Presented to the Interdisciplinary Studies Program:

UNIVERSITY OF OREGON
APPLIED INFORMATION MANAGEMENT

Applied Information Management and the Graduate School of the University of Oregon in partial fulfillment of the requirement for the degree of Master of Science

Information Visualization: Concepts and Techniques Enabling Exploration, Discovery, and Insight within Large Data Sets (Big Data)

CAPSTONE REPORT

Brian T. Fabrizio
Knowledge Base Analyst
Givaudan

University of Oregon
Applied Information Management
Program

February 2013

Continuing Education
1277 University of Oregon
Eugene, OR 97403-1277
(800) 824-2714

Approved by

Dr. Linda F. Ettinger
Senior Academic Director, AIM Program

Information Visualization: Concepts and Techniques Enabling Exploration, Discovery, and Insight within Large Data Sets (Big Data)

Brian T. Fabrizio

Givaudan

Abstract

This annotated bibliography examines information visualization concepts and techniques supporting exploration, discovery, and crystallization of new knowledge based on visual recognition of hidden patterns and structures in big data (Offenhuber, 2010), and effectiveness evaluation methodologies within specific domains. Literature published between 2000 - 2012 suggests considerations for software development teams tasked with creating information visualization solutions. Results include a taxonomy of terms and concepts, a concept map, and over a dozen images demonstrating visualization techniques.

Keywords: information visualization, visualization techniques, information design, structured data, unstructured data, big data, software development

Table of Contents

List of Figures

Introduction to the Annotated Bibliography

Problem

Businesses, industries, governments, universities, scientists, consumers, and nonprofits of all sizes are generating data at unprecedented levels and at an incredible pace (Gordon-Murnane, 2012). Manyika et al. (2011) state that collecting, storing, and mining what is being termed as big data, for insights, can create significant value for organizations, enhancing productivity, innovation, and competitiveness. While exact definitions may vary, big data can be generally defined as massive amounts of stored content (structured or unstructured) requiring analysis in real time to get useful and actionable answers (Arnold, 2012).

According to Boeri (2012) and Fekete, Vanwijk, Stasko, and North (2008), traditional business intelligence (BI) applications, geared for automatically finding interesting facts in large datasets, fail to address the three main dimensional challenges of big data: (a) variety of data formats, (b) velocity of incoming data, and (c) volume of overall data. Few (2006), more specifically notes that BI vendors have made tremendous progress in handling large data repositories; however, little progress has been made in exploring the information efficiently, effectively, and with high levels of cognition. Arnold (2012) asserts that the challenges of increasingly massive structured and unstructured data flows and repository sizes will lead us to the requisite implementation of specialized tools needed to cope with, and respond to, the future information environment we are creating.

Business context. The world's leading food and beverage companies are in the business of continuously providing innovative flavor and application solutions for a wide range of consumer products around the world (Khatchadourian, 2009). Increasing volumes of structured and unstructured data related to commercial food and beverage projects enter the organizations on a daily basis through commercial and proprietary global project management software. Each project is manually assessed according to (a) strategic business and service-grid rules, (b) technical competencies, and (c) multifaceted customer-supplied criteria. Decisions are concurrently made in terms of appropriate research and development (R&D) (Spencer, 2012) and portfolio management (PM) personnel (Mills, 2012) for flavor or finished product submissions to commercial product development customers throughout the project's lifecycle.

Based on over twenty-five years of food and beverage product portfolio management experience, Mills (2012), a senior director with one the world's largest flavor companies, identifies four primary challenges faced by application development and flavor portfolio knowledge managers during an R&D lifecycle: (a) establishment of comprehensive structured and unstructured information views related to portfolio products, consumer understanding, sensory, marketing and research data that is stored in varying formats; (b) the generation of what are termed "real-time data views" (delivered immediately) based on continual organizational data inflows; (c) the isolation of existing product profiles that specifically meet customer requirements and restrictions amidst numerous enterprise databases; and (d) the ability to pull all forms of organizational data, information, and knowledge into a single view, enabling rapid insight and discovery of existing and market-proven customer solutions.

Cognos, IBM's business intelligence (BI) and performance management suite, is a popular software reporting application commonly used within the food and beverage industry to provide business users without technical knowledge the ability to extract structured corporate data from database systems, analyze that data, and assemble reports (Rouse, 2010). R&D application scientists and portfolio knowledge base managers manually search isolated enterprise databases for flavor profile information via keywords that meet project requirements, defined in part by a set of predetermined project restrictions, as a way to determine relevance on a project-by-project basis; factors include existing customer diversity, application diversity, sales history, regulatory restrictions, consumer understanding, and sensory studies. Exhaustive result sets are frequently returned, requiring additional manual scanning, filtering, organizing, and analysis by R&D and PKBM specialists to determine information relevance.

According to Steinke (2012), senior vice president of flavor and application development with a world leading food and beverage research organization, these isolated, manual searches, are problematic because they: (a) jeopardize project submission deadlines due to non-optimized procedures and views, (b) do not provide existing or derived qualitative and quantitative information assisting in holistically informed screening, (c) produce highly variable decisions due to considerable reliance on tacit knowledge, (d) limit the organization's ability to focus on novel product creation, and (e) do not include unstructured or real-time information that is typically associated with active customer projects.

Purpose

Efficiently and effectively exploring large information collections becomes increasingly difficult as the volume and differing types of information grow (Kreusler, Lopez, & Schumann, 2000). The combination of (a) cognitive complexity, (b) broad range of data domains, and (c) varying display methods and technologies has driven a reevaluation of presentation and interaction solutions for human information consumption following computer-automated data aggregation (Arias-Hernandez, Dill, Fisher, & Green, 2011). As Few (2006) points out, without strategic, coordinated, and standardized data handling methodologies such as data warehousing, requisite information is spread across too many disparate sources, often generating incomplete and unreliable results that do not lend to meaningful information visualizations.

Information visualization is a process of constructing visual presentations of abstract quantitative data (Ahokas, 2008). These visualizations reveal hidden patterns to domain experts by presenting data based on computational intelligence methods such as data mining, machine learning, inductive reasoning strategies, expert systems, fuzzy logic and neural networks (Kovalerchuk, 2003). As Zhang (2001) points out, these methods often feed data into simulation and decision support systems that generate reports, but lack discovery and interpretation capabilities within specific and complex information consumption environments. To support discovery within large data collections and exploratory processes of R&D specialists, information visualizations help influence and guide the questions that are asked, constraints that are addressed, and tasks that arise during the projects lifecycle (Fekete, Vanwijk, & Stasko, 2008). This sort of human-information feedback relationship makes it possible for researchers and application specialists to obtain internal mental models of the information content in large datasets; models which subsequently can be used for characterization, prediction, and/or decision making (Nagel, 2006).

The purpose of this Annotated Bibliography is to identify (a) information visualization concepts and techniques that can support the exploration, discovery, and crystallization of new knowledge based on visual recognition of hidden patterns and structures in big data (Offenhuber, 2010) and (b) evaluation methodology that can be used to determine the effectiveness of information visualization within a specific information domain. Haroz and Whitney (2012) specifically identify limitations of attention as strongly modulating the effectiveness of information visualizations, particularly the ability to: (a) establish a qualitative information overview, (b) detect unexpected information, (c) spot subtle trends, and (c) notice random anomalies. As noted by Fekete, Vanwijk, and Stasko (2008) good information visualization allows users to forget the system, focusing on exploration, discovery, and insight. With this goal in mind, the four abilities presented by Haroz and Whitney (2012) are pre-selected as a way to frame the identification of new information visualization techniques and supporting display technologies.

Research Questions

Main question. How can information visualization techniques and supporting display technologies aid members of software development teams in creating software that enables exploring and discovering new knowledge and insight within big data systems as noted by Offenhuber, (2010).

Sub-questions.

How can information visualization techniques and supporting display technologies support visual recognition of hidden patterns and structures in information exploration?

How can information visualization techniques and supporting display technologies support visual recognition of hidden patterns and structures in information discovery?

What specific information visualization techniques are currently recommended for this work?

How can information visualizations be evaluated in terms of decision support effectiveness?

Audience

Kreusler, Lopez, and Schumann (2000) assert that achieving flexible, useable, and valuable visualizations (i.e., preprocessing large quantities of heterogeneous information, displaying information context, and supporting a variety of exploration tasks) carry over entirely new qualities of problems with information presentation. Few (2006) notes that preattentive processing, the early stage of visual perception that rapidly occurs below the level of consciousness, is tuned to detect certain sets of visual attributes based on specific visualization methods and plays a large role in easily assimilating meaning from information visualizations. As Bihanic and Polacsek (2012) confirm, visual perception of information amplifies cognition by expressing semantic connections between data.

Forsell (2012) notes that information visualization design must support perceivable, comprehensible, and usable representations of data. The primary audience for this annotated bibliography is software development teams tasked with designing information visualization software applications. A secondary audience is top- level executives responsible for strategic investment, planning, and execution of organizational information interaction technology initiatives.

Significance

Research by Manyika et al. (2011) reveals that the United States alone faces a shortage of 140,000 to 190,000 people with analytical expertise and 1.5 million managers and analysts with the skills to understand and make decisions based on presentations of big data. For information visualization to be truly effective, solution designers must be aware, and design accordingly, based on the ways visualization can support, and negatively attack, the perceptual, cognitive, and motor capabilities of human end users, impairing or preventing the assimilation of valuable information and knowledge (Conti, Ahamad, & Stasko, 2005).

As Arnold (2012) states, the value of big data lies in analyzing information that is holistically accessed, organized and visualized easily for insight discovery and rapid decision making purposes. Lin (1997) identifies specific cases where information visualization value can be directly measured: (a) when there is a good underlying structure so that information items close to one another can be inferred to be similar, (b) when users are unfamiliar with a collections contents, (c) when users have limited understanding of how a system is organized and prefer a less cognitively loaded method of exploration, (d) when users have difficulty verbalizing the underlying information need, and (e) when information is easier to recognize than describe.

Gordon-Murnane (2012) points out that big data is a significant concern for organizations of all sizes, spurring new innovations and new product opportunities in order to achieve cost savings and efficiencies, while using predictive analytics that enable businesses to understand what their customers want now as well is in the future. Within an enterprise environment, information visualizations must convey objective and clear business insight that users can quickly understand, as opposed to subjective impressions (Gaviria, 2008).

Research Delimitations

This section of the paper focuses on factors that determine the scope of this Annotated Bibliography, including the literature collection timeframe, topic focus, literature sources, intended audience, and key concepts (Simon, 2011).

Timeframe. Information visualization is moving out of research laboratories with a growing number of commercial products, additions to statistical packages, and commercial development environments geared towards users such as CEOs, executives, politicians, and administrators (Plaisant, 2004). As Keim (2002) points out, the last decade has seen a large number of new information visualization techniques supported by modern visualization technologies. To ensure that visualization methods supported by modern technology are represented, literature no older than 2000 is reviewed.

Topic focus. Commercial and research environments place significant importance on information evaluation involving metrics that can be explored, assessed, and judged with clarity, assurance, and accuracy (Fekete, Wijk, Stasko, & North, 2008). According to Li, Feng, and Li (2011), many consumers of data derived for commercial insight and discovery lack the mathematical mindset to benefit from traditional information presentations seen in research and business domains. This Annotated Bibliography focuses on information visualization techniques easily perceived by any user with ancillary mention of visualization display technologies that enable such visualizations.

Literature sources. In order to provide the grounds for legitimization of the research questions proposed in the study as well as validate the approach proposed by the study (Levy & Ellis, 2006), resources supporting this Annotated Bibliography consist primarily of peer-reviewed articles, proceedings, and published books located within the University of Oregon (UO) online library. Peer-reviewed literature is preferred as manuscripts are circulated, discussed, and edited according to the documents ability to (a) contribute significantly to the content area covered by the journal, (b) communicate with clarity and conciseness, and (c) follow style guidelines (American Psychological Association, 2009). Menachemi and Ginter (n.d.) outline the anatomy of scholarly literature suitable for peer review noting that such a document must (a) frame the manuscript around an important topic, (b) synthesize previous literature in a succinct and meaningful way, and (c) communicate the academic and practical implications of the research.

Literature related to aspects of the information visualization life-cycle including (a) methodology and techniques, (b) perceptual considerations, (c) visualization technologies, and (c) intelligent information generation systems are retrieved from the following databases based on topic breadth and relevance:

- JSTOR
- Academic Index.net
- IEEE Computer Society
- Academic Search Premier
- Web of Science
- CiteSeerX

Intended audience. Information visualization designers assert that the graphical presentation of information is their fundamental goal, which they achieve by applying principles basic to art and design – namely, hierarchies of importance, spatial relationships, layering, contrast versus analogy, legibility, and readability (Stone, n.d.). As a complement to traditional software development methodologies, Zhang, Johnson, Malin, and Smith (2002) recommend a human-centered visualization framework consisting of (a) functional analysis, (b) user analysis, (c) task analysis, and (d) representational analysis that provides requirements for application implementation. This Annotated Bibliography primarily serves as a reference for software development teams that require insight to proven information visualization techniques for large data sets.

Reading and Organization Plan Preview

All literature included in this Annotated Bibliography follows standardized, yet flexible, review, encoding, and analysis processes based on keyword/phrase existence and frequency. Frequency counts are used to generate treemap visualizations indicating relative frequencies between keywords and phrases (a) within a given document and (b) across the entire Annotated Bibliography. The list of keywords and phrases used to generate the treemaps is standardized and consistently applied to all documents within the Annotated Bibliography. Each document is initially scanned electronically and highlighted according to pre-defined keyword and concept-phrase lists. Subsequent manual review is performed to determine document relevance and to enable concept mapping (Novak & Cañas, 2006). Selected references are coded as described in a conceptual analysis process, using an established set of key words and phrases in order to identify concepts articulated in the set of research questions (Busch, De Maret, Flynn, Kellum, Le, Meyers, Saunders & White, 2005).

Four primary thematic content sections are established for the organization and presentation of the results of the coding and analysis process (Busch et al., 2005). The first two thematic sections are based on two sub-question topics: (a) information visualization interaction, navigation, and exploration and (b) information visualization discovery and practices. A third thematic content section is related to examples of visualization techniques with mention of visualization creation software as a way to provide a practical, applied component. A fourth thematic content section focuses on qualitative visualization evaluation criteria in support of decision-making.

Definitions

The following list of definitions is provided to ensure that technical terms and associated concepts related to information visualization techniques and technologies are understood as these are used within the context of the Annotated Bibliography. Definitions are secured from peer-reviewed literature selected for use within the Annotated Bibliography, as well as domain experts located on the World Wide Web. Figure 1 provides a definition taxonomy (organized view of terms based on central concept) of all definitions listed in this paper.

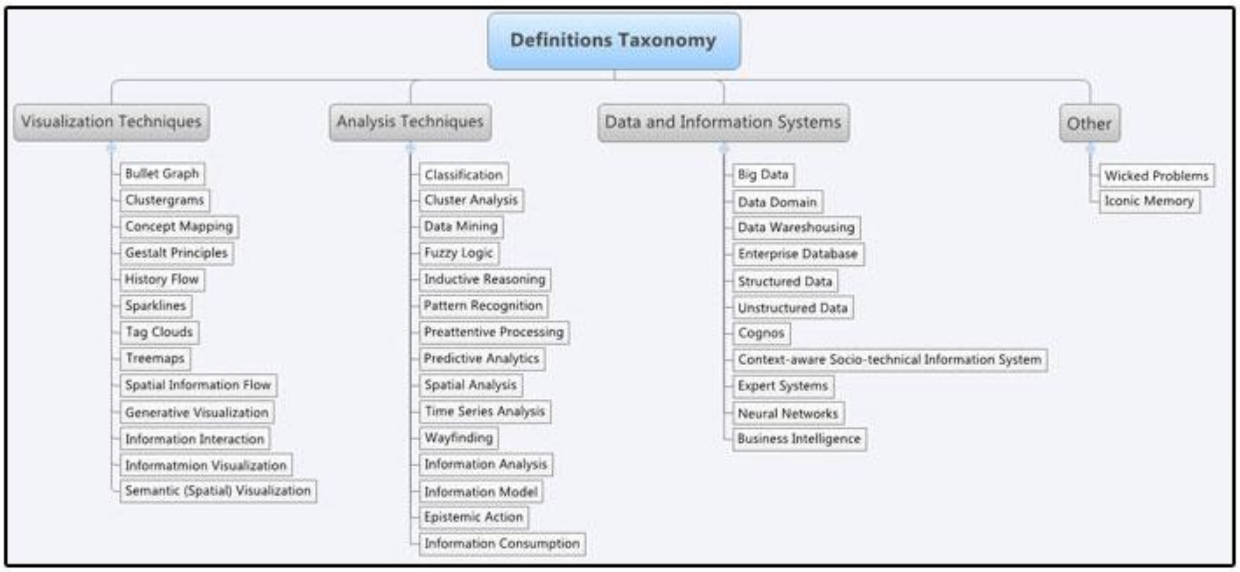

***Figure 1*.** **Definitions taxonomy**.

Big Data – Refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze (Manyika et al., 2011). This definition is intentionally subjective and incorporates a moving definition of how big a dataset needs to be in order to be considered big data – i.e., we assume that, as technology advances over time, the size of datasets that qualify as big data will also increase (Manyika et al., 2011, p. 11).

Bullet Graph – A visual data representation that displays both the actual, and target values, of a performance measure along a qualitative range (Ahokas, 2008, p. 51).

Business Intelligence – Business intelligence (BI) is the process of gathering correct information in the correct format at the correct time; and delivering the results for decision-making purposes, having a positive impact on business operations, tactics, and strategy in the enterprise (Zeng, Li, & Duan, 2012, p. 297).

Classification - A set of data analysis techniques used to identify the categories in which new data points belong, based on a training set containing data points that have already been categorized (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Byers, 2011, p. 28).

Cluster Analysis – A statistical method for classifying objects that splits a diverse group into smaller groups of similar objects, whose characteristics of similarity are not known in advance (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Byers, 2011, p. 28).

Clustergrams – A visualization technique used for cluster analysis displaying how individual members of a data set are assigned to clusters as the number of clusters increases (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Byers, 2011, p. 34).

Cognos – Business intelligence software that provides reports, analysis, dashboards and scorecards to help support the way people think and work when they are trying to understand business performance (Cognos Business Intelligence, n.d.).

Concept Mapping – A graphical tool for organizing and representing knowledge that includes concepts, usually enclosed in circles or boxes of some type, with relationships between concepts indicated by a connecting line linking two concepts (Novak & Cañas, 2006).

Context-aware Socio-technical Information Systems – An emerging human-centered computational environment that possesses some awareness of the tasks users are performing and some understanding of the knowledge background of individual users as well as their location in the world (Fischer, 2012, p. 287).

Data Domain – Refers to the range of acceptable values which a particular row or field in a database can contain (Wisegeek, 2012).

Data Mining – Business intelligence technology that enables in-depth analysis of data including the ability to build predictive models (Chaudhuri, Dayal, & Narasayya, 2011, p. 97).

Data Warehousing – A data warehouse is a subject-oriented, integrated, time-varying, non-volatile collection of data that is used primarily in organizational decision making (Inmon, 1992).

Enterprise Database – An organized / structured collection of data based on a representation of the structure, activities, processes, information, resources, people, behavior, goals, and constraints of a business, government, or other enterprise (Fox & Gruninger, 1998, p. 109).

Epistemic Action - Such an action is defined as an activity intended to uncover new information (Ware, 2012).

Expert Systems – An expert system is computer software that applies human expertise to solve problems, and successful applications result in numerous, diverse benefits (Duchessi & O’Keefe, 1995).

Generative Visualization – The automatic generation of a visualization based on user-specific requirements (Muller, Kovacs, Schilbach, & Eisenecker, 2011, p. 47).

Gestalt Principles – A set of principles that describe the way in which different objects are grouped together by the human visual system (Ahokas, 2008, p. 8).

Fuzzy Logic – The development of computational methods that can perform reasoning and problem solving tasks that require human intelligence (Yen, 1998, p. 1).

History Flow – A visualization technique that charts the evolution of a document as it is edited by multiple contributing authors (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Byers, 2011, p. 35).

Iconic Memory – Very short-term (less than 1 second) memory buffers that store information about the position, shape, color, and texture of objects in the visual field from one fixation to the next (Ware, 2012).

Inductive Reasoning – Entails using existing knowledge or observations to make predictions about novel cases (Hayes, Heit, & Swendsen, 2010, p. 1).

Information Analysis – The science of evaluating information content, and refining information to build informed portfolios (Grinold & Kahn, 1992, p. 1).

Information Consumption – Data, in the form of signals intended to covey meaning, received by humans through the senses of sight and hearing (Bohn & Short, 2010, p. 1).

Information Interaction – The ability of an information visualization to be engaged in the sense that the information seeker can: (a) display more data as needed, (b) hide the visualization when not needed, and (c) interface with the visualization by means of commands to help in the thinking process (Ware, 2012, p. 345).

Information Visualization – [#1] Process of constructing a visual presentation of abstract quantitative data (Ahokas, 2008). The visual presentation reveals hidden patterns to domain experts by presenting data based on computational intelligence methods such as data mining, machine learning, inductive reasoning strategies, expert systems, fuzzy logic and neural networks (Kovalerchuk, 2003, p. 1).

[#2] An inductive method in sense making that is focused on generating new insight and ideas – the seeds of theories – using human perception as a very fast filter (Fekete, Vanwijk, Stasko, & North, 2008).

Information Model – A representation of concepts, relationships, constraints, rules, and operations to specify data semantics (language) for a chosen domain of discourse (Lee, 1999).

Neural Networks – A technological method of incorporating, processing, and formalizing qualitative knowledge indicative of human intelligence by way of recognizing patterns, using language, learning, forgetting, generalizing, visualizing, observing, moving, and so on (Zahedi, 1991, p. 25-28).

Pattern Recognition – A set of machine learning techniques that assign some sort of output value (or label) to a given input value (or instance) according to a specific algorithm. Classification techniques are an example categorized (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Byers, 2011, p. 30).

Preattentive Processing – The early stage of visual perception that rapidly occurs below the level of consciousness which is tuned to detect a specific set of visual attributes (Few, 2006, p. 81).

Preattentive Attributes – The four categories of preattentive processing: (a) color, (b) form, (c) spatial position, and (d) motion (Ware, 2012).

Predictive Analytics - The process of making predictions based on the analysis of structured data typically stored in computerized databases (Elkan, 2012, p. 5).

Semantic (Spacialized) Visualization – The process of visually representing meaning within a specific spatial frame of reference (information domain) (Fabrikant & Buttenfield, 2001).

Sparklines – A compressed visualization presenting data-intense, word-sized graphics that are usually embedded in a context of words, numbers, and images (Ahokas, 2008, p. 49).

Spatial Analysis – A set of data analysis techniques, some applied from statistics, which analyze the topological, geometric, or geographic properties encoded in a data set (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Byers, 2011, p. 30).

Spatial Information Flow – A visualization technique that depicts spatial information flow between locations (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Byers, 2011, p. 36).

Structured Data – Collections of inherently related data stored within relational databases based upon carefully planned information modeling (Lu, He, Zhao, Meng, & Yu, 2007, p. 376).

Tag Clouds – A visualization of text in which words that appear most frequently are larger than words that appear less frequently (smaller) (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Byers, 2011, p. 34).

Time Series Analysis – A set of data analysis techniques from both statistics and signal processing for analyzing sequences of data points, representing values at successive times, to extract meaningful characteristics. Often used to predict future values of a time series based on past values (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Byers, 2011, p. 31).

Treemaps – A space-constrained visualization of hierarchical structures enabling users to compare nodes and sub-trees at varying depths while helping them spot patterns and exceptions (Schneiderman, 1998).

Unstructured Data – Collections of non-inherently related data, not stored within relational database systems such as video, sound, document and image files (Smullen, Tarapore, & Grurumurthi, 2007, p. 1).

Wayfinding - The organization and communication of our dynamic relationship to space and the environment (Center for Inclusive Design and Environmental Access, n.d.).

Wicked Problems – Problems characterized by large and uncertain data and resources that present potentially profound consequences (Arias-Hernandez, Dill, Fisher, & Green, 2011).

Research Parameters

Researchers and practitioners in the field of information visualization have long identified the need to evaluate visual data representations, interaction techniques, and visualization systems (Lam, Bertini, Isenberg, & Plaisant, 2011). Studying information visualization implementations is difficult due to factors such as human experience and analysis skills, as well as sheer data volume and its presentation influence, within a given visualization scenario or domain (Isenberg, Zuk, Collins, & Carpendale, 2008). This section details the research process used to locate, review, and organize this Annotated Bibliography.

Search Strategy

The search strategy used to locate references for use in this annotated bibliography is based on two primary methods: (a) manual search within established research databases for defined topic documents based on main research question, and (b) subsequent software-assisted regular expression (regex) document scanning for existing and additional keywords / topics in accordance with evolving sub-questions (Goyvaerts, 2012). Software supporting regular expressions (pattern recognition) is used to scan multiple documents simultaneously, highlighting embedded (a) keywords and (b) potential topic phrases and sentences for additional search. Such searches provide occurrence counts of keywords throughout all literature in seconds, assisting in quickly determining potential relevancy. General Internet searches broaden business and technology concepts useful in additional literature searches. Searching for keywords / topics within and between research documents provides metrics useful in quantifying and qualifying materials.

Search terms. All database keyword searches are done in the University of Oregon libraries. The primary databases used are Academic Search Premier, AcademicIndex.net, IEEE

Computer Society, JSTOR, Web of Science, and CiteSeerX. Preliminary reference consideration is based on a set of criteria that includes (a) year of publication; (b) topic relevance related to primary theme, breadth of topic information; and (c) breadth of linkage to relevant, secondary / associated topics (Creswell, 2009). A growing list of top-level key terms is based on initial readings of peer-reviewed material, along with specialized Internet sites devoted to subject matter by practicing professionals:

- Information visualization
- Information visualization techniques
- Information design
- Visualization technologies
- Structured data
- Unstructured data
- Big data

Documentation Approach

The overall documentation approach is based on progressive literature keyword searches and document collection creation. UO library search results based on primary keywords are first scanned based on title and abstract to determine if literature focus seems relevant to research focus. Potential documents and associated American Psychological Association (APA) citations are saved and labeled within the university system according to keyword(s) used to generate overall result list. All electronic documents are downloaded in original document format to local file folder scheme based on (a) research concept and (b) primary and secondary research questions. The folder hierarchy is intentionally kept very shallow and moves from concept to question to facilitate logical, as opposed to theoretical, lookup of documents when necessary.

Local document copies are imported into Evernote (Pachikov, 2012) providing (a) visualized presentation of documents (b) rapid access to documents without file system browsing, and (c) advanced tagging capabilities with a controlled vocabulary (predetermined list of approved terms for labeling) (Moreville & Rosenfeld, 2006). Once established, tagged documents are easily searched and linked based on multiple concepts, providing a more detailed visualization of associated concepts.

Literature Evaluation Criteria

Levy and Ellis (2006) assert that an effective literature review should (a) methodically analyze and synthesize quality literature, (b) provide a firm foundation to a research topic, (c) provide a firm foundation to the selection of research methodology, and (d) demonstrate that the proposed research contributes something new to the overall body of knowledge or advances the research field's knowledge-base.

Relevance and credibility. Literature contributing to this study is reviewed and included based on a subset of qualitative evaluation considerations, as noted by Meyrick (2006):

- Time frame of studies.
- Epistemological or theoretical topic stance.
- Clear research objectives communicated.
- Research / experimental methods appropriate to central question.
- Level of research / experimental detail indicative of current topic breadth.

- Data collection representative of sufficient sample or group sizes and populations.
- Sampling detail described and representative of current topic themes and trends.
- Clear and understandable transition from data analysis to conclusions.
- Systematic data analysis transparently outlined and described.
- Internal and external validations of conclusions.
- Presence in credible database that places high value on peer-reviewed literature.
- Breadth of citations given to industry experts across multiple publications.
- Accuracy and consistency of material with similar sources.
- Professional affiliations of authors during literature creation.

Reading and Organization Plan

Detailed conceptual analysis is performed on all literature included in this Annotated Bibliography. To ensure all material is systematically and thoroughly reviewed, coded, analyzed and organized, a standardized, detailed plan is outlined.

Reading plan. Based on a process described by Busch et al. (2012), this Annotated Bibliography utilizes a conceptual analysis approach. Keyword and concept phrases are identified and grouped, focusing on two broad categories: (a) visualization techniques that support information exploration and discovery within big data systems, and (b) visualization technologies that support information exploration and discovery within big data systems. All concepts are coded on a fixed list of keywords and phrases related to visualization concepts, techniques and technologies in accordance with the research sub-questions designed for the study. Coding flexibility is allowed in order to ensure unpredicted/unknown concepts are captured. Lists are based on controlled vocabularies and synonym rings, ensuring that coding is standardized, yet flexible enough to establish cross-references (Moreville & Rosenfeld, 2006).

Concepts are coded based on existence and frequency in order to gauge overall concept focus and determine meaning. Identification of the existence and frequency of distinct and relevant concepts provides for document-based treemaps (visual grid denoting concept frequency based on geometric proportions) diagraming the conceptual weight of all keywords and concepts within the given document (Lais, 2001). Both automatic and manual coding is performed. Automated scanning is done by computer software (EditPad) that (a) requires manual input of keywords and phrases and (b) highlights the existence of such entries. Manual analysis is done based off automated scanning in order to review the context of identified keywords and concepts. Different colors are used to highlight differing concepts. The list of keywords used for coding and treemap generation includes:

- Information Visualization
- Infoviz (popular shorthand term for information visualization)
- Information
- Visualization
- Big Data
- Exploration
- Interaction
- Discovery
- Analysis
- Decision Making

- Techniques

Analysis of literature is designed to identify the breadth, depth and applicability of visualization techniques and technologies in direct relation to big data. Information is discarded that (a) does not focus on large data sets, and (b) is too specific in terms of information domain. Once segmented, deep reading (analysis) reveals if keyword and phrase locations within the document correlate with detailed content that supports addressing the research sub-questions.

Organization plan. A thematic approach (Busch et al., 2012) is used to present information within this Annotated Bibliography in accordance with the two sub-questions: (a) how can information visualization techniques and supporting display technologies support information exploration, and (b) how can information visualization techniques and supporting display technologies support information discovery (Busch et al., 2012)? Reference organization is based on four primary themes; a simple concept map is developed and presented in the Conclusion, to characterize key cross-links (i.e., relationships between concepts), as described by Novak and Cañas (2006). Two themes, presented by Few (2009), support each research sub-question by focusing on (a) information visualization interaction, navigation, and exploration; and (b) information visualization discovery and practices. A third theme is related to information visualization design and implementation. A fourth theme is related to information visualization effectiveness evaluation and decision support considerations. Anticipated detailed topics to be discussed within these themes include: (a) the challenges of information visualization with respect to rapidly growing information repositories (Plaisant, 2004), (b) classifications of information visualization rendering techniques (Keim, 2002), (c) information visualization interaction techniques (Yi, Kang, Stasko, & Jacko, 2007), (d) the detection of subtle trends and random anomalies noted by Haroz and Whitney(2012), (e) preattentive attributes (visual cues) contributing to rapid information visualization interpretation (Few, 2009), and (f) evaluation criteria used to determine the quality and applicability of information visualizations within a specific information environment (Isenberg, Zuk, Collins, & Carpendale, 2008).

Annotated Bibliography

Peer-reviewed journal articles and conference proceedings provide the majority of literature selected to support this Annotated Bibliography. A total of 34 key references are read and analyzed prior to inclusion, ensuring content that represents (a) current research, (b) prevalent techniques, and (c) supporting technologies in the field of information visualization, as described in the set of delimitations for this study. References are organized according to four themes: (a) information visualization interaction, navigation, and exploration, which examines manual interfacing with visualization solutions; (b) information visualization discovery and practices, which examines the generation of insight based on visualizations; (c) information visualization design and implementation, which examines techniques used in creating effective visualizations; and (d) visualization implementation effectiveness evaluation and decision-making, which examines visualization solution effectiveness in supporting decision-making processes. All ideas, suggestions, insights, and recommendations noted in summary paragraphs for each bibliographic entry are those of the cited author(s).

Many of the references in this set are downloaded from CiteSeerX, an online scientific digital library focusing primarily on computer and information science literature. Articles available in the database are peer-reviewed publications. The database strives to remain compliant with the Open Archives Initiative Protocol for Metadata Harvesting, a standard proposed by The Open Archive Initiative for content dissemination. When available, the original source for the article is noted. To serve as a visual reference, concept map node groupings in Figure 2 are colored based on each primary theme denoting concepts addressed in each annotation. Each concept map node represents a specific information visualization concept that plays an integral role in establishing effective visualization solutions.

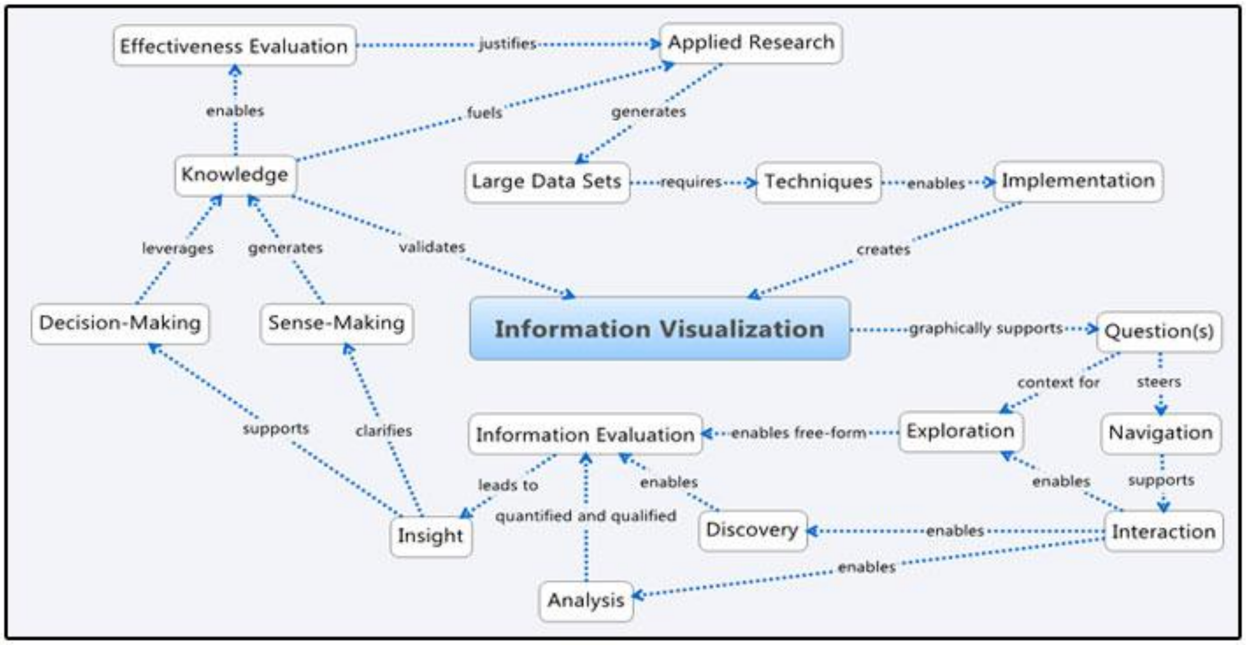

Figure 2. **Information visualization concept map.**

Theme 1: Information Visualization Interaction, Navigation, and Exploration

References within this theme focus on concept nodes regarding (a) question(s), (b) navigation, (c) interaction, and (d) exploration. These nodes are highlighted in green (see Figure 3) within the larger concept map.

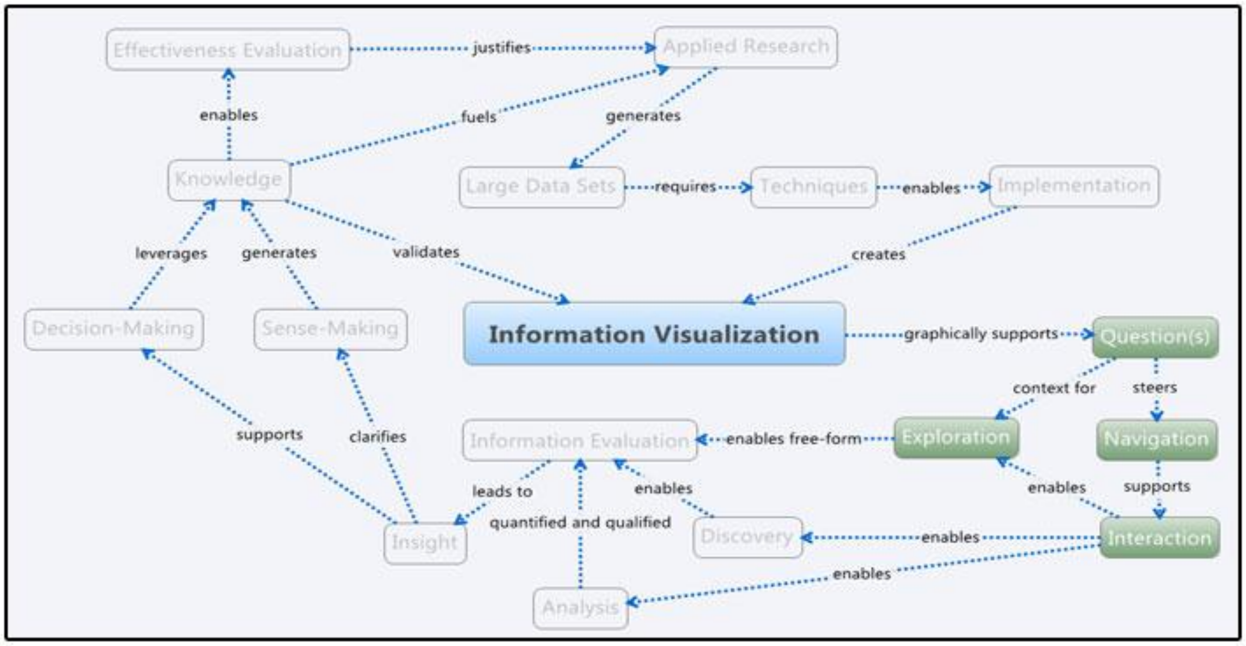

Figure 3. **Visualization interaction, navigation, and exploration concepts.**

Arias-Hernandez, R., Dill, J., Fisher, B., & Green, T. (2011). Visual analytics and human-computer interaction. *Interactions, 18*(1), 51-55.

Abstract. This article discusses the ways in which visual analytics research contributes to ongoing efforts in human-computer interaction that address cognitive task performance and how it is affected by highly interactive “human-information discourse” with visualization of data, information, and knowledge. Our conclusion is that the challenges posed by visual analytics will require HCI practitioners and researchers to expand their collaboration with cognitive scientists and visualization and computation researchers.

Summary. This article focuses primarily on the challenges faced by practitioner’s coupling human-computer interaction principles with visual analytics practices within “wicked problem” domains. Challenges noted by the authors in “wicked problems” include (a) multiple decision makers, (b) highly coordinated activities, (c) cognitive complexity, (d) broad range of technological applications, (e) large data sets, (f) graphical validity, (g) systems thinking, (h) situational constraints, and (i) individual differences.

The authors stress that visual analytics solutions must address tasks that are ill-defined, focusing on detecting the expected and discovering the unexpected. It is stressed that technology alone cannot support rapid decision-making in complex environments and that human insight must be incorporated. The authors conclude that visual analytical challenges will require human-computer interaction practitioners to expand collaboration efforts with cognitive scientists, visualization and computational researchers in order to enable rapid decision-making amid massive data sets.

Credibility. *Interactions* magazine is published bi-monthly by the Association for Computing Machinery (ACM), the largest educational and scientific computing society in the world. Richard Arias-Hernandez holds a PhD in science and technology studies from Rensselaer Polytechnic Institute and is a postdoc fellow at the SCIENCE lab, School of Interactive Arts and Technology, Simon Fraser University. John Dill holds a PhD in philosophy from Caltech and is a professor Emeritus in the School of Interactive Art and Technology at Simon Fraser University. Brian Fisher is an associate professor of Interactive Arts and Technology and Cognitive Science at Simon Fraser University and associate director of the Media and Graphics Interdisciplinary Centre at University of British Columbia. Tera Marie is a PhD student and graduate research assistant at the School of Interactive Arts and Technology, Simon Fraser University. Her research focuses on applied cognitive science for visual interfaces with an emphasis on the impact of individual differences on cognition and development of cognitive models for mixed initiative interfaces. The article contains three endnote references to three publications: (a) IEEE Press, (b) Information Visualization, and (c) Topics in Cognitive Science.

Bresciani, S., & Eppler, M. (2009). The benefits of synchronous collaborative information visualization: Evidence from an experimental evaluation. Retrieved November 22, 2012 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.164.1010.

Abstract. A great corpus of studies reports empirical evidence of how information visualization supports comprehension and analysis of data. The benefits of visualization for synchronous group knowledge work, however, have not been addressed extensively. Anecdotal evidence and use cases illustrate the benefits of synchronous collaborative information visualization, but very few empirical studies have rigorously examined the impact of visualization on group knowledge work. This experiment is designed to analyze the impact of visualization on knowledge sharing in situated work groups. The experimental study consists of evaluating the performance of 131 subjects (all experienced managers) in groups of five (for a total of 26 groups), working together on a real-life knowledge-sharing task. Comparisons are done concerning the control condition (no visualization provided), with two visualization supports: (a) optimal and (b) suboptimal visualization (based on a previous survey). The facilitator of each group was asked to populate the provided interactive visual template with insights from the group, and to organize the contributions according to the group consensus. Evaluations of the results are done via both objective and subjective measures. Statistical analysis clearly shows that interactive visualization has a statistically significant, objective and positive impact on the outcomes of knowledge sharing, but that the subjects seem not to be aware of this. In particular, groups supported by visualization achieved higher productivity, higher quality of outcome and greater knowledge gains. No statistically significant results

could be found between an optimal and a suboptimal visualization though (as classified by the pre-experiment survey).

Summary. This experimental study focuses on the use of interactive visualization templates to support group-based collaboration and decision-making processes in a strategic business setting. The main question centers on how real-time, interactive data visualization enables participant contributions in knowledge sharing teams. Co-located groups are studied, with three primary sub-groups providing variant forms of information visualization support: (a) optimal visualization support, (b) suboptimal visualization support, and (c) no visualization support. The overall task for each group is to share and document knowledge and experiences concerning business strategy implementation problems. The authors hypothesize that any form of interactive information visualization will increase (a) productivity, (b) outcome quality, (c) learning, (d) satisfaction, and (e) participation. Study results show that information visualization leads to objectively better collaboration results. The authors conclude that this study is relevant in that it provides initial evidence suggesting that interactive visualization support solutions, compared to unsupported, leads to statistically significant performance / collaboration enhancement.

Credibility. Sabrina Bresciani is a PhD who, at the time of publication, was associated with the University of Lugano. She is currently an assistant professor at the University of St. Gallen, Switzerland, adjunct professor at Franklin College Switzerland and visiting professor at Universidad del Pacifico in Peru. Dr. Bresciani has completed (a) 5 journal articles, (b) 4 book chapters, (c) 25 conferences, (d) 1 book review, (e) 2 case studies, (f) 1 working paper, and (g) 4 other publications. Martin Eppler is a PhD professor in the School of Management, University of St. Gallen, Switzerland. He has 179 publications including (a) journal articles, (b) books, (c) book chapters, and (d) conference papers. There are a total of 49 references supporting the article comprised primarily of papers focused on information visualization and information management. CiteSeerX is an online scientific digital library focusing primarily on computer and information science literature. Articles available in the database are peer-reviewed publications. The database strives to remain compliant with the Open Archives Initiative Protocol for Metadata Harvesting, a standard proposed by The Open Archive Initiative for content dissemination.

Fisher, D., DeLine, R., Czerwinski, M., & Drucker, S. (2012). Interactions with big data analytics. Retrieved November 21, 2012 from http://dl.acm.org/citation.cfm?id=2168943

Abstract. Here we report on the state of the practice of big data analytics, based on a series of interviews we conducted with 16 analysts. While the problems uncovered are pain points for big data analysts (including HCI practitioners), the opportunity for better user experience around each of these areas is vast. It is our hope that HCI researchers will not only turn their attention toward designs that improve the big data research experience, but that they will also cautiously embrace the big data available to them as a converging line of evidence in their iterative design work. The big data user experience challenge will affect every one of us.

Summary. This article discusses big data (extremely large data sets) and the challenges of (a) creating optimal user interaction experiences and (b) deriving meaning from data, when it comes to leveraging such repositories. Big data is defined as data that cannot be handled and processed in an easy manner. The authors focus on the state of the practice of big data analytics and the myriad of opportunities that exist for data analysts/consumers. Two primary types of big data analysts/consumers are identified: (a) corporate analytics teams and (b) academic analytics teams. Numerous processes and goals of big data analytics are identified for each group: (a) exploratory and demand-driven approach, (b) clear communication in a succinct and actionable presentation form, (c) the production of high confidence under tight deadlines, and (d) the need to preserve institutional memory. To accomplish these goals, definitive analyst workflows are defined: (a) acquiring data, (b) choosing an architecture, (c) shaping the data to the architecture, (d) writing and editing analysis code, and (e) reflecting and iterating over the results. The authors stress that future analysis solutions must be interactive, allowing users to iteratively pose questions and see rapid responses in a visual manner. It is communicated that big data analytics cannot be ignored and that these analytics present new challenges when interacting with computing technology in order to derive meaning from increasing levels of streaming data.

Credibility. Each contributing author is employed by Microsoft. Danyel Fisher is a researcher working on information visualization. His focus is on novel ways to explore and interact with data. Robert DeLine is a principal researcher working at the intersection of software engineering and human-computer interaction. He leads a group responsible for designing development tools in a user-centered fashion, based on cases studies and application prototyping. Mary Czerwinski is a research manager in the Visualization and Interaction for Business and Entertainment group.   Her research focuses on novel information visualization and interaction techniques. Steven M. Drucker is a principal researcher and manager of the VUE group, focusing on human-computer interaction when dealing with large amounts of information. He is also an affiliate professor at the University of Washington Computer Science and Engineering Department. The authors cite a total of 12 references focusing primarily on data management and information visualization or large data sets. The ACM digital library represents the world's largest educational and scientific computing society, housing peer reviewed journals, magazines, conferences, workshops, and electronic forums.

Mouine, M., & Lapalme, G. (2012). Using clustering to personalize visualization (Ed). *2012 16th International Conference on Information Visualization* (pp. 258).n.d.: IEEE. doi: 10.1109/IV.2012.51.

Abstract. The goal of our work is to propose models or methods to personalize the visualization of a large amount of weather information in a simple way and to make sure that a user can analyze all needed information. The visualization is prepared for each user according to an automatically detected profile based on clustering. Clustering is used to group users who share similar information visualization analysis preferences, and then set the visualization variables according to the visualization preferences of those users.

Summary. This article discusses the implementation of on-demand, interactive information visualizations based on large data sets. Predetermined user profiles are used to provide automated generation of information visualizations. The overall goal is to create an information product that enables rapid cognition without manual scanning of detailed data sets. The authors note that interaction plays a key role, allowing the user to iterate over the data, changing views based on objective exploration. Clustering (associating similar user profiles into groups) is done in order to personalize visualizations according to similar user profiles and preferences. Information visualization interactivity sessions are used to collect user preference data, subsequently leveraged to train the system to predict visualization rendering preferences for a given reporting need. The hope of this approach is to propose models or methods where large amounts of information that change continuously through time can be visualized in a rapid manner based on unique user preferences.

Credibility. Both Mohamed Mouine and Guy Lapalme were associated with the University of Montreal at the time of article publication. The Institute of Electrical and Electronics Engineers (IEEE) digital library houses peer- reviewed publications, conferences, and technology standards. The organization's overriding mission is to help set standards dedicated to advancing technological innovation and excellence. The authors reference seven publications in support of this paper. The references focus on (a) information visualization, (b) data mining, and (c) clustering algorithms.

Offenhuber, D. (2010). Visual anecdote. *Leonardo, 43*(4), 367-374.

Abstract. The discourse on information visualization often remains limited to the exploratory function – its potential for discovering patterns in the data. However, visual representations also have a rhetorical function: they demonstrate, persuade, and facilitate communication. In observing how visualization is used in presentations and discussions, I often notice the use of what could be called "visual anecdotes." Small narratives are tied to individual data points in the visualization, giving human context to the data and rooting the abstract representation in personal experience. This paper argues that these narratives are more than just illustrations of the dataset; they constitute a central epistemological element of the visualization. By considering these narrative elements as parts of the visualization, its design and knowledge organization appear in a different light. This paper investigates how the "story" of data representation is delivered. By means of ethnographic interviews and observations, the author highlights the different aspects of the visual anecdote, a specific point where the exploratory and the rhetorical functions of visualization meet.

Summary. This article examines how information visualization fulfills both an exploratory and rhetorical role in information when communicating meaning through anecdotes. The exploratory role refers to patterns and structures hidden in visualizations that must be discovered via interactive techniques. The rhetorical role also refers to the mobility and immutability of the language used to enable discovery. More specifically, the author states that a rhetorical role focuses on the social narrative of visualizations located outside the inherent data itself. The visualization should support a social narrative that is understood by everyone interacting with the visualization. Small narratives are tied to individual data points in the visualization, giving human context to the data and rooting the abstract representation in personal experience. In essence, visual anecdotes represent accounts of single rhetorical incidents, which are short, seductive, humorous, and sometimes invented. The author concludes that anecdotes even go beyond inherent rhetorical principles of visual communication by covering the non-visual and performance aspects of the presentation.

Credibility. Dietmar Offenhuber is a professor at the Art University Linz and Key Researcher at the Ludwig Boltzmann Institute for Media Art Research. From 2006 to 2008, he worked as a researcher at the MIT Media Lab. *Leonardo* is a professional and research journal providing an international publication channel for artists who use science and developing technologies in their work. *Leonardo* journals are published by The MIT Press, a university press that publishes titles in (a) the arts and humanities, (b) economics,

(c) international affairs, (d) history, (e) political science, and (f) science and technology. The author provides 17 references, focusing primarily on information visualization.

Yi, J., Kang, Y., Stasko, J. & Jacko, J. (2007). Toward a deeper understanding of the role of interaction in information visualization. Retrieved January 2, 2013 from http://www.cc.gatech.edu/~john.stasko/papers/infovis07-interaction.pdf

Abstract. Even though interaction is an important part of information visualization (Infovis), it has garnered a relatively low level of attention from the Infovis community. A few frameworks and taxonomies of Infovis interaction techniques exist, but typically focus on low-level operations and do not address the variety of benefits interaction provides. After conducting an extensive review of Infovis systems and their interactive capabilities, we propose seven general categories of interaction techniques widely used in Infovis: (a) Select, (b) Explore, (c) Reconfigure, (d) Encode, (e) Abstract/Elaborate, (f) Filter, and (g) Connect. These categories are organized around a user's intent while interacting with a system rather than the low-level interaction techniques provided by a system. The categories can act as a framework to help discuss and evaluate interaction techniques and hopefully lay an initial foundation toward a deeper understanding and a science of interaction.

Summary. The main goal of this article is to identify the fundamental ways that interaction is used in information visualization systems and the benefits it provides. The authors note that historically, visualization representation received the majority of research focus, ignoring the fact that representation and interaction are not mutually exclusive subjects. A call for the formal establishment of a "science of interaction" is mentioned, along with clearer definitions of what constitutes valuable information visualization interaction. For the purposes of this paper, visualization interaction is defined as features that provide users with the ability to directly or indirectly manipulate and interpret representations. Static visualizations are noted as having limited value as the number of visualized variables (data sets) grows. An extensive literature review of (a) visualization techniques and (b) visualization software was conducted reviewing a wide range of interaction techniques detailed by numerous publication authors. It was discovered that for different representation techniques, different interaction techniques are used to perform a similar task or achieve a similar goal. Seven categories of visualization interaction are identified and detailed: (a) selection, (b) exploration, (c) reconfiguration, (d) encoding, (e) abstraction/elaboration, (f) filtering, and (g) connection. The authors conclude that two main contributions to the general information visualization domain provide (a) importance of information visualization interaction techniques and inherent complexity and (b) categorization and detailed discussion of information visualization techniques.

Credibility. All four authors are affiliated with the Georgia Institute of Technology. Ji Soo Yi is with the Health Systems Institute & H. Milton Stewart School of Industrial and Systems Engineering. Youn ah Kang and John Stasko are with the School of Interactive Computing & GVU Center. Julie Jacko is with the Health Systems Institute, The Wallace H. Coulter Department of Biomedical Engineering, and College of Computing at Emory University as well. The article's research is supported by GVU Center Seed Grant, the National Science Foundation, and the National Visualization and Analytics Center. The authors cite 56 references for this paper focusing on (a) information visualization, (b) information exploration, and (c) information interfaces.

Theme 2: Information Visualization Discovery and Practices

References within this theme focus on concept nodes regarding (a) information evaluation, (b) discovery, (c) analysis, and (d) insight. These nodes are highlighted in orange (see Figure 4) within the larger concept map.

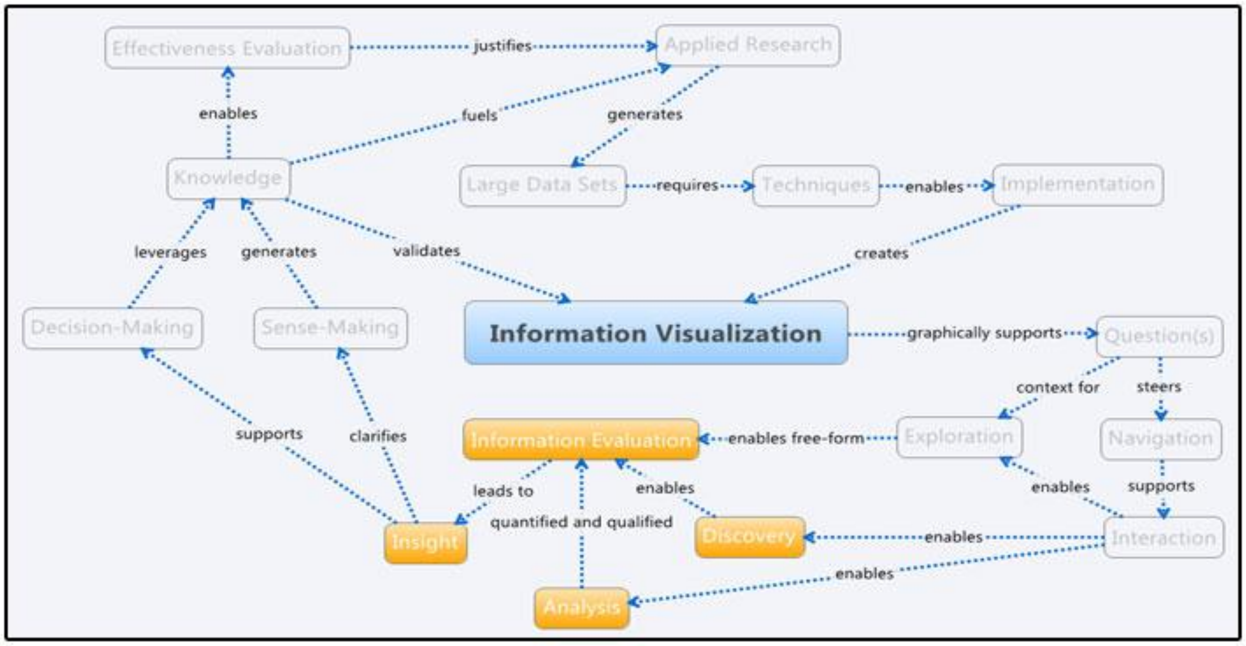

Figure 4. **Visualization discovery and practices concepts.**

Dadzie, A., Lanfranchi, V., & Petrelli, D. (2009). Seeing is believing: Linking data with knowledge. *Information Visualization*, *8*(3), 197-211.

Abstract. The analysis of data using a visual tool is rarely a task done in isolation; it tends to be part of a wider goal: that of making sense of the current situation, often to support decision-making. A user-centered approach is needed in order to properly design interaction that supports sense-making incorporating visual data analysis. This paper reports the experience gained in X-Medla, a project that aims to support knowledge management (KM), sharing and reuse across different media in large enterprises. A report of the user-centered design approach adopted and the design phases that led to the first prototype is provided. A user evaluation was conducted to assess the design and how different levels of data, information and knowledge were mapped using alternative visual tools. The results show that a clear separation of the visual data analysis from other sense-making sub-tasks helps users in focusing their attention. Users particularly appreciated the data analysis across different media and formats, as well as the support for contextualizing information within the broader perspective of KM. Further work is needed to develop more fully intuitive visualizations that exploit the richer information in multimedia documents and make the multiple connections between data more easily accessible.

Summary. The goal of this article is the discussion of user-centered design related to visualization software supporting sense-making activities. The authors support the concept of hierarchical layers of data, information, and knowledge. They postulate that each layer requires differing types of visualizations to communicate inherent meaning within a given cognitive step. Three primary challenges are identified for visualization tools supporting each layer: (a) supporting sense-making, (b) identifying and designing the most appropriate visualization for the type of task being performed in relation to the hierarchical layer, and (c) accommodating sense-making in complex situations involving teams of experts engaged in collective discussions that lead to final decisions. The authors argue that visualization is not an isolated activity and should be done within the context of human factors and knowledge management, putting the user in the driving seat in order to harness optimal interaction between the human and the machine to obtain truly effective analysis of large data sets. Considerations are identified that need to be made in order to create optimal visualizations allowing the extraction and sharing of knowledge within large data sets including (a) targeting the end user, (b) the expertise the end user brings in their domain, and (c) the resources available for collecting, processing and analyzing the data. It is noted that traditional iterative visualization software design does not accommodate design ideas directed by interaction with end users and results in implementations from software designers based at times on assumptions. They recommend cycles of visualization application design fueled by validation from end users based on functional prototypes. To support this form of iteration, the authors leverage specialized software that allows them to capture feedback, track activity and log interaction tendencies through what are termed interactive *knowledge objects.* Extensive data is collected resulting in alternative methods for representing knowledge visually: (a) employing semantic web technologies; (b) allowing users to control interactive visualizations, supporting exploratory and directed analysis of large, complex, data, and (c) using working prototypes that integrate differing and specialized technologies.

Credibility. All authors are affiliated with the University of Sheffield. Aba-Sah Dadzie and Vitaveska Lanfranchi are affiliated with the Department of Computer Science, while Deniela Petrelli is affiliated with the Department of Information Studies. *Information Visualizations* is a peer-reviewed journal that adheres to strict publication guidelines for both editors and reviewers. Strict ethical policies are enforced to ensure articles do not violate ethical publication standards. The authors provide a listing of 26 references used throughout the article. Reference topics include (a) information visualization, (b) knowledge management, (c) information analytics and analysis, (d) information interaction, and (e) human-computer interaction. Research funding is sponsored by the European Commission as part of the Information Society Technologies (IST) program.

Gordon-Murnane, L. (2012). Big data. *Online, 36*(5), 30-34.

Abstract. The article discusses the opportunities for librarians and information professionals of the projected big data generation between 2015 and 2016. A forecast from Cisco revealed that the annual global Internet Protocol (IP) traffic will grow to 1.3 zettabytes. The reasons for this data explosion are cited including the widespread accessibility and affordability of new digital devices that make access to the Internet easy. A definition of big data is also provided.

Summary. This article discusses opportunities for librarians and information professionals within the rapidly emerging profession of big data management. The author identifies multiple sources contributing to the rapid increase in data volumes including (a) email, (b) searching, (c) browsing, (d) blogging, (e) tweeting, (f) buying, (g) sharing, (h) texting, and (i) networked sensors. Additionally, the author notes that improved tools exist to (a) store, (b) aggregate, (c) combine, (d) analyze, and (e) extract new insight. The author states that organizations are becoming more pressured to leverage this data in order to (a) spur new innovations, (b) spur new product opportunities, (c) achieve cost savings and efficiencies, and (d) use predictive analytics to understand customer behavior. It is noted that ignoring big data management issues puts organizations at risk in competitive markets. The article identifies two primary reasons why librarians and information professionals are well situated to address the growing demands of effective big data management issues and strategy implementation: (a) they have the requisite skills, knowledge, and service mentality to help information consumers capitalize on big data assets, and (b) they understand how to exploit/leverage differing information types. In conclusion, it is recommended that librarians and information professionals embrace all opportunities with regards to big data management across business domains.

Credibility. Laura Gordon-Murnane is an information professional and freelance writer. *Online* is a magazine written for both novice and experienced information professionals and librarians in academic, corporate, government, research, and web information management. Publications focus on practical articles, product reviews, case studies, and informed opinions about selecting, using, manipulating, and managing digital information products. The author acknowledges 18 publications in the Recommended Reading section based on in-text references to industry experts and professionals. References focus on big data and data sciences.

Shrinivasan, Y., & Wijk, J. (2008). Supporting the analytical reasoning process in information visualization. Retrieved November 21, 2012 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.184.2649.

Abstract. This paper presents a new information visualization framework that supports the analytical reasoning process. It consists of three views: a data view, a knowledge view, and a navigation view. The data view offers interactive information visualization tools. The knowledge view enables the analyst to record analysis artifacts such as findings, hypotheses and so on. The navigation view provides an overview of the exploration process by capturing the visualization states automatically. An analysis artifact recorded in the knowledge view can be linked to a visualization state in the navigation view. The analyst can revisit a visualization state from both the navigation and knowledge views to review the analysis and reuse it to look for alternate views. The whole analysis process can be saved along with the synthesized information. A user study is presented and discussed, reviewing the perceived usefulness of a prototype based on this framework that we have developed.

Summary. The authors of this paper present an information visualization framework based on general analytical reasoning in combination with information visualization. Three forms of visualization are presented based on (a) data views, (b) knowledge views, and (c) navigation views. The authors note that analysis is often unsystematic, continuously emerging, and emergent. A major selling point of the framework is that analysts can save the analysis processes used to generate the view, along with the view, and provide the entire lifecycle product to the customer/consumer. In this sense, a visualization user must have an overview of not only what has been found, but what has been done to establish what is found. A formal use case was conducted, based on proprietary data analysis software, revealing that analysts benefited from the three primary views. It was found overall that analytical reasoning is enhanced by providing analysts (a) reasoning abilities facilitated by extending visualization support based on mental models that link analysis artifacts to the visualizations and (b) the ability to revisit the visualization states in order to review and validate the findings for reuse or alternate view generation.

Credibility. Yedendra B. Shrinivasan completed his PhD in visual analytics in the Department of Mathematics and Computer Science at the Technische Universiteit Eindhoven. His research interests include data visualization, visual analytics and human-computer interaction. Jack van Wijk is a full professor in visualization at the Department of Mathematics and Computer Science of Eindhoven University of Technology. He holds an MS degree in industrial design engineering and PhD in computer science.

CiteSeerX is an online scientific digital library focusing primarily on computer and information science literature. Articles available in the database are peer-reviewed publications. The database strives to remain compliant with the Open Archives Initiative Protocol for Metadata Harvesting, a standard proposed by The Open Archive Initiative for content dissemination. The research paper is supported by the VIEW program of the Netherlands Organization for Scientific Research (NWO) under research grant number 643.100.502. The authors list 37 references in the paper, focusing on (a) information interaction, (b) information visualization, (c) cognitive science, (d) human-computer interaction, (e) visual analytics, and (f) information exploration.

Theme 3: Information Visualization Design and Implementation

References within this theme focus on concept nodes regarding (a) applied research, (b) large data sets, (c) techniques, and (d) implementation. These nodes are highlighted in black (see Figure 5) within the larger concept map.

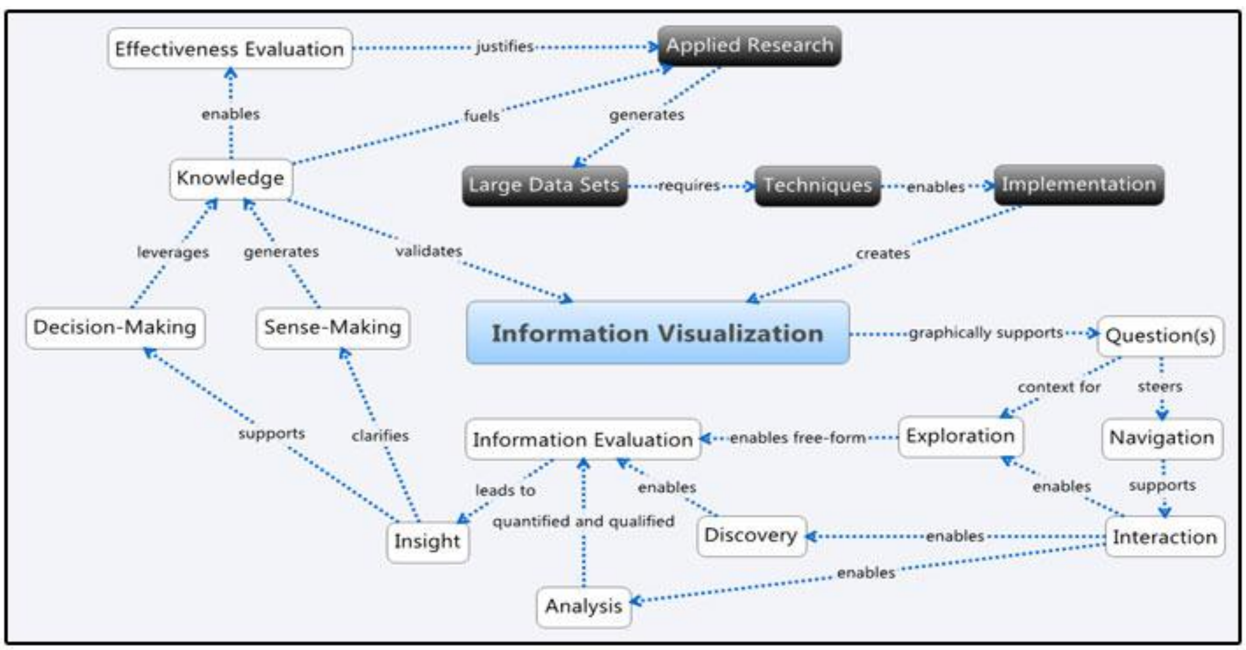

***Figure 5*. Visualization design and implementation concepts.**

Bermudez, J., Agutter, J., Foresti, S., Westenskow, D., & Syroid, N., et al. (2005). Between art, science and technology: Data representation architecture. *Leonardo, 38*(4), 280-285.

Abstract. As our civilization continues to dive deeper into the information age, making sense of complex data becomes critical. This work takes on this challenge by means of a novel method based on complete interdisciplinarity, design process and built-in evaluations. The result is the design, construction, testing and deployment of data environments supporting real-time decision-making in such diverse domains as anesthesiology and live art performance. Fundraising success, technology licensing, market implementation and many live art performances provide evidence of the great potential of committed interdisciplinary work for advancing science, art and technology while benefiting society at large.

Summary. This article discusses the importance of integrating art, science, and technology in order to rectify inadequacies in modern day data and information visualization practices. The authors clearly state that there is a direct correlation between how data is represented and the meaning that can be extracted from the visual representation. These inadequacies of information visualization techniques appear to be attributed to (a) the persistence of early 20th-century quantitative methods, (b) a naïve understanding of human cognition, and (c) the use of simplistic representation spaces that are inadequate to address the complexity of modern day information environments. A strong recommendation is made to couple art, science, and technology in order to push the horizons of each discipline while introducing artistic expertise into the practice of information visualization across industries. The way that data is presented has an overwhelming weight in how a system or situation is perceived and what ultimately drives decision-making processes. The authors define data representation architecture (information visualization) as the organizational, functional, experiential, and media-technological order defining the interaction between data representation and user. It is suggested that new visualization approaches are created that enable data-based decision-making with less cognitive effort and prerequisite training. "Selected depiction" is noted as a means to deliver visualizations designed for particular information consumption.

Credibility. All authors are affiliated with the University of Utah. Julio Bermudez and Jim Agutter represent the College of Architecture and Planning. Stefano Foresti represents the Center for High Performing Computing. Dwayne Westenskow and Noah Syroid represent the School of Medicine. Frank Drews represents the Department of Psychology, while Elizabeth Tashjian represents the School of Business. *Leonardo* journals are published by The MIT Press, a university press that publishes titles in (a) the arts and humanities, (b) economics, (c) international affairs, (d) history, (e) political science, and (f) science and technology. The authors reference 15 papers, focusing (a) cognition, (b) information architecture, and (c) graphic display design.

Bihanic, D., & Polacsek, T. (n.d.). Models for visualization of complex information systems. In The French National Network of Complex Systems (Ed). *2012 16th International Conference on Information Visualization* (pp. 130). n.d.: The French National Network of Complex Systems. doi: 10.1109/IV.2012.

Abstract. Modular approaches (objects, components and service-based) offer efficient solutions to face the complexity of Information Systems. However, today they involve major problems in dealing with data/information and decision-making (from the standpoint of the end- user). In this article, we aim to initiate a dialogue between digital/interaction design and model-driven engineering through the study of a new data representation-visualization approach based on generative visualization. Due to recent advances in graphical user interfaces and interaction techniques, it is expected that the world of modeling could benefit from creating views of existing models defined in a modeling language complying with Meta- Object Facility (like UML).

Summary. In this article the authors provide a detailed example of creating data models suitable for generating visual representations of data within complex information systems. The authors note that organizations capable of integrating complex system data are able to leverage up to 50 percent more of their big data for business intelligence and analytics. According to the authors, a complex system features some of the following properties: (a) heterogeneousness (composition of several entities or agents), (b) flow processing (relations or interactions between system elements), (c) size (processing capabilities based on system complexity), (d) hierarchical organization (network of interrelated system elements); and (e) evolution (feedback mechanisms, collective behavior, and emergent properties). The authors of this article consider current IS (information system) complexity when dealing with large, interrelated data repositories, and present a new approach to graphically visualizing large-scale information system data. It is stated that visualization "amplifies cognition" by expressing semantic connections between data (p. 131). The goal is to provide sufficient reasoning as to why customized data visualizations for specific tasks are required in order to provide quick access to relevant data within complex information systems. The authors state that the development of specific, customized views and viewpoints is critical to ensure optimal use of complex systems and the data they provide, subsequently amplifying cognition. It is noted that the last fifteen years has seen a convergence of new visual/graphical information representations, theories, and paradigms based on contributions from computer science, informatics engineering, cognitive science, and experimental cognitive psychology. They conclude that predefined information visualizations are not appropriate for interacting with complex system data and that semantic information needs to be added to data modeling efforts in order to generate contextual views of data.

Credibility. David Bihanic is Associate Professor at the University of Clermont-Ferrand. He specializes in information interaction design. He examines new paradigms of information visualization and manipulation of large, complex databases. He holds a PhD in interaction /GUI design. At the time of publication, he was affiliated with the CALHISTE Laboratory, University of Valenciennes and Hainaut-Cambresis. Thomas Polacsek is affiliated with the Information Processing and Modeling Department at the ONERA French Aerospace Lab in France. The article is supported by the French National Network of Complex Systems (RNSC) and the French Institute of Complex Systems of Ile-de-France(ISCPIF). The article lists 19 references focusing on (a) information visualization, (b) visualization taxonomies, (c) spatial memory, (d) software architecture, and (e) modeling languages.

Conti, G., Ahamad, M., & Stasko, J. (2005).). Attacking information visualization system usability overloading and deceiving the human. Retrieved December 12, 2012 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.111.772

Abstract. Information visualization is an effective way to easily comprehend large amounts of data. For such systems to be truly effective, the information visualization designer must be aware of the ways in which their system may be manipulated and protect their users from attack. In addition, users should be aware of potential attacks in order to minimize or negate their effect. These attacks target the information visualization system as well as the perceptual, cognitive and motor capabilities of human end users. To identify and help counter these attacks we present (a) a framework for information visualization system security analysis, (b) a taxonomy of visualization attacks, and (c) technology independent principles for countering malicious visualizations. These themes are illustrated with case studies and working examples from the network security visualization domain, but are widely applicable to virtually any information visualization system.

Summary. The goal of this article is to (a) identify potential attacks (intentional or unintentional) of human and computer information visualization capabilities that can negatively affect the ability to easily comprehend or handle large amounts of visual data and (b) develop principles to counter or mitigate potential attacks. The authors base the article on a comprehensive analysis of weaknesses regarding both visualization systems and their supporting data flow, while considering (a) data sources, (b) data communications, (c) data storage, (d) processing, (e) presentation, and (f) human interpretation. The authors note that research into the manipulation and deployment of human and computer information visualization systems is relatively uncommon when considering that such techniques as distraction, misinformation and disinformation are frequently encountered. Most essays on the topic focus on non-interactive information visualizations involving speaker to audience scenarios, indicating that attacks are coordinated by the creator of the visualization system, business presentation, advertisement, or statistical report, again either intentionally or unintentionally. The authors present (a) a framework for human information visualization system security analysis, (b) a taxonomy of visualization attacks, and (c) technology independent principles for countering malicious visualizations. In order to validate the article's contributions to the field of information visualization, they provide examples of visualization attacks analogous to those seen over computer networks with traditional intrusion detection systems, highlighting the possibility of targeting (a) human perception capabilities and (b) computer visualization system functionality. The authors argue that the only path to secure and reliable information visualizations is through understanding the types of threats and countermeasures currently available. They specifically mention performing a threat analysis focusing on (a) identifying assets you wish to protect, (b) brainstorming known threats to the system, (c) ranking the threats by severity, (d) choosing how to respond to threats, and (e) choosing techniques and technologies (if any) to mitigate the threats. It is described how attacks can be accomplished by diagraming and detailing various types of attacks focused on (a) human memory, (b) human attention, (c) human visual perception, (d) human motor resources, (e) specific human targets, (f) computer hardware and software, (g) computer processing software, (h) computer-generated visualizations, (i) computer data generation capabilities, (j) computer storage, and (k) human and computer communication channels. Specific types of attacks include (a) denial of information (DoI) , (b) "Cry Wolf", (c) displacement, (d) color mapping, (e) extreme information overload, (f) round-off, (g) jamming, (h) occlusion, (i) labeling, (j) auto scale, (k) sensor blindness, (l) selective sensor blindness, (m) spoofing source identity, (n) sampling rate adjustment, and (o) poisoned data insertion. It is concluded that the best form of protection from such attacks is user education including how to conduct a threat analysis, and the creation of visualization systems or techniques that are resilient to such attacks.

Credibility. At the time of publication, all authors were affiliated with the College of Computing at the Georgia Institute of Technology. Gregory Conti is Associate Professor in the Department of Electrical Engineering and Computer Science at West Point. He holds a PhD in computer science and has published numerous journal articles regarding a wide range of technological topics. Mustaque Ahamad is a PhD and professor in the College of Computing at the Georgia Institute of Technology. His research interests are in distributed operating systems, computer security, and fault-tolerant systems. John Stasko is Professor and Associate Chair in the School of Interactive Computing at the Georgia Institute of Technology. He holds a PhD in computer science from Brown University. He is a faculty investigator in the Department of Homeland Security's VACCINE Center of Excellence focusing on developing visual analytics technologies and solutions for grand challenge problems in homeland security, and in the NSF FODAVA Center exploring the foundations of data analysis and visual analytics. The article lists 40 references focusing on such topics as (a) data visualization, (b) countering information attacks, (c) scientific visualization, (d) human-computer interaction, (e) cognitive processing, and (f) network security. The paper was presented at the Symposium on Usable Privacy and Security (SOUPS) 2005, July 6-8, 2005, Pittsburgh, PA, USA. The symposium is designed to bring together an interdisciplinary group of researchers and practitioners in human computer interaction, security, and privacy. The program features technical papers, workshops and tutorials, a poster session, panels and invited talks.

Fekete, J., & Plaisant, C. (2002).). Interactive information visualization of a million items. Retrieved November 23, 2012 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.58.2019.

Abstract. Existing information visualization techniques are usually limited to the display of a few thousand items. This article describes new interactive techniques capable of handling a million items (effectively visible and manageable on screen). Available hardware-based techniques are evaluated with newer graphics cards, as well as new animation techniques and non-standard graphical features such as stereovision and overlap count. These techniques have been applied to two popular information visualizations: treemaps and scatter plot diagrams; but are generic enough to be applied to other 2D representations as well.

Summary. The goal of this article is to address the currently assumed limits of information visualization scalability (size of visible items) by representing a million items (object displayed as a distinguishable, contiguous area using a single visualization technique). The authors note this research is especially relevant as (a) data sets are growing exponentially across domains, (b) designers of visualizations are often required to produce displays on small devices such as smart phones and Personal Data Assistants (PDA), and (c) current visualization systems are limited to about 10,000 items on standard display devices utilizing non-optimized graphical rendering technology. The authors clarify the difference between information visualization and scientific visualization. It is pointed out that information visualization maps the attributes of an abstract data structure to visual attributes such as position, color, and size. Scientific visualization renders data that usually has some intrinsic value and representational characteristics. The authors also note that, as often as possible, such large scale visualization should rely on preattentive processing attributes associated with human recognition of graphical objects (Few, 2009). Modern mapping techniques exist for information visualizations that typically have a strong interaction component allowing users to rapidly explore data sets via (a) point-and-click, (b) drag-and-drop, (c) predefined filter lists, or (d) dynamic/real time queries. The techniques (treemaps and scatter plots) used to visualize a million items in the paper rely on (a) computer hardware acceleration (leveraging physical computer hardware in place of software to accomplish graphics processing), (b) non-standard visual attributes such as stereovision or synthetic overlap and animation, and (c) animation utilizing multiplexing in order to analyze the data across several views and mappings without losing context. The authors represent items on the screen with (a) saturated colors, (b) color intensity, (c) quadrilateral sizes within treemaps, and (d) positioning in scatter plots. In order to overcome the 10,000 item limit, the authors note specifically that (a) space multiplexing, (b) time multiplexing, (c) overlapping, or (d) space deformation techniques can be used. The authors provide explanations of each technique. In conclusion, the authors clearly prove that one million items can be displayed on a single visual screen. Techniques are detailed, noting that commonly available graphical acceleration hardware can be used. They note that more experiments are required in order to understand how humans can (a) actually cope and (b) generate insight cognitively with such large data representations.

Credibility. Jean-Daniel Fekete is a Senior Research Scientist and Director de Recherche 2e Classe at the University of Paris. He is the Scientific Leader of the INRIA Project Team AVIZ. He holds a PhD in computer science and has 137 publications. Catherine

Plaisant holds a PhD and is Associate Director of Research of the Human-Computer Interaction Lab at the University of Maryland Institute for Advanced Computer Studies. She has over 100 technical publications on subjects such as information visualization, digital libraries, universal access, image browsing, and evaluation methodologies. The article is supported in part by Chevron Texaco. There are 25 cited references focusing on issues such as (a) visual information seeking, (b) human factors in computing systems, (c) display of quantitative information graphics, and (d) the exploration of dynamic graphs.

Few, S. (2009). Information visualization. In N. Wishner (Eds.), *Now you see it: Simple visualization techniques for quantitative analysis*. Oakland, CA: Analytics Press.

Abstract. This book focuses on data exploration, leading to discovery, and data sense-making, leading to understanding. Compared to data presentation, data exploration and sense-making requires extensive interaction with data and a richer set of graphic displays. Numerous examples of data interaction techniques are presented along with detailed discussions related to visualization implementation and analysis.

Summary. This book is a comprehensive reference of techniques and tools useful in the exploration and analysis of information. There are three primary sections to the book: (a) building core skills for visual analysis, (b) honing skills for diverse types of visual analysis, and (c) further thoughts and hopes for information visualization. The book compares numerous types of visualizations, highlighting which sort of visualization works best for a given exploration, discovery, analysis, and presentation solution. The author notes that the solution to making sense from the increasing volumes of information in our lives does not need to be complicated.   It is stressed that human

interpretation is critical to information sense-making, and our eyes serve as the primary tool by which to accomplish such goals. The solutions presented do not rely on highly complicated or advanced solutions and are intentionally kept simple for any level of information exploration and analysis experience.

Credibility. Stephen Few is the founder and principal of Perceptual Edge, an organization that focuses on the tools and techniques of visual business intelligence. Few has worked for over 25 years as a teacher, writer, consultant, and innovator, primarily in the fields of business intelligence and information design. He publishes the monthly Visual Business Intelligence Newsletter and teaches in the MBA program at the University of California, Berkeley.

Fischer, G. (2012). Context-aware systems: The 'right' information, at the 'right' time, in the 'right' place, in the 'right' way, to the 'right' person. Retrieved November 12, 2012 from http://l3d.cs.colorado.edu/~gerhard/papers/2012/paper-AVI-context-aware.pdf.

Abstract. Based on the assumption that the scarce resource for many people in the world today is not information but human attention, the challenge for future human-centered computer systems is not to deliver more information "to anyone, at any time, and from anywhere," but to provide *"the 'right' information, at the 'right' time, in the 'right' place, in the 'right' way, to the 'right' person."* This article develops a *multidimensional framework* for *context aware systems* to address this challenge, transcending existing frameworks that limited their concerns to particular aspects of context-awareness and paid little attention to potential pitfalls. The framework is based on insights derived from the development and assessment of a variety of different systems that we have developed over the last twenty years to explore different dimensions of context awareness. Specific

challenges, guidelines, and *design trade-offs (promises* and *pitfalls)* are derived from the framework for designing the next generation of context-aware systems. These systems will support advanced interactions for assisting humans (individuals and groups) to become more knowledgeable, more productive, and more creative by emphasizing context awareness as a fundamental design requirement.

Summary. This article discusses the concept of *context-aware socio-technical information systems*. The general premise is that future information systems should have a working knowledge of individual user backgrounds as well as their location within the world in order to provide contextual support regarding information insight generation. The authors define the concept of context and discuss problems that context awareness can address. A context-aware socio-technical system is described as taking place in a certain context that refers to the physical and social situation in which computational devices and environments are embedded. The overall context is determined by (a) the people involved (including their background knowledge and their interactions); (b) the objective of the interaction (including the tasks to be carried out), and (c) the time and place where the interactions occur. The authors describe three primary aspects to socio-technical systems: (a) how the information is obtained, (b) how the information is represented, and (c) the objectives and purposes for which the context information is used. The problems addressed by these types of systems include (a) information overload and human attention, (b) differing aspects of context, (c) unarticulated design intent that can affect information interpretation, and (d) information access and delivery. The authors list and describe scenarios and examples of exceptional context-aware systems. Scenarios include (a) active help systems providing incremental learning of high functionality environments, (b) domain oriented design environments, (c) critiquing systems, and (d) cultures of participation (requiring co-design of systems). Additionally, the authors provide a multidimensional framework for context-aware systems, indicating how the different parts of the paper relate to each other. Important design considerations that could establish future promises or pitfalls for such systems are noted: (a) filter bubbles and groupthink, (b) making information relevant to the task at hand versus serendipity, (c) intrusiveness, (d) remembering and forgetting, and (e) privacy. The authors conclude by stressing that such systems will not be perfect and only partially aware of the total problem-solving processes a human user may require. Additionally, they note that the ultimate goal of human-centered computing is that it will serve the benefit of users (acting as individuals or in teams) by empowering them, improving their information interaction experience, and making them more productive and creative by integrating social and technical dimensions.

Credibility. Gerhard Fischer is the director of the Center for Lifelong Learning and Design, professor in the Department of Computer Science, and fellow of the Institute of Cognitive Science at the University of Colorado. His research interests include (a) lifelong learning, (b) design, (c) distributed intelligence, and (d) human-computer interaction. This article was obtained online, directly from the University of Colorado. There are 45 references listed in the article related to topics such as (a) information interaction, (b) context-aware applications, (c) human-computer interaction, and (d) ubiquitous computing.

Gaviria, A. (2008). When is information visualization art?: Determining the critical criteria. *Leonardo, 41*(5), 479-48.

Abstract. This paper initially examines the differences between functional and aesthetic forms of visualization for information visualization. The author then shows such a dual categorization to be ineffective as a critical scheme for evaluating artwork that utilizes comparable visualization techniques. Adopting Joline Blais and Jon Ippolito's classification of artistic production, the author argues for the use of "genre art" and "research art" as more suitable criteria for the analysis and assessment of such artwork.

Summary. Traditionally, information visualization products with clearly different representational attributes have been lumped together with artistic and academic analysis techniques. Because information visualizations strive to provide functional purpose, this paper discusses the differences between pure aesthetic visualizations indicative of art and representations that strive to provide clearly defined functions. The author makes a clear distinction between functional and aesthetic visualizations by stating that functional information visualizations aim to convey a message or delineate patterns hidden in the represented data through metaphors that users can quickly understand, while aesthetic information visualizations are more concerned with presenting a subjective impression of a data set by eliciting a visceral or emotive response from the user (p. 479). Where functional visualizations strive to rapidly communicate meaning and insight, aesthetic visualizations attempt to elicit instinctive and emotional responses centered on interest, attention, enjoyment, and curiosity. One very important assertion made by the author notes that interpretations of aesthetic visualizations are also biologically controlled across individuals, cultures, and time. The application of aesthetic symbols in functional visualization can have significantly different results on users based on these considerations. It is pointed out that knowledge gleaned from artistic visualization is

reached obliquely, through associations that may be elaborate and complex and not at first distinctly understood, but whose theoretical implications are, in some sense, significant to the related field or related topics and tasks.

Credibility. Andres Ramirez Gaviria is a conceptual artist who uses digital technology as his laboratory for the analysis of historical "truths" (Trantow & Schurl, n.d.). *Leonardo* is a professional and research journal providing an international publication channel for artists who use science and developing technologies in their work. *Leonardo* journals are published by The MIT Press, a university press that publishes titles in (a) the arts and humanities, (b) economics, (c) international affairs, (d) history, (e) political science, and (f) science and technology. There are 10 references listed focused on (a) information visualization, (b) aesthetics in usability, and (c) digital information graphics.

Keim, D. (2000). An introduction to information visualization techniques for exploring large databases. Retrieved November 12, 2012 from http://www.inf.ethz.ch/personal/peikert/SciVis/Literature/keim-Tutorial2000.pdf.

Abstract. These tutorial notes provide visual examples of information visualization techniques used for the presentation of information enabling visual information exploration over large databases. The tutorial presents the reader with visual examples of each technique, along with notes describing the construction of each graphical representation. An extremely diverse number of visualizations are demonstrated, providing the readers with a solid reference for both basic and advanced visualization techniques.

Summary. This tutorial provides extensive details with respect to various information visualization techniques for large data sets. A brief overview of information visualization

goals is listed including (a) explorative analysis, (b) confirmative analysis, and (c) presentation. The primary focus of the tutorial is on explorative analysis. As such, the author notes that such analysis consists of (a) establishing a starting point, (b) defining a process, and (c) generating a result. Data exploration is defined as the process of searching and analyzing databases to find implicit but potentially useful information. The author classifies data visualization techniques as being (a) geometric, (b) icon-based, (c) pixel-oriented, (d) hierarchical, (e) distortion, and (f) dynamic/interaction. Keim provides examples of visualization techniques for each classification (see Figures 6 - 15).

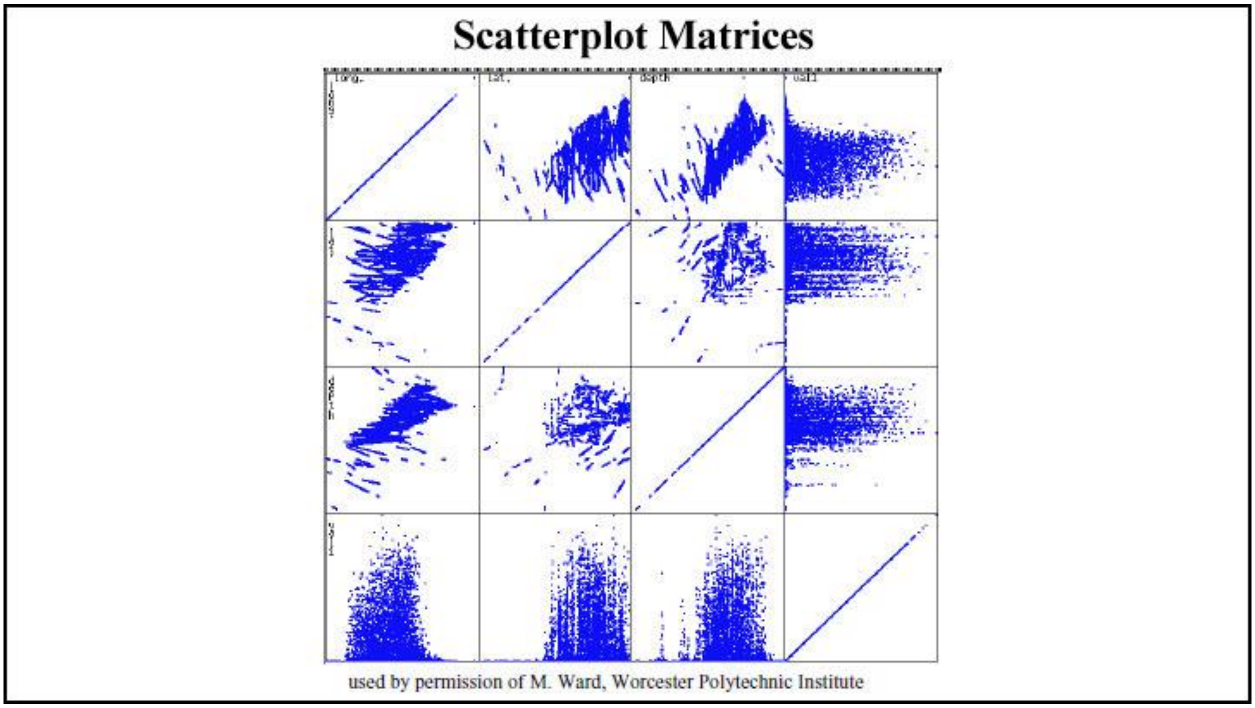

Figure 6. **Scatterplot matrices.** This visualization demonstrates a geometric technique, depicting how multiple visualizations can be aligned in a single visual field. Few (2006) notes that such visualization organization provides highly effective and rapid comparison of data sets representing similar data domains (Few, 2006).

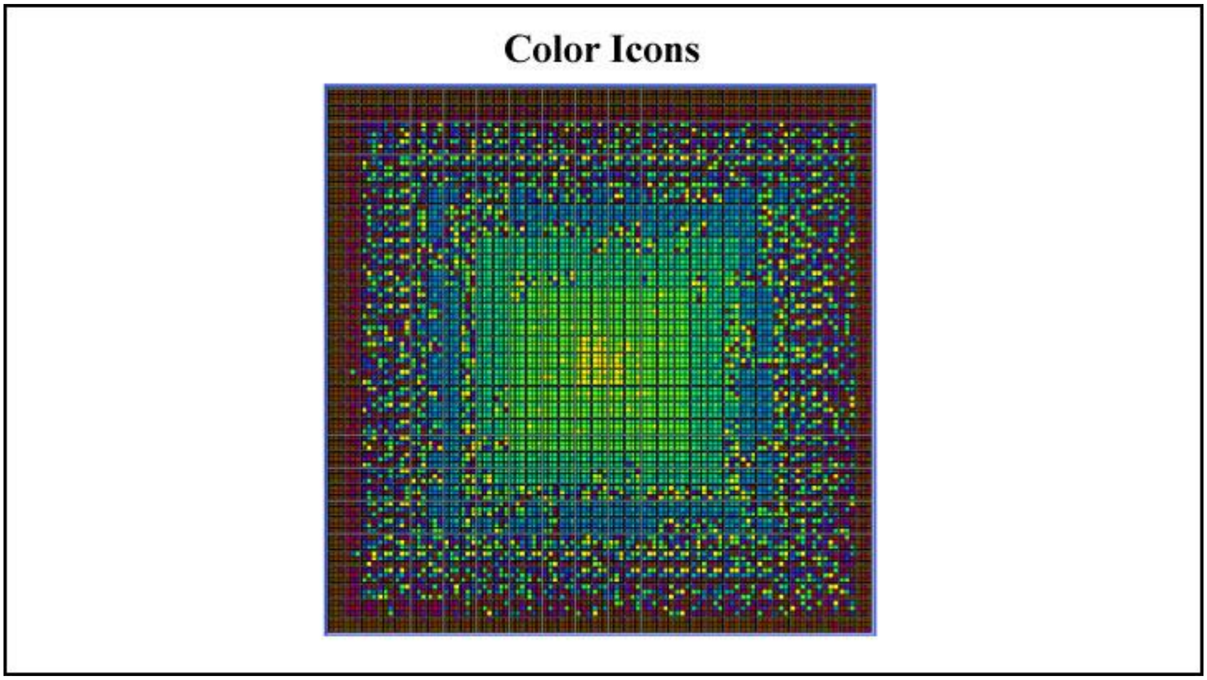

***Figure 7*. Color icons.** This visualization demonstrates an icon-based technique.

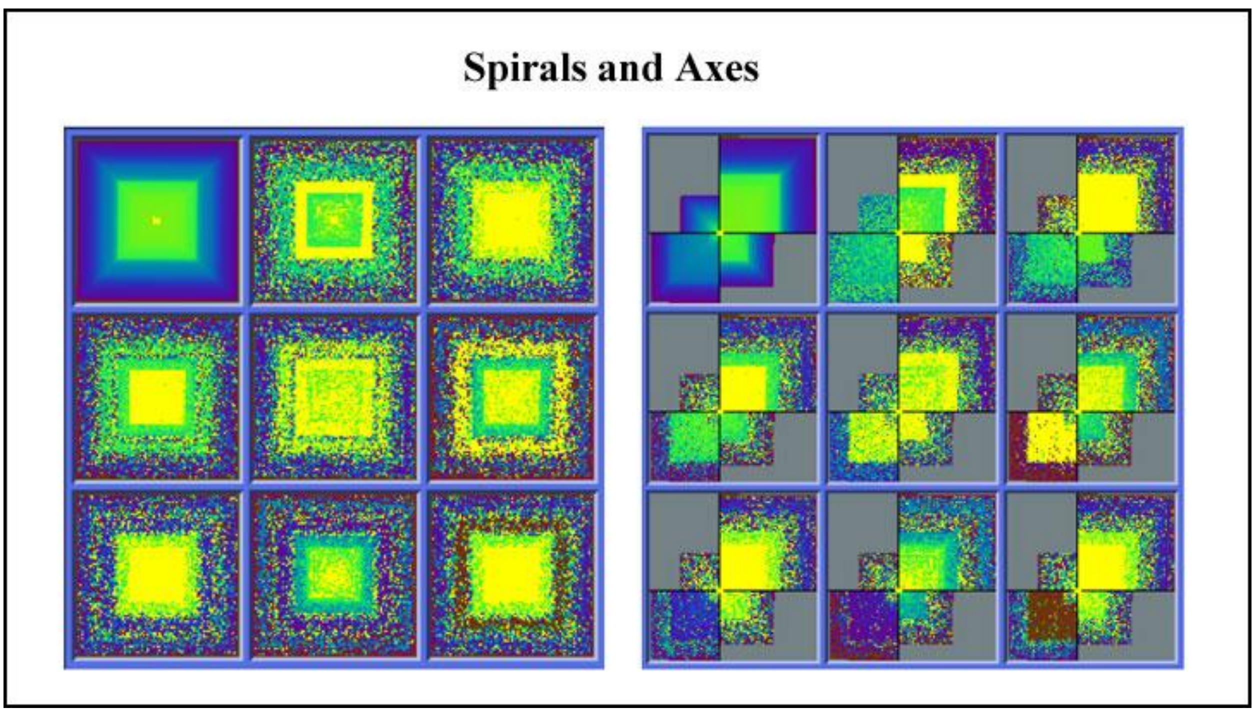

***Figure 8*. Spirals and Axes.** This visualization demonstrates pixel-oriented techniques.

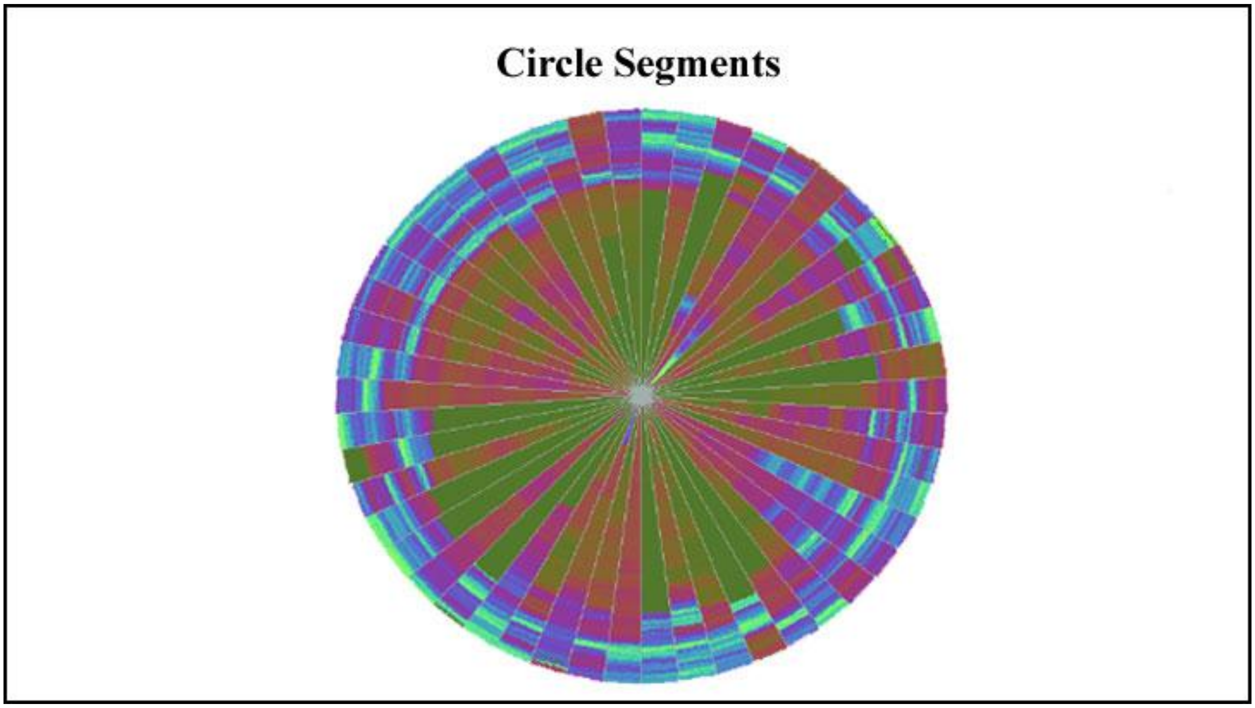

Figure 9. **Circle segments.** This visualization demonstrates pixel-oriented techniques.

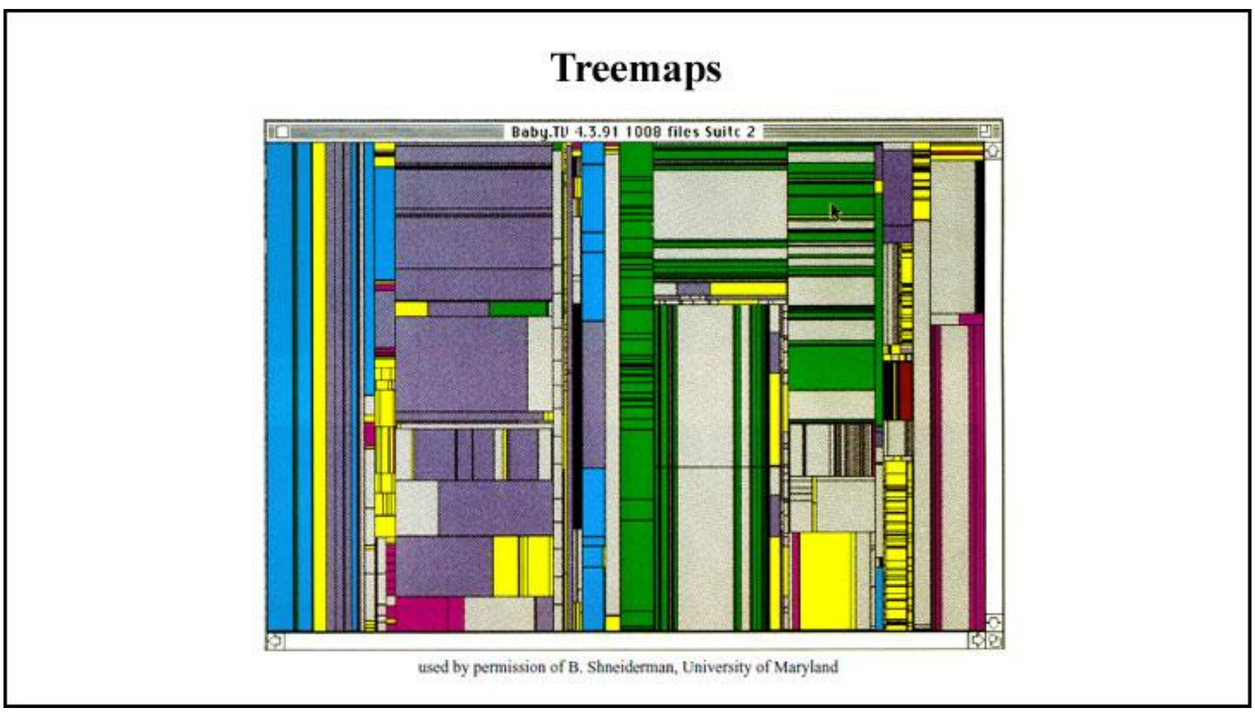

Figure 10. **Treemaps.** This visualization demonstrates hierarchical techniques.

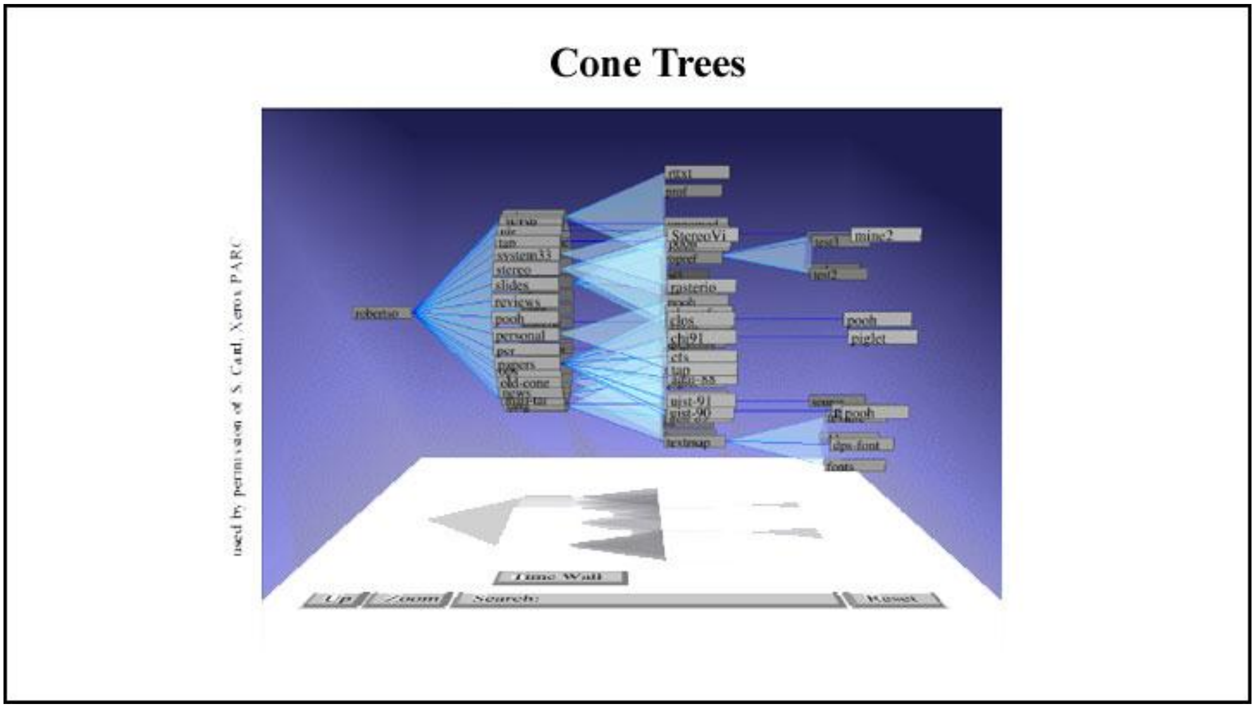

***Figure 11*. Cone trees.** This visualization demonstrates hierarchical techniques.

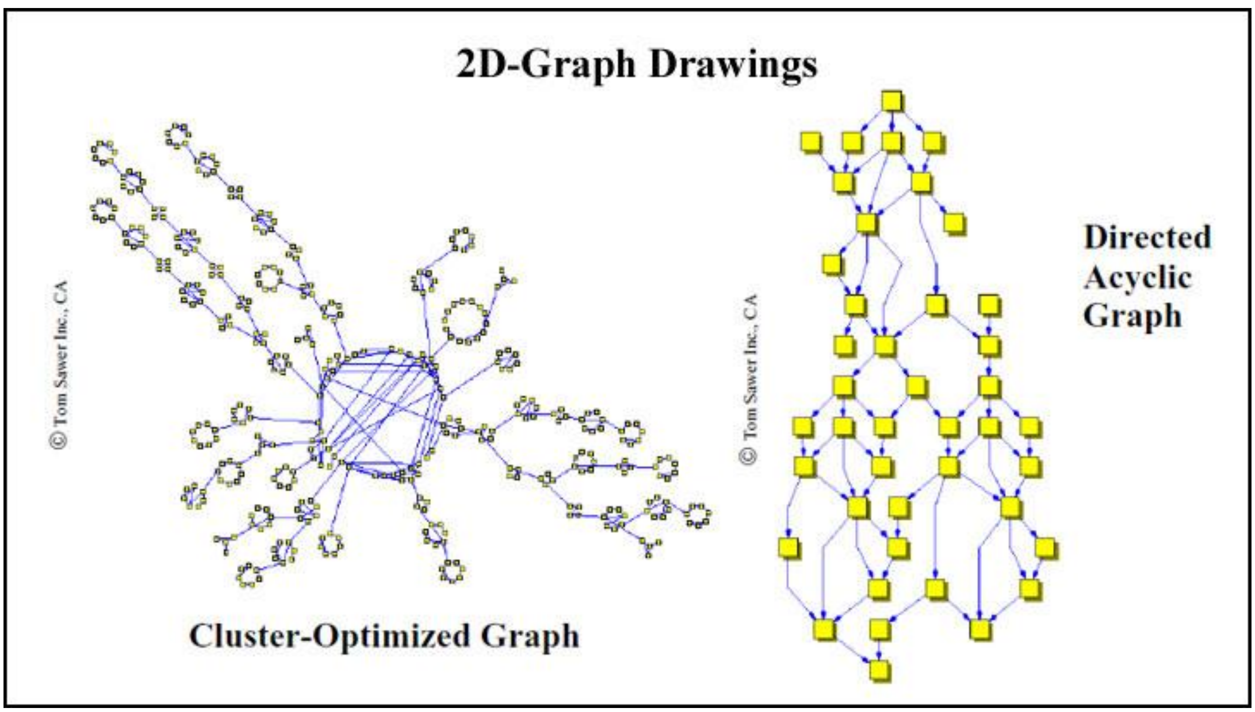

Figure 12. **2-D Graph drawings.** This technique demonstrates graph-based techniques.

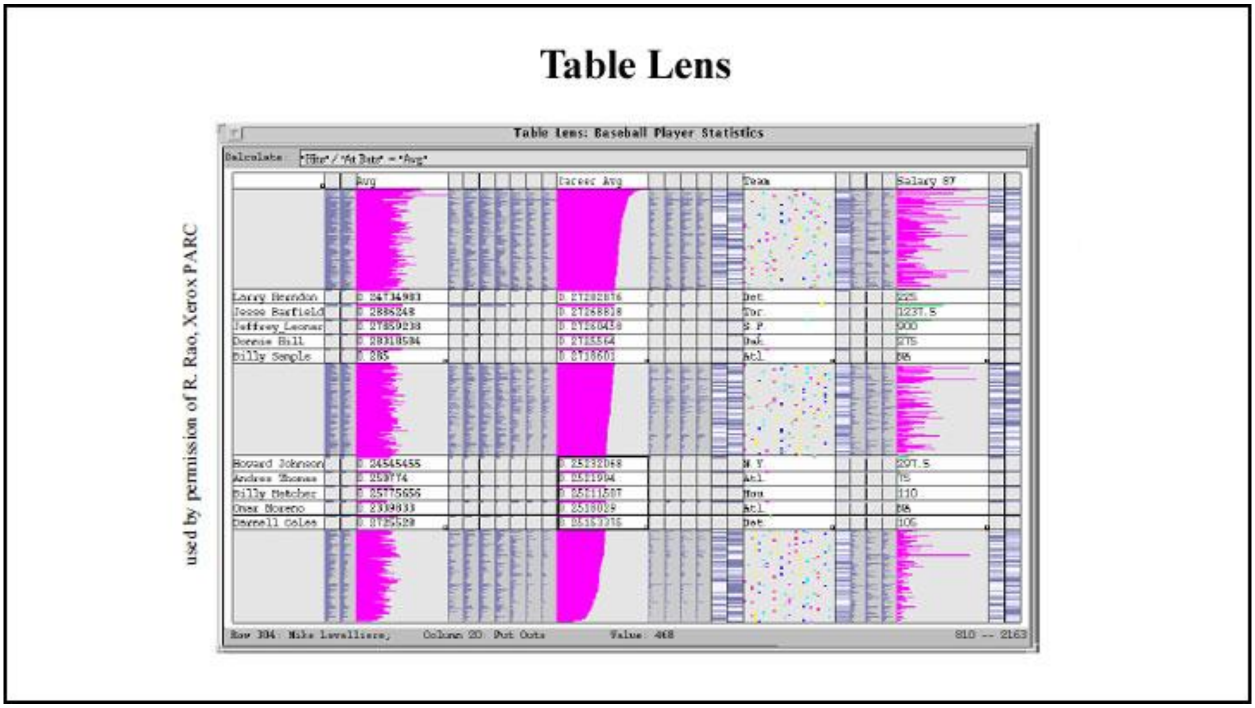

Figure 13. **Table lens.** This visualization demonstrates distortion techniques.

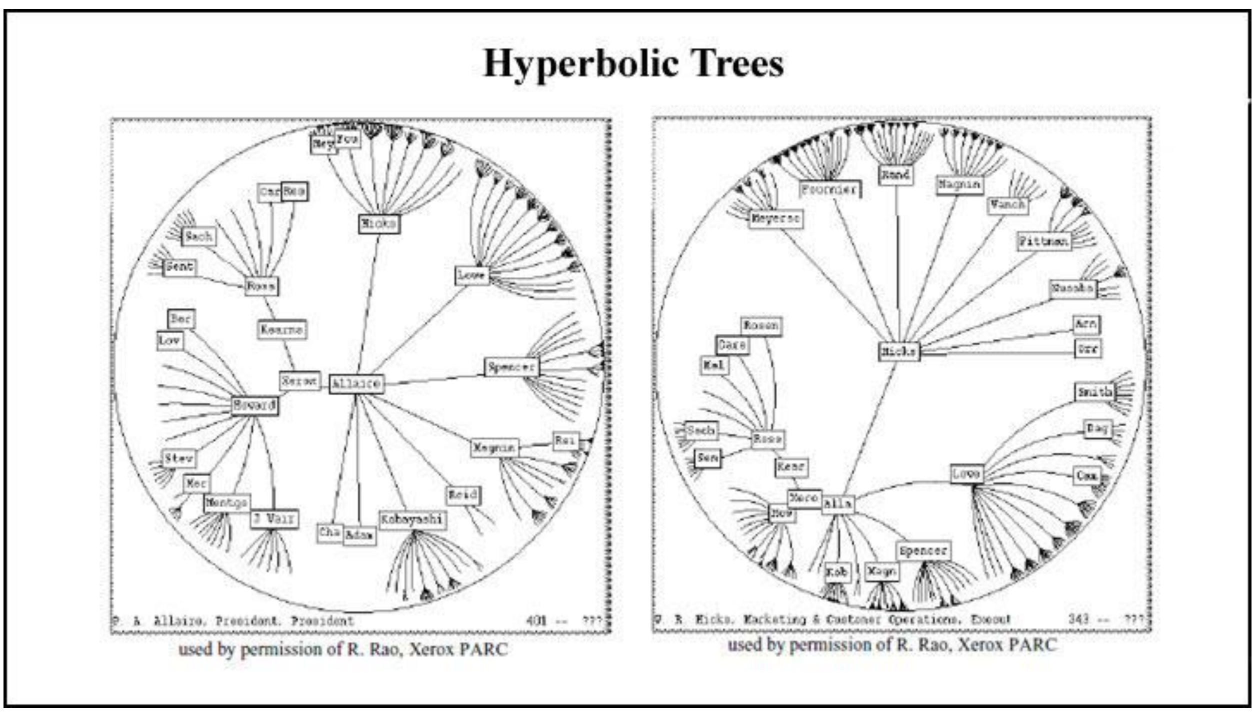

Figure 14. **Hyperbolic trees.** This visualization demonstrates distortion techniques.

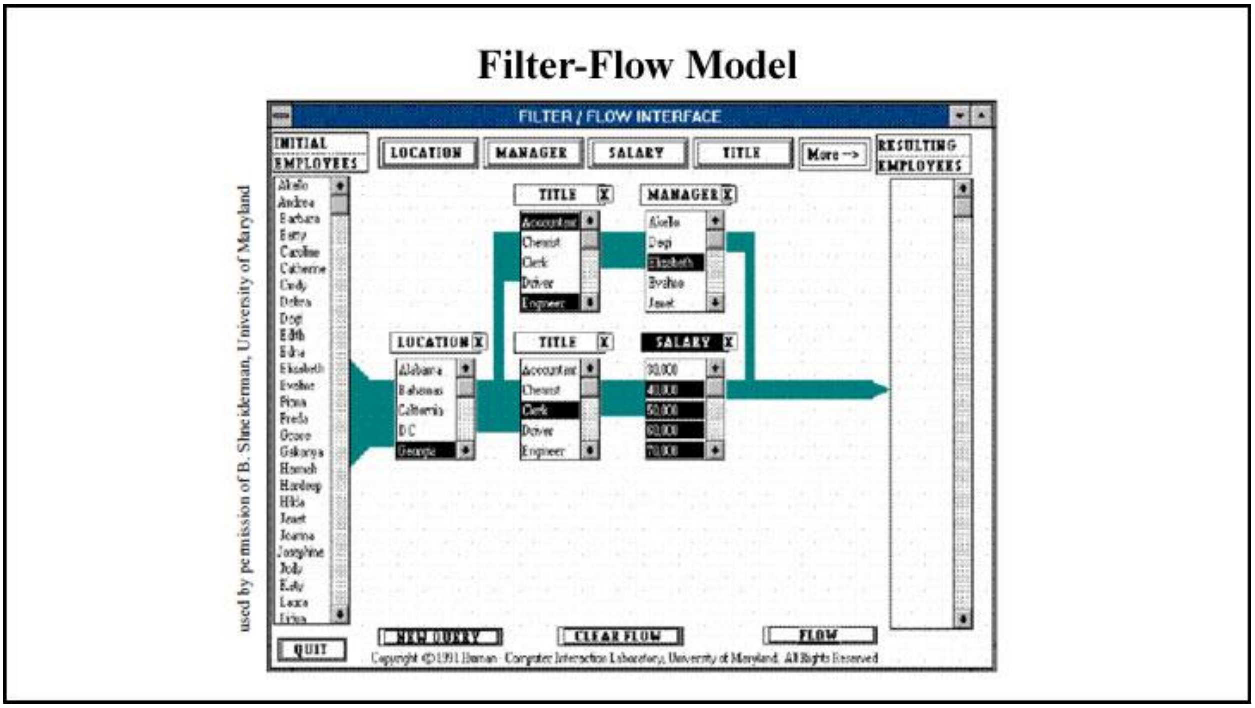

Figure 15. **Filter-flow model.** This visualization demonstrates dynamic/interaction techniques.

Of key importance is the authors diagraming of exploratory data visualization dimensions (see Figure 10). The visual provides a tool for determining when a technique from each classification may prove most relevant.

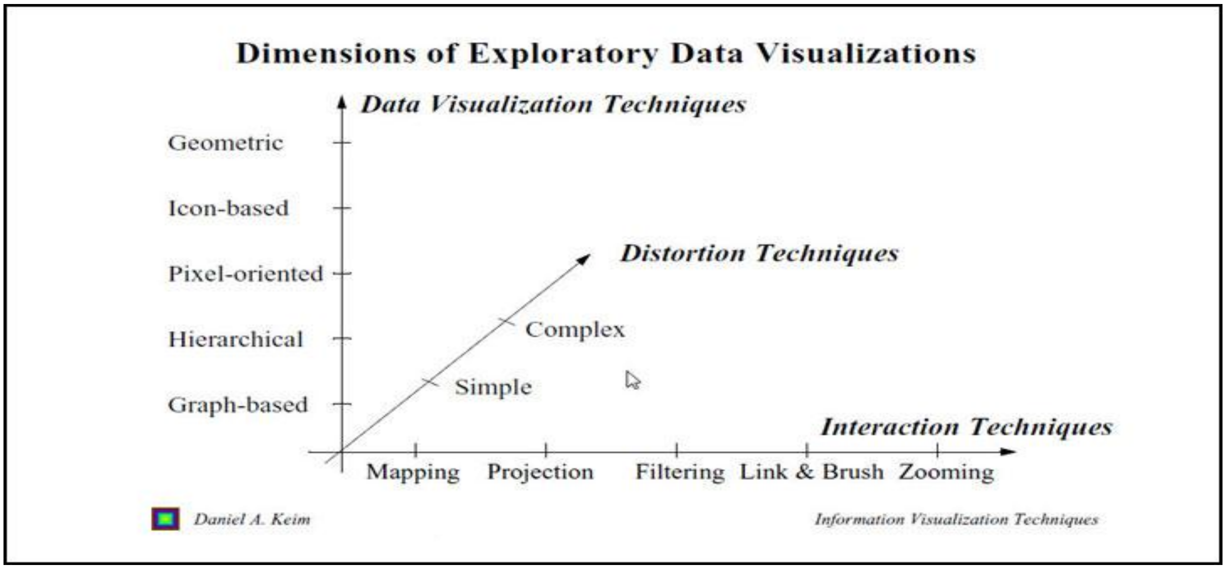

***Figure 16*. Dimensions of exploratory data visualizations.** This figure illustrates data visualization (distortion) and interaction techniques most appropriate for the level of complexity associated with the data set.

Credibility. Daniel Keim is a PhD in Computer and Information Science at the Universität Konstanz in Germany. His areas of research include (a) data mining and knowledge discovery, (b) high-dimensional indexing, and (c) visualization of large databases. The Institute of Electrical and Electronics Engineers (IEEE) digital library houses peer- reviewed publications, conferences, and technology standards. The organization's overriding mission is to help set standards dedicated to advancing technological innovation and excellence. There are over 160 references listed that focus on topics such as (a) information seeking, (b) dynamic queries, (c) interactive visualization, (d) network visualization, (e) interface design, (f) exploratory visualization, and (g) interactive statistical graphs

Keim, D. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, *7*(1), 1.

Abstract. Never before in history has data been generated at such high volumes as it is today. Exploring and analyzing the vast volumes of data becomes increasingly difficult. Information visualization and visual data mining can help to deal with the flood of information. The advantage of visual data exploration is that the user is directly involved in the data mining process. There are a large number of information visualization techniques which have been developed over the last decade to support the exploration of large data sets. In this paper, we propose a classification of information visualization and visual data mining techniques which is based on the data type to be visualized, the visualization technique and the interaction and distortion technique. We exemplify the classification using a few examples, most of them referring to techniques and systems presented in this special issue.

Summary. This paper discusses the importance and benefits of visual data mining based on targeted information visualizations. It is noted that visual data exploration aims to integrate the human in data exploration processes, allowing for the application of perceptual abilities. Human must be brought into exploration process. This form of exploration is especially beneficial when the data sets are (a) extremely large and (b) multidimensional. The main advantages of visual exploration/data mining compared to automated data mining are (a) visual data exploration is intuitive and requires no understanding of complex mathematical or statistical algorithms and (b) visual data exploration can easily deal with noisy, non-homogeneous data. Overall, visual exploration is faster and provides better results. The authors note three types of classifications used to organize modern information visualization: (a) the data type to be visualized, (b) the visualization technique, and (c) the interaction and distortion technique

used. Each classification is discussed in detail providing both textual and visual examples. The authors conclude that future work will involve the tight integration of visualization techniques with traditional techniques from such disciplines as statistics, machine learning, operations research, and simulation.

Credibility. Daniel Keim is a PhD in Computer and Information Science at the Universität Konstanz in Germany. His areas of research include (a) data mining and knowledge discovery, (b) high-dimensional indexing, and (c) visualization of large databases. The Institute of Electrical and Electronics Engineers (IEEE) digital library houses peer- reviewed publications, conferences, and technology standards. The organization's overriding mission is to help set standards dedicated to advancing technological innovation and excellence. The article contains 67 references on topics such as (a) information visualization, (b) treemaps, (c) visualization in functional design, (d) design considerations for information exploration, and (e) multidimensional information visualization.

Moreville, P. & Rosenfeld, L. (2006). *Information architecture for the world wide web.* Sebastopol, CA: O'Reilly Media, Inc.

Abstract. This book focuses on numerous aspects related to the discipline of information architecture. The profession can be broken out into the following definitions: (a) the structural design of shared information environments; (b) the combination of organization, labeling, search, and navigation systems within web sites and intranets, (c) the art and science of shaping information products and experiences to support usability and findability, and (d) the emerging discipline and community of practice focused on bringing principles of design and architecture to the digital landscape.

Summary. This book discusses the practice of information architecture. The author notes that a single definition does not exist for the profession. The following definitions are provided: (a) the structural design of shared information environments; (b) the combination or organization, labeling, search, and navigation systems within web sites and intranets, (c) the art and science of shaping information products and experiences to support usability and findability, and (d) an emerging discipline and community of practice focused on bringing principles of design and architecture to the digital landscape. As such, the book goes into great detail on the following: (a) presentation, (b) organization, (c) navigation, and (d) search of information organizations and graphic representations. The author discusses such topics as (a) controlled vocabularies, (b) synonym rings, (c) progressive disclosure, (d) taxonomies, and (e) metadata. Additionally, the book discusses information architecture research and considerations when applying information architecture in a corporate setting. Multiple visual examples of information architecture techniques are presented to clarify applied concepts.

Credibility. Peter Morville is the president of Semantic Studios, an information architecture and strategy consultancy. He holds an advanced degree in library and information science from the University of Michigan, where he now teaches graduate level information architecture. His work has been featured in numerous publications including Business Week, Fortune, MSNBC, and the Wall Street Journal. Lou Rosenfeld is an information architecture consultant and manages Rosenfeld Media, which publishes user experience books and provides user experience training. He has also consulted for organizations such as Paypal, Accenture, AT&T, Caterpillar, Lowes, the Centers for

Disease Control, Ford, and Microsoft. Rosenfeld holds a Masters in Information and Library Studies.

Nagel, H. (2006). Scientific visualization versus information visualization. doi: 10.1.1.164.319

Abstract. While scientific visualization techniques are used for the clarification of well-known phenomena, information visualization techniques are used for searching for interesting phenomena. There are important differences between the two fields of research which are compared, with focus on explaining Information Visualization techniques for the needs of scientific visualization researchers. This paper lists the main characteristics of the two fields of research, describes common Information Visualization techniques and discusses differences and similarities in the software that are commonly used in the two fields of research.

Summary. This paper summarizes the most important and seemingly irreconcilable differences between information and scientific visualization. The author notes that consideration of the techniques used in each field can lead to new ways of visualizing data that could benefit both communities. Information visualization is defined as the use of interactive visual representations of abstract, non-physically based data to amplify cognition. Scientific visualization is defined as the use of interactive visual representations of scientific data, typically physically based, to amplify cognition. It is noted that a key use of information visualization is to generate internal mental models of the information content in data sets; models which subsequently can be used for characterization, prediction, and/or decision making. The author lists and details the most common visual data exploration techniques including (a) icons and glyphs, (b) brushing and linking, and (c) The Grand Tour. It is also noted that unique software exists

for each visualization domain. Scientific visualization typically enables linking and interaction in relation to visual data objects. Information visualization typically provides a “single-pass” solution where the visualization can be interacted with in a limited fashion. The overriding conclusion drawn is that techniques from both communities need to be analyzed together in order to create solutions that complement each visualization need.

Credibility. Henrik Nagel is a senior engineer at NTNU – Trondheim, the Norwegian University of Science and Technology in the Department of Computer and Information Science. He holds a Post Doc at NTNU and is an Assistant Professor at Aalborg University. The article was retrieved using Google Search. There are eight references listed focusing on (a) information visualization, (b) visual data mining, and (c) visualization of multidimensional data.

Ware, C. (2012). *Information visualization: Perception for design*. Boston: Morgan Kaufmann.

Abstract. Information Visualization: Perception for Design is a comprehensive guide to what the science of human perception tells us about how we should display information. The human brain is a super-computer for finding patterns in information. Our understanding of visual data and visual information is greatly enhanced or impeded by the way information is presented. It is essential that visual data be designed in such a way that key information and important patterns will stand out. It is only by understanding how perception works that the best visualizations can be created. Colin Ware outlines the key principles for a wide range of applications and designs, providing designers with the tools to create visualizations of improved clarity, utility and persuasiveness. The book continues to be the key resource for practical design guidelines, based on perception, which can be applied by practitioners, students and researchers alike. Packed with over 400 informative full color illustrations which are key to understanding of the subject, this is a book about what the science of perception can tell us about visualization. The purpose of this book is to extract from that large body of research literature those design principles that apply to displaying information effectively.

Summary. This chapter, from the book Information Visualization: Perception for Design, focuses on information visualization interactions. Specifically, the chapter discusses what are called espistemic actions that center on low-level interaction (data manipulation) and exploration. Such an action is defined as an activity intended to uncover new information. To accommodate such actions, the chapter focuses on numerous issues related to each action: (a) choice reaction time; (b) two-dimensional positioning and selection, (c) hover queries, (d) path tracing, (e) spatial navigation metaphors; (f) wayfinding, cognitive maps, and real maps; (g) landmarks, border, and place, (h) frames of reference, (i) distortion techniques, (j) rapid zooming techniques, and (k) multiple simultaneous views.

Credibility. Colin Ware is the Director of the Data Visualization Research Lab, which is part of the Center for Coastal and Ocean Mapping at the University of New Hampshire. He is a professor in both the Ocean Engineering and Computer Science Departments. He specializes in advanced data visualization and has a special interest in applications of visualization to Ocean Mapping. Ware has published over 140 articles in scientific and technical journals.

Zhang, J., Johnson, K., Malin, J. & Smith, J. (2002). Human-centered information visualization. Retrieved November 22, 2012 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.121.6064.

Abstract. Information visualization should be human-centered because by definition it is the process of designing information to match the processing characteristics of the human visual system. This is largely true for many special purpose visualization products regardless of whether the human-centered approach is taken deliberately in the design process. However, this type of human-centered visualization is typically only at the level of representations, which are relatively independent of tasks, users, and functions. In this paper we argue that human-centered visualization should be considered not just at the level of representations but also at the levels of functions, users, and tasks. This multiple level approach is important for the design of complex information systems that support multiple types of users performing varieties of tasks in different contexts to achieve different goals. We will first describe a framework of multilevel human-centered visualization. Then we will use one simple example to demonstrate the concept of this multilevel human-centered approach, focusing on the relations between tasks and representations.

Summary. This article argues that human-centered visualization should be considered not just at the level of visual representations, but also at the level of user tasks. The authors note that visualizations take advantage of human's powerful visual systems by (a) bypassing the inherent bottleneck of human memory, and (b) leveraging easily distinguishable physical symbols. The authors provide an informative list of inherent visualization features that make information visualizations extremely useful: (a) they provide short-term or long-term memory aids, (b) they provide information that can be directly perceived and utilized with little cognitive processing, (c) they provide knowledge and skills that are unavailable from internal representations, (e) they support perceptual objects that enable easy feature recognition and inference generation, (f) they anchor and structure cognitive behavior without conscious awareness, (g) they change the nature of a task by generating more efficient action sequences, (h) they aid processibility by limiting abstractions, and (i) they determine decision making strategies through accuracy maximization and effort minimization. The authors provide a detailed, hierarchical framework of human-centered visualization based on (a) functional analysis, (b) user analysis, (c) task analysis, and (d) representational analysis. To provide context, the authors present a case study focusing on information displays for aviation.

Credibility. Jiajie Zhang is a PhD and Director at the National Center for Cognitive Informatics and Decision Making in Healthcare. He is also a professor in the School of Biomedical Informatics at the University of Texas at Houston. His research interests include biomedical informatics, human-centered computing, cognitive science, and decision making. Kathy Johnson is affiliated with the School of Health Information Sciences at the University of Texas at Houston and the NASA Space Center. Jane Malin is an Expert Consultant at the NASA Johnson Space Center. She previously worked as a Principal Scientist at Lockheed EMSCO. Jack Smith is a PhD and professor at the University of Texas at Houston in the School of Biomedical Informatics. CiteSeerX is an online scientific digital library focusing primarily on computer and information science literature. Articles available in the database are peer-reviewed publications. The database strives to remain compliant with the Open Archives Initiative Protocol for Metadata

Harvesting, a standard proposed by The Open Archive Initiative for content dissemination. The article lists 24 references focusing on topics such as (a) information visualization, (b) information displays and decision processes, (c) information processing and human-machine interaction, and (d) cognitive tasks.

Theme 4: Information Visualization Decision Support and Effectiveness Evaluation

References within this theme focus on concept nodes regarding (a) decision-making, (b) sense-making, (c) knowledge, and (d) effectiveness evaluation. These nodes are highlighted in purple (see Figure 17) within the larger concept map.

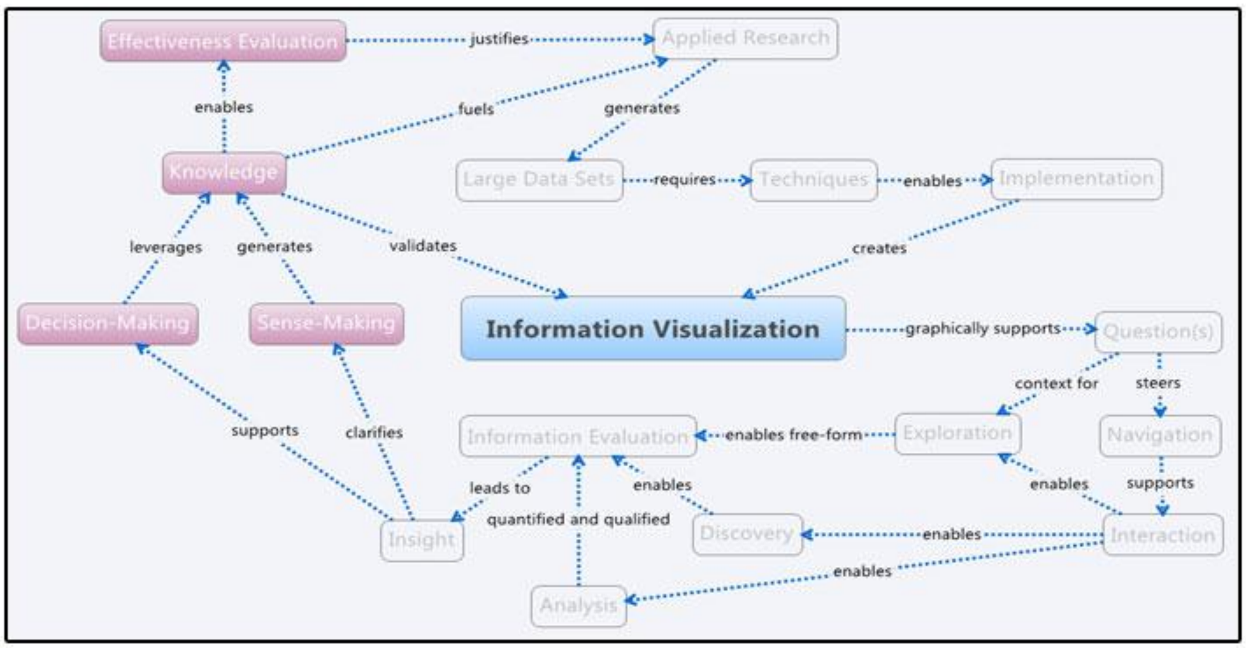

Figure 17. **Visualization decision support and effectiveness evaluation concepts.**

Ahokas, T. (2008). Information visualization in a business decision support system. Retrieved November 21, 2012, from https://helda.helsinki.fi/handle/10138/21417.

Abstract. Information visualization is a process of constructing a visual presentation of abstract quantitative data.   The characteristics of visual perception enable humans to recognize patterns, trends and anomalies inherent in the data with little effort in a visual display. Such properties of the data are likely to be missed in a purely text-based presentation. Visualizations are therefore widely used in contemporary business decision support systems. Visual user interfaces called dashboards are tools for reporting the status of a company and its business environment to facilitate business intelligence (BI) and performance management activities. In this study, we examine the research on the principles of human visual perception and information visualization as well as the application of visualization in a business decision support system. A review of current BI software products reveals that the visualizations included in them are often quite ineffective in communicating important information. Based on the principles of visual perception and information visualization, we summarize a set of design guidelines for creating effective visual reporting interfaces.

Summary. This thesis focuses on information visualizations within decision support systems (DSS). The author identifies two main objectives for the paper: (a) a review of literature on the principles of information visualization and the underlying principles of visual perception in order to understand which factors influence the effectiveness of visual presentations of abstract quantitative data, and (b) to learn how these principles can be utilized in designing visual reporting and analysis interfaces in business intelligence and performance management systems. The thesis is broken out into five sections: (a) principles of visual perception that may be regarded as most important for understanding information visualization, (b) principles of information visualization; (c) an introduction to decision support systems, with an emphasis on business intelligence and performance management, (d) discussion of how the principles of information visualization can be used to design effective visual interfaces in business intelligence and performance management systems, and (e) a summary of findings. Regarding visual perception, the author provides examples of numerous grouping (Gestalt) principles, which enables rapid recognition of comparisons.

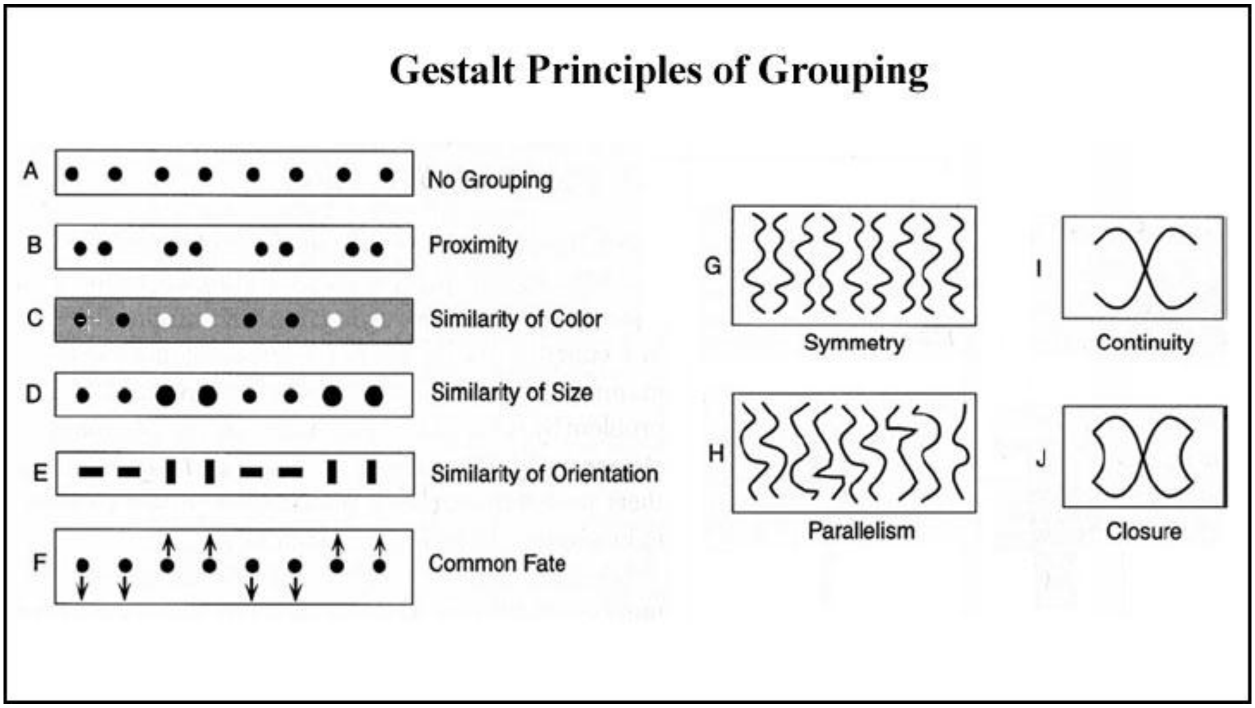

Figure 18. **Examples of Gestalt grouping principles.**

The author states that color differentiation within visualizations is essential for effective representations. He points out that the three color properties of (a) hue (humans can generally distinguish 200 variants), (b) saturation, and (c) lightness are used extensively to create visualizations that easily distinguishable by everyone; including individuals with color blindness. Regarding visual attention and memory, the author states that attention provides the interface between the physiological visual system and the cognitive processes in the human mind. Three memory subsystems are identified with relation to visual perception and memory: (a) iconic memory, (b) visual working memory, and (c) visual long-term memory. Preattentive processing (early stage of visual perception that rapidly occurs below the level of consciousness) is noted as a first step in visualization processing leading to attention and cognition. In an effort to maximize information visualization effectiveness, the author recommends keeping several aspects of visualization creation in mind at all times: (a) color coding , (b) scaling, (c) data set ordering, (d) small multiples, and (e) visualization interaction. The key to estimating visualization effectiveness is noted as the accuracy of judging different visual attributes based on the specific scale of the encoded variable(s). With respect to information visualizations used within decision support systems, the author notes dashboards (graphical interfaces for reporting business metrics) as advantageous to presenting large amounts of information in a concise fashion. It is noted that information dashboards are the most important part of the information processing chain in the sense that they represent (a) organized representations of core data and (b) serve as the interface through which users interact with the DSS and the underlying data. To support information dashboard development, the author details numerous graphical elements supporting sense making. He notes that the visualization technique used is highly dependent on the information need of the systems users. The most frequently used techniques – regardless of appropriateness - include (a) line and bar graphs, (b) pie charts, (c) sparklines, (d) gauges, and (e) bullet graph.

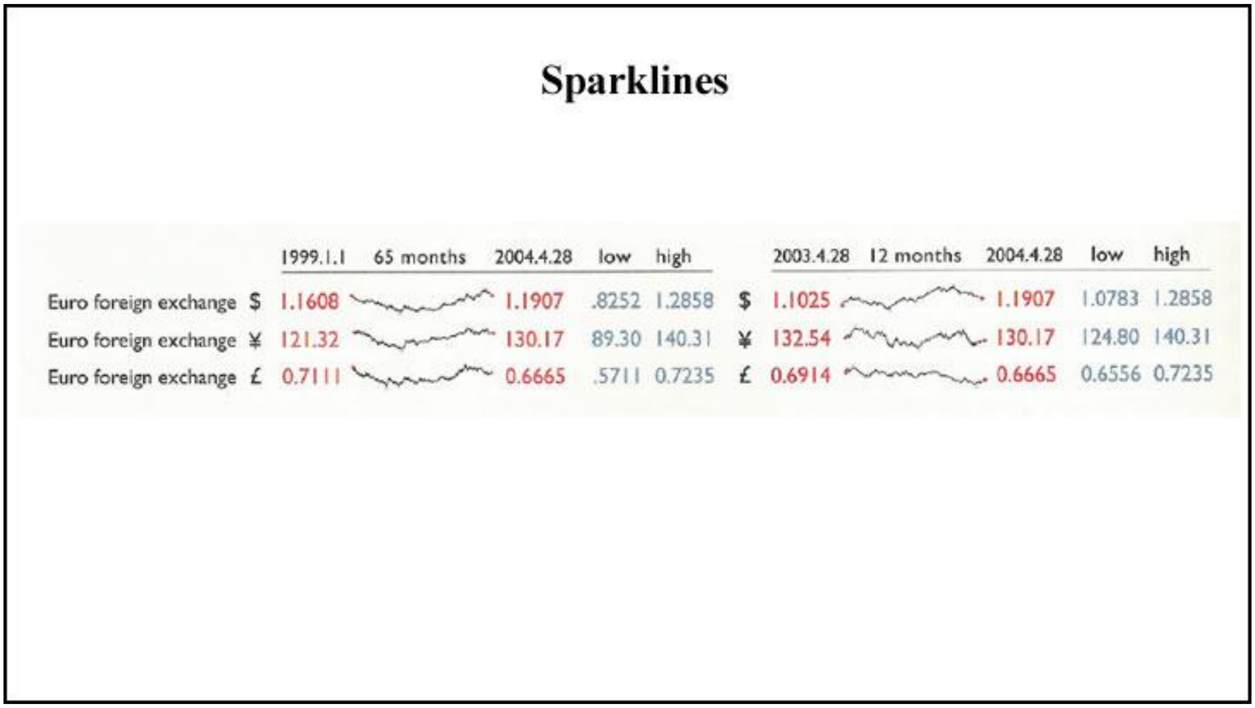

Figure 19. **Sparklines.**

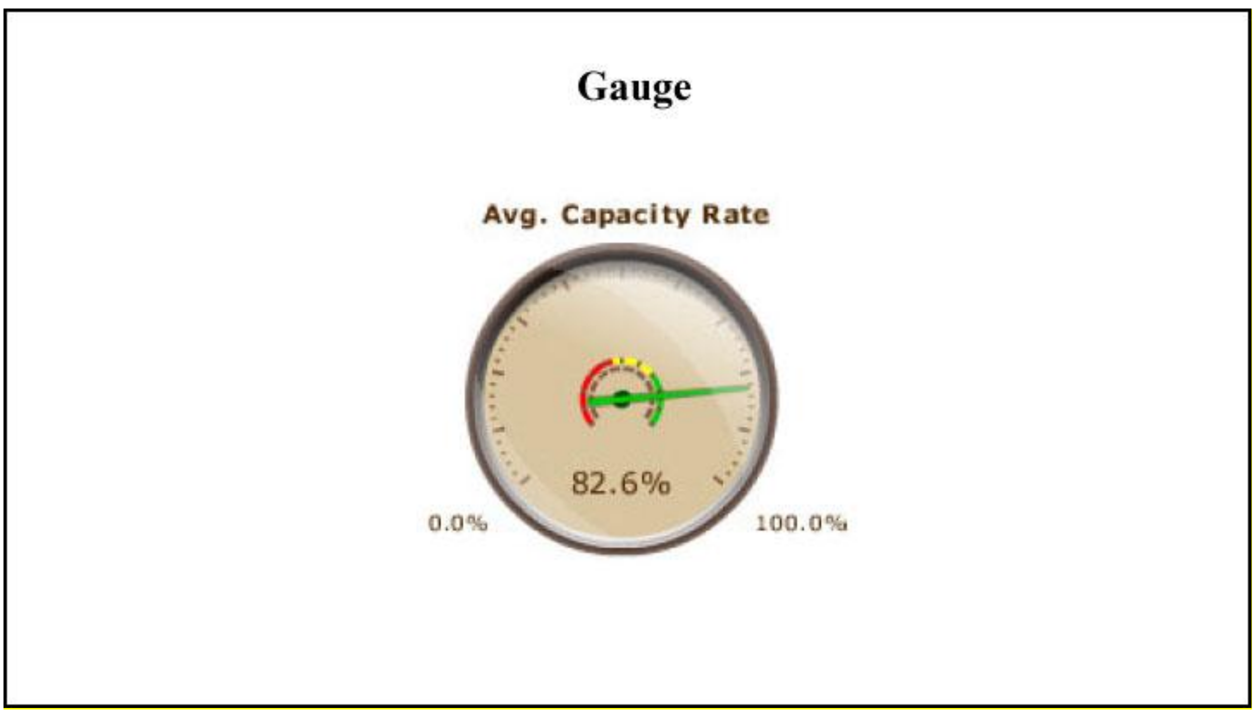

***Figure 20*. Gauge.**

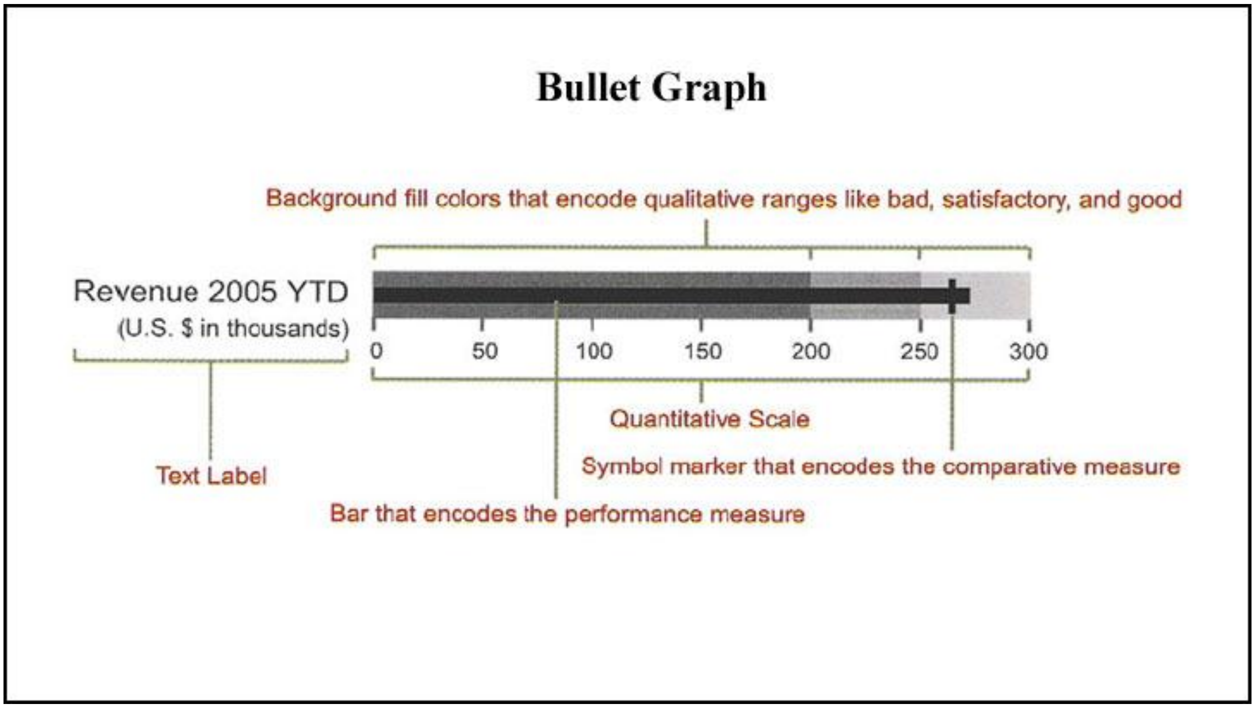

***Figure 21*. Bullet graph.** This visualization is provided by Few (2006).

When designing dashboards, the author recommends four guidelines to keep in mind: (a) provide context to ground the user so they are not lead astray accidentally, (b) avoid distracting or confusing graphic perspectives such as three-dimensional (3D) effects that can occlude data points, (c) provide interactive visualizations that support navigation within a data set in order to filter or drill down into finer detail that supports questions or hypothesis, and (d) weigh the value of using either graphs or tables based on the number of data points and communication intention of the unique visualization. Five examples of information dashboards are presented. An analysis of each dashboard visualization is provided, highlighting (a) techniques mentioned throughout the paper and (b) common mistakes made based on poor/misinformed design decisions. In conclusion, the author identifies two findings based on the research: (a) business intelligence (BI); and meaningful decision making, means more than merely automated technological solutions void of active human perception and cognition, contrary to common messages received by mainstream software vendors; and (b) the rather quick emergence of visual reporting and analysis software tools seems to be the result of improved data collection and processing capabilities, providing a powerful channel (visual perception) for absorbing large amounts of information with little cognitive effort.

Credibility. At the time of writing, Tapio Ahokas was a graduate student in the Department of Computer Science at the University of Helsinki. This article represents his Master's thesis. The department has numerous research groups focusing on (a) adaptive computing, (b) interactive systems, (c) data mining theory and application, and (d) machine learning. The article is a Master's thesis published by the Department of Computer Science, University of Helsinki. There are 56 references listed focusing on topics such as (a) decision support, (b) information visualization, (c) visualization research, (d) data analysis, (e) cognitive psychology, (f) business intelligence, and (g) decision-making.

Fabrikant, S., & Buttenfield, B. (2001). Formalizing semantic spaces for information access. *Annals of the Association of American Geographers, 91*(2), 263-280.

Abstract. An exponentially growing volume of digital information makes extraction of relevant items increasingly difficult. This article documents the adoption of information visualization tools by researchers in the disciplines of geography, computer science, and information science to facilitate exploration of very large data archives. Graphic depiction of database content (or the database "semantics") can be based on a spatial or even a geographic metaphor. Such depictions, often called information spaces or information worlds, provide one example of "spatialization". Various forms of spatialized views are critiqued in this article. To date, systematic approaches to the creation of spatialized views have lacked solid theoretical foundations. Three spatial frames of reference are presented to formalize and visualize semantic spatialized views: geographic space, cognitive space, and Benediktine space (named after Michael Benedikt). Application to an example of a very large online catalog (GEOREF) highlights the underlying assumptions of the space types and demonstrates what spatial properties are preserved for each proposed approach.

Summary. This article explores the application of cognitive and spatial concepts related to the exploration, navigation, and knowledge extraction of very large datasets such as online archives and digital libraries. The article also documents the adoption of information organization and visualization tools by researchers in geography, computer science, and information science to facilitate the exploration of very large datasets. The authors point out that recent empirical testing indicates that the generation of information visualizations substantially helps people access data in online archives more effectively than traditional text-based representations containing alpha-numerically scrolling lists, keyword query boxes, and similar tools. It is communicated that a conceptual shift is emerging within the field of information retrieval, placing less importance on known-item searches based on explicit keywords. Sense-making, and subsequent decision-making, is being challenged by such information interaction processes as (a) interactive data mining, (b) information grazing, and (c) information foraging. In these cases, the information seeker becomes the main actor and the concept of the search task becomes the center of research investigation. It is noted that spatial metaphor (visualization of abstract, multidimensional data) in information visualization allows the viewer's intrinsic comfort with everyday concepts of human spatial orientation and wayfinding to guide their exploration and interpretation of the representation. Ordination is discussed, which involves the reduction of high volumes of data into smaller, manageable units, creating a visualization solution space with the lowest number of data dimensions necessary to describe complex phenomenon while still providing inherent meaning. Three primary spatial frames (visualization metaphors) of reference to formalize semantic spatialization of large information spaces are discussed: (a) geographic/location, (b) cognitive (interpretation of representations), and (c) Benediktine (based on human-computer interaction). As a means to provide an applied example, the authors describe and illustrate spatialized views of a semantic information space related to a digital library archive. In conclusion, the authors state that the use of graphical metaphors provides a viable solution to overcome some of the access problems associated with increasing difficulties generating insight and meaning within large information repositories.

Credibility. Sarah Fabrikant is a PhD and Head of Geographic Information Visualization & Analysis in the Department of Geography at the University of Zurich. She is actively involved in research projects focusing on topics such as (a) animated visual analytics, (b) neuroimaging leverage colors, (c) time information visualization and analysis, and (d) measuring the complexity of map designs based on metrics and perception. Barbara Buttenfield is a PhD and Professor of Geography at the University of Washington. She teaches courses in Geographic Information Science, GIS Modeling, Cartography and Information Design. She is the past President of the American Cartographers

Association. The article lists over 30 references focusing issues such as (a) visual information seeking, (b) collaboration in virtual environments, (c) proactive maps for exploratory visualization, (d) sense-making research, (e) mental maps, and (f) the cognitive aspects of human-computer interaction.

Fekete, J., Vanwijk, J., Stasko, J., & North, C. (2008). The value of information visualization. Retrieved November 20, 2012 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.140.2850.

Abstract. Researchers and users of Information Visualization are convinced that it has value. This value can easily be communicated to others in a face-to-face setting, such that this value is experienced in practice. To convince broader audiences, and also, to understand the intrinsic qualities of visualization is more difficult, however. In this paper we consider information visualization from different points of view, and gather arguments to explain the value of our field.

Summary. This paper discusses the attribution of quantifiable value to information visualization. The authors note that information visualization systems/solutions are best applied for unguided exploratory tasks that involve interfacing with extremely large information spaces (data sets). The assertion is made that information visualization is primarily about developing insights from collected data, not about the immediate and holistic understanding of an entire information domain. The core benefits provided by visuals seem to hinge upon their acting as a frame of reference, or temporary storage area, for human cognitive processes. In essence, when viewed from an exploratory perspective, it is noted that information visualization can provide insight as to how emergent theories arise as opposed to clearly highlighting when they arise. The authors

note that information visualization is an inductive method in sense making that is focused on generating new insight and ideas - the seeds of theories - using human perception as a very fast filter. In accordance, they admit that activities like (a) exploration, (b) browsing, (c) gaining insight, and (d) asking better questions are not necessarily easily measured quantifiably. It is suggested that by identifying situations where browsing and exploratory information interaction is helpful in achieving a measurable goal, such as sense-making, it should be easier to determine information visualization value. The authors provide several visual examples where visualization enhances cognition and supports sense-making. Discussion as to what information visualization can accomplish in stark contrast to more traditional and well known fields such as statistics and data mining is captured. Numerous means by which visuals can amplify cognition are highlighted: (a) increasing memory and processing resources available, (b) reducing search for information, (c) enhancing the recognition of patterns, (d) enabling perceptual inference operations, (e) using perceptual attention mechanisms for monitoring, (f) encoding information in a medium enabling manipulation, and (g) increasing perceptual support. Economic value is attributed to information visualization by suggesting that visualization profit is the difference between the value of the increase in knowledge and the costs made to obtain insight.   The authors provide a detailed analysis of an algorithm that can be used to directly measure visualization value. In conclusion the authors note that although they have described the challenges and posed a number of answers for defining information visualization value, they assert that attributing value to information visualization is not easy and ultimately, the larger community of information visualization researchers and practitioners must create techniques and systems that clearly illustrate the value of the field.

Credibility. Jean-Daniel Fekete is a Senior Research Scientist and Director de Recherche 2e Classe at the University of Paris. He is the Scientific Leader of the INRIA Project Team AVIZ. He holds a PhD in computer science and has 137 publications. Jack van Wijk is a full professor in visualization at the Department of Mathematics and Computer Science of Eindhoven University of Technology. He holds an MS degree in industrial design engineering and PhD in computer science. John Stasko is Professor and Associate Chair in the School of Interactive Computing at the Georgia Institute of Technology. He holds a PhD in computer science from Brown University. He is a faculty investigator in the Department of Homeland Security's VACCINE Center of Excellence focusing on developing visual analytics technologies and solutions for grand challenge problems in homeland security, and in the NSF FODAVA Center exploring the foundations of data analysis and visual analytics. Chris North is an Associate Professor in the Department of Computer Science and Center for Human-Computer Interaction at Virginia Tech. His primary research areas include visual analytics, information visualization, and human-computer interaction. CiteSeerX is an online scientific digital library focusing primarily on computer and information science literature. Articles available in the database are peer-reviewed publications. The database strives to remain compliant with the Open Archives Initiative Protocol for Metadata Harvesting, a standard proposed by The Open Archive Initiative for content dissemination. The article lists 32 references focusing on topics such as (a) graphs in statistical analysis, (b) information visualization, (c) interactive visual exploration, (d) preattentive processing in vision, and (e) the visual display of quantitative information.

Forsell, C. (2012). Evaluation in information visualization: Heuristic evaluation (Ed). *2012 16th International Conference on Information Visualization* (pp. 136).n.d.: IEEE. doi: 10.1109/IV.2012.33.

Abstract. This paper presents a review of heuristic evaluation and recommendations for how to apply the method for information visualization evaluation. Heuristic evaluation is a widely known and popular method within the area of human-computer interaction and the information visualization community now also recognizes its usefulness. However, in this area it is not applied to the same extent. In its original form the method has limitations that need to be considered in order for it to be optimal for information visualization. The aim with this paper is to provide the reader with knowledge about the methods, and awareness of what issues, that call for refined or supplemental actions and resources in order for it to generate as valid and useful results as possible. The paper also discusses the research challenges for future work in how to further improve the method.

Summary. This article discusses the evaluation of information visualizations based on heuristic criteria. Heuristic evaluation is defined as an informal and subjective analytical method belonging to the larger family of usability inspection methods. The evaluators are typical experts in usability. Evaluation in the context of the article is defined as the systematic acquisition and assessment of information to provide useful feedback about some object. The author states that the purpose of any information visualization process is the promotion of insight and understanding in relation to the graphically represented data set. It is noted that evaluation can take place at any stage of visualization development, utilizing a wide range of methods. The goal of this article is to provide (a) a relatively simple review of information visualization heuristic techniques, (b) note some of the limitations of these techniques in order establish optimal evaluations, and (c) suggest how to proceed in heuristic evaluation keeping certain issues in mind. The key contribution of the article is a single, overall evaluation method for information visualization researchers, developers, and consumers. A significant benefit to scientific, heuristic evaluation of information visualizations is the ability to base the next generation of development on proven success with greater confidence and possible industry, academic, or public adoption. Based on research by cited authors, the overview of evaluation characteristics is classified based on (a) whether representative users participate or not in the evaluation, (b) the context of the evaluation, and (c) the types of data/results that are obtained. Heuristic evaluation is typically done by experts in usability. The evaluator utilizes a set of heuristic principles or guidelines for good design and implementation. The author cites Ten Usability Heuristics that establish a set of evaluation criteria, and generalized process flow, that have proven highly successful over time: (a) visibility of system status, (b) match between the system and the real world, (c) user control and freedom, (d) consistency and standards, (e) error prevention, (f) recognition rather than recall, (g) flexibility and efficiency, (h) aesthetic and minimalist design, (i) help users recognize, diagnose, and recover from errors, and (j) help and documentation functions. While adhering to aspects of this proposed process flow, the author also points out issues/approaches that could be addressed during the evaluation: (a) clear descriptions of heuristic application and interpretation, (b) inclusion of data experts or domain experts, (c) collaboration of evaluators in pairs, (d) creation of task scenarios, and (e) use in conjunction with user studies prior to summative empirical work. In conclusion, the author notes that there is no consensus as to what kind of heuristics prove most useful in information visualization evaluation and that future studies are needed to begin determining how many evaluators are really needed for successful heuristic evaluation.

Credibility. At time of writing, Camilla Forsell was associated with C-Research at Linkoping University in Sweden. The article is published in proceedings from the 2012 16th International Conference on Information Visualization. The Institute of Electrical and Electronics Engineers (IEEE) digital library houses peer- reviewed publications, conferences, and technology standards. The organization's overriding mission is to help set standards dedicated to advancing technological innovation and excellence. The article lists 50 references including topics on (a) information visualization for interaction, (b) heuristic evaluation in information visualization, (c) information seeking, and (d) animated transitions statistical data graphics.

Haroz, S., & Whitney, D. (2012). How capacity limits of attention influence information visualization effectiveness. *IEEE Transactions on Visualization and Computer Graphics*, *18*(12).

Abstract. In this paper, we explore how the capacity limits of attention influence the effectiveness of information visualizations. We conducted a series of experiments to test how visual feature type (color vs. motion), layout, and variety of visual elements impacted user performance. The experiments tested users' abilities to (a) determine if a specified target is on the screen, (b) detect an oddball, deviant target, different from the other visible objects, and (c) gain a qualitative overview by judging the number of unique categories on the screen. Our results show that the severe capacity limits of attention strongly modulate the effectiveness of information visualizations, particularly the ability to detect unexpected information. Keeping in mind these capacity limits, we conclude with a set of design guidelines which depend on a visualization's intended use.

Summary. This article represents formalized research based on visual searching and rapid counting capabilities in relation to varying types of information layouts. The authors ran three experiments with similar stimuli using either colored or moving visual features within a geometric grid representation. The visualization tasks included (a) detect a unique target with a known appearance (e.g. find the red object); (b) detect a unique target with an unknown appearance (e.g. find a unique or oddball target); and (c) determine and compare the number of visual categories (e.g. determine extent of heterogeneity or consistency). A primary goal of this research is quantification of the influence of grouped vs. random visualization arrangement on visual search and subitizing (rapid counting). It is hypothesized that the effect would vary by task and would modulate the impact of capacity limits. All experiments were divided into four blocks: (a) color – grouped (b) color – random, (c) motion – grouped, and (d) motion – random. The research provides insight that can help guide visualization design: (a) cluttered and randomly organized arrangements significantly impairs search for information targets (collective data points), (b) knowing a target's appearance in advance dramatically improves accuracy and speed of target detection, and (c) a compromise must be made between the number of nominal (named) categories and the perceptual complexity of a visualization. The authors note that searching for a known target is not always a sufficient test of visualization effectiveness and evaluations of visualizations should make certain to test user performance for more attentionally demanding user goals. Three main conclusions are provided: (a) grouping is far more beneficial for oddball search compared with known-target search, (b) accessing overall information (like heterogeneity or number of categories) is better for grouped displays, and (c) for difficult tasks, aim to reduce variety in the entire view rather than optimizing small regions.

Credibility. Steve Haroz is a PhD student in the Computer Science Graduate Group, UC Davis, working with Dr. David Whitney. His research applies visual perception and cognition to visualization and human-computer interaction (Haroz, .n.d.). David Whitney is an Associate Professor in the UC Berkeley Department of Psychology. His research interests include (a) visual and visuomotor localization, (b) motion perception, (c) object recognition, (d) perceptual and motor crowding, and (e) visual impairments. The Institute of Electrical and Electronics Engineers (IEEE) digital library houses peer-reviewed publications, conferences, and technology standards. The organization's overriding mission is to help set standards dedicated to advancing technological innovation and excellence. The article lists 35 references focusing on topics such as (a) scientific visualization, (b) information visualization valuation, (c) preattentive vision, (d) visual working memory, and (e) short-term memory.

Isenberg, P., Zuk, T., Collins, C., & Carpendale, S. (2008). Grounded evaluation of information visualizations. Retrieved November 4, 2012 from http://www.aviz.fr/~isenberg/publications/papers/Isenberg_2008_GEO.pdf.

Abstract. We introduce *grounded evaluation* as a process that attempts to ensure that the evaluation of an information visualization tool is situated within the context of its intended use. We discuss the process and scope of grounded evaluation in general, and then describe how qualitative inquiry may be a beneficial approach as part of this process. We advocate for increased attention to the field of qualitative inquiry early in the information visualization development life cycle, as it tries to achieve a richer understanding by using a more holistic approach considering the interplay between factors that influence visualizations, their development, and their use. We present three case studies in which we successfully used observational techniques to inform our understanding of the visual analytics process in groups, medical diagnostic reasoning, and visualization use among computational linguists.

Summary. This article focuses on the use of grounded evaluations regarding information visualizations. The goal of the article is to discuss the context of grounded evaluation in terms of the information visualization development life cycle, while focusing on the evaluation in the initial phase of information visualization development, the exploration of the problem space, and where grounded evaluation begins. To conclude, the authors provide three case studies about their experience with qualitative evaluations which form the first stage of grounded evaluation. The authors define grounded evaluation as a process that attempts to ensure that the context of information visualization uses grounding concepts, enabling further evaluations. The authors note that any form of information visualization evaluation should be grounded in the context in which they are used to assist users. Contexts can include (a) the size and complexity of data sets and tasks, (b) personal data exploration experience, (c) current stress level, (d) environmental distractions, (e) cognitive processing capabilities, and (f) analysis processes within a given work environment. They advocate an increase in attention to the field of qualitative inquiry, which tries to achieve a richer understanding by using a more holistic approach considering interplay between factors that influence visualizations, their development, and use. Qualitative research methods are listed which can help in gaining an understanding of factors influencing information visualization use and design: (a) action research, (b) ethnography, (c) conversation analysis, and (d) grounded theory. It is noted that gathering data for analytical/evaluation purposes relies heavily on coding. Coding is defined as the process of subdividing and labeling raw data, then reintegrating collected codes to form a theory. The authors note that grounded evaluations can, and should, take place during the development life cycle of information visualizations. It is concluded that qualitative evaluations are a powerful method which can provide useful design and evaluation criteria that should be applied throughout the life cycle of information visualization development.

Credibility. Petra Isenberg, Torre Zuk, and Sheelagh Carpendage were all affiliated with the University of Calgary at the time of writing. Petra Isenberg is a research scientist at INRIA in the AVIZ research group headed by Jean-Daniel Fekete. Her main research interests are information visualization, visual analytics, computer-supported cooperative work, and human-computer interaction. She has a PhD in Computational Visualistics from the University of Calgary. Torre Zuk and Sheelagh Carpendale were affiliated with the Department of Computer Science, while Christopher Collins was associated with the Department of Computer Science at the University of Toronto at the time of writing. Aviz is a multidisciplinary team of INRIA aiming at improving the analysis and visualization of large and complex data sets by combining analysis methods with interactive visualizations (Aviz, n.d.). The article lists 34 references focused on such topics as (a) qualitative data coding and analysis, (b) qualitative inquiry, (c) exploratory visual information analysis, (d) usability inspection methods, and (e) visualization analysis.

Kovalerchuk, B. (2003). Visualization and decision-making using structural information. doi: 10.1.1.133.6423.

Abstract. Visualization and computational intelligence methods such as data mining and knowledge discovery are important tools for decision support in imaging science. The goal of this paper is to discuss the blending of these areas to improve decision making for a variety of applications. The paper offers a conceptual model for integrating decision-making models, models for discovery of relations (computational intelligence models) and methods of visual correlation. Analysts use visual correlation for discovering relations and decision-makers use them for visualizing the decision under consideration. The model reflects this difference. The blending of these models is illustrated in the example of stopping cholera in London in the 19th century. The last part of the paper is devoted to systematization of visual correlation (VC) methods and their efficiency.

Summary. The goal of this article is to discuss the blending of data mining and knowledge discovery based on a blending of (a) decision-making model integration, (b) model creation for the discovery of information relations, and (c) methods of visual analysis based on structural information. Such blending is illustrated in the example of stopping cholera in London in the 19th century. The author notes that analysts use models of visual correlation for discovering relations and decision-makers use them for visualizing the decision under consideration. It is noted that mass media information visualizations typically do not provide graphic representations that are helpful for decision making. The two primary reasons for this are noted as (a) decision making is not a mass media goal and (b) "rich" information is actually scarce. The article goes into great detail describing the concepts of discovered relations/patterns (DRP) and a decision making model (DMM). A DMM is not formulated in terms of any preexisting relations. It is developed through the addition of new information objects to the DMM, which grows in a treelike fashion, creating a form of structured information. Once established, relationships can be analyzed, and even rearranged, supporting DRP. A detailed DRP can enable well-informed decision making.

Credibility. Boris Kovalerchuk is a Professor in the Department of Computer Science at Central Washington University. He holds a dual PhD in computer science and applied mathematics. Dr. Kovalerchuk has published a book, five book chapters and more than sixty papers in journals and conference proceedings. His research has been supported by grants from the (a) US National Research Council, (b) Advanced Research & Development Activity (ARDA), (c) National Imagery and Mapping Agency, (d) Office of Naval Research, (e) NASA, (f) International Science Foundation, (g) International Research and Exchange Board, (h) Austrian Cybernetics Society, (i) Commission of European Community, (j) Macarthur Foundation, and (k) S. Scott Cancer Center. The article was retrieved via Google Search. There are seven references in the article focusing on (a) performance visualization, (b) data mining, and (c) knowledge discovery.

Lam, H., Bertini, E., Isenberg, P., Plaisant, C., & Carpendale, S. (2011). Seven guiding scenarios for information visualization evaluation. Retrieved November 22, 2012 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.188.3308.

Abstract. We take a new, scenario based look at evaluation in information visualization. Our seven scenarios, evaluating visual data analysis and reasoning, evaluating user performance, evaluating user experience, evaluating environments and work practices, evaluating communication through visualization, automated evaluation of visualizations, and evaluating collaborative data analysis were derived through an extensive literature review of over 800 visualization publications. These scenarios are described through their goals, the types of questions they embody and illustrated through example studies. Through this broad survey and the distillation of these scenarios we make two contributions. One, we encapsulate the current practices in the information visualization research community and, two, we provide a different approach to reaching decisions about what might be the most effective evaluation of a given information visualization. For example, if the research goals or evaluative questions are known they can be used to map to specific scenarios, where practical existing examples can be considered for effective evaluation approaches.

Summary. The authors take a broad community-based approach to creating linkages between information visualization goals and evaluation approaches. The article takes a descriptive, as opposed to prescriptive, approach. The authors focus on *evaluation scenarios* based on a literature review of over 800 papers (345 with evaluations). A list of the seven most commonly encountered evaluation scenarios are discussed in great detail: (a) evaluating environments and work practices, (b) evaluating visual data analysis and reasoning, (c) evaluating communication through visualization, (d) evaluating collaborative data analysis, (e) evaluating user performance, (f) evaluating user experience, and (g) automated evaluation of visualizations. The goal of the authors work is to (a) encourage the selection of evaluation methods based on specific evaluation goals, (b) diversify the evaluation methods potentially used, and (c) act as a first step to develop a repository of examples and scenarios as a reference. Five differing stages of visualization development are identified including (a) predesign, (b) design, (c) prototype, (d) deployment, and (e) re-design. The authors make note that evaluation is not restricted to the analysis of specific visual representations. They encourage the focus of analysis to be on a visualizations role such as data analysis, or on specific environments to which visualizations might be applied. Related work focusing on evaluation taxonomies, systematic reviews, and evaluation methodologies are provided. The authors detail their methodology to derive the outlined scenarios: (a) compiling and evaluation dictionary, (b) open coding and tagging (based on 17 predefined tags), and (c) scenario development. The authors conclude by stating that their scenario approach can be used as a starting point for expanding the range of evaluation studies and opens new perspectives and insights on information visualization evaluation.

Credibility. The article is a Technical Report published by the Department of Computer Science, University of Calgary. Heidi Lam has a PhD in computer science from the University of British Columbia. Her thesis focuses on the visual exploratory analysis of large data sets. She currently works at Google and her research interests include information visualization and human-computer interaction. Enrico Bertini is a PhD in computer engineering and is Assistant Professor of Computer Science and Engineering at NYU Poly. His research focuses on information visualization, human-computer interaction, and visual analytics. Petra Isenberg is a research scientist at INRIA in the AVIZ research group headed by Jean-Daniel Fekete. Her main research interests are information visualization, visual analytics, computer-supported cooperative work, and human-computer interaction. She has a PhD in Computational Visualistics from the University of Calgary. Catherine Plaisant holds a PhD and is Associate Director of Research of the Human-Computer Interaction Lab at the University of Maryland Institute for Advanced Computer Studies. She has over 100 technical publications on subjects such as information visualization, digital libraries, universal access, image browsing, and evaluation methodologies. Sheelagh Carpendale was affiliated with the Department of Computer Science at the University of Calgary. The articles lists 94 references including topics such as (a) information visualization evaluation methods, (b) information visualization exploratory analysis, (c) asynchronous collaborative information visualization, (d) visualization tools, and (e) data analysis.

Li, T., Feng, S., & Li, L. (2001). Information visualization for intelligent decision support systems. *Knowledge-Based Systems, 14*(5-6), 259-262.

Abstract. To work efficiently with decision support systems (DSS), most users benefit from representation conversion, i.e. translating the specific outcome from the DSS, normally portrayed in a numerical format, into the universal language of the visual. In general, interpretation of data is much more intuitive if the results from the DSS are translated into charts, maps, and other graphical displays because visualization exploits our natural ability to recognize and understand visual patterns. In this paper we discuss the concept of visualization user interface (VUI) for DSS. A proprietary software system known as AniGraftool is introduced as an example of an information visualization application for DSS. In addition, a visualized information retrieval engine based on fuzzy control is proposed.

Summary. This article discusses the use of information visualizations in designing effective decision support systems. The authors note that traditional information presentation is geared towards mathematically inclined users, and leveraged numerical presentations. It is noted that information visualizations allow users to extract useful information from complex and/or voluminous data sets. They state that such visualizations are not limited to the graphical display of data, but now encompass a broader spectrum, including the design of graphical interfaces used to input and access such data (interaction). The authors provide a process flow of the "visualization pipeline", consisting of five stages: (a) simulating, (b) preparing, (c) mapping, (d) rendering, and (e) interpreting an information visualization, which represents the process of presenting scientific data.

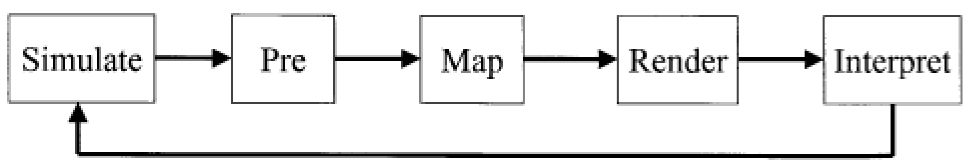

Figure 22. **Visualization pipeline.** Demonstrates a circular approach to information visualization the decision support systems must enable to provide efficient and effective decision-making.

A proprietary software system known as AniGraftool is introduced as an example of an information visualization application for DSS that addresses each stage of the visualization pipeline.

Credibility. At the time of publication, Tong Li and Shan Feng were affiliated with the Department of Automatic Control Engineering at the Huazhong University of Science and Technology in China. Xia Li was affiliated with both the Department of Information

Systems and Decision Sciences and National Key Laboratory for Manufacturing Systems Engineering. The article is published in Elsevier, a global leader in medical and health information (Elsevier, 2013). The article lists 10 references, including such topics such as (a) scientific visualization, (b) situation assessment and prediction in intelligence domains, (c) visualization and decision support, (d) visual thinking, and (e) visualization toolkits.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. Retrieved November 5, 2012 from http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation.

Abstract. This report seeks to understand the state of digital data, how different domains can use large datasets to create value, the potential value across stakeholders, and the implications for the leaders of private sector companies and public sector organizations, as well as for policy makers. We have supplemented our analysis of big data as a whole with a detailed examination of five domains (health care in the United States, the public sector in Europe, retail in the United States, and manufacturing and personal location data globally). This research by no means represents the final word on big data; instead, we see it as a beginning. We fully anticipate that this is a story that will continue to evolve as technologies and techniques using big data develop and data, their uses, and their economic benefits grow (alongside associated challenges and risks).

Summary. The goals of the article are to (a) understand the state of digital data, (b) determine how different domains can use large datasets to create value, (c) determine the

potential value across stakeholders, and (d) identify the implications for the leaders of private sector companies and public sector organizations, as well as policy makers. The author's note that the scale and scope of changes that big data brings are greatly expanding as technology accelerates data collection and accessibility. The means to extract insight from data are markedly improving as software available to apply increasingly sophisticated techniques combines with growing computing power, however, as of 2010, more than 4 billion people, or 60 percent of the world's population, were using mobile phones, and about 12 percent of those people had smartphones. More than 30 million networked sensor nodes are now present in the transportation, automotive, industrial, utilities, and retail sectors. The number of these data sensors is increasing at a rate of more than 30 percent a year. Throughout the paper, the authors discuss four primary challenges that should be addressed in the wake of big data: (a) a shortage of analytical and managerial talent necessary to make the most of big data, (b) the need to ensure that the right infrastructure is in place and that incentives and competition are in place to encourage continued innovation, (c) that the economic benefits to users, organizations, and the economy are properly understood, and (d) that safeguards are in place to address public concerns about big data. The authors identify and discuss five ways that big data creates value for those capable of managing its multiple needs: (a) creating transparency across the organization, (b) enabling experimentation to discover needs, expose variability, and improve performance, (c) segmenting populations to customize actions, (d) replacing/supporting human decision-making with automated algorithms, and (e) innovating new business models, products, and services. A total of twenty five data analysis techniques are listed and defined. The techniques most applicable in creating information visualizations include (a) classification, (b) cluster analysis, (c) data mining, (d) pattern recognition, (e) spatial analysis, and (f) time series analysis. For a full review of data analysis techniques, please refer to the article. Regarding visualization techniques, the authors demonstrate and explain modern methods such as (a) tag clouds, (b) clustergrams, (c) history flow, and (d) spatial information flow.

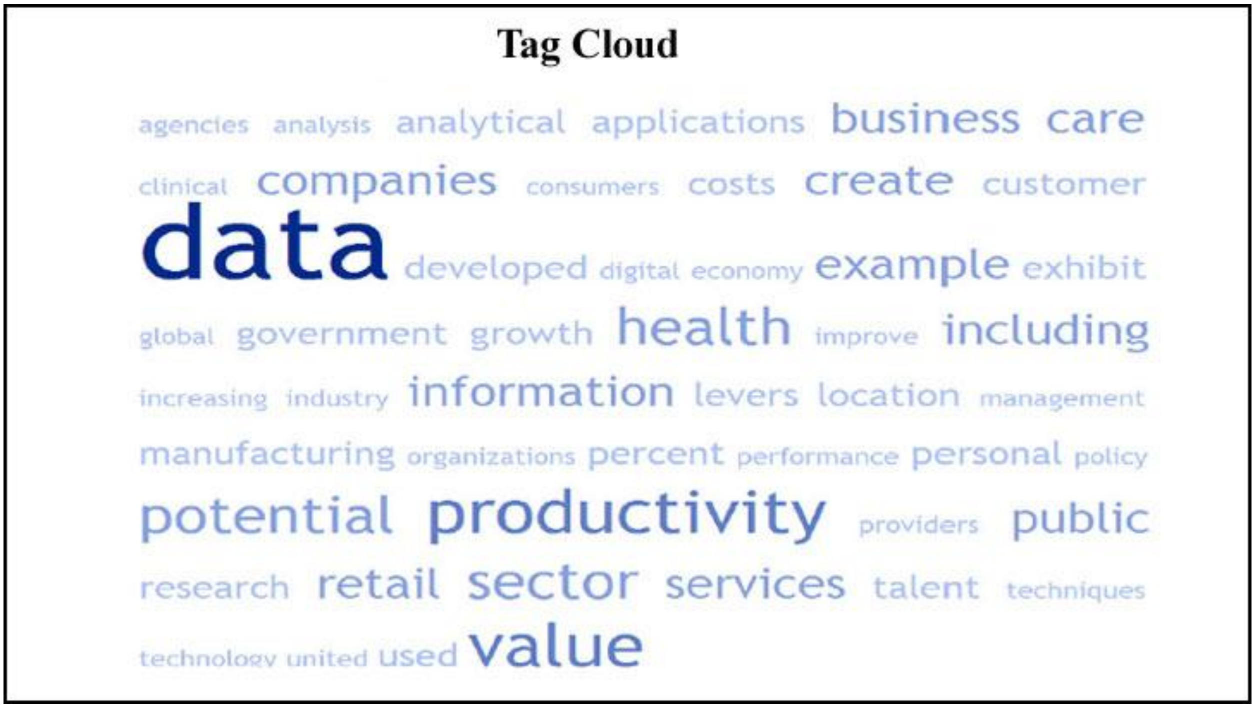

***Figure 23*. Tag cloud.** A visualization of text in which words or concepts that appear most frequently are displayed in a larger font size than words that appear less frequently (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Byers, 2011, p. 34).

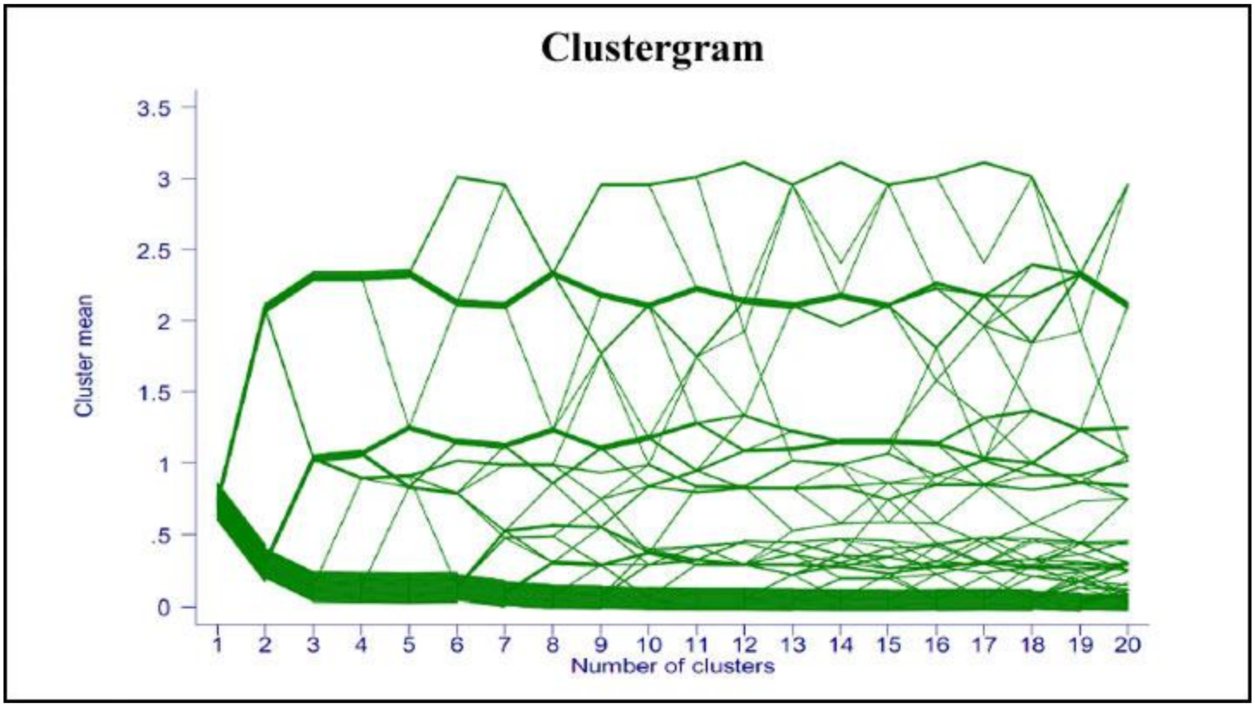

Figure 24. **Clustergram.** A visualization technique used for cluster analysis displaying how individual members of a data set are assigned to clusters as the number of clusters increases (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Byers, 2011, p. 34).

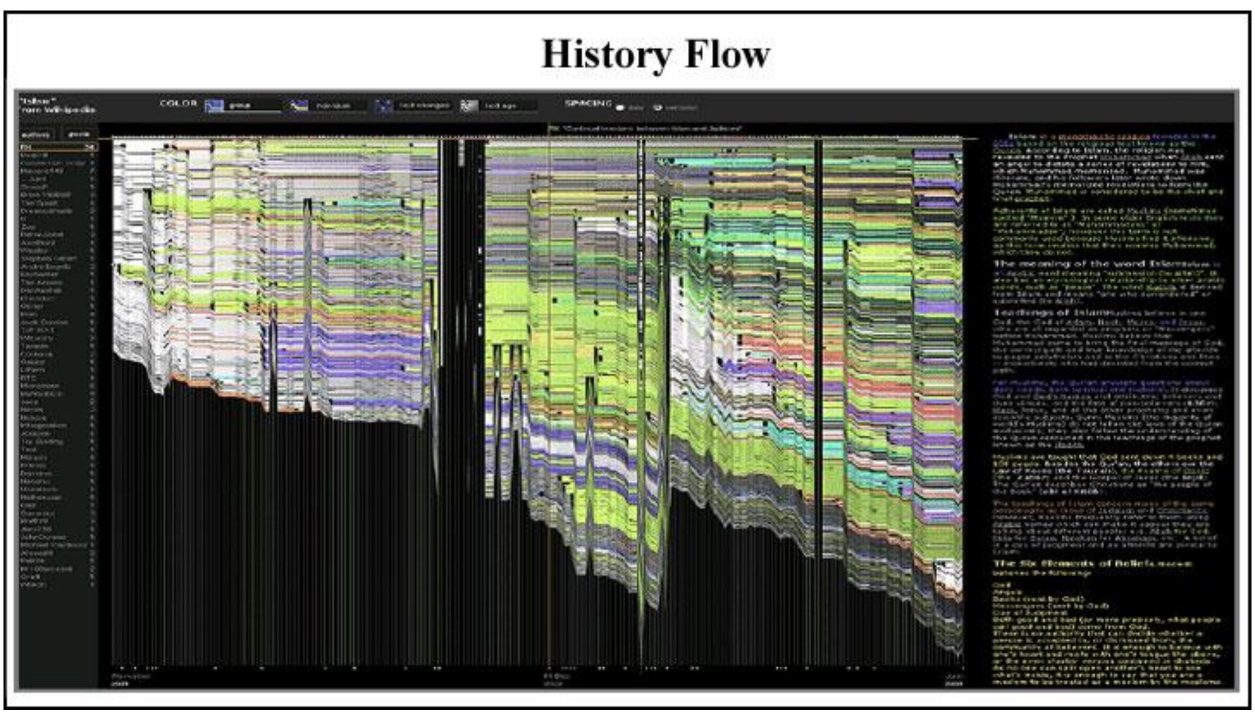

Figure 25. **History flow.** A visualization technique that charts the evolution of a document as it is edited by multiple contributing authors (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Byers, 2011, p. 35).

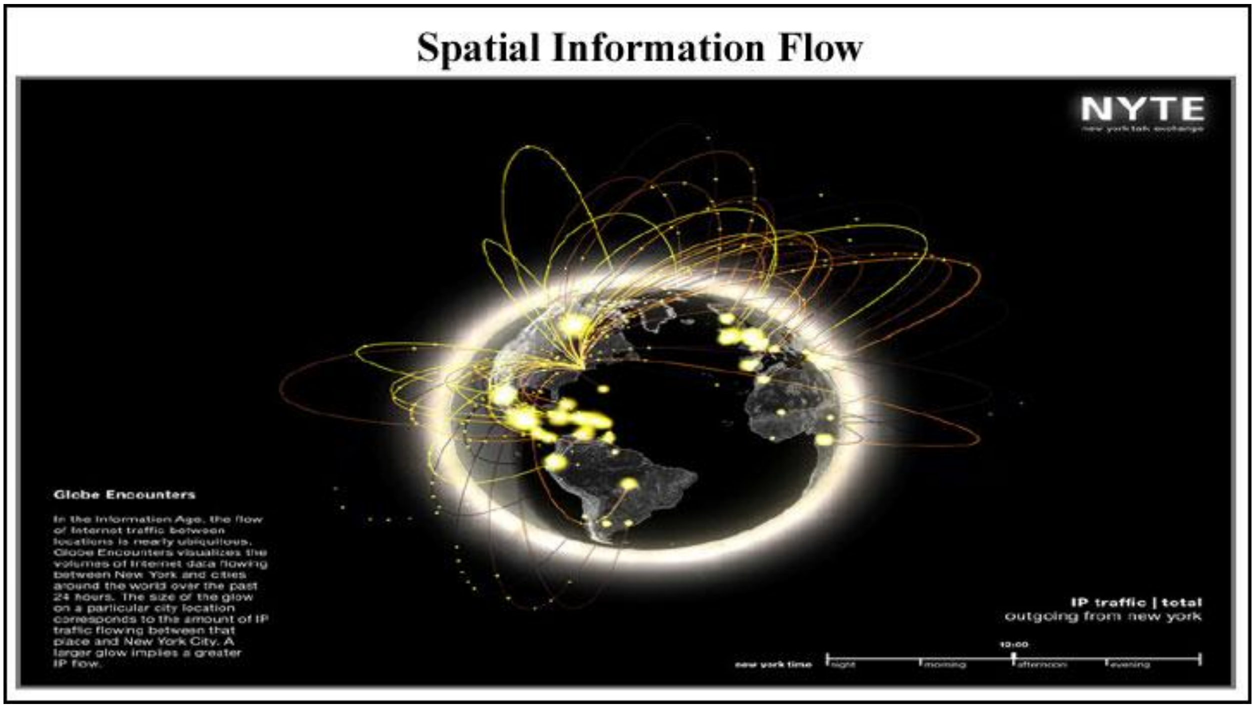

Figure 26. **Spatial information flow.** A visualization technique that depicts spatial information flow between locations (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Byers, 2011, p. 36).

Credibility. James Manyika is a PhD and director for the McKinsey Global Institute (MGI) business and economics research arm. He is also a senior partner and has lead research on areas including (a) growth, (b) productivity, (c) competitiveness, (d) labor markets, and (e) technology. In 2011, Dr. Manyika was appointed by the US Secretary of Commerce to serve on a 15-member innovation advisory board to advise the Secretary and report to Congress on US economic competitiveness and innovation. Michael Chui is a PhD and principal of the McKinsey Global Institute within the business and

economics arm. He leads research on the impact of information technologies and innovation on business, the economy, and society. Brad Brown leads McKinsey Global Institutes Insurance Operations and Technology Practice in North America. Brown's primary responsibilities include working with corporations to restructure their operations, IT organizations, and systems plans. Jacques Bughin is a PhD, Senior Partner and Director with McKinsey Global Institute. He is a core leader of the Telecom/media/ high tech practice as well as strategy practices. He has published more than 50 articles in international journals such as Management Science and the Journal of Industrial Economics. Richard Dobbs is a Director with McKinsey Global Institute in the business and economics research arm. While at MGI, he has led research on global economic trends including (a) urbanization, (b) resource markets, (c) capital markets, and (d) productivity and growth, with a focus on Asia. Charles Roxburgh is a Director of the McKinsey Global Institute's business and economics research arm. His work at MGI has focused on global capital markets trends, including (a) debt and deleveraging, and (b) how sovereign wealth funds, hedge funds, and private equity have fared in the financial crisis. Angel Hung Byers was also affiliated with MGI, however specific details are not available. There are 114 references listed focusing on topics such as (a) data mining association rules between sets of items in large databases, (b) big data, (c) machine learning, and (d) data processing.

Plaisant, C. (2004). The challenge of information visualization evaluation. Retrieved November 11, 2012 from http://delivery.acm.org/10.1145/990000/989880/p109-plaisant.pdf?ip=208.100.40.42&acc=ACTIVE%20SERVICE&CFID=173238642&CFTOKEN=76074580&__acm__=1358947059_a3e9c7323f2a86b67300b4122816b4ae.

Abstract. As the field of information visualization matures, the tools and ideas described in our research publications are reaching users. The reports of usability studies and controlled experiments are helpful to understand the potential and limitations of our tools, but we need to consider other evaluation approaches that take into account the long exploratory nature of user's tasks, the value of potential discoveries or the benefits of overall awareness. We need better metrics and benchmark repositories to compare tools, and we should also seek reports of successful adoption and demonstrated utility.

Summary. This article summarizes current information visualization evaluation practices, reviews challenges specific to information visualization, proposes initial steps such as the development of benchmarks and repositories, refined evaluation methodologies and toolkits, and dissemination of success stories. The author notes that evaluation approaches, which take into account the long exploratory nature of user tasks and potential discoveries or benefits of overall awareness, need to be considered. Plaisant states that information visualization has demonstrated faster task completion in laboratory settings, but the challenge appears to be convincing business managers of the potential value in an applied domain. Current evaluation practices included in this summary are (a) controlled experiments comparing design elements, (b) usability evaluation, (c) controlled experiments comparing two or more tools, and (d) case studies. Based on such approaches, it is noted that providing evidence of information visualization usefulness is difficult and presents specific challenges. Plaisant argues that the effectiveness of visualizations are partially determined by (a) looking at the same data from different perspectives over a long period of time, (b) answering questions you didn't know you had, and (c) factoring in the chances of discovery and the benefits of awareness. In order to improve information visualization evaluation and facilitate adoption, the author suggests (a) repositories of data and tasks, (b) case studies and success stories, and (c) the strengthening of the role of visualization toolkits. In conclusion, the author suggests that information visualization solutions must be carefully integrated into visualization tools that solve real problems in order to increase adoption in commercial settings.

Credibility. Catherine Plaisant holds a PhD and is Associate Director of Research of the Human-Computer Interaction Lab at the University of Maryland Institute for Advanced Computer Studies. She has over 100 technical publications on subjects such as information visualization, digital libraries, universal access, image browsing, and evaluation methodologies. . The ACM digital library represents the world's largest educational and scientific computing society, housing peer reviewed journals, magazines, conferences, workshops, and electronic forums. There are 38 references listed in the article covering topics such as (a) information visualization, (b) evaluation of visual interface effectiveness, (c) human-computer interaction, (d) interactive data mining, and (e) data exploration utilizing dynamic queries.

Somervell, D., McCrickard, S., North, C., & Shukla, M. (2002). An evaluation of information visualization in attention-limited environments. doi: 10.1.1.104.4322.

Abstract. People often need to quickly access or maintain awareness of secondary information while busy with other primary tasks. Information visualizations provide rapid, effective access to information, but are generally designed to be examined by users as the primary focus of their attention. The goal of this research is to discover how to design information visualizations intended for the periphery and to understand how quickly and effectively people can interpret information visualizations while they are busily performing other tasks. We evaluated how several factors of a visualization (visual density, presence time, and secondary task type) impact people's abilities to continue with a primary task and to complete secondary tasks related to the visualization. Our results suggest that, with relaxed time pressure, reduced visual information density and a single well-defined secondary task, people can effectively interpret visualizations with minimal distraction to their primary task.

Summary. The goal of this article is to explore the use of information visualizations as secondary displays and discover how to design visualizations intended for users who are performing other tasks preoccupying their attention. The authors attempt to answer two primary questions: (a) how quickly and effectively can people interpret an information visualization while performing other tasks and (b) how can peripheral visualizations be designed to reduce distraction while maintaining awareness? Several factors of visualization were evaluated including (a) visual density, (b) presence time, and (c) secondary task type. The authors note that most information visualization evaluations focus on situations in which a user explores the information as their only task. The authors hypothesize that peripheral visualizations will have some benefit in terms of user performance in assimilating information; however they also expect that peripheral visualizations will need to be significantly different from the content primarily engaging user attention. In order understand how people interact with peripheral visualizations in attention-limited environments, the authors conduct a pilot study in which participants play a video game (primary task) while simultaneously attempting to interpret an information visualization (secondary task). Regarding the information visualization during each round of testing, all participants are challenged to (a) identify the quadrant of the visualization containing an object of certain color and shape, (b) count the occurrences of certain objects, (c) determine the pattern formed by the certain objects, and (d) identify the quadrant containing a cluster of certain objects. The authors list the most interesting findings as (a) peripheral visualizations can be introduced without hindering primary task performance; (b) interpreting complex visualizations within one second in a dual-task scenario cannot be done effectively, but with relaxed time constraints can be very effective, (c) lower density displays can result in performance that is as good or better than high density displays in a dual-task scenario, and (d) finding clusters of visually similar items is easier than locating a single item. The authors state that a better understanding of the effects of visualizations as secondary displays will impact the increasing development of desktop information management tools.

Credibility. At the time of writing, all authors were affiliated with the Department of Computer Science at Virginia Polytechnic Institute and State University. Scott McCrickard is an Associate Professor with research interests in (a) human-computer interaction, (b) design methods in human-computer interaction, (c) user and task modeling, (d) notification systems, (e) interfaces for mobile and ubiquitous computing devices, and (f) peripheral and secondary displays. Chris North is an Associate Professor with primary research interests in (a) information visualization and evaluation, (b) visual analytics, and (c) human-computer interaction. His applied research interests include (a) intelligence analysis, and (b) cyber security. Maulik Shukla is a graduate student with six publications. Detailed information for Jacob Somervell is not available. The article was retrieved via Google Search. The paper is published in the proceedings of the Joint

Eurographics IEEE TCVG Symposium on Visualization (VISSYM). There are 25 references focusing on such topics as (a) information exploration, (b) the psychology of vision, (c) high-speed visual estimation using preattentive processing, and (d) information processing.

Conclusion

This annotated bibliography presents 34 references including peer-reviewed journal articles, conference proceedings, and published books. Each reference is reviewed in a four-part annotation, including (a) a bibliographic citation in APA format, (b) an abstract pulled directly from the reference, (c) a description of the credibility of the reference, and (d) a summary of the content, as pertains to this study.

The sheer volume of data and information available today that require handling, processing, and analysis is growing at an exponential rate across industries (Fisher, DeLine, Czerwinski, & Drucker, 2012). In addition, the scale and scope of changes that big data brings are greatly expanding as technology accelerates structured and unstructured data collection; these changes provide increasing information accessibility and management challenges while attempting to derive meaning, enable sense-making and generate knowledge (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Byers, 2011). Decision support systems are typically used to attempt and support the communication of inherent meaning within large data sets (Li, Feng, & Li, 2001). The overall goal of information visualization, regardless of implementation environment, is to enable rapid cognition without the manual scanning of detailed data sets often found in traditional information presentation solutions such as proprietary or commercial decision support systems (Mouine, Lapalme, 2012).

The purpose of this study, described in detail in the Introduction section of this document, is to identify (a) information visualization techniques that can support the exploration, discovery, and crystallization of new knowledge based on visual recognition of hidden patterns and structures in big data (Offenhuber, 2010) and (b) evaluation criteria that can be used to determine the effectiveness of information visualization within a specific information domain. The primary audience for this annotated bibliography is software development teams challenged with designing and implementing information visualization software that handles large data sets. The Definitions section of this document includes a taxonomy (see Figure 1) designed to categorize high level information visualization terminology related to (a) implementation techniques, (b) analysis techniques, (c) data types and repositories, and (d) cognitive considerations. Additionally, Figure 16 within the Annotated Bibliography provides considerations regarding the categories of visualization techniques best suited for the overall complexity of the data set(s) under investigation.

References are categorized into four sections in the Annotated Bibliography section of this document, each one focusing on a specific aspect of information visualization: (a) information visualization interaction, navigation, and exploration; (b) information visualization discovery and practices; (c) information visualization design and implementation; and (d) information visualization decision support and effectiveness evaluation. As the concept map in Figure 27 denotes, this categorization scheme addresses topics which support each concept node. Each colored concept map *node grouping* identifies a specific Annotated Bibliography theme: (a) black nodes represent information visualization design and implementation, (b) green nodes represent information visualization interaction, navigation, and exploration, (c) orange nodes represent information visualization discovery and practices, and (d) purple nodes represent information visualization decision support and effectiveness evaluation.

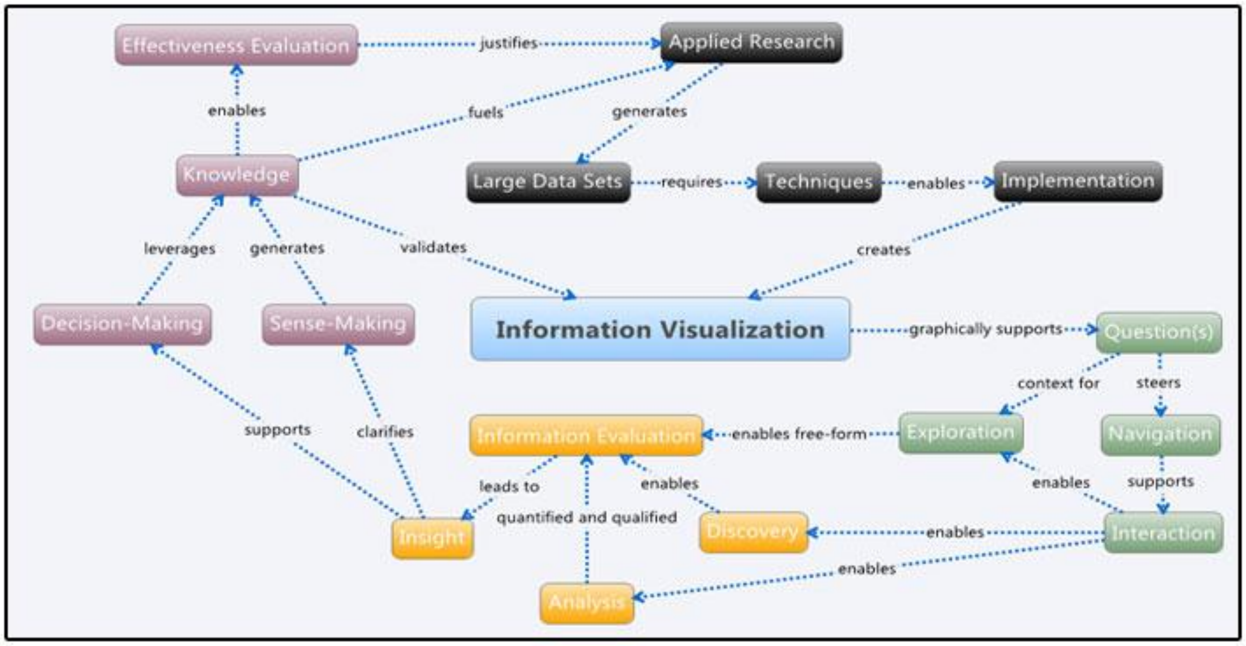

Figure 27. **Color coded information visualization concept map node groupings.**

By conceptually associating nodes in such groupings, the concept map demonstrates that theoretical and applied components of a graphical visualization or visualization system (a) can be analyzed or derived based on a linear conceptual-to-applied-to-evaluative flow, with each node grouping feeding a successive node grouping; (b) can be addressed in isolation, either theoretically or in an applied manner; and (c) can be viewed as a circular loop, ultimately enabling continued implementations based on applied, quantified, and qualified research. To better understand the importance of each node grouping, the concept map is broken out into isolated node grouping views, detailing the importance of each node within the associated grouping.

Theme 1: Information Visualization Interaction, Navigation, and Exploration

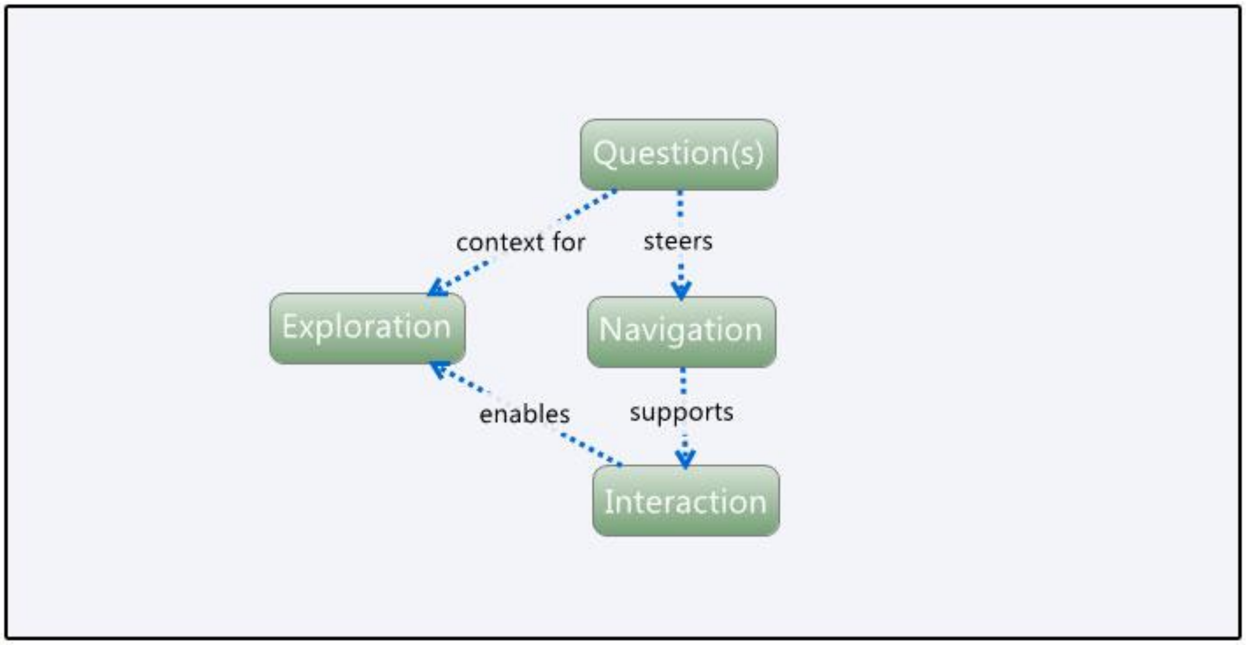

***Figure 28*. Node grouping detail (see Figure 27): Visualization interaction, navigation, and exploration concepts.**

Information visualizations set the context for data exploration. As noted by Offenhuber (2010), initial question(s) steer navigation (transversal) through a given data set, supporting exploration as well as providing a rhetorical role not available with table-based presentations. Questions also steer navigation functionality design, which in turn supports various forms of interaction, enabling visual information exploration in a non-scripted, free-flowing manner.

Large and often times complex data sets, typically defined by both structured and unstructured data, are frequently associated with equally complex problems (a) involving multiple decision makers, (b) requiring highly coordinated activities, (c) containing cognitive complexity, (d) spanning a broad range of technological applications, (e) requiring graphical validity, (f) residing within situational constraints, and (i) facing individual interpretation differences (Arias-Hernandez, Dill, Fisher, & Green, 2011). These environmental factors influence the types of questions users pose when interpreting, or deriving meaning from, a graphical visualization. When addressing such complex problems, Fisher, DeLine, Czerwinski, and Drucker (2012) stress three areas for consideration and suggest that users attend to (a) exploratory and demand-driven approaches, (b) clear communication in a succinct and actionable presentation form, (c) production of high confidence under tight deadlines, and (d) the need to preserve institutional memory.

According to Bresciani and Eppler (2009), it is widely hypothesized that interactive information visualizations optimally increase (a) productivity, (b) outcome quality, (c) learning, (d) satisfaction; and (e) participation, leading to better collaboration and question answering over static (non-interactive) implementations. Visualization navigation, interaction, and exploration serve to enhance cognitive recognition of hidden meaning. Seven widely-used categories of visualization interaction are noted by Yi, Kang, and Jacko (2007) for consideration to help reveal such meaning: (a) selection, (b) exploration, (c) reconfiguration, (d) encoding, (e) abstraction/elaboration, (f) filtering, and (g) connection. These interactive techniques allows users to iterate over data sets, changing views based on objective exploration, while enabling the personalization of a visualization according to personal preferences and specific user profiles (Mouine & Lapalme, 2012). Such personalization potentially enhances cognitive processing capabilities, increasing the effectiveness of visualizations during exploratory phases, and provides the ability to create clear use cases for information visualization implementation.

Theme 2: Information Visualization Discovery and Practices

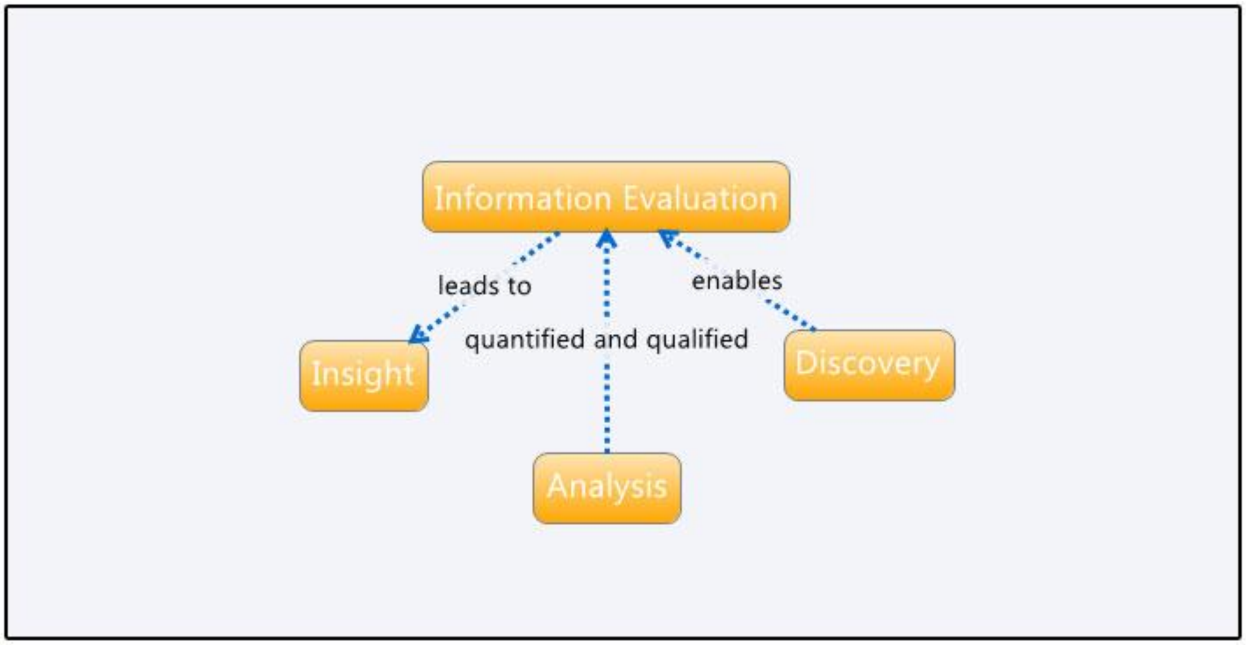

Figure 29. **Node grouping detail (see Figure 27): Visualization discovery and practices concepts.**

The discovery of new information and its analysis, based on both quantified and qualified criteria, leads to an evaluation of information visualization value in terms of generating novel or hidden insight. Gordon-Murnane (2012) identifies multiple sources contributing to the rapid increase in personal and organizational data volumes: (a) email, (b) searching, (c) browsing, (d) blogging, (e) tweeting, (f) buying, (g) sharing, (h) texting, and (i) networked sensors. Increasing volume, coupled with diverse data types, introduces numerous considerations that must be taken into account when creating optimal visualization solutions. Dadzie, Lanfranchi, and Petrelli (2009) identify such considerations as (a) targeting the end user's needs; (b) acknowledging the expertise the end user brings in their domain; and (c) the resources available for collecting, processing, and subsequently analyzing the requisite data. It is noted by Shrinivasan (2008) that analysis is often unsystematic, continuously emerging, and emergent. This poses significant challenges, especially for organizations that are feeling increasing pressure from a market perspective, to leverage large data in order to (a) spur new innovations, (b) spur new product opportunities, (c) achieve cost savings and efficiencies, and (d) use predictive analytics to understand customer behavior (Gordon-Murnane, 2012).

Evaluating inherent information value based on such seemingly ad-hoc analysis processes leaves little confidence in securing new insight. As interactive visualizations appear to best support analysis and subsequent insight, Dadzie, Lanfranchi, and Petrelli (2009) recommend that organizations deploy cycles of visualization application design fueled by validation from end users based on functional prototypes. They state that such development methodologies best support visualization software development that allows development teams to capture user feedback, track activity, and log interaction tendencies throughout the visualization interaction life cycle. Such immersive development cycles increase the chances of ensuring users glean insight, deriving informational value from large and expensive data repositories.

Theme 3: Information Visualization Design and Implementation

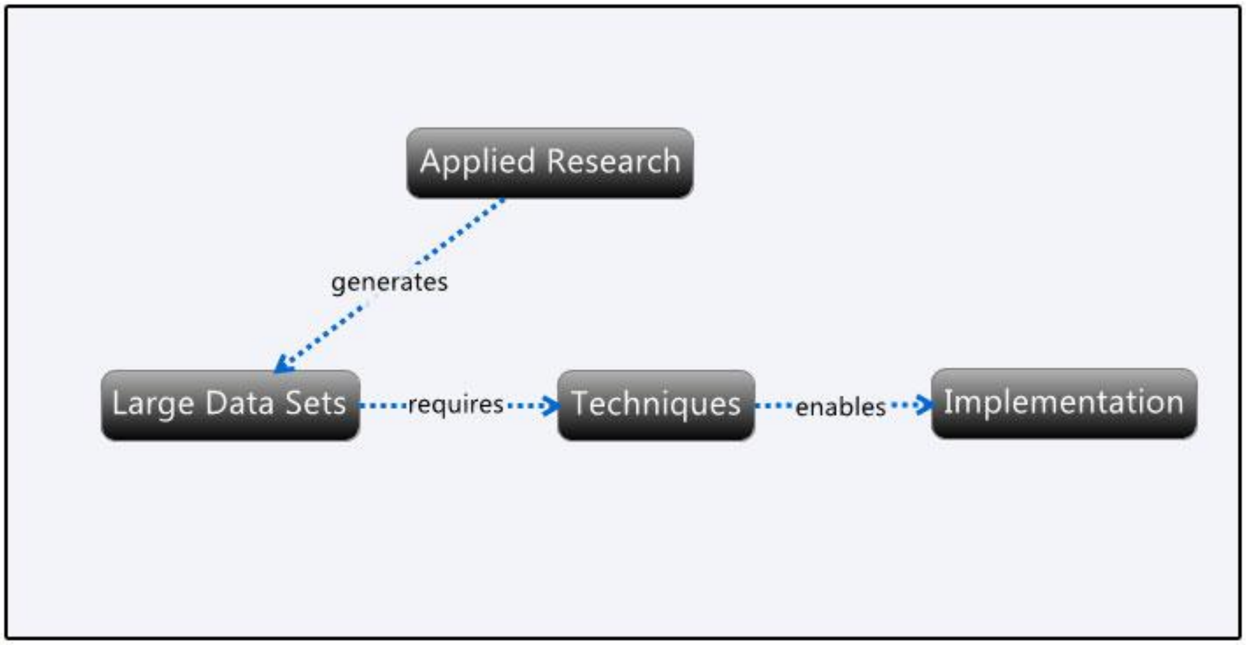

Figure 30. **Node grouping detail (see Figure 27): Visualization design and implementation concepts.**

Applied research in the field of information visualization collects vast quantities of data, providing large, domain specific data sets. Visualization techniques are used to generate inclusive visualization implementations that allow users to rapidly assimilate data in a cognitive manner. Gaviria (2008) decrees that functional information visualizations should aim to convey a message or delineate patterns hidden in the represented data through metaphors that users can quickly understand. Zhang, Johnson, Malin, and Smith (2002) highlight inherent visualization features for consideration, that make information visualizations extremely useful: (a) they provide short-term or long-term memory aids, (b) they provide information that can be directly perceived and utilized with little cognitive processing, (c) they provide knowledge and skills that are unavailable from internal representations, (e) they support perceptual objects that enable easy feature recognition and inference generation, (f) they anchor and structure cognitive behavior without conscious awareness, (g) they change the nature of a task by generating more efficient action sequences, (h) they aid processibility by limiting abstractions, and (i) they determine decision making strategies through accuracy maximization and effort minimization.

Bermudez et al., (2005) note that the current inadequacies of information visualization techniques and implementations appear to be attributed to (a) the persistence of early 20th-century quantitative methods, (b) a naïve understanding of human cognition, and (c) the use of simplistic representation spaces that are inadequate to address the complexity of modern day information environments. As Conti, Ahamad, and Stasko (2005) point out, such naivety when implementing information visualizations, either intentional or unintentional, poses significant threats to effective visualization processing and comprehension affecting (a) human memory, (b) human attention, (c) human visual perception, and (d) human motor resources.

Fekete and Plaisant (2002) stress that further research is needed to not only overcome such cognitive processing challenges, but to push the limits of visualization scalability (size of visible items) beyond current limitations of roughly 10,000 objects, accommodating exponentially growing data sets and smaller screen sizes associated with smart phones and tablets. As society moves to increasingly mobile and *socially aware technological solutions* (devices that can track social interactions, location, habits, etc.), access to such disparate data in an easily digested manner, at any time, becomes increasingly valuable. Fischer (2012) points out that when designing visualizations and visualization systems to support such technological domains, development teams should consider (a) filter bubbles and groupthink, (b) making information relevant to the task at hand versus serendipity, (c) intrusiveness, (d) remembering and forgetting, and (e) privacy.

Theme 4: Information Visualization Decision Support and Evaluation Effectiveness

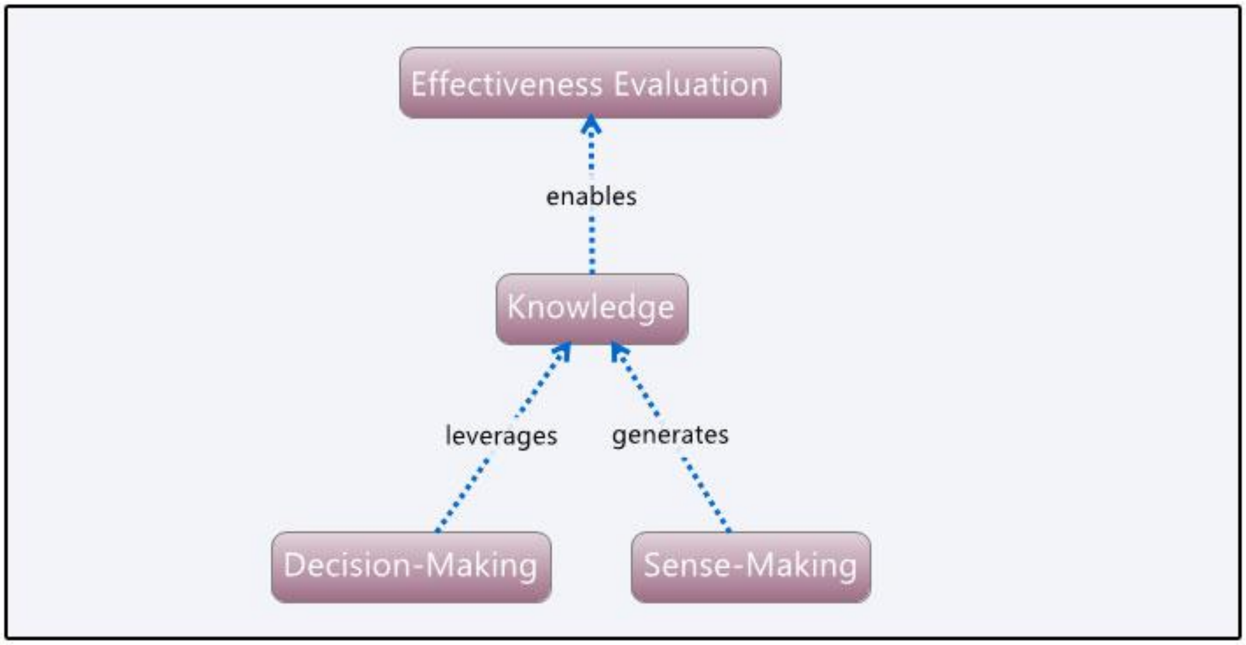

Figure 31. **Node grouping detail (see Figure 27): Visualization decision support and effectiveness evaluation concepts.**

An evaluation of information visualization effectiveness is based on its ability to reveal potentially undiscovered knowledge. Such knowledge is generated through the visualizations ability to provide improved decision-making and aid in sense-making. When designing a visualization or visualization system, Ahokas (2008) recommends keeping several design concepts in mind: (a) color coding, (b) scaling, (c) data set ordering, (d) small multiples, and (e) visualization interaction. The author also recommends the following guidelines when creating any sort of visualization dashboard: (a) provide a context to ground the user so they are not led astray accidentally, (b) avoid distracting or confusing graphic perspectives such as three-dimensional (3D) effects that can occlude data points, (c) provide interactive visualizations that support questions or hypothesis, and (d) weigh the value of using either graphs or tables based on the number of data points and communication intention of the unique visualization(s). These design concepts and guidelines should help alleviate potential cognitive confusion and ambiguity.

When developing visualization systems, Lam, Bertini, Isenberg, Plaisant, and Carpendale (2011) identify five linear stages: (a) predesign, (b) design, (c) prototype, (d) deployment, and (e) re-design. This linear systems design approach aligns closely with the linear visualization pipeline approach (see Figure 22) presented by Li, Feng, and Li (2001).

Lam, Bertini, Isenberg, Plaisant, and Carpendale (2011) list seven of the most commonly used evaluation scenarios: (a) evaluating environments and work practices, (b) evaluating visual data analysis and reasoning, (c) evaluating communication through visualization, (d) evaluating collaborative data analysis, (e) evaluating user performance, (f) evaluating user experience, and (g) automated evaluation of visualizations. When deriving value and determining the effectiveness of a visualization, Fabrikant and Buttenfield (2001) point out that empirical testing indicates that the generation of information visualizations, which consider and address the potential pitfalls noted by Ahokas (2008), substantially helps people access data in online archives more effectively than traditional text-based representations containing alpha-numerically scrolling lists, keyword query boxes, and similar tools. In such a massive data environment, this reinforces the need to create visualizations which accommodate representations that exceed 10,000 objects (Fekete & Plaisant, 2002).

According to Forsell (2012) *heuristic evaluation* provides the ideal solution to determining visualization effectiveness as such evaluation provides next-generation development to leverage proven success with greater confidence and possible industry, academic, or public adoption. Forsell provides a set of nine proven heuristic criterion for any visualization: (a) visibility of system status, (b) match between the system and the real world, (c) user control and freedom, (d) consistency and standards, (e) recognition rather than recall, (g) flexibility and efficiency, (h) aesthetic and minimalist design; (i) help users recognize, diagnose, and recover from errors, and (j) help and documentation functions.

In addition to heuristic evaluation, Isenberg, Zuk, Collins, and Carpendale (2008) suggest the use of *grounded evaluation*, in which a visualizations effectiveness is based on an implementation context such as (a) the size and complexity of data sets and tasks, (b) personal data exploration experience, (c) current stress level, (d) environmental distractions, (e) cognitive processing capabilities, and (f) analysis processes within a given work environment. Such evaluation considerations help address concerns researched by Haroz and Whitney (2012), involving a user's ability to effectively utilize a visualization within an environment that competes with the users attention to a given visualization. Haroz and Whitney (2012) conclude through multiple experiments that (a) grouping is far more beneficial for oddball search compared with known-target search, (b) accessing overall information (like heterogeneity or number of categories) is better for grouped displays, and (c) for difficult tasks, aim to reduce variety in the entire view rather than optimizing small regions.

Plaisant (2004) admits that information visualizations have demonstrated great potential in laboratory settings, but the true challenge appears to be convincing *visualization consumers* of their effectiveness and value in an applied domain. Plaisant argues that the effectiveness of visualizations is partially determined by (a) looking at the same data from different perspectives over a long period of time, (b) answering questions you didn't know you had, and (c) factoring in the chances of discovery and the benefits of awareness. To improve information visualization evaluations moving forward and facilitate widespread adoption, Plaisant suggests three proactive considerations: (a) maintaining collective repositories of data and tasks, (b) documenting case studies and success stories, and (c) strengthening of the role and applicability of visualization toolkits.

References

Ahokas, T. (2008). Information visualization in a business decision support system. Retrieved November 21, 2012, from https://helda.helsinki.fi/handle/10138/21417.

American Psychological Association. (2009). Publication manual of the American Psychological Association (6th ed.). Washington, DC: Author.

Arias-Hernandez, R., Dill, J., Fisher, B., & Green, T. (2011). Visual analytics and human-computer interaction. *Interactions, 18*(1), 51-55.

Arnold, E. (2012). The big deal about big data. *Information today*, *29*(9), 1-34.

Aviz, (n.d.). Visual Analytics Project. Retrieved January 18, 2012 from http://www.aviz.fr/.

Bermudez, J., Agutter, J., Foresti, S., Westenskow, D., & Syroid, N., et al. (2005). Between art, science and technology: Data representation architecture. *Leonardo*, *38*(4), 280-285.

Bihanic, D., & Polacsek, T. (2012). Models for visualization of complex information systems. In The French National Network of Complex Systems (Ed). *2012 16th International Conference on Information Visualization* (pp. 130). n.d.: The French National Network of Complex Systems. doi: 10.1109/IV.2012.

Boeri, R. (2012). Making sense of big social data. *EContent*, *35*(9), 29.

Bohn, R., & Short, J. (2010). How much information? 2009 Report on American consumer. Retrieved January 9, 2013 from http://hmi.ucsd.edu/pdf/HMI_2009_ConsumerReport_Dec9_2009.pdf.

Bresciani, S., & Eppler, M. (2009). The benefits of synchronous collaborative information visualization: Evidence from an experimental evaluation. Retrieved November 22, 2012 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.164.1010.

Brynjolfsson, E., Hitt, L., & Kim, H. (2011). Strength in numbers: How does data-driven decision making affect firm performance? Retrieved December 5, 2012 from http://executive.mit.edu/mysloan/documents/detail/?url=http%3A%2F%2Fexecutive.mit.edu%2Fresource%2Fdocuments%2FStrengthinNumbers.pdf.

Busch, C., De Maret, P., Flynn, T., Kellum, R., Le, S., Meyers, B., Saunders, M., & White, R. (2005). Content analysis. Writing@CSU. Colorado State University. Retrieved December 10, 2012 from http://writing.colostate.edu/guides/research/.

Center for Inclusive Design and Environmental Access (n.d.). Wayfinding. Retrieved January 20, 2013 from http://www.ap.buffalo.edu/idea/udny/section4-1c.htm.

Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM, 54*(8), 88-98. doi:10.1145/1978542.1978562.

Cognos Business Intelligence. Retrieved November 23, 2012 from http://www-142.ibm.com/software/products/us/en/business-intelligence/.

Conti, G., Ahamad, M., & Stasko, J. (2005). Attacking information visualization system usability overloading and deceiving the human. Retrieved December 12, 2012 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.111.772.

Creswell, J. W. (2009). Research design: Qualitative, quantitative, and mixed methods approaches. *Research questions and hypotheses* (pp. 129). Thousand Oaks, CA: Sage Publications, Inc.

Dadzie, A., Lanfranchi, V., & Petrelli, D. (2009). Seeing is believing: Linking data with knowledge. *Information visualization, 8*(3), 197-211.

Duchessi, P., & O'Keeffe, R. (1995). Evolutionary steps in expert systems projects. *Interfaces, 25*(5), 194-208.

Elkan, C. (2012). Predictive analytics and data mining. Retrieved November 24, 2012 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.164.416.

Elsevier. (2013). World class information. Retrieved January 22, 2013 from http://www.elsevier.com/about/mission/world-class-information.

Fabrikant, S., & Buttenfield, B. (2001). Formalizing semantic spaces for information access. *Annals of the Association of American Geographers, 91*(2), 263-280.

Fekete, J., & Plaisant, C. (2002). Interactive information visualization of a million items. Retrieved November 23, 2012 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.58.2019.

Fekete, J., Vanwijk, J., Stasko, J., & North, C. (2008). The value of information visualization. Retrieved November 20, 2012 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.140.2850.

Few, S. (2006). Clarifying the vision. In C. Wheeler (Eds.), *Information dashboard design: The effective visual communication of data* (pp. 6-7). Sebastopol, CA: O'Reilly.

Few, S. (2009). Information visualization. In N. Wishner (Eds.), *Now you see it: Simple visualization techniques for quantitative analysis* (pp. 11). Oakland, CA: Analytics Press.

Fischer, G. (2012). Context-aware systems: The 'right' information, at the 'right' time, in the 'right' place, in the 'right' way, to the 'right' person. Retrieved 12 November, 2012 from http://l3d.cs.colorado.edu/~gerhard/papers/2012/paper-AVI-context-aware.pdf.

Fisher, D., DeLine, R., Czerwinski, M., & Drucker, S. (2012). Interactions with big data analytics. Retrieved November 21, 2012 from http://dl.acm.org/citation.cfm?id=2168943.

Forsell, C. (2012). Evaluation in information visualization: Heuristic evaluation (Ed). *2012 16th International Conference on Information Visualization* (pp. 136).n.d.: IEEE. doi: 10.1109/IV.2012.33.

Fox, M., & Gruninger, M. (1998). Enterprise modeling. *American Association for Artificial Intelligence.na* (n.a.), 109. Retrieved November 12, 2012 from http://www.eil.utoronto.ca/enterprise-modelling/papers/fox-aimag98.pdf.

Gaviria, A. (2008). When is information visualization art?: Determining the critical criteria. *Leonardo, 41*(5), 479-482.

Grinold, R., & Kahn, R. (1992). Information analysis. *Journal of Portfolio Management, 18*(3), 14-21.

Gopalkrishnan, V., Guszcza, J., Lewis, H., & Steier, D. (2012). Big data, big business: Bridging the gap. Retrieved November 20, 2012 from http://www.deloitte.com/assets/Dcom-Singapore/Local%20Assets/Documents/Deloitte%20Analytics/2012/Big%20Data,%20Big%20Business%20Bridging%20the%20Gap%20-%20Deloitte.pdf.

Gordon-Murnane, L. (2012). Big data. *Online*, *36*(5), 30-34.

Goyvaerts, J. (2012). Regular expression basic syntax reference. Available from http://www.regular-expressions.info/reference.html.

Haroz, S. (n.d.). About me. Retrieved January 22, 2013 from http://steveharoz.com/.

Haroz, S., & Whitney, D. (2012). How capacity limits of attention influence information visualization effectiveness. *IEEE Transactions on Visualization and Computer Graphics, 18*(12).

Hayes, B., Heit, E., & Swendsen, H. (2010). Inductive reasoning. *WIREs Cognitive Science, 1*(na), 278.

Inmon, W. (1992). Building the data warehouse: Getting started. Retrieved November 25, 2012 from http://inmoncif.com/inmoncif-old/www/library/whiteprs/ttbuild.pdf.

Isenberg, P., Zuk, T., Collins, C., & Carpendale, S. (2008). Grounded evaluation of information visualizations. Retrieved November 4, 2012 from http://www.aviz.fr/~isenberg/publications/papers/Isenberg_2008_GEO.pdf.

J. Steinke (personal communication, November 7, 2012).

Keim, D. (2000). An introduction to information visualization techniques for exploring large databases. Retrieved November 12, 2012 form http://www.inf.ethz.ch/personal/peikert/SciVis/Literature/keim-Tutorial2000.pdf.

Keim, D. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, *7*(1), 1.

Khatchadourian, R. (2009). The taste makers. The secret world of the flavor factory. Retrieved November 20, 2012 from http://www.newyorker.com/reporting/2009/11/23/091123fa_fact_khatchadourian.

Kovalerchuk, B. (2003). Visualization and decision-making using structural information. doi: 10.1.1.133.6423.

Kreusler, M., Lopez, N., & Schumann, H. (2000). A scalable framework for information visualization. doi: 10.1.1.27.4514.

Lais, S. (2001). Treemaps bloom. *Computerworld*, *35*(27), 46.

Lam, H., Bertini, E., Isenberg, P., Plaisant, C., & Carpendale, S. (2011). Seven guiding scenarios for information visualization evaluation. Retrieved November 22, 2012 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.188.3308.

Lee, Y. T. (1999). Information modeling: From design to implementation. Retrieved November 4, 2012 from http://www.mel.nist.gov/msidlibrary/doc/tina99im.pdf.

Levy, Y., & Ellis, T. (2006). A systems approach to conduct an effective literature review in support of information systems research. *Informing Science Journal*, *9*(n.d.), 184.

Li, T., Feng, S., & Li, L. (2001). Information visualization for intelligent decision support systems. *Knowledge-Based Systems*, *14*(5-6), 259-262.

Lin, X. (1997). Map displays for information retrieval. *Journal of the American Society for Information Science*, *48*(1), 40-54.

Lu, Y., He, H., Zhao, H., Meng, W., & Yu, C. (2007). Annotating structured data of the deep web. Retrieved November 5, 2012 from http://www.cs.binghamton.edu/~meng/pub.d/paper279_Lu_ICDE07-Camera.pdf.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. Retrieved from http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation.

Marinai, R. (n.d.) Introduction to document analysis and recognition. Retrieved December 3, 2012 from http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CC0QFjAA&url=http%3A%2F%2Fciteseerx.ist.psu.edu%2Fviewdoc%2Fdownload%3Fdoi%3D10.1.1.157.4490%26rep%3Drep1%26type%3Dpdf&ei=khG9UI3tOcfr0QGuwYDIDA&usg=AFQjCNE9foyaAyKPt9pJRGENMnr5qtsFmA.

Menachemi, N., & Ginter, P. (n.d.). The anatomy of an empirical research manuscript suitable for the peer-reviewed health administration literature. Retrieved December 2, 2012 from http://www.ache.org/pubs/JHM/56-1/ResourceGuide-forAuthors.pdf.

Meyrick, J. (2006). What is good qualitative research? A first step towards a comprehensive approach to judging rigour/quality. *Journal of Health Psychology, 11*(5), 803.

Moreville, P., & Rosenfeld, L. (2006). *Information architecture for the world wide web.* Sebastopol, CA: O'Reilly Media, Inc.

Mouine, M., & Lapalme, G. (2012). Using clustering to personalize visualization (Ed). *2012 16th International Conference on Information Visualization* (pp. 258).n.d.: IEEE. doi: 10.1109/IV.2012.51.

Muller, R., Kovacs, P., Shilbach, J., & Eisenecker, U. (2011). Generative software visualization: Automatic generation of user-specific visualizations. Retrieved January 25, 2013 from http://141.44.32.131/iwde-2011/files/pdf/iwde2011_M%C3%BCller.pdf.

Nagel, H. (2006). Scientific visualization versus information visualization. doi: 10.1.1.164.319.

Novak, J. & Canas, A. (2006). The theory underlying concept maps and how to construct and use them. Retrieved January 4, 2012 from http://cmap.ihmc.us/publications/researchpapers/theorycmaps/theoryunderlyingconceptmaps.htm.

Offenhuber, D. (2010). Visual anecdote. *Leonardo, 43*(4), 367-374.

Pachikov, S. (2012). Evernote. Retrieved January 9, 2013 from http://evernote.com/.

Plaisant, C. (2004). The challenge of information visualization evaluation. Retrieved November 11, 2012 http://delivery.acm.org/10.1145/990000/989880/p109-

plaisant.pdf?ip=208.100.40.42&acc=ACTIVE%20SERVICE&CFID=173238642&CFTOKEN=76074580&__acm__=1358947059_a3e9c7323f2a86b67300b4122816b4ae.

Rasmussen, N., Bansal, M., Chen, C. (2009). Business dashboards: A visual catalog for design and deployment. Retrieved from http://www.orbis.eblib.com.libproxy.uoregon.edu/patron/FullRecord.aspx?p=427546&userid=D01nBugYM5g5rb9fcnAyGA%3d%3d&tstamp=1353355438&id=63561BB2895CE8DEE606589D269035DD675F122E&conl=uoregon.

Rouse, M. (2010). Cognos. *Definition.* November 18, 2012. Retrieved from http://searchcio.techtarget.com/definition/Cognos.

S. Mills (personal communication, April 10, 2010).

Schneiderman, B. (1998). Treemaps for space-constrained visualization hierarchies. Retrieved January 4, 2013 from http://www.cs.umd.edu/hcil/treemap-history/.

Shrinivasan, Y. & Wijk, J. (2008). Supporting the analytical reasoning process in information visualization. Retrieved November 21, 2012 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.184.2649.

Simon, M. (2011). *Assumptions, limitations, and delimitations.* (Doctoral dissertation). Retrieved December 13, 2012 from http://dissertationrecipes.com/wp-content/uploads/2011/04/AssumptionslimitationsdelimitationsX.pdf. (n.d.).

Simpson, S. (2011). A taste of the flavor company sector. Retrieved November 8, 2012 from http://seekingalpha.com/article/259259-a-taste-of-the-flavor-company-sector

Smith, M. K. (2003). Michael Polanyi and tacit knowledge. *The Encyclopedia of Informal Education.* Retrieved November 23, 2012 from www.infed.org/thinkers/polanyi.htm.

Smullen, C., Tarapore, S., & Gurumurthi, S. (2007). A benchmark suite for unstructured data processing. Retrieved January 4, 2012 from http://www.cs.virginia.edu/~gurumurthi/papers/snapi07.pdf.

Somervell, D., McCrickard, S., North, C., & Shukla, M. (2002). An evaluation of information visualization in attention-limited environments. doi: 10.1.1.104.4322.

Stone, M. (n.d.). Information visualization: challenge for the humanities. Retrieved December 1, 2012 from http://www.clir.org/pubs/resources/promoting-digital-scholarship-ii-clir-neh/stone11_11.pdf.

Trantow, B., & Schurl, K. (2006). Erzahlungen -35/65+ Zwel Generationen. Retrieved January 20, 2013 from http://www.andresramirezgaviria.com/texts/texts-katrin-bucher-trantow/.

Ware, C. (2012). Interacting with visualizations. In Dunkerley, M., & Scherer, H. (Eds.), *Information visualization: Perception for design* (pp. 345-374). Boston: Morgan Kaufmann.

Wisegeek. (2012). wiseGEEK.com. Retrieved November 26, 2012 from http://www.wisegeek.com/what-is-a-data-domain.htm.

Yen, J. (1998). Fuzzy logic – A modern perspective. Retrieved November 25, 2012 from http://dl.acm.org/citation.cfm?id=627980.doi: 10.1109/69.755624.

Yi, J., Kang, Y., Stasko, J. & Jacko, J. (2007). Toward a deeper understanding of the role of interaction in information visualization. Retrieved January 2, 2013 from http://www.cc.gatech.edu/~john.stasko/papers/infovis07-interaction.pdf.

Zahedi, F. (1991). An introduction to neural networks and a comparison with artificial intelligence and expert systems. *Interfaces, 21*(2), 25 – 38.

Zeng, L., Li, L., & Duan, L. (2012). Business intelligence in enterprise computing environment. *Information Technology & Management*, *13*(4), 297-310.

Zhang, J., Johnson, K., Malin, J. & Smith, J. (2002). Human-centered information visualization. Retrieved November 22, 2012 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.121.6064.