"Discovery consists of seeing what everybody has seen and thinking what nobody has thought."

—Albert von Szent-Gyorgyi (1893–1986)

3

Visual Representations and Interaction Technologies

The use of visual representations and interactions to accelerate rapid insight into complex data is what distinguishes visual analytics software from other types of analytical tools. Visual representations translate data into a visible form that highlights important features, including commonalities and anomalies. These visual representations make it easy for users to perceive salient aspects of their data quickly. Augmenting the cognitive reasoning process with perceptual reasoning through visual representations permits the analytical reasoning process to become faster and more focused.

It is a challenge to create well-constructed visual representations. In the field of scientific visualization, data often correspond to real-world objects and phenomena, meaning that there are natural visual representations. In scientific visualization, the goal is to mimic these real-world representations as faithfully as computationally feasible. However, most visual analytics problems deal with abstract information so the researcher is left to select the best representation for the information.

Visual representations invite the user to explore his or her data. This exploration requires that the user be able to interact with the data to understand trends and anomalies, isolate and reorganize information as appropriate, and engage in the analytical reasoning process described in Chapter 2. It is through these interactions that the analyst achieves insight.

This chapter discusses important aspects of visual representations and interaction techniques necessary to support visual analytics. It covers five primary topics. First, it addresses the need for scientific principles for depicting information. Next, it focuses on methods for interacting with visualizations and considers the opportunities available given recent developments in input and display technologies. Third, it addresses the research and technology needed to develop new visual paradigms that support analytical reasoning. Then, it discusses the impact of scale issues on the creation of effective visual representations and interactions. Finally, it considers alternative ways to construct visualization systems more efficiently.

Visual analytics tools can support people working under great time pressure, whether they are analysts, emergency management and response staff, or border personnel. Well-crafted visual representations can play a critical role in making information clear. The visual representations and interactions we develop must readily support users of varying backgrounds and expertise. In an emergency situation, for example, personnel may need to use unfamiliar systems to gain the insight that they need to respond. Visual representations and interactions must be developed with the full range of users in mind, from the experienced user to the novice working under intense time pressure, so that visual analytics tools can achieve their promise.

Developing Principles for Depicting Information

The design of visual representations of information has been ongoing for centuries. Over the past 20 years, driven by the ever-increasing speed and availability of computers, information visualization researchers have invented dynamic and interactive computer-mediated visual metaphors for representing abstract information. A new discipline is rapidly emerging around the creation of computer-mediated visual representations to support display and analysis of information.

Some of these new techniques work well and have generated great excitement. However, the number of successful new computer-mediated visual representations today is small compared to the number of highly evolved and widely used metaphors created by human information designers. Human-designed visualizations are still much better than those created by our information visualization systems.

State of the Art

The creation of computer-mediated visual representations has much in common with other emerging disciplines. An emerging discipline progresses through four stages. It starts as a craft and is practiced by skilled artisans using heuristic methods. Later, researchers formulate scientific principles and theories to gain insights about the processes. Eventually, engineers refine these principles and insights to determine production rules. Finally, the technology becomes widely available. The challenge is to move from craft to science to engineering to systems that can be widely deployed.

Today, we are still in the early stages of development of the discipline. We lack fundamental understanding of the basic principles for effectively conveying information using graphical techniques. Without fundamental knowledge of what makes certain representations effective, it is not possible to efficiently construct new representations for new classes of information or to know that the new representations will work as designed. Poorly designed visualizations may lead to an incorrect decision and great harm. (A famous example is the poor visualizations of the O-ring data produced before the disastrous launch of the Challenger space shuttle, as discussed more fully in Tufte [1997] and Chapter 5.) Thus, we need to develop scientific principles for effectively conveying information.

Cognitive scientists have studied visual representations and the larger class of *external aids to cognition*. An external aid to cognition is an artifact that helps us reason about the world. In the physical world, we build and use power tools to extend our physical abilities. In the same way, in the world of information, we build cognitive tools to extend our reasoning abilities. Visual representations are the equivalent of power tools for analytical reasoning.

A first step in developing principles for visual representations is to understand how they enable cognition [Card, 1999; Norman, 1993]. Some basic principles for developing effective depictions include the following (adapted from [Norman, 1993]):

- **Appropriateness Principle** – The visual representation should provide neither more nor less information than that needed for the task at hand. Additional information may be distracting and makes the task more difficult.
- **Naturalness Principle** – Experiential cognition is most effective when the properties of the visual representation most closely match the information being represented. This principle supports the idea that new visual metaphors are only useful for representing information when they match the user's cognitive model of the information. Purely artificial visual metaphors can actually hinder understanding.
- **Matching Principle** – Representations of information are most effective when they match the task to be performed by the user. Effective visual representations should present affordances suggestive of the appropriate action.

Another prominent cognitive scientist has suggested the following two basic principles [Tversky et al., 2002]:

- **Principle of Congruence** – The structure and content of a visualization should correspond to the structure and content of the desired mental representation. In other words, the visual representation should represent the important concepts in the domain of interest.
- **Principle of Apprehension** - The structure and content of a visualization should be readily and accurately perceived and comprehended.

The subjects of mental representations and reasoning are the main focus of cognitive science, so the principles for depicting information must be based on research in cognitive science. The apprehension principle underlies the importance of research in perception. These meta-principles underscore that the biggest challenge in choosing a visual representation is to find the *right* one (not just any one) for the reasoning task at hand.

The next step to take in developing a set of design principles is to formally define the different types of visualizations. The French cartographer Bertin has developed a system for characterizing representations of charts, maps, and networks [Bertin, 1981]. Bertin considered the space of possible visual representations as a visual language. The spatial and visual attributes of the image encode the information using the rules of the language. Bertin's system has since been used to define a design space of information visualizations. Examples of extensions are Mackinlay [1986], Roth et al. [1991], MacEachren [1995], Card et al. [1999], and Stolte [2002]. Wilkinson [1999] has developed an extensive grammar for graphics. Others have tried to develop taxonomies of visual techniques [Shneiderman, 1996; Spence, 2000; Ware, 2000]. The most notable is Shneiderman's taxonomy, which breaks down visualization by the characteristics of the data (1D, 2D, nD, network, etc.). Although these design spaces and taxonomies are very promising, we are far from having a complete, formally developed theory of visual representations.

After defining the space of visualizations, Bertin developed design principles for choosing among the possibilities. He argued that the properties of the visual representation should match the properties of the data representation. For example, color represents nominal data well because hue is not naturally ordered. Although Bertin provides a set of principles loosely based on perception, it is important to realize that his system is not based on rigorous experiments involving human subjects. Bertin also did not emphasize the importance of the task when choosing a visual representation.

Another notable attempt to provide design principles for statistical graphics is the work of Cleveland [1985]. Some of Cleveland's recommendations are based on experiments in graphical perception [Cleveland & McGill, 1984]. Perceptual design principles have also been developed for color [Rogowitz & Treinish, 1993] and motion [Bartram & Ware, 2002]. However, scientific principles are rare, and most recommendations are based on general principles of graphic design. Tufte's three outstanding books on information presentation [1983, 1990, 1997] also stress the importance of using principles regularly practiced by graphic designers. Best practices have also been developed for different domains such as statistical graphics [Cleveland, 1985] and cartography [MacEachren, 1995]. Tukey [1977] was also a strong advocate of using graphics in data analysis and developed many visual representations that are now common in statistics. To move the field of visual analytics forward, we need to perform more research in developing scientifically tested design principles.

Recent work in the information visualization community has attempted to systematically apply design principles to the automatic generation of visualizations. Mackinlay [1986] developed Automated Presentation Tool (APT) that automatically designed charts based on Bertin's and Cleveland's ideas. APT searches over a space of possible visual representations, evaluates them based on expressiveness and effectiveness criteria, and chooses the best one. This work has been extended by Roth and colleagues [Roth, 1991; Zhou, 1999]. There has also been recent work on using cognitive design principles for automatically producing route maps [Agrawala et al., 2003] and assembly instructions [Heiser et al., 2004]. In the future, most visualizations will be generated by machines as users interact with the information, so automating the presentation of information will become increasingly important.

Technology Needs

The systems described here are only initial steps toward solving the major problems in creating a complete set of cognitive, perceptual, and graphic design principles. The creation of analysis systems that are based on cognitive, perceptual, and graphic design principles will dramatically improve the efficiency, effectiveness, and capabilities of analysts, decision makers, scientists, and engineers.

Recommendation 3.1

Conduct research to formally define the design spaces that capture different classes of visualizations.

We must both characterize the taxonomy of visual representations that must be considered and describe the range of design parameters associated with these representations. While there is a body of work available to draw upon, as described above, this work has not been targeted specifically at visual representations in support of analysis. We must build upon and extend this work to create a formal definition of the range of available visual representations.

Recommendation 3.2

Develop a set of scientifically based cognitive, perceptual, and graphic design principles for mapping information to visual representations.

Using the taxonomy created above, we must define the set of principles for selecting the most promising visual representations to support a specific combination of analytic task and data characteristics. These principles must be verified through user testing. In addition, we must develop and test principles for selection of the specific visualization properties to support specific tasks and data characteristics.

The use of design patterns has become an accepted and useful technique in areas such as object-oriented design and software engineering [Gamma, 1994]. We must investigate whether we can develop a set of visual design patterns for both designing new visualizations and determining the types of visualizations most useful for particular analytic tasks. One potential approach is to develop a library of common visualization design patterns from which developers could draw to build new visualizations.

A Science of Interaction

Visual analytics is not simply about presenting information. Rather, an analysis session is more of a dialogue between the analyst and the data, where the visual representation is simply the interface or view into the data. In an analysis dialogue, the analyst observes the current data representation, interprets and makes sense of what he or she sees, and then thinks of the next question to ask, essentially formulating a strategy for how to proceed [Card et al., 1999; Spence, 2000]. Undoubtedly, new questions occur to the analyst and new factors must be considered. Thus, a different perspective on the data will be needed and new variables will need to be considered. The manifestation of this dialogue is the analyst's interactions with the data representation. How does the analyst request other perspectives on the data? How does the analyst filter out unwanted details? How does the analyst request new visual representations of the data?

State of the Art

Too often in the visual analytic process, researchers tend to focus on visual representations of the data but interaction design is not given equal priority. We need to develop a "science of interaction" rooted in a deep understanding of the different forms of interaction and their respective benefits. The mantra by Shneiderman

[1996] of "*Overview first, zoom and filter, details on demand*" is well-accepted, but what are the next steps, or additional different steps?

There are at least three ways to look at the science of interaction. First, we can look at interaction from the point of view of human time constants. This is an important viewpoint because all interaction is constrained and driven by what the user is cognitively and perceptually capable of doing. Second, we can look at how interaction is used to accomplish tasks such as data manipulation, manipulation of visual mappings, navigation, and dialogue. Third, we can look at the nature of the interaction itself, including the differences between interactions in 2-dimensional (2D) and 3-dimensional (3D) environments and the effects of the devices used for interaction. Each of these viewpoints yields different insights into the current state of the art, as described below.

Levels of interaction: human time constants

Analysis of human time constants for human-computer interaction was initially discussed by Card et al. [1983], considered from a cognitive science point of view by Newell [1990], and discussed from an information visualization point of view by Card et al. [1999]. Newell describes four bands of time scales for human action (biological, cognitive, rational, and social) ranging from 100 microseconds to months. For purposes of a science of interaction for analytical reasoning, the two bands of greatest focus are Newell's cognitive (100 milliseconds to 10 seconds) and rational (minutes to hours) bands. Card describes three distinct bands within Newell's cognitive band. Note that these time constants represent approximate time ranges. That is, when we say ~100 milliseconds, we mean somewhere in the range of 50 to 300 milliseconds.

~100 milliseconds. Card refers to this as the *perceptual fusion time constant*, while Newell refers to it as the *deliberate act time constant*. This time constant is the rate necessary to produce the perception of a smooth animation. In animation, 10 frames per second equates to 100 milliseconds per frame. In interaction design, this time constant is the rate necessary to create the perception of an immediate response. Users expect to see an immediate response when they move a dynamic query slider [Ahlberg, 1994]. Likewise, as users brush over items of interest, they expect to see immediate corresponding highlighting of the linked items [Cleveland, 1999]. This time constant is also important because minimum human motor response time is around 250 milliseconds.

~1 second. Card refers to this as the *unprepared response time*, while Newell refers to it as the *operation time*. For our purposes, this constant represents the necessary rate of response to simple user actions. For example, clicking a web link should produce the display of the next web page within 1 second to be effective. If the response might take more time, it is important to provide some kind of feedback in the 1-second timeframe to reassure the user that something is happening. This time constant is also important for interactive animation, like user-initiated transition animations (transitions from one complex structure to another or one viewpoint to another). It has been demonstrated that providing a 1-second transition animation can reduce user task performance time compared to providing no transition animation [Robertson et al., 2002].

~10 seconds. Both Card and Newell refer to this as the *unit task time*. This is the time within which users expect more complex user-initiated activities to complete (e.g., a complex search). Again, if an activity of this kind will take more than 10 seconds to complete, it is important to provide the user with feedback within this 10-second timeframe.

~100 seconds (minutes to hours). This is referred to as the *rational band*. Higher-level reasoning processes, including the analytic reasoning and human-information discourse processes described in Chapter 2, take place in this band. Interaction techniques in this timeframe rely heavily on complex cognitive processing and are greatly affected by attentional resource demands such as interruptions and shifts in focus. These techniques are the least well understood and developed.

Uses of interaction

Card et al. [1999] identify three primary uses of interaction for information visualization: to modify data transformation (filtering), to modify visual mappings, and to modify view transformation (i.e., navigation). For visual analytics, we add a fourth use, which is for human-information discourse, a higher-level user dialogue with the information.

Interactions for modifying data transformation (filtering). Several common techniques are in use today, including direct manipulation, dynamic queries [Ahlberg, 1994], brushing [Cleveland & McGill, 1984], and details-on-demand.

Interactions for modifying visual mappings. Dataflow systems [Haeberli, 1988] and Pivot Tables are two examples of techniques that allow the user to interactively change mappings between the data and their visual representations.

Interactions for modifying view transformation (navigation). Interaction techniques range from simple approaches like direct selection for selecting and highlighting objects of interest, to more complex camera control techniques in 3D environments. They also include techniques for panning and zooming [Bederson et al., 1996] as well as for achieving a balance between overview and detail [Plaisant et al., 1995].

Interaction for human-information discourse. The least well understood use of interaction is to support a true human-information discourse in which the mechanics of interaction vanish into a seamless flow of problem solving. Interactions are needed to support processes such as comparing and categorizing data, extracting and recombining data, creating and testing hypotheses, and annotating data.

To date, there has been no foundational work to characterize the design space of these interaction techniques. We really do not know if the techniques that have been created thus far are the best or most appropriate techniques.

Nature of interactions

The nature of an interaction is affected by whether it takes place in a 2D or 3D environment. The best developed interaction techniques have been for 2D visualizations and 2D graphical user interfaces. While a lot of work has been devoted to interaction techniques for 3D virtual environments, they are not nearly as well developed as the 2D techniques. 3D manipulation and navigation techniques tend to be harder to use and harder to learn. One promising approach for simplifying interaction in 3D environments has been to identify cases where 3D visual representations are used, but the interactions are constrained so that 2D interaction techniques can be used. An example of this is the Data Mountain [Robertson et al., 1998], where a 3D visual representation is used but object manipulation takes place on a tilted plane.

Interaction can also be greatly affected by the display and interaction devices used for visual analytics tasks. A wide range of display configurations will be used to support visual analytics; hence, interaction techniques should be designed so that they are similar across different devices ranging from large shared displays to desktop environments to field-portable devices. While this is technically challenging to do, it has been done in at least one case. DateLens [Bederson et al., 2004] is a scalable calendar system that works on everything from a personal digital assistant (PDA) display to a wall-sized display, scaling the visual representation to the appropriate size for the device and using the same interaction technique at all scales.

While most interaction techniques use a single modality (or human sense), there is work that suggests that multimodal interfaces can overcome problems that any one modality may have. For example, voice and deictic (e.g., pointing) gestures can complement each other and make it easier for the user to accomplish certain tasks [Oviatt, 1999]. Sonification can be used to enhance visualization either by redundant encoding of visual information in the auditory channel [Robertson, 1998] or by use of sound to represent data values [Smith et al., 1990].

Technology Needs

Although a lot of isolated design work has been done in specific aspects of interaction science, little systematic examination of the design space has been done. As a field, we are in a transition phase in which researchers are beginning the foundational work to understand that design space. Creating a science of interaction is critical because the large-scale nature of the analytic problem and the compressed timeframe for analysis require that we identify and develop the correct interaction techniques for any given human timeframe, interaction use, or interaction environment.

Basic interaction techniques

To achieve successful adoption, visual analytics software must support both basic interactions and highly sophisticated interactions that support the analytic reasoning process. Before these more sophisticated interactions can be addressed systematically, work is needed to create a scientific understanding about the basic interactions that are used to support simpler operations. This understanding will form the foundation for research into more sophisticated interactions.

Recommendation 3.3

Create a new science of interaction to support visual analytics.

The grand challenge of interaction is to develop a taxonomy to describe the design space of interaction techniques that supports the science of analytical reasoning. We must characterize this design space and identify under-explored areas that are relevant to visual analytics. Then, R&D should be focused on expanding the repertoire of interaction techniques that can fill those gaps in the design space.

Interaction techniques for human-information discourse

Existing work on interaction techniques for human-computer interaction and information visualization has focused on cognitive time bands, interaction for data manipulation, visual mapping manipulation, and navigation. The discussion on analytic discourse and sense-making in Chapter 2 makes it clear the higher-level dialogue between analyst and information, or human-information discourse, is of vital importance. This discourse involves the rational time band and higher-level uses of interaction, but neither has been sufficiently explored.

Recommendation 3.4

Expand the science of interaction to support the human-information discourse needed for analytical reasoning. In particular, identify and develop interaction techniques that support higher-level reasoning and that address the rational human timeframe.

Human beings are very skilled at analyzing complex situations using a combination of their available information and their combined knowledge and experience. However, there are inherent human tendencies that analysts must recognize and overcome. Interaction techniques must be developed that support an analytic discourse and help compensate for human limitations, including:

- **Information overload in complex situations**. Techniques are needed to help analysts simplify their cognitive load without compromising the analyst's effectiveness and to help compensate for faulty memory.
- **Overcoming biases**. Biases affect the way data are interpreted. Biases about the reliability of different sources may lead people to discount information from sources that aren't considered reliable. People often see what they expect to see and tend to ignore evidence that is contradictory to a preferred theory. If they form a preliminary judgment too early in the analytical process, they may hold firm to it long after the evidence invalidates it [Heuer, 1999].
- **Satisficing**. People settle for a "good enough" answer, sometimes stopping their analytical process before they identify critical information that would lead them to a different conclusion [Heuer, 1999].

New interaction techniques are needed to support the user in evaluating evidence, challenging assumptions, and finding alternatives. Analytical environments should support the user in identifying and understanding all relevant information to reach a solid conclusion rapidly. The tools we create need to establish a correct balance between structure and intuition.

Leveraging New Media to Support Interaction

While it is not expected that the visual analytics research community will need to focus on inventing new display technologies, the ability to harness the power of new display and interaction technologies invented by others will be one key to success. Advances in the past decade have led to new discoveries in terms of the capacity and manner in which visualizations can be displayed and interacted with. Traditional desktop displays, which appeared stuck for several years with Cathode Ray Tube (CRT) technology at resolutions below a mega-pixel, are now rapidly changing in terms of resolution and form factor. As these and even more dramatic changes occur, it will be important for visual analytic researchers to remain abreast of these changes and to use new techniques for expressing data with the new media.

Harnessing new display technologies

On the desktop, a variety of technologies exist to enable display and interaction with visualizations. These range from the traditional single-mega-pixel CRT computer monitors to 3D Liquid Crystal Displays (LCDs) to multi-mega-pixel LCDs.

While stereovision has been available on the desktop for many years, displays that free the user from goggles are recent introductions to the marketplace and potentially will have a greater penetration and therefore more availability to users of visual analytics. As these displays merge with traditional 2D displays and the prices normalize, an opportunity exists for researchers to use this technology in their applications.

While stereo on the desktop has been available for years, this technology is still relatively unexplored. However, users quickly adopt improvements in screen resolution and size. Multi-mega-pixel displays are offered by most monitor vendors and are being rapidly adopted at both homes and offices. Improvements in LCD technologies are providing displays of up to 60 inches while resolutions have approached nearly 10 mega-pixels (although not at such extreme display sizes). Other technologies like Liquid Crystal on Silicon (LCOS) are expected to take such improvements to LCDs to the next level. These improvements are only expected to increase and should be considered by those developing visual analytic techniques.

Form factor and resolution are only a few of the expected improvements in the coming years to display technologies. We believe that new technologies like the Organic Light-Emitting Diode (OLED) displays will improve viewing angle, weigh less, and be more cost effective, brighter, and power efficient [Tang & Van Slyke, 1987].

Recommendation 3.5

Develop visual representations and interaction techniques that exploit new display devices for visual analytics.

Mobile technologies will play a role in visual analytics, especially to users that are on the front line of homeland security. First responders now use technologies like cell phones and PDAs; in the future, they will use new technologies like foldable displays, electronic inks, or virtual retinal displays [Wang et al., 1999; Kollin, 1993]. These technologies, which allow flexible, lightweight, and wearable options for users, will allow information to rapidly be disseminated to users in field. Researchers must devise new methods to best employ these technologies and provide a means to allow data to scale between high-resolution displays in command and control centers to field-deployable displays.

Scaling to multiple devices and device configurations

To support homeland security missions, visual analytics applications must support applications ranging from an operations center using shared large-screen displays or potentially augmented reality displays, to the individual analyst working at a desktop computer to first responders and border personnel using handheld, field-portable devices. Users need to have a consistent set of interactions that they can count on regardless of the device they are using. This is especially true in emergency situations, when users' attention is directed toward the immediate situation and mechanics of the computer system must be second nature to the user.

Recommendation 3.6

Develop interaction techniques that scale across different devices and are used in the same way on platforms ranging from handheld, field-portable devices to wall-sized displays.

Many common desktop computing interaction methods are not currently portable to other devices. Research is needed to develop interaction techniques that both optimize the opportunities offered by new devices and provide consistency of operation across devices.

For large and multiple displays, there is a benefit to seeing more information for more people and to enabling interactive group collaboration. Furthermore, increased screen real estate enables new research efforts into peripheral awareness of information [Greenberg & Rounding, 2001; Cadiz et al., 2002; Stasko, 2004]. Extra displays could be used to help analysts stay aware of information and might facilitate "noticing" important facts. The availability of added display space may foster the development of new information representations, ones that simply were not practical on traditional, single-monitor systems. Multiple or distributed display environments, however, present a whole new set of challenges for interaction and navigation. Navigating with a mouse over a large display area can become slow and tiresome, for example. Thus, new interaction techniques for these environments are needed [Baudisch et al., 2003; Hutchings, 2004; Robertson et al., 2004].

Multimodal interaction

Voice and gesture complement each other and, when used together, can create an interface more powerful than either modality alone. Oviatt [1999] shows how natural language interaction is suited for descriptive techniques, while gestural interaction is ideal for direct manipulation of objects. Unlike gestural or mouse input, voice is not tied to a spatial metaphor. Voice can interact with objects regardless of degree of visual exposure, particularly valuable in a graphical environment where objects may be hidden inside each other or occluded by other objects.

Users prefer using combined voice and gestural communication over either modality alone when attempting graphics manipulation. Hauptman and MacAvinney [1993] used a simulated speech and gesture recognizer in an experiment to judge the range of vocabulary and gestures used in a typical graphics task. Three different modes were tested: gesture only, voice only, and gesture and voice recognition. Users overwhelmingly preferred combined voice and gestural recognition because of the greater expressiveness possible. Users were also able to express commands with greatest sufficiency using combined input.

Some tasks are inherently graphical; others are verbal; and yet others require both vocal and gestural input to be completed. Allowing both types of input maximizes the usefulness of the environment by broadening the range of tasks that can be done intuitively. Also, allowing both types of input would enable analysts to vocally annotate how discoveries were made and replay sequences to others.

There are also psychological reasons for integrating speech and gesture recognition into a virtual environment. Experiments in cognitive psychology have shown that a person's ability to perform multiple tasks is affected by whether these tasks use the same or different sensory modes, for example visuo-spatial or verbal modes. According to the multiple resource theory of attention [Kinsbourne & Hicks, 1978; Wickens, 1980], the brain modularizes the processing of different types of information—when different tasks tap different resources, much of the processing can go on in parallel. Such is the case with speech and visuo-spatial modalities. Thus, by adding speech input to the visual environment, users should be able to perform visuo-spatial tasks at the same time as giving verbal commands with less cognitive interference.

Experimental evidence supports this theory. Investigating attentional capacity, Treisman and Davis [1973] found that the ability to concentrate on more than one task at a time was expanded when the tasks were presented in separate perceptual channels and people responded to them across different response channels. This is in part due to the spatial and verbal information being stored separately in human memory [Baddeley & Hitch, 1974]. Martin [1989] finds that the benefits of multiple modalities previously demonstrated with separate, multiple tasks also extend to single-task, multiple-component situations more typical of human-computer interactions.

The level to which multimodal interfaces can bring benefit to the analytical reasoning process has yet to be fully explored, however.

Recommendation 3.7

Investigate the applicability of multimodal interaction techniques for visual analytics.

The combination of gestural and voice interaction has produced some promising research results, but it has not been widely adopted in application. Visual analytics supports much more complex reasoning tasks than have been the subject of previous research on multimodal interfaces. Additional study is needed to see how multimodal interfaces affect the effectiveness and efficiency of the analytical process. We should investigate the use of multimodal interfaces for both individual analysts and collaborative teams of analysts. In addition, research is needed to determine the value of multimodal interfaces in field conditions faced by border personnel and by distributed teams working in noisy and time-sensitive emergency situations.

New Visual Paradigms to Support Analytic Reasoning

Over the last 20 years, numerous new visual representations, interaction methods, software tools, and systems have been developed. Often these representations and tools have been developed without considering the analytical reasoning tasks that they support. Meaningful visualization techniques for complex information must be task-driven.

In this section we consider various reasoning tasks of critical importance and outline the types of visual representations that need to be developed.

Organizing Large Collections of Information

Today, two primary ways to organize information spaces exist: the graphical desktop user interface and information displayed by search engines. The desktop interface consists of hierarchically organized folders containing documents. It was designed in an era when floppy disks were common and file systems contained thousands of documents. A desktop interface to handle larger collections of information is badly needed. Search engines developed by companies such as Google™, Inktomi®, Yahoo!®, and Microsoft® Corporation organize large interconnected document collections. The interfaces are query-driven and show only small result sets. No information is given about the space of all documents. Search technology is now being deployed to help organize the desktop, e-mail, and other information spaces.

Information visualization systems, such as IN-SPIRE™, allow thousands of documents to be visually, succinctly described, navigated, and accessed. The ThemeView™ 3D visual landscape shown in Figure 3.1 reflects the high-dimensional properties and relationships of sets of documents by showing clusters of themes and their strengths. The complex content is visible. Exploration, summarization, comparison, trends over time, tracking, and many other operations are more efficient. The strengths of visualization and document vector mathematics are combined to achieve a new analytical capability for massive data [Hetzler & Turner, 2004].

SeeSoft [Eick et al., 1992] provides a visual representation of all source code in a large software system, as shown in Figure 3.2. The system represents lines of code from software files in successively smaller fonts, ultimately representing the most deeply indented lines as individual rows of pixels whose indentation and line length tracks the original text. The color of each row encodes an attribute such as the age, author, or function of the corresponding line of code. This representation captures the sequential nature of source code, shows critical loops in the code, and allows overlays of additional information. SeeSoft shows the information inside each source code file, as well as relationships across the files. For example, the lines of code written by a single developer can be shown both within a file and across multiple files.

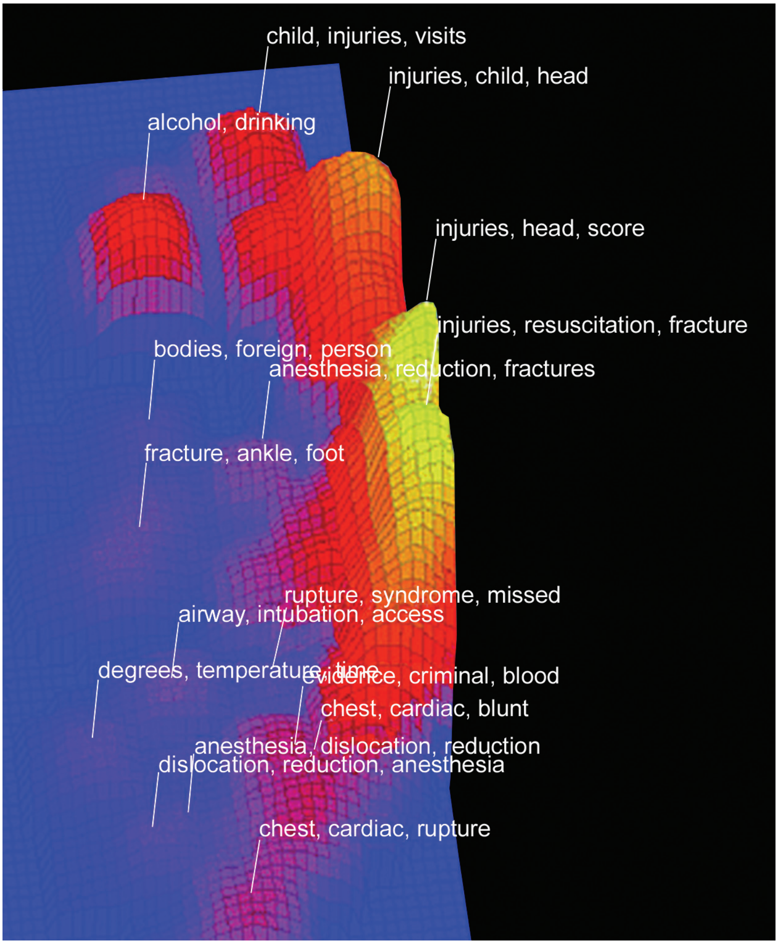


***Figure 3.1**. The IN-SPIRE software's ThemeView landscape shows relationships among documents. High peaks represent prominent themes. Peaks close together represent clusters of similar documents.*

Figure 3.2 shows an entire module from this software system with color showing age of each line. The reduced representation provides an overview, shows the new code, old code, and frequently changed code. This example highlights how an effective representation of a domain, such as software in this example, becomes a critical visual tool to enable understanding of the structure of the code and patterns of maintenance in the code that were not possible before it was created.

Another example of a set of integrated views of an information space is the Command Post of the Future system described in Chapter 5. Used in military applications, it presents a comprehensive map-based view of the operational area. The map provides a unified information space that is populated with intelligence assessments and planned activities.

***Figure 3.2**. Left: SeeSoft representation for software source code where successively smaller fonts lead to a "reduced" single row of pixels per line of source code. Right: Module in a software system with color encoding the age of lines of code.*

These examples illustrate the value of showing large integrated views of dynamic information spaces. However, much more work is needed to support the full range of analytic tasks faced by the analyst and accommodate the demands of analyzing massive and complex data collections.

Recommendation 3.8

Create visual analytic tools that provide integrated views of large-scale information spaces, support coordinated viewing of information in context, and provide overview and detail.

Integrated views of data can support and improve perception and evaluation of complex situations by not forcing the analyst to perceptually and cognitively integrate multiple separate elements. Visualization systems should present all the relevant information required for a decision maker or operator to efficiently and correctly comprehend and act in a complex situation. Systems that force a user to view sequence after sequence of information are time-consuming and error-prone [Kapler & Wright, 2004; Mooshage et al., 2002].

Similarly, combining or merging interactions and controls with visible representations can also speed access, control, and manipulation operations. Often, users experience a cognitive separation between the task they want to accomplish and the mechanics for accomplishing the task. New techniques can be invented that do away with the separation of "what I want and the act of doing it" [Van Dam, 2001]. Integrating views, interactions, and analytics provides significant productivity improvement for the combined human, analytical, and data system.

To provide the maximum information for the user, visual representations are routinely combined with textual labels. The generation of meaningful labels and their placement on the display represents an often-overlooked challenge. Labels should be visible without overwhelming the display or confusing users. The number of labels, their placement, and their content are all areas for further investigation.

Cartographers have long wrestled with this problem and their work offers valuable lessons, but interactive systems and dynamic data bring new challenges and opportunities to labeling.

Analysts working with large data sets want to gain a global understanding of the data and then focus on particular data items and their attributes. Alternating among levels of detail generates moment-to-moment insights and spurs new questions. A number of visualization and user interface techniques have been developed to support coordinated views of both overview and detail. Greene et al. [2000] found that previews and overviews help users quickly understand the scope of their data set and discriminate between interesting and uninteresting content. Existing techniques include the use of multiple coordinated windows at different focus levels, panning and zooming operations, and so-called "fisheye" views [Plaisant et al., 1995; Furnas, 1986]. Fisheye techniques are "focus+context" views, i.e., one in which both the overview and detail are presented side by side in the same view. The power of "focus+context" comes from the ability to subjugate unimportant information for contextual navigation while using attentive focus for effectively communicating information. While existing techniques in this area are very useful, new methods must be developed that take advantage of multiple and large-area computer displays to assist analysts with inquiries on the massive data sets evident in visual analytics.

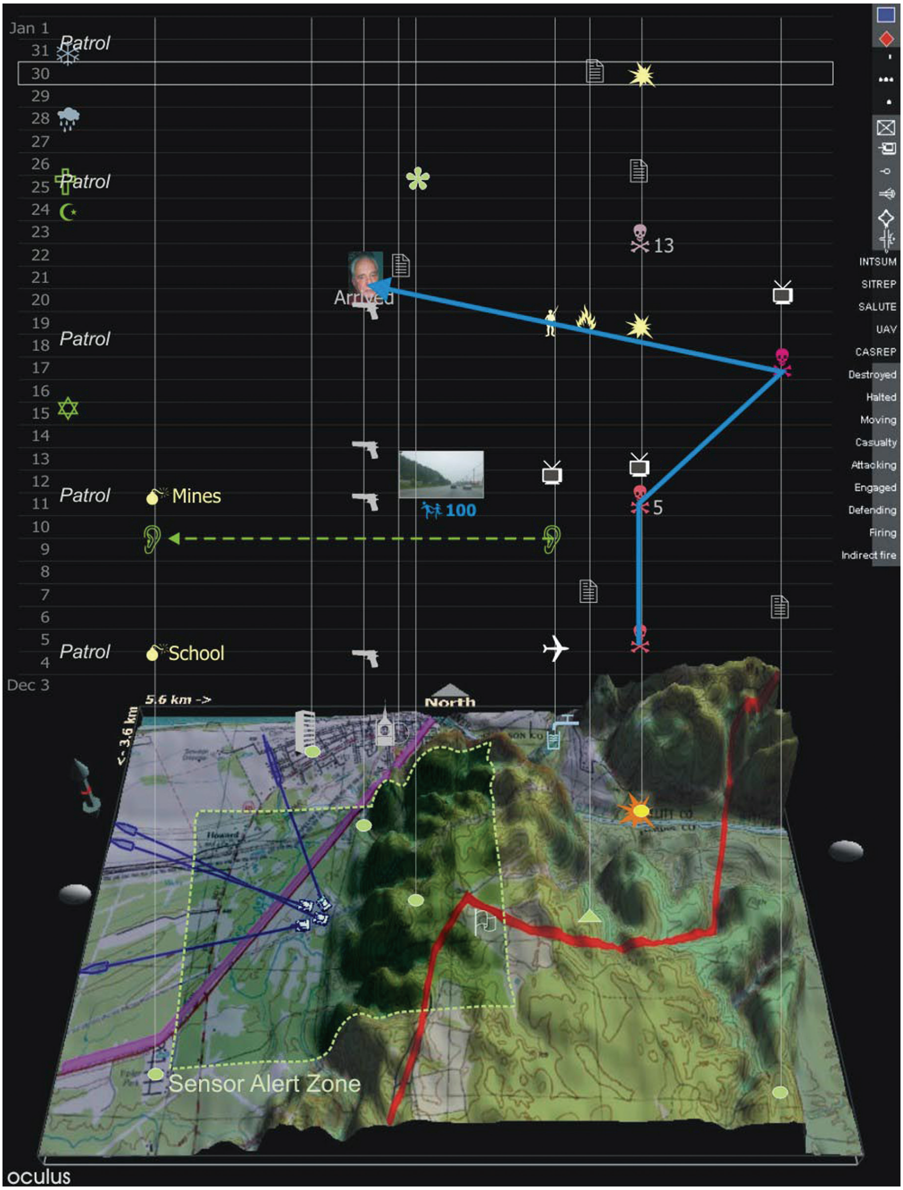
Reasoning about Space and Time

Because of our daily need to navigate and reason about the world around us, people are particularly good at reasoning about space and time. Maps, which are one of the earliest visual inventions of the human race, abstract the space around us in ways that support various forms of reasoning. Modern geographic information systems provide access to large amounts of geospatial information, including satellite imagery, digital terrain models, and detailed maps of roads and cities. To support analysis, this information must be studied in a temporal context.

Cartographers have developed representations such as flow maps [Dent, 1999] that show migration patterns on top of maps. The common weather map is an example of a flow map. Understanding how best to combine time and space in visual representations needs further study. For example, in the flow map, spatial information is primary (i.e., it defines the coordinate system of the visualization). Why is this the case, and are there visual representations where time is foregrounded that could also be used to support analytical tasks?

An example of an innovative system is GeoTime™, shown in Figure 3.3. GeoTime works with the spatial inter-connectedness of information over time and geography within a single, highly interactive 3D view. Events are represented within an X,Y,T coordinate space. Patterns of activity among people, places, and activities can be analyzed. Connectivity analytical functions help find groups of related objects [Kapler & Wright, 2004].

While systems such as GeoTime show great promise, the challenge of integrated spatial and temporal reasoning is still a substantial one.



©2004 Oculus Info Inc.

***Figure 3.3**. GeoTime provides an integrated view for analysis of a combination of temporal and geospatial data.*

Recommendation 3.9

Develop tools that leverage humans' innate abilities to reason about space and time.

Analyzing observations over time and geography typically requires multiple, separate tools, e.g., a map for geospatial information and a timeline for the temporal information. Representations of time that support temporal reasoning are less studied and less developed than representations of geospatial data. Navigation and other problems that involve reasoning about space are well studied; however, reasoning about sequence of events is not as well understood.

Not only must we deepen the research understanding about temporal reasoning, but we must create task-appropriate methods for integrating spatial and temporal dimensions of data into visual representations.

Abstraction – Changing to the Appropriate Representation

To show what is important and why it is important is exceedingly difficult. Illustrators have successfully developed a powerful set of principles for concisely conveying complex information in an appropriate way. The objective is "to create an abstraction that conveys key ideas while suppressing irrelevant detail." The challenge is to be able to assess a situation, extract key features, and visually represent those features and their combinations effectively. This needs to be done dynamically as conditions, interests, and tasks change [Foley, 2000; Smallman et al., 2001].

There are two interrelated issues in dealing with abstraction. The first is the development of an analytic capability to transform data from one representation to another. Selecting relevant information, filtering out unneeded information, performing calculations, sorting, and clustering are all components of data abstraction. Second is the development of techniques for visual abstraction. Visual abstraction involves developing effective representations for different types of information. Visual abstraction also involves the control of emphasis and level of detail. Different representations of the same object may be needed at different levels of detail, depending on the importance of that object for the given task. Secondary visual attributes can also be used to connote additional attributes that are important in reasoning, such as the quality of the data or the confidence in the assessment.

One particular challenge is to develop automatic, user-driven techniques for changing representation. The Pad++ system uses a zoomable interface to navigate a document collection. As the user zooms into a folder or document, more detail is shown. The system changes representations based on the *semantics* of the data, and hence it is possible to do meaningful *semantic zooms*. Other examples of multi-scale interfaces include Woodruff [1998] and Stolte et al. [2002].

Rendering techniques that support cognitive abstraction from visual representations and improve understanding of complex information spaces must also be incorporated into visual analytic solutions. These techniques include advanced rendering methods that aid 3D perception (e.g., advanced illumination, shading, texturing, transparency, shadowing) as well as illustrative rendering techniques that support design and illustration principles (e.g., effective design [Agrawala et al., 2003; Heiser et al., 2004], volume illustration [Rheingans & Ebert, 2002; Svakhine et al., 2003; Treveat & Chen, 2000]).

Recommendation 3.10

Develop visual representation methods for complex information that provide the appropriate level of abstraction.

Research is necessary to:

- Identify alternative visual representations of data that best support different analytical tasks
- Develop transformation methods that allow the user to move among alternative visual representations to facilitate exploration and discovery
- Provide level of emphasis and detail appropriate to the user's data and task.

Uncertainty – Understanding Incomplete or Erroneous Information

Reasoning and working with uncertain information is common in most visual analytics applications. To reach the appropriate conclusions, analysts must remain fully aware of the uncertainties and conflicts present in their information. However, representation of uncertainty is not often considered in current visual analytics systems.

Recommendation 3.11

Develop visual representations that illustrate uncertain, missing, and misleading information, and that allow the analyst to understand the uncertainty inherent in visual analytics applications.

There is no accepted methodology to represent potentially erroneous information, such as varying precision, error, conflicting evidence, or incomplete information. There is no agreement on factors regarding the nature of uncertainty, quality of source, and relevance to a particular decision or assessment. Nevertheless, interactive visualization methods are needed that allow users to see what is missing, what is known, what is unknown, and what is conjectured, so that they may infer possible alternative explanations. Uncertainty must be displayed if it is to be reasoned with and incorporated into the visual analytics process. In existing visualizations, much of the information is displayed as if it were "true" [Finger & Bisantz, 2002].

Integrating Powerful Analysis Tools with Visualization

Data representation and transformation (described more fully in Chapter 4) evolved from the statistics, pattern recognition, and machine-learning communities and have long been a staple in analysis. These approaches are powerful, automated, and quantitative.

Unfortunately, data transformation approaches, by themselves, are insufficient to provide the required insights. Visual analytics couples these computational capabilities with a human decision maker. The hybrid system is more powerful than either the machine or the analyst working alone.

Many possible data transformations may be applicable to a particular problem, but it is not necessarily clear which ones will be of most value in facilitating insight. Visual analytics offers advantages to the user because it provides visual cues that can help the analyst formulate a set of viable models. Also, because visual analytics is qualitative as well as quantitative, there are no assumptions of exact parameters and well-defined boundaries between what is interesting and what is not. A priori criteria of significance may be manipulated based on the judgment of the analyst. The weaknesses of visual analytics are that there are often infinite possibilities in terms of mappings and views, and there is a high potential for information overload in dense information fields.

Many analytic packages support multiple visual representations and computational techniques, although generally each will emphasize one over the other. For example, plots are routinely used to confirm analysis and sampling, and clustering algorithms are often used for data reduction prior to visual exploration. The problem is that the communication between the two forms of analysis is often a thin, one-directional channel.

Recommendation 3.12

Develop visual analytic methods that combine data transformations with interactive visual tools, leveraging powerful computational methods that are developed for continuous and discrete data analysis with human cognitive and perceptual abilities.

An ideal environment for analysis would have a seamless integration of computational and visual techniques. For instance, the visual overview may be based on some preliminary data transformations appropriate to the data and task. Interactive focusing, selecting, and filtering could be used to isolate data associated with a hypothesis, which could then be passed to an analysis engine with informed parameter settings. Results could be superimposed on the original information to show the difference between the raw data and the computed model, with errors highlighted visually. This process could be iterated if the resulting model did not match the data with sufficient accuracy, or the analyst could refocus on a different subspace of information.

An environment that strongly links data transformation and visualization will result in more powerful analysis process that allows the user to draw on the strengths of each approach.

Monitoring Streams of Data – Assessing Situations and Detecting Changes

Many analytic activities involve monitoring a stream of information. The analyst is required to identify and respond to major new developments. We need a new visualization paradigm that will enable analysts to extract relevant information from information streams, gain situational awareness, and formulate appropriate actions.

Streaming data are particularly important for homeland security applications. With the decreasing cost of silicon, it has become cost-effective to deploy new sensor systems that are capable of collecting massive information streams. The systems will collect much more data than can be combined and warehoused in a centralized system. Thus we need to develop fundamentally new techniques to visualize and analyze data in motion.

As an example, we consider an ongoing project at University of Illinois-Chicago National Center for Data Mining. In this project, shown in Figure 3.4, traffic data from the tri-state region (Illinois, Indiana, and Wisconsin) are collected from hundreds of embedded sensors. The sensors are able to identify vehicle weights and traffic volumes. There are also cameras that capture live video feeds, Global Positioning System (GPS) information from selected vehicles, textual accident reports, and weather information. The research challenge is to integrate this massive information

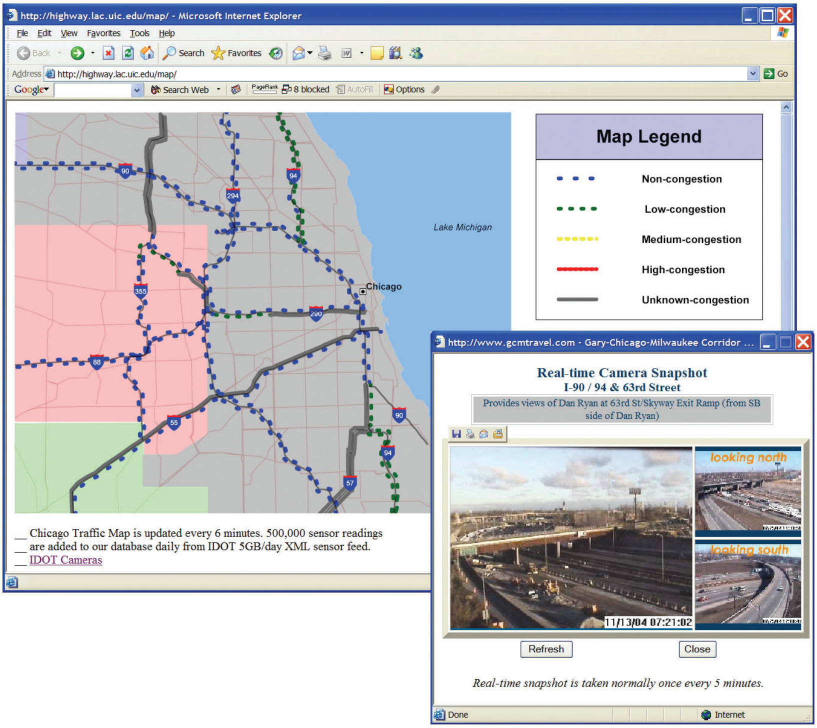


Figure 3.4. Real-time view of Chicago traffic flows that integrates congestion levels, flow, vehicle types, video feeds, and textual accident reports.

flow, provide visualizations that fuse this information to show the current state of the traffic network, and develop algorithms that will detect changes in the flows. Part of this project will involve characterizing normal and expected traffic patterns and developing models that will predict traffic activity when stimulus to the network occurs, as would be expected if there were a terrorist attack. The changes detected will include both changes in current congestion levels and differences in congestion levels from what would be expected from normal traffic levels.

Another example is *Smart Money* magazine's Map of the Market [Wattenberg, 1999], shown in Figure 3.5. It allows the user to monitor the performance of hundreds of stocks in real time as trading is underway. The Map of the Market uses the Treemap visualization technique [Johnson & Shneiderman, 1991; Bederson et al., 2002]. Each rectangle represents a stock (company) and the rectangle's size corresponds to the market capitalization of the company. The color of a rectangle denotes the stock's performance in a given period of time (red-decline and green-advance). The display is interactive and the viewer can easily change the segment of time being reviewed and can focus in on a particular market segment. One of the strengths of the visualization is that it provides a global impression of how the market is doing as a whole as well as the details of individual companies.

As these two examples show, real-time visualization can be a powerful tool in gaining insight from streaming data. However, real-time analytical systems for streaming data are still in their very early stages, as most visual analytics tools are targeted at static data sets.

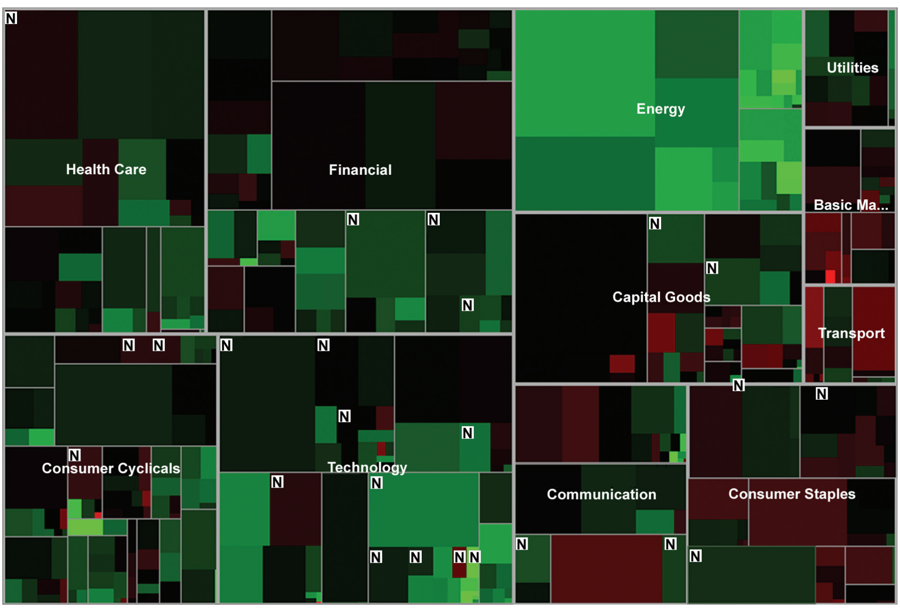


Figure 3.5. Smart Money *magazine's Map of the Market illustrates both high-level overviews and company-level details about stock market activity.*

Recommendation 3.13

Develop visual representations and new analysis techniques for streaming data, including data collected by sensors. Develop visual analytic techniques that detect and show changes in the streams and that integrate with predictive modeling tools.

The research challenge is to create a new class of visualizations for streaming data. Three significant problems must be addressed for streaming data visualizations:

1. Provide **situational awareness** for data streams.
2. Show **changes** in the state of the system and help users identify when the changes are significant.
3. **Fuse** various types of information to provide an integrated view of the information.

Visual representations by themselves are insufficient to answer many analytic questions and must integrate with algorithms for change detection, forecasting, and predictive modeling tools.

It is also important to note that these types of analysis activities rarely occur in a quiet, private setting free of interruptions or distractions. Instead, they often take place under extreme pressure in shared workspaces such as command and control centers. We need a better understanding of human attention and how it affects the analysis activities that a person may be performing. How can we facilitate an analyst acquiring the previous state in an analytic process when some interruption or distraction occurs? Can we design visualizations and systems that are more pliable to the interruptions that are bound to occur, that is, techniques that better facilitate analysts reorienting themselves and resuming prior activities?

Handling Scale

As described in Chapter 1, our ability to collect data is increasing at a faster rate than our ability to analyze it. In this section, we address the issues of visual scalability, information scalability, and software scalability that were raised in Chapter 1. Recall that visual scalability is the capability of visualization representations and visualization tools to display massive data sets effectively. Information scalability is the capability to extract relevant information from massive data streams. Software scalability is the capability of the software to accommodate data sets of varying sizes. We wish to avoid the hidden costs that arise when we build and maintain monolithic, non-interacting, non-scalable software models.

Analytic scalability is the capability of the mathematical algorithms to efficiently accommodate large data sets. As data set sizes and complexity increase, new analytical approaches and algorithms are needed that can handle the increased complexity. The computational complexity of many current visual analytics algorithms is such that these algorithms cannot process data as rapidly as they are received.

Scalability of Visual Representations

The state of the art for representing an information space typically consists of a static representation of the space that users interact with and manipulate to discover patterns within the information. The challenge is that the complexity of large information spaces can overwhelm any single representation. A new class of dynamic and scalable visual representations is needed to enable rich analysis of these information spaces.

The choice of visual representation affects visual scalability. Some visual metaphors scale well in some circumstances, while others do not. To illustrate, we describe six common visual metaphors, which are shown in Figure 3.6 using Visual Insights' ADVIZOR. These metaphors are chosen to be intentionally simple. They represent only a single dimension of the information space and illustrate one approach toward visual scalability.

Bar charts (top left in Figure 3.6) are collections of vertical bars arranged in a window. Two data attributes can be encoded in the bar height and color, and bars can be clustered or stacked to increase the number of attributes. In a bar chart, the minimum possible thickness for each bar is a single pixel, as is the minimum separation between adjacent bars. Assuming a window width of 1000 pixels, at maximum zoom a bar chart can display at most 500 bars. However, especially when there is little structure to index the bars, 50 bars is more realistic. One approach to increasing the scalability of bar charts is to employ dynamic transitions to new representation as the size and complexity of the information space increases. For example, the ADVIZOR bar chart can transition into a smoothed **histogram** (Figure 3.6, bottom right) when the number of bars exceeds the number of available pixels.

Landscapes (middle left in Figure 3.6) are a 3D version of **matrix views** (top right in Figure 3.6). They show 2D tabular data using glyphs of skyscraper-like towers arranged on a grid. Usually, as in Figure 3.6, a landscape is viewed from an angle (from straight overhead, it becomes a matrix view). The height, color, and shape of the towers can potentially encode three data attributes, depending on the nature of the attributes. For example, numerical attributes map well onto height, somewhat well onto color, and poorly onto shape. Categorical attributes map best onto color.

Landscapes can show hundreds to thousands of data elements. Limiting factors are the number of pixels used to render each 3D glyph (typically several hundred), occlusion caused by tall bars in front obscuring short bars in the back and, as for matrix views, how well the numbers of index values match the screen aspect ratio.

Relationship views (middle right in Figure 3.6) show both characteristics of individual data elements and pairwise relationships among them. Nodes correspond to data elements, whose attributes become visual characteristics such as size, color, and shape. The relationships among nodes are encoded as visual characteristics of links (width, color, pattern). For example, the network view in Figure 3.6 shows characteristics of automobile traffic in Chicago by zones (such as the central business district, the large node at the center). Node sizes are number of destination trips, and link widths show zone-to-zone flows. Network views can usefully display a graph with tens to thousands of nodes, with strong dependence on the connectivity, number of links, and inherent structure of the graph. Scalability decreases dramatically as connectivity increases, because many of the connecting links overplot,

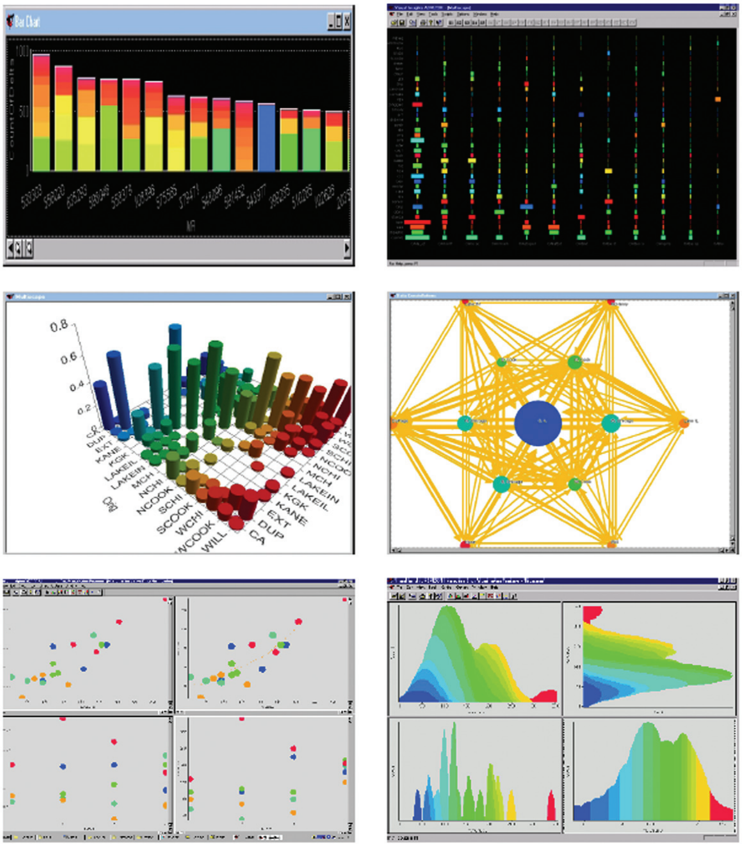


Figure 3.6*. Scalability of common visual representations. Top Left: Bar chart and Top Right: Matrix view, both illustrated with software change data. Middle Left: Landscape and Middle Right: Network view, both illustrated with zone-to-zone traffic flows in metropolitan Chicago. Lower Left: Scatterplot. Lower Right: Histogram.*

causing the display to become confusing. Graph layout algorithms that attempt to minimize overplotting can overcome this to some extent. Their effect is visual accessibility to data rather than display of structure, because distances may not encode relationships between nodes. Visual scalability is limited if layout algorithms destroy or distort "real" relationships (for example, geography) among nodes.

Scatterplots (bottom left in Figure 3.6) can display 100,000 points or more, depending on the data pattern. The primary factor limiting scatterplot scalability is point overplotting: as the number of points increases, points overplot, not only making structure in the data, such as trends or concentrations of points, harder and harder to identify, but also rendering access to details of the data impossible.

More recent work by Fekete and Plaisant [2002] has addressed the challenge of scaling visual representations. They are studying innovative approaches to represent one million discrete items visually without use of aggregation techniques. They are investigating both visual attributes and interaction techniques, such as animation, to facilitate data set exploration. Work by Munzner et al. [2003] on TreeJuxtaposer provided tools for comparing trees of several hundred thousand nodes. The visualization technique, called Accordion Drawing, has recently been extended to work on trees of up to 15 million nodes [Beermann, 2005].

Technology Needs

To scale our visual representations to meet ever-escalating data volumes, we must advance the state of the art in several major areas: visual representation of large data collections, support for multi-type information synthesis, and support for visualization of high-dimensional spaces.

Visual representation of large data collections

We need to extend the state of the art for visual representations to be able to explore heterogeneous multi-source, multi-dimensional, time-varying information streams. We must develop new visual methods to explore massive data in a time-critical manner. We must develop new methods to address the complexity of information and create a seamless integration of computational and visual techniques to create a proper environment for analysis. We must augment our methods to consider visual limits, human perception limits, and information content limits.

Recommendation 3.14

Develop a science of visual scalability that includes new ways to define it, metrics to quantify it, and techniques to increase visual representation scalability.

It is difficult to increase what we cannot measure, so the first step toward increasing visual representation scalability must be to develop ways to measure it. Current scalability metrics do not capture what is important for visual scalability. Thus we must establish metrics that allow us to evaluate both visual metaphors and data representations as they apply to scalable algorithms. The best measurement will evaluate the representations according not only to scale but also to the number of insights, actions, or value achieved for the analyst.

Existing visual representations commonly support display of several orders of magnitude less data than needed to fully represent available data. For the field of visual analytics to achieve its potential, we need to develop new representations and techniques that support display of much greater data volumes. One approach to this is to create approaches that dynamically change form as the size and complexity of the information space increases.

Visual representations to support synthesis

Synthesis includes the capability to fuse the relevant information from divergent multi-source, multi-dimensional, time-varying information streams. This is a grand challenge in visual analytics. Not only must researchers produce new visual representations and data representations for specific data types or information streams but also we must develop methods that synthesize the relevant information into a single information space and develop new visual metaphors that allow the analyst to "look inside" this complex, time-varying space.

Recommendation 3.15

Develop visual representations for information synthesis. The representations should combine relevant information from heterogeneous multi-source, multi-dimensional, time-varying information streams into a single seamless visual representation.

Many visual analytics problems will involve heterogeneous data that must be integrated, synthesized, and viewed at multiple layers. Current visual representations often focus on a single attribute and do not enable analysts to understand the richness in complex heterogeneous information spaces. New representations are needed to help analysts understand complex heterogeneous information spaces. For example, in a crisis management center, analysts need to integrate information involving geospatial locations, text, flows, weather, video feeds, and radiological sensors into a single analytic environment to support real-time emergency management.

Scaling the number of dimensions

Analysis of large information spaces often translates into the analysis of data scattered in very-high-dimensional (VHD) spaces, consisting of hundreds or thousands of variables. Interesting structures in these spaces may be nonplanar or nonlinear, suggesting that the analyst will require more sophisticated tools for analysis.

The challenges of scaling to deal with high-dimensional data affect both the visual representation and interaction techniques and the fundamental data representations and transformations that underlie those visual representations. We address the visual representation and interaction challenge here. The data representation and transformation issues are described in Chapter 4.

Recommendation 3.16

Research and develop visual representation and interaction methods for very-high-dimensional, large information spaces.

Visually representing only a couple dimensions of a high-dimensional space is not necessarily effective in conveying the important content of that space. Not only can this scaled-down representation obscure the complex relationships that may exist within the data but it can also deceive the user with its simplicity. New visual representations and interactions are needed that help represent complex relationships without oversimplifying information. For example, Yang et al. [2004] developed a visualization that represents each of the dimensions of a high-dimensional space, along with all of the values for those dimensions, in a single display. Dimensionality reduction techniques (described in Chapter 4) should go hand in hand with visual representations that help users understand the complexity of their information.

Novel Systems for Generating Visualizations

Our ability to experiment with and evaluate new interactive visual representations depends on our ability to create systems using them. This section focuses on the issues involved with the software that will create visual representations for the analyst's use.

Creating effective visualization representations is a labor-intensive process that requires a solid understanding of the visualization pipeline, characteristics of the data to be displayed, and the tasks to be performed by the analyst. Current visualization software generally has been written in environments where at least some of this necessary information was missing.

In general, it is not possible for the data analyst, who has the best understanding of the data and task, to construct new tools. The development of a visualization application requires a firm understanding of issues of perception, data, information structures, human-computer interaction, and graphics, not to mention the knowledge of the wide range of possible visual mappings available. He or she also has little time to learn how to interpret new visualizations and determine when to use them. Instead, the visual analytics community needs technology to support the rapid creation of visual methods that are tuned to the data, tasks, and users involved in the analysis.

State of the Art

There have been four general approaches for constructing visualization software. The first, and most common, is to build a general-purpose visualization tool that targets a particular domain. Examples of systems following this approach include IN-SPIRE visual text analysis software [Hetzler & Turner 2004]; OpenDx (formerly IBM Visualization Data Explorer [http://www.opendx.org]); the general-purpose AVS software [Haeberli, 1988 and http://www.avs.org]; XmdvTool [Ward, 1994] and Spotfire [http://www.spotfire.com] for multi-dimensional data; and Rivet [Bosch, 2000].

A second broad approach for constructing visualizations involves visualization toolkits. The most obvious approach is to build component-based visualization libraries such as the InfoVisToolkit [Fekete, 2004 and http://ivtk.sourceforge.net/], ILOG's library, (www.ilog.com), Visual Decisions' In3D, or AT&T Bell Laboratories' Vz. These libraries simplify software construction by providing high-level programming constructs for creating visualizations. However, programming is still required to use them.

A third related approach at a higher level of abstraction is to build visualization components that work well together. Examples of this approach include North's Snap-Together Visualizations [North et al., 2002] and Eick's ADVIZOR visualization components [Eick, 2000]. Although visual components provide more capability than object libraries, the currently available technologies are not sufficient. It is very difficult to create reusable software in general and even harder to create reusable user interface software that includes visualization software.

A fourth approach toward constructing visualizations involves systems that automatically generate visualization software. This approach includes the generation of visualizations based on creating a large database of examples that can be queried based on user needs [Zhou et al., 2002] and the use of rule-based techniques to match task and data characteristics to appropriate visualizations [Mackinlay, 1986; Roth & Mattis, 1991; Zhou, 1999]. Taxonomies of methods have also been used as a mechanism to facilitate the rapid development of effective visualizations [Chi, 2000; Fujishiro et al., 2000]. Analysis of domains [Espinoza et al., 1999], user tasks [Casner, 1991], and data characteristics [Zhou & Feiner, 1996] have also been used in the design of visual presentations.

Technology Needs

Research is needed to move beyond the current state of handcrafting special-purpose visual representations to reach a future in which visual analytics software can rapidly adapt to new data and analytical needs.

Recommendation 3.17

Develop tools and techniques to incrementally automate the common tasks involved with creating visualizations.

We believe that it will be quite difficult to provide a complete solution for the problem of generating visual representations to support visual analytics. However, we think that there is an opportunity for semi-automatic methods that help users with many of the routine tasks involved with creating visualizations. There is a vast difference in quality between visualizations created by skilled artists and those created using widely available visualization software. We need to develop incremental techniques and software to reduce this gap.

Although we need new and novel visual representations, we also need robust, easy-to-use software that implements well-known metaphors. For example, the community needs software to produce visualizations for timelines, graphs, trees, and geospatial data.

Recommendation 3.18

Develop high-quality visualization components for well-known visual representations.

Currently, each visualization system re-implements these basic visual representations at great expense and effort. We need reusable visualization components that embody well-known visual representations. In many ways, these components are like the mathematical software libraries that are now widely distributed.

We need to create the next-generation technology for producing visual analytics systems. The current generation of visual analytics tools has been developed at great expense and targets a narrow range of specific problems. To achieve their potential, visual analytics tools need to target a much broader range of problems; therefore, we must reduce the development costs to create these tools. New ideas and technologies are needed to produce these tools. We are optimistic that this is now possible with the emergence of standards such as XML and web services.

Summary

Visual representations and interaction technologies provide the mechanism for allowing the user to see and understand large volumes of information at once. Scientific principles for depicting information must provide the basis for visual representations, and principles are needed for new interaction approaches to support analytical techniques. Together, these foundations provide the basis for new visual paradigms that can scale to support analytical reasoning in many situations.

Visual design theory is more mature than interaction theory, so investments in the further development of interaction theory should take priority. Interaction theory must take into account the time constraints associated with varying levels of urgency in an analytic task. The application of visual representations and interactions must necessarily be adapted to fit the needs of the task at hand. The issues of scale also profoundly affect the design of visual representations and interactions and must be considered explicitly in the design of new visual representation and interaction techniques.

Creating effective visual representations is a labor-intensive process. We need new methods for constructing visually based systems that simplify the development process and result in better-targeted applications.

Summary Recommendations

The following high-level recommendations summarize the detailed recommendations from this chapter. These actions are necessary to advance the science of visual representations in support of visual analytics.

Recommendation

Create a science of visual representations based on cognitive and perceptual principles that can be deployed through engineered, reusable components. Visual representation principles must address all types of data, address scale and information complexity, enable knowledge discovery through information synthesis, and facilitate analytical reasoning.

Visual representations and interaction techniques provide the analyst and the first responder with their understanding of developing situations so that they may take action. A science of visual representations has been developed to support scientific applications, but different visual representations are needed to address the diverse data types that are relevant to homeland security missions. These data must be combined and presented to the user in a way that allows the user to understand their meaning, regardless of the data type or format of the original data. The goal is to expose all relevant data in a way that facilitates the reasoning process to enable action.

Recommendation

Develop a new suite of visual paradigms that support the analytical reasoning process.

These visualizations must:

- Facilitate understanding of massive and continually growing collections of data of multiple types
- Provide frameworks for analysis of spatial and temporal data
- Support understanding of uncertain, incomplete, and often misleading information
- Provide user- and task-adaptable, guided representations that enable full situation awareness while supporting development of detailed actions
- Support multiple levels of data and information abstraction
- Facilitate knowledge discovery through information synthesis, which is the integration of data based on their meaning rather than the original data type.

No one visual paradigm can address all possible tasks and situations. Therefore, we recommend developing a suite of visual paradigms that address multiple situations ranging from vulnerability analysis to real-time monitoring to emergency response support. The scale of data, especially in the forms of sensor, text, and imagery, is rapidly growing. Data are continually growing and changing, and visual representations must help analysts understand the changing nature of their data and the situations they represent. Likewise, many data are associated with a particular place and time. Representing these spatial and temporal qualities is necessary to provide analytical understanding. Furthermore, the visualization process is complicated by the need to support understanding of missing, conflicting, and deceptive information in an analytic discourse that is guided by the individual's knowledge and his or her task.

Recommendation

Develop a new science of interactions that supports the analytical reasoning process. This interaction science must provide a taxonomy of interaction techniques ranging from the low-level interactions to more complex interaction techniques and must address the challenge to scale across different types of display environments and tasks.

Interaction is the fuel for analytic discourse. Although the fundamental principles of interaction have been around for more than a decade, they do not address the needs for higher-order interaction techniques, such as task-directed or hypothesis-guided discourse, to support the analysis process. A new scientific theory and practice are critical to address the complexity of homeland security needs for analysis, prevention, and response. These interaction techniques must adapt to the particular dimensions of the analytical situation, ranging from longer-term analytical assessments to urgent and highly stressful emergency response support tasks. These interactions must be adaptable for use in platforms ranging from the large displays in emergency management control rooms to field-deployable handheld devices in the hands of first responders. This is a high priority for initial investments.

References

Agrawala M, D Phan, J Heiser, J Haymaker, J Klingner, P Hanrahan, and B Tversky. 2003. "Designing Effective Step-by-Step Assembly Instructions." *ACM Transactions on Graphics (TOG)*. 22(3):828-837.

Ahlberg C and B Shneiderman. 1994. "Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays." In *Proceedings of CHI'94*, pp. 313-317. ACM Press, Boston.

Baddeley AD and GJ Hitch. 1974. "Working Memory." *The Psychology of Learning and Motivation*, Vol. 8, ed. GH Bower. Academic Press, New York.

Barlow T and P Neville. 2001. "Visualization for Decision Tree Analysis in Data Mining." In *Proceedings of the IEEE Symposium on Information Visualization 2001*, pp. 149-152. IEEE Computer Society, Washington, D.C.

Bartram L and C Ware. 2002. "Filtering and Brushing with Motion." In *Information Visualization* 1(1):66-79. Palgrave Macmillan, Basingstoke, New York.

Baudisch P, E Cutrell, D Robbins, M Czerwinski, P Tandler, B Bederson, and A Zierlinger. 2003. "Drag-and-Pop and Drag-and-Pick: Techniques for Accessing Remote Screen Content on Touch and Pen-Operated Systems." In *Proceedings of INTERACT 2003*, pp. 57-64.

Bederson B, B Shneiderman, and M Wattenberg. 2002. "Ordered and Quantum Treemaps: Making Effective Use of 2D Space to Display Hierarchies." *ACM Transactions on Graphics* 21(4):833-854.

Bederson B, A Clamage, M Czerwinski, and G Robertson. 2004. "DateLens: A Fisheye Calendar Interface for PDAs." *ACM Transactions on Computer-Human Interaction* 11(1):90-119.

Bederson B, J Hollan, K Perlin, J Meyer, D Bacon, and G Furnas. 1996. "Pad++: A Zoomable Graphical Sketchpad for Exploring Alternate Interface Physics." *Journal of Visual Languages and Computing* 7(1):3-31.

Beermann D, T Munzner, and G Humphreys. In Press. "Scalable, Robust Visualization of Large Trees." *Proceedings of EuroVis 2005*.

Bertin J. 1981. *Graphics and Graphic Information Processing*. Walter De Gruyter Inc., Berlin.

Bosch R, C Stolte, D Tang, J Gerth, M Rosenblum, and P Hanrahan. 2000. "Rivet: A Flexible Environment for Computer Systems Visualization." *Computer Graphics* 34(1):68-73.

Cadiz JJ, G Venolia, G Jancke, and A Gupta. 2002. "Designing and Deploying an Information Awareness Interface." In *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*, pp. 314-323.

Card SK, JD Mackinlay, and B Shneiderman. 1999. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufman, Los Altos, California.

Card S, T Moran, and A Newell. 1983. *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Casner S. 1999. "A Task-Analytical Approach to the Automated Design of Graphic Presentations." *ACM Transactions on Graphics* 10(2):111-151.

Chi E. 2000. "A Taxonomy of Visualization Techniques Using the Data State Reference Model." In *Proceedings of the IEEE Symposium on Information Visualization 2000*, pp. 69-75.

Cleveland WS. 1985. *The Elements of Graphing Data*. Hobart Press, Summit, New Jersey.

Cleveland W and M McGill. 1984. "Graphical Perception: Theory, Experimentation and Application to the Development of Graphical Methods." *Journal of the American Statistical Association* 79(387):531-554.

Dent BD. 1999. *Cartography: Thematic Map Design*. Wm. C. Brown Publishers, Dubuque, Iowa.

Eick SG. 2000. "Visual Discovery and Analysis." *IEEE Transactions on Visualization and Computer Graphics*. 6(1):44-58.

Eick SG, JL Steffen, and EE Sumner. 1992. "Seesoft – A Tool for Visualizing Line Oriented Software Statistics." *IEEE Transactions on Software Engineering* 18(11):957-968.

Espinosa O, C Hendrickson, and G Garrett Jr. 1999. "Domain Analysis: A Technique to Design a User-Centered Visualization Framework." In *Proceedings of the 1999 IEEE Symposium on Information Visualization 1999 (InfoVis 99)*, pp. 44-52.

Fekete J. 2004. "The InfoVis Toolkit." In *Proceedings of the 10th IEEE Symposium on Information Visualization (InfoVis'04)*, pp. 167-174. IEEE Press, Piscataway, New Jersey.

Fekete J and C Plaisant. 2002. "Interactive Information Visualization of a Million Items." In *Proceedings of the IEEE Symposium on Information Visualization (InfoViz 2002)*, pp. 117-124. IEEE Computer Society, Washington, D.C.

Finger R and A Bisantz. 2002. "Utilizing Graphical Formats to Convey Uncertainty in a Decision-Making Task." *Theoretical Issues in Ergonomic Science* 3(1):1-25.

Foley J. 2000. "Getting There: The Top Ten Problems Left." *IEEE Computer Graphics and Applications 20(1)*:66-68.

Fujishiro I, R Furuhata, Y Ichikawa, and Y Takeshima. 2000. "GADGET/IV: A Taxonomic Approach to Semi-Automatic Design of Information Visualization Applications Using Modular Visualization Environment." In *Proceedings of the IEEE Symposium on Information Visualization 2000*, pp. 77-83. IEEE Computer Society, Washington, D.C.

Furnas G. 1986. "Generalized Fisheye Views." In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, pp. 16-23. ACM Press, New York.

Gamma E, R Helm, R Johnson, and J Vlissides. 1994. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley, Reading, Massachusetts.

Greenberg S and M Rounding. 2001. "The Notification Collage: Posting Information to Public and Personal Displays." In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, pp. 515-521. ACM Press, New York.

Greene S, G Marchionini, C Plaisant, and B Shneiderman. 2000. "Previews and Overviews in Digital Libraries: Designing Surrogates to Support Visual Information-Seeking." *Journal of the American Society for Information Science* 51(3):380-393.

Haeberli P. 1988. "ConMan: A Visual Programming Language for Interactive Graphics." *Proceedings of the 15th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 103-111. ACM Press, New York.

Hauptmann AG and P McAvinney. 1993. "Gestures with Speech for Graphics Manipulation." *International Journal of Man-Machine Studies* 38:231-249.

Haykin S, ed. 2000. *Unsupervised Adaptive Filtering*. Wiley-Interscience, New York.

Heiser J, D Phan, M Agrawala, B Tversky, and P Hanrahan. 2004. "Identification and Validation of Cognitive Design Principles for Automated Generation of Assembly Instructions." In *Proceedings of the Working Conference on Advanced Visual Interfaces*, pp. 311-319. ACM Press, New York.

Hetzler E and A Turner. 2004. "Analysis Experiences Using Information Visualization." *IEEE Computer Graphics and Applications* 24(5):22-26.

Heuer RJ. 1999. *Psychology of Intelligence Analysis*. U.S. Government Printing Office, Washington, D.C.

Hutchings D and J Stasko. 2004. "Shrinking Window Operations for Expanding Display Space." In *Proceedings of Advanced Visual Interfaces*, pp. 350-353. ACM Press, New York.

Johnson B and B Shneiderman. 1991. "Tree-Maps: A Space Filling Approach to the Visualization of Hierarchical Information Structures." In *Proceedings of the 2nd IEEE Conference on Visualization '91*, pp. 284-291, 1991, San Diego, California. IEEE Computer Society, Washington, D.C.

Kapler T and W Wright. 2004. "GeoTime Information Visualization." *IEEE InfoVis 2004*. Oculus Info Inc. Available at http://www.oculusinfo.com/papers/KaplerWright_GeoTime_InfoViz_Final_Conf.pdf.

Kinsbourne M and RE Hicks. 1978. "Functional Cerebral Space; a Model for Overflow, Transfer and Interference Effects in Human Performance: a Tutorial Review." In *Attention and Performance VII*, ed. J Requin. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Kollin J. 1993. "A Retinal Display for Virtual-Environment Applications." In *Proceedings of SID-International Symposium, Digest of Technical Papers*, pp. 827.

MacEachren AM. 1995. *How Maps Work: Representation, Visualization and Design*. The Guilford Press, New York.

Mackinlay J. 1986. "Automating the Design of Graphical Presentations of Relational Information." *ACM Transactions on Graphics* 5(2):110-141.

Martin GL. 1989. "The Utility of Speech Input in User-Computing Interfaces." *International Journal of Man-Machine Studies* 30:355-375.

Mooshage O, H Distelmaier, and M Grandt. 2002. "Human Centered Decision Support for Anti-Air Warfare on Naval Platforms." In *Proceedings of the RTO HFM Symposium on the Role of Humans in Intelligent and Automated Systems*, RTO/NATO 2002 RTO-MP.

Munzner T, G Guimbretiere, S Tasiran, L Zhang, and Y Zhou. 2003. "TreeJuxtaposer: Scalable Tree Comparison Using Focus+Context with Guaranteed Visibility." *ACM Transactions on Graphics* 22(3):453-462. ACM Press, New York.

Newell A. 1990. *Unified Theories of Cognition*. Harvard Press, Cambridge, Massachusetts.

Norman DA. 1993. *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. Perseus Books, New York.

North C, N Conklin, K Indukuri, and V Saini. 2002. "Visualization Schemas and a Web-Based Architecture for Custom Multiple-View Visualization of Multiple-Table Databases." *Information Visualization* 1(3/4):211-228. Palgrave MacMillan, Basingstoke, New York.

Oviatt SL. 1999. "Ten Myths of Multimodal Interaction." *Communications of the ACM* 42(11):74-81.

Plaisant C, D Carr, and B Shneiderman. March 1995. "Image-Browser Taxonomy and Guidelines for Designers." *IEEE Software* 12(2):21-32.

Rheingans P and DS Ebert. 2002. "Volume Illustration: Non-Photorealistic Rendering of Volume Models." *IEEE Transactions on Visualization and Computer Graphics* 7(3):253-265.

Robertson G, K Cameron, M Czerwinski, and D Robbins. 2002. "Animated Visualization of Multiple Intersecting Hierarchies." *Journal of Information Visualization* 1(1):50-65.

Robertson G, M Czerwinski, K Larson, D Robbins, D Thiel and M van Dantzich. 1998. "Data Mountain: Using Spatial Memory for Document Management." In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology*, pp. 153-162. ACM Press, New York.

Robertson G, E Horvitz, M Czerwinski, P Baudisch, DR Hutchings, B Meyers, D Robbins, and G Smith. 2004. "Scalable Fabric: Flexible Task Management." In *Proceedings of the Working Conference on Advanced Visual Interfacesence*, pp. 85-89. ACM Press, New York.

Rogowitz BE and LA Treinish. 1993. "An Architecture for Perceptual Rule-Based Visualization." In *Proceedings of the 4th IEEE Conference on Visualization*, pp. 236-243.

Roth S and J Mattis. 1991. "Automating the Presentation of Information." In *Proceedings of the IEEE Conference on Artificial Intelligence Applications*, pp. 90-97.

Sharp Systems of America. 2004. "Sharp's 3D LCD Technology Now Available for Desktop Computers." Available at http://www.sharpsystems.com/news/press_releases/release.asp?press=59.

Shneiderman B. 1996. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations." In *Proceedings of the 1996 IEEE Symposium on Visual Languages*, pp. 336-343. IEEE Computer Society, Washington, D.C.

Smallman H, M St. John, H Oonk, and M Cowen. 2001. "SYMBICONS: A Hybrid Symbology That Combines the Best Elements of SYMBOLS and ICONS." In *Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomics Society*, pp. 114-119.

Smith S, RD Bergeron, and G Grinstein. 1990. "Stereophonic and Surface Sound Generation for Exploratory Data Analysis." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People*, pp. 125-132. ACM Press, New York.

Spence R. 2000. *Information Visualization*. ACM Press, New York.

Stasko J, T Miller, Z Pousman, C Plaue, and O Ullah. 2004. "Personalized Peripheral Information Awareness Through Information Art." In *Proceedings of UbiComp (2004)*, pp. 18-35.

Stolte C, D Tang, and P Hanrahan. 2002. "Polaris: A System for Query, Analysis and Visualization of Multidimensional Relational Databases." *IEEE Transactions on Visualization and Computer Graphics* 8(1):52-65.

Svakhine N and DS Ebert. 2003. "Interactive Volume Illustration and Feature Halos." *Pacific Graphics 2003*.

Tang CW and SA Van Slyke. 1987. "Organic Electroluminescent Diodes." *Applied Physics Letters* 51(12):913-915.

Treisman A and A Davies A. 1973. "Divided Attention to Ear and Eye." pp. 101-117 in *Attention and Performance IV*, ed. S Kornblum. Academic Press, New York and London.

Tufte E. 1983. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, Connecticut.

Tufte E. 1990. *Envisioning Information*. Graphics Press, Cheshire, Connecticut.

Tufte E. 1997. *Visual Explanations: Images and Quantities, Evidence and Narrative*. Graphics Press, Cheshire, Connecticut.

Tukey JW. 1977. *Exploratory Data Analysis*. Addison-Wesley, Reading, Massachusetts.

Tversky B, JB Morrison, and M Betrancourt. 2002. "Animation: Can It Facilitate?" *International Journal of Human-Computer Studies* 57:247-262. Academic Press Inc., Duluth, Minnesota.

van Dam A. 2001. "User Interfaces: Disappearing, Dissolving and Evolving." *Communications of the ACM* 44(3):50-52.

Wang JF, GE Jabbour, EA Mash, J Anderson, Y Zhang, PA Lee, NR Armstrong, N Peyghambarian, and B Kippelen. 1999. "Oxadiazole Metal Complex for Organic Light-Emitting Diodes." *Advanced Materials* 11(15):1266-1269.

Ward MO. 1994. "XmdvTool: Integrating Multiple Methods for Visualizing Multivariate Data." In *Proceedings of the Conference on Visualization '94*, pp. 326-333. IEEE Computer Society Press, Los Alamitos, California.

Ware C. 2000. *Information Visualization: Perception for Design*. Academic Press, distributor. Published by Morgan Kaufmann, San Francisco.

Wattenberg M. 1999. "Visualizing the Stock Market." In *Proceedings of SICGHI '99 Extended Abstracts*, pp. 188-189, Pittsburgh, Pennsylvania.

Wickens CD. 1980. "The Structure of Attentional Resources." In *Attention and Performance VIII*, eds. R Nickerson and R Pew. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Wilkinson L. 1999. *The Grammar of Graphics*. Springer-Verlag, New York.

Woodruff A, J Landay, and M Stonebraker. 1998. "Constant Density Visualizations of Non-Uniform Distributions of Data." In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology*, pp. 19-28. ACM Press, New York.

Yang J, A Patro, S Huang, N Mehta, MO Ward, A Elke, and E Rundensteiner. 2004. "Value and Relation Display for Interactive Exploration of High Dimensional Datasets." In *Proceedings of the IEEE Symposium on Information Visualization 2004 (InfoVis 2004)*, pp. 73-80. IEEE Computer Society, Washington, D.C.

Zhou M. 1999. "Visual Planning: A Practical Approach to Automated Visual Presentation." In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pp. 634-641.

Zhou M and S Feiner. 1996. "Data Characterization for Automatically Visualizing Heterogeneous Information." In *Proceedings of the 1996 IEEE Symposium on Information Visualization*, pp. 13-20. IEEE Computer Society, Washington, D.C.

Zhou M, M Chen, and Y Feng. 2002. "Building a Visual Database for Example-Based Graphics Generation." In *Proceedings of the 2002 IEEE Symposium on Information Visualization*, pp. 23-30. IEEE Computer Society, Washington, D.C.