JESSICA HULLMAN – RESEARCH STATEMENT

By leveraging our perceptual faculties, information visualizations improve our ability to understand and make decisions from data. My research focuses on enhancing the value that statistical graphics and other communicative visualizations can have for conveying data to broad audiences. New technologies for production are needed to support widespread visual communication of data, as the complex design processes and tacit knowledge that professional designers use to produce custom visualizations do not scale to the large amounts of data online. Successful design practice also requires anticipating how an audience will interpret a data visualization. Yet much remains to be learned about how "extra-representational" aspects of interaction affect interpretation—from the presence of other media to social factors to the cognitive tendencies of users. I **design, implement, and evaluate systems and approaches that combine knowledge of visualization design and interpretation to enhance the use of data visualization for presentation and communication**. This entails studying the characteristics of successful communicative visualizations along with evidence of their design, and demonstrating principled approaches for modeling and automating visualization practice. My recent research explores **modeling and automating the design of communicative visualizations for providing context for data and text,** and **examines interpretation** as mediated by cognitive and extra-representational factors.

My methods extend primarily from HCI, visualization, and data mining, but I complement these approaches with findings and experimental methodologies based in individual and group psychology, decision science, and design research. The novel visualization techniques I propose are informed by close study of professional visualization practice and controlled studies of contextual dependencies in interpretation. My focus on online practice makes my research outcomes useful for news content providers, professional and amateur visualization designers, and designers of visualization systems.

SUPPORTING THE DESIGN OF *CONTEXTUALIZING* VISUALIZATIONS

A key affordance of communicative data visualizations is an ability to provide critical *context* when they appear with news and other text articles, by summarizing relevant data and emphasizing information via text annotation. Yet the costs of professional design resources limit the appearance of these visualizations to only a handful of articles and news sources among the many available online. My recent work has explored automating the production of customized interactive visualizations to **augment large numbers of online news articles with visualizations that provide context**.

For example, I led the creation of the *Contextifier* system [1], a tool that demonstrates automated, annotated visualizations for financial news articles (Fig. 1). Contextifier leverages a corpus of articles to automatically create and annotate customized stock visualizations to

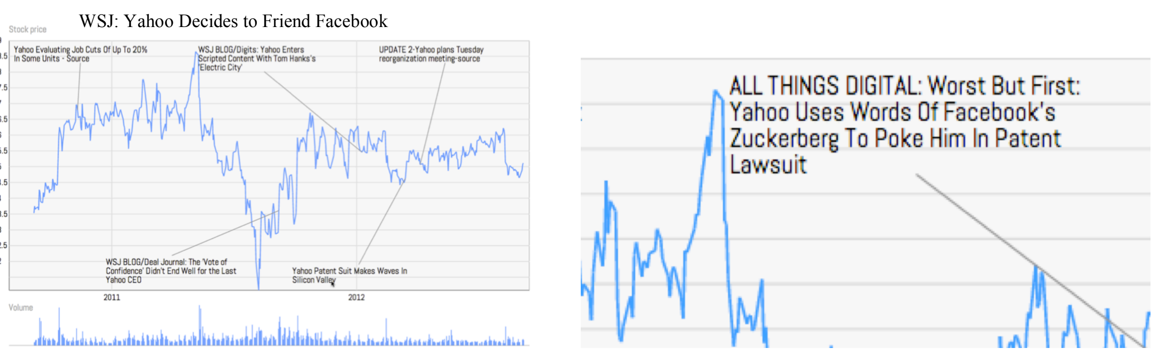

Fig. 1: The Contextifier system automatically produces annotated visualizations to accompany news (left), selecting annotations likely to be relevant to an article (right) and related to salient data.

accompany online news about companies. Annotations are selected to add topical information given the discussion in the original article, and to explain visually-salient points in the stock series, mimicking strategies that we observed designers using in a large sample of New York Times and Guardian visualizations.

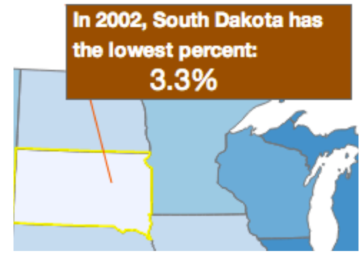

Fig. 2: Close-up of an annotation added to an extreme value in a map produced automatically by NewsViews.

Contextifier demonstrates how a broader perspective on a company's current and prior performance can be supported via automatically generated annotated visualizations. A more general principle of effective human-generated contextualizing visualizations is the way that they enable users to make relevant data comparisons implied by a text article. My collaborators and I created the *NewsViews* system [2] to explore how visualizations could be produced in bulk to support common comparisons implied in text articles, including geographic comparisons (Fig. 2). NewsViews' combines novel algorithms and cartographic best practices to develop important criteria for automatic, annotated map generation, which we realize in a generalizable design pipeline. This work experimentally validated key features for successful contextualizing visualizations, including relevance to available contextual information, variable selection, visual "interestingness," and annotation quality.

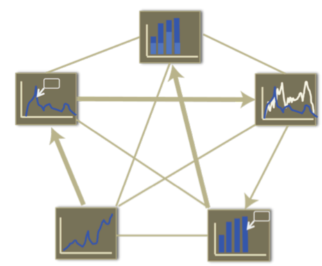

Fig. 3: A graph-based algorithm applied to visualization transitions can help designers find effective designs for presenting a set of visualizations.

Visualizations are also frequently presented as *sets* of related graphics that provide context to a multi-faceted data set, such as interactive and static slideshow visualizations or narrative-style reports created by researchers and other data workers. I collaborated with Nicholas Diakopoulos to develop a framework for understanding **how these "narrative" visualizations prompt particular interpretations of data** via rhetorical strategies that interact with aspects of the interpretation context like cultural knowledge [3]. My dissertation research characterizes inevitable design trade-offs that affect visualization interpretation among broad audiences. I **propose a graph-based approach for modeling and optimizing the design of popular "sequenced" narrative visualizations** like animations or slideshows [4] (Fig. 3). This model, which directly addresses a gap in current visualization systems around designing *sets* of visualizations as well as singular graphics, is based in insights gained from close examination of professional artifacts and validated via controlled user studies. This work demonstrates how properties of effective narrative visualizations can be modeled and measured, which is a theme in much of my work on communicative visualization.

VISUALIZING UNCERTAINTY

A common trade-off that must be addressed in designing visualizations for online audiences is the need to balance accurate presentation of the data as an *approximation* of a real-world phenomena, versus designing a visualization that can be understood by the majority of users. Uncertainty remains a challenging concept for most individuals to grasp [5]. Expert scientists and novice analysts alike often

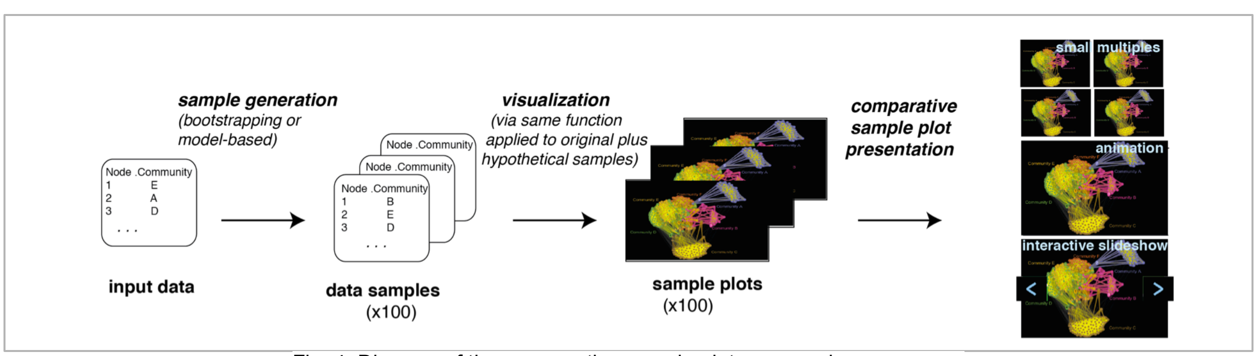

Fig. 4: Diagram of the comparative sample plots approach.

fail to accurately translate common graphical symbols of uncertainty, like error bars showing confidence intervals around predicted values in a bar chart [6]. Additionally, many complex graphical formats, like network diagrams or choropleth maps, lack a common visual uncertainty representation despite their prevalence online today. My recent research has outlined an alternative, generalizable method to visualizing uncertainty called *comparative sample plots* [7] (Fig. 4). The comparative sample plots procedure generates, visualizes, and presents *hypothetical samples* to visualization users directly as small multiples, animations, or via interactive formats. The technique is based on a hypothesis that the ability to directly make *visual comparisons* between multiple realizations of data improves users' abilities to estimate forms of uncertainty like reliability. We find that in multiple cases, this method improves users' abilities to judge *where* a given outcome is located relative to a distribution. In my dissertation, I evaluate the technique's usefulness for complex formats like network diagrams depicting predicted communities in a social graph like a friendship network (Fig. 4), maps displaying predicted voting outcomes (Fig. 5) that are prevalent in the popular press around election times, and treemaps. Compared to a single baseline data visualization, comparative sample plots reduce overconfidence in estimates of how reliable a perceived outcome is by upwards of 35%. The technique can also reduce the likelihood of misinterpreting visualized data values by over 50%.

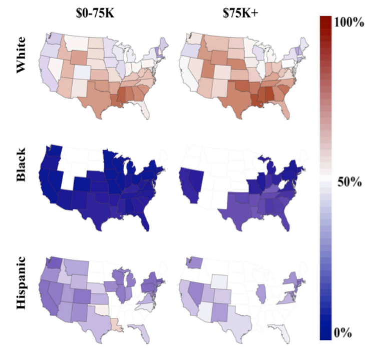

Fig. 5: A choropleth map grid of voting patterns by ethnicity (rows) and income levels (columns) presents challenges for uncertainty glyphs like error bars.

A collective agent, like a crowd of citizen analysts examining data to help a researcher evaluate a hypothesis, can also benefit from the production of multiple hypothetical samples. In a recent project [8], I investigated mechanisms for distributing visualized data to a group of crowd workers for analysis. These include presenting a single visualization to each crowd member, presenting small sets of related visualizations to each member (Fig. 6), and presenting a single, randomly drawn hypothetical sample plot to each member. This work exposes the negative impacts to crowd accuracy that can result from order effects in cases where each member analyzes multiple visualizations. It also surfaces the potential for decreased crowd accuracy due to individuals' oversensitivity to the features and framing of a singular visualization.

Fig. 6: Final judgments of the number of sources of a fatal disease differed based on the viewing order of the plots (in which deaths appear as red dots).

OTHER WORK

I have also contributed findings related to **collaborative visual analytics**, such as the influence of social information on visual judgments as simple as comparing the lengths of bars in a bar chart [99]. This work demonstrated how individuals can be easily influenced by prior users' judgments, even when those judgments were biased. Insensitivity to the number of prior responses presented among those who used the social information suggests that dynamics resembling information cascades can reduce the group judgment accuracy based on only a few initially-biased responses. By better understanding influence dynamics that affect interpretation, we can design mechanisms for collaborative visualization systems that reduce the likelihood of socially-biased judgments, like algorithms for withholding initial signals or varying design properties to balance systematic biases associated with particular visualization designs. I am currently exploring how end-users who comment on multimedia visualization news presentations distribute their attention across multiple representations and aspects of context. This work aims to enable sophisticated commenting technologies that better support comment re-use in social environments.

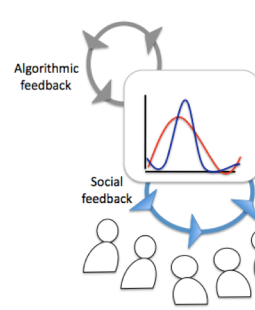

Fig. 7: Social feedback (i.e., comments) influences visualization users' judgments. Algorithms can adjust the presentation based on social signals.

LONG TERM RESEARCH AGENDA

I will continue to contribute approaches and interactive systems that support communicative visualization practice and data-based reasoning, based on a deep understanding of how visual features interact with users' cognitive processes.

Context-adaptive interactive visualization

I will use my future work to demonstrate how canonical approaches to the automated design of *singular* visualizations (e.g., [10]) can be adapted to support more efficient design of *sets of* visualizations created for presentation. I see this goal, in conjunction with continuing work toward understanding how visualization interacts with text, as a critical next step in order for visualization research to progress toward tools that support storytelling and context provision.

In addition to my research around communicative visualization production, I believe there is a need for *context-sensitive* interactive tools and approaches in presenting visualizations online. For example, features like animated transitions and text annotations that are used in communicative visualizations are not typically adjusted for a particular user's interaction trajectory. My future research will explore the design space of "intelligent visualization interactions" that adjust the presentation to support comprehension by modeling a user's developing understanding of visualized data.

Bridging the semantic gap

A gap exists in the visualization toolset when it comes to understanding how to support semantic processing of presented data among diverse audiences of end-users. Complex cognitive processing occurs as users compare their existing mental models to presented information, informed by "top-down" factors like learning strategies or prior knowledge [11, 12]. My future work will contribute knowledge and applications that target the gap between low-level interpretation of visualization features (such as axes or visual mark decoding), and the conclusions drawn by an end-user. I intend to embark on future study aimed at providing solutions to help people understand unfamiliar scales that are applied to data, such as very large or small magnitudes. I will also study and design techniques for leveraging graphical formats and framing to reduce the likelihood that users jump to causal conclusions about correlated data, such as the associations that are often discussed in media reports (e.g., "drinking beer is associated with good health").

Visual statistical reasoning

My research in visualizing uncertainty and interest in data reasoning [13] motivate future investigation of *how* users develop statistical understandings. For example, I am interested in how visualization users infer notions of data reliability and "fit" even in the absence of explicit uncertainty representations or statistics background. I will conduct future study informed by observations from my prior work that suggested that individuals rely on differences in the magnitude of data signals to infer reliability in the absence of explicit information. The knowledge gained from controlled studies of the heuristics that drive common visual-statistical judgments will inform the development of design interventions for improving data reasoning among diverse end-user audiences.

CONCLUSION

Visualization is indispensable in today's data intensive society. My work at the intersection of information visualization and HCI contributes solutions that enhance and help to scale effective visualization production and insight generation among broad audiences of users. My research has implications for a number of data and tool providers, and has attracted the attention of industry stakeholders in addition to academic awards. My focus on supporting practice through empirically-derived understanding of design and interpretation ensures the continued relevance of my research trajectory. As I transition into a faculty role, I believe I can help shape understanding in communicative visualization practice including automated visualization production, visual statistical reasoning and social interpretation.

REFERENCES

1. **Hullman, J.**, Diakopoulos, N., and Adar, E. Contextifier: automatic generation of annotated stock visualizations. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, ACM (2013), 2707–2716.
2. Gao, T., **Hullman, J.,** Adar, E., Hecht, B., and Diakopoulos, N. NewsViews: An automated pipeline for creating custom geovisualizations for news. Under review.
3. **Hullman, J.** and Diakopoulos, N. Visualization rhetoric: Framing effects in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 17*, 12 (2011), 2231–2240.
4. **Hullman, J**., Drucker, S., Henry Riche, N., Lee, B., Fisher, D., and Adar, E. A deeper understanding of sequence in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 19*, 12 (2013), 2406–2415.
5. Tversky, Amos, and Daniel Kahneman. Belief in the law of small numbers. *Psychological Bulletin 76*, 2 (1971), 105.
6. Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological methods 10,* 4, 389.
7. **Hullman, J.**, Adar, E., Resnick, P., Ghitza, Y., and Gelman, A. Comparative sample plots: Visualizing uncertainty for complex visualizations. Under review.
8. **Hullman, J.**, Krupka, E., and Adar, E. Evaluating approaches to crowdsourced visual analytics: The benefits of a distributed resampling mechanism. In revise and resubmit for *ACM TOCHI*.
9. **Hullman, J**., Adar, E., and Shah, P. The impact of social information on visual judgments. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2011), 1461–1470.
10. Mackinlay, J. Automating the design of graphical presentations of relational information. *ACM Transactions of Graphics (TOG) 5,* 2 (1986), 110-141.
11. **Hullman, J.,** Adar, E., and Shah, P. Benefitting infovis with visual difficulties (Honorable Mention Award). *IEEE Transactions on Visualization and Computer Graphics (TVCG) 17*, 12 (2011), 2213–2222.
12. **Hullman, J.** How prior knowledge affects the processing of visualized data. Short paper at *Many People, Many Eyes* (workshop), *ACM CHI 2013*.
13. **Hullman, J**., Rhodes, R., Rodriguez, F., and Shah, P. Research on graph comprehension and data interpretation: Implications for score reporting. In Zapata, D., & Zwick, R. (Eds.) *Test Score Reporting: Perspectives from the ETS Score Reporting Conference.* Princeton, NJ: Educational Testing Service, 2011.