Mini Project 1: Life Expectancy

Ashwed Patil, Vatsal Jatakia, Saniya Ambavanekar, Akshay Naik (Team Hong Kong)

February 23rd, 2018

Introduction

Using principles from Exploratory Data Analysis, this project investigates the relationship between life expectancy in 142 countries and GDP Per Capita through the years 1952 to 2007. Our primary research question is to see whether increase in life expectancy after World War II can be largely explained by increases in GDP Per capita. We explore this question using the **gapminder** dataset.

1. GDP and Life Expectancy in 2007

Life Expectancy by GDP Per Capita in 2007
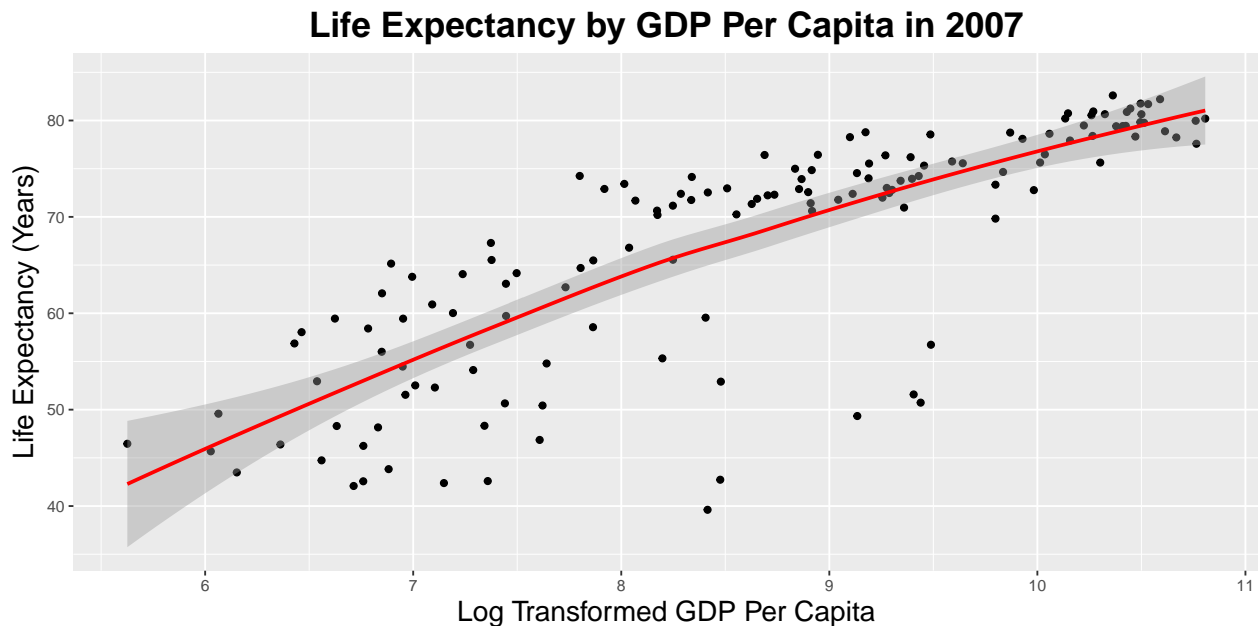
Because the GDP Per Capita data is significantly right skewed as evident from the density plot (see Appendix), we log transformed it. The above plot shows the relationship between life expectancy and log transformed GDP Per Capita in 2007. The data points represent the respective countries and the curve captures the general trend in the relationship. The curve seems to be fairly close to a straight line and hence it would be safe to assume that there is a linear relationship between life expectancy and GDP Per Capita.

We constructed a linear model which is specified by *Life Expectancy = GDP Per Capita + intercept.* The residual curve of the model (see Appendix for the residual plot) wiggles around zero and the R^2 came out to be 0.656 which indicates that the model captured 66% of the variance in life expectancy for the year 2007.

Thus, we can say that in 2007, people in richer countries lived longer than those who lived in poor countries.

Is the pattern same or different for every continent? Can the differences between continents be simply described by an additive or multiplicative shift, or is it more complicated than that?

The plot on the next page shows the relationship between life expectancy and GDP Per Capita for each continent in 2007. Each line describes the trend for the respective continent. The trend is similar for every continent, countries with high gdp per capita have higher life expectancy. However, Africa seems to lag

behind other continents significantly in terms of life expectancy as evident from the graph. From the graph, we can conclude the difference between Africa and other continents can be roughly described by an additive shift.

Life Expectancy Vs GDP Per Capita By Continent
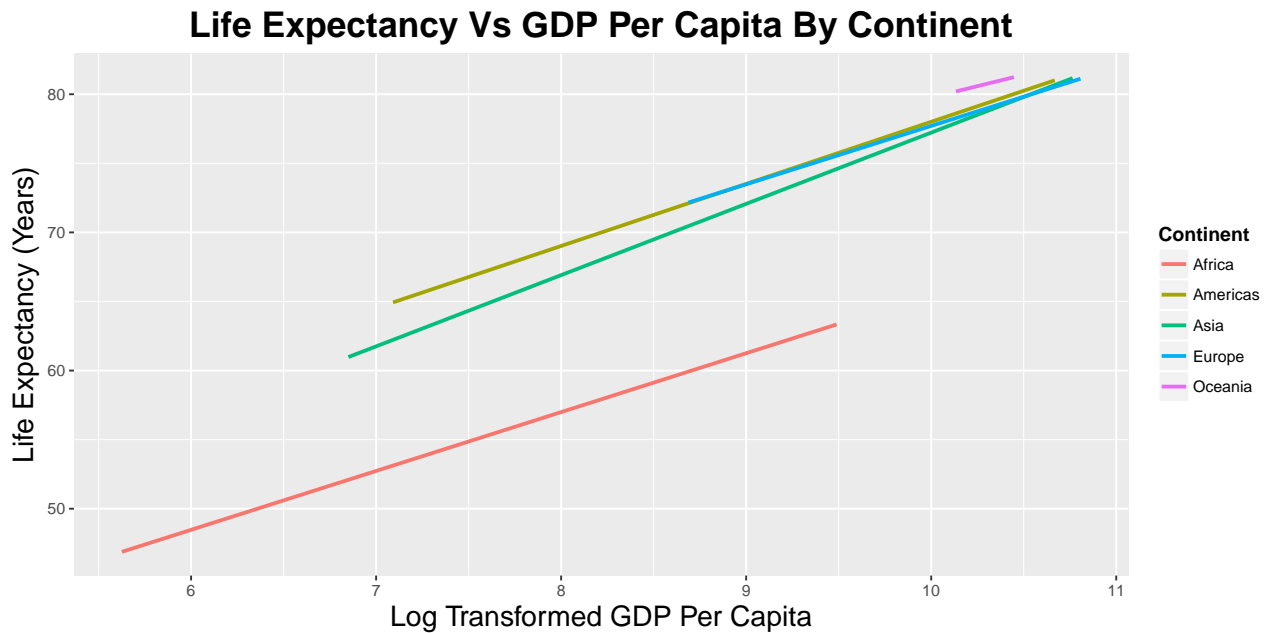
2. Life Expectancy Over Time By Continent

Life Expectancy Change over time in each continent
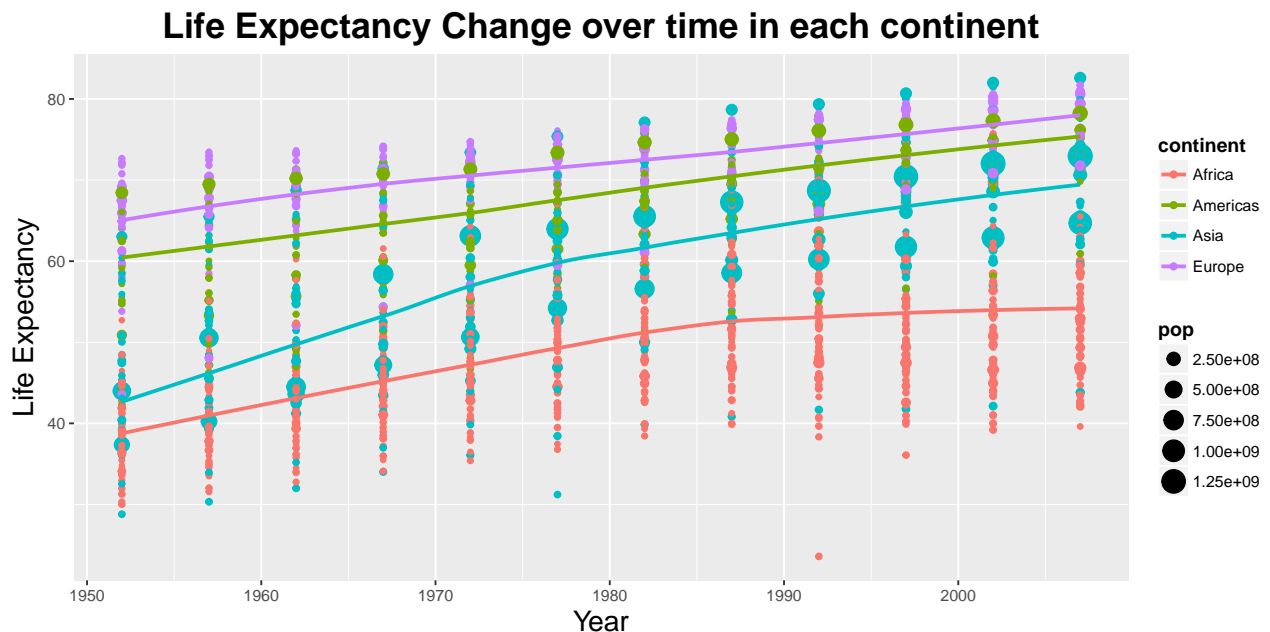
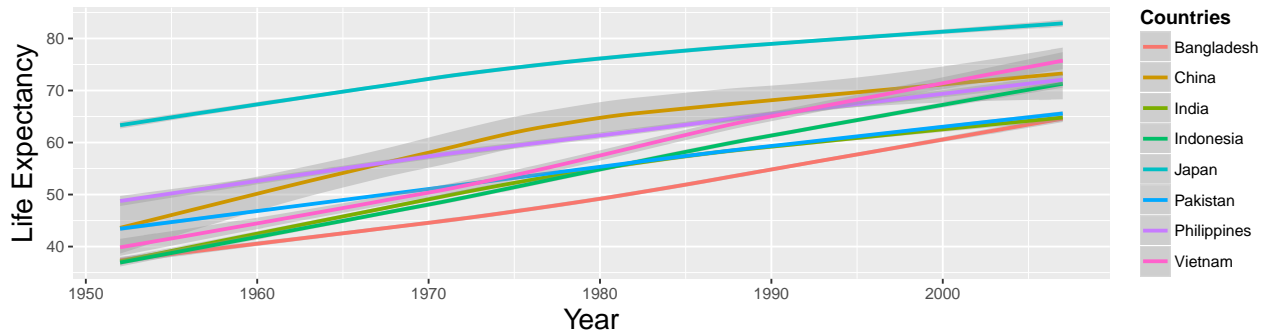From the above graph we can see that Asia's overall life expectancy has increased significantly from around 46 to 72 over the period 1950-2007.Alongside, Americas and Africa have also shown significant improvement in their life expectancies. We can observe that incase of Asia the increase in trend is mainly due to the countries with very large population.

Let us inspect the life expectancy growth in Asia in more depth.Asia is noted for its size and huge population, but this is characterized by some countries with a very high population while many of them being sparsely

populated.Hence we decided to have a population threshold (50 million) and observe life expectancies of only those countries with very high population as these are ones which will have the most interesting observations for life expectancy.

Life Expectancy of Countries in Asia

From the above graph we can see that there are 8 countries that are above our population threshold and all of the have shown a significant increase in their life expectancies.

## Changes in the Life Expectancy of the Continents
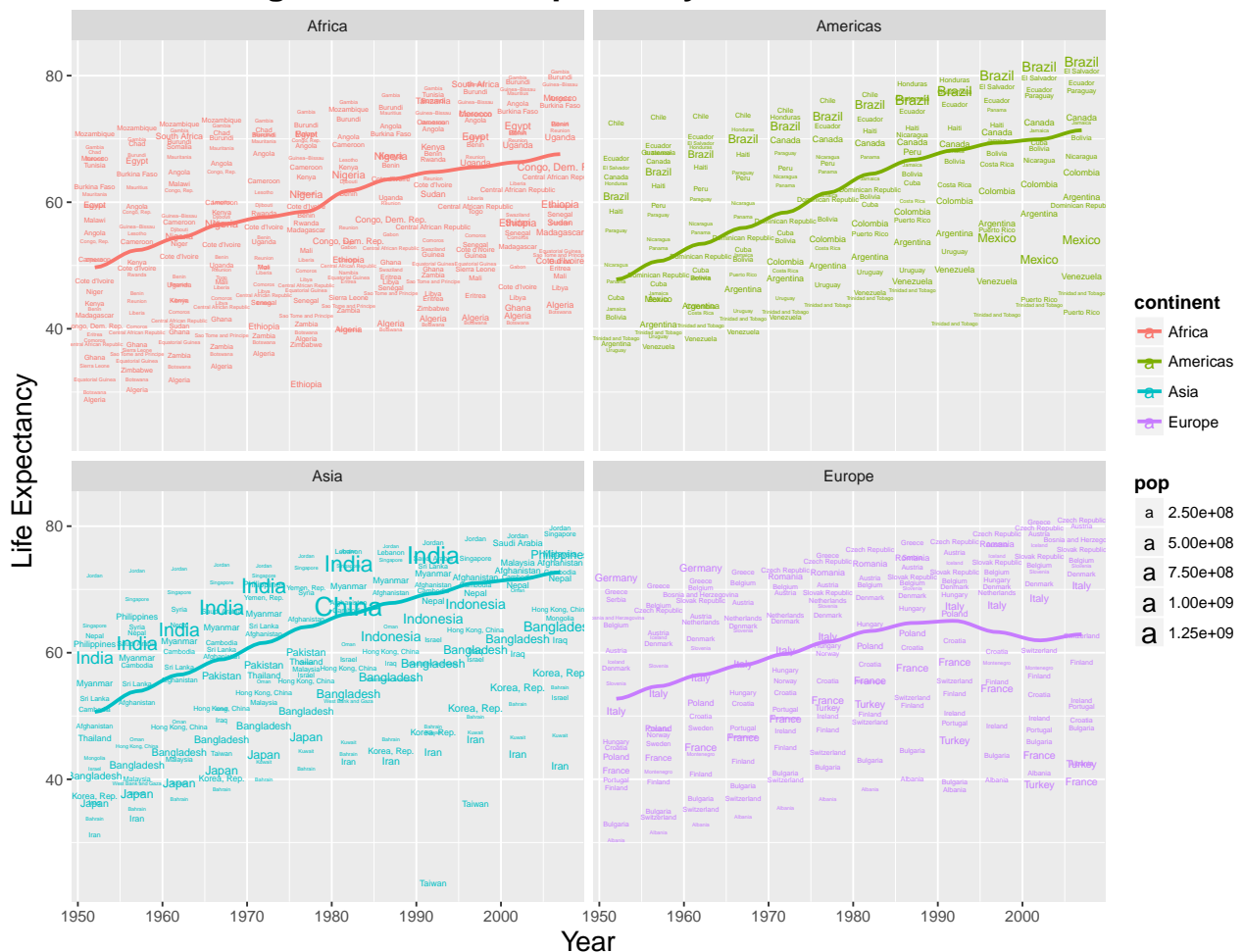
From the above plots we can observe that the life expectancies of all the continents have shown a somewhat linear increase over time. We can observe that the Asia and America are having almost similar growth in Life Expectancy. As well as the increase in Life Expectancy of Europe decreases and slowly grows in between 1990 and 2000. In Asia and America we can see that majorly the increase in life expectancy is because of the

countries with very large popluation.

3. Changes in the relationship between GDP and Life Expectancy Over Time

Life Expectancy Vs. GDP Per Capita over time along different continents
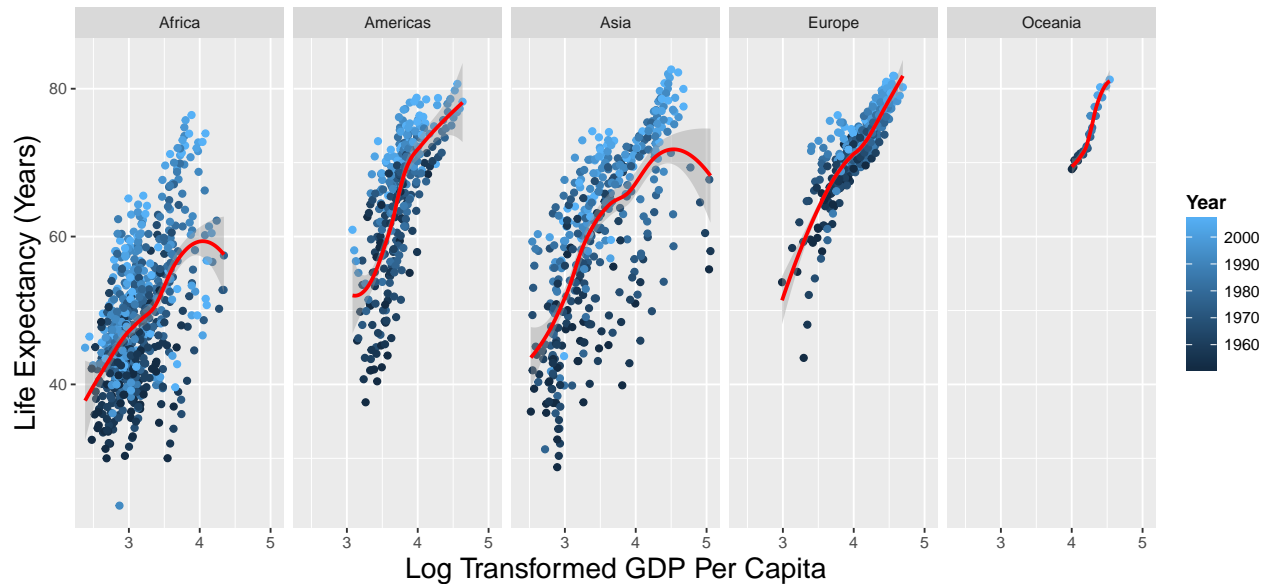
Observing the above plot, we observe a similar trend in case of all the continents for life expectancy and log transformed GDP Per Capita. Further, the increasing trend also suggests that the there is a positive relationship between these two variables except in the case of Asia and Africa where it increases upto a certain point and then decreases. But the overall trend is quite similar across all the continents. If we see the legend, as the time increases, the points become darker. In the plot, as life expectancy increases, there is a transition of shade from dark to light, which indicates the presence of some kind of a time effect. But, we need to investigate more in order to reach any conclusion.

To check the time effect, we evaluated 2 models, one excluding and the other including the year variable. We observe that on adding the year, we see a significant improvement of almost 10% in the R^2 value of the model. Hence, there is some time effect in the relationship between the 2 variables.

We will now try to check how much of the variation in lifeExp can be explained by GDP alone. To check this, we fit a loess model for all the continents.

```
##   Continent  R_square
## 1    Africa 0.2860303
## 2      Asia 0.4753904
## 3  Americas 0.5042927
## 4    Europe 0.7265084
```

The table above shows that for each continent the r-square is between 25-70% which suggests that GDP per capita alone is not enough to explain all of the variations in life expectancy. However adding the year pushes the r-square range to 45-80%, but this is again not enough to explain the variance of the life expectancy entirely.

The next step in this investigation is to check for any convergence. To do so, we will make the following plot.

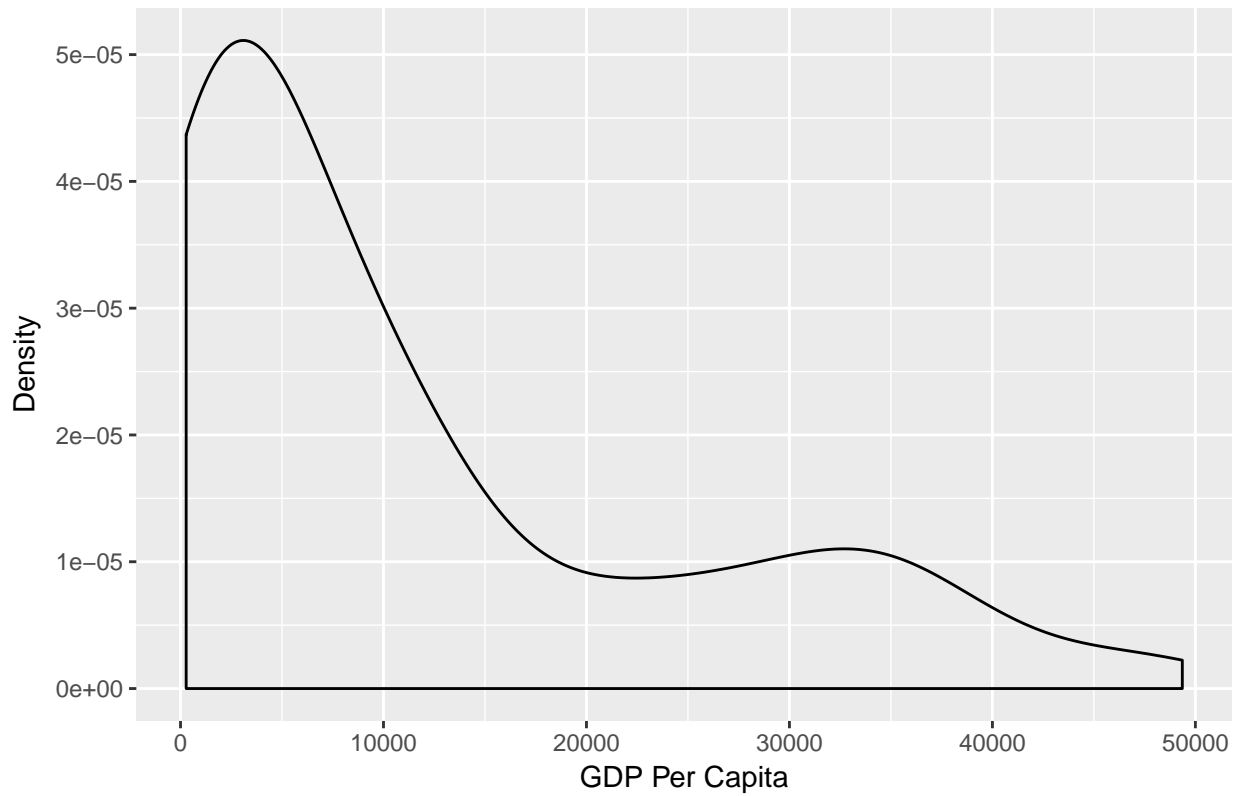## Life Expectancy Vs. GDP Per Capita over time for different continents
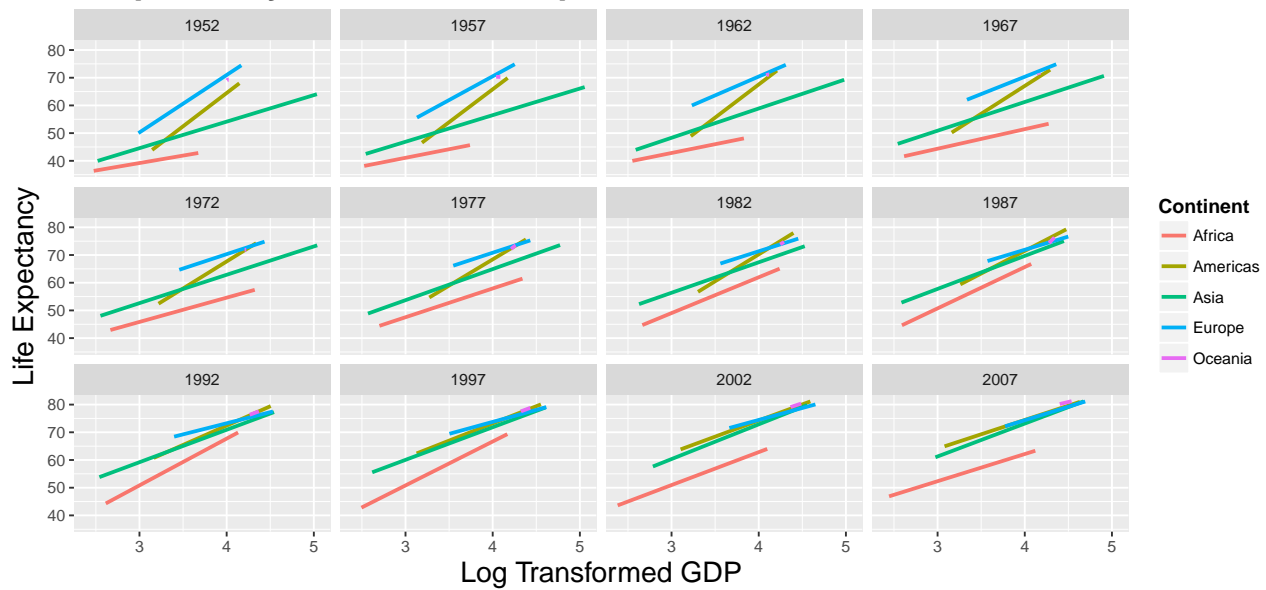
In the above plot, we observe that initally in the 1950's to 1960's the slopes of trend lines for all the continents in quite different, but as time is progressing, the trend lines for all the other continents except Africa are coming together and aligning themselves together giving us a convergence effect. These lines after convergence are fairly parallel to the line for Africa, which indicates that Africa has lagged behind in terms of other continent to improve it's life expectancy.

Conclusion

From our analysis of the gapminder dataset, we concluded that GDP Per capita cannot be used as the sole parameter to explain changes in life expectancy. We observed that life expectancy varies farily linearly with gdp per capita for all the continents. There is an additive shift for each continent over years which is because of the time effect on life expectancy. Except Africa, the Life Expectanct vs. GDP relationship converges for all other continents as time progresses while African nations have been still lagging behind others in terms of improving their life expectancies.

Appendix

Density Plot of GDP Per Capita

Residual Plot