Data Visualization II

Xinyi Li xinyili@princeton.edu
Nick Budak nbudak@princeton.edu
Center for Digital Humanities

Workshop description
Visual display of data and information can be incredibly powerful. Last semester the CDH showed you how to start making your own data visualizations; join Xinyi Li and Nick Budak for this follow-up workshop and learn how to appreciate and critique visualizations critically. This workshop will present a critique framework of key aspects to consider when assessing and constructing visualizations. We will turn a critical eye to various examples, and the framework will be a tool to help you unpack the strengths, weaknesses, and hidden arguments in visual presentations.

Why visualize?
- Visuals help us think
- Provide a frame of reference, a temporary storage area

What is Data
1. DIKW pyramid
    a. data **know-nothing**
    Recordings or perception of evidence. Product of observation. It has no meaning by itself.
    b. information **know-what**
    Data that have been given meaning by way of relational connection. Do not have to be useful (for the user)
    c. knowledge **know-how**
    Internalization of useful information. when it's under 30 degree, …
    d. wisdom **know-why**
    Consists of making the best use of knowledge
    a wise decision may be made with incomplete information

Data Visualization Definition
(i) Interpreting information in visual terms by forming a mental picture based on data.
(ii) Applying suitable methods to put data into visible form.

– Hinterberger H. (2009) Data Visualization. In: LIU L., ÖZSU M.T. (eds) Encyclopedia of Database Systems. Springer, Boston, MA
http://www.infovis-wiki.net/index.php?title=Information_Visualization

Some contentions

Some include the **computer** and **cognition** in their definition of visualization

"The use of computer supported, interactive, visual representations of data to amplify cognition"

Arnheim R. Visual Thinking. University of California Press, Berkeley, CA, 1969.

how data visualization differs from information visualization:

"data visualization is for exploration, for uncovering information, as well as for presenting information. It is certainly a goal of data visualization to present any information in the data, but another goal is to display the raw data themselves, revealing the inherent variability and uncertainty"

Unwin A., Theus M., and Hofmann H. Graphics of Large Datasets: Visualizing a Million. Springer Series in Statistics and Computing, Berlin, 2006.

Information Visualization

Non-spatial data: Items, entities, things which do not have a direct physical correspondence

Taking things without a direct physical correspondence (non-spatial) and mapping them to a 2-D or 3-D physical space

Giving information a visual representation that is useful for analysis and presentation

Broader understanding

- projects that represent data or information in graphical ways
- mapping from structured and unstructured data to a graphical form
- mapping from data variables to visual elements
- Intention of doing interpretations and visual analysis

- "visualization designer are like ancient cartographers"
- produce representations of the world, as we know it
- use our representation to explore the world and to understand it

Critique Culture

- negative connotations; merit recognition; practice of doubt
- method of systematic analysis of a written, oral and visual discourse

Data Viz Critique

- Intersection of traditions:
    - literary and artistic critics, "*a given domain's informed exercise of judgment*"; public
    - Art and design studio; assessment tool involving knowledge transfer; reflective; the goal of improving a work in progress; private
    - Science, computer graphics, HCI; remove human judgment, direct comparison with previous work; "**redesign**" to critique

- Data viz paper started to include “evaluation”; new and not as easy as usability testing in interactive design; “usable” ≠ good; art critic
- Too nice and polite / constructive and thoughtful feedback
- Field is young, less hierarchies, doctrine, and unquestionable authorities

Critique Framework

- different professionals different criteria for success for a visualization
- some key aspects of visualization

1. Intention

- Visualization as **Medium** or **tool**
  Representative*, *Generative
  Explanatory, Exploratory
  - **Medium**
    the designer observes reality and produces a visual representation with the objective of communicating a message to an audience
  - Representations of knowledge already known
    - **Static representation**
    - Communication; Telling a story, making a point
  - **Tool**
    in this process the designer observes and designs a visual representation in order for the audience (and himself as well) to perform a specific task
    - Example: constellation
  - Knowledge generators create new information through their use
    - **Dynamic generators**, open-ended relation
    - Generating insights; Helping to raise questions and supply answers

- Spectrum, Transparent analysis or Arguments

  Example: Boris Müller *for Poetry on the Road* literature festival
  etchings

2. Data/information coding
Translation from statistic parameters into graphical elements

Scaling
Bar, line

- Scale of one axis in relation to another
- Use of broken or continuous metrics
  area
- Linear vs. Quadratic Change
  - Example: Florence Nightingale, “Diagram of the Causes of Mortality in the Army in the

East – Notes on matters, affecting the health, efficiency and hospital administration of the British Army" (1858)
- demonstrating that more soldiers died of preventable epidemic diseases (blue) than battlefield wounds (red) or other causes (black) during the Crimean War (1853-56).
- area is what the eye sees

Relationship
The Part-Whole Relationship

- Compare the parts to the whole. If the parts do not sum up to a meaningful whole, they cannot be represented in a pie chart.
- the parts must be mutually exclusive
- we are not very good at measuring angles, numbers need to be shown
- we look at the angle in the center. *donut chart* is missing a circular area in the center.

Perceptual Properties

- NO 3D pie chart with **Perspective distortion**
- **Perception distortion**
  The Ebbinghaus Illusion
- **Perceived weight**

- **Spatialized Relations / Spatial Distribution**
- hierarchy, juxtaposition, proximity, grouping, separation by lines, alignment, connectivity, orientation
  - Does the spatialized relations help the production of meaning?
  - Gestalt laws
  - Unintentional meaning
  - If auto generated, spatial distribution is often determined by computer software, no by *semantic* value

- **Color and Brightness**
  1. **Sequential** schemes are suited to ordered data that progress from low to high. Lightness steps dominate the look of these schemes, with light colors for low data values to dark colors for high data values.
  2. **Diverging** schemes put equal emphasis on mid-range critical values and extremes at both ends of the data range. The critical class or break in the middle of the legend is emphasized with light colors and low and high extremes are emphasized with dark colors that have contrasting hues.
  3. **Qualitative** schemes do not imply magnitude differences between legend classes, and hues are used to create the primary

visual differences between classes. Qualitative schemes are best suited to representing nominal or categorical data.

Brewer, Cynthia A. 1994. Color use guidelines for mapping and visualization. Chapter 7 (pp. 123-147) in Visualization in Modern Cartography

Color scheme generation tool
http://colorbrewer2.org/#type=sequential&scheme=YlOrBr&n=3

3. Visual Language
"qualities" of the info/data

- Use semantically-resonant colors

- color theme, typography, and other visual elements
- visual forms and rules are coherent
- referent of the style, associations

Example with good visual
https://frightgeist.withgoogle.com

4. Interaction

*"The essence of data graphics is **visual comparisons**. This is still true for interactive graphics. Design a graphic that informs even when completely static. Then use interactivity — judiciously, powerfully, consistently — to subset the data space according to the reader's interests. Don't make interactivity a barrier to information."*
Bret Victor, How Many Households

- Are the interactions meaningful and necessary?
- Control level of details
- Undo and Redo

Example:
Anne Luther, *Collecting Contemporary Art: a visual analysis of a qualitative investigation into patterns of collecting and production.*
http://anneluther.info/research-1

5. Technical properties (digital visualizations)

one overarching characteristic unique to digital visualizations is the potential for the use of *libraries* - distributed software that aids in the creation of a visualization

- "Responsiveness" / user interface plasticity
    - How does the UX differ across different media?
    - "content is like water" - medium affects visualization
    - allowing designs to scale dynamically using a grid system
    - allowing static content to scale dynamically
    - media queries
    - **using a library might help us be more responsive**
- "Size"
    - Tufte measured this in raw quantities like *numbers*, and *ink*
    - He wanted to "maximize data-ink": a vs b
    - For digital visualizations, perhaps we can measure size in raw quantities like *bytes*
    - Size affects:
        - Initial load time for the visualization
        - Portability of the visualization
    - **using a library might increase overall size, affecting other properties - or if it's optimized it might replace larger parts of our own visualization**
- Complexity
    - Tufte thought about this as the number of lines a printer or artist would have to physically draw: how many times does a ruler touch paper?
    - A potential benchmark for complexity is LOC (source lines of code): how many times does the visualization architect have to instruct the computer?
        - Some estimation models used by professionals are explicitly based on LOC, e.g. COCOMO, which predicts complexity to **develop** & **maintain**
    - Complex visualizations might sacrifice other metrics to be more powerful or beautiful
    - **using a library might let us abstract out parts of our visualization, reducing complexity - or it might create complex interdependence**
        - This leads to the SLOC vs LOC distinction - how many lines of code are the *source* of your visualization?
- Locality
    - Digital content in the 21st century is almost always interlinked and hypertextual; unlike a visualization on paper we have to consider where things "live"
    - This include the data, the visualization, and potentially the libraries or tools used to make the visualization
    - Making content broadly available usually costs resources (hosting, domain)
    - Asynchronous fetching of resources can lead to bad UX, e.g. the visualization is ready but nothing has appeared because the underlying data is hosted in a slow location -> this can even create race conditions, "breaking" visualizations
    - **where does a library or tool "live"? do we have control over it?**
- Ownership

- Questions of rights can become complex when we are dealing with separate licenses for the data, a visualization, and any number of tools used to work with either of the two
- Modern licenses intended for digital distribution sometimes explicitly interact or have far-ranging implications; e.g. copyleft
- **Libraries can have rights different from the data itself, sometimes impacting the redistribution of data or even the usefulness of the visualization**

Case studies

Design heavy
Translating Worlds
https://www.behance.net/gallery/28096955/Translating-worlds-Corriere-della-Sera-La-Lettura

- x-axis by GDP per capita and on y-axis by total population speaking the language
- Per each language are displayed: number of books translated from the language, number of books translated to the language, connection between the language and the main translation.

One of the data source is a digital network:
http://language.media.mit.edu

https://www.behance.net/gallery/24692089/From-Vincent-to-Theo-Corriere-della-Sera-La-Lettura

Interactive
http://www.storiesbehindaline.com
a visual narrative of six asylum seekers' routes. They travelled from their hometown to Italy. This project wants to tell their stories through the data that shaped their personal travelling line.

Visualization of temporal events
Wind Map by Fernanda Viegas and Martin Wattenberg
http://hint.fm/wind/

- Exhibited at moma
- Abstract info has no graphical representation.
  Show the unseen, movement

Digital, traditional charts
https://public.tableau.com/profile/publish/IndicatorsofGlobalGrowthandDevelopment/Story#!/publish-confirm

Static vs interactive
Distribution of Slaves in 1860
https://public.tableau.com/en-us/s/gallery/1860-census-slavery-map?gallery=votd

https://www.census.gov/history/pdf/1860_slave_distribution.pdf

- State border
- Filled by patterns
- Reversed text color on dark background
- Rounded to 1 decimal place
- *Number distribution is determined by software*

https://public.tableau.com/en-us/s/gallery/alpine-skiing-paralympic-winter-games?gallery=votd

Static, original
https://www.behance.net/gallery/20364255/An-Information-Design-made-by-a-non-Designer

http://www.beautifulinenglish.com
analyzed all the single word translations of nouns & adjectives into English delivered through Google Translate for 10 of the most popular languages on Google.

Moore to look

Map of the First
http://infowetrust.com/scroll/

Tools
https://www.datawrapper.de