Life Expectancy

Stanislav Fatkhutdinov, Betapho Hannah, Meryl Patrick

University of Maryland Global Campus

Assignment 5

DATA 610 Section 9040

Decision Management Systems

Vahe Heboyan

Life Expectancy

**Introduction:**

In assignment 5, we will be using the dataset Life Expectancy. This data belonged to the World Health Organization (WHO) and was gathered for health purposes. The dataset is made up of the factors affecting the life expectancy of 193 countries in both developed and developing countries. In this assignment, we will focus on life expectancy between the years 2000-2015. The data consists of 22 columns and 2938 rows. We will analyze factors such as immunization, economy, mortality rates, and other relevant data that affect life expectancy in these countries.

Many of the 22 columns can be described by their title alone; some require context to be interpreted. Polio, Diphtheria, and Hepatitis B columns describe the coverage percentage of immunization. Measles represents the number of cases per 1000 people, and HIV/AIDS defines deaths in the first four years per 1000 births. Alcohol data described as liters of pure alcohol consumed per individual.

Four variables describe economic data. GDP represents a general measure of the country's market value; percentage expenditure represents the percent of this GDP spent on health. Total expenditure represents the percentage of total government expenditure on health. Income composition of resources (ICOR) is a 0 to 1 measure of the availability of resources and can imply the significant quality of life factors impacting each individual in a population.

**Data Refinement:**

The examined data set was large, with nearly three thousand entries with 22 variables for each. With such vastness and complexity, missing values should be expected. In the case of infant deaths, under-five deaths, percentage expenditure, it is assumed that an entry of zero is unlikely to be accurate. These values are replaced with NULL to avoid imprecise analytics. This assumption cannot be applied to zero representing numbers of measles cases as some countries have successfully eliminated the disease. As of 2019, seven states have eliminated the measles virus (Sharma, 2019); however, when filtering nations reported zero measles cases, the data set reports 23 countries. This discrepancy implies that many instances of no reported measles cases should instead be considered as missing values.

Two variables in this dataset appear to be complete; however, some reported entries are very improbable—the first being population. For example, the range 34 to 393 was declared as the population of Maldives; while this nation of islands is relatively small, its people should be closer to 300,000 (Worldbank, 2020). The smallest country included in this dataset is Nauru, with about 10,000 (Vincej, 2020). By filtering out smaller populations than the smallest possible value, these inaccurate small sums will be eliminated from the dataset.

The second variable to be treated with suspicion is BMI. Body mass index is calculated by dividing one’s weight in kilograms by their height in meters squared. Most adults have a BMI between the range of 18 to 30 (CDC, 2020). Extreme examples may fall outside of this range. However, this dataset includes a range from 0.9 to 87, which is likely physically impossible at both ends and more improbable to represent the average of an entire population. It is possible to filter only to represent a feasible BMI range. As the source of this inaccuracy cannot be identified, it may be best not to rely on this variable for insights.

This dataset contains an error in the column’s title representing the prevalence of thinness in ages 10 to 19. The column is erroneously titled thinness 1 to 19, which may lead to an inaccurate data interpretation.

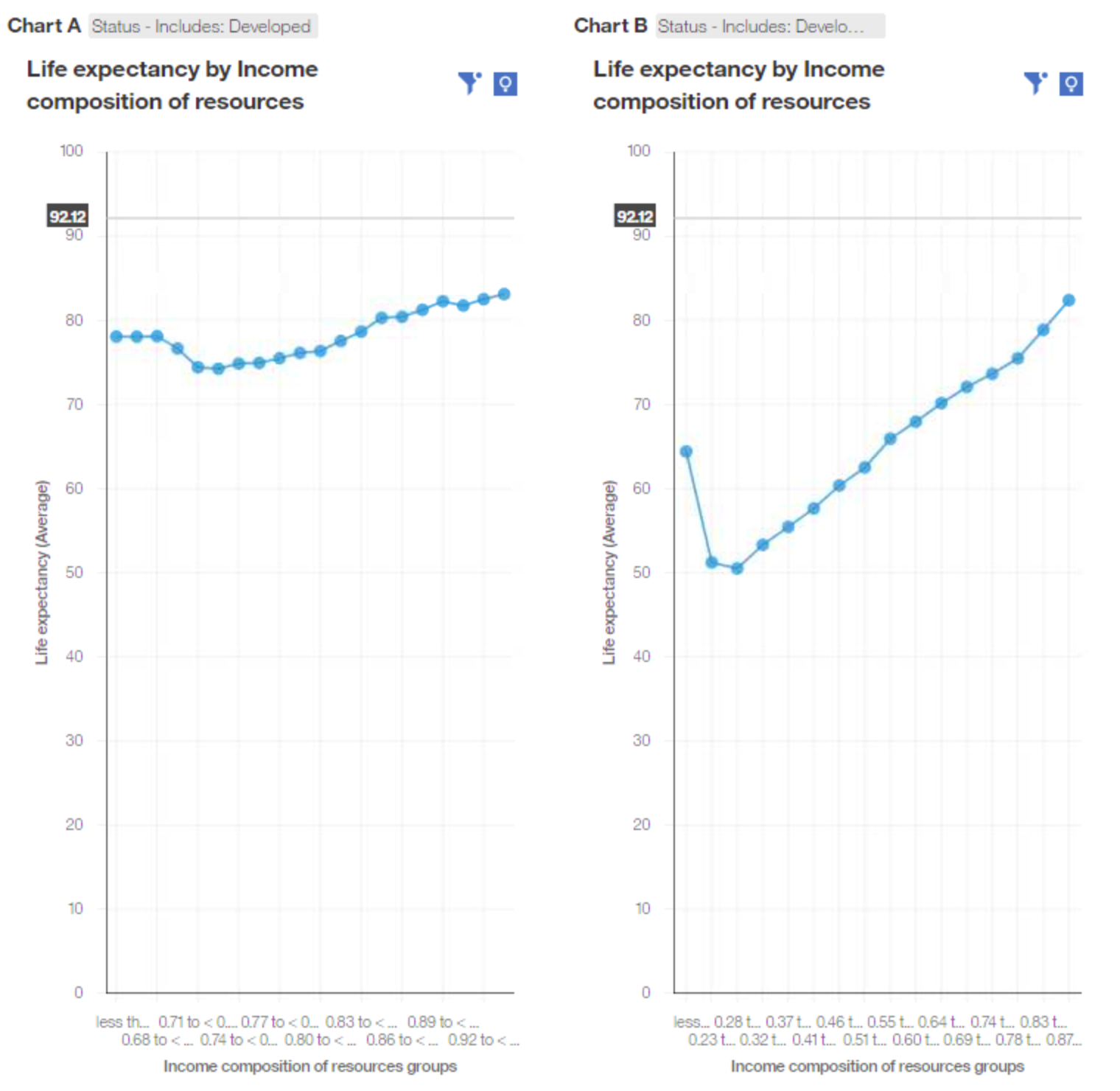
**Data Exploration:**

The 193 countries in this data set are divided economically into developed and developing countries. The ratio of developing countries to developed countries leans heavily toward not only the number of developing countries but the populations of those countries. This status is determined by the United Nations country classification based on necessary economic country conditions (UN, 2014).



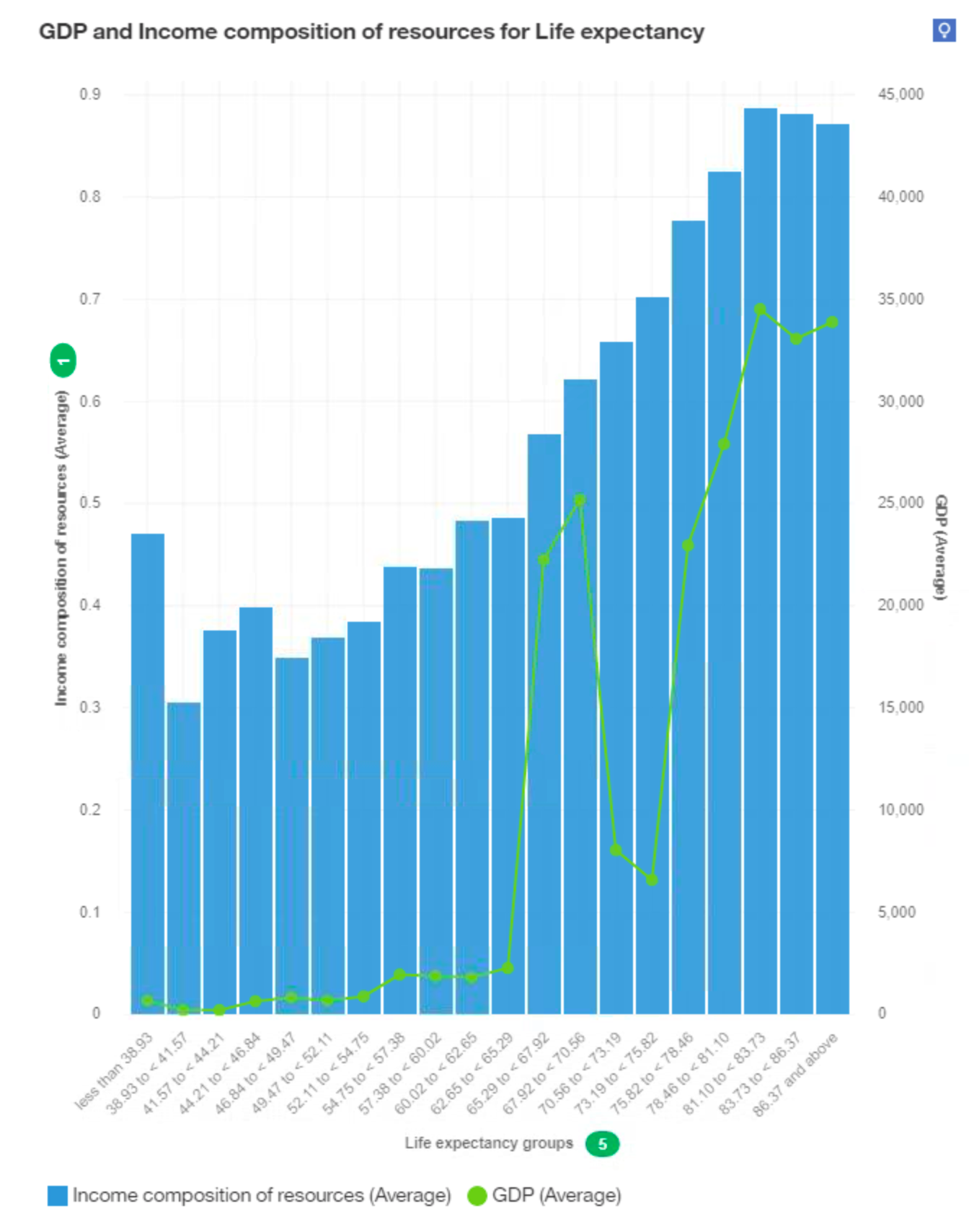
Figure 1. *Word Cloud of Included Countries. Word cloud of included countries separated by developing and developed. The country’s font size is a representation of the relative average population.*

When examining predictive drivers for a country’s life expectancy, the most substantial single driver is identified as the income composition of resources with a predictive power of 64%. GDP yields a predictive power of only 29%, suggesting that how a country's resources are spent may be more significant to its citizens’ quality of life over the amount of market wealth present in that country.

****

*Figure 2. Life expectancy by ICOR. Average life expectancy by income composition of resources grouped by 20. The left image contains the trend for developed countries; the right image includes developing countries’ direction.*

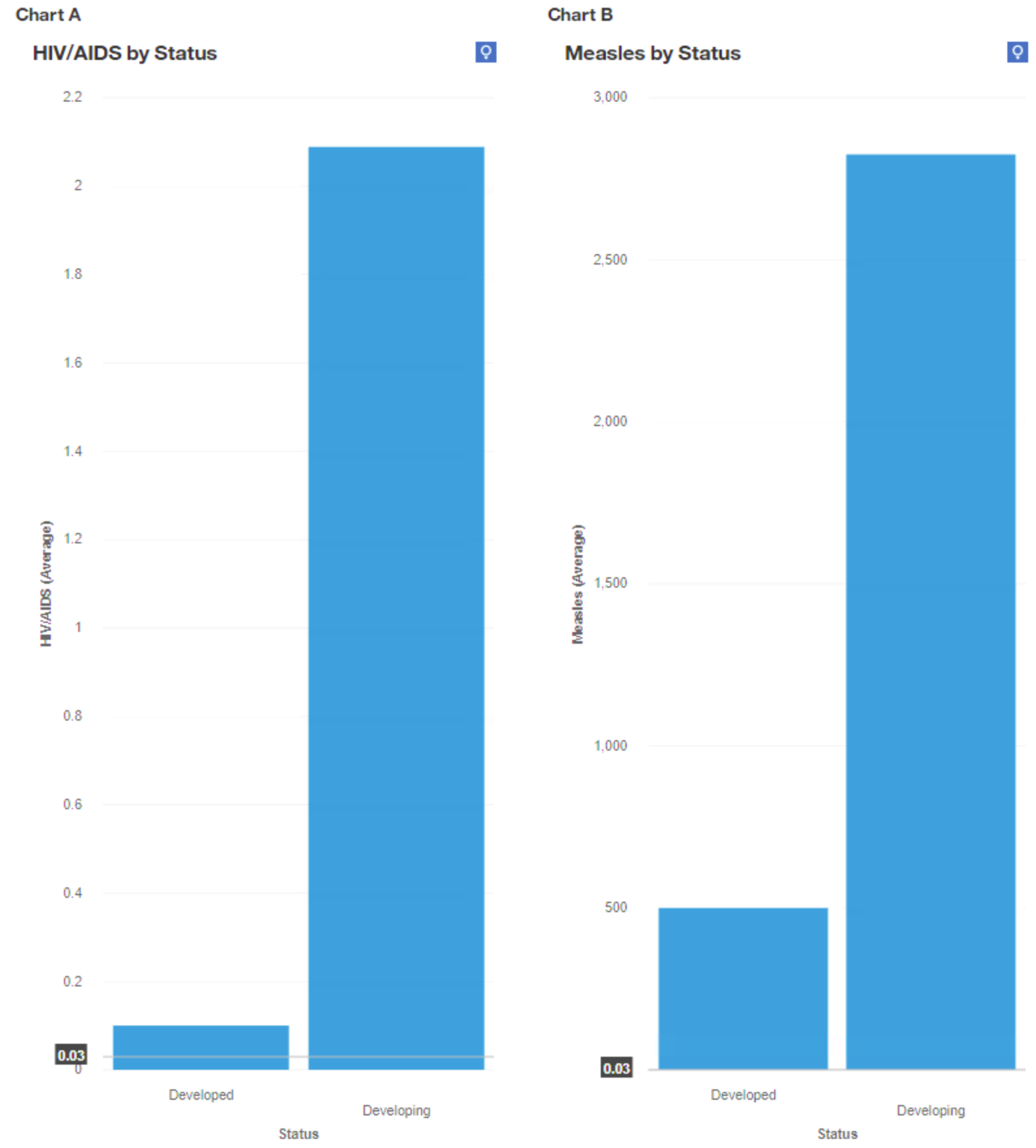
Overall, wealth does lead to greater availability of resources but does not guarantee this to be true.



*Figure 3. ICOR by Life expectancy with GDP. Average ICOR and Average GDP are compared to Life expectancy grouped by 20.*

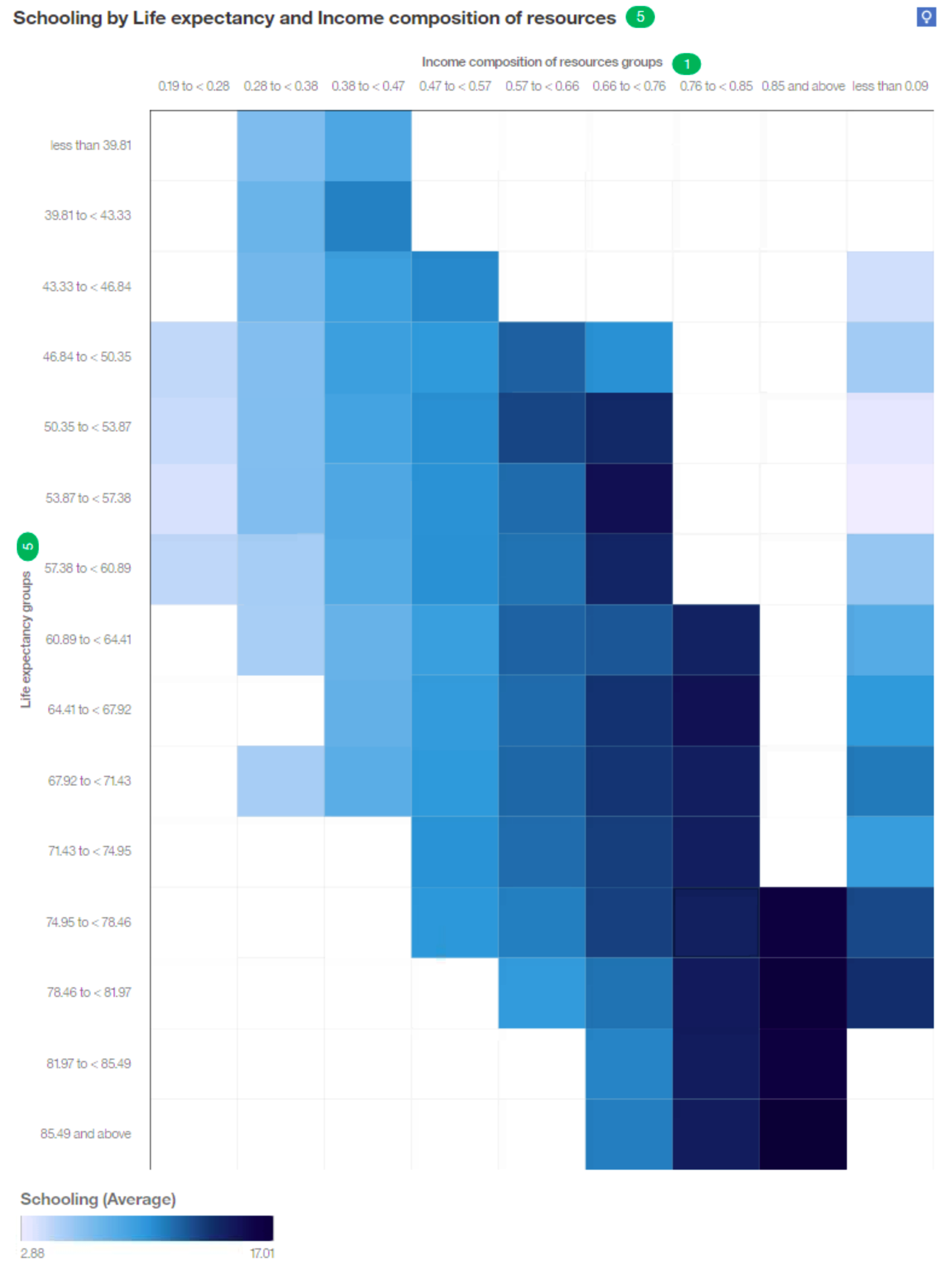
Countries with high GDP are expected to have a higher life expectancy. Countries with a life expectancy between 70 and 85 years can be observed to have a much lower GDP than expected, given their citizens’ longevity. These countries do fit the trend of the high-income composition of resources.

When visualized by developed and developing countries, it can be observed that developing countries are more susceptible to diseases. This relationship may intensify the importance of investing in the medical infrastructure in underprivileged countries.



*Figure 4. Average HIV and Measles by Status. The left image indicates average HIV/Aids-related deaths in children under four in developed vs. developing countries. The right image suggests the average number of measles cases per population of 1000.*

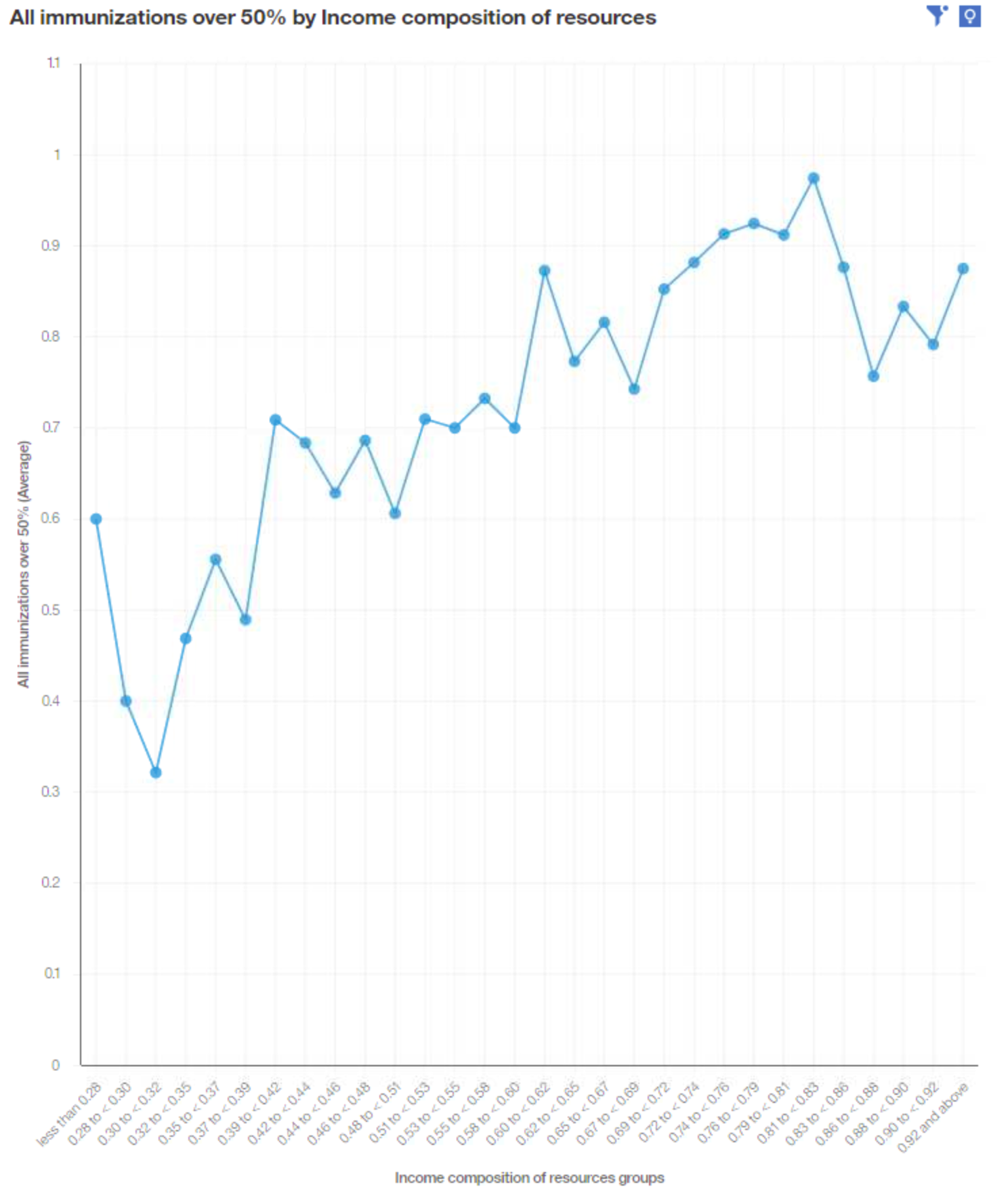
Developing countries have a much higher prevalence of diseases in this data set; this is likely related to the importance of available resources and average life expectancy observed in Figure 2. Resources that impact the economic independence of an individual are also critical to the quality and longevity of life.



*Figure 5. Heat Map of Schooling to ICOR and Life expectancy. Heat map describing the average years of schooling by average life expectancy grouped by ten and income composition of resources grouped by 10.*

In Figure 5, a clear pattern describing the qualities of countries with more years of education score higher on available resources and have a higher average life expectancy. This observation is substantiated by a 2011 study which found that within the United States, individuals with higher educational attainment were likely to live longer (Hummer, 2011).

Income composition of resources (ICOR) measures the general availability of resources. A conditional statement was added to determine if this is related to the availability of immunizations. If a row contained an immunization rate of over 50% for polio, diphtheria, and hepatitis B, this row equated to 1. If not, the row equates to 0, immunization rates for years 2000 to 2015 can be averaged to identify a percentage of overall immunization availability.



*Figure 6. Immunization rate by ICOR. A calculated average immunization rate by income composition of resources grouped by 30.*

A positive trend can be observed between the average rate of immunization over 50% and the income composition of resources.

**Decision Tree Models:**

A factor that reflects both the health and wealth of an individual is thinness prevalence. To understand the drivers, which impact the percentage of thinness prevalence ages 5 to 9 and ages 10 to 19 have been combined into average thinness among young people. In this decision tree, factors that directly relate to body composition have been filtered out not to count these measurements double. This includes thinness for 5 to 9, 10 to 19, and BMI.



*Figure 7. Average Thinness Decision Tree. Decision tree of average thinness prevalence percentage. Thinness in 5 to 9, thinness in 10 to 19, and BMI have been excluded as drivers.*

Much like many other factors in this data set, the most vital individual driver is the Income composition of resources, followed by measles cases per 1000 individuals. Countries with the highest-scoring ICOR greater than 0.797 and highest alcohol consumption, more significant than 2.38, are predicted to have the lowest average thinness prevalence of 1.14%. As predicted thinness prevalence rises, more health-related factors are considered. Countries with an ICOR less than 0.797 and more significant than 0.624, Alcohol consumption less than 5.36, and measles greater than three are predicted to have a higher average thinness prevalence of 5.44%. Four variables predict those anticipated to have the highest prevalence of thinness of 18.35%, accounting for 2% of the dataset. ICOR is less than 0.624 and more significant than 0.461, Measles greater than 750, HIV less than 1.6, and more substantial than 0.2, Adult Mortality less than 261.

These predictions can be useful in the context of humanitarian aid. By isolating factors that define countries with high rates of underweight citizens, programs to benefit overall health and quality of life can be developed. Resources that may have been wasted can be avoided. For example, high rates of alcohol consumption are often associated with poor health. It can be seen that it is a factor that defines populations of a healthy weight does not impact overall health as much as disease prevention and immunizations may. Resource composition is clear to be a significant factor in the well being of a country's populace. Data-driven predictive analytics may guide how to invest in a community efficiently and effectively.

Media 1 represents the life expectancy decision tree. The most important predictors are Income composition of resources (ICOR), HIV/AIDS, and Adult Mortality. This breakdown follows the following rules: the income composition of resources ≧ 0.79 has the highest predictive value of 81%, meaning that the larger the income composition of resources is, the higher the life expectancy. The second highest factor in predicting life expectancy is Adult Mortality rates. The higher the adult mortality number, the longer the longevity. The next factor playing a role in determining life expectancy is HIV/AIDS rates. Unsurprisingly, the life expectancy goes up when the rates of HIV/AIDS are down. Media 1 demonstrates the change in the decision tree depending on the countries’ development status throughout the given timeframe.

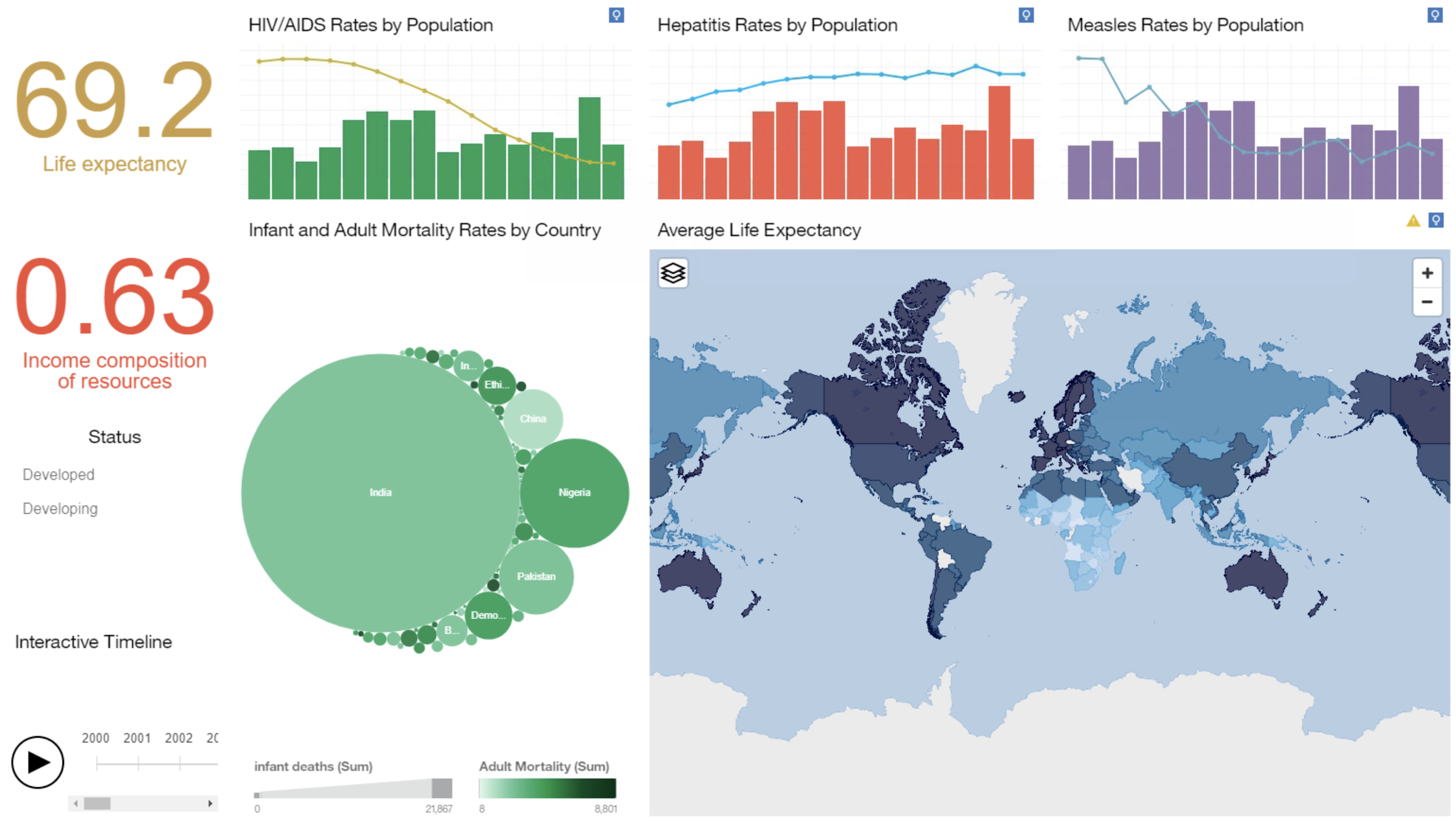
[](https://drive.google.com/file/d/1qazy5tyGSkPOMbhl9be5OPmpRHMb7tOq/view?usp=sharing)

*Media 1. Life Expectancy Decision Tree Timeline*

Both decision trees make predictions about general health, life expectancy, and thinness prevalence. The importance of the plurality of HIV/AIDs related deaths to both health indicators exemplifies the extent of disease management in populations.

**Dashboard Assembly:**

Media 2 represents the interactive dashboard to interact with the life expectancy data quickly. The dashboard contains information regarding life expectancy broken down by countries and adult and infant mortality rates. The dashboard also includes interactive filters that allow the user to sort by countries, years, or the developed-undeveloped country status. The timeline allows the user to see the changes in the entire world or a selected state and see if the HIV/AIDS, Hepatitis, and Measles Rates affect the population numbers. The Life Expectancy and the ICOR tickers reflect on how much a particular country spends on healthcare and how it affects life expectancy.

[](https://drive.google.com/file/d/14is668y0PPRWhZoycnI-jMd40LNr9-Yb/view?usp=sharing)

*Media 2. Interactive Dashboard Visualization.*

As seen in the video demonstration of the dashboard features, quick insights are easily gleaned regarding the population’s life expectancy. For example, infant mortality is highest in India; however, their adult mortality is ten times lower. Another insight shows that despite the relative increase in Hepatitis B vaccination and Measles rates, India’s population remains unaffected.

The strongest predictor of the life expectancy of any given country represented by the ICOR index, the higher the number of income composition of resource index is, the higher the life expectancy. The downward line of HIV/AIDS and Measles rates are trending globally; however, neither of them affects the life expectancy as much as the ICOR index.

Another insight a user can see from this dashboard, for example, is that Romania, a developed country, has some of the low infant mortality rates. However, their adult mortality rates are one of the highest in all the developed countries. As demonstrated in Media 2, their life expectancy was declining through the given timeline.

**Storybook presentation:**

[****](https://drive.google.com/file/d/1w_tnwtLs1dOeczlY5J36Iw6U9a86AjC4/view?usp=sharing)

*Media 3. Life Expectancy Storyboard (click to play the video)*

The created storybook outlines the observed data and patterns found between developing and developed countries and the factors impacting average life expectancy. Using data from the World Health Organization collected from the year 2000 to 2015. We focus on several metrics, including gross domestic product, income composition of resources, and a metric representing the availability of resources. We discuss diseases by identifying the prevalence of HIV/AIDs as well as Measles in each population as well as the coverage percentage of immunization of Diphtheria, Hepatitis B, and Polio. We also compare the differences in infant mortality rates and adult mortality, defined as deaths between ages 15 and 60 per 1000 individuals.

The life expectancy trends upwards as the ICOR index and GDP increase. The difference in life expectancy between developing and developed countries is evident in the different “starting points” as the lowest life expectancy for developed countries is higher than that of developing countries. It can be seen that GDP does not rise with the same consistency as ICOR for groups with higher expected life spans, implying that wealth on its own does not fully account for the health of a population. Even as the world becomes overall less diseased, a large disparity in infant and adult mortality can be seen between developed and developing countries.

Adult mortality, the probability of death between the ages of 15 and 60, is drastic when comparing developed and developing countries. The prevalence of infant death is even more drastic than a country’s status; developed countries are almost imperceivable.

Disease prevalence of measles is much higher in developing countries than in developed countries. The number of measles cases decreased in both populations over time, likely due to healthcare advances.

Vaccination coverage of polio, hepatitis B, and diphtheria are lower in all cases for developing countries, exemplifying the available resource disparity between the two groups.

Thinness prevalence for both children aged 5 to 9 and 10 to 19 is low in nations with a high life expectancy regardless of status. A high prevalence of thinness may be an example of food resource availability reducing hunger and underweight citizens in nations with increased life expectancy. Though thinness prevalence is lower in all groups of developed countries, highlighting that a gap in resources may still exist between these groups even with a similar life expectancy.

A social resource, average years of attended schooling, shows a pattern of high life expectancy in nations with higher years of school attended. Once a value of 14 years of school attended, developed and developing countries appear indistinguishable when these two factors are compared.

**Conclusion:**

Human lives are immensely complex; this data set could not possibly describe all possible factors which determine the lifespan of an individual. With a data-oriented approach, key indicators that define healthy long-life spans in a country can be identified. Only with these indicators, such as education and immunization coverage, can an actionable strategy be created. By addressing the discrepancies of resources in developing countries, there is hope for life expectancies to be increased. The data we used for this assignment gives us an insight into the discrepancy of the available resources needed to improve life expectancy. From the insights above, one can see that the average life expectancy for the developed countries is 79.2, while that of developing countries is 67.08. This discrepancy in resources between the developed and the developing countries have dramatically affected life expectancy. Even though the data shows a vast disparity in resources, there is still hope and a future for the developing countries to increase their life expectancy by increasing targeted resources.

**References:**

About Adult BMI. (2020, September 17). Retrieved November 03, 2020, from <https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html>

Hummer, R. A., & Hernandez, E. M. (2013). The Effect of Educational Attainment on Adult Mortality in the United States. *Population Bulletin*, *68*(1), 1–16.

Population Total. (n.d.). Retrieved November 03, 2020, from https://data.worldbank.org/indicator/SP.POP.TOTL?end=2013

Sharma, S. (2019, July 9). Step up efforts to eliminate measles: WHO. Retrieved November 03, 2020, from <https://www.who.int/southeastasia/news/detail/09-07-2019-sri-lanka-eliminates-measles>

UN (2014), *World Economic Situation and Prospects 2014*, UN, New York, <https://doi.org/10.18356/ee1a3197-en>.

Vincej, V. (2020, June 26). 10 Smallest Countries in the World by Population - [2020]. Retrieved November 03, 2020, from https://www.travelinglifestyle.net/smallest-countries-world-by-population/

**Appendix A**

**Calculations**

AllImmunizations greater than 50% coverage

IF ((Polio) > 50 and (Hepatitis\_B) > 50 and (Diphtheria) > 50 )

THEN (1)

ELSE (0)

Average Thinness Prevalence

(thinness\_1\_19 + thinness\_5\_9\_years) / 2