

Analisi e Visualizzazione di Reti Complesse

DV04 - Tabular Data

Prof. Rossano Schifanella

DV04 - Tabular data | Plotting



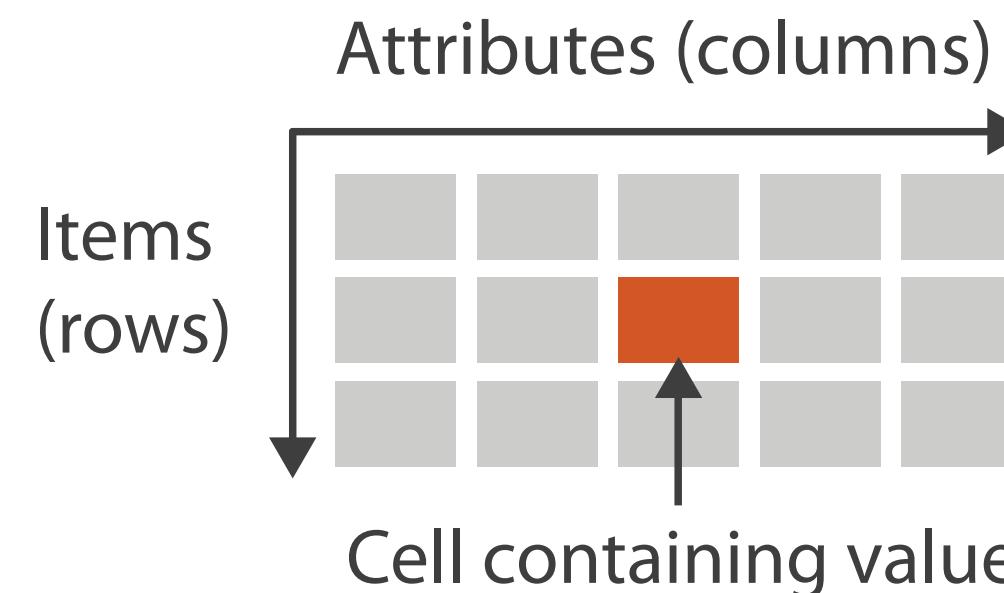
Tables

Focus on Tables

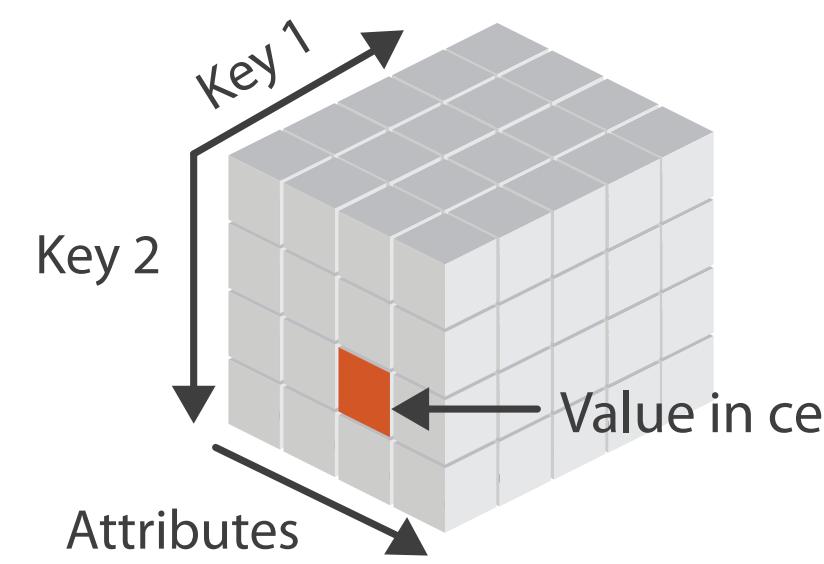


Dataset Types

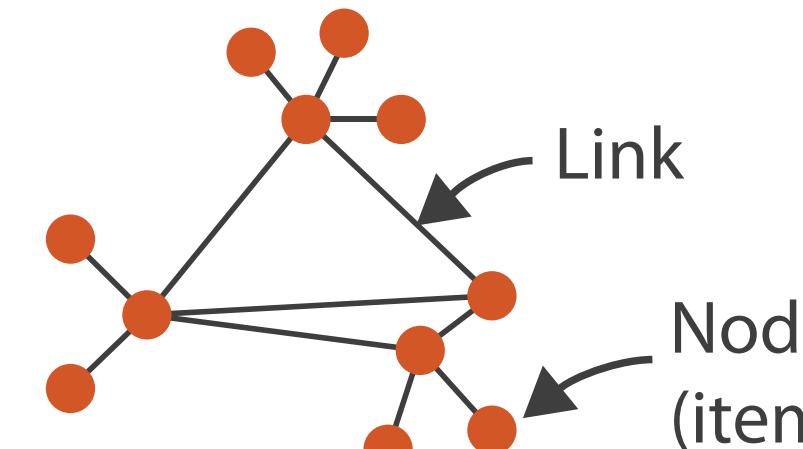
→ Tables



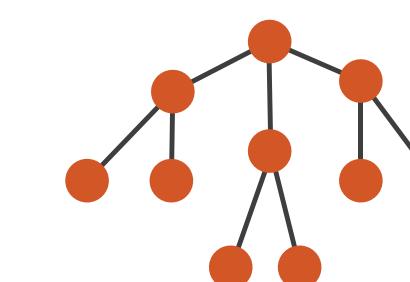
→ Multidimensional Table



→ Networks

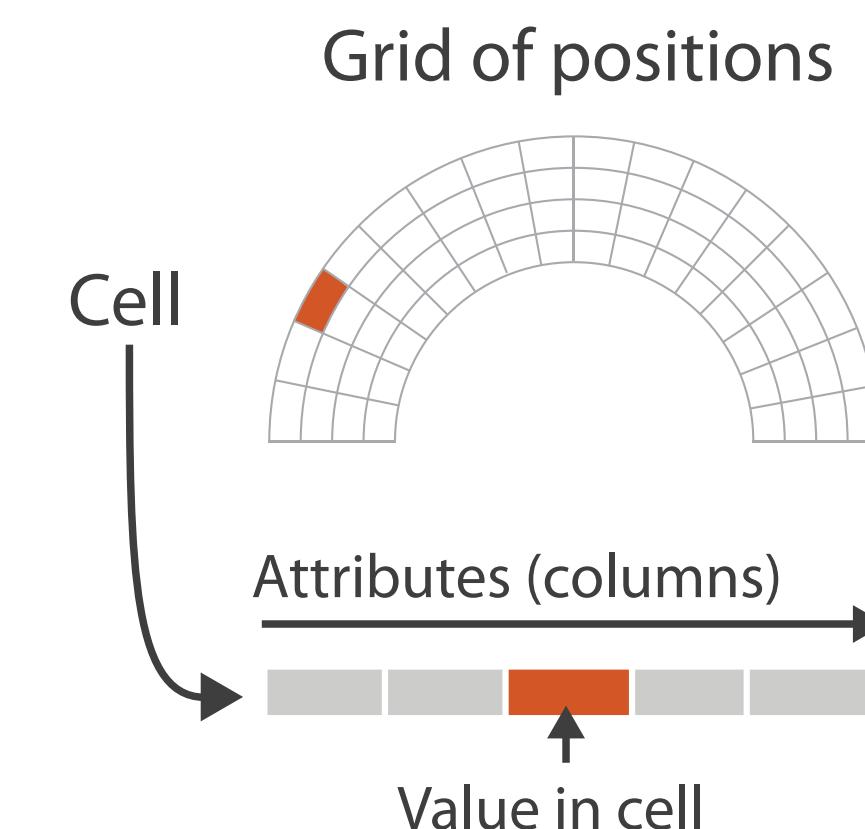


→ Trees

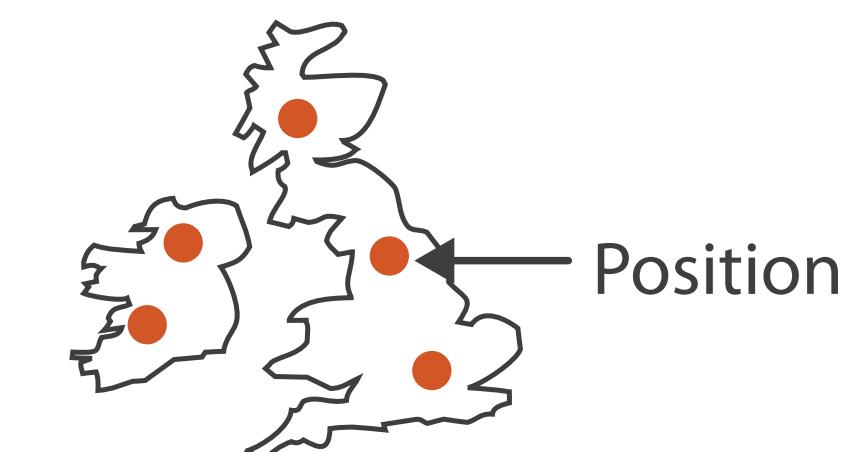


→ Spatial

→ Fields (Continuous)



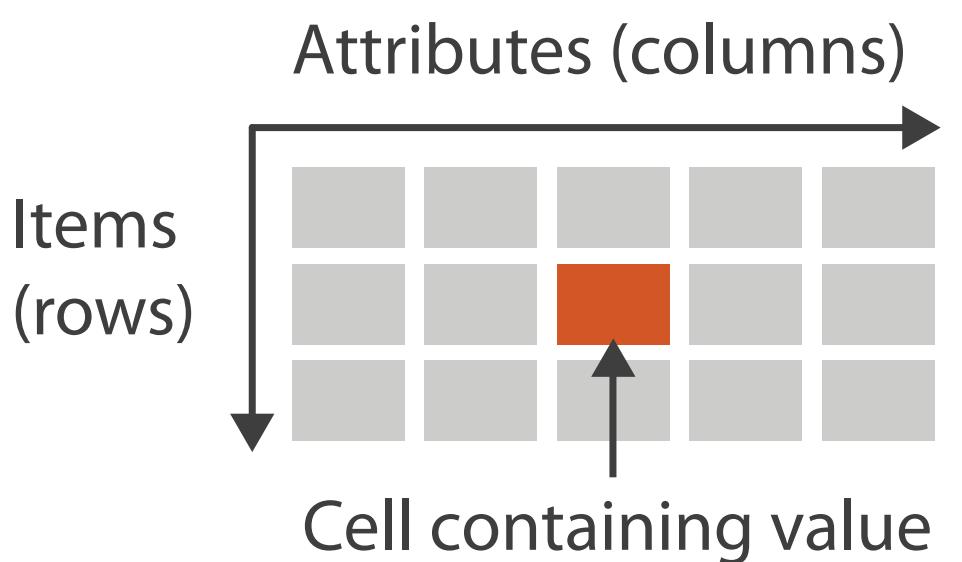
→ Geometry (Spatial)



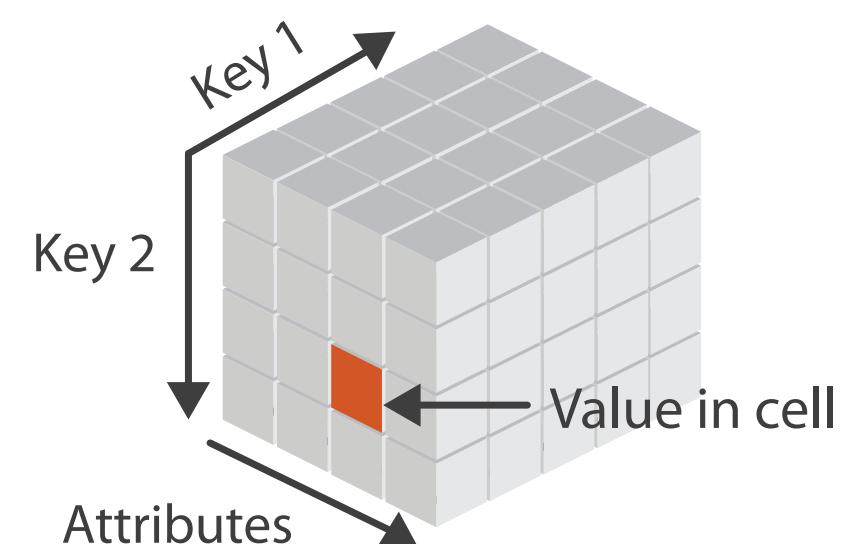
Keys and values

- **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

→ Tables



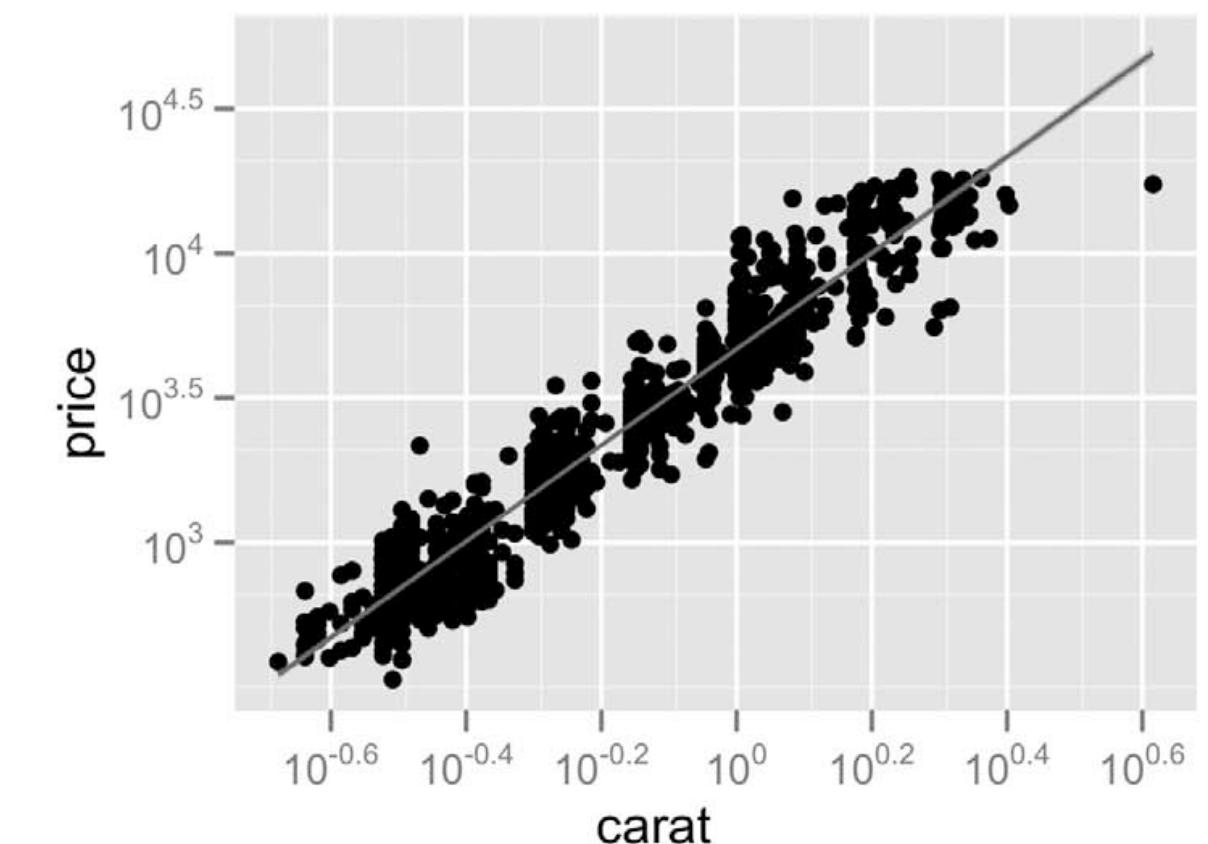
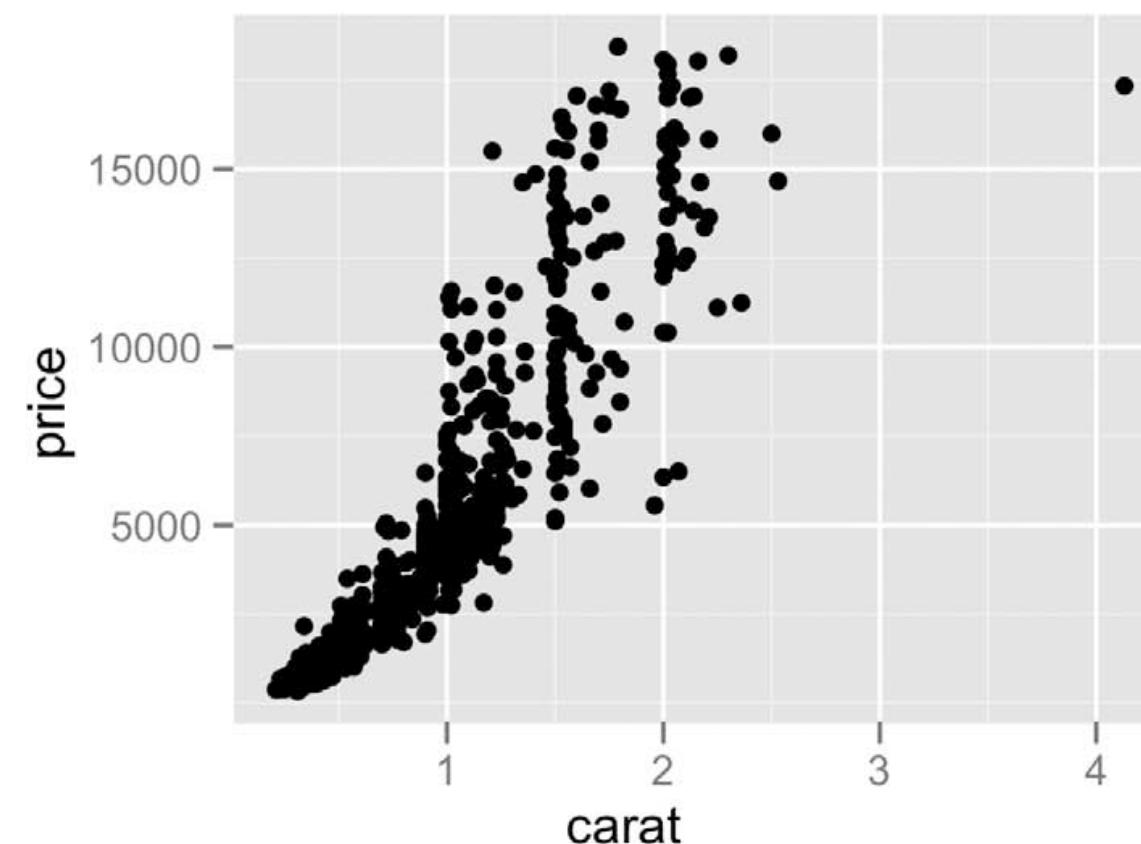
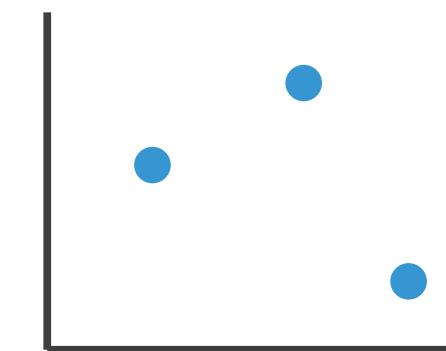
→ *Multidimensional Table*



Idiom: scatterplot

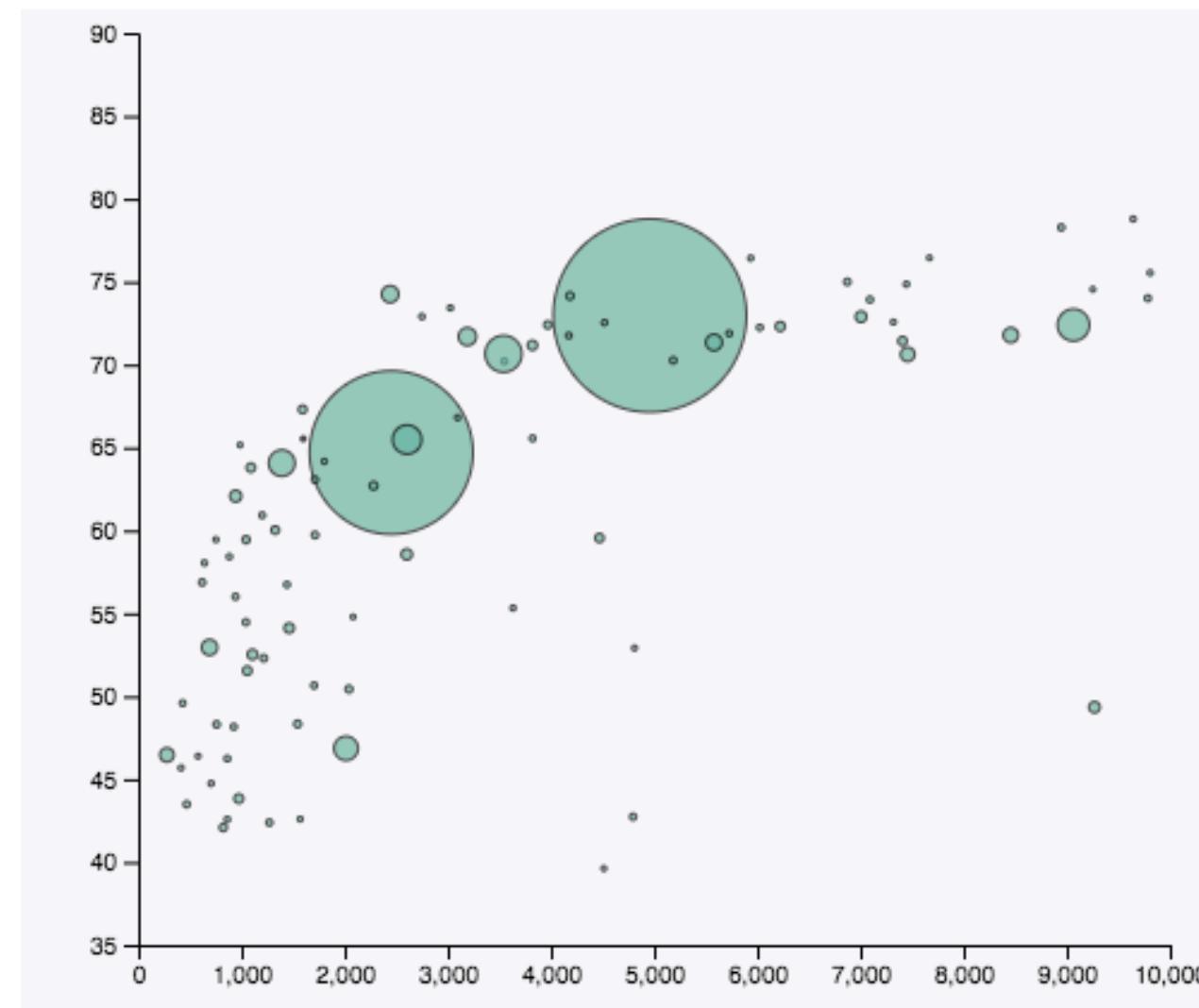
- express values (magnitudes)
 - quantitative attributes
- no keys, only values
 - data
 - 2 quant attrs
 - mark: points
 - channels
 - horiz + vert position
 - tasks
 - find trends, outliers, distribution, correlation, clusters
 - scalability
 - hundreds of items

→ Express Values

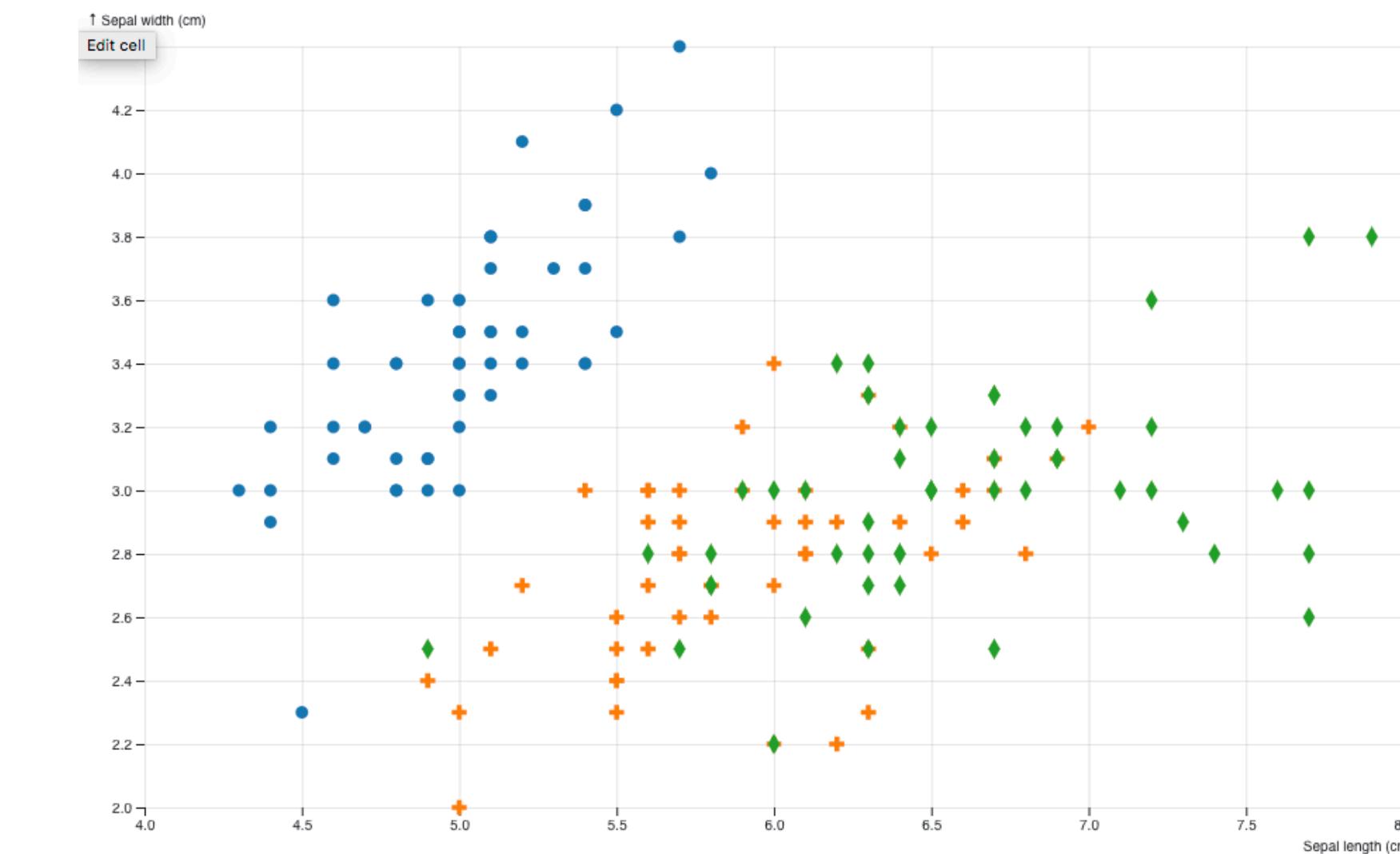


Scatterplots: Encoding more channels

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

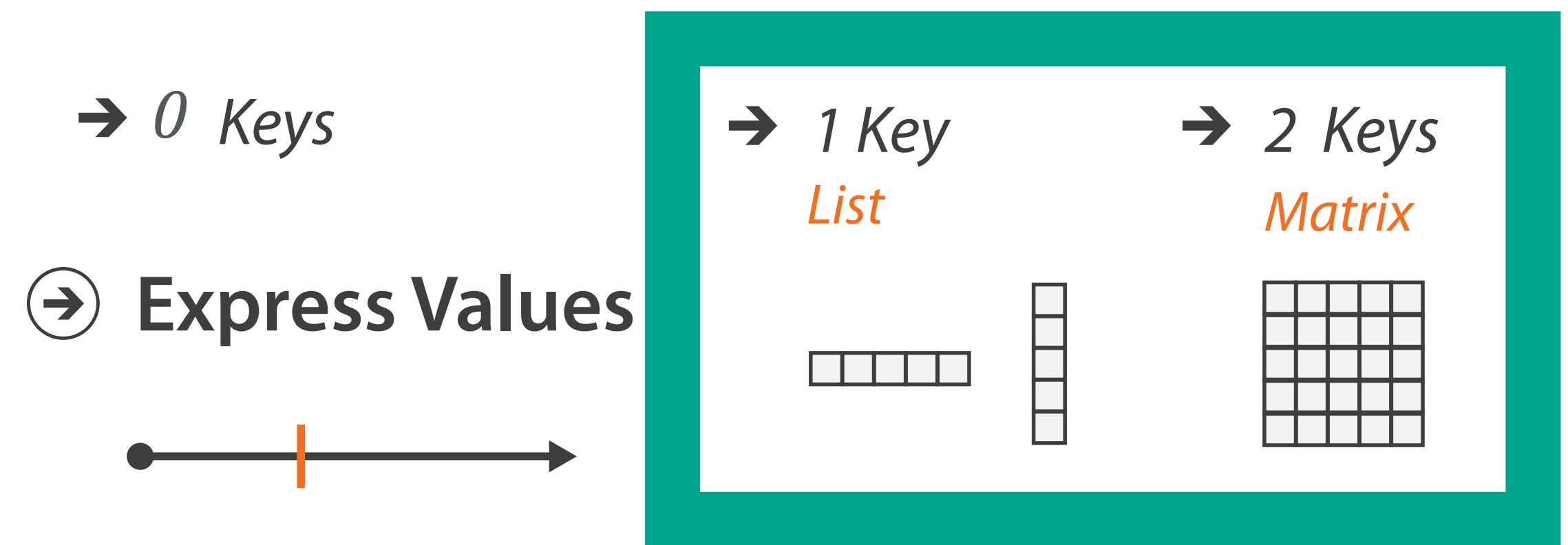


https://www.d3-graph-gallery.com/graph/bubble_basic.html



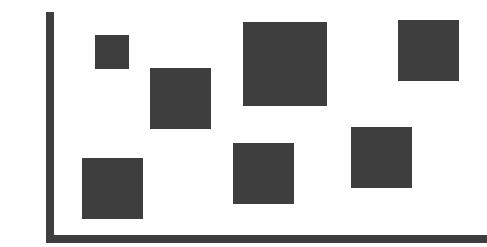
<https://observablehq.com/@d3/scatterplot-with-shapes>

Some keys

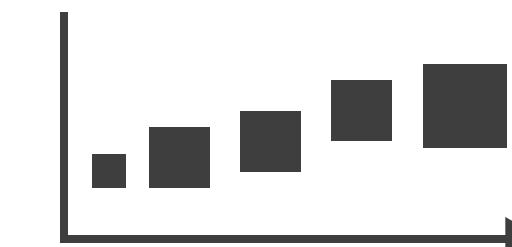


Regions: Separate, order, align

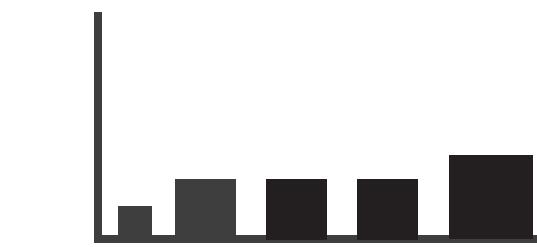
→ Separate



→ Order



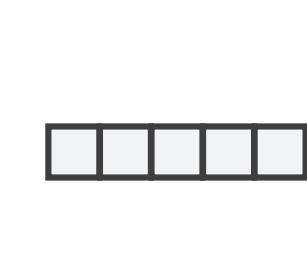
→ Align



- **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

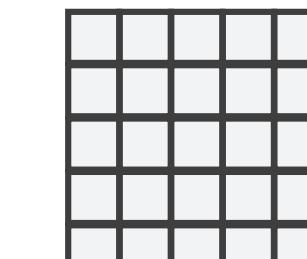
→ 1 Key

List



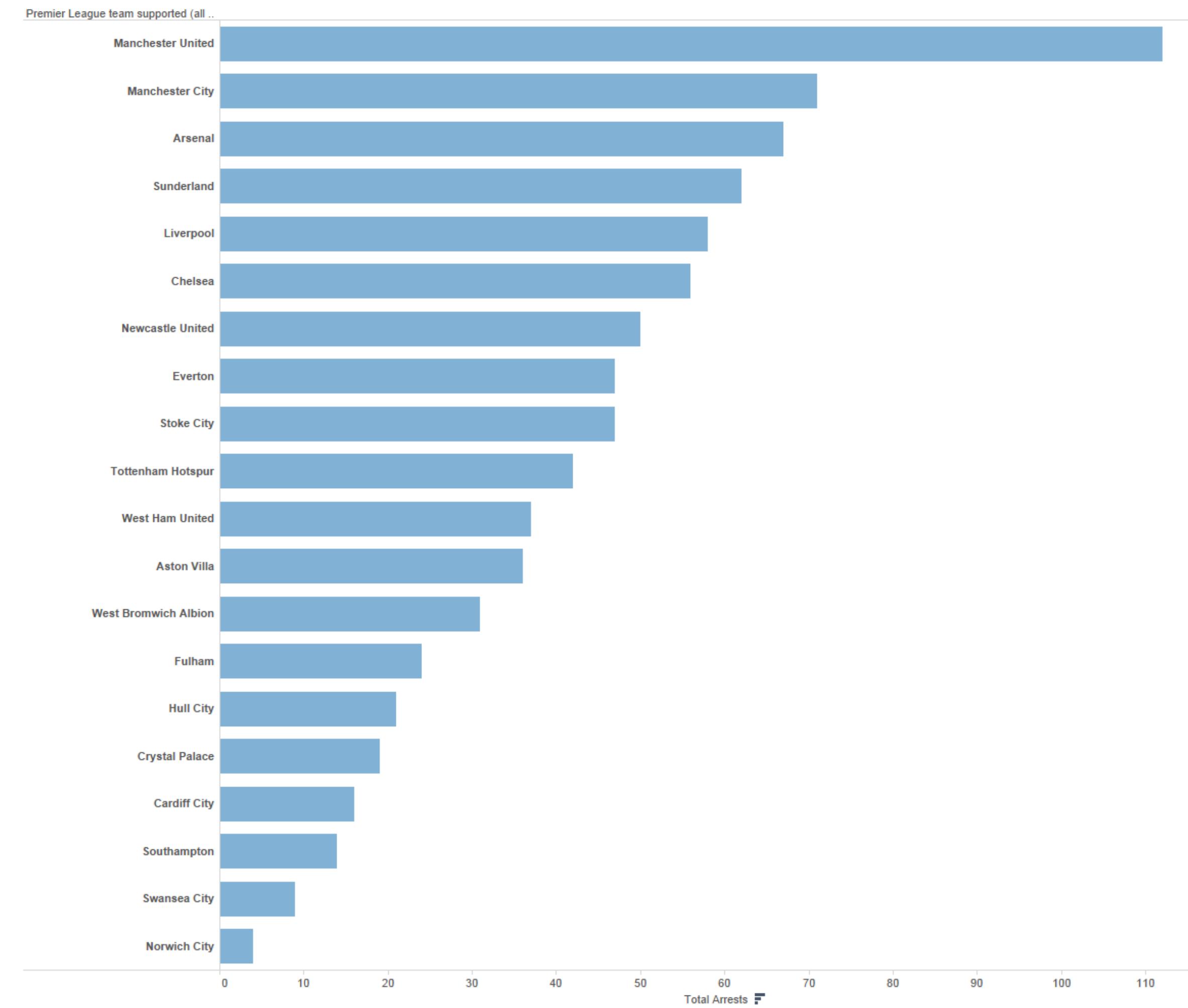
→ 2 Keys

Matrix



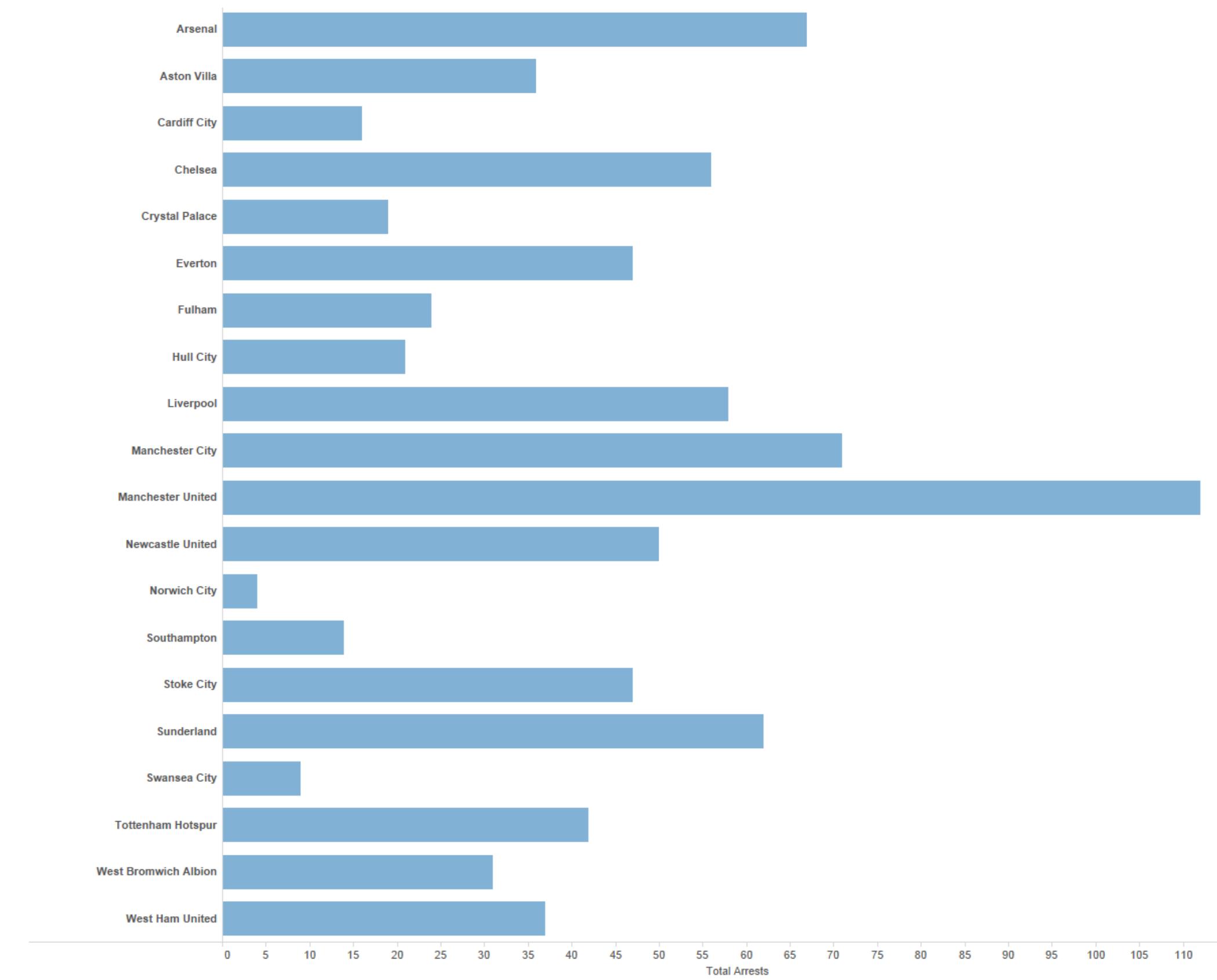
Separated and aligned and ordered

- best case



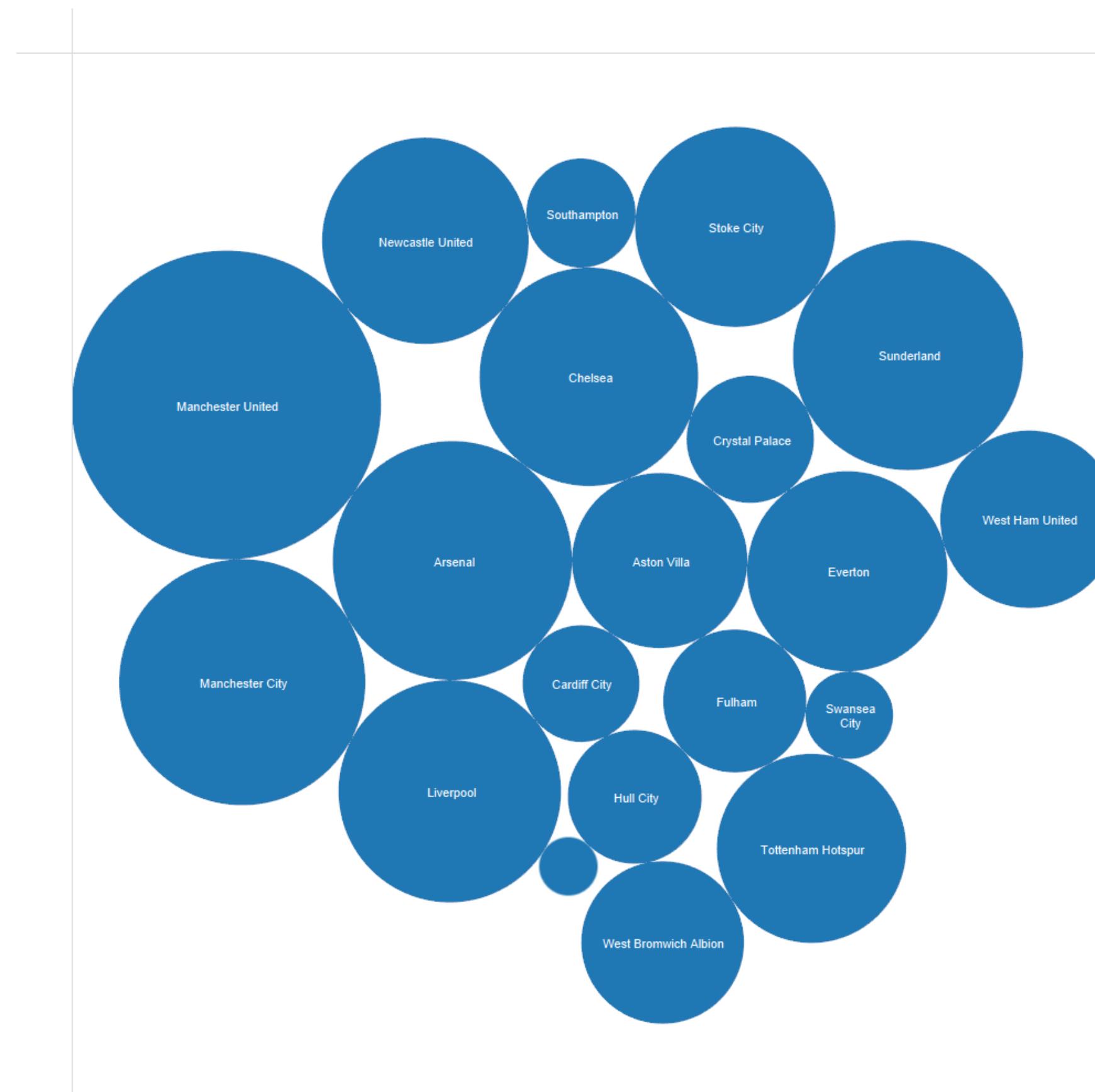
Separated and aligned but not ordered

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



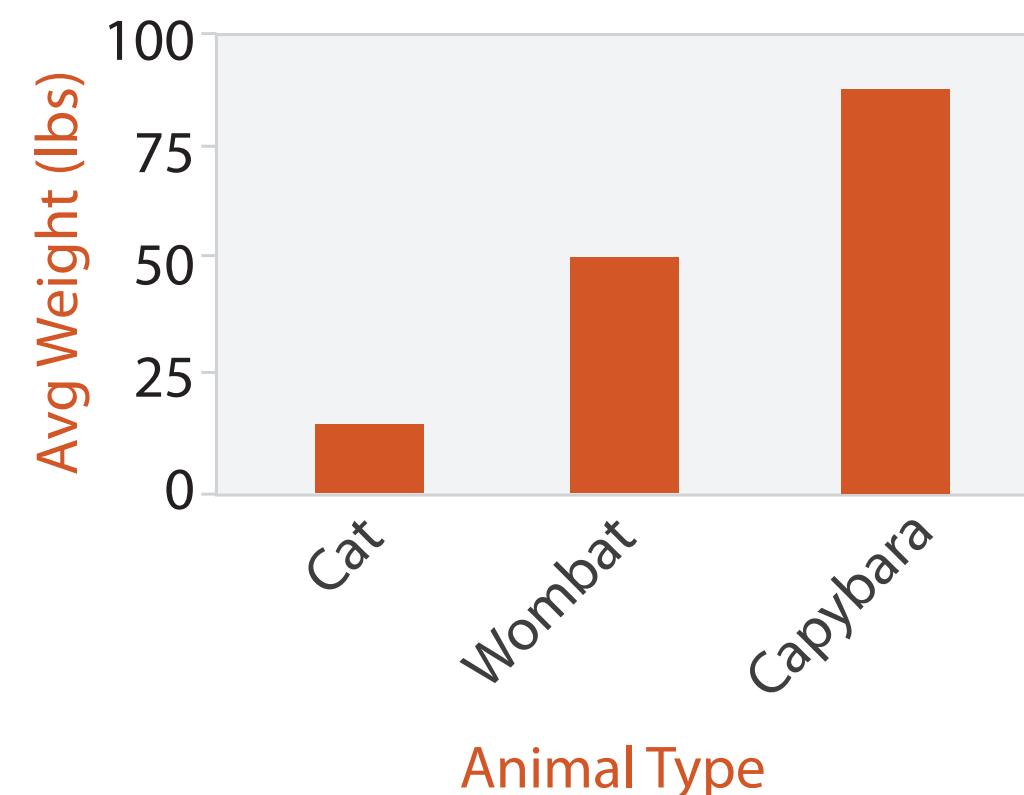
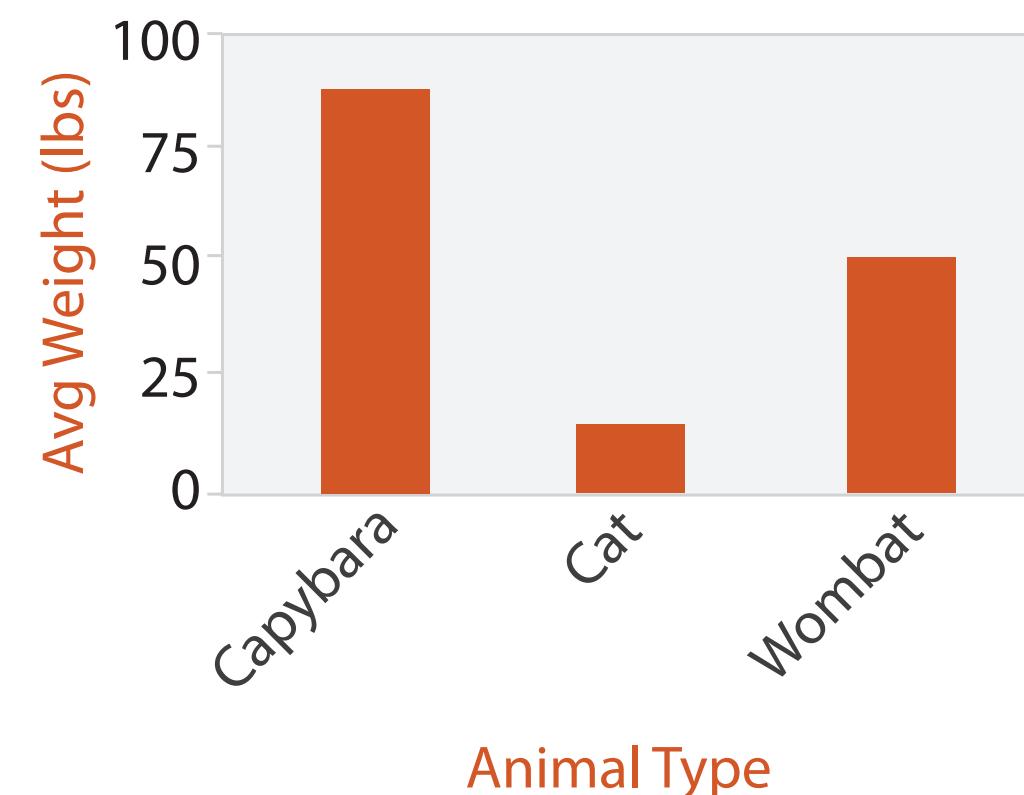
Separated but not aligned or ordered

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



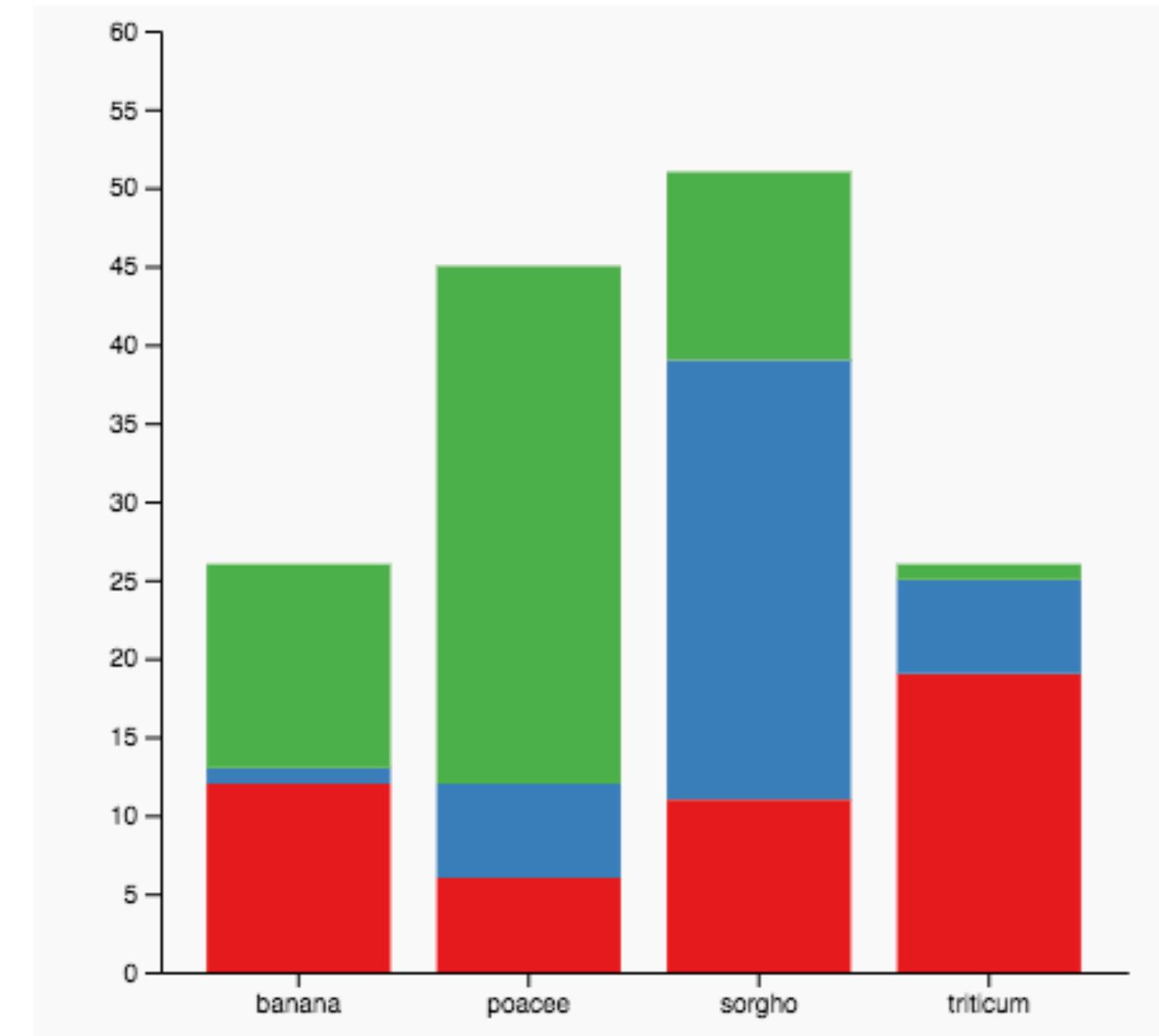
Idiom: bar chart

- 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



Idiom: stacked bar chart

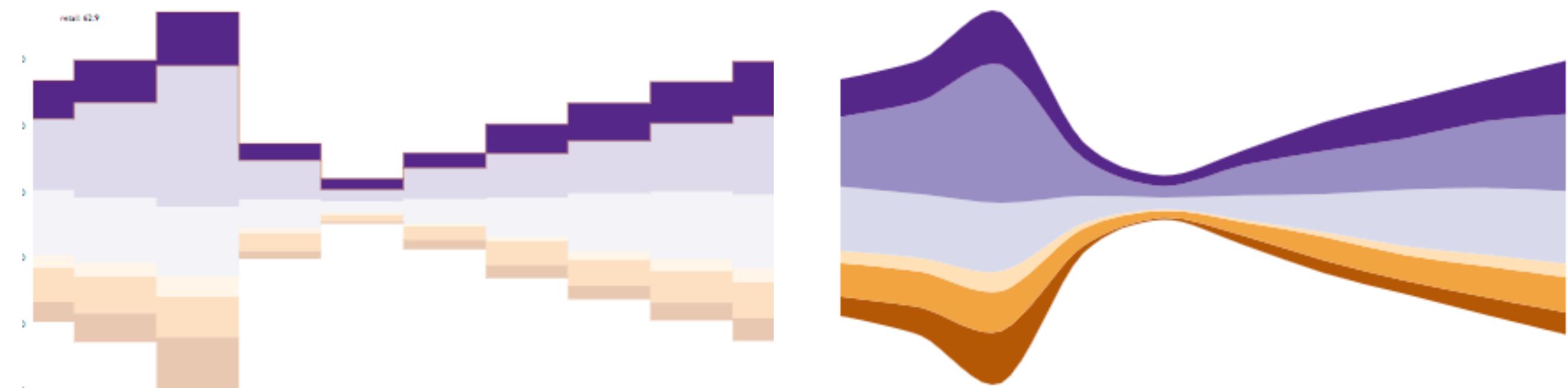
- one more key
 - data
 - 2 categ attrib, 1 quant attrib
 - mark: vertical stack of line marks
 - glyph: composite object, internal structure from multiple marks
 - channels
 - length and color hue
 - spatial regions: one per glyph
 - aligned: full glyph, lowest bar component
 - unaligned: other bar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



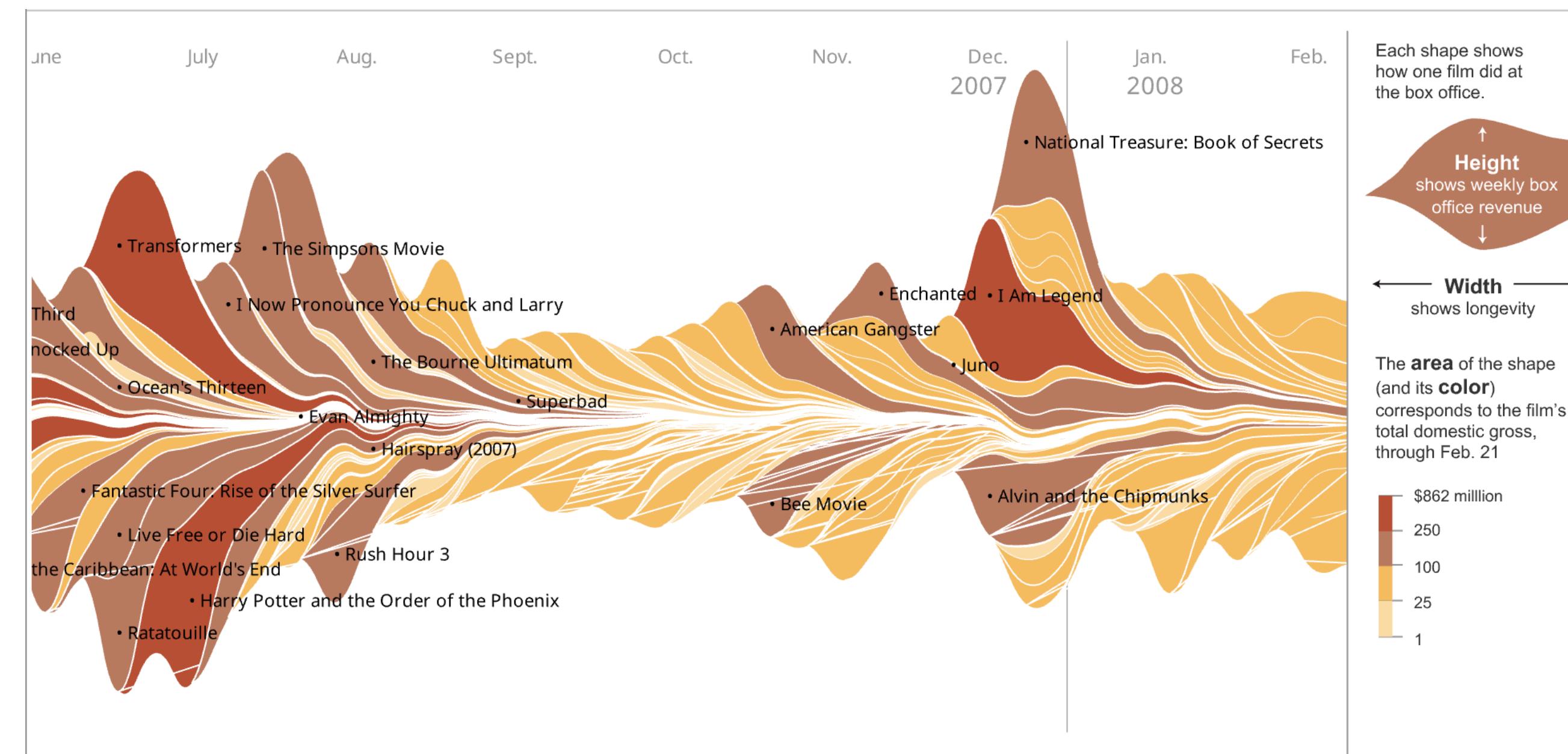
https://www.d3-graph-gallery.com/graph/barplot_stacked_basicWide.html

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

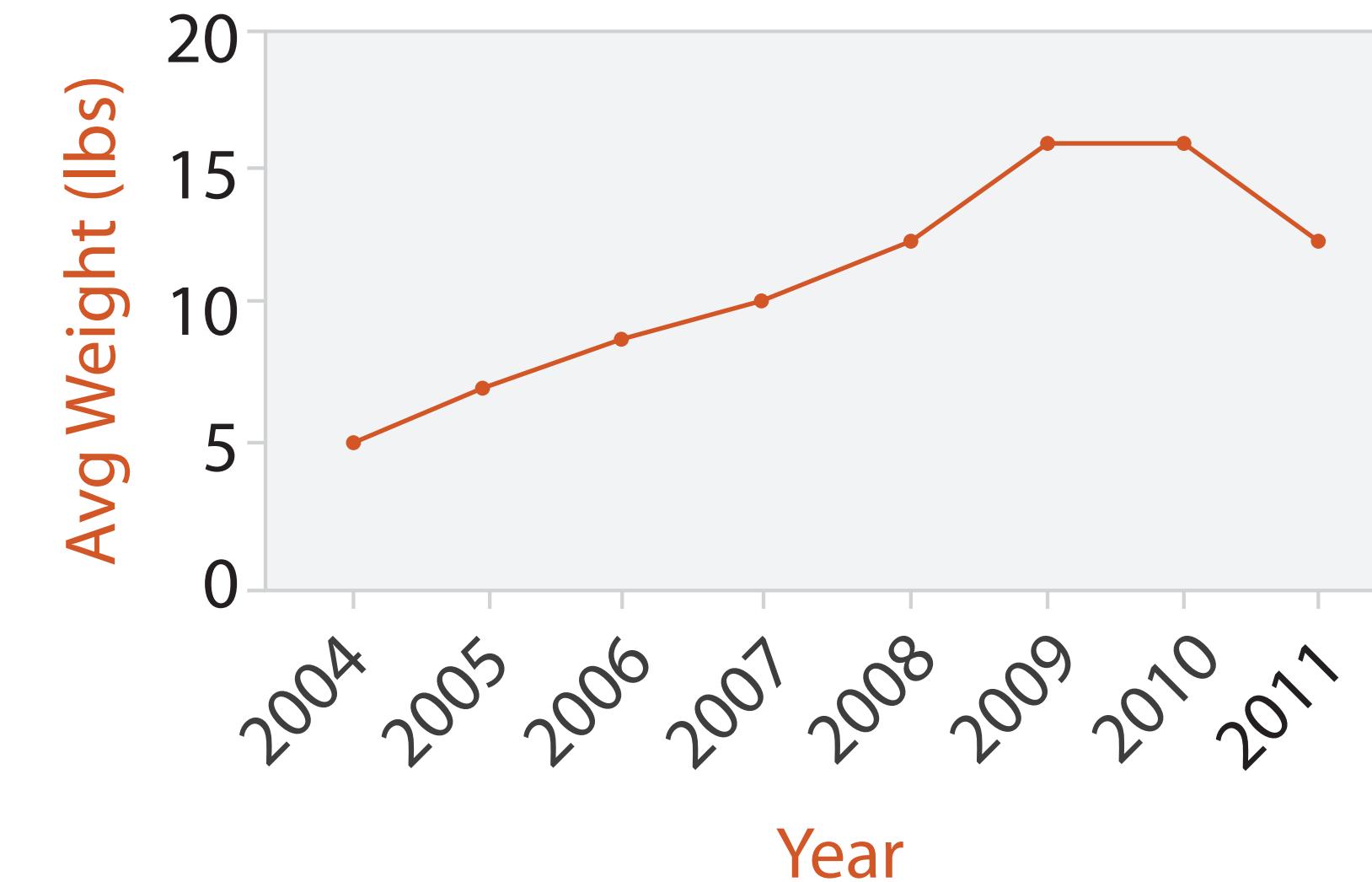
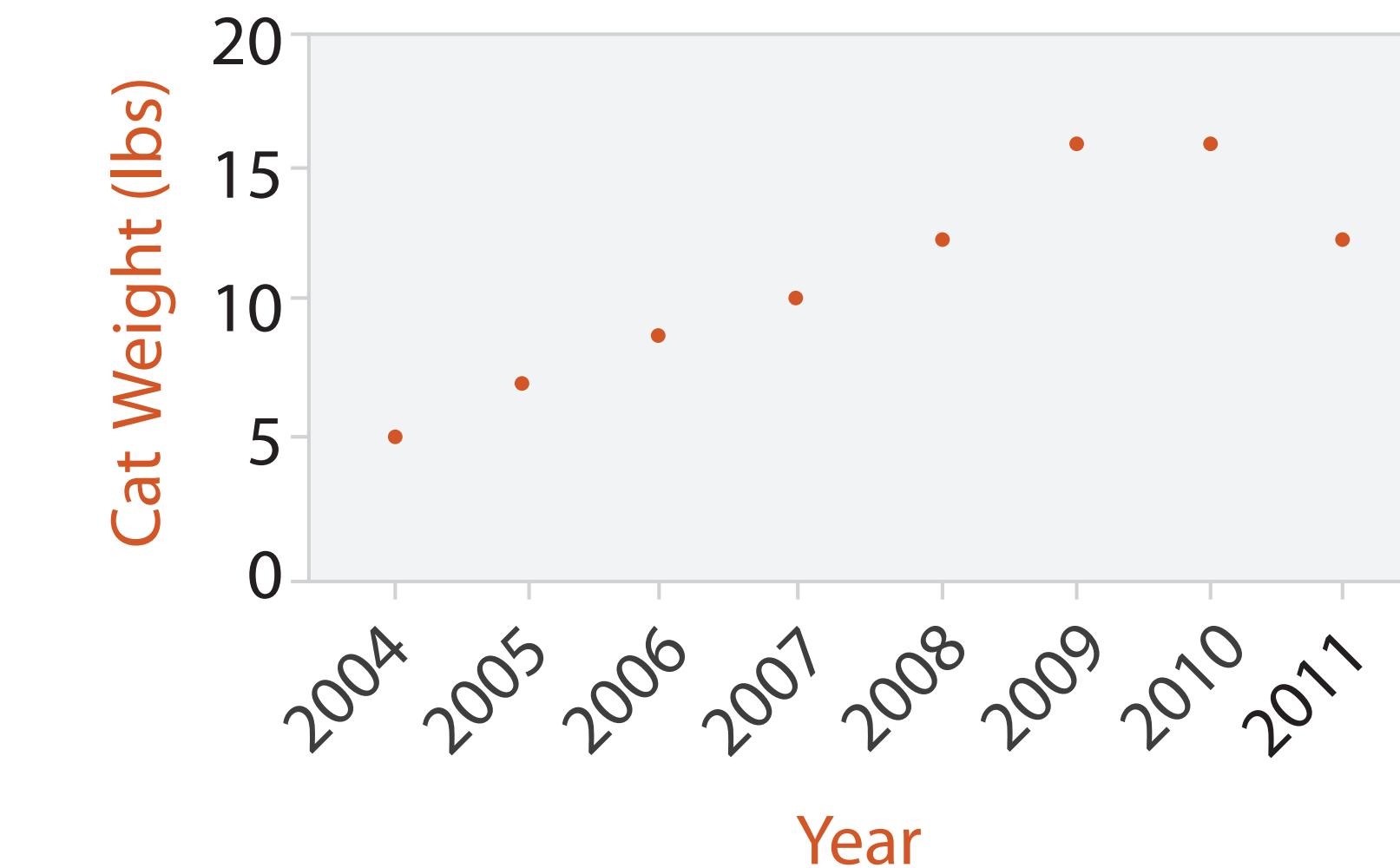


[Stacked Graphs Geometry & Aesthetics. Byron and Wattenberg. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2008) 14(6): 1245–1252, (2008).]



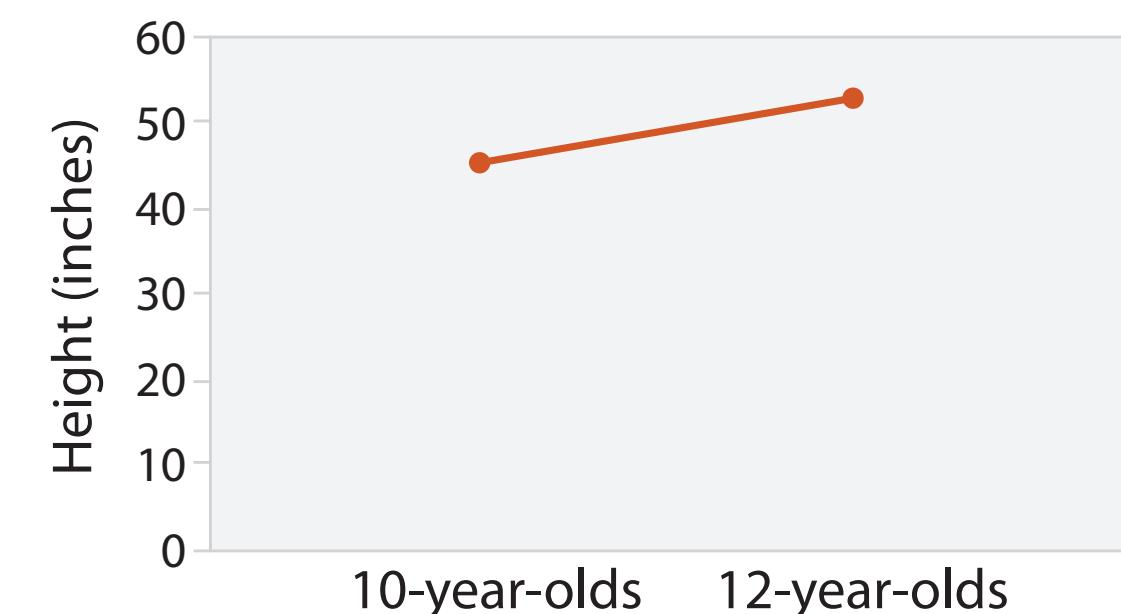
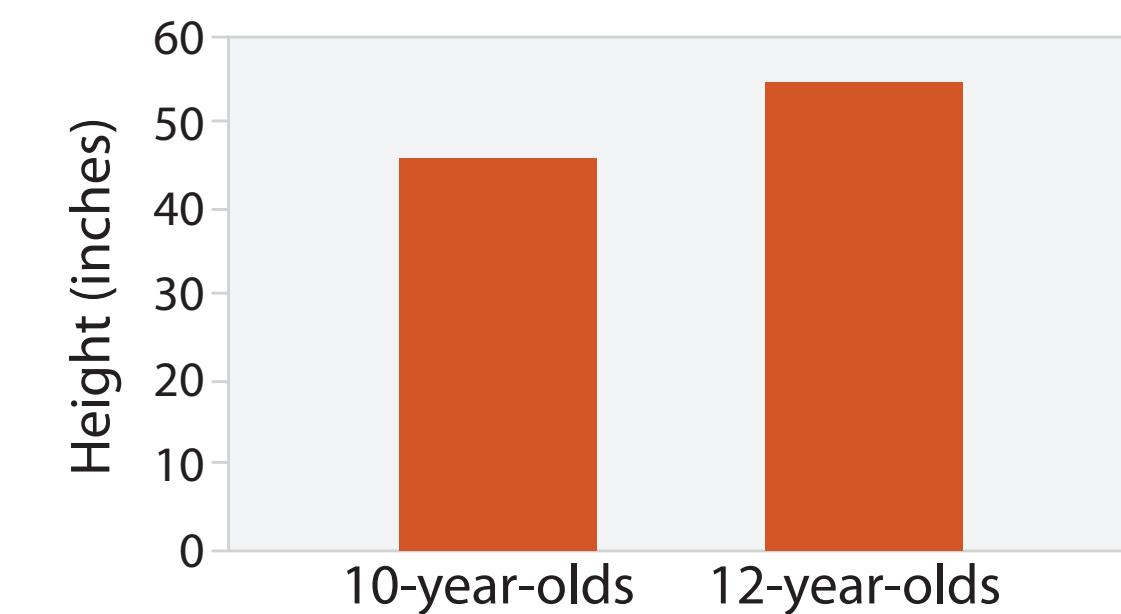
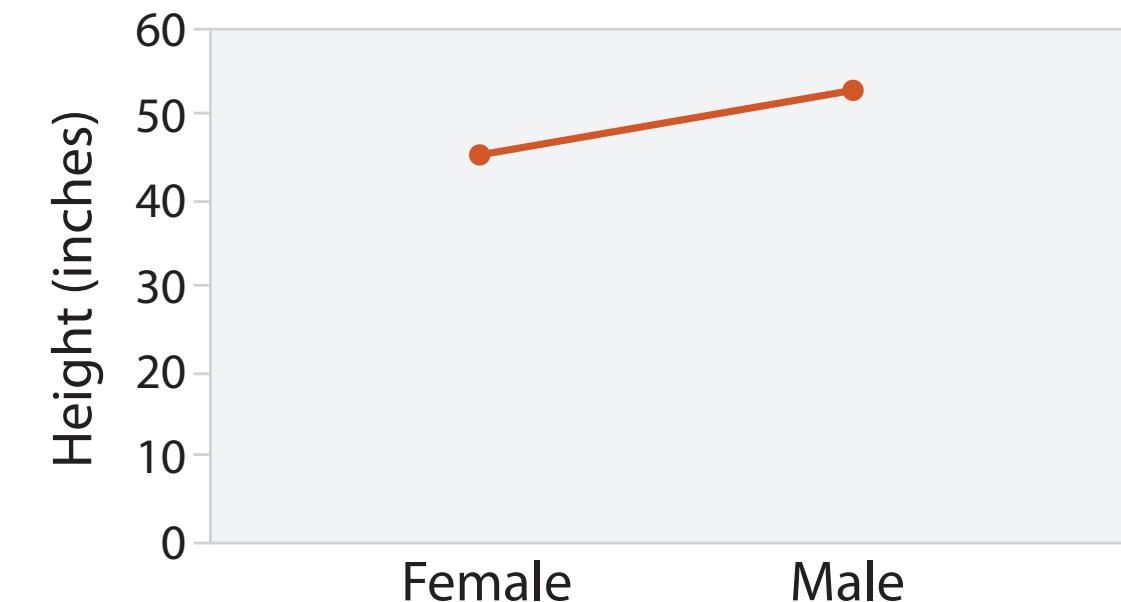
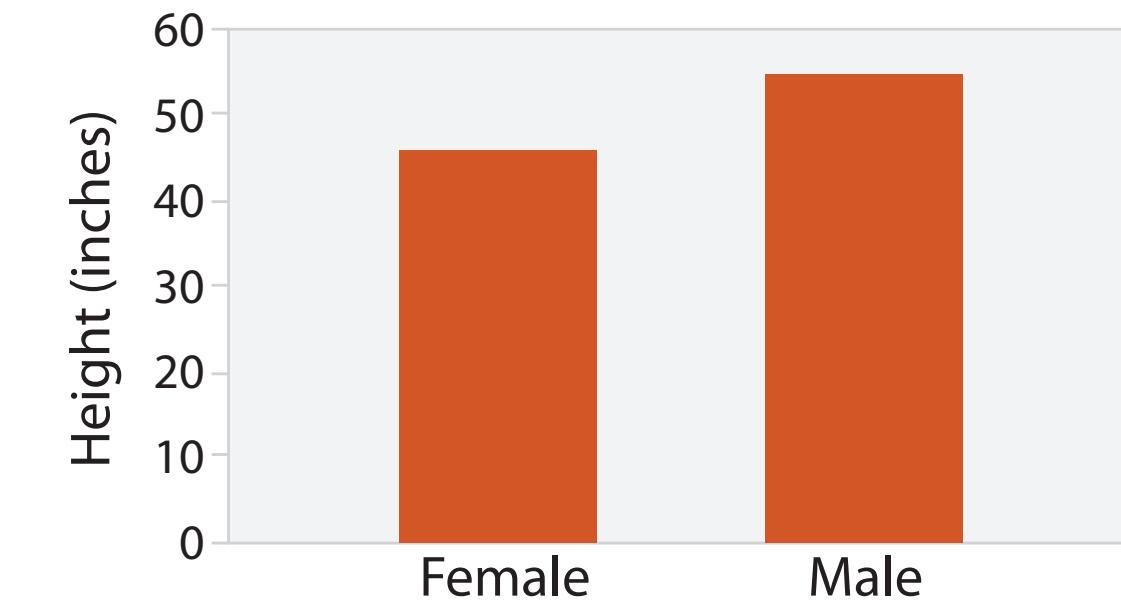
Idiom: dot / line chart

- 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



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 attrs
 - violates expressiveness principle
 - implication of trend so strong that it overrides semantics!
 - “The more male a person is, the taller he/she is”



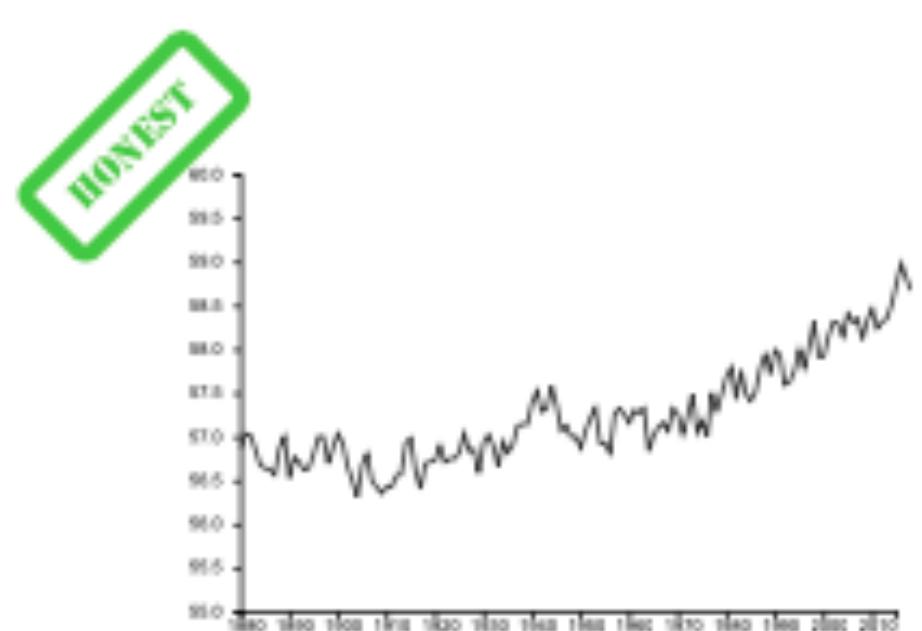
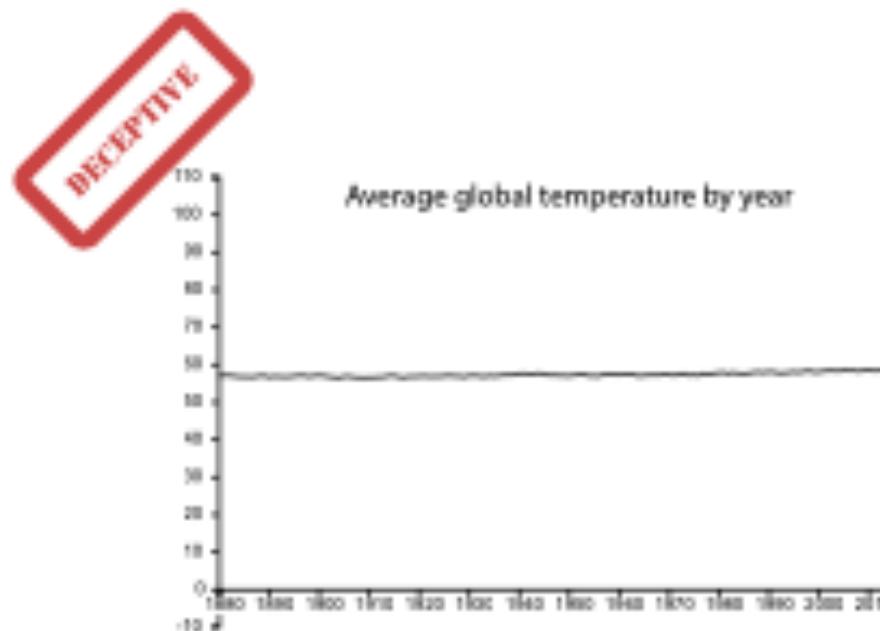
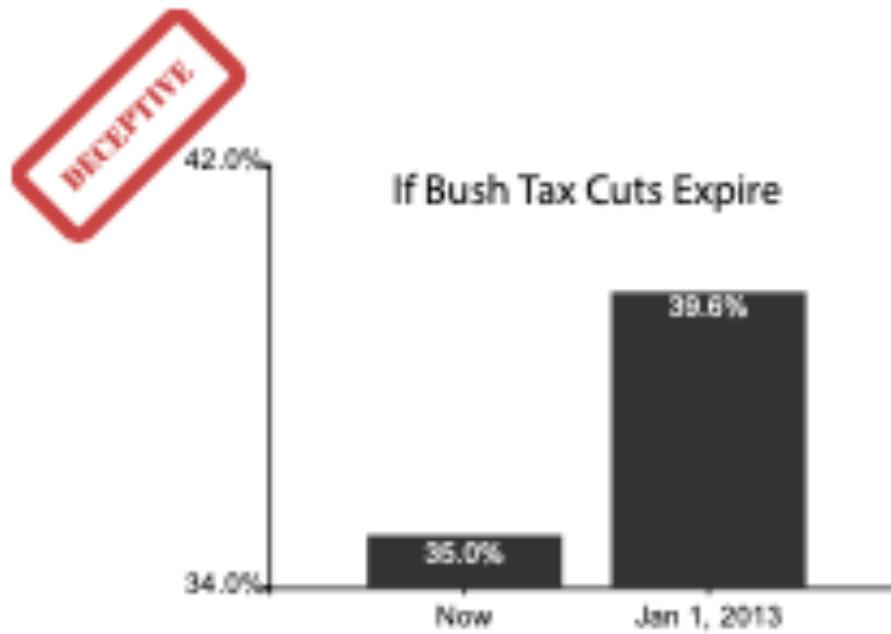
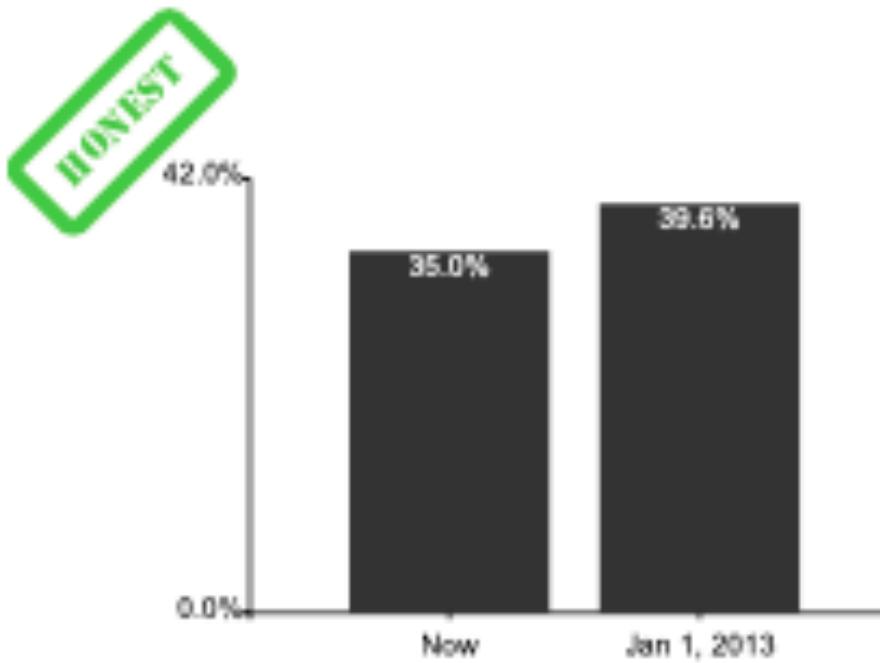
[Bars and Lines: A Study of Graphic Communication. Zacks and Tversky. Memory and Cognition 27:6 (1999), 1073–1079.]

Chart axes: label them!

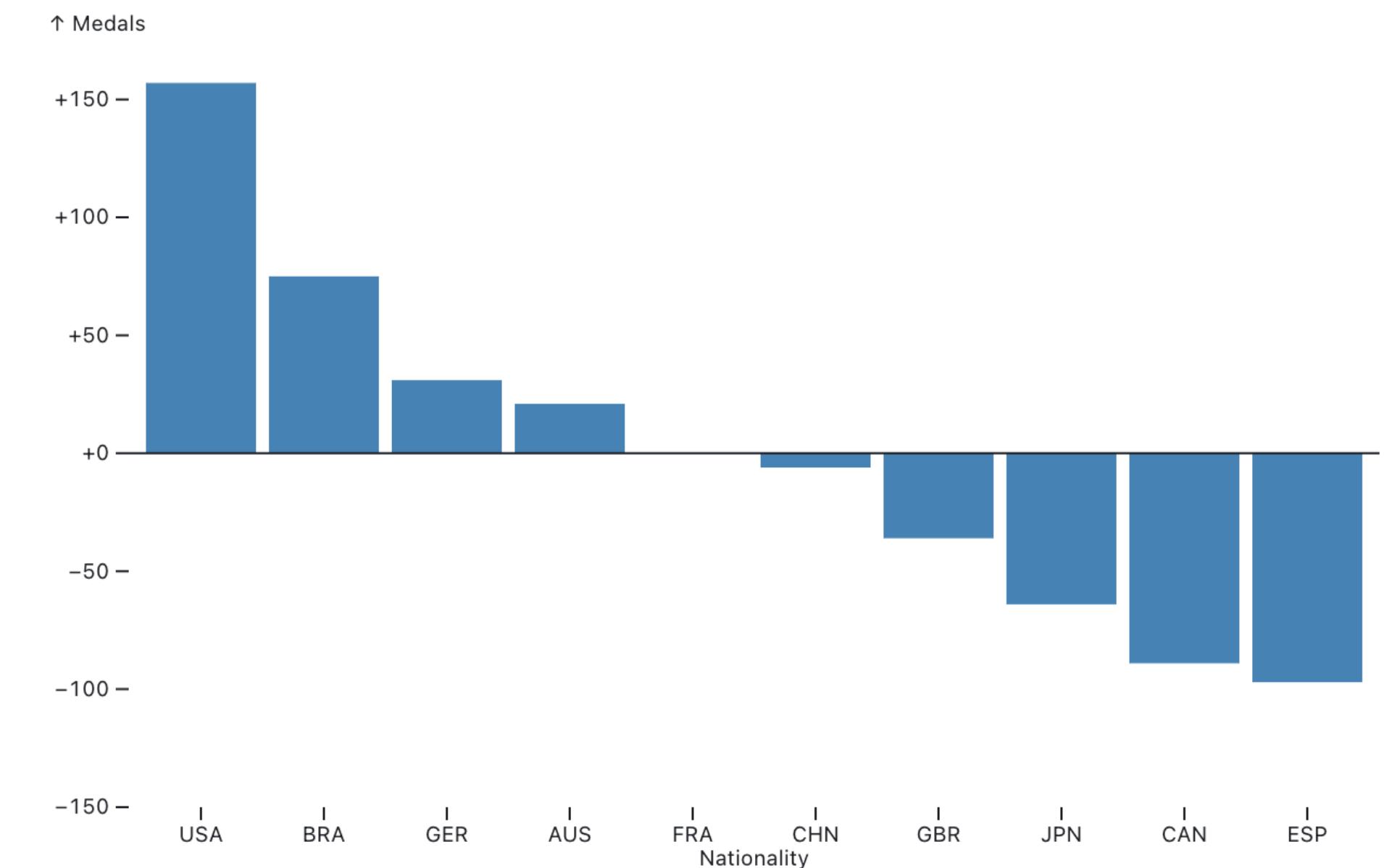
- best practice to label
 - few exceptions: individual small multiple views could share axis label

Chart axes: avoid cropping y axis

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



- Relative axis instead of cropping
 - # medals compared to France (reference)

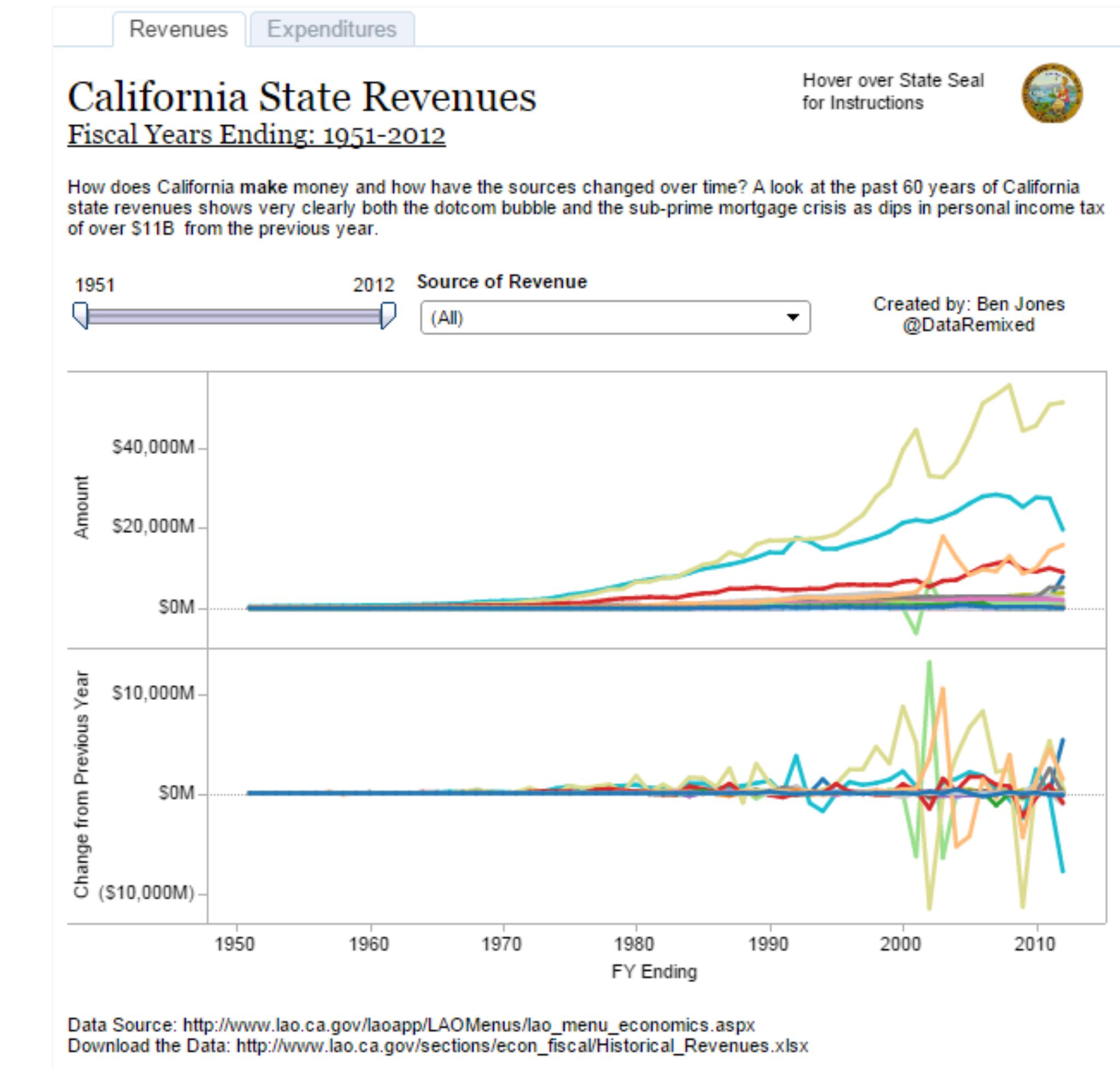


<https://observablehq.com/blog/never-okay-crop-y-axis-except-when-it-is>

<https://observablehq.com/blog/why-never-crop-y-axis-part-2-how-to-and-what-to-do-instead>

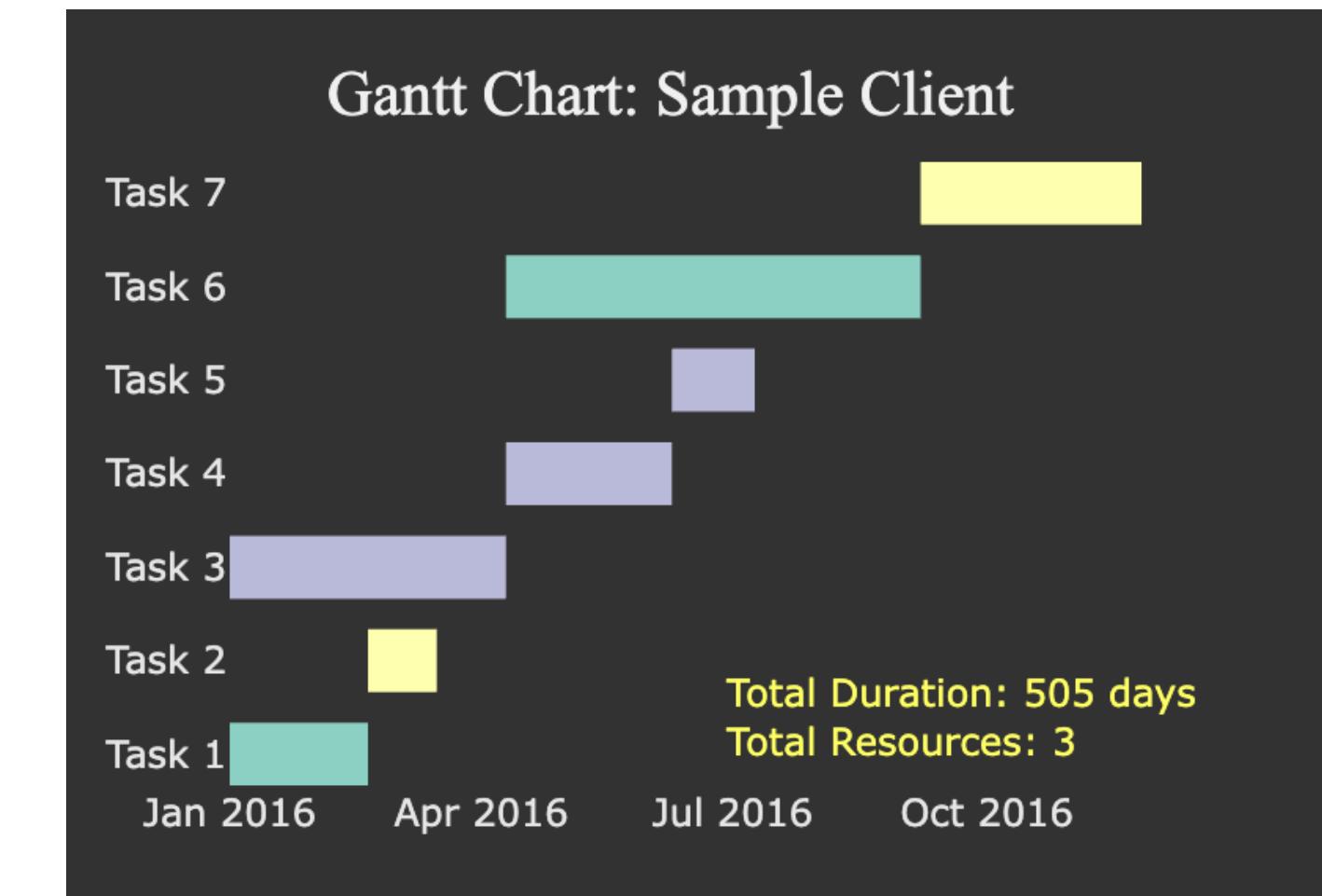
Idiom: Indexed line charts

- data: 2 quant attrs
 - 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



Idiom: Gantt charts

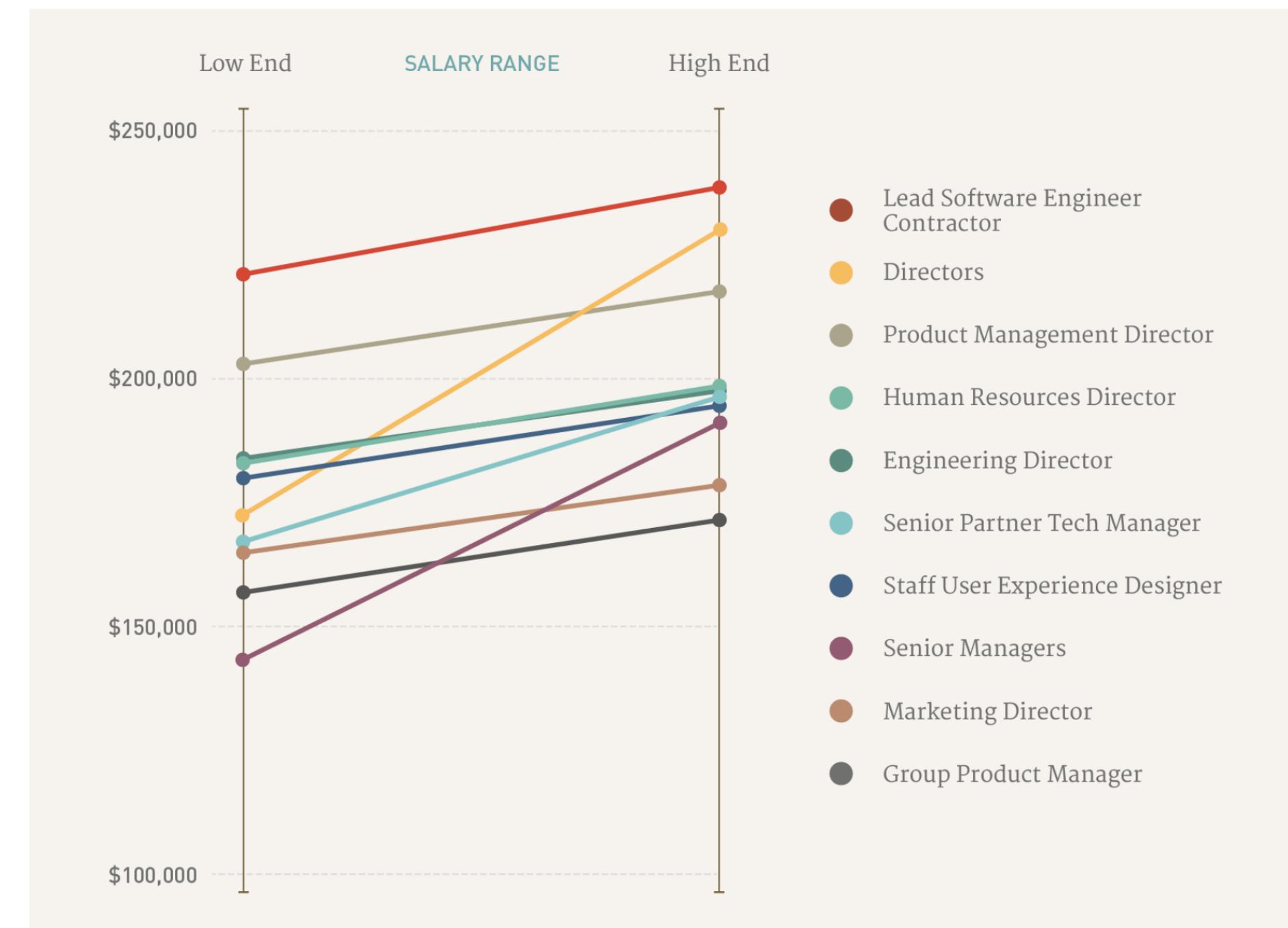
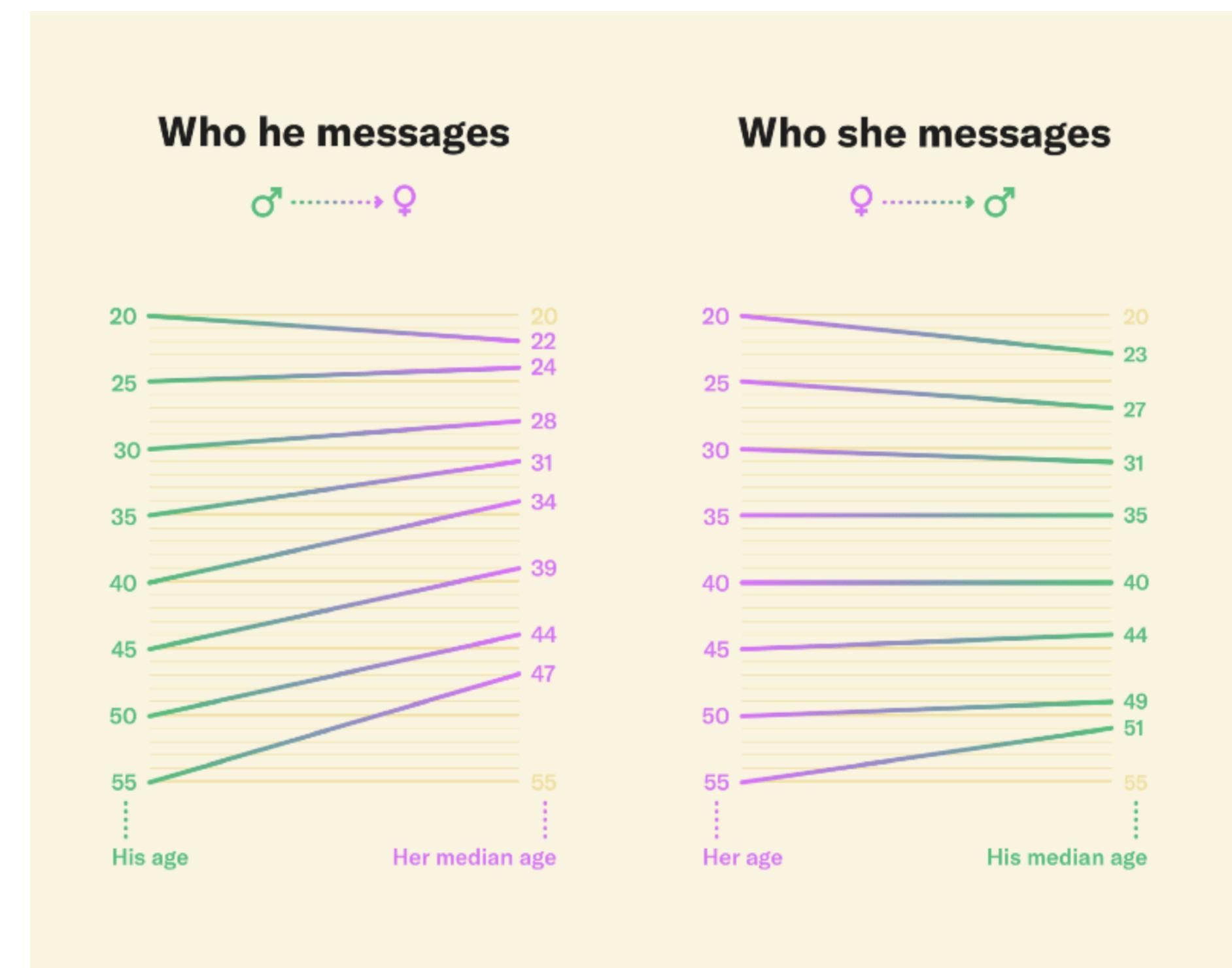
- one key, two (related) values
 - data
 - 1 categ attrib, 2 quant attrs
 - mark: line
 - length: duration
 - channels
 - task
- emphasize temporal overlaps & start/end dependencies between items
- scalability
 - dozens of key levels [bars]
 - hundreds of value levels [durations]



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

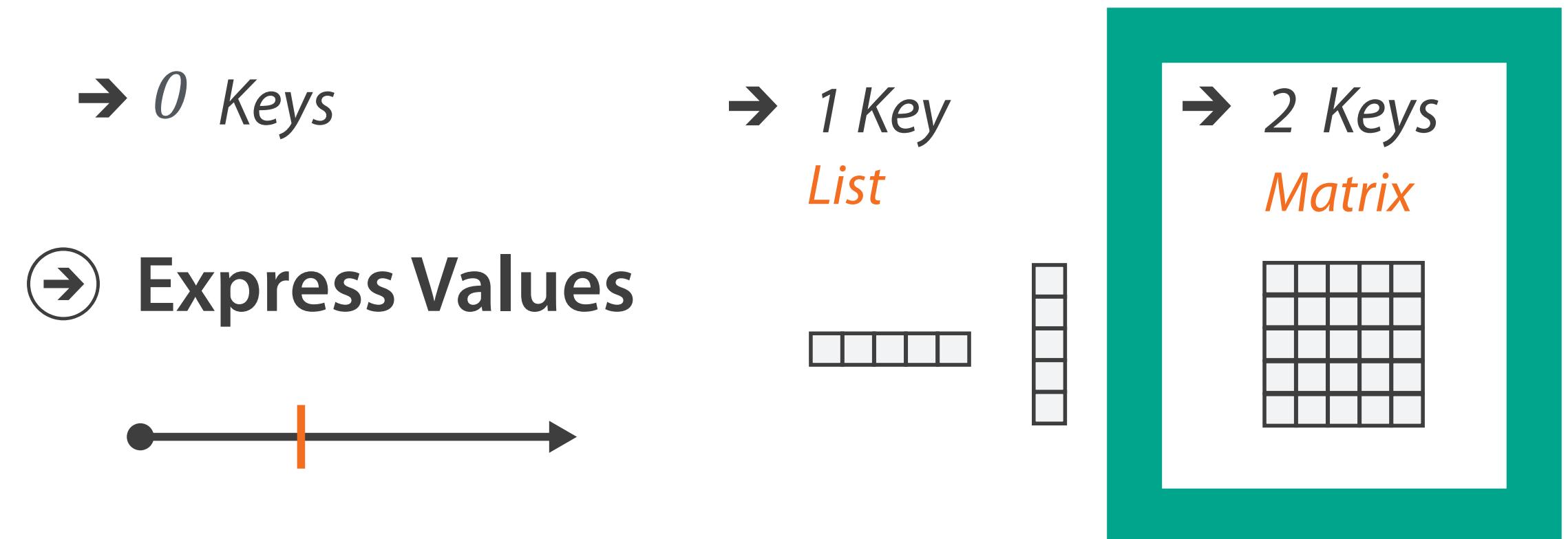
Idiom: Slopegraphs

- two values
 - data
 - 2 quant value attrs
 - (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



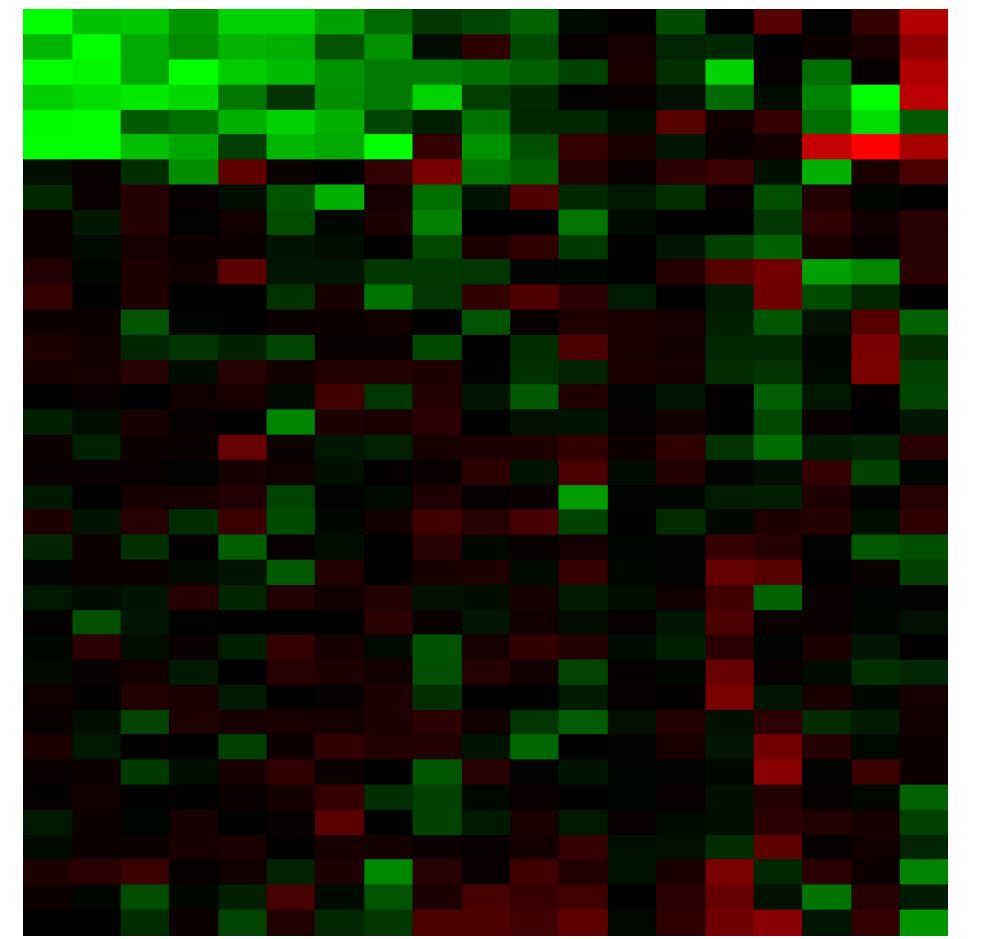
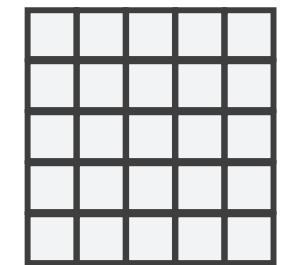
2 Keys

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



Idiom: heatmap

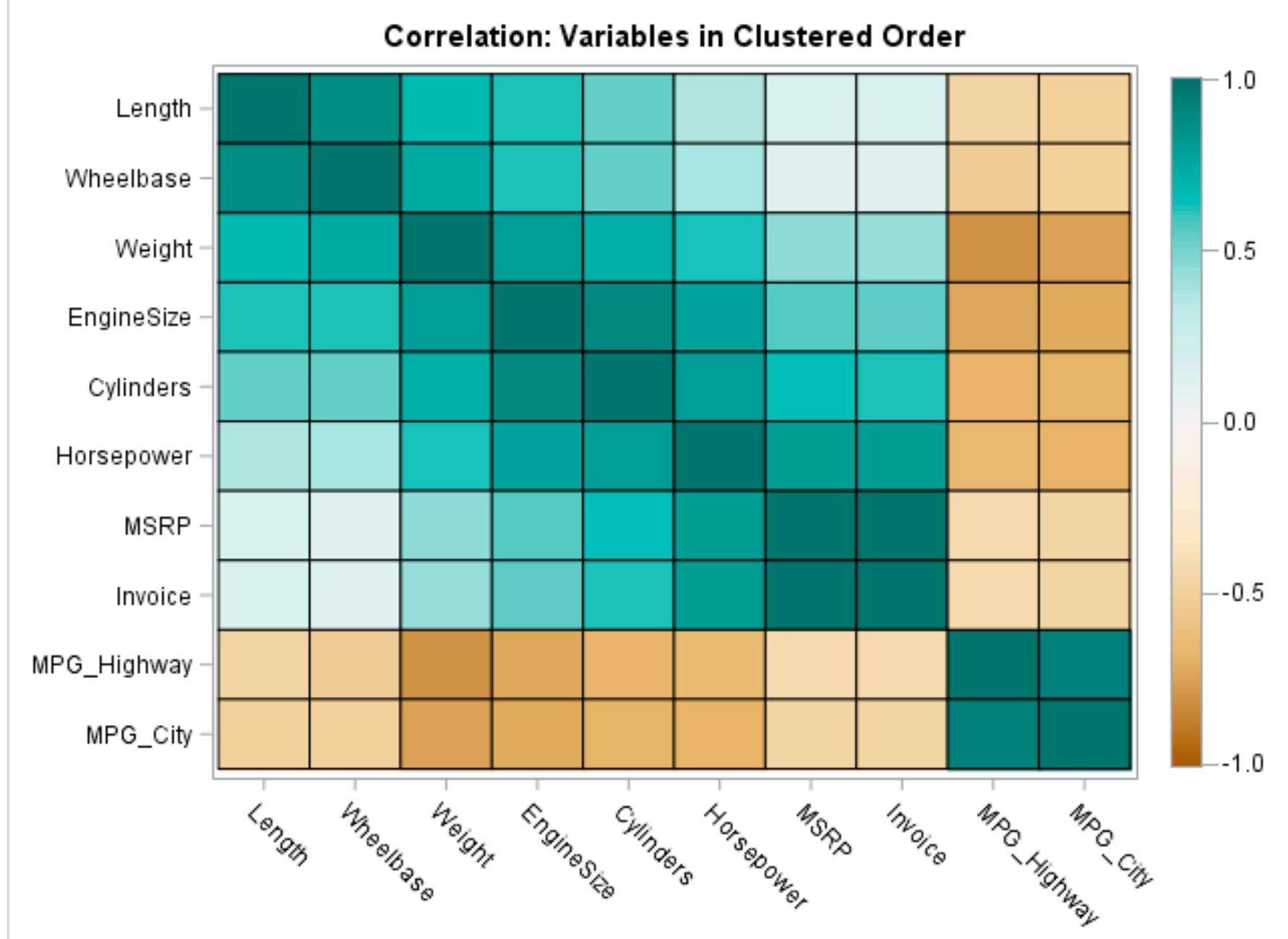
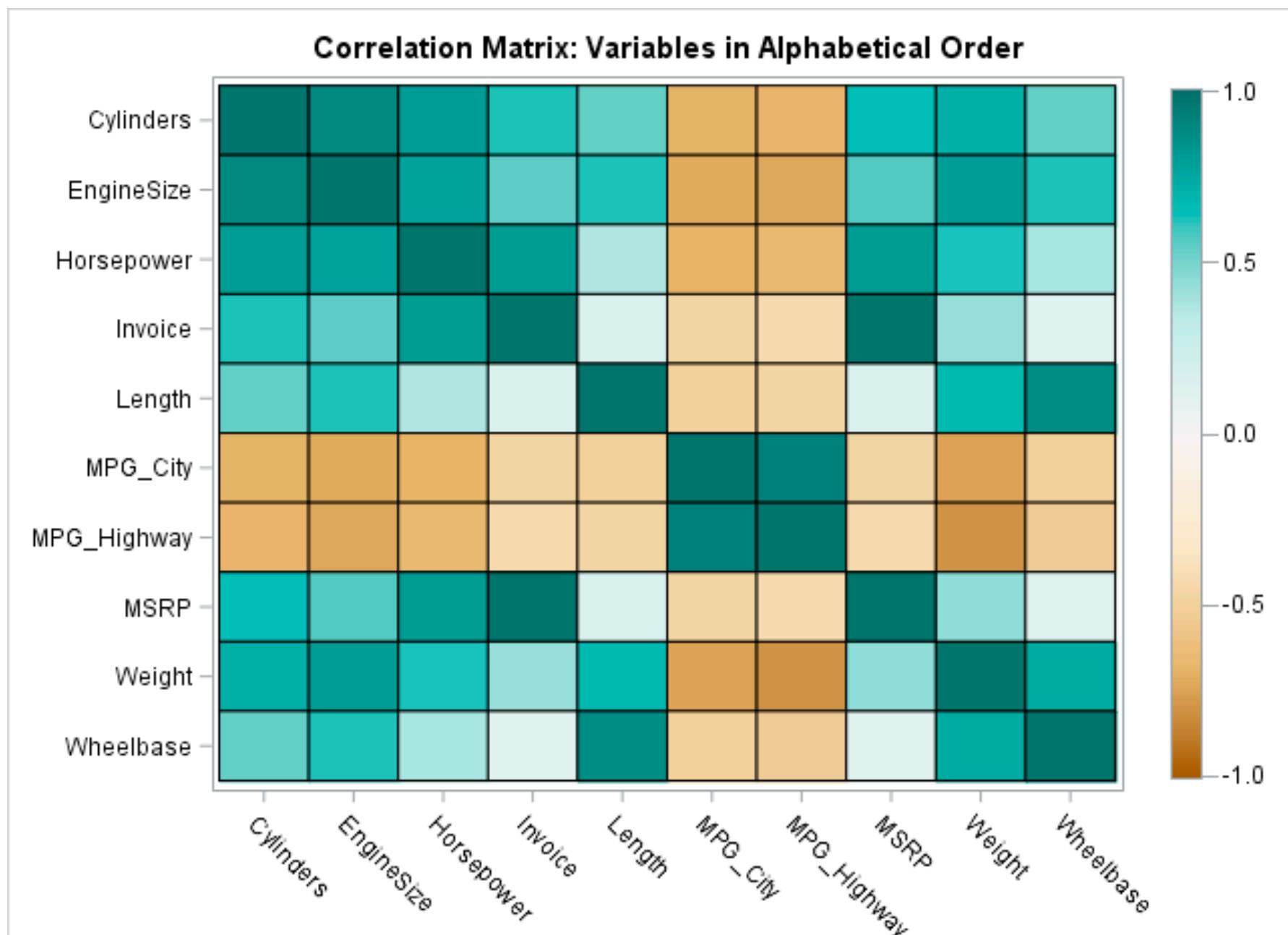
→ 2 Keys
Matrix



- 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

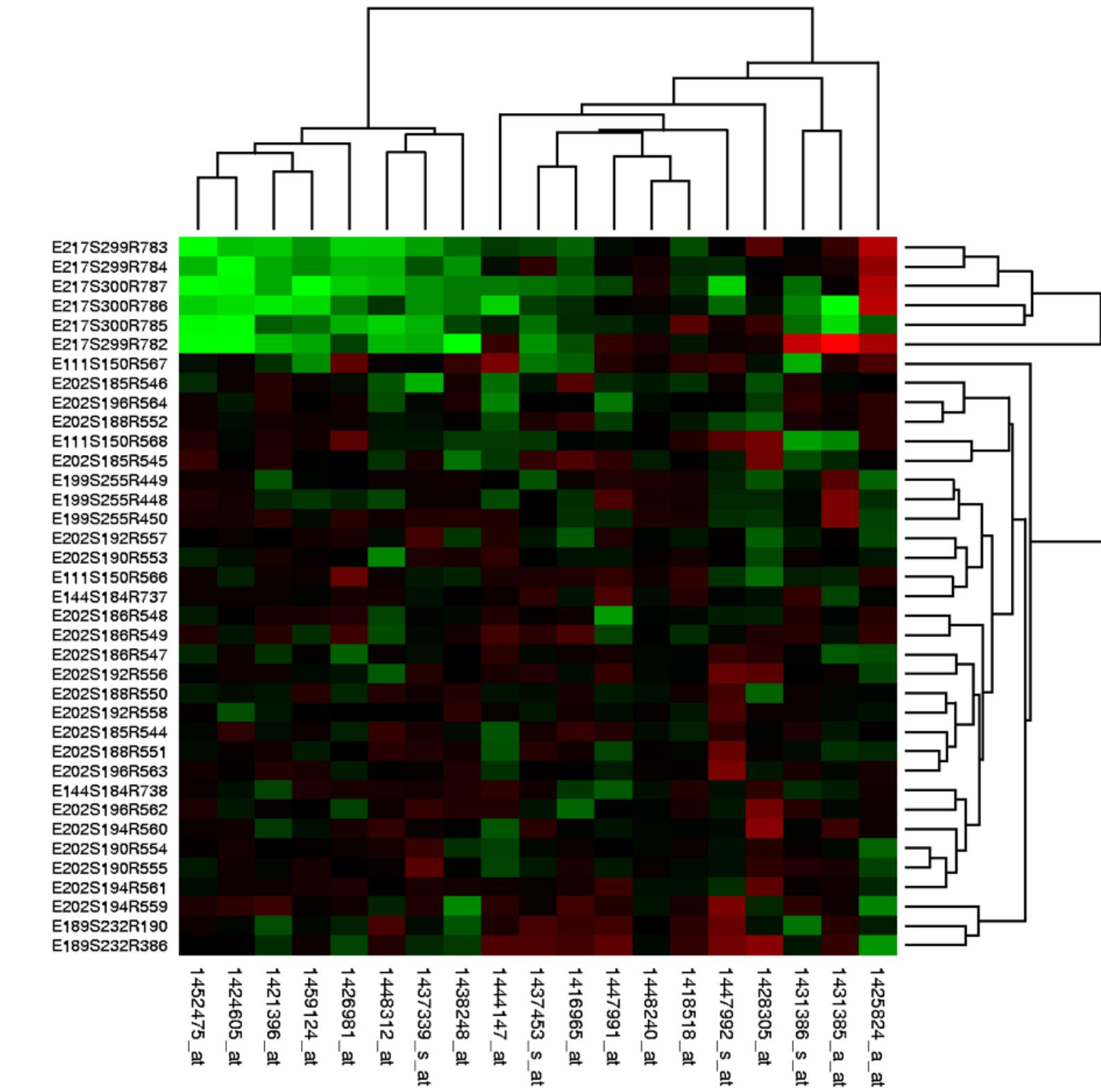
Heatmap reordering

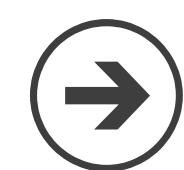
- in addition
 - derived data
 - 2 cluster hierarchies
 - dendrogram
 - parent-child relationships in tree with connection line marks
 - leaves aligned so interior branch heights easy to compare
 - heatmap
 - marks (re-)ordered by cluster hierarchy traversal
 - task: assess quality of clusters found by automatic methods



Idiom: cluster heatmap

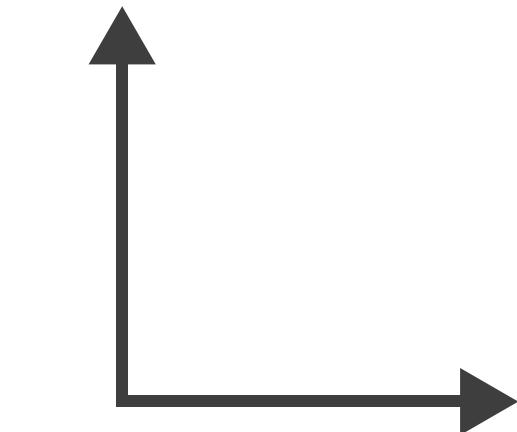
- in addition
 - derived data
 - 2 cluster hierarchies
 - dendrogram
 - parent-child relationships in tree with connection line marks
 - leaves aligned so interior branch heights easy to compare
 - heatmap
 - marks (re-)ordered by cluster hierarchy traversal
 - task: assess quality of clusters found by automatic methods



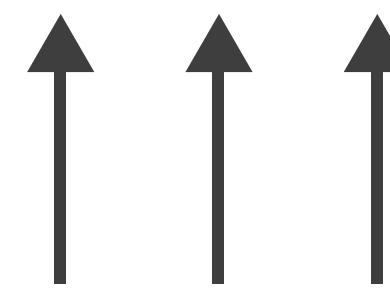


Axis Orientation

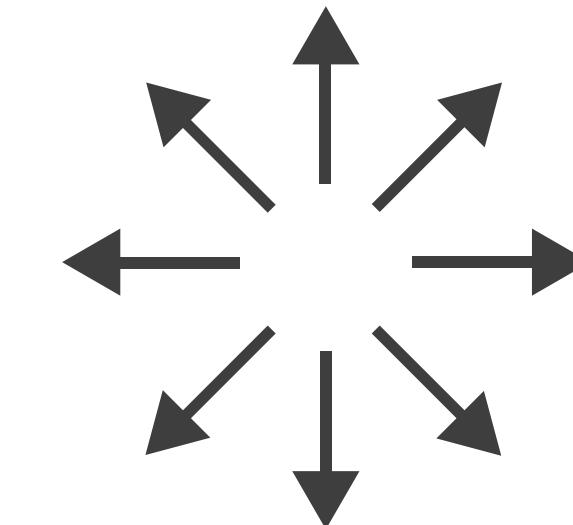
→ Rectilinear



→ Parallel

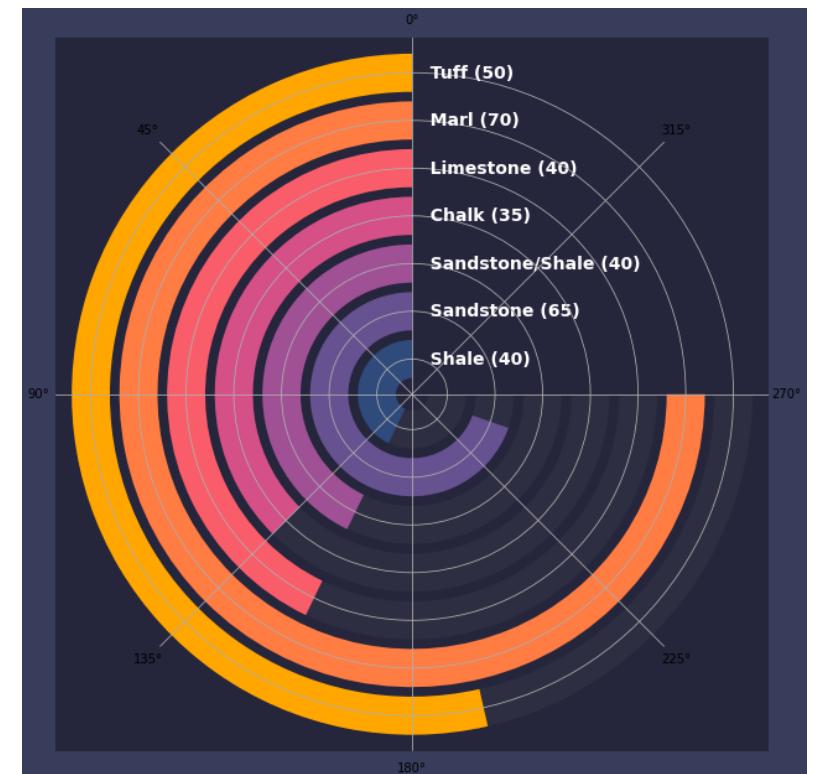
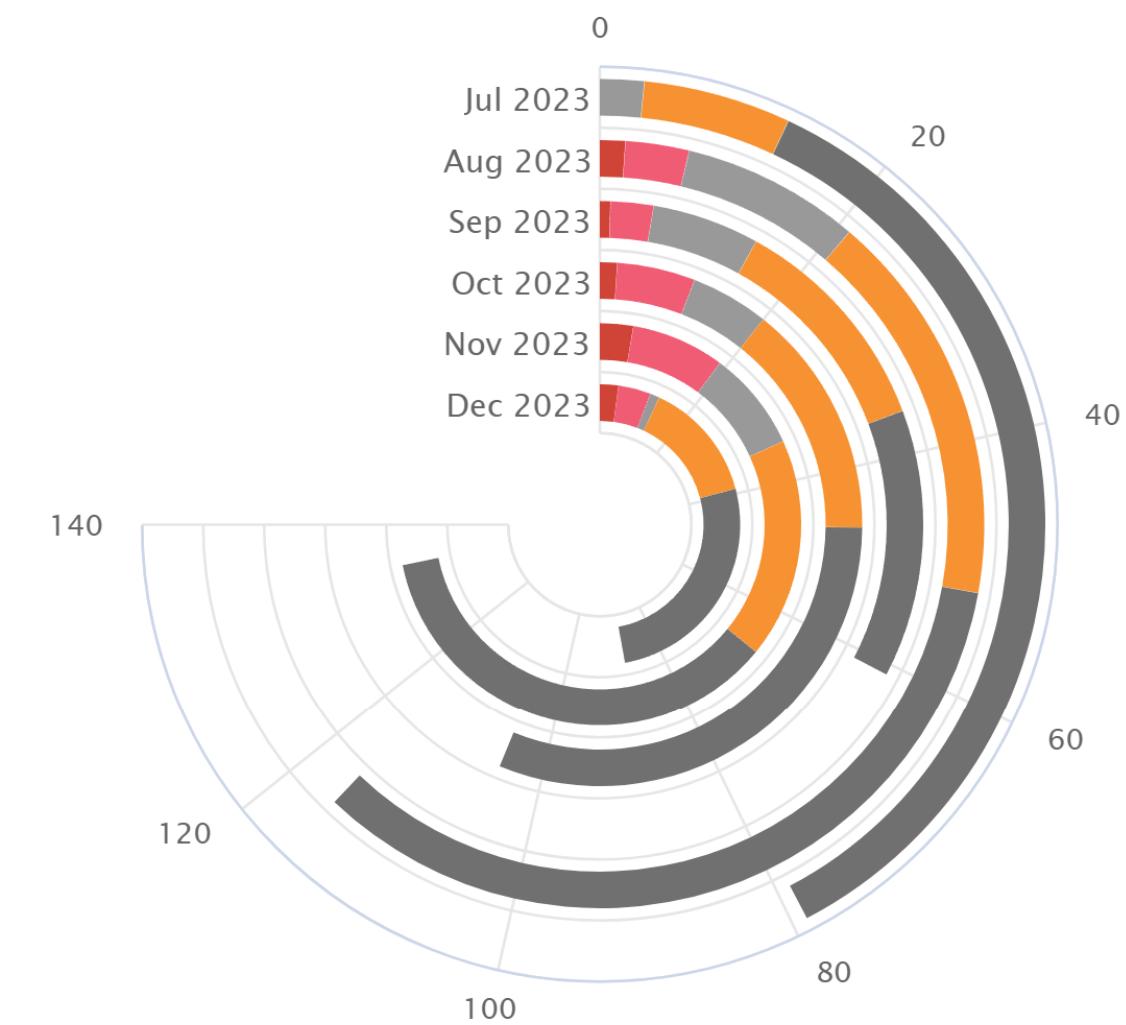
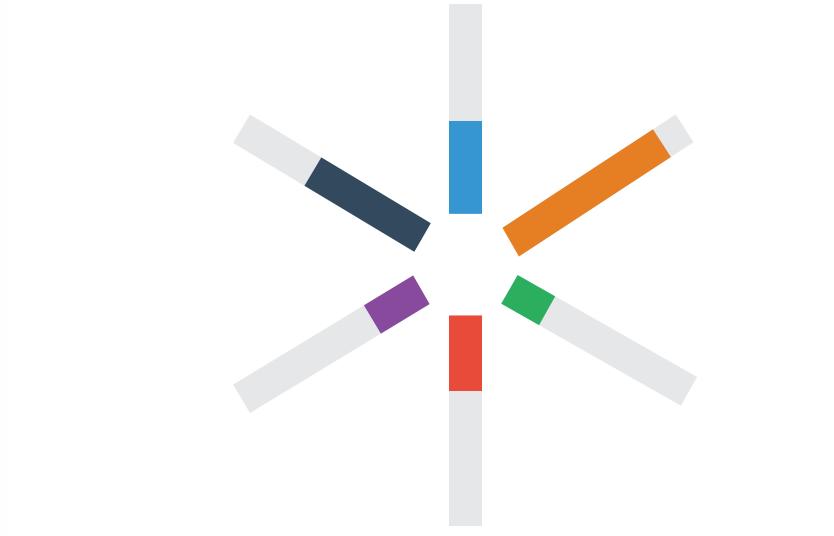
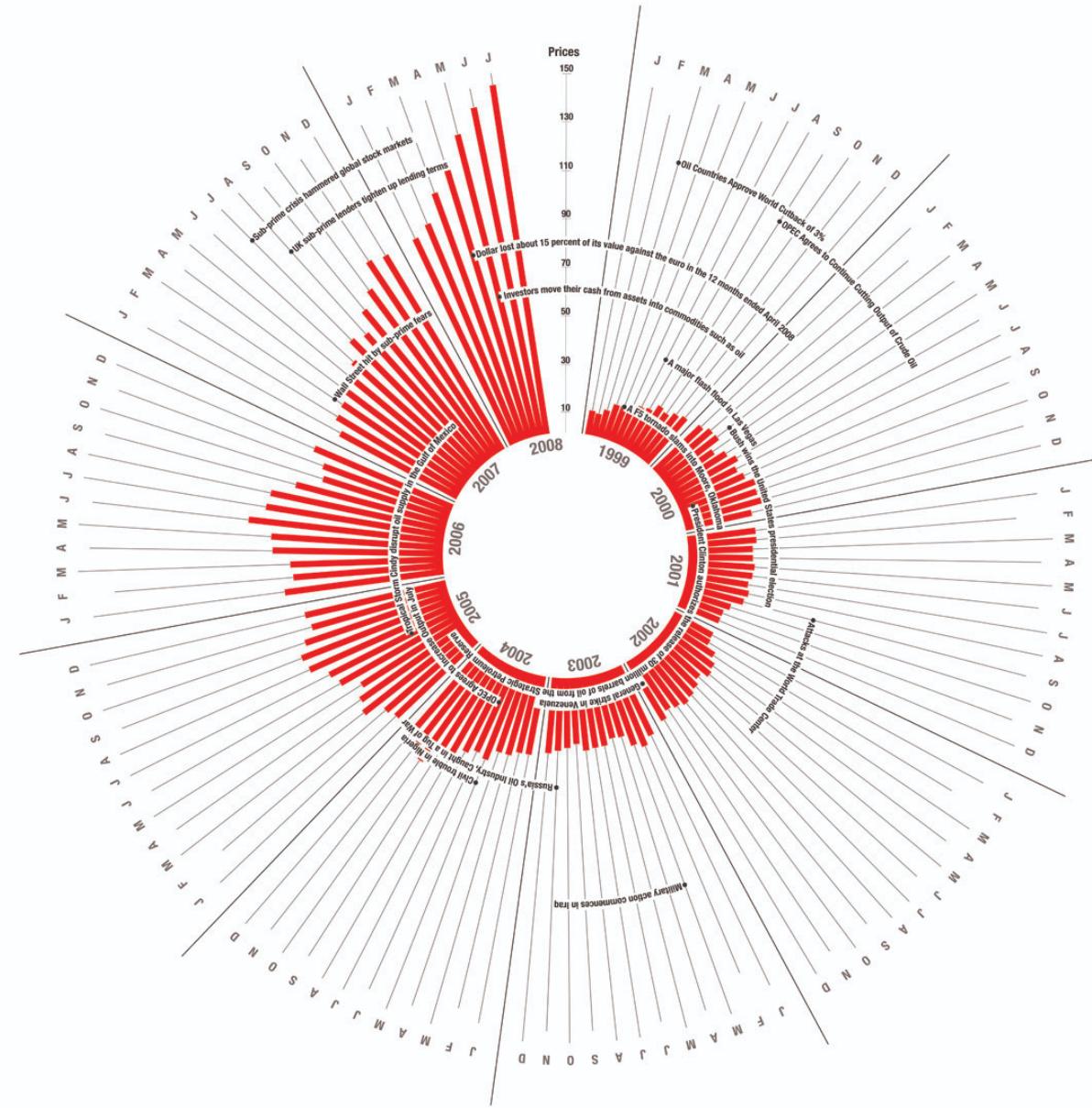


→ Radial



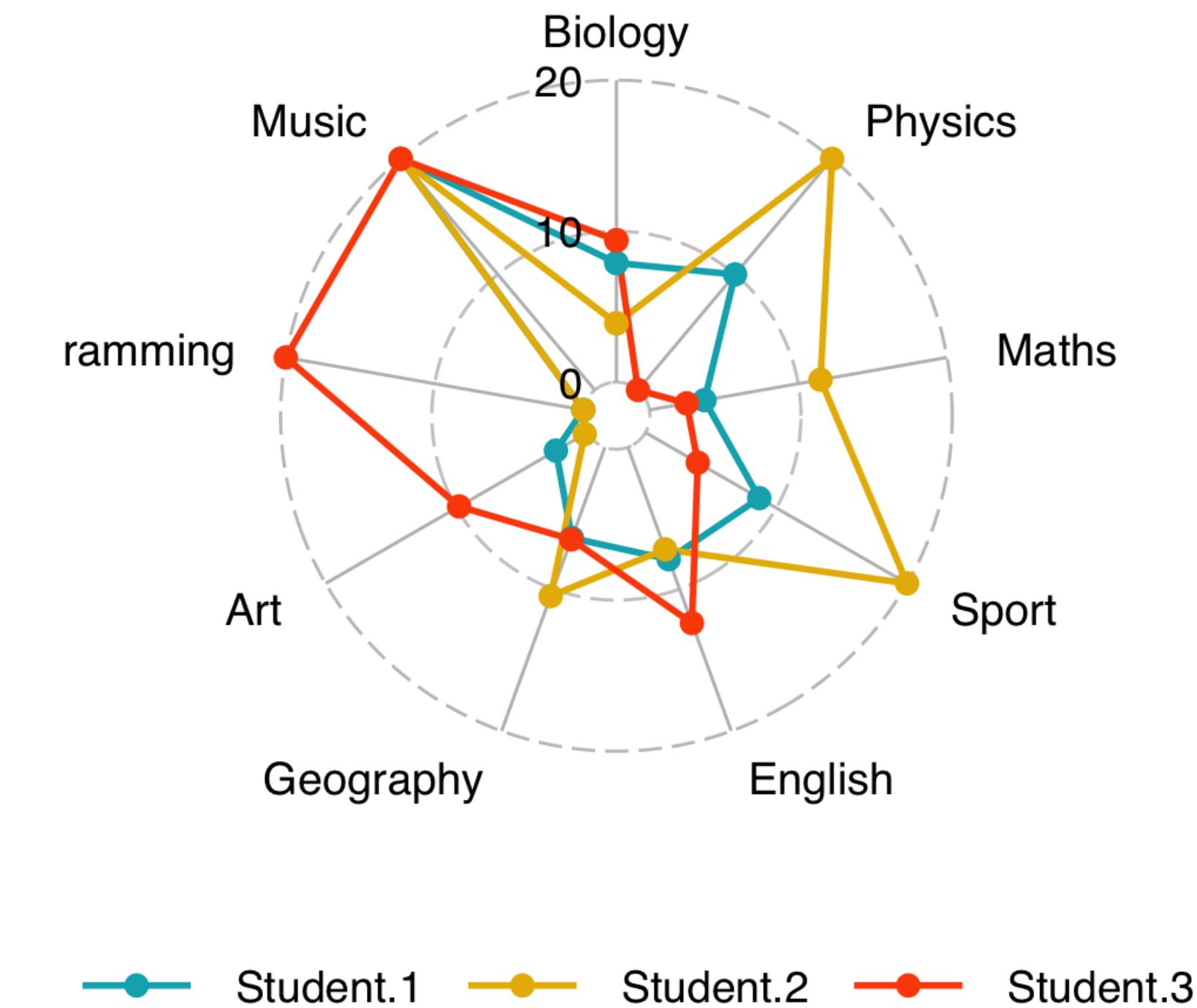
Idioms: radial bar chart, star plot

- 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



Idiom: radar plot

- radial line chart
 - point marks, radial layout
 - connecting line marks



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



Os sinal da bússola eleitoral

Disputa de 2010 foi parecida com a de 2006

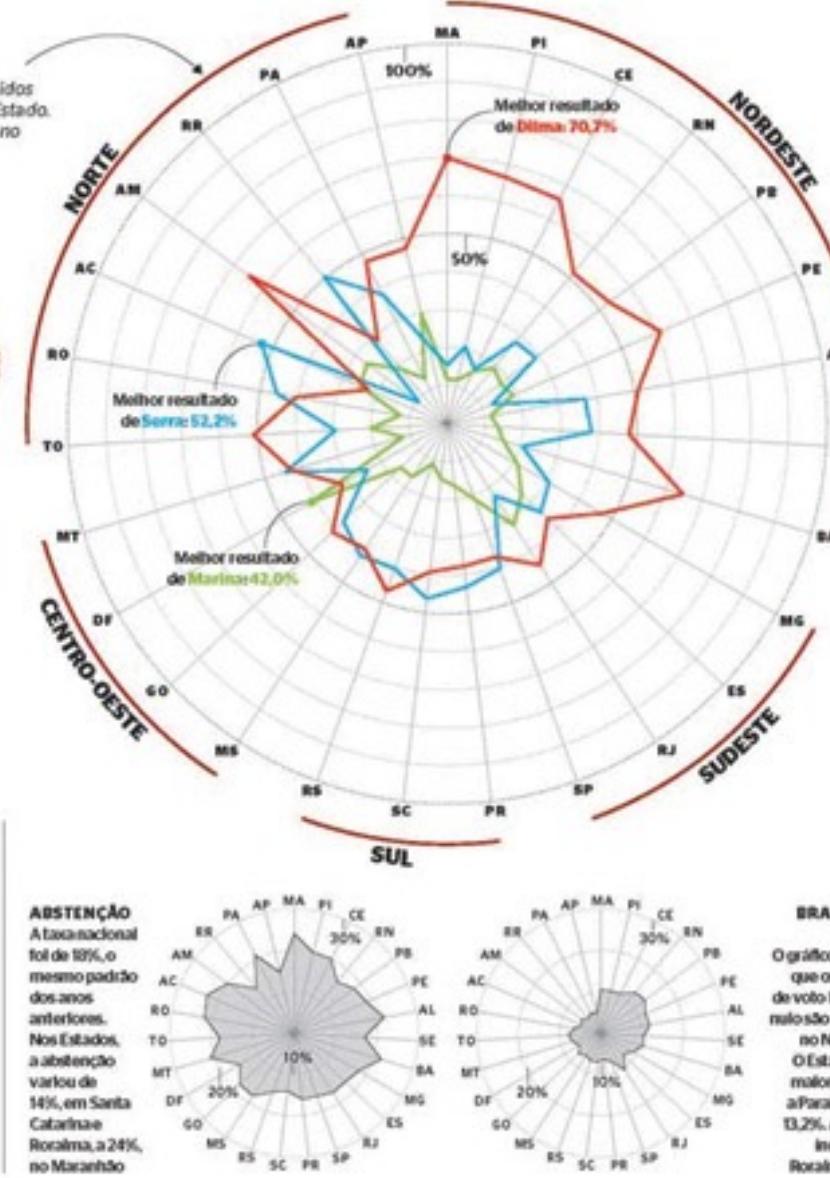
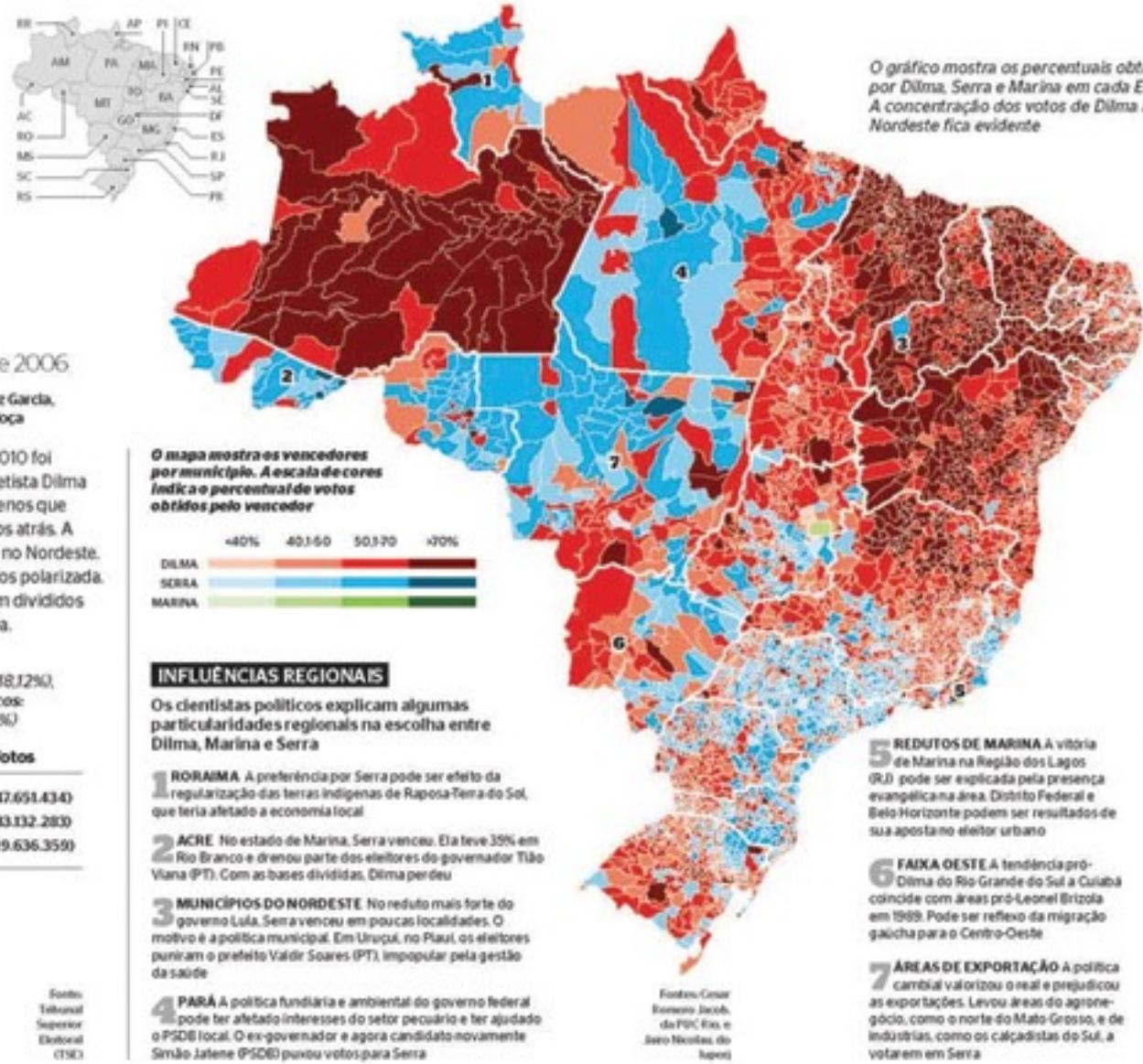
Alberto Cairo, Alexandre Mansur, Carlos Eduardo Cruz Garcia, Eliezer Barreira Junior, Marco Vergotti e Ricardo Mendoza

O PRIMEIRO TURNO da eleição presidencial de 2010 foi muito parecido com o da disputa de 2006. A petista Dilma Rousseff teve apenas 1,7 ponto percentual a menos que o índice obtido pelo presidente Lula quatro anos atrás. A concentração maior de seus votos também foi no Nordeste. Dessa vez, porém, a disputa foi um pouco menos polarizada. Os votos que provocaram segundo turno foram divididos entre o tucano José Serra e a verde Marina Silva.

Eleitores: 135.804.433; **abstêncio:** 24.610.296 (18,12%); **votos válidos:** 101.590.152 (91,36%); **votos brancos:** 3.479.340 (3,13%) e **votos nulos:** 6.124.254 (5,51%).

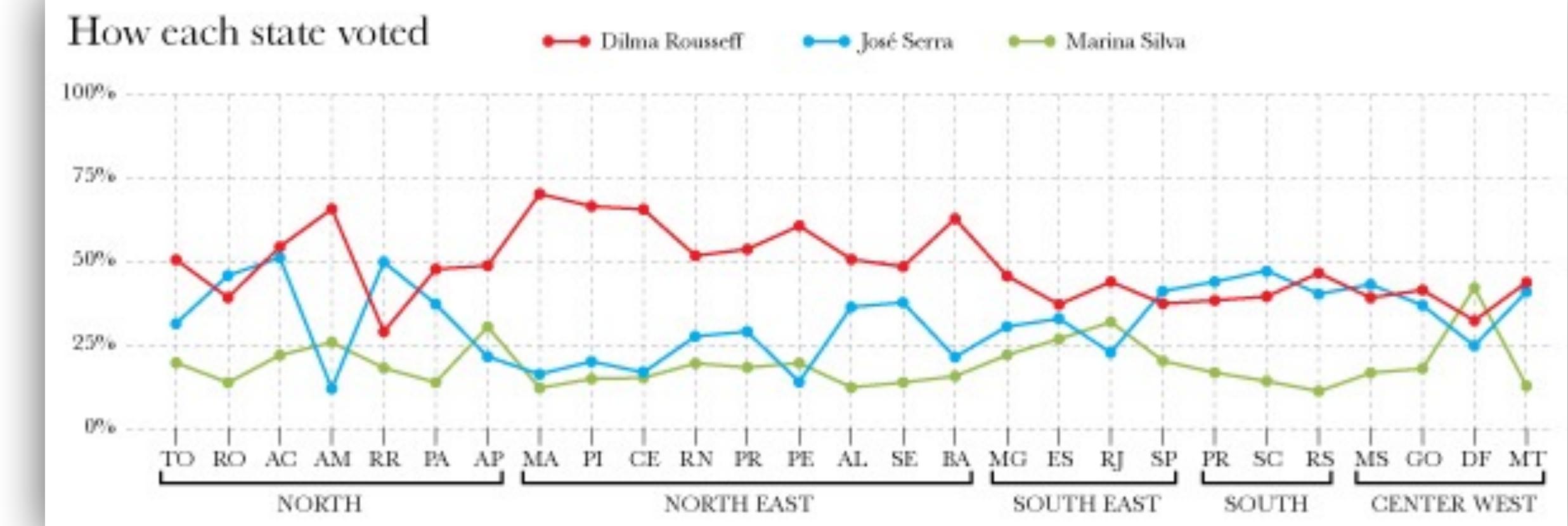
| Candidatos | % | Votos |
|---------------------|-------|--------------|
| Dilma Rousseff (PT) | 46,9% | (47.651.433) |
| José Serra (PSDB) | 32,6% | (33.132.283) |
| Marina Silva (PV) | 19,3% | (19.636.359) |

| Outros candidatos | % | Votos |
|--------------------------|-------|----------|
| Piñio (PSOL) | 0,87% | (86.850) |
| José Maria Eymael (PSDC) | 0,69% | (69.359) |
| Zé Maria (PSTB) | 0,68% | (68.609) |
| Levy Fidelix (PRTB) | 0,66% | (67.960) |
| Ivan Pinheiro (PRB) | 0,49% | (49.136) |
| Rui Costa Pimenta (PDC) | 0,01% | (12.206) |



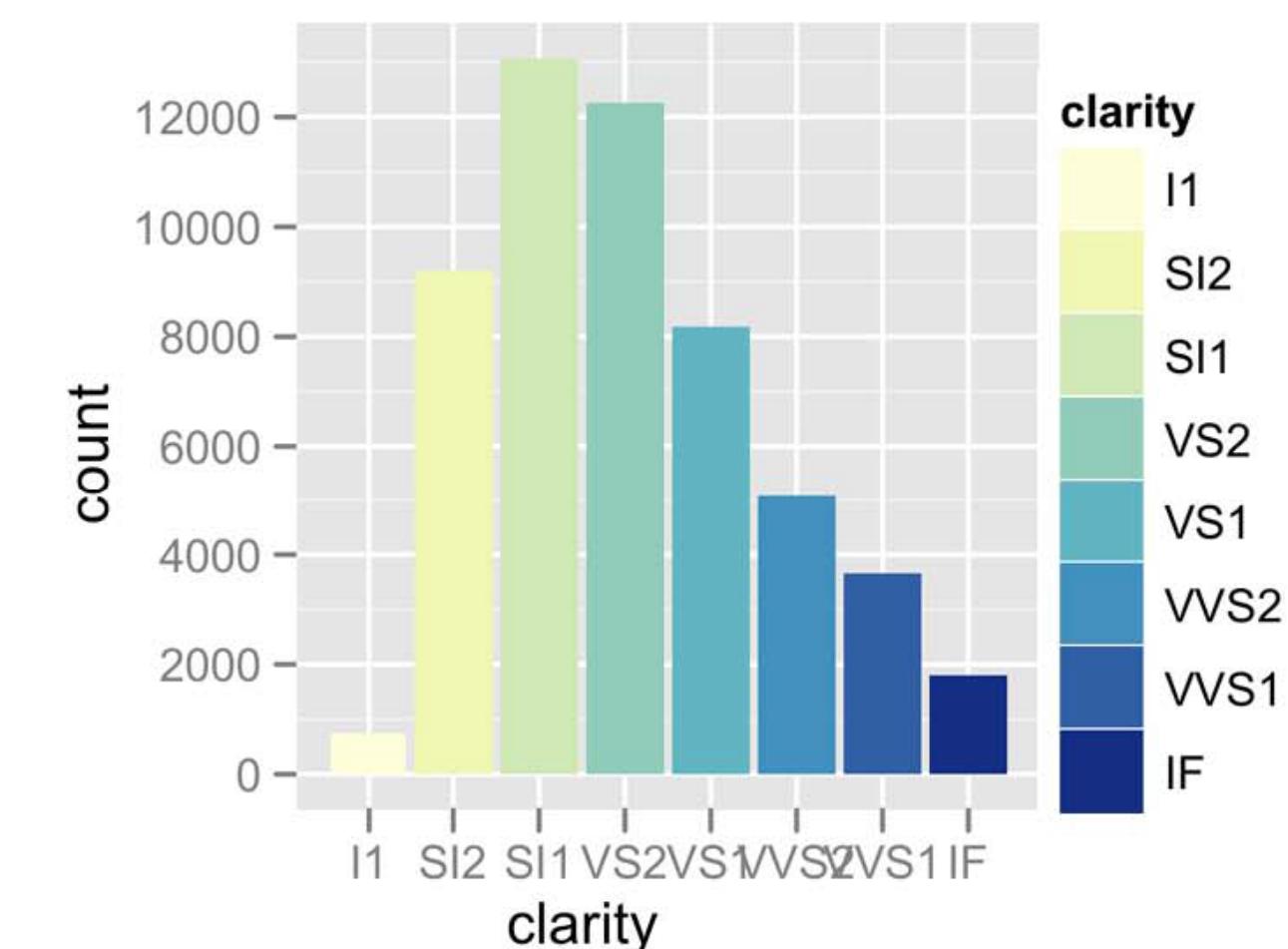
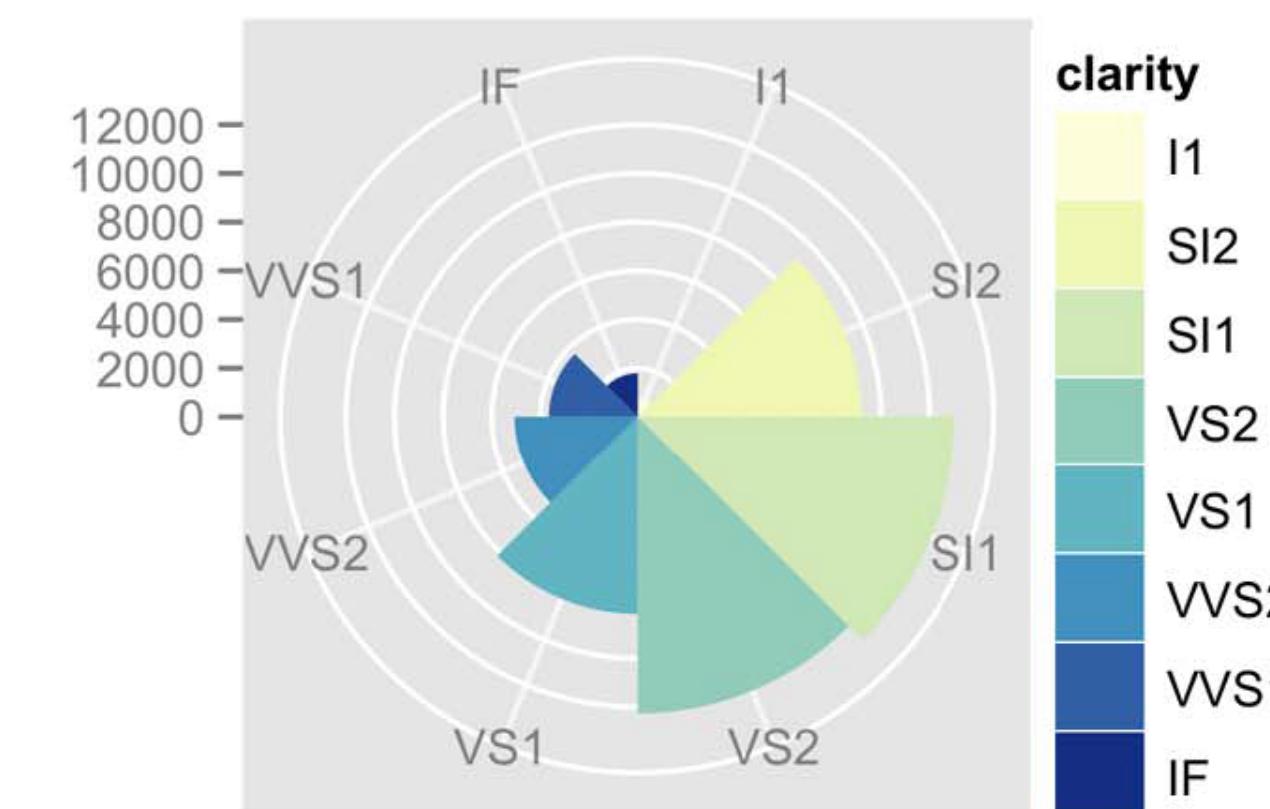
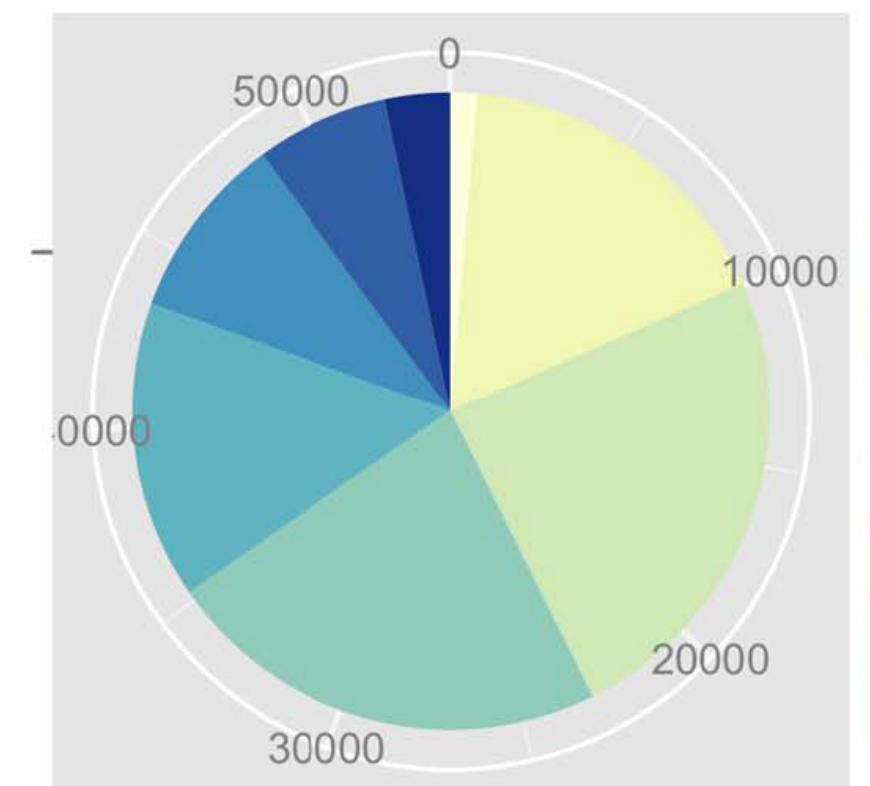
original
difficult to interpret

redesign for
rectilinear



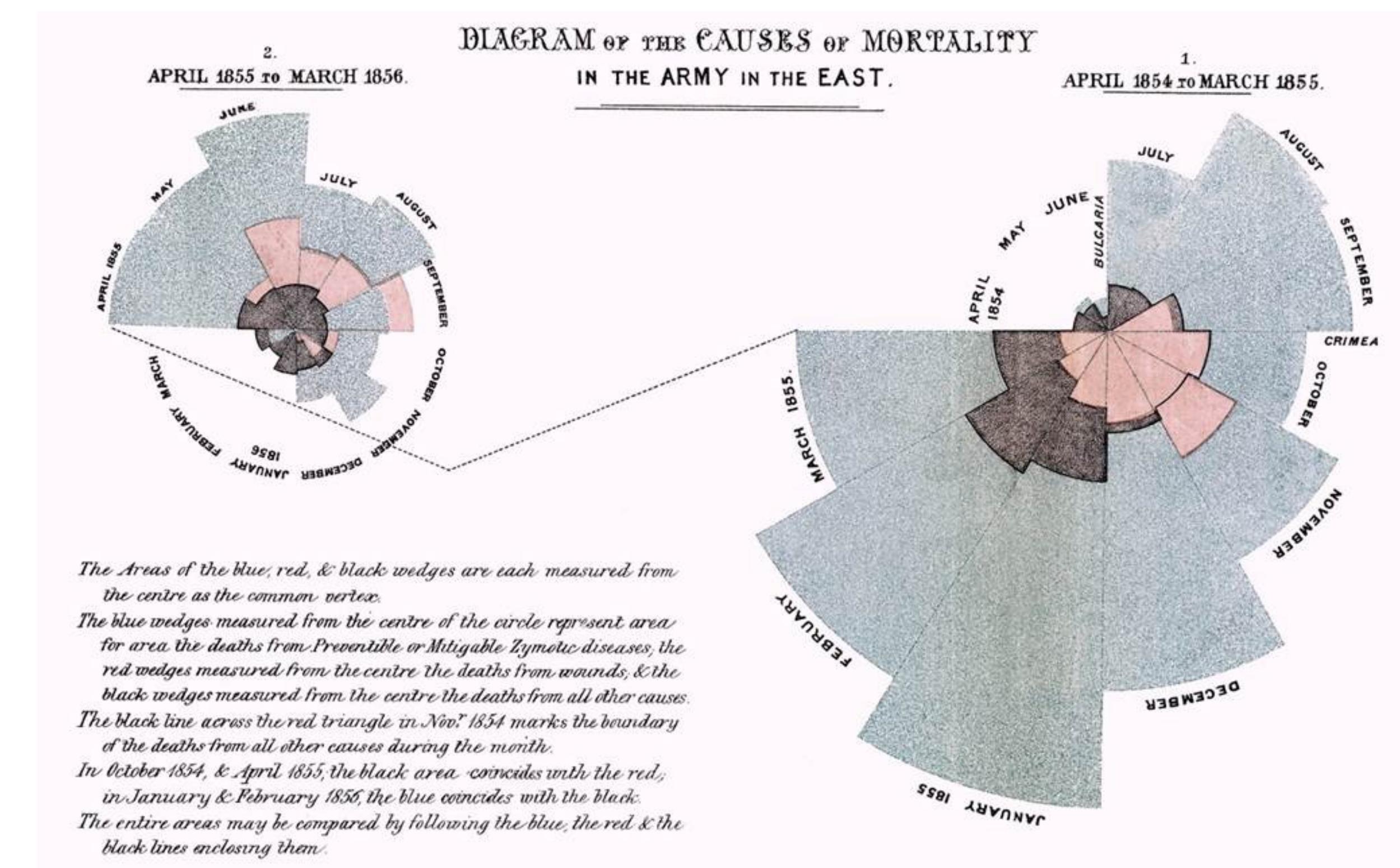
Idioms: pie chart, coxcomb chart

- 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



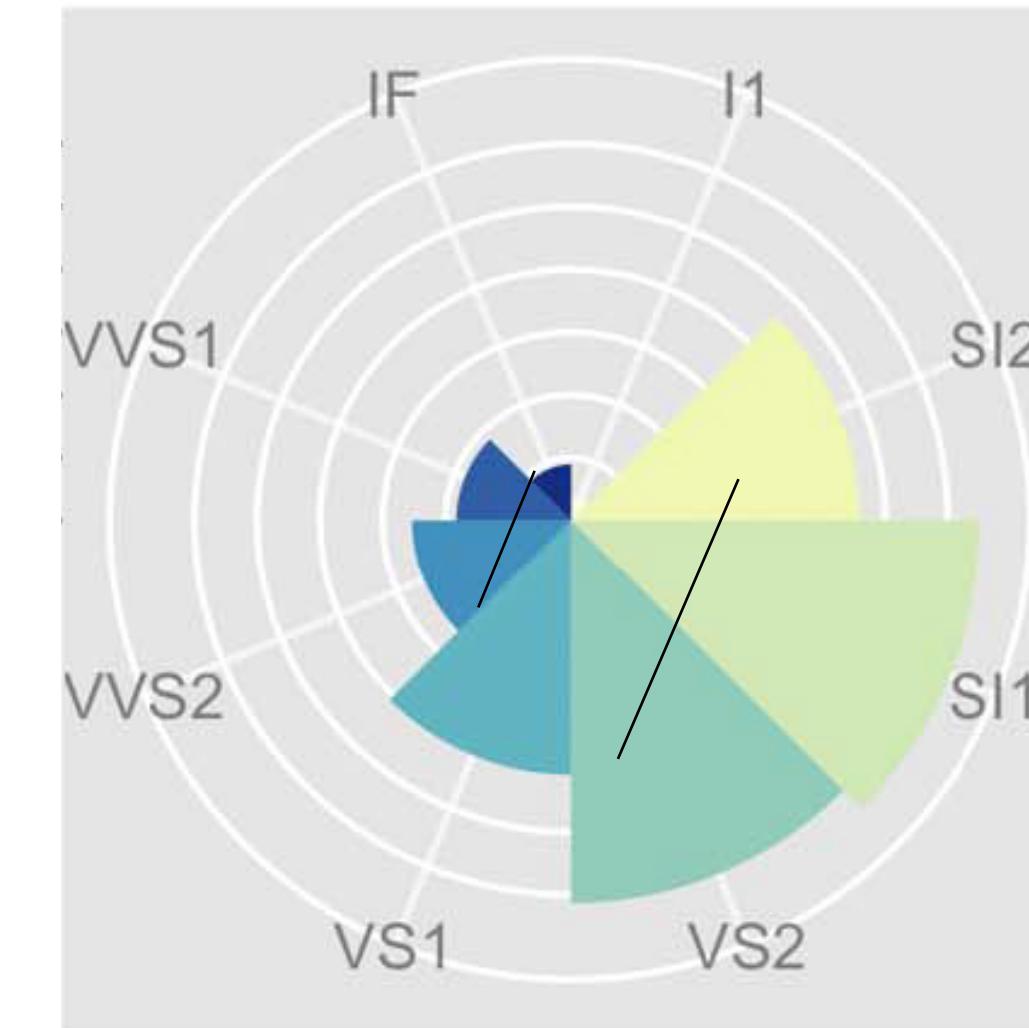
Coxcomb / rose / polar area chart

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

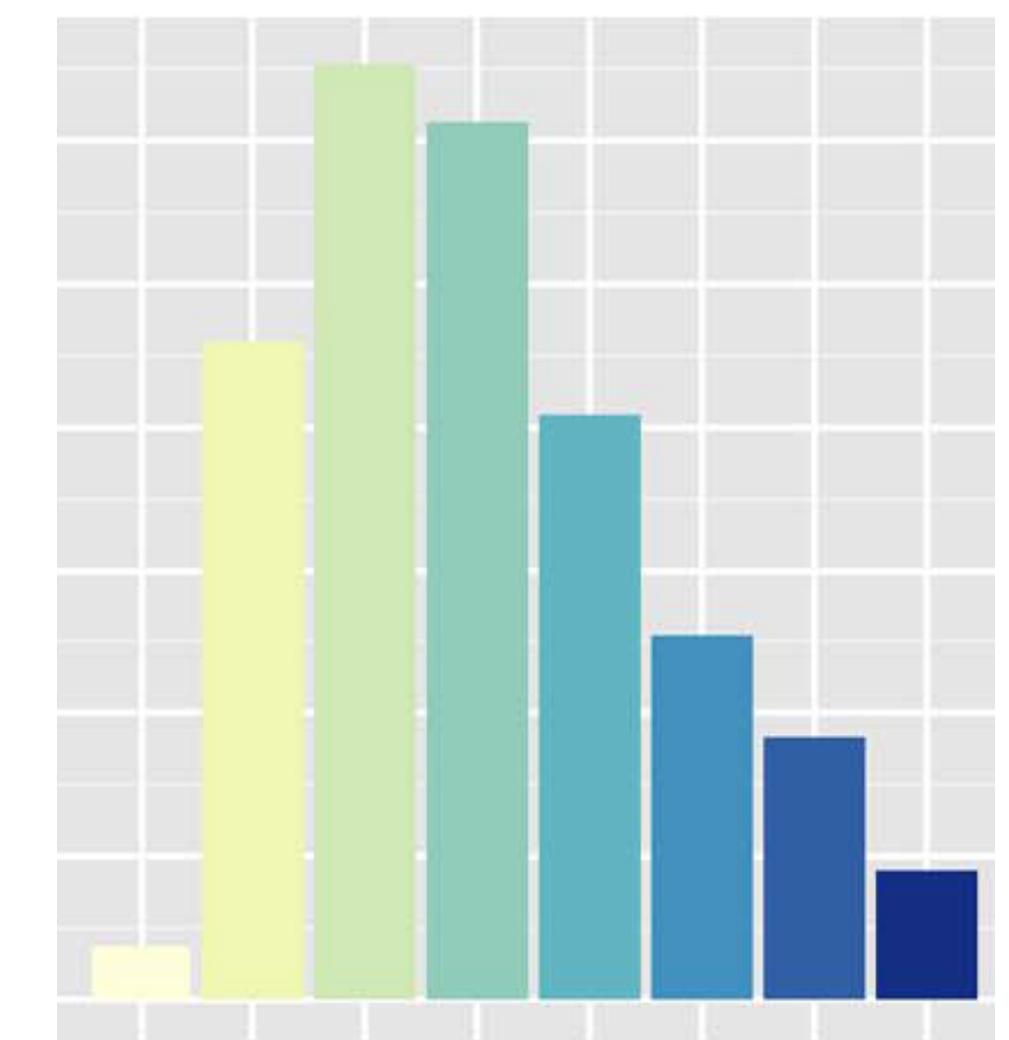


Coxcomb: perception

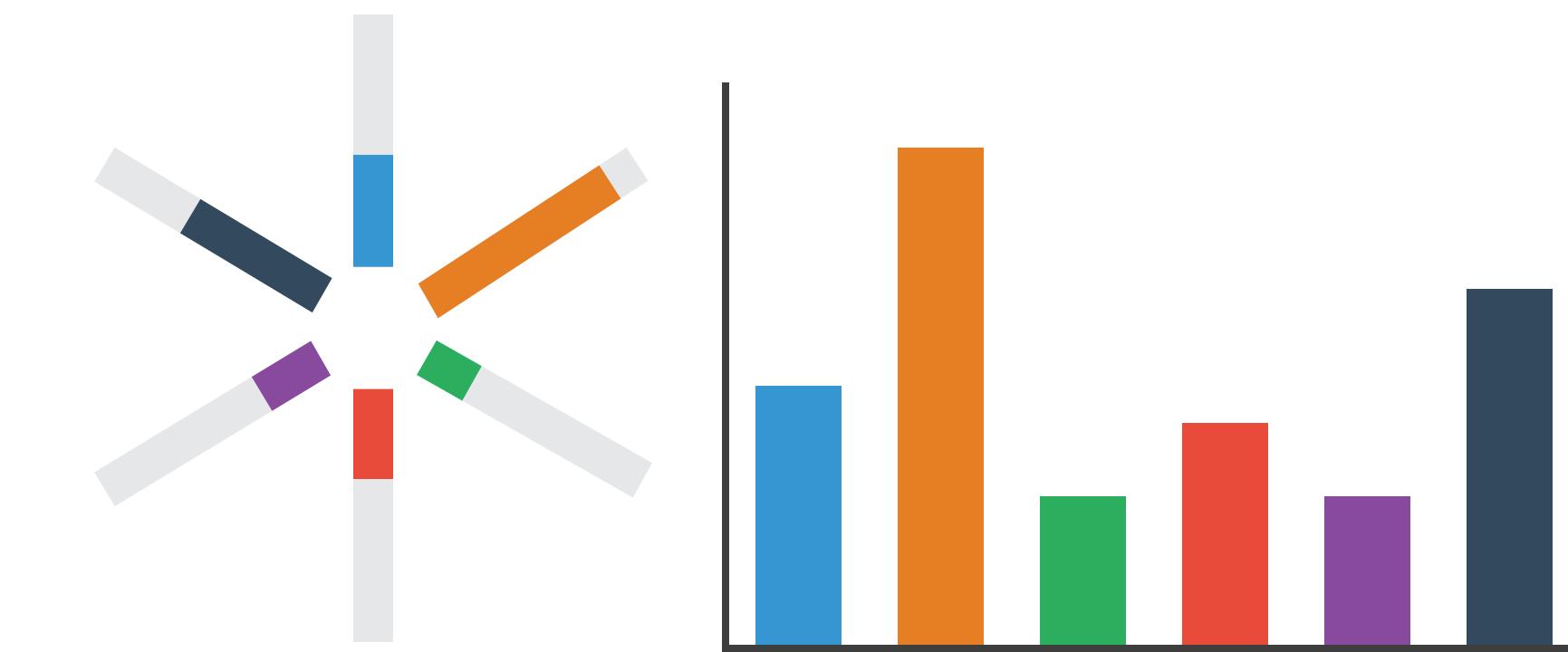
- encode: 1D length
- decode/perceive: 2D area
- nonuniform line/sector width as length increases
 - so area variation is nonlinear wrt line mark length!
- bar chart safer: uniform width, so area is linear with line mark length
 - both radial & rectilinear cases



nonuniform width as
length increases



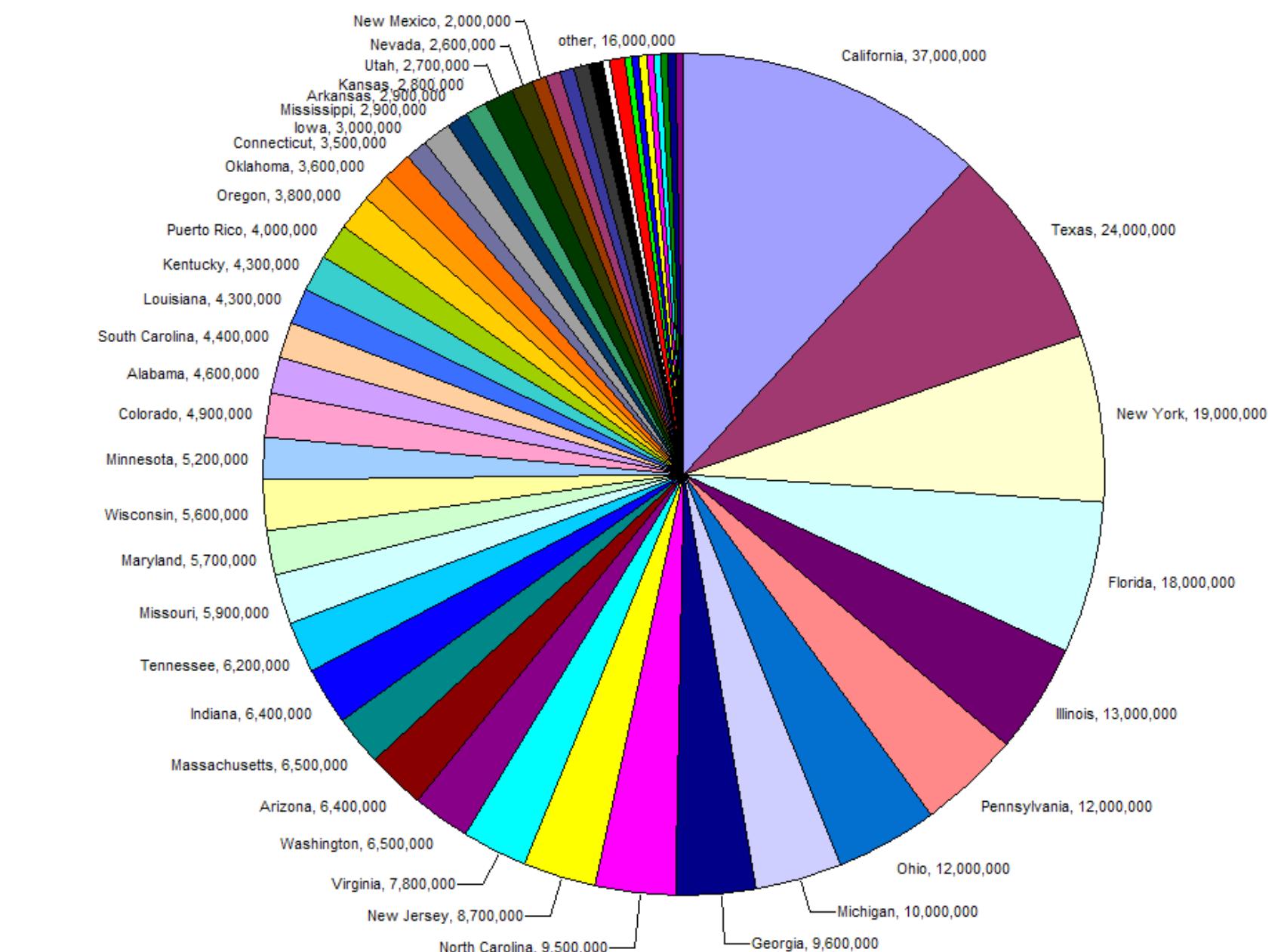
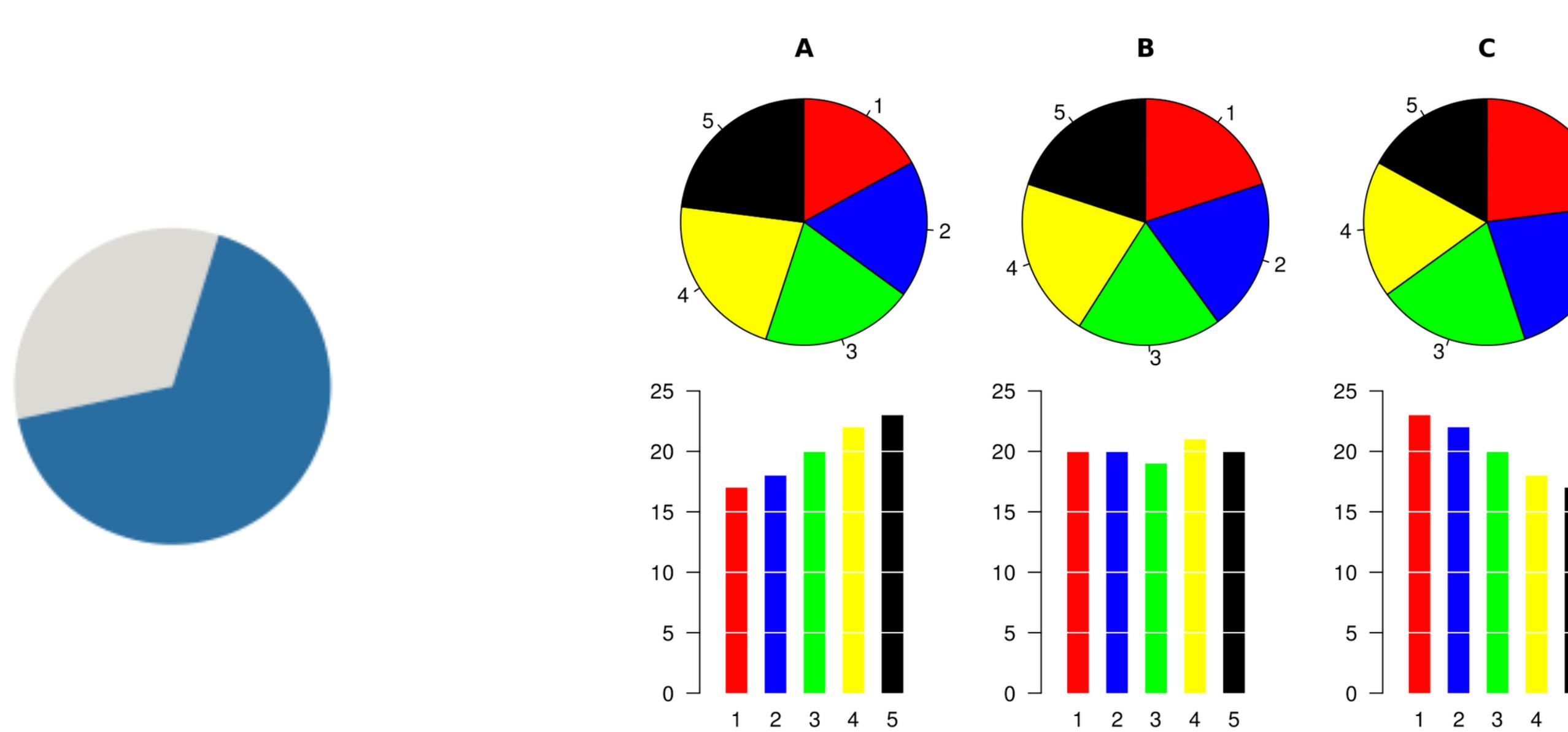
uniform width as
length increases



radial & rectilinear bars: uniform width as
length increases

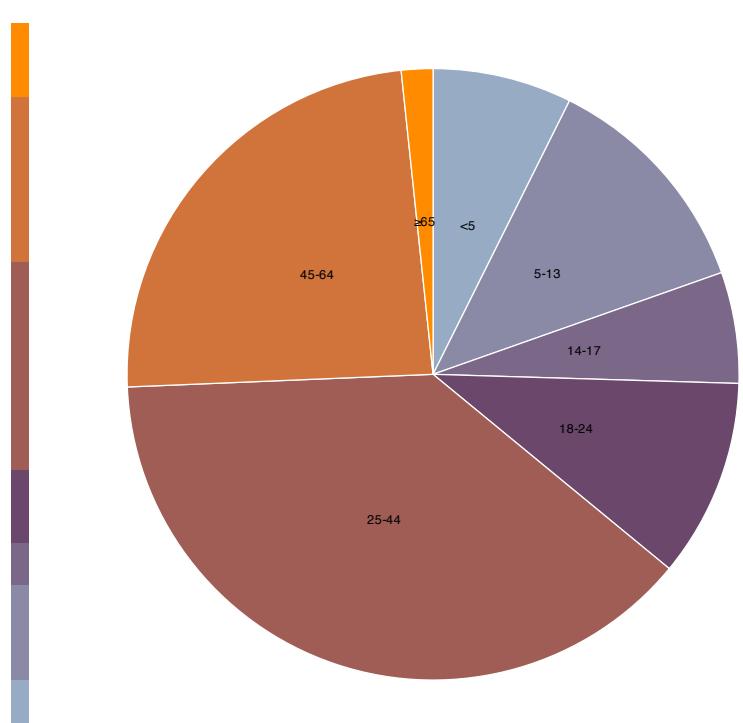
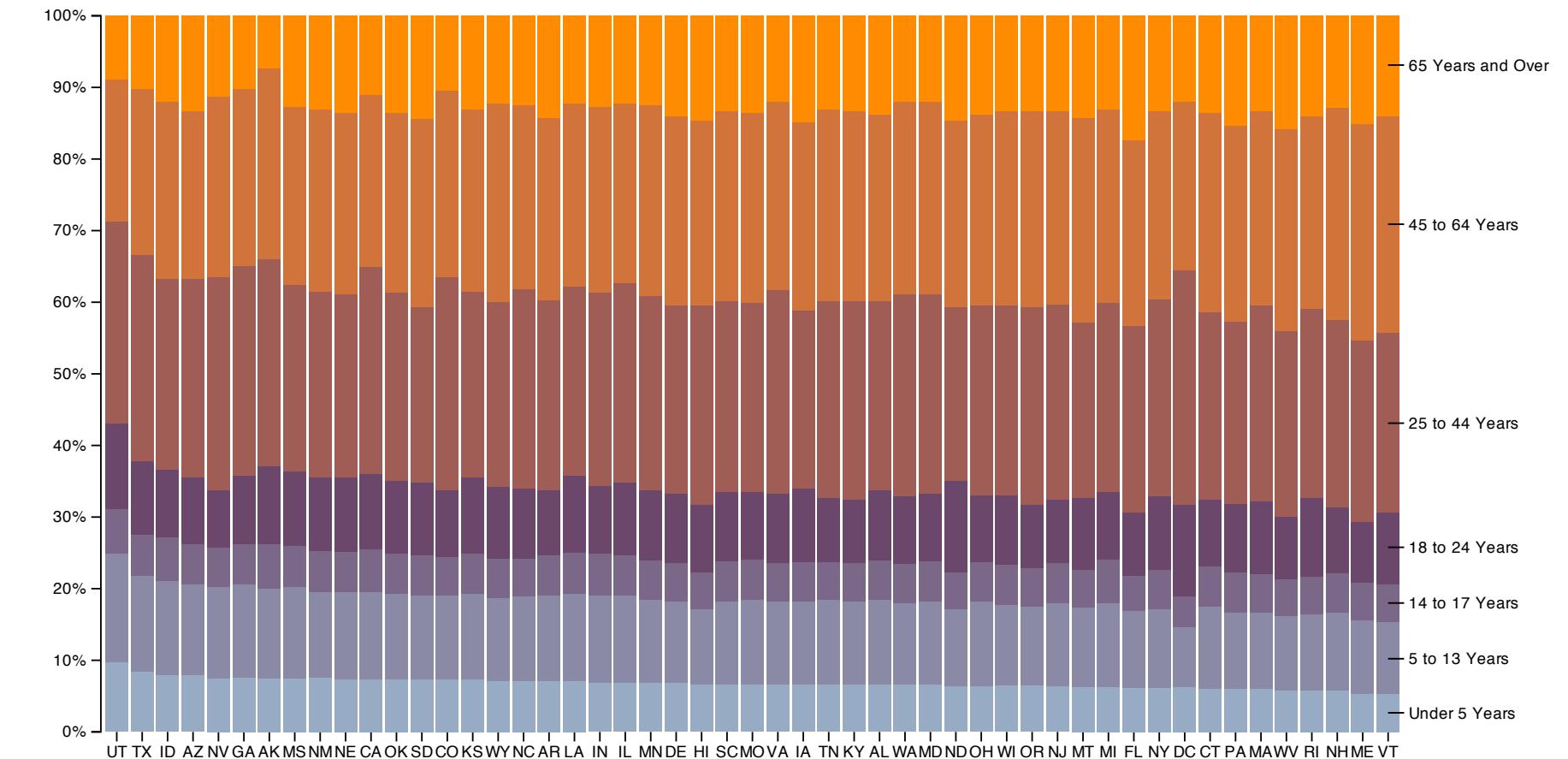
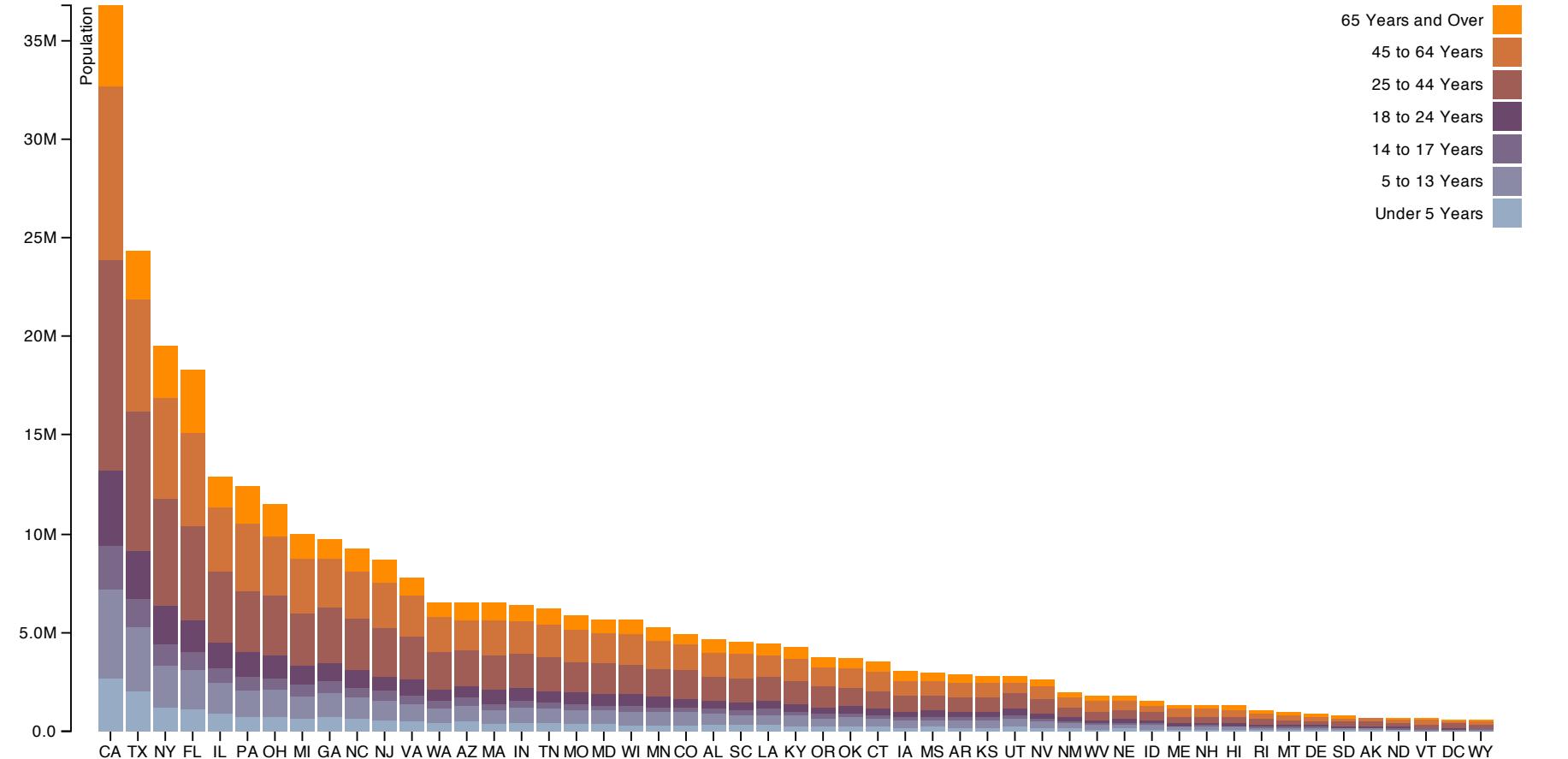
Pie charts: best practices

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



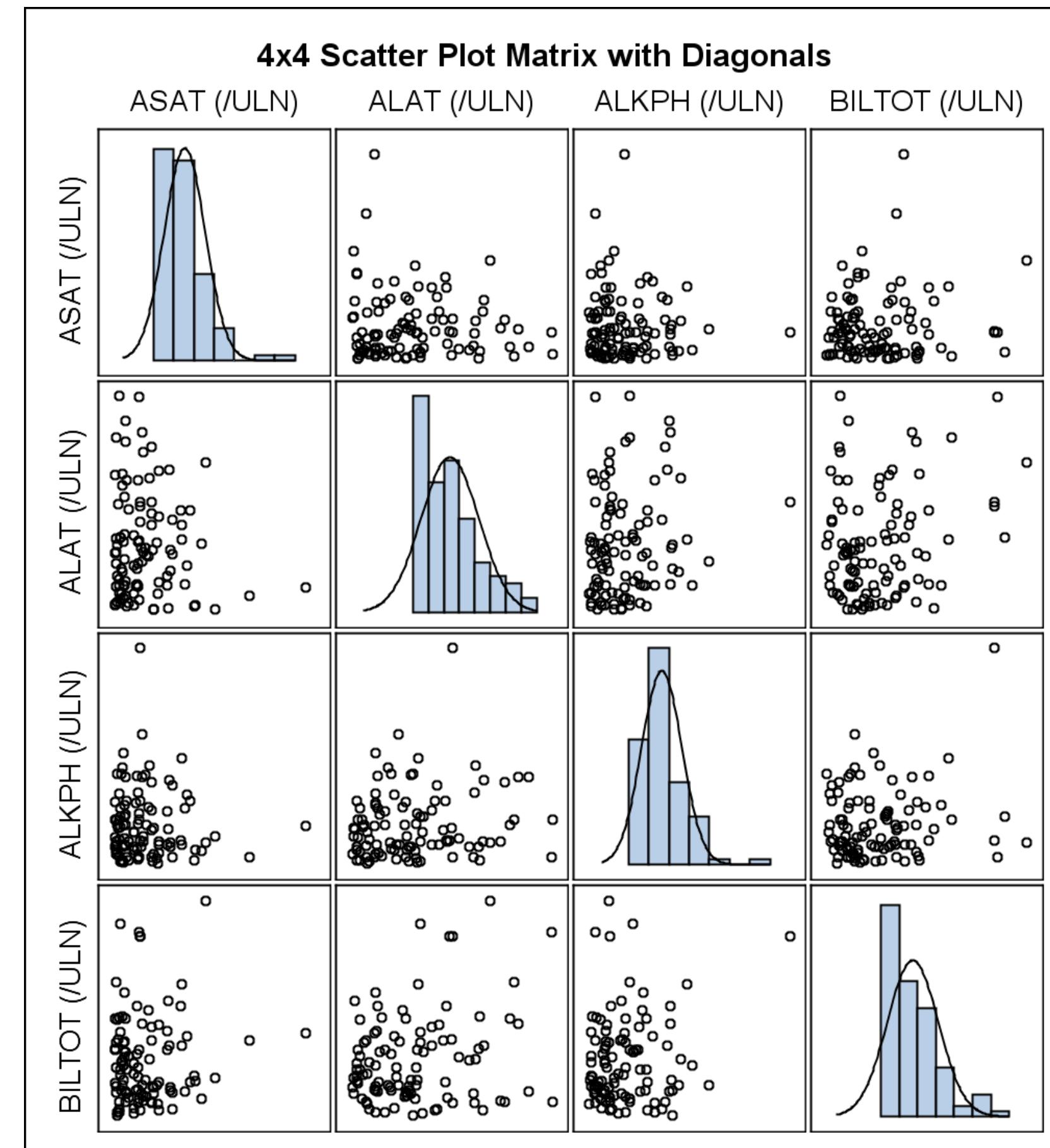
Idioms: normalized stacked bar chart

- 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



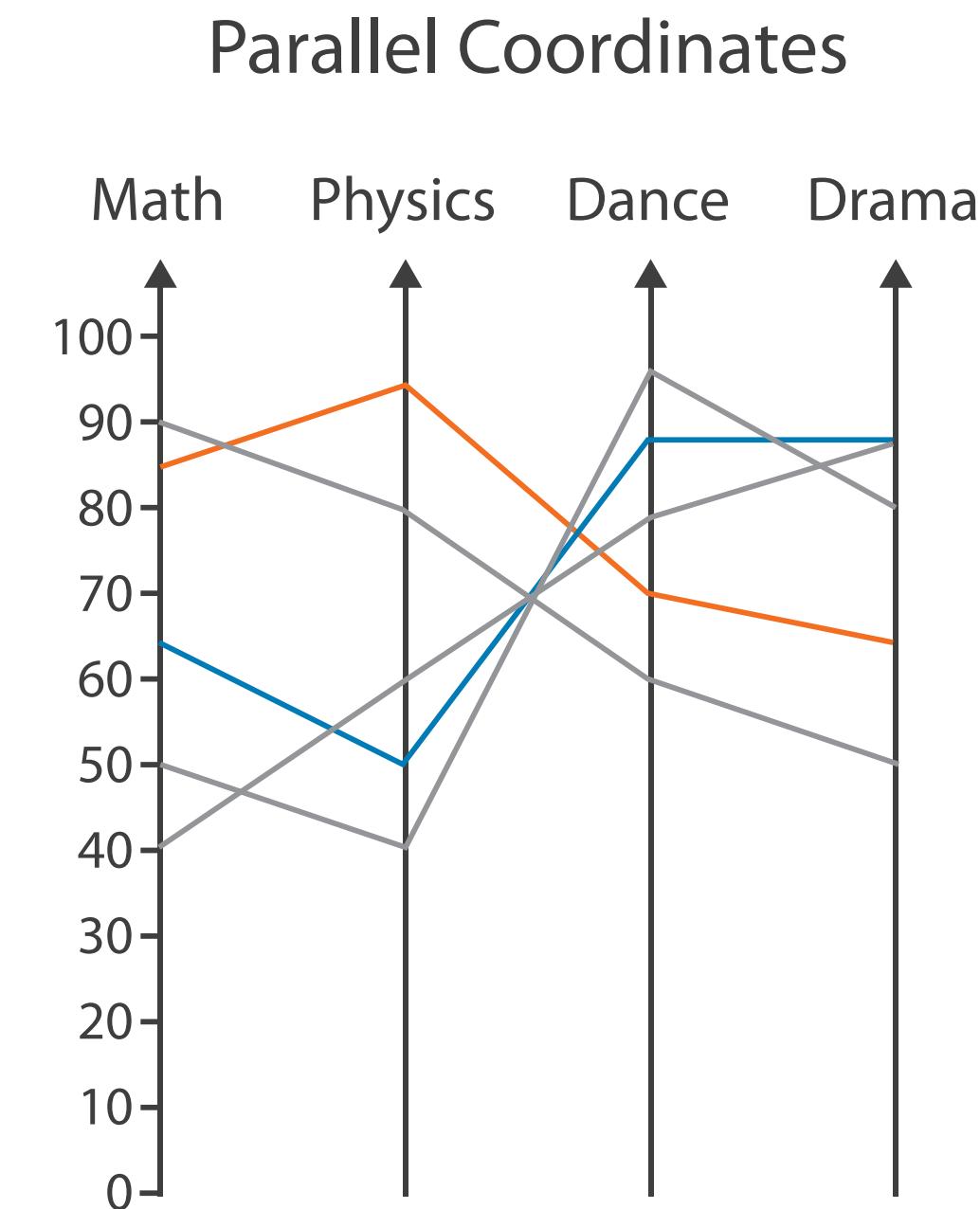
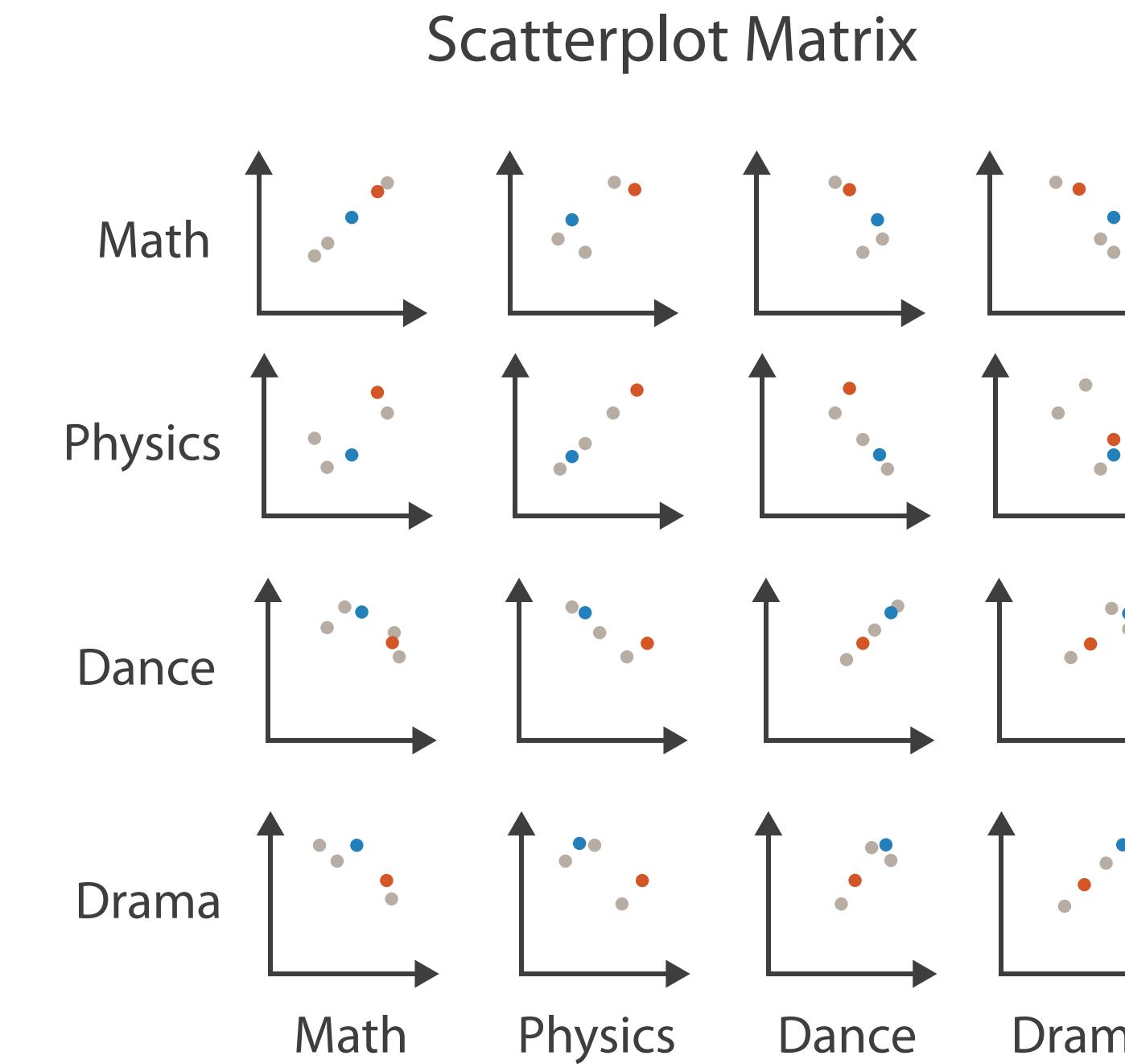
Idiom: SPLOM

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



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 attrs
 - hundreds of items



Table

| | Math | Physics | Dance | Drama |
|---|------|---------|-------|-------|
| 1 | 85 | 95 | 70 | 65 |
| 2 | 90 | 80 | 60 | 50 |
| 3 | 65 | 50 | 90 | 90 |
| 4 | 50 | 40 | 95 | 80 |
| 5 | 40 | 60 | 80 | 90 |

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

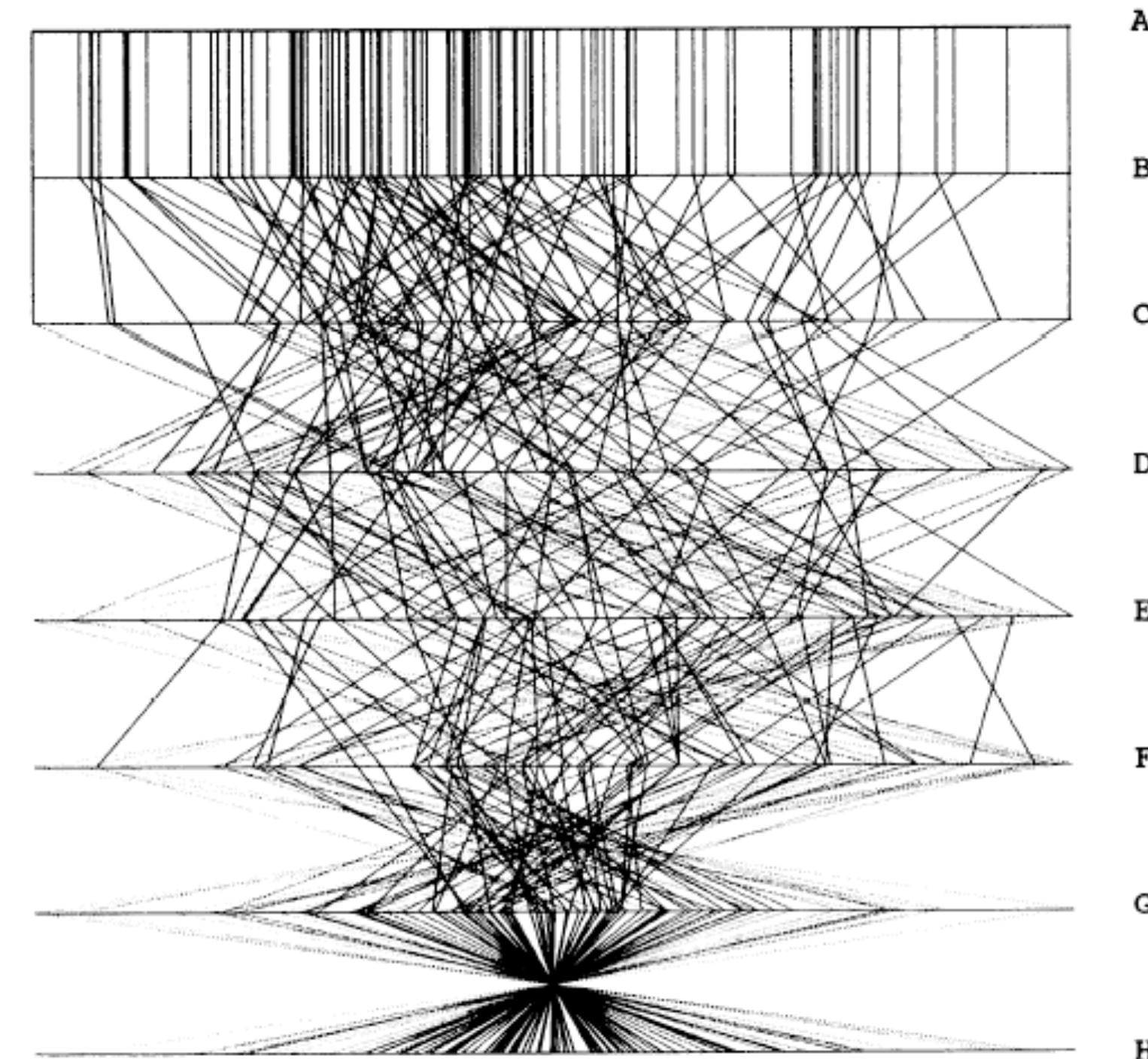
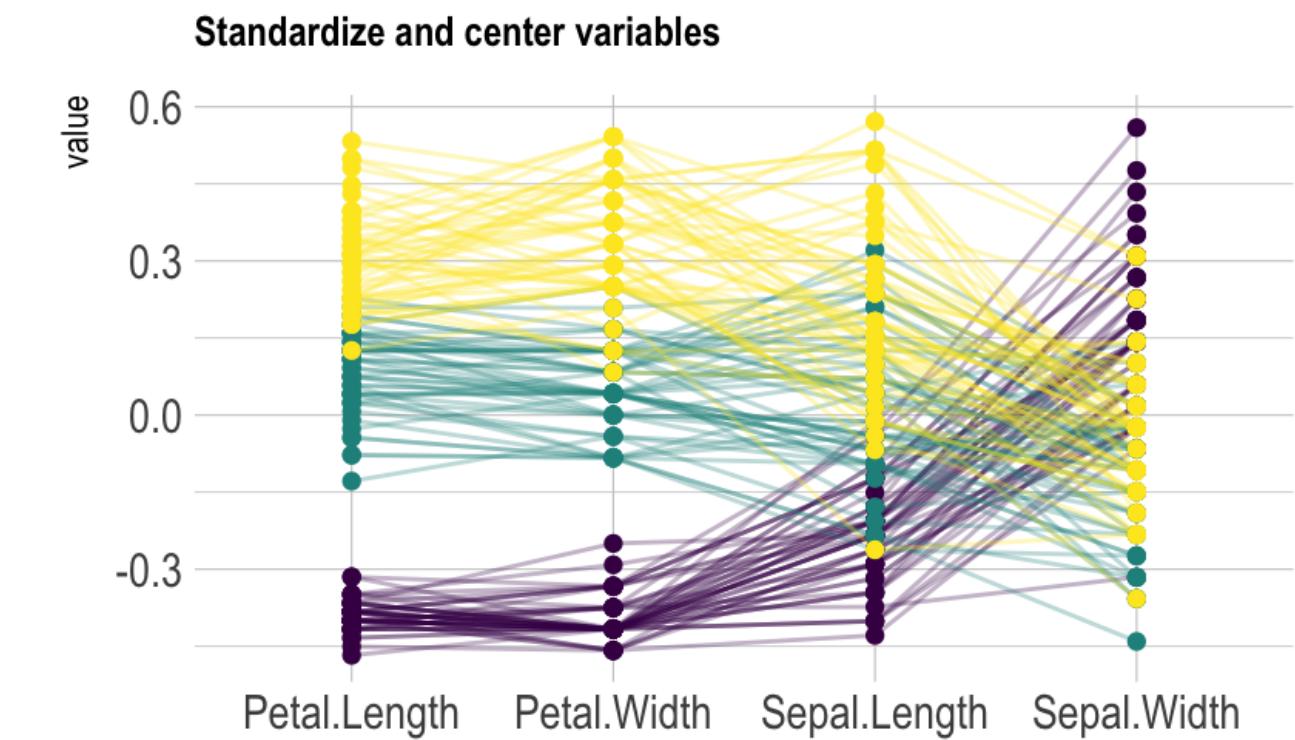
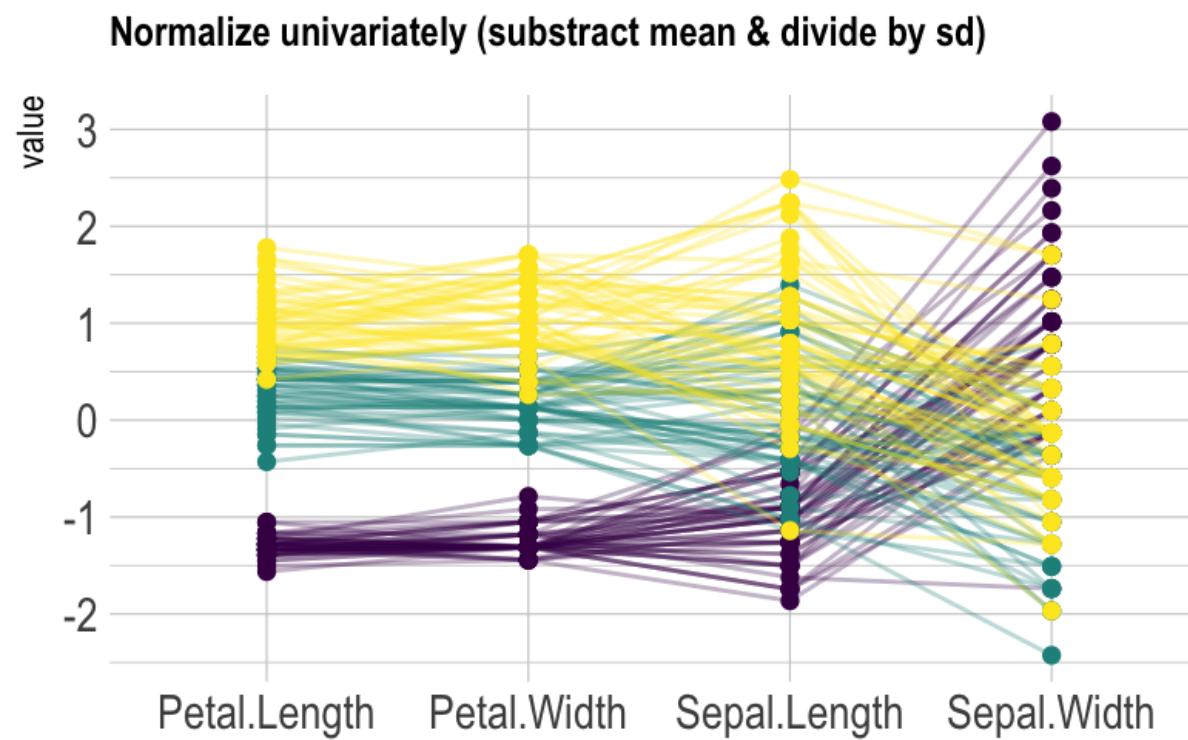
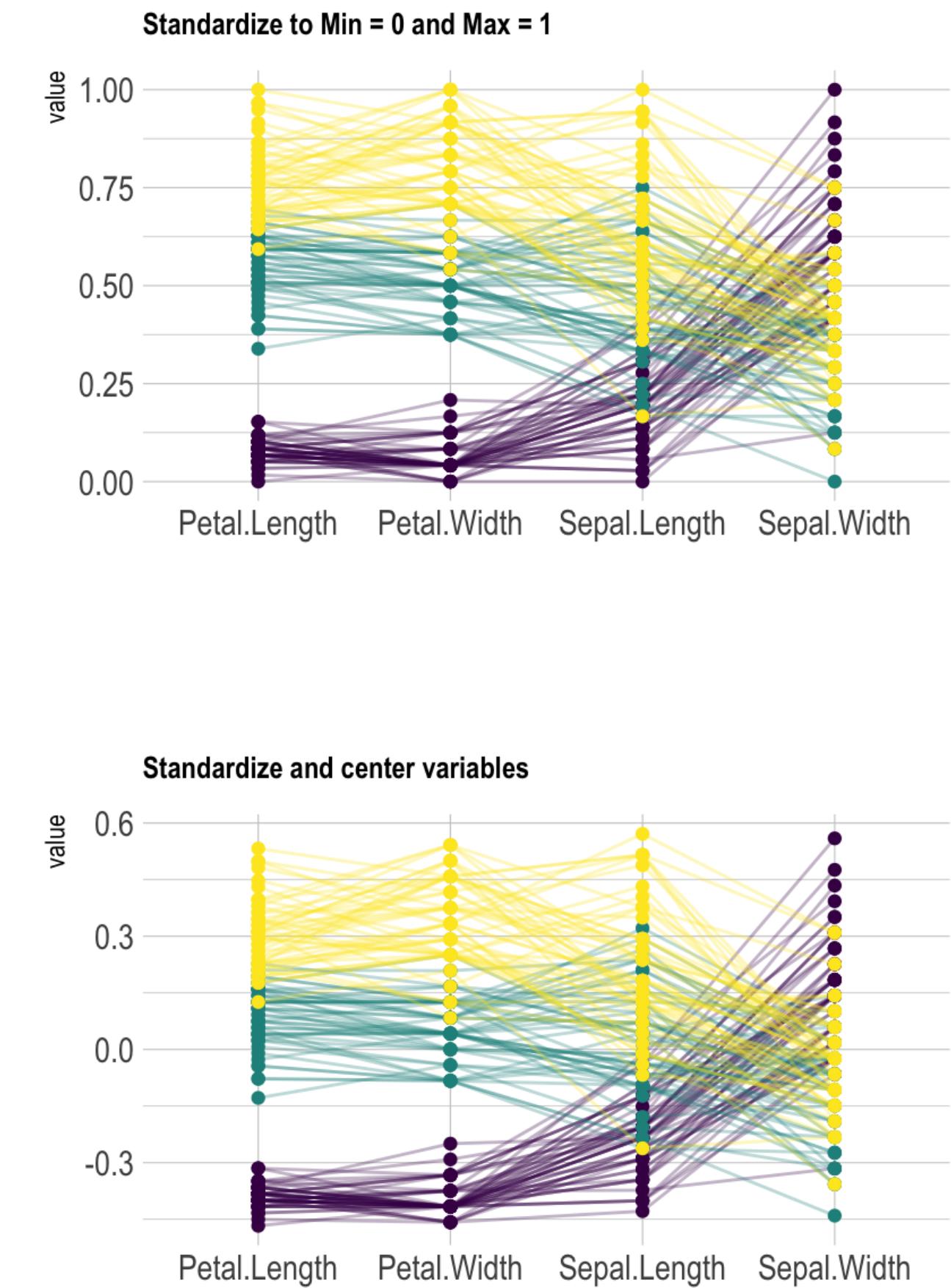
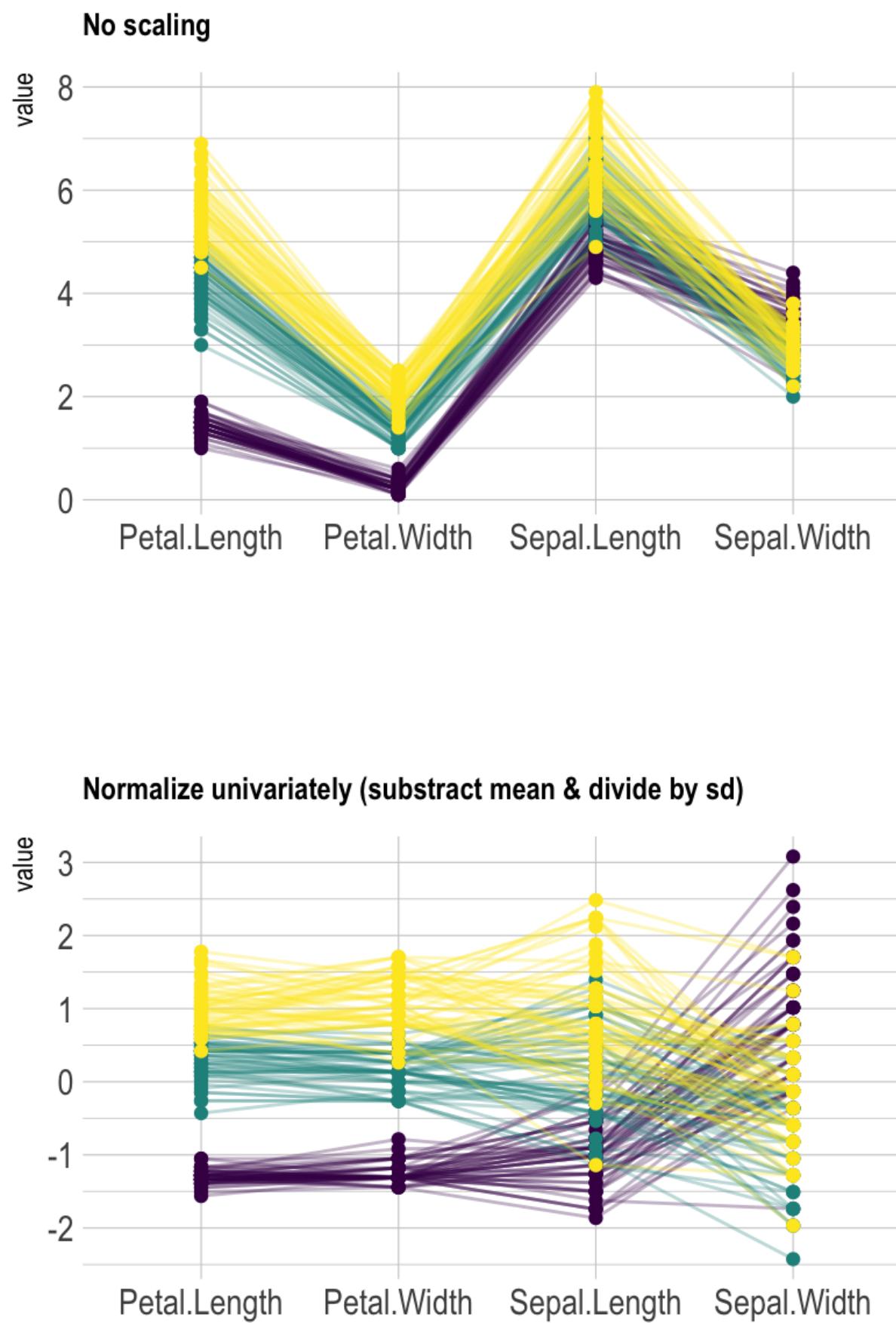


Figure 3. Parallel Coordinate Plot of Six-Dimensional Data Illustrating Correlations of $\rho = 1, .8, .2, 0, -.2, -.8, \text{ and } -1$.

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

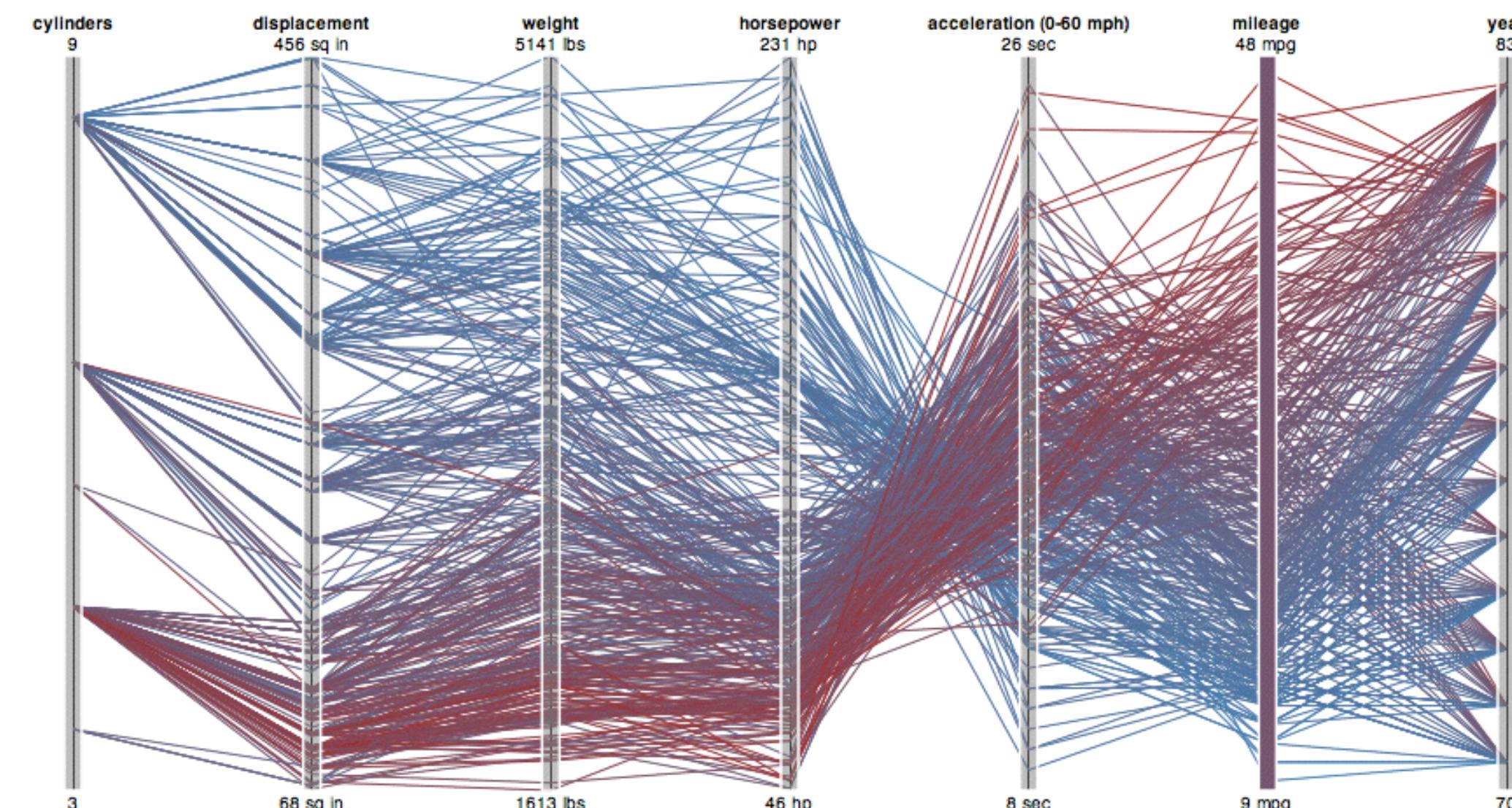
Task: Correlation

- 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



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

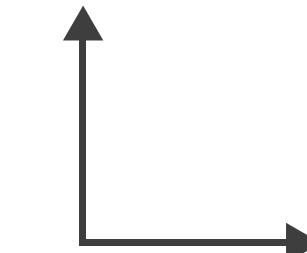


Orientation limitations

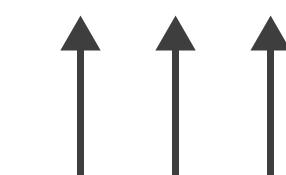
- rectilinear: scalability wrt #axes
 - 2 axes best, 3 problematic, 4+ impossible
- parallel: unfamiliarity, training time
- radial: perceptual limits
 - polar coordinate asymmetry
 - angles lower precision than length
 - nonuniform sector width/size depending on radial distance
 - frequently problematic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

④ Axis Orientation

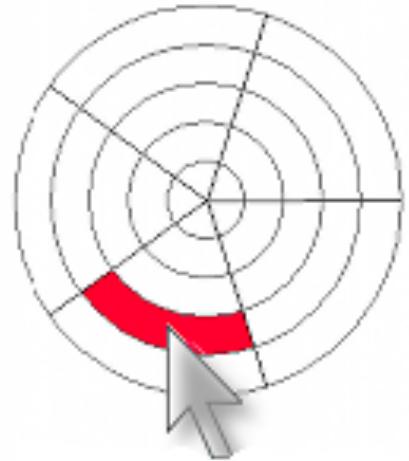
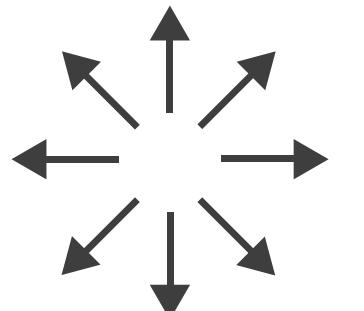
→ Rectilinear



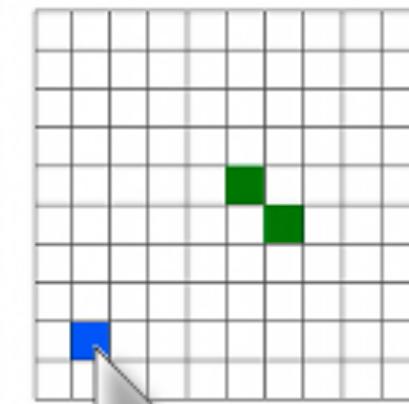
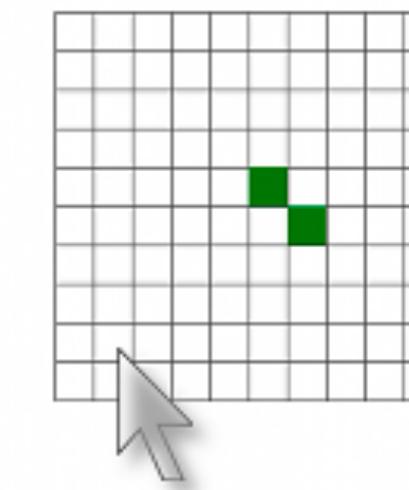
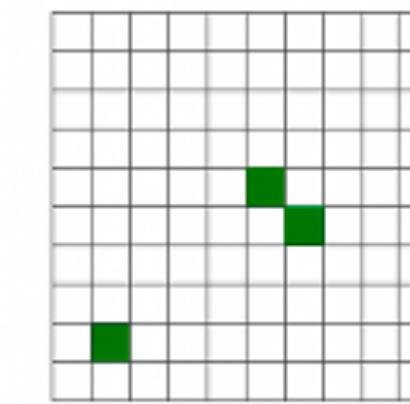
→ Parallel



→ Radial



clicked at wrong cell

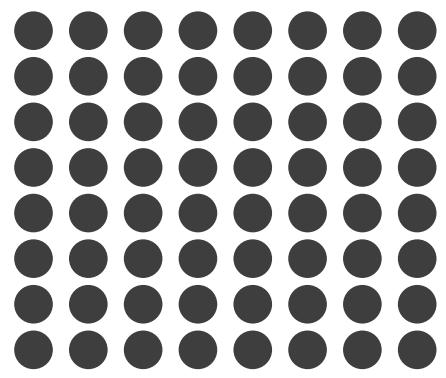


clicked at correct cell

Layout density

➔ Layout Density

→ Dense



→ Space-Filling

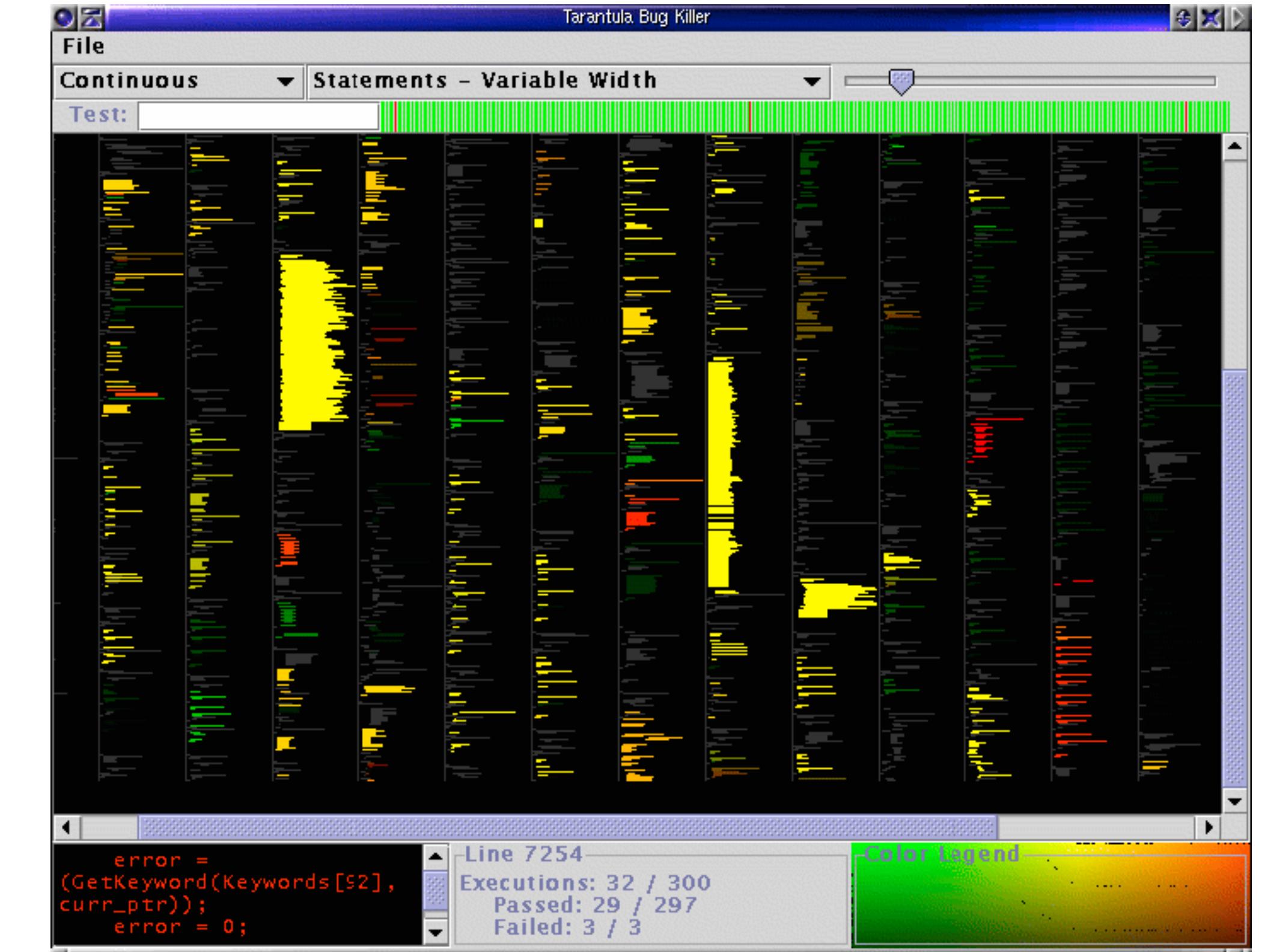
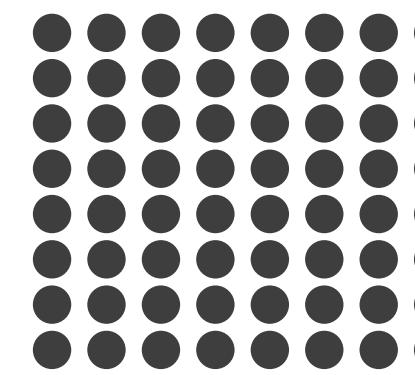


Idiom: Dense software overviews

- 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

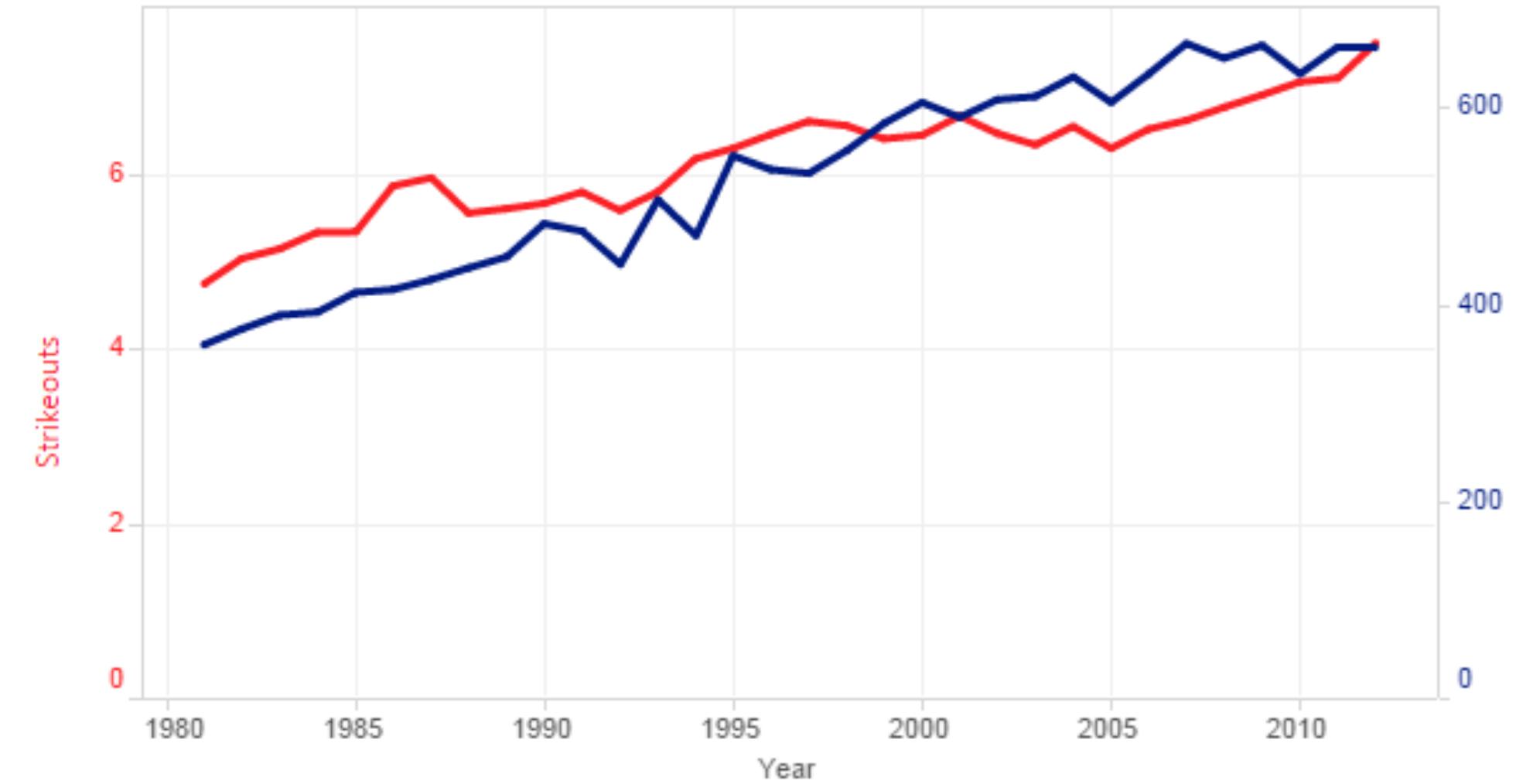
→ Layout Density

→ Dense



Idiom: dual-axis line charts

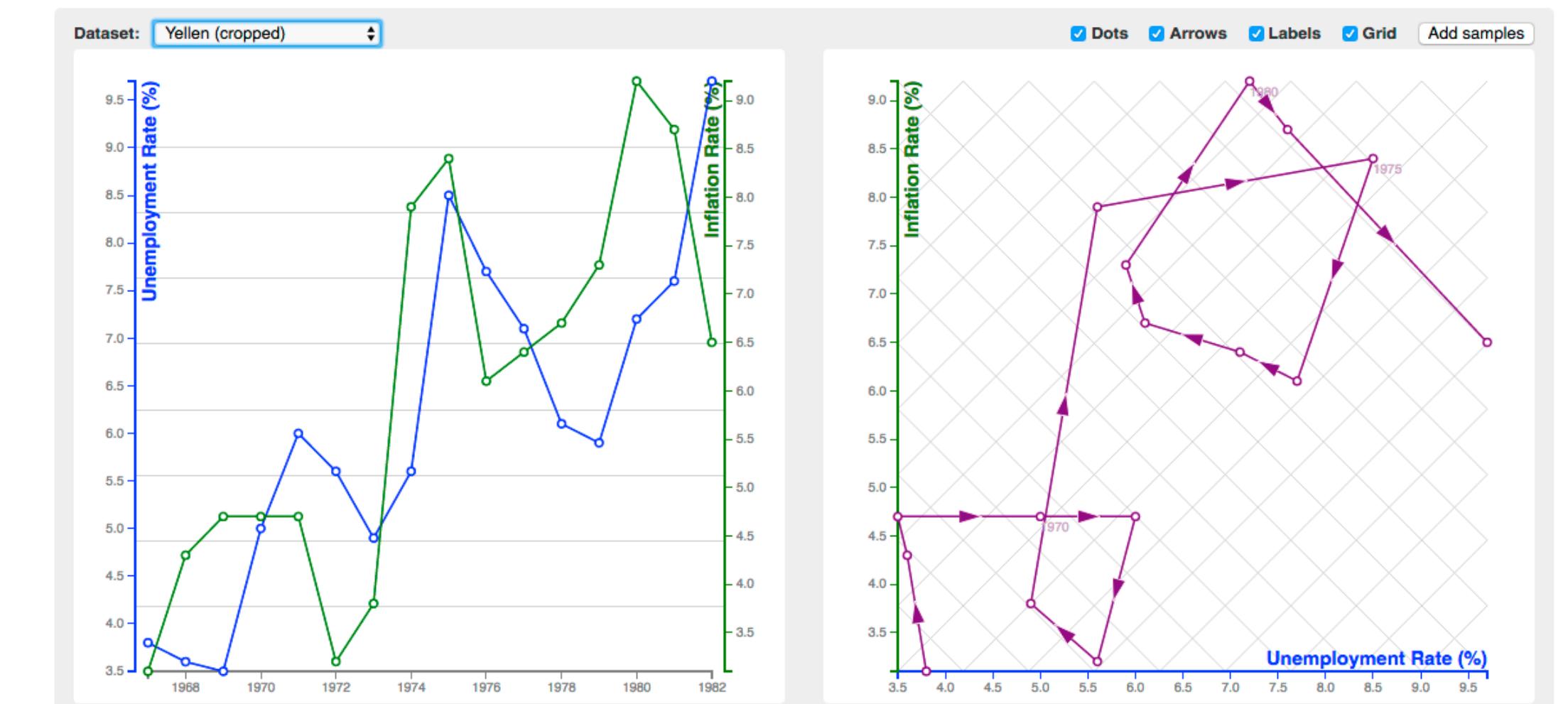
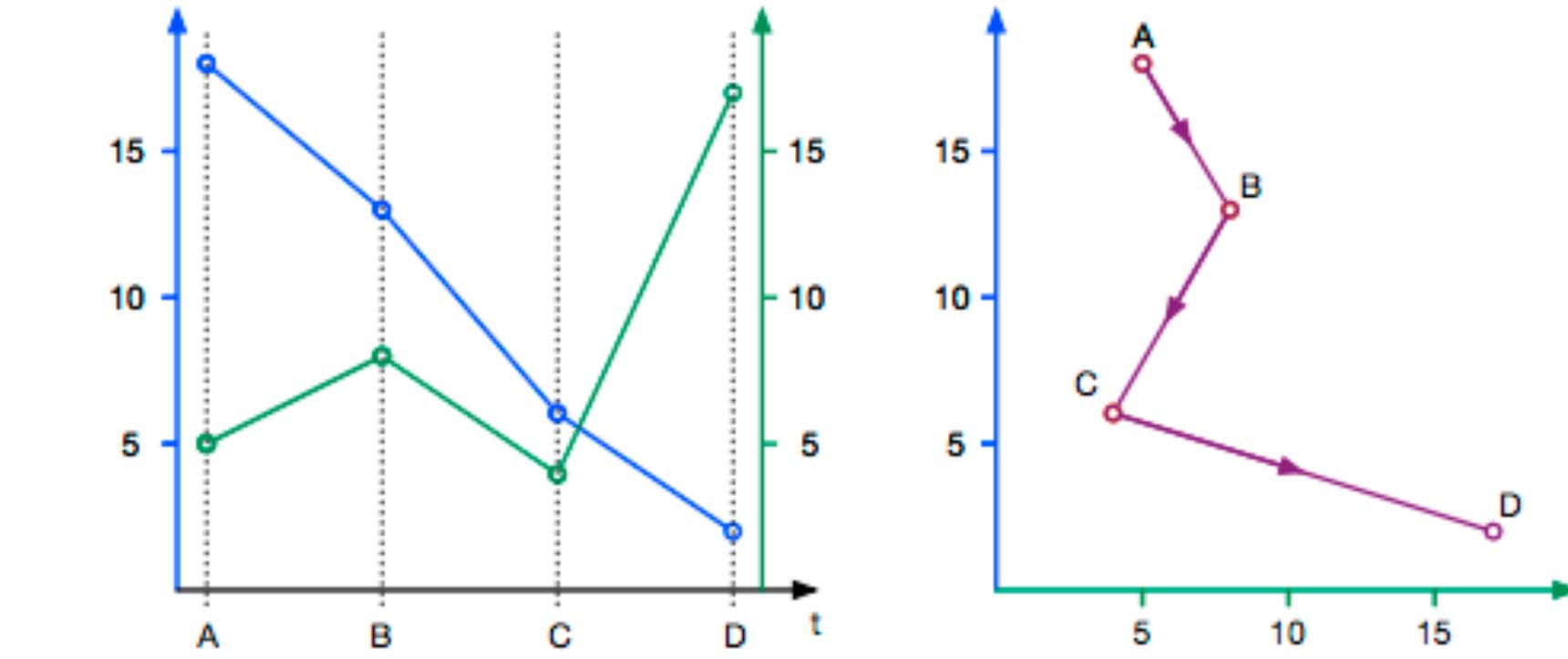
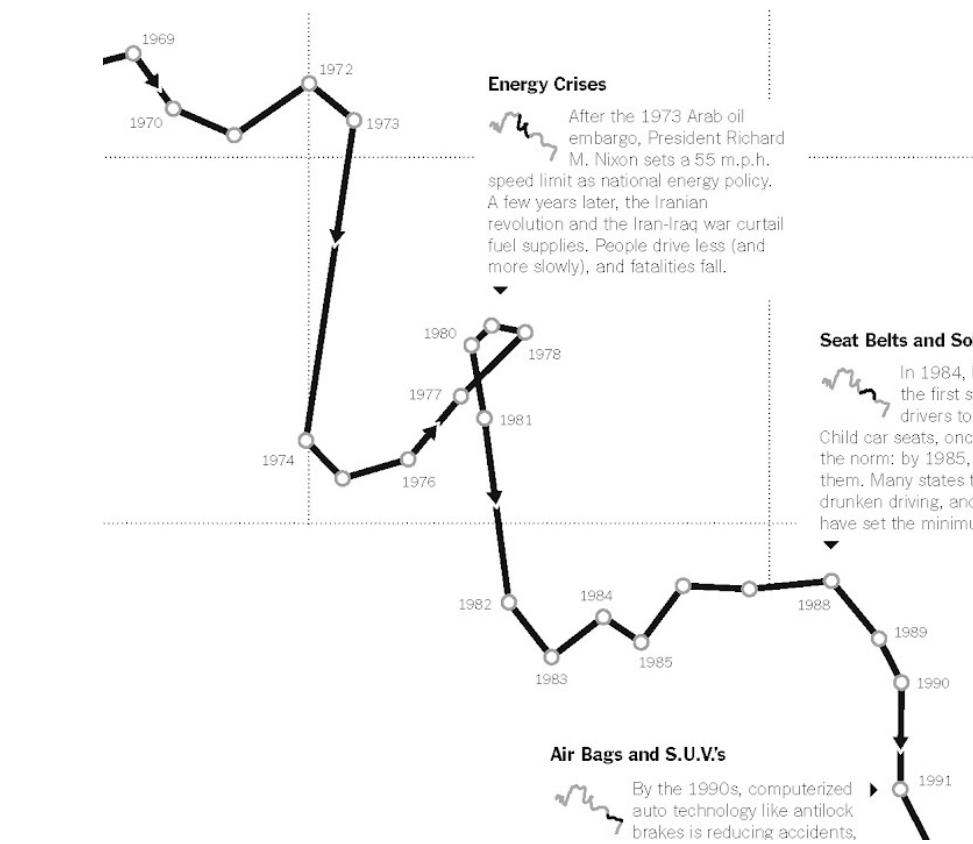
- controversial
 - acceptable if commensurate
 - beware, very easy to mislead!



Source | <http://www.baseball-reference.com/leagues/MLB/pitch.shtml> Ben Jones (@DataRemixed) | 5/4/2013

Idiom: connected scatterplots

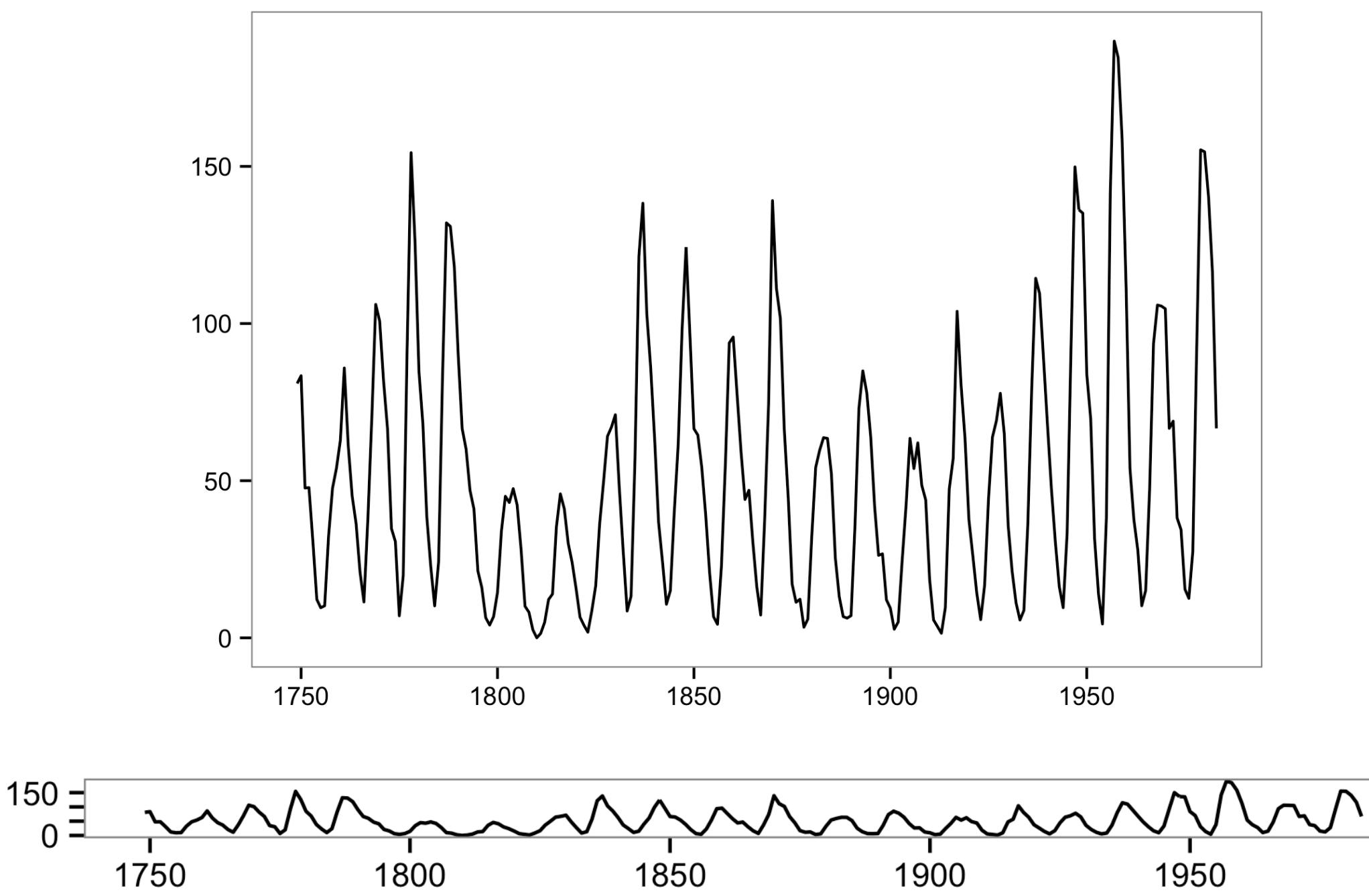
- scatterplot with line connection marks
 - popular in journalism
 - horiz + vert axes: value attrs
 - line connection marks: temporal order
 - alternative to dual-axis charts
 - horiz: time
 - vert: two value attrs
 - empirical study
 - engaging, but correlation unclear



Choosing line chart aspect ratios

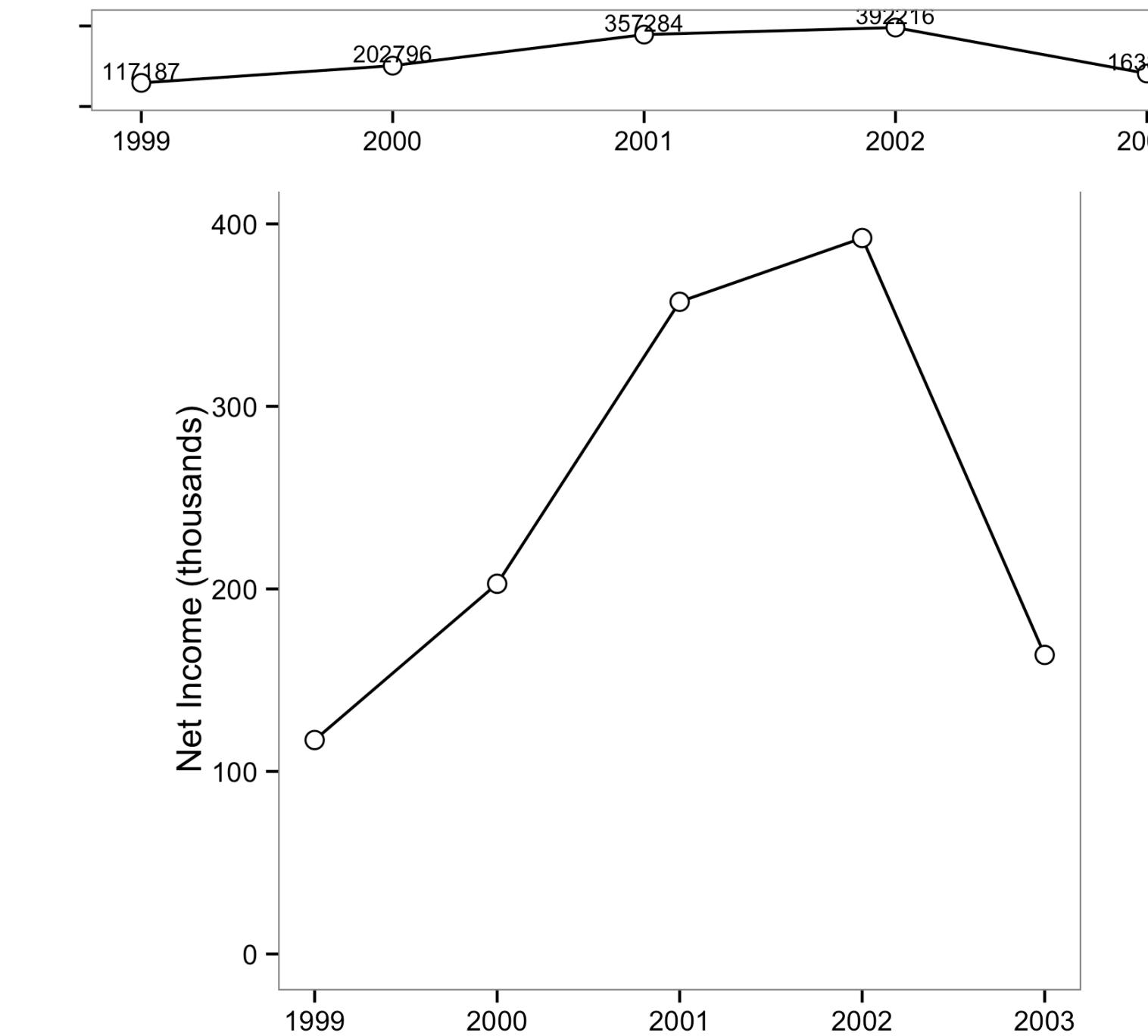
- 1: banking to 45 (1980s)
 - Cleveland perceptual argument: most accurate angle judgement at 45

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

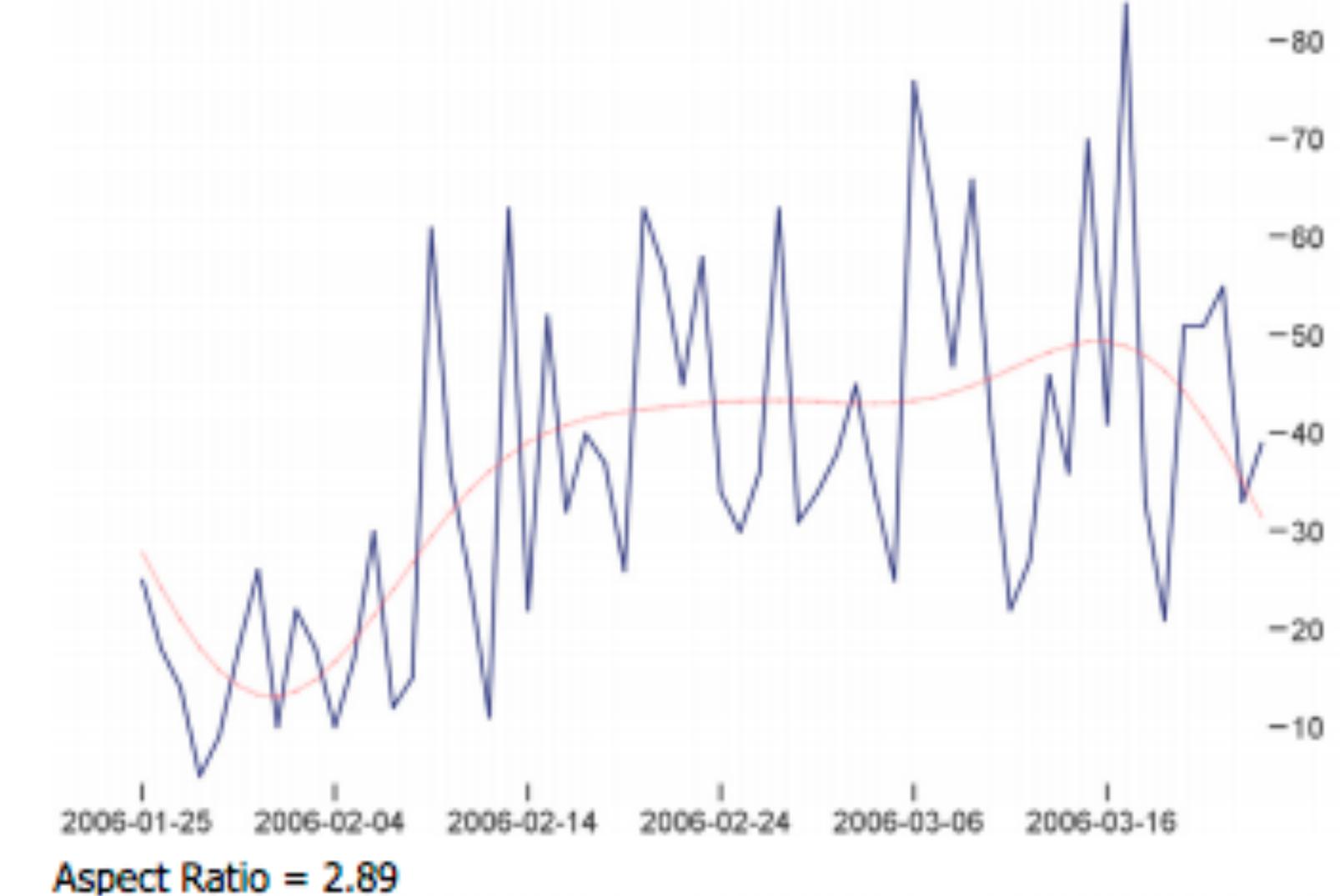
Downloads of the prefuse toolkit

Aspect Ratio = 1.44

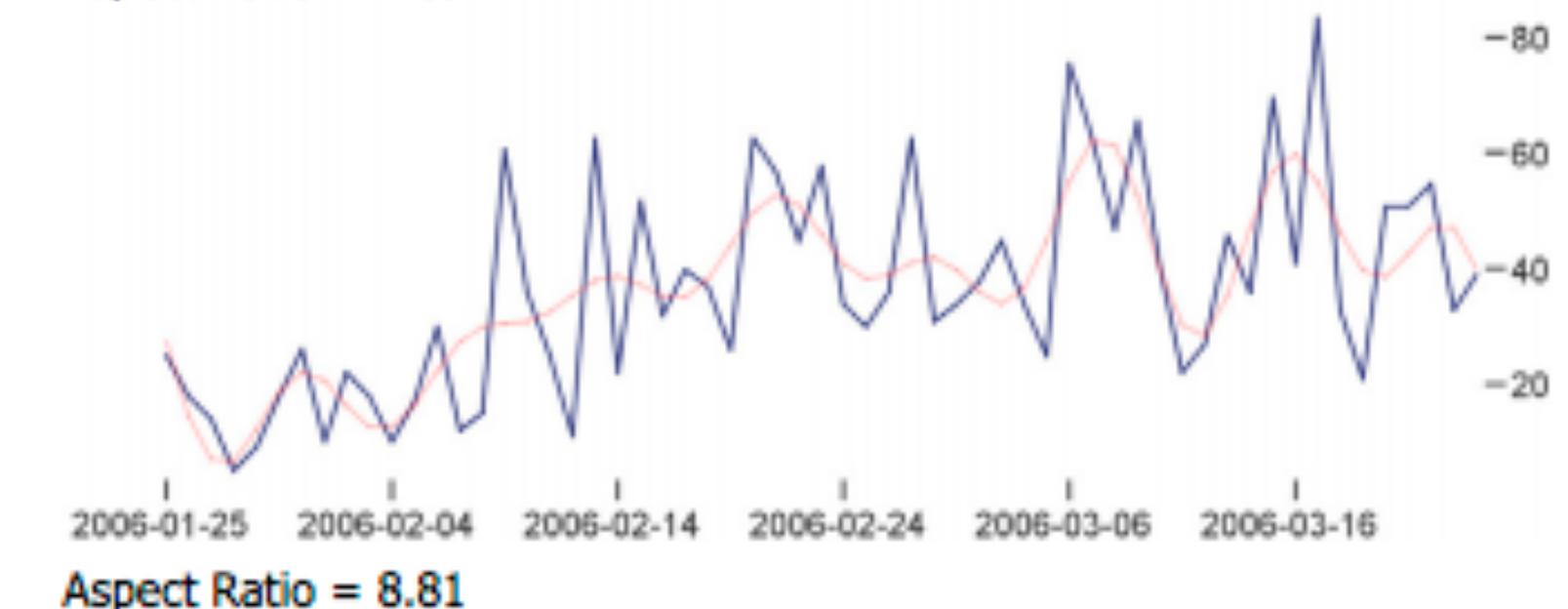
Choosing line chart aspect ratios

- 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

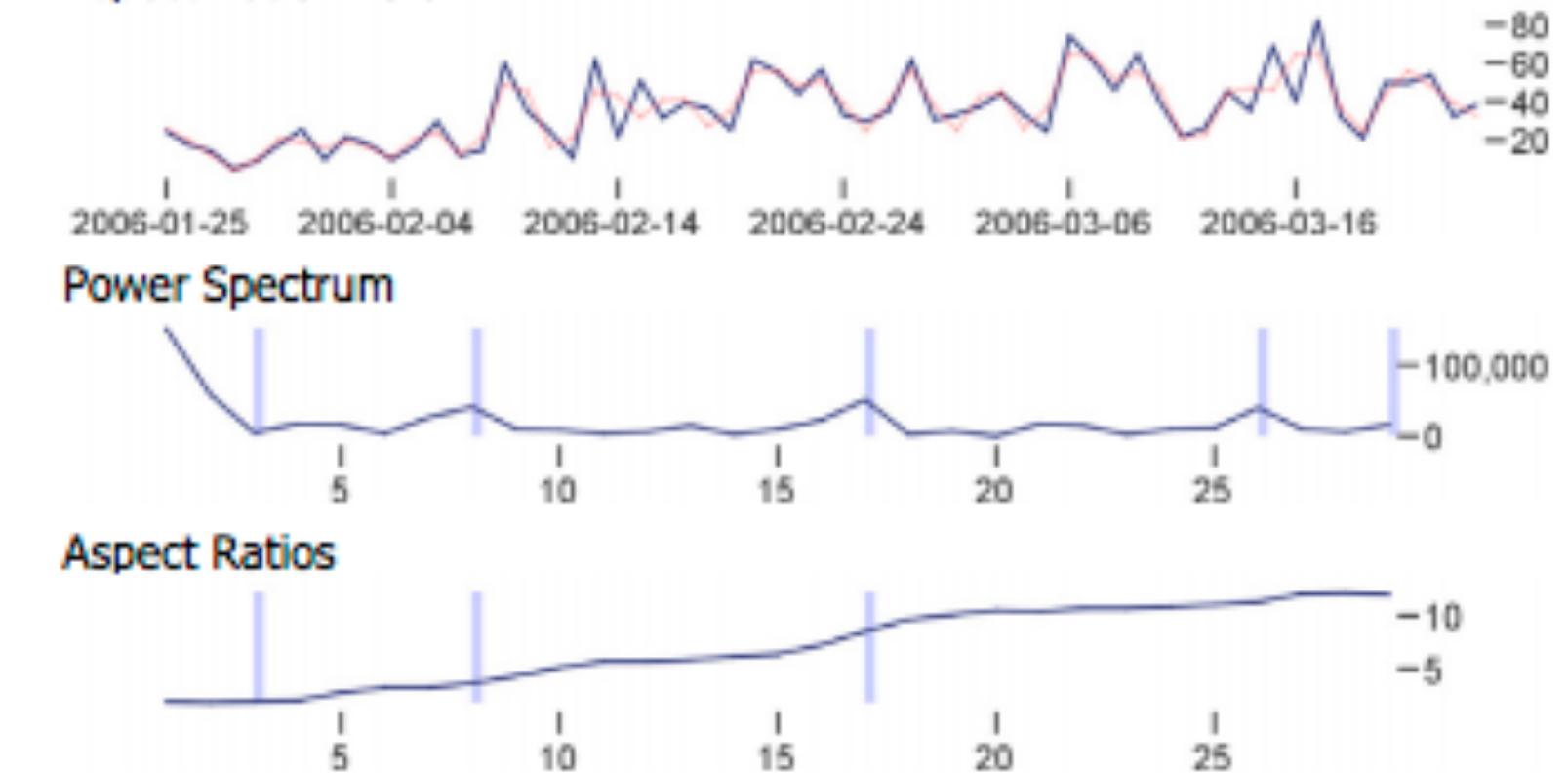
overall



weekly



daily



Visual vocabulary

Designing with data

There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

FT graphic: Alan Smith; Chris Campbell; Ian Bott; Liz Faunce; Graham Parrish; Billy Ehrenberg-Shannon; Paul McCallum; Martin Stabe
Inspired by the Graphic Continuum by Jon Schwabish and Severino Ribecca

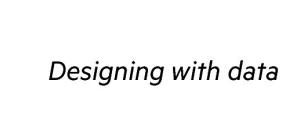
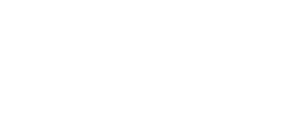
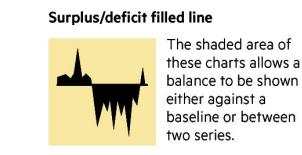
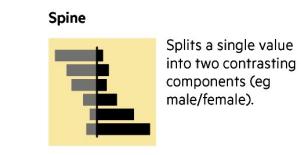
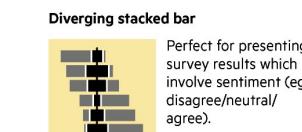
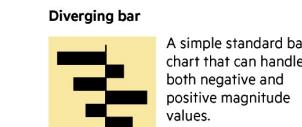


ft.com/vocabulary

Deviation

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (positive/neural/negative).

Example FT uses
Trade surplus/deficit, climate change



There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

FT graphic: Alan Smith; Chris Campbell; Ian Bott; Liz Faunce; Graham Parrish; Billy Ehrenberg-Shannon; Paul McCallum; Martin Stabe
Inspired by the Graphic Continuum by Jon Schwabish and Severino Ribecca

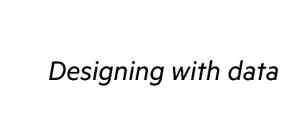
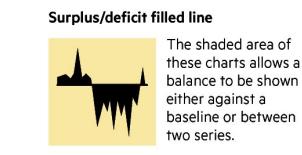
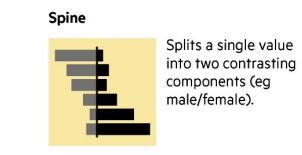
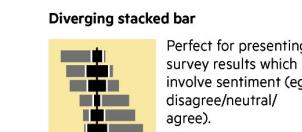
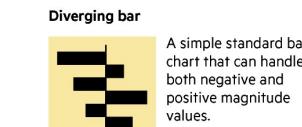


ft.com/vocabulary

Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e. one causes the other).

Example FT uses
Inflation and unemployment; income and life expectancy



There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

FT graphic: Alan Smith; Chris Campbell; Ian Bott; Liz Faunce; Graham Parrish; Billy Ehrenberg-Shannon; Paul McCallum; Martin Stabe
Inspired by the Graphic Continuum by Jon Schwabish and Severino Ribecca

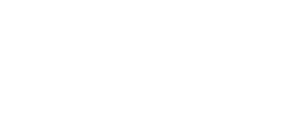
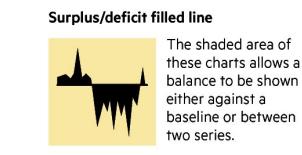
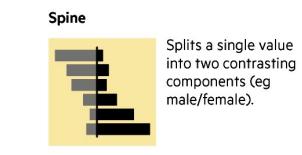
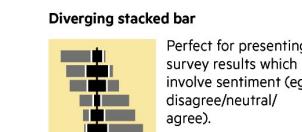
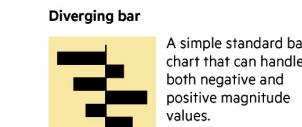


ft.com/vocabulary

Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

Example FT uses
Wealth, deprivation, league tables, constituency election results



There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

FT graphic: Alan Smith; Chris Campbell; Ian Bott; Liz Faunce; Graham Parrish; Billy Ehrenberg-Shannon; Paul McCallum; Martin Stabe
Inspired by the Graphic Continuum by Jon Schwabish and Severino Ribecca

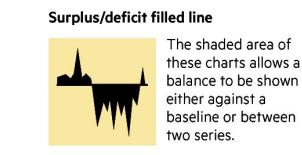
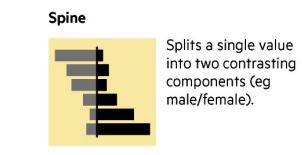
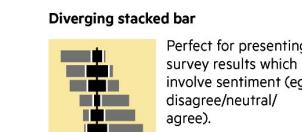
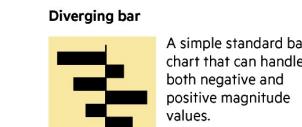


ft.com/vocabulary

Distribution

Show values in a dataset and how often they occur. The shape (or 'skew') of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.

Example FT uses
Income distribution, population (age/sex) distribution, revealing inequality



There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

FT graphic: Alan Smith; Chris Campbell; Ian Bott; Liz Faunce; Graham Parrish; Billy Ehrenberg-Shannon; Paul McCallum; Martin Stabe
Inspired by the Graphic Continuum by Jon Schwabish and Severino Ribecca

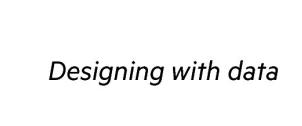
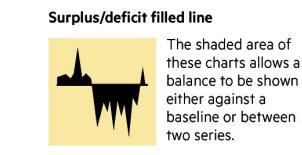
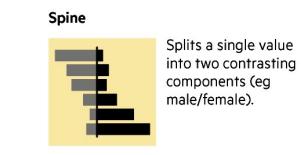
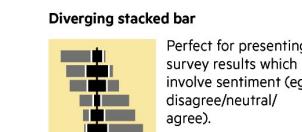
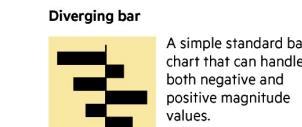


ft.com/vocabulary

Change over Time

Give emphasis to changing trends. These can be short (intra-day) movements or extended series traversing decades or centuries. Choosing the correct time period is important to provide suitable context for the reader.

Example FT uses
Share price movements, economic time series, sectoral changes in a market



There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

FT graphic: Alan Smith; Chris Campbell; Ian Bott; Liz Faunce; Graham Parrish; Billy Ehrenberg-Shannon; Paul McCallum; Martin Stabe
Inspired by the Graphic Continuum by Jon Schwabish and Severino Ribecca

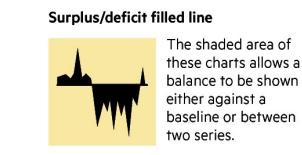
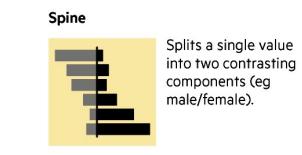
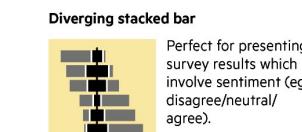
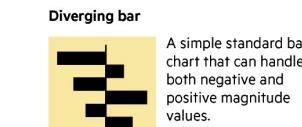


ft.com/vocabulary

Magnitude

Show size comparisons. These can be relative (just being able to see larger/bigger) or absolute (need to see fine differences). Usually these show a 'counted' number (for example, barrels, dollars or people) rather than a calculated rate or per cent.

Example FT uses
Commodity production, market capitalisation, volumes in general



There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

FT graphic: Alan Smith; Chris Campbell; Ian Bott; Liz Faunce; Graham Parrish; Billy Ehrenberg-Shannon; Paul McCallum; Martin Stabe
Inspired by the Graphic Continuum by Jon Schwabish and Severino Ribecca

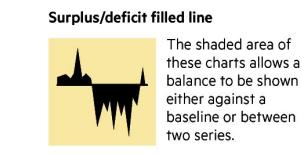
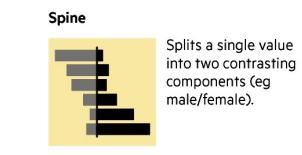
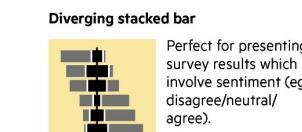
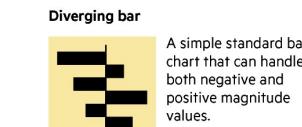


ft.com/vocabulary

Spatial

Show from locator maps only used when precise locations or geographical patterns in data are more important to the reader than anything else.

Example FT uses
Population density, natural resource locations, natural disaster risk/impact, catchment areas, variation in election results



There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

FT graphic: Alan Smith; Chris Campbell; Ian Bott; Liz Faunce; Graham Parrish; Billy Ehrenberg-Shannon; Paul McCallum; Martin Stabe
Inspired by the Graphic Continuum by Jon Schwabish and Severino Ribecca

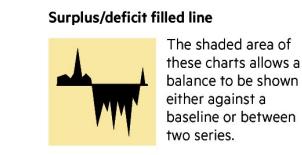
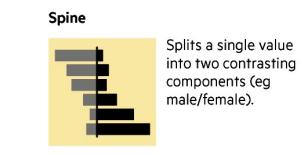
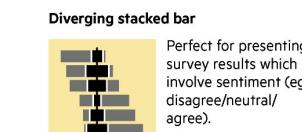
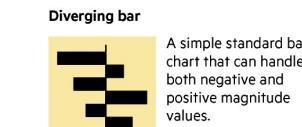


ft.com/vocabulary

Flow

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations.

Example FT uses
Movements of funds, trade, migrants, lawsuits, information; relationship graphs.



There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

FT graphic: Alan Smith; Chris Campbell; Ian Bott; Liz Faunce; Graham Parrish; Billy Ehrenberg-Shannon; Paul McCallum; Martin Stabe
Inspired by the Graphic Continuum by Jon Schwabish and Severino Ribecca



ft.com/vocabulary



© Financial Times

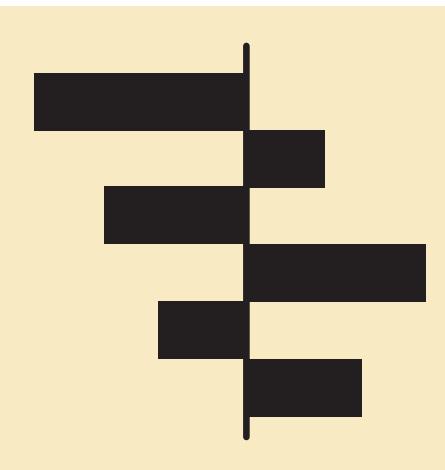
Deviation

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (positive/neutral/negative).

Example FT uses

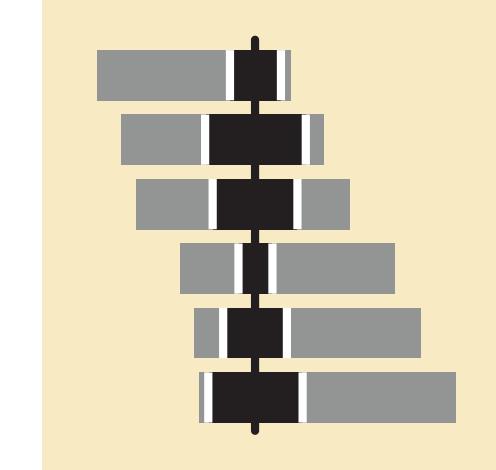
Trade surplus/deficit, climate change

Diverging bar



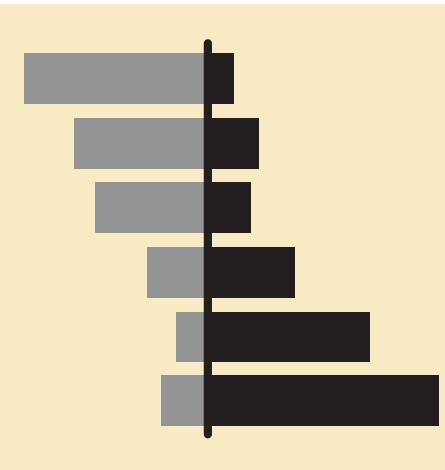
A simple standard bar chart that can handle both negative and positive magnitude values.

Diverging stacked bar



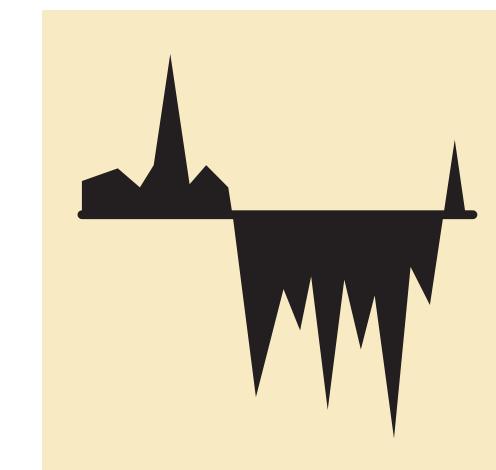
Perfect for presenting survey results which involve sentiment (eg disagree/neutral/agree).

Spine



Splits a single value into two contrasting components (eg male/female).

Surplus/deficit filled line



The shaded area of these charts allows a balance to be shown – either against a baseline or between two series.

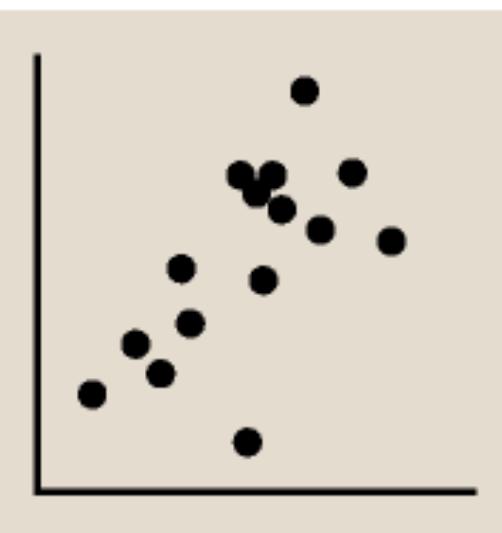
Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e. one causes the other).

Example FT uses

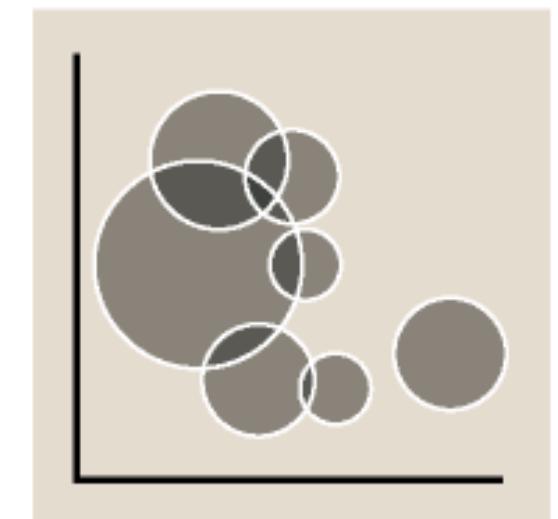
Inflation and unemployment, income and life expectancy

Scatterplot



The standard way to show the relationship between two continuous variables, each of which has its own axis.

Bubble



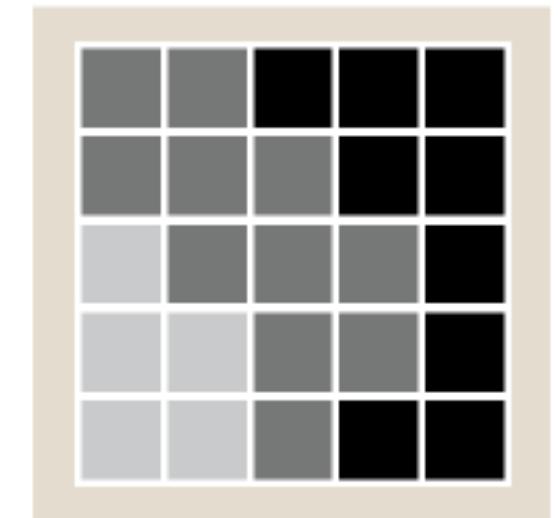
Like a scatterplot, but adds additional detail by sizing the circles according to a third variable.

Column + line timeline



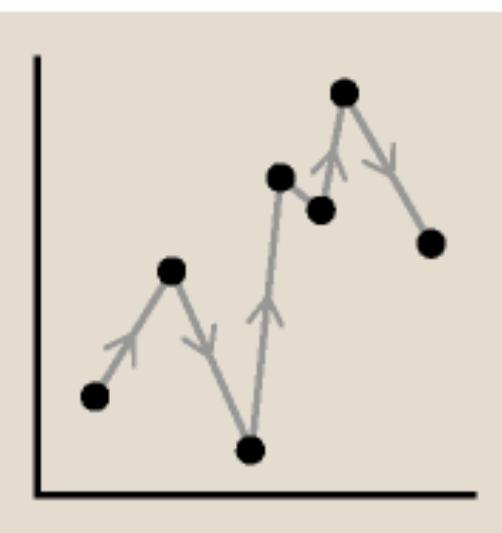
A good way of showing the relationship between an amount (columns) and a rate (line).

XY heatmap



A good way of showing the patterns between 2 categories of data, less effective at showing fine differences in amounts.

Connected scatterplot



Usually used to show how the relationship between 2 variables has changed over time.

Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

Example FT uses

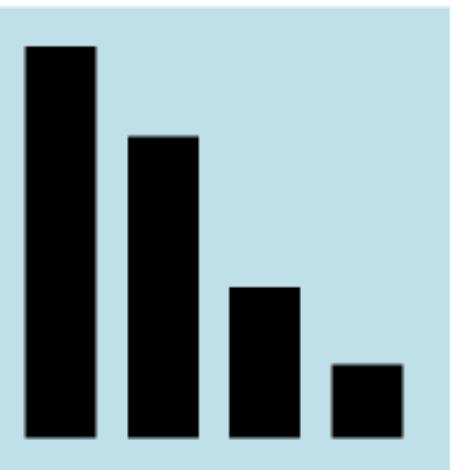
Wealth, deprivation, league tables, constituency election results

Ordered bar



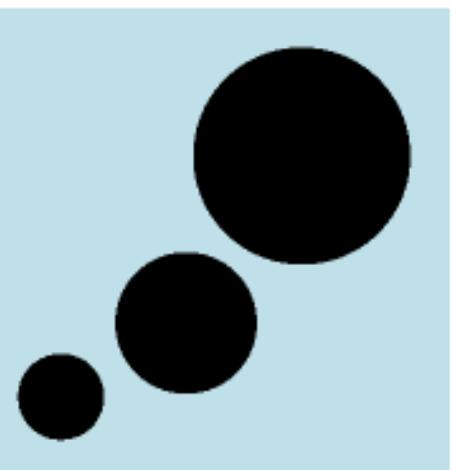
Standard bar charts display the ranks of values much more easily when sorted into order.

Ordered column



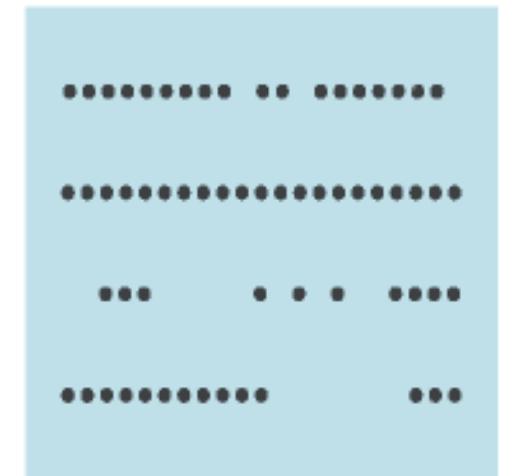
See above.

Ordered proportional symbol



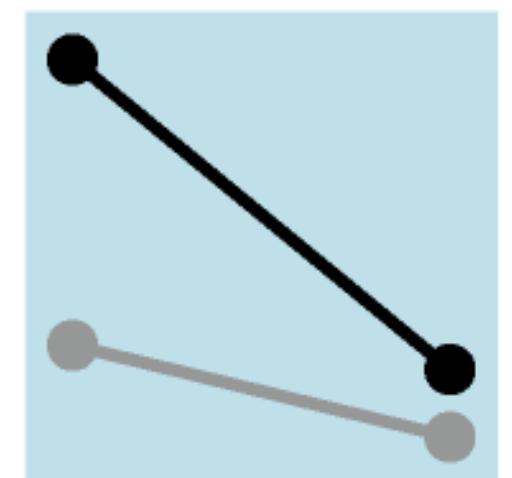
Use when there are big variations between values and/or seeing fine differences between data is not so important.

Dot strip plot



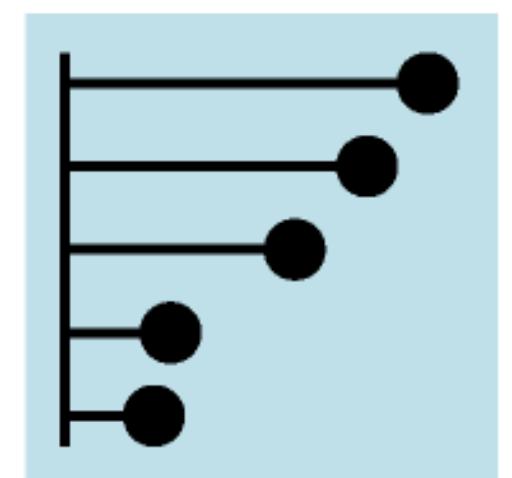
Dots placed in order on a strip are a space-efficient method of laying out ranks across multiple categories.

Slope



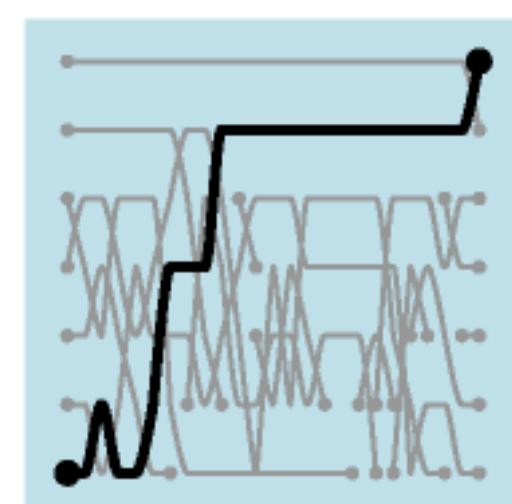
Perfect for showing how ranks have changed over time or vary between categories.

Lollipop



Lollipops draw more attention to the data value than standard bar/column and can also show rank and value effectively.

Bump



Effective for showing changing rankings across multiple dates. For large datasets, consider grouping lines using colour.

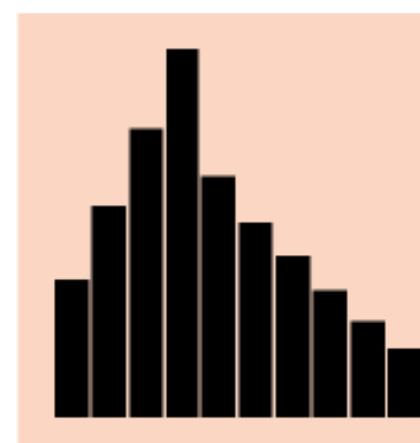
Distribution

Show values in a dataset and how often they occur. The shape (or 'skew') of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.

Example FT uses

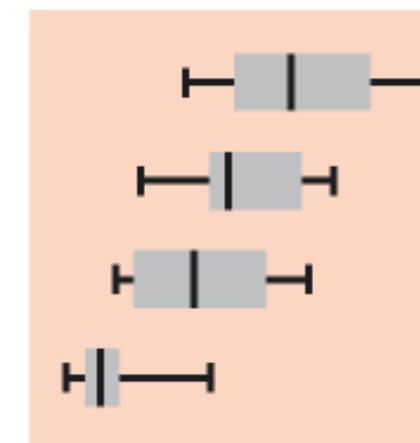
Income distribution, population (age/sex) distribution, revealing inequality

Histogram



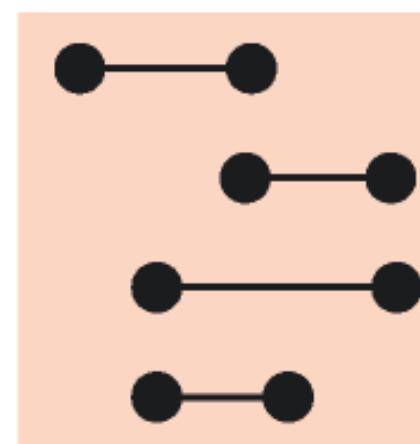
The standard way to show a statistical distribution - keep the gaps between columns small to highlight the 'shape' of the data.

Boxplot



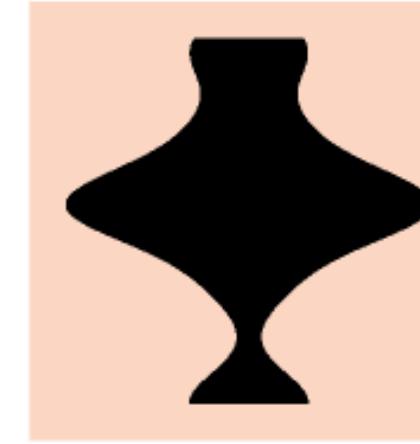
Summarise multiple distributions by showing the median (centre) and range of the data

Dot plot



A simple way of showing the change or range (min/max) of data across multiple categories.

Violin plot



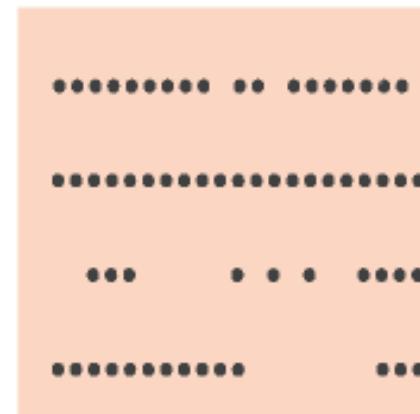
Similar to a box plot but more effective with complex distributions (data that cannot be summarised with simple average).

Frequency polygons



For displaying multiple distributions of data. Like a regular line chart, best limited to a maximum of 3 or 4 datasets.

Dot strip plot



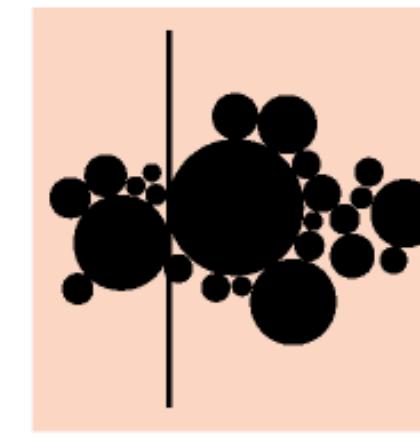
Good for showing individual values in a distribution, can be a problem when too many dots have the same value.

Population pyramid



A standard way for showing the age and sex breakdown of a population distribution; effectively, back to back histograms.

Beeswarm



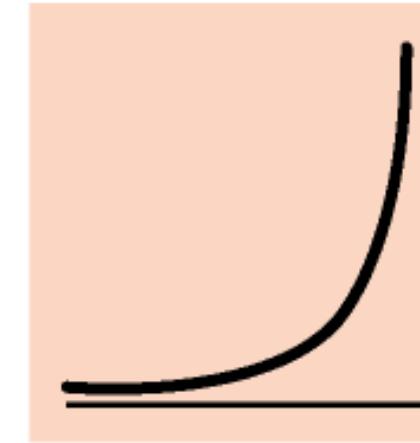
Use to emphasise individual points in a distribution. Points can be sized to an additional variable. Best with medium-sized datasets

Barcode plot



Like dot strip plots, good for displaying all the data in a table, they work best when highlighting individual values.

Cumulative curve



A good way of showing how unequal a distribution is: y axis is always cumulative frequency, x axis is always a measure.

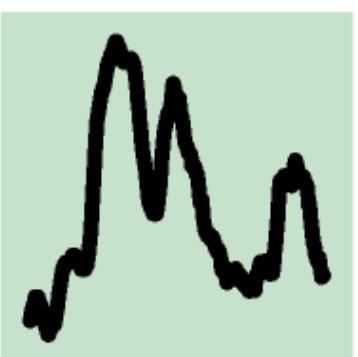
Change over Time

Give emphasis to changing trends. These can be short (intra-day) movements or extended series traversing decades or centuries: Choosing the correct time period is important to provide suitable context for the reader.

Example FT uses

Share price movements, economic time series, sectoral changes in a market

Line



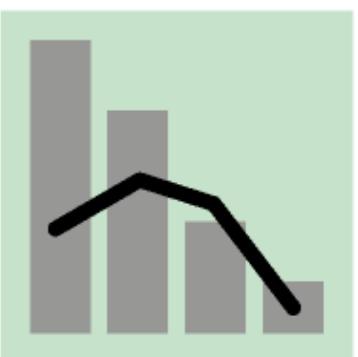
The standard way to show a changing time series. If data are irregular, consider markers to represent data points.

Column



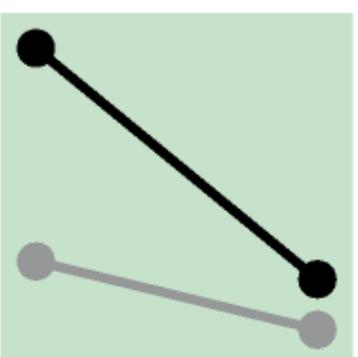
Columns work well for showing change over time - but usually best with only one series of data at a time.

Column + line timeline



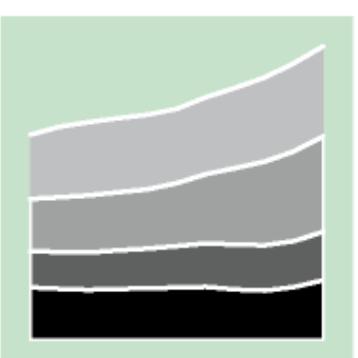
A good way of showing the relationship over time between an amount (columns) and a rate (line).

Slope



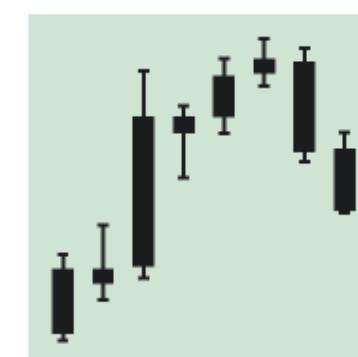
Good for showing changing data as long as the data can be simplified into 2 or 3 points without missing a key part of story.

Area chart



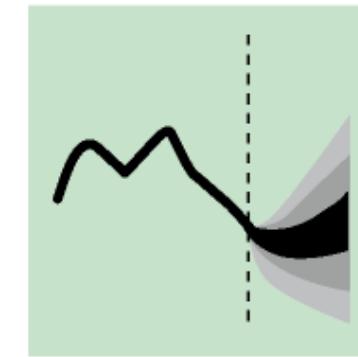
Use with care – these are good at showing changes to total, but seeing change in components can be very difficult.

Candlestick



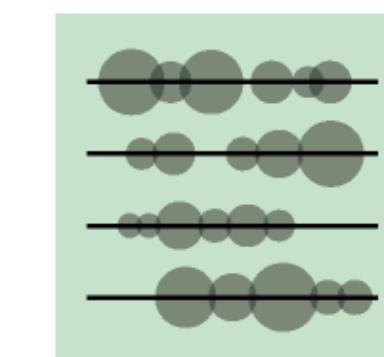
Usually focused on day-to-day activity, these charts show opening/closing and high/low points of each day.

Fan chart (projections)



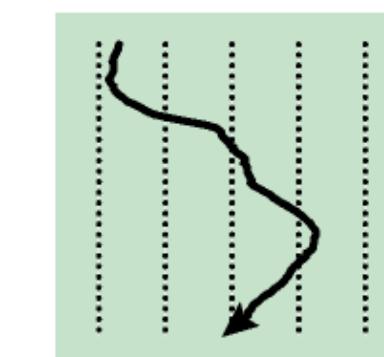
Use to show the uncertainty in future projections - usually this grows the further forward to projection.

Circle timeline



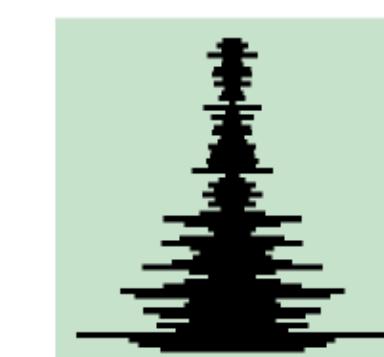
Good for showing discrete values of varying size across multiple categories (eg earthquakes by continent).

Vertical timeline



Presents time on the Y axis. Good for displaying detailed time series that work especially well when scrolling on mobile.

Seismogram



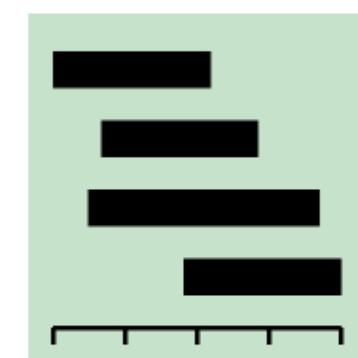
Another alternative to the circle timeline for showing series where there are big variations in the data.

Streamgraph



A type of area chart; use when seeing changes in proportions over time is more important than individual values

Priestley timeline



Great when date and duration are key elements of the story in the data.

Magnitude

Show size comparisons. These can be relative (just being able to see larger/bigger) or absolute (need to see fine differences). Usually these show a 'counted' number (for example, barrels, dollars or people) rather than a calculated rate or per cent.

Example FT uses

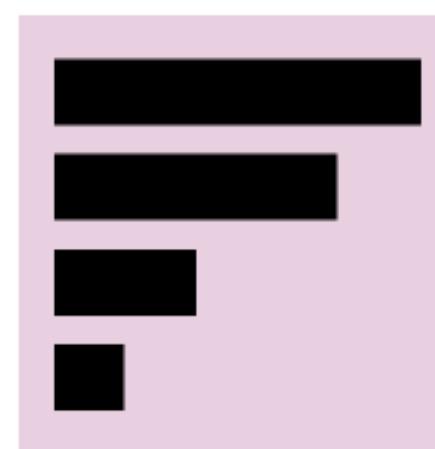
Commodity production, market capitalisation, volumes in general

Column



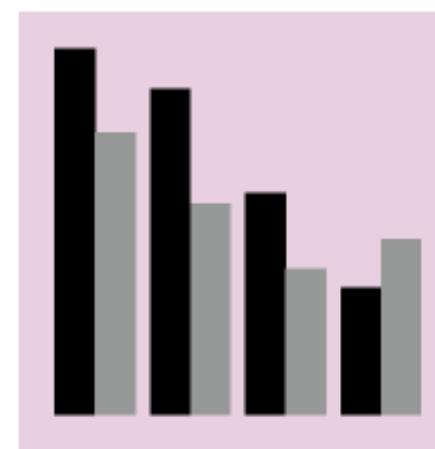
The standard way to compare the size of things. Must always start at 0 on the axis.

Bar



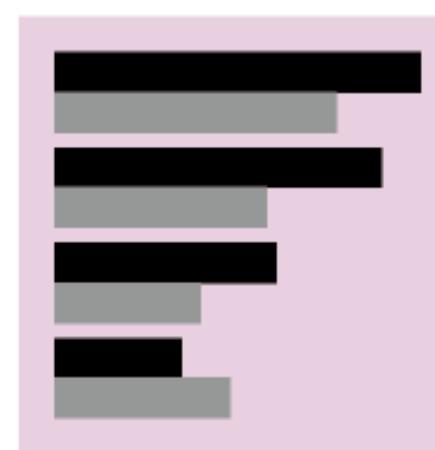
See above. Good when the data are not time series and labels have long category names.

Paired column



As per standard column but allows for multiple series. Can become tricky to read with more than 2 series.

Paired bar



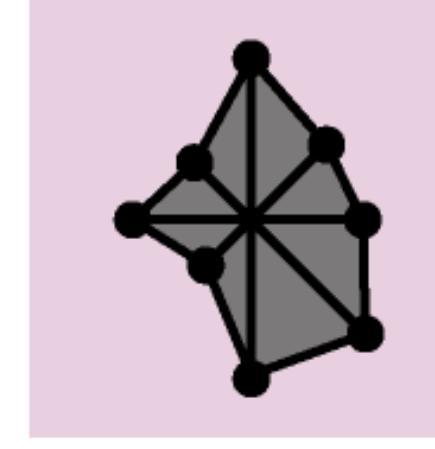
See above.

Marimekko



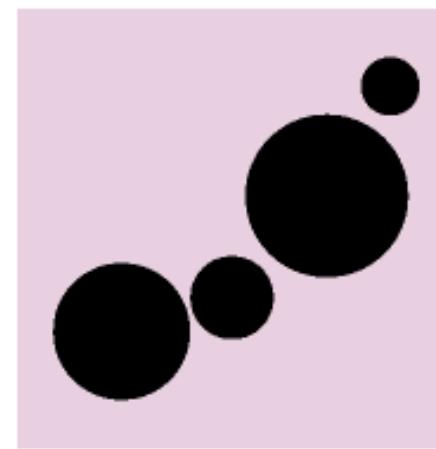
A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.

Radar



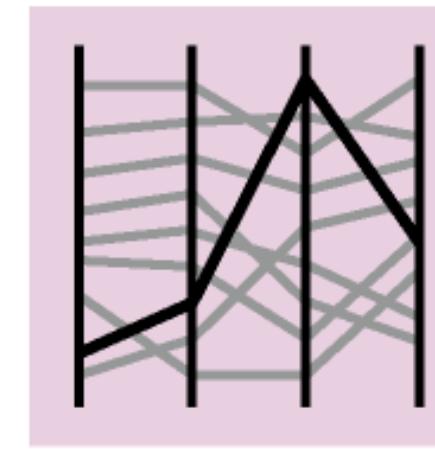
A space-efficient way of showing value of multiple variables – but make sure they are organised in a way that makes sense to reader.

Proportional symbol



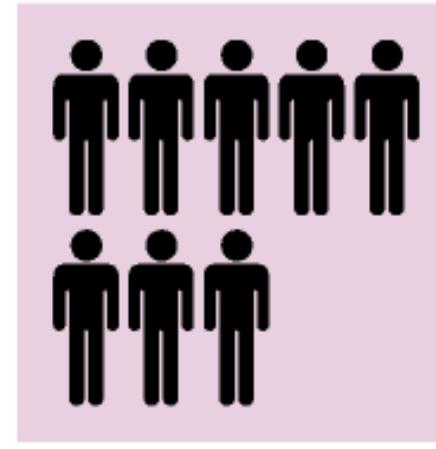
Use when there are big variations between values and/or seeing fine differences between data is not so important.

Parallel coordinates



An alternative to radar charts – again, the arrangement of the variables is important. Usually benefits from highlighting values.

Isotype (pictogram)



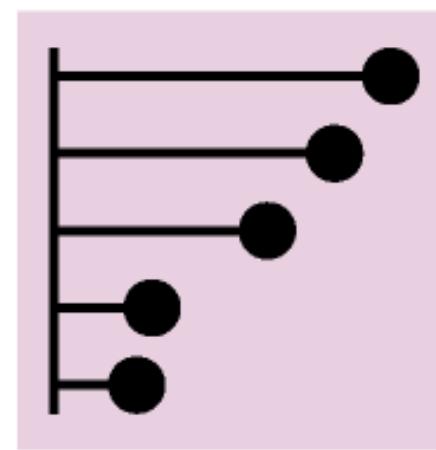
Excellent solution in some instances – use only with whole numbers (do not slice off an arm to represent a decimal).

Bullet



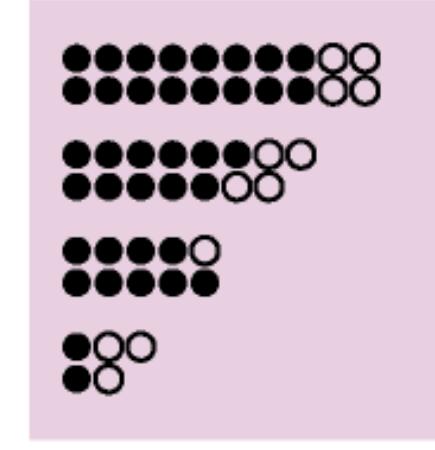
Good for showing a measurement against the context of a target or performance range.

Lollipop



Lollipop charts draw more attention to the data value than standard bar/column – does not have to start a zero (but preferable).

Grouped symbol



An alternative to bar/column charts when being able to count data or highlight individual elements is useful.

Part-to-whole

Show how a single entity can be broken down into its component elements. If the reader's interest is solely in the size of the components, consider a magnitude-type chart instead.

Example FT uses

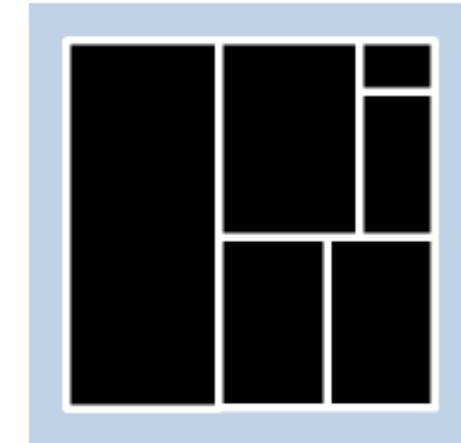
Fiscal budgets, company structures, national election results

Stacked column/bar



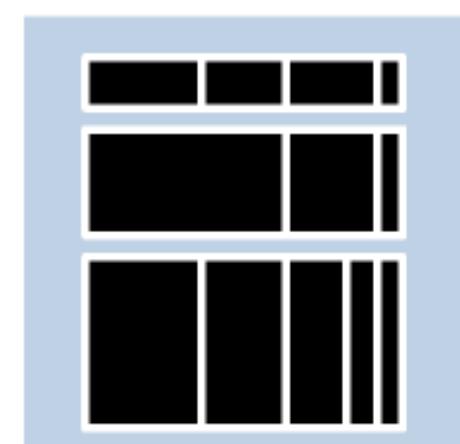
A simple way of showing part-to-whole relationships but can be difficult to read with more than a few components.

Treemap



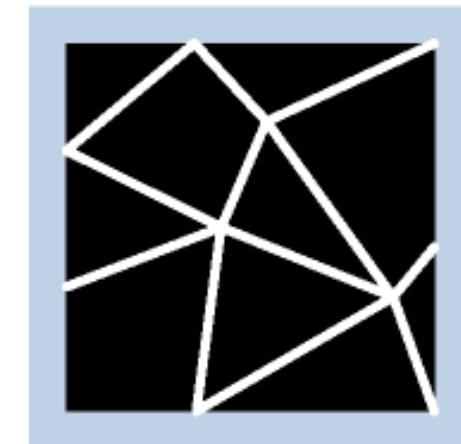
Use for hierarchical part-to-whole relationships; can be difficult to read when there are many small segments.

Marimekko



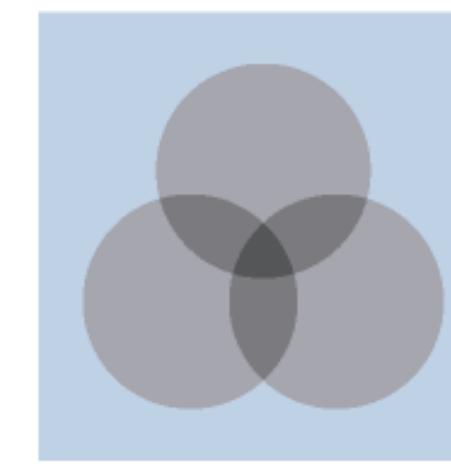
A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.

Voronoi



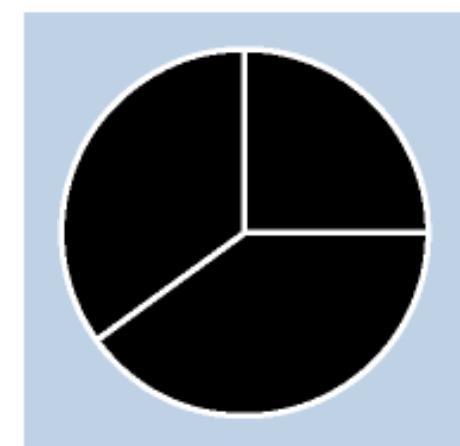
A way of turning points into areas – any point within each area is closer to the central point than any other centroid.

Venn



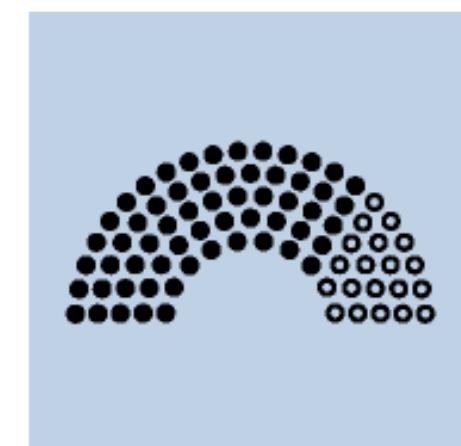
Generally only used for schematic representation.

Pie



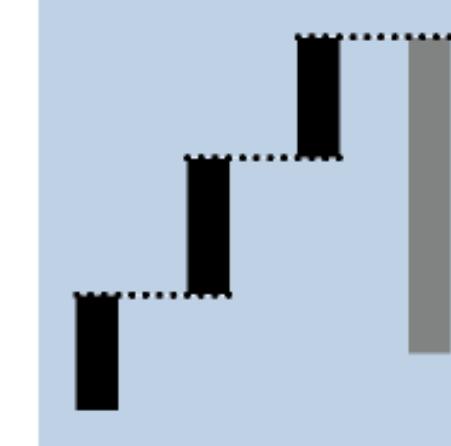
A common way of showing part-to-whole data – but be aware that it's difficult to accurately compare the size of the segments.

Arc



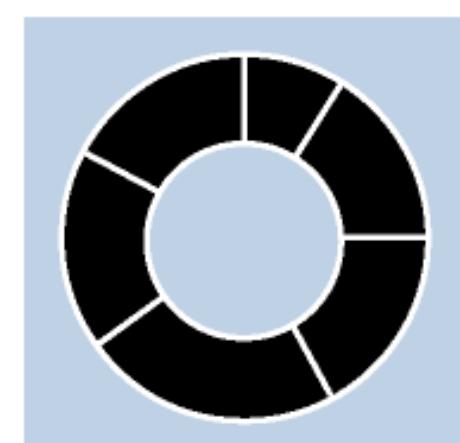
A hemicycle, often used for visualising parliamentary composition by number of seats.

Waterfall



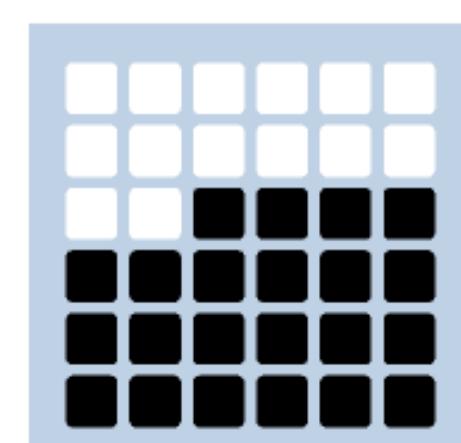
Can be useful for showing part-to-whole relationships where some of the components are negative.

Donut



Similar to a pie chart – but the centre can be a good way of making space to include more information about the data (eg total).

Gridplot



Good for showing % information, they work best when used on whole numbers and work well in small multiple layout form.

Spatial

Aside from locator maps only used when precise locations or geographical patterns in data are more important to the reader than anything else.

Example FT uses

Population density, natural resource locations, natural disaster risk/impact, catchment areas, variation in election results

Basic choropleth (rate/ratio)



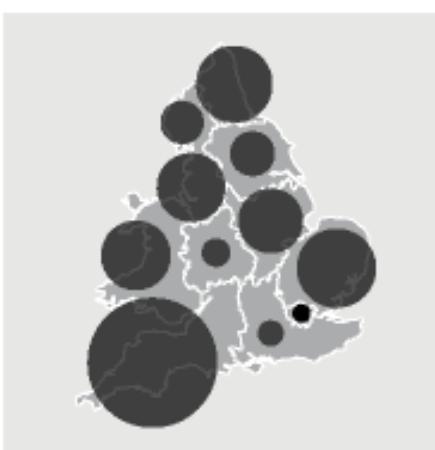
The standard approach for putting data on a map – should always be rates rather than totals and use a sensible base geography.

Equalised cartogram



Converting each unit on a map to a regular and equally-sized shape – good for representing voting regions with equal value.

Proportional symbol (count/magnitude)



Use for totals rather than rates – be wary that small differences in data will be hard to see.

Scaled cartogram (value)



Stretching and shrinking a map so that each area is sized according to a particular value.

Flow map



For showing unambiguous movement across a map.

Dot density



Used to show the location of individual events/locations – make sure to annotate any patterns the reader should see.

Contour map



For showing areas of equal value on a map. Can use deviation colour schemes for showing +/- values

Heat map



Grid-based data values mapped with an intensity colour scale. As choropleth map – but not snapped to an admin/political unit.

Flow

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations.

Example FT uses

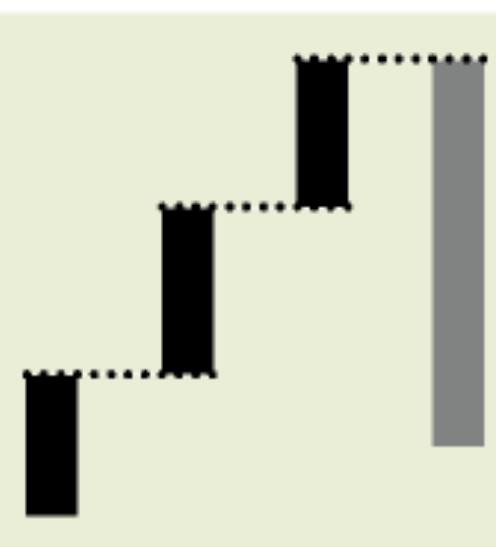
Movement of funds, trade, migrants, lawsuits, information; relationship graphs.

Sankey



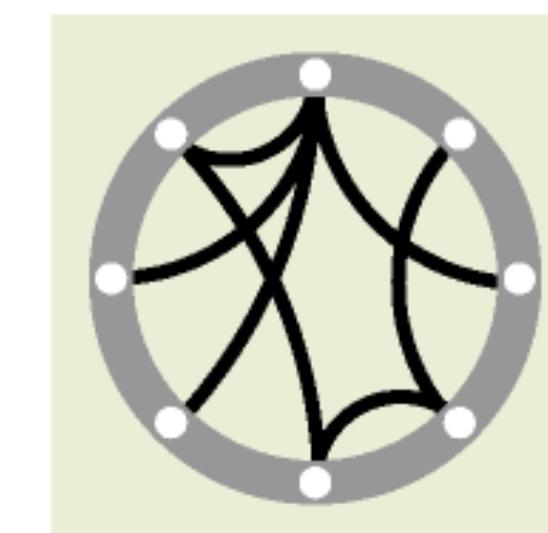
Shows changes in flows from one condition to at least one other; good for tracing the eventual outcome of a complex process.

Waterfall



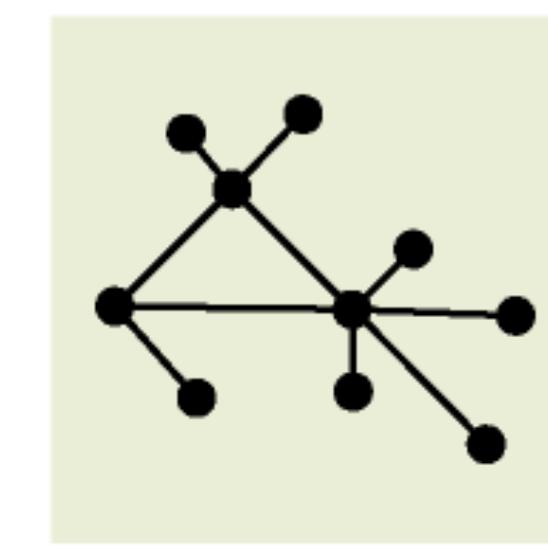
Designed to show the sequencing of data through a flow process, typically budgets. Can include +/- components.

Chord



A complex but powerful diagram which can illustrate 2-way flows (and net winner) in a matrix.

Network



Used for showing the strength and inter-connectedness of relationships of varying types.

Reading Material

[dv3] Chapter 7 - Arrange Tables

Questions?

:::::::



@rschifan



rossano.schifanella@unito.it



<http://www.di.unito.it/~schifane>