Tracking Energy Production Through GDP Per Capita

Ethan Ericson

University of Arkansas

DASC 1003/H – Intro to Data Science

Dr. Schubert

10/11/23

Introduction

Overview and Descriptive Model

In the following study, the relationship between gross energy production and GDP per capita is explored through comparison of UN-sourced datasets. The gross production of energy is the total amount of energy a country can generate while the gross domestic product per person, otherwise known as the GDP per capita, is an indicator of a country's development. This study's exploratory data analysis examines the relationship between a country's total amount of energy production, measured in Terawatt-hours, and its developmental status, measured by GDP per capita, to explore correlations between gross energy production and GDP per capita. These correlations are then used to predict future trends between gross energy production and GDP per capita in specific countries as well as globally. By predicting the behavior of gross energy production compared to GDP per capita, a deeper understanding of GDP per capita's role in influencing gross energy production can be gained, allowing for prescriptive actions to be taken that can affect a country's gross energy production.

Methods

To explore the relationship between gross energy production and GDP per capita, two UN-datasets: "GDP per capita, PPP (current international $)" and "Total Electricity" were used. After downloading both datasets as Excel files, the data cleaning process began. Cleaning consisted of constructing a table for each of the two data sets, inner merging those tables on the "Country or Area" and "Year" columns, and then renaming the columns "Country or Area", "Quantity", and "Value" to "Country", "Terra-watt hours", and "GDP Per Capita" respectively. The "Commodity – Transaction" column was deleted, and the "Terra-watt hours" column units were converted to terra-watt hours instead of millions of kilowatt-hours. This concluded the data cleaning process, which was followed by data visualization.

Data visualization was performed in Python using the modules pandas, matplotlib.pyplot, and seaborn. The following code chunks, Chunk 1 and Chunk 2, detail the process by which visualization was processed.

Chunk 1

Importing and sorting countries by region.

Reading the Datasets

```python
merged_sheet = pd.read_excel("./I-EDA Project.xlsx", sheet_name = "GDP and Energy") # Reading in Excel Merged Sheet

del merged_sheet["Gross Energy Production.Quantity"]    # Deleting non-Terrawatt-hours column

# Creating region filters:

Africa_countries = ["Egypt", "South Africa", "Nigeria", "Ethiopia", "Ghana"]
Asia_countries = ["China", "India", "Indonesia", "Japan", "Pakistan"]
Caribbean_countries = ["Haiti", "Dominican Republic", "Trinidad and Tobago", "Puerto Rico", "Jamaica"]
Central_America_countries = ["Guatemala", "El Salvador", "Honduras", "Panama", "Costa Rica"]
Europe_countries = ["Netherlands", "Ukraine", "Germany", "United Kingdom", "Greece"]
North_America_countries = ["United States", "Canada", "Mexico"]
Oceania_countries = ["Australia", "New Zealand", "Papua New Guinea", "Solomon Islands", "Fiji"]
South_America_countries = ["Brazil", "Argentina", "Chile", "Colombia", "Peru"]

# Filtering the data set by region

Africa = merged_sheet[merged_sheet['Country'].isin(Africa_countries)]
Asia = merged_sheet[merged_sheet['Country'].isin(Asia_countries)]
Caribbean = merged_sheet[merged_sheet['Country'].isin(Caribbean_countries)]
Central_America = merged_sheet[merged_sheet['Country'].isin(Central_America_countries)]
Europe = merged_sheet[merged_sheet['Country'].isin(Europe_countries)]
North_America = merged_sheet[merged_sheet['Country'].isin(North_America_countries)]
Oceania = merged_sheet[merged_sheet['Country'].isin(Oceania_countries)]
South_America = merged_sheet[merged_sheet['Country'].isin(South_America_countries)]
```

[20] ✓ 0.3s

Note. The figure details the process of importing the countries' data from an excel sheet, and then sorting the countries and their data based on world region.

As shown above in Chunk 1, data was imported from the cleaned excel sheet using pandas. The resulting data frame from this import, "merged_sheet", was sorted by world region through the creation of custom region filters. These sorted data frames were then used in the creation of visualizations, as shown in Chunk 2:

Chunk 2

Creating plots for each world region using seaborn.

Starting Visualizations

Africa

```
# Using seaborn to create the scatter plot for each region:

lm = sns.lmplot(x='GDP Per Capita', y='Terrawatt-hours', hue='Country', data= Africa, scatter = False, ci = None)

# Customizing the plot
plt.subplots_adjust(top=0.9)  # Adjust the title position
plt.suptitle('Gross Energy Production by Country in Africa, 1990 - 2020')


# Showing the plot
plt.show()
```

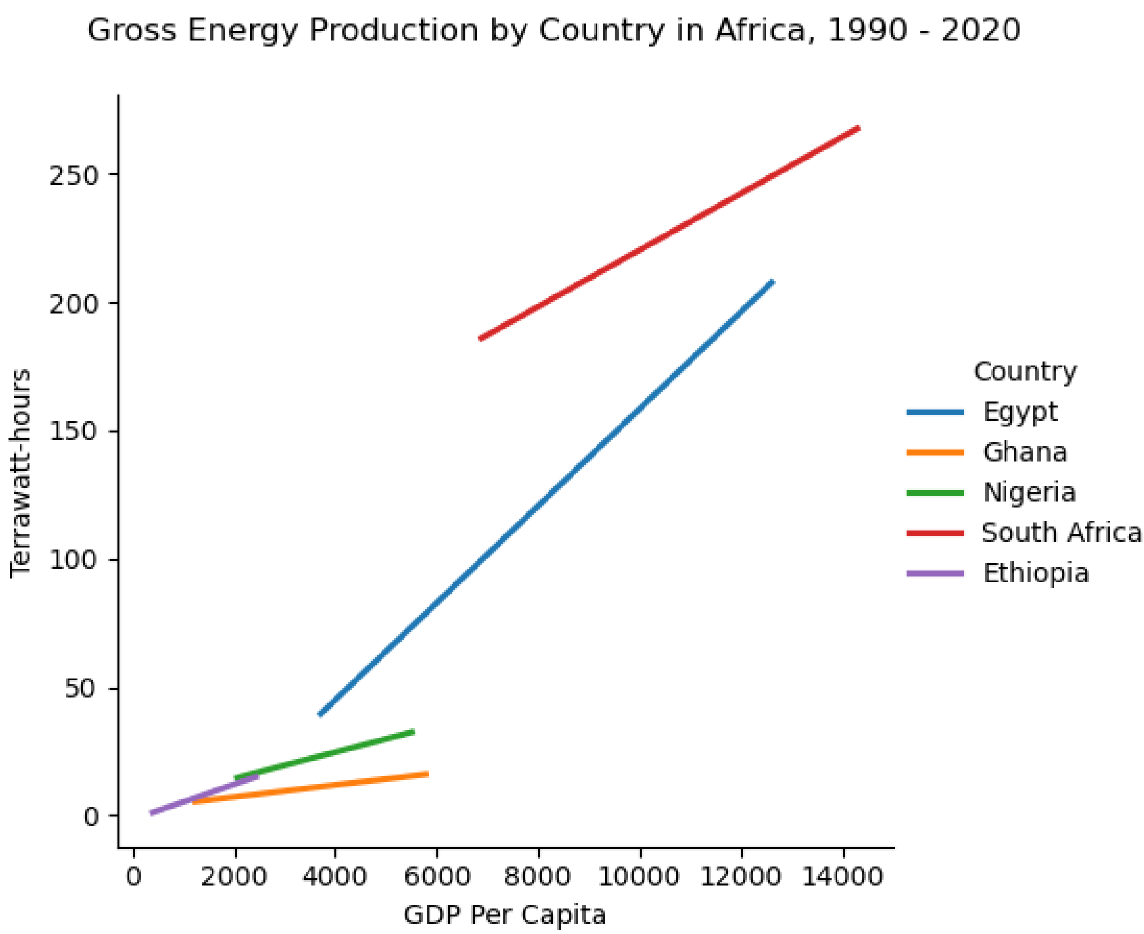
✓ 0.3s

Note. The figure details the process of crafting visualizations from each sorted data frame created in Chunk 1.

The code detailed in Chunk 2 uses seaborn to create a visualization with a regression line for each country within the specified region. The variable “lm” designates the created plot for the region. The plot is given its title and displayed as an output following the execution of the rest of the code. This chunk of code was repeated for each region until all regions had been graphed, with each country within the region plot modeled by a regression line. The resulting output of Chunk 2 is displayed as Plot 1 below:

Plot 1

A scatterplot of GDP Per Capita and Gross Energy Production modeling several countries by the regression lines fitting their data from 1990 to 2020.



Note. This visualization represents the relationship between GDP Per Capita and Terra-watt hours for five countries in Africa, modeled by the regression line of their respective data.

With visualizations created, the lengths and slopes of each regression line were analyzed, and research conducted to provide potential explanations for variation among countries and regions. This information facilitated the creation of a predictive data model focused both on individual countries and the global relationship between GDP per capita and gross energy production. The prediction of future trends led to the development of a prescriptive model to impact gross energy production based off GDP per capita.

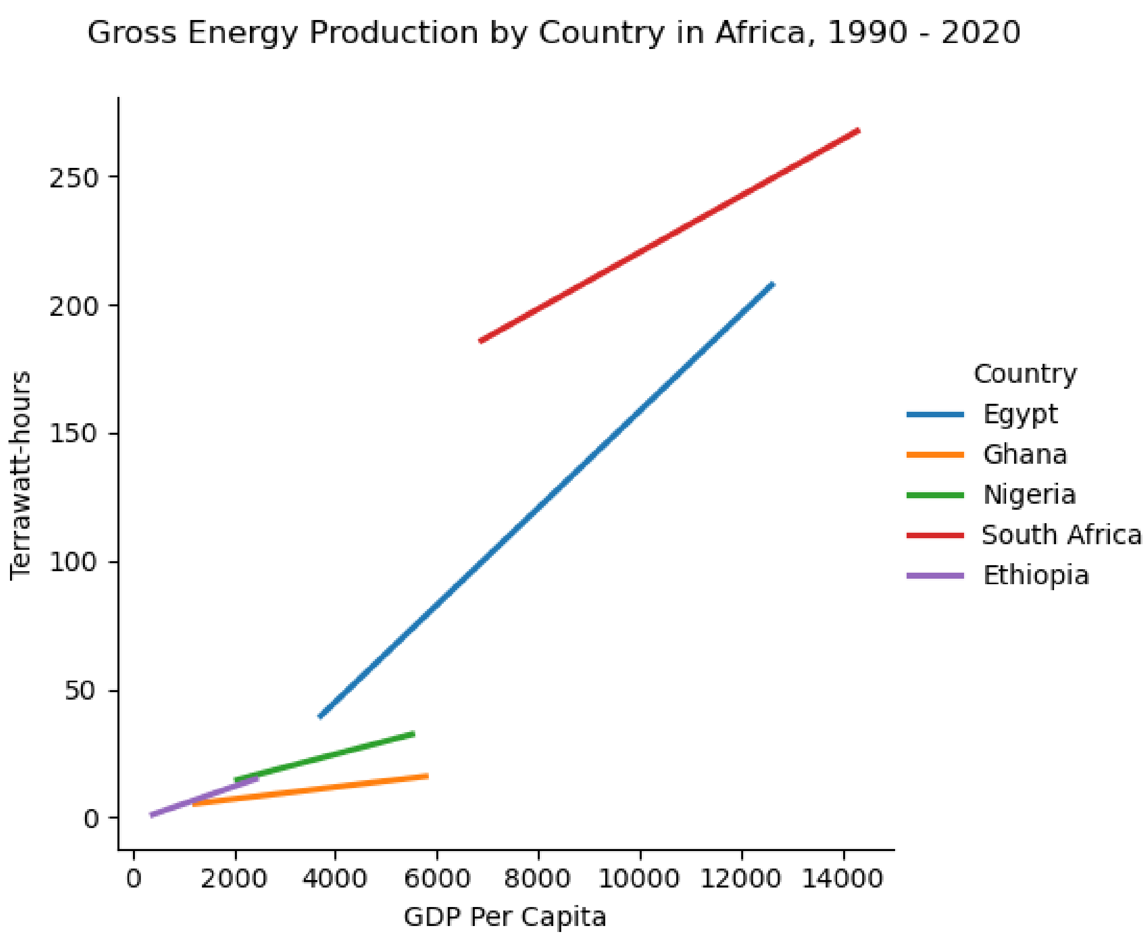
Data Analysis

Three of the visualizations created through the exploratory data analysis described in the previous section will be listed in this portion of the report. Important observations will be highlighted and explained, followed by an evaluation of the drivers behind the noted trends.

Africa

Plot 1

A scatterplot of GDP Per Capita and Gross Energy Production modeling several countries by the regression lines fitting their data from 1990 to 2020.



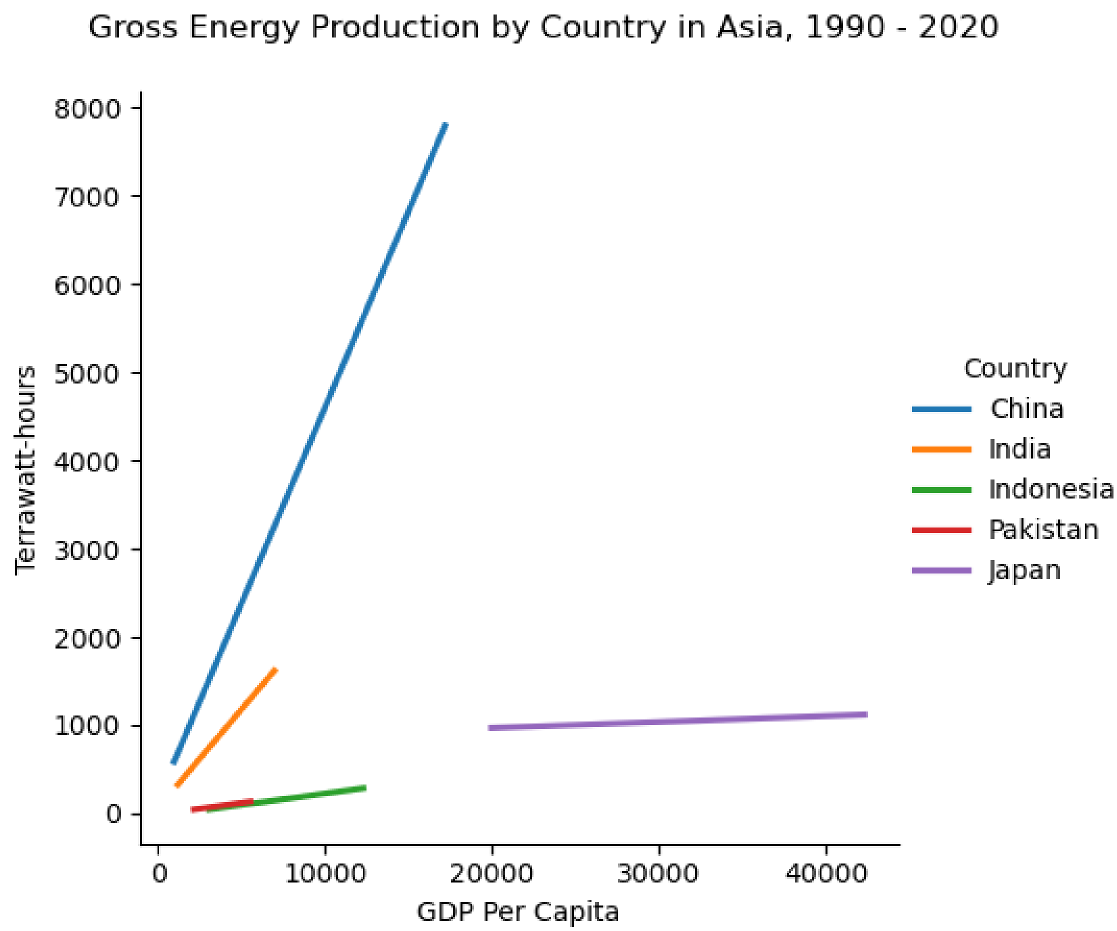
Note. This visualization represents the relationship between GDP Per Capita and Terra-watt hours for five countries in Africa, modeled by the regression line of their respective data.

In the plot of Africa's countries, several interesting trends can be observed. First, all regression lines in the plot display a positive correlation between GDP per capita and terra-watt hours, indicating that from the sample of African countries, GDP per capita is positively correlated to gross energy production. Also, Egypt has a regression line that is approximately 1-1 between GDP per capita and gross energy production. This means that, in Egypt, the growth of gross energy production closely mirrors the growth in GDP per capita. In Ethiopia, however, gross energy production is not as closely related to GDP per capita, as the slope of the regression line is shallower than Egypt's. A shallower regression line indicates that the country does not experience a high growth in gross energy production for a significant growth in GDP per capita. Ethiopia's regression line is also much shorter than Egypt's, illustrating that Ethiopia has not experienced as much growth in either GDP per capita or gross energy production as Egypt has. These observations indicate that GDP per capita's level of correlation with gross energy production varies per country in Africa, and that African countries with a lower GDP per capita produce less energy.

Asia

Plot 2

A scatterplot of GDP Per Capita and Gross Energy Production modeling several countries by the regression lines fitting their data from 1990 to 2020.



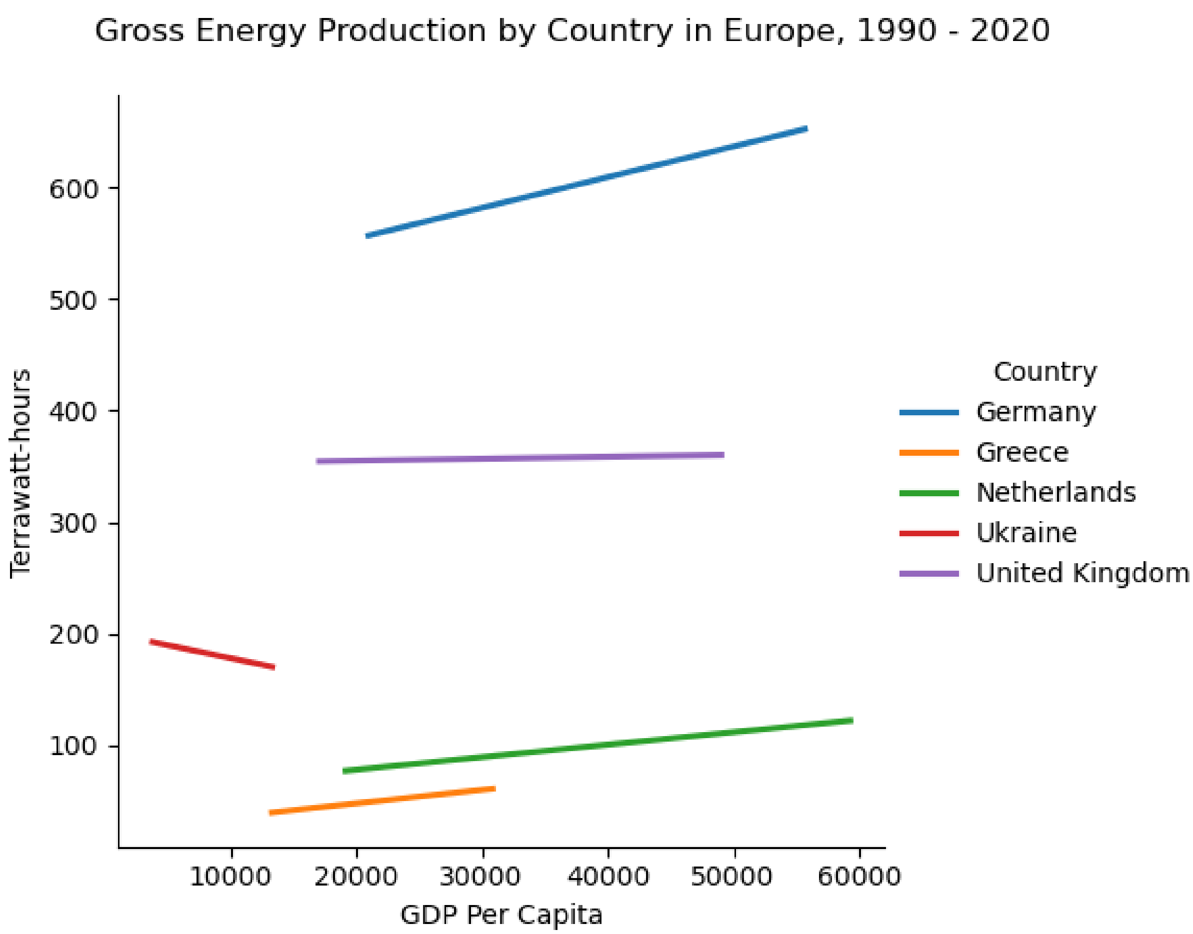
Note. This visualization represents the relationship between GDP Per Capita and Terra-watt hours for five countries in Asia, modeled by the regression line of their respective data.

In this plot of GDP per capita and terra-watt hours among the countries of Asia, similar trends to those in Plot 1 are observed. There are varying lengths of lines, and countries with a low GDP per capita are generally producing less energy than countries with higher GDP's. However, there is an exception to this trend: Japan. Japan has a very high GDP, put is not producing more energy than India, who has a lower GDP per capita. This exception indicates that there is not a purely direct connection between GDP per capita and gross energy production. Other factors, such as the smaller size of Japan compared to India, may play a role in the amount of possible energy that Japan can produce. Notice that in 1990, at the beginning of the regression lines for India and Japan, Japan produces more energy than India does, but as time increases along the regression line, India is able to expand its energy production potential faster than Japan, potentially due to Japan's lack of space for the capital to produce energy.

Europe

Plot 3

A scatterplot of GDP Per Capita and Gross Energy Production modeling several countries by the regression lines fitting their data from 1990 to 2020.



Note. This visualization represents the relationship between GDP Per Capita and Terra-watt hours for five countries in Europe, modeled by the regression line of their respective data.

Plot 3 depicts immediate differences from the first two plots. First, not all the regression lines demonstrate a positive correlation between GDP per Capita and gross energy production, as seen with Ukraine. Second, there are three countries: The Netherlands, United Kingdom, and Germany that all have similar GDP's per capita in 1990 to 2020 but have very different levels of energy production. The Ukraine anomaly can be explained by the Russo-Ukrainian gas dispute of January 2009, in which Russian shut off gas to Ukraine after failing to agree on a price for Russian gas. The gas shut off lasted 13 days and severely impacted the energy production of Ukraine and other Eastern Europe countries (Stern, Pirani, Yafimava, 2009). Because of this significant drop in energy production, the regression line of Ukraine is pulled into a negative slope. Additionally, Ukraine has seen declining energy production as the result of the Russian annexation of the Crimean Peninsula in 2014, which resulted in the loss of 80% of Ukraine's oil and natural gas deposits in the Black Sea (U.S. Energy Information Administration, 2021). The combination of these two external factors results in a negative correlation between GDP per capita and gross energy production in Ukraine, demonstrating that GDP per capita is not the only factor in a country's energy production. The second anomaly, the presence of three difference levels of energy production despite similar GDP's per capita in The Netherlands, United Kingdom, and Germany, can be explained by the differing economic priorities of those countries. GDP per capita represents the success of a country's economy, and therefore the flexibility of its resources for various investments. The difference in gross energy production in these three countries is due to those countries' decisions to invest varying levels of economic resources into gross energy production before the start of 1990. Assuming all countries start at the origin of zero GDP per capita and zero terra-watt hours of energy production, these three countries must have invested different amounts of economic resources to gross energy production to arrive at their respective points in 1990. This observation demonstrates that GDP per capita's influence on gross energy production is determined by the willingness of a country to devote economic resources to gross energy production.

Conclusion

Predictions and Prescriptions

Based on the data gathered and visualized in the last section, we can draw three primary conclusions:

1. For countries investing a significant portion of their economic resources, represented by GDP per capita, into energy production, gross energy production will be positively correlated and proportional with GDP per capita.
2. Countries can choose not to invest a significant portion of their economic resources into energy production, weakening the relationship between GDP per capita and gross energy production.
3. Factors other than GDP per capita have effects on gross energy production within a country, such as the space available for energy producing facilities and the loss or gain of territory containing materials used in the production of energy.

These conclusions can help us predict the future relationship between GDP per capita and gross energy production within a country, as well as globally. China, for example, invests a significant amount of their economic resources into energy production, represented by strong positive correlation between China's GDP per capita and gross energy production in Plot 2. Should this relationship continue, we can expect that increasing the GDP per capita within China will greatly increase its energy production, and vice versa. This knowledge is beneficial should China's gross energy production ever need to be manipulated to reach certain values. Additionally, on a global setting, we can predict that certain regions, such as Africa and Asia, will continue to see an increase in gross energy production as long as the GDP per capita of its constituent countries increases and countries continue investing similar portions of their economic resources into energy production. Other regions, such as Europe, cannot have their gross energy production predicted as reliably, due to the variable nature of how countries within Europe invest their economic resources. Therefore, it is more difficult to prescribe a single factor to influence gross energy production in all countries of the world than it is to prescribe a single factor to just one country. Overall, gross energy production, while generally sharing a positive correlation with GDP per capita, can be influenced by several factors that make it difficult to manipulate by just changing one variable.

Reflections

This report was very enjoyable to work with and very educational in many visualization techniques. If I were to redo this assignment, I would pick datasets that addressed a clear problem in society and narrow the range of data I was observing. It was difficult to maintain a sense of purpose with this assignment as I was purely looking for a potential relationship instead of searching for a solution to a problem, and the amount of data I was dealing with initially was overwhelming. That being said, I am glad that I had the experience of working with bigger datasets than desired, as that will likely be more similar to my experiences in the real world. This project was a great experience for me and helped introduce me to the world of a data scientist.

Bibliography

Ember. (2022, March 29). *Egypt | Electricity Transition*. Ember. https://ember-climate.org/countries-and-regions/countries/egypt/

Our World in Data. (2020). *GDP per capita*. Our World in Data. https://ourworldindata.org/grapher/gdp-per-capita-worldbank

Stern, J., Pirani, S., & Yafimava, K. (2009, February). *The Russo-Ukrainian gas dispute of January 2009: a comprehensive assessment*. Oxford Institute for Energy Studies. https://www.oxfordenergy.org/publications/the-russo-ukrainian-gas-dispute-of-january-2009-a-comprehensive-assessment/

U.S. Energy Information Administration. (2021, August). *International - U.S. Energy Information Administration (EIA)*. Www.eia.gov. https://www.eia.gov/international/analysis/country/UKR

United Nations. (n.d.-a). *UNdata | record view*. Data.un.org. Retrieved October 11, 2023, from https://data.un.org/Data.aspx?d=EDATA&f=cmID%3aEL

United Nations. (n.d.-b). *UNdata | record view | Series not found!* Data.un.org. https://data.un.org/Data.aspx?d=WDI&f=Indicator_Code%3ANY.GDP.PCAP.PP.CD