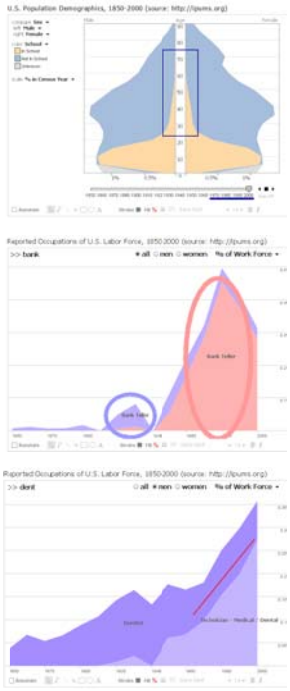
TECHNOLOGY-MEDIATED SOCIAL PARTICIPATION

Jeffrey Heer Stanford University



Annotated views from social data analysis in sense.us.

1. The rise of adult education from the 1970s onward
2. Reversal of the dominant gender of bank tellers
3. Stratification of dentistry into dentists and hygienists

Analysts in all areas of human knowledge, from science and engineering to economics, social science and journalism are drowning in data. New technologies for sensing, simulation, and communication are helping people to both collect and produce data at exponential rates [4]. As a result we must rethink how we design the tools and techniques for exploring, analyzing and communicating this abundance of data.

Throughout the data lifecycle, such *sensemaking* is often a collaborative process. As different analysts contribute to data acquisition, cleaning, analysis, and interpretation they contribute contextual knowledge that deepens understanding. Studies of information workers [5] have demonstrated that sensemaking is highly iterative, as an insight gained from a visualization may suggest the need for additional corroborating data or highlight a data cleaning error. At times people may disagree on how to interpret data, but then work together to reach consensus. Many documents and data sets are so large that thorough exploration by a single person is unlikely. Moreover, social interaction depends on an understanding of social context, including the skills, inclinations, past actions, and relationships among collaborators [2]. The distributed and often asynchronous nature of computer-mediated collaboration, as well as the sheer volume of data and people at hand, complicates the process of making sense of the social environment.

I am particularly interested in the design and evaluation of techniques for facilitating collaborative sensemaking. New tools that address social interaction and data analysis in an integrated fashion are an important class of social participation technologies and are central to improving our ability to turn data into knowledge. Such tools may impact diverse fields including business, intelligence, science, politics, and public policy.

SOCIAL DATA ANALYSIS

To explore the potential of incorporating social interaction with visual analysis, we built Sense.us, a web application for collaborative sensemaking of 150 years of United States census data [3]. Sense.us integrates visualizations of demographic data with features for collective analysis. Users can attach commentary and annotations to views, share collections of views, and engage in discussion. Novel bookmarking and indexing features facilitate view sharing and reduce cross-talk between related visualization states.

We studied usage of the system through a live deployment and a series of laboratory studies, and conducted a content analysis of recorded usage. We found that users often combined their knowledge in cycles of observation and hypothesis to make sense of trends in the data. For example, one observer noted a decline in the number of dentists in the labor force. Other people then hypothesized possible explanations, including fluoridation of the water supply and an increasing stratification between dentists and hygienists over the last century. In other cases, users explored topics such as changing gender roles, the impact of technology on the job market, and correlations among the wax and wane of occupations. We observed that social features helped mobilize users in the process of identifying interesting trends and generating hypotheses, and that exposing social activity regularly catalyzed new explorations by collaborators.

Building on these observations, recent attention has focused on the critical role of social cognition in the process of data analysis, investigating tools for sharing, annotation, and group deliberation of visualized data. We are currently witnessing a flowering of data sharing and visualization sites on the web, including web services such as IBM's Many-Eyes [6], Swivel.com, Data360.org, and Google Fusion Tables; commercial products such as Spotfire Decision Posters and Tableau Server; and a variety of public data repositories including data.gov and usaspending.gov.

FUTURE DIRECTIONS

Despite these initial successes, social data analysis tools are still in their infancy and there is much we do not understand. How can we best support people as they gather and curate information, engage in social interpretation and deliberation, and then disseminate, verify, and build support for identified findings?

For example, the activity and commentary of many actors is itself an unruly data set. How can the contributions of various actors be synthesized in a manner that supports group sensemaking? Sites such as Wikipedia rely on human editing to integrate contributions, whereas structured approaches such as Luis von Ahn's "games with a purpose" afford statistical aggregation. One engineering and design effort is to explore systems that lie on the spectrum between these examples, providing representations (e.g., of hypotheses and evidence [1]) that facilitate aggregation without unduly inhibiting communication. A related issue is how to combine the efforts of human and machine collaborators. How can

human activity guide automated processes, and conversely, how should automated techniques better suggest interesting observations or action items?

I am interested in discussing both research and educational programs that will enable our data-deluged generation to address these and other questions. Topics of interest include:

- **Collaborative information foraging**. How do groups most effectively collect, curate, and annotated information resources?
- **Assessment of mediated social environments**. How does one make sense of social context in networked systems? How do visualized activity cues affect one's actions?
- **Discussion and synthesis**. How can user contributions be better integrated?
- **Dissemination and influence**. How do findings and opinions disseminate in social media, what factors affect their influence, and how might this be shaped by tools?
- **Research methodology**. What are the most appropriate methods—or combinations of methods—to empirically assess these issues?

REFERENCES

1. D. Billman, G. Convertino, J. Shrager, P. Pirolli, and J. P. Massar. Collaborative intelligence analysis with cache and its effects on information gathering and cognitive bias. In *Human Computer Interaction Consortium*, 2006.
2. J. Caroll, M. B. Rosson, G. Convertino, and C. H. Ganoe. Awareness and teamwork in computer-supported collaborations. *Interacting with Computers*, **18**(1):21–46, 2005.
3. J. Heer, F. B. Viégas, and M. Wattenberg. Voyagers and voyeurs: Supporting asynchronous collaborative information visualization. *Proc. ACM CHI*, pp. 1029–1038, 2007.
4. P. Lyman and H. Varian. How much information? 2003. http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/.
5. P. Pirolli and S. K. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. *Proc. of International Conference on Intelligence Analysis*, 2005.
6. F. B. Viégas, M. Wattenberg, F. van Ham, J. Kriss, and M. McKeon. Many Eyes: A site for visualization at internet scale. *IEEE Transactions on Visualization and Computer Graphics*, **12**(5):1121–1128, Nov/Dec 2007.