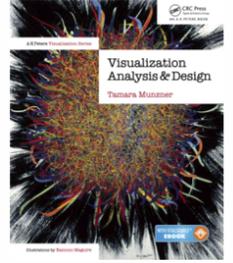


Visualization Analysis & Design



Tamara Munzner
Department of Computer Science
University of British Columbia

City University London
February 3 2015, London UK

<http://www.cs.ubc.ca/~tmm/talks.html#vad15london>

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.

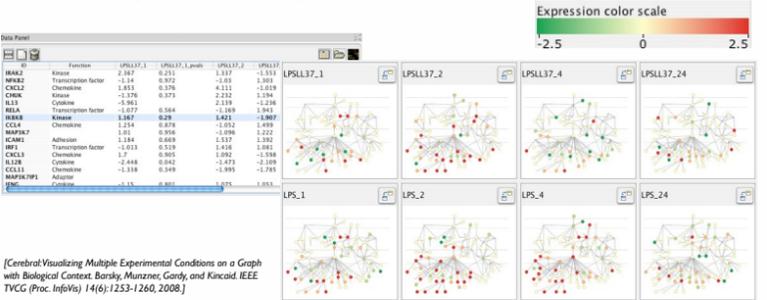
Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

- 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.

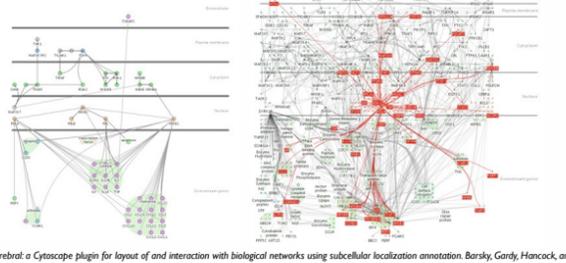
- external representation: replace cognition with perception



Why have a computer in the loop?

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

- beyond human patience: scale to large datasets, support interactivity
 - consider: what aspects of hand-drawn diagrams are important?



Why depend on vision?

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

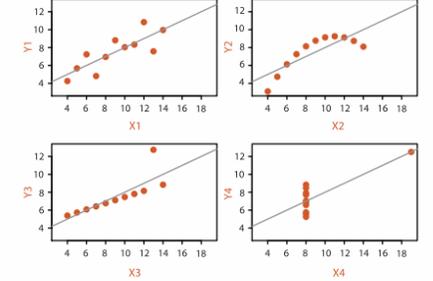
- human visual system is high-bandwidth channel to brain
 - overview possible due to background processing
 - subjective experience of seeing everything simultaneously
 - significant processing occurs in parallel and pre-attentively
- sound: lower bandwidth and different semantics
 - overview not supported
 - subjective experience of sequential stream
- touch/haptics: impoverished record/replay capacity
 - only very low-bandwidth communication thus far
- taste, smell: no viable record/replay devices

Why show the data in detail?

- summaries lose information
 - 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 |



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)
 - representations 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: enable entirely new kinds of analysis
 - faster: speed up existing workflows

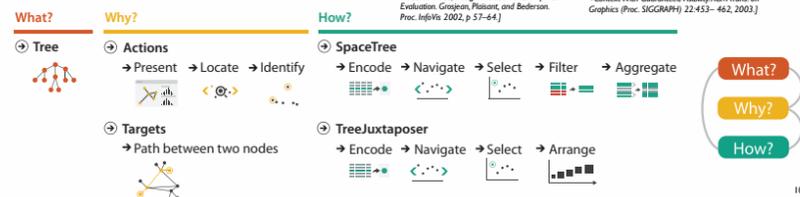
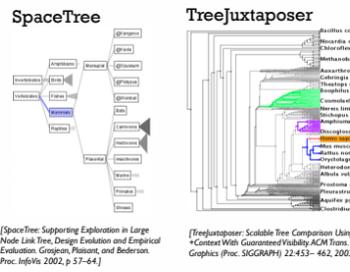
Why are there resource limitations?

Vis designers must take into account three very different kinds of resource limitations: those of computers, of humans, and of displays.

- computational limits
 - processing time
 - 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

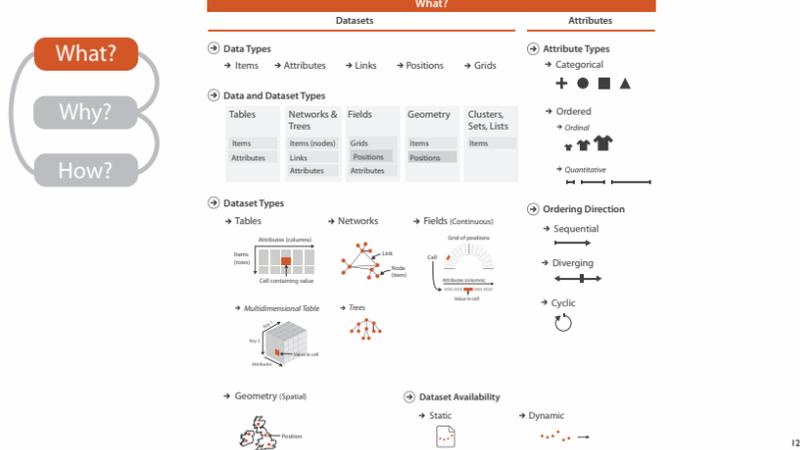
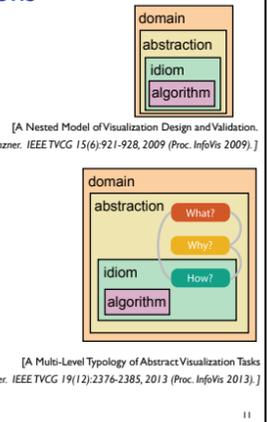
Why analyze?

- imposes a structure on huge design space
 - scaffold to help you think systematically about choices
 - analyzing existing as stepping stone to designing new

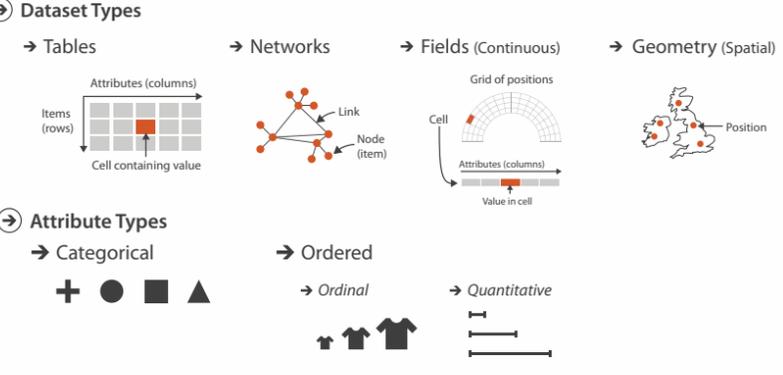


Analysis framework: Four levels, three questions

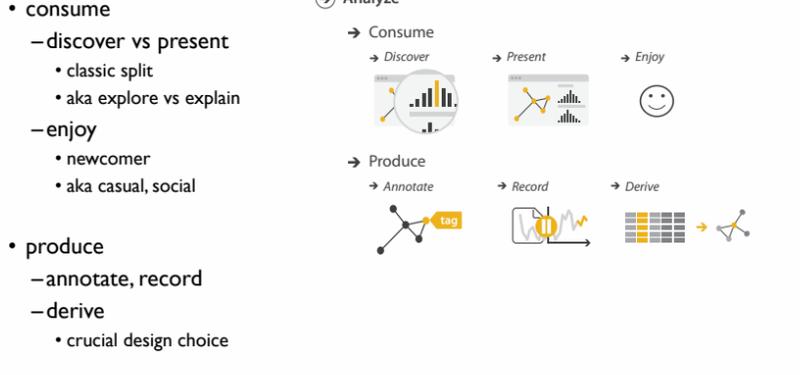
- domain situation
 - who are the target users?
- abstraction
 - translate from specifics of domain to vocabulary of vis
 - **what** is shown? **data abstraction**
 - **why** is the user looking at it? **task abstraction**
- idiom
 - **how** is it shown?
 - **visual encoding idiom**: how to draw
 - **interaction idiom**: how to manipulate
- algorithm
 - efficient computation



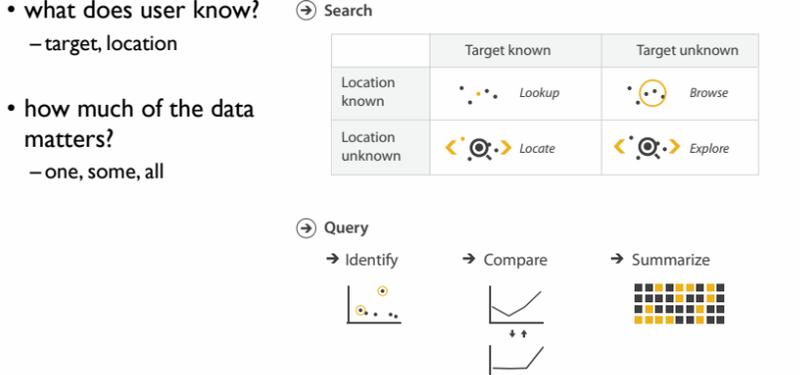
Dataset and data types



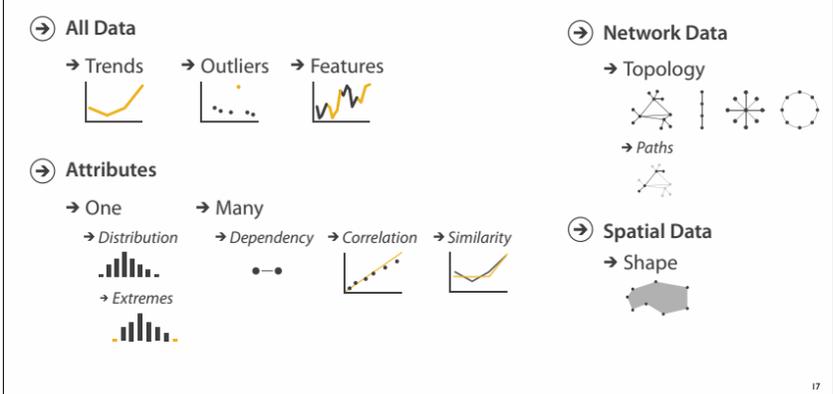
Actions, high-level: Analyze



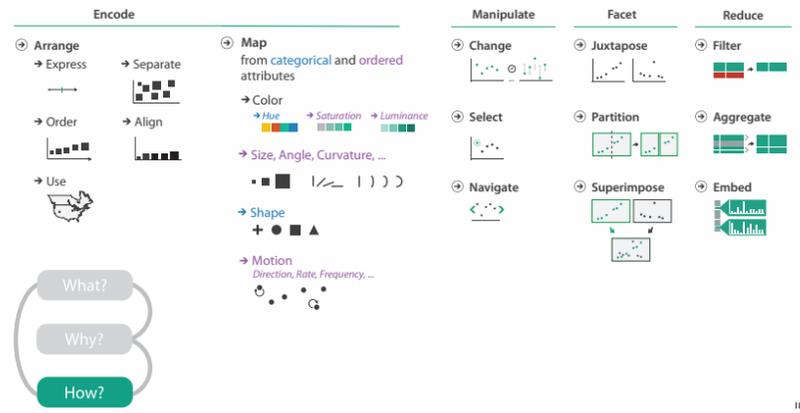
Actions: Mid-level search, low-level query



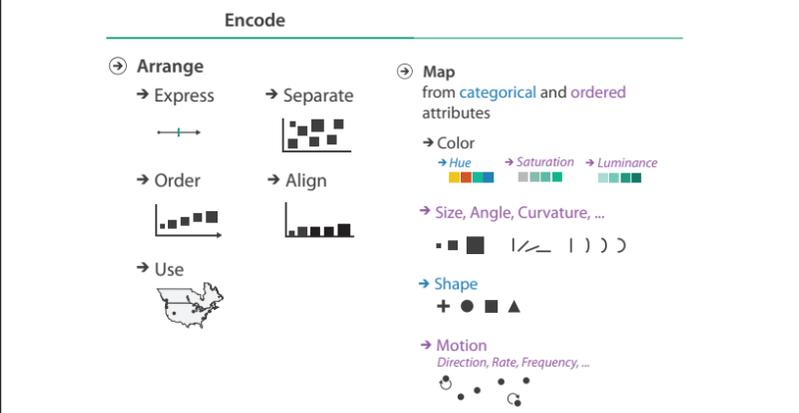
Targets



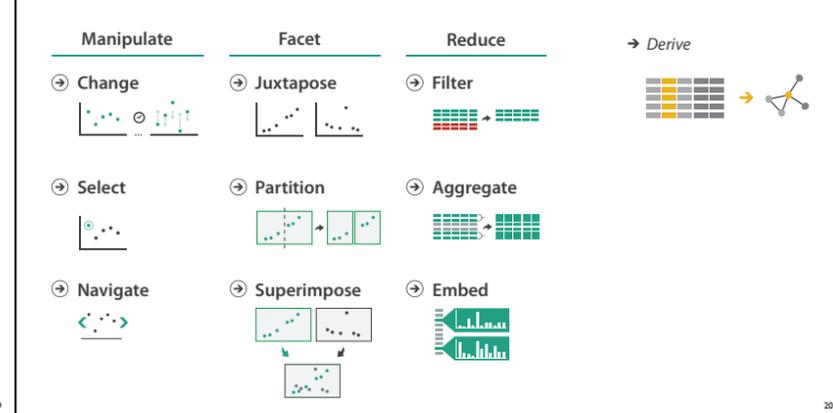
How?



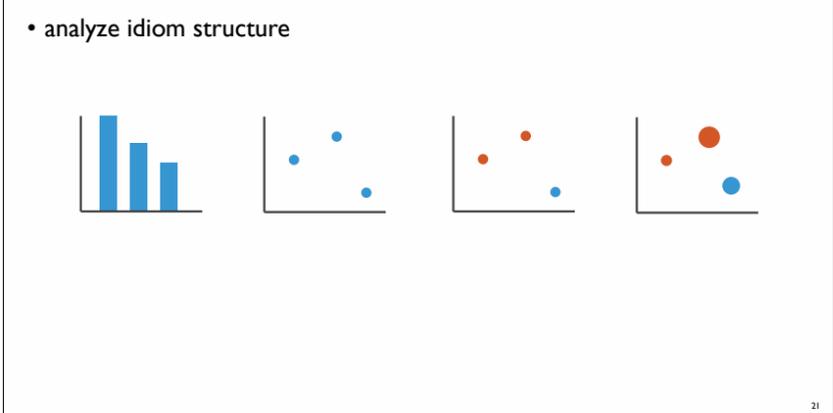
How to encode: Arrange space, map channels



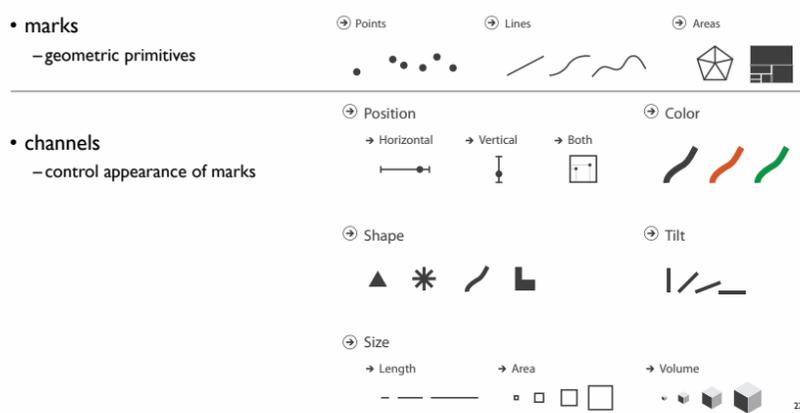
How to handle complexity: 3 more strategies + 1 previous



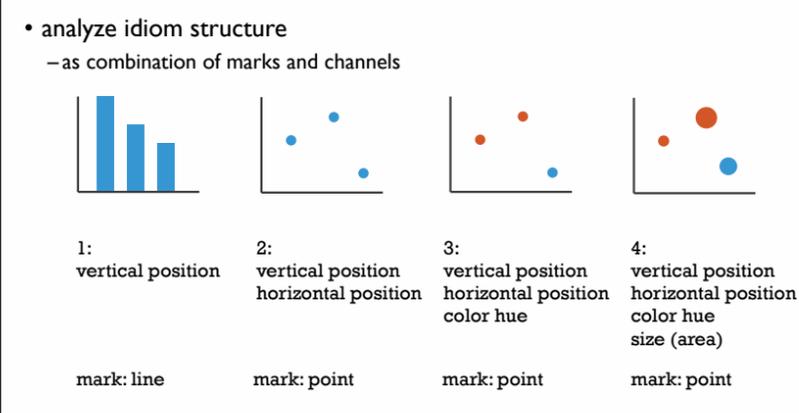
Encoding visually



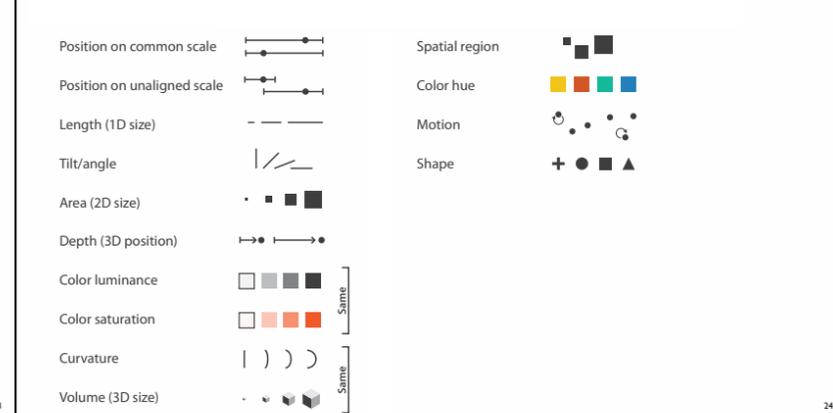
Definitions: Marks and channels



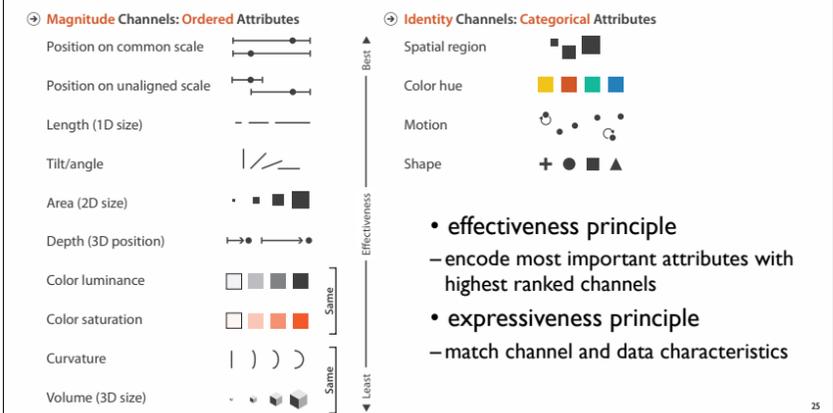
Encoding visually with marks and channels



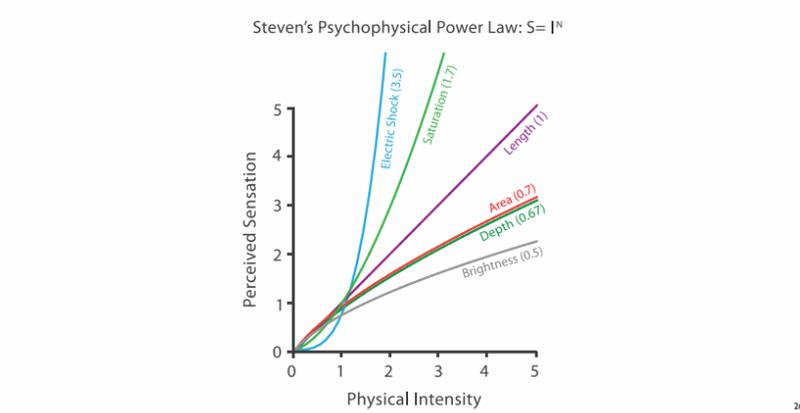
Channels



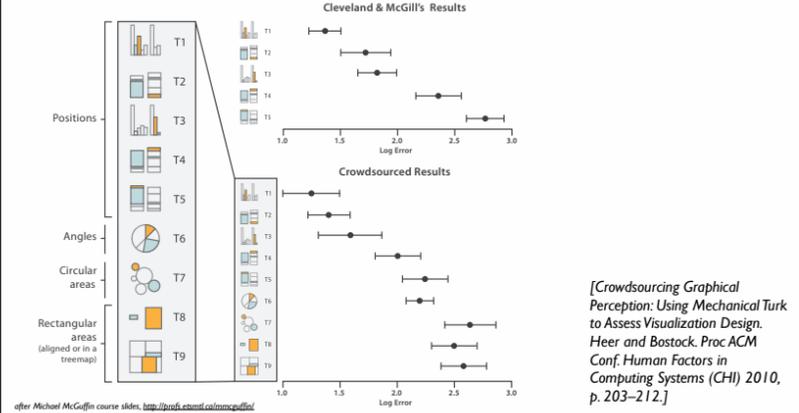
Channels: Rankings



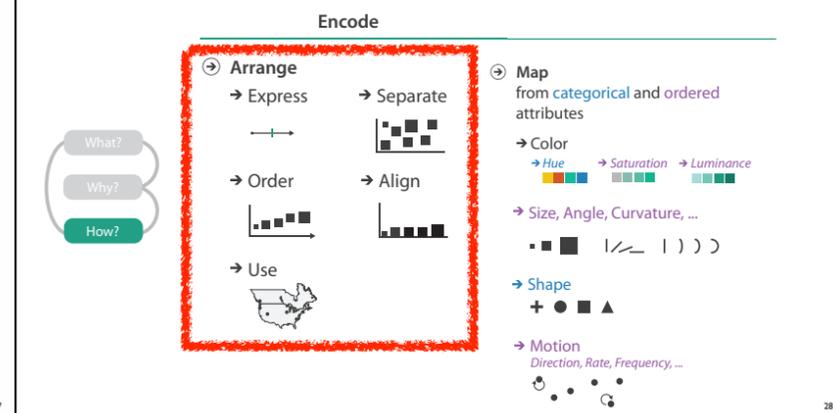
Accuracy: Fundamental Theory



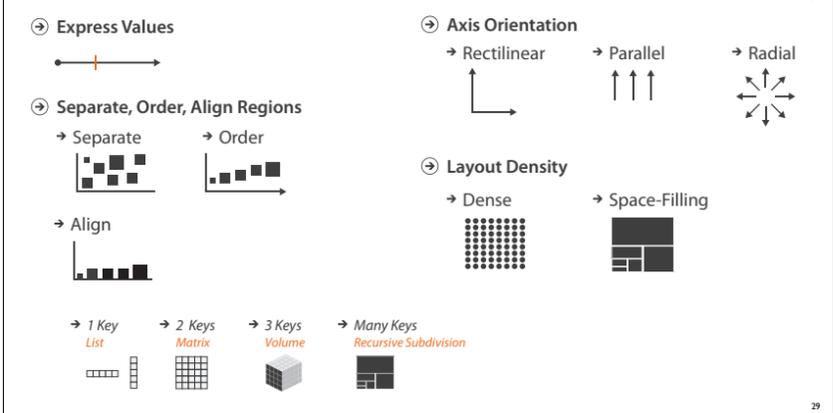
Accuracy: Vis experiments



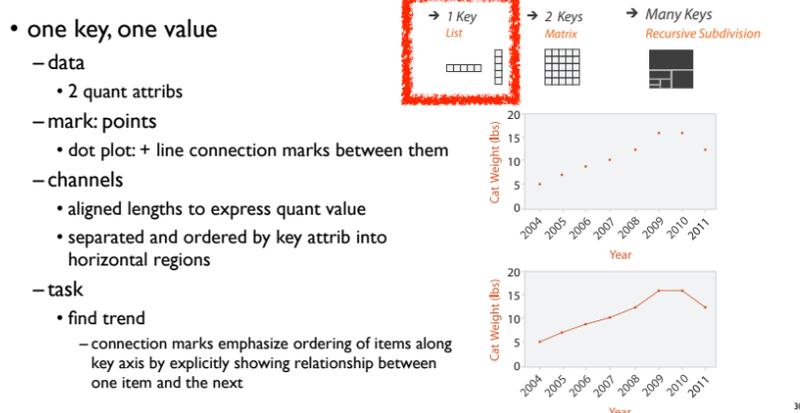
How to encode: Arrange position and region



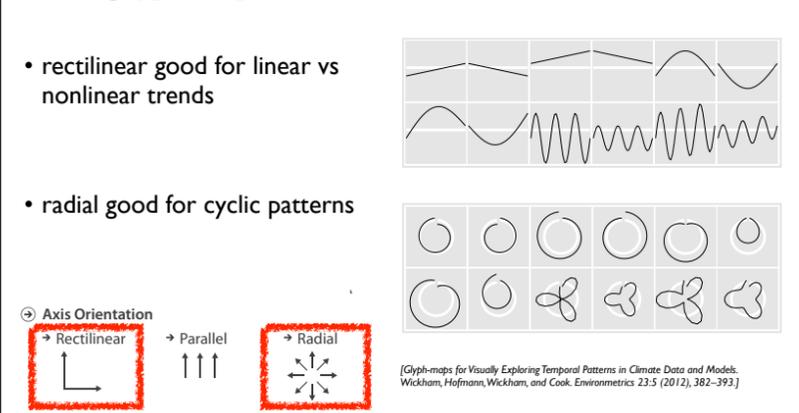
Arrange tables



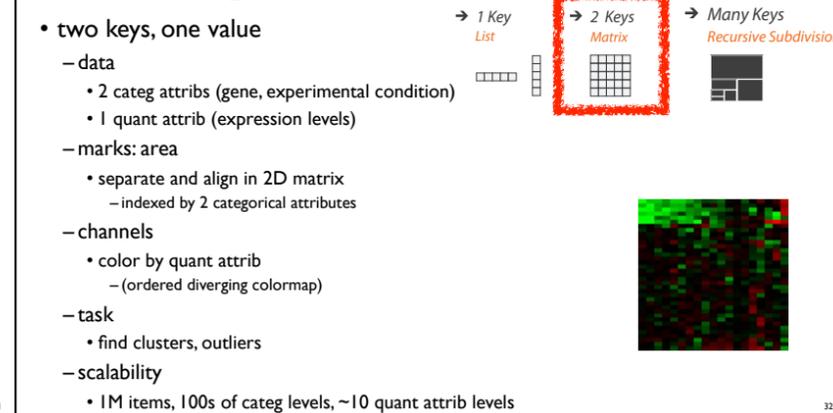
Idioms: dot chart, line chart



Idiom: glyphmaps

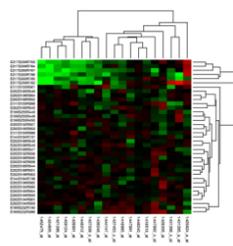


Idiom: heatmap



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



33

Arrange spatial data

Use Given

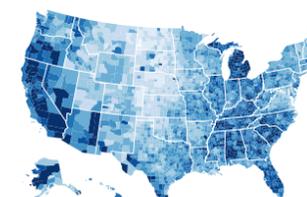
- Geometry
 - Geographic
 - Other Derived
- Spatial Fields
 - Scalar Fields (one value per cell)
 - Isocontours
 - Direct Volume Rendering
 - Vector and Tensor Fields (many values per cell)
 - Flow Glyphs (local)
 - Geometric (sparse seeds)
 - Textures (dense seeds)
 - Features (globally derived)



34

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://blacks.org/mbostock/4060606>

35

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

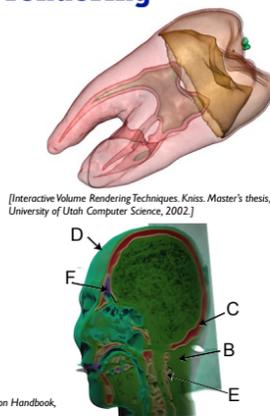


Land Information New Zealand Data Service

36

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



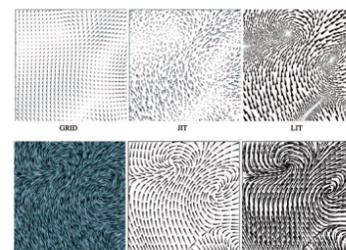
[Interactive Volume Rendering Techniques. Kniss. Master's thesis, University of Utah Computer Science, 2002.]

[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.]

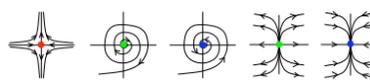
37

Idioms: vector glyphs

- tasks
 - finding critical points, identifying their types
 - identifying what type of critical point is at a specific location
 - predicting where a particle starting at a specified point will end up (advection)



[Comparing 2D vector field visualization methods: A user study. Laidlaw et al. IEEE Trans. Visualization and Computer Graphics (TVCG) 11:1 (2005), 59-70.]

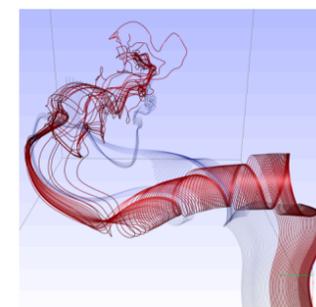


[Topology tracking for the visualization of time-dependent two-dimensional flows. Tricache, Wischgoll, Scheuermann, and Hagen. Computers & Graphics 26:2 (2002), 249-257.]

38

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. McLaughlin, Jones, Laramée, Malki, Masters, and Hansen. IEEE Trans. Visualization and Computer Graphics 19:8 (2013), 1342-1353.]

39

Arrange networks and trees

Node-Link Diagrams

Connection Marks

✓ NETWORKS ✓ TREES



Adjacency Matrix

Derived Table

✓ NETWORKS ✓ TREES



Enclosure

Containment Marks

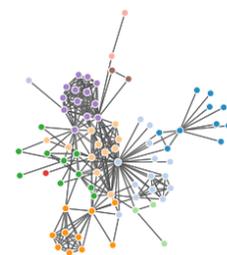
✗ NETWORKS ✓ TREES



40

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 $E < 4N$

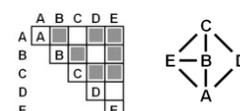


<http://mbostock.github.com/d3/ex/force.html>

41

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.]

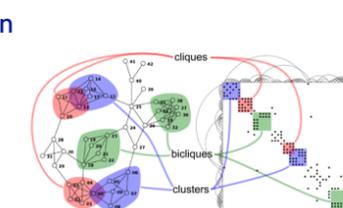


[Points of view: Networks. Gehlenborg and Wang. Nature Methods 9:115.]

42

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. Choniem, Fekete, and Castagliola. Information Visualization 4:2 (2005), 114-135.]

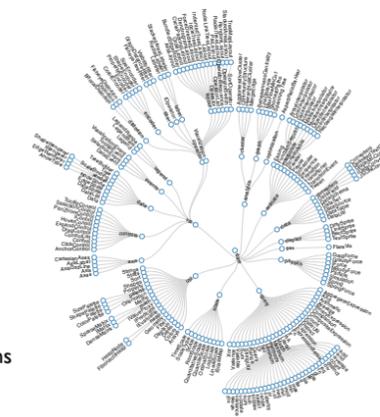


<http://www.michaelmcguffin.com/courses/vis/patternsInAdjacencyMatrix.png>

43

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

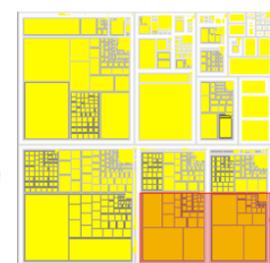


<http://mbostock.github.com/d3/ex/tree.html>

44

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

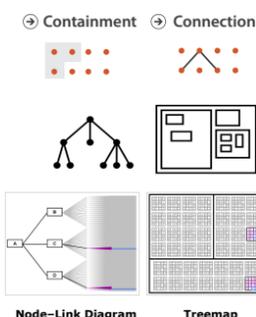


http://tulip.lsbri.fr/Documentation/3_7/userHandbook.html#ch06.html

45

Connection vs. containment comparison

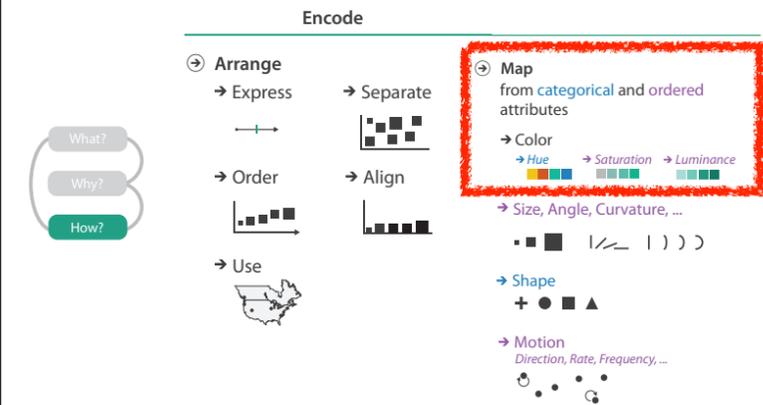
- marks as links (vs. nodes)
 - common case in network drawing
 - 1D case: connection
 - ex: all node-link diagrams
 - emphasizes topology, path tracing
 - networks and trees
 - 2D case: containment
 - ex: all treemap variants
 - emphasizes attribute values at leaves (size coding)
 - only trees



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

46

How to encode: Mapping color



Color: Luminance, saturation, hue

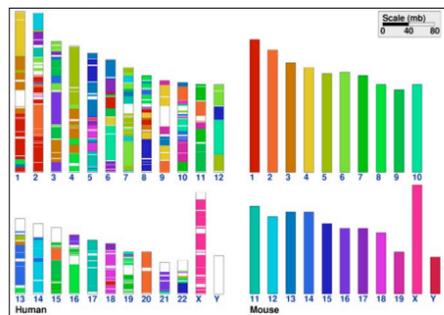
- 3 channels
 - identity for categorical
 - hue
 - magnitude for ordered
 - luminance
 - saturation
- better match for visual encoding than RGB color space from graphics



48

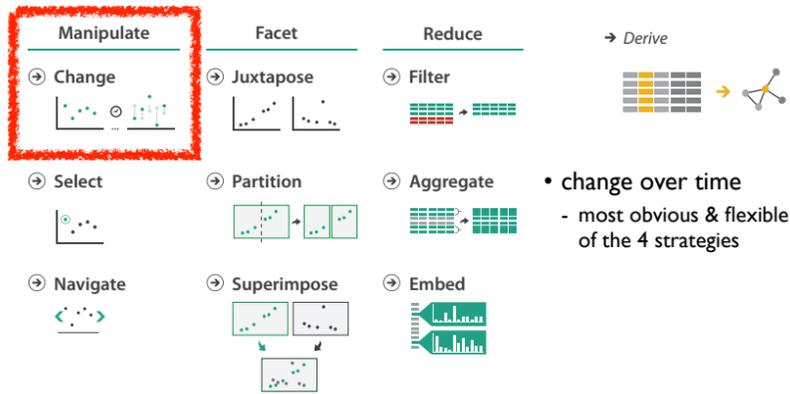
Categorical color: Discriminability constraints

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



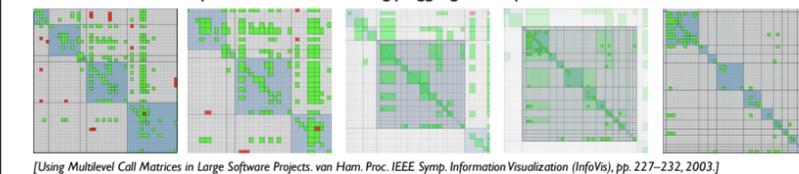
[Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. Sinha and Meller. BMC Bioinformatics, 8:82, 2007.]

How to handle complexity: 3 more strategies + 1 previous



Idiom: **Animated transitions**

- smooth transition from one state to another
 - alternative to jump cuts
 - support for item tracking when amount of change is limited
- example: multilevel matrix views
 - scope of what is shown narrows down
 - middle block stretches to fill space, additional structure appears within
 - other blocks squish down to increasingly aggregated representations

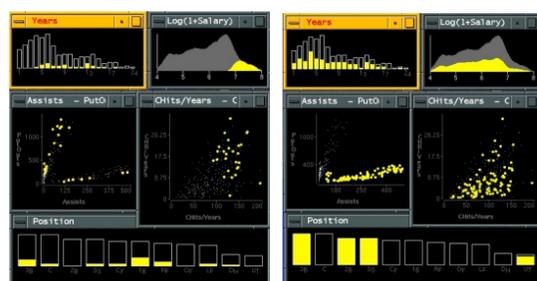


[Using Multilevel Call Matrices in Large Software Projects. van Ham. Proc. IEEE Symp. Information Visualization (InfoVis), pp. 227-232, 2003.]

Idiom: **Linked highlighting**

System: **EDV**

- see how regions contiguous in one view are distributed within another
- powerful and pervasive interaction idiom
- encoding: different
- **multiform**
- data: all shared



[Visual Exploration of Large Structured Datasets. Wills. Proc. New Techniques and Trends in Statistics (NTTS), pp. 237-246. IOS Press, 1995.]

Idiom: **bird's-eye maps**

System: **Google Maps**

- encoding: same
- data: subset shared
- navigation: shared
- bidirectional linking
- differences
- viewpoint
- (size)
- **overview-detail**

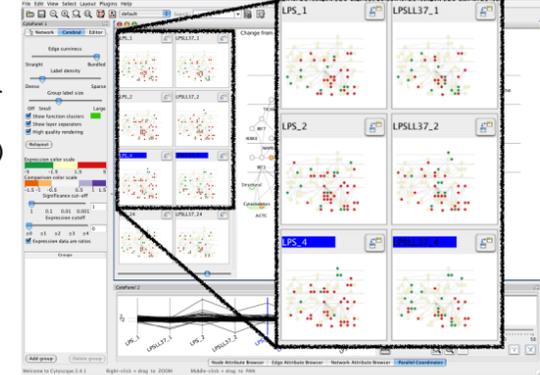


[A Review of Overview+Detail, Zooming, and Focus+Context Interfaces. Cockburn, Karlson, and Bederson. ACM Computing Surveys 41:1 (2008), 1-31.]

Idiom: **Small multiples**

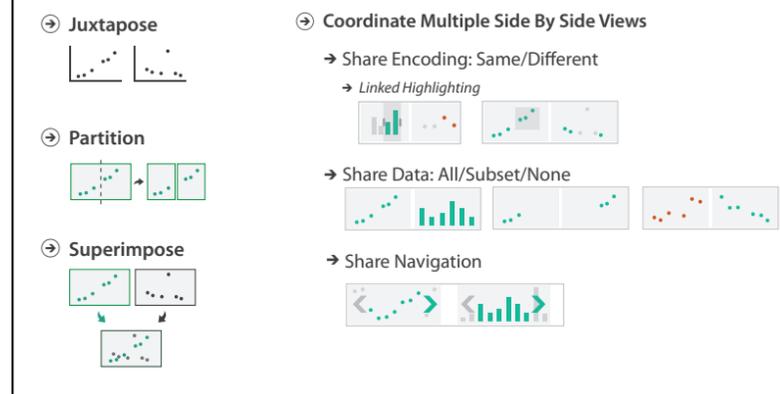
System: **Cerebral**

- encoding: same
- data: none shared
- different attributes for node colors
- (same network layout)
- navigation: shared



[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.]

Facet



Partition into views

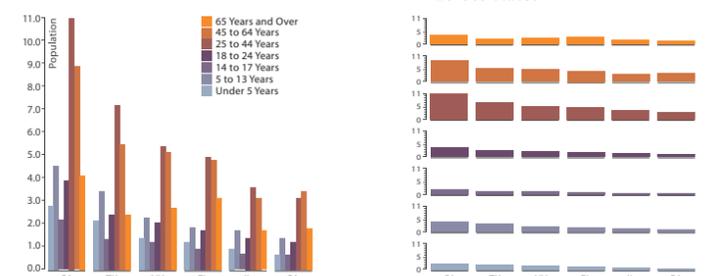
- how to divide data between views
- encodes association between items using spatial proximity
- major implications for what patterns are visible
- split according to attributes
- design choices
- how many splits
- all the way down: one mark per region?
- stop earlier, for more complex structure within region?
- order in which attribs used to split
- how many views

Partition into Side-by-Side Views



Partitioning: List alignment

- single bar chart with grouped bars
- split by state into regions
- complex glyph within each region showing all ages
- compare: easy within state, hard across ages



57

Partitioning: Recursive subdivision

System: **HIVE**

- split by type
- then by neighborhood
- then time
- years as rows
- months as columns

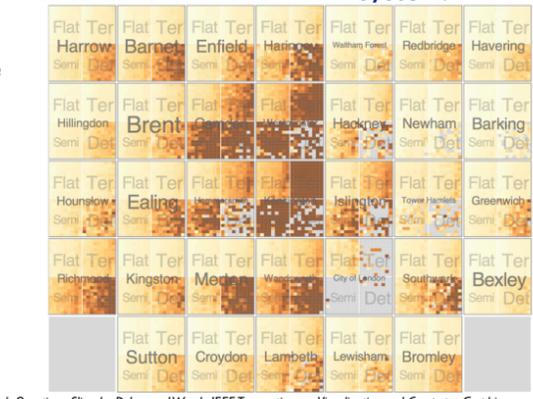


[Configuring Hierarchical Layouts to Address Research Questions. Slingsby, Dykes, and Wood. IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2009) 15:6 (2009), 977-984.]

Partitioning: Recursive subdivision

System: **HIVE**

- switch order of splits
- neighborhood then type
- very different patterns



[Configuring Hierarchical Layouts to Address Research Questions. Slingsby, Dykes, and Wood. IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2009) 15:6 (2009), 977-984.]

Partitioning: Recursive subdivision

System: **HIVE**

- size regions by sale counts
- not uniformly
- result: treemap

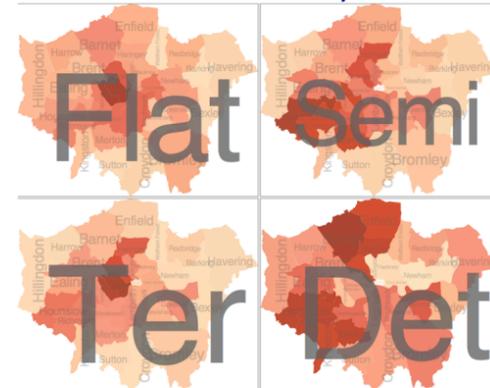


[Configuring Hierarchical Layouts to Address Research Questions. Slingsby, Dykes, and Wood. IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2009) 15:6 (2009), 977-984.]

Partitioning: Recursive subdivision

System: **HIVE**

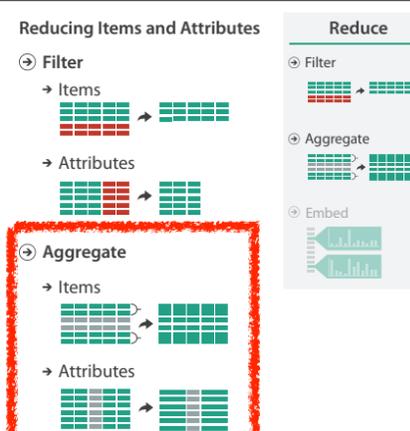
- different encoding for second-level regions
- choropleth maps



[Configuring Hierarchical Layouts to Address Research Questions. Slingsby, Dykes, and Wood. IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2009) 15:6 (2009), 977-984.]

Reduce items and attributes

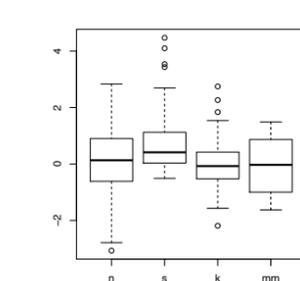
- reduce/increase: inverses
- filter
- pro: straightforward and intuitive
- to understand and compute
- con: out of sight, out of mind
- aggregation
- pro: inform about whole set
- con: difficult to avoid losing signal
- not mutually exclusive
- combine filter, aggregate
- combine reduce, facet, change, derive



63

Idiom: **boxplot**

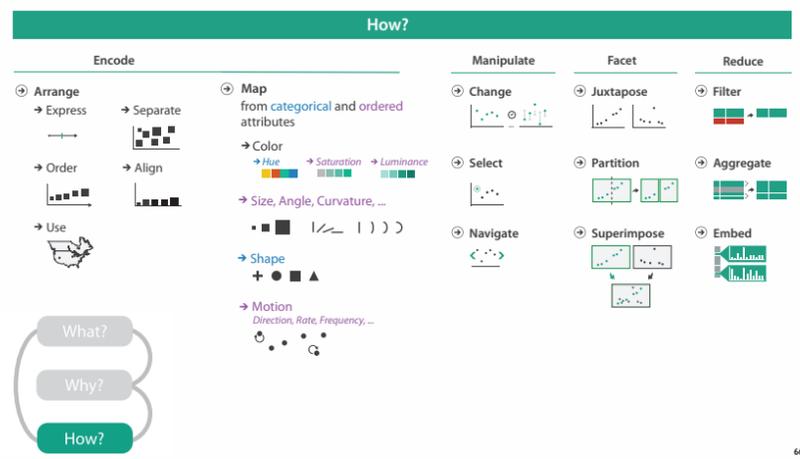
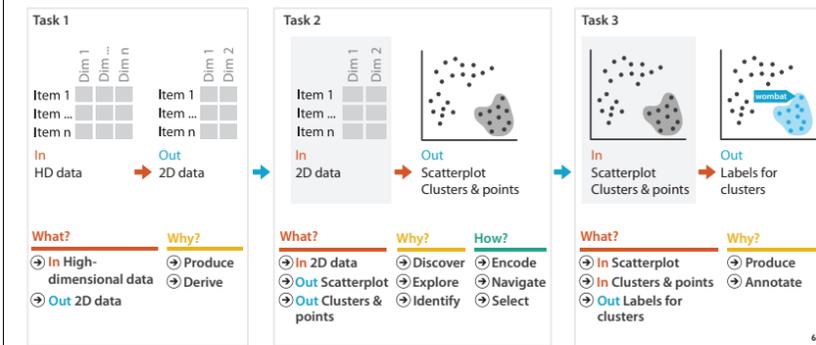
- static item aggregation
- task: find distribution
- data: table
- derived data
- 5 quant attribs
- median: central line
- lower and upper quartile: boxes
- lower upper fences: whiskers
- values beyond which items are outliers
- outliers beyond fence cutoffs explicitly shown



[40 years of boxplots. Wickham and Stryjewski. 2012. had.co.nz]

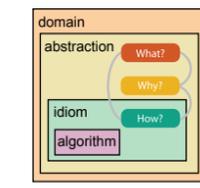
Idiom: Dimensionality reduction for documents

- attribute aggregation
 - derive low-dimensional target space from high-dimensional measured space



Analysis with 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**
 - **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



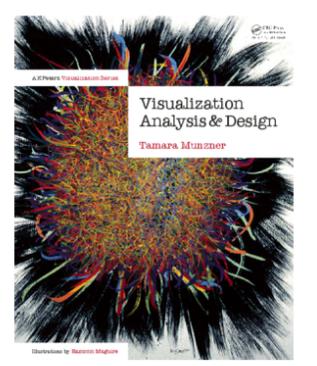
Choosing appropriate validation methods for each level



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

More Information

- this talk
 - <http://www.cs.ubc.ca/~tmm/talks.html#vad15london>
- papers, videos, software, talks, full courses
 - <http://www.cs.ubc.ca/group/infovis>
 - <http://www.cs.ubc.ca/~tmm>
- book (including tutorial lecture slides)
 - <http://www.cs.ubc.ca/~tmm/vadbook>
- acknowledgements
 - illustrations: Eamonn Maguire



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