TNM111 - Assignment 2

Wilhelm Björkström (wilbj602) & Marcus Sjöblom (marsj402), MT4

Task 1

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

The InfoVis Model is a model that describes the process of how information visualizations are produced and how they continuously adapt to changes during the process. Figure 1 is an illustration of all the different stages of the process, from a collection of raw data to a visualization where conclusions and trends can be observed in a structured manner.

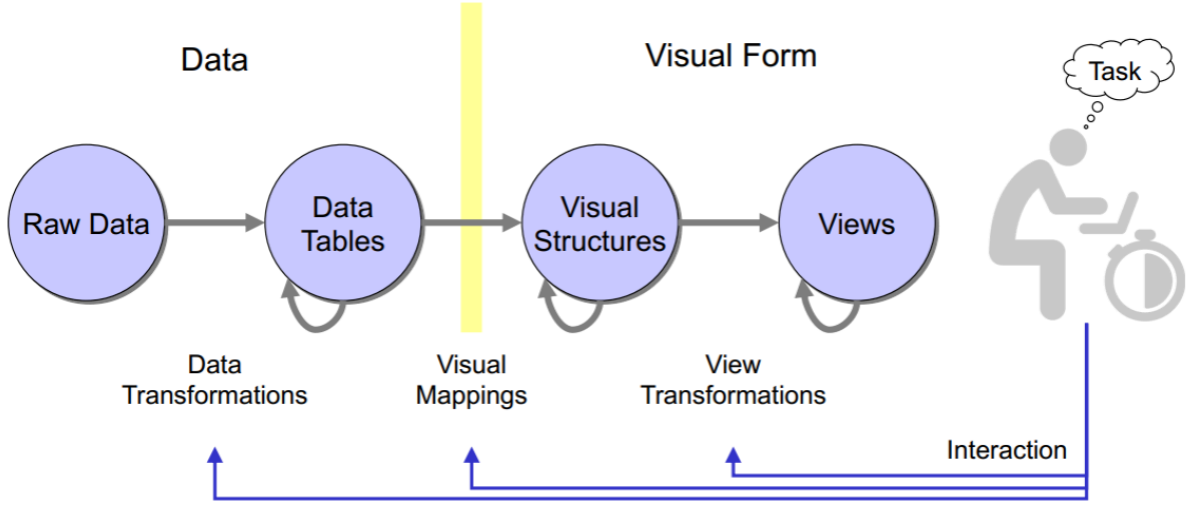


Figure 1: The InfoVis Reference Model pipeline.

Data Transformation

Visualization is typically built upon a foundation of raw data, and this foundation can take various forms. Common sources of raw data include survey results, sensor readings, and temperature data. These types of data lend themselves well for visualization. The raw data is often collected in an unsorted manner which means that some sort of sorting and structuring is required, which is exactly what happens in the next step, data transformation.

Data transformation consists of building and structuring data tables to make the data readable but also accessible to be used by different applications like visualizations. An example data table can be examined in Figure 2. Structures can vary but data tables often consist of Objects (Grey), Variables (green), values (magenta) and metadata (brown). Variables can be of type Nominal (red), ordinal (green) and quantitative (blue). All of these contain different sorts of information and should therefore be ordered correctly. A lot of errors when creating data tables stem from raw data being labeled as incorrect type or straight up missing. Data also comes with different dimensionalities which is good to keep in account when using it for visualizations. They are called univariate data (1D), bivariate data (2D), trivariate data (3D) and multivariate data (≥4D).

Model	V-Type	120d	T6
Manufac.	N	BMW	Volvo
Series	N	1	XC60
ID	O	36983	43832
HP	Q	190	350

Figure 2: A possible data table.

Visual Mappings

After the data has been transformed to a structural manner it is time to visualize the data. This is done in the Visual Mappings part of the InfoVis Model pipeline. The aim is to visually structure the data in a manner which reflects the structure of the table. A simple example of this type of work is drawing out points in a coordinate system from a table with x- and y-values. The core idea is thereby to show a visual representation of the input data as simple, without unnecessary visual aspects, for the user.

The process of finding a good way of implementing Visual Mappings is not straightforward and requires planning. Focus should thereby be put in to create a good visual perception for the user. Aspects such as visual structures in the plot, techniques to show information in a compact and efficient way and much more. Summarizing the step, it is a critical step for informative visualization requiring good planning and structure.

View Transformations

The View Transformations aspect of the pipeline is a way of giving the user the possibility to modify and show more of the visual mapping. This can be done with many different methods, like brushing and linking. In other words, View transformations can be viewed as the “refining” step in the visualization. By zooming, cutting out unnecessary information and moving angles, the information can be more clearly represented by users.

Interaction

Interaction is one of the most valuable aspects of InfoVis. To be able to actively change and manipulate the data transformations, visual mappings and view transformations in real time provides a lot of opportunities for the user to find and the sought after knowledge. Interaction with the data transformations can revolve around selecting different kinds of data, filtering the available data. Interacting with the visual mapping can present the viewer with different relations between data by simply changing labels. Through interaction with view transformations, specific areas can be highlighted in a visualization. Employing various techniques, such as distortion, further enhances the viewing experience. These methods contribute to an improved and more nuanced understanding of the presented data.

Strength of the model

The InfoVis Reference model strength is its adaptability for very specific tasks. If a user has a certain goal in mind, very easily can the model change and iterate to show what the user is looking for. The model is also very straightforward and easy to understand.

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

Range Sliders manipulate the raw data, filtering out the unnecessary data decided by the slider. While being a relatively simple addition to a visualization, it can change the content by quite a lot and give opportunity to more effective visualizations. A slider can be extended to have two anchor points, giving the ability to choose certain regions of the data to view and analyze.

Possible improvements when implementing a range slider is to have more feedback about the data set directly connected to the slider. A simple slider only showcasing the range and the slider does not give any information about the dataset. it is possible to showcase what range the data set is in (if the slider is not normalized to it) by using colors in the slider. A different change could also be to show how much of the data set is in a certain range of the slider, like seen in Figure 3. There are many more different ways of improving the slider, and it all comes down to what the end visualization goal is.

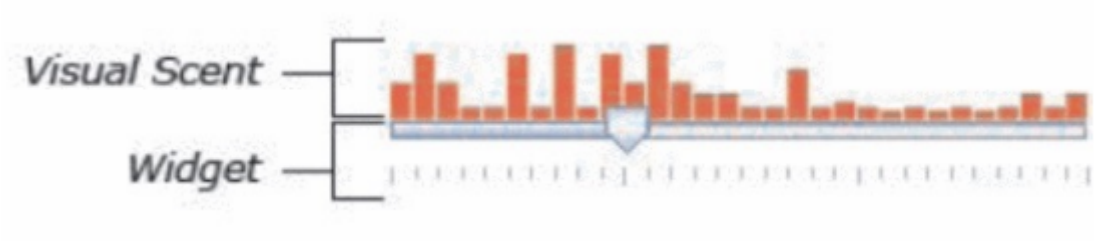


Figure 3: An example of an improved range slider.

Pros

- Logical selection for numerical/measurable ranges.
- An innovative way of looking at the data for the user with lots of freedom in the visualization.
- Real-time feedback for the user when changing the range.
- Can be a compact solution for showcasing data.

Cons

- May be inconvenient and hard to read for large ranges.
- Requires good and clear labels.
- Most suited for numerical ranges.
- Requires space from the UI.
- Classes must be able to sort in a range.

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?

Brushing and linking

Brushing

Involves selecting an item by reflecting over a region or simply clicking on specific data points. After the data points have been selected they are visually highlighted.

Linking

Once the data points have been brushed they are synchronized in different views of the same data. The data points will therefore be highlighted on every view showcasing the data.

Advantages given by the method

- It is much easier to study a dataset from different views to reveal patterns or certain relationships between the data.
- If there are a lot of windows showcasing the data, all at once, it can be really hard for the user to identify what data points are the same.
- Different types of methods for showcasing data can be connected to the same data making it easier to explore and understand the data.
- Multidimensional data can be divided into different visualizations (for example a number of 2D plots) making it easier to analyze data across multiple dimensions simultaneously and still have the multidimensional connection.
- The user may choose to study the data from their specific preference and still be able to see the changes across all visualizations of the data.

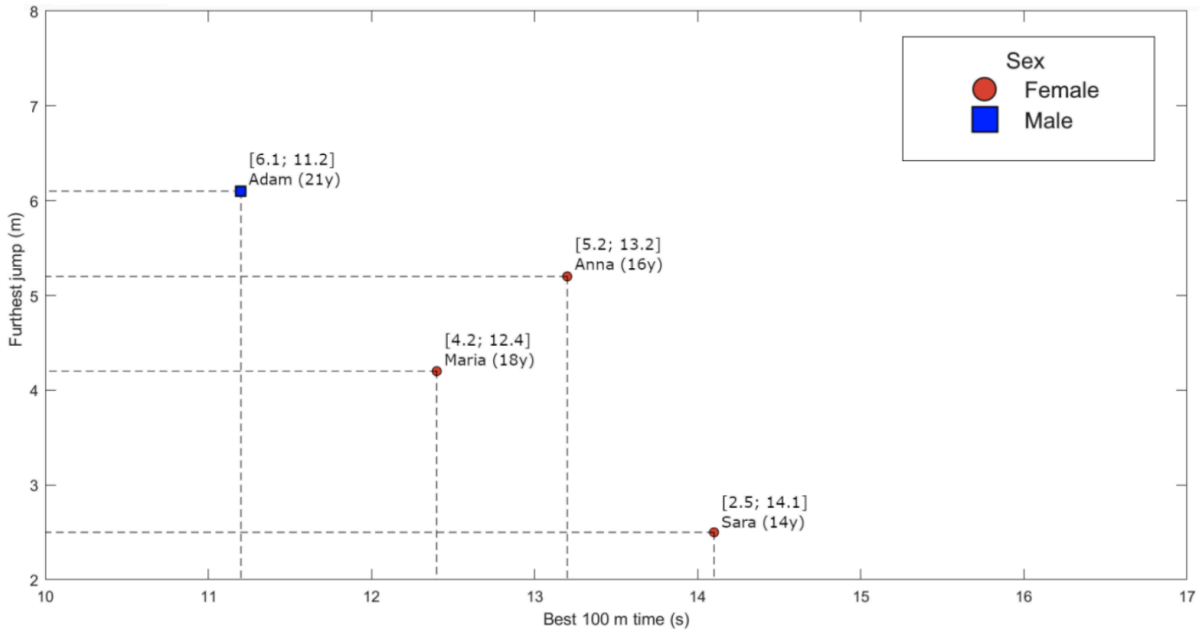
Task 2

Sketch a suitable visual encoding for the table below. The sketch should include all of the items and attributes for that table as a static visual encoding: do not assume any interaction will take place. You may use a single view or multiple views. You can sketch it using pen and paper, or any sketching software of your choice. Analyze the sketch, considering whether it is suitable according to the alignment between chosen visual features:

- Attribute types/values (in terms of expressiveness and effectiveness).
- Semantics of the attributes (an important part of data abstraction).
- Information density (we strive for a balance between maximizing information density and avoiding cognitive overload).

Name	Age	Best 100 m (s)	Furthest Jump (m)	Sex
Anna	16	13.2	5.2	Female
Maria	18	12.4	4.2	Female
Sara	14	14.1	2.5	Female
Adam	21	11.2	6.1	Male

Result:



Task 3

In this assignment, you will implement a scatter plot visualization, but without the usage of higher-level libraries to understand how the visualization is created from scratch. You are therefore not allowed to use any visualization libraries for this assignment.

You can use whichever programming language you prefer (such as HTML/JavaScript or Python), but without using visualization libraries such as D3.js or Matplotlib. However, you are allowed to use libraries to read CSV files and process data (such as pandas or csvreader). You are also allowed to use libraries such as Tkinter to draw and render GUI elements (buttons, shapes/colors for the data points, and straight lines for axes and grids).

Your application should be able to load and visualize the data sets found in the “Assignment 2” folder on Lisam. Your visualization tool should at least be able to:

- Draw the x- and y-axis and the ticks and tick values.
- Display a legend that shows the categorical information.
- Display the categorical information of the data points by using different shapes to represent the points.
- Display the data points correctly for the axes.
- Set the value range automatically based on the data values present in the data set.

After implementing the basic scatter plot, you should also add the following two features when interacting with a data point:

- When left-clicking with the mouse on a data point, a new grid system will be used where the selected point will become the new origin. The other points should get distinct colors depending on which quadrants they are located in. This new grid system is deactivated when the user leftclicks on the selected point again. Do not forget to mark the selected point somehow, e.g., stroke or highlight around the shape.
- When right-clicking with the mouse (or by using any other interaction of your choice, e.g., ctrl+leftclick) on a data point, the nearest five geometrically neighboring points,

based on Euclidean distance, will be highlighted with a color of your choice. This feature will be deactivated when the user right-clicks on the selected point again.

Your submission should contain a zip file with your code, instructions on running it if needed, and a short reflection on how you interpret the data with your implemented scatter plot.

Result: