Data Preparation and Visualization

Visualization of Renewable and Total Energy Consumption

Master in Data Science and Engineering
1º Semester - 2024/2025

Manuel Silva – up202108874

João Viterbo Vieira – up202107689

João Soares – up202105364

Index

1. Introduction

Energy consumption in Europe has seen significant shifts in recent years, driven by a growing focus on sustainable energy sources. Europe consumes roughly 13,000 TWh of energy annually, with renewable sources — such as wind, solar, and hydropower — accounting for 22% to 24% of this total. Countries like Sweden lead the way, deriving over 50% of their energy from renewables, while others remain more reliant on fossil fuels, reflecting the region's diverse energy landscape. This transition is central to the European Union's goals of reducing greenhouse gas emissions, with a target of 42.5% of energy consumption coming from renewable sources.

Through these visualizations, we aim to explore how different EU countries are **progressing toward this target**. We'll also compare their **total energy consumption** and **population size** to gain deeper insights into their energy profiles.

2. Research Questions

The primary research question guiding our project is the following:

How does energy consumption influence the proportion of renewable energy used?

Additionally, we explored how **population size** impacts the percentage of energy derived from renewable sources. Through these visualizations, we address several key questions:

- Which **countries and regions** have the **highest and lowest** percentages of renewable energy consumption?
- Which countries **fall short** of the European Union's renewable energy consumption **target**?
- Does **population size** affect the **total energy consumed**?
- How do **different regions** of Europe compare in terms of overall **energy consumption and renewable energy adoption**?

These insights aim to provide a clearer picture of the factors shaping renewable energy usage across Europe.

3. Methodology

To create the visualizations, we used the programming language R along with libraries such as *ggplot2* and *tidyverse*. The code was collaboratively developed and shared among team members using *GitHub*.

3.1 Data Understanding

This study relied on three datasets sourced from Eurostat, each contributing for the analysis:

- **Renewable Energy Data**: Contains the share of renewable energy in total energy consumption, broken down by country and year. The dataset was last updated on 19/09/2024.

- **Total Energy Consumption Data**: Captures energy consumption by end-users, including industry, transport, households, services, and agriculture, categorized by country and year. Data is measured in terawatt-hours (TWh) and was last updated on 24/05/2024.
- **Population Size Data**: Provides the population size for each country, defined as the number of usual residents as of January 1 of the respective year. The dataset was last updated on 08/11/2024.
- **Distribution of Renewable Energy Consumption Across Sectors**: Contains the information of renewable energy consumption in the sectors of Biofuels in Transport, Electricity and Heating/Cooling in 2022 across multiple countries.

These datasets form the foundation of our analysis, enabling a comprehensive exploration of renewable energy trends across Europe.

3.2 Data Preparation

To ensure the datasets were ready for analysis, a series of data preparation steps were implemented:

3.2.1 Data Cleaning

The initial cleaning phase addressed issues such as missing values and inconsistent column names, like we can see in the following steps:

1. Rows containing ":" were replaced with NA to standardize missing data representation.
2. The first column in all datasets was renamed to "Country" for consistency across tables.
3. Country population values were explicitly converted to numeric format to ensure accurate analysis.

3.2.2 Data Transformation

Datasets were reshaped into a long format to facilitate easier merging and analysis.

1. The pivot_longer function converted year columns into a single "Year" column for each dataset.
2. All year values were cast as integers, and variables like "PopulationDensity" were converted to appropriate data types.

3.2.3 Data Integration

Datasets were combined into a single dataset containing all relevant information.

1. Using *inner_join*, the renewable energy, total energy consumption, and population density datasets were merged based on shared "Country" and "Year" columns.
2. This integration created a comprehensive dataset enabling analysis of energy consumption, renewable energy share, and population metrics in tandem.

3.2.4 Data Reduction

The data was filtered to include only the relevant entities for the analysis.

1. Only one year was retained, reflecting the most up-to-date data.
2. Non-European Union countries, as well as Luxembourg, Lithuania, Malta, and Slovakia, were excluded from the analysis.
3. A "Region" column was added, categorizing countries into Northern, Eastern, Western, and Southern Europe, ensuring a meaningful regional comparison.

4. Data Analysis and Visualizations

This section presents the visualizations used in our analysis, explaining their design and relevance to the research questions. The chosen visualizations aim to explore the relationships between energy consumption, renewable energy, and regional disparities within the European Union (EU).

4.1. Scatter Plot

The scatter plot is our primary visualization, as it effectively represents relationships between multiple continuous and categorical variables. This visualization enabled a comprehensive exploration of correlations between renewable energy share and total energy consumption across EU countries, while also highlighting regional and demographic disparities.

In this analysis, the scatter plot has three continuous variables and one categorical variable, specifically:

- **Renewable Energy Percentage**: The share of renewable energy within a country's total energy consumption.
- **Total Energy Consumption**: The overall energy usage measured in TWh.
- **Country Population**: The total population of each country.
- **European Region**: A categorical variable indicating the region (e.g., Northern, Southern, Eastern, and Western Europe).

The scatter plot encodes the dataset's variables as follows:

- **X-axis:** Represents Total Energy Consumption on a numeric scale.
- **Y-axis:** Represents Renewable Energy Percentage on a numeric scale, illustrating the proportion of renewable energy within a country's total energy consumed
- **Color:** Differentiates European regions (e.g., Northern, Southern, Eastern, and Western Europe).
- **Size:** The size of each data point corresponded to the population of the represented country, emphasizing demographic impact.

Additional encodings and features were implemented for clarity and interpretability:

- **Labels:** Countries with extreme values (e.g., the highest and lowest renewable energy percentages) are directly labeled, ensuring these key data points are easily identifiable.
- **The plot includes contextual reference lines:**

    - **EU Renewable Energy Target (2030):** Highlighted using a dashed line, this provides a benchmark for evaluating progress toward future goals.
    - **EU Average Renewable Energy Share:** Another dashed line marks the average share, serving as a comparative benchmark for individual countries.
- **Color Palette:** A colorblind-friendly palette was chosen to ensure accessibility for a broader audience.

4.1.1 Scatter Plot GIF

To enhance the understanding of temporal trends in energy consumption and renewable energy, the scatter plot was animated. This animation visualizes the changes across the years, providing a dynamic perspective on the progression of energy metrics in EU countries. The GIF includes the following features:

- **Temporal Dynamics**: The animation transitions through yearly data, enabling viewers to observe how renewable energy percentages and total energy consumption evolved over time for each country.
- **Key Reference Lines**: The dashed line indicating the the EU Average Renewable Energy Share remain fixed, but the EU Renewable Energy Target line changes across time

This animated visualization not only makes the data more engaging but also provides valuable insights into how countries progress toward renewable energy targets over time. It highlights the pace of change in different regions and countries, encouraging a deeper understanding of the challenges and achievements within the EU's energy landscape.

4.2. Heatmap

While not the primary visualization, the heatmap offered an intuitive geographical representation of renewable energy adoption across Europe.

- **Color Intensity**: The map used color intensity to represent the ratio of renewable energy consumption, with stronger colors indicating higher percentages of renewables in total energy consumption.
- **Labels:** Select countries with notable values were labeled to display their total energy consumption and renewable energy percentage directly on the map.

This visualization provided an overview of regional patterns, emphasizing the distribution of renewable energy adoption across the continent.

4.3. Stacked Bar Graph

To analyze the sectoral distribution of renewable energy consumption, we included a stacked bar graph. The chart highlights how renewable energy is utilized across three primary sectors: **electricity**, **biofuels in transport**, and **heating and cooling**. The scatter plot encodes the dataset's variables as follows:

- **X-Axis:** Displays the percentage of total renewable energy consumption, divided by sector.
- **Y-Axis:** Lists the EU countries included in the analysis.

Additional encodings and features were implemented for clarity and interpretability:

- **Color Scheme:** Distinct colors represent each sector:
    - **Electricity:** Green
    - **Heating and Cooling:** Blue
    - **Biofuels in Transport:** Red

Additionally, **Portugal** and the **EU mean values** are emphasized using distinct shades to facilitate easy comparison with broader trends. The chart aims to break down the contribution of each sector to a country's total renewable energy consumption. This visualization allows for:

1. A detailed understanding of how countries prioritize renewable energy across sectors
2. A direct comparison of Portugal's renewable energy profile with EU averages and other countries.

5. Insights

5.1. Insights from the Scatter Plot

- **Sweden's Leadership**: Sweden stands out with the highest renewable energy share (66%) and relatively low total energy consumption, far surpassing the EU Renewable Target for 2030.
- **Germany's High Consumption, Low Renewables**: Germany has the highest total energy consumption but falls below the EU average in renewable energy share, highlighting a gap in its green energy transition.
- **Northern Europe Outperforms**: Countries like Finland, Latvia, and Denmark exceed the EU Renewable Target, showcasing Northern Europe's strong renewable energy adoption.
- **Southern Europe's Lagging Performance**: Southern European countries like Italy and Spain have moderate renewable shares but are below the EU target, despite significant energy consumption levels.
- **Eastern Europe Shows Promise**: Nations like Romania and Slovenia approach the EU average in renewable share despite lower total energy consumption, indicating a potential for growth in green energy initiatives.

5.1.1. Insights from the Scatter Plot GIF

- **Steady Growth in Renewable Energy**: The animation highlights a consistent increase in the share of renewable energy in many countries across the years.
- **Impact of COVID-19 on Energy Consumption**: A noticeable dip in total energy consumption occurs in 2020, coinciding with the global COVID-19 pandemic. This reduction reflects the economic slowdown and decreased industrial and transport activities during lockdowns.
- **Regional Trends Amid Recovery**: As the years progress beyond 2020, energy consumption begins to recover, but the proportional share of renewables remains higher than pre-pandemic levels in several countries.

- **EU-Wide Challenges and Achievements**: The EU's average renewable energy share steadily moves closer to its 2030 target throughout the animation.

5.2. Insights from the Heatmap

- **High Renewable Energy Share in Northern Europe**: Sweden leads with 66% renewable energy usage, demonstrating significant reliance on sustainable energy sources.
- **Western European Nations' Moderate Use**: Countries like Germany (20.8%) and France (20.3%) maintain moderate renewable energy shares despite their high total energy consumption.
- **Southern Europe's Mixed Performance**: While Portugal shows a substantial 34.7% renewable energy share, neighboring Spain and Italy lag behind with 22.1% and 19.1%, respectively.
- **Eastern Europe Struggles**: Poland reports a low renewable energy share (16.9%) despite a relatively high total energy consumption (71,229 TWh).
- .**Ireland's Renewable Energy Challenges**: Ireland has one of the lowest renewable shares (13.1%), reflecting significant room for improvement in renewable energy.

5.3. Insights from the Stacked Bar Graph

- **Sectoral Contributions:** The graph reveals that renewable electricity is the dominant sector in most countries, with biofuels in transport contributing smaller shares.
- **Sector-Specific Disparities:** Some countries show significant reliance on specific sectors, such as heating/cooking in Northern Europe, reflecting climatic and infrastructural factors. Mediterranean countries, such as Spain and Italy, show a distinct emphasis on renewable electricity, likely due to the abundant availability of solar energy.
- **Highlighting EU Mean:** By comparing individual country profiles to the EU mean, the graph underscores the variation in sectoral reliance across member states.
- **Variability in Energy Profiles:** EU countries show diverse energy profiles, with some excelling in balanced renewable adoption. For example, Sweden combines strong contributions across electricity, heating, and transport, while Ireland focuses almost exclusively on renewable electricity, leaving gaps in other sectors.

6. Conclusion

The visualization of renewable and total energy consumption across the European Union provides valuable insights into the current state and progression towards sustainable energy goals. Through meticulous data preparation and the application of effective visualization techniques, this study has shed light on the complex relationships between energy consumption, renewable energy adoption, population size, and regional disparities within the EU.

The data preparation process was crucial in ensuring the accuracy and reliability of the analysis. By handling missing values, standardizing column names, transforming datasets into long formats, and integrating multiple data sources, we established a solid foundation for meaningful visualization. Filtering the data to focus on relevant years and EU member states

enhanced the relevance of the findings, while variable selection minimized computational complexity without sacrificing essential information.

The visualizations developed — namely the scatter plot, heatmap, and stacked bar graph — served as powerful tools to explore and communicate the nuances of renewable energy consumption across different dimensions. The scatter plot effectively illustrated the relationship between total energy consumption and the proportion of renewable energy used, highlighting how population size and regional factors influence this dynamic.

The heatmap offered a geographical perspective, emphasizing regional trends and disparities in renewable energy adoption. It visually confirmed the leadership of Northern European countries and the lag in some Southern and Eastern European nations.

The stacked bar graph provided a sectoral analysis of renewable energy consumption, revealing the dominant role of renewable electricity and the varying contributions of heating, cooling, and biofuels in transport across different countries.

In conclusion, these visualizations addressed our research questions. The visualization of renewable and total energy consumption provides a comprehensive overview of Europe's progress towards sustainable energy goals. It highlights successes and identifies areas needing improvement. Future work could expand on this foundation by incorporating additional variables such as economic indicators to further enrich the analysis.

7. Appendix

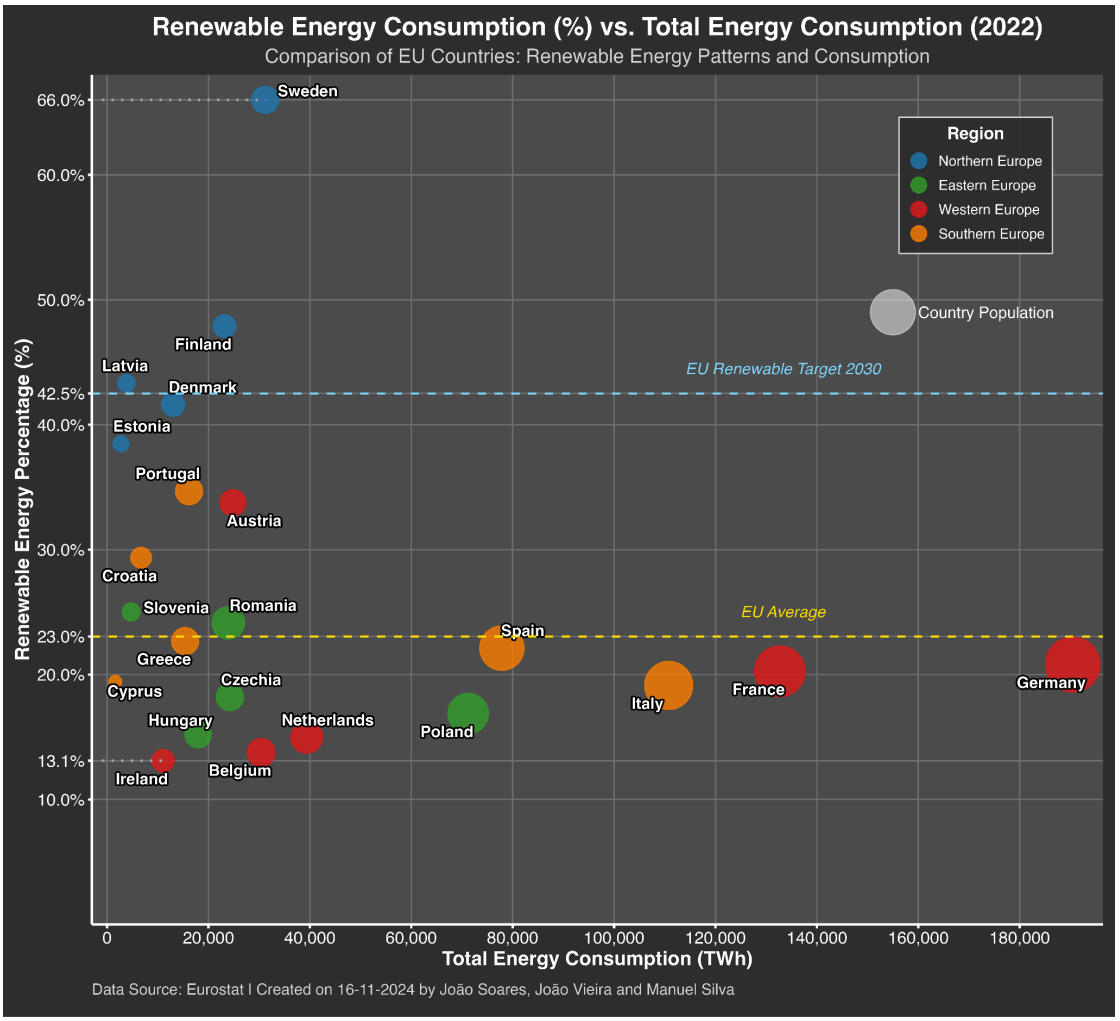

Image 1: Scatter plot

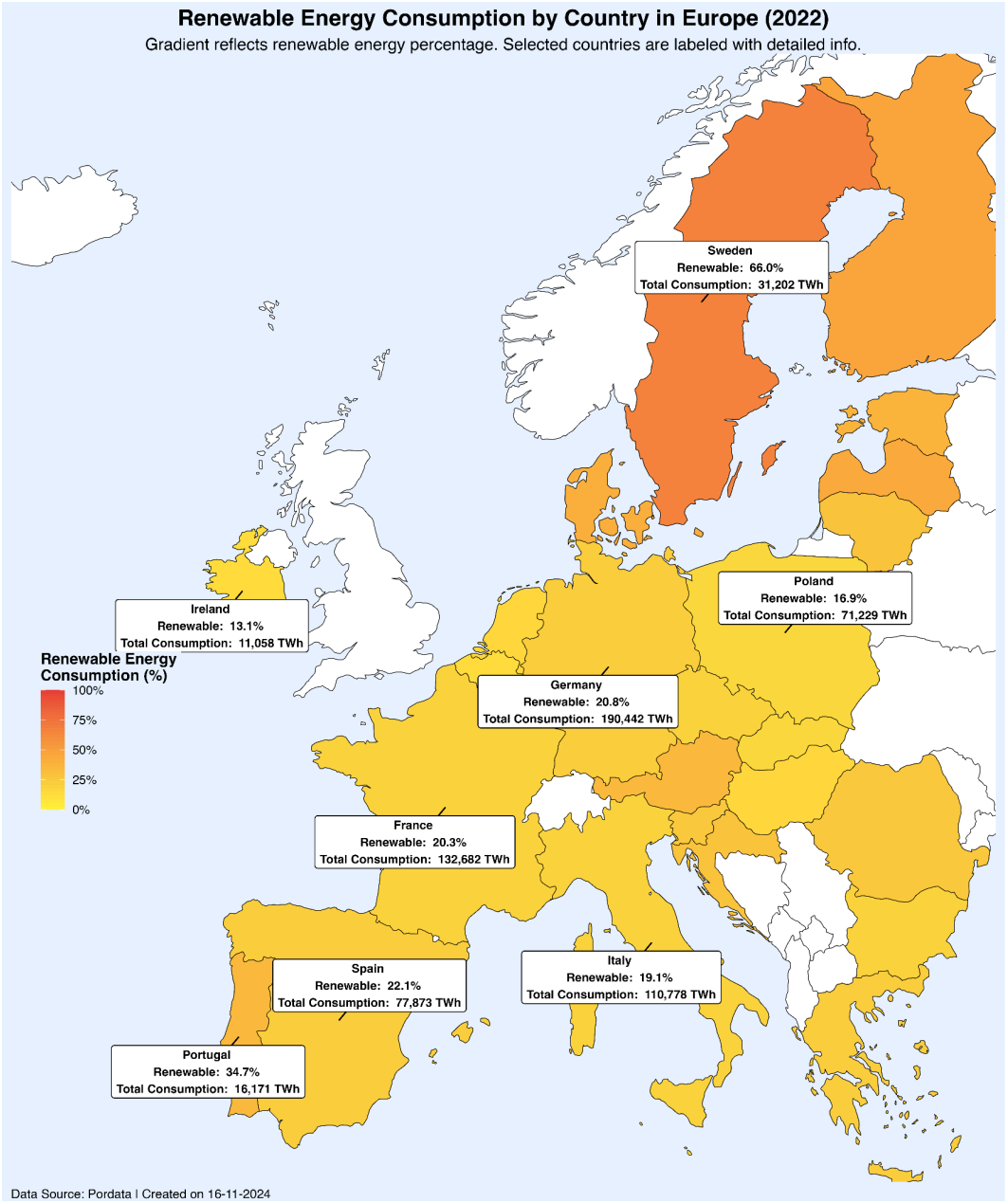

Image 2: Heatmap

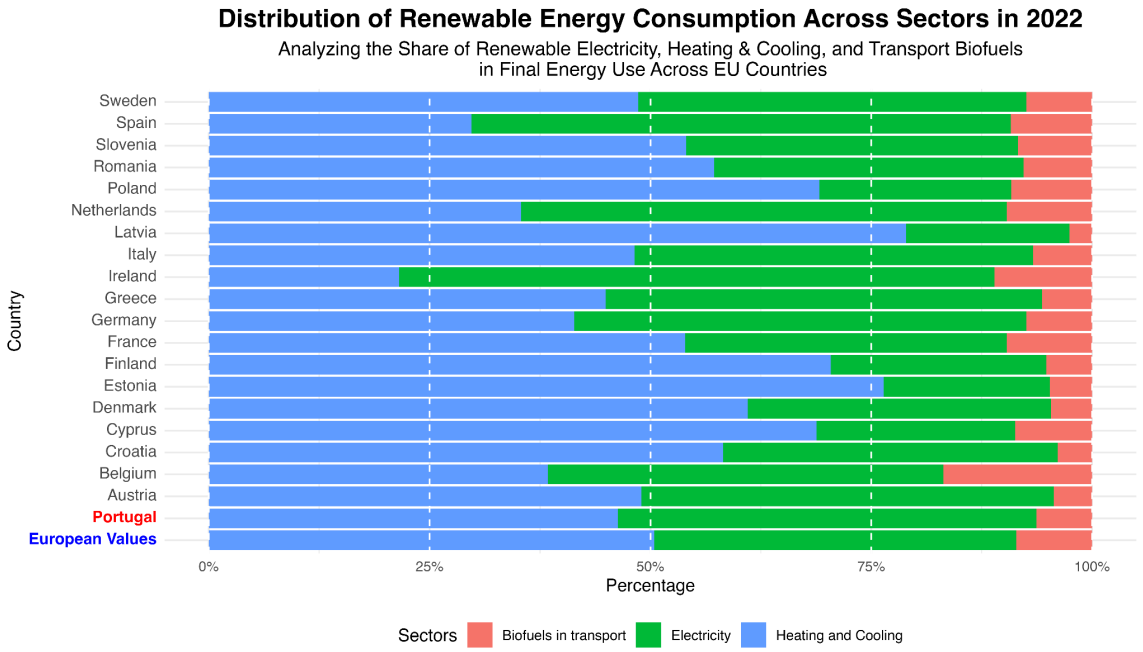

Image 3: Stacked Bar Graph