Introduction

Foundations and Frontiers in Visual Analytics

Joe Kielman\textsuperscript{a}
Jim Thomas\textsuperscript{b,}\textsuperscript{*} and
Richard May\textsuperscript{b}

\textsuperscript{a}US Department of Homeland Security, Science and Technology Directorate, Washington, DC, USA.
\textsuperscript{b}Pacific Northwest National Laboratory, PO Box 999, K7-28, Richland, WA 99352, USA.

\textsuperscript{*}Corresponding author.
E-mail: jim.thomas@pnl.gov

This article is a product of a workshop on the Future of Visual Analytics, held in Washington, DC on 4 March, 2009. Workshop attendees included representatives from the visual analytics research community across government, industry and academia. The goal of the workshop, and the resulting papers, was to reflect on the first 5 years of the visual analytics enterprise and propose research challenges for the next 5 years. The article incorporates input from workshop attendees as well as from its authors.

This introduction and the future vision section for this special issue of Information Visualization hopes to set the stage for an emerging worldwide effort to advance the state of the science of visual analytics. We present some of the driving needs followed by some principles and methods for advancing this science through partnerships among national laboratories, academia, industry and the international science community. Also presented is a selection of the many successes the science, engineering and industrial communities have had in taking core scientific research to end users in the field during these early years. These stories are followed by some thoughts on frontiers and the future vision for visual analytics. Finally, we introduce the eight papers in this special issue, each one addressing part of that vision.

Background of Visual Analytics

The formation of the U.S. Department of Homeland Security (DHS) National Visualization and Analytics Center™ (NVAC™)\textsuperscript{1} in March 2004 resulted in increased interest in the field of visual analytics. In 2005, a diverse team of academic and laboratory researchers, government managers, and industry scientists turned a vision into a science direction – one published in the book \textit{Illuminating the Path: The R&D Agenda for Visual Analytics}.\textsuperscript{2} Shortly after that book’s publication, five university-led Regional Visualization and Analytics Centers (RVACs) were established at Stanford University, the University of North Carolina Charlotte with Georgia Tech, Penn State University with Drexel University, Purdue University, and University of Washington. Also, at that same time, many other researchers around the world were developing similar or complementary visions and offering new opportunities for collaboration. Special issues of magazines and journals provided early outlets for emerging research and applications within visual analytics.\textsuperscript{3–6} Also in 2005, NVAC began hosting semi-annual Consortiums to bring academia, industry and national laboratories together with end users, government sponsors and international partners to advance this new, potentially significant field of research.

To further build the scientific community, in 2006 IEEE launched the Symposium on Visual Analytics Science and Technology (VAST), the first international symposium dedicated to advances in visual analytics science and technology. Since then, several topical workshops have been held on financial analytics, composition and active products, and mathematic foundations of visual analytics. The latter topic set the stage for the
current joint National Science Foundation (NSF) and DHS program on the Foundations of Data and Visual Analytics (FODAVA).\textsuperscript{7} While all of these early efforts provided a research foundation in the United States, many other individual scientists and research groups were establishing visual analytics centers around the world and hosting conferences with visual analytics as a topic of interest.

Recently, DHS hosted a Future of Visual Analytics Workshop to look back and look forward within the science and the applications for visual analytics; participants are listed in the acknowledgements. Out of the workshop discussions, and the resultant topic teams formed, came many of the sources for the papers included in this Special Issue.

**Driving Needs**

The drivers for visual analytics are many and diverse. One of the primary reasons first advanced for the development and use of visual analytics techniques was the need to manage the massive amounts of data (or the so-called data deluge) overwhelming analysts. It has been estimated (University of California, Berkeley and the IDC) that by 2010 the amount of digital data produced annually will be almost 1ZB. Such numbers are staggering and demonstrate that we live, work and must survive in a digital world encompassing databases, documents, video, imagery, audio and even sensor data. This situation will only get more complex in the future, and how we navigate through all the diverse, diffuse data presented to us will depend on precisely honed selection, synthesis, analysis and communication tools.

Going forward, however, the problem is evolving from one of data deluge, from which selected nuggets must be extracted, to one where we are presented with an almost infinite landscape of indistinguishable data elements. Jorge Luis Borges described this situation best in his 1941 short story ‘The Library of Babel.’\textsuperscript{8} That notion of vast indeterminacy should be how we view the digital era rather than the notion of a tsunami of data. What this means is that we should be looking to techniques that provide ‘precise’ data, which is immediate, relevant and understandable to individual users, groups, or communities of interest.

Beyond the challenges presented by the scale and form of those data, there are others associated with their ultimate use or utility. Early in the development of the new science of visual analytics, it was clear that most of the tools and technology had limited use because of complex interfaces, limited access to data, single data type analytics, limited analytical reasoning support and little or no technology for capturing the analytic processes into tailored active decision packages. Thus, the visual analytics agenda\textsuperscript{a,d} described in *Illuminating the Path* was established as an initial science vision. That agenda has served us well and will be advanced as the technology and needs evolve. Nevertheless, new uses and users for visual analytics technologies are already evident and will drive much of the future work for this new science. For example, the users of visual analytics in the future will broaden from government analysts or analysts in general to a much more diverse set of professionals in public safety, business, health and manufacturing. The uses of visual analytics will also broaden and will be applied to much more difficult concerns. Terrorism and regional security had been the focus of much of the NVAC’s work in the early years, with some consideration for large, single-event natural or manmade disasters. Global security, with issues as diverse as climate change, resource scarcity, economic stability and the problems of ‘slow’-moving catastrophes associated with declines in food, energy and environmental health can all benefit from the proper applications of visual analytics techniques. The last section of this article, as well as the final article in this Special Issue, describe such topics in more detail.

**Success Stories: Research to Reality (R2R)**

In this section, we present success stories that can help those interested in this young science understand and fuel the potential impacts of the frontiers of visual analytics. Such stories lay the foundation for continued development and broader application of visual analytics science and technology. There are many measures of early success in the field of visual analytics. A simple one is the number of peer-reviewed papers that have been published. NVAC and the RVACs have published 222 peer-viewed papers since 2005; this number does not include all the wonderful visual analytics work done by others outside of the VAC enterprise. Over 100 invited plenary and keynote presentations have been given by NVAC and RVAC science leaders. Cooperative activities are another measure of success. DHS’s Science and Technology Directorate has formed formal partnerships with other government agencies such as NSF and entered into collaborative agreements with Canada and Germany, with many more coming.

An essential element in the rapid rise of visual analytics has been the ability to transition research to user applications and the development of community resources that drive continued advancement of visual analytics. More than 100 local, state, and federal agencies are already using visual analytics. Through the DHS/Pacific Northwest National Laboratory (PNNL) R2R Program, several of those technologies have been successfully commercialized. Developing tools and technologies to address real-world problems allows user communities to better understand and manage complex analytical tasks and provides researchers with valuable lessons learned, which in turn guide future research efforts. Five such success stories of the many that already exist today are briefly described below.
Foundations and Frontiers in Visual Analytics

Figure 1: Visual analytics supports emergency responders in the office and in the field.

Story 1 – Visual analytics for law enforcement

A central challenge in visual analytics is the creation of accessible, walkup usable, widely distributable analysis applications that bring the benefits of visual discovery to as broad a user base as possible. The Scalable Reasoning System (SRS)\(^9\) provides web-based and mobile interfaces for visual analysis through a service-oriented analytic framework (see Figure 1). The goal of SRS is the straightforward deployment of pervasive visual analytic environments that can be rapidly deployed in a platform-agnostic fashion across an enterprise. The SRS lightweight platform, which uses web distribution and simplified interfaces to visual analytics, is being used at three regional law enforcement agencies, by more than 10,000 users as part of their daily activities.

Users at the Port Authority of New York and New Jersey (PANYNJ), the Automated Regional Justice Information System (ARJIS) in San Diego, and the Seattle Police Department (SPD) are using and evaluating SRS visual analytics techniques to improve law enforcement practices and help save lives. Visual analytics tools allow officers to analyze the narrative information present in incident reports and case files; this information has been historically underutilized because it is not readily available to officers on the street and in command or analysis centers. SRS enables those officers and analysts to explore relationships in the information in fundamentally new ways. Furthermore, law enforcement officers deal with large quantities of heterogeneous data from multiple sources with varying levels of uncertainty and indeterminate quality. The decisions made by those officers directly and often immediately affect public safety and societal quality of life issues. Such decisions are either tactical, with only minutes or seconds available to analyze and respond to a situation, or strategic, where complex relationships are analyzed to help develop proactive plans to reduce crime and counter terrorism threats.

Developing an analytical tool suite to support law enforcement and counter-terrorism field units brought together different ongoing research threads in a single, yet powerful deployable application. Early work in the use of visual analytics on mobile appliances\(^11\) looked at how to provide easy-to-use interactions to visual analytics and how to better distribute quickly digestible knowledge to a wide range of users. Visual analytics researchers also had been looking at the importance of and challenges associated with decomposable reasoning artifacts and the distribution of knowledge through analytic environments that attempt to bridge the divides among information retrieval, information analysis and information dissemination tools.\(^10\) It is the combination of these research areas that drove the development of new analytical tools and the creation of SRS in order to support the demands of law enforcement professionals. Nevertheless, many of the analytical needs of law enforcement and counter-terrorism field units are common to those found in other domains. The lessons learned from the efforts for the PANYNJ, ARJIS and the SPD will inform future work in those domains.

Story 2 – Visual analytics for critical infrastructure protection

Visual analytics inherently requires interactivity between the analyst and the visual representations. However, maintaining interactivity becomes a significant or even impossible challenge as the quantity of information or the complexity of the system increases. Large semantic graphs or networks illustrate this challenge. Analysts use semantic graphs to organize concepts and relationships as graph nodes and links and as a way of discovering key trends, patterns and insights. The Have Green\(^12\) framework was designed to maintain interactive analysis of semantic graphs with up to one million nodes. Have Green fills the theoretical and developmental limitations found in many related systems by creating an analytical environment in which analysts (not researchers) conduct network analysis in terms and concepts that are intuitive and meaningful to them.

A variant of Have Green, GreenGrid,\(^13\) was created to allow engineers to explore and monitor the North American Electricity Infrastructure. The system, commonly referred to as the electrical grid is made up of generators, distribution centers, transmission lines, consumers and other elements that can be treated as a graph. For many years, the electrical grid has been limited to depictions as information on top of a geographic layout or a quasi circuit diagram. While these traditional visualizations are valuable tools, they are but a subset of the possible ways of depicting and exploring the possible relationships in the high-dimensional information space that is the electrical power grid. GreenGrid’s interactive exploration,
weighted-graph design and linked visualization approach represent a significant innovation in applying visualization in the electric power industry (see Figure 2). GreenGrid provides previously unavailable freedom for engineers to express the power grid layout in terms of variables of their choosing. GreenGrid was used to analyze the electricity blackout that occurred within the Western U.S. power grid on 10 August 1996. Comparisons of GreenGrid results with the conclusion of the post-disturbance analysis revealed that many of the disturbance characteristics could be readily identified using its advanced visual analytics tools.\textsuperscript{13}

**Story 3 – Visual analytics for financial fraud analytics**

In some analytical tasks, the anomaly being sought is remarkably similar to routine information that is not of interest. The challenge is to detect – in a timely and reliable fashion – those minute differences in millions of records covering arbitrary periods of time. This is the case with financial fraud analytics. Banks have the difficult problem of monitoring massive flows of transactions. They must also be sensitive to the frequent changes and adaptation of strategy that perpetrators of fraud use to avoid detection. When bank investigators see potentially fraudulent activity, such as money laundering, they must investigate further, spending considerable time to winnow out the few cases that necessitate action. This is a visual analytics problem requiring visual exploration, discovery and analysis.

Researchers created WireVis\textsuperscript{14} to combine the trade-craft honed by financial fraud analysts and advanced visual analytic techniques. A common approach used by financial fraud analysts involves keywords found in transactional records. WireVis augments this approach with a highly interactive system using linked visualizations that give the analysts simultaneous information overviews and detailed exploratory capabilities. WireVis uses a range of visualization techniques such as heatmaps and
network maps as a medium for interaction, letting investigators explore the relationships between keywords and transactions. The WireViz team augmented these capabilities with a new technique called *strings and beads* to better detect minute transaction anomalies (see Figure 3).

WireVis was created in close collaboration with Bank of America, which sees great promise for the use of visual analytical tools in the area of financial fraud analysis. Tools like these have potential for both exploratory analysis and training in a wide variety of transactional analytics tasks.

**Story 4 – Visual analytics for real-time situation assessment**

Large interaction displays provide opportunities for co-located synchronous collaboration by providing groups of people with walkup interfaces. Large interactive displays can also be designed as ambient information portals providing casual access to continually updated information through robust analytical capabilities usable by individuals and groups alike. Ambient displays located in gathering locations such as kiosks have been used for many years to provide vital information. This concept was expanded in a technology called *the Assessment Wall* (see Figure 4) to incorporate visual analytics techniques in appealing, walkup appliances, offering real time multiple source situation assessment. The system is deployed and being tested in a wide variety of analytical environments,
such as group interaction spaces and officer break rooms in law enforcement agencies. In all of them, the result is the same – the Assessment Wall helps users easily monitor and analytically explore continually updating streaming information sources through easy-to-learn, engaging, high-value interactions.

The Assessment Wall occupies a singular position among large displays for work and ambient systems. In an ambient capacity it conveys important themes that change with the data, persistent query matches and patterns in time, any of which can be perceived at a glance. Unlike many ambient systems, it also supports interaction that helps the discovery of relationships between concepts and temporal patterns without training or special equipment. With this balance of ambient and interactivity, novice users can maintain awareness of the data, dig deeper into topics of interest and use the system as a focal point for discussion and impromptu collaboration on work-related data.

Usability studies guided the initial Assessment Wall design choices, and the first installations at a large government client provided the opportunity for user feedback and observing the system in use. Several ongoing projects will use the Assessment Wall in operational environments to support a variety of missions. Continued operational use promises many future opportunities to support effective display and analysis while maintaining a lightweight and intuitive interface.

Story 5 – Better data + better evaluations = better science

As researchers develop visual analytic environments, it is vitally important to develop metrics and methodologies to help researchers measure the progress and understand their impact on the users who will work in such environments. It is equally important for the advancement of the science to have measures of ground truth to evaluate these metrics and methodologies. The Threat Stream Generator (TSG) is a new method and toolset for creating realistic, synthetic test data for evaluating visual analytics applications and techniques. The distinct advantage to the TSG sets is that known data, in the form of a pre-determined threat such as terrorist activity or a law enforcement concern, is created and translated into data cues. The cues are inserted into the data set with a known expressivity – the number of cues and their subtlety of representation are controlled. TSG has produced many data sets that are available to the visual analytics community for evaluation. Hundreds of researchers and analysts have used and relied on these data sets and an evolving understanding of evaluation methodologies to guide and accelerate the advancement of visual analytics science.

For the past 3 years, the TSG team and those developing evaluation methodologies have run the VAST Symposium Challenge. The purpose of the VAST Challenge is to provide representative tasks and data sets that participants can use to see how well tools work. As part of its format, the VAST Challenge provides an open and encouraging environment for researchers and analysts to work together in order to better understand each other. In 2008, the VAST Challenge had an unprecedented 73 entries from 13 countries (see Figure 5) and the data sets were downloaded by over 400 research groups from 36 countries.

Figure 5: VAST 2008 Challenge participants work through TSG data sets to find solutions.

Frontiers and Future Vision

We have described just a few of the successes of visual analytics above. Most concerned homeland security applications, as befitted the genesis of NVAC and the source of funding for its initial 5 years. The future for visual analytics will be much more interesting, in our view, and include many more and diverse applications. Several of these are as follows: Synthetic Materials, Disease Development, Economic Stability, Climate Change, Environmental Stability, ‘Slow’ Catastrophes, Genetics and Epigenetics, Threat (All-Hazards) Awareness, Human Learning and Financial Policy. All of these reflect concerns or problems that bedevil the modern world and deserve new or better approaches, including technologies and techniques. In our estimation, only visual analytics can provide the set of capabilities that will enable scientists as well as policymakers to address such concerns in ways that take advantage of all the data necessary to understand them.

Likewise, this broader perspective requires that more and diverse disciplines be considered as contributing to our development and use of visual analytics. These include, but are not limited to, the following: Cognitive and Perceptual Science, Data Transformation and Representation, Dimensionality Reduction, Smart Query,
Our vision is that future visual analytics technologies will affect every interface between the human and the digital universe. From the situation room, the office, the home, the desktop, the collaborative teams, to individual mobile devices – all of these will be directly or indirectly influenced and enabled by visual analytics science and technology. There will be typeless data interfaces seamlessly supporting text, imagery, and video. There will be walkup-usable, mixed-initiative interfaces allowing immediate use of technology with successively more complex tools as the situation demands.

The issue of scale will change from a limitation to an advantage: the opportunities of scale will be an enabling foundation. Our digital universe will exhibit scale tolerance and provide synthesized nuggets of knowledge, with known provenance, toward problems and situation analytics, or even personalized knowledge packages.

The interaction and analytic methods will match and complement cognitive reasoning: for the field of visual analytics, this means that the tools and approaches to be adopted in the future need to incorporate elements from cognitive or perceptual science, taking advantage of millions of years of human evolutionary history or modern neuroscience and psychology principles. The critical challenges of uncertainty will be reduced by effective analytic interaction from source data to multiple scales, abstractions, views and levels of knowledge.

Our visual analytical sciences will enable prospective and proactive thinking well beyond today’s confirmatory and exploratory methods. As a foundational statement, we want our tools to help us detect the expected and discover the unexpected™.

We expect in the future that an international set of partnerships among academia, industry and applied research laboratories will contribute a diversity in cultures, thought and applications. This diversity will be considered during the investigation, development and implementation of visual analytics techniques.

We expect a broadening of applications and domains interested in deploying visual analytic solutions. And finally, we expect a sustained educational workforce relying on a common, comprehensive curriculum to provide the talent needed for research, engineering, training and support.

Is this all possible? 5 years ago when visual analytics was born, the founders had an aggressive vision that is playing out faster than anyone expected. Yes, it is possible, the opportunity is before us, and the following papers provide increased detail about what is needed to achieve this vision.

Introduction to Papers

The first paper, by George Robertson’s team, presents the many issues and challenges offered by scale and complexity. The second paper, from Bill Ribarsky’s team, presents new thinking on the science of analytical reasoning. The third, by Bill Pike’s team, covers the science of interaction. Without appropriate information representations and integration, the science of interaction has limited capability; so the fourth paper, from Dave Kasik’s team, discusses data transformation and representation for information generation. The fifth paper, by Nancy Chinchor’s team, covers the science of analytic reporting. The sixth paper, from Jean Scholtz’s team, presents technology transition progress. The seventh paper, by Pak Chung Wong’s team, looks at the growth of the science of visual analytics compared to other related fields. The final paper discusses several aspects of challenges, observations from deployed systems, and future needs for the emerging field of visual analytics.

Acknowledgements

We want to thank the leaders of Information Visualization and Palgrave, specifically Chaomei Chen and Ben Shneiderman, for this opportunity to present a suite of papers defining a new direction within the science of visualization. We especially want to thank Carrie Almquist, Sharon Eaton, Fran Stanley and Kelly Sandretto for their organizational and editing support in the preparation of the special issue. The following individuals participated in the DHS Future of Visual Analytics Workshop held in Arlington, Virginia on 4 March 2009: Carrie Almquist, Neville Clarke, Kris Cook, John Dill, David Ebert, Brian Fisher, Howard Greenblatt, Peter Highnam, David Kasik, Daniel Keim, Joseph Kielman, Doug Lemon, Alan MacEachren, Emile Morse, Terry Nipp, Chris North, Theresa O’Connell, Haesen Park, Bill Pike, Bill Pottinger, Larry Rosenblum, Randy Rohrer, George Robertson, Bill Ribarsky, Emily Saulsgiver, Jean Scholtz, Yukiko Sekine, John Stasko, Anita Street, Jim Thomas, Alan Turner, Mark Whiting, Kirsten Whitley, Pak Wong. This work has been supported by the NVAC™ located at the Pacific Northwest National Laboratory (PNNL) in Richland, WA. NVAC is sponsored by the US Department of Homeland Security Science and Technology Directorate. The PNNL is managed for the US Department of Energy by Battelle Memorial Institute under Contract DE-AC05-76RL01830.

References


