Chapter 1. The Fabric of Visualization*

G. ELISABETA MARAI,† University of Illinois at Chicago, USA
TORSTEN MÖLLER,‡ University of Vienna, Austria

2020

*Part of the book: Foundations of Data Visualization, M. Chen, H. Hauser, P. Rheingans, G. Scheuermann (editors), Mathematics and Visualization Series, Springer 2020
†gmarai@uic.edu
‡Torsten.Moeller@UniVie.ac.at

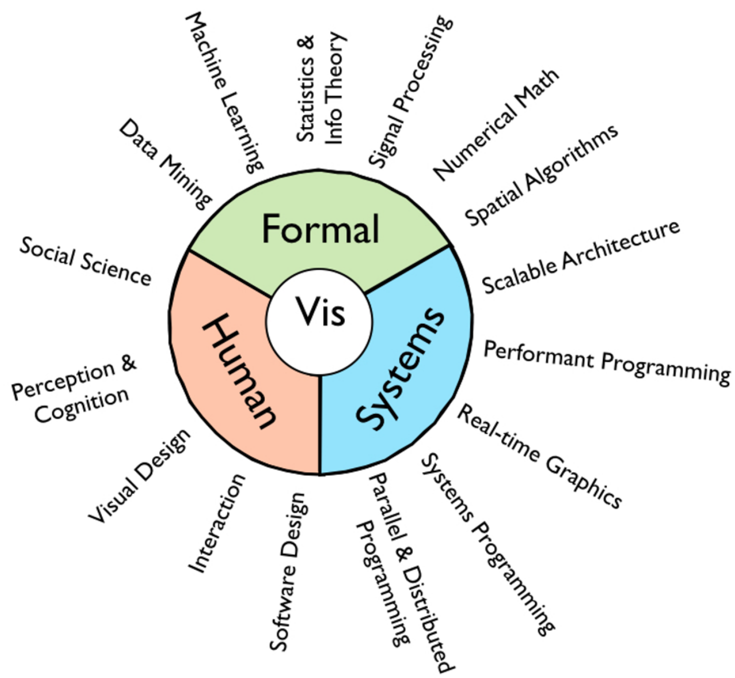

Figure 1: The visualization theory foundations draw on multiple domains.

Abstract

The visualization theory foundations draw on several domains, from signal processing to software design and perception. This chapter describes the landscape of visualization foundations along three aspects: a Humans aspect, a Systems aspect, and a Formal aspect, along with the domains the visualization foundations are rooted in. This chapter further provides definitions for the visualization, theory foundation, theory, model, and concept terms, and a discussion of theory granularity, from grand theories to middle-range theories and to practice theories. The chapter further discusses several challenges related to the theory fabric of visualization that result from the diversity of our roots. The chapter ends with a discussion of possible evaluation criteria for theory components, with respect to the range of theories and models, from mathematical frameworks to guidelines and best practice advice presented in this book.

1 Visualization: definition and essential aspects

Since year 2000, the term ”visualization” appears to be ten times more frequently used in books than ”computational biology”, and about five times more frequently than ”compilers” [1]. No doubt, this use is compounded by the dual dictionary definition of the term, where visualization is defined [2] as either:1) [*mass noun*] The representation of an object, situation, or set of information as a chart or other image (e.g., ”video systems allow visualization of the entire gastrointestinal tract”), or [*count noun*] a chart or other image that is created as a visual representation of an object, situation, or set of information (e.g., ”3D visualizations for architectural design”) or 2) The formation of a mental image of something (e.g., ”visualization is a helpful technique for relieving stress”).

In fact, the emphasis on visualization as a computing discipline started relatively recently in 1987, with the publication of ”Visualization in Scientific Computing”, a special issue of Computer Graphics [3]. Since then, the term has been continuously revisited and redefined, to clarify, for example, that interactive visualization was distinct from digital weather animations.

In this book, we refer to visualization along its first dictionary interpretation above, and as more precisely defined by Munzner: ”Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively” [4]. In this textbook definition of visualization, we note the explicit reference to computer-based systems, to analysis tasks involving datasets, and to humans who perform the tasks and their visual system. In fact, these three axes capture the essential aspects on which the theory foundations of visualization stand: the Systems, Formal/Analysis, and Human aspects.

In turn, these aspects draw on several domains, from signal processing to perception. Figure 1 illustrates the landscape of visualization foundations along the three aspects (Systems, Formal, and Humans), as well as the domains the visualization foundations are rooted in, with their respective connections to specific disciplines: the System aspect connects to Computer Science and Engineering; the Formal aspect connects to Formal Sciences; and the Humans aspect connects to Social Sciences. Clearly, some of these roots are further connected to each other—for example, software design in visualization involves both Systems and Humans, and Data Mining is related to Data Base Programming (Systems), as well as to Machine Learning and Statistics (Formal).

Terminology

At the center of Fig. 1 lie the existing visualization theory components (including principles, frameworks, models, guidelines), whether specific to the visualization field or borrowed from related fields. Before we discuss the challenges related to these theory components, as well as possible criteria for their evaluation, we need terminology. In the following sections, **theory foundation** denotes the set of concepts, theories, and models on which the practice of visualization is based, as well as a system of ideas intended to explain phenomena or observations related to visualization. A **concept** is an abstract idea or a general notion. A **theory** provides a description of concepts and their relationship, in order to help us understand a phenomenon or observation. A **model** is a system or prototype used as an example to follow or imitate; unlike a theory, a model usually involves some meaningful arrangement or sequence of concepts. A **theory component** is any aspect of a theory foundation above, be it a concept, a theory, or a model.

Furthermore, following similar definitions in other fields [5], theories can have different granularity, spanning the formal to practice space. **Grand theories** have the broadest scope and present general concepts and propositions or principles. Grand visualization theories consist of conceptual frameworks intended to be pertinent to all instances of visualization. Theories at this level may both reflect and provide insights useful for practice but are not designed for empirical testing. **Middle-range theories** are narrower in scope than grand theories—they are simple, straightforward, general, and consider a limited number of variables and limited aspect of reality, they present concepts at a lower level of abstraction, and guide theory-based research and visualization practice strategies. The functions of middle-range theories are to describe, explain, or predict phenomena and observations. Middle-range theories offer an effective bridge between grand visualization theories and visualization practice. One of the hallmarks of middle-range theory compared to grand theories is that middle-range theories are more tangible and verifiable through testing. **Practice theories** have the most limited scope and level of abstraction and are developed for use within a specific range of visualization situations. Visualization practice theories may provide guidelines for visualization design and implementation, and predict outcomes and the impact of visualization practice. The capacity of these theories is limited, and they analyze a narrow aspect of a visualization phenomenon or observation. Visualization practice theories are usually defined to an exact community (e.g., broad audiences, domain expert audiences etc.).

2 Theory foundation challenges

Because the visualization theory foundations draw on several domains (Fig. 1), it is important to be aware of the effect these roots have on the theory fabric of visualization. First, while several books, including this collection, seek to formulate theory components in the form of concepts, theories and models, it is important to be aware that, with such complex roots, the landscape of visualization theory is sparsely populated. For example, it is unlikely that exactly three types of visual analysis workflows exist, although only three are currently documented (”Overview-first”, ”Search-first”, and ”Details-first”) [6, 7, 8]. Second, given the root diversity of visualization, theoreticians should be open to multiple points of view. In fact, as a result of their adaptation to the visualization research context, some theories and models may disagree with each other, some may complement each other, some may be incorrectly framed into, transferred to, or applied to the visualization field, and some concepts may be duplicated. Third, given that the visualization field itself is still evolving, we should be prepared to revisit our theory foundation periodically. For example, from a practical perspective, some of the visualization theories are based on observations, and may be later contradicted by other observations. Fourth, we note that while some of the visualization roots are mature (e.g., numerical math), others are still, themselves, evolving at a fast pace (e.g., social science). In consequence, there is always a chance that visualization theory components will ignore important new developments in our root disciplines. For these reasons, the visualization field should never overlook or dismiss a challenge to our theory foundation. Furthermore, given the currently high entry bar for new theory contributions, we could also carefully consider how to evaluate and challenge an existing theory component.

In this section, we illustrate these issues by examining first the source of one example foundational component, showing how that component entered the visualization field, what evidence supported it at the time, and how it was later challenged.

A Trajectory: User- and Activity-Centered Design

To illustrate some of the challenges associated with having multiple roots as a field, let us consider the historical background and trajectory of a core theory concept in visualization, that of domain characterization in the visualization design process.

Visualization relies significantly on data from other domains. As a consequence, as we train in visualization research, we also train into interdisciplinary collaboration. Whether we seek to design a novel technique, or adapt or extend an existing one, the users' needs, goals, constraints need to be first discussed. In the visualization literature, this first step is known as "characterizing the domain", and it is particularly important, since subsequent layers in the design process depend on it [4]. This step is also notoriously difficult: from one end, designers may lack the domain knowledge to extract or even understand the domain experts' needs; from the other end, the domain expert may not be able to articulate those needs, or have the time to apprentice a visualization researcher.

While the concept of domain characterization exists in both the software engineering and in the interaction design literature (as "requirements engineering"), specific models of this process have been constructed in the visualization field, perhaps because visualization design relies on human visual system and depends heavily on data. These models tend to rely on the User-Centered-Design or Human-Centered-Design (HCD) paradigm introduced by human computer interaction research. In this paradigm, as described by Don Norman in his "The Design of Everyday Things" 2002 book [9], we (the designer) start the design process by observing the user, we generate ideas, then prototype, after which we test the prototype with the user, and reiterate through this process. The core of this paradigm is that a deep, detailed knowledge of the *user* is necessary in order to design.

Yet several aspects of this HCD paradigm, as adapted into the visualization field, come at odds with either other roots of visualization, or with empirical observations in the field. For example, if, as proposed in the HCD-derived literature, the value of a visualization is measured in its number of users [10], then the relative value of a visual computing project commissioned by the two researchers who will find a cure for Alzheimer's disease would be really questionable, despite its transformative impact. A step further, the software design literature emphasizes writing functional specifications (a layman description of the function of a software system, without any implementation or design details) before prototyping [11]; that essential stage does not appear in HCD models, despite the fact that most visualization systems are a form of software. Furthermore, HCD models of domain characterization also lead to starkly lower rates of project success than those observed in software engineering [12]. Last but not least, visualization domain characterization models based on the HCD paradigm emphasize a so-called visualization triad, Humans-Data-Tasks, and in doing so tend to lose sight of the user workflows and context of the user activity. Workflows are not just sequences of tasks; they are sets of interrelated processes [13].

Interestingly, by 2005 Norman was already cautioning the interaction community against HCD: ”HCD has become such a dominant theme in design that it is now accepted by interface and application designers automatically, without thought, let alone criticism. That’s a dangerous state—when things are treated as accepted wisdom” [13]. Around that time, Norman started advocating for an alternative model, called Activity Centered Design (ACD). Rooted in Russian and Scandinavian Activity Theory, ACD focuses on activities, not on the individual person: ”...because people are quite willing to learn things that appear to be essential to the activity, activity should be allowed to define the product and its structure” (D. Norman, The Design of Everyday Things Revised and Expanded, 3rd ed) [14]. Because activities are performed by humans, ACD can be regarded as an enhancement of Human-Centered Design. However, note that ACD specifically ranks activities before users, and by extension, before data and users. Furthermore, activities are a higher level concept than tasks: with increasing granularity, users have activities (problems) and tasks. An activity is a high-level structure such as ”go shopping” or ”understand the relationship between E.coli genomes”, while a task is a lower-level component of an activity such as ”drive to market”, ”find a shopping basket”, ”use a shopping list to guide the purchases”, respectively ”load the complete E.coli dataset (673 genomes)”, ”locate an ortholog cluster in the 673 genomes”, ”examine the gene neighborhood of the ortholog cluster” [15, 12] etc. An activity is a collected set of tasks, but all performed together, potentially as part of a workflow, toward a common high-level goal. In contrast, a task is an organized, cohesive set of operations directed toward a single, low-level goal [12].

Most notably, the concepts underlying ACD resemble the software engineering emphasis on the functionality of a software system. In particular, because designers and domain experts can agree, during the initial requirement engineering stage of the design, on the activities to be supported by a system, these functionality-related requirements can be verified and formally approved by the ”user” before the ideation stage. After all, a functional specification describes the functionality and features of a product, and it does not concern itself with how the product is implemented, the underlying algorithms, exact interactions, or visual encodings used. Therefore, functional specifications can effectively ensure that the designers are not solving the wrong problem. They can also help the designers avoid situations where the way the data is shown does not fit correctly the user workflow—**before** the prototyping stage.

While the interaction design domain was coming to terms with the ACD paradigm, the visualization field continued to develop theories and guidelines using the HCD paradigm and the Humans-Data-Tasks triad for at least another decade. A first model unifying via ACD the interaction design roots and the software design roots of domain characterization was proposed only in 2017 [12]. In this activity-centered model, functional specifications explicitly capture the user activities and workflows, as determined during the requirements session, in the form of designer-written scenarios. Asking the user to review these scenarios is a unique opportunity to verify that the visualization designers are not solving the wrong problem, before the prototyping stage. When evaluated on a set of 75 visualization projects, this ACD visualization model correlated with a 63% success rate, compared to a 25% success rate using HCD models [12], marking a wealth of missed opportunities in the field.

The example described above illustrates how theory components gleaned from across the fabric of visualization may disagree with each other, and how they, and the entire field, may benefit from being reconciled under a visualization-specific framework. Given the heavy evidence accompanying the 2017 illustrative model (the experience of many young designers, as opposed to the experience of one to a few authors, in the case of earlier guidelines; or in contrast to considering the explicit incorporation for the first time of user workflows into a model as sufficient merit, as argued below), the example also illustrates the high entry-barrier to new theory components. In general, once theory components are established in the visualization field (e.g., the ”Overview-first” mantra [6]), alternatives (e.g., the ”Search-first” mantra [7], and later the ”Details-first” mantra [8]) are introduced with difficulty.

3 Evaluation of a theory component

Once a theory component takes root within a field, the entry barrier for new theory components goes high. In particular, in our experience, appropriately or not, new models and theories are often required during the review process to provide, along with the theory component itself, some form of evaluation of that element. This request comes in contrast to earlier published works; for example, the ”Overview-first” mantra was not accompanied by evidence [6].

One way to address this conundrum is by considering several possible evaluation criteria for theory components. We note that theory components can be supported by empirical evidence, or be mathematically provable. Neither form of evidence is infallible: empirical evidence may be contradicted by later observations, and mathematical proofs often make assumptions (which by definition are statements taken to be true, without proof). In the assessment of the science philosopher Karl Popper, a theory in the empirical sciences can never be proven, although it can be falsified [16]. For example, the statement "All swans are white" can hold true in certain parts of the world, and be falsified once a black swan is observed.

With this observation in mind, in the visualization literature, a model or theory have been acceptably supported by as little as one to a few concrete examples coming from the experience of one to a few authors [17, 18, 19, 20], and sometimes by no evidence [6]. A theory component has also been acceptably supported by evidence from other reports in the literature, by direct comparison against an accepted alternative theory component [12], or by mathematical proof [21]. In general, the visualization literature captures some of the different types of evaluation and concepts used in the field [22, 23], without addressing the theoretical underpinnings of these types and concepts.

In principle, there are many possible evaluation criteria for theory components. Because this topic has not been explicitly discussed in the visualization community, it is worth looking at how other disciplines handle the same issue of theory evaluation. For example, one of the disciplines with keen interest in this topic is nursing. As in visualization, nursing theory includes both theories unique to the field, and theories that have been borrowed from related sciences by practitioners to explain and explore phenomena specific to nursing. Furthermore, nursing theories also span a wide range, from grand theories to practitioner guidelines, and they are also largely based on observations and phenomena, and on a mix of quantitative and qualitative data. The seminal nursing-theory evaluation work of Fawcett and Rizzo-Parse [24, 5, 25] proposes as evaluation criteria the following elements: significance, internal consistency, parsimony, testability, empirical adequacy, and pragmatic adequacy [26], which we briefly discuss below.

Significance: The criterion of significance focuses on the context of the theory component. This criterion requires justification of the importance of the theory component to the discipline, and is met when the origins of the theory component are explicit, when antecedent knowledge is cited, and when the special contributions made by the theory component are identified.

Internal consistency: This criterion focuses on both the context and the content of the theory component. The criterion requires all philosophical claims, conceptual model, and concepts and propositions, to be consistent with each other, the linkages between concepts to be specified and that no contradictions in propositions are evident. The concepts also need to reflect semantic clarity (for example, explicit definitions are given) and semantic consistency (the same term and the same definition are used consistently for each concept in the entirety of the author's discussion).

Parsimony: This criterion assesses whether the content of a theory component is stated clearly and concisely. The fewer the concepts and propositions needed to fully explicate the theory component, the better.

Testability: Theory components may be amenable to direct empirical testing. Such an approach would require the concepts to be observable and the propositions to be measurable. The criterion of testability for middle-range theories may be met, for example, when specific instruments or experimental protocols have been developed to observe the concepts and statistical techniques are available to measure the assertions made by the propositions. Descriptions of personal experiences may be used, although they are not mandatory, to evaluate the testability of grand theories.

Empirical adequacy: This criterion requires the assertions made by the theory component to be congruent with empirical evidence, determined by means of a systematic review. If the empirical data conform to the theoretical assertions, it may be appropriate to tentatively accept the assertions as reasonable or adequate.

Pragmatic adequacy: This criterion focuses on the utility of the theory component for visualization practice. The criterion requires that practice theories be used in the real world of visualization practice, while the extent to which a grand theory or a middle-range theory meets this criterion could be determined, for example, by reviewing all descriptions of the use of the theory in practice.

This list of potential evaluation criteria is not exhaustive; other criteria could be heurism (the amount of research and new thinking stimulated by the theory; whether other theorists quote the theory and use it as a springboard to create their own theories), tests-of-time (a theory's durability over time) and so on. Furthermore, some of the criteria above might be more appropriate for specific theory components; for example, pragmatic adequacy (or actionability) seems a good fit with the guidelines and best-practices described in the latter half of this book, while parsimony and internal consistency may be more appropriate for the mathematical frameworks and formal models described in the first half. Overall, we note the need for a wider range of evaluation criteria of theory components, and for a better understanding of where specific evaluation criteria may apply.

4 Conclusion

The visualization field draws on a multitude of domains, connected to different branches of science and engineering. Accordingly, many of the visualization theory foundations draw from principles in these domains and sciences. In this chapter, we organized these aspects along three main axes: Systems, Formal and Humans, and we used these axes to describe the fabric of the visualization field and its ties to multiple science and engineering domains. Building theory foundations for visualization is the collective responsibility of the visualization community. Our different roots mean that our different sub-communities have different contributions to this space, and at the same time, that our resulting space coverage is in consequence sparse.

As a result of their adaptation to the visualization research context, we noted that some of the resulting theories and models may disagree with each other, some may complement each other, and some may be incorrectly framed into, transferred to, or applied to the visualization field. These complications affect the fabric of visualization, and lead in some cases to terminology overloading and duplication. The field is still very young and fragmented, and there is a need to reconcile conflicting views in the existing theory landscape. To maintain growth and intellectual diversity, we furthermore need to keep an open mind with respect to existing guidelines, accept challenges to our existing theory components, and lower the entry barrier for alternative theories.

There are multiple resources available that discuss theory components of visualization, including software design and user-centric design frameworks, mathematical and systems engineering frameworks, and perceptual and cognitive frameworks [4, 27, 28]. This book itself contributes additional theory components, and is not an exhaustive resource either. All these existing and proposed theory components can be evaluated along a multitude of criteria. While mathematical frameworks may score highly along parsimony and consistency criteria, guidelines and best-practice advice also have complementary value along pragmatic adequacy criteria. In general, the field stands to benefit from a wide range of theory contributions and from a wider range of evaluation criteria.

References

[1] Google, "Google Books N-Gram Search," http://books.google.com/ngrams/, accessed: 2019-04-05.

[2] O. Dictionary, "Oxford dictionaries," *Language Matters*, 2014.

[3] B. H. McCormick, T. A. DeFanti, and M. D. Brown, "Visualization in scientific computing: a synopsis," *IEEE Computer Graphics and Applications*, vol. 7, no. 7, pp. 61–70, 1987.

[4] T. Munzner, *Visualization analysis and design*. AK Peters/CRC Press, 2014.

[5] R. R. Parse, *Nursing science major paradigms, theories, and critiques*. W B Saunders Co, 1987.

[6] B. Shneiderman, "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations," in *Proc. IEEE Symp. Visual Lang.*, 1996, pp. 336–.

[7] F. van Ham and A. Perer, """Search, Show Context, Expand on Demand": Supporting Large Graph Exploration with Degree-of-Interest," *IEEE Trans. Vis. Comp. Graph.*, vol. 15, no. 6, pp. 953–960, 2009.

[8] T. Luciani, A. Burks, C. Sugiyama, J. Komperda, and G. Marai, "Details-First, Show Context, Overview Last: Supporting Exploration of Viscous Fingers in Large-Scale Ensemble Simulations," *IEEE Trans. Vis. Comp. Graph.*, vol. 25, no. 1, pp. 1–11, 2019.

[9] D. A. Norman, *The Design of Everyday Things*. New York, NY, USA: Basic Books, Inc., 2002.

[10] J. J. Van Wijk, "The value of visualization," in *Visualization, 2005. VIS 05. IEEE*, 2005, pp. 79–86.

[11] J. Spolsky, "Painless Functional Specifications Part 1: Why Bother?" in *Joel on Software*. Springer, 2004, pp. 45–51.

[12] G. E. Marai, "Activity-centered domain characterization for problem-driven scientific visualization," *IEEE Trans. Vis. Comp. Graph.*, vol. 24, no. 1, pp. 913–922, 2018.

[13] D. A. Norman, "Human-centered design considered harmful," *interactions*, vol. 12, no. 4, pp. 14–19, 2005.

[14] D. Norman, *The design of everyday things: Revised and expanded edition*. Basic books, 2013.

[15] J. Aurisano, K. Reda, A. Johnson, E. G. Marai, and J. Leigh, "BactoGeNIE: a large-scale comparative genome visualization for big displays," *BMC Bioinformatics*, vol. 16, no. 11, p. S6, 2015.

[16] B. Gower, *Scientific Method: An Historical and Philosophical Introduction*. Psychology Press, 1997.

[17] D. Lloyd and J. Dykes, "Human-Centered Approaches in Geovisualization Design: Investigating Multiple Methods Through a Long-Term Case Study," *IEEE Trans. Vis. Comp. Graph.*, vol. 17, no. 12, pp. 2498–2507, 2011.

[18] T. Munzner, "A Nested Model for Visualization Design and Validation," *IEEE Trans. Vis. Comp. Graph.*, vol. 15, no. 6, pp. 921–928, 2009.

[19] A. J. Pretorius and J. J. Van Wijk, "What Does the User Want to See?: What Do the Data Want to Be?" *Information Vis.*, vol. 8, no. 3, pp. 153–166, 2009.

[20] M. Sedlmair, M. Meyer, and T. Munzner, "Design Study Methodology: Reflections from the Trenches and the Stacks," *IEEE Trans. Vis. Comp. Graph.*, vol. 18, no. 12, pp. 2431–2440, 2012.

[21] M. Chen and H. Jaenicke, "An information-theoretic framework for visualization," *IEEE Trans. Vis. Comp. Graph.*, vol. 16, no. 6, pp. 1206–1215, 2010.

[22] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale, "Empirical studies in information visualization: Seven scenarios," *IEEE Trans. Vis. Comp. Graph.*, vol. 18, no. 9, pp. 1520–1536, 2011.

[23] T. Isenberg, P. Isenberg, J. Chen, M. Sedlmair, and T. Möller, "A systematic review on the practice of evaluating visualization," *IEEE Trans. Vis. Comp. Graph.*, vol. 19, no. 12, pp. 2818–2827, 2013.

[24] J. Fawcett and S. Desanto-Madeya, *Contemporary nursing knowledge: Analysis and evaluation of nursing models and theories*. FA Davis, 2012.

[25] R. R. Parse, "Parse's criteria for evaluation of theory with a comparison of Fawcett's and Parse's approaches," *Nursing Science Quarterly*, vol. 18, no. 2, pp. 135–137, 2005.

[26] J. Fawcett, "Criteria for evaluation of theory," *Nursing science quarterly*, vol. 18, no. 2, pp. 131–135, 2005.

[27] E. R. Tufte, *The visual display of quantitative information*. Graphics press Cheshire, CT, 2001, vol. 2.

[28] C. Ware, *Information Visualization: Perception for Design*, ser. 3^{rd} edition. Morgan Kaufmann, 2012.