Dataset Types

- Items
- Tables
- Attributes (columns)
- Value in cell

Trees

Link

Dataset types

• flat table
– one item per row
– each column is attribute
– cell holds value for item-attribute pair

• grid types

Data abstraction: Three operations

- translate from domain-specific language to generic visualization language
- identify dataset type(s), attribute types
- identify cardinality

– how many items in the dataset?
– what is cardinality of each attribute?
– number of levels for categorical data
– range for quantitative data

– consider whether to transform data
– guided by understanding of task

Visualization Analysis & Design

Data Abstraction (Ch 2)

Visualization Analysis & Design

Analysis: Nested Model (Ch 4)

Dataset types

• multidimensional tables
– indexing based on multiple keys
– eg genes, patients

Spatial fields

– attribute values associated w/ cells
– cell contains value from continuous domain
– eg temperature, pressure, wind velocity

– measured or simulated

– major concerns
– sampling
– where attributes are measured
– interpolation
– how to model attributes elsewhere
– grid types

– major divisions
– attributes per cell
– scalar (1), vector (2), tensor (3+)

Spatial fields

– attribute values associated w/ cells
– cell contains value from continuous domain
– eg temperature, pressure, wind velocity

– measured or simulated

– major concerns
– sampling
– where attributes are measured
– interpolation
– how to model attributes elsewhere
– grid types

Nested model

– downstream: cascading effects

– upstream: iterative refinement

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
– schedule
• this week
– async read & comment & respond
– VAD Ch. 3: 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 33
• Canvas link added
– future assignment handin
– soon marks for sync & async participation (posted weekly)

Mini-Lecture, Q&A: Round 1

Table

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Age</th>
<th>Shirt Size</th>
<th>Favorite Fruit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Amy</td>
<td>26</td>
<td>S</td>
<td>Apple</td>
</tr>
<tr>
<td>2</td>
<td>Miles</td>
<td>31</td>
<td>M</td>
<td>Orange</td>
</tr>
<tr>
<td>3</td>
<td>Clive</td>
<td>34</td>
<td>M</td>
<td>Orange</td>
</tr>
<tr>
<td>4</td>
<td>Leonard</td>
<td>24</td>
<td>L</td>
<td>Grape</td>
</tr>
<tr>
<td>5</td>
<td>Ernest</td>
<td>32</td>
<td>L</td>
<td>Orange</td>
</tr>
<tr>
<td>6</td>
<td>Susan</td>
<td>28</td>
<td>M</td>
<td>Olive</td>
</tr>
<tr>
<td>7</td>
<td>George</td>
<td>26</td>
<td>S</td>
<td>Orange</td>
</tr>
<tr>
<td>8</td>
<td>Wendy</td>
<td>33</td>
<td>L</td>
<td>Peach</td>
</tr>
<tr>
<td>9</td>
<td>Sam</td>
<td>32</td>
<td>M</td>
<td>Tangerine</td>
</tr>
</tbody>
</table>

Dataset types

- template: 
- one key per row
- each item holds attribute
- cell contains value from continuous domain

– eg temperature, pressure, wind velocity
– interpolation
– how to model attributes elsewhere
– grid types
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?
  - qual: 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

Paper types: Validation
- design studies
  - qualitative discussion of result images/videos
  - abstraction & idiom validation for technique papers: computational benchmarks
  - idiom validation for technique papers: controlled experiments
  - evaluation
    - (controlled experiment as primary contribution)
- system
  - show power for developer using system

Analysis examples: Single paper includes only subset of methods

- Justification of choices?
- Analysis of results?
- Computational benchmarks?
  - 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

Visualization is suitable when there is a need to augment human capabilities

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

Why?

- 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
  - strategy: model creation to provide understanding
  - during algorithm creation to refine, debug, set parameters
  - before or during deployment to build trust and monitor the system

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

Reading visualization papers
- one strategy: multiple passes
  - title
  - abstract, authors/affiliation
  - flip through, glance at figures, notice structure from section titles
  - skim intro, results/discussion (maybe conclusion)
  - fast read to get big ideas
  - if you don’t get something, just keep going
  - second pass to work through details
  - linear parts may cast light on earlier parts for badly structured papers
  - third pass to dig deep
  - if you're presenting it to class,
    - decide when to stop reading: is this relevant to my current concerns?

Data abstraction: Three operations

- translate from domain-specific language to generic visualization language
  - identify dataset type(s), attribute types
  - identify cardinality
    - how many items in the dataset?
    - what is cardinality of each attribute?
    - number of levels for categorical data
    - range for quantitative data
  - consider whether to transform data
    - guided by understanding of task

Defining visualization (vis)

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

Exercise: Abstractions

- translate from domain-specific language to generic visualization language
- identify cardinality
- consider whether to transform data

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)
    - googledocs worksheets, as before
    - document in your group’s googledocs, w/ next as you go
    - reportbacks, as before (intermediate and final)
    - I’ll flip through googledocs, some questions for group spokesperson

Backup/Reference Slides

Ch 1. What’s Vis, and Why Do It?
Why use an external representation?

- external representation: replace cognition with perception

Why depend on vision?

- human visual system is high-bandwidth channel to brain
  - overview possible due to background processing
  - difficult to control
  - processing time

Why analyze?

- computational limits
  - processing time
  - system memory
  - data limits
  - human attention and memory

Why represent all the data?

- display limits

Why does data mean?

- What does this sequence of six numbers mean?
  - two points far from each other in 3D space, with 15 links between them, and a weight of 100001 for the link?
  - something else??

Further reading

- Visualization Analysis & Design, Munzner. AK Peters Visualization Series. CRC Press, 2014
- Current approaches to shape blindness. Davis J, Simon. Visual Cognition 7:123 (2003), 1-16

Visualization Analysis & Design

Data Abstraction (Ch 2)
Dataset types

- **Flat table**
  - one item per row
  - each column is attribute
  - cell holds value for item-attribute pair

Items & Attributes

- **Item**: individual entity, discrete
  - e.g. patient, car, stock, city
  - “independent variable”

What does data mean?

14, 2.6, 30, 30, 15, 100001

- What does this sequence of six numbers mean?
  - two points far from each other in 3D space?
  - two points close to each other in 3D space, with 15 links between them, and a weight of 100001 for the link?
  - something else?

Basil, 7, S, Pear

- What about this data?
  - food shipment of produce (basil & pear) arrived in satisfactory condition on 7th day of month

Now what?

- **Semantics**: real-world meaning
- **Mathematical interpretation of data**

Items & Attributes

- **Item**: individual entity, discrete
  - e.g. patient, car, stock, city
  - “independent variable”

Dataset types

- **Table**
  - one item per row
  - each column is attribute
  - cell holds value for item-attribute pair

What about this data?

- food shipment of produce (basil & pear) arrived in satisfactory condition on 7th day of month
- Bad Point neighborhood of city had 7 inches of snow cleared by the Pear Creek Limited snow removal service

- Item, link, attribute, position, (grid)

14, 2.6, 30, 30, 15, 100001

- What does this sequence of six numbers mean?
  - two points close to each other in 2D space, with 15 links between them, and a weight of 100001 for the link?
  - something else?

Basil, 7, S, Pear

- What about this data?
  - food shipment of produce (basil & pear) arrived in satisfactory condition on 7th day of month
  - Bad Point neighborhood of city had 7 inches of snow cleared by the Pear Creek Limited snow removal service

Now what?

- **Mathematical interpretation of data**
- **Semantics**: real-world meaning

Items & Attributes

- **Item**: individual entity, discrete
  - e.g. patient, car, stock, city
  - “independent variable”

Other data types

- **Links**
  - express relationship between two items
  - e.g. friendship on Facebook, interaction between proteins

- **Positions**
  - spatial data location in 2D or 3D
  - pixels in photo, voxels in MRI scan, latitude/longitude

- **Grids**
  - sampling strategy for continuous data

What does data mean?

14, 2.6, 30, 30, 15, 100001

- What does this sequence of six numbers mean?
  - two points close to each other in 3D space, with 15 links between them, and a weight of 100001 for the link?
  - something else?
Data and Dataset Types

Data Types

Dataset Availability

Geometry

Attributes (columns)

Position

Trees

Attributes

Datasets

Networks

Dynamic

Attributes

Grids

Positions

Items

Positions

Items

Grid of positions

Position

What?

Cell

Grid of positions

Continuous

Value in cell

sets, lists

Clusters,

Items

Grids

Cyclic

Quantitative

Data vs conceptual model, example

• data model: floats
  – 32.52, 54.06, -14.35, ...
• conceptual model
  – temperature
  – continuous to 2 significant figures: quantitative
  – task: forecasting the weather
  – hot, warm, cold-ordinal
  – task: deciding if bath water is ready
  – above freezing, below freezing: categorical
  – task: decide if I should leave the house today

Data abstraction: Three operations

• translate from domain-specific language to generic visualization language
• identify dataset type(s), attribute types
• identify cardinality
  – how many items in the dataset?
  – what is cardinality of each attribute?
  – number of levels for categorical data
  – range for quantitative data
• consider whether to transform data
  – guided by understanding of task

Data vs conceptual models

• data model
  – mathematical abstraction
  – sets with operations, eg floats with + −
  – variable data types in programming languages
• conceptual model
  – mental construction (semantics)
  – supports reasoning
  – typically based on understanding of tasks [stay tuned!]
  – data abstraction process relies on conceptual model
  – for transforming data if needed

Data vs conceptual model, example

• data model: floats
  – 32.52, 54.06, -14.35, ...
• conceptual model
  – temperature
  – continuous to 2 significant figures: quantitative
  – task: forecasting the weather
  – hot, warm, cold-ordinal
  – task: deciding if bath water is ready

Derived attributes

• derived attribute: compute from originals
  – simple change of type
  – acquire additional data
  – derived attribute: compute from originals
  – simple change of type
  – acquire additional data

Analysis example: Derive one attribute

• derived attribute: compute from originals
  – simple change of type
  – acquire additional data
  – complex transformation
Design process

Task Abstraction (Ch 3)

From domain to abstraction
- domain characterization: details of application domain
  - group of users, target domain, their questions & data
  - varies wildly by domain
  - must be specific enough to get traction

Actions: Search
- what does user know?
  - target, location

Actions: Search
- what does user know?
  - target, location

Actions: Search
- what does user know?
  - target, location

Actions: Search
- what does user know?
  - target, location

Actions: Search
- what does user know?
  - target, location

Further reading, full Ch 2
- 1996 IEEE Visualization.
- The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization Ben Shneiderman. Proc. 1996 IEEE Visualization Languages}

From domain to abstraction
- domain characterization: details of application domain
  - group of users, target domain, their questions & data
  - varies wildly by domain
  - must be specific enough to get traction
- domain questions/problems
  - break down into simpler abstract tasks
- abstraction: data & task
  - map what and why into generalized terms

•{domain characterization: details of application domain
- group of users, target domain, their questions & data
- varies wildly by domain
- must be specific enough to get traction
- domain questions/problems
- break down into simpler abstract tasks
- abstraction: data & task
- map what and why into generalized terms

A very high-level pattern
- {action, target} pairs
- discover distribution
- compare trends
- locate outliers
- browse topology

Actions: Analyze
- analyze
- high-level choices
- search
- find a known/unknown item
- query
- find out about characteristics of item
- targets
- what is being acted on

• very high-level pattern
- {action, target} pairs
- discover distribution
- compare trends
- locate outliers
- browse topology

Actions: Search
- what does user know?
- target, location

Actions: Search
- what does user know?
- target, location

Actions: Search
- what does user know?
- target, location

Actions: Search
- what does user know?
- target, location

Actions: Search
- what does user know?
- target, location

Actions: Search
- what does user know?
- target, location

Visual encoding/interaction idiom

Algorithm

Domain situation

Data/task abstraction

Visual encoding/interaction idiom

Algorithm

Domain situation
Why is validation difficult?
• solution: use methods from different fields at each level

Domain situation
• observation
  – translate from specific domain to vocabulary of vis
  – 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

Solution: use methods from different fields at each level

Analysis framework: Four levels, three questions
• domain situation
  – who are the target users?

Nested model
• downstream: cascading effects
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.

Usability

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/