

Project Report

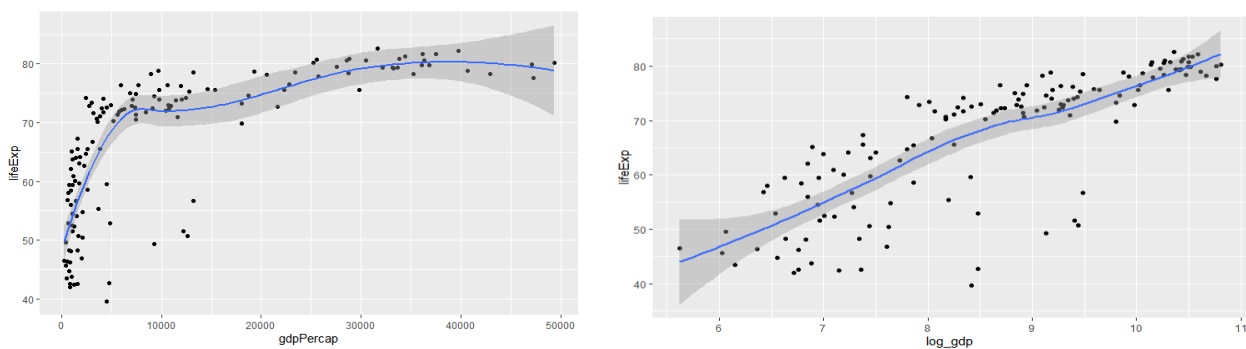
Introduction:

The dataset that is used in this project is from gap minder. It basically explains the life expectancy of various countries on different continents over the years along with population and GDP per capita. Here we are trying to understand whether life expectancy is largely dependent on GDP data or whether some other factors also influence it. Basically, we are trying to understand the factors on which the life expectancy of the countries is majorly dependent. As the data has the year, continents, and GDP as the major factors that could influence life expectancy, we are trying to understand the individual factors that influence the response variable. So, first, we are trying to keep the year constant and then analyze the relationship between life expectancy with GDP and also compare that with different continents. Then to analyze the effect of the year alone, we are trying to get insights into the relationship between life expectancy in different years and also compare that with different continents. Finally, we are having all the variables and try to fit the model that explains the relationship best between the GDP, Year with life expectancy.

First Level Analysis:

With year constant, understand the relationship between GDP and life expectancy and compare that with different continents. As suggested, we have tested the above scenario for the 2007 year, and the results are described below,

How does life expectancy vary with GDP per capita in 2007? Life expectancy has a logarithmic growth with respect to the GDP per capita. This can be seen by plotting a scatter plot between life expectancy and GDP per capita.



The logarithmic growth can be verified by taking the log value of gdpPercap and plotting that to verify if we get a straight line. And yes, we got a nearly linear relationship with loess.

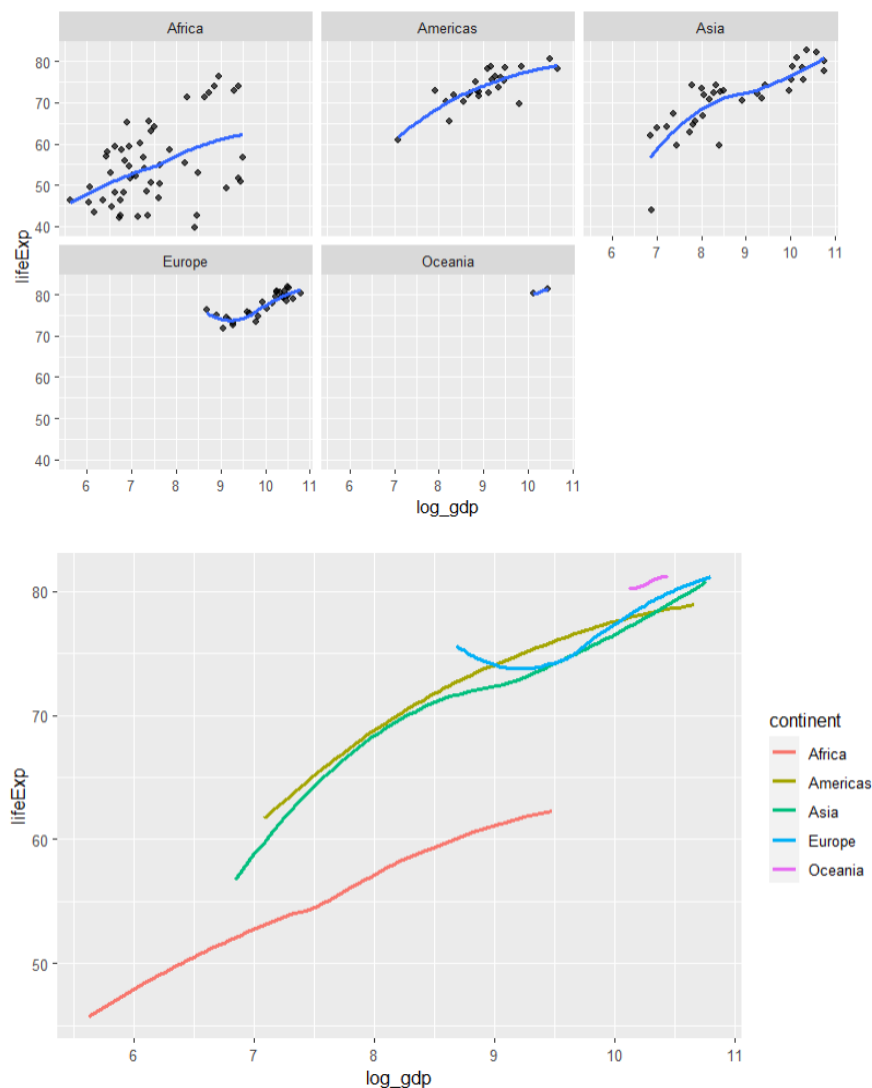
LOESS:

In LOESS, a weighted regression is performed on a small subset of data points within a window centered around each data point. The weights are based on the distance of the data point from the center of the window, with closer points receiving higher weights. This means that points that are closer to the center of the window have a greater influence on the fitted curve than points further away. So, the span is a parameter that determines the window size.

Can the trends be well-described by a simple model such as a linear model, or is a more complicated model required? If we do a simple transformation of taking the log of GDP and then

plotting that with life expectancy could be largely represented as a nearly linear model. So, yes we could represent via linear model if we could perform transformations on GDP per capita.

Is the pattern the same or different for every continent? The pattern looks largely similar, with some differences between the models of different continents. Asia and America represent a largely similar model with Europe having a curvature model, but all three have a very similar range of data points in the plot.



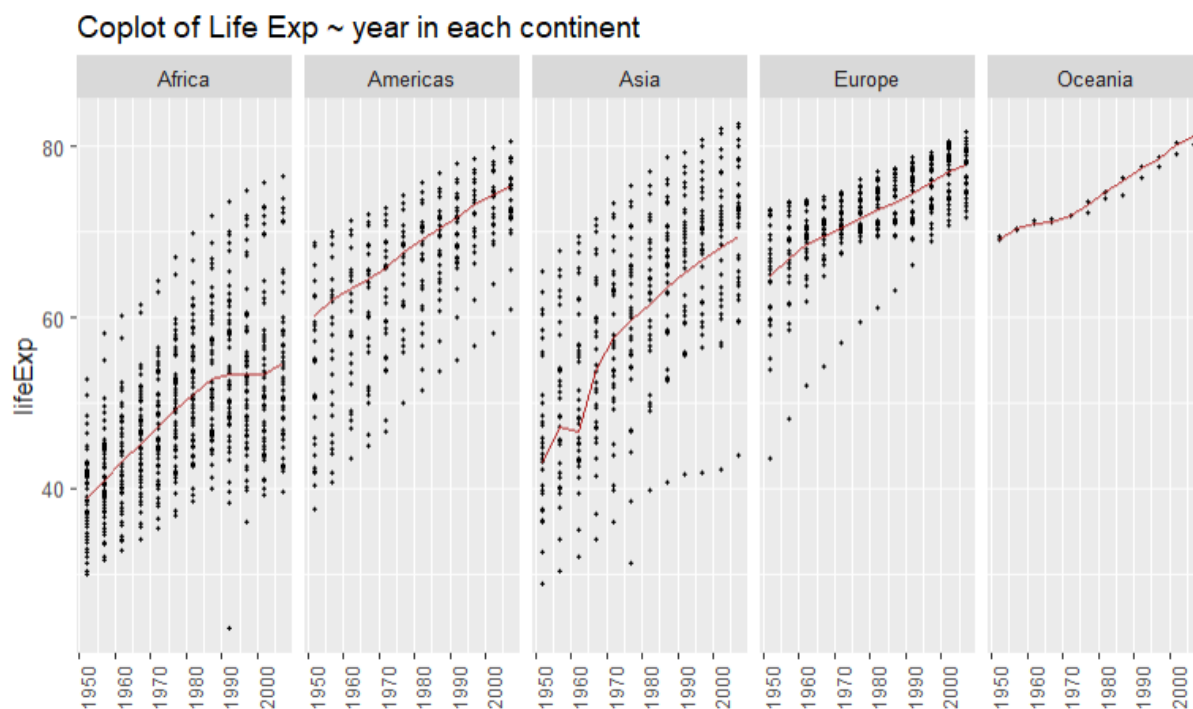
Here we have displayed the relationship between Log (GDP) and life expectancy in 2007 faceted by continents. The graph on the right explains the models of different continents corresponding to the year 2007 between life expectancy and log (GDP). The models are made based on loess function for individual continents with span 1 and degree 2.

If some continents are different, which ones? Africa and Europe are the ones, which are quite different. The average life expectancy was already high in Europe and it was at the same value for quite a period of years before it starts increasing. Whereas in Africa, we see an increase in life expectancy with an increase in GDP per capita, but Africa started with a very low value of life expectancy in the initial years, so the model though similar to America the position of the model is different. We can't able to confirm the model for Oceania, as there is very minimum data available. The Asia and Americas represent largely similar models with a slight difference in the bend in the Asia model.

Can differences between continents be simply described by an additive or multiplicative shift, or is it more complicated than that? The additive shift can make the African model similar to the Americas and Asia. But with Europe since the model itself is different, this process will not make the model equivalent to the other. As earlier said, the Oceania model is not verifiable so we can't able comment on that.

Second Level Analysis:

Here we plotted weighted average life expectancy with years for different continents. The weights correspond to the population of the respective countries.



Africa saw a decrease after 1990 because of suddenly reduced life expectancy in some countries but increased after 2000. America saw a general increasing trend but Asia managed to catch up with it even with a slight dip after 1960, but the overall trend was still general. While there were countries in Europe with less life expectancy than other countries, they still managed to catch up with Oceania.

How has average life expectancy changed over time in each continent? The overall life expectancy has observed an increase in each continent. Since 1952, Oceania has seen an increase of a little more than 10 units; while Africa, Europe, and America observed a change of 15 units; Asia saw an increase of 25 units till 2007.

Have some continents caught up (at least partially) to others? While Africa started with 40, they increased to 55, other continents like Asia started with 45 and are now at par with America (75 in 2007) with just 5 units short, and Europe started with 65 and has now caught up to Oceania till almost 80. So, after years passed, now Americas, Asia, and Europe have similar life expectancies. Oceania also has a similar but slightly higher life expectancy than the others. In Africa, it seems some countries have similar life the others but most countries have yet to catch up to others.

If so, is this just because of some countries in the continent, or is it more general? In the case of the Americas, the variance between countries was initially large but later it became smaller explaining that towards the end the trend is explained overall and not just by some countries. But in Africa, the variance of life expectancy increased over time explaining that few countries in the

continent have shown a better rate of improvement in life expectancy than the rate of increase for other countries. So, in Europe and the Americas, the trend is explained by most of the countries, but in Africa, since the variance is increased, the trend can't be explained overall. In Asia, there is a slight decrease in the variance of life expectancy between countries, so the trend could be attributed to the continent.

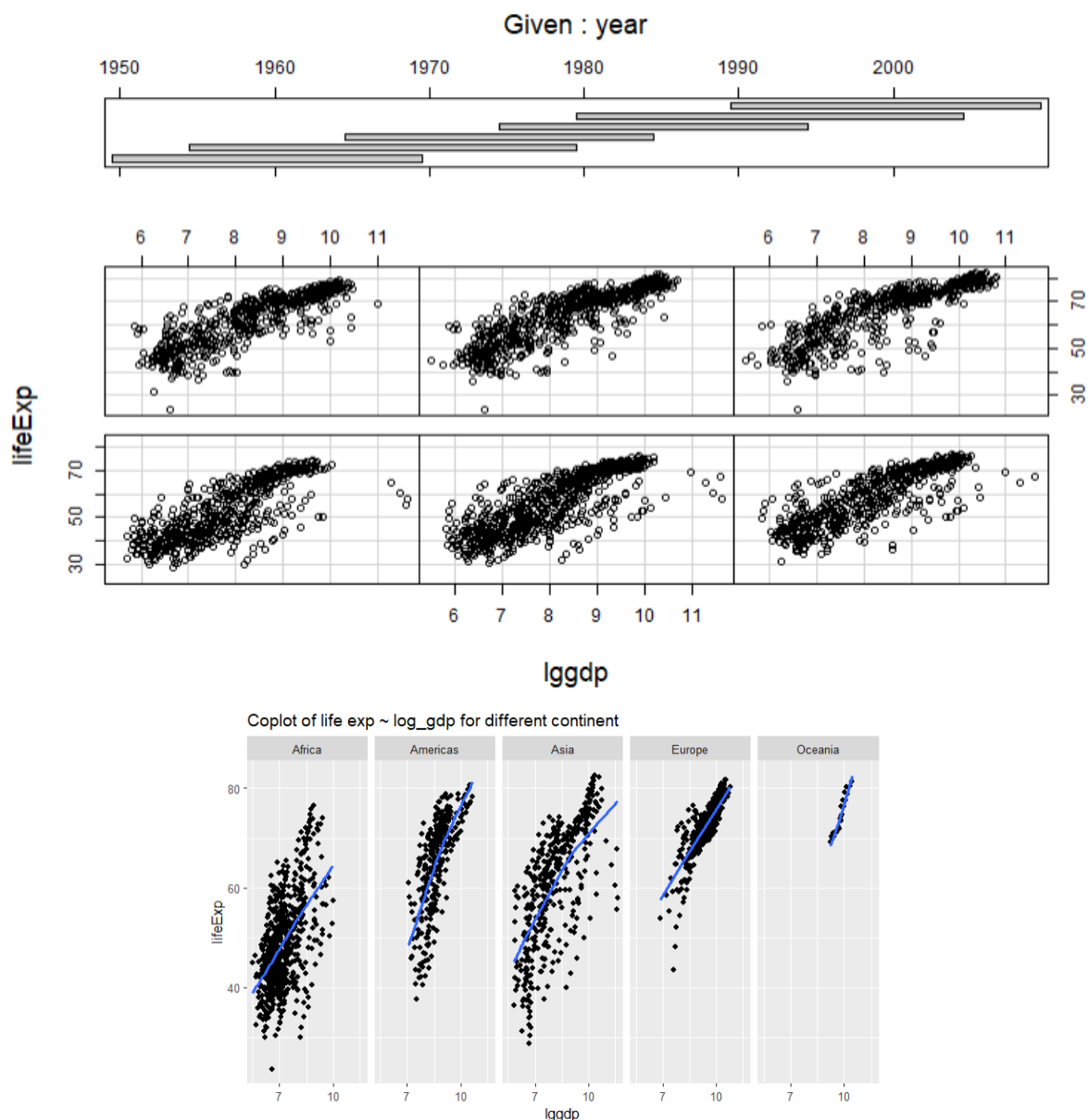
Have the changes been linear, or has it been faster/slower in some periods for some continents?

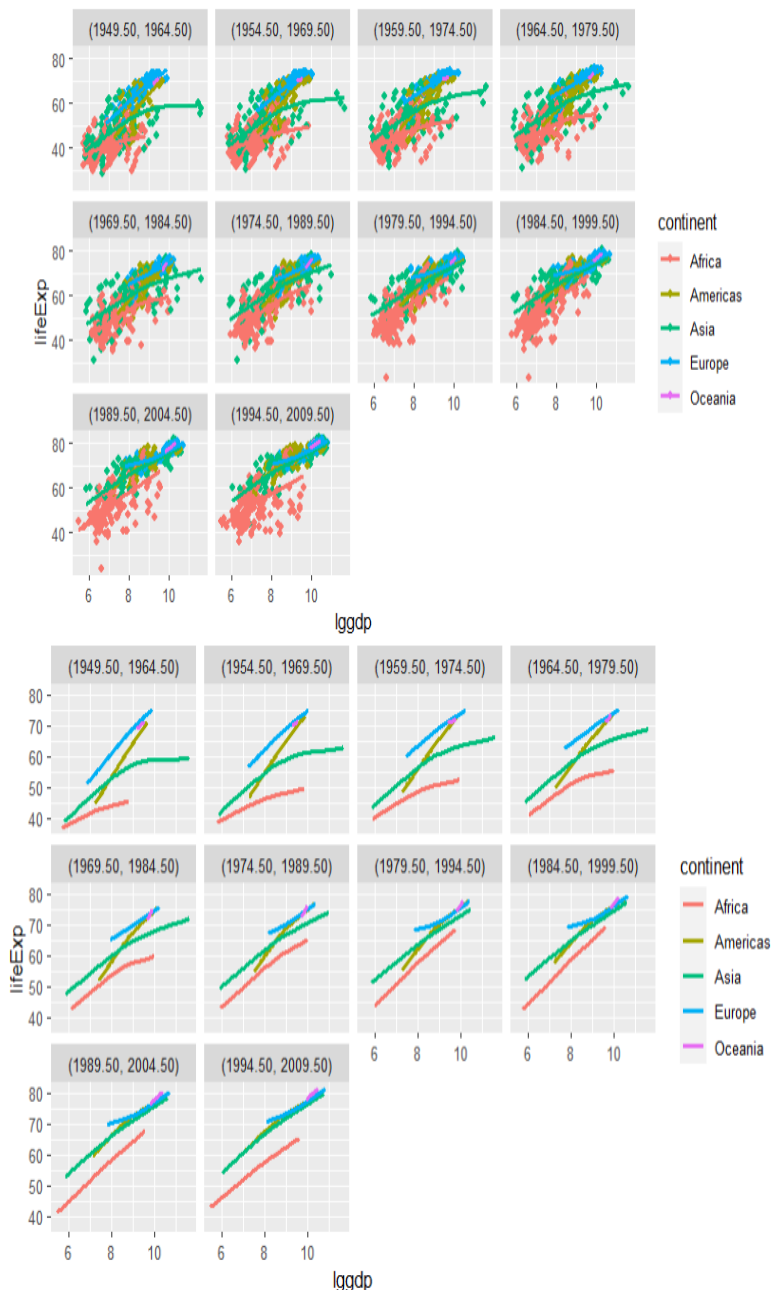
What might explain periods of faster/slower change? While America, Europe, and Oceania has seen a linear increase with almost the same slopes, Oceania saw a slow increase in the period of 1960-1970, but still maintained the highest spot than other continents probably because of the presence of both countries being developed. America on the other hand saw a generic trend with some countries

Final Level Analysis:

Changes in the relationship between GDP and life expectancy over time:

We started with coplot by which we tried to understand the relationship between GDP and life expectancy at various ranges of time. Overall, the plot gave a similar trend in which it is increasing.





Here we are trying to understand the relationship between life expectancy with GDP for different continents at various times. The model that we used is LOESS, with span 1 and degree 1 as this fits the data more appropriately than the other span and degree we tried.

How has the relationship between GDP and life expectancy changed in each continent? First, we can't able comment on the Oceania model as we have very fewer data. We could see a noticeable difference in the models of different continents at different times. In the case of the Americas, the model starts shifting upwards and the range initially was between 45-70 and it got changed slowly and ended up between 60-80 life expectancy. Also, the relationship was very linear but as time passes it became slightly curved. In the Asia case, the curved relationship started to become a linear relation and in Europe initially, it was linear but toward the end, it starts to flatten, though the relation is not exactly a flattened relationship, it became flattened compared to the initial one. In Africa, the relationship changed from curvature to a linear one, but the range of the model of this continent is always below the range of other continent models.

Can changes in life expectancy be entirely explained by changes in GDP per capita? As explained above, just GDP can't explain the change in Life expectancy as you can see in the above plots which clearly explain that life expectancy changes with time in addition to the GDP parameter.

Does it look like there's a time effect on life expectancy in addition to a GDP effect? Yes Absolutely. The time effect can be very well seen if you compare the life exp vs GDP relationship until 1964 with the 1994-2009 range data.

Has there been "convergence" in the sense that perhaps GDP and/or continent don't matter as much as it used to? Are there exceptions to the general patterns? In the initial years, we don't see any convergence. But as years passed the models of Europe, the Americas, and Asia starts to have similar ranges. The Americas and Asia models seem to overlap in the final years and Europe also seems to have a similar relationship but the range of the Europe model is less and it is within the range of the other two models. So, we could see the convergence of Europe, Asia, Americas during the final years, the African model also has a similar relationship but the range of that model is below the other one. So, during the final years, the GDP and continent effect on life expectancy is minimal if you compare Asia, Europe, and the Americas. But with Africa, we could still see a GDP effect on life expectancy.

Conclusion:

Based on the above three-level analysis, let's try to explain whether the increase in life expectancy since World War 2 be largely explained by increases in GDP per capita. We can't conclude that based on just GDP alone. Time is the main factor that determines the answer to this problem. If we take the final years where the relationship is largely linear, we could say Yes to this question. But if you take the initial years, the Asia and Africa model would say that though the GDP is increasing, we could not see a considerable increase in the life expectancy as the model for this continent starts to flatten out after some GDP per capita value. So, the Time factor needs to be considered to analyze the larger issue.

Team Members:

We are a group of three and we have done the mini-project as a team.

Rakesh Santha Kumaran (rsantha@iu.edu)

Aakriti Sachdeva (aasach@iu.edu)

Shricharan Baskaran (shrib@iu.edu)