European Wind and Renewable Energy Trends (2016-2018)

Stat 184

Jacob Gavin, Sahana Ramachandran, Linh Tran, Yuv Boghani

Analyzing European Energy Production: Wind, Renewable, and Fossil Fuels (2016-2018)

1. Introduction

In recent years, there has been a significant shift in the European Union's (EU) energy landscape, driven by efforts to reduce carbon emissions and transition towards more sustainable energy sources. The EU has increasingly focused on expanding renewable energy, such as wind, solar, and hydro, while gradually decreasing its reliance on fossil fuels. This shift is evident in the energy production data from 2016 to 2018, sourced from Eurostat, which tracks energy generation across various sources including conventional thermal, nuclear, hydro, wind, solar, and geothermal energy.

Our project aims to explore these changes in the EU's energy production between 2016 and 2018, identifying key trends and understanding how the energy mix is evolving.

Specifically, we focus on how the growth of renewable energy sources, such as wind and solar, has impacted the overall energy supply and the contributions of traditional energy sources.

By analyzing this data, we aim to answer questions about the sustainability and resilience of the EU's energy sector, the role of different energy sources in meeting demand, and the challenges and opportunities that lie ahead in the transition to cleaner energy.

2. Research Questions

1. How much do the two sources of energy (renewable and non-renewable) contribute to net production in the EU across 2016-18?

2. What are the trends and variations in energy production among different renewable sources over 2016-18?

3. What are the net trade balances for different countries in the EU as of 2018, and how do they compare geographically?

4. Which EU countries are ranked in the top 5 for the highest and lowest renewable energy production in 2018, and what are their key sources?

5. How did the total renewable energy production and imports impact an EU country's net trade balance in 2018?

3. Data Description and Provenance

The dataset was sourced from Eurostat, the statistical office of the European Union, which collects and provides comprehensive and standardized data for EU member countries.

Eurostat aims to support evidence-based policymaking and research across Europe. This dataset focuses on energy production trends, examining renewable and non-renewable energy sources, as well as cross-border energy flows.

The dataset contains energy production data for European countries from 2016 to 2018, classified by energy sources such as wind, nuclear, solar, and thermal. It provides a basis for examining energy sustainability, shaping policy decisions, and understanding the transition to renewable energy.

Data Dictionary:

- country: Two-letter ISO code for the country
- country_name: Full name of the country
- type: Energy production source
- 2016, 2017, 2018: Annual energy production data in Gigawatt hours (GWh)
- net_production, imports, exports: Metrics on total production and cross-border energy flow

The FAIR principles ensure that this dataset is Findable through its clear documentation and structured format from Eurostat. It is Accessible in open formats like .csv and .xlsx, ensuring compatibility with tools like R and Python.

The use of standardized units (e.g., Gigawatt hours) and ISO country codes ensures Interoperability across platforms, while its consistent formatting and clear variable descriptions make it Reusable for various analysis, including energy sustainability research.

The CARE principles emphasize Collective Benefit by supporting renewable energy policies that address climate change. Authority to Control is respected by following Eurostat's guidelines on data use, ensuring ethical and responsible sharing.

The dataset promotes Responsibility and Ethics by fostering transparency and sustainable energy practices, benefiting society as a whole.

To conclude this section, future data colleciton could include extending the dataset to include more recent years, or tracking other variables, such as greenhouse gas emissions.

This would provide a more holistic view of which countries are "net zero" in terms of emissions. By integrating the data we might be able to gain a better insight into the effectiveness of renewable energy initiatives, and locate where countries need to improve their efforts.

In addition, a comparative analysis between EU and non-EU countries could also be useful in tracking how the continent as a whole is doing compared to other areas of the world.

4. Data Wrangling and Cleaning

To clean the data before it was ready for visualization, a few measures had to be taken, using `dplyr` and `stringr` packages in R. First, rows with a missing value in the 4th column were removed, as they were not suitable to be used for visualization. To make the countries easier to work with, a new column, "country" was created by removing all digits from the country codes, that way only the human readable part was left over. An additional check was performed, that will replace the value of "country" to NA if the length of the value was less than or equal to 1. Fills were then completed on rows without a country code, with the last non-NA value above it. Additionally, columns …1, …2, and …14 through …18 were removed from the data set, as they were considered extraneous.

First, rows with a missing value in the 4th column were removed, as they were not suitable to be used for visualization. To make the countries easier to work with, a new column, "country" was created by removing all digits from the country codes, that way only the human readable part was left over.

An additional check was performed, that will replace the value of "country" to NA if the length of the value was less than or equal to 1. Fills were then completed on rows without a country code, with the last non-NA value above it. Additionally, columns …1, …2, and …14 through …18 were removed from the data set, as they were considered extraneous.

Then, the first column is selected for more cleaning. Digits are removed, the substring "of which:" is removed, and periods (".") are removed, and extra whitespace is removed.

The data is then stepped through by blocks of 34 rows, corresponding to the start and end rows for each country. columns 1,2, and 3 are converted into rows, and new column ranges are computed.

The last two rows are then removed, and a stats list is appended to each row. `glue` is then used to create an excel-style range from the row and column bounds. The column names are set to years, and rows are more narrowly filtered.

The data was structured in a way that adhered to the TIDY principles covered in class, with one atomic record in a single row. this allowed analysis to be done

The original data set was composed of multiple tables, one for each country in the European Union that was being tracked. An example of one such table tracking UK energy production can be seen below.

	2016	2017	2018	2017/2016	2018/2017
Total net production	324,274	323,440	317,376	-0.3%	-1.9%
Of which:					
Conventional Thermal	203,165	189,297	180,838	-6.8%	-4.5%
Nuclear	65,149	63,887	59,098	-1.9%	-7.5%
Hydro	8,287	8,723	7,679	5.3%	-12.0%
of which: pumped	2,949	2,862	2,516	-2.9%	-12.1%
Wind	37,263	50,004	56,904	34.2%	13.8%
Solar	10,411	11,525	12,857	10.7%	11.6%
Geothermal	0	0	0	N/A	N/A
Other	0	0	0	N/A	N/A
Imports	20,018	18,167	21,332	-9.2%	17.4%
Exports	2,273	3,407	2,225	49.9%	-34.7%
Pump eng. Absorbed	4,014	3,859	3,391	-3.9%	-12.1%
Energy Supplied	338,005	334,340	333,092	-1.1%	-0.4%

Given that this data is quite compartmentalized, it makes sense to move the data into a format that is more easy to operate on, which is how we ended up with the two new csv data sets.

Moving the data into multiple files allows us to manipulate it with more precision. Ensuring that when performing analysis, we do not have to sift through and filter rows that are not applicable to the research question that we are trying to answer.

Below are shown the heads of the two "new" csv files that were made from data derived from the original Excel file.

Data set 1, (country_totals.csv)

```
  country country_name                 type level X2016     X2017
1      BE      Belgium Total net production Total 82520 82948.500
2      BE      Belgium              Imports Total 14648 14189.400
3      BE      Belgium              Exports Total  8465  8167.800
```

```
4        BE      Belgium Energy absorbed by pumping Total  1475  1485.400
5        BE      Belgium           Energy supplied Total 87228 87484.700
6        BG     Bulgaria      Total net production Total 41221 41351.303
7        BG     Bulgaria                   Imports Total  4568  3705.423
8        BG     Bulgaria                   Exports Total 10940  9185.794
9        BG     Bulgaria Energy absorbed by pumping Total   918   949.527
10       BG     Bulgaria           Energy supplied Total 33931 34921.405
       X2018
1  69212.347
2  21635.908
3   4308.347
4   1347.901
5  85192.007
6  41705.000
7   2223.000
8  10030.000
9    425.000
10 33473.000
```

Data set 2, (energy_types.csv)

```
   country country_name                 type   level X2016    X2017     X2018
1       BE      Belgium Conventional thermal Level 1 30728 31316.00 30092.635
2       BE      Belgium              Nuclear Level 1 41430 40128.50 26995.628
3       BE      Belgium                Hydro Level 1  1476  1360.90  1239.248
4       BE      Belgium   Pumped hydro power Level 2  1110  1093.20   983.190
5       BE      Belgium                 Wind Level 1  5340  6387.90  7177.346
6       BE      Belgium                Solar Level 1  3070  3264.30  3488.979
7       BE      Belgium           Geothermal Level 1     0     0.00     0.000
8       BE      Belgium                Other Level 1   476   490.90   218.509
9       BG     Bulgaria Conventional thermal Level 1 18909 20234.21 19334.000
10      BG     Bulgaria              Nuclear Level 1 14932 14718.37 15290.000
```

5. Exploratory Data Analysis (EDA)

Below is Summary Statistics for Conventional Thermal Energy Production a for Energy Types produced in the years 2016, 2017, and 2018

```
Attaching package: 'dplyr'
```

```
The following objects are masked from 'package:stats':

    filter, lag


The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union

  Statistic Year_2016  Year_2017  Year_2018
1      Min.      0.00      0.000      0.000
2   1st Qu.   4241.00   4394.819   4003.505
3    Median  18065.01  20135.846  16350.754
4      Mean  48687.88  49696.821  46065.660
5   3rd Qu.  47817.00  47246.628  48042.915
6      Max. 390141.00 376128.000 320437.701
```

This allows us to get a broad overview of what the landscape of energy production is like in the EU for the years specified.

In addition, below is a summary table for Total energy produced in the years 2016, 2017, and 2018

```
  Statistic Year_2016  Year_2017  Year_2018
1      Min.    807.16   1594.459   1888.398
2   1st Qu.  11365.00  11315.800  10834.143
3    Median  41221.00  41351.303  41705.000
4      Mean 101391.35 102397.845 101018.719
5   3rd Qu. 149067.00 144905.800 146845.187
6      Max. 614155.00 619059.000 571799.713
```

This is yet another summary table that will allow us to get an overview of what energy production looks like in the EU, however this data set does not concern itself necessarily with the TYPE of energy produced, rather the net production, imports, and exports of a given country. This data lends itself more to the economic factors surrounding energy production, compared to the previous summary table, which is grouped by TYPES of energy produced

Below is a summary of the most lucrative methods of energy production in the EU from 2016-2018

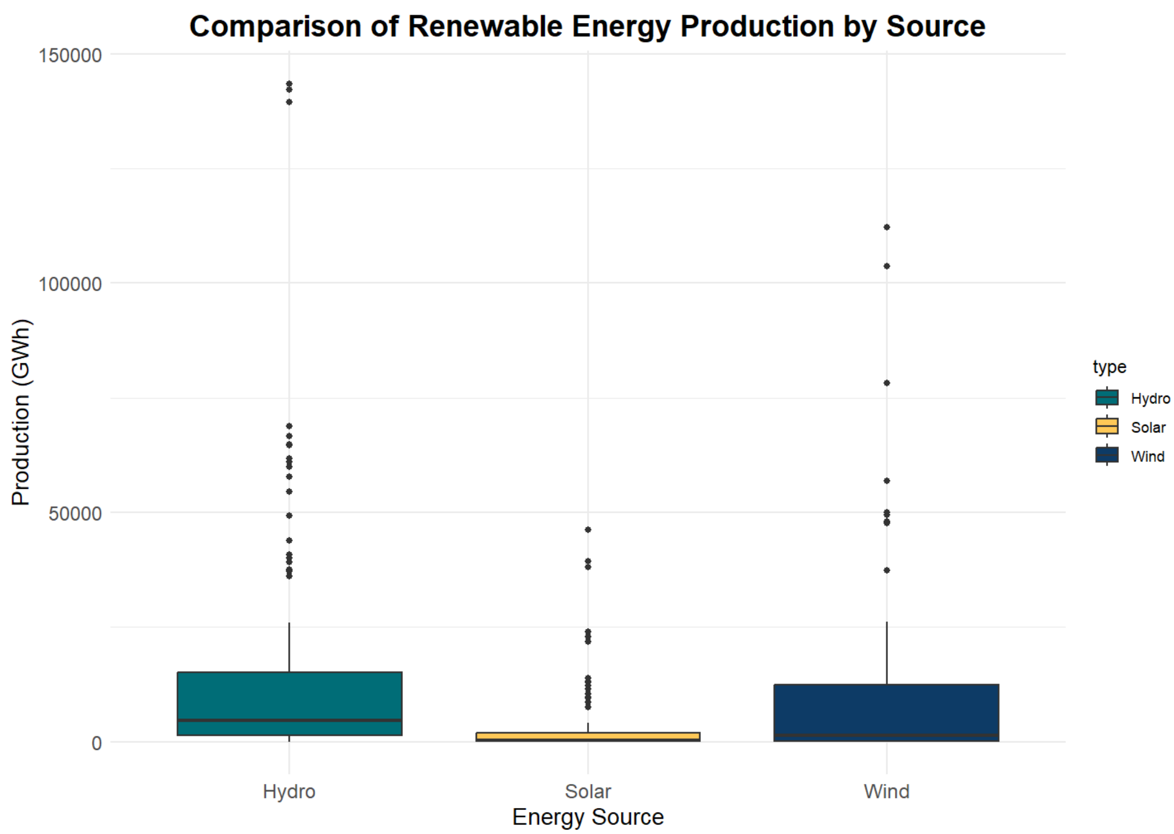
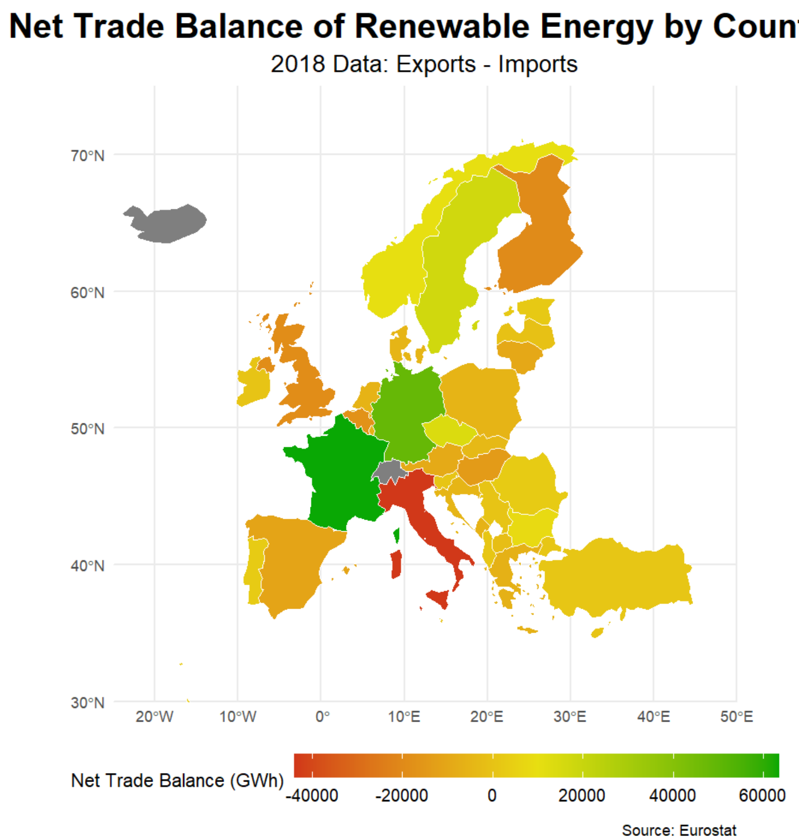


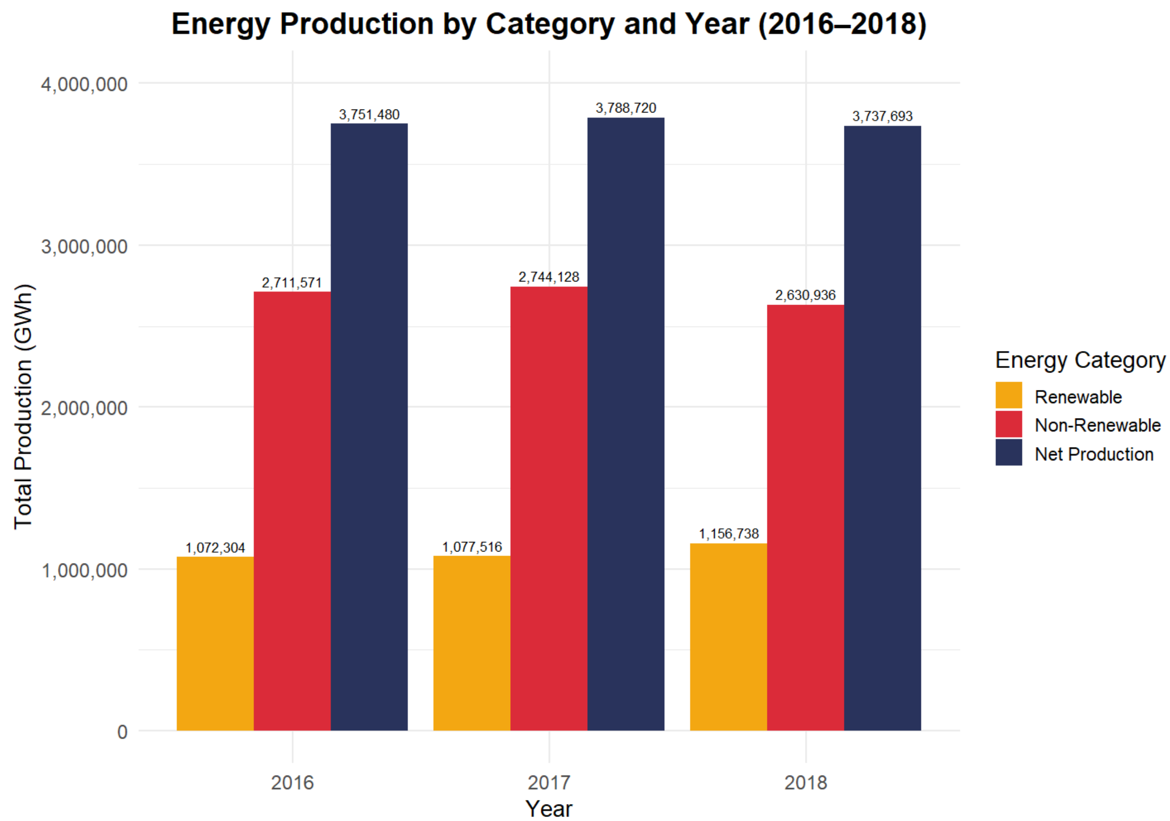
Figure 1: Energy Production Boxplot

As you can see, much of the data lies on the lower end of the boxplot, highlighting that there are a few "outliers" that account for vastly more energy production than even the rest of the top 1/4 producers. This is true for each of the three types of renewable energy being tracked.

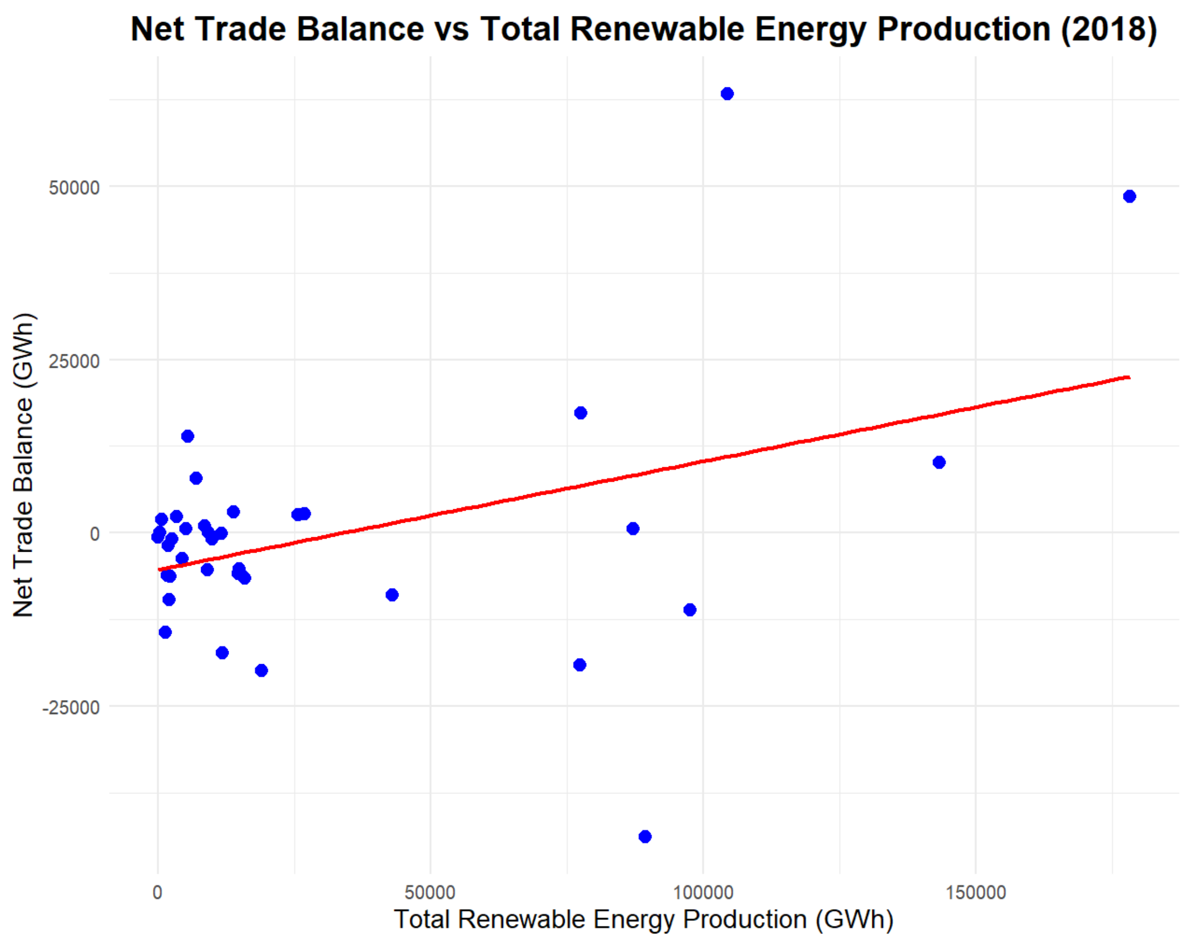


This color-coded map displays which countries in the EU contribute and benefit most from power sharing. At the top of the charts is France, and what looks to be the most reliant on outside energy is Italy. This visualization provides a purely visual insight into the way that power is distributed among countries, and which countries are "sucklers" on the European power grid.

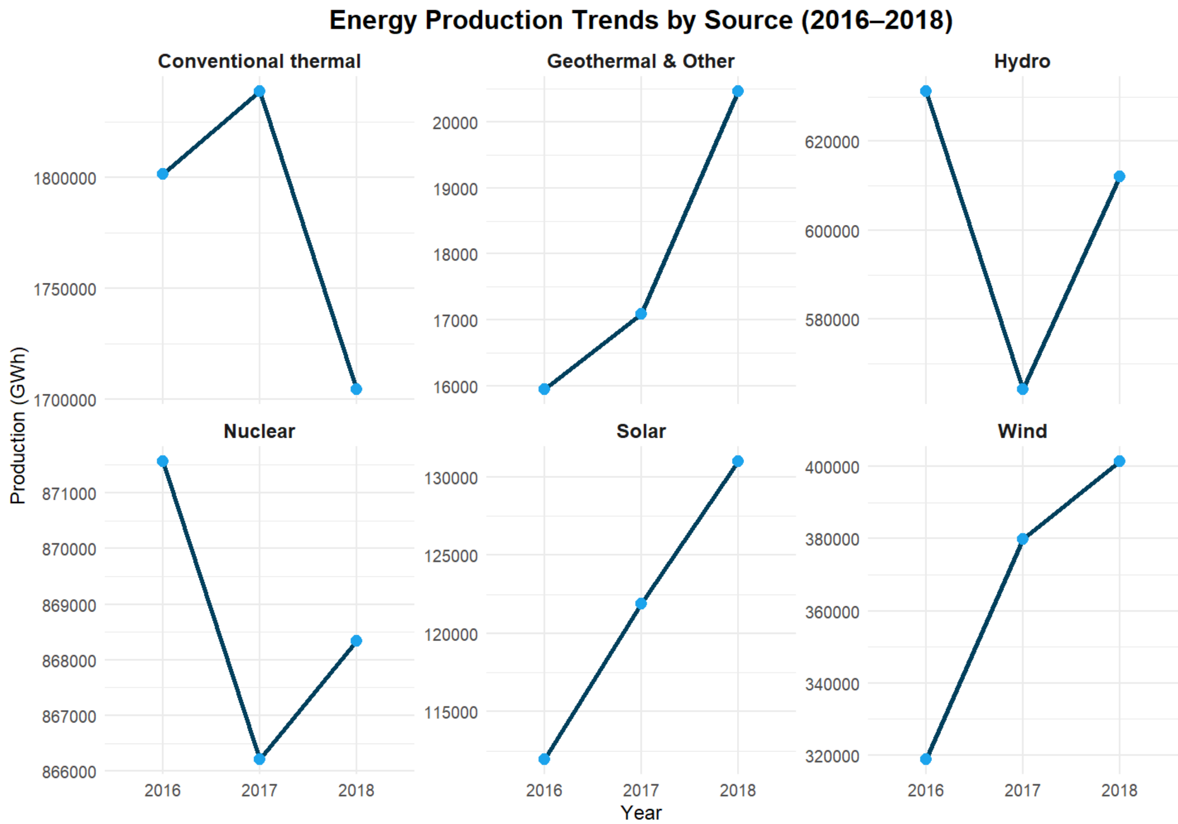
This visualization provides key insights into the steps that individual countries are taking towards self-sustaiment as well as moving towards renewable energy. Naturally the shift to renewable energy is something that will take quite some time, and data visualization like this will allow us to better track the progress of countries.



Attached above is a bar chart of energy production across the years 2016-2018. You can see that there is an increase in renewable energy production across the three years, signifying more of an emphasis being placed on renewable energy. However, the visualization provides more insight than just that. The increase in renewable energy from 2016 to 2017 is a mere ~5,000 GWh, but the increase in renewable energy production from 2017 to 2018 is ~77,000 GWh. This shows that not onlt is there an increase in production of renewable energy, the rate at which countries are switching to renewable energy is ALSO increasing! This is good to see, that the switch to renewable energy is actually picking up steam. Early reports show that this trend is still being followed, although that data is outside of the scope of this report.



This is perhaps one of the most profound findings of the study. Although the coefficient of correlation is relatively weak, there is in fact a correlation between countries that generate enough power to have a surplus, and countries that produce more renewable energy. This has already taken into account the proportions of renewable vs total energy production. Naturally, this is not a purely causative relationship, as there are many other factors to take into account, such as GDP, Leadership, and other factors that may confound this correlation, but the data shows that countries that produce more renewable energy tend to have a surplus of energy, hinting at its effectiveness and necessity going forward in the journey to a more renewable power grid.



This is another very interesting visual. This graphic provides insight not only into the amount of each kind of energy being produced, but also highlights the ways in which their production is changing, such as the “geothermal / other” seeing a sharp uptick from 2017 to 2018, or the dip and recovery that was experienced by both nuclear and Hydro energy. While surely there are many many factors that influence how much of a certain power type is produced, this visual gives an insight into the short-term trends of the energy production types. That is to say, this gives us insight into the “What” but not necessarily the “Why”

```
# A tibble: 8 x 4
  type                 Total_2016 Total_2017 Totat_2018
  <chr>                     <dbl>      <dbl>      <dbl>
1 Conventional thermal   1801452.   1838783.   1704429.
2 Geothermal               10282.     11395.     12444.
3 Hydro                   631264.    564250.    611994.
4 Nuclear                 871560.    866212.    868334.
5 Other                     5670.      5692.      8019.
6 Pumped hydro power       32889.     33444.     50154.
7 Solar                    111948.    121911.    130968.
8 Wind                     318810.    379965.    401332.
```

Below is shown a table of the totals of the energy produced through the years 2016-2018

Type of Energy	2016 Total	2017 Total	2018 Total
Conventional Thermal	1,801,451.654	1838783.375	1704429.420
Geothermal	10,281.953	11,394.997	12,443.754
Hydro (traditional)	631,263.647	564,249.938	611,993.531
Nuclear	871,560.205	866,211.980	868,333.983
Hydro (pumped)	32,889.277	33,443.817	50,153.928
Solar	111,948.354	121,910.507	130,968.490
Wind	318,809.583	379,964.520	401,332.380
Other	5,670.308	5,692.438	8,109.144

6. Analysis of Research Questions

1. How much do the two sources of energy (renewable and non-renewable) contribute to net production in the EU across 2016-18?

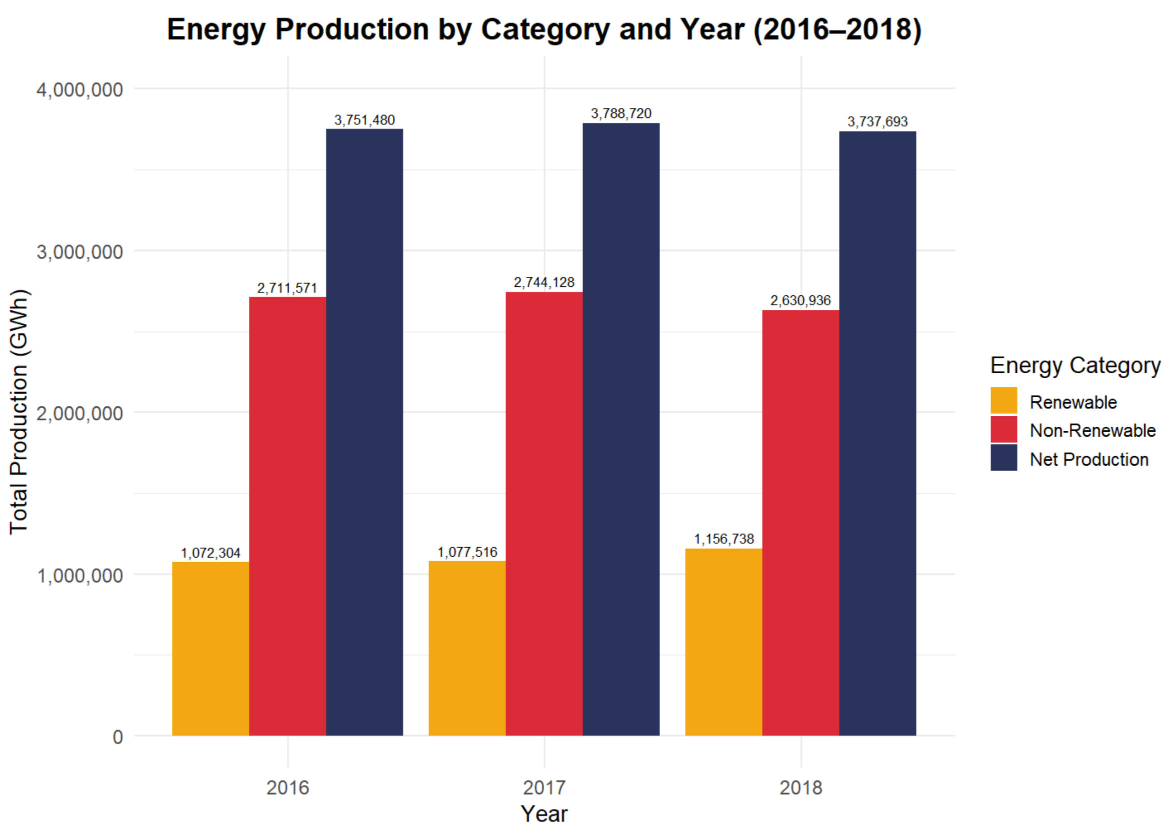
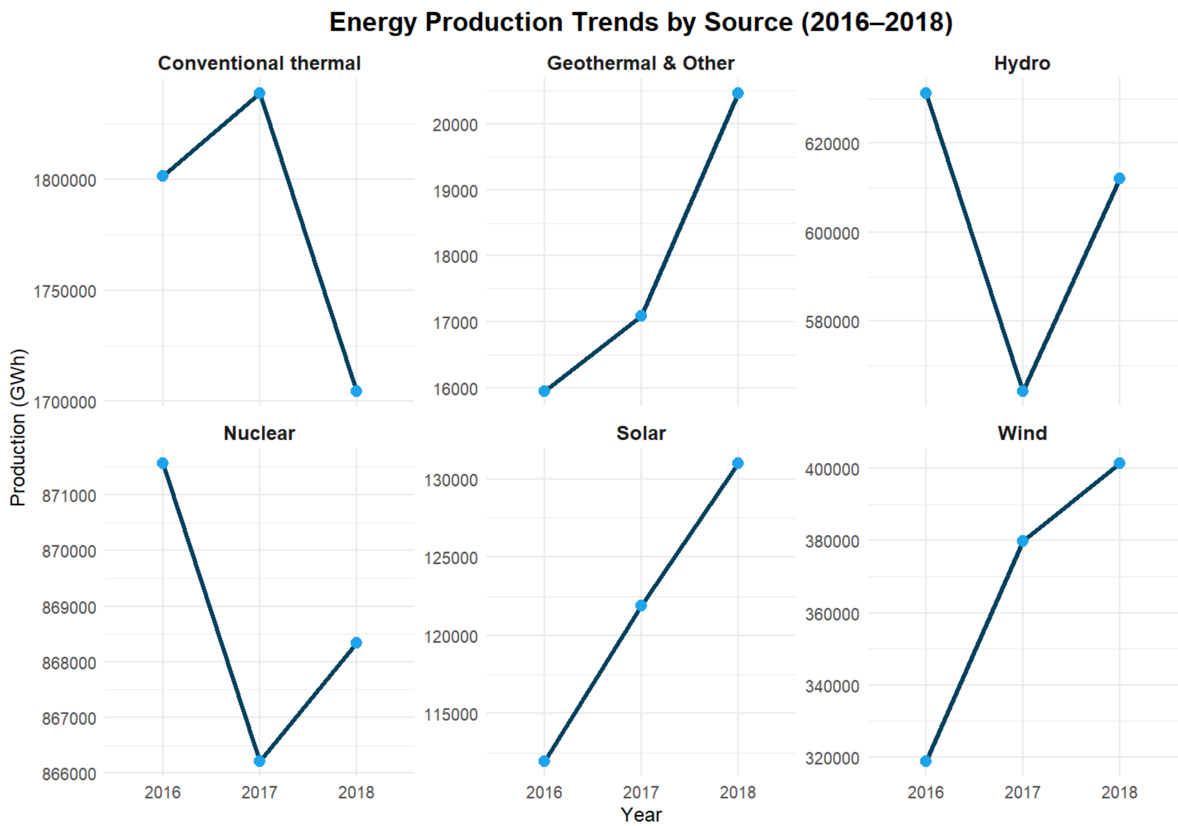
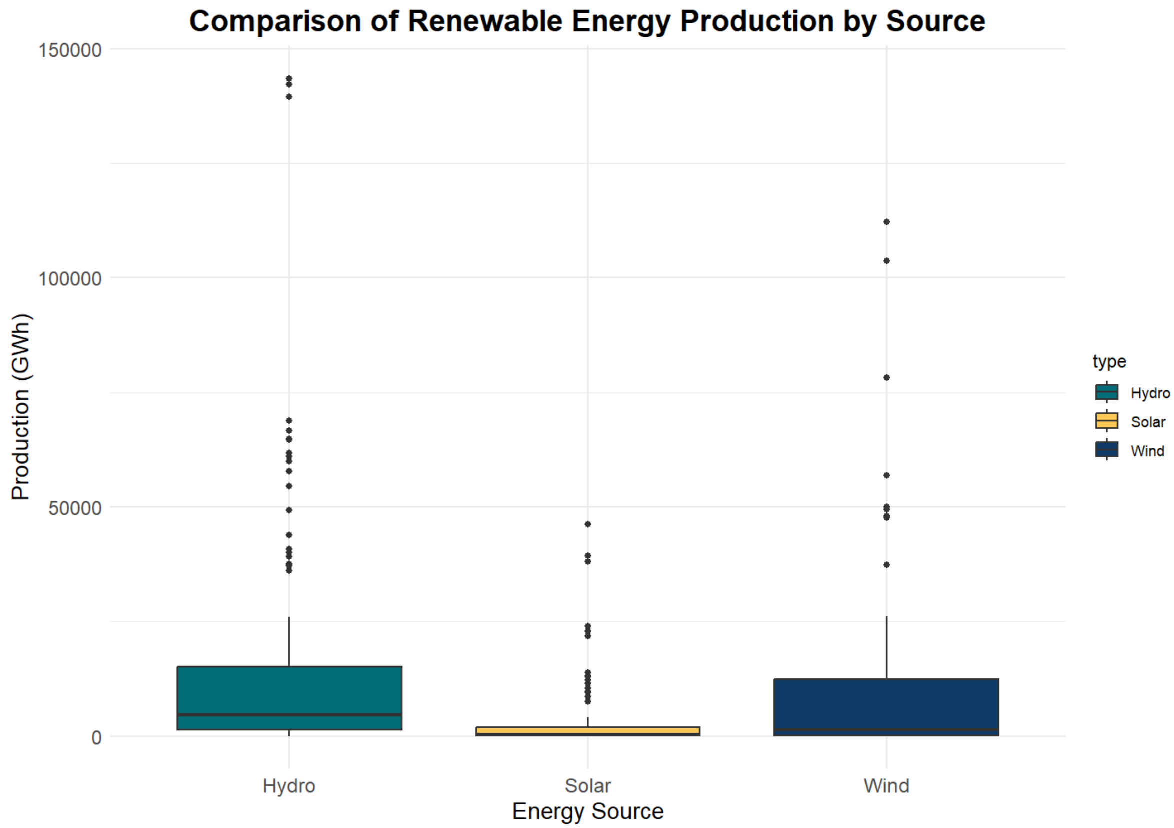


Figure 2: Energy Production by Category and Year

Renewable energy production increased steadily from 1,072,304 GWh in 2016 to 1,156,738 GWh in 2018, indicating a gradual shift toward sustainable sources. Non-renewable energy, while dominant, decreased slightly from 2,711,571 GWh in 2016 to 2,630,936 GWh in 2018. This reflects their efforts to switch to greener sources of energy. Net production remained relatively stable, with 3,751,480 GWh in 2016 and 3,737,693 GWh in 2018. Non-renewable energy consistently contributed around 70-72% of total production, while renewable energy accounted for approximately 28-30%. This shift shows progress in energy diversification driven by policy initiatives and advancements in renewable technologies. However, the dominance of non-renewables reiterates ongoing challenges in fully transitioning to sustainable energy production.

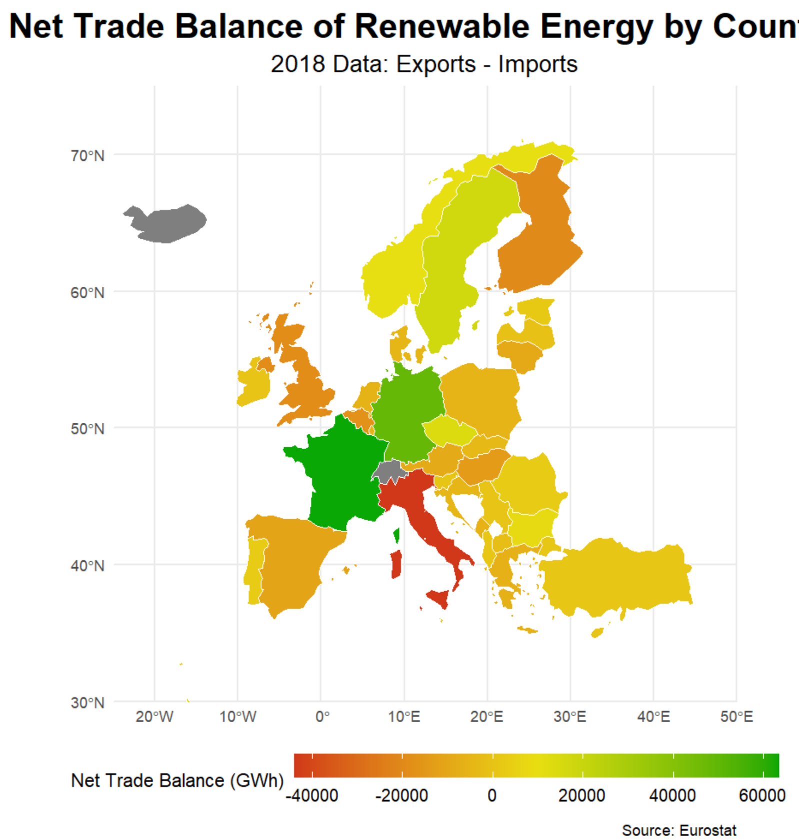
2. What are the trends and variations in energy production among different renewable sources over 2016-18?

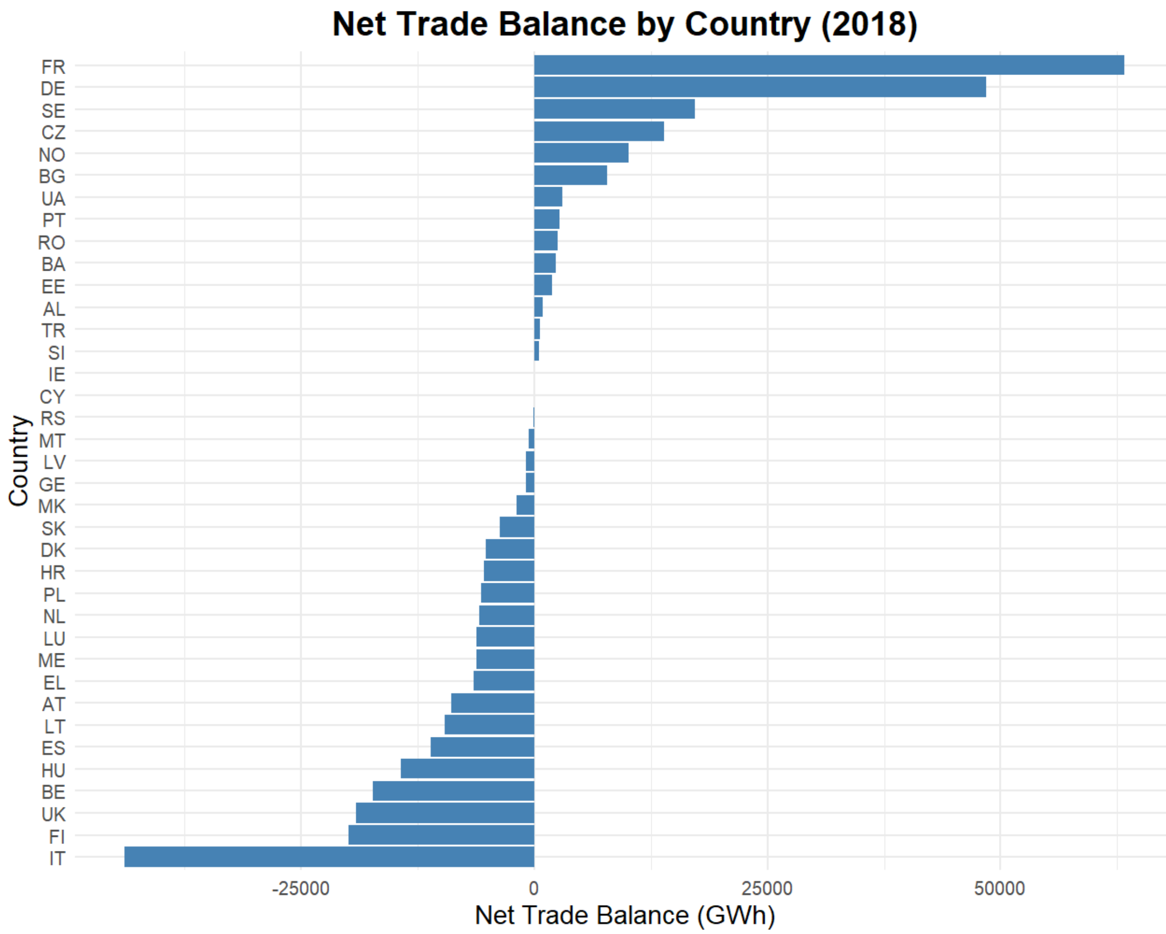




The data highlights the production trends and variations across renewable energy sources, namely hydro, solar, and wind. The trends show a sharp fluctuation in hydro energy, starting above 620,000 GWh in 2016, with a drop to 580,000 GWh in 2017, then rebounding by 2018. Solar energy shows steady growth, from approximately 115,000 GWh in 2016 to around 130,000 GWh in 2018. There is also significant growth in wind energy production, starting from approximately 320,000 GWh in 2016 and surpassing 400,000 GWh by 2018. The box plot further illustrates variability, showing that hydro has the highest production levels and variability, with some outliers. Wind follows in terms of production levels but exhibits more consistent outputs. Solar remains the smallest contributor among the three despite consistent growth. These insights show a growing reliance on wind and solar energy, aligning with technological advancements and policy shifts toward sustainable sources. Hydro, while the most substantial renewable energy source, demonstrates variability, likely influenced by climatic factors like rainfall. Together, these trends highlight the importance of diversifying renewable energy sources to ensure consistent production levels.

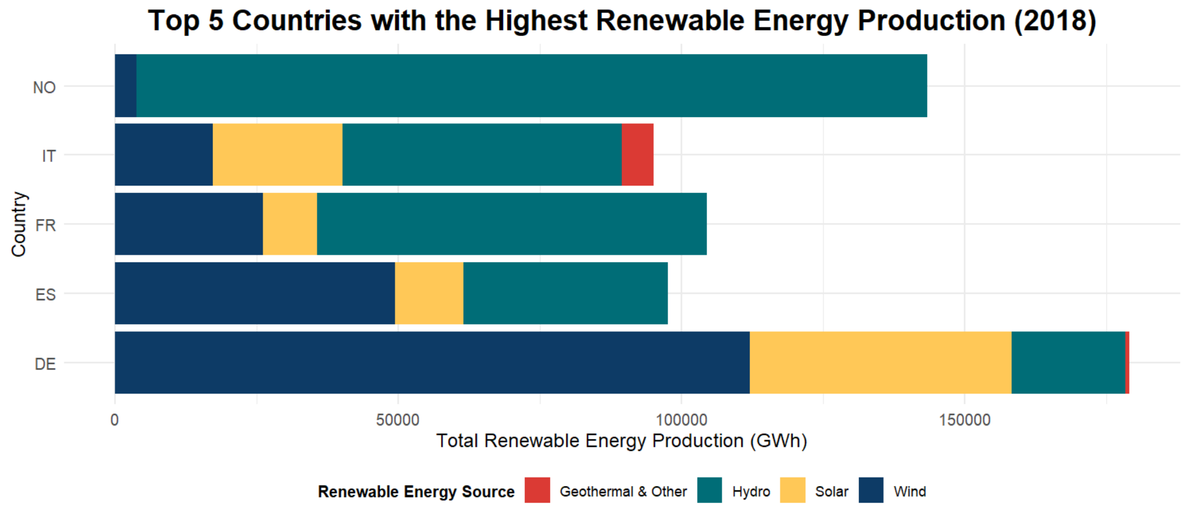
3. What are the net trade balances for different countries in the EU as of 2018, and how do they compare geographically?

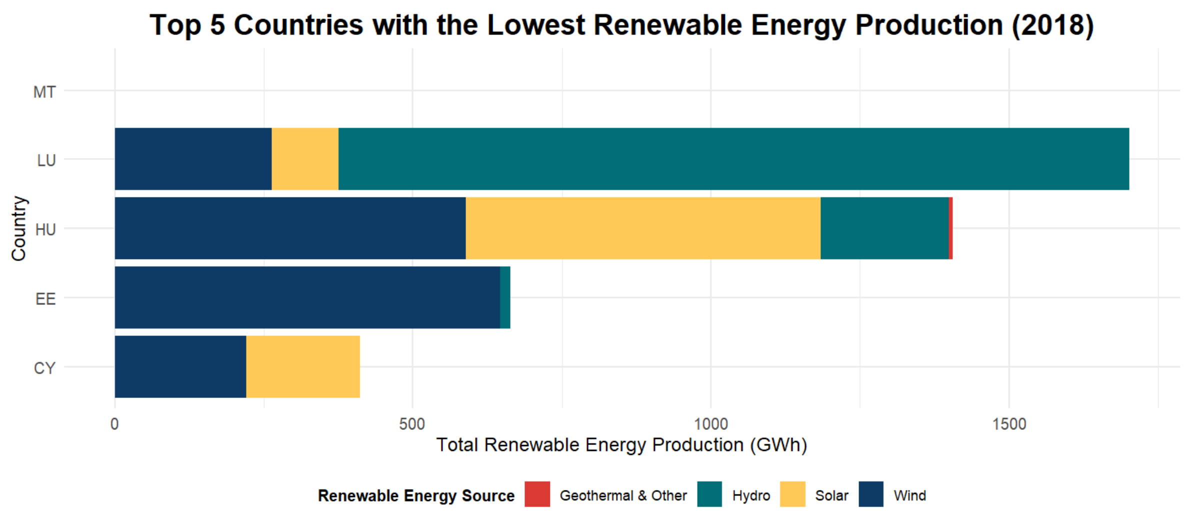

Net Trade Balance of Renewable Energy by Country
2018 Data: Exports - Imports
70°N
60°N
50°N
40°N
30°N
20°W
10°W
0°
10°E
20°E
30°E
40°E
50°E
Net Trade Balance (GWh)
-40000
-20000
0
20000
40000
60000
Source: Eurostat



The visualizations depict the net trade balances of renewable energy among EU countries in 2018. The bar chart shows the trade balance in GWh for each country. France, Germany, and Sweden had the highest positive trade balances, being significant exporters of renewable energy. Conversely, countries like Italy and Finland had the largest negative balances, being heavily reliant on imports to meet energy demands. The map visualization provides a geographic perspective, with green regions representing countries with an energy surplus and red regions indicating deficits. Particularly, Western and Northern European countries, such as France, Germany, and Sweden, emerge as renewable energy hubs, while Southern European countries, including Italy, appear to be more dependent on imports. This reflects the regional disparities in renewable energy production and consumption. These insights underline the interdependence among EU countries in renewable energy trade and indicate the collaborative energy strategies used to balance production and consumption.

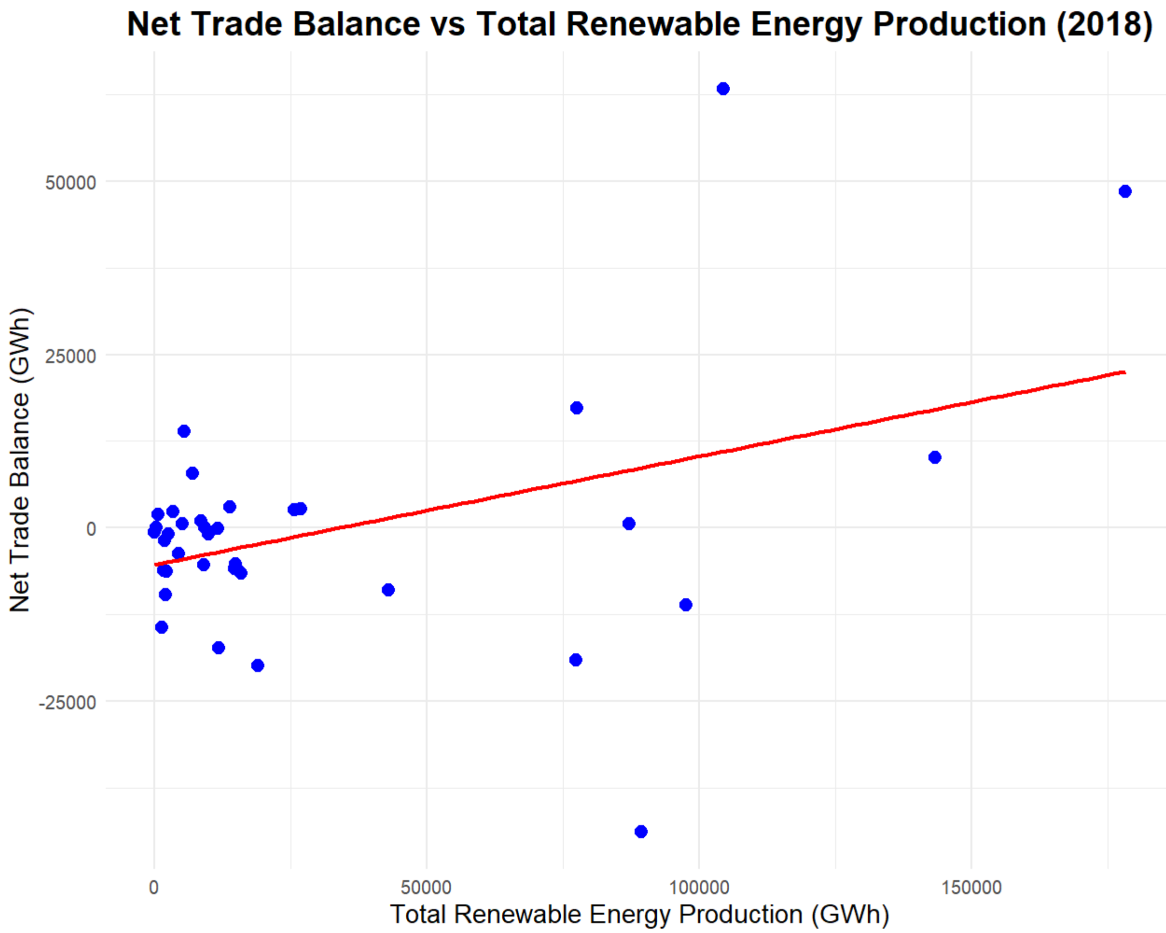
4. Which EU countries are ranked in the top 5 for the highest and lowest renewable energy production in 2018, and what are their key sources?

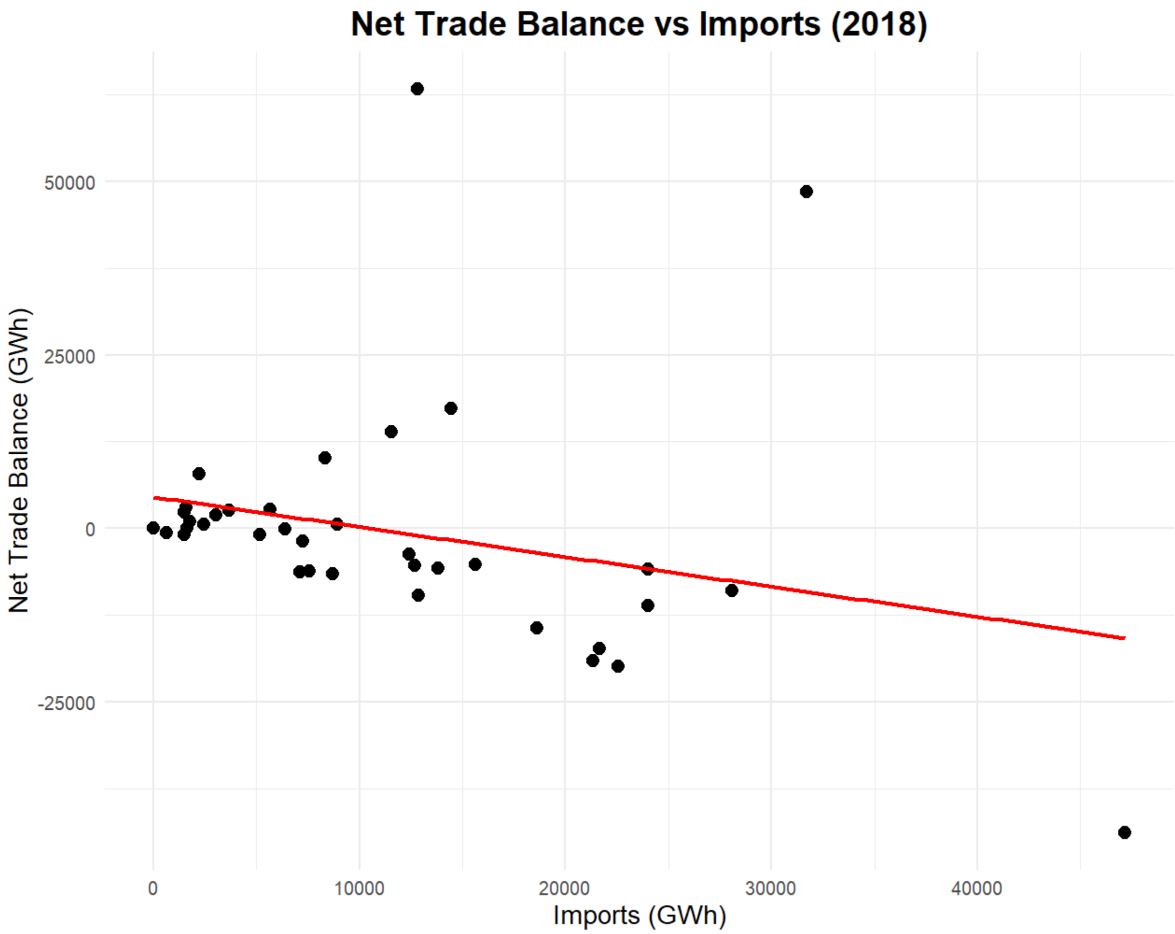




The graphs show the inconsistency in renewable energy production among EU countries in 2018. The bar chart for the top 5 countries with the highest renewable energy production shows that Norway dominates with the largest share, primarily driven by hydro energy, followed by Germany, where wind energy contributes the most. Other mentions include France, Spain, and Italy, with a mix of hydro, solar, and wind energy sources. Conversely, the chart for the lowest producers shows that Malta, Luxembourg, Hungary, Estonia, and Cyprus generated minimal renewable energy. These countries relied primarily on hydro and wind, with very little contribution from solar or geothermal sources. The stark discrepancies between the highest and lowest producers underlines the geographic and infrastructural variations in renewable energy potential and adoption. Countries like Norway and Germany leverage their natural resources and advanced infrastructure to lead in production, whereas smaller nations like Malta and Luxembourg face challenges due to limited resources or geographical constraints. These findings further emphasize the need for regional cooperation to bridge the renewable energy gap across the EU.

5. How did the total renewable energy production and imports impact an EU country's net trade balance in 2018?





The first scatterplot illustrates a positive correlation of 0.385 between net trade balance and total renewable energy production in 2018. As the production increases, net trade balances also tend to rise, as indicated by the upward-sloping trendline. This suggests that countries with higher renewable energy production are better positioned to generate surplus, or being less reliant on imports, ultimately contributing to a stronger trade balance. However, the variability in data points indicates that other factors, such as domestic consumption and export policies, also influence net trade balance. The second scatterplot examines the relationship between net trade balance and imports, showing a negative correlation of -259l. The downward-sloping trendline indicates that higher energy imports are associated with lower or negative net trade balances. This pattern highlights the economic disadvantage of reliance on imports, as it exacerbates a country's trade deficit. These visualizations emphasize both the critical role of renewable energy production in bolstering trade balance and the economic challenges posed by energy import dependency. They highlight the importance of policies promoting renewable energy production to enhance trade stability and reduce reliance on external energy sources.

This is an important question to answer because it allows to get a better read into the energy “leadership” of a country, ie. which countries are “leading the charge” in energy production. This likely corresponds to bargaining power in the continental stage, as energy is one of the most important commodities for a country to have.

Answering this question is key to gauging how close the European union is to it’s sustainable energy goals. Regardless of actual renewable energy production, we are able to see the rate at which countries are making a shift to renewable energy, which serves as a marker of whether or not a country is on track to meet the goals set by and for them, in terms of renewable energy. This is especially important when considering how much of a push the EU is making towards renewable energy. Discuss how your data addresses each research question. Analyze the results with the help of the visualizations and any relevant statistics. Use narrative text to provide context and explain what the visualizations reveal about the data.

7. Additional Analysis (Optional)

If applicable, apply more advanced statistical methods (e.g., hypothesis testing, regression analysis, ANOVA) to provide further insights. Mention any trends or anomalies discovered in the data and discuss their implications.

8. Conclusion

The analysis of European energy production from 2016 to 2018 highlights a transformative shift towards renewable energy, driven by the European Union’s sustainability goals and the global urgency to reduce carbon emissions. While conventional thermal energy and nuclear power still dominate the overall energy mix, the rapid growth of wind and solar energy production signals a strong commitment to renewable sources. Wind energy, for eg, grew by 34.2% between 2016 and 2017, with an additional 13.8% increase from 2017 to 2018, reflecting significant investment in wind infrastructure. Solar energy production also demonstrated steady growth, rising by 10.7% and 11.6% year-over-year during the same period. This trend underscores the increasing focus on harnessing clean energy sources to transition towards a more sustainable and resilient energy system. However, this shift is not uniform across all member states, as larger economies like France and Germany continue to dominate energy production, creating a notable disparity in output across the EU.

France’s heavy reliance on nuclear power and Germany’s diversified energy mix place these nations at the forefront of the EU’s energy landscape, while smaller or less industrialized countries contribute significantly less to the overall production. Such disparities emphasize the need for more balanced energy policies that promote investment in renewable energy infrastructure across all member states, particularly in underrepresented regions. Moreover, while the analysis provides valuable insights into energy production trends, it leaves out critical factors like consumption patterns and the socio-economic impacts of the energy transition, which are equally important in understanding the complete picture of Europe’s energy landscape.

Moving forward, addressing regional inequalities and maintaining the momentum in renewable energy adoption will be essential to achieving a cohesive and sustainable energy future for the European Union.

Research Questions:

1. How have renewable energy trends evolved across European countries between 2016 and 2018? 2. What factors contribute to the disparities in energy production across EU member states?

Limitations:

1. The analysis relies on Eurostat data, which may have reporting inconsistencies or gaps for certain countries. 2. The study focuses on production data, leaving out consumption patterns or economic impacts of the energy transition.

1. How does the total energy production vary across European countries in 2018? Which are the top 5 most produced ?

In 2018, total energy production varied substantially across European countries, largely depending on their resource availability, infrastructure, and policy focus. France, Germany, the United Kingdom, Italy, and Spain emerged as the top five energy producers, with France notably leading due to its extensive nuclear capacity and Germany following closely with a more diversified energy mix.

2. What is the proportion of energy production for the European countries in 2018?

The proportion of energy production was unevenly distributed, with a few large economies dominating the output. France held a significant share, followed by Germany, while other countries produced comparatively smaller proportions. Overall, a handful of major players accounted for the bulk of Europe's total energy supply in 2018.

3. How have renewable energy trends changed over time (2016–2018) for European countries?

Between 2016 and 2018, renewable energy production in Europe increased steadily, with wind and solar sources experiencing particularly strong growth. Not only did the volume of renewable energy rise, but the rate of growth also accelerated over this period, reflecting the EU's strengthening commitment to reducing fossil fuel dependence and achieving sustainability targets.

9. References

European Commission. (n.d.). Electricity generation statistics – first results. Eurostat Statistics Explained. Retrieved December 18, 2024, from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Archive:Electricity_generation_statistics_%E2%80%93_first_results For the GitHub repository: Ford, R. (2020, August 4). TidyTuesday dataset: Renewable energy. GitHub. Retrieved December 18, 2024, from https://github.com/rfordatascience/tidytuesday/blob/main/data/2020/2020-08-04/Electricity_generation_statistics_2019.xlsx

10. Appendix: Code

```
# BOAST Coding Style
# Load Required Libraries
library(tidyverse)
library(readxl)
library(countrycode)
library(glue)
library(ggplot2)
library(dplyr)
library(eurostat)
library(tidyr)

# Set Working Directory
#setwd("C:/Users/linht/Github/Sec2__FP_MichaelAmbrose_JacobGavin_SahanaRamachandran_LinhTran_

# Load and Prepare Raw Data ----
raw_code <- countrycode::codelist %>%
  select(country_name = country.name.en, country = eurostat)

file_path <- "..data/Electricity_generation_statistics_2019.xlsx"
raw_excel <- read_excel(file_path, sheet = 3)

# Clean and Transform Raw Data
raw_excel_clean <- raw_excel %>%
  filter(!is.na(...4)) %>%  # Remove rows with missing data
  mutate(
    country = str_remove_all(...4, "[:digit:]"),  # Extract country names
    .before = ...1
  ) %>%
  mutate(
    country = if_else(
```

```
    str_length(country) > 1, country, NA_character_  # Validate country names
  ),
  country = str_extract(country, "[:alpha:]+")  # Retain only alphabetic characters
) %>%
fill(country) %>%  # Fill down country names
select(-c(...1, ...2, ...14:...18))  # Drop irrelevant columns

# Extract Row Statistics
row_stat <- read_excel(
  path = file_path,
  sheet = 3,
  range = "C48:C61",
  col_names = FALSE
)[[1]][c(1, 3:14)] %>%
  str_remove("[:digit:]") %>%
  str_remove("of which: ") %>%
  str_remove("\\.") %>%
  str_trim()  # Clean string data

# Create Country Ranges
country_range <- tibble(
  row_start = seq(from = 46, to = 454, by = 34),
  row_end = seq(from = 61, to = 469, by = 34)
) %>%
  mutate(
    col1 = 4,
    col2 = col1 + 5,
    col3 = col2 + 5
  ) %>%
  pivot_longer(
    cols = col1:col3,
    names_to = "col_var",
    values_to = "col_start"
  ) %>%
  mutate(col_end = col_start + 2) %>%
  select(-col_var) %>%
  slice(-n(), -(n() - 1)) %>%
  mutate(row_stat = list(row_stat))

# Define Function for Processing Country Data
get_country_stats <- function(row_start, row_end, col_start, col_end, row_stat) {
  col_range <- glue("{LETTERS[col_start]}{row_start}:{LETTERS[col_end]}{row_end}")
```

```r
  raw_data <- suppressMessages(read_excel(
    file_path, sheet = 3, col_names = FALSE, range = col_range
  ))
  country_data <- raw_data %>%
    set_names(nm = c(2016:2018)) %>%
    filter(!is.na(`2016`), `2016` != "2016") %>%
    mutate(
      country = if_else(is.na(`2017`), `2016`, NA_character_),
      .before = `2016`
    ) %>%
    fill(country) %>%
    filter(!is.na(`2017`)) %>%
    mutate(
      type = row_stat,
      .after = country,
      level = c("Total", "Level 1", "Level 1", "Level 1", "Level 2",
                "Level 1", "Level 1", "Level 1", "Level 1", "Total",
                "Total", "Total", "Total")
    ) %>%
    mutate(across(c(`2016`:`2018`), as.double))  # Convert columns to numeric
  return(country_data)
}

# Process and Organize Data
all_countries <- country_range %>%
  pmap_dfr(get_country_stats) %>%
  left_join(raw_code, by = "country") %>%
  select(country, country_name, everything())

country_totals <- all_countries %>%
  filter(level == "Total")  # Filter for total production data

energy_types <- all_countries %>%
  filter(level != "Total")  # Filter for specific energy types

# Save Processed Data
write.csv(country_totals, "data/country_totals.csv", row.names = FALSE)
write.csv(energy_types, "data/energy_types.csv", row.names = FALSE)

# Compute Summary Statistics
summary_table <- read.csv("data/country_totals.csv") %>%
  filter(type == "Total net production") %>%  # Filter for 'Total net production'
```

```
  rename_with(make.names) %>%  # Ensure column names are syntactically valid
  summarise(
    Statistic = c("Min", "1st Qu.", "Median", "Mean", "3rd Qu.", "Max"),
    Year_2016 = summary(.$X2016)[1:6],
    Year_2017 = summary(.$X2017)[1:6],
    Year_2018 = summary(.$X2018)[1:6]
  )

# Calculate Energy Categories and Combine Data
combinedEnergyData <- bind_rows(
  energy_types %>%
    filter(type %in% c("Hydro", "Wind", "Solar", "Geothermal")) %>% # Select renewable sourc
    pivot_longer(cols = c(`2016`, `2017`, `2018`), names_to = "year", values_to = "productio
    group_by(year) %>%
    summarize(totalProduction = sum(production, na.rm = TRUE)) %>% # Calculate total renewab
    mutate(category = "Renewable"),
  energy_types %>%
    filter(!type %in% c("Hydro", "Wind", "Solar", "Geothermal")) %>% # Select non-renewable s
    pivot_longer(cols = c(`2016`, `2017`, `2018`), names_to = "year", values_to = "productio
    group_by(year) %>%
    summarize(totalProduction = sum(production, na.rm = TRUE)) %>% # Calculate total non-ren
    mutate(category = "Non-Renewable"),
  country_totals %>%
    filter(type == "Total net production") %>%
    pivot_longer(cols = c(`2016`, `2017`, `2018`), names_to = "year", values_to = "productio
    group_by(year) %>%
    summarize(totalProduction = sum(production, na.rm = TRUE)) %>% # Calculate total net pro
    mutate(category = "Net Production")
) %>%
  mutate(category = factor(category, levels = c("Renewable", "Non-Renewable", "Net Productio

# Plot Bar Chart
## Energy Production by Category and Year (2016-2018)
ggplot(combinedEnergyData, aes(x = year, y = totalProduction, fill = category)) +
  geom_bar(stat = "identity", position = "dodge") + # Create a grouped bar chart
  geom_text(
    aes(label = scales::comma(totalProduction)),
    position = position_dodge(width = 0.9),
    vjust = -0.5,
    size = 3
  ) +
  scale_fill_manual(values = c("Renewable" = "#f3a712", "Non-Renewable" = "#db2b39", "Net Pr
```

```r
  scale_y_continuous(
    limits = c(0, 4000000),  # Define y-axis limits
    breaks = seq(0, 4000000, by = 1000000),  # Define axis intervals
    labels = scales::comma_format() # Format y-axis labels as comma-separated numbers
  ) +
  labs(
    title = "Energy Production by Category and Year (2016-2018)", # Add chart title and axis
    x = "Year",
    y = "Total Production (GWh)",
    fill = "Energy Category"
  ) +
  theme_minimal(base_size = 14) +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5), # Center and style the
    axis.title = element_text(size = 14),
    axis.text = element_text(size = 12)
  )


# Prepare Data
energy_bar_data <- energy_types %>%
  filter(type != "Pumped hydro power") %>%  # Exclude "Pumped hydro power" as it's not a prin
  mutate(
    type = ifelse(type %in% c("Geothermal", "Other"), "Geothermal & Other", type)  # Group "(
  ) %>%
  group_by(type) %>% # Group by energy source
  summarise(
    `2016` = sum(`2016`, na.rm = TRUE), # Aggregate production for each year
    `2017` = sum(`2017`, na.rm = TRUE),
    `2018` = sum(`2018`, na.rm = TRUE)
  ) %>%
  pivot_longer(
    cols = c(`2016`, `2017`, `2018`),  # Convert year columns to long format
    names_to = "Year",
    values_to = "Production"
  )
# Create Small Multiples Charts
## Energy Production Trends by Source (2016-2018)
ggplot(energy_bar_data, aes(x = Year, y = Production, group = type)) +
  geom_line(size = 1.2, color = "#003f5c") +  # Dark blue for lines
  geom_point(size = 3, color = "#1ca3ec") +  # Lighter blue for points
  facet_wrap(~ type, scales = "free_y") +
```

```r
  labs(
    title = "Energy Production Trends by Source (2016-2018)",
    x = "Year",
    y = "Production (GWh)"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 16, face = "bold"),
    strip.text = element_text(size = 12, face = "bold"),
    axis.text = element_text(size = 10),
    axis.title = element_text(size = 12)
  )

# Define Custom Color Palette
custom_colors <- c(
  "Hydro" = "#006d77",
  "Wind" = "#0d3b66",
  "Solar" = "#ffc857",
  "Geothermal & Other" = "#db3a34",
  "Conventional thermal" = "#9966cc",
  "Nuclear" = "#6a0572"
)

# Boxplot for Selected Renewable Energy Sources
selected_energy_data <- energy_types %>%
  filter(type %in% c("Hydro", "Wind", "Solar")) %>%  # Filter only "Hydro," "Wind," and "Sola
  pivot_longer(cols = c(`2016`, `2017`, `2018`),
               names_to = "Year", values_to = "Production")


# Visualize with Boxplot
ggplot(selected_energy_data, aes(x = type, y = Production, fill = type)) +
  geom_boxplot() +
  scale_fill_manual(values = custom_colors) +
  labs(
    title = "Comparison of Renewable Energy Production by Source",
    x = "Energy Source",
    y = "Production (GWh)"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
```

```r
    axis.title = element_text(size = 14),
    axis.text = element_text(size = 12)
  )

# Net Trade Balance Calculation and Plotting
data_clean <- country_totals %>%
  filter(type %in% c("Imports", "Exports")) %>%
  pivot_longer(
    cols = c(`2016`, `2017`, `2018`),
    names_to = "Year",
    values_to = "Value"
  ) %>%
  pivot_wider(
    names_from = type,
    values_from = Value
  ) %>%
  mutate(Net_Trade_Balance = Exports - Imports) %>%
  filter(Year == "2018")

ggplot(data_clean, aes(x = reorder(country, Net_Trade_Balance), y = Net_Trade_Balance)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  coord_flip() +
  labs(
    title = "Net Trade Balance by Country (2018)",
    x = "Country",
    y = "Net Trade Balance (GWh)"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    axis.title = element_text(size = 12),
    axis.text = element_text(size = 10)
  )

# Geospatial Mapping for Net Trade Balance
renewable_countries <- energy_types %>%
  filter(type %in% c("Wind", "Solar", "Hydro", "Geothermal & Other", "Nuclear", "Conventiona
  select(country) %>%
  distinct()

renewable_trade_balance <- country_totals %>%
  filter(country %in% renewable_countries$country) %>%
```

```r
  filter(type == "Imports" | type == "Exports") %>%
  pivot_wider(
    names_from = type,
    values_from = c(`2016`, `2017`, `2018`)
  ) %>%
  mutate(
    net_trade_balance_2018 = `2018_Exports` - `2018_Imports`
  ) %>%
  select(country, net_trade_balance_2018)

europe_map <- get_eurostat_geospatial(resolution = "60", nuts_level = 0)

europe_map <- europe_map %>%
  left_join(renewable_trade_balance, by = c("CNTR_CODE" = "country"))

ggplot(europe_map) +
  geom_sf(aes(fill = net_trade_balance_2018), color = "white", size = 0.2) +
  scale_fill_gradientn(
    colors = c("#d13819", "#e8df11", "#09a804"),
    name = "Net Trade Balance (GWh)"
  ) +
  labs(
    title = "Net Trade Balance of Renewable Energy by Country",
    subtitle = "2018 Data: Exports - Imports",
    caption = "Source: Eurostat"
  ) +
  coord_sf(
    xlim = c(-25, 50),
    ylim = c(30, 75),
    expand = FALSE
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 20, face = "bold"),
    legend.position = "bottom"
  )

# Summarize renewable energy production
data_clean <- energy_types %>%
  mutate(type = ifelse(type %in% c("Geothermal", "Other"), "Geothermal & Other", type)) %>%
  filter(type %in% renewable_sources) %>%
  group_by(country, type) %>%
```

```
  summarise(
    `2016` = sum(`2016`, na.rm = TRUE),
    `2017` = sum(`2017`, na.rm = TRUE),
    `2018` = sum(`2018`, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  pivot_longer(
    cols = c(`2016`, `2017`, `2018`),
    names_to = "Year",
    values_to = "Production"
  )

# Total renewable production per country
total_renewable_production <- data_clean %>%
  filter(Year == "2018") %>%
  group_by(country) %>%
  summarise(Total_Production = sum(Production, na.rm = TRUE), .groups = "drop") %>%
  arrange(desc(Total_Production))

# Top 5 and bottom 5 countries for renewable production
top_5_countries <- total_renewable_production %>% slice_head(n = 5)
bottom_5_countries <- total_renewable_production %>% slice_tail(n = 5)

data_top_5 <- data_clean %>% filter(country %in% top_5_countries$country, Year == "2018")
data_bottom_5 <- data_clean %>% filter(country %in% bottom_5_countries$country, Year == "2018

# Plot renewable energy trends
top_5_most_plot <- ggplot(data_top_5, aes(x = country, y = Production, fill = type)) +
  geom_bar(stat = "identity", position = "stack") +
  coord_flip() +
  scale_fill_manual(values = custom_colors) +
  labs(
    title = "Top 5 Countries with the Highest Renewable Energy Production (2018)",
    x = "Country",
    y = "Total Renewable Energy Production (GWh)",
    fill = "Renewable Energy Source"
  ) +
  theme_minimal()

top_5_least_plot <- ggplot(data_bottom_5, aes(x = country, y = Production, fill = type)) +
  geom_bar(stat = "identity", position = "stack") +
  coord_flip() +
```

```r
  scale_fill_manual(values = custom_colors) +
  labs(
    title = "Top 5 Countries with the Lowest Renewable Energy Production (2018)",
    x = "Country",
    y = "Total Renewable Energy Production (GWh)",
    fill = "Renewable Energy Source"
  ) +
  theme_minimal()

# Display plots
print(top_5_most_plot)
print(top_5_least_plot)

# Merge renewable production with trade balance
combined_data <- total_renewable_production %>%
  left_join(net_trade_data, by = "country") %>%
  drop_na()

# Correlation and Visualizations
# Correlation between total renewable production and net trade balance
correlation_renewable <- cor(combined_data$Total_Production, combined_data$Net_Trade_Balance)
print(paste("Correlation between Total Renewable Energy Production and Net Trade Balance: ",

# Scatter plot: Net trade balance vs renewable production
plot_renewable_vs_trade <- ggplot(combined_data, aes(x = Total_Production, y = Net_Trade_Bala
  geom_point(color = "blue", size = 3) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(
    title = "Net Trade Balance vs Total Renewable Energy Production (2018)",
    x = "Total Renewable Energy Production (GWh)",
    y = "Net Trade Balance (GWh)"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    axis.title = element_text(size = 14),
    axis.text = element_text(size = 10)
  )

# Correlation between imports and net trade balance
data_imports <- country_totals %>%
  filter(type %in% c("Imports", "Exports")) %>%
```

```r
  pivot_longer(
    cols = c(`2016`, `2017`, `2018`),
    names_to = "Year",
    values_to = "Value"
  ) %>%
  filter(Year == "2018") %>%
  pivot_wider(
    names_from = type,
    values_from = Value
  ) %>%
  mutate(Net_Trade_Balance = Exports - Imports) %>%
  select(country, Net_Trade_Balance, Imports)

correlation_imports <- cor(data_imports$Net_Trade_Balance, data_imports$Imports, use = "compl
print(paste("Correlation between Imports and Net Trade Balance: ", round(correlation_imports

# Scatter plot: Net trade balance vs imports
plot_imports_vs_trade <- ggplot(data_imports, aes(x = Imports, y = Net_Trade_Balance)) +
  geom_point(color = "black", size = 3) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(
    title = "Net Trade Balance vs Imports (2018)",
    x = "Imports (GWh)",
    y = "Net Trade Balance (GWh)"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5),
    axis.title = element_text(size = 14),
    axis.text = element_text(size = 10)
  )

# Display Plots
print(plot_renewable_vs_trade)
print(plot_imports_vs_trade)
```