SWEN 422 Visualisation - Final Report

Owen Bannister - 300172912

Abstract

Visualising information about cars and their attributes in its database form is difficult[**?**]. Representing this information graphically can resolve this by allowing users to see trends and relationships in the data. This report details the design, implementation, and evaluation of a visualisation attempting to address this need. The created visualisation successfully allows users to view the trends and relationships in the data. The user evaluation discovered aspects that could be improved in the visualisation.

1 Introduction

This report describes an information visualisation that uses user interaction techniques to explore a large amount of multivariate data to convey relationships.

The system I designed was to solve the problem that there is currently no visual means of viewing the relationships between the multivariate data contained in the dataset. The main goal was to create a system that conveyed the relationships between a car's attributes and the its miles per gallon which states how economical a vehicle is.

The data chosen to visualise was a dataset of cars and their attributes that was originally used as a training dataset for Artificial Intelligence work, to discover if the miles per gallon (MPG) of the cars could be predicted.

A challenge that I faced when designing this visualisation was selecting the best visual option to represent the data and achieve my goal of conveying the information to users. I addressed this challenge by creating an explanatory visualisation that contains methods of interaction that allowed users to explore the trends in the data easily and effectively.

Choosing a design that was simple and straightforward to understand was important as users should be able to easily understand what is being displayed and how to interact with it. This lead to design features such as data filtering, modifiable layouts of the data, and multiple selection techniques. These are all discussed later in this report.

dataexample.jpg

Figure 1: Example from dataset

2 Target Users

The target users of this system are interested laypeople with a limited knowledge of cars. There are multiple possible uses of this system ranging from simple curiosity to using it as a searching tool for making a vehicle purchase.

3 Description of system

An important aspect of the system was that it needed to have an extremely low learning curve. This was so that a user could understand and use the visualisation within moments of starting it up. To achieve this, the system has a simple method of visualising each car and its attributes, all interactions were kept simple with filtering, selecting, and changing the layout of the graph nodes being the main interactions[**?**].

The system created is made up of 3 main Panels[**?**], the Graph Panel, the Interaction Panel, and the List Panel. Together these panels form a spatially separated layout which ensures that each of the visualisation elements remain distinct and easily visible to users.

Figure 2: Overview of visualisation 1

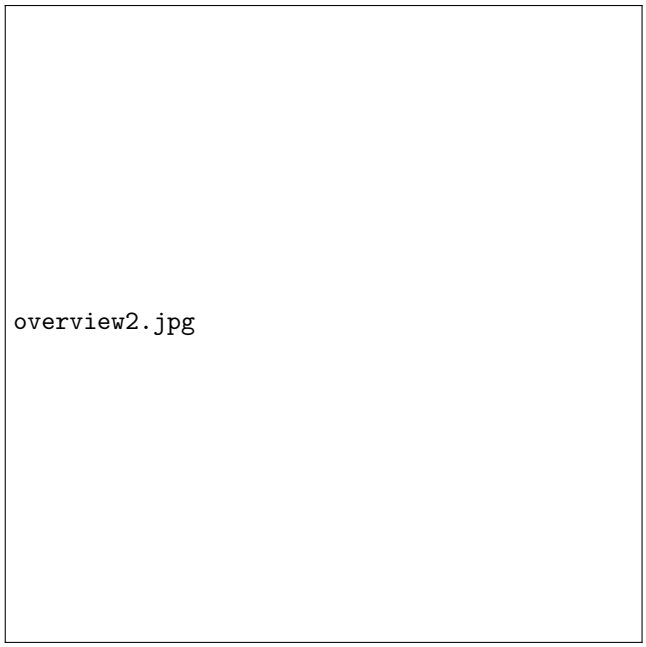

Figure 3: Overview of visualisation 2

The next 3 subsections focus on each of the 3 main panels in the system in more detail, and discuss how they contribute to the visualisation.

3.1 Graph Panel

The Graph Panel displays the information about the cars. It uses a simple x,y coordinate system in a scatter plot graph that the target audience should be familiar with. The car nodes positions on the graph correspond to the attributes selected by the user. As the focus of the visualisation is to display the MPG of each car, it was important to prominently display this attribute. To do this the MPG of each car is displayed in its colour with red being the lowest MPG and green being the highest with colours in between denoting the different ranges of MPG. The size of the Nodes on the graph also relates to the MPG, again with the smallest being the lowest and largest being the highest MPG.

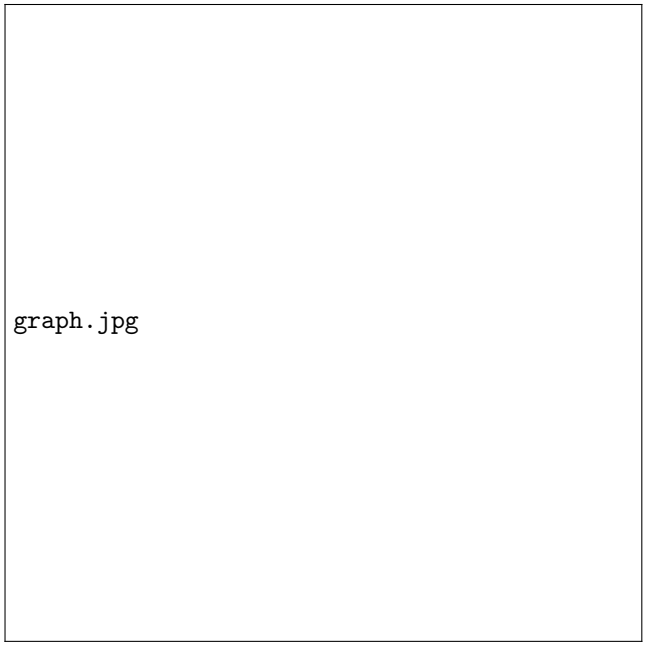

Figure 4: Graph panel

As the above image shows, many of the nodes intersect and overlap. This meant that it was important to be able to filter the data as well as re-arrange it on the x and y axis so that each element could be easily selected. To do this the Interaction Panel was needed.

3.2 Interaction Panel

The interaction panel allows users to filter the data by selecting the range for each attribute of cars. It also allows users to select which attribute will be displayed on each axis. Finally it allows users to access help which triggers a pop up detailing how to use the system.

As the images above of the Graph and Interaction panels show, there are not many details available about each node on the graph. This would mean that it is very difficult for users to find certain cars on the graph that they wish to examine further. This meant that a way to select nodes on the graph by name would be important. To allow this the list panel was created.

3.3 List Panel

The list panel allows users to see the name of all cars in the dataset and the colour depicting the MPG, which also corresponds to the colour of the nodes on the graph. This was so that cars could be searched for based on the name which users may know.

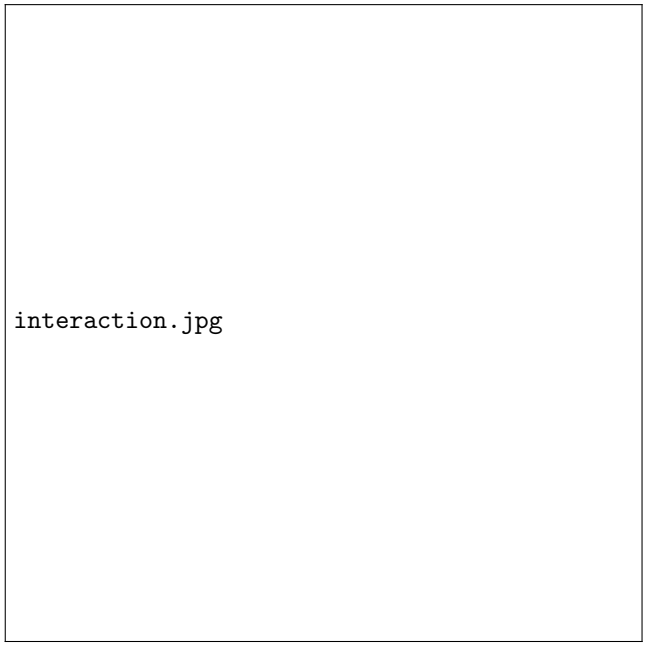

Figure 5: Interaction panel

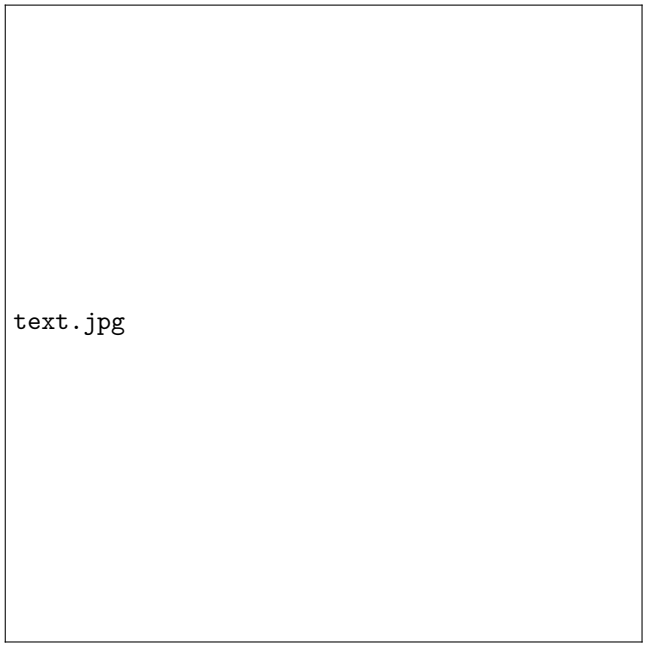

Figure 6: Text panel

3.4 Interaction Techniques

The interaction techniques in the system can be broken down into 3 types, filtering elements, modifying the axis, and selecting elements on the graph. These interactions are discussed in more detail in the following subsections.

3.4.1 Filtering Elements

Filtering graph nodes by their attributes allows users to dynamically drill down into the data to display a subset of the larger dataset. This is important because when all elements are displayed on the graph the amount of overlap and clustering of nodes means that accurate selection is difficult. Allowing users to filter the data means that this problem can be avoided as it allows the user to specify the ranges of data that they wish to see.

3.4.2 Modifiable Axis

Allowing users to select the attributes on the x and y axis of the graph means that the relationships between each of the attributes for each car can be explored at the users own pace. By ensuring that the nodes on the graph dynamically transition to their new locations of the graph whenever a new attribute was chosen ensures that the user to understand how the graph is changing as they can track the nodes as they move. The axis of the graph also changes when this selection is made to maintain the correct x and y scales for the data.

3.4.3 Selection Of Element

Allowing users to select elements of the graph was vital as there needed to be a way of accessing all of the information about every node on the graph. Selection was the most sensible option for this, as everyones natural intuition when trying to interact with something by mouse, is to point and click. This coupled with the cursor changing to indicate the interactive elements made this very effective.

4 Justification of Design

For the technology for my visualisation I chose D3.js[**?**]. D3 is a JavaScript library that allows displaying of digital data in dynamic graphics. Embedded within an HTML web page, the JavaScript D3.js library uses pre-built JavaScript functions to select elements, create Scalable Vector Graphic (SVG) objects, style them, and add transitions, dynamic effects and tooltips. Large datasets can be easily bound to SVG objects using simple D3 functions to generate rich charts and diagrams. This meant that it would be able to create a dynamic visualisation that allowed effective interaction with the data.

I chose to use a 2D graph because there were inherent relations among the data being visualised. This meant that the graphs would be extremely well suited towards displaying the relations in the data. As the dataset is not overly large the graph would not experience problems of performance loss, or issues of usability and understandability. Usability would suffer with large datasets due to the inability to select nodes, and understandability would suffer due to not being able to discern between nodes and their edges. Both of these problems lead to loss of comprehension about the information the visualisation is attempting to provide. This means that for visualisations with smaller datasets such as this, graphs are suitable [**?**].

A key component of the design was to allow users a way of filtering and comparing the data to view trends. This was because there was a large amount of overlap in the attributes. Without a third dimension to allow the data to be separated further there was the need to filter it so that the different nodes on the graph could be distinguished. Using simple interaction components such as sliders,

drop-downs, and buttons should not induce large cognitive load on the users as they try to understand how to use the system.

The colours chosen to represent the MPG of cars were based on the fact that humans can perceive the most shades of green out of any other colour. By using both red and green as well as the size to represent the MPG of each car, it helps users to see the MPG differences even when the MPG of two cars is very similar.

The layout of the system was chosen as it allowed the graph layout to be the center of a users focus, closely followed by the interaction panel and the text panel. The importance of having the interaction panel next to the graph panel was so that users could interact with the data while simultaneously seeing the changes in the graph.

5 Changes to experiment

The only modification to the experiment was changes to the user questionnaire. This was because the questionnaire did not allow users to record their experiences in their own words. To remedy this I included 4 sections:

1. What did you like about the visualisation tool?
2. What were the strengths of the visualisation tool?
3. What were the weaknesses of the the visualisation tool?
4. What could be improved for the visualisation tool?

6 Interpretation of Evaluation Results

The results of the qualitative user evaluation showed that the visualisation was easy to use, simple to understand, and effective at displaying the trends in the data [**?**]. Each participant was able to answer all of the questions that were posed in the questionnaire. This showed that they were able to use the system as intended. The heuristics that part of the evaluation was based on proved to be valuable as it outlined how usable users found the system [**?**][**?**]. The next 2 subsections detail the strengths and weaknesses that the participants discovered in the visualisation.

6.1 Strengths of the system

The participants found that one of the main strengths of the system was the ability to change the axis, which caused trends to be obvious if they were configured well. They also found that the colours used to show the MPG of the cars was effective and the fade outs and animated transitions helped them to understand what was happening and changing. The help documentation for the users which was very simple and short was found to be helpful as it summarised how to carry out each of the different types of interaction.

6.2 Weaknesses of the system

Weaknesses found in the system were that some elements would be drawn half on the screen and without a side scroller they found it difficult to understand. The participants would also forget to reset the filters when trying to answer the questions about the trends in the data. This meant that they would not see all of the graph nodes as some would be filtered out. A solution to this would be to enlarge the canvas so that nothing was ever drawn out of bounds. Another weakness was that due to the filter sliders having different ranges of data, the participants tried to click on where they assumed the slider should be but due to the differing data ranges often got it wrong. Due to the JQuery plugin used for the sliders, it was not possible to enter in the range they wanted by keyboard. Doing this would have solved the problem. A further weakness was that the axis labels were too small and so users faced a small amount of confusion about which attributes they were viewing. The solution to this is to make the labels more prominent.

There were other weaknesses that required larger modifications to the system to solve. These are discussed in the following section.

7 Proposed Improvements for System

It was observed that users wanted to compare different cars with the graph, with the current system this meant selecting different nodes on the graph and remembering the attributes of the cars. To make this easier for users it would be beneficial to allow the selection of multiple graph nodes for comparison. An easy way to do this in the current implementation would be to click on the first node to compare, drag to the next and release to make the 2 selections.

Using a method of displaying the trends in the data was proposed by both participants. This was because it would remove the need for making assumptions about the trends, and users could see actual figures about the trends in data. Pearson's Correlation Coefficient would be suitable for this purpose and would allow a line to be drawn on the graph showing the trend.

The list panel was criticised as being overly cluttered, hard to easily scan, and taking up too much screen space. A proposed solution to this was to provide a drop down for each letter of the alphabet. This would allow for all cars in the database to still be selected by name, but would only require a very small amount of screen space.

Including an image of each car was suggested as it would improve the understandability of how how the attributes looked in a car, ie faster looking cars have less MPG.

Another proposal was to modify the Graph panel to support 3D so that a third axis could be added to provide more complex comparisons of the data. However this would require a change in technology as D3 does not support 3D.

Other proposals for improving the graph was that the graphs axis should scale when data is filtered. This would mean that the data would spread out and become even less cluttered as more elements are filtered out, doing this would provide a zooming effect. Another was that when a node of the graph is selected to provide horizontal and vertical lines extending to the axis of the graph to show the exact value of the node. This would allow users to see the values contributing to its placement with more accuracy.

8 Completed Improvements for System

From the suggested improvements in the previous sections I have carried out 2 at this stage:

When a node on the graph is selected, horizontal and vertical lines are drawn.

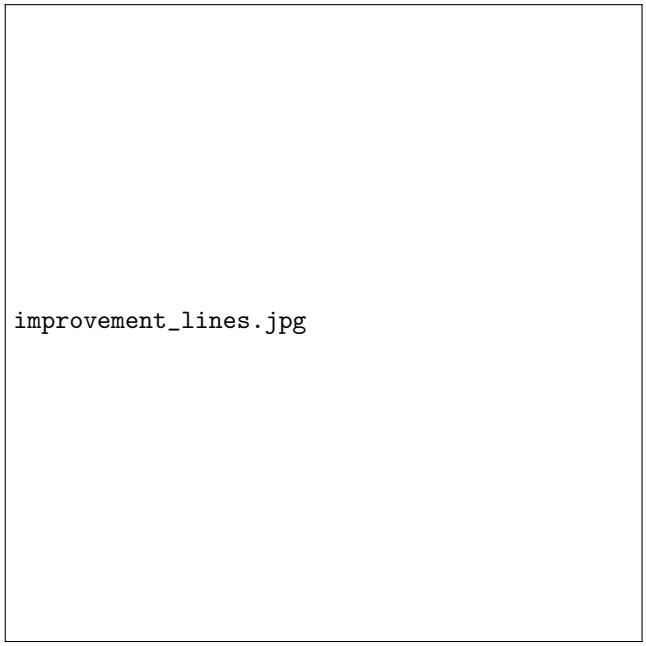

Figure 7: Lines drawn aligning to axis when node selected

All axis labels increased in size to ensure users are aware of the attributes contributing to the layout.

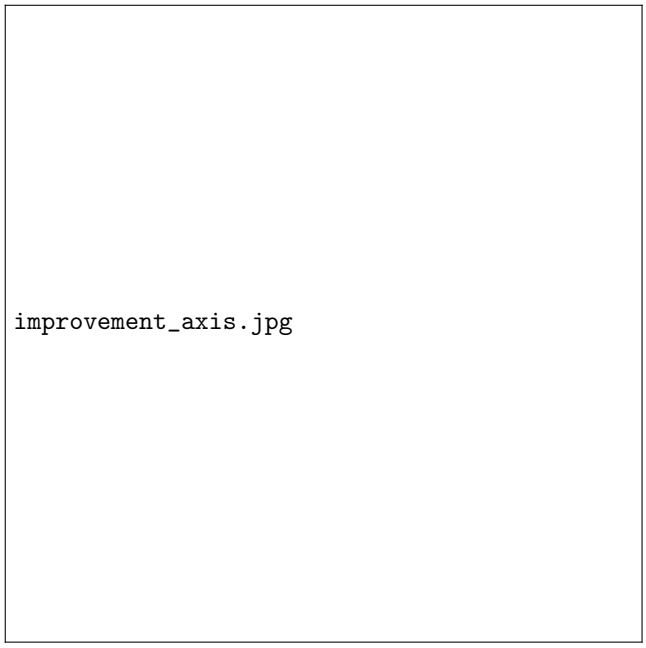

Figure 8: Size of axis labels increased

9 Limitations of the Evaluation

There were 2 limitations in my evaluation. The first was that further training and introduction to the system would have been beneficial. This is because users the participants tried to use the system to answer the questions in the questionnaire straight away before they knew how to use the interaction techniques. This lead to issues where the users tried a mismatch of the axis selection and filtering to solve a question. This issue was obvious as the users took much longer to answer the first few questions, but sped up dramatically after they knew how to use the system. The amount of time required seemed to be about 2 minutes of experimenting with the different interactions.

The second limitation was that users did not know enough about cars to easily predict relationships. I had overestimated what I assumed was "common knowledge" about cars and their attributes. However users were still able to answer the questions regarding each of the attributes without this knowledge, albeit more slowly. A solution to this would be to provide a description of what each attribute is and what it means in the context of a car.

10 Summary

The system created fulfilled its goal of allowing users to view the trends in the data to make assumptions about the relationships between a cars attributes and its MPG. The interaction techniques used were effective at allowing users to explore the data in ways that helped to convey the information in the dataset.

The user evaluation showed that users were able to use the visualisation as intended and could answer the questions provided. This showed that they could use the system to view the trends which was the main goal of the visualisation.

11 Appendix

Appendix 1. User Questionnaires