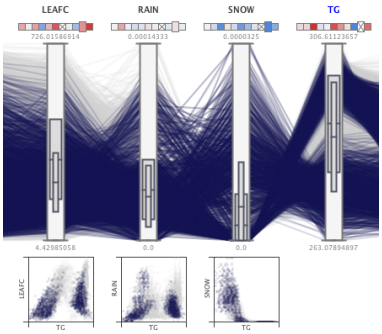
Statement of Research
Chad A. Steed

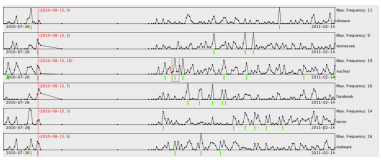
My primary research area is visual analytics—a multi-disciplinary field that combines automated analysis methods with interactive visualizations to amplify cognition in complex data sets. With so-called big data analysis, the volume, velocity, and sheer complexity of the data makes serendipitous discoveries difficult, at best. But a carefully designed visual analytics approach reduces knowledge discovery timelines by integrating visualization theory with other disciplines such as data mining, data management, human-computer interaction, and human perception and cognition. In my research, I explore these areas in a variety of application domains such as social media, text analysis, health care, climate study, intelligence, and military applications. My work focuses on data intensive, extreme scale analysis issues that are at the forefront of computer science.

The fundamental hypothesis of my research is the idea that intelligently combining the strengths of humans with computers is the most effective approach to understanding today's data. Indeed, given the exponential data growth, visual analytics is arguably the only viable solution to solving the critical technological challenges in our society. In terms of the volume of information and the speed of input, the human visual channel far surpasses all other human sensory inputs (about 70% of human sensory receptors are in the eyes and nearly half of the cerebral cortex is devoted to visual processing). Therefore, the effective combination of our high bandwidth visual channel, creativity, and background knowledge with the ever-increasing computational power of machines will facilitate the development of more intuitive frameworks that effectively guide the user to the most significant features. To this end, I strive to design and development interactive techniques that connect the user to the data behind the visualization to help them *see* the values, structure, and associations in the data set.

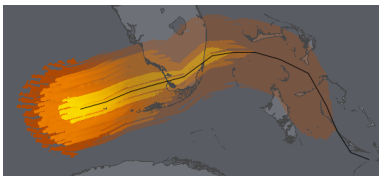
My Active Research Areas:



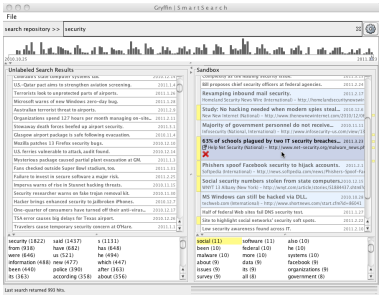
Multivariate Visual Analytics: One of the most challenging tasks in exploring multivariate data is the identification and quantification of the associations between a set of interrelated variables. For extreme scale, real-world data sets, such as those we encounter in investigations of climate simulations, this task is even more daunting due to the complexity and number of different variables under consideration. Often times, the user has some understanding of expected relationships; but unexpected discoveries are nearly impossible to achieve with conventional approaches. Despite these challenges, such tasks are paramount in activities such as uncertainty quantification and exploratory analysis for extreme scale multivariate data sets. My approach to this problem is to design intuitive methods to augment multivariate visualization techniques, such as parallel coordinates, with new interaction techniques and graphical statistical indicators. In a number of published case studies with domain experts, our interactive visualizations have facilitated the discovery of previously unknown relationships in complex multivariate data sets, thereby corroborating the notion that visual analysis techniques are superior to conventional approaches. Results of this type are the central promise of visual analytics research.



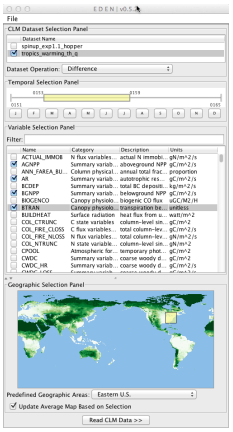
Text Visual Analytics: As we rush to meet the challenges of so-called "big data" analysis, we realize that the bulk of data (common estimates say over 80%) are stored as unorganized, free-form text which is difficult to parse and often incomplete. Finding relevant and unexpected nuggets of knowledge in a large text corpus remains a significant challenge. In my research, I have explored visual analytics solutions for text mining to enable near real-time situational awareness and exploration of social media streams and search-based investigative analysis of large document collections. My search-based techniques leverage intelligent user interfaces and intuitive focus+context and drill-down/roll-up capabilities. In my research on social media streams, I have devised new spatio-temporal techniques that are designed to work with high velocity, high throughput streams to highlight surprising events, trends, and abnormalities. These visualizations are augmented with information from sentiment analysis, statistical analytics, and topic modeling algorithms to reveal additional insight. The results of these activities have proven effective for intelligence, cyber security, biomedical, and biosurveiliance.



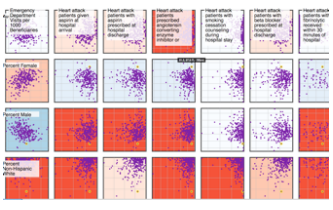
Illustrative Visualization: The advancements in geographical tagging of information combined with increased dimensionality places severe limits on the capacity of conventional maps and geospatial information display to foster creative analysis. Consequently, the user is forced to reduce the problem to fit the limitations of the tools, which results in eliminating potentially useful parameters. To address these challenges, I have devised new geographical visualization techniques to better support interactive visual analysis of geospatial patterns and trends through the use of illustrative rendering techniques from the overarching field of computer graphics. These illustrative visualizations are inspired by artistic brush strokes to convey the complexity of the information using icons, geometric shapes, textures, and other visual features. Such visualization techniques facilitate the encoding of up to 8 variables in a single display, thereby alleviating the perceptual challenges caused by short term visual memory. In addition, this area of my research allows me to draw from my passion and training in drawing, painting, and graphic design to build efficient and aesthetically pleasing displays.



Intelligent User Interfaces: As the complexity of information increases, intelligent user interfaces are needed to assist users in the analysis process. An intelligent user interface adapts to the user based on interactions with the display. I have developed systems that use semi-supervised machine learning algorithms to re-rank document search results based on the user's interactions Given a list of textual items from a search, the user's interactions are captured and used to label documents as relevant or irrelevant for the given subject. These labeled items are then examined and used to re-sort the remaining unlabeled documents in a manner that moves potentially relevant items to more visible locations, thereby increasing the likelihood of finding relevant documents buried in the lower portions of a large search hit list. I am also particularly excited about the potential of such personalized interfaces in a variety of application areas.



Data Management for Visualization: A critical component of visual analytics is efficient data management. An integrated and consistent data model is a necessary precondition for any form of data analysis. Due to the growth of the Internet, database research has moved from a focus on uniform, structured data to heterogeneous data. Finding effective representations for different data types such as unstructured text, numeric data, graphs, audio and video signals, and images is a key challenge for today's database technologies. Heterogeneous data also require data cleaning–dealing with missing or inaccurate data values. Furthermore, new dynamic sources of data such as streaming sources and sensor networks account for an additional set of challenges. In my investigations, I devise data management solutions that employ intelligent data analysis techniques for fusing information and feeding visualizations. These investigations rely on utilizing new data management research and dynamic change detection methods in a manner that supports the full spectrum of analysis–from millions of items to the components of a single item.



Visual Analytics for the Web: Advances in web browser technologies have made it possible to move analysis applications into the Internet cloud and on leadership class supercomputers. As more applications are moving from thick client models to thin client approaches, several challenges must be addressed. For example, level-of-detail algorithms which were pioneered in early computer graphics systems are needed to seamlessly deal with the limited number of objects that can be accommodated in the web browser's document object model. I investigate ways to provide varying levels-of-data detail and analytics for the user's current "zoom" level into the data set. Furthermore, I am exploring collaborative and social interaction techniques to take advantage of simultaneous access by multiple users in the distributed analysis scenario inherent to the web application platform. This area of my research encapsulates advances that I have devised in these other areas to support visual analytics on the web.

Collaborative and Multidisciplinary Research:

In my career, I have enjoyed mutually beneficial relationships with several leading domain scientists in cross-cutting research projects with both national and global implications. I am motivated to seek these collaborations because: (1) I believe the results have more relevance in directly contributing to the greater good of humanity; and (2) there is a growing gap between viable visualization techniques and real-world analysis. Although there have been many new visualization approaches introduced in the literature in recent years, relatively few of these approaches have been brought to bear in data-rich domains where they are needed the most. In contrast, my investigations have been successful in connecting these communities by simply responding to the needs of experts. For example, I have worked extensively with experts in climate modeling, intelligence, military, and health care services to formulate novel approaches for data analysis. The 2011 DOE ASCR Workshop on Exascale Data Management, Analysis, and Visualization Report[1] highlights the importance of such collaborations by recommending "close collaboration between science domain experts and data analysis experts." My collaborators frequently offer testimony to the success of these efforts and together we have written several publications on our results. In the future, I intend to continue these fruitful collaborations, as well as seek out new collaborators, especially those located at my home institution.

Future Research Plans:

I intend to continue these threads of research in a manner that continually weaves them together under a unified theme of a personalized visual analytics interface. I will continue to investigate advanced methods for machine learning and data management to improve the successful visual analytics systems that I have already devised. I also plan to focus more on the basic, fundamental science of human perception and cognition. I will do this by conducting formal evaluations of the techniques using empirical and/or insight-based evaluations, and I will integrate the resulting knowledge into theories of perception and cognition. Although the opportunity to plan and execute such evaluations is particularly enticing to me, such research activities are difficult in the national laboratory environment and will require partnerships with academia. In addition, I plan to collaborate with domain experts in the field of human-computer interaction, computer graphics, data mining, as well as educators and practitioners in fine art and graphic design.

[1] http://science.energy.gov/ /media/ascr/pdf/program-documents/docs/Exascale-ASCR-Analysis.pdf