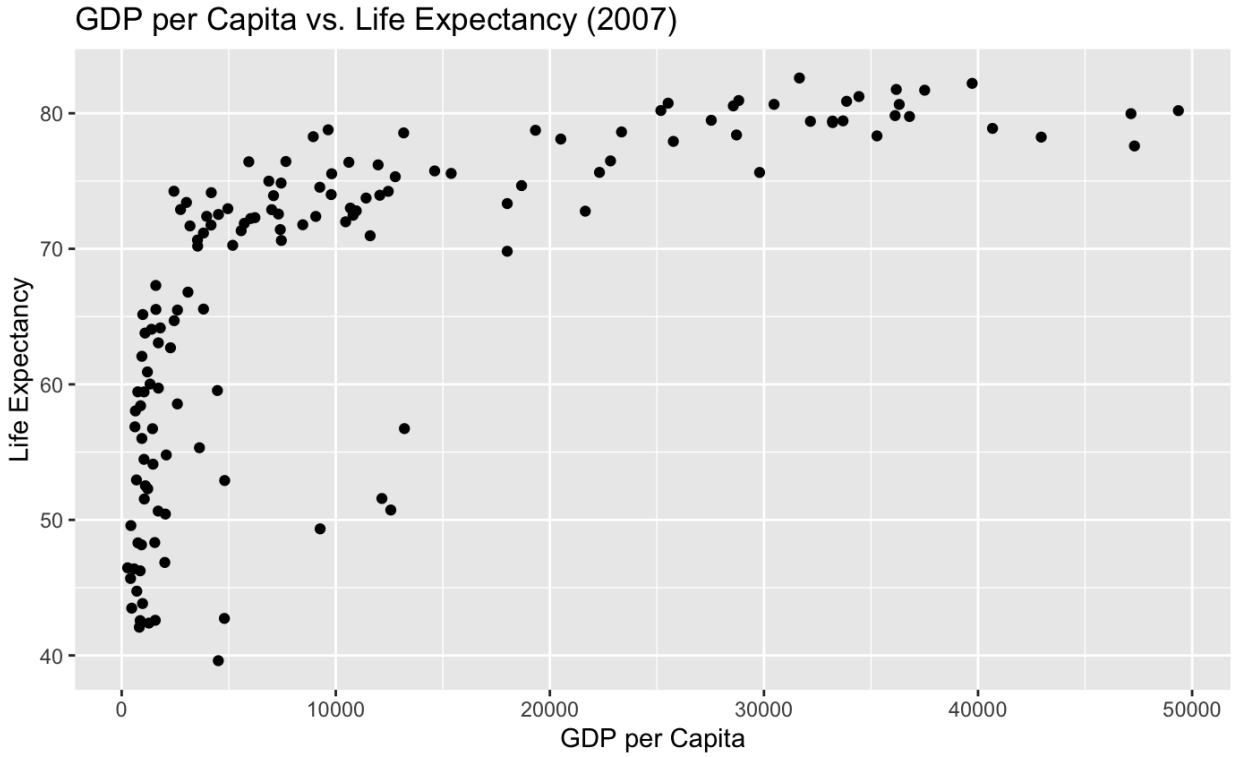
Mini Project 1: Life Expectancy

Implemented by
Ritwik Srivastava (ritsriv)
Sushant Kotwal(skotwal)

The second world war had a sharp effect on the entire world in a lot of factors, the most notable being economic and happiness index effects. In this report, we analyse how life expectancy has changed since the end of the second world war and what factors contributed or did not contribute to this change. For our analysis, we use a dataset called gapminder. The gapminder dataset is made up of 5 continents which are: Asia, Africa, Americas, Europe and Oceania. Each continent has data for the countries in the continent which includes “Life Expectancy”, “Population” and “GDP per Capita” for the years 1952 to 2007. Our first approach was to compare GDP per capita with the life expectancy of each country and the kind of correlation that the two factors have.

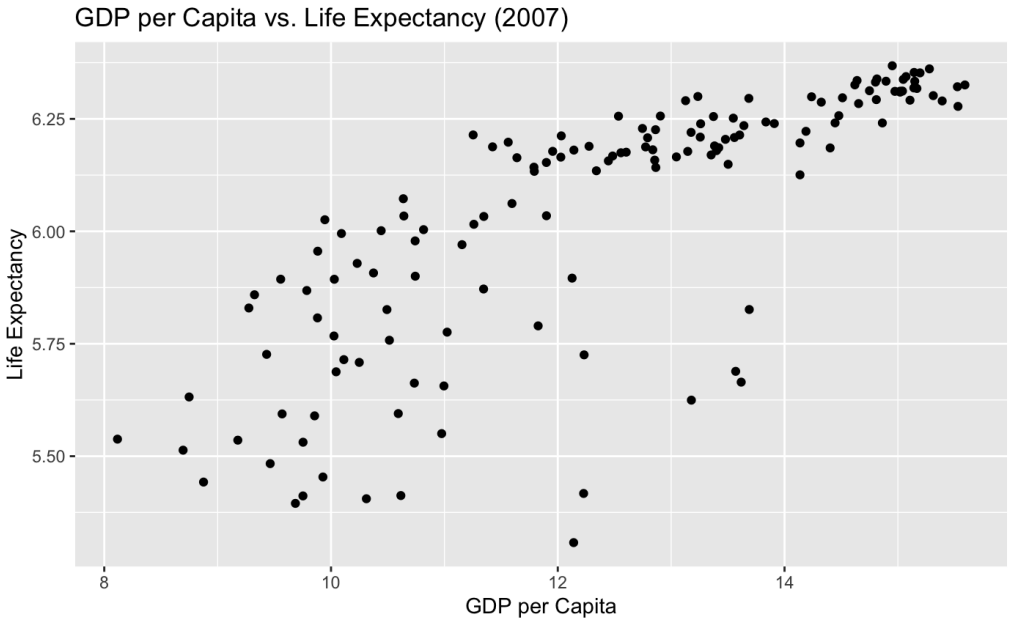
1. Analysis of GDP per Capita and Life Expectancy



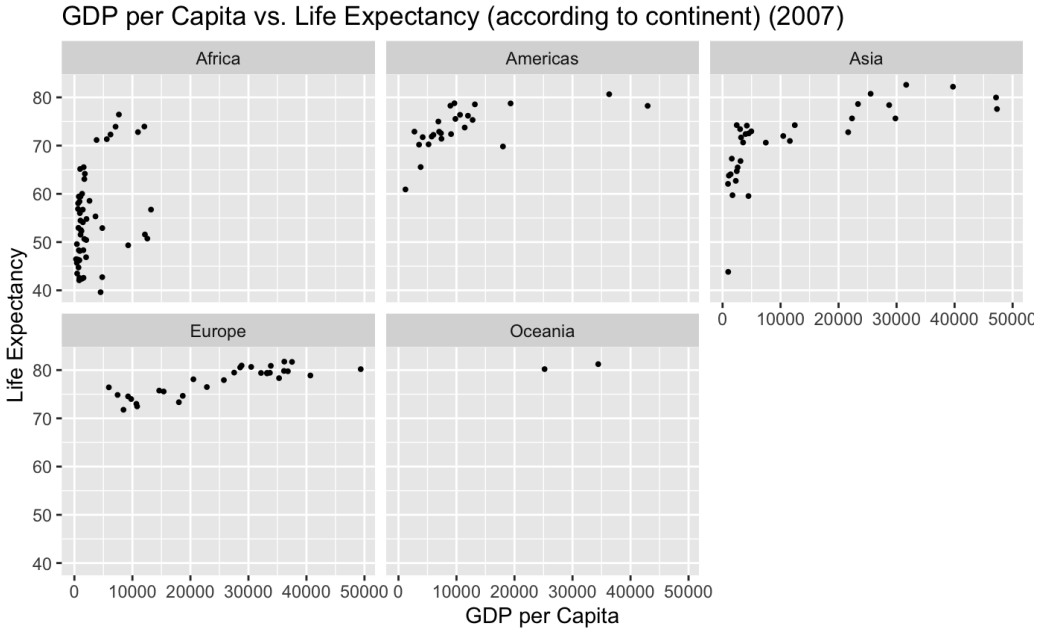
A comparison plot of GDP per capita and life expectancy shows quite a few things, the most notable being the way life expectancy increases as a function of GDP per capita. For reference, the x and y axes represent GDP per capita and life expectancy respectively, with the dots representing a particular country. We see here that there are quite a few countries that have a

sub-10000 GDP per capita. However, this does not change the fact that life expectancy has a steep increase over the same range of GDP per capita, signifying a relationship between the two that does not have the characteristics of a linear trend.

Seeing as the trend follows in this initial comparison, a logarithmic fit might be a more plausible explanation.



A more balanced result is shown in a linear format when the scales of the two attributes are changed to logarithmic.

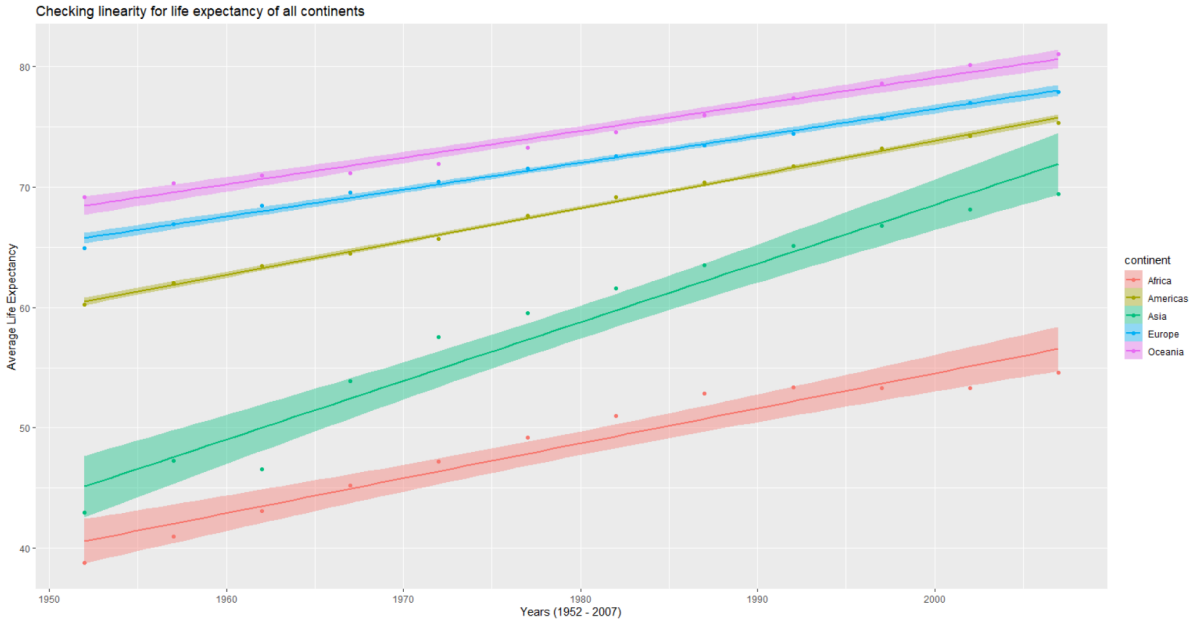


If we look solely at the data for 2007, there is not much data to get to a definitive conclusion, but from the noticeable trends, some similarities and differences between continents are pretty clear:

1. Africa has a pretty low life expectancy with most countries lying in the range of 40-65. There is also a very low average GDP per capita in Africa. This could indicate a positive correlation between life expectancy and GDP per capita. However, even the countries that have a life expectancy > 70 tend to have a not-very-high GDP per capita.
2. America and Asia both follow a very similar pattern of mid-high GDP per capita and life expectancy. This makes sense as few countries in Latin America are not as developed as the US or Canada, explaining the points at the lower ends of the graph. Similarly, Asia has both very developed countries or countries that are still developing, again describing the points at the bottom.
3. As most of Europe is considered to be developed and self-sustainable, most countries in Europe have a mid-high GDP per capita and a very high life expectancy owing to a stable life led by most Europeans.
4. Oceania has only two data points so we can't come to a rigid conclusion in this case.

2. How does the Life Expectancy change in each continent

Since for the first part we analyzed the relationship between life expectancy and GDP for every continent, there is a question of how important is **Time** for determining the change in **Life Expectancy** for all continents
So we check the relationship between Time and Average Life Expectancy for each continent from the years 1952 - 2007 using the below plot.



This graph shows Time frame on x-axis and according change in Average life expectancy on y-axis for each continent. We used weighted averages for calculating the Average Life Expectancy for each continent based on population of each country in the continent. This is because **countries like Bahrain and Kuwait have very low population compared to countries like India and China which are two of the highest populated countries in Asia.** We used the linear modeling on our data to check how linearly the model fits to the data and difference between the expected value and actual value for average life expectancy is depicted using the shading around the line of regression.

We then made the following observations from this plot:

1. We see that the average life expectancy for all continents is increasing progressively, especially for continents Europe, Americas and Oceania.
2. Africa has the lowest average life expectancy beginning at the year 1952 which was around 39 years and it increases steadily till 1987. After 1987 we see a period of stagnation from 1992 to 1997 and a minor decrease in average life complexity till 2002. Again it increases from 2002 to 2007. For 2007, the average life expectancy for Africa stands at around 54 years.
3. Asia has the second lowest average life expectancy in 1952 which was around 43 years. There is progressive increase in the the value except for the period 1957-1962. Again after 1962 there is rapid increase in the life expectancy till 2007 and Americas, Europe and Oceania show a relatively high life expectancy in 1952 which is in the range of 60-70 years.
4. For Oceania, there is a stagnation period from 1962-1967 and thereafter there is an increasing trend. It reaches highest life expectancy in 2007 which is around 82 years
5. There is not any sign of dip in the trend for Americas and Europe. They have the highest life expectancy which is roughly 77 and 75 years respectively.

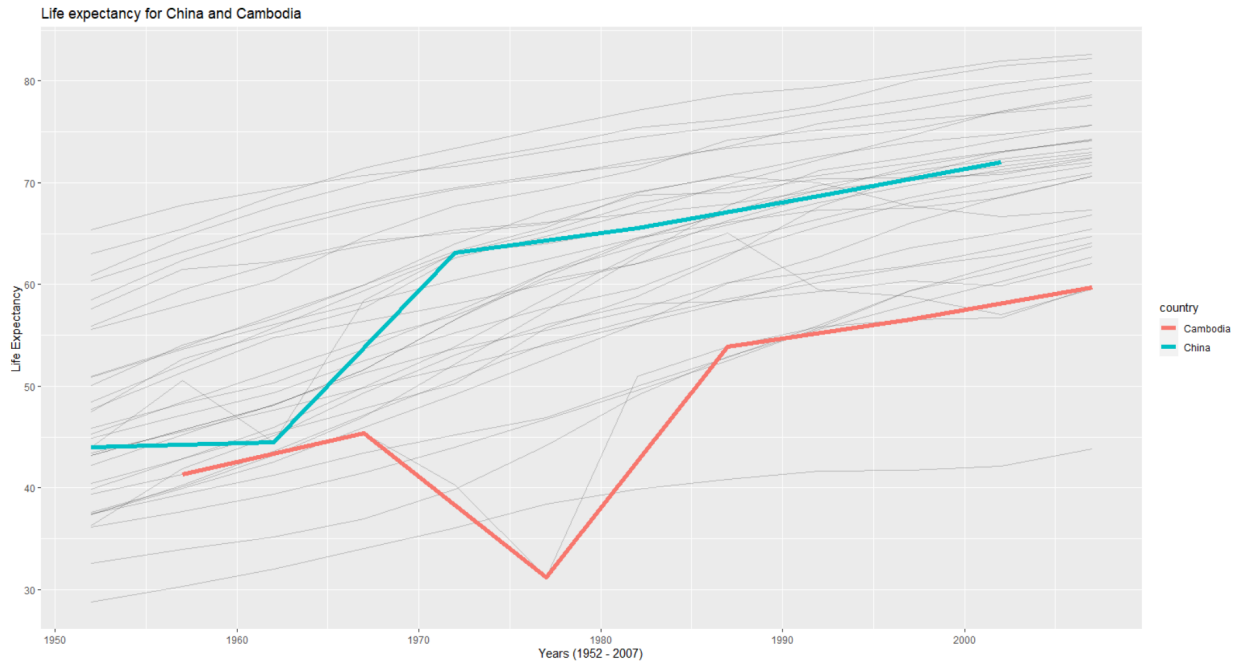
Judging from our observations, we see that Asia which started at 43 years of average life expectancy in 1952 ended up with almost 69. So we can make out from the above graph that **Asia starting from a low life expectancy has partially caught up to Americas,Europe and Oceania.**

This increase in **life expectancy can be termed to be linear for America, Europe and Oceania as they show a very good fit to the linear model**. We see that for Africa and Asia the increase has not been linear and it shows a significant amount of standard deviation from the linear model. Africa shows a slow at the beginning from the period 1952-1962 and after 1970-1987 there was a fast increase in life expectancy. It seems that for the period 1992-2002 there is no increase because the population for that period is not showing any increase. Same can be said about Asia for the period 1957-1962 where there is a major decrease in life expectancy.

We now check for the continent which has caught up partially to other continents which is Asia. The decline in particular for 1957 - 1962 can be explained for Asia only because of China.

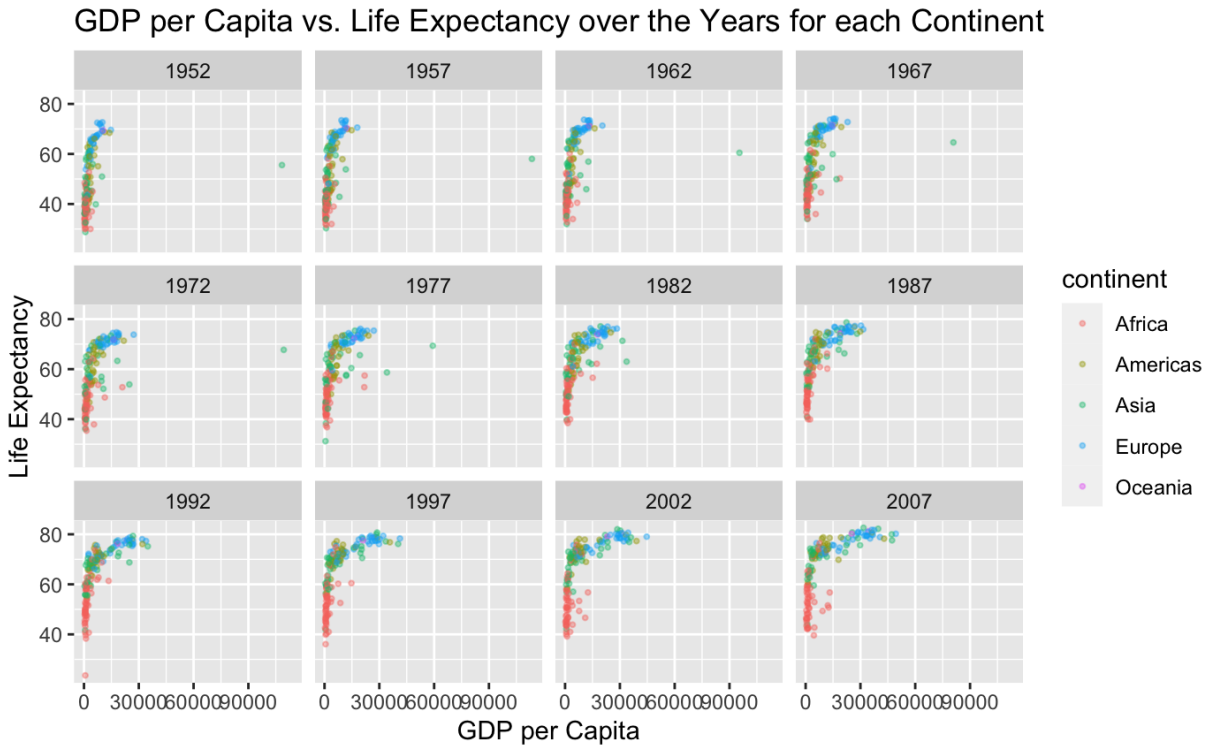
Referring Appendix Plot 1 we see that China and Cambodia are the only countries that experience a major dip in life expectancy.

For simplicity we plot the change of life expectancy for both the countries in below plot:



In this plot, the grey lines in the background reflect the life expectancy for all other countries and we are not highlighting those since we only want to focus on China and Cambodia. For year 1962 we see that only China has a dip and not other countries yet life expectancy for continent is declining for that year. This is because China is the highest populated country in Asia. When we look at Cambodia for 1977 we see that it has the deepest decline among all countries across all years. In spite of that there is no decline in life expectancy of Asia for 1977 and on the other hand there is an increase. This is because Cambodia is among the low populated counties.
So we can make a conclusion that countries may or may not affect the life expectancy of the continent but it depends on the population and it is the one making the difference.

Now Let us see how the Life Expectancy changes according to both Time and GDP



The above plot describes the dependency of GDP per capita on life expectancy and how, over the years, that dependency has changed. The x-axis represents an increasing GDP per capita and the y-axis represents an increasing life expectancy, with a separate plot for each year in 5-year gaps and the colours represent continents.

Firstly, it would be fair to say that on average, yes, life expectancy has increased over time since the second world war. The cluster of dots (countries) in the first plot (1952) can be seen starting from a sub-40 value and caps somewhere around the 70-year mark, which is quite low if you think about it. However, as time passes, this cluster can be seen moving up to the point in 2007 that the lowest life expectancy for a country is 40 and the highest is 80+ (with quite a few countries in that ballpark).

On the other hand, GDP per capita compared with life expectancy doesn't quite tell the same story. There is a gradual increase in life expectancy over time, yes, but the increase in GDP per capita does not share the same trend. For example, looking at certain countries in Africa (highlighted in red) a lateral increase over the x-axis, signifying an increase in GDP, does not reflect the same on the y-axis aka life expectancy. Moreover, for the continents of Asia and the Americas, the cluster of countries constantly remains somewhere around the 20000 mark for GDP per capita, and yet shows a rise in life expectancy. However, there is a notable exception to this trend. European countries, that were the most affected during the second world war, faced a dip in their economy as well as life expectancy due to the war. But, over the years (most

notably years 1982-2002), the plots show a steady increase in GDP as well as life expectancy. This shows that Europe has had a better standard of living post the war, resulting in a higher life expectancy. Additionally, we can also conclude that a better standard of living is also a direct result of a higher GDP per capita.

Appendix: