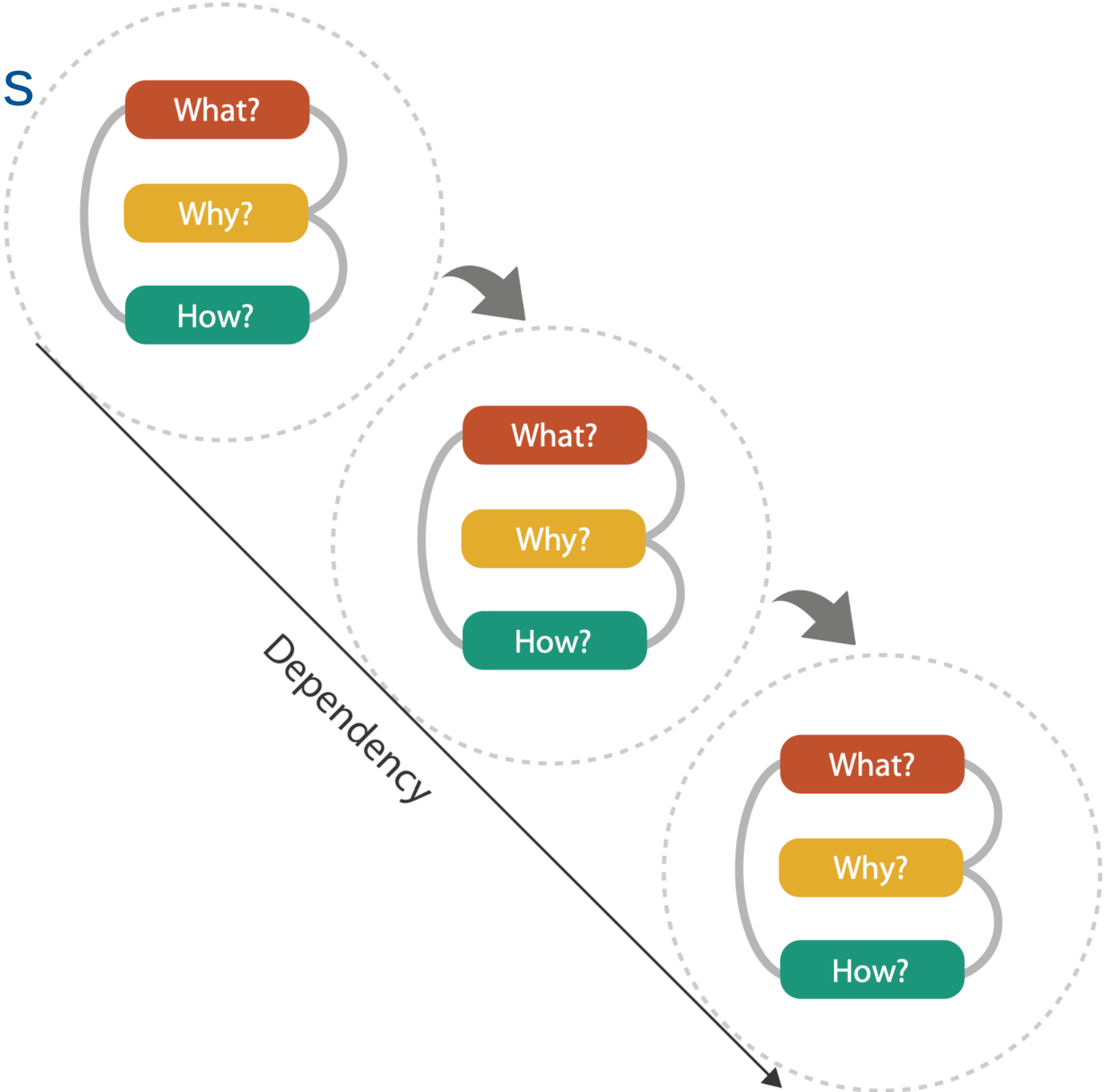
Visualization for Data Science
The Zoo II

Means and ends



What?

Why?

How?

What?

Datasets

Data Types

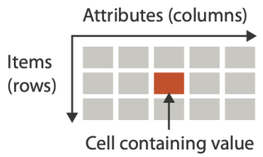
- Items
- Attributes
- Links
- Positions
- Grids

Data and Dataset Types

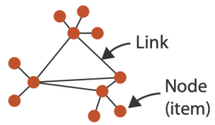
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Dataset Types

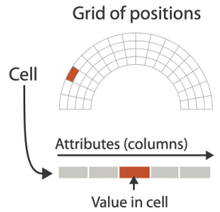
- Tables



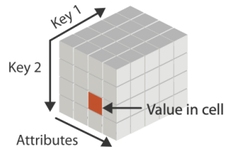
- Networks



- Fields (Continuous)



- *Multidimensional Table*



- *Trees*

- Geometry (Spatial)

Dataset Availability

- Static

- Dynamic

Attributes

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

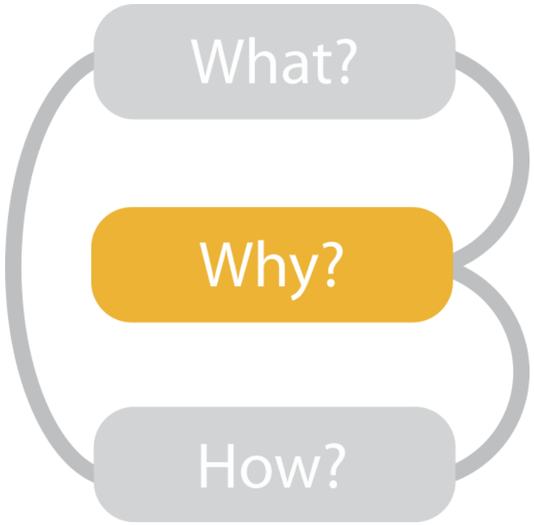
  - *Quantitative*

Ordering Direction

- Sequential

- Diverging

- Cyclic



- {action, target} pairs
  - *discover distribution*
  - *compare trends*
  - *locate outliers*
  - *browse topology*

Why?

Actions

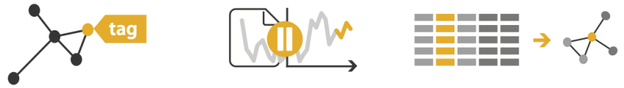
Analyze

→ Consume

→ *Discover* → *Present* → *Enjoy*

→ Produce

→ *Annotate* → *Record* → *Derive*

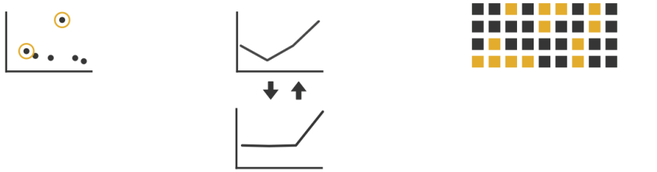


Search

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

Query

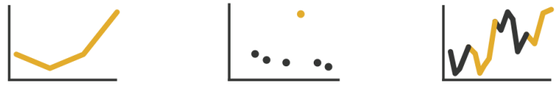
→ Identify → Compare → Summarize



Targets

All Data

→ Trends → Outliers → Features



Attributes

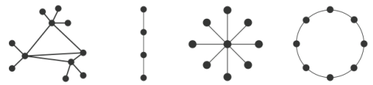
→ One → Many

→ *Distribution* → *Dependency* → *Correlation* → *Similarity*

→ *Extremes*

Network Data

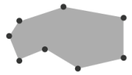
→ Topology



→ *Paths*

Spatial Data

→ Shape



How?

Encode

Arrange

- Express
- Separate
- Order
- Align
- Use

Map

from categorical and ordered attributes

- Color
  - *Hue*
  - *Saturation*
  - *Luminance*
- Size, Angle, Curvature, ...
- Shape
- Motion

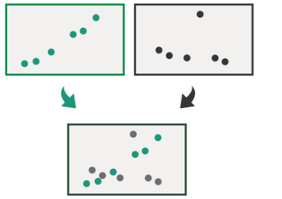
Direction, Rate, Frequency, ...

Manipulate

- Change
- Select
- Navigate

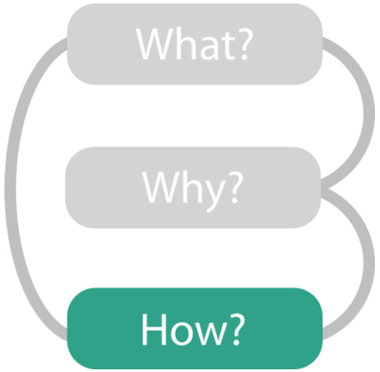
Facet

- Juxtapose
- Partition
- Superimpose



Reduce

- Filter
- Aggregate
- Embed



Encode

Arrange

→ Express

→ Separate

→ Order

→ Align

→ Use

Accuracy: User studies

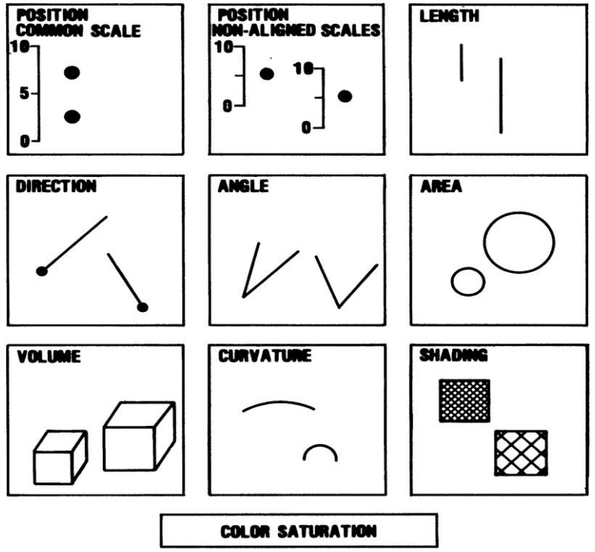


Figure 1. Elementary perceptual tasks.

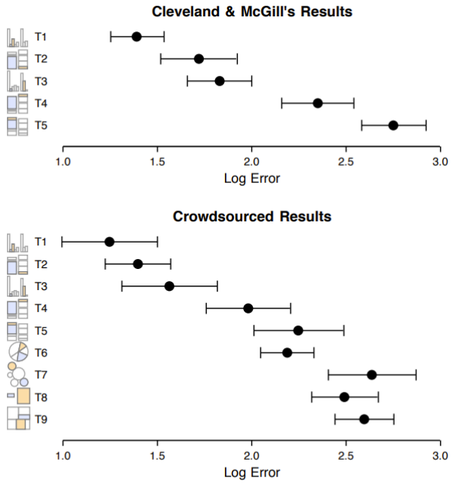
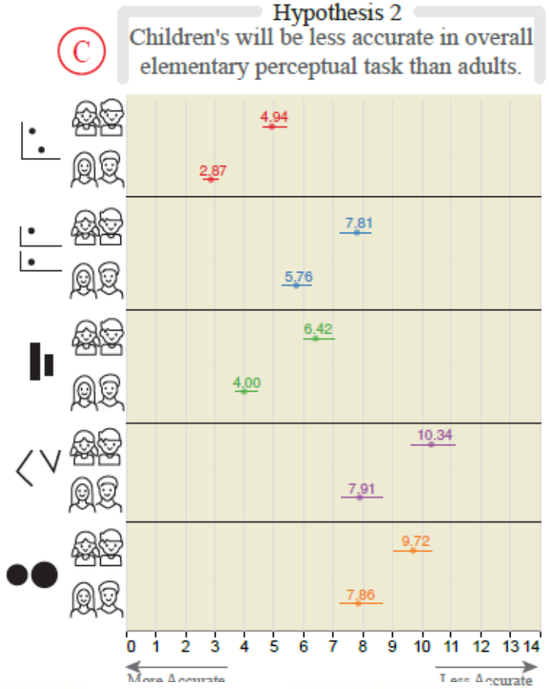
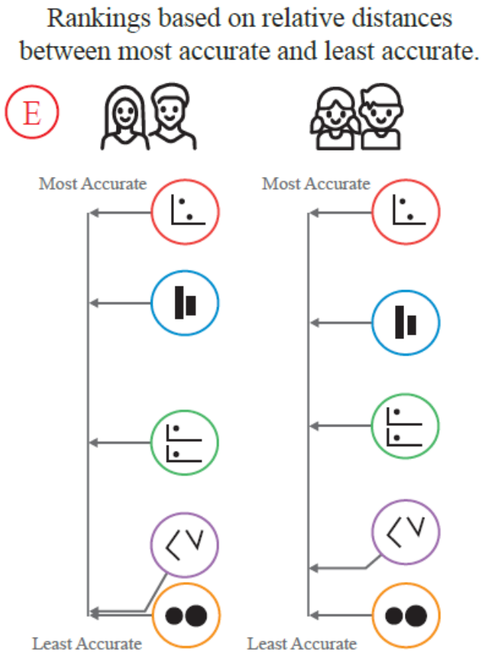


Figure 4: Proportional judgment results (Exp. 1A & B). Top: Cleveland & McGill's [7] lab study. Bottom: MTurk studies. Error bars indicate 95% confidence intervals.





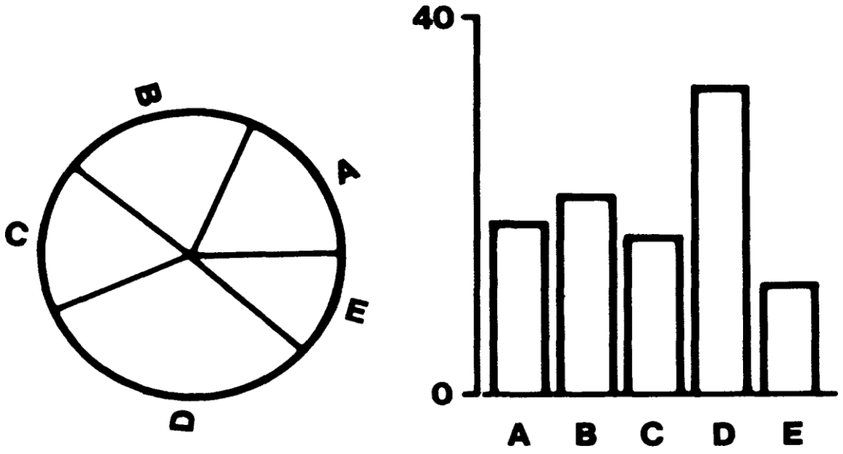
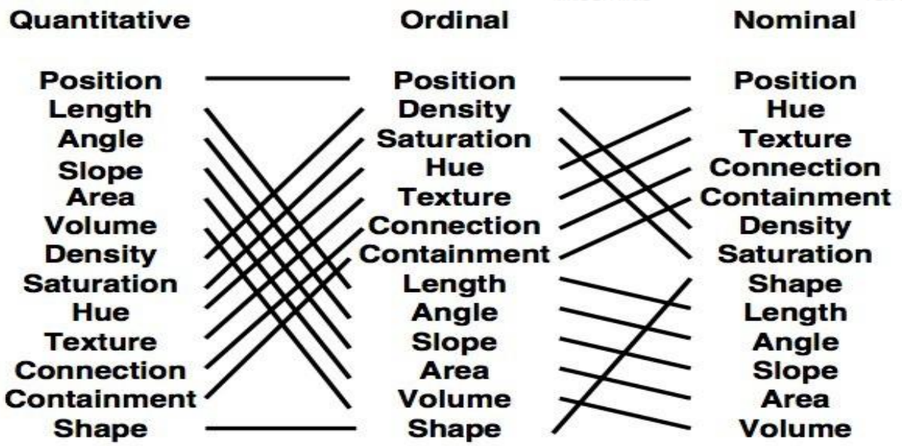


Figure 3. Graphs from position–angle experiment.



[Cleveland & McGill, 1984]
[Mackinlay, 1986]
[Heer & Bostock, 2010]
[Panavas et al., 2022]

Express Values

→ *0 Keys*

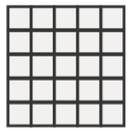
x, y both quantity

→ *1 Key*

List

→ *2 Keys*

Matrix



Learning Outcomes

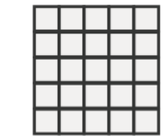
- Select a visualization based on the data and task
- Describe visualizations in terms of keys and values that they encode

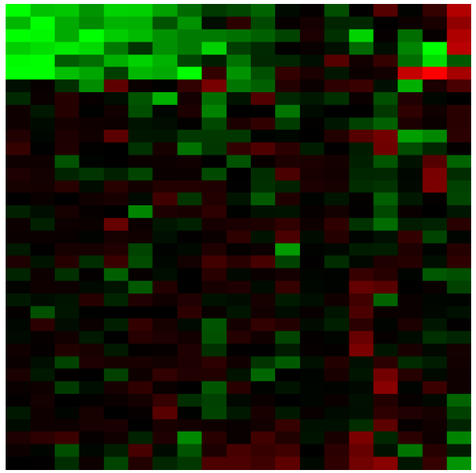
Idiom: **heatmap**

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

➔ *2 Keys*

Matrix

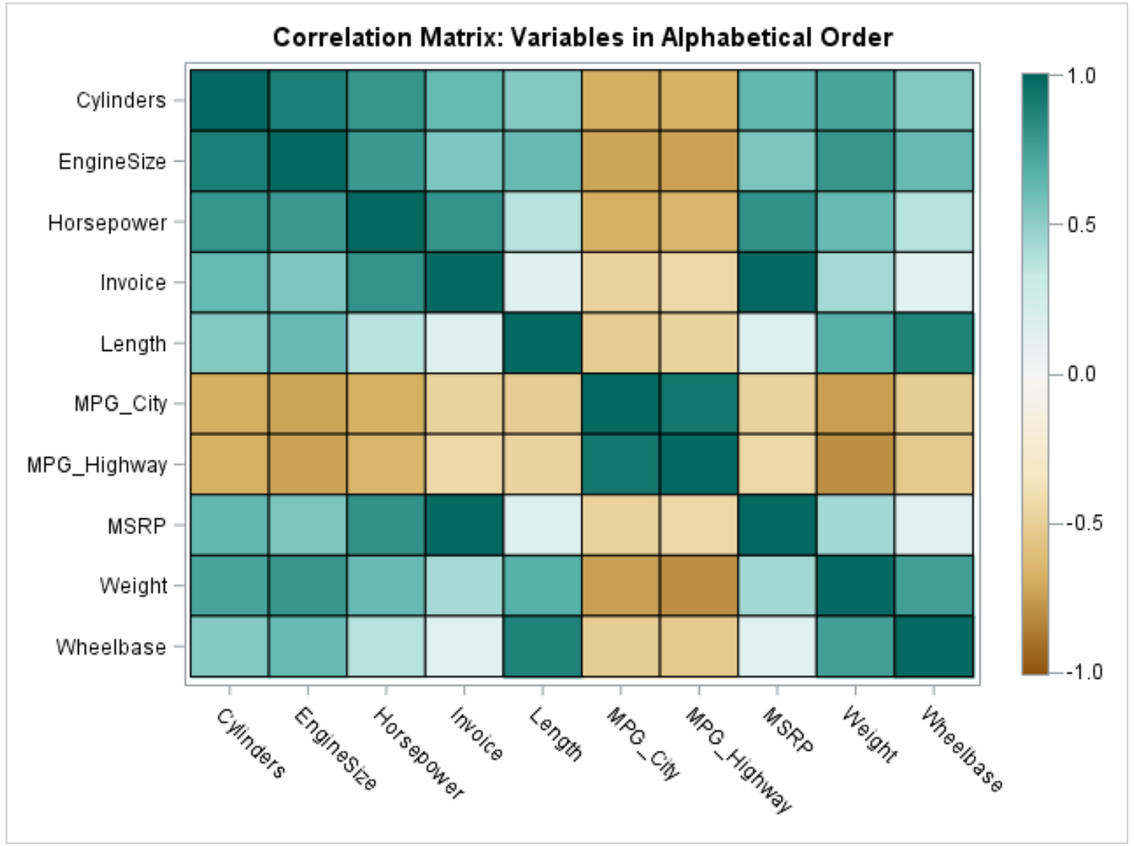


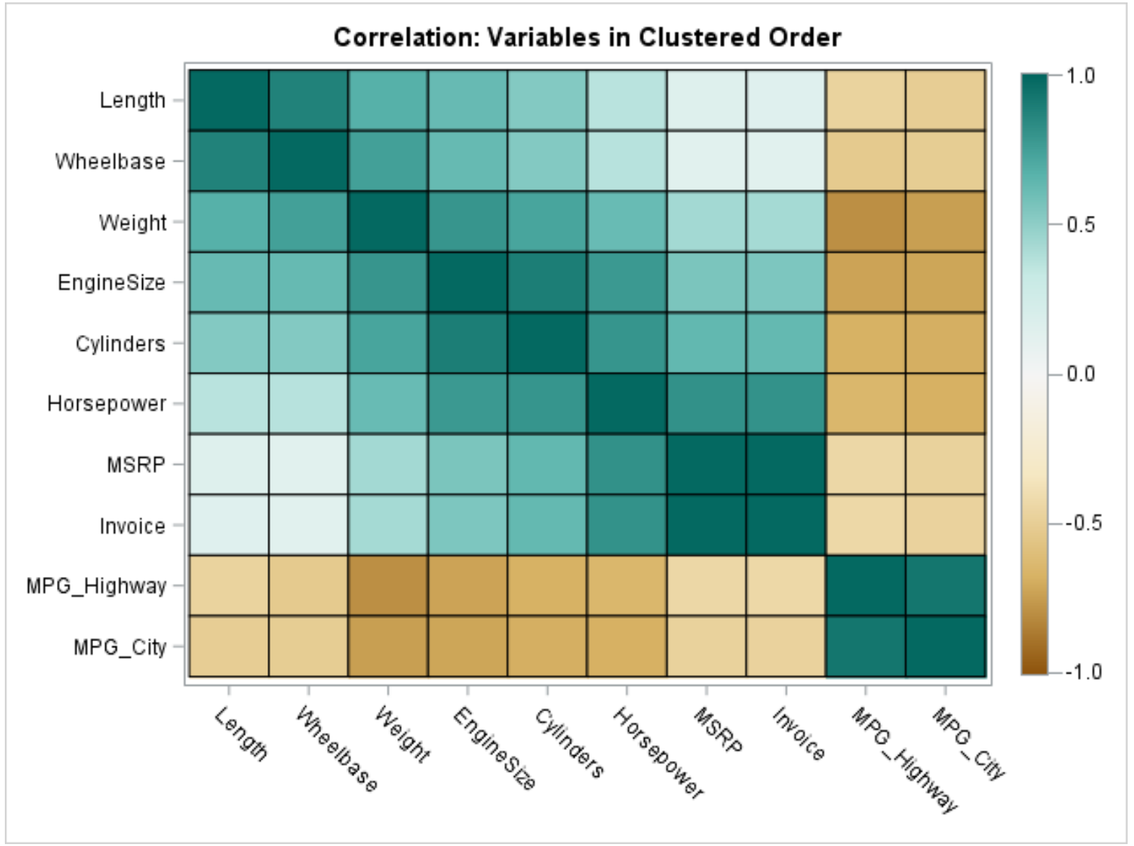


How many shades of red can you see

A. 2
B. 3
C. 4
D. 5
E. > 5

Heatmap reordering



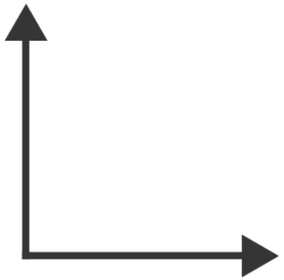


https://blogs.sas.com/content/iml/2018/05/02/reorder-variables-correlation-heat-map.html

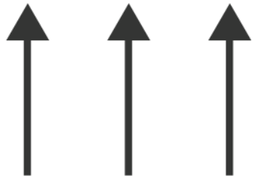
https://bost.ocks.org/mike/miserables/

Axis Orientation

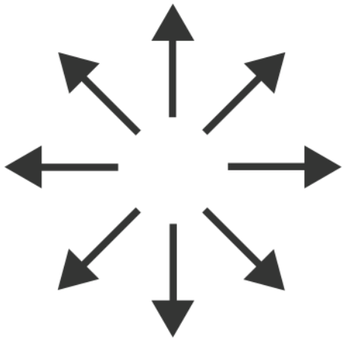
- Rectilinear



- Parallel

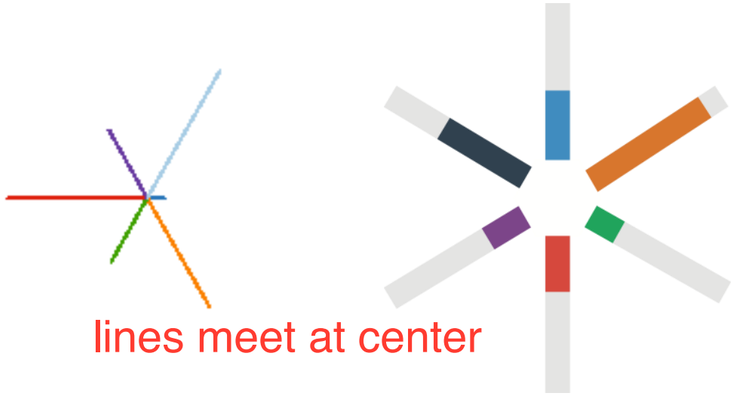


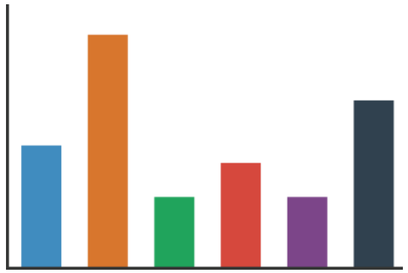
- Radial



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





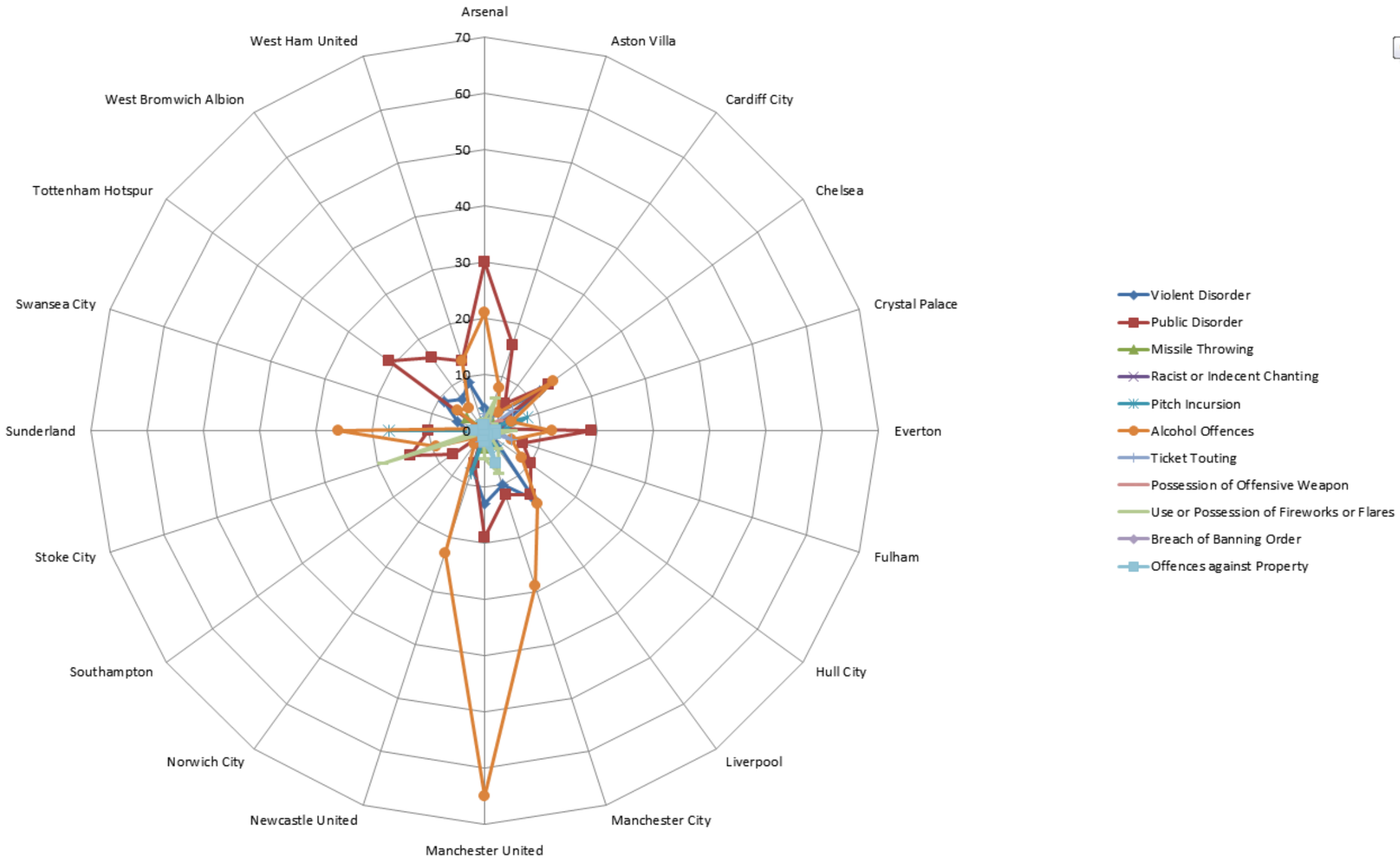
- accuracy
  - length not aligned with radial layouts
    - less accurately perceived than rectilinear aligned

[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.]

Idiom: **radar plot**

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

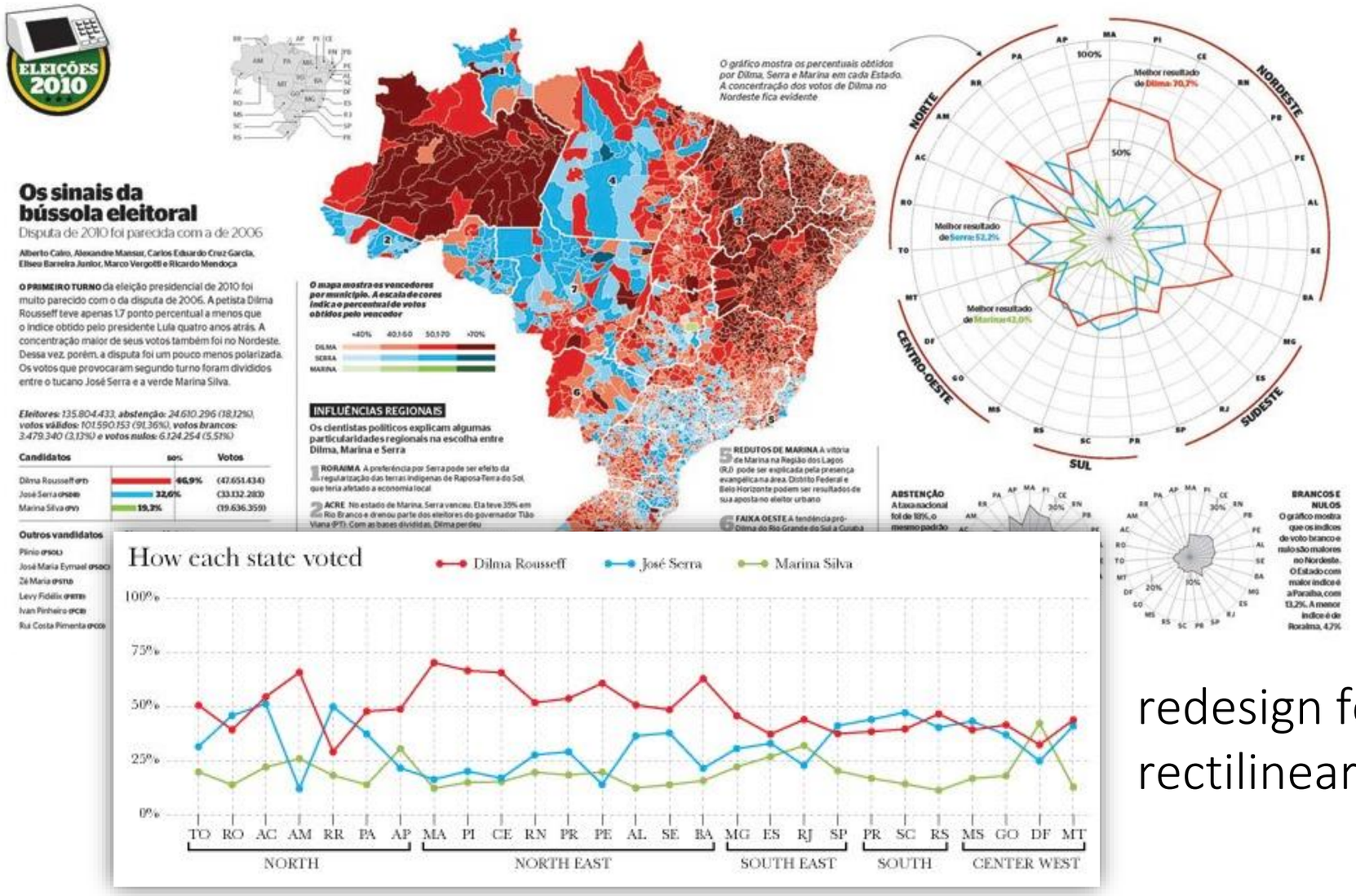
no distinct ordering for the category on the circle



shape depicts the meaning

not able to see all data point

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

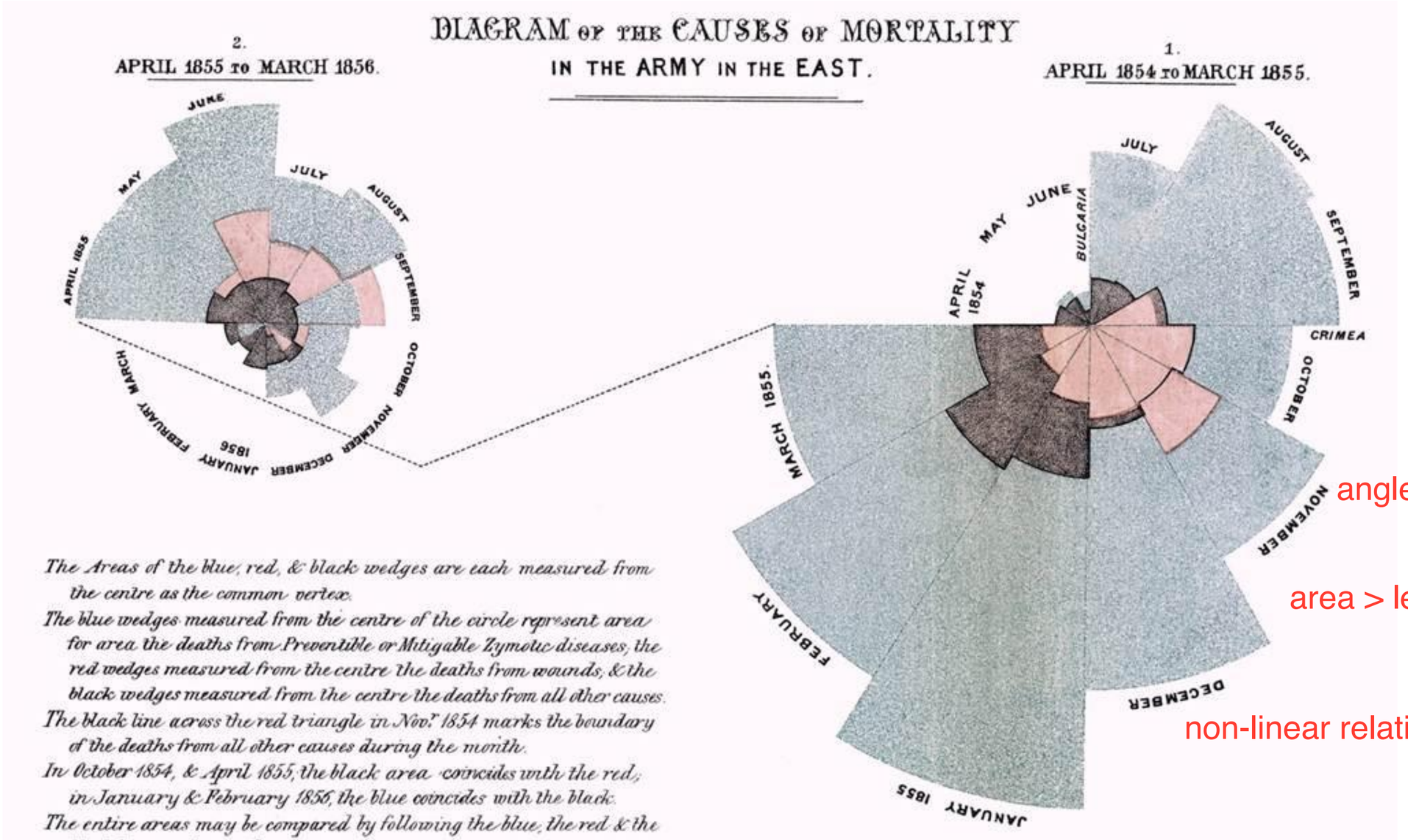


original
difficult to interpret

redesign for
rectilinear

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

Coxcomb / nightingale rose / polar area chart



angles are the same

area > length

non-linear relationship

Florence Nightingale

https://www.youtube.com/watch?v=u6XqiDccroM

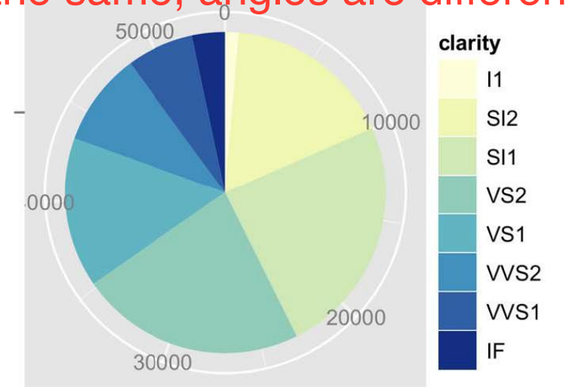
As we watch the video pay attention to the following

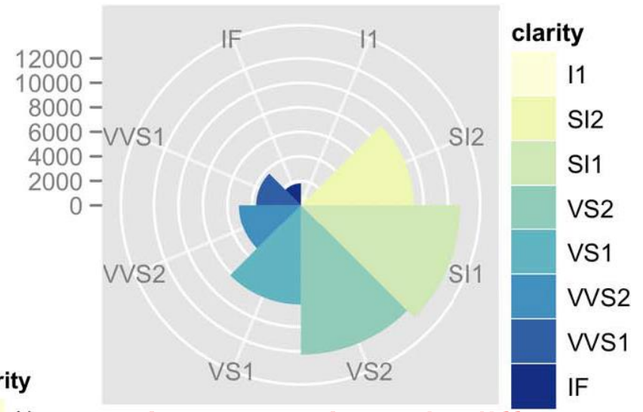
1. The Structure
2. The data represented
3. The case for visualization that Florence Nightingale made
4. The case for sanitation

Idioms: **pie chart, coxcomb chart**

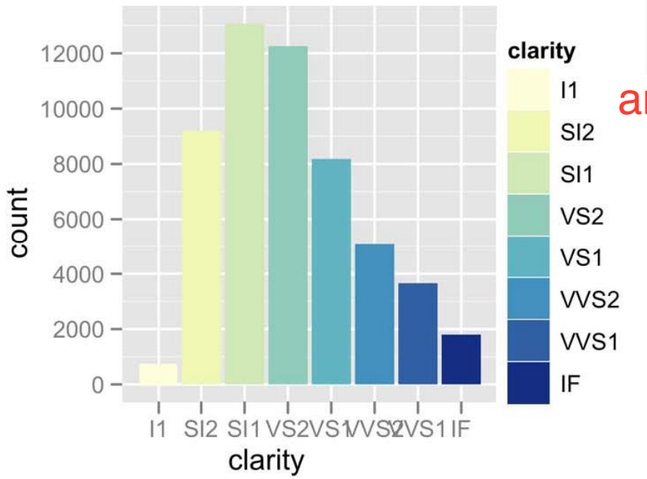
- pie chart
  - interlocking area marks with angle channel: **2D area varies**
    - separated & ordered radially, uniform height
  - accuracy: area *less accurate* than rectilinear aligned line length
  - **task: part-to-whole judgements**
- coxcomb chart
  - line marks with length channel: **1D length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

length (radius) is the same, angles are different





angle same, length different



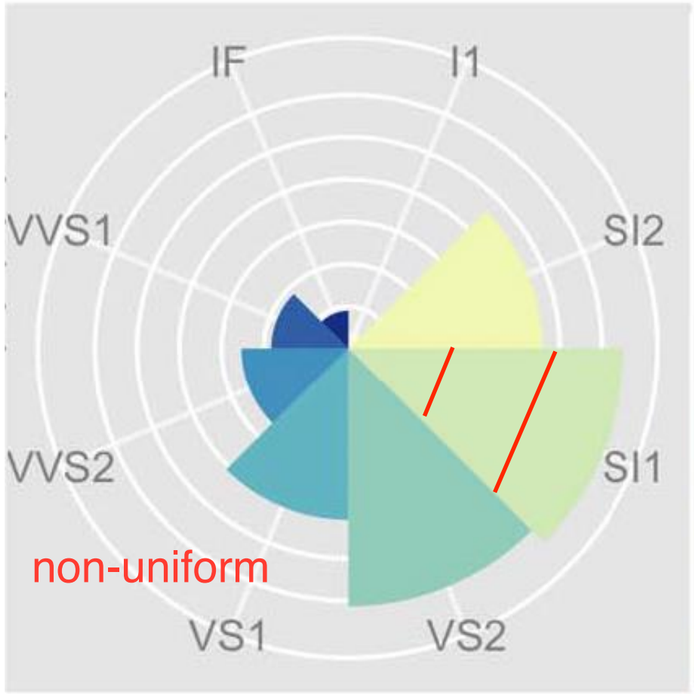
[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]

Coxcomb: perception

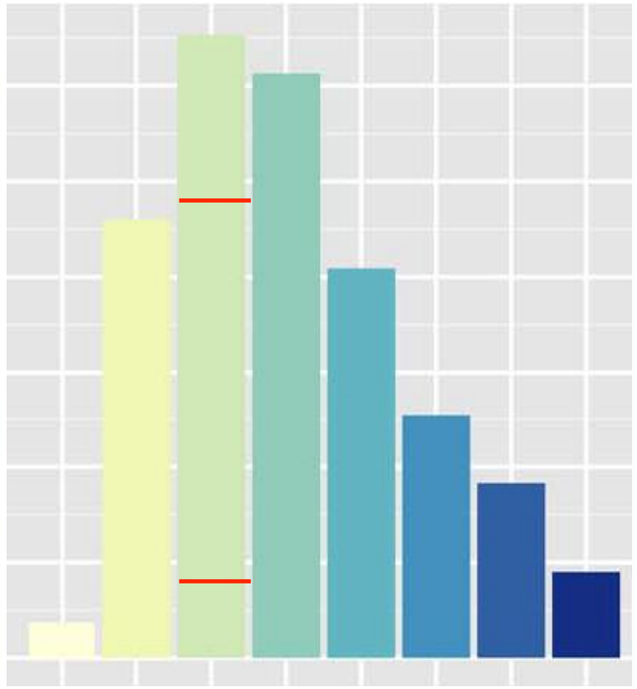
- encode: 1D length
- decode/perceive: 2D area

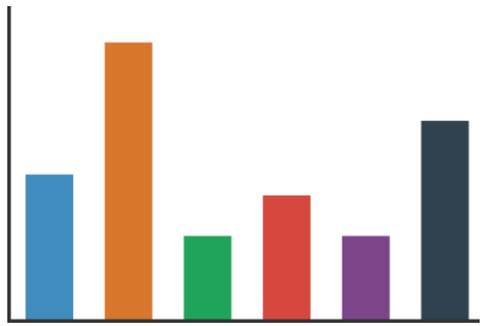
- nonuniform line/sector width as length increases
  - so area variation is nonlinear wrt line mark length!

- bar chart safer: uniform width, so area is linear with line mark length
  - both radial & rectilinear cases



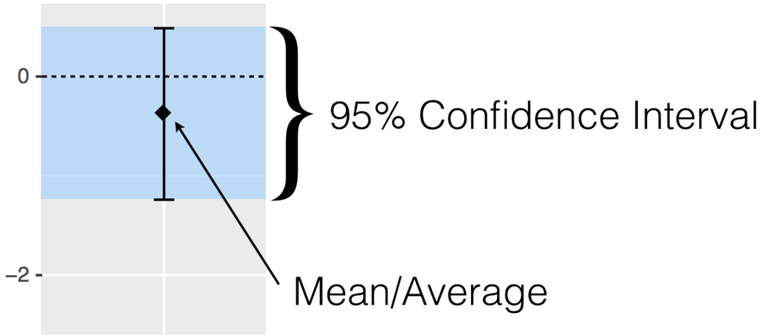
nonuniform width as length increases


uniform width as length increases

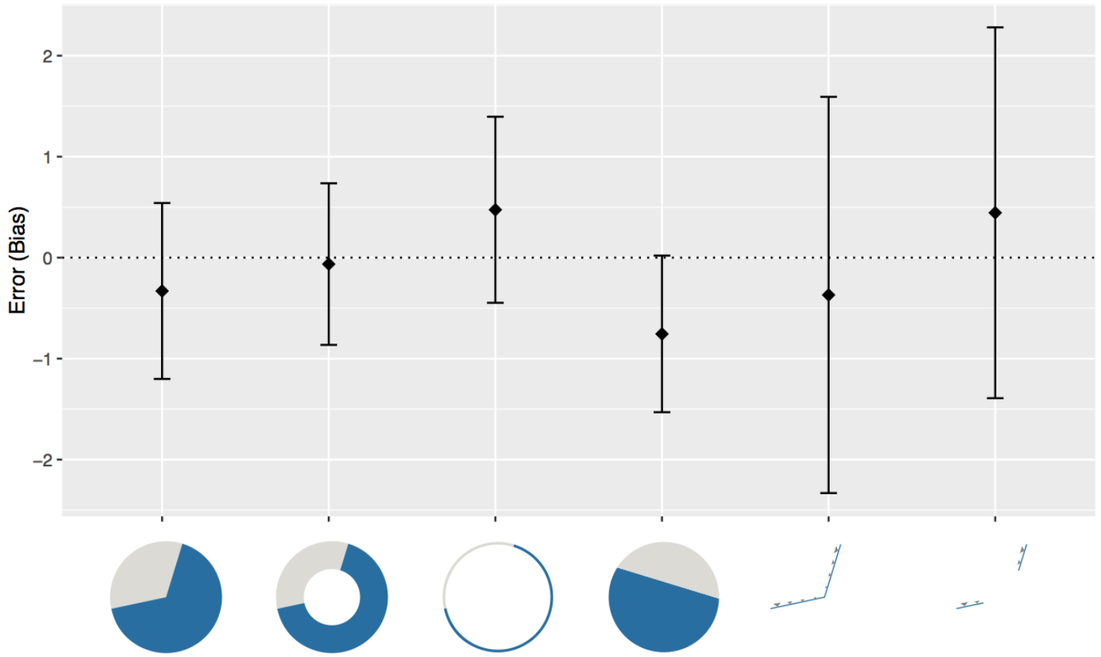

radial & rectilinear bars: uniform width as length increases

Pie charts: perception

- some empirical evidence that people respond to arc length
- what we encode is different that what people perceive, most people do not read pie charts by angle
  - It could be areas, arc length alone, or combination
- donut charts no worse than pie charts and this is a direct result from the fact that most people don’t read it by angle



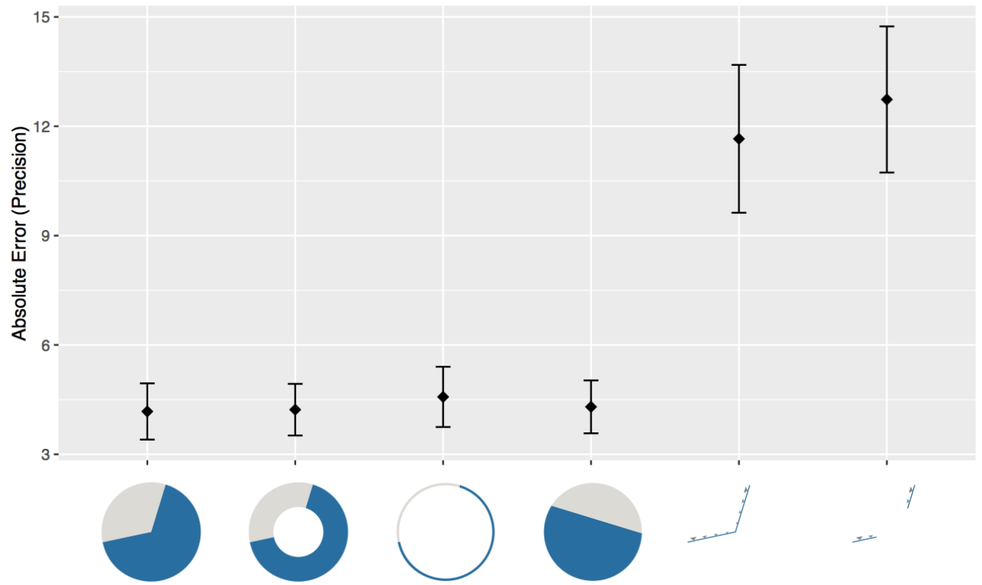
Pie charts: perception



Pie donut Arc length (thin donut) Pie-diff (area-only) arcs arcs unconnected

Each of these bars shows us the signed error. That means we can get a sense of the deviation from the real value (the lengths of the bars), but especially the bias: were people systematically over- or underestimating?

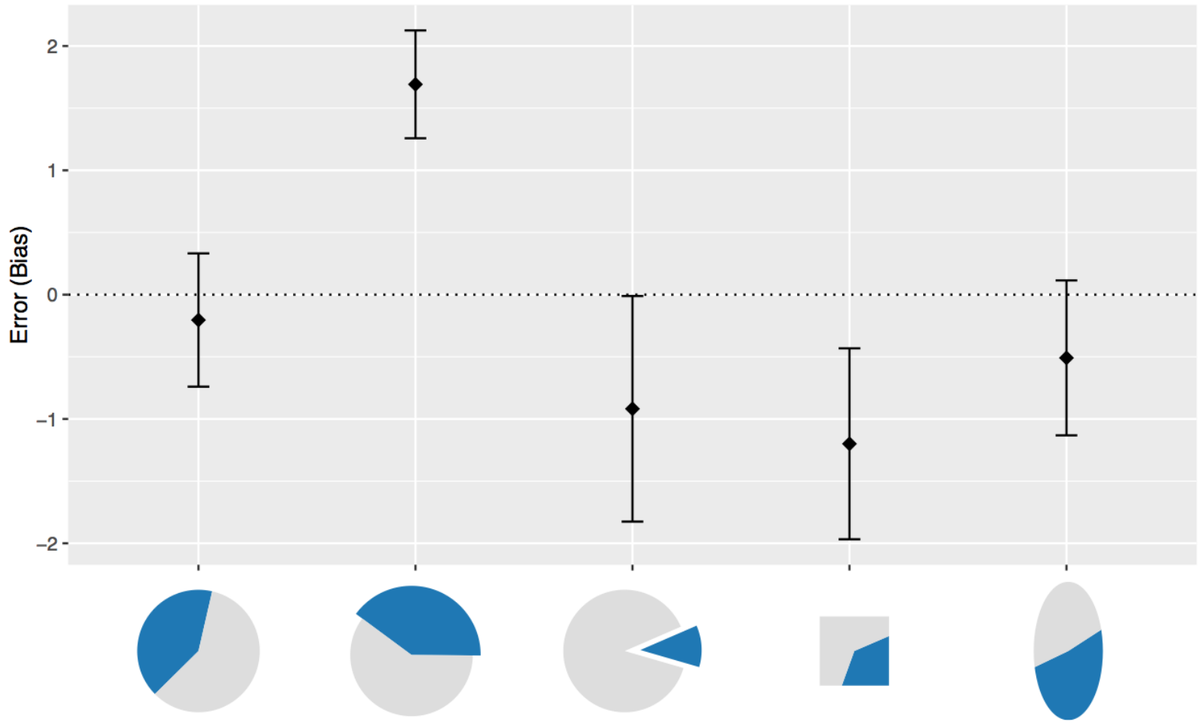
arc length/ angle/ area

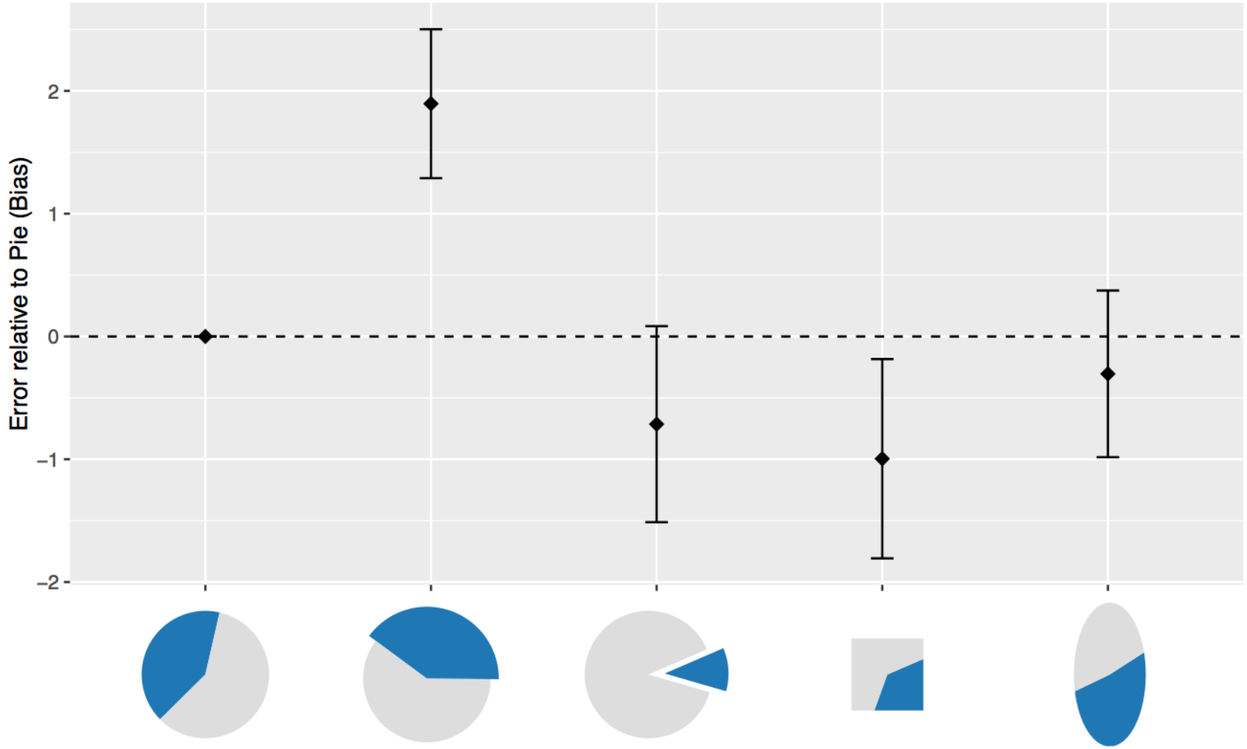


More interesting than signed error (where we take the average of all the errors, both above and below) is absolute error, where we count all deviations in the same direction. That means errors don't even out, so we can see how far off people are, no matter if above or below the correct value.

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

Pie charts variations



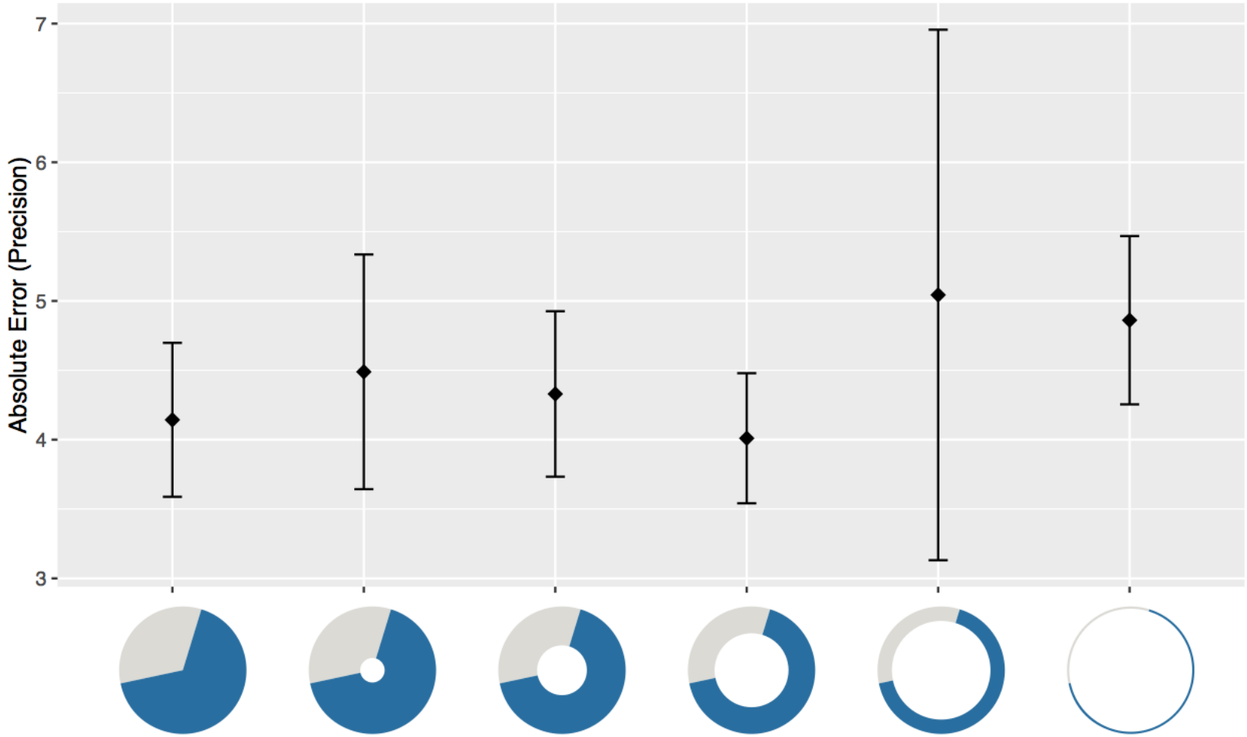


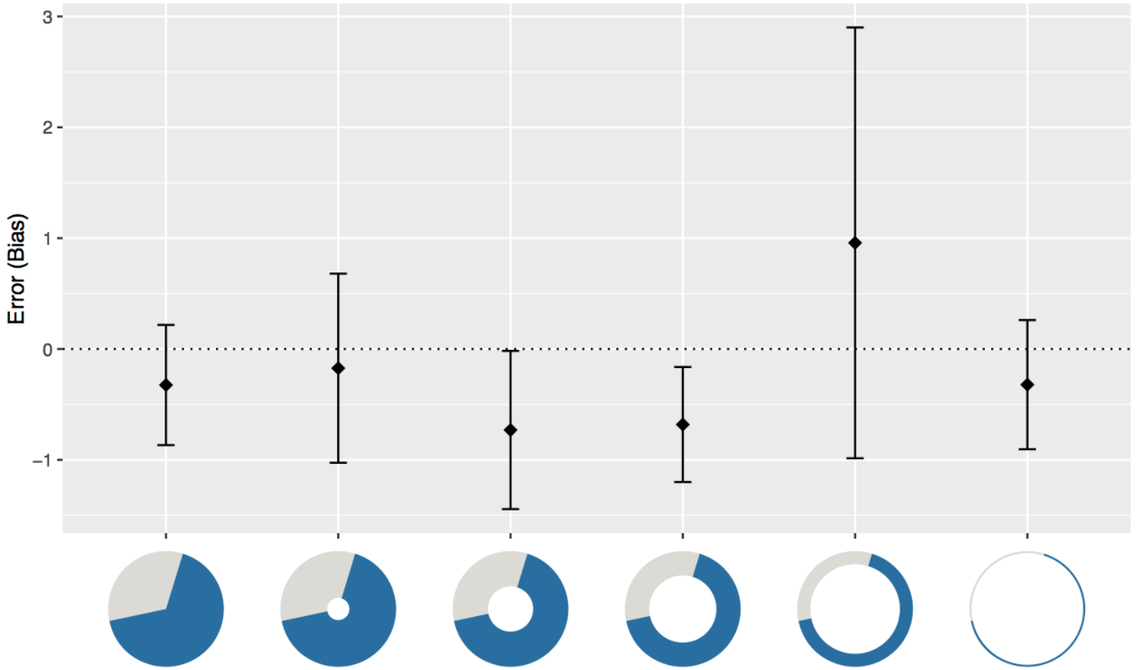
If the decomposed pie charts in the first study didn't convince you, this definitely should: the larger slice gets overestimated systematically. This is exactly what you'd expect if pie charts were read by arc length or area (since those are larger in comparison due to the larger radius), but not if you're in the angle camp. This is the smoking gun, right there!

overestimate

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

Donut chart Radii

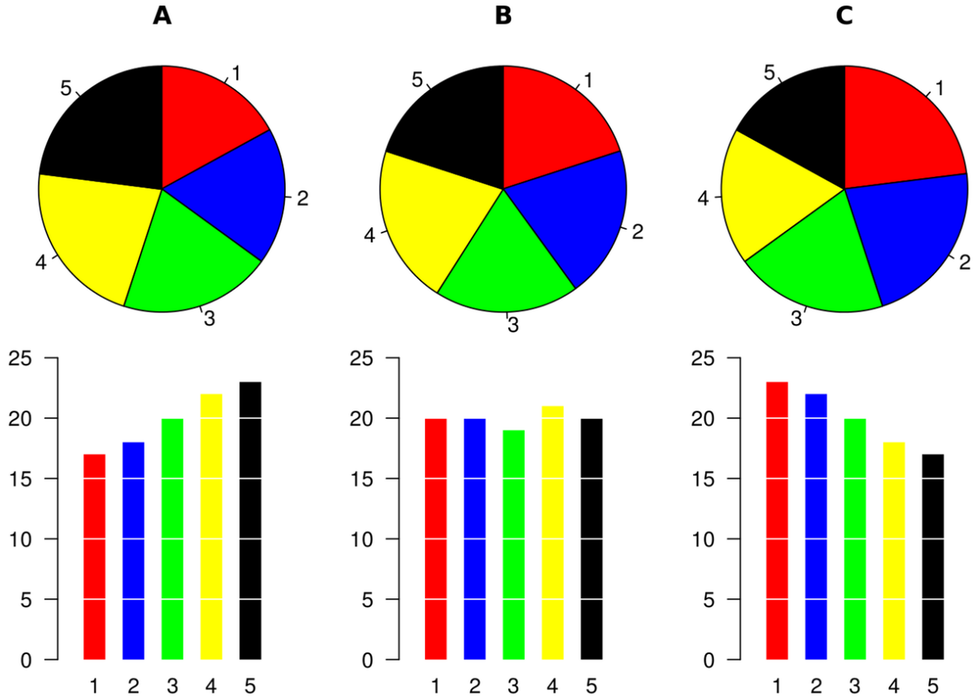


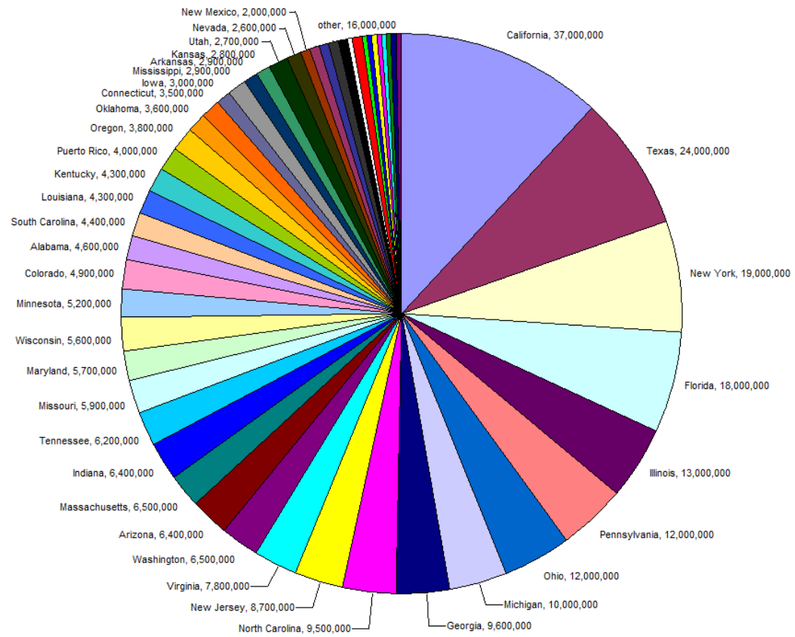


https://eagereyes.org/blog/2016/an-illustrated-tour-of-the-pie-chart-study-results

Pie charts: best practices

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



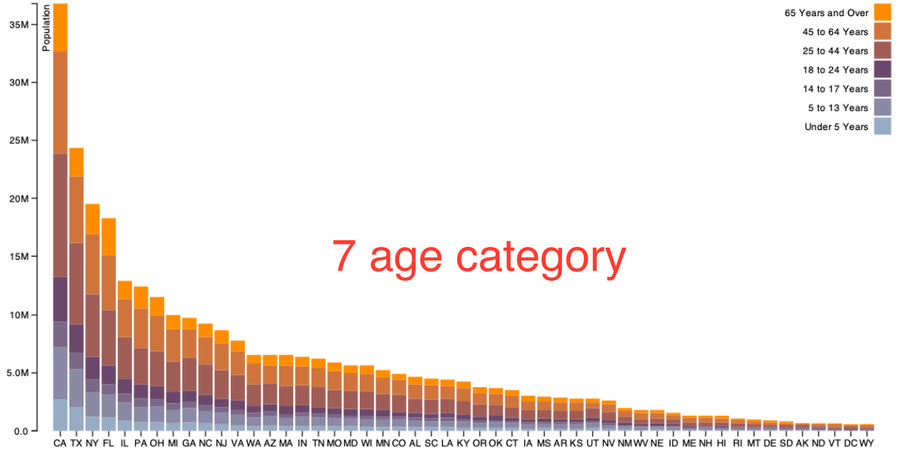


LARGE AMOUNT AND SIMLIAR
use pie chart with ordering

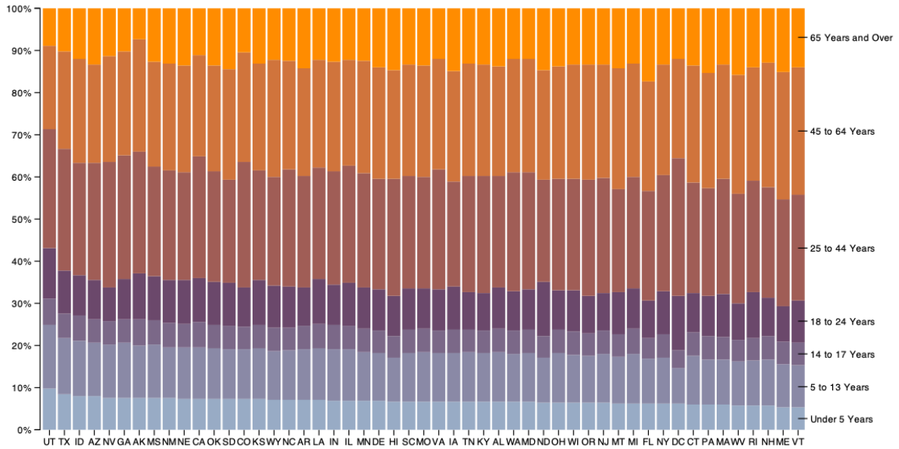
https://eagereyes.org/pie-charts

Idioms: **normalized stacked bar chart**

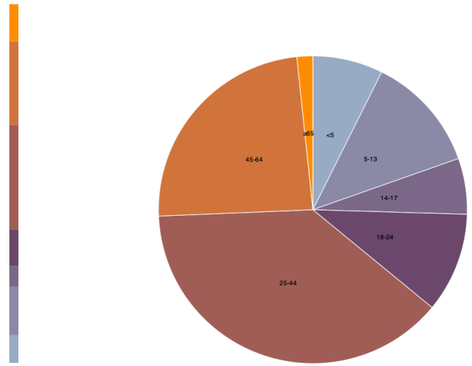
- task
  - part-to-whole judgements
- normalized stacked bar chart
  - stacked bar chart, normalized to full vert height
  - single stacked bar equivalent to full pie
    - high information density: requires narrow rectangle
- pie chart
  - information density: requires large circle



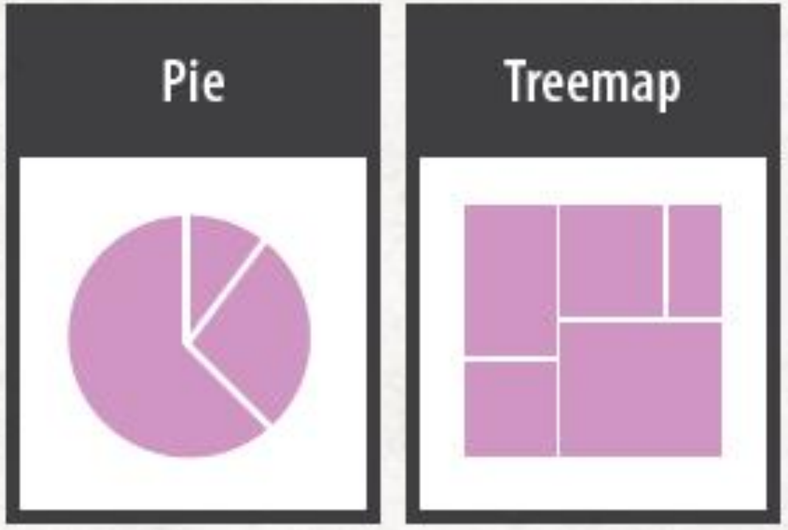
7 age category



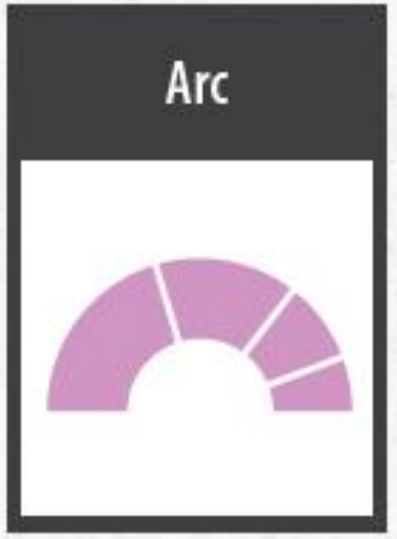
50 states



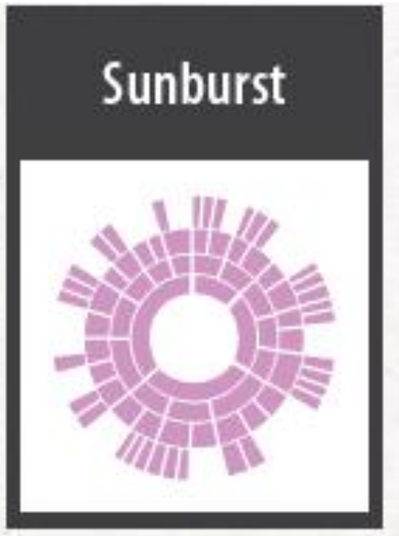
1 state

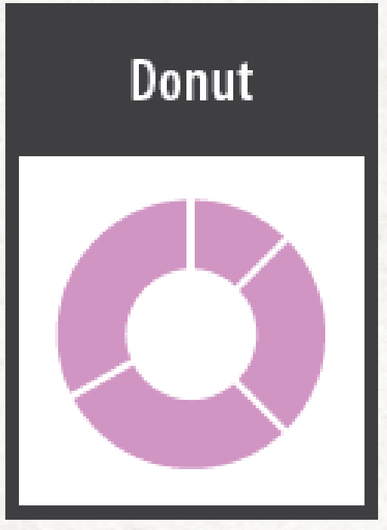

Pie

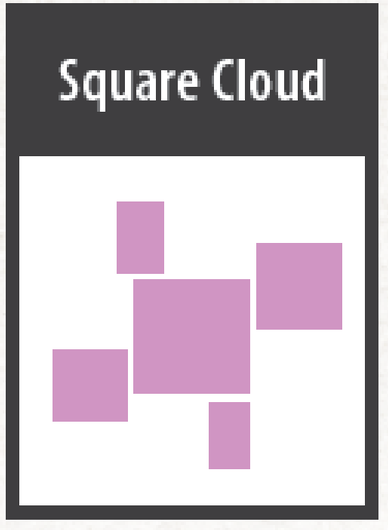
Treemap


Arc


Voronoi

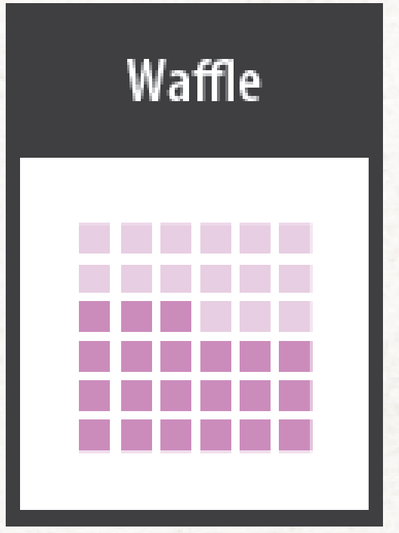

Sunburst


Donut


Square Cloud


Nightingale

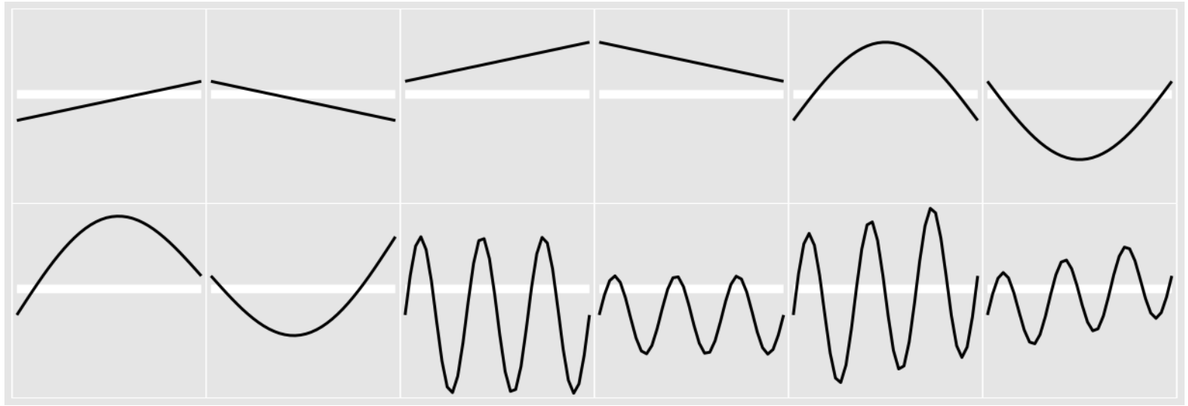

Triangle
Treemap

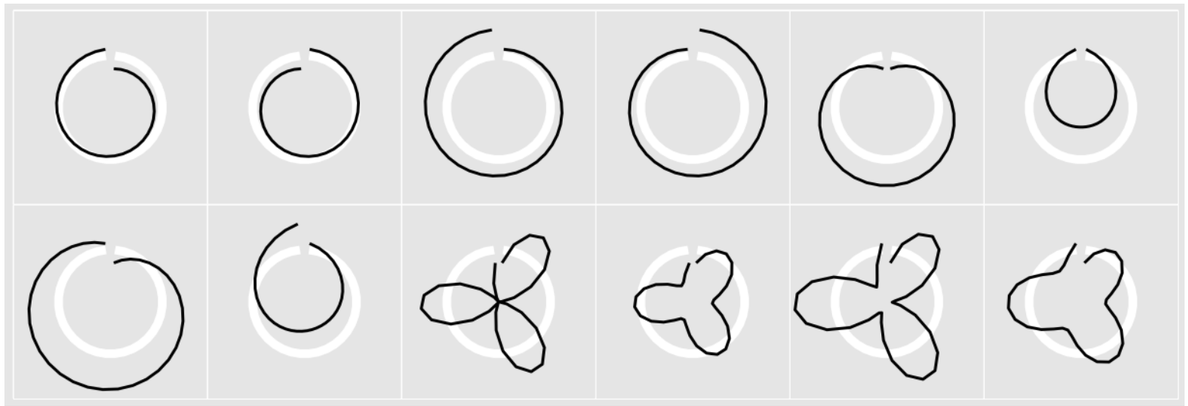

Waffle

Idiom: **glyphmaps**

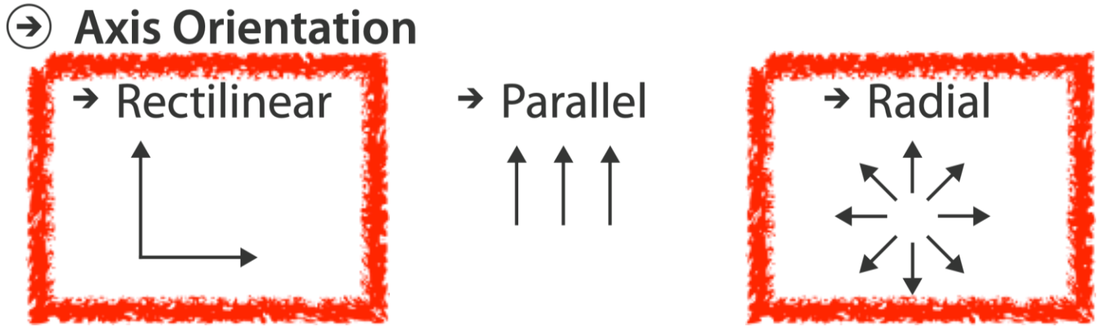
- rectilinear good for linear vs nonlinear trends

- 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.]



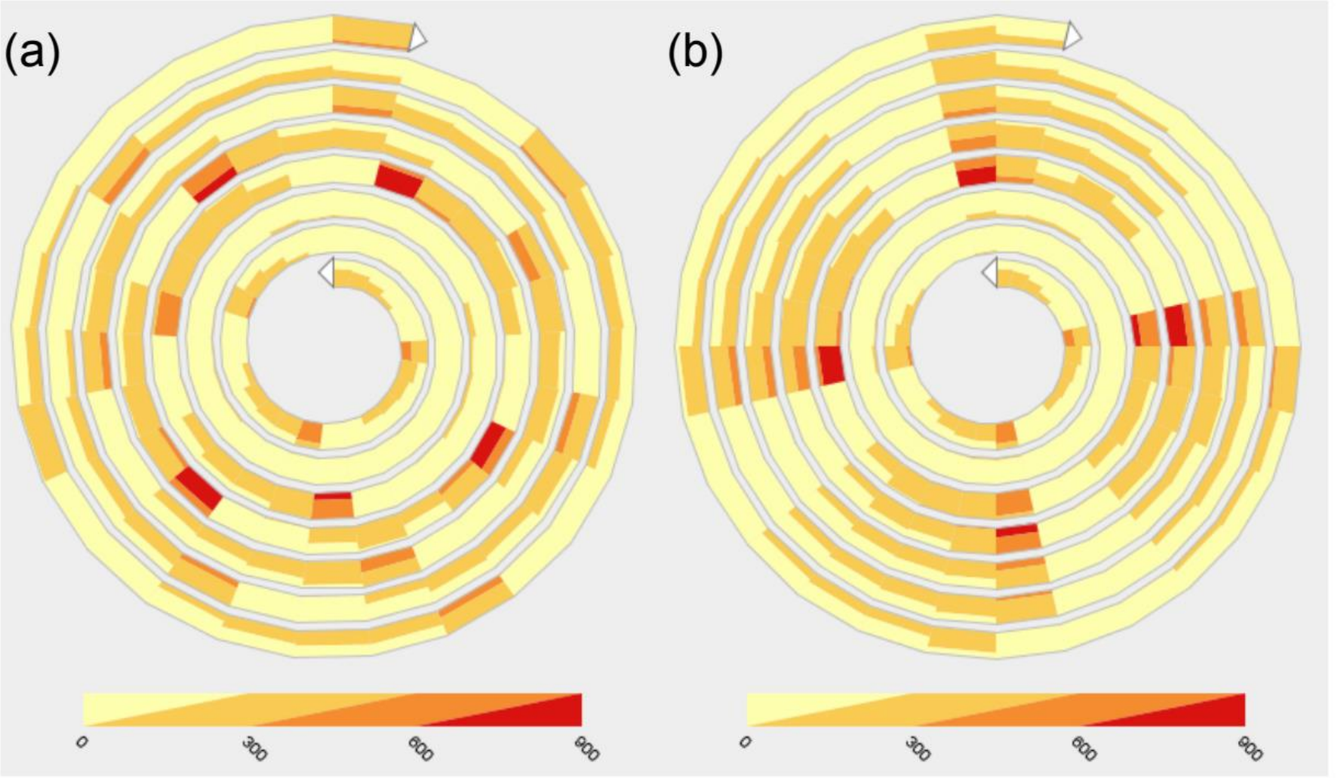
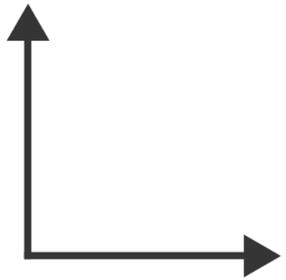


Figure 6: *Finding a pattern – (a) Cycle length = 25; (b) Cycle length = 28*

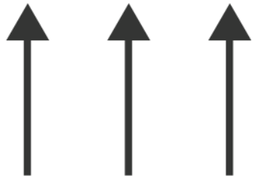
Tominski & Schumann https://ep.liu.se/ecp/034/013/ecp083413.pdf

Axis Orientation

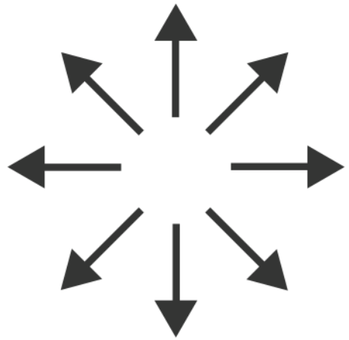
- Rectilinear



- Parallel



- Radial



Idiom: **SPLOM**

scatterplot matrix (SPLOM)

- rectilinear axes, point mark
- all possible pairs of axes
- scalability
  - one dozen attribs
  - dozens to hundreds of items

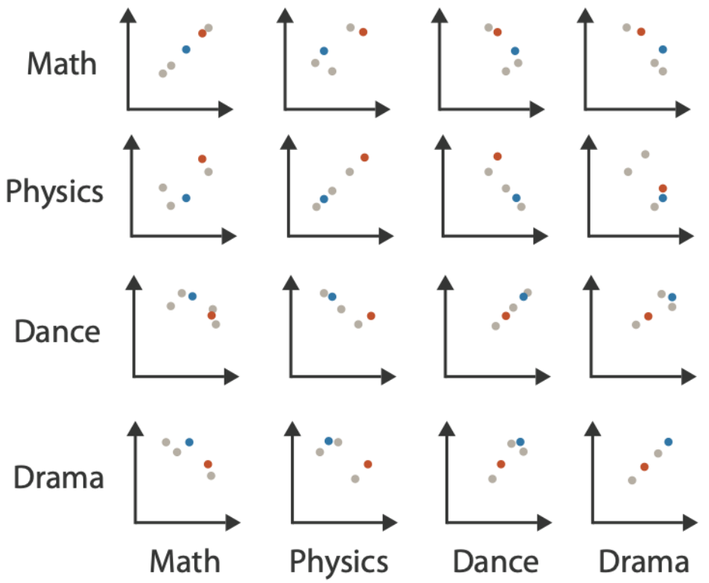
Wilkinson et al., 2005

number limit

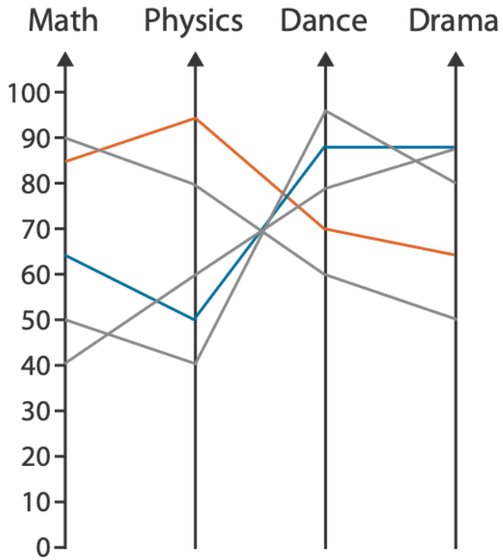
Idioms: **parallel coordinates**

- 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

Scatterplot Matrix



Parallel Coordinates



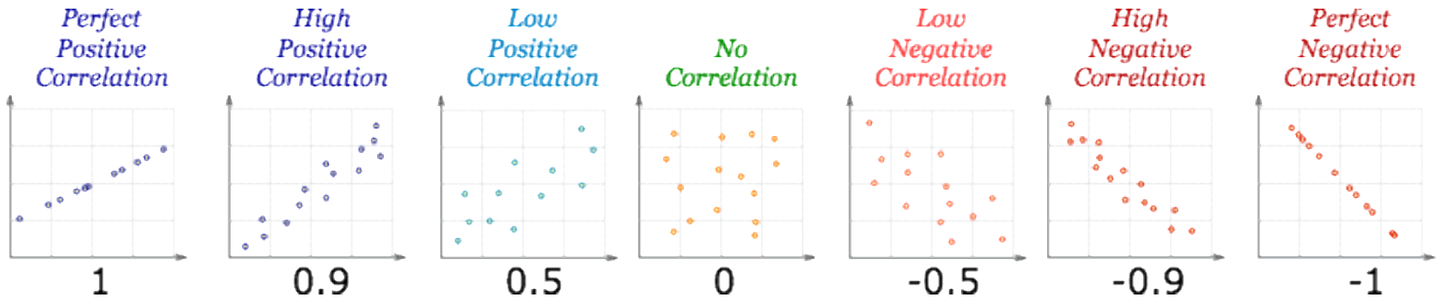
Table

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: spread out
- parallel coordinates
  - positive correlation
    - parallel line segments
  - negative correlation
    - all segments cross at halfway point
  - uncorrelated
    - scattered crossings



https://www.mathsisfun.com/data/scatter-xy-plots.html

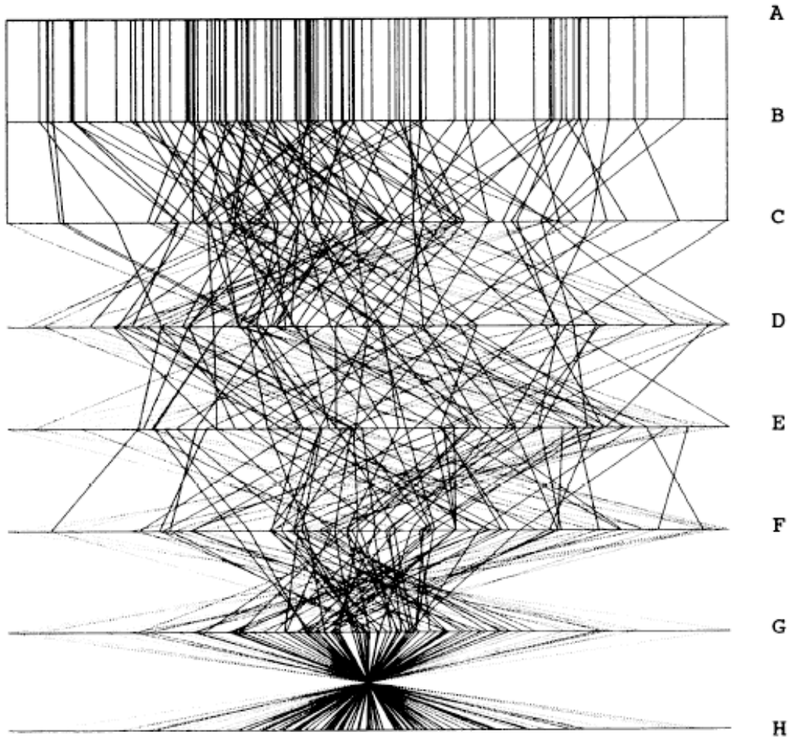
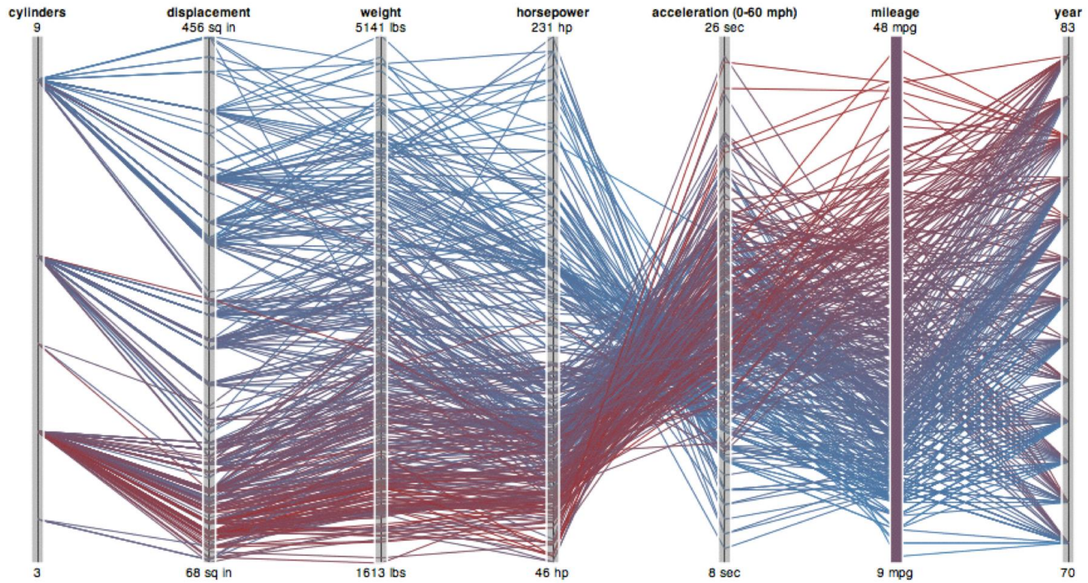


Figure 3. Parallel Coordinate Plot of Six-Dimensional Data Illustrating Correlations of $\rho = 1, .8, .2, 0, -.2, -.8,$ and -1.

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

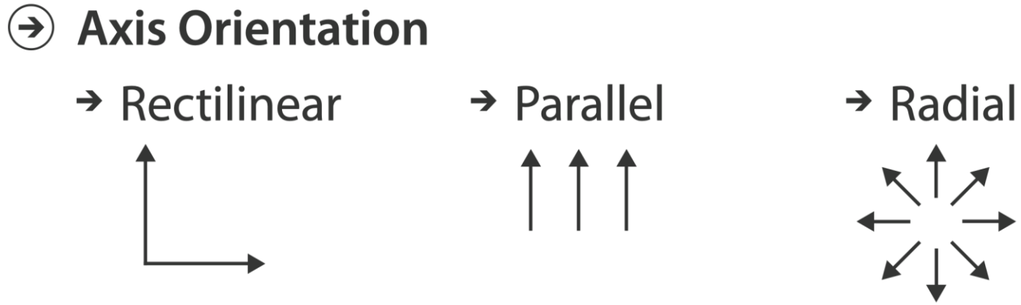
Parallel coordinates, limitations

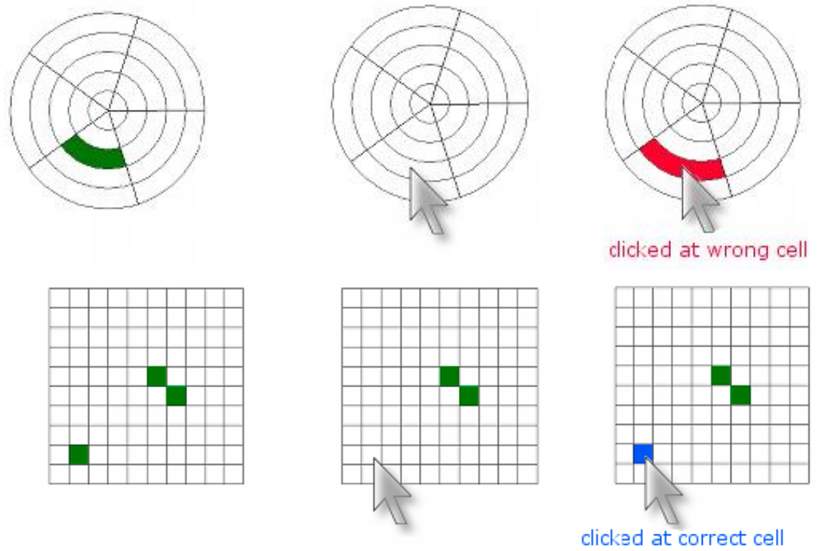
- visible patterns only between neighboring axis pairs
- how to pick axis order?
  - usual solution: reorderable axes, **interactive** exploration
  - same weakness as many other techniques
    - downside of interaction: human-powered search
  - some algorithms proposed, none fully solve



Orientation limitations

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





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

For example https://observablehq.com/@d3/zoomable-sunburst

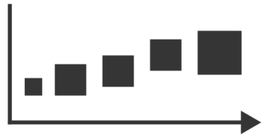
Arrange tables

Express Values

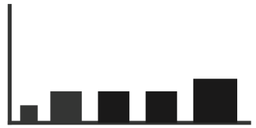
Separate, Order, Align Regions

Separate

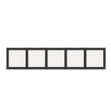
Order



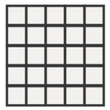
Align



1 Key
List

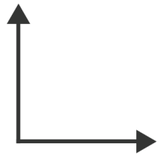


2 Keys
Matrix



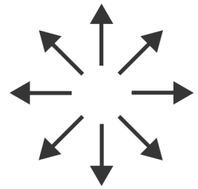
Axis Orientation

Rectilinear



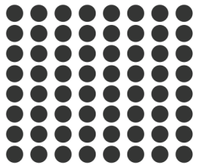
Parallel

Radial



Layout Density

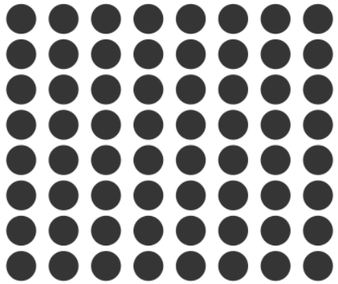
Dense



Layout density

Layout Density

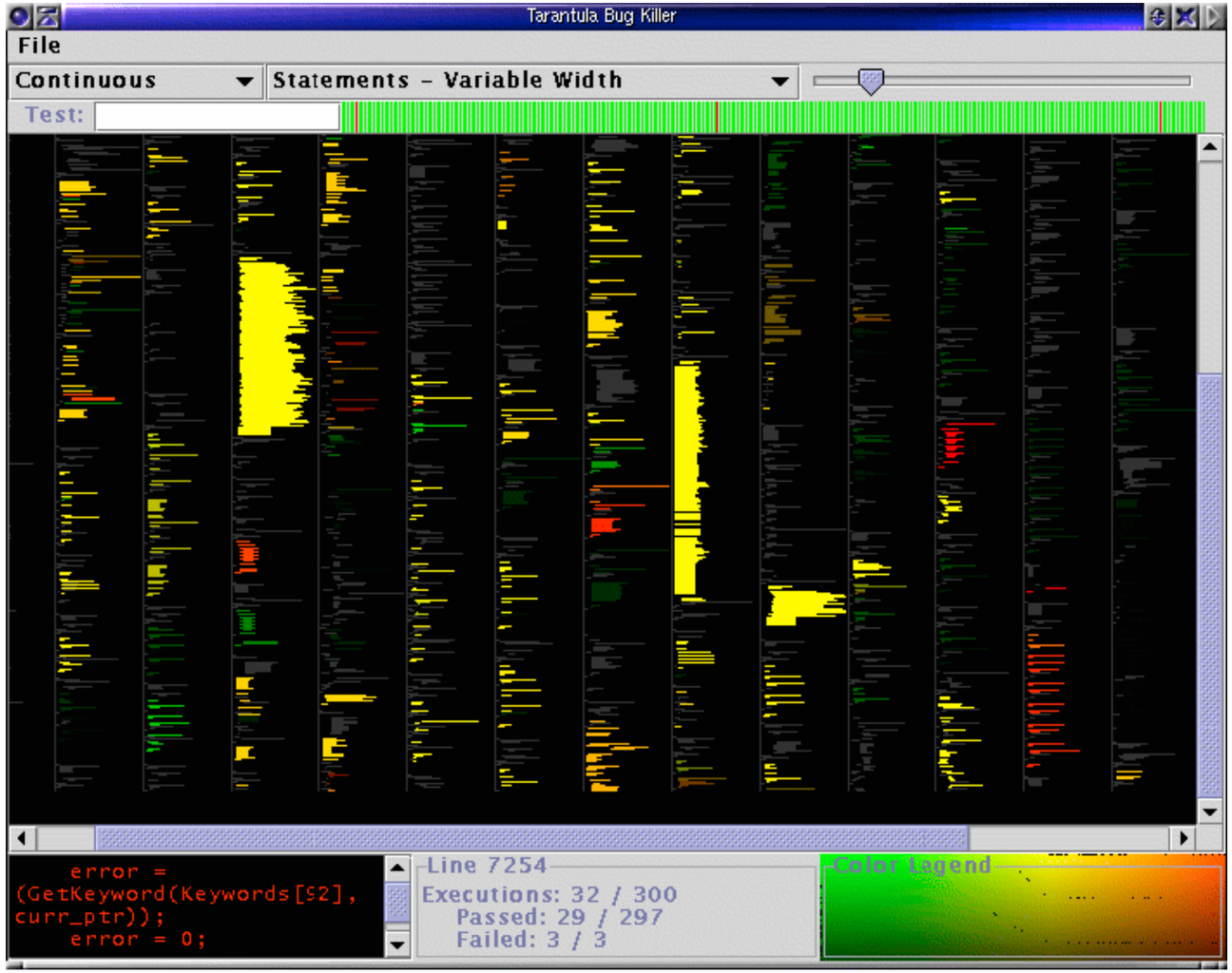
- Dense



- Space-Filling

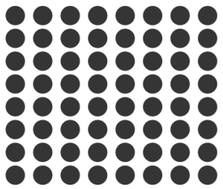
Idiom: Dense software overviews

- 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



➔ **Layout Density**

➔ Dense



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

How?

Encode

Arrange

- Express
- Separate
- Order
- Align
- Use

Map

from categorical and ordered attributes

- Color
  - *Hue*
  - *Saturation*
  - *Luminance*
- Size, Angle, Curvature, ...
- Shape
- Motion

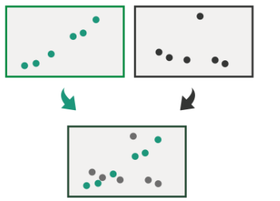
Direction, Rate, Frequency, ...

Manipulate

- Change
- Select
- Navigate

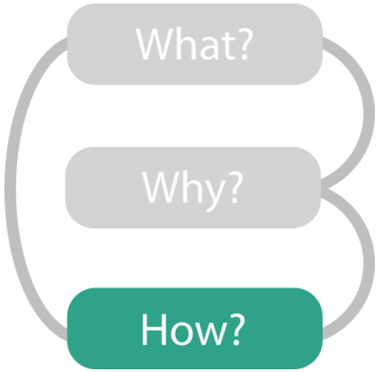
Facet

- Juxtapose
- Partition
- Superimpose



Reduce

- Filter
- Aggregate
- Embed



How?

Encode

Arrange

- Express
- Separate
- Order
- Align
- Use

Map

from categorical and ordered attributes

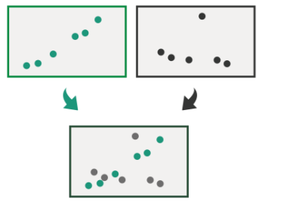
- Color
  - *Hue*
  - *Saturation*
  - *Luminance*
- Size, Angle, Curvature, ...
- Shape
- Motion
  Direction, Rate, Frequency, ...

Manipulate

- Change
- Select
- Navigate

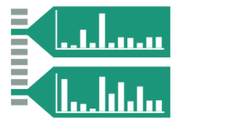
Facet

- Juxtapose
- Partition
- Superimpose



Reduce

- Filter
- Aggregate
- Embed



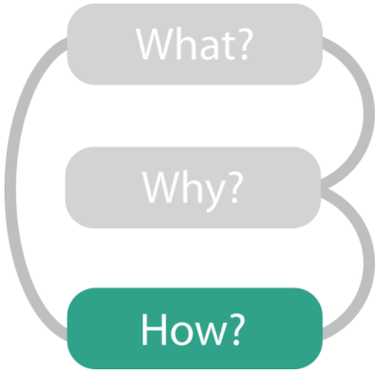
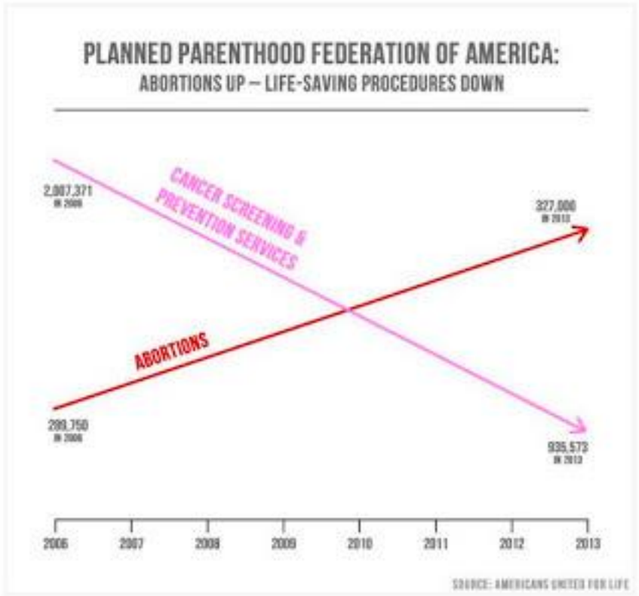
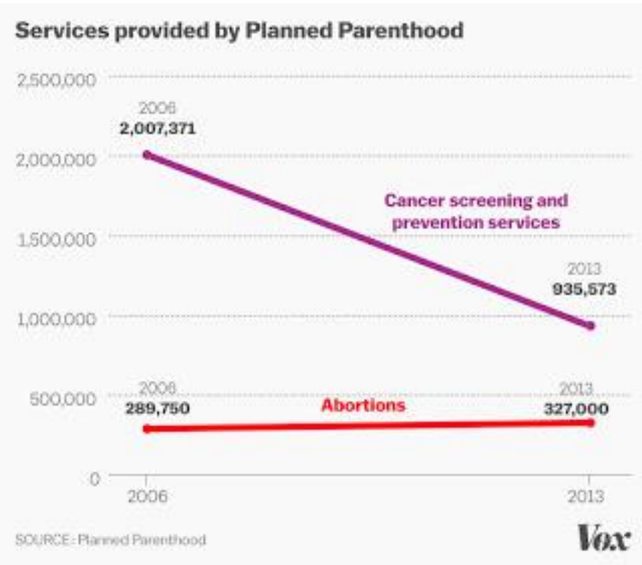
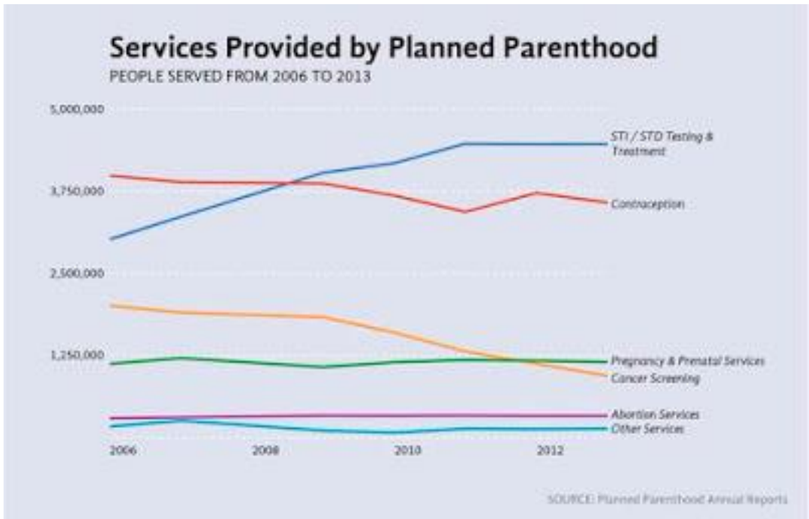


Chart axes

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



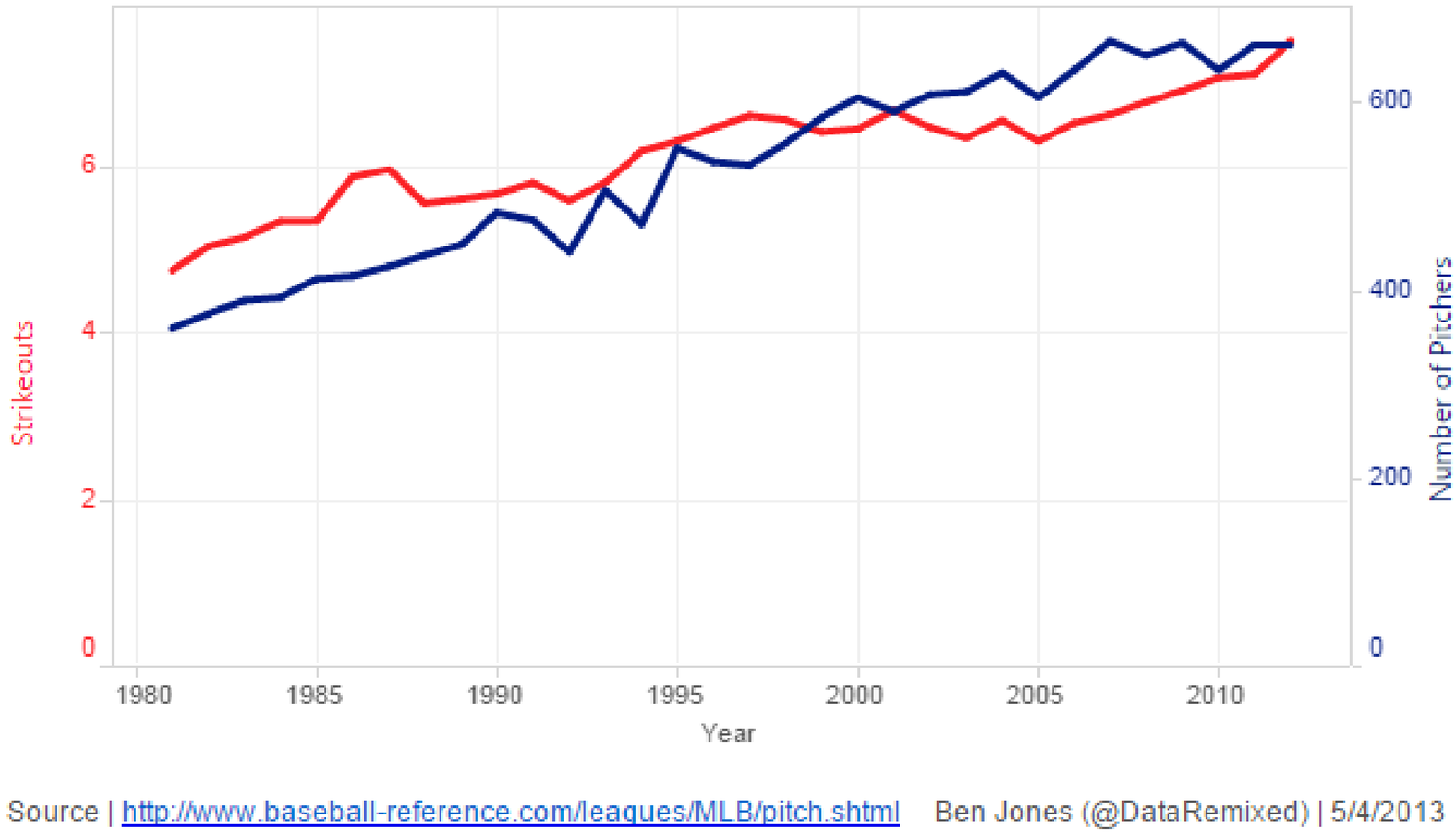




http://www.thefunctionalart.com/2015/10/if-you-see-bullshit-say-bullshit.htm

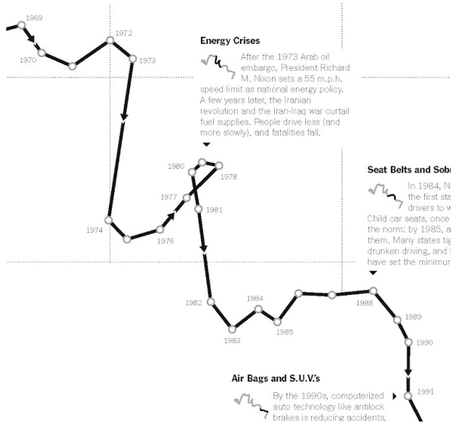
Idiom: **dual-axis line charts**

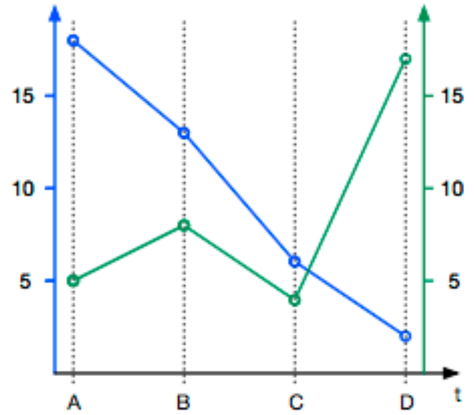
- controversial
  - acceptable if commensurate
  - beware, very easy to mislead!

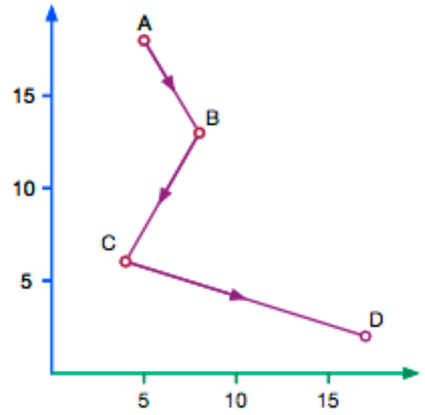


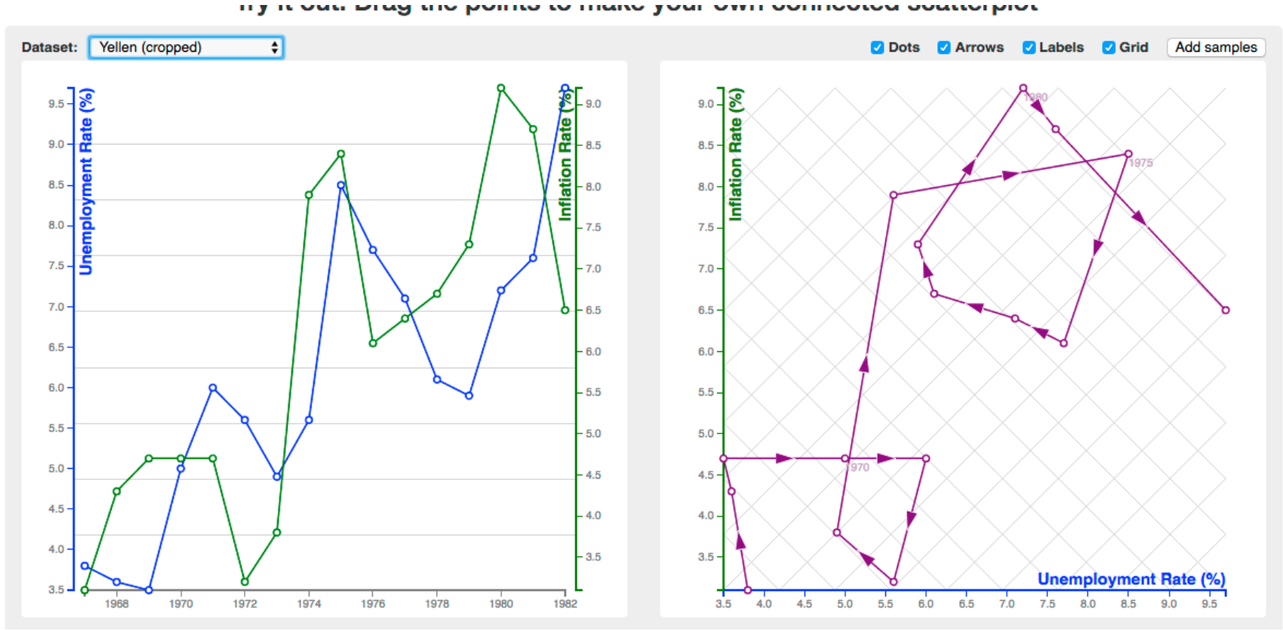
Idiom: **connected scatterplots**

- 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









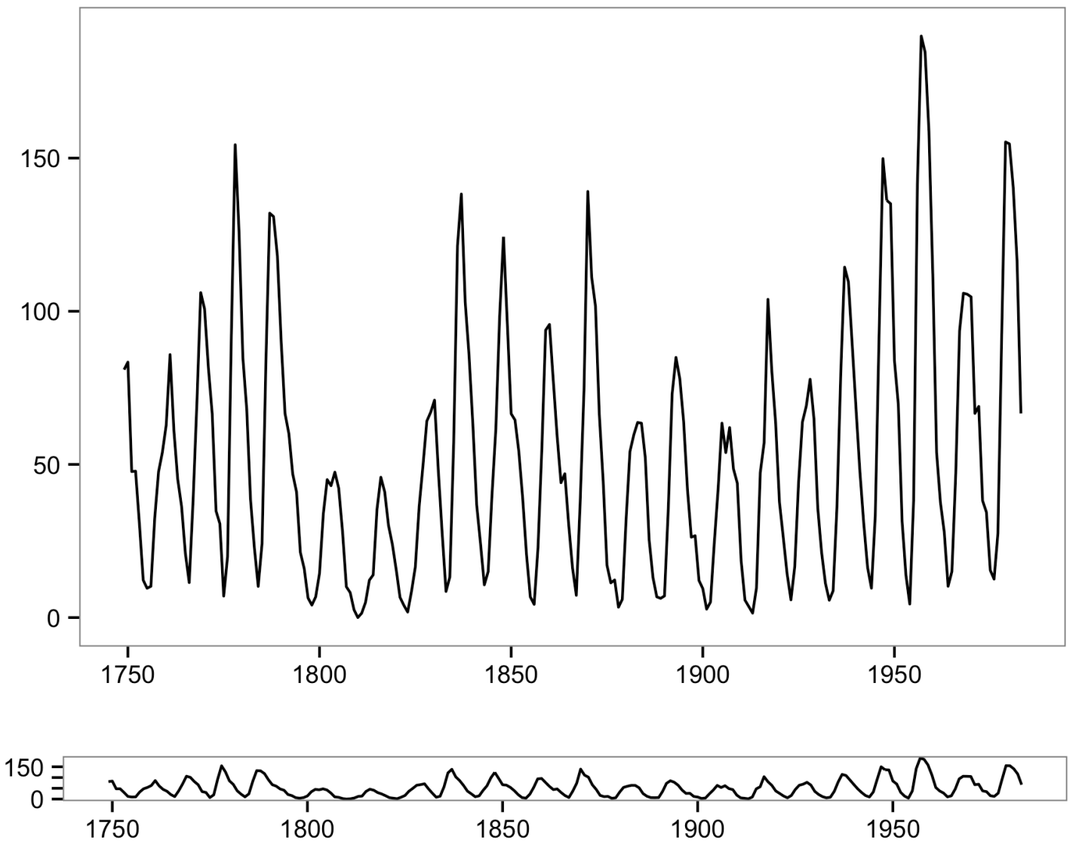
[The Connected Scatterplot for Presenting Paired Time Series. Haroz, Kosara and Franconeri. IEEE TVCG 22(9):2174-86, 2016.]

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

Choosing line chart aspect ratios

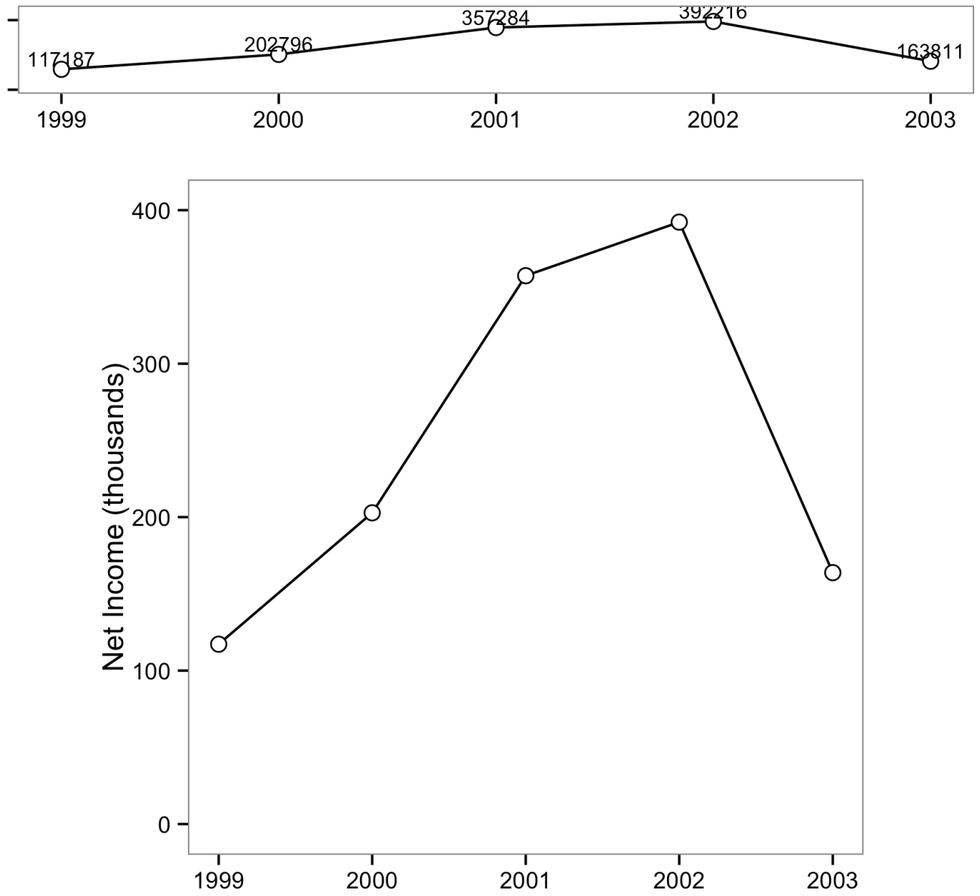
- 1: banking to 45 (1980s)
  - Cleveland perceptual argument: most accurate angle judgement at 45

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

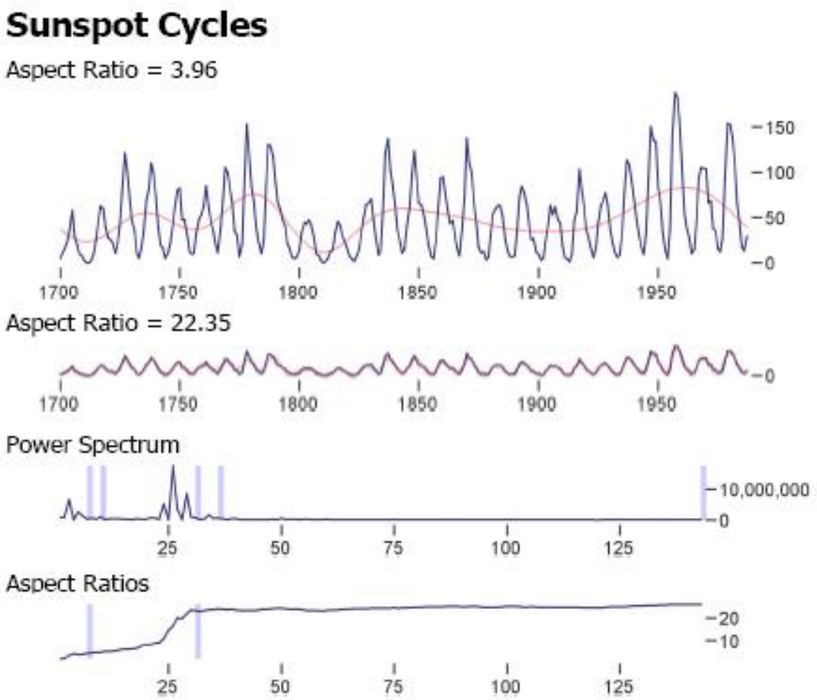
Fig 7.2 Annual Report: Aspect Ratio 2



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

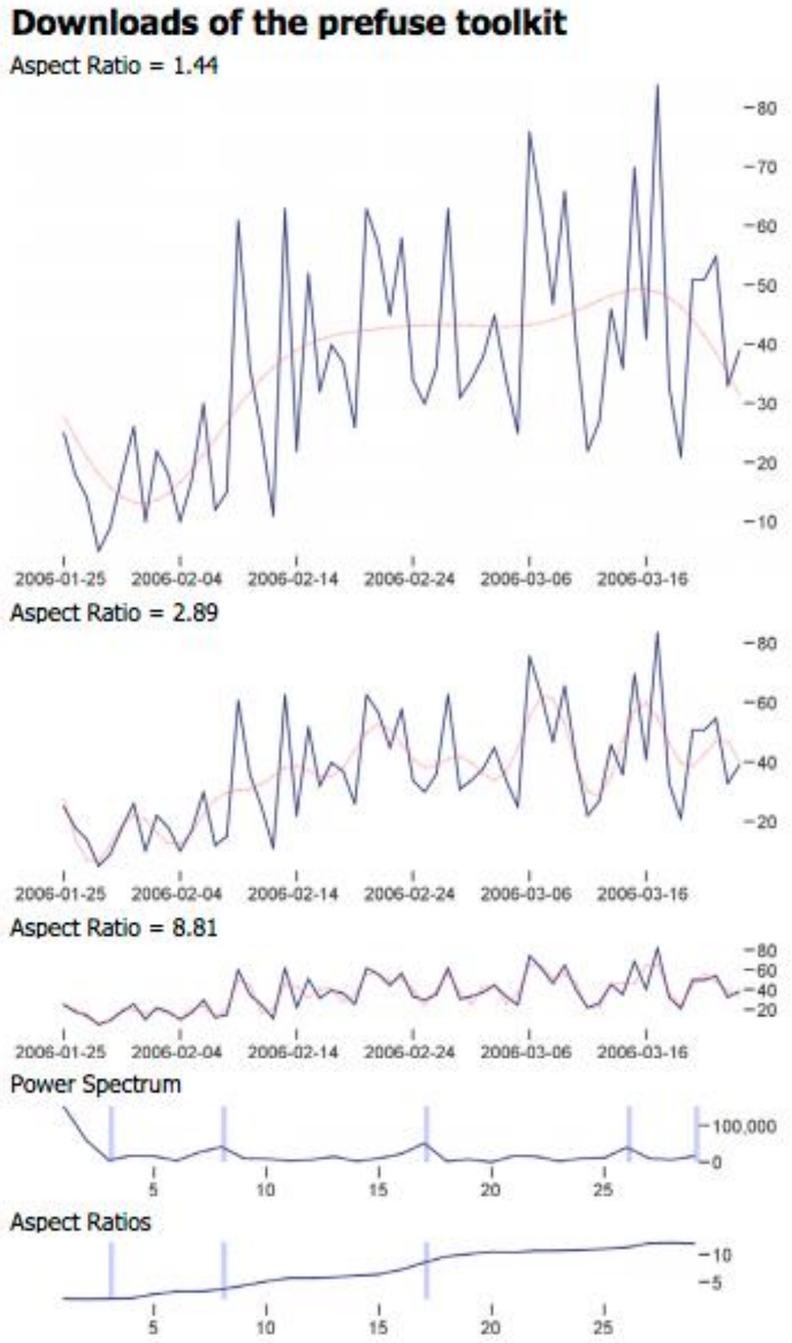


[Multi-Scale Banking to 45 Degrees. Heer and Agrawala, Proc InfoVis 2006]

overall

weekly

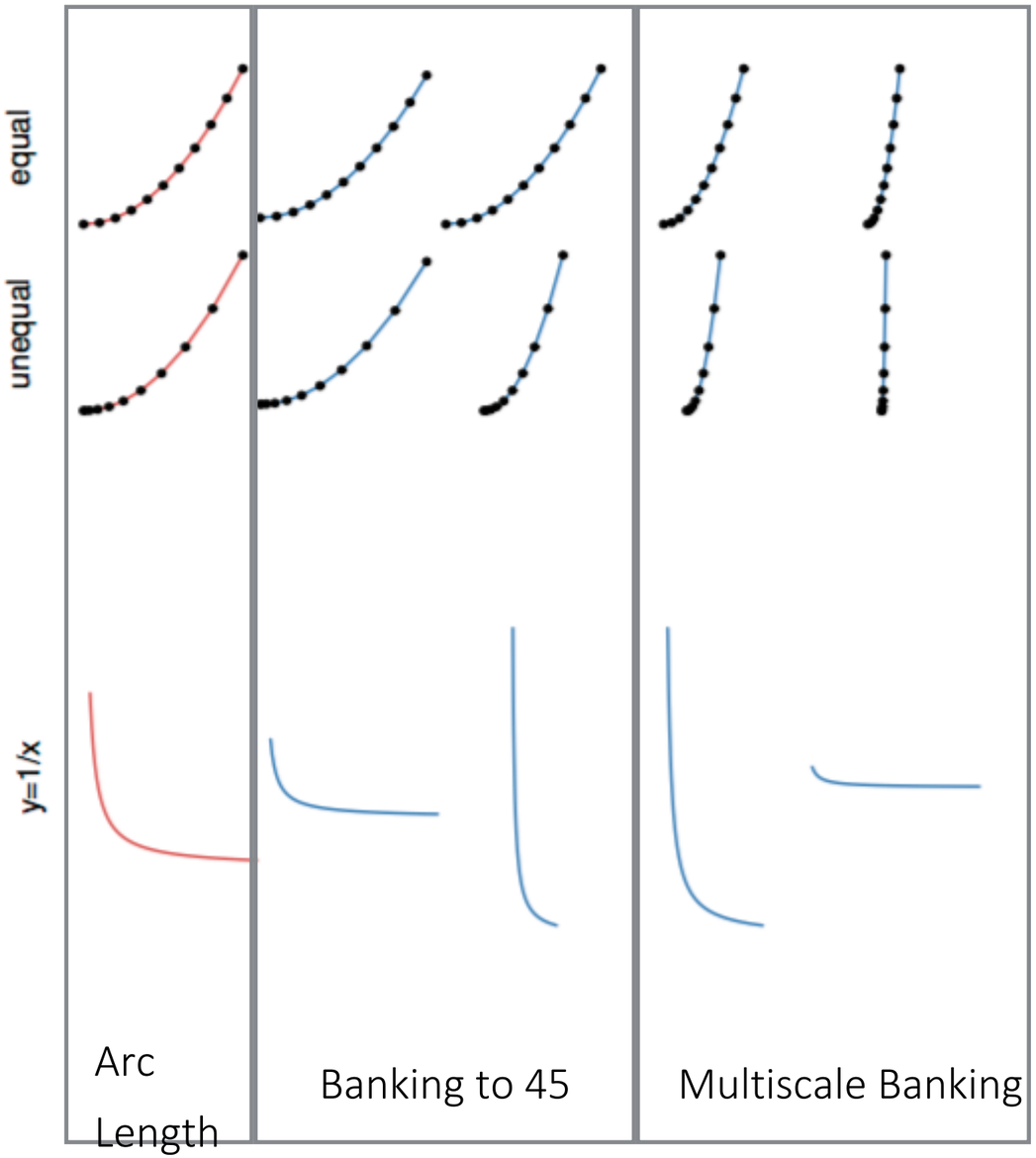
daily



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

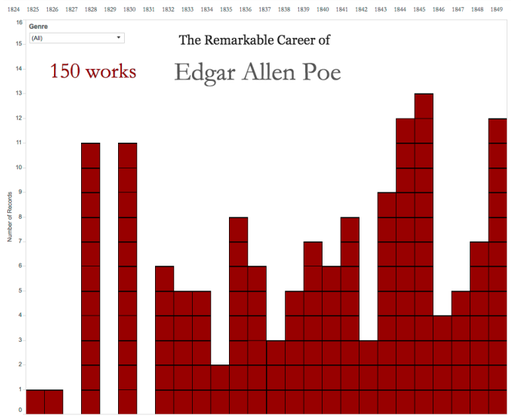
- meta-points from this progression
  - young field; prescriptive advice changes rapidly
  - reasonable defaults required deep dive into perception meets math



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

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

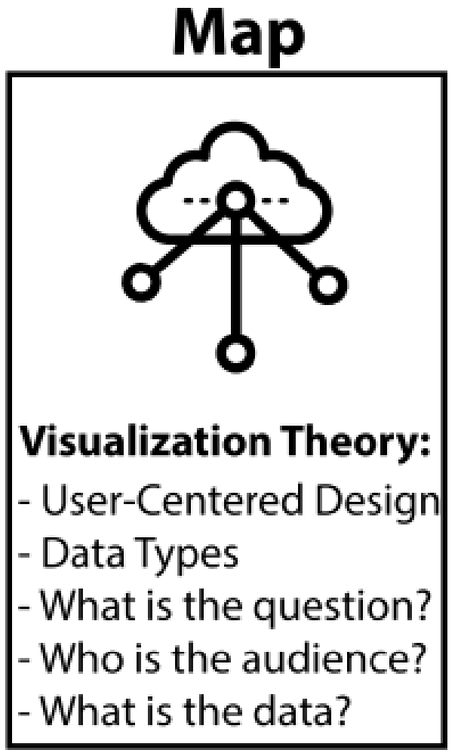
[Slide inspired by Ben Jones]

Data and Task Abstraction Group Exercise – 20 minutes

- Dataset
- https://www.thesquirrelcensus.com/data

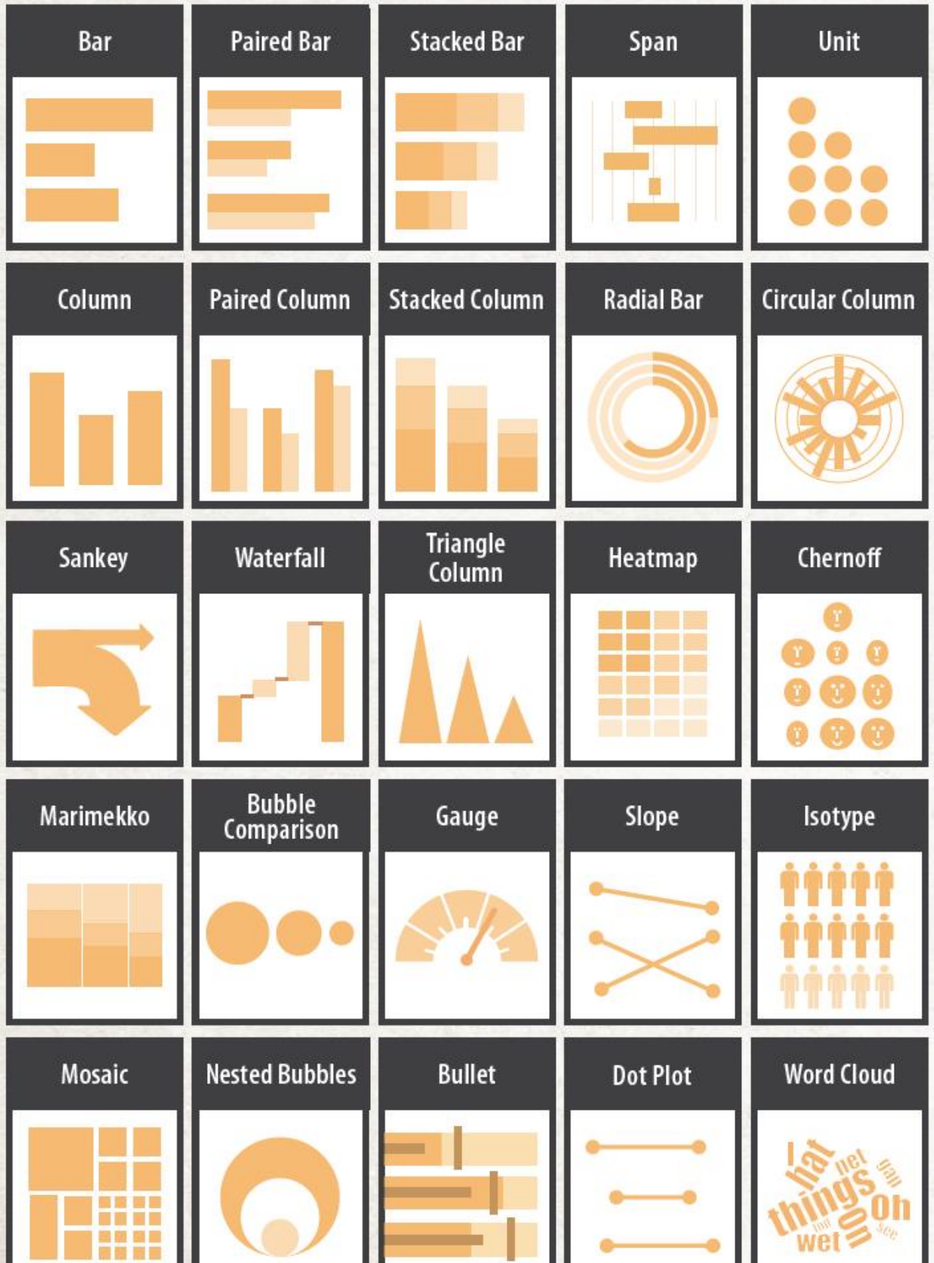
Squirrel Design Sprint (20 minutes)

- Group size 3 – 4
- Choose a CEO (aka the Decider – has the final say on decisions)
- Choose a Manager (aka the Facilitator – responsible for keeping track of time and organizing the sprint process)
- Main Task
  - Using the provided dataset identify
  - Who the audience is
  - What questions you want to answer (identify tasks that users wish to perform)
  - Make sure that there is variability in your tasks, use Lec09 to make your questions as specific as possible
- Submit to Gradescope https://help.gradescope.com/article/m5qz2xsnjy-student-add-group-members
  - Your submission must include who the CEO is, who the Manager is and a grouping of your questions based on the Task Abstraction Model (Analyze, Search, Query)
  - Also post to jamboard (see EdStem for link) don't add names



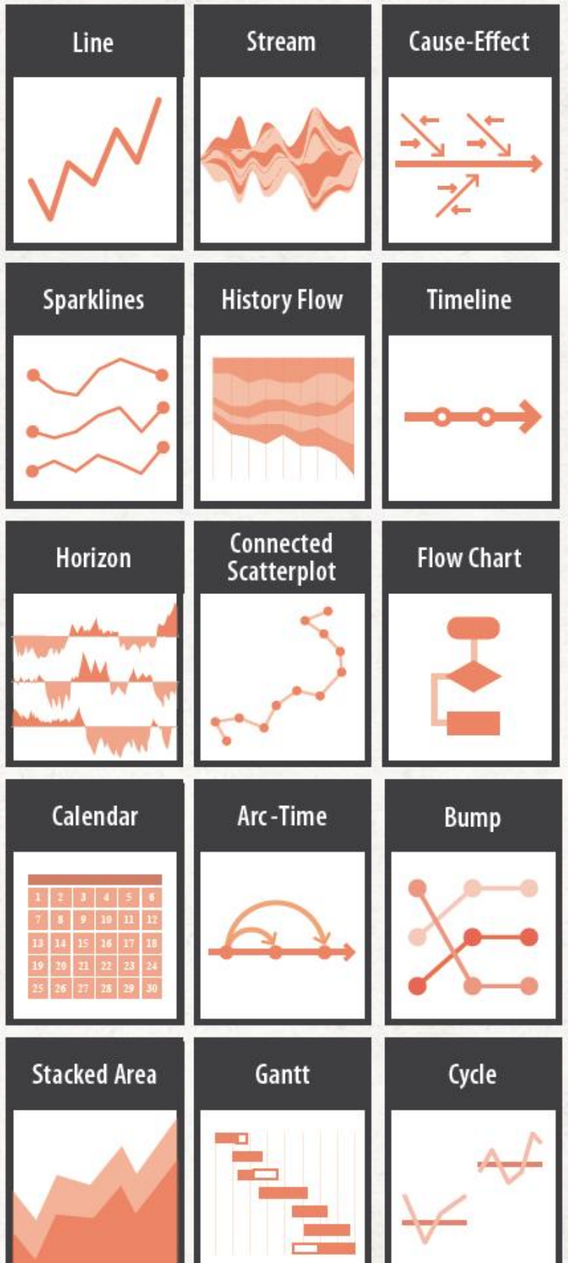
COMPARING CATEGORIES

Compare values across categories



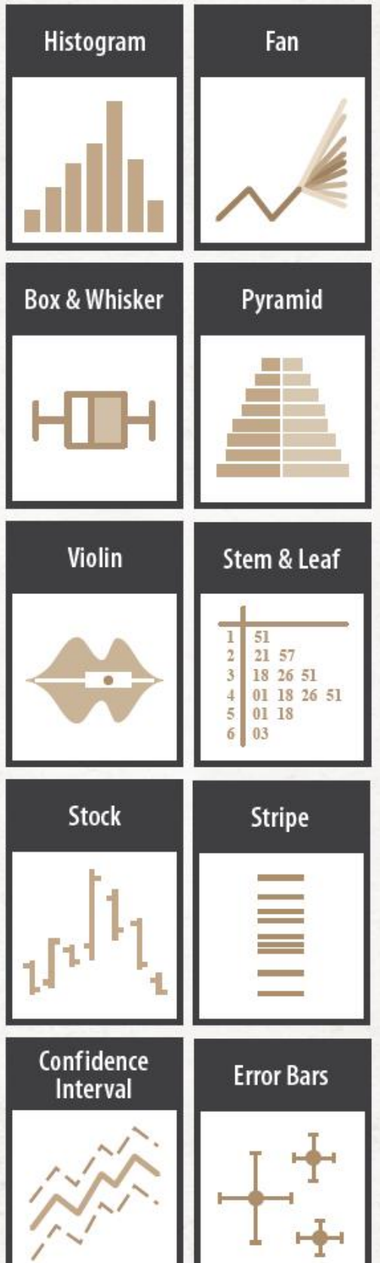
TIME

Track changes over time



DISTRIBUTION

Representation of the distribution of data



GEOSPATIAL

Relates data to its geography

- Map
- Flow Map
- Icon Map
- Choropleth
- Map with Columns
- Isopleth
- Cartogram
- Map with Pie Charts
- Contour
- Non-Contiguous Cartogram
- Bubble Map
- Dorling Map
- Connection Map
- Point Map
- Subway Map

PART-TO-WHOLE

Relates the part of a variable to its total

- Pie
- Treemap
- Donut
- Square Cloud
- Arc
- Voronoi
- Nightingale
- Triangle Treemap
- Sunburst
- Waffle

RELATIONSHIP

Illustrates correlations or relationships between variables

- Scatterplot
- Arc-Connection
- Dendrogram
- Word Tree
- Circle Packing
- Chord
- Tree
- Correlation Matrix
- Bubble
- Hive
- Double Tree
- Radar
- Parallel Coordinates
- Force-Directed
- Network
- Venn Diagram

@jschwabish @SR_Visual_Info