Literature Review to propose a Visual Design for an interactive Visualization

Conny Walchshofer

April 2019

Tabular data plays an important role in many different fields, such as biology or computer science. During a project, however, a single table can change over time. The aim of this document is to propose a visual design for an interactive visualization that allows a user to explore the evolution of a single table over time. It contains a quick literature review in order to find out how other scientists have approached similar research problems before.

Goal: Demonstrate your ability to design a novel visualization.

1 Structure of a Table

This section focuses on different structures of a data table, on how to choose an appropriate design layout as well as on how to deal with adding and removing columns or rows.

Starting with static tabular displays, design guidelines are described and quantitatively evaluated by Eisl et al. In particular, their design guidelines are developed by taking the cognitive load into consideration. Numerous pre-attentive attributes (e.g. color-coding, font-size, line-width,...) therefore play a main role in the perceived usefulness of the tabular data. Emerged from these finding, a guideline for static and perceptually optimised financial data can be retrieved from Figure 1. This guideline can be predominantly used for a structured data set. When dealing with semi-structured or unstructured data sets, this manual can no longer be entirely used. Transferring the given data set from the field of accounting to biological or computer scientific table data, more interaction in terms of re-arranging or re-ordering data for an effective and efficient analysis is required because not only presenting, but exploring the given data set is of main objective.

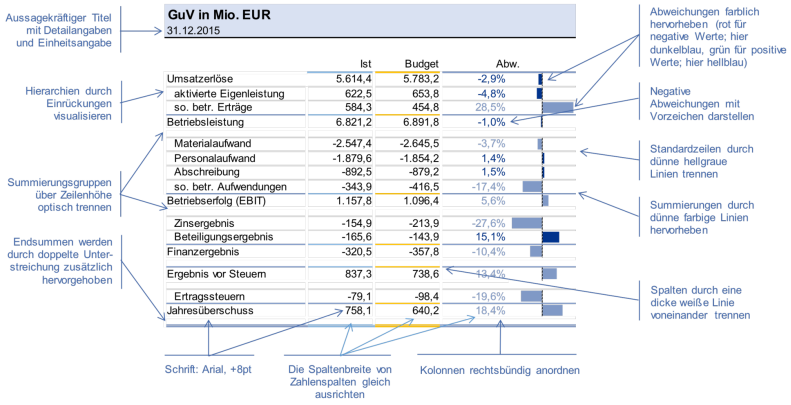


Figure 1: Design Guideline for a tabular display of the data in the fields of controlling, finance, and accounting [1]

> "Overview first, zoom and filter, then details on demand" [8]

The highly cited Shneiderman's mantra [8] emphasized the role of visualization in the knowledge discovery process. With the aim of combining analytical approaches with advanced visualization techniques, Keim grounded and expanded the mantra by the following statement: "Analyze first, show the Important, Zoom, filter and analyze further, Details on demand" [4].

According to scientific literature, three types of tabular data visualizations are famous [2]:

1. Overview Technique
2. Projection Technique
3. Tabular Technique

The term *overview technique* refers to high-level trends and patterns across attributes. Figure 2 shows exemplary used graphical displays for this technique. The first visualization represents a Scatter Plot, the second one a parallel coordinates plot, and the latter a star plot.

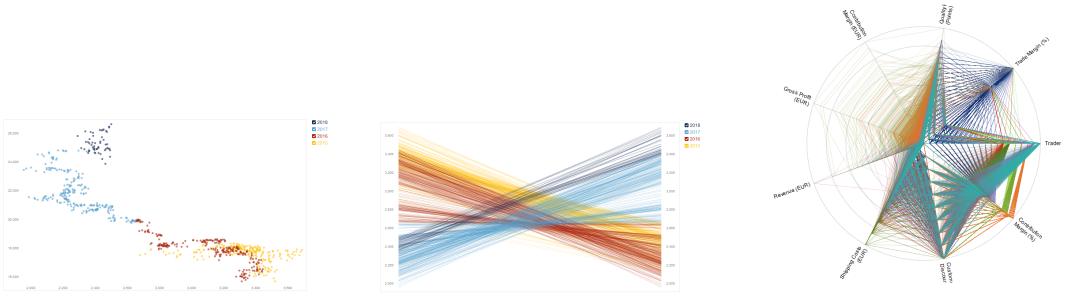


Figure 2: Overview Technique: Left: Scatter Plot; Middle: Parallel Coordinates; Right: Star Plot

The second technique, namely *projection technique* describes a lower-dimensional projection of a high-dimensional data set and the last technique, known as *tabular technique* deals for instance with matrix-reordering techniques or ordering a table by sorting. Deciding for using one technique does not exclude the opportunity for using a second or third technique at the same time. The major challenge in linking the three techniques is maintaining the proper relation. These approaches are name hybrid approaches [2].

It is of utmost importance to provide both, either an overview of the used data as well as a detailed view [8]. In particular when dealing with interactive tables, it is necessary to always provide information about the current position as well as past interactions (history).

Moreover, 12 requirements for dealing with an effective scalable and interactive table were published by Furmanova et al. [2]. The given list contains interaction techniques (e.g. sorting, filtering, and grouping) as well as more advanced ones (e.g. interactive refinement and visual feedback).

1. Encoding
2. Sorting
3. Filtering
4. Grouping
5. Aggregating
6. Multiform
7. Combining Columns

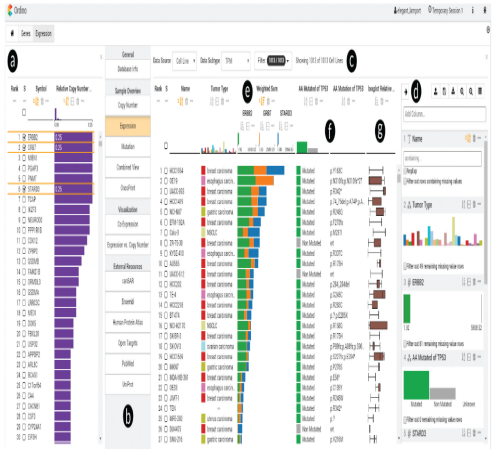
8. Transforming data
9. Bidirectional Matrix operations
10. Interactive Refinement and visual Feedback
11. Showing an overview of items
12. Showing details of items

1.1 General Layout

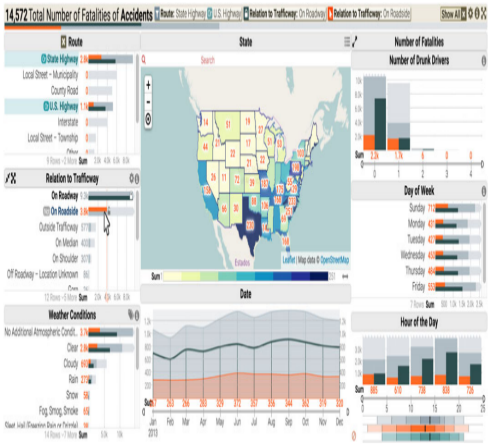
The following subsection displays different interactive table layouts for biological or computer scientific table data. Table 1 The displayed data range from geographical data containing information about car accidents (Keshif by Yalcin et al. (2018)) over population multivariate, tabular data (TACO by Niederer et al. (2017)) to genomic alteration and gene expression data of breast cancer cell lines (Ordino by Streit et al. (2919)). The selection only contains four possible representations of interactive analytic tools.

Furmanova (2017) [2]

Streit (2019) [9]



Yalcin (2018) [10]



Niederer (2018) [5]

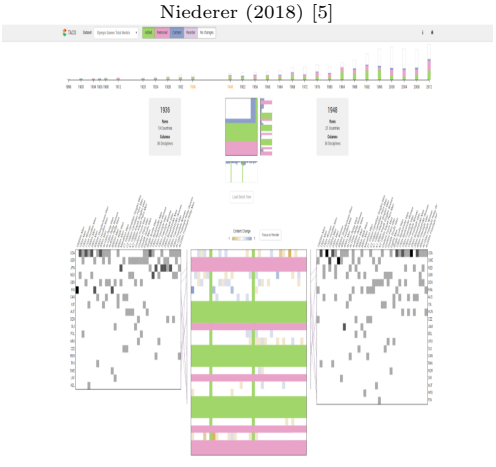


Table 1: Different exploration tools in order to analyze geographical, multivariate tabular data. The displayed screenshots make use of numerical indicators as well as (stacked) bar charts, line charts, and/or heatmaps. These tools particularly can be interactively used to explore the data.

1.2 Add/remove Columns and/or Rows

Adding complexity by appending or removing attributes/parameters in rows can be seen as a drill-down or drill-up features with hierarchical data according to Ross et al. [7]. In the publication *The Data Warehouse Toolkit*, he therefore points out to the feature as following:

> Drilling down in a data mart is nothing more than adding row headers from the dimension tables. Drilling up is removing row headers. We can drill down or up on attributes from more than one explicit hierarchy and with attributes that are part of no hierarchy. *[7] p.44*

2 Content of the Table

This section presents possibilities of displaying data in a tabular view. The display can either range from numerical indicators over sparklines in order to indicate a chronological trend to connecting line charts with the purpose of visually recognizing correlations. In particular, the content of cells can either be single data values (string or number) or multiple data values. Furthermore, using histrograms (bar charts) or box plots help to visualize cells that need to be summarized [9].

3 Order of the Columns and Rows

This section will show the latest research and different techniques used to generate further insights into the given data set by ordering and arranging columns and rows user-specifically. Because not only adding and removing columns and/or rows increase the complexity of a tabluar visualization, but reordering and therefore prioritizing columns/rows, supports the user in decision-making processes more effectively and efficiently.

In contrast to traditional tables, interactive ones allow to rank a table by a single or multiple columns. One of the most well-known example therefore is the so-called Table Lens, originally invented from Rao and Card in 1994 [6]. When deciding to chose another order of columns or rows, it is important to provide visual feedback to the user in terms of animated transitions. Smooth transitions keep the cognitive load low by showing the changes resulting from using filter, sorting, or aggregational operations.

4 Additional Considerations

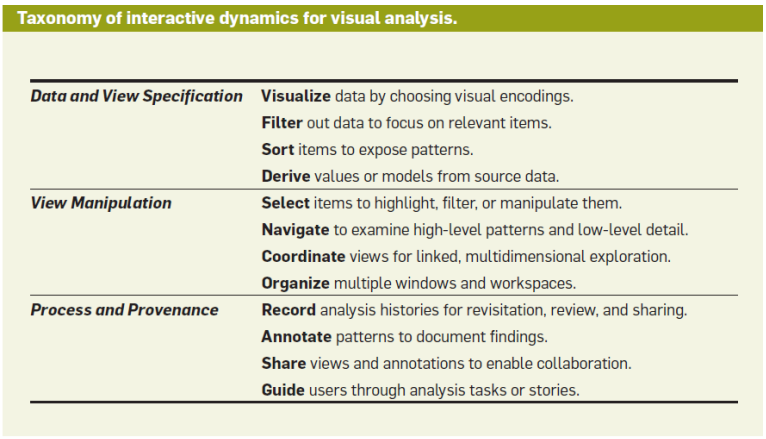
Most of the work in this section focuses on different data types and large amount of data. Additionally, it highlights difficulties when dealing with more columns than rows and vice versa.

4.1 Data Type

Dealing with heterogeneous data that comprise categorical, numerical, and textual vectors can be challenging due to flexible aggregating data subsets in both, columns and rows [2]. In contrast, using hierarchical data simplifies operations such as filtering, sorting, and flexible encoding of the subsets.

4.2 Big Data

Using the buzzword Big Data evokes numerous diverse reactions and associations. With regard to interactive visualizations, it is evident that users cannot deal with large amounts of tabular data when using a static visualization. Therefore, interaction is the key to explore the given



Taxonomy of interactive dynamics for visual analysis.

Data and View Specification	**Visualize** data by choosing visual encodings.
	Filter out data to focus on relevant items.
	Sort items to expose patterns.
	Derive values or models from source data.
View Manipulation	**Select** items to highlight, filter, or manipulate them.
	Navigate to examine high-level patterns and low-level detail.
	Coordinate views for linked, multidimensional exploration.
	Organize multiple windows and workspaces.
Process and Provenance	**Record** analysis histories for revisitation, review, and sharing.
	Annotate patterns to document findings.
	Share views and annotations to enable collaboration.
	Guide users through analysis tasks or stories.

Figure 3: Taxonomy of interactive dynamics for visual analysis introduced by Heer and Shneiderman [3]

data set. Famous interaction techniques described within the taxonomy of interactive dynamics for visual analysis contain Filtering, Sorting, Navigate, as well as the usage of annotations with the objective of guiding the users through analysis tasks or stories [3]. Figure 3 displays the introduced taxonomy, categorized by data and View specification, view manipulation, and process and provenance in addition to the corresponding interaction techniques.

4.3 More Columns than Rows / More Rows than Columns

In particular, interactive table visualizations are designed to operate with a numerous rows where the full-content can be explored by using the interaction technique of scrolling or drilling-down. Dealing with large and complex tables, Taggle presents a promising technique for presenting and exploring the data. It notably deals with individual columns of either categorical, numerical data and/or homogeneous matrices [2].

5 Conclusion

The short literature review about the proposition of a visual design for an interactive visualization highlights important aspects in terms of visual design and interaction techniques supporting users to explore the given data set. By transferring the focus from only traditional numerical indicators (e.g. dashboards or reports) to interactive tools containing pre-attentive attributes and a large variation of graphical displays (e.g. bar charts, line charts, landmaps), provide the user to individually explore the data. By making use of interaction techniques, the identification of outliers or correlations support the user in decision-making processes. Moreover, the aspect of a high performance system need to be taken into account when dealing with different data types and Big Data.

References

[1] Eisl, C., Perkhofer, L., and Hofer, P. *Exzellenz im Reporting Design: Leitfaden für messbar bessere Berichte.* Haufe, 2018.

[2] Furmanova, K., Gratzl, S., Stitz, H., Zichner, T., Jaresova, M., Lex, A., and Streit, M. Taggle: Combining overview and details in tabular data visualizations.

[3] Heer, J., and Shneiderman, B. Interactive dynamics for visual analysis. *Communications of the ACM 55*, 4 (2012), 45.

[4] Keim, D. A. Visual analytics: Scope and challenges. *VIsual Data Mining*, 4404 (2008), 76–90.

[5] Niederer, C., Stitz, H., Hourieh, R., Grassinger, F., Aigner, W., and Streit, M. Taco: Visualizing changes in tables over time. *IEEE transactions on visualization and computer graphics 24*, 1 (2018), 677–686.

[6] Rao, R., and Card, S. K., Eds. *The table lens: merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information* (1994), Citeseer.

[7] Ross, M., and Kimball, R. *The data warehouse toolkit: The complete guide to dimensional modeling*. Wiley, Hoboken, N.J., 2013.

[8] Shneiderman, B. The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization*. Elsevier, 2003, pp. 364–371.

[9] Streit, M., Gratzl, S., Stitz, H., Wernitznig, A., Zichner, T., and Haslinger, C. Ordino: a visual cancer analysis tool for ranking and exploring genes, cell lines, and tissue samples. *Bioinformatics (Oxford, England)* (2019).

[10] Yalcin, M. A., Elmqvist, N., and Bederson, B. B. Keshif: Rapid and expressive tabular data exploration for novices. *IEEE transactions on visualization and computer graphics 24*, 8 (2018), 2339–2352.