

COMP6214 Open Data Innovation Coursework 1

Assignment:	Open Data Innovation	Lecturer:	Heather Packer hp3@ecs.soton.ac.uk	Weighting:	20%
Deadlines:	15/Mar/2018 @16:00	Feedback:	01/May/18	Effort:	30 hours

Instructions

This coursework has two parts, the first requires you to clean the provided dataset, and the second requires you to create one visualisation using D3.

You must use the provided dataset for both cleaning and generating your visualisation. 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.

Part 1: Clean the dataset

You will be required to clean the dataset and then perform simple manipulations to prepare it for creating your visualisation. You will be assessed on your ability to identify and handle a number of different types of errors in the dataset. These errors should be accounted for these either through pre-processing (using tools such as Open Refine). You must provide the names of the tools you used, which errors you found, and how you cleaned the error. There are 5 marks available for this suggesting you should look for 5 or more different types errors. It is not necessary to find all the errors.

Part 2: Create and host your visualisation

Using this dataset you must build a website to show your visualisation of the data. You should aim to develop multidimensional (greater than 2 dimensions) visualisations that enable rich exploration of the data. The visualisation 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 shows
- How the visualisation is interactive
- Who the intended audience is

You must use the D3 visualisation library for this task, and your code must dynamically generate the visualisation not just present a pre-rendered image. Your visualisation should have suitable interactivity that allows for manipulation, filtering, and detailed analysis of the visualisation.

You must also host a working copy of your visualisation somewhere online (e.g. your ECS web space) and ensure it is accessible between the handin and feedback date outlined at the top of this document. It is your responsibility to keep this visualisation accessible.

In the submission you will include all code for the visualisation, so that another running installation could be easily established. Please do not include the D3 library. The submission will also include a text file, which should be kept concise (it is not a report) and should not be longer than 500 words. The accompanying text file should contain:

- A short description of how to host a copy of your visualisation.
- 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

You must submit a zip archive containing:

- 1) Your cleaned csv file (part 1)
- 2) A text file, detailing tool the tools, which errors you found and how you cleaned them. (part 1)
- 3) Source code for your D3 visualisation, including any accompanying CSS or JavaScript files (part 2)
- 4) A text file, containing the URL and instructions to run your code. (part 2)

Your zip file will 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 is 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