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
KLANSDGLKNASDGNDFVMD
GJERKJVERJVJKENJLVNEKVJEN
VJEAJVJNDJVNAAJBVRKLVLJKD

How many R's are there?

GLNSAGGKLDSANGNASDGN
KLANSDGLKNASDGNDFVMD
GJERKJVERJVJKENJLVNEKVJEN
VJEAJVJNDJVNAAJBVRKLVLJKD

How Much Data Is There?

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

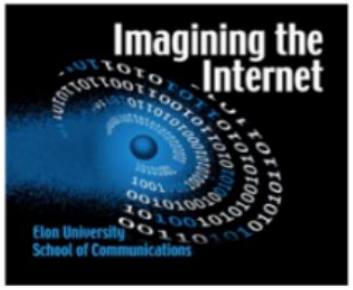
-Forbes.com

- A zettabyte is one trillion gigabytes
  - Zettabyte ~= 1, 000, 000, 000, 000, 000, 000, 000
  - 200x all words ever spoken by humans
- Current annual data creation rate is16.3ZB
- Most of that is not scientific data
  - It is not specifically associated with a physical domain

So…Maybe That Much Data is a Problem?

PewResearchCenter

Pew Internet
Pew Internet & American Life Project

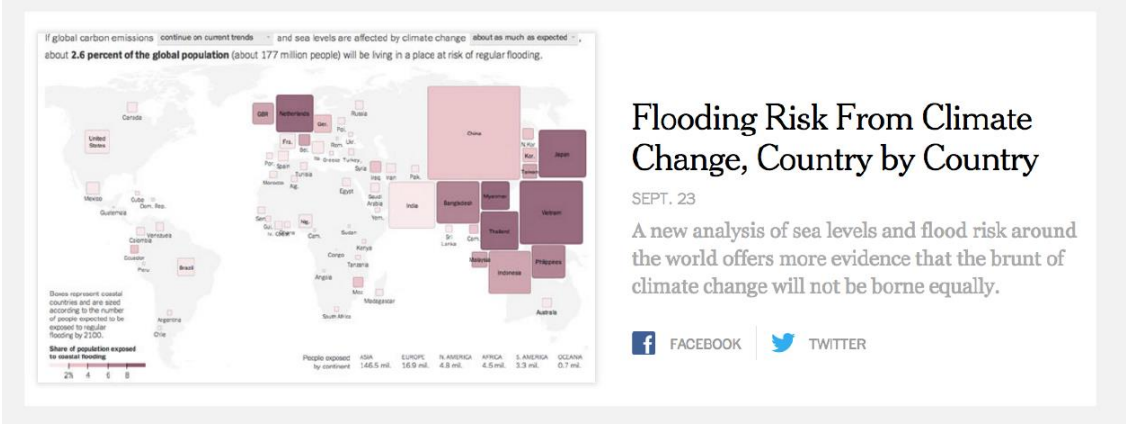


Big Data: Experts say new forms of information analysis will help people be more nimble and adaptive, but worry over humans' capacity to understand and use these new tools well

- 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.

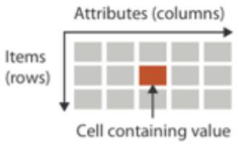
Benefits of Visualization

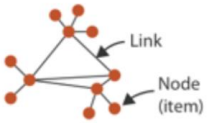
- Adapt data to a form better processed by people
  - Maximize the use of limited perception and cognition
- Ultimate goals
  - Explore the data
    - Find patterns
    - Fit a function
  - Tell a story
    - Convince others
    - Spread information



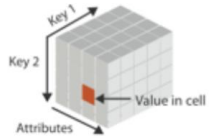
Dataset Types

Tables	Networks & Trees	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Items	Items
Attributes	Links	Positions	
	Attributes		





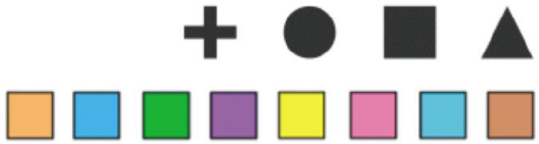
→ *Multidimensional Table*



→ *Trees*

Attribute Types

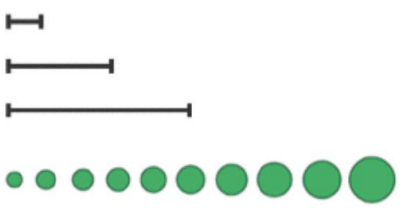
➔ Categorical
no implicit ordering



➔ Ordered

➔ *Ordinal*

➔ *Quantitative*
meaningful magnitude,
can do arithmetic

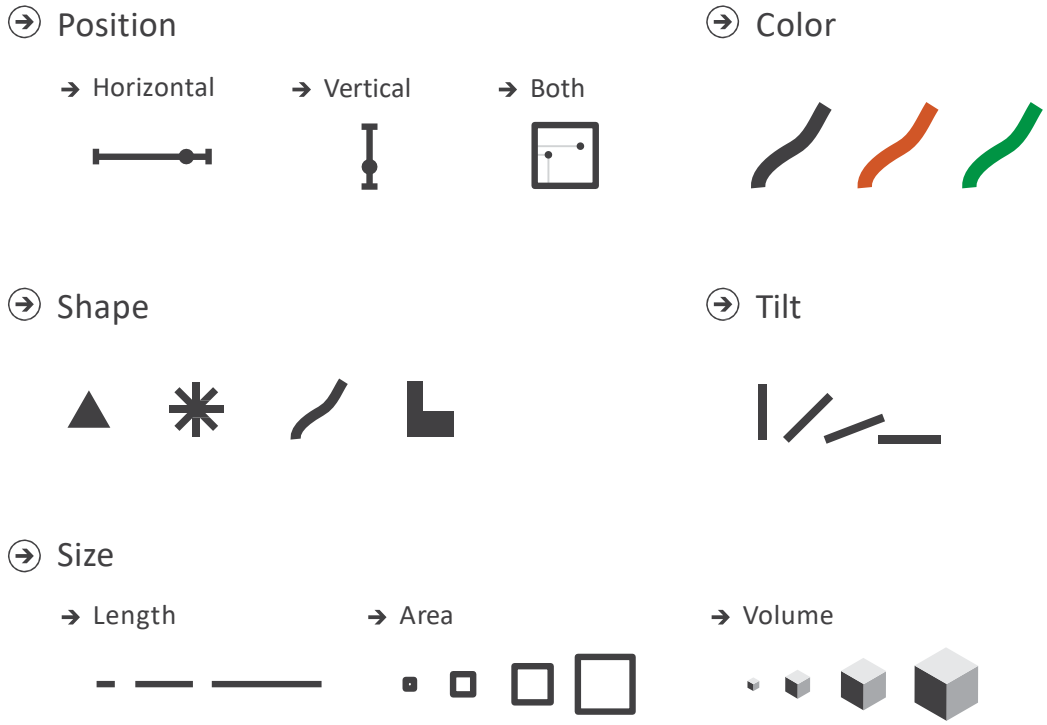


- Categorical also referred to as "nominal"
- What operations are possible for each type?

Definitions: Marks and Channels

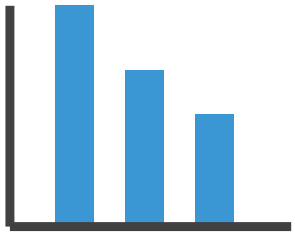
- marks
  - geometric primitives

- channels
  - control appearance of marks



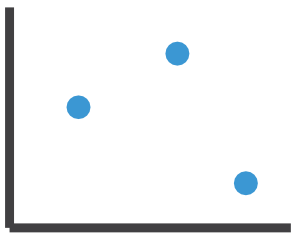
Encoding visually with marks and channels

- analyze idiom structure
  - as combination of marks and channels



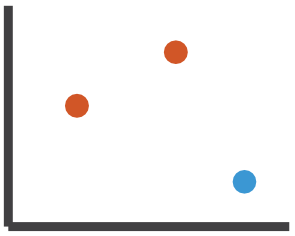
1:
vertical position

mark: line



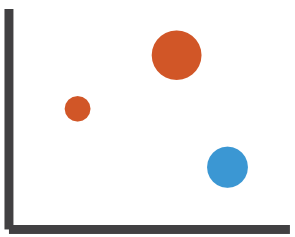
2:
vertical position
horizontal position

mark: point



3:
vertical position
horizontal position
color hue

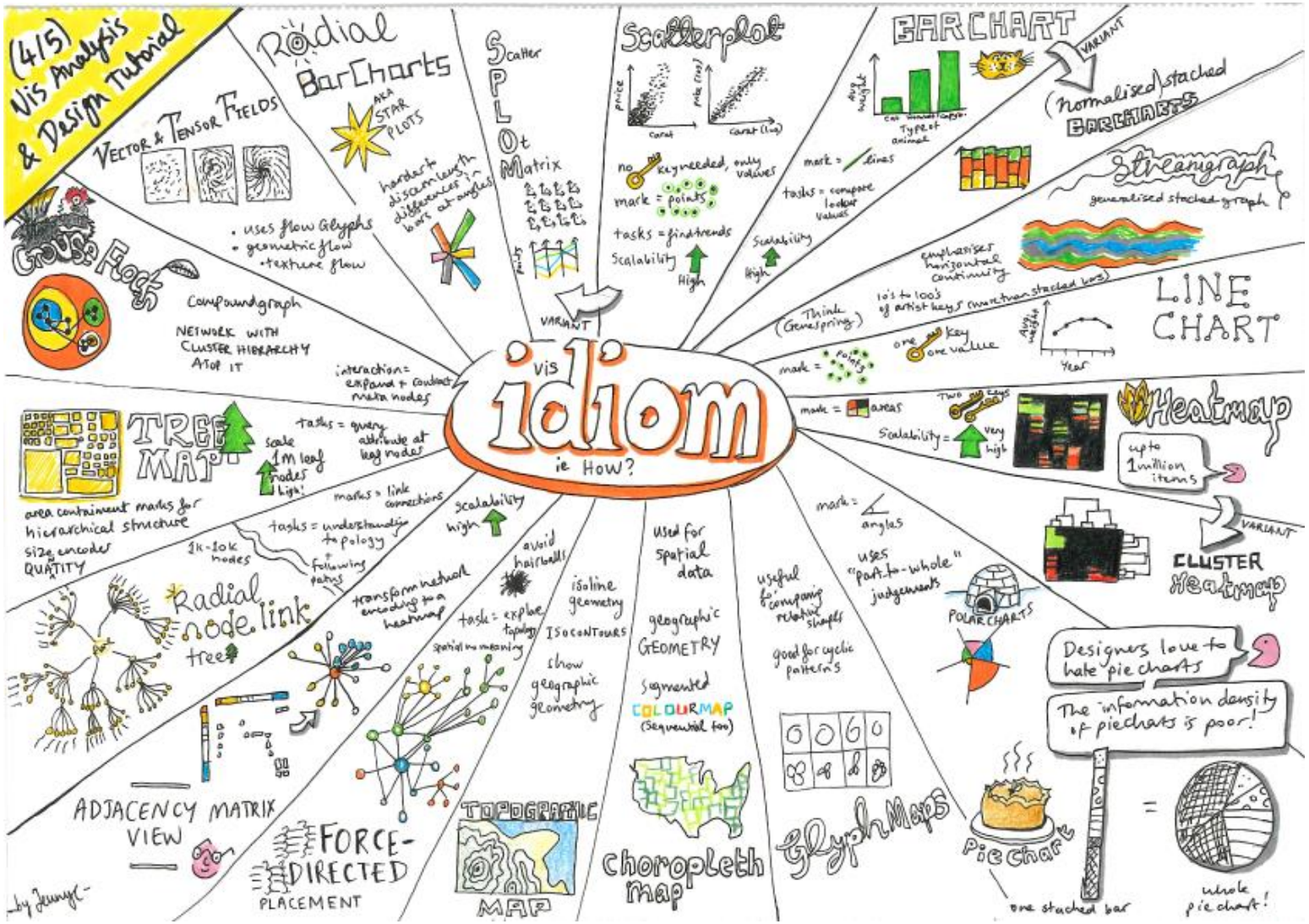
mark: point



4:
vertical position
horizontal position
color hue
size (area)

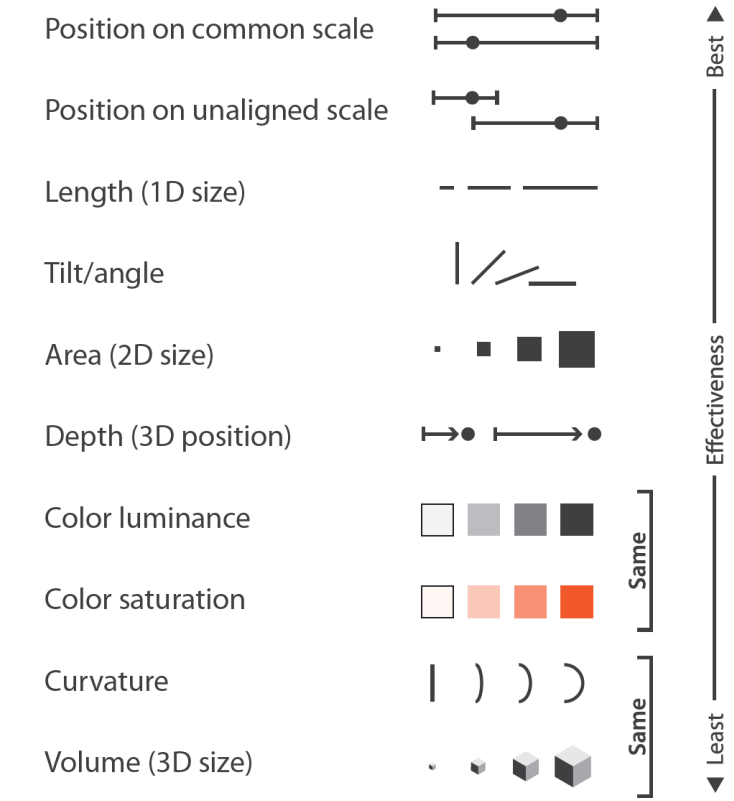
mark: point

What is an Idiom?

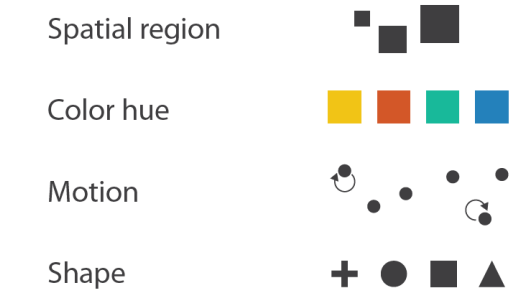


Channels: Rankings

Magnitude Channels: Ordered Attributes

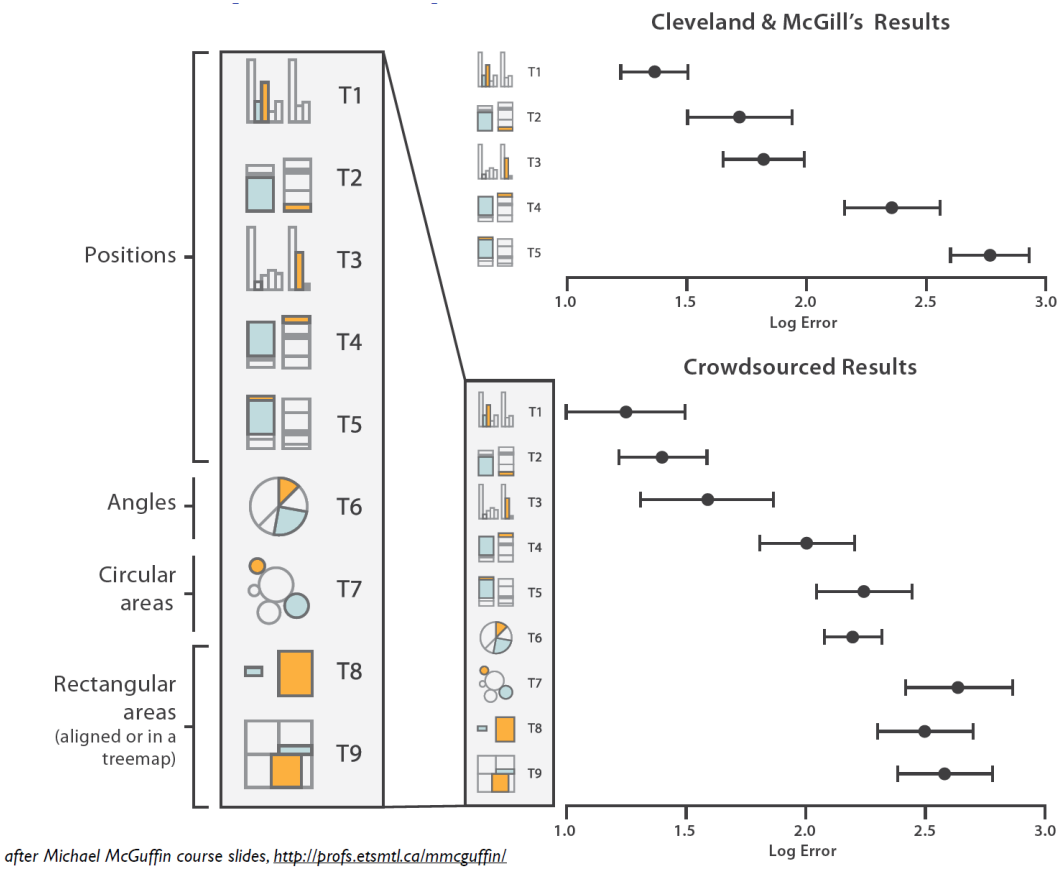


Identity Channels: Categorical Attributes



- effectiveness principle
  - encode most important attributes with highest ranked channels
- expressiveness principle
  - match channel and data characteristics

Accuracy: Visualization Experiments

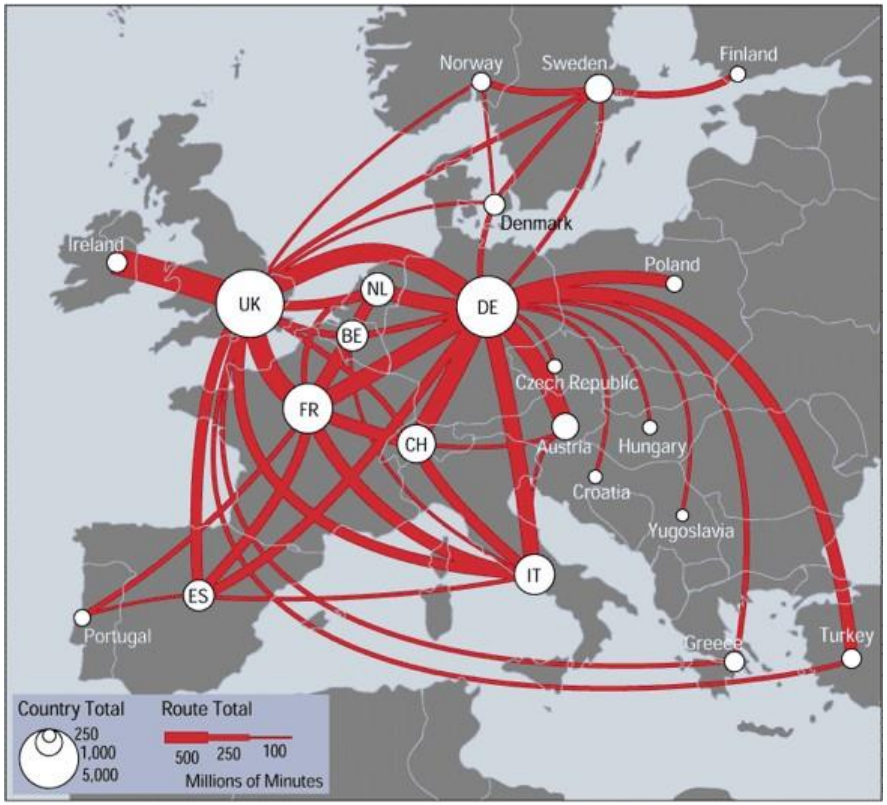


after Michael McGuffin course slides, http://profs.etsmtl.ca/mmcguffin/

[Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. Heer and Bostock. Proc ACM Conf. Human Factors in Computing Systems (CHI) 2010, p. 203–212.]

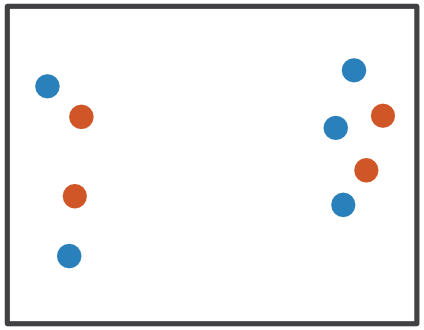
Discriminability: How many usable steps?

- must be sufficient for number of attribute levels to show
  - linewidth: few bins



Separability vs. Integrality

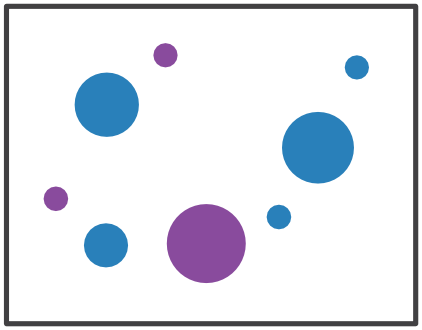
Position
✚ Hue (Color)



Fully separable

2 groups each

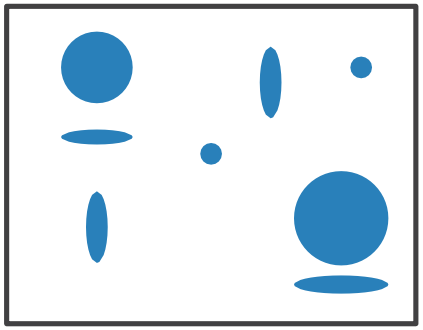
Size
✚ Hue (Color)



Some interference

2 groups each

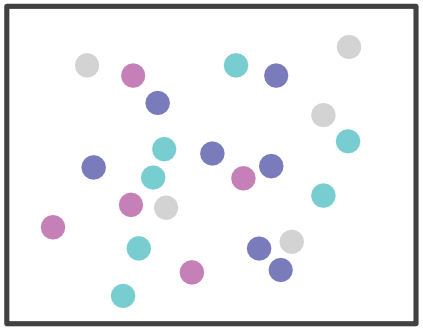
Width
✚ Height



Some/significant interference

3 groups total: integral area

Red
✚ Green

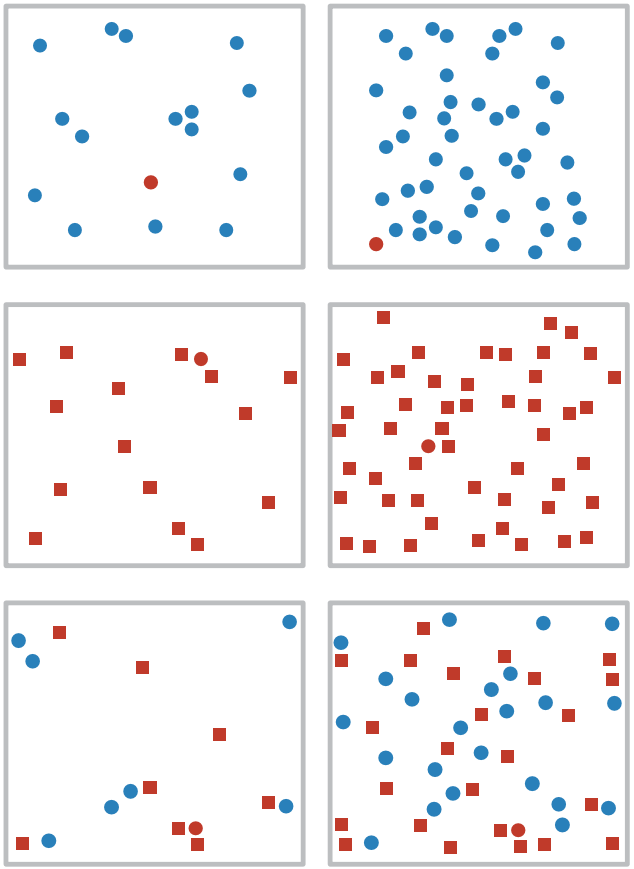


Major interference

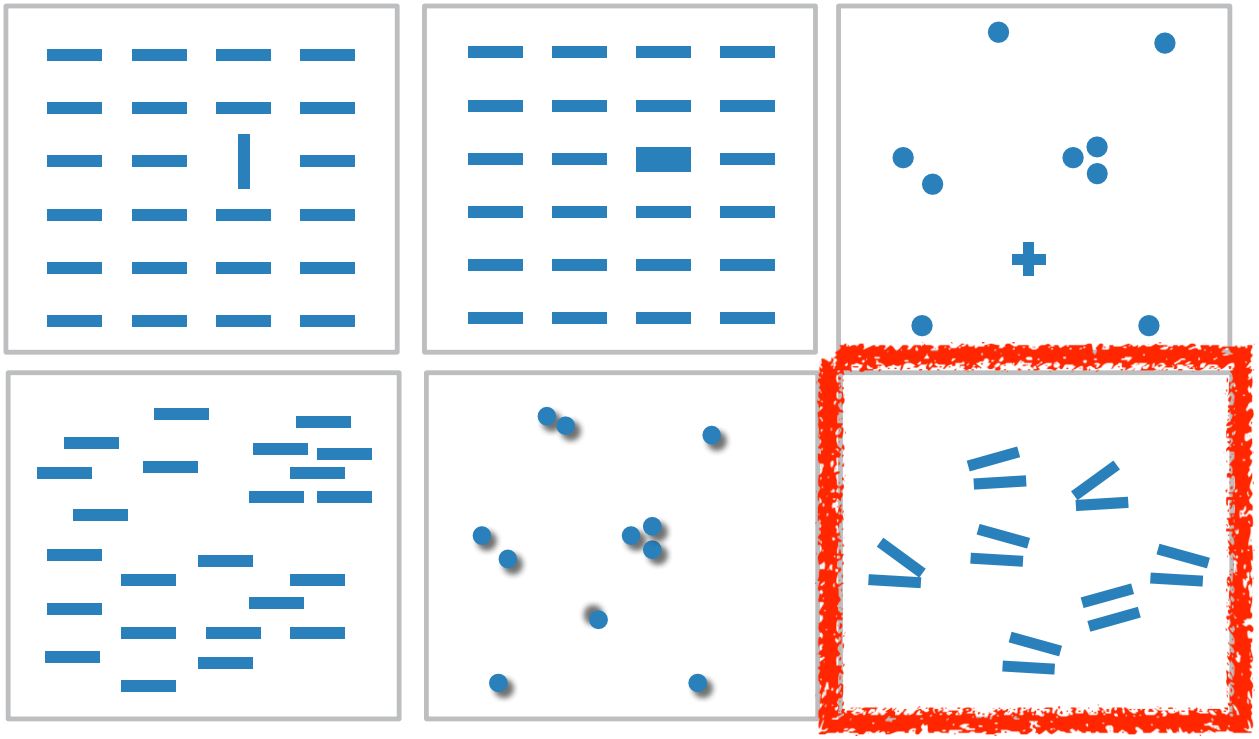
4 groups total: integral hue

Popout

- 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



Popout



- 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

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

Identity Channels: Categorical Attributes

Spatial region

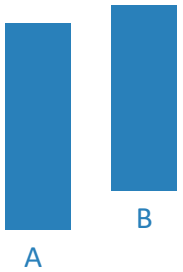
Color hue

Motion

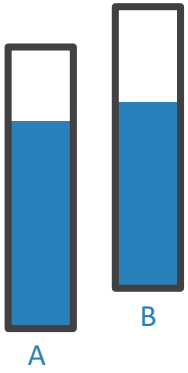
Shape

Relative vs. Absolute Judgements

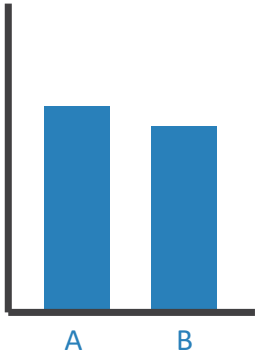
- perceptual system mostly operates with relative judgements, not absolute
  - that's why accuracy increases with common frame/scale and alignment
  - Weber's Law: ratio of increment to background is constant
    - filled rectangles differ in length by 1:9, difficult judgement
    - white rectangles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



length



position along unaligned common scale

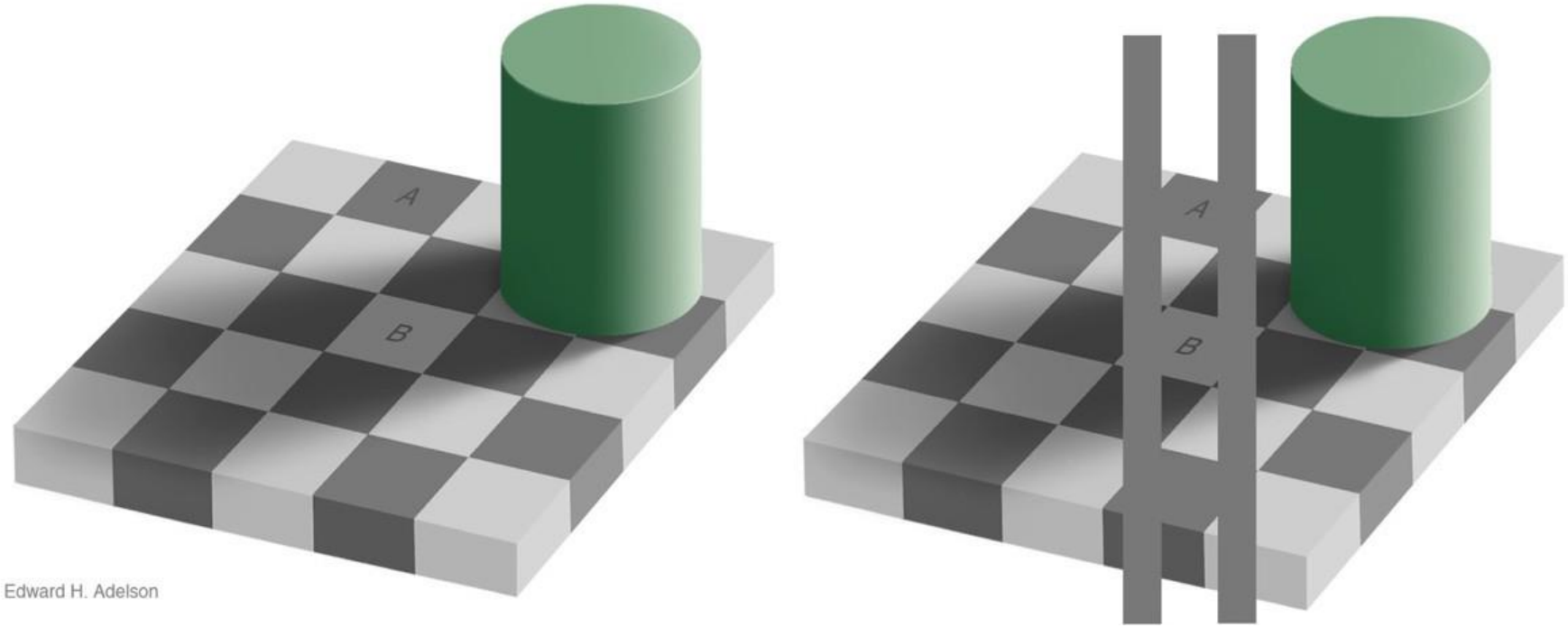


position along aligned scale

after [Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. Cleveland and McGill. Journ. American Statistical Association 79:387 (1984), 531–554.]

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)

No Unjustified 3D

- high-ranked spatial position channels: **planar** spatial position
  - not depth!

➔ **Magnitude Channels: Ordered Attributes**

Position on common scale

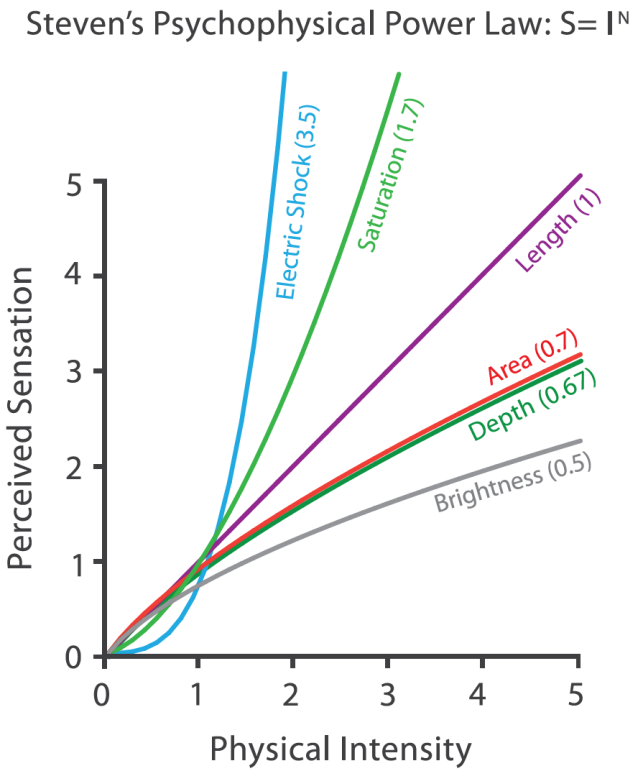
Position on unaligned scale

Length (1D size)

Tilt/angle

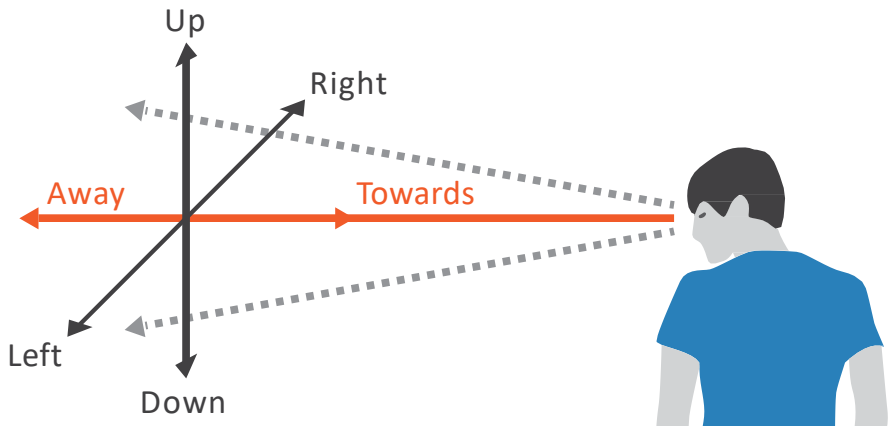
Area (2D size)

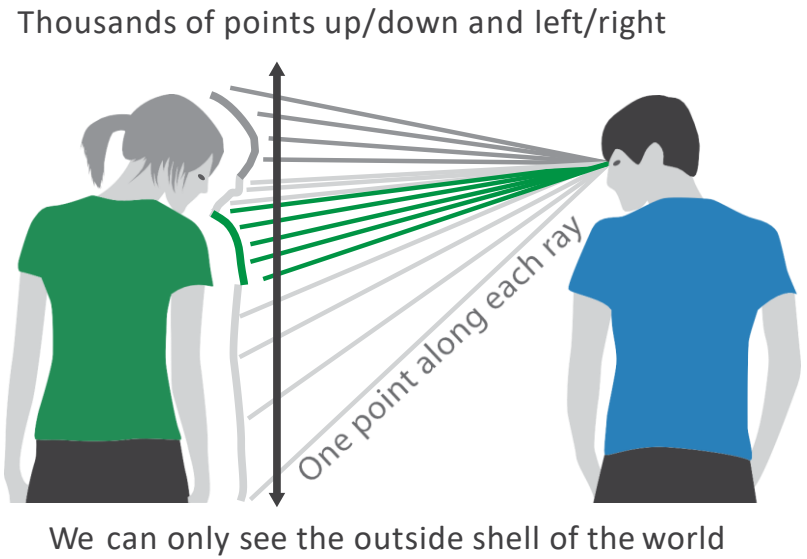
Depth (3D position)



Danger of Depth

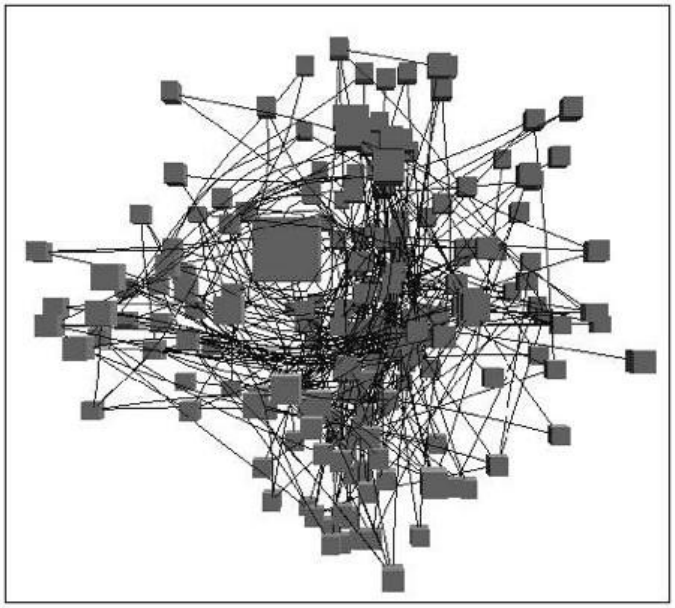
- we don't really live in 3D: we ***see*** in 2.05D
  - acquire more info on image plane quickly from eye movements
  - acquire more info for depth slower, from head/body motion





Occlusion Hides Information

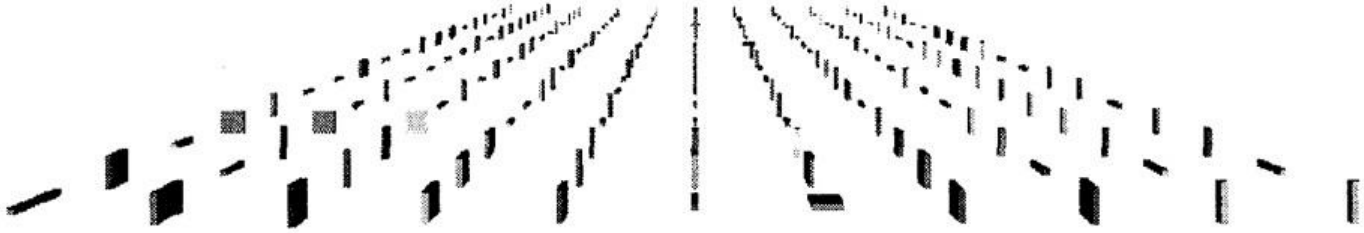
- occlusion
- interaction complexity



[DistortionViewingTechniques for 3D Data. Carpendale et al. InfoVis 1996.]

Perspective Projection Loses Information

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

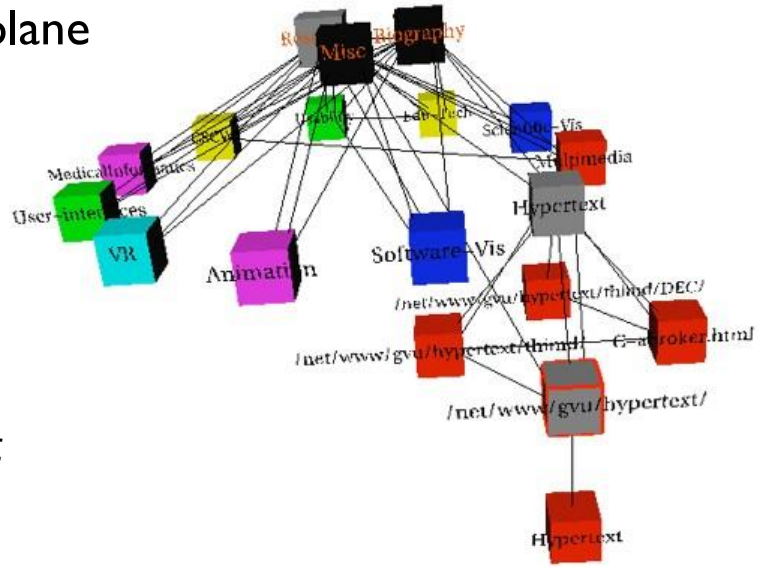


[Visualizing the Results of Multimedia Web Search Engines. Mukherjea, Hirata, and Hara. InfoVis 96]

Tilted Text Is Not Legible

- text legibility
  - far worse when tilted from image plane
- further reading

*[Exploring and Reducing the Effects of Orientation on Text Readability in Volumetric Displays.
Grossman et al. CHI 2007]*



[Visualizing the World-Wide Web with the Navigational View Builder. Mukherjea and Foley. Computer Networks and ISDN Systems, 1995.]

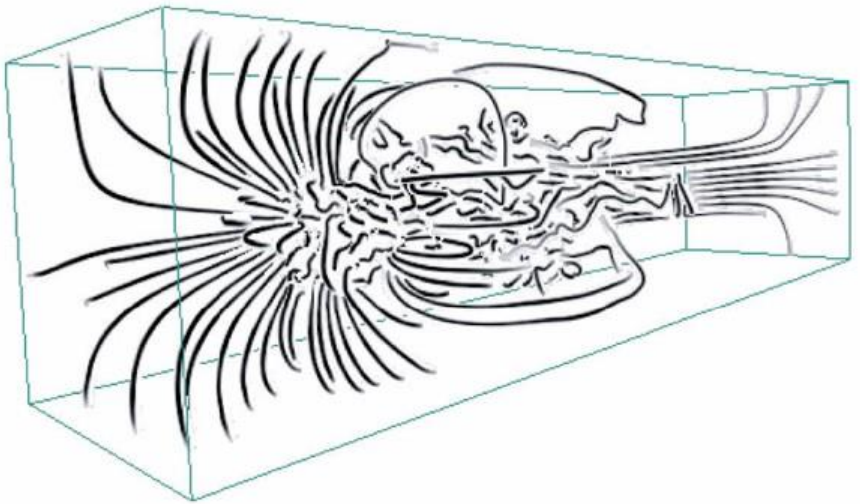
Justified 3D: Shape Perception

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

Targets

Spatial Data

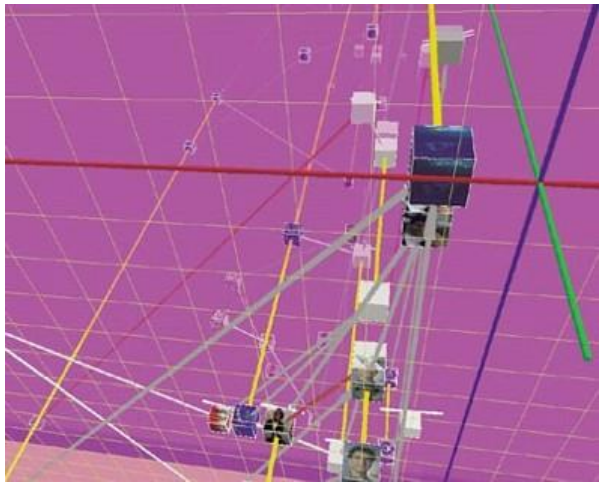
→ Shape



[Image-Based Streamline Generation and Rendering. Li and Shen. IEEE Trans. Visualization and Computer Graphics (TVCG) 13:3 (2007), 630–640.]

No Unjustified 3D

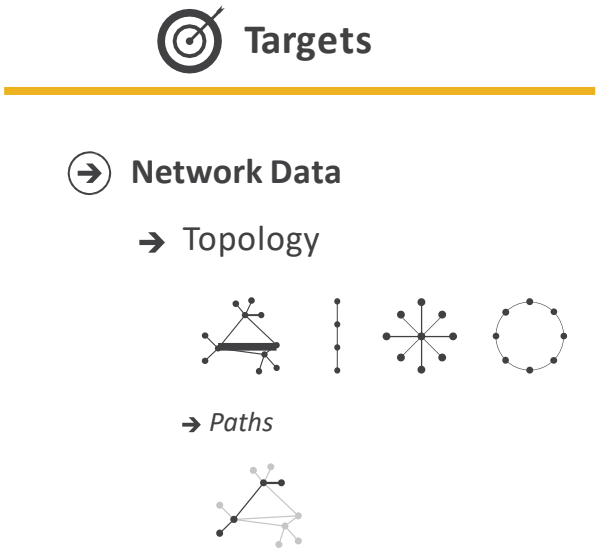
- 3D legitimate for true 3D spatial data
- 3D needs very careful justification for abstract data
  - enthusiasm in 1990s, but now skepticism
  - be especially careful with 3D for point clouds or networks



[WEBPATH-a three dimensionalWeb history. Frecon and Smith. Proc. InfoVis 1999]

No Unjustified 2D

- 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

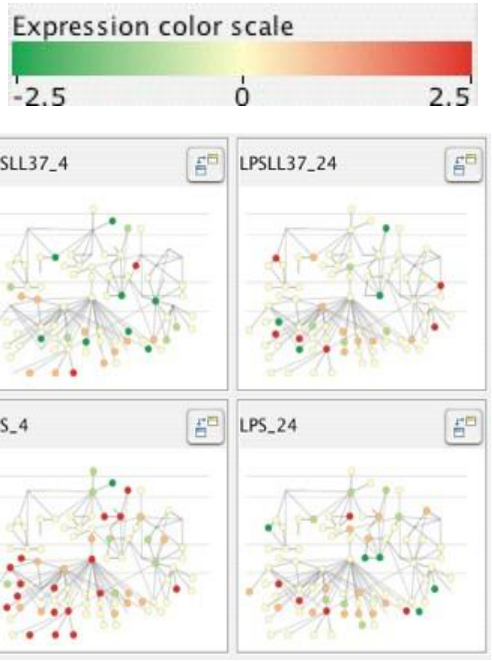


Eyes Beat Memory

- 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

Eyes Beat Memory Example

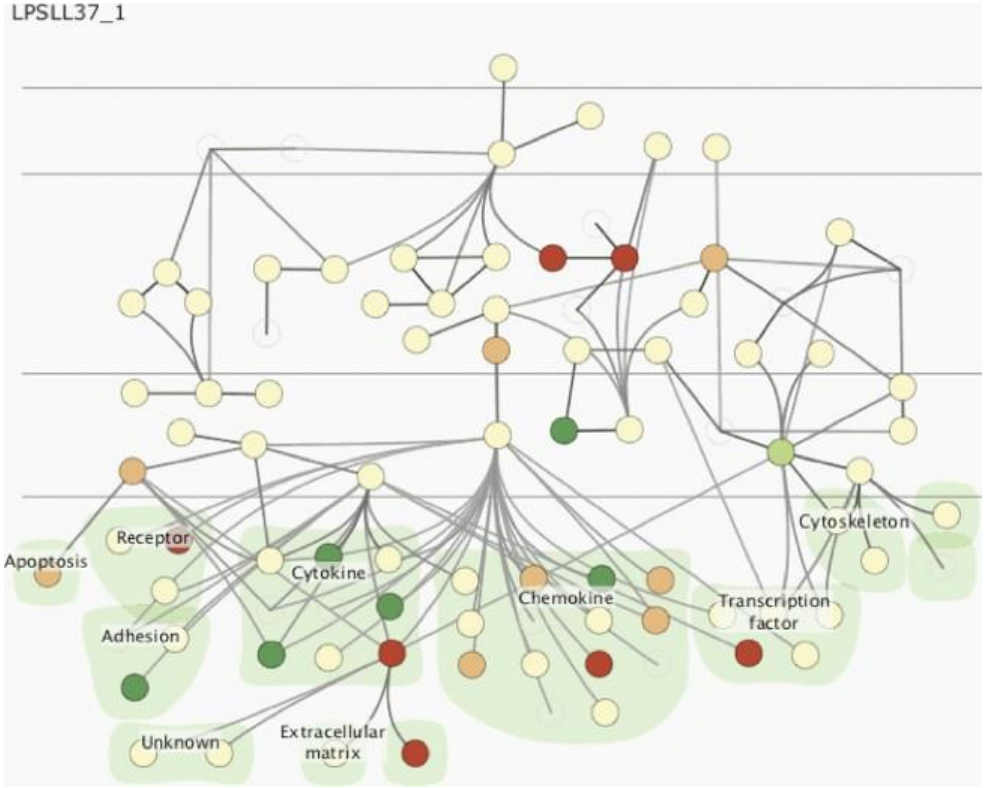
- small multiples: one graph instance per experimental condition
  - same spatial layout
  - color differently, by condition



[Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2008) 14:6 (2008), 1253–1260.]

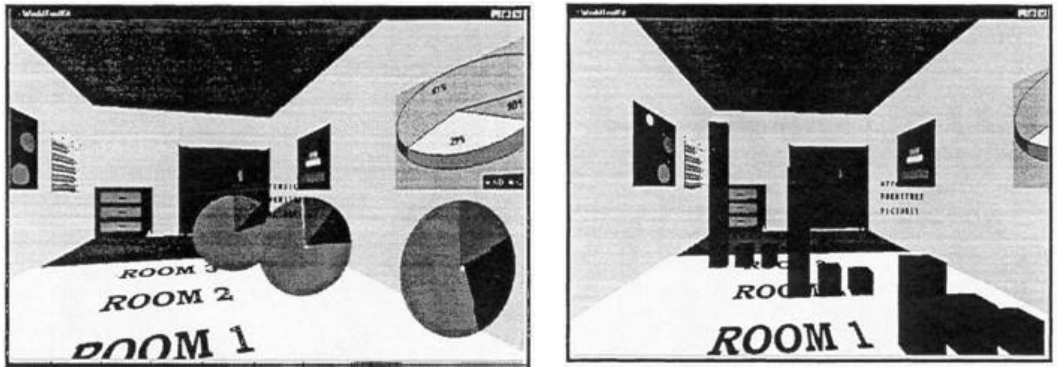
Why Not Animation

- 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



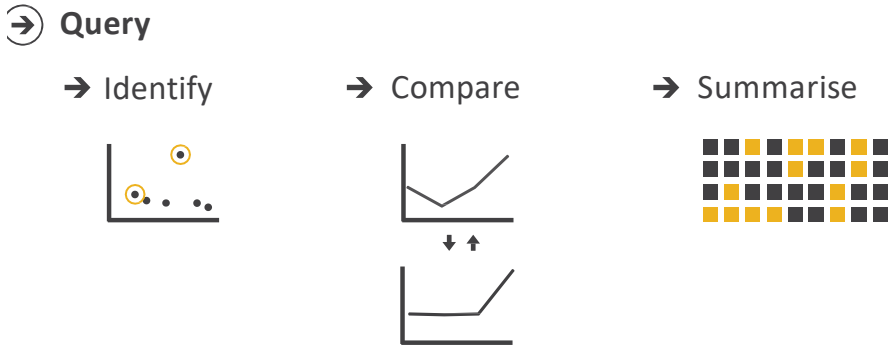
[Development of an information visualization tool using virtual reality. Kirner and Martins. Proc. Symp Applied Computing 2000]

Overview First, Zoom and Filter, Details on Demand

- influential mantra from Shneiderman

[The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Shneiderman. Proc. IEEE Visual Languages, pp. 336–343, 1996.]

- 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?

[Search, Show Context, Expand on Demand: Supporting Large Graph Exploration with Degree-of-Interest. van Ham and Perer. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2009) 15:6 (2009), 953–960.]

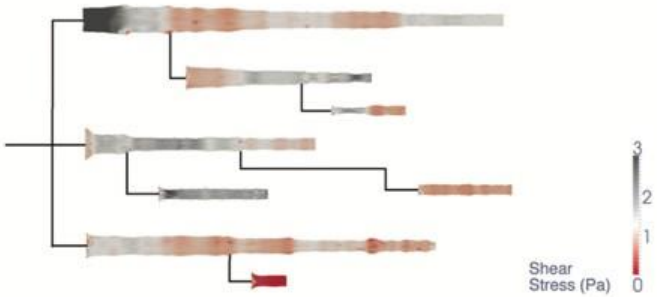
Function First, Form Next

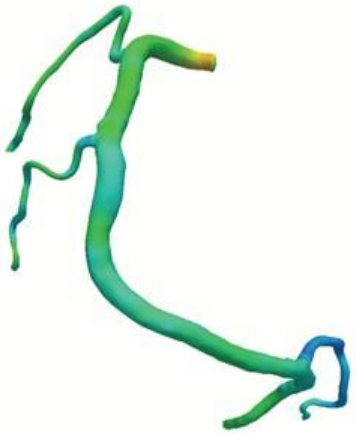
- start with focus on functionality
  - straightforward to improve aesthetics later on, as refinement
  - if no expertise in-house, find good graphic designer to work with

- dangerous to start with aesthetics
  - usually impossible to add function retroactively

HemoViz Design Study

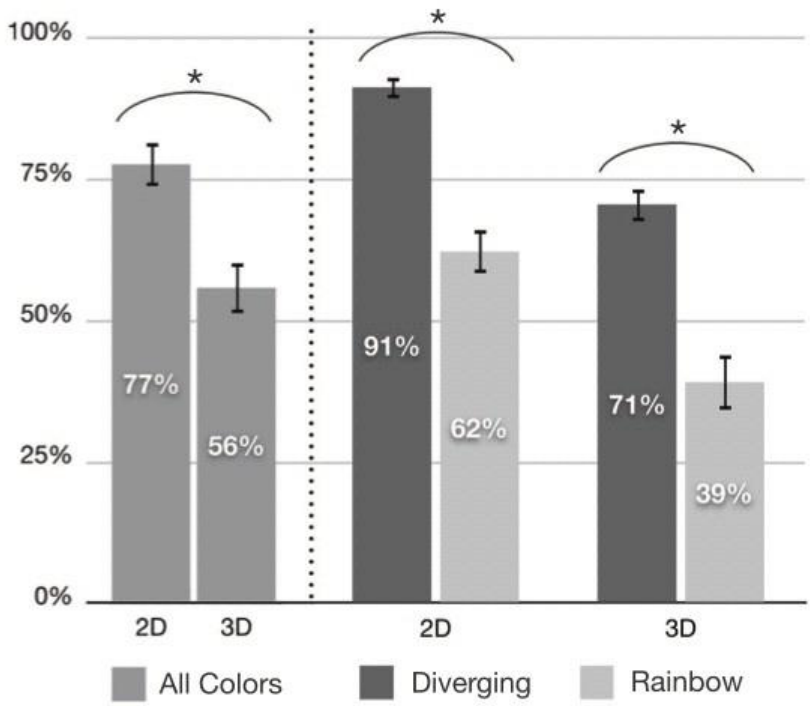
- formative study with experts
  - –task taxonomy
- HemoViz design
- deploy attempt fails
  - –experts balk: demand 3D and rainbows
- quantitative user study
  - –med students, real data
  - – 91% with 2D/divergingvs 39% with 3D/rainbows
  - –experts willing to use

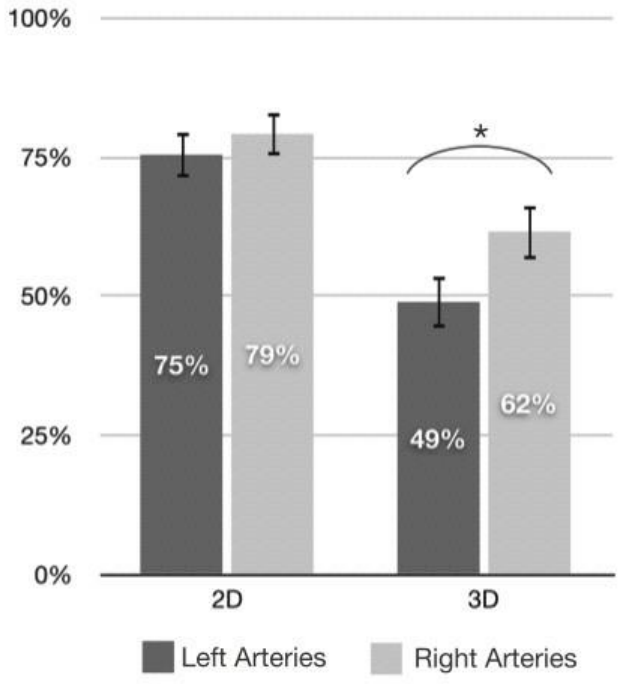




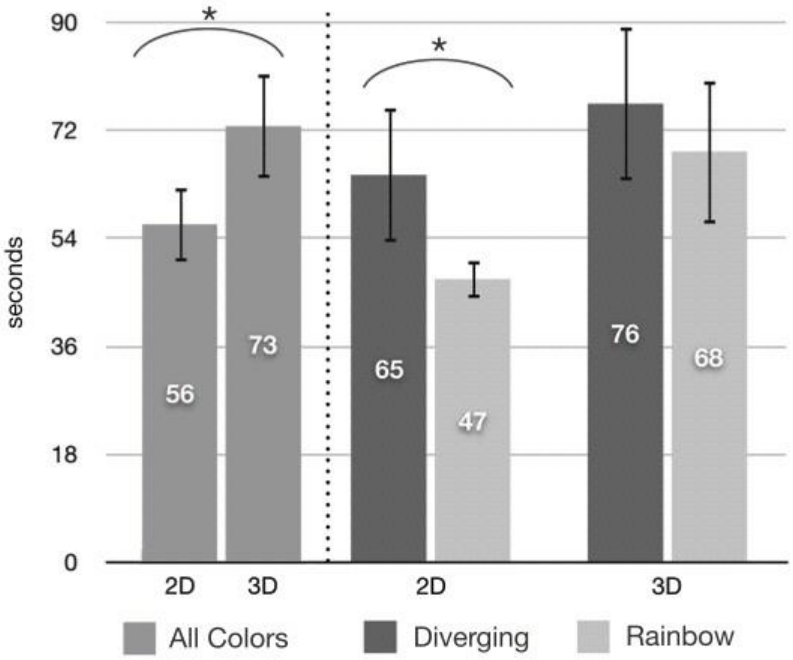
[Fig 1. Borkin et al. Artery Visualizations for Heart Disease Diagnosis. Proc InfoVis 2011.]]

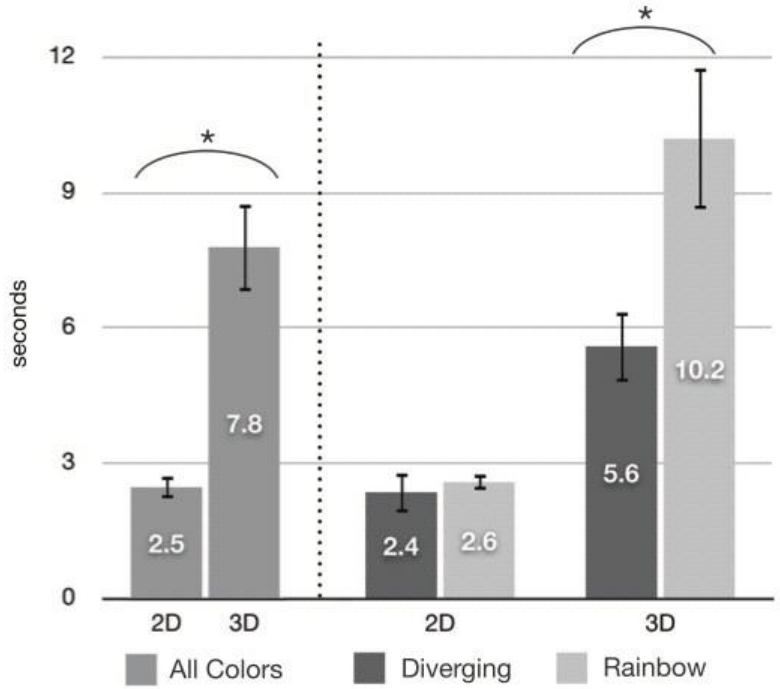
Study Results: Error





Study Results: Time





Critique

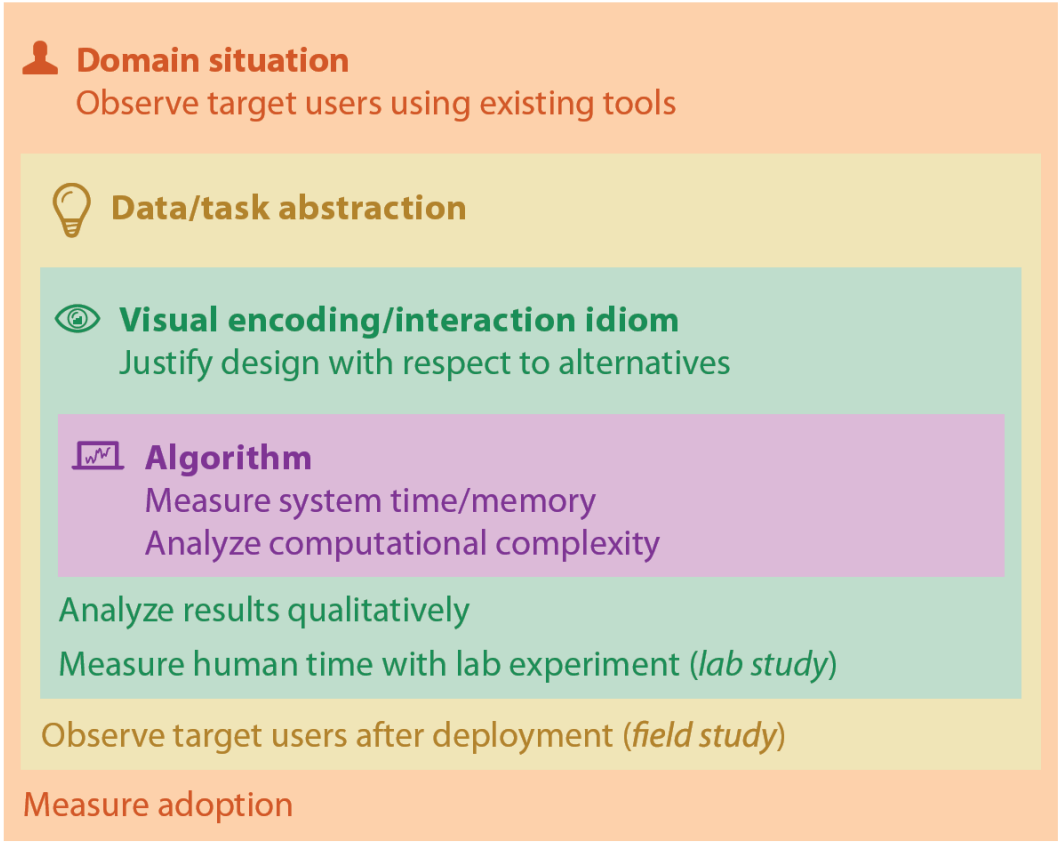
- many strengths
  - careful and well justified design, convincing human-subjects experiment
    - bringing visualization best practices to medical domain
- limitation
  - paper does not clearly communicate why colormap is diverging not sequential
    - answer by email
    - doctors care about extremely 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

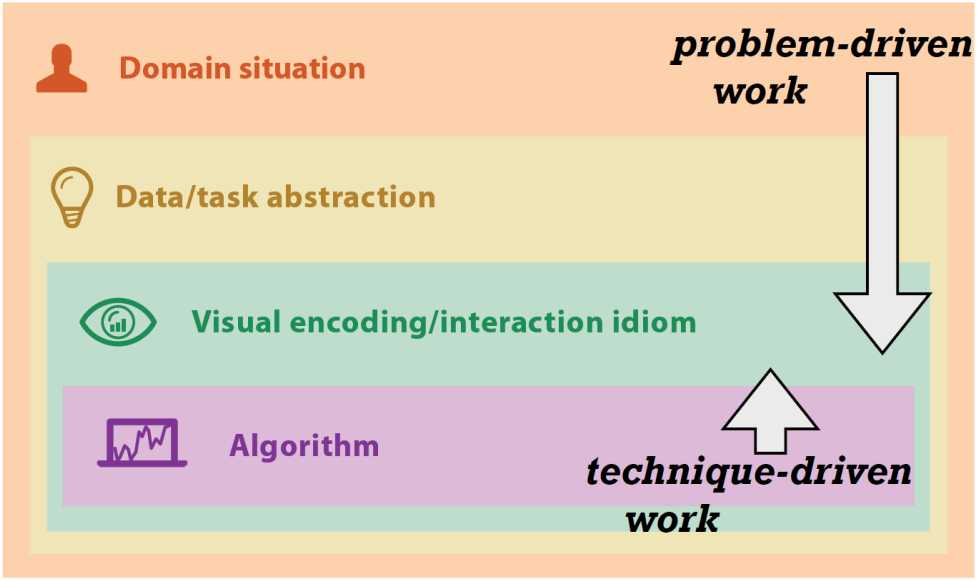
– different threats to validity at each level

Validation



mismatch: cannot show idiom good with system timings
mismatch: cannot show abstraction good with lab study

Visualization Research and Development



- 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