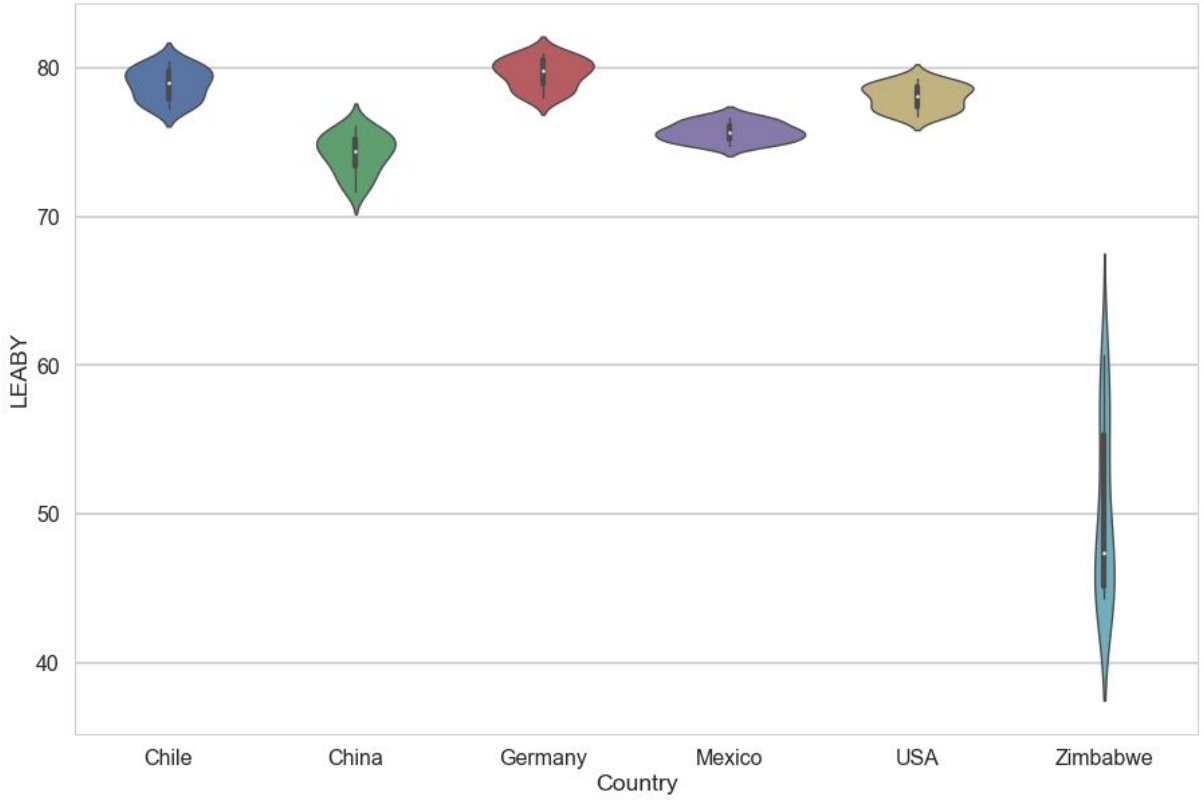
Data Science Can't Replace Brain Usage

Today, we will analyse a bit the correlation between "Gross Domestic Product" and "Life Expectancy at Birth" in different countries as part of a codecademy project. Before looking into the data of half a dozen countries, let's consider, what we should expect. To do so, we begin by having a closer look into the two values, which we shall consider:

Gross Domestic Product (GDP) is the sum of all production, that takes place in a country. While in the details a fairly abstract thing, much of it is quite easy to understand. If you build a car and sell it, the paid price is added to the sum. If you cut someone's hair and get paid for it, your pay is added to the sum. The 'Gross' is there to explain, that no depreciation is subtracted as e.g. in the Net Domestic Product. The Domestic Product is not entirely the same as income, as for example payments to other countries might have to be made, but the difference between Gross National Income and GDP is only large for a small number of special countries like Ireland, where major US corporations run their whole EU business through. GDP is easier and more precise to measure. So is GDP wealth and wealth something we expect to be strongly correlated to life expectancy? Not directly. First of all, it would make sense to divide by the number of people, that live in the country. GDP per capita would be the reasonable measure of wealth to be correlated to life expectancy. Luckily at least the number of people doesn't usually dramatically shift, so at least changes in GDP are probably correlated to life expectancy. However, there is a second thing to consider. GDP is strongly affected by changes of prices. If GDP increases, was it the result of e.g. the building of more cars and fancier hair cuts, or was simply the price increased? To make things internationally comparable in first order - and that was the case in the data of codecadamy - one can use the GDP transformed into a common currency, e.g. the US-Dollar. Currencies with high inflation have a tendency to devalue to the dollar, where the prices are relatively stable. This however doesn't distinguish between major economic crises and currency devaluations, as there are other factors for currency movements as well. The 2008/09 financial crisis had substantial negative impacts in some European countries on life expectancy with a reduction of GDP of ~5%. In the last year of the dataset, which we will analyse (2000 - 2015), there was a reduction in the GDP of Germany, which was not at all the result of an economic crisis, but the result of a devaluation of the Euro with respect to the US-Dollar of actually more than 5%. Such a devaluation is much less harmful than the crisis of 2008/2009 as most product bought and sold in Germany are not imports from the Dollar region, but Eurozone products. Therefore serious analysis would actually use a concept called purchasing power parity. As bad as this works in the real world due to

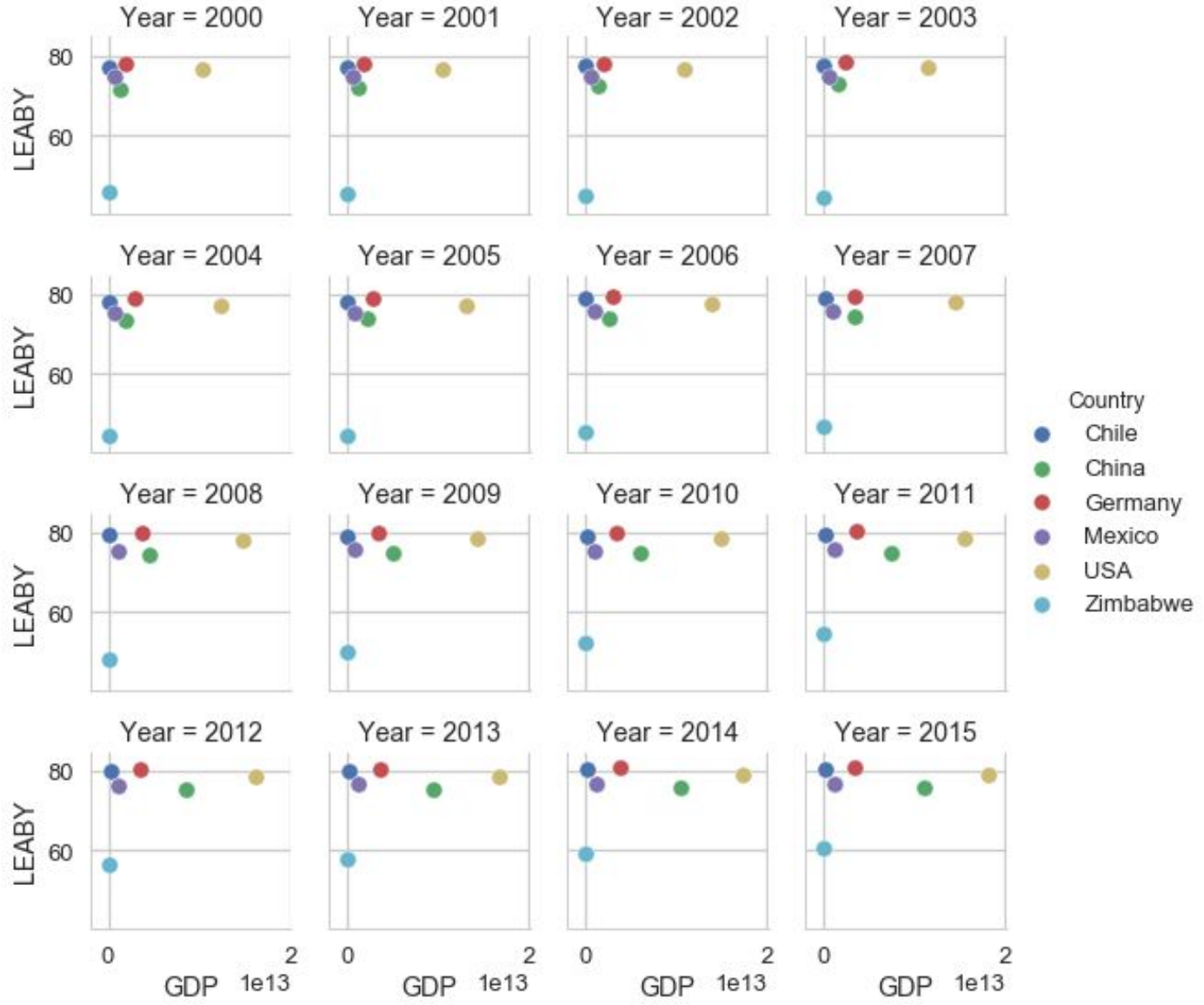
different product baskets purchased, it gives a much better feeling for the actual economic performance of a country. So what we expect to actually be strongly correlated with life expectancy in some corridor is GDP per capita in purchasing power parity units. In the book “The Spirit Level”, the authors show, that up until very roughly 15,000 “USD” in GDP per year, life expectancy is strongly correlated with this value, but then the correlation breaks down. Essentially, once people have the basics, other things become important than pure wealth of the country, e.g. the level of hierarchy and distribution of the income.

Life Expectancy at Birth in Years (LEABY) is more easily intuitively to grasp. One caveat is, that actually early child death has a fairly large influence and especially in very poor countries LEABY is strongly influenced by child death. The actual calculations work with looking into the death rate at different ages and from that calculating an expected survival length. In the first plot we show, we see the variation of LEABY in the time between the years 2000 and 2015 for different countries. Zimbabwe as the only really poor and as well politically quite unstable country is a clear outlier. We will later look into the development over time and then understand a bit more, what is going on.



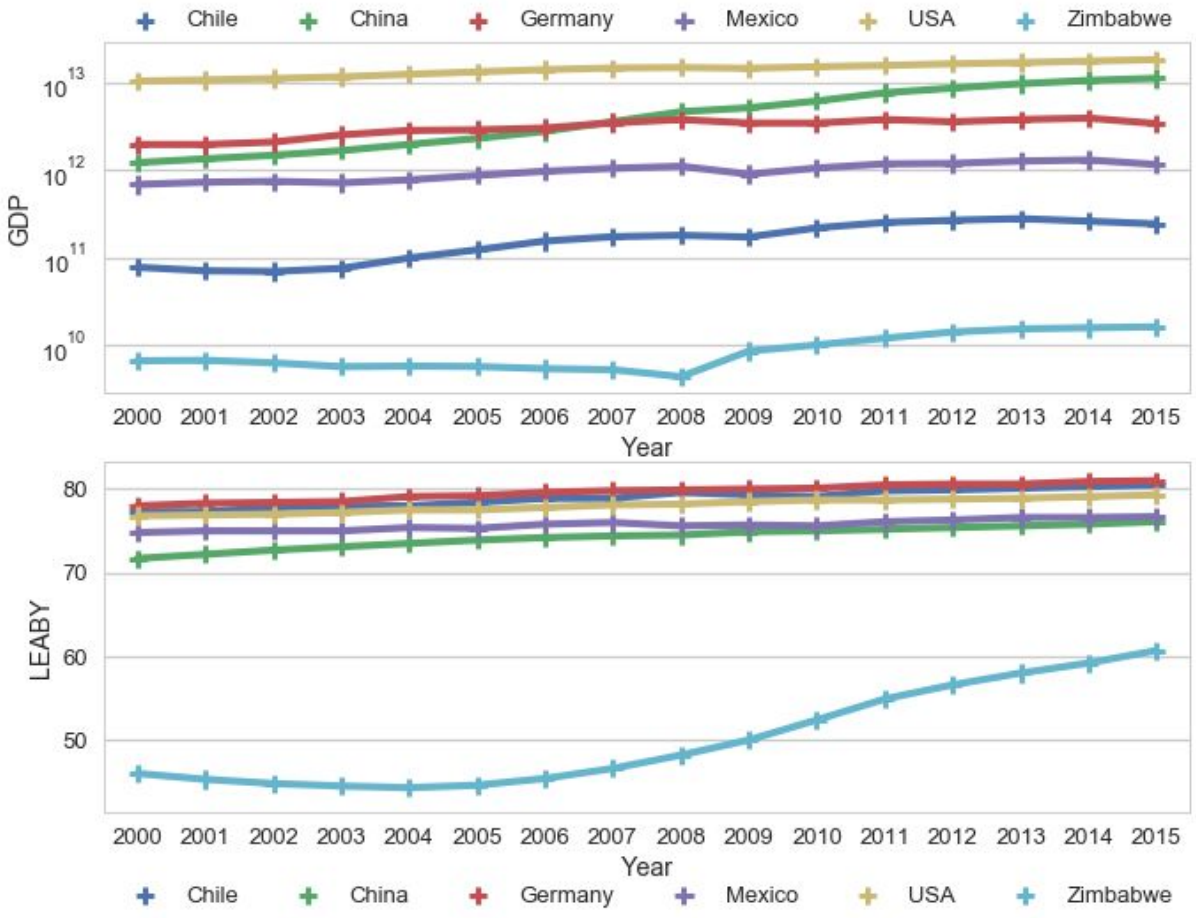
The next set of plots shows LEABY on the y-axis with the x-axis being the GDP for different countries and years. The most visible effects are on GDP in China and LEABY in Zimbabwe. However, as China already had a relatively high

LEABY for its wealth before the GDP growth sprint in the last two decades, there is only a small increase in LEABY visible. On a linear scale, the GDP of Zimbabwe remains so small, that despite substantial growth, it remains invisible…



Finally, we simply want to look into the time series of GDP and LEABY for the chosen countries. As we know, that the absolute value of GDP is quite useless for our purposes, but growth might tell us a bit more given the slow change in population. And as the usual understanding of economics is, that GDP growth is exponential, we resort to logarithmic plotting. Remember: An exponential curve in a logarithmic representation becomes a straight line with the exponent transforming into the steepness. The different colours in the plot represent different countries. GDP is in dollar. 10^{12} corresponds to a trillion, so in 2000 the United States started of at about 10 trillion dollars. The USA is the

first yellowish line and has steady, but slow growth. Germany, the red line, begins with a higher GDP as China, but then falls below. Remember, that the final dip here is a currency devaluation effect, not a reduction in real growth. China is the most spectacular as consistent large grower. Mexico and Chile more or less grow similar to USA and Germany. Zimbabwe first shrinks and then grows rather rapidly. Despite 2008 being a year of world wide economic crisis, the 2008 change here is positive and mostly correlated to domestic politics with increased stability after 2008.



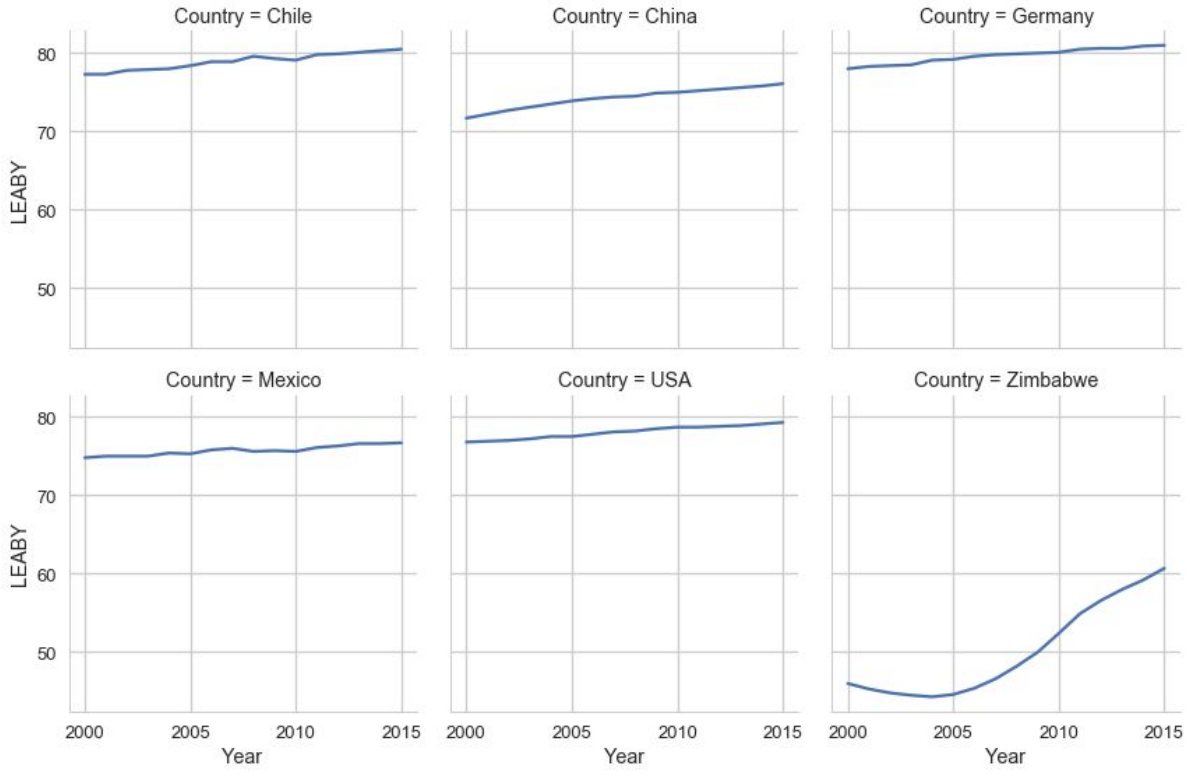
In the LEABY plot, we see Germany, Chile, and USA performing similarly, starting with already large values and slowly growing further. Actually the USA is slightly underperforming on this metric despite a more consistent GDP development. The reason is mainly to be found in the inequality as demonstrated convincingly in aforementioned book “The Spirit Level”. China is still poor enough, that the large growth corresponds as well to a faster increase in LEABY relative to the more developed countries and we can expect, that China will reduce the GAP to the USA further in the future, if GDP growth continues as in the last two decades. Mexico however, demonstrates the reason for the fear of the “middle income” trap, which in reality however is mostly a statistical artefact. Mexico is as well still poor enough, that strong GDP growth would likely lead to higher LEABY, but Mexico isn’t growing

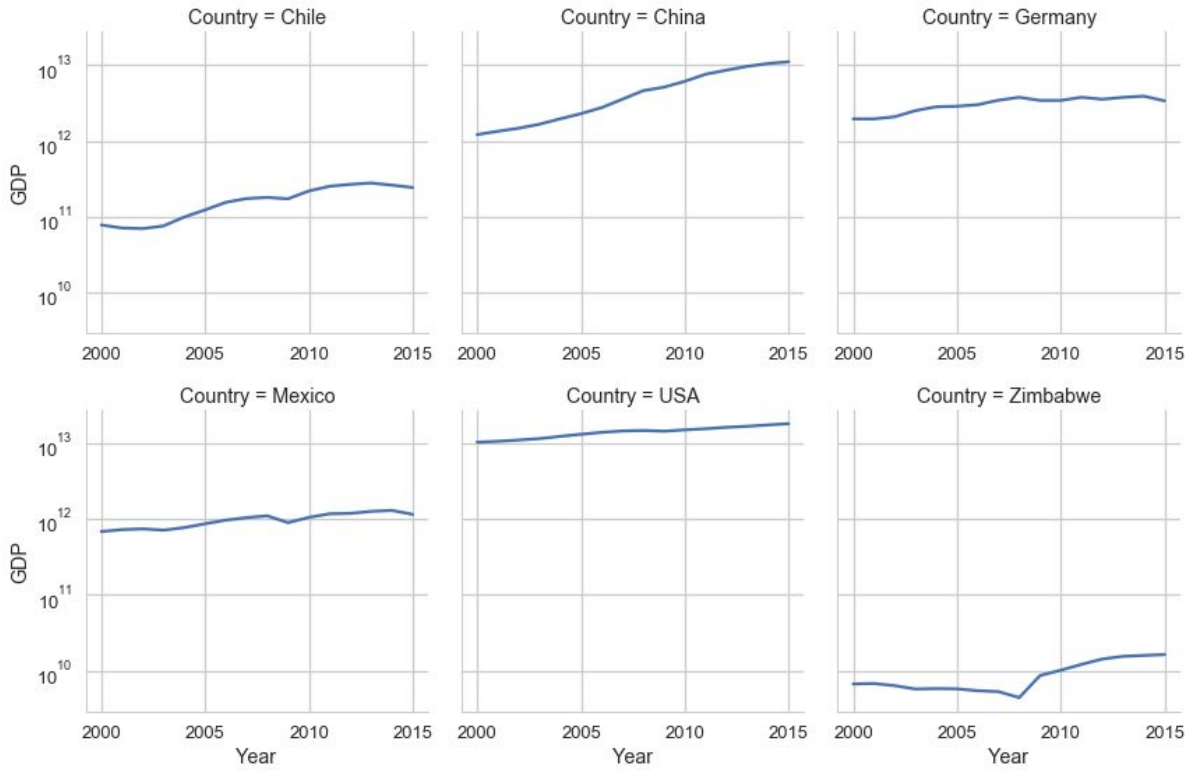
sufficiently at the moment any more. Without a deep analysis only scanning international news, the main reason for both sluggish growth of GDP and LEABY seems to be gang induced violence.

Zimbabwe's economy slumped until about 2008. It actually turns out, that in that year a power sharing agreement between the dictator Mugabe and an opponent came into being, that probably reduced the political risk, leading to a growing GDP. LEABY already had a few years of growth before albeit from a very low level and after a decline at the beginning of the millenium due to a power struggle, which Mugabe in the end won.

Overall, the analysis using GDP at market rate turns out to be clearly flawed and the kind of conclusions, that can be drawn are very limited. GDP per capita at purchasing power parity would have helped to give a clearer picture of the relation between wealth and health and helped more to identify times of economic crises and their potential influence on LEABY. Such understanding however requires a broader understanding of real world events and purely focusing on analysing inadequate data series can't replace this kind of more thorough way of thinking.

Backup with further plots:




Country = Chile
Country = China
Country = Germany
Country = Mexico
Country = USA
Country = Zimbabwe
GDP
10^{13}
10^{12}
10^{11}
10^{10}
2000
2005
2010
2015
Year