NoMoreBarPlots

Process book

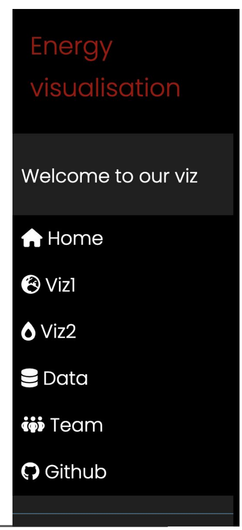
Being able to propose an environment that facilitates the visualization, and thus the understanding, of all our data was the main challenge we faced in the early stages of this project. From the beginning of our reflection, it appeared to us that the visualization of all our data could not be done with a single visualization. Indeed, the visualization of almost a hundred energy sources in more than a hundred countries and all this over several years did not seem to be feasible in a single visualization format. As a result, we decided to offer two visualizations, each offering a different perspective on our data. Thus, the first visualization offers a way to visualize the main energy flows in each country and their evolution over time. The second visualization changes the point of view by proposing an animation allowing to see the evolution of a flow/quantity of an energy for all countries, as well as to compare and relate them.

What are the energy exchanges between countries ? Which types of energy are most present, in which countries and the diversification over time?

In what areas are the different types of energy used? How does this evolve over time and how does it differ according to the wealth of the countries for example? Are the most developed countries more advanced in the production of renewable energy?

Website structure

The two visualizations that we have created have been incorporated into a website that we have also created. The structure of the website is shown in the adjacent figure.



The index presents an introduction to the project, our motivations as well as the problems posed. This page allows to go to the different visualizations by clicking on one of the two videos.
Visualization 1 is first composed of a map that allows the user to choose the country from which he wants to see the energy flows. This visualization will be explained in detail in the following part.
Visualization 2 will be completely defined in a dedicated section.
The data page gathers the information concerning the origin of the data and the choice of the datasets used.
Team page is a description of the group members with a description for each of them.
Finally you can see in the sidebar the link to the corresponding GitHub page.

NoMoreBarPlots

One of the many original elements of our website is the introduction, it appears slowly and consecutively as if to guide the user in the discovery of our website. It is however possible to skip this progressive introduction thanks to a button.
Concerning the other specificities of our website and the difficulties encountered, the map was one of the major points.
Indeed, it was created with Leaflet. However, since we needed the name of the country to plot the visualization of the selected country, the examples provided in the course and on Leaflet were not enough.
In order to avoid using an API to find the country name from the coordinates, we used a GEOJson file. This file contains tiles of all countries in the world (at least almost all). The use of this file allowed to zoom on the tile on which the user clicked and to retrieve the name of the associated country.
Another difficulty encountered was to display the visualization 1 but to be able to go back on the map to choose another country to explore. Many transitions have been used here.
Concerning the sidebar, we had initially chosen to use the Leaflet sidebar but as it was not compatible with the other non Leaflet pages of the website. This is why we implemented this one.
Finally the last great difficulty was to transfer the visualizations coded on observablehq in .js files and to find the equivalent functions to those which exist in observablehq but not in javascript. This took us many hours to figure out the problems and to come out with effective solutions.

Visualization 1

The first visualization is an interactive Sankey flow visualization by country of data on production, final consumption, losses, imports and exports of the energies present in the country.

Motivation

The idea is to provide the user with three key pieces of information about a country's energy network.
First, we want to show how independent each country is in terms of energy. Indeed, energy independence and the pressure that one country can exert on another because it depends on it for energy is very powerful. Unfortunately our dataset did not have the information of which countries are involved in these exchanges (who gives to whom) but the fact of being able to already see the quantity of imports and exports of each country gives an idea of this concept.
Second, we want to show how much energy a country consumes and throws away.
And finally, we want to show which energies are used by each country. The idea is to find out whether the energy used in each country comes from green production or even fossil fuels.
Thanks to data over several years, it is also possible to see whether countries have made efforts in terms of reducing their consumption and transitioning to green energies.

Data pre-processing

A lot of data preprocessing was required to achieve this visualization.

First, we sorted through the 75 datasets of the different energies. Some datasets do not contain information about production, final consumption, losses, imports and exports. These were immediately discarded. 58 datasets still remained. These are sometimes very specific and not so meaningful to the general public. This number is also too high for the visualization to remain visible so we decided to keep only the main subcategories of each type of energy defined in the guidelines given by the dataset, as an example in the table below the retained category is "Hard Coal".

Table A.1: Correspondence between SIEC codes and the Questionnaire product codes

SIEC Headings			Correspondences
Section / Division / Group	Class		Codes for the UN Energy Questionnaire
0		**Coal**	
01		**Hard coal**	CL
011	0110	Anthracite	AT
012		Bituminous coal	
	0121	Coking coal	CC
	0129	Other bituminous coal	OB

These main categories are sometimes available as is in the dataset, other times they are divided into subcategories. In the second case, we summed the small categories to get the total values of the large one. We also summed the variable that indicated the estimate, initially equal to NaN if the value is not estimated and 1 if it is. So we now have an estimation scale that is no longer binary. Also, the energies were not all expressed in the same unit, so we converted everything into kWh to be able to display the energy flows on the same graph and be able to compare them. It was then necessary to rearrange the dataset by country rather than by energy, keeping only the information on production, final consumption, losses, imports and exports. Finally, it was necessary to rearrange the data so that it was understandable as input to a sankey chart. The standard format for a sankey chart input is a dataset defining the links of the sankey (ie for each link define source, target, value). A "Total amount of energy" node was artificially created to be able to visually separate inputs from energy outputs. To this is added in our case, the year and the estimate of the value.

Another unexpected preprocessing was to adjust the country names of the map used compared to those in our dataset. In particular, there are countries whose name was shortened in the map dataset (e.g. country name without "Republic of") or specific regions that did not exist (e.g. Hong Kong, Macao, French Polynesia).

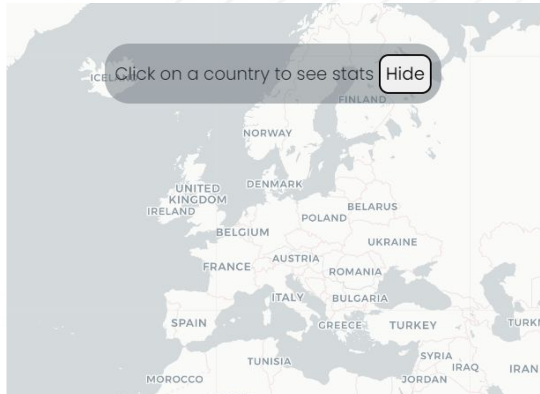
The entire preprocessing was done using Jupyter Notebook (python programming language).

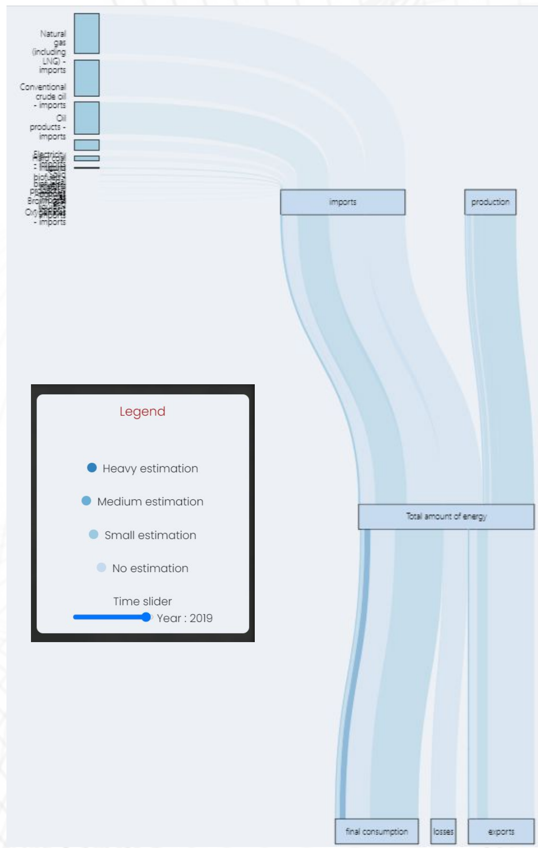
Visualization

The main idea of visualization has always been to first have a map of the world. The user must click on a country of which he wants to see the visualization, the instruction is given to him at the top of the map. The map zooms to the country the user has chosen and fades into the background. The sankey plot appears accompanied by the name of the country on which we clicked, a legend and a time slider. The legend defines the different possible colors of the flows according to the rate of estimation of the values of these, it is a panel of blue which goes from the lightest if there is no estimate to the darkest which represents a great estimation rate. The time slider allows you to browse all the years whose data is in the dataset of the selected country.

The Sankey chart is quite particular. It is a sankey chart which at first glance only appears vertical but thanks to a mouse over on the nodes of production, final consumption, losses, imports and exports the details of the quantities for each energy appear: the rates of production, final consumption and losses appear vertically but import and export rates appear on the sides (left and right respectively). The vertical structure of the plot was chosen to have an intuition of the flows that start from imports and production (top, energy inputs) and go to final consumption, losses and exports (bottom, energy output). We wanted to have the exports and imports entering and leaving the sides, to convey the fact of displacement/exchange of these flows in particular. The fact that the plot first presents only general information on production, final consumption, losses, imports and exports makes it possible to focus on these quantities, so that it is more readable at first glance. Then the user can go into the details of the specific energies.

Thanks to the reset button the user can return to the world map to be able to choose another country.





Technical challenges

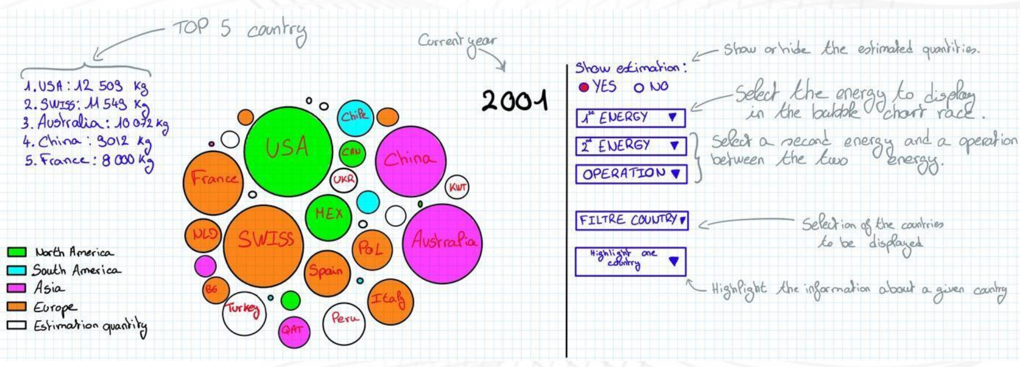
Given the choice to have a "cross" and non-linear sankey chart, it was difficult to find examples on which to base ourselves at the beginning. We really wanted to draw the lines of the energies of the imports and exports that arrive from the sides to underline the entry and exit of the country of these flows. Normally the sankey charts are horizontal, sometimes vertical but in any case all the lines are on the same axis. The only example similar to what we wanted to do was found on observablehq.
Also, we wanted the flow color to be proportional to the flow value estimation rate. It was not easy to introduce this constraint because the data usually accepted by a sankey chart function are the sources, the targets and the values of the links.
In the same way we had to filter the data inputs according to the year selected by the slider.
The slider also caused us a lot of trouble because not all countries have the same interval of years available, so we had to adapt the minimum and maximum years of the slider according to the country selected. A function did this but these values were not loaded in time by the algorithm so the first click on a country did not work, we had to add a delay to successfully solve this problem.
In addition, the mouse over technique allowed us to make the graph more readable and to show the detail of the energies in a second step. The fact that the plot first presents only general information on production, final consumption, losses, imports and exports makes it possible to focus on these quantities, so that it is more readable at first glance. Then the user can go into the details of the specific energies.
The legend defines the different possible colors of the flows according to the rate of estimation of the values of these, it is a panel of blue which goes from the lightest if there is no estimate to the darkest which represents a great estimation rate.
When many small energy flows are present, the names of the energies are not always visible. It is very difficult to adjust this, we had thought of zooming in on the energy labels to make it more readable but we did not have time to implement this solution.
A question that always challenges us is: why for some countries the sum of input flows (imports, production) is not equal to the sum of output flows (final consumption, losses, exports). We tried to understand by reading the documentation of our dataset but we did not find the answer.
Compared to the plans announced in Milestone 2, we have achieved all the objectives we set ourselves apart from making the flow dark when the mouse is on it. We decided not to implement this anymore since we already have another mouseover event which is more essential.

Visualization 2

As explained earlier, our idea of the second visualization was to provide a global view of all the countries by focusing on one energy flow at a time. Our first idea was to use an interactive 3D map to visualize the flow of an energy between the different countries, but this quickly seemed to be a dead end as we did not have the direction/provenance of the different energy flows between the countries, so we would have had to cross-reference our dataset with another dataset to get this information. Following this, we slightly reoriented our idea by no longer trying to visualize the flows of an energy between countries but rather by focusing on the visualization of the quantities linked to these flows. After this fine-tuning, our ideas were directed towards a visualization in the spirit of a bar chart race. We found that orienting the visualization around a race between the quantities of each country's flows was interesting as it would allow these quantities to be compared in a dynamic and entertaining way. However, this visualization format was a real challenge to our group identity. We, the NoMoreBarPlots group, could not afford to include bar plots in these visualizations. Moreover, the bar chart race being quite classical, reproducing a visualization that could be obtained by a copy-paste from D3.js was not motivating us at all. Inspired by this idea of a bar chart race, we came up with our final idea, which was to combine the bar chart race with a bubble chart to create a bubble chart race. This visualization format was even a better option for us than the bar chart race, because as we wanted to visualize a certain amount for all countries, the bubble was much more representative of the country than a bar would have been. Moreover, as for some energies we had data on more than a hundred countries, we would have had to limit ourselves to displaying, for example, a Top 20 of the countries to remain readable on one page, which would have made us "lose" a certain amount of information, whereas the bubble chart format allowed us to make better use of the visualization space.

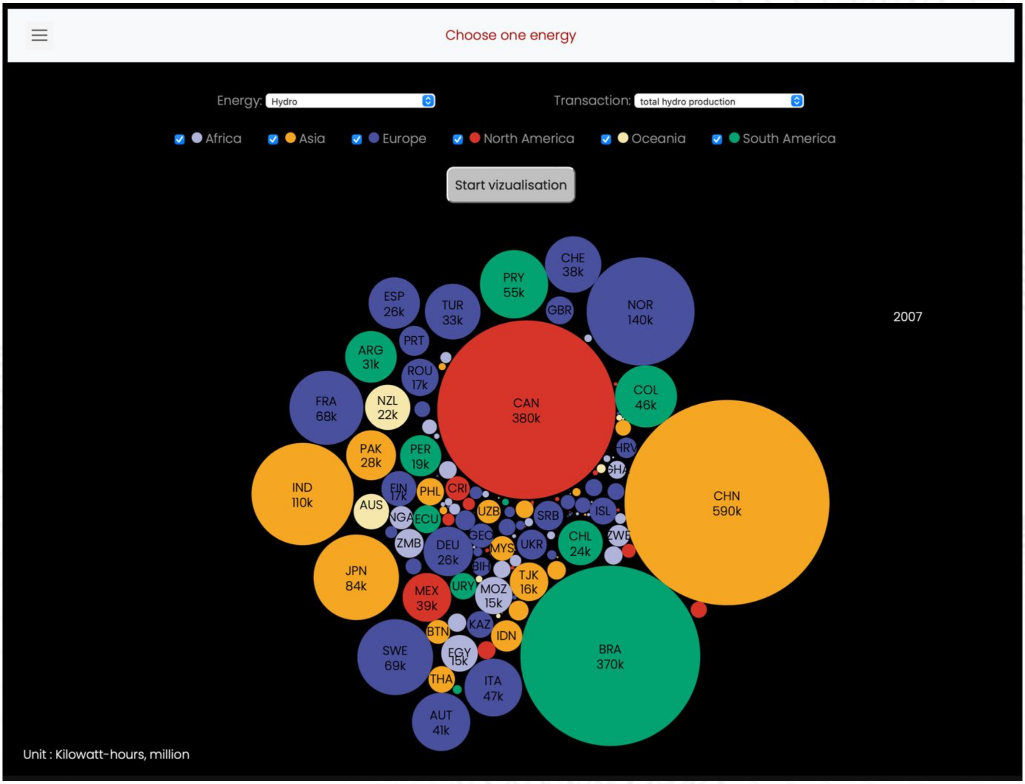
Sketches/plans

When we started to implement the second visualization, we aimed to get a layout like the one in the sketch below. However, we soon realized that this layout greatly limited the space that our visualization could occupy and made it less readable and therefore less relevant. The main problems were the options on the right and the displays on the left taking up too much space in the width of the window.



NoMoreBarPlots

To address this problem, we modified the layout of the visualization as shown in the visualization below. Rather than using the limited width of the window to display the option bar and the legend for the continents, we took advantage of the fact that we could scroll the window vertically to display the option bar and then the visualization, which could then occupy all the available space in the window. To clean up the design, we combined the color code for the continents with the option to select the continents you wish to display in the visualization. In addition, we have removed the option to choose to display the countries for which quantities were estimated, to force this option to be displayed. We thought that adding a button for this was not necessary, but that displaying the estimated data was still important. However, these estimated data are rarely observable as they appear mainly for countries with very low energy quantities/flows, so the white bubbles are hardly discernible. Moreover, instead of displaying a Top 5 of countries, we have chosen to display the quantities, with a reduced format directly, in the bubbles when the size of the bubbles allows it and we have also decided to limit ourselves to the display of an abbreviation for the name of the countries. Nevertheless, we have added the option to display the name and the complete quantity for each country by moving the mouse cursor over the desired bubble.



Finally, if we had the time, we wanted to implement the possibility of carrying out operations between different energy flows (to be able to compare, for example, the imports and exports of an energy ...). However, we did not have enough time to finalize this option. Implementing this option would not have been long, we would have had the time to do it, but what was missing was the time to determine a consistent and coherent way to transform a negative quantity (which could appear if we consider the difference between two energy flows) into a radius which must necessarily be positive.

Peer assessment

Maxime Dulon
Responsible for the website (including Leaflet map). He made the structure and transferred the visualizations from observablehq on it. He also did data processing for visualization 1, conversion of data into kWh and rearrangement of dataset into countries was done by him.

Francesca Paola Nicoletti
Responsible for first visualization. Data cleaning and data processing and rearranging for the first visualization were made by her as well as the entire first visualization code.

Kieran Vaudaux
Responsible for second visualization. Data cleaning and data processing for the second visualization were made by him as well as the entire second visualization code.