Life Expectancy GDP Project

My objective with this project is to identify the relationship between GPD and Life expectancy in 6 countries. Then, I will plot my findings into meaningful visualizations that will be self-evident and easy to understand.

I’ve started my initial analysis by starting with an overview of the data, ensuring all of the data is accounted for and there are no Nan or Null values present. Changing column names to lowercase and reducing longer labeled columns to a shorthand version of themselves. Within the initial overview, we find four columns: country, year, life expect, and GDP.

Directing my attention towards the Dtype values I can concur that each column was properly imported over from the CSV file and that no type modifications will be necessary. I will however tack on another column to better represent GDP, as values reaching into the trillions is quite difficult to read at face value.

A quick glance over the .info() presents a few key interests. From the minimum and Maximum rows, we can see that our data spans over 15 years and will start in the 2000s and end in 2015. In addition, we also find that the population minimum life expectancy is 44 with a maximum of 81 and a mean value of approximately 73 years.

To begin my descent into the data, an initial scatter plot with a hue defining the 6 countries, GDP on the Y plot and years on the X plot. This will give me an idea of which countries will pertain to the higher and lower GDP of brackets. First impressions tell me that the United States and China have the higher GDP, while Zimbabwe and Chile have the lowest GDP. Next, I created two scatterplots for each country side by side that help visualize how GDP and Life expectancy progressed between 2000s and 2015. To really dig into our data, we want to look into why our data is the way it is and if there are any outliers or reasons for skewed data.

Looking at Zimbabwe, in the 2000s the contemporary lack of agricultural expertise from land redistribution caused export losses and negatively affected market confidence. (Agricultural export is one of Zimbabwe’s primary sources of GDP) Government spending is 29.7% of GDP. State enterprises are strongly subsidized. By 2008, unemployment had risen to 94%. At their current rate of development (2014), it would take 190 years for the country to double its per capita GDP.

Seeing that due to an Economic fallout across Zimbabwe during the periods in our data, we may find it difficult to see a stable relationship when correlating it with much more established and booming Countries.

Understanding that we are trying to find a relationship between two quantitate variables, we have to make a decision on what type of test will best summarize our dataset. First thing, I created a scatter plot with Life expectancy on the Y axis as it’s the dependent value, and GDP on the X axis since it is the independent value. Lastly plotting a line across the scatterplot to indicate the correlation between the two values, we can see that all countries plotted here show a positive linear relationship. Following the plots, I figured that a linear regression model would offer helpful insight into finding a correlation between the two variables. I created an OLS regression model for the two leading countries in GDP, and the two countries with the least GDP.

As we can see, there is a clear indicator between GDP and Life expectancy, and this is further reinforced by having multiple results from different samples.