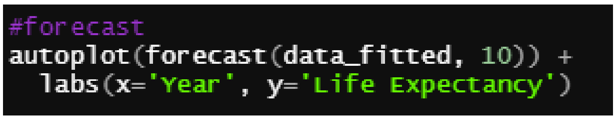
```
#Testing Goodness of fit
checkresiduals(data_fitted)
```



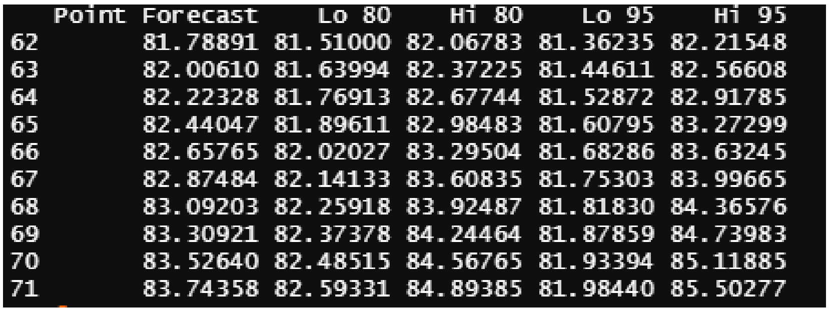
The residuals across the dataset, shown as the top plot, are shown to have a near-constant mean (stationary), the next graph (bottom left) shows no autocorrelation of the residuals (no significant peak), and the histogram shows an approximately normal distributed residuals. This all shows that the fit of the model is good.

Forecast using ARIMA

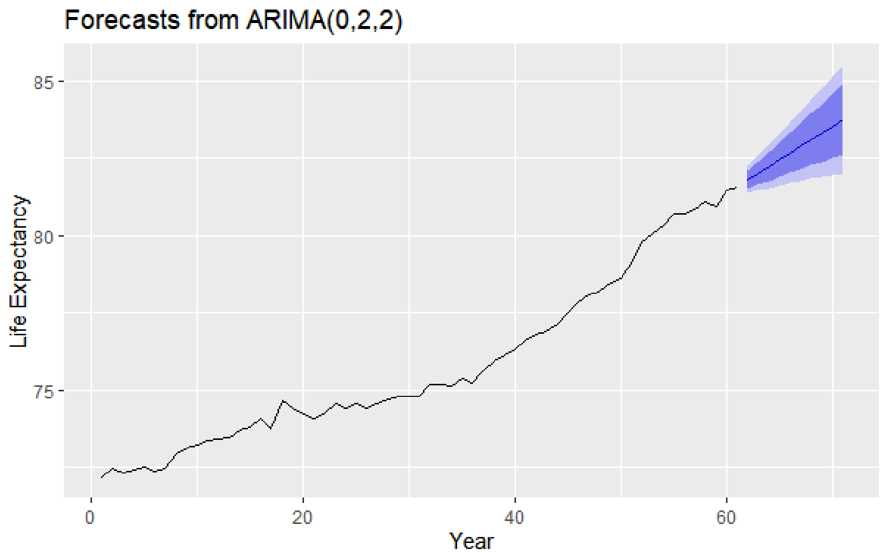
A forecast for the next 10 years was done, and the code, the table showing the predicted values with the upper and lower confidence values, and the plotted graph are as shown below:



```
#forecast
autoplot(forecast(data_fitted, 10)) +
  labs(x='Year', y='Life Expectancy')
```

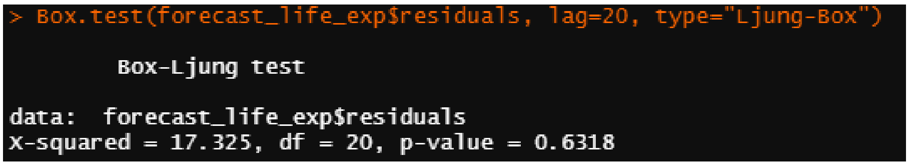


```
   Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
62       81.78891 81.51000 82.06783 81.36235 82.21548
63       82.00610 81.63994 82.37225 81.44611 82.56608
64       82.22328 81.76913 82.67744 81.52872 82.91785
65       82.44047 81.89611 82.98483 81.60795 83.27299
66       82.65765 82.02027 83.29504 81.68286 83.63245
67       82.87484 82.14133 83.60835 81.75303 83.99665
68       83.09203 82.25918 83.92487 81.81830 84.36576
69       83.30921 82.37378 84.24464 81.87859 84.73983
70       83.52640 82.48515 84.56765 81.93394 85.11885
71       83.74358 82.59331 84.89385 81.98440 85.50277
```



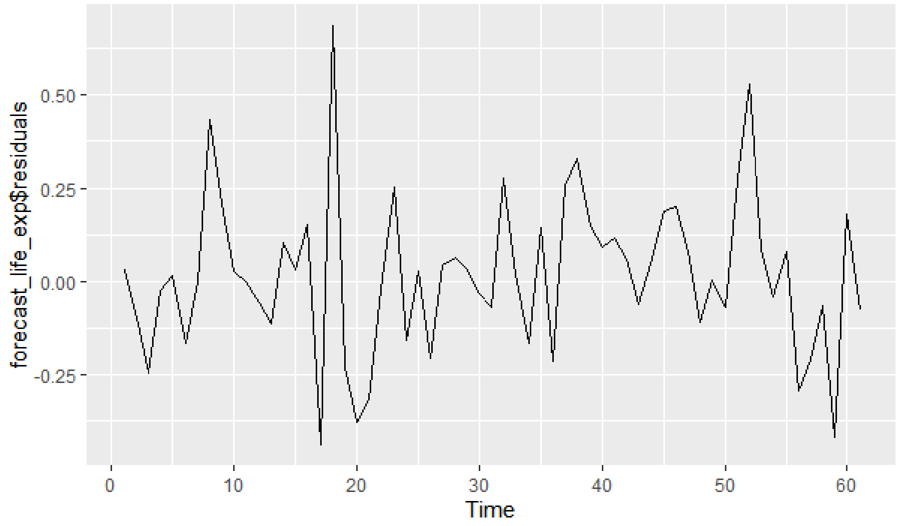
Forecast errors (Ljung-Box Test)

A Box-test was done for the Arima model, from the result shown, with the p-value at 0.6318, it can be concluded that there is very little evidence for non-zero autocorrelations in the forecast errors at lags 1-20.



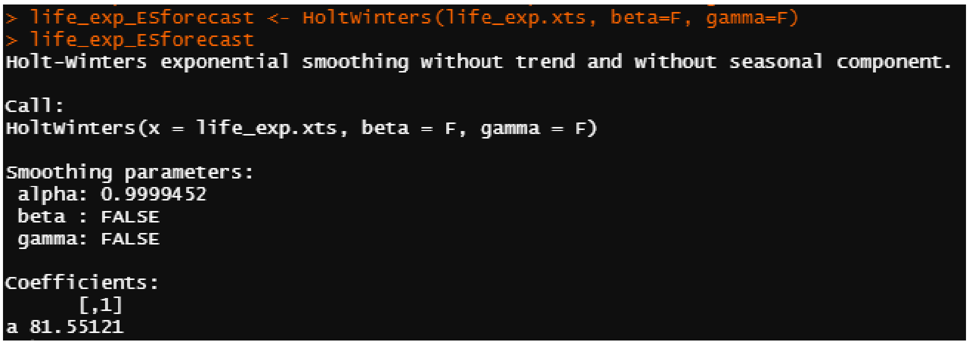
```
> Box.test(forecast_life_exp$residuals, lag=20, type="Ljung-Box")

        Box-Ljung test

data:  forecast_life_exp$residuals
X-squared = 17.325, df = 20, p-value = 0.6318
```

A plot of the residuals shows a normal distribution and constant variance. See below:



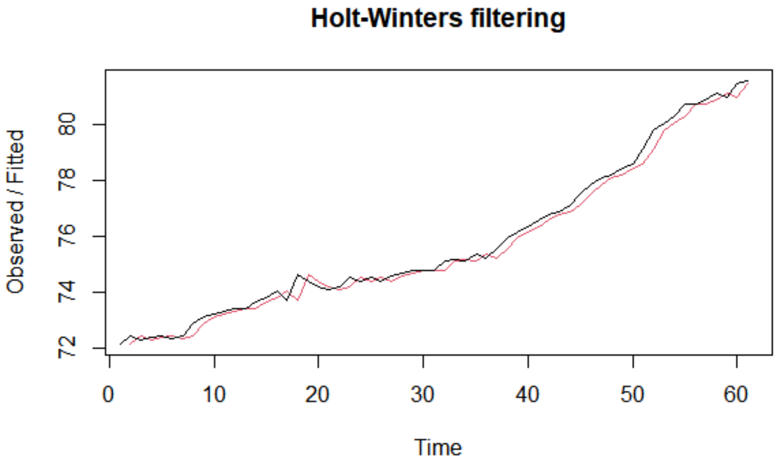
4.4.4 Using the Exponential Smoothing Time series model

The Simple Exponential Smoothing forecast model was also implemented, and the Holt-Winters function was used.

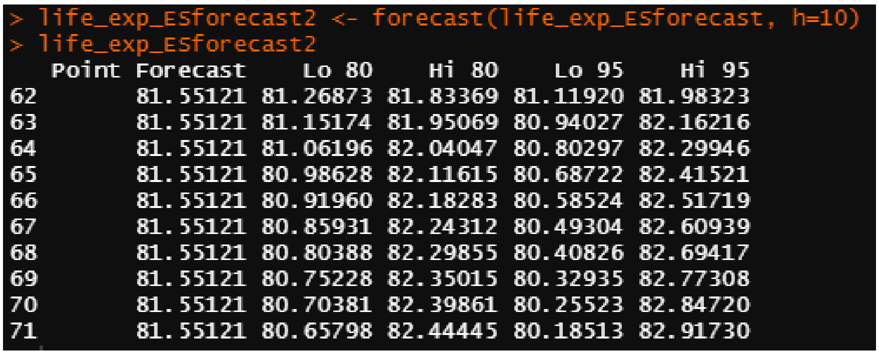


```
> life_exp_ESforecast <- HoltWinters(life_exp.xts, beta=F, gamma=F)
> life_exp_ESforecast
Holt-Winters exponential smoothing without trend and without seasonal component.

Call:
HoltWinters(x = life_exp.xts, beta = F, gamma = F)

Smoothing parameters:
 alpha: 0.9999452
 beta : FALSE
 gamma: FALSE

Coefficients:
      [,1]
a 81.55121
```

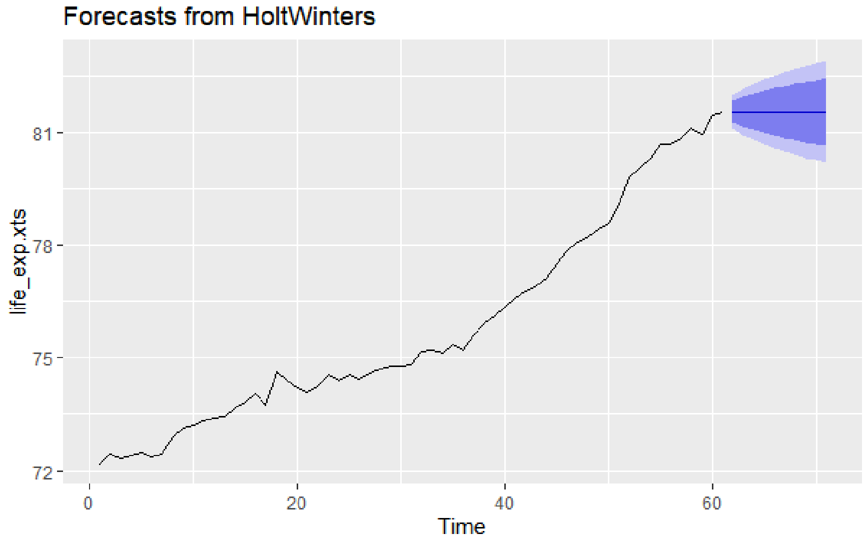
Next, the original time series was plotted against the forecasts (depicted by the red line) as shown below:



The forecast for the life expectancy in Denmark over the next 10 years according to the data, showing forecast values with the upper & lower confidence level, and the visual plot is as shown:



```
> life_exp_ESforecast2 <- forecast(life_exp_ESforecast, h=10)
> life_exp_ESforecast2
   Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
62       81.55121 81.26873 81.83369 81.11920 81.98323
63       81.55121 81.15174 81.95069 80.94027 82.16216
64       81.55121 81.06196 82.04047 80.80297 82.29946
65       81.55121 80.98628 82.11615 80.68722 82.41521
66       81.55121 80.91960 82.18283 80.58524 82.51719
67       81.55121 80.85931 82.24312 80.49304 82.60939
68       81.55121 80.80388 82.29855 80.40826 82.69417
69       81.55121 80.75228 82.35015 80.32935 82.77308
70       81.55121 80.70381 82.39861 80.25523 82.84720
71       81.55121 80.65798 82.44445 80.18513 82.91730
```



4.4.5 *Similar Research:*

ARIMA model:

(Torri & Vaupel, 2012) in their journal article, use the univariate ARIMA model to forecast the life expectancies in the US and Italy for 134 and 73 years respectively. The Dickey-Fuller test was used to test for stationarity of the data. The results obtained generated stochastic forecasts of the indicator, life expectancy, with manageable error bands.

Exponential Smoothing model:

This article (Shang et al., 2011) uses the exponential smoothing state space models for the forecasting of the principal component scores. The results showed a weighting that favored the more recent data as a result of the exponential smoothing done.

4.5 Comparative Analysis

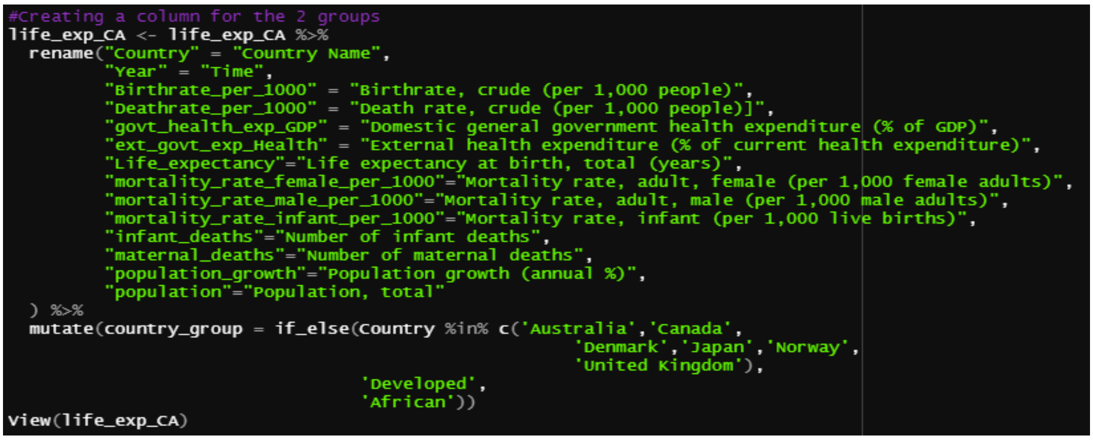
The main hypothesis testing approaches used in this section are the

1. **Mann-Whitney (Wilcoxon rank sum) Test:** This is the ideal test to be used when the data is continuous and not normally distributed.
2. **Independent Two Sample T-test result:** This can also be used in cases where the dependent variable is categorical and has 2 values. Also, when the samples being tested are independent of each other.
3. **Paired T-test:** This is used when the variables being tested are taken out from the same observations or sample space. In this case, the variables are paired or dependent on each other.

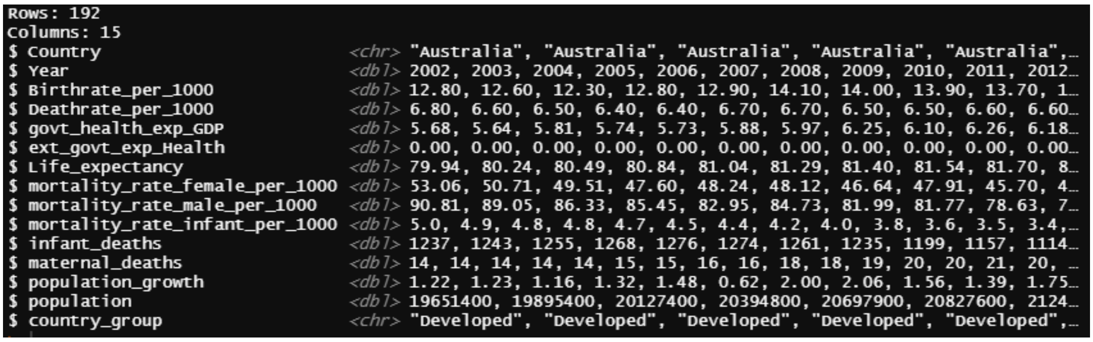
Further explanations and analysis are shown below.

This report seeks to compare the life expectancy (healthcare system in extension) of 2 groups of countries namely the Developed and the African countries so the T-test is optimal.

First, a column showing which group the countries belong to was created as shown below:



```
#Creating a column for the 2 groups
life_exp_CA <- life_exp_CA %>%
  rename("Country" = "Country Name",
         "Year" = "Time",
         "Birthrate_per_1000" = "Birthrate, crude (per 1,000 people)",
         "Deathrate_per_1000" = "Death rate, crude (per 1,000 people)]",
         "govt_health_exp_GDP" = "Domestic general government health expenditure (% of GDP)",
         "ext_govt_exp_Health" = "External health expenditure (% of current health expenditure)",
         "Life_expectancy"="Life expectancy at birth, total (years)",
         "mortality_rate_female_per_1000"="Mortality rate, adult, female (per 1,000 female adults)",
         "mortality_rate_male_per_1000"="Mortality rate, adult, male (per 1,000 male adults)",
         "mortality_rate_infant_per_1000"="Mortality rate, infant (per 1,000 live births)",
         "infant_deaths"="Number of infant deaths",
         "maternal_deaths"="Number of maternal deaths",
         "population_growth"="Population growth (annual %)",
         "population"="Population, total"
  ) %>%
  mutate(country_group = if_else(Country %in% c('Australia','Canada',
                                                 'Denmark','Japan','Norway',
                                                 'United Kingdom'),
                                  'Developed',
                                  'African'))
View(life_exp_CA)
```



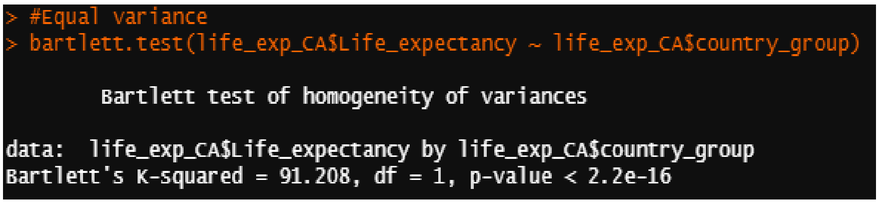
```
Rows: 192
Columns: 15
$ Country                        <chr> "Australia", "Australia", "Australia", "Australia", "Australia",…
$ Year                           <dbl> 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012…
$ Birthrate_per_1000             <dbl> 12.80, 12.60, 12.30, 12.80, 12.90, 14.10, 14.00, 13.90, 13.70, 1…
$ Deathrate_per_1000             <dbl> 6.80, 6.60, 6.50, 6.40, 6.40, 6.70, 6.70, 6.50, 6.50, 6.60, 6.60…
$ govt_health_exp_GDP            <dbl> 5.68, 5.64, 5.81, 5.74, 5.73, 5.88, 5.97, 6.25, 6.10, 6.26, 6.18…
$ ext_govt_exp_Health            <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00…
$ Life_expectancy                <dbl> 79.94, 80.24, 80.49, 80.84, 81.04, 81.29, 81.40, 81.54, 81.70, 8…
$ mortality_rate_female_per_1000 <dbl> 53.06, 50.71, 49.51, 47.60, 48.24, 48.12, 46.64, 47.91, 45.70, 4…
$ mortality_rate_male_per_1000   <dbl> 90.81, 89.05, 86.33, 85.45, 82.95, 84.73, 81.99, 81.77, 78.63, 7…
$ mortality_rate_infant_per_1000 <dbl> 5.0, 4.9, 4.8, 4.8, 4.7, 4.5, 4.4, 4.2, 4.0, 3.8, 3.6, 3.5, 3.4,…
$ infant_deaths                  <dbl> 1237, 1243, 1255, 1268, 1276, 1274, 1261, 1235, 1199, 1157, 1114…
$ maternal_deaths                <dbl> 14, 14, 14, 14, 14, 15, 15, 16, 16, 18, 18, 19, 20, 20, 21, 20, …
$ population_growth              <dbl> 1.22, 1.23, 1.16, 1.32, 1.48, 0.62, 2.00, 2.06, 1.56, 1.39, 1.75…
$ population                     <dbl> 19651400, 19895400, 20127400, 20394800, 20697900, 20827600, 2124…
$ country_group                  <chr> "Developed", "Developed", "Developed", "Developed", "Developed",…
```

4.5.1 *Test prerequisites*

To carry out a valid T-test, some data attribute conditions need to be known. These include:

Equality of variance of the 2 groups

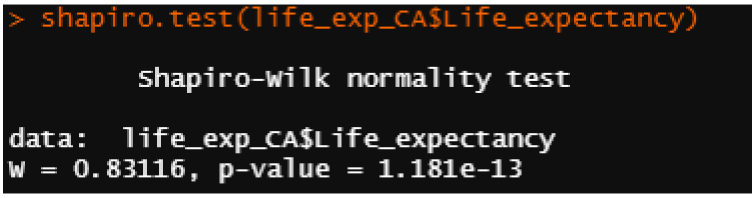
This was done using the Bartlett test. The results show a p-value less than 0.05 means the null hypothesis (the variances are equal) is rejected. So the variance is not equal.



```
> #Equal variance
> bartlett.test(life_exp_CA$Life_expectancy ~ life_exp_CA$country_group)

	Bartlett test of homogeneity of variances

data:  life_exp_CA$Life_expectancy by life_exp_CA$country_group
Bartlett's K-squared = 91.208, df = 1, p-value < 2.2e-16
```

Test for normal distribution

This was tested using the Shapiro-Wilk test. The p-value shown below is less than 0.05 which implies that the data is not distributed normally.

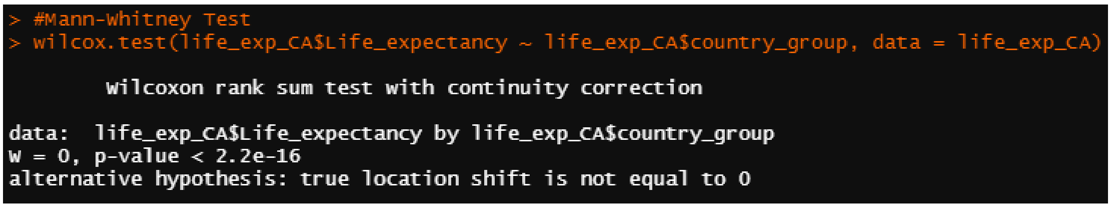


```
> shapiro.test(life_exp_CA$Life_expectancy)

	Shapiro-Wilk normality test

data:  life_exp_CA$Life_expectancy
W = 0.83116, p-value = 1.181e-13
```

Since the data is continuous and not normally distributed, the ideal method to test the hypothesis is the Mann-Whitney test.

4.5.2 Mann-Whitney (Wilcoxon rank sum) Test

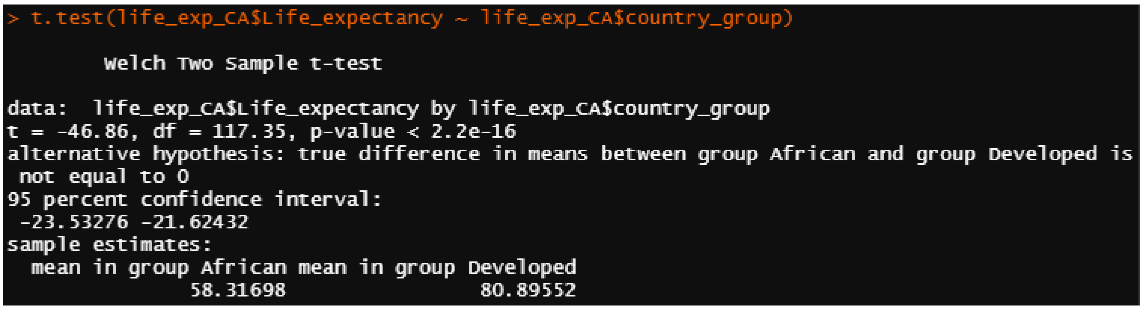
This test result shown in the image below shows a p-value less than 0.05, therefore, the null hypothesis is rejected, which means there is a significant difference between the life expectancy of the 2 groups of countries.



```
> #Mann-Whitney Test
> wilcox.test(life_exp_CA$Life_expectancy ~ life_exp_CA$country_group, data = life_exp_CA)

	Wilcoxon rank sum test with continuity correction

data:  life_exp_CA$Life_expectancy by life_exp_CA$country_group
W = 0, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

4.5.3 Independent Two Sample T-test result

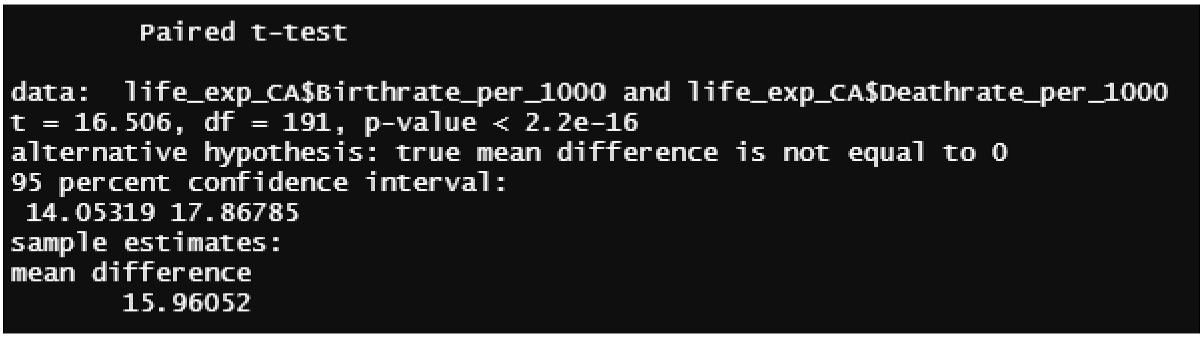
Next, the T-test was run. Afterward, the results show a p-value less than 0.05 which means that the null hypothesis, which states that the difference in means between the 2 groups is equal to 0. Therefore, statistically, there is a marked difference between the means of both groups:



```
> t.test(life_exp_CA$Life_expectancy ~ life_exp_CA$country_group)

	Welch Two Sample t-test

data:  life_exp_CA$Life_expectancy by life_exp_CA$country_group
t = -46.86, df = 117.35, p-value < 2.2e-16
alternative hypothesis: true difference in means between group African and group Developed is
 not equal to 0
95 percent confidence interval:
 -23.53276 -21.62432
sample estimates:
  mean in group African mean in group Developed
               58.31698                80.89552
```

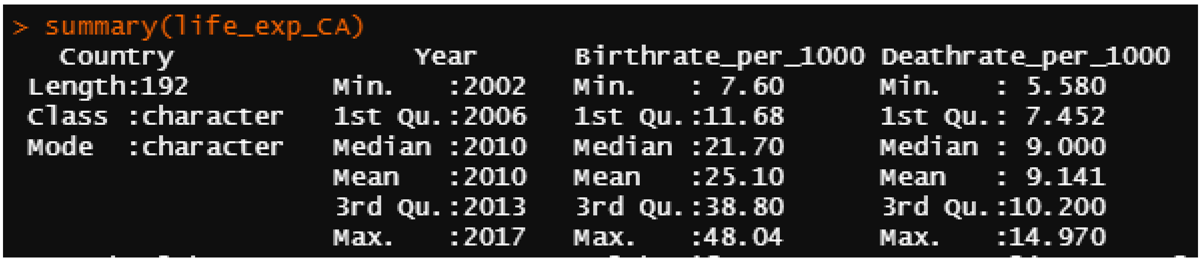
4.5.4 Paired T-test

This is a comparison between the birth rate and the death rate. Since this is done on the same sample, it is specified to be a paired test. The results below show a p-value less than 0.05, which means the null hypothesis was rejected and there is a significant difference (15.96052) between the mean birth and death rate.



```
        Paired t-test

data:  life_exp_CA$Birthrate_per_1000 and life_exp_CA$Deathrate_per_1000
t = 16.506, df = 191, p-value < 2.2e-16
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 14.05319 17.86785
sample estimates:
mean difference 
       15.96052 
```

A summary function shows that birthrate has the higher mean among the 2 indicators:



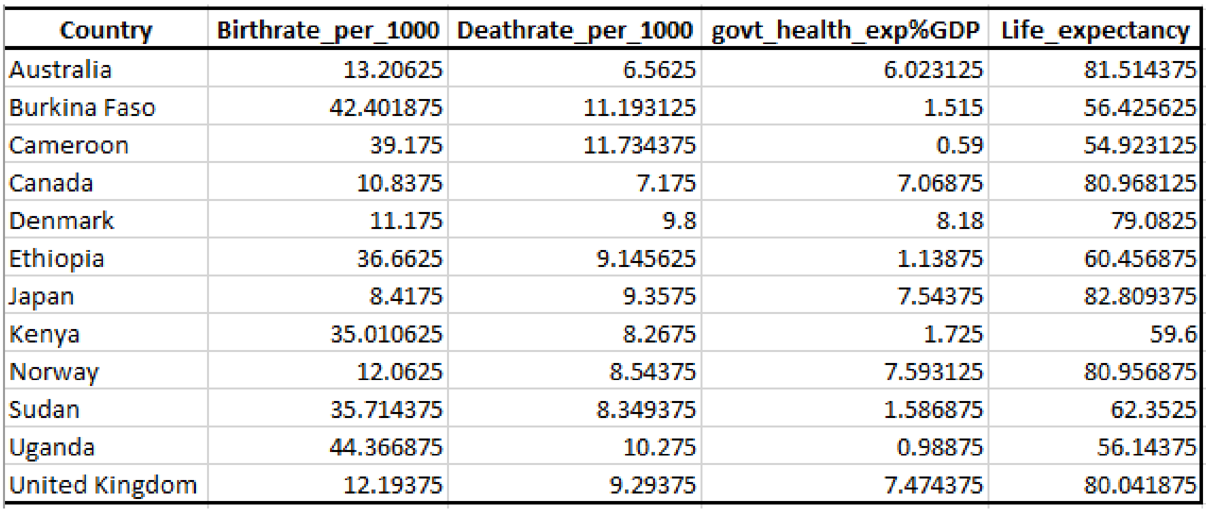
```
> summary(life_exp_CA)
   Country               Year      Birthrate_per_1000 Deathrate_per_1000
 Length:192         Min.   :2002   Min.   : 7.60      Min.   : 5.580    
 Class :character   1st Qu.:2006   1st Qu.:11.68      1st Qu.: 7.452    
 Mode  :character   Median :2010   Median :21.70      Median : 9.000    
                    Mean   :2010   Mean   :25.10      Mean   : 9.141    
                    3rd Qu.:2013   3rd Qu.:38.80      3rd Qu.:10.200    
                    Max.   :2017   Max.   :48.04      Max.   :14.970    
```

5. Discussion

As shown in the introductory section, the life expectancy (at birth) of the people in a country can be regarded as a true test of the state of the healthcare system of that country.

The primary objective of this study was to compare the health systems of the African group of countries to the Developed ones using life expectancy as the measuring indicator. While the secondary objective sought to establish a disparity between the birth and death rates of the countries.

The results of the preliminary descriptive statistics shown below, focusing mainly on the mean, show a marked difference in some of the key indicators. It can be seen that life expectancy and government health expenditure are higher in the developed countries than in the African group, but the birthrate in the African group is higher than that of the developed countries. The birth rate can be seen to be higher than the death rate in all of the groups selected.



Country	Birthrate_per_1000	Deathrate_per_1000	govt_health_exp%GDP	Life_expectancy
Australia	13.20625	6.5625	6.023125	81.514375
Burkina Faso	42.401875	11.193125	1.515	56.425625
Cameroon	39.175	11.734375	0.59	54.923125
Canada	10.8375	7.175	7.06875	80.968125
Denmark	11.175	9.8	8.18	79.0825
Ethiopia	36.6625	9.145625	1.13875	60.456875
Japan	8.4175	9.3575	7.54375	82.809375
Kenya	35.010625	8.2675	1.725	59.6
Norway	12.0625	8.54375	7.593125	80.956875
Sudan	35.714375	8.349375	1.586875	62.3525
Uganda	44.366875	10.275	0.98875	56.14375
United Kingdom	12.19375	9.29375	7.474375	80.041875

The correlation analysis conducted revealed a strong positive relationship (coefficient of 0.937) between government health expenditure and life expectancy, which means that as the health expenditure was increasing, life expectancy was also increasing. Conversely, this analysis reveals a strong negative relationship between life expectancy and mortality rates of females, males, and infants (correlation coefficient of -0.981, -0.978, and -0.974 respectively), which means as the former is increasing, the latter is reducing. This is logical as mortality is a measure of the death rate and should go in the opposite direction to life expectancy.

For regression analysis, 3 variations were employed. These include the ordinary least square regression method which is ideal when there are several combinations of variables to be considered. The backward stepwise regression method, a time-saver method that automatically tests various model combinations of variables by starting with a model containing all the variables, calculating the AIC score, then removing a variable, and iterating the process until it returns the model with the least AIC value. The robust method, which takes outliers and other influencing observations into consideration, was also utilized.

Of all the 3 regression models applied, the stepwise regression method was the highest performing as it gave the highest adjusted R-squared value of 0.9965, It also showed the least amount of residual error at 0.6964

Next, a time series analysis was done with a key focus on one of the developed countries, Denmark. Additional data was utilized, increasing the period of focus to 60 years to get more accurate predictions.

The time series forecast methods utilized were the **ARIMA** method and the **Exponential smoothing time series model**. The ARIMA model showed a result with a manageable amount of error while the exponential smoothing model showed a weighing pattern that favors the more recent data because of the exponential smoothing done.

The hypotheses tested in the comparative analysis part were:

1. There is a difference in the life expectancies of the 2 groups of countries studied.
2. The birth rate and death rate in the dataset used in this report are not equal.

The results gotten show that both hypotheses stated are not rejected.

A major limitation faced during this research is the unavailability of data in most of the African countries selected as a result of poor data collection.

6. Conclusion

There is a telling lack of investment in the health sector by the government of the selected African countries and this is evident by the low numbers of the government health expenditure percentage. This, in turn, is a major contributor to the consistently relatively low life expectancy as compared to their counterparts in the Developed group of countries.

The descriptive analysis, particularly the mean showed that the life expectancy was higher in the developed country while the correlation analysis revealed the relationships between the life expectancy and the government health expenditure spending & mortality rates. The regression further reinforced these relationships and extracted the variables with the most significant effect on life expectancy, then the time series forecast showed a glimpse into the life expectancy pattern of Denmark in future years.

It was not surprising to see that the birth rate of the countries was generally higher than their death rates, while the mortality rates showed an inverse relationship with life expectancy.

The main research objective of comparing the life expectancy and hence the healthcare of the 2 groups of countries was a success and it was proven that the developed countries had a better healthcare performance than their African cohorts.

The secondary objective aimed at finding out the difference in birth and death rate was also achieved and revealed that the birth rate was indeed higher than the rate of death.

References

Aday, L. A., & Andersen, R. (2021). A framework for the study of access to medical care. *Health services research*, 208.

Analyzing healthcare open data with power BI. (2017, March 24). Retrieved from Microsoft Power BI Community: https://community.powerbi.com/t5/Data-Stories-Gallery/Analyzing-Healthcare-Open-Data-with-Power-BI/m-p/148492

Asiskovitch, S. (2010). Gender and health outcomes: The impact of healthcare systems and their financing on life expectancies of women and men. *Social Science & Medicine*, 886-895.

Berger, M. C., & Messer, J. (2002). Public financing of health expenditures, insurance, and health outcomes. *Applied Economics, 34*(17), 2105-2113. https://doi.org/https://doi.org/10.1080/00036840210135665

Burgers, G. R., Dobrow, M., Minhas, R., Wendt, C., Cohen, A. B., & Luxford, K. (2014). Healthcare system performance improvement: a comparison of key policies in seven high-income countries. *Journal of health organization and management*.

CD-Editorial. (2022, July 5). *Top 7 tips for designing effective power BI dashboards*. Retrieved from Calculate Data: https://www.calculatedata.com/designing-effective-power-bi-dashboard-top-7-tips/

Compton, J. (2022, May 19). *Top 10 tips for designing power BI dashboards*. Retrieved from Pragmatiq: https://www.pragmatiq.co.uk/top-10-tips-for-designing-power-bi-dashboards/

Croxton, F. E., & Stein, H. (1932). Graphic Comparisons by Bars, Squares, Circles, and Cubes. *Journal of the American Statistical Association, 27*(177), 54-60.

Davidiseminger. (n.d.). *Microsoft Learn: Build skills that open doors in your career*. Retrieved from Measures in power BI desktop - Power BI: https://learn.microsoft.com/en-us/power-bi/transform-model/desktop-measures

Deshpande, N., Kumar, A., & Ramaswami, R. (2014). *The effect of national healthcare expenditure on life expectancy.*

Hakim, A. R., Warsito, B., & Yasin, H. (2020). Live expectancy modelling using spatial durbin robust model. *In Journal of Physics: Conference Series, 1655*(1), 012098.

Hassouna, F., & Al-Sahili, K. (2020). Practical minimum sample size for road crash time-series prediction models. *Advances in civil engineering*.

Jaba, E., Balan, C. B., & Robu, I.-B. (2014). The relationship between life expectancy at birth and health expenditures estimated by a cross-country and time-series analysis. *Procedia Economics and Finance, 15*, 108-114. https://doi.org/https://doi.org/10.1016/S2212-5671(14)00454-7

Kaya Samut, P., & Cafri, R. (2016). Analysis of the efficiency determinants of health systems in OECD countries by DEA and panel tobit. *Social Indicators Research, 129*(1), 113-132.

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American statistical association, 87*(419), 659-671. https://doi.org/https://doi.org/10.1080/01621459.1992.10475265

Liu, T., Yang, S., Peng, R., & Huang, D. (2021). A Geographically Weighted Regression Model for Health Improvement: Insights from the Extension of Life Expectancy in China. *Applied Sciences, 11*(5), 2022.

Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks, CA: Sage Publications. Retrieved from https://stats.oarc.ucla.edu/r/dae/tobit-models/

MaggiesMSFT. (n.d.). *Create report tooltip pages in power BI - Power BI*. Retrieved from Microsoft Learn: Build skills that open doors in your career: https://learn.microsoft.com/en-us/power-bi/create-reports/desktop-tooltips?tabs=powerbi-desktop

Owusu, P. A., Sarkodie, S. A., & Pedersen, P. A. (2021). Relationship between mortality and health care expenditure: Sustainable assessment of health care system. *Plos one, 16*(2). https://doi.org/https://doi.org/10.1371/journal.pone.0247413

Pacáková, V., Jindrová, P., & Zapletal, D. (2016). Comparison of Health Care Results in Public Health Systems of European Countries. *In European Financial Systems 2016: proceedings of the 13th International Scientific Conference.* Masarykova univerzita.

Retail analysis -dynamic Measure&TopN selection. (2022, November 9). Retrieved from Microsoft Power BI Community: https://community.powerbi.com/t5/Data-Stories-Gallery/Retail-Analysis-Dynamic-Measure-amp-TopN-Selection/td-p/121080

Sales analysis dashboard. (2022, December 4). Retrieved from Microsoft Power BI Community: https://community.powerbi.com/t5/Data-Stories-Gallery/Sales-Analysis-Dashboard/td-p/1567208

Samadder, S. R., Nagesh Kumar, D., & Holden, N. M. (2014). An empirical model to predict arsenic pollution affected life expectancy. *Population and Environment, 36*(2), 219-233. https://doi.org/https://doi.org/10.1007/s11111-014-0212-5

Shang, H. L., Booth, H., & Hyndman, R. J. (2011). Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods. *Demographic Research, 25*, 173–214. https://doi.org/https://doi.org/10.4054/demres.2011.25.5

Sokolowski, A. (1999). Regional Differences in Living Standards in Eastern European. *Research Support Scheme, Open Society*.

Stiefel, M. C., Perla, R. J., & Zell, B. L. (2010). A healthy bottom line: healthy life expectancy as an outcome measure for health improvement efforts. *The Milbank Quarterly, 88*(1), 30-53. https://doi.org/https://doi.org/10.1111/j.1468-0009.2010.00588.x

Torri, T., & Vaupel, J. W. (2012). Forecasting life expectancy in an international context. *International Journal of Forecasting, 28*(2), 519-531.

WHO. (2000). *World health report: health systems improving performance.* Geneva, Switzerland.

Winter, J. (2021, May 13). *To measure table or NOT to measure table?* Retrieved from Greyskull Analytics: https://greyskullanalytics.com/to-measure-table-or-not-to-measure-table/

WOW2022 Week43 | Power BI: Advanced conditional formatting. (2022, October 30). Retrieved from Microsoft Power BI Community: https://community.powerbi.com/t5/Data-Stories-Gallery/WOW2022-Week43-Power-BI-Advanced-Conditional-Formatting/m-p/2872916