Worldwide Energy Use by GDP

DNSC 6211: Programming for Analytics

Amit Talapatra

December 9th, 2015

Abstract

Climate change, and its relation to energy use, is a concern for all of the countries of the world. Both large and small economies are considering energy policy decisions that balance promoting growth in their economies with the environmental damage caused by some readily available energy sources. This analysis uses data provided by the World Bank to visualize the relationship between GDP (Gross Domestic Product) and the types of energy sources that make up a country's energy needs. Looking at 106 countries, this analysis uses k-means clustering to summarize 22 years of energy use data against GDP as a line plot for each of three energy use categories: Alternative/Nuclear Energy, Combustible Renewables and Waste, and Fossil Fuels. The charts presented in this report can be used to illustrate how energy use by type differs in economies of varying size.

Contents

1 Introduction

The week that this report was submitted, delegates from 196 countries met in Paris for the 2015 United Nations Climate Change Conference. The goal of this meeting was for participating countries to develop a binding agreement to address climate change at a global level. Although climate change is a global problem, developing countries are more likely to face the brunt of its effects. Countries with less developed economies are disproportionately affected by sea level rises compared to more developed areas, and often lack the infrastructure to manage rising sea levels and other weather phenomena attributed to climate change. The World Bank has estimated that hundreds of millions of people could be displaced by sea level rise in the next century (Dasgupta, 2007). Given this threat, both developing countries and wealthier nations must consider the pros and cons of fossil fuels and low carbon alternatives. Fossil fuels, which emit CO2 in combustion and contribute to climate change, also present a readily available way to build an economy on a less expensive energy source, and were arguably a key contributor to the current economies of many developed nations, but this benefit must be weighed against the damage fossil fuel use could do in the long term. For economies in developing nations to grow, they need reliable sources of energy. This poses the question: what types of energy sources do economies of varying size use?

This report examines the relationship between national Gross Domestic Product (GDP) and energy use proportions by type for three different categories recorded by the World Bank using a sample of 106 nations. The goal of this analysis is to distill from this data the differences in national energy profiles over a wide range of economies.

2 Background

To investigate the relationship between national economies and energy use proportions, this study used the World Banks data domain (data.worldbank.org), as it is one of the most complete sources of country data, and as this is a topic that the World Bank has conducted research on.

Many countries in this dataset were missing data on their energy use proportions. Choosing the subset of data for this analysis required balancing the size of the timescale such that data would be available for enough countries to show a relatively complete picture of the world. Choosing a larger timescale would mean more countries would have missing data and be left out of the analysis, but choosing a timescale that is too small might produce results that do not accurately reflect the relationship between energy type and economy size. The timescale chosen for this analysis was 1990-2012, which meant that 106 countries had complete datasets and could be included in the analysis.

After the missing datasets were removed from the sample, the plots produced had too many data points to see a clear relationship between energy use and GDP. The cast majority of the countries selected were in the bottom 10% of the GDP scale. To condense this into a single line plot, a k-means cluster method was used to group countries into distinct sub-groups based on GDP. This made it possible to find an average GDP and average energy use proportions for each subgroup, allowing for a chart that shows a single line for each of the three energy types examined.

3 Method

The primary goal of this project was to produce a plot showing how the three energy proportions (Alternative/Nuclear, Combustibles, and Fossil Fuels) differ for countries with economies of varying size. While this analysis succeeds in distilling the mass of data into a simple line plot for each of the three energy source types, these trends are significantly skewed by outlier countries, and there is a wide range of external factors between a nations GDP and its energy use profile. For example, countries like the US and Japan skew this relationship heavily on the high-GDP end, and Iceland significantly skews the trend for alternative energy on the low-GDP end. A question that should be explored for a follow-up analysis is: how do these trends look when normalized by region and countries with similar access to natural energy sources. This analysis instead focuses on the preceding step, comparing GDP and energy use among all types of countries.

3.1 Workflow

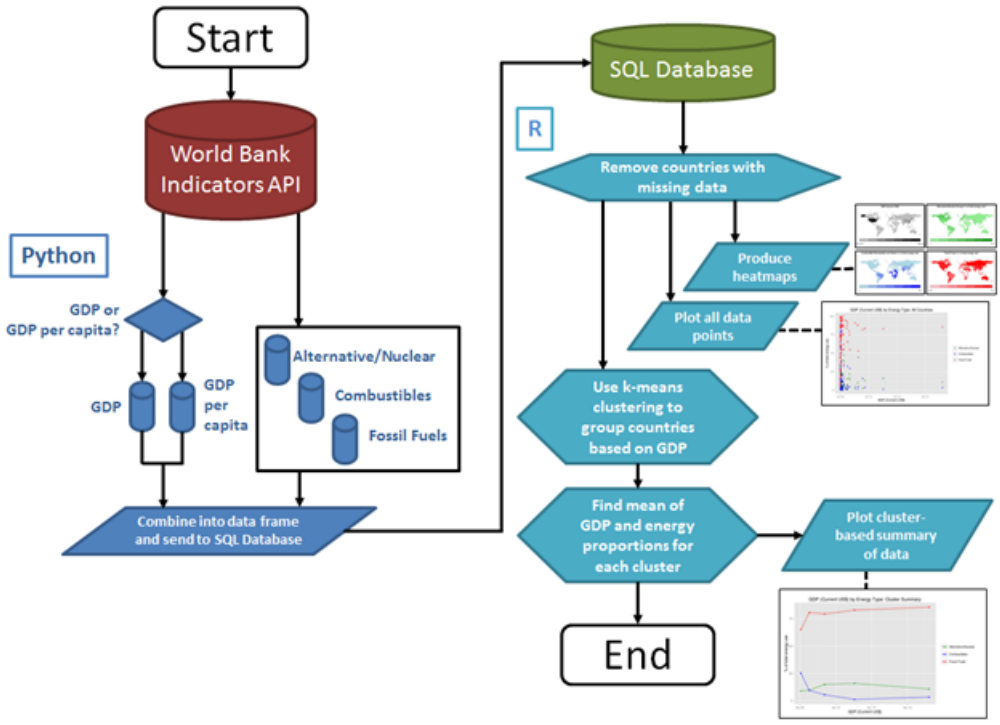

Figure 1: Project Workflow

The project workflow is illustrated in Figure 1. This project demonstrates pulling data from an API, interworking Python and R code, using a SQL database as a local repository of data, generation of heat maps in R, plotting using the R library ggplot, and the use of a k-means cluster analysis to summarize a larger dataset.

First the data was pulled from the World Bank Indicators API using the pandas Python

package. The python code was originally designed to use GDP data, and was also modified to use GDP per capita. Both results are used in this analysis. The GDP data was then put into a data frame alongside energy use percentage data for the three categories: Alternative/Nuclear, Combustibles, and Fossil Fuels. This combined data frame was then stored in a SQL database for easy local access.

The next stage was exploratory data analysis. R was chosen to visualize the data because of the availability of tools like ggplot and rworldmap. To visualize the data, first heat maps were created to show the concentrations of countries with high GDP, and to highlights the countries with higher dependency on each type of energy source. This exploratory analysis was accompanied by plots of all of the energy use data points against national GDPs. As this data did not show a clear relationship, an additional task was set to create an average line plot for each energy type. Doing so required subsetting the countries by GDP so a k-means clustering method was used to create five distinct subsets of nations. For each one, an average was created for GDP and energy use data and plotting these points formed the final line plots summarizing this relationship. This analysis was run for both GDP and GDP per capita.

3.2 Project Structure

This analysis used the following datasets from the World Bank Indicators Database. Each title is followed by its corresponding indicator key. These keys can be used to find their corresponding datasets on data.worldbank.org. GDP, or Gross Domestic Product, is used to quantify the size of a nations economy. GDP per capita divides GDP by a countrys mid-year population, effectively normalizing a nations economy by the number of people that contribute to it. The three energy source datasets present a clear way to classify a nations energy use with respect to its effect on carbon emissions and its environmental impacts.

GDP (current US$) - NY.GDP.MKTP.CD

GDP per capita (current US$) - NY.GDP.PCAP.CD

Alternative and nuclear energy (% of total energy use) - EG.USE.COMM.CL.ZS : Energy sources that do not produce carbon dioxide for energy generation. Includes hydropower, nuclear, geothermal, and solar (among others).

Combustible renewables and waste (% of total energy) - EG.USE.CRNW.ZS : Includes energy from industrial and municipal waste, solid and liquid biomass, and biogas.

Fossil fuel energy consumption (% of total) - EG.USE.COMM.FO.ZS : Includes coal, oil, petroleum, natural gas, and derivative products used in energy production.

3.3 Maps and Plots

3.3.1 Exploratory Data Analysis

For exploratory data visualization, this analysis began with a number of world heat maps to highlight countries that are strong outliers in the variables of interest. Note that some countries are not colored in as they did not have complete datasets for the date range (1990-2012) in this analysis.

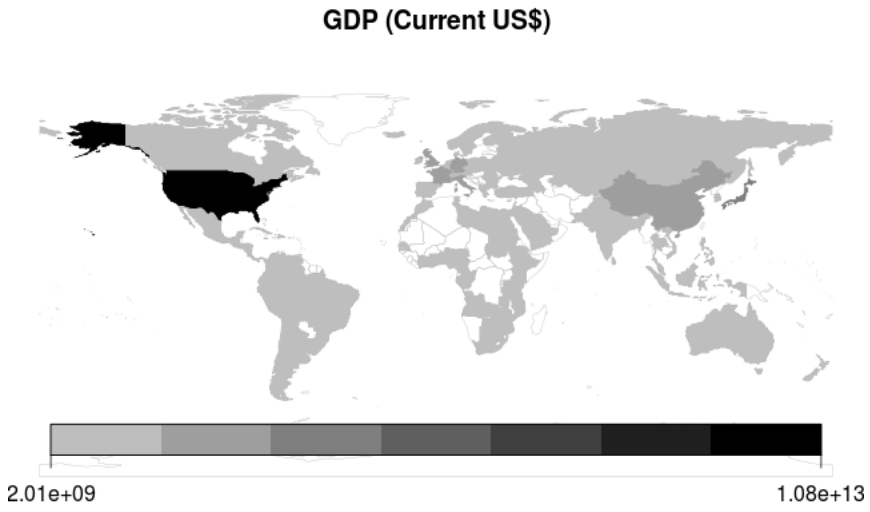

Figure 2: Average GDP 1990-2012, World Heat Map

As shown in Figure 2, when looking at GDP, we see that the United States is on the high end of the GDP range, followed by Japan, China, and several European countries.

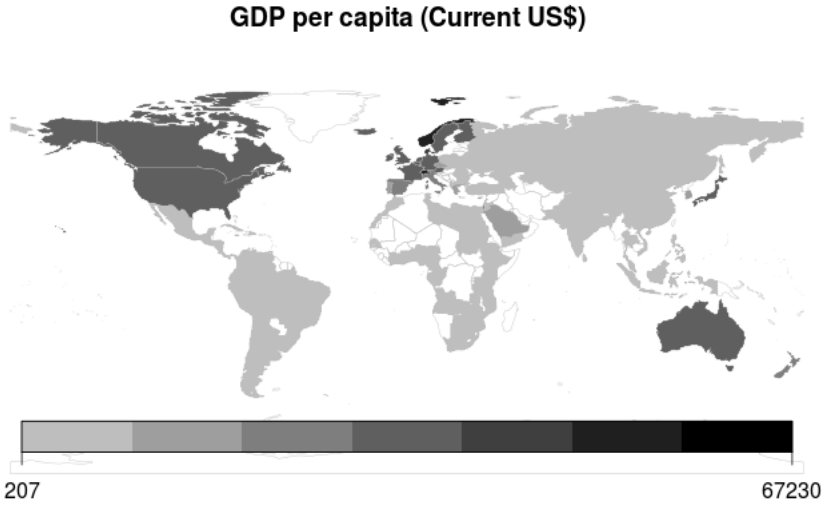

Figure 3: Average GDP per capita 1990-2012, World Heat Map

Figure 3 shows GDP when normalizing by population (per capita), the GDP disparity diminishes significantly and more countries are visible as being on the high end of the GDP range.

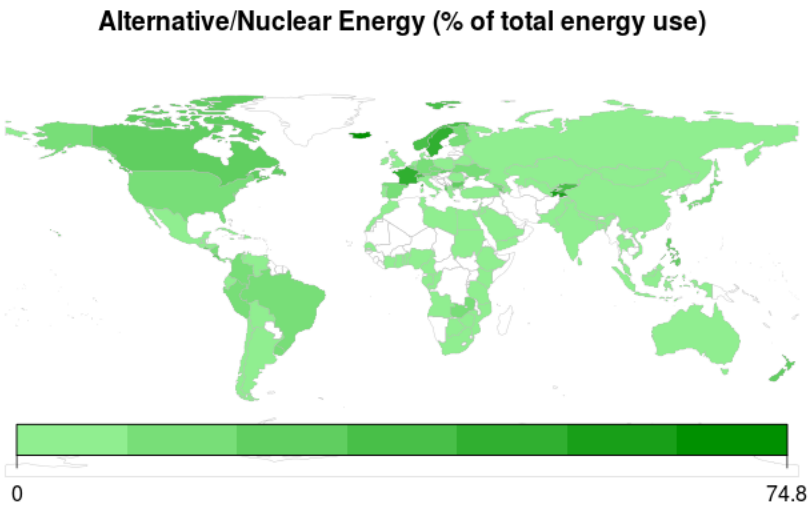

Figure 4: Alternative/Nuclear Energy (% of Total Energy Use) Average 1990-2012, World Heat Map

Looking at alternative and nuclear energy percentages (Figure 4): France, Iceland, and Sweden seem to be strong outliers, and we see some correlation with the previous GDP maps when looking at first world countries in regions like Europe and North America.

As illustrated in Figure 5, the area of combustible renewables and waste-based energy generation is largely localized in Africa.

Finally, Figure 6 shows that fossil fuel use is ubiquitous throughout much of the world, much more so than alternative/nuclear or combustible/waste energy.

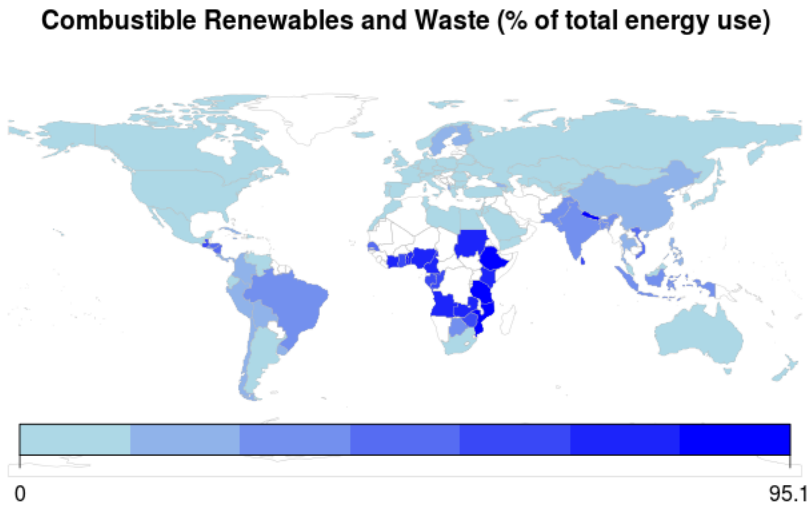

Figure 5: Combustible Renewables and Waste (% of Total Energy Use) Average 1990-2012, World Heat Map

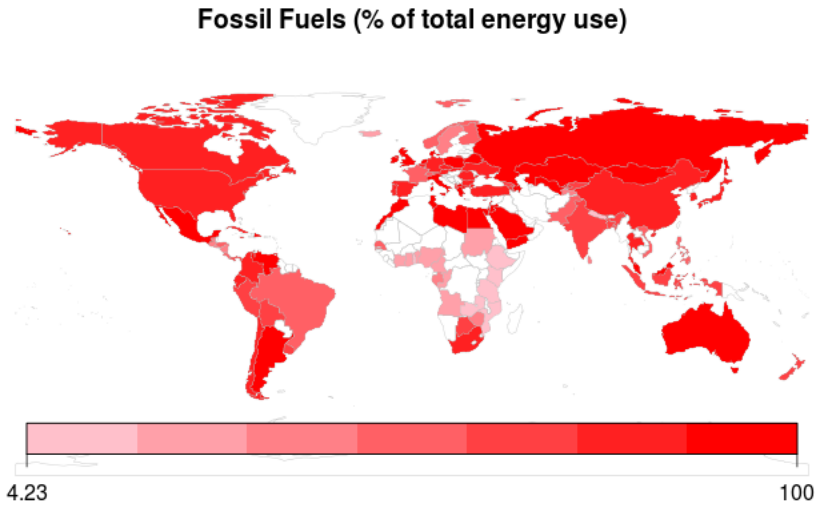

Figure 6: Fossil Fuels(% of Total Energy Use) Average 1990-2012, World Heat Map

3.3.2 Final Results

Figures 7 and 8 show the final results for GDP by Energy Type before and after the cluster analysis. In the line plot, we see that fossil fuel use tends to be larger in wealthier nations. Alternative/nuclear energy peaks towards the middle of the GDP axis, and combustible are significantly more present in lower GDP countries but their use diminishes as GDP increases

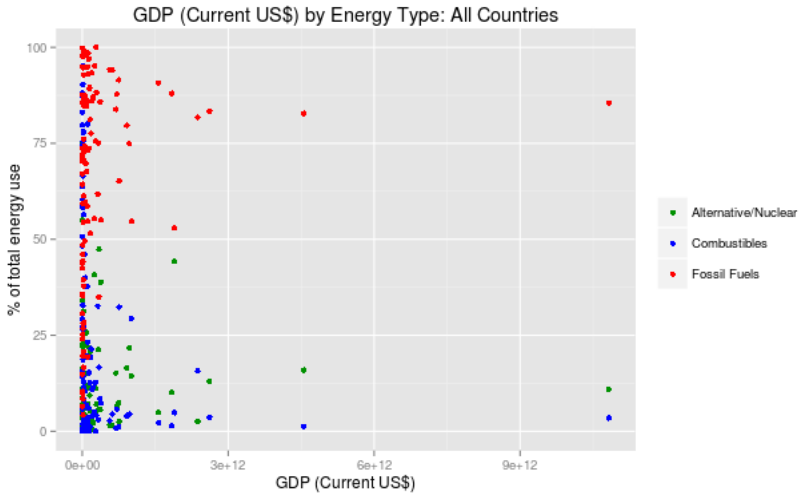

Figure 7: GDP by Energy Type Average 1990-2012, Scatter Plot

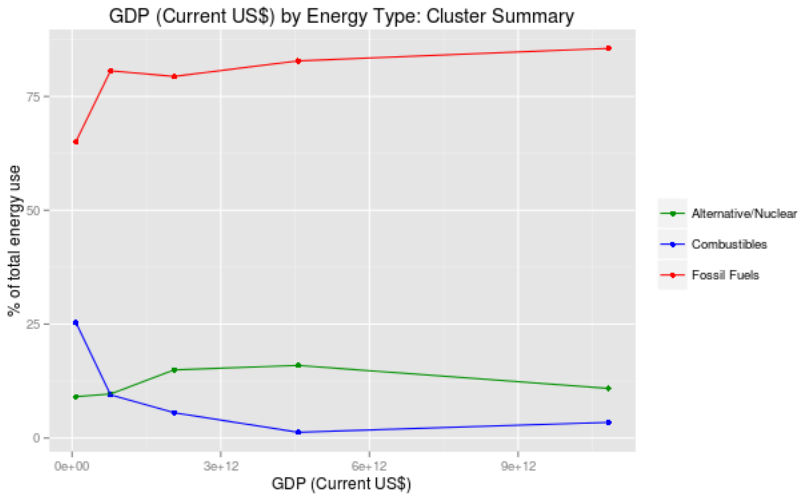

Figure 8: GDP by Energy Type Average 1990-2012, Cluster Summary

and is only slightly larger at the right-most side of the graph.

Using the scatter plot for comparison, we also see that much of the trend on the right side of the line plot is influenced by very few points. The right-most points on the scatter plot in each of the three categories corresponds to the United States, and the second-highest GDP points for each category represent Japan. These two countries are almost entirely responsible for the trends we see on the right side of the plot. As explained earlier in the Method section, a

follow-up analysis might focus on countries that share the same reason to normalize for external factors that cause outlier countries to skew the trend.

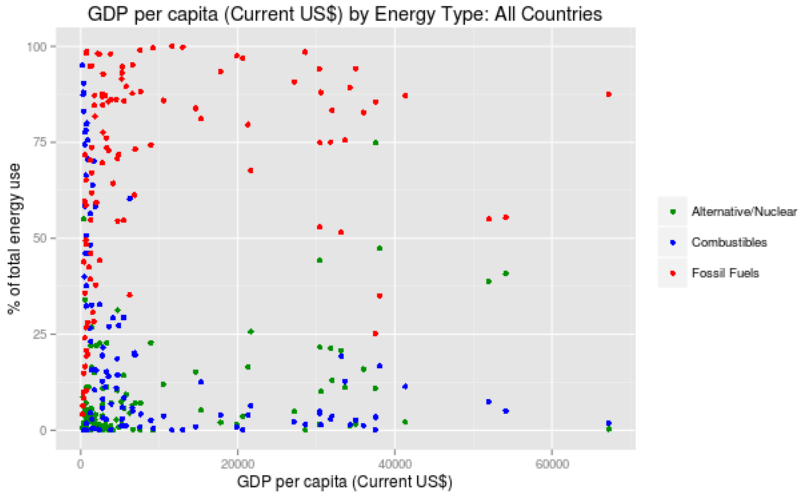

Figure 9: GDP per capita by Energy Type Average 1990-2012, Scatter Plot

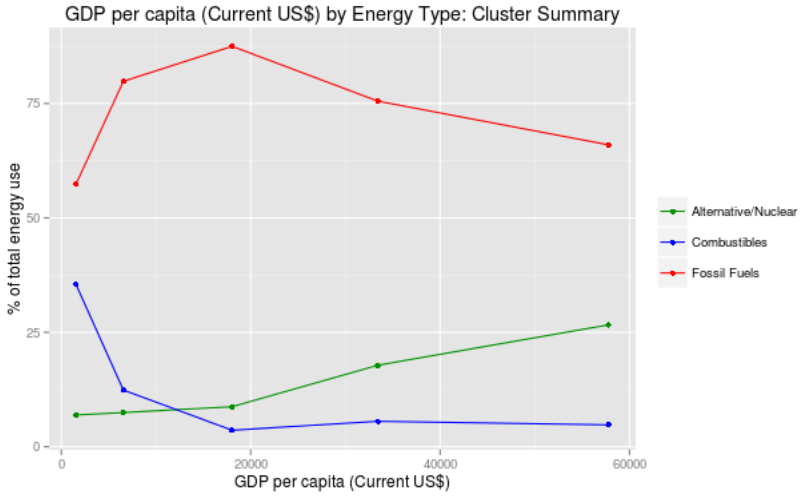

Figure 10: GDP per capita by Energy Type Average 1990-2012, Cluster Summary

The final set of plots (Figures 9 and 10), shows the scatter and line plot from before using GDP per capita. This normalizes the data by size of populations, giving us more pronounced

changes in the lines= plots for each energy type. Here, there is a clear peak in fossil fuel use for countries closer the middle of the GDP per capita range. It is also interesting to note that the alternative/nuclear category steadily increases with GDP per capita, averaging at over 25% of a nations total energy use for countries on the high end of the GDP per capita spectrum. The trend for combustibles and waste stays mostly the same relative to the original GDP plot, but the peaks and dips in the line plot are more pronounced.

4 Discussion

This project provides a starting point for understanding the breakdown of the worlds energy use. There are several improvements discussed earlier in this report that would be necessary to produce defendable conclusions from this analysis, as there are many external factors affecting the relationship explored here. With some revision to the methodology, the relationship between the size of a nation's economy and its energy use could be used as a benchmark to measure whether countries are working to the extent of their means when responding to climate change by way of energy policy decisions. The use of k-means clustering demonstrated here could also be used to visualize other trends that are not immediately apparent due to the large input of data points.

4.1 Learnings

In completing this project, I furthered my knowledge of integrating several different data analysis tools. I had not previously used SQL databases as a means to share data between Python and R, and doing so showed me the power of combining multiple tools to accomplish a single task. This project also expanded my knowledge of available R packages which can be used to transform and visualize data in effective ways. I have a much stronger working knowledge of "ggplot," and "rworldmap" which will now be more accessible to me for future projects.

4.2 Challenges

As discussed at the beginning of the 'Method' section, the ultimate result produced from this analysis shows a relationship between GDP and energy use that is also affected by a host of external factors, as evident from the large number of outliers that skew the lines shown in the cluster summary. An appropriate next step would be to filter the data into smaller subsets of countries that share other similar factors. This would help cancel out the effects of externalities and better show the relationship between GDP and energy use. This additional level of analysis was left out of this project because the project's primary requirement was demonstration of the technologies used. Another future step would be to further refine the date range selected to include data for more countries in this analysis, or to find another solution for estimating missing values.

5 Conclusion

This project ultimately produced line plots showing the relationship between GDP and the breakdown of energy use on a global scale. In doing so, the project demonstrates integrated use of several data analysis technologies including: Python, R, databases, pulling data from online data sources, cluster analysis, and data visualization. While there are a number of next steps that would need to be pursued to produce defendable conclusions from the trends shown

here, these plots and maps are effective visuals for highlighting the disparity in energy use between countries on the low and high ends of the GDP spectrum.

6 Additional Resources

A short video presentation on this analysis is viewable at the following link:

`https://www.youtube.com/watch?v=-Rwh1bIYOao&feature=youtu.be`

7 Works Cited

Dasgupta, S., Laplante, B., Meisner, C., Wheeler, D., & Yan, J. (2007, February). The Impact of Sea Level Rise on Developing Countries:A Comparative Analysis. World Bank Policy Research Working Paper.