07. Tabular Data

ID 413: Information Graphics and Data Visualization
Spring 2025
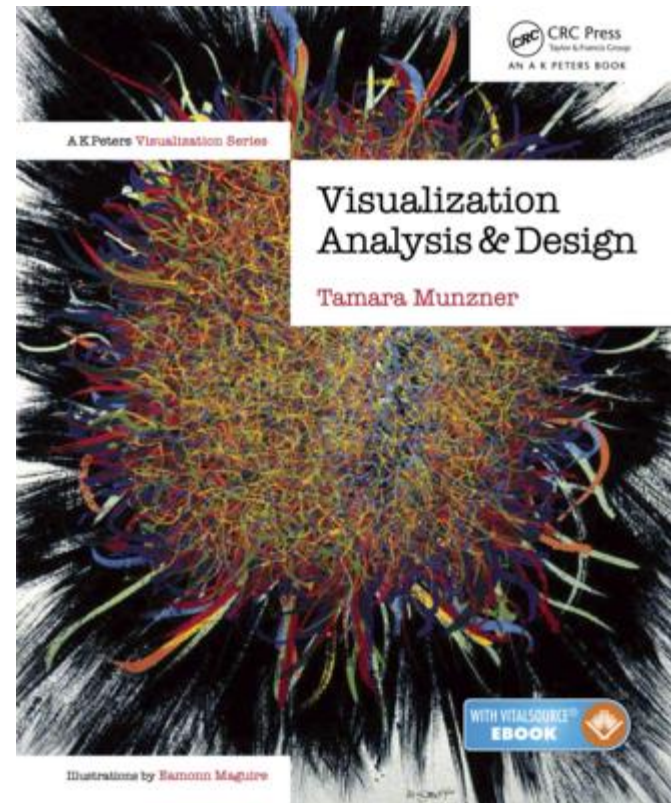
Venkatesh Rajamanickam (@venkatrajam)
venkatra@iitb.ac.in
http://info-design-lab.github.io/

Visualization Analysis & Design

Tamara Munzner

A K Peters Visualization Series
CRC Press, 2014

Chapter 7 (Arrange Tables)

What?

Datasets

Attributes

➔ 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

Attributes (columns)
Items (rows)
Cell containing value

➔ Networks

Link
Node (item)

➔ Fields (Continuous)

Grid of positions
Cell
Attributes (columns)
Value in cell

➔ *Multidimensional Table*

Key 1
Key 2
Value in cell
Attributes

➔ *Trees*

➔ Dataset Availability

➔ Static

➔ Dynamic

➔ Geometry (Spatial)

Position

➔ Attribute Types

➔ Categorical

➔ Ordered

➔ *Ordinal*

➔ *Quantitative*

➔ Ordering Direction

➔ Sequential

➔ Diverging

➔ Cyclic

What?

Why?

How?

3

Focus on Tables

→ **Dataset Types**

→ Tables

Attributes (columns)

Items (rows)

Cell containing value

→ *Multidimensional Table*

Key 1

Key 2

Value in cell

Attributes

→ Networks

Link

Node (item)
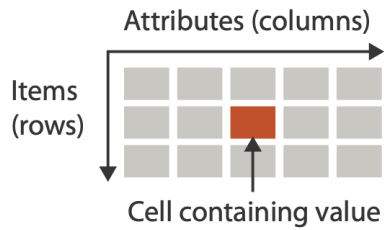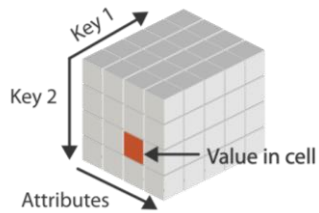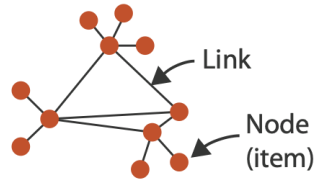
→ *Trees*

→ Spatial

→ Fields (Continuous)
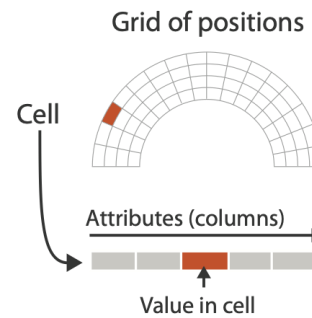
Grid of positions

Cell

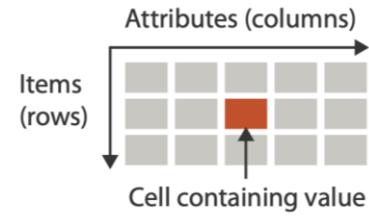Attributes (columns)

Value in cell

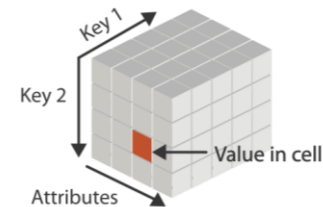→ Geometry (Spatial)

Position

Keys and values

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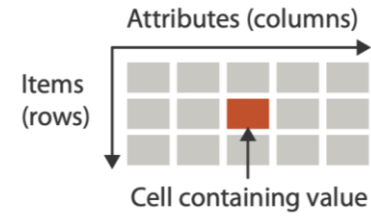 - dependent attribute, value of cell

→ Tables

Attributes (columns)

Items (rows)

Cell containing value

→ *Multidimensional Table*

Key 1

Key 2

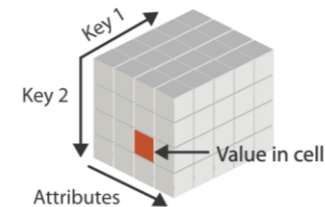Value in cell

Attributes

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, ...

➔ Tables

Attributes (columns)

Items (rows)

Cell containing value

➔ Multidimensional Table

Key 1

Key 2

Attributes

Value in cell

➔ 0 Keys
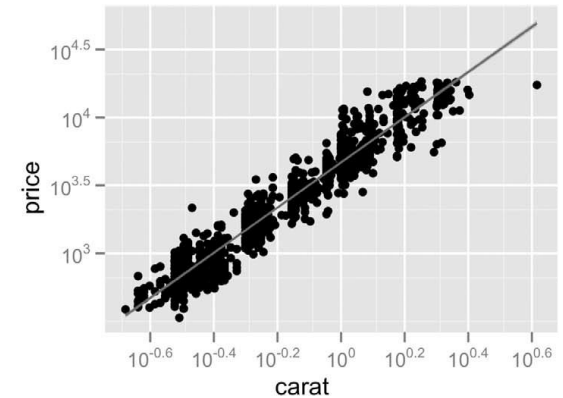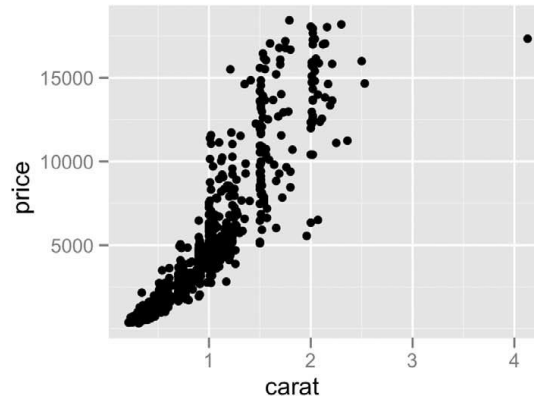
⊙ Express Values

➔ 1 Key
List

➔ 2 Keys
Matrix

6

Idiom: **scatterplot**

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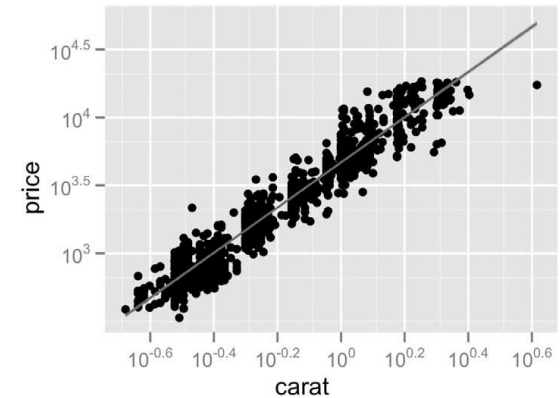 - quantitative attributes
- no keys, only values

Express Values

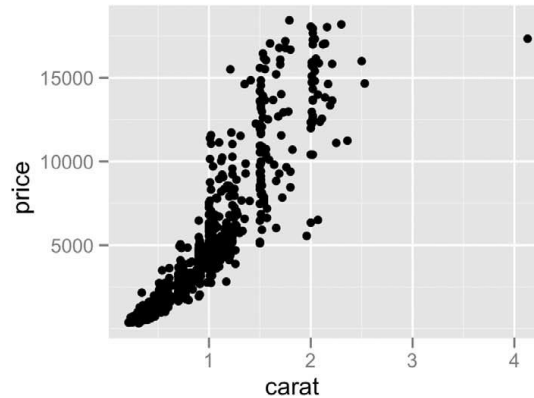*[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]*

Idiom: **scatterplot**

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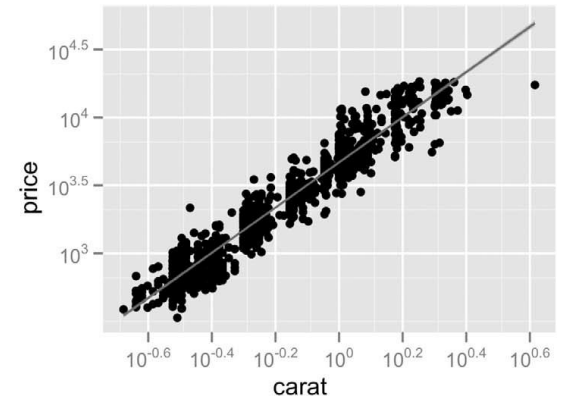 - mark: points
  - channels
    - horiz + vert position

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

Idiom: **scatterplot**

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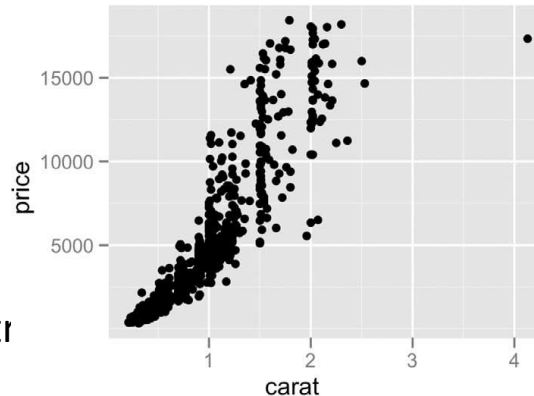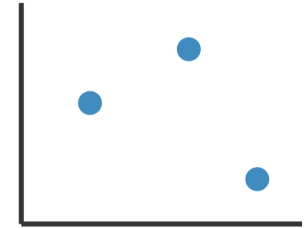 - channels
    - horiz + vert position
  - tasks
    - find trends, outliers, distr
  - scalability
    - hundreds of items

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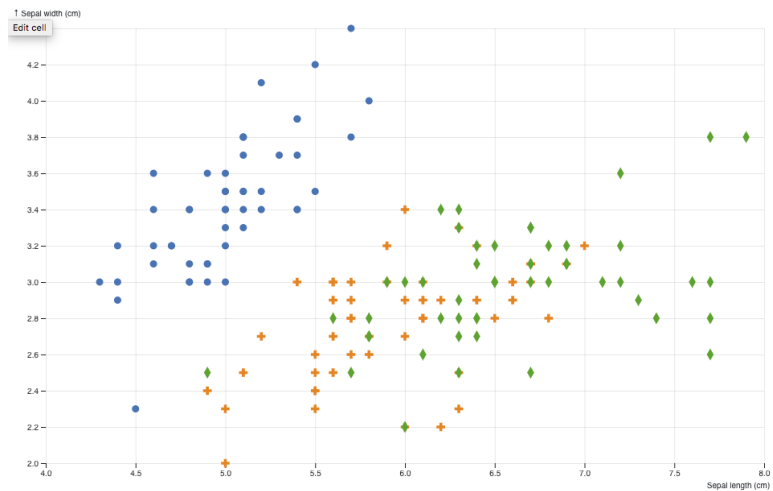
Scatterplots: Encoding more channels

- additional channels viable since using point marks
  - color
  - size (1 quant attribute, used to control 2D area)
    - note radius would mislead, take square root since area grows quadratically
  - shape
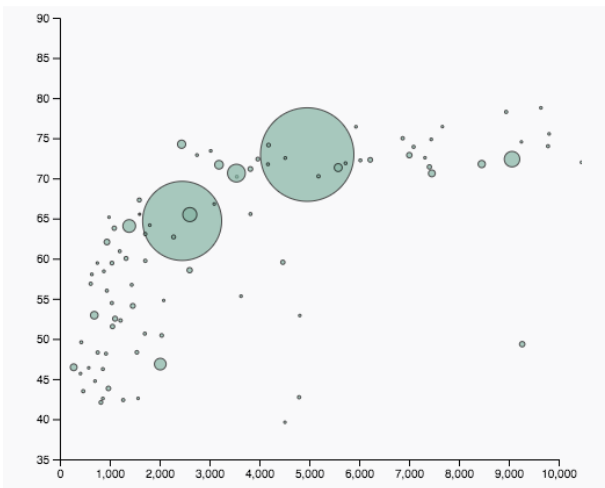
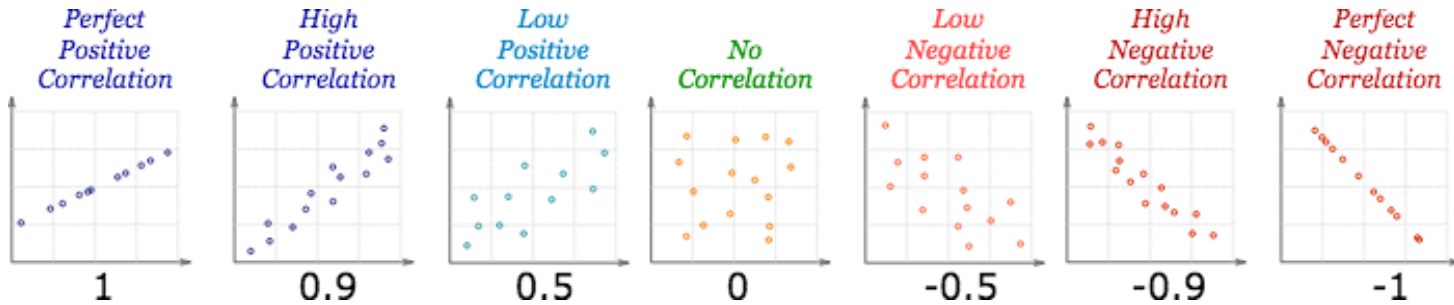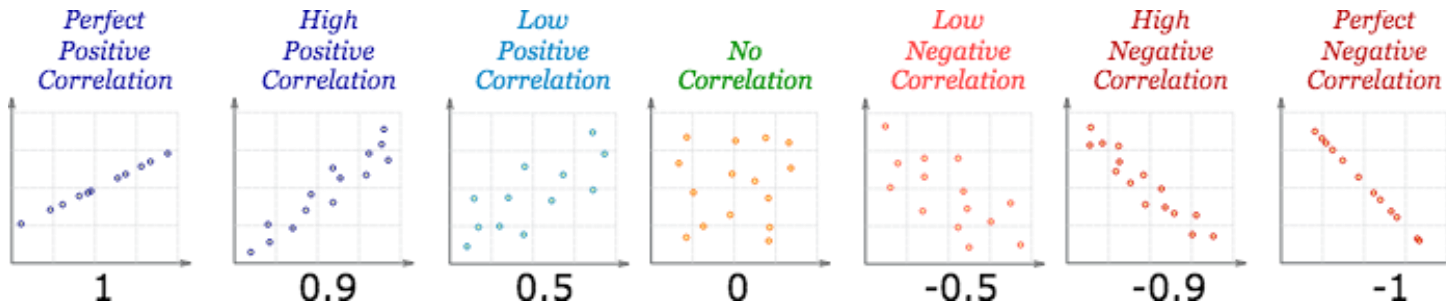https://www.d3-graph-gallery.com/graph/bubble_basic.html

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

Scatterplot tasks

Scatterplot tasks

- correlation

Scatterplot tasks

- correlation

- clusters/groups, and clusters vs classes

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

13

Some keys

→ *0 Keys*

⊙ **Express Values**

→ *1 Key*
List

→ *2 Keys*
Matrix
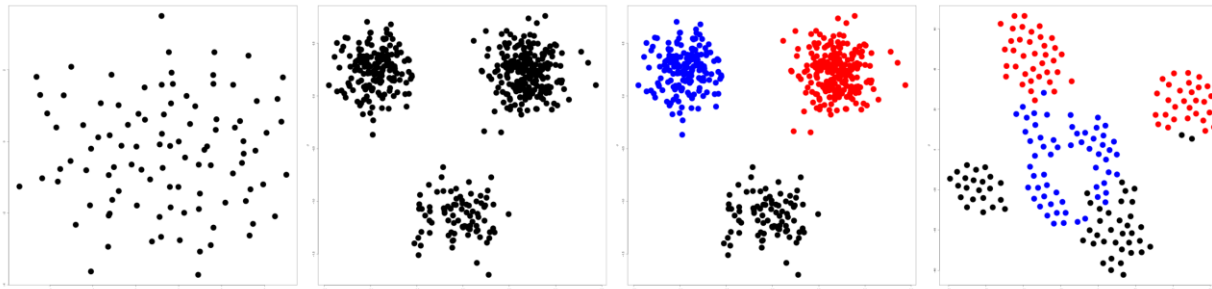
Some keys: Categorical regions

→ Separate

→ Order

→ Align

Regions: Separate, order, align
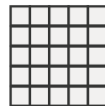

→ Separate   → Order   → Align

- 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
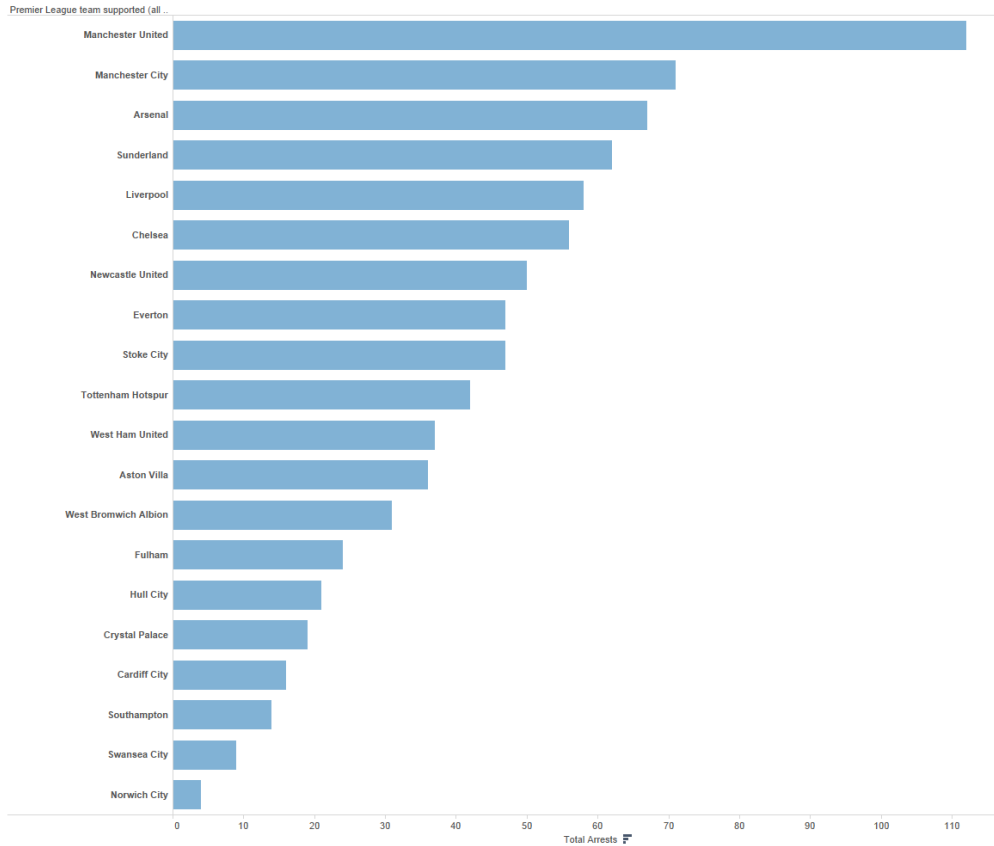- use ordered attribute to order and align regions
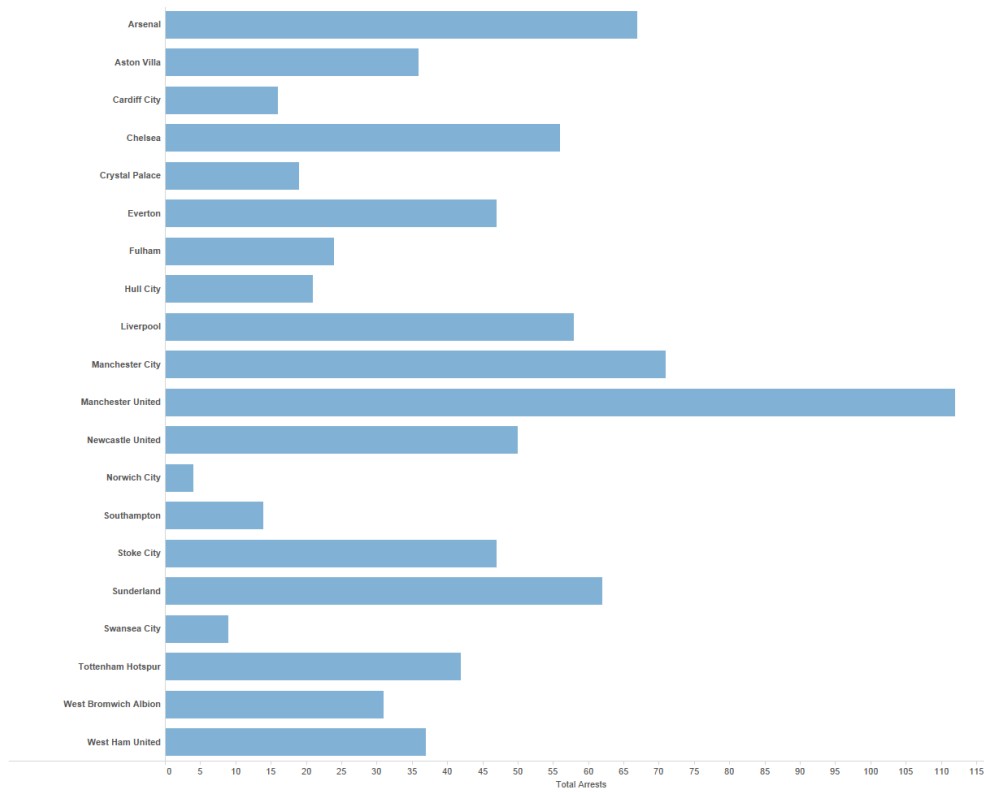

→ 1 Key   → 2 Keys
List   Matrix

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
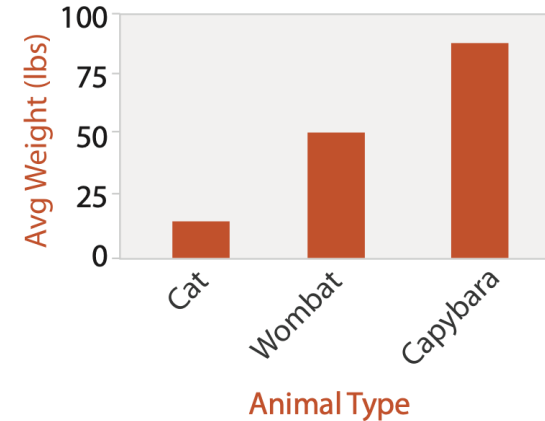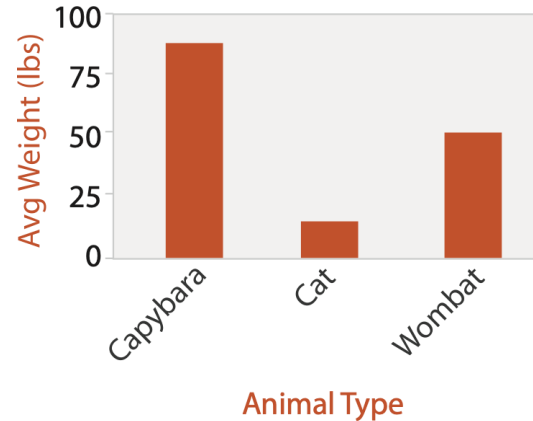
      »by label (alphabetical), by length attrib (data-driven)
 - task
    - compare, lookup values
  - scalability
    - dozens to hundreds of levels for key attrib [bars], hundreds for values

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 mu
  – 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]
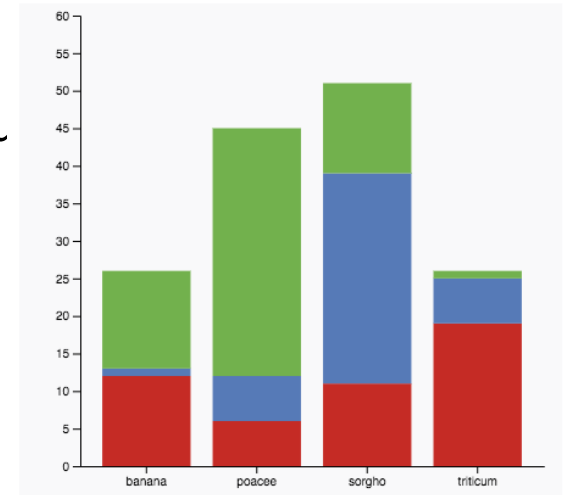
# Idiom: **streamgraph**

- generalized stacked graph
  - emphasizing horizontal continuity
    - vs vertical items
  - data
    - 1 categ key attrib (movies)
    - 1 ordered key attrib (time)
    - 1 quant value attrib (counts)
  - derived data
    - geometry: layers, where height encodes counts
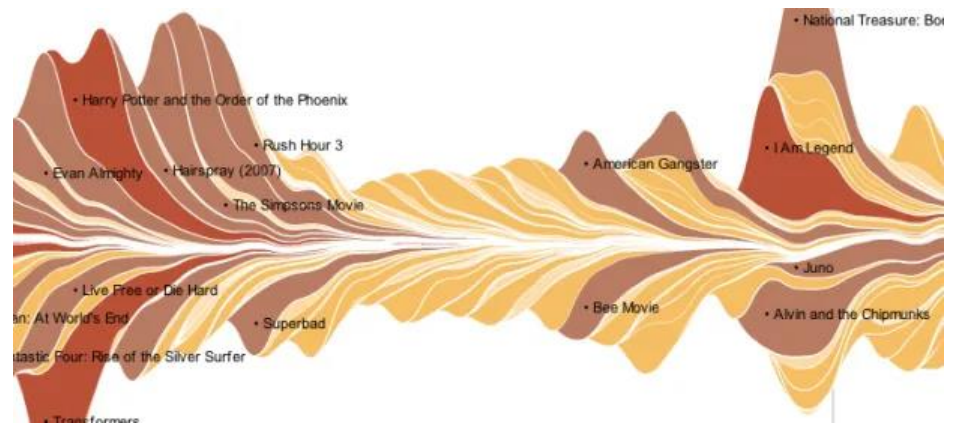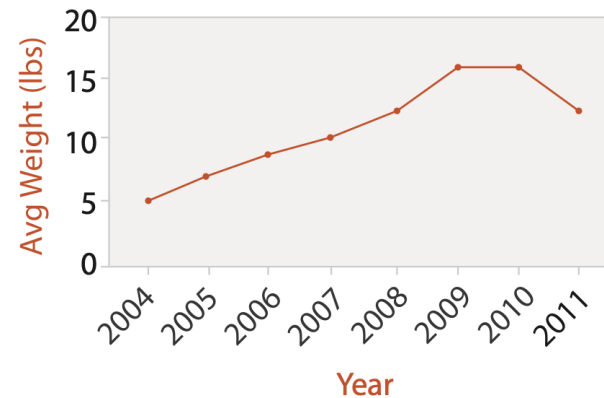   - 1 quant attrib (layer ordering)
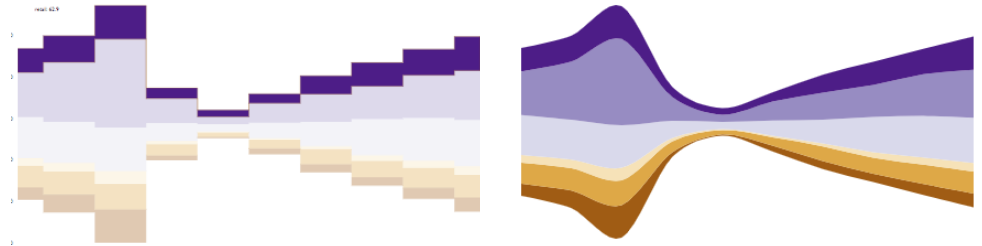
[Stacked Graphs Geometry & Aesthetics. Byron and Wattenberg. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2008) 14(6): 1245–1252, (2008).]

Idiom: **streamgraph**

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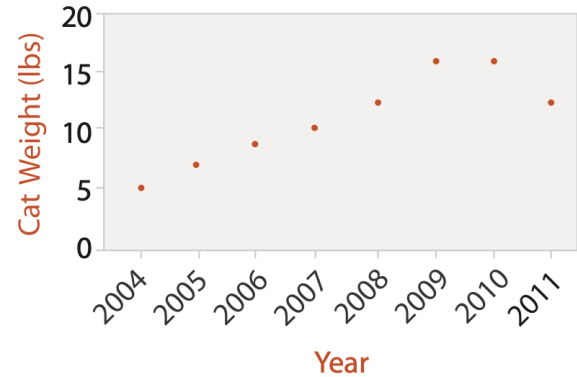   - 1 quant attrib (layer ordering)

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

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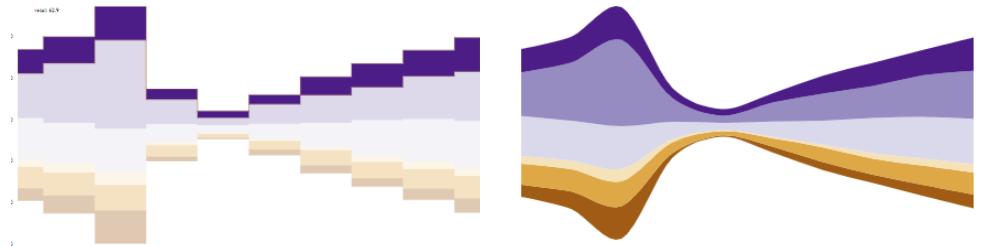
Idiom: **dot / line chart**

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

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
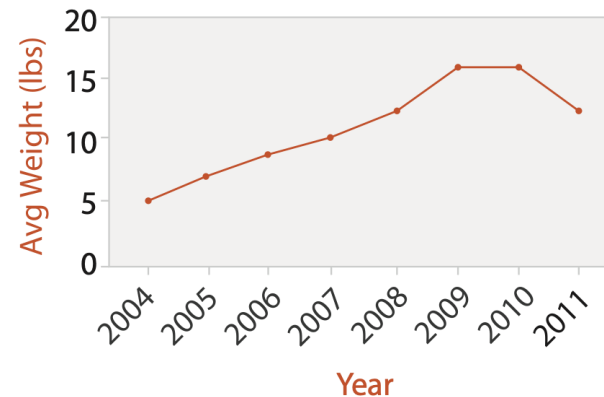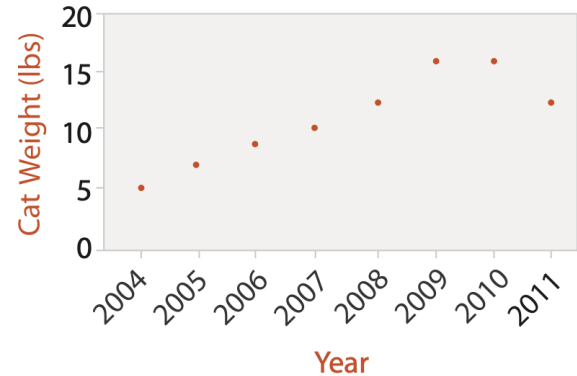
Choosing bar vs line charts

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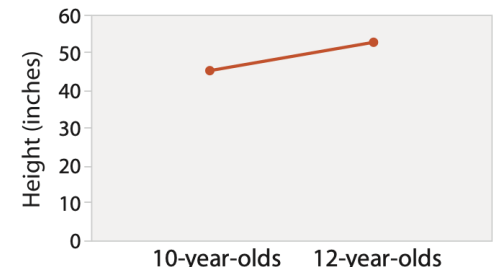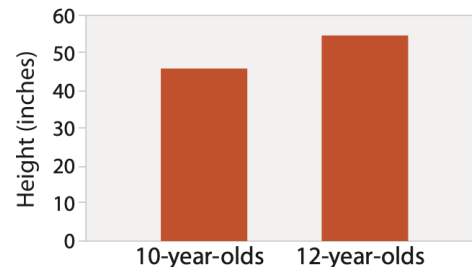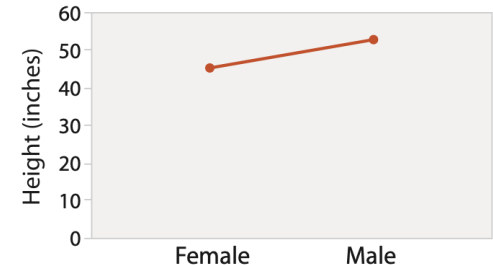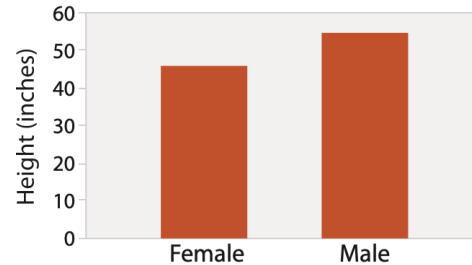
after [Bars and Lines: A Study of Graphic Communication. Zacks and Tversky. Memory and Cognition 27:6 (1999), 1073–1079.]

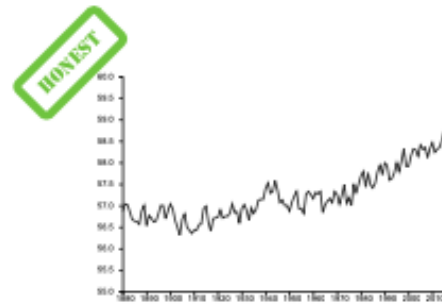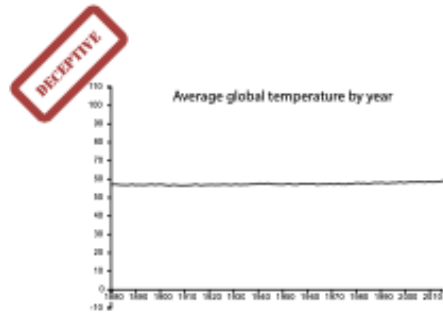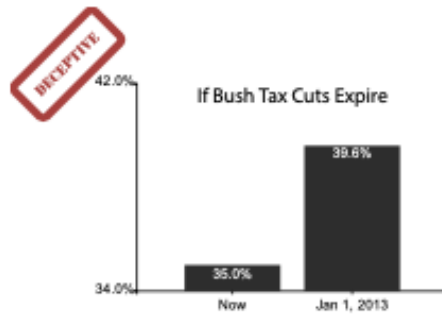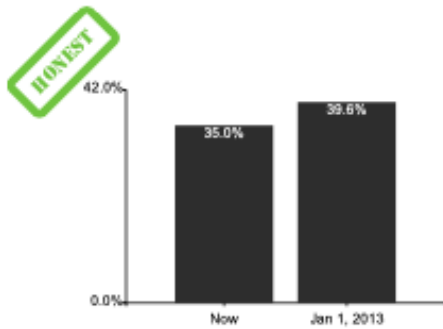# • include 0 at bottom left or slope misleads
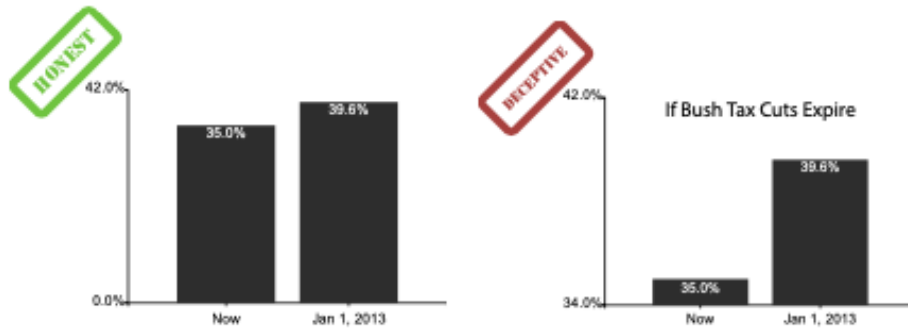
[Truncating the Y-Axis: Threat or Menace?
Correll, Bertini, & Franconeri, CHI 2020.]

Chart axes: avoid cropping y axis

- include 0 at bottom left of eads
  - some exceptions (arbitrary 0, natters)
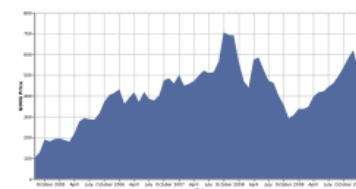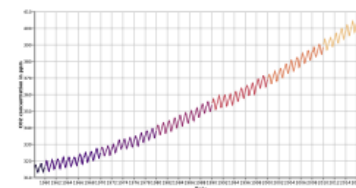

(a) Statistical process charts rely on comparison to an expected value, and so deviations from that value, not from zero, are important

(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.

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

Idiom: **Indexed line charts**

- data: 2 quant attribs
  - 1 key + 1 value

- derived data: new quant value attrib
  - index
  - plot instead of original value

- task: show change over time
  - principle: normalized, not absolute

- scalability
  - same as standard line chart

https://public.tableau.com/profile/ben.jones#!/vizhome/CAStateRevenues/Revenues
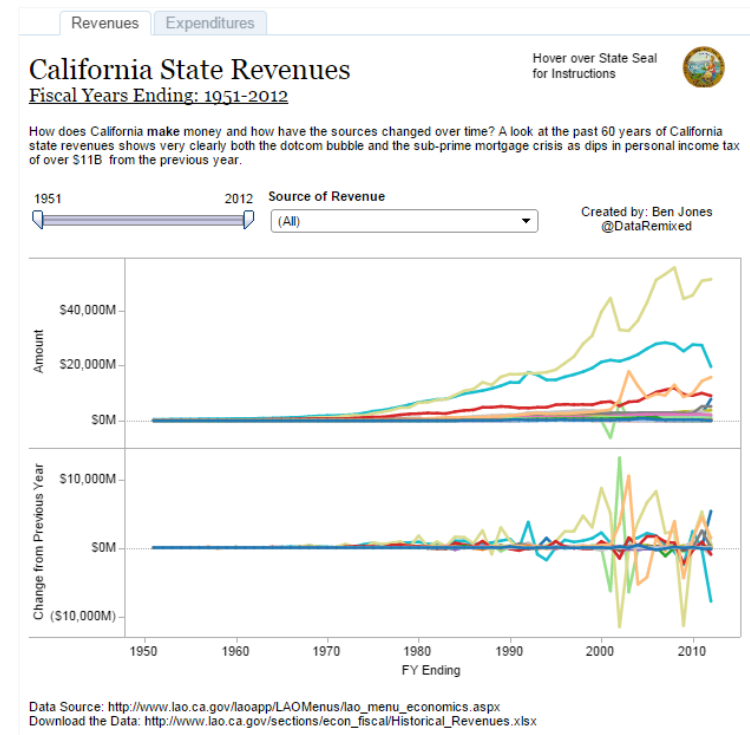
Idiom: **Gantt charts**

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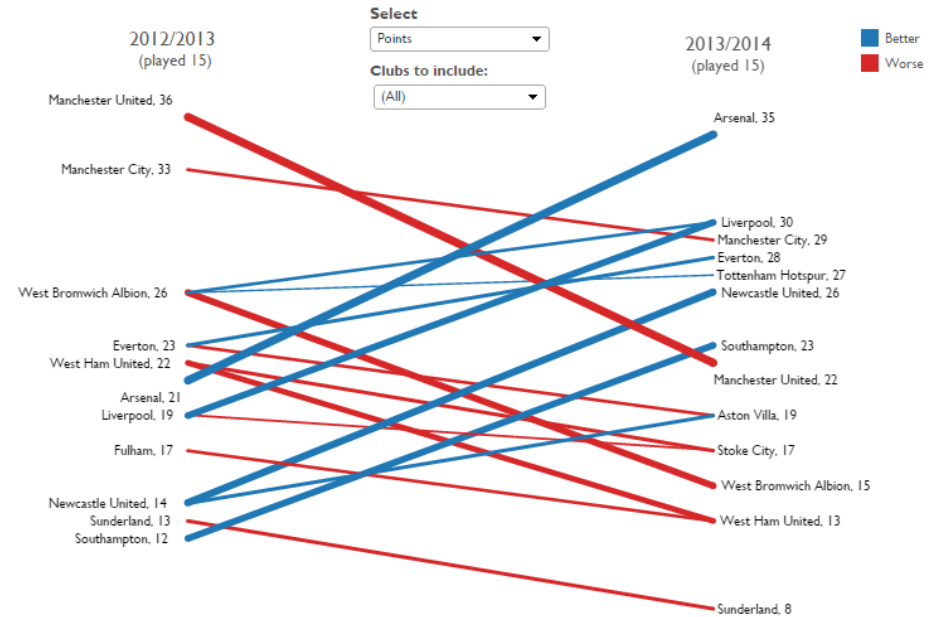   - hundreds of value levels [durations]
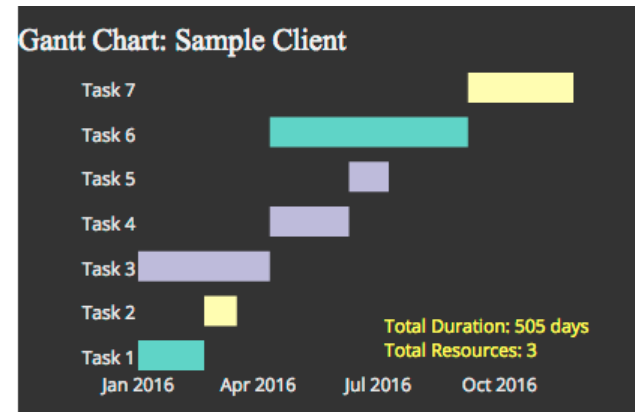
Idiom: **Slopegraphs**

- 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

Barclay's Premier League Tables: Comparing 2012/2013 Starts to 2013/2014 Starts

2012/2013 (played 15)   2013/2014 (played 15)

Select
Points

Clubs to include:
(All)

Better
Worse

Manchester United, 36

Manchester City, 33

West Bromwich Albion, 26

Everton, 23
West Ham United, 22

Arsenal, 21
Liverpool, 19

Fulham, 17

Newcastle United, 14
Sunderland, 13
Southampton, 12

Arsenal, 35

Liverpool, 30
Manchester City, 29
Everton, 28
Tottenham Hotspur, 27
Newcastle United, 26

Southampton, 23

Manchester United, 22

Aston Villa, 19

Stoke City, 17

West Bromwich Albion, 15

West Ham United, 13

Sunderland, 8

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

2 Keys

→ *0 Keys*

⊖ **Express Values**

→ *1 Key*
*List*

→ *2 Keys*
Matrix
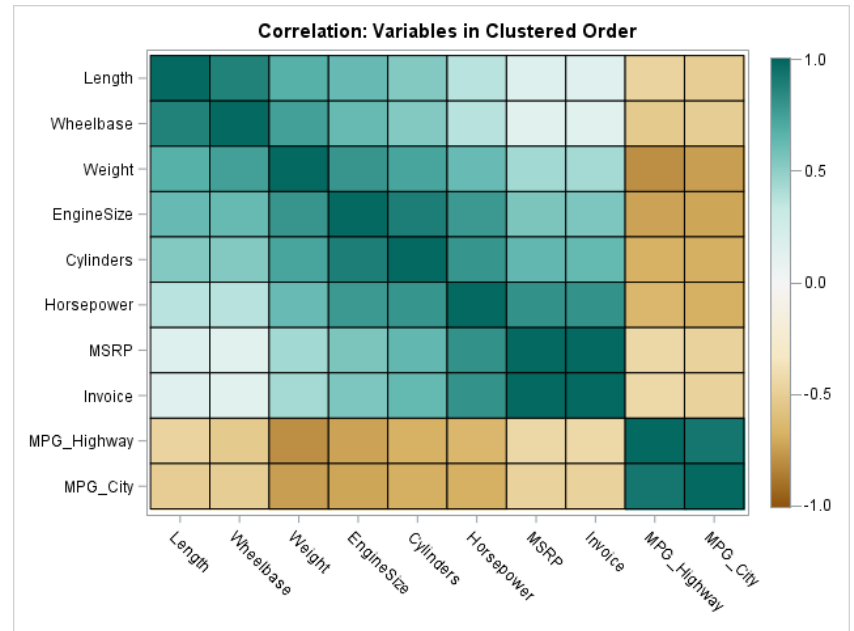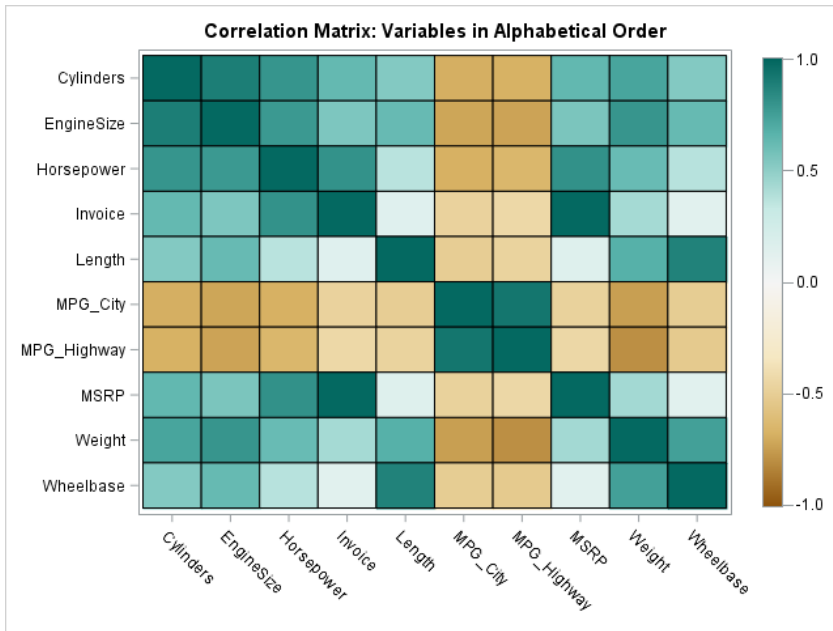
Idiom: **heatmap**

- 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
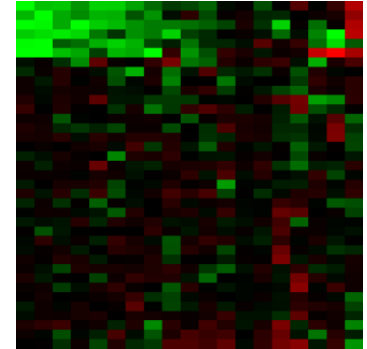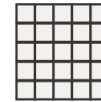
Heatmap reordering

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
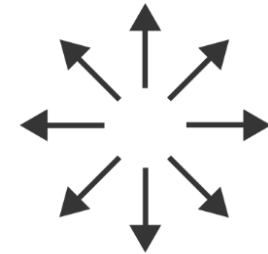   - task: assess quality of clusters found by automatic methods
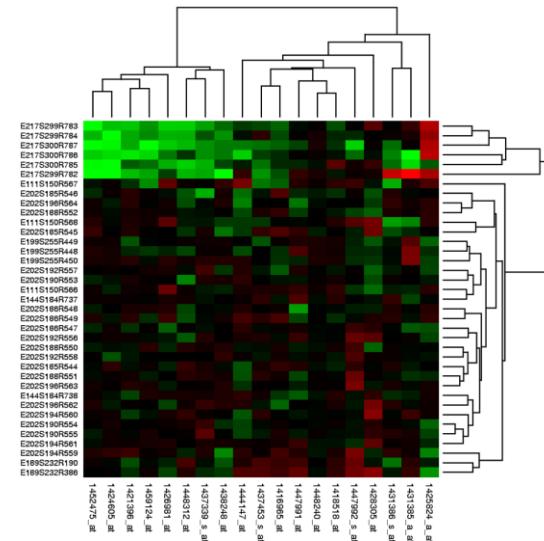
# → **Axis Orientation**
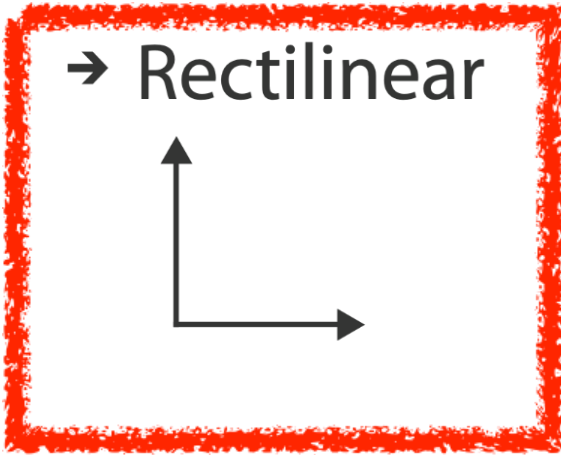
→ Rectilinear

→ Parallel
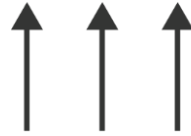
→ Radial

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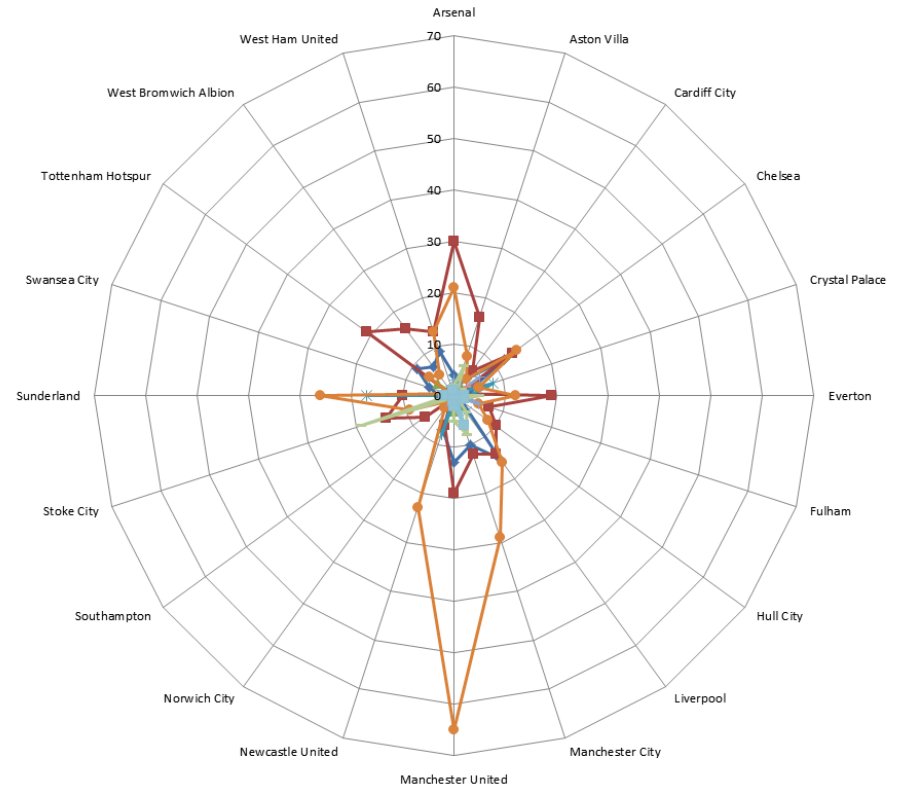   - less accurately perceived than rectilinear aligned
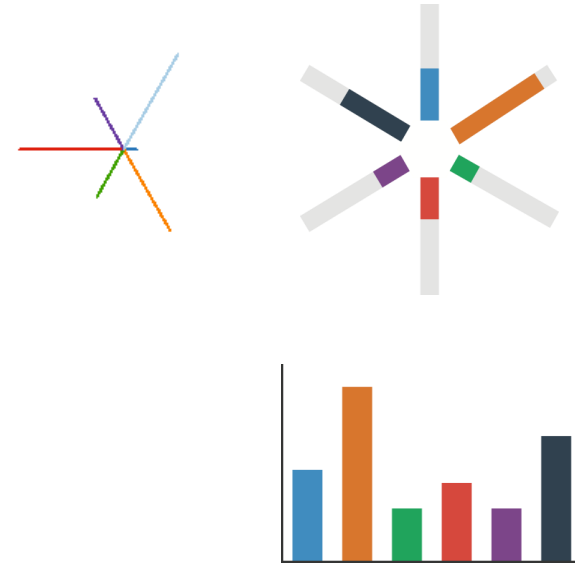
[Vismon: Facilitating Risk Assessment and Decision Making In Fisheries Management. Booshehrian, Möller, Peterman, and Munzner. Technical Report TR 2011-04, Simon Fraser University, School of Computing Science, 2011.]

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

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**
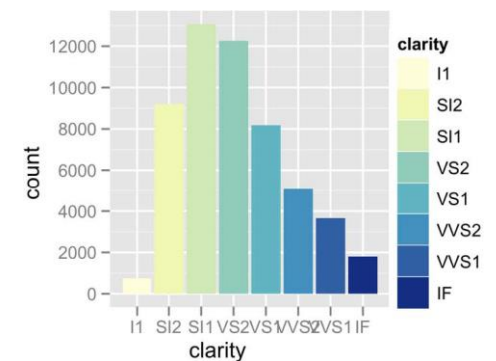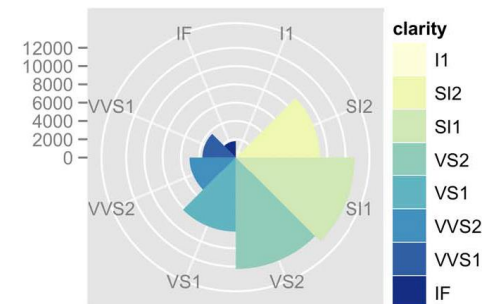- coxcomb chart
  - line marks with length channel: **1D length varies**
    - separated & ordered radially, uniform width
  - direct analog to radial bar charts
- data
  - 1 categ key attrib, 1 quant value attrib

[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]
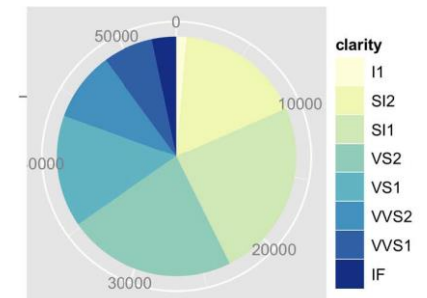
Coxcomb / nightingale rose / polar area chart

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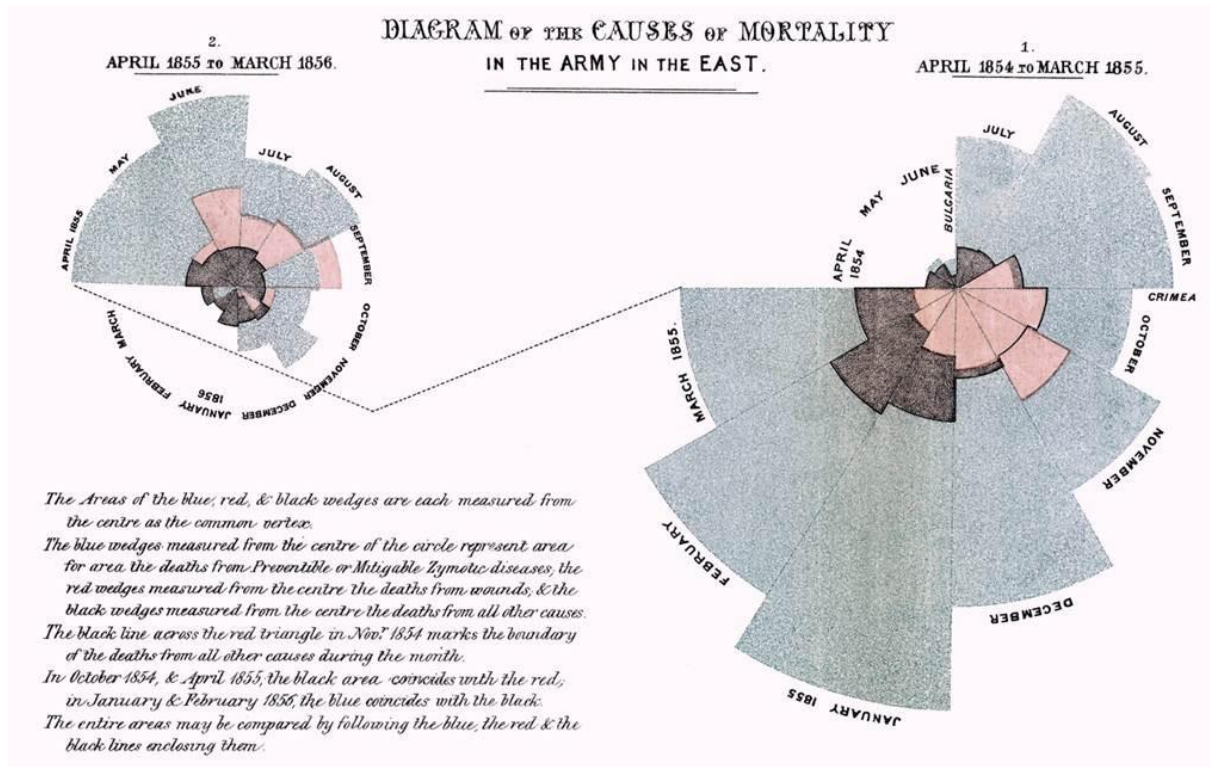
Coxcomb: perception

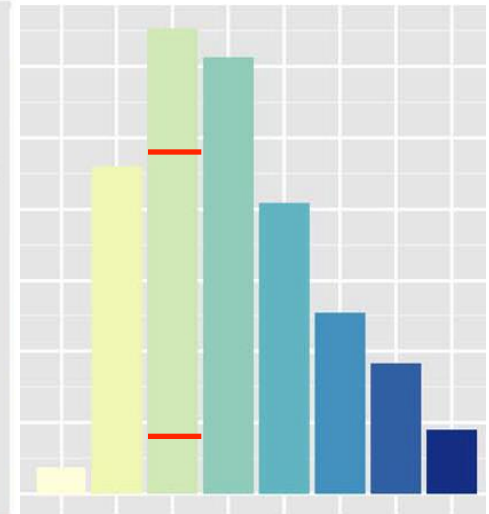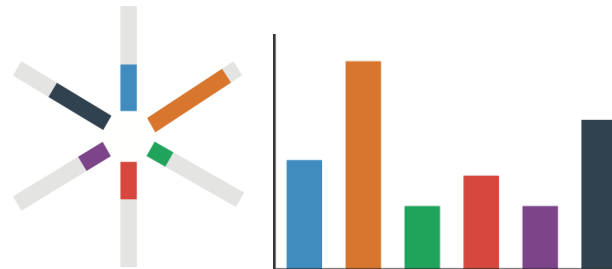- encode: 1D length
- decode/perceive: 2D area

- 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
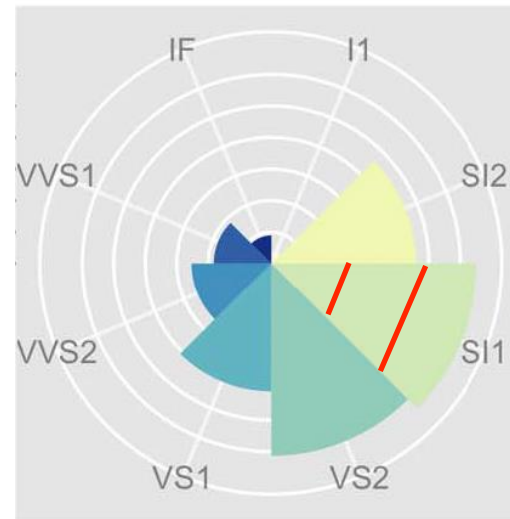
nonuniform width as length increases

uniform width as length increases

radial & rectilinear bars: uniform width as length increases

42

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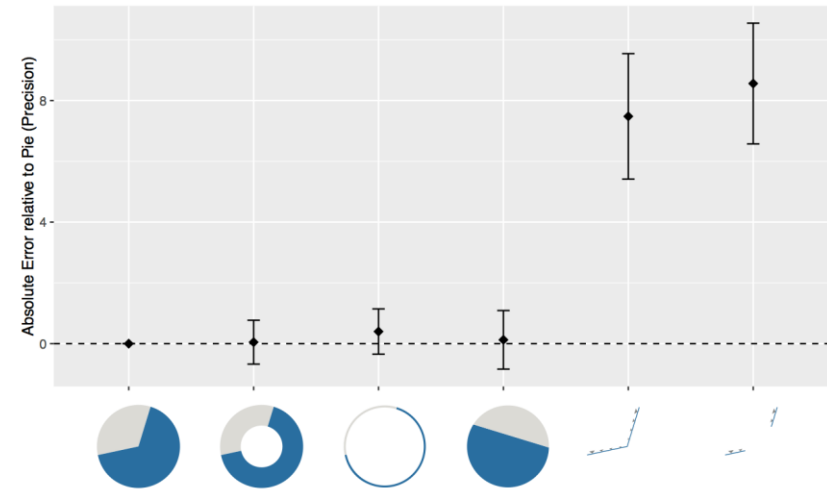
[Arcs, Angles, or Areas: Individual Data Encodings in Pie and Donut Charts. Skau and Kosara. Proc. EuroVis 2016.]

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
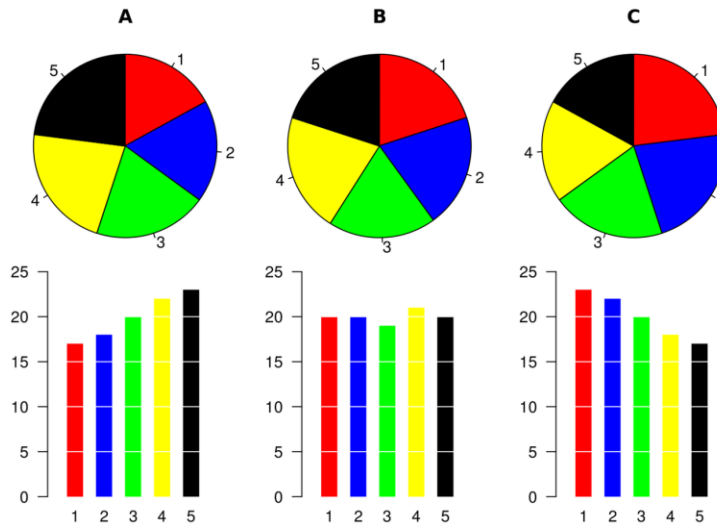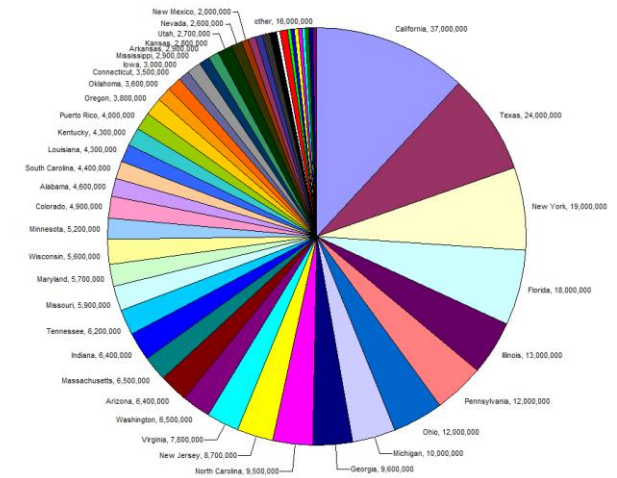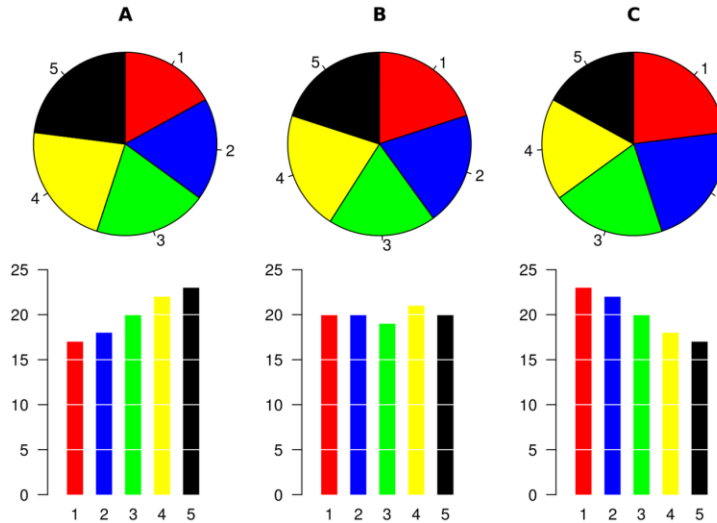
Pie charts: best practices

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

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

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

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

http://bl.ocks.org/mbostock/3886394.
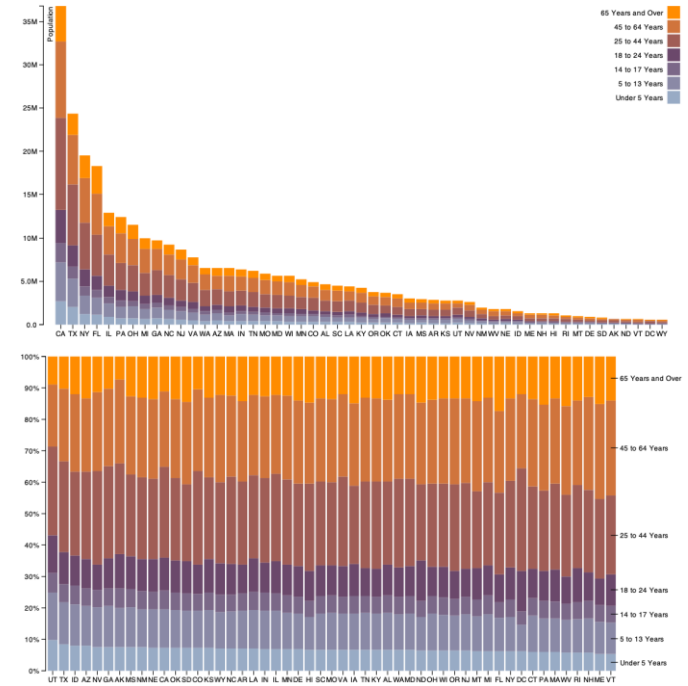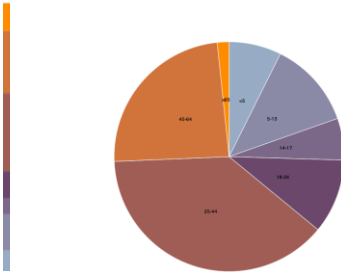
Idiom: **glyphmaps**

- rectilinear good for
  linear vs nonlinear trends

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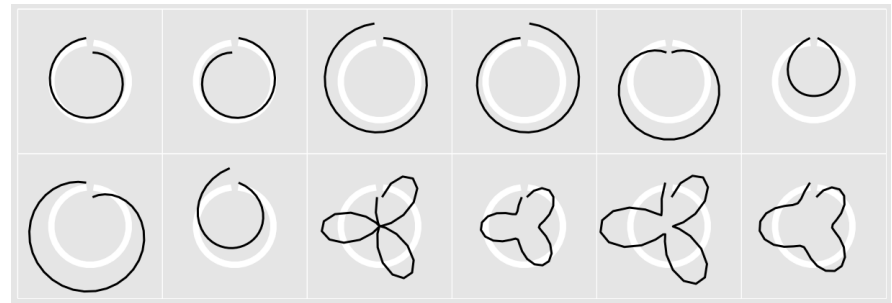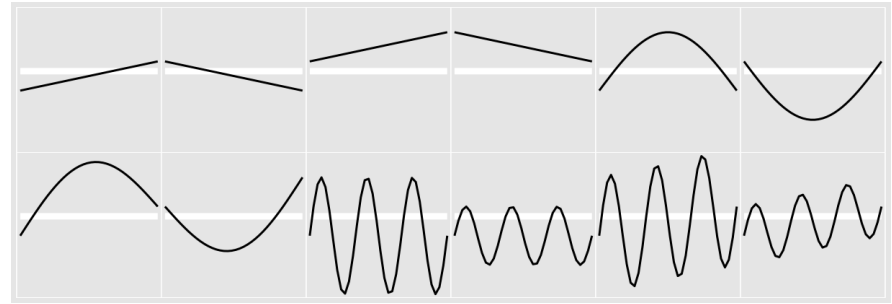
[Glyph-maps for Visually Exploring Temporal Patterns in Climate Data and Models. Wickham, Hofmann, Wickham, and Cook. Environmetrics 23:5 (2012), 382–393.]

Axis Orientation
→ Rectilinear   → Parallel   → Radial

➔ Axis Orientation
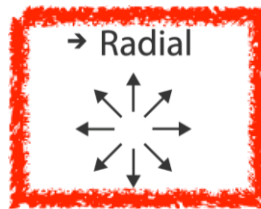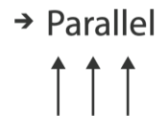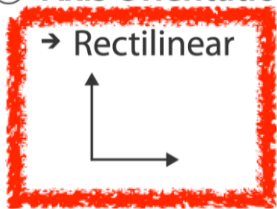
➔ Rectilinear   ➔ Parallel   ➔ Radial
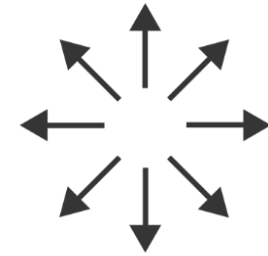
Idiom: **SPLOM**

- scatterplot matrix (SPLOM)
  - rectilinear axes, point mark
  - all possible pairs of axes
  - scalability
    - one dozen attribs
    - dozens to hundreds of items
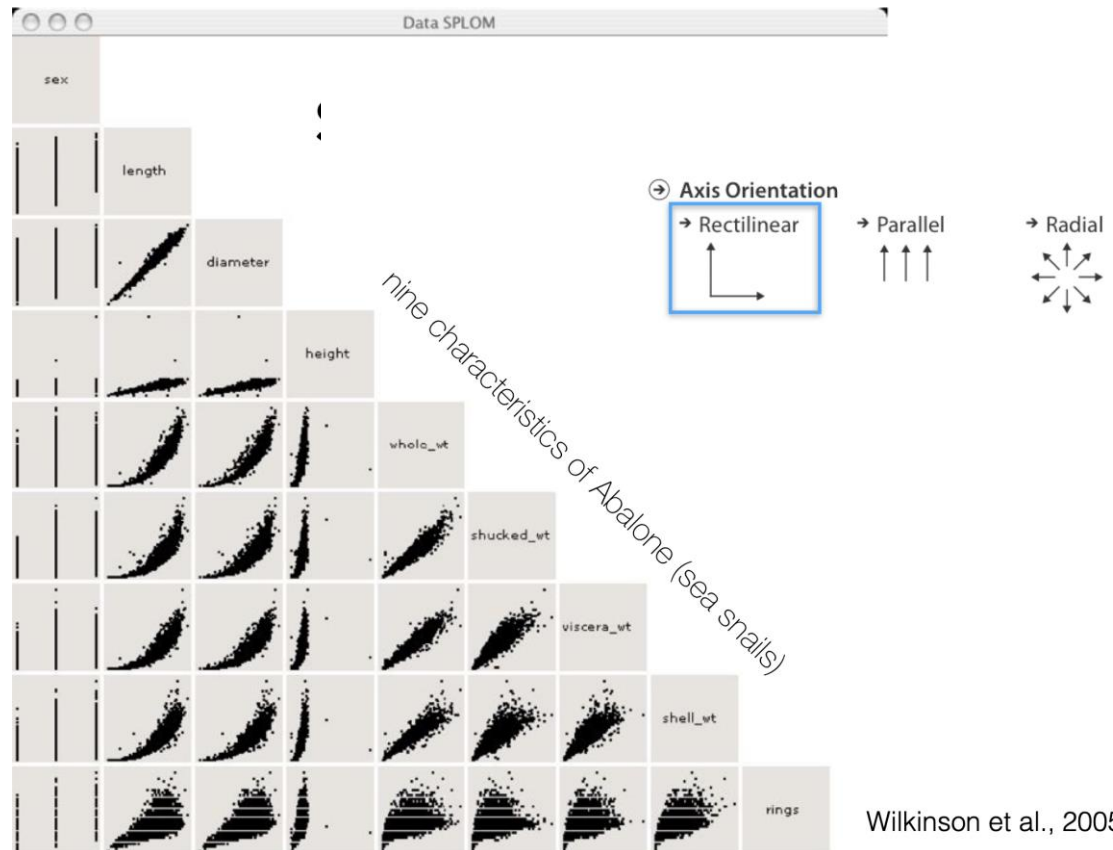
Wilkinson et al., 2005

Idioms: **parallel coordinates**

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

Scatterplot Matrix

	Math	Physics	Dance	Drama
Math				
Physics				
Dance				
Drama				

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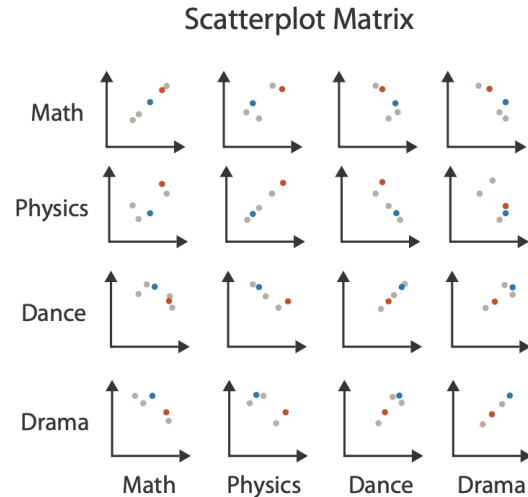
after [Visualization Course Figures. McGuffin, 2014.
http://www.michaelmcguffin.com/courses/vis/]

Idioms: **parallel coordinates**

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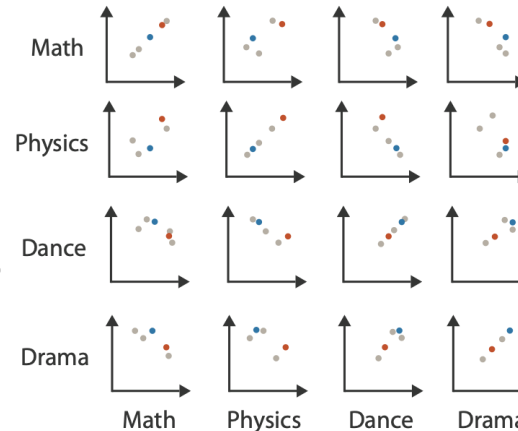- alternative: line up axes in parallel to show many attributes with position
  - item encoded with a line with n segments
  - n is the number of attributes shown
- parallel coordinates
  - parallel axes, jagged line for item
  - rectilinear axes, item as point
    - axis ordering is major challenge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

Scatterplot Matrix

Parallel Coordinates
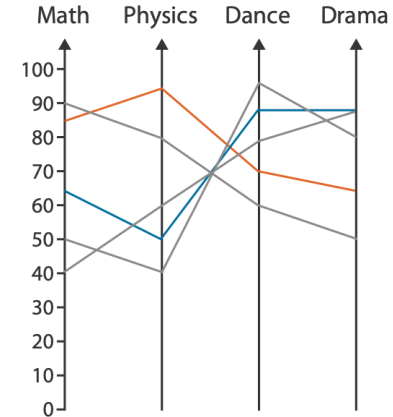
Table

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/]

52

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
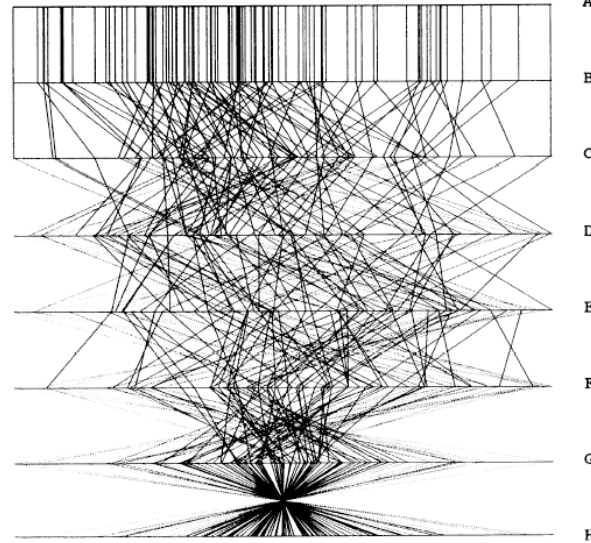 - uncorrelated
    - scattered crossings


https://www.mathsisfun.com/data/scatter-xy-plots.html
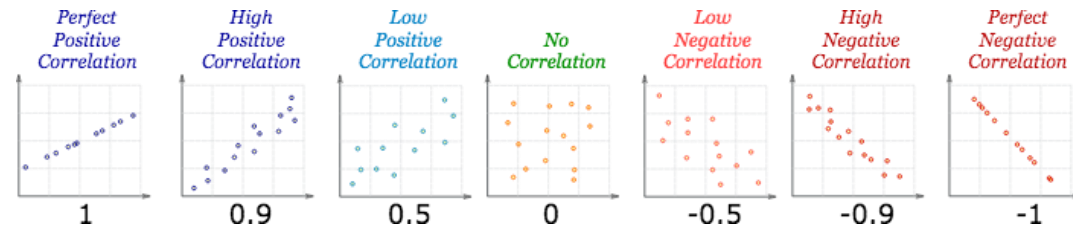
Figure 3. Parallel Coordinate Plot of Six-Dimensional Data Illustrating Correlations of $\rho = 1, .8, .2, 0, -.2, -.8,$ and -1.

[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
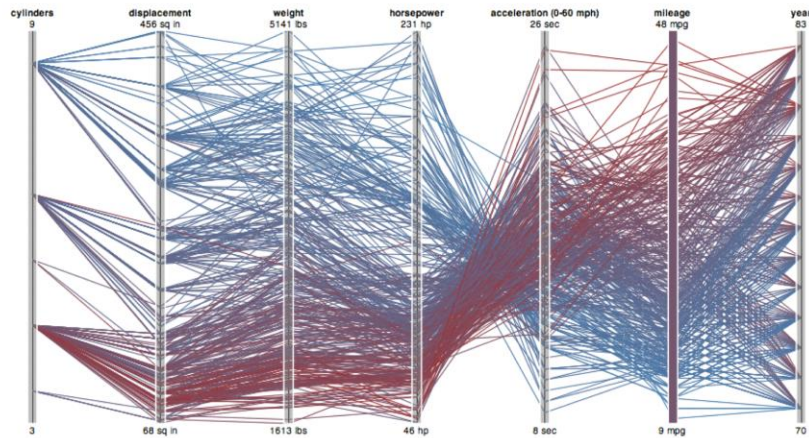
Orientation limitations

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

Axis Orientation
→ Rectilinear

Orientation limitations

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

⊕ **Axis Orientation**
  → Rectilinear   → Parallel

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!
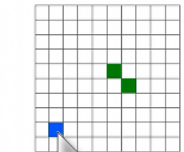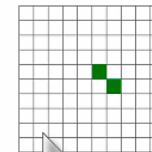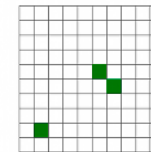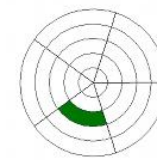   - for 2 attribs of very unequal importance

Axis Orientation
→ Rectilinear   → Parallel   → Radial

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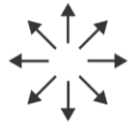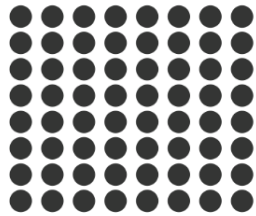
[Uncovering Strengths and Weaknesses of Radial Visualizations - an Empirical Approach. Diehl, Beck and Burch. IEEE TVCG (Proc. InfoVis) 16(6):935--942, 2010.]
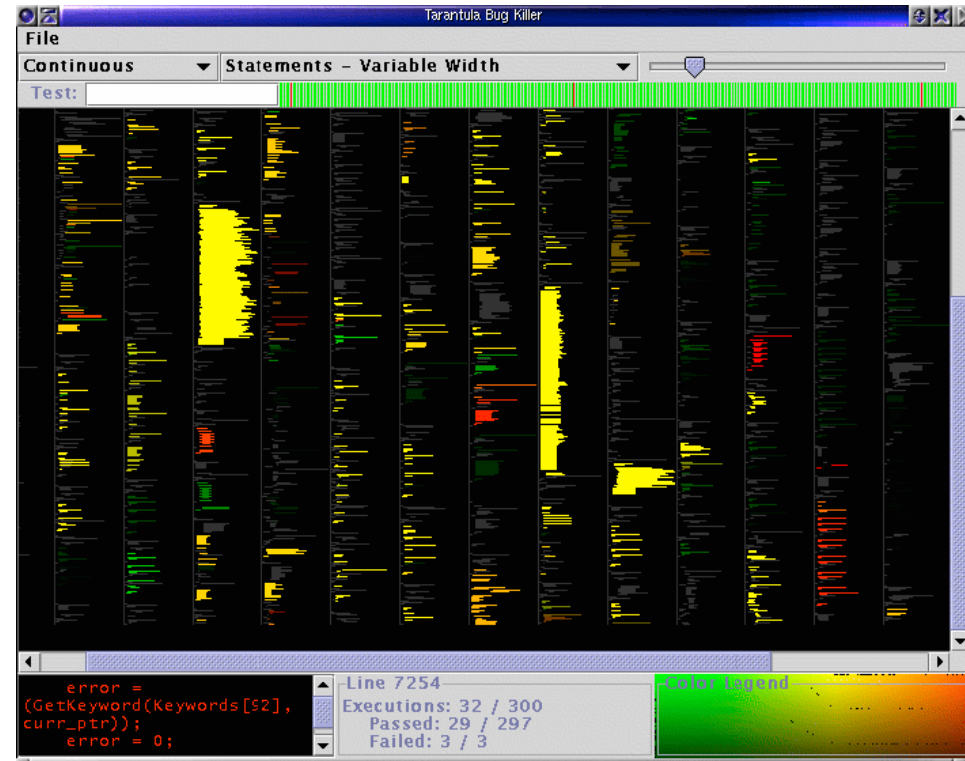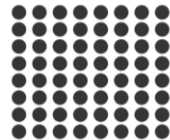
Layout density

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.]

I II III IV V VI VII VIII IX X XI XII XIII XIV XV

Reset   Pause   Slow   Fast

these spheres. It was really curious to note in cases of difficulty, as when two pieces of comb met at an angle, how often the bees would pull down and rebuild in different ways the same cell, sometimes recurring to a shape which they had at
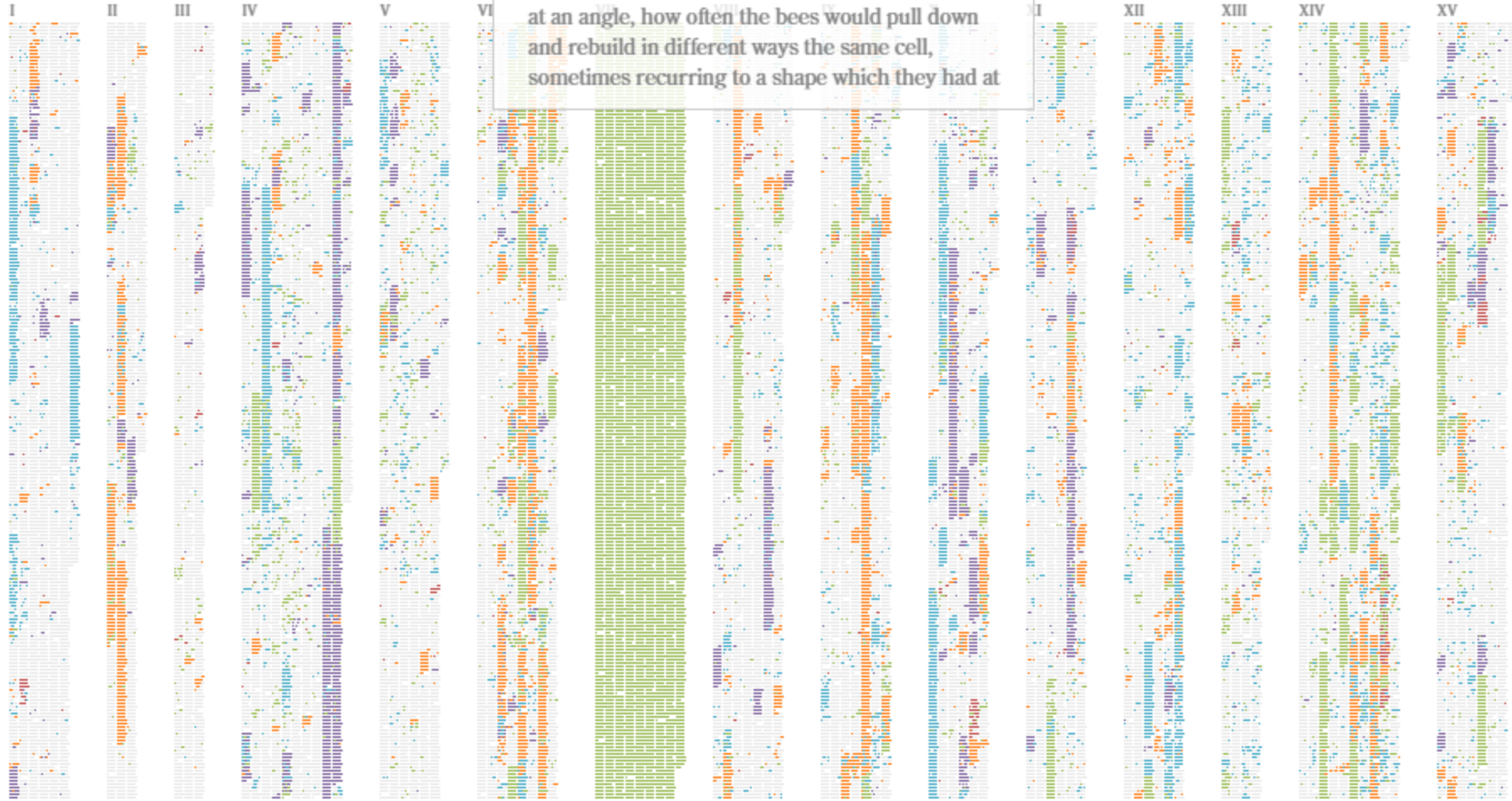
First Edition (1859)   Second Edition (1860)   Third Edition (1861)   Fourth Edition (1866)   Fifth Edition (1869)   Sixth Edition (1872)

https://benfry.com/traces/

Arrange tables

⊕ Express Values

⊕ Separate, Order, Align Regions

→ Separate

→ Order

→ Align

→ *1 Key*
List

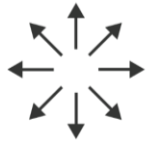→ *2 Keys*
Matrix

⊕ Axis Orientation

→ Rectilinear

→ Parallel

→ Radial

⊕ Layout Density

→ Dense