VISUALIZING THE DATA VISUALIZATION NETWORK: THE DVMAP PROJECT

Rosário Durão, PhD
Tie Wei, PhD
Kristina Henneke, MA
Karen M. Balch, BGS
Maxwell Hill, BSc candidate
Rachel Rayl, BSc candidate
New Mexico Institute of Mining and Technology, USA

Abstract

Data visualization is a familiar buzzword. Experts in the humanities, social and natural sciences, as well as technology, along with semi-experts and the general public, reach people everywhere with trends and conclusions drawn from visualized data. Governments, industries, businesses, sciences, marketers, academics, students and others value data visualization methods and tools as critical, applicable tools for understanding the world, which provide rich information analyses for specialists and generalists alike.

Unfortunately, no single resource offers a space where people working in the multifaceted field of data visualization can share projects they are working on, tools they created, educational opportunities in the field, nor where they (and their work) are situated geographically. A research group at New Mexico Institute of Mining and Technology seeks to fill this gap with a repository of data visualization resources called "Data Visualization Map (DVMap)." The DVMap is an interactive network data and geographic representation graph that provides a data visualization space for people across the world to share, view and/or collaborate on projects and publications; tools deployed or under development; educational opportunities in data visualization, such as formal programs, summer seminars, conferences; and the geographical locations of the users, projects, tools and educational opportunities.

Given the necessity of this repository, this paper outlines the structure, underlying methodology, and anticipated outcomes for the DVMap data visualization network. The paper also accounts for limitations of the project and the potential problems of creating a map that wants to share work – especially work in progress – with everyone.

Keywords: Data visualization network repository, Interactive network data graph, Geographic representation graph, Information analysis, Research tool, Collaboration tool

Introduction

This is the digital universe. It is growing 40% a year into the next decade, expanding to include not only the increasing number of people and enterprises doing everything online, but also all the "things" – smart devices – connected to the Internet, unleashing a new wave of opportunities for businesses and people around the world. Like the physical universe, the digital universe is large – by 2020 containing nearly as many digital bits as there are stars in the universe. It is **doubling in size every two years** (Turner, Gantz, Reinsel, and Minton 2014)

Data visualization is a familiar buzzword, reflecting the need to help sort the influx of global data into understandable and therefore usable information. Experts, semi-experts and the general public consume and produce data visuals, which governments, industries, businesses, sciences, marketers, academics, students and others use to understand and make decisions about the world.

Yet, very few of us - if any - know who is working on data visualization, the projects they are working on and the tools they deployed or are developing, the educational opportunities in the field, or where they (and their work) are situated geographically. Instead, there are a few lists of tools (e.g., Kirk 2009-2014; National Cancer Institute; Datavisualization.ch), people (e.g., Franchi 2013; Information Management 2014), projects (e.g., NASA Scientific Visualization Studio; visualizing.org), and educational opportunities (e.g., Nielsen 2013). There are almost no references to where the people working on data visualization and the projects and educational opportunities are in the world. Ortiz's "Data visualization references network" is the most complete aggregation of data visualization "blogs, studios, people, tools, books" so far, yet it contains a reduced number of items per topic, and it does not cover data visualization projects and educational opportunities. Also, with the partial exception of Ortiz's network graph, none of the existing resources represent the data for easy viewing, understanding, navigation, exploration, interconnectivity and charting on a map of the world. Nor do the existing resources pay special attention to the contribution and data visualization needs of governments, industries, businesses, sciences, marketers and engineers. In sum, no online resource currently aggregates the complexity of the data visualization network and does it in a manner that makes it easy for users with different needs and preferences to query the data or visually explore it, to see how each element relates to the entire network, and to understand where the elements stand on the world map.

A research group from the Visualizing STEM Research Synergy Cluster of the Humanizing Tech/nology research project at the New Mexico Institute of Mining and Technology seeks to fill this gap with a repository of data visualization resources called "Data Visualization Map (DVMap)." The DVMap is a data visualization space for anyone across the world with Internet access to share, collaborate and visualize projects and publications; tools deployed or under development; educational opportunities in data visualization, such as formal programs, individual courses, summer seminars, workshops; and the geographical locations of the users, projects, tools and educational opportunities.

The DVMap is an interactive network data and geographic representation graph. It is a network data graph because of the volume and complexity of items it displays (people, projects, tools, educational opportunities and places) and the relations among them. It is a geographic representation graph because the majority of these items are geographically located. The DVMap is explorable, searchable and interactive because it is centered on the preferences and needs of the users. With this in mind, we found the Max Planck Research Networks Prototype Moritz Stefaner created for the Institute to be a source of inspiration because of its similar purpose: "reveals how Max Planck Institutes collaborate with each other, and with their international partners." (Max Planck Research Networks; Stefaner 2012)[7].

In this paper, we outline the structure, underlying methodology and anticipated outcomes of the project. We also account for limitations and the potential problems of creating a map that wants to share work – especially work in progress – with everyone.

[7] A similar visualization tool, though targeted at visualizing digital library collections, is Thudt, Hinrichs, and Carpendale's The Bohemian Bookshelf information visualization (Thudt, Hinrichs, and Carpendale 2012; http://www.alicethudt.de/BohemianBookshelf/index.html)

DVMap structure - The users' view

The DVMap is centered on the user in two ways: users can engage with the interface to either explore the data visualization network, or search it for known items; and users generate (most or all of) the content that is displayed on the interface. In this section, we describe the structure of the DVMap from the perspective of both types of users.

Using the DVMap

When users access the website, they see an interface like the sample page in Fig. 1. The interface is dark colored to afford greater contrast with the data, and it has five components: A, B, C, D, E, and F. The topmost bar (component A) contains the title and navigation functions; panel B displays the overview map, C is the details section of the map, and D describes the information of the active node; E contains the search results. The bar at the bottom (F) contains the search boxes and selection functions.

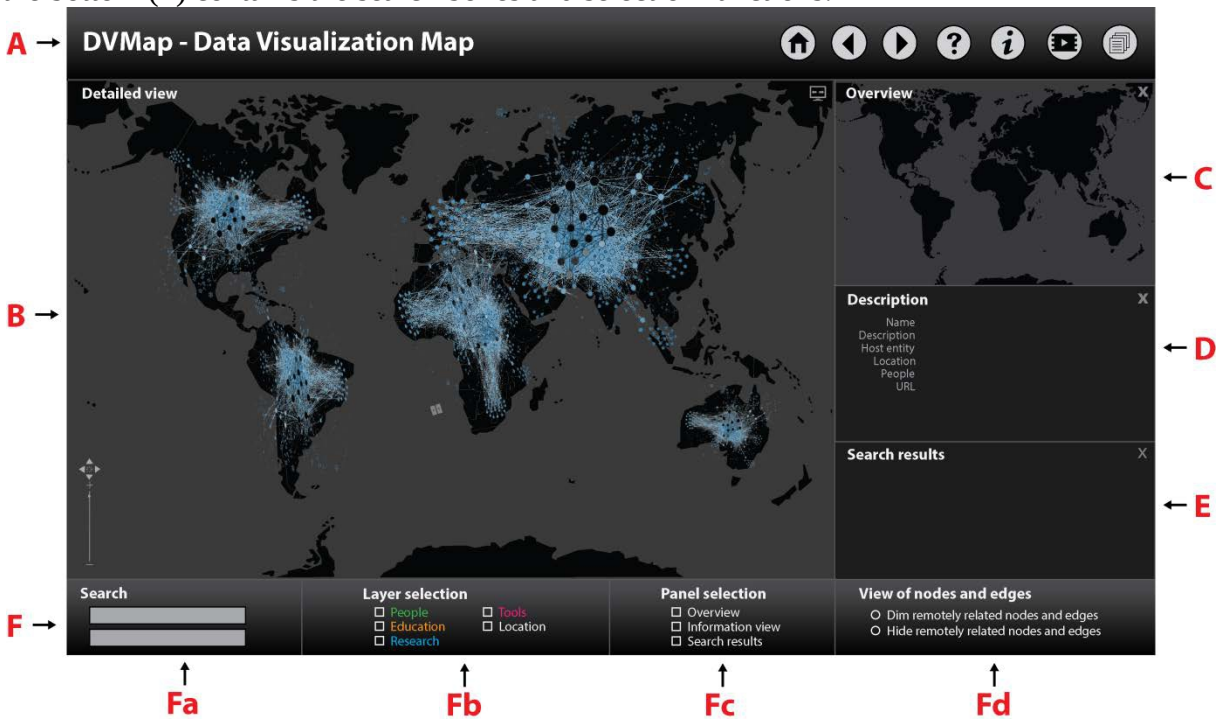

Fig. 1. Sample DVMap landing page, with page sections labeled.

The top bar (A) contains the project title on the left and the navigation menu on the right. Because the top bar is common to the whole website, clicking on the title always takes the user back to the home page. The menu is a graphics navigation bar; we chose this type of navigation bar to help users from different cultures and nationalities use the site with the same ease and user friendliness overall. The icons are: home, back, forward, about us, help, video tutorials, and add resource. Clicking on the about us, help, video tutorial, and add resource icons opens new pages. The help page provides written help on specific topics such as using the panels. The tutorials page duplicates this function in the form of brief videos. Providing the written and graphical help modes caters to users who are more writing centered and users who are more visually centered. Clicking on the add resource icon opens up the questionnaire where users contribute new content for the DVMap database (see Appendix). The menu supports addition of other icons, if necessary.

The center panels are a slight variation on Schneiderman's "Overview first, zoom and filter, then details-on-demand" visual information seeking mantra (1996), with panel C representing the overview, panel B the zoom and filter, and panel D the details-on-demand.

The search results panel (E) also filters the data. The default view displays all the panels (see Fig. 1); users can hide panels C, D and/or E from the bottom bar (section Fc).

Panel C displays a world map where, via a red rectangular frame and a node (see Fig. 2), the user keeps track of both the area of the world he is zooming in on in panel B, and where the active node is on the panel B map. Users may hide panel C by clicking on the *x* on the upper-right corner.

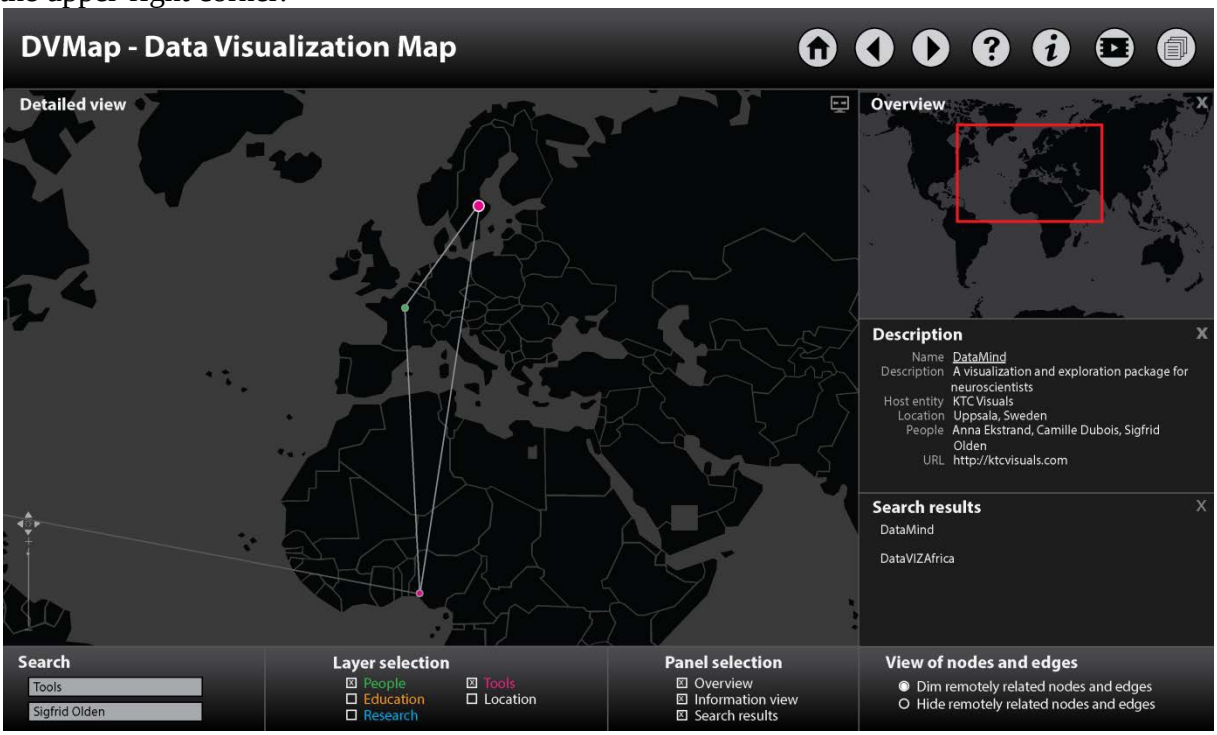

Fig. 2. Sample DVMap landing page with simulated use scenario.

Panel B displays the actual network on a world map. Here, users can explore the network, zoom in on details, and "filter out uninteresting items." (Schneiderman 1996). The network nodes and edges are color-coded for the different overlapping layers of content: green for people, orange for education, blue for research, pink for tools, and off-white for places. When a user clicks on a node, they activate the node, which becomes bigger, and both the node and the edges leading off it brighter. At the same time, by default the more remotely related nodes and edges become dimmer (users can dim or hide them from the bottom bar, section Fd). When a node becomes active, the information related to it is displayed in panel D, allowing the user to learn more about the item and navigate its connections and overlaps. Users can also move between the different layers either by zooming in on the corresponding color nodes and edges, or by selecting the desired layer, or layers in section Fb. Users can control zooming and panning in different ways: via the map control widget on the lower-left of the panel, via a mouse, or via touch. Users can also view the map in full screen by clicking on the icon on the upper-right corner of the panel.

Panel D displays the information related to the node: names of people, educational resources, research projects, tools and places. This information is displayed as a form, with the items that are linked to nodes[8] underlined.

Panel E displays the search results. Like panel D, the items that link to nodes are underlined.

[8] The items that link to nodes (here called "nodable" items) are names of people, educational, research, and tool resources, and their geographic location.

The bar at the bottom (F) is the search and options part of the DVMap. In section a, users can enter up to two search queries, and the results will be displayed in panel E. The results include both nodable and non-nodable items because search queries search the whole DVMap database and display all the information related to the query items.

Section b allows users to filter their analysis by category. An example of this is users wanting to explore the tools network. They would select the "Tools" checkbox and the corresponding network would be highlighted. If they also want to explore the network of people, they would also select the "People" checkbox. Users can select any combination of categories to analyze. They can also apply this type of selection to filter the search results.

Section c allows users to hide or show panels C, D and/or E. They can also resize panels D and E.

In section d, users select whether they want to dim or hide the remotely related nodes and edges.

Contributing to the DVMap

The content displayed in panels B, C, D and E is generated (wholly or predominantly[9]) by users. When they click on the add resource icon in the top bar, contributors are taken to the "DVMap questionnaire" (Appendix). Fig. 3 depicts the structure of the questionnaire.

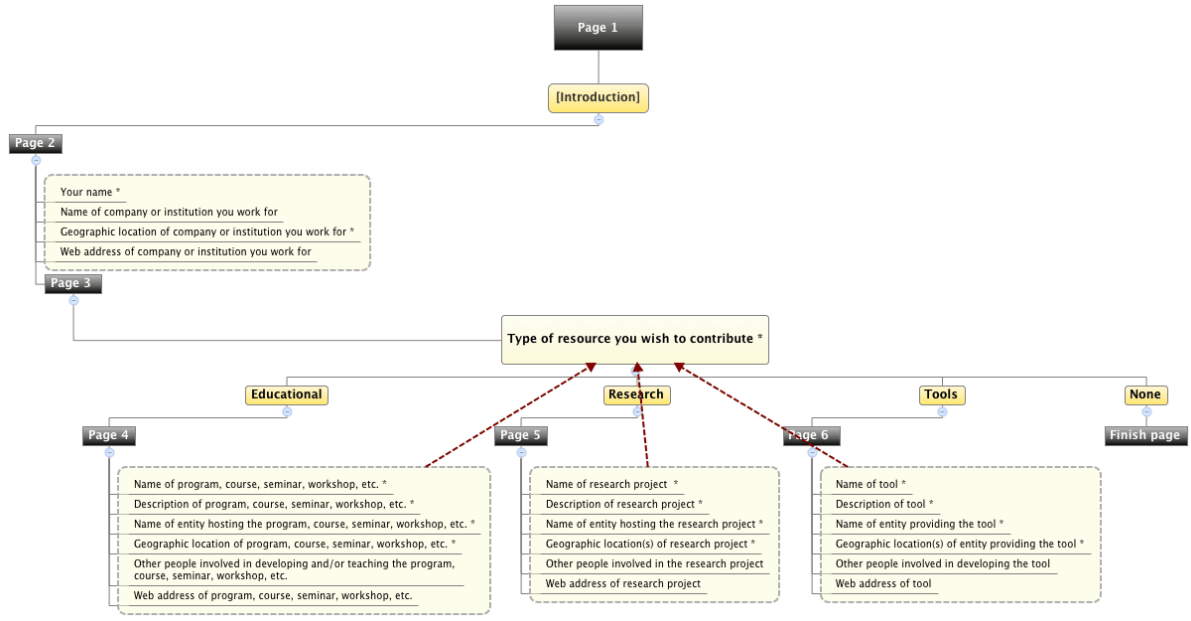

Fig. 3. Questionnaire structure.

The questionnaire contains seven pages, with the first page being the introduction and the last the finish page. After reading the introduction and signing the disclaimer by clicking the "Next" button, contributors are taken to page 2 where they provide their name and information about their workplace. Of the four questions in this section, only the contributors' name and geographic location of their workplace are required. Contributors are then taken to page 3, where they select the type of resource they wish to contribute (educational, research or tools) or the "None" button that takes them to the finish page. Selecting one of the other options takes contributors to the corresponding question page (Educational to page 4, Research to page 5, and Tools to page 6). After completing one of

[9] Our intent is that all input be user-generated. However, if the quality and amount of user input is unsatisfactory, we will revisit this decision.

these pages, contributors are taken back to page 3, where they repeat the selection and response sequence until they submit all the desired information.

DVMap methodology - The developers' view

The methodology, or developers' view, includes the content processing cycle (Fig. 4) and the current and future development of the project.

Content processing cycle

Content processing is a three-stage process that begins with users logging into the website, creating an account that is then added to the database, and filling out the questionnaire. This information forms a nugget in the data base. The answers are verified manually to prevent duplicate, inconsistent or incomplete information from being displayed on the interface. The information is then processed into the node/network graph via neo4j, and displayed on the interface for users to explore and search. This allows users to extract specific set(s) of data.

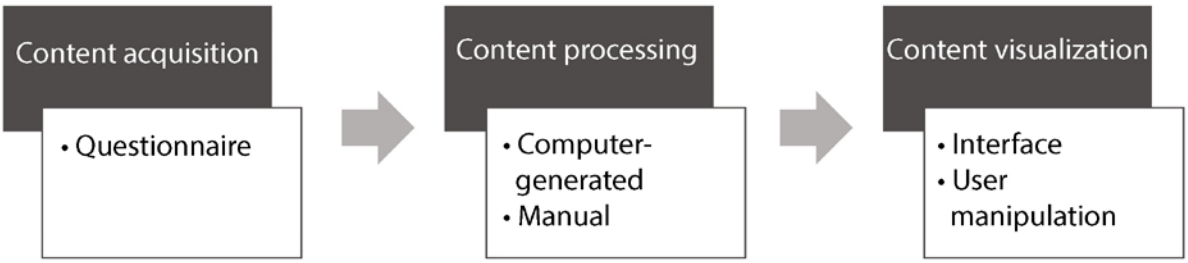

Fig. 4. DVMap content processing cycle

Current and future development

To date, we devised the questionnaire (Fig. 3 and Appendix) and samples of the website (Figs. 1 and 2). We also answered seven questionnaires with simulated data, and one with actual data about data visualization courses and projects at NMT. From the responses, we are developing a sample interactive map (Fig. 2), which will be available on the project website at http://blogs.nmt.edu/digitalhumanities/cluster-i/data-visualization-map-dvmap/

Next, we will produce a prototype of the website and perform user experience testing with faculty and students at NMT. We will use the results to produce a beta version of the website and start collecting data from actual contributors via individualized invitations, announcements among data visualization and professional communication communities, publications, and conference presentations. We plan on doing two further rounds of beta testing and integration of the results. Beta testing will be performed with actual users elicited via announcements in data visualization and professional communication communities. We expect to perform the tests at 6-month intervals. We expect to have a fully functional website within two years.

Anticipated outcomes

We anticipate two major outcomes. First, the DVMap will provide a unique interactive repository of data visualization resources for anyone with an Internet connection to share, view and/or collaborate on. Second, make the resource available for free to the academic and research communities, and on a paid basis for other users. Creative ways of generating income, in addition to public or private funding, is important to the sustainability of our effort. Income generating products may include tailored training sessions, project consulting, and more. We will closely follow the strategies of online education or online business in general, and adopt the ones that match the DVMap.

Together, these outcomes will (a) help researchers and practitioners better communicate their findings and applications with peers, other specialists and the public at large, and they will facilitate governments, industries, businesses, scientists, marketers, academics, students and others searching for new work and networking opportunities in or involving data visualization; and (b) ensure the long-term economic and academic viability of the project.

Some of the limitations of the DVMap are that it is not being designed for tablet and mobile devices, nor for different languages and writing systems. The decision to have users contribute the data to populate the network, while circumventing legal and ethical problems, can also be a limitation if the entries are too few and far apart, or are not of required standard.

Some of the problems we anticipate are related to user input. An inevitable problem is people entering incomplete data, inconsistent data, data that is out of date, or two or more people submitting the same data in different ways. Solving this issue is time-consuming and costly as it will require a potentially large pool of people dedicated to verifying the data submitted. A second issue is that almost every day new people become involved with data visualization and new resources are developed that could, and ideally should, be entered into the database; yet, if we depend exclusively on user input to populate the repository, it might quickly, even chronically, be inadequate in size and outdated in time. Both these issues will result in the repository not fulfilling its aims of being a reliable source of information about the data visualization network. (Aylett et al 2012).

A further group of problems is structural: these have to do with the task of maintaining a repository that can register and process increasing amounts data, to do it in such a way as to prevent the above mentioned user input problems, and to do it with as little human intervention as possible. We also foresee problems with the data processing as the larger the data sets become, the longer they take to process, making the real-time updates to the DVMap more cumbersome. Another problem has to do with whether to categorize the data or not. For instance, is it preferable to leave the data relative to tools general, or will it meet users' requirements better if we create taxonomies based on input, output, field/discipline, or others? Categorizing based on input would allow for creation of more specific output forms, such as a historical timeline, that users are already familiar with (Friedman 2007). However, categorizing based on the common output of the tool allows us to potentially streamline our categories based on the current classifications (Meirelles 2013), to which current users of data visualization would already be more accustomed. Finally, categories based on intended field of use, such as science and medicine for mapping cancer genomes, could allow users to better search for tools more relevant to their field. (Zoss 2014; StatSoft, Inc. 2013).

We also anticipate some people issues. One issue is making the website accessible to people with different visual capabilities. For instance, is the dark background adequate for all visualization users and tools? Also, will color-coding the layers suffice, or should we code the nodes for shape also? Another issue is making the website accessible to people from multiple cultures, nationalities and communities of practice. Examples of issues we would need to address in this case would be, as already mentioned, catering to different languages and writing systems.

Conclusion

The DVMap is a repository of data visualization resources that fills a gap in the knowledge about the people, projects, tools, educational opportunities and geographical location of the data visualization network. This makes it a useful tool for governments, industries, businesses, sciences, marketers, academics, students and the general public. In

this paper, we described the rationale, structure, layout and the anticipated outcomes and problems of the DVMap project.

References:

Aylett, R., Bental, D., Stewart, R., Forth J.. and Wiggins, G. Supporting Serendipitous Discovery. Digital Futures (Third Annual Digital Economy Conference), 23rd – 25th October 2012, Aberdeen, UK. Available at http://www.serena.ac.uk/papers/#sthash.lOtufHnu.dpuf

Datavisualization.ch. Datavisualization.ch Selected Tools. Available at http://selection.datavisualization.ch/

Franchi, Francesco. About information design / data visualization, Fracesco Franchi blog, 2013. Available at www.ffran.ch/id.pdf

Friedman, Vitaly. Data Visualization: Modern Approaches. *Smashing Magazine*, 2007. Available at http://www.smashingmagazine.com/2007/08/02/data-visualization-modern-approaches/

Information Management. 11 Data Visualization Experts You Should Follow on Twitter. Information Management, 2014. Available at http://www.information-management.com/gallery/11-data-visualization-experts-you-should-be-following-on-twitter-10024000-1.html

Kirk, Andy. Essential Collection of Visualisation Resources. Visualising Data. 2009-2014. Available at http://www.visualisingdata.com/index.php/resources/

Max Planck Research Networks. Available at http://max-planck-research-networks.net/

Meirelles, Isabel. *Design for Information*. Rockport, 2013.

NASA Scientific Visualization Studio. Available at http://svs.gsfc.nasa.gov/

National Cancer Institute. Other GIS Tools. Surveillance Research Program, Division of Cancer Control & Population Sciences, U.S. Department of Health & Human Services, National Institutes of Health, National Cancer Institute, USA.gov. Available at http://gis.cancer.gov/tools/other_tools.html

René Clausen Nielsen. Going to Data Visualization School [Web log post]. January 11, 2013. Available at http://worldofdata.org/2013/01/11/going-to-data-visualization-school/

Ortiz, Santiago. Data visualization references network: A navigable selection of sites and resources about data visualization. Santiago Ortiz website, 2002-2013. Available at http://moebio.com/datavisnetwork/

Schneiderman, Ben. The eyes have it: A task by data type taxonomy for information visualizations. *IEEE Symposium on Visual Languages, 1996. Proceedings*, 1996, 336-343, 1996. doi: 10.1109/VL.1996.545307

StatSoft, Inc. How to Visualize Data (Graph Types). Electronic Statistics Textbook. StatSoft, 2013. Available at http://www.statsoft.com/Textbook/Graphical-Analytic-Techniques

Stefaner, Moritz. Eyeo Festival - Moritz Stefaner 2012. Available at http://vimeo.com/46893010

Thudt, Alice, Uta Hinrichs, and Sheelagh Carpendale. The Bohemian Bookshelf: Supporting Serendipitous Book Discoveries through Information Visualization. In CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2012. Available at http://innovis.cpsc.ucalgary.ca/innovis/uploads/Publications/Publications/ThudtCHI2012.pdf

Turner, Vernon, John F. Gantz, David Reinsel, and Stephen Minton. The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things [White paper]. EMC Corporation, 2014. Available at http://idcdocserv.com/1678

visualizing.org. Available at http://visualizing.org/

Zoss, Angela. Introduction to Data Visualization. Duke University Libraries, 2014. Available at http://guides.library.duke.edu/print_content.php?pid=355157&sid=2904910&mode=g

Appendix

o DVMap questionnaire

The DVMap is a repository of information about people working in Data Visualization, as well as the projects, tools, and educational opportunities they develop and participate in.

If you are a practitioner, researcher, or faculty member and you developed or participated in Data Visualization educational opportunities, research projects, and tools, please consider answering this questionnaire.

Your answers will help us create an online interactive network map of the people, projects, tools, and educational opportunities in Data Visualization. Your answers will also help other faculty members, practitioners, and researchers, as well as companies, institutions, and students

- better communicate their data visualization findings and applications,
- facilitate searching for new work and networking opportunities in and involving data visualization, and
- provide easy and comprehensive access to a repository of data visualization tools and learning opportunities.

Disclaimer

The DVMap research group, the Humanizing Tech/nology project and the New Mexico Institute of Mining and Technology accept no liability for the information submitted in this questionnaire. Only submit information which you are authorized to share.

o [2] Your name *

o [3] Name of company or institution you work for

o [4] Geographic location of company or institution you work for *
(please provide name of city/town, state (if applicable), country; e.g., “Lagos, Nigeria” or “Albuquerque, NM, United States”)

o [5] Web address of company or institution you work for

o [6] Type of resource you wish to contribute *

Educational
(“educational resources” include academic programs [e.g., BA, MS, PhD], individual courses [e.g., “Data Visualization”], summer seminars, workshops, and other educational opportunities)

Research
(“research resources” are typically research projects)

Tools
(“tool resources” include tools for visually displaying quantitative and qualitative data)

None

[7] Name of program, course, seminar, workshop, etc. *

[8] Description of program, course, seminar, workshop, etc. *
(100 words max.)

[9] Name of entity hosting the program, course, seminar, workshop, etc. *

[10] Geographic location of program, course, seminar, workshop, etc. *
(please provide name of city/town, state (if applicable), country; e.g., “Lagos, Nigeria” or “Albuquerque, NM, United States”)

[11] People involved in developing and/or teaching the program, course, seminar, workshop, etc.
(please use commas to separate the names; e.g., “Samantha Smith, Jacques Renault”)

[12] Web address of program, course, seminar, workshop, etc.

[Page 5]
[13] Name of research project *

[14] Description of research project *
(100 words max.)

[15] Name of entity hosting the research project *

[16] Geographic location(s) of research project *
(please provide name of city/town, state (if applicable), country; e.g., “Lagos, Nigeria” or “Albuquerque, NM, United States”)

[17] People involved in the research project
(please use commas to separate the names; e.g., “Samantha Smith, Jacques Renault”)

[18] Web address of research project

[Page 6]
[19] Name of tool *

[20] Description of tool *
(100 words max.)

[21] Name of entity providing the tool *

[22] Geographic location(s) of entity providing the tool *
(please provide name of city/town, state (if applicable), country; e.g., “Lagos, Nigeria” or “Albuquerque, NM, United States”)

[23] People involved in developing the tool
(please use commas to separate the names; e.g., “Samantha Smith, Jacques Renault”)

[24] Web address of tool

[Finish page]
Thank you for taking the time to complete this questionnaire.