

# Mini Project

*Srinithish Kandagadla and Vishal Singh*

*October 3, 2019*

## Abstract

In this project we study the change in two of the main indicators of development level of a country i.e. Life Expectancy and Gross Domestic Product (GDP) per capital for years after the Second World War. The two metrics are defined as follows- 1. **Life Expectancy** - It refers to the number of years a person is expected to live. Life expectancy is based on an estimate of the average age that members of a particular population group will be when they die. 2. **Gross Domestic Product** - It is the total monetary or market value of all the finished goods and services produced within a country's borders in a specific time period. **GDP per capita** is a measure of a country's economic output that accounts for its number of people. We measure these metrics across years and each other to see the trend and how they effect each other. Further, we study their effects across various continent to measure how have the different countries in the continents have performed and the overall performance of the continents. This helps us measure if some parts of the world are performing better than the others and if all the countries have performed similarly in a continent or if some countries are better than others. In general, we find that both the GDP per capita and Life Expectancy has increased after World War 2, this highlights a high correlation between the two metrics. We zoom in on all the continents and countries to provide a more granular view on these metrics.

## Data

We obtain the data from the gapminder website, we download the tables which have the longitudinal data for the metrics - Life Expectancy, Gross Domestic Product per capita, and Population. The table for Life Expectancy contains data for different countries for years 1800-2018. The value for each year appears as a seperate column. We used the data for post WW2 years for our project, this mean only years after 1945 were used. The table for GDP/per capita contains data for different countries for years 1800-2040. The years 1945-2018 are filtered for our analysis. This table also contains data for different years as different columns. This table contains inflation-adjusted value and can therefore be compared with each other directly. The table for population contains country population for years 1800-2100 in different columns. We filter for years 1945-2100 We also use the country-continent mapping form the gapminder data in the gapminder library. This table is joined with the various tables to aggregate the information at the continent level

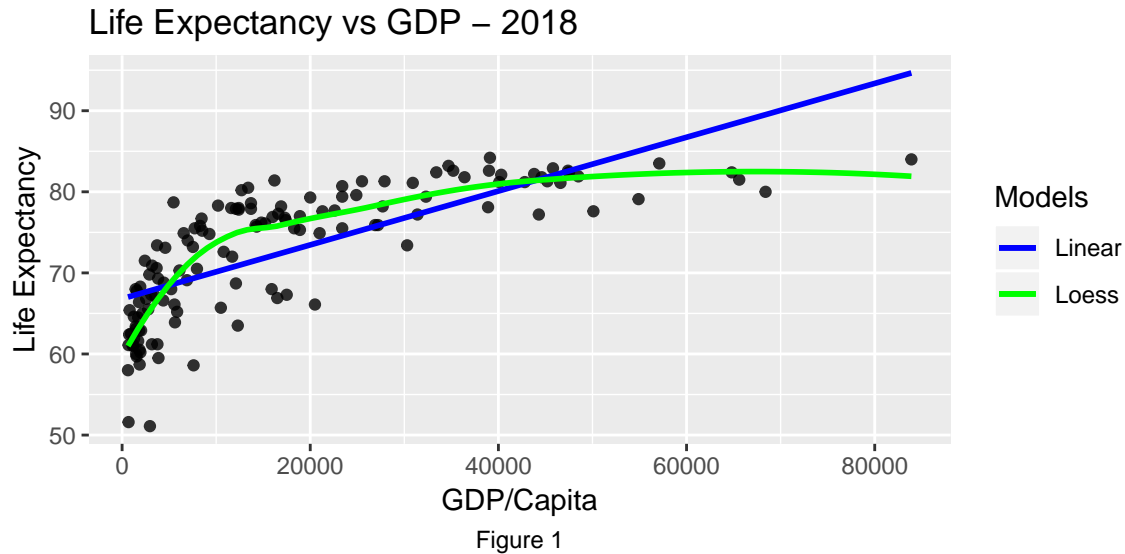
## Problem Statement

We broadly look at 3 problem statements to measure the metrics in question and explain the change w.r.t. time and each other 1. How were the GDP and Life Expectancy related for the year 2018 ? 2. How did the Average Life Expectancy change w.r.t. to time for different continents ? 3. How has the relationship between GDP per capita and Life Expectancy changed over years? We further drill down on these problem statements to look at the metrics at a more granular level for different countries.

Table 1: Table 1 - Continent statistics for year 2018

| Continent | Avg Life Expectancy | Avg GDP   | Life Expectancy Variance | GDP variance | Correlation (LE and GDP) |
|-----------|---------------------|-----------|--------------------------|--------------|--------------------------|
| Africa    | 64.97451            | 5100.392  | 31.343137                | 31233179     | 0.5491858                |
| Americas  | 76.70833            | 16757.083 | 13.581667                | 150263091    | 0.5036482                |
| Asia      | 74.45185            | 20630.000 | 33.856439                | 447932362    | 0.6773218                |
| Europe    | 80.11000            | 34923.333 | 6.344379                 | 206712195    | 0.7378562                |
| Oceania   | 82.35000            | 41100.000 | 0.605000                 | 44180000     | 1.0000000                |

## Question 1



In Figure 1 we look at Life Expectancy vs GDP per capital for the year 2018 for all the countries. We try to fit a **Linear Model** and **Loess Model** on this data. We can see that the data can be better explained using a loess model as life expectancy plateaus with increase in GDP beyond GDP > 20000. A linear model does not fit well to this data as can be seen from the R-squared value of 0.54.

**Overall Average Life Expectancy - 72.63**

**Overall Average GDP - 13307.63**

From Figure 1 we can see that for the year 2018 the relationship between GDP and life expectancy can be explained using a linear regression line for all the continents. From Table 1 we can also observe good correlation (> 0.50) between the two variables for all the continents. We observe the following pattern for the different continents-

1. In Africa we can see GDP < 15000 and life expectancy < 75 for most of the countries. The mean Life Expectancy and GDP is lowest amongst all the continents, these metrics are also lower than the overall averages. Only a few countries are above the overall average for GDP - Botswana, Gabon, Libya and Mauritius; and only 3 countries - Algeria, Egypt and Mauritius are above in term of average Life Expectancy than overall average. From table 1, we can observe least variance in the GDP for countries in Africa continent. The slope for Africa is the highest which means that effect of increase in Life Expectancy is the highest with increase in GDP.
2. For America, we can see a couple of outliers - US, Canada for GDP which have values for GDP > 40k. The mean life expectancy is well above overall average, the GDP per capita is close to the world average. We see a lot of countries in the range of 78-80+ for this continent which states that countries not just US and Canada, but others as well have a good status of health, this is also supported by the

fact that variance for Life Expectancy is quite low. We can see a strong correlation (0.50) between the variable for this continent as well.

3. We can see a lot of countries with a very high GDP in Asia, these values might have been reported wrongly as they don't match with the data online. We can see a very high value of correlation between life expectancy and GDP for Asia (0.75), this can be evident from cases like Japan where life expectancy and GDP both are high and Afghanistan where both the metrics are low.
4. We can see the highest mean values of life expectancy and GDP in the case of Europe and the value for correlation is also the highest for this continent indicating a strong relationship

The slope for the linear regression line between life expectancy and GDP is the highest for Africa indicating that Life Expectancy increase the most with increase in GDP in this continent. The pattern can be described using linear model for all the continents.

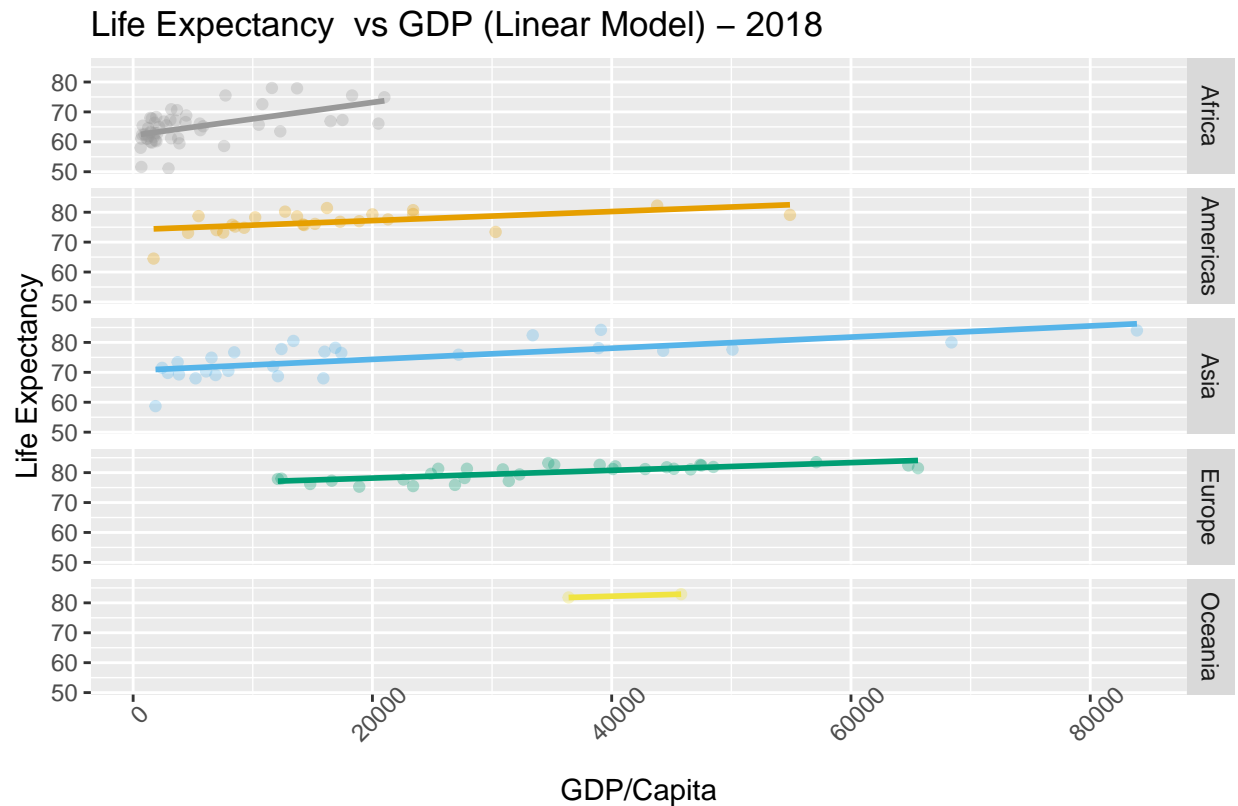


Figure 2

## Question 2

2. We can see from the Figure 3 that a linear model explains very well the increase in life expectancy for years after the second world war in 1945. This has been the case for all the continents, and we can see that all the continents have caught up in term of life expectancy except for Africa which has a slightly lower average life expectancy in the year 2018 (highest). We can see a great improvement in the average life expectancy in case of Asia (35 to 73) and Africa (35 to 66) We can observe a slight dip for Europe in years after the war, this could be because Europe was at the center of the war and could have suffered from the aftermath of the war. We can also see a slight upwards bump in case of Asia in the 1960s, this could be due to improvement of economy in the middle east due to the rise in exports of oil. There is a downwards bump for Asia prior to this period, this could be due to incorrect data as the magnitude of the bump is big. A liner model is enough to explain the increase in the life expectancy over the years. Overall we can say that the world is doing better than what it was in 1945 in terms of

life expectancy . As evident from the table, we can see there is a lot of variation in the case of Asia for both life expectancy and GDP this is due to countries having extreme values for GDP like Afghanistan and the middle eastern countries, from the plot we can see the maximum amount of deviation from the standard regression line in case of Asia.

### Average Life Expectancy (1945 – 2018)

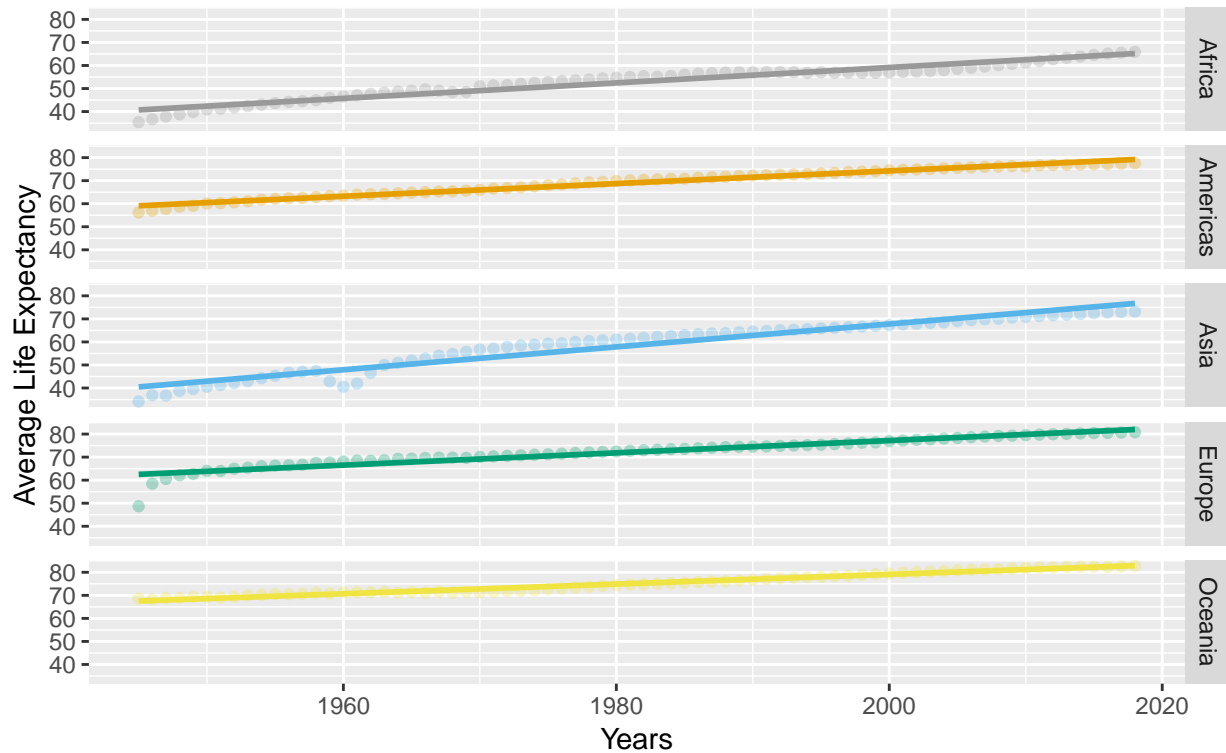


Figure 3

Since, the continents have more or less caught up in term of life expectancy we look at the data for recent years to check if its due to all the countries in the continent. In Figure 4, we have plotted the boxplots for different continents for data after year 2014, we see that that the interquartile range is the highest for Asia and Africa and lowest for Americas and Europe, this shows that the disparity is highest in continents of Asia and Africa and lowest in America and Europe. In the case of Asia, the life expectancy of some countries is among the highest e.g.- Japan, Singapore and lowest e.g.- Afghanistan

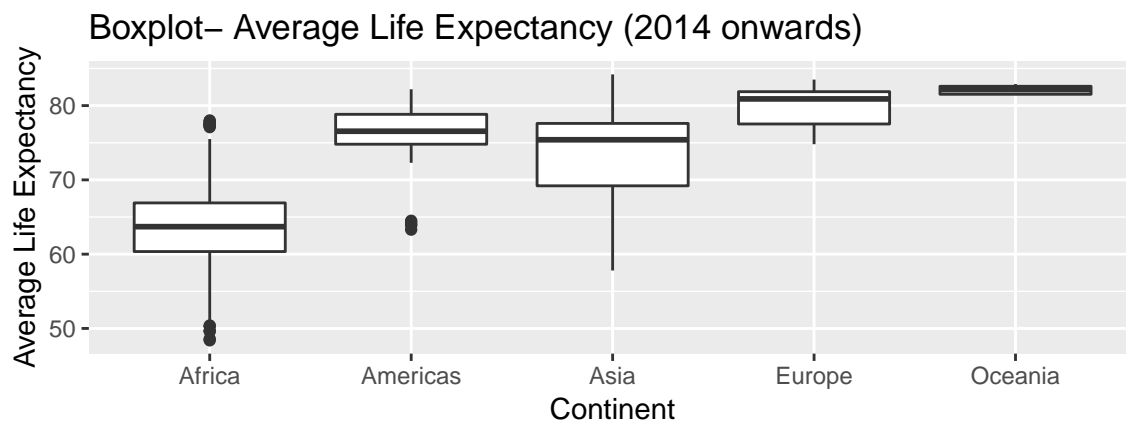


Figure 4

Next in Figure 5, we look at breakdown of slopes over different periods after WW2 for different continents

- In case of Africa, we can see the greatest slope in the period of 2003-2018, which means the life expectancy increased the most in this period
- For America, slope for the Avg Life Expectancy vs Time graph has decreased over time, this means that the improvement in life expectancy is approaching stagnation
- We saw the greatest increase in life expectancy for the case of Asia in the period of 1960 -1974, it has continued to show a high positive slope ever since
- We see an increase in life expectancy over years for Europe as well \*The number of data points for Oceania are few but we can see a positive slope for all time periods

### Linear Model of Avg Life Expectancy vs GDP over years

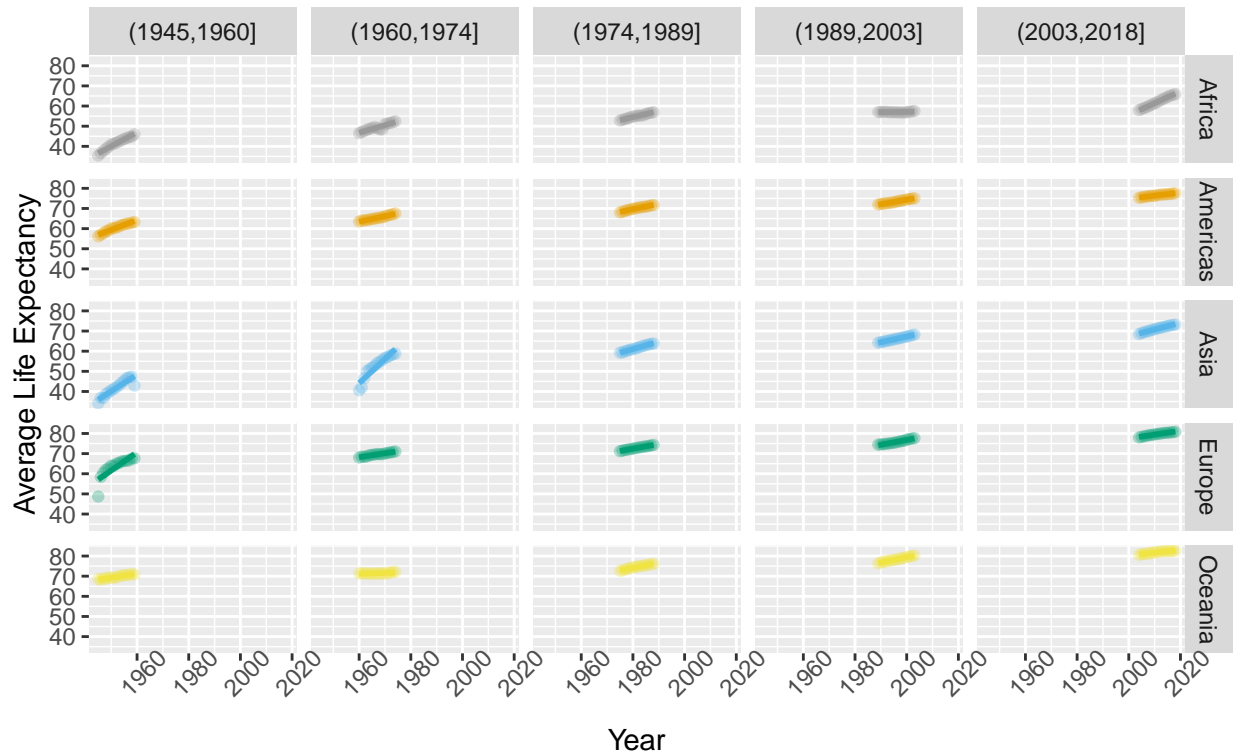


Figure 5

### Question 3

In Figure 6, we look at Weighted Average Life Expectancy vs Weighted Average GDP per capita across different continents for different years. The weighted average GDP per capita is less for Asia as the countries with largest populations (India, China) have low GDP. We can deduce the following for different continents from the figure-

- In case of Africa, increase of Average GDP over the years has resulted in increase in average Life Expectancy, there has been a stark increase in the avg life expectancy over the years
- For America, there has been a steady improvement in Life Expectancy with increase in GDP over the years
- For Asia, there is a sharp increase average life expectancy with increase in GDP in years until 1980 after that the increase in life expectancy has slowed with increase in average GDP
- In case of Europe there has been a steady increase in Life Expectancy with increase in GDP over the years until the life expectancy has reached ~80 years

- There has been a slow increase in average life expectancy for Oceania in years prior to 1970 with increase in GDP, the rate increased during years between 1970 - 1990 but has slowed since then

Overall we can say that prior to 1980s, both average GDP and average Life Expectancy increased rapidly for most countries, which meant that GDP was Life Expectancy a lot during years. The increase in Life Expectancy has slowed since then with more and more countries reaching a life expectancy of 80

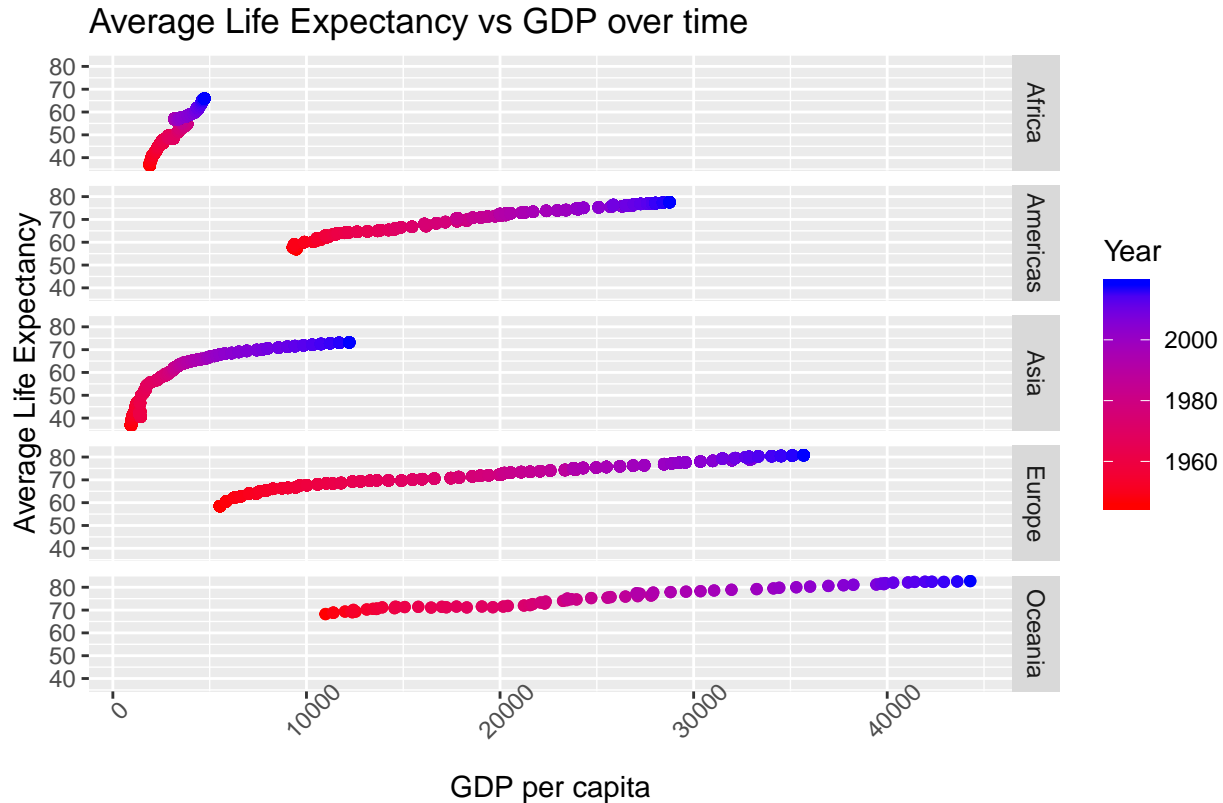


Figure 6

## References

- <https://ourworldindata.org/life-expectancy-how-is-it-calculated-and-how-should-it-be-interpreted>
- <https://www.investopedia.com/terms/g/gdp.asp>
- [https://rmarkdown.rstudio.com/authoring\\_basics.html](https://rmarkdown.rstudio.com/authoring_basics.html)
- <https://www.gapminder.org/data/>
- <https://rmarkdown.rstudio.com/lesson-7.html>

Appendix

