Stat 470/670 Mini Project 1: Life Expectancy

Team members: Akhila Sakiramolla (asakiram), Mitali Tavildar (mtavilda), Riya Shetty (rishett)

Introduction:
The aim of this Mini Project is to understand and learn about life expectancy and its relationship to GDP per capita. For this, we have made use of the package gapminder which for each of 142 countries, over 5 continents, provides values for life expectancy, GDP per capita, and population, for a particular number of years, from 1952 to 2007.

Data Description:
The 'Gapminder' data provides values for life expectancy, GDP per capita, and population, every five years, from 1952 to 2007 for each of 142 countries. GDP per capita measures the value of everything produced in a country during a year, divided by the number of people. The unit is in international dollars, fixed to 2017 prices. The data is adjusted for inflation and differences in the cost of living between countries, known as PPP dollars. Following is the table that describes the gapminder data.

Metric	Life Expectancy	Population	GDP Per Capita
Min	23.60	60,010	241
Median	60.71	7.02M	3,531
Mean	59.47	29.6M	7,215
Max	82.60	1.32B	113,523

Problem Statement:
Can the increase in life expectancy since World War 2 be largely explained by increases in GDP per capita?

Findings:
1) We initially wanted to understand the trend that Life Expectancy follows over the years across continents. From this analysis, we observed that Americas, Europe, and Oceania seem to have a linear increase over years without the impact of any significant factors, whereas Asia (1950 - 1965) and Africa (1990 - 2000) were impacted during the respective time periods due to some countries.
2) We then wanted to observe if any other factors are affecting Life Expectancy apart from changes within country. We wanted to see the affect of GDP on Life Expectancy across continents. We first wanted to study the relationship between GDP and Life Expectancy in 2007 to see what the recent trend looked like and if it is being directly affected by GDP. We observed that there is no linear relationship between GDP and Life Expectancy in 2007.
3) We wanted to check if a similar pattern is observed across the past years so we studied the relationship between GDP and Life expectancy over time for each continent and observed that changes in life expectancy cannot be entirely explained by changes in GDP per capita for continents like Europe and Africa. Whereas, for Asia and America it sort of follows a linear relationship.

4) We further wanted to see if there is effect of time on the relationship between GDP and Life expectancy and observed that the relationship seems to converge for all the continents over time with an exception of Africa.

Conclusion:

From all our analyses, we finally conclude that increase in life expectancy since World War 2 cannot entirely be explained by increase in GDP per capita but both the GDP and time have an affect on increase of Life Expectancy.

Find the detailed analysis for the conclusion and each of the steps, below:

Analysis of Life Expectancy over time across continents

1) From **Fig-1** we can observe that for the continents Americas, Europe, Oceania the Life expectancy (weighted mean) is increasing approximately linearly over the years. We can say that with assurity because Europe and Americas have a substantial number of data points (countries) as compared to Oceania. Since Oceania has only 2 countries, based on their trend, it is visible but not reliable to say that the trend is a linear upward going one. Whereas for Africa, it is increasing linearly but is constant from 1990 to 2000 and for Asia, we observe a dip in the initial years, between 1960 - 1970 after which the increase is linear. For the continents Americas, Europe, and Oceania, life expectancy ranges from approximately around 60 to 80. But for continents like Africa, Asia the scale for the weighted mean of life expectancy ranges from around 40 to 55 & 40 to 70 respectively.
2) We can observe that even though in the initial years the life expectancy values for Asia are small (starting between 40 - 45), compared to Americas, Europe and Oceania (life expectancy starting range is from 60 -70 and towards 2007 the range is 75 - 85), it seems to have caught up for Asia towards 2007 as the value increased and now lies in the range of 60-70. Whereas for Africa the starting range is low around 40 and towards 2007, it is still below 50 and not close to other continents. These changes are not general across continents and may be attributed to various factors like rapid changes in life expectancy within countries, time, etc., We dig deeper to check if countries in these continents are the reasons for the same, particularly for Asia and Africa.
3) For Asia, at the beginning of the trend graph till the year 1970, we observe some abnormality, followed by an approximately linear trend. (**Fig-1**). When we dig deeper and look at the distribution of Life expectancy of countries within Asia (**Fig-2)**, We observe that the life expectancy value in the beginning years is widely different for different countries, for some it is at a very low value below 30 and for some at an extreme high at 60. We see that the trend graph for China is similar to the overall trend graph for the whole of Asia between 1950 and 1965. During this period, the Life expectancy of Yemen, Rep., Oman and Hong Kong, China is low but increased suddenly and linearly, which may be the reason for the sudden dip and subsequent linear increase for Asia.
4) During 1990-2000, we observed that in Africa (**Fig-1**), there is a change in the linear path of the graph. To justify the flattening of Life expectancy, we plotted graphs for countries within Africa for this particular period (1990 - 2000) to look at the average life expectancy as well. We observe (**Fig-1 in Appendix**) that for most of the countries, the Life expectancy value lies between 50 - 60, which could justify a flattened graph around 53 for the continent of Africa.

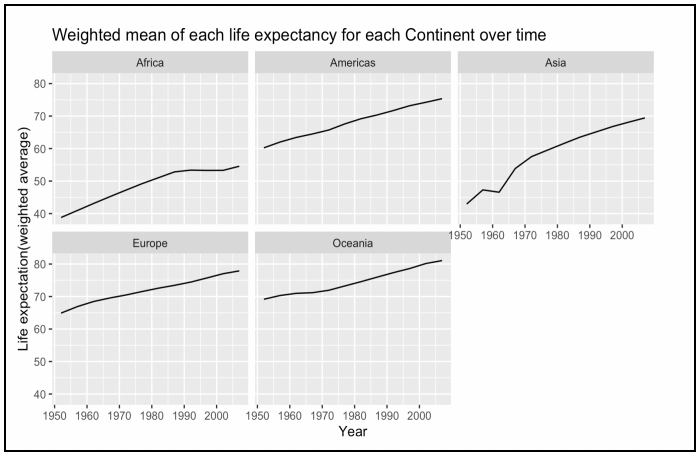

Fig - 1 Life Expectancy over years across continents

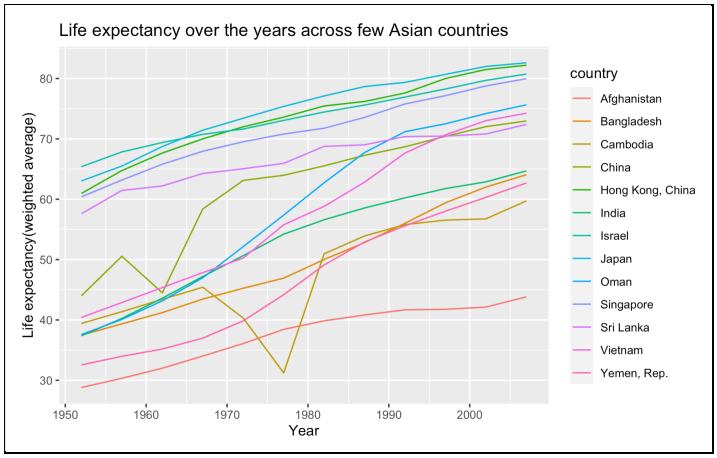

Fig - 2 Life Expectancy over years across a few Asian countries

Analysis of the relationship between GDP and life expectancy over continents in most recent times (2007)

1) From **Fig-3**, we notice that the relationship between GDP and Life Expectancy seems to be increasing upto a certain point over time but this relationship is not linear. It can be noticed that the trends cannot be well-described by a simple model such as a linear model, a more complicated model is required like loess, which seems to fit.(**Fig-2 in Appendix**). The points are not spread across the linear line which proves that the graph is not linear. As the data is not evenly spread with huge variation, this can affect the linear model's predictions as they can easily affect the direction/slope of the line of best fit.
2) This trend is also observed in the plot of GDP and Life Expectancy for Africa, where a linear model cannot be fit, whereas for the rest of the continents we can fit a linear model (**Fig-3 in Appendix).** After log transformation of GDP (**Fig-4 in Appendix**), we observe a similar pattern as in the case of GDP without log transformation, for the rest of continents except Africa.
3) We notice that the median for each and every continent varies by a huge difference (**Fig-5 in Appendix**). The Americas and Asia have almost the same median for Life Expectancy even though the spread of GDP per capita is more for Asia. This difference is because of the varying population within the countries on these continents. We can also notice that the graph for Oceania seems linear, this is because there are just 2 countries in Oceania, and the Life Expectancy can't be drawn over GDP per capita for these points for the year 2007 in a log graph.
4) From the ecdf it seems that the bulk of distributions are pretty similar but offset from each other with an additive shift. It can be said since most curves are mostly shifted along the x-axis from each other and overall shape is same for each value of the Life Expectancy.
5) From the graphs, we can comment that the continents follow a linear pattern when we take the log of the GDP percentage. Hence, it can be concluded that the graph follows an additive shift approximately. For the continent of Oceania, it is difficult to decide whether it is a multiplicative shift or an additive shift since the data points for Oceania are less and the data seem to be very limited. Hence, Oceania can be considered as an outlier and requires a more complicated method to estimate the shift.

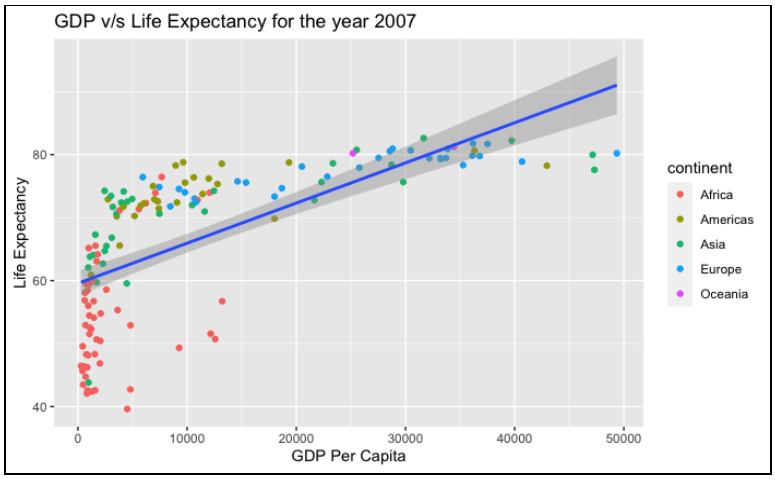

Fig - 3 GDP v/s Life Expectancy across continents

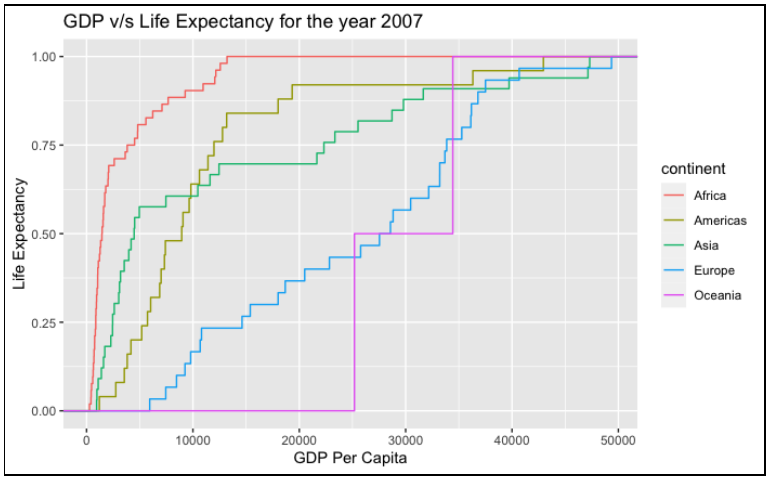

Fig - 4 ECDF plot of GDP v/s Life Expectancy across continents

Analysis of relationship between Life Expectancy and GDP over time across continents

1) We plot the increase in Life expectancy over years in (**Fig-5**), and the corresponding shift in GDP per capita. There is certainly an increase in Life expectancy over time, but a very slight overall increase in the GDP per capita (**Fig-6 in Appendix**). There seems to be a direct linkage between Life expectancy and GDP for most of the continents. For most continents, GDP appears to have a slight effect, as an increase in one corresponds with increase in Life expectancy. We fit a linear model for each continent, to plot Life expectancy vs GDP per capita. For the purpose of the same, we considered the log of GDP per capita. We observe that over the years, the relationship between the Life expectancy and GDP becomes linear for all the continents.It seems to converge for all the continents with the exception of Africa.
2) For Asia (**Fig-6**), we observe that Life expectancy and GDP do not have a linear relationship over the years. We can also observe this from their individual distributions (**Fig-7,8 in Appendix**). Life expectancy is linearly increasing over years, whereas GDP does not have a linear change. It increased from 1952 to 1957, then became constant from 1957 to 1967, then there was a sudden peak in 1972 and so on. We can conclude that life expectancy cannot be entirely explained by GDP for Asia as it doesn't have a direct causal effect from the below graphs.
3) For Europe (**Fig-7**), we observe that Life expectancy and GDP sort of have a linear relationship over the years. We can also observe this from their individual distributions (**Fig-9,10 in Appendix**). Life expectancy is linearly increasing over years, whereas GDP is also linearly increasing with just one phase during 1987 to 1992 where it seems to be constant. We can conclude that life expectancy may be entirely explained by GDP for Europe as they exhibit a linear relationship from the below graphs.
4) For Africa (**Fig-8**), we observe that Life expectancy and GDP sort of have a linear relationship till 1977 and then do not change linearly after that. We can also observe this from their individual distributions (**Fig-11, 12 in Appendix**). Life expectancy is linearly increasing till 1987, then it is constant till 2002 and then increases again till 2007, whereas GDP is also linearly increasing till 1977 and then doesn't change linearly. We can conclude that life expectancy may not be entirely explained by GDP for Europe as they do not exhibit a completely linear relationship from the above graphs.

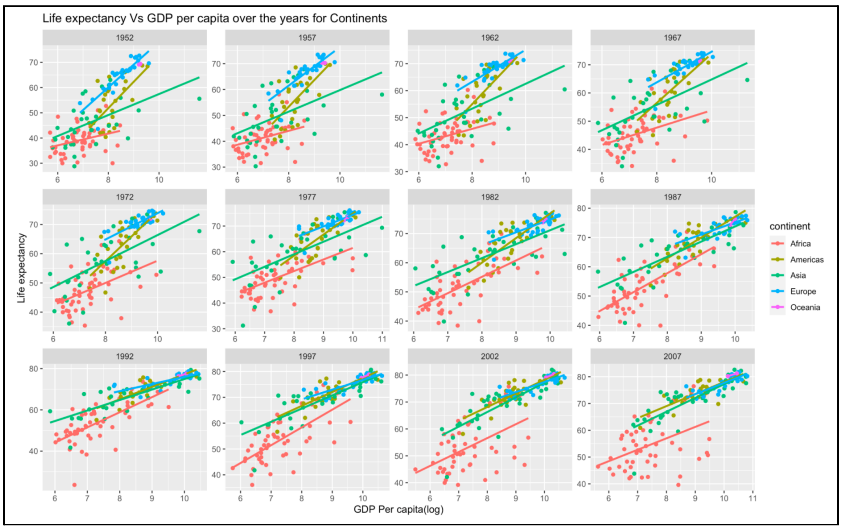

Fig - 5 Life Expectancy vs GDP over years across continents

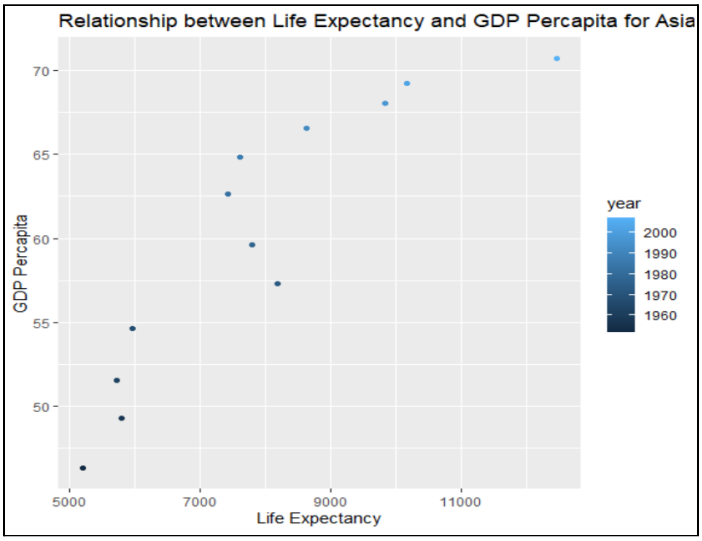

Fig - 6 GDP Percapita vs Life Expectancy over years for Asia

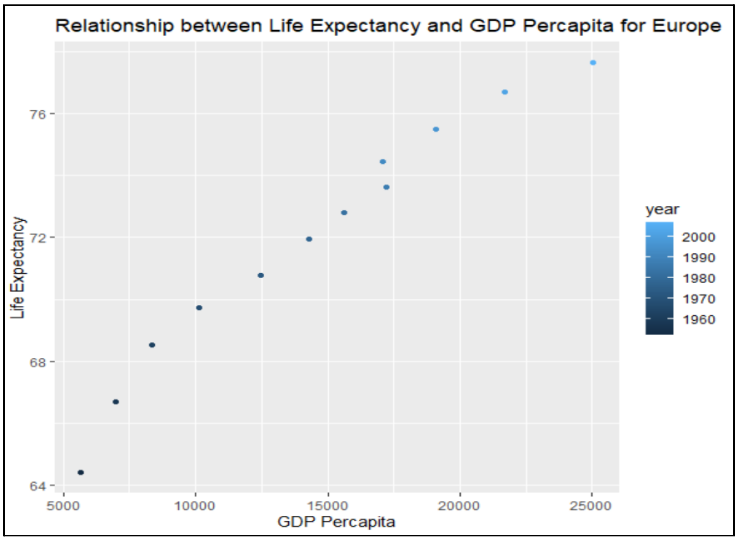

Fig - 7 GDP Percapita vs Life Expectancy over years for Europe

5) For the Americas (**Fig-9**), we observe that Life expectancy and GDP sort of have a linear relationship over the years till 2002 and then a sudden increase in 2002. We can also observe this from their individual distributions (**Fig-13, 14 in Appendix**). Life expectancy is linearly increasing over years, whereas GDP is also sort of linearly increasing with a sudden increase in 2002. We can conclude that life expectancy may be entirely explained by GDP for Americas as they exhibit a linear relationship from the below graphs.

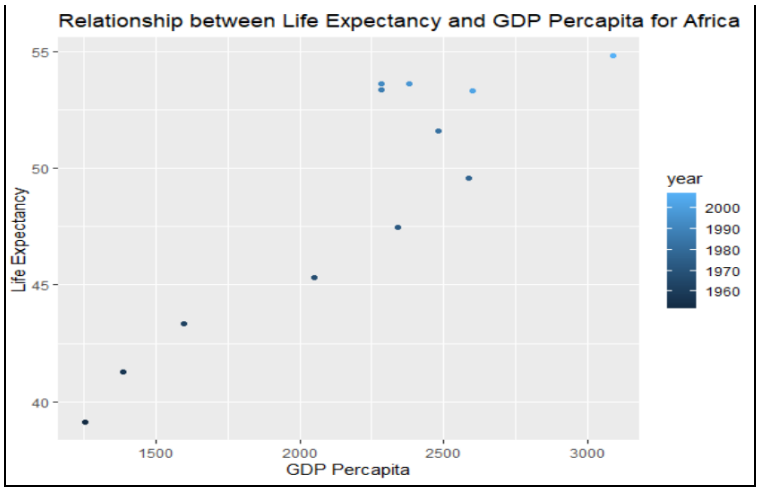

Fig - 8 GDP Percapita vs Life Expectancy over years for Africa

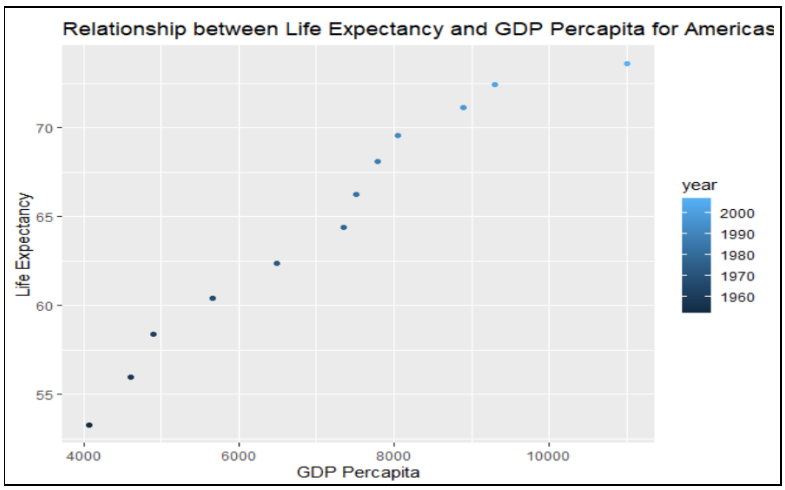

Fig - 9 GDP Percapita vs Life Expectancy over years for Americas

Appendix

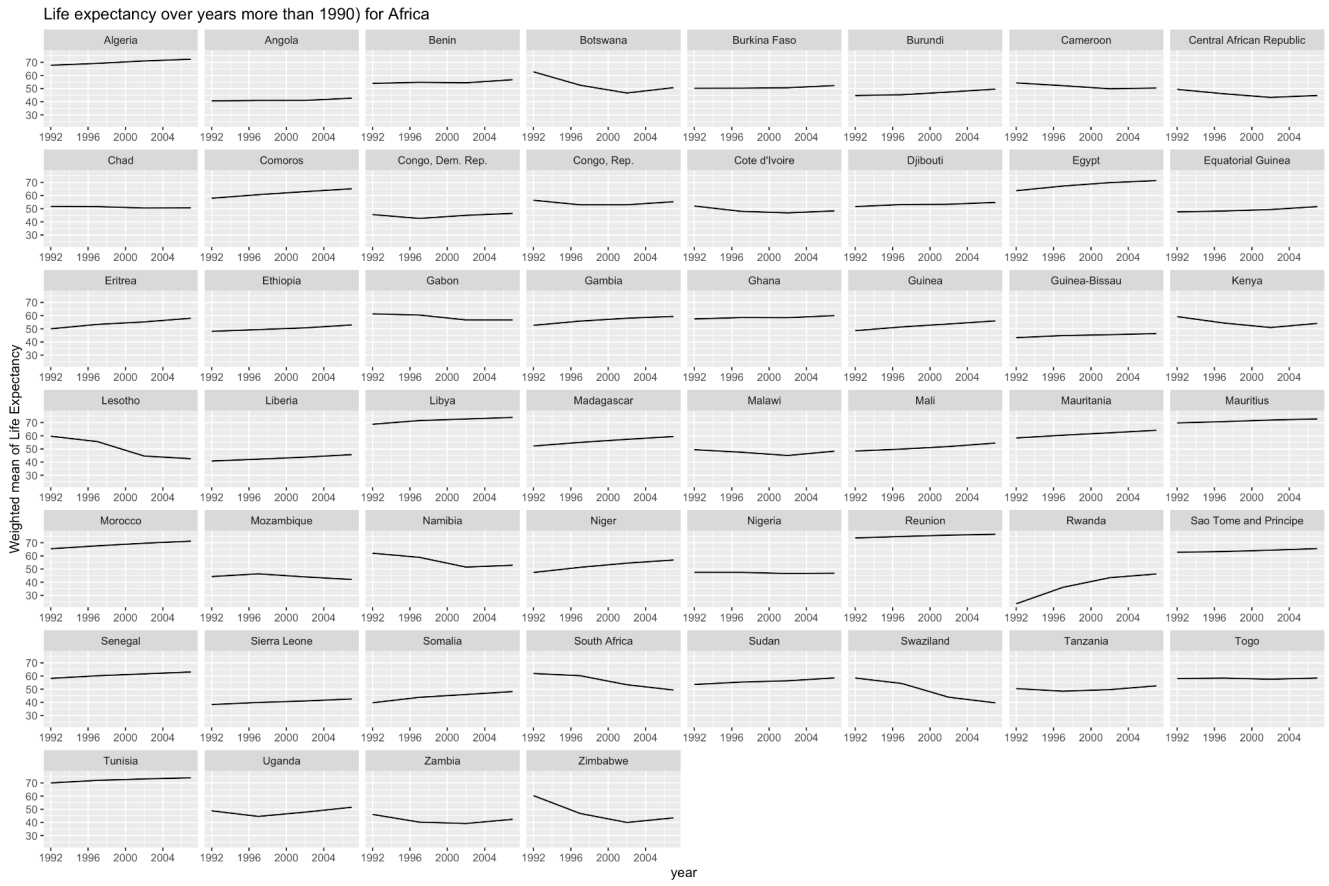

Fig - 1 Life Expectancy over years across counties with Africa (1990 - 2000)

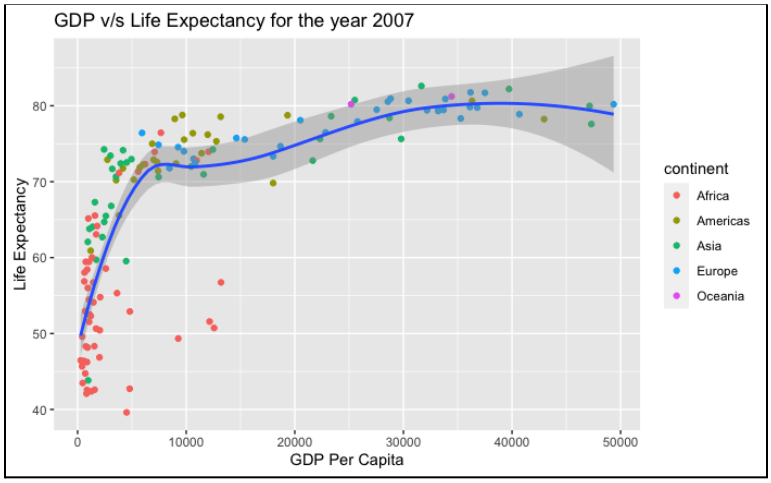

Fig - 2 GDP Percapita vs Life Expectancy across continents

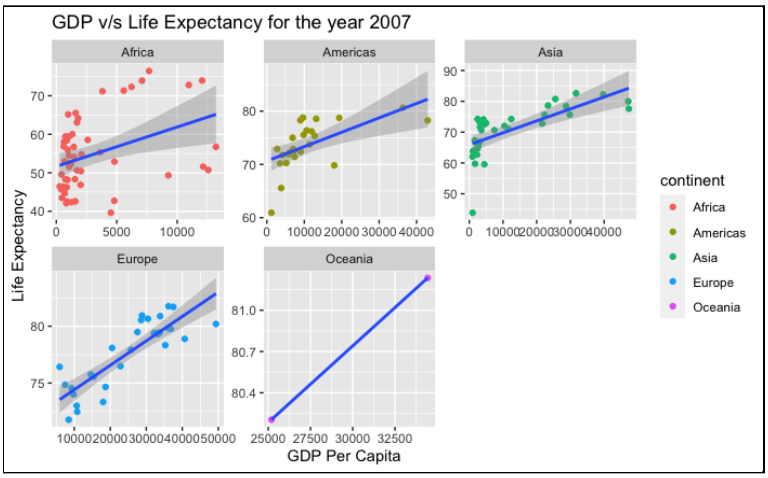

Fig - 3 GDP Percapita vs Life Expectancy across continents for the year 2007

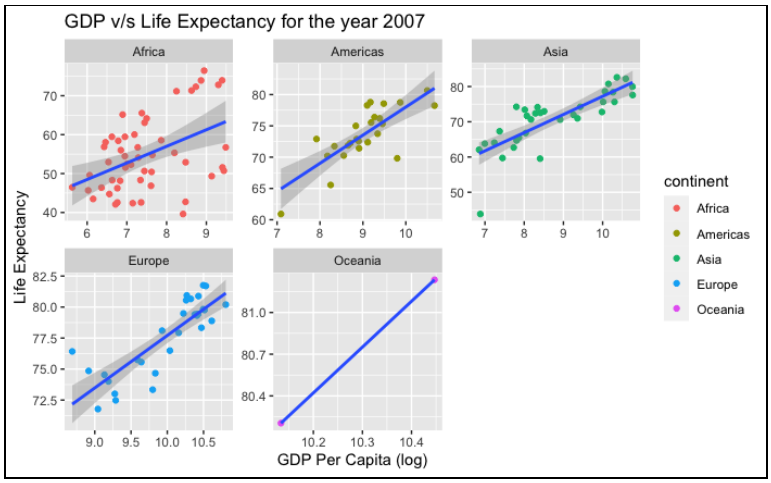

Fig - 4 log(GDP Percapita) vs Life Expectancy across continents for 2007

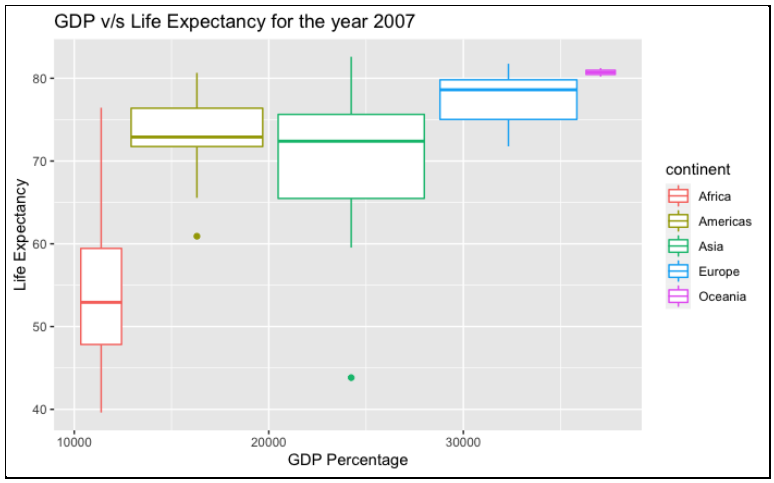

Fig - 5 Box plot of GDP vs Life expectancy across continents

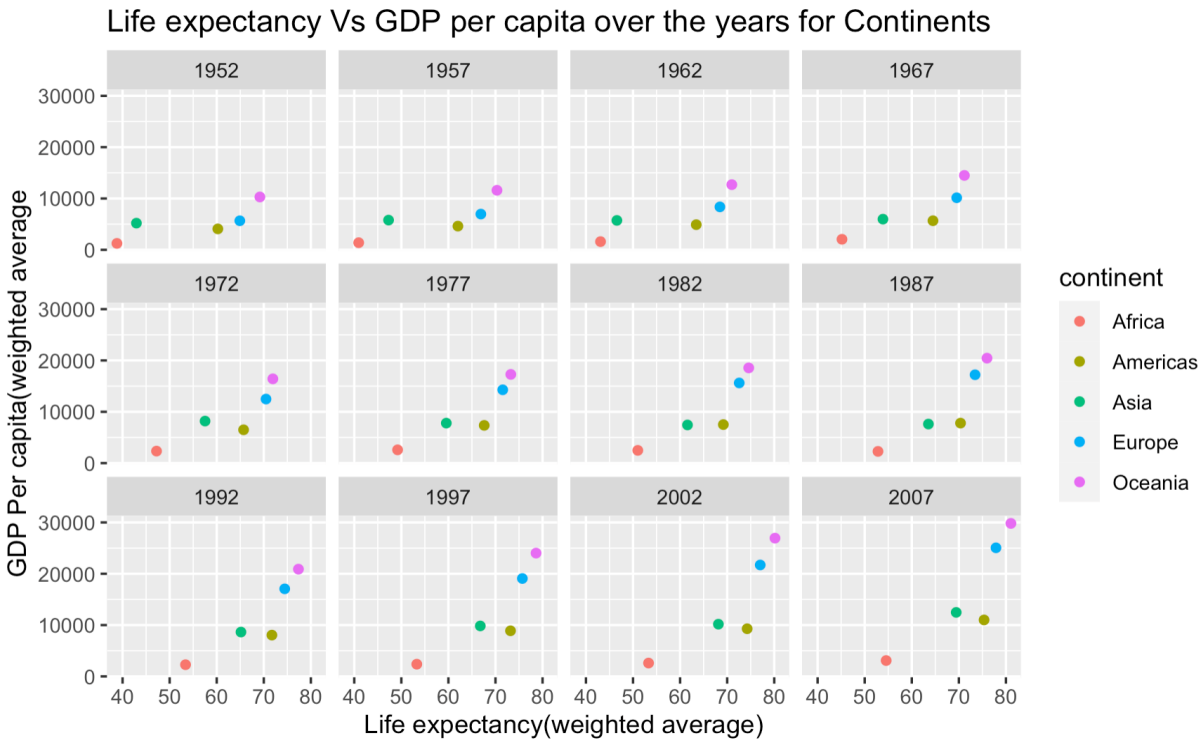

Fig - 6 Life Expectancy vs GDP over years across continents

ASIA

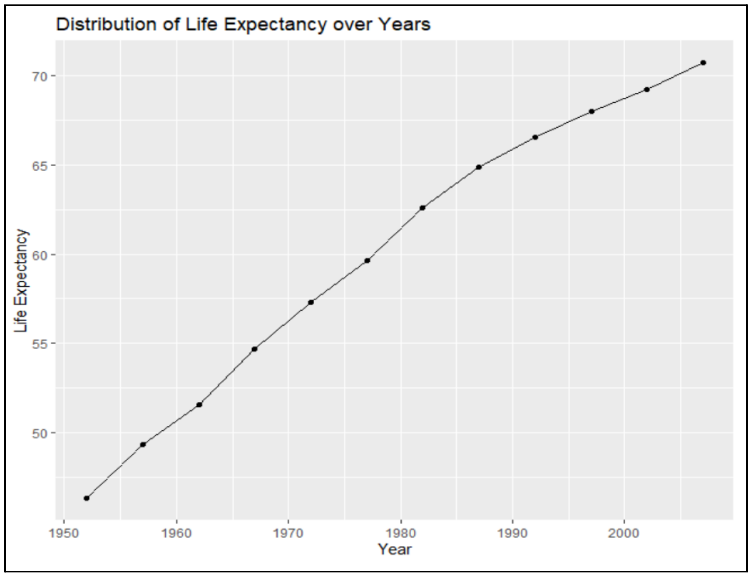

Fig - 7 Distribution of life Expectancy over the years

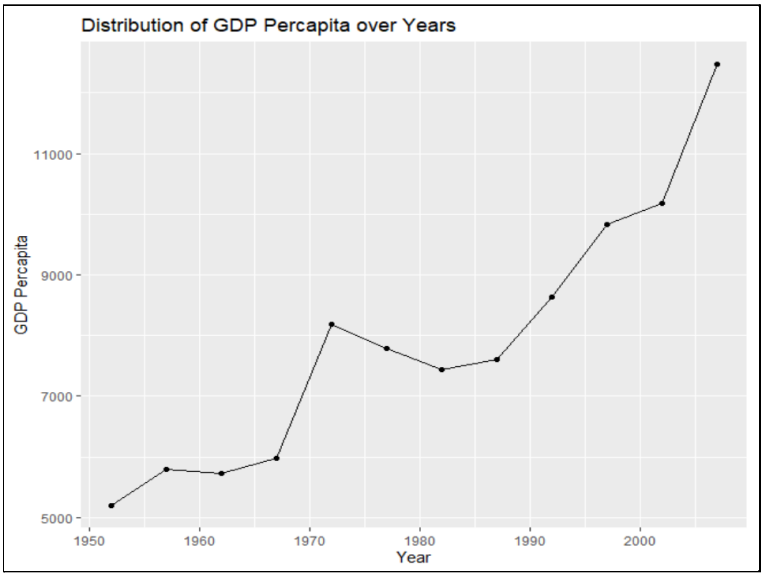

Fig - 8 Distribution of life Expectancy over the years

EUROPE

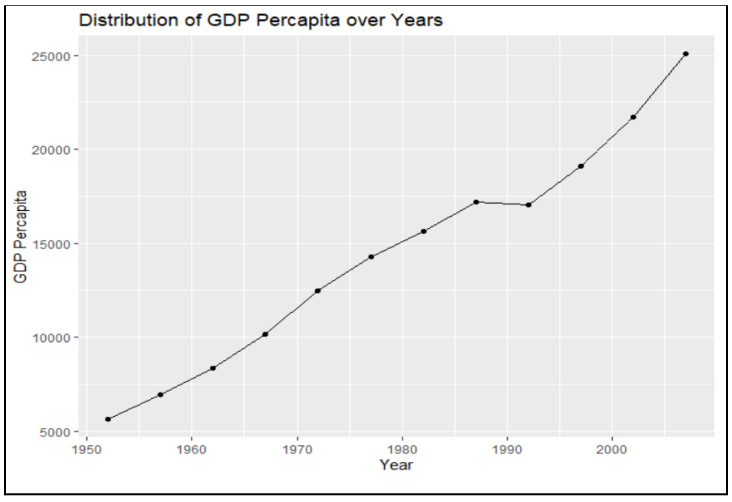

Fig - 9 Dsitribution of GDP per capita over years

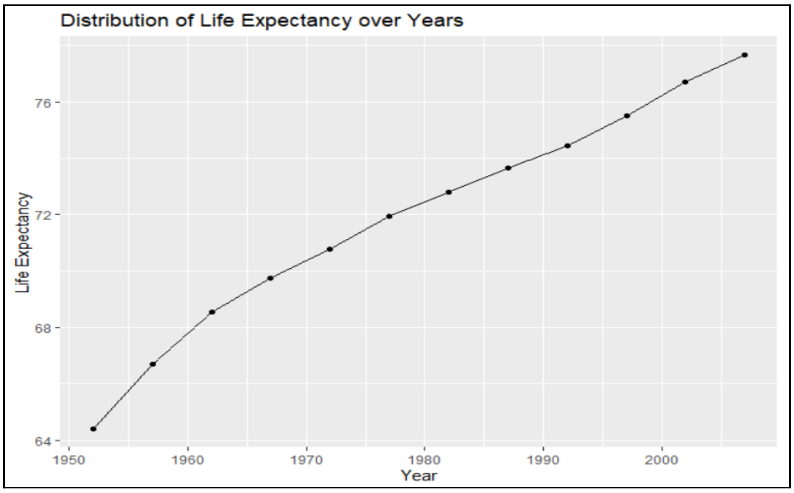

Fig - 10 Dsitribution of GDP per capita over years

AFRICA

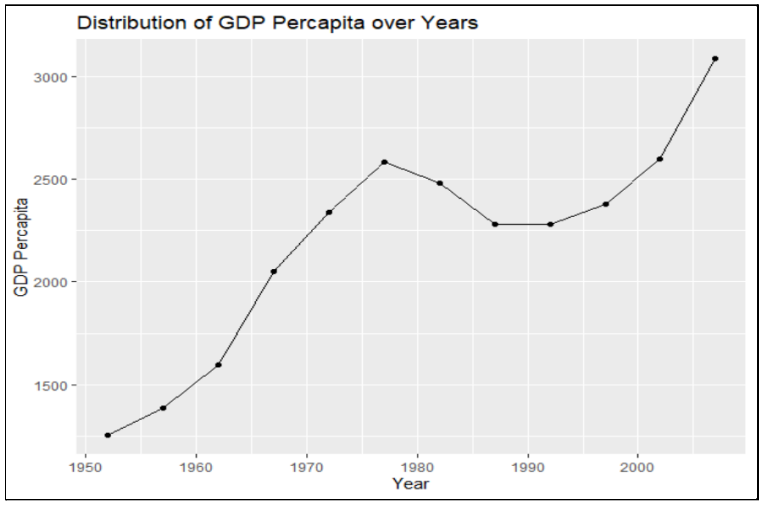

Fig - 11 Dsitribution of GDP per capita over years

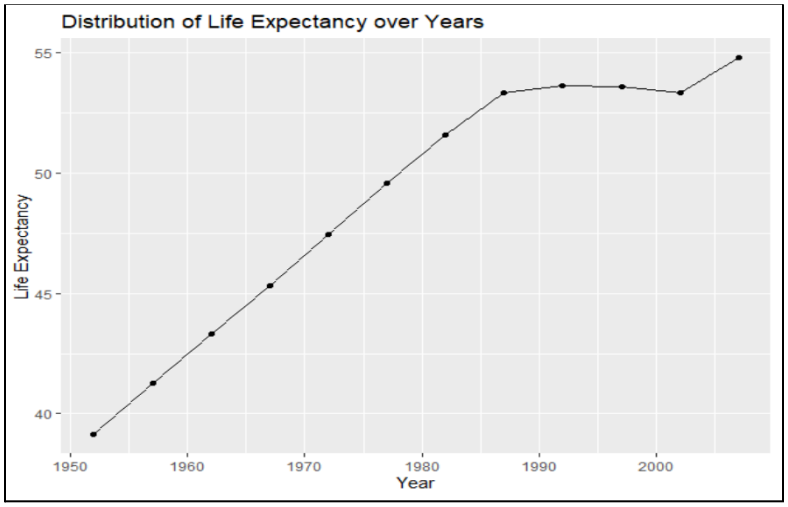

Fig - 12 Dsitribution of GDP per capita over years

AMERICAS

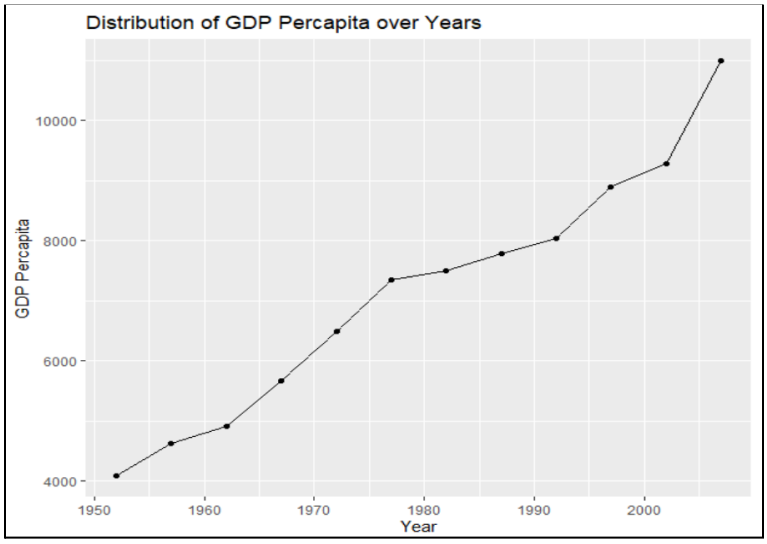

Fig - 13 Dsitribution of GDP per capita over years

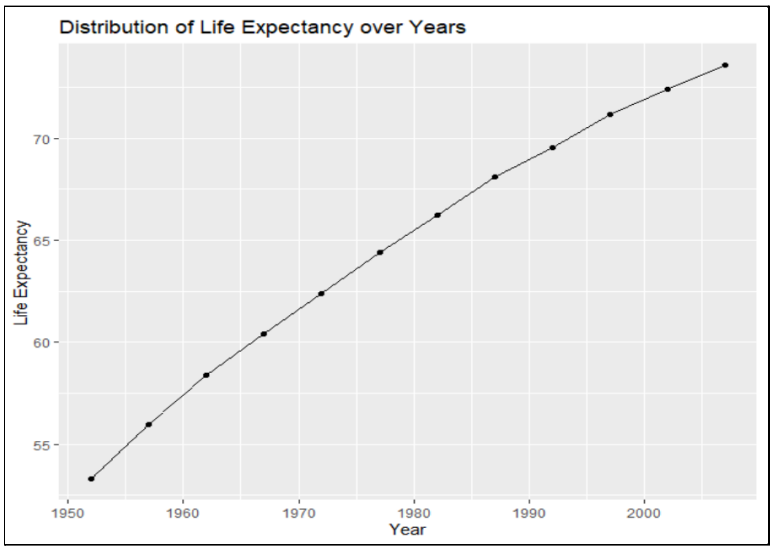

Fig - 14 Dsitribution of Life Expectancy over years