Designing Effective Data Visualizations

Katy Börner
Professor of Information Science
Director, Cyberinfrastructure for Network Science Center
School of Informatics and Computing and Indiana University Network Science Institute
Indiana University, USA

Emory University
(Atwood Chemistry Center, Room 316)
February 4, 2016

The Power of Data Visualizations

Find your way

Find collaborators, friends

Identify trends

Terra bytes of data

Descriptive & Predictive Models
Mapping the Evolution of Co-Authorship Networks
Ke, Visvanath & Börner. 2004. Won 1st prize at the IEEE InfoVis Contest.
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.

Supported by NIH/NCI Contract HHSN261200800812
The Global 'Scientific Food Web'


Contributions:
Comprehensive global analysis of scholarly knowledge production and diffusion on the level of continents, countries, and cities.

Quantifying knowledge flows between 2000 and 2009, we identify global sources and sinks of knowledge production. Our knowledge flow index reveals, where ideas are born and consumed, thereby defining a global 'scientific food web'.

While Asia is quickly catching up in terms of publications and citation rates, we find that its dependence on knowledge consumption has further increased.
Places & Spaces: Mapping Science Exhibit
Places & Spaces: Mapping Science Exhibit, online at http://scimaps.org

Clickstream Map of Science

This is the first map created from large-scale, automatically, including usage data. It visualizes the collective flow of scientists' movements from one journal to another in their online navigation behavior.

Data: WOS 2004 - 2007


Language Communities of Twitter

Language Communities of Twitter - Eric Fischer - 2012
Places & Spaces at Northwestern University
May 14 - September 23, 2015

Places & Spaces Exhibit at the David J. Sencer CDC Museum, Atlanta, GA
January 25-June 17, 2016
Illuminated Diagram Display on display at the Smithsonian in DC. http://scimaps.org/exhibit_info/#ID
Science Maps in “Expedition 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.expedition-zukunft.de
Places & Spaces Digital Display in North Carolina State’s brand new Immersion Theater

Ingo Gunther’s Worldprocessor globe design on display at the Giant Geo Cosmos OLED Display at the Museum of Emerging Science and Innovation in Tokyo, Japan

http://scimaps.org/call

Themes for the upcoming iterations/years are:
- 11th Iteration (2015): Macroscopes for Interacting With Science
- 12th Iteration (2016): Macroscopes for Making Sense of Science
- 13th Iteration (2017): Macroscopes for Forecasting Science
- 14th Iteration (2018): Macroscopes for Economic Decision Makers
- 16th Iteration (2020): Macroscopes for Scholars

Plug-and-Play Macroscopes
Microscopes, Telescopes, Macroscopes

The Infinitely Great

Telescope

Stars

Galaxies

Society

The Infinitely Complex

NATURE

Technology

The Infinitely Small

Microscope

Cells

Macroscopes

Different datasets/formats.
Diverse algorithms/tools written in many programming languages.

Plug-and-Play Macroscopes

IS

CS

Bio

SNA

Physics
Plug-and-Play Macroscopes

Common algorithm/tool pool
Easy way to share new algorithms
Workflow design logs
Custom tools

EpiC
Converters
Sci2
NWB

TexTrend

IS
CS
Bio
SNA
Phys
Information Visualization Framework
&
IVMOOC

Tasks

Levels

Micro: Individual Level
about 4,000 records
page 5

Mesoscale: Local Level
about 1,000–50,000 records
page 8

Macro: Global Level
more than 100,000 records
page 30

Types

- Statistical Analysis
- Temporal Analysis
- Geospatial Analysis
- Textual Analysis
- Network Analysis

See page 5
Visualization Framework

<table>
<thead>
<tr>
<th>Insight Need Types</th>
<th>Data Scale Types</th>
<th>Visualization Types</th>
<th>Graphic Symbol Types</th>
<th>Graphic Variable Types</th>
<th>Interaction Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>page 26</td>
<td>page 28</td>
<td>page 30</td>
<td>page 32</td>
<td>page 34</td>
<td>page 26</td>
</tr>
<tr>
<td>• categorize/clusters</td>
<td>• nominal</td>
<td>• table</td>
<td>• geometric symbols</td>
<td>• spatial</td>
<td></td>
</tr>
<tr>
<td>• order/rank/sort</td>
<td>• ordinal</td>
<td>• chart</td>
<td>• point</td>
<td>• position</td>
<td></td>
</tr>
<tr>
<td>• distributions</td>
<td>• interval</td>
<td>• graph</td>
<td>• line</td>
<td>• retrieved</td>
<td></td>
</tr>
<tr>
<td>(also outliers, gaps)</td>
<td>• ratio</td>
<td>• map</td>
<td>• area</td>
<td>• form</td>
<td></td>
</tr>
<tr>
<td>• comparisons</td>
<td>• network layout</td>
<td>• surface</td>
<td>• volume</td>
<td>• color</td>
<td></td>
</tr>
<tr>
<td>• trends</td>
<td></td>
<td>• linguistic symbols</td>
<td>• number</td>
<td>• optics</td>
<td></td>
</tr>
<tr>
<td>• process and time</td>
<td></td>
<td>• punctuation marks</td>
<td>• text</td>
<td>• motion</td>
<td></td>
</tr>
<tr>
<td>• geospatial</td>
<td></td>
<td>• pictorial symbols</td>
<td>• numerical</td>
<td>• rotation</td>
<td></td>
</tr>
<tr>
<td>• compositions</td>
<td></td>
<td>• images</td>
<td>• icons</td>
<td>• disturbance</td>
<td></td>
</tr>
<tr>
<td>• also of text</td>
<td></td>
<td>• statistical glyphs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• correlations/relationships</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

See page 24

Visualization Framework

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>selection</td>
<td>categorize</td>
<td>table</td>
<td>compare</td>
<td>compare</td>
<td>tool</td>
<td>categorize/cluster</td>
<td>order/rank/sort</td>
<td></td>
</tr>
<tr>
<td>order</td>
<td>rank</td>
<td>distribution</td>
<td>compare</td>
<td>compare</td>
<td>tool</td>
<td>distributions</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>and contrast</td>
<td>data values</td>
<td>tool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>time series</td>
<td>compare</td>
<td>tool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>patterns</td>
<td>and contrast</td>
<td>tool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>time over time</td>
<td></td>
<td>tool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>process and time</td>
<td></td>
<td>tool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>geospatial</td>
<td>generate</td>
<td>tool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>spatial</td>
<td>maps</td>
<td>tool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>relations</td>
<td></td>
<td>tool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>location</td>
<td></td>
<td>tool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>part-to-whole</td>
<td></td>
<td>tool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>proportions</td>
<td></td>
<td>tool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>form and structure</td>
<td></td>
<td>tool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>hierarchy</td>
<td></td>
<td>tool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>associations</td>
<td></td>
<td>tool</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

See page 24
### Visualization Framework

<table>
<thead>
<tr>
<th>Insight Need Types</th>
<th>Data Scale Types</th>
<th>Visualization Types</th>
<th>Graphic Symbol Types</th>
<th>Graphic Variable Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>page 26</td>
<td>page 28</td>
<td>page 30</td>
<td>page 32</td>
<td>page 14</td>
</tr>
<tr>
<td>• category/corpus</td>
<td>• nominal</td>
<td>• table</td>
<td>• geometric symbols</td>
<td>• spatial position</td>
</tr>
<tr>
<td>• order/indirect</td>
<td>• ordinal</td>
<td>• chart</td>
<td>• point</td>
<td>• retinal form</td>
</tr>
<tr>
<td>• distributions</td>
<td>• interval</td>
<td>• graph</td>
<td>• line</td>
<td>• color</td>
</tr>
<tr>
<td>(also outliers, gaps)</td>
<td>• ratio</td>
<td>• map</td>
<td>• area</td>
<td>• optics</td>
</tr>
<tr>
<td>• comparisons</td>
<td>• network layout</td>
<td>• surface</td>
<td>• volume</td>
<td>• motion</td>
</tr>
<tr>
<td>• trends</td>
<td></td>
<td>• lingustic symbols</td>
<td>• text</td>
<td></td>
</tr>
<tr>
<td>(process and time)</td>
<td></td>
<td>• numerals</td>
<td>• punctuation marks</td>
<td></td>
</tr>
<tr>
<td>• geospatial</td>
<td></td>
<td>• pictorial symbols</td>
<td>• images</td>
<td></td>
</tr>
<tr>
<td>• compositions</td>
<td></td>
<td>• icons</td>
<td>• statistical glyphs</td>
<td></td>
</tr>
<tr>
<td>(also of text)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• correlation/relationships</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

### Graphic Variable Types Versus Graphic Symbol Types

<table>
<thead>
<tr>
<th>Variables</th>
<th>Geometric Symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Z</td>
<td></td>
</tr>
<tr>
<td>Color</td>
<td></td>
</tr>
<tr>
<td>Shape</td>
<td></td>
</tr>
<tr>
<td>Rotation</td>
<td></td>
</tr>
<tr>
<td>Num</td>
<td></td>
</tr>
<tr>
<td>Value</td>
<td></td>
</tr>
<tr>
<td>Class</td>
<td></td>
</tr>
<tr>
<td>Saturation</td>
<td></td>
</tr>
</tbody>
</table>

See page 24
Course Schedule

Part 1: Theory and Hands-On
- Session 1 – Workflow Design and Visualization Framework
- Session 2 – “When:” Temporal Data
- Session 3 – “Where:” Geospatial Data
- Session 4 – “What:” Topical Data

Mid-Term
- Session 5 – “With Whom:” Trees
- Session 6 – “With Whom:” Networks
- Session 7 – Dynamic Visualizations and Deployment

Final Exam

Part 2: Students work in teams on client projects.

Final grade is based on Class Participation (10%), Midterm (30%), Final Exam (30%), and Client Project(30%).
Needs-Driven Workflow Design

**Stakeholders**

**Validation**

**Interpretation**

Types and levels of analysis determine data, algorithms & parameters, and deployment

**Data**

**READ**

**ANALYZE**

**VISUALIZE**

**DEPLOY**

Visually encode data

Overlay data

Select visualization type

Load **One** File and Run **Many** Analyses and Visualizations

<table>
<thead>
<tr>
<th>Times Cited</th>
<th>Publication Year</th>
<th>City of Publisher</th>
<th>Country</th>
<th>Journal Title (Full)</th>
<th>Title</th>
<th>Subject Category</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>2011</td>
<td>NEW YORK</td>
<td>USA</td>
<td>COMMUNICATIONS OF THE ACM</td>
<td>Plug-and-Play Macroscopes</td>
<td>Computer Science</td>
<td>Borner, K</td>
</tr>
<tr>
<td>18</td>
<td>2010</td>
<td>MALDEN</td>
<td>USA</td>
<td>CTS-CLINICAL AND TRANSLATIONAL SCIENCE</td>
<td>Advancing the Science of Team Science</td>
<td>Research &amp; Experimental Medicine</td>
<td>Falk-Kraenzlein, HJ</td>
</tr>
</tbody>
</table>
Load **One** File and Run **Many** Analyses and Visualizations

<table>
<thead>
<tr>
<th>Times Cited</th>
<th>Publication Year</th>
<th>City of Publisher</th>
<th>Country</th>
<th>Journal Title (Full)</th>
<th>Title</th>
<th>Subject Category</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>2011</td>
<td>NEW YORK</td>
<td>USA</td>
<td>COMMUNICATIONS OF THE ACM</td>
<td>Plug-and-Play Macrosopes</td>
<td>Computer Science</td>
<td>Borner, K</td>
</tr>
<tr>
<td>18</td>
<td>2010</td>
<td>MALDEN</td>
<td>USA</td>
<td>CTS-CLINICAL AND TRANSLATIONAL SCIENCE</td>
<td>Advancing the Science of Team Science</td>
<td>Research &amp; Experimental Medicine</td>
<td>Falk-Krauszinski, HJ</td>
</tr>
<tr>
<td>13</td>
<td>2010</td>
<td>WASHINGTON</td>
<td>USA</td>
<td>A Multi-Level Systems Perspective for the Science of Team Science</td>
<td>Call Biology</td>
<td>Research &amp; Experimental Medicine</td>
<td>Borner, K</td>
</tr>
</tbody>
</table>

Co-author and many other bi-modal networks.

---

**Course Schedule**

**Part 1: Theory and Hands-On**
- **Session 1** – Workflow Design and Visualization Framework
- **Session 2** – “When:” Temporal Data
- **Session 3** – “Where:” Geospatial Data
- **Session 4** – “What:” Topical Data

**Mid-Term**
- **Session 5** – “With Whom:” Trees
- **Session 6** – “With Whom:” Networks
- **Session 7** – Dynamic Visualizations and Deployment

**Final Exam**

**Part 2: Students work in teams on client projects.**

Final grade is based on Class Participation (10%), Midterm (30%), Final Exam (30%), and Client Project(30%).
IVMOOC App – More than 60 visualizations

The “IVMOOC Flashcards” app can be downloaded from Google Play and Apple iOS stores.

References


All papers, maps, tools, talks, press are linked from http://cns.iu.edu
These slides are at http://cns.iu.edu/docs/presentations
CNS Facebook: http://www.facebook.com/cnscenter
Mapping Science Exhibit Facebook: http://www.facebook.com/mappingscience