# COMP30750 Visual Exploration Tool Design Document

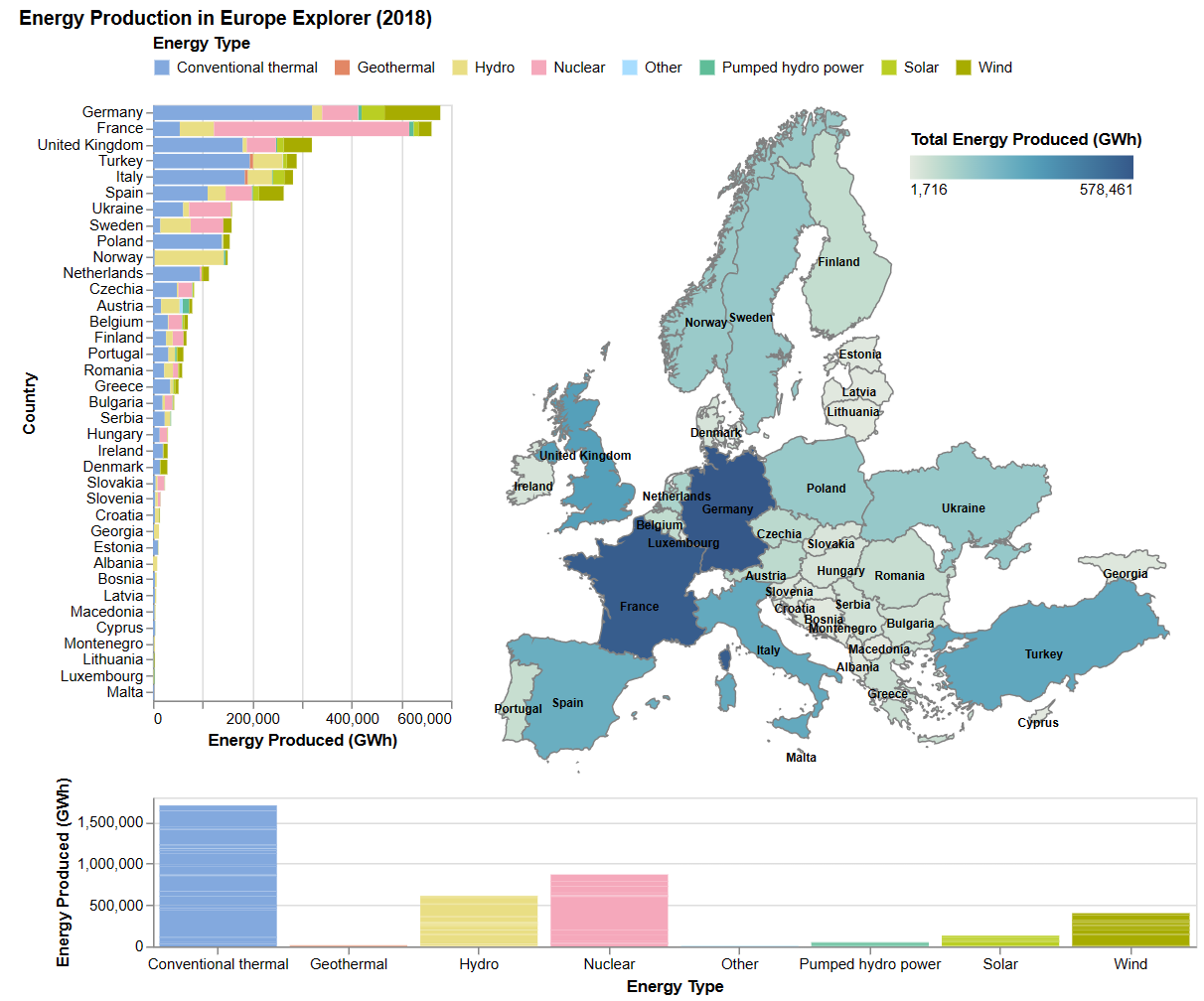
Student Number: 21334466

Github repository of used datasets: [ataha03/infovis (github.com)](https://github.com/ataha03/infovis)

## Title

Energy Production in Europe Explorer (2018)

## Screenshots



## Dataset overview

The dataset I picked is about [energy production in Europe](https://github.com/rfordatascience/tidytuesday/blob/master/data/2020/2020-08-04/energy_types.csv). There are 37 countries and 8 energy types in the dataset. Each row specifies a country, its country code, the type of energy shown, and the energy produced by the country in 2016, 2017, and 2018.

I decided to only use the information in the 2018 column, as it is the most recent.

I used python to derive two datasets from the original data. I provided a file called “Dataset Pre-processing with Python” detailing the steps I took.

The first dataset, “energy\_types.csv”, fills in one missing value in the dataset.

The second dataset, “energy\_types\_2018.csv”, is a transposed version of the first dataset, using only 2018 values. The energy types become separate columns, and the values for these columns are taken from the 2018 column.

The bar charts use “energy\_types.csv”, and the map uses “energy\_types\_2018.csv”.

Having a second version of the dataset was necessary so that the user could hover over a country, and see the production of each energy type by the country, as well as the total energy production of that country.

## Design Considerations

### Overall Goal

My overall goal with this tool was to explore energy production across the European continent, and how sources of energy from each country differ from each other based on geography.

### Vertical Bar Chart

This is a stacked bar chart that displays the amount of energy production in each country by energy type. It is sorted by total energy production per country. This makes it easy to compare countries with the highest energy production. Alternatively, the countries could have been sorted alphabetically for easier look-up. Countries can be looked up easily by selecting them on the map, so I chose to sort by energy production in order to show more information about the countries, and allow for better comparison.

Marks and channels:

* The marks are lines/bars
* The colour channel encodes energy type
  + Colour is a very effective encoding for nominal attributes like energy type
* The length/x-axis channel encodes energy production of the country
  + Length was chosen as an encoding because it is among the most accurate encodings. It is easy to compare different countries/energy types, and roughly estimate ratios of countries (e.g. Turkey produces approximately half the amount of energy that France produces)
  + Bar charts use redundancy to encode position on the x-axis and the length of the line, which is very effective
* The y-axis channel encodes the name of the country
  + I chose to place the country names on the y-axis because the large amount of countries are easier to read vertically than horizontally

I will discuss the colour palette choice in the horizontal bar chart section.

I initially wanted to create a parallel coordinates chart to compare the energy production of each country by energy type. I thought it would be a suitable chart since the dataset is high dimensional (“energy\_types\_2018.csv” has 10 columns). The parallel lines would represent each energy type, and the countries would be plotted as lines across the axes. I decided against this as there are too many countries to plot and compare clearly.

### Horizontal Bar Chart

This chart is very similar to the previous bar chart. However, when nothing is selected, the bars are stacked and they encode the total energy produced for a specific energy type in all of Europe. When a country is selected, bar length encodes the energy produced in that country. There are 8 bars, representing the 8 energy types. Users can hover over a bar for a selected country to see exactly how much energy (of a specific energy type) was produced by that country.

Marks and channels:

* The marks are lines/bars
* The colour channel and the x-axis encode energy type
  + Colour encodes energy type to match vertical bar chart
  + The redundant encoding of energy type links the two charts as they show related information. This redundancy emphasises the data and makes it clear where each energy type is shown.
  + Both bar charts share the same legend at the top
* The length/y-axis channel encode energy production of a specific energy type
  + This could be for all countries, or for a selected country

I chose a colour-blind friendly palette for the bar charts. After doing some research, I found out that red should not be used, especially with green, as red-green colorblindness is the most common. I decided on Paul Tol’s light colour scheme. Tol is a Dutch instrument scientist who created many colour palettes for his colour blind colleagues.

These colours are:

* Distinct for all people, including colour-blind readers,
* Distinct from black and white,
* Distinct on screen and paper,
* Match well together

*Tol, P. (2021) “Colour schemes and templates”’. Available at URL:* [*Paul Tol's Notes (sron.nl)*](https://personal.sron.nl/~pault/)

I am aware that encoding 8 different categories with colour is not ideal, so care was taken to ensure the colours aren’t too similar. They should be distinct enough to recognise the energy types. The redundant encodings of colour and position on the x-axis should also help if the colours are indiscernible.

I also considered using a radar chart where the axes are the different energy types, and data would be shown for all countries/a selected country. I decided against using this chart because it is difficult to compare non-adjacent attributes, and because the shape is entirely dependent on the ordering of the attributes (much like parallel coordinates.) Data is much easier to compare on a bar chart as all the data share the same axis.

### Choropleth Map

The choropleth map shows the locations of the countries in Europe. It is coloured by the total amount of energy produced per country. This makes it easy to see which countries produce the most and least energy overall. It is also easier to spot patterns in energy production in relation to geographical location. By hovering over a specific country, a user can find more information about energy production per energy type, as well as the total energy produced in gigawatts. This was done using tooltips. I chose a choropleth map because it is easy to show regional variations, which means it is great for comparing countries.

I initially wanted the user to be able to select an energy type from a dropdown menu to change the choropleth map colour by the amount of energy each country produces in that specific energy type. However, after consulting with demonstrators, it was clear that it would be too complex to properly graph.

I used two external datasets to create this map. The first is a [topojson of europe](https://raw.githubusercontent.com/ataha03/infovis/main/europe.topojson), which I used to map out the polygons of the countries in Europe, and the second is [a list of countries and their latitude and longitude coordinates](https://raw.githubusercontent.com/ataha03/infovis/main/Countries-Eurasia.csv), which I used to display the names of each country.

A few changes had to be made to these datasets to normalise them (e.g. changing “Bosnia and Herzegovina” and “Bosnia & Herzegovina” to “Bosnia” for standardisation purposes, or changing “U.K.” to “UK”.)

The original datasets:

[map-of-europe/TopoJSON/europe.topojson at master · leakyMirror/map-of-europe (github.com)](https://github.com/leakyMirror/map-of-europe/blob/master/TopoJSON/europe.topojson)

[raw.githubusercontent.com/ajturner/acetate/master/places/Countries-Eurasia.csv](https://raw.githubusercontent.com/ajturner/acetate/master/places/Countries-Eurasia.csv)

Marks and channels:

* The marks are the areas of the countries
* The colour channel encodes total energy produced per country
  + I used a monochromatic scale for the map legend, which is colour blind friendly
* Horizontal and vertical positions encode longitude and latitude of countries
  + Shows their geographical location of the countries

Note: Some countries are missing from the dataset - namely Iceland, Belarus, Russia, Switzerland, and Moldova. I chose to exclude them from the map to avoid confusion in the energy production. If I included them, some users may think that they don’t produce any energy.

#### Projection

I chose to project this map using Conic Conformal. I believe this was the best option for a projection of Europe.

The conic conformal projection preserves angles (since it is conformal) and shapes in relatively small regions. As Europe is a small continent, this projection maintains the shapes of the countries well. Conic Conformal is also great for regions that are long in the east-west dimension, like the United States or Europe, because distortion is constant along common parallels (east-west lines on a map). The disadvantages of this projection is that it does not display the range of the entire world, and is generally unsuitable for large regions. It is also unsuitable for long north-south regions.

Some alternative projections I considered for this map are stereographic and transverse mercator.

The stereographic projection is very similar to the conformal conic projection, in that it is ideal for small regions, and that it preserves angles and shapes. However, it produces extreme distortion near the equator. It also distorts the area and distance of regions away from the centre of the map, leading to peripheral countries appearing larger. Initially I thought that the transverse mercator projection would be well suited to this visualisation. However, after doing some research, I found out that it distorts east-west spanning regions like Europe. Therefore, I found that these projections were not the best option.

### Interaction Considerations

The main interaction in this exploration tool is selection and linking.

Users can select a country from the vertical bar chart, or directly from the map. This will change all the charts. The map and the vertical bar chart will show which country was selected by making other countries transparent. The horizontal bar chart will transform completely as it is filtered for a specific country, going from a stacked bar chart on the energy production of every country, to an unstacked bar chart showing the energy production of the selected country.

Users can double click anywhere to undo selection. Users may also hover over the map or the horizontal bar chart to see more information.

I believe that these interactive tools allow the user to explore energy production levels in Europe efficiently. (E.g. how much nuclear energy does France produce, relative to other energy types and other countries?)

### Other Alternatives

Here are other charts I considered for this project:

#### Slope charts

Since the dataset includes data for multiple years (2016, 2017, 2018), a slope chart would have been a suitable option for this project.

This could be used to visualise how the energy production of countries vary year by year. I didn’t go with this option because the great number of countries would result in overplotting. I also didn’t know how I could create other charts relating to the slope chart, and plotting the data over multiple years would be quite complex.

#### Cartograms

I considered using Dorling cartograms and tile grid maps instead of the choropleth map.

These cartograms have an advantage over choropleths. Choropleths can distort attributes as areas are not uniformly populated, with big size differences among areas, so larger areas may be given more importance as they take up more space visually. However, since we are not dealing with populations in this project, I thought that this disadvantage was not too costly.

Additionally, using size (for Dorling cartograms) and colour to represent the energy production of each country would have been effective, but it was at the cost of matching real life geography.

#### No Map

Another possibility is not using a map at all. Reading quantitative data from colour is difficult, it is much easier to read on a bar chart using length. It is also easier to see rankings and approximate values and ratios with a bar chart than a map.

However, since bar charts are already given beside the map, no readability is lost by including a map. On the contrary, a map allows for analysing geographic trends and spotting high and low areas in a region.