Name: Kun Zhang
StudentID: 4081903

Life Expectancy Data Visualization

The Life Expectancy dataset explores the key factors that affect life expectancy in different countries, focusing on health, economic, and social aspects. The figures in this report highlight both positive and negative influences on life expectancy. Variables like schooling and GDP have positive effects, while diseases are linked to shorter life expectancy.

The report uses visualizations such as **word clouds**, **scatter plots**, **networks**, **heatmap** and other types of graphics to help readers better understand the key factors that influence life expectancy. These visualizations make it easier to identify actionable insights for improving quality of life.

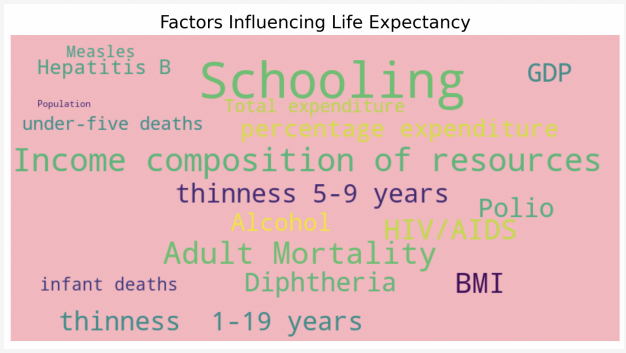

Figure 1: factors influencing life expectancy

Figure 1 I use **Wordcloud** to generate an overview of which variables have the most influence (positive or negative) on life expectancy. From the figure we can see schooling, income composition of resources, adult mortality BMI has either positively or negatively affects our life expectancy. Next I will use scatter plots to find out what elements have the positive and negative correlation with life expectancy.

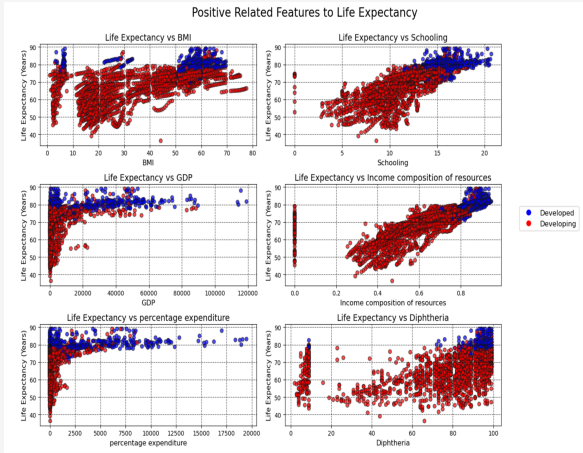

Figure 2: positive related features to life expectancy

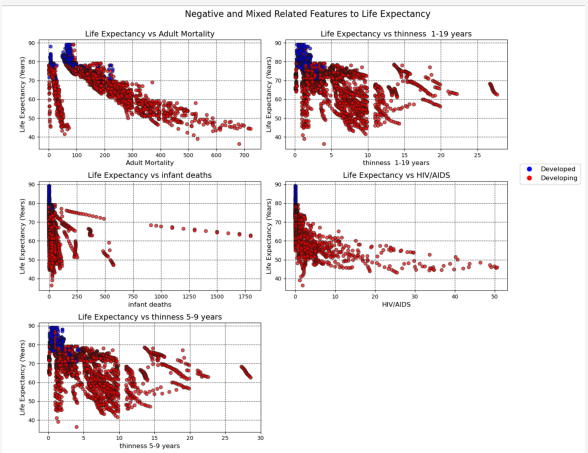

Figure 3: negative related features to life expectancy

Name: Kun Zhang
StudentID: 4081903

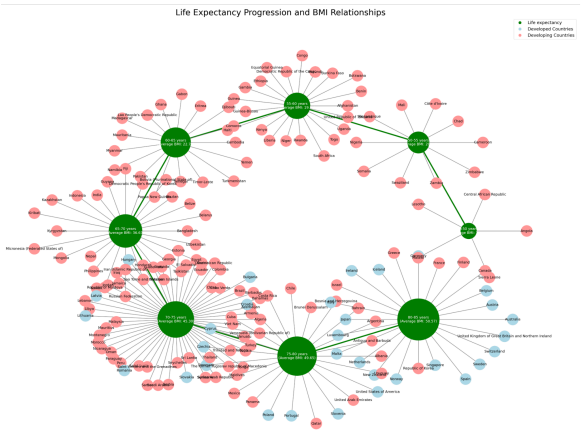

Figure 4: life expectancy progression and BMI relationship

Figure 2 & 3 use **Scatter plots** shows:

- BMI, Schooling, GDP, Income Composition of Resources, Percentage Expenditure, and Diphtheria show a generally **positive correlation** with life expectancy.Countries with higher levels of these indicators tend to have longer life expectancies, particularly in developed nations.
- Adult Mortality, Thinness (1-19 years and 5-9 years), Infant Deaths, and HIV/AIDS demonstrate a **negative correlation** with life expectancy. Higher rates of these factors are associated with shorter life expectancies, with developing countries being more affected.
- Data points from developed countries (blue) generally cluster towards higher life expectancies, while developing countries (red) show more variation and are often associated with lower life expectancies.

Scatter plots were created for the factors that correlated to life expectancy, and Each subplot examines the relationship between life expectancy and a single factor to highlight trends and differences.

Figure 4 tells that:

- Countries are grouped by life expectancy into clusters (e.g., less than 50, 50-55 years, 60-65 years, 70-75 years etc).
- The nodes of each cluster displays the average BMI and includes connected countries with similar characteristics.
- **Developed countries** (blue) are predominantly associated with **higher life expectancy** clusters (e.g., 70-75 years and above).
- **Developing countries** (red) are distributed across all clusters but are more concentrated in **lower life expectancy categories**.

A **network graph** was generated by using **netwrokX** to visualize relationships between countries based on life expectancy and average BMI. The nodes represent age groups and average BMI, the vertices represent the countries that are located in the certain clusters.

Name: Kun Zhang
StudentID: 4081903

From the analysis, the data show that developed countries generally have a higher life expectancy compared to developing countries. Next, we aim to examine the distribution of life expectancy for both developed and developing countries to gain deeper insights.

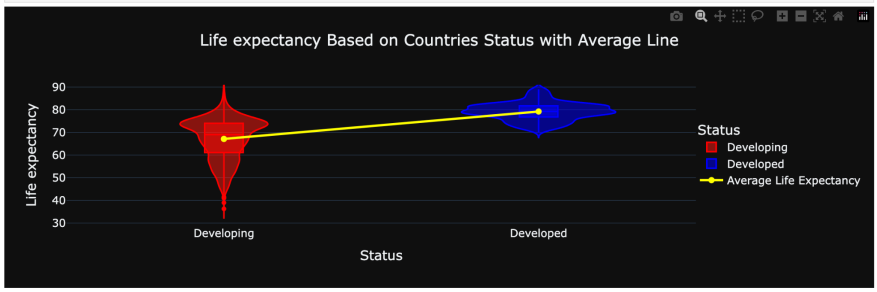

Figure 5: life expectancy with average line

From figure 5,

- Developed countries show a **higher median** and narrower spread in life expectancy, indicating greater consistency across these nations.
- Developing countries have a **wider distribution**, reflecting significant variation in life expectancy across these nations.
- The average life expectancy line shows that life expectancy is significantly higher in developed countries.
- The density of the violin shapes indicates that a larger proportion of developed countries cluster around higher life expectancy values (e.g., 75-85 years), while developing countries have a wide range.

This **Violin Plot** was created to visualize the distribution of life expectancy within each category. The shape of the violin reflects the density of the BMI; wider sections indicate higher concentrations of values.
Line Chart represents the mean value of both developed and developing countries, the line added in the figure is to enhance the visual comparison.

Next, we will identify the locations with higher and lower life expectancy on the world map. Additionally, we will determine the top 15 and bottom 15 countries in terms of life expectancy and identify the continents where they are located.

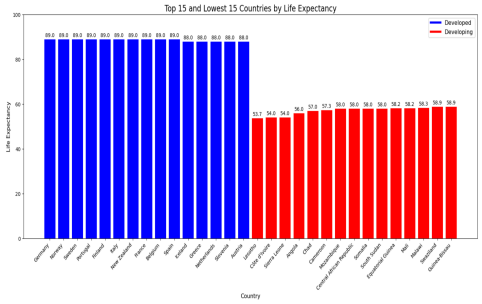

Figure 6: top and lowest 15 countries' life expectancy

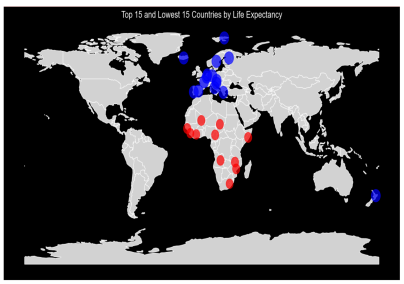

Figure 7: top and lowest 15 countries location

Name: Kun Zhang
StudentID: 4081903

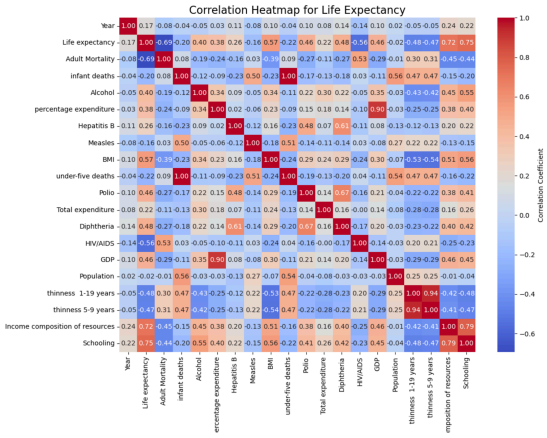

Figure 8: correlation heatmap for life expectancy

Figure 6 shows:

- Top 15 countries (blue): Life expectancy > 80 years, indicating better health and socioeconomic conditions.
- Bottom 15 countries (red): Life expectancy < 55 years, reflecting significant health challenges.

Figure 7 indicates:

- Top countries are concentrated in Europe, North America, and parts of Oceania.
- Bottom countries are mostly located in Africa.

Figure 8 shows:

- The heatmap highlights the strong positive impact of factors such as **schooling**, **GDP**, and **income composition of resources** on life expectancy.
- Negative factors like **adult mortality**, **infant and under-five deaths**, and **HIV/AIDS** significantly reduce life expectancy.
- Economic and health indicators (e.g., **BMI**, **alcohol consumption**) show moderate to strong associations with life expectancy, suggesting their nuanced role in overall health outcomes.

The **bar chart** and the **map** were created to identify countries with the highest and lowest life expectancy and to visualize their locations on the world map. A **heatmap visualization** was created using **Seaborn**, to highlight the strength of these relationships.

The presented figures reflect the overall health and well-being of populations, highlighting disparities between developed and developing countries. They help identify global inequalities. By presenting these insights, we empower policymakers, organizations, and communities to prioritize resources to put priority on factors like education, healthcare and GDP, and advocate for equitable health systems worldwide.

Github link: https://github.com/lovestefanie7/Data-Visualization-Final