

GDP and Energy

DNSC 6211: Programming for Analytics

Wendy Zhang

Abstract

My project is intended to analyze the relationship between energy use and GDP. I wanted to explore if there is an apparent linear relationship with GDP, and if so, how it varied across the time. I collected access to electricity and GDP data for all the countries from the World Bank. I plotted the data in different ways to analyze the relationship between these two variables. I was expecting that countries with higher national GDP would have more access to electricity. This means that the two variables should have a positive relationship. However after plotting it I was unable to see a clear pattern that defines the trend. To further investigate the issue I added energy use per GDP unit to my data set to obtain more insight into the relationship between energy use and GDP.

Contents

1	Introduction	3
2	Background	3
3	Method	3
4	Organization	3
4.1	Workflow	4
4.2	Project structure	4
4.3	Figures and Tables	5
5	Discussion	9
5.1	Learnings	9
5.2	Challenges	9
6	Conclusion	10

1 Introduction

Intuitively countries with higher GDP are generally well developed countries, or at least better developed countries, even after taking the population into consideration. We know that there are many parts of the world where people still don't have access to the basic resources, i.e., food, water, electricity. I wanted to examine whether we can predict access to electricity using a country's GDP. Or at the minimum be able to see a consistent increasing trend of the access with the increase in GDP. Furthermore to assist with exploring the relationships I added energy per GDP use data to examine how efficient and/or proportional are the two variables in relation with each other.

2 Background

I started with the questions and assumptions outlined in the introduction section and searched for dataset that would serve the purpose intended with the analysis. I chose the datasets due to a variety of factors that include an effort to make relevant comparison and making sure data is available within the same time frame, and an intent to keep the data consistent and complete. I didn't give up on any dataset. On the contrary I added one more variable to my dataset as I was performing the analysis. The result of the analysis, while unable to confirm a straightforward and consistent linear relationship I was looking to see, did provide some insight into the trend and elicited the questions and observations that there may be many other variables and factors that would affect the relationship. I was able to better understand the relationship by graphing all three variables separately and looking at the overall trend.

3 Method

I intended to investigate the relationship between GDP, access to electricity and energy use per unit. As discussed previously, I expected to see an apparent relationship, i.e., I expected the access rate to go up with higher level of GDP. However, I did not expect to see this relationship throughout the entire time period under observation. For example, there could be fluctuations in the trend due to a variety of other factors that were not taken into consideration in our study. I observed that access to electricity increased dramatically at higher GDP levels, however, this trend is not increasing at a constant rate once GDP reached certain levels. At lower level of GDP there were a lot of fluctuations in access data. While I was expecting such fluctuations and anomalies the extent of such are not explained by the data I have. GDP vs energy per GDP unit had many fluctuations as well. At higher GDP levels energy use per unit wasn't consistently higher. Nonetheless lower GDP levels exhibited much more and concentrated fluctuations.

4 Organization

Interworking python and R

I wrote all my functions to pull and analyze the data in python. Then I import ggplot in Python and used it to create the line graphs. My python code also generated a csv file called df.csv, which was used to read in R to create the world heat maps. So I was able to use R plotting functions within python to see the graphs. In addition to that I wrote separate code in R for plotting as supplements to what I had created in python.

4.1 Workflow

The following diagram shows the project workflow.

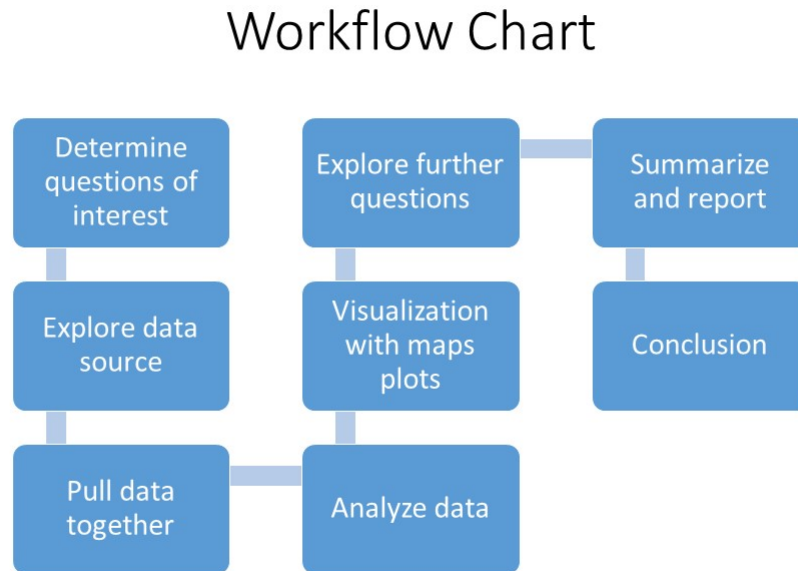


Figure 1: The project workflow

4.2 Project structure

I retrieved my data from the World Bank API. I used indicators to read data into a data frame then store it in a database in preparation of my analysis. There were multiple years of data available for GDP but there is only year 2012 data for access to electricity. Therefore I had to limit my analysis to 2012. While analyzing the relationship between GDP and access to electricity, since I wasn't able to see a clear trend I wanted to explore if some relationship can be obtained by adding another variable and whether the observed trend would be consistent with what I had already seen. After deciding to add an additional observation variable: energy use per GDP unit, I retrieved 2012 data for all countries that had available information.

<http://data.worldbank.org/indicator/EG.ELC.ACCS.RU.ZS>

I obtained GDP and energy use data from the links below.

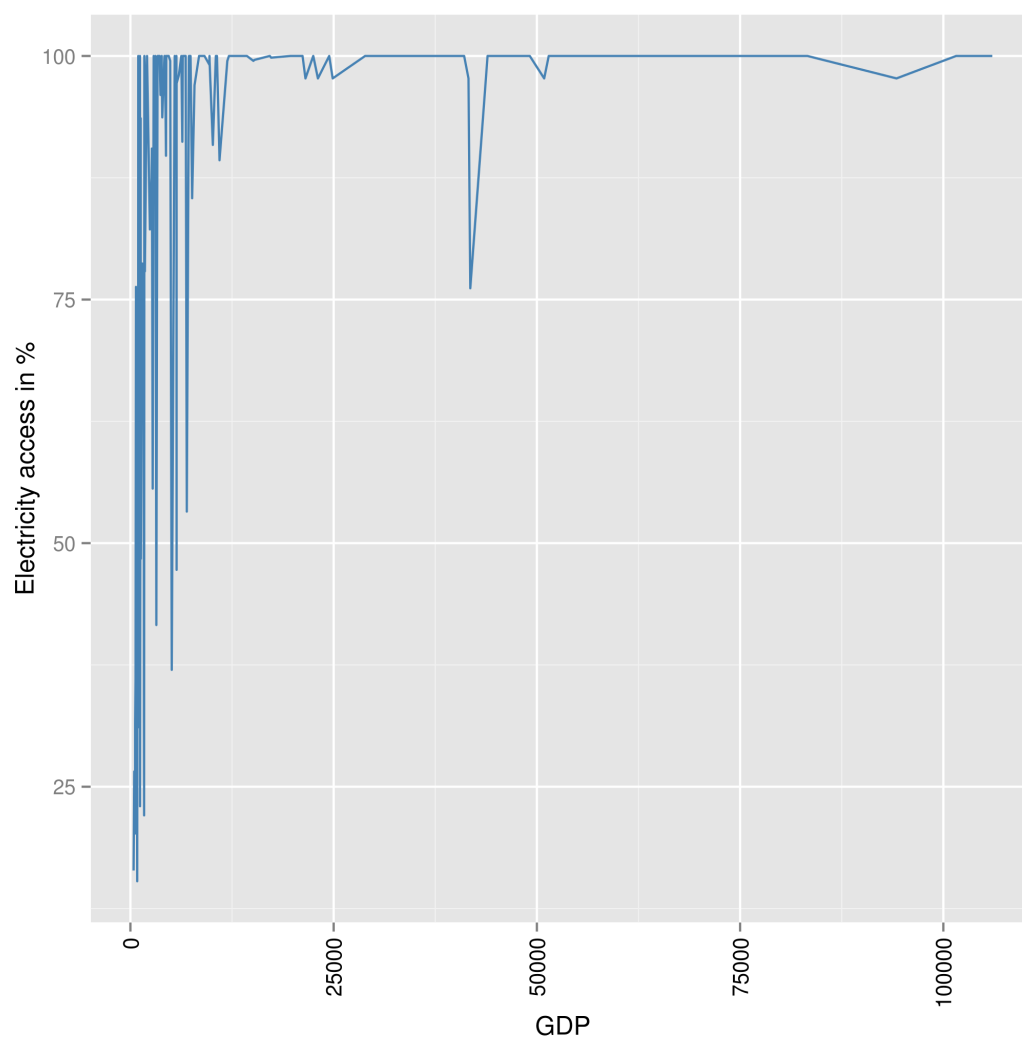
<http://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD>

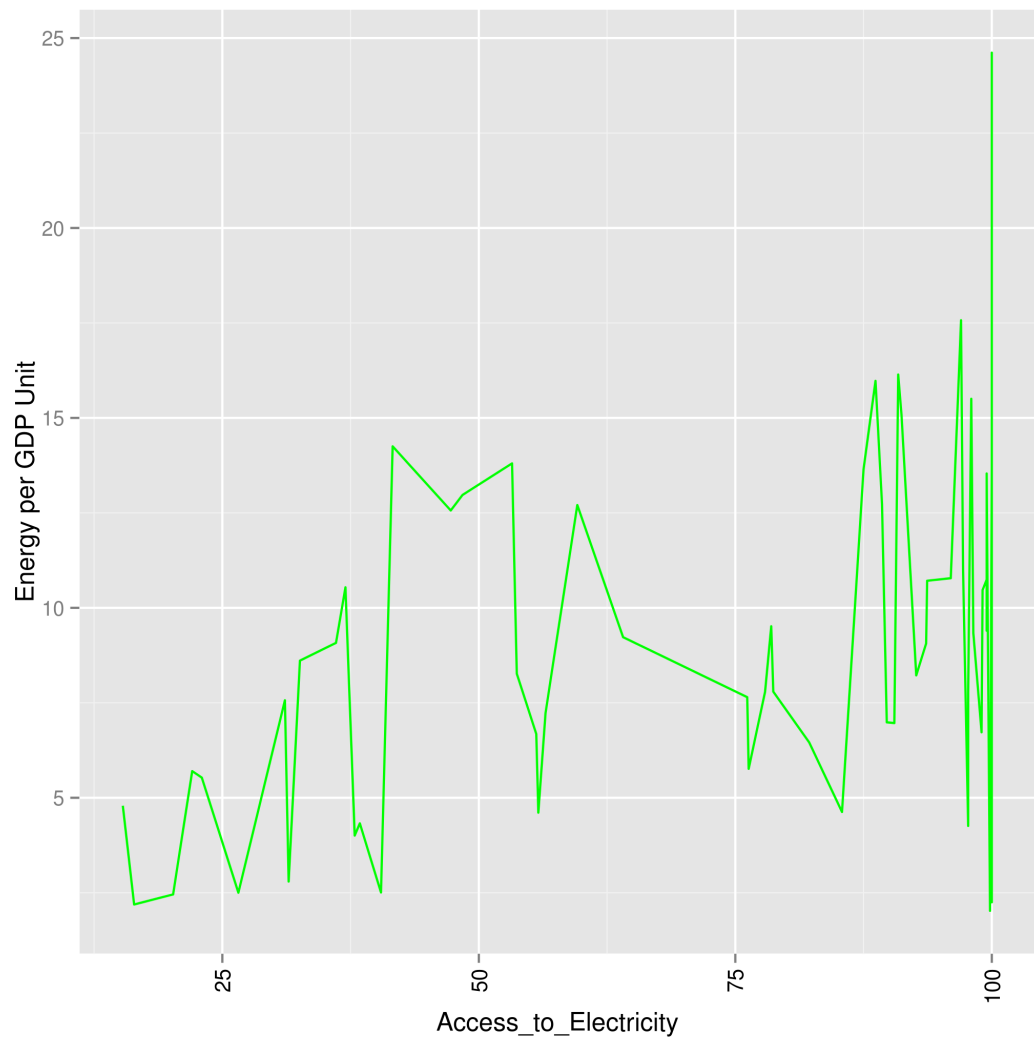
<http://data.worldbank.org/indicator/EG.GDP.PUSE.KD.PP.KD/countries>

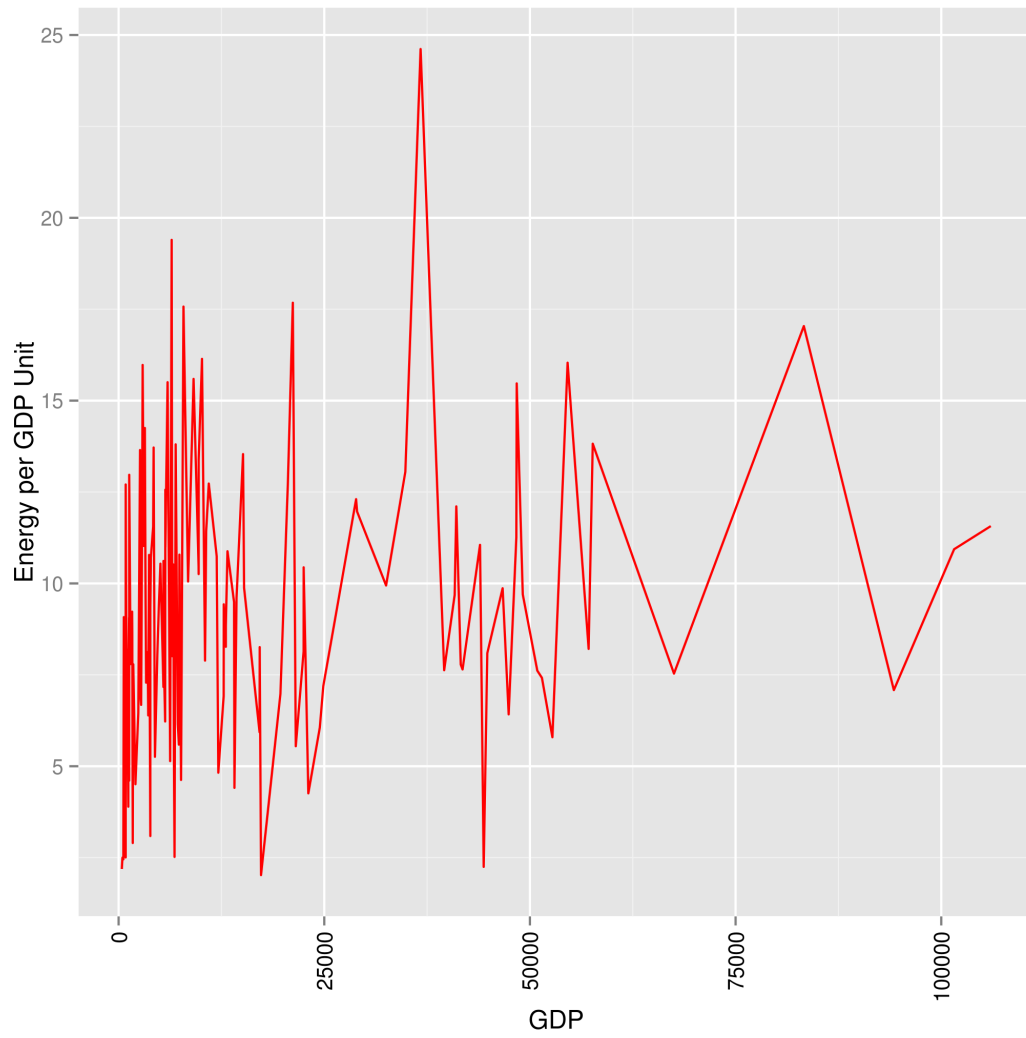
My github link containing all the files is below:

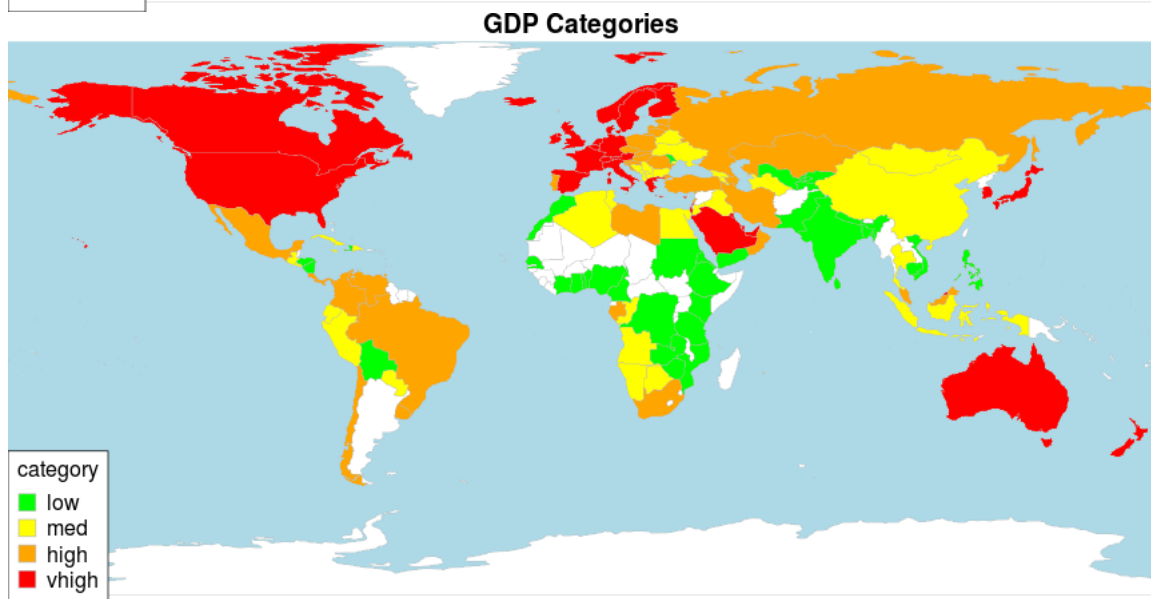
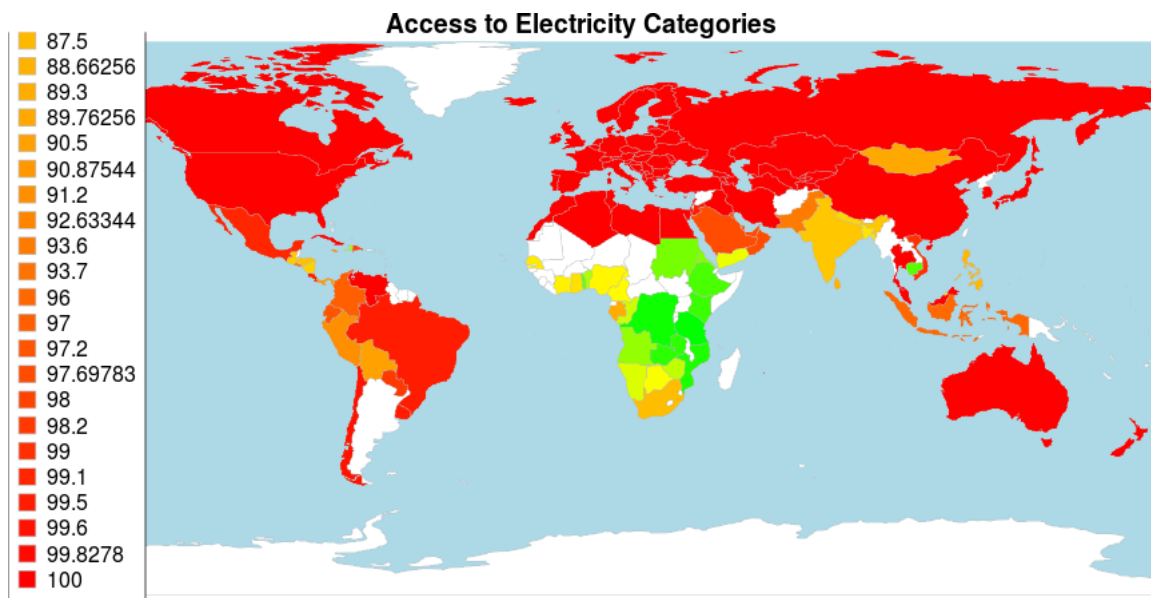
https://github.com/Bobo1980/Individual_Project

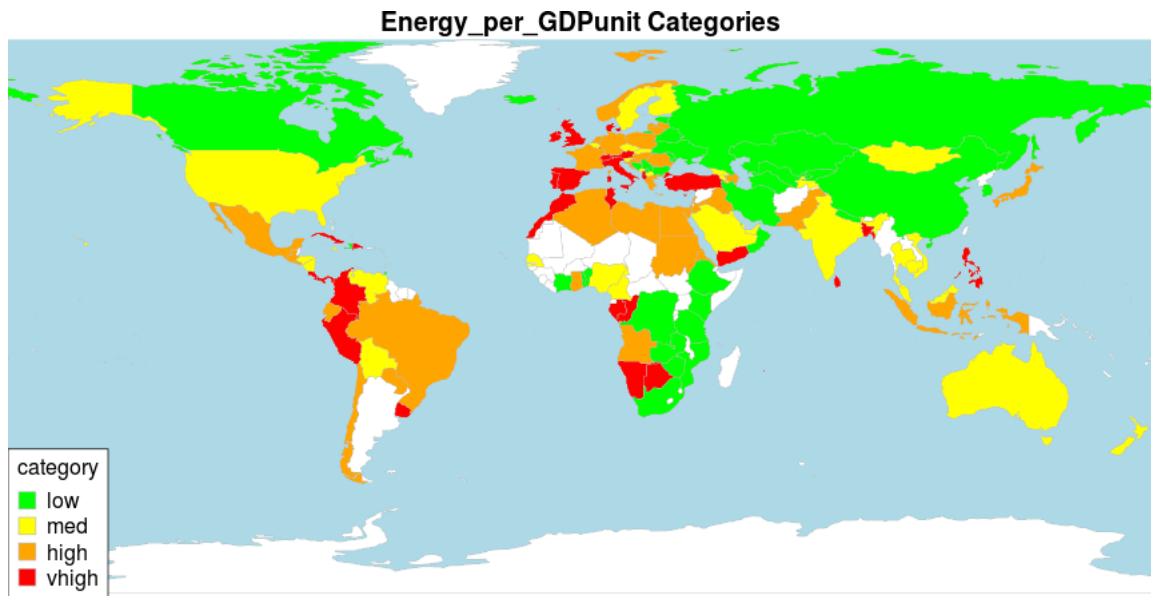
4.3 Figures and Tables











The figures above show the analyses performed and their results that were previously discussed. The heat maps indicate the spread and differences among all the countries for the variables that was exploring.

5 Discussion

Although my project didnt find a conclusive, clear and apparent relationship, it did provide valuable insight while exploring the relationship. There could be so many other variables and factors that affected and will continue to affect the fluctuations and the extent of them. It provides proof and opportunity to further explore the questions and elicit the question of what other data we could be and should be looking into if we were really to understand the effect of GDP on those two variables and how they are affecting each other. Higher GDP may or may not mean more access to resources or more efficient use of them, or at least not in a very consistent and expected manner. Given this observation, what can we do and how can we reverse the trend? I think this is one of the main selling points of my research. The other is that I also used R to plot the world maps of GDP level, access to electricity and energy use per unit. This gave me a clear visual image of the geographical spread and whether they are consistent.

5.1 Learnings

One of the better moments was being able to utilize R to create world heat maps that allowed visualization of the data observe the relationship among the three variables I was exploring.

5.2 Challenges

One of the challenges was getting enough data for the question I was interested in. I only had 2012 data for access to electricity and that limited what data I could use for GDP and energy use per GDP unit. Although an apparent trend wasnt obvious in this research it is entirely

possible that given enough data for multiple years we might be able to observe something more conclusive. I was hoping to be able to plot multiple year but the data source was limited.

6 Conclusion

While GDP levels seem to have an overall positive relationship with access to electricity and energy use per unit, the extent of fluctuations and condensed aggregation at lower GDP level make it conclusive. I did see a positive relationship but it was not consistent. In conclusion, given the data I was able to find some explorations and findings were useful to further research to better answer the question and determine the relationships. To further investigate the questions of interest we will need to obtain more data and include more factors in our analysis to get a clearer picture.