Visualization Analysis & Design

Full-Day Tutorial

Tamara Munzner
Department of Computer Science
University of British Columbia

Outline
• Session 1 8:30-10:00am
 Visualization Analysis Framework
 – Introduction: Definitions
 – Analysis: What, Why, How
 – Marks and Channels
 – Spatial Layout
 – Arrange Tables
 – Arrange Networks and Trees
• Session 3 1:00-2:30pm
 Guidelines and Examples
 – Rule of 3:3:3
 – Rules of Thumb
 – Q&A

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

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

• don’t need vis when fully automatic solution exists and is trusted
• many analysis problems ill-posed
• don’t know exactly what questions to ask in advance
• possibilities
– long-term use for end users (e.g. exploratory analysis of scientific data)
– presentation of known results
– stepping stone to better understanding of requirements before developing models
– helps developers of automatic solution refine/debug, determine parameters
– help end users of automatic solution verify/believe results

Why is validation difficult?
• solution: use methods from different fields at each level
  – idiosyncrasies don’t serve all tasks equally
  – challenge: recast tasks from domain-specific vocabulary to abstract forms
  – most possibilities ineffective
  – validation is necessary but tricky
  – increases chance of finding good solutions if you understand full space of possibilities
  – what counts as effective
– novel enabling entirely new kinds of analysis
– faster speed up existing workflows

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

Human visual system is high-bandwidth channel to brain
– overview possible due to background processing
– subjective experience of seeing everything simultaneously
– subjective experience of sequential stream
– touch/haptic: impoverished record/replay capacity
– only very low-bandwidth communication thus far
– taste, smell, no viable record/replay devices

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
– 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
– subjective experience of seeing everything simultaneously
– most possibilities ineffective for hands-free task
– most possibilities ineffective for human limits
– subjectivity of visual perception
– human limits
– human time
– human the argument
– human time with lab

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
– subjective experience of sequential stream
– sound: lower bandwidth and different semantics
– subjective experience of sequential stream
– touch/haptic: impoverished record/replay capacity
– only very low-bandwidth communication thus far
– taste, smell. no viable record/replay devices

Why analyze?
• imposes structure on huge design space
– scaffolding to help you think systematically about choices
– assessing existing as stepping stone to designing new
– most possibilities ineffective for particular task/data combination
– space

Further reading
– Chap. 1: What’s Vis, and Why Do It?
– Chap. 2: Why is validation difficult?
– Chap. 3: Why is validation difficult?

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

Why focus on tasks and effectiveness?
Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

• tasks serve as constraint on design (as does data)
– idioms do not serve all tasks equally
– challenge: recast tasks from domain-specific vocabulary to abstract forms
• most possibilities ineffective
– validation is necessary but tricky
• increases chance of finding good solutions if you understand full space of possibilities
– what counts as effective
– novel enabling entirely new kinds of analysis
– faster speed up existing workflows

Who are the target users?
• solution: use methods from different fields at each level

Why adoption?
• solution: use methods from different fields at each level
How?
Encode Manipulate Facet Reduce

• many channels: tilt, size, shape, proximity, shadow direction, ...
• but not all! parallel line pairs do not pop out from tilted pairs

Outline
Magnitude Channels: Ordered Attributes
Identity Channels: Categorical Attributes
Spatial region
Color...size)
Tilt/angle
Area (2D size)
Depth (3D position)
Color luminance
Color saturation
Curvature
Volume (3D size)

Further reading

Relative vs. absolute judgements
perceptual system mostly operates with relative judgements, not absolute
– that's why accuracy increases with common frame/scale and alignment
– Weber's Law: ratio of increment to background is constant
– filled rectangles differ in length by 1/9, difficult judgement
– white rectangles differ in length by 1/2, easy judgement

Further reading

Outline
Session 1: 8:30-10:30am
Visualization Analysis Framework
– Introduction: Definitions
• Analysis: What, Why, How
– Marks and Channels
Session 2: 10:30am-12:00pm
Spatial Layout
– Arrange Tables
– Arrange Spatial Dies
– Arrange Networks and Trees

Further reading

Further reading

Further reading

Further reading

Further reading

Further reading
No text content could be extracted from the image.
Connection vs. adjacency comparison

- Adjacency matrix strengths
  - predictability: scalability supports recording
  - some topology tasks trainable
- Node-link diagram strengths
  - topology understanding path tracing
  - instructions, no reasoning needed

Empirical study

- Node-link best for small networks
- Matrix best for large networks

If tasks don’t involve topological structure!

[For readability of graphs using node-link and matrix-based representations a combined approach and cantilever solutions!]

Further reading

- Ock-Hands Networks and Trees

- Tree drawing idioms comparison

- Data shown
  - link relationships
  - tree depth
  - drawing order
- Design choices
  - design vs. containment link marks
  - reducer vs. radial layout
  - spatial position channels
- Considerations
  - redundant? arbitrary?
  - avoid wasting space

Categorical vs. ordered colors

- 3 channels
  - Identity for categorical
  - hue
  - magnitude for ordered
  - luminance
  - saturation
- RGB poor for encoding
- HSV better but beware
- lightness ≈ luminance

Idiom: force-directed placement

- Visual encoding
  - link connection marks, node point marks
- Considerations
  - Spatial position: no meaning directly encoded?
  - link to external reference?
  - proximity semantics?
  - numerous meaningful
- Node-link arbitrary artifact of layout algorithm
  - sensor width
  - Long edges more visible than short
- Tasks
  - Explore topology; follow paths, clusters
  - Scarcity
- Node-link density E < N

Idiom: radial node-link tree

- Data
- Visual encoding
  - link connection marks
  - point node marks
- Considerations
  - Instructions, no reasoning
  - Angular proximity siblings
  - Distance from center: depth in tree
- Tasks
  - Understanding topology, following paths
  - Scarcity
  - IK → N nodes

Idiom: tree map

- Data
- 1 tree
- 1 quant attrib at leaf nodes
- encoding
  - Area: contains a hierarchical structure
  - Resilient orientation
  - size encodes quant attrib
- Tasks
  - Notify attribute at leaf nodes
  - Scarcity
  - 1M leaf nodes

Idiom: sfdp (multi-level force-directed placement)

- Data
- Original network
- Derived cluster hierarchy dominates

- Considerations
  - Better algorithm for some encoding technique
  - Some additional use of space
- Node hierarchies used for spatial/palatability thus not shown explicitly
- More an algorithm vs. encoding in algorithms
  - Only colors show presence/absence of edge
  - IK nodes, IM edges

Idiom: adjacency matrix view

- Data
  - Network
  - Transform into same data/encoding as heatmap
- Derived data: table from network

- 1. Quant strata
  - Weighted edge between nodes
- 2: Categorize edge node list x 2
  - Visual encoding
  - Cell shows presence/absence of edge
- Scarcity
  - IK nodes, IM edges

Link marks: Connection and containment

- Marks as links (vs. nodes)
  - Common case in network drawing
  - 1D case: connection
  - ECC all node-link diagram
  - Emphasizes topology path tracing
  - Node and tree
  - 2D case: containment
  - ECC all spanning trees
  - Emphasizes containment always leaves (size encoding)
  - Only trees

Outline

- Session 1 8:30–10:00am Presentation: Analysis Framework
  - Introduction: Definitions
  - Analysis: What, Why How Marks and Channels

- Session 2 10:30am–12:00pm Spatial Layout
  - Across Tables
  - Across Spatial Data
  - Across Networks and Trees

Idiom design choices: Encode

- Map from categorical and colored attributes
  - Color
  - Hue
- Angle, Curvature...

Spectral sensitivity

- Color perception before optic nerve
  - One achromatic luminance channel L
  - Edge detection through luminous contrast
- Two chromatic channels R-G and Y-B
- "Color blind" if one axis has degraded acuity
- 8% of men are red/green color deficient
- Blue/yellow is rare

Opponent color and color deficiency

Idiom: Visible Spectrum

- Luminance
  - Saturation
- Hue
- Source of the RGB color
  - L from HLS
  - All the same
- Luminance values

Connection
- Node-Link Diagrams
- Adjacency Matrix
- Illuminance

Tree drawing idioms comparison

- Data shown
  - Link relationships
  - Tree depth
  - Drawing order
- Design choices
  - Connection vs containment link marks
  - Reducer vs radial layout
  - Spatial position channels
- Considerations
  - Redundant? Arbitrary
  - Avoid wasting space
**Bivariate**

- **Diverging**
- **Categorical**
- **Sequential**
- **Categorical**

**ColorBrewer**
- [http://www.colorbrewer2.org](http://www.colorbrewer2.org)
- saturation and area example: size affects salience!

1. **Bezold Effect**: Outlines matter
   - **color constancy**: simultaneous contrast effect

2. **Color/Lightness constancy**: Illumination conditions
   - **Image courtesy of John McCann**

3. **Colormaps**
   - Categorical
   - Ordered
   - Sequential
   - Diverging

4. **Color/Lightness constancy**: Illumination conditions
   - **Image courtesy of John McCann**

5. **Categorical color**: Discriminability constraints
   - noncontiguous small regions of color: only 6-12 bins

6. **Ordered color**: Rainbow is poor default
   - problems
   - benefits
   - alternatives

7. **Ordered color**: Rainbow is poor default
   - problems
   - benefits
   - alternatives

8. **Ordered color**: Rainbow is poor default
   - problems
   - benefits
   - alternatives

9. **Ordered color**: Rainbow is poor default
   - problems
   - benefits
   - alternatives

**Designing for color deficiency**
- Check with simulator
- Avoid encoding by hue alone
- Redundantly encode
  - vary luminance
  - change shape

10. **Color deficiency**: Reduces color to 2 dimensions
   - **Image courtesy of John McCann**

11. **Color/Lightness constancy**: Illumination conditions
    - Image courtesy of John McCann

12. **Colormaps**
    - Categorical
    - Ordered
    - Sequential
    - Diverging

13. **Colormaps**
    - Categorical
    - Ordered
    - Sequential
    - Diverging

14. **Colormaps**
    - Categorical
    - Ordered
    - Sequential
    - Diverging

15. **Colormaps**
    - Categorical
    - Ordered
    - Sequential
    - Diverging

16. **Colormaps**
    - Categorical
    - Ordered
    - Sequential
    - Diverging

17. **Colormaps**
    - Categorical
    - Ordered
    - Sequential
    - Diverging

18. **Colormaps**
    - Categorical
    - Ordered
    - Sequential
    - Diverging

19. **Colormaps**
    - Categorical
    - Ordered
    - Sequential
    - Diverging

20. **Colormaps**
    - Categorical
    - Ordered
    - Sequential
    - Diverging

21. **Colormaps**
    - Categorical
    - Ordered
    - Sequential
    - Diverging

22. **Colormaps**
    - Categorical
    - Ordered
    - Sequential
    - Diverging

23. **Colormaps**
    - Categorical
    - Ordered
    - Sequential
    - Diverging

24. **Colormaps**
    - Categorical
    - Ordered
    - Sequential
    - Diverging
Maneulc at e  Fac et  Re du ce
C hange
S ele c t
N a v ig at e
Ju x tapose
P a r tition
Superimpose
F il t er
A g g r eg at e
Embed

Deriv e
At t ribute  Red uction
Slice
Cut
Project

Chang e  o v e r  T ime
S ele c t

Further reading
• Visualization Analysis and Design. Munzner. AK Peters Visualization Series, CRC Press, 2014
  – Chap 10: Map Color and Other Channels
  • Visual Thinking for Design. Ware, Morgan Kaufmann, 2008.
  – Enc o de
A r r ange
Exp r ess S epa r a t e
O r der A li g n
U se
M anipul at e F a c et R edu c e
C hange
S ele c t
N a v ig at e
Ju x tapose
P a r tition
Superimpose
F il t er
A g g r eg at e
Embed

How to handle complexity: 1 previous strategy + 3 more
• derive new data to show within view
• change view over time
• facet across multiple views
– reduce items/attributes within single view

Manipulate

1) Change over time
   • change any of the other choices
     – encoding itself
     – parameters
     – arrange: rearrange, reorder
     – aggregation level, what is filtered...
     – interaction entails change

2) Select
   • data: tables with many attributes
   • task: compare rankings
   System: LineUp
   

Idiom: Re-encode

1) Change over time
   • change any of the other choices
     – encoding itself
     – parameters
     – arrange: rearrange, reorder
     – aggregation level, what is filtered...
     – interaction entails change

2) Select
   • data: tables with many attributes
   • task: compare rankings
   System: Tableau

Idiom: Animated transitions

1) Change over time
   • change any of the other choices
     – encoding itself
     – parameters
     – arrange: rearrange, reorder
     – aggregation level, what is filtered...
     – interaction entails change

2) Select
   • data: tables with many attributes
   • task: compare rankings
   System: Tableau

Idiom: Realign

1) Change over time
   • change any of the other choices
     – encoding itself
     – parameters
     – arrange: rearrange, reorder
     – aggregation level, what is filtered...
     – interaction entails change

2) Select
   • data: tables with many attributes
   • task: compare rankings
   System: LineUp

Idiom: Reorder

1) Change over time
   • change any of the other choices
     – encoding itself
     – parameters
     – arrange: rearrange, reorder
     – aggregation level, what is filtered...
     – interaction entails change

2) Select
   • data: tables with many attributes
   • task: compare rankings
   System: Tableau

Navigate: Changing item visibility

1) Change over time
   • change viewpoint
     – changes which items are visible within view
     – camera metaphor
       • zoom
       • pan/translate
   – articulated
   – constrained navigation
     • often with customized transitions
     • often based on selection set

Idiom: Exposed transitions

1) Change over time
   • change any of the other choices
     – encoding itself
     – parameters
     – arrange: rearrange, reorder
     – aggregation level, what is filtered...
     – interaction entails change

2) Select
   • data: tables with many attributes
   • task: compare rankings
   System: Tableau

Attribute Reduction
Slice
Cut
Project

Map other channels

1) Size, Angle, Curvature...
2) Map Color
   • value
   – primary vs secondary: semantics (eg source/target)
   – line vs column: visual emphasis
   – click vs hover: heavyweight, lightweight

2) Angle
   – secondary: semantics (eg source/target)
   – angular value

2) Curvature
   – primary: semantic

Further reading
• Orderd color: Rainbow is poor default
  – problems
    • perceptually unordered
    • non-monotonically increasing luminance
  – benefits
    • fine-grained structure visible and readable
  – alternatives
    • large-scale structure fewer hues
    • fine structure multiple hues with monotonically increasing luminance
  • colorful, perceptually uniform, color blind safe, monotonically increasing luminance

Map other channels

1) Size, Angle, Curvature...
• size
   – length accurate, 2D area ok, 3D volume poor
• angle
   – non-linear accuracy
• curvature
   – highly separable against static
   – biary gray for highlighting
   – use with care to avoid irritation

2) Map Color
   – value
   – primary vs secondary: semantics (eg source/target)
   – line vs column: visual emphasis
   – click vs hover: heavyweight, lightweight

2) Angle
   – secondary: semantics (eg source/target)
   – angular value

2) Curvature
   – primary: semantic

Vividis

• colorful, perceptually uniform, colorblind-safe, monotonically increasing luminance

Additional Reading:
http://www.cs.ubc.ca/~tmm/talks.html#vad16act
Partition into views

- **how to divide data between views**
  - splits into regions by attributes
  - encodes association between items through proximity
  - order of splits has major implications for what patterns are visible

  - no strict dividing line
    - view: high-level
      - contiguous region in which visually encoded data is shown on the display
    - glyph: small/large
      - adapt to visual structure that arises from multiple marks

**Partition into Side-by-Side Views**

- **how to coordinate views**
  - juxtapose and coordinate views

**Coordinate views: Design choice interaction**

**Partitioning: List alignment**

- **small bar chart with grouped bars**
  - split by state into regions
  - compare glyphs within each region showing age
  - compare easy within state, hard across age

**Partitioning: Recursive subdivision**

- **split by neighborhood**
  - type of view
  - then state
    - years as rows
    - months as columns
  - color by price
  - neighborhood patterns
    - where it’s expensive
    - where you pay much more

**System: HIVE**

Outlook

- **Session 1: 8.30am-10.00am**
  - Visualisation Analysis Framework
  - Introduction: Definitions
  - Analysis: What, Why, How
  - Marks and Choreography

- **Session 2: 10.30am-12.00pm**
  - Spatial Layout
  - Arrange Tables
  - Arrange Spatio Data
  - Arrange Networks and Trees

- **Session 3: 1.00pm-3.00pm**
  - Color & Interaction
  - Reduce: Alter Graphics
  - Rules of Thumb
  - Q&A
### Eyes beat memory

- **principle**: external cognition vs. internal memory
  - easy to compare by moving eyes between side-by-side views
  - harder to compare visible item to memory of what you saw
- **implications for animation**
  - great for choreographed storytelling
  - great for transitions between two states
  - poor for many states with changes everywhere
- **small multiples** instead
  - literal: 
    - animation
    - abstract:
  - show time with time
  - show time with space

### Why not animation?

- **disparate frames and regions**: comparison difficult
  - vs contiguous frames
  - vs small region
  - vs coherent motion of group
- **safe special case**
  - animated transitions

### Resolution beats immersion

- **immersion typically not helpful for abstract data**
- **resolution much more important**
  - pixels are the scarcest resource
  - desktop also better for workflow integration
- **virtual reality for abstract data very difficult to justify**
- **overview = summary**
- **microcosm of full vs design problem**

### Further reading

- **Visualization and Design**
  - Tamara Munzner: CRC Press, 2014
  - Chap 1: Rule of Thumb
  - Chap 2: Have Time Requirements

- **Additional resources**
  - [this talk](http://www.cs.ubc.ca/group/infovis)
  - [book page](http://www.cs.ubc.ca/group/infovis)
  - [20% promo code for book+ebook combo](http://www.cs.ubc.ca/group/infovis) 
  - [book page (including tutorial lecture slides)](http://www.cs.ubc.ca/group/infovis)
  - [PPTX slides](http://www.cs.ubc.ca/group/infovis)

---

<table>
<thead>
<tr>
<th>No unjustified 2D</th>
<th>Eyes beat memory</th>
<th>Why not animation?</th>
<th>Resolution beats immersion</th>
<th>Further reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>• 3D legitimate for true 3D spatial data</td>
<td>• principle: external cognition vs. internal memory</td>
<td>• disparate frames and regions: comparison difficult</td>
<td>• overview = summary</td>
<td><strong>No unjustified 3D</strong></td>
</tr>
<tr>
<td>• 3D needs very careful justification for abstract data</td>
<td>• 0.1 seconds: perceptual processing</td>
<td>• vs contiguous frames</td>
<td></td>
<td><strong>Eyes beat memory</strong></td>
</tr>
<tr>
<td>– enthusiasm in 1990s, but now skepticism</td>
<td>• immediately: immediate response</td>
<td>– vs small region</td>
<td></td>
<td><strong>Why not animation?</strong></td>
</tr>
<tr>
<td>– be especially careful with 3D for point clouds or networks</td>
<td>• 10 seconds: brief tasks</td>
<td>– vs coherent motion of group</td>
<td></td>
<td><strong>Resolution beats immersion</strong></td>
</tr>
<tr>
<td>• consider whether network data requires 2D spatial layout</td>
<td></td>
<td></td>
<td></td>
<td><strong>Further reading</strong></td>
</tr>
<tr>
<td>– especially if rendering is central to task.</td>
<td></td>
<td></td>
<td></td>
<td><strong>No unjustified 3D</strong></td>
</tr>
<tr>
<td>– arranging as network means lower information density and harder label lookup compared to text list</td>
<td></td>
<td></td>
<td></td>
<td><strong>Eyes beat memory</strong></td>
</tr>
<tr>
<td>• benefits outweigh costs when topological structure/context important for task</td>
<td></td>
<td></td>
<td></td>
<td><strong>Why not animation?</strong></td>
</tr>
<tr>
<td></td>
<td>• – literal:</td>
<td></td>
<td></td>
<td><strong>Resolution beats immersion</strong></td>
</tr>
<tr>
<td></td>
<td>– animation</td>
<td></td>
<td></td>
<td><strong>Further reading</strong></td>
</tr>
<tr>
<td></td>
<td>– abstract</td>
<td></td>
<td></td>
<td><strong>No unjustified 3D</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Eyes beat memory</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Why not animation?</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Resolution beats immersion</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Further reading</strong></td>
</tr>
</tbody>
</table>