TNM111 - Assignment 2

Linnéa Svensson - linsv482

Dasmit Sethi - dasse788

Task 1

1. Explain in detail the infovis reference model, what are its strengths?

The infovis reference model is a way to think about infovis by dividing it into four different components: **raw data**, **data tables**, **visual structure** and **views**. The **raw data** is "all the info". This then has to be transformed in some way to fit into **data tables**. The **data tables** allow the end visualization to easily retrieve the data. They contain information about how the data elements are connected with each other. The **data tables** are then connected to the **visual structures** through visual mapping. The **visual structures** are the different graphical elements such as axes, plot points, substrate, shapes, legends, animations, colours etc. This mapping can for example say that all data in column A of a table is to be represented as squares in a scatterplot, while all data in column B is to be represented as circles. **Views** is the “camera” of the viewer. It’s what is currently visible. It’s where the viewer has panned to, the level of the zoom, which things are seen as in the foreground and which are seen as the background etc.. When **interaction** is involved, the viewer can often interact by panning and zooming (views), highlighting (visual mapping) or filtering (data transformation)[1]. One of its strengths is that the way the model divides the visualization into different parts can help with the planning of how to visualize a given dataset, and which kinds of interaction would affect which parts in what way.

2. What kind of interaction is supported by range sliders? Is there a way to improve them to show more information? Make a short list of pros and cons.

Range sliders allow the viewer to put a top and a bottom threshold for **filtering** in **dynamic queries**, affecting the **data transform** between the raw data and the data tables of the visualization. This can be useful when the viewer has requirements that allow for leeway. For

[1] http://www.ifs.tuwien.ac.at/~silvia/wien/vu-infovis/articles/card_2007_hci-handbook_infovis.pdf viewed 2023-01-30 12:40.

example, they might be looking to buy a sofa big enough to fit at least four of their friends at the same time, but small enough to fit into their living room. One way to **show more information** in range sliders is to provide a so-called **scent** of how the data is distributed within the available range, for example by showing numbers, using more space for categories with more matches, showing small bar charts or using different colours inside the slider to signify distribution. This can be seen on websites comparing the prices of hotels or computer parts[2][3].

Pros:

Allows imprecise queries

Allows filtering

Easy to change the query

Cons:

Takes up space

Without a scent it can be difficult at first for the viewer to figure out what thresholds to use

3. In most visualization systems selecting or highlighting a data object in a specific view leads to a highlight in another view. What is this interaction technique called? What are its advantages?

It's called brushing and linking and is a form of direct selection. One advantage is that it **keeps the display in place** while still providing insight by allowing the viewer to highlight certain aspects. Brushing and linking can be done by using a scented range slider. Another advantage is that it can bring **correlations** to light.

Task 2

Analyze the sketch considering whether it is suitable according to the alignment between chosen visual features:

Attribute types/values (in terms of expressivenes and effectiveness)

[2] https://www.inet.se/kategori/260/for-intel-processor viewed 2023-01-30 22:50.

[3] https://www.trivago.se/sv/lm/hotell-stockholm-sverige?search=200-33024;dr-20230214-20230215 viewed 2023-01-30 22:53.

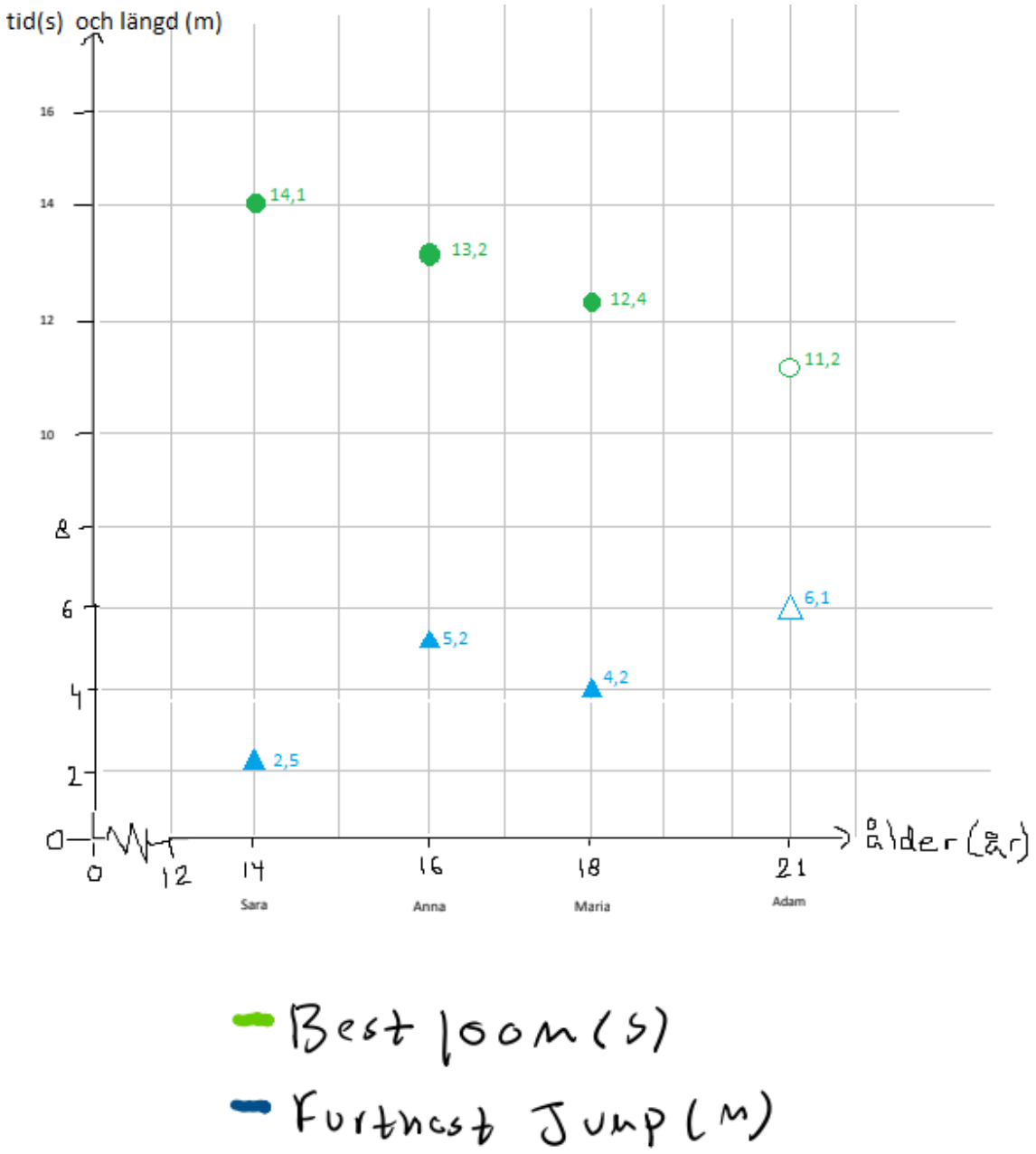
The sketch is **expressive** in that it shows all the data values in the table and nothing else through the visual structures. It can be argued that it is also **effective** because the use of different shapes and colours makes the distinction between different data types clear, and it also allows the data to be read if there was a case with overlapping data along the Y axis of the same individual. The labeled numbers also contribute to this distinction. This works as long as the individuals are all of different ages, since the X axis is quantitative. However, if two individuals were the same age, it would be hard to tell which points belong to which person. This could be solved by having only the shapes represent different data types, and using colours to differentiate individuals instead. Since the data table the sketch is based on is rather sparse, it is quite fast to interpret with this mapping. Had the table been bigger, this mapping might get a bit too cluttered if kept non-interactive. In that case, allowing interaction through panning, zooming and filtering might make it more effective. In this case, the values for number of seconds to run 100 m and the lengths of the jumps were close enough to each other to be able to be displayed along the same X axis in the same plot. Had the 100 m times been closer to each other, the jumps and the run times might have been better to show separately, for example in two adjacent plots. Alternatively, interaction through the use of a magic lens could have been used to show the differences.

Semantics of the attributes (an important part of data abstraction)

The shape and colour of the plot points have nothing to do with the real world physical counterparts they represent. The time attribute does not really have a good visual representation in the physical world, except possibly an hourglass. This goes for both the time of the 100 m sprints and the age of the athletes, but the hourglass would probably not be suitable for the age since a timeglass is usually associated with smaller time units, such as an hour, instead of years. The distance jumped can in the real world be seen with the naked eye, and it could have been represented in the diagram as bars in a bar chart, but in our case that would lead to overlapping elements and reduced visibility. In our case the abstract semantics we chose helps the visualization be easier to read and feel less cluttered than if we had used hourglass or ruler symbols to represent the data points and axes.

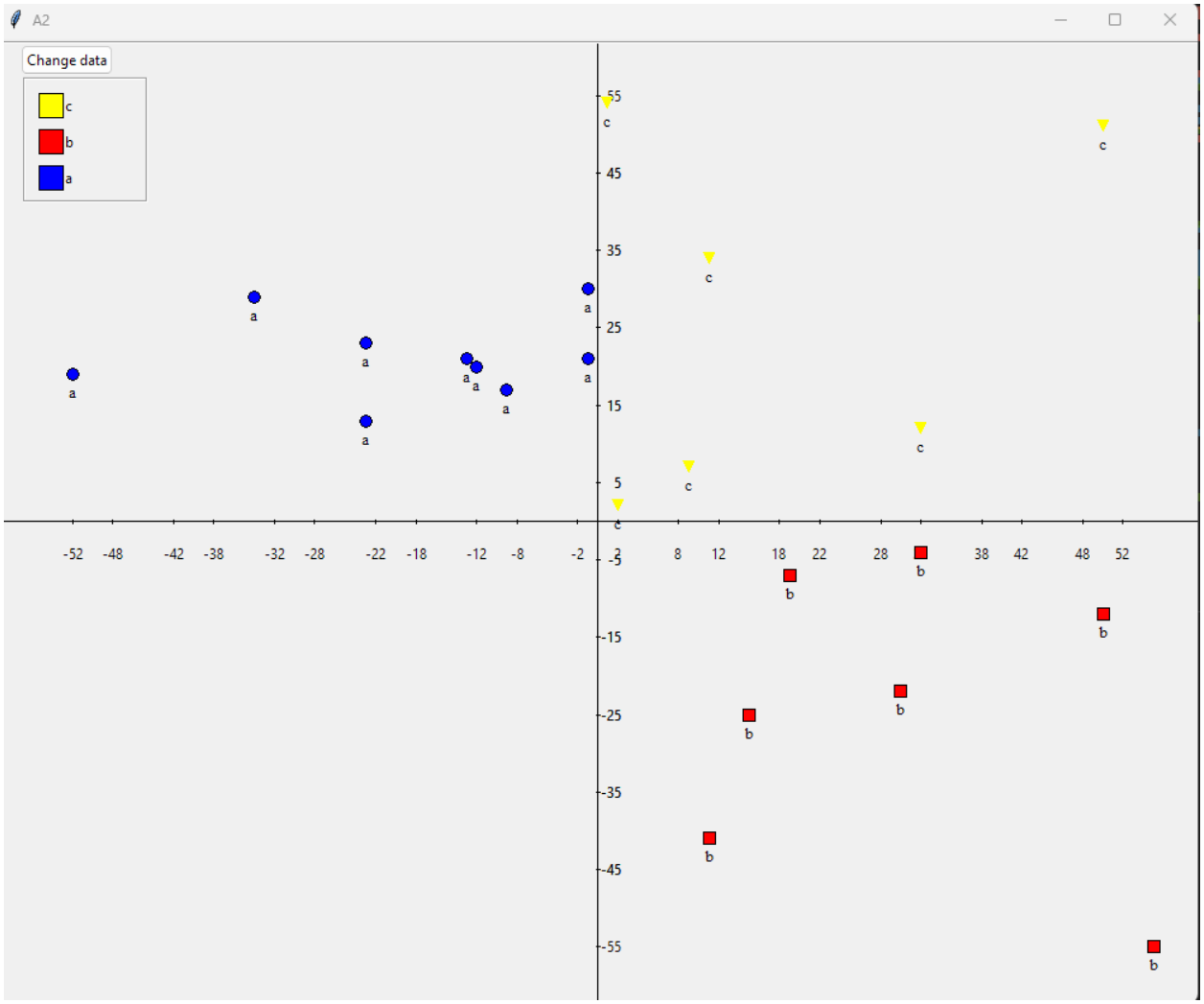
Information density (we strive for a balance between maximizing information density and avoiding cognitive overload)

It is about as **dense** as it can get based on the given data. One **cognitive load** present is that if the viewer wants to see the performance of a specific individual without knowing their age, they would have to manually scan through the X-axis in search of their name. Another **cognitive load** is that the viewer has to look at the legend to understand the placement of the points along the Y-axis.

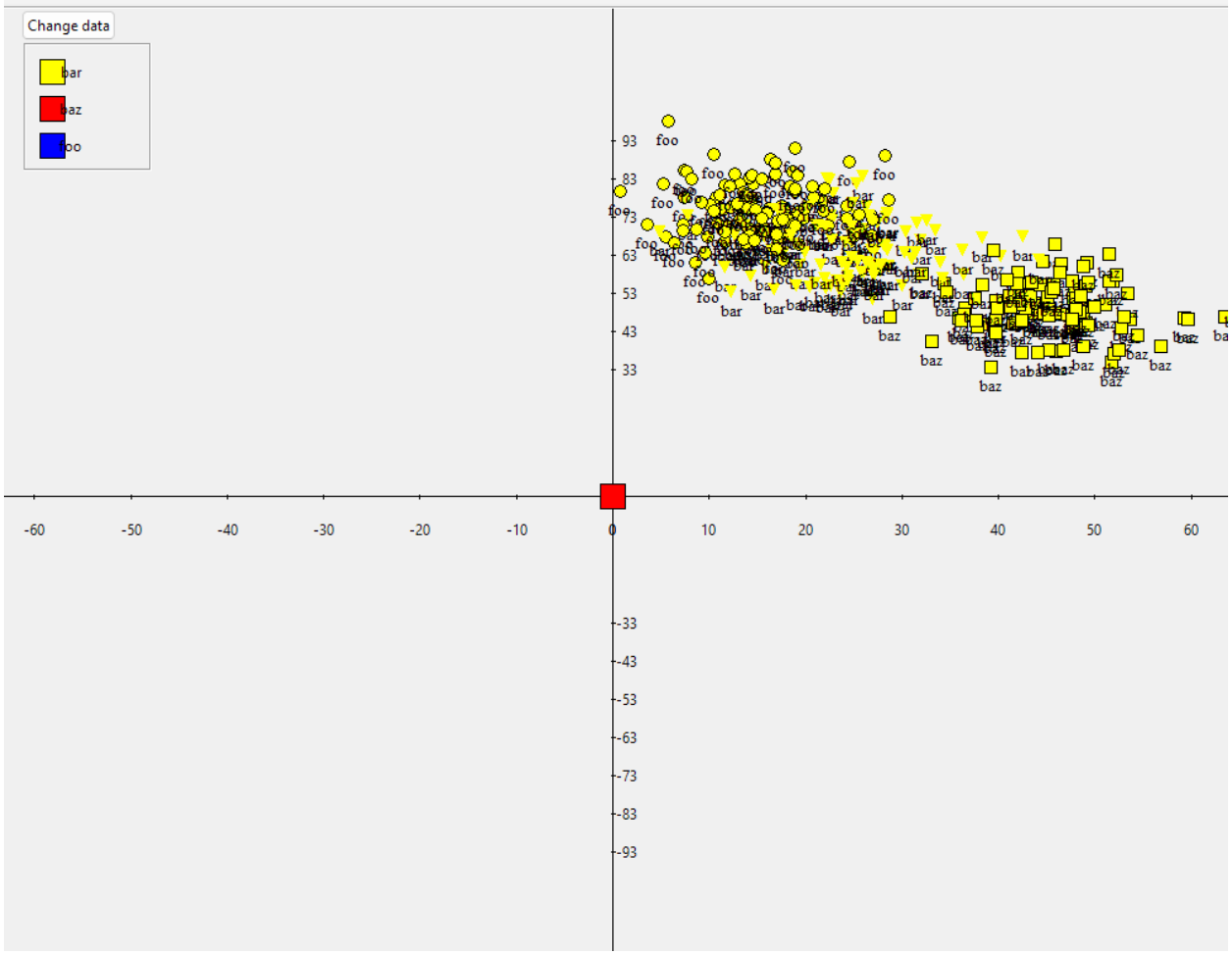


Task 3

The result from the implementation is shown below.



The figure above shows the points from ‘data1.csv’, where the points belonging to a, b, c are ‘together’ in the first, second and third quadrant. This dataset has both positive and negative values which were transformed accordingly.



In this figure the result of ‘data2.csv’ is shown. There are a lot more data points in this dataset which are easy to see, but what is also noticeable is that there are ‘clumped’ together, indicating that scatter plot might not be the best visualization tool for this kind of data. There are also no negative values which end up wasting screen space.