Lectures 3/4: Spatial Layout of Tables

Tamara Munzner
Department of Computer Science
University of British Columbia

DSCI 531: Data Visualization 1
Lecture 3: 23 September 2016
Lecture 4: 28 September 2016

https://github.ubc.ca/ubc-mds-2016/DSCI_531_viz-1_students

Recap Channel Rankings

Magnitude Channels: Ordered 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

Effectiveness: Best ▲ ▼ Least

Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape

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

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

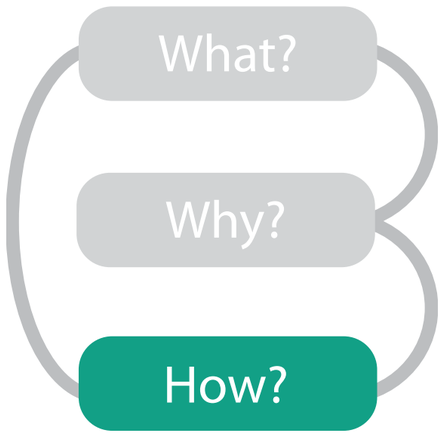
- **Change**
- **Select**
- **Navigate**

Facet

- **Juxtapose**
- **Partition**
- **Superimpose**

Reduce

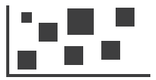
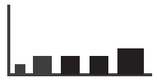
- **Filter**
- **Aggregate**
- **Embed**



- arrange: set spatial position

How?

Encode

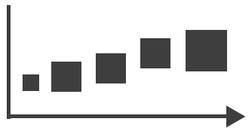
- **Arrange**
  - Express
    
  - Separate
    
  - Order
    
  - Align
    

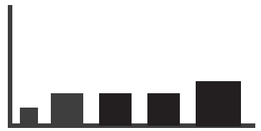
Encode tables: Arrange space

Encode

- **Arrange**
  - Express
  - Separate

  - Order
  - Align

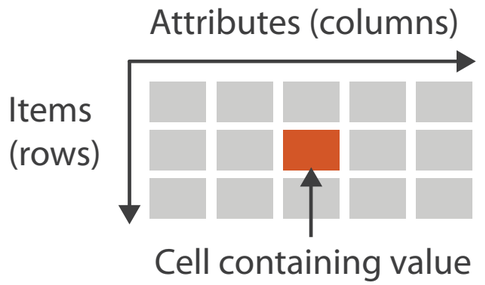




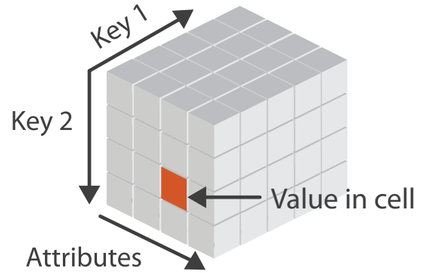
Keys and values

- 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



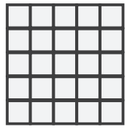
➔ *Multidimensional Table*



➔ **Express Values**

➔ *1 Key*
List

➔ *2 Keys*
Matrix



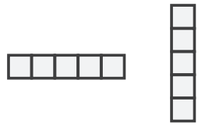
➔ *3 Keys*
Volume

➔ *Many Keys*
Recursive Subdivision

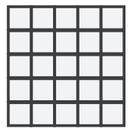
0 Keys

Express Values

→ *1 Key*
List



→ *2 Keys*
Matrix

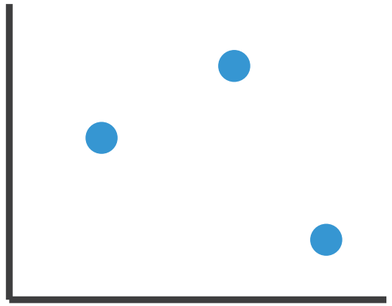


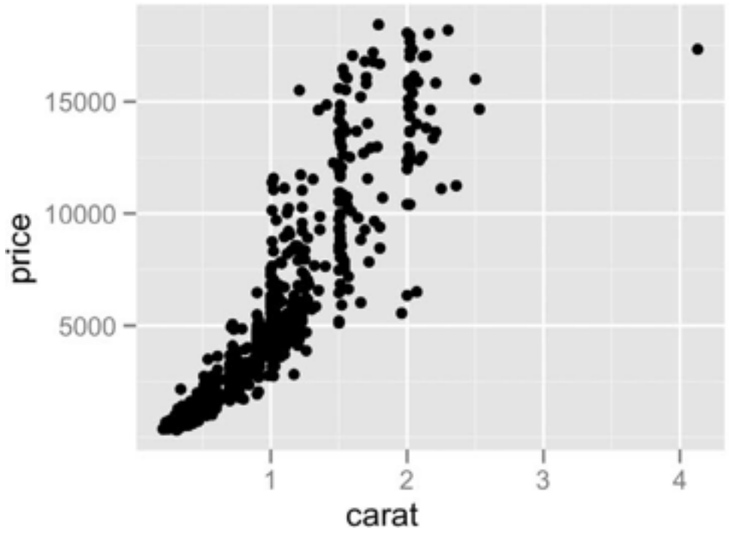
→ *3 Keys*
Volume

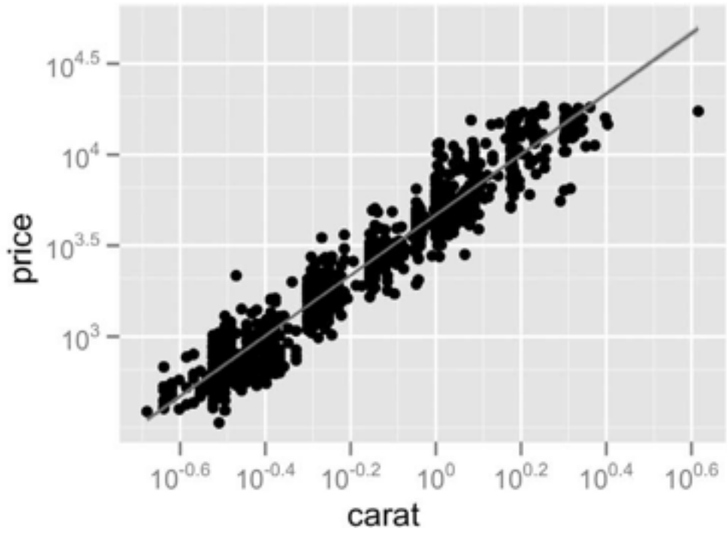
→ *Many Keys*
Recursive Subdivision

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





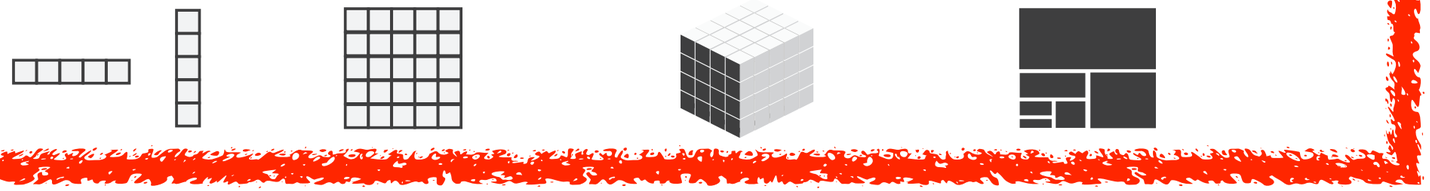


[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]

Some keys

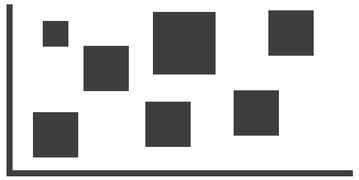
Express Values

- *1 Key*
  - *List*
- *2 Keys*
  - *Matrix*
- *3 Keys*
  - *Volume*
- *Many Keys*
  - *Recursive Subdivision*



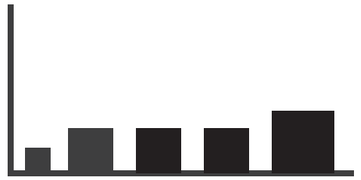
Some keys: Categorical regions

➔ Separate



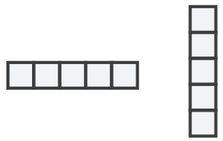
➔ Order

➔ Align

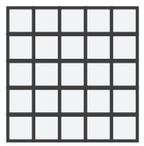


- **regions**: contiguous bounded areas distinct from each other
  - using space to ***separate*** (proximity)
  - following expressiveness principle for categorical attributes
- use ordered attribute to ***order*** and ***align*** regions

➔ *1 Key*
List



➔ *2 Keys*
Matrix

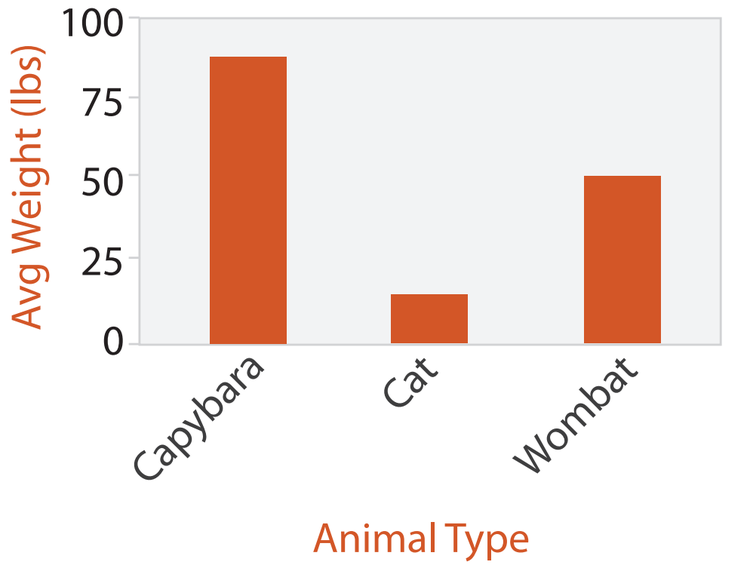


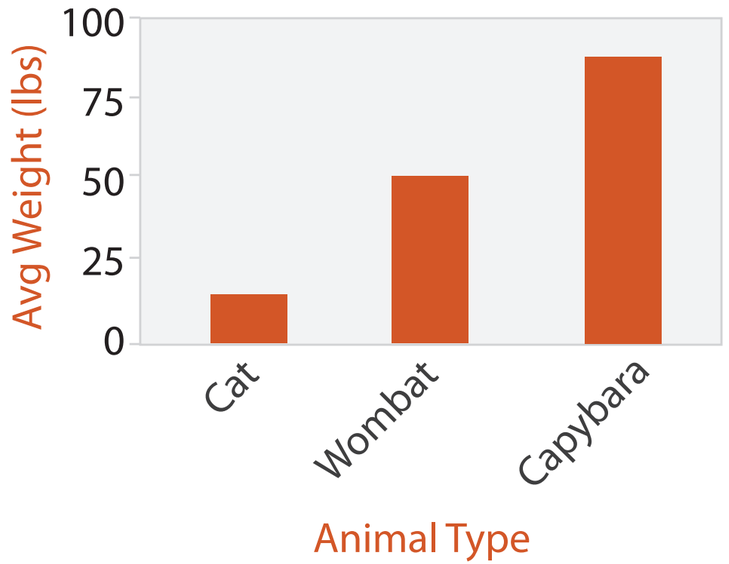
➔ *3 Keys*
Volume

➔ *Many Keys*
Recursive Subdivision

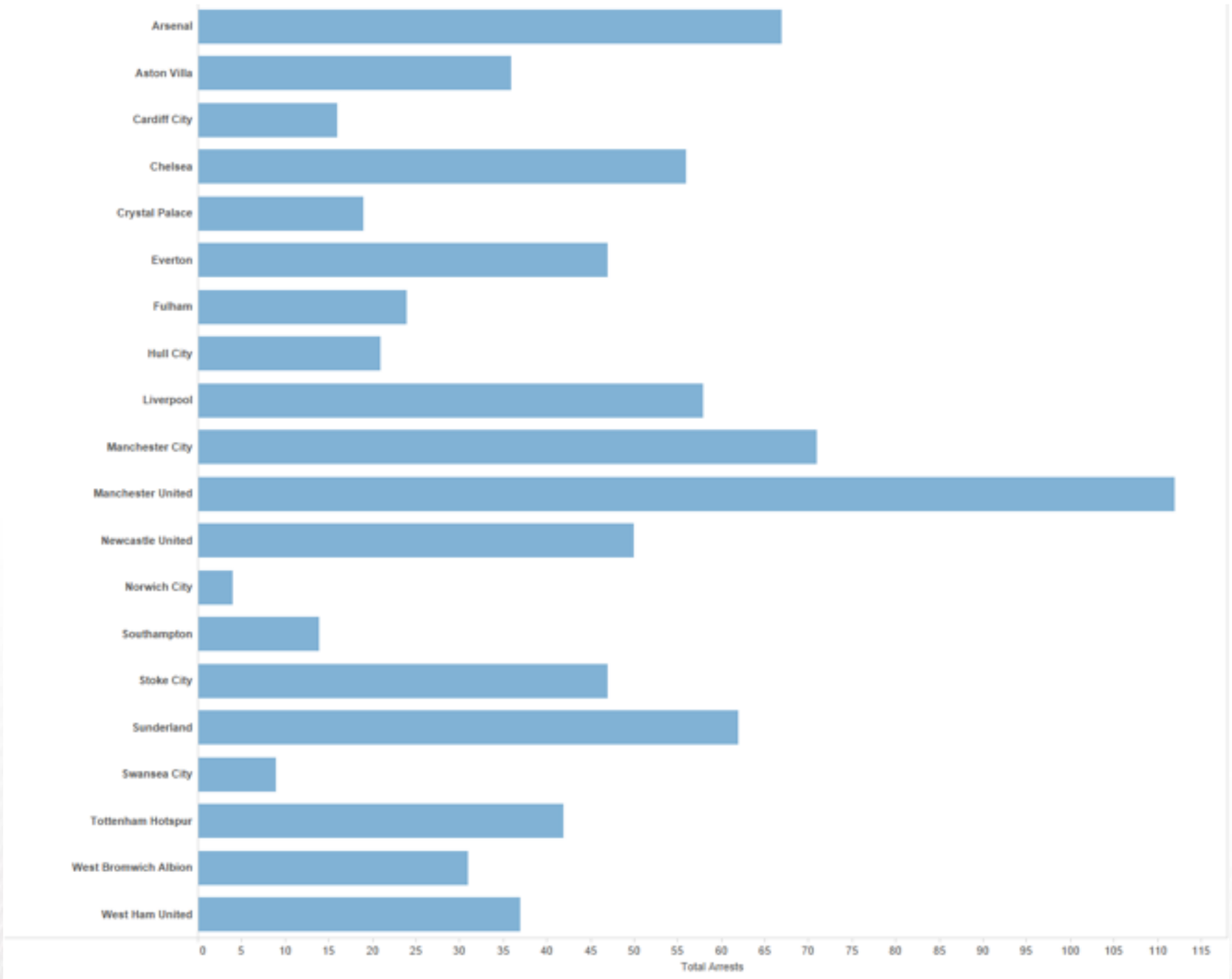
Idiom: **bar chart**

- 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





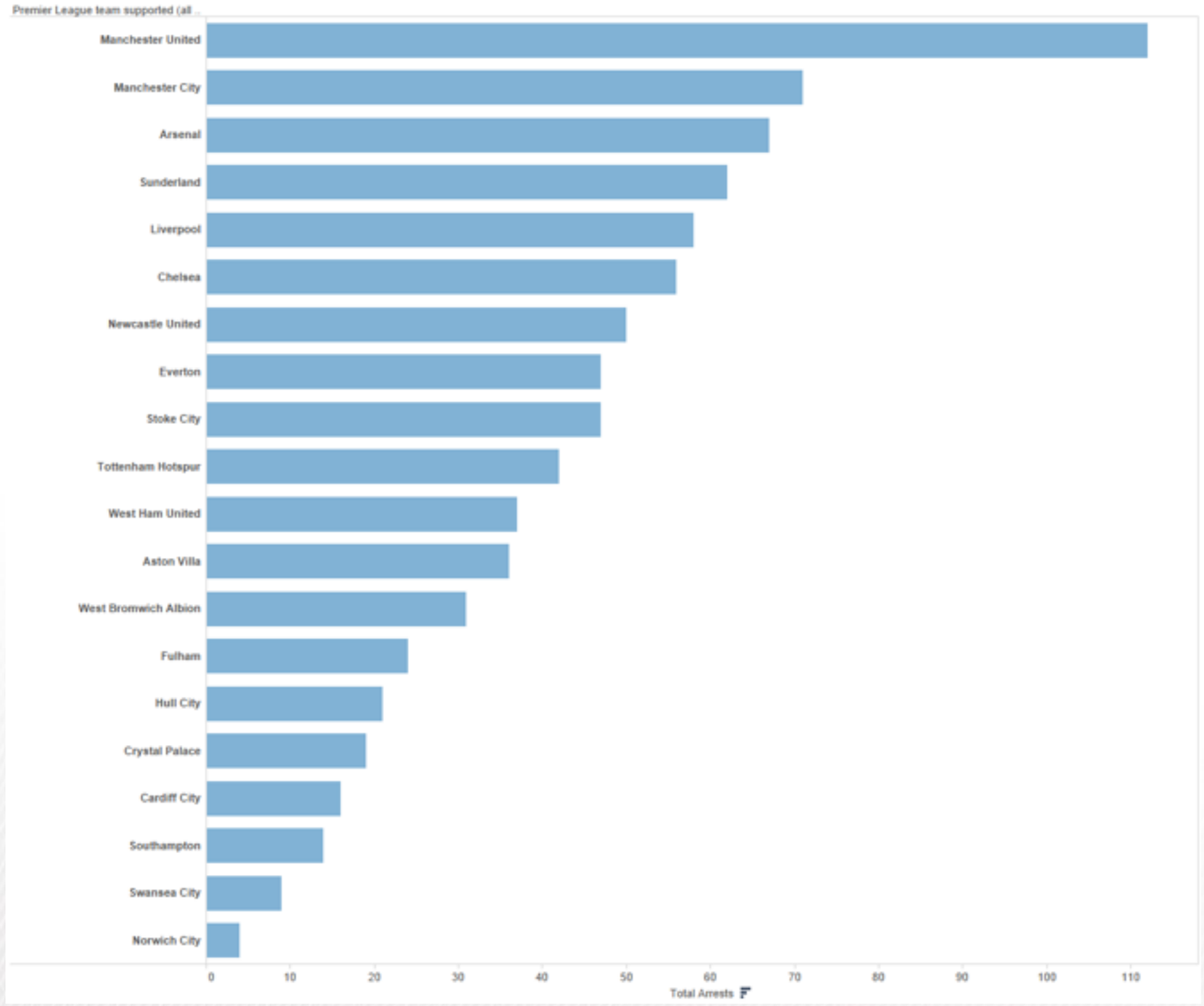
Separated and Aligned but not Ordered



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

[Slide courtesy of Ben Jones]

Separated, Aligned and Ordered



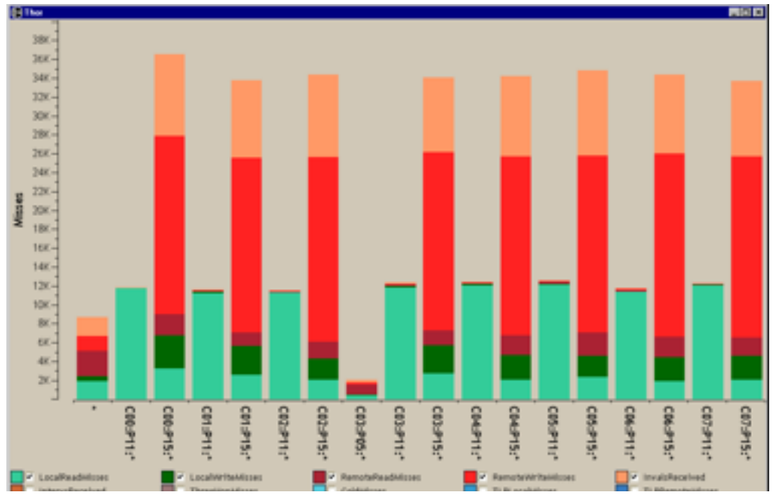
[Slide courtesy of Ben Jones]

Separated but not Ordered or Aligned

LIMITATION: Hard to make comparisons

[Slide courtesy of Ben Jones]

Idiom: **stacked bar chart**

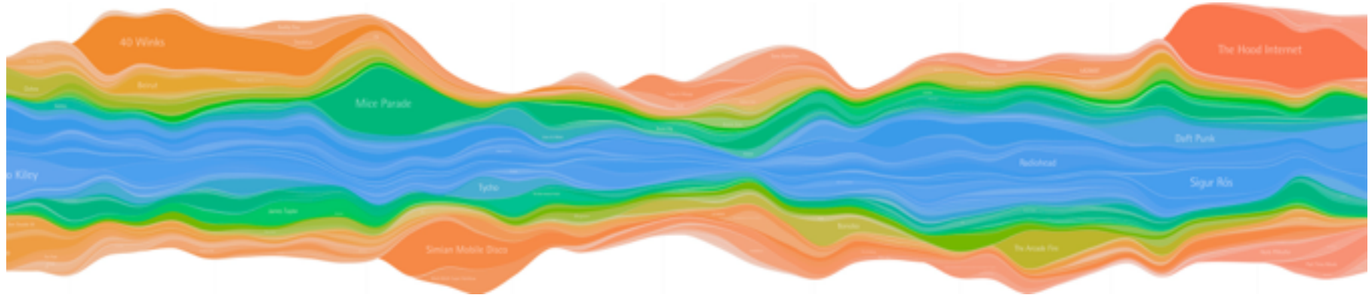


- one more key
  - –data
    - 2 categ attrib, 1 quant attrib
  - –mark: vertical stack of line marks
    - **glyph**: composite object, internal structure from multiple marks
  - –channels
    - length and color hue
    - spatial regions: one per glyph
      - – aligned: full glyph, lowest bar component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

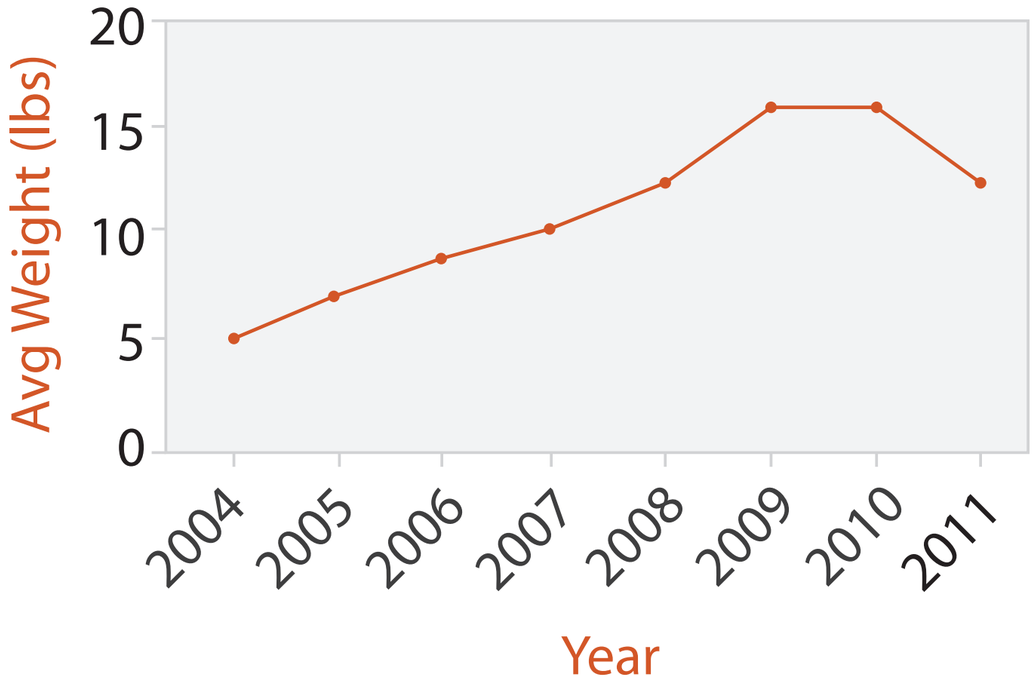
Idiom: **streamgraph**

- 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).]

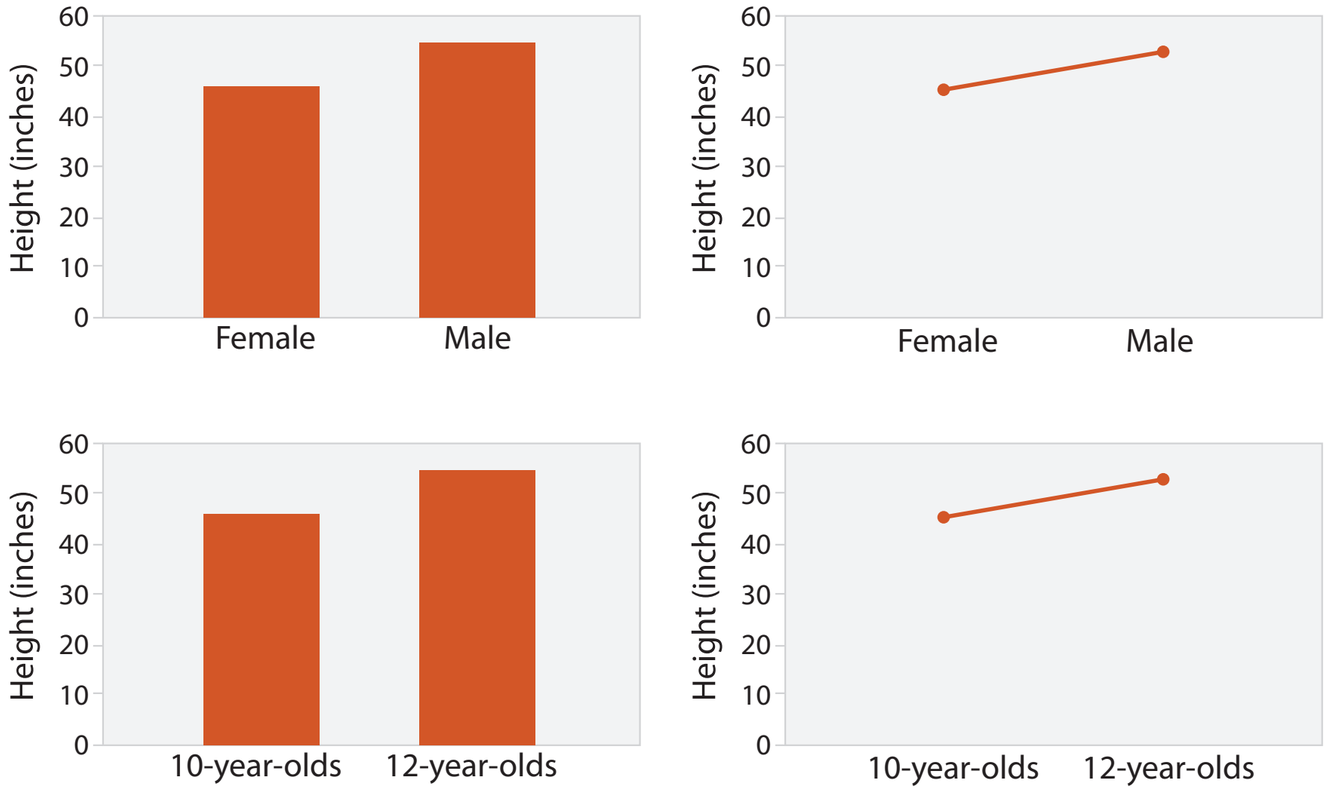
Idiom: **line chart / dot plot**



- 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
  - –scalability
    - hundreds of key levels, hundreds of value levels

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"

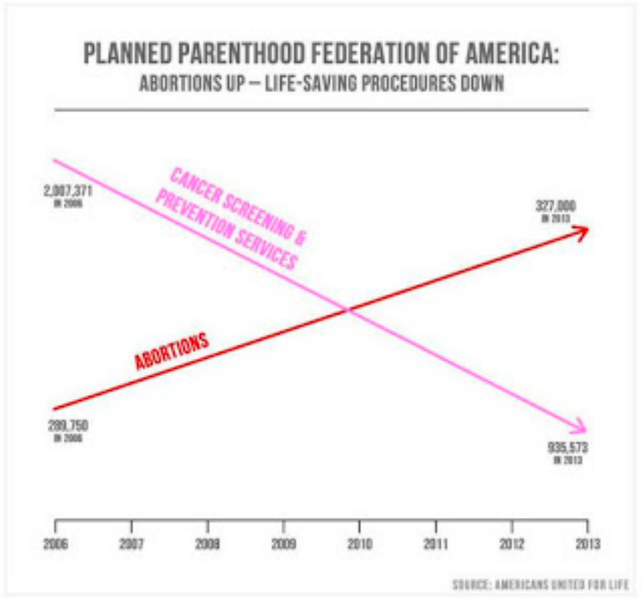


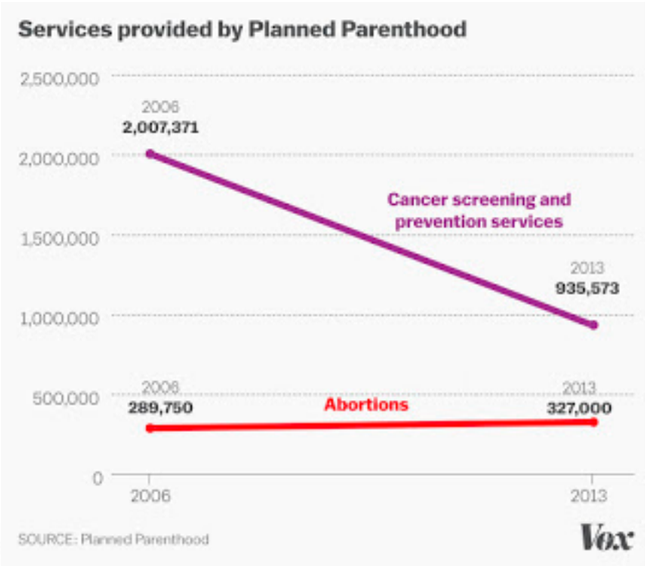
after [Bars and Lines: A Study of Graphic Communication. Zacks and Tversky. Memory and Cognition 27:6 (1999), 1073–1079.]

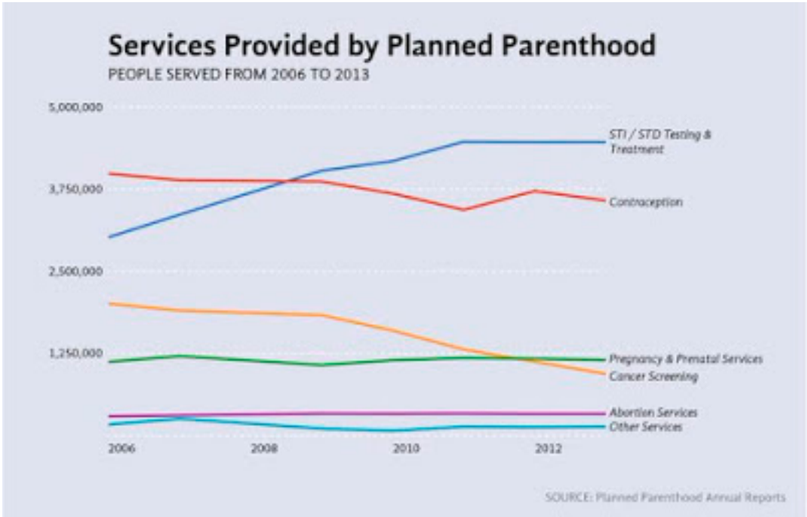
Chart axes

- labelled axis is critical
- avoid cropping y-axis
  - include 0 at bottom left
  - or slope misleads

- dual axes controversial
  - acceptable if commensurate
  - beware, very easy to mislead!



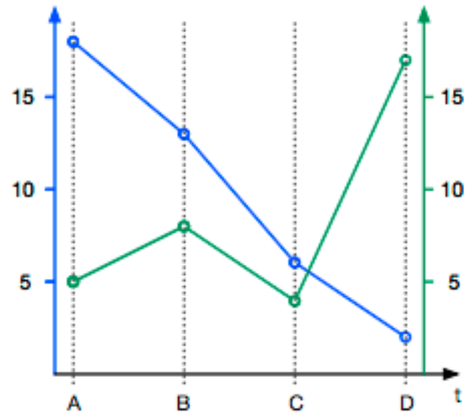


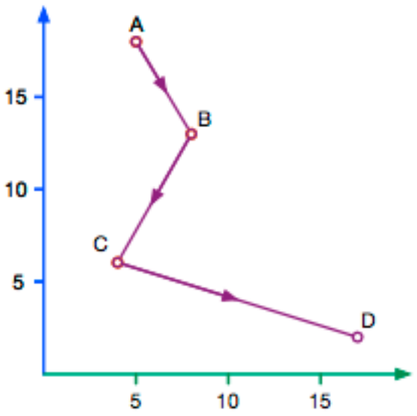


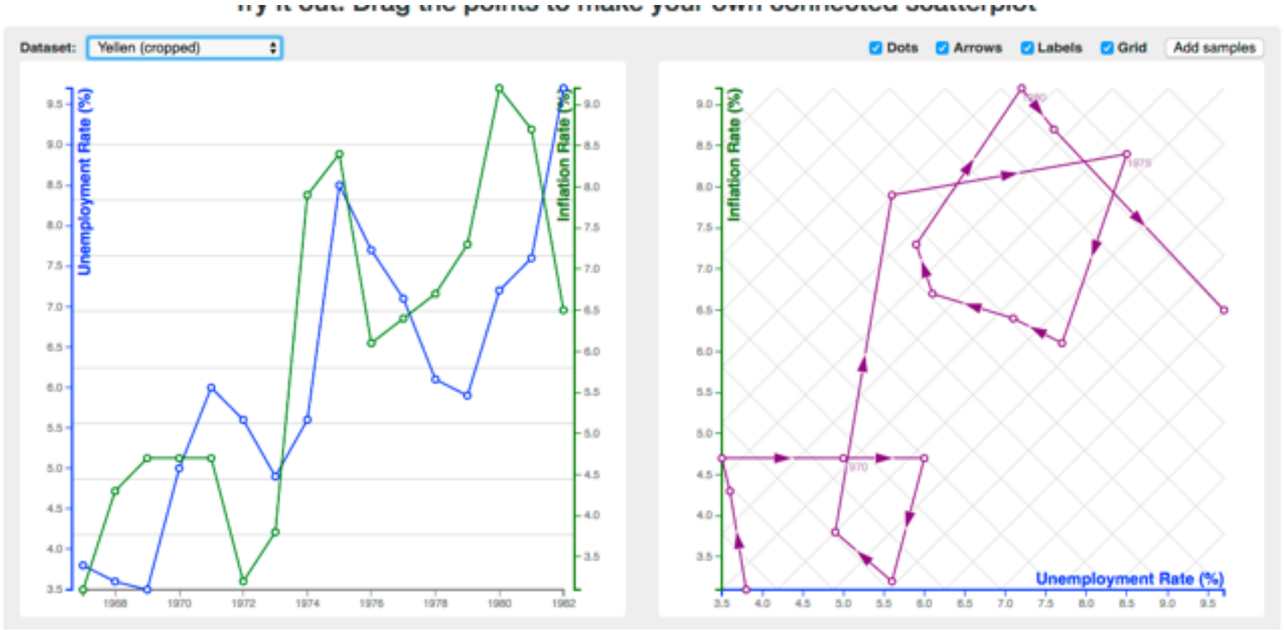
http://www.thefunctionalart.com/2015/10/if-you-see-bullshit-say-bullshit.html

Idiom: **connected scatterplots**

- scatterplot with line connection marks
  - popular in journalism
  - horiz + vert axes: value attribs
  - line connection marks: temporal order
  - alternative to dual-axis charts
    - horiz: time
    - vert: two value attribs
- empirical study
  - engaging, but correlation unclear







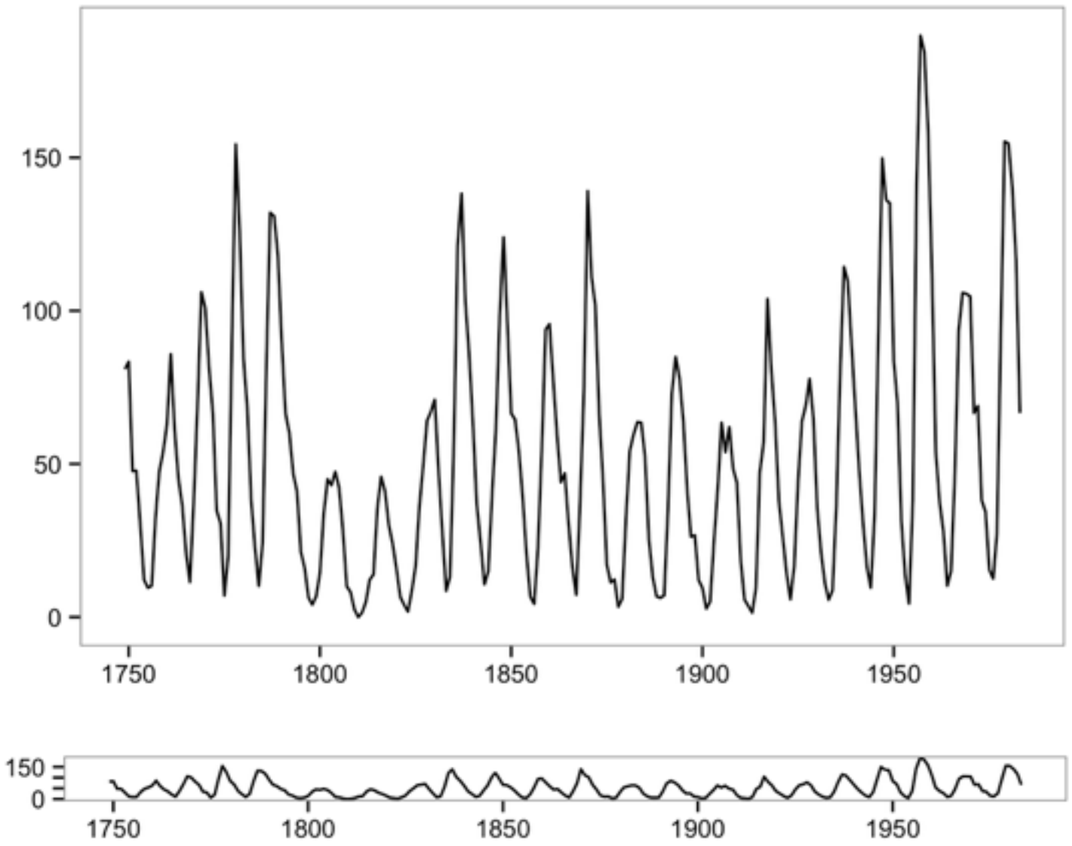
[The Connected Scatterplot for Presenting Paired Time Series. Haroz, Kosara and Franconeri. IEEE TVCG 22(9):2174-86, 2016.]

http://steveharoz.com/research/connected_scatterplot/

Choosing line chart aspect ratios

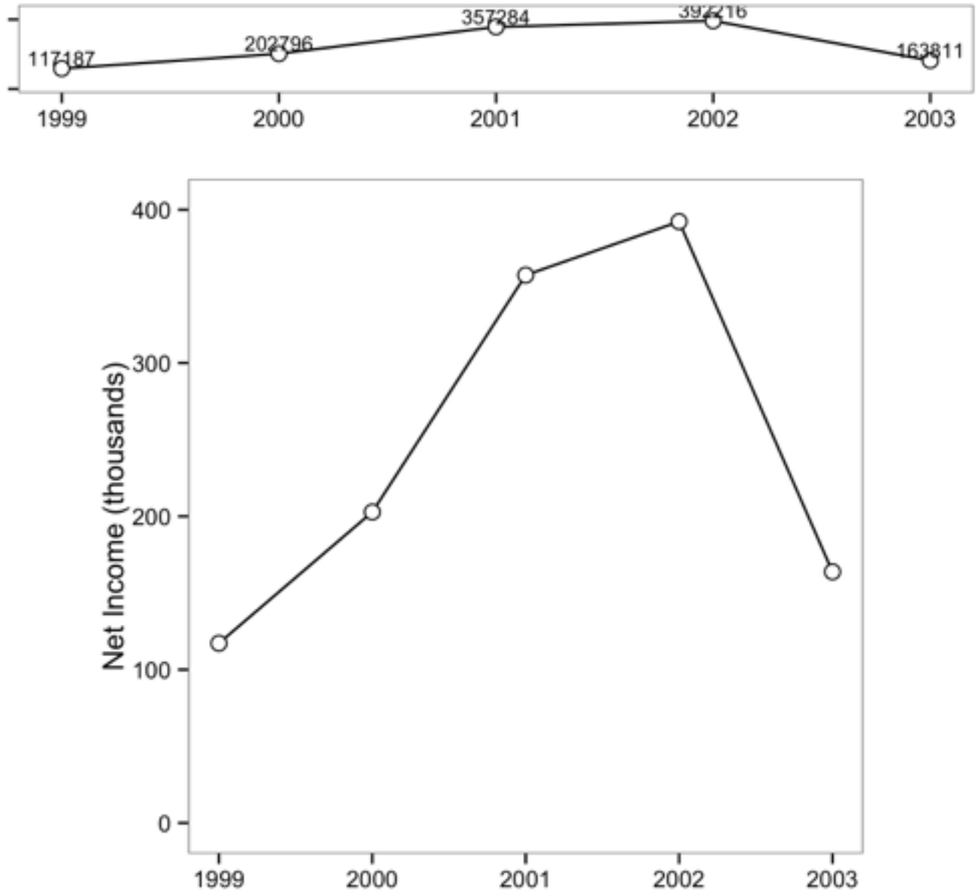
- 1: banking to 45 (1980s)
  - Cleveland perceptual argument: most accurate angle judgement at 45

Fig 7.1 Sunspot Data: Aspect Ratio 1



https://github.com/jennybc/r-graph-catalog/tree/master/figures/fig07-01_sunspot-data-aspect-ratio-1

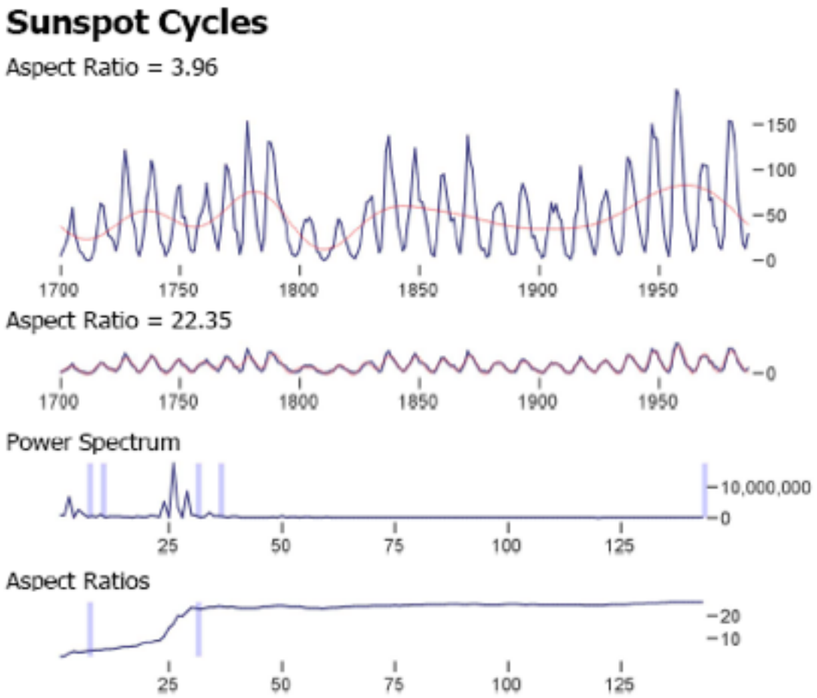
Fig 7.2 Annual Report: Aspect Ratio 2



https://github.com/jennybc/r-graph-catalog/tree/master/figures/fig07-02_annual-report-aspect-ratio-2

Choosing line chart aspect ratios

- 2: multi scale banking to 45 (2006)
  - frequency domain analysis to find ratios
    - FFT the data, convolve with Gaussian to smooth
  - find interesting spikes/ranges in power spectrum
    - cull nearby regions if similar, ensure overview
  - create trend curves (red) for each aspect ratio

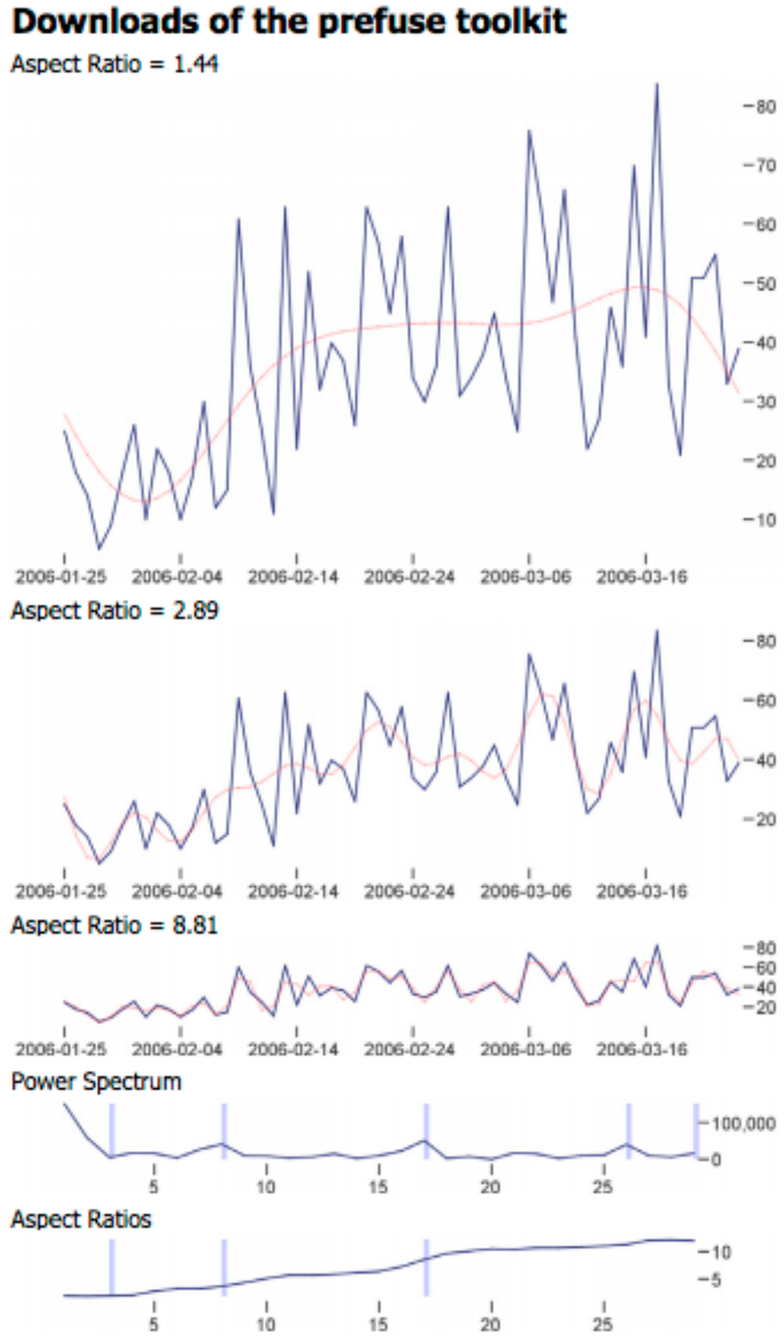


[Multi-Scale Banking to 45 Degrees. Heer and Agrawala, Proc InfoVis 2006]

overall

weekly

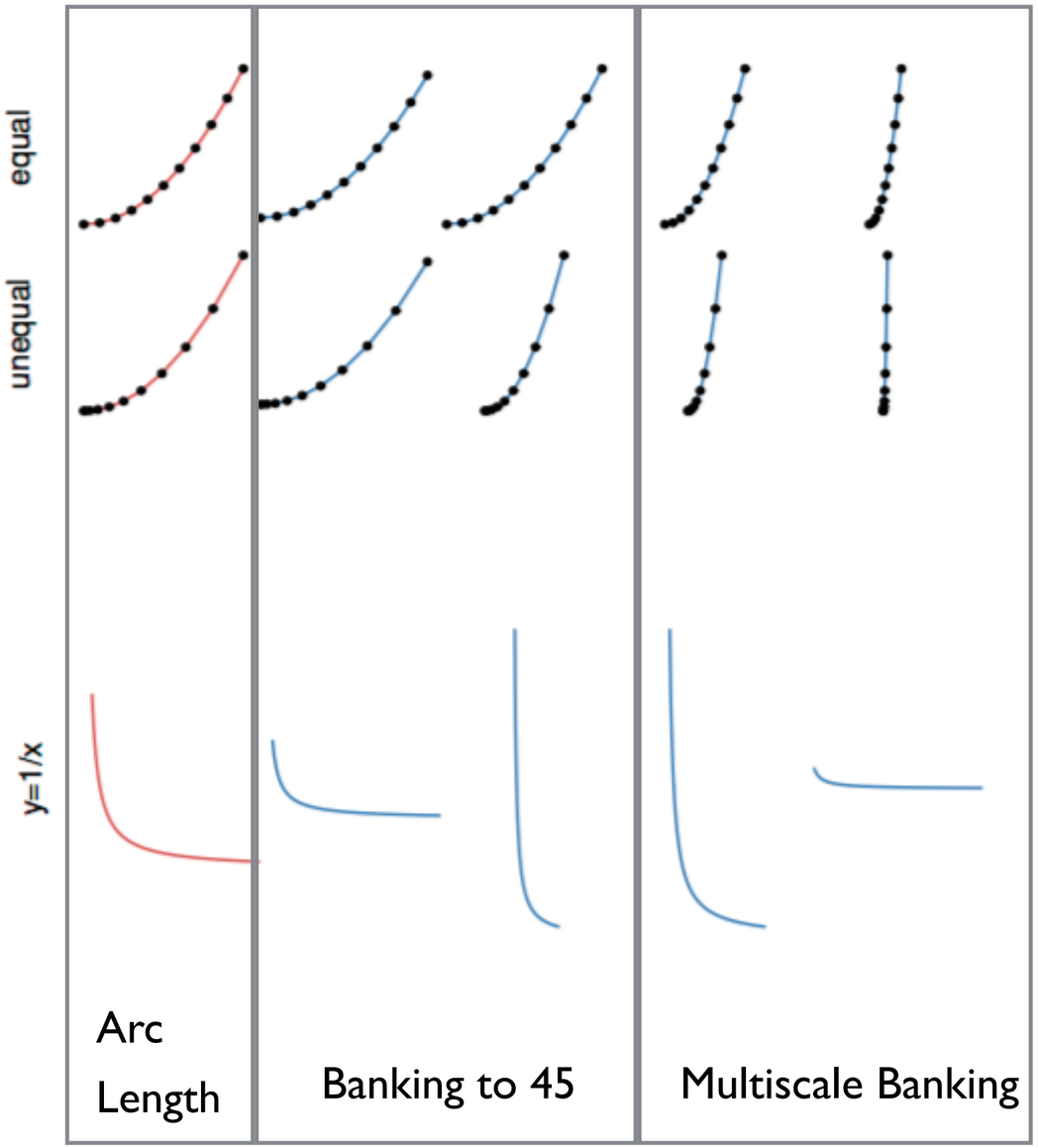
daily



Choosing line chart aspect ratios

- 3: arc length based aspect ratio (2011)
  - minimize the arc length of curve while keeping the area of the plot constant
  - parametrization and scale invariant
  - symmetry preserving
  - robust & fast to compute

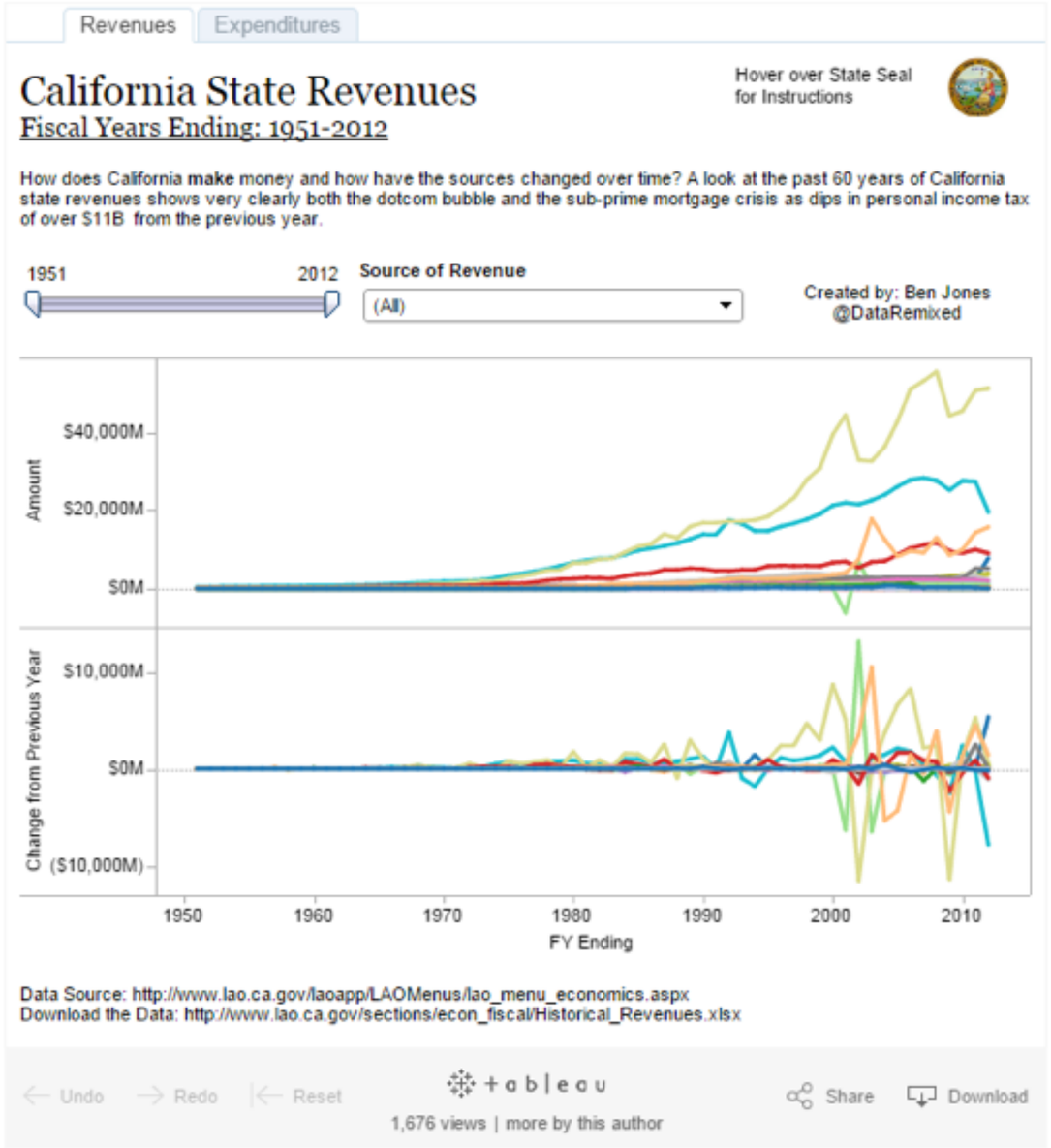
- meta-points from this progression
  - young field; prescriptive advice changes rapidly
  - reasonable defaults required deep dive into perception meets math



[Arc Length-Based Aspect Ratio Selection. Talbot, Gerth, and Hanrahan. Proc InfoVis 2011]

Idiom: **Indexed line charts**

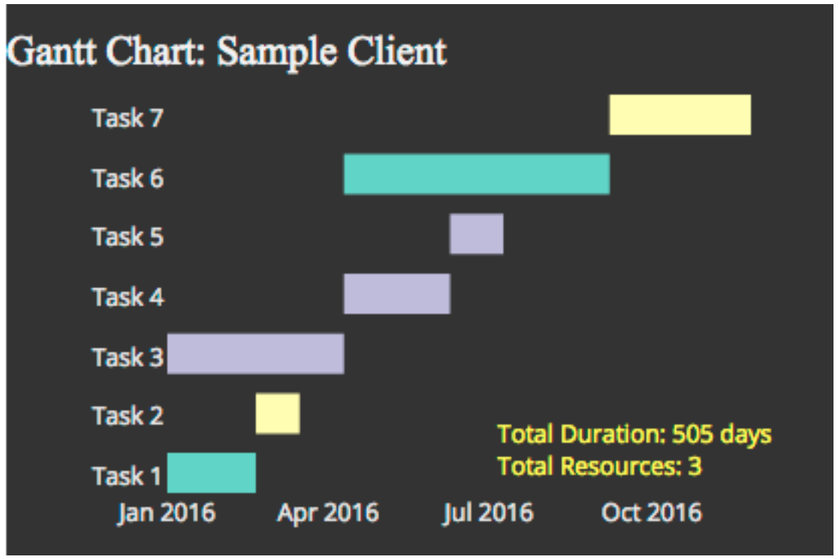
- data: 2 quant attires
  - 1 key + 1 value
- derived data: new quant value attrib
  - index
  - plot instead of original value
- task: show change over time
  - principle: normalized, not absolute
- scalability
  - same as standard line chart



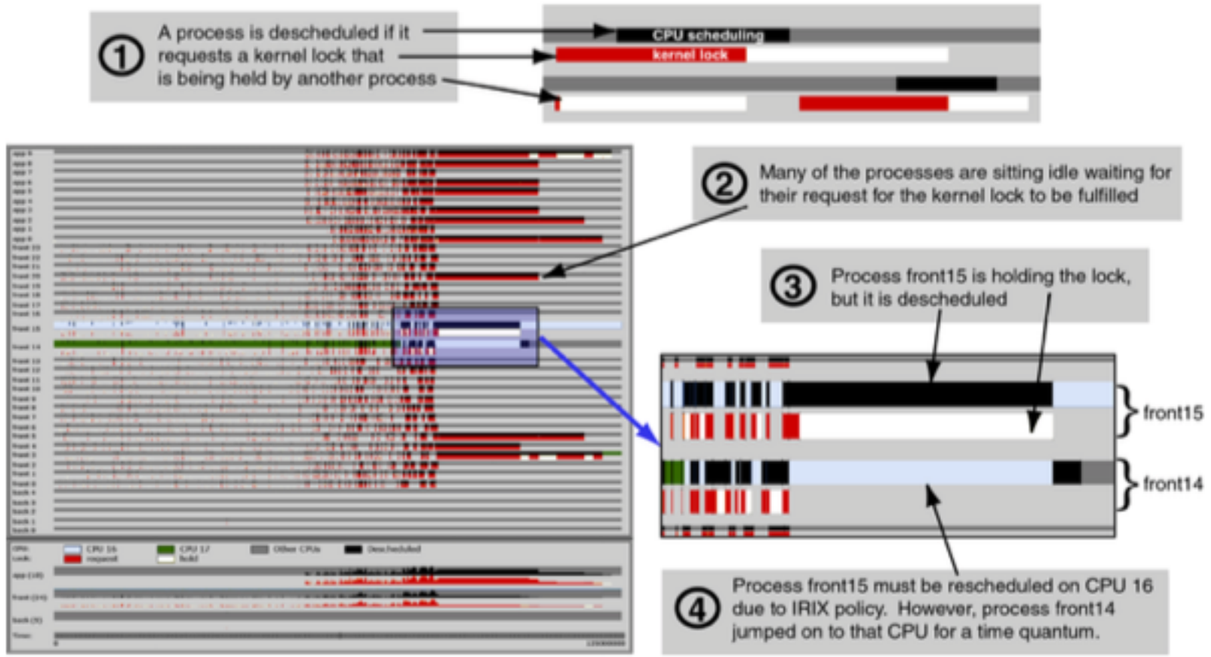
https://public.tableau.com/profile/ben.jones#!/vizhome/CAStateRevenues/Revenues

Idiom: **Gantt charts**

- one key, two (related) values
  - –data
    - 1 categ attrib, 2 quant attribs
  - –mark: line
    - length: duration
  - –channels
    - horiz position: start time (+end from duration)
  - –task
    - emphasize temporal overlaps, start/end dependencies between items
  - –scalability
    - dozens of key levels
    - hundreds of value levels



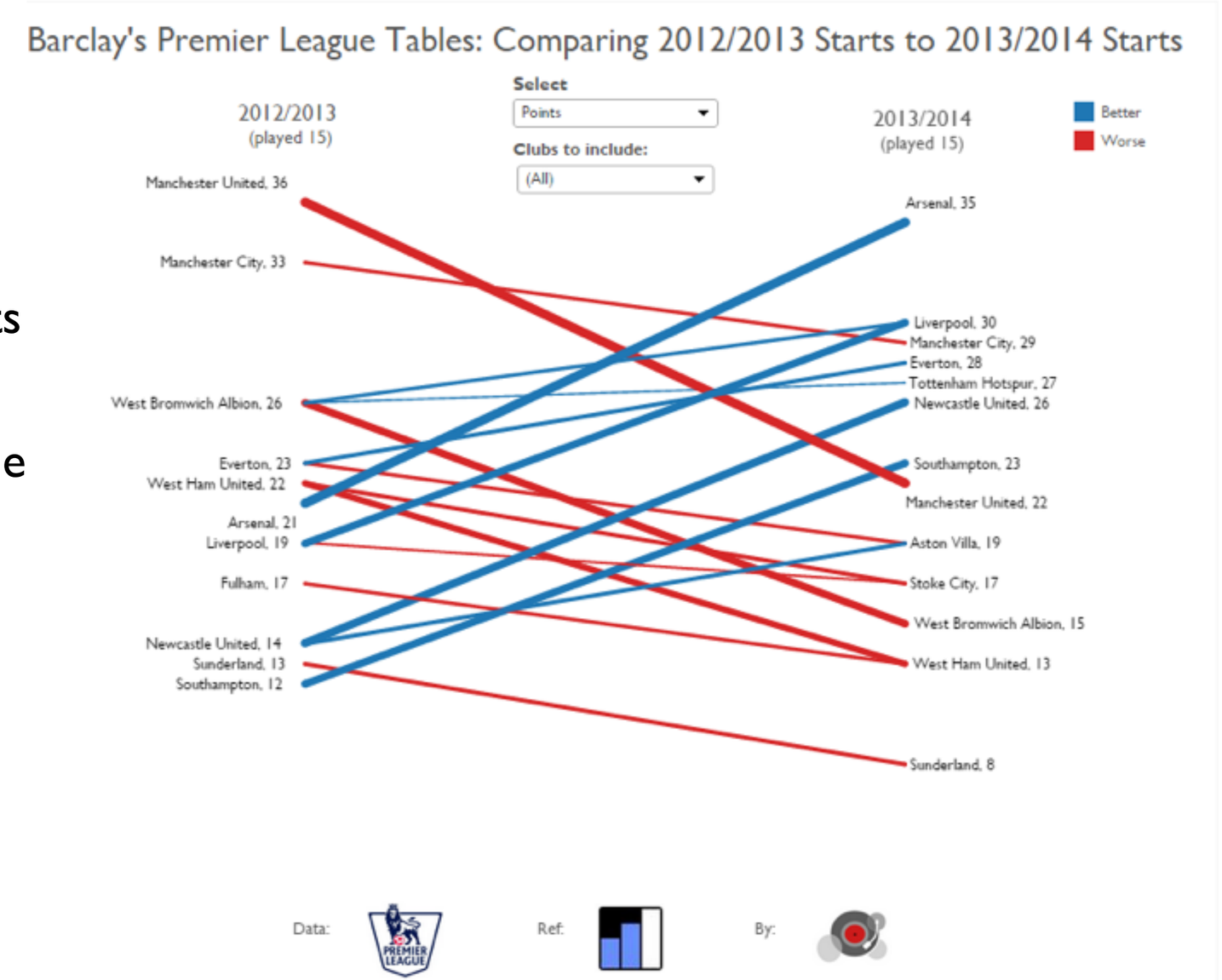
https://www.r-bloggers.com/gantt-charts-in-r-using-plotly/



[Performance Analysis and Visualization of Parallel Systems Using SimOS and Rivet: A Case Study. Bosch, Stolte, Stoll, Rosenblum, and Hanrahan. Proc. HPCA 2000.]

Idiom: **Slopegraphs**

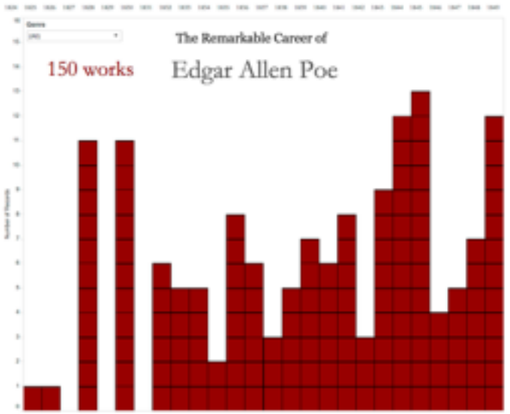
- two values
  - data
    - 2 quant value attribs
  - mark: point + line
    - line connecting mark between pts
  - channels
    - 2 vertical pos: express attrib value
  - task
    - emphasize changes in rank/value
  - scalability
    - hundreds of value levels



https://public.tableau.com/profile/ben.jones#!/vizhome/Slopegraphs/Slopegraphs

Breaking conventions

- presentation vs exploration
  - engaging/evocative
  - inverted y axis
    - blood drips down on Poe



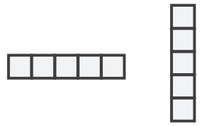
https://public.tableau.com/profile/ben.jones#!/vizhome/EdgarAllanPoeBoring/EdgarAllenPoeBoring

https://public.tableau.com/profile/ben.jones#!/vizhome/EdgarAllanPoeViz/EdgarAllanPoeViz

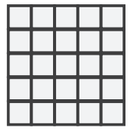
2 Keys

Express Values

→ *1 Key*
List



→ *2 Keys*
Matrix

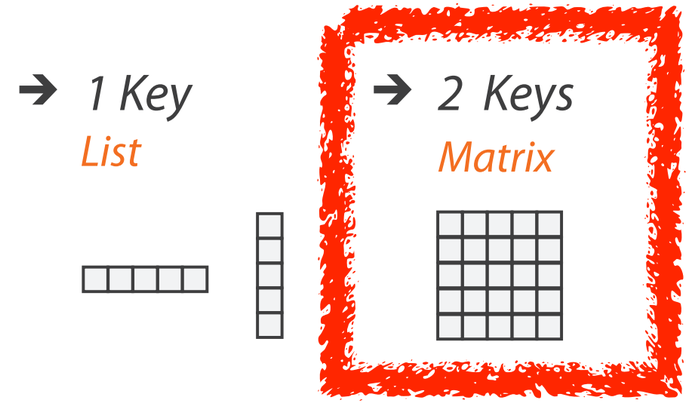


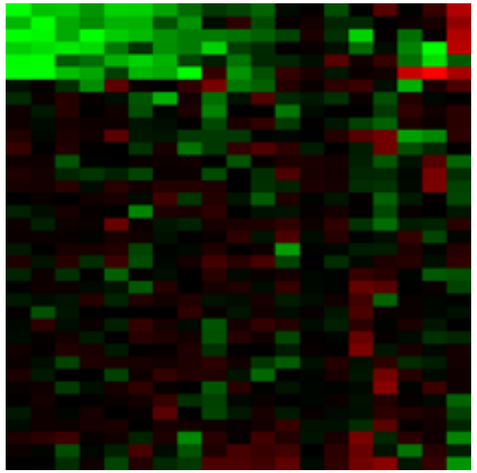
→ *3 Keys*
Volume

→ *Many Keys*
Recursive Subdivision

Idiom: **heatmap**

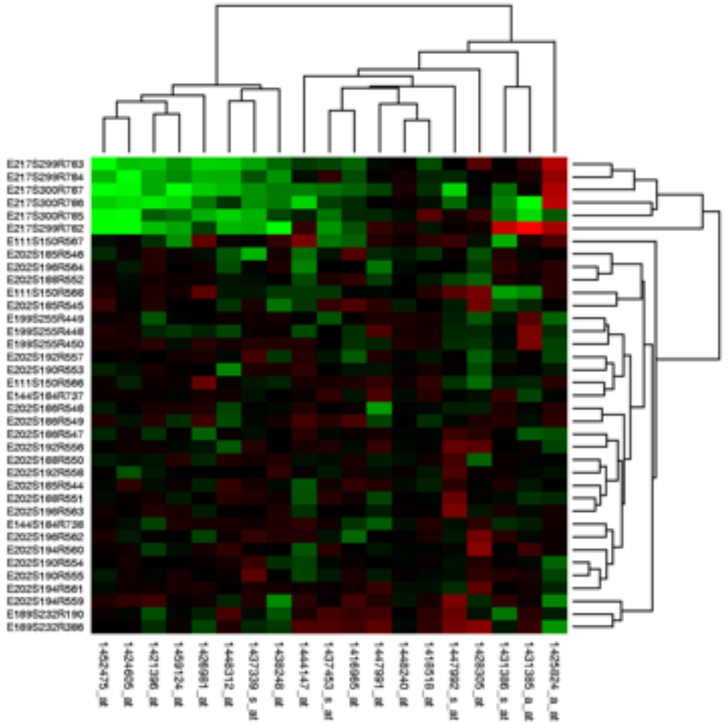
- two keys, one value
  - data
    - 2 categ attribs (gene, experimental condition)
    - 1 quant attrib (expression levels)
  - marks: area
    - separate and align in 2D matrix
      - indexed by 2 categorical attributes
  - channels
    - color by quant attrib
      - (ordered diverging colormap)
  - 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
    - task: assess quality of clusters found by automatic methods



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, ...

Viz-1

Manipulate

Change

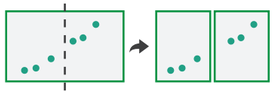
Select

Navigate

Facet

Juxtapose

Partition



Superimpose

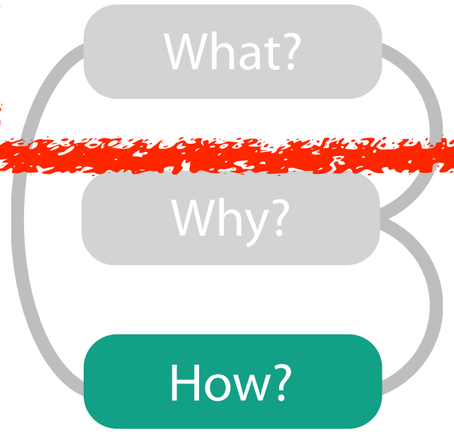
Reduce

Filter

Aggregate

Embed

Viz-2



- creating static single view
- handling complexity

How to handle complexity: 1 previous strategy + 3 more

→ *Derive*

- derive new data to show within view

- change view over time
- facet across multiple views

- reduce items/attributes within single view

Manipulate

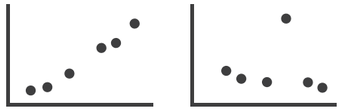
→ Change

→ Select

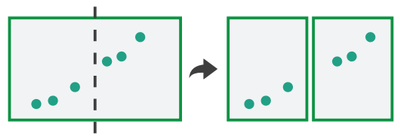
→ Navigate

Facet

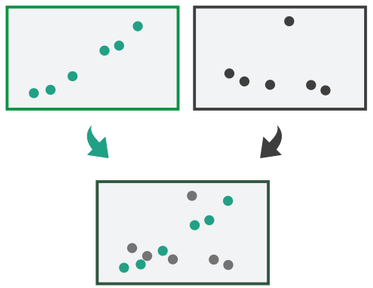
→ Juxtapose



→ Partition



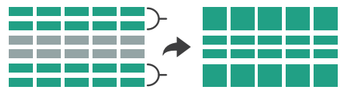
→ Superimpose



Reduce

→ Filter

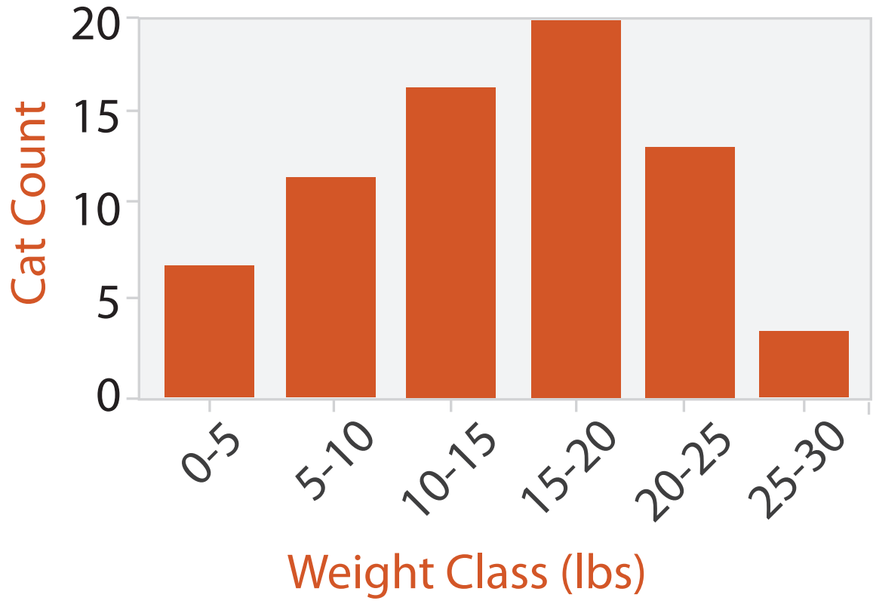
→ Aggregate



→ Embed

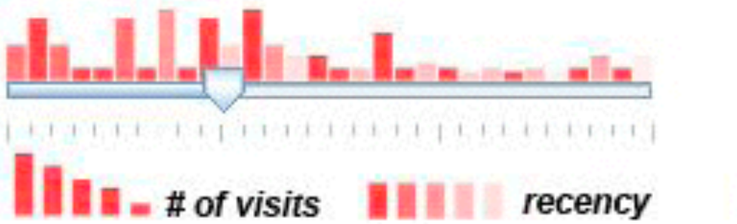
Idiom: **histogram**

- static item aggregation
- task: find distribution
- data: table
- derived data
  - new table: keys are bins, values are counts
- scalability
  - depends on bin size, not original table size
- bin size crucial
  - pattern can change dramatically depending on discretization
  - opportunity for interaction: control bin size on the fly

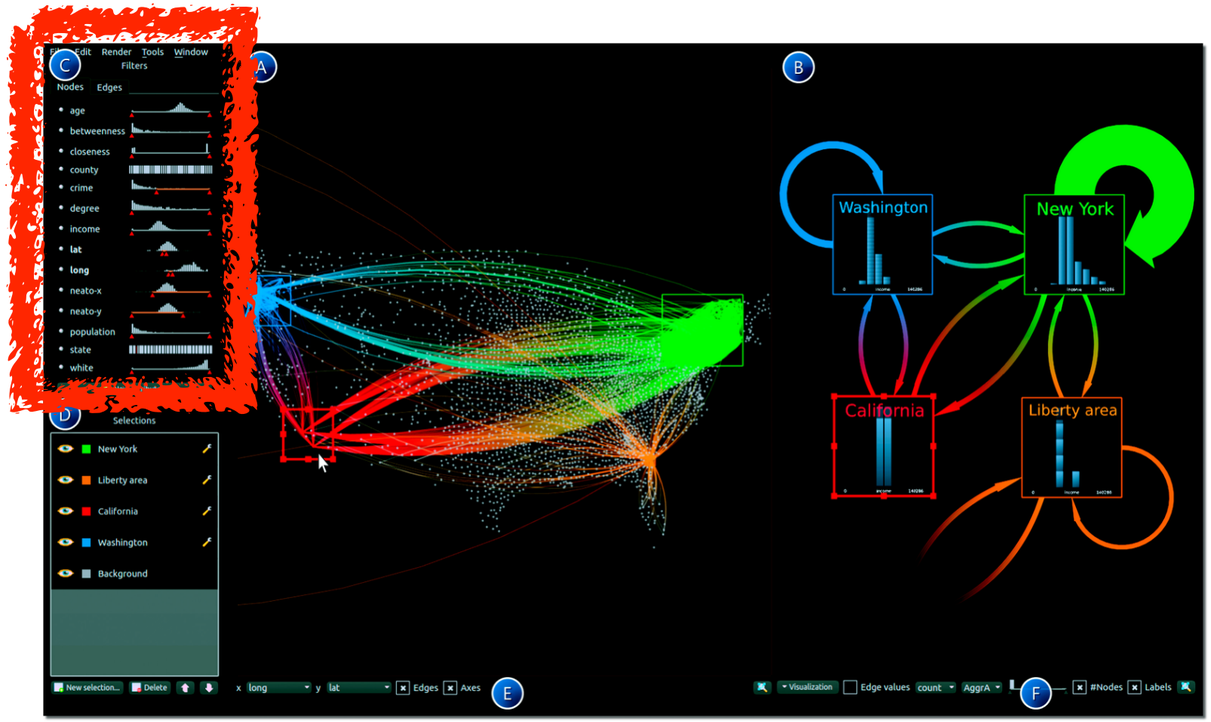


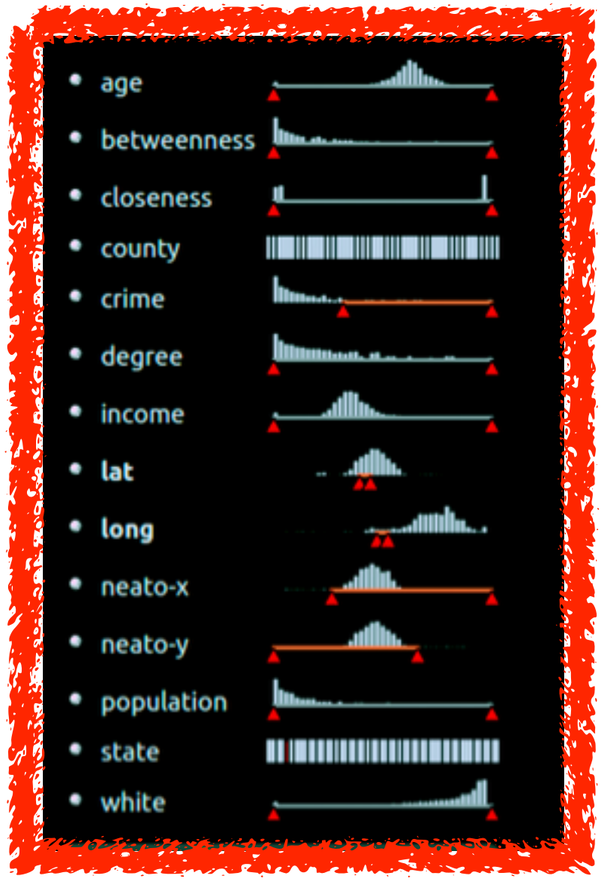
Idiom: **scented widgets**

- augmented widgets show information scent
  - cues to show whether value in drilling down further vs looking elsewhere
- concise use of space: histogram on slider



[Scented Widgets: Improving Navigation Cues with Embedded Visualizations. Willett, Heer, and Agrawala. IEEE TVCG (Proc. InfoVis 2007) 13:6 (2007), 1129–1136.]

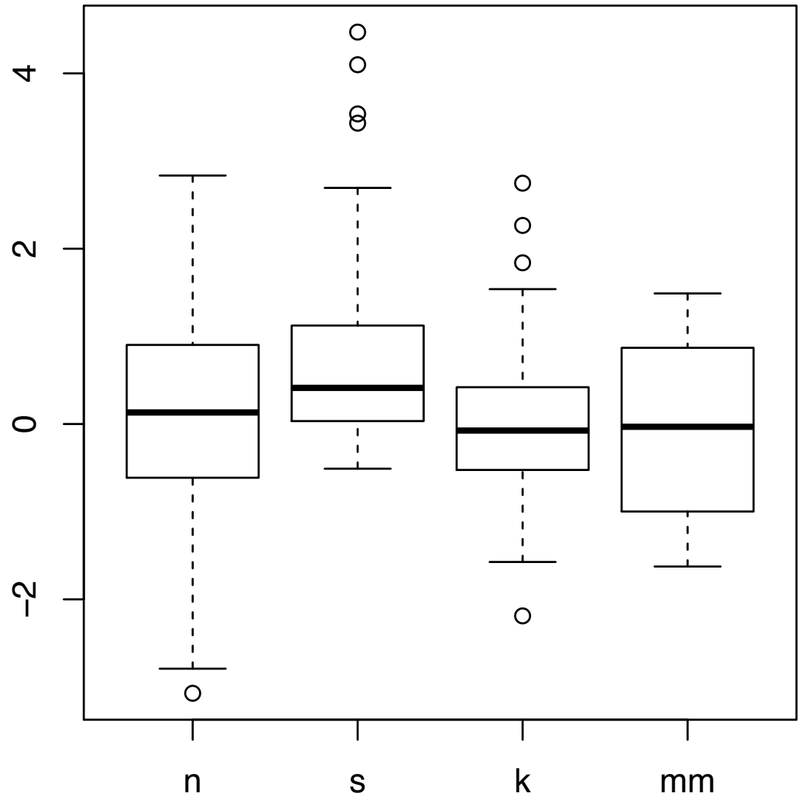




[Multivariate Network Exploration and Presentation: From Detail to Overview via Selections and Aggregations. van den Elzen, van Wijk, IEEE TVCG 20(12): 2014 (Proc. InfoVis 2014).]

Idiom: **boxplot**

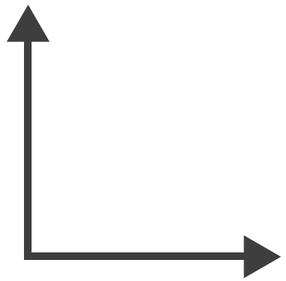
- static item aggregation
- task: find distribution, find outliers
- data: table
- derived data
  - 5 quant attribs
    - median: central line
    - lower and upper quartile: boxes
    - lower upper fences: whiskers
      - values beyond which items are outliers
  - outliers beyond fence cutoffs explicitly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



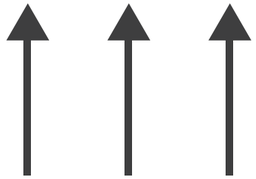
[40 years of boxplots. Wickham and Stryjewski. 2012. had.co.nz]

Axis Orientation

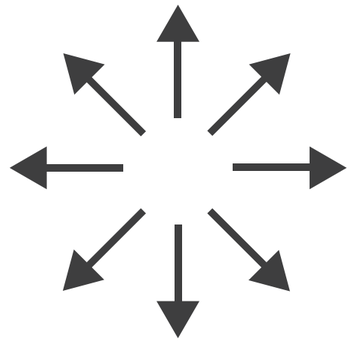
- Rectilinear



- Parallel



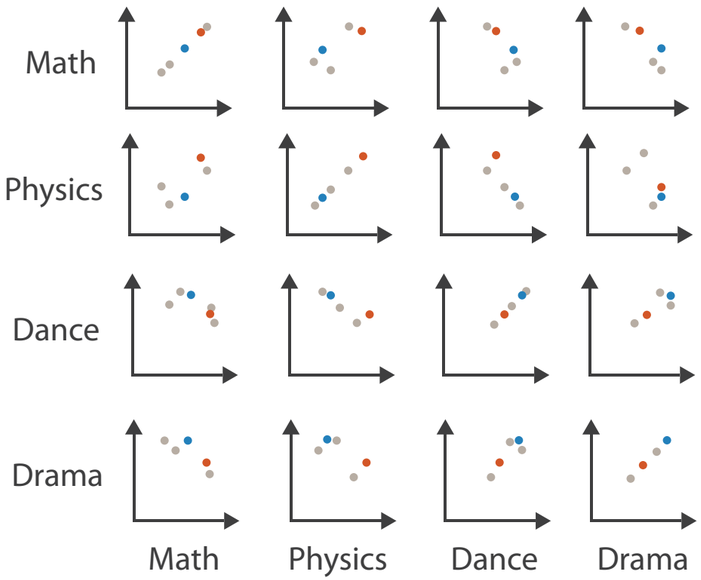
- Radial



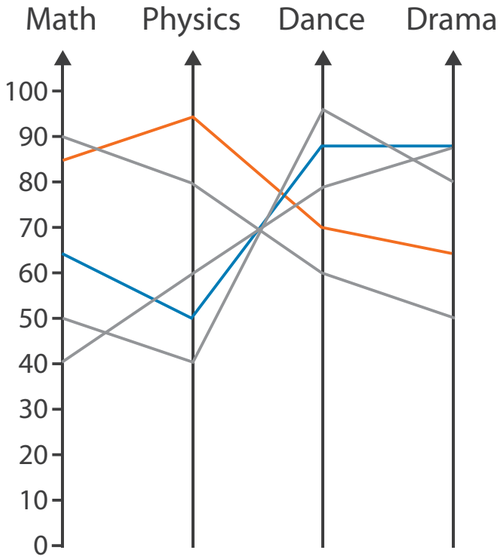
Idioms: **scatterplot matrix, parallel coordinates**

- 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



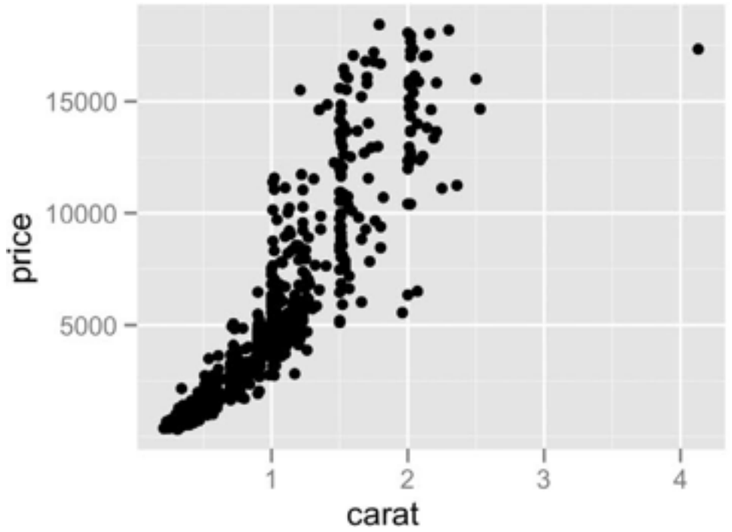
Table

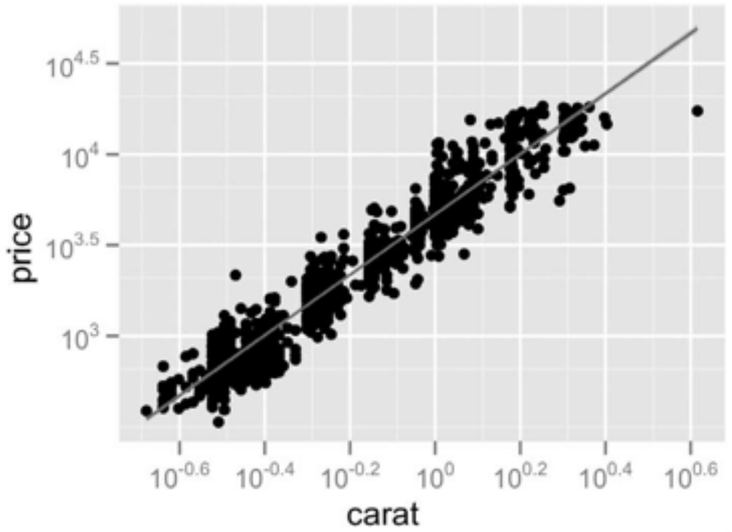
Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

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.]

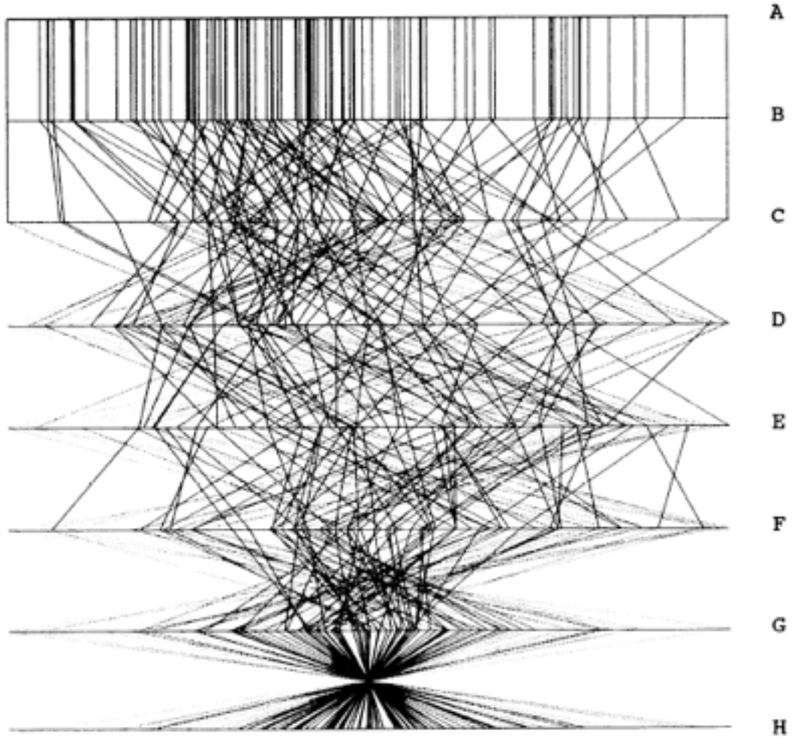


Figure 3. Parallel Coordinate Plot of Six-Dimensional Data Illustrating Correlations of $\rho = 1, .8, .2, 0, -.2, -.8,$ and -1.

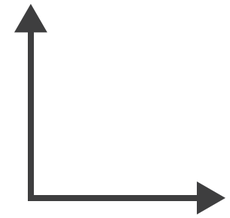
[Hyperdimensional Data Analysis Using Parallel Coordinates. Wegman. Journ. American Statistical Association 85:411 (1990), 664–675.]

Orientation tradeoffs/limitations

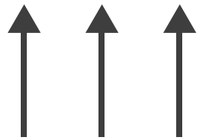
- rectilinear: scalability wrt #axes
  - 2 axes best
  - 3 often problematic
    - more later
  - 4+ impossible
- parallel: unfamiliarity, training time

➔ **Axis Orientation**

➔ Rectilinear

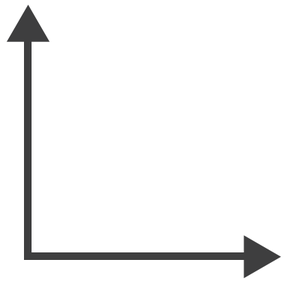


➔ Parallel

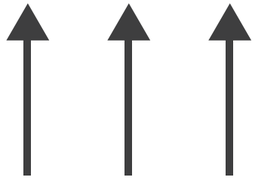


Axis Orientation

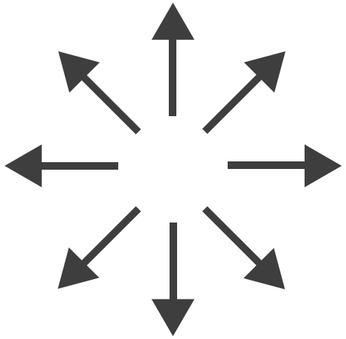
- Rectilinear



- Parallel

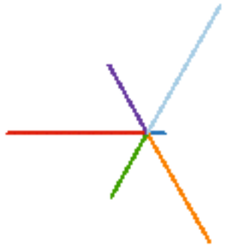


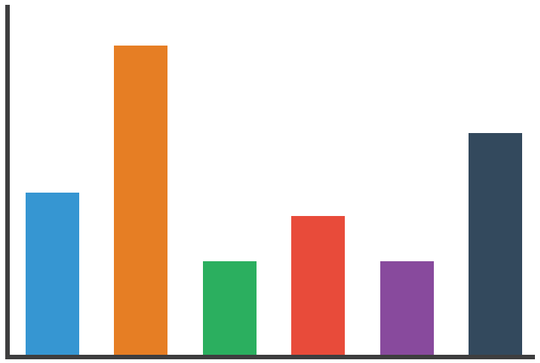
- Radial



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

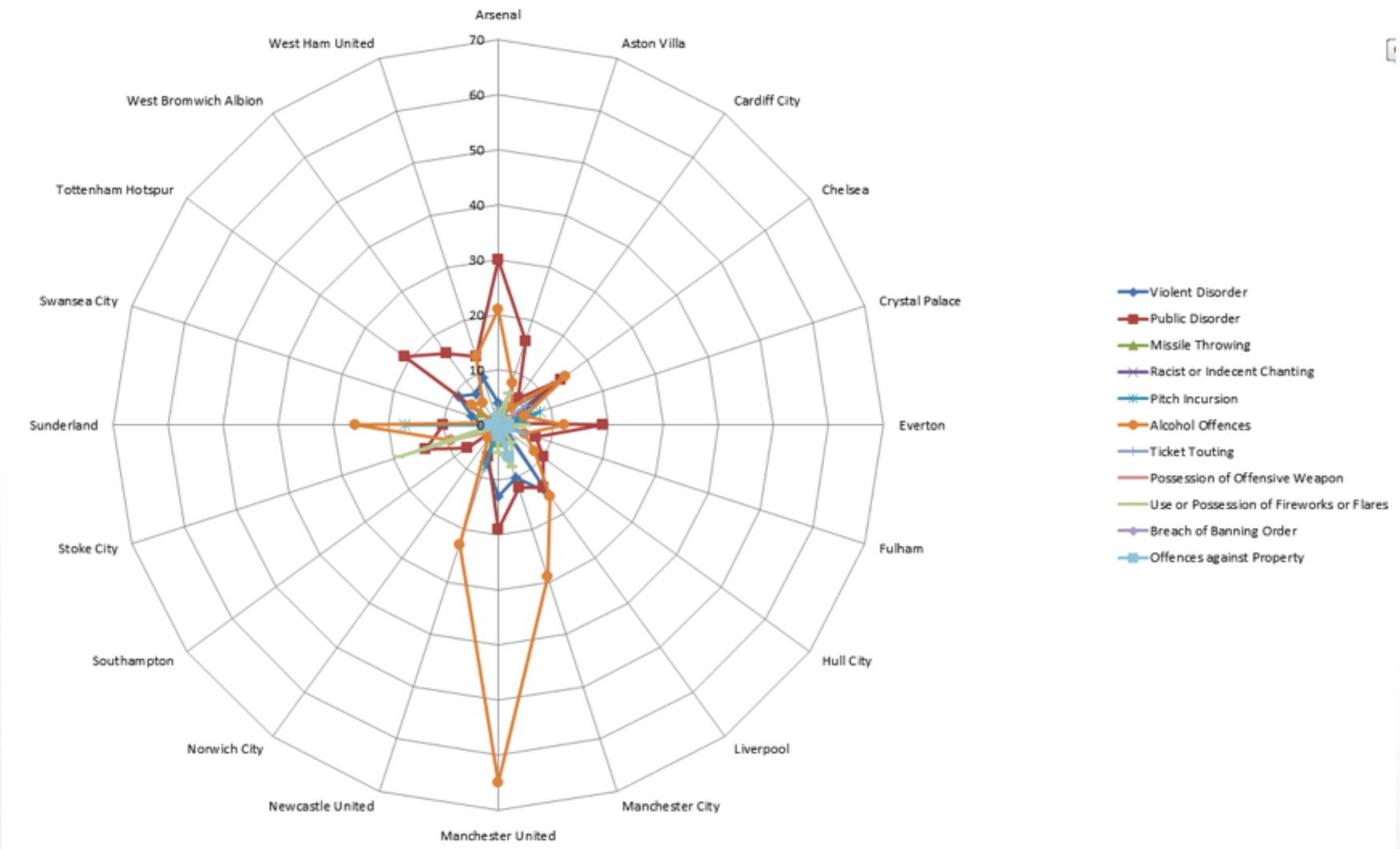




- accuracy
  - length unaligned with radial
    - less accurate than aligned with rectilinear

[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.]

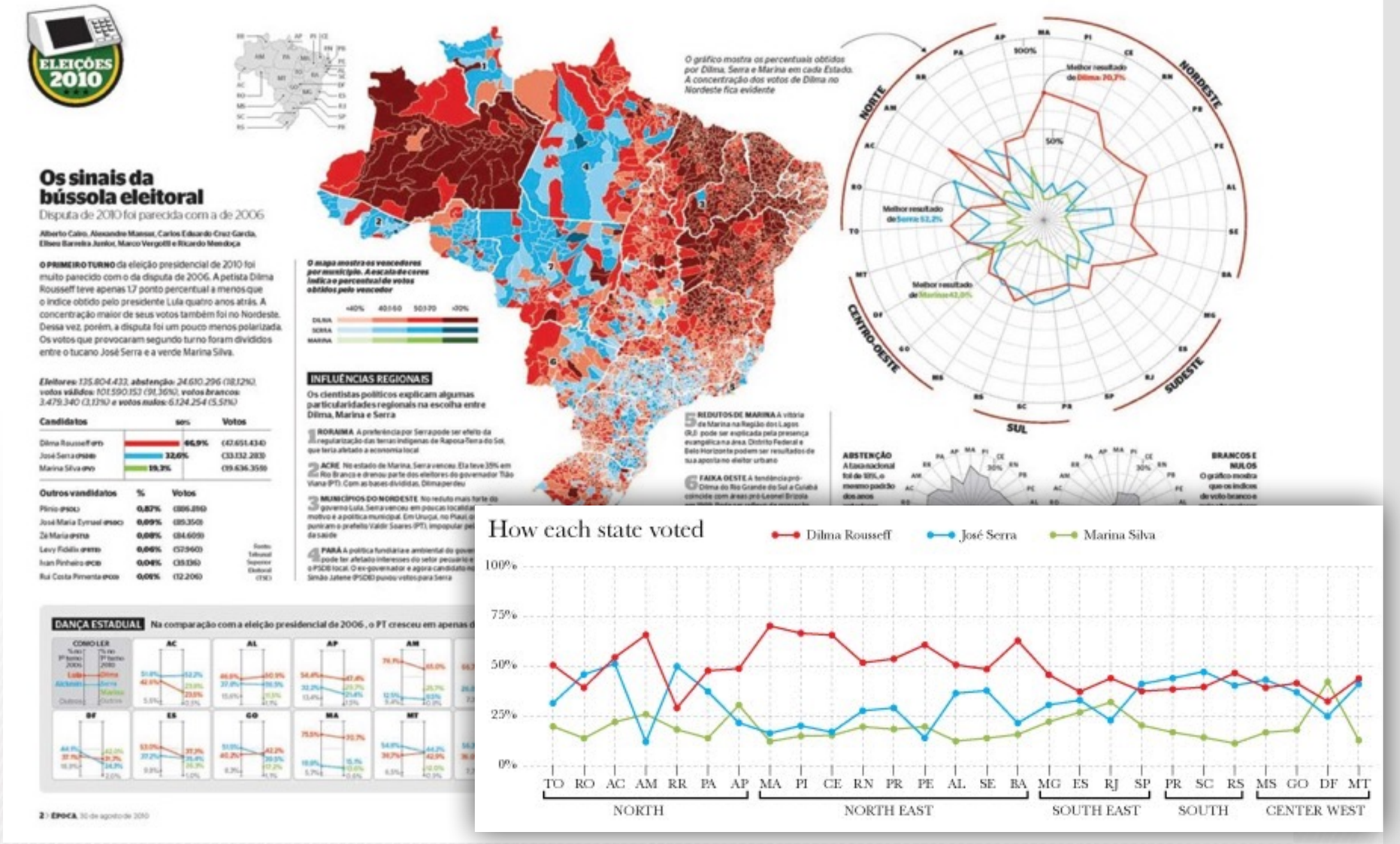
Radial Orientation: Radar Plots



LIMITATION: Not good when categories aren't cyclic

[Slide courtesy of Ben Jones]

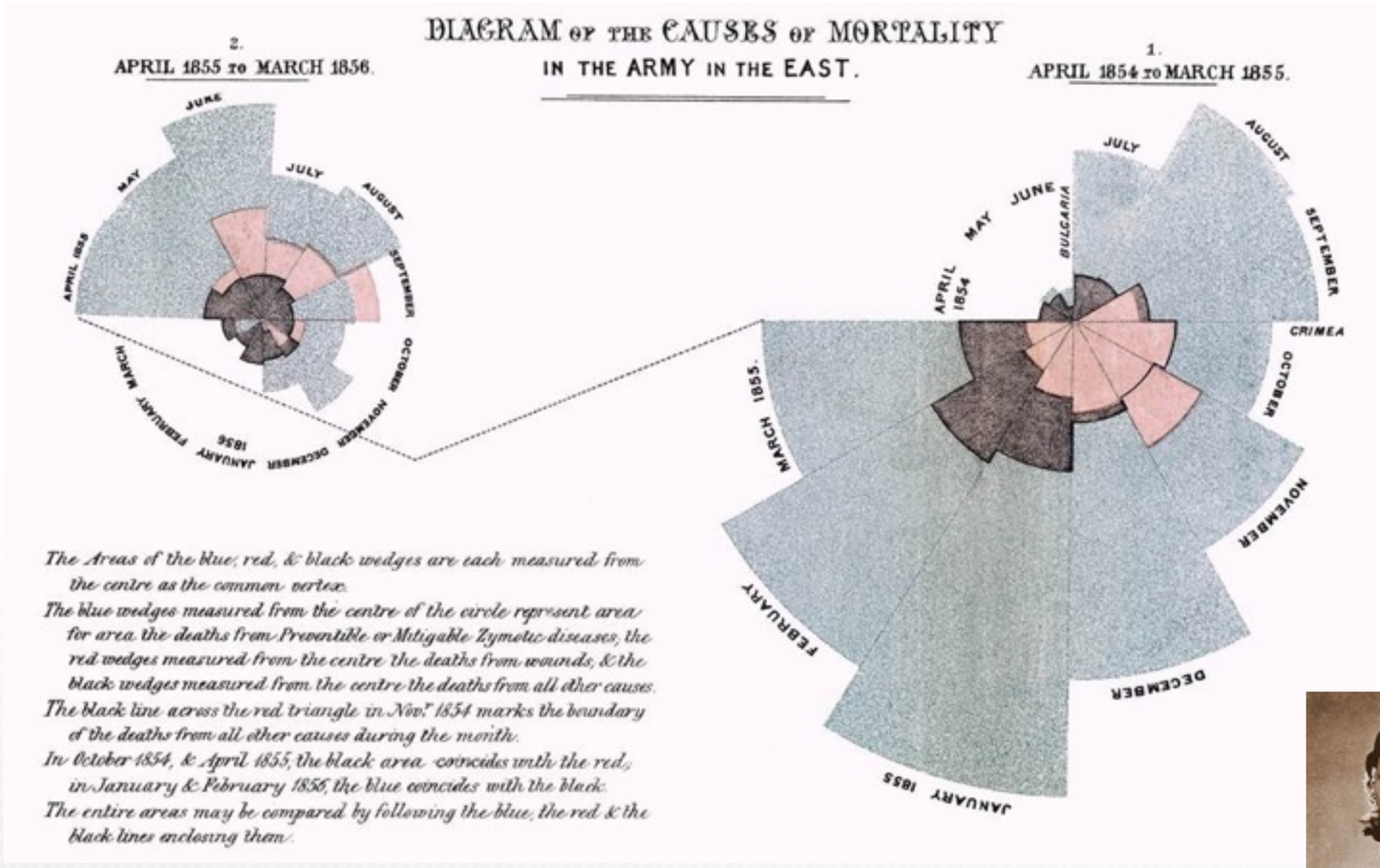
“Radar graphs: Avoid them (99.9% of the time)”



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

[Slide courtesy of Ben Jones]

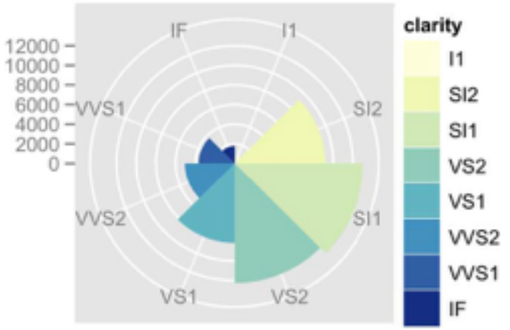
"Diagram of the causes of mortality in the army in the East" (1858)

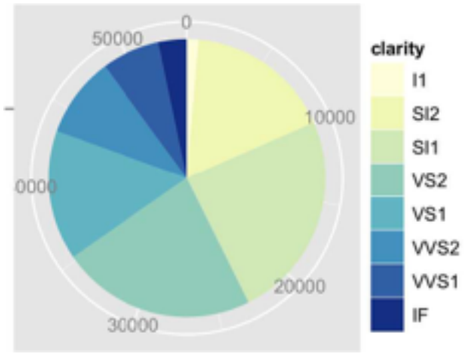


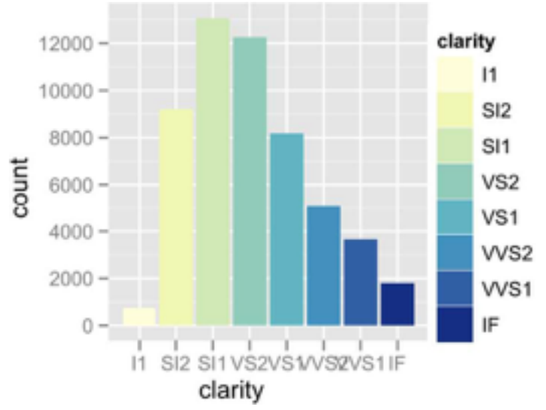
[Slide courtesy of Ben Jones]

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

- polar area chart
  - area marks with length channel
  - direct analog to bar charts
- pie chart
  - area marks with angle channel
  - accuracy: less accurate than aligned line length
- data
  - 1 categ key attrib, 1 quant value attrib
- task
  - part-to-whole judgements
    - note parts must add up to 100% whole!



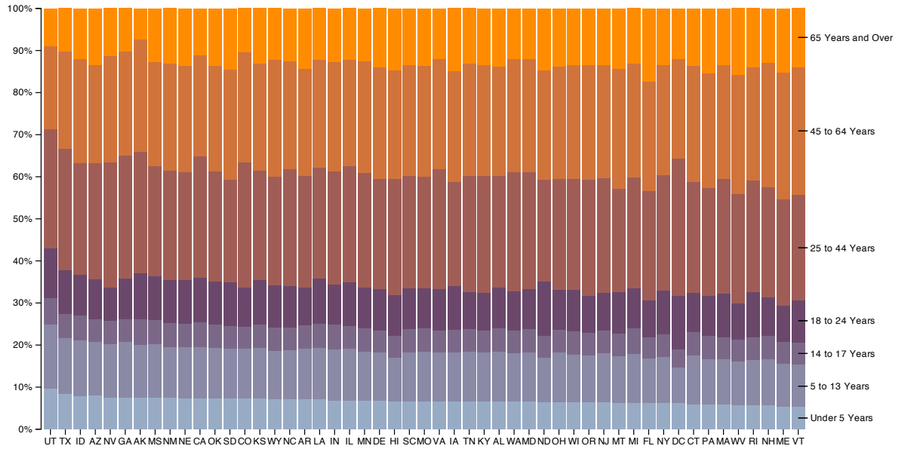


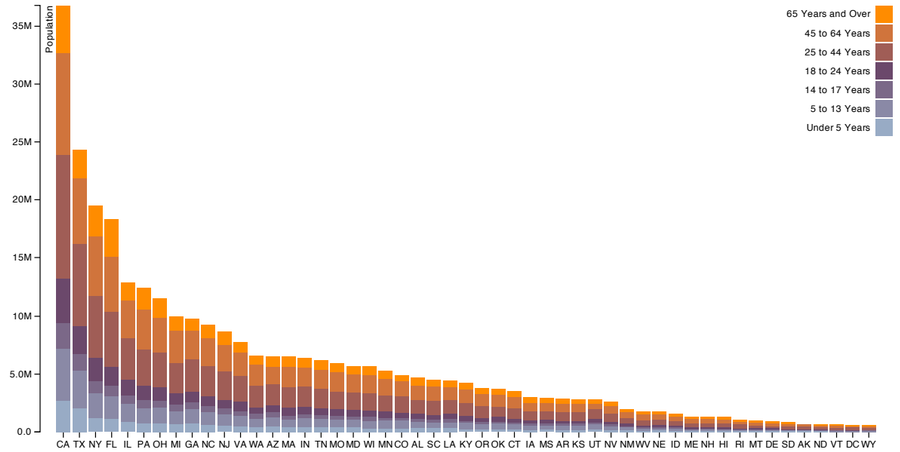


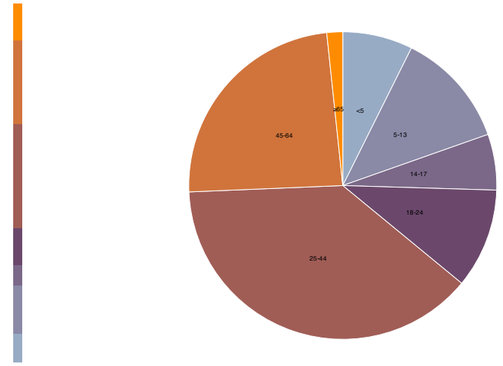
[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]

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
  - poor information density: requires large circle







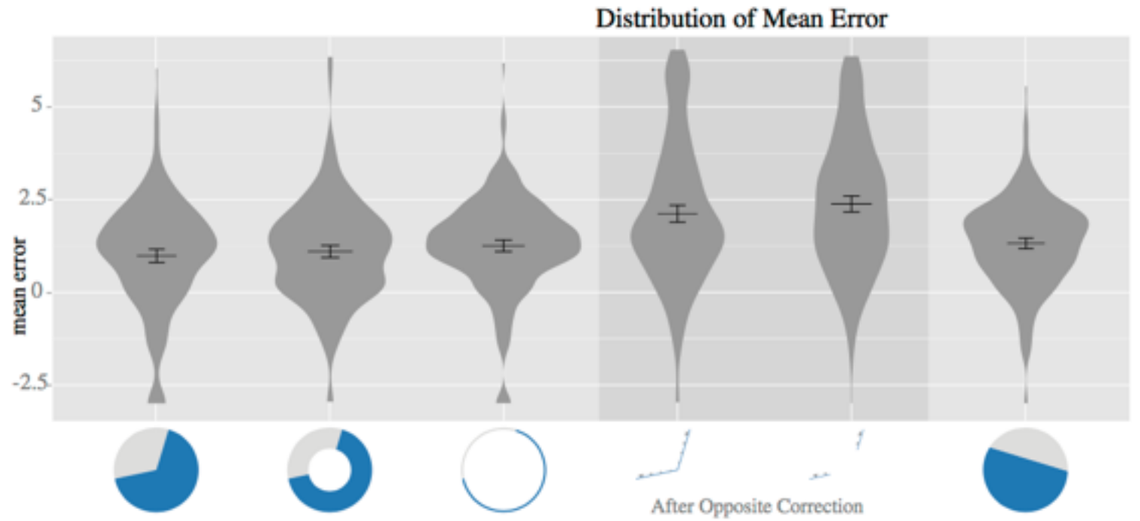
http://bl.ocks.org/mbostock/3887235,
http://bl.ocks.org/mbostock/3886208,
http://bl.ocks.org/mbostock/3886394.

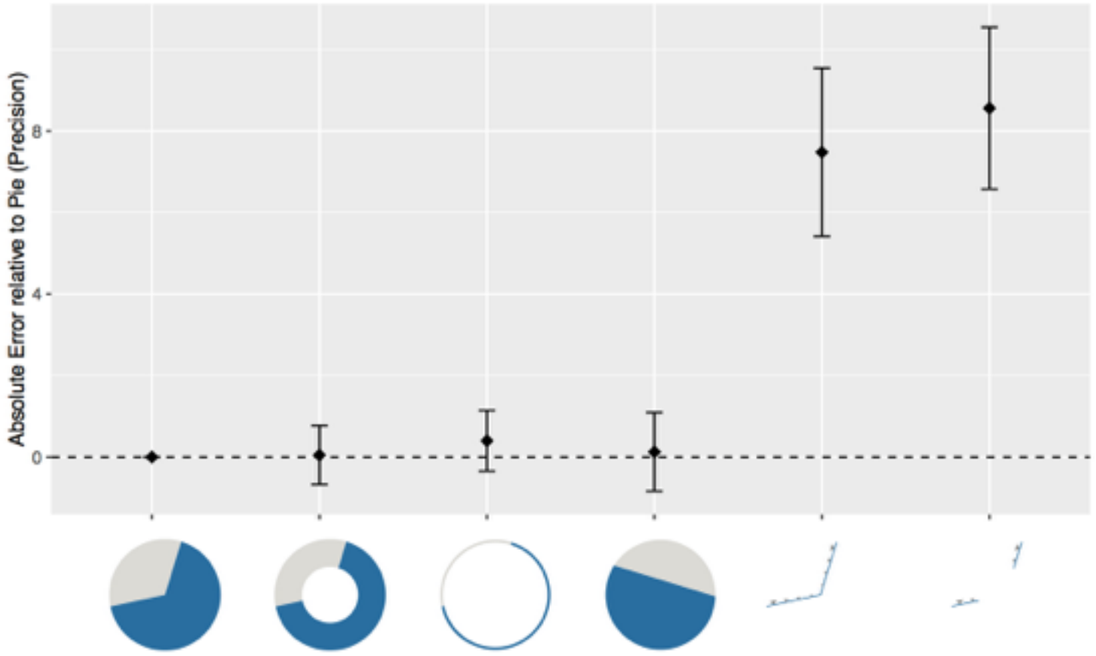
Pie chart perception

- some empirical evidence that people respond to arc length
  - not angles
  - maybe also areas?…
- donut charts no worse than pie charts

[Arcs, Angles, or Areas: Individual Data Encodings in Pie and Donut Charts. Skau and Kosara. Proc. EuroVis 2016.]

- meta-points
  - redesign of paper figures in later blog post
    - violin plots good for analysis but too detailed for presentation
  - my advice: still dubious for pie/donut charts
    - sometimes ok if just 2 attribs

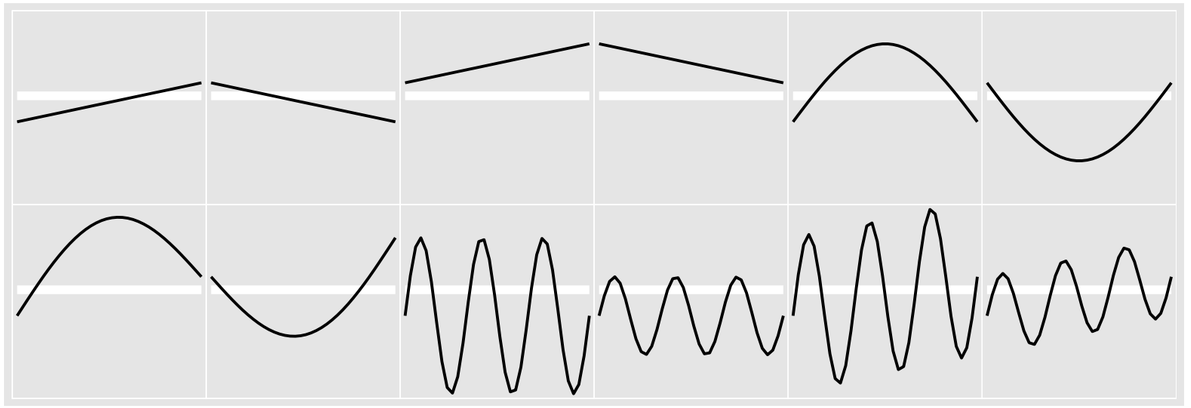




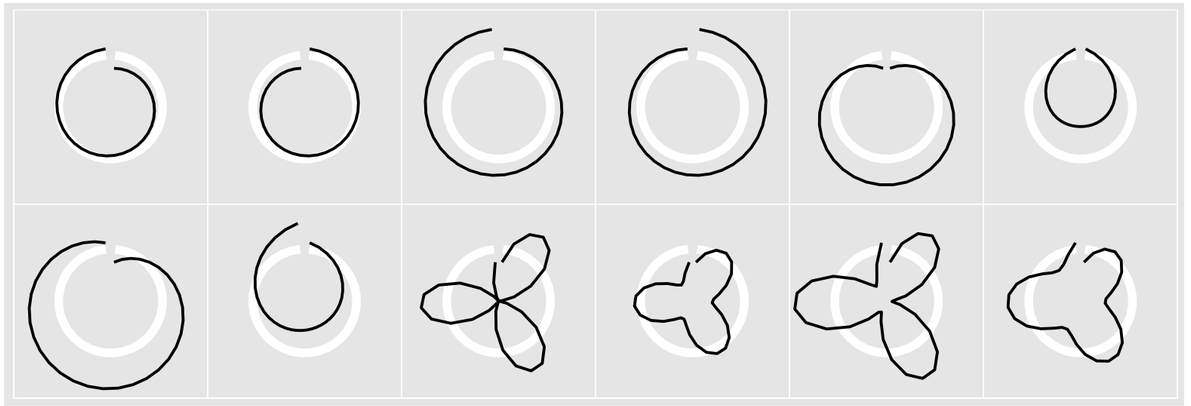
https://eagereyes.org/blog/2016/an-illustrated-tour-of-the-pie-chart-study-results

Idiom: **glyphmaps**

- rectilinear good for linear vs nonlinear trends



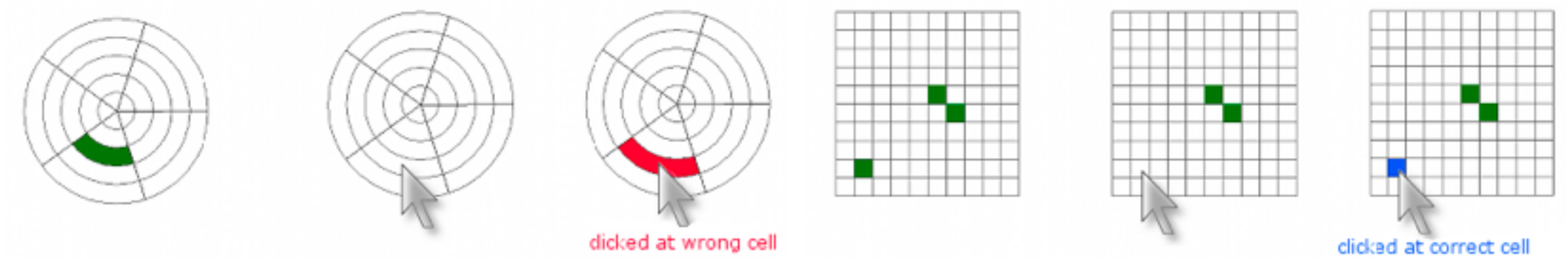
- radial good for cyclic patterns
  - use case: finding periodicity



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

Radial orientation

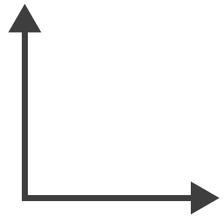
- perceptual limits
  - –polar coordinate asymmetry
    - angles lower precision than lengths
    - frequently problematic
    - sometimes can be deliberately exploited!
      - for 2 attribs of very unequal importance



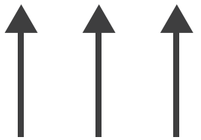
[Uncovering Strengths and Weaknesses of Radial Visualizations - an Empirical Approach. Diehl, Beck and Burch. IEEE TVCG (Proc. InfoVis) 16(6):935–942, 2010.]

Axis Orientation

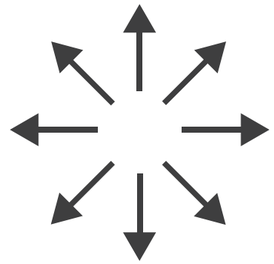
- Rectilinear



- Parallel

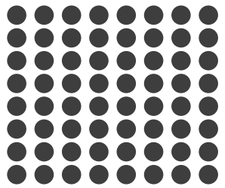


- Radial

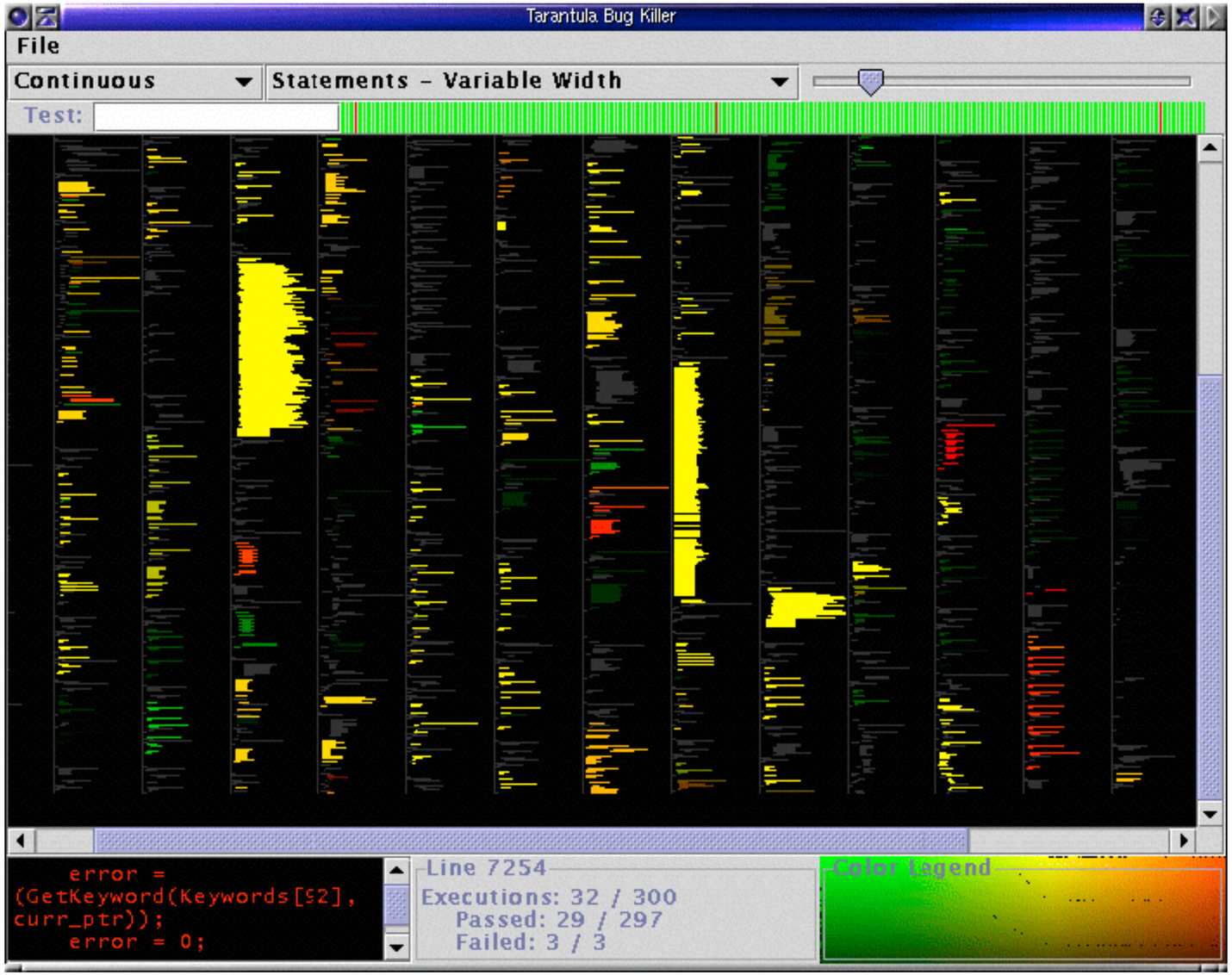


Idiom: Dense software overviews

- Layout Density
  - Dense



- data: text
  - text + 1 quant attrib per line
- derived data:
  - one pixel high line
  - length according to original
- color line by attrib
- scalability
  - 10K+ lines



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