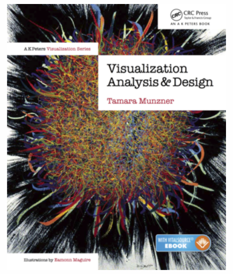
Visualization



Tamara Munzner
Department of Computer Science
University of British Columbia

UBC 314 Computer Graphics, Jan-Apr 2016

http://www.ugrad.cs.ubc.ca/~cs314/Vjan2016

Defining visualization (vis)

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

Why?...

Why have a human in the loop?

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

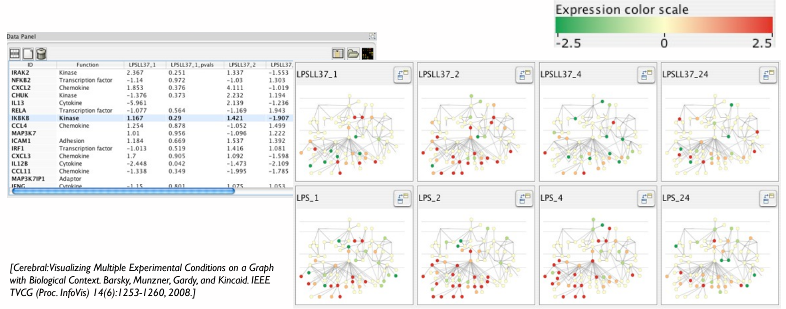
Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

- don't need vis when fully automatic solution exists and is trusted
- many analysis problems ill-specified
  - 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
  - help developers of automatic solution refine/debug, determine parameters
  - help end users of automatic solutions verify, build trust

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



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

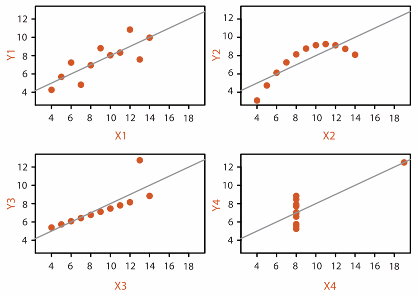
Why represent all the data?

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

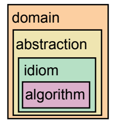
Anscombe's Quartet

Identical statistics	
x mean	9
x variance	10
y mean	8
y variance	4
x/y correlation	1

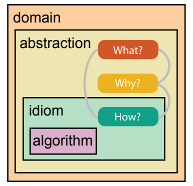


Analysis framework: Four levels, three questions

- *domain* situation
  - who are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don't just draw what you're given: transform to new form
  - **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



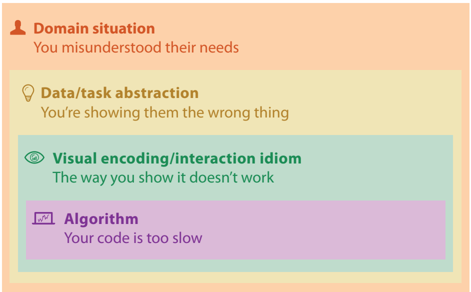
[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]



[A Multi-Level Typology of Abstract Visualization Tasks Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

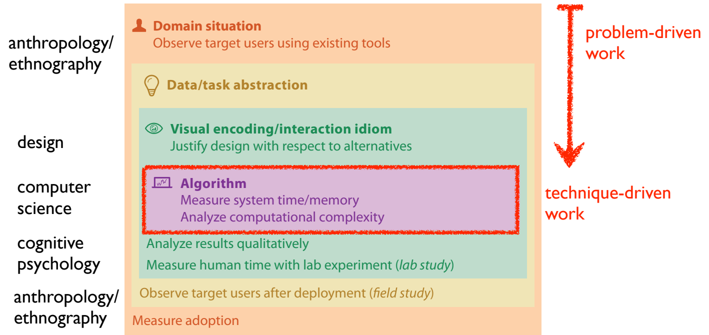
Why is validation difficult?

- different ways to get it wrong at each level

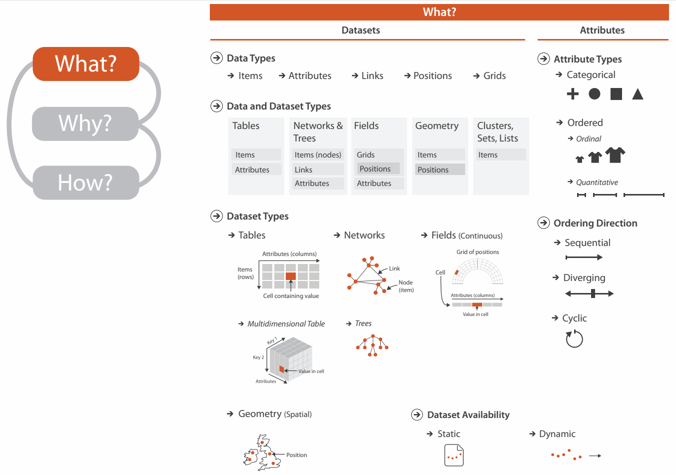


Why is validation difficult?

- solution: use methods from different fields at each level



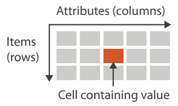
[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]



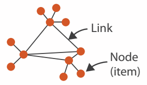
Types: Datasets and data

Dataset Types

→ Tables



→ Networks



→ Spatial

→ Fields (Continuous)

→ Geometry (Spatial)

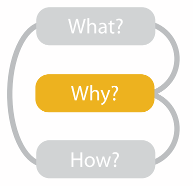
Attribute Types

→ Categorical

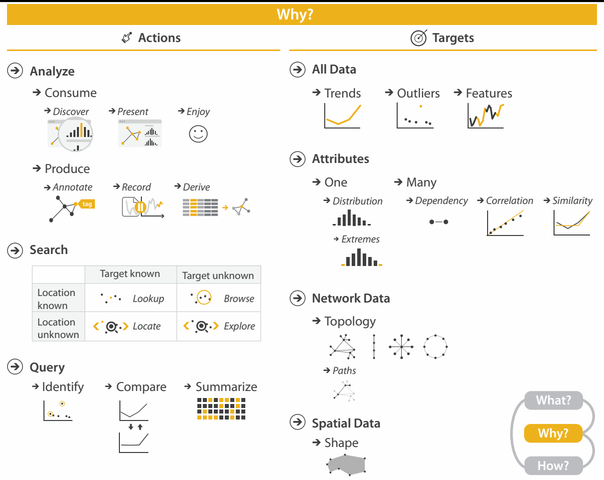
→ Ordered

→ *Ordinal*

→ *Quantitative*

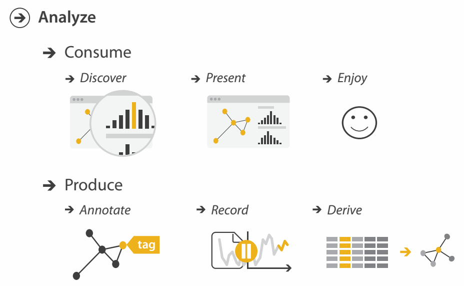


- {action, target} pairs
  - *discover distribution*
  - *compare trends*
  - *locate outliers*
  - *browse topology*



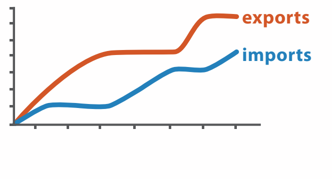
Actions: Analyze

- consume
  - discover vs present
    - classic split
    - aka explore vs explain
  - enjoy
    - newcomer
    - aka casual, social
- produce
  - annotate, record
  - derive
    - crucial design choice

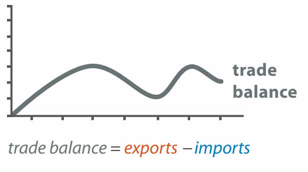


Derive

- 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



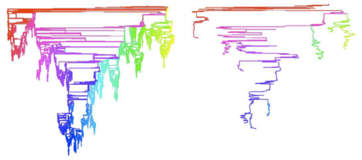
trade balance = *exports* − *imports*

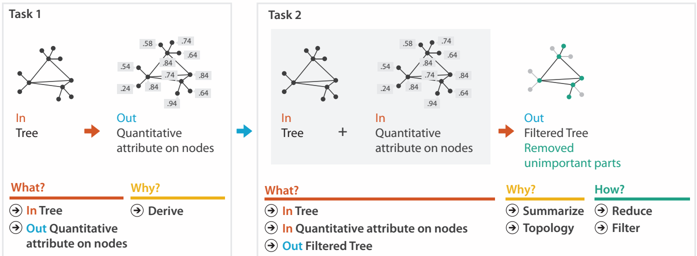
Derived Data

Analysis example: Derive one attribute

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

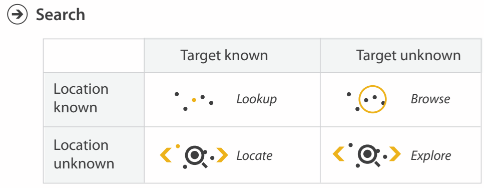
[Using Strahler numbers for real time visual exploration of huge graphs. Auber. Proc. Intl. Conf. Computer Vision and Graphics, pp. 56–69, 2002.]

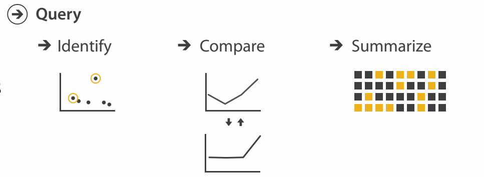




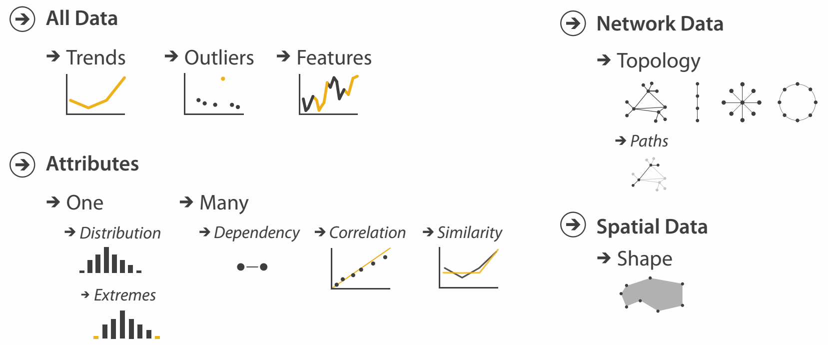
Actions: Search, query

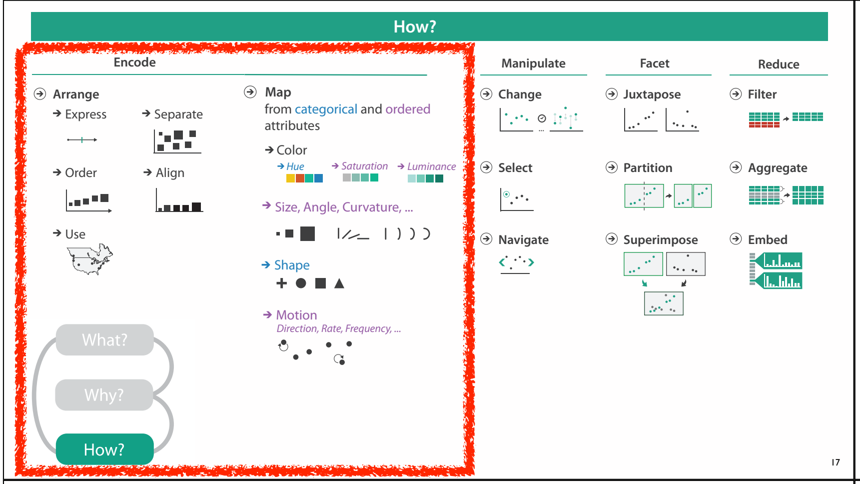
- what does user know?
  - target, location
- how much of the data matters?
  - one, some, all
- independent choices for each of these three levels
  - analyze, search, query
  - mix and match

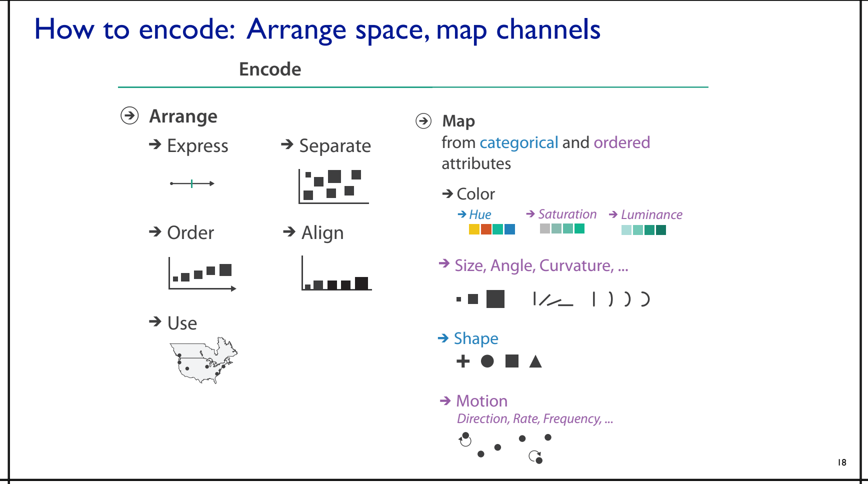


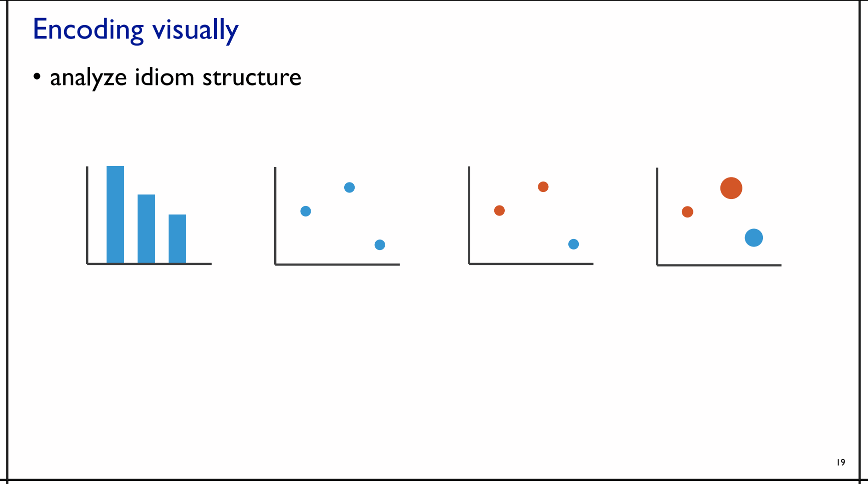


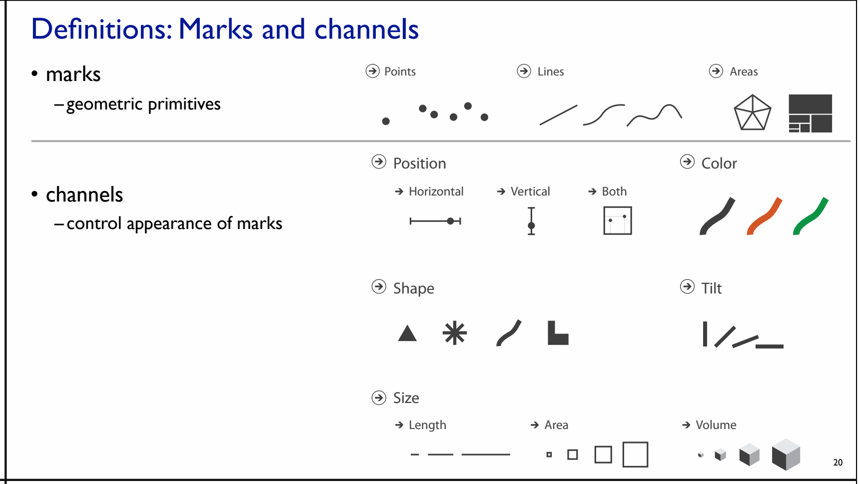
Targets

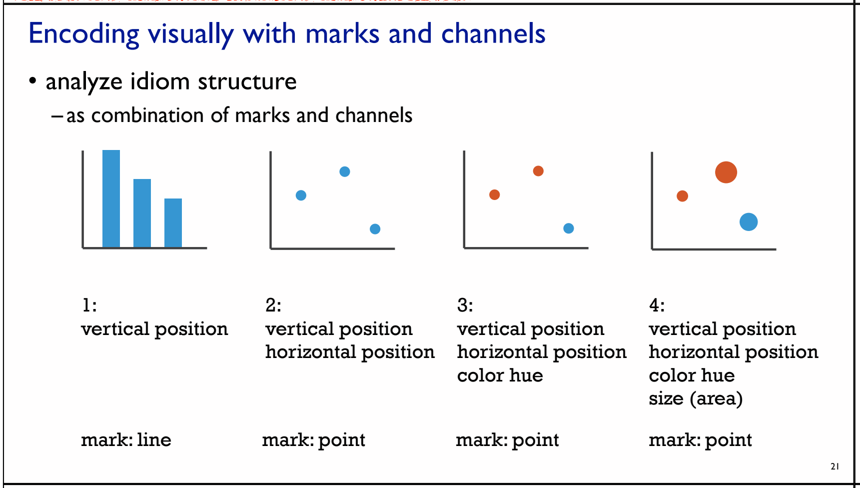


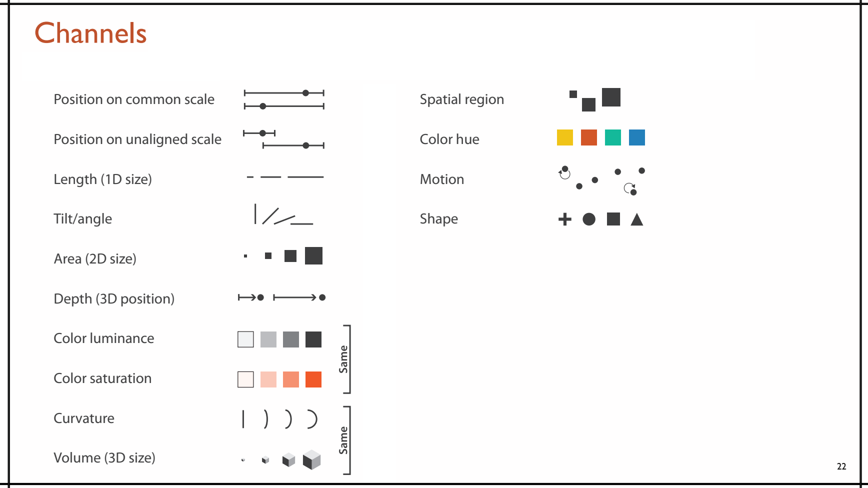

How?
Encode
Manipulate
Facet
Reduce
Arrange
Express
Separate
Order
Align
Use
Map
from categorical and ordered attributes
Color
Hue
Saturation
Luminance
Size, Angle, Curvature, ...
Shape
Motion
Direction, Rate, Frequency, ...
Change
Select
Navigate
Juxtapose
Partition
Superimpose
Filter
Aggregate
Embed
What?
Why?
How?
17

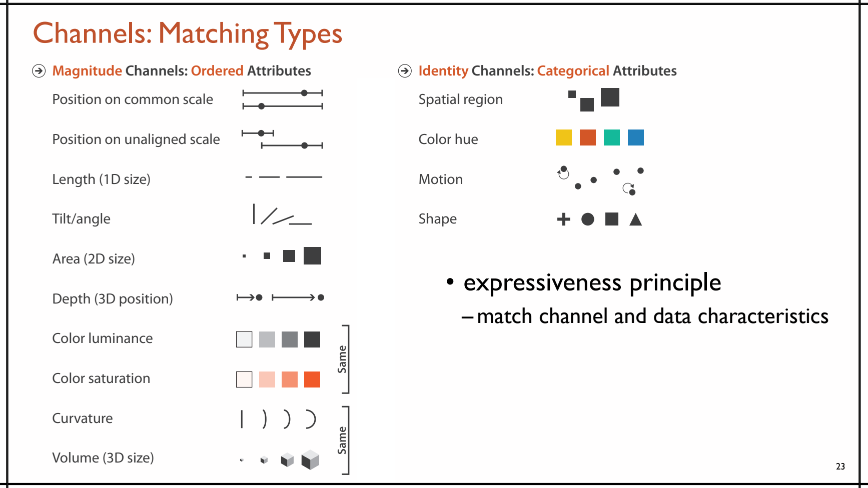

How to encode: Arrange space, map channels
Encode
Arrange
Express
Separate
Order
Align
Use
Map
from categorical and ordered attributes
Color
Hue
Saturation
Luminance
Size, Angle, Curvature, ...
Shape
Motion
Direction, Rate, Frequency, ...
18

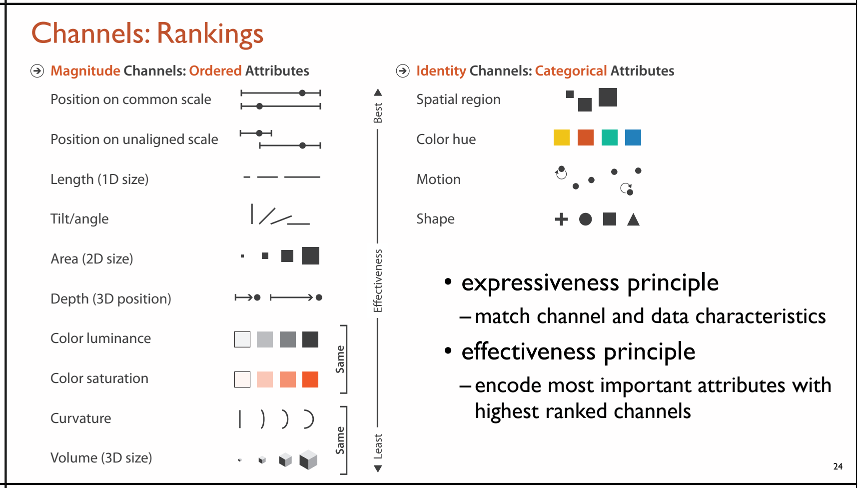

Encoding visually
• analyze idiom structure
19

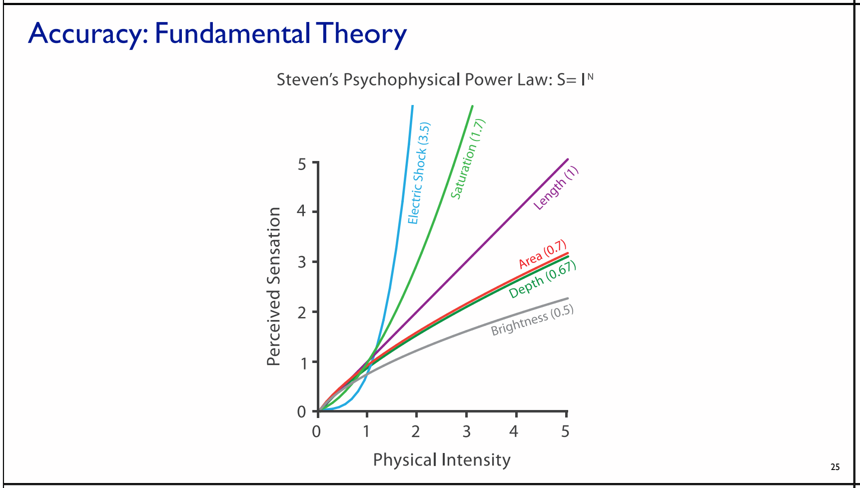

Definitions: Marks and channels
• marks
– geometric primitives
Points
Lines
Areas
• channels
– control appearance of marks
Position
Horizontal
Vertical
Both
Color
Shape
Tilt
Size
Length
Area
Volume
20

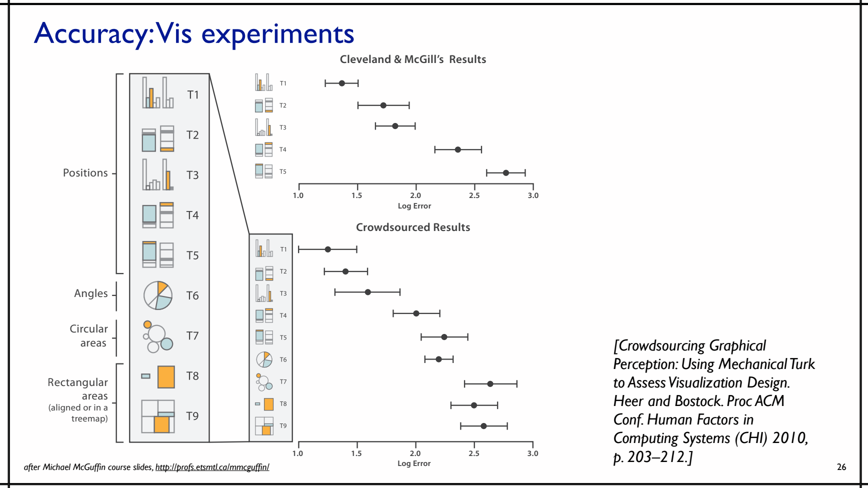

Encoding visually with marks and channels
• analyze idiom structure
– as combination of marks and channels
1:
vertical position
2:
vertical position
horizontal position
3:
vertical position
horizontal position
color hue
4:
vertical position
horizontal position
color hue
size (area)
mark: line
mark: point
mark: point
mark: point
21

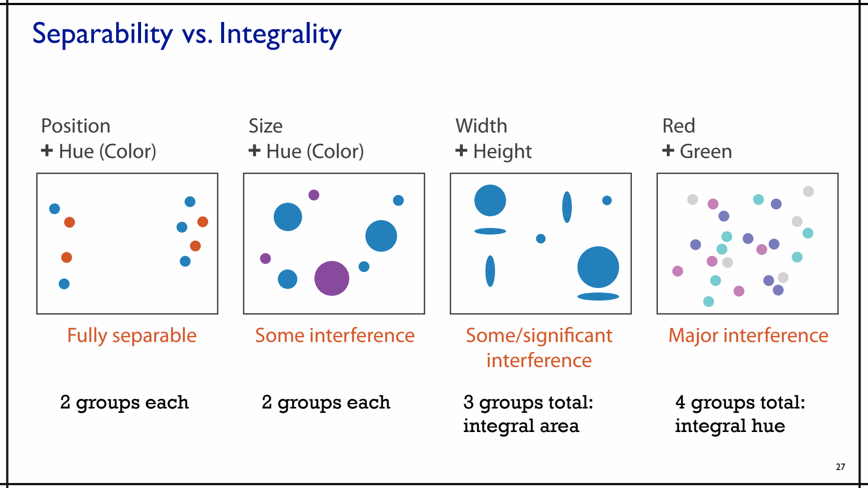

Channels
Position on common scale
Position on unaligned scale
Length (1D size)
Tilt/angle
Area (2D size)
Depth (3D position)
Color luminance
Color saturation
Curvature
Volume (3D size)
Same
Same
Spatial region
Color hue
Motion
Shape
22

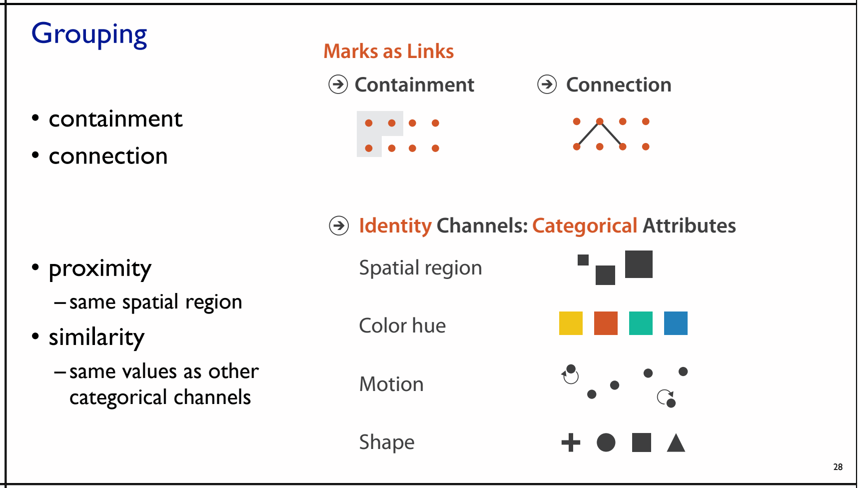

Channels: Matching Types
Magnitude Channels: Ordered Attributes
Identity Channels: Categorical Attributes
Position on common scale
Position on unaligned scale
Length (1D size)
Tilt/angle
Area (2D size)
Depth (3D position)
Color luminance
Color saturation
Curvature
Volume (3D size)
Same
Same
Spatial region
Color hue
Motion
Shape
• expressiveness principle
– match channel and data characteristics
23

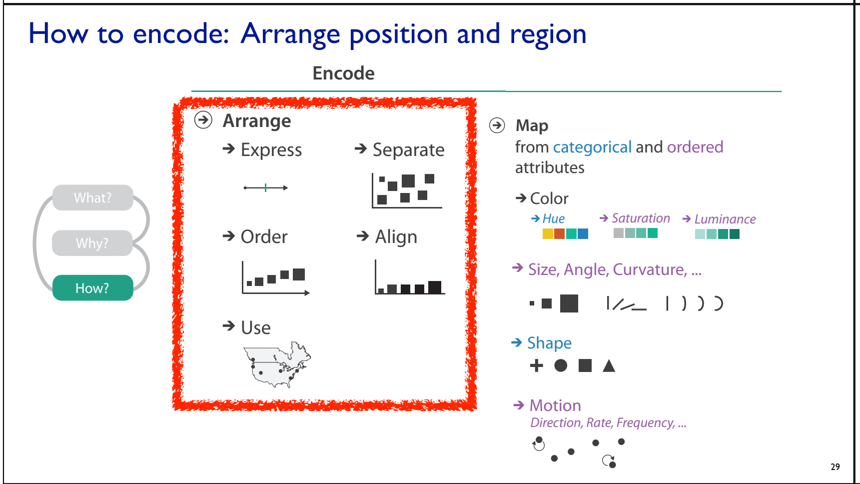

Channels: Rankings
Magnitude Channels: Ordered Attributes
Identity Channels: Categorical Attributes
Position on common scale
Position on unaligned scale
Length (1D size)
Tilt/angle
Area (2D size)
Depth (3D position)
Color luminance
Color saturation
Curvature
Volume (3D size)
Best
Effectiveness
Least
Same
Same
Spatial region
Color hue
Motion
Shape
• expressiveness principle
– match channel and data characteristics
• effectiveness principle
– encode most important attributes with highest ranked channels
24

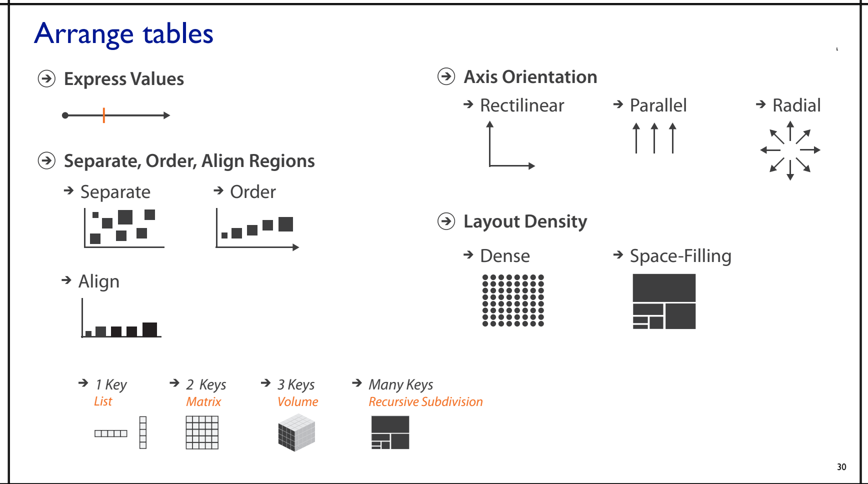

Accuracy: Fundamental Theory
Steven's Psychophysical Power Law: S= I^N
Electric Shock (3.5)
Saturation (1.7)
Length (1)
Area (0.7)
Depth (0.67)
Brightness (0.5)
Perceived Sensation
Physical Intensity
25

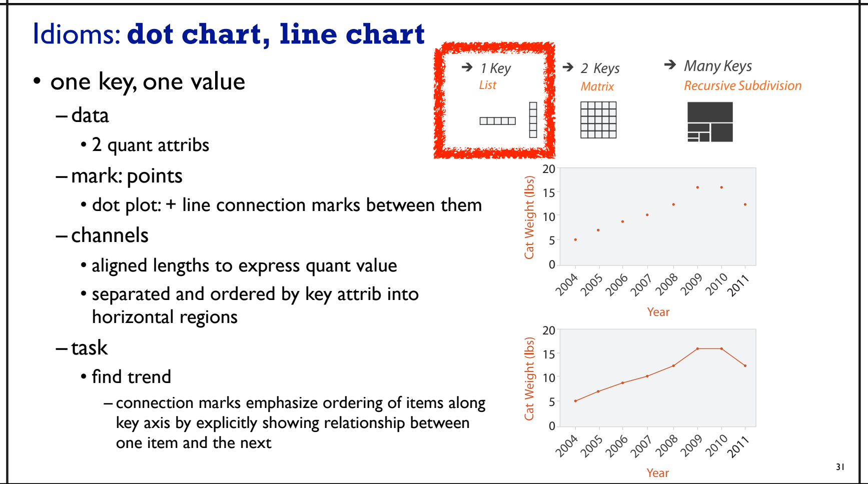

Accuracy: Vis experiments
Cleveland & McGill's Results
Crowdsourced Results
Positions
Angles
Circular areas
Rectangular areas
(aligned or in a treemap)
Log Error
[Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. Heer and Bostock. Proc ACM Conf. Human Factors in Computing Systems (CHI) 2010, p. 203–212.]
after Michael McGuffin course slides, http://profs.etsmtl.ca/mmcguffin/
26

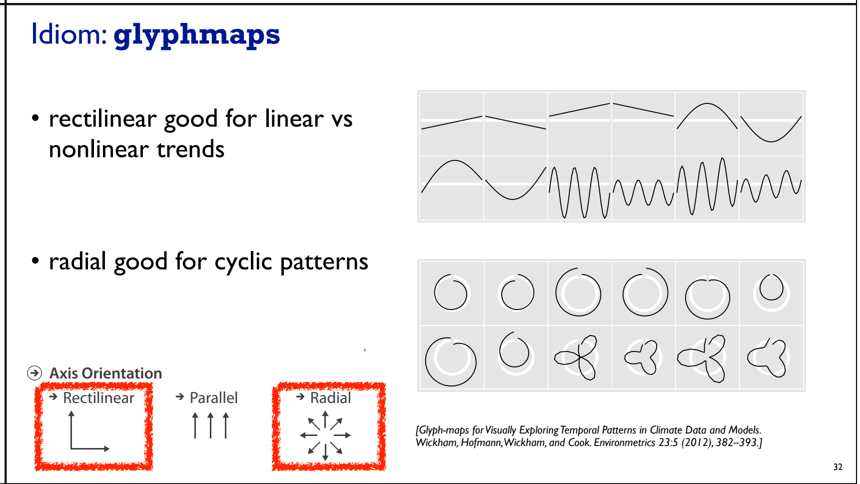

Separability vs. Integrality
Position
+ Hue (Color)
Size
+ Hue (Color)
Width
+ Height
Red
+ Green
Fully separable
Some interference
Some/significant interference
Major interference
2 groups each
2 groups each
3 groups total: integral area
4 groups total: integral hue
27


Grouping
• containment
• connection
• proximity
– same spatial region
• similarity
– same values as other categorical channels
Marks as Links
Containment
Connection
Identity Channels: Categorical Attributes
Spatial region
Color hue
Motion
Shape
28

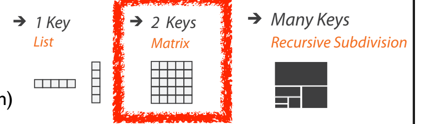

How to encode: Arrange position and region
Encode
Arrange
Express
Separate
Order
Align
Use
Map
from categorical and ordered attributes
Color
Hue
Saturation
Luminance
Size, Angle, Curvature, ...
Shape
Motion
Direction, Rate, Frequency, ...
What?
Why?
How?
29


Arrange tables
Express Values
Separate, Order, Align Regions
Separate
Order
Align
Axis Orientation
Rectilinear
Parallel
Radial
Layout Density
Dense
Space-Filling
1 Key
List
2 Keys
Matrix
3 Keys
Volume
Many Keys
Recursive Subdivision
30

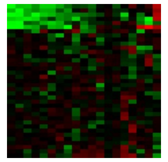

Idioms: dot chart, line chart
• one key, one value
– data
• 2 quant attribs
– mark: points
• dot plot: + line connection marks between them
– channels
• aligned lengths to express quant value
• separated and ordered by key attrib into horizontal regions
– task
• find trend
– connection marks emphasize ordering of items along key axis by explicitly showing relationship between one item and the next
1 Key
List
2 Keys
Matrix
Many Keys
Recursive Subdivision
Cat Weight (lbs)
Year
31


Idiom: glyphmaps
• rectilinear good for linear vs nonlinear trends
• radial good for cyclic patterns
Axis Orientation
Rectilinear
Parallel
Radial
[Glyph-maps for Visually Exploring Temporal Patterns in Climate Data and Models. Wickham, Hofmann, Wickham, and Cook. Environmetrics 23:5 (2012), 382–393.]
32

Idiom: **heatmap**

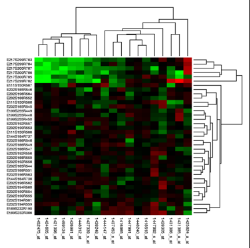


- two keys, one value
  - data
    - 2 categ attribs (gene, experimental condition)
    - 1 quant attrib (expression levels)
  - marks: area
    - separate and align in 2D matrix
      - indexed by 2 categorical attributes
  - channels
    - color by quant attrib
      - (ordered diverging colormap)
  - task
    - find clusters, outliers
  - scalability
    - 1M items, 100s of categ levels, ~10 quant attrib levels



Idiom: **cluster heatmap**

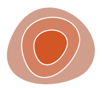
- in addition
  - derived data
    - 2 cluster hierarchies
  - dendrogram
    - parent-child relationships in tree with connection line marks
    - leaves aligned so interior branch heights easy to compare
  - heatmap
    - marks (re-)ordered by cluster hierarchy traversal



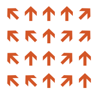
Arrange spatial data

- Use Given
  - Geometry
    - *Geographic*
    - *Other Derived*

  - Spatial Fields
    - *Scalar Fields (one value per cell)*
      - *Isocontours*
      - *Direct Volume Rendering*

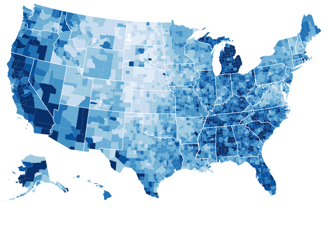


    - *Vector and Tensor Fields (many values per cell)*
      - *Flow Glyphs (local)*
      - *Geometric (sparse seeds)*
      - *Textures (dense seeds)*
      - *Features (globally derived)*



Idiom: **choropleth map**

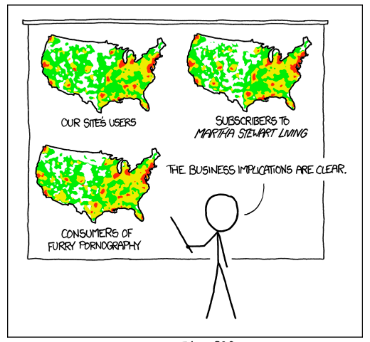
- ***use*** given spatial data
  - when central task is understanding spatial relationships
- data
  - geographic geometry
  - table with 1 quant attribute per region
- encoding
  - use given geometry for area mark boundaries
  - sequential segmented colormap



http://bl.ocks.org/mbostock/4060606

Population maps trickiness

- beware!

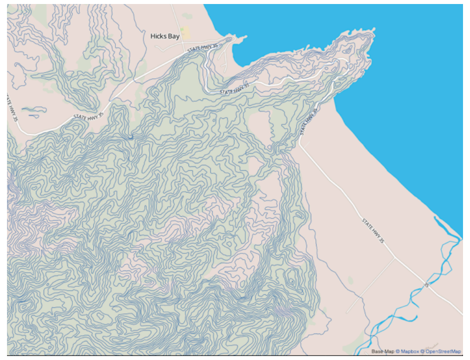


PET PEEVE #208:
GEOGRAPHIC PROFILE MAPS WHICH ARE
BASICALLY JUST POPULATION MAPS

[https://xkcd.com/1138]

Idiom: **topographic map**

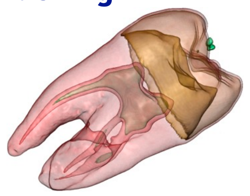
- data
  - geographic geometry
  - scalar spatial field
    - 1 quant attribute per grid cell
- derived data
  - isoline geometry
    - isocontours computed for specific levels of scalar values



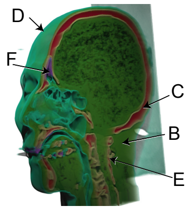
Land Information New Zealand Data Service

Idioms: **isosurfaces, direct volume rendering**

- data
  - scalar spatial field
    - 1 quant attribute per grid cell
- task
  - shape understanding, spatial relationships
- isosurface
  - derived data: isocontours computed for specific levels of scalar values
- direct volume rendering
  - transfer function maps scalar values to color, opacity
    - no derived geometry



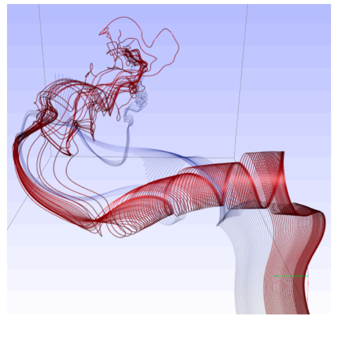
[Interactive Volume Rendering Techniques. Kniss. Master's thesis, University of Utah Computer Science, 2002.]



[Multidimensional Transfer Functions for Volume Rendering. Kniss, Kindlmann, and Hansen. In The Visualization Handbook, edited by Charles Hansen and Christopher Johnson, pp. 189–210. Elsevier, 2005.]

Idiom: **similarity-clustered streamlines**

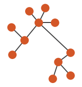
- data
  - 3D vector field
- derived data (from field)
  - streamlines: trajectory particle will follow
- derived data (per streamline)
  - curvature, torsion, tortuosity
  - signature: complex weighted combination
  - compute cluster hierarchy across all signatures
  - encode: color and opacity by cluster
- tasks
  - find features, query shape
- scalability
  - millions of samples, hundreds of streamlines



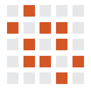
[Similarity Measures for Enhancing Interactive Streamline Seeding. McLoughlin, Jones, Laramee, Malki, Masters, and. Hansen. IEEE Trans. Visualization and Computer Graphics 19:8 (2013), 1342–1353.]

Arrange networks and trees

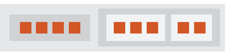
- **Node–Link Diagrams**
  Connection Marks
  ✔ NETWORKS ✔ TREES



- **Adjacency Matrix**
  Derived Table
  ✔ NETWORKS ✔ TREES

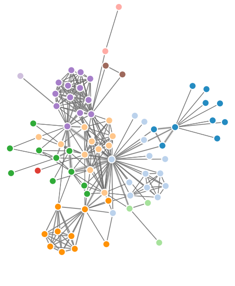


- **Enclosure**
  Containment Marks
  ✕ NETWORKS ✔ TREES



Idiom: **force-directed placement**

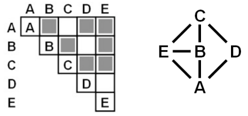
- visual encoding
  - link connection marks, node point marks
- considerations
  - spatial position: no meaning directly encoded
    - left free to minimize crossings
  - proximity semantics?
    - sometimes meaningful
    - sometimes arbitrary, artifact of layout algorithm
    - tension with length
      - long edges more visually salient than short
- tasks
  - explore topology; locate paths, clusters
- scalability
  - node/edge density E < 4N



http://mbostock.github.com/d3/ex/force.html

Idiom: **adjacency matrix view**

- data: network
  - transform into same data/encoding as heatmap
- derived data: table from network
  - 1 quant attrib
    - weighted edge between nodes
  - 2 categ attribs: node list x 2
- visual encoding
  - cell shows presence/absence of edge
- scalability
  - 1K nodes, 1M edges



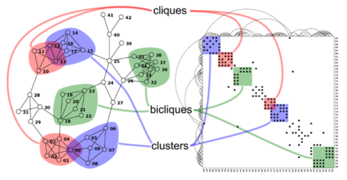
[NodeTrix: a Hybrid Visualization of Social Networks. Henry, Fekete, and McGuffin. IEEE TVCG (Proc. InfoVis) 13(6):1302-1309, 2007.]

[Points of view: Networks. Gehlenborg and Wong. Nature Methods 9:115.]

Connection vs. adjacency comparison

- adjacency matrix strengths
  - predictability, scalability, supports reordering
  - some topology tasks trainable
- node-link diagram strengths
  - topology understanding, path tracing
  - intuitive, no training needed
- empirical study
  - node-link best for small networks
  - matrix best for large networks
    - if tasks don't involve topological structure!

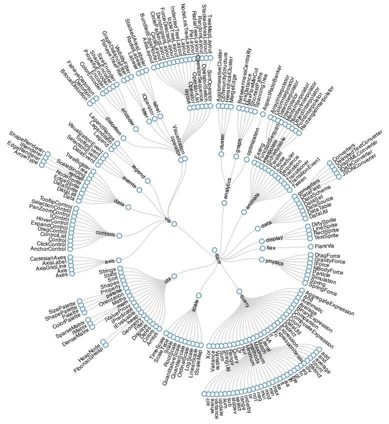
[On the readability of graphs using node-link and matrix-based representations: a controlled experiment and statistical analysis. Ghoniem, Fekete, and Castagliola. Information Visualization 4:2 (2005), 114–135.]



http://www.michaelmcguffin.com/courses/vis/patternsInAdjacencyMatrix.png

Idiom: **radial node-link tree**

- data
  - tree
- encoding
  - link connection marks
  - point node marks
  - radial axis orientation
    - angular proximity: siblings
    - distance from center: depth in tree
- tasks
  - understanding topology, following paths
- scalability
  - 1K - 10K nodes



http://mbostock.github.com/d3/ex/tree.html

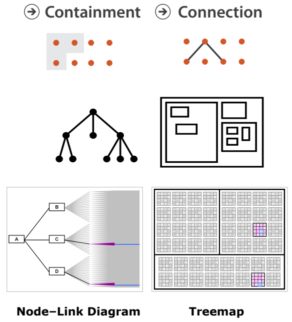
Idiom: **treemap**

- data
  - tree
  - 1 quant attrib at leaf nodes
- encoding
  - area containment marks for hierarchical structure
  - rectilinear orientation
  - size encodes quant attrib
- tasks
  - query attribute at leaf nodes
- scalability
  - 1M leaf nodes

http://tulip.labri.fr/Documentation/3_7/userHandbook/html/ch06.html

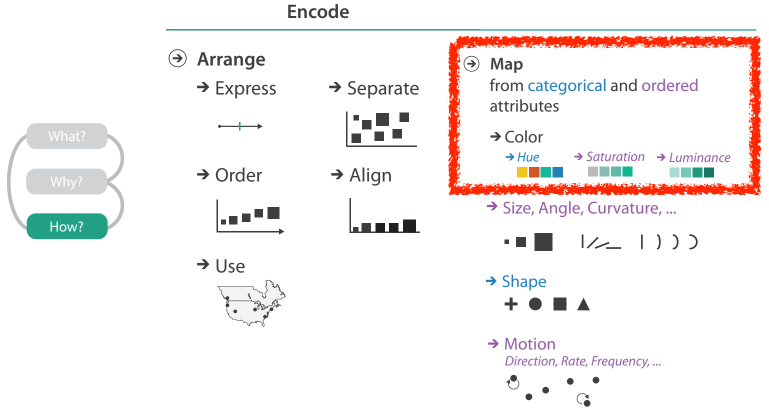
Connection vs. containment comparison

- marks as links (vs. nodes)
  - common case in network drawing
  - 1D case: connection
    - ex: all node-link diagrams
    - emphasizes topology, path tracing
    - networks and trees
  - 2D case: containment
    - ex: all treemap variants
    - emphasizes attribute values at leaves (size coding)
    - only trees



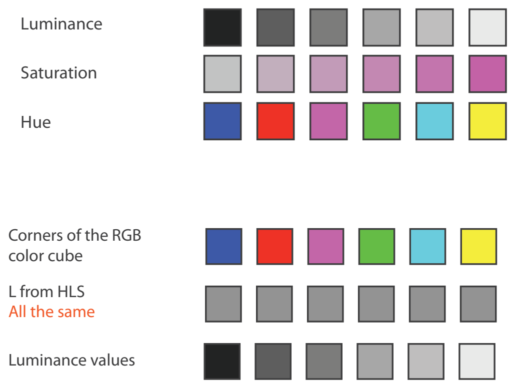
[Elastic Hierarchies: Combining Treemaps and Node-Link Diagrams. Dong, McGuffin, and Chignell. Proc. InfoVis 2005, p. 57-64.]

How to encode: Mapping color



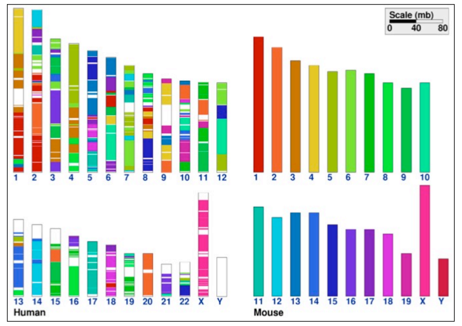
Color: Luminance, saturation, hue

- 3 channels
  - identity for categorical
    - hue
  - magnitude for ordered
    - luminance
    - saturation
- RGB: poor for encoding
- HSL: better, but beware
  - lightness ≠ luminance



Categorical color: Discriminability constraints

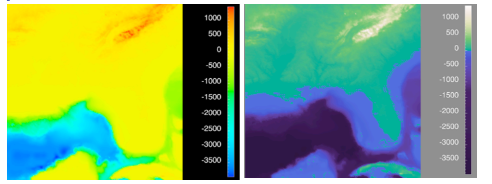
- noncontiguous small regions of color: only 6-12 bins



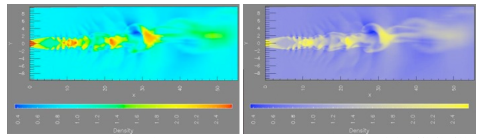
[Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. Sinha and Meller. BMC Bioinformatics, 8:82, 2007.]

Ordered color: Rainbow is poor default

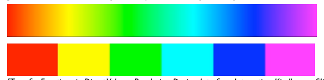
- problems
  - perceptually unordered
  - perceptually nonlinear
- benefits
  - fine-grained structure visible and nameable
- alternatives
  - fewer hues for large-scale structure
  - multiple hues with monotonically increasing luminance for fine-grained
  - segmented rainbows good for categorical, ok for binned



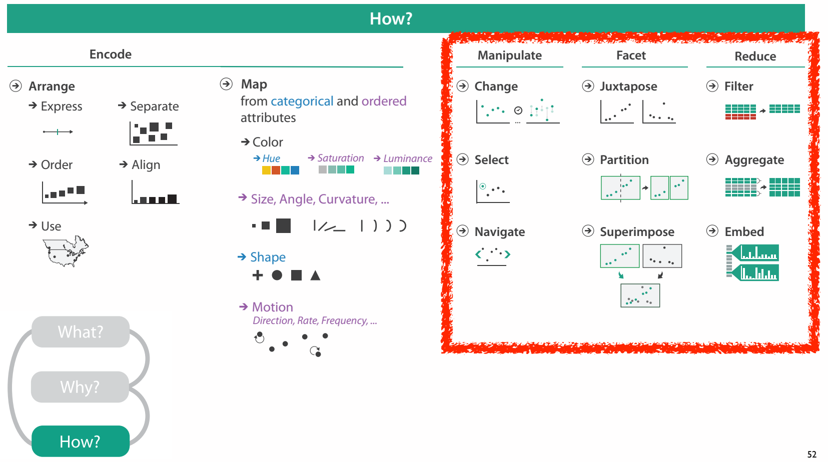
[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998. http://www.research.ibm.com/people/l/lloydt/color/color.HTM]



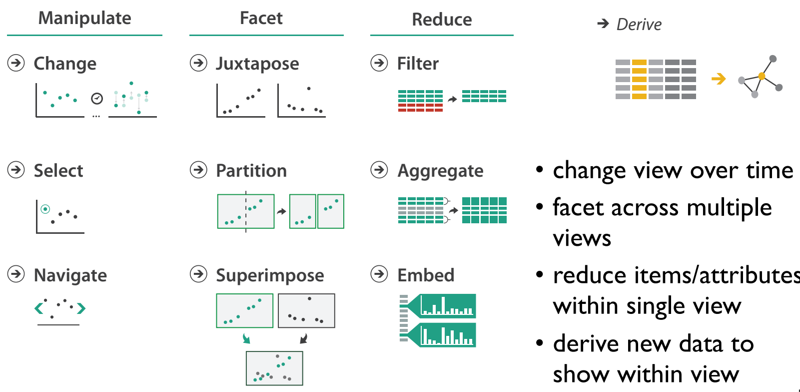
[A Rule-based Tool for Assisting Colormap Selection. Bergman, Rogowitz, and Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]



[Transfer Functions in Direct Volume Rendering: Design, Interface, Interaction. Kindlmann. SIGGRAPH 2002 Course Notes]



How to handle complexity: 3 more strategies + 1 previous

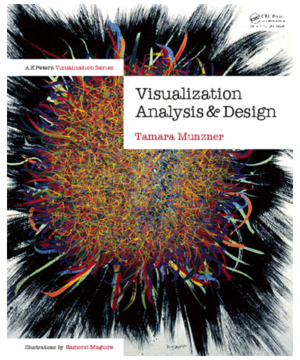


- change view over time
- facet across multiple views
- reduce items/attributes within single view
- derive new data to show within view

More Information

@tamaramunzner

- book page (including tutorial lecture slides)
  http://www.cs.ubc.ca/~tmm/vadbook
  - illustrations: Eamonn Maguire
- grad class CPSC 547
  - usually taught fall term



Visualization Analysis and Design.
Munzner. A K Peters Visualization Series, CRC Press, Visualization Series, 2014.