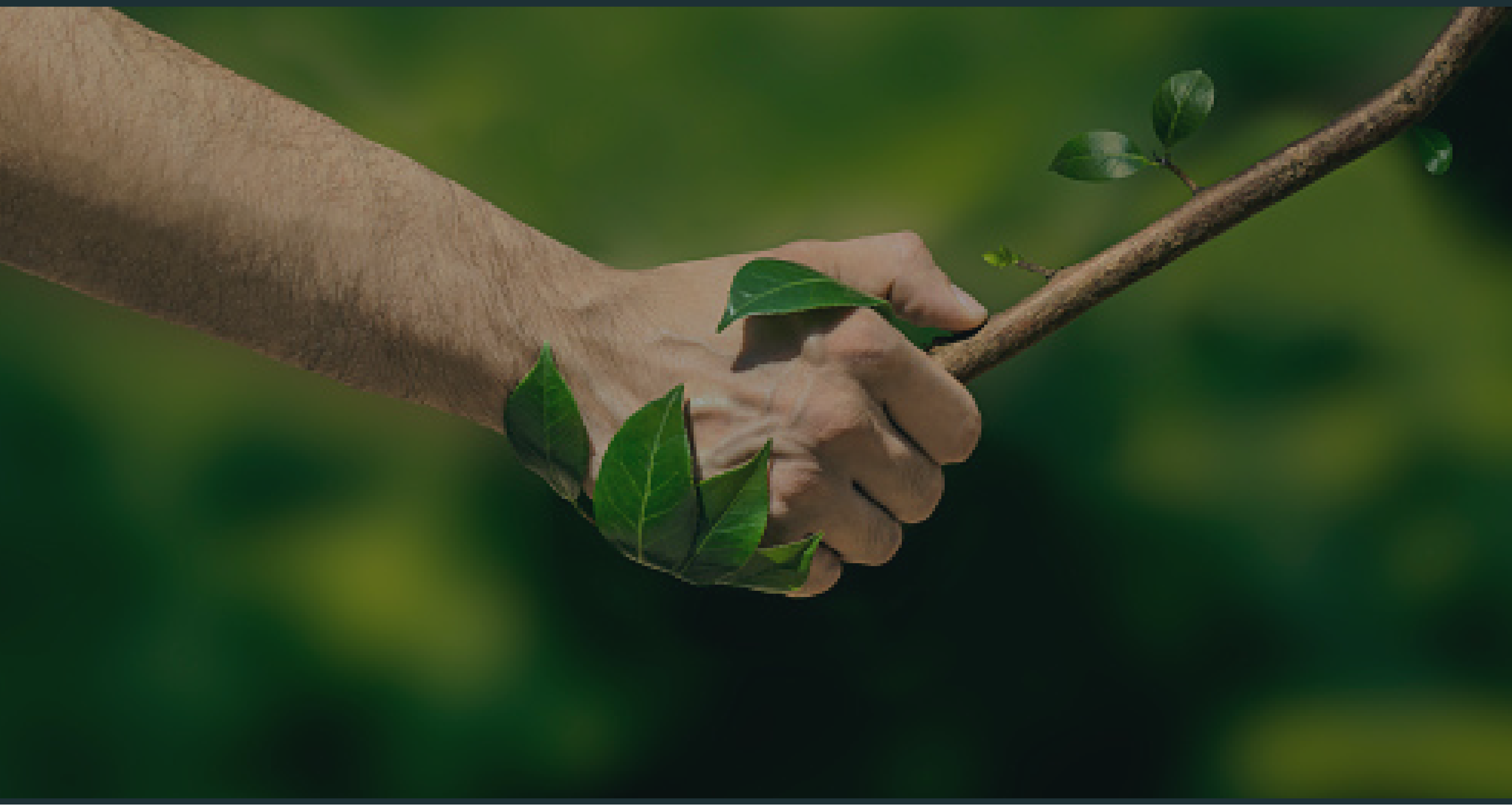
Msc. Data Science and Advanced Analytics
Martim Santos, 20220540
Miguel Cruz, 20221391
Rúben Serpa, 20221284
Sara Galguinho, 20220682

2022 - 2023

EXPLORING TRENDS IN ELECTRICITY CONSUMPTION OVER TIME AND SPACE



INTRODUCTION

The energy industry plays a big role when discussing environmental concerns. The energy sector accounts for about two-thirds of global greenhouse gas emissions attributed to human activity (Why Does Energy Matter?, n.d.). Nowadays, it is vital to have social and environmental awareness and understanding this industry is the first step towards a greener tomorrow.

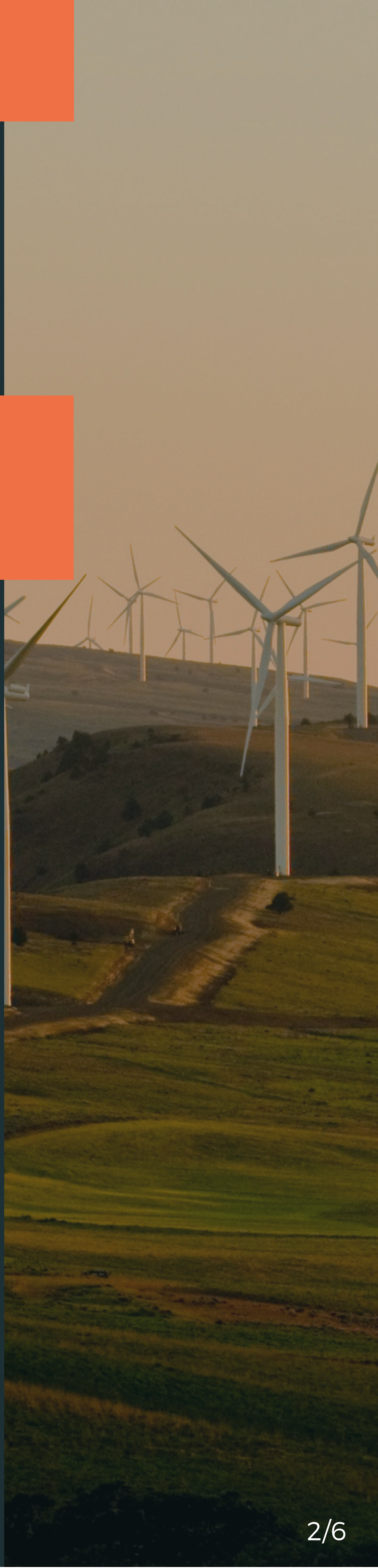
INSPIRATION: THE WILLOW PROJECT

The inspiration for this work comes after the news that on March 13, the Bidden administration approved the implementation of the Willow Project that consists of creating three new oil drill pads in Alaska. According to Nilsen (2023), the project would generate enough oil to release 9.2 million metric tons of planet-warming carbon pollution a year – equivalent to adding 2 million gas-powered cars to the roads. While countries around the world are hand-in-hand trying to find greener solutions around energy production, there are still attitudes that go against the current.

To further understand electricity consumption trends and greener solution preferences around the world, it was decided to create a reader-driven visualization that shows general patterns in a first instance and then lets the user play around with more detailed information. The visualization is comprised of various graphics and diagrams that utilize different energy sources by geographical area and are under a thirty-five-year timeline (1985-2020).

The code and visualization can be accessed on GitHub and Tableau Public through the following links:

https://public.tableau.com/app/profile/miguel.cruz7210/viz/ElectricityConsumption_16809870452720/Overview



DATA & METHODS

In order to construct a proper visualization for the project, we resorted to Python and Tableau. We started by downloading an Energy dataset from Github where the link can be found in the reference page. This dataset is a collection of key metrics maintained by Our World in Data. It is updated regularly and includes data on energy consumption, energy mix, electricity mix and other relevant metrics (owid, 2023). It was built upon several datasets and processing steps, listed and described in the Github page.

Its first version was made available on September 9, 2020, and since it is continuously updated with energy data and most metrics are published on an annual basis, it is important to mention that we have downloaded the data on March 14, 2023, so our visualizations will not consider any data published after this date.

The original dataset had 21890 rows, 129 columns/features and covered information from the year 1900 until 2022. For the project, we decided to use data regarding the period of 1985-2020. Our decision was based on the fact that until 1985 a considerable number of countries did not have data and the period of 2021–2022 also had flaws, probably due to external factors. Regarding the high number of features, we also decided to maintain only those that we thought would be helpful to answer our questions (considering we focused on electricity consumption) and thus produce good visualizations. Having that, we ended up with 10400 rows and 22 features.

When analyzing our data, we noticed some problems. The column 'country' had some other names that were not countries, like regions and continents. To address this issue, we created a function that deleted those names. We also observed that there was a large number of missing values in all the features, apart from 'year' and 'country'. To solve this, we assumed that those missing values existed because there was no information available and we filled those cells with 0.

Finally, to get a better understanding of our data and build more appealing and informative visualizations, we created new features, changed the format of our dataset and separated our information into two new datasets: 'ds_total', 'ds_pc'.

A brief description of the variables is provided in the tables below:

General Features	Description
Country	Country location
Year	Year of observation
GDP	Total real gross domestic product, inflation-adjusted
Continent	Continent location

Table 1 - General features description

Features	Dataset	Description
Electricity	ds_total	Source of the electricity generated (Biofuel, Coal, Gas, Hydrogen, Nuclear, Oil, Other Renewables, Solar, Wind)
	ds_pc	Source of the electricity generated per capita
Consumption	ds_total	Primary energy consumption, measured in terawatt-hours
	ds_pc	Per capita primary energy consumption, measured in kilowatt-hours
Energy Category	ds_total ds_pc	Type of energy: Renewable or Non Renewable

Table 2 - Specific features description of each dataset

VISUALIZATION & INTERACTION CHOICES

All visualizations described below were filtered by year, where the user chooses a desired interval, and some by type of energy using a color hue identity channel (both the scatter plot and treemap do not have the type of energy filter). The encoding was the following: Yellow for total electricity consumption; Blue for renewable energy; Red for non-renewable energy. Yellow was chosen due to its usual association with energy, being a perfect color to symbolize all energy sources. Blue is easily associated with cleanliness and the environment being a smart choice to symbolize renewable energy. Red is usually associated with negative things and danger being a great choice to represent non-renewable energy and negative impacts associated with these forms of energy, such as pollution, climate change, and resource depletion. The color palette was chosen based on Color Psychology (ColorPsychology.org, 2023).

1° Dashboard - Overview

The goal of this dashboard is to provide a brief overview of the electricity consumption trends around the world and guide the user to what it is going to see in the following dashboards.

2° Dashboard - Global

On the second dashboard, a line chart was created in order to visualize electricity consumption over the time period chosen. This type of chart was chosen due to its effectiveness for showing changes over time, highlighting trends and patterns in the data. In terms of encoding, it uses a point mark with the horizontal spatial position channel that represents a temporal scale, where each point is separated by year. It also uses a vertical spatial position channel that shows global electricity generation values (quantitative variable), measured in terawatt-hours. The values are then connected creating a line mark.

The second plot is a horizontal bar chart that provides a quick comparison of the top 10 countries with the highest electricity consumption. It uses both spatial position channels and length channels to encode the data. The x-axis represents the quantitative variable of electricity consumption, measured in terawatt-hours. The y-axis represents a categorical variable of country names. The user can also organize the countries listed in the bar chart by ascending or descending alphabetical order. Both charts information change depending on the filter selected. It is also important to note that both graphs were chosen in accordance with the effectiveness of channels (Munzner, 2014, p. 94).

Four key facts were described using general indicators. The first two describe, on a global scale, the year in which the biggest percentage of renewable/non-renewable energy was registered, respectively (2020 and 2007). The last two correspond to the overall most popular type of renewable/non-renewable energy sources for the global time period analyzed (Hydrogen and Coal). These major indicators, shown through text, were chosen to communicate some data insights in a quick and effective way.

3° and 5° Dashboard – General Map and Map per Capita

A choropleth map shows a quantitative attribute encoded as color over regions delimited as area marks, where the shape of each region is determined by using given geometry (Munzner, 2014, p. 181). In this third dashboard, through a choropleth map, we aimed to give the viewer a sense of geographic information about the countries and their total values of electricity consumption.

The fifth, and last dashboard page in the project consists of another choropleth map that shows the electricity consumption per capita in each country. The map was designed to be interactive, allowing the user to explore different variables and patterns in a more specific approach. Moreover, when hoovering over the map it is also possible to read specific details about the country's consumption or in the second case the country's consumption per capita.

4° Dashboard – Detail

This dashboard is highly effective in allowing the user to explore with much more detail different aspects of electricity consumption trends and preferences. The bubble chart was designed to compare electricity consumption, gross domestic product (GDP) and population across countries, regions and time. Each point mark represents a country, with horizontal and vertical spatial position encoding the primary quantitative attributes of GDP and electricity consumption. The color channel is used to categorize the continent that the mark belongs to and also allows for the ability to select specific continents for viewing (using popout), and the size channel for quantitative population attribute. With the help of the two labels (population and continent) we hope to give the user the possibility to identify relationships between different variables and patterns that may not be immediately apparent.

The second plot, a treemap, encodes different types of energy in a rectangular layout using area marks for each type of energy. Above that, these area marks are contained depending on if the type of energy is from a renewable source or non-renewable. This allows the user to effectively visualize complex datasets and quickly identify the relative proportions of different energy sources being used in a particular country. For a more detailed, country-based analysis, we also included a dropdown menu where the user can choose which country to see.

CONCLUSION

For the final insights, this project provided a comprehensive and engaging overview of electricity consumption trends and preferences around the world. The creation of an interactive multipage dashboard enabled us to display important information that is needed to understand and advocate towards sustainable transformation in the energy industry.

Some limitations included initial lack of advanced knowledge on the software used, high percentage of missing values and extended multivariate data. For future improvements, it would be wise to connect external data that is reliable and would allow for some imputation of missing values as well as diversifying the components that would play on our dashboards and utilizing other possible variables that could also be of high importance for the matter of analysis.

REFERENCES

ColorPsychology.org. (2023, February 1). Color Psychology 101: A Beginner's Guide to the Meaning of Colors. Color Psychology. https://www.colorpsychology.org/

Environment, U. (2017). Why does energy matter? UNEP - UN Environment Programme. https://www.unep.org/explore-topics/energy/why-does-energy-matter

Nilsen, E. (2023, March 14). The Willow Project has been approved. Here's what to know about the controversial oil-drilling venture. CNN; CNN. https://edition.cnn.com/2023/03/14/politics/willow-project-oil-alaska-explained-climate/index.html

Munzner, T. (2014). Visualization Analysis and Design. CRC Press.

owid. (2023, March 6). GitHub - owid/energy-data: Data on energy by Our World in Data. GitHub. https://github.com/owid/energy-data