Course Logistics
Async so far

• last week
  – async read only
    • Course Logistics (no comments, no responses)
  – async read & comment
    • VAD Ch 1: Why Visualization? (comments only, no responses)
  – async discuss
    • self-intros

• this week
  – async read & comment & respond
    • VAD Ch 2: Data Abstraction
    • VAD Ch 3: Task Abstraction
    • paper: Nested Model [basis for VAD Ch 4]
Updates

• All students moved from waitlist to registered
• Official enrolment now 38
• Very likely to move to Forestry (FSC) 2330 starting next week
  – especially if ventilation here in SWNG 207 remains terrible!
• Stay tuned for Canvas marks updates
Discussion: Round 1
Exercise: Abstractions
Now: In-class design exercise, in small groups

• Abstractions
  – practice with data & task abstractions, on concrete example: Aid to Countries
  – crucial ideas: determine cardinalities/ranges
    – precondition for all decisions about visual encoding

• Small-group exercise: 60-ish min
  – breakout groups (4 people/group)
  – googledoc worksheets, as before
  – document in your group's googledoc w/ text as you go!
  – reportbacks, as before (intermediate and final)
    – I'll flip through googledocs, some questions for group spokesperson
Discussion: Round 2
Next week

• to read & discuss (async, before next class)
  – VAD book, Ch 5: Marks & Channels
  – VAD book, Ch 6: Rules of Thumb
  – paper: Design Study Methodology
Backup/Reference Slides
Ch 1. What's Vis, and Why Do It?
Visualization defined & motivated

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

• human in the loop needs the details
  – doesn't know exactly what questions to ask in advance
  – longterm exploratory analysis
    • speed up through human-in-the-loop visual data analysis
  – presentation of known results
  – stepping stone towards automation: refining, trustbuilding
  – interplay between human judgement and automatic computation

• intended task, measurable definitions of effectiveness

short version: alternate to next 3 slides
Defining visualization (vis)

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Why?...
Visualization (vis) defined & motivated

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

• human in the loop needs the details & no trusted automatic solution exists
  – doesn't know exactly what questions to ask in advance
  – exploratory data analysis
    • **speed up** through human-in-the-loop visual data analysis
  – present known results to others
  – stepping stone towards automation
    – before model creation to provide understanding
    – during algorithm creation to refine, debug, set parameters
    – before or during deployment to build trust and monitor
Why use an external representation?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- external representation: replace cognition with perception

Why depend on vision?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

• human visual system is high-bandwidth channel to brain
  – overview possible due to background processing
    • subjective experience of seeing everything simultaneously
    • significant processing occurs in parallel and pre-attentively

• sound: lower bandwidth and different semantics
  – overview not supported
    • subjective experience of sequential stream

• touch/haptics: impoverished record/replay capacity
  – only very low-bandwidth communication thus far

• taste, smell: no viable record/replay devices
Why represent all the data?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- summaries lose information, details matter
  - confirm expected and find unexpected patterns
  - assess validity of statistical model

**Anscombe’s Quartet**

<table>
<thead>
<tr>
<th>Identical statistics</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>x mean</td>
<td>9</td>
<td>9.5</td>
</tr>
<tr>
<td>x variance</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>y mean</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>y variance</td>
<td>3.75</td>
<td>3.75</td>
</tr>
<tr>
<td>x/y correlation</td>
<td>0.816</td>
<td>0.816</td>
</tr>
</tbody>
</table>

https://www.youtube.com/watch?v=DbJyPELmhJc

Same Stats, Different Graphs
Visualization defined & motivated

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- suitable when human in the loop needs details
- interplay between human judgement and automatic computation

Anscombe’s Quartet

<table>
<thead>
<tr>
<th></th>
<th>x mean</th>
<th>y mean</th>
<th>x variance</th>
<th>y variance</th>
<th>x/y correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>9</td>
<td>7.5</td>
<td>10</td>
<td>3.75</td>
<td>0.816</td>
</tr>
</tbody>
</table>

Datasaurus Dozen

Matejka & Fitzmaurice
Why focus on tasks and effectiveness?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

• effectiveness requires match between data/task and representation
  – set of representations is huge
  – many are ineffective mismatch for specific data/task combo
  – increases chance of finding good solutions if you understand full space of possibilities

• what counts as effective?
  – novel: enable entirely new kinds of analysis
  – faster: speed up existing workflows

• how to validate effectiveness
  – many methods, must pick appropriate one for your context
What resource limitations are we faced with?

Vis designers must take into account three very different kinds of resource limitations: those of computers, of humans, and of displays.

- **computational limits**
  - processing time
  - system memory

- **human limits**
  - human attention and memory

- **display limits**
  - pixels are precious resource, the most constrained resource
  - **information density**: ratio of space used to encode info vs unused whitespace
    - tradeoff between clutter and wasting space, find sweet spot between dense and sparse
Why analyze?

- imposes structure on huge design space
- scaffold to help you think systematically about choices
- analyzing existing as stepping stone to designing new
- most possibilities ineffective for particular task/data combination
### How?

#### Encode

<table>
<thead>
<tr>
<th>Arrange</th>
<th>Map</th>
<th>Manipulate</th>
<th>Facet</th>
<th>Reduce</th>
</tr>
</thead>
<tbody>
<tr>
<td>Express</td>
<td>Color</td>
<td>Change</td>
<td>Juxtapose</td>
<td>Filter</td>
</tr>
<tr>
<td>Separate</td>
<td>Hue</td>
<td>Select</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Align</td>
<td>Saturation</td>
<td>Partition</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Luminance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Size, Angle, Curvature, ...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shape</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ ● ■ △</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Motion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Direction, Rate, Frequency, ...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Use</th>
<th>Navigate</th>
<th>Superimpose</th>
<th>Embed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt; ⬇️ ⬆️</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Further reading

  – Chap 1: What’s Vis, and Why Do It?


Ch 2. What: Data Abstraction
**Datasets**

- **Data Types**
  - Items
  - Attributes
  - Links
  - Positions
  - Grids

- **Data and Dataset Types**
  - Tables
  - Networks & Trees
  - Fields
  - Geometry
  - Clusters, Sets, Lists

**Attributes**

- **Attribute Types**
  - Categorical
  - Ordered
  - Quantitative

- **Dataset Availability**
  - Static
  - Dynamic

**Dataset Types**

- **Tables**
  - Items
  - Attributes
  - Cell containing value

- **Networks**
  - Nodes
  - Links
  - Grids

- **Fields (Continuous)**
  - Attributes (columns)
  - Value in cell

- **Geometry (Spatial)**
  - Clusters, Sets, Lists

- **Dataset Availability**
  - Static
  - Dynamic
Types: Datasets and data

**Dataset Types**
- Tables
- Networks

**Attribute Types**
- Categorical
- Ordered
  - Ordinal
  - Quantitative

**Ordering Direction**
- Sequential
- Diverging
- Cyclic

**Spatial**
- Fields (Continuous)
- Geometry (Spatial)
Three major datatypes

Dataset Types

- **Tables**
  - Attributes (columns)
  - Items (rows)
  - Cell containing value
  - Multidimensional Table

- **Networks**
  - Link
  - Node (item)
  - Trees

- **Spatial**
  - Fields (Continuous)
  - Geometry (Spatial)
  - Grid of positions
  - Cell
  - Attributes (columns)
  - Value in cell

- **Visualization vs Computer Graphics**
  - Geometry is design decision
Dataset and data types

Data and Dataset Types

<table>
<thead>
<tr>
<th>Tables</th>
<th>Networks &amp; Trees</th>
<th>Fields</th>
<th>Geometry</th>
<th>Clusters, Sets, Lists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items</td>
<td>Items (nodes)</td>
<td>Grids</td>
<td>Items</td>
<td>Items</td>
</tr>
<tr>
<td>Attributes</td>
<td>Links</td>
<td>Positions</td>
<td>Positions</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Attributes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data Types

- Items
- Attributes
- Links
- Positions
- Grids

Dataset Availability

- Static
- Dynamic
Attribute types

- **Attribute Types**
  - Categorical
  - Ordered
  - Ordinal
  - Quantitative

- **Ordering Direction**
  - Sequential
  - Diverging
  - Cyclic
Further reading, full Ch 2


• The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations Ben Shneiderman, Proc. 1996 IEEE Visual Languages


• Visualization of Time-Oriented Data. Wolfgang Aigner, Silvia Miksch, Heidrun Schumann, Chris Tominski. Springer 2011.
Ch 3. Why: Task Abstraction
• \{action, target\} pairs
  – discover distribution
  – compare trends
  – locate outliers
  – browse topology

**Actions**
- **Analyze**
  - Consume
    - Discover
  - Produce
    - Annotate
    - Record
    - Derive

- **Search**
  - Target known
    - Location known
    - Lookup
    - Location unknown
    - Locate
    - Explore
  - Target unknown
    - Enjoy

- **Query**
  - Identify
  - Compare
  - Summarize

**Targets**
- **All Data**
  - Trends
  - Outliers
  - Features

- **Attributes**
  - One
    - Distribution
  - Many
    - Dependency
    - Correlation
    - Similarity

- **Network Data**
  - Topology
    - Paths

- **Spatial Data**
  - Shape
Actions: Analyze, Query

- analyze
  - consume
    • discover vs present
      - aka explore vs explain
    • enjoy
      - aka casual, social
  - produce
    • annotate, record, derive

- query
  - how much data matters?
    • one, some, all

- independent choices
  - analyze, query, (search)
Actions: Analyze

• consume
  – discover vs present
    • classic split
    • aka explore vs explain
  – enjoy
    • newcomer
    • aka casual, social

• produce
  – annotate, record
  – derive
    • crucial design choice
Derive

• don’t just draw what you’re given!
  – decide what the right thing to show is
  – create it with a series of transformations from the original dataset
  – draw that

• one of the four major strategies for handling complexity

Original Data

Derived Data
Analysis example: Derive one attribute

- Strahler number
  - centrality metric for trees/networks
  - derived quantitative attribute
  - draw top 5K of 500K for good skeleton


Task 1

- **What?**
  - In Tree
  - Out Quantitative attribute on nodes

- **Why?**
  - Derive

Task 2

- **What?**
  - In Tree
  - In Quantitative attribute on nodes
  - Out Filtered Tree

- **Why?**
  - Summarize
  - Reduce
  - Filter

- **How?**
  - Topology
Actions: Search, query

• what does user know? ➔ Search
  – target, location

• how much of the data matters?
  – one, some, all

• independent choices for each of these three levels
  – analyze, search, query
  – mix and match

<table>
<thead>
<tr>
<th></th>
<th>Target known</th>
<th>Target unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location known</td>
<td>Lookup</td>
<td>Browse</td>
</tr>
<tr>
<td>Location unknown</td>
<td>Locate</td>
<td>Explore</td>
</tr>
</tbody>
</table>
Why: Targets

🔄 All Data

➡️ Trends

➡️ Outliers

➡️ Features

🔄 Attributes

➡️ One

➡️ Distribution

➡️ Extremes

➡️ Many

➡️ Dependency

➡️ Correlation

➡️ Similarity

🔄 Network Data

➡️ Topology

➡️ Paths

🔄 Spatial Data

➡️ Shape
Further reading

  – Chap 2: What: Data Abstraction
  – Chap 3: Why: Task Abstraction


Further reading, full Ch 3


- What does the user want to see?: what do the data want to be? A. Johannes Pretorius and Jarke J. van Wijk. Information Visualization 8(3):153-166, 2009.

- Chapter 1, Readings in Information Visualization: Using Vision to Think. Stuart Card, Jock Mackinlay, and Ben Shneiderman, Morgan Kaufmann 1999.


Ch 4. Analysis: Four Levels for Validation
How to evaluate a visualization: So many methods, how to pick?

• Computational benchmarks?
  – quant: system performance, memory

• User study in lab setting?
  – quant: (human) time and error rates, preferences
  – qual: behavior/strategy observations

• Field study of deployed system?
  – quant: usage logs
  – qual: interviews with users, case studies, observations

• Analysis of results?
  – quant: metrics computed on result images
  – qual: consider what structure is visible in result images

• Justification of choices?
  – qual: perceptual principles, best practices
Nested model: Four levels of visualization design

- **domain situation**
  - who are the target users?

- **abstraction**
  - translate from specifics of domain to vocabulary of visualization
    - **what** is shown? data abstraction
    - **why** is the user looking at it? task abstraction

- **idiom**
  - **how** is it shown?
    - **visual encoding** idiom: how to draw
    - **interaction** idiom: how to manipulate

- **algorithm**
  - efficient computation
Nested model: Four levels of visualization design

- **domain situation**
  - who are the target users?

- **abstraction**
  - translate from specifics of domain to vocabulary of visualization
    - **what** is shown? **data** abstraction
    - **why** is the user looking at it? **task** abstraction
      - often must transform data, guided by task

- **idiom**
  - **how** is it shown?
    - **visual encoding** idiom: how to draw
    - **interaction** idiom: how to manipulate

- **algorithm**
  - efficient computation
Different threats to validity at each level

- cascading effects downstream

- **Domain situation**
  - You misunderstood their needs

- **Data/task abstraction**
  - You're showing them the wrong thing

- **Visual encoding/interaction idiom**
  - The way you show it doesn't work

- **Algorithm**
  - Your code is too slow
Interdisciplinary: need methods from different fields at each level

• mix of qual and quant approaches (typically)

Mismatches: Common problem

- **Domain situation**
  - Observe target users using existing tools

- **Data/task abstraction**
  - **Visual encoding/interaction idiom**
    - Justify design with respect to alternatives
  - **Algorithm**
    - Measure system time/memory
    - Analyze computational complexity
  - Analyze results qualitatively
  - Measure human time with lab experiment (*lab study*)
  - Observe target users after deployment (*field study*)
  - Measure adoption

benchmarks can't confirm design

lab studies can't confirm task abstraction

Analysis examples: Single paper includes only subset of methods

- observe and interview target users
- justify encoding/interaction design
- measure system time/memory
- qualitative result image analysis

- observe and interview target users
- justify encoding/interaction design
- qualitative result image analysis
- field study, document deployed usage

An energy model for visual graph clustering. (LinLog)
Noack. Graph Drawing 2003
- qualitative/quantitative image analysis

Effectiveness of animation in trend visualization.
- lab study, measure time/errors for operation

Interactive visualization of genealogical graphs.
McGuffin and Balakrishnan. InfoVis 2005.
- justify encoding/interaction design
- qualitative result image analysis
- test on target users, get utility anecdotes

- justify encoding/interaction design
- computational complexity analysis
- measure system time/memory
- qualitative result image analysis
Further reading

  – Chap 4: Analysis: Four Levels for Validation


• How to do good research, get it published in SIGKDD and get it cited!, Eamonn Keogh, SIGKDD Tutorial 2009.


• Externalisation - how writing changes thinking. Alan Dix. Interfaces, Autumn 2008.
Guerilla/Discount Usability

• grab a few people and watch them use your interface
  – even 3-5 gives substantial coverage of major usability problems
  – agile/lean qualitative, vs formal quantitative user studies
    • goal is not statistical significance!

• think-aloud protocol
  – contextual inquiry (conversations back and forth) vs fly on the wall (you’re silent)
Further reading, usability

• 7 Step Guide to Guerrilla Usability Testing, Markus Piper
  – https://userbrain.net/blog/7-step-guide-guerrilla-usability-testing-diy-usability-testing-method

• The Art of Guerrilla Usability Testing, David Peter Simon
  – http://www.uxbooth.com/articles/the-art-of-guerrilla-usability-testing/

• Discount Usability: 20 Years, Jakob Nielsen
  – https://www.nngroup.com/articles/discount-usability-20-years/

• Interaction Design: Beyond Human-Computer Interaction

• About Face: The Essentials of Interaction Design

• Task-Centered User Interface Design. Lewis & Rieman, 1994
  – http://hcibib.org/tcuid/