Visual Design Principles

Katy Börner
Cyberinfrastructure for Network Science Center, Director
Information Visualization Laboratory, Director
School of Library and Information Science
Indiana University, Bloomington, IN
katy@indiana.edu

With special thanks to the members at the Cyberinfrastructure for Network Science Center; the Sci2, NWB, and EpiC team; and the VIVO Collaboration

3rd International Meeting on Visualizing Biological Data (VIZBI 2012)
EMBL, Heidelberg, Germany
http://vizbi.org/2012

March 8, 2012

Overview

1. **Motivation:** Design informative and visually pleasing visualizations that make a difference.

2. **Theory:** Learn from and combine approaches from psychology, cartography, computer science, information visualization, statistics, graphic design.

3. **Practice:** Sample visualizations designed for experts and a general audience + plug-and-play macroscope tools that commoditize data mining and visualization.
Overview

1. **Motivation:** Design informative and visually pleasing visualizations that make a difference.

2. **Theory:** Learn from and combine approaches from psychology, cartography, computer science, information visualization, statistics, graphic design.

3. **Practice:** Sample visualizations designed for experts and a general audience + plug-and-play macroscope tools that commoditize data mining and visualization.

http://www.home-2009.com

Yann Arthus-Bertrand
http://www.malarialifecycle.com

The Whole Brain Catalog:
http://wholebraincatalog.org

Edward R. Tufte

Envisioning Information

Edward R. Tufte
All three care deeply about

1. Data,
2. Existing expertise and insight needs, and
3. Are able to acquire the resources it takes to

 Spent months/years to deeply understand the problem/possible solutions.
 Render data optimally for the human perceptual-cognitive system – given our current understanding of human perception/cognition and technology.

The result are insightful yet perceptually stunning, intellectually stimulating, and emotion provoking imagery.

Photography, scientific visualization, animation— x-y-z positions exist.

Data visualization, data graphics— Mostly without x-y-z positions.
Today’s massive amounts of streaming data cannot be rendered by hand.

How can computers be used to render informative and visually pleasing visualizations that make a difference?

Need BIG data analyses and visualizations that conform to human visual perception and cognitive processing.
Hubble Telescope
Original total cost estimate: US$400 million
Construction costs: US$2.5 billion
Cumulative costs: US$4.5 and $6 billion with €593 million EU contribution
http://www.spacetelescope.org/about/faq

Overview

1. **Motivation:** Design informative and visually pleasing visualizations that make a difference.

2. **Theory:** Learn from and combine approaches from psychology, cartography, computer science, information visualization, statistics, graphic design.

3. **Practice:** Sample visualizations designed for experts and a general audience + plug-and-play macroscope tools that commoditize data mining and visualization.
**Needs-Driven Workflow Design** using a modular data acquisition/analysis/modeling/visualization pipeline as well as modular visualization layers.

**Visualization Layers:** Reference System (samples)

<table>
<thead>
<tr>
<th>1D</th>
<th>2D</th>
<th>Circular</th>
<th>Tabular</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aravind</td>
<td>Baldauf</td>
<td>Becerra</td>
<td>Benachenhou</td>
</tr>
<tr>
<td>Brinkmann</td>
<td>Brocks</td>
<td>Brown</td>
<td>Caetano-Anolles</td>
</tr>
</tbody>
</table>

Examples taken from [http://vizbi.org/Posters/2012](http://vizbi.org/Posters/2012)
Visualization Layers: Projection/Distortion


Visualization Layers: Raw Data

Examples taken from Wikipedia and http://www2.warwick.ac.uk/fac/sci/moac/people/students/peter_cock/r/ramachandran/
Graphic Design

Semiology of Graphics

- visual encoding
  - points, lines, areas
  - patterns, trees/networks, grids
  - positional: XY
    - 1D, 2D, 3D
  - retinal: Z
    - size, lightness, texture,
      - colour, orientation, shape,
  - temporal:
    - animation

See Jessie Kennedy’s Tutorial slides on “Visualization Principles”
http://mkweb.bcgsc.ca/vizbi/2012

Visualization Layers: Aggregation/Clustering

Aggregate over time, (geo) space, semantic similarity, network structure

Cluster using
- data mining on data itself.
- exogenous classifications, ontologies, knowledge, etc.
- Idiographic data classification using
  natural breaks
  quantiles
  means and standard deviations
  equal intervals
  etc.

Examples taken from http://vizbi.org/Posters/2012

Visualization Layers: Combination

Small Multiples  Multiple (coupled) windows

Change over time

Examples taken from http://vizbi.org/Posters/2012
Visualization Layers: Interaction

Search, filter, select
Zoom, Pan
Pruning
Brushing
Details on demand
Focus & context
User status

Examples taken from http://vizbi.org/Posters/2012

Visualization Layers: Legend Design

Show scale for objects that have a real-world size.
Grid lines to help people visually track data.
Info on “How to read the map” and major insights to take away.

Data details, analysis+modeling algorithms and parameters, credits, author names.
Visualization Layers: Deployment

Paper printout
(static, high resolution)

Interactive displays
(dynamic, low resolution)

Illuminated Diagram
Animations
Hands-on exercises

Needs-Driven Workflow Design using a modular data acquisition/analysis/modeling/visualization pipeline as well as modular visualization layers.
Type of Analysis vs. Level of Analysis

<table>
<thead>
<tr>
<th>Type of Analysis</th>
<th>Micro/Individual (1-100 records)</th>
<th>Meso/Local (101–1,000 records)</th>
<th>Macro/Global (10,000 &lt; records)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical Analysis/Profiling</td>
<td>Individual person and their expertise profiles</td>
<td>Larger labs, centers, universities, research domains, or states</td>
<td>All of NSF, all of USA, all of science.</td>
</tr>
<tr>
<td>Temporal Analysis (When)</td>
<td>Funding portfolio of one individual</td>
<td>Mapping topic bursts in 20-years of PNAS</td>
<td>113 Years of Physics Research</td>
</tr>
<tr>
<td>Geospatial Analysis (Where)</td>
<td>Career trajectory of one individual</td>
<td>Mapping a states intellectual landscape</td>
<td>PNAS publications</td>
</tr>
<tr>
<td>Topical Analysis (What)</td>
<td>Base knowledge from which one grant draws.</td>
<td>Knowledge flows in Chemistry research</td>
<td>VxOrd/Topic maps of NIH funding</td>
</tr>
<tr>
<td>Network Analysis (With Whom?)</td>
<td>NSF Co-PI network of one individual</td>
<td>Co-author network</td>
<td>NIH’s core competency</td>
</tr>
</tbody>
</table>

Overview

1. **Motivation:** Design informative and visually pleasing visualizations that make a difference.

2. **Theory:** Learn from and combine approaches from psychology, cartography, computer science, information visualization, statistics, graphic design.

3. **Practice:** Sample visualizations designed for experts and a general audience + plug-and-play macroscope tools that commoditize data mining and visualization.
Different Stakeholder Groups and Their Needs

Funding Agencies
- Need to monitor (long-term) money flow and research developments, identify areas for future development, stimulate new research areas, evaluate funding strategies for different programs, decide on project durations, funding patterns.

Scholars
- Want easy access to research results, relevant funding programs and their success rates, potential collaborators, competitors, related projects/publications (research push).

Industry
- Is interested in fast and easy access to major results, experts, etc. Influences the direction of research by entering information on needed technologies (industry-pull).

Advantages for Publishers
- Need easy to use interfaces to massive amounts of interlinked data. Need to communicate data provenance, quality, and context.

Society
- Needs easy access to scientific knowledge and expertise.

Scholars Have Different Roles/Needs

Researchers and Authors—need to select promising research topics, students, collaborators, and publication venues to increase their reputation. They benefit from a global view of competencies, reputation and connectivity of scholars; hot and cold research topics and bursts of activity, and funding available per research area.

Editors—have to determine editorial board members, assign papers to reviewers, and ultimately accept or reject papers. Editors need to know the position of their journals in the evolving world of science. They need to advertise their journals appropriately and attract high-quality submissions, which will in turn increase the journal’s reputation.

Reviewers—read, critique, and suggest changes to help improve the quality of papers and funding proposals. They need to identify related works that should be cited or complementary skills that authors might consider when selecting project collaborators.

Teachers/Mentors—teach classes, train doctoral students, and supervise postdoctoral researchers. They need to identify key works, experts, and examples relevant to a topic area and teach them in the context of global science.

Inventors—create intellectual property and obtain patents, thus needing to navigate and make sense of research spaces as well as intellectual property spaces.

Investigators—scholars need funding to support students, hire staff, purchase equipment, or attend conferences. Here, research interests and proposals have to be matched with existing federal and commercial funding opportunities, possible industry collaborators and sponsors.

Team Leads and Science Administrators—many scholars direct multiple research projects simultaneously. Some have full-time staff, research scientists, and technicians in their laboratories and centers. Leaders need to evaluate performance and provide references for current or previous members; report the progress of different projects to funding agencies.
### Type of Analysis vs. Level of Analysis

<table>
<thead>
<tr>
<th>Analysis Type</th>
<th>Micro/Individual (1-100 records)</th>
<th>Meso/Local (101–10,000 records)</th>
<th>Macro/Global (10,000 &lt; records)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistical Analysis/Profiling</strong></td>
<td>Individual person and their expertise profiles</td>
<td>Larger labs, centers, universities, research domains, or states</td>
<td>All of NSF, all of USA, all of science.</td>
</tr>
<tr>
<td><strong>Temporal Analysis (When)</strong></td>
<td>Funding portfolio of one individual</td>
<td>Mapping topic bursts in 20-years of PNAS</td>
<td>113 Years of PNAS Research</td>
</tr>
<tr>
<td><strong>Geospatial Analysis (Where)</strong></td>
<td>Career trajectory of one individual</td>
<td>Mapping a state’s intellectual landscape</td>
<td>PNAS publications</td>
</tr>
<tr>
<td><strong>Topical Analysis (What)</strong></td>
<td>Topical structure from which one grant draws.</td>
<td>Knowledge flows in Chemistry research</td>
<td>VxOrd/Topic maps of NIH funding</td>
</tr>
<tr>
<td><strong>Network Analysis (With Whom?)</strong></td>
<td>NSF Co-PI network of one individual</td>
<td>Co-author network NIH’s core competency</td>
<td>NIH’s core competency</td>
</tr>
</tbody>
</table>

---

### Mapping Indiana’s Intellectual Space

Identify:
- Pockets of innovation
- Pathways from ideas to products
- Interplay of industry and academia
Mapping the Evolution of Co-Authorship Networks
Mapping Transdisciplinary Tobacco Use Research Centers Publications

Compare R01 investigator based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.

Zoss & Börner, forthcoming.

Supported by NIH/NCI Contract HHSN261200800812

Mapping Topic Bursts

Co-word space of the top 50 highly frequent and bursty words used in the top 10% most highly cited PNAS publications in 1982-2001.

Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions


Research questions:
1. Does space still matter in the Internet age?
2. Does one still have to study and work at major research institutions in order to have access to high quality data and expertise and to produce high quality research?
3. Does the Internet lead to more global citation patterns, i.e., more citation links between papers produced at geographically distant research instructions?

Contributions:
- Answer to Qs 1 + 2 is YES.
- Answer to Qs 3 is NO.
- Novel approach to analyzing the dual role of institutions as information producers and consumers and to study and visualize the diffusion of information among them.
VIVO International Researcher Network

VIVO: A Semantic Approach to Creating a National Network of Researchers (http://vivoweb.org)

- Semantic web application and ontology editor originally developed at Cornell U.
- Integrates research and scholarship info from systems of record across institution(s).
- Facilitates research discovery and cross-disciplinary collaboration.
- Simplify reporting tasks, e.g., generate biosketch, department report.

Funded by $12 million NIH award.

**Cornell University:** Dean Krafft (Cornell PI), Manolo Bevia, Jim Blake, Nick Cappadona, Brian Caruso, Jon Corson-Rikert, Elly Cramer, Medha Devare, John Fereira, Brian Lowe, Stella Mitchell, Holly Mistlebauer, Anup Sawant, Christopher Westling, Rebecca Younes. **University of Florida:** Mike Conlon (VIVO and UF PI), Cecilia Botero, Kerry Britt, Erin Brooks, Amy Buhler, Ellie Bushhousen, Chris Case, Valrie Davis, Nita Ferree, Chris Haines, Rae Jesano, Margeaux Johnson, Sara Kreinstein, Yang Li, Paula Markes, Sara Russell Gonzalez, Alexander Rockwell, Nancy Schaefer, Michele R. Tennant, George Hack, Chris Barnes, Narayan Raum, Brenda Stevens, Alicia Turner, Stephen Williams. **Indiana University:** Katy Borner (IU PI), William Barnett, Shanshan Chen, Ying Ding, Russell Duhamon, Jon Dunn, Micah Linneweber, Nianli Ma, Robert McDonald, Barbara Ann O’Leary, Mark Price, Yuyin Sun, Alan Walsh, Brian Wheeler, Angela Zoss. **Ponce School of Medicine:** Richard Noel (Ponce PI), Ricardo Espada, Damaris Torres. **The Scripps Research Institute:** Gerald Joyce (Scripps PI), Greg Dunlap, Catherine Dunn, Brant Kelley, Paula King, Angela Murrell, Barbara Noble, Cary Thomas, Michael Irvin. **Washington University, St. Louis:** Rakesh Nagarajan (WUSTL PI), Kristi L. Holmes, Sunita B. Koul, Leslie D. McIntosh. **Weill Cornell Medical College:** Curtis Cole (Weill PI), Paul Albert, Victor Brodsky, Adam Cheriff, Oscar Cruz, Dan Dickinson, Chris Huang, Itay Klar, Peter Michilini, Grace Migliorisi, John Ruffing, Jason Specland, Tru Tran, Jesse Turner, Vinay Varughese.
Temporal Analysis (When) Temporal visualizations of the number of papers/funding award at the institution, school, department, and people level
**Topical Analysis (What)** Science map overlays will show where a person, department, or university publishes most in the world of science. (in work)

**Network Analysis (With Whom?)** Who is co-authoring, co-investigating, co-inventing with whom? What teams are most productive in what projects?
Geospatial Analysis (Where)  Where is what science performed by whom? Science is global and needs to be studied globally.
VIVO On-The-Go

Overview, Interactivity, Details on Demand come to commonly used devices and environments

Mapping Science Exhibit – 10 Iterations in 10 years
http://scimaps.org/
Read about and zoom into maps at http://scimaps.org/exhibit_info

Call for Maps for the 8th Iteration of the Places & Spaces: Mapping Science Exhibit on "Science Maps for Kids" (2012)

http://scimaps.org/call
Mapping Science Exhibit at MEDIA X was on May 18, 2009 at Wallenberg Hall, Stanford University, 

Science Maps in “Expediton Zukunft” science train visiting 62 cities in 7 months 12 coaches, 300 m long 
Opening was on April 23rd, 2009 by German Chancellor Merkel
http://www.expediton-zukunft.de
Illuminated Diagram Display
soon on display at the Smithsonian in DC.
http://scimaps.org/exhibit_info/#ID

Geographic Map: Where Science Gets Done

Science Map: How Scientific Disciplines Relate

About
This Illuminated Diagram display adds the flexibility of an interactive program to the incredibly high data density of a print. This technique is generally useful when there is too much pertinent data to be displayed on a screen but the data is relatively stable. The computer can direct the eye to what’s important by using projections or screens as smart spotlights,3 arranging the research impact of individuals, giving a “grand tour” of science, or highlighting query results (as when you touch the touch screen) with an overlay of moving light.

Cloning
Cloning is the process of making a genetically identical copy. Cloning can refer to techniques of producing a genetic copy of an organism by replacing the nucleus of an unfertilized ovum with the nucleus of a body cell from the organism. The reconstituted egg containing the DNA from a donor cell must be treated with chemicals or electric current in order to stimulate cell division. Once the cloned embryo reaches a suitable stage, it is transferred to the uterus of a female host where it continues to develop until birth.

The first adult mammal cloned was Dolly the Sheep in 1997.

Interact
Select any location on the Geographic Map (by brushing your finger over an area on the touch screen) and topics studied in that area will highlight on the Science Map; the brighter a topic, the more papers on that topic originate in the selected area. Conversely, touching a scientific area on the Science Map illuminates places on the Geographic Map where that topic is studied. People and topic buttons support the exploration of publication output by selected Nobel laureates and particular lines of research using MEDLINE data from 2000-2009.

http://scimaps.org
Ingo Gunther's Worldprocessor globe design now on display at the Giant Geo Cosmos OLED Display at the Museum of Emerging Science and Innovation in Tokyo, Japan
References


Overview

1. **Motivation:** Design informative and visually pleasing visualizations that make a difference.

2. **Theory:** Learn from and combine approaches from psychology, cartography, computer science, information visualization, statistics, graphic design.

3. **Practice:** Sample visualizations designed for experts and a general audience + plug-and-play macrocope tools that commoditize data mining and visualization.
Designing “Dream Tools”

Many of the best micro-, tele-, and macroscopes are designed by scientists keen to observe and comprehend what no one has seen or understood before. Galileo Galilei (1564–1642) recognized the potential of a spyglass for the study of the heavens, ground and polished his own lenses, and used the improved optical instruments to make discoveries like the moons of Jupiter, providing quantitative evidence for the Copernican theory.

Today, scientists repurpose, extend, and invent new hardware and software to create “macroscopes” that may solve both local and global challenges.

Plug-and-play macroscopes empower me, my students, colleagues, and 100,000 others that downloaded them.
Macrosopes

Decision making in science, industry, and politics, as well as in daily life, requires that we make sense of data sets representing the structure and dynamics of complex systems. Analysis, navigation, and management of these continuously evolving data sets require a new kind of data-analysis and visualization tool we call a macroscope (from the Greek macros, or “great,” and skopein, or “to observe”) inspired by de Rosnay’s futurist science writings. Macrosopes provide a “vision of the whole,” helping us “synthesize” the related elements and enabling us to detect patterns, trends, and outliers while granting access to myriad details. Rather than make things larger or smaller, macrosopes let us observe what is at once too great, slow, or complex for the human eye and mind to notice and comprehend.

While microscopes and telescopes are physical instruments, macrosopes resemble continuously changing bundles of software plug-ins. Macrosopes make it easy to select and combine algorithm and tool plug-ins but also interface plug-ins, workflow support, logging, scheduling, and other plug-ins needed for scientifically rigorous yet effective work.

They make it easy to share plug-ins via email, flash drives, or online. To use new plugins, simply copy the files into the plug-in directory, and they appear in the tool menu ready for use. No restart of the tool is necessary. Sharing algorithm components, tools, or novel interfaces becomes as easy as sharing images on Flickr or videos on YouTube. Assembling custom tools is as quick as compiling your custom music collection.
CIShell (http://cishell.org) is an open source software specification for the integration and utilization of datasets, algorithms, and tools.

It extends the Open Services Gateway Initiative (OSGi) (http://osgi.org), a standardized, component oriented, computing environment for networked services widely used in industry since more than 10 years.

Specifically, CIShell provides “sockets” into which existing and new datasets, algorithms, and tools can be plugged using a wizard-driven process.
About the Cyberinfrastructure Shell

The Cyberinfrastructure Shell (CfShell) is an open source, community-driven platform for the integration and utilization of datasets, algorithms, tools, and computing resources. Algorithms integration support is built in for Java and most other programming languages. Being Java based, it will run on almost all platforms. The software and specification is released under an Apache 2.0 License.

CfShell is the basis of Network Workbench, TexTrend, SciT and the upcoming Epic tool.

CfShell supports remote execution of algorithms. A standard web service definition is in development that will allow pools of algorithms to transparently be used in a peer-to-peer, client-server, or web front-end fashion.

CfShell Features

A framework for easy integration of new and existing algorithms written in any programming language

Using CfShell, an algorithm writer can fully concentrate on creating their own algorithm in whatever language they are comfortable with. Simple tools are provided to then take their algorithm and

Getting Started...

- Documentation & Developer Resources
- Download

Getting Involved...

- Contact Us
The Network Workbench (NWB) tool supports researchers, educators, and practitioners interested in the study of biomedical, social and behavioral science, physics, and other networks.

In February 2009, the tool provides more than 169 plugins that support the preprocessing, analysis, modeling, and visualization of networks.

It has been downloaded more than 110,000 times since December 2006.

---

Computational Proteomics
What relationships exist between protein targets of all drugs and all disease-gene products in the human protein–protein interaction network?


---

Computational Economics
Does the type of product that a country exports matter for subsequent economic performance?


Computational Social Science
Studying large scale social networks such as Wikipedia

Computational Epidemics
Forecasting (and preventing the effects of) the next pandemic.


Sci² Tool – “Open Code for S&T Assessment”
Users come from 40+ countries

OSGi/CIShell powered tool with NWB plugins and many new scientometrics and visualizations plugins.

Sci Maps

GUESS Network Vis

Sci² Tool Vis cont.

**Sci² Tool**

File | Preprocessing | Modeling | Analysis
--- | --- | --- | ---

**Visualization** | **Scientometrics** | **Help**
--- | --- | ---
GUESS
GnuPlot
Radial Tree/Graph (prefuse alpha)
Radial Tree/Graph with Annotation (prefuse beta)
Tree View (prefuse beta)
Tree Map (prefuse beta)
Force Directed with Annotation (prefuse beta)
Fruchterman-Reingold with Annotation (prefuse beta)
Erl. (View)
Specified (prefuse beta)
Horizontal Line Graph
Circular Hierarchy
Geo Map (circle annotations)
Geo Map (region coloring annotations)
Image Viewer
RefViewer

---

75

---

76
A number of other projects recently adopted OSGi and/or CIShell:

- **Cytoscape** ([http://cytoscape.org](http://cytoscape.org)) Led by Trey Ideker at the University of California, San Diego is an open source bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data (Shannon et al., 2002).

- **MAEviz** ([https://wiki.ncsa.uiuc.edu/display/MAE/Home](https://wiki.ncsa.uiuc.edu/display/MAE/Home)) Managed by Jong Lee at NCSA is an open-source, extensible software platform which supports seismic risk assessment based on the Mid-America Earthquake (MAE) Center research.

- **Taverna Workbench** ([http://taverna.org.uk](http://taverna.org.uk)) Developed by the myGrid team led by Carol Goble at the University of Manchester, U.K. is a free software tool for designing and executing workflows (Hull et al., 2006). Taverna allows users to integrate many different software tools, including over 30,000 web services.

- **TEXTrend** ([http://textrend.org](http://textrend.org)) Led by George Kampis at Eötvös Loránd University, Budapest, Hungary supports natural language processing (NLP), classification/mining, and graph algorithms for the analysis of business and governmental text corpuses with an inherently temporal component.

- **DynaNets** ([http://www.dynanets.org](http://www.dynanets.org)) Coordinated by Peter M.A. Sloot at the University of Amsterdam, The Netherlands develops algorithms to study evolving networks.


As the functionality of OSGi-based software frameworks improves and the number and diversity of dataset and algorithm plugins increases, the capabilities of custom tools will expand.
All papers, maps, tools, talks, press are linked from http://cns.iu.edu

CNS Facebook: http://www.facebook.com/cnscenter
Mapping Science Exhibit Facebook: http://www.facebook.com/mappingscience