

Taggle: Scaling Table Visualization through Aggregation

Katarina Furmanova*
Masaryk University

Miroslava Jaresova†
Masaryk University

Bikram Kawan‡
Johannes Kepler
University Linz
Alexander Lex**
University of Utah

Holger Stitz§
Johannes Kepler
University Linz
Marc Streit††
Johannes Kepler
University Linz

Martin Ennemoser¶
Johannes Kepler
University Linz

Samuel Gratzl||
Johannes Kepler
University Linz

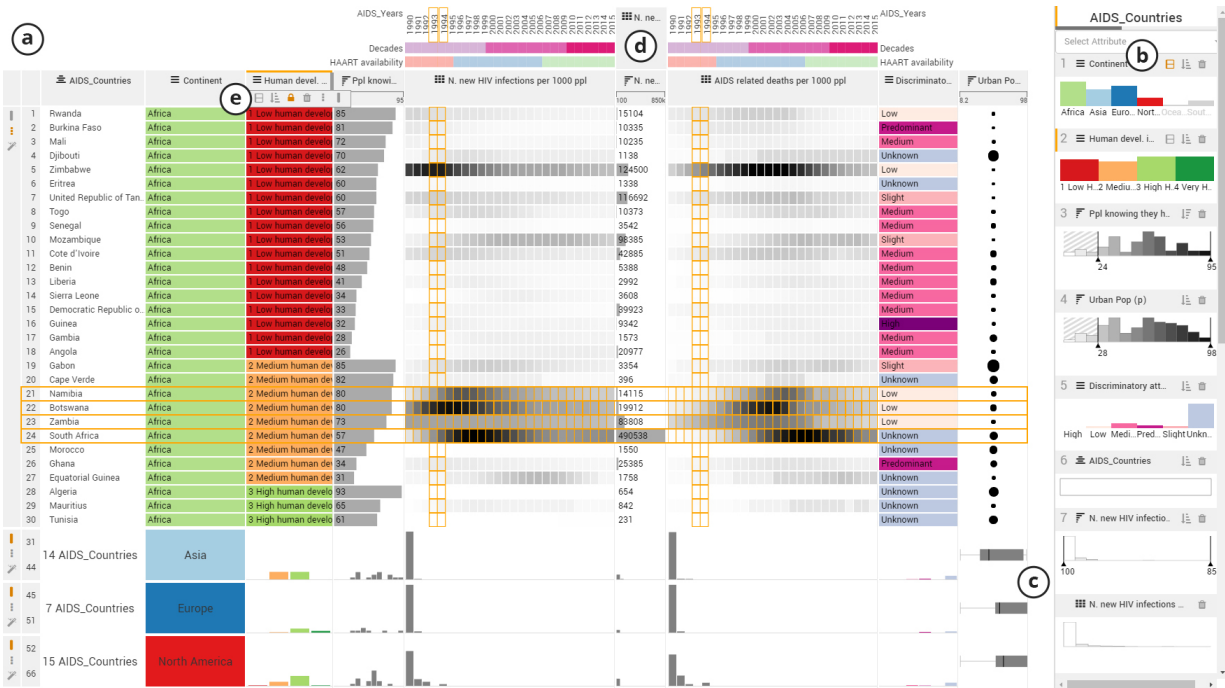


Figure 1: The Taggle interface consisting of a table view (a) and a data selection panel (b) showing a dataset on AIDS in several countries. The data selection panel consists of attribute filter views that allow users to filter out records by interacting with the histograms. The selected rows indicate the relationship between new infections and death rate over time. The rows of countries in Asia, Europe, and North America (c) have been aggregated to histograms and box-plots. Matrix columns can be aggregated using statistical metrics (d). The visualizations of attributes can be changed via a toolbar (e) for both aggregated and non-aggregated data subsets.

ABSTRACT

Tabular data visualizations are easy to understand and powerful at communicating patterns in datasets, especially when paired with interaction techniques such as sorting. In this work we present Taggle, a novel visualization technique for large and complex tables. We consider datasets that are composed of columns of categorical or numerical data, in addition to homogeneous matrices. The key contribution of Taggle is its ability to aggregate data subsets (rows and columns) on demand based on user-defined grouping rules.

Different visual representations for individual cells and aggregated subsets are available. The aggregation strategy is complemented by the ability to sort hierarchically and by rich data selection and filtering capabilities. We demonstrate the usefulness of Taggle using an AIDS dataset for 160 countries.

Keywords: Tabular data, multi-dimensional data, aggregation, hierarchical grouping and sorting.

1 INTRODUCTION

Visualization of tabular or multi-dimensional data is important in many application domains and a mainstay of visualization research. Not considering projection techniques, there are two types of tabular data visualization techniques: (1) those that position marks for each cell based on the cell value, such as parallel coordinates or scatterplot matrices, and (2) those that retain item position across columns and use marks within the cells, such as heatmaps or techniques like the Table Lens [5] and LineUp [2].

We find that while there is a rich body of work on the first class of visualization techniques, there is significantly fewer work on tabular layouts, with some notable exceptions [1–3, 5].

One significant drawback of tabular data visualizations is that by

*e-mail: furmanova@mail.muni.cz

†e-mail: jaresova@mail.muni.cz

‡e-mail: bikram.kawan@jku.at

§e-mail: holger.stitz@jku.at

¶e-mail: martin.ennemoser@jku.at

||e-mail: samuel.gratzl@jku.at

**e-mail: alex@sci.utah.edu

††e-mail: marc.streit@jku.at

drawing each item in a row, the available screen space is quickly exhausted, even for medium-sized datasets. A partial remedy are fish-eye techniques [2, 5], which, however, have other drawbacks. Our solution to this problem is the selective aggregation of rows encompassing diverse data types based on user driven selections. The goal of Taggle is to provide a high-level overview of large and heterogeneous tabular datasets, while also allowing users to drill-down to the individual items while preserving the context.

2 TAGGLE

Taggle is designed for exploration and presentation of tables that are a combination of both heterogeneous vectors and homogeneous matrices.

Vectors are columns in the table where all associated records are of the same type and semantics, such as the name, gender, or age column in a table of people. Taggle supports three types of attributes: categorical, numerical, and text.

Matrices are composed of vectors of the same semantics and data type. An example are blood pressure values collected at multiple time points in a table of people, or the GDP of countries over multiple years. While it would be possible to treat matrices as a list of vectors, Taggle handles them separately as the homogeneity of the data is an opportunity for aggregation. The columns in matrices can also be associated with vectors describing a common property of the column, such as the decade associated with a year.

2.1 Visualization and Interaction Design

The Taggle interface consists of two parts, as shown in Figure 1: (a) the main *Table View*, which visualizes the data, and (b) a *Data Selection Panel*, used for selecting the vectors and matrices to show in the Table view and for filtering.

For each vector or matrix present in the table view, the side panel shows a histogram of the data. The entries are grouped by shared keys. For example in the AIDS dataset shown in Figure 1, one key is “countries”, while another key is “years”. While vectors can have only one key, matrices have two keys.

The table view encodes each selected vector or matrix using a mark appropriate for the data type, such as, bar plots for numerical data, colored cells for categorical data, text, and heat maps for matrices. The visual encoding can be changed on-demand [4]. For example, the default bar plot can be replaced by a proportional symbol plot, if desired.

Sorting Taggle supports hierarchical sorting, where ties, as they are common when sorting based on a categorical vector, can be broken using other vectors. To separate the layout of the table from the order in which vectors are used for sorting, we use the order of histograms in the side panel to define the sorting: first, the rows are sorted by the first entry in the side panel, and in case of ties, the data from the second entry is used, etc.

Grouping and Aggregation Being able to stratify tables into meaningful groups is not only an important feature for structuring tabular data, but also an essential prerequisite for aggregation operations in Taggle. We leverage categorical vectors to group datasets, which allows us, for example, to group countries by continent. The grouping is defined using the sorting interface described above. To enable a more fine-grained control of what constitutes a group for the purpose of aggregation, we use the hierarchical sorting introduced previously. Grouping by multiple hierarchically sorted categorical attributes then results in the Cartesian product of their individual categories as groups. These groups can be leveraged to aggregate the data: instead of representing each row in the dataset with a row in the table, we can aggregate the grouped rows and represent them in a single row. For example, instead of bar plots for individual items, we show a histogram representing the whole group. The visual encoding of grouped cells can also be altered, to, for example,

replace a histogram with a box-plot (see Figure 1 (c)). Aggregation increases the scalability of Taggle significantly and also provides an overview for comparing groups.

Filtering Filtering is an important operation to drill down into a dataset. In Taggle, filters are defined by interacting with the histograms in the side panel, either by brushing a range in case of numerical data, or by selecting categories to filter out. Textual data is filtered by string matching.

3 IMPLEMENTATION

Taggle is implemented as a HTML5 web application based on the Phovea Platform¹ using TypeScript. The code is open source and available on Github². A demo version can be accessed at <https://taggle.caleydoapp.org>. We acquired the AIDS dataset present in the demo from UNAIDS³.

4 DISCUSSION AND CONCLUSION

In this poster we present Taggle, a scalable tabular data visualization technique for combining, filtering, and aggregating attributes of a heterogeneous dataset.

Taggle is unique among tabular data visualization techniques due to its ability to dynamically aggregate subsets of the table and due to its flexibility with respect to data types. We argue that Taggle is useful for both interactive exploration and presentation purposes.

During the design process we also investigated methods to automatically aggregate rows using various algorithms, with the goal of increasing scalability. We found, however, that users had difficulties understanding the dramatic changes in the visual representation that might occur without having explicitly triggered them.

In the future, we plan to integrate methods that allow us to aggregate and/or sort based on logical combinations of columns. For example, sorting could be defined as a function of multiple columns, instead of individual columns and tie breakers. Similarly, features for comparison, either for different sortings or groupings, or to compare data items are desirable.

ACKNOWLEDGMENTS

This work was supported in part by Boehringer Ingelheim Regional Center Vienna, the State of Upper Austria (FFG #851460), and the US National Institutes of Health (U01 CA198935).

REFERENCES

- [1] S. Gratzl, N. Gehlenborg, A. Lex, H. Pfister, and M. Streit. Domino: Extracting, Comparing, and Manipulating Subsets across Multiple Tabular Datasets. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '14)*, 20(12):2023–2032, 2014. doi: 10.1109/TVCG.2014.2346260
- [2] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit. LineUp: Visual Analysis of Multi-Attribute Rankings. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '13)*, 19(12):2277–2286, 2013. doi: 10.1109/TVCG.2013.173
- [3] Z. Gu, R. Eils, and M. Schlesner. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32(18):2847–2849, Sept. 2016. doi: 10.1093/bioinformatics/btw313
- [4] A. Lex, H.-J. Schulz, M. Streit, C. Partl, and D. Schmalstieg. Vis-Bricks: Multifunctional Visualization of Large, Inhomogeneous Data. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '11)*, 17(12):2291–2300, 2011. doi: 10.1109/TVCG.2011.250
- [5] R. Rao and S. K. Card. The table lens: merging graphical and symbolic representations in an interactive focus + context visualization for tabular information. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '94)*, pp. 318–322. ACM, 1994. doi: 10.1145/191666.191776

¹<http://phovea.caleydo.org>

²<https://github.com/Caleydo/taggle>

³<http://aidsinfo.unaids.org/>