Visual Analytics in Support of Data-Driven Decision Making

Katy Börner  @katycns

Victor H. Yngve Distinguished Professor of Intelligent Systems Engineering & Information Science Director, Cyberinfrastructure for Network Science Center School of Informatics, Computing, and Engineering Indiana University Network Science Institute (IUNI) Indiana University, Bloomington, IN, USA

Knowledge Network Talk, The Boeing Company

February 4, 2020
Overview

Intro to Data-Driven Decision Making Using (Interactive) Maps

- *Places & Spaces: Mapping Science* exhibit ([http://scimaps.org](http://scimaps.org)).

Data Visualization Literacy (DVL)


Visual Analytics Certificate

- Learn to render data into actionable insights in 6 weeks! Class begins **March 2, 2020**. Get course info at [https://boeing.cns.iu.edu](https://boeing.cns.iu.edu)

Learning Analytics

Data-Driven Decision Making Using (Interactive) Maps

http://scimaps.org
Data-Driven Decision Making

Most decision makers prefer

• orderly, predictable conditions,

• little disruption, and

• sufficient resources (e.g., money, talent, compassion)

to invent and implement desirable futures.

They want to understand the likely impact of decisions (e.g., hiring, purchasing, strategy changes) BEFORE writing a check and/or starting implementation.
Data Access and Actionable Visualizations

Decision makers have a deep interest in—and are willing to pay for—

• easy-to-use,
• near-real-time access to
• data, models, and visualizations

that help them make sense of, communicate, and proactively manage science, technology, and education and increase return on investment (ROI).

Visualizations that show the structure and dynamics of science, technology, and education are actively being researched and developed.
Use Cases: Industry

• Determine how to utilize limited resources to increase innovation and labor productivity; improve inventory turnover, asset utilization, supply-chain management, traffic optimization, error and attack tolerance.

• Research collaboration and workforce development decisions require knowing where the most productive research is being done and the best experts are trained, as well as how that production has changed over time and across individuals and institutions.

• Data-driven strategic planning, hiring, and resource allocation.

Use Cases: Government

• *The Promise of Evidence-Based Policymaking* report from 2017 opens with this statement: “The American people want a government that functions efficiently and responsibly addresses the problems that face this country.” The report argues for a future in which “rigorous evidence is created efficiently, as a routine part of government operations, and used to construct effective public policy.”

• Short-term goal: Improve access to data to improve the quantity and quality of evidence that informs important program and policy decisions.

• Longer-term goal: Simulate the impact of different policy decisions (e.g., alternative retirement age, funding schemas, tax rates) BEFORE they are implemented.
Use Cases: Academia

• Evaluate the scientific impact of scholars, journals, academic institutions, or nations.
• Quantify and predict scientific research, impact, and outcomes.
• Support the selection of candidate faculty members by universities.
• Identify the best reviewers.
• Prioritize the development of research fields in which a country or region should invest.

Map of Scientific Collaborations from 2005-2009

Computed Using Data from Elsevier's Scopus

VII.6 Stream of Scientific Collaborations Between World Cities - Olivier H. Beauchesne - 2012
Sources and Sinks of Life Time in U.S. Air Travel

A Topic Map of NIH Grants 2007

Bruce W. Herr II (ChalkLabs & IU), Gully Burns (ISI), David Newman (UCI), Edmund Talley (NIH)

The National Institutes of Health (NIH) is organized as a multitude of Institutes and Centers whose missions are primarily focused on distinct diseases. However, disease etiologies and therapies float scientific boundaries, and thus there is tremendous overlap in the kinds of research funded by each Institute. This creates a daunting landscape for decisions on research directions, funding allocations, and policy formulations. Shown here is devised an interactive topic map for navigating this landscape, online at www.nlm.nih.gov. Institute abbreviations can be found at www.nih.gov/ficd.

Topic modeling, a statistical technique that automatically learns semantic categories, was applied to assess projects in terms used by researchers to describe their work, without the biases of keywords or subject headings. Grant similarities were derived from their topic mixtures, and grants were then clustered on a two-dimensional map using a force-directed simulated annealing algorithm. This analytic creates an interactive environment for assessing grant relevance to research categories and to NIH Institutes in which grants are localized.

Cardiac Diseases Research

An area of the map focused on cardiovascular functions and dysfunctions. Cardiac failure (primarily heart Failure) and NIDR (National Institute of Dental and Craniofacial Research) are the two major medical emergencies associated with ischemia, which results from restricted blood supply. Also located in this area are grants focused on Nitric Oxide (NO) Signaling, a major biochemica1 pathway for vasodilation, and grants on Hemodynamics, Sickle Cell Disease, and anemia.

Neural Circuits Research

An area of the map focused on neural circuits, which shows the diversity of topics and NIH Institutes that fund research in this area, such as Cardiopulmonary Regulation, primarily funded by NIDR, Brain Processing, primarily funded by NIDR and NIBIB, primarily funded by NIMH. For color scaling, we assigned in the upper left treat.
Impact

The United States Patent and Trademark Office classifies and catalogs patents by subject matter. Inventors are categorized in a taxonomy that groups related inventions, making it easy to search for patents within a specific category. This visualization displays the US Patent Hierarchy, which is a graphical representation of the relationships between different patents. Each node represents a patent, and the edges connect related patents, showing the hierarchical structure of the patent classification system.

The US Patent Hierarchy

Prior Art

New patents often build on older ideas from many different categories. How does this happen? How are patents classified? The US Patent Hierarchy provides a visual representation of the relationships between patents. Each node represents a patent, and the edges connect related patents, showing the hierarchical structure of the patent classification system. This visualization helps to understand the relationships between patents and the evolution of ideas in science and technology.

MACROSCOPES FOR INTERACTING WITH SCIENCE

Earth
Weather on a worldwide scale

AcademyScope
Exploring the scientific landscape

Mapping Global Society
Local news from a global perspective

Charting Culture
2,600 years of human history in 5 minutes

PLACES & SPACES
MAPPING SCIENCE
scimaps.org
The Megaregions of the US

This is the Roanoke (Raleigh) megaregion.

101st Annual Meeting of the Association of American Geographers, Denver, CO. 
April 5th - 9th, 2005 (First showing of Places & Spaces)

University of Miami, Miami, FL. 
September 4 - December 11, 2014.

Duke University, Durham, NC. 
January 12 - April 10, 2015

http://scimaps.org

The David J. Sencer CDC Museum, Atlanta, GA. 
Data Visualization Literacy Framework

Data Visualization Literacy (DVL)

Data visualization literacy (ability to read, make, and explain data visualizations) requires:

• literacy (ability to read and write text in titles, axis labels, legends, etc.),
• visual literacy (ability to find, interpret, evaluate, use, and create images and visual media), and
• mathematical literacy (ability to formulate, employ, and interpret math in a variety of contexts).

Being able to “read and write” data visualizations is becoming as important as being able to read and write text. Understanding, measuring, and improving data and visualization literacy is important to strategically approach local and global issues.
DVL Framework: Desirable Properties

• Most existing frameworks focus on **READING**. We believe that much expertise is gained from also **CONSTRUCTING** data visualizations.

• Reading and constructing data visualizations needs to take human perception and cognition into account.

• Frameworks should build on and consolidate prior work in cartography, psychology, cognitive science, statistics, scientific visualization, data visualization, learning sciences, etc. in support of a de facto standard.

• Theoretically grounded + practically useful + easy to learn/use.

• Highly modular and extendable.
DVL Framework: Development Process

• The initial DVL-FW was developed via an extensive literature review.

• The resulting DVL-FW typology, process model, exercises, and assessments were then tested in the *Information Visualization* course taught for more than 17 years at Indiana University. More than 8,500 students enrolled in the IVMOOC version ([http://ivmooc.cns.iu.edu](http://ivmooc.cns.iu.edu)) over the last six years.

• The FW was further refined using feedback gained from constructing and interpreting data visualizations for 100+ real-world client projects.

• Data on student engagement, performance, and feedback guided the continuous improvement of the DVL-FW typology, process model, and exercises for defining, teaching, and assessing DVL.

• The DVL-FW used in this course supports the systematic construction and interpretation of data visualizations.
Data Visualization Literacy Framework (DVL-FW)

Consists of two parts:

**DVL Typology**
Defines 7 types with 4-17 members each.

<table>
<thead>
<tr>
<th>Insight Needs</th>
</tr>
</thead>
<tbody>
<tr>
<td>categorize/cluster</td>
</tr>
<tr>
<td>order/rank/sort</td>
</tr>
<tr>
<td>distributions (also outliers, gaps)</td>
</tr>
<tr>
<td>comparisons</td>
</tr>
<tr>
<td>trends (process and time)</td>
</tr>
<tr>
<td>geospatial</td>
</tr>
<tr>
<td>compositions (also of text)</td>
</tr>
<tr>
<td>correlations/relationships</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Scales</th>
</tr>
</thead>
<tbody>
<tr>
<td>nominal</td>
</tr>
<tr>
<td>ordinal</td>
</tr>
<tr>
<td>interval</td>
</tr>
<tr>
<td>ratio</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>statistical</td>
</tr>
<tr>
<td>temporal</td>
</tr>
<tr>
<td>geospatial</td>
</tr>
<tr>
<td>topical</td>
</tr>
<tr>
<td>relational</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Visualizations</th>
</tr>
</thead>
<tbody>
<tr>
<td>table</td>
</tr>
<tr>
<td>chart</td>
</tr>
<tr>
<td>graph</td>
</tr>
<tr>
<td>map</td>
</tr>
<tr>
<td>tree</td>
</tr>
<tr>
<td>network</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Graphic Symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td>geometric symbols</td>
</tr>
<tr>
<td>point</td>
</tr>
<tr>
<td>line</td>
</tr>
<tr>
<td>area</td>
</tr>
<tr>
<td>surface</td>
</tr>
<tr>
<td>volume</td>
</tr>
<tr>
<td>linguistic symbols</td>
</tr>
<tr>
<td>text</td>
</tr>
<tr>
<td>numerals</td>
</tr>
<tr>
<td>punctuation marks</td>
</tr>
<tr>
<td>pictorial symbols</td>
</tr>
<tr>
<td>images</td>
</tr>
<tr>
<td>icons</td>
</tr>
<tr>
<td>statistical glyphs</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Graphic Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>spatial</td>
</tr>
<tr>
<td>position</td>
</tr>
<tr>
<td>retinal</td>
</tr>
<tr>
<td>form</td>
</tr>
<tr>
<td>color</td>
</tr>
<tr>
<td>optics</td>
</tr>
<tr>
<td>motion</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>zoom</td>
</tr>
<tr>
<td>search and locate</td>
</tr>
<tr>
<td>filter</td>
</tr>
<tr>
<td>details-on-demand</td>
</tr>
<tr>
<td>history</td>
</tr>
<tr>
<td>extract</td>
</tr>
<tr>
<td>link and brush</td>
</tr>
<tr>
<td>projection</td>
</tr>
<tr>
<td>distortion</td>
</tr>
</tbody>
</table>

**DVL Workflow Process**
Defines 5 steps required to render data into insights.

![Workflow Diagram]

- **Interpret**
- **Acquire**
- **Analyze**
- **Visualize**
- **Deploy**

**Stakeholders**
Data Visualization Literacy Framework (DVL-FW)

Consists of two parts that are interlinked:

DVL Typology +
DVL Workflow Process
### Typology of the Data Visualization Literacy Framework

<table>
<thead>
<tr>
<th>Insight Needs</th>
<th>Data Scales</th>
<th>Analyses</th>
<th>Visualizations</th>
<th>Graphic Symbols</th>
<th>Graphic Variables</th>
<th>Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• categorize/cluster</td>
<td>• nominal</td>
<td>• statistical</td>
<td>• table</td>
<td>• geometric symbols</td>
<td>• spatial</td>
<td>• zoom</td>
</tr>
<tr>
<td>• order/rank/sort</td>
<td>• ordinal</td>
<td>• temporal</td>
<td>• chart</td>
<td>point</td>
<td>position</td>
<td>• search and locate</td>
</tr>
<tr>
<td>• distributions (also outliers, gaps)</td>
<td>• interval</td>
<td>• geospatial</td>
<td>• graph</td>
<td>line</td>
<td>retinal</td>
<td>• filter</td>
</tr>
<tr>
<td>• comparisons</td>
<td>• ratio</td>
<td>• topical</td>
<td>• map</td>
<td>area</td>
<td>form</td>
<td>• details-on-demand</td>
</tr>
<tr>
<td>• trends (process and time)</td>
<td></td>
<td>• relational</td>
<td>• tree</td>
<td>surface</td>
<td>color</td>
<td>• history</td>
</tr>
<tr>
<td>• geospatial</td>
<td></td>
<td></td>
<td>• network</td>
<td>volume</td>
<td>optics</td>
<td>• extract</td>
</tr>
<tr>
<td>• compositions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>motion</td>
<td>• link and brush</td>
</tr>
<tr>
<td>(also of text)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• projection</td>
</tr>
<tr>
<td>• correlations/relationships</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• distortion</td>
</tr>
</tbody>
</table>

## Typology of the Data Visualization Literacy Framework

<table>
<thead>
<tr>
<th>Insight Needs</th>
<th>Data Scales</th>
<th>Analyses</th>
<th>Visualizations</th>
<th>Graphic Symbols</th>
<th>Graphic Variables</th>
<th>Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>categorize/clustering</td>
<td>nominal</td>
<td>statistical</td>
<td>table</td>
<td>geometric symbols</td>
<td>spatial</td>
<td>zoom</td>
</tr>
<tr>
<td>order/rank/sort</td>
<td>ordinal</td>
<td>temporal</td>
<td>chart</td>
<td>point</td>
<td>position</td>
<td>search and locate</td>
</tr>
<tr>
<td>distributions (also outliers, gaps)</td>
<td>interval</td>
<td>geospatial</td>
<td>graph</td>
<td>line</td>
<td></td>
<td>filter</td>
</tr>
<tr>
<td>comparisons</td>
<td>ratio</td>
<td>topical</td>
<td>map</td>
<td>area</td>
<td></td>
<td>details-on-demand</td>
</tr>
<tr>
<td>trends (process and time)</td>
<td></td>
<td>relational</td>
<td>tree</td>
<td>surface</td>
<td></td>
<td>history</td>
</tr>
<tr>
<td>geospatial</td>
<td></td>
<td></td>
<td>network</td>
<td>volume</td>
<td></td>
<td>extract</td>
</tr>
<tr>
<td>compositions (also of text)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>link and brush</td>
</tr>
<tr>
<td>correlations/relationships</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>projection</td>
</tr>
</tbody>
</table>

<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>category</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>categorize/cluster</td>
</tr>
<tr>
<td>order</td>
<td>rank</td>
<td>ranking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>order/rank/sort</td>
</tr>
<tr>
<td>distribution</td>
<td>distribution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>distributions (also outliers, gaps)</td>
</tr>
<tr>
<td>compare</td>
<td>nominal comparison &amp; deviation</td>
<td>differences</td>
<td>compare and contrast</td>
<td>compare data values</td>
<td>comparison</td>
<td>comparisons</td>
<td></td>
<td></td>
</tr>
<tr>
<td>time series</td>
<td>patterns over time</td>
<td>time</td>
<td>process and time</td>
<td>track rises and falls over time</td>
<td>trend</td>
<td>trends (process and time)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>geospatial</td>
<td>spatial relations</td>
<td>location</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>geospatial</td>
</tr>
<tr>
<td>quantity</td>
<td>part-to-whole</td>
<td>proportions</td>
<td>form and structure</td>
<td>see parts of whole, analyze text</td>
<td>composition</td>
<td>compositions (also of text)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>association</td>
<td>correlate</td>
<td>correlation</td>
<td>relationships hierarchy</td>
<td>relations between data points</td>
<td>relationship</td>
<td>correlations/relationships</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Typology of the Data Visualization Literacy Framework

<table>
<thead>
<tr>
<th>Insight Needs</th>
<th>Data Scales</th>
<th>Analyses</th>
<th>Visualizations</th>
<th>Graphic Symbols</th>
<th>Graphic Variables</th>
<th>Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• categorize/clustering</td>
<td>• nominal</td>
<td>• statistical</td>
<td>• table</td>
<td>• geometric symbols</td>
<td>• spatial</td>
<td>• zoom</td>
</tr>
<tr>
<td>• order/rank/sort</td>
<td>• ordinal</td>
<td>• temporal</td>
<td>• chart</td>
<td>• point</td>
<td>• position</td>
<td>• search and locate</td>
</tr>
<tr>
<td>• distributions (also outliers, gaps)</td>
<td>• interval</td>
<td>• geospatial</td>
<td>• graph</td>
<td>• line</td>
<td>• retinal</td>
<td>• filter</td>
</tr>
<tr>
<td>• comparisons</td>
<td>• ratio</td>
<td>• topical</td>
<td>• map</td>
<td>• area</td>
<td>• form</td>
<td>• details-on-demand</td>
</tr>
<tr>
<td>• trends (process and time)</td>
<td></td>
<td>• relational</td>
<td>• tree</td>
<td>• surface</td>
<td>• color</td>
<td>• history</td>
</tr>
<tr>
<td>• geospatial</td>
<td></td>
<td></td>
<td>• network</td>
<td>• volume</td>
<td>• optics</td>
<td>• extract</td>
</tr>
<tr>
<td>• compositions (also of text)</td>
<td></td>
<td></td>
<td></td>
<td>• linguistic symbols</td>
<td>• motion</td>
<td>• link and brush</td>
</tr>
<tr>
<td>• correlations/relationships</td>
<td></td>
<td></td>
<td></td>
<td>• text</td>
<td></td>
<td>• projection</td>
</tr>
</tbody>
</table>

### Graphic Symbol Types

<table>
<thead>
<tr>
<th>Spatial Position</th>
<th>Graphic Symbol Types</th>
<th>Linguistic Symbols</th>
<th>Pictorial Symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Geometric Symbols</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Point</strong></td>
<td><img src="image" alt="Graphic Symbols" /></td>
<td><img src="image" alt="Text" /></td>
<td><img src="image" alt="Pictorial Symbols" /></td>
</tr>
<tr>
<td><strong>Line</strong></td>
<td><img src="image" alt="Graphic Symbols" /></td>
<td><img src="image" alt="Text" /></td>
<td><img src="image" alt="Pictorial Symbols" /></td>
</tr>
</tbody>
</table>

#### Qualitative

- **Form**
  - Also called: Categorical Attributes
  - Identity Channels

- **Shape**
  - ![Shape Symbols](image)

- **Value**
  - ![Value Symbols](image)

- **Color**
  - ![Color Symbols](image)

- **Saturation**
  - ![Saturation Symbols](image)

- **Granularity**
  - ![Granularity Symbols](image)

- **Pattern**
  - ![Pattern Symbols](image)

- **Blur**
  - ![Blur Symbols](image)

- **Speed**
  - ![Speed Symbols](image)

#### Quantitative

- **Also called:** Ordered Attributes
- **Magnitude Channels**

---

Data Visualization Literacy Framework (DVL-FW)

Implemented in Make-A-Vis (MAV) to support learning via horizontal transfer, scaffolding, hands-on learning, etc.
Visual Analytics Certificate

https://boeing.cns.iu.edu
Visual Analytics Certificate

Duration: 6 weeks x 5 hours = 30 hours (3 CEUs)
Format: Online using Canvas
Type: Executive Education

Learn to render data into actionable insights in 6 weeks!

For US-based Boeing employees:
Begin the enrollment process by completing the pre-enrollment form (https://iu.co1.qualtrics.com/jfe/form/SV_ekXpn97sf6gvZAx). NOTE: You will need your voucher to complete the form. Over the days following submission, look for emails about next steps and course information to finalize your enrollment.

For non-US based Boeing employees and others not using the voucher system:
Register for the course online through Indiana University's IU Expand portal. Go to: https://expand.iu.edu/browse/sice/cns/courses/visual-analytics-boeing-2. Payment is via credit card.
Instructors

Katy Börner
Instructor
Victor H. Yngve Distinguished Professor of Engineering and Information Science at the School of Informatics, Computing, and Engineering, Founding Director of the Cyberinfrastructure for Network Science Center (http://cns.iu.edu) at Indiana University.

Michael Ginda
Assistant Instructor
Data analyst and research assistant with the Cyberinfrastructure Center for Network Science. He holds a Master’s degree in Library Science from Indiana University.

Andreas Bueckle
Assistant Instructor
PhD student in Information Science at Indiana University focused on information visualization.

- Research focus on development of data analysis and visualization techniques for information access, understanding, and management.
- Cyberinfrastructures development for large-scale scientific collaboration and computation.
- Research focus on knowledge representation and organization, metadata, and information networks.
- Lead instructional designer.
- Research focus on information visualization, specifically interactive and augmented reality.
- Videography and photography.
Schedule

• Week 1 – Introduction to Visual Analytics

• Week 2 – When: Temporal Data Analysis and Visualization
• Week 3 – Where: Geospatial Data Analysis and Visualization
• Week 4 – What: Topical Data Analysis and Visualization
• Week 5 – With Whom: Network Analysis and Visualization

• Week 6 – Value Creation via Data-Driven Decision Making and Future Developments
## Schedule

- Highly structured course design.
- Each week features the same modules that build on each other to support scaffolding.
- Materials are designed to support transfer—building on previous knowledge and expertise to promote deeper levels of learning.
- Students immediately apply theory to gain hands-on knowledge and expertise.
- Social learning is supported.

### Course Overview

<table>
<thead>
<tr>
<th>Schedule</th>
<th>Theory</th>
<th>Hands-on</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Week 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Week 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Week 3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Week 4</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Week 5</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Week 6</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Post</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Transfer

- **Pre**
- **Week 1**
- **Week 2**
- **Week 3**
- **Week 4**
- **Week 5**
- **Week 6**

### Scaffolding

- **Pre**
- **Week 1**
- **Week 2**
- **Week 3**
- **Week 4**
- **Week 5**
- **Week 6**

### Theory

- **Pre**
- **Week 1**
- **Week 2**
- **Week 3**
- **Week 4**
- **Week 5**
- **Week 6**

### Hands-on

- **Pre**
- **Week 1**
- **Week 2**
- **Week 3**
- **Week 4**
- **Week 5**
- **Week 6**
Schedule: Your Project

• Week 1 – Introduction to Visual Analytics

• Week 2 – When: Temporal + Identify your very own “When” questions
• Week 3 – Where: Geospatial + Identify your “Where” questions
• Week 4 – What: Topical + Identify your “What” questions
• Week 5 – With Whom: Network + Identify your “With Whom” questions

• Week 6 – Value Creation via Data-Driven Decision Making and Future Developments + Submit Visualization (Specification)
Grading and Completion Criteria

Final grade is based on:

- Data Visualization Literacy Framework Quizzes (30%)
- Make-A-Vis Tasks (30%)
- “Your Project” Work (20%)
- Participation (20%)

Students must achieve a score of 70% or higher to pass the course.
Case Studies

• Guiding Professional Training Choices
• Optimizing Career Trajectories within Boeing
• Visualizing Business Process Outsourcing/Supply Chain Management

Based on MyProject submissions in Beta Run of the Visual Analytics Certificate we are also exploring case studies on
• Airspace and Operational Efficiency Visualizations
• Safety and Health Visualizations
• Visualizing Software Development Progress Using GitHub Activity Data
Case Study 1: Career Trajectories within Boeing

Many employees at Boeing have been with the company for several years if not decades. During their tenure at Boeing, many held diverse jobs, with different associated skill sets. Using anonymized Boeing HR data, we will plot the trajectories within Boeing over the last 20 years.

- Different types of “reference systems” might be valuable, including
- **WHEN-Time**: A stream graph, color coded by job types, and thickness coded by the number of employees
- **WHERE-Geo**: A geospatial map of the Boeing campus
- **WHAT-Topic**: A topic map of skills, see [https://www.pnas.org/content/115/50/12630](https://www.pnas.org/content/115/50/12630)
- **NETWORK**: A network layout of all major career trajectories.

“Data overlays” of interest would help reveal typical salary, mean age, types of skills, typical number of years active, etc. associated with certain job types. They would make visible entry jobs that effectively start a successful career at Boeing. They would show major “career highways” that a large number of employees have travelled. Visualizations would identify jobs (and associated skills) that slow down or accelerate career growth; or lead to employees resigning.

Different combinations of reference systems and data overlays will be explored with Boeing experts, see below list of most relevant experts. The 2-3 best will be prototyped and one will be selected for use in the course. Associated exercises and assessments will be tested in Run1-Boeing and refined and optimized for Run2-Boeing.

**Experts most relevant**: HR experts and those involved in hiring and promotion decision making + optimizing the Boeing workforce for maximum competitiveness.
Case Study 2: Guiding Professional Training within Boeing

Employees at Boeing have access to 50,000 courses and degrees that take between 30mins to 3 years (for a degree) to complete and are provided by IEEE, Coursera, and others. Most employees are overwhelmed by this rich set of options. Few explore the space of training opportunities and a rather small percentage completes any of the courses.

Boeing does own access log files for all courses that Boeing employees and employees of Boeing suppliers can take. Using these log files and metadata for all courses, the landscape of training opportunities can be plotted and data on course quality, costs, time invested can be overlaid.

Different types of “reference systems” might be valuable, including

• WHEN-Time: A stream graph, color coded by course types, and thickness coded by the number of employees taking the course OR the skills trained.
• WHERE-Geo: A geospatial (world) map of work locations of Boeing/supplier employees that took the course.
• WHAT-Topic: A topic map of skills, see https://www.pnas.org/content/115/50/12630
• NETWORK: A network layout of all major course taking trajectories.

“Data overlays” of interest would help reveal course quality, costs, time commitment, #employees that took the course. They would make visible courses that have high/low completion rates. They would show major “learning highways” that a large number of employees have travelled. If data on employee job positions/salaries can be linked, visualizations would help identify courses (and associated skills) that slow down or accelerate career growth; or lead to employees resigning.

Different combinations of reference systems and data overlays will be explored with Boeing experts, see below list of most relevant experts. The 2-3 best will be prototyped and one will be selected for use in the course. Associated exercises and assessments will be tested in Run1-Boeing and refined and optimized for Run2-Boeing.

Experts most relevant: HR experts and others to guide employees in selecting best training options. Employees interested to invest their family time wisely to advance their expertise and career options. Plus, anyone interested to optimize employee attention usage and in developing “signals” that employees can use to maximize ROI for workforce training.
Case Study 3: Optimizing Business Process Outsourcing

This case study builds on insights gained from *Case Study: Supply Chain Dreams and Nightmares* by Gerard Chick. Kogan Page. (c) 2016.

Advanced business process outsourcing (BPO) is needed to procure a complex product system such as a commercial airliner that involves xx suppliers in xx countries producing xx parts that all need to fit together in highest quality—and within time and budget constraints. For the 787 Dreamliner, the BPO model can be easily as complex as the model of the aircraft itself. A detailed understanding and optimization of the BPO is required to meet quality goals as well as budget and time constraints.

Given the maxim that the performance of the prime manufacturer (here Boeing) can never exceed the capabilities of the least proficient supplier, and the fact that delays in the earlier parts of the supply chain percolate through the entire supply chain, it seems desirable to visualize the network of all suppliers and intermediate products together with associated costs, current delays, any quality issues that arose, etc. so all involved gain a more complete understanding of this multi-level socio-technical systems problem and its current solution.

Boeing might not be able to release the complete supply BPO model for the 787 Dreamliner. However, an example such as all parts needed for an important subcomponent of an aircraft that is not produced anymore or a toy example such as all ingredients needed for a amazing recipe (that employees might like to cook with their families) could serve to illustrate how to analyze and visualize BPO data. Together with information on how to read in Boeing data, employee-students could re-run the visualizations with Boeing internal data.

- Different types of “reference systems” might be valuable, including
  - WHEN-Time: Gantt chart
  - WHERE-Geo: A geospatial (world) map of Boeing suppliers
  - NETWORK: A tree visualizations showing part-of relationships of all airplane parts.

“Data overlays” of interest would help reveal costs, delays, quality for each supplier/subcomponent. They would make visible suppliers that cause delays, increase costs, and are the reason for quality problems. In addition, costs for transporting components and managing teams that work on outsourced components should be shown.

Different combinations of reference systems and data overlays will be explored with Boeing experts, see below list of most relevant experts. The 2-3 best will be prototyped and one will be selected for use in the course. Associated exercises and assessments will be tested in Run1-Boeing and refined and optimized for Run2-Boeing.

**Experts most relevant:** System engineering experts that develop, optimize, and ensure compliance to BPO.
Resources


Visual Analytics Certificate

Learn to render data into actionable insights in 6 weeks!
Class begins **March 2, 2020**. Get course info at [https://boeing.cns.iu.edu](https://boeing.cns.iu.edu)

**For US-based Boeing employees:**
Begin the enrollment process by completing the pre-enrollment form ([https://iu.co1.qualtrics.com/jfe/form/SV_ekXpn97sf6gvZAx](https://iu.co1.qualtrics.com/jfe/form/SV_ekXpn97sf6gvZAx)). NOTE: You will need your voucher to complete the form. Over the days following submission, look for emails about next steps and course information to finalize your enrollment.

**For non-US based Boeing employees and others not using the voucher system:**
Register for the course online through Indiana University's IU Expand portal. Go to: [https://expand.iu.edu/browse/sice/cns/courses/visual-analytics-boeing-2](https://expand.iu.edu/browse/sice/cns/courses/visual-analytics-boeing-2). Payment is via credit card.
Learning Analytics

Background and Motivation

In Fall 2017, our team began working with The Boeing Company to leverage our expertise in visual analytics to study data produced by students taking MITxPro online courses to understand

- the structure of courses resources,
- student engagement and learner trajectories, and
- student performance


1,611 Boeing engineers registered; 1,565 were active and generated nearly 31 million click event records while accessing videos, projects, and assessments. Some students generated over 100,000 separate events.

All but 255 engineers passed the course, resulting in a completion rate of 84.1%.
Course Structure

Course Structure Tree Diagram shows 5-level hierarchical structure of the Architecture of Complex Systems course. Nodes are ordered based on the sequence of learning modules presented to learners in the course.

Insights: Course structure allows for analysis and visualizations at multiple levels of granularity, temporality. Modules presented to students share similar lengths.
Student Engagement Predictions

Instructors Temporal Predictions are represented in a temporal bar graph that compares course instructors *estimated* time learners would need to complete course materials, and the average time taken by learners in the course computed from data.

**Insights:** Instructor’s temporal estimates are accurate but did not account for studying activity of students in their estimates.
Learning Trajectories

#Event Interactions | Node Color - Module Type | Edge Color - Transitions
--- | --- | ---
21 | Course Mechanics and Information Pages* | Forward Direction
1 | Video Modules | Backward Direction
| HTML/Text Modules | Open Assessment Modules
| Ungraded Problem Modules |
Student Engagement and Performance

Students Interacting with a Course Module.

Scatter graph showing the percentage of the learners in the Architecture of Complex System course accessing modules by certificate group and module type.

Insights: Clear difference in access patterns by students across the course by certificate and non certificate earners, as well as subtle differences between module types. Most notably, few of the students that do not earn a certificate do access the Open Assessment Modules.
Thank You

Q&A

https://boeing.cns.iu.edu