INFORMATION VISUALIZATION

Guest Lecturer: Ashleigh Faith
SAE International Lead Taxonomist
WHAT IS INFORMATION VISUALIZATION (INFORVIS)?
Early Examples
“Information visualization is the use of computer-supported, interactive, visual representations of abstract data to amplify cognition” –i.e something without physical form. -Stephen Few

And/Or

“Visualization is an activity which a human being engages in, and that it is a cognitive activity: in other words, it goes on in the mind...[and results in internal insight and understanding]. –Robert Spence

And/Or

“Visualization is ...a process of transforming information into a visual form enabling the viewer to observe, browse, make sense, and understand the information. It typically employs computers to process the information and computer screens to view it using methods of interactive graphics, imaging, and visual design.1 Growing access to information from education, archives, records, social media, government, astronomy, geology, medicine, and news offers an increasingly widening pool of data that can be combined to create impressive visuals ranging from cartography to cartoons.2 [However,] visualization is more than pretty pictures. This is not to lessen the importance of visual aesthetics...However, the beauty of an effective visualization is more than skin deep.”3

1 Infovis.org
3 Infovis.org
What is records intelligence?

Records intelligence refers to whatever supports access, delivery, presentation, visualization, and exploration of information for the purpose of leveraging data analytics.

-James Kobielus, Forrester analyst
London Cholera Map by John Snow 1854
Gapminer by Hans Rosling 2007

Between 10 and 40% of the children died.

We now have a completely different world!
Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Paris, le 20 Novembre 1869.

Les nombres d'hommes pratiques sont représentés par les longues des zénes coloris à raison d'un millième pour dix mille hommes; ils sont de plus écrits en travers des zénes. Le rouge désigne les hommes qui tombèrent en Russie; le noir ceux qui sont vivants. Les renoncements que l'on devra à dresser la carte ont été pris avec les encouragements de MM. Chrest. de Pluig, de Forissier, de Chambray et le journal intime de Jollif, pharmacien de l'Armée depuis le 23 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai juxtaposé les corps de Blau, Thieme et du Marshal Davoust qui arrivèrent réunis un à Moscou, et en retour vers Ofscha et Wiedare, avec toujours marchés avec l'armée.
The areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent areas for all the deaths from Preventible or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds; & the black wedges measured from the centre the deaths from all other causes.

The black line across the red triangle in Nov 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855, the black area coincides with the red; in January & February 1855, the blue coincides with the black.

The entire areas may be compared by following the blue, the red & the black lines enclosing them.
A NEW CHART OF HISTORY.
If students aren't taught the language of sound and images, shouldn't they be considered as illiterate as if they left college without being able to read or write? - George Lucas

The top 3 points to learn about infovis:

1. Multimodal fluency: Teach a basic visual design vocabulary
2. Design context

BREAK

3. Visual judgment: Develop constructive critics of visual information

---

Typical Infovis components

- Software that will compute the information
- Metadata tags on content (taxonomy/ontology)
- A dataset (is it clean or automated?)
- Scope of information to extract. It is helpful to write this out and get approval on it beforehand (in the workplace).
  - What questions are you trying to answer?
  - Have you asked your stakeholders what their records intelligence needs are?
  - What are your restrictions?
  - What will the output be?
- Understand the time and resource commitment for the project.
- Identify where this fits into the workflow (refer back to previous workflow and retention schedule lecture)
Design Context (Spence):

- **Selection:** what data are you trying to visualize? What does your audience need to know about the dataset? What should and should not be included? Should the data be visualized? How clean is the data? Can it be automatically gathered?

- **Representation:** what can be represented effectively? Should the visual be interactive? Can the visualization evolve with the data and does it need to?

- **Scale and dimensions:** what will represent the data best? Line, Bar, pie, bubble, geospatial, temporal...etc.?

- **Externalization:** how will the user understand the visual? What is the audience? What will the visual be used for?

- **Mental models:** colors, patterns, X and Y axis, type of visualization

- **Invention, experience, and skill:** find a tool you can work with that fits what you want to do and what your skill level is
Few and Spencer’s 4 Types of Data

- **Nominal**
  - 4.1
  - 27
  - 102
  - 3.14
  - -0.1
  - 16

- **Ordinal data**
  - Monday
  - Wednesday
  - Tuesday
  - Thursday

- **Categorical data**

- **Interval**
TAKE A BREAK
5 models of visualization

- Statistical (representative numbers)-
- Temporal (When) —
- Geospatial (Where)-
- Topical (What)-
- Network (With whom?)-
Name that Model
Black text indicates a year with detect values.

Blue text indicates a year with only nondetect values.

Analyte concentrations averaged over the specified year starting in the center at 0 and ending at the outer ring with a maximum of 200.

Grey radials indicate years prior to the start of sampling.

Grey text indicates a year when sampling data was not present.

Sample Location Name
Impact of Air Travel on Global Spread of Infectious Diseases - Vittoria Colizza, Alessandro Vespignani, and Elisha F. Hardy - 2007
M/RCPL
Mansfield/Richland County Public Library
43 West Third Street • Mansfield, OH • 44902

Our Library is busy!

74,000 registered users
over two thousand five hundred visitors daily

- 5,600 library programs & events
- 117,000 participants
- 13,000 e-books downloaded

However:

Lost 35%
$1.7 million per year

State funding continues to decline. State funding of public libraries has declined 35% since 2001. Since 2010, our Library has lost almost $1.7 million per year in state funding and it is not likely to return.

Without the levy,
our Library will have to cut $6.7 million more per year, meaning the Board of Trustees will have to make tough decisions:
- Cutting hours of operation, including closing some days/s at every location
- Eliminating updates to technology
- Eliminating adult programming and severely reducing children’s and teen programming
- Further layoffs and reductions in staff
- Slashing the budget for books, movies, music, deep web resources

With the levy,
our Library keeps services at current levels:
- Keeps enough materials like books, DVDs and references for our customer needs
- Maintains technology
- Keeps our children’s programming intact
- Continues services like Books-To-Go for older adults
- Maintains the critical 2:1:1 line for those in need
- Keeps hours of operation that serve all of our customers
- Protects opportunities to ensure our students can compete in the global economy

Our Library needs this issue to maintain the quality of our community and our property value. The current levy expires next year. The 2 mill renewal + 1.9 additional levy will cost $1.54 per month more per $100,000 in home valuation to keep existing Library services.
5 models of visualization

- Everyone wins!
- Visualization models can be combined for more detailed information extraction