Eurographics Conference on Visualization (EuroVis), Posters Track (2017)
A. Puig Puig and T. Isenberg (Editors)

Visualization Taxonomy based on the Specification of User's Goal and Data Dimensions

P. Kaur [†1] and B. König-Ries [‡ 1]

[1]Heinz-Nixdorf Endowed Chair for Distributed Information Systems, Department of Mathematics and Computer Science, Friedrich Schiller University, Jena, Germany

Abstract

In this paper we present a visual task taxonomy based on user goals derived via our extensive study, different data dimensions, and available visualization types. This taxonomy will be used in the creation of a visualization system that will assist users in the selection of more appropriate visualizations based on the specification of the task they want to perform and the available data dimensions.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [Computer Graphics]: User Interfaces—Theory and methods

Two important factors that guide visualization recommendation systems are properties of the data (in particular data dimensions) and the goal the user aims to achieve with the visualization. For example, a visualization that provides an overview of the dataset will be different from a visualization to represent data distribution. Instigating the user's information seeking goal is important in the visualization design process. However, these domain-independent information seeking goals (for example as provided by studies from [RM90,ZF98,AES05,SNHS13]) are very generic and therefore different visualizations can be used to represent them in a different way. This might not be sufficiently helpful for the user. Consider the goal to show distribution again. If a user is interested in viewing distribution over the spatial scale, then distribution maps are suitable. If he/she is interested in distribution among two variables, then XY scatterplots are suitable. Whereas if he/she is interested in distribution among multiple variables, then this might lead to further sub goal of either correlation or clustering. Correlation among multiple variables is represented by a scatterplot matrix or parallel coordinates. Clustering among multiple variables is represented by biplots or dendrograms and their variants. Moreover, one visualization can represent more than one user goal depending upon the type of the data. For example, a treemap can represent hierarchical levels with hierarchical data and can represent part-to-the whole composition relationships when provided with categorical data. The advent of different visualizations has made it possible now to classify the visualizations according to the data and user goal they represent. This is different from the situation many years back where the designer had to create specific visualization systems for specific goals.

This work is inspired by a mind map presented in [Abe06] http://tinyurl.com/AbelaWork where the author has presented the four main user goals which he has decomposed into different data domains to suggest the visualization. This mind map covers limited user goals and limited data domains. However, it provides us a clue of hierarchically decomposing the goals into subgoals and data domains for effective data visualization suggestion. In contrast to, e.g. [NK15], which focuses on geographical data, we aimed for a taxonomy that is not based on a specific data domain. Unlike [LPP*06], we also tried to avoid limitation to some specific set of visualizations like network visualizations. The typology presented by [BM13] helps users in formulating a scientific inquiry which leads to a generic user task. Our taxonomy identifies and classifies such generic user tasks or user goals and recommends the related visualization based on different specification of each task. For example, a specification for the composition task is different from that of the specification of the distribution or the comparison task. Each task has a different specification and is represented differently for different data domains. Once the visualization is presented, then a user may choose different operators like filter, zoom etc to get insight into the data and the graphics or can change his task and choose another visualization.

For the creation of this taxonomy, we have reviewed more than 5000 statistical/scientific data visualization images from different scientific journals. Through this review, we have generated a database of more than 75 different statistical/scientific data visualizations. Further this review has led us to understand 6 basic/general user goals or user intentions for visualizing data. These goals can further vary into 9 subgoals. Representing these subgoals into different data dimensions generates different visualizations. For example, a user goal of distribution can be decomposed into clustering, correlation, and distribution itself. Here, the visualiza-

† www.fusion.cs.uni-jena.de/fusion/members/pawandeep-kaur/
‡ www.fusion.cs.uni-jena.de/fusion/members/birgitta-konig-ries/

DOI: 10.2312/eurp.20171161

tions for clustering of two or more variables will be different. The visualization for correlation among one, two and more variable will be different. The visualization for the distribution of one, two and more variables, will be different. The visualization for the spatial and temporal variables will be different. Hence all these different visualizations represent different facets of one main goal, i.e Distribution. To further clarify this point, let us look at the example of network visualizations. Someone interested in understanding the network might want to know the hierarchy that exists within the network or might be interested in the flow that exists within the network. The visualization for representing the hierarchy is different from the visualizations for representing the flow. Furthermore, the visualizations that represents the flow into temporal, non-temporal and spatial domains are respectively different.

Due to the limitation of the page size, we have provided the taxonomy as a list view in Figure 1. Its full version is available at http://tinyurl.com/visTaxom . It has 6 main goals:

1. **Network** is either in the form of Hierarchy or a Flow. Flow can be seen on the temporal, non-temporal and spatial scales
2. **Range** is to visually understand the extent to which the variation is possible
3. **Comparison** can be viewed between two or more temporal and non-temporal entities. Furthermore, the change in one entity over temporal or spatial scales can also be visually depicted under the goal of comparison
4. **Overview** of one variable and overview of multiple variables can be displayed according to different visualizations
5. **Composition** can be defined either as part-to-the-whole or part-to-the-part relationship. This further can be represented differently on temporal, non-temporal and spatial dimensions
6. **Distribution** can be viewed for one, two or multiple variables with different visualizations. Geospatial distribution can also be represented by different versions of distribution maps e.g. Choropleth Maps, Dot Map or Bubble Map. Distribution then leads to either depict correlation or clustering. Correlation can further be visualized for one, two or multiple variables. Whereas clustering can only happen between two or multiple variables

We will be using this taxonomy in our ongoing project on visualization recommendation for biodiversity data. This taxonomy will guide the selection of visualization once a user goal is provided to it. In our work, the user goal will be derived partially from the domain knowledgebase and from the user interaction. For a given dataset, the metadata and a user goal, this taxonomy will filter down the possible visualizations that specifically matches to the provided dataset, the metadata and the intended task. By implementing this taxonomy as an interactive system, it will enable user to explore the large space of visualizations based on different dimensions.

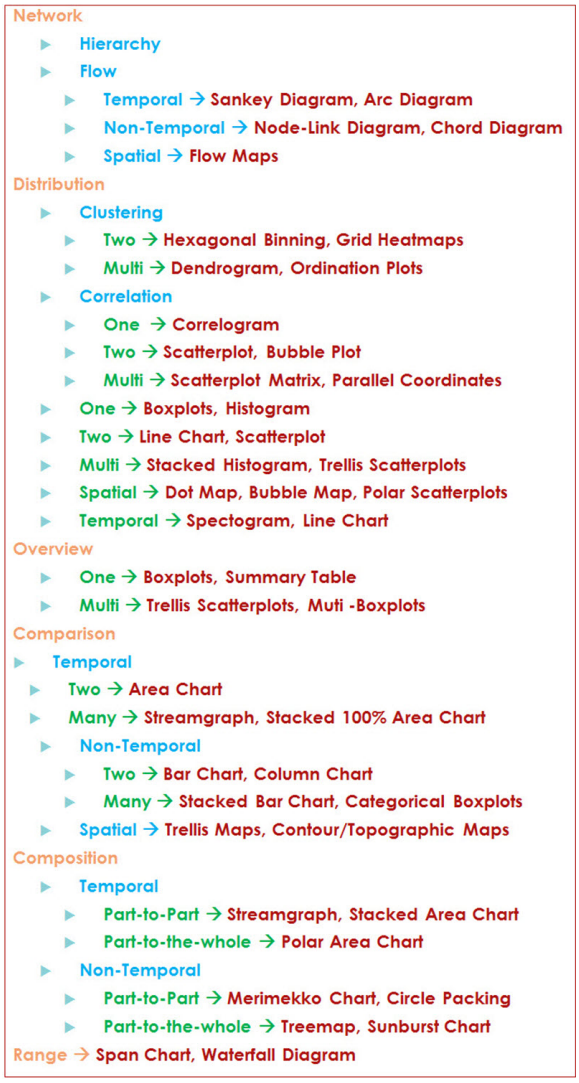

Figure 1: *A Visual Task Taxonomy*

Acknowledgment

The work has been funded by the DFG Priority Program 1374 "Infrastructure-Biodiversity-Exploratories" (KO 2209 / 12-2).

References

[Abe06] ABELA A.: Chart suggestions - a thought-starter, 2006. URL: http://tinyurl.com/AbelaWork/ . 1

[AES05] AMAR R., EAGAN J., STASKO J.: Low-level components of analytic activity in information visualization. In *Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization* (2005), pp. 111–117. doi:10.1109/INFOVIS.2005.24 . 1

[BM13] BREHMER M., MUNZNER T.: A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2376–2385. doi:10.1109/TVCG.2013.124 . 1

[LPP*06] LEE B., PLAISANT C., PARR C. S., FEKETE J.-D., HENRY N.: Task taxonomy for graph visualization. In *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization* (2006), pp. 1–5. doi:10.1145/1168149.1168168 . 1

[NK15] NUSRAT S., KOBOUROV S.: Visualizing cartograms: Goals and task taxonomy. *arXiv* (2015). 1

[RM90] ROTH S. F., MATTIS J.: Data characterization for intelligent graphics presentation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (1990), pp. 193–200. doi:10.1145/97243.97273 . 1

[SNHS13] SCHULZ H.-J., NOCKE T., HEITZLER M., SCHUMANN H.: A design space of visualization tasks. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (Dec. 2013), 2366–2375. doi:10.1109/TVCG.2013.120 . 1

[ZF98] ZHOU M. X., FEINER S. K.: Visual task characterization for automated visual discourse synthesis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (1998), pp. 392–399. doi:10.1145/274644.274698 . 1