Visualizing (:) A New Data Support Role for Duke University Libraries

Angela Zoss
Data Visualization Coordinator
Duke University Libraries

December 10, 2013
CNI Fall 2013 Membership Meeting
What is Data Visualization?

http://guides.library.duke.edu/datavis/
What is Data Visualization?

State of the Union Address, 2002 vs. 2011

President Bush, January 29, 2002

http://guides.library.duke.edu/datavis/
What is Data Visualization?

http://guides.library.duke.edu/datavis/
What is Data Visualization?

http://guides.library.duke.edu/datavis/
Why Visualize?

• Explore data, uncover hidden patterns

Anscombe’s Quartet

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of x</td>
<td>9 (exact)</td>
</tr>
<tr>
<td>Variance of x</td>
<td>11 (exact)</td>
</tr>
<tr>
<td>Mean of y</td>
<td>7.50 (to 2 decimal places)</td>
</tr>
<tr>
<td>Variance of y</td>
<td>4.122 or 4.127 (to 3 decimal places)</td>
</tr>
<tr>
<td>Correlation between x and y</td>
<td>0.816 (to 3 decimal places)</td>
</tr>
<tr>
<td>Linear regression line</td>
<td>$y = 3.00 + 0.500x$ (to 2 and 3 decimal places, respectively)</td>
</tr>
</tbody>
</table>

http://en.wikipedia.org/wiki/Anscombe%27s_quartet
Why Visualize?

- Translate something typically invisible into the visible

Economic Indicators

http://blogs.library.duke.edu/data/2012/11/12/adding-colored-regions-to-excel-charts/
Why Visualize?

• Communicate results, contextualize data, tell a story, or possibly even mobilize action around a problem

Hans Rosling: The River of Myths

http://www.youtube.com/watch?v=Owll-dwh-bk
Visualization at Duke

No single centralized community, but plenty of distributed groups and projects:

- Visualization Technology Group, c. 2001
- Visualization Friday Forum, c. 2002
- Duke immersive Virtual Environment (DiVE), c. 2006
- Data & GIS Services, c. 2007
- Visual Studies Initiative, c. 2007
- The Wired! Group, c. 2009
- LINK MediaWall, c. 2009
- Media Arts + Sciences Mellon Grant, c. 2013
- M.A. in Historical and Cultural Visualization, c. 2014
Visualization at Duke

Types of visualization work:

http://virtualreality.duke.edu/
Visualization at Duke

Types of visualization work:

http://visualizingvenice.org/
Visualization at Duke

Types of visualization work:

http://research.duke.edu/stories/chemists-lasers-could-id-ancient-artists-intent
Visualization at Duke

Types of visualization work:

Visualization at Duke

Types of visualization work:

- How Intellectual: 69% expected, 42% moderate

- How Intellectual Duke is: 55% very, 45% moderate

- How do students feel about their intellectual experiences?
  - Intellectual climate in the classroom: 78%
  - Social climate of the school: 58%
  - Faculty availability: 56%
  - Faculty help on the weekend: 35%
  - Faculty availability limited to weekends: 7%

- 50% think faculty encourage intellectual climate “moderately” or “very much”

- Matters a lot to your purpose for being at Duke:
  - 84% working on or planning relationships
  - 81% preparing for a career

- How many times per semester do you read a book that’s not assigned for class?
  - Never: 35%
  - 1-2: 43%
  - 3-5: 15%

- Finding your social and intellectual niche takes time...
  - Days: 8%
  - Weeks: 16%
  - Months: 36%
  - Years: 39%

- Satisfied with intellectual experiences outside of class:
  - Never: 16%
  - Slightly: 35%
  - Moderately: 30%
  - Very: 19%

Visualization at Duke

Types of visualization work:

Who could support visualization?

- Faculty/department?
- College/school?
- Campus-wide organization?
Similar support at Duke

• Data and GIS Services (DUL)
• Research and Instructional Services (DUL)
• Center for Instructional Technology (DUL)
• Multimedia Project Studio (OIT)
• Link Teaching and Learning Center (OIT)
• Social Science Research Institute (campus wide)
Data Visualization Coordinator Position

- Started June 2012
- Dual report
  - Data & GIS Services, Duke University Libraries
  - Research Computing, Office of Information Technology
- Objectives:
  - instruction and outreach
  - consultation
  - develop new visualization services, spaces, programs
After 18 Months:
What has been most successful?

• Visualization workshop series
• Online instructional material
• Just-in-time consulting
• Ongoing visualization seminar series
• Student data visualization contest
Workshops

• Visualization software (Tableau, d3)
• Data processing (text analysis, network analysis)
• Best practices (designing academic figures/posters, top 10 dos and don’ts for charts and graphs)

http://guides.library.duke.edu/profile.php?uid=62927
New technique: GitHub/Gist

• Using GitHub to share sample code
• Using Gist and bl.ocks.org to see the visualization right away

http://bl.ocks.org/dukevis/6768900
A Few Basic Tips

From workshop,
“Top 10 Dos and Don’ts for Charts and Graphs”

http://guides.library.duke.edu/topten
Simplify less important information.

Chart elements like gridlines, axis labels, colors, etc. can all be simplified to highlight what is most important/relevant/interesting.


http://vis4.net/blog/posts/doing-the-line-charts-right/
Don’t use 3D effects.

Source: http://blog.visual.ly/2ds-company-3ds-a-crowd/

http://www.slideshare.net/jschwabish/making-excel-graphs-better/18
Color

Don’t use rainbows for ordered, numerical variables.

“Rainbow Color Map (Still) Considered Harmful”
D Borland, RM Taylor, UNC-CH
http://www.renci.org/~borland/pdfs/RainbowColorMap_VisViewpoints.pdf
Solution: Single hue, varying luminance
Just-in-time Consulting

• Weekly walk-in consulting hours in the Data & GIS Services computer lab
• Additional appointments outside of walk-in hours
• Detailed support and troubleshooting via email
Weekly Visualization Seminars

• Visualization Friday Forum, in its 12\textsuperscript{th} year
• Informal discussion environment
• Lunch provided
  (through funding from OIT Research Computing and administrative support from Computer Science department)
• Speakers from across and outside of Duke
• Live streaming and archived video

\url{http://vis.duke.edu/FridayForum}
Student Data Visualization Contest

• Goal: to advertise new services, take a survey of visualization at Duke
• Open to Duke students, any type of visualization
• Judged on insightfulness, narrative, aesthetics, technical merit, novelty
• Awards: Amazon gift cards, public exhibits, posters

Wide range of submissions

In February 2012, Panda Properties Sino, LLC, announced plans to construct a new “NC Chinese” that would serve as a venue for the Chinese and broader Asian-American community in the Triangle. The idea was to create a space that uses a combination of community center, cultural hub, retail outlet mall, and restaurant row. This map demonstrates how NC Chinese would have fit into an existing spatial network comprised of Chinese restaurants, language schools, and political/community organizations. The NC Chinese project sailed in the summer of 2012.

Mapping Chinatown
Sabrina McCutchan

The separation plot navigation tool, spNavigate, provides researchers with a graphical interface for exploring the results of logistic and probit regression models. Every observation in a data set is ordered by predicted probability given a binomial regression model and then colored according to the dependent variable and unidirectional parameters. The plot on the left represents observations; those with low predicted probability are on bottom and those with high predicted probability are on top. Blue arrows represent movements of the dependent variable; colored lines represent observations of event occurrence. The panel on the right provides users with a scrollable version of the plot that has been magnified so that identifying characteristics of each observation are visible.

spNavigate
Benjamin Radford

Duke Intellectual Climate Report 2012
Amanda Peralta

Grand Prize Winners

Limbique
Pinar Yoldas and David Paulsen

ACC Basketball Tournament Series Records
Volodymyr Zavidovych

After 18 Months:
What has been most challenging?
After 18 Months: 
What has been most challenging?

• Marketing and outreach
After 18 Months:
What has been most challenging?

• Marketing and outreach
• Staying current
After 18 Months:
What has been most challenging?

• Marketing and outreach
• Staying current
• Project work, priorities
After 18 Months: What has been most challenging?

- Marketing and outreach
- Staying current
- Project work, priorities
- Disciplinary silos and conventions
After 18 Months:
What has been most challenging?

- Marketing and outreach
- Staying current
- Project work, priorities
- Disciplinary silos and conventions
- Curriculum and skill gaps
Common skill gaps

- Visualization types and tools
- Spreadsheet and/or database familiarity
- Scripting
- Robust data management practices
- Basic graphic design

Resource: https://github.com/veltman/learninglunches
Hopes for the Future
Hopes for the Future

• Active student training program
  (courses, independent studies, employment)
Hopes for the Future

- Active student training program (courses, independent studies, employment)
- Additional physical and digital exhibit opportunities
Hopes for the Future

• Active student training program (courses, independent studies, employment)
• Additional physical and digital exhibit opportunities
• Continued project and workshop development
Advice: What should a coordinator know?

• Data transformations
• Range of visualization types, tools
• Range of teaching strategies
• Marketing
Advice: What should a coordinator do?

• Find access points to different user communities
• Use events to build a community, make the service more visible
• Collaborate on research projects
• Stockpile interesting datasets
• Beware of unmanaged screens
• Block out plenty of quiet time for the above
Advice: How should an organization establish a new visualization support program?

- Identify potential early adopters
- Budget for a few events, marketing materials, etc.
- Involve other service points to pave the way for referrals
- Provide a research or practice support system for the coordinator
- Expect high demand!
QUESTIONS?

angela.zoss@duke.edu
twitter.com/duke_vis