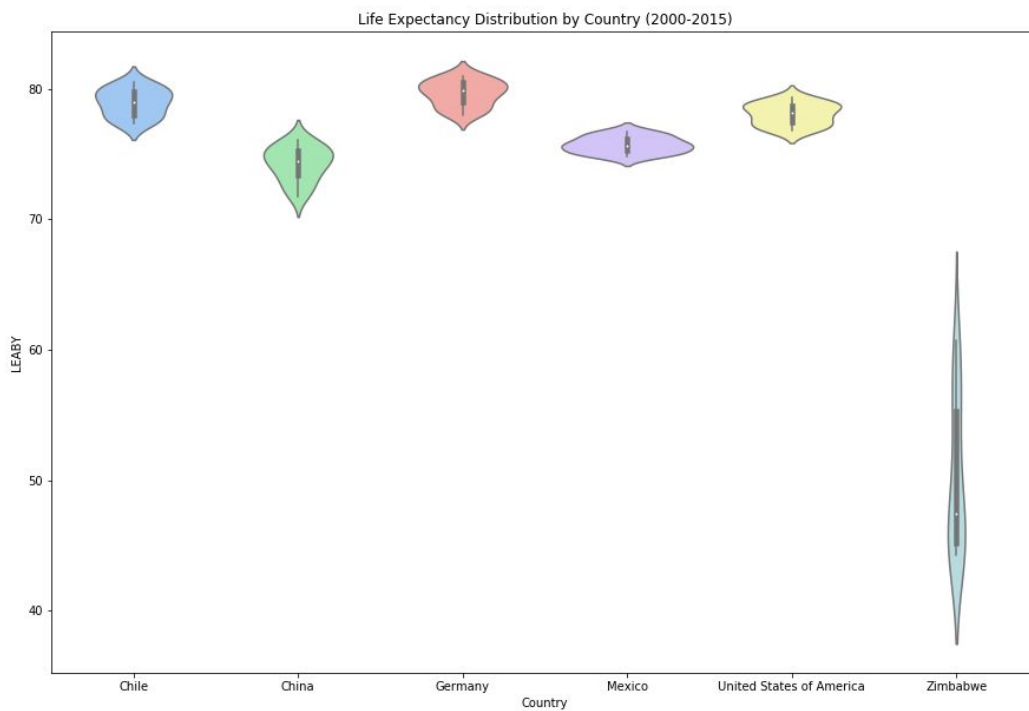


## How I Learned to Better Understand Data Using Python

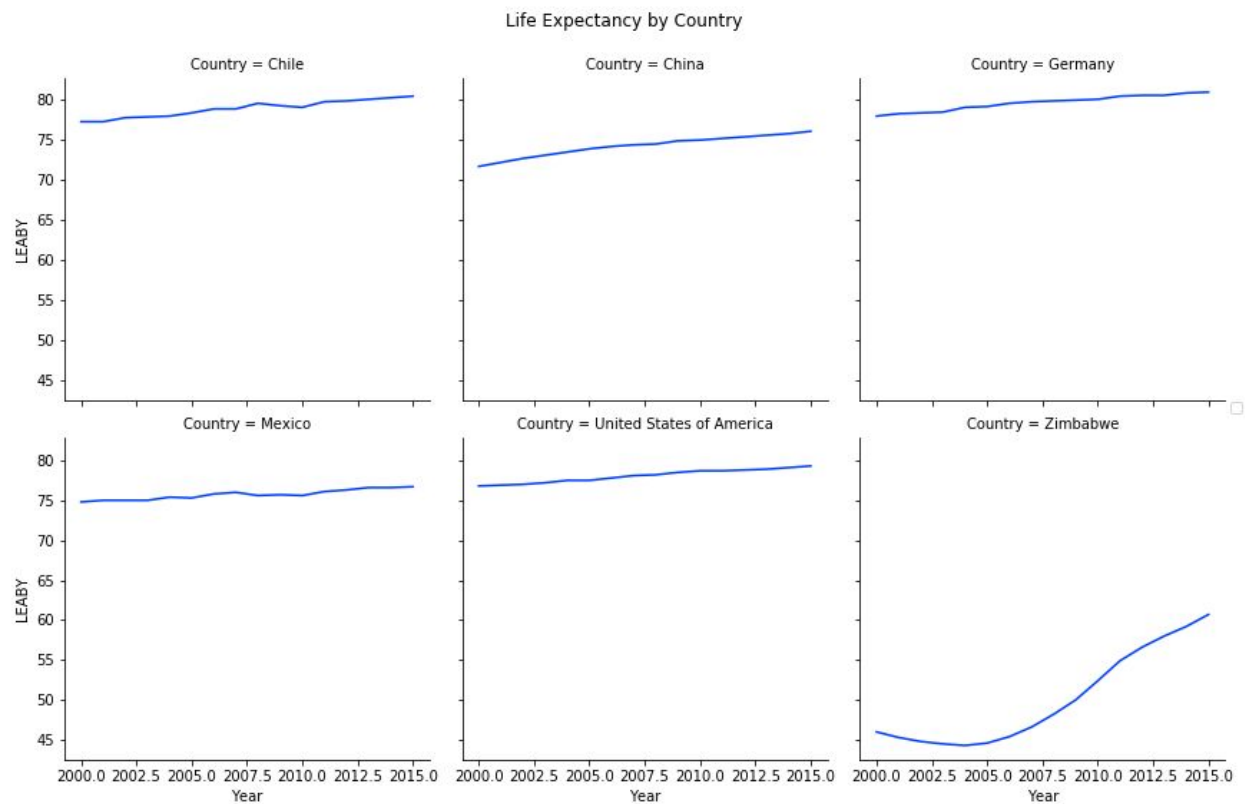
While taking a Codecademy course on “Data Visualization with Python”, I was provided with a dataset for six countries that included GDP and life expectancy for the years 2000-2015. We were asked to perform some analysis using Python to create charts and graphs that help to better understand the data. While this was ultimately a programming exercise, what I found studying the data was fascinating.

The countries included in the dataset were Chile, China, Germany, Mexico, USA and Zimbabwe. While 15 years is a short period in history, there were some pretty drastic changes in some of these countries. A violin plot of the life expectancy data quickly highlights the leaps and bounds in life expectancy in Zimbabwe during this period. The tall, slender life expectancy violin for Zimbabwe shows a drastic change without revealing the direction of the trend. Nevertheless, it stands in stark contrast to the bloated, floating violins representing the other countries.



It wasn't until I plotted the life expectancy data in line chart format that the story of Zimbabwe became clear. The overall trend is a marked increase in life expectancy for Zimbabweans, but there is an interesting dip around 2002-2005 that shows that life expectancy had been previously decreasing. World Health Organization data shows that it had indeed decreased from 65 in 1990 and was attributed to several crises including HIV/AIDS,

malnutrition, malaria among others.<sup>1</sup> It is noteworthy that there is a trend of improving life expectancy across all countries represented in the data, but none so marked as that of Zimbabwe.

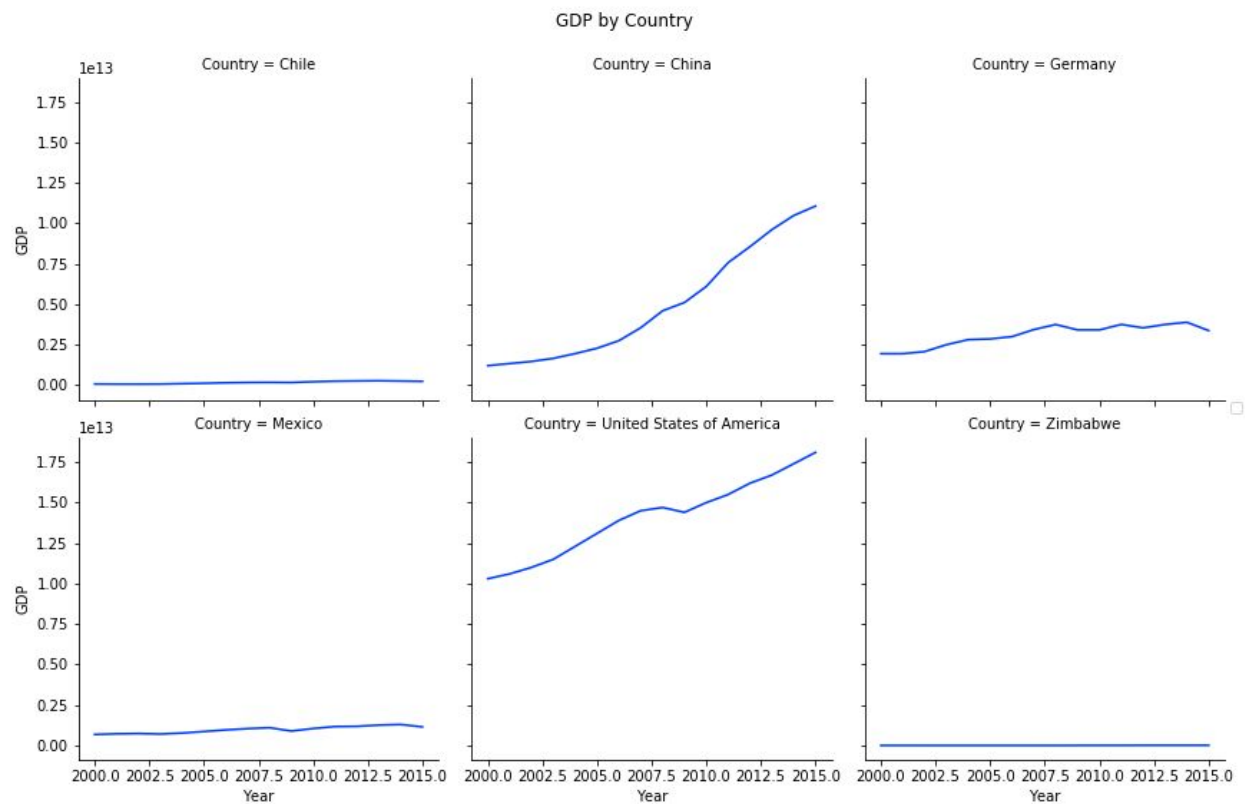


Plotting the changes in GDP over the same time period revealed strong growth for the United States which experienced an almost doubling of GDP. However, China stood out as the biggest success by growing its GDP over 9X from 2000 to 2015. There are several reasons for China's exceptional growth, one of which is heavy state investment.<sup>2</sup>

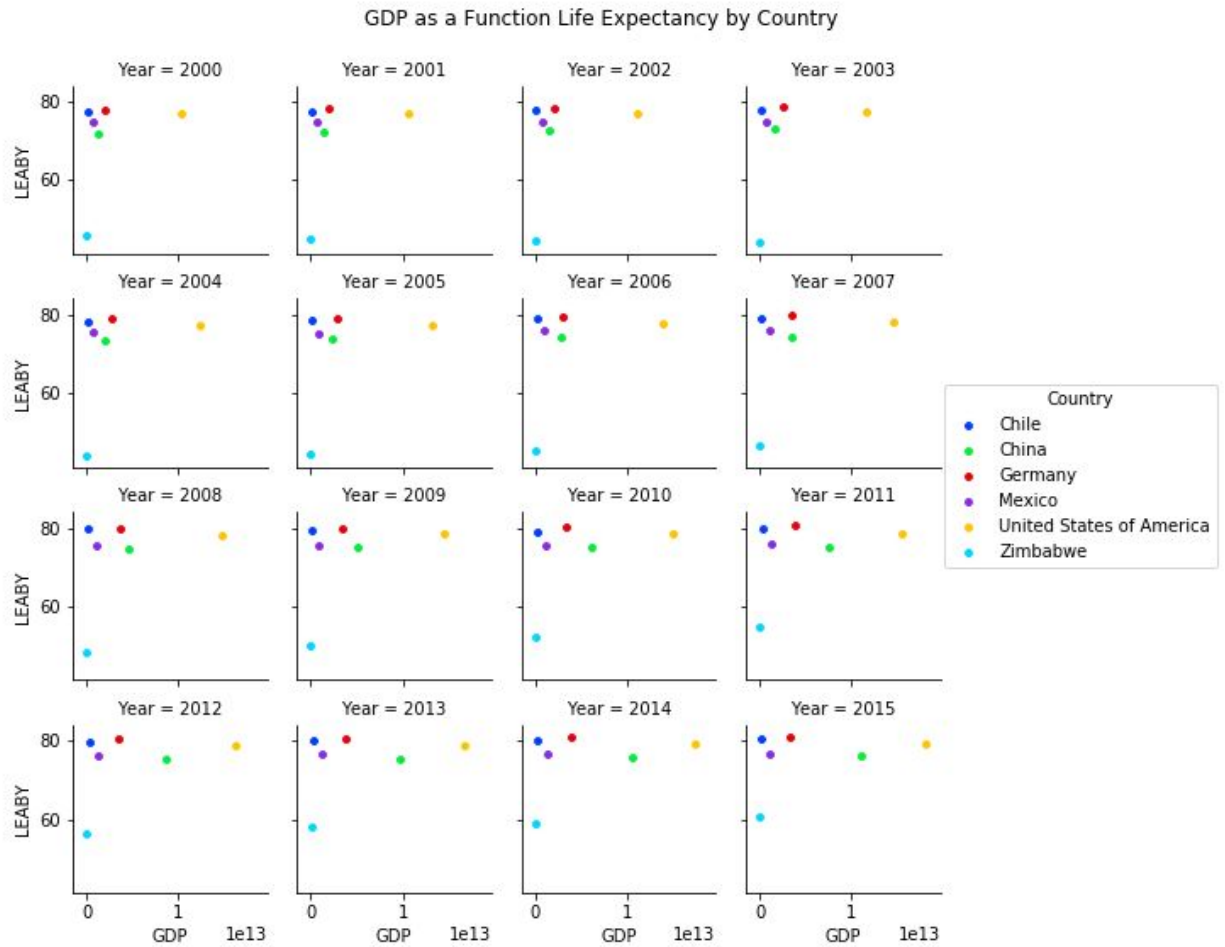
Due to the exceptionally high GDP of countries such as the USA and China, moderate GDP growth in the other countries is somewhat, if not completely obscured. This is a result of the huge scale of the GDP in China and USA as compared to the other countries. Zimbabwe, for example, actually experienced over 2.4X growth of their GDP while Chile's GDP grew over 3X. Plotting these data separately would allow for better scaling and would paint a more accurate picture of their growth.

<sup>1</sup> [http://www.who.int/hac/crises/zwe/zimbabwe\\_profile\\_dec2008.pdf](http://www.who.int/hac/crises/zwe/zimbabwe_profile_dec2008.pdf)

<sup>2</sup> [https://www.huffingtonpost.com/john\\_ross/china-india-growth\\_b\\_11655472.html](https://www.huffingtonpost.com/john_ross/china-india-growth_b_11655472.html)



A juxtaposition of the GDP and the life expectancy data in the form of a scatter plot reveals how trends in both categories play out in different countries. I plotted GDP as a function of life expectancy for each country across all 15 years as a nice way to highlight the marked improvements in life expectancy in Zimbabwe over the same period as the huge gains in GDP in China. While there are many factors at play, and neither data is completely independent of the other, it does lead one to wonder if priorities are a factor. While Zimbabwe faced an extreme health crisis and was forced to work to improve in this area, China focused on strengthening their economy.



All in all, I found this exercise to be rewarding and was able to learn some valuable skills in both plotting and interpreting data.