

**Helena Van Ess**

**October 21, 2023**

**Life Expectancy Among Countries**

**A person and person walking up a bar graph

Description automatically generated**

**What is your project’s context?**

**The data is obtained from the website Kaggle and the Global Health Observatory (GHO) data repository under World Health Organization (WHO) and the United Nations that collect the health and other metrics of the countries around the world. Within the 193 countries that are listed in this data, there is representative health data for each of the countries among a 15-year period between 2000 and 2015. This data emphases the importance of the public health and the resources available to allow for a long-life expectancy. This data is important because it allows for the analysis of the provided metrics to see the relationship between the metrics and the life expectancy. We want to understand the differences between those countries with a high life expectancy and the countries with a lower life expectancy. Being able to analyze this and show this will allow for changes to be made at a leader and government level to increase the expectancy. The data needs to be stressed and understood because of the drastic differences among key metrics.** I see the main purpose of this project to see the main causes for low life expectancy. I would like to see which factors are closest related to the life expectancy and which are not. I would also like to see from 2000-2015 if the life expectancy is rising or falling in countries with a lower life expectancy in the year 2000. The population in each country will also be interesting to compare to life expectancy as well. **By analyzing and understanding the morality rate and the healthcare expenditure, one can hope to see the keys to a longer life expectancy. One constraint to keep in mind is this data is only from the year 2000 to the year 2015. However, this 15-year period allows for one to see how life expectancy has changed overtime with respects to certain metrics. These metrics will be prepared through line graphs to show data overtime, text fields, charts, and geographical graphs as well. The leaders of the countries that have a lower life expectancy should see the difference between them and the countries whose life expectancy has grown overtime. The life expectancy does not occur equally through the countries. With the proper analysis, the leaders of the countries provided in the data can show key metrics that should be altered in efforts to increase life expectancy within their country.**

**Audience**

My audience is the leaders and the people of power within the countries. That being said, I want my audience to see the severity of the relationship between their country and other country’s life expectancy. The focus of the data will be the top and the bottom countries to see what metrics differ from each other. This will begin conversation of what needs to be changed in order to increase the life expectancy of the countries that have a lower life expectancy. The audience impacts my work as a data visualization designer by stressing what aspects need to be shown in order to show the differences in the life expectancy. As an analysis presenter I need to be able to show the metrics that are completely different than those that have good life expectancy. Then I need to properly portray those metrics to address the factors that can increase life expectancy.

**Data Set**

      It is important to note that within this data set there was no evident errors found after obtaining the data from **the World Health Organization (WHO). This ensures that the data is credible, and the data has been cleaned through R software by using Missmap command. The data set included 22 metrics and dimensions for each of the 193 countries. It is worth noting that some countries have very different metrics, however, this does not deem the countries to be outliers that need to be taken out. It is important to be able to see these countries to see what is so different between them and others.** Within this dataset there are some metrics and dimensions that need to be explained before analyzing this information. One dimension that could use of explanation is status. The status of the country is either presented by “Developed” or “Developing.” The metrics that are going to be used should also be cleared up. The life expectancy is the life expectancy in age. The adult morality rate is the rate among both sexes. This is the probability of the adults dying between 15 and 60 years of age and this data is per 1000 population. The total expenditure is the expenditure on health as a percentage of total government expenditure. The population is the total population of the country. The infant deaths are the number of infant deaths per 1000 population. The under-five-deaths is also per 1000 population.

**Constraints**

There were only a couple issues I encountered while working with this data source. There were a couple mistakes in the metrics in the dataset that I had to ensure I knew exactly what it was supposed to mean. For an example, there is a thinness percentage for 5-9 years old and 1-19 years old. That being said the 1-19 years old didn’t make much sense when analyzing the data. I did some research within the dataset and the credible website the data was obtained from and found out that it was supposed to be thinness 10-19 years old. In addition to the mistake in metric names, there are some countries that are either not included in the dataset and some that have missing data for some metrics. Luckily, the data that is missing is from some of the less known countries like Vanuatu, Tonga, Togo, and Cabo Verde. In addition to the less known countries, the metrics that are missing are also the metrics that I am not planning to use in my project like Hepatitis B and GDP. Another issue that I came across is that the data is until 2015. This is about 8 years ago but the analysis through the 15-year period from 2000 to 2015 should allow for a good conclusion based on how the metrics relate to life expectancy.

**Data Source Name:** Life Expectancy

**Total # of records (rows):** 2,938

**Dimensions you plan to use:**

Country, status

**Metrics you plan to use:**

Year, life expectancy, population, infant deaths, under-five-deaths, adult morality, total expenditure.

Metrics that also might be used but still need to be assessed:

Alcohol, BMI, schooling

**Dashboard**

**1-sentence summary of what your dashboard will contain:**

Life expectancy among the top five and bottom five countries.

**List at least 5 different data elements that will be in your dashboard (charts, tables, etc.)**

1. List: Top five countries life expectancy and bottom five life expectancy and if they are developing or developed.
2. Chart: Line chart showing the life expectancy overtime for the desired countries. (line for each country)
3. Chart: Death rate among top five and bottom five countries.
4. Chart: Total expenditure on healthcare compared to life expectancy
5. Chart: Life expectancy among countries (geographical)
6. Chart: Morality rate among desired countries.

**Analysis Presentation**

**1-sentence summary of what your analysis will focus on:**

Analysis of life expectancy among countries with the relation to healthcare expenditure and morality rate.

**List the specific questions you will ask/answer with your analysis:**

* Is there a relationship between healthcare expenditure and life expectancy?
* What is the trend over time for the top and bottom country’s life expectancy?
* What does the life expectancy look like among countries across the world?
* Is there a relationship between morality rate and life expectancy?
* Is there a relationship between morality rate and healthcare expenditure?
* Does population have a relationship with life expectancy?