



Visual Analytics Lab (VAL)

2015

Visual analytics: The role of design and art in the emerging field of big data

Diamond, Sara

Suggested citation:

Diamond, Sara (2015) Visual analytics: The role of design and art in the emerging field of big data. Stream 1: Enterprise and Engagement. Available at <http://openresearch.ocadu.ca/id/eprint/1084/>

Open Research is a publicly accessible, curated repository for the preservation and dissemination of scholarly and creative output of the OCAD University community. Material in Open Research is open access and made available via the consent of the author and/or rights holder on a non-exclusive basis.

The OCAD University Library is committed to accessibility as outlined in the [Ontario Human Rights Code](#) and the [Accessibility for Ontarians with Disabilities Act \(AODA\)](#) and is working to improve accessibility of the Open Research Repository collection. If you require an accessible version of a repository item contact us at repository@ocadu.ca.

Visual Analytics – the Role of Design and Art in the Emerging Field of Big Data

Dr. Sara Diamond
OCAD University

Abstract

Driven by the increasing complexity of data sets the need for sophisticated analytics algorithms coupled with visualization of both data and information is growing exponentially in every discipline and industry. Artists, designers and visual thinkers have an important role to play in the presentation and interpretation of data. The Visual Analytics Lab (VAL) at OCAD University is a preeminent research lab for innovation and training in information and scientific visualization and visual analytics. As well as its perspective on the field, two brief case studies are provided, one for health care and the second for media navigation and analysis.

1. Visual Analytics and Big Data

Unstructured data such as social media provide new challenges in volume and quality (Aggarwal, 2012). The field of visual analytics addresses the growth of big data, a paradigm shift in which instruments and systems produce large quantities of data that is meaningless without extraction and analysis. Ware and Bobick (2005) explain that understanding data requires interaction between human cognition, computer memory and its related algorithms, and the physical actions of the user. Visual analytics supports a mix of cognitive and perceptual reasoning; it a practice that includes data science, cognitive science and visual and design practices and aesthetics. For these reasons it is a new interdisciplinary category of knowledge that draws from university research capacity yet challenges disciplinary and academic unit boundaries in the same moment. For example, the Centre for Information Visualization and Data Driven Design is a three-university network (York University, OCAD University, and University of Toronto) which includes scientists and designers and works with data sets from actual users from healthcare, earth science, communications and social science and culture³⁰. Users tend to either develop hypotheses from data or illustrate and test their theories with data that they extract (Diamond, 2010). Visual analytics thinking has previously separated the presentation of data from analytics, however these two approaches have recently moved into proximity because analysts need to explain their findings to non-analyst users including leaders and general audiences using sophisticated digital design techniques (Turner, 2012). Data analysts argue for the removal of the subjectivity that visual expression brings through visual simplicity, yet every aesthetic decision (to use a graph instead of a diagram) and convention carries cultural history and knowledge behind it (Diamond, 2010).

³⁰ See <http://www.civ-ddd.ca/> for information regarding research themes and researchers. The team collaborates with experts across the laboratory and around the world and in our immediate vicinity, such as York University; University of Toronto; Dalhousie University; University of British Columbia; Simon Fraser University and members of the CANVAC Visual Analytics Network across Canada; University of Bergen, Norway; University of Brasilia; UNICAMP; UFSCar, and the Brazilian BRAVA network.

A fundamental question is when and where the human fits in the loop. As systems become increasingly automatic it is critical for users to maintain histories, personalize their interfaces, understand the quality of data being used to build assumptions and provide a human layer of understanding. In response to the big data configuration several concepts emerge that recognize this complexity. “Deep data”, “is a combination of experts' domain knowledge of a subject with data points that confirm or contradict human understanding”. It is meant to, “make people and communities see *themselves*” (Raghavan, 2014). A similar concept is that of, “thick data”, suggesting the profound need for ethnographic as well as design-thinking lenses to be placed against the automatic results of big data analytics, because humans are unpredictable and contextually motivated (Wang, 2013). Deciding the degree to which creative processes can and should be automated requires a sophisticated understanding of humans in their diversity. For example, news media outlets such as our partner, the Globe and Mail, are increasingly relying on recommender engines that suggest content to their readers, and define the placement of content on the page relative to advertising. The fundamental question arises of where human editorial decisions then fit into this process.

As an art and design university OCAD University has engaged with visual analytics as an extension of information visualization and infographics. The visual component of data analytics is important because images create a bridge between the empirical world and the viewer, revealing patterns and unexpected relationships, hence evoking interpretation. Effective representations can allow the monitoring of simultaneous data sources. The visual can support comparative analysis, the recognition of anomalies or problems as these emerge and can support prediction. Visualizations effectively use human visual perception in order to navigate large quantities of data, compressing it into a screen space. Visualizations can provide a beautiful engaging experience for users that provoke new ways of interpreting data.

2. The Visual Analytics Laboratory at OCAD University

With the help of industry and Not-for-Profit partners, VAL offers students and research members an advanced, data secure environment to design novel visualizations and develop new visual analytical and visualization tools including using real-time visualization. VAL's philosophy is that actual data sets should be used within research in order to test hypotheses and methods and to ensure that outcomes are meaningful for user groups, or audiences. For these reasons it has a wide range of partners who are data holders or users and include public and private sectors. Industrial partners are wide-ranging, from Boeing Aircraft, the Globe and Mail (Canada's national newspaper), to N-Logic (the leading analyst of consumer trends in media consumption), IBM, Autodesk and visualization start-ups, long-term care facilities, transit authorities, cities, and energy conservation companies such as Zero Footprint.

OCAD University faculty are primarily designers, artists, digital media makers and cultural theorists. However, our team includes two experts in cognitive science as well as faculty who have statistical analysis, engineering and programming skills. We rely on partnerships with data scientists, and domain experts, whether academic or industrial, to complete our research and industrial development outputs.

Our sustainability relies on provincial support through the Centre for Information Visualization and Data Driven Design <http://www.civ-ddd.ca/> in partnership with York University and University of Toronto, and other Ontario grants, as well as Canadian research grants, foundation and direct industry funding. VAL's research and development activities focus on four main areas:

- **Visual Design:** Developing new ways to tell stories with data through unique visual representations and appropriate metaphors. Depiction of complex patterns and relationships through appealing visuals to enable effective information delivery. Experiments in mass personalization.
- **Analytic Methods:** Designing algorithms and software to guide users in exploring and gaining valuable information from visualizations; enabling users to create and test hypotheses, communicate results, and challenge assumptions.
- **Natural User Interactions:** Emerging multimodal techniques and hardware for developing natural interfaces and data representation/expression that brings users closer to their data. The team creates new interactive visual environments for the web, mobile devices, and desktop application; sonic, tactile and physical data expressions that allow collaboration and multi-sensory exploration of data.
- **User Experience:** VAL helps people work with information. The user-centered design approach used in the lab includes collaborating closely with users to understand their problems, test solutions, and deliver usable and useful software products. However, we also challenge concepts of user-centric design in an era of large data sets, looking at how to understand patterns that emerge from the data and might challenge users' assumptions.

The discussion now provides two case study examples of the VAL partnered research.

3. Case Study One: Extracting Data Sources in the Hospital environment: Specific Methodologies

With support from Boeing and Canada's MITACs³¹ internship funding, and through a partnership with the CAIS hospital in Brazil, the team set out to create a system that facilitates long-term care workers to input client data as they create it to mobile devices, primarily the tablet, but also smartphones. Through research and a literature review it became apparent that care givers

³¹ <https://www.mitacs.ca/en> is a significant source of industry- partnered internships.

need to be able to describe and respond to patient behavior, and then use visualizations to show patterns over time, to “enable a user to extract information necessary for rapid, actionable decision making in real-time on the site” (Pattath, 2010). Data analytics can support human memory by providing longitudinal data on professional interactions.

The principle of these tools is knowledge management (Eppler, 2011), and further, the support of “knowledge intensive conversations” (Mengis & Eppler, 2005, p. 4). The flexible, adaptive nature of conversations allows insight and innovation to occur and casual, yet valuable information to be captured. Viégas and Wattenberg use the notion “communication-minded visualization” (2006, p. 801) to describe online conversation tools that enable collaboration. Such tools must meet the fundamental needs of collaborative systems: balancing effort and benefit, engendering critical mass adoption through appropriate social context, supporting social processes and providing appropriate levels of privacy (Viégas & Wattenberg, p. 805-6). Effective tools bring together emotional and cognitive responses, creating stronger communities and collaborations (Bresciani, S., Tan, M. & Eppler, M.J. (2012). The goal was to sustain the existing strong bonds and caring in the long-term care environment, while creating a means to enhance condition analysis.

The team first undertook participant observation, a method used in social anthropology and ethnography, wherein researchers join the community they study participate in typical activities and observe its participants allows behavior that is not mediated by technology to be analyzed, including work flow, and then translated into digital tools. At a later stage the team shifted to participatory design methods first discussed by Schuler and Namioka (1993) to work closely with the staff in the hospital to define culturally appropriate visual interfaces to facilitate their entry and analysis of patient care on one hand and provide efficient support for work processes and flow. The intention was “use testing” (Greenberg & Buxton, 2008, p. 116) that is discovering whether an imagined tool is useful to a user group, a concept analogous to Nielson's (1993) view that technology should advance users' goals. Cockton (2008) argues that it is important to include “design purpose” in analyzing usability. The team then used a task based approach to developing interactive visualizations for data exploration and analysis, drawing from Pattath's (2011) deployment of application scenarios and tasks, after which visual analytic solutions are created and tested. Resulting tools run on small scale devices as well as larger displays.

The table below demonstrates the central findings regarding the qualities needed to successfully support team collaboration in long-term care using a tablet-based visual analytics solution.

The design points described above were first applied and tested through three low fidelity prototype designs – *Metas da Marta*, *Corpose* and *Procare* for 7” tablets as shown in following figure. *Metas da Marta* allowed multimodal data capture using multimedia controls like text, photos and videos, and provided a means for caregivers to collaborate in setting specific care goals and related tasks with each other, and then visualize the results of these collaborations

over time. *Corpose* was a natural user interface that allowed the user to photograph a resident and then sketch and notate treatments on the image, capturing change over time. In *Procure*, formal health record data could be captured and shared within the team through an optimized messaging system, visualized over a period of time and maintained using a tablet. It also allowed for emergency alerts³².

Key Point	Observation	Outcome
Multimodal data collection	In daily workflow, care is delivered based on formal health record data, recordings of treatments and their results and informal oral conversations about residents' condition.	Interface design makes use of audio, video, image, and text provided by the mobile device for data capturing.
Visual methods analytics	Caregivers analyze health records and personal notes, to create person-centric care and to conduct self-analysis of condition of each resident under their care. They require a means to analyze treatment results and the impacts on clients over time.	Visual analytics methods are applied to support immediate analysis of changes in residents as well as providing an analysis of change over time in relation to a care goal.
Support of collaboration through social media	Supportive family-care model is followed at CAIS for care delivery. Personal observations are exchanged through paper records and oral conversations in a timely manner with other caregivers.	Instant text messaging service for information sharing coincides with visual analytics results of care and condition system.
Interaction through natural user interfaces	Long-term care settings feature efficiency and freedom of movement during daily workflow. Apart from formal data entry, data capturing also includes note-taking and sketching on notepad.	Application of interactive techniques such as gesture, motion, and sketch, relevant to NUI design that aim to improve expressivity and intuitiveness of tablets and interfaces.
Optimized communication through interface	Caregivers interact continually with residents and colleagues in contextually-relevant ways throughout work.	Collaborative communication requires careful interaction design that includes sharing of information through social media component.

Table 1: Design Strategy

³² The challenges, design issues, feedback, and considerations of the five-point design strategy and the prototypes were presented at CANVAS 2013 in Halifax, at BRAVA in Brazil in May, 2013 and then in an intensive workshop with CAIS in the summer of 2013 and were summarized through a paper presented at the CSCWD 2014 conference.



Figure 1: Prototypes designs:
(a) *Metas da Marta* (b) *Corpose* and (c) *Procure*

Following our first observational work and preliminary prototype development we returned to the CAIS and led a workshop that reviewed the prototypes and gathered feedback for a second round of design. The second design, entitled the *Care and Condition Monitor*, was founded on a “care and condition” strategy, which appeared to be valuable to many long-term care contexts, not only that of the CAIS, and could potentially involve families or care recipients. It would enable workers to capture longitudinal data from informal data sources and to visually analyze the condition of residents.

Condition is measured by five core metrics of health and wellness – general health, morale, cognition, mobility and physical body structure. These metrics were extracted from CAIS medical records and correlate with criteria found in other support technologies and healthcare providers. Care, on the other hand, is individualized and organized around care goals tailored to needs of each resident and translated into action as treatments³³. User interaction to support the analysis and review of longitudinal data is provided through the ability to scrub back and forth in time in any visualization. This is important because there are multiple visualizations associated with the data. Finally, the process of creating visualization engages the caregivers, facilitating collaboration, as communication is integral to the cyclical relationship between a resident’s conditions based on care services. Condition, which is recorded by all caregivers, prompts the creation of new care goals, which are delivered by a team of caregivers, for improving condition.

The circular layout spatializes the social relationship between the resident and his/her circle of care. The individual care or condition metrics that the central circle is derived from are arranged around one side of the circle and the caregiver identified (for the sake of transparency). The state of each metric (green, yellow, red) is indicated around the opposite side. Curved lines join

³³ Although a work-in-progress, CCM illustrates and presents methods to collect and analyze informal and qualitative health care data in long-term care environments. Future steps include, development of a high-fidelity prototype for testing and evaluation of the proposed system. Detail description of methods used to design this tool is presented in conferences - IEEE Vis 2013, CSCWD 2014, MobileHCI 2014, IEEE Vis 2014 and workshops at CANVAS 2013, again at CANVAS 2014, and at the CAIS hospital.

the metrics, on one side, to the states, on the other. Examples of visualizations are shown below:

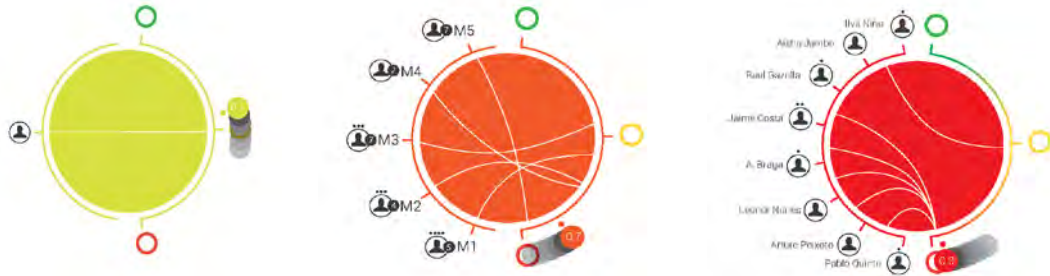


Figure 2: Care and Condition Monitor Interface Example

After initial user testing of the tool components of the interface are being refined and potentially applied to other health care data scenarios.

4. Case Study Two: Globe and Mail – Supporting the Digital Transition

In a partnership with the Globe and Mail the VAL continues to investigate the visualization of users' navigation through digital media properties. The team searched for a model that optimizes potential relationships between categories of knowledge rather than mimicking print media. The foundation for exploration is found not necessarily in the cohesion of content; but rather the fragmented, heterogeneous, organic connections between content (Bostock, Ogievetsky & Heer, 2011). Every topic or branch of knowledge contains concepts, language, and themes, which define the boundaries of a specific topic (Foucault, 2006). These boundaries define what Foucault refers to as *discourse*, and each discourse defines the way in which a topic can be discussed. Discourse is plural; it does not consist of one statement, text, or source (Hall & Foucault, 2001) Discursive Formations provide a model for thinking about relations found within the text of articles—based on a consistency between statements instead of keywords—that reduce a document down to one or more themes.

Our final conceptual foundation attempted to describe the space in which discourses meet—the underlying mechanisms at work in Foucault's *Discursive Formations*. Gilles Deleuze and Felix Guattari's concept of the *Rhizome* is a conceptual tool for characterizing a model of knowledge in which discourse is rhizomatic, meaning that it is non-hierarchical, shapeless, and comprised of further dimensions. To complicate things further, a discourse is itself but one dimension of a larger Rhizome. There are other properties of a Rhizome that are important to consider for a model of navigation; any point can connect to any other point irrespective of their commensurability or concomitance. It has no origin, no beginning nor end, and consists of multiple entries and exit points. A real world example of this pattern of navigation can be found in Wikipedia, whereby user defined

hyperlinks embedded in wiki pages can connect to any other wiki page.

Inspired by the models of Discursive Formations and the Rhizome the team sketched an exploratory interface specifically based upon the navigation of news media in digital format. Discursive Formations are the result of the overlap and regularity between discourses. A discourse is plural and does not consist of one statement, text, or source. This idea is reflected in how we aggregate and relate data through this sketch. To complement the associations we use the concept of the Rhizome to structure the connections between nodes; namely the Rhizomes ability and nature to make both concrete and abstract associative connections.

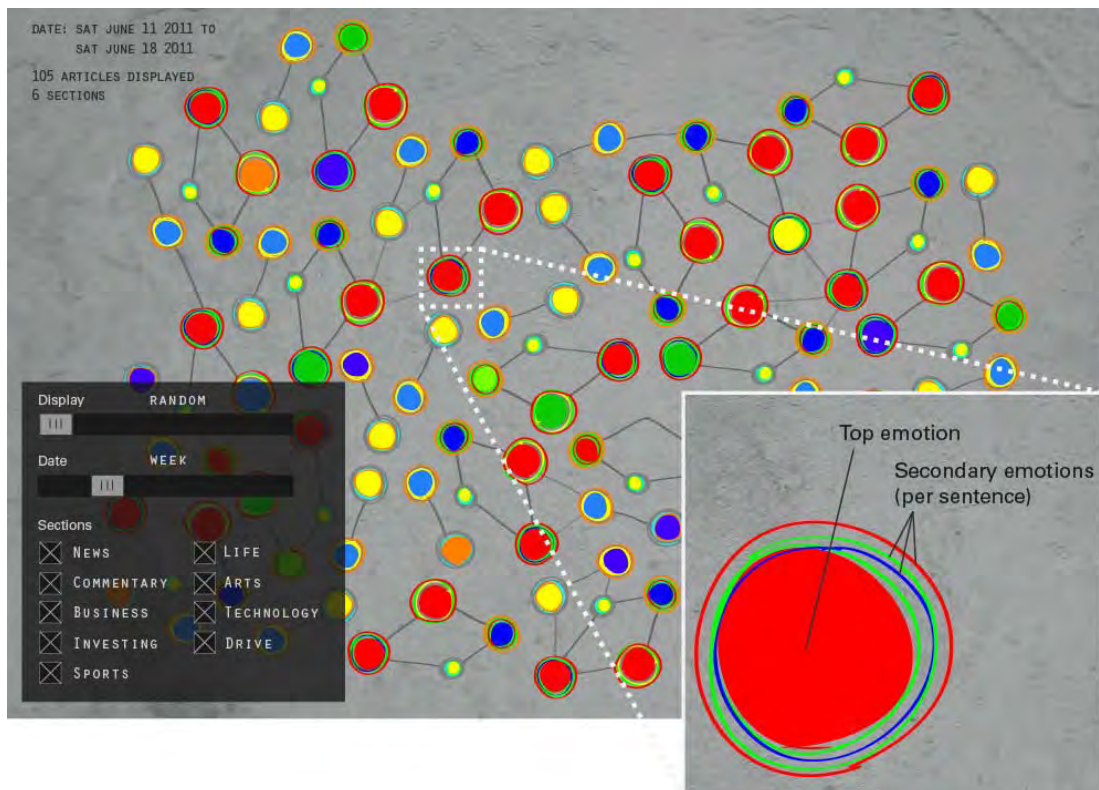


Figure 3: This sketch version features a network structure of nodes connected by section. Users can customize the display using the panel in the lower left corner. The arbitrary colours of each node represent the dominant emotions

Cell Chamber is comprised of news articles, which represent a single node which are described as cells. Each article represents an organic assemblage of concepts, interviews, imagery, and captions – components that together compose the entire story. Each cell has the ability to connect to any other cell in the network through heterogeneous connections built from the data that each article contains. Cell Chamber can provide an overall view of the article network or offer the user a single article as a starting point, which expands to more related articles as the user navigates based on relationships alone. Clicking on each cell expands into secondary rings that surround the article. The smaller rings signify each sentence in the article and how it adds up to the overall emotion of an article. Clicking on any of these rings shows the

associated metadata from the article, such as writing location, connections, geospatial, keywords, and emotional score. Making the data visible offers the reader a critical analysis of an article and its authors who contribute to the overall stance of an article.

For example, a reader may navigate based upon the writing location of the article. Such a view will reveal patterns that may emerge among author groups who are closer in a physical network. Article locations will also paint a better picture of the environmental factors and events surrounding the author at the time of writing. This brings the user closer to the origin point of the article, allowing for personal curiosity to take over the interpretation of the article and the factors surrounding it. Cells may also be aggregated based on time, positioning article relations on a scale of weeks or months.

Cell Chamber offers navigation between articles from different contexts. The organic nature of the visualization allows users to flow from one context to another, as cells are connected through different parameters that contrast the typical connections of related articles often established through keywords alone. Further work currently explores novel patterns of navigation in relation to users' preferences, recommender engine suggestions and editorial choice.

Summary

These case studies validate approaches to visual analytics and data visualization that draw from cultural theory, design methods such as participatory design or personas, and data analytics. These represent unique aesthetic approaches which provide use value to VAL's partners.

Bibliography

- Arunachalan, B., Diamond, S., Stevens, A., Ghaderi, M. & Talaie, B., "Visualization of Residents in Long-Term Care Centres through Mobile Natural User Interfaces (NUI)", *LDAV 2013, IEEE VIS 2013*, Atlanta, October, 2013
- Arunachalan, B., Diamond, S., Szigeti, S., Chevalier, F., Stevens, A., Ghaderi, M., Talaie, B., Reilly, D., "Designing Portable Solutions to Support Collaborative Workflow in Long-Term Care: A Collaborative Strategy", *IEEE CSCWD 2014*, Taipei, May/2014.
- Bostock, M., Ogievetsky, V., & Heer, J. (2011) D3: Data-driven documents. In *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, 2011
- Bresciani, S., Tan, M. & Eppler, M.J. (2012) Augmented Communication with Visualization: Effects on Emotional and Cognitive Response/ADIS, *International Journal on WWW/Internet* Vol. 9 (2)
- Cockton, G. (2008) Revisiting Usability's Three Key Principles *CHI 2008 Proceedings, April 5-10, Florence*, Florence: ACM, pp. 2473 – 2484.
- Diamond, S. (2010) Many Points of Light: Visualization and the Converging Aesthetics of Art, Design and Science, In *C Theory* ed. Kroker, A. & M. L. Reprinted 2012, in *Critical Digital Studies Reader*. ed. Kroker, A. & M. L. Toronto: UT Press.
- Eppler, M.J. (2011) What makes an Effective Knowledge Visualization? A Review of Seminal Concepts in *Proceedings of the 15th International*

- Conference on Information Visualization*, IEEE, London.
- Foucault, M. *Archaeology of Knowledge*. Routledge, 2006.
- Greenberg, S. & Buxton, B. (2008) Usability Evaluation Considered Harmful (Some of the Time). *CHI 2008 Proceedings, April 5-10, Florence*. Florence: ACM, pp. 111-120.
- Hall, S. Foucault: Power, Knowledge and Discourse. In *Discourse Theory and Practice: A Reader*, Sage, 2001.
- Kovalerchuk, Boris & Schwing, James (2004). *Visual and Spatial Analysis: Advances in Data Mining, Reasoning, and Problem Solving*
- Nielson, Jakob (1993) *Usability Engineering*. Mountain View: SunSoft.
- Oliver, S.; Gali, Guia, Chevalier, F. & Diamond, S. (2012) Fracturing Media Paradigms : Online Navigation of Online News Media. ACM. *DIS 2012*, June 11-15, 2012, Newcastle, UK
- Pattath, A. (2010) *Interactive Visualization for Mobile Visual Analytics*. Dissertation. Purdue University. Available at <http://gradworks.umi.com/34/53/3453347.html>
- Pridham, H., Stevens, A., Szigeti, S., Diamond, S., Arunachalan, B., “Fostering Insight and Collaboration in Long-Term Healthcare through the Collection and Visualization of Qualitative Healthcare Data”, *Mobile HCI*, ACM, Toronto, September/2014.
- Raghavan, Badri (2013) Deep Data Trumps Big Data Available at <http://www.informationweek.com/big-data/big-data-analytics/deep-data-trumps-big-data/d/d-id/1297588> *Information Week*
- Scharmer, Otto (2014) From Big Data to Deep Data Available at http://www.huffingtonpost.com/otto-scharmer/from-big-data-to-deep-dat_b_5599267.html, Posted: 07/18/2014 10:42 am EDT
- Schuler, D. & Namioka, A., eds. (1993) *Participatory Design: Principles and practices*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Szigeti, S, et al., Natural User Interface for Mobile Health Care Systems, *GRAND 2013*, Canada, May/2013.
- Stevens, Anne, Pridham, Hudson, Steve Szigeti, Steve, Diamond, Sara & Arunachalan, Bhuvaneshwari (2014) The Care and Condition Monitor: A Tablet Based Tool for Visualizing Informal Qualitative Health Care Data, *InfoVis 2014*, IEEE VIS 2014, Paris, November/2014.
- Thomas, James J. & Cook, Kristen A. (Ed.) (2005). *R&D Agenda for Visual Analytics*. National Visualization and Analytics Center. p.4
- Turner, Delaney (2012) Exploratory or Presentation? Visualization serves two masters| Oct 17 2012 | Available at *information-insights ibmsoftware*
- Viégas, F.B. & Wattenberg, M. (2005) *Artistic Data Visualization: Beyond VisualAnalysis*. Unpublished paper available from Cambridge, MA: VisualCommunication Lab, IBM Research
- Wang, Tricia (2013) Big Data Needs Thick Data. *Ethnography Matters*. Available at <http://ethnographymatters.net/blog/2013/05/13/big-data-needs-thick-data/>
- Ware, C. & Bobrow, R. (2005) Supporting Visual Queries on Medium-size Node-link Diagrams. *Information Visualization*, 2005, 4, London: Palgrave, pp. 49-58.