Life Expectancy

Apurva Gupta,Shailendra Patil,Surbhi Paithankar

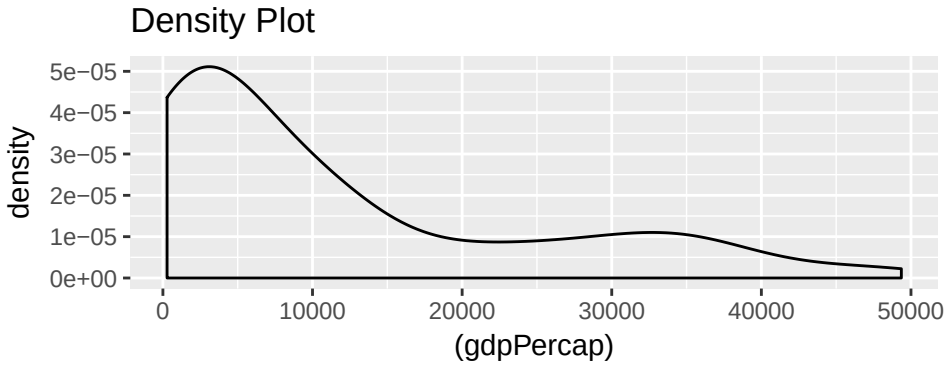
INTRODUCTION

In this project, we want to know if the increase in life expectancy since world war 2 can be explained by increases in the GDP per capita.The basic intuition behind this question is the post war economy. The United States, Soviet Union, Western European and East Asian countries in particular experienced unusually high and sustained growth. The economy had rejuvenated after the great depression during the inter-war period. ThinkTank wants to analyze if this increase in GDP can be attributed to increase in life expectancy. For answering this, we analyse the different types of relationships with respect to time and continents.

Question 1: GDP and Life Expectancy in 2007

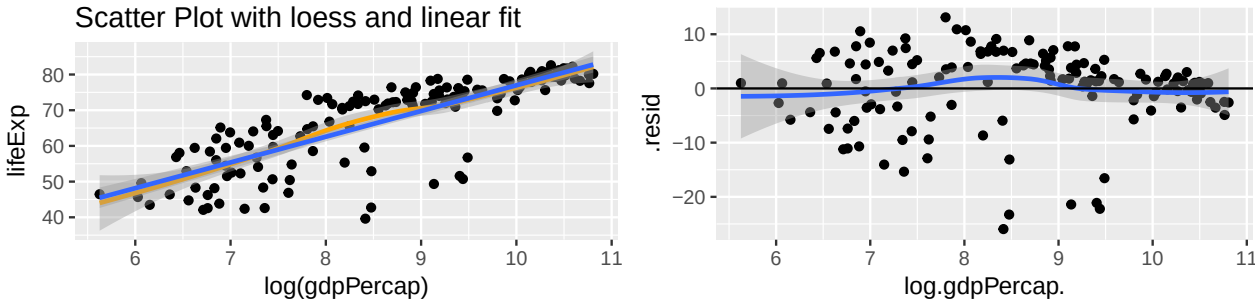
Let us analyze the data for 2007

Let us check the distribution for gdpPercap in 2007 before trying out the scatter plot



From the above graph it is evident that the data is right skewed and hence we can apply log transformation on data.

Now let us check the scatter plot for transformed data usin both default smoothing and linear smoothing



The smooth curve almost looks like a linear relation, so we can say that there is a linear relation between the transformed gdpPercap and lifeExp in 2007. Most of the residuals are scattered around zero, with few exceptions being very large negative values. But those can be ignored considering the fact majority of points are scattered well around zero. And the horizontal line is within the confidence band as well. Hence in 2007 the linear fit is a good approximations, we can say the transformed log gdp has a near linear relation with life expectancy.

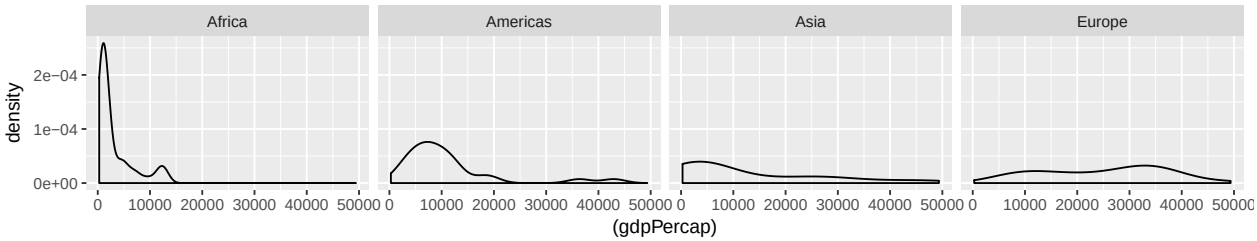
Checking the trend for different continents

Checking the summary statistic for continents

```
##    Africa Americas      Asia    Europe   Oceania
##        52       25        33        30         2
```

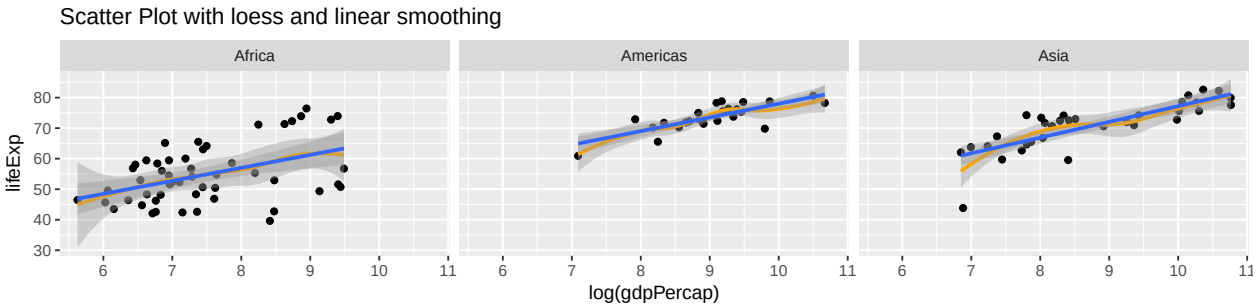
We can see that for 2007 Oceania has just 2 data points , hence we cannot find any relation between gdpPercap and lifeExp in 2007 for Ocenia, for this particular analysis we will ignore Ocenia.Remaining continents have a good amount of data for 2007, so we will check the trend for these continents.

Let us check the gdpPercap distribution for each continent



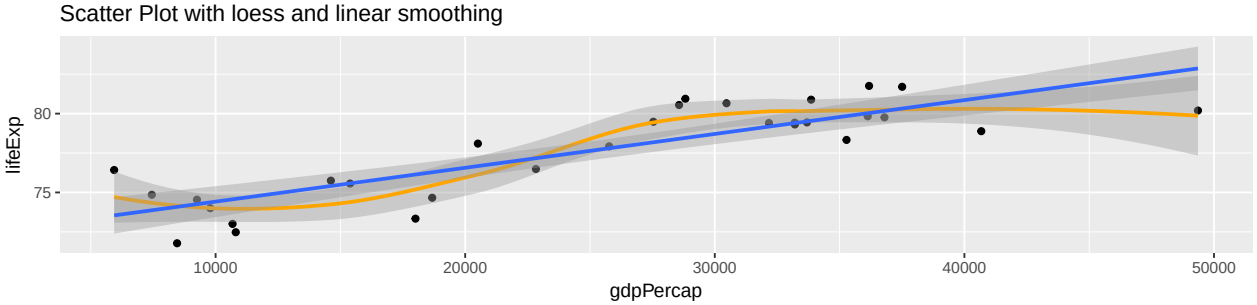
For Africa, Americas, Asia the data is right skewed , hence we will log transform the data for just these continents

Checking scatter plots for Africa, Americas and Asia



Here ignoring few points from Asia where there is a clutter between 8 and 9 and few points away from the linear fit We can say that for Asia and America there is a linear relationship between transformed gdpPercap and lifeExp. And for Africa the relation is not exactly linear. But as we just want to know the trend and not predict values, we can assume it to be linear trend. For Africa, Americas and Asia, it looks like the linear trends have almost similar slope and hence we can say that it close to a additive shift

Checking scatter plots for Europe

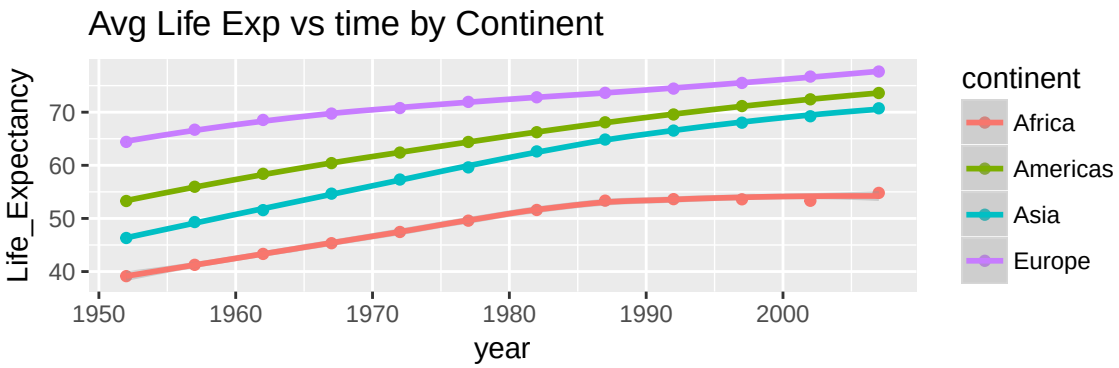


It looks like the "loess" is a better fit in the above graph. Loess is a form of local polynomial regression,

meaning that at every xx-value, it fits a weighted polynomial model "locally": data at nearby xx-values will be weighted heavily, while data at far away xx-value will be downweighted or not considered at all if they fall outside a "neighborhood. But for Europe it something complicated and its not a simple additive or multiplicative shift

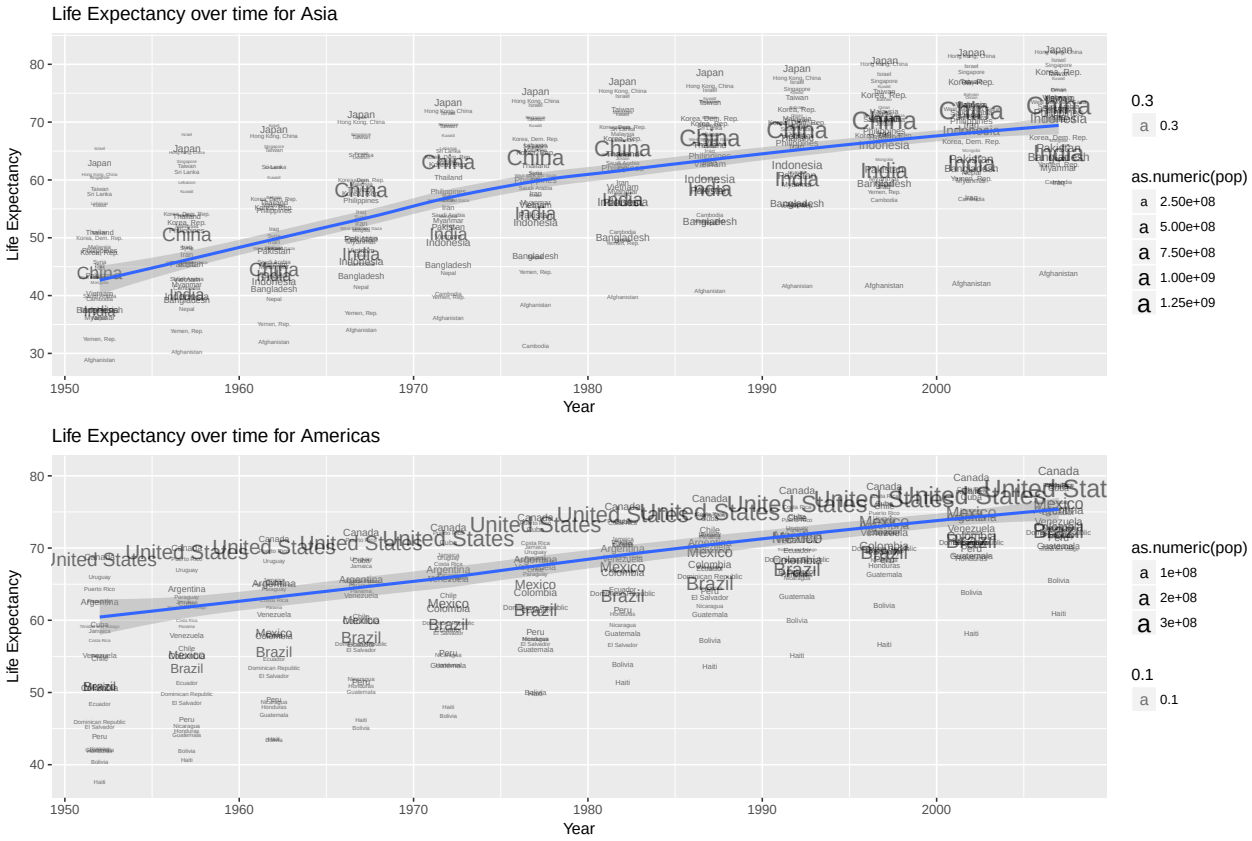
Question 2: Life expectancy over time by continent:

When analyzing average life expectancy for each continent , we need to perform the weighed averaging. This is because the population of countries in every continent varies significantly.



From the above graph, we can observe that the life expectancy for all the continents looks to be increasing.In the period of mid 1950s to 1980s ,Asia and America looks to be catching up with the continent with higher life expectancy, Europe. On the other hand, the life expectancy of Africa increased till mid 1980s, however it converges after that.

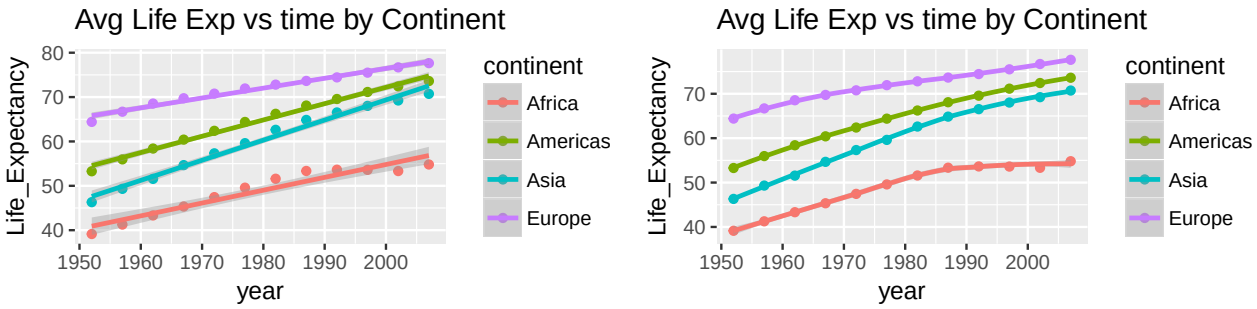
Now lets analyze the countries that contribute to the growth in life expectancy for Asia and Americas.



From the above graphs, we observe that China & India are countries that are significantly contributing to the increase in the life expectancy for Asia. Similarly US and Brazil are prime contributors for Americas. This is attributed to the fact that the population of these countries have been dramatically increasing over the years.

Fitting a Model

Now lets try to see how a linear models fits the data compared to the default loess model.

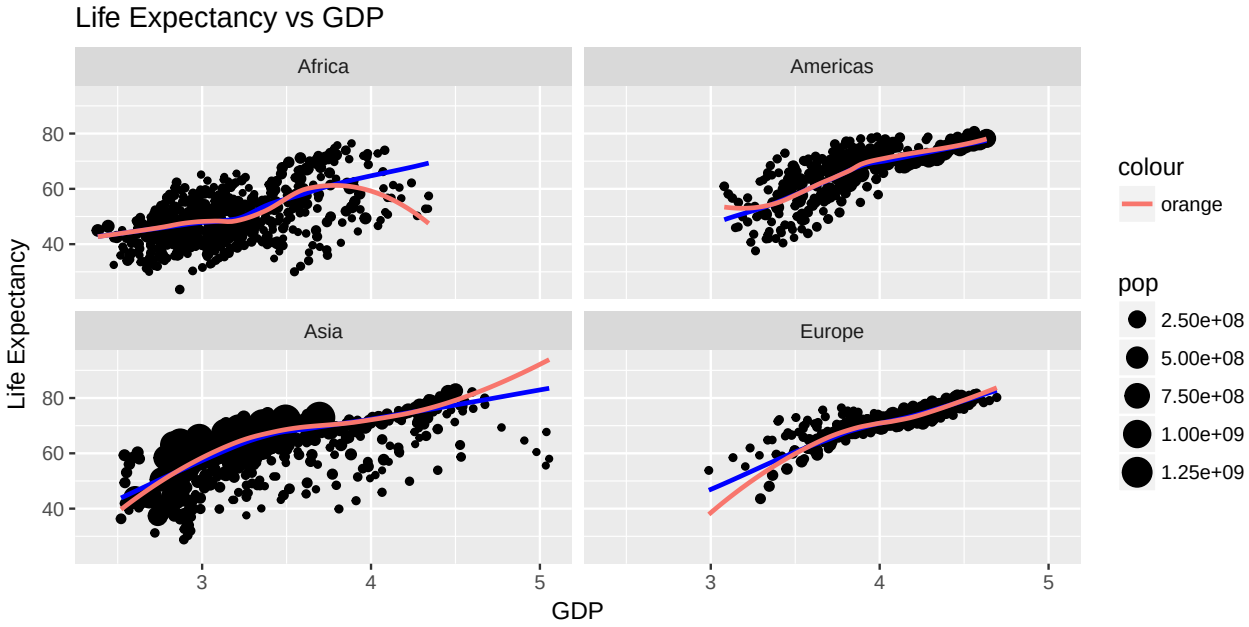


In the above graphs we observe that the linear model fits the data well for continents having steady upward trend. However for the continent like Africa and Asia where the growth gets slower after certain point is not captured well by the linear model. Hence loess model looks to convey the information better regarding the periods of change. Additionally we can observe that rate of change in life expectancy for Africa has slowed down from mid 0f 1980's onwards.This is not explained in the given data as to why there is such a slow down and this can be attributed to external factors. Further investigation revealed, in the 1970s decade the true HIV/AIDS epidemic broke out in Africa. This might be one of the reasons for the slow change.

Question 3: Changes in relationship between GDP and life expectancy over time

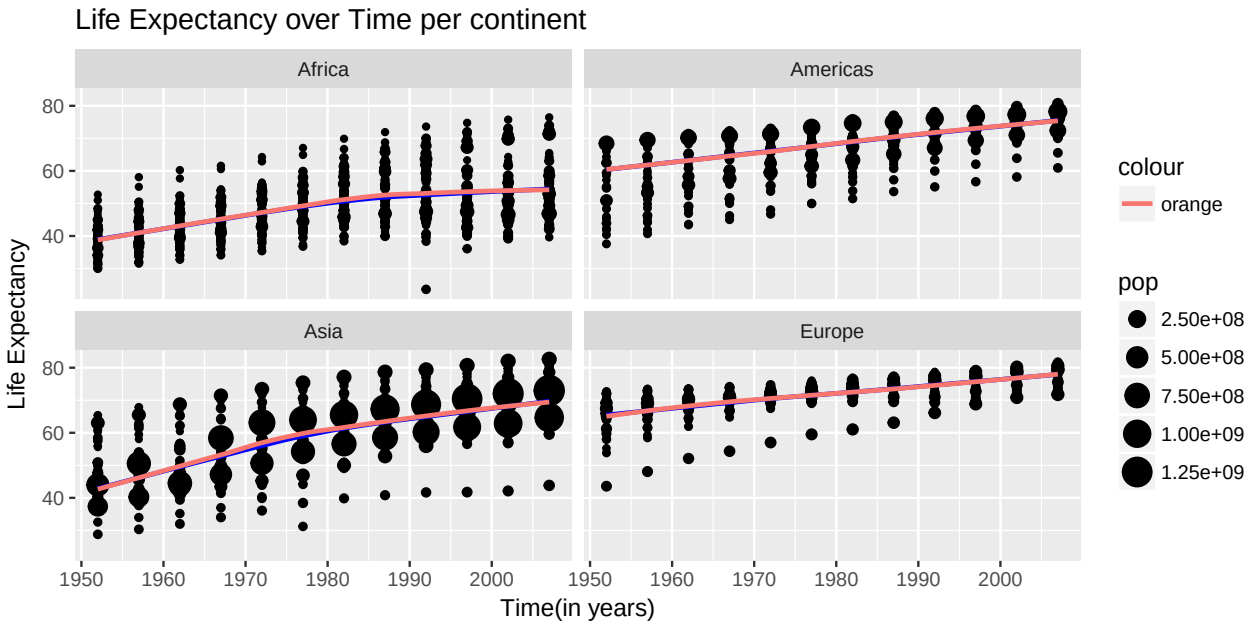
As we have seen before GDP is right skewed , hence for our analysis we will use log transformed value of GDP.

Relationship between GDP and life Expectancy



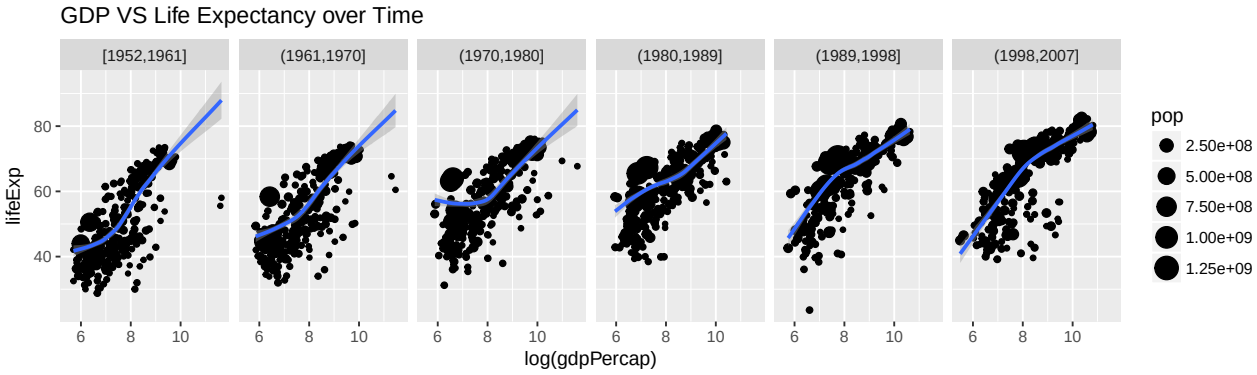
From the above graph, we can observe that there is definitely a trend between GDP and Life expectancy. As GDP increases,life expectancy also increases.Blue line gives a better fit i.e. degree 1 loess fit is better than the default degree 2 fit.

Relationship between Life Expectancy and Time



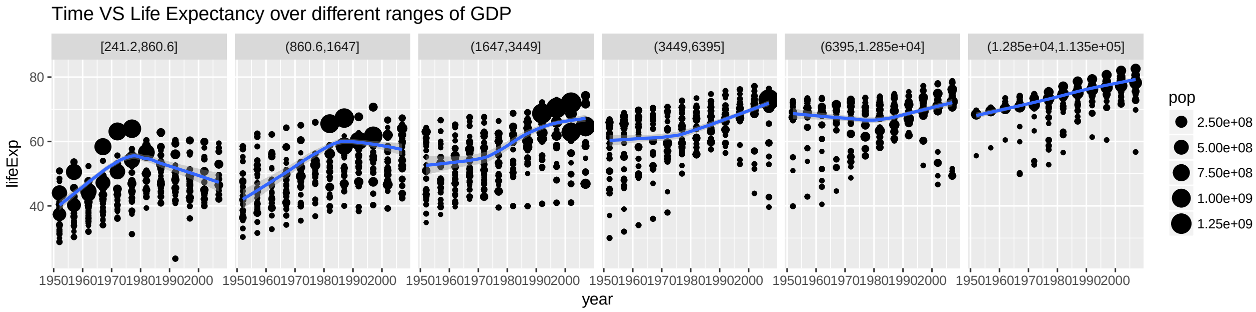
We can observe a pattern in Life expectancy vs Time. Over the Time, Life expectancy increases gradually for all the continents.From 1980s onwards,the rate of increase in life expectancy for Africa is less than the rate of increase for other continents.

Let us analyze the relationship between GDP and Life expectancy for different time frames.



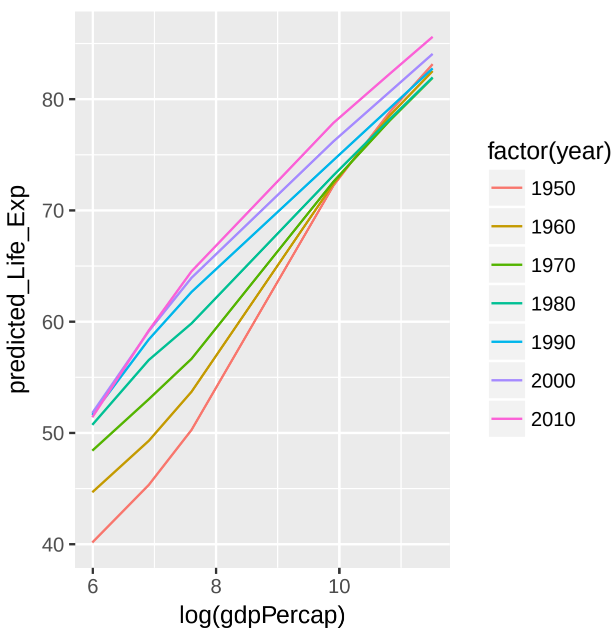
From the above graph, we can see that for initial years(before 1980), Life expectancy increases rapidly. After 1980, the trend seems to slow down.

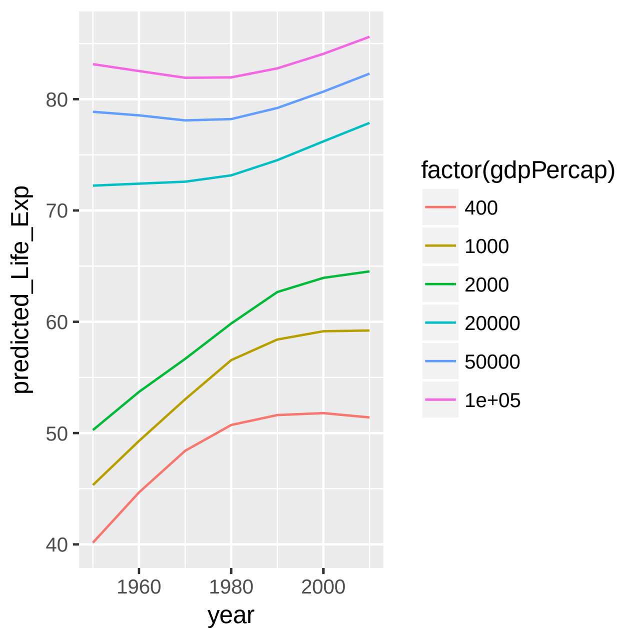
Let us analyze the relationship between Time and Life expectancy over different ranges of GDP.



There seems to be an interesting trend here. For countries with lower GDPs', till 1980, Life expectancy increases rapidly.After that period, life expectancy decreases rapdily. As we progress towards the countries with higher GDP, trend seems to stabilize and time doesnot make a huge impact on Life Expectancy. Also, we can observe that there is no additive shift between different time frames and different GDP ranges.This indicates that fitting a model with interactions will be safer.

Fitting Model





The first graph explains us that as GDP increases, life expectancy increases. But after a certain point of time, Time does not have any effect on Life expectancy. The second graph tells us that for lower GDP countries, as time progresses, life expectancy seems to stagnate i.e. there is not much increase in life expectancy over the time. But, for higher GDP countries we observe a different trend. Life expectancy keeps on increasing over the time.

Conclusion

From all the data analysis we can observe that there is a steady increase in the life expectancy over all the continents except Africa(post 1980).For countries with lower GDP,we observe that life expectancy doesn't increase beyond a certain time .But for higher GDP countries,such effect is not observed and the life expectancy increases over the time.