Matthew Brehmer

Research Statement - January 2016

Data Visualization Researcher

Vancouver, British Columbia

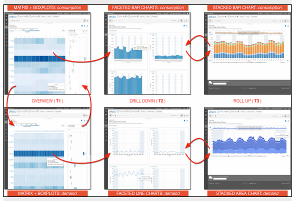
mattbrehmer.github.io

mattbrehmer@gmail.com

@mattbrehmer

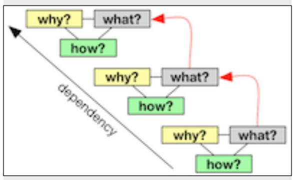
Our ability to gain an understanding of phenomena in our lives, our world, and our universe is accelerated by the visualization of data. Likewise, our ability to communicate our newfound understanding is made possible by visualizing these phenomena and presenting these visualizations to others. The science and practice of visualizing data is maturing: while the visualization research community continues to develop and refine techniques for graphically encoding data and for interacting with these visualizations, there is a growing community of visualization practitioners, and with it, a growing audience for visualization content, particularly in the domains of journalism, public administration, and business intelligence. In response, the visualization research community must contribute guidance, ensuring that appropriate visualization techniques are chosen for particular data analysis tasks or communication intents. My academic background is in human-computer interaction and cognitive science, and over the last several years I have drawn from these disciplines to provide such guidance. My research has been grounded in the design and evaluation of visualization tools and techniques that address a specific application domain or a particular form of data. This applied research has not only inspired new and ongoing visualization technique development; it has also led to considerations for introducing people to novel or unfamiliar visualization techniques.

Applied Visualization Design and Evaluation



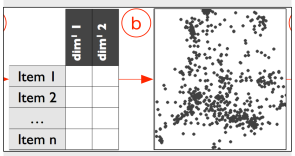
*Visualizing energy usage in large building portfolios [**InfoVis15**].*

In technique-oriented visualization research, the development of a technique is motivated by an abstract data analysis task or communication intent, in combination with a known data abstraction; for instance, consider the task of comparing multiple arrays of quantitative values, where each value is mapped to a point in time. Conversely, applied visualization research is motivated by a specific domain problem; for instance, the domain problem that we addressed in our **[InfoVis15]** paper was one faced by energy analysts at large organizations, in which they needed to understand how energy was being consumed by the organization's buildings, reliably diagnose anomalous energy spikes or outages, and confidently decide upon measures to conserve energy. In applied projects such as this one, the visualization practitioner should recast the specific domain problem as a set of task and data abstractions, thus enabling them to consider existing candidate visualization techniques or decide that a new technique is warranted. As a consequence, this abstraction process allows for cross-pollination between application domains, where techniques originally developed for one domain can be appropriated by another, owing to similarities in data and task abstractions.



*Our typology of abstract visualization tasks is a tool for mapping domain problems to visualization techniques [**InfoVis13**].*

However, the process of decomposing and abstracting domain-specific data analysis tasks or communication intents is seldom straightforward. Not only can tasks have complex hierarchical and sequential relationships, but visualization practitioners lack a shared vocabulary to describe these tasks in an abstract way. This is why we developed a typology of abstract visualization tasks **[InfoVis13]**. This work integrated the vocabulary used throughout the visualization literature, and was informed by approaches to task analysis used in human-computer interaction and cognitive work analysis. Our typology provides a framework from which to consider multiple levels of abstraction, intended to help practitioners map domain problems to a set of appropriate visualization techniques; it also encourages practitioners to think about sequential dependencies between tasks.



*We interviewed individuals who visualized dimensionally-reduced data in their ongoing workflows [**BELIV14a**].*

My dissertation research demonstrates the efficacy of this approach to task analysis and abstraction in three applied visualization design and evaluation projects. In our **[InfoVis14]** paper, we applied this approach to the post-deployment case study evaluation of Overview, a visualization tool that allows investigative journalists to analyze large collections of text documents in an efficient way. Over the course of multiple deployments and six case studies, our task abstraction revealed two dominant use cases for Overview: one that involves browsing documents and identifying themes, and another that involves a more systematic search for evidence, a search that cannot be easily formulated as a keyword query. The latter task was not anticipated at the outset of the project, and the designers of Overview acted upon this finding to better support this task in later versions of Overview. In our **[BELIV14a]** paper, we again applied this approach to task analysis in a formative evaluation, an interview study in which we spoke to ten data analysts spanning six application domains; each interviewee was accustomed to working with high-dimensional quantitative data and used dimensionality reduction and visualization techniques as part of their ongoing workflows. We used our typology to identify six abstract task sequences relevant to the visualization of dimensionally-reduced data, and we recommend that those developing new visualization techniques for this type of data use these task sequences as motivation and as a way to frame their contributions.

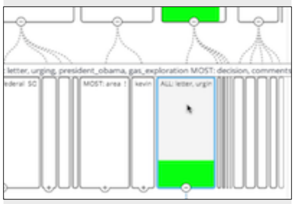
Finally, we used our approach to task analysis during the requirements gathering phase of a visualization design project in the energy domain **[InfoVis15]**; based on interviews with nine energy analysts and consultation with our collaborators, we identified three recurring energy analysis tasks and subsequently developed a set of candidate visualizations to address these tasks. Some of these visualizations were adopted by our collaborator and integrated into their commercial energy analysis software tool.

Future research: I will continue to engage in applied visualization design and evaluation projects. This entails establishing new collaborations and learning about new application domains. As in our **[InfoVis15]** paper about visualization for energy analysis, these projects allow me to embed myself into a domain environment, work alongside collaborators, and regularly consult with the people who will use the new visualization tool or technique. Universities and large organizations are ideal environments for fostering these collaborations, as many groups may benefit from working with an embedded visualization researcher-pracitioner. The visualization research community will stand to benefit from these applied projects as well: they provide use cases for existing visualization techniques, inspiration for novel technique development, and documentation of an applied design methodology. In addition to continued methodological development relating to visualization task analysis and abstraction, I aim to contribute methodological guidance relating to work domain analysis for visualization, requirements elicitation, and the use of visualization artefacts in interviews and focus groups. We provide methodological guidance of this nature in our **[InfoVis15]** and **[BELIV14b]** papers. However, since each design project presents unique problems and constraints on research, there will continue to be a need for methodological development. Finally, on the evaluation front, I will continue to study the adoption and appropriation of visualization tools and techniques following their deployment or dissemination. In our **[InfoVis14]** and **[BELIV14a]** papers, we studied self-initiated rather than researcher-prompted use of visualization in the wild: people using visualization with their own data as part of their regular workflows. This approach yields phenomena that are impossible to replicate or observe in a laboratory setting with prescribed tasks or data that is not personally relevant to the individual. Though this form of evaluation is time consuming and may span months and years following the deployment of a visualization tool, the findings can reveal the impact of visualization as it is used in practice, and can serve to validate or refute initial hypotheses and conventional wisdom.

Visualization in the Newsroom: Storytelling and Investigative Analysis

*TimeLineCurator is a visualization tool for rapidly wrangling and curating a timeline dataset extracted from unstructured text [**VAST15**].*

One of the most vibrant and exciting application domains for visualization is that of journalism. Over the past decade, large newsrooms have recruited visualization practitioners to design compelling visual stories for the news-reading public, raising the public awareness of this powerful storytelling medium. In smaller newsrooms that do not have the resources or staff to perform custom visualization development, commercial visualization tools and web-based chart libraries have made it possible to embed a set of commonly-used interactive charts into online news stories. However, these tools have limited expressivity and require clean and formatted data; since journalists work on tight deadlines, these are significant limitations. We recently addressed the problem of tedious data wrangling and formatting in the context of visualizing timeline data: our **[VAST15]** paper presented TimeLineCurator, a tool for rapidly curating and visualizing timelines, in which structured categorical event data is extracted from unstructured text documents and plotted along a visual timeline. This project was directly motivated by journalists who were unable or restricted in their ability to visualize timelines in an efficient manner. TimeLineCurator was deployed and promoted within the journalism community in early 2015 and we have received positive feedback based on early usage.



*Overview is a document collection visualization tool for investigative journalists [**InfoVis14**].*

As the use of visualization for presentation and storytelling in journalism increases, so too does the use of visualization for behind-the-scenes investigative analysis. The Overview project **[InfoVis14]** is one such instance of a visualization tool that is intended to support journalists' analysis work, again with consideration for their deadline-driven schedule. The skill set of journalists and especially so-called "data journalists" is rapidly evolving, but there remains a wide range in terms of technical expertise, and thus there is a need for accessible and interactive visualization analysis tools to support investigative journalism.

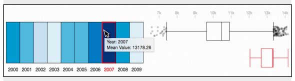
Future research: In terms of visualization for presentation and storytelling, journalists will continue to be in need of expressive and efficient tools. TimeLineCurator **[VAST14]** addressed the efficiency issue for a particular form of data. Expressivity remains a challenge, one that I plan to address through the design and evaluation of visual storytelling authoring tools. In terms of investigative analysis, the best way to determine what tools journalists need is to continue to network and speak with them about their workflows, a process that I am familiar with as a result of my role in the Overview project **[InfoVis14]**. Overview addressed text-based data; tool support is also required for the other forms of data that journalists encounter, providing us with many opportunities for applied visualization design. Ideally, our design solutions will generalize to other domains that involve both storytelling and investigative analysis, including the digital humanities, public administration, and business intelligence.

Visualization Technique Design Space Exploration & Ensuring Learnability

One of the additional outcomes of applied visualization design work is motivation for new technique development. The applied work described above has inspired me to develop novel yet learnable approaches to visualizing rankings, distributions, and timelines.

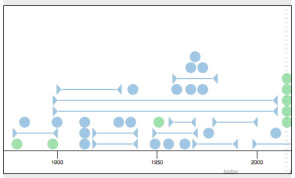
*SoundConsensus is a music review visualization that encodes rank, absolute value, and rank consensus [**SC**].*

Visualizing rank and rank change: In our **[InfoVis15]** paper, the energy analysts that we worked with oversaw the energy use of many buildings, and they had to decide which buildings should be prioritized for energy efficiency retrofits. To do so, they considered a ranking of buildings based on a single aggregate energy performance metric. Consider the case where ranks are computed based on annual average energy consumption values, the resultant ranking may differ from one year to the next, and the ranking may be differ substantially at a seasonal or monthly granularity. Thus we considered ways to visualize multiple rankings. Bump charts and slope graphs are existing visualization techniques that conveys rank change over time, however they do not directly encode the absolute value that corresponds to each rank position, a value that is important to consider along with relative rank. Drawing from recently proposed techniques in the visualization community, we began to explore the design space of rank-based visualizations and generated designs that encode rank, rank change, and rank consensus (or rank variability), shown alongside absolute values. In addition to implementing these techniques in the energy domain, I have also applied these techniques to university rankings **[UR]** and music reviews **[SC]**. I will continue to explore the design space of rank-based visualizations and seek out representative data.



*Juxtaposed time-series grid and boxplots, interactively linked with brushing [**InfoVis15**, **BP**].*

Visualizing distributions: Another question posed by energy analysts during their work relates to whether a building's energy performance is "normal" or "anomalous". To address this question, analysts have resorted to using line charts and bar charts with overlaid or juxtaposed estimated or baseline values. We learned that these estimated values are not always trusted; the analysts would have preferred to compare observed values to historical values, but there was no way of doing so with their bar and line chart visualizations without resorting to excessive panning or toggling between multiple visualizations. In response, we designed compact visualizations of distributions based on box-and-whisker plots (boxplots). Conventional boxplots are static charts, so we extended them with interaction and juxtaposed them with a grid-based time-series visualization that allows for the user to quickly compare distributions to identify normal and anomalous values **[InfoVis15, BP]**. As with visualization of rankings, I will continue to explore the design space of interactive boxplots and other graphical depictions of distribution.



*A timeline visualization in TimeLineCurator [**VAST15**].*

Visualizing timelines: Our experience with the TimeLineCurator project **[VAST15]** revealed that the set of existing visualization tools for presenting and analyzing timelines does not permit the designer or analyst much leeway in terms of visual expressiveness and interactivity. Depictions of categorical events occurring at points in time have been represented in many ways over the course of human history, from pictographic scenes spanning walls and columns in Ancient Egypt and Greece to the chronological tabular ledgers set down during the middle ages to the simplified linear charts of Joseph Priestley in the 18th century. More recently, designers of static information graphics have expanded the design space of visual timelines, making use of visual forms such as circles and spirals to reveal cyclical patterns of events. In future work, I hope to capitalize on this large design space and develop alternatives to the set of existing visualization tools that address timeline data, providing storytellers as well as analysts with more options in terms of visual forms and interactivity, including the selective annotation and progressive disclosure of events along the timeline.

Future research: In a controlled laboratory evaluation, it is possible to demonstrate the perceptual effectiveness of a novel visualization technique over previous techniques. However, this result does not guarantee adoption by practitioners, nor does this result guarantee that an audience will be able to learn and interpret the novel visualization should it be adopted into a presentation. In order to increase the likelihood that these novel visualization techniques are understood, we need to consider and design approaches to introducing and teaching novel visualizations. “Novel" is relative, as individuals working in some application domains may be accustomed to working with a wider repertoire of visualization techniques. Juxtaposing while interactively linking and brushing between a more familiar visualization and a novel visualization is one way to foster the interpretation of the novel visualization. In our work with energy analysts described above **[InfoVis15]**, we found that by juxtaposing and linking two novel visualizations without a familiar visualization was more effective than presenting either of these novel visualizations in isolation.

I will continue to examine the role of juxtaposition and interactive brushing between familiar and novel visualizations. Other researchers have also recently explored the role of animated transitions between familiar and novel visualizations, and comparing the effectiveness of juxtaposition versus animated transitions is an interesting research direction to pursue for a variety of data types, analysis tasks, and presentation intents.

References

Note: copies of all publications listed below are available at cs.ubc.ca/~brehmer/

[InfoVis15] **Matches, Mismatches, and Methods: Multiple-View Workflows for Energy Portfolio Analysis.**
Matthew Brehmer, Jocelyn Ng, Kevin Tate, & Tamara Munzner.
In *IEEE Trans. Visualization and Computer Graphics (TVCG) / Proc. InfoVis 2015, 22(1), p. 449–458.*

[VAST15] **TimeLineCurator: Interactive Authoring of Visual Timelines from Unstructured Text.**
Johanna Fulda, Matthew Brehmer, & Tamara Munzner.
In *IEEE Trans. Visualization and Computer Graphics (TVCG) / Proc. Visual Analytics Science & Technology (VAST 2015), 22(1), p.300–309.*

[InfoVis14] **Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool For Investigative Journalists.**
Matthew Brehmer, Stephen Ingram, Jonathan Stray, & Tamara Munzner.
In *IEEE Trans. Visualization and Computer Graphics (TVCG) / Proc. InfoVis 2014, 20(12), p. 2271–2280.*

[BELIV14b] **Pre-Design Empiricism for Information Visualization: Scenarios, Methods, and Challenges.**
Matthew Brehmer, Sheelagh Carpendale, Bongshin Lee, & Melanie Tory.
In *Proc. ACM Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization (BELIV 2014), p. 147–151.*

[BELIV14a] **Visualizing Dimensionally-Reduced Data: Interviews with Analysts and a Characterization of Task Sequences.**
Matthew Brehmer, Michael Sedlmair, Stephen Ingram, & Tamara Munzner.
In *Proc. ACM Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization (BELIV 2014), p.1–8.*

[InfoVis13] **A Multi-Level Typology of Abstract Visualization Tasks.**
Matthew Brehmer & Tamara Munzner.
IEEE Trans. Visualization and Computer Graphics (TVCG) / Proc. InfoVis 2013, 19(12), p. 2376–2385.

[UR] **UniversityRanks**
Matthew Brehmer (2014).
http://people.cs.ubc.ca/~brehmer/demo/universities/

[SC] **SoundConsensus**
Matthew Brehmer (2014).
http://mattbrehmer.github.io/SoundConsensus/

[BP] **Juxtaposed Focus + Context Box Plots**
Matthew Brehmer (2014).
http://bl.ocks.org/mattbrehmer/287e44c9a12151967874