CX3002 M4C GDP vs. Life Expectancy

# Introduction

There are two parts to my research. The first will consist of looking at 40 random countries, all over the world, and comparing their GDP and Life Expectancy in 2011, to see if there is a correlation, and how strong that correlation is, between the wealth and the health of a country. Part two will consist of selecting 6 countries, each from a different continent or geographical region, and focusing on these countries’ GDP, compared to their life expectancy, over a period of 50 years. I will also be comparing the growth of GDP with the increase of life expectancy, and how these are impacted by the location of these countries, and other variables. I will be looking at the period between 1960 and 2009, this is because the GPD data is only available since 1960 and choosing 2009 as the stopping point means that this time frame can be split nicely into 5 decades. I have selected these countries because they have a complete data set between the desired time frame for GDP and Life Expectancy, they are in completely different areas, and (out of a lack for a better method for picking randomly) they were first alphabetically. These countries are:

Algeria (28.0339° N, 1.6596° E)

Argentina (38.4161° S, 63.6167° W)

Australia (25.2744° S, 133.7751° E)

Bangladesh (23.6850° N, 90.3563° E)

Belgium (50.5039° N, 4.4699° E)

Benin (9.3077° N, 2.3158° E)

For the first part, my hypothesis is that GDP and Life Expectancy have a very high positive correlation (at least 0.8). I believe that a rise of GDP will lead to a rise in sanitary and health standards, therefore causing longer life expectancy. However, it could also work the other way, with higher GDP being attributed to higher life expectancy and the ability to work longer or harder.

For part two, the sample size is much smaller, so the correlation between GDP and Life Expectancy may not be the same as for part one. However, I think the sample will be sufficient to find out if the rate of increase to GDP influences the rate of increase to Life Expectancy. I believe that there will be a moderate positive correlation between these two data sets. I think that the rate of growth of infrastructure and wealth in a country will have a positive impact on how quickly life expectancy will rise, as health is one of the primary things that countries spend money on, regardless of how much wealth they possess. I will be working with 6 different countries, over 5 different time periods, resulting in 30 data entries. I worked out the mean value for GDP and Life Expectancy of each decade and will be using and comparing the results

# Part One:

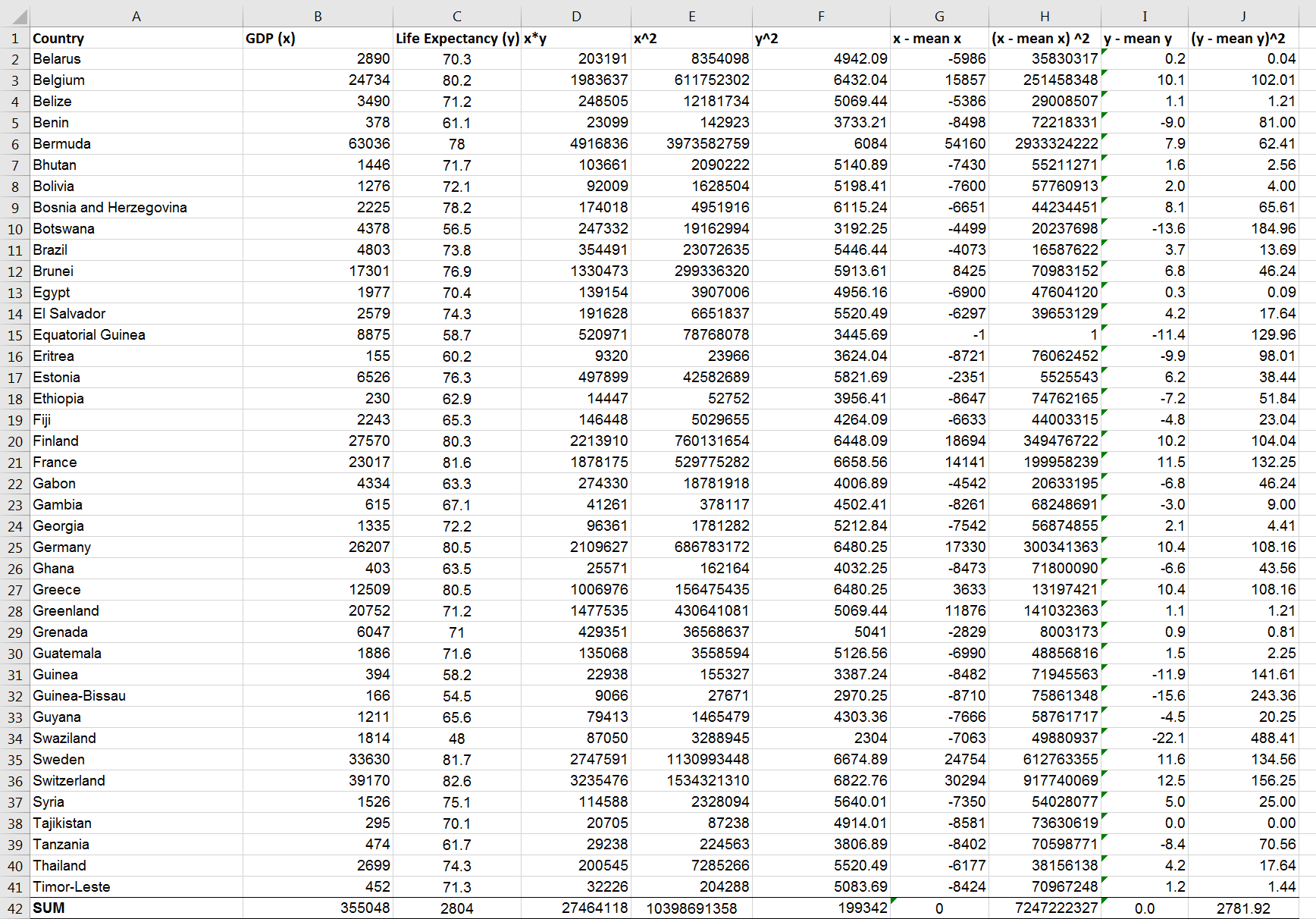
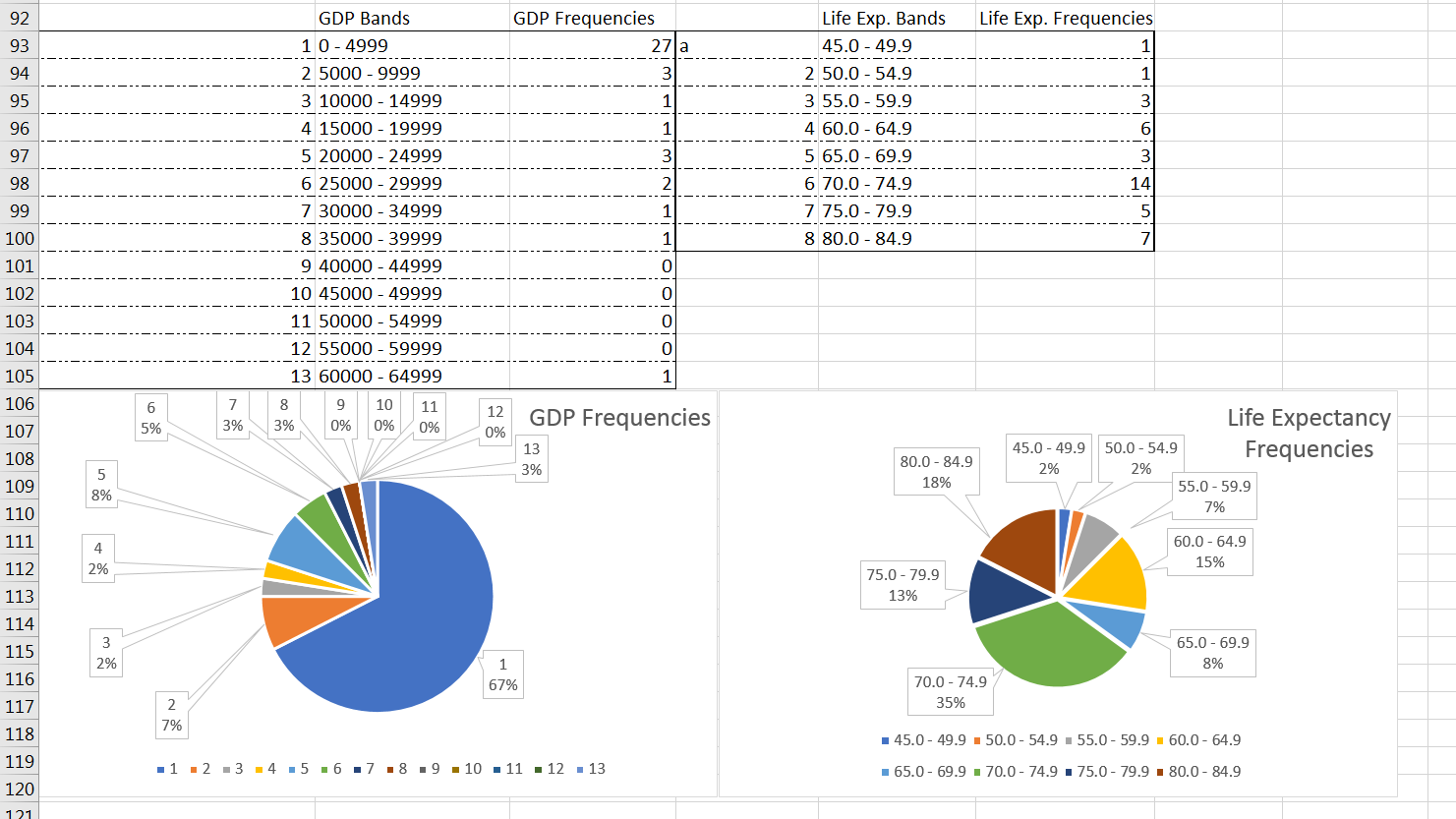
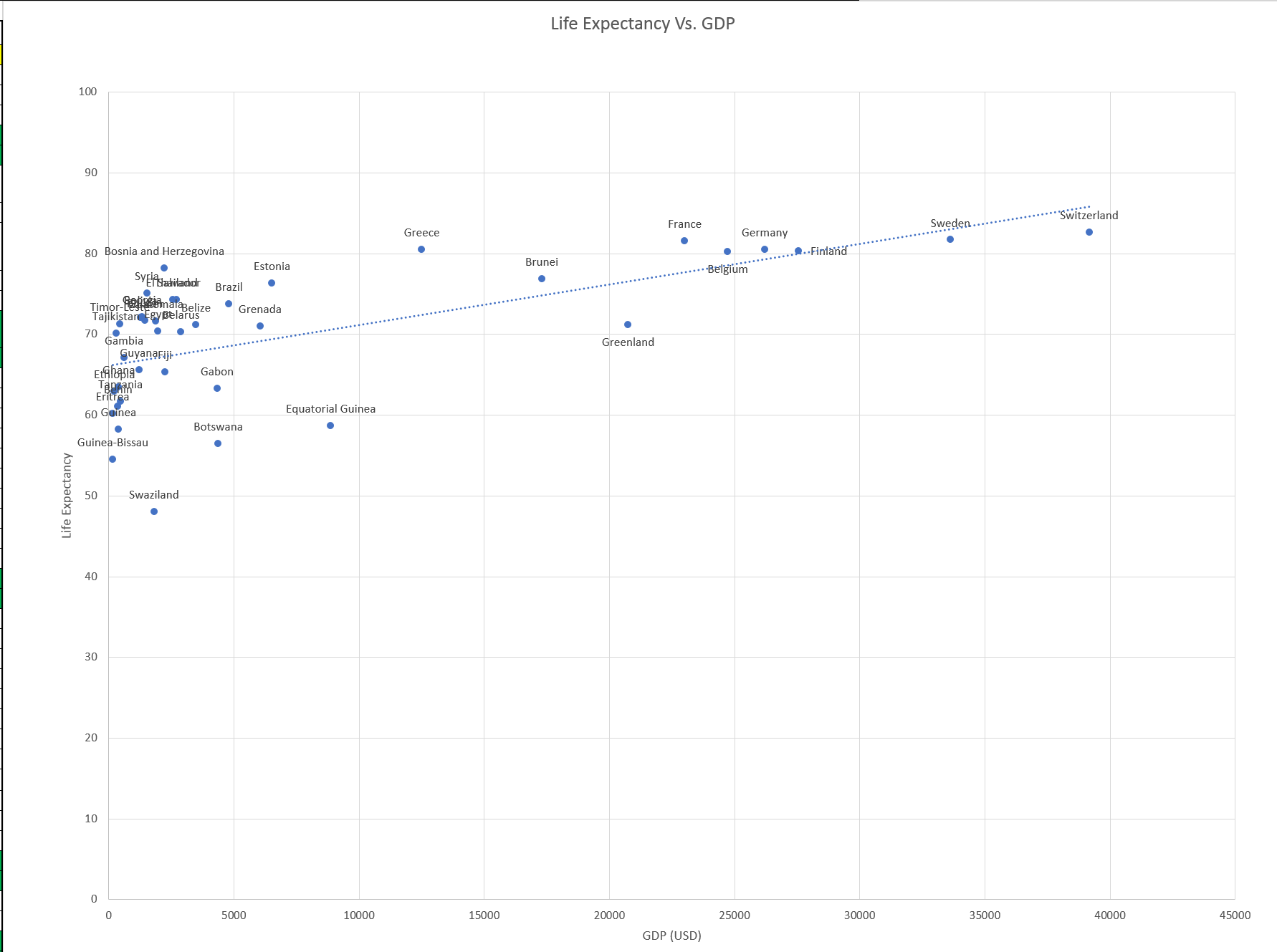
These are the countries that I chose at random to compare the GDP per capita (US$, inflation-adjusted) and Life Expectancy. The data used here is taken directly from Gapminder, from 2009 (as this is one of the newest entries). The data provider for the GDP per capita is the World Bank, while the Life Expectancy is provided for by various sources, although these are not listed.

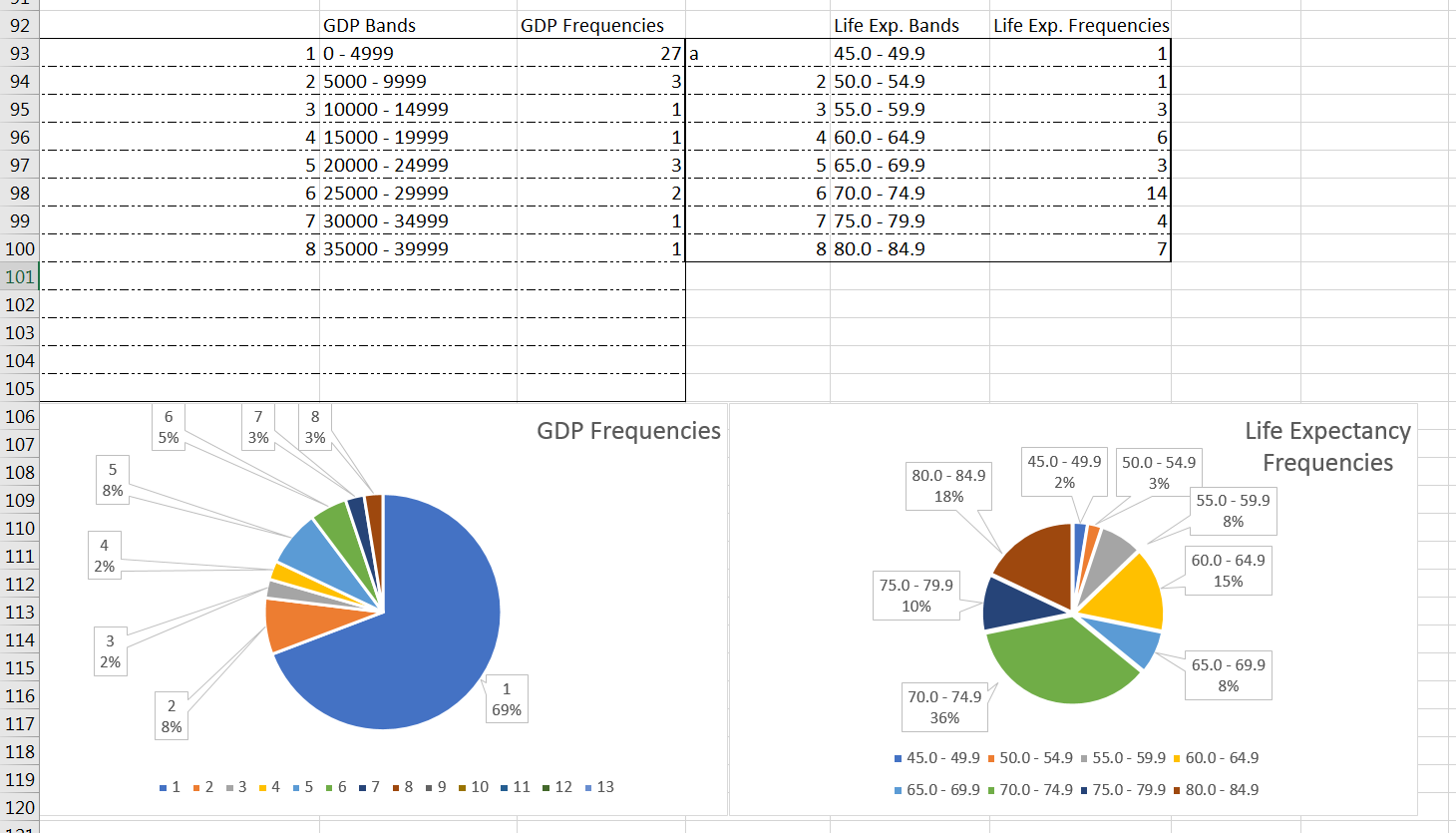
Table above – Original working out, including the outlier ‘Bermuda’



Charts above – Pie charts of the 40 countries with their respective bands and frequency tables.

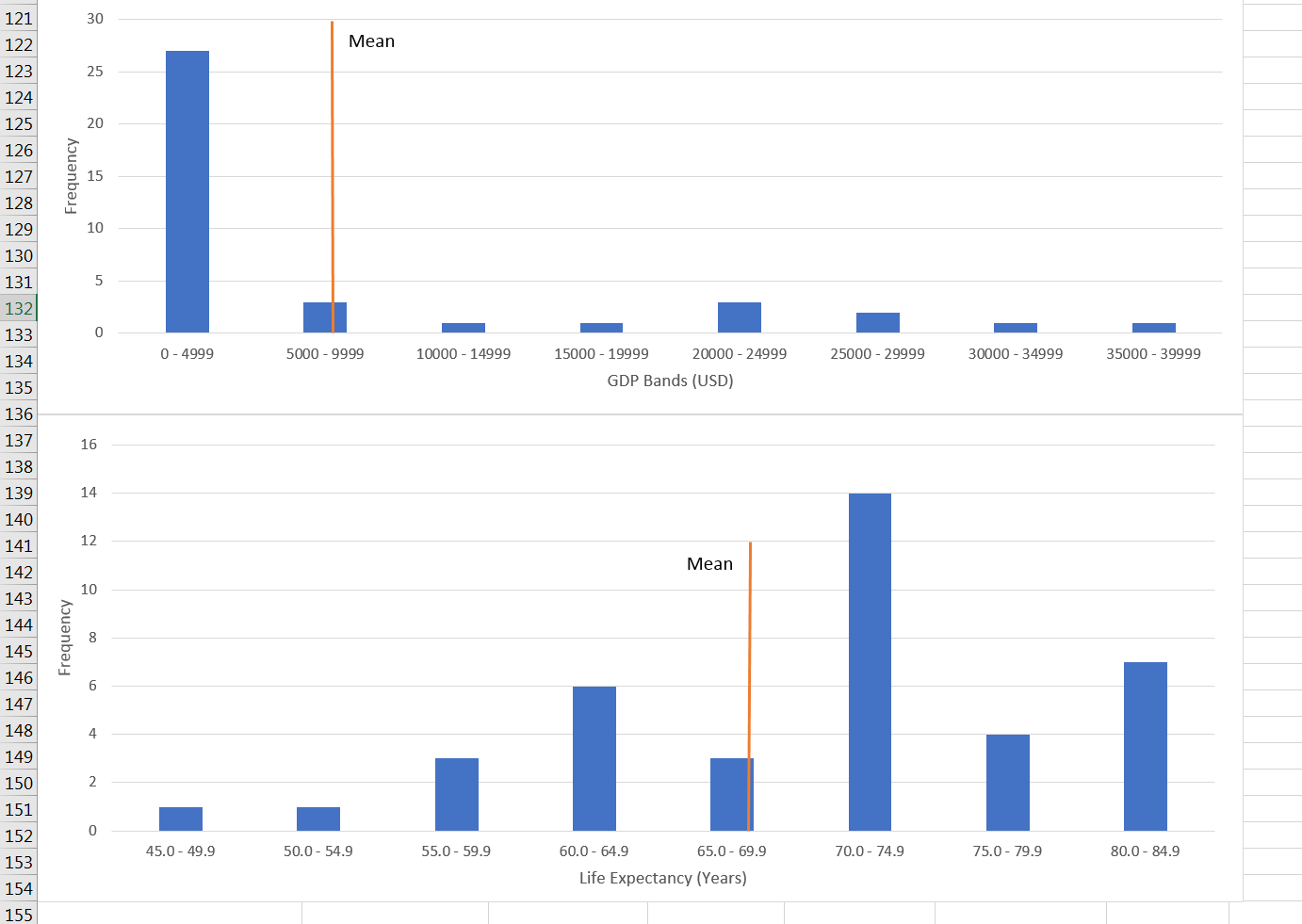
As the above frequency tables and pie charts show, the sample I used has a nice range of values. I have identified only one possible outlier, that being the 13th GDP Band, and the country being Bermuda. The reason for this outlier is that Bermuda is one of the biggest tax havens on the planet, and as such it would have an abnormally high GDP. As such, any further working out will be corrected for 39 countries, completely ignoring Bermuda. The rest of the entries do not seem to be outliers, especially considering the sample size. Bermuda was the only data entry that stood out as definitely being an unreliable source. 

As you can see on the graph above, the sample consists of quite a wide range of data. There are some interesting points to pick out here. For example, Central European and Scandinavian (*orange band*) countries can be seen as having a much higher GDP than South American, African or Asian (100x more in some cases), but they do not necessarily have much higher Life Expectancy. On the other hand, the very lowest countries (*blue band*) in terms of GDP have a trend of lower Life Expectancy. This hints at other factors playing key roles to the increase of Life Expectancy, but also suggests that past a certain point, the impact that GDP has is decreased dramatically. Comparing the blue band to the green, you can see that there is some clustering visible. This would suggest that a difference of GDP, the green band being approx. 5-7x the amount of the blue, can cause an increase of about 8 years. However, the difference between the green and orange bands, which is perhaps 10x as much, only yields an increase of maybe 6-8 years.



Charts above – Pie charts now representing 39 countries.

Graphs below – The data frequency tables represented in bar graphs, with the location of the mean approximately labelled.



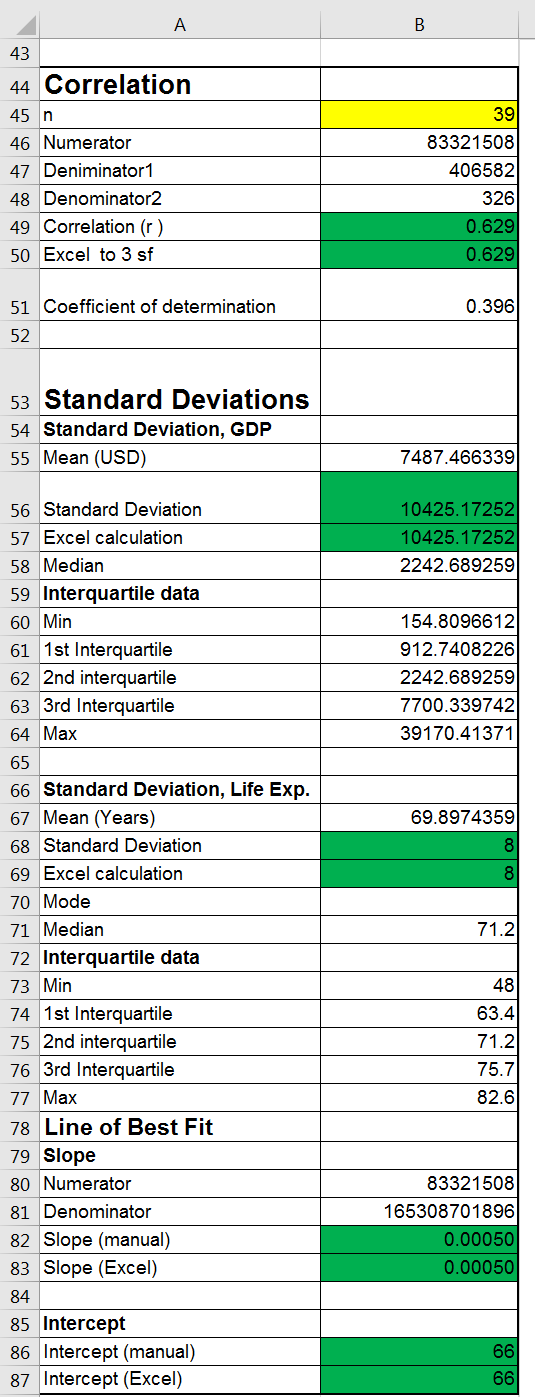


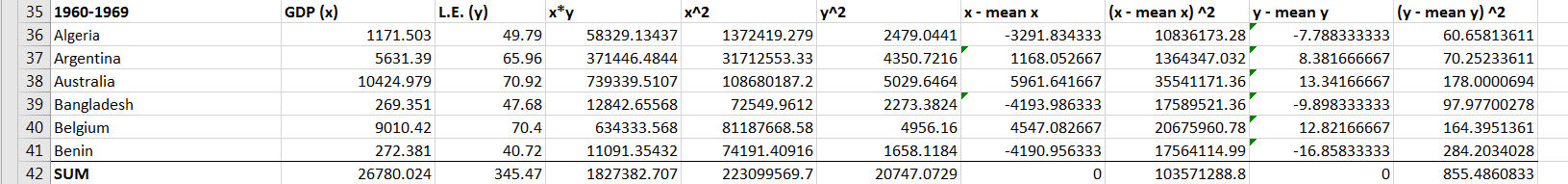
Table left – working out the correlation, standard deviations for GDP and Life Expectancy, and the line of best fit, manually and using the Excel functions.

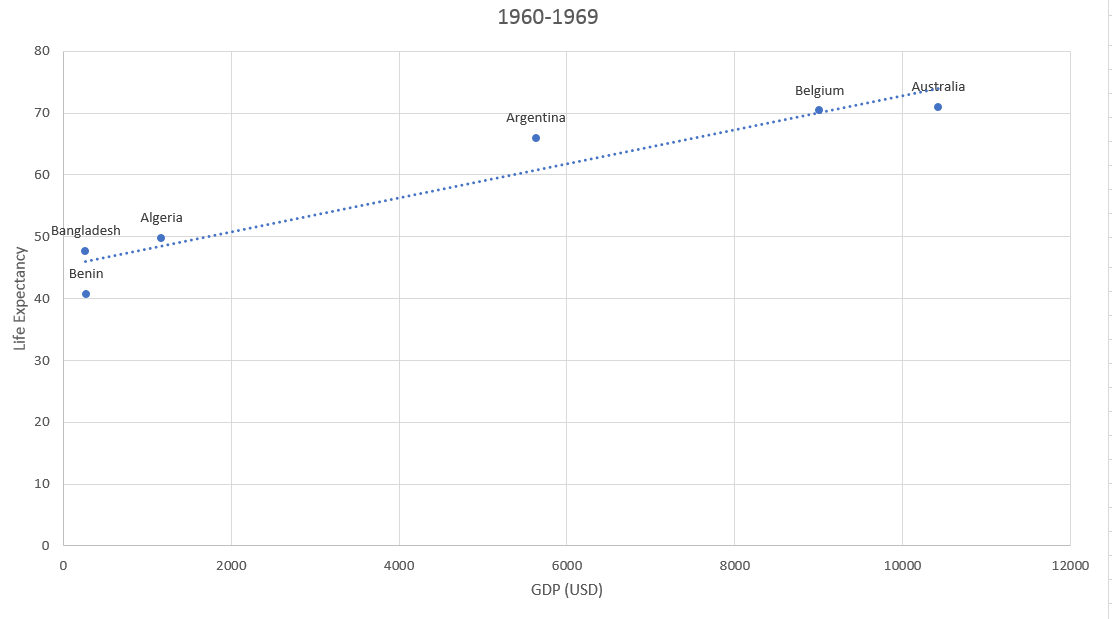
Looking at the data from this sample, I can say that my hypothesis was mostly correct. There is a moderate positive correlation between the GDP and Life Expectancy of a country. The data shows that just under 40% of the Life Expectancy is caused by the GDP. This is moderately high, although like I said in my hypothesis, I was expecting a higher value still. I believe other factors, such as diets, environment and even latitude, also cause an impact on Life Expectancy.

# Part Two:

For part two, I picked out data from 1960 to 2009 for 6 countries; Algeria, Argentina, Australia, Bangladesh, Belgium and Benin for Life Expectancy and GDP per capita (US$). As mentioned before, I grouped these data values by decade, and took the mean of each decade to represent the respective period. My hypothesis is as follows: first, I believe there will be a strong positive correlation between the GDP per capita and the Life Expectancy. I’d expect this correlation to be as strong as in part one, if not stronger (at least to begin with). Second, I think that the correlation will decrease over time, with a much stronger correlation between the two datasets in the 60s, and a weaker (but still moderately strong and positive) correlation in the 2000s.

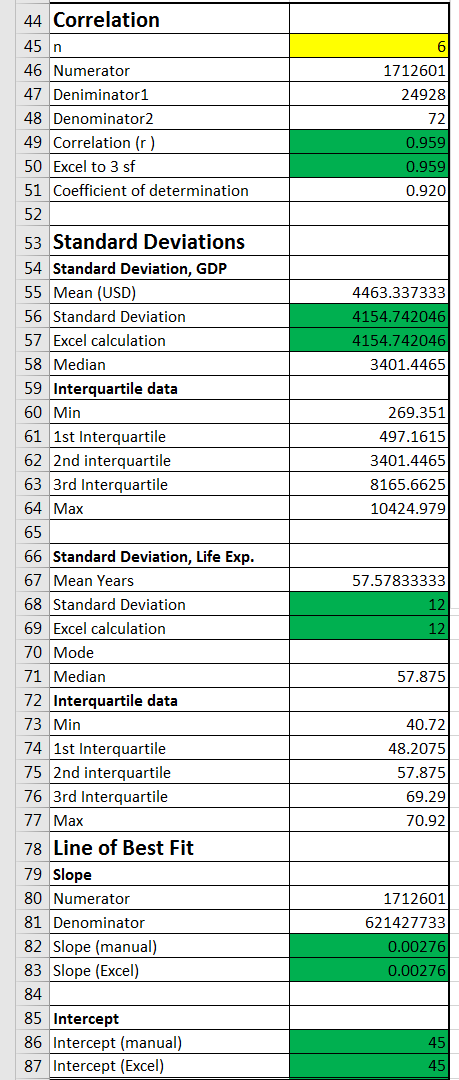
## 1960-1969



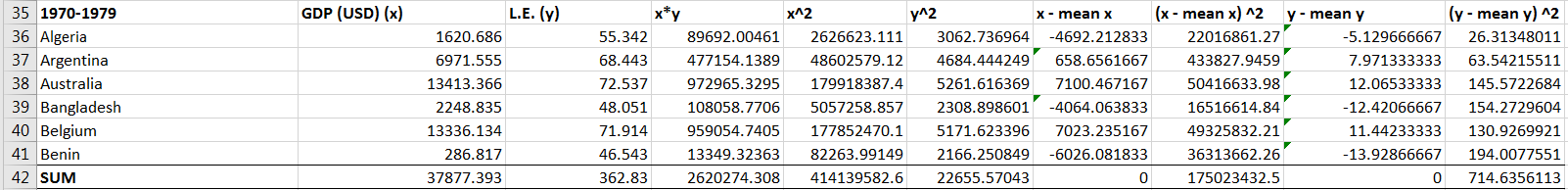


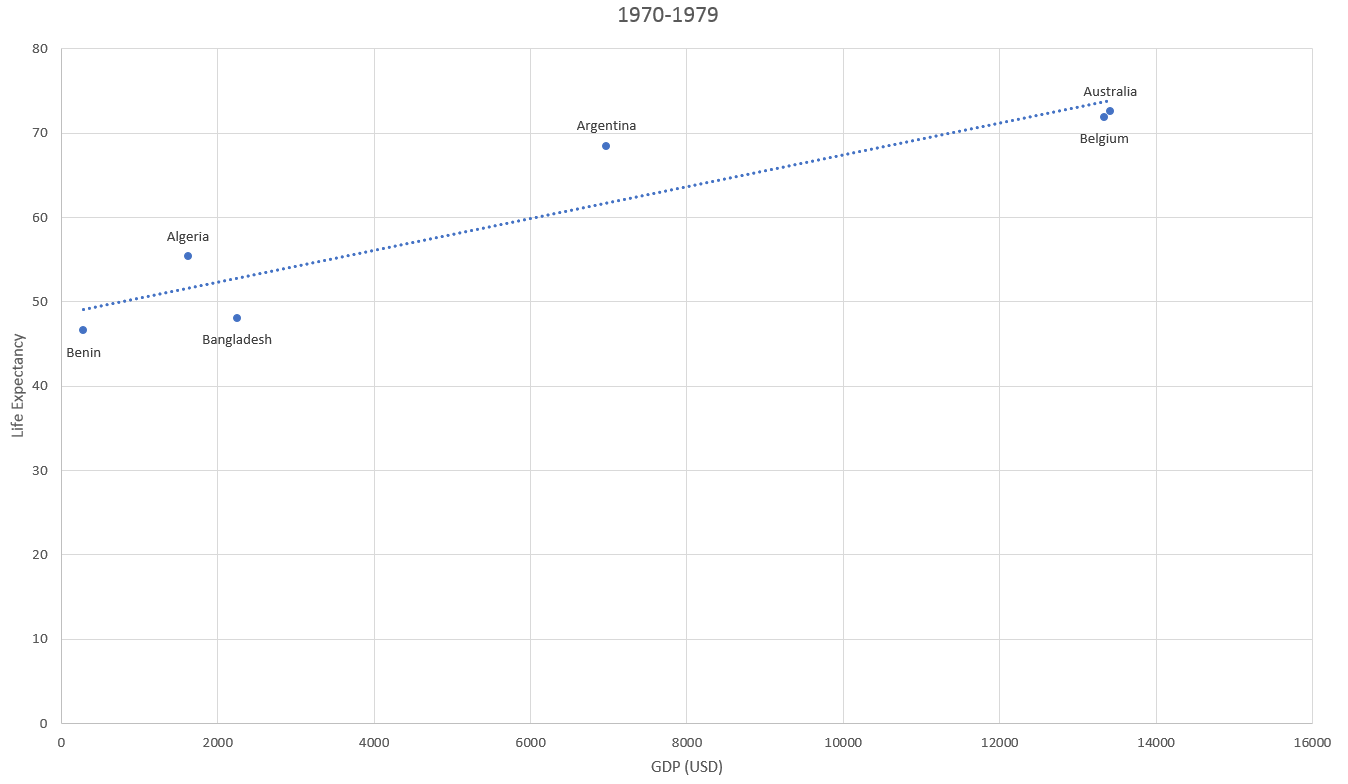
As the table below shows, the correlation for this period is 0.96, which is an almost perfect positive correlation. The mean GDP is 4463.34, and the standard deviation is 4154.74. In a normal distribution, one should expect approximately 80% of the data to be within 1 Standard Deviation of the mean, and in this case, that would seem to be true. The same can be said for the Life Expectancy, with a mean of 57.58 and a Standard Deviation of 12.   
Also, the line of best fit has a slope of 0.00276. This suggests that for every 1 USD increase to the GDP per capita, an increase of 0.00276 can be expected to the Life Expectancy.

**1960-1969**



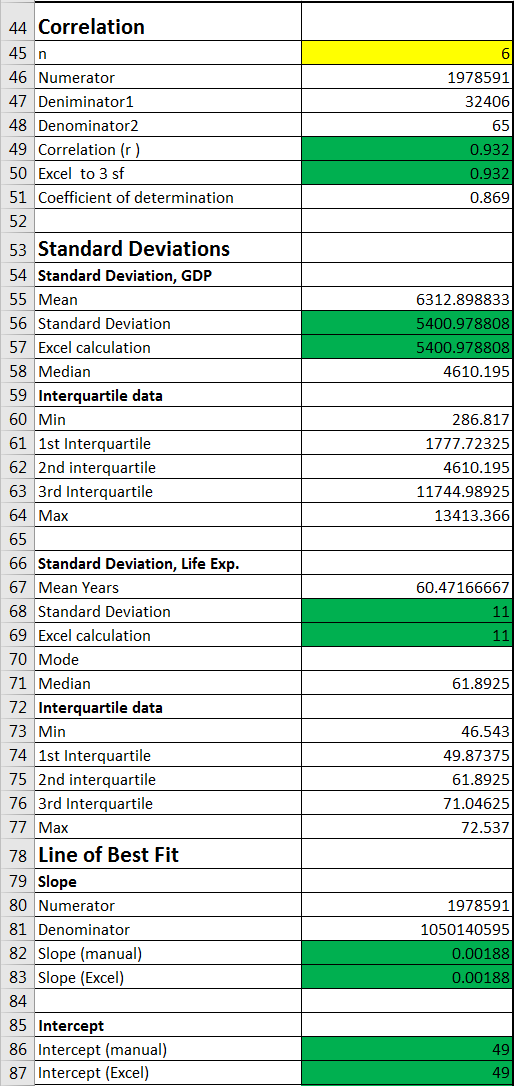
## 1970-1979



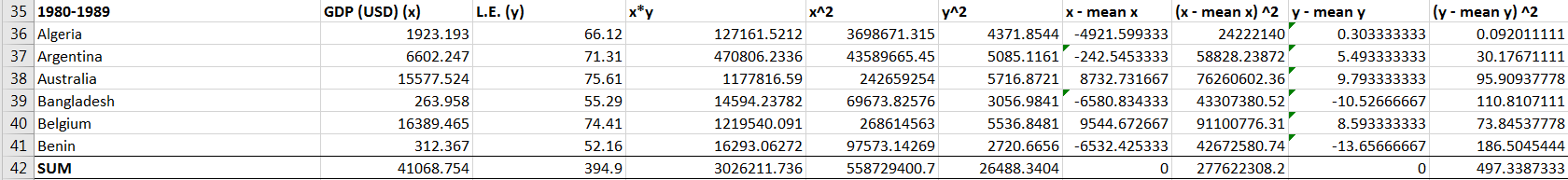


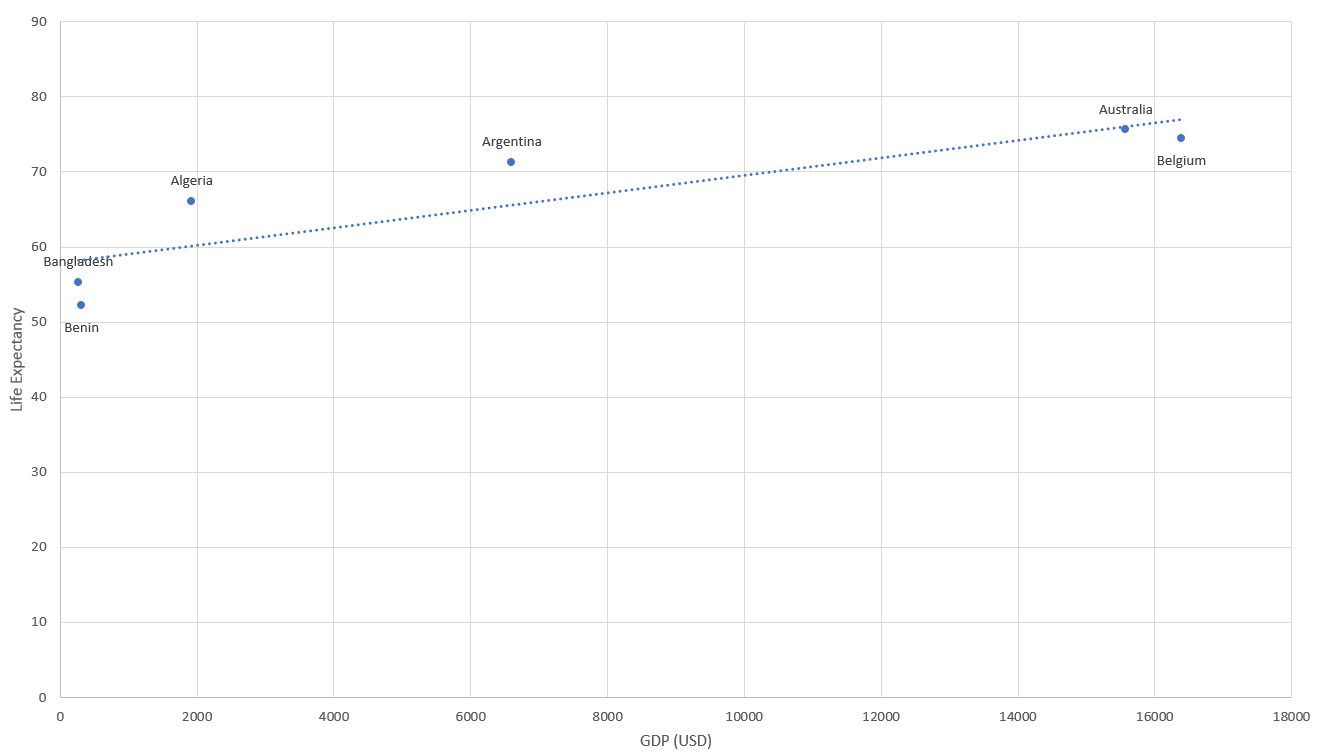
As with the previous decade, the correlation is strong and positive, although it is slightly smaller (0.03 less). The mean GDP has risen by approx. 50%, but the Standard Deviation has also increased substantially. This implies that the countries with less GDP have less GDP growth than more wealthy countries. This is the only way that data could become more spread out. On the other hand, the Life Expectancy mean has increased by approx. 2 ½ years, and the Standard Deviation has decreased by a year. The gradient of the line of best fit is also slightly smaller, implying that during this period, wealth may have had a smaller impact on Life Expectancy.

**1970-1979**



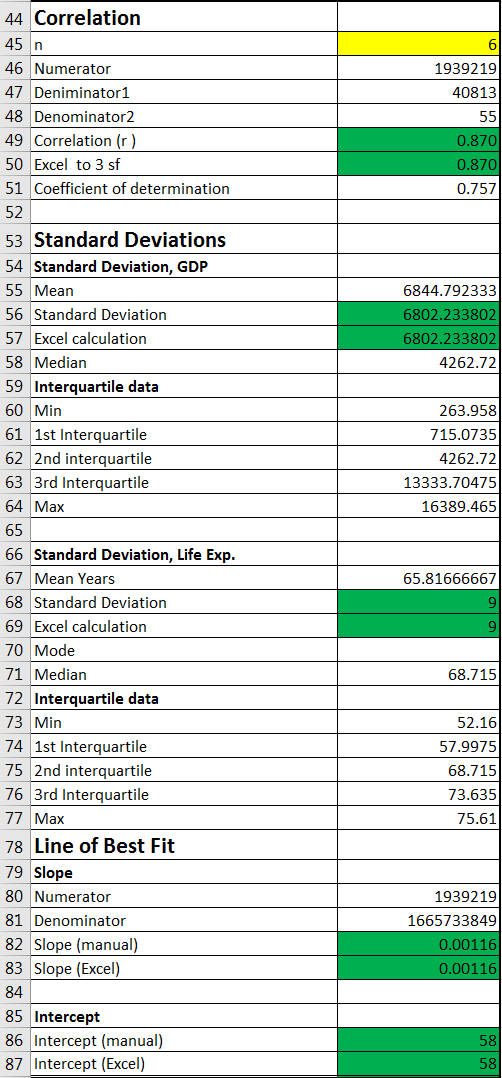
## 1980-1989



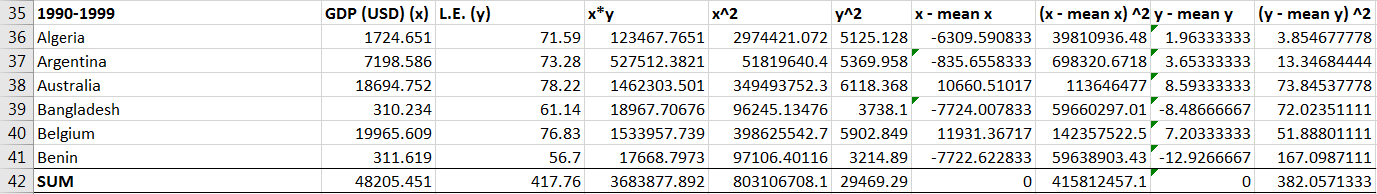


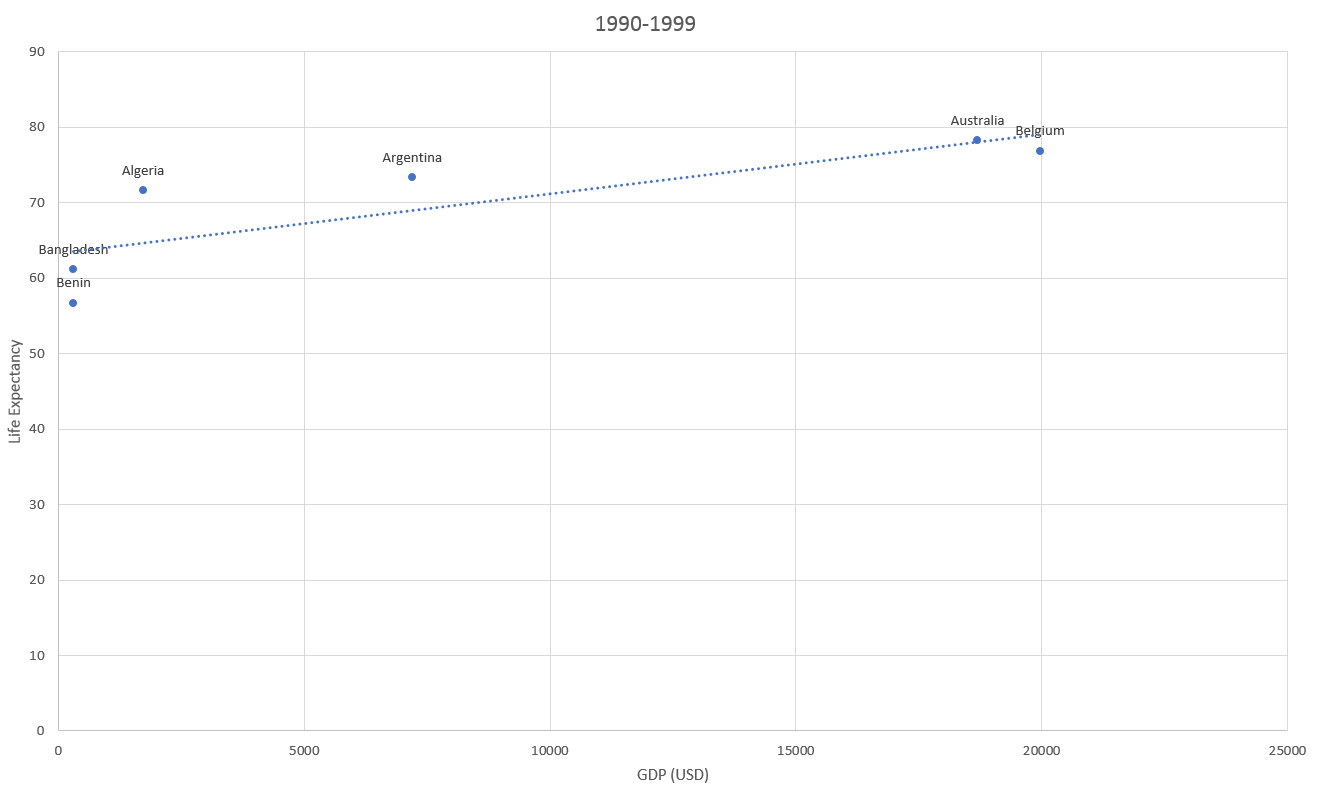
As with the previous set of data, the 1980-1989 period also sees a decrease of correlation between GDP per capita and Life Expectancy. However, this decrease is quite sizable, falling from 0.932 to 0.870. This is still a rather strong positive correlation but it is no longer perfect. The mean GDP has increased by 500, whereas the standard deviation increased by almost 1500. This suggests that majority of the entries are not being increased by a substantial amount, and in fact are just getting more spread out from each other. For Life Expectancy, the mean rises once again, this time to almost 66 years, and the standard deviation is reduced to 9. This means that the majority of countries are having their Life Expectancy increased, and that the gap between each country is being slowly filled. The line of best fit gradient is, once again, smaller than previously. It is now at 0.00116, which is less than half of the original amount in the 1960-1969 data set.

**1980-1989**



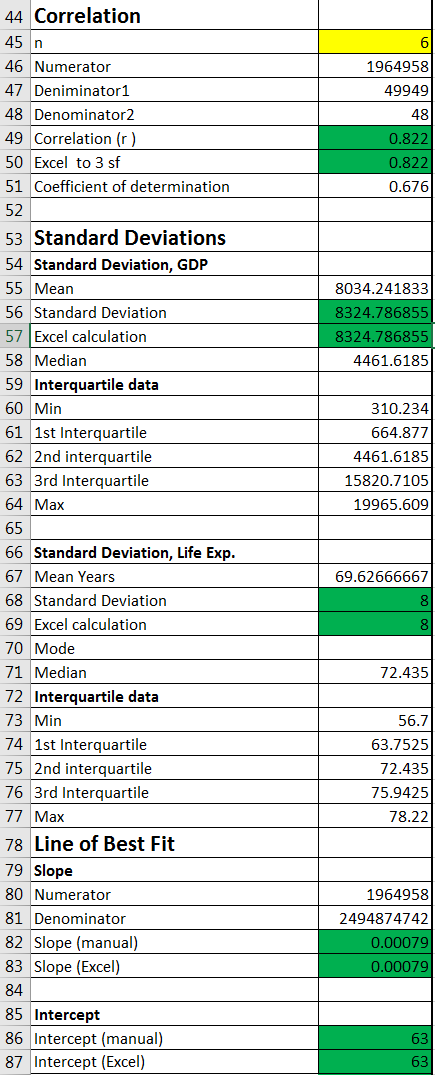
## 1990-1999



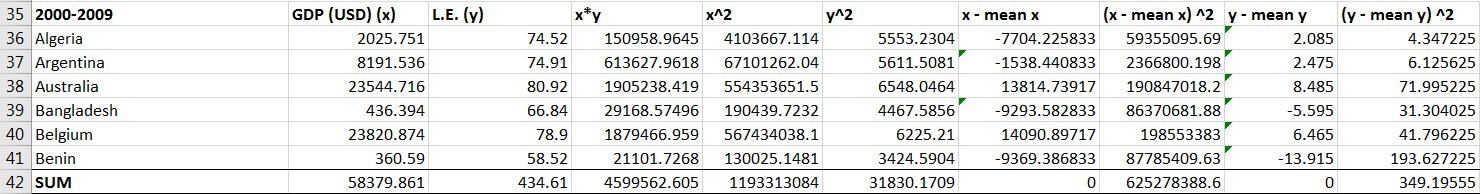


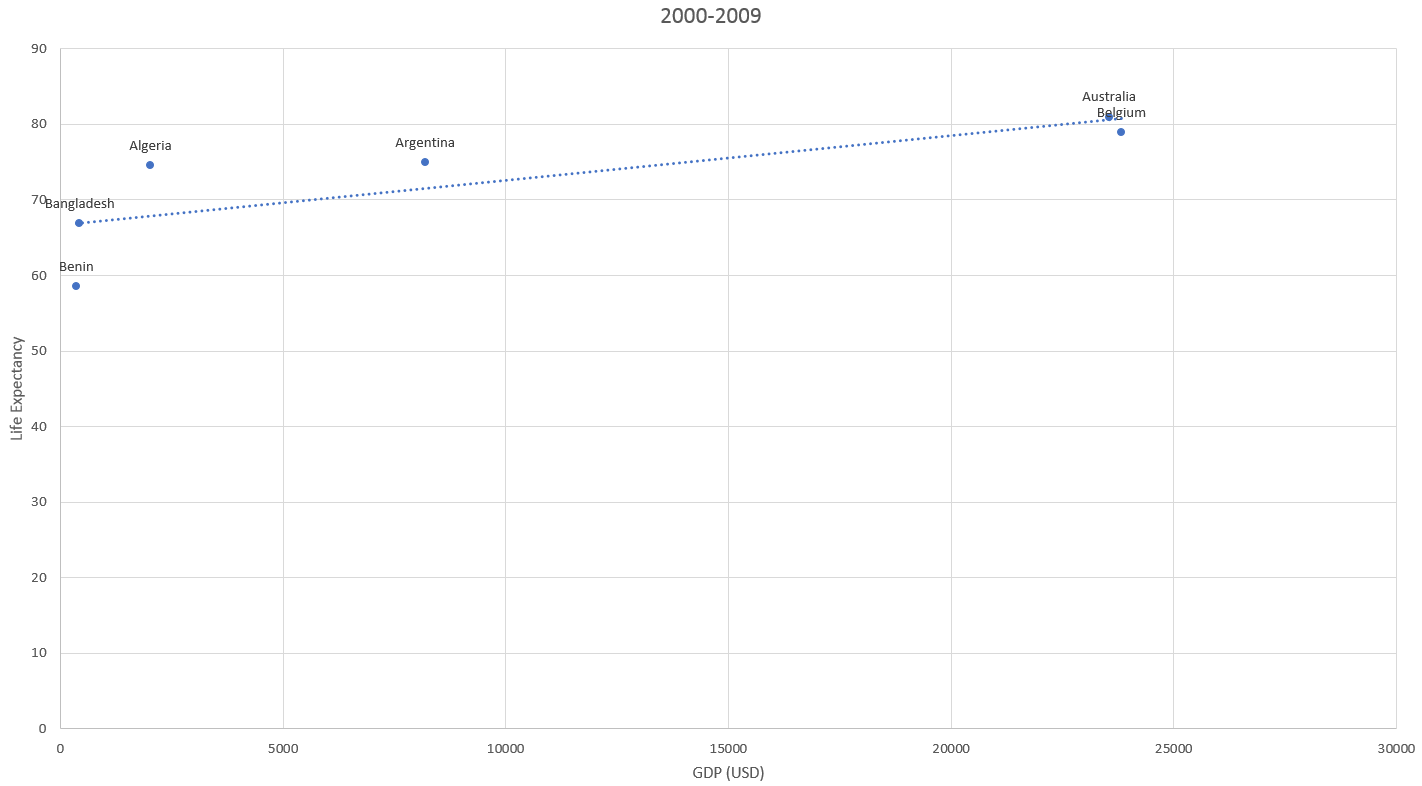
During this period, the Standard Deviation for GDP per capita overtakes the mean GDP. This happened because the two wealthiest countries, Australia and Belgium, are growing much faster than the rest. This is causing a huge gap between the data entries. The correlation is, once again, less than in the previous period. Life Expectancy is still steadily rising, with a small decrease to the standard deviation. The gradient of the line of best fit has almost halved since the previous time band, and the intercept of that line is still rising as it has been from the very beginning.

**1990-1999**



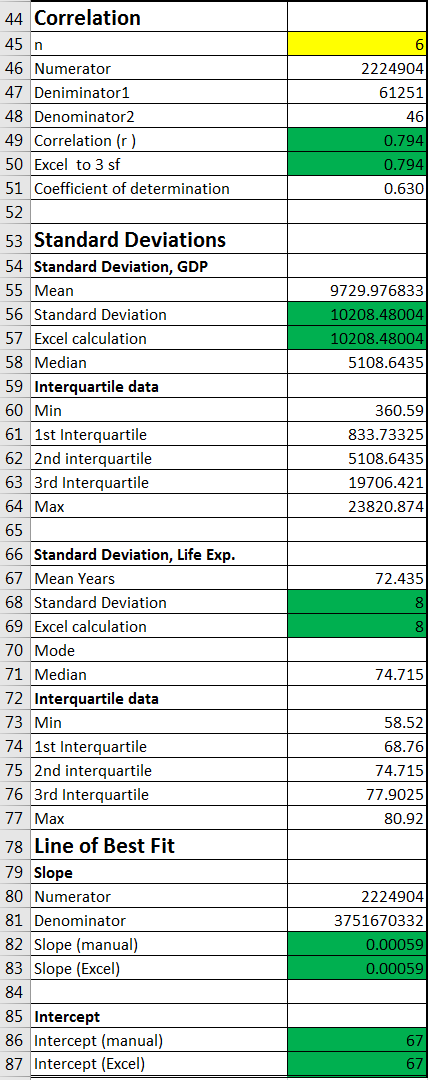
## 2000-2009



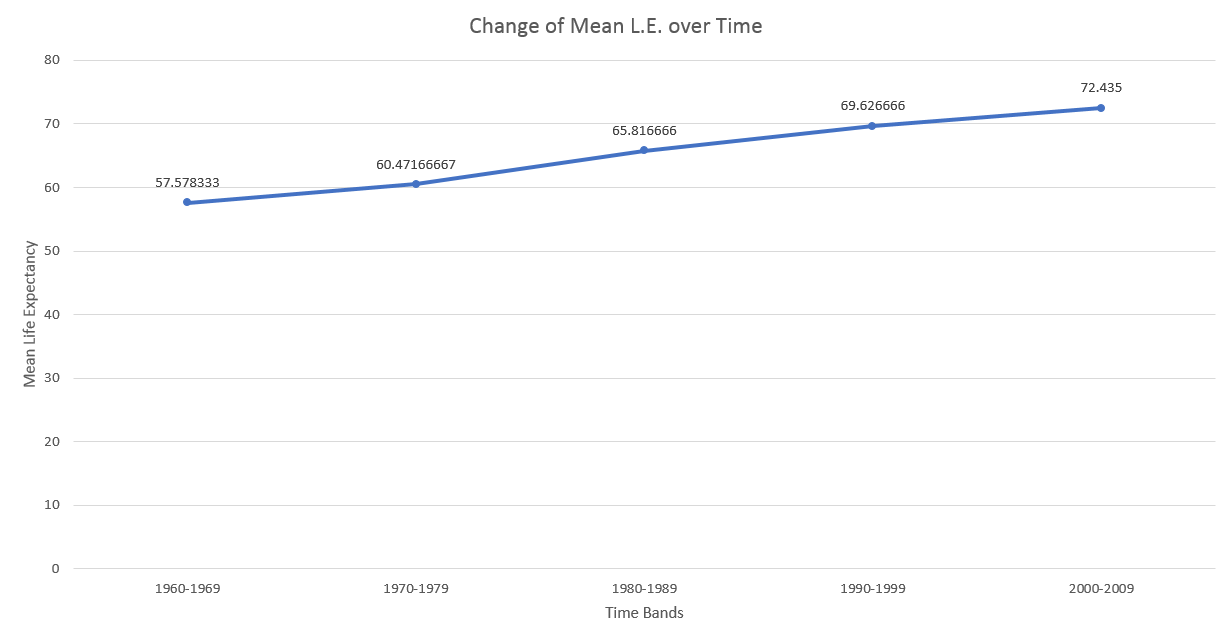


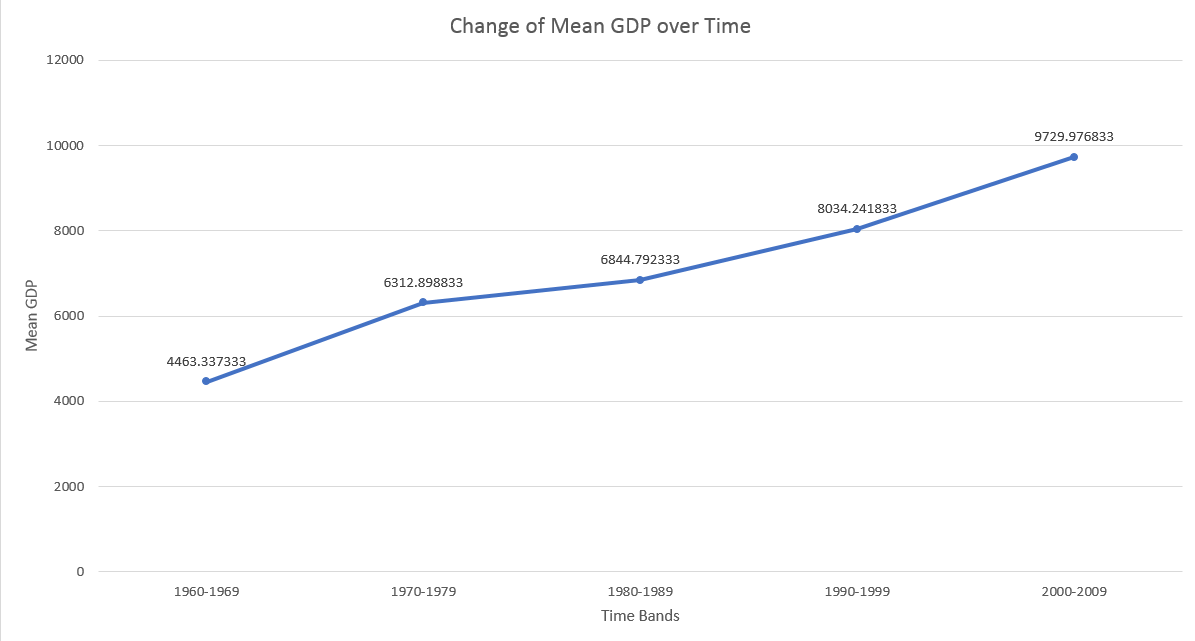
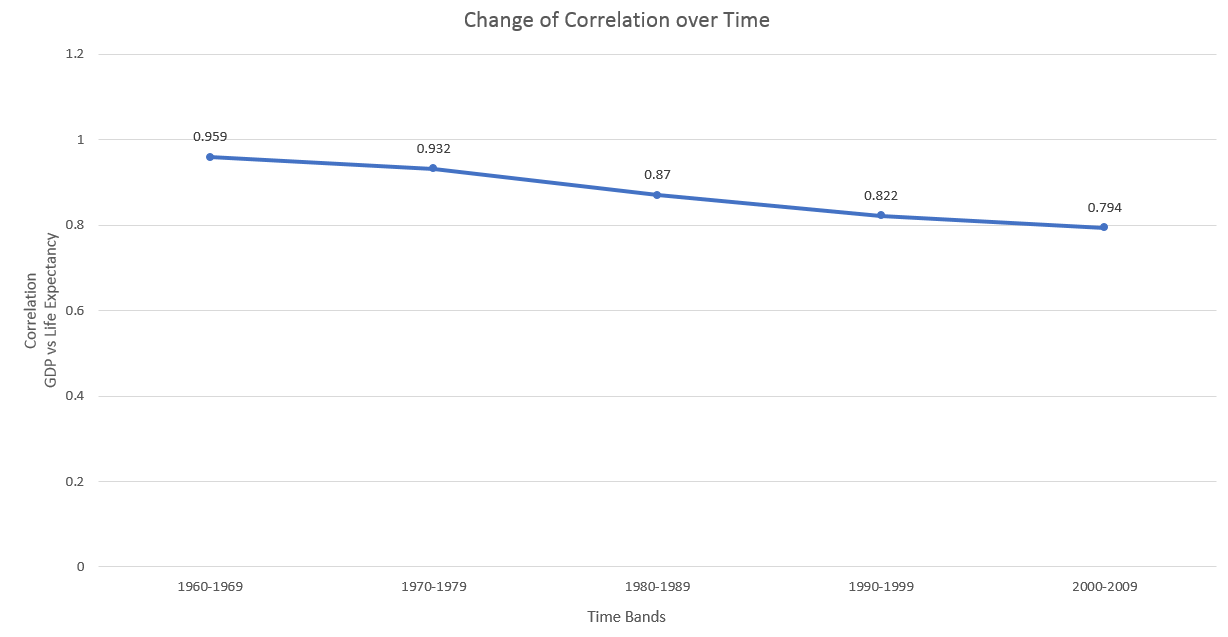
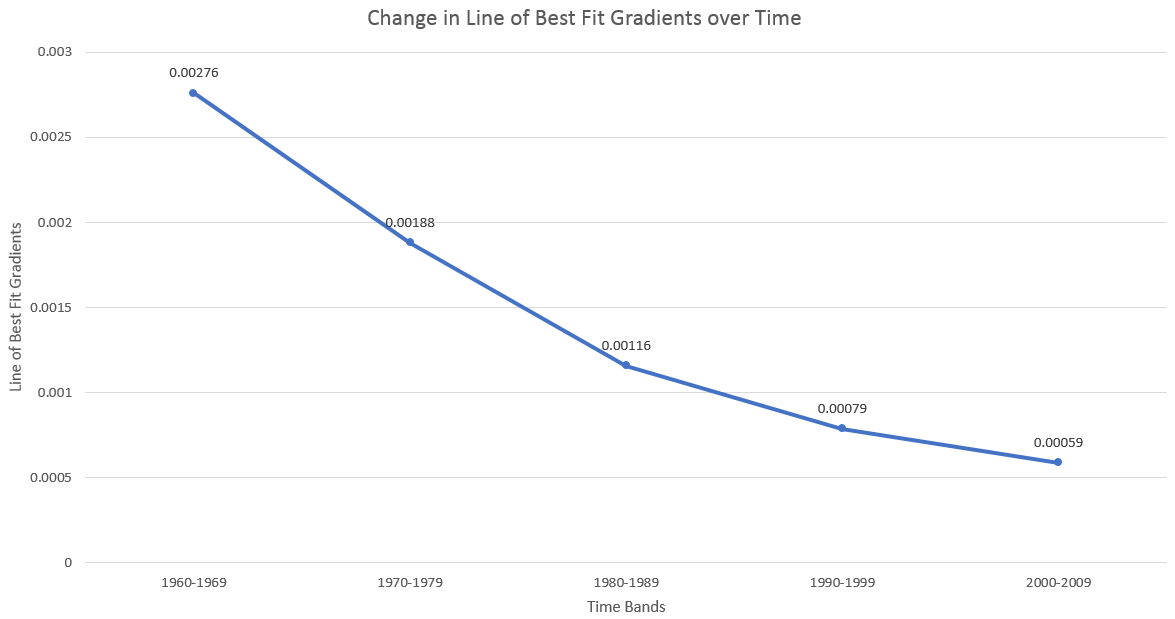
As the last dataset, it follows the trend set by the previous 4 to the dot. The correlation is, once again, smaller than before. It drops below 0.8, and so I believe it is no longer a strong positive correlation, but a moderate positive correlation. The GDP mean is 9729.98, while the standard deviation is 10208.48. This means that less than one standard deviation is needed to get to 0 USD GDP, and the value cannot go any lower. The Life Expectancy standard deviation remains the same as the last data set, but the mean is increased to 72.435.   
The line of best fit gradient is reduced yet again, although by a small amount this time. The gradient is 0.00059, which is almost 1/5th of the original gradient from 1960-1969.

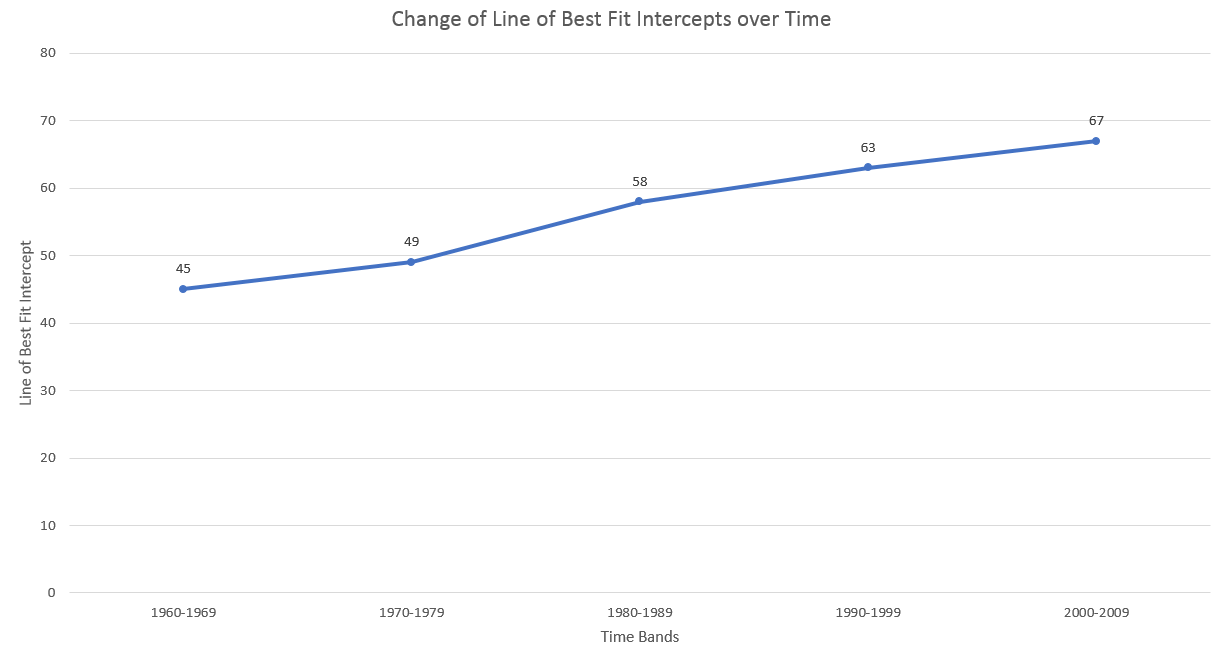
**2000-2009**



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1960-1969 | 1970-1979 | 1980-1989 | 1990-1999 | 2000-2009 |
|  |  |  |  |  |







The above graphs show a couple of interesting points. For example, the “Change of Line of Best Fit Intercepts over Time” graph very closely resembles the “Change of Mean Life Expectancy over Time” graph.

If I was to extrapolate the “Change in Line of Best Fit Gradients over Time” graph following the current trend its following, in less than 4-5 decades the gradient would reach 0 or very close to 0.

# Conclusions

Part One: As I mentioned before, I think my hypothesis was part right. The correlation is in fact positive, and there is enough trust in the data to say that there is some causation between the GDP per capita and Life Expectancy, although it is not the only factor. I believe the sample I selected reflected the population well, however the only way to be sure would be to include the whole population in future attempts to replicate this.

Part Two: In this part of my research, my hypothesis was more correct than earlier. The correlation did in fact start off extremely high and positive, and began decreasing as time went on. It went from almost certain causation (0.96) to highly contributing factor (0.8). I believe this is because the standards of health care and the use of technology for treatments went up drastically since the 1960s, and the cost of production on medicines was reduced, meaning that poorer countries had access to better or more drugs. I also believe that one of the main explanation for this is the rise of non-profit organisations and charities which are dedicated to bringing better health care to areas which cannot afford it or cannot access it for many reasons. I only selected 6 countries, although I made sure that they were from different geographical areas, to see if there was a chance at my hypothesis being right. After the data that I gathered, I would repeat the process but with more data entries. With enough time available, it would be preferable to gather the entire population.

Of course, with both parts there is the problem of outliers, and these would mainly be in the form of tax havens such as Bermuda, Luxembourg and others, but there would also be temporary outliers, such as wars, famines, natural disasters and other things of that nature.