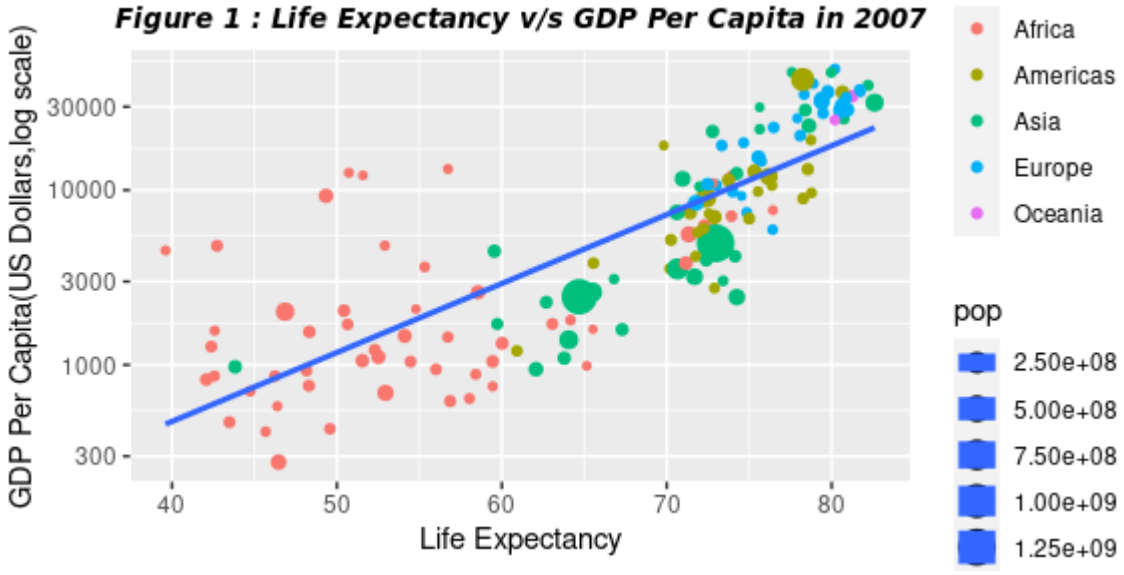
EDA-Mini Project

Introduction

Using the gapminder data set, which is accessible in the R package of the same name, we investigate the link between per capita GDP and life expectancy between 1952 and 2007. We discover that both wealth and time have significant apparent effects on life expectancy, while the details differ significantly between continents.

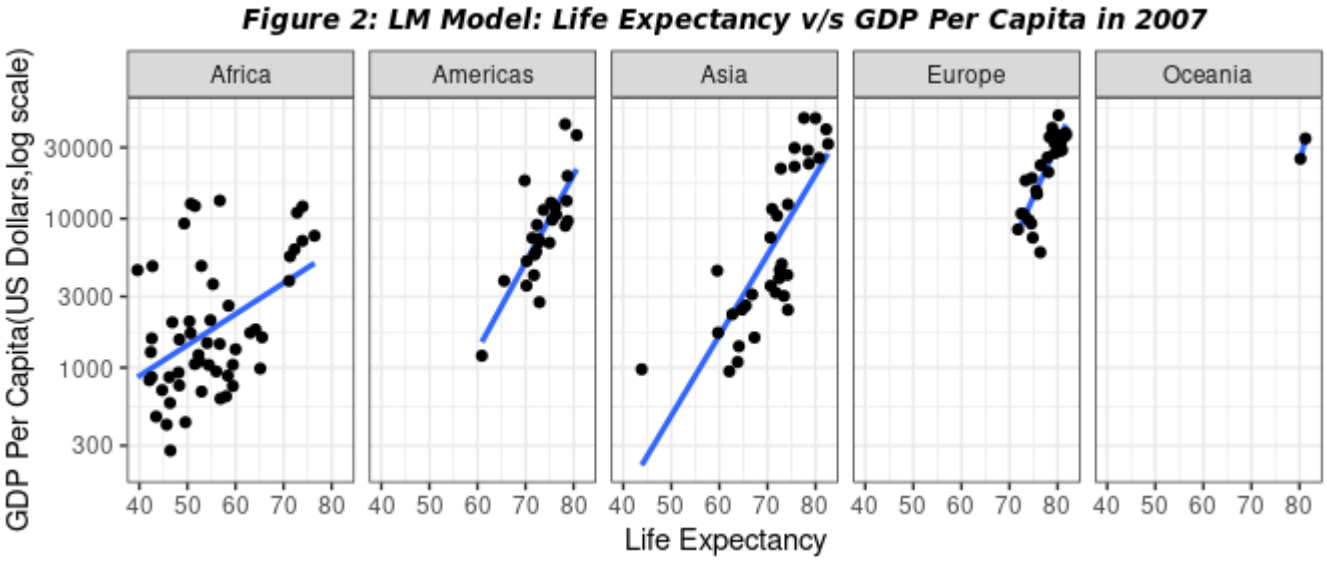
GDP and life expectancy in 2007

We begin our analysis by examining the relationship between GDP per capita and the life expectancy in the year 2007, the most recent year. This will help us understand if there is any trend now and then we can compare it to historical data available to us.



- In Figure 1, bigger points correspond to bigger country population. The distribution of GDP per capita among countries is strongly right-skewed, that is why we have taken the base 10 lof of the variable. We have plotted the graph between 'GDP Per Capita' and 'Life Expectancy'.
- We can see that for the countries having GDP per capita below 5000 have life expectancies less than 70 disregarding few outliers. On the other hand, there is very little impact of GDP rise in life expectancy for countries having GDP per capita of more than 10000.
- As there are many countries in different continents and each country has varying population, there is huge variation seen in the **Life Expectancies** and **GDP Per Capita**.
- We may be able to interpret some relationship between Life Expectancies and GDP Per Capita if we consider it individually for all the continents.
- The linear model fails to capture the trend between GDP and life expectancies if we consider the entire dataset together. A more complex model is needed. But for that let's consider the continent wise modeling.

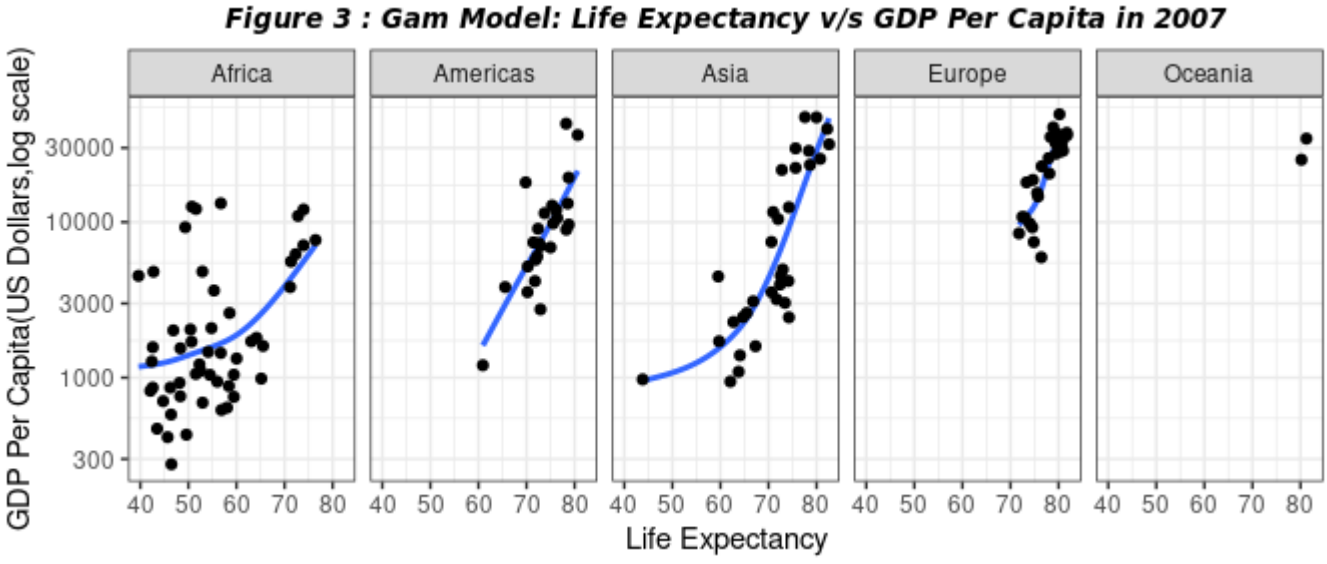
Relationship between GDP and life expectancy for every continent in 2007



In Figure 2, we have tried to encapture the relationship between **Life Expectancies** and **GDP Per Capita** using LM Model, in addition to segregating different continents. We can see that for continents Americas and Europe, linear fit captures the trend between GDP and Life Expectancy approximately, but for the continent Africa a more complicated model is required. If we remove a few outlier countries from Asia, we believe that an LM model will fit the relation in a better way. We are not considering Oceania for our analysis as it contains only two countries and thus we don't have enough data.

Fitting Generalized Additive Model (GAM)

Let us try another complex model and see if it captures the relationship between **Life Expectancies** and **GDP Per Capita** in a better way. We will try using GAM or GLM which will help us show the additive or multiplicative shift. Generalized Additive Models (GAMs) are an adaptation that allows us to model non-linear data while maintaining explainability. A GAM is a linear model with a key difference when compared to Generalized Linear Models such as Linear Regression. A GAM is allowed to learn non-linear features. We simply replace beta coefficients from Linear Regression with a flexible function which allows nonlinear relationships.

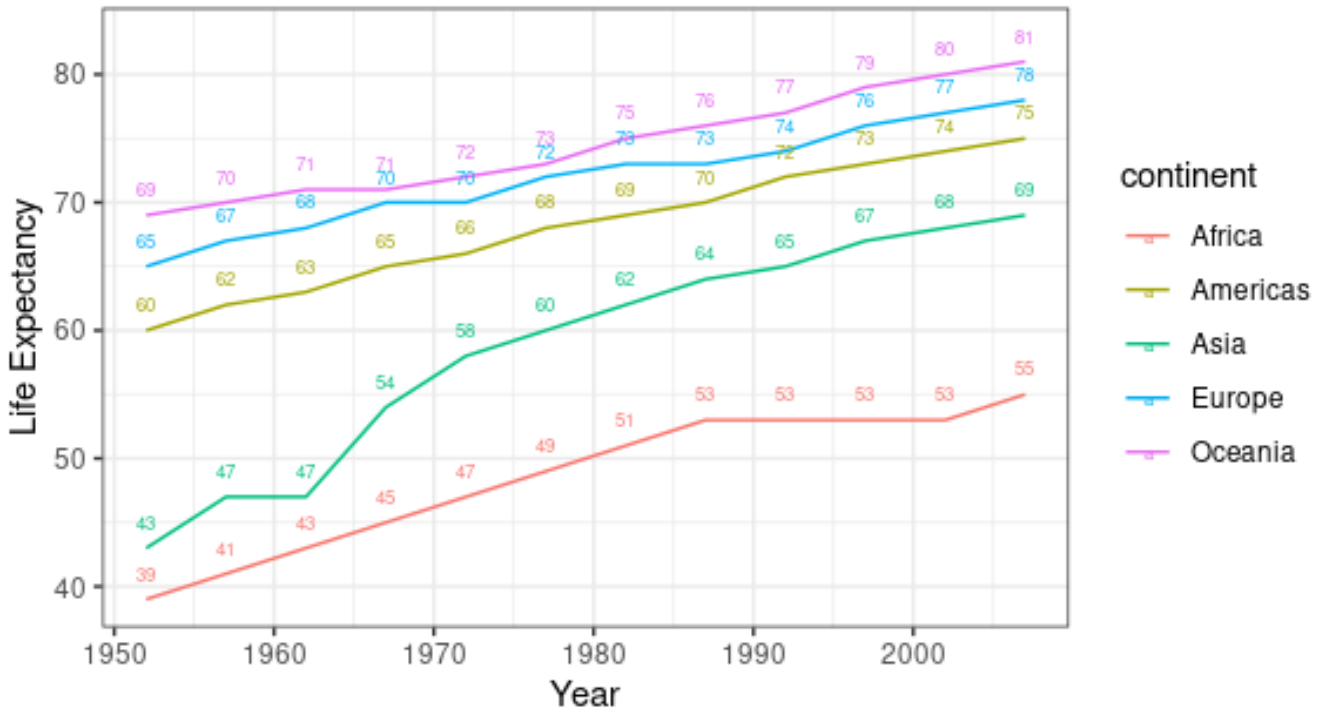


As we can see in the Figure 3, the GAM model works better than the LM Model and captures the relationship in a better way even for Asia. However, we are still unable to find any trends for the African continent. Still this model in a whole is helping us find the relationship in a much better way and that is the reason why we will be using gam model moving forward.

Life expectancy over time by continent

We will here look at average life expectancy change over time in each continent to find if there is any trend other than GDP that is effecting the change in life expectancy whether positively or negatively. For analyzing this we have created temporary dataset by using a function subsetting and grouping the data by the continents and year and taking the weighted average in mind for the population, thereby finding the average life Expectancy for each continent with the year.

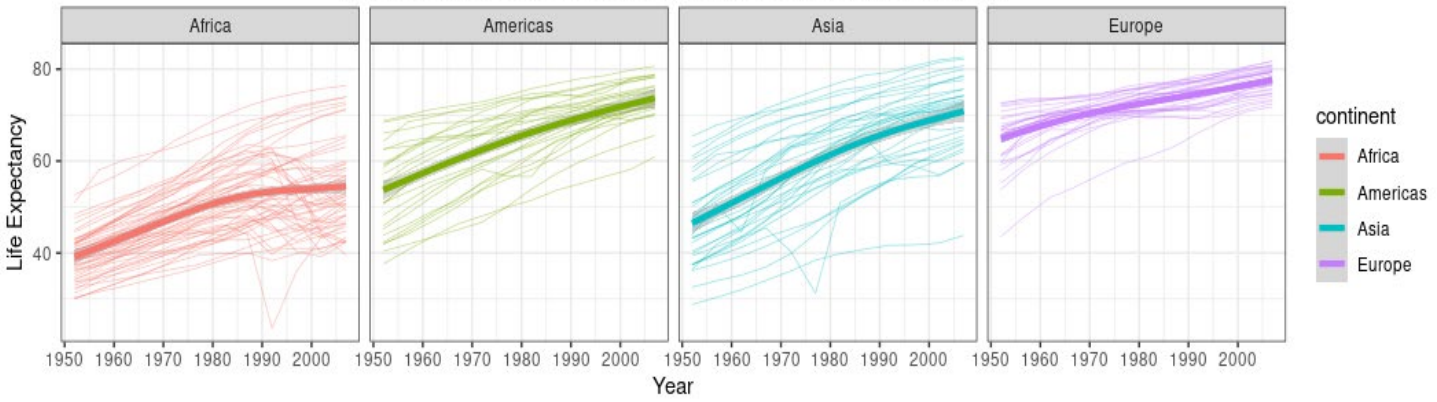
Figure 4 : Weighted Average Life Expectancy by Continent vs Year



In figure 4 we see that the average life expectancy in Oceania, Europe, and the Americas, the increase has been quite consistent, but when we look at Africa, it grew at much faster rate till 1987 then there is not much growth in the Life expectancy for the African continent with respect to time. This might be attributed to AIDS virus. Although the trend is quite uneven, Asia has almost caught with the Americas. However, there was a notable decrease between 1957 and 1962. Because China and India combined have made up about two-thirds of Asia's population throughout the study period, we split these two countries off from the rest of Asia. This can be attributed to the war that had been happening between India and China during that period of time.

Let's dive deeper into this reasoning and look if this change is because of some countries or is it more general. We will look at the pace of the change also to support our statement above.

Figure 5: Plot between Life Expectancy of each country and Year



- As already mentioned in Figure 4 that there is a shift in rate of increase of Average Life expectancy for African continent after 1987, and before 1987 there was a group effect of all the countries in pulling up the average life expectancy. Average Life expectancy didn't grow much after 1987 in Africa and we can attribute this to a group of countries like Rwanda as can be seen in the plot above that they are holding down the increase in average life expectancy. This might because of the civil war.
- For America and Asia, there is a linear trend in all the countries together except few outliers like China and India. All the countries together are contributing towards the change in Europe, since the spread is not much and life expectancy between countries are tightly closed over the time.

We can look into GDP per capita and the total population of the country and this can explain these periods of faster/ slower change. With the data we are given there two are the only possibilities we can look into. Let's move forward with relationship between GDP and life expectancy overtime.

Changes in the relationship between GDP and life expectancy overtime:

Figure 6.a: GDP v/s Life Expectancy

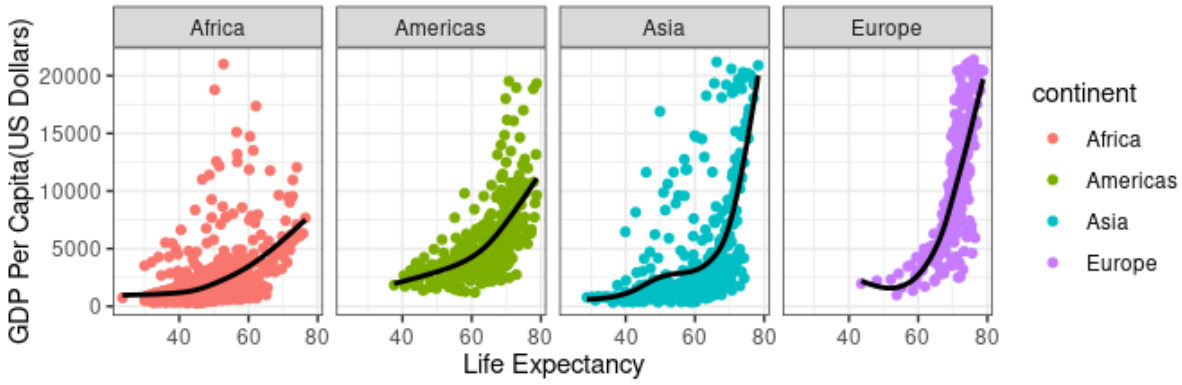
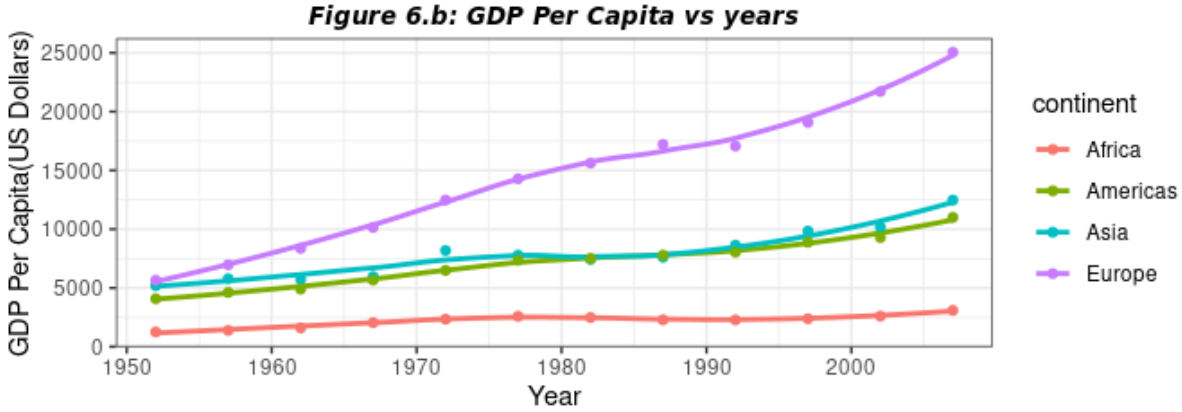


Figure 6.b: GDP Per Capita vs years



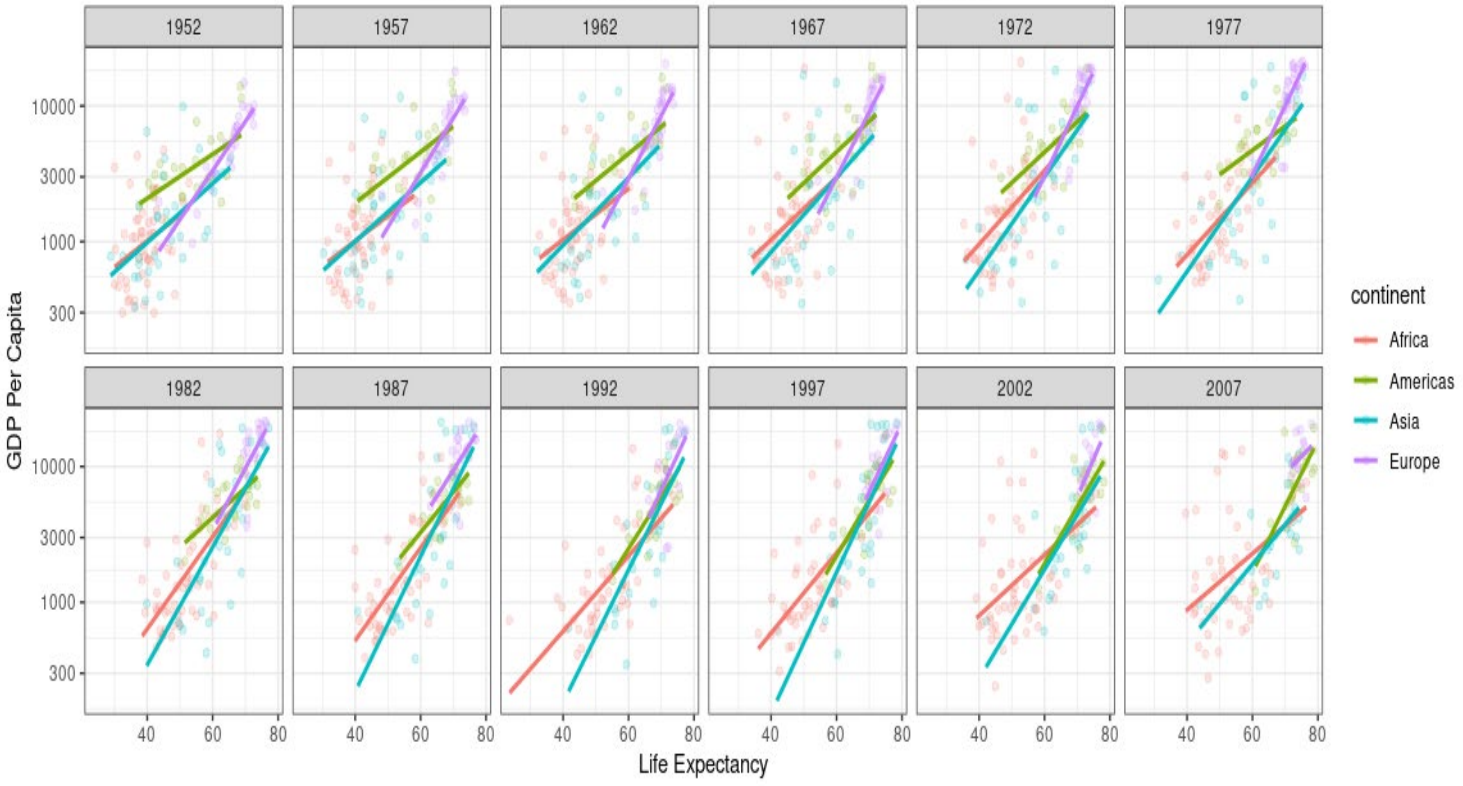
Here we have created temporary dataset by using a function subseting and grouping the data by the continents and year and taking the weighted average in mind for the population, thereby finding the average GDP per capita for each continent with the year. In Figure 6, we are plotting the effect of Average Life Expectancy with respect to the change in GDP after removing a few countries from each continent that are outliers. We have also eliminated Oceania from this analysis.

Here, we can see that there is some trend between life expectancy and GDP for countries when separated by continent. We need to look at the increase of GDP with time to understand whether there is a in a more robust manner.

As we can see from three plots above there is not much increase in GDP for Africa but still there was some increase in Life Expectancy for Africa. For Americas and Asia there is a trend upto a certain limit of GDP, till there the average life expectancy increases with increase in GDP, but after a threshold increase in GDP does not much effect in the increase in the avg. life expectancy as we can say that there are some other factors maybe temperature and kind of living environment.

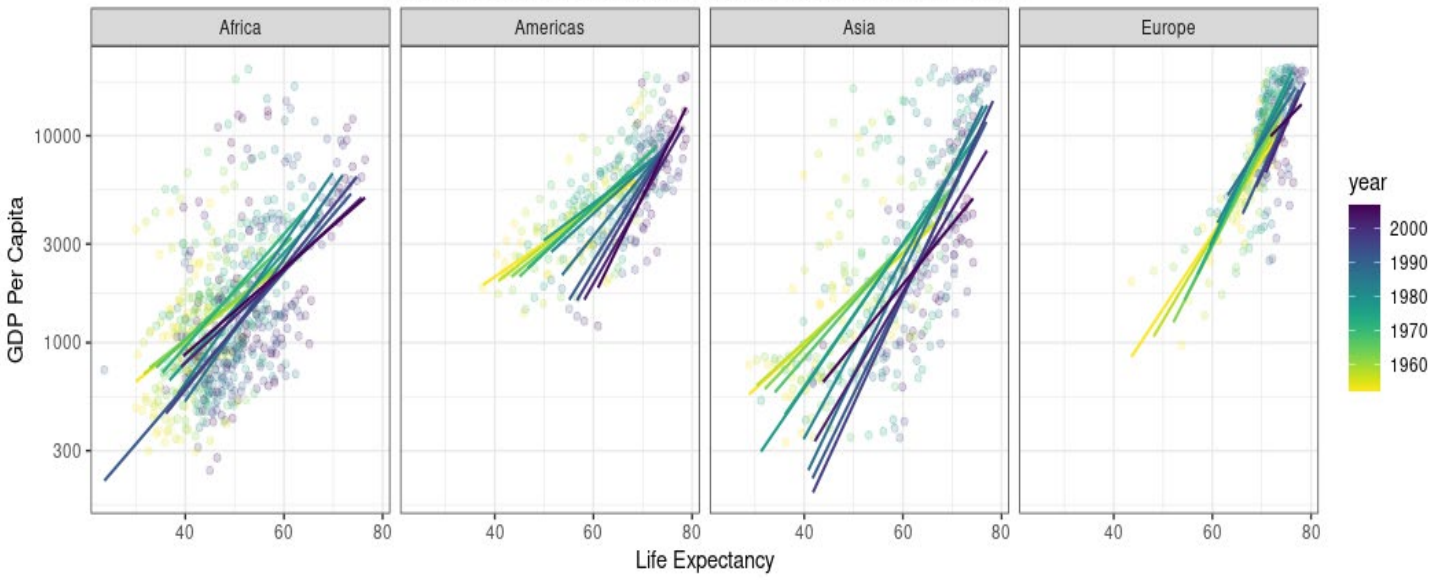
Yes, we do see a time effect on the life expectancy for different continents. On an average, the life expectancy has increased globally with time in almost the same manner.

Figure 7: Life Expectancy v/s GDP Per Capita for each year



In Figure 7, if we compare the years 1952 and 2007, we see that for Europe and Americas, the increase in Life Expectancy with respect to GDPs has been considerably high for countries with lower GDPs. If we see Asia, we can conclude that the slope of GDP has been relatively constant. While the slope in Africa was quite steep in the initial years, but it started getting flatten after a while. Finally, we can say that from the faceted plots by time, we see that over time the average life expectancy of the world has increased despite the gdp or continent but at certain age it get's flattened even though the gdp is increasing, so we can say that perhaps GDP and/or continent don't matter as much.

Figure 8: Life Expectancy v/s GDP Per Capita for each country



In Figure 8, we have excluded Oceania from the analysis as it contains only 2 countries. In general for Americas, we can see that over the time the GDPs remained more or less the same, however, we can see changes with respect to Life Expectancies. For Asia and Africa, there has been an increase in Life Expectancies as well as the GDPs over the years. In case of Europe, collective growth can be observed for large as well as small GDP countries.

Conclusion

- Life expectancy rises as per capita GDP rises. While there isn't enough information to pinpoint the exact structure of the relationships, they are fairly equivalent across all continents. The biggest regional difference is that Africa's curve is around 10 years lower than the rest of the world's.
- All continents' weighted average life expectancy has risen. The growth has been continuous in the more developed continents — Europe, the Americas, and Oceania. The gap between Asia and the more developed continents has narrowed over time, but at a slightly unequal rate at initially due to the outsized impact of political situations in China and, to a lesser extent, India (though the convergence has been observed in the rest of Asia as well.) Africa was catching up to the industrialized world until 1990, when it reached a plateau in terms of life expectancy.
- The trend line connecting per capita GDP to life expectancy has risen on all continents. This implies that life expectancy has both a time and a wealth influence. With the exception of Africa, there has been convergence between affluent and poor continents, as well as rich and poor countries within continents.