Ch 7+8: Tables, Spatial Data

Tamara Munzner
Department of Computer Science
University of British Columbia

CPSC 547, Information Visualization
Day 8: 6 October 2015

http://www.cs.ubc.ca/~tmm/courses/547-15

News

- clarification on artery vis
  - diverging colormap since doctors care about high and low values
    - not much about the ones in the middle
  - personal communication with Borkin, not clearly stated in paper
- second guest lecture today from Kosara
  - vis for presentation (versus discovery/exploration)
- then continue with lecture/discussion
  - catch up on chapters, leave papers for Thu
- remember
  - I have office hours on Tuesdays
  - pitches are coming up Thu Oct 22
  - start talking to me about project ideas!

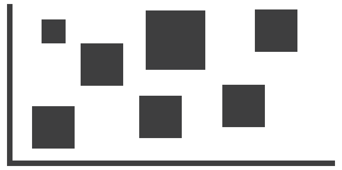
VAD Ch 7: Arrange Tables

Encode

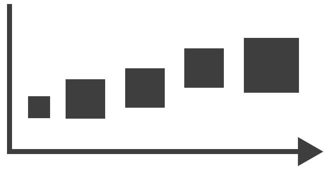
➔ **Arrange**

- ➔ Express

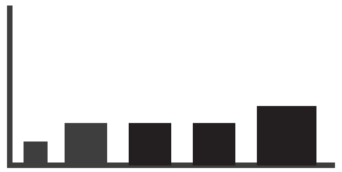
- ➔ Separate



- ➔ Order



- ➔ Align



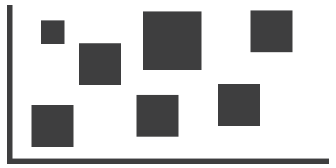
- ➔ Use

Arrange tables

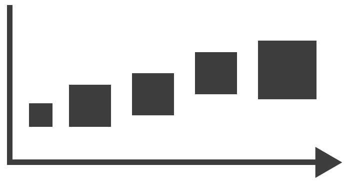
➔ Express Values

➔ Separate, Order, Align Regions

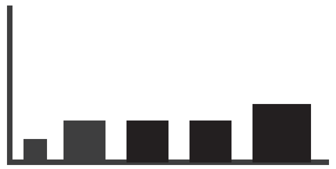
→ Separate



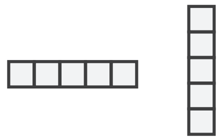
→ Order



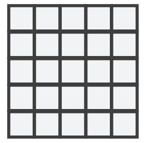
→ Align



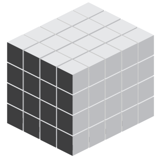
→ *1 Key*
List



→ *2 Keys*
Matrix



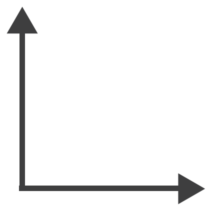
→ *3 Keys*
Volume



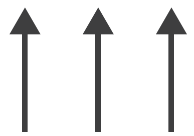
→ *Many Keys*
Recursive Subdivision

➔ Axis Orientation

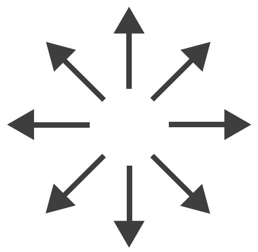
→ Rectilinear



→ Parallel

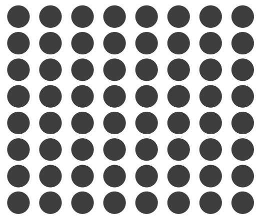


→ Radial



➔ Layout Density

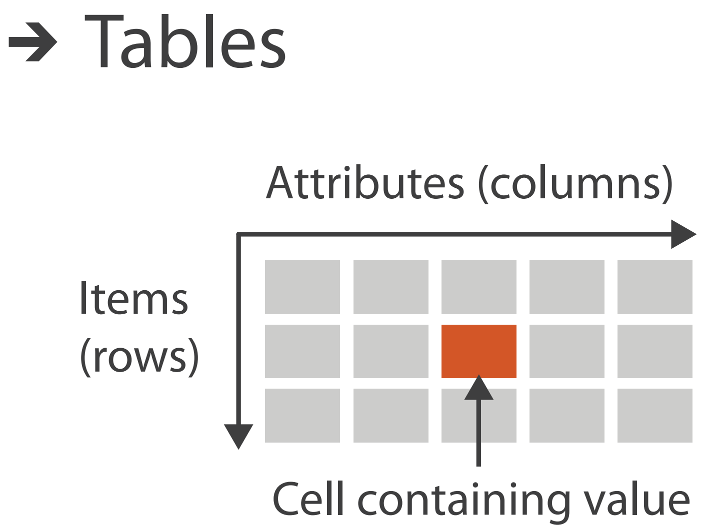
→ Dense

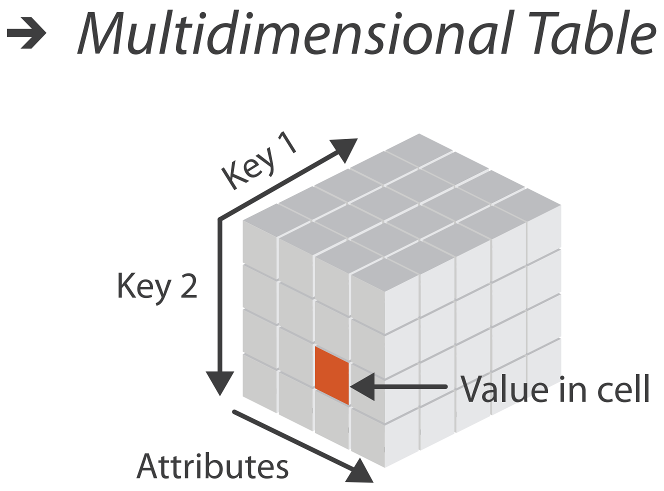


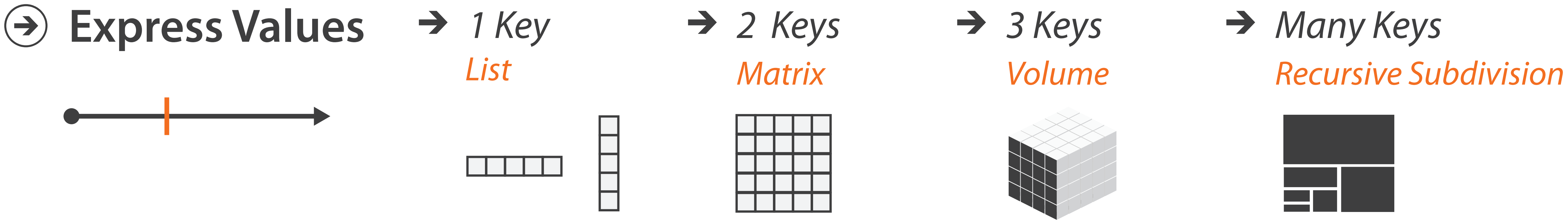
→ Space-Filling

Keys and values

- key
  - independent attribute
  - used as unique index to look up items
  - simple tables: 1 key
  - multidimensional tables: multiple keys
- value
  - dependent attribute, value of cell
- classify arrangements by key count
  - 0, 1, 2, many...

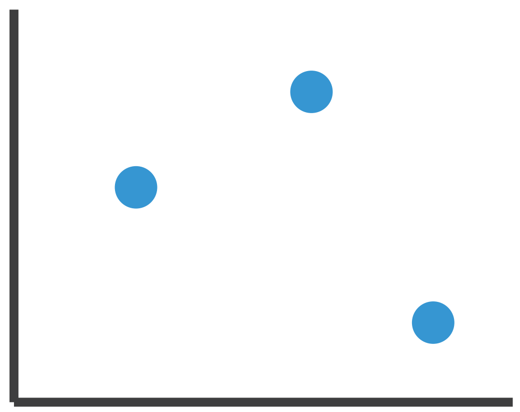




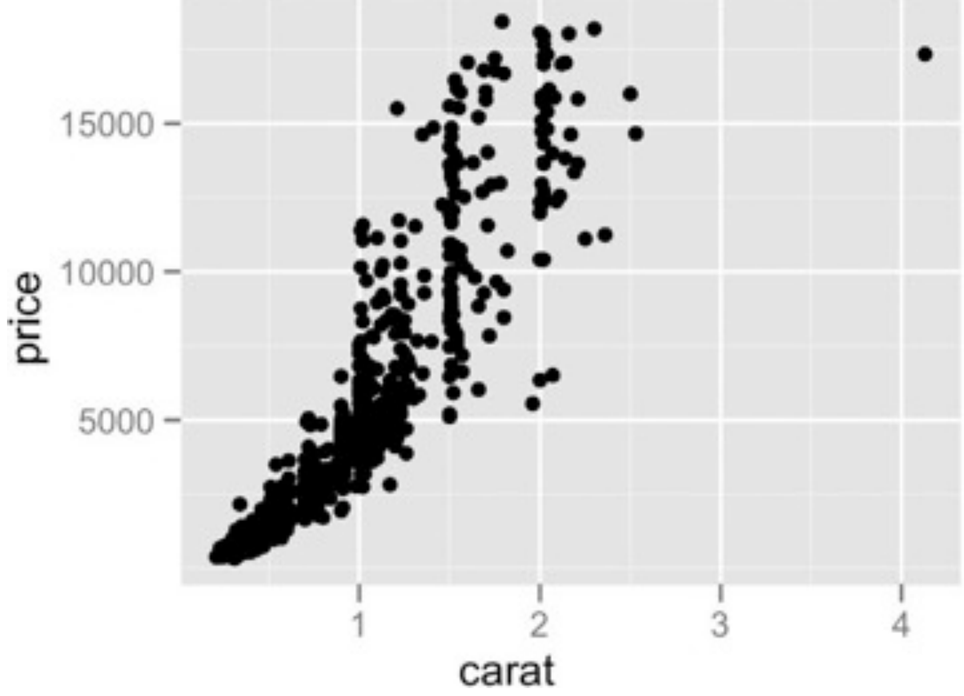


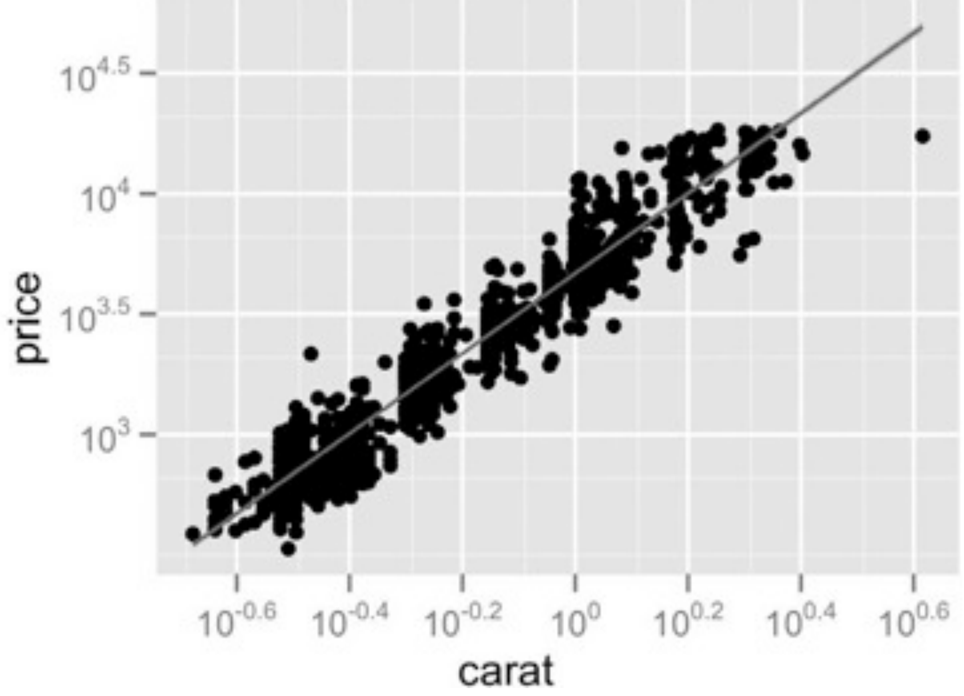
Idiom: **scatterplot**

➔ Express Values



- ***express*** values
  - quantitative attributes
- no keys, only values
  - data
    - 2 quant attribs
  - mark: points
  - channels
    - horiz + vert position
  - tasks
    - find trends, outliers, distribution, correlation, clusters
  - scalability
    - hundreds of items



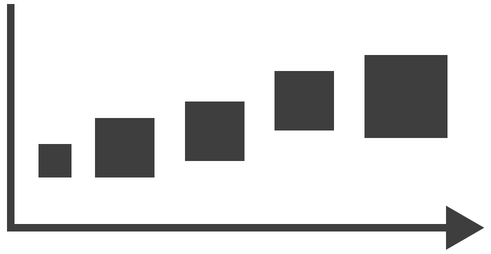


[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]

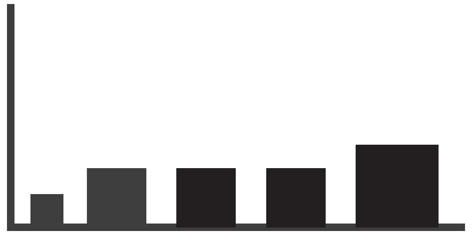
Some keys: Categorical regions

➔ Separate

➔ Order

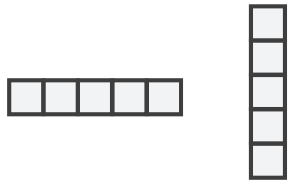


➔ Align

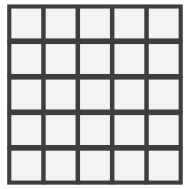


- **regions**: contiguous bounded areas distinct from each other
  - using space to ***separate*** (proximity)
  - following expressiveness principle for categorical attributes
- use ordered attribute to ***order*** and ***align*** regions

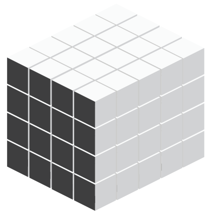
➔ *1 Key*
List



➔ *2 Keys*
Matrix



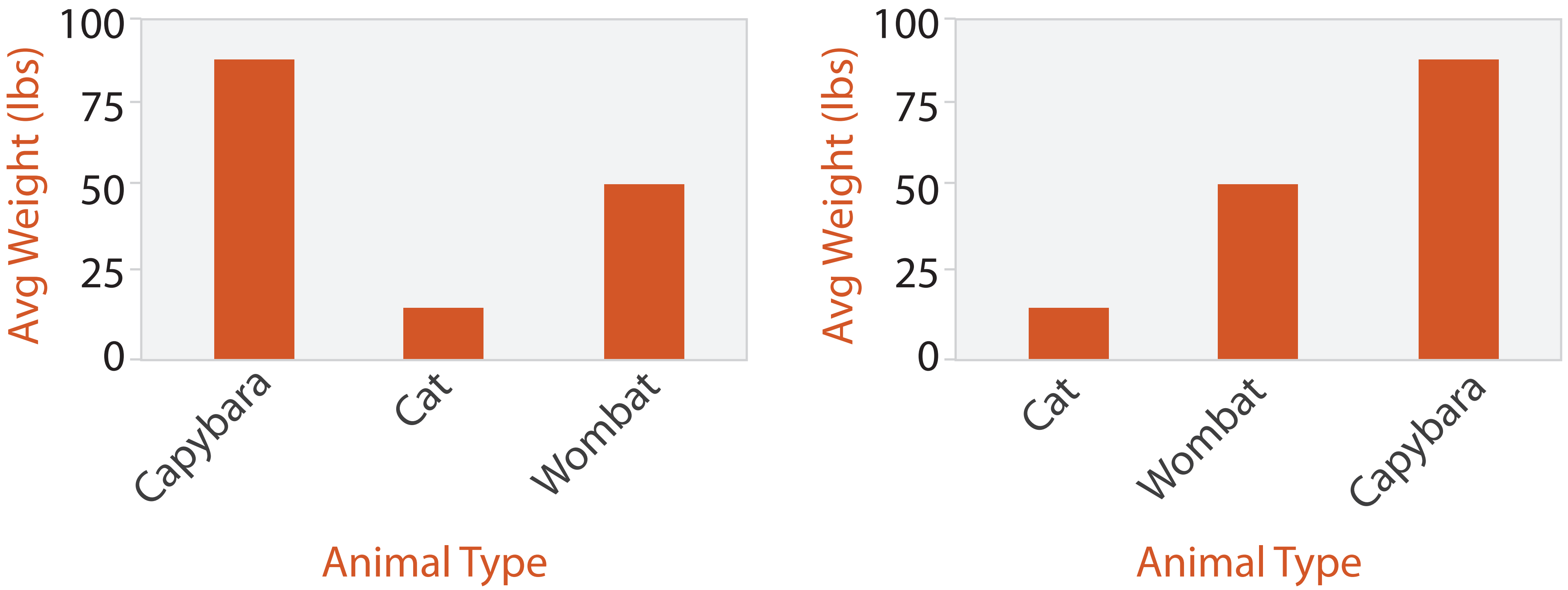
➔ *3 Keys*
Volume



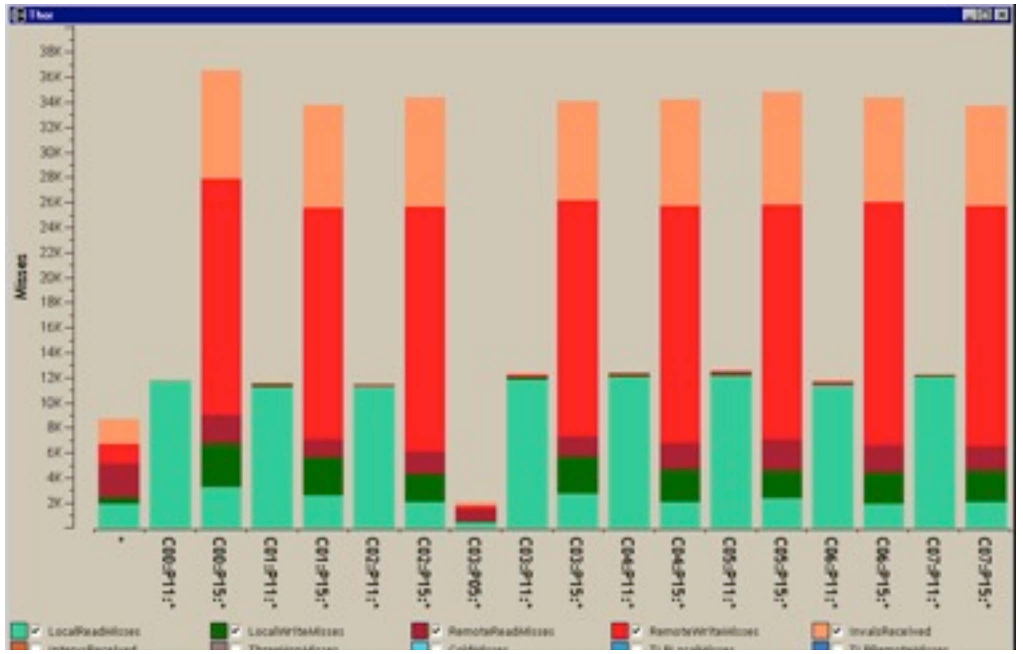
➔ *Many Keys*
Recursive Subdivision

Idiom: **bar chart**

- one key, one value
  - data
    - 1 categ attrib, 1 quant attrib
  - mark: lines
  - channels
    - length to express quant value
    - spatial regions: one per mark
      - separated horizontally, aligned vertically
      - ordered by quant attrib
        - » by label (alphabetical), by length attrib (data-driven)
  - task
    - compare, lookup values
  - scalability
    - dozens to hundreds of levels for key attrib



Idiom: **stacked bar chart**

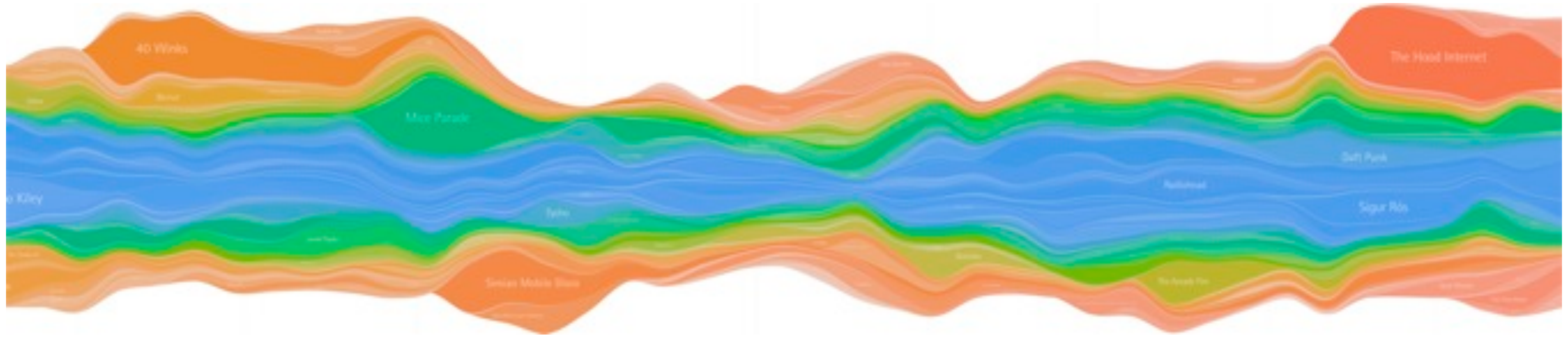


- one more key
  - data
    - 2 categ attrib, 1 quant attrib
  - mark: vertical stack of line marks
    - **glyph**: composite object, internal structure from multiple marks
  - channels
    - length and color hue
    - spatial regions: one per glyph
      - aligned: full glyph, lowest bar component
      - unaligned: other bar components
  - task
    - part-to-whole relationship
  - scalability
    - several to one dozen levels for stacked attrib

[Using Visualization to Understand the Behavior of Computer Systems. Bosch. Ph.D. thesis, Stanford Computer Science, 2001.]

Idiom: **streamgraph**

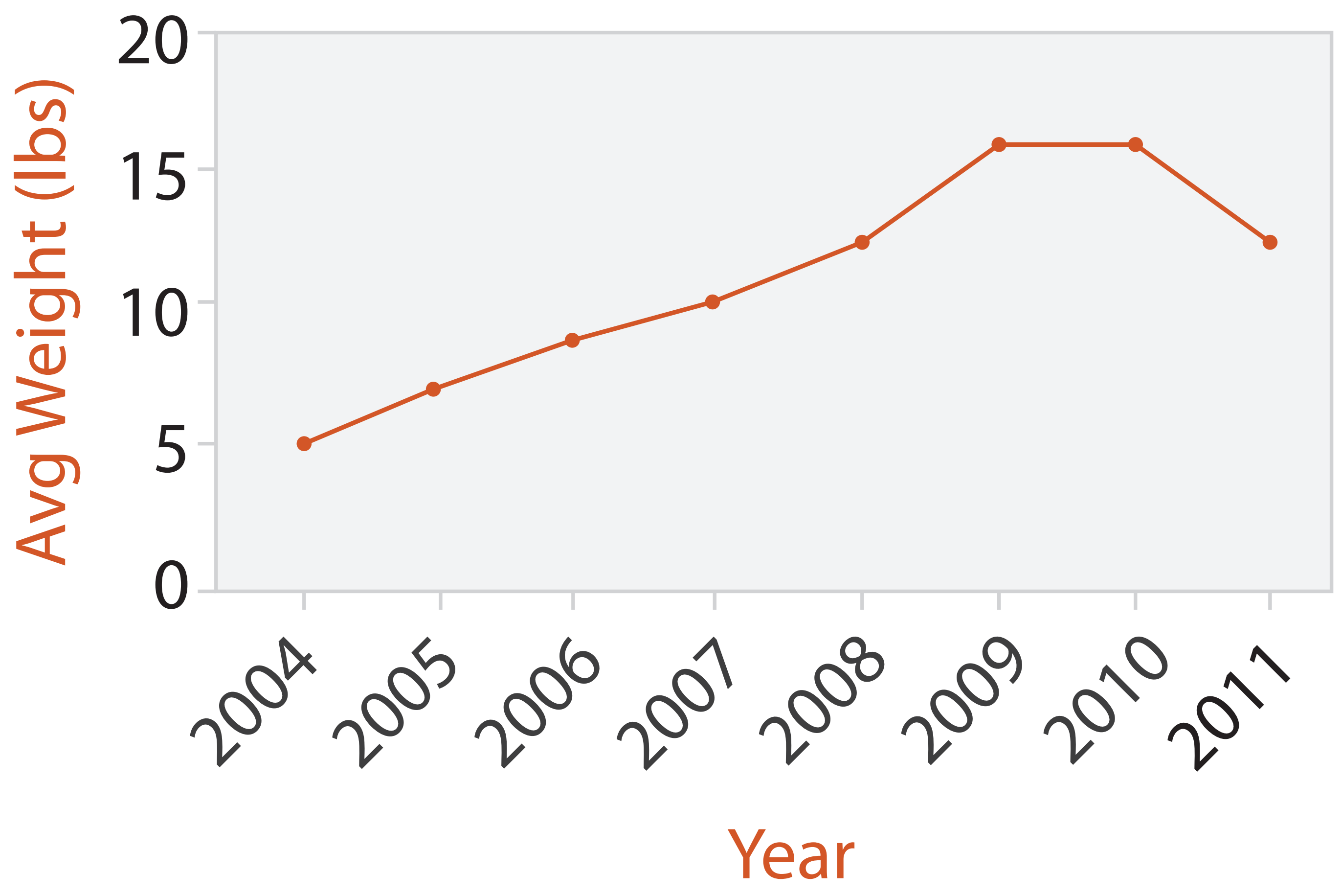
- generalized stacked graph
  - emphasizing horizontal continuity
    - vs vertical items
  - data
    - 1 categ key attrib (artist)
    - 1 ordered key attrib (time)
    - 1 quant value attrib (counts)
  - derived data
    - geometry: layers, where height encodes counts
    - 1 quant attrib (layer ordering)
  - scalability
    - hundreds of time keys
    - dozens to hundreds of artist keys
      - more than stacked bars, since most layers don't extend across whole chart



[Stacked Graphs Geometry & Aesthetics. Byron and Wattenberg. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2008) 14(6): 1245–1252, (2008).]

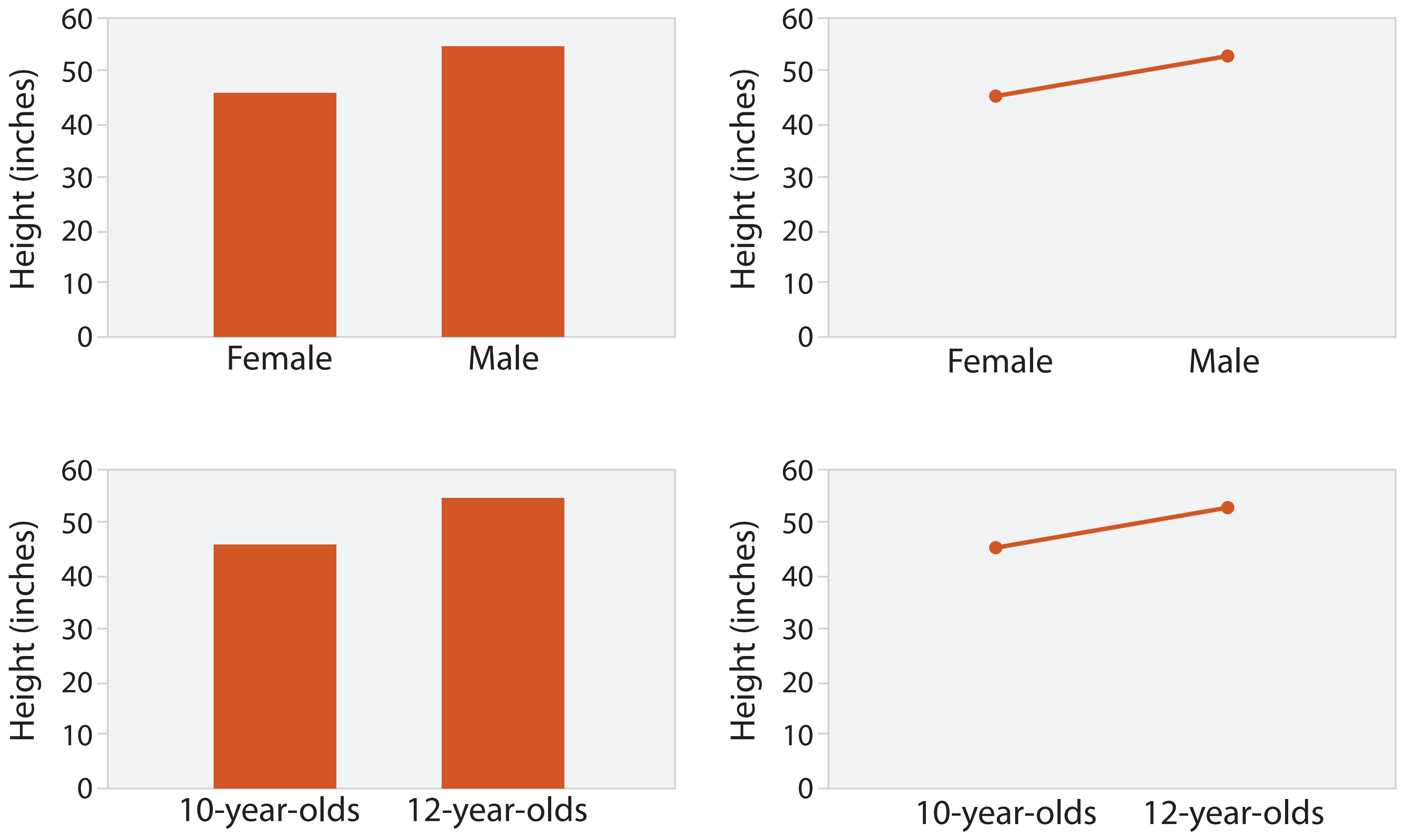
Idiom: **line chart**

- one key, one value
  - data
    - 2 quant attribs
  - mark: points
    - 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



Choosing bar vs line charts

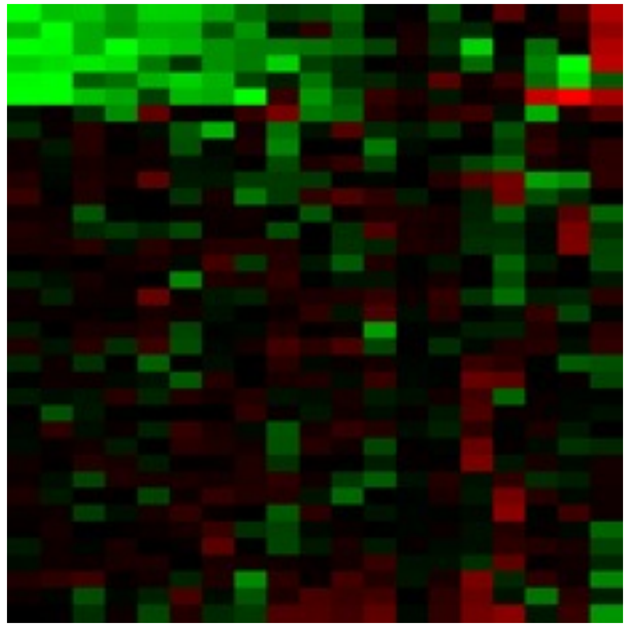
- depends on type of key attrib
  - bar charts if categorical
  - line charts if ordered
- do not use line charts for categorical key attribs
  - violates expressiveness principle
    - implication of trend so strong that it overrides semantics!
      - "The more male a person is, the taller he/she is"

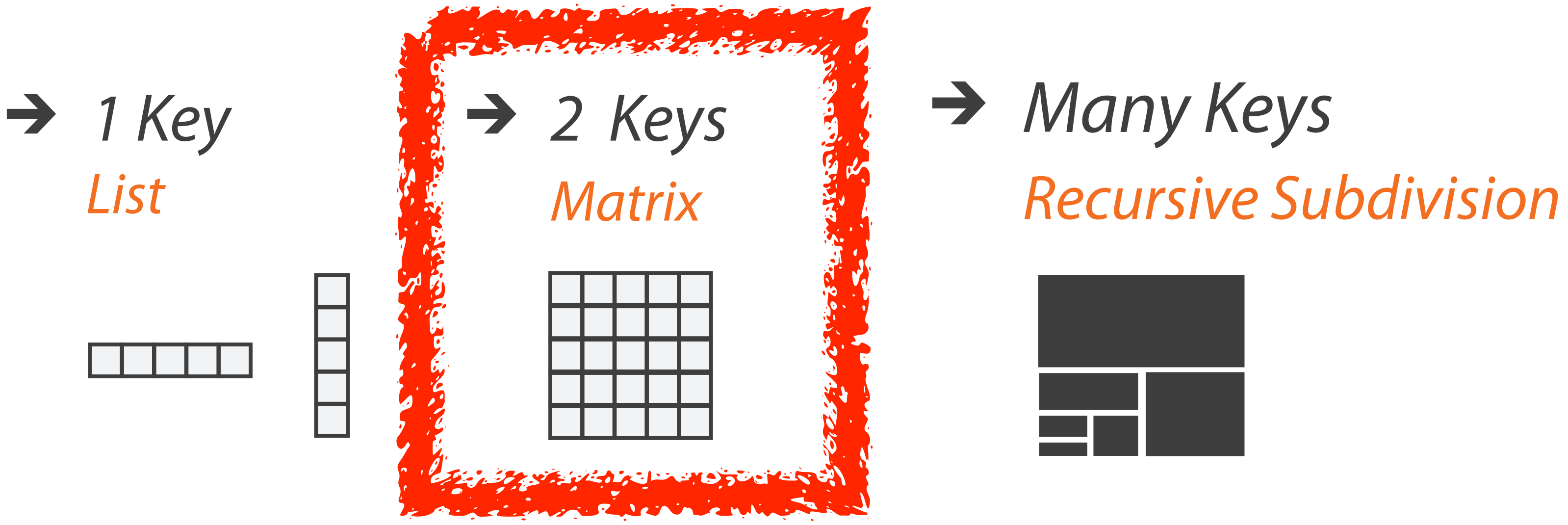


after [Bars and Lines: A Study of Graphic Communication. Zacks and Tversky. Memory and Cognition 27:6 (1999), 1073–1079.]

Idiom: **heatmap**

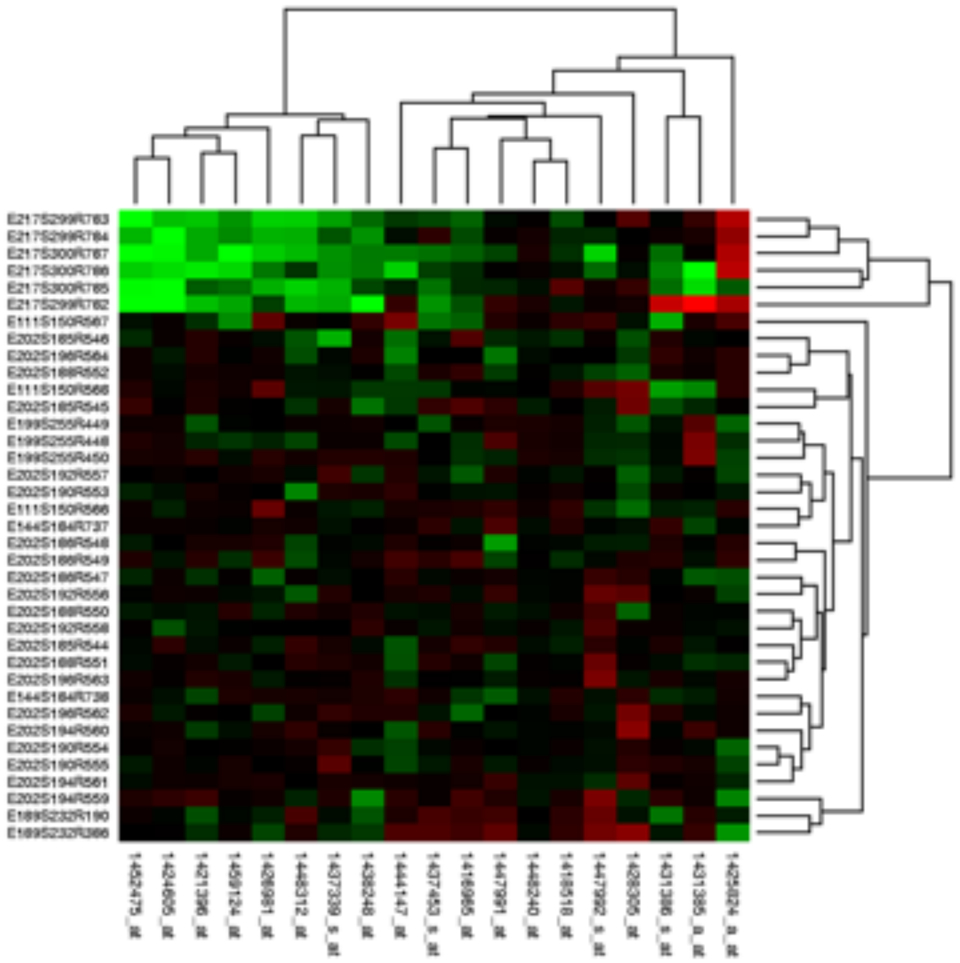
- two keys, one value
  - data
    - 2 categ attribs (gene, experimental condition)
    - 1 quant attrib (expression levels)
  - marks: area
    - separate and align in 2D matrix
      - indexed by 2 categorical attributes
  - channels
    - color by quant attrib
      - (ordered diverging colormap)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





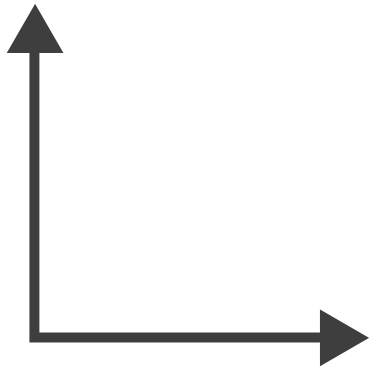
Idiom: **cluster heatmap**

- 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

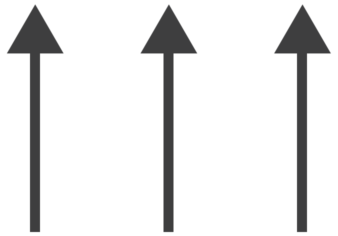


Axis Orientation

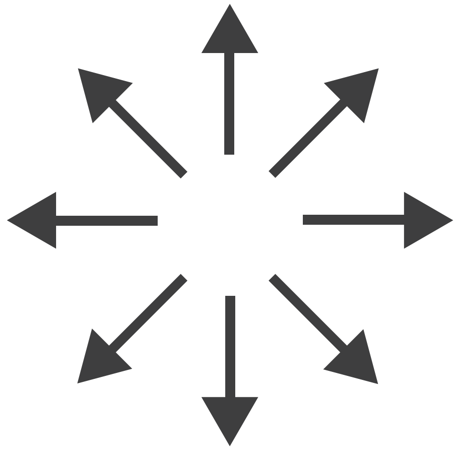
- Rectilinear



- Parallel

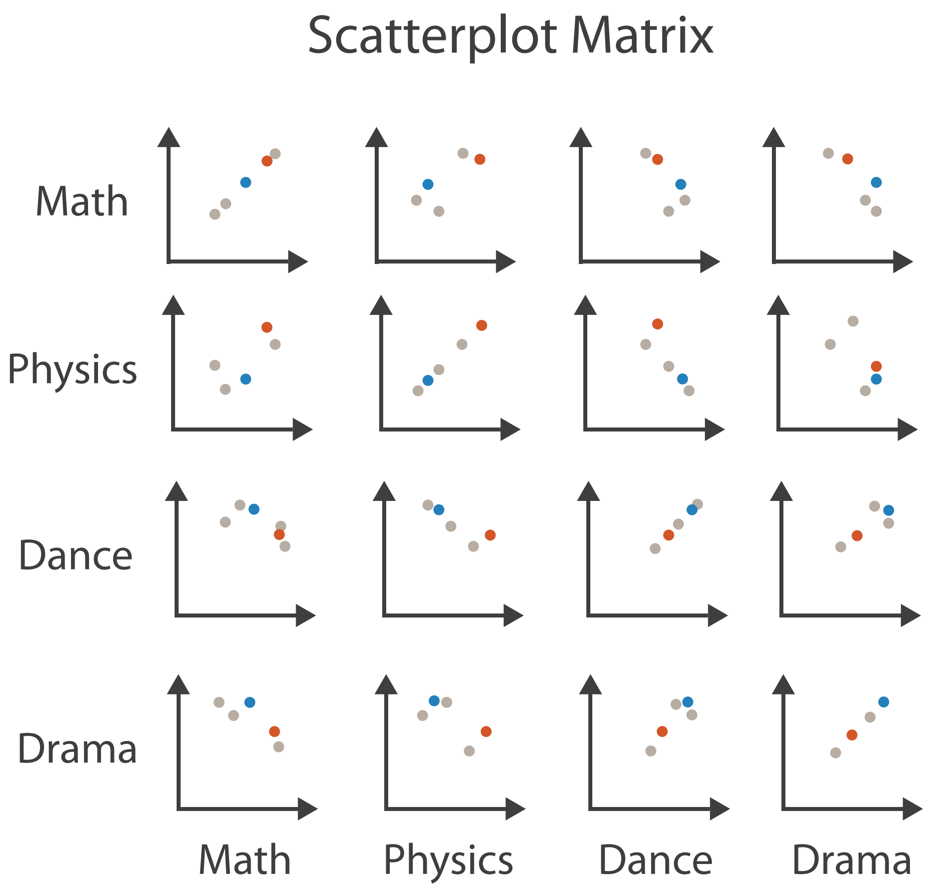


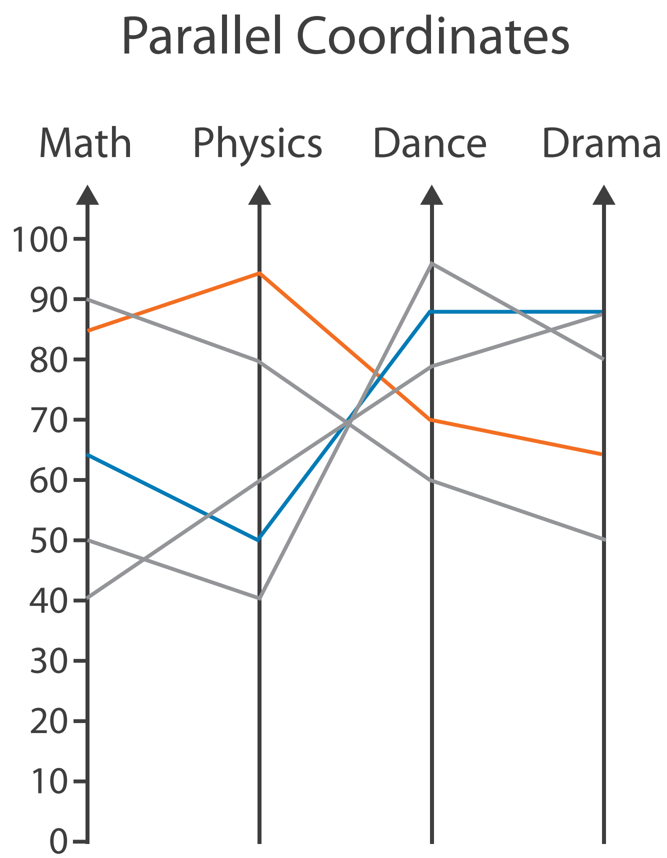
- Radial



Idioms: **scatterplot matrix, parallel coordinates**

- scatterplot matrix (SPLOM)
  - rectilinear axes, point mark
  - all possible pairs of axes
  - scalability
    - one dozen attribs
    - dozens to hundreds of items
- parallel coordinates
  - parallel axes, jagged line representing item
  - rectilinear axes, item as point
    - axis ordering is major challenge
  - scalability
    - dozens of attribs
    - hundreds of items





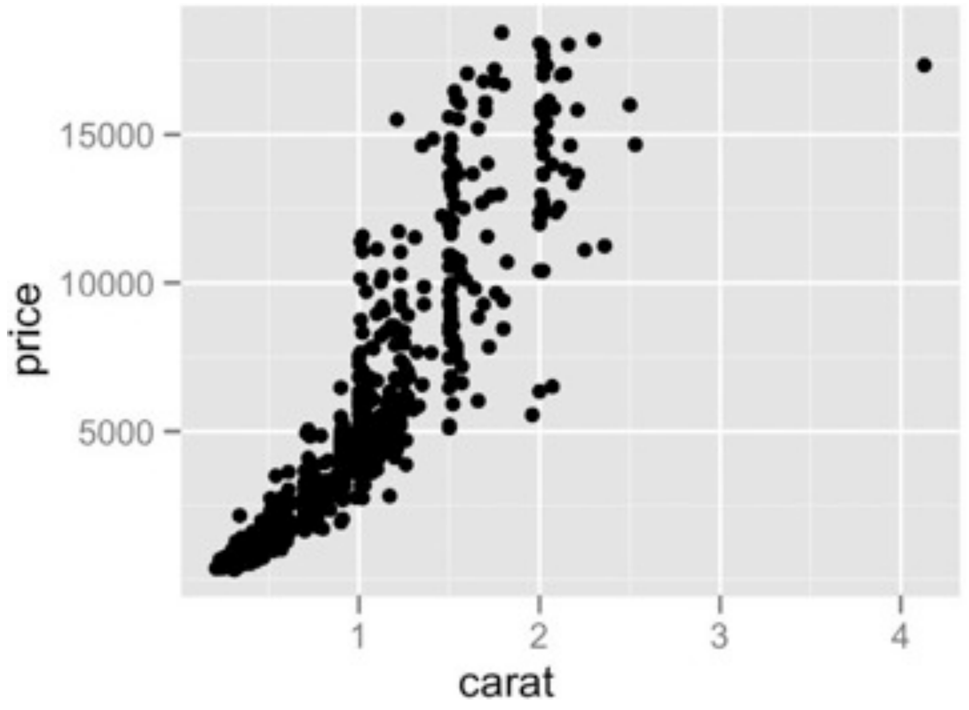
Table

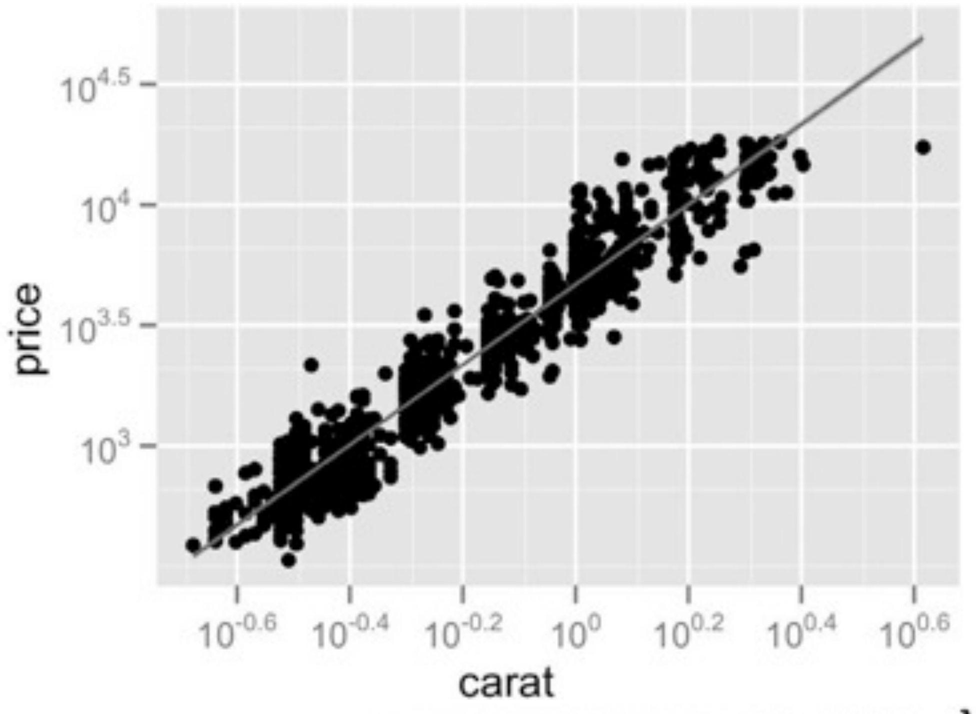
Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

after [Visualization Course Figures. McGuffin, 2014. http://www.michaelmcguffin.com/courses/vis/]

Task: Correlation

- scatterplot matrix
  - positive correlation
    - diagonal low-to-high
  - negative correlation
    - diagonal high-to-low
  - uncorrelated
- parallel coordinates
  - positive correlation
    - parallel line segments
  - negative correlation
    - all segments cross at halfway point
  - uncorrelated
    - scattered crossings





[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]

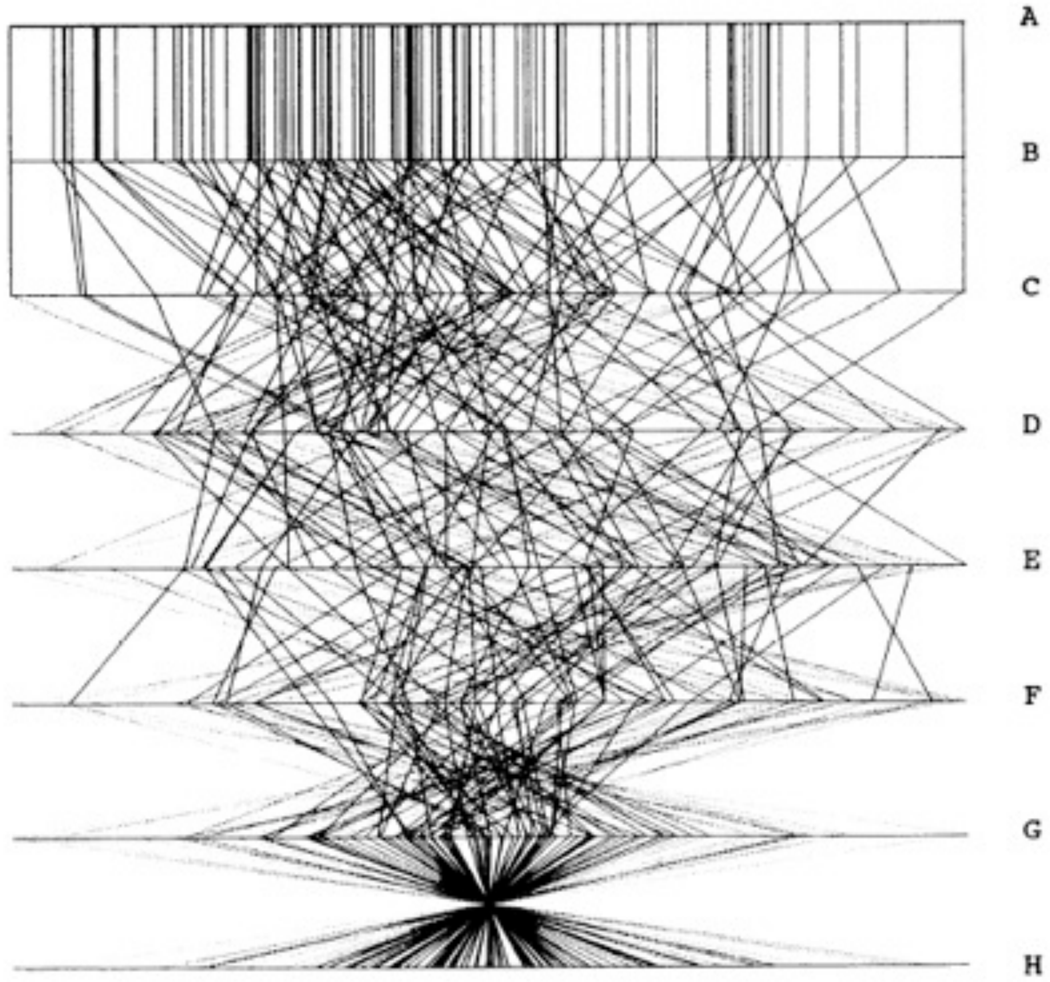
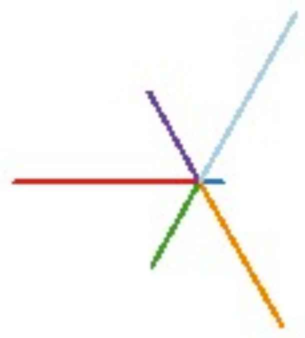


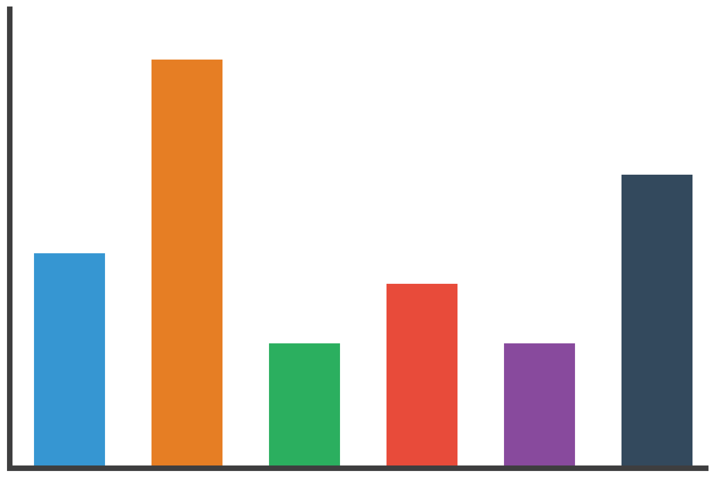
Figure 3. Parallel Coordinate Plot of Six-Dimensional Data Illustrating Correlations of ρ = 1, .8, .2, 0, −.2, −.8, and −1.

[Hyperdimensional Data Analysis Using Parallel Coordinates. Wegman. Journ. American Statistical Association 85:411 (1990), 664–675.]

Idioms: **radial bar chart, star plot**

- radial bar chart
  - radial axes meet at central ring, line mark
- star plot
  - radial axes, meet at central point, line mark
- bar chart
  - rectilinear axes, aligned vertically



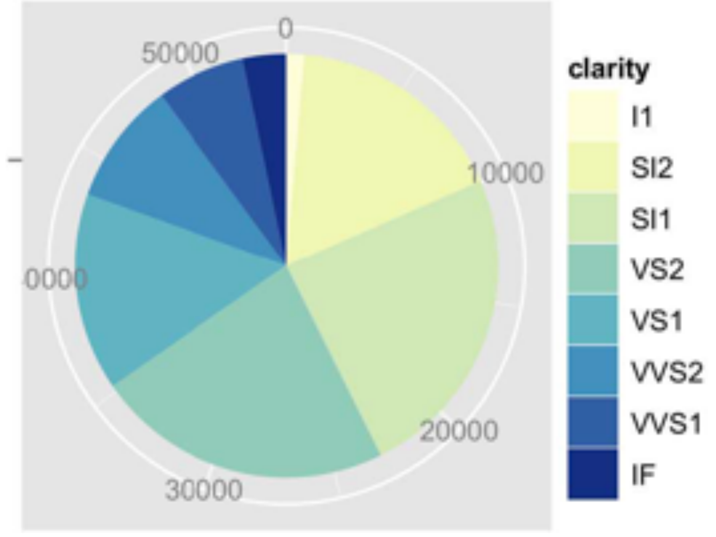


- accuracy
  - length unaligned with radial
    - less accurate than aligned with rectilinear

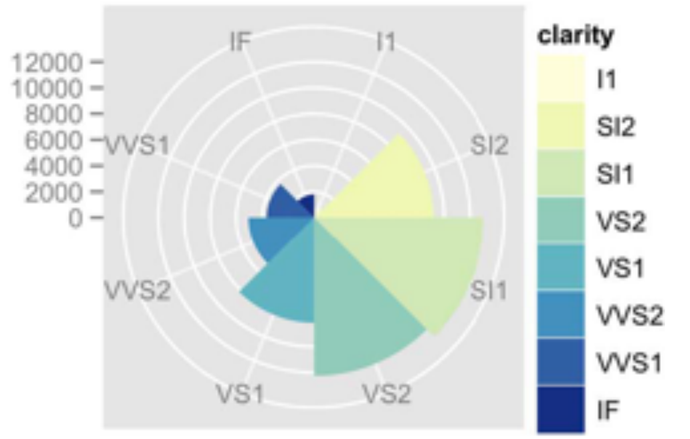
[Vismon: Facilitating Risk Assessment and Decision Making In Fisheries Management. Booshehrian, Möller, Peterman, and Munzner. Technical Report TR 2011-04, Simon Fraser University, School of Computing Science, 2011.]

Idioms: **pie chart, polar area chart**

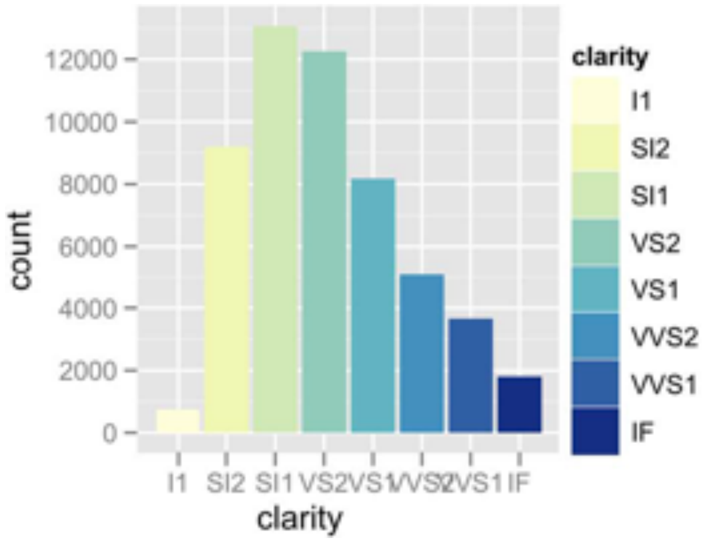
- pie chart
  - area marks with angle channel
  - accuracy: angle/area much less accurate than line length



- polar area chart
  - area marks with length channel
  - more direct analog to bar charts



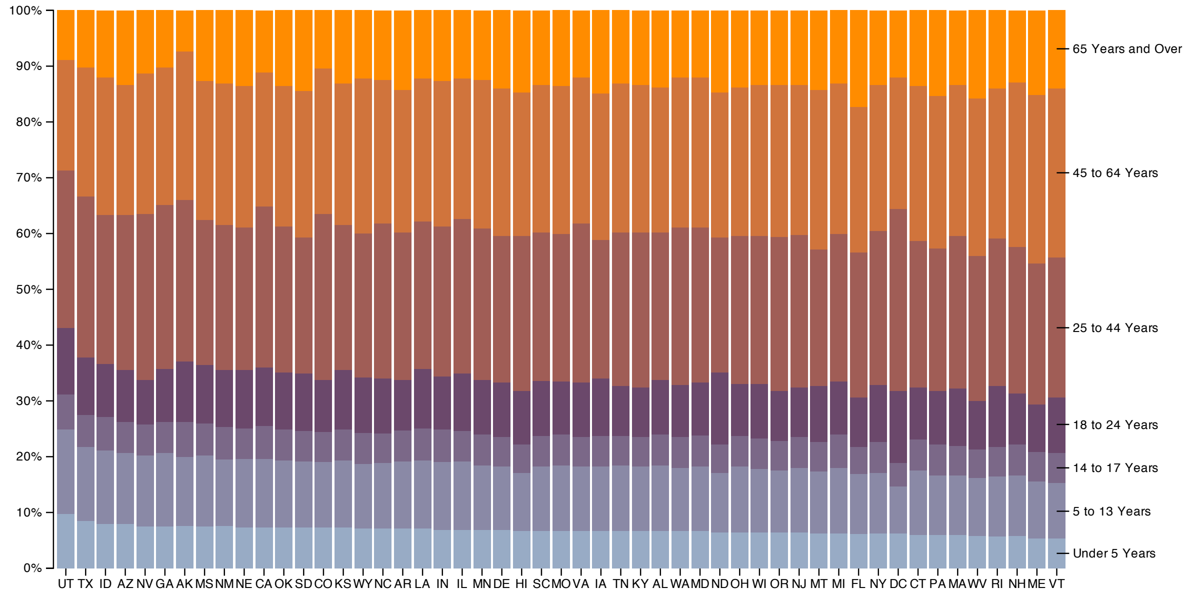
- data
  - 1 categ key attrib, 1 quant value attrib
- task
  - part-to-whole judgements

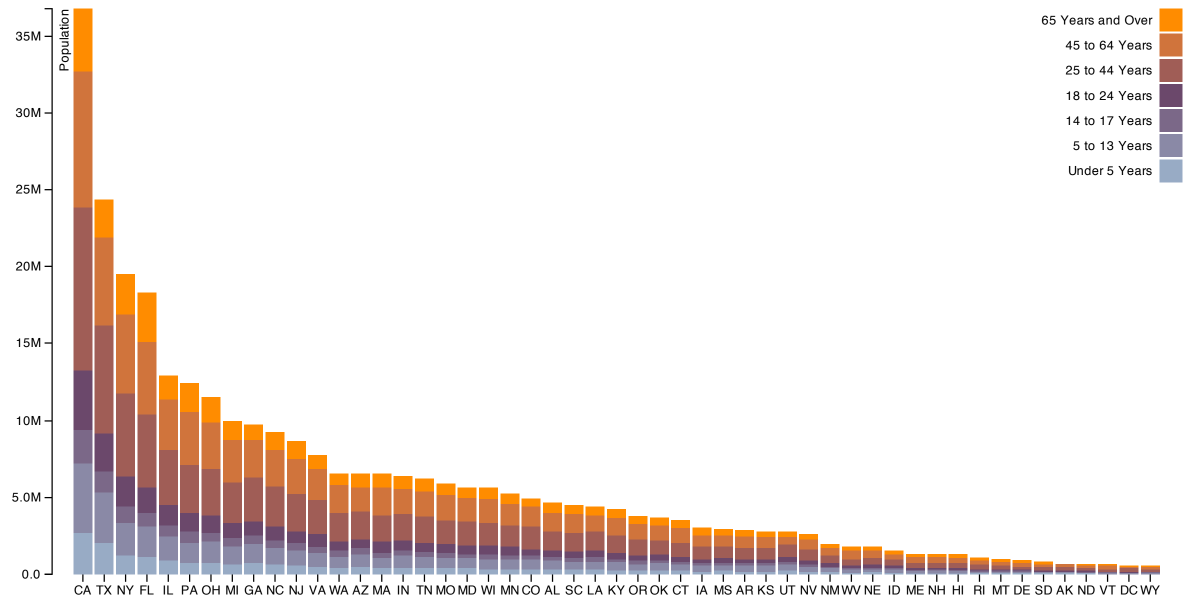


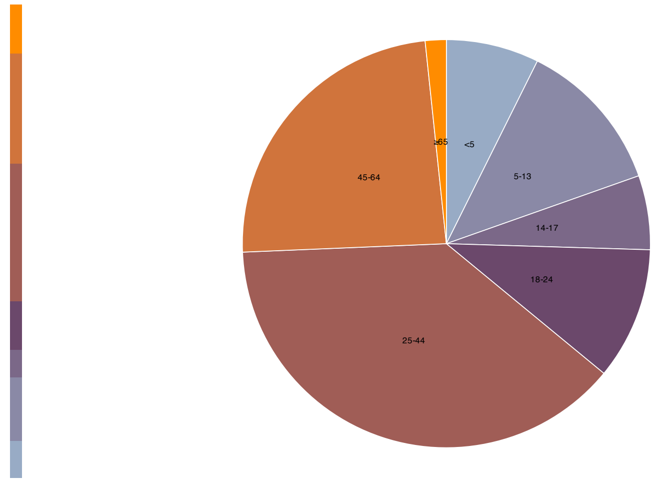
[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]

Idioms: **normalized stacked bar chart**

- task
  - part-to-whole judgements
- normalized stacked bar chart
  - stacked bar chart, normalized to full vert height
  - single stacked bar equivalent to full pie
    - high information density: requires narrow rectangle
- pie chart
  - information density: requires large circle



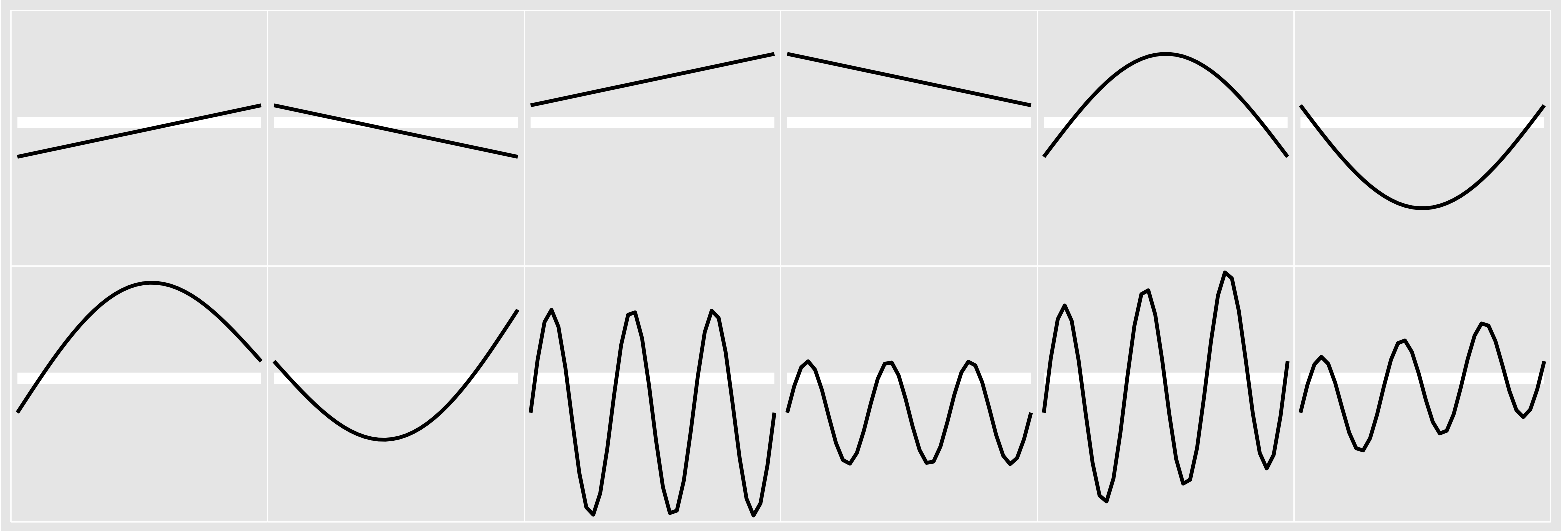




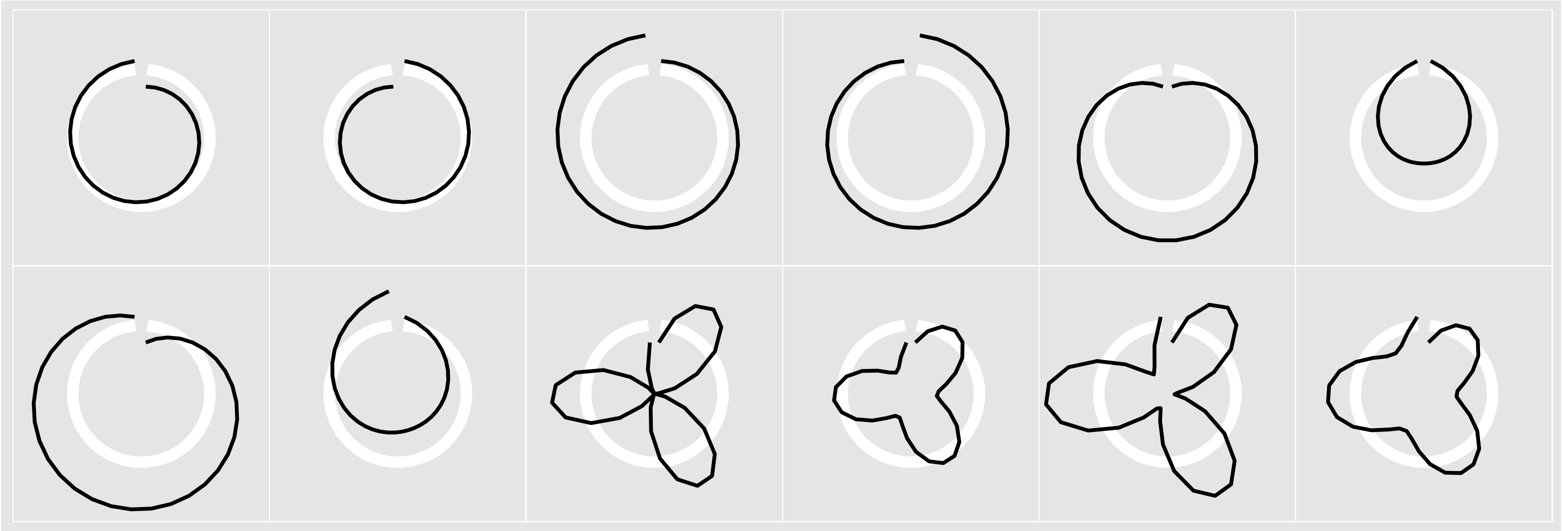
http://bl.ocks.org/mbostock/3887235,
http://bl.ocks.org/mbostock/3886208,
http://bl.ocks.org/mbostock/3886394.

Idiom: **glyphmaps**

- rectilinear good for linear vs nonlinear trends



- radial good for cyclic patterns

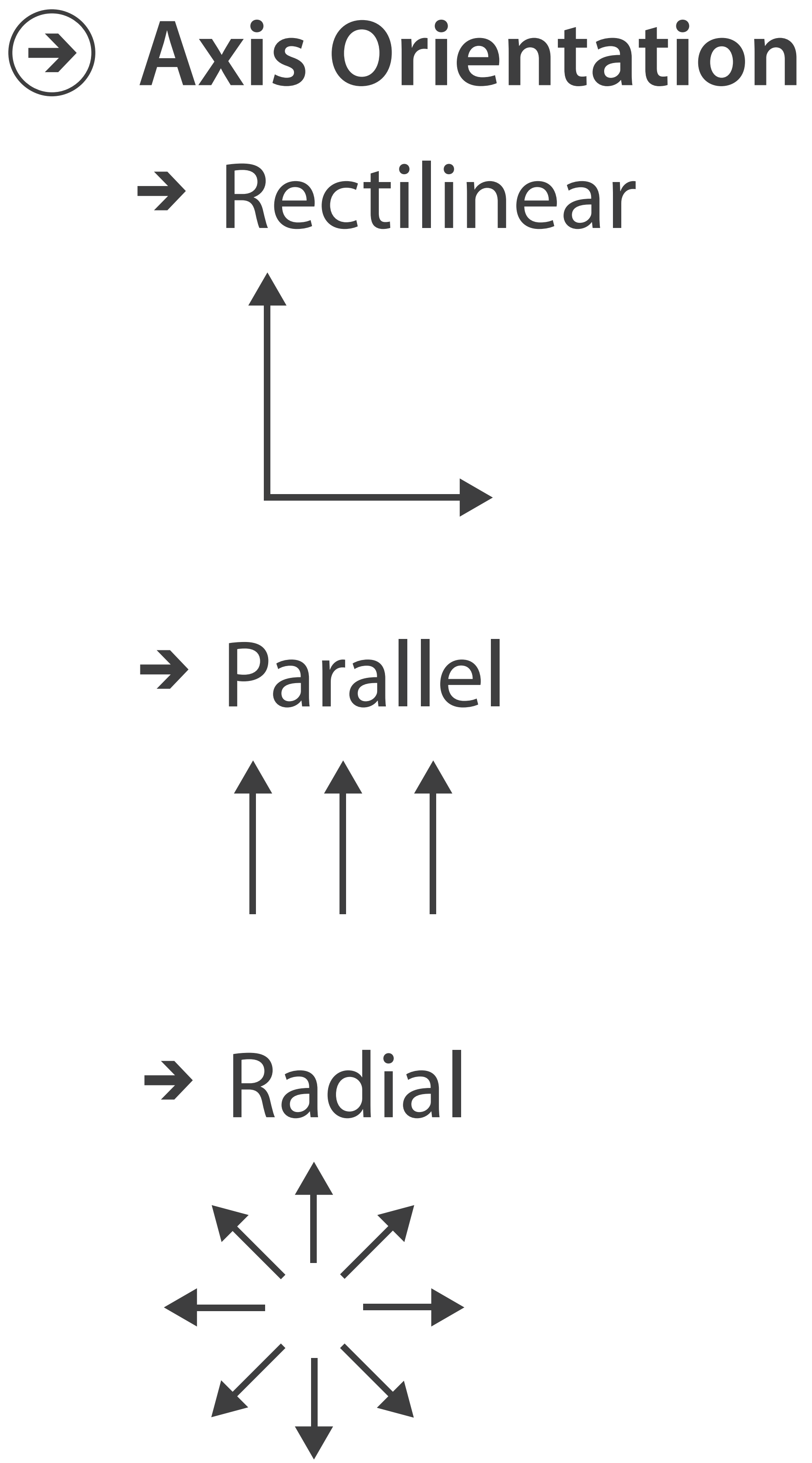


[Glyph-maps for Visually Exploring Temporal Patterns in Climate Data and Models. Wickham, Hofmann, Wickham, and Cook. Environmetrics 23:5 (2012), 382–393.]

Orientation limitations

- rectilinear: scalability wrt #axes
  - 2 axes best
  - 3 problematic
    - more in afternoon
  - 4+ impossible
- parallel: unfamiliarity, training time
- radial: perceptual limits
  - angles lower precision than lengths
  - asymmetry between angle and length
    - can be exploited!

[Uncovering Strengths and Weaknesses of Radial Visualizations - an Empirical Approach. Diehl, Beck and Burch. IEEE TVCG (Proc. InfoVis) 16(6):935--942, 2010.]



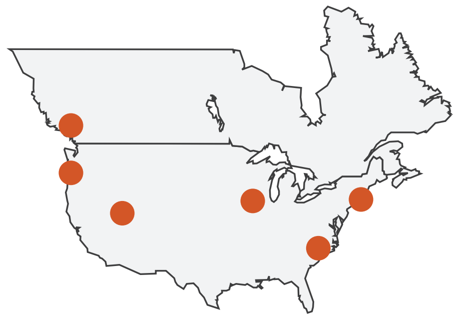
Further reading

- Visualization Analysis and Design. Munzner. AK Peters / CRC Press, Oct 2014.
  - *Chap 7: Arrange Tables*
- Visualizing Data. Cleveland. Hobart Press, 1993.

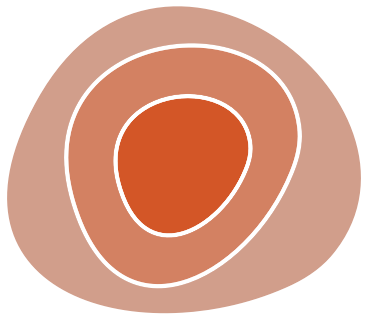
Arrange spatial data

➔ Use Given

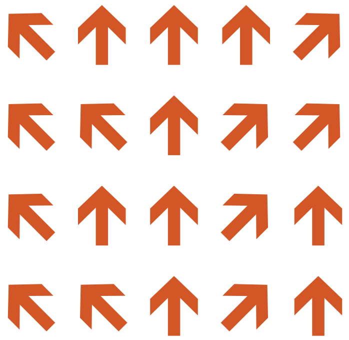
- ➔ Geometry
  - ➔ *Geographic*
  - ➔ *Other Derived*



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

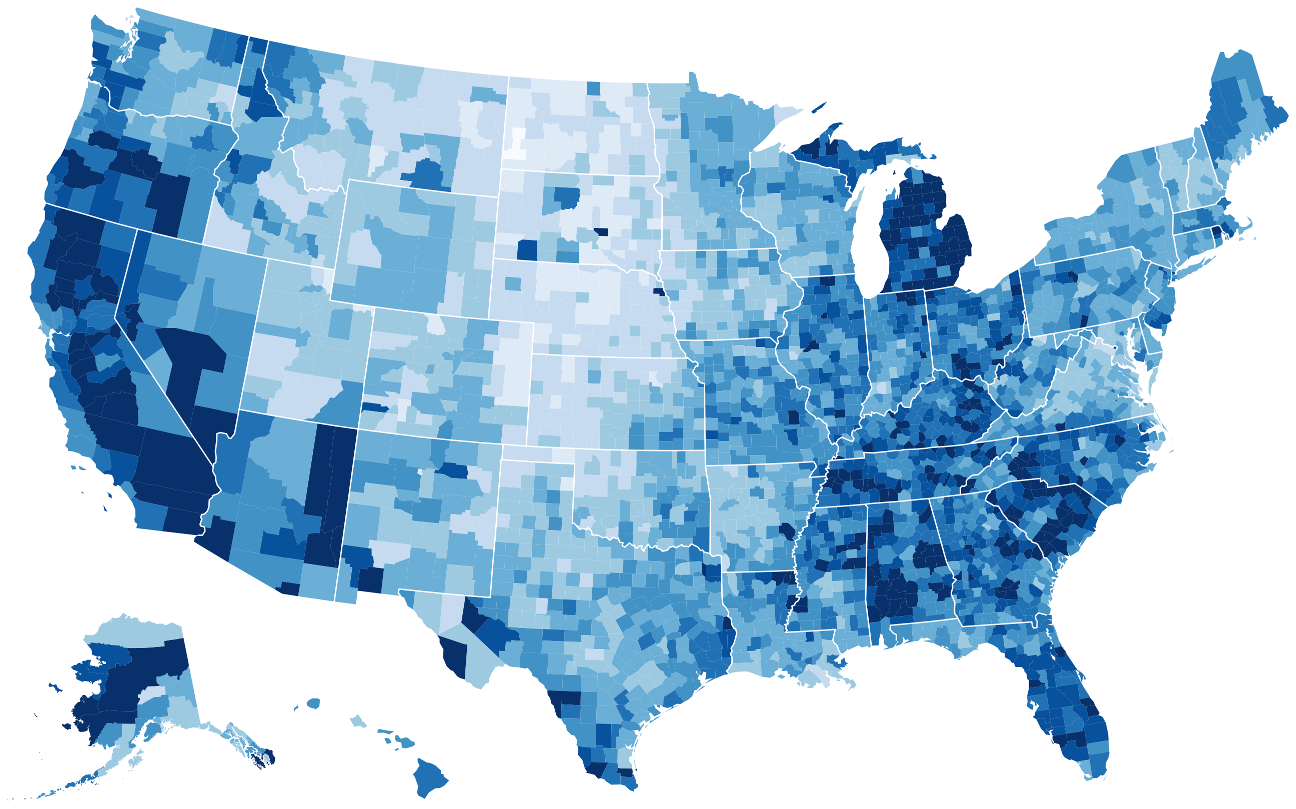


  - ➔ *Vector and Tensor Fields (many values per cell)*
    - ➔ *Flow Glyphs (local)*
    - ➔ *Geometric (sparse seeds)*
    - ➔ *Textures (dense seeds)*
    - ➔ *Features (globally derived)*



Idiom: **choropleth map**

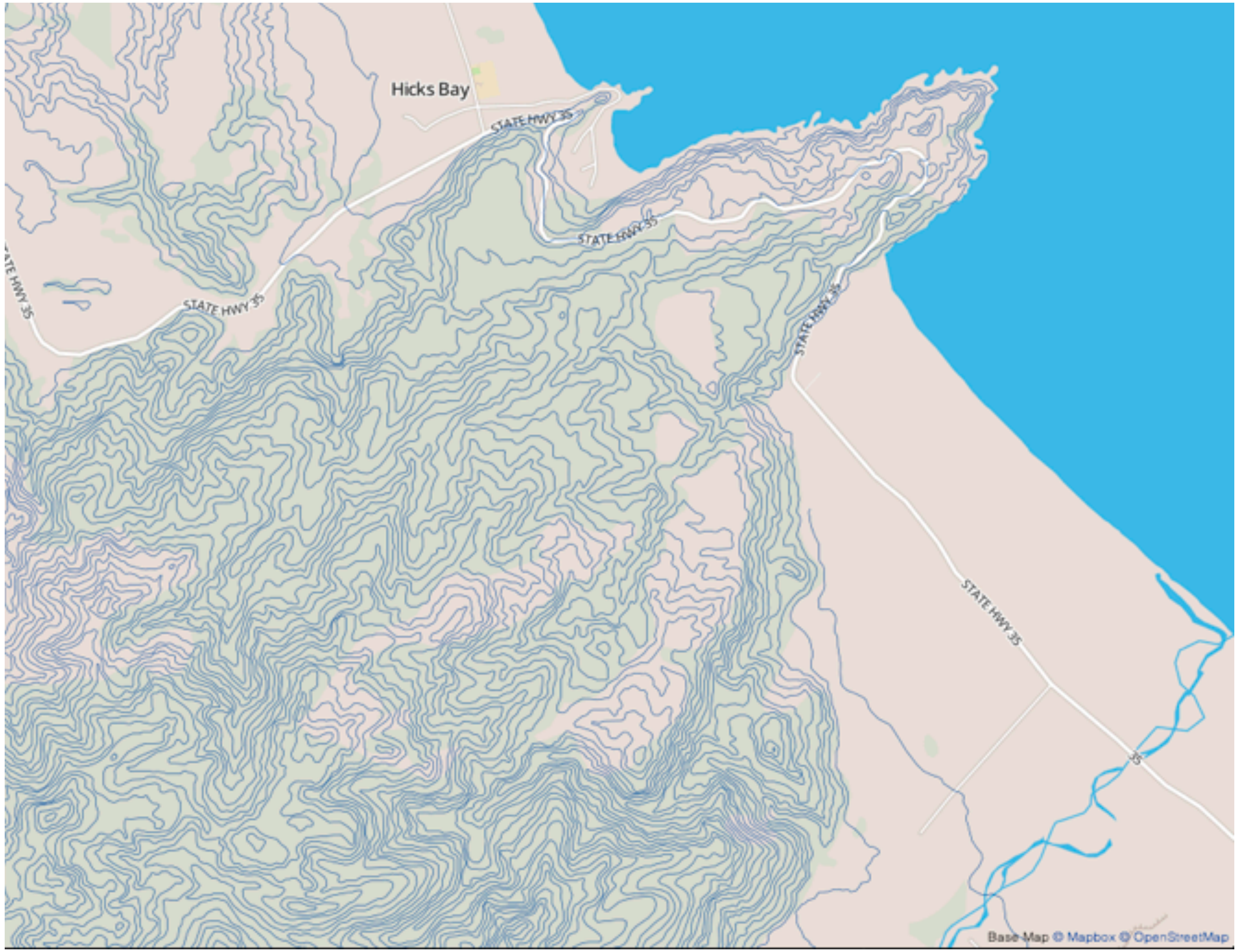
- ***use*** given spatial data
  - when central task is understanding spatial relationships
- data
  - geographic geometry
  - table with 1 quant attribute per region
- encoding
  - use given geometry for area mark boundaries
  - sequential segmented colormap



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

Idiom: **topographic map**

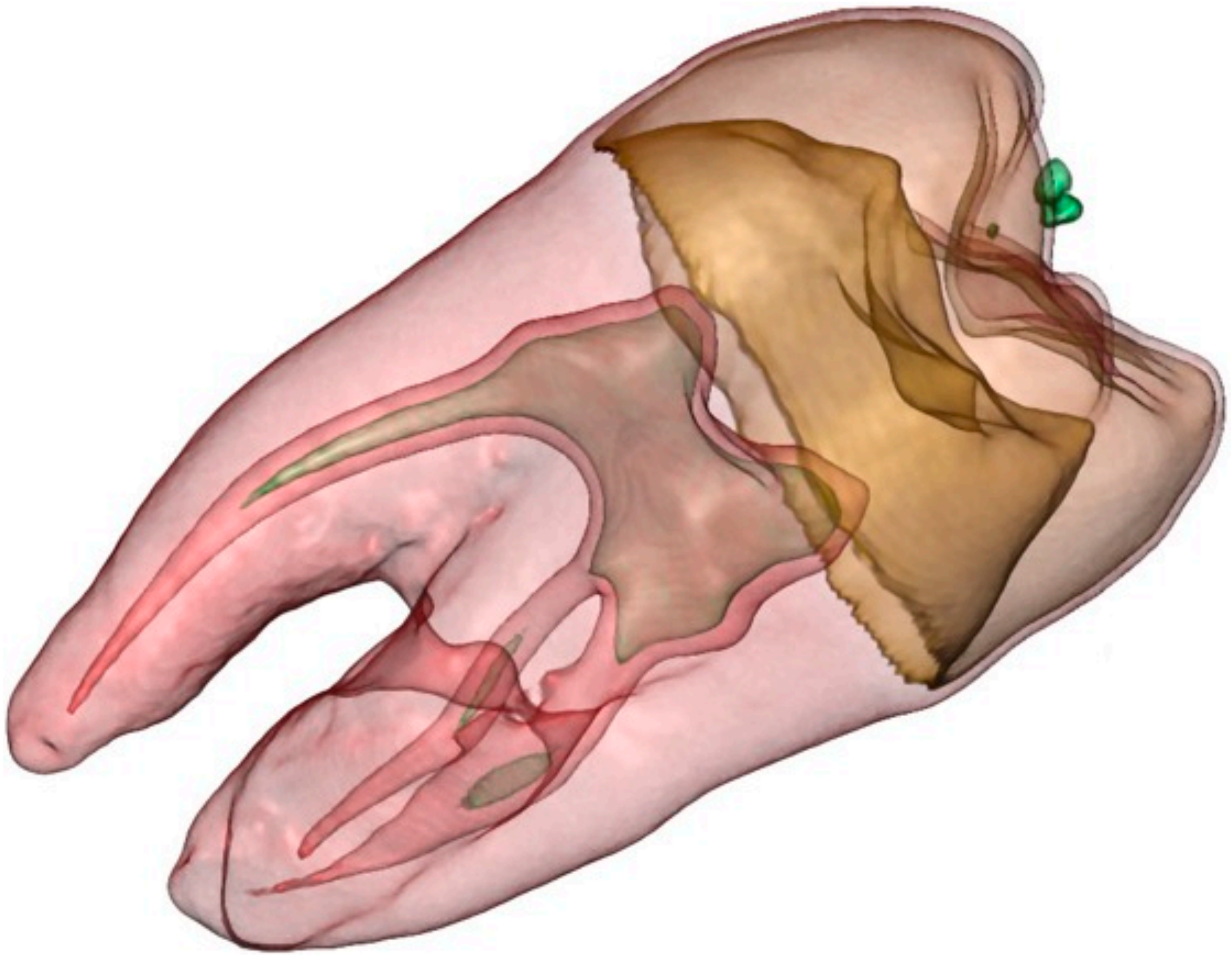
- data
  - geographic geometry
  - scalar spatial field
    - 1 quant attribute per grid cell
- derived data
  - isoline geometry
    - isocontours computed for specific levels of scalar values



Land Information New Zealand Data Service

Idiom: **isosurfaces**

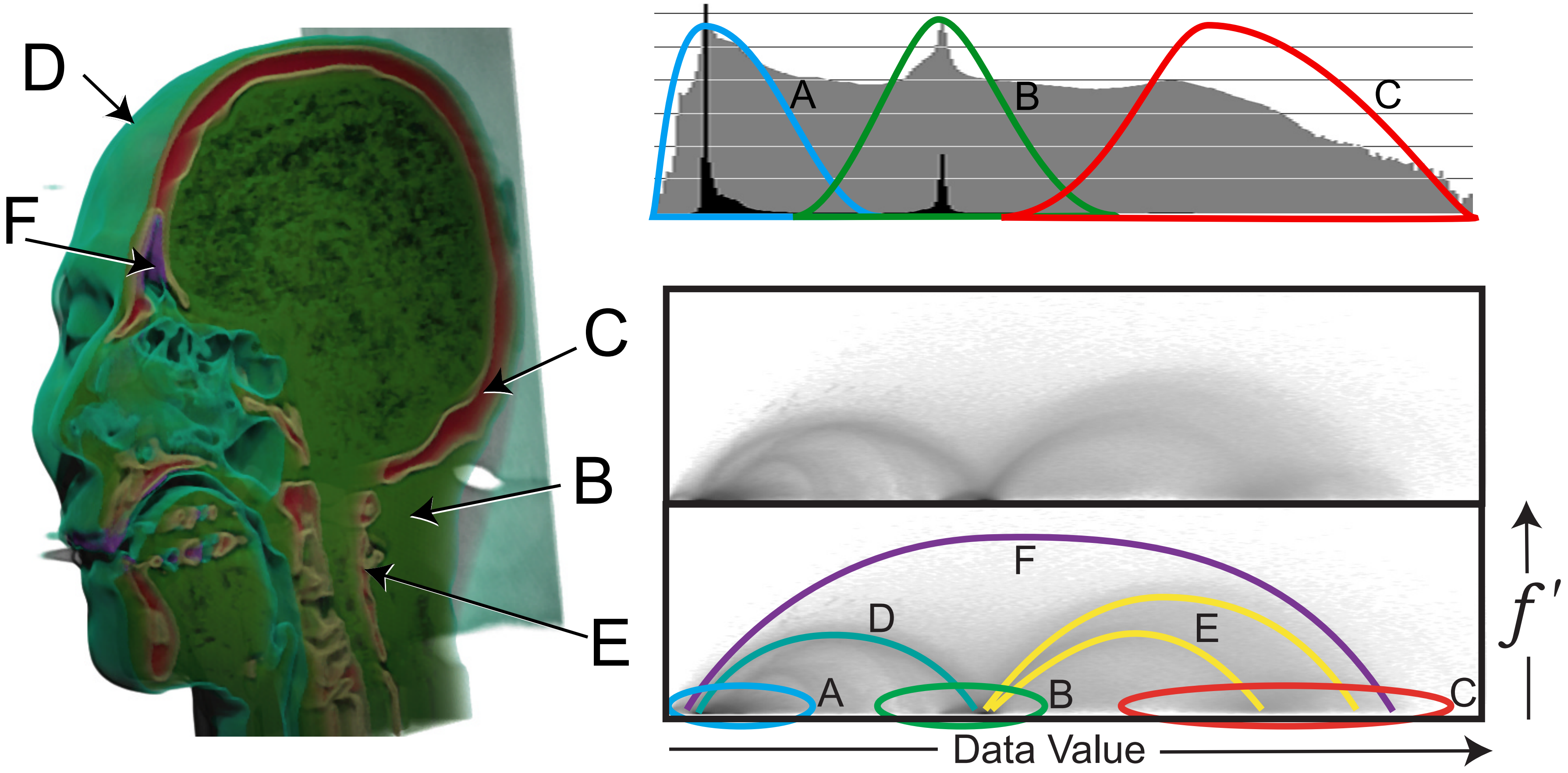
- data
  - scalar spatial field
    - 1 quant attribute per grid cell
- derived data
  - isosurface geometry
    - isocontours computed for specific levels of scalar values
- task
  - spatial relationships



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

Idioms: **DVR, multidimensional transfer functions**

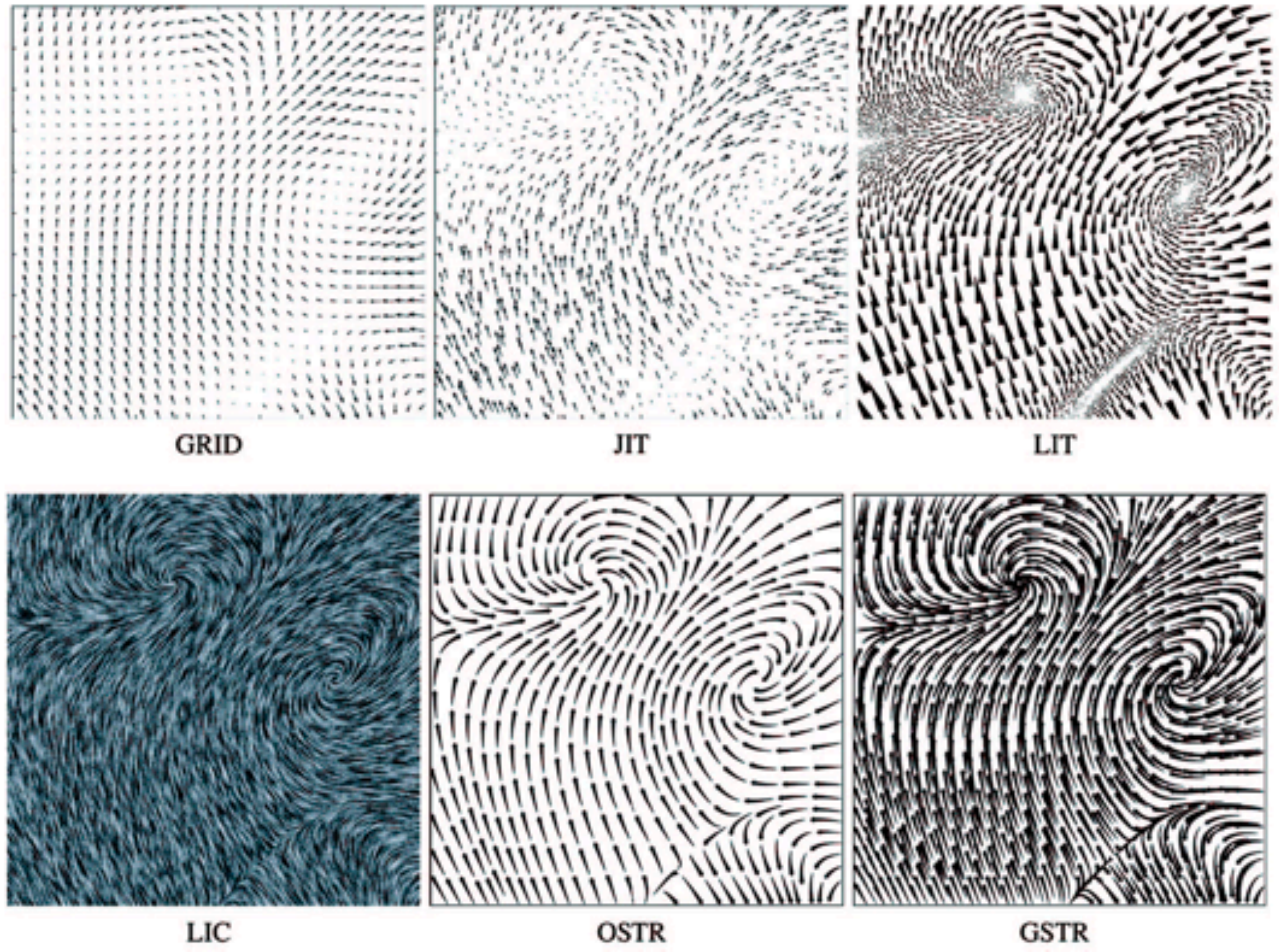
- direct volume rendering
  - **transfer function** maps scalar values to color, opacity
    - no derived geometry
- multidimensional transfer functions
  - derived data in joint 2D histogram
    - horiz axis: data values of scalar func
    - vert axis: gradient magnitude (direction of fastest change)
    - [more on cutting planes and histograms later]



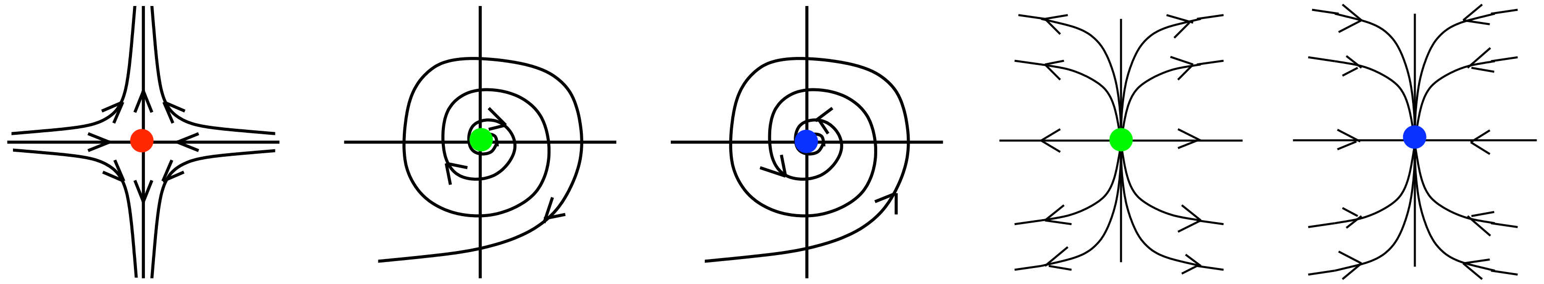
[Multidimensional Transfer Functions for Volume Rendering. Kniss, Kindlmann, and Hansen. In The Visualization Handbook, edited by Charles Hansen and Christopher Johnson, pp. 189–210. Elsevier, 2005.]

Vector and tensor fields

- data
  – many attribs per cell
- idiom families
  – flow glyphs
    - purely local
  – geometric flow
    - derived data from tracing particle trajectories
    - sparse set of seed points
  – texture flow
    - derived data, dense seeds
  – feature flow
    - global computation to detect features
      – encoded with one of methods above



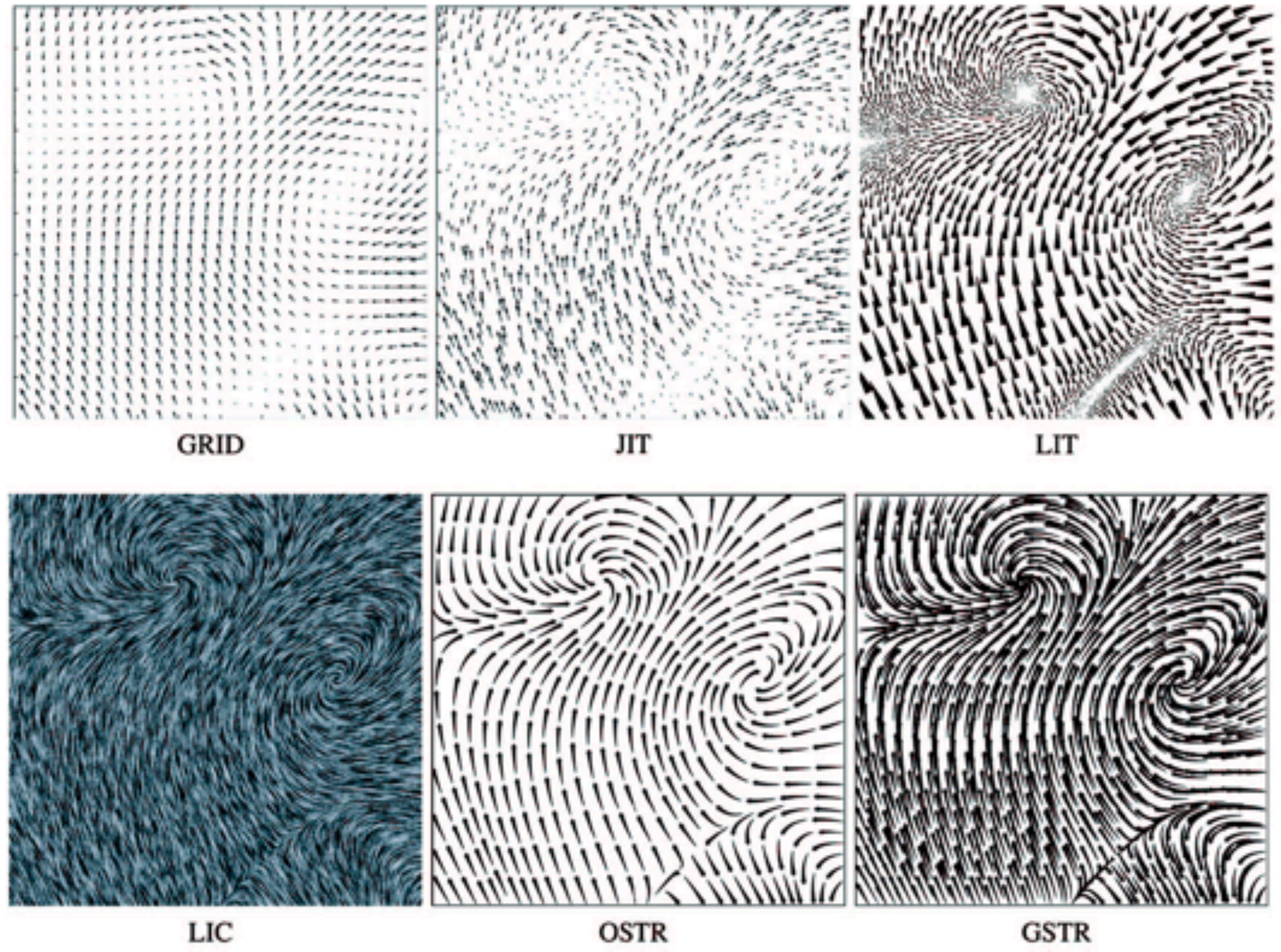
[Comparing 2D vector field visualization methods: A user study. Laidlaw et al. IEEE Trans. Visualization and Computer Graphics (TVCG) 11:1 (2005), 59–70.]



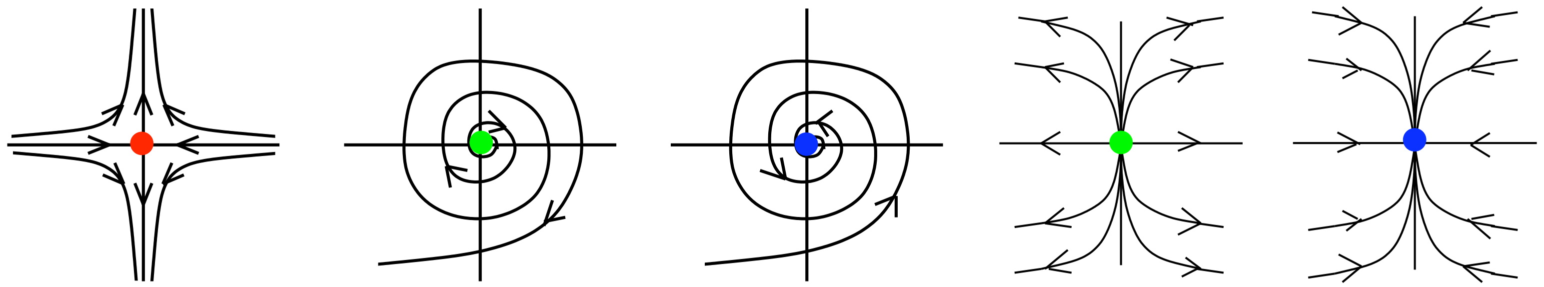
[Topology tracking for the visualization of time-dependent two-dimensional flows. Tricoche, Wischgoll, Scheuermann, and Hagen. Computers & Graphics 26:2 (2002), 249–257.]

Vector fields

- empirical study 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)



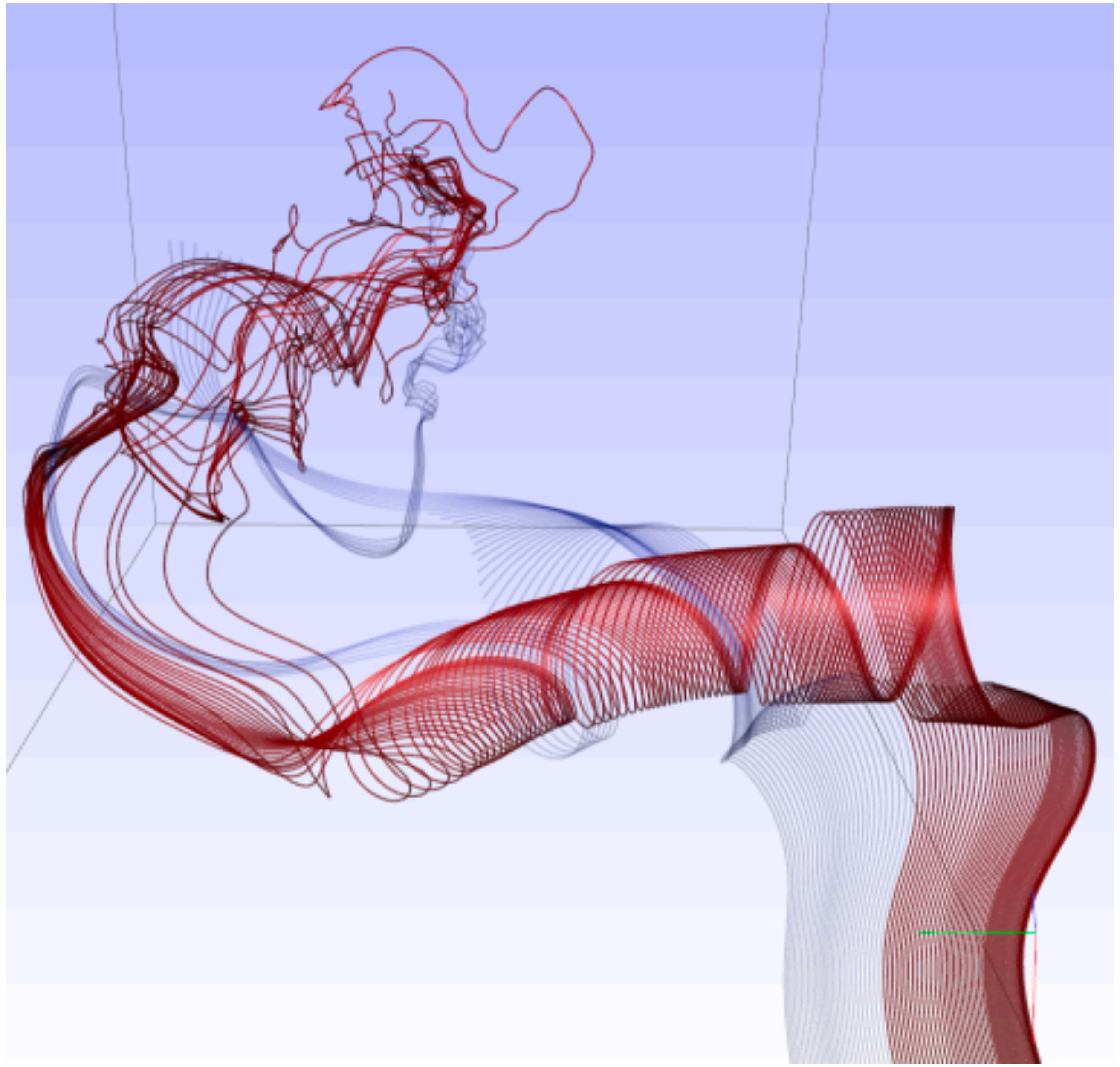
[Comparing 2D vector field visualization methods: A user study. Laidlaw et al. IEEE Trans. Visualization and Computer Graphics (TVCG) 11:1 (2005), 59–70.]



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

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.]

Further reading

- Visualization Analysis and Design. Munzner. AK Peters / CRC Press, Oct 2014.
  - *Chap 8: Arrange Spatial Data*
- How Maps Work: Representation, Visualization, and Design. MacEachren. Guilford Press, 1995.
- Overview of visualization. Schroeder and. Martin. In The Visualization Handbook, edited by Charles Hansen and Christopher Johnson, pp. 3–39. Elsevier, 2005.
- Real-Time Volume Graphics. Engel, Hadwiger, Kniss, Reza-Salama, and Weiskopf. AK Peters, 2006.
- Overview of flow visualization. Weiskopf and Erlebacher. In The Visualization Handbook, edited by Charles Hansen and Christopher Johnson, pp. 261–278. Elsevier, 2005.

Next Time

- to read
  - VAD Ch. 9: Networks
  - Topological Fisheye Views for Visualizing Large Graphs, Emden Gansner, Yehuda Koren and Stephen North. IEEE TVCG 11(4):457-468, 2005.
    - paper type: technique