Codecademy 'Data Visualization with Python' Capstone: Life expectancy vs GDP

Heather Akers-Healy

Dec 19 · 6 min read

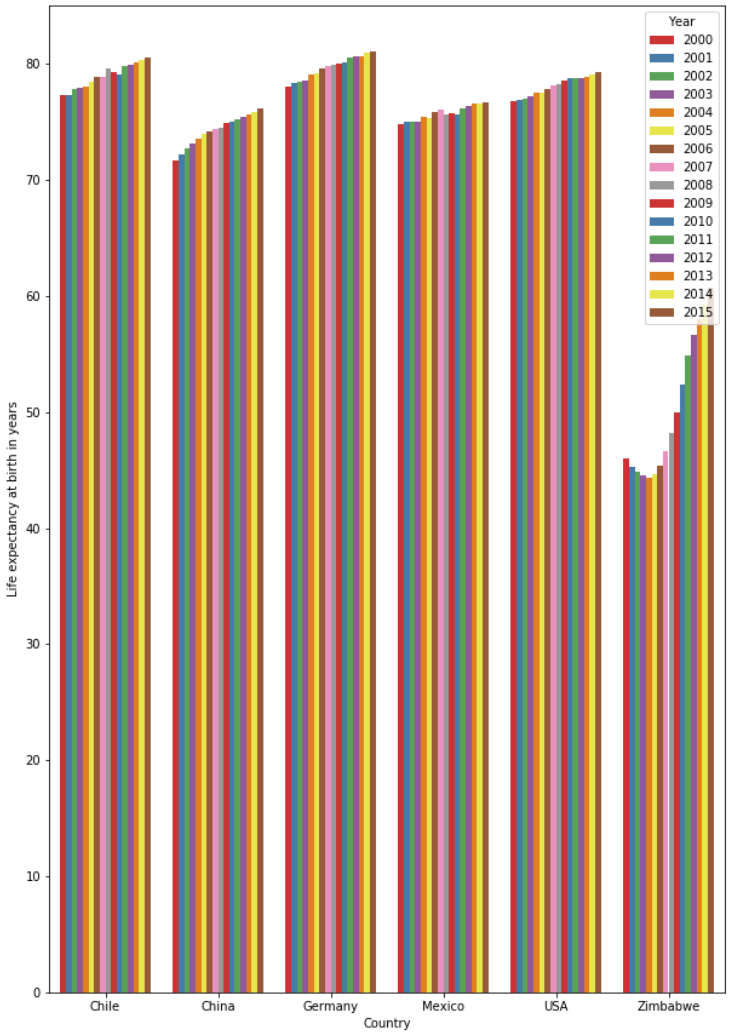
Introduction

This post documents my final project for the Codecademy 'Data Visualization with Python' course. We were given a csv file that contains data on the GDP and average life expectancy for Mexico, China, the United States, Chile, Zimbabwe and Germany. This data was collected by the World Bank, OECD National Accounts and the World Health Organization. We plotted a series of graphs using the Seaborn library for Python and were asked to draw conclusions on the following questions:

- 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?

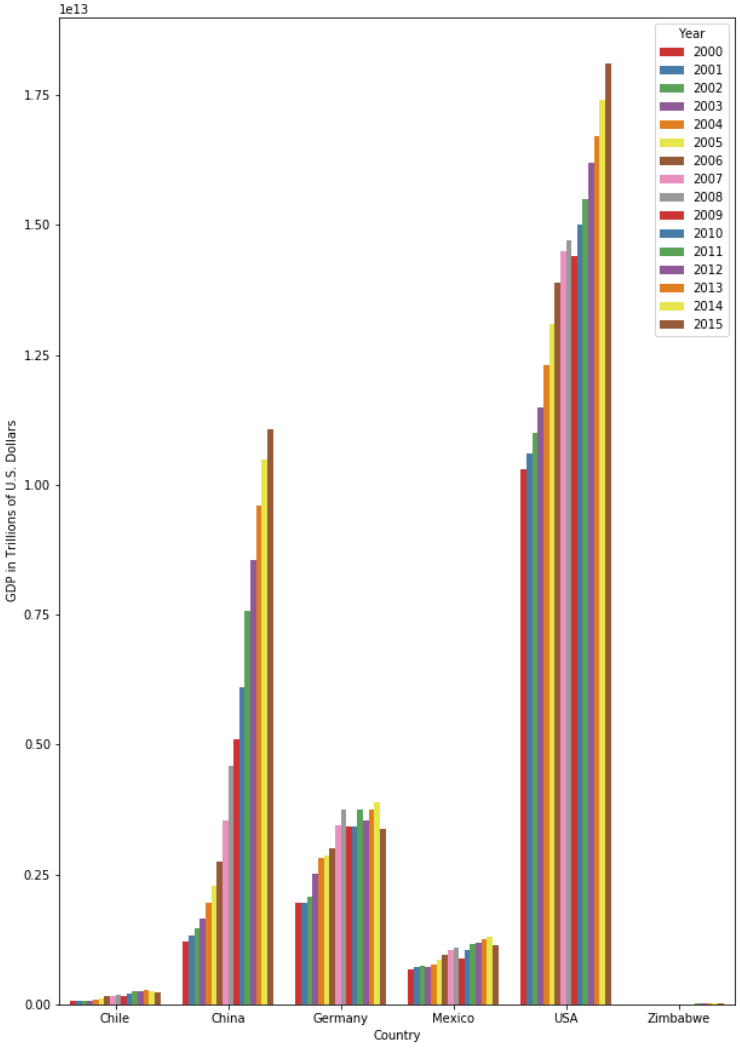
Has life expectancy increased over time in the six nations?

By plotting the data in a bar chart by country and year 2000 through 2015, we can see that each country shows a very clear increasing trend in life expectancy over time. Zimbabwe shows the most dramatic improvement, with an increase from an average life expectancy of around the mid-40's in 2000 to around 60 years in 2015.



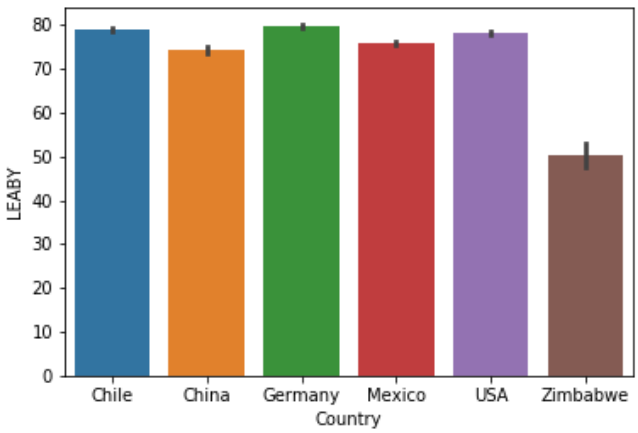
Has GDP increased over time in the six nations?

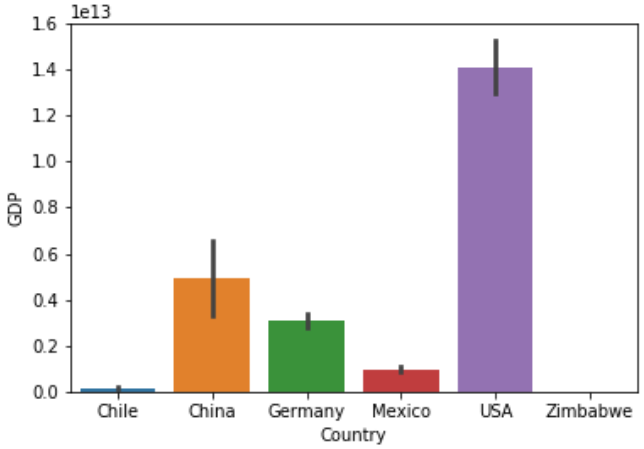
GDP is a monetary measure of the market value of all final goods and services produced in a time period. A bar chart plotting the GDP for years 2000 through 2015 by country shows that each country has also had an increase in GDP over time, though this is difficult to see for Zimbabwe given the scale of the graph.



What is the average life expectancy in these nations? Is there a correlation between GDP and life expectancy of a country?

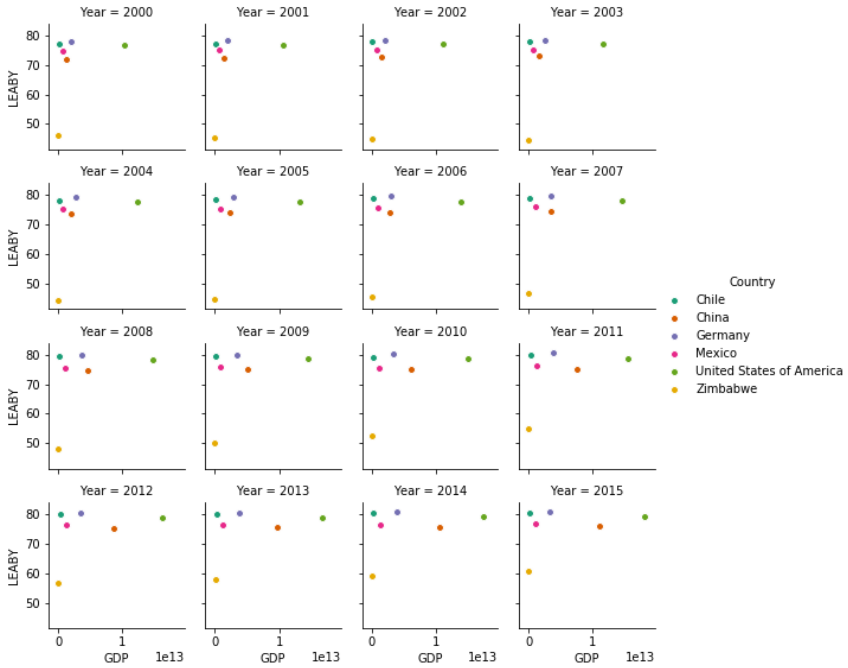
In order to answer this question, we were first asked to plot two simple bar charts of the GDP and life expectancy for each country.



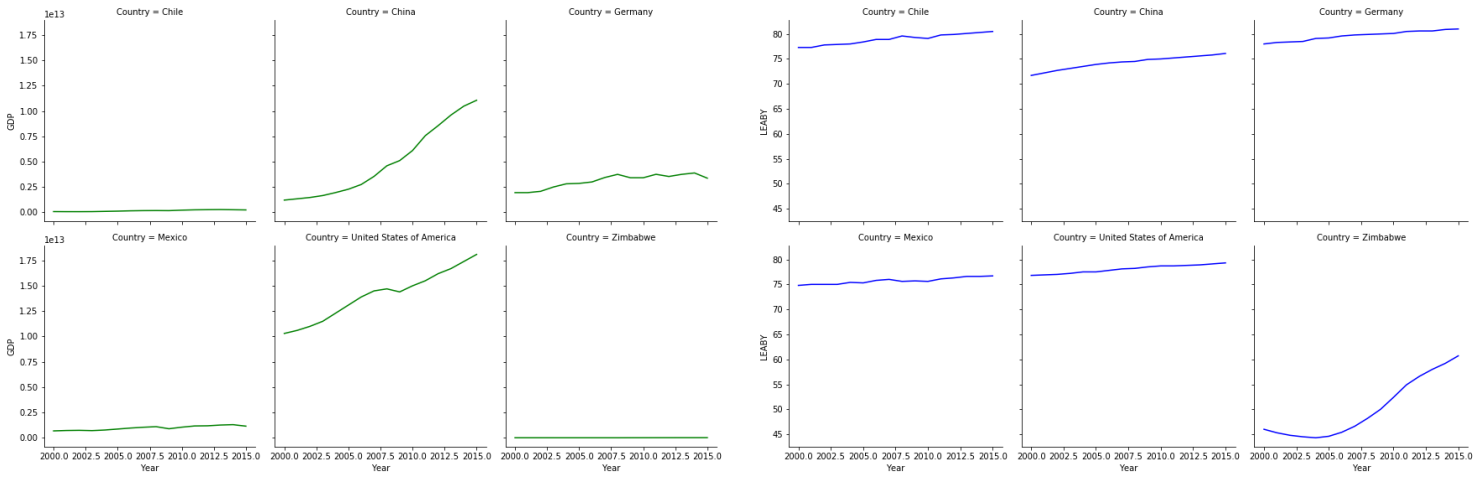


Based on an examination of the two above bar graphs, we might conclude that there is no correlation between life expectancy and GDP of a country. The United States, for example, has both a high GDP and high life expectancy. But Chile has a low GDP and a high life expectancy. This suggests that there are other factors than GDP in the average life expectancy of any given country.

However, if we plot the data using other methods, visual trends emerge that suggest there may be a correlation between life expectancy and GDP. For example, a scatter plot grid over our period of 2000–2015 demonstrates that life expectancy and GDP are increasing together over time. (Also refer back to the bar graphs over time in the first two sections.)



Unfortunately the scatter plots of this data are difficult to read. A better representation of this trend emerges when we plot GDP and life expectancy side by side over time with line graphs.

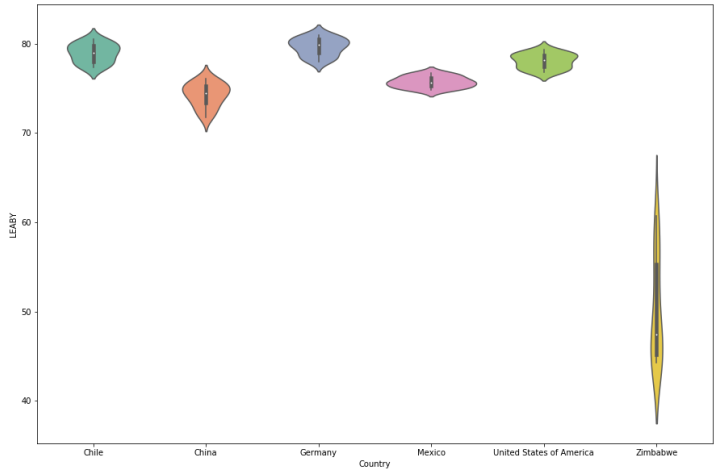


For each country, GDP and life expectancy are clearly in an increasing trend together over time, although in some countries like Zimbabwe, GDP is growing at a much slower rate than life expectancy. (Again, it's difficult to see Zimbabwe's growth given the scale of the graph.)

However remember the rule 'correlation does not imply causation.' Each country has a unique history and social and economic profile that influences both life expectancy and GDP.

What is the distribution of life expectancy?

In order to show the distribution of life expectancy in each country, we plotted a violin graph.



For each country, the length of the violin plot shows the range of life expectancy over time from the minimum to the maximum. The white circle in the center of each plot represents the median for that country, and the black bar represents the interquartile range (the middle 50%). The varying thickness of the plots demonstrates how common any given age is within the data. For example, based on examining the thickest part of the plots, the most common life expectancy from 2000–2015 in Chile was around 80, while in Mexico it appears to have been around 75. Zimbabwe displays the longest distribution, as we would expect from the corresponding bar graph of life expectancy over time.

Limitations of the data and further research

There are a few limitations in the conclusions we can draw from the data. While we can speculate that there is indeed a relationship between GDP and life expectancy, there are compounding factors to

consider. These include income levels by individual, whether a country has universal public healthcare, and access to modern medical treatments.

Revisiting the example of Chile, which has a low GDP and high life expectancy, it appears that their large public health system may be a strong factor in their relatively high life expectancy. Only 19% of Chile's population has private health insurance.[1] In contrast, the United States has a high GDP and high life expectancy, but their average life expectancy is actually slightly lower than that of Chile. The difference here may be that there are more limitations placed on public health care and access in the United States. As of 2015 in the United States, 67.2% of the population was covered by private health insurance, which may not offer the same level of care as Chile's public health system.[2] Compared to other countries, Zimbabwe's life expectancy has increased dramatically, which may be a function of increased access to modern care over time instead of simply an increase in GDP.

Beyond the data limitations, this exercise was also limited by the graphs we were asked to plot. Dual axis graphs or a split scale would help with visualizing data at different scales, such as plotting Zimbabwe's GDP growth on the same graph as the United States. Instead of plotting separate line graphs, we might have instead plotted GDP and life expectancy growth on the same graph to get a better view of the trend lines in comparison to each other.

[1] "Starting in 1990, the civilian government increased public funding, especially for hospitals, without further reform for more than a decade. In the early 2000s, President Ricardo Lagos strengthened the public-sector. Private health care became more and more expensive and as of 2015 covers only 19% of the population, down from a peak of 26%." https://en.wikipedia.org/wiki/Healthcare_in_Chile

[2] "In 2015, private health insurance coverage continued to be more prevalent than public coverage, at 67.2 percent and 37.1 percent, respectively." https://www.census.gov/library/publications/2016/demo/p60-257.html