

# Mini-project 1: Life expectancy

## Problem 1

### Introduction:

In this section, we aim to find out how Life Expectancy varies with GDP per capita in the year 2007 and also analyze the trends by continent.

### Strategy:

- 1) First, we shall load the required libraries. Next, we create a data frame called "In2007", which will capture necessary information like GDP per capita, life expectancy, population for the year 2007.
- 2) Then we tried to capture the trend of Life expectancy vs GDP Per capita with linear and loess smoothing but neither seem to fit the data properly.
- 3) We then perform a Log2 transformation on the GDP per capita data and plotted using linear smoothing. (Figure 1)
- 4) Next, we plotted a graph to understand the trend of Life expectancy vs GDP Per capita in Asia because there seems to be unusual activity in the few of the years. (Figure 3)

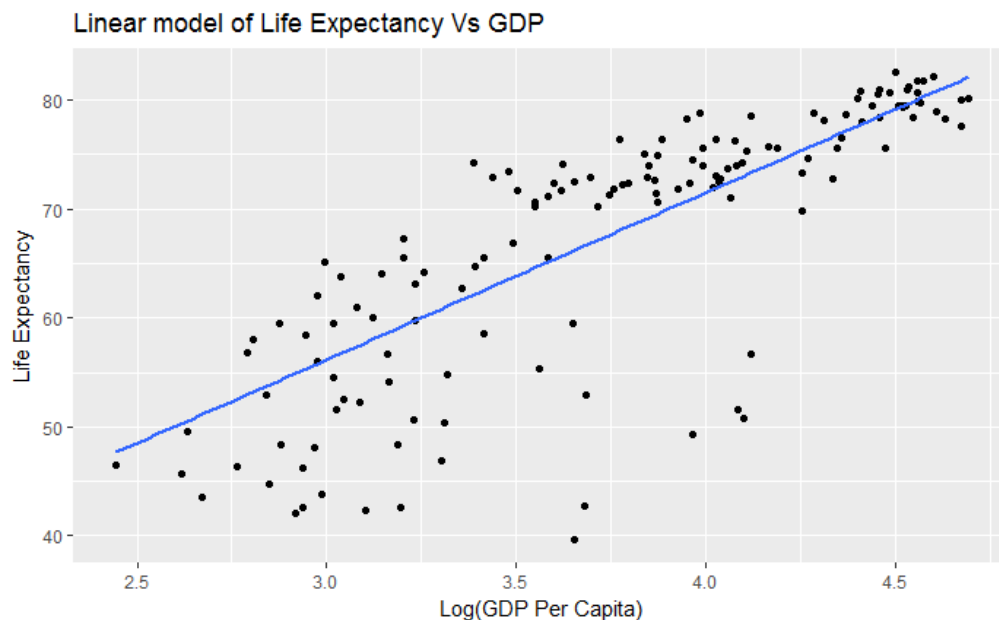


Figure 1: Linear Model of Life Expectancy Vs GDP PerCapita

### Conclusions/Inferences:

- 1) Firstly, without transforming the GDP data we could not fit the model properly. After performing a log2 transformation, the **linear model seems to do a good job describing the data**. When we try any complicated models like loess, there doesn't seem to be any change in the fit.
- 2) When checked for accuracy, linear model explains 65.4% of the variance in life expectancy in 2007. (Figure 2)
- 3) The overall graph increases linearly with the increasing GDP per capita. This indicates that Life expectancy increases with increasing GDP.
- 4) Regarding the trend of Life expectancy vs GDP Per capita per each continent:
  - a. There are only two data points for continent Oceania, so nothing concrete can be said about Oceania. America, Asia and Europe have linearly increasing data, and all seem to converge to a constant life expectancy towards the end. Africa has different distribution, but it is also increasing.
  - b. By looking at the plot, the trend for all continents look almost similar and could be explained by an additive shift. This is evident when the slopes and intercept values are compared. All of them have a slope around 3 with varying intercepts ranging from 22 to 35 indicating an additive shift.

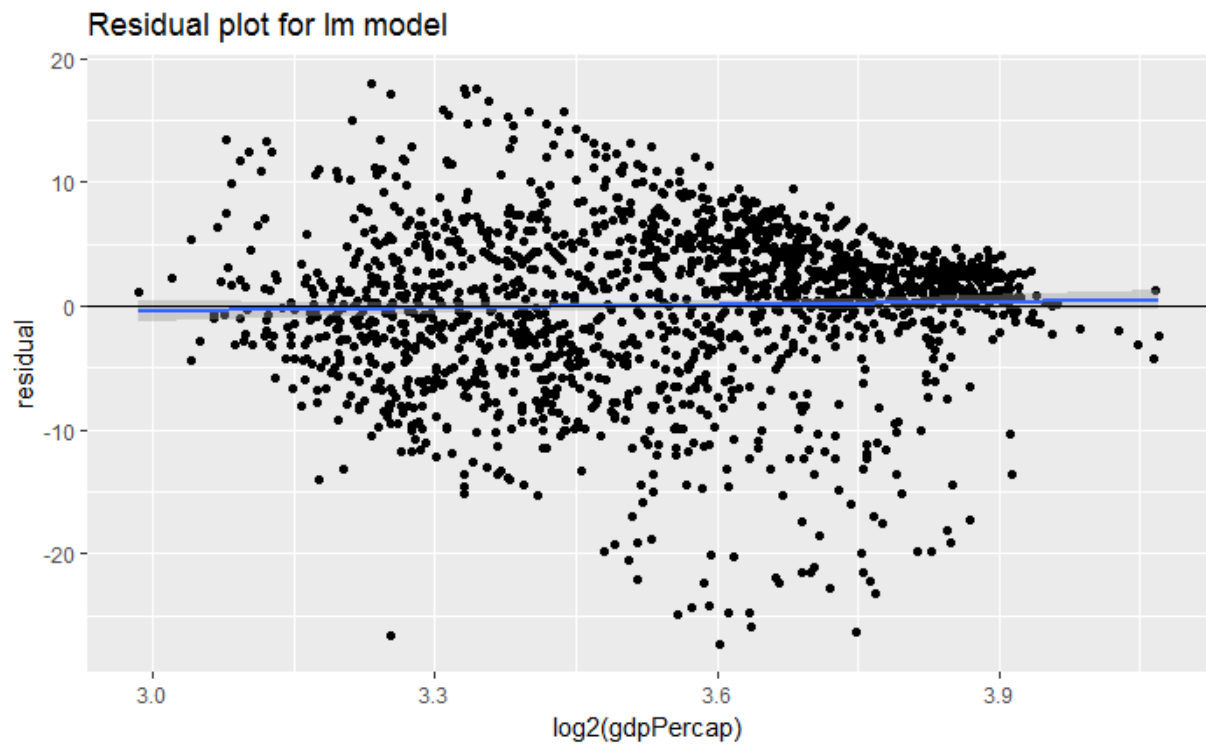


Figure 2: Residual plot

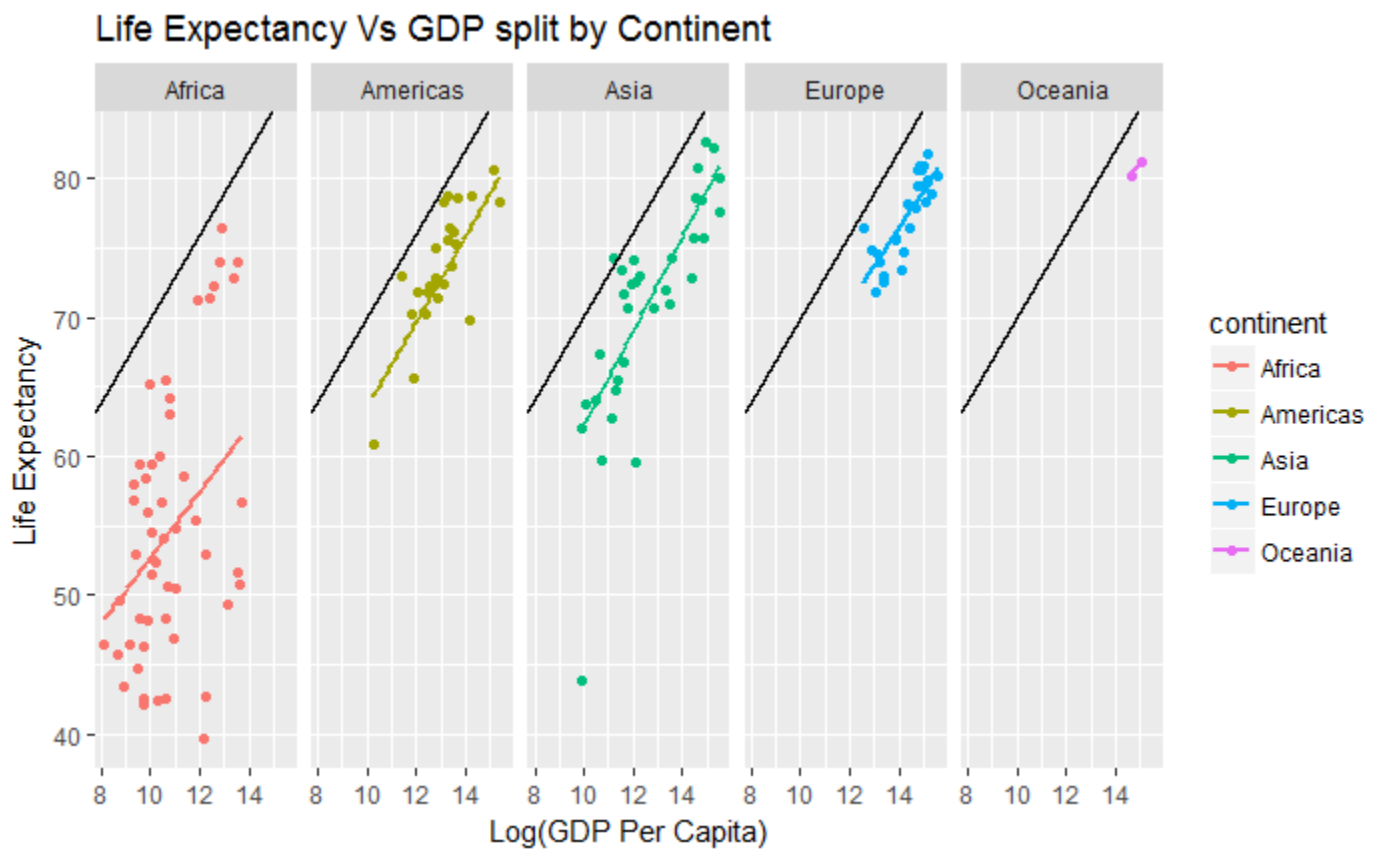


Figure 3: Life Expectancy vs GDP separated by Continent

## Problem 2

### Introduction:

The aim of this section is to find if there are any trends in Life expectancy over time for each continent, by the end of this section we should be able to explain the questions that the researcher has with regards to changes in Life Expectancy vs time.

### Strategy:

- 1) First, we shall create a data frame called "avgLifeExpectancy" which captures the essential data like weighted average of life Expectancy for each continent by year, year and continent.
- 2) Next, we shall plot average Life Expectancy vs time as per continent. (Figure 4)
- 3) To understand the anomalies in the trend for Asia and Africa, we draw plots for their most populated countries.

### Conclusions/Inferences:

- 1) Yes, the weighted average Life Expectancy is increasing overtime for each continent with almost a similar slope.
- 2) For the continents - America and Europe, the increase is almost identical and the life expectancy in general is very high in these 2 continents compared to the others. This could be due to the presence of highly developed countries with better health conditions.
- 3) Asia has the most rapid rate of increase (Figure 5). This can be explained by the rapidly developing countries like India, China, Singapore, Japan etc., with drastic changes in population and GDP. But there seems to be a dip in the growth around 1962, which when inspected was found to be caused by China. This is definitely because of the Great Chinese Famine between 1959 and 1961. And since we had taken weighted averages of the life expectancy China had a great impact on the overall trend for the continent.
- 4) The continent with the least life expectancy is Africa (Figure 6). This can be explained by the low standard of living. The growth in life expectancy seems to slow down after the 1990s which can be due to the high number of diseases becoming predominant in the continent. When we look at the country wise distribution, this stagnation in growth seems like a general phenomenon with exceptions like Egypt.

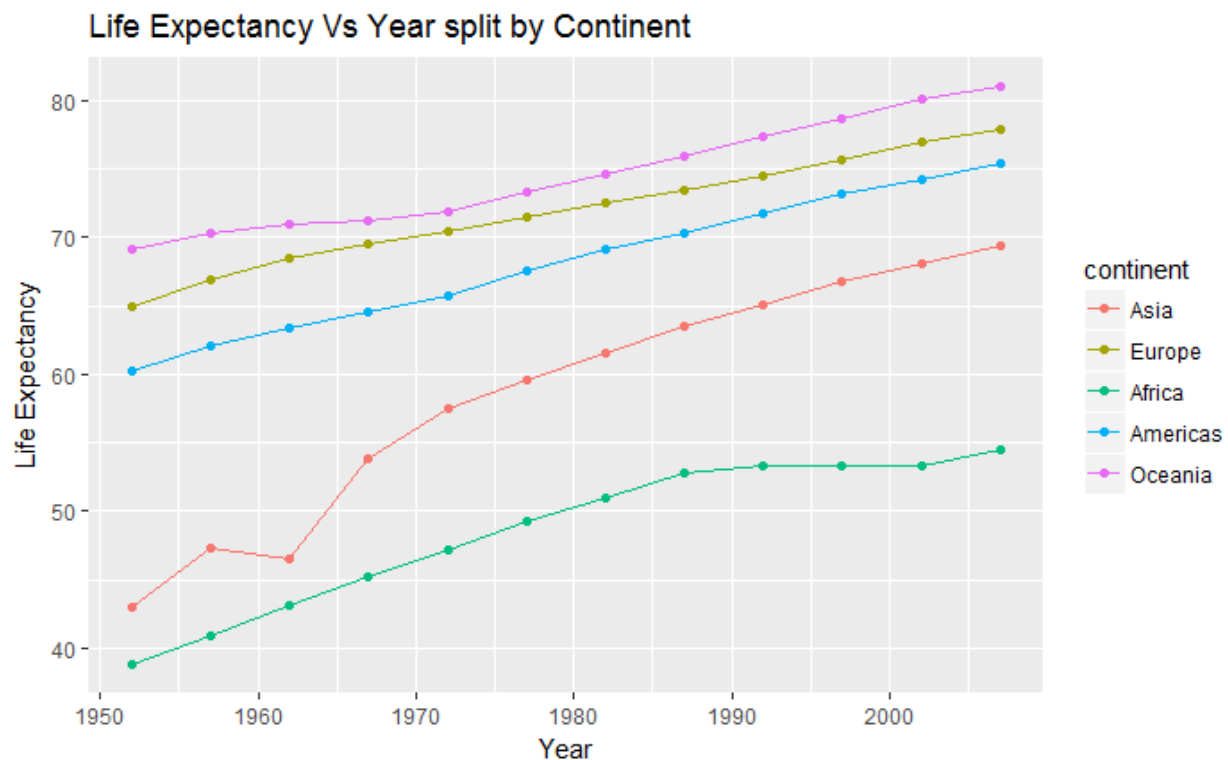


Figure 4: Life Expectancy vs Year separated by Continent

## Life Expectancy Vs Year in Asia split by the most populated countries

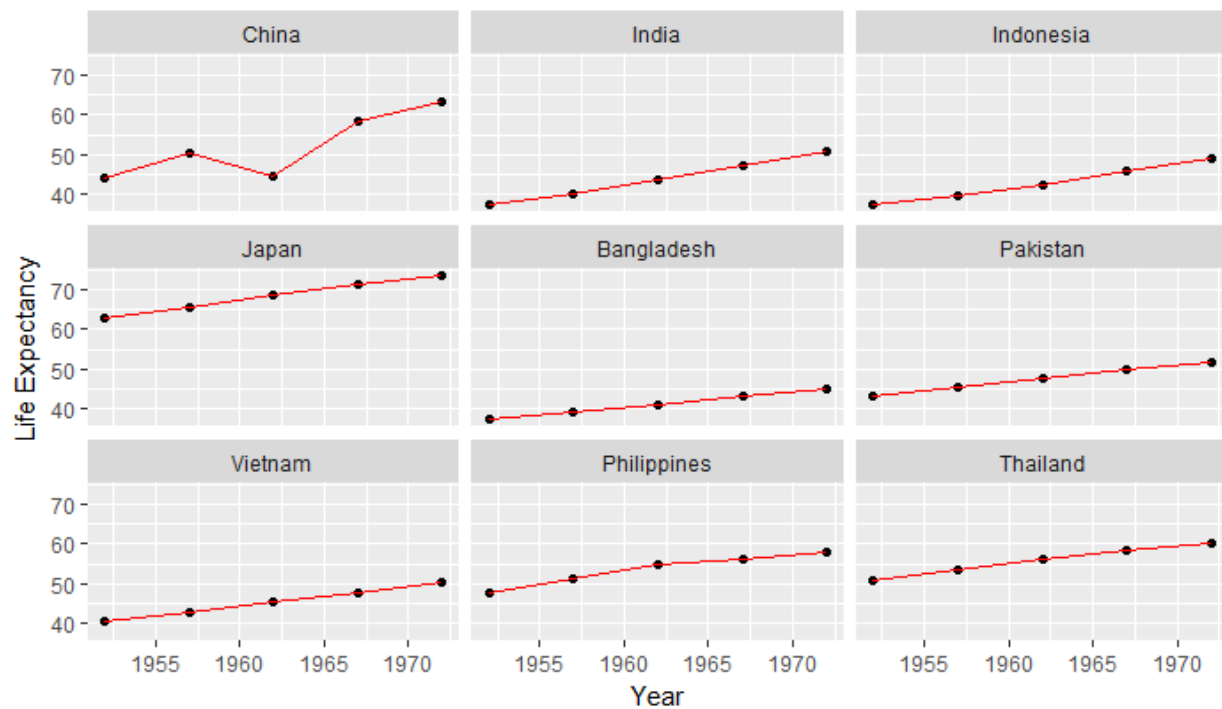


Figure 5: Asia – Life Expectancy vs Year

## Life Expectancy Vs Year in Africa split by the most populated countries

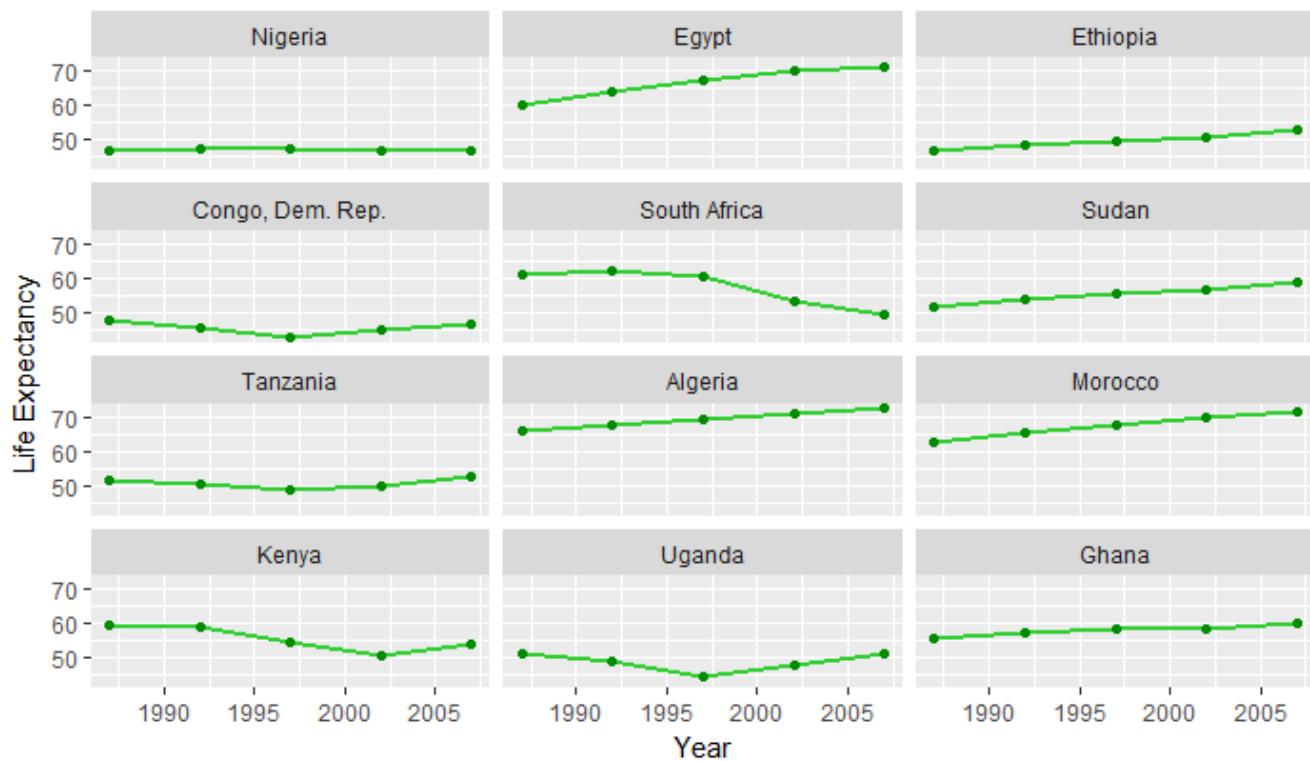


Figure 6: Africa – Life Expectancy vs Year

### Problem 3

#### Introduction:

In this section we aim to explore the relationship between Life Expectancy, GDP Per capita and time using trivariate analysis. We shall answer questions like does Life Expectancy only depend on GDP Per capita or time also? does the effect of GDP Per capita wear off? Are there any exceptions.

#### Strategy:

- 1) First, we use the weighted average of the life expectancy and the GDP and plot it by factoring it on the year.
- 2) Since the weighted averages were tricky to explain, we used the normal life expectancy values and the transformed GDP values and plotted a graph.

#### Conclusions/Inferences:

- 1) The first plot (Figure 7) looked very hard to explain but it has a lot of information. Each colored point indicates a year and looks like it follows an increasing trend in case of GDP with different slopes for different continents. Similarly, the life expectancy also increases over the time. But this doesn't fully explain the relationship between life expectancy, GDP and time.
- 2) When we look at the 2<sup>nd</sup> graph (Figure 8) everything becomes much clearer. Wherever there is an increase in the GDP, the life expectancy also seems to be increasing. But, we cannot say if time is also a definitive factor. Though as time passes, the GDP does seem to improve in almost all continents. So, we can indirectly say that time also has a part in effecting the life expectancy.
- 3) The slope for the individual continents are completely different during the beginning years but seem to converge to almost identical life expectancies towards the end.
- 4) The effect of the GDP per capita on life expectancy doesn't seem to wear off at all, in fact as you can see in case of Africa, as the GDP stopped growing, the life expectancy seems to have become stagnant as well.
- 5) Checking the slopes of each continent - in the beginning i.e., the 50s to 80s, Americas seems to have the highest slope indicating that with little increase in GDP, there seems to be a high increase in life expectancy. Asia's GDP vs life expectancy remained constant but has steadily increase to converge to a similar life expectancy as that of Europe and Americas. Africa had the highest changes with increasing and decreasing slopes.

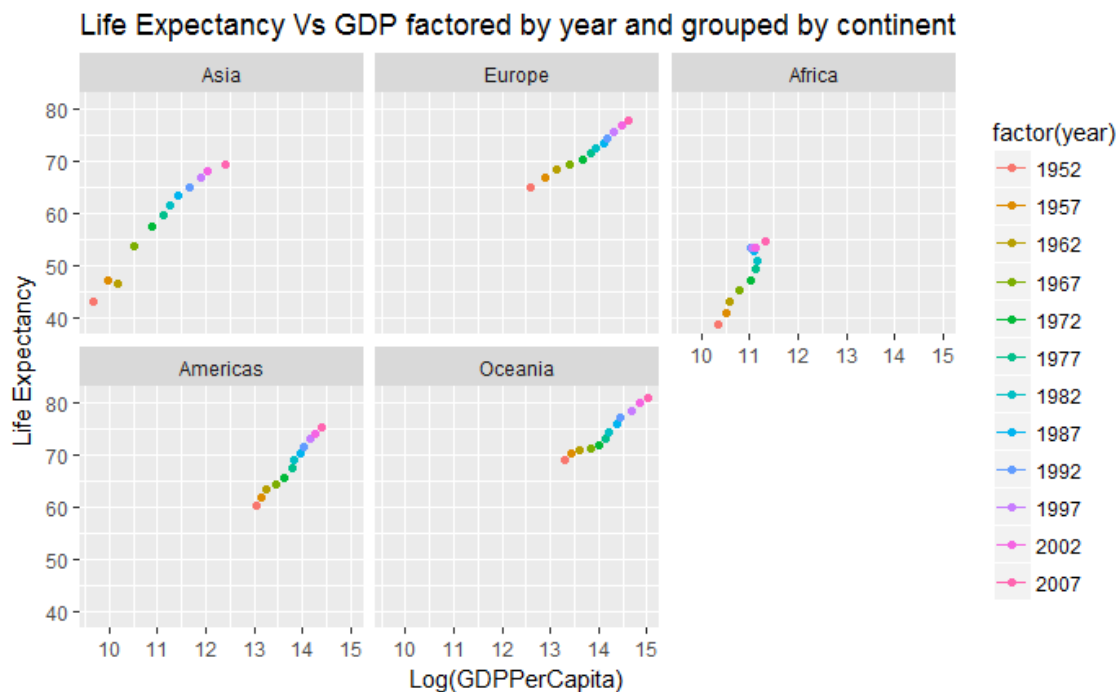


Figure 7: weighted Life expectancy vs weighted GDP factored by Year



Figure 8: Life Expectancy vs GDP factored by Year

Appendix

