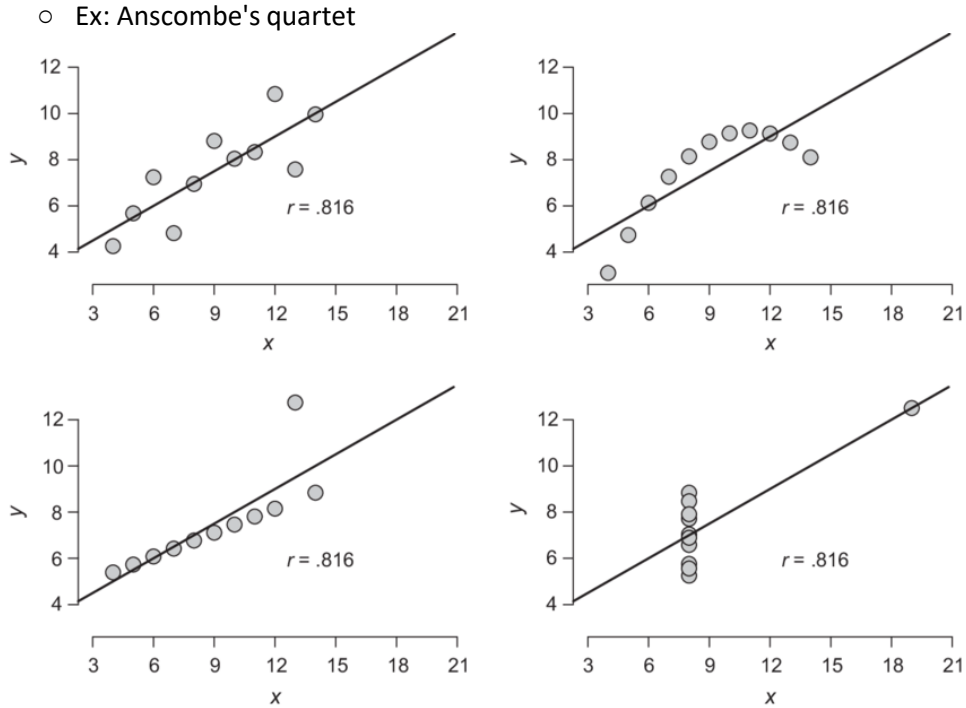
Tamara Ch. 1 - What's Vis, and Why Do It?

Wednesday, February 19, 2020 2:16 PM

- Data vis is the design and implementation of computer-based systems that provide visual represntations of datasets
- The goal of vis is to allow humans to carry out data anlysis taks more effectively
- Augmenting human abilities vs. strictly computational decision making models
- Limitations for data vis:
  - Computational complexity
  - Gradient on display (2D or 3D?)
  - The people (the human perceptual system)
- Humans vs. computers
  - Defining the problem first is important when deciding if you need a human or computer in the loop
  - Do you want a completely automated system?
  - Computation HFT (human intervation would slow this down)
  - Transitional backtesting (maybe you want to refine a model that will eventually be totally automated)
  - Long-term TSA (the goal is not to ever get rid of humans in the loop)
- Vis with computers versus by hand
  - You can display better precision with computers
  - Can have dynamic displays (interactive or animated)
- External representation and visual engagement
  - We use exernal representation to engage our perceptual system
  - Allows us to more easily make powerful inferences and engage both short and long term memory
  - We process visual information sequentially and in parallel
    - Parallel processing means intepreting information across entire field of vision
  - We can extract some "statistics" visually without scanning sequentially
  - Taste, smell and touch are not as efficient as sight, but they may aid vision in a multimodal approach that can amplify memory formation and other cognitive function
- Summary statistics can hide the true nature of a dataset
  - Ex: Anscombe's quartet



- Interaction

    - A single static view can only show one aspect of the data but in reality datasets have multiple dimensions
    - Interaction allows the designer to display more dimensions of the data without overwhelming the user
- Vis design space
    - Picking a specific vis idiom from a large space of options
    1. Encode info with spatial position
    2. Facet
    3. Reduce
- Tasks and effectiveness
    - Understand the type of data
    - What specific analysis tasks are normally run on the data?
    - Better to find a satisfactory solution than the optimal one
    - There will always be tradeoffs in the design choices
    - Possible space --> known space --> consideration space --> proposal space --> selected solution
- Validating this choice is difficult
- Resource limitations
    - Computational capacity
    - Human perceptual and cognitive capacity (memory and attention, change blindness)
    - Display capacity (screen resolution, information density)
- Analyzing for effective vis system design
    - What data does the user see?
    - Why does the user intend to use a vis tool?
    - How are the visual encodings and interaction idioms constructed?
- Discussion
    - Do these principles work for 3D designs?
        - There is a chapter that talks about techniques, and they do talk about 3D
        - The consensus is that it is much less effective as an ecoding, even though it is attention-grabbing
    - Effectiveness in the context of assessing if it is a good "graph." Are there domain-specific tests for task effectiveness?
        - Tamara says that measuring effectiveness is a very subjective endeavor
        - In chapter 4, she talks about nested task hierarchy. Part of what she talks about is validity. What you evaluate is dependent on what is valid.
        - She provides a methodology for thinking about this process

- Chapter 4: The Nested Model
    - Interface design in general has a nested set of properties, and it makes sense to go from outside-in
    1. Domain situation (politics, genomics, network modeling?)
        i. Who are the users/audience? What do they know? What language do they use? What is the setting? Is there lots of time to process information, or does it need to be rapid?
    2. Data/task abstraction (What's the goal? What data is needed?)
        i. Don't solve a problem that people don't care about
        ii. People don't want to "navigate the web," for example. Building a topological network isn't helpful because people really just want to find their exact
    3. Visual encoding/interaction idiom
        i. Perceptual and design limitations come into play here
    4. Algorithm
        i. This is the last step!
        ii. Attention to computation and performance
    - At any stage of this process you can ask yourself, "is this something people want?" or "Is this an appropriate data abstraction?"
    - Algorithm is arguably the "easiest" part, because the goal is clear: speed
    - Tamara breaks down these four nested categories into smaller, incremental steps,

and their possible pitfalls

- Once you implement the system, you can validate, working your way out of the nested model
  - Is it efficient enough? (algorithm)
  - Test on any users (visual encoding/interaction idiom)
  - Test on target users (data/task abstraction)
  - Observe adoption rates (domain situation)
- It is easy to start at the visual encoding step because we are trained in the technical part, and that can be a major "trap," even for experienced researchers

02/19/2020

COMS W6998
Comparison Discussion

-Presenter Celia Arsen
-Scribe Carmine Elvezio

Discussion Notes

- Note: Celia: Objects - items that we might want to compare
- Q: Prof: They're not really evaluations as much as examples of how not to use the system. Did the authors talk about how to utilize the framework in creating the use cases
    - Celia: The authors talk about strategies and considerations for using the considerations (e.g., impact of visual design)
- Q: Is the theoretical framework important?
- Note: Yiru: Before or after you design your visualization, something to assess against to determine whether you're doing a good job.
- Note: It's hard to determine
    - Celia: They did a lot of important work that could be useful to a lot of people in a lot of different ways
    - The emphasis on identifying the challenges is a helpful way of thinking about it
        - Mirrors the out-to-in nested modality
- Q: Do they really learn from this and then design their system or was it visa versa?
    - Celia: I've done that myself
    - Prof: Related work: the one that motivates your problem
    - Prof: This paper won best paper. Took a while to appreciate. Talked about visual encodings and the value of having a language. What doesn't really exist is a language for describing the task you want to perform.
        - There are a bunch of papers trying to get towards describing the tasks and how you would want to do something.
        - Each of these papers says they're the best for their tasks. Are they all actually doing the same thing when you pull them altogether?
    - A: Naming them having the structure as is done in the paper helps to organize your thoughts better.
    - Celia: People who are working on the same thing can communicate about it in the same way.
    - Prof: Now we can have a language to talk about graphs in the comparison task they're trying to accomplish.
- Q: Talking about more interactive visualizations - does this framework extend to that - how might it be extended or applied to account for that?
    - A: Deka: The Tamara book chapter talks about this about. The bit on subset of data would discuss that here.
    - Prof: Scroll bar is scanning sequentially. That seems to fit into this framework.
        - That falls into the strategy of subsetting
- Q: Do people think this framework is comprehensive?

        - Celia: Could you use this framework to understand and approach any data vis task?
        - Deka: Is all data vis is a comparison task?
        - Celia: He didn't say it directly but he implies it.
        - Prof: He said it to me directly.
        - Prof: What about a simple task with one value?
        - A: It might have a comparison implicitly?
- Q: Yiru: Visual Design is very simple - there are more things to consider?
    - A: Deka: He calls out the other stuff and wants to focus on his own thing
    - Celia: He didn't connect the challenges to the solutions, but that wasn't his intent.
    - Does he leave open the possibility to scale down?
        - Prof: Might be scan sequentially and select subset.
    - Deka: Data transformation- is it covered here?
        - A: This is more of a philosophical paper
        - Celia: Here's the methodology but then you need to think about what the solution is
        - Prof: Can you write this down in some type of grammar?
    - Celia: This can be used as a pre-Draco thing - as input considerations for Draco
        - Draco is the second most inner part of the iterative design model shown in the beginning of the class
        - Prof: They're different things and why we looked at stuff in the order in which we have in the class so far.
- Prof Discussion Slides:
    - Titanic Death chart
        - What are the targets
            - Passengers, gender, if survived
        - Actions
            - Celia: They're comparing between survived and not, and genders.
            - Comparing the ratio of male and female
    - NYT Changing Nature of Middle-Class Jobs
        - Let's list all of the comparisons
            - How would we articulate this as a set of comparison tasks
                - 1980 - 2012 (of some sort)
                - Jobs/thousand
                - Gender breakdown
                - Occupational categories
                - Job Types
                - Implicit: gender parity and expectation
        - "Is there a better re-design if that is what I'm trying to express"
        - One of the least clear things in this chart is the gender changes over time
        - A: I feel like it's hard to understand which jobs are which
            - A: It is doable with a mouse hover
- Prof: Why would you use interaction at all?
    - Interaction helps and goes towards the human constraint part of this.
    - John et al. graph - you have to memorize as you try to analyze
        - Interaction would help as you can drag things around
    - It helps you to navigate to filter and re-specify the task
- Prof: Why is Scale the thing?
    - It is a little generic (by intent)
        - Celia: The author knows as you progress more data you use in more disciplines will not have been designed for analysis
            - We're scraping this from natural processes

                - Are the three challenges what it will continue to be? If another axis emerges, he has established a way to think about it.
                - All these people are designing tools for data exploration
        - Prof: If you were designing a tool to identify interesting relationships
            - Some domains or statistics are actually the ones that matter
- Deka: It's hard to say If a particular target has a relationship or not
- Prof: The way I interpreted the entire paper is simple: I have two things I want to compare, and the action is the line between them.
    - Deka: Comparison is the whole thing
    - There are only so many things you can talk about when it comes to two nodes and an edge
- Celia: What do people think of the evaluation he provided?
    - The TopicModelling one was weird
    - Prof: I tried to reconstruct what the set of comparisons was and determine that from the chart as opposed to reading the evaluation itself
    - Subtle differences only a domain expert would care about
- Celia: I'm not really sold on the "few" case studies and then it's "proven." Giving the framework to other designers and seeing if they can use it to improve their design.
    - A: Prof: Good for future work.
    - Prof: Some subset of this could make sense as a tool for doing layout or something. Possibly adding something on top of Draco.
    - Celia: Not everything needs to be a tool. Helping to understand your process of research.
    - Deka: Understanding your task to make better visualizations
- Yiru: Section 4 says difficulty can come from complexity or size off items
    - Prof: Comparing two really large source files where it might not be natural to break up.
    - Yiru: How do you know which aspects to compare if it can't be broken up?
    - Prof: It's probably hierarchical
        - Nodes can be compared, but sets of nodes might also be comparable.
        - A source file is also potentially describable in the same diagram.
            - Many to many relationship of characters

Presentation Notes

- Considerations for visualizing comparison
- Michael Gleicher - University of Wisconsin
- Lack of abstract scalable framework for design comparison tasks
- Context
    - Graph - Napoleon's march through Russia
    - Drawing: comparison of items such that one can compare size
- Contribution: Framework that abstracts comparison challenges
- The **4 Considerations**
    - Elements, Challenges, Strategies, Designs
    - Elements
        - Target: Set of Items that are relation
        - Action that we want to perform
        - Other things that we can do besides just identifying a difference
    - Comparative Challenges
        - # of items being compared
        - The size or complexity of individual items

        - The size or complexity of the relationships
        - This is described as a set of three axes
    - Identify Comparative Strategies
        - Sequential Scan
        - Select Subset
        - Summarize
    - Identify Comparative Visual Design
        - Juxtaposition
        - Superposition
        - Explicit encoding
- Evaluation
    - Sequence Surveyor
        - Using the framework to improve over Mauve
        - Use all three strategies
    - LayerCake
    - Topic Model Comparison