

TNM111: Assignment 2

Adrian Szuter (adrsz812) & Sofia Lundström (soflu921)

1. Explain in detail the InfoVis Reference Model. What are the strengths of this model?

The InfoVis model describes how information is visualized by different steps. One of these steps is data transformation where raw data is transformed into data tables which is a more structured form. The reason for this is to easier map a visual form. Data transformation can be used to build but also improve data tables. The raw data can consist of spreadsheets or program output and becomes data tables with different variables and attributes. The second stage is visual mapping where mapping to a visual structure should reflect the table of data. The structured data is then assigned visual elements such as marks like points or graphical elements like colors. An effective mapping is easy to interpret and is an effective visualization where the user has a low cognitive load. For expressive mapping it should only represent the information given by the data, no extra variables or more dimensions than needed. After the visual mapping where the information has been visualized by using scatter plot or bar charts for example, the next step is view transformation. The goal is to improve usability by allowing dynamic adjustments such as location probes, viewpoint controls and distortion techniques. The final step is interaction from the user in data transformation, visual mapping and view transformation to achieve the best visualization of the information. The interaction allows the user to engage with the data to learn new information that was not obvious at first glance.

The fact that the model is divided into several steps is its main strength since this allows us to create a model/visualization in an incremental and controlled way. Another strength this model has is that it can be used for a lot of different visualization techniques and data types. The model works for both a bar chart and a more complex visualization in more dimensions.

2. What kind of interactions are 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 are a form of dynamic queries which make it possible to filter data in different ways. Usually they are used for, as the name suggests, range filtering, this can for example be price or time period filtering. When dealing with range sliders it can often be beneficial to encode information about the dataset, into the slider itself. For example, you can include the distribution of the dataset. This makes it easier for the user to explore and experiment with the dataset.

Pros:

- It is considered to be a natural kind of interaction
- Easy to use and work with
- Makes it easy to explore the dataset with

Cons:

- Interface needs space on the screen
- It can be tricky to only display a single value (boolean values)

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?

This is called Brushing & Linking and it is commonly used in InfoVis. It allows for a consistent display of the dataset over multiple views. This in turn helps us get a better understanding of

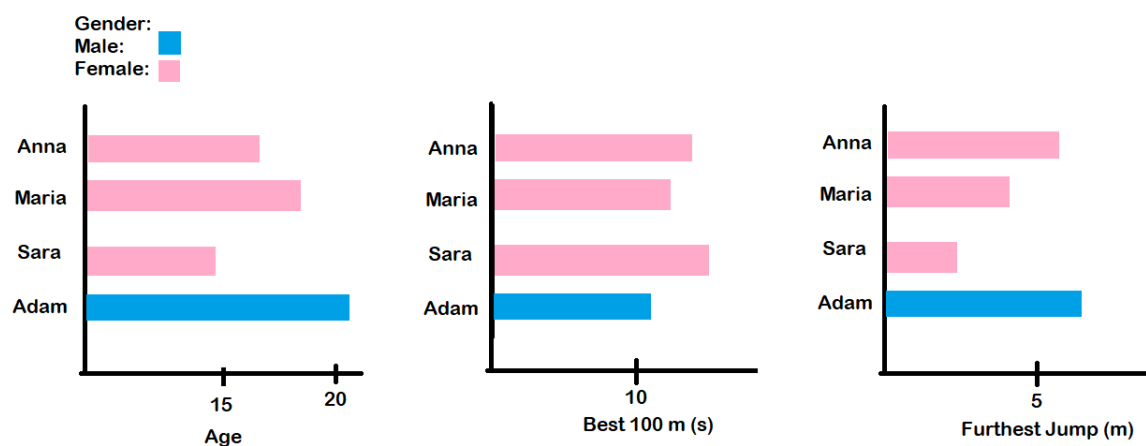
the dataset by making it possible to view relationships between different dimensions at the same time. It also helps with search and exploration of the dataset for the same reason as stated above.

Task 2:

Since all attributes are represented in the visualizations the expressiveness is good. For effectiveness, the bar charts are easy to compare for each individual. It also helps that the colors chosen for the gender are very distinct. However, a scatter plot or parallel coordinates would be able to better showcase the relationships between different attributes and thus make it easier to see correlations.

The visualization does a good job representing nominal attributes such as the name and sex with labels and color coding. Ordinal data is also represented as the age which has a bar chart. Quantitative attributes are clearly distinguished in separate charts and align with the visualizations goals.

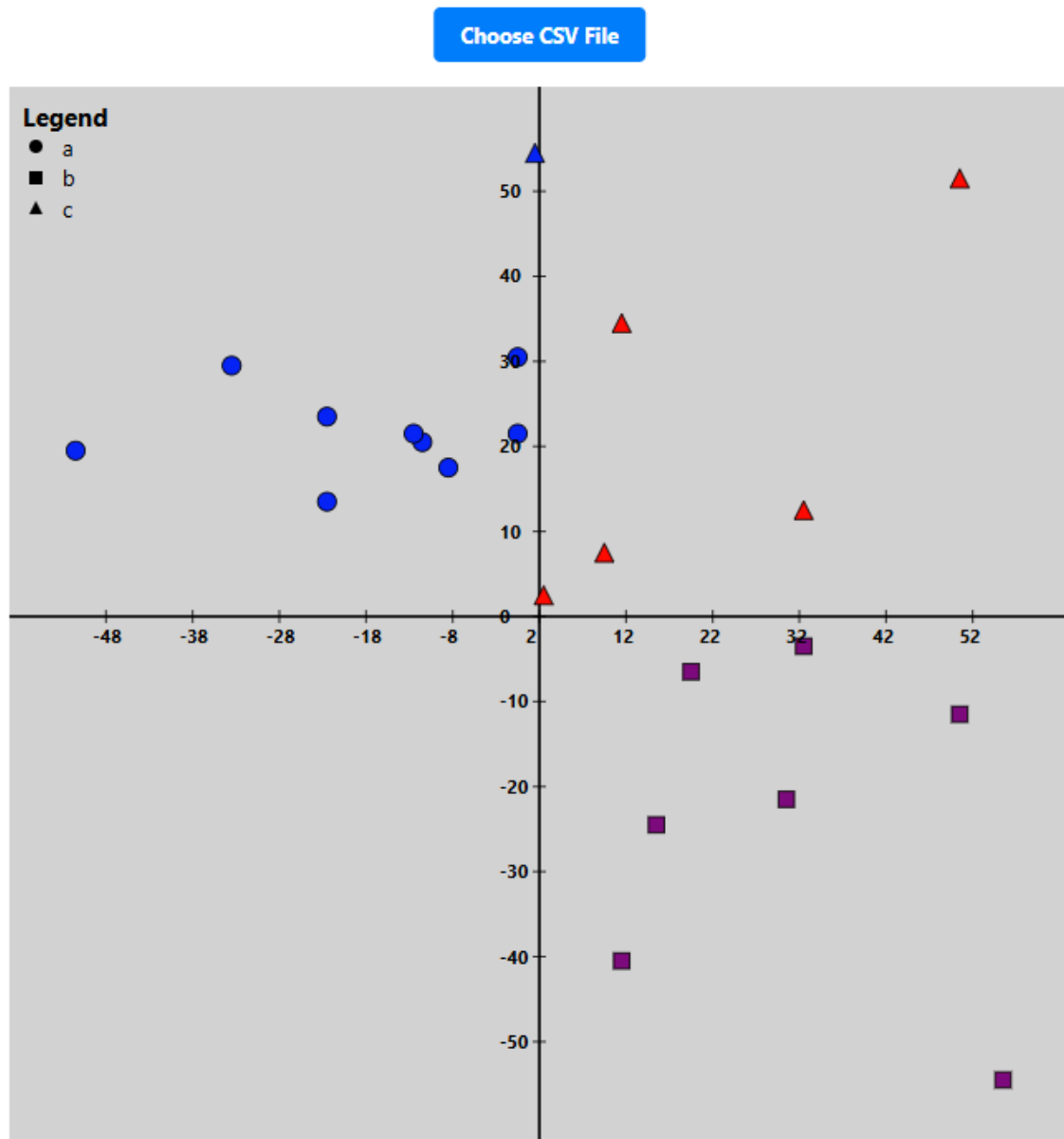
There is a good balance between information density and cognitive load since we have three different bar charts to avoid clutter and using color to make gender distinct.



Task 3:

From the implemented scatter plot some conclusions can be made about the data. From the scatter plot for data2 we can see that the values of the same label are closely related to each other and that they can be grouped together. We could see this thanks to the different quadrants having different colors and the different labels having different shapes which thus also allowed us to see that the different groups are distinct from each other. It is important to note that our axes do not have a positive and negative side since the center of the axes are not at origo. Another conclusion that can be made about data2 is that the category bar's points are usually between the points for foo and baz. Before any point has been selected in data1, we can observe how the different categories almost are divided by the four quadrants which can be seen by the different shapes for each category and the different colors for each quadrant. The axes for data1 are almost in origo so it is possible to see a positive and negative side to both x and y. A conclusion can be made that for the category a, the x value is usually negative while the y value is positive. For category b you can tell that most of the x

and y values are positive since they are in the upper right quadrant. Category c has x values that are positive and y values that are negative.



Choose CSV File

