# CSC337 Coursework 1

Matt Hall (961500)

## Part 1, design 1

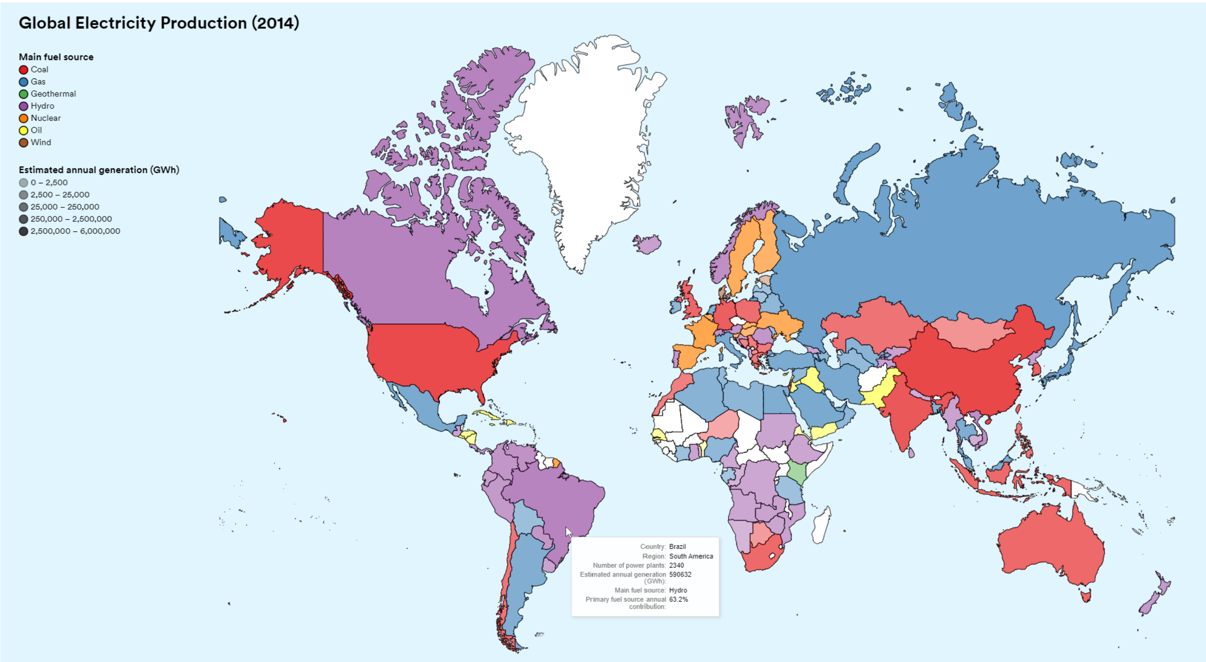



Figure 1: Global electricity production (Design 1)

**Description**

**Visual Design Type:** Choropleth map

**Name of Tool:** Altair

**Country:** Worldwide

**Year:** 2014

**Visual Mappings:**

- **Hue**: The hue of a country represents its main fuel type; the fuel type it uses to generate the most electricity.
- **Saturation**: The saturation of a country represents the amount of electricity it produces across all fuel types. It is banded into five discrete values along an exponential scale.
- **Tooltip**: Tooltips are used to offer detail-on-demand to the user, showing precise information about a country's electricity production.

**Unique Observation:** The visualisation allows the user to quickly and easily grasp the global distribution of fuel types whilst also providing a useful summary of the state of each country's power generation. For example, it is evident that coal, gas, and hydroelectric power are the major fuels across the world and that China and the United States generate the most electricity by a wide margin.

Countries' colour can be compared at a glance to get an idea of how big and varied the energy market is in a specific region. We can hover over Brasil for example, to see that hydroelectric power provides around two-thirds of the country's power - and by looking at the saturation of neighbouring countries we can see that Brasil is also the largest producer in the region of South America.

The contrasting hues also make it easy to spot countries with less common fuel types, for example we can see that Kenya is the only country in the world to favour geothermal power, and that Estonia is almost completely wind powered.

**Data Preparation:** In order to generate the map, the provided GPPD data set was combined with an open source GeoJSON map of the world so that each country's shape could be coloured. The power generation figures for each country were gathered by combining the 2014 US EPA Global Estimates (`'estimated_generation_gwh'`) with the 2014 national estimates (`'global_generation_2014'`).

Data was also aggregated through several `groupby` commands to form country-level generation and capacity figures, total number of power plants per country, and for finding out which fuel type produces the most power per country.

## Part 1, design 2

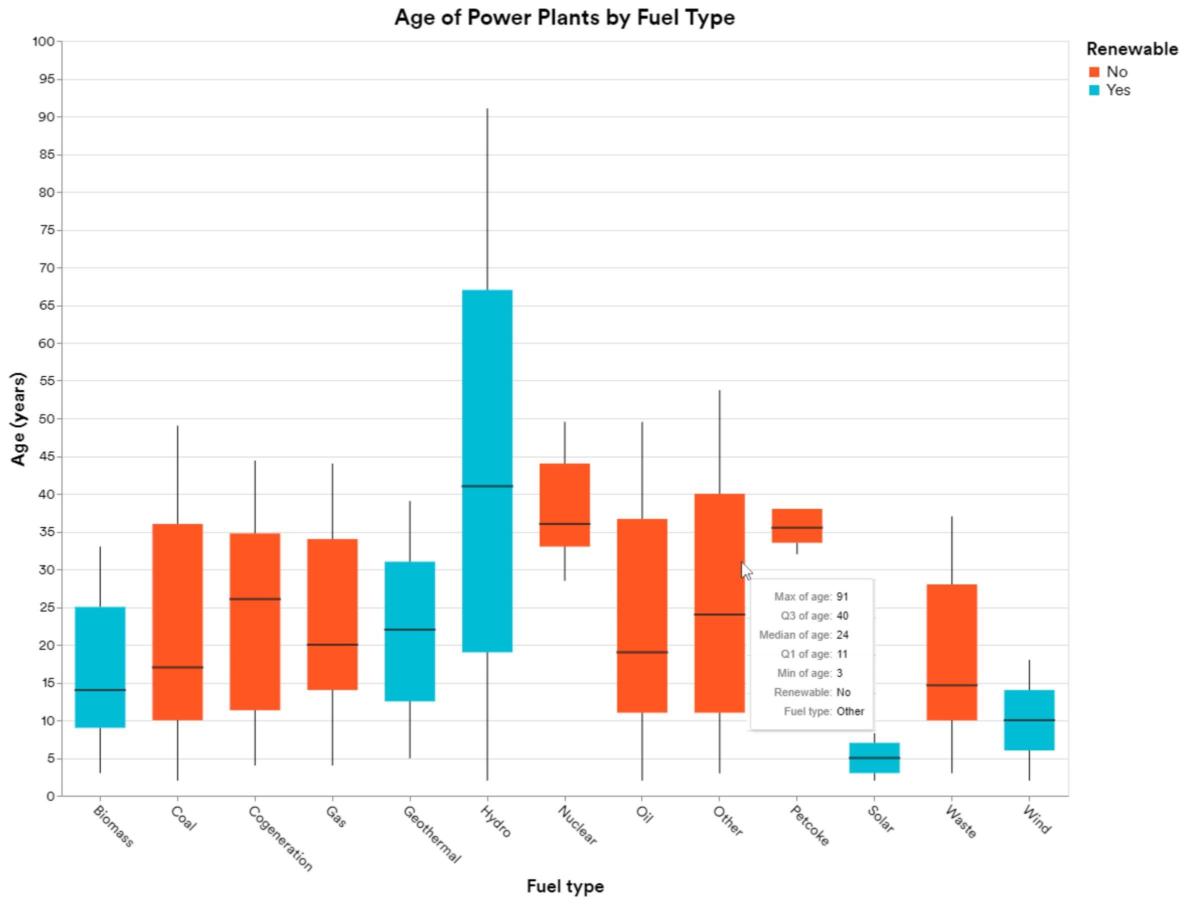



Figure 2: Age of power plants by fuel type (Design 2)

**Description**

**Visual Design Type:** Box plot

**Name of Tool:** Altair

**Country:** Worldwide

**Year:** N/A

**Visual Mappings:**
- **Boxes**: The coloured boxes illstrate the variability, or spread (interquartile range) of the ages of each fuel type.
- **Whiskers**: A relative minimum and maximum power plant age for each fuel type is illustrated by the extent of the whiskers (vertical black lines).
- **Median**: The horizontal black line within each box indicates its median, or roughly the average age of power plants of each fuel type.
- **Hue**: The hue of a box is either blue or orange, indicating if the fuel type is renewable or non-renewable respectively.
- **Tooltip**: Tooltips are used to offer detail-on-demand to the user, showing precise information about the distribution of the ages.

**Unique Observation:** Power plants have lifespans, and it is crucial to ensure that plants are as modern and efficient as they can be. This chart shows the general range of ages of power plants worldwide such that opportunities to enrich or update ageing plants can be identified, and also allows the user to keep an eye on emerging fuel types. We can easily identify that hydroelectric power plants have been the longest-running source of energy, and plants have been built regularly for most of hydro power's history.

The visualisation makes it plain to see the usage and evolution of fuel generation over the years, as the box plots highlight clearly the eras at which different fuel types were more prevalent. We see that highly-polluting petroleum coke plants were built mostly in a five-year period 35 years ago, whereas more modern fuels such as solar or wind have only been established in the last seven years or so. In fact, all renewable sources with the exception of hydro power have generally been built sooner than those non-renewable.

**Data Preparation:** This chart required only the provided GPPD data set, so preparation was minimal; age was calculated for each plant based on the year of commissioning, a simple query was used tell apart renewable and non-renewable sources.

Additionally, outliers have been hidden on the chart as they introduced almost random clutter with little meaning.

## Part 1, design 3

**Description**

**Visual Design Type:** Area chart

**Name of Tool:** Altair

**Country:** United States

**Year:** 2001-2018

**Visual Mappings:**
- **Area**: Power generation contribution for each fuel type is shown by the area occupied by each stacked band of colour. At any given year, the bands combine to 100 percent of power generated.
- **Hue**: Fuel type is distinguished using a categorical hue scale.

**Unique Observation:** This chart illustrates how electricity production in the US has evolved over an 18-year period. It is not concerned with actual production amounts, instead showing a normalised percentage contribution of each fuel type to the total power generation. Unnormalised charts may deceive users as changes in production of one fuel type can appear to have cascading effects on others.

We can then see clearly that renewable sources, especially wind and solar have slowly risen from almost nothing - whereas oil has largely died out. Whilst nuclear power remains largely the same, gas has overtaken coal as the dominant fuel type - presumably as the United States look to reduce their energy emissions.

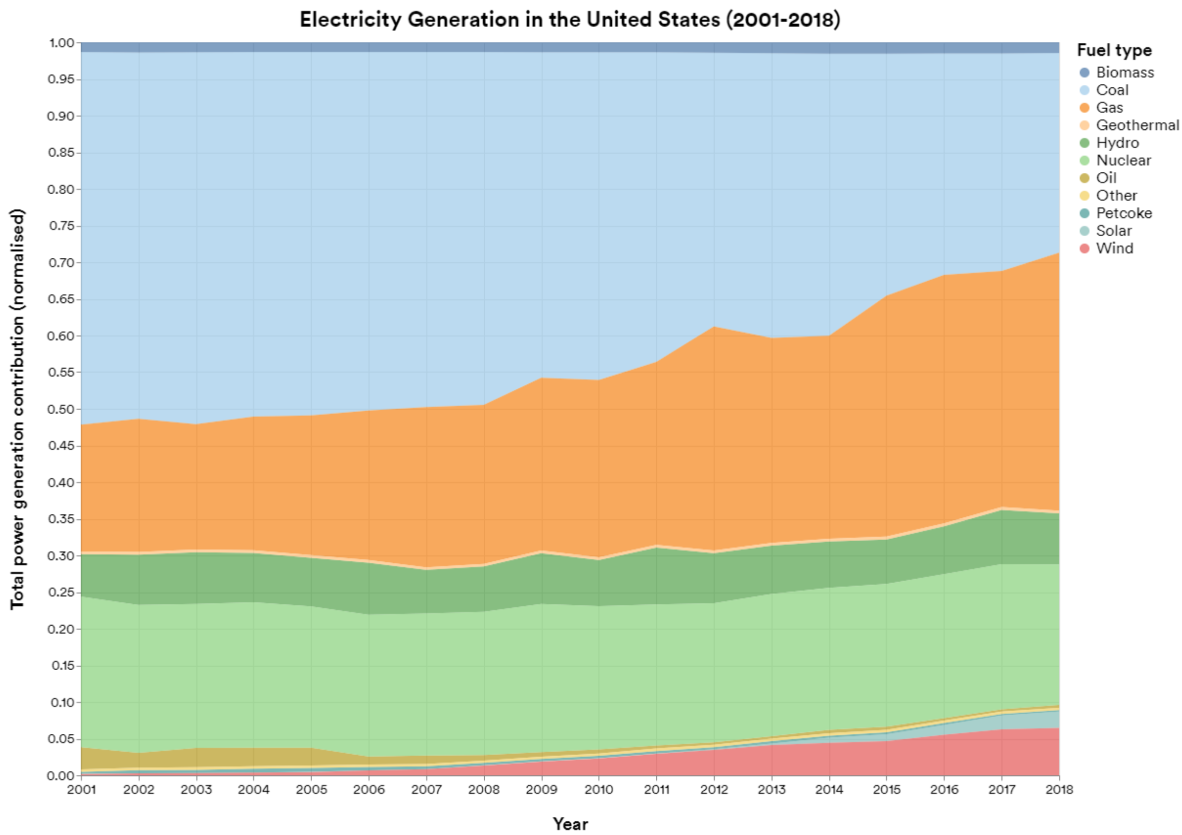



Figure 3: Electricity generation in the United States (Design 3)

**Data Preparation:** The provided GPPD data set was combined with a second which detailed electricity generation in the US by fuel type, dating back to 2001. Generation data was aggregated by year, then normalised in order to determine percentage contribution of each fuel source rather than amount produced.

## Part 1, design 4

**Description**

**Visual Design Type:** Interactive scatter plot

**Name of Tool:** Altair

**Country:** Worldwide

**Year:** 2014

**Visual Mappings:**

- **Position**: Countries are represented as points on the chart, where their national power generation is compared with their national $CO_2$ emissions.
- **Hue**: The main fuel type of each country is indicated by the hue of the chart point.
- **Tooltips**: Tooltips are used to offer detail-on-demand to the user, showing exact generation and emissions figures.
- **Line of regression**: A computed line of regression is drawn over the points in order to evaluate countries' performance relative to the average.
- **Scroll**: An interactive chart means the user can drag and zoom the chart, allowing for points to be inspected with greater granularity.

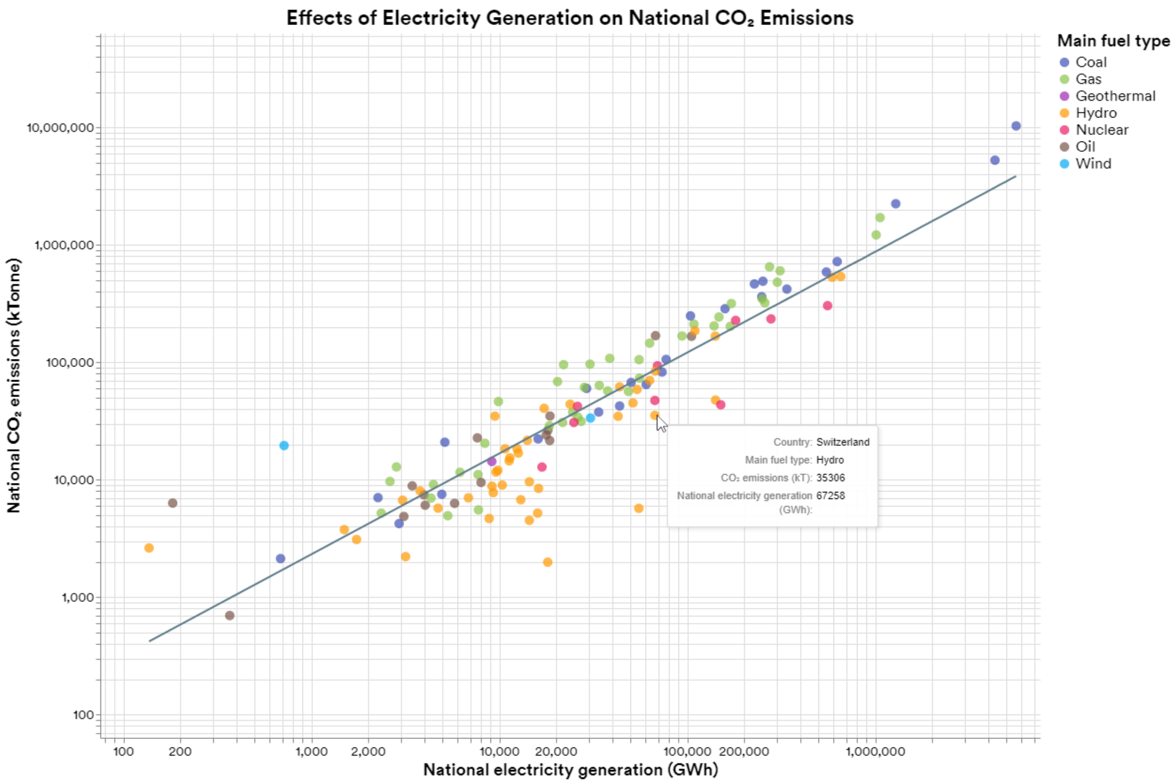



Figure 4: Effects of electricity generation on national $CO_2$ emissions (Design 4)

**Unique Observation:** The chart shows a strong correlation between national electricity generation and $CO_2$ emissions, and since the scale of the chart is logarithmic we can deduce that increased electricity generation correlates with an exponential increase in $CO_2$ emissions on a national level.

It is easy to compare countries in both generation and emissions, and the line of regression gives the user a baseline average from which countries' performance can be evaluated. For example, we can see that Switzerland with hydro power generates a similar amount of power to Finland with nuclear power, however Switzerland produces around 10,000 fewer kilotonnes of $CO_2$.

Additionally, more sweeping observations can be found, such as the trend for countries with gas as their main fuel tend to have above average $CO_2$ emissions as the green points largely appear above the trend line. Inversely, hydro power countries with yellow points tend to have less than average emissions.

**Data Preparation:** The provided GPPD data set was combined with a *UN Environment Programme* data set in order to append $CO_2$ emissions figures. This merging was achieved via a third table, which was used to translate between different ISO country code formats.

Similarly to design 1, power generation figures for each country were derived from the combination of two variables in the GPPD data set, and data was aggregated through several `groupby` commands to find the main fuel type for each country.

The points were observed to trend exponentially, so the scale is drawn logarithmically such that points now appear linear and compact.

# Part 1, design 5

**Description**

**Visual Design Type:** Scatter plot matrix (SPLOM)

**Name of Tool:** Altair

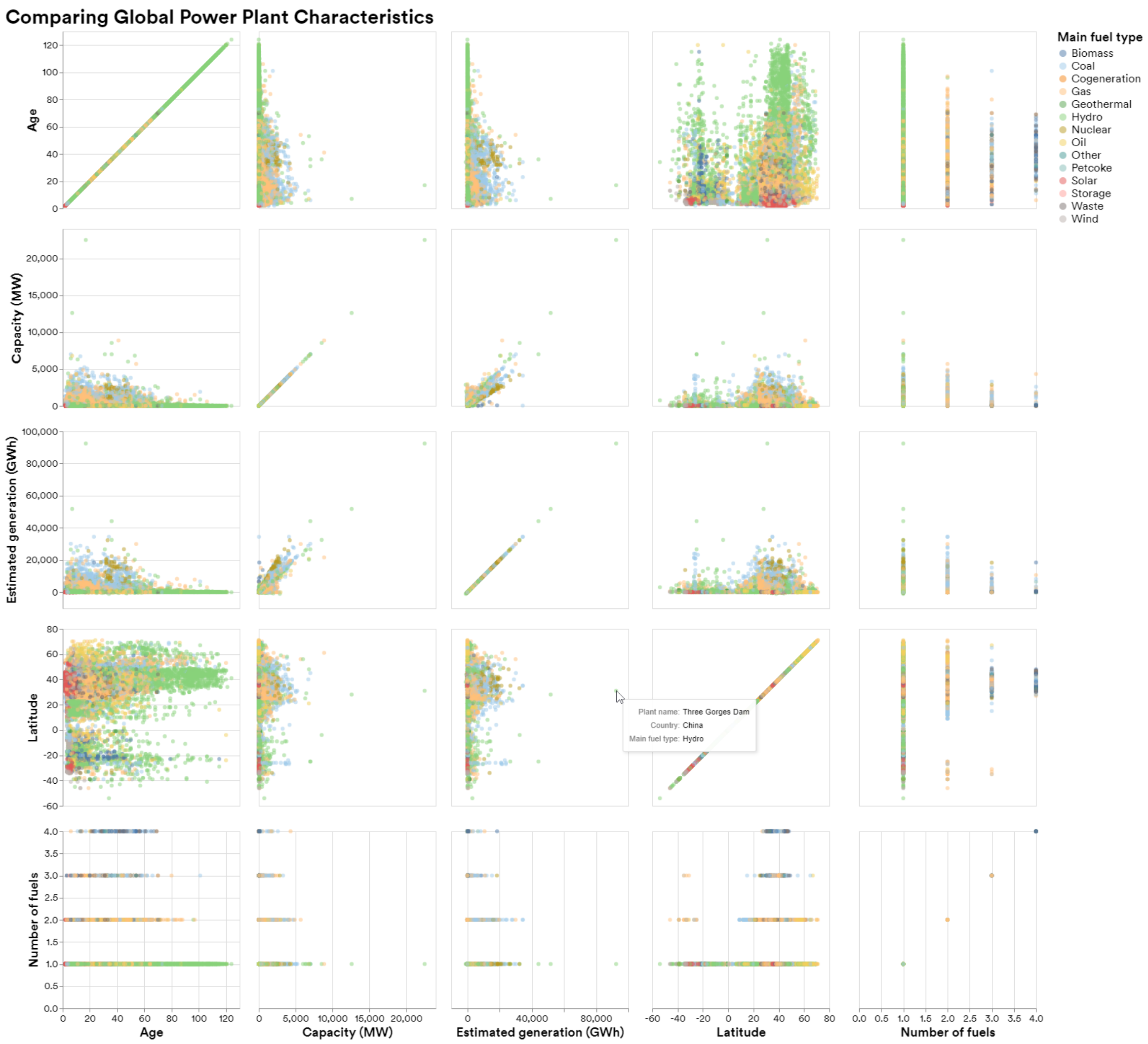



Figure 5: Comparing global power plant characteristics (Design 5)

**Country:** Worldwide

**Year:** 2014

**Visual Mappings:**

- **Attributes**: List of five quantitative attributes for each power plant are compared in every combination to indicate correlation, trends, and outliers.
- **Hue**: Power plant main fuel type.
- **Tooltip**: Tooltips are used to offer detail-on-demand to the user, showing individual plant names and locations.

**Unique Observation:** Correlation between fields is made apparent using this chart, where we can see that there is a distinct lack of power plants located near the equator, and what is there have been constructed recently. The northern hemisphere generates more electricity in general, however global plant capacity can be seen to be proportional to power generated as the two rows appear very similar.

From this perspective, we can also spot the Three Gorges Dam as an obvious outlier to most of our plots with its extremely high capacity and generation. The dam is shown to operate on three fuel types in total and it follows the trend that power plants with more fuel types tend to have greater capacities.

**Data Preparation:** From the provided GPPD data set, fields were combined and derived to form five key attributes which represent each power plant in the data set. These attributes were then used to sequentially generate charts and populate the SPLOM.

# Part 2, Treemap

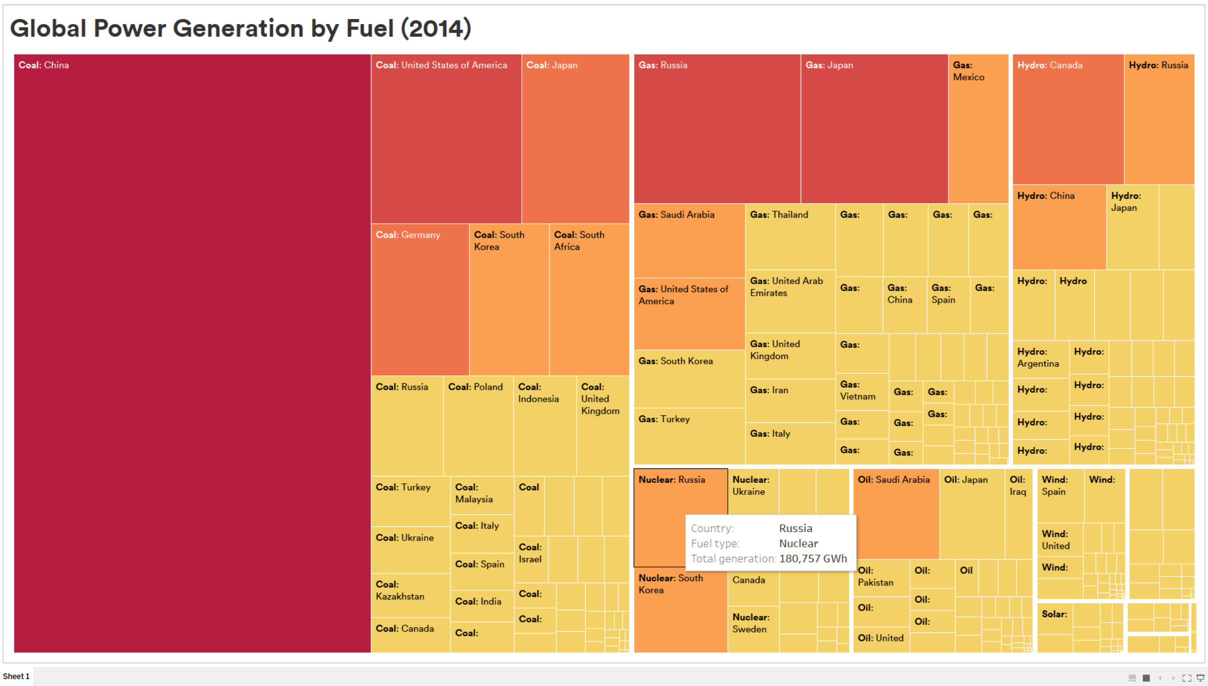



Figure 6: Treemap

Apologies, however I didn't leave enough time to complete this part to the best of my abilities. Blank fields are in italics.

- **Name of Tool**: Tableau
- **Country**: Worldwide
- **Year**: 2014
- **Data Preparation**: Total annual generation per fuel type was summed per country.
- **Color**: Power generation is represented by a banded linear colour map. The upper bound has been limited due to China's enormous generation skewing the scale.
- **Hierarchy**: Each fuel type leaf node is made up of the countries which produce it.
- Leaf node size is mapped to the total annual generation per fuel type - the fuel type which accounts for the most power generation is largest.
- *How are the leaf nodes laid out or positioned?*
- Internal nodes are mapped to the countries which produce power with the relevant fuel type.
- Internal node size is mapped to the amount of power produced by the country.
- *Which treemap node layout algorithm is used?*

# Part 3

See next few pages

VisGuides

# Multiple colour encodings for countries in a choropleth map

Visual Design

**matt** 1d

**Question**:
I'm working on a visualisation that aims to find interesting patterns and trends in a world power plant data set, which contains all manner of statistics and production figures on the plants. As part of this, I have chosen to draw a choropleth map of the world in order to illustrate the most used power sources globally.

The problem is that I'm starting to doubt my use of visual encodings - specifically the hue and saturation of the countries. Here's where I am so far:

**Name of Tool:** Altair
**Dataset:** Global Power Plant Database
**Country:** Worldwide
**Year:** 2014
**Visual Mappings:**

- The hue of a country represents its main fuel type; the fuel type it uses to generate the most electricity.
- The luminance of a country represents the amount of electricity it produces across all fuel types. It is banded into five discrete values along an exponential scale.
- Tooltips are used to offer detail-on-demand to the user, showing precise information about a country's electricity production. (so just imagine tooltips 😺)

Output:

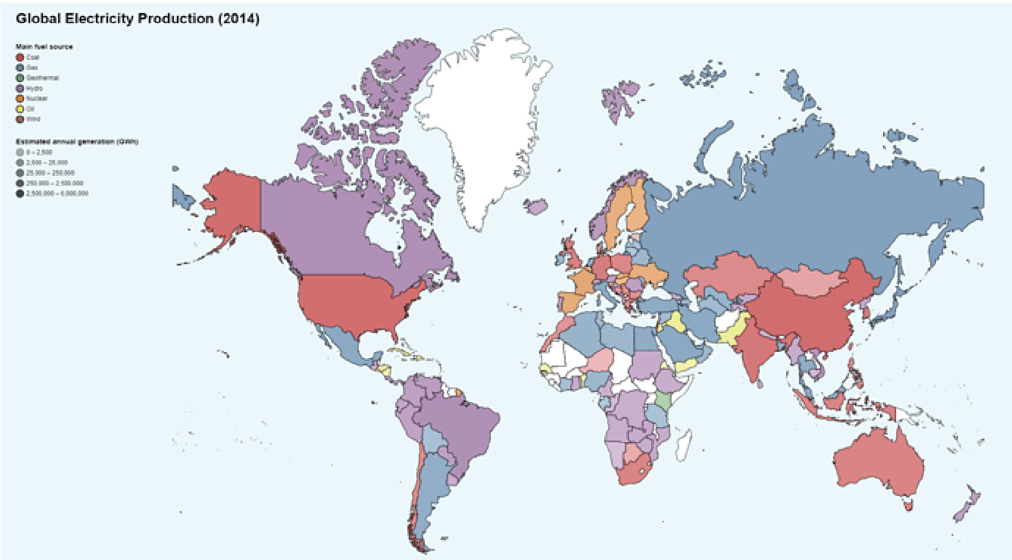
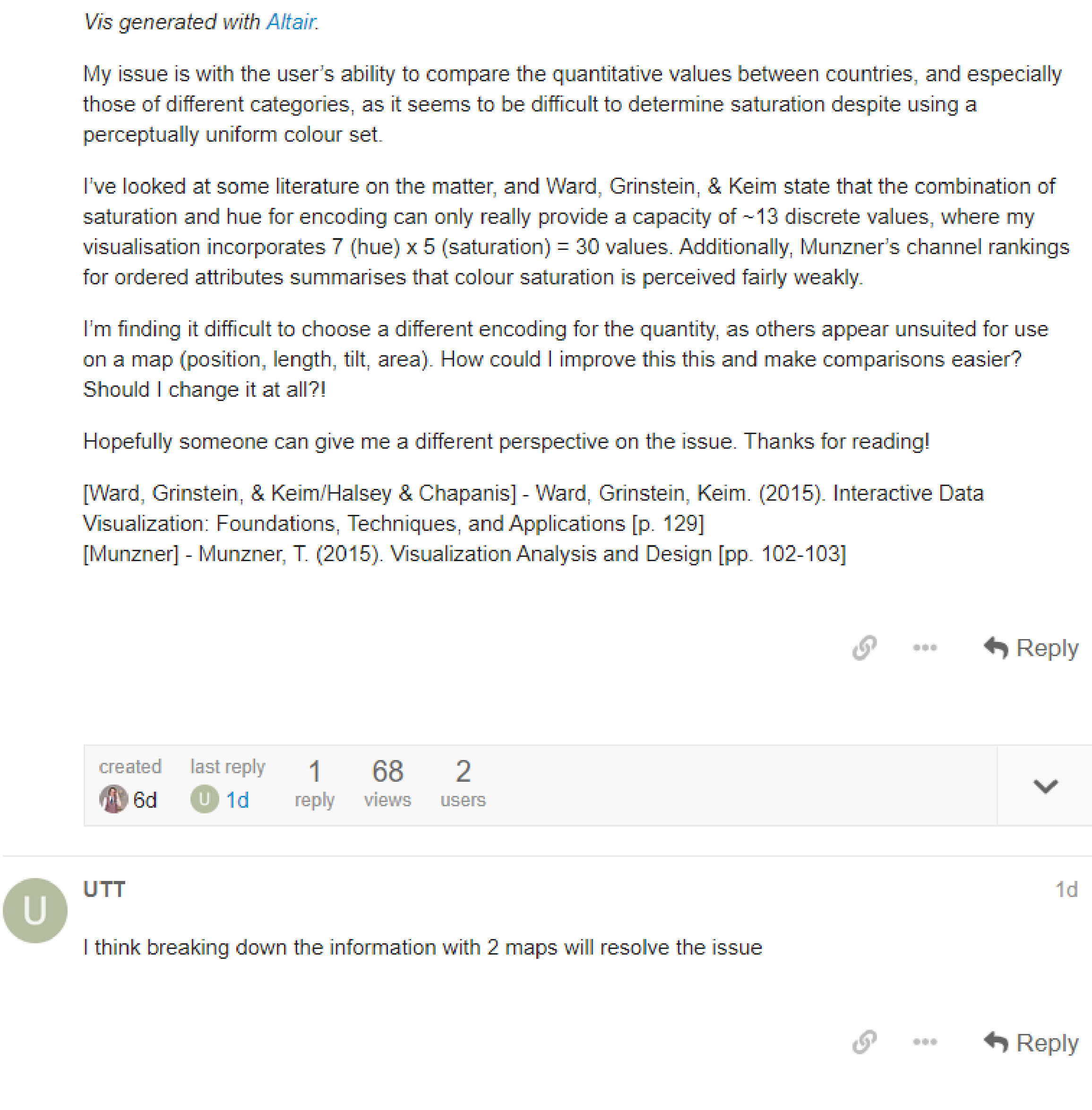



Figure 7: VisGuides screenshot: my post (1 of 3)



*Vis generated with Altair.*

My issue is with the user's ability to compare the quantitative values between countries, and especially those of different categories, as it seems to be difficult to determine saturation despite using a perceptually uniform colour set.

I've looked at some literature on the matter, and Ward, Grinstein, & Keim state that the combination of saturation and hue for encoding can only really provide a capacity of ~13 discrete values, where my visualisation incorporates 7 (hue) x 5 (saturation) = 30 values. Additionally, Munzner's channel rankings for ordered attributes summarises that colour saturation is perceived fairly weakly.

I'm finding it difficult to choose a different encoding for the quantity, as others appear unsuited for use on a map (position, length, tilt, area). How could I improve this this and make comparisons easier? Should I change it at all?!

Hopefully someone can give me a different perspective on the issue. Thanks for reading!

[Ward, Grinstein, & Keim/Halsey & Chapanis] - Ward, Grinstein, Keim. (2015). Interactive Data Visualization: Foundations, Techniques, and Applications [p. 129]
[Munzner] - Munzner, T. (2015). Visualization Analysis and Design [pp. 102-103]

Reply

created 6d | last reply 1d | 1 reply | 68 views | 2 users

UTT 1d

I think breaking down the information with 2 maps will resolve the issue

Reply

Figure 8: VisGuides screenshot: my post continued, and a response (2 of 3)

**matt** 4d

I like the look of the map overall; the presentation is very clear and well laid out.

I think your issue with the colour stems from the use of the diverging colour scheme, in which the two different colour of the countries seem to indicate a categorical attribute rather than the obviously quantitative percentage being illustrated.

You mention having issues with single hue scales, however I would still suggest them to be the best in this situation. Potentially binning/thresholding the percentages into discrete bands would make the different saturations stick out more - I believe Tableau has an option 2 to use 'stepped colour' when adjusting quantitative colour scales.

Alternatively, you could try using a multi-hue colour scheme 2 to potentially make the difference stick out more.

Hope this helps! 😀

Figure 9: VisGuides screenshot: me answering a question (3 of 3)