

# The Information Visualization MOOC

**Katy Börner**

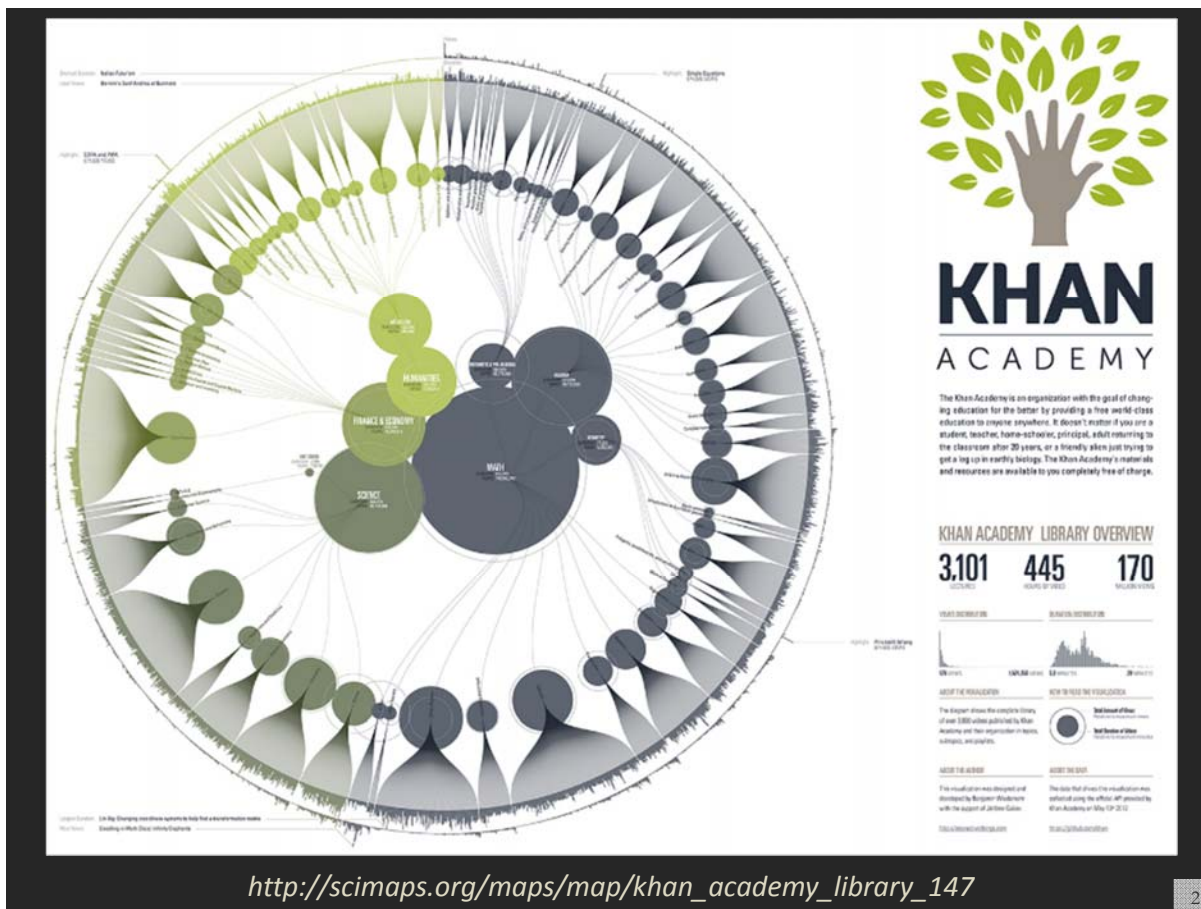
Cyberinfrastructure for Network Science Center, Director  
Information Visualization Laboratory, Director  
School of Library and Information Science  
Indiana University, Bloomington, IN  
[katy@indiana.edu](mailto:katy@indiana.edu)



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*

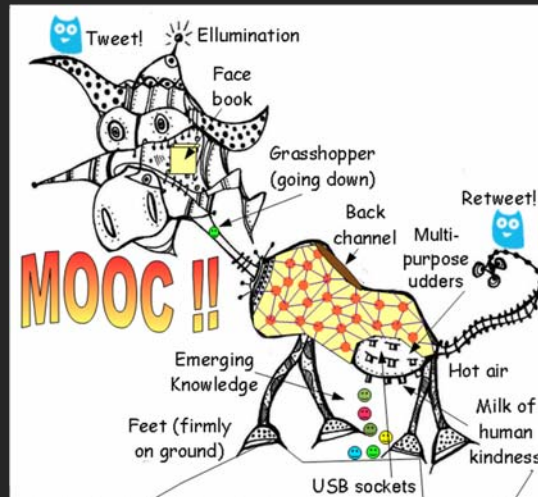
*March 25-26, 2013*



# MOOCs

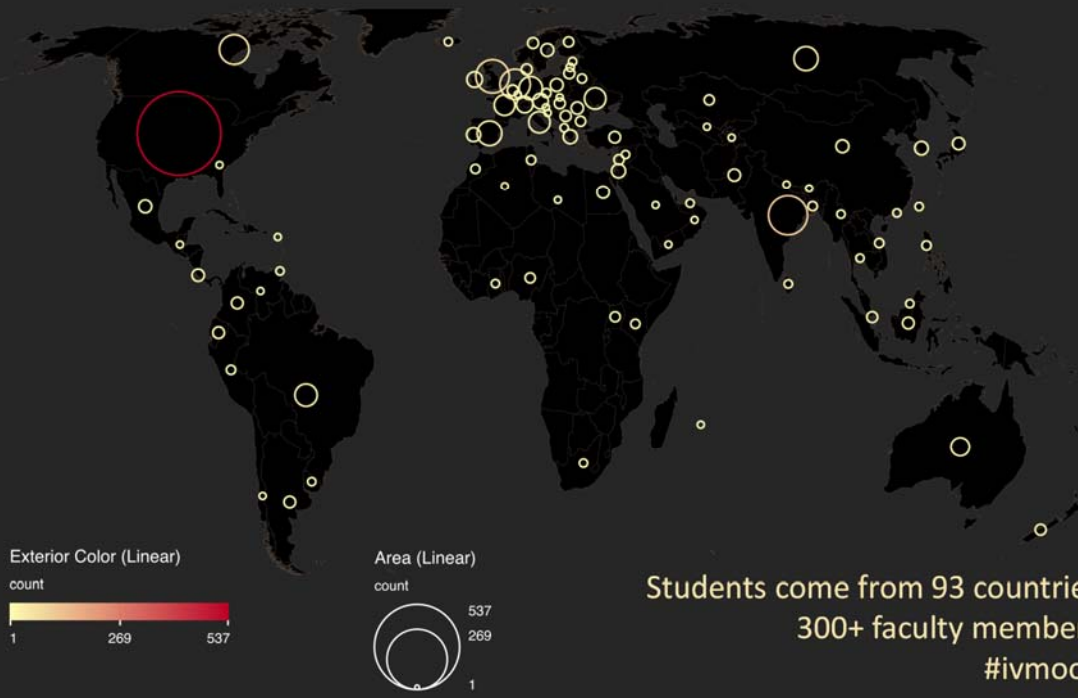
In 2012, Google hosted three massive open online courses (MOOCs) collectively reaching over 400,000 registrants.

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



## Information Visualization MOOC

INDIANA UNIVERSITY CNS



### Overview

This course provides an overview about the state of the art in information visualization. It teaches the process of producing effective visualizations that take the needs of users into account.

Among other topics, the course covers:

- Data analysis algorithms that enable extraction of patterns and trends in data
- Major temporal, geospatial, topical, and network visualization techniques
- Discussions of systems that drive research and development.

Please watch the introduction video to get better acquainted with the course.

Everybody who registers gains free access to the Scholarly Database (26 million paper, patent, and grant records) and the Sci2 Tool (100+ algorithms and tools).

Katy Börner, Ph.D.  
Indiana University



Go To The Course

ivmooc.cns.iu.edu

## 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



7

## 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**

8



# Different Question Types



Terabytes of data

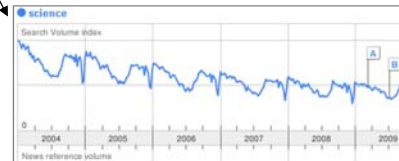
Descriptive & Predictive Models



Find your way

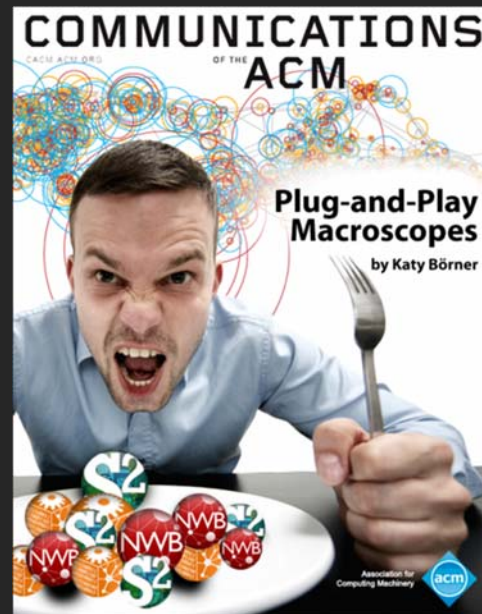


Find collaborators, friends



Identify trends

Plug-and-Play Macroscopes  
cishell.org



Börner, Katy. (March 2011). Plug-and-Play Macroscopes. *Communications of the ACM*, 54(3), 60-69. <http://www.scivee.tv/node/27704>

## 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)

11

## Theory Unit Structure

Each theory unit comprises:

- Examples of best visualizations
  - Visualization goals
  - Key terminology
  - General visualization types and their names
- 

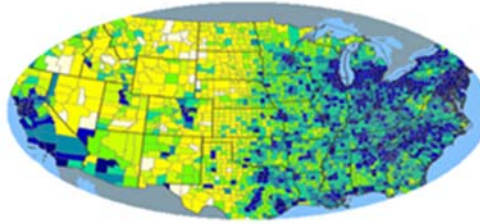
- Workflow design
    - Read data
    - Analyze
    - Visualize
- 

- Discussion of specific algorithms

12

## Different Levels of Abstraction/Analysis

Macro/Global  
Population Level



Meso/Local  
Group Level



Micro  
Individual Level



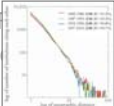
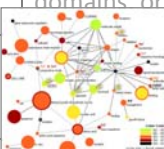


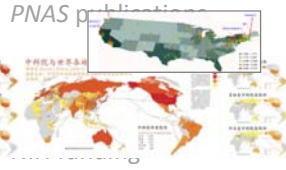

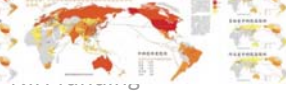



13

## Type of Analysis vs. Level of Analysis

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

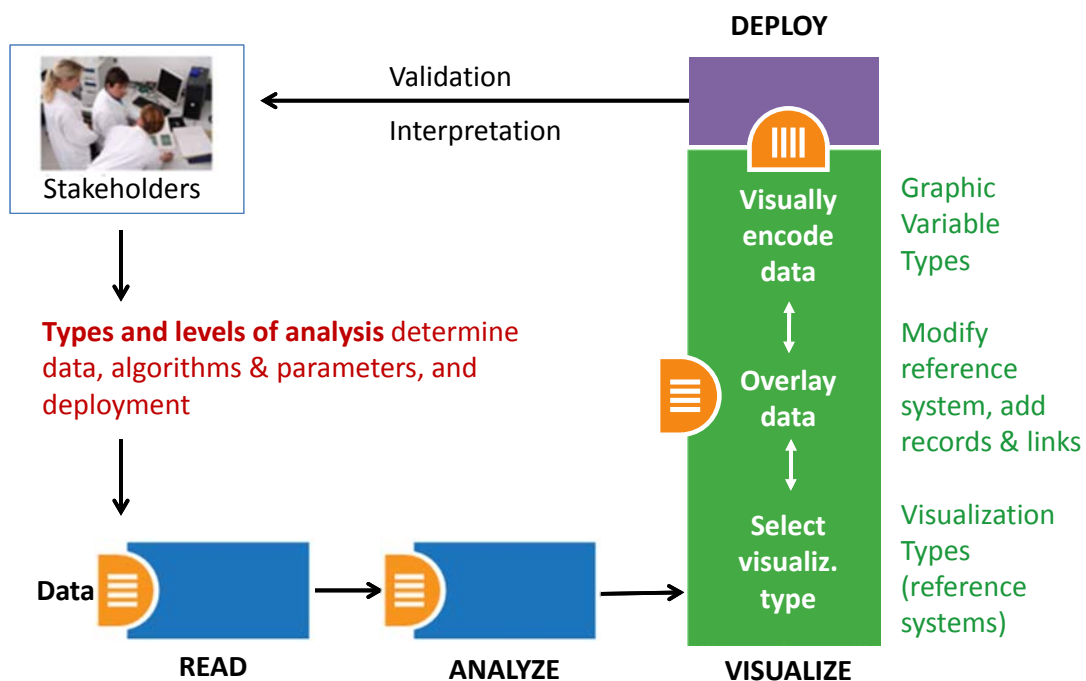
14

# Type of Analysis vs. Level of Analysis

	<b>Micro/Individual</b> (1-100 records)	<b>Meso/Local</b> (101-10,000 records)	<b>Macro/Global</b> (10,000 < records)
<b>Statistical Analysis/Profiling</b>	Individual person and their expertise profiles	Larger labs, centers, universities, research domains or states	All of NSF, all of science 
<b>Temporal Analysis (When)</b>	Funding portfolio of one individual	Topic bursts of PNAS 	113 years of PNAS research 
<b>Geospatial Analysis (Where)</b>	Career trajectory of one individual	Wrapping a state intellectual landscape 	PNAS publication 
<b>Topical Analysis (What)</b>	 s.	Knowledge flow in chemistry research	
<b>Network Analysis (With Whom?)</b>	NSF network of one 	NIH's network 	NIH's network 

15

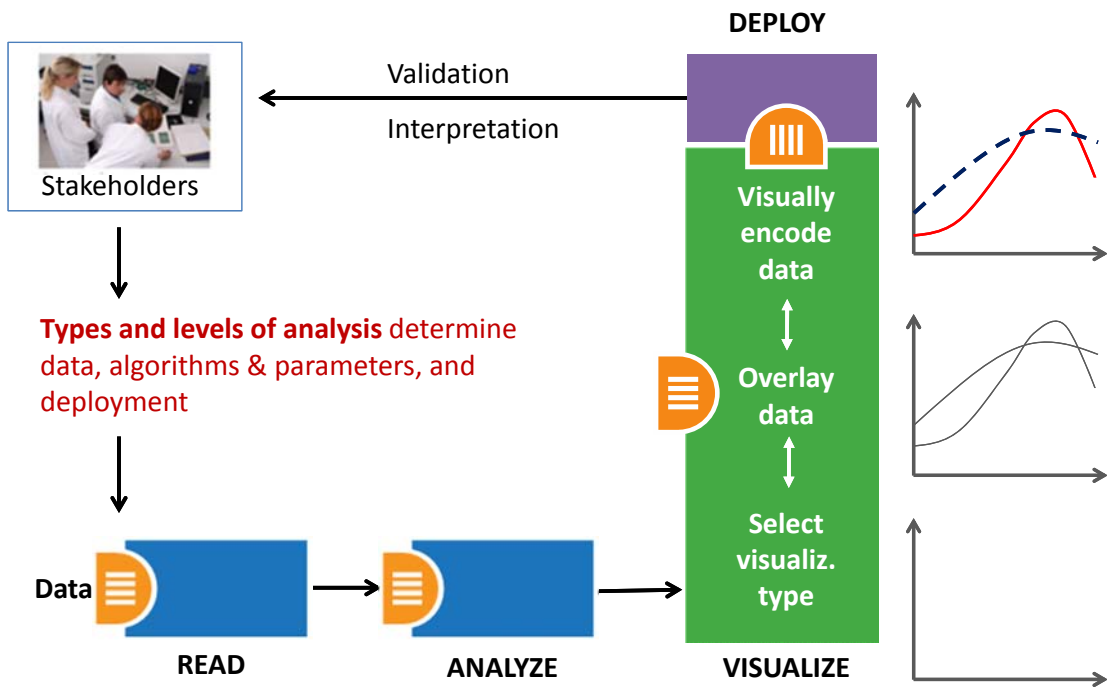
# Needs-Driven Workflow Design



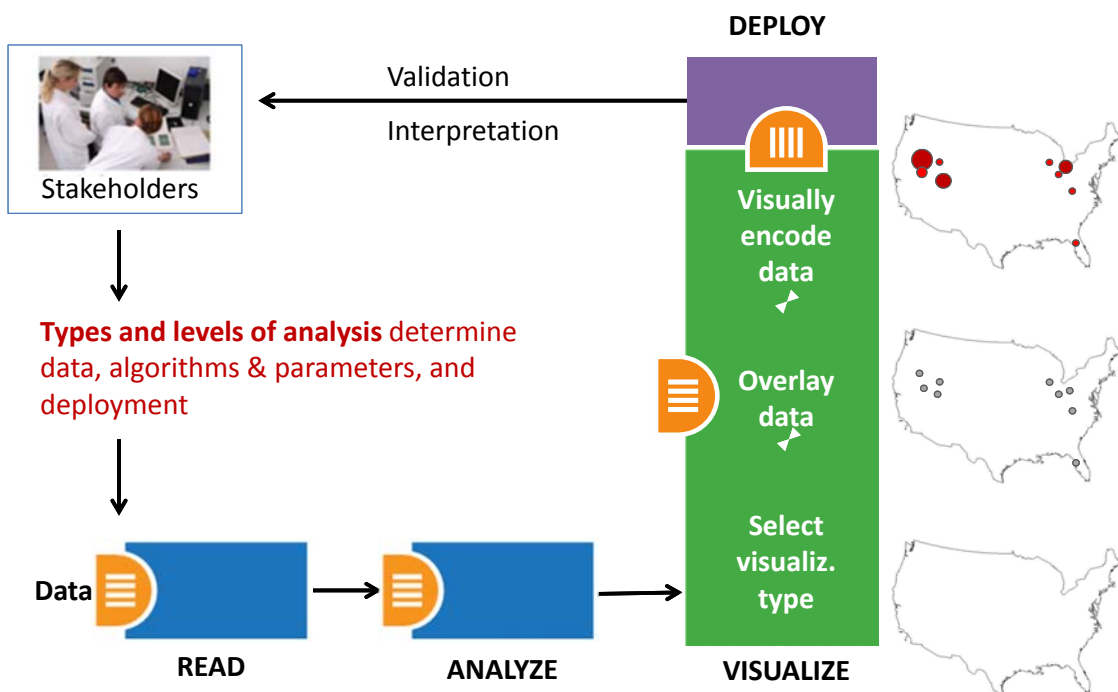
16






# Needs-Driven Workflow Design



# Needs-Driven Workflow Design





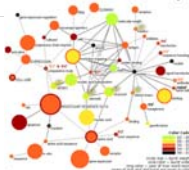
## Visualization Types vs. Data Overlays

Visualization Type	Chart	Table	Graph	Geospatial Map	Network Graph
Modify / <b>visually encode</b> base map.					
Place and <b>visually encode</b> records/nodes.					
Place and <b>visually encode</b> links.					

Plus, add a title, labels, legend, explanatory text, and author info.

19

## Visualization Types vs. Data Overlays

Visualization Type	Chart	Table	Graph	Geospatial Map	Network Graph
Modify / <b>visually encode</b> base map.					
Place and <b>visually encode</b> records/nodes.					
Place and <b>visually encode</b> links.					

Plus, add a title, labels, legend, explanatory text, and author info.

20

## 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.



21

## Grading

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



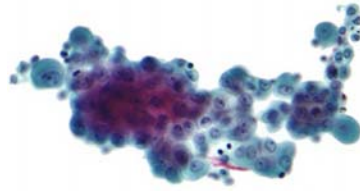
Final grade is based on Midterm (**30%**), Final (**40%**), Client Project (**30%**).

- 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.

22

Diogo Carmo



# Mesothelioma

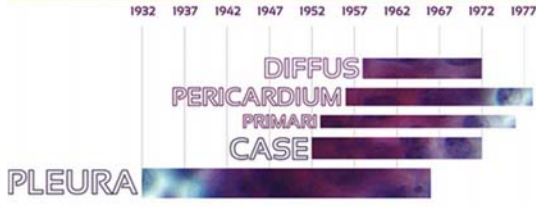
Main title topics in Medline papers

Mesothelioma is a rare form of cancer that develops from transformed cells originating in the mesothelium, the protective lining that covers many of the internal organs of the body. It is usually caused by exposure to asbestos.

The most common anatomical site for the development of mesothelioma is the pleura (the outer lining of the lungs and internal chest walls), but it can also arise in the peritoneum (the lining of the abdominal cavity), and the pericardium (the sac that surrounds the heart) or the tunica vaginalis (a sac that surrounds the testis).

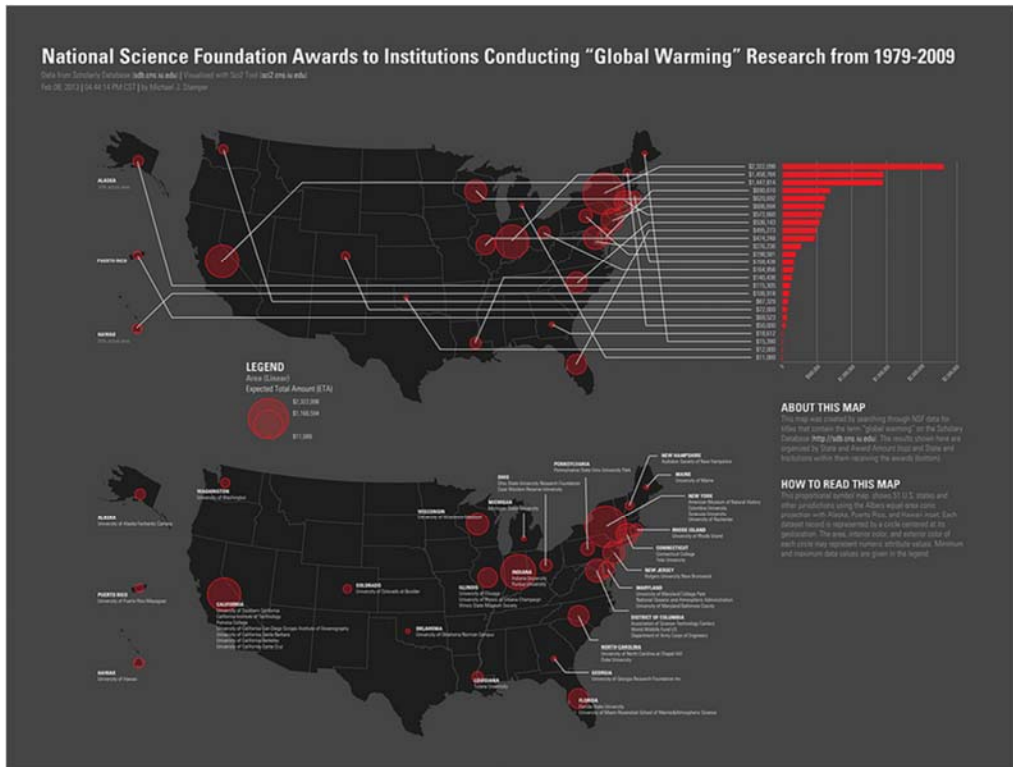
Most people who develop mesothelioma have worked in jobs where they inhaled asbestos, or were exposed to asbestos dust and fibers in other ways. It has also been suggested that working children of a family member who worked with asbestos increases their risk for developing mesothelioma. Unlike lung cancer, there seems to be no association between mesothelioma and tobacco smoking, but smoking greatly increases the risk of other asbestos-induced cancers. Some people who were exposed to asbestos have inherited damage to their asbestos-related disease, including mesothelioma. Compensation via asbestos funds or class action lawsuits is an important issue in law practices regarding mesothelioma.

# MALIGNANT PLEURAL CYSTIC BENIGN DIAGNOSIS



**How To Read This Map**  
This temporal bar graph visualizes diagnoses each record as a horizontal bar with a specific start and end date and a label on its left side. The area of each bar encodes the number of records and target magnitude - in the scope identified in the labels (target words were summed).

Author: Diogo Carmo <http://diogo.carmo@indiana.edu> | Visualization software: Sci2 Tools, (2009) Science of Science (Sci2) Tool, Indiana University and SciTech Strategies, <http://sci2.ucis.edu/> | Dataset: Medline Papers, as available in Pubmed Database <http://pubmed.ncbi.nlm.nih.gov/> | Text and Images: Wikipedia Mesothelioma article, available at <http://en.wikipedia.org/wiki/Mesothelioma> | Font: Sansation, by Bernd Mohr © 2011 - All Rights Reserved. This font family is licensed and available at <http://font Bundles.com>.

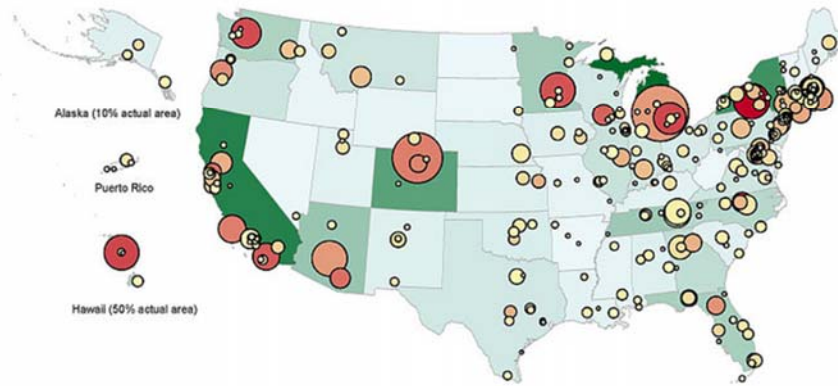


mjstamper ivmooc



NSF Awards 1976-2010 with "ecology" in the title

[Sandra M. Chung](#)



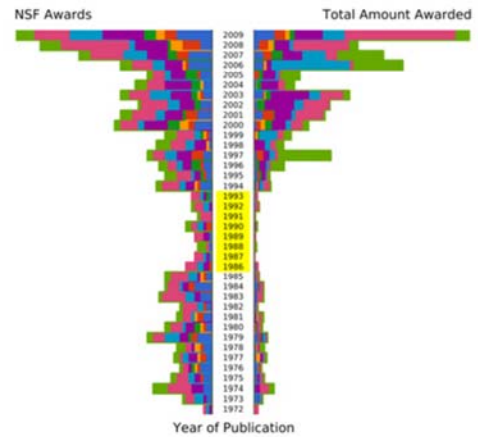
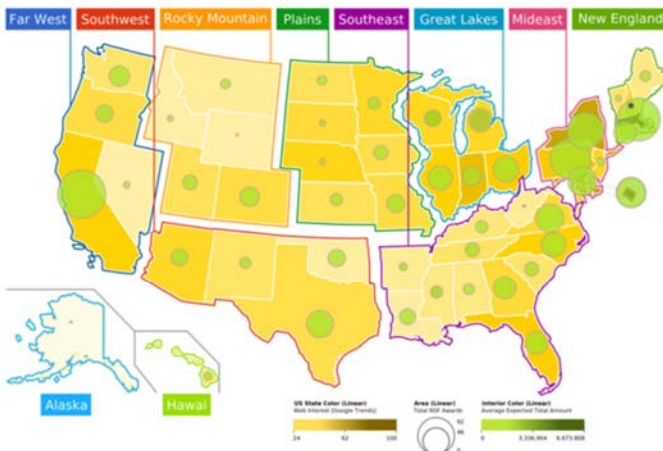
**How to Read this Map**

This proportional symbol map shows 52 U.S. states and other jurisdictions using the Albers equal-area conic projection with Alaska, Puerto Rico, and Hawaii inset. Each dataset record is represented by a circle centered at its geolocation. The area, interior color, and exterior color of each circle may represent numeric attribute values. Minimum and maximum data values are given in the legend.

Data retrieved from Scholarly Database (<http://sdb.cns.iu.edu/>).  
Choropleth generated by Sandra Chung (2013) using "Sci2".  
"Sci2 Team. (2009). Science of Science (Sci2) Tool. Indiana University and SciTech Strategies, <http://sci2.cns.iu.edu>.

**Innovation & Entrepreneurship**

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

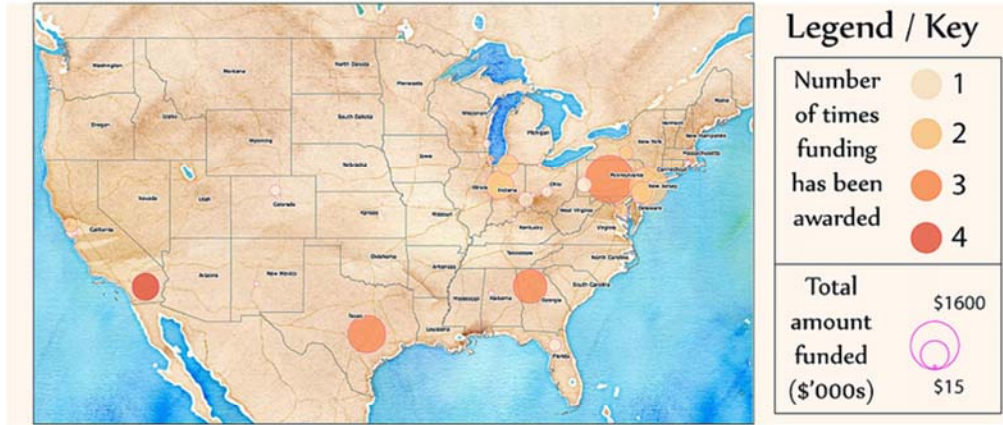


Author: Diogo Carmo (<http://diogo.carmo.org/>) Visualization software: Sci2 Team (2009). Science of Science (Sci2) Tool. Indiana University and SciTech Strategies, <http://sci2.cns.iu.edu/>; Google Sheets was used to produce the bar graphs; Dataset: National Science Foundation (NSF) Awards, as available in Scholarly Database <http://sdb.cns.iu.edu/>; Title - innovation (if entrepreneurship, Google Trends, for innovation (if entrepreneurship) [From: Sanjour, by Bernd Meinig © 2014 - All Rights Reserved. This font family is licensed and is available at <http://fontbundles.net/> and Dribbble, by Shutterstock, Inc. © 2014. All Rights Reserved. This font family is available at <http://fontbundles.net/>

[Diogo Carmo](#)

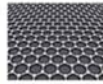
# NSF Funding - Graphene Projects 2004-2010

#ivmooc Week 3 homework @jonopatterson



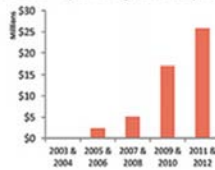
## What is Graphene?

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 yet is also one of the strongest. It conducts electricity as efficiently as copper and outperforms all other materials as a conductor of heat. Graphene is almost completely transparent, yet so dense that even the smallest atom helium cannot pass through it.



Originally thought to be unstable in its free state it proved to be quite the opposite when isolated by Andre Geim and Konstantin Novoselov at the University of Manchester in 2003. 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.

## Total NSF Spending on Graphene

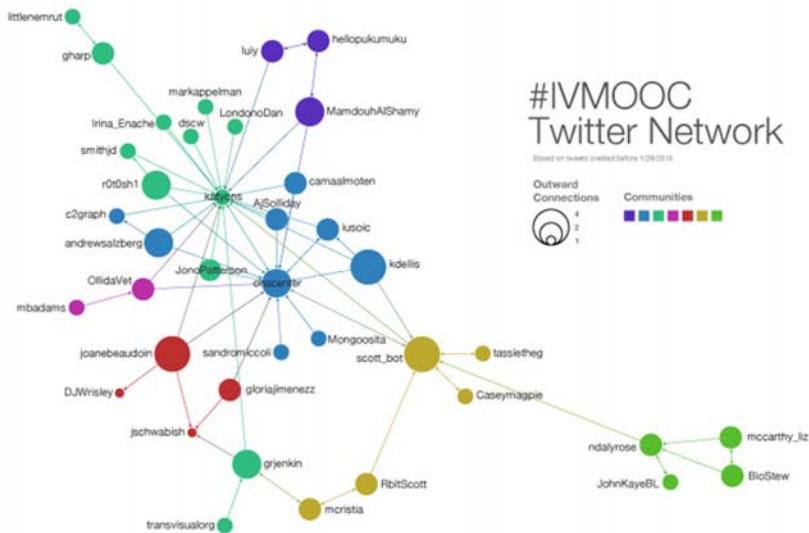


## How to read this map

The map shows NSF funding awards to US institutions. Each circle corresponds to an institution. The depth of colour represents the number of times funding was awarded to the same institution for different projects. The circle size indicates the amount awarded in \$'000s.

Sources: NSF, Manchester University  
Open Street Mapping by CC, Styled by Stamen

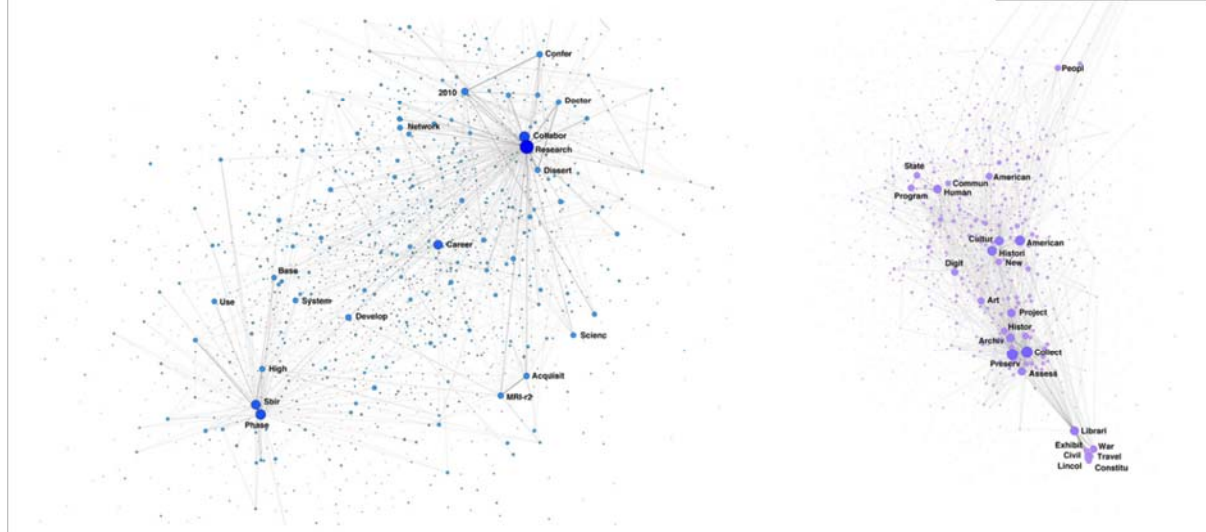
[JonoPatterson](#)



[camaal](#)

## NSF and NEH Funding (2010)

### Word Co-Occurrence in Grant Titles



Word Co-Occurrence in NSF and NEH Funded Grant Titles (2010)

By tzepei ★ Favorite □ 1 comment

29

## References

Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains**. In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Volume 37, Chapter 5, pp. 179-255.

<http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf>

Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl\_1).

Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007). **Network Science**. In Blaise Cronin (Ed.), *ARIST*, Information Today, Inc., Volume 41, Chapter 12, pp. 537-607. <http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf>

Börner, Katy (2010) **Atlas of Science**. MIT Press. <http://scimaps.org/atlas>

Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2012) **Models of Science Dynamics**. Springer Verlag.



30

## 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).

