

# Life Expectancy

*Apurva Gupta, Shailendra Patil, Surbhi Paithankar*

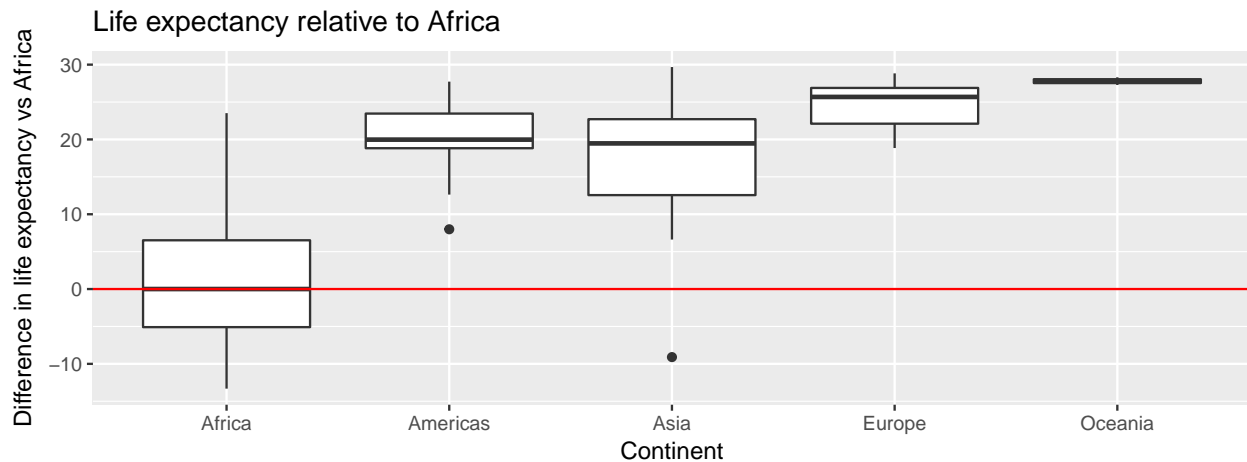
## INTRODUCTION

In this project, ThinkTank wants to analyse 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 life expectancy can be attributed to rise in GDP. For answering this, we will analyse the different types of relationships with respect to time and continents.

## Question 1: GDP and Life Expectancy in 2007

##	continent	Mean_LifeExp	Median_LifeExp
## 1	Africa	54.80604	52.9265
## 2	Americas	73.60812	72.8990
## 3	Asia	70.72848	72.3960
## 4	Europe	77.64860	78.6085
## 5	Oceania	80.71950	80.7195

We can observe that the mean and median life expectancy of Africa is less when compared to other continents.

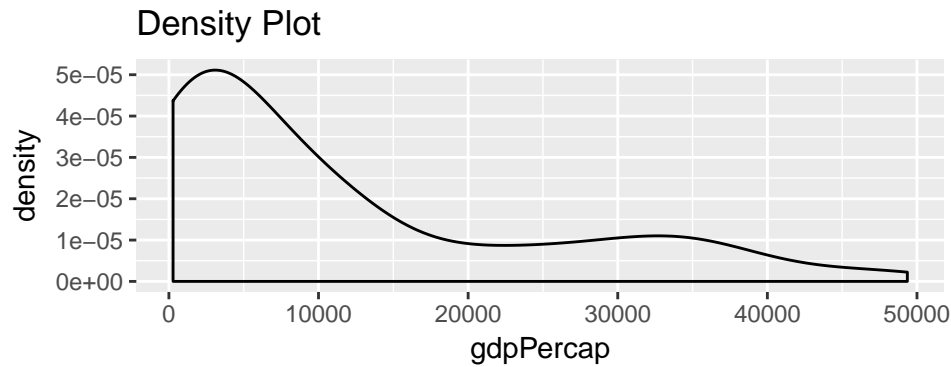


Let us now compare the distribution of life expectancy for different continents. Oceania has the highest mean and median life expectancy. But as Oceania has only 2 countries, we can ignore it from our analysis. Therefore, Europe has the highest mean and median life expectancy of 77.64860 and 78.6085 respectively.

The median life expectancy for Americas and Asia is 20 years greater than Africa. Europe has the highest median life expectancy around 25 years greater. As Oceania has just 2 countries, comparing the life expectancy wouldn't be accurate.

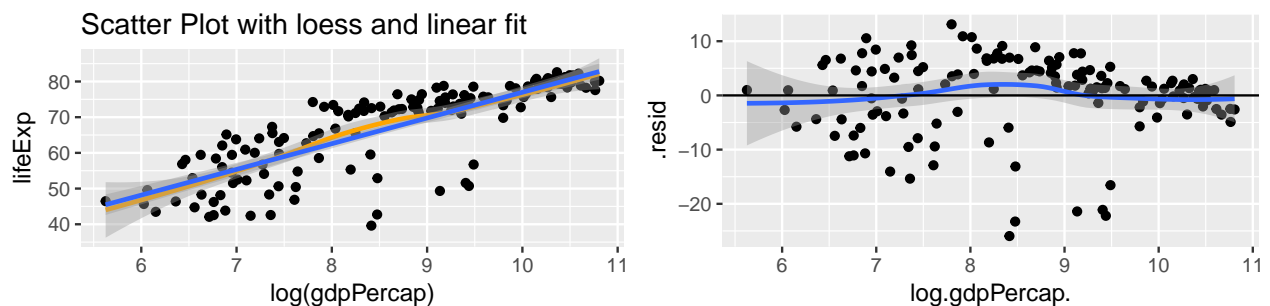
## Fitting the model

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 to the data.

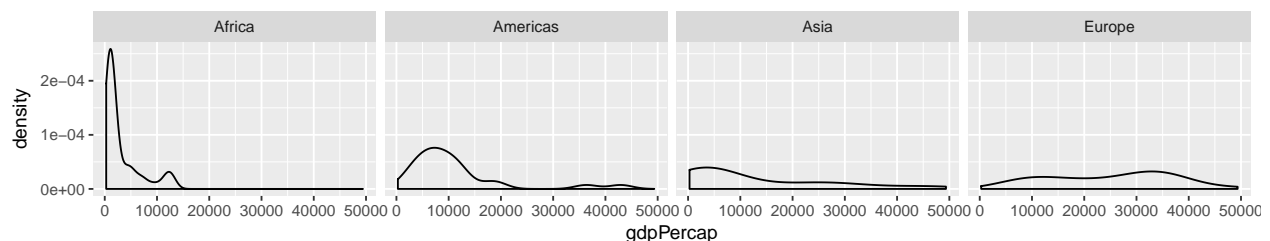
Now let us check the scatter plot for transformed data using both default smoothing and linear smoothing



```
##      (Intercept) log(gdpPercap)
##          4.949612      7.202802
```

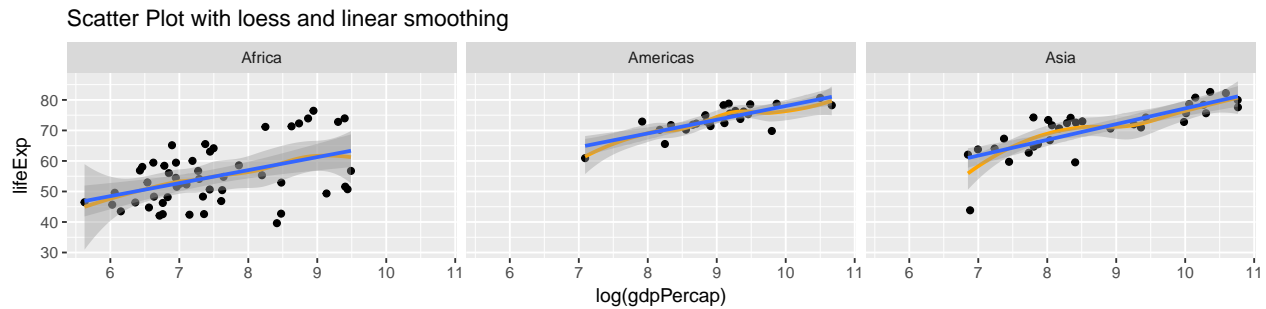
The smooth curve almost looks like a linear relation, so we can say that there is a linear relation between the transformed  $\text{gdpPercap}$  and  $\text{lifeExp}$  in 2007. For every unit increase in  $\log(\text{gdpPercent})$ , the life expectancy increases by 7.20 years approximately. Looking at the residuals, we see that they are scattered around zero, with few exceptions being very large negative values. But those can be ignored considering the fact that majority of points are scattered well around zero & the horizontal line is within the confidence band as well. Hence in 2007 the linear fit is a good approximation, indicating that the transformed  $\log \text{gdp}$  has a near linear relation with life expectancy.

Let us check the  $\text{gdpPercap}$  distribution for each continent



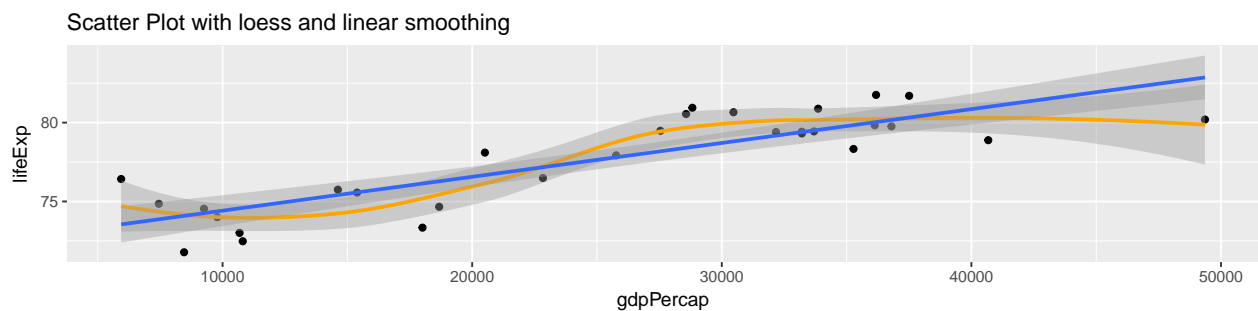
For Africa, Americas, Asia the data is right skewed, hence we will log transform the data for just these continents

Checking scatter plots for Africa, Americas and Asia



As seen from above graphs, we can say that there is a linear relationship between Life Expectancy and tranformed GDP distribution. For Africa, Americas and Asia, it looks like the linear trends have almost similar slope and hence we can say that it is close to an additive shift.

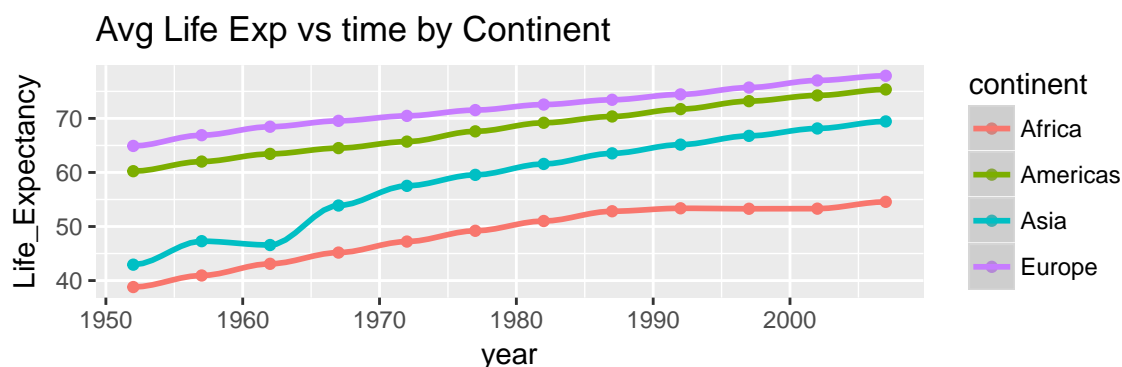
### Checking scatter plots for Europe



It looks like the “loess” is a better fit in the above graph for Europe. 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. Hence for Europe it is something complicated and its not a simple additive or multiplicative shift.

### Question 2: Life expectancy over time by continent:

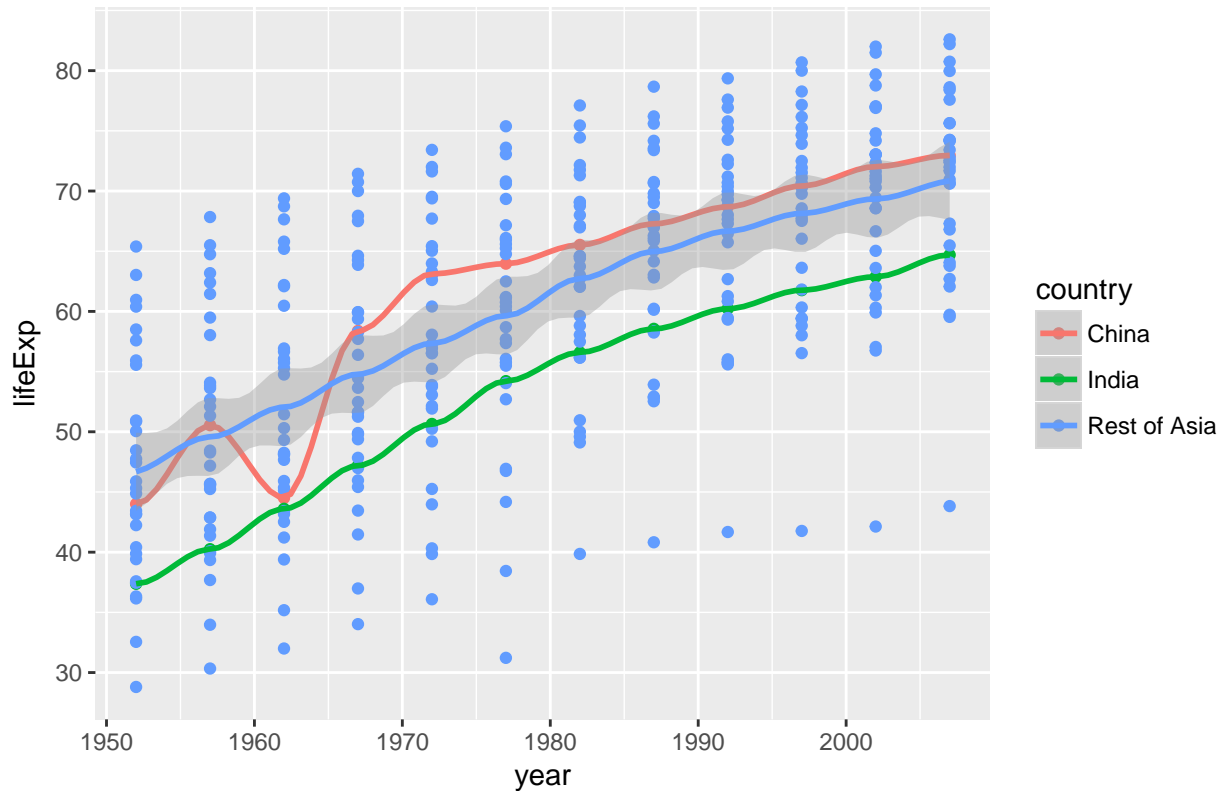
When analyzing average life expectancy for each continent , we need to perform the weighed averaging. This is because the population of countries in every continent varies significantly.



From the above graph, we can observe that the life expectancy of Americas and Europe increases over time. For asia around 1962 there is a sudden dip in the life expectancy and after that the life expectancy has increased over time. For Africa, the life expectancy increase over time, but after 1980 we observe that the increase in life expectancy is less compared to before. Now lets go ahead and check for Asia separately to see

what might be affecting the life expectancy. Additionally, let us take China and India separately due to their higher populations.

### Avg Life Exp vs time for Rest of Asia compared to Indian and China

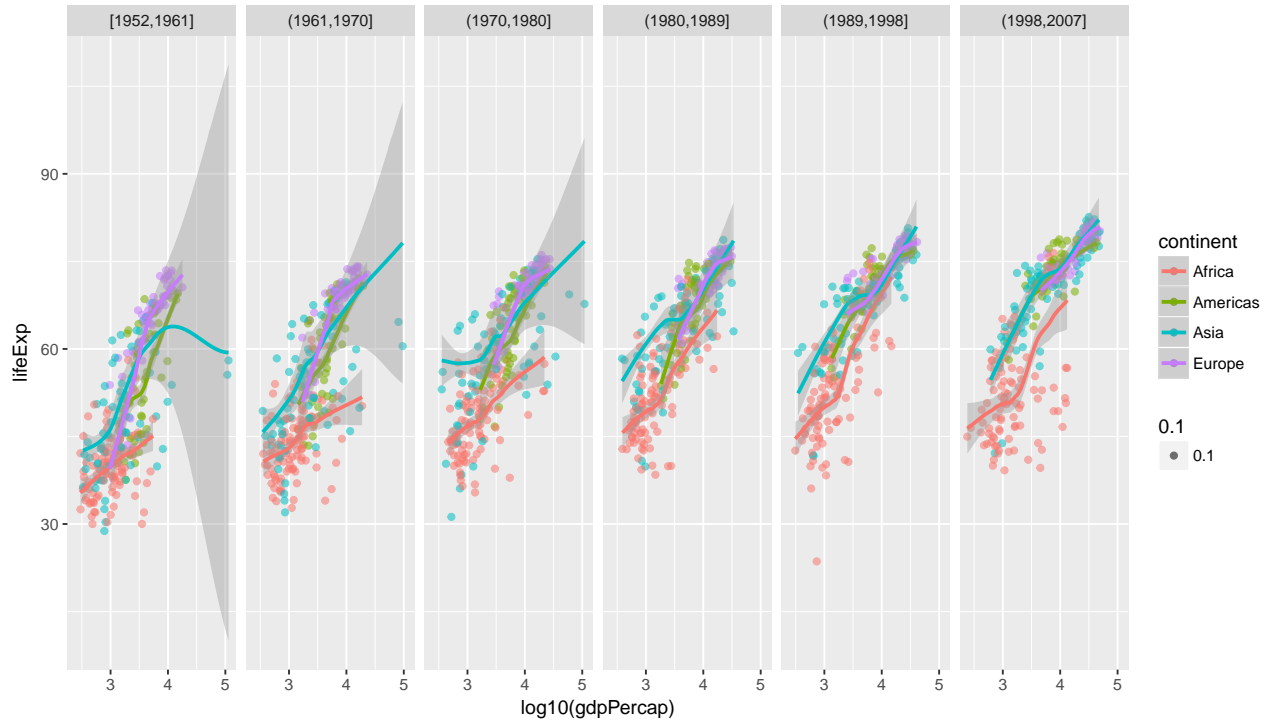


We observe that the overall trend in Asia is affected to a greater extent because of China. We can see a dip in life expectancy somewhere around 1962. So this affect can also be seen in the Average Life expectancy of Asia suggesting China has greater affect on average life expectancy of Asia. There is no reason why there was decrease for China in the given data , there can be external factors contributing to this. Further investigation revealed that from 1958-1961 China had the greatest famine which is popularly known as “Great Chinese Famine” which resulted in the decline of life expectancy significantly. Additionally we can also observe that rate of change in life expectancy for Africa had slowed down from mid of 1980’s onwards. This is not explained in the given data as to why there is such a slow down and this can also 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.

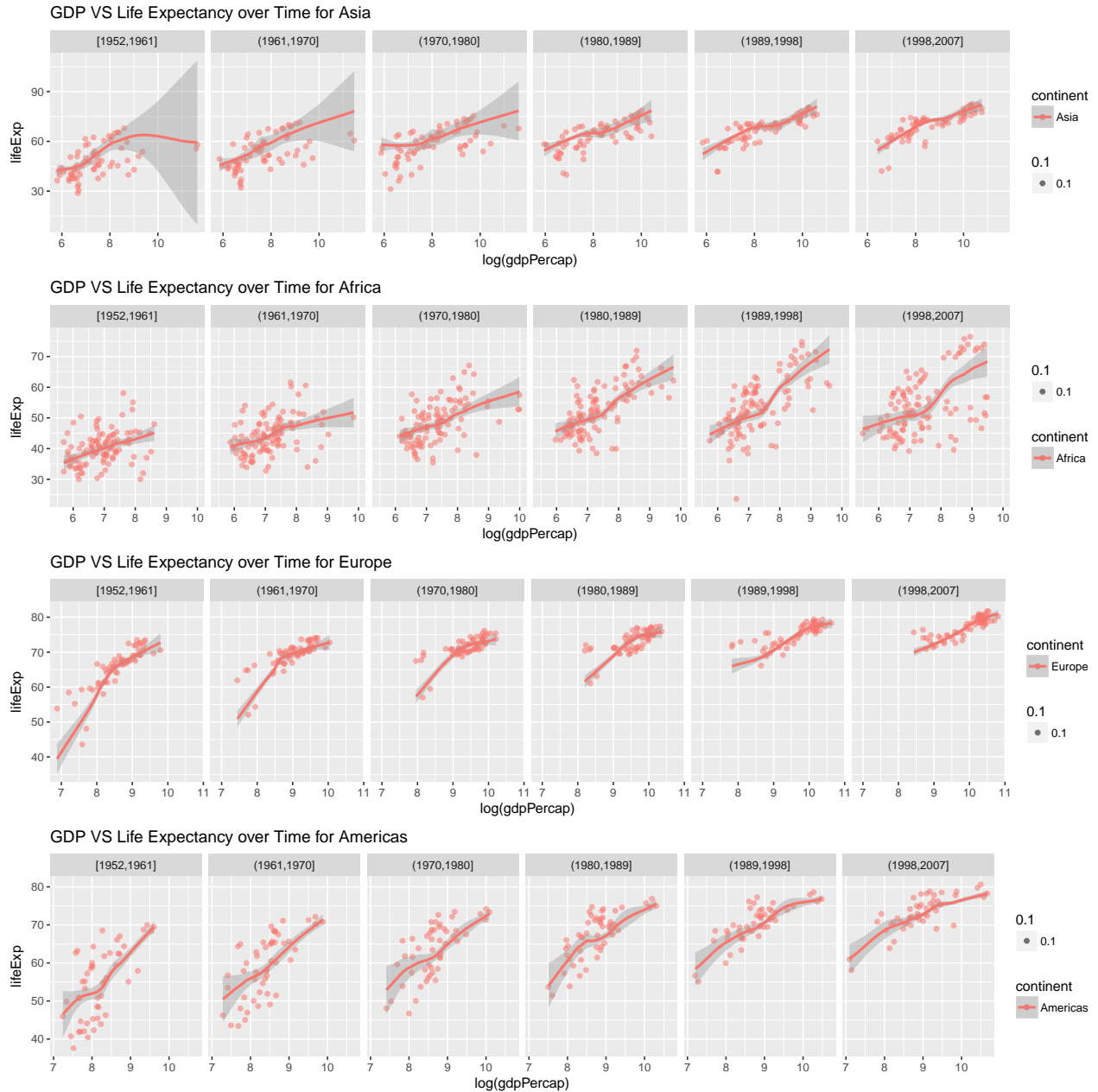
### QUESTION 3: Relationship between GDP and life Expectancy

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

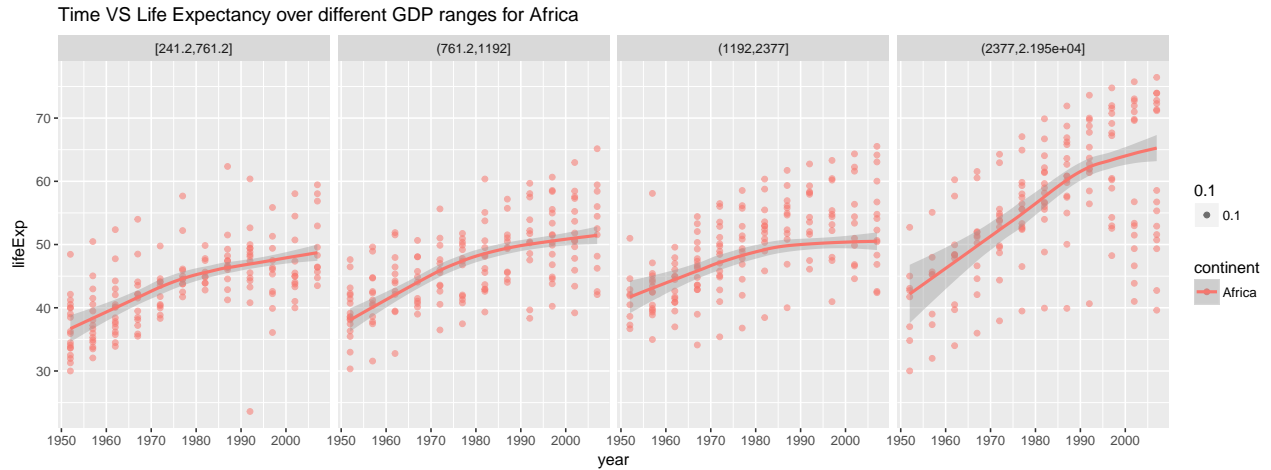
GDP VS Life Expectancy over Time for different continents



If we consider individual block of years, we see that the general trend is life expectancy increases with an increase in  $\log_{10}(\text{gdpPerCap})$ . We see that the slopes for first 3 blocks is something complex with different slopes. There is an increase in magnitude of Life expectancy as time progresses. Hence we can say that along with gdp, time also has an effect on Life expectancy. After 1980, we see that the slopes are almost similar, and the affect of time on life expectancy is not much compared to earlier periods, thereby suggesting convergence. In the graph, we can see that life expectancy of Africa is lowest amongst all the continents over the time. Let us get a broader picture of how life expectancy has changed with gdp and time for different continents.



For Asia, apart from the 2 outliers in the first and second group of years, we see there is a linear trend throughout. The trends almost seem to converge with time. On further investigating those two points represented Kuwait. By 1952, Kuwait became the largest oil exporter in the Persian Gulf. This massive growth attracted many foreign workers, especially from Palestine, Egypt, and India. This was one among many reasons why Kuwait had high GDP in that period. For Europe, apart from the outliers in lower gdp, we can see a general linear trend i.e life expectancy increase with  $\log(\text{gdpPercap})$  and also the magnitude increases over time. After a period the trend almost converges and effect of time is not much. Similarly we can say that even for Americas life expectancy increase with  $\log(\text{gdpPercap})$  with increase in magnitude of life expectancy over time. And after a period the trend seems to converge. When we check graphs for Africa, we see a lot of variance in data, and the trend is not simpler, in some parts there is a big increase in life expectancy with gdp and some part the increase is not much. In order to check this, let us check the trends for Africa for different gdp groups and see if there is any trend.



For Africa, we see that for countries with low gdp, the increase in life expectancy over time has not increased much. But for countries with higher GDP we see a greater affect of time on life expectancy. Life expectancy increases by a greater extent with time for these countries. So this might suggest that in poor countries where GDP is low, there might be no sufficient measure taken to improve life expectancy.

## Conclusion

We observed that for 2007, ignoring Oceania , Median life expectancy was greater for Europe and for Americas and Asia it was almost similar. Africa had the least life expectancy. Europe had a median life expectancy of 25 years greater compared to Africa. The overall trend what we observed in 2007 was life expectancy increased with increase in  $\log_e(\text{gdpPercap})$ . We also observe that the life expectancy has increased over time for all the continents, with few exceptions. For Asia, there was a dip around 1960's whereas for Africa, we see that after 1970's the increase is less compared to previous periods. We see that along with GDP & continent, time also has an affect on life expectancy. Upto a certain period we observe that the magnitude of life expectancy increases with time , and after that we see a convergence i.e time doesn't have much effect on life expectancy as before for the continents except Africa. In conclusion, we can say that the increase in life expectancy since World War 2 can be largely explained by increase in GDP per capita along with some other external factors.