

Mini Project 1

Meghana Boinpally (meboin), Raghav Chegu Shyam Kumar (racheg), Shefali Luley (sluley)

Introduction:

In the following analysis, we are going to answer a specific research question posed to us - Can the increase in life expectancy since World War 2 be largely explained by increases in GDP per capita?

To answer the above question, the **gapminder** dataset is used as it contains the required information from the year 1952 to the year 2007. This EDA will thus give an overview on how the relationship has changed between the life expectancy and GDP per capita over the years.

Part 1: GDP and Life Expectancy in 2007

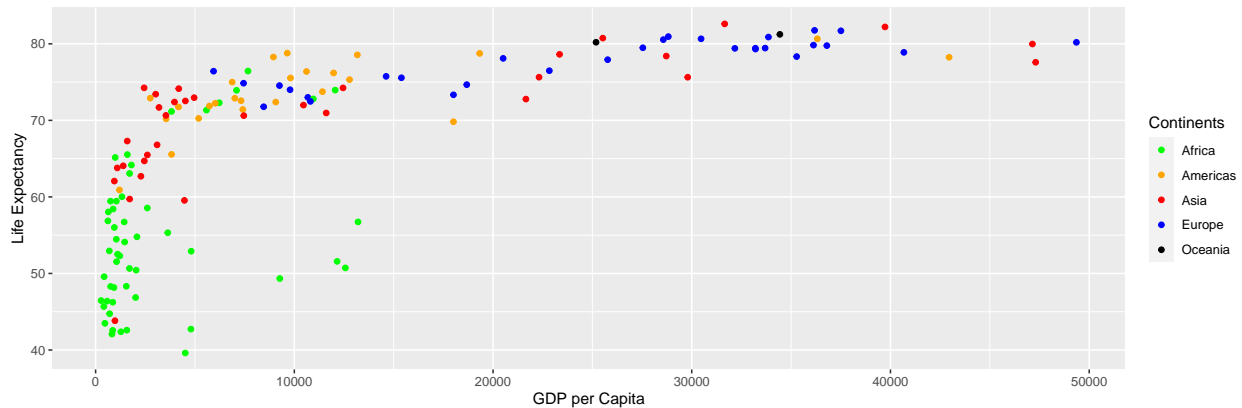


Figure 1: Life Expectancy vs GDP per Capita

Life Expectancy varying with GDP per capita in 2007

- We can observe in the Figure 1 that the overall life expectancy of the countries is not particularly linear with the GDP per capita.
- It can also be noted that the continents with higher GDP per capita, such as Europe and America, also have longer life expectancy.
- In this case, the life expectancy also seems to be positively correlated with GDP. We also notice that where GDP is extremely low as we can spot in the continent of Africa, the life expectancy is also very less.
- Here, we can observe that the GDP is uncorrelated with life expectancy below certain a threshold (around 5000).

Patterns observed across Continents

- It can be observed that every continent has a different pattern, yet there are a few similarities.
- In Africa, we can notice that the countries are underdeveloped, as seen by their low GDP per capita and lower life expectancy.

- In America, Europe, and Asia, there is linear dependency between the life expectancy and the GDP per capita. However, we can see that the initial values in Asia's distribution closely reflect the pattern seen in Africa, while the rest follows the pattern seen in Europe.
- We cannot establish any concrete relationship for Oceania because there are just two data points.

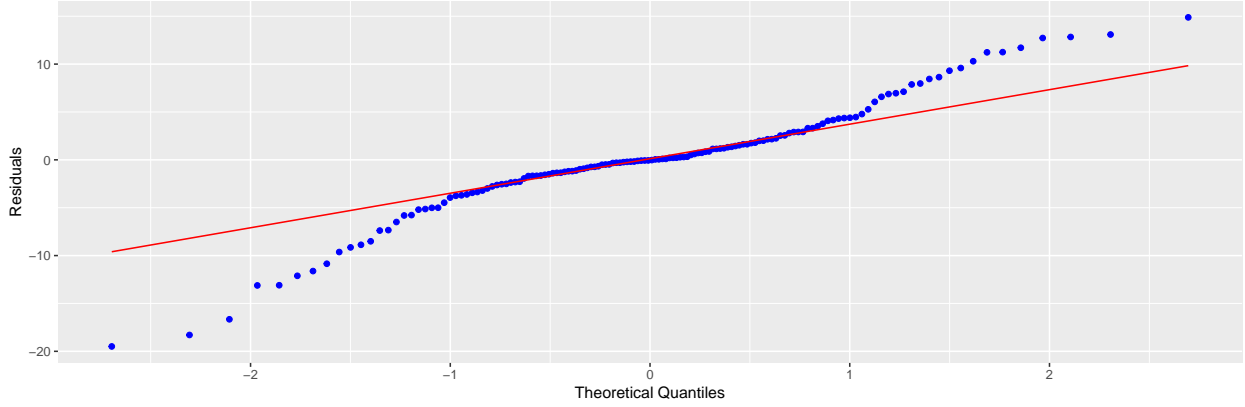


Figure 2: Normal Q-Q plot - Linear Model Residuals

Trends explained by Models

- From Figure 2, it can be observed that the graph follows a log distribution and fitting a linear model on the untrained parameters will not fit the data.
- To address this issue, we used the log transformation on the variable GDP per capita. After running the model, we can observe in Figure 2 that the linear model does not perfectly fit the data because the linear model's residuals do not represent a normal distribution.
- This asserts that a simple linear model will not fit the data and require a complex model can be used to better characterize the data.

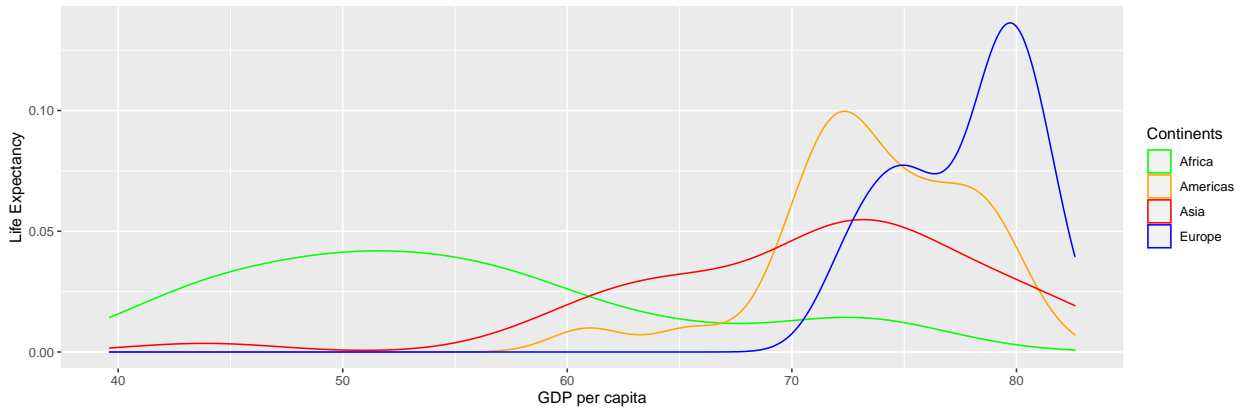


Figure 3: Density distributions of Life Expectancy vs GDP across Continents

Quantifying the difference between the continent's distributions

- In Figure 3, it can be observed that the distribution of each continent is not similar in any aspect and displays lot of variance between them.
- The observed differences cannot be explained with a simple additive or multiplicative shift and needs a more complicated explanation.

- For example, Africa and Asia have a significant difference between their centers and the number of points in the common region is very low.
- The Oceania continent is removed from this analysis as it is skewing the distribution graph with just it's two data points.

Part 2: Life Expectancy over time by Continent

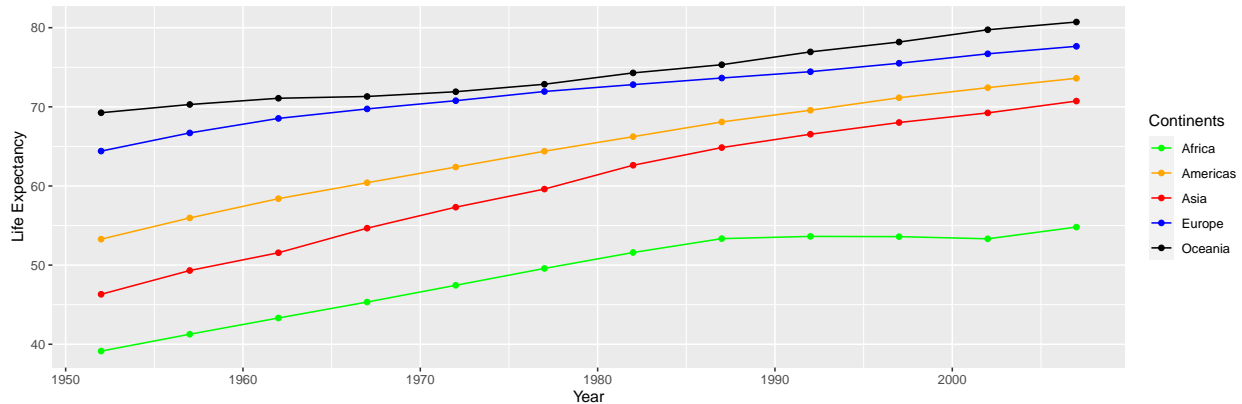


Figure 4: Life Expectancy of the different Continents over Time

Average Life Expectancy over Time across Continents

- In Figure 4, we observe that every continent's life expectancy increases over the years indicated by the positive slope of each continent. This growth can fairly seem somewhat linear.
- It is also worth noting that over time, the life expectancy of developed continents has grown far more slowly than that of developing continents, possibly because life expectancy has hit a saturation point.
- From the 1950s through the 1990s, the life expectancy in Africa increased dramatically, rising from roughly 35 to over 50 years old after which the rate of growth is close to 0.
- From the 2000s forward, America appears to have partially caught up with Europe.

Countries Contribution to the Average Life Expectancy

Figure 5 shows that the life expectancy increases in all the continents over the years, this can be determined by the uniformly increase in the median of life expectancy over years in every continent. To determine whether this increase is uniform for every country or because only a few selected countries progressing much faster than others we analyze the box plot. This can be determined by the interquartile range, lower and higher end in each of the box plots.

- In America, over the years we see that the inter-quartile range decreases as the median life expectancy increases which show us that all the countries are uniformly developing.
- For Africa, the median life expectancy is slowly increasing over the years but it can also be observed that there is an increase in the inter-quartile range over the years. This implies that there might be some countries in Africa that are developing at a faster pace compared to others.
- Europe is following a similar trend to the other continents where the median life expectancy is increasing over the years. The countries are developing which can be indicated by the small inter-quartile range.
- It can be observed in Asia that the overall trend in median life expectancy seems to increase although there is a slight decrease in the interquartile range over the years. This indicates that over the years the countries are catching up in the development index.
- Finally, although America and Asia show a similar trend, due to the slight variation of their inter-quartile ranges, there is difference in the overall development of the countries.

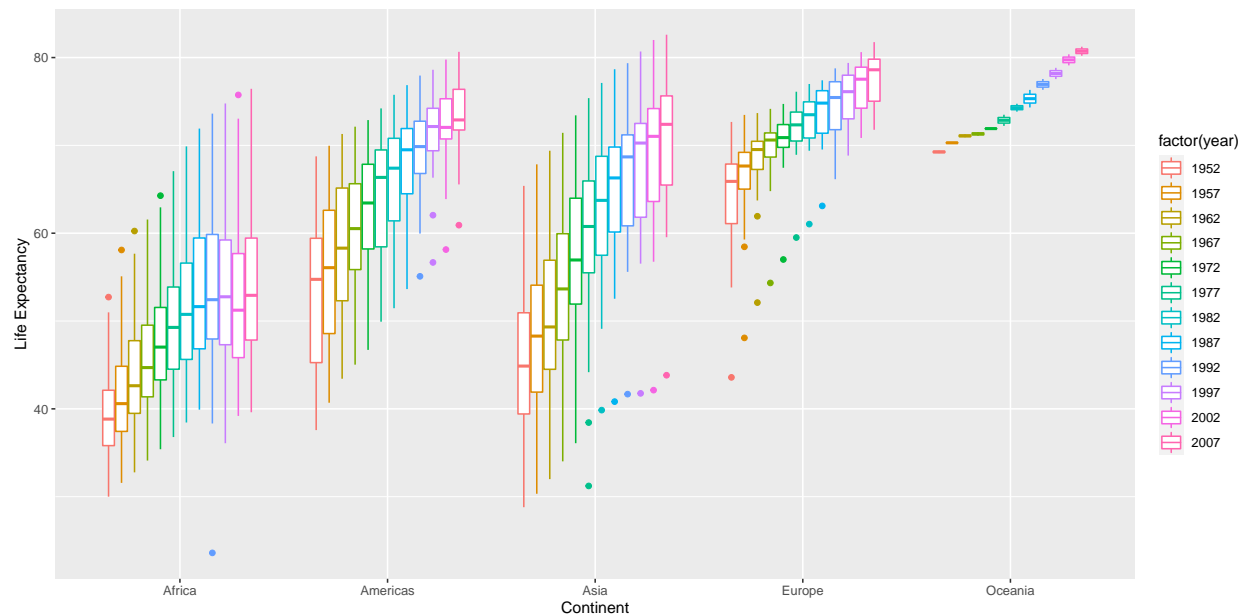


Figure 5: Life Expectancy and GDP of the different Continents over Time

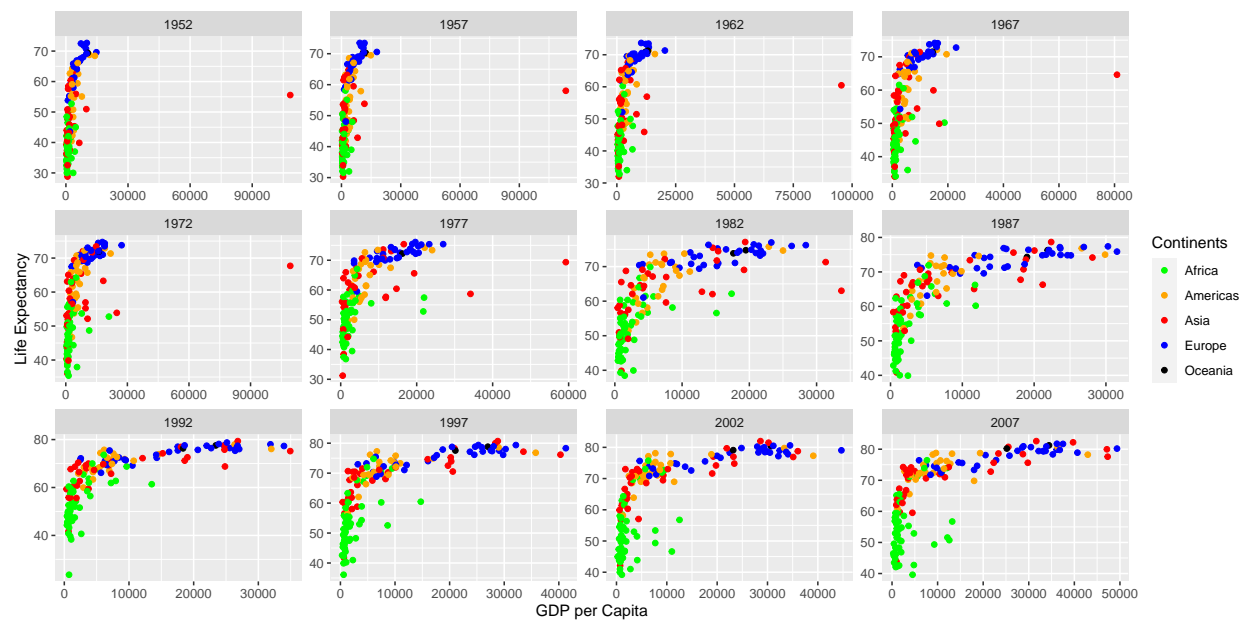


Figure 6: Life Expectancy and GDP of the different Continents faceted by Year

Part3: Changes in the relationship between GDP and Life Expectancy over time

Relationship between GDP and Life Expectancy:

GDP has continuously increased over time, with notable gains in continents such as Europe and America. The relationship between GDP and life expectancy appear to have a linear effect from around the age of 70.

- Referring Figure 5, we can see that the relation between GDP and life expectancy varies between continents. In comparison to other continents, Africa's GDP has barely increased.
- A much lower level of GDP, as seen in Africa, translates to a very low level of Life Expectancy, with a little increase in GDP and Life Expectancy through time.
- In America, Figure 5 shows the inter-quartile range shrinks as median life expectancy rises over time. This demonstrates that all countries are developing in the same way.
- Europe regardless of a large increase in GDP has always had a better level of life expectancy than compared to others. GDP has increased significantly in European countries over the years, but because life expectancy is likely to be capped after a certain age, the growth in life expectancy is less in the count, but significant if compared to the average human life.
- In Asia, there is an overall growth, but several countries are still falling behind. Few countries appear to be catching up to those in Europe and America.

Time Effect on Life Expectancy:

In Figure 6, it can be seen that there is not much change till the year 1972 when the maximum life expectancy exceeded the threshold of 70. Over time, Life Expectancy increases but stagnates after reaching an upper threshold which can be explained due to other factors.

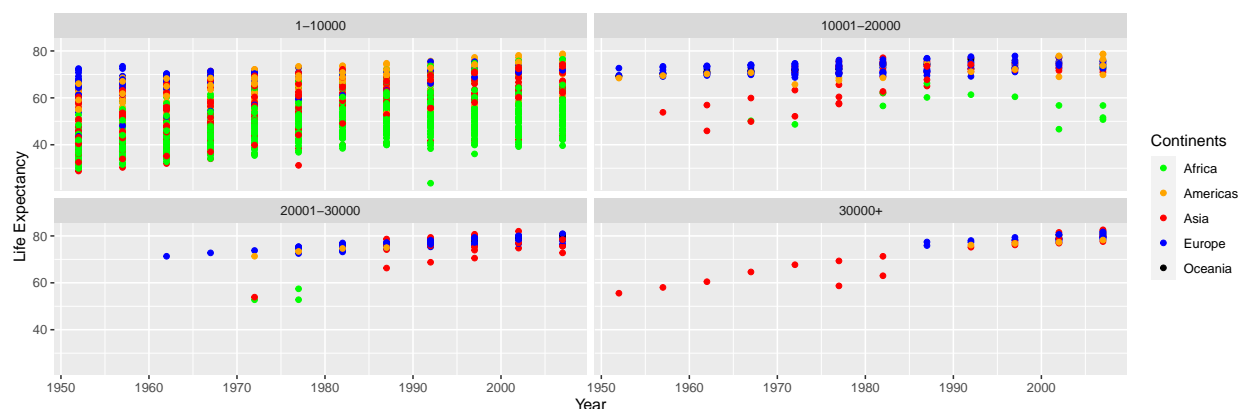


Figure 7: Life Expectancy vs Year of the different Continents across GDP ranges

Dependency of Life Expectancy on GDP and Continent

- The GDP has been divided into four ranges in the graph above, and it can be seen that as the GDP range increases, the variation of the points diminishes, indicating that the impact of GDP on Life Expectancy could be negligible or has converged but cannot be concluded as the number of data points in the higher ranges are low.
- The same cannot be said for the Continent, as statistics reveal that Life Expectancy differs by continent. For example, although the average Life Expectancy of the two continents is similar, the difference is still considerable due to other factors, indicating that the Continent does matter when forecasting Life Expectancy.

Summary:

We can conclude from the foregoing study that improvements in GDP per capita cannot account for the majority of the rise in life expectancy since World War II. In 2007, we saw that different continents had different patterns and that a simple linear model could not generalize the pattern across all continents. Moreover, it can be noted that the GDP and Life Expectancy follow a linear relationship from the age of around 70. Furthermore, the evidence is consistent with the overall patterns reported throughout continents. Africa, on the other hand, has not developed in terms of GDP and Life Expectancy when compared to the other continents, and its growth is variable. Furthermore, while Oceania follows the general pattern, the data points are insufficient to draw any conclusions.