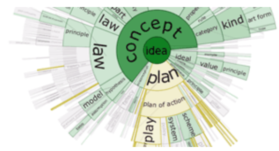
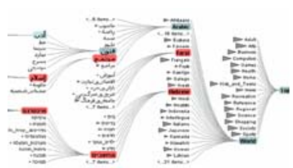
RESEARCH STATEMENT

Jeffrey Michael Heer



Visualization of word relations in the WordNet database, built using our open-source tools.

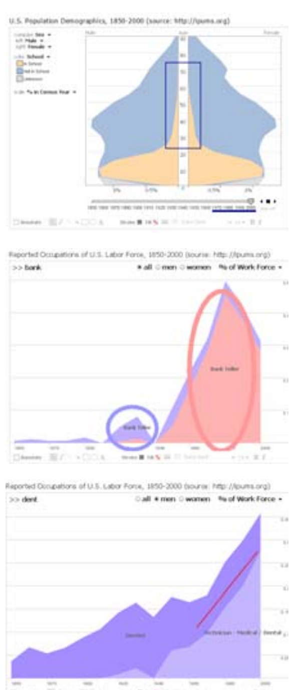
New technologies for sensing, simulation and communication are helping people collect and produce data at exponential rates. Yet acquiring and storing this data is, by itself, of little value. We must make sense of the data in order to produce real value from it. Such *sensemaking* — turning data sets into knowledge — is a fundamental challenge for both computer systems and human-computer interaction research. It requires integrating data analysis algorithms with human judgments about the meaning and significance of patterns in the data. **The goal of my research group is to enhance our collective ability to analyze and communicate data through the design of interactive visual analysis tools**.



Degree-of-Interest Tree visualization of dmoz.org.

My research in *visualization techniques* uses insights from studies of human perception and cognition to design visual representations and interaction techniques for data. We first conduct perceptual studies to characterize the effectiveness of visual encoding choices. We then apply the findings to create new visualizations and automated design methods. For example, our "degree-of-interest" trees build a model of user's interest across a data set using input such as search queries and items clicked. This model then determines what information is shown and how it is displayed, smoothly animating the visualization in response to a user's changing focus of attention. We have applied this technique to explore data sets containing on the order of a million items, including the analysis of time-varying hierarchies such as the political and military organization charts of nations.

Of course, improved visualization techniques are of little use if they never make it into the hands of designers and developers. My group also researches *software architectures* that support the design of novel, customized visualizations. With these frameworks, developers can rapidly craft visualizations by composing fine-grained "building blocks" for visual encoding and interaction. This research has produced open-source toolkits that have been downloaded over 100,000 times and are widely used across academia and industry. Both myself and others have used these systems to build visual analysis applications for domains such as protein interaction pathways, census data, social networks, and the Enron e-mail corpus.



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

By deploying such systems, we can also study how users conduct real-world analyses. For example, we observe that analysis is often a social process: the magnitude of data and the diversity of expertise needed to fully analyze it require that our information interfaces enable us to work together to effectively forage, analyze, point, argue, and disseminate. Our research on *collaborative visual analysis* explores how interfaces can catalyze social interpretation and deliberation. One such system is *sense.us*, a web site for social exploration of 150 years of U.S. census data. Our studies of system usage found that social features can improve hypothesis generation and that awareness of others' activity catalyze new explorations by collaborators.

We also observe that analysts spend much of their time cleaning and reformatting data to make it suitable for analysis. As a result, domain experts often spend more time manipulating data than they do exercising their specialty, while less technical users (e.g., non-programmers) may be excluded. In response, we are investigating new methods for *interactive data cleaning and transformation*. With our *Wrangler* system, users construct data transformation scripts in a direct manipulation interface. Wrangler significantly accelerates data cleaning by using novel programming-by-demonstration techniques to automatically suggest applicable transforms and visualize their effects. The result is not simply transformed data, but a reusable program that can be run on other platforms (e.g., a MapReduce cluster) to transform data at scale.

Our ongoing research seeks to further advance these themes by studying the perceptual, cognitive, and social factors underlying successful analyses and then designing improved interfaces to help us make sense of the increasing scale and diversity of available data.