COMP6214 Open Data Course Work 1

Assignment:	*Open Data Innovation*	**Lecturer:**	*Sc2*	**Weighting:**	*20%*
Deadlines:	24/Mar/2017 @16:00	**Feedback:**	1/May/7	**Effort:**	*30 hours*

Instructions

This coursework requires you to take a provided dataset and create two of different and appropriate interactive web visualisations. You will be required to clean and perform simple manipulation upon the dataset to prepare it for visualising as well as fixing errors.

The hand-in will be a copy of the web applications source and a text file explaining the visualisation with a link to a working version hosted by you.

It is a requirement of this coursework that you must use the provided data set. The dataset should not be considered "real" data. The dataset has been heavily modified in order to evaluate your ability to clean, manipulate and visualise such data. The data is provided in csv and can be found via the course website. Using this dataset you must build a website that must show at least one visualisation of the data. You should aim to develop multidimensional (greater than 2 dimensions) visualisations that enable rich exploration of the data. The visualisations should be appropriate to the dataset and appropriate for the target audience or use case of your choosing. The web page must include a description of the visualisation. This description should detail:

- What the visualisation is showing
- How the visualisation is interactive
- Who the intended audience is

It is recommend that you use the D3 visualisation library for this task, however you may construct your visualisation in any way you choose, but, the application itself must dynamically generate the visualisation not just present a pre-rendered image.

You will be rewarded for a robust application that can cope with identified data errors, run in a stable fashion, and has suitable interactivity that allows for manipulation, filtering, and detailed analysis of the visualisation. You will also be assessed on your ability to identify and handle a number of different errors in the dataset. These errors should be accounted for these either through pre-processing (using tools such a Google refine) or with your visualisation application itself. You must detail how you went about finding errors and how you handled them in your accompanying text file, or clearly on your website.

The hand in for this coursework is a source code zip and a text file. You must also host working copy of your visualisation somewhere (e.g. your ECS web space) and ensure it is accessible between handin and the feedback date outlined at the top of this document. It is your responsibility to keep this visualisation up and accessible.

The source code zip must contain all code for the application in such a way that another running installation could be easily established. This text file should be kept concise (it is not a report) and should not be longer than 500 words without good reason. The accompanying text file should contain:

- A short description of how to host a copy of your visualisation.
- A short description relating to each error you found and how you have accounted for each.
- You should also detail how you manipulated the dataset in order to clean the data.
- There are 5 marks available for this suggesting you should look for at least 5 different errors or places where the data will need manipulating.
- An overview of the audience and use case for each visualisation and why your visualisation is appropriate both to this audience and the data. (or a pointer to this data on your website)

- A description of the interaction each visualisation provides and why this interaction is appropriate both to the audience and the data. (or a pointer to this data on your website)
- Any details of any other operation you feel exceptional, such as data enrichment.

Submission

Three things must be submitted:

1. A zip archive containing your solution.
2. A short text file containing the content required above.
3. A URL (in the text file) to your hosted solution that you must ensure is available until the feedback date.

The first two items should be submitted electronically via handin.ecs.soton.ac.uk. We recommend you host your solution on your ECS web space (as no marks are lost for ECS being offline).

Relevant Learning Outcomes

1. Identify innovation opportunities for open data.
2. Be able to apply appropriate validation, cleaning and transformation to use, reuse and combine a multitude of complex datasets.
3. Critically evaluate a large range of Infographics and interaction techniques suitable for different tasks. Marking Scheme

Criterion	Description	Outcomes	Total
Visualisation Choice	Each visualisation was appropriate to the data and audience. Each visualisation was well presented showing careful design. The visualisation presented multi-dimensional data.	1,2,3	5
Implementation	Good use was made of an appropriate library for presenting dynamically loading visualisations. The implementation is considered elegant and maintainable, is robust and performs smoothly and without errors	2,3	4
Interactivity	The visualisation provides interaction that is appropriate for the audience. The interactivity allows powerful filtering, selection and analysis of the data.	2,3	4
Cleaning and Manipulation	The student has identified a number of different types of errors in the dataset. The student has applied suitable techniques to fix errors and manipulate the dataset ready to be visualised.	2	5
Completion	The student has shown exceptional ability to provide complex, yet highly suitable visualisations that have potential to be used commercially.	1,2,3	2

Further Resources

- *Getting Started with D3*, Mike Dewar, O'Reilly Media, 2012
- *Information Visualization: Perception for Design,* Colin Ware, Morgan Kaufmann, 2004 •

Visualising Data, Ben Fry, O'Reilly Media, 2007