Project

# **Investing in Health, Investing in Life: A Socioeconomic Analysis of Strategies for Global Life Expectancy Improvement**

**BANL-6900-02**

**Group15**

Anirudh. Sunkara and Sumukh. Ramagiri

# **About the Dataset**

**Link:**[**https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who**](https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who)

**Context**

Although there have been lot of studies undertaken in the past on factors affecting life expectancy considering demographic variables, income composition and mortality rates. It was found that the effects of immunization and human development index were not taken into account in the past. Also, some of the past research was done considering multiple linear regression based on data set of one year for all the countries. Hence, this gives motivation to resolve both the factors stated previously by formulating a regression model based on mixed effects and multiple linear regression while considering data from a period of 2000 to 2015 for all the countries. Important immunization like Hepatitis B, Polio and Diphtheria will also be considered. In a nutshell, this study will focus on immunization factors, mortality factors, economic factors, social factors and other health-related factors as well. Since the observations this dataset are based on different countries, it will be easier for a country to determine the predicting factor which is contributing to lower life expectancy. This will help in suggesting a country which area should be given importance in order to efficiently improve the life expectancy of its population.

## **Content**

The project relies on accuracy of data. The Global Health Observatory (GHO) data repository under World Health Organization (WHO) keeps track of the health status as well as many other related factors for all countries The data sets are made available to public for the purpose of health data analysis. The data-set related to life execptancy publicy and health factors for 193 countries has been collected from the same WHO data repository website and its corresponding economic data was collected from United Nation website. Among all categories of health-related factors only those critical factors were chosen which are more representative. It has been observed that in the past 15 years , there has been a huge development in health sector resulting in improvement of human mortality rates especially in the developing nations in comparison to the past 30 years. Therefore, in this project we have considered data from years 2000-2015 for 193 countries for further analysis. The individual data files have been merged together into a single data-set. On initial visual inspection of the data showed some missing values. As the data-sets were from WHO, we found no evident errors. Missing data was handled in R software by using Missmap command. The result indicated that most of the missing data was for population, Hepatitis B and GDP. The missing data were from less known countries like Vanuatu, Tonga, Togo, Cabo Verde etc. Finding all data for these countries was difficult and hence, it was decided that we exclude these countries from the final model data-set. The final merged file(final dataset) consists of 22 Columns and 2938 rows which meant 20 predicting variables. All predicting variables was then divided into several broad categories:​Immunization related factors, Mortality factors, Economical factors and Social factors.

## **Acknowledgements**

The data was collected from WHO and the United Nations website with the help of Deeksha Russell and Duan Wang.

## **Introduction**

The life expectancy dataset consists of variables that collectively help us to understand about health and survival across different countries and regions. This dataset not only includes economic indicators, which often serve as broad measures of a country’s overall wealth and capacity for providing services to its population, but also encompasses specific health-related factors that can have direct impacts on lifespan. Among the economic indicators, Gross Domestic Product (GDP) and percentage of health expenditure offer insights into the financial priorities and capabilities of nations. Additionally, the dataset includes detailed health metrics such as vaccination rates for diseases like Hepatitis B, Polio, and Diphtheria, as well as other critical factors like alcohol consumption and infant mortality rates. These variables are pivotal in understanding the direct interventions and lifestyle factors that contribute to longer or shorter life expectancies.

## **Objectives**

The objective of this analysis is to dissect and interpret how various economic, healthcare, and lifestyle factors influence life expectancy across different countries. By applying both descriptive and inferential statistical methods, the study aims to:

**Quantifying Relationships:** Applying linear regression models to quantitatively assess how life expectancy responds to changes in economic conditions (GDP and health expenditure), public health policies (vaccination rates), and lifestyle choices (alcohol consumption).

**Identifying Key Predictors:** Determining which factors are most strongly associated with longer or shorter life expectancies. This involves evaluating the statistical significance and strength of relationships between life expectancy and predictors such as GDP, vaccination rates, and alcohol consumption.

**Applying Data Exploration:** Using histograms to visualize the distribution of life expectancy and scatter plots to explore relationships between life expectancy and other continuous variables like GDP and vaccination coverage. These visual tools are crucial for identifying trends, outliers, and potential insights in the data.

**Health and Economic Insights:** Integrating findings from statistical models and visualizations to draw insights about the health and economic status of populations. These insights can guide public health decisions and economic policies aimed at improving life expectancy.

**Policy Recommendations:** Based on the analysis, to provide actionable recommendations for policymakers and health organizations to target areas where health interventions or economic improvements could yield significant benefits in public health outcomes.

library(readr)  
setwd("/Users/sumukhramagiri/Desktop/Capstone")  
Life\_expectancy <- read.csv("Life\_Expectancy\_Data.csv")  
missing\_values <- is.na(Life\_expectancy)  
sum(missing\_values)

[1] 0

summary(Life\_expectancy)

Country Year Status Life.expectancy  
 Length:1649 Min. :2000 Length:1649 Min. :44.0   
 Class :character 1st Qu.:2005 Class :character 1st Qu.:64.4   
 Mode :character Median :2008 Mode :character Median :71.7   
 Mean :2008 Mean :69.3   
 3rd Qu.:2011 3rd Qu.:75.0   
 Max. :2015 Max. :89.0   
 Adult.Mortality infant.deaths Alcohol percentage.expenditure  
 Min. : 1.0 Min. : 0.00 Min. : 0.010 Min. : 0.00   
 1st Qu.: 77.0 1st Qu.: 1.00 1st Qu.: 0.810 1st Qu.: 37.44   
 Median :148.0 Median : 3.00 Median : 3.790 Median : 145.10   
 Mean :168.2 Mean : 32.55 Mean : 4.533 Mean : 698.97   
 3rd Qu.:227.0 3rd Qu.: 22.00 3rd Qu.: 7.340 3rd Qu.: 509.39   
 Max. :723.0 Max. :1600.00 Max. :17.870 Max. :18961.35   
 Hepatitis.B Measles BMI under.five.deaths  
 Min. : 2.00 Min. : 0 Min. : 2.00 Min. : 0.00   
 1st Qu.:74.00 1st Qu.: 0 1st Qu.:19.50 1st Qu.: 1.00   
 Median :89.00 Median : 15 Median :43.70 Median : 4.00   
 Mean :79.22 Mean : 2224 Mean :38.13 Mean : 44.22   
 3rd Qu.:96.00 3rd Qu.: 373 3rd Qu.:55.80 3rd Qu.: 29.00   
 Max. :99.00 Max. :131441 Max. :77.10 Max. :2100.00   
 Polio Total.expenditure Diphtheria HIV.AIDS   
 Min. : 3.00 Min. : 0.740 Min. : 2.00 Min. : 0.100   
 1st Qu.:81.00 1st Qu.: 4.410 1st Qu.:82.00 1st Qu.: 0.100   
 Median :93.00 Median : 5.840 Median :92.00 Median : 0.100   
 Mean :83.56 Mean : 5.956 Mean :84.16 Mean : 1.984   
 3rd Qu.:97.00 3rd Qu.: 7.470 3rd Qu.:97.00 3rd Qu.: 0.700   
 Max. :99.00 Max. :14.390 Max. :99.00 Max. :50.600   
 GDP Population thinness..1.19.years  
 Min. : 1.68 Min. :3.400e+01 Min. : 0.100   
 1st Qu.: 462.15 1st Qu.:1.919e+05 1st Qu.: 1.600   
 Median : 1592.57 Median :1.420e+06 Median : 3.000   
 Mean : 5566.03 Mean :1.465e+07 Mean : 4.851   
 3rd Qu.: 4718.51 3rd Qu.:7.659e+06 3rd Qu.: 7.100   
 Max. :119172.74 Max. :1.294e+09 Max. :27.200   
 thinness.5.9.years Income.composition.of.resources Schooling   
 Min. : 0.100 Min. :0.0000 Min. : 4.20   
 1st Qu.: 1.700 1st Qu.:0.5090 1st Qu.:10.30   
 Median : 3.200 Median :0.6730 Median :12.30   
 Mean : 4.908 Mean :0.6316 Mean :12.12   
 3rd Qu.: 7.100 3rd Qu.:0.7510 3rd Qu.:14.00   
 Max. :28.200 Max. :0.9360 Max. :20.70

### **1. Setting Up the Environment:**

To begin, I set the working directory in R to the folder containing the dataset. This is crucial as it locates all file operations within this designated workspace, ensuring easy access to the dataset and related files.

### **2. Loading the Dataset:**

Using the readr package, a well-known package in the R community for its speed and efficiency, I load the dataset into the R environment. The read.csv function from this package is ideal for reading comma-separated values (CSV) files into R.

### **3. Checking for Missing Values:**

The immediate step after loading the data is to ensure its quality, which primarily involves checking for missing values. Missing data can significantly impact the analysis, leading to biased or incorrect conclusions. Here, I use the is.na function combined with sum to quickly tally up any missing entries in the dataset.

This process revealed that there are no missing values in the dataset ([1] 0), indicating that the dataset is complete and no immediate cleaning or deletion of data is required.

### **4. Understanding the Data Structure:**

Understanding the distribution, range, and general characteristics of the data is fundamental. The summary () function in R provides a quick and comprehensive overview of each variable within the dataset. It gives insights into mean, median, quartiles, and range, which are very useful for planning further analysis.

From the summary, I can observe various aspects such as:

Life expectancy ranges from a minimum of 44.0 to a maximum of 89.0 years.

Variables like GDP and population exhibit wide ranges, suggesting varying economic conditions and country sizes.

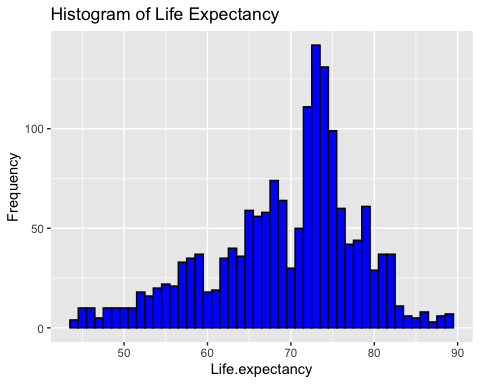
Health-related indicators such as adult mortality, infant deaths, and vaccination rates (e.g., Hepatitis B, Polio, Diphtheria) also vary significantly, which could be crucial in understanding public health trends.

When doing data analysis, it’s important for an analyst to start by asking specific questions. These questions help guide the analysis, making it easier to spot important trends and make predictions. This focused approach helps the analyst to concentrate on what matters most for the business, leading to useful insights that can help in making decisions. By asking the right questions from the start, the analyst can choose the best methods and tools to get clear and relevant results, which are crucial for understanding past events and predicting what might happen next.

**What is the distribution of life expectancy across the countries in our dataset, and are there any immediate patterns or anomalies in the data that could indicate areas of interest for further investigation?**

This question aims to establish a foundational understanding of the life expectancy variable, which is crucial for further detailed analysis. It helps identify the range, central tendency, and spread of life expectancy values, which are essential for formulating hypotheses about health outcomes and associated factors.

library(ggplot2)  
ggplot(Life\_expectancy, aes(x = `Life.expectancy`)) +  
 geom\_histogram(binwidth = 1, fill = "blue", color = "black") +  
 labs(title = "Histogram of Life Expectancy", x = "Life.expectancy", y = "Frequency")



The life expectancy histogram clearly shows that most countries have life expectancy ranging from 64.4 to 75.0 years, indicating a widespread level of health that many nations attain. This consistency implies that a substantial number of nations have comparable healthcare access, economic circumstances, and public health policies that contribute to the lifespan of their inhabitants. This graphic highlights the effectiveness of existing health standards and serves as a reference point for global health.  
  
Nevertheless, the prevalence of a bias towards the lower range of life expectancy demonstrates that not all nations adhere to this worldwide benchmark, with many countries severely falling behind. These departures from the norm often arise due to a range of obstacles, such as inadequate healthcare facilities, increased illness prevalence, economic hardships, or socio-political instability. Determining these anomalies is essential as it guides the allocation of specific health initiatives and foreign assistance to concentrate efforts on elevating these countries towards the larger global health benchmarks, with the goal of reducing gaps and enhancing global health equality.

**How does Gross Domestic Product (GDP) correlate with life expectancy across different countries, and can we observe any patterns that might suggest a relationship between a country’s economic status and its public health outcomes?**

This question help us learn about the relationship between economic wealth and health, providing insights into how financial resources and programs might influence public health at the national level. Understanding this relationship can guide policies on healthcare investment and economic development.

library(dplyr)

Warning: package 'dplyr' was built under R version 4.2.3

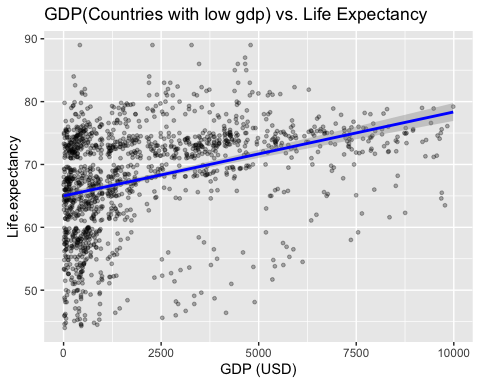
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':  
  
 filter, lag

The following objects are masked from 'package:base':  
  
 intersect, setdiff, setequal, union

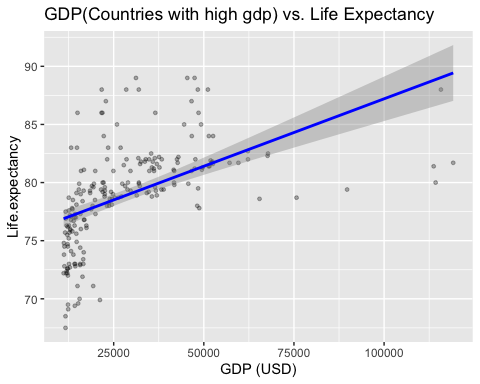
library(ggplot2)  
low\_gdp <- Life\_expectancy %>%  
 filter(GDP < 10000)  
ggplot(low\_gdp, aes(x = GDP, y = `Life.expectancy`)) +  
 geom\_point(alpha = 0.3, size = 1) +  
 geom\_smooth(method = "lm", color = "blue") +  
 labs(title = "GDP(Countries with low gdp) vs. Life Expectancy", x = "GDP (USD)", y = "Life.expectancy")

`geom\_smooth()` using formula = 'y ~ x'



high\_gdp <- Life\_expectancy %>%  
 filter(GDP > 10000)  
  
ggplot(high\_gdp, aes(x = GDP, y = `Life.expectancy`)) +  
 geom\_point(alpha = 0.3, size = 1) +  
 geom\_smooth(method = "lm", color = "blue") +  
 labs(title = "GDP(Countries with high gdp) vs. Life Expectancy", x = "GDP (USD)", y = "Life.expectancy")

`geom\_smooth()` using formula = 'y ~ x'



Upon reviewing the relationship between the GDP and life expectancy, it’s clear there’s a link between a country’s wealth (i.e measured in GDP) and the health of its people, indicated by life expectancy. In countries with lower GDPs, a small increase in wealth tends to correspond with a noticeable rise in how long people live. This suggests that in poorer countries, even a slight economic boost can lead to better health outcomes.

As countries get richer, however, the situation changes. In wealthier nations, the life expectancy still goes up as the GDP increases, but the change isn’t as large. This points to the idea that once a country reaches a certain level of wealth, just adding more money doesn’t make as big of a difference to people’s health. Instead, these countries may need to look at other ways to keep improving the health of their citizens.

It means that for poorer countries, focusing on growing the economy could be a powerful way to improve public health. In contrast, richer countries might need policies that target specific health issues more directly, rather than relying on economic growth alone to drive health improvements. Understanding this nuanced relationship can help tailor your strategies to be effective in different economic contexts, ensuring that investments are made where they can have the greatest impact on people’s health.

**What is the impact of vaccination rates on life expectancy among different countries, and can we quantify this impact using a linear model to understand the potential benefits of increasing vaccination coverage?**

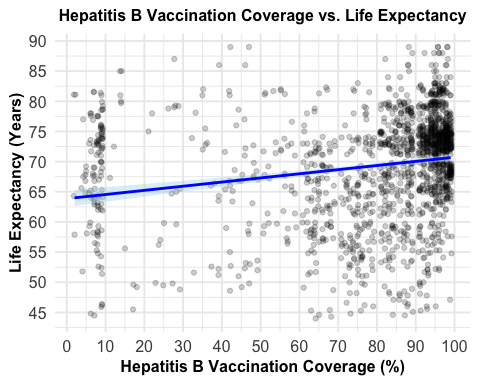
This question is designed to explore how public health initiatives, specifically vaccination programs, correlate with improvements in life expectancy. The analysis will help in assessing the effectiveness of vaccinations as a public health strategy.

model\_Hepatitis.B <- lm(`Life.expectancy` ~ `Hepatitis.B`, data = Life\_expectancy)  
summary(model\_Hepatitis.B)

Call:  
lm(formula = Life.expectancy ~ Hepatitis.B, data = Life\_expectancy)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-25.337 -4.924 1.788 5.225 22.254   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 63.860800 0.690540 92.480 < 2e-16 \*\*\*  
Hepatitis.B 0.068691 0.008295 8.281 2.49e-16 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 8.622 on 1647 degrees of freedom  
Multiple R-squared: 0.03997, Adjusted R-squared: 0.03939   
F-statistic: 68.58 on 1 and 1647 DF, p-value: 2.492e-16

library(dplyr)  
library(ggplot2)  
  
  
ggplot(Life\_expectancy, aes(x = `Hepatitis.B`, y = `Life.expectancy`)) +  
 geom\_jitter(alpha = 0.2, size = 1.5, width = 0.3, height = 0) +   
 geom\_smooth(method = "lm", color = "blue", fill = "lightblue") +   
 labs(title = "Hepatitis B Vaccination Coverage vs. Life Expectancy",  
 x = "Hepatitis B Vaccination Coverage (%)",  
 y = "Life Expectancy (Years)") +  
 theme\_minimal(base\_size = 14) +   
 theme(  
 plot.title = element\_text(size = 12, face = "bold", hjust = 0.5),  
 axis.title = element\_text(size = 12, face = "bold"),  
 axis.text = element\_text(size = 12)  
 ) +   
 scale\_x\_continuous(breaks = seq(0, 100, by = 10)) +   
 scale\_y\_continuous(breaks = seq(40, 90, by = 5))

`geom\_smooth()` using formula = 'y ~ x'



From the scatter plot of Hepatitis B vaccine coverage versus life expectancy, there is a noticeable trend that can be derived. As vaccination coverage rises, there is a general tendency towards longer life expectancy. This tendency is obvious despite the vast dispersion of data points at the bottom end of the vaccination coverage range.  
  
From a glance at the linear trend line overlaying on the plot, it’s evident there is a positive correlation between vaccination rates and life expectancy. Countries with larger percentages of their people immunized against Hepatitis B likely to have a longer average lifetime. The clustering of data points around higher vaccination rates shows that when coverage approaches 100%, life expectancy of that country is most likely to be in the higher range.  
  
In terms of public health implications, the plot supports the concept that immunizations are a significant tool for improving health outcomes. Increasing vaccination coverage seems to be connected with lifespan gains. Quantifying this influence using a linear model would enable us to anticipate how much life expectancy may rise with changes in vaccination rates.

In short, plot reveals a great investment opportunity in public health: boosting immunization programs might lead to determined gains in life expectancy. This coincides with global health policies highlighting the value of immunizations as a cost-effective strategy to increase population health and lifespan. This emphasizes the potential advantages of expanding efforts to enhance vaccination coverage as a method to improve the general health and well-being of people.

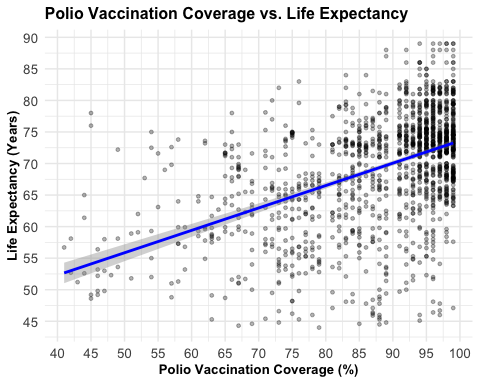
### **Linear Model Analysis of Polio Vaccination Coverage vs. Life Expectancy**

#Linear model for Polio vaccination  
model\_polio <- lm(`Life.expectancy` ~ Polio, data = Life\_expectancy)  
summary(model\_polio)

Call:  
lm(formula = Life.expectancy ~ Polio, data = Life\_expectancy)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-25.541 -4.684 1.585 5.088 25.260   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 58.585628 0.789397 74.22 <2e-16 \*\*\*  
Polio 0.128244 0.009123 14.06 <2e-16 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 8.315 on 1647 degrees of freedom  
Multiple R-squared: 0.1071, Adjusted R-squared: 0.1066   
F-statistic: 197.6 on 1 and 1647 DF, p-value: < 2.2e-16

library(dplyr)  
library(ggplot2)  
filtered\_data <- Life\_expectancy %>%  
 filter(Polio > 40, Polio <= 100, `Life.expectancy` > 40)  
#Scatter plot with regression line for Polio and Life Expectancy  
ggplot(filtered\_data, aes(x = Polio, y = `Life.expectancy`)) +  
 geom\_point(alpha = 0.3, size = 1) +  
 geom\_smooth(method = "lm", color = "blue") +  
 labs(title = "Polio Vaccination Coverage vs. Life Expectancy",  
 x = "Polio Vaccination Coverage (%)",  
 y = "Life Expectancy (Years)") +  
 theme\_minimal() +  
 theme(plot.title = element\_text(size = 12, face = "bold"),   
 axis.title = element\_text(size = 10, face = "bold"),   
 axis.text = element\_text(size = 10)) +  
 scale\_x\_continuous(breaks = seq(0, 100, by = 5)) +   
 scale\_y\_continuous(breaks = seq(40, 90, by = 5))

`geom\_smooth()` using formula = 'y ~ x'



The plot describes the relationship between Polio vaccine coverage and life expectancy across several nations. A positive trend is clearly visible from the plot: which indicates as the percentage of the population covered by Polio vaccine grows, there’s an equal gain in life expectancy. This pattern is shown by the blue trend line, which slopes higher as vaccination rates increase.  
  
The clustering of data at higher vaccine coverage rates near the 100% mark is especially dense, showing that nations with near universal Polio vaccination coverage likely to have greater life expectancy. Conversely, there is higher variability in life expectancy with lower vaccination rates, which may be related to a combination of reasons including the existence of Polio itself or additional indicators of healthcare availability and quality.  
  
From these facts, we may draw a strong correlation between vaccination programs and increasing life expectancy. The findings collected from this study might imply that a focus on improving Polio vaccination coverage could be a critical public health strategy with the potential to benefit the general health and lifespan of a community.  
  
It would be underlined that expenditures in vaccination programs are not merely a measure of avoiding sickness but also a substantial contribution to prolonging the average lifetime of a nation’s population. It’s a testimony to the multifarious advantages of such health initiatives: they’re not just immediate and disease-specific but also contribute to the long-term health and vitality of a country’s inhabitants. This data may influence choices on health financing, demonstrating that increasing immunization efforts is likely to yield rewards in the health and well-being of the population.

### **Linear Model Analysis of Diphtheria Vaccination Coverage vs. Life Expectancy**

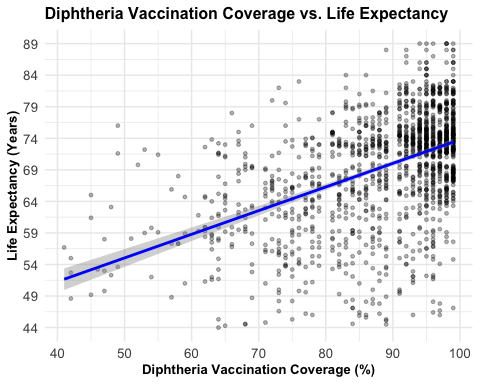
# Linear model for Diphtheria vaccination  
model\_diphtheria <- lm(`Life.expectancy` ~ Diphtheria, data = Life\_expectancy)  
summary(model\_diphtheria)

Call:  
lm(formula = Life.expectancy ~ Diphtheria, data = Life\_expectancy)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-25.476 -4.702 1.510 5.010 21.055   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 57.592533 0.820252 70.21 <2e-16 \*\*\*  
Diphtheria 0.139145 0.009442 14.74 <2e-16 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 8.271 on 1647 degrees of freedom  
Multiple R-squared: 0.1165, Adjusted R-squared: 0.116   
F-statistic: 217.2 on 1 and 1647 DF, p-value: < 2.2e-16

# Scatter plot with regression line for Diphtheria and Life Expectancy  
filtered\_data <- Life\_expectancy %>%  
 filter(Diphtheria > 40, Diphtheria <= 100, `Life.expectancy` > 40, `Life.expectancy` <= 100)  
  
# Creating the plot with the filtered data  
ggplot(filtered\_data, aes(x = Diphtheria, y = `Life.expectancy`)) +  
 geom\_point(alpha = 0.3, size = 1, width = 0.2) +  
 geom\_smooth(method = "lm", color = "blue", se = TRUE) +  
 labs(title = "Diphtheria Vaccination Coverage vs. Life Expectancy",  
 x = "Diphtheria Vaccination Coverage (%)",  
 y = "Life Expectancy (Years)") +  
 theme\_minimal() +  
 theme(plot.title = element\_text(size = 12, face = "bold"),  
 axis.title = element\_text(size = 10, face = "bold"),  
 axis.text = element\_text(size = 10)) +  
 scale\_x\_continuous(breaks = seq(0, 100, by = 10)) +   
 scale\_y\_continuous(breaks = seq(min(filtered\_data$`Life.expectancy`),   
max(filtered\_data$`Life.expectancy`), by = 5))

Warning in geom\_point(alpha = 0.3, size = 1, width = 0.2): Ignoring unknown  
parameters: `width`

`geom\_smooth()` using formula = 'y ~ x'



The scatter plot relating Diphtheria vaccine coverage versus life expectancy reveals a positive association, illustrated by an upward-sloping trend line. Higher rates of Diphtheria immunization results with improved life expectancy, and this trend is upward and constant throughout the range of vaccine coverage.  
  
Observing the dense clustering of data points around the upper end of vaccination coverage, and as we approach the 100% level, we can observe that life expectancy is higher. It’s also remarkable that there’s less dispersion in life expectancy as vaccination coverage improves, showing that high vaccination coverage is a significant predictor of higher life expectancy results.  
  
These facts, reveal a strong correaltion between vaccination coverage for Diphtheria and the total life expectancy of a country’s population. These findings underscore the practical advantages of immunization programs. Increasing Diphtheria vaccination rates appears to be strongly connected with longer lives, demonstrating that public health expenditures in immunization might give large benefits in terms of population health. Such data might be essential in pushing for the growth and sustainability of vaccination programs as a strategic component of enhancing public health and longevity within a community. This emphasizes the wider benefit of preventive healthcare initiatives and the need of their sustained support and execution at the national level.

# Creating the multiple regression model  
economic\_model <- lm(Life.expectancy ~ GDP + percentage.expenditure + Income.composition.of.resources,   
 data = Life\_expectancy)  
summary(economic\_model)

Call:  
lm(formula = Life.expectancy ~ GDP + percentage.expenditure +   
 Income.composition.of.resources, data = Life\_expectancy)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-20.6638 -2.6176 0.6954 3.1106 26.6891   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 4.879e+01 5.578e-01 87.466 <2e-16 \*\*\*  
GDP 7.984e-05 4.682e-05 1.705 0.0884 .   
percentage.expenditure 2.293e-04 2.984e-04 0.768 0.4424   
Income.composition.of.resources 3.152e+01 9.052e-01 34.825 <2e-16 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 5.986 on 1645 degrees of freedom  
Multiple R-squared: 0.5379, Adjusted R-squared: 0.537   
F-statistic: 638.2 on 3 and 1645 DF, p-value: < 2.2e-16

### **Outcomes from the Multiple Regression Model**

The multiple regression analysis reveals an intriguing picture of how economic factors influence life expectancy across countries. Most notably, it indicates that while GDP has some impact, it’s the income composition of resources that truly drives changes in life expectancy. The statistically significant coefficient for income composition suggests that factors such as the quality of education, access to healthcare, and income distribution have profound effects on how long people live. This reinforces the idea that human development and the equitable distribution of economic gains are fundamental to achieving longer, healthier lives for the population.

On the other hand, the percentage of GDP expenditure, potentially related to health spending, doesn’t show a significant impact on life expectancy as per this model. This could be interpreted as a call for not just increased spending but also for smarter, more effective use of these funds. It’s not just about the amount of money spent on health but how it’s spent that may lead to better health outcomes.

These findings suggest a multi-faceted approach for policymakers looking to improve the health and longevity of their populations. Instead of focusing solely on economic growth, there should be a concerted effort to ensure that the benefits of increased wealth are shared across society. This includes investing in quality education, improving healthcare access, and ensuring that the fruits of economic advancement reach all segments of the population. Such comprehensive strategies are likely to be more successful in raising life expectancy than a narrow focus on GDP growth alone.

In developed countries, higher life expectancy may often correlate with better healthcare and education systems, while in developing countries, improvements might hinge on basic healthcare access and education. Thus, enhancing life expectancy globally involves a multifaceted approach that combines economic, educational, and nutritional strategies to create a holistic improvement in public health.

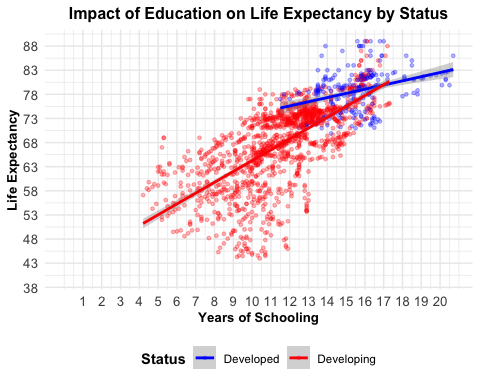
Let’s see if our assumptions are actually true:

# Making sure that Status is factorized   
Life\_expectancy$Status <- as.factor(Life\_expectancy$Status)  
  
# Build the multiple regression model  
model <- lm(Life.expectancy ~ Schooling + Status + BMI, data = Life\_expectancy)  
summary(model)

Call:  
lm(formula = Life.expectancy ~ Schooling + Status + BMI, data = Life\_expectancy)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-21.9555 -3.1650 0.9208 3.9798 14.0177   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 46.086512 1.004697 45.871 < 2e-16 \*\*\*  
Schooling 1.797312 0.068527 26.228 < 2e-16 \*\*\*  
StatusDeveloping -2.261188 0.471722 -4.793 1.79e-06 \*\*\*  
BMI 0.088173 0.008724 10.107 < 2e-16 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 5.819 on 1645 degrees of freedom  
Multiple R-squared: 0.5632, Adjusted R-squared: 0.5624   
F-statistic: 707 on 3 and 1645 DF, p-value: < 2.2e-16

library(ggplot2)  
# Scatter plot for Schooling vs Life Expectancy with Status distinction  
ggplot(Life\_expectancy, aes(x = Schooling, y = Life.expectancy, color = Status)) +  
 geom\_point(alpha = 0.3, size = 1) +  
 geom\_smooth(aes(group = Status), method = "lm", se = TRUE) +  
 labs(title = "Impact of Education on Life Expectancy by Status",  
 x = "Years of Schooling",  
 y = "Life Expectancy") +  
 theme\_minimal() +  
 theme(  
 plot.title = element\_text(size = 12, face = "bold", hjust = 0.5),  
 legend.title = element\_text(face = "bold"),  
 legend.position = "bottom",  
 axis.title = element\_text(size = 10, face = "bold"),  
 axis.text = element\_text(size = 10)  
 ) +  
 scale\_color\_manual(values = c("Developed" = "blue", "Developing" = "red")) +  
 scale\_x\_continuous( breaks = seq(min(Life\_expectancy$Schooling, na.rm = TRUE, by = 1),   
max(Life\_expectancy$Schooling, na.rm = TRUE), by = 1),  
limits = c(0, NA)) +  
scale\_y\_continuous( breaks = seq(min(Life\_expectancy$Life.expectancy, na.rm = TRUE, by = 3),   
max(Life\_expectancy$Life.expectancy, na.rm = TRUE), by = 5),  
limits = c(40, NA))

`geom\_smooth()` using formula = 'y ~ x'

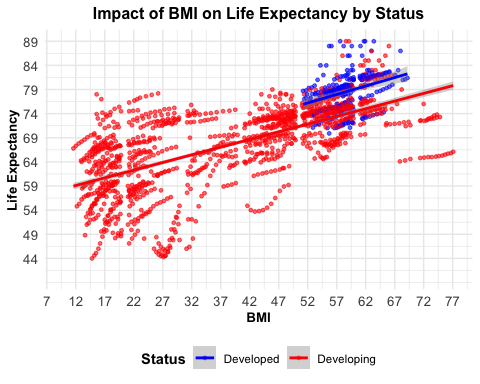


# Scatter plot for BMI vs Life Expectancy with Status distinction  
ggplot(Life\_expectancy, aes(x = BMI, y = Life.expectancy, color = Status)) +  
 geom\_point(alpha = 0.6, size = 1) +  
 geom\_smooth(aes(group = Status), method = "lm", se = TRUE) +  
 labs(title = "Impact of BMI on Life Expectancy by Status",  
 x = "BMI",  
 y = "Life Expectancy") +  
 theme\_minimal() +  
 theme(plot.title = element\_text(size = 12, face = "bold", hjust = 0.5),  
 legend.title = element\_text(face = "bold"),  
 legend.position = "bottom",  
 axis.title = element\_text(size = 10, face = "bold"),  
 axis.text = element\_text(size = 10)  
 ) +  
 scale\_color\_manual(values = c("Developed" = "blue", "Developing" = "red")) +  
 scale\_x\_continuous(  
 breaks = seq(min(Life\_expectancy$BMI, na.rm = TRUE),   
 max(Life\_expectancy$BMI, na.rm = TRUE),   
 by = 5),limits = c(10, NA)   
 ) +  
 scale\_y\_continuous(  
 breaks = seq(min(Life\_expectancy$Life.expectancy, na.rm = TRUE), max(Life\_expectancy$Life.expectancy, na.rm = TRUE), by = 5),limits = c(40, NA)  
 )

`geom\_smooth()` using formula = 'y ~ x'

Warning: Removed 165 rows containing non-finite values (`stat\_smooth()`).

Warning: Removed 165 rows containing missing values (`geom\_point()`).



The analysis of the relationship between socio-economic factors such as education and BMI—and life expectancy in developed and developing nations provide vital insights for establishing successful public health policy.

Education appears as a significant factor of life expectancy, with both developed and developing countries demonstrating a positive association between years of schooling and longevity. However, the effect is more obvious in underdeveloped nations where gains in education levels lead to large improvements in life expectancy. This shows that educational expenditures in these locations positively generate considerable public health benefits, perhaps delivering dramatic improvements to societal well-being.

From the other plot, displaying the relationship between BMI and life expectancy shows a more nuanced picture. In developing nations, a higher BMI often corresponds with longer life expectancy, perhaps reflecting improvements in nutritional quality and overall health. Conversely, in industrialized nations, the advantages of a higher BMI plateau and even reverse when values near the obese range, demonstrating that excessive weight increase is connected to unfavorable health effects. This underlines the necessity for complex health strategies that address under-nutrition in poorer locations while tackling obesity in more wealthy ones. On obtaining this balance can highly impact the nations improve their population health.

Overall, the results support our assumptions that account the particular socio-economic and health settings of nations. Promoting education and controlling nutritional health via focused interventions highly boost global health outcomes, emphasizing the impact between education, nutrition, and life expectancy in varied economic circumstances.

### **Conclusion**