

# Information Visualization

## Tables, LineUp, Bertifier

### In Class: Pitches

**Tamara Munzner**  
 Department of Computer Science  
 University of British Columbia

Week 4, 29 Sep 2021  
<https://www.cs.ubc.ca/~tmm/courses/547-21>

### Plan for today

- Pitches
  - 70-80 min pitches (2 min each)
  - up to 2 min Q&A after each
- (break in the middle)
- remaining time: Reading Q&A
  - Tables
  - LineUp
  - Bertifier

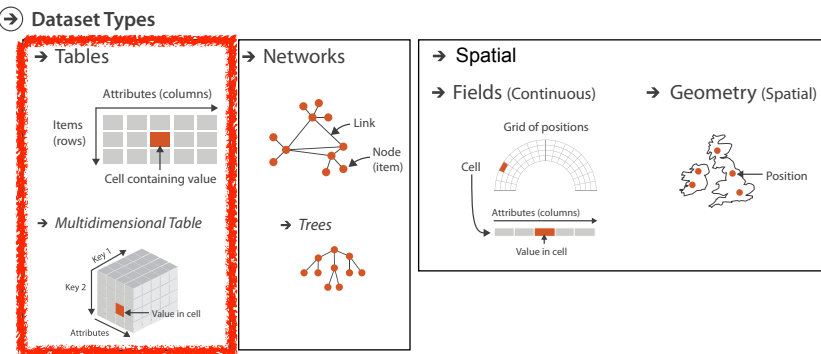
### Next week

- to read & discuss (async, before next class)
  - VAD book, Ch 10: Color
  - paper: Artery Viz [design study & quant evaluation]
  - paper: Rainbows Revisited [evaluation]

### Q&A / Backup Slides

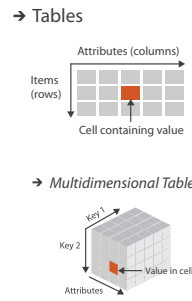
## Tables

### Focus on Tables



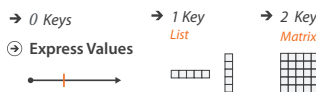
### Keys and values

- 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



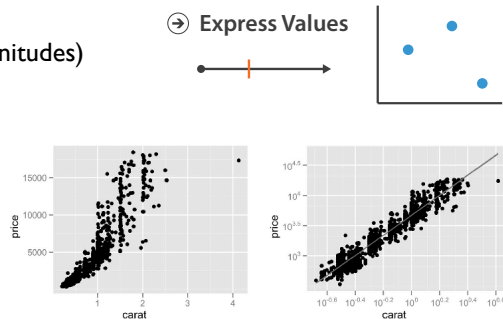
### Keys and values

- 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 keys used
  - 0, 1, 2, ...



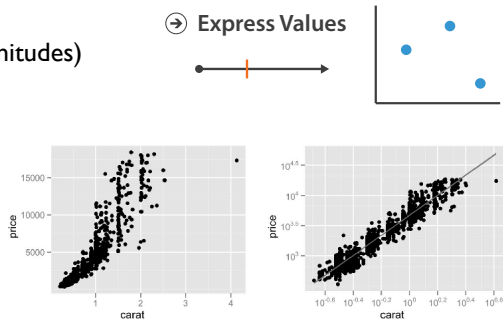
### Idiom: scatterplot

- **express** values (magnitudes)
  - quantitative attributes
- no keys, only values



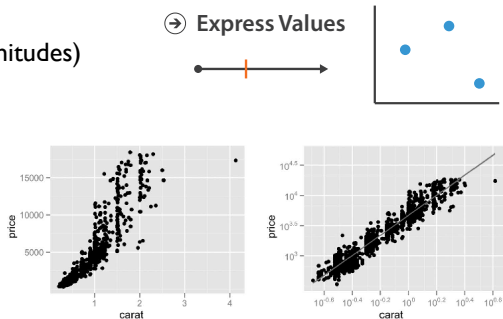
### Idiom: scatterplot

- **express** values (magnitudes)
  - quantitative attributes
- no keys, only values
  - data
    - 2 quant attribs
  - mark: points
  - channels
    - horiz + vert position



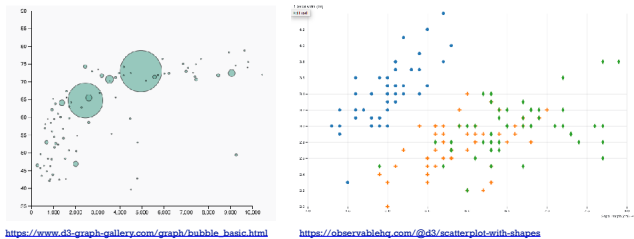
### Idiom: scatterplot

- **express** values (magnitudes)
  - 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



### Scatterplots: Encoding more channels

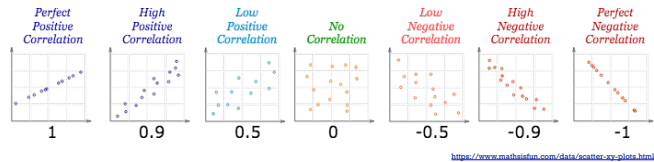
- additional channels viable since using point marks
  - color
  - size (1 quant attribute, used to control 2D area)
    - note radius would mislead, take square root since area grows quadratically
  - shape



### Scatterplot tasks

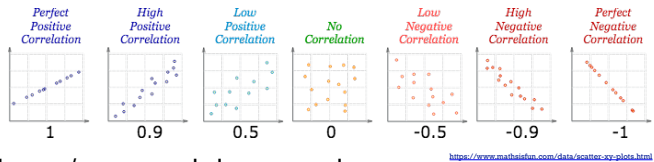
### Scatterplot tasks

- correlation

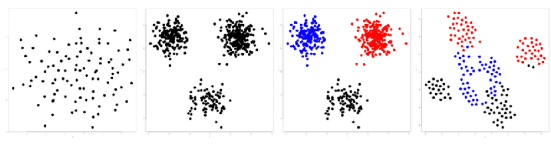


### Scatterplot tasks

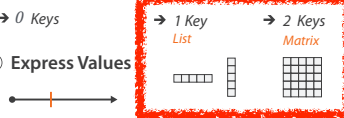
- correlation



- clusters/groups, and clusters vs classes



### Some keys



## Some keys: Categorical regions



## Regions: Separate, order, align



- regions: contiguous bounded areas distinct from each other
  - separate into spatial regions: one mark per region (for now)
- use categorical or ordered attribute to separate into regions
  - no conflict with expressiveness principle for categorical attributes
- use ordered attribute to order and align regions

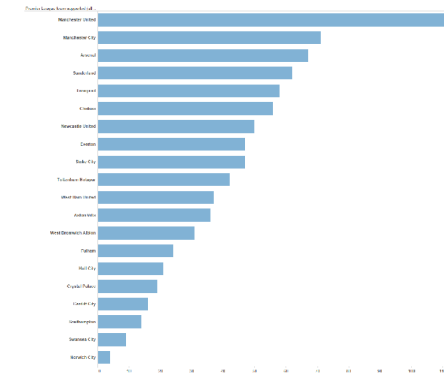


17

18

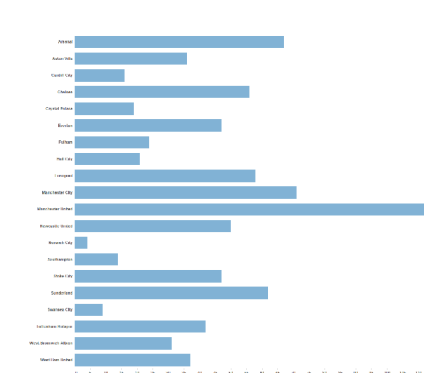
## Separated and aligned and ordered

- best case



## Separated and aligned but not ordered

- limitation: hard to know rank. what's 4th? what's 7th?

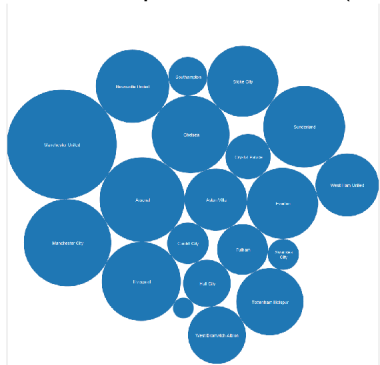


19

20

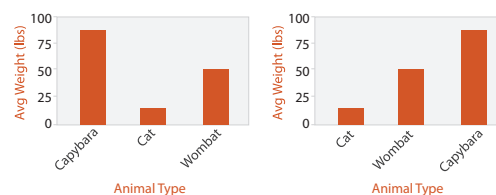
## Separated but not aligned or ordered

- limitation: hard to make comparisons with size (vs aligned position)



## 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 [bars], hundreds for values

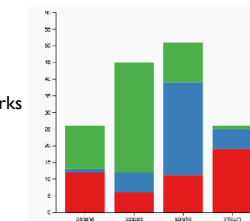


21

22

## 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: asymmetric
    - for stacked key attrib, 10-12 levels [segments]
    - for main key attrib, dozens to hundreds of levels [bars]

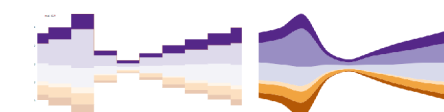


[https://www.d3-graph-gallery.com/graph/barplot\\_stacked\\_basicWide.html](https://www.d3-graph-gallery.com/graph/barplot_stacked_basicWide.html)

23

## Idiom: streamgraph

- generalized stacked graph
  - emphasizing horizontal continuity
    - vs vertical items
  - data
    - 1 categ key attrib (movies)
    - 1 ordered key attrib (time)
    - 1 quant value attrib (counts)
  - derived data
    - geometry: layers, where height encodes counts
    - 1 quant attrib (layer ordering)



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

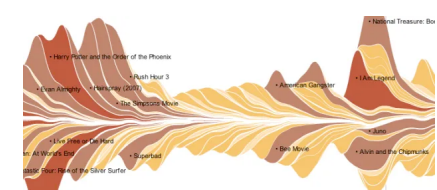
24

## Idiom: streamgraph

- generalized stacked graph
  - emphasizing horizontal continuity
    - vs vertical items
  - data
    - 1 categ key attrib (movies)
    - 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 movies keys
      - more than stacked bars: most layers don't extend across whole chart



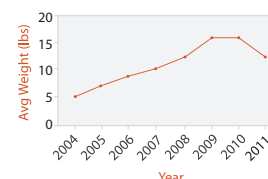
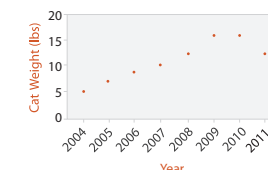
[Stacked Graphs Geometry & Aesthetics, Byron and Wattenberg, IEEE Trans Visualization and Computer Graphics (Proc. InfoVis 2008) 14(6): 1245–1252, (2008).]



<https://flowingdata.com/2008/02/25/ebb-and-flow-of-box-office-receipts-over-past-20-years/>

## Idiom: dot / line chart

- one key, one value
  - data
    - 2 quant attribs
  - mark: points
    - AND line connection marks between them
  - channels
    - aligned lengths to express quant value
    - separated and ordered by key attrib into horizontal regions

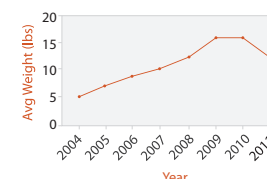
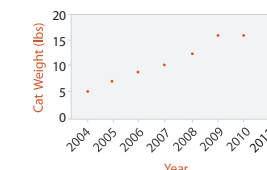


25

26

## Idiom: dot / line chart

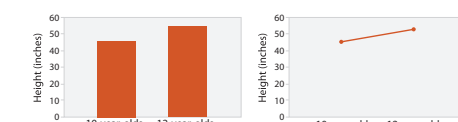
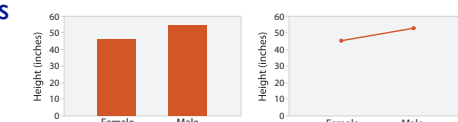
- one key, one value
  - data
    - 2 quant attribs
  - mark: points
    - AND line connection marks between them
  - channels
    - aligned lengths to express quant value
    - separated and ordered by key attrib into horizontal regions
  - task
    - find trend
      - connection marks emphasize ordering of items along key axis by explicitly showing relationship between one item and the next
  - scalability
    - hundreds of key levels, hundreds of value levels



27

## 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”

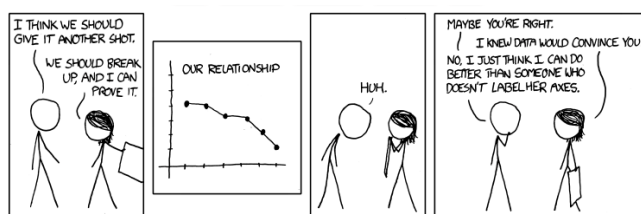


after [Bars and Lines: A Study of Graphic Communication, Zacks and Tversky, Memory and Cognition 27:6 (1999), 1073–1079.]

28

## Chart axes: label them!

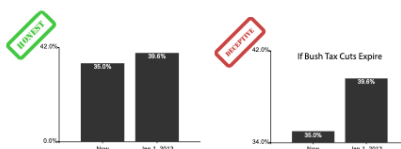
- best practice to label
  - few exceptions: individual small multiple views could share axis label



<https://xkcd.com/833/>

## Chart axes: avoid cropping y axis

- include 0 at bottom left or slope misleads

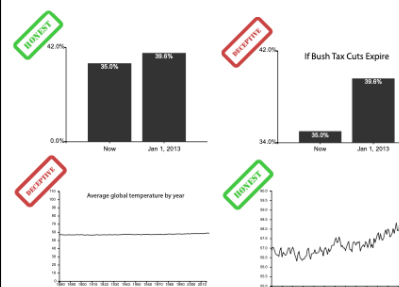


[Truncating the Y-Axis: Threat or Menace? Correll, Bertini, & Franconeri, CHI 2020.]

29

## Chart axes: avoid cropping y axis

- include 0 at bottom left or slope misleads
  - some exceptions (arbitrary 0, small change matters)



[Truncating the Y-Axis: Threat or Menace? Correll, Bertini, & Franconeri, CHI 2020.]

30

## Chart axes: avoid cropping y axis

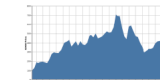
- include 0 at bottom left or slope misleads
  - some exceptions (arbitrary 0, small change matters)



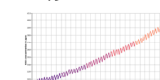
(a) Statistical process charts rely on comparison to an expected value, and so deviations from that value, not from zero, are important.



(b) Index charts compare to an indexed value rather than zero.



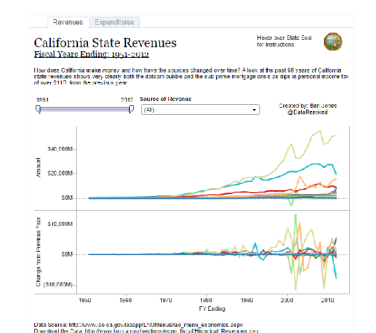
(c) Stock charts must show small differences in stock value, so these can translate to enormous monetary gains or losses.



(d) Climate Anomaly charts rely on both highlighting variation from a non-zero expected value but also emphasize the potentially disastrous impact of even minute changes in climate.

## Idiom: Indexed line charts

- data: 2 quant attribs
  - 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

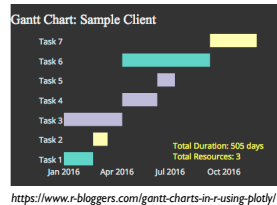


<https://public.tableau.com/profile/benjen#vizhome/CASStateRevenues/Revenues>

31

## Idiom: Gantt charts

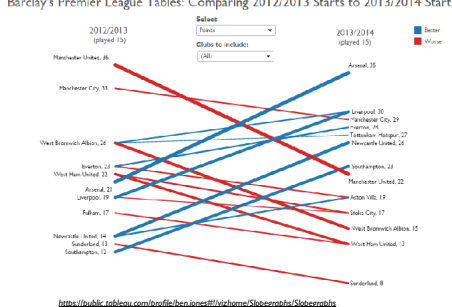
- one key, two (related) values
  - data
    - 1 categ attrib, 2 quant attribs
  - mark: line
    - length: duration
  - channels
    - horiz position: start time (+end from duration)
  - task
    - emphasize temporal overlaps & start/end dependencies between items
  - scalability
    - dozens of key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



33

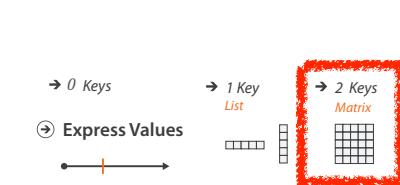
## Idiom: Slopegraphs

- two values
  - data
    - 2 quant value attribs
    - (1 derived attrib: change magnitude)
  - mark: point + line
    - line connecting mark between pts
  - channels
    - 2 vertical pos: express attrib value
    - (linewidth/size, color)
  - task
    - emphasize changes in rank/value
  - scalability
    - hundreds of value levels
    - dozens of items



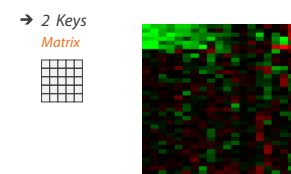
34

## 2 Keys



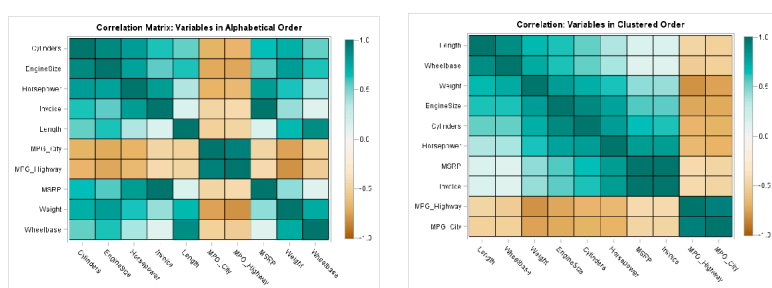
## Idiom: heatmap

- two keys, one value
  - data
    - 2 categ attribs (gene, experimental condition)
    - 1 quant attrib (expression levels)
  - marks: point
    - 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



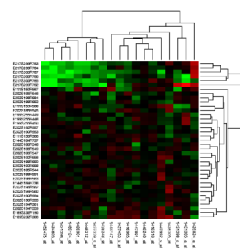
36

## Heatmap reordering



## 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

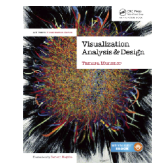


38

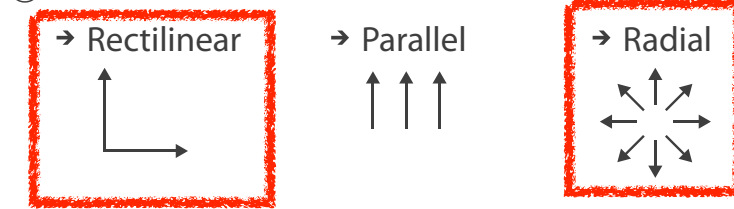
## Visualization Analysis & Design

### Tables (Ch 7) II

Tamara Munzner  
Department of Computer Science  
University of British Columbia  
@tamaramunzner



## Axis Orientation

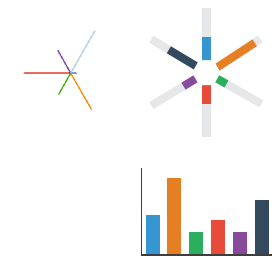


<https://blogs.sas.com/content/iml/2018/05/02/reorder-variables-correlation-heat-map.html>

37

## Idioms: radial bar chart, star plot

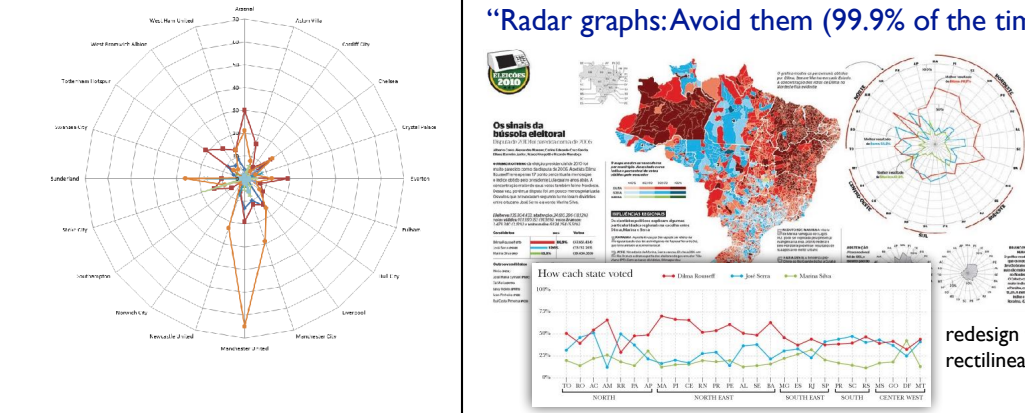
- star plot
  - line mark, radial axes meet at central point
- radial bar chart
  - line mark, radial axes meet at central ring
  - channels: length, angle/orientation
- bar chart
  - rectilinear axes, aligned vertically
- accuracy
  - length not aligned with radial layouts
  - less accurately perceived than rectilinear aligned



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

## Idiom: radar plot

- radial line chart
  - point marks, radial layout
  - connecting line marks
- avoid unless data is cyclic



<http://www.thefunctionalart.com/2012/11/1/radar-graphs-avoid-them-999-of-the-time.html>

42

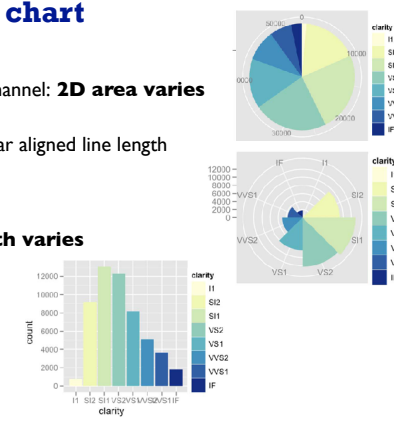
## “Radar graphs: Avoid them (99.9% of the time)”

original difficult to interpret

redesign for rectilinear

## Idioms: pie chart, coxcomb chart

- pie chart
  - interlocking area marks with angle channel: 2D area varies
  - separated & ordered radially, uniform height
  - accuracy: area less accurate than rectilinear aligned line length
  - task: part-to-whole judgements
- coxcomb chart
  - line marks with length channel: 1D length varies
  - separated & ordered radially, uniform width
  - direct analog to radial bar charts
- data
  - 1 categ key attrib, 1 quant value attrib

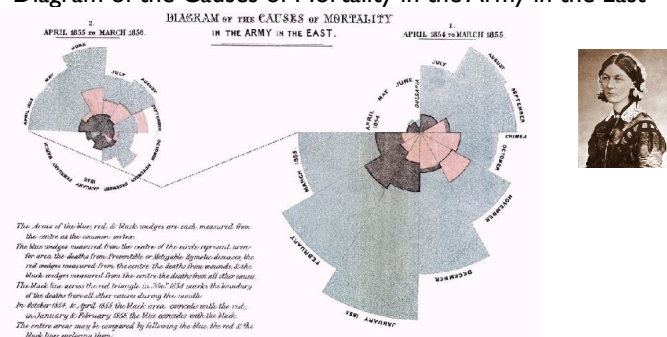


[A layered grammar of graphics, Wickham, Joun. Computational and Graphical Statistics 19:1 (2010), 3–28.]

44

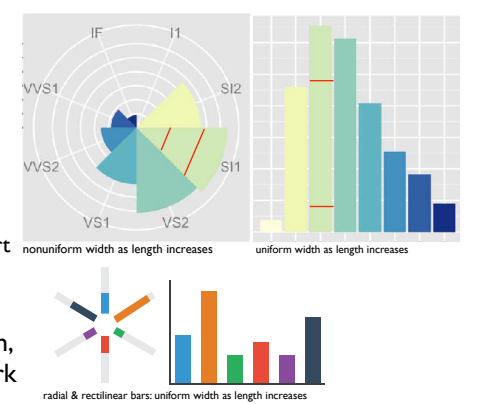
## Coxcomb / nightingale rose / polar area chart

- invented by Florence Nightingale:  
Diagram of the Causes of Mortality in the Army in the East



## Coxcomb: perception

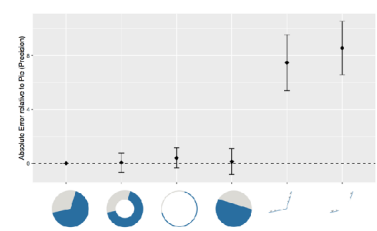
- encode: 1D length
- decode/perceive: 2D area
- nonuniform line/sector width as length increases
  - so area variation is nonlinear wrt line mark length!
- bar chart safer: uniform width, so area is linear with line mark length
  - both radial & rectilinear cases



46

## Pie charts: perception

- some empirical evidence that people respond to arc length
  - decode/perceive: 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.]

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

## Pie charts: best practices

- not so bad for two (or few) levels, for part-to-whole task

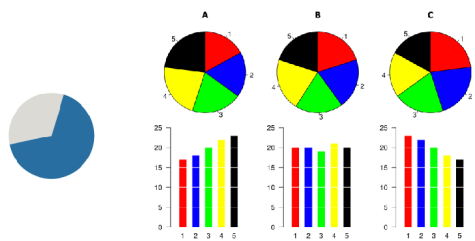


<https://eagereyes.org/pie-charts>

48

## Pie charts: best practices

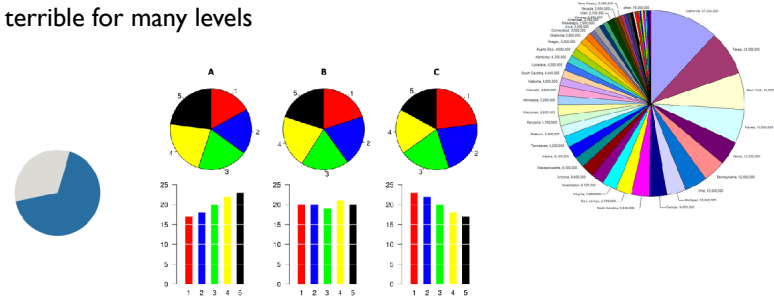
- not so bad for two (or few) levels, for part-to-whole task
- dubious for several levels if details matter



<https://eagereyes.org/pie-charts>

## Pie charts: best practices

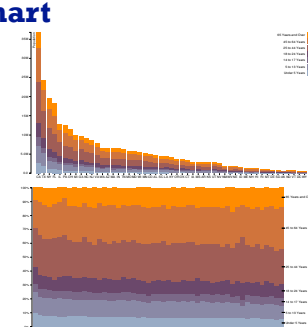
- not so bad for two (or few) levels, for part-to-whole task
- dubious for several levels if details matter
- terrible for many levels



<https://eagereyes.org/pie-charts>

## Idioms: normalized stacked bar chart

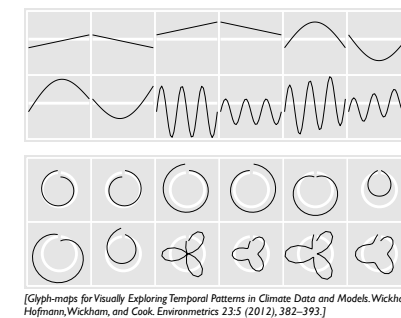
- task
  - part-to-whole judgements
- normalized stacked bar chart
  - stacked bar chart, normalized to full vert height
  - single stacked bar equivalent to full pie
    - high information density: requires narrow rectangle
- pie chart
  - information density: requires large circle



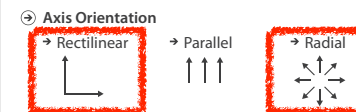
<http://nl.ocks.org/imbostock/3886208>  
<http://nl.ocks.org/imbostock/3887235>  
<http://nl.ocks.org/imbostock/3886394>

## Idiom: glyphmaps

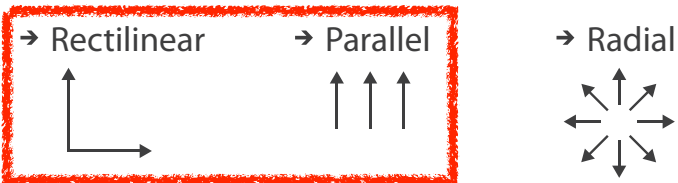
- rectilinear good for linear vs nonlinear trends
- radial good for cyclic patterns
  - evaluating periodicity



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

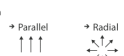
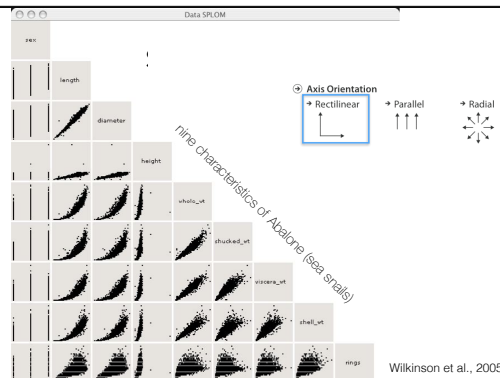


## Axis Orientation



## Idiom: SPLOM

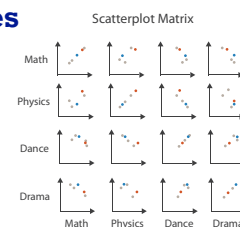
- scatterplot matrix (SPLOM)
  - rectilinear axes, point mark
  - all possible pairs of axes
  - scalability
    - one dozen attribs
    - dozens to hundreds of items



Wilkinson et al., 2005

## Idioms: parallel coordinates

- scatterplot limitation
  - visual representation with orthogonal axes
  - can show only two attributes with spatial position channel



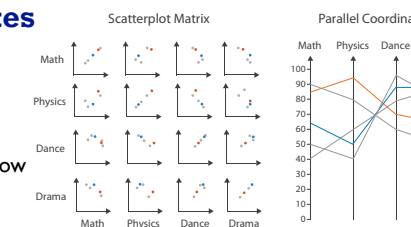
Table

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

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

## Idioms: parallel coordinates

- scatterplot limitation
  - visual representation with orthogonal axes
  - can show only two attributes with spatial position channel
- alternative: line up axes in parallel to show many attributes with position
  - item encoded with a line with n segments
  - n is the number of attributes shown
- parallel coordinates
  - parallel axes, jagged line for item
  - rectilinear axes, item as point
    - axis ordering is major challenge
  - scalability
    - dozens of attribs
    - hundreds of items



Table

	Math	Physics	Dance	Drama
Math	85	95	70	65
Physics	90	80	60	50
Dance	65	50	90	90
Drama	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: spread out
- parallel coordinates
  - positive correlation
    - parallel line segments
  - negative correlation
    - all segments cross at halfway point
  - uncorrelated
    - scattered crossings

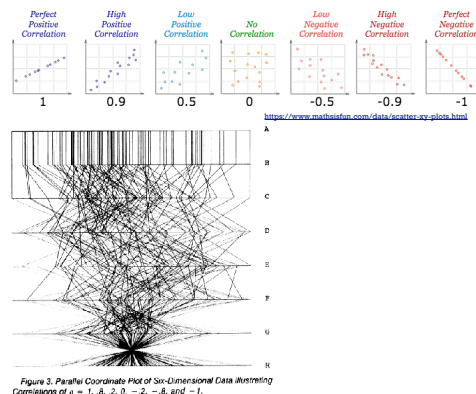
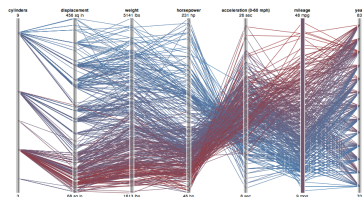


Figure 3. Parallel Coordinates Plot of Six-Dimensional Data illustrating Correlations of  $\rho = 1, .8, .5, 0, -.5, -.8, \text{ and } -1$ .  
 [Hyperdimensional Data Analysis Using Parallel Coordinates. Wegman, *Journ. American Statistical Association* 85:411 (1990), 664-675.]

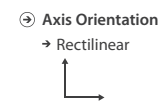
## Parallel coordinates, limitations

- visible patterns only between neighboring axis pairs
- how to pick axis order?
  - usual solution: reorderable axes, interactive exploration
  - same weakness as many other techniques
    - downside of interaction: human-powered search
  - some algorithms proposed, none fully solve



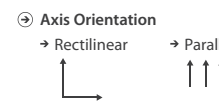
## Orientation limitations

- rectilinear: scalability wrt #axes
  - 2 axes best, 3 problematic, 4+ impossible



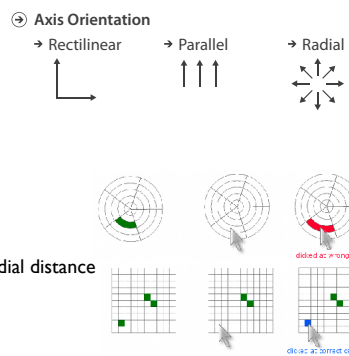
## Orientation limitations

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



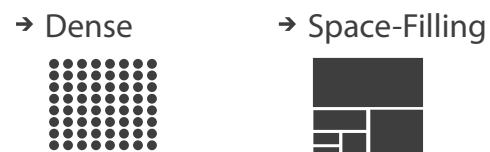
## Orientation limitations

- rectilinear: scalability wrt #axes
  - 2 axes best, 3 problematic, 4+ impossible
- parallel: unfamiliarity, training time
- radial: perceptual limits
  - polar coordinate asymmetry
    - angles lower precision than length
    - nonuniform sector width/size depending on radial distance
  - frequently problematic
    - but sometimes can be deliberately exploited!
      - for 2 attribs of very unequal importance



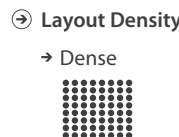
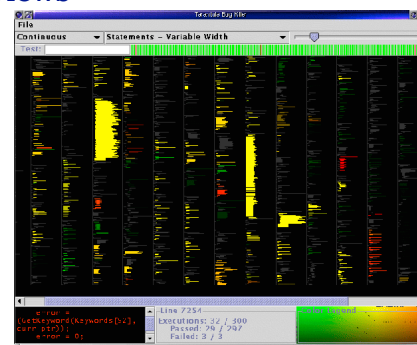
## Layout density

### Layout Density



## Idiom: Dense software overviews

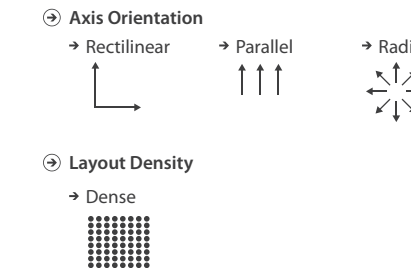
- data: text
  - text + 1 quant attrib per line
- derived data:
  - one pixel high line
  - length according to original
- color line by attrib
- scalability
  - 10K+ lines



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

## Arrange tables

- Express Values
  - Separate, Order, Align Regions
    - Separate
    - Order
  - Align
- 1 Key List
  - 2 Keys Matrix



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

Encode	Manipulate	Facet	Reduce
<ul style="list-style-type: none"> <li>Arrange               <ul style="list-style-type: none"> <li>Express</li> <li>Separate</li> <li>Order</li> <li>Align</li> <li>Use</li> </ul> </li> <li>Map from categorical and ordered attributes               <ul style="list-style-type: none"> <li>Color                   <ul style="list-style-type: none"> <li>Hue</li> <li>Saturation</li> <li>Luminance</li> </ul> </li> <li>Size, Angle, Curvature, ...</li> <li>Shape                   <ul style="list-style-type: none"> <li>+</li> <li>●</li> <li>■</li> <li>▲</li> </ul> </li> <li>Motion                   <ul style="list-style-type: none"> <li>Direction, Rate, Frequency, ...</li> </ul> </li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Change</li> <li>Select</li> <li>Navigate</li> </ul>	<ul style="list-style-type: none"> <li>Juxtapose</li> <li>Partition</li> <li>Superimpose</li> </ul>	<ul style="list-style-type: none"> <li>Filter</li> <li>Aggregate</li> <li>Embed</li> </ul>

What?  
Why?  
How?

Encode	Manipulate	Facet	Reduce
<ul style="list-style-type: none"> <li>Arrange               <ul style="list-style-type: none"> <li>Express</li> <li>Separate</li> <li>Order</li> <li>Align</li> <li>Use</li> </ul> </li> <li>Map from categorical and ordered attributes               <ul style="list-style-type: none"> <li>Color                   <ul style="list-style-type: none"> <li>Hue</li> <li>Saturation</li> <li>Luminance</li> </ul> </li> <li>Size, Angle, Curvature, ...</li> <li>Shape                   <ul style="list-style-type: none"> <li>+</li> <li>●</li> <li>■</li> <li>▲</li> </ul> </li> <li>Motion                   <ul style="list-style-type: none"> <li>Direction, Rate, Frequency, ...</li> </ul> </li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Change</li> <li>Select</li> <li>Navigate</li> </ul>	<ul style="list-style-type: none"> <li>Juxtapose</li> <li>Partition</li> <li>Superimpose</li> </ul>	<ul style="list-style-type: none"> <li>Filter</li> <li>Aggregate</li> <li>Embed</li> </ul>

What?  
Why?  
How?