|  |
| --- |
| **Information Visualization project** |
| Dependency of European Union on Energy imports from Russia |
|  |
| Dávid Juraj Szücs  Erick Fernando Villa Rodriguez  Esteban Velásquez Rendón |

Table of Contents

[1 Project presentation 3](#_Toc104758026)

[1.1 Target user 3](#_Toc104758027)

[1.2 Target task 3](#_Toc104758028)

[2 Datasets 3](#_Toc104758029)

[3 Data preprocessing 5](#_Toc104758030)

[4 Platform 7](#_Toc104758031)

[Python 7](#_Toc104758032)

[Panel 7](#_Toc104758033)

[Bokeh 7](#_Toc104758034)

[Plotly and Plotly express 7](#_Toc104758035)

[5 Visualization 8](#_Toc104758036)

[5.1 Components of the visualization 8](#_Toc104758037)

[5.1.1 Choropleth Map 8](#_Toc104758038)

[5.1.2 Treemap 8](#_Toc104758039)

[5.1.3 Timeline graph 9](#_Toc104758040)

[5.1.4 Summary table 9](#_Toc104758041)

[5.1.5 Control elements 9](#_Toc104758042)

[5.2 Interactivity 9](#_Toc104758043)

[5.2.1 Selecting a Country: Clicking on the map 9](#_Toc104758044)

[5.2.2 Selecting Dataset and Year: Via Radio buttons and Year slider -> effect on which graphs 10](#_Toc104758045)

[6 Evaluation 11](#_Toc104758046)

[6.1 External feedback 12](#_Toc104758047)

[7 Conclusion 13](#_Toc104758048)

# Project presentation

This paper presents the report that follows the visualization prepared for the project: Dependency of the European Union on energy imports from Russia.

Energy security is one of the most important elements in assuring the basic needs of a country, such as its stability, and economic development. This has been put in question in recent years and especially in the last few months with the shift in the geopolitical situation with regards to Russia. This visualization tool thus proposes to bring the information on the energy imports from Russia relative to the total import. This information is to be presented with a help of a choropleth map, a treemap, and a line graph with various interactivity levels.

The general overview of the project is summarized in Table 1.

Table 1: General overview of the project

|  |  |  |
| --- | --- | --- |
| **Title** | Dependency of the European Union on energy imports from Russia | |
| **Target user** | General European public |
| **Target task** | Informative |
| **Main data source** | Eurostat |
| **Main platforms used** | Python, Bokeh, Panel, Plotly |

## Target user

As specified in Table 1, the main target user for the visualization is the general public primarily located in EU countries. This mainly includes the adult citizens and residents that are allowed to vote in these countries. However, this tool can also be used as a quick reference tool for governmental bodies or politicians.

## Target task

This visualization tool aims to raise awareness of the scope to which countries of the European Union are dependent on the energy imports from the Russian Federation. Even though the general public has no direct possibility of choosing the source location of their energy, informed citizens can create political pressure and influence the decisions that the country takes indirectly via democratic elections.

# Datasets

We have used four main datasets with actual data on energy imports and energy dependency, and one additional dataset with countries’ polygons for the purposes of drawing a map in our visualization.

Besides the shapefile dataset of EU countries that was taken from the European Environmental Agency, all datasets were obtained from Eurostat. Eurostat is the main statistical office of the European Union which is interconnected with the EU’s national statistical offices via the European Statistical System (ESS). Eurostat provides high-quality data collection in a consistent and normative way that is bound by legislation for all member countries [1]. This assures a reliable source for the purposes of the visualization for this project.

Table 2 summarizes all datasets used with their respective formats and sources. The first three datasets provide absolute data on energy import for three respective resources, namely, oil and petroleum products, solid fossil fuel, and natural gas. All three datasets contain these import values including their importers and exporters. The fourth dataset specifies the values of energy import dependency of each individual country in the EU with respect due their own production or reserves.

Table 2: Summary of the main source datasets used for the visualization.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Dataset | Format | Source |
| 1. | Imports of oil and petroleum products by partner country 2000-2020 | .tsv | Eurostat [2] |
| 2. | Imports of solid fossil fuels by partner country 2000-2020 | .tsv | Eurostat [3] |
| 3. | Imports of natural gas by partner country 2000-2020 | .tsv | Eurostat [4] |
| 4. | Energy imports dependency - 2000-2020 | .tsv | Eurostat [5] |
| 5. | Shapefile with border lines of the European countries | .shp | European Environmental Agency [6] |

For all available datasets, Eurostat provides a tool called Data Browser that enables custom personalization of the raw data based on the desired filter and a mid-level exploratory browser for a quality check before further manipulation with data. This tool is depicted in a print screen in Figure 1 for one of the chosen datasets. The tool shows the possibility of data selection/filtering in the stage before downloading the dataset. Customization was applied to all datasets by choosing the desired geographical entities, the period between 2000-2020, annual frequency, and units. After having chosen the desired selection, the custom datasets were downloaded via the download button in .tsv format.

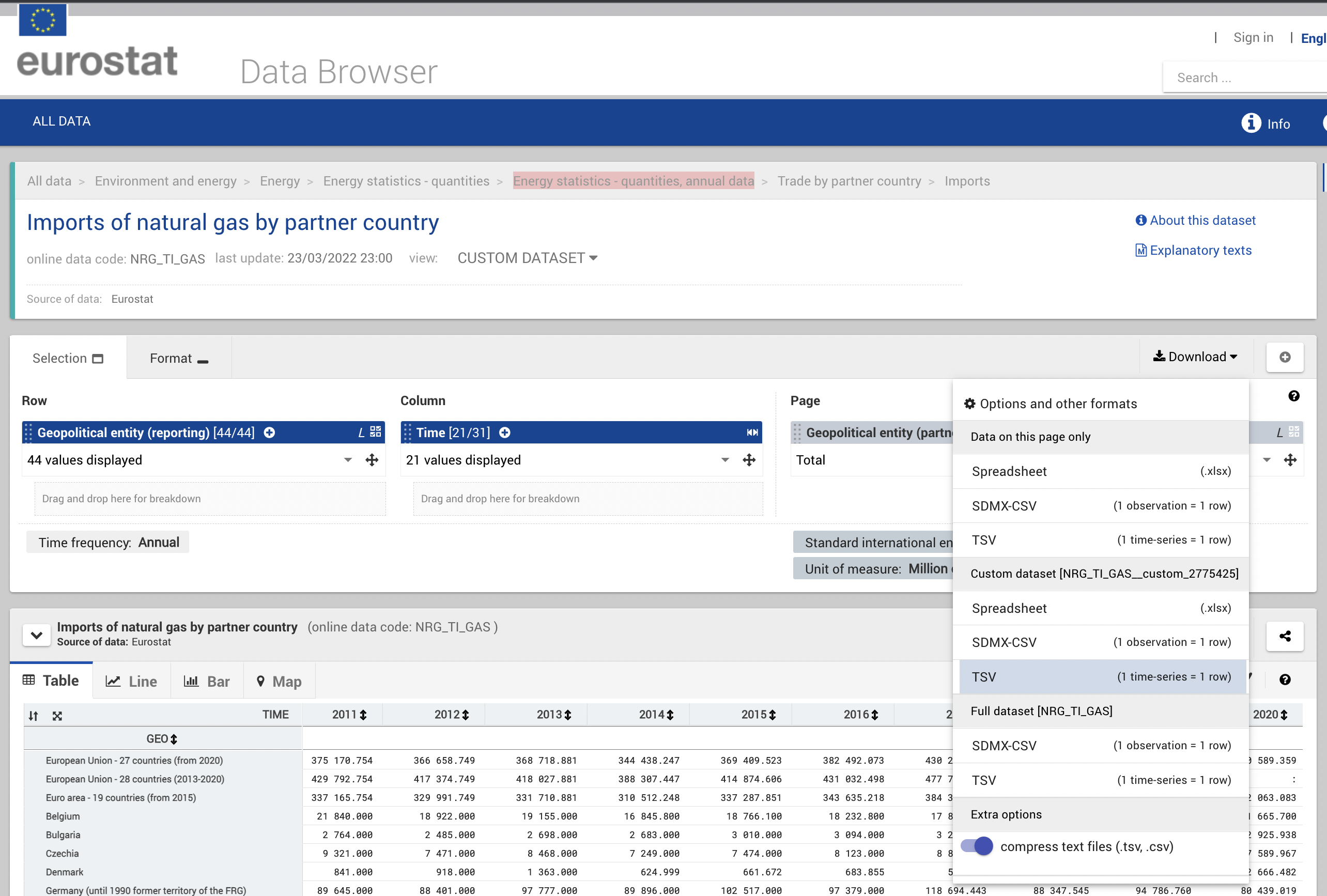


Figure 1: Print screen of Eurostat's Data Browser with exploration option of one of the datasets [4].

# Data preprocessing

To prepare the data ready for the applications in the Bokeh server, preprocessing steps had to be taken. Both algorithmic and formatting operations were applied to the datasets retrieved from the sources described in Section 2. Most of these operations were done using Pandas library in Python and were all done in steps in Jupyter Notebook, in the code file named *dataframes.ipynb*.

Figure 2 shows the tsv format in which the datasets 1-4 in Table 2 are downloaded. Besides the values of the energy product imports, it also shows frequency, product code, partner exporter, unit of the product, and geographical region (all merged in the first column). The first step in preprocessing the data was to split the first column and remove the unnecessary new columns created. Additional formatting operations done included replacing “:” with NaN, replacing country codes with the country names, reshaping the dataframe to the one with values in one column, merging the dataset with polygons of the dataset with the countries’ shapefiles (dataset 5 in Table 2), or expanding the dataframe by country’s continent.

Obrázok, na ktorom je text, potvrdenie

Automaticky generovaný popis

Figure 2: Example of one of the TSV files downloaded from Eurostat.

The results of these formatting processing steps are demonstrated in the examples in Figure 3 and Figure 4. Figure 3 shows an example of the final form of a dataframe needed for the map visualization including the data of the countries’ polygons. Figure 4 shows an example of the final form of a dataframe needed for the treemap with the categorical information about the continent of the exporter country.

|  |  |
| --- | --- |
| Figure 3: Example of the dataframe format needed for our map visualization | Figure 4: Example of the dataframe needed for our treemap visualization |

To get the relative values of dependency on Russia, algebraic preprocessing had to be done as well. This was done by combining and aggregating downloaded datasets from Eurostat. In the example of natural gas imports in Figure 5, rectangles filled with orange color represent those datasets that were downloaded from Eurostat, and the ones filled with white color stand for datasets aggregated. The aggregating of the two top datasets (Imports of natural gas by Russia and Imports of natural gas from TOTAL) is done by division with the result of a new dataframe: Relative dependency on imports from Russia. This dataset represents the percentage of the natural gas being imported from Russia relative to the country’s total natural gas import. If we want to represent the total relative dependency from Russia on natural gas, the previous relative dataset needs to be multiplied by the import dependency that also considers the domestic production of natural gas in the concerned country.

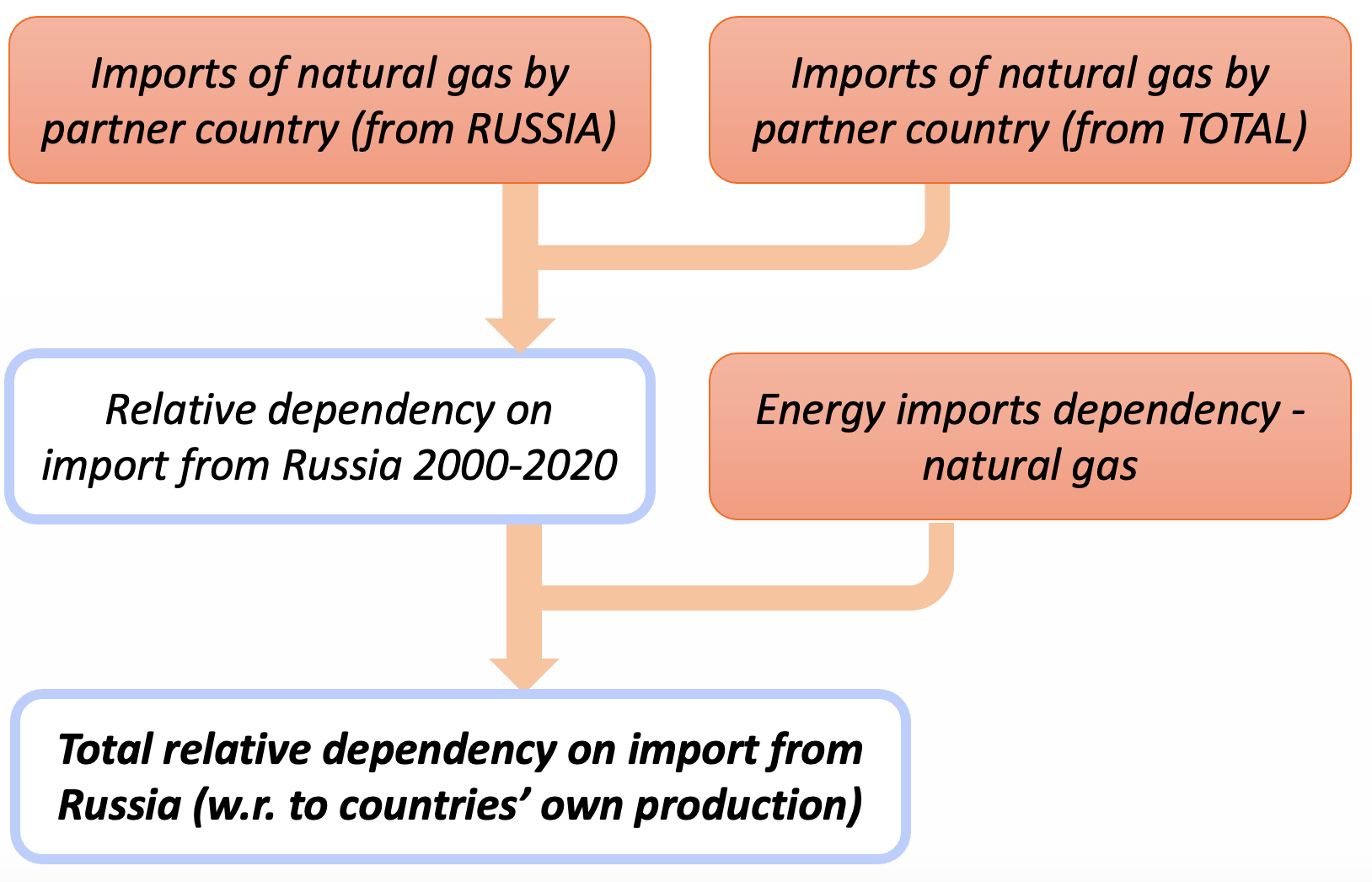


Figure 5: An example of the datasets aggregation methods for natural gas imports.

The same workflow of algebraic processing is applied to all two remaining datasets as well (oil and petroleum products, and solid fossil fuels). A tangible example of the natural gas imports from Russia to The Netherlands following the workflow in Figure 5 is demonstrated in Figure 6. The result of total relative imports from Russia in each respective dataset are used in the main visualizations.

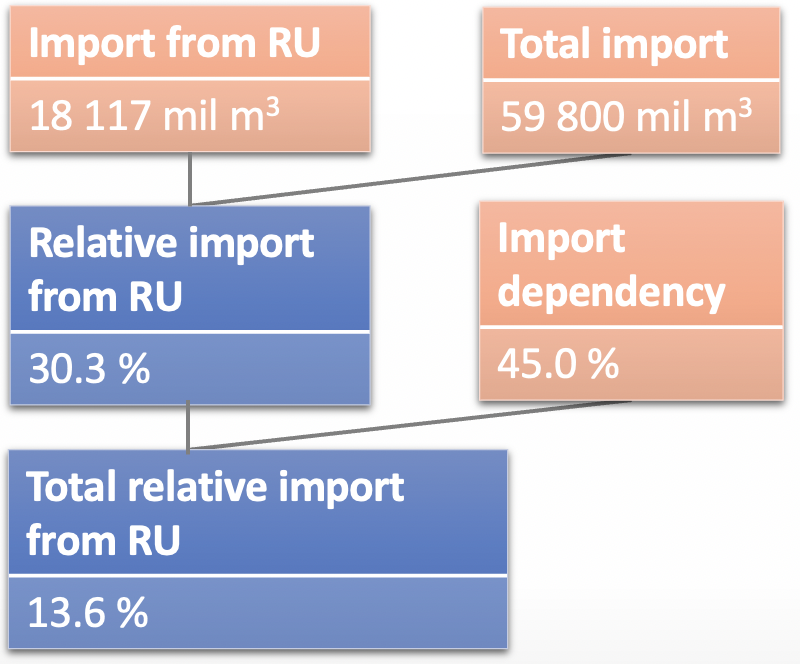


Figure 6: An example of the algebraic processing workflow on an example of The Netherlands in 2020 on Natural gas imports.

# Platform

Python

Python has a number of libraries with diverse functionalities for displaying data. All of these libraries have unique characteristics and can support various graph types. Some of the libraries that were used on the project are Pandas, Bokeh and Plotly. As mentioned before Python was also used in data preprocessing.

Panel

To be able to use multiple frameworks on our platform and to keep the interactivity of the dashboard, it was necessary to use Panel library to divide the screen into rows and columns allowing us to embed visualization elements from other platforms such as Bokeh (Choropleth map), Plotly (treemap) and Bokeh graph (timeline plot). Panel also supports other widgets like radio buttons, sliders, dropdown menus, images, and Strings, and it is well known for its flexibility. Thanks to its easy-to-implement stretching size functionalities, the panel dashboard was able to fit into the screen.

Bokeh

It is an open-source platform that facilitates the creation of plots but can handle custom specialized features, high interactivity, and widgets for the data. It creates an interactive web application by connecting front-end events to python code, therefore a Bokeh server was implemented, allowing high-level python models, then converting these objects to JSON to pass them to its client library, BokehJS.

Figure 7 discusses the structrure of the Bokeh server used in our project.

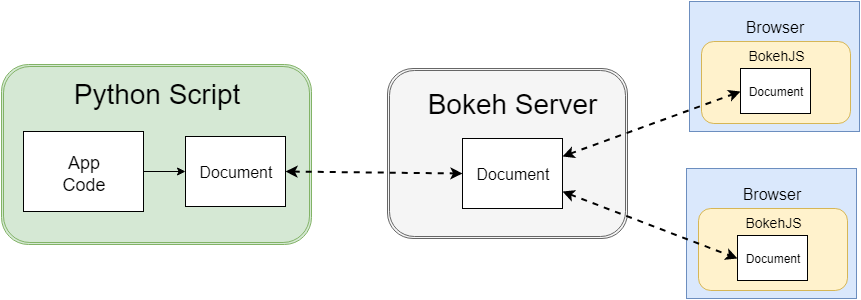


Figure 7: Bokeh workflow [7]

Plotly and Plotly express

Plotly is an open-source graphing library for Python however in the project the Plotly express was used. It offers many built functions and more high-level features than just Plotly, being more suitable for direct visualization purposes. In our case, it was used to create the treemap figure and has a straightforward implementation with its flexible input formats, styling control, and automatic hover and label parameters.

* To initialize the visualization it is necessary to follow the instructions stated inside the ***README.txt*** file provided with the code.

# Visualization

The platforms discussed in the previous chapter were used in the project and their implementations to create each of the visualizations are described in the next sub-sections.

## Components of the visualization

### Choropleth Map

Choropleth maps are thematic maps that allow visualizing variables to change across geographic areas. It uses different shades, patterns, and sequential color palettes for geographic areas based on statistical data. Data is categorized with light colors or shades to represent lower numerical data values and darker areas represent higher values (as the color bar indicates). The data was prepared to display the values as in percentages from the three main datasets after performing the preprocessing step (Natural Gas, Oil Petrol, and Solid Fuel datasets).

It was the best approach to represent the imports share of the European Union from Russia on a map as it provides users with a visual geographical reference that in other graphical formats would be rather hard to convey. The non-EU countries were colored with a gray color representing the NaN values and each dataset uses a different color palette (greens, blues, and oranges).

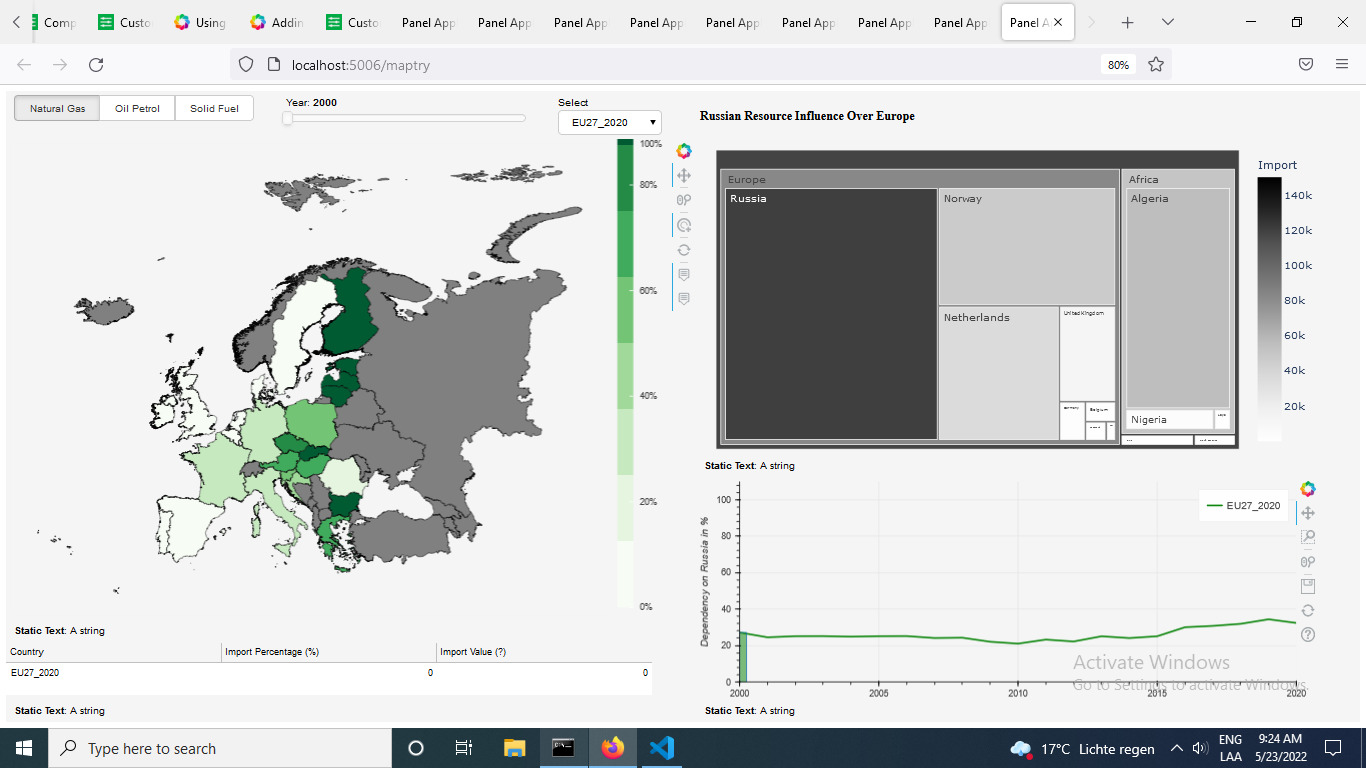


Figure 8: Choropleth map displating Natural gas EU 2020 dataset

### Treemap

After some feedback, the use of a pie chart to represent the ratio difference between the share of a specific country for energy resources was highly discouraged. So, it was decided to display categorical data in hierarchical visualization. The selected component was a treemap because it allows showing the composition of a whole for many components using the space efficiently. It also allows to display the absolute import divided it into smaller parts for every exporter country; in this sense, hierarchical data is nested into blocks (rectangles). Each category (continent) is given a main rectangular box that then is divided into smaller rectangles representing sub-branches. A leaf node's rectangle has an area proportional to a specified dimension of data (import from a specific exporter country to the selected country in the EU).

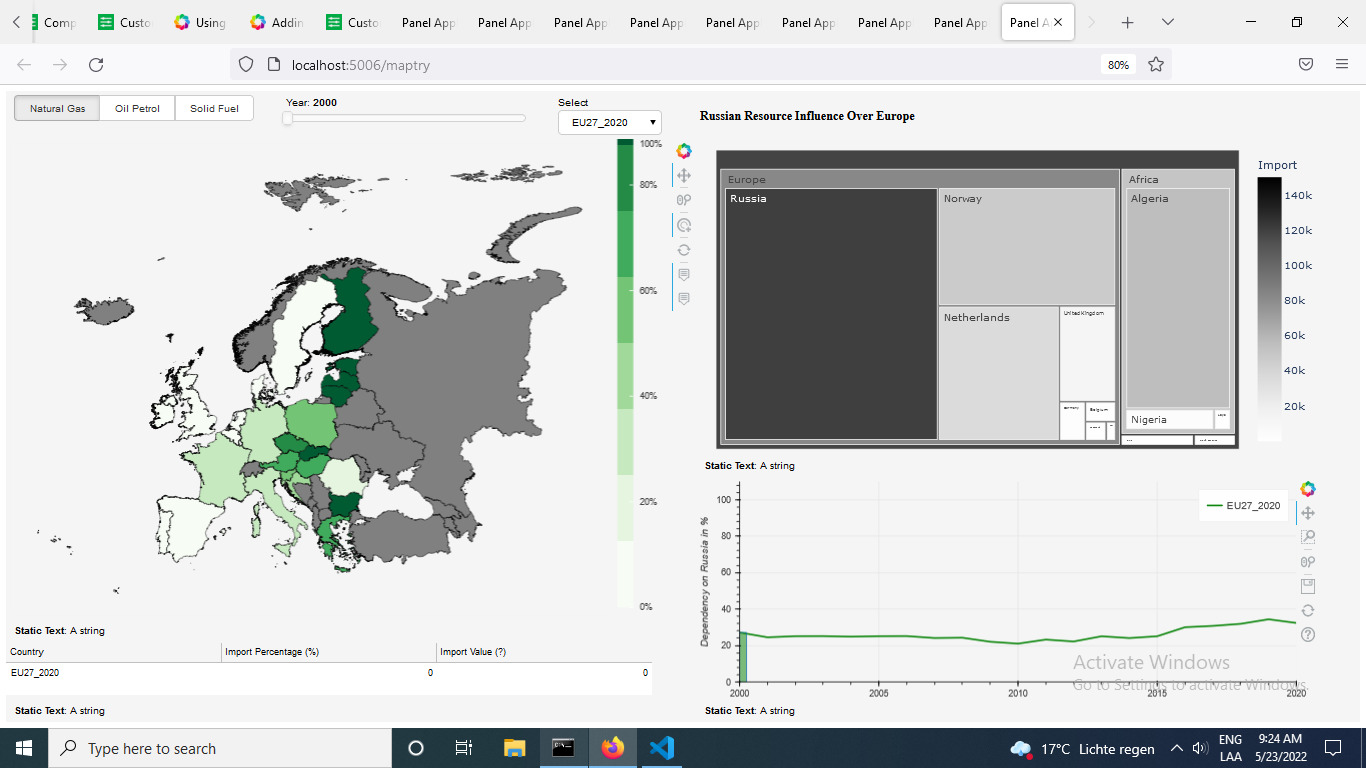


Figure 9: Treemap on Natural gas EU 2020 dataset

### Timeline graph

As a way to visualize values of imports and the trend over the years for specific countries, a timeline graph that has all the information of the selected dataset and selected country from 2000 until 2020 was implemented, this highlighted area under the curve matches the selected year. The default value is the European Union but once a country is clicked in the choropleth map, the information of that country is then displayed in the timeline graph.

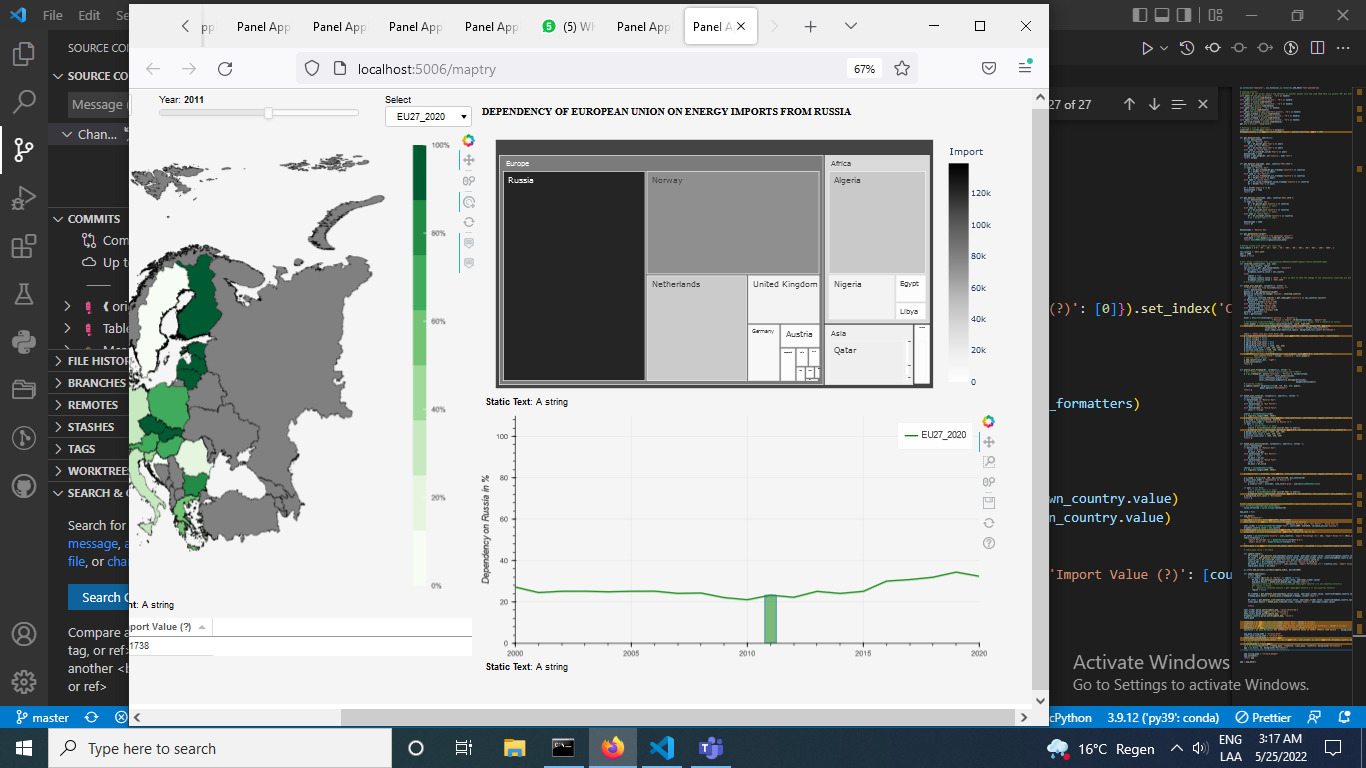


Figure 10: Timeline graph highlighting year 2011 for Natural gas import and trend over the years in the EU

### Summary table

A table displaying the relative and absolute values of energy import for the selected dataset, year and country is provided as a direct feedback for the users. The units change according to the data set and the value according to the selected country.

### Control elements

1. Radio buttons

Three radio buttons were used to give the user the ability to select between 3 different datasets (Natural Gas, Oil Petrol, and Solid Fuel).

1. Slider

A slider ranging from 2000 to 2020 was implemented to be able to select a specific year of the import data from which the user is interested. The year 2020 was chosen as the default because it represents the most recent available year of the dataset.

## Interactivity

### Selecting a Country: Clicking on the map

To select a country on the dashboard, the user must directly click on the country on the map, resulting in highlighting such a country and darkening all other parts of the map. To return to the standard visualization the user just needs to click on the sea.

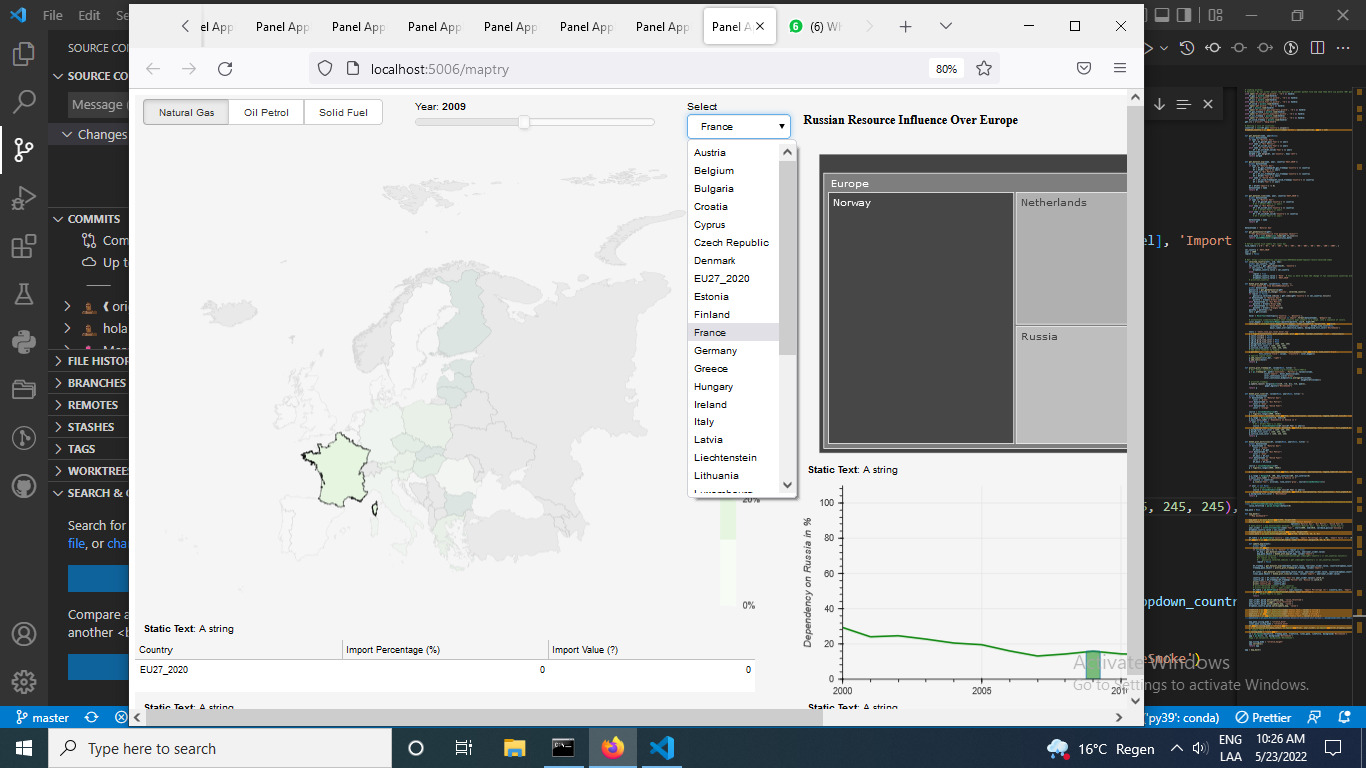


Figure 11: Country selection

### Selecting Dataset and Year: Via Radio buttons and Year slider -> effect on which graphs

All the components in the upper left side of the dashboard affect the graphs, the radio buttons change the dataset represented on the plots and change the color palette of the choropleth map and time plot (while keeping the color palette consistent on these two).

|  |
| --- |
|  |

Figure 12: Color palette change according to the dataset selection

To let users identify the main focus areas and avoid detracting them with too many colors on different graphs, we proposed only applying matching colors on the choropleth map and the timeline graph. In this sense, the treemap graph makes only use of the gray palette (independent of the selected dataset). The default visualization of the treemap displays all the regions in the world so to have a general idea about the share of imports worldwide, this data changes also according to the specific dataset and year.

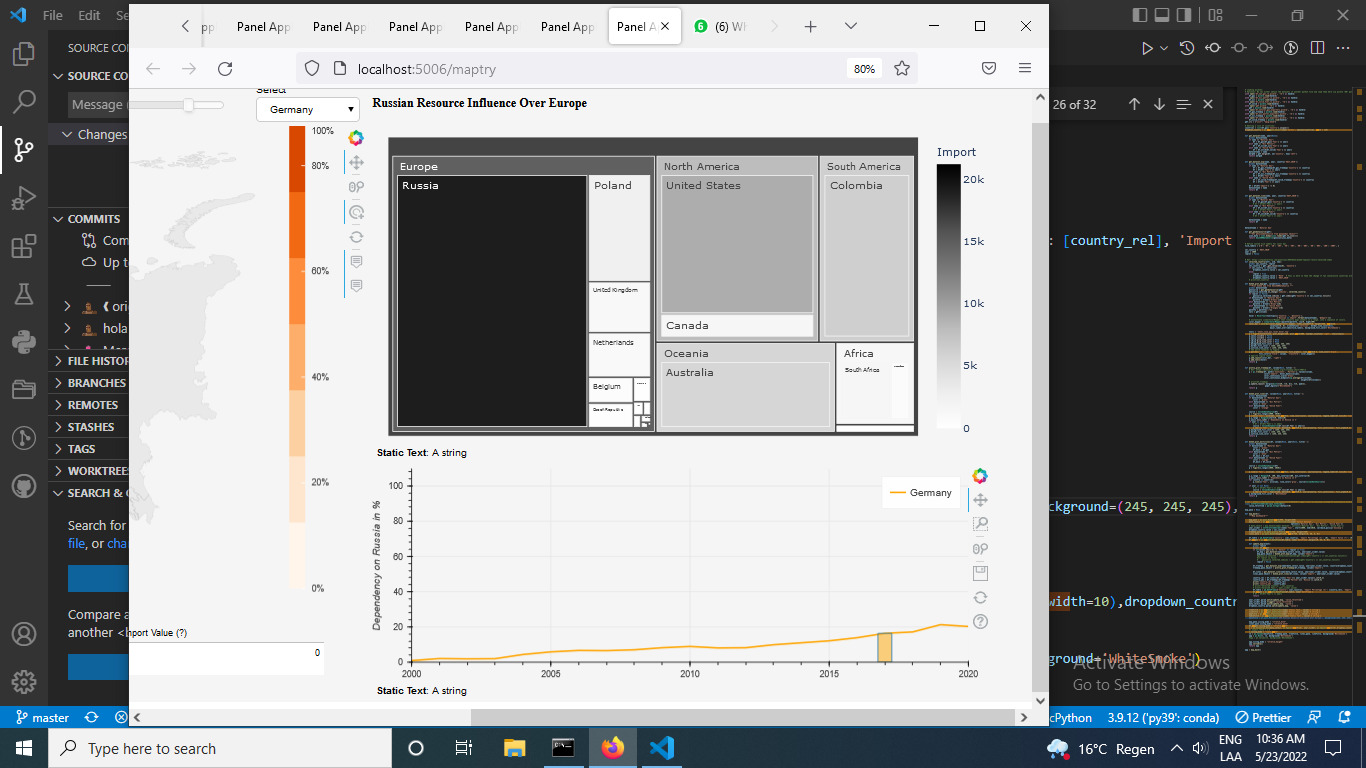


Figure 13:default treemap showing share of natural gas according per regions worldwide

However, if a country is selected, the treemap will display just the partner countries that contribute to the import.



Figure 14: Specific import partners from France for natural gas

The plotline palette of colors follows the same interactivity as the choropleth map, with a specific color according to the data set. There is a highlighted area under the curve that changes according to the selected year on the selector and just a single label displaying the selected country.

|  |
| --- |
|  |

Figure 15: Timeplot interactivity for different datasets and years on France

# Evaluation

After the interim presentation, the internal feedback pointed out the overuse of the same palettes of colors in the choropleth map, the bad use of visualization for hierarchical data in the pie chart, and excessive labels on the bar chart.

Following an internal discussion with the team, it was decided to change the palette of the choropleth map according to the dataset and change the saturation of the color according to the level value, as higher values should be more visually distinct than those representing less important information. Also, the option to highlight countries was added, the user can reduce the contrast of unimportant items by clicking on a country. The background is adjusted to increase the luminance contrast of critical areas, resulting in easier readability. As reviewed in the lectures, light colors will be best because there is more room in color space in the high-lightness region than in the low-lightness region.

The aim of the bar and pie chart was that data elements and data patterns could be quickly perceived but this solution was confusing for the user, so they were changed, opting for a better hierarchical visualization like a treemap. The timeline plot changed for a specific individual selection over time instead of multiple bar charts with messy labels. Regarding the treemap, the multicolored palette was changed to a greyscale palette to avoid the over-saturation of colors. The color saturation encodes numerical quantity, using a greater saturation to represent greater numerical quantities. The final solution offers an elegant grey palette of colors. Users can correlate easily the dark shades with higher values and sizes, without being misleading with multiple colors, and in the timeline plot, there is only one label, leaving out any possibility of confusion about what country trend about the year the user may be seeing.

## External feedback

For external feedback, the UEQ provided in class was used to evaluate our solution to understand the values of the mean and variance. Ten samples were taken from people ranging from around 25-50 years old indistinctly of the gender.

The following scale represents the feature performance:

|  |  |
| --- | --- |
| Value | Evaluation |
| > 0,8 | Positive |
| between -0.8and 0.8 | Neutral |
| < -0,8 | Negative |

The range of the scales is between -3 (horribly bad) and +3 (extremely good) and quite good values are already valued at +1.5. According to this overview, the worst features are easiness to learn, creativity, and speed of interaction. On the other hand, the strength is located in efficiency, organization, and understandability

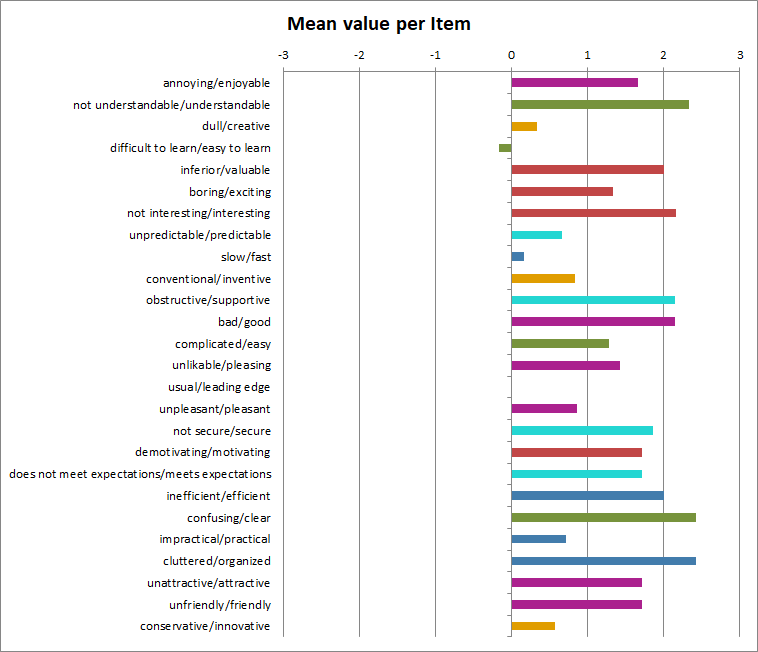


Figure 16: Per question evaluation.

From the previous aspects, all of them can be summarized into six paramters where the solution had a good performance in general (shown in Table 3).

Table 3: Final division per visualization category.

|  |  |  |
| --- | --- | --- |
| **Scale** | **Mean** | **Comparisson to benchmark** |
| **Attractiveness** | 1.64 | **Good** |
| **Perspicuity** | 1.64 | **Above Average** |
| **Efficiency** | 1.46 | **Above Average** |
| **Dependability** | 1.75 | **Excellent** |
| **Stimulation** | 2.00 | **Excellent** |
| **Novelty** | 0.54 | **Below Average** |

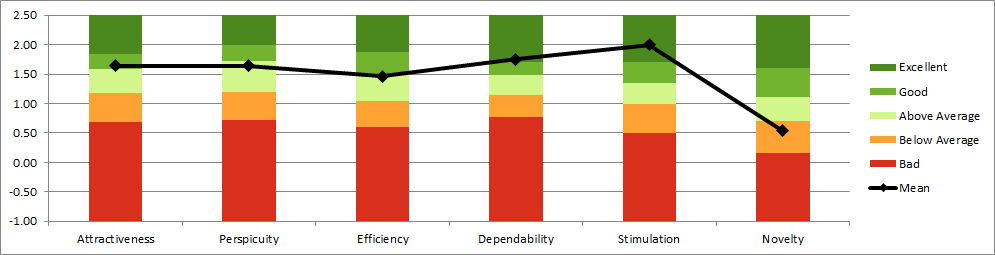


Figure 17: Final division per UEQ category

The goal of our implementation was to have a clear and organized visualization for very complex data and datasets, something that would interest our users about it. The stimulation and attractiveness features reflect that the interactivity, layout, and color design was the right one for our purposes of interesting the user on the data. Perspicuity and efficiency were double edge features as they share some of our worst and best features such as speed and clearness. Finally, we understand that novelty was low rated as since the beginning our intention was never to be innovative or offer a new way of visualizing the data.

# Conclusion

In terms of user experience, we consider that our solution accomplished its goal: a straightforward visualization of complex datasets on energy imports for the general public. Thanks to the panel various types of frameworks were used and the user could visualize the most dependent country's changes over time, and the biggest exporters, all on one page. The main focus was the interactivity and visualization of the information but in the process of the development, the theoric aspects of the class were not initially taken into consideration resulting in inefficient presentations of data, however, after interim presentation and the feedback received both from internal and external soruce, we implemented better and more complete solution. Speed of the interactivity remained still slighty underperformed (0.5s-1s to change the energy resource dataset) possibly because of the unsufficient database performance. UEQ questionnaire also demostrated that it can take some time for the user to understand the visualization. The questionnaire, however, was constructed the way that did not consider the visualization as a part of whole digital article, website, or other content, which would give more sense to the visualisation rather that a stand-alone solution. These remarks, alongside with the better novelty in the design could deliver increased user experience.

References

[1] European Parliament, *Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics and repealing Regulation (EC, Euratom) No 1101/2008 of the European Parliament and of the Council on the transmission of data subject to statistical confidentiality to the Statistical Office of the European Communities, Council Regulation (EC) No 322/97 on Community Statistics, and Council Decision 89/382/EEC, Euratom establishing a Committee on the Statistical Programmes of the Eur…*. Brussels: European Parliament and the Council, 2015, pp. 001–001 [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1435820363605&uri=CELEX:02009R0223-20150608. [Accessed: 25-May-2022]

[2] Eurostat, “Imports of oil and petroleum products by partner country,” *Energy statistics - quantities, annual data*, 2022. [Online]. Available: https://ec.europa.eu/eurostat/databrowser/view/NRG\_TI\_OIL\_\_custom\_2774870/default/table?lang=en. [Accessed: 23-May-2022]

[3] Eurostat, “Imports of solid fossil fuels by partner country,” *Energy statistics - quantities, annual data*, 2022. [Online]. Available: https://ec.europa.eu/eurostat/databrowser/view/NRG\_TI\_SFF\_\_custom\_2775364/default/table?lang=en. [Accessed: 23-May-2022]

[4] Eurostat, “Imports of natural gas by partner country,” *Energy statistics - quantities, annual data*, 2022. [Online]. Available: https://ec.europa.eu/eurostat/databrowser/view/NRG\_TI\_GAS\_\_custom\_2512612/default/table?lang=en. [Accessed: 23-May-2022]

[5] Eurostat, “Energy imports dependency,” *Energy statistics - quantities, annual data*, 2022. [Online]. Available: https://ec.europa.eu/eurostat/databrowser/view/NRG\_IND\_ID\_\_custom\_2512645/default/table?lang=en. [Accessed: 23-May-2022]

[6] European Environment Agency, “EEA coastline - Polygon,” *GIS data*, 2022. [Online]. Available: https://www.eea.europa.eu/data-and-maps/data/eea-coastline-for-analysis-2/gis-data/eea-coastline-polygon. [Accessed: 23-May-2022]

[7] Bokeh, “Running a Bokeh server — Bokeh 2.4.3 Documentation,” 2022. [Online]. Available: https://docs.bokeh.org/en/latest/docs/user\_guide/server.html. [Accessed: 29-May-2022]