

Drawing Distinctions: The Visualization of Classification

Martin J. Eppler

(mcm *institute*, University of St. Gallen, Switzerland
epplerm@gmail.com)

Peter Stoyko

(Canada School of Public Service, Ottawa, Canada)

Abstract: Classifying phenomena is a key step to building new knowledge, especially in the early stages of a research process. It can bring about multiple advantages and insights, such as overview and comparison. Yet it also poses several risks and constraints. Thankfully, challenges can be over-come by re-classifying items in a domain with alternative classification principles, which lead to new insights or perspectives, as well as highlight previously neglected considerations. This process can be supported by graphic representations. Visualizing the drawn (and redrawn) distinctions can make a classification accessible and versatile, which makes it easier to compare with other classifications. Visualizing classifications can augment the entire research process, including hypothesis formation, testing, interpretation and result reporting. There is no systematic overview of methods to represent (especially qualitative) classifications graphically. This paper fills that gap in the literature. We distinguish between four types of visual classifications, based on their differing ability to emphasize hierarchies or group relations. We label these four types as compilations, configurations, layers, and trees. We analyze their benefits for the research process and point out potential risks to consider when using visualization for classifications purposes in social science research.

Keywords: Classification, taxonomy, typology, graphic representation, research

Categories: A.1, E.1, M.4, M.9

1 Introduction

A key step of many social science research endeavors is the systematic classification of different items in a domain or area. This can be done deductively, in the form of a theory-based typology, or inductively, through the creation of an empirically grounded taxonomy. Examples abound. In qualitative research, classificatory frameworks are a widely used form (see [Marradi, 90], [Nowotny, 71]). In quantitative research, cluster analysis is a handy example of an approach that is explicitly supporting classification. Highly cited cases include Bloom's classification of educational objectives [Bloom, 56], Searle's classification of speech acts [Searle, 75], Shrivastava's classification of organizational learning systems [Shrivastava, 83], Holt's typology of consumption practices [Holt, 95], or Shneiderman's taxonomy of information visualizations [Shneiderman, 96]. Classification is so often integral to the way that social scientists perceive, interpret, and organize subject matter.

Many researchers take a step further and represent classifications using graphics. Graphics may show the hierarchy of grouping, emphasize underlying attributes of

items, or highlight relations among groups (see [Card, 99] or [Tufte, 90]). Many different graphical formats are employed. These range from simple two-by-two matrices and tree structures to complex grids and coordinate systems or may even include elaborate visual metaphors. Arguably the most famous graphical classification schema for research is Mendeleev's periodic table that systematically classifies the basic elements in chemistry in a u-shaped grid structure. (The table organizes elemental atoms by the number of protons in their nucleus.) Highly cited graphic classifications from the social sciences include Michael Porter's five-part classification of market elements [Porter, 80], Mintzberg's classification of organizational configurations [Mintzberg, 79], or Rogers' S-curve typology of innovation acceptance styles [Rogers, 95]. Despite the growing popularity of classification visuals, this method is not being used to its full potential. Moreover, it is often used inappropriately.

In this article, we argue that the visualization of classification is crucial for the advancement of social scientific research methodology as a discipline. Yet, despite its importance, visual classification is an insufficiently researched subject. We submit that different graphic forms provide significant advantages when drawing distinctions. For example, whereas a matrix or grid forces the researcher to combine two or more attributes to form a group or type, a Venn diagram enables the researcher to reflect on items that exhibit one, two, three or more attributes. As another example, a tree structure emphasizes hierarchy and sub-groups, while a pyramidal classification emphasizes how one class builds or depends on another. Researchers may not be aware of these and other properties of diagrams when developing or presenting their classification. As a consequence, they may not fully benefit from the versatility that visualizations can provide for the development of typologies or taxonomies. This is especially so when switching between different visual formats. The switching gives researchers the opportunity to extend their classifications or detect inconsistencies, among other benefits.

We present a systematic overview of the formats available for depicting classifications visually and highlighting the properties of each format. Our article is structured as follows. In the next section, we briefly define classification and describe its role as part of the scientific method. Section three offers an overview of visual forms for classification, including examples and guidelines for their use within the research process. Section four is dedicated to the risks and limitations of using visualization for classification-oriented research. The final section summarizes our contribution and draws some conclusions. In terms of limitations of this paper we highlight the fact that we focus mostly on qualitative classifications and their graphic rendering.

2 Classification as a Scientific Method

Classification, according to [Bailey, 94], designates the ordering of entities into groups or classes on the basis of their similarity. The research goals of classification are the minimization of within-group variance and maximization of between-group variance. In other words, classification aims for within-group homogeneity and between-group heterogeneity within a given domain. This task can be achieved in various ways. A *quantitative* approach usually consists of proceeding inductively and developing classes from empirical data. Cluster analysis is the most obvious example

because it is a statistical procedure that is explicitly designed to place cases into similar groupings. For instance, marketing researchers use cluster analysis to group consumers into various categories according to their spending habits, income levels, and lifestyle attributes (among other variables). These groupings can then be used to develop a formal taxonomy that can be further tested and applied practically. A *qualitative* approach may consist of a deductive, theory-based, “top-down” framework into which items are placed. The latter classification is often referred to as a *typology*, while the former is usually labeled as a *taxonomy*, although numerous studies use both terms loosely or even interchangeably.

There are several noteworthy qualifications about the distinction between typology and taxonomy. [Doty, 94: 232] define typologies as “conceptually derived interrelated sets of ideal types each of which represents a unique combination of the attributes that are believed to define the phenomenon.” A taxonomy, according to [Sanchez, 93], is simply an empirically derived grouping. While a typology may be oriented towards some normative end (such as market success in the case of a typology of business strategies), taxonomies tend to be descriptive. The goals of a taxonomy, according to [Chandra, 05:247], are to introduce structure into a body of facts, as well as to build a unified and homogeneous view of the domain of interest.

In contrast to mere categorization, classifications tend to be more systematic and stable, and lead to groups with clear boundaries and inclusion or exclusion rules [Jacob, 04: 528]. According to [Bowker, 99] classification refers to the spatio-temporal segmentation of the world. This segmentation, according to most scholars in the field, has to exhibit the following traits:

1. there are consistent, unique classificatory principles in operation;
2. the categories are mutually exclusive (non-overlapping);
3. the system is complete (all items can be placed in a group).

These are minimal requirements for a sound classification based on the original work of [Sneath, 73]. [Gregory, 06] goes further and states that a high quality classification must also be useful (i.e., it should aid analysis), contain meaningful and natural category labels and groupings, and consist of hierarchies that are appropriate (i.e., most important divisions are shown at the highest levels). In addition, the logic for placing elements in categories should be clear, as should be the characteristics that define each category. Likewise, [Meyer, 07: 28] mentions validity, simplicity, relevance, and difference as key traits of a typology. The validity of a typology depends on the unambiguous definition of categorization criteria (referred to as classification principles above). It requires that a typology be consistent with established theories in a field. The criterion of simplicity refers to the conciseness and comprehensibility of a classification. A classification, in other words, should contain a limited number of categories that are easy to distinguish. The criterion of relevance implies that only the most influential items should be included in a typology. According to Meyer, focusing on the important has a higher priority than being complete in the case of a typology. Difference, finally, refers to the fact that a typology should highlight the major differences in views, while paying less attention to issues on which more agreement exists. The focus should be on contradictory

assumptions and conflicting norms. In other words, only the aspects with a high discriminatory value should be considered. This is reminiscent of the definition of information as “a difference that makes a difference.”

There are various reasons why a researcher may want to develop a classification. The table below summarizes the main benefits that can result from a classification. The compilation is based on [Bailey, 94], [Bowker, 99], [Jacob 04], [Doty, 94], [Gregory, 06], [Sneath, 73] and [Meyer, 07]. We tentatively distinguish between simple benefits for description or representation and benefits for further analysis and evaluation of items or groups in a domain.

Classification benefits for representation	Classification benefits for analysis
Parsimonious description of items in a group; structured overview of a domain, reducing complexity of an area	Enabling new perspectives on a domain
Highlighting differences and similarities among items	Facilitating comparisons among items
Highlighting under-explored areas	Helping to understand relationships among items and groups
Presenting an exhaustive list of dimensions	Generating types as criteria for measurement and analysis
Inventory and management of types	Prediction of future instances of identified types

Table 1: A compilation of classification benefits

In spite of these numerous advantages, the classification approach has also been heavily criticized [Bowker, 98]. The primary risks associated with classifications are their tendency to: render concrete those things that are in short-term flux, long-term evolution, or poorly understood; and to reify things that are not so neat and tidy in the first place. Classifications may have an inertia that marginalizes alternative viewpoints and under-emphasize important attributes. [Chandra, 05] criticize the loss of information because of the generalization inherent in classifications. Classifications can also lead to framing effects. These effects limit creativity and “thinking outside of the box” (or rather “outside the tree structure”). A classification may lead to stereotypical thinking and false dichotomies.

How can researchers best profit from classification? How can they do so while avoiding drawbacks and pitfalls? One answer is to make use of the versatility and conciseness of visualization. In the next section, we explore this possibility and show how to use diagrams to develop and present research-based classifications. The adequate use of visualization can improve the quality of classifications and help to avoid frequent classification errors, such as: mixing levels of abstraction; switching classification principles; including unrelated items; attributing elements to the wrong group; or generating an incomplete classification system; to name but a few common shortcomings of research classification.

3 The Visualization of Classification

The visual depiction of groupings based on similarity can take on many forms and shapes. This ranges from traditional tree structures to elaborate logic or set diagrams (such as Venn or Euler diagrams). It can also include multi-dimensional scaling cubes and classifications based on parallel coordinate clusters. There are many two- and three-dimensional shapes that can be used fruitfully to make distinctions between groups and show how they differ. A key distinction is the one between *hierarchical differences* (based on levels of abstraction) and *relational aspects* (such as similarity or relative importance or size). Note that some visualization formats for classification are not so easily applied in the social sciences (such as cladograms that depict the evolution of an object, particularly a biological organism).

In our survey of visual classification formats, we distinguish between four kinds of visual representations found within the social sciences. These are based on the aforementioned distinction of whether (and to what degree) each visual emphasizes hierarchy or other relations among groups (see Figure 1).

- First, we have grouped those visualizations together that do not provide a hierarchic structure and do not emphasize relationships between elements or their groups. We label these simple groupings as *Compilations*, emphasizing the fact that they only provide a relatively unstructured compilation of items based on their similarity. Typical representatives of this group are the (two-by-two) matrix or a grid.
- *Configurations* is the label we have chosen for visualizations that do not provide a hierarchy, but emphasize (to varying degrees) relationships among groups or classes. A configuration is a particular arrangement of items that has an explicit structure which facilitates comparisons between items or groups, or shows some overall pattern among the groups. Typical forms in this group are Venn and Euler diagrams or structures that are based on a metaphor (such as a temple, ladder or puzzle) to convey an additional insight about the groups' relations.
- There are also groupings that only emphasize hierarchy which we label *Layers*. Simply put, layers group items hierarchically without showing additional relationships between the groups. The pyramid would be one simple example of visualizations in this group.
- The last set of classification graphics can be labeled as *Trees*. Trees show both hierarchy (in terms of levels of abstraction) and relationships, such as relative size of a group (as with a tree map) or closeness of groups (as with a dendrogram).

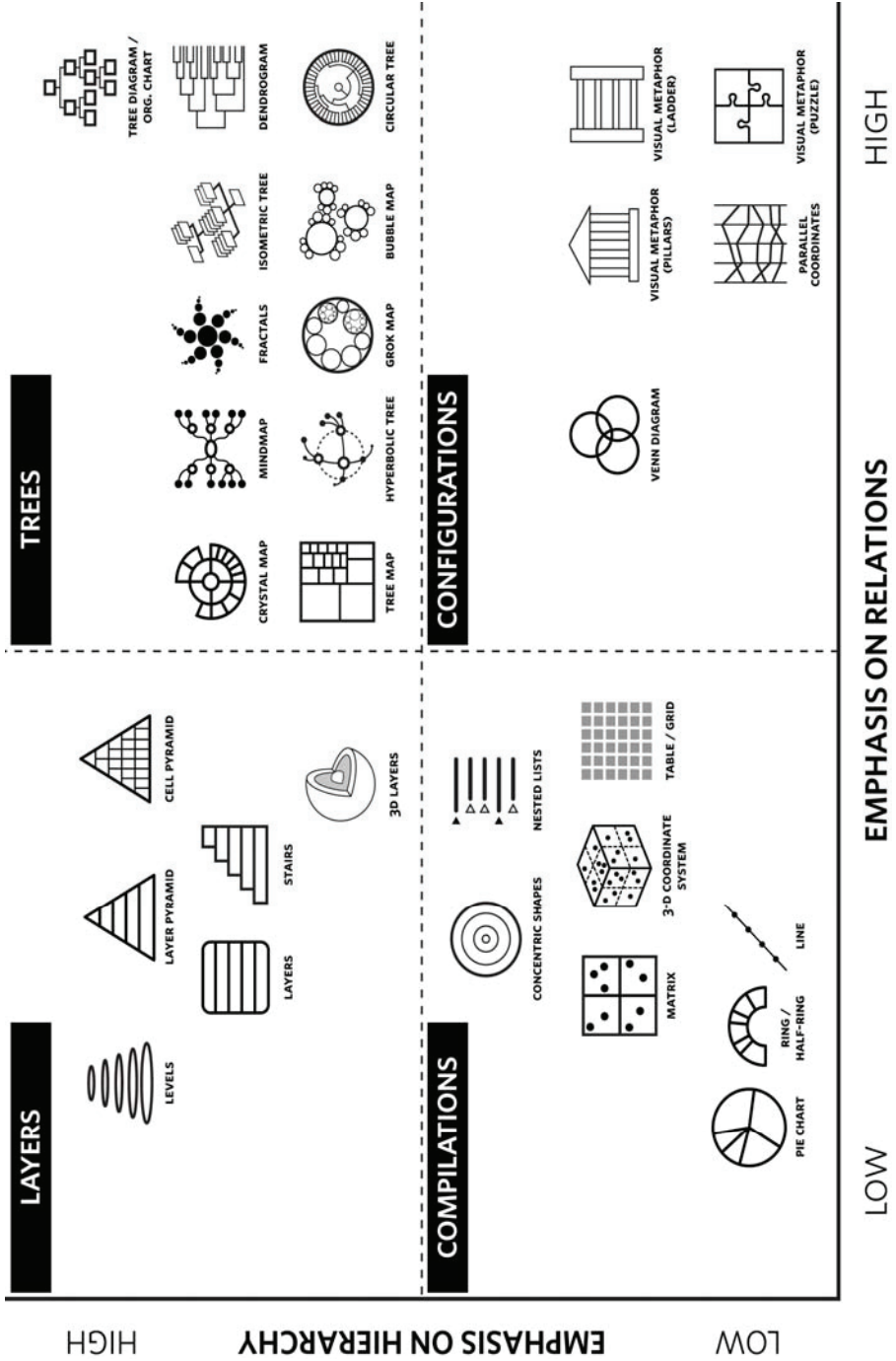


Figure 1: An overview on different graphic representations of classifications

The positioning of the different visualization formats in our framework is still tentative and ultimately depends on the actual use and adaptation of a graphic classification template to a particular research context. It should thus not be over interpreted or read with expectations of high accuracy. Nevertheless, the grouping provides an informative first segmentation of the various visual forms available to represent classifications graphically.

Deciding when to apply each of the four types of visual classifications requires a consideration of at least three factors:

The first factor is the *nature of the research* to be considered. For example, the analytical approach may be inductive or deductive, or the nature of the research observations may be qualitative or quantitative. Some formats (such as dendrograms) are useful for inductive, quantitative purposes, whereas others (such as matrices) may be more conducive to qualitative deductive work.

The second factor is the *type of classification scheme* involved. Certain research techniques lend themselves to particular classification schemes, and vice versa. A normative typology can, for example, be represented in a pyramid, while an empirical taxonomy could be first depicted in a multi-dimensional-scaling (MDS) cube.

The final factor is the choice of *visual technique* to be considered as it relates to the researcher's analytical goals. For example, the researcher may want to show an overview or focus on particular details.

The table below summarizes the key contributions of visualized classifications for each generic research step in the social science research process, and describes the benefits that can be achieved when visualizing classification in each phase.

Research Phase	Benefit of Visualizing Classifications
1. Exploratory	As in grounded theory, a visualized classification helps the researcher to detect and (re-) organize recurring patterns.
2. Error Detection	Visualized classifications can be used by different coders to detect and register errors.
3. Theory Development	Visualization helps the researcher determine whether a classification scheme complements the larger theory prior to the testing of that theory. It helps researchers draw connections between different theories and see differences and similarities.
4. Data Collection	Visualization can improve the collection process by helping researchers record observations in a way that complements human cognition.
5. Theory Testing	Identifying correspondence between data points and categories (and finding outliers) can be done more efficiently with visual methods that map or match data to a classification scheme.
6. Reporting	In many cases, visualization communicates a classification scheme (or research results involving the scheme) more efficiently and accurately

Table 2: The role of classification visualization in different research phases

4 Limitations and Risks of Visualizing Classifications

As with any analytical technique, there are limitations and risks associated with visualization in the classification context. Most of these risks are inherent to visualization, although a few are posed by common misuses or misunderstandings of visualization techniques. This list is by no means exhaustive. It is a list of some of the “marquee” risks and limitations that researchers ought to consider carefully.

Risk	Explanation
Reinforcing the reification fallacy/ fallacy of misplaced concreteness	Treating an abstract concept or classification scheme as a concrete, real entity is a type of fallacious reasoning. This is called the <i>fallacy of reification</i> or, using Whitehead's term [97], the <i>fallacy of misplaced concreteness</i> . Visualization can lead to greater risk of this fallacy occurring because it makes the intangible seem more tangible.
Reducing flexibility/locking in classifications	Visualization can potentially undermine the evolution of a classification scheme when the graphic does not easily accommodate adding, subtracting, and combining categories. A particular choice of graphic can also restrict expansion of a classification scheme because it is difficult to change the spatial dimensions (e.g., the length of a list) or spatial resolution (e.g., the density of points). A related problem is the premature visualization of a classification scheme that restricts flexibility before the classification is finalized.
The problem of visual forced-fitting.	When an established classification scheme encounters a novel instance that is not easily accommodated, it may be necessary to change the scheme. When change is necessary but a researcher insists on situating the new case within the inadequate scheme, it is called the <i>problem of forced fitting</i> . There is a case of forced fit that is specific to visualization: the situation whereby a novel instance seems to fit logically into a classification scheme but does not fit well within a visualization.
The problematic handling of miscellany in visuals	Many forms of visualization do not accommodate outliers very well, and may do much worse with large numbers of disparate outliers. In some cases, the choice of visual may result in the exclusion of outliers.
Trivialization and over-simplification	The choice for a specific graphic format may lead to neglecting certain aspects in the classification.
Misrepresentation of similarities or differences	A classification scheme may indicate something about the <i>degree</i> of within-group similarities and differences, or the degree of between-group similarities and differences. Yet, the corresponding visual may suggest something else about these similarities and differences.

Table 3: Potential risks of visualizing classifications

Considering these risks in the graphic rendering of a classification can consequently help to profit more fully from the aforementioned benefits.

5 Conclusion

In this paper, we have outlined the crucial role that graphic representations can play in the development, refinement, validation and communication of research-based classifications. We have given a tentative overview on possible graphic forms that can be used by researchers to construct or convey their taxonomies or typologies. We have also outlined how the functions of graphic representations of classification change according to the research step supported by them. Another contribution of our work has been the identification of key advantages and risks that come about through the use of visuals in this context.

One main implication of this paper is that we have articulated a re-classification imperative. That is to say, researchers should resist a premature closure in their choice of a graphic schema for their classification. They should keep their options open according to their ongoing observations and insights. Most importantly, we urge researchers to consider the benefits of visualization on the way that classifications are formed. While some visualizations invite the researcher to combine classification criteria (such as a matrix or a Venn diagram), they may also neglect to ask for different hierarchical levels. Conversely, some visualizations may emphasize a tree-like structure and neglect overlapping areas or groupings based on mixed criteria. The intent of this article is to sensitize researchers to these (often implicit) issues of graphic representation of classification. We acknowledge that our list of graphical forms is not exhaustive and is only a first step towards improving the way that classification graphics are used in the social sciences. We also acknowledge that our positioning of these forms along two axes (hierarchy and relationships) is a first, rough segmentation that must be subject to subsequent analyses and refinement.

In future research, we would consequently like to more systematically compare the different graphic forms for classifications (in terms of their respective functions and limitations) and describe their features more formally (such as their level of complexity, their biases, their affordances, etc.), as well as match them more accurately with different research scenarios or needs.

One feasible research method to do so could be performance experiments, where different groups of researchers need to make sense of the same data set, using differing visual templates for their classification tasks. The respective objective and self-reported measures could then be used to derive a performance- and task-based ranking of different visual schemas for classification.

References

- [Bailey, 94] Bailey, K.D.: "Typologies and Taxonomies: An Introduction to Classification Techniques", Thousand Oakes: Sage. (1994)
- [Bowker, 99] Bowker, G. C. and Star, S.L.: "Sorting Things Out, Classification And Its Consequences", Cambridge: MIT Press. (1999)

- [Card, 99] Card, S. K., Mackinlay, J. D. & Shneiderman, B. (Eds.): "Readings in information visualization: Using vision to think"; Morgan Kaufman. (1999)
- [Chandra, 05] Chandra, C., Tumanyan, A. "Supply chain system taxonomy: A framework and methodology"; *Human Systems Management*, 24(4), (2005), 245-258.
- [Doty, 94] Doty, D.H. Glick, W.H.: "Typologies as a unique form of theory building: towards improved understanding and modelling"; *Academy of Management Review* 19(2), (1994), 230–251.
- [Gregory, 06] Gregory, S.: "The nature of theory in information systems"; *MIS Quarterly*, 30 (3), (2006), 611-642.
- [Holt, 95] Holt, H.B.: "How consumers consume: A typology of consumption practices"; *Journal of Consumer Research*, 22(1), (1995), 1-17
- [Jacob, 04] Jacob, E. K.: "Classification and categorization: a difference that makes a difference"; *Library Trends*, 52(3), (2004), 515-540.
- [Marradi, 90] Marradi, A.: "Classification, Typology, Taxonomy"; *Quality & Quantity*, XXIV, 2 (May), (1990), 129-157.
- [Meyer 2007] Meyer, R.: "Mapping the Mind of the Strategist"; PhD thesis, Erasmus University Management School, The Netherlands.
- [Mintzberg, 79] Mintzberg, H. "The Structuring of Organizations"; New York: Prentice Hall. (1979)
- [Nowotny, 71] Nowotny, H.: "The Uses of Typological Procedures in Qualitative Macro-sociological Studies"; *Quality & Quantity*, VI, 1 (June), (1971), 3-37.
- [Porter, 80] Porter, M.: "Competitive Strategy"; New York: The Free Press. (1980)
- [Rogers, 95] Rogers, E.: "The Diffusion of Innovations"; Free Press. (1995)
- [Sanchez, 93] Sanchez, J.C.: "The long and thorny way to an organizational taxonomy"; *Organization Studies* 14(1), (1993), 73-92.
- [Shneiderman 96] Shneiderman, B.: "The Eyes have it: a task by data type taxonomy of information visualizations"; *Proc. IEEE Symposium on Visual Languages '96*, IEEE, Los Alamos, CA (September 1996), 336-343.
- [Sneath, 73] Sneath, P.H., Sokal, R.R.: "Numerical taxonomy: The principles and practice of numerical classification"; San Francisco: Freeman. (1973)
- [Searle, 75] Searle, J.: "A taxonomy of illocutionary acts. Language, Mind and Knowledge"; *Minnesota Studies in the Philosophy of Science*, (1975), 344-369.
- [Tufte, 90] Tufte, E.: "Envisioning Information"; Cheshire: Graphics Press. (1990)
- [Whitehead, 97] Whitehead, A. N.: "Science and the Modern World"; Free Press. (1997 /1925)