Final report

**MIT 5032 R Analytics**

Life Expectance

(WHO)

Professor

**Dr. Naveen Kumar**

Team

**Finley, Ben N**

**Georgestone, Irene K**

**Jasthi, Srija**

**Russell, Braum P**

Table of Contents

[PROBLEM STATEMENT 3](#_Toc1139874841)

[DATA SOURCES AND DESCRIPTION 3](#_Toc1063680759)

[DATA PREPROCESSING 3](#_Toc1507115403)

[HYPOTHESES AND INTUITION 4](#_Toc784319861)

[SUMMARY AND CONCLUSIONS OF HYPOTHESES 21](#_Toc1872248521)

[KEY LEARNINGS AND LIMITATIONS 22](#_Toc600300101)

[REFERENECES 22](#_Toc717902167)

# **PROBLEM STATEMENT**

By utilizing R analytics, this study seeks to create a spectrum of correlations related to life expectancy and evaluate the interconnections between these variables. This will aim to enhance our understanding of factors influencing life expectancy in developing and developed countries.

# DATA SOURCES AND DESCRIPTION

**The team got the dataset from Kaggle. This data contains information from the World Health Organization's (WHO), about life expectancy in 193 countries. This dataset, which consists of 2938 rows and 22 columns containing 20 predicting variables, serves as the foundation for testing theories about how social determinants, mortality rates, immunization rates, and economic factors affect life expectancy.**

**The dataset was selected based on its compliance with the project's requirements. At least three of the data types discussed in the class are present in the data. The data is categorical ordinal, interval continuous, and interval discrete. The figure below shows a snippet of data from Kaggle.**

**A screenshot of a graph

Description automatically generated**

# **DATA PREPROCESSING**

1. **Missing Value Handling:** 
   * Initial checks revealed missing values in a number of different columns.
   * To ensure data completeness without distorting statistical measures, missing values for numerical columns were substituted with their corresponding means.
   * To maintain the integrity of the distributions of categorical data, mode values were used to replace missing values in categorical columns.
2. **Data Integrity Assurance:**

* By taking these actions, the dataset was prepared for analysis by ensuring that missing values were properly addressed.

**3. Frequency Distribution and Summary Statistics:**

* For numerical columns, summary statistics were produced, offering crucial information on central distributions and tendencies.
* The frequency distributions for the categorical columns provided a clear picture of the incidence of different categorical values.

1. **Column Removal**:

* To streamline the data attributes, the 'BMI' column was eliminated from the dataset because it may not be as relevant for analysis.

These preprocessing steps are a first step that guarantees the dataset is complete, clean, and ready for further analysis. Missing value handling, summary statistics, and column refinement all help to improve the quality of data, which makes it easier to conduct accurate and well-informed analyses of the health-related variables influencing life expectancy in various countries.

# **HYPOTHESES AND INTUITION**

Hypothesis **1:**

Analyzing life expectancy across continents and development statuses reveals a prevalent trend: developed countries consistently showcase higher life expectancies compared to their developing counterparts, indicating that countries categorized as developed consistently exhibit longer life expectancies than those classified as developing.

**Intuition 1:**

This trend is strongly linked to the advanced healthcare infrastructure and overall superior health outcomes found in developed nations. Factors such as robust healthcare systems, better access to medical facilities, and advanced disease management contribute significantly to the higher life expectancies witnessed in these countries.

**Total instances of developed and developing countries**

A graph of different colored bars

Description automatically generated

The bar chart showcases the count of developing versus developed countries by continent, reflecting data from the years 2000 to 2015. It is evident from the graph that:

* **Africa** has an overwhelming majority of developing countries, with 864 instances counted, and a small number of developed countries. This significant difference is indicative of the development challenges faced by the continent and could be a contributing factor to the lower life expectancy observed in the previous visualization.
* **Asia** follows with 704 instances of developing countries and a moderate count of developed ones. Asia's large population and diversity of economies might explain the high number of developing countries.
* **Europe** presents a contrasting picture, with a higher count of developed countries (416) compared to developing ones (210). This aligns with Europe's higher median life expectancy and indicates a strong association between development status and health outcomes.
* **The Americas and Oceania** show a mixed composition, with Oceania having a lower overall count but a relatively balanced distribution between developed and developing countries.

The graph clearly delineates the disparity between continents in terms of development status. Africa's high count of developing nations could be directly linked to its lower life expectancy figures, while Europe's dominance in the developed category correlates with its higher life expectancy trends. This visualization further substantiates the hypothesis that a country's development status encompassing economic, social, and health infrastructure has a profound impact on the life expectancy of its population. The stark contrast between Africa and Europe in terms of the number of developed and developing countries provides a clear depiction of the global development divide and its implications for public health.

**Median life expectancy by continents**

A graph with different colored squares

Description automatically generated

The box plot visualization illustrates life expectancy data across continents:

* **Africa:** The box plot for Africa displays a median life expectancy below 60 years. The length of the box and the whiskers suggest a wide range of life expectancies among African countries, with several outliers indicating that a few countries have life expectancies that are significantly lower than the rest. This variance could be due to differences in access to healthcare, prevalence of infectious diseases, and socio-economic factors characteristic of many developing nations within the continent.
* **America:** The median life expectancy for the Americas is above 70 years, reflecting a better health outcome than Africa. The range is narrower than Africa’s, implying less variability in life expectancy. However, it's important to note that the Americas encompasses both developed and developing countries, which may contribute to the spread of data points.
* **Asia:** With a median also above 70 years, Asia's life expectancy is similar to the Americas. The continent’s large size and diversity mean that it includes both countries with high life expectancies and those with lower figures. The spread of life expectancies in Asia is quite extensive, reflecting this diversity.
* **Europe:** Europe stands out with the highest median life expectancy, surpassing 75 years. The box plot indicates a higher standard of living and well-developed healthcare systems typical of developed countries. The range is relatively tight, suggesting that there's less disparity among European countries compared to continents with more developing nations.
* **Oceania:** Oceania’s life expectancy is just under 70 years. The range is minimal, which could be due to the smaller number of countries within this dataset, or it could indicate that these countries have similar health outcomes.

Overall, the graph supports the hypothesis by showing a clear trend where continents with a higher number of developed countries, such as Europe, have a higher median life expectancy. In contrast, continents with a larger proportion of developing countries, like Africa, display lower median life expectancies and greater variability. This aligns with the assertion that development status, which includes factors such as economic stability, healthcare infrastructure, and education, is a significant determinant of life expectancy. The visualization provides empirical evidence that development status is correlated with health outcomes, as seen in the distribution and median values of life expectancy across the continents.

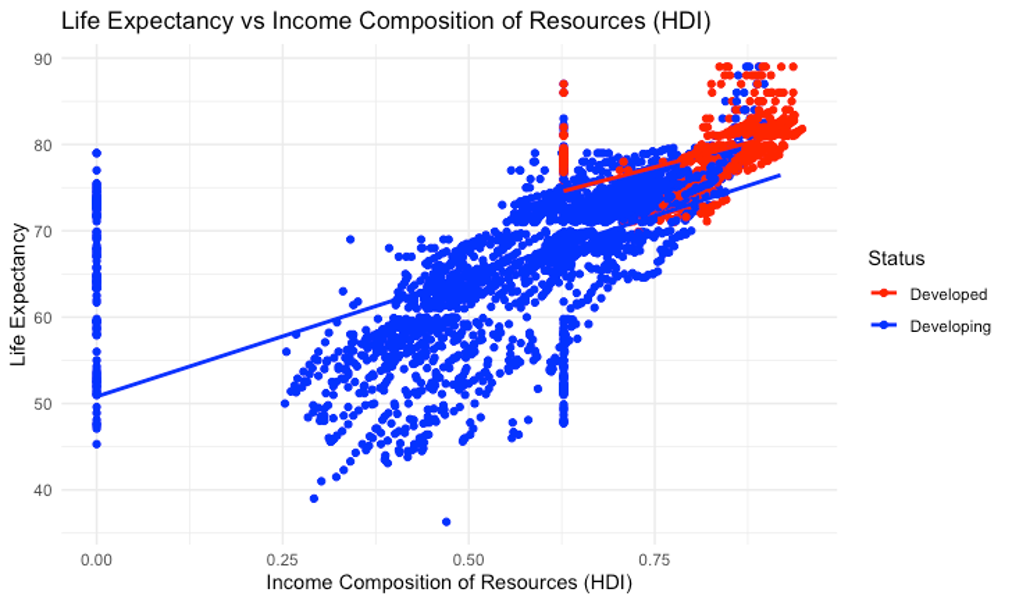
**Vaccination Rate effect on Life expectancy**

A diagram of a vaccination rate

Description automatically generated

Life expectancy tends to rise with increasing rates of Hepatitis B vaccination. The idea that developed nations have longer life expectancies due to higher vaccination rates is supported by this trend. The scatter plot makes a distinct trend evident: We can see from the plot that the majority of developed nations are concentrated near the top of the distribution for both life expectancy and vaccination rates. Strong healthcare systems, which are more prevalent in developed nations, are thought to improve health outcomes, and this pattern lends support to that theory. The significance of funding healthcare policies and infrastructure is highlighted by this analysis. Increasing immunization rates is one way developing nations can try to raise their citizens' average life expectancy.

**Income Composition effects of Life expectancy**



The upward slope of the regression line in the graph demonstrates a strong positive correlation between HDI and life expectancy, affirming the hypothesis that comprehensive development in developed nations is a major determinant of health outcomes. Red points, representing developed countries, are predominantly clustered in the upper-right quadrant, indicating high HDI and high life expectancy. Blue points, representing developing countries, are dispersed more widely, with many falling into the lower HDI and lower life expectancy areas. This visual analysis supports the hypothesis by illustrating that developed countries with advanced resources tend to have citizens with longer lifespans. Higher HDI reflects superior health outcomes, with developed nations leading in life expectancy, underscoring the vital role of comprehensive development in public health. Investment in human development, encompassing education, income, and healthcare, is crucial for enhancing population health and longevity.

Hypothesis 2:

Alcohol consumption rates will be higher in developing countries compared to developed countries and due to the higher alcohol usage, the corresponding life expectancy will be lower in developing countries.

Intuition 2:

Due to developing countries having more stress factors that cause anxiety and depression, individuals will use alcohol to cope with living in these conditions. Since alcohol is widely known for causing adverse health effects, having higher consumption rates will lead to having lower life expectancies.

Approach 2:

Stated in hypothesis 2 is two statements that needed to be determined valid in order to prove the hypothesis. First, we needed to determine that life expectancy is higher in developed countries compared to developing countries needed to be determined. The steps stated to approach this is listed below in the figure.

A white rectangle with black text

Description automatically generated

The second part of the hypothesis would be that alcohol consumption was indeed higher in developing countries. This basis of the hypothesis, if proven, can be used for fact that if life expectancy is higher in developed countries, this could be one of the determining factors. Steps of how this was analyzed is shown below.

A white background with black text

Description automatically generated

Analysis

\*\*For information, life expectancy is in units of years and alcohol consumption is in units of liters consumed per capita.

**A red squares with black text

Description automatically generated**

The bar chart above shows the comparison of life expectancy in developing countries to developed countries. For further information, there were a total of 512 developed countries and 2426 developing countries given in our data set. The median life expectancy for developing countries is 69.05 and the median life expectancy for developed countries was 79.25. This visual proves the first aspect that life expectancy is higher in developed countries than developing countries. The median point was used in this determination to account for outliers in the life expectancy year data. R code that calculated median variables for life expectancy in developed and developing countries is shown below.

A screen shot of a computer code

Description automatically generated

A graph of alcohol consumption

Description automatically generated

Shown above is a histogram that shows the frequency of how many times developed countries had alcohol consumption in certain liter consumption bins. The histogram shows that alcohol consumption for developed countries is left-skewed. This means that when we increase the bin including how many liters of alcohol are consumed per capita, the count of countries will increase. It can be seen that the mode alcohol consumption is 10 - 12.5 liters. 0-2.5 and 2.5-5 looks to be very low regarding the frequency and then the bins just continuously increase, until it reaches 12.5 -15 liters.

A graph of alcohol consumption

Description automatically generated

The second histogram (shown above) shows the frequency of how many times developing countries had alcohol consumptions in certain liter consumption bins. The histogram presents that alcohol consumption for developing countries is right skewed. Alcohol consumption data being right skewed means that when we increase the bin including how many liters of alcohol are consumed per capita, the count of countries will decrease. It can be seen that the mode alcohol consumption for developing countries is 0-1 liters. After this bin, the frequency looks to drop by almost 50% and continuously decreases except in 2 bins.

A screenshot of a computer code

Description automatically generated

A close-up of a black text

Description automatically generated

\*For further analysis, the two figures above show the median alcohol consumption for developed and developing countries and the R code that was created to perform the analysis.

From the analysis, hypothesis two will have to be rejected. It can be confirmed that life expectancy is higher in developed countries compared to developing countries, as we learned that the median life expectancy was 69.05 for developing years and 79.25 years for developed countries meaning life expectancy is around 10 years higher in developed countries than developing. It is then confirmed that alcohol consumption is not higher in developing countries than developed countries. The median alcohol consumption in developing countries is 2.950 liters of alcohol per capita and 10.205 liters of alcohol per capita for developed countries. Also, developed country alcohol consumption is right skewed and developing country alcohol consumption is left skewed. These are two ways that both confirm that alcohol consumptions is not higher in developing countries. The original reasoning was stated to be due to individuals having higher anxiety and depression rates leading to using alcohol to cope, but reasoning, not originally analyzed, this could be incorrect is due to the lack of local resources/stores in developing countries to purchase alcohol and the lack of funds to actually buy the liquor.

Hypothesis 3:

Developing countries in the West African Region have an increase in life expectancy over the years due to the influence of an increase in total health expenditure and a decrease in cases of measles.

Intuition 3:

The hypothesis assumes that there has been a growing focus on health awareness and increased funding in the West African Region over time. As a result, more resources are being allocated towards improving healthcare infrastructure. Enhanced healthcare typically leads to more effective disease management and a healthier population.

Approach 3:

Following the cleaning of the dataset, I further filtered the data to tailor it to the analysis of my hypothesis. The first thing I did was to create a vector variable for the list of West African countries. I then attempted to filter the data even further to subset West African countries between 2000 and 2015. I quickly realized I would need to install certain packages and libraries to successfully run the filter function. The purpose of this code is to prepare a dataset specific to West African countries over the 2000 and 2015 time, which would then use the ‘ggplot’ package and library for further analysis and visualization. The snippet below shows the codes used to execute this data processing stage.

A screenshot of a computer

Description automatically generated

Specific Outcomes and Insights

The precedent of my hypothesis assumed that West African countries have an increase in life expectancy over the years between 2000 and 2015. To ensure this, I started with an initial analysis to see if life expectancy indeed improved. The figure below shows the outcome of the analysis, which is denoted by the upward trend of the graph, showing that the average life expectancy in the region did in fact increase over time.

A graph with a line going up

Description automatically generated

Following this precedent, I then went ahead to compare and visualize the first independent variable over time. This comparison was between the number of measles count in west African countries over time. The aim of the heat map is to see if there are any major changes in the number of measles cases within the 15-year period. The visualization showed little to no changes, where countries stayed around the 500,000 or less counts of cases. There is a case of a drastic decrease in count of measles in Nigeria over the period, as is highlighted by the change in color from red to blue between 2000 and 2015. The figure below demonstrates this, and the code used to create the heatmap.

A screenshot of a graph

Description automatically generated

The same method was conducted to visualize the changes in the total expenditure spent on health. Like the measles count over the years, visualization in the previous figure, the total expenditure spent on health is well about 15% or less for these countries with most of them being below 10%. Sierra Leone and Liberia seem to have a higher percentage of total expenditure spent on health. The heatmap color gradient below shows this distribution of total expenditure over the years for each country.

A graph of a graph with numbers and a number of people

Description automatically generated with medium confidence

Though both graphs above are aesthetically pleasing visuals, they tell us little about the actual impact that measles and total expenditure spent on health, may have on the increasing life expectancy in West African region over the 15 year period. In order to get a better analysis, a scatter plot was graphed to show the relationship between these two independent variables and the life expectancy.

Using ggplot, a scatter plot that compares the measles count for all West African countries over the 15 year period was graphed.

A graph with black dots and a blue line

Description automatically generated

The graph depicts a negative linear correlation wherein as the number of measles count increases, the life expectancy decreases. We can deduce that measles outbreak does have a negative effect on the life expectancy in the West African region. The data points are concentrated in the life expectancy range of 50 to 65 years, at measles counts of below 250,000 counts. This number is high, especially compared to regions with developed countries like North America for example. Though a correlation, it does not imply that reduction in the number of measles causes an increase in life expectancy in the region. However, this analysis does support the hypothesis that a decrease in disease outbreaks, like measles, has an overall positive effect on the life expectancy of countries in the region.

The next part of this analysis compares the total expenditure spent on health versus life expectancy. Below is a scatterplot of the outcome.

A graph showing a line going up

Description automatically generated

The scatter plot above shows a downward sloping trend which suggests that as total expenditure increases, life expectancy decreases. However, this is not a typical relationship. One would expect the life expectancy of a country to increase as total expenditure spent on health increases. There is a cluster of data points concentrated between the 50 to 65 years life expectancy, where the total expenditure is mostly between 2.5% and 7%. Given this, one can deduce that health expenditure does not influence life expectancy due to several possible reasons. One of the reasons is what was shown in the heat map over total expenditure over years, being there is not enough increase over the years as expenditure stayed about the same. There may be a rise in other health or non-health factors that have a bigger influence. Increase in fundings does not necessarily mean more money was placed into health care expenditure. It is important to note that between the 2013 to 2015 period, the Ebola epidemic broke out in some West African countries. So even though there might have been a boost in fundings, the outbreak of other diseases might have influenced how much of those fundings go to Ebola related cases, or just prioritizing other basic needs of people that does not include health.

My hypothesis was not proven true. However, I took another step to calculate the correlation between measles and total expenditure with life expectancy and the image below shows the result.

A screenshot of a computer

Description automatically generated

The p value is less than 0.001 which means that both measles count and total expenditure are significant in the prediction of life expectancy. The adjusted R square also shows that about 13.5% of the change in life expectancy in the West African countries can be explained by measles and total expenditure. Even though my hypothesis was not fully correct, the analysis shows great insights, using ggplots, to understand how life expectancy is affected by measles and total expenditure spent on health in the region.

Hypothesis 4:

There is a higher life expectancy in developed and developing countries with higher GDP and more schooling years. Both developed and developing countries will be analyzed to look for stronger or weaker trends depending on the status of the country while analyzing these two trends of schooling years and GDP.

Intuition 4:

More schooling years and higher GDP are typically seen as trends that are associated with more developed countries. It would be interesting to analyze whether these factors impact the life expectancy in countries that are developing as well. It would reveal whether these trends have impacts on life expectancy.

Approach 4:

The first step in the approach for hypothesis four was creating a subset for the status of each country, which was developed and developing. Then, the ggplot library was used to create graphs for each status together and separately. This yielded three graphs each for the factors of GDP and schooling years. After that, a line of best fit was created for each graph and a correlation coefficient was produced for each with the cor() function. Then, analysis was done on the results and conclusions were made for the hypothesis. After all of this, the analysis felt like it was lacking, so a linear regression model was used to analyze GDP and schooling years in developed and developing countries to get a higher understanding of the variables and create a higher level of analysis. Below are some code snippets of the library, the subset, and an example of the code for creating a scatter plot.

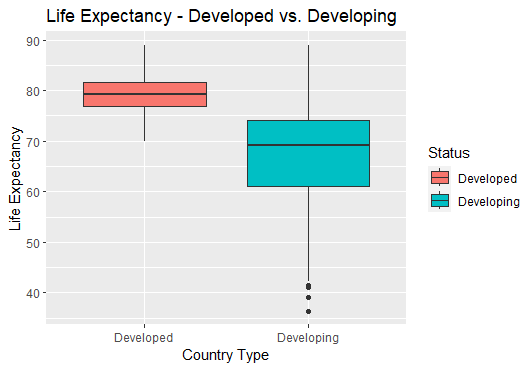
A screenshot of a computer

Description automatically generated

A screen shot of a computer

Description automatically generated

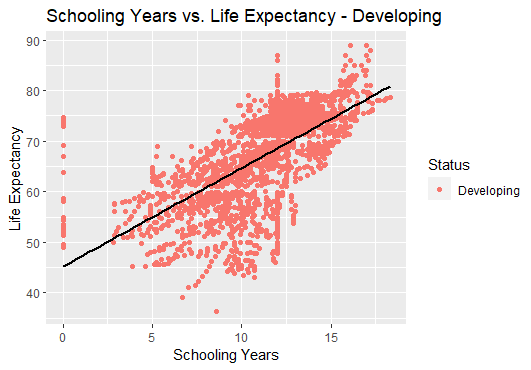
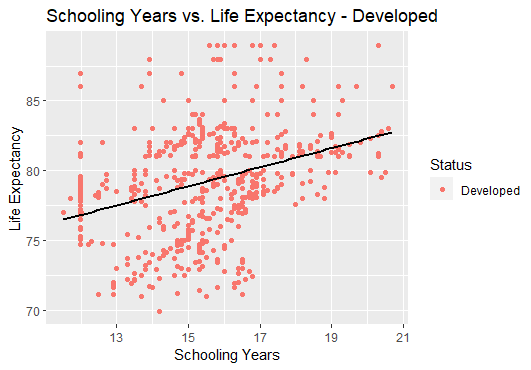
**Establish a Basis**

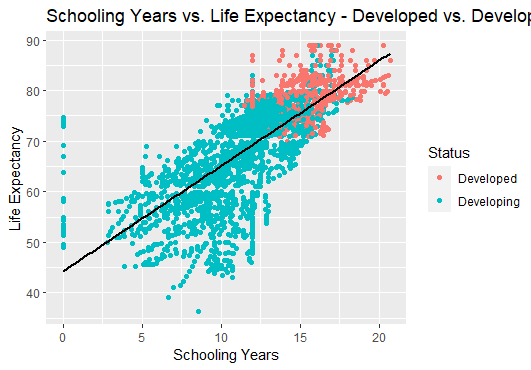


Above is a boxplot of the life experiences of developed and developing countries. The purpose of this is to establish a basis for expectations of life expectancies when looking at the differences between developed and developing countries. For developed countries, there is a higher overall life expectancy, and it has a more concise and smaller spread of life expectancies. In developing countries, there is a lower overall life expectancy, and the data has a larger spread of overall life expectancy.

**Life Expectancy vs. Schooling Years**

The number of schooling years on average for a country is a massive indicator of development in a country. It can indicate how focused a country is on education for its people and shows that a country has the money to spend on education. This factor can be an indicator of societal progress, meaning a more educated general population and socioeconomic growth of the country.





The scatter plots depicted above show three different situations of the same data. There are two that depict the number of schooling years on average for developing and developed countries. There is another one that includes both. These graphs show that schooling has a greater impact on developing countries when it comes to life expectancy. The correlation coefficient for developed countries for schooling and life expectancy is ~0.64, which indicates that there is a relationship between the two variables. The correlation coefficient for developed countries is ~0.35, which indicates a weak relationship between the two variables, at best. This leads to the conclusion that schooling has a greater impact on life expectancy in developing countries than in developed countries. While there is a relationship between the variables for each status, the relationship for developing countries is stronger. For only developing countries, the number of schooling years could theoretically be used as a predictor for life expectancy.

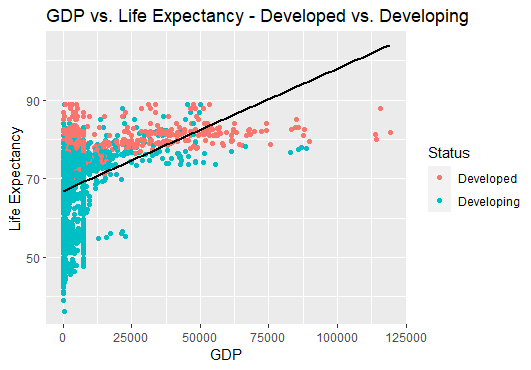
**Life Expectancy vs. GDP**

GDP is a factor that is commonly associated with developed countries. The more money a country makes, the more infrastructure can be created and more can be spent on people. This depends on whether the money the government has is spent on improving the lives of people. There will be three scatter plots shown below depicting Life Expectancy vs. GDP for developing countries, developed countries, and both developed and developing countries together.

A graph with red dots and a black line

Description automatically generatedA graph with red dots and black line

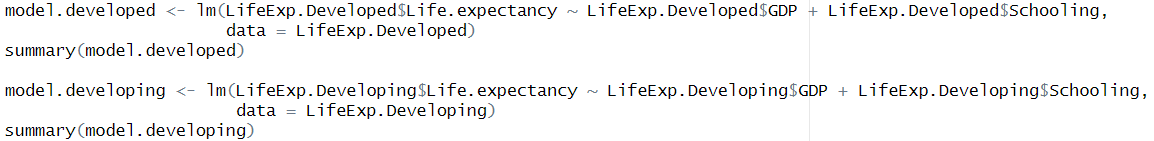
Description automatically generated



The results of these graphs are slightly different and are also underwhelming. Neither developed nor developing countries seemed to have a strong relationship. The correlation coefficient for GDP and life expectancy in developing countries was ~0.35, which is not a very strong correlation. The correlation coefficient for GDP and life expectancy in developed countries was ~0.36, which is not a very strong correlation once again. The correlation coefficients for both situations are relatively weak. This means that a conclusion that there is not a strong correlation for GDP and life expectancy, meaning that it may not be a strong predictor of life expectancy. Something important to remember for all these correlation coefficients is that correlation does not imply causation. All the factors contributing to life expectancy in these countries make looking at everything multidimensional.

**Regression Model**

To look more in depth at the data, the next step is to use a linear regression model. A linear regression model is a statistical method that models the relationship between one or more independent variables with a single dependent variable. This is accomplished by creating a linear equation for the data. The model’s goal is to make predictions about the relationship between variables. This can be accomplished rather easily using R. The code below is used to create a linear regression model for life expectancy in developed and developing countries:



This uses the lm() function to create a linear regression model. The results of this code for developed countries are shown below.

A white text with black numbers

Description automatically generated

The default for significance levels in this function is 0.05. From the p-value, R-squared value, and other significant statistical values in the results, it can be concluded that both GDP and schooling years are statistically significant in predicting life expectancy in developed countries. The results for developing countries are shown in the image below.

A screenshot of a computer

Description automatically generated

These results show that the GDP and number of schooling years are both significant in predicting life expectancy in developing countries. This is according to the p-value, R-squared value, and other significant statistical values in the results. This means that despite lower correlation coefficients, the factors of GDP and schooling years are still significant predictors of life expectancy in both developed countries and developing countries. There are also indications that there are other factors that contribute to life expectancy, which makes sense because there are many other factors included in the data set.

# **SUMMARY AND CONCLUSIONS OF HYPOTHESES**

**HYPOTHESIS 1**

*If life expectancy is indeed higher in developed countries compared to developing countries, what are factors that could contribute to this?*

**Answer:** Life expectancy is proven to be higher in developed countries. It is proven to show that the more developed countries a continent has than developing countries, the higher the median life expectancy is going to be. It is also proven that higher hepatitis B vaccinations have higher corresponding life expectancies and also income composition has a positive linear correlation with life expectancy.

**HYPOTHESIS 2**

*Could lower alcohol consumption rates add to the reasoning why life expectancy is higher in developed countries than developed countries?*

**Answer:** No, this cannot be determined. While life expectancy was determined to be higher in developed countries, alcohol consumption was higher in developing countries and not developed countries, not allowing us to perform this analysis.

**HYPOTHESIS 3**

*Could life expectancy in developing countries in the West Region of Africa be increasing over the years due to increasing healthcare expenditures?*

**Answer:** Although it is proven that life expectancy is increasing over the years in developing countries in the West Region of Africa, our analysis has not proven it is because of higher healthcare expenditures. It was found that measle counts have a negative correlation with life expectancy, but the analysis was not strong enough to say that this is the reason. It was then seen that life expectancy and total expenditure actually showed a negative correlation, which proves the exact opposite of what the hypothesis was stated.

**HYPOTHESIS 4**

*Is life expectancy in developing countries and developed countries change impacted by the factors of GDP and number of schooling years?*

**Answer:** Yes, to an extent. GDP and number of schooling years can be used as predictors for life expectancy in both developed and developing countries. This is evidenced slightly by the correlation coefficients found for each factor and more so by the linear regression model findings for both developed and developing countries. There are other factors that contribute to life expectancy, so it’s important to remember that correlation does not mean causation.

# KEY LEARNINGS AND LIMITATIONS

The things that the group learned from this project included a multitude of programming skills and analytical concepts. There were many different limitations learned during the working phase of the project. There was data quality, meaning our data might not be of perfect quality. Kaggle, where the data was found, does not only contain perfect quality data, which should be considered when looking at the results. Another limitation to consider is data bias. There may be bias in any data set gathered. The is exemplified by the fact that there is more data on developing countries than developed countries in the data set used. There are also temporal limitations in two different aspects. There are limitations on the data, which was gathered over a period of time. There could have been differences in the data for data 5 years before and after the period when the data was gathered. There was also a time limit for the completion of this project. This class lasts for about 8 weeks, so there was only a small amount of time to complete this project. Another limitation is the fact that correlation does not imply causation. This means that while there may be correlations between the variables, this does not absolutely imply causation between the variables. There are many factors to consider when looking at the life expectancies of countries, and finding a correlation between variables doesn’t necessarily mean that life expectancies are impacted directly by a single variable. There are also human limitations. The group members completing this project have had limited time this semester to learn R and have had limited time to learn analytics techniques to complete this project. The knowledge of this group is nowhere near complete in terms of R and analytics, so that is a limitation to consider when looking at this project.

Some things the group has learned and improved upon over the course of the project has been devising methods of analysis, data cleaning, communication of results, handling wrong or missing data, data filtering and partitioning, collaboration with code and documentation, data interpretation, data exploration techniques. Every group member devised their own method for analyzing each hypothesis and did their own analysis of the data. Each group member communicated their results with the rest of the group and to the class through presentations and group meetings. Each group member also created their own analyses and documented the purpose of each part of their code. Each member’s code also needed to be combined into one R file, which took collaboration in making sure variable names and other parts of each member’s code did not interfere with each other member’s code. Data cleaning was a vital part of data exploration and making data useful for creating plots and for general analysis. It was also involved in handling wrong or missing data, which in turn was useful for data interpretation. Data filtering was another useful technique for creating analyses for specific parts of code. These learning are all interconnected for exploring data, creating useful graphs, and data analysis.

# 

# REFERENECES

Kumarajarshi, Arshi. "Life Expectancy (WHO)." Kaggle, <https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who>.