Week 2:
Chart Types and Best Practices

Tamara Munzner
Department of Computer Science
University of British Columbia

JRNL 520H, Special Topics in Contemporary Journalism: Data Visualization
Week 2: 19 September 2017

www.cs.ubc.ca/~tmm/courses/journ17

How?

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, ...

Manipulate

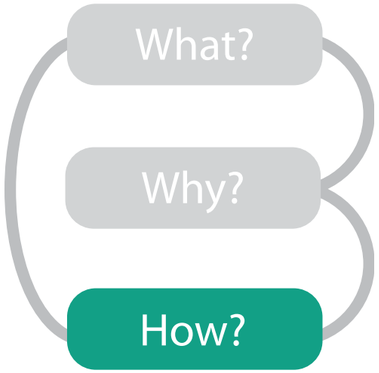
- Change
- Select
- Navigate

Facet

- Juxtapose
- Partition
- Superimpose

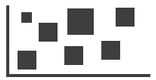
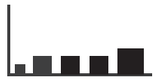
Reduce

- Filter
- Aggregate
- Embed



How?

Encode

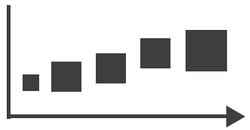
- **Arrange**
  - Express
    
  - Separate
    
  - Order
    
  - Align
    

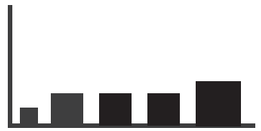
Encode tables: Arrange space

Encode

- **Arrange**
  - Express
  - Separate

  - Order
  - Align

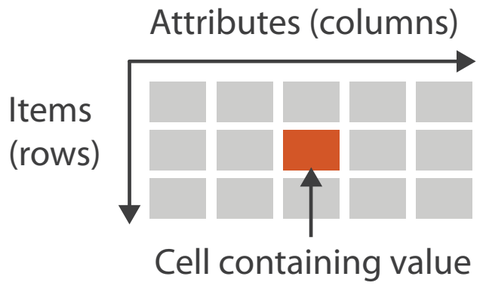




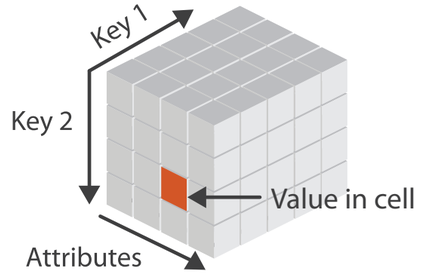
Keys and values

- 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...

➔ Tables



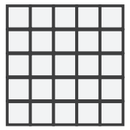
➔ *Multidimensional Table*



➔ **Express Values**

➔ *1 Key*
List

➔ *2 Keys*
Matrix

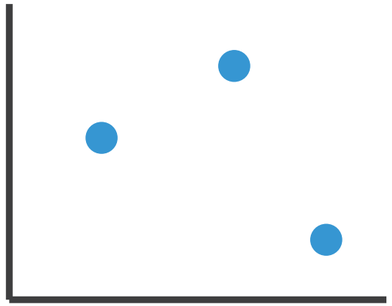


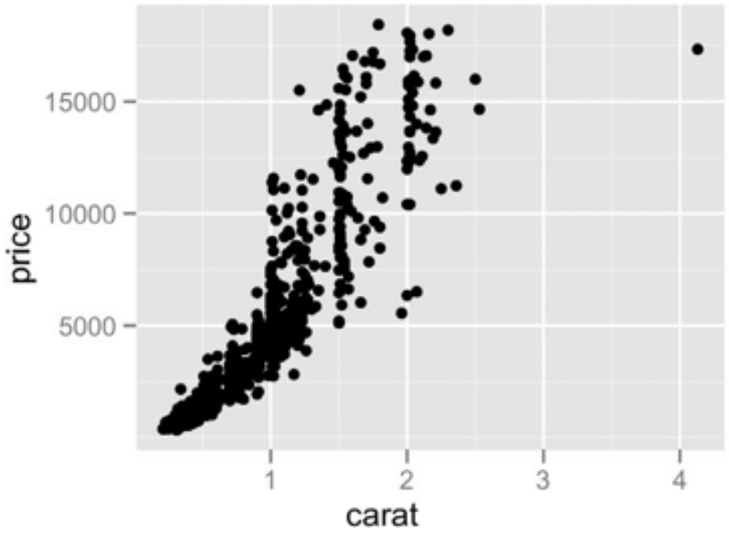
➔ *3 Keys*
Volume

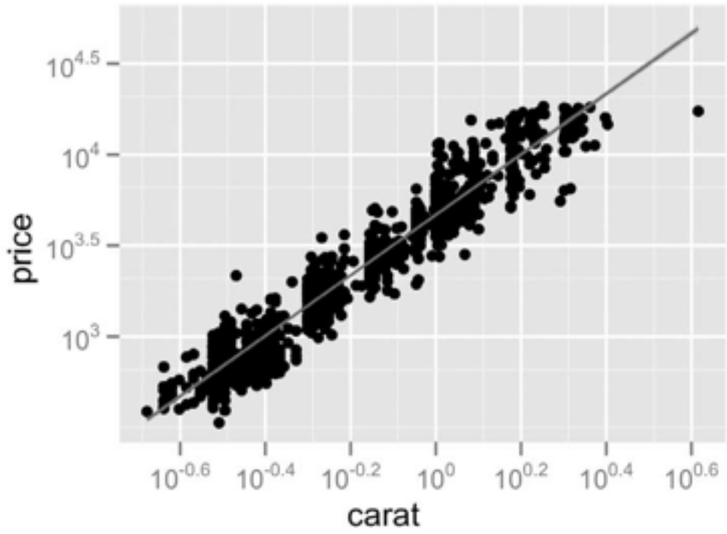
➔ *Many Keys*
Recursive Subdivision

Idiom: **scatterplot**

- ***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



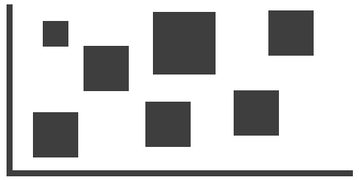




[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]

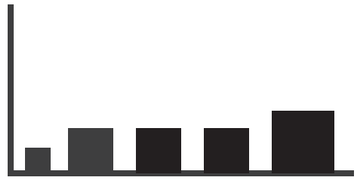
Some keys: Categorical regions

➔ Separate



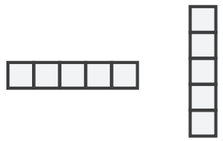
➔ Order

➔ Align

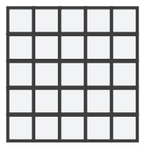


- **regions**: contiguous bounded areas distinct from each other
  - using space to ***separate*** (proximity)
  - following expressiveness principle for categorical attributes
- use ordered attribute to ***order*** and ***align*** regions

➔ *1 Key*
List



➔ *2 Keys*
Matrix

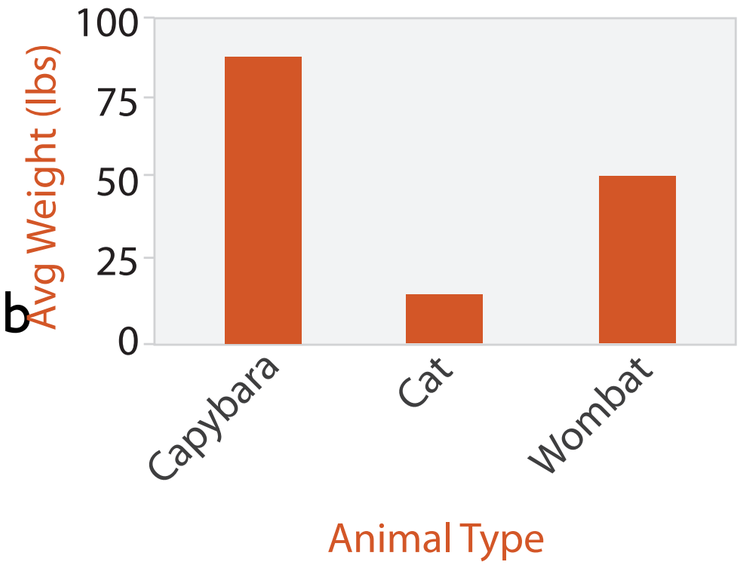


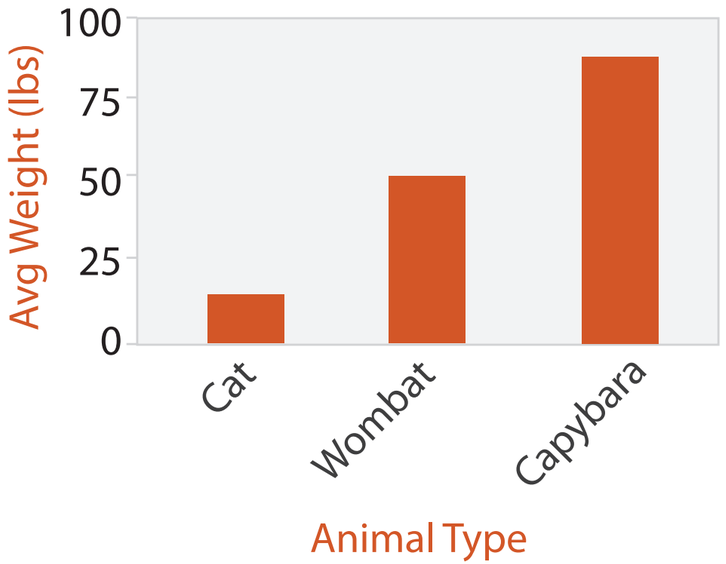
➔ *3 Keys*
Volume

➔ *Many Keys*
Recursive Subdivision

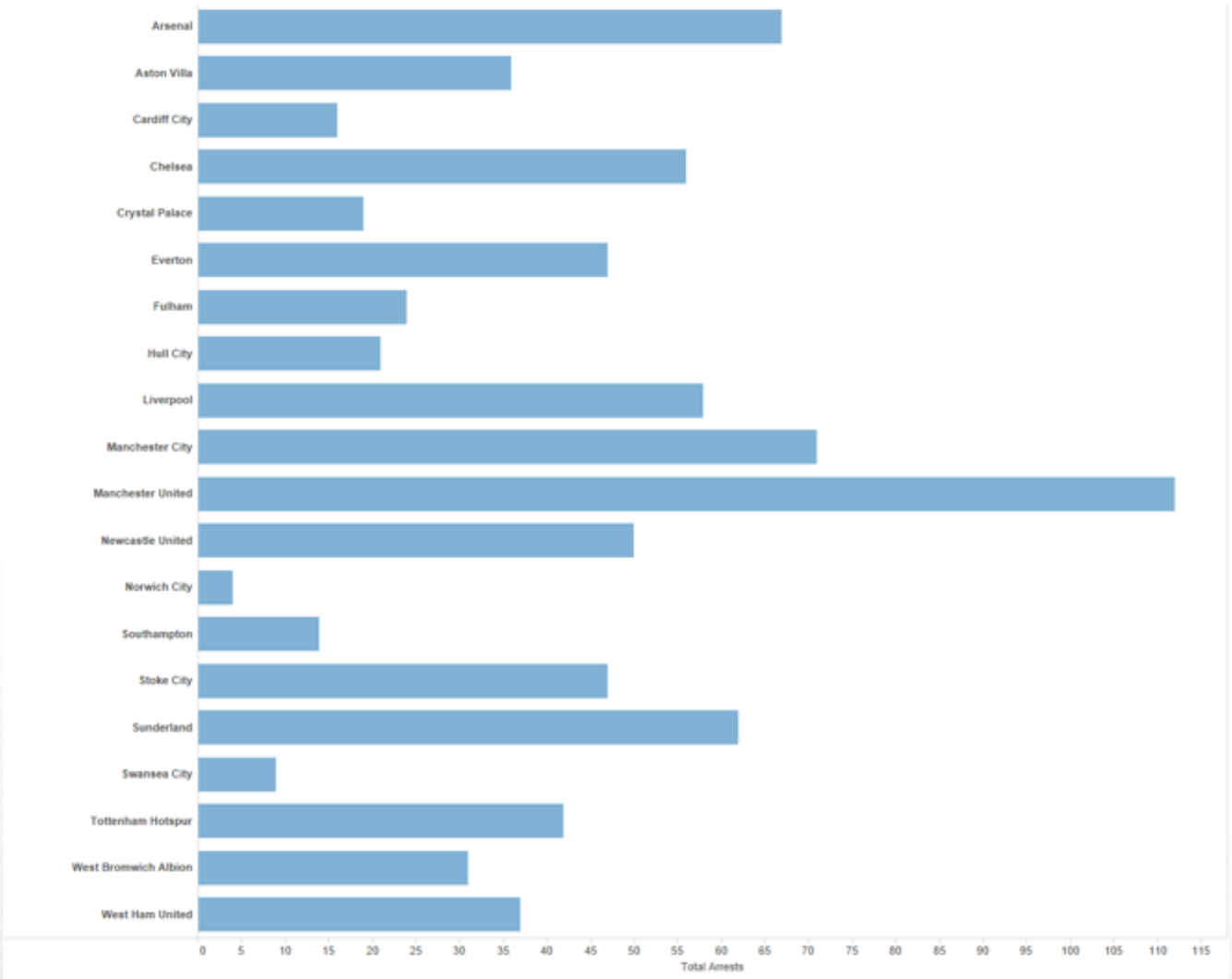
Idiom: **bar chart**

- 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





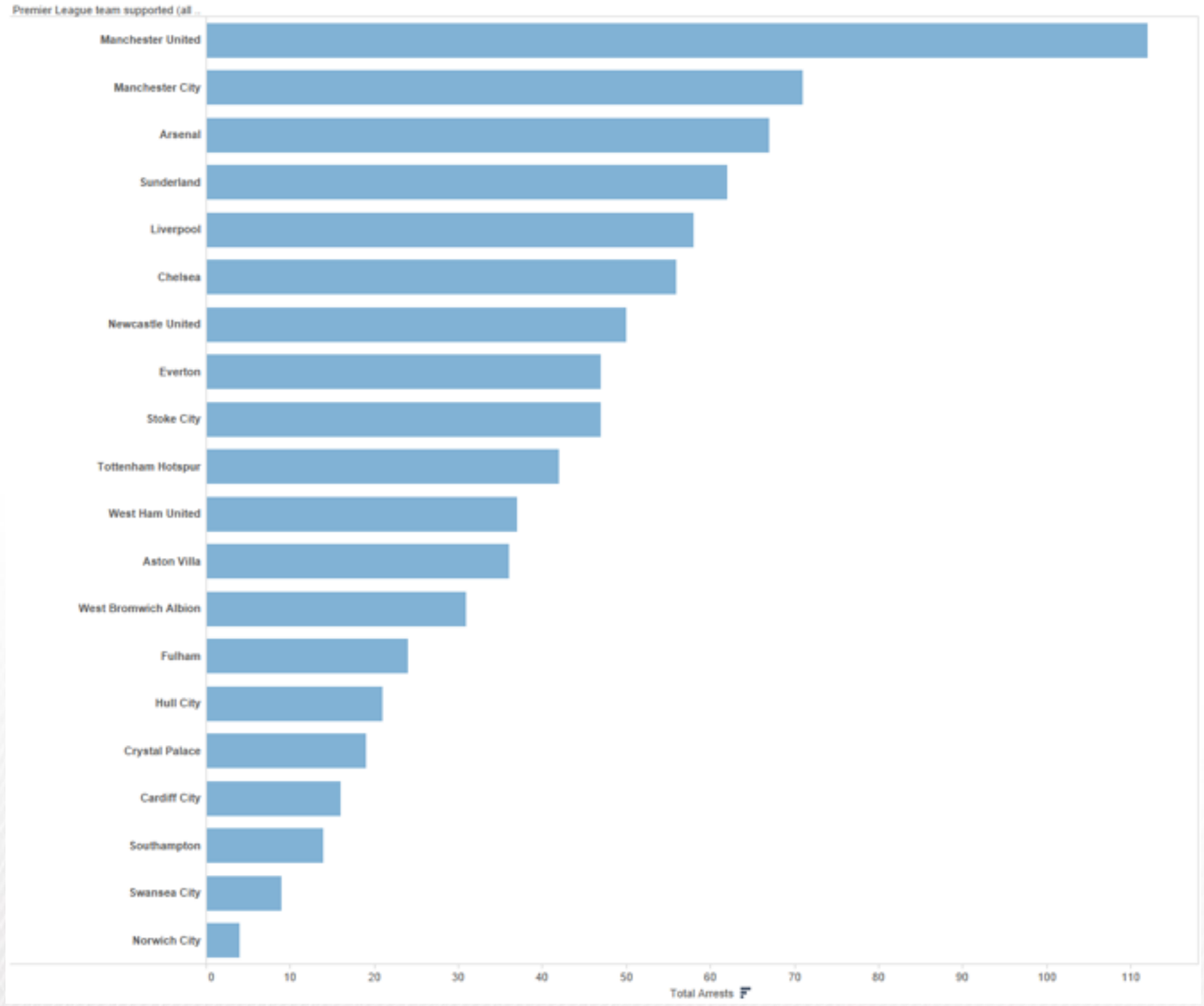
Separated and Aligned but not Ordered



LIMITATION: Hard to know rank. What's the 4th most? The 7th?

[Slide courtesy of Ben Jones]

Separated, Aligned and Ordered



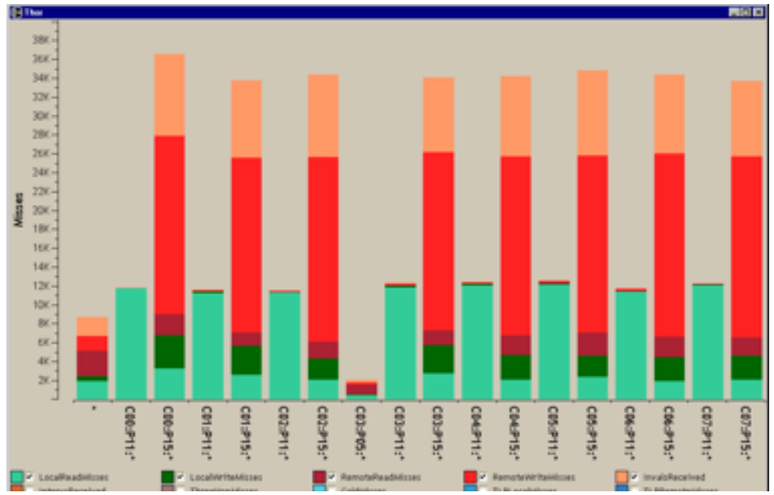
[Slide courtesy of Ben Jones]

Separated but not Ordered or Aligned

LIMITATION: Hard to make comparisons

[Slide courtesy of Ben Jones]

Idiom: **stacked bar chart**

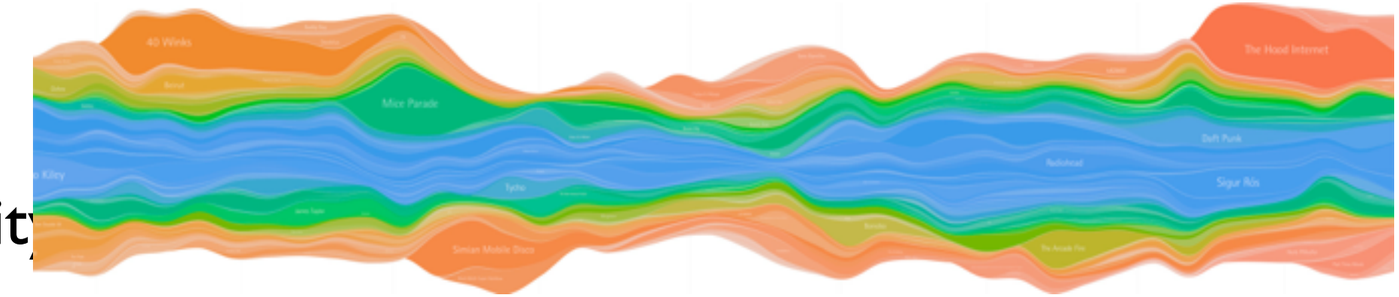


- one more key
  - –data
    - 2 categ attrib, 1 quant attrib
  - –mark: vertical stack of line marks
    - **glyph**: composite object, internal structure from multiple marks
  - –channels
    - length and color hue
    - spatial regions: one per glyph
      - – aligned: full glyph, lowest bar component
      - – 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.]

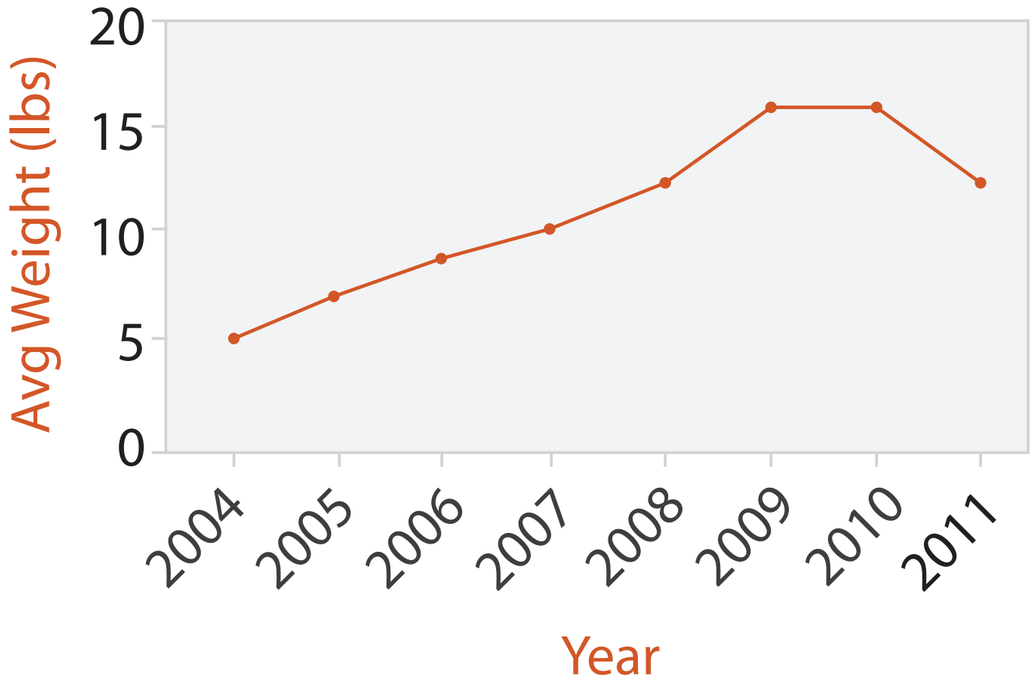
Idiom: **streamgraph**

- 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).]

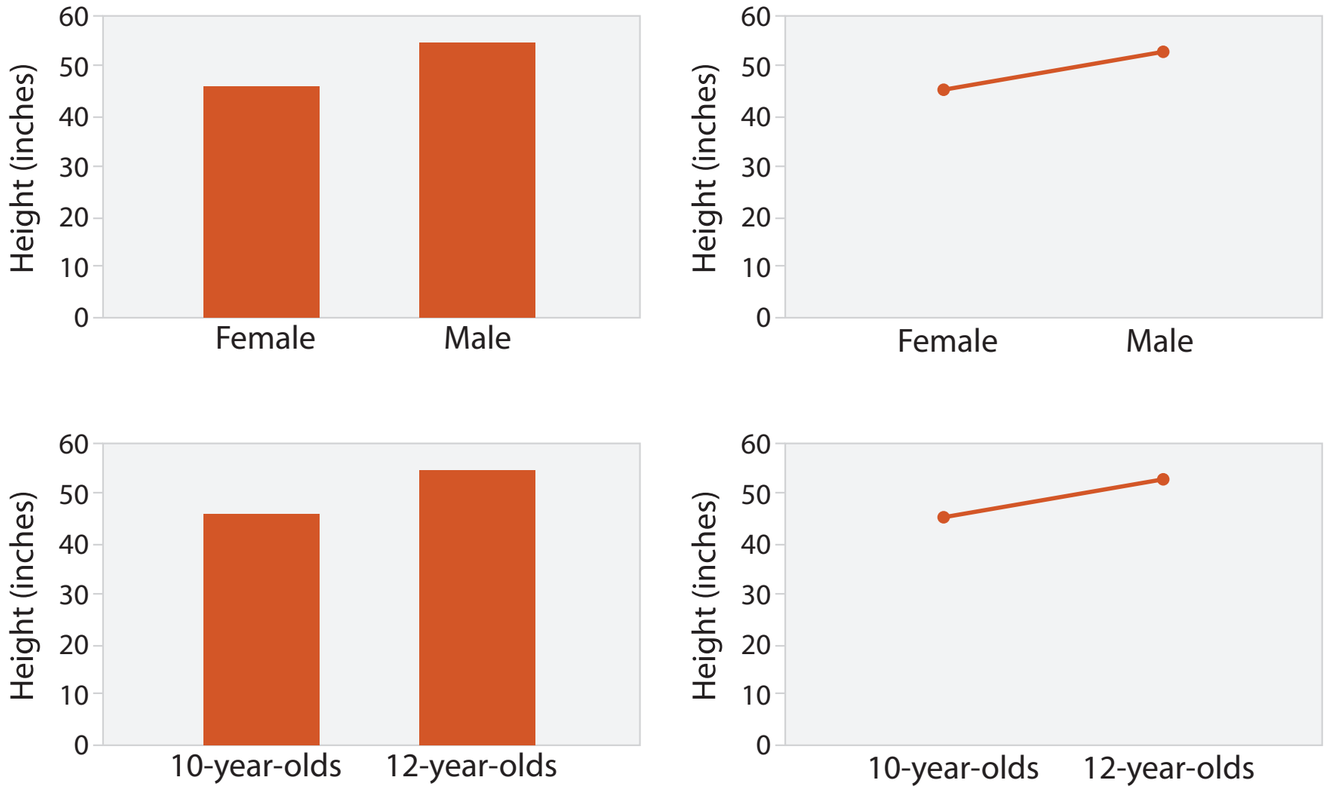
Idiom: **line chart**



- 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

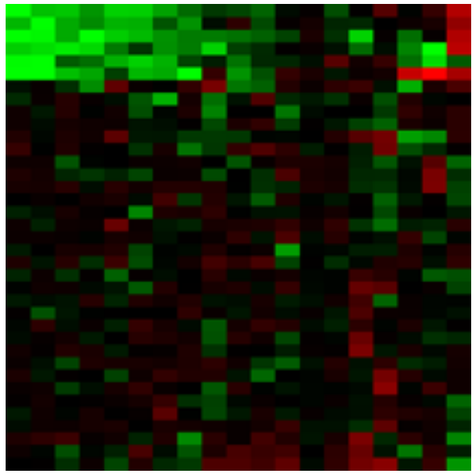
- depends on type of key attrib
  - bar charts if categorical
  - line charts if ordered
- do not use line charts for categorical key attribs
  - violates expressiveness principle
    - implication of trend so strong that it overrides semantics!
      - "The more male a person is, the taller he/she is"

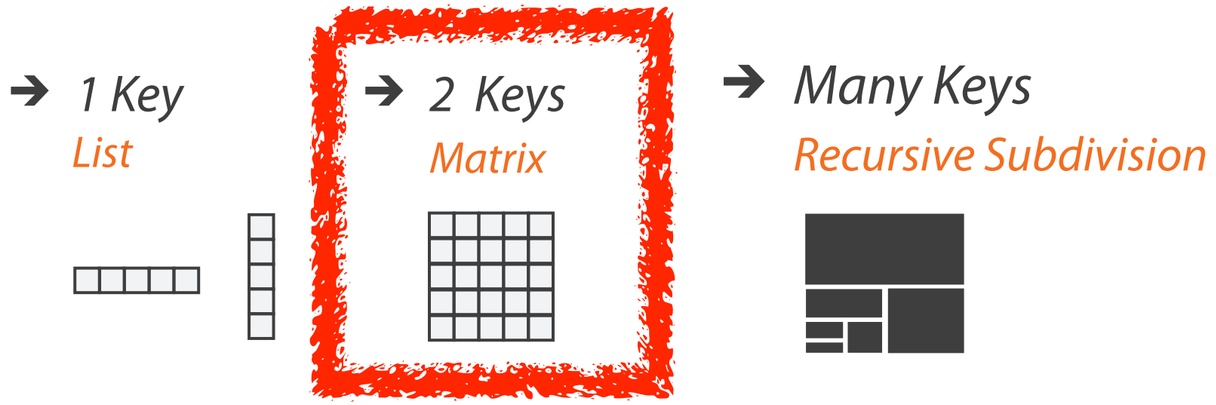


after [Bars and Lines: A Study of Graphic Communication. Zacks and Tversky. Memory and Cognition 27:6 (1999), 1073–1079.]

Idiom: **heatmap**

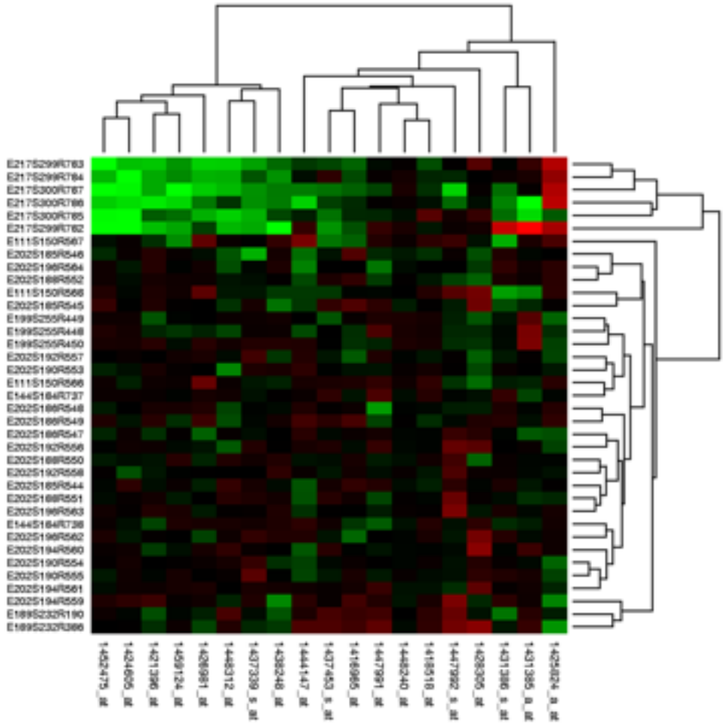
- two keys, one value
  - –data
    - 2 categ attribs (gene, experimental condition)
    - 1 quant attrib (expression levels)
  - –marks: area
    - separate and align in 2D matrix
      - – indexed by 2 categorical attributes
  - –channels
    - color by quant attrib
      - – (ordered diverging colormap)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





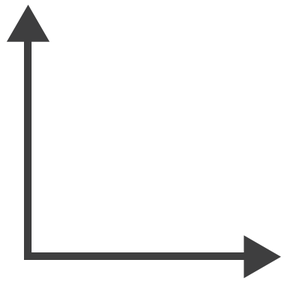
Idiom: **cluster heatmap**

- 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

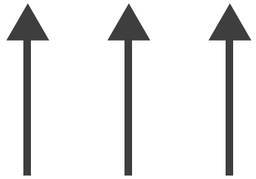


Axis Orientation

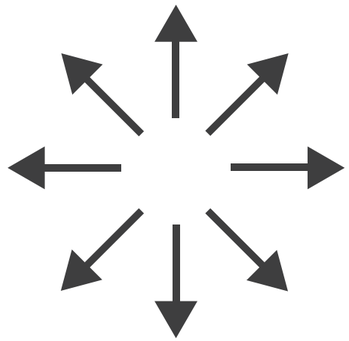
- Rectilinear



- Parallel



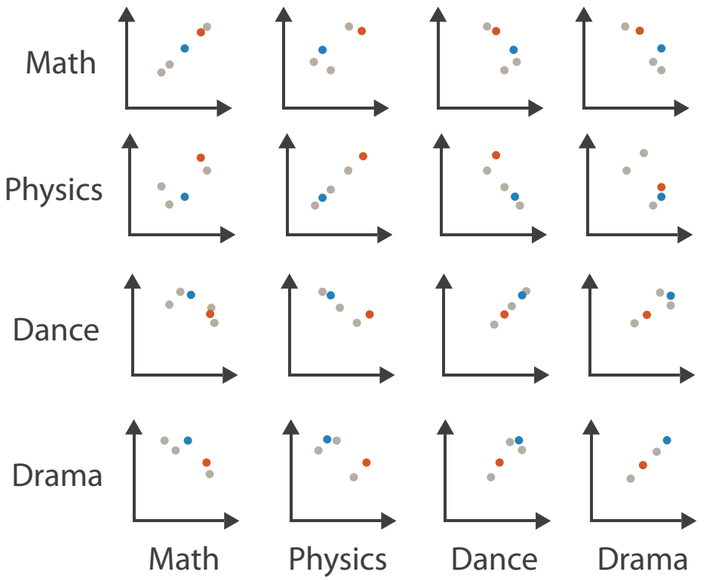
- Radial



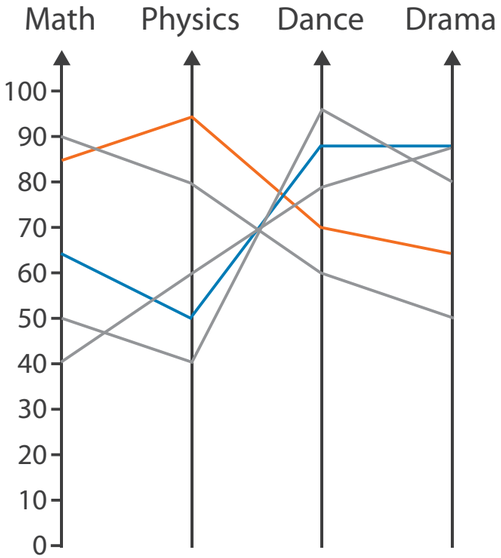
Idioms: **scatterplot matrix, parallel coordinates**

- 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

Scatterplot Matrix



Parallel Coordinates



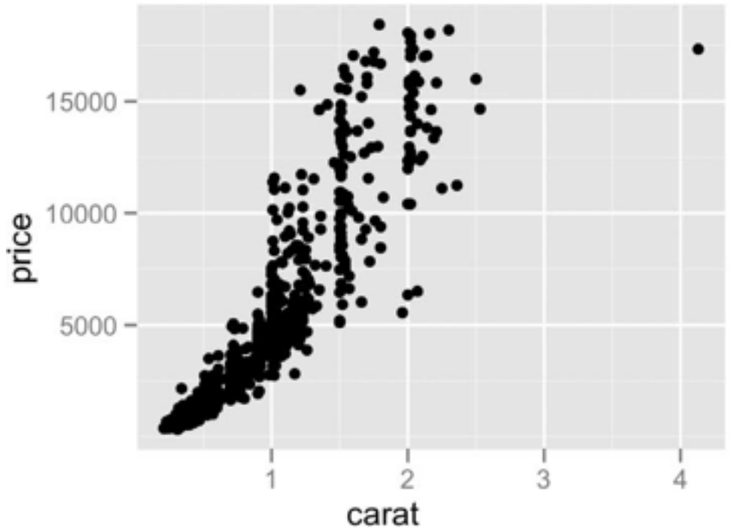
Table

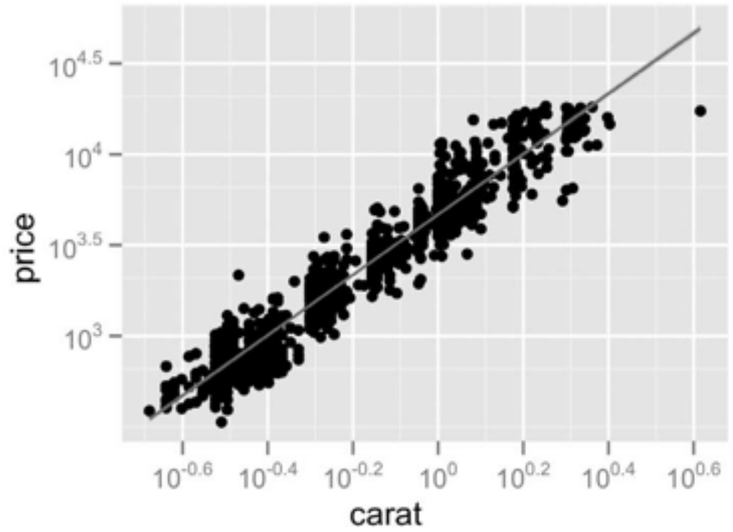
Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

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.]

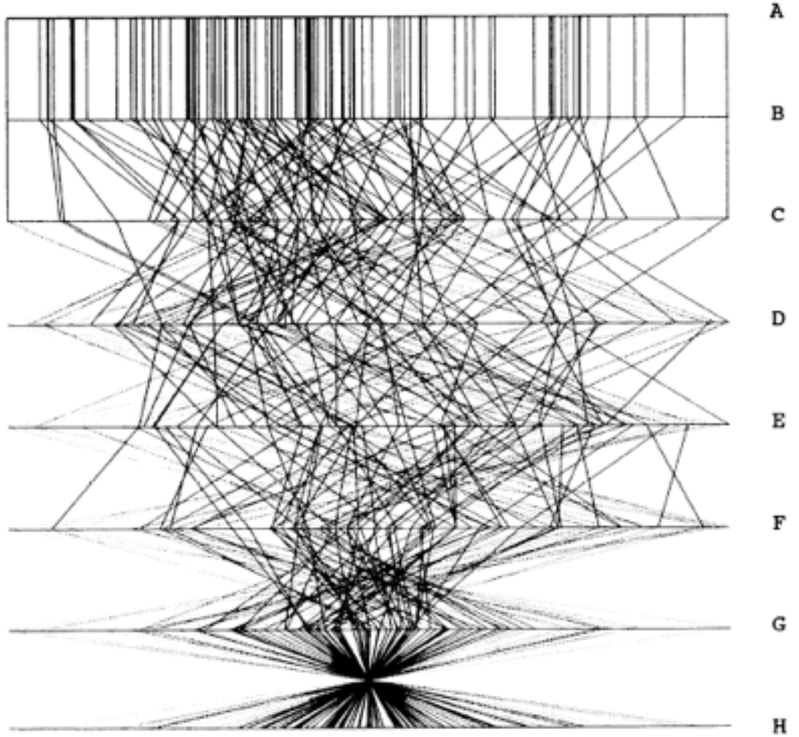
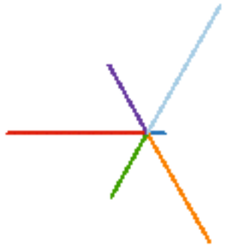


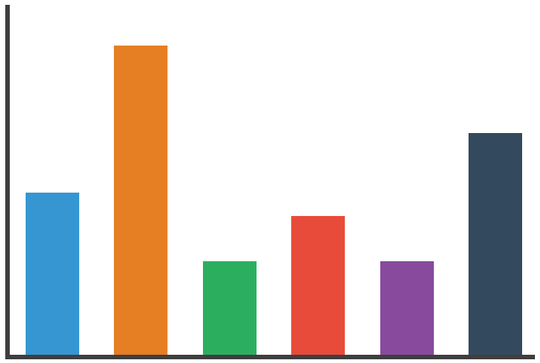
Figure 3. Parallel Coordinate Plot of Six-Dimensional Data Illustrating Correlations of $\rho = 1, .8, .2, 0, -.2, -.8,$ and -1.

[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

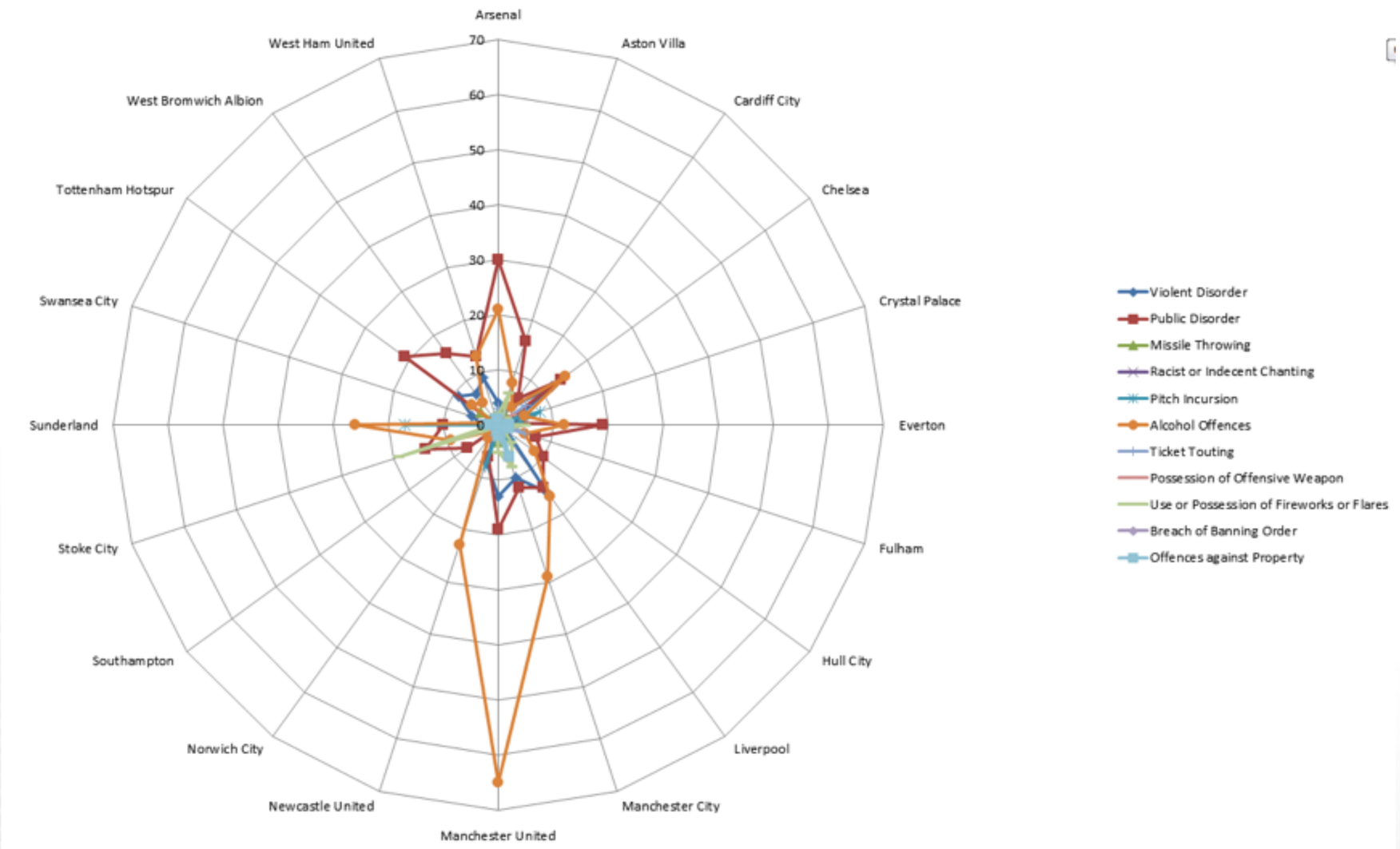




- accuracy
  - length unaligned with radial
    - less accurate than aligned with rectilinear

[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.]

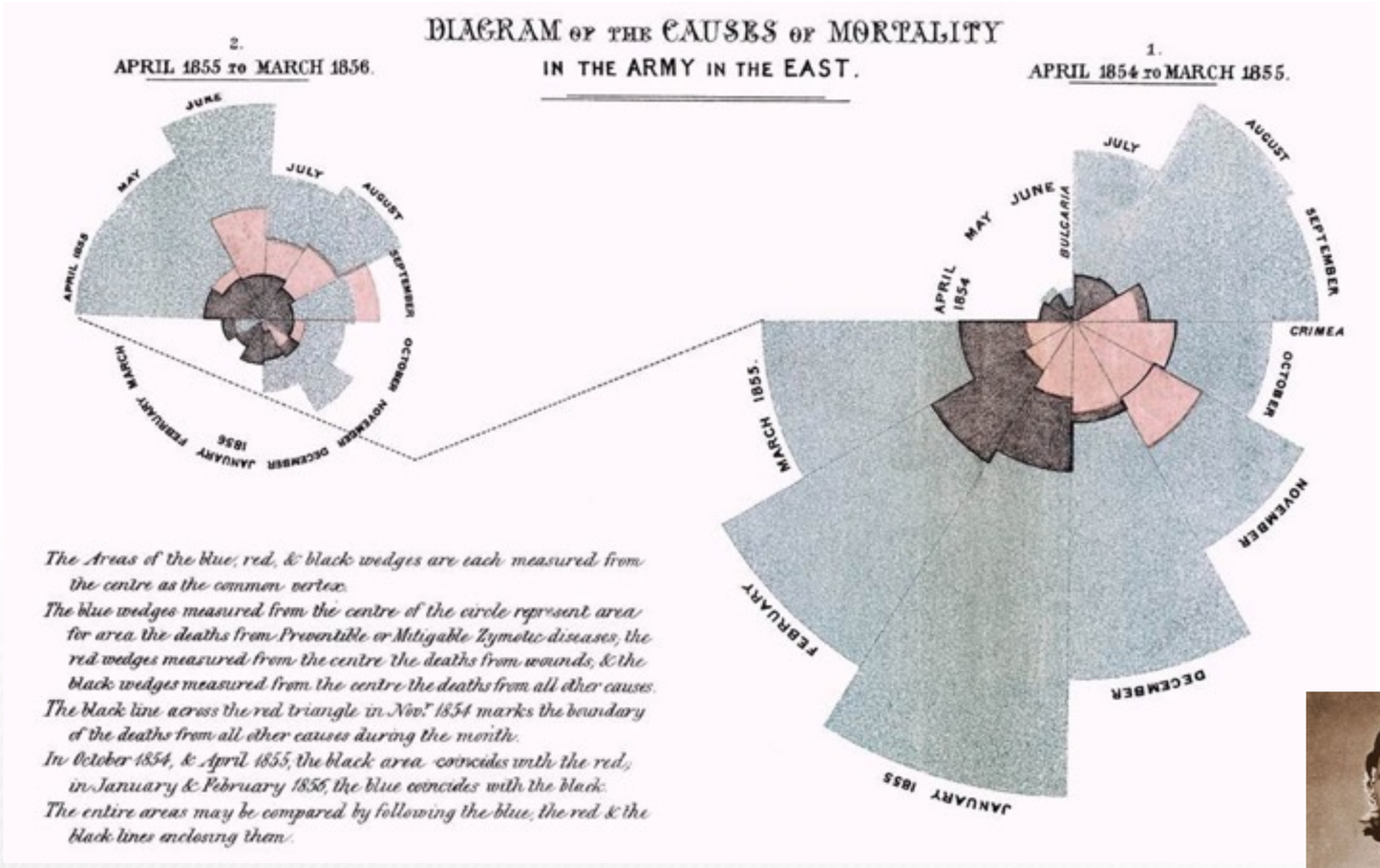
Radial Orientation: Radar Plots



LIMITATION: Not good when categories aren't cyclic

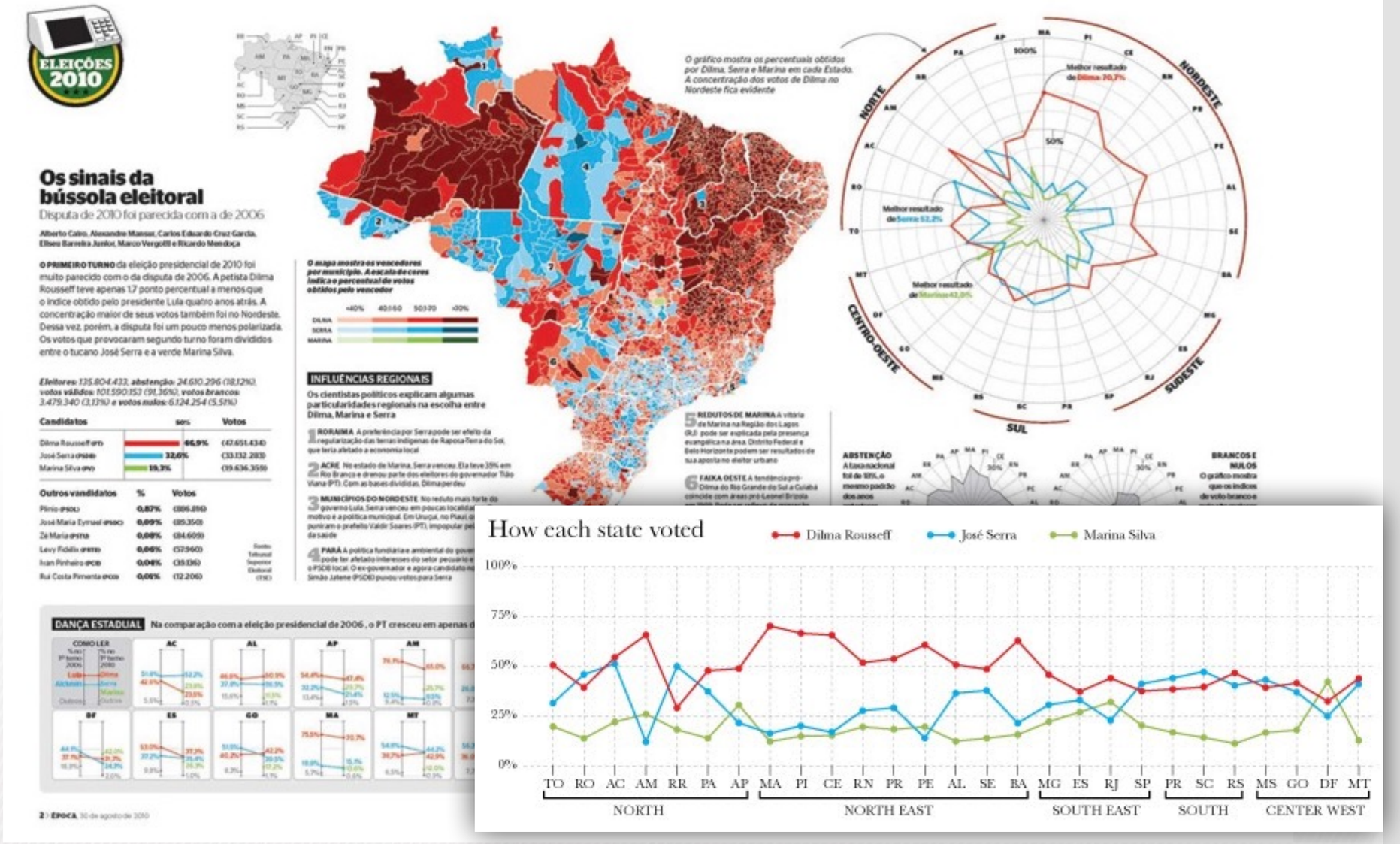
[Slide courtesy of Ben Jones]

"Diagram of the causes of mortality in the army in the East" (1858)



[Slide courtesy of Ben Jones]

“Radar graphs: Avoid them (99.9% of the time)”

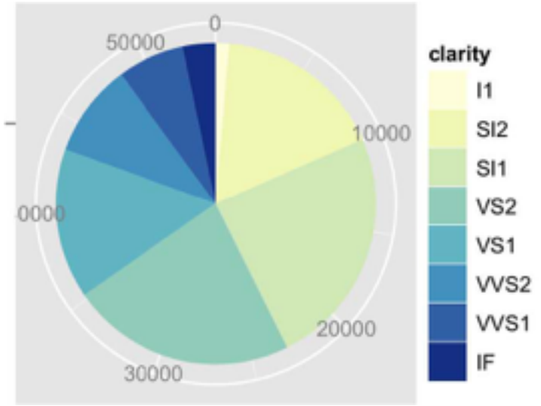


http://www.thefunctionalart.com/2012/11/radar-graphs-avoid-them-999-of-time.html

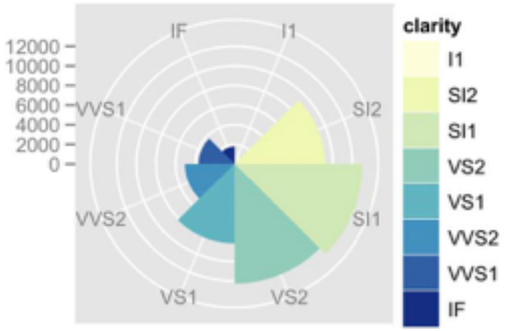
[Slide courtesy of Ben Jones]

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

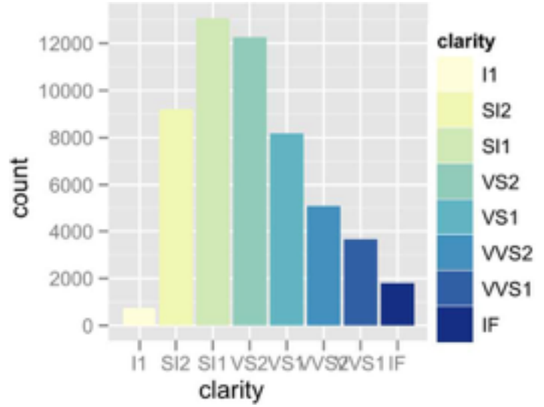
- pie chart
  - area marks with angle channel
  - accuracy: angle/area much less accurate than line length



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



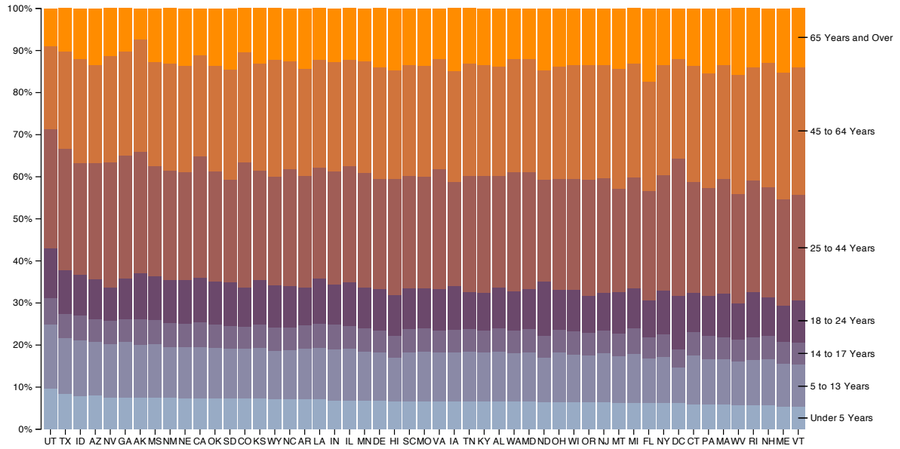
- data
  - 1 categ key attrib, 1 quant value attrib
- task
  - part-to-whole judgements

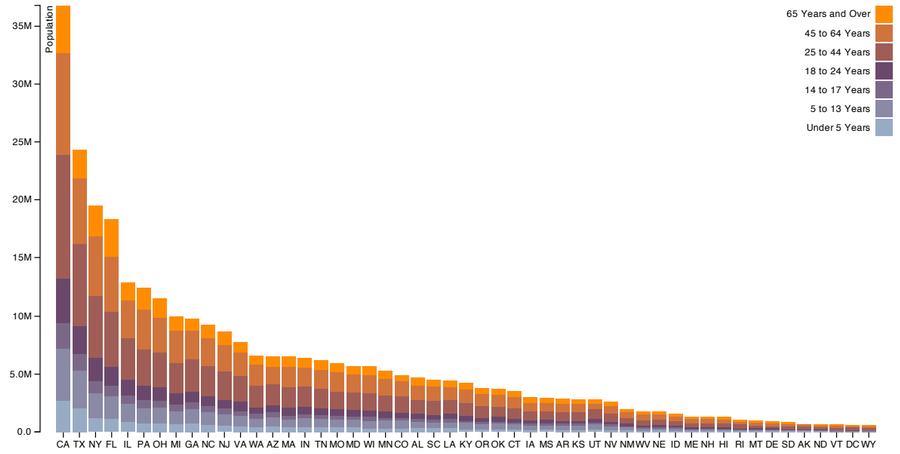


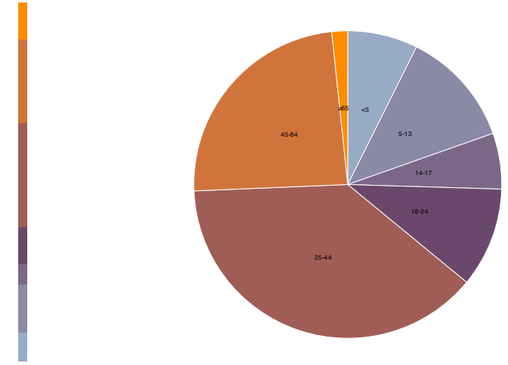
[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]

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



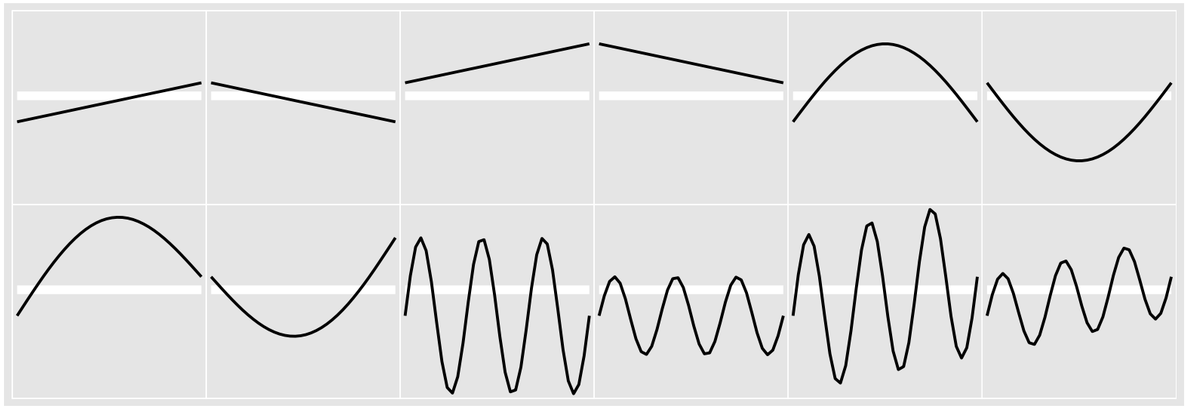




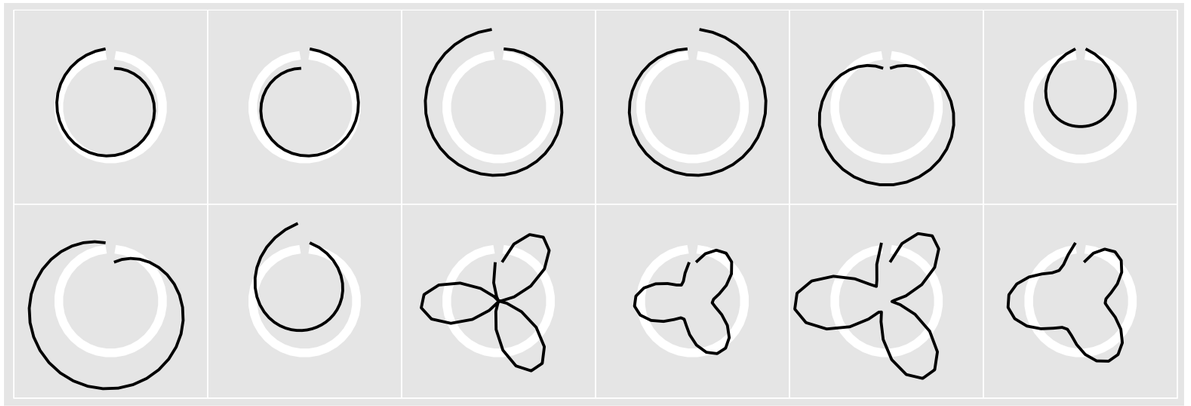
http://bl.ocks.org/mbostock/3887235,
http://bl.ocks.org/mbostock/3886208,
http://bl.ocks.org/mbostock/3886394.

Idiom: **glyphmaps**

- rectilinear good for linear vs nonlinear trends



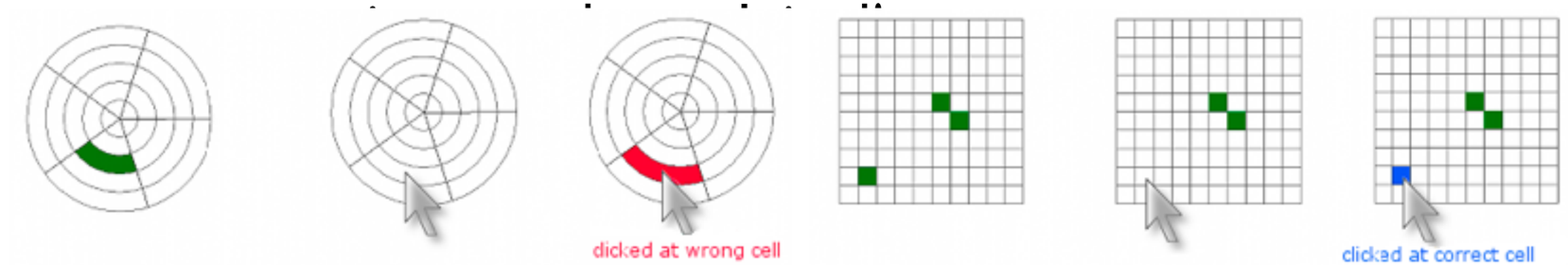
- radial good for cyclic patterns



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

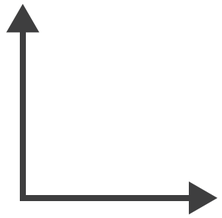
Orientation limitations

- rectilinear: scalability wrt #axes
  - 2 axes best
  - 3 problematic
    - more in afternoon
  - 4+ impossible
- parallel: unfamiliarity, training time
- radial: perceptual limits
  - asymmetry: angles lower precision than lengths

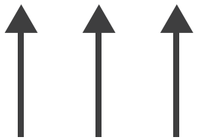


Axis Orientation

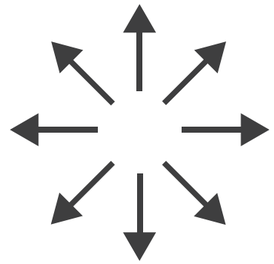
- Rectilinear



- Parallel



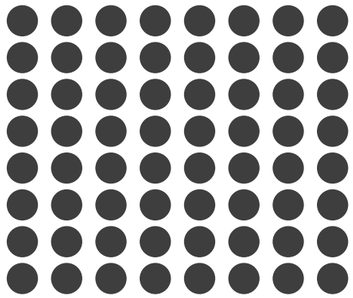
- Radial



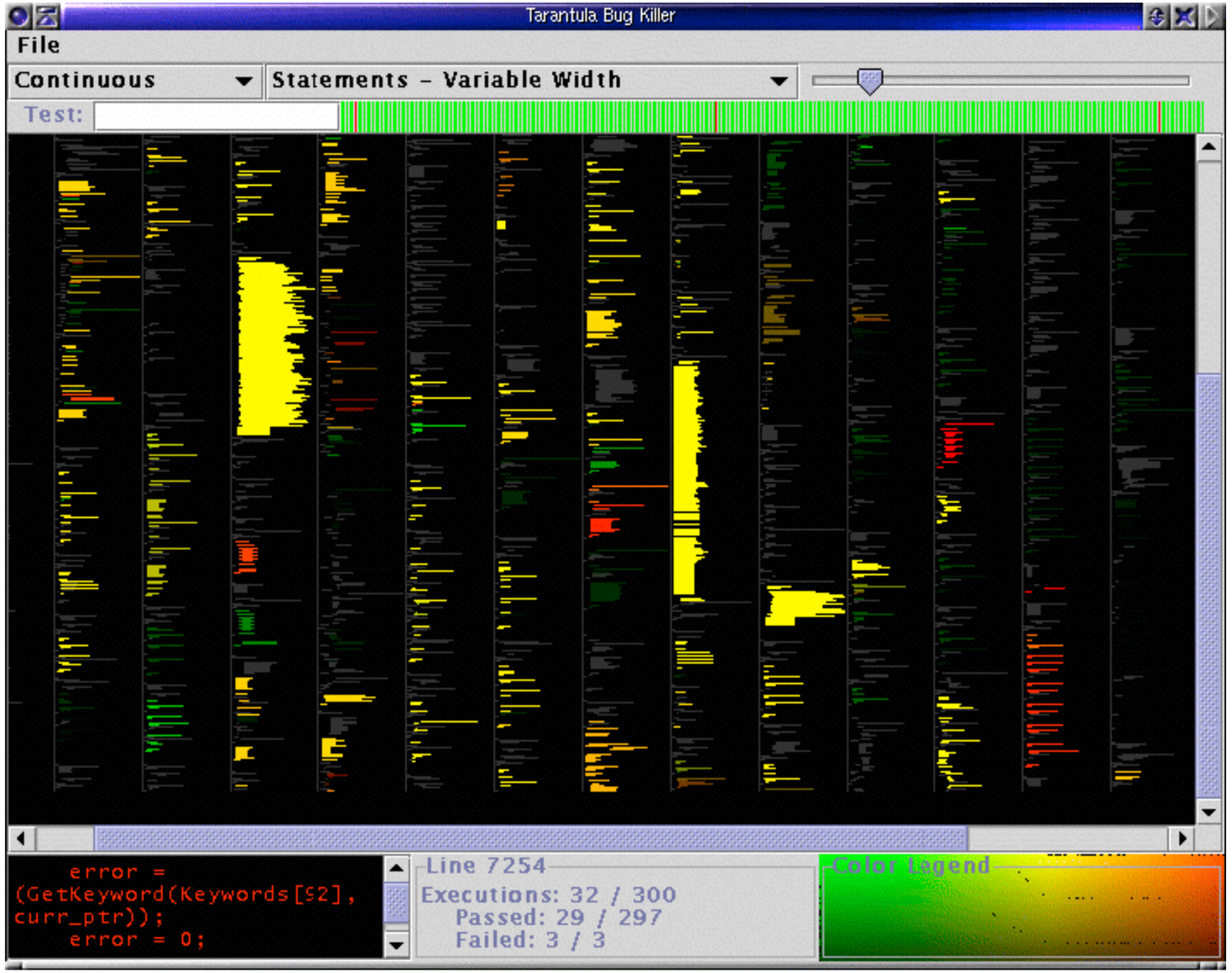
[Uncovering Strengths and Weaknesses of Radial Visualizations - an Empirical Approach. Diehl, Beck and Burch. IEEE TVCG (Proc. InfoVis) 16(6):935–942, 2010.]

Layout Density

- Dense

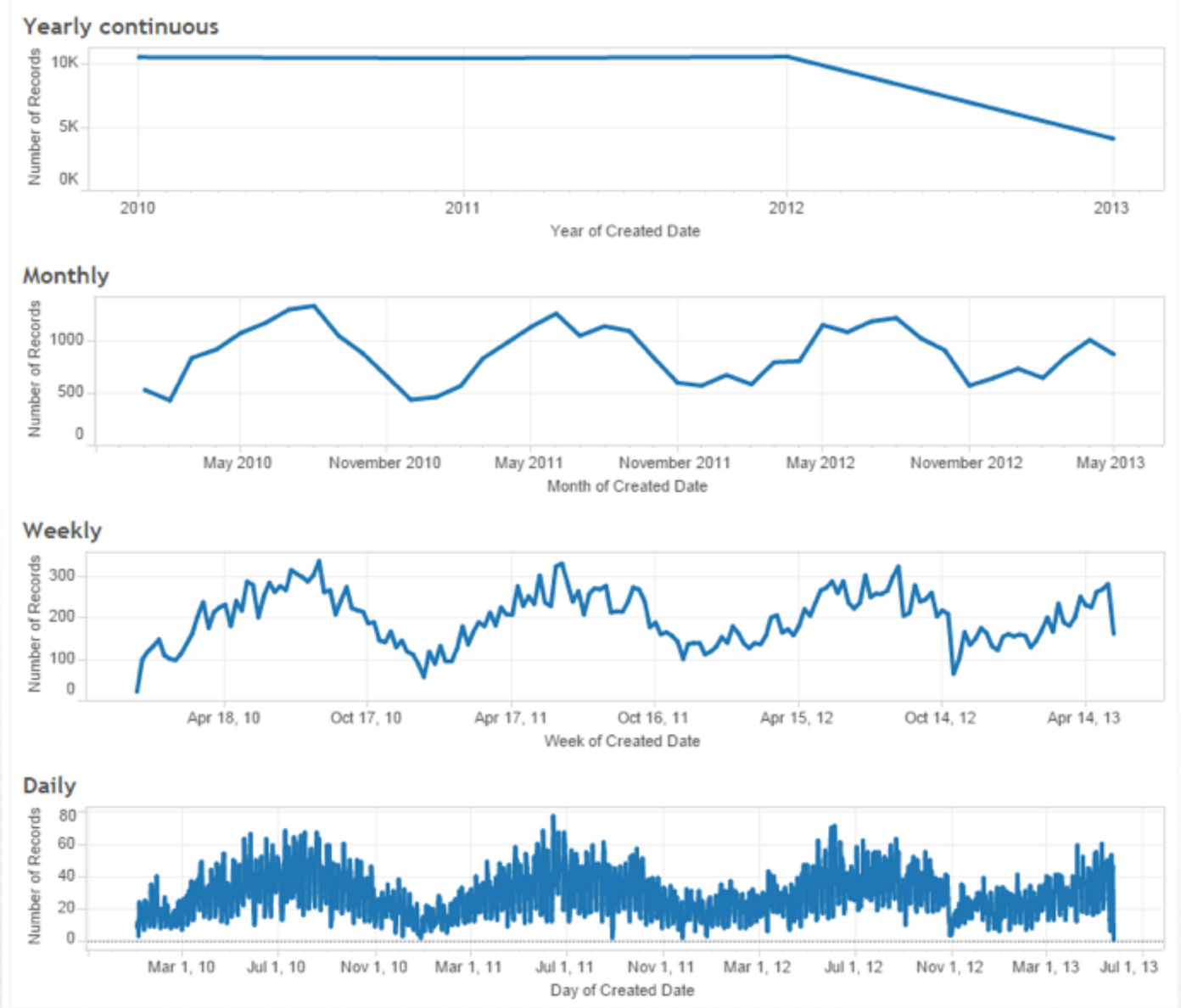


dense software overviews



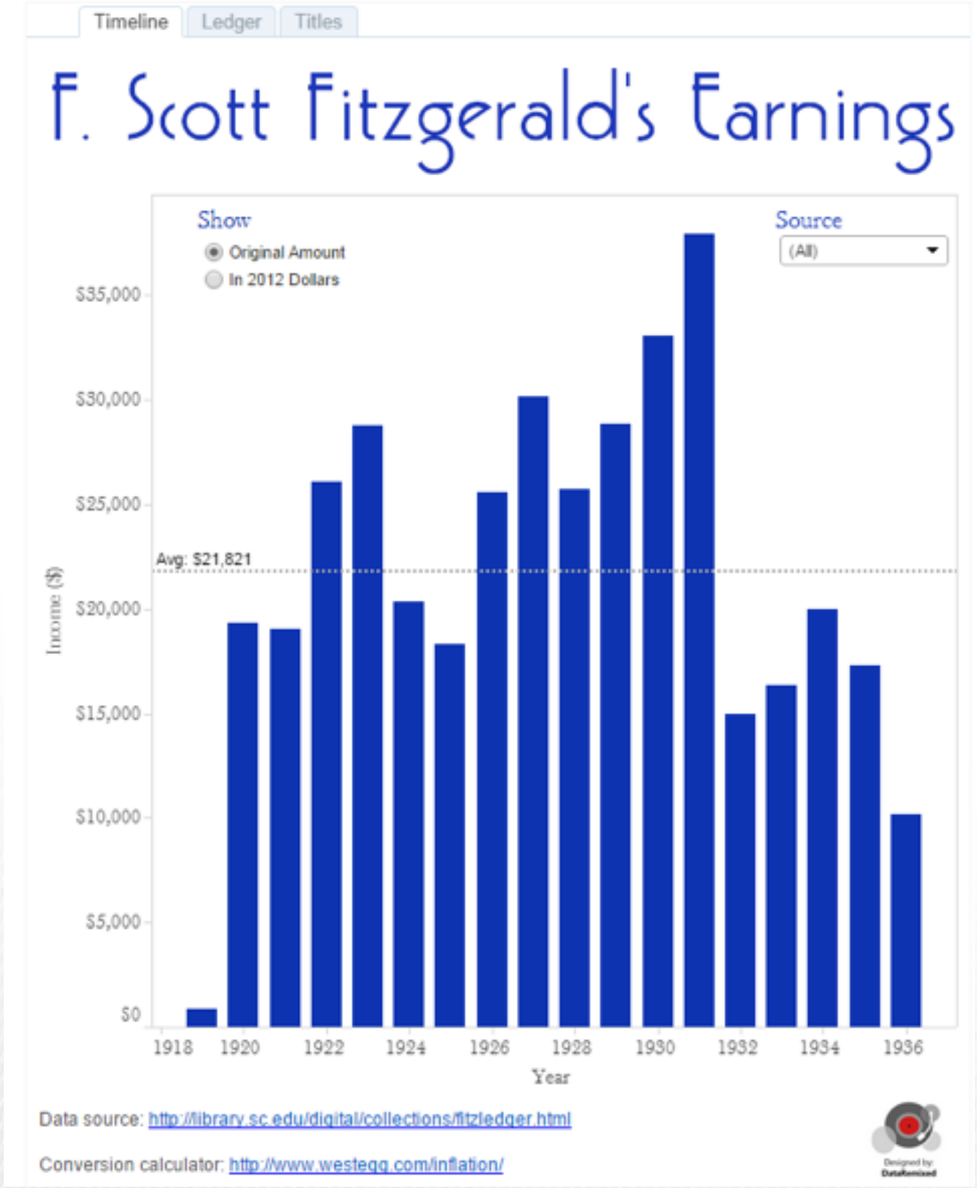
[Visualization of test information to assist fault localization. Jones, Harrold, Stasko. Proc. ICSE 2002, p 467-477.]

Basic Timelines – Working with Dates



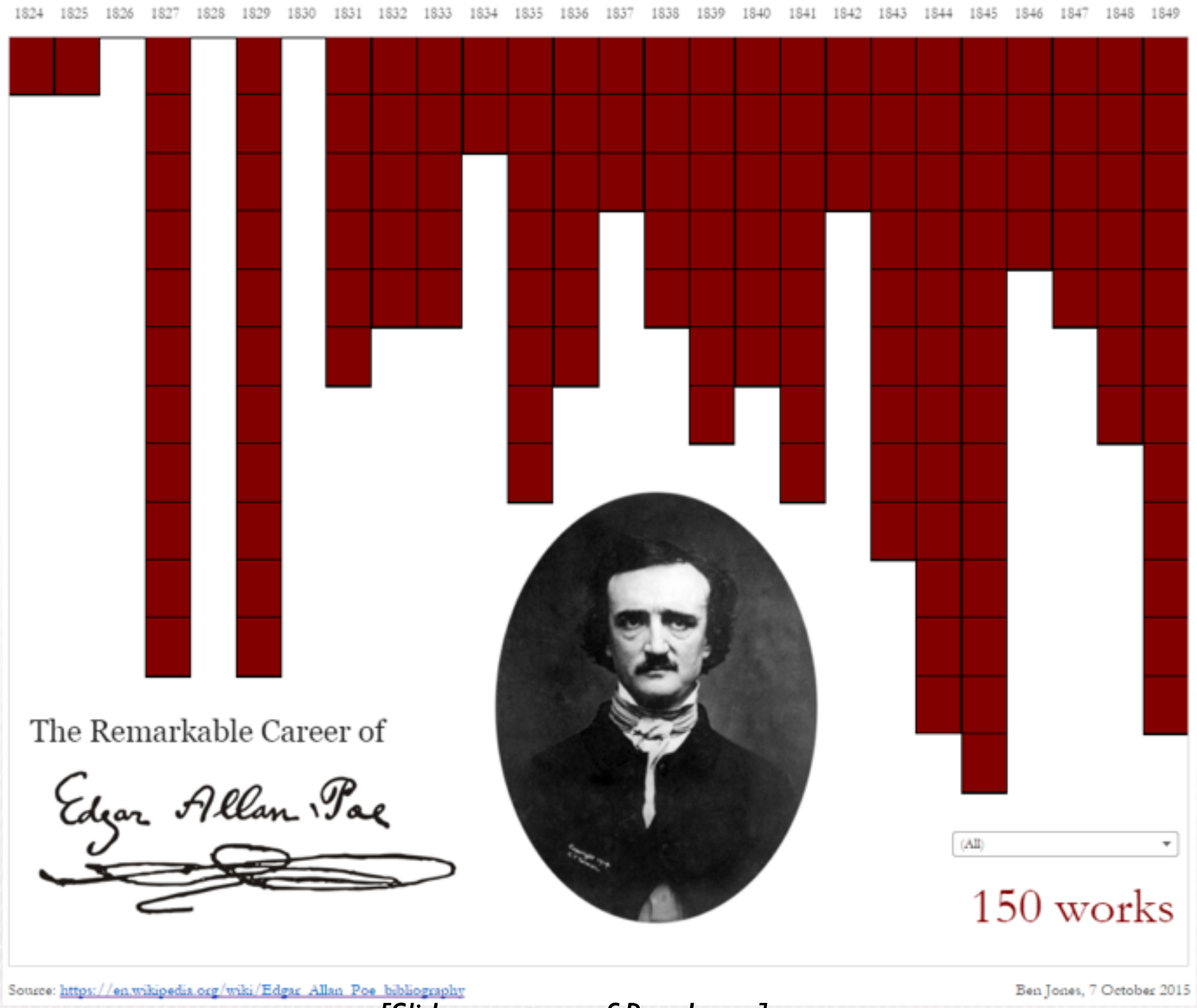
[Slide courtesy of Ben Jones]

Column Charts



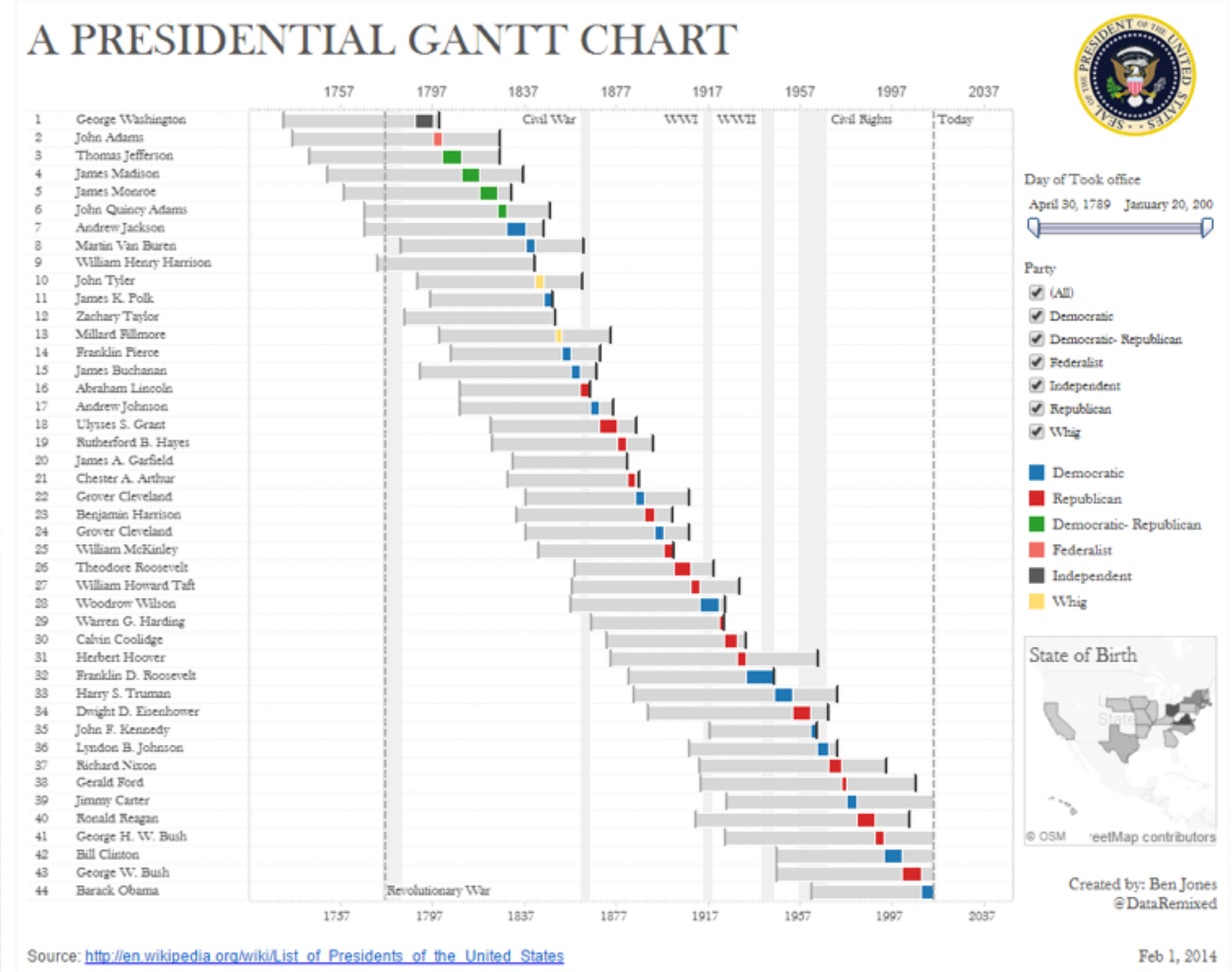
[Slide courtesy of Ben Jones]

Inverted Column Charts



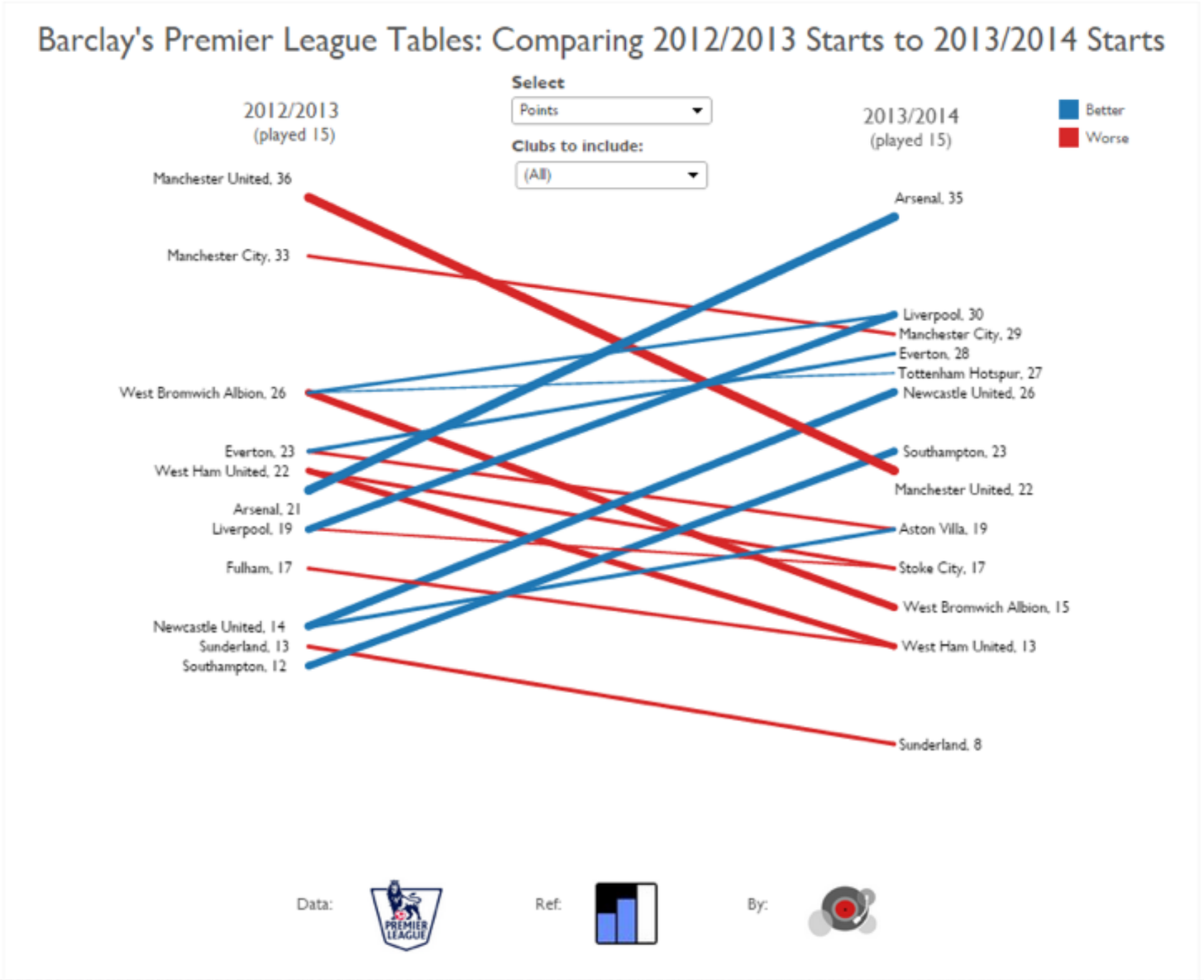
[Slide courtesy of Ben Jones]

Gantt Charts



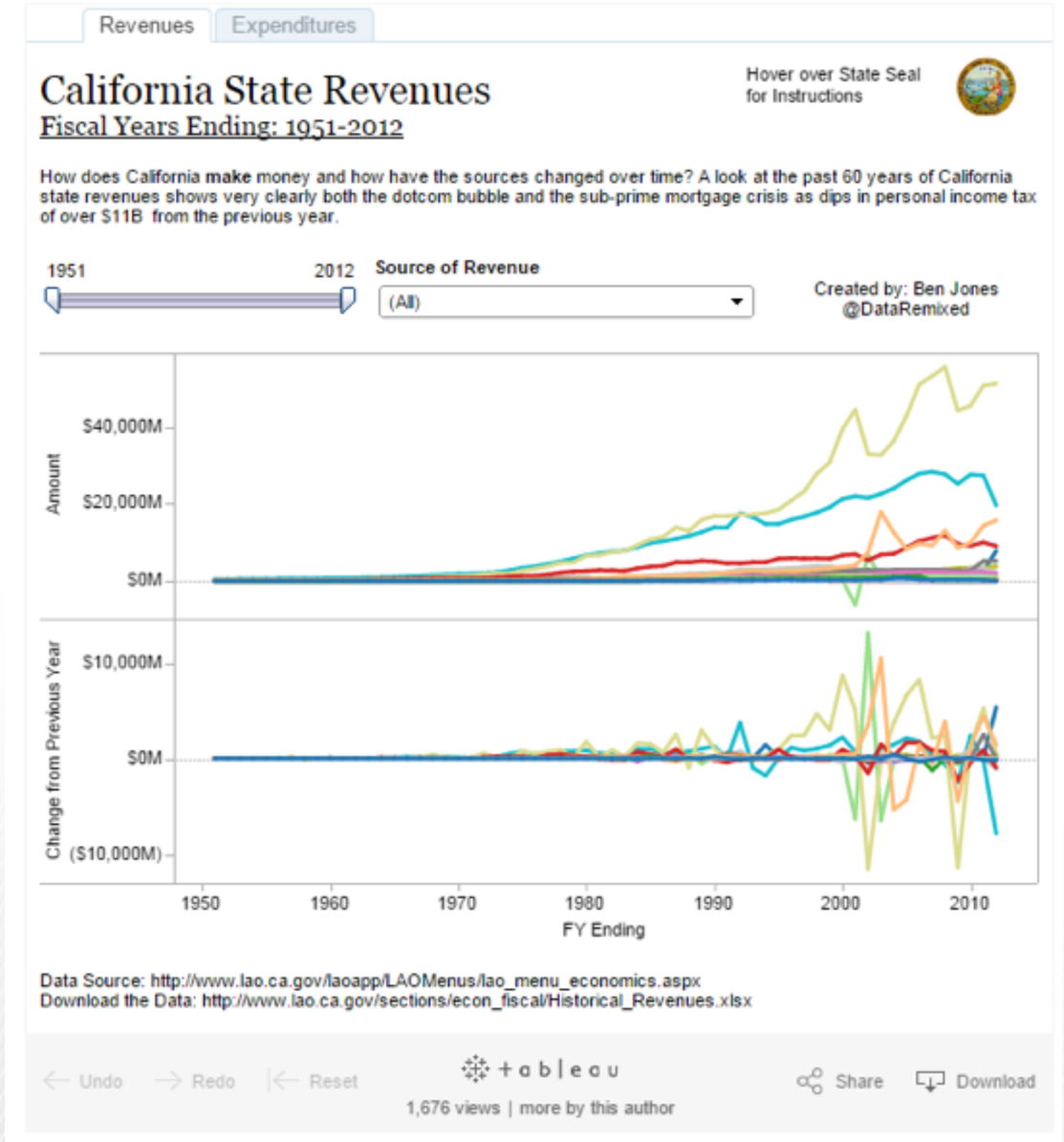
[Slide courtesy of Ben Jones]

Slopegraphs



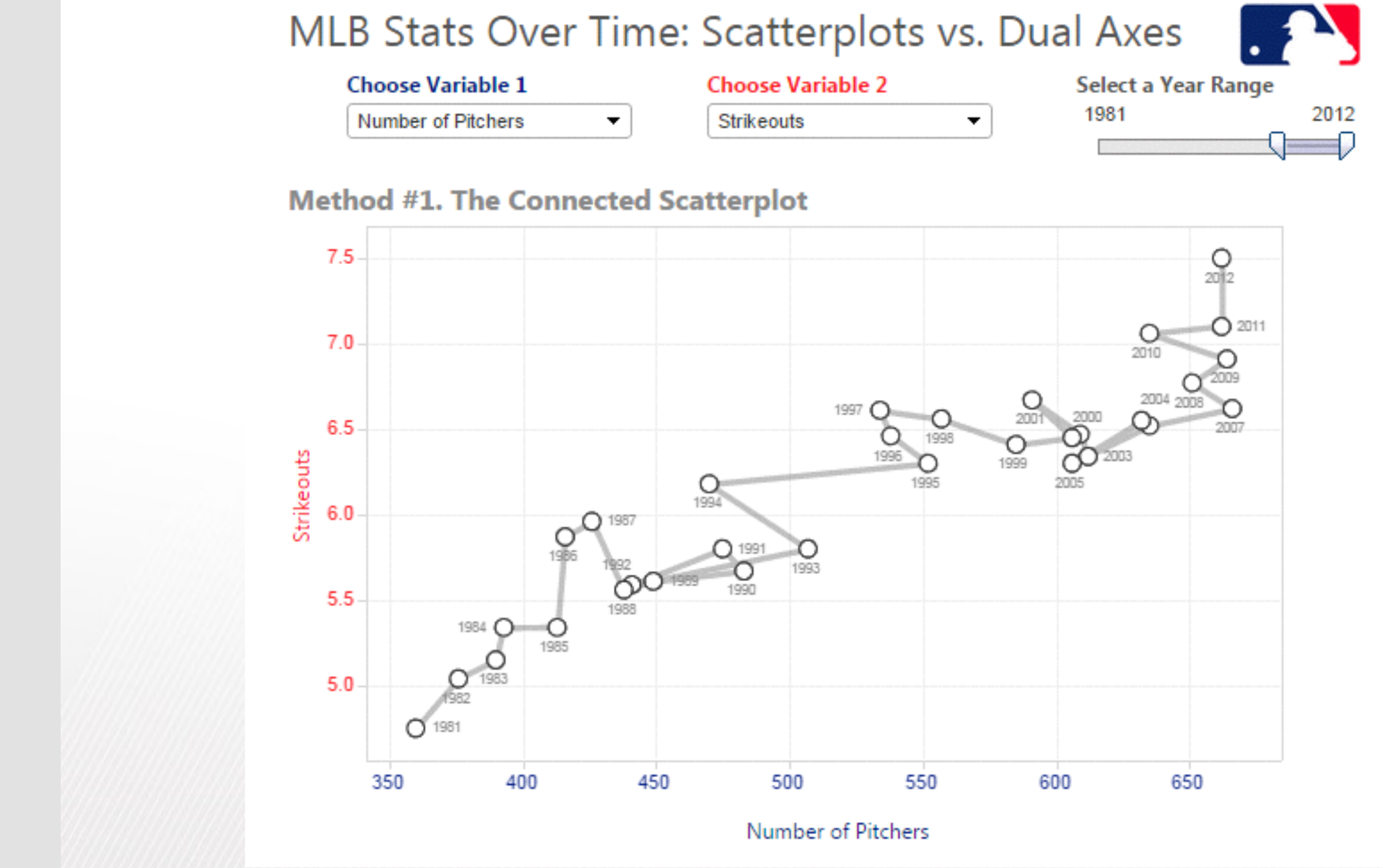
[Slide courtesy of Ben Jones]

Change from Previous



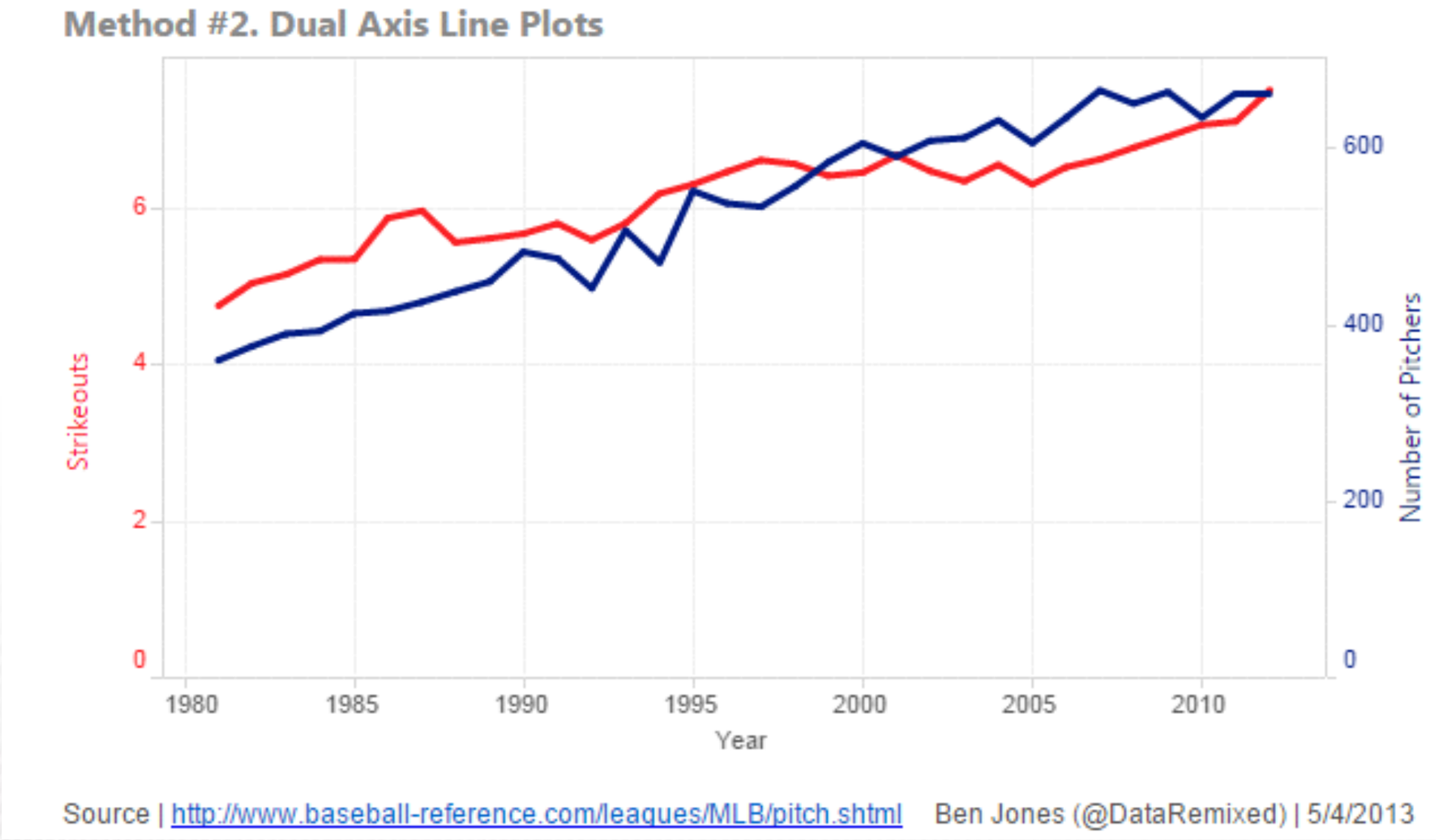
[Slide courtesy of Ben Jones]

Connected Scatterplots



[Slide courtesy of Ben Jones]

Dual Axis Line Plots



[Slide courtesy of Ben Jones]

Best Practices

- meaningful title
- axis labels
- include legend when necessary

Rules of Thumb

- No unjustified 3D

- Resolution over immersion
- Overview first, zoom and filter, details on demand
- Responsiveness is required
- Function first, form next

No unjustified 3D: Power of the plane

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

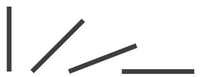
➔ **Magnitude Channels: Ordered Attributes**

Position on common scale

Position on unaligned scale

Length (1D size)

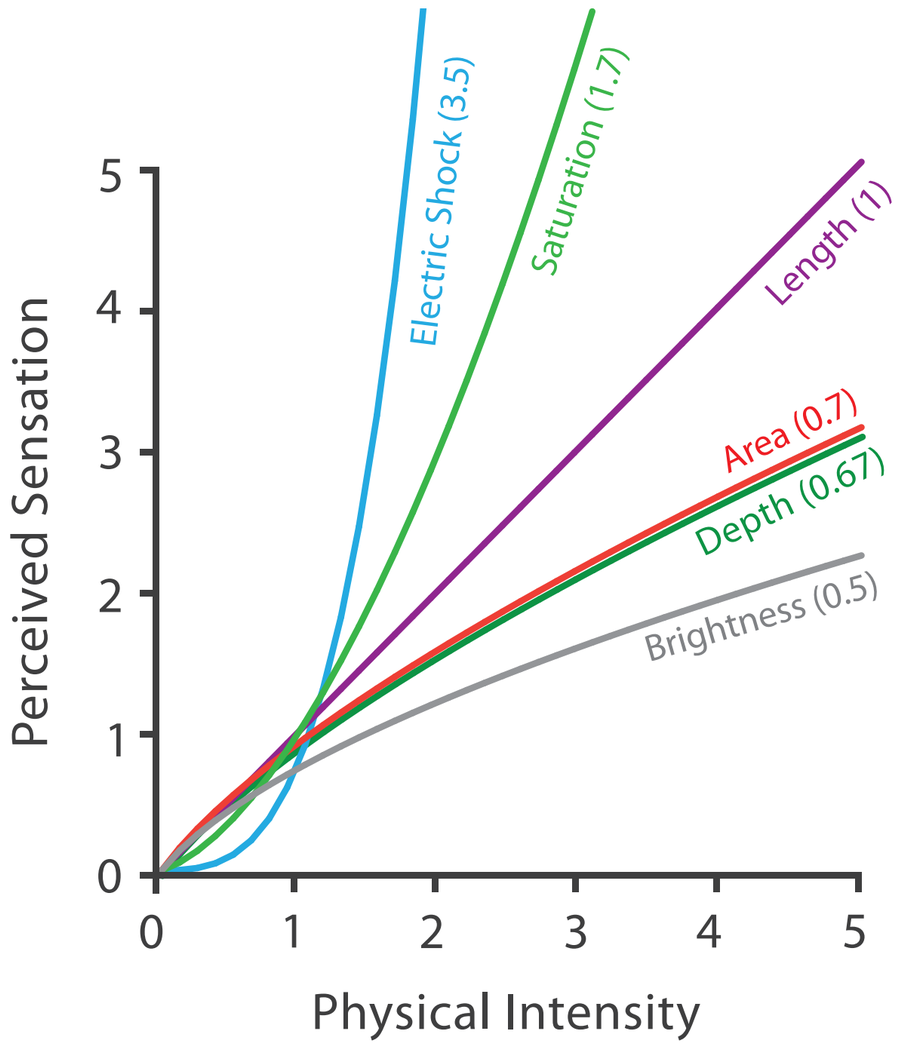
Tilt/angle



Area (2D size)

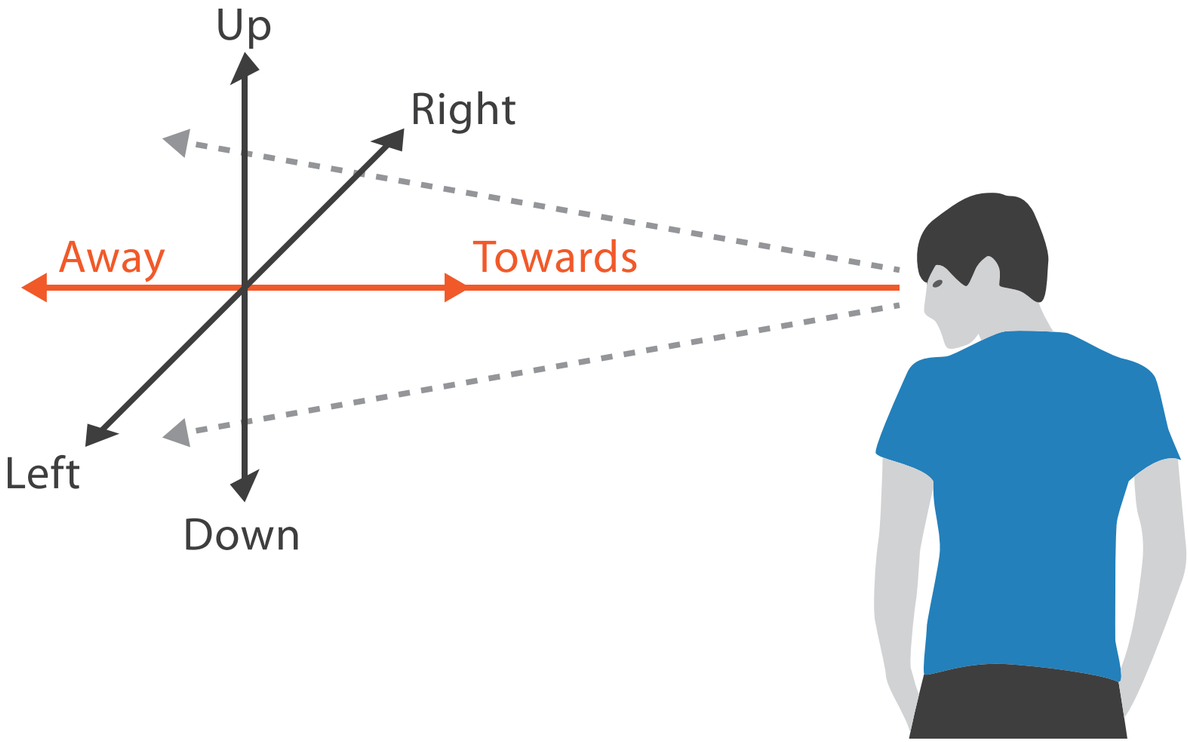
Depth (3D position)

Steven's Psychophysical Power Law: $S = I^N$

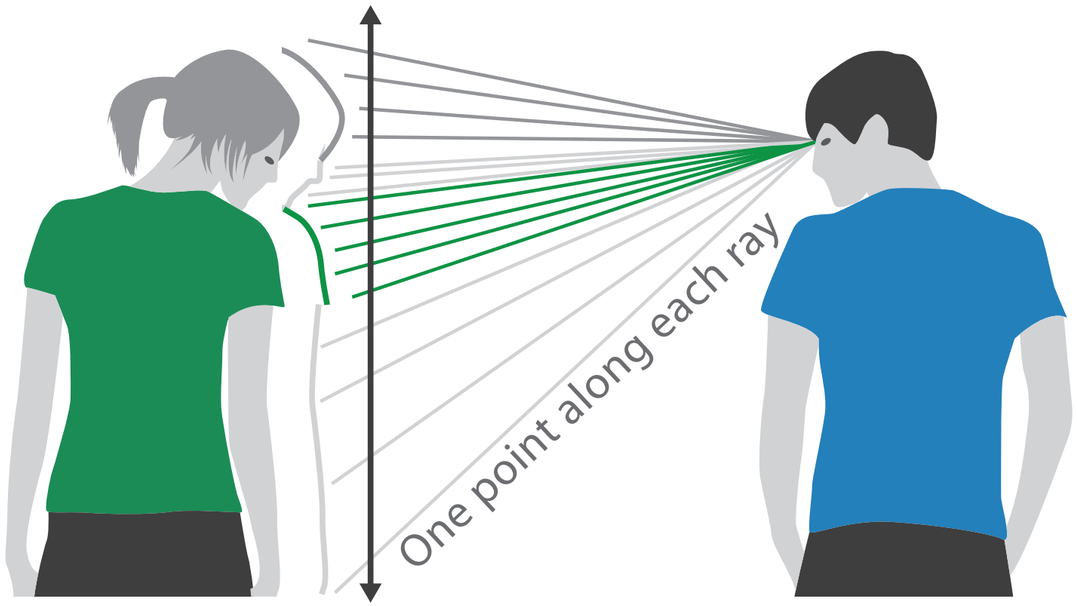


No unjustified 3D: Danger of depth

- 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



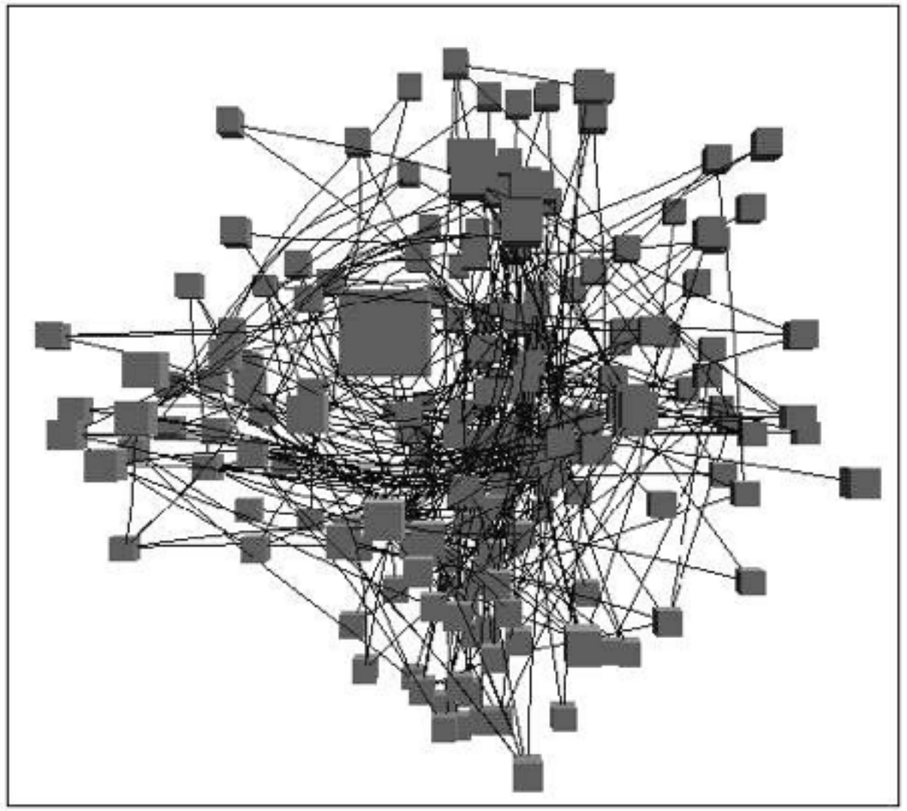
Thousands of points up/down and left/right



We can only see the outside shell of the world

Occlusion hides information

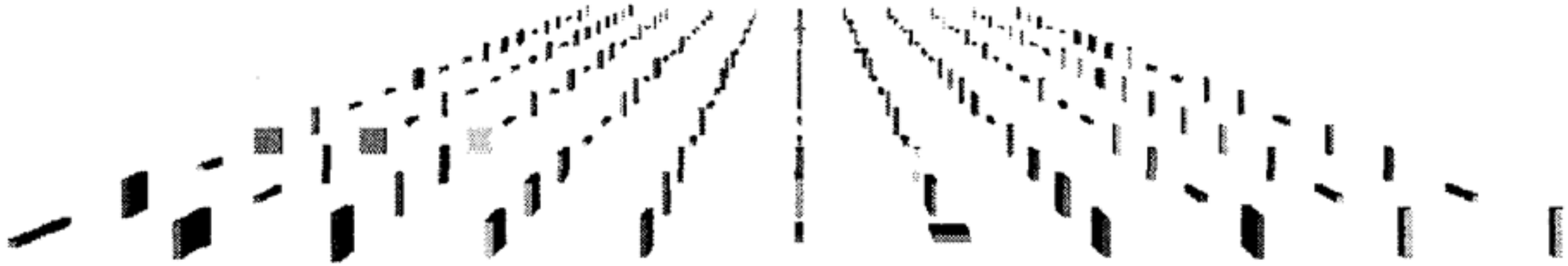
- occlusion
- interaction complexity



[Distortion Viewing Techniques for 3D Data. Carpendale et al. InfoVis1996.]

Perspective distortion loses information

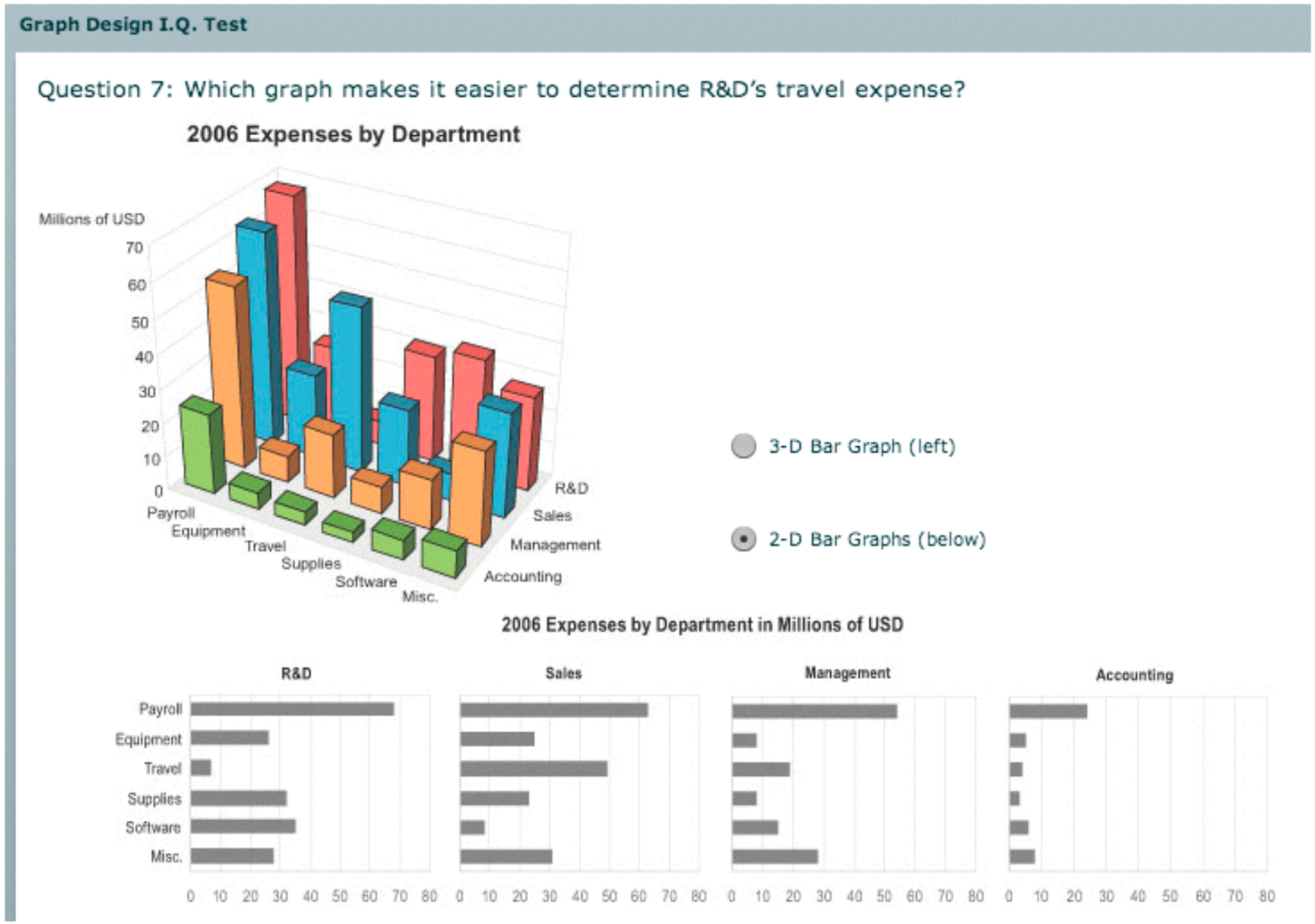
- perspective distortion
  - interferes with all size channel encodings
  - power of the plane is lost!



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

3D vs 2D bar charts

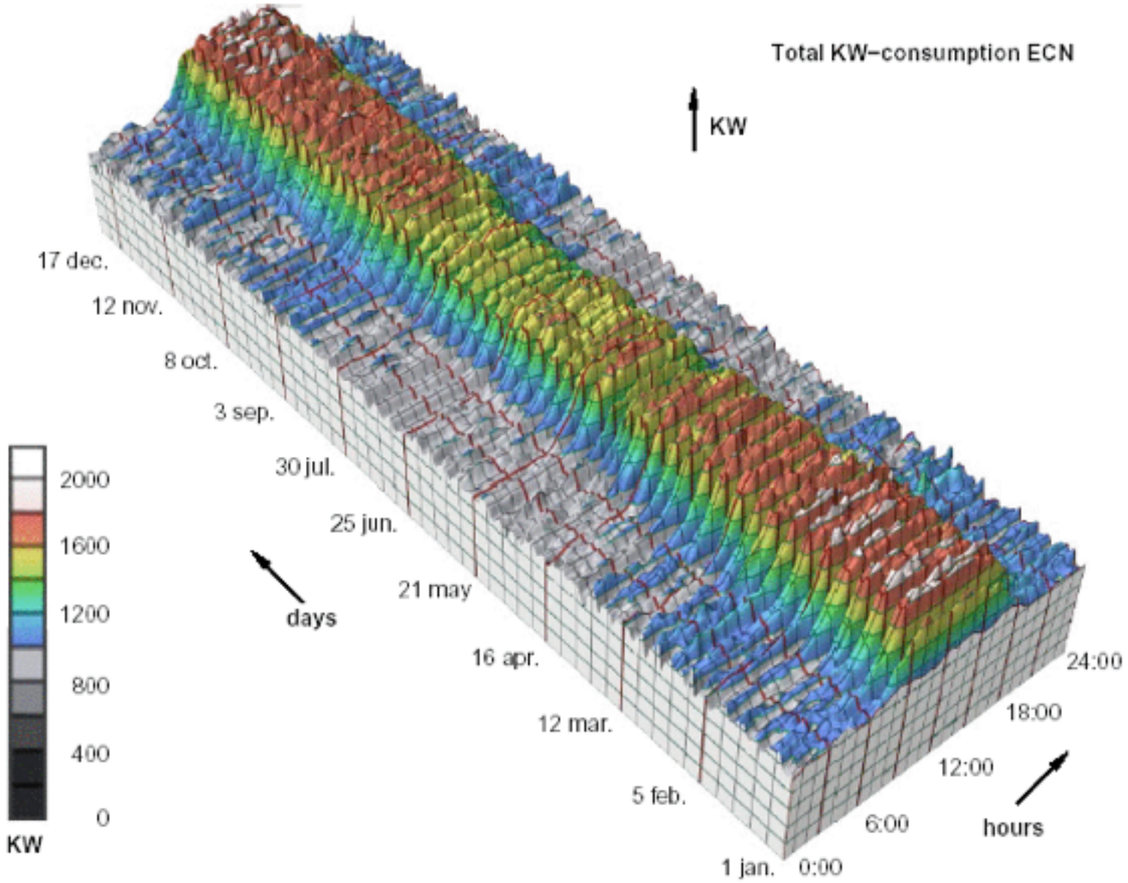
- 3D bars never a good idea!



[http://perceptualedge.com/files/GraphDesignIQ.html]

No unjustified 3D example: Time-series data

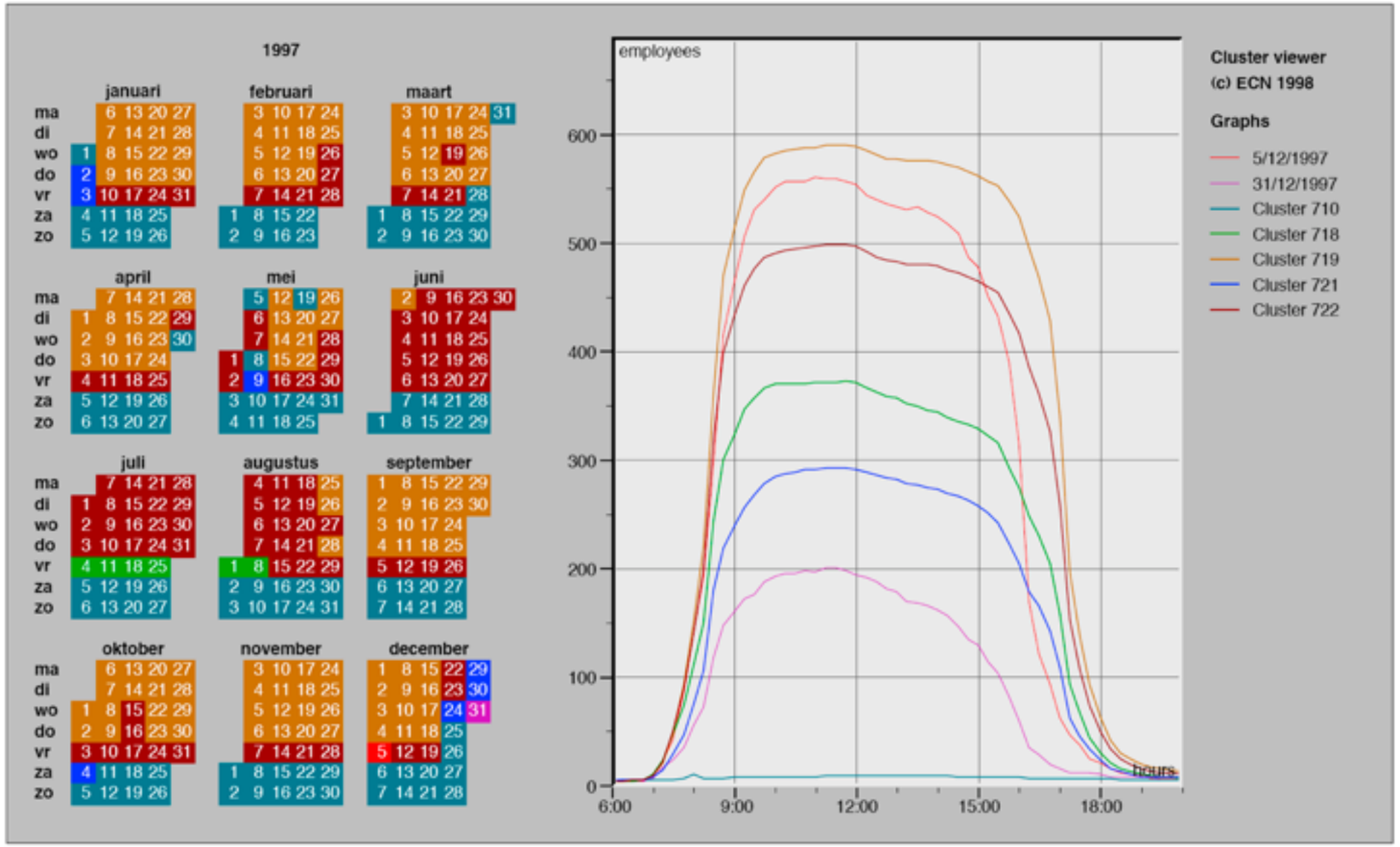
- extruded curves: detailed comparisons impossible



[Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc. InfoVis 99.]

No unjustified 3D example: Transform for new data abstraction

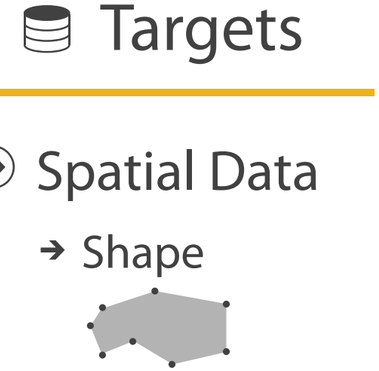
- derived data: cluster hierarchy
- juxtapose multiple views: calendar, superimposed 2D curves

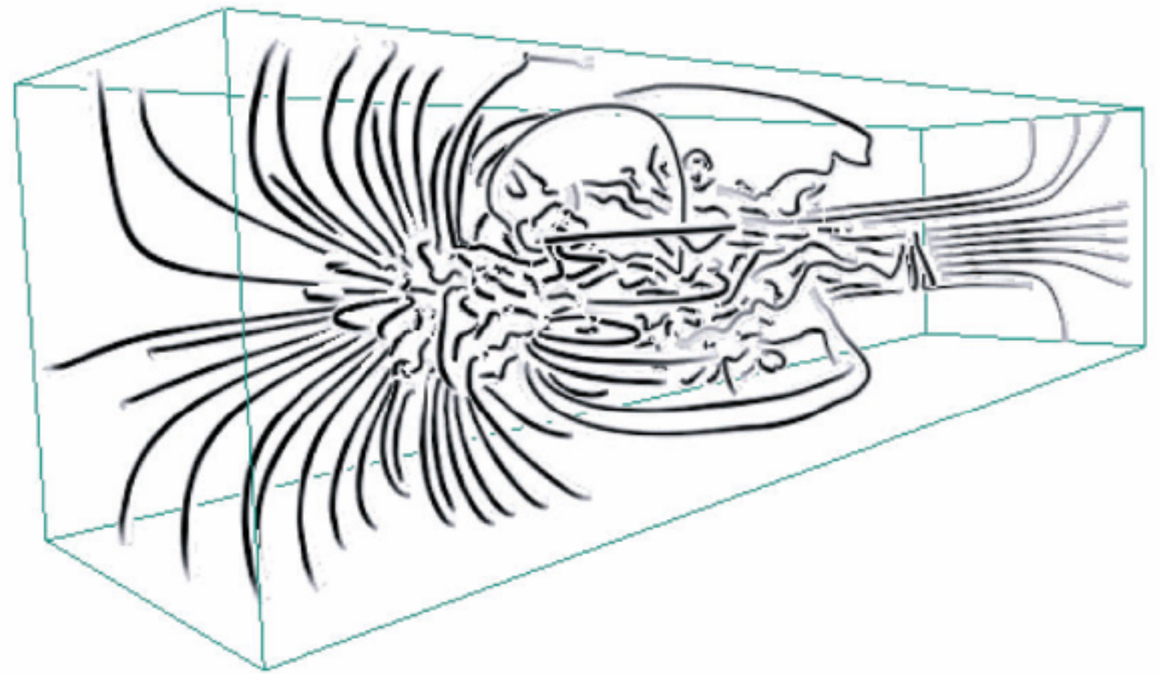


[Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc. InfoVis 99.]

Justified 3D: shape perception

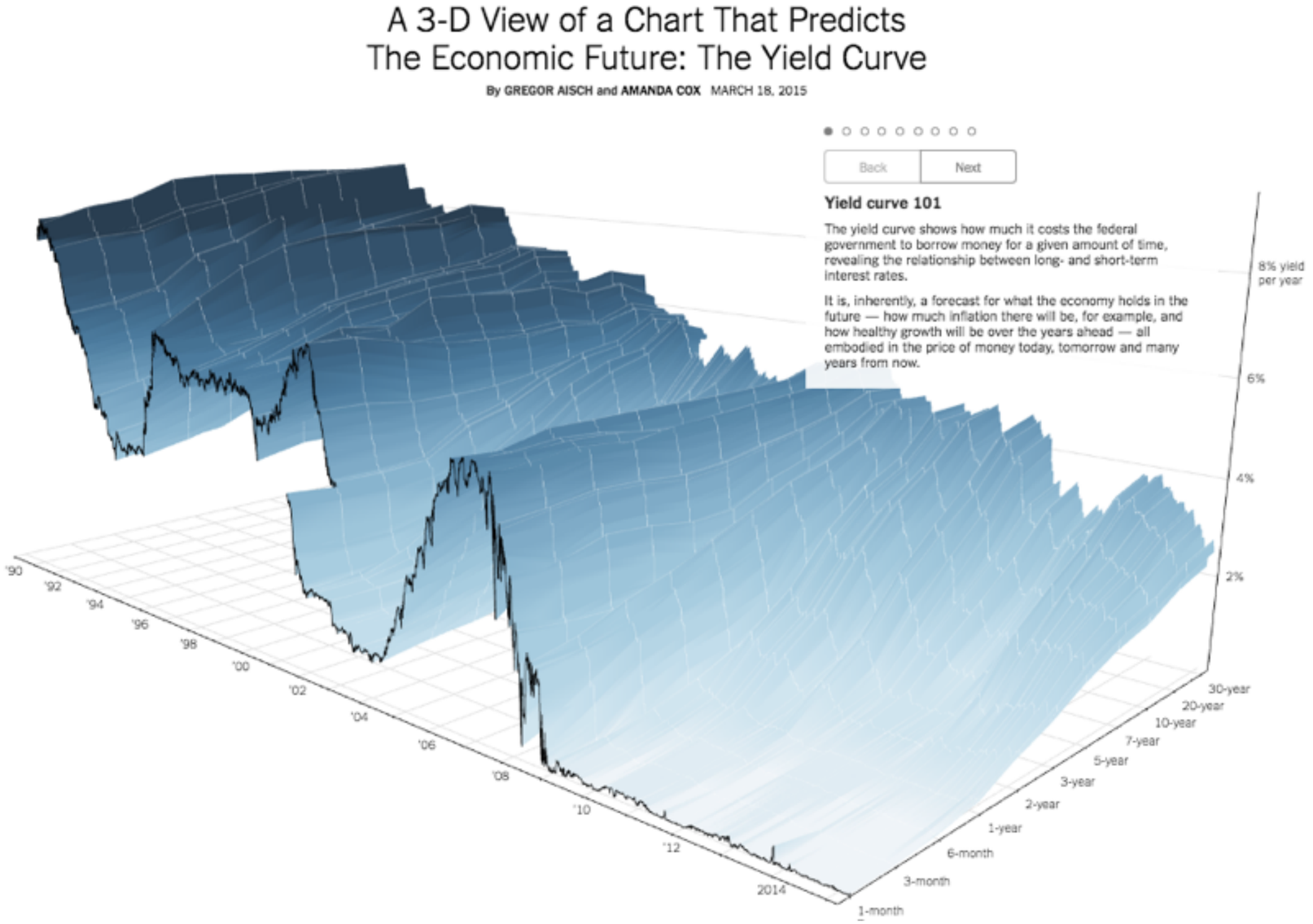
- benefits outweigh costs when task is shape perception for 3D spatial data
  - interactive navigation supports synthesis across many viewpoints





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

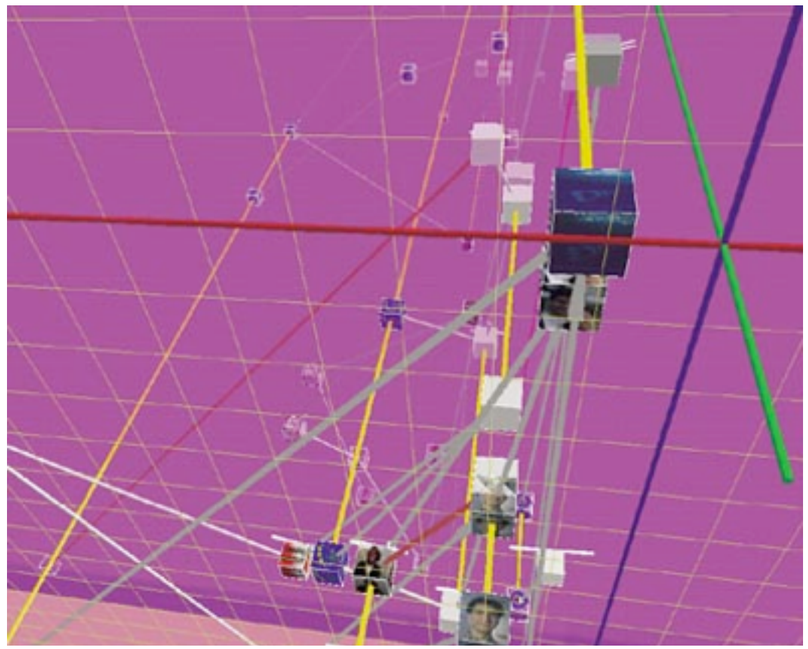
Justified 3D: Economic growth curve



http://www.nytimes.com/interactive/2015/03/19/upshot/3d-yield-curve-economic-growth.html

No unjustified 3D

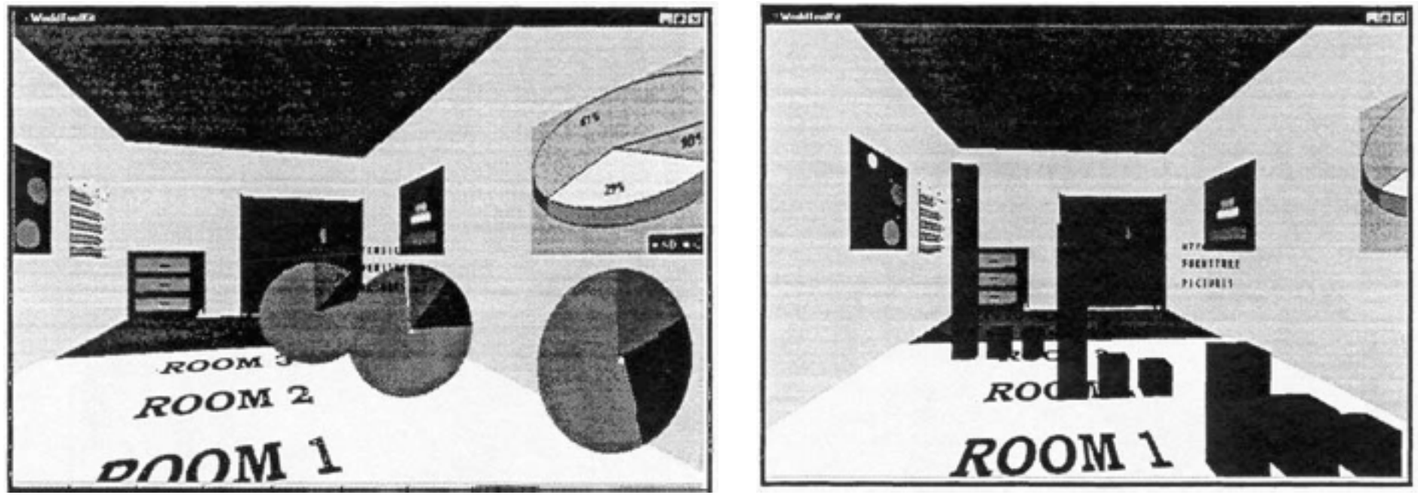
- 3D legitimate for true 3D spatial data
- 3D needs very careful justification for abstract data
  - enthusiasm in 1990s, but now skepticism
  - be especially careful with 3D for point clouds or networks



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

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



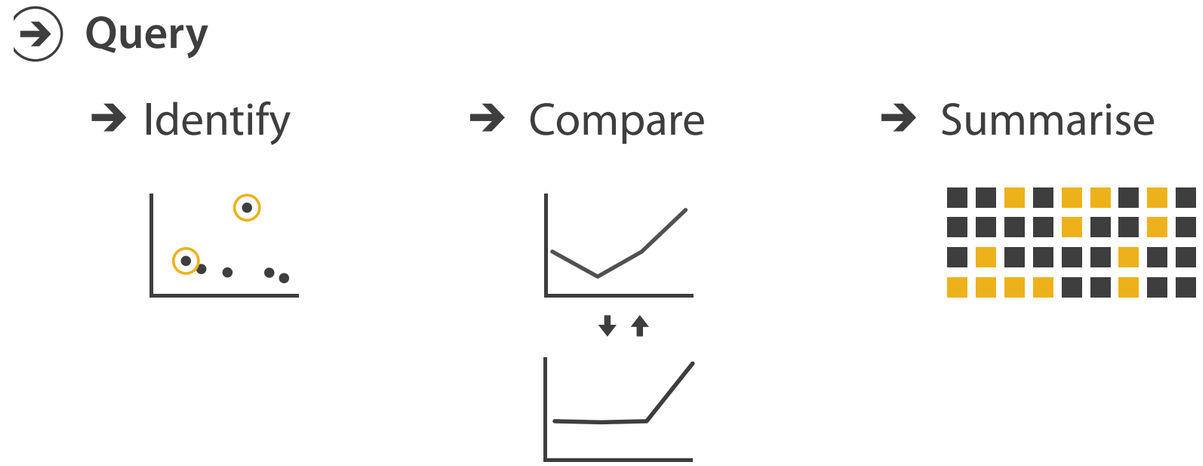
[Development of an information visualization tool using virtual reality. Kirner and Martins. Proc. Symp. Applied Computing 2000]

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



Responsiveness is required

- three major categories
  - 0.1 seconds: perceptual processing
  - 1 second: immediate response
  - 10 seconds: brief tasks
- importance of visual feedback

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