

# Life Expectancy Exploratory Analysis

Simran Gaglani

Chapman University, CADS

CS 510 Reproducible Research

## Abstract

Life expectancy is a critical factor and indicator of the health around a given population. Thus, identifying and analyzing the variables which impact life expectancy can advance the overall health of a country into the future. Throughout the course of history, much research has been conducted around this subject. According to studies, the average life expectancy has increased over the course of history (Chetty *et al.* (2016)). In the past, many people and countries around the world endured a lack of access to resources, disease immunizations, or education, leading to a higher mortality rate, and a shorter lifespan. However as time progressed, countries continued to make improvements in their standard of living, vaccines, and overall development. This report aims to study how life expectancy has changed over 15 years as a result of development, and what variables have the most impact on it. The data-set has international information aggregated by the World Health Organization (WHO), and consists of time series data by country. Immunization values are also utilized, like Hepatitis B and Polio. The separation by country can facilitate the identification of predicting variables upon overall life expectancy. Specific countries can also be pinpointed which need to be prioritized in terms of analysis and health benefits. The top countries with the lowest life expectancy are developing countries with less economic resources and less education. The relationships between GDP, education, income, and life expectancy are explored.

**Keywords:** Life Expectancy; Status; Population Health; Developing

## I. Introduction

In order to record and observe global health status, the Global Health Observatory (GHO) under the World Health Organization (WHO) collects information pertaining to population health and other factors on behalf of all countries to advance research on life expectancy. While there are numerous variables that influence the average life span for people around the world, scientific breakthroughs and progress have made huge impacts upon human development and well-being(Chetty *et al.* (2016)). As science and technology improved, overall international inequality slowly decreased. As a result, there is a change in life expectancy. The improvements can be mapped and analyzed to uncover the underlying pattern behind life expectancy fluctuations. Analyzing a 15-year span from 2000 to 2015, we can uncover how world development variables can impact life expectancy.

## II. Background

Here are the different variables which are provided within the data-set. Each of these factors contain values which will be used to analyze different trends in life expectancy in developing and developed countries with differing economic circumstances. Figure 1 illustrates the information of the data-set.

- country (Nominal): Country
- year (Ordinal): Calendar year
- status (Nominal): 'Developing' or 'Developed'
- life.expectancy (Ratio): Life expectancy in years
- adult.mortality (Ratio): Adult mortality rate per 1000
- infant.deaths (Ratio): Number of infant deaths per 1000
- alcohol (Ratio): Alcohol consumption rate

Rows: 2,938	
Columns: 22	
\$ Country	<chr> "Afghanistan", "Afghanistan", "Afghani...
\$ Year	<int> 2015, 2014, 2013, 2012, 2011, 2010, 20...
\$ Status	<fct> Developing, Developing, Developing, De...
\$ Life.expectancy	<dbl> 65.0, 59.9, 59.9, 59.5, 59.2, 58.8, 58...
\$ Adult.Mortality	<int> 263, 271, 268, 272, 275, 279, 281, 287...
\$ infant.deaths	<int> 62, 64, 66, 69, 71, 74, 77, 80, 82, 84...
\$ Alcohol	<dbl> 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0...
\$ percentage.expenditure	<dbl> 71.279624, 73.523582, 73.219243, 78.18...
\$ Hepatitis.B	<int> 65, 62, 64, 67, 68, 66, 63, 64, 63, 64...
\$ Measles	<int> 1154, 492, 430, 2787, 3013, 1989, 2861...
\$ BMI	<dbl> 19.1, 18.6, 18.1, 17.6, 17.2, 16.7, 16...
\$ under-five.deaths	<int> 83, 86, 89, 93, 97, 102, 106, 110, 113...
\$ Polio	<int> 6, 58, 62, 67, 68, 66, 63, 64, 63, 58...
\$ Total.expenditure	<dbl> 8.16, 8.18, 8.13, 8.52, 7.87, 9.20, 9...
\$ Diphtheria	<int> 65, 62, 64, 67, 68, 66, 63, 64, 63, 58...
\$ HIV.AIDS	<dbl> 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1...
\$ GDP	<dbl> 584.25921, 612.69651, 631.74498, 669.9...
\$ Population	<dbl> 33736494, 327582, 31731688, 3696958, 2...
\$ thinness.1.19.years	<dbl> 17.2, 17.5, 17.7, 17.9, 18.2, 18.4, 18...
\$ thinness.5.9.years	<dbl> 17.3, 17.5, 17.7, 18.0, 18.2, 18.4, 18...
\$ Income.composition.of.resources	<dbl> 0.479, 0.476, 0.470, 0.463, 0.454, 0.4...
\$ Schooling	<dbl> 10.1, 10.0, 9.9, 9.8, 9.5, 9.2, 8.9, 8...

Figure 1 Data Summary

- percentage.expenditure (Ratio): Expenditure on health as a percentage of gdp
- hepatitis.b (Ratio): Number of 1 year olds with Hepatitis B immunization
- measles (Ratio): Number of reported Measles cases per 1000
- bmi (Interval/Ordinal): Average Body Mass Index (BMI) of population
- under-five.deaths: number of people under the age of 5 deaths per 1000
- polio (Ratio): Number of 1 year olds with Polio immunization
- total.expenditure (Ratio): Government expenditure on

health as a percentage of total government expenditure

- diphtheria (Ratio): immunization rate of 1 year olds for Diphtheria
- hiv/aids (Ratio): Deaths per 1000 live births by HIV/AIDS for people under 5
- gdp (Ratio): Gross Domestic Product per capita
- population (Ratio): Population
- thinness.1-19.years (Ratio): Rate of thinness among people aged 10-19
- thinness.5-9.years (Ratio): Rate of thinness among people aged 5-9
- income.composition.of.resources (Ratio): Income composition of resources
- schooling (Ratio): Average number of years of schooling

### III. Goals

Life expectancy as a variable can highlight the health of a population. The analysis of life expectancy over the course of 15 years (from 2000 - 2015) illustrates the overall behavior of population health during that specific frame of time. It can also be used to discover the status of which countries have low or high life expectancy. Farther analysis will showcase which variables impact life expectancy the most, and the relationship between gdp, income, schooling, and life expectancy. Finally, a linear regression model will facilitate the testing of the different variables within the data-set.

### IV. Change in Life Expectancy Over Time and Status

Data is aggregated to group the average life expectancy by calendar year. In the beginning, there was an initial drop in population from the year 2000 to about year 2004. However, in the continuing years there was a huge spike in growth for life expectancy that showed a continuing trend into the future. Figure 2 illustrates the increase in life expectancy over the course of these 15 years.

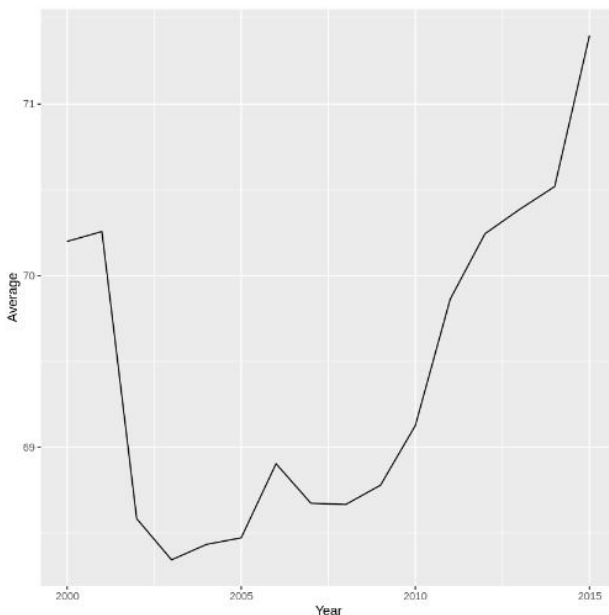


Figure 2 Change in Life Expectancy

### V. High vs Low Life Expectancy

Examples of how nations can improve their standard of living and health include increasing access to clean drinking water, advancing education for women and minorities, advancing health care, or the igniting civil/human rights movements. Life expectancy vary drastically from country to country, which lead scientists to wonder what factors cause these shifts in health patterns across the globe. According to some studies, the overwhelming inequities in population health can be attributes to wealth imbalances. Usually, richer countries have a higher average life expectancy than poorer countries. Furthermore, the "Preston Curve" discovered a cross-section logarithmic curve affiliation between income and life span, stating that life expectancy increases swiftly with income at first, before it plateaus. (Freeman and H.A. (2020)).

#### Low Life Expectancy and Status

On the other hand, analysis around countries with a lower on average life expectancy can portray different parts of the world which need support to improve the factors which may lead to an increase in population health. Figure 3 shows top 10 countries with the lowest average life expectancy.

#### High Life Expectancy and Status

Historically, developed countries with more resources experience a longer life expectancy(Leon (2011)). The data can be filtered to show countries with the highest life expectancy, and their status. We can use this data to pinpoint countries with a good population health. Figure 4 shows top 10 countries with the highest average life expectancy.

Country	Average.Life.Expectancy
<chr>	<dbl>
Mozambique	53.40714
Nigeria	52.84000
Chad	52.28571
Central African Republic	51.41667
Swaziland	50.82000
Angola	50.67500
Malawi	50.26154
Zimbabwe	49.38667
Lesotho	48.60833
Sierra Leone	48.42500

Figure 3 Countries with the Lowest Life Expectancy.

#### Compare and Contrast

Sierra Leone has the lowest average life expectancy at 48.4, while Ireland has the highest average life expectancy at 83.4.

Country	Average.Life.Expectancy
<chr>	<dbl>
Ireland	83.44000
Canada	82.23333
France	82.20667
Italy	82.15333
Spain	82.02000
Australia	81.90714
Sweden	81.90000
Austria	81.48000
Netherlands	81.32500
Greece	81.23333

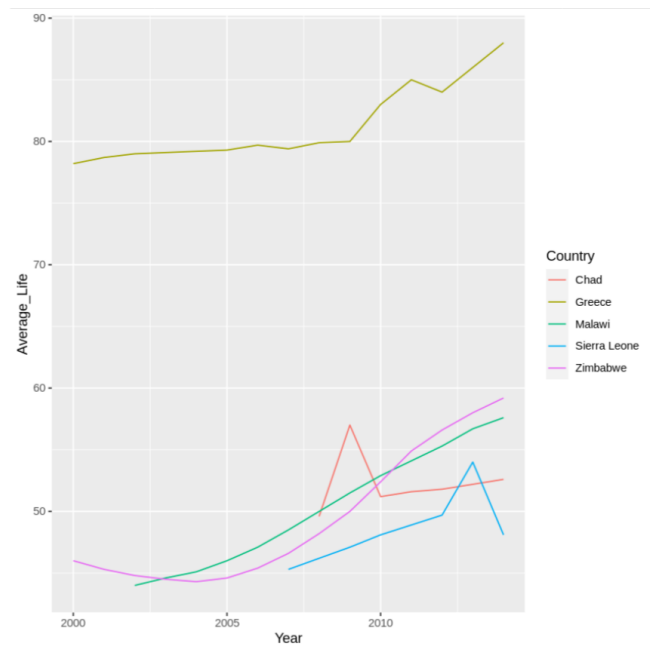
**Figure 4** Countries with the Highest Life Expectancy.

Country	Year	Status
<chr>	<int>	<fct>
Angola	2014	Developing
Central African Republic	2014	Developing
Chad	2014	Developing
Lesotho	2014	Developing
Malawi	2014	Developing
Mozambique	2014	Developing
Nigeria	2014	Developing
Sierra Leone	2014	Developing
Swaziland	2014	Developing
Zimbabwe	2014	Developing

**Figure 5** Countries with the Lowest Life Expectancy are also Developing Countries.

## VI. Status

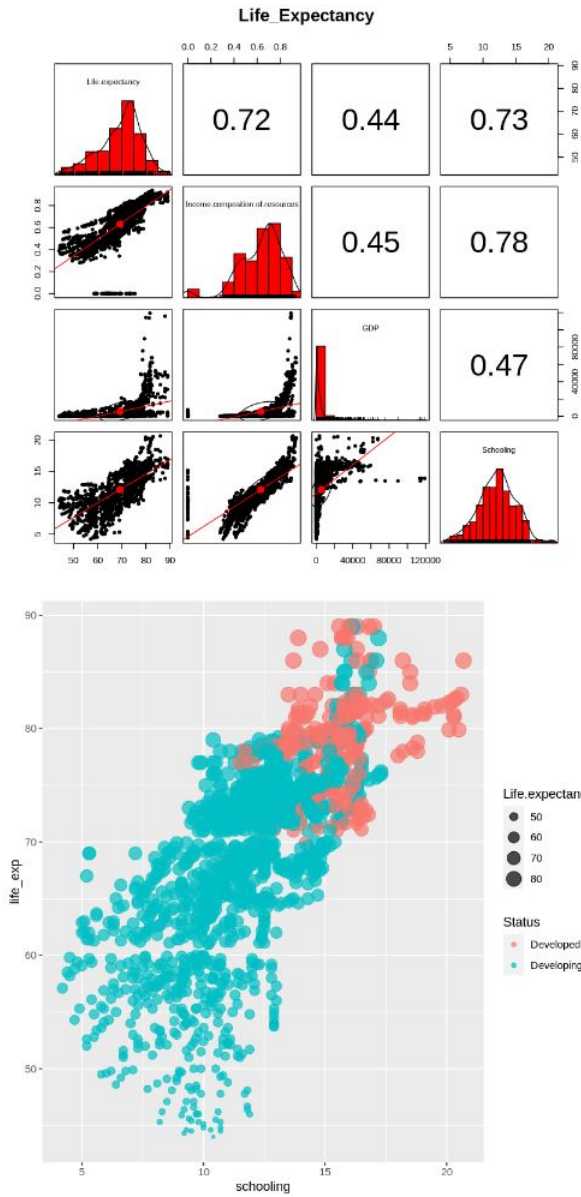
After finding the average life expectancy by country, it was found that all countries with the lowest life expectancy have the status of 'Developing'. Figure 5 illustrates this trend. This pattern articulates that these countries need support in increasing overall health of the population to improve life expectancy. To farther illustrate the disparity between the nations, figure 6 shows a graph which compares the average life expectancy of Greece, a developed country with a high life expectancy, vs some other countries like Zimbabwe and Sierra Leone with a lower average life expectancy. Greece is represented by the top yellow line on the graph while the other countries lie below; we can visualize the disproportion and imbalance between the nations. To reduce this inequity between developing and developed countries, specific variables need to be investigated and implemented which can beneficially impact life expectancy for countries in need.



**Figure 6** The Life Expectancy of Developing Countries - Chad, Zimbabwe, Malawi, and Sierra Leone - Against the Developed Country of Greece, Plotted on a Graph.

## VII. Variables Which Impact Life Expectancy

With an increase in health care, world development, and a decrease in mortality rates, there are a plethora of different factors which may lead to a longer or shorter life expectancy by country. This data-set includes immunization values which enhance the record's effectiveness in measuring the change in health patterns for a specific population. According to [Freeman and H.A. \(2020\)](#), some major factors which yield a healthy population include low health care costs, political support for health care causes, education, and community participation. Based on this information and our analysis, critical factors like education, GDP, and income are taken into consideration when analyzing population health development for this data-set. Figure 7 shows all the variables which impact life expectancy in a correlation matrix.



**Figure 7** Variables Impacting Life Expectancy

Because schooling and life expectancy had a positive and significant correlation, the second graph shows the relationship between the two.

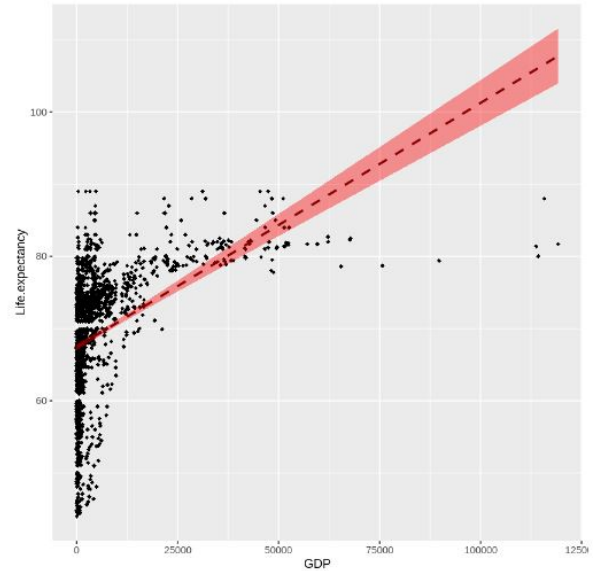
### Correlation Matrix

To compare and contrast critical variables within the data, correlation values for different factors need to be extracted. The correlation matrix displays the correlation between all the possible pairs of values. This tool is utilized to summarize the patterns within the data for population health. This analysis also allows for the identification of structured determinants which influence the overall population well-being.

### GDP and Life Expectancy

Since the countries from this data-set with the lowest life expectancy all have the status of developing, the relationship between GDP and Life expectancy should be explored. The correlation

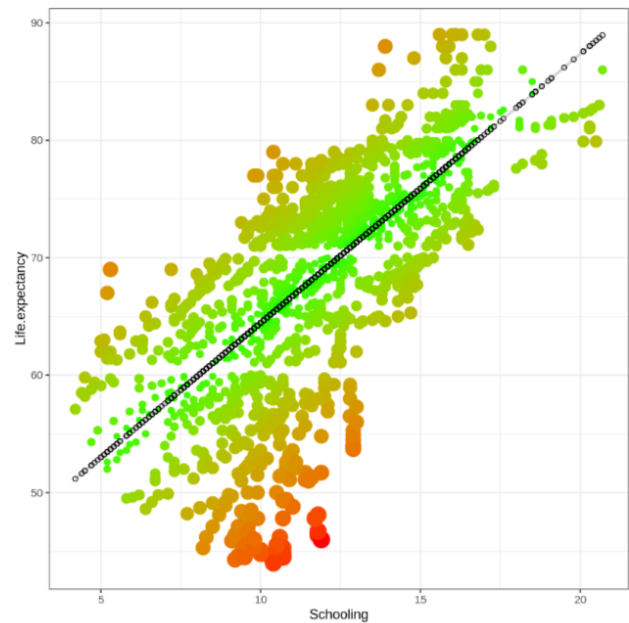
coefficient between the two factors is 0.44. Thus, the correlation between life expectancy and GDP is positive, but not significant. Figure 8 shows that the relationship between GDP and life expectancy.



**Figure 8** Correlation is Positive but Not Significant: .44

### Education and Life Expectancy

The correlation coefficient between schooling (education) and life expectancy is 0.727. With a strong and positive correlation, schooling has a direct relationship with lifespan by country. Figure 9 shows the positive and significant correlation. More education equals higher life expectancy.

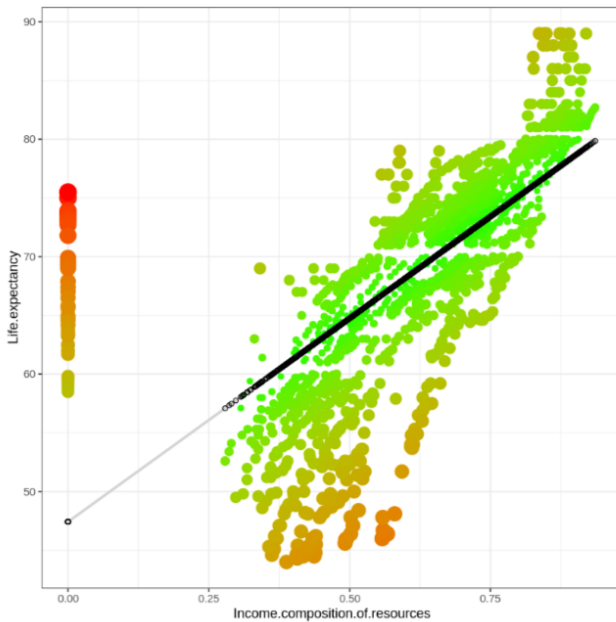


**Figure 9** Correlation between Schooling and Life Expectancy is Positive and Significant: .727. Countries with Higher Schooling Have Higher Life Expectancy.



## Income Composition and Life Expectancy

Within earlier analysis, it can be seen that Income Composition also has a strong relationship with life expectancy. The correlation coefficient between the two variables is 0.721 - positive and significant. Figure 10 portrays this relationship.



**Figure 10** Correlation between Income Composition and Life Expectancy is Positive and Significant: .721.

## VIII. Linear Regression

Figure 11 shows a linear regression model of schooling and income. There is R-Squared value of 0.588; approximately 58% of variation in life expectancy can be explained by our model. We also have the f-statistic and p-value to test the overall significance. Here, we reject the null hypothesis that there's no difference between the means and conclude that a significant difference does exist. The residual standard error value of 5.95 gives us an idea of how far observed the life expectancy is from the predicted or fitted y value.

```
Residuals:
    Min       1Q   Median       3Q      Max
-21.6304  -2.4647   0.7075   3.3022  17.8130

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    41.38866    0.61928   66.83  <2e-16
life_clean$Schooling  1.32507    0.08032   16.50  <2e-16
life_clean$Income.composition.of.resources  1.22628    1.22628   15.31  <2e-16

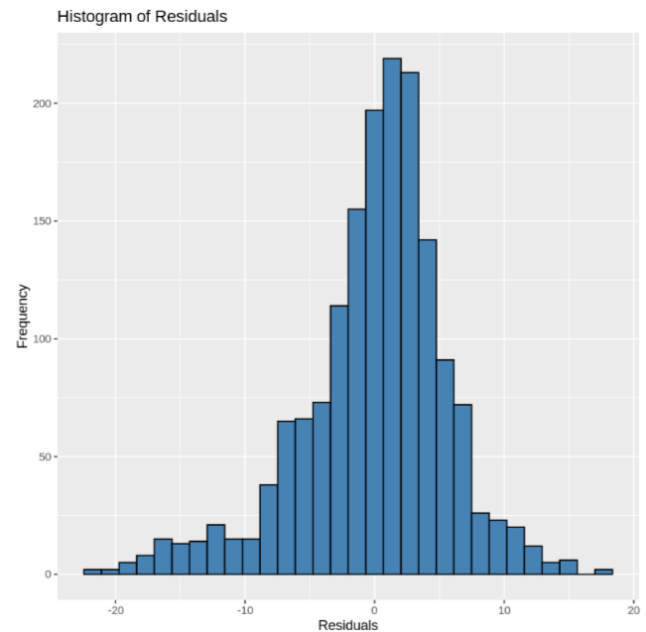
(Intercept)          ***
life_clean$Schooling  ***
life_clean$Income.composition.of.resources ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.649 on 1646 degrees of freedom
Multiple R-squared:  0.5881,    Adjusted R-squared:  0.5876
F-statistic: 1175 on 2 and 1646 DF,  p-value: < 2.2e-16
```

**Figure 11** Linear Regression Model of Schooling, Income Composition, and Life Expectancy

## Residuals

Figure 12 shows the residuals for the model plotted into a histogram, and is distributed around 0.



**Figure 12** Residuals Graphed

## IX. Conclusion

Conclusions drawn from the analysis can confirm a large uptick in life expectancy from 2005 to 2015. prior to this period, there was a drop in life expectancy from 2001 to 2003. Additionally, Ireland has the highest on average life expectancy while Sierra Leone has the lowest life expectancy on average. While looking into the correlation between variables, it was found that GDP is positively correlated with life expectancy, although not significant. However, Schooling and Income Composition do have a significant and positive correlation with life expectancy. They share a direct relationship. Finally, the linear regression model yielded the R-squared value of .588, and a p-value of under .05. Schooling and Income have the largest impact upon life expectancy, and can be manipulated in reality to help struggling countries enhance in overall population health. Farther analysis which can be done includes investigating the relationship between immunization variables to shed light on the impact of diseases upon life expectancy during the span of 15 years.

## Literature cited

- Chetty R, Stepner M, Abraham S, Lin S, Scuderi B, Turner N, Bergeron A, Cutler D. 2016. The Association Between Income and Life Expectancy in the United States, 2001-2014. *JAMA*. 315:1750–1766.
- Freeman, T. G, H.A. B. 2020. Why do some countries do better or worse in life expectancy relative to income? An analysis of Brazil, Ethiopia, and the United States of America. *International Journal for Equity in Health*. .
- Leon DA. 2011. Trends in European life expectancy: a salutary view. *International Journal of Epidemiology*. 40:271–277.