Visualization Task Abstraction from Multiple Perspectives

Matthew Brehmer

Abstract— Applied information visualization design and evaluation involves a mapping between a domain problem and appropriate visual encoding and interaction techniques. This mapping translates a domain-specific workflow into *abstract visualization tasks*, which facilitates the comparison of visualization usage across application domains; the mapping also applies in reverse, whenever practitioners aim to contextualize novel visualization techniques. I propose four perspectives on visualization task abstraction: the *synthesis* perspective describes our recent typology of abstract visualization tasks. The remaining three perspectives involve validating this typology in applied settings: a post-deployment evaluation of visualization utility in the domain of investigative journalism, an interview study spanning 11 scientific research domains pertaining to visualization tasks for dimensionally-reduced data, and a visualization design study project in the domain of energy usage analysis.

Index Terms—Task and requirements analysis, visualization design studies, evaluation beyond time and error.

1 Introduction

How do we evaluate the utility or usefulness of visualization techniques and tools in an applied domain context, particularly if these techniques and tools are used for exploratory data analysis?

This question directed my initial thinking when I entered the PhD program in late 2011. At the time, I had read an early manuscript of Lam et al.'s survey of visualization evaluation [7], in which the authors read and coded over 800 recent visualization papers that report an evaluation component; while many of these papers evaluate human perceptual performance or visualization usability, relatively few of these papers document an attempt to evaluate visualization *utility*, and fewer still comment on *adoption*: whether a deployed visualization tool was incorporated into the regular data analysis workflows of real users. The findings of this survey prompted me to ask: *why is the study of visualization utility and adoption reported so infrequently?* and *if this research is difficult to conduct, what makes it so difficult?*

Initially, I focused my attention on the series of ACM BELIV workshops on novel evaluation techniques for visualization, from which a number of methodologies and methods for evaluating visualization for exploratory data analysis have been proposed. I attended the 2012 BELIV workshop, where there was substantial discussion pertaining to a need for a better shared understanding of visualization users' *tasks*, and that the effective use and generalizability of these evaluation methodologies for assessing utility or usefulness rely upon this understanding.

I also gravitated toward visualization design studies, in which visualization techniques and tools are developed for specific target users having domain-specific data and tasks; design studies also incorporate many of the evaluation methodologies from the BELIV community. In a design study, it is critical that the researchers have correctly abstracted the users' domain specific tasks and mapped these to appropriate visual encoding and interaction techniques [11], often borrowing techniques originally applied to other domains. However, this abstraction and mapping is seldom straightforward: initial designs often fail to address users' tasks, inappropriate evaluation methods are selected, or prematurely deployed visualization tools fail to be adopted by target users [15].

Problem statement: By early 2013, my thinking had coalesced into this problem statement: *applied visualization design and evaluation is difficult because mapping users' tasks to visual encoding and interaction techniques requires multiple levels of abstraction, and practitioners would benefit from a domain-agnostic, consistent, and validated instrument for assisting in this mapping*.

There is currently little agreement in the visualization literature as to the appropriate granularity for describing a *task*. For instance, *finding an extreme value* is fairly concrete, while *exploring* or *integrating insights* are quite abstract, and *comparing sequence variants in a human genome* is quite domain-specific. Relating these descriptions is difficult, though not impossible; however, visualization practitioners lack a shared lexicon when describing these relations between levels of abstraction.

Multiple perspectives: My proposed thesis examines visualization task abstraction from multiple perspectives. The first perspective, described in Section 2.1, documents the *synthesis* of related work and extant theory relating to task analysis. The result of this synthesis was a new theoretical instrument, a typology for characterizing abstract visualization tasks at multiple levels of specificity [3].

Proposing this conceptual typology is not enough: it is necessary to validate this typology as a pragmatic tool [1]; this validation will address the typology's ability to describe or *analyze* existing interactions between target domain users and visualizations, to *generate* new designs, and to *evaluate* these designs. The three forms of validation are intertwined, as the ability to *generate* or *evaluate* implies the ability to *analyze*; I propose to address all three types of validation.

The remaining three perspectives aim to validate this typology in applied settings spanning more than a dozen application domains. In Section 2.2, I describe how the typology was used to *evaluate* a visualization tool that was adopted by investigative journalists [2]. In Section 3.1, I propose the use of the typology to *analyze* findings from a interview study about visualizing dimensionally-reduced data in 11 scientific research domains. Finally, in Section 3.2, I describe the ongoing use of the typology to *design* and *evaluate* prototype visualization designs in the domain of energy usage analysis.

2 Completed Work

The first two perspectives, *synthesis* and *journalism design study*, are documented in published and submitted research papers [2, 3].

2.1 Perspective 1: Synthesis

My motivation for developing an instrument for characterizing tasks was pragmatic: we had amassed observational data of visualization usage in the *journalism design study* and *interview study* projects (described below), and we were struggling to describe and compare tasks between users, visualization tools, and domains. We required a systematic tool for analyzing tasks abstractly, allowing us to both describe and evaluate visualization designs that address these tasks.

Methodology: We began with a comprehensive review of previous work that characterized user tasks, goals, intentions, and interactions. This review included over two dozen extant classification systems and

• *Matthew Brehmer is with the University of British Columbia. E-mail: brehmer@cs.ubc.ca.*

Manuscript received 10 May 2014; date of current version 31 July 2014. For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

theoretical frameworks from the literatures of visualization, human-computer interaction, information retrieval, communications, and cartography; a full list of references is provided in our paper [3]. We examined the vocabulary and definitions used in this body of previous work, and after multiple rounds of coding, we had grouped similar terms, selected representative terms for each group, and arranged these representative terms into multiple levels of abstraction. We reasoned about how tasks could be described using this arrangement of terms, either in isolation, or as a sequence of interdependent tasks.

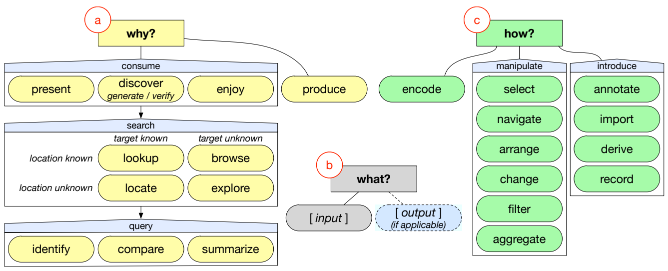

Fig. 1. Our proposed multi-level typology of abstract visualization tasks [3], which characterizes *why* a task is undertaken (a), *what* dependencies a task might have (b), and *how* the task is supported (c) in terms of visual encoding and interaction idioms.

Results and contributions: The result of our synthesis was a typology of abstract visualization tasks, illustrated in Figure 1. This typology allows for succinct descriptions of tasks, in which a task description is comprised of *why* a task is undertaken at multiple levels of specificity, *what* dependencies a task might have in terms of *input* and *output*, and *how* the task is supported in terms of visual encoding and interaction *idioms*; given this structure, it is possible to describe sequences of interdependent tasks. Our typology has since proven to be useful in our subsequent design study work discussed below, as well as in recent work by others [12, 13].

Feedback from panelists: I would appreciate feedback on how to validate or extend this work beyond the ongoing and proposed work described in Section 3:

- *Aside from design studies and interview studies, by what other means could we validate or refine this typology?*

2.2 Perspective 2: Journalism Design Study

Since 2010, we have collaborated with a software developer / journalist to build and deploy a visualization application that would allow journalists to systematically explore large text document collections, those attained from Freedom of Information Act requests or from whistleblowers, collections ranging in size from hundreds to tens of thousands of documents. Since this time, the tool has been deployed as a web-based application called *Overview*[1] (shown in Figure 2). I led a post-deployment evaluation of *Overview*, in which I analyzed its adoption and use by self-initiated investigative journalists.

Methodology: We conducted case studies of six journalists who used *Overview* to conduct an investigation involving a large document collection; these journalists agreed to participate in our research after they had contacted the *Overview* team for technical support. In five of these cases, the investigation resulted in a published story. Once their investigation was complete, we interviewed these six journalists about the form and provenance of their documents, the objectives of their investigation, and their use of *Overview*; we also collected their logged interaction data and their annotated document collections. We used our task typology [3] to better understand *why* and *how Overview* was adopted by these journalists to perform their investigations.

Results: The analysis of journalists' use of *Overview* revealed that our initial understanding of their task was insufficient: the task of *"exploring"* a document collection, which appears often in previous work on visualizing document data, is both too vague and too narrow to capture how journalists actually used our application. Instead, we identified

[1] http://overviewproject.org

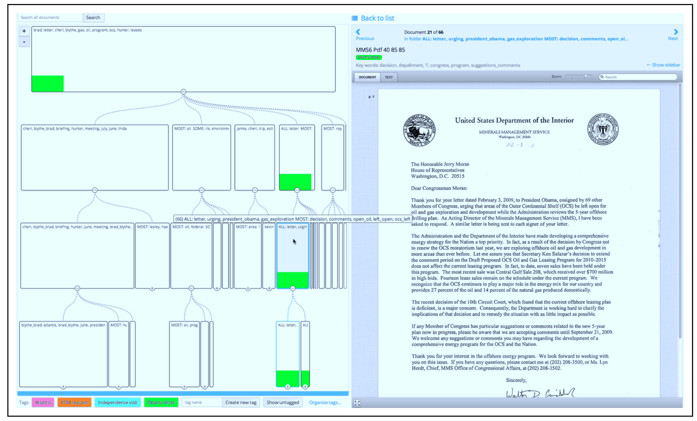

Fig. 2. *Overview* [2] is a multiple-view application intended for the systematic search, annotation, and reading of a large collection of text documents, which visualizes hierarchical clusters of documents as a tree (left). We used our task typology [3] to better understand *why* and *how Overview* was adopted by self-initiated journalists to perform real investigations leading to published stories.

two different tasks using the vocabulary and structure of our typology: one of *generating* hypotheses and *summarizing* the contents of a document collection, and another of *locating* and *identifying* specific evidence in order to *verify* or *refute* prior hypotheses.

Contributions: Given our more precise understanding of journalists' tasks, we were able to rigorously analyze the rationale for *Overview*'s visual encoding and interaction design choices. This analysis generalizes beyond the domain of journalism and speaks to the design of visualization techniques and tools addressing document data or hierarchically-clustered data in general. In addition, our analysis of real world visualization usage is a form of validation for our proposed task typology. Finally, we were able to reflect upon *Overview*'s design and evaluation process, as well as the value, logistics, and limitations of studying visualization adoption.

Feedback from panelists: A limitation of adoption-phase research is that a set of specific target users cannot be identified in advance, in contrast to the typical design study chronology [15]. As a result, there is an inherent selection bias in our case studies, because they largely represent successful cases; this elicits the following question, which is relevant to the future evaluation of *Overview* and to the *energy analysis design study* project described in the next section:

- *How can we study cases in which a deployed visualization tool was used briefly and then abandoned as being unsuitable for the problem at hand?*

3 Ongoing and Proposed Work

The remaining two perspectives, *interview study* and *energy analysis design study*, are currently the focus of ongoing work.

3.1 Perspective 3: Interview Study

This project focuses on our typology's ability to *analyze* current data analysis practices "in the wild", by way of a interview study. In particular, we are using the typology to examine the data analysis tasks of scientific researchers in several domains, and specifically tasks related to the analysis of high-dimensional data; we seek to better understand this data, the transformations applied to it, as well as *why* and *how* visualization is used throughout analysts' workflows.

Methodology: The original focus of this research was to examine the use of dimensionality reduction techniques in scientific research [14]. Our data collection and analysis methodology, included 24 field interviews with scientific researchers and a literature survey spanning 11 domains including chemistry, biology, computer science, and policy analysis. Our approach was similar to Kandel et al.'s recent interview-based study characterizing data analysis and visualization use among enterprise data analysts [6]; we see our work as being complementary to Kandel et al.'s findings, given that both projects address data analysis and visualization use in real work settings for a broad group of domains.

Current status: Using a qualitative coding approach in the spirit of grounded theory, we developed a characterization of high-dimensional data analysis and dimensionality reduction in scientific research, as depicted in Figure 3. We distinguished between analysis tasks relating to learning about the data's dimensionality and those relating to learning about clusters of items in the data.

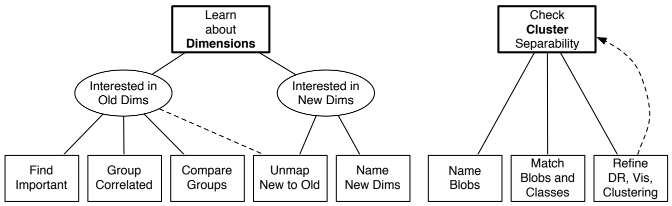

Fig. 3. A characterization of dimensionality-reduction tasks in scientific research [14], having implications for the design and evaluation of visual encoding and interaction techniques, particularly those relating to filtering, aggregating, deriving, and annotating data.

Planned work and expected contributions: After several rounds of iteration and review, our original paper [14] was rejected. Reviewers commented that our characterization of dimensionality reduction does not adequately relate to previous characterizations in the statistics and machine learning literature, and that our findings do not adequately refer back to visualization design. We have recently submitted a substantial and much-abbreviated revision to the 2014 BELIV workshop [4].

We no longer attempt to characterize the use of dimensionality-reduction techniques themselves; instead, we retain a focus on the outcome of these techniques, on dimensionally-reduced data. We position this work as a *data type-specific characterization of tasks*, grounded in observations of real-world analyst behaviour. In our typology paper [3], we encouraged the further characterization of tasks specific to data type, as these are complementary to our data type-agnostic typology; an example would include Lee et al.'s taxonomy of graph-specific tasks [9]. To our knowledge, no classification of visualization tasks for dimensionally-reduced data exists; we hope that our classification would be an appreciated addition to existing characterizations of data type-specific tasks.

With the advent of our task typology, we have a new theoretical lens with which to approach the considerable amount of data that we collected, and to better relate our findings to specific visual encoding and interaction techniques, thereby validating and informing visualization technique research. We now have a richer understanding of *why* these techniques are applied in sequential activities, such as in the sequence depicted in Figure 4, and we can now characterize *what* the *inputs* and *outputs* of these tasks are [4]. Using these characterizations, we can suggest guidelines for designing visualizations for these sequences, as well as evaluate how existing systems address these tasks.

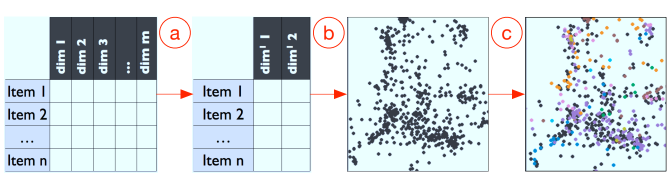

Fig. 4. A task sequence in which (a) data is reduced to 2D, (b) encoded in a scatterplot to *verify* visible clusters, and (c) colour-coded according to preexisting class labels to *match* clusters and classes.

Feedback from panelists: I would appreciate any feedback on our shift in focus, from a characterization of dimensionality reduction to a one of visualization tasks for dimensionally-reduced data. Another focus could involve a survey of *how* these task sequences are supported by existing data analysis tools and *how* dimensionally-reduced data is visually encoded and manipulated.

- *How could we integrate these foci in a longer journal paper?*

3.2 Perspective 4: Energy Analysis Design Study

In 2013, we initiated a visualization design study project that provided an opportunity to validate the *generative* potential of our typology [3]. We entered a collaboration with an industry partner that develops energy usage reporting software for commercial clients. Many of these client organizations have designated energy analysts who oversee large portfolios of buildings; these analysts are responsible for identifying cost saving opportunities, diagnosing erratic energy usage behaviour, and attempting to understand the role of fluctuating external factors such as weather, occupancy, operating hours, and equipment usage within buildings. Tools and techniques for addressing these tasks for single buildings already exist, however they do not scale to portfolios of hundreds or thousands of buildings, potentially spanning large geographical regions. In addition, many commercial buildings are now outfitted to report energy usage at the granularity of minutes, rather than months, which is still typical of residential buildings. We conjectured that an application integrating interactive visualization while considering these issues of scale could address analysts' tasks.

Methodology: We began by interviewing six energy analysts from commercial client organizations who were current users of our industry partner's software, asking them about their roles and responsibilities, their technical background, their portfolio of buildings, and the limitations of current tools. We used our typology [3] to identify and abstract the tasks of these analysts. We focused on a subset of these tasks, which could be summarized as follows: given a portfolio of buildings, an analyst will perform a multi-faceted comparison of absolute, normalized, and ranked energy usage performance over time, faceted by performance measure, time interval, or by shared item attributes, such as building location, occupant count, or size, then drilling down from the entire portfolio to groups of items, and from groups to single items.

This abstraction has informed our ongoing process of mapping these tasks to a set of appropriate visual encoding and interaction techniques. Among the techniques that we have considered thus far include those for performing multiple comparisons between aggregate and individual items over time, for identifying cyclic and acyclic events using meaningful temporal granularities, and for identifying differences in multiple lists of ranked items while simultaneously identifying the cause of rank changes [5].

Current status: Over the course of two months, we designed and implemented over a dozen interactive visualization *data sketches* [10] to address the tasks of these analysts, following a process of rapid iteration in which functional sketches featuring target analysts' data were used to further refine our understanding of analysts' tasks and context of use; three of these data sketches are shown in Figure 5. In early 2014, we demonstrated these sketches to four groups of analysts; the energy usage data used in these demonstrations was collected from their own building portfolios.

Planned work and expected contributions: Our industry partner has integrated ideas from our data sketches into their product development cycle, and we expect to evaluate deployed designs with current and prospective users beginning in the Fall of 2014. Throughout this time, higher-fidelity prototype designs will allow us to discover how our visual encoding designs are used in series or in conjunction as part of analysts' analysis workflows. In the more immediate future, we will continue to explore the design space of visual encodings.

While the visualization research community values design studies, it is often difficult to predict what a design study's contributions will be at early phases of design, prior to subsequent phases of deployment, analysis, and reflection [15]. Nevertheless, one possible contribution is further validation of our typology [3], with a specific focus on descriptions of analysts' workflows as sequences of interdependent tasks.

Feedback from panelists: The current set of data sketches serve to explore the design space of visual encoding techniques; we have yet to fully discover the design space of interactive techniques. While analysts' reaction to our sketches have been largely positive, we cannot fully understand how these visualizations might be used in users' complex, iterative, and interactive analysis workflows encompassing

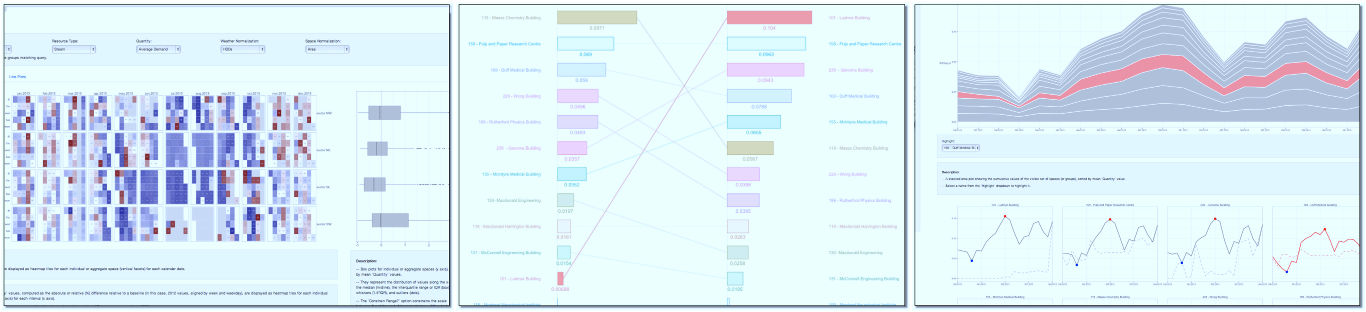

Fig. 5. Visualization data sketches for analyzing the energy usage of building portfolios. *Left*: calendar-based tilemaps are aligned to box plots encoding the difference between average energy demand in 2012 and 2013 for four groups of buildings; *Centre*: a visualization inspired by Gratzl et al.'s *LineUp* [5], encoding the rank (y axis) and absolute value of 12 buildings' energy consumption (bar lengths) in 2012 and 2013 (x axis), with rank deltas encoded as alpha-varying bump plot lines connecting the bars; *Right*: time-series stacked area plot of 12 buildings' combined energy consumption in December 2013 (top), with small multiple plots for each building (bottom)

drill-down and backtracking operations: a user might start with a portfolio of items, select an arbitrary group of items, then select a single item from within this group, and subsequently retrace these steps, ideally without losing context. We are currently considering a class of techniques proposed in previous work that allow the user to alternate between the analysis of multiple items and the detailed analysis of a single item, providing a visual history of previously viewed visualizations [16]. However, implementation of these techniques requires significant development effort, and we risk ignoring viable alternatives:

- *If our rapidly-developed "data sketches" [10] serve to explore the space of visual encoding design, is there an analogous way to develop "interaction sketches" capable of loading real data that serve to explore the space of possible workflows?*

There exists a dependency between visual encoding and interaction design: once a set of visual encoding designs are encapsulated in a *view*, practitioners must consider how these views are to be coordinated or sequenced. Considerations for the design of multiple-view systems are well-documented [8], and include questions such as *how many discrete views are appropriate*, *how should views be arranged or sequenced* and *how should views be coordinated*, such as with linking and brushing techniques. Exploring the design space of view sequencing and coordination within interactive workflows should not necessitate high-fidelity prototyping, and yet low-fidelity storyboarding techniques are limited in that they would require static views of a single dataset: *does a middle-ground exist?*

Another open question relates to the structure of the data being considered: analysts' tasks include multiple comparisons, both within and between individual or aggregate items. Some of these comparisons involve time-varying values, which are aggregated by month, week, or day depending on a user-defined time window, while others involve comparisons of variability using distributions of values removed their temporal context. The juxtaposition of time-varying and summarized data is illustrated by the sketch in Figure 5 (left), which features four rows of calendar-based tilemaps aligned to box plots, where each row is a group of buildings. This form of juxtaposition has been a source of confusion among prospective users.

- *For data that can be aggregated in terms of items and in terms of time, do effective combinations of visual encoding and interaction techniques exist for facilitating multiple simultaneous comparisons of statistical summaries and time-varying values?*

4 Milestones and Conclusion

The four perspectives on visualization task abstraction outlined in this paper were described in a thesis proposal document and defended before my supervisory committee in May 2014. My remaining milestones include the completion of the two ongoing and proposed projects described in Section 3, as well as writing and defending my dissertation, with a target completion date in at end of 2015.

Our typology of abstract visualization tasks has already proven useful in two design studies, at both formative and summative stages of design and evaluation. Upon the completion of my thesis, I hope to demonstrate the pragmatic value of this domain-agnostic analysis tool for applied visualization research and design.

Acknowledgments

I thank my supervisory committee: Tamara Munzner, Joanna McGrenere and Ron Rensink, my collaborators: Stephen Ingram, Michael Sedlmair, Jonathan Stray, and the product development team at Pulse Energy, as well as my funding sources: NSERC and Mitacs.

References

[1] M. Beaudouin-Lafon. Designing interaction, not interfaces. In *Proc. ACM AVI*, pages 15–22, 2004.

[2] M. Brehmer, S. Ingram, J. Stray, and T. Munzner. Overview: The design, adoption, and analysis of a visual document mining tool for investigative journalists. *To appear in IEEE TVCG (Proc. InfoVis)*, 2014.

[3] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE TVCG (Proc. InfoVis)*, 19(12):2376–2385, 2013.

[4] M. Brehmer, M. Sedlmair, S. Ingram, and T. Munzner. Visualizing dimensionally-reduced data: Interviews with analysts and a characterization of task sequences. Submitted to *the 2014 ACM BELIV Workshop*.

[5] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit. LineUp: Visual analysis of multi-attribute rankings. *IEEE TVCG (Proc. InfoVis)*, 19(12):2277–2286, 2013.

[6] S. Kandel, A. Paepcke, J. M. Hellerstein, and J. Heer. Enterprise data analysis and visualization: An interview study. *IEEE TVCG*, 18(12):2917–2926, 2012.

[7] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE TVCG*, 18(9):1520–1536, 2012.

[8] H. Lam and T. Munzner. A guide to visual multi-level interface design from synthesis of empirical study evidence. *Synthesis Lectures on Visualization*, 1(1):1–117, 2010.

[9] B. Lee, C. Plaisant, C. S. Parr, J. D. Fekete, and N. Henry. Task taxonomy for graph visualization. In *Proc. ACM BELIV Workshop*, 2006.

[10] D. Lloyd and J. Dykes. Human-centered approaches in geovisualization design: investigating multiple methods through a long-term case study. *IEEE TVCG (Proc. InfoVis)*, 17(12):2498–2507, 2011.

[11] M. Meyer, M. Sedlmair, P. S. Quinan, and T. Munzner. The nested blocks and guidelines model. *Information Visualization*, In press. doi.org/10.1177/1473871613510429.

[12] B. Mirel and C. Görg. Scientists' sense making when hypothesizing about disease mechanisms from expression data and their needs for visualization support. *BMC Bioinformatics*, 15(1):117–128, 2014.

[13] B. Saket, P. Simonetto, and S. Kobourov. Group-level graph visualization taxonomy. In *Proc. EuroVis Short Papers*, 2014.

[14] M. Sedlmair, M. Brehmer, S. Ingram, and T. Munzner. Dimensionality reduction in the wild: Gaps and guidance. Technical report, University of British Columbia, 2012. http://goo.gl/Ngyyz1.

[15] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE TVCG (Proc. InfoVis)*, 18(12):2431–2440, 2012.

[16] S. van den Elzen and J. J. van Wijk. Small multiples, large singles: A new approach for visual data exploration. *Computer Graphics Forum (Proc. EuroVis)*, 32(3):191–200, 2013.