

STAT – S670: Exploratory Data Analysis

Mini Project 1: Life Expectancy

Group Members: Saranjeet Singh Saluja, Sricharraan Ramaswamy, Athulya Anand

Introduction:

The question to be answered through this project is can the increase in life expectancy be largely explained by increases in GDP per capita? Through a series of graphs and related inferences, we will try to identify if it is true and answer the questions.

Graphs:

Exploring 2007 data:

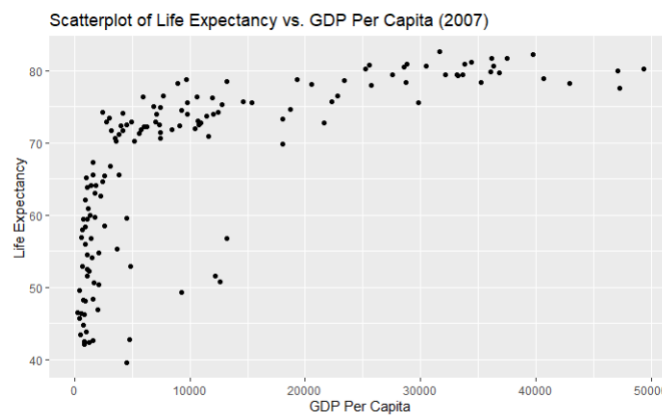


Fig 1. Relation between Life Expectancy and GDP

Q. How does life expectancy vary with GDP per capita in 2007? Can the trends be well-described by a simple model such as a linear model, or is a more complicated model required?

The above graph shows the relationship between GDP per capita and Life Expectancy (2007). There is a non-linear relationship between the variables in the form of an inverse-L shape or logarithmic. There are outliers, but the general pattern appears logarithmic. The majority of the values of Life expectancy are between 40 - 70 (some are around 85). Thus, the transformation of such a variable does not make sense. However, the GDP values range from around 300-50000. Applying a logarithmic transformation might help. Applying log transformation to GDP per cap variable:

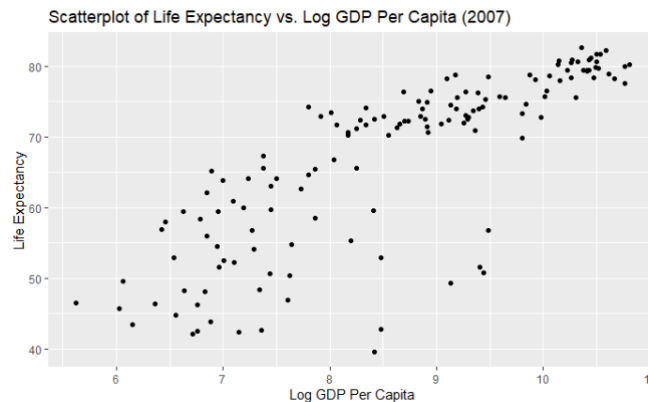


Fig 2. Relation between Log Life Expectancy and GDP

Based on the above, Life expectancy can be determined by a simple linear model. However, it most probably won't be the perfect model.

Q. Is the pattern the same or different for every continent? If some continents are different, which ones? Can differences between continents be simply described by an additive or multiplicative shift, or is it more complicated than that?

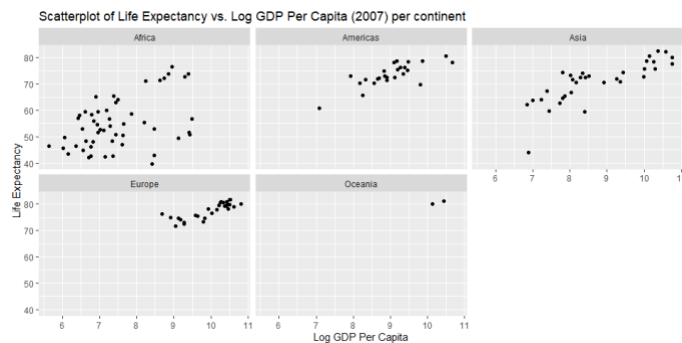


Fig 3. Relation between Life Expectancy and Log GDP for each continent

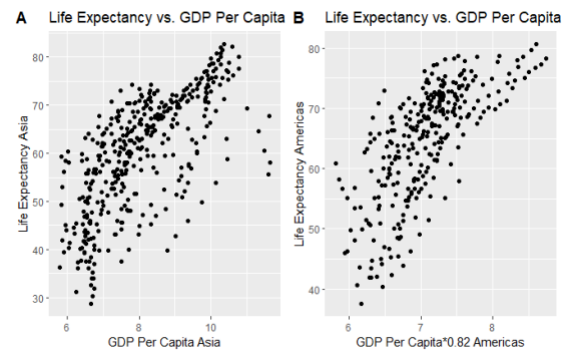


Fig 4. Deciphering similarities between Asia and America

The African continent has majorly a low GDP per capita (barring some outliers) but their Life Expectancy is mostly between 41-70. The life expectancy of the Americas, Asia, and Europe appear to have a linear relationship with GDP per capita. The trend is not the same for every continent except Asia and the Americas as they look similar. The difference in the Asian and Americas continents can be explained by a multiplicative shift.

Exploring Life Expectancy over time and continent:

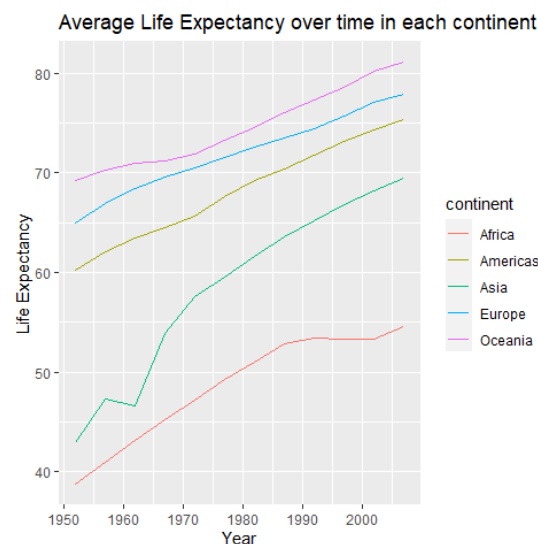


Fig 5. Relation between Average (Weighted) Life Expectancy and Year

Q. How has average life expectancy changed over time in each continent? Have some continents caught up (at least partially) to others?

Africa: The continent of Africa is following a linear trend till 1987 and after that, the growth of life expectancy has been constant till 2002 and a slight growth between 2002 and 2007.

Asia: Initially, the plot is disruptive, and it is very tough to conclude it but the growth has been exponential till 1972 and from there Asia is trying to play the catch up compared to its counterparts.

Americas: The growth of life expectancy has been linear throughout the timeline except for the year 1972 when there is a small dip in the growth.

Europe: The life expectancy of Europe has been linear throughout the timeline.

Oceania: Oceania has been in the forefront throughout the timeline and has a linear growth expect between 1962 - 1967 where the growth has been stagnant.

Q. If so, is this just because of some countries in the continent, or is it more general? Have the changes been linear, or has it been faster/slower in some periods for some continents? What might explain periods of faster/slower change?

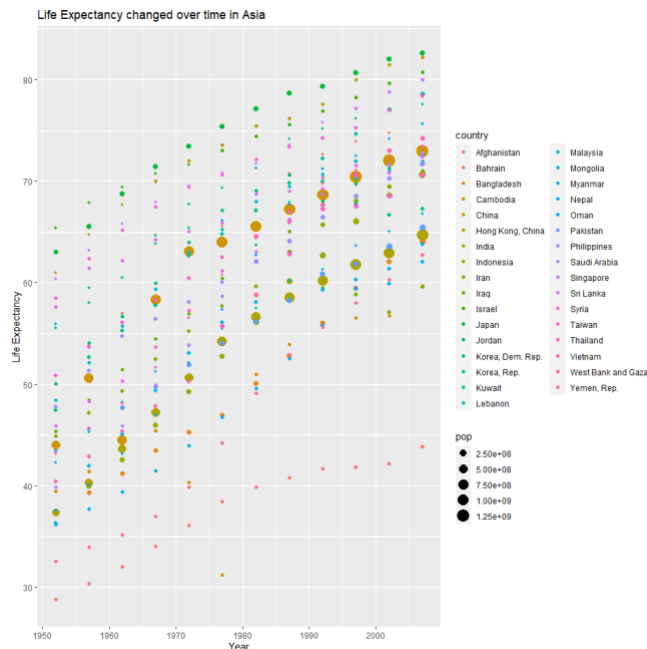


Fig 6. Deciphering which countries in Asia contributed to GDP

From the first graph, we can see that Asia is trying to catch up with other continents barring Africa. So, after taking a closer look at the countries that contribute to the average life expectancy of the continent, we can see that the life expectancy of the countries like Japan, China, and India play a major role in the growth of the life expectancy of the continent since the population of these countries are quite large and in weighted average population plays a major role in calculating the average life expectancy of the continent.

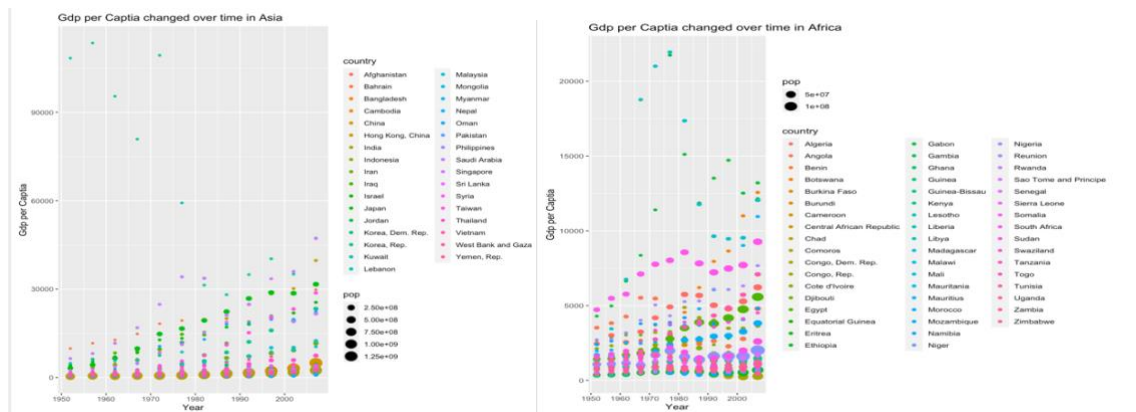


Fig 7. GDP per Capita changed over time in Asia and Africa

In Fig. 5, we can see the faster growth of life expectancy in the African continent whereas after 1987 the growth became stagnant whereas, from 1962 the continent of Asia has had a faster growth rate of life expectancy till 2007.

Factor like GDP per Capita play a major role in deciding the rate of growth of the continent which is depicted in the continent of Africa and Asia in Fig. 7

Relation between GDP and Life Expectancy throughout the years:



Fig 8. Year-wise Relationship between Life Expectancy and GDP per capita per country in the continent

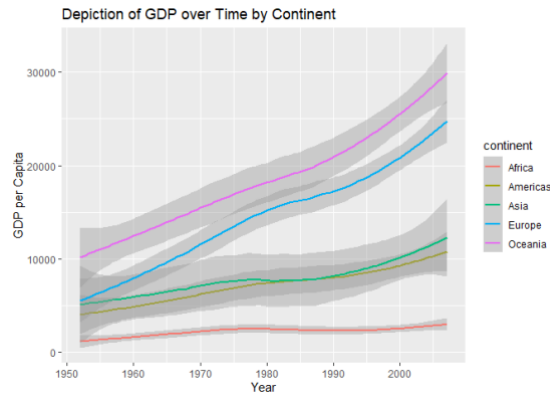


Fig 9. GDP over the years

Q. How has the relationship between GDP and life expectancy changed in each continent? Can changes in life expectancy be entirely explained by changes in GDP per capita? Does it look like there's a time effect on life expectancy in addition to a GDP effect? Has there been "convergence" in the sense that perhaps GDP and/or continent don't matter as much as they used to? Are there exceptions to the general patterns?

Fig 8., From 1952 – to 1977, it can be observed that African nations, along with many Americas and Asian nations, have a low GDP and low Life Expectancy. However, the distinction between the continents becomes clearer from 1977 onwards as the Asian and Americas nations have an increased GDP per capita and potentially a higher Life Expectancy.

In a nutshell, we can conclude the relationship between the GDP per capita and Life Expectancy is very different in each continent. However, the data points depicting Americas and Asia nations fall on approximately the same positions (especially in 2007), thereby indicating a similar linear relationship as explained by the multiplicative shift earlier. Thus, we can infer from the plot that changes in life expectancy cannot be entirely explained by changes in GDP per capita. This is solely because there are multiple factors like the population size, per capita income, longevity, etc which affect this relationship. Noticeably, Asia has the largest population size as compared to other continents which could be one of the factors affecting its life Expectancy.

Fig 9., The above graph portrays the GDP over Time by continent. We can infer that Africa has a constant curve with a very slight drift between 1970 and 1980, following which the curve is constant. Again, we can notice a slight dip around 1995. The curve of Americas and Asia is intersecting in 1985 and both follow an approximate similar curve. However, we cannot infer if it is linear or continuous. The curve depicting Europe is linear and continuous, it follows a smooth curve with a very slight drift in 1980. Oceania follows a slight exponential curve.

In addition to this, we can also confirm that there has been a convergence between all the continents except for Africa. This is because, as you can see all other continents meet at one point between 1950 – 1960, whereas the African continent follows a constant trend. Thus, the growth in GDP per capita is independent of the growth in each of the continents.

In conclusion, this graph tells us that over time the GDP per capita has increased for almost all continents and that time has a major role to play in altering the Life Expectancy and the GDP Per Capita of each continent.

Modeling:

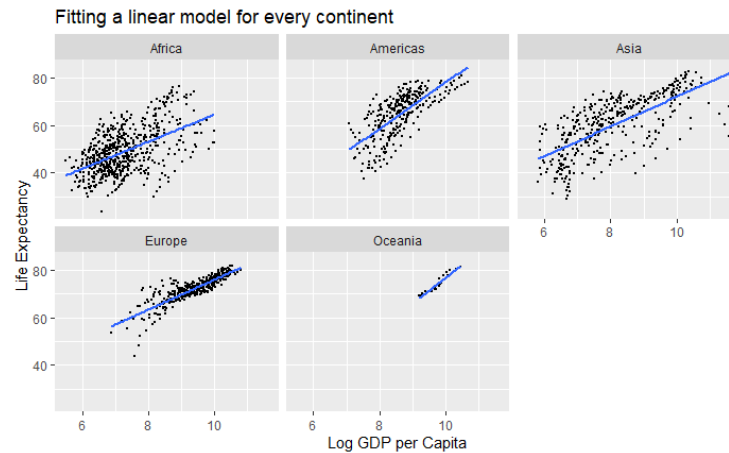


Fig 10. Fitting a simple linear regression model for every continent

On simply fitting a simple linear regression model, we can see that for the continents Americas, Europe and Oceania, the model fits quite well with the single predictor Log GDP per Capita. However, it does not fit as well with Asia and Africa, and a more complicated model such as LOESS might be required if the error tolerance is to be kept too low.

Interpretation of the model:

Consider fitting the simple linear regression model for Europe:

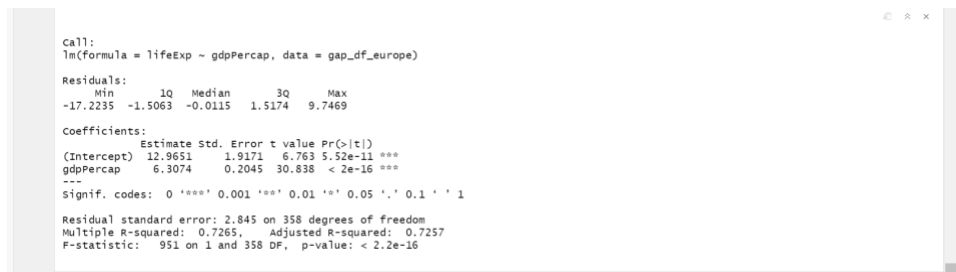


Fig 11. Summary of the simple linear regression model (Europe)

As we can observe, the coefficient is 6.3074, which will be divided by 100 [1]. We get the value of the new coefficient as 0.063074. Now this implies that, for every 1% increase in the independent variable, our dependent variable increases by about 0.06.

Similarly, if we carry out fitting a simple linear regression model for every continent, the GDP per cap variable is statistically significant for every continent.

Conclusion:

As per our analysis, GDP per capita plays an important role in determining Life Expectancy irrespective of the time. GDP per capita is an important predictor variable for Life Expectancy. Even though it can be argued that GDP per capita does not entirely define Life Expectancy, it can be said that majorly does affect Life Expectancy.

Reference/s:

[1] <https://data.library.virginia.edu/interpreting-log-transformations-in-a-linear-model/>