Correlation between Life Expectancy and GDP: A Study

Caroline Hussey

Codecademy Data Analysis Project

May 2021

Project Objectives

- Has life expectancy increased over time in the six nations?
- Has GDP increased over time in the six nations?
- Is there a correlation between GDP and life expectancy of a country?
- What is the average life expectancy in these nations?
- What is the distribution of that life expectancy?

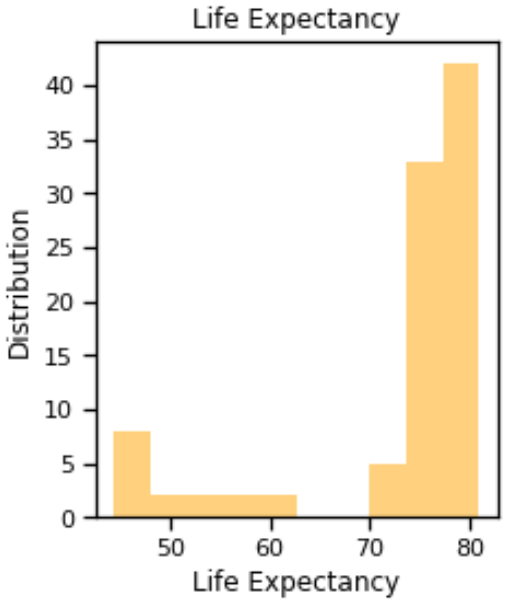
Data Sources

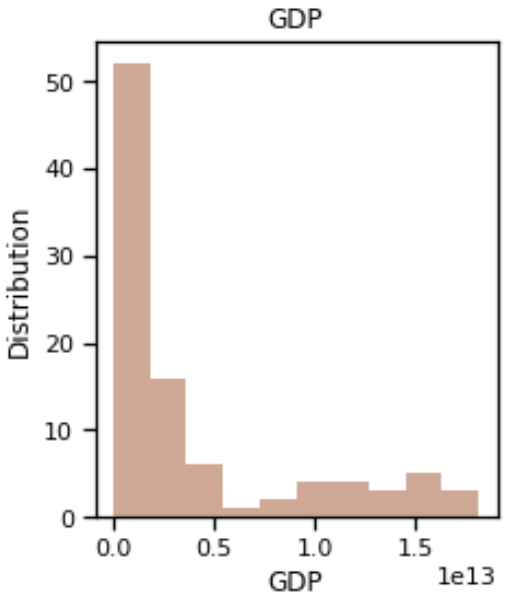
- GDP Source:[World Bank](https://data.worldbank.org/indicator/NY.GDP.MKTP.CD)national accounts data, and OECD National Accounts data files.
- Life expectancy Data Source: [World Health Organization](http://apps.who.int/gho/data/node.main.688)

Methods and Languages

- Jupyter Notebook
- Python
- Pandas
- NumPy
- Data Visualisations
- Hypothesis Testing – Pearsonr, one-sample ttest

Life Expectancy and GDP Distributions

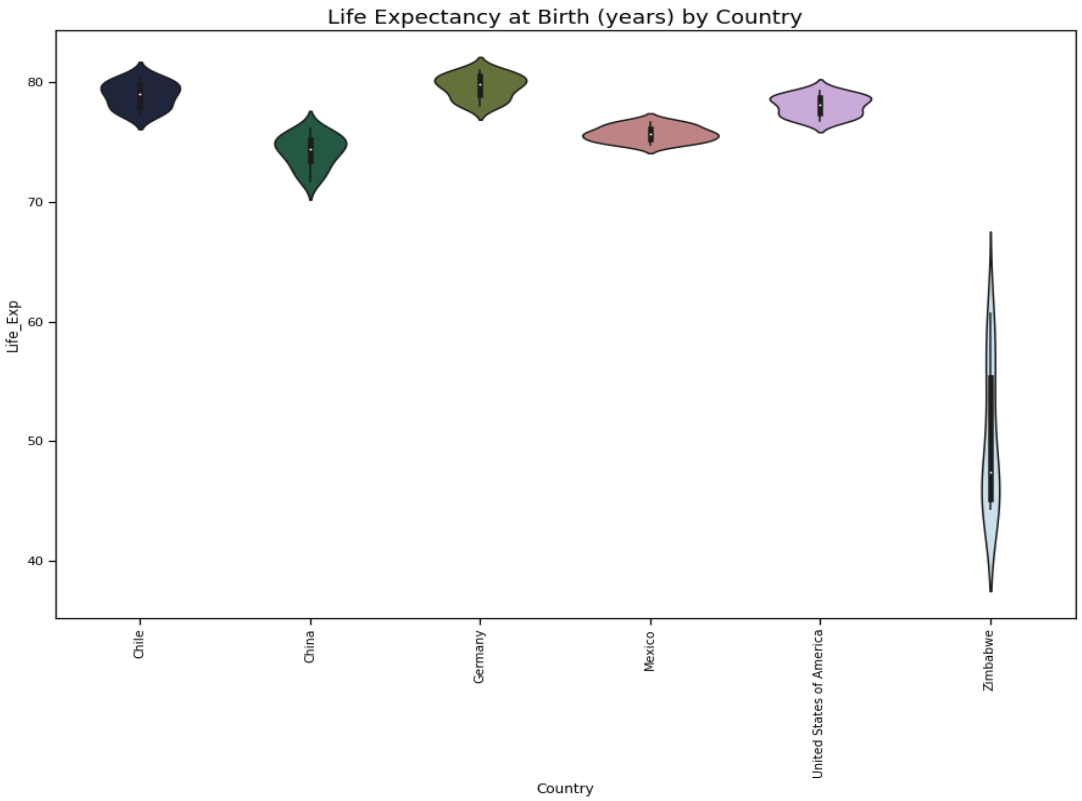




The distribution of the life expectancy graph shows a right skew

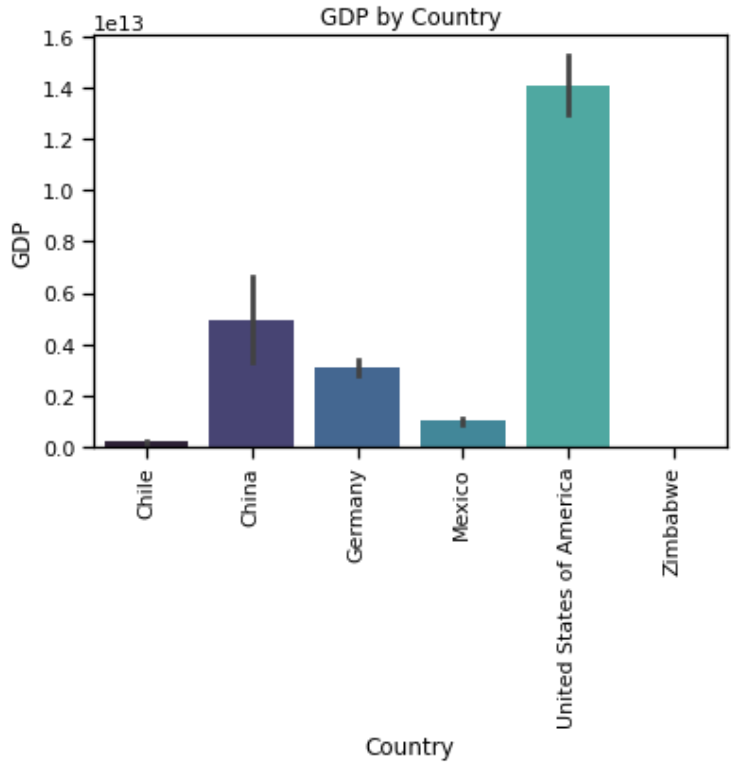
The distribution of GDP shows a left Skew

Violin Plot – Life Expectancy Means and Quartiles



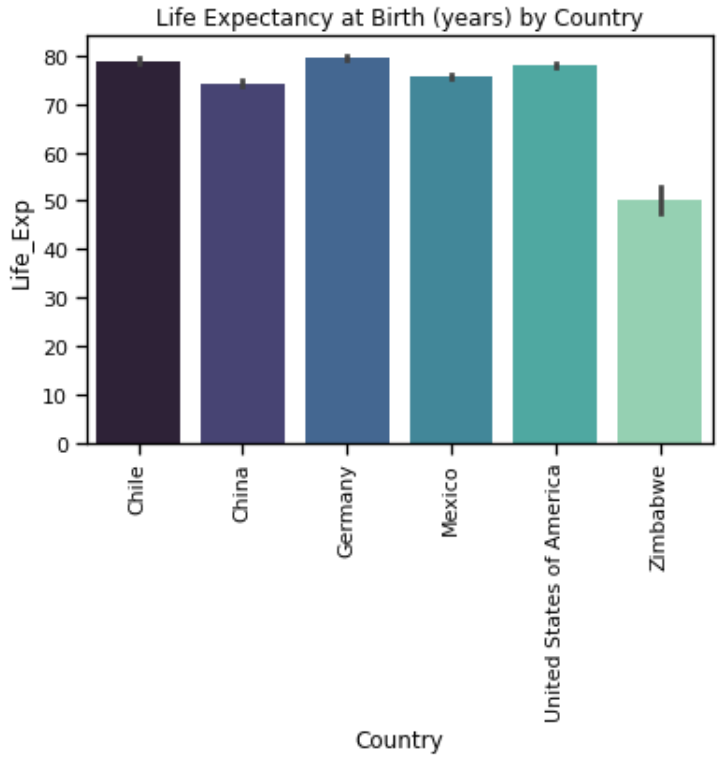
- The violin plot shows that the life expectancy in Chile, China, Germany, Mexico and the US all have a range between 70 – 81 with a mean between 75 – 80.
- Life expectancy in Zimbabwe shows a much wider range – from 40 – 70 – with an average around 50 years old.
- The violin plot shows that the maximum average life expectancy for Chile, China, Germany, Mexico and the US to be around 80 years old, whereas the maximum average life expectancy in Zimbabwe is 10 years younger than that - 70 years old.

GDP by Country



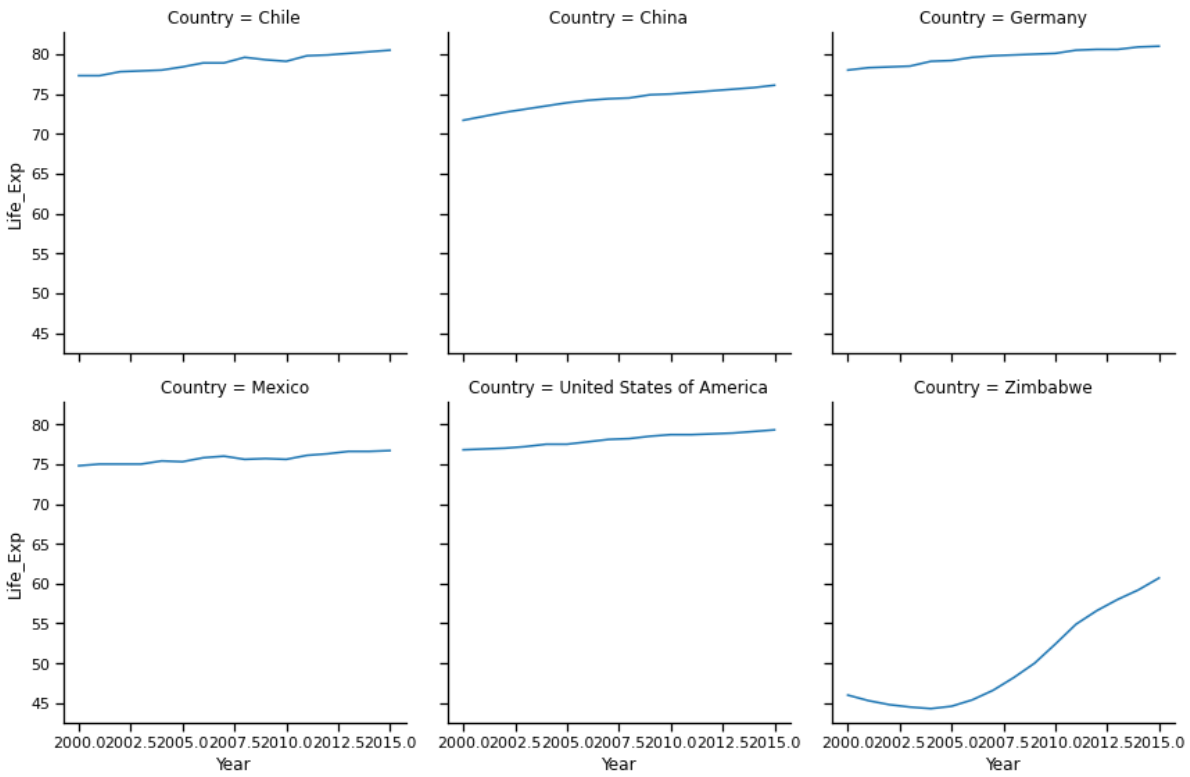
GDP is much higher in the US than the other countries represented in this dataset, whereas Chile and Zimbabwe show very low GDP.

Life Expectancy by Country

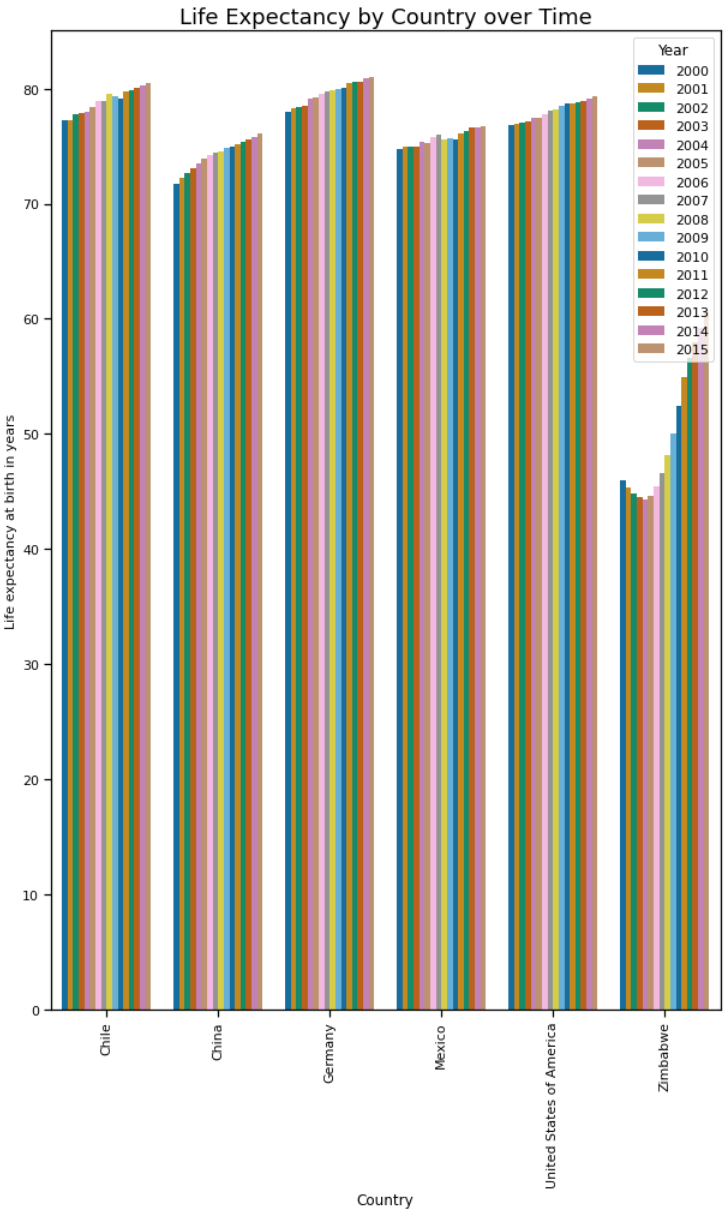


Overall Life expectancy appears to be fairly evenly spread between Chile, China, Germany, Mexico and US, with a slightly lower life expectancy in China. However, the life expectancy appears significantly lower in Zimbabwe than in the other countries explored in this dataset.

Life Expectancy Over Time by Country

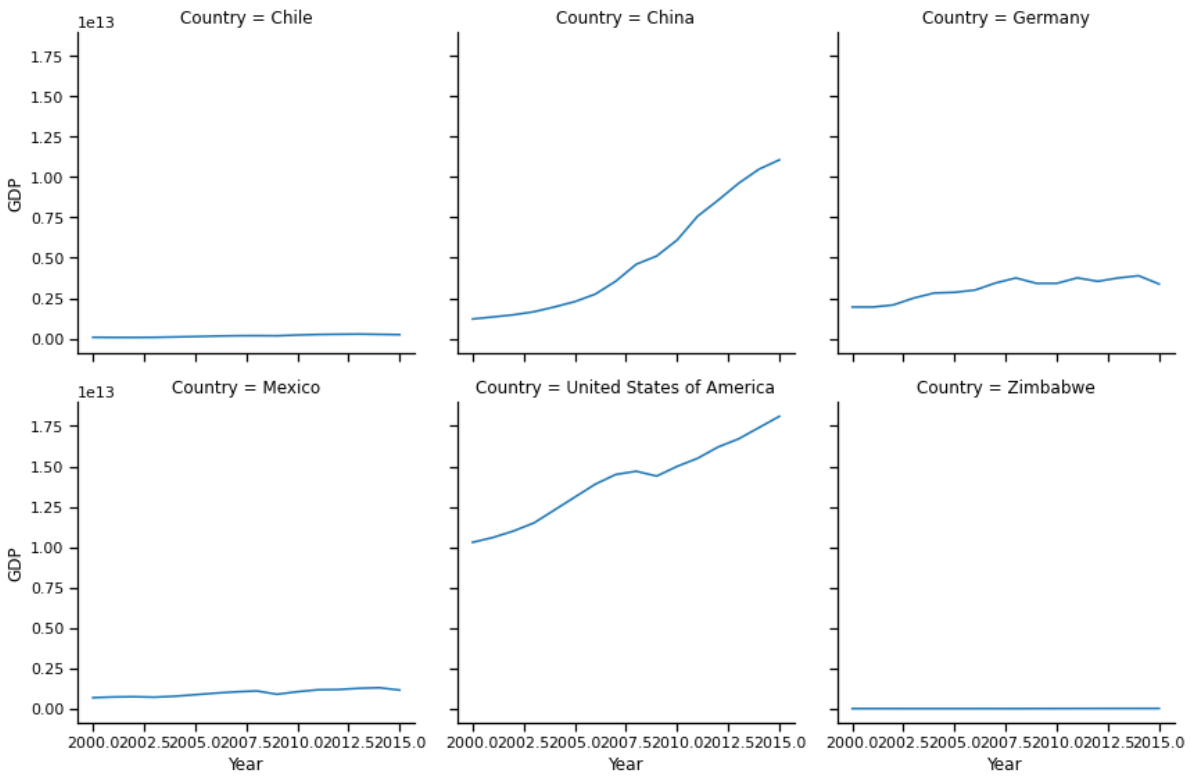


The line graph with facet grid shows changes over time more clearly. Here we can see there is a small but visible upward change in life expectancy over time in Chile, China, Germany, Mexico and US whereas Zimbabwe shows a significant change over time, particularly from 2005.

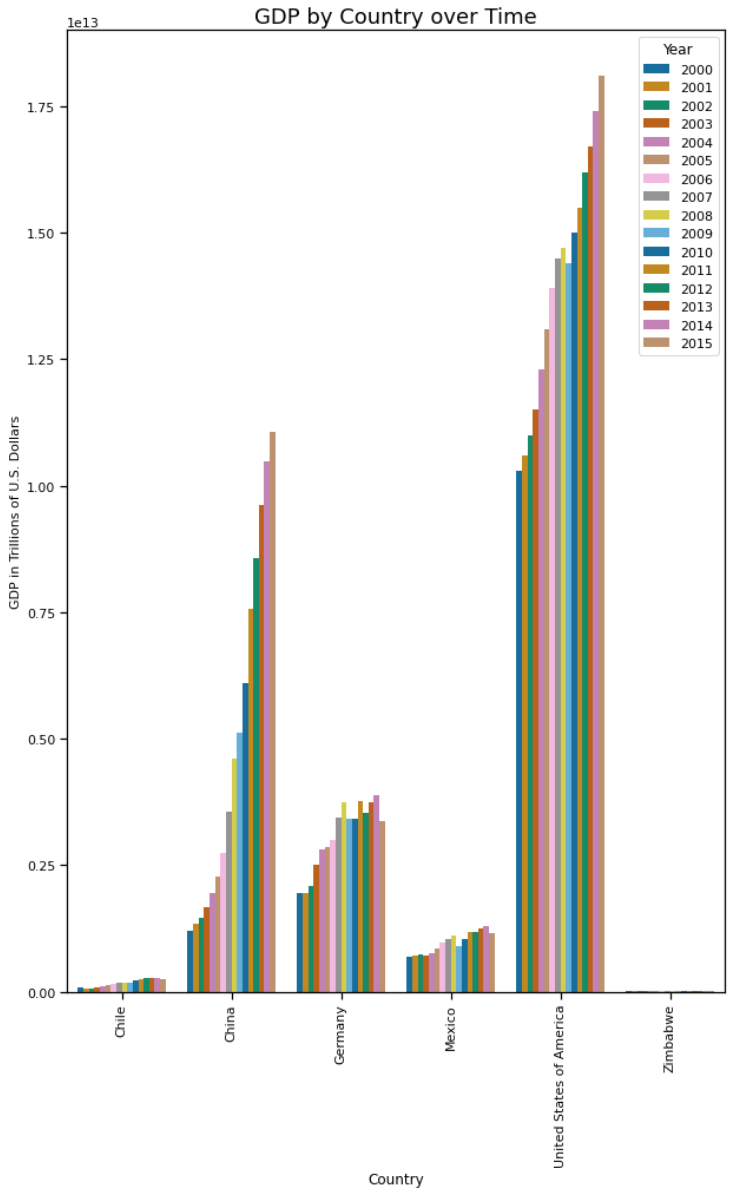


Life Expectancy shows an upward trend in all countries over time, with the largest change in Zimbabwe. Life Expectancy in Zimbabwe showed a downward trend until 2004, however it has increased significantly from 2005 until 2015.

GDP Over Time by Country

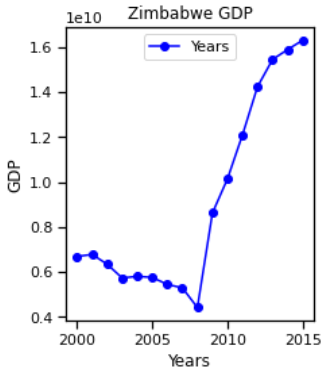


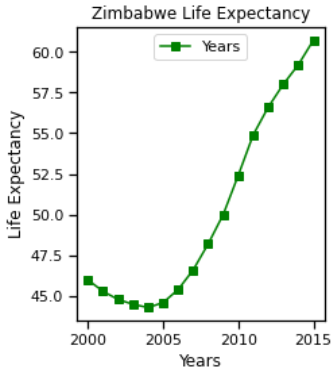
The facet grid / line graph shows changes over time for GDP more clearly. Here we can see there is only a small change in GDP over time in Chile, Germany and Mexico while the US and China showing a significant change over time. Zimbabwe shows no tangible change in GDP over time.



GDP trends over time vary significantly from country to country. Chile, China, Germany, Mexico and US all show an upward trend over time with China and the US showing a higher increase over time. GDP in Chile and Zimbabwe are steadily lower and significant change is not detectable in this graph.

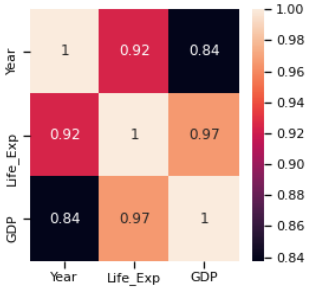
Zimbabwe: A Closer Look

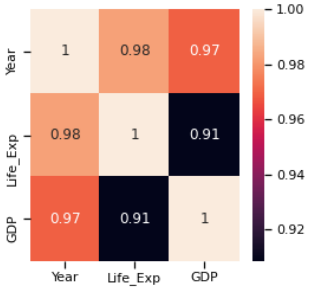


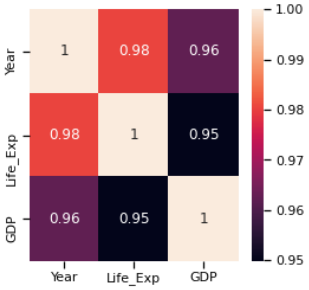


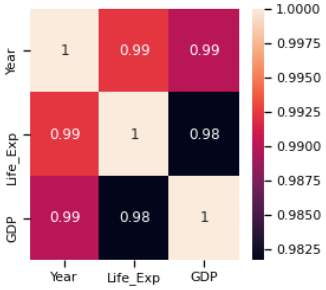
By taking a closer look at Zimbabwe's GDP and Life Expectancy, we can see that these factors are growing just as they are for the GDP and Life Expectancy in the US, China, Chile, Mexico and Germany. However, the scale for GDP is much smaller in Zimbabwe, and much higher in the US, resulting in a 'squished' chart. One way around this is normalizing the data prior to analysis.

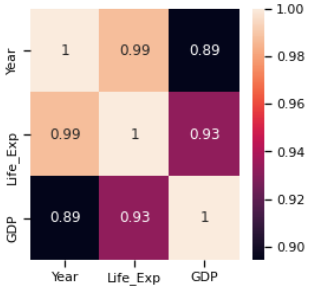
Correlation Heatmap

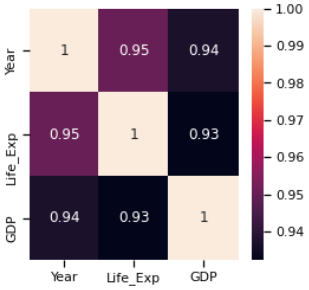












Considering a significance threshold of 0.05, The hypothesis test for USA, China, Chile, Mexico and Germany show that the country's average life expectancy is not significantly different from the global average. However, Zimbabwe's one way t-test demonstrates that the country's average life expectancy is significantly different from the global average.

One-Sample t-test

- Each country's average life expectancy is compared to the global average
- The global average is calculated by taking an average of the dataset after the country being tested is removed.
- Calculation is carried out using SciPy's built in ttest_1samp (from SciPy's stats module)
- Zimbabwe, Chile, Mexico, Germany and the US results from the one-way t-test demonstrate that all these countries have an average life expectancy that is different from the global average, of variable margins.
- It is likely that there is an outlier in the dataset that is misleading the one-sample ttest.
- Zimbabwe's p-value is significantly lower than all of the others.

P-values

- Zimbabwe: 1.4706352812307376e-87
- China: 0.1792814551743123
- Chile: 1.0687862179218956e-07
- Mexico: 0.008303280131667177
- Germany: 4.630744217572992e-09
- USA: 3.027609873792618e-06

One-Sample t-test (excluding Zimbabwe from the global mean)

- Each country's average life expectancy is compared to the global average
- This time Zimbabwe is removed from the global average. This is because we know Zimbabwe's Life expectancy is an outlier of this dataset and we want to examine if removing this data any impact on the p-values of the other five nations and the global average.
- Calculation is carried out using SciPy's built in ttest_1samp (from SciPy's stats module)
- Significance threshold: 0.05. Anything above that value is not significantly different from the global average. Anything under that is significantly different from the global average.
- Results from these one-way t-tests now show that all countries have an average life expectancy that is different from the global average, of variable margins.
- Removing the outlier has changed all of the p-values, but it has not explained why almost every country has shown a significant difference from the global average.

P-values

- Zimbabwe: 1.4706352812307376e-87
- China: 3.4180837175058165e-26
- Chile: 1.8307526230180135e-09
- Mexico: 3.184676093840348e-09
- Germany: 2.2258542168736216e-16
- USA: 0.004051893742867191

Project Findings

Has life expectancy increased over time in the six nations?

The analyses shows that there has been an increase in life expectancy over time in every country, with the highest noted increase in Zimbabwe.

Has GDP increased over time in the six nations?

The analyses shows that there has been an increase in GDP over time in every country. Increases are marginal in both Chile and Zimbabwe while China and the US have enjoyed a marked increase in GDP.

What is the average life expectancy in these nations?

The average life expectancy in Chile, China, Germany, Mexico and US is between 75 - 80 years old. In Zimbabwe the average life expectancy is much lower, averaging around 48 - 50 years old. Zimbabwe shows a much larger range in life expectancy, with the lower end of the range being 38 years old and the higher life expectancy being 78 years old.

What is the distribution of that life expectancy?

The distribution of the data shows a right skewed distribution for Life Expectancy and a left skewed distribution for GDP.

Is there a correlation between GDP and life expectancy of a country?

By calculating the covariance and correlation, we can see there is a clear association between GDP and Life Expectancy. This is clearer when correlation and covariance are calculated at a country level, rather than on the dataset as a whole. the reason for this is likely to be due to the large variance seen between some countries.