

## 05839 Homework 2 Write Up

### Description of the goals of this project:

As climate change becomes a more pressing issue, the need to cut emissions consequently becomes extremely urgent. Energy production and consumption accounts for an astounding 75% of the global annual CO<sub>2</sub> emissions. Therefore, the best way to cut down global emissions significantly and quickly is by transitioning our energy system from one powered by fossil fuels such as coal, gas, and oil to one powered by renewable sources such as solar, wind, and geothermal. This project aims to understand the current energy production system around the globe and investigate the trend in the transition to green and renewable energy.

There are three questions that I want to enable users to answer:

- 1. Where do countries get their energy from?**
- 2. Are countries starting the transition to renewable energy production? If so, when did the transition start?**
- 3. Does public opinion on climate change have an impact on the transition to renewable energy production?**

To help users answer these questions, I used a dataset from Kaggle: <https://www.kaggle.com/mathurinache/global-power-plant-database>. This dataset contains a comprehensive list of power plants around the globe, including their fuels, location, capacity, etc. Some of the power plants also include their commissioning date and other useful information. I also extracted poll results for public opinion on climate change from <https://www.aei.org/research-products/report/aei-public-opinion-study-polls-on-the-environment-energy-global-warming-and-nuclear-power-2/> and [https://www.pewresearch.org/fact-tank/2019/04/18/a-look-at-how-people-around-the-world-view-climate-change/ft\\_19-04-18\\_climatechangeglobal\\_since2013concerns/](https://www.pewresearch.org/fact-tank/2019/04/18/a-look-at-how-people-around-the-world-view-climate-change/ft_19-04-18_climatechangeglobal_since2013concerns/).

### Design Rationale:

#### Part 1:

The first question involves displaying a breakdown of each country's energy sources, and it would be nice to have some kind of visual comparison between different countries and continents. To answer this question, I need to display the composition of each country's power plants (e.g. percentage of coal plants, percentage of gas plants, etc.). To represent the result effectively, I used a Treemap Chart. This allows clear visualization of hierarchical data, and more importantly, it allows the display of many countries' data simultaneously within a relatively small space. Using Treemap Charts, not only the user can see the composition of a country's energy sources visually, but the user can also compare country by country, and even continent by continent at the same time. The user could click into a rectangle to see a more detailed breakdown. I considered using a lot of pie charts side by side, but given that there are almost 200

countries, this will take up an unacceptable amount of space. However, just in case a user wants to see the composition for an individual country, I added a user select box where a user can select any country and that country's energy composition will be displayed as a pie chart.

## **Part 2:**

The second question involves helping users identify the trend in a country's power plant commissioning history and the trend in the globe's power plant commissioning history. For the first part, I used a stacked bar chart to show a country's commissioning history. This allows clear visualization of the trend over time by examining the length of the bars where each small bar in a stacked bar represents a type of power plant. An entire stacked bar represents all of the power plants commissioned for that year. I considered using a filled area chart because I thought it would give a better and more fluid visual. However, the visual was not as clear as I expected, and it required a lot of extra processing to reach the desired results. The stacked bar chart is a nice balance between visual clarity and the amount of processing required. I added a multi-select box so that the user could potentially select multiple countries to compare their commissioning history side-by-side and discover which country is transitioning faster and which is lagging.

To examine the global trend, I added a separate line graph that shows the moving average of each type of power plant commissioned over the years around the globe. I used a line graph because it is easy to see the trend. For interaction, I used a slider where the user can select the time range. This allows the user to identify trends within a specific timeframe and compare it with other periods to see how global energy production evolves.

## **Part 3:**

The third question involves examining the relationship between a country's power plant commissioning history and its public opinion on climate change. For this question, I specifically focused on US data because they are the most complete and accessible. I once again used a stacked bar chart to show the commissioning history and added line graphs on top of it using a secondary y-axis. The line graphs show public opinion trends over time. They share the same x-axis. This arrangement allows the user to easily discover any correlation between trends in the bar chart and trends in the line graph. I added a multi-select box so that the user can select one or more public opinion trends to add on top of the stacked bar chart for better visual clarity if desired.

## **Development Process:**

I did this assignment as a solo project.

First I needed to research what is an interesting dataset to explore, what questions I should ask, and what visualization I need to create. I browsed Kaggle and other websites such as data.gov to find ideas. This took me around 2.5 - 3 hours.

After finalizing the topic, I performed data cleaning on the dataset. This includes removing items that are irrelevant, such as storage facilities, and filling in some N/A columns. One interesting

thing I did was estimating the commissioning year for some of the power plants. The majority of wind farms, solar farms, and nuclear power plants were missing commissioning years in the dataset. Based on my research, most of the wind farms in the world are commissioned between 2000 and now. Therefore, when there is a wind farm that is missing a commissioned date, I filled it with a random value between 2000 and 2018 (since this dataset only covers power plants up to 2018). Similarly, solar farms started to take off starting from 2010, so when there is a solar farm that is missing a commissioned date, I will fill it with a random value between 2010 and 2018. Most of the nuclear reactors are built from 1980 - 2015. A random value will be chosen between that range. I also fixed the random seed so that the result is repeatable. Thinking about how to clean data and researching how to do it took me around 3 hours.

Then the majority of the time was spent on actually implementing the design. I had to research the Streamlit APIs and find ways to manipulate the data frames such that I can get the desired results. Some of the Streamlit APIs were poorly documented on the official website so I had to search places like Stackoverflow for clearer documentation. I used a few types of graphs and it took me a while to put everything together. These parts took me around 9 to 10 hours.

Finally, around 3 hours were used to clean up everything (such as fixing axis labels, add in more explanatory texts, adjusting color schemes, etc.) and produce the write-up.