Assignment One (A1): Information Visualisation (Individual)

Daniel Archambault

Electronic Submission Deadline:
Wednesday, Nov. 9th at 11am

Learning Outcome(s): To develop ones ability to analyse and visualise data in a programming language (Python and Altair)

1 Overview

In this assignment, you will be designing and implementing a visualisation of a data set individually and making an argument for its effectiveness. Both your designs and your prototype should be effective according to the module material. The assignment assesses your ability to demonstrate the use module material on a data set of your choice. Select your data set and task to demonstrate your knowledge and effective application of the more advanced concepts in the module. This means a series of non-interactive bar charts will not score high on the assignment. You must demonstrate knowledge of the module material.

This assignment is individual. This means that you are not to collaborate on the production of your designs and implementation. You can talk about strategies with other students, but all of the work for this assignment (both design and programming) must be done individually and must be your own work.

All written parts of this assignment should presented in a single PDF document as part of your submission. Word documents are not acceptable. **Word documents will receive a mark of 0.** Code for your submission should be submitted using a ZIP file. Include the data set as part of your ZIP. Your prototype will be implemented using jupyter notebooks and Altair. No other programming languages are acceptable for this assignment.

1.1 Altair Tutorial

This part of the assignment is not assessed, but is critical to complete it successfully.

In this module, I will not teach you Python or Altair. You will be responsible for leaning how to program in this language using this visualisation library on your own. However, I can help explain language details if you have percise questions in my office hours. Please come with partially written code that we can discuss.

Information on how to install jupyter notebook and the Altair visualisation libraries is available here:

`https://altair-viz.github.io/getting_started/installation.html`

I will ask you to use jupyter notebooks for this assignment. Please install jupyter notebooks and Altair following these instructions. I will also post instructions for installation on lab machines which you can use for your own machine of a similar configuration.

You are expected to become proficient in both of these for the module. Please make use of the following tutorials:

`https://altair-viz.github.io/altair-tutorial/README.html`

I would spend a bit of time doing these tutorials and familiarising yourself with how to use python and Altair so that you are able to begin to program in it comfortably. In my lectures, I will signpost you to when you should be doing these tutorials. These tutorials are required as a minimum, but you should exceed them to achieve higher marks (ie you will need to learn much more). The more you demonstrate that you've gone above and beyond the tutorials and sought external resources to the module, the greater chance for a higher mark.

This part of the assignment is not worth any marks, but is probably the most important part of the assignment.

1.2 Choose a Data Set

Choose a data set that interests you from the following repository (UCI Machine Learning Repository):

`http://archive.ics.uci.edu/ml/index.php`

You can choose any data set **except the wine data set**. Make sure that your data set is not too complex and not too simple (more than four data columns and a variety of data types). You can choose data types that we haven't covered in the module yet, but you would be responsible for learning effective representations for this data. If you have issues selecting a data set, please come and see me during my office hours and we can discuss it.

Given this data set, describe how the data set is structured and describe a task you think the user should be able to undertake with this data (no more than two concise sentences). At the end of your description, provide a link to the page that supplies and describes your data. Your description of the data and the user task will be worth **10 marks.** A concise description of the data is worth five marks and a concise and focused description of the task is worth five marks. For the data description, concentrate on the meaning of the columns and not facts about the columns. For the data description, a bullet point list is okay.

1.3 Create a Design

Given the chosen data set, create **two visualisation designs** that would help explore the contents of this data set. This designs should support an intended task that the user has when exploring the data. The design you create must be accompanied by a 3/4 of a page argument as to why it would be effective for this data. You should use the course material and citations from the scientific literature that goes beyond the class material to support your arguments. Make sure you clarify how the interaction would work in your paper prototype. A collection of non-interactive bar charts will not acheive a high mark. More effort to seek correct and relevant external sources will be needed to achieve high marks. Your design should be as realistic as possible, but should be paper based. Designs drawn on paper by hand will generally receive lower marks. Designs made with relevant prototyping tools that you have sought externally will receive higher marks.

The design is worth a total of **40 marks** for this section of the assignment. There will be **10 marks** for each design and **10 marks** for each argumentation with respect to the literature.

1.4 Implementation

Implement your selected design as faithfully as possible using jupyter notebooks and Altair. No other languages or environments will be acceptable.

A number of tutorials and resources are here:

`https://github.com/uwdata/visualization-curriculum/` and `https://altair-viz.github.io/`

Your implementation should use a relevant data visualisation technique as covered in the module. If you don't use them or use very little of them, you will receive lower marks.

Use your developed tool to explore your data to support the intended task that it was developed for. When you find interesting features of the selected data set, take a screen capture of these features and include them in your report. If these features are known about your data set, provide citations for the know feature and argue using the image that your visualisation well represents this feature.

The quality of the prototype will be worth **30 marks**. There will be **15 marks** for the visual design and **15 marks** for interaction with the visual design.

The description of the discovered features will be worth **20 marks**. If you are unable to get your implementation working by the deadline, explain how your code works and what you think you need to do to get your code working for partial marks.

1.5 Submission

All written parts of this assignment should presented in a single PDF document as part of your submission. Word documents are not acceptable. **Word documents will receive a mark of 0.** Your python notebook file and your data should be submitted using a

ZIP file. Your program should easily run via jupyter notebooks.

Both PDF and ZIP should be uploaded and submitted to canvas by the indicated deadline. No hard copy submission is required for this assignment.

By submitting to Canvas you acknowledge that you have read and understood the following:

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`.

If you do not understand this text, please come see me during office hours.