Visualization for Data Science
The Zoo

Office hours

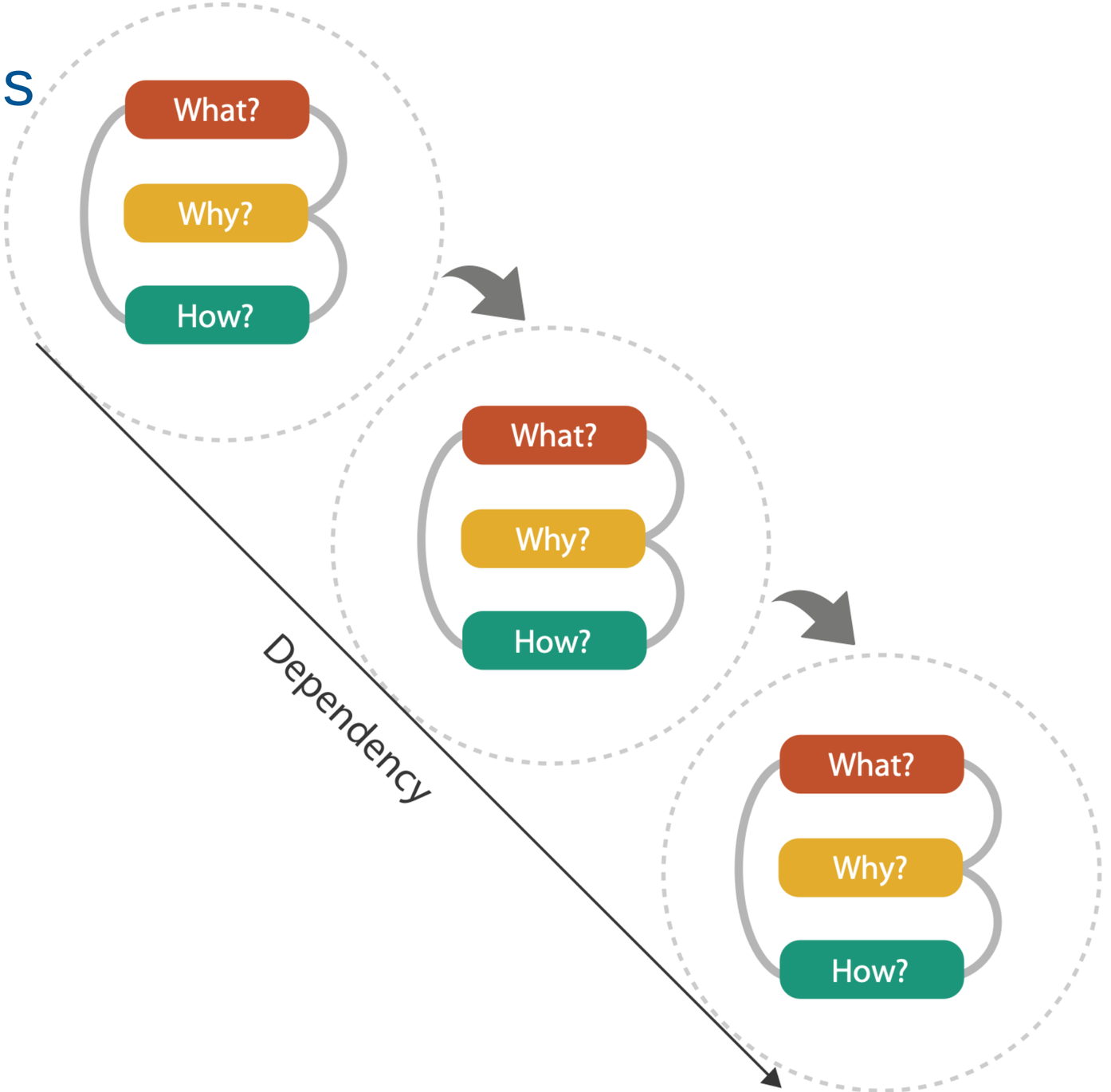
- TA Office Hours
  - Tuesdays in RM 238 at 5pm
  - Wednesdays Online on Zoom at 5pm
  - Saturday Online on Zoom at 2pm
  - Sunday Online on EdStem at 4pm
- Instructor Office Hours
  - **In Person RM 202 from 1 – 2:45pm on Wednesdays**

Abstraction

- these {action, target} pairs are good starting point for vocabulary
  - but sometimes you'll need more precision!
- rule of thumb
  - systematically remove all domain jargon

- interplay: task and data abstraction
  - need to use data abstraction within task abstraction
    - to specify your targets!
    - but task abstraction can lead you to transform the data
  - iterate back and forth
    - first pass data, first pass task, second pass data, ...

Means and ends



What?

Why?

How?

What?

Datasets

Data Types

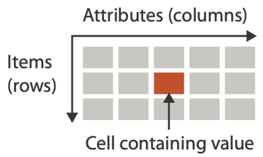
- Items
- Attributes
- Links
- Positions
- Grids

Data and Dataset Types

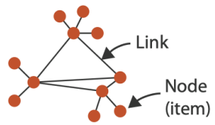
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Dataset Types

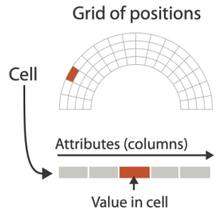
- Tables



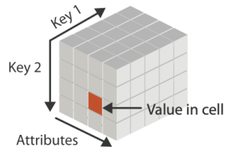
- Networks



- Fields (Continuous)



- *Multidimensional Table*



- *Trees*

- Geometry (Spatial)

Dataset Availability

- Static

- Dynamic

Attributes

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

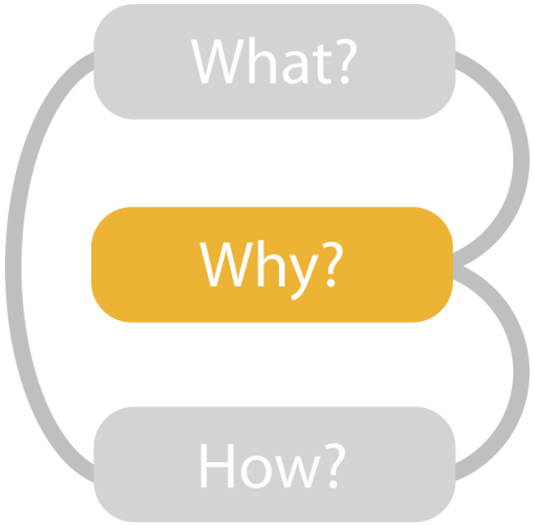
  - *Quantitative*

Ordering Direction

- Sequential

- Diverging

- Cyclic



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

Why?

Actions

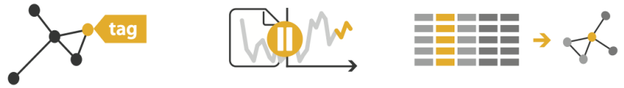
Analyze

→ Consume

→ *Discover* → *Present* → *Enjoy*

→ Produce

→ *Annotate* → *Record* → *Derive*

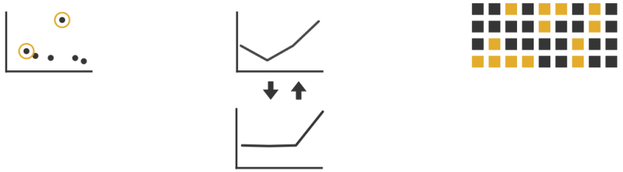


Search

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

Query

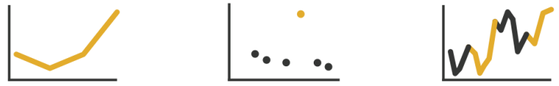
→ Identify → Compare → Summarize



Targets

All Data

→ Trends → Outliers → Features



Attributes

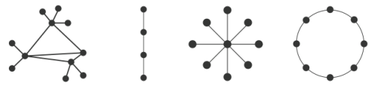
→ One → Many

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

→ *Extremes*

Network Data

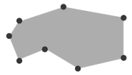
→ Topology



→ *Paths*

Spatial Data

→ Shape



DATA

NYC OPEN DATA WEEK MULTI-PARK SQUIRREL COUNT

The Squirrel Census is a multimedia science, design, and storytelling project focusing on the Eastern gray (*Sciurus carolinensis*). They count squirrels and present their findings to the public.

On March 1, 2020 — with the help of 72 volunteer Squirrel Sighters, as well as NYC Open Data — they performed a sample count in 24 New York City parks, and gathered other material data. Four hundred and thirty-three squirrel sightings were tallied. The methodology was less focused on total squirrel numbers per hectare and more attuned to the stories — of squirrels, humans, and parks.

The data is organized into three sets below.

USER GUIDE

PARK DATA

SQUIRREL DATA

STORIES

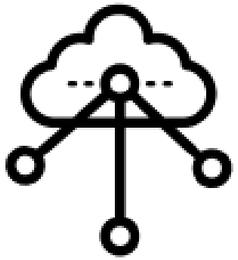
Data and Task Abstraction Group Exercise – 20 minutes

- 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)

Map



Visualization Theory:

- User-Centered Design
- Data Types
- What is the question?
- Who is the audience?
- What is the data?

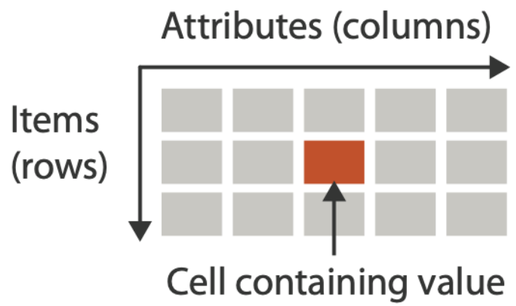
Learning Outcomes

- Analyze a dataset using both the task and data abstraction frameworks presented so far
- Select a visualization based on the data and task
- Describe visualizations in terms of keys and values that they encode

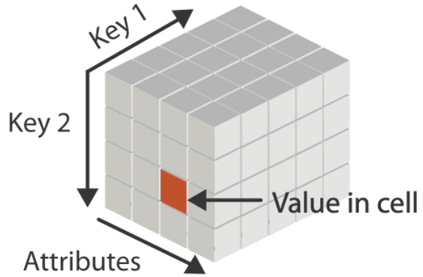
Focus on Tables

Dataset Types

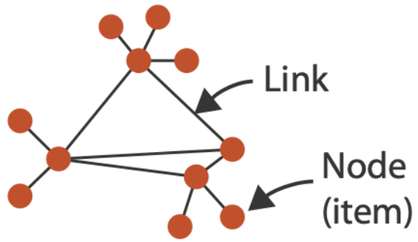
Tables



Multidimensional Table



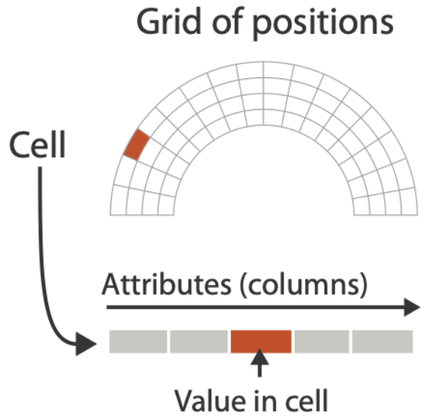
Networks



Trees

Spatial

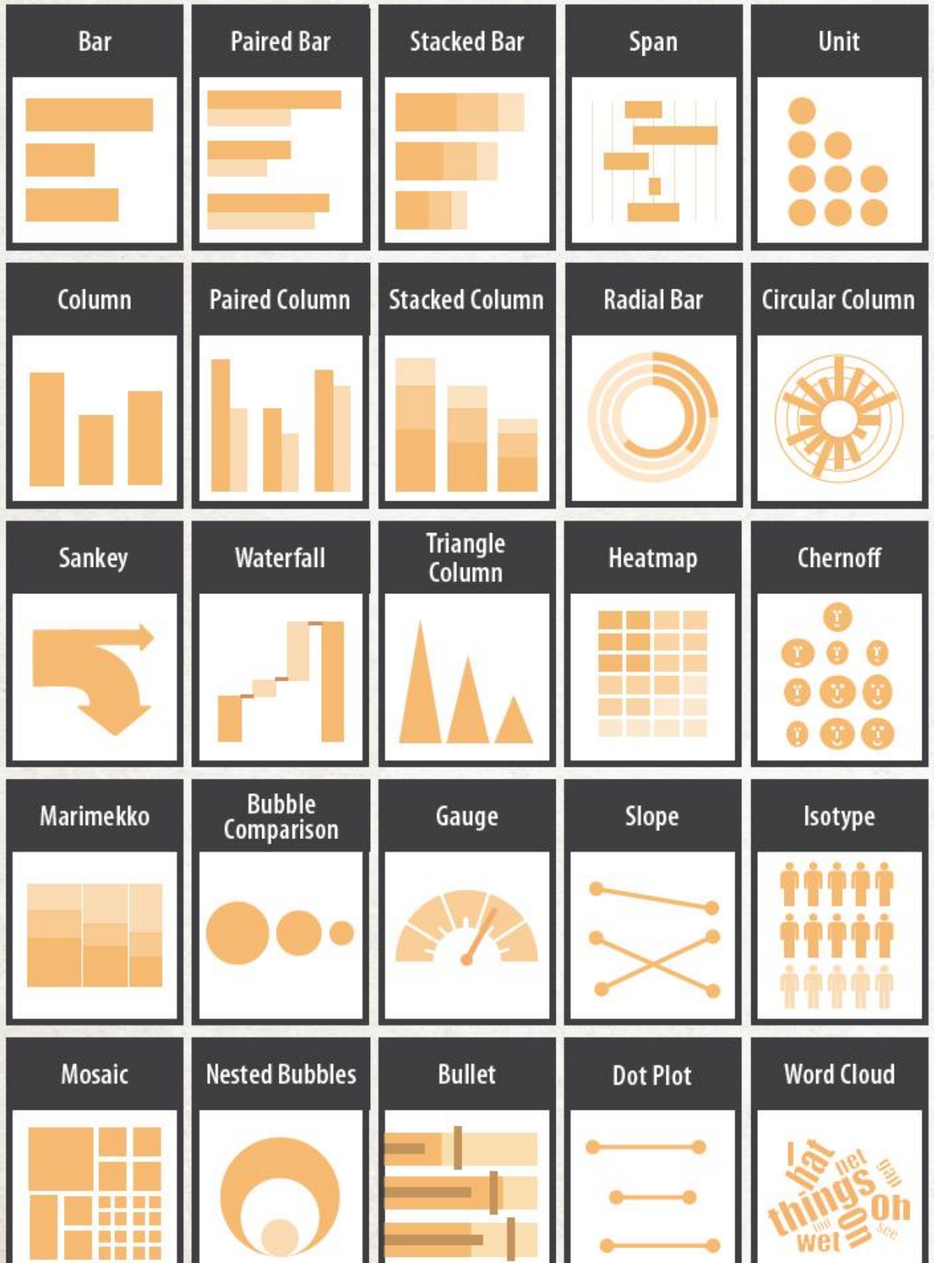
Fields (Continuous)



Geometry (Spatial)

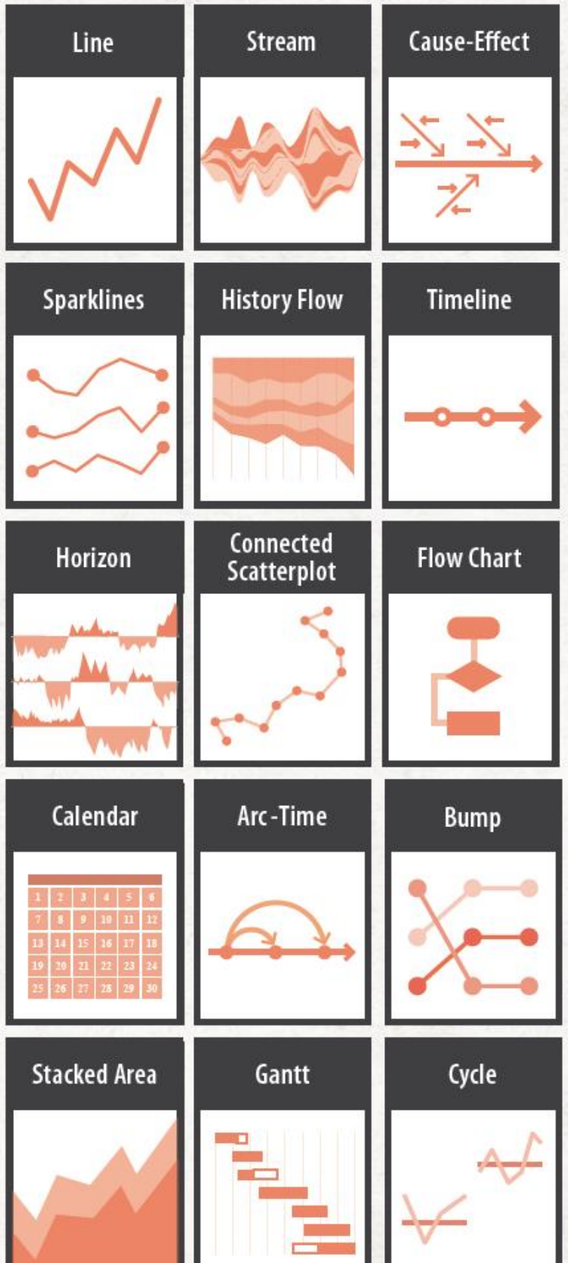
COMPARING CATEGORIES

Compare values across categories



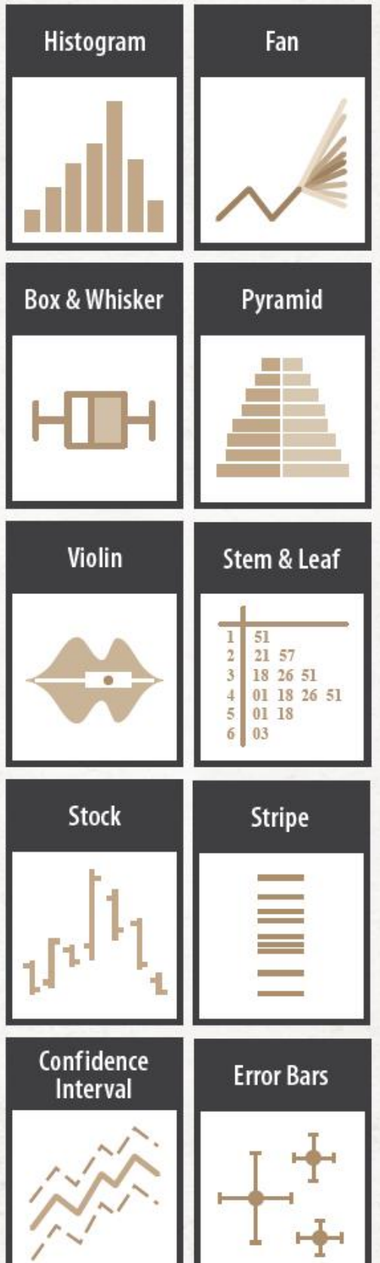
TIME

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

How?

Encode

Arrange

- Express
- Separate
- Order
- Align
- Use

Map

from categorical and ordered attributes

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

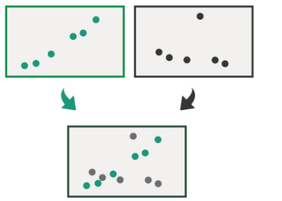
Direction, Rate, Frequency, ...

Manipulate

- Change
- Select
- Navigate

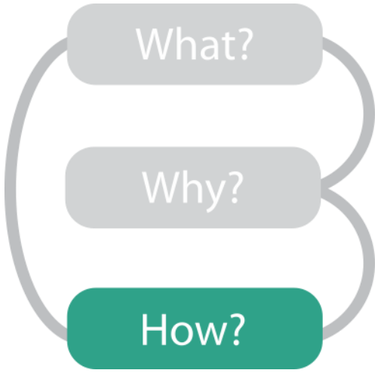
Facet

- Juxtapose
- Partition
- Superimpose



Reduce

- Filter
- Aggregate
- Embed



Keys and values

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

➔ *0 Keys*

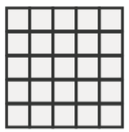
➔ **Express Values**

➔ *1 Key*

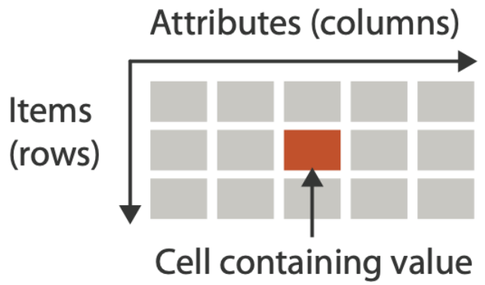
List

➔ *2 Keys*

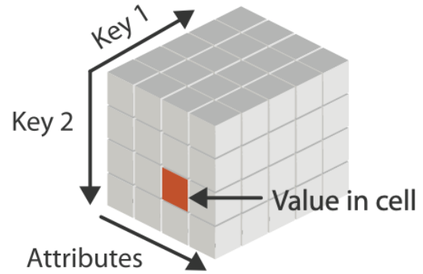
Matrix



➔ Tables



➔ *Multidimensional Table*



How?

Encode

Arrange

- Express

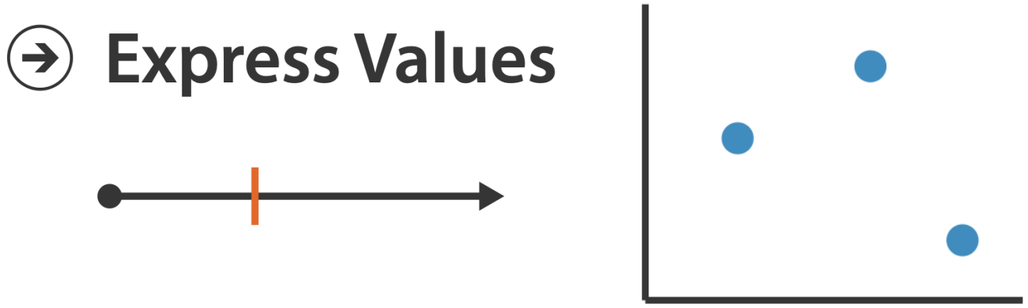
- Separate

- Order

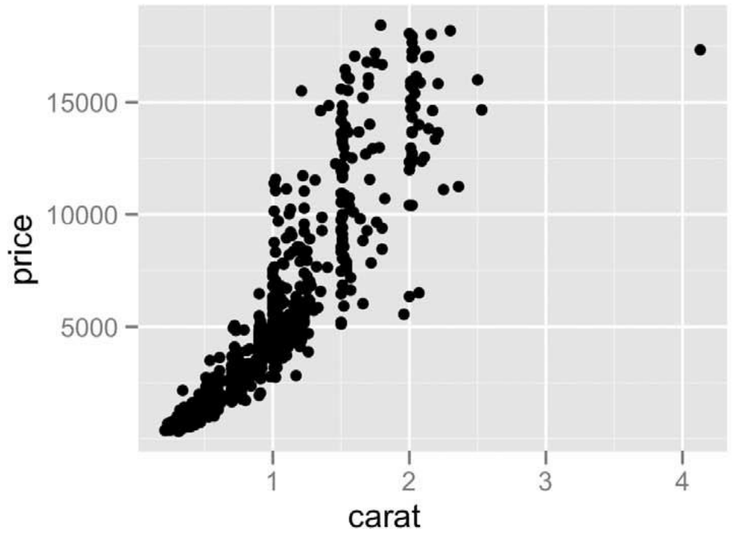
- Align

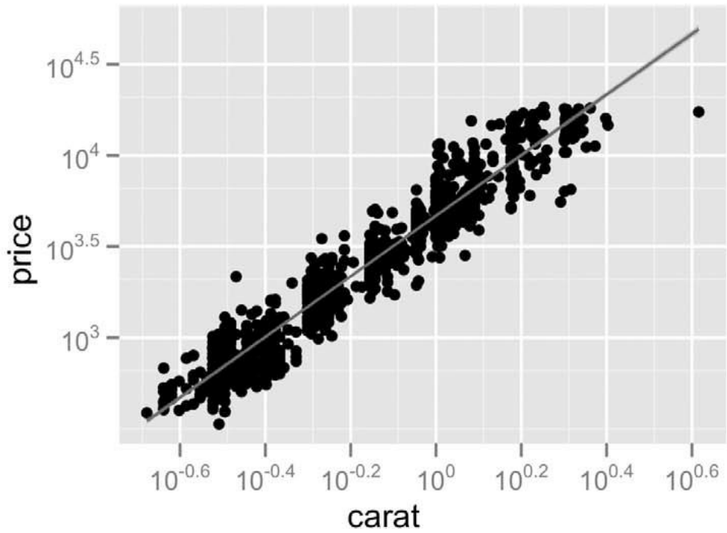
- Use

Idiom: **scatterplot**



- **express** values (magnitudes)
  - quantitative attributes
- no keys, only values
  - data
    - 2 quant attribs
  - mark: points
  - channels
    - horiz + vert position
  - tasks
    - Providing overviews, characterize distributions, find trends, outliers and correlation, locate clusters
  - scalability
    - hundreds of items

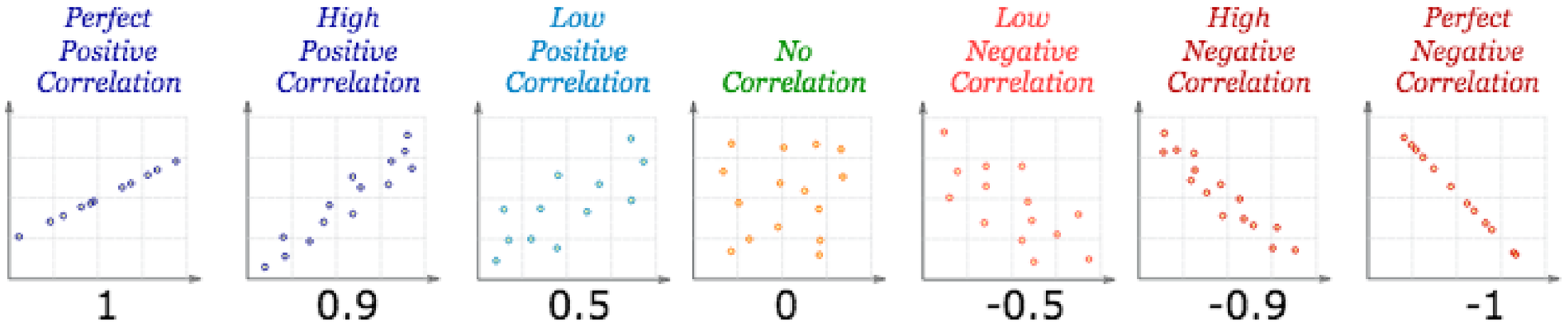




[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]

Scatterplot tasks

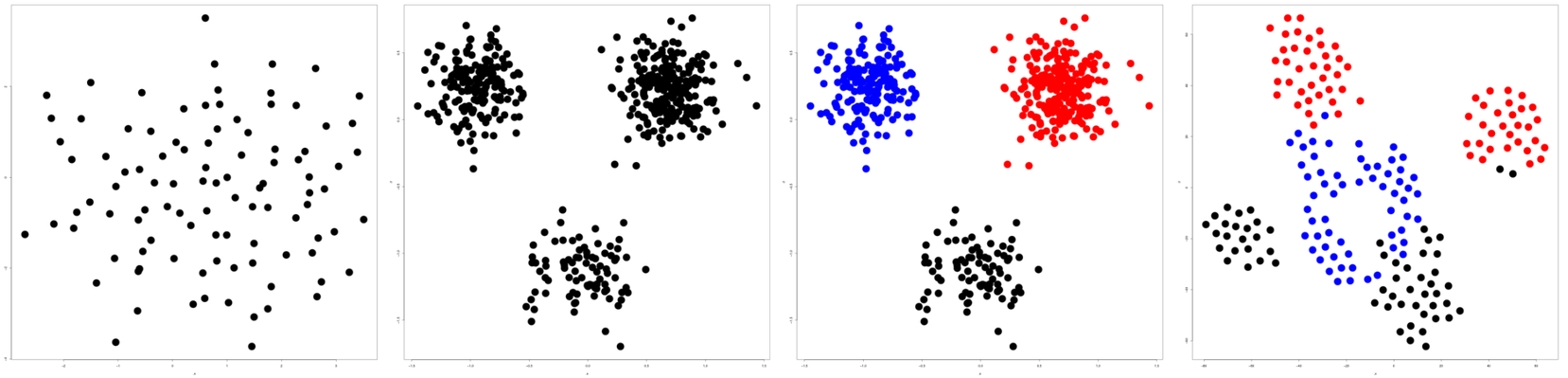
- correlation



sym

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

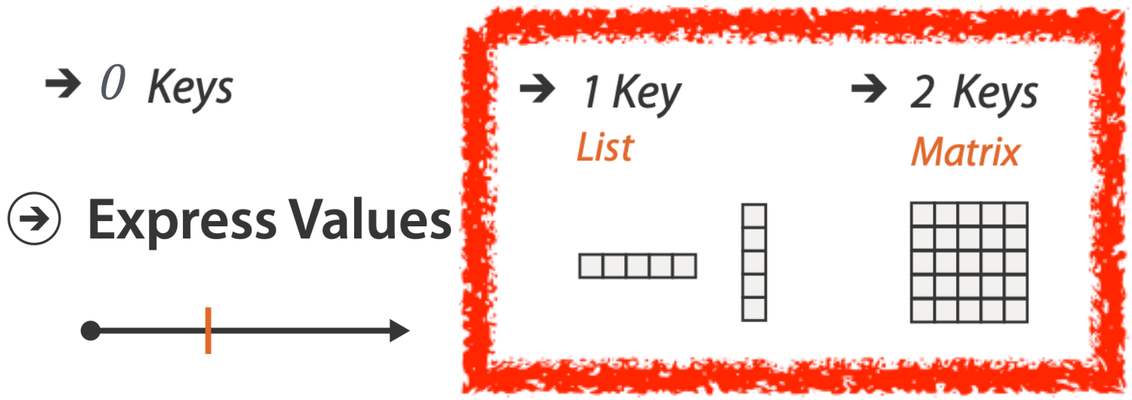
- clusters/groups, and clusters vs classes



pro

https://www.cs.ubc.ca/labs/imager/tr/2014/DRVisTasks/

Some keys



Some keys: Categorical regions

➔ Separate

➔ Order

➔ Align

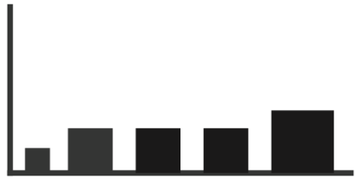
The **expressiveness principle** dictates that the visual encoding should express **all of, and only**, the information in the dataset attributes. The most fundamental expression of this principle is that **ordered data** should be shown in a way that our perceptual system **intrinsically senses as ordered.** Conversely, **unordered data** should not be shown in a way that perceptually implies an **ordering** that does not exist.

Regions: Separate, order, align

➔ Separate

➔ Order

➔ Align



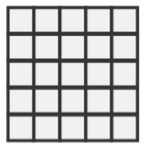
- regions: contiguous bounded areas distinct from each other
  - separate into spatial regions: one mark per region (for now)
- use categorical or ordered attribute to separate into regions
  - no conflict with expressiveness principle for categorical attributes
- use ordered attribute to order and align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

➔ *1 Key*

List

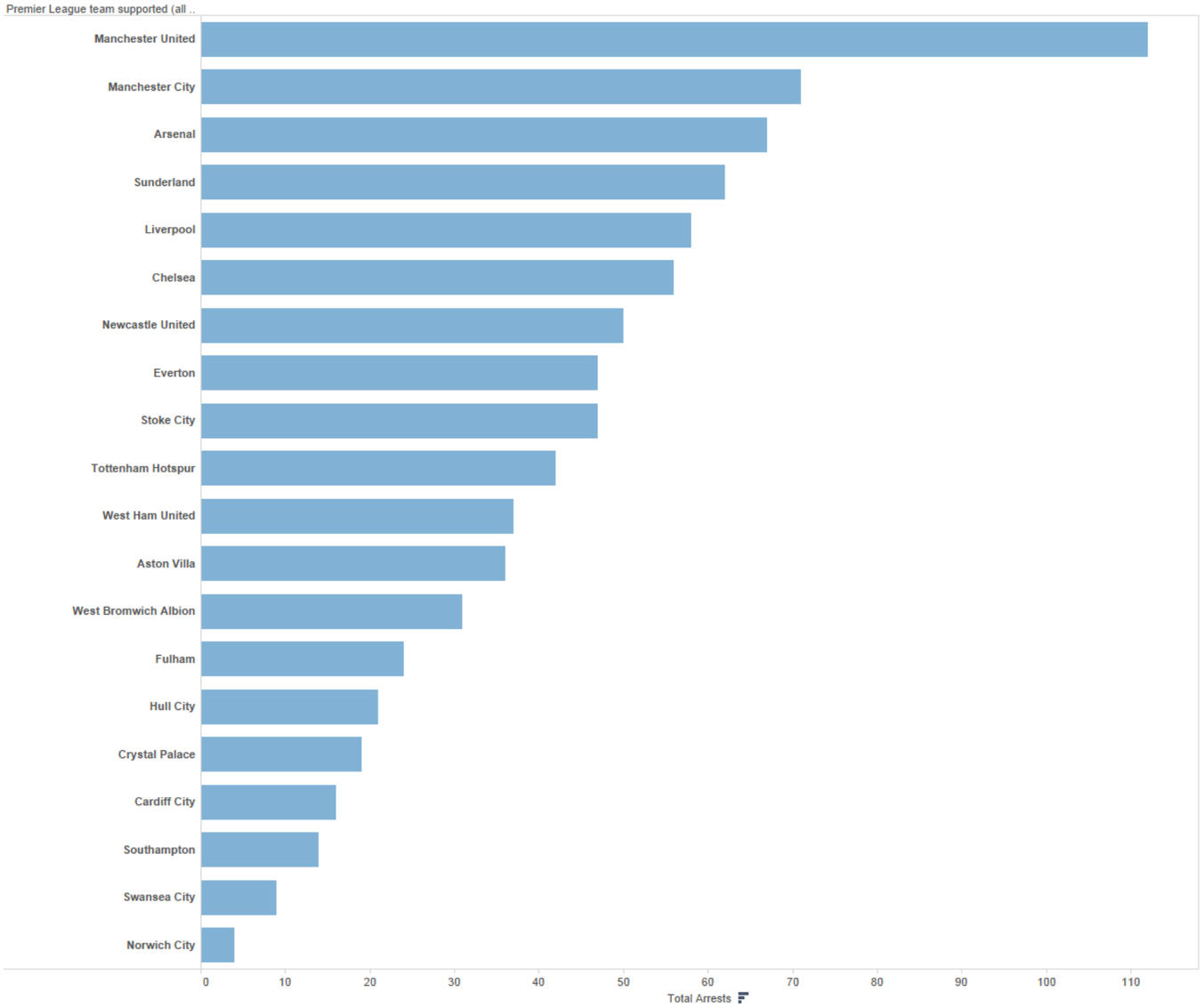
➔ *2 Keys*

Matrix



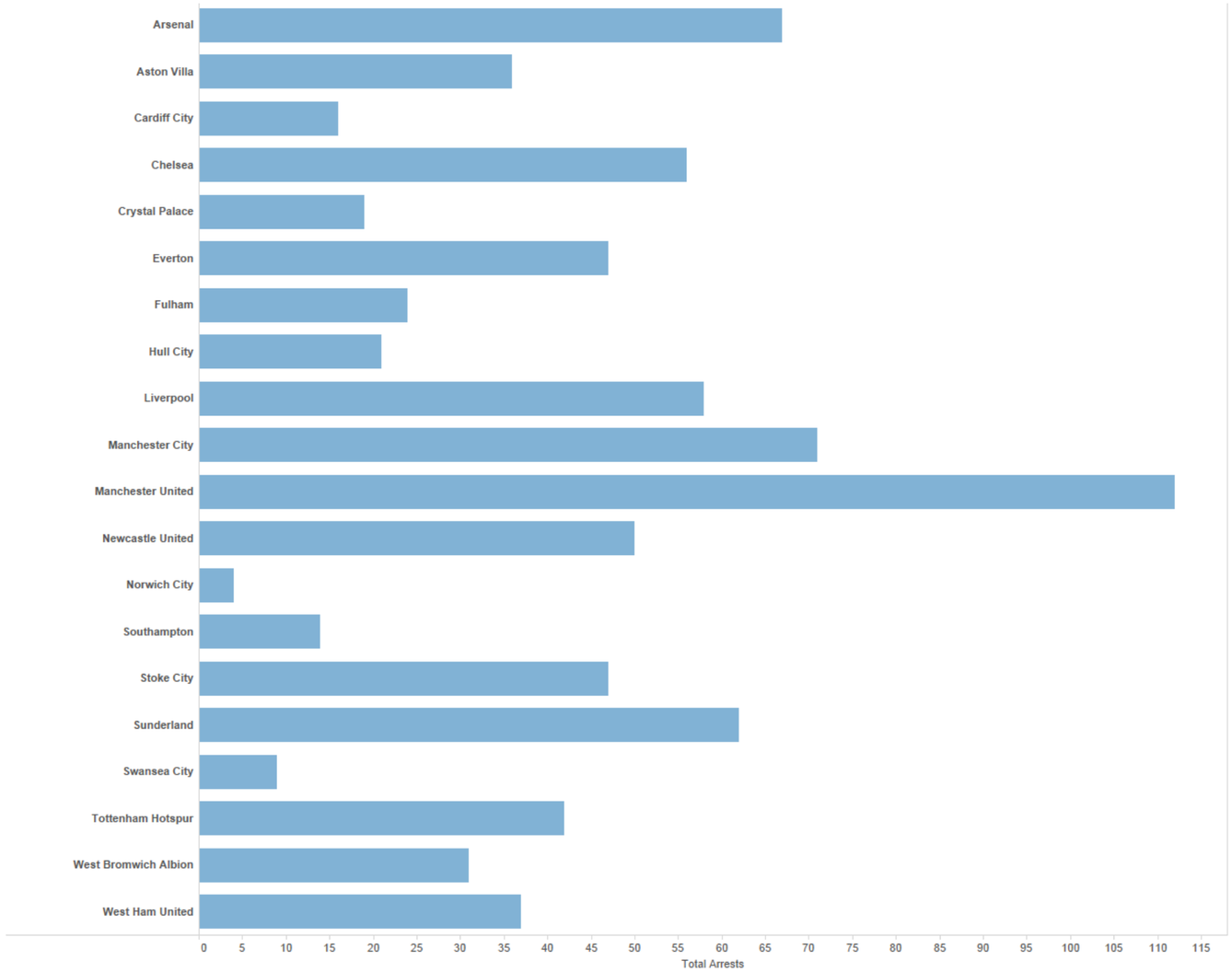
Separated and aligned and ordered

best case



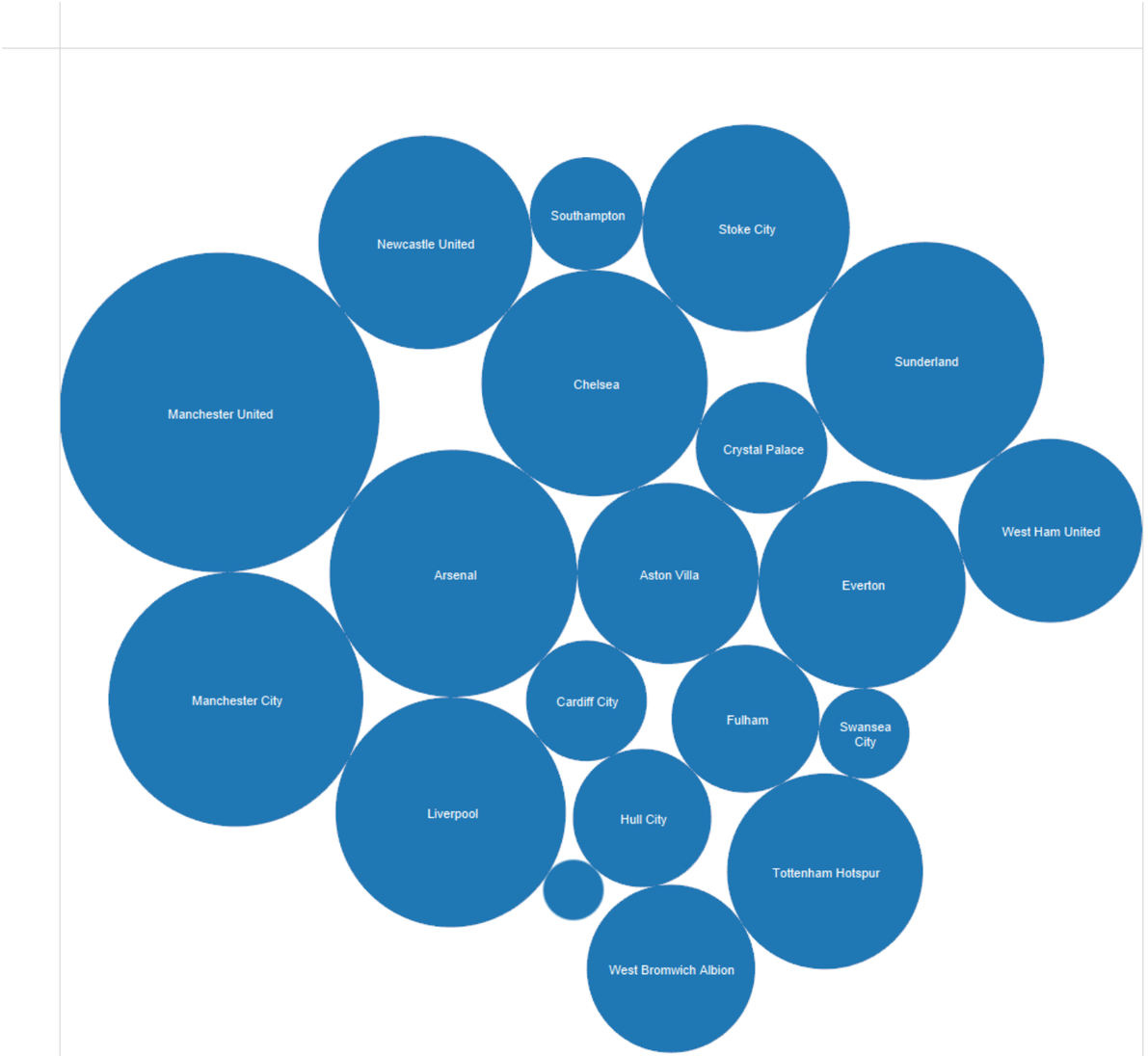
Separated and aligned but not ordered

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



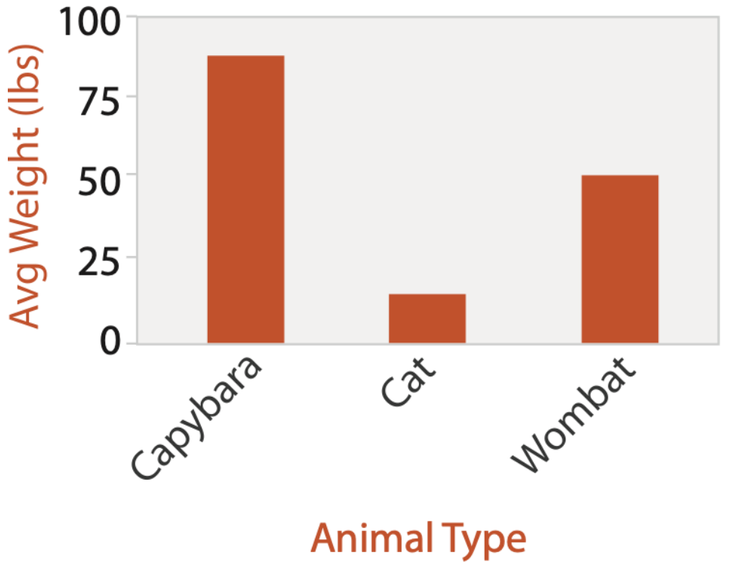
Separated but not aligned or ordered

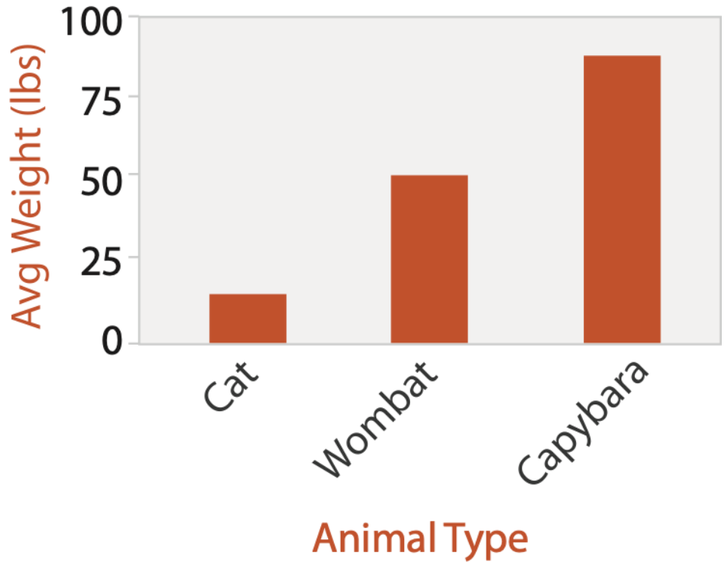
limitation: hard to make comparisons with size (vs aligned position)



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

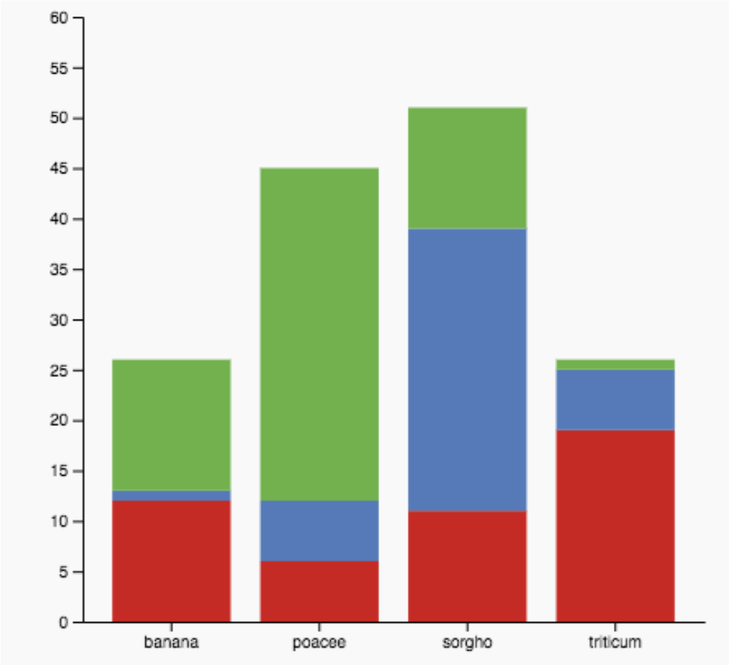




common axis

Idiom: **stacked bar chart**

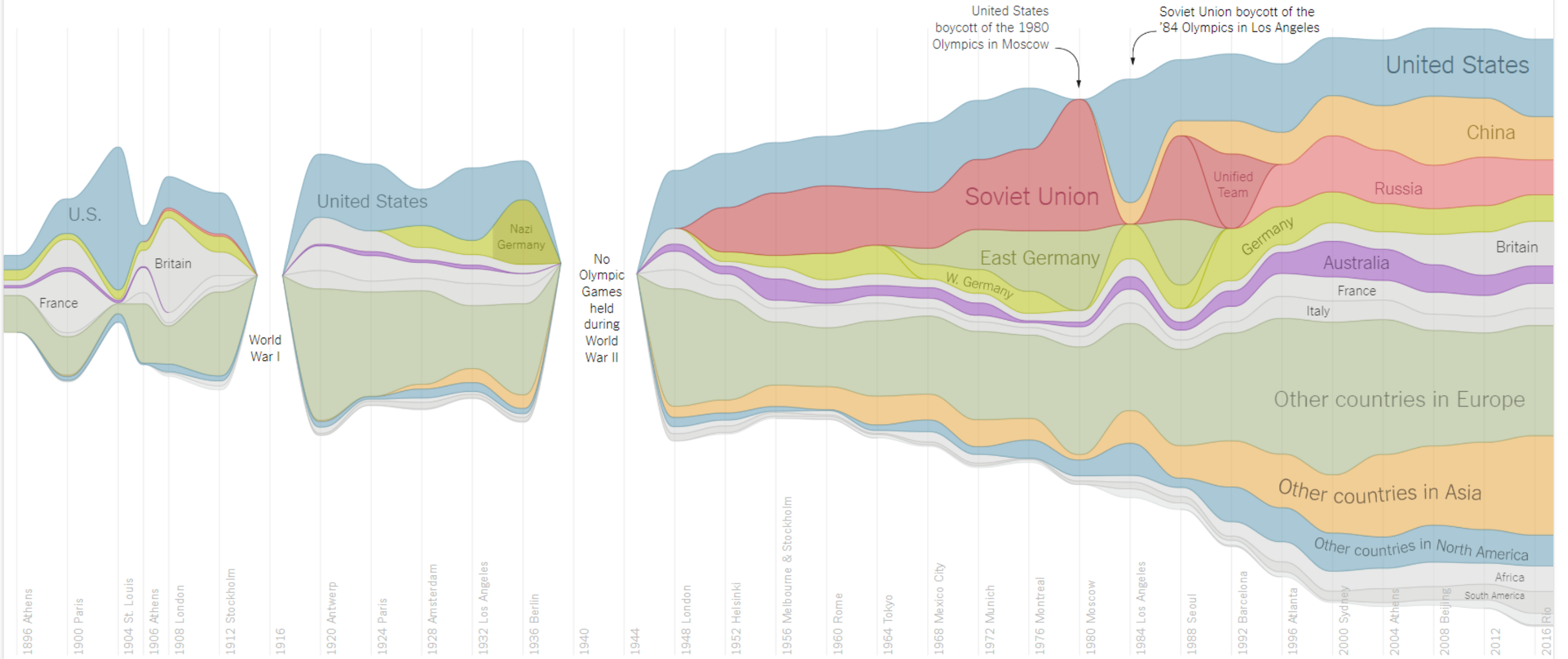
- one more key
  - data
    - 2 categ attrib, 1 quant attrib
  - mark: vertical stack of line marks
    - **glyph**: composite object, internal structure from multiple marks
  - channels
    - length and color hue
    - spatial regions: one per glyph
      - aligned: full glyph, lowest bar component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

Have Dominated the Summer Olympics

By GREGOR AISCH and LARRY BUCHANAN UPDATED August 22, 2016

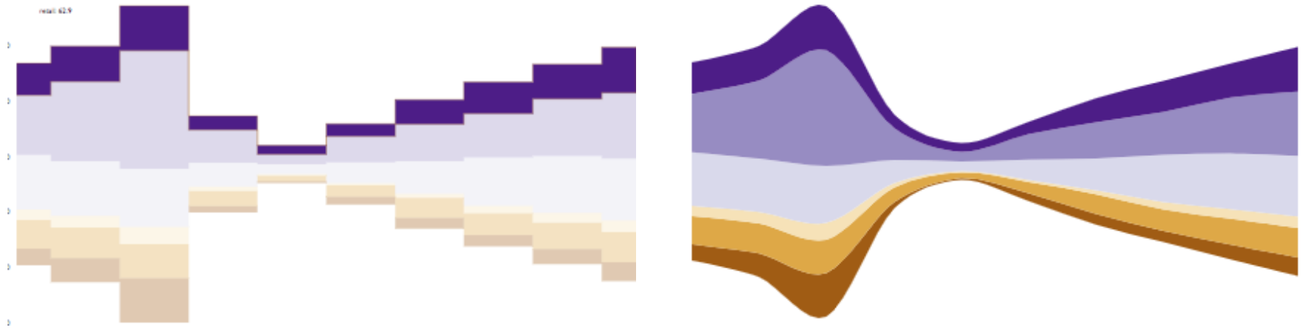


Separated and aligned and ordered

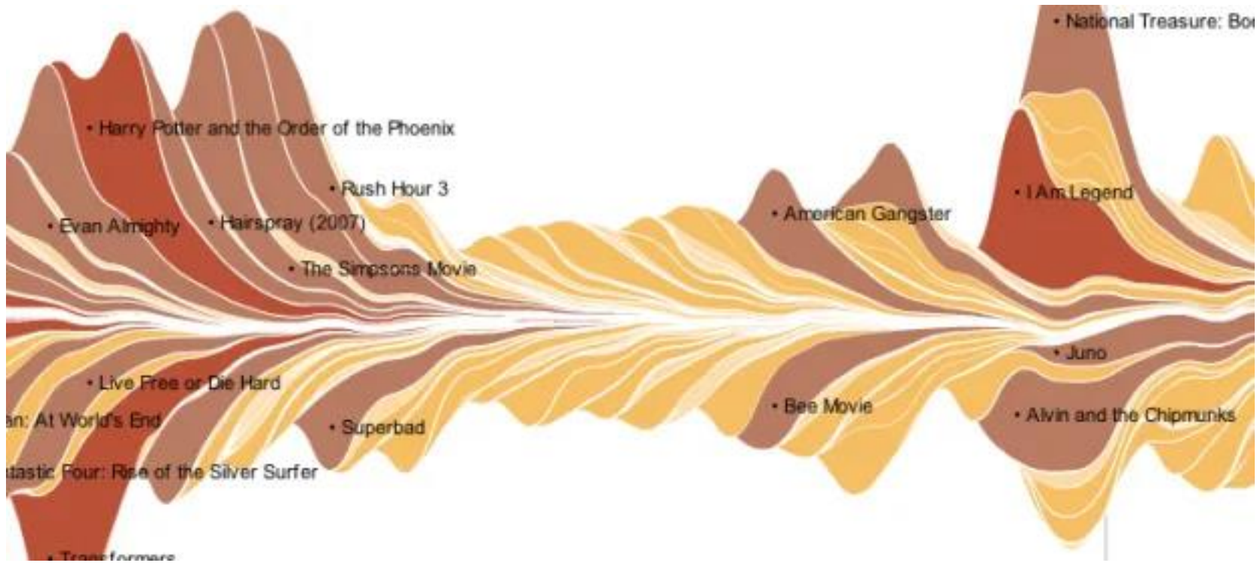
http://www.nytimes.com/interactive/2016/08/08/sports/olympics/history-olympic-dominance-charts.html

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



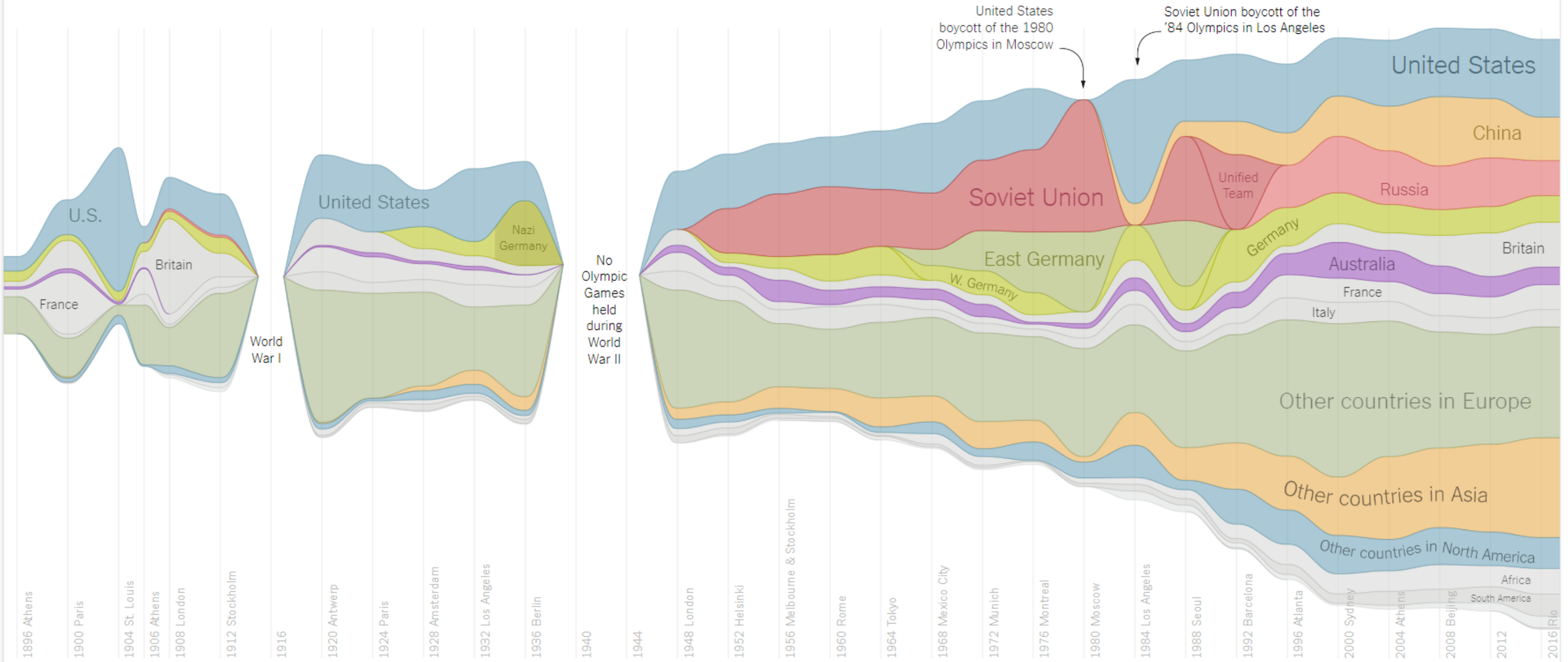
[Stacked Graphs Geometry & Aesthetics. Byron and Wattenberg. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2008) 14(6): 1245–1252, (2008).]



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

Have Dominated the Summer Olympics

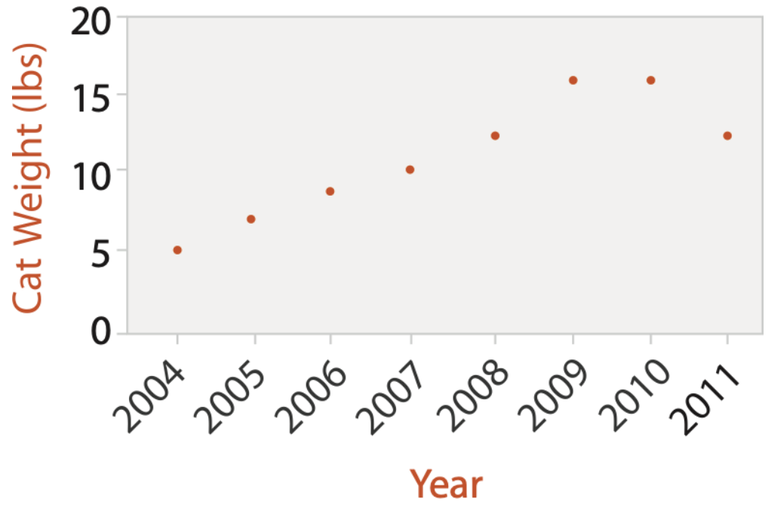
By GREGOR AISCH and LARRY BUCHANAN UPDATED August 22, 2016

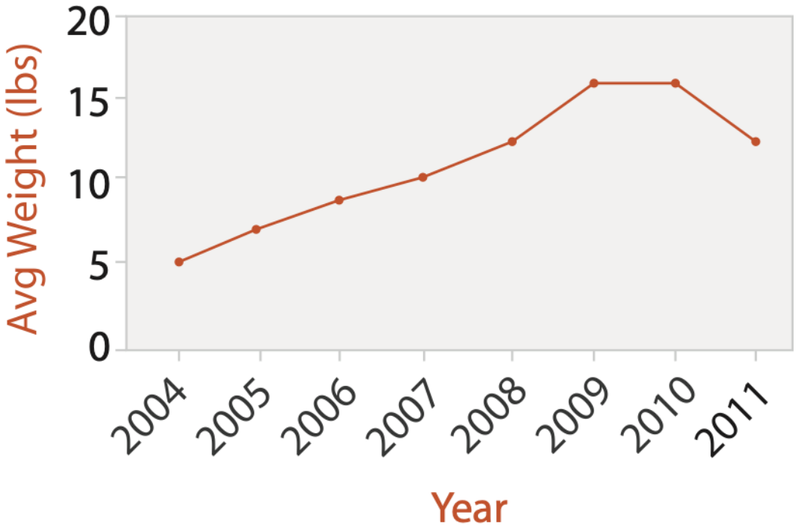


http://www.nytimes.com/interactive/2016/08/08/sports/olympics/history-olympic-dominance-charts.html

Idiom: **dot / line chart**

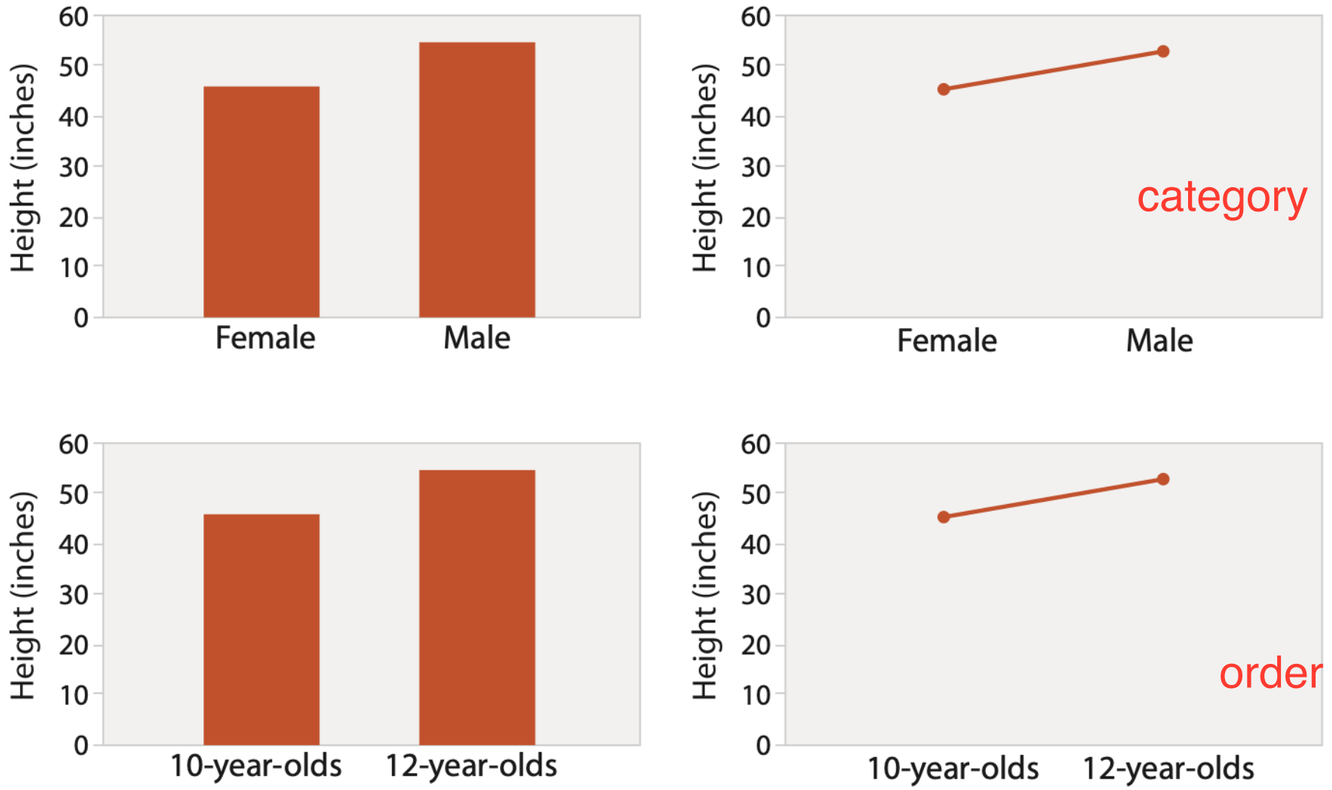
- one key, one value
  - data
    - 2 quant attribs
  - mark: points *AND line connection marks between them*
  - channels
    - aligned lengths to express quant value
    - separated and ordered by key attrib into horizontal regions
  - task
    - find trend
      - connection marks emphasize ordering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

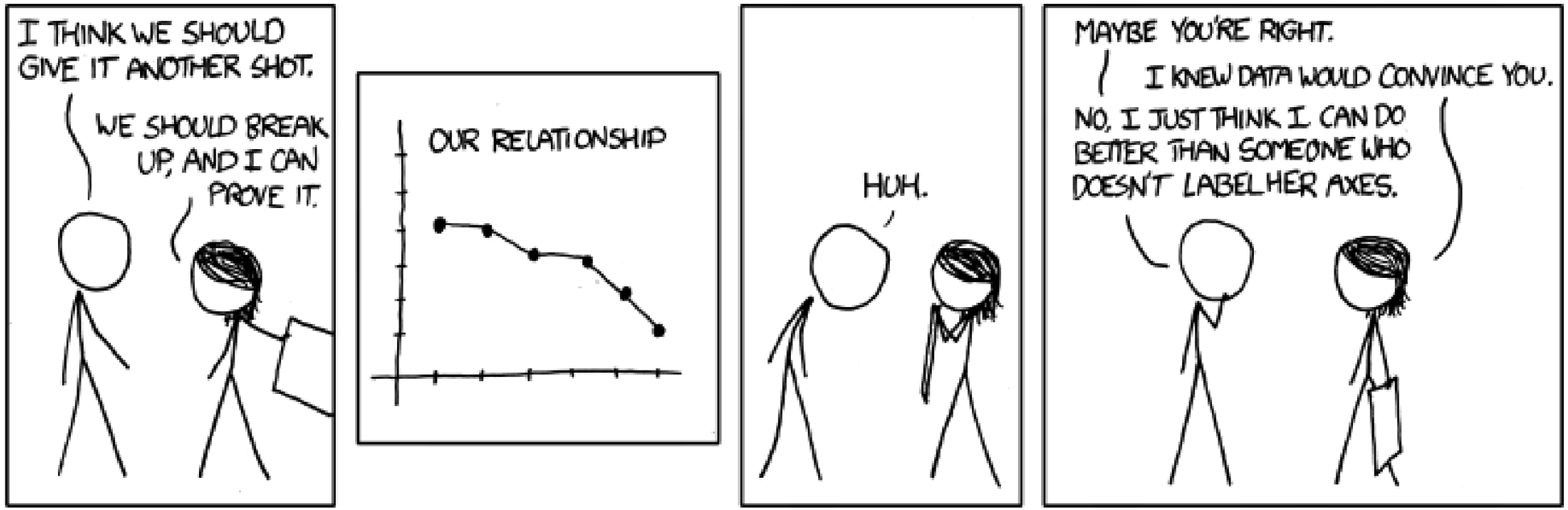
- depends on type of key attrib
  - bar charts if categorical
  - line charts if ordered
- do not use line charts for categorical key attribs
  - violates expressiveness principle
    - implication of trend so strong that it overrides semantics!
      - "The more male a person is, the taller he/she is"



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

Chart axes: label them!

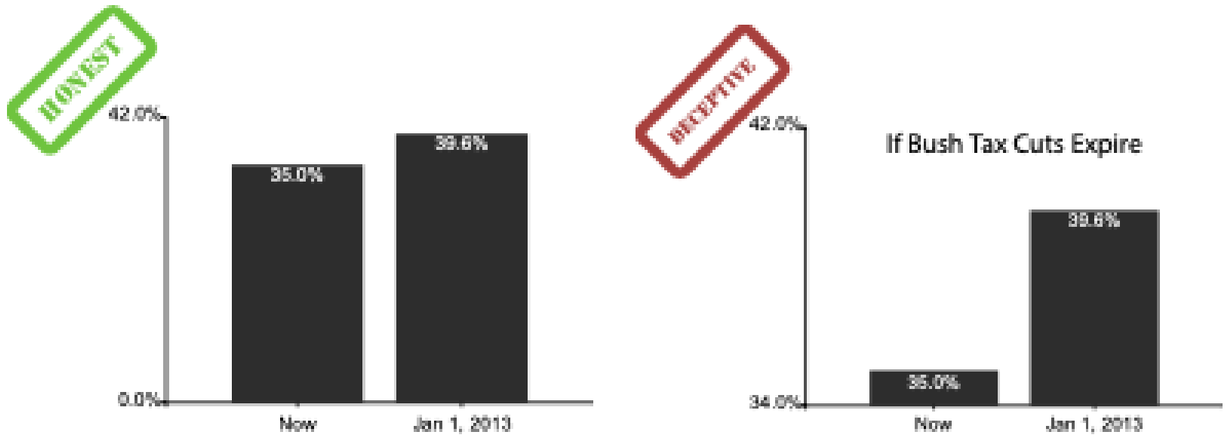
- best practice to label
  - few exceptions: individual small multiple views could share axis label



https://xkcd.com/833/

Chart axes: avoid cropping y axis

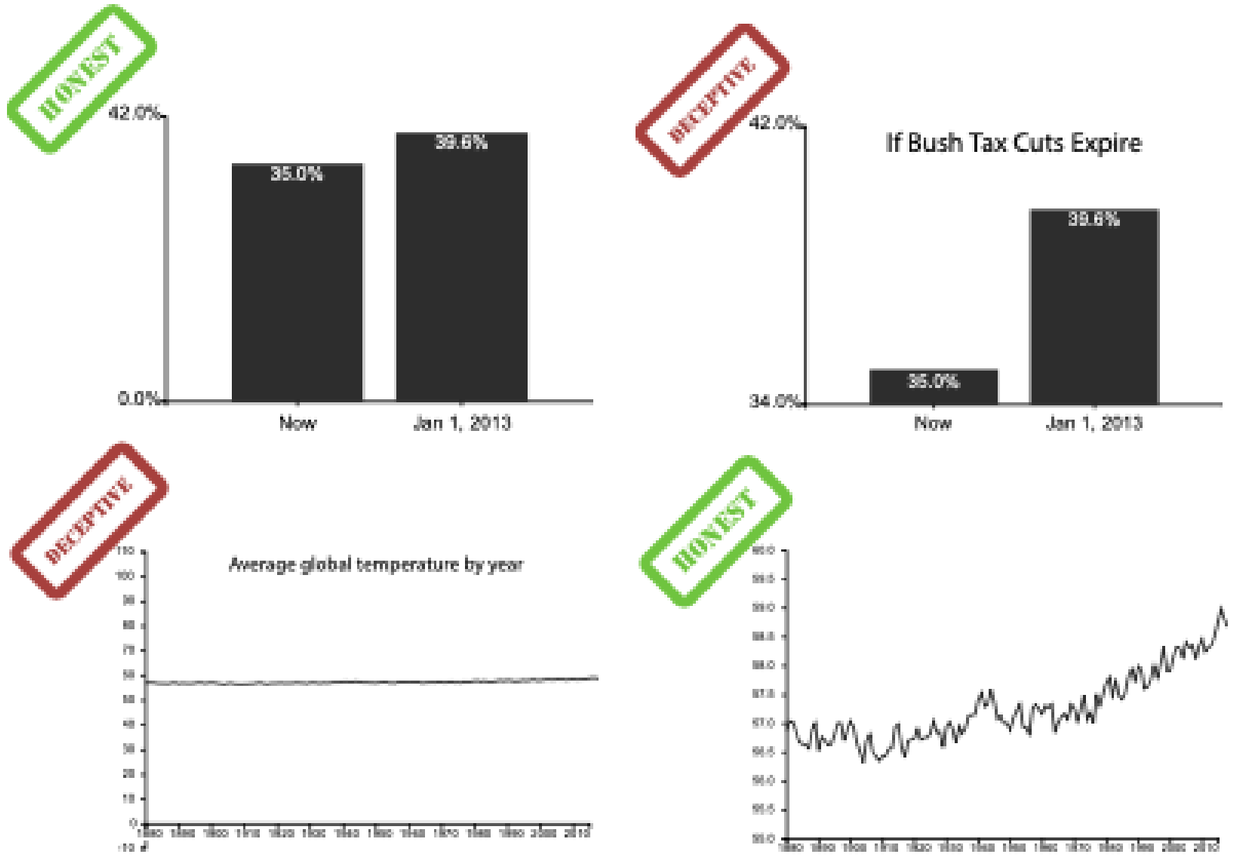
- include 0 at bottom left or slope misleads



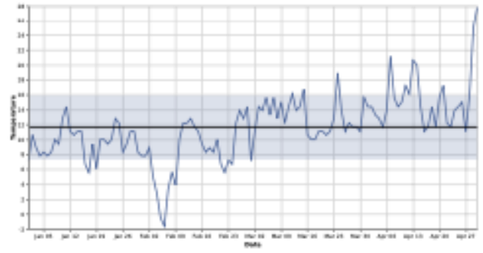
[Truncating the Y-Axis: Threat or Menace?
Correll, Bertini, & Franconeri, CHI 2020.]

Chart axes: avoid cropping y axis

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

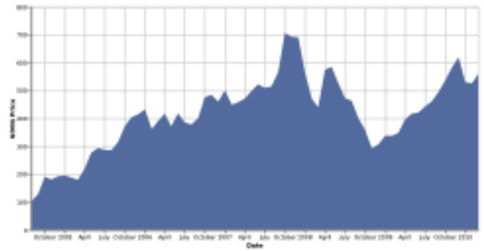


[Truncating the Y-Axis: Threat or Menace?
Correll, Bertini, & Franconeri, CHI 2020.]

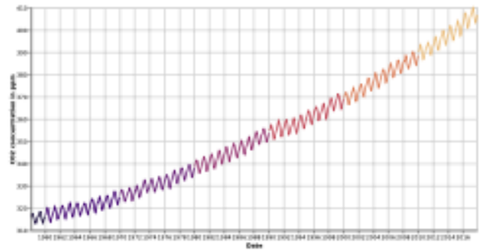


(a) Statistical process charts rely on comparison to an expected value, and so deviations from that value, not from zero, are important

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



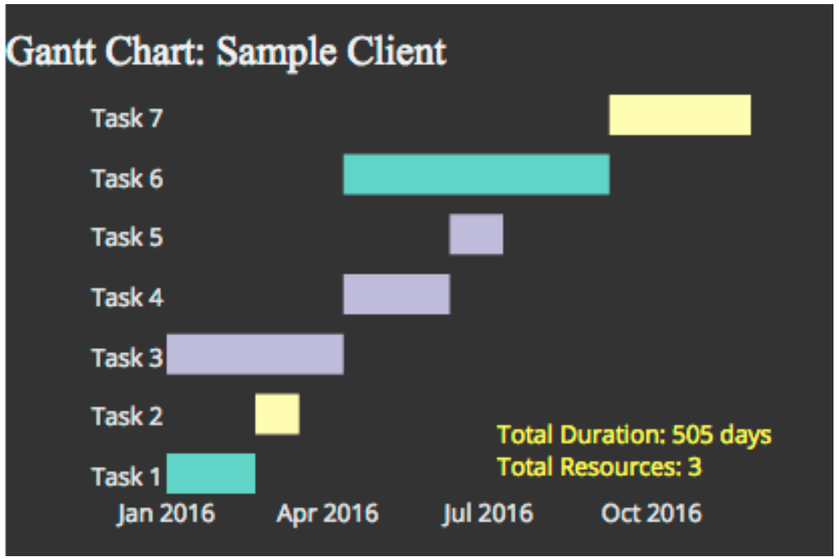
(c) Stock charts must show small differences in stock value, as these can translate to enormous monetary gains or losses.



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

Idiom: **Gantt charts**

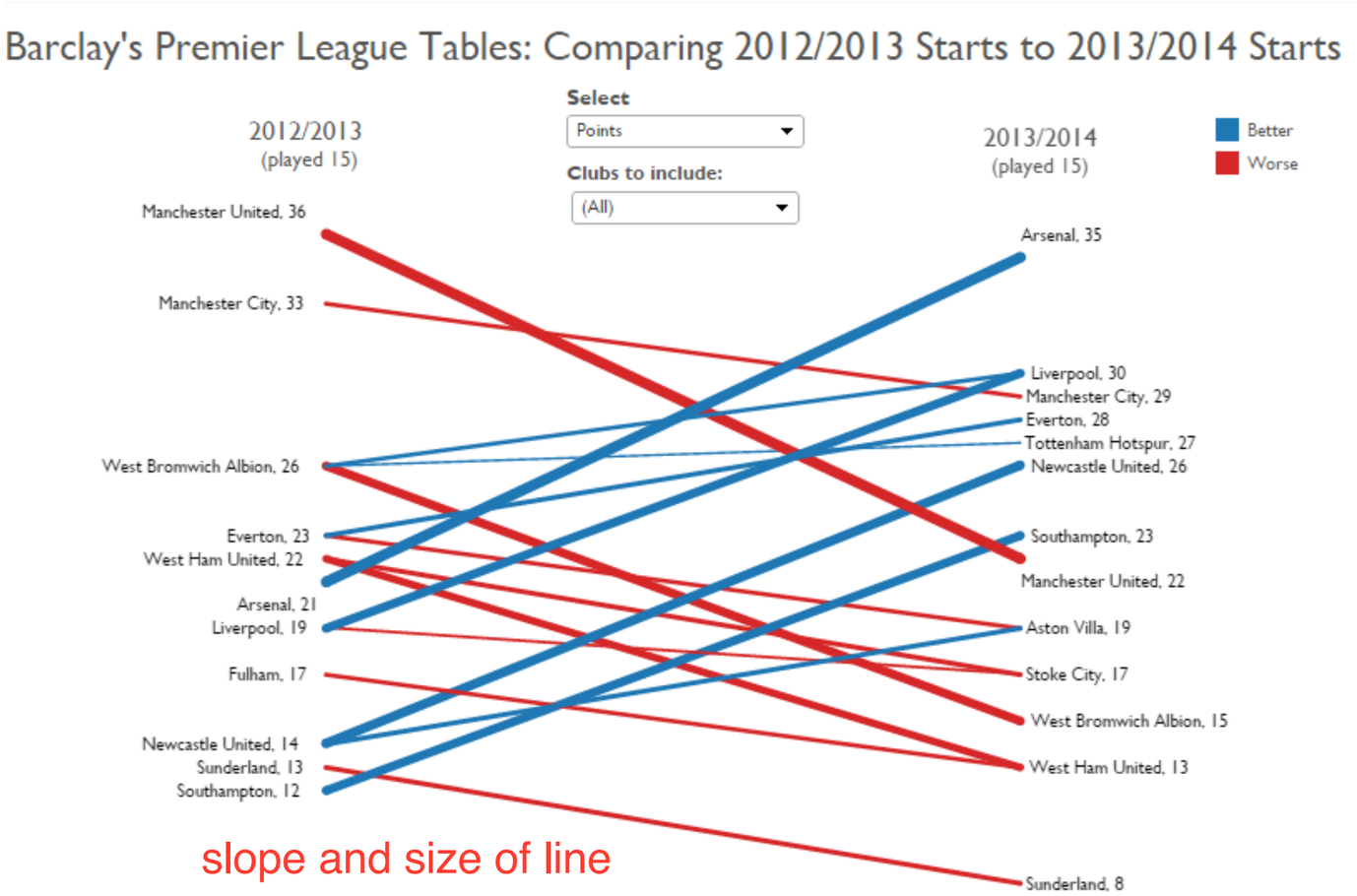
- one key, two (related) values
  - data
    - 1 categ attrib, 2 quant attribs
  - mark: line
    - length: duration
  - channels
    - horiz position: start time (+end from duration)
  - task
    - emphasize temporal overlaps & start/end dependencies between items
  - scalability
    - dozens of key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



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

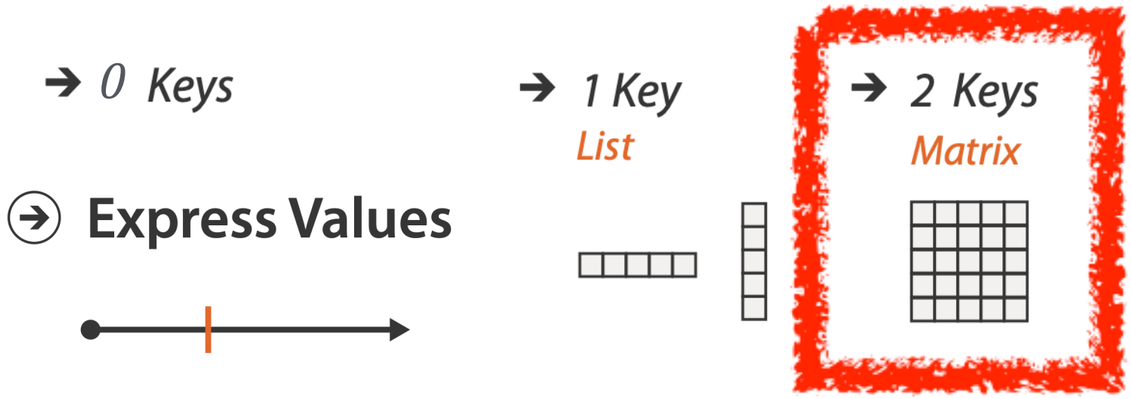
Idiom: **Slopegraphs**

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



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

2 Keys

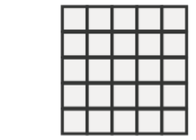


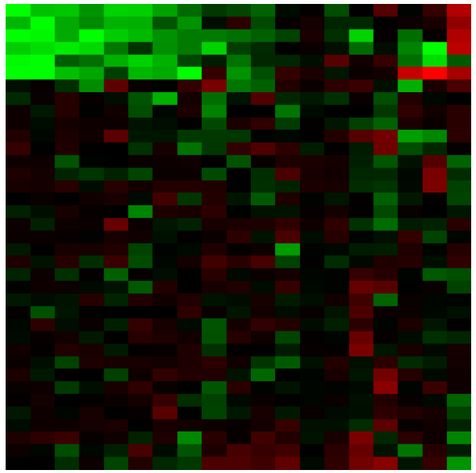
Idiom: **heatmap**

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

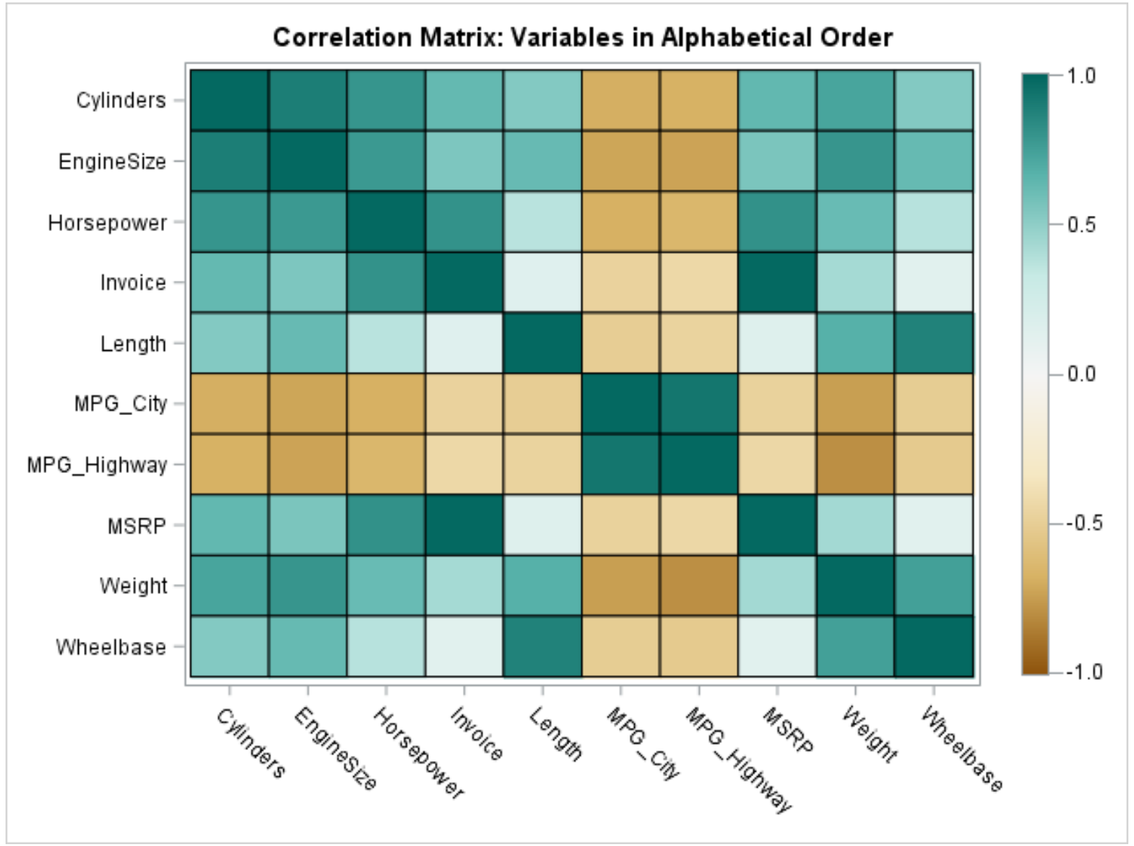
➔ *2 Keys*

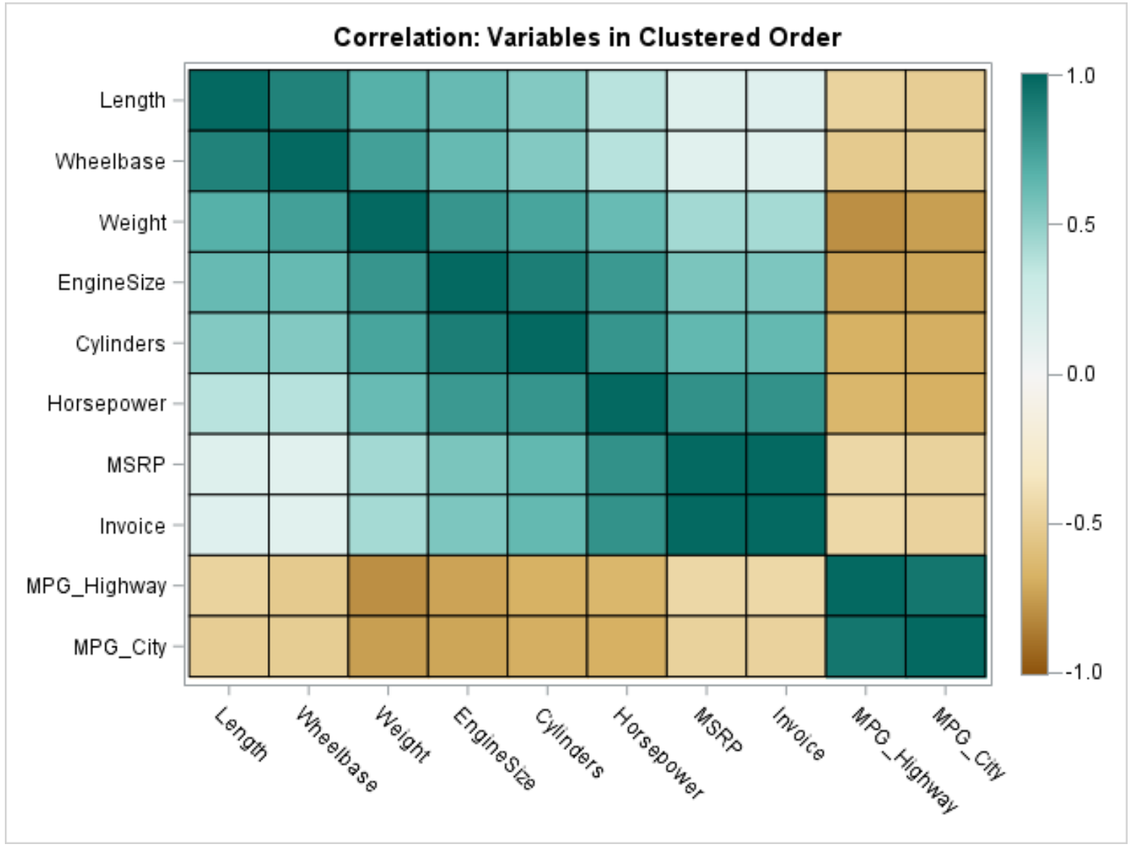
Matrix





Heatmap reordering

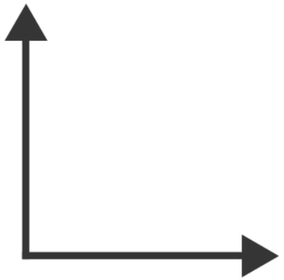




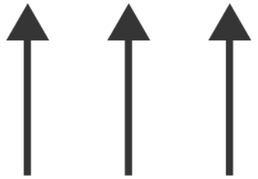
https://blogs.sas.com/content/iml/2018/05/02/reorder-variables-correlation-heat-map.html

Axis Orientation

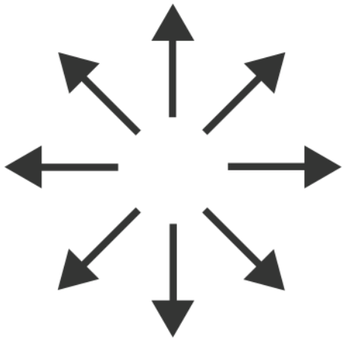
- Rectilinear



- Parallel

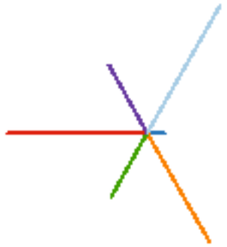


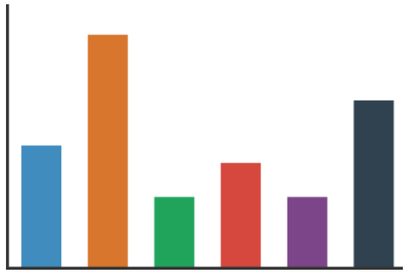
- Radial



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



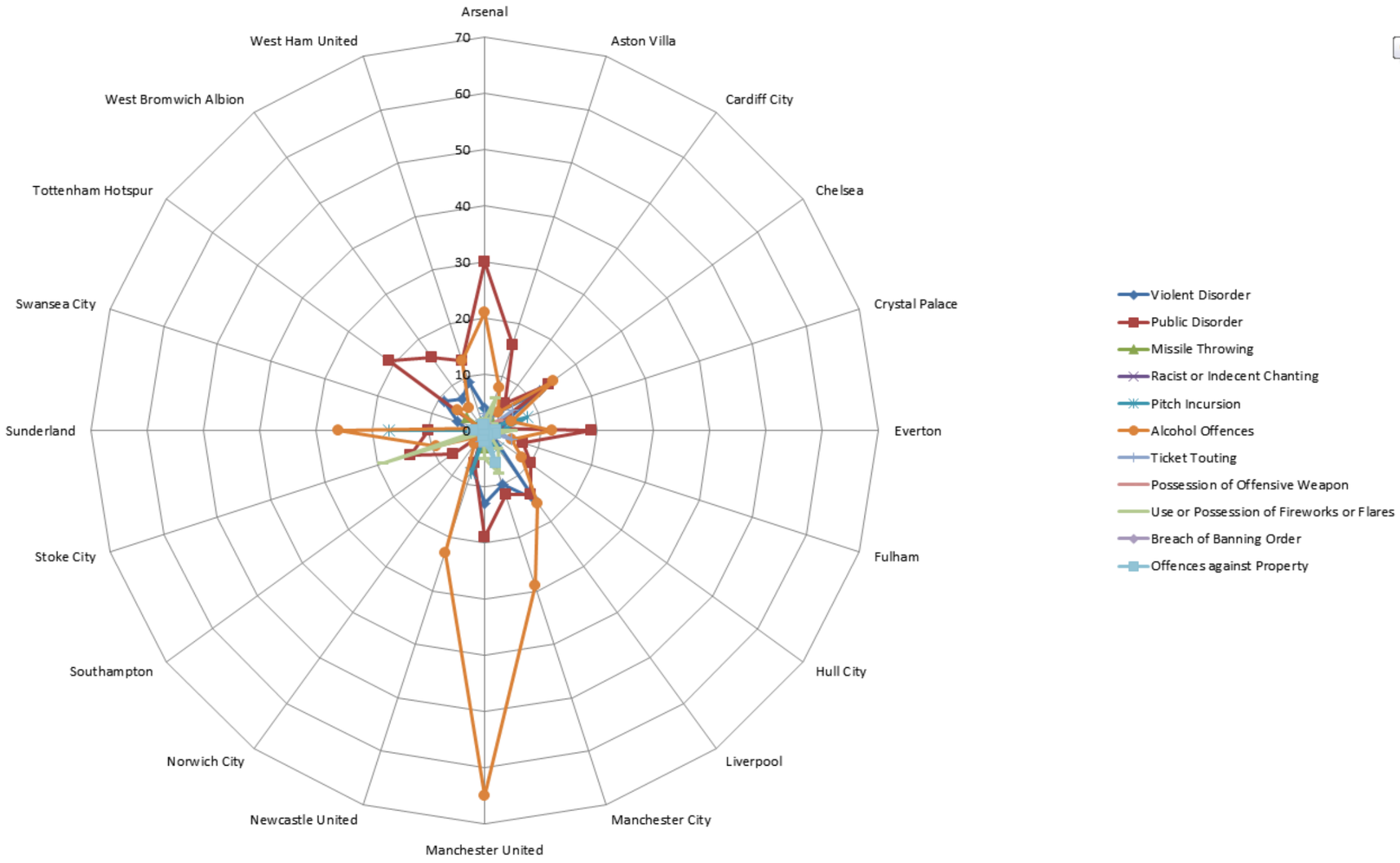


- accuracy
  - length not aligned with radial layouts
    - less accurately perceived than rectilinear aligned

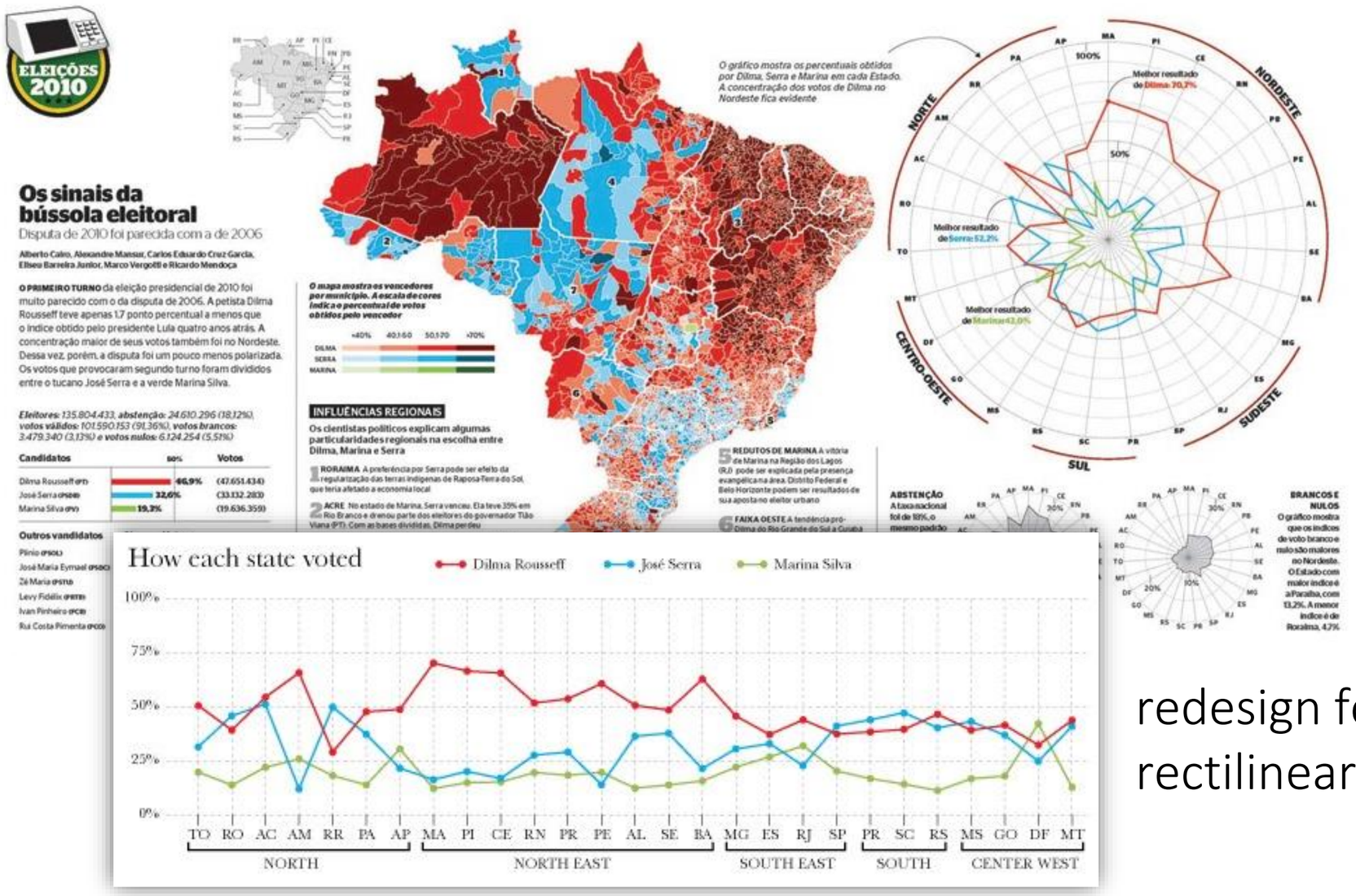
[Vismon: Facilitating Risk Assessment and Decision Making In Fisheries Management. Booshehrian, Möller, Peterman, and Munzner. Technical Report TR 2011-04, Simon Fraser University, School of Computing Science, 2011.]

Idiom: **radar plot**

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



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



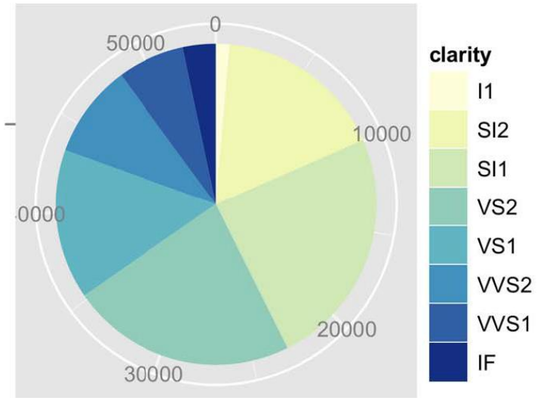
original
difficult to interpret

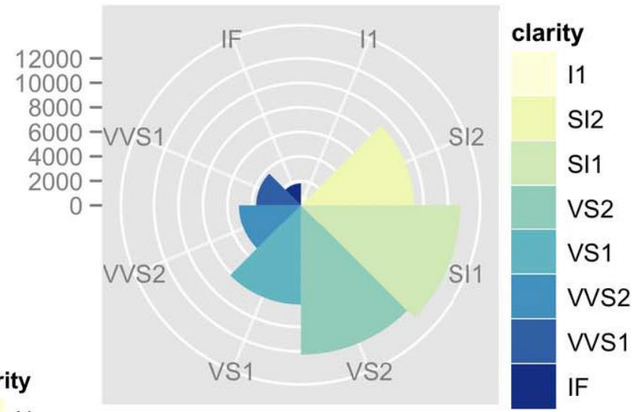
redesign for
rectilinear

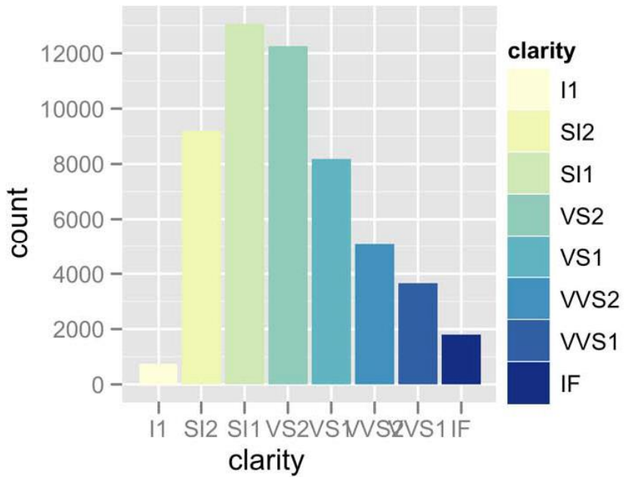
http://www.thefunctionalart.com/2012/11/radar-graphs-avoid-them-999-of-time.html

Idioms: **pie chart, coxcomb chart**

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



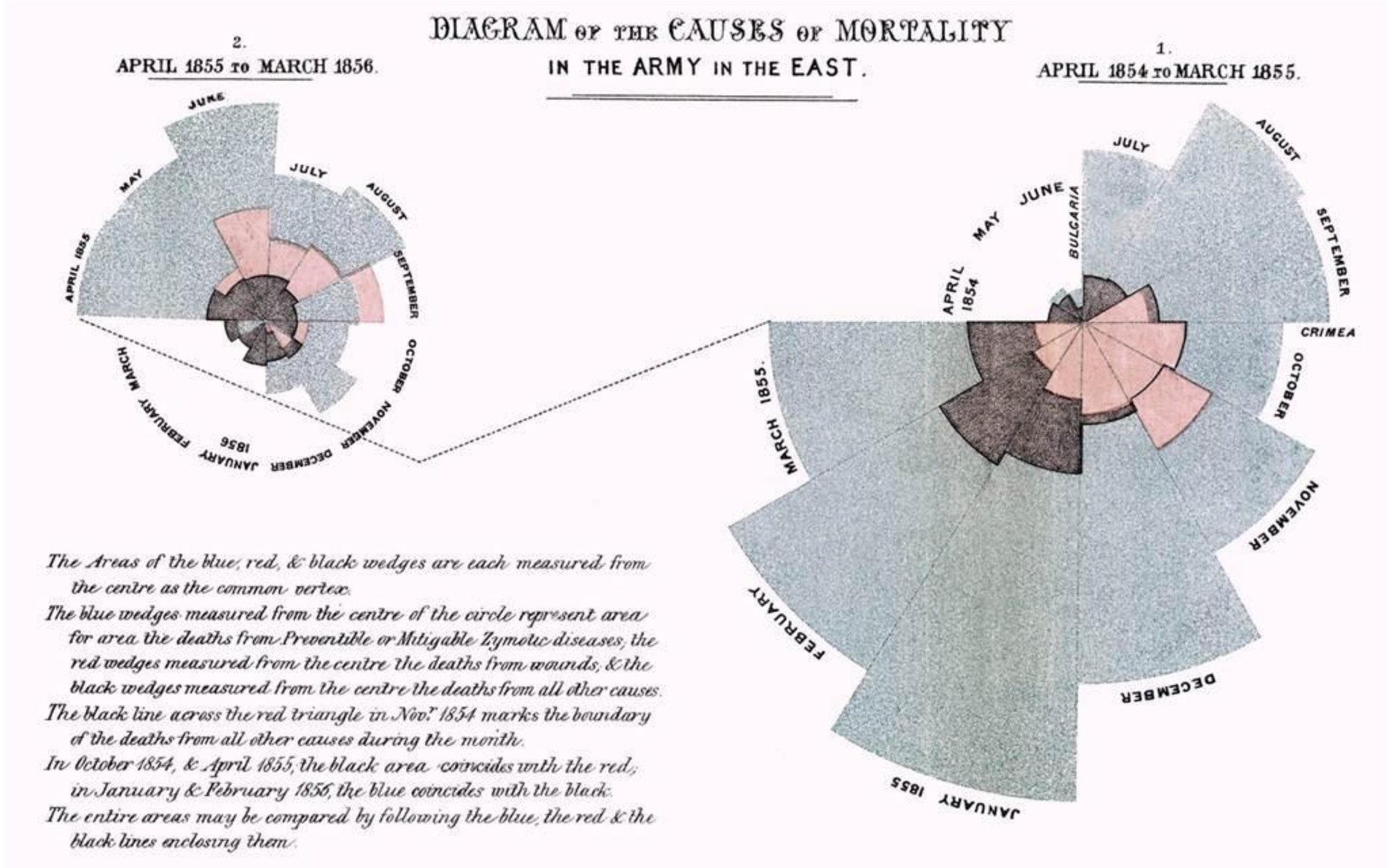




[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]

Coxcomb / nightingale rose / polar area chart

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

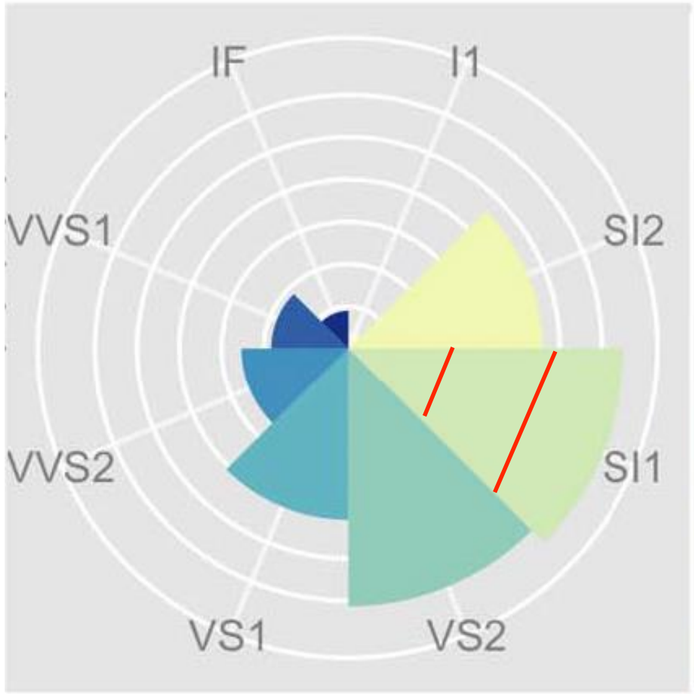


Coxcomb: perception

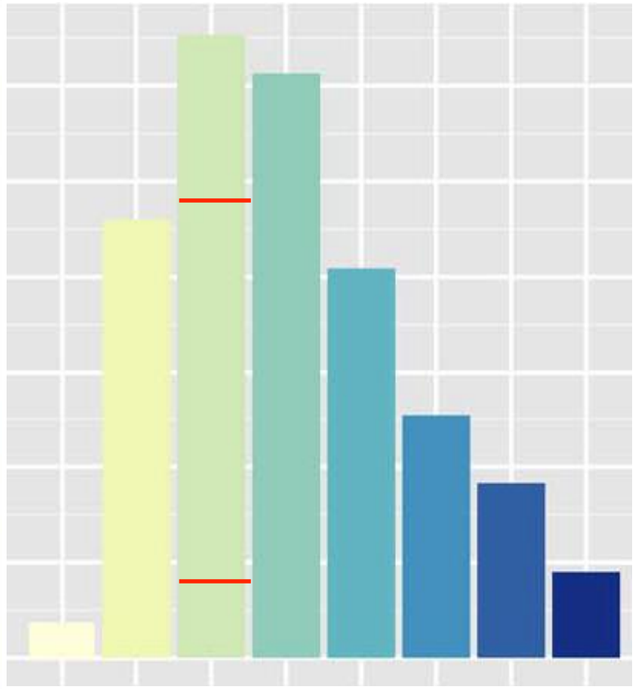
- encode: 1D length
- decode/perceive: 2D area

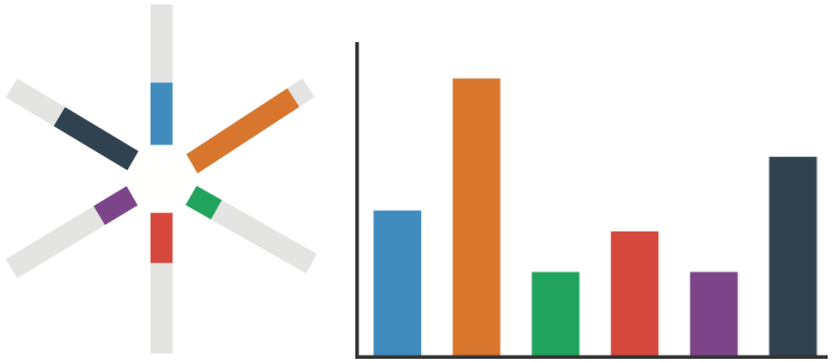
- nonuniform line/sector width as length increases
  - so area variation is nonlinear wrt line mark length!

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



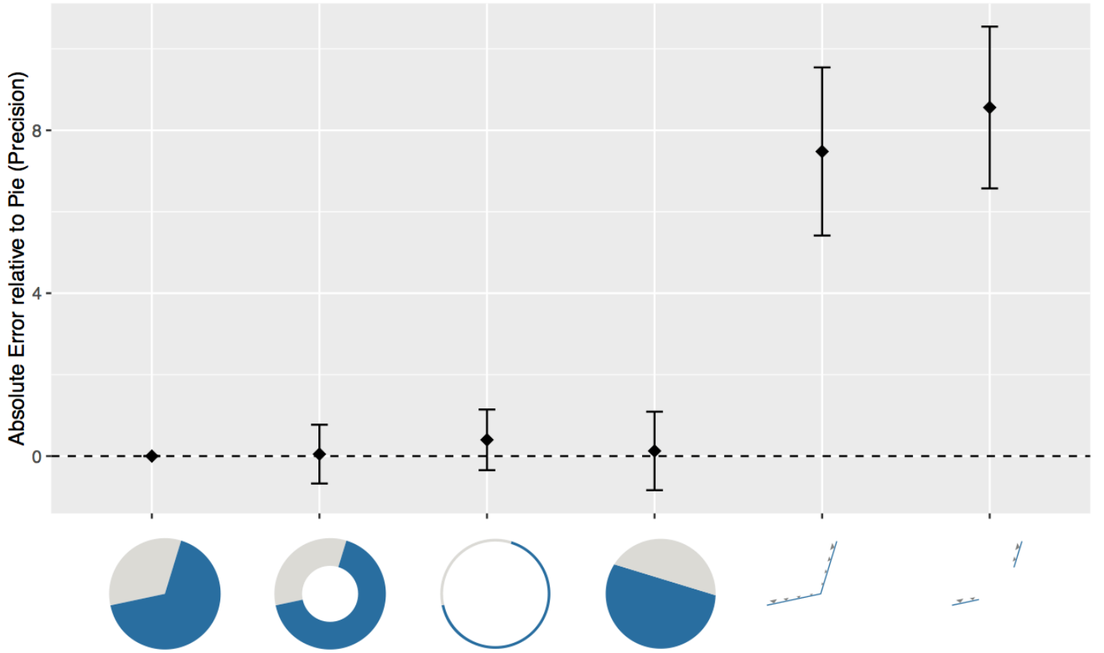
nonuniform width as length increases


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
  - decode/perceive: not angles
  - maybe also areas?…
- donut charts no worse than pie charts

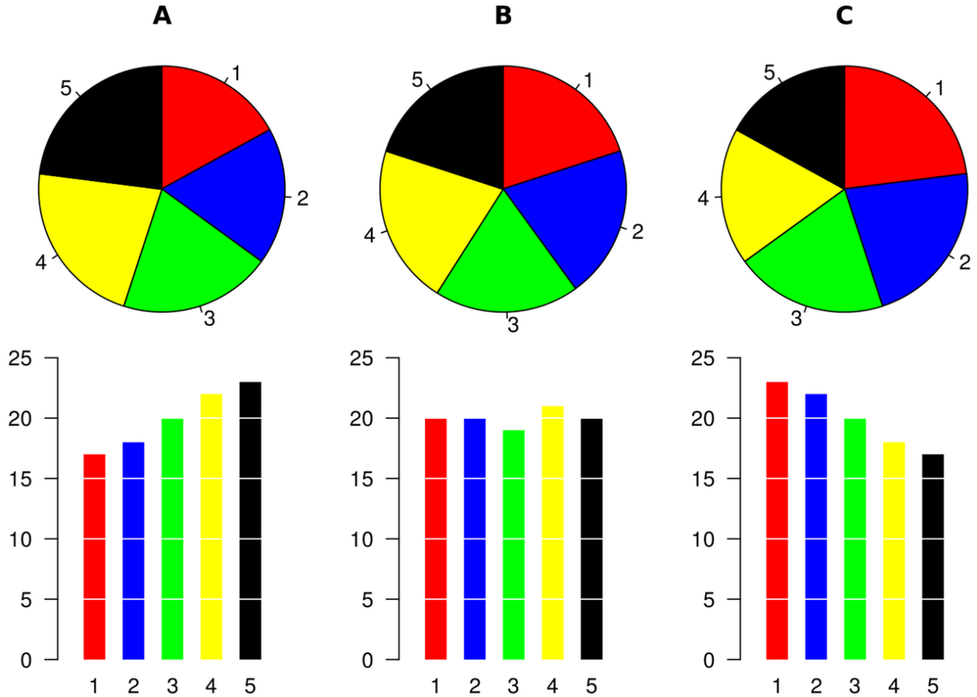


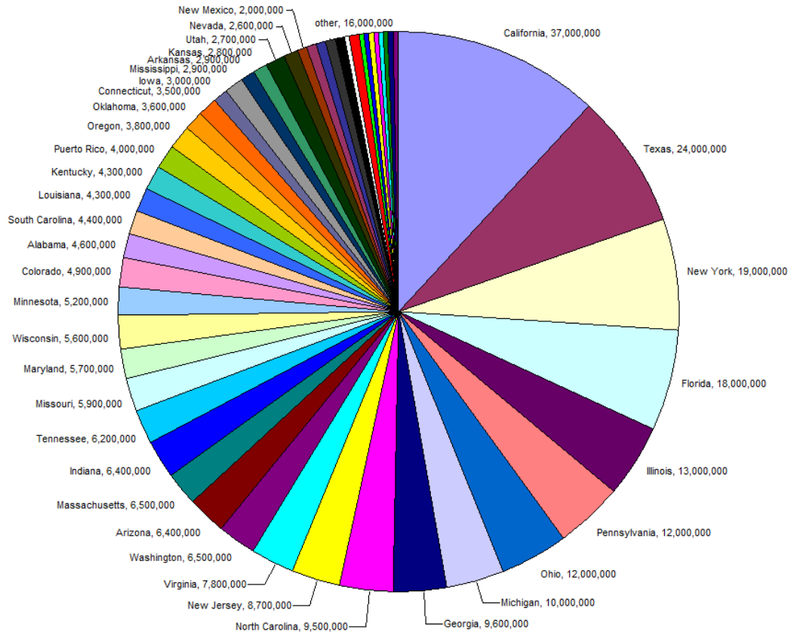
[Arcs, Angles, or Areas: Individual Data Encodings in Pie and Donut Charts. Skau and Kosara. Proc. EuroVis 2016.]

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

Pie charts: best practices

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

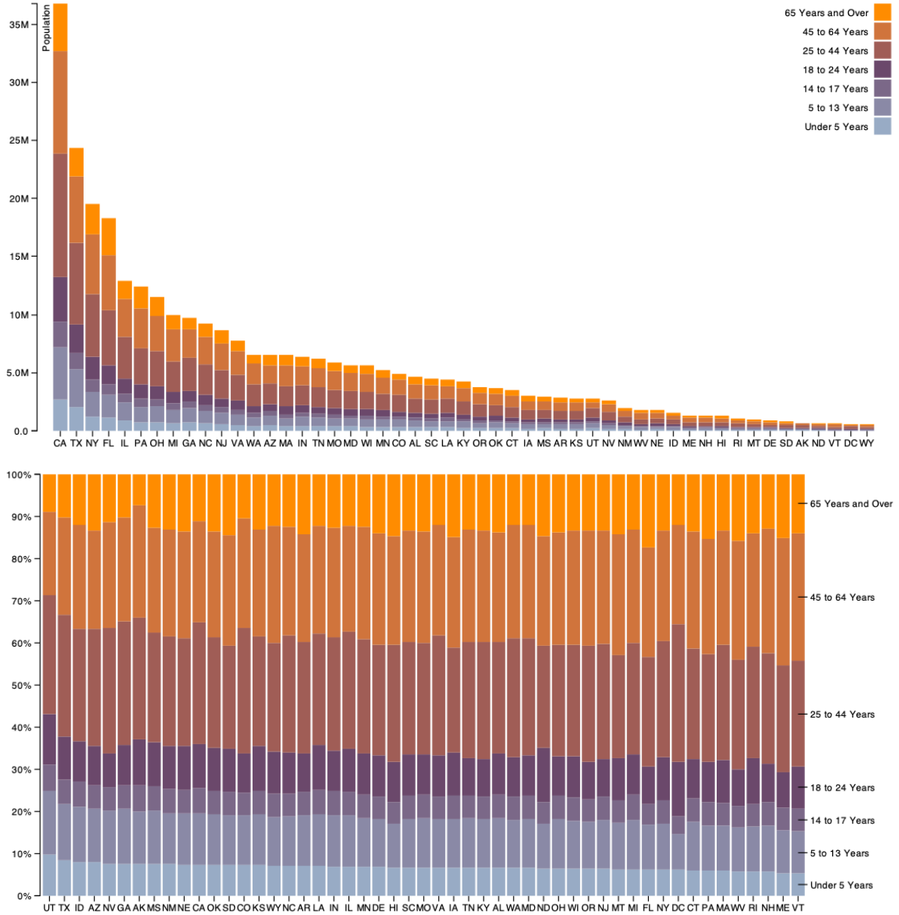


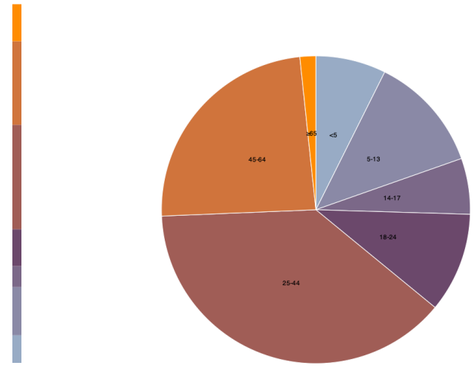


https://eagereyes.org/pie-charts

Idioms: **normalized stacked bar chart**

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





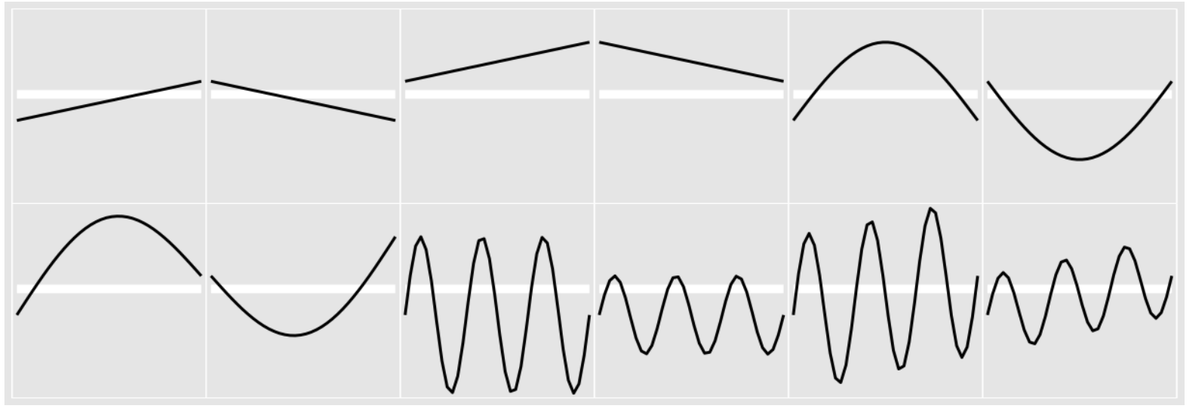
http://bl.ocks.org/mbostock/3886208,

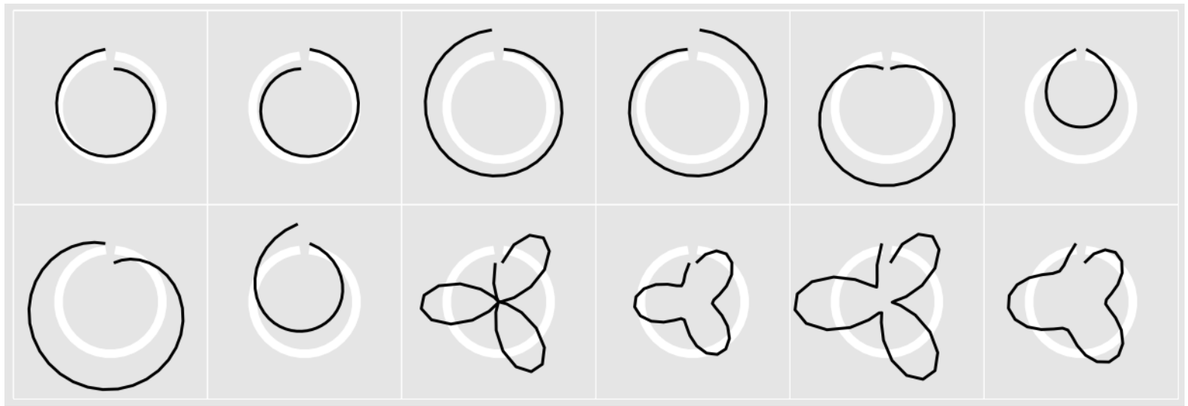
http://bl.ocks.org/mbostock/3887235

Idiom: **glyphmaps**

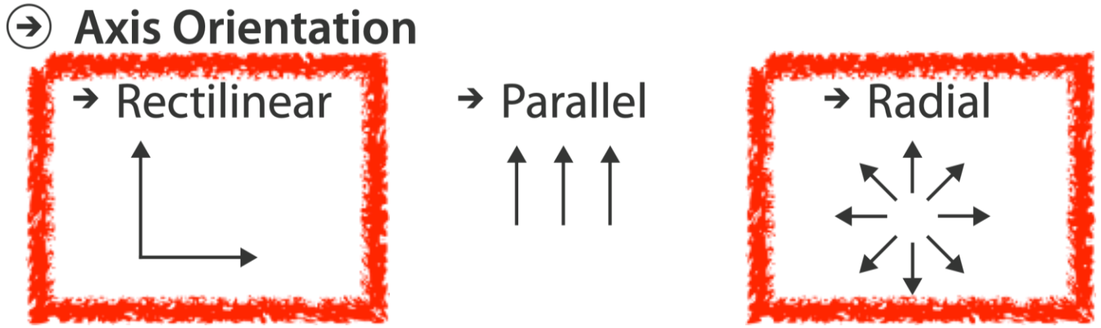
- rectilinear good for linear vs nonlinear trends

- radial good for cyclic patterns
  - evaluating periodicity



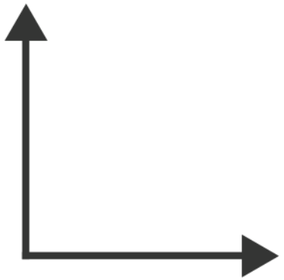


[Glyph-maps for Visually Exploring Temporal Patterns in Climate Data and Models. Wickham, Hofmann, Wickham, and Cook. Environmetrics 23:5 (2012), 382–393.]

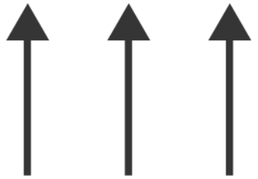


Axis Orientation

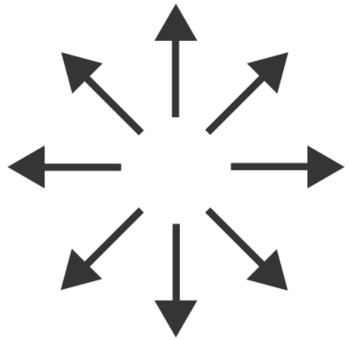
- Rectilinear



- Parallel



- Radial



Idiom: **SPLOM**

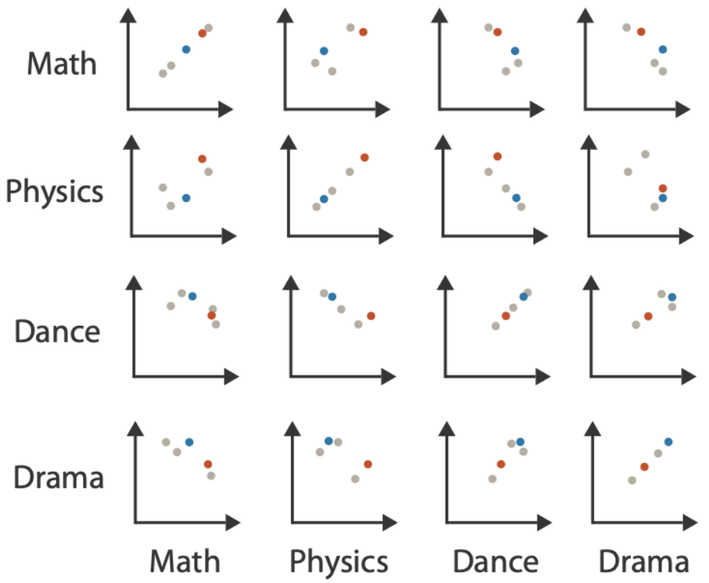
- scatterplot matrix (SPLOM)
  - rectilinear axes, point mark
  - all possible pairs of axes
  - scalability
    - one dozen attribs
    - dozens to hundreds of items

Wilkinson et al., 2005

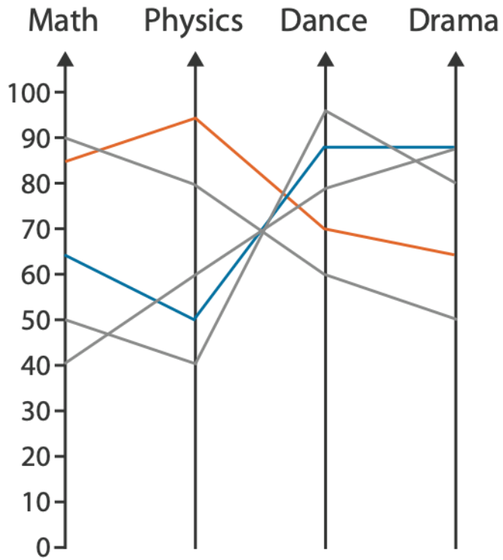
Idioms: **parallel coordinates**

- 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



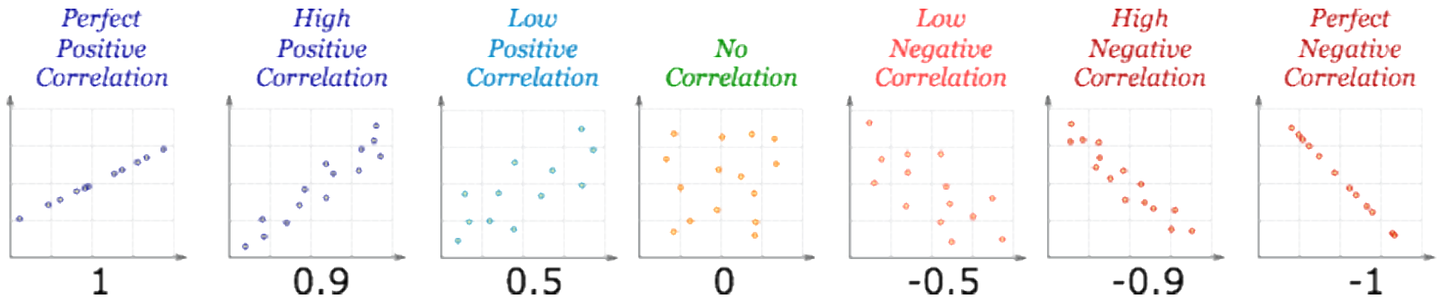
Table

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

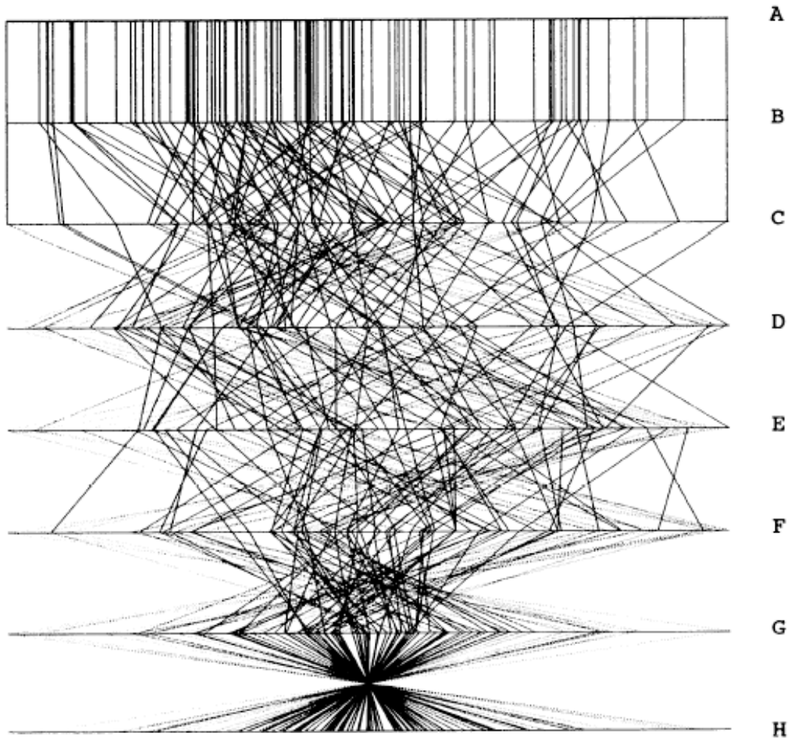
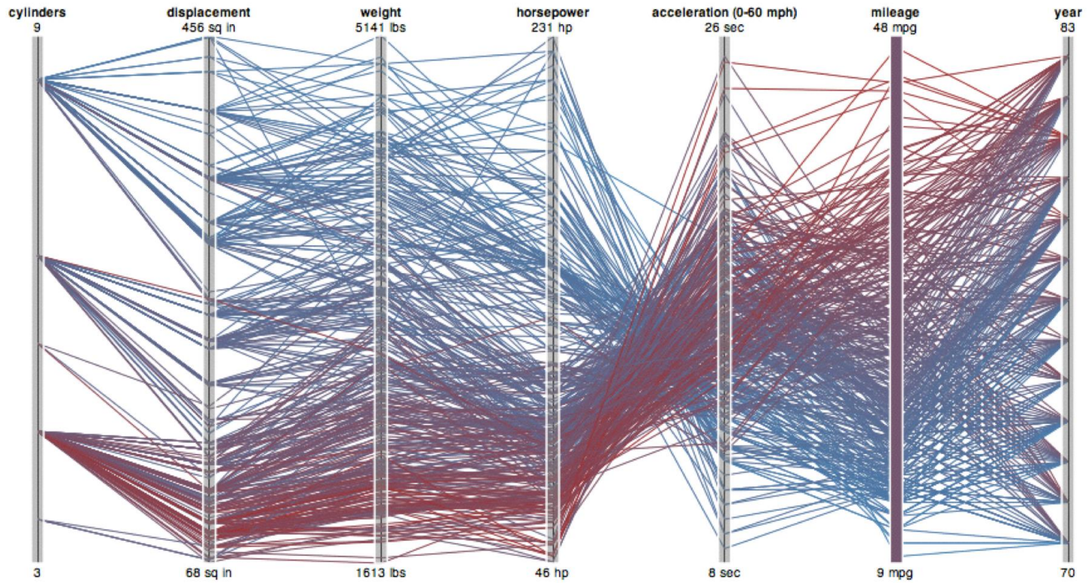


Figure 3. Parallel Coordinate Plot of Six-Dimensional Data Illustrating Correlations of $\rho = 1, .8, .2, 0, -.2, -.8,$ and -1.

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

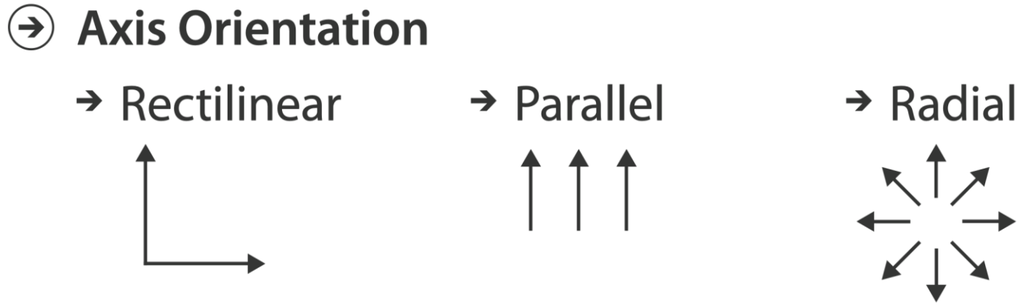
Parallel coordinates, limitations

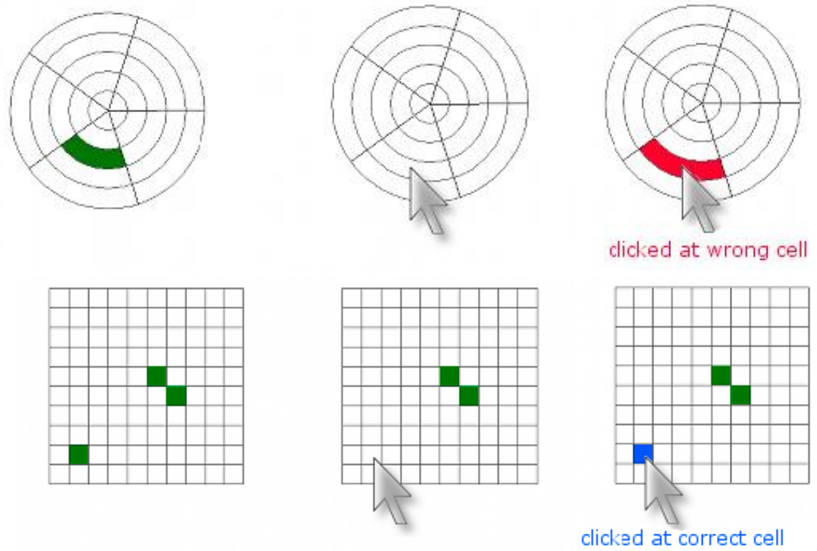
- visible patterns only between neighboring axis pairs
- how to pick axis order?
  - usual solution: reorderable axes, interactive exploration
  - same weakness as many other techniques
    - downside of interaction: human-powered search
  - some algorithms proposed, none fully solve



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



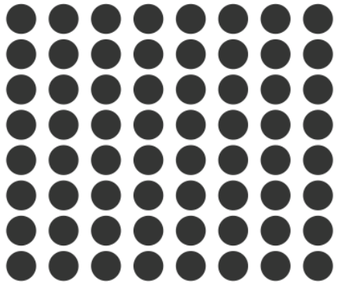


[Uncovering Strengths and Weaknesses of Radial Visualizations - an Empirical Approach. Diehl, Beck and Burch. IEEE TVCG (Proc. InfoVis) 16(6):935--942, 2010.]

Layout density

Layout Density

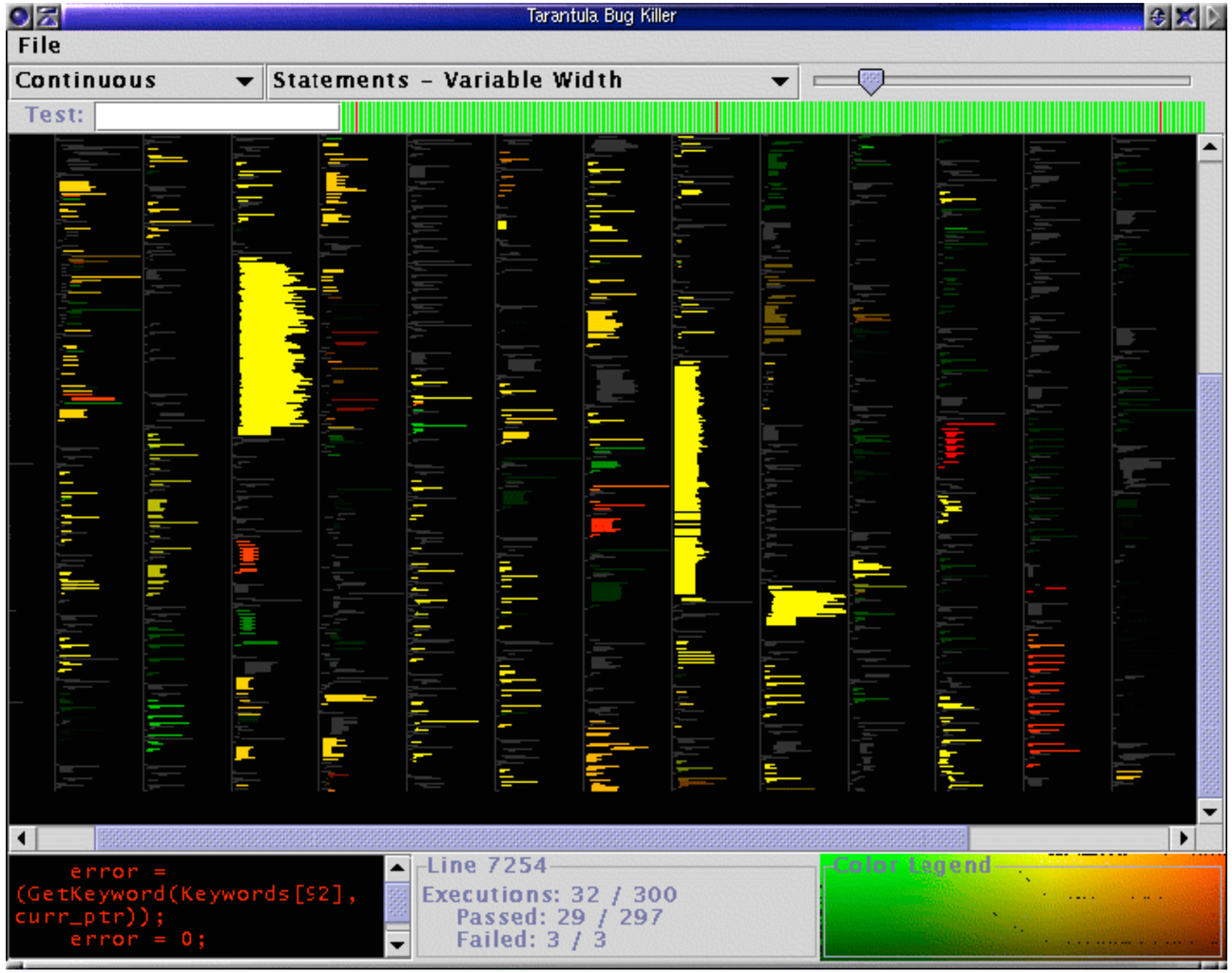
- Dense



- Space-Filling

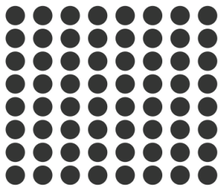
Idiom: Dense software overviews

- 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.]

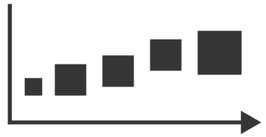
Arrange tables

Express Values

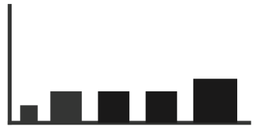
Separate, Order, Align Regions

- Separate

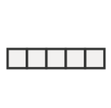
- Order



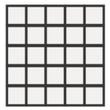
- Align



- *1 Key*
  List

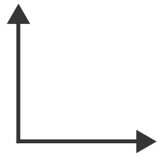


- *2 Keys*
  Matrix



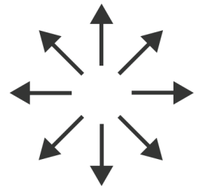
Axis Orientation

- Rectilinear



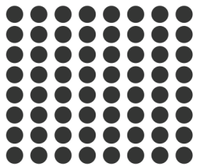
- Parallel

- Radial



Layout Density

- Dense



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

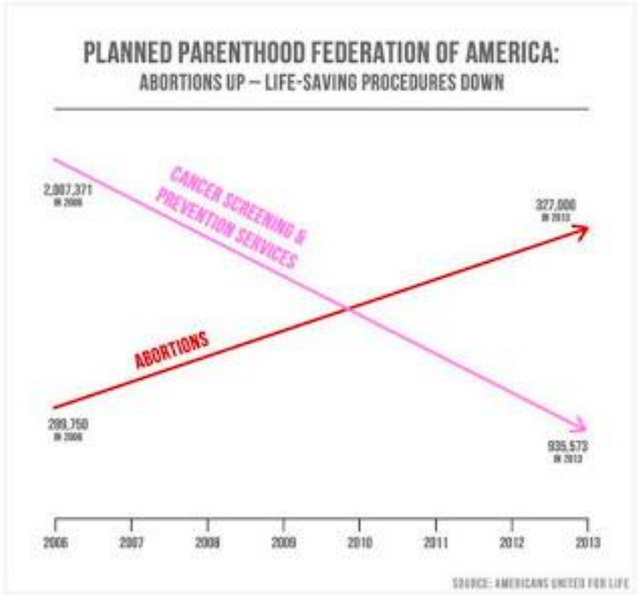
What?

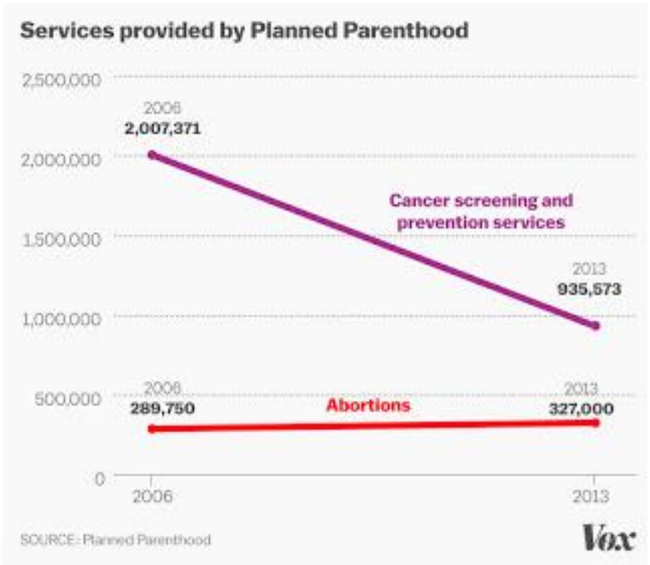
Why?

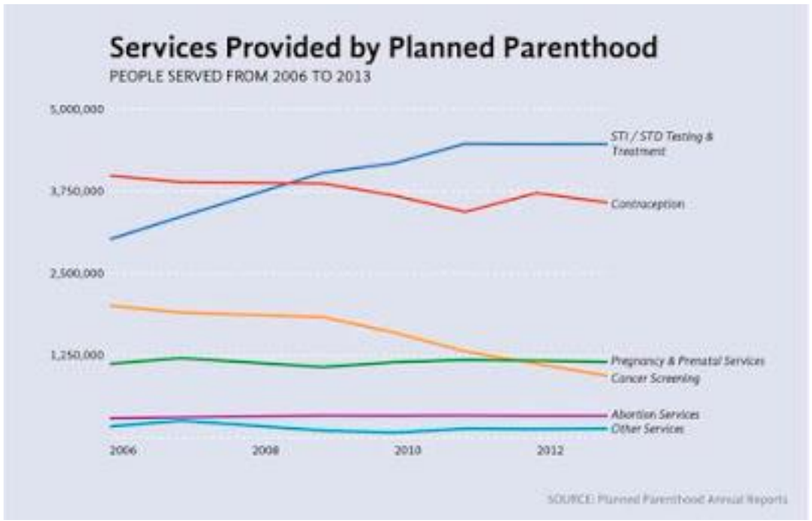
How?

Chart axes

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



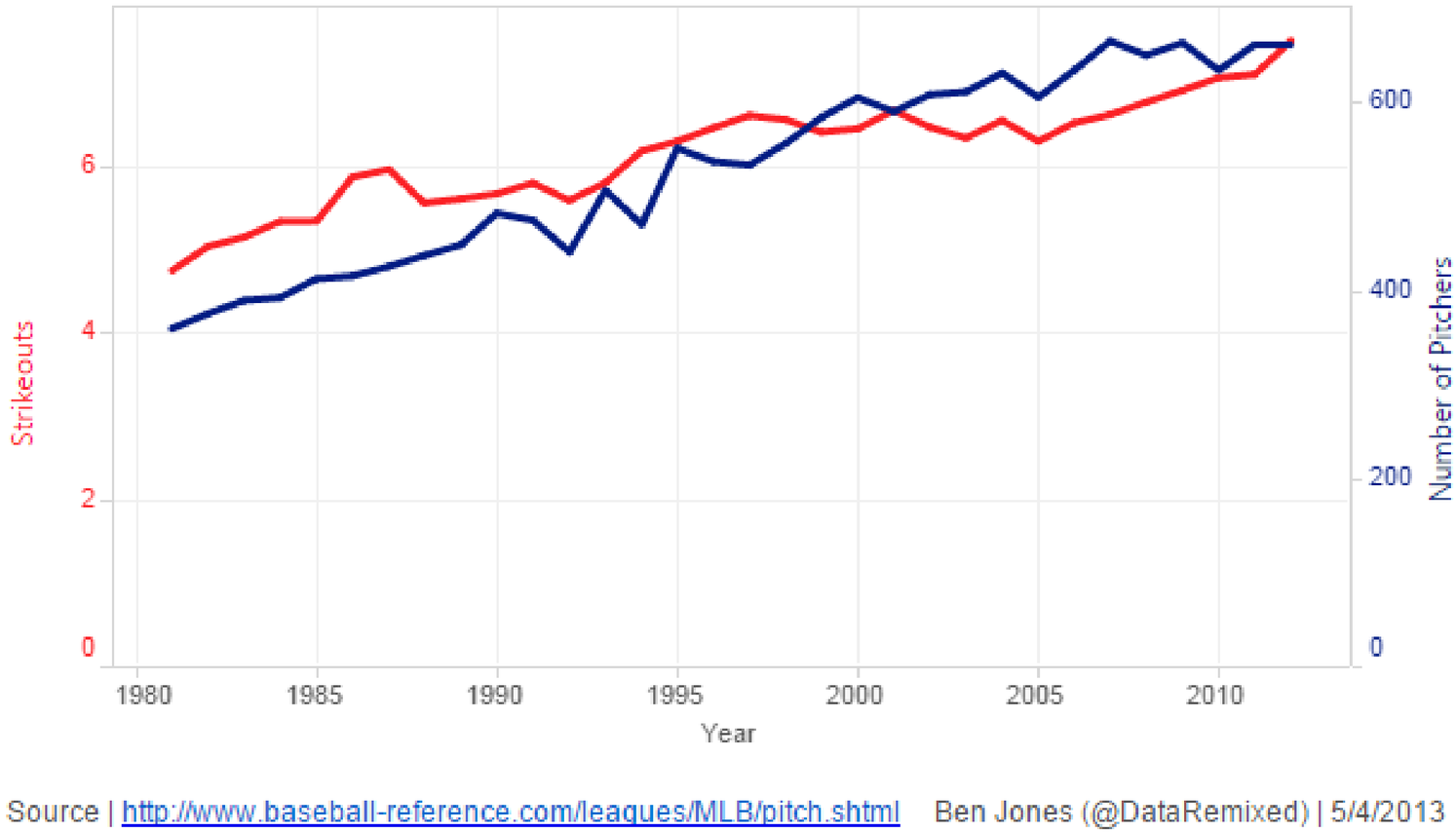




http://www.thefunctionalart.com/2015/10/if-you-see-bullshit-say-bullshit.htm

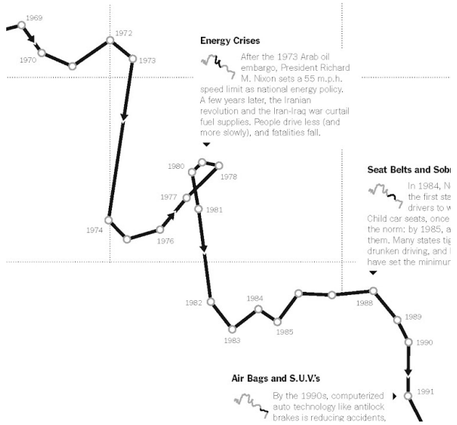
Idiom: **dual-axis line charts**

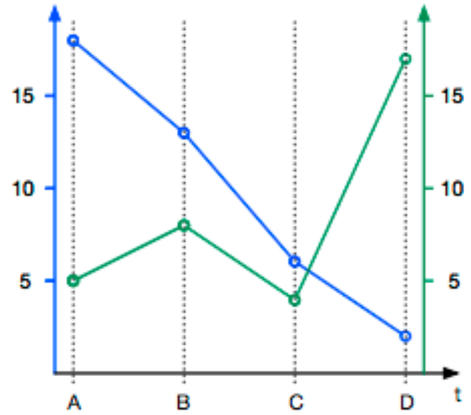
- controversial
  - acceptable if commensurate
  - beware, very easy to mislead!

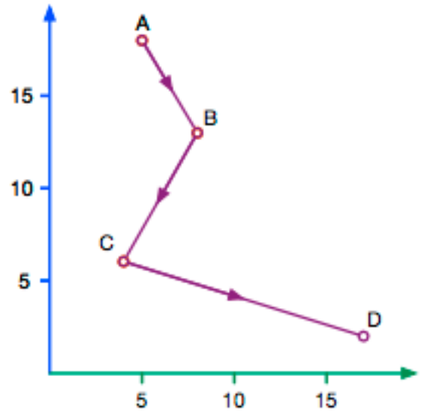


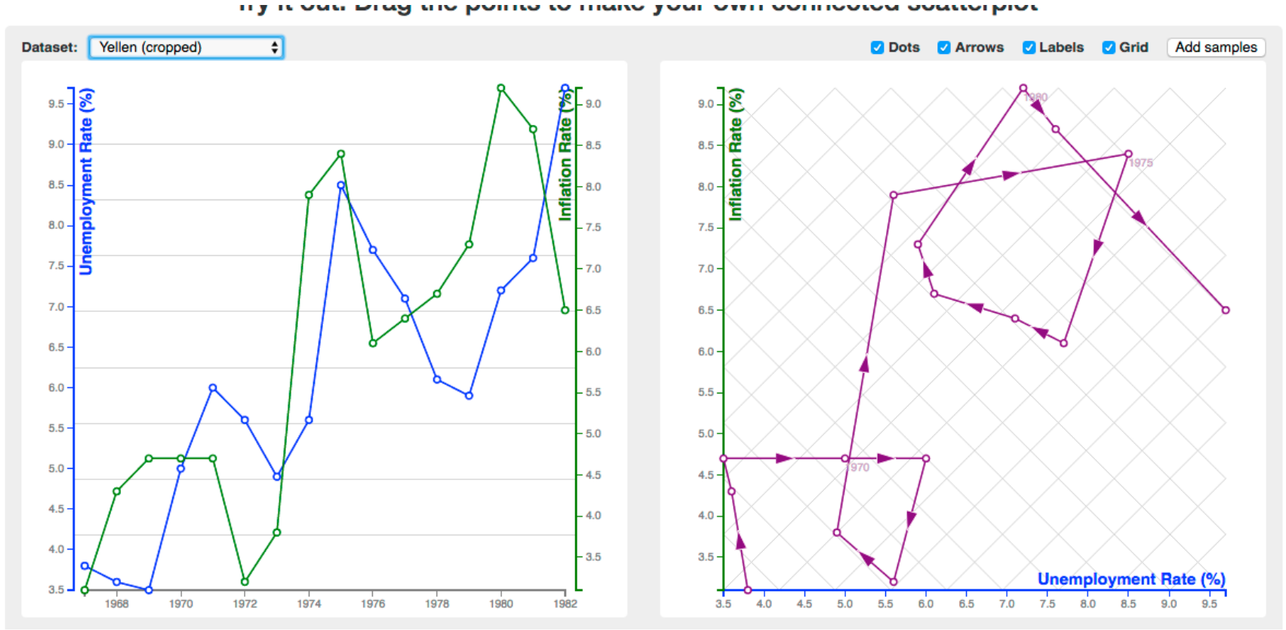
Idiom: **connected scatterplots**

- 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









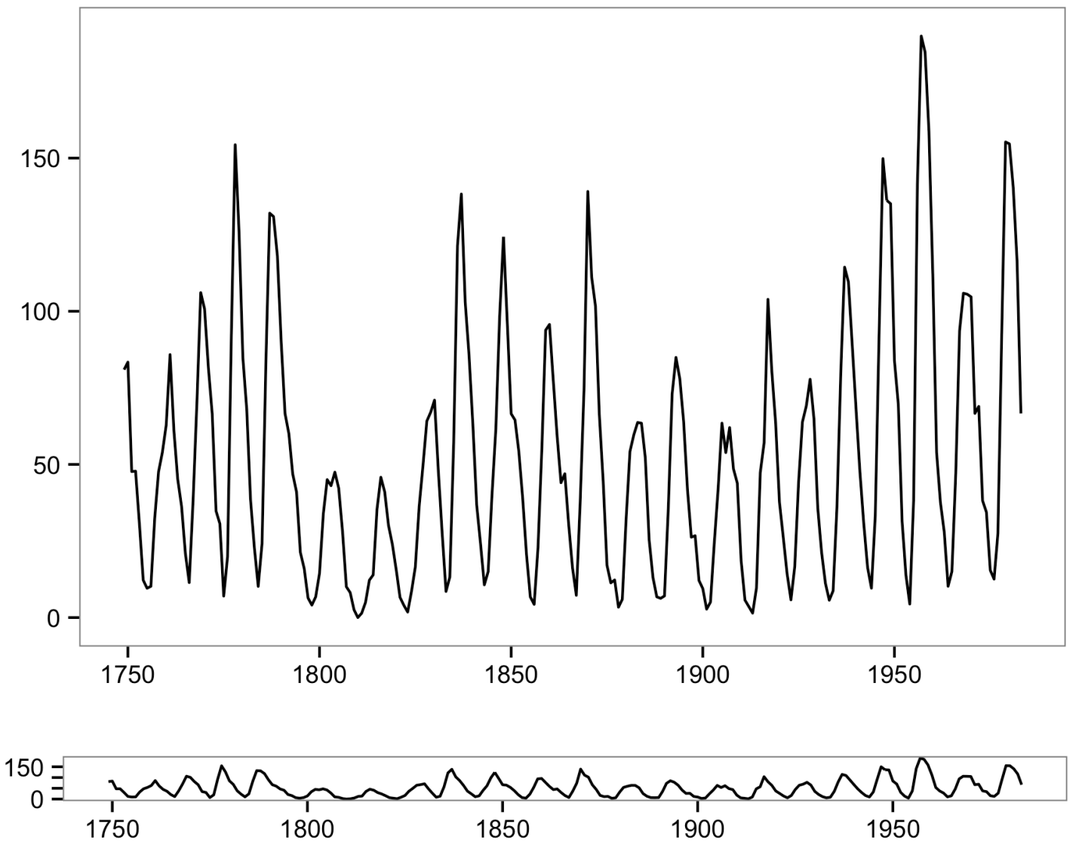
[The Connected Scatterplot for Presenting Paired Time Series. Haroz, Kosara and Franconeri. IEEE TVCG 22(9):2174-86, 2016.]

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

Choosing line chart aspect ratios

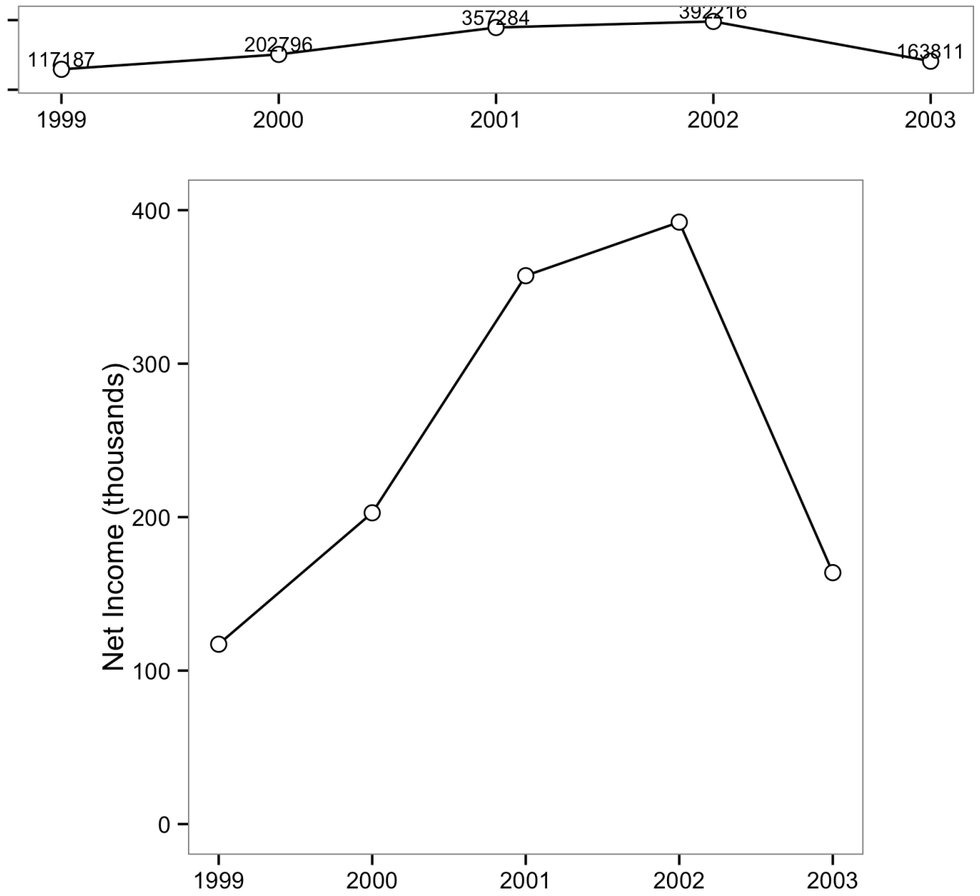
- 1: banking to 45 (1980s)
  - Cleveland perceptual argument: most accurate angle judgement at 45

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

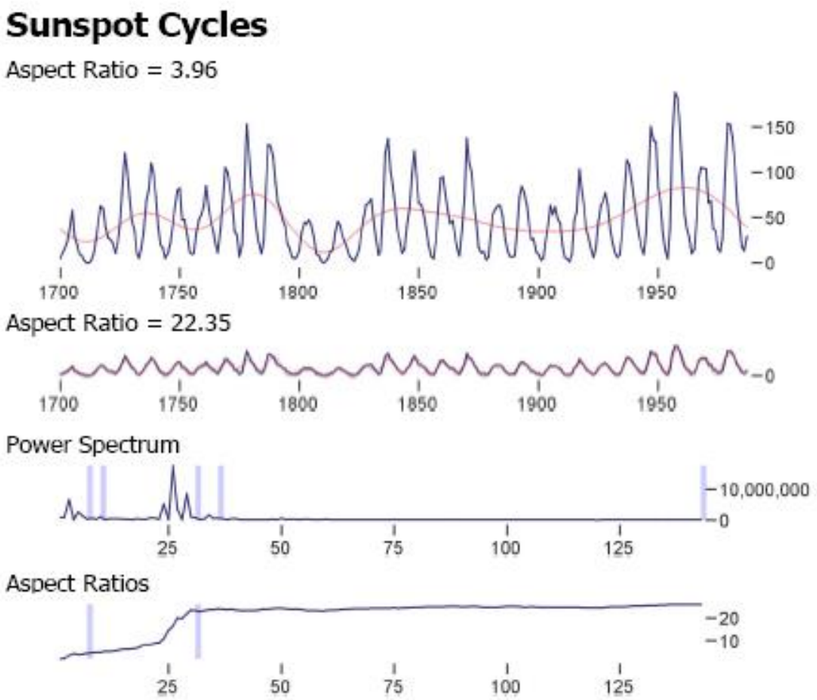
Fig 7.2 Annual Report: Aspect Ratio 2



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

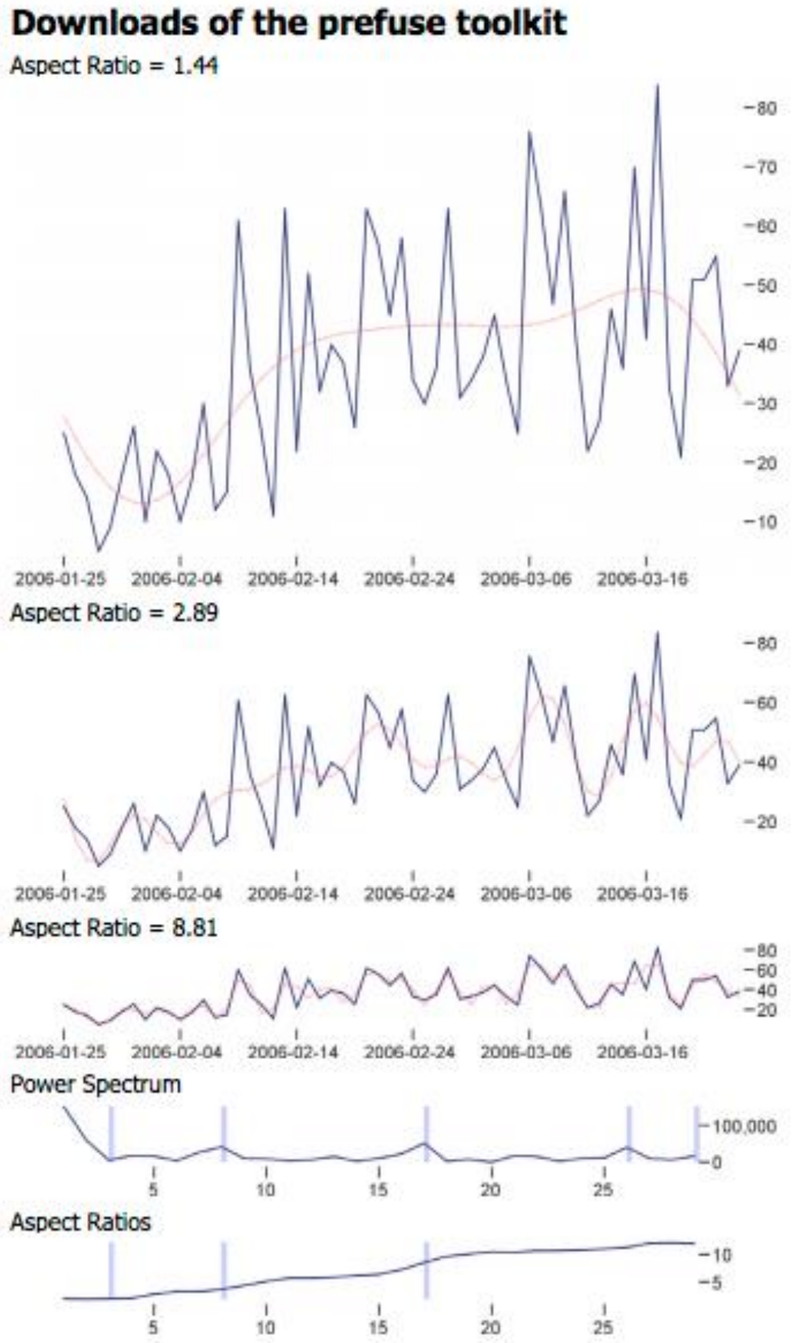


[Multi-Scale Banking to 45 Degrees. Heer and Agrawala, Proc InfoVis 2006]

overall

weekly

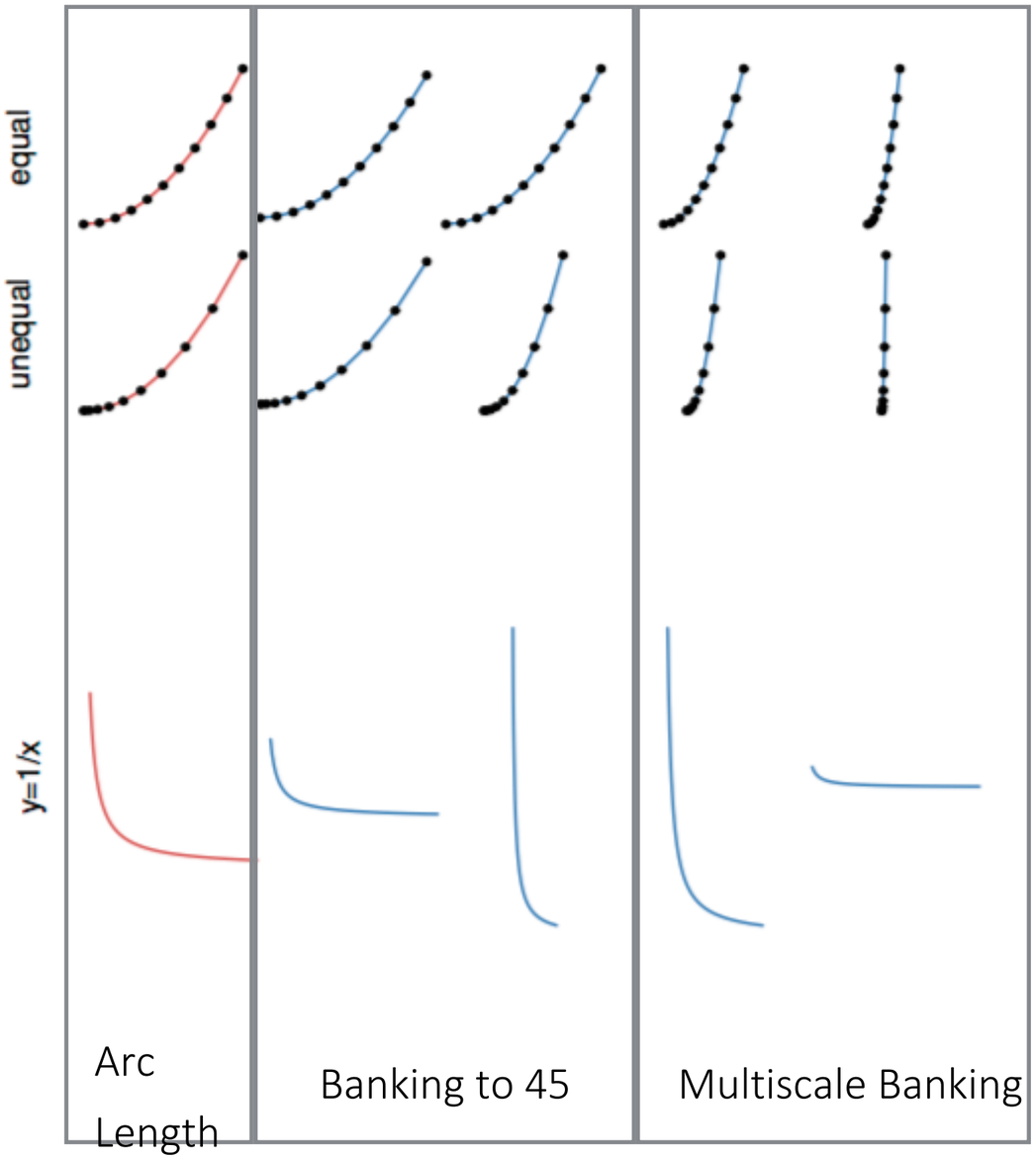
daily



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

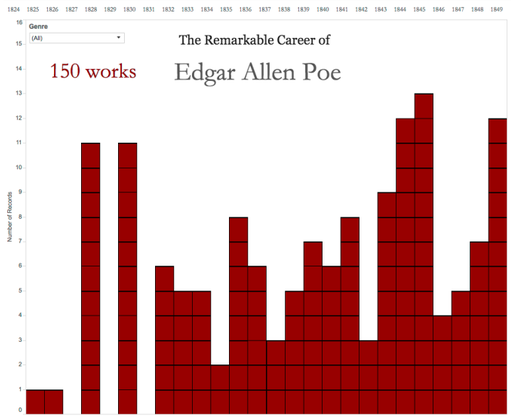
- meta-points from this progression
  - young field; prescriptive advice changes rapidly
  - reasonable defaults required deep dive into perception meets math



[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]