The Information Visualization MOOC

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

Exploiting Big Data Semantics for Translational Medicine Workshop
Indiana University Bloomington, Indiana Memorial Union

March 25-26, 2013

http://scimaps.org/maps/map/khan_academy_library_147
MOOCs

In 2012, Google hosted three massive open online courses (MOOCs) collectively reaching over 400,000 registrants.

By the end of 2013 more than 250 courses will be run using the Google, Coursera, Udacity, EdX, and other platforms.
The Information Visualization MOOC
ivmooc.cns.iu.edu

Students come from 93 countries
300+ faculty members
#ivmooc
Instructors

Katy Börner – Theory Parts
Instructor, Professor at SLIS

David E. Polley – Hands-on Parts
CNS Staff, Research Assistant with MIS/MLS
Teaches & Tests Sci2 Tool

Scott B. Weingart – Client Work
Assistant Instructor, SLIS PhD student

Course Schedule

Course started on January 22, 2013
• Session 1 – Workflow design and visualization framework
• Session 2 – “When:” Temporal Data
• Session 3 – “Where:” Geospatial Data
• Session 4 – “What:” Topical Data

Mid-Term

Students work in teams with clients.
• Session 5 – “With Whom:” Trees
• Session 6 – “With Whom:” Networks
• Session 7 – Dynamic Visualizations and Deployment

Final Exam
Different Question Types

- Find your way
- Find collaborators, friends
- Identify trends

Terabytes of data

Descriptive & Predictive Models

Plug-and-Play Macroscopes
cishell.org

Unit Structure

The course and each unit has three components:

**Theory**: Videos and Slides
Self-Assessment (not graded)

**Hands-on**: Videos and Slides & Wiki pages with workflows
Homework (not graded)

**Client Work**: Using Drupal Marketplace (peer review)

Theory Unit Structure

Each theory unit comprises:
- Examples of best visualizations
- Visualization goals
- Key terminology
- General visualization types and their names

- Workflow design
  - Read data
  - Analyze
  - Visualize

- Discussion of specific algorithms
### Different Levels of Abstraction/Analysis

- **Macro/Global Population Level**
  - Population-level analysis
  - Focuses on large-scale data
  - Examples include national funding trends, state-level economic indicators

- **Meso/Local Group Level**
  - Group-level analysis
  - Focuses on medium-sized data sets
  - Examples include city-level housing trends, community healthcare patterns

- **Micro Individual Level**
  - Individual-level analysis
  - Focuses on small data sets
  - Examples include personal health records, individual academic publications

### Type of Analysis vs. Level of Analysis

<table>
<thead>
<tr>
<th></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 physics 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>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><strong>Network Analysis (With Whom?)</strong></td>
<td>NSF Co-PI network of one individual</td>
<td>Co-author network</td>
<td>NIH’s core competency</td>
</tr>
</tbody>
</table>
### Type of Analysis vs. Level of Analysis

<table>
<thead>
<tr>
<th>Level of Analysis</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 physics research</td>
</tr>
<tr>
<td><strong>Geospatial Analysis (Where)</strong></td>
<td>Career trajectory of one individual</td>
<td>Mapping a states intellectual landscape</td>
<td>PNAS publications</td>
</tr>
<tr>
<td><strong>Topical Analysis (What)</strong></td>
<td></td>
<td>Knowledge flow in chemistry research</td>
<td></td>
</tr>
<tr>
<td><strong>Network Analysis (With Whom?)</strong></td>
<td>NSF co-author network of one individual</td>
<td>NIH’s core competency</td>
<td></td>
</tr>
</tbody>
</table>

### Needs-Driven Workflow Design

**Stakeholders**

**Data**

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

**Read**

**Analyze**

**Visualize**

**Deploy**

- Visually encode data
- Overlay data
- Select visualization type
- Modify reference system, add records & links
- Graphic variable types

**Validation**

**Interpretation**
**Needs-Driven Workflow Design**

- **Stakeholders**
  - Validation
  - Interpretation

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

- **Data**
  - **READ**
  - **ANALYZE**

- **DEPLOY**
  - Visually encode data
  - Overlay data
  - Select visualiz. type

**Needs-Driven Workflow Design**

- **Stakeholders**
  - Validation
  - Interpretation

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

- **Data**
  - **READ**
  - **ANALYZE**

- **DEPLOY**
  - Visually encode data
  - Overlay data
  - Select visualiz. type
### Visualization Types vs. Data Overlays

<table>
<thead>
<tr>
<th>Visualization Type</th>
<th>Chart</th>
<th>Table</th>
<th>Graph</th>
<th>Geospatial Map</th>
<th>Network Graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modify / <strong>visually encode</strong> base map.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Place and <strong>visually encode</strong> records/nodes.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Place and <strong>visually encode</strong> links.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Plus, add a title, labels, legend, explanatory text, and author info.
IVMOOC Social Media Stream

Before, during, and after the course, please use tag “ivmooc” on

- **Twitter** to share links to insightful visualizations, conferences and events, or relevant job openings.
- **Flickr** to upload your own visualizations or tag visualizations by others.

We hope to use this course to create a unique, real-time data stream of the best visualizations, experts, and companies that apply data mining and visualization techniques to answer real-world questions.

Grading

All students are asked to create a personal profile to support working in teams.

Final grade is based on Midterm (30%), Final (40%), Client Project (30%).

- Weekly self-assessments are not graded.
- Homework is graded automatically.
- Midterm and Final test materials from theory and hands-on sessions are graded automatically.
- Client work is peer-reviewed via online forum.

All students that receive more than 80% of all available points get an official certificate/badge.
Sandra M. Chung

How to Read this Map
This proportional symbol map shows 50 U.S. states and other jurisdictions using the Albers equal-area conic projection with Alaska, Puerto Rico, and Hawaii inset. Each distinct record is represented by a circle centered at its projection. The area, interior color, and exterior color of each circle correspond with the values. Minimum and maximum data values are given in the legend.


Diogo Carmo

Innovation & Entrepreneurship
NSF Funding Across the US, from 1972 to 2009, and Current Web Interest

Legend
- Interior Color (Linear)
  - 10
  - 15
  - 25
  - 30

U.S. State Color (Linear)
- Number of records
- 10,000
- 15,000
- 20,000
- 25,000
- 30,000

How to Read this Map
This proportional symbol map shows 50 U.S. states and other jurisdictions using the Albers equal-area conic projection with Alaska, Puerto Rico, and Hawaii inset. Each distinct record is represented by a circle centered at its projection. The area, interior color, and exterior color of each circle correspond with the values. Minimum and maximum data values are given in the legend.


**NSF Funding - Graphene Projects 2004-2010**

**Legend / Key**
- Number of times funding has been awarded:
  - 1
  - 2
  - 3
  - 4
- Total amount funded ($'000s)
  - $1600
  - $15

**What is Graphene?**
Graphene is a two-dimensional material consisting of a single layer of carbon atoms arranged in a honeycomb or chicken wire structure. It is the thinnest material known and is also one of the strongest. It conducts electricity as effectively as copper and is more effective than some other materials as a conductor of heat. Graphene is almost completely transparent, yet so dense that even the smallest atom reflects over 99% of its light.

Originally thought to be unstable, it was first observed in 2004 by Andre Geim and Kostya Novoselov at the University of Manchester in 2004. The results of this work, which were published in 2004, heralded a new dawn in the study of two-dimensional materials and of graphene in particular.

**How to read this map**
The map shows NSF funding awards to US institutions. Each circle corresponds to an institution. The depth of the circle represents the number of times funding was awarded to the same institution for different projects. The color of the circle indicates the amount awarded ($'000s).

**IVMOOC Twitter Network**

![Twitter Network Diagram]
References


http://scimaps.org/atlas

Acknowledgments

We would like to thank Miguel I. Lara and his colleagues at the Center for Innovative Teaching and Learning for instructional design support, Samuel Mills for designing the web pages, Robert P. Light and Thomas Smith for extending the GCB platform, and Mike Widmer and Mike T. Gallant for adding the Forum. Support comes from CNS, CITL, SLIS, SOIC, and Google.

The tool development work is supported in part by the Cyberinfrastructure for Network Science Center and the School of Library and Information Science at Indiana University, the National Science Foundation under Grants No. SBE-0738111 and IIS-0513650, the US Department of Agriculture, the National Institutes of Health, and the James S. McDonnell Foundation.

Visualizations used in the course come from the Places & Spaces: Mapping Science exhibit, online at http://scimaps.org, and from the Atlas of Science: Visualizing What We Know, MIT Press (2010).