CS 519: Scientific Visualization

What is Information Visualization?

Eric Shaffer

Some (many!) slides adapted from work by
Professor Tamara Munzner
University of British Columbia
A visualization is a visual representation of abstract data to aid human cognition

- Must be based on data
- The results must be readable, recognizable and useful
How many R’s are there?

GLNSAGGKLDSANGNASDGN
KLANSDDLKNASDGNDFVMD
GJERKJVERJVJKENJLVNEKVKVJEN
VJEAJVJNDJVNAAJBVRKMLVLJKD
How many R’s are there?

GLNSAGGKLDSANGNASDGN
KLANSDGLKNASDGNDFVMD
GJERKJVERRJVJKENJLVNEKJEN
VJEAJVJNDJVNAAJBVVRKLVLJKD
How Much Data Is There?

What Will We Do When The World's Data Hits 163 Zettabytes In 2025?

-Forbes.com

- A zettabyte is one trillion gigabytes
  - Zettabyte ~= 1,000,000,000,000,000,000,000,000
  - 200x all words ever spoken by humans

- Current annual data creation rate is 16.3ZB

- Most of that is not scientific data
  - It is not specifically associated with a physical domain
So...Maybe That Much Data is a Problem?

Visualization is a proven technique to aid human comprehension of complex data.

- It can aid in many analytical activities
  - Filtering, determining provenance, identifying critical events...etc.

Big Data: Experts say new forms of information analysis will help people be more nimble and adaptive, but worry over humans’ capacity to understand and use these new tools well.
Benefits of Visualization

- Adapt data to a form better processed by people
  - Maximize the use of limited perception and cognition
- Ultimate goals
  - Explore the data
    - Find patterns
  - Fit a function
  - Tell a story
    - Convince others
  - Spread information
Dataset Types

Tables
- Items
- Attributes

Networks & Trees
- Items (nodes)
- Links
- Attributes

Geometry
- Items
- Positions

Clusters, Sets, Lists
- Items
Attribute Types

- **Categorical**
  - no implicit ordering
  - Examples: +, ○, □, △, ○

- **Ordered**
  - **Ordinal**
    - Examples: T-shirts, stars
  - **Quantitative**
    - meaningful magnitude, can do arithmetic
    - Examples: ★★★★★, ★★★★☆, ■■■■■

- Categorical also referred to as “nominal”
- What operations are possible for each type?
Definitions: Marks and Channels

• marks
  – geometric primitives

• channels
  – control appearance of marks
- analyze idiom structure
  - as combination of marks and channels

1: vertical position
mark: line

2: vertical position
horizontal position
mark: point

3: vertical position
horizontal position
color hue
mark: point

4: vertical position
horizontal position
color hue
size (area)
mark: point
What is an Idiom?
**Channels: Rankings**

<table>
<thead>
<tr>
<th><strong>Magnitude Channels: Ordered Attributes</strong></th>
<th><strong>Identity Channels: Categorical Attributes</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Position on common scale</td>
<td>Spatial region</td>
</tr>
<tr>
<td>Position on unaligned scale</td>
<td>Color hue</td>
</tr>
<tr>
<td>Length (1D size)</td>
<td>Motion</td>
</tr>
<tr>
<td>Tilt/angle</td>
<td>Shape</td>
</tr>
<tr>
<td>Area (2D size)</td>
<td></td>
</tr>
<tr>
<td>Depth (3D position)</td>
<td></td>
</tr>
<tr>
<td>Color luminance</td>
<td></td>
</tr>
<tr>
<td>Color saturation</td>
<td></td>
</tr>
<tr>
<td>Curvature</td>
<td></td>
</tr>
<tr>
<td>Volume (3D size)</td>
<td></td>
</tr>
</tbody>
</table>

- **effectiveness principle**
  - encode most important attributes with highest ranked channels

- **expressiveness principle**
  - match channel and data characteristics
Accuracy: Visualization Experiments


after Michael McGuffin course slides, http://profs.etsmtl.ca/mmCGuffin/
Discriminability: How many usable steps?

- must be sufficient for number of attribute levels to show
  - linewidth: few bins
Separability vs. Integrality

Position
- Hue (Color)
  - Fully separable
    - 2 groups each

Size
- Hue (Color)
  - Some interference
    - 2 groups each

Width
- Height
  - Some/significant interference
    - 3 groups total: integral area

Red
- Green
  - Major interference
    - 4 groups total: integral hue
• find the red dot
  – how long does it take?
• parallel processing on many individual channels
  – speed independent of distractor count
  – speed depends on channel and amount of difference from distractors
• serial search for (almost all) combinations
  – speed depends on number of distractors
• many channels: tilt, size, shape, proximity, shadow, direction, ...
• but not all! parallel line pairs do not pop out from tilted pairs
Grouping

- containment
- connection

Marks as Links

- Containment
- Connection

Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape

- proximity
  - same spatial region
- similarity
  - same values as other categorical channels
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

---

Relative vs. Absolute Judgements

- perception of luminance is contextual based on contrast with surroundings

http://persci.mit.edu/gallery/checkershadow
Principles Based on Practice

• No unjustified 3D
  – Power of the plane, dangers of depth
  – Occlusion hides information
  – Perspective distortion loses information
  – Tilted text isn’t legible

• No unjustified 2D

• Eyes beat memory

• Resolution over immersion

• Overview first, zoom and filter, details on demand

• Function first, form next

• (Get it right in black and white)
• high-ranked spatial position channels: **planar** spatial position – not depth!

**Magnitude** Channels: **Ordered** Attributes

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)

![Graph](image)

Steven’s Psychophysical Power Law: \( S = I^N \)
Danger of Depth

- we don’t really live in 3D: we see in 2.05D
  - acquire more info on image plane quickly from eye movements
  - acquire more info for depth slower, from head/body motion

We can only see the outside shell of the world
Occlusion Hides Information

- occlusion
- interaction complexity

Perspective Projection Loses Information

- perspective distortion
  - interferes with all size channel encodings
  - power of the plane is lost!

[Visualizing the Results of Multimedia Web Search Engines. Mukherjea, Hirata, and Hara. InfoVis 96]
• text legibility
  – far worse when tilted from imageplane

• further reading


• benefits outweigh costs when task is shape perception for 3D spatial data
  – interactive navigation supports synthesis across many viewpoints

No Unjustified 3D

• 3D legitimate for true 3D spatial data
• 3D needs very careful justification for abstract data
  – enthusiasm in 1990s, but now skepticism
  – be especially careful with 3D for point clouds or networks

• consider whether network data requires 2D spatial layout
  – especially if reading text is central to task!
  – arranging as network means lower information density and harder label lookup compared to text lists
• benefits outweigh costs when topological structure/context important for task
  – be especially careful for search results, document collections, ontologies
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
    • consider small multiples instead

<table>
<thead>
<tr>
<th>literal</th>
<th>abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>animation</td>
<td>small multiples</td>
</tr>
<tr>
<td>show time with time</td>
<td>show time with space</td>
</tr>
</tbody>
</table>
• small multiples: one graph instance per experimental condition
  – same spatial layout
  – color differently, by condition

Why Not Animation

- disparate frames and regions: comparison difficult
  - vs contiguous frames
  - vs small region
  - vs coherent motion of group

- change blindness
  - even major changes difficult to notice if mental buffer wiped

- safe special case
  - animated transitions
Resolution Beats Immersion

• immersion typically not helpful for abstract data
  – do not need sense of presence or stereoscopic 3D
• 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 First, Zoom and Filter, Details on Demand

• influential mantra from Shneiderman


• overview = summary
  – microcosm of full vis design problem
  
• nuances
  – beyond just two levels: multi-scale structure
  – difficult when scale huge: give up on overview and browse local neighborhoods?

Function First, Form Next

• start with focus on functionality
  – straightforward to improve aesthetics later on, as refinement
  – if no expertise in-house, find good graphic designer to work with

• dangerous to start with aesthetics
  – usually impossible to add function retroactively
HemoViz Design Study

• formative study with experts
  – task taxonomy
• HemoViz design
• deploy attempt fails
  – experts balk: demand 3D and rainbows
• quantitative user study
  – med students, real data
  – 91% with 2D/diverging vs 39% with 3D/rainbows
  – experts willing to use

[Fig 1. Borkin et al. Artery Visualizations for Heart Disease Diagnosis. Proc InfoVis 2011.]
Study Results: Error
Study Results: Time

Bar charts showing time in seconds for 2D and 3D visualization for different color schemes:
- **All Colors**
- **Diverging**
- **Rainbow**

Left chart:
- 2D: All Colors - 56, Diverging - 65, Rainbow - 47
- 3D: All Colors - 73, Diverging - 76, Rainbow - 68

Right chart:
- 2D: All Colors - 2.5, Diverging - 2.4, Rainbow - 2.6
- 3D: All Colors - 7.8, Diverging - 5.6, Rainbow - 10.2

* indicates statistically significant differences.
Critique

- **many strengths**
  - careful and well justified design, convincing human-subjects experiment
    - bringing visualization best practices to medical domain
- **limitation**
  - paper does not clearly communicate why colormap is diverging not sequential
    - answer by email
    - doctors care about extremely high and extremely low ESS (scalar) values
      - high values (top of scale, dark grey): extreme blood flow patterns may relate to heart malfunctions - but not imminently life threatening and don't indicate plaque locations
      - low values (bottom of scale, dark red): very diseased regions with lots of plaque, docs care a lot!
      - much debate from doctors on where is boundary between “normal” and “low” ESS values
        » most think below 3 Pa are indicative of disease but many argue other values in the 2-4 range.
        » all docs agree that values below 2 Pa are increasingly dangerous disease levels.
        » thus map has transition at 3 Pa for the diverging point and truly red below 2 Pa
    - why continuous not segmented?
      - doctors gain tremendous insight by seeing the subtle patterning of the ESS values
      - particularly varying values in red region - patterns help them understand disease progression and severity
        » especially useful for deciding what types of interventions to prescribe for the patient
Is Your Visualization Effective?

Four levels of design problems
- different threats to validity at each level

- **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
Validation

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

**Data/task abstraction**

**Visual encoding/interaction idioms**
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

mismatch: cannot show idiom good with system timings
mismatch: cannot show abstraction good with lab study
In either case, visualization research benefits from a scientific approach

- Scientific? Maybe “systematic” or “engineering” would be more apt
- Create a taxonomy of visual elements
  - Understand which elements are appropriate for a given problem