

# Assignment 1 — Information visualization

Deadline 11 a.m. on Tuesday 25/02/2020

Dr T. Torsney-Weir

By submitting this coursework, electronically and/or hardcopy, you state that you fully understand and are complying with the university's policy on Academic Integrity and Academic Misconduct. The policy can be found at <https://myuni.swansea.ac.uk/academic-life/academic-misconduct/>.

**Recommended hours:** 20–25 hours

**Credits for CSCM37** 15% of module

## Motivation

You've been working at Global Power Plant Planning Production (GPPPP), a company who contracts new green power plants around the world. Your boss comes by one day and asks, that since you do “something with computers,” if you could look over current power plants around the world and find good places for some new power plants. They mention that there's some data online at <http://datasets.wri.org/dataset/globalpowerplantdatabase> for you to use.

You have a look at the dataset and it is a lot of complex data. Maybe you can use visualization to help make some sense of the complex data.

## Understanding the data

As with real-world projects, acronyms and special terminology are prevalent. Thus, you may need to conduct some research, like reading the data description online or googling terms you're unsure of. You may also ask questions on <http://VisGuides.org> if you need help.

## Goals of the analysis

The goal is to create visualizations that maximize our understanding of the data. In addition to some obvious factual information, such as:

1. What is the oldest power plant in each country?
2. What is the distribution of fuel types around the world?
3. Which fuel types are responsible for most/least of the power generation?

There are also some less obvious questions, such as:

1. Are there any interesting patterns or trends in the data?
2. What are the outliers?
3. What insight can we gain from multi-variate visualizations of the data?
4. Are there any relationships between the fuel types?

Your boss shows you a bar chart they made in excel and asks for, “something like this.” However, you know from your insightful vis lectures that old-fashioned pie charts, bar charts, bubble charts, and line graphs might not be very insightful or aesthetically impressive. We resolve to do better (and maybe get a promotion)!

## Useful Links to Visualization Tools

Here is a list of visualization tools that you may use for this assignment. You can potentially use something else but you **must** clear it with me first.

- Data Driven Documents (JS): <http://d3js.org/>
- Vega (JS): <https://vega.github.io/vega/>
- Altair (Python): <https://altair-viz.github.io/>
- Matplotlib (Python): <https://matplotlib.org/>
- ggplot (R): <https://ggplot2.tidyverse.org/>
- Gadfly.jl (Julia):
- Lyra: <https://idl.cs.washington.edu/projects/lyra/> (sort of an open-source version of Tableau)
- Tableau: <http://www.tableausoftware.com/> (free trial available)

To help you get started using the tools, here are two live, hands-on, video tutorials of how to use Tableau:

- *Visualization in Practice: A Live Introduction and Hands-on Demo of Tableau* <https://youtu.be/vah0Y6YDhts>
- *Applied Visualization using Tableau: A Hands-On Tutorial and Demonstration* <https://youtu.be/bIz0BuEvHQ4>

Both carefully prepared tutorials start from the very beginning with a description of a sample data set and proceed step-by-step creating a selection of visual designs.

A paper called, *Visual Analysis of Document Triage Data*(Geng et al. 2011), provides a nice overview of free information visualization tools and their features. If you have any questions or concerns, please do not hesitate to contact the module lecturer or the module teaching assistant.

## Submission template

There is a submission template located on Blackboard or at [https://github.com/SwanseaU-TTW/csc337\\_coursework1](https://github.com/SwanseaU-TTW/csc337_coursework1). There are template files in markdown and LaTeX. If you use the markdown files you will need to have pandoc installed. Be sure to enter your name and student id in the `submission.md`

and/or `submission.tex` file. You can use the `md_build.sh` or `md_build.bat` for creating a final submission document if you are using markdown.

## Part 1: A Real-World Challenge (maximum marks 50)

Your task is to produce **five different** visual designs which can convey some meaningful and hopefully interesting insight about the data. You are required to supply five unique observations about the data (and not just repeat the same observation over and over). For each design, fill out a description template as below.

You are allowed and encouraged to further abstract the data, e.g., aggregating data records into groups, combining some attributes together, or add new data, e.g., cities, geographic regions. In fact, you can do anything you like to the data and you are expected to make changes, e.g., perhaps combining some spreadsheets together or creating new spreadsheets.

### Description Template

Provide the following information for each visual design you create:

**Image:** The visualization itself as an image

**Visual Design Type:** The name/type of the visual design

**Name of Tool:** The tool that was used to generate the image

**Country:** Name of country(s) data shown

**Year:** the year(s) or time-span of data shown

**Visual Mappings:** Each of the visual design mappings. Include the data mapping information about color, shape, size, position (x,y axes), and any other visual mappings.

**Unique Observation:** Things we can learn from the visualization, e.g, from this visualization we can see this pattern...

**Data Preparation:** Any modifications to the original data that had to be performed to generate your beautiful image.

A good observation requires interpretation of the resulting image that you generate.

### What is better?

Create visualizations that:

- convey information and knowledge
- aid discovery of: patterns, deviation, hierarchy, relationships and association
- identify relationships between data attributes
- depict data at different scales

- separate noise from the signal
- are intuitive to laypersons and easy to learn
- are aesthetically pleasing

**Examples:** Simply creating more standard pie charts, line charts, and bar charts is not good enough. Some examples of more advanced visual designs are treemaps, parallel coordinates, matrix charts, scatter plots etc. You will see many examples in the lectures. If you'd like to get a head start, you may watch the course lectures from last year on YouTube ([https://www.youtube.com/playlist?list=PLZo40sVmw\\_4No1\\_K-Xllf04O977PYngq1](https://www.youtube.com/playlist?list=PLZo40sVmw_4No1_K-Xllf04O977PYngq1)), look at the slides on Blackboard, or read the recommended text book (Ward, Grinstein, and Keim 2015). You are also encouraged to ask questions in lectures.

Describe the insight that your visualizations provide. What can we learn from your visualizations? How are they better than a standard line, pie, or bar chart? Also, a common mistake that students make is using a treemap to show non-hierarchical data. Treemaps are a hierarchical data visualization technique. Without a hierarchy, they may be even less effective than a bar chart. However, you can *create a hierarchy* from your data to visualize. See the example submission and Geng et al. (Geng et al. 2011) for examples.

**Digital Maps:** You are encouraged to explore the use of digital maps in your visualizations. Using digital maps will yield bonus points. For a good source of digital maps, please visit the following URL: <http://commons.wikimedia.org/>. You can then do a search for the relevant maps.

**Hands-On Tutorials:** For extra help, you may also watch either one of the following live, hands-on, video tutorials of how to create a visual design including a geo-spatial map:

**Visualization in Practice: A Live Introduction and Hands-on Demo of Tableau** <https://youtu.be/vah0Y6YDhts>

*Applied Visualization using Tableau: A Hands-On Tutorial and Demonstration* <https://youtu.be/bIz0BuEvHQ4>

You may certainly use the digital, geo-spatial maps from these tutorials in your submission. However, you will not receive credit for the bar charts and line charts demonstrated in the tutorial. Also, a paper called, *Liquid Diagrams: Information Visualisation Gadgets* (Andrews and Lessacher 2010) discusses the use of digital maps for visualization.

## Part 2: More Depth (maximum marks: 20)

For part two, you are to go into greater depth on one specific visual design than in part 1. You are to create a treemap, again using the power plant data. However, the description of your treemap will go into much more depth than in

part 1. For this part, you provide the following information:

- **Image:** the treemap image you are describing
- **Name of Tool:** The tool that was used to generate the treemap
- **Country:** Name of country(s) data shown
- **Year:** the year(s) or time-span of data shown
- **Data Preparation:** A helpful description of how you prepared the data
- **Color:** what is color mapped to?
- **Hierarchy:** What is the data hierarchy contained in the treemap?
- What leaf node size is mapped to?
- How are the leaf nodes laid out or positioned?
- What are internal nodes mapped to?
- What is internal node size mapped to?
- Which treemap node layout algorithm is used?

Some of these questions may require some research or extra background reading to answer. The data set used must contain at least 1,000 data records (rows). Also, you may not simply re-use a treemap from part 1. If you used a treemap in part 1, then create a different one for part 2.

### Part 3: Getting Professional Help (maximum marks: 30)

For Part 3, you are to request professional help from VisGuides.org on your assignment. VisGuides is a democratic discussion forum about visualization guidelines. It welcomes all students, scientists, researchers, practitioners, designers, engineers, and users who are interested in data visualization and visual analytics regardless of their background. Anyone may post a visualization question or provide answers to visualization questions on VisGuides.org. You will have to register with the site.

Your task is to choose an image from Parts 1 or 2 and upload a description of the image and the image itself to VisGuides.org. You then pose a question(s) on the site in order to help you complete the assignment.

The description you provide should contain all of the information from the **description template** from Part 1.

Make sure you tell the readers which dataset you are trying to explore. Add a reference to the literature you used when creating the visual design. It could be a paper, a blog, a book, or something that they can use to support their questions. Your description also contains the URL(s) of the data file(s) you used to generate the image. Here are some possible example questions you could ask on VisGuides:

- Does my visual design make sense?
- Is my choice of color map optimal?
- How can my visual design be improved?

- Are there any other types of visual layouts that would be better for this challenge?
- How else could I visualize the relationships between countries?

These are just some ideas to help get you started. You are allowed to post any question(s) you like.

When you have completed this task, you are to take a screenshot of your post and include it in Part 3 of your submission. The screenshot includes:

- 1) your description,
- 2) your image,
- 3) your question and
- 4) your Visuguides.org username.

The assessment is based on the screenshot you provide. More points will be given to screenshots that also contain an answer from another user on VisGuides.org. Thus, you may not want to wait until the last day before the deadline to do this. If you are very confident, you may also answer another student's questions and receive credit for this too.

## Submission

You are required to select an appropriate visualization tool for depicting the dataset concerned. Below is a list of possible tools that you can use. If you want to use a tool or technology not on this list then please clear it with me first. **If not, you will receive a 0 for the assignment.**

You are required to submit a document, which contains your name, student number, and **seven examples** with complete template information as in the example submission. Five examples are from Part 1, your treemap from Part 2, and a screenshot of your VisGuides post from Part 3.

*Make sure each of your visual design types is distinct for Part 1.* In other words, two Treemap layouts are two instances of one visual design.

You are required to upload a single PDF file of your report via blackboard. Any other file format (zip file, word, powerpoint, html, etc) is not acceptable and will be marked with 0 points.

## References

Andrews, Keith, and Martin Lessacher. 2010. "Liquid Diagrams: Information Visualisation Gadgets." *2010 14th International Conference Information Visualisation*, 104–9. <https://doi.org/10.1109/IV.2010.100>.

Geng, Zhao, Robert S. Laramee, Fernando Loizides, and George Buchanan. 2011. "Visual Analysis of Document Triage Data." In *Proceedings of the International*

*Conference on Imaging Theory and Applications and International Conference on Information Visualization Theory and Applications*, 1:151–63. Algrave, Portugal. <https://doi.org/10.5220/0003320401510163>.

Ward, Matthew O., Georges Grinstein, and Daniel Keim. 2015. *Interactive Data Visualization: Foundations, Techniques, and Applications, Second Edition - 360 Degree Business*. 2nd ed. USA: A. K. Peters, Ltd.