

[Modified slides from]
Visualization Analysis & Design
Full-Day Tutorial

Tamara Munzner

Department of Computer Science
University of British Columbia

*Sanger Institute / European Bioinformatics Institute
June 2014, Cambridge UK*

<http://www.cs.ubc.ca/~tmm/talks.html#minicourse14>

Outline

- **Visualization Analysis Framework**

Session 1 9:30-10:45am

- Introduction: Definitions
- Analysis: What, Why, How
- Marks and Channels

- **Idiom Design Choices, Part 2**

Session 3 1:15pm-2:45pm

- Manipulate: Change, Select, Navigate
- Facet: Juxtapose, Partition, Superimpose
- Reduce: Filter, Aggregate, Embed

- **Idiom Design Choices**

Session 2 11:00am-12:15pm

- **Arrange Tables**
- Arrange Spatial Data
- Arrange Networks and Trees
- Map Color

- **Guidelines and Examples**

Session 4 3-4:30pm

- Rules of Thumb
- Validation
- BioVis Analysis Example

How?

Encode

→ Arrange

→ Express



→ Separate



→ Order



→ Align



→ Use



What?

Why?

How?

→ Map

from **categorical** and **ordered** attributes

→ Color

→ Hue



→ Saturation



→ Luminance



→ Size, Angle, Curvature, ...



→ Shape



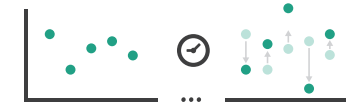
→ Motion

Direction, Rate, Frequency, ...

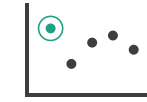


Manipulate

→ Change



→ Select



→ Navigate

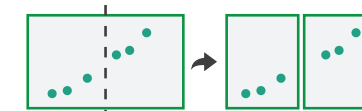


Facet

→ Juxtapose



→ Partition



→ Superimpose



Reduce

→ Filter



→ Aggregate



→ Embed



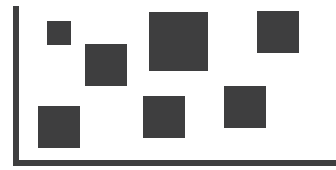
Arrange tables

→ Express Values

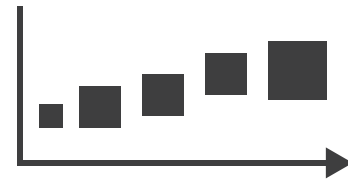


→ Separate, Order, Align Regions

→ Separate



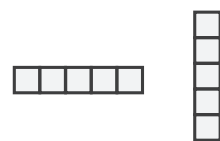
→ Order



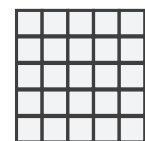
→ Align



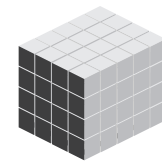
→ 1 Key *List*



→ 2 Keys *Matrix*



→ 3 Keys *Volume*

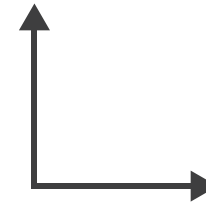


→ Many Keys *Recursive Subdivision*

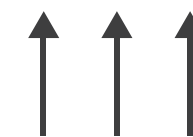


→ Axis Orientation

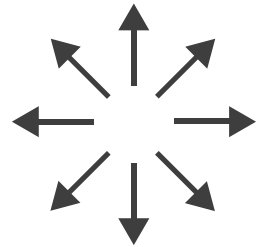
→ Rectilinear



→ Parallel

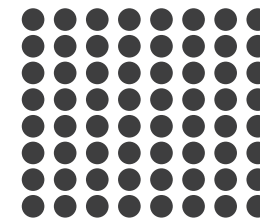


→ Radial



→ Layout Density

→ Dense



→ Space-Filling



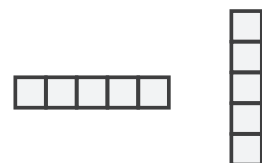
Keys and values

- **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 key count**
 - 0, 1, 2, many...

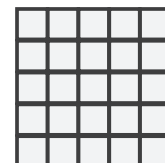
⇒ Express Values



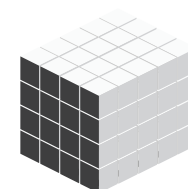
→ 1 Key
List



→ 2 Keys
Matrix



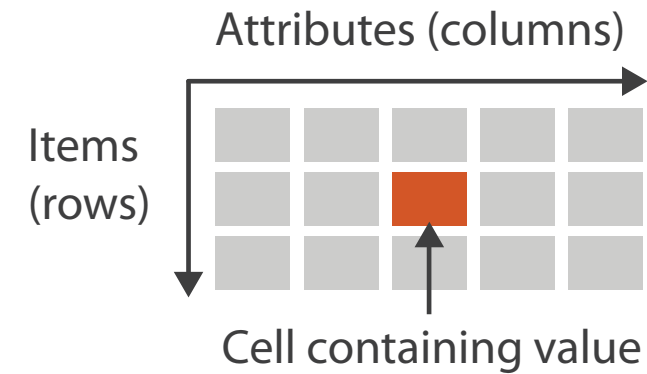
→ 3 Keys
Volume



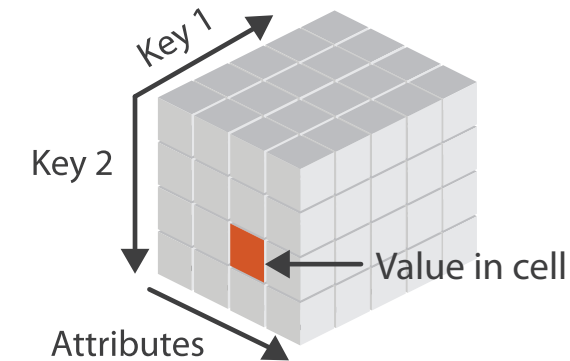
→ Many Keys
Recursive Subdivision



→ Tables



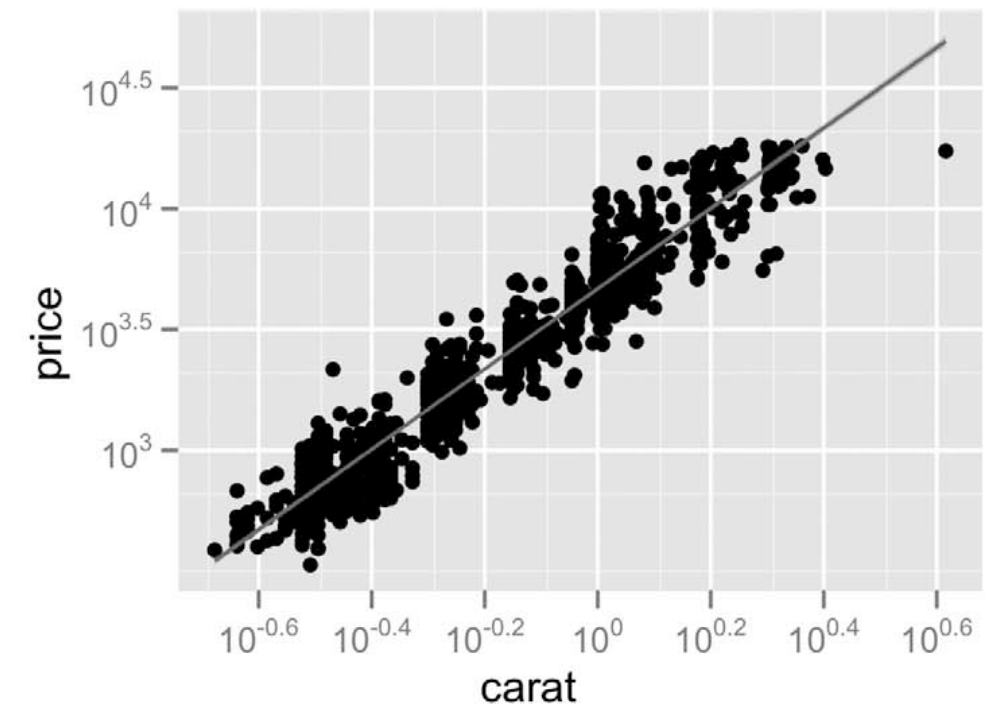
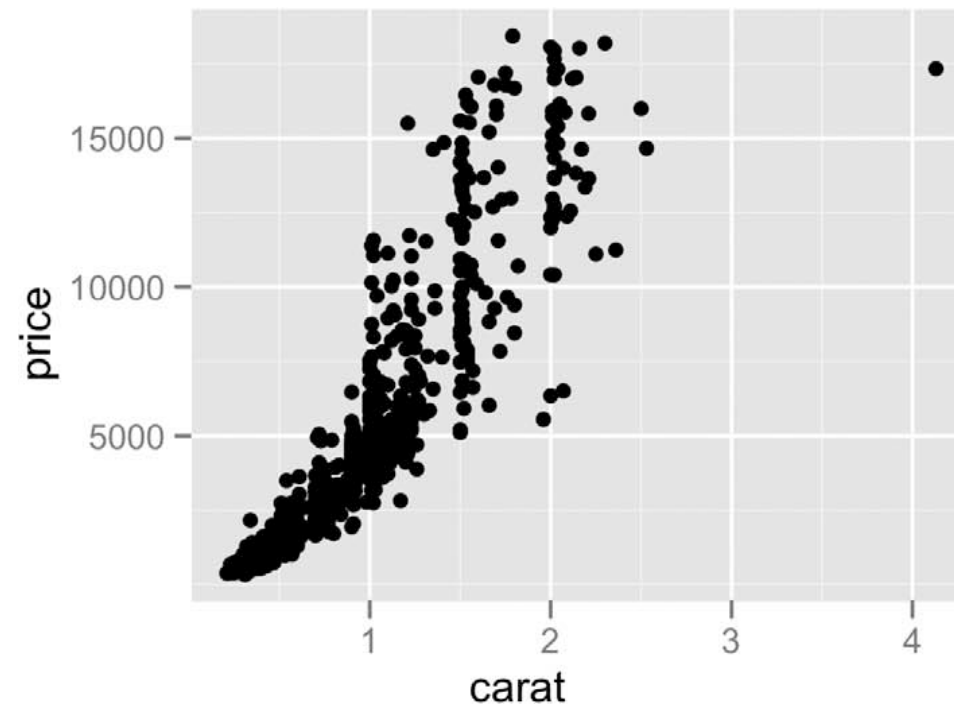
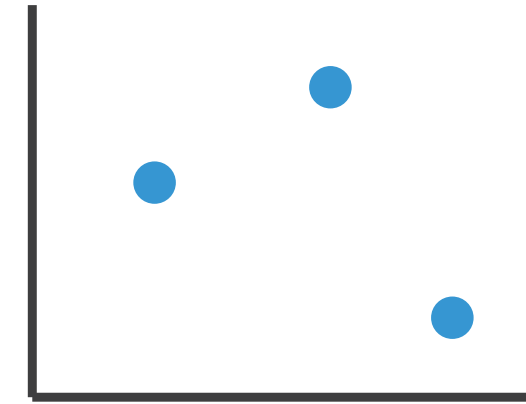
→ *Multidimensional Table*



Idiom: scatterplot

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

⊞ Express Values

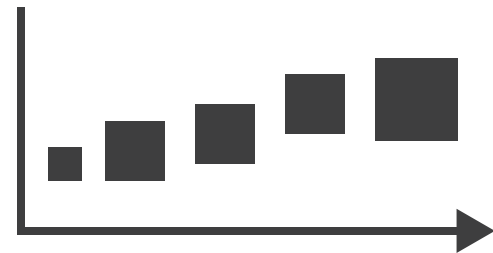


Some keys: Categorical regions

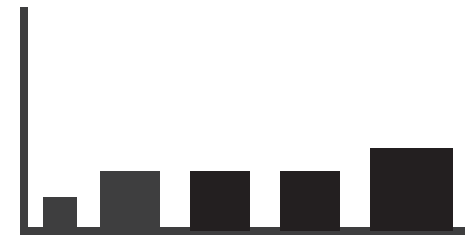
→ Separate



→ Order

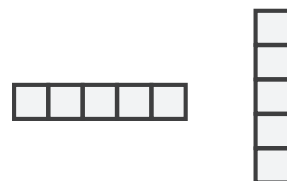


→ Align

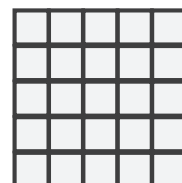


- **regions:** contiguous bounded areas distinct from each other
 - using space to *separate* (proximity)
 - following expressiveness principle for categorical attributes
- use ordered attribute to *order* and *align* regions

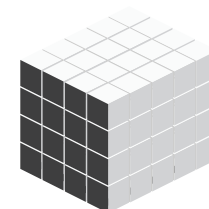
→ 1 Key
List



→ 2 Keys
Matrix



→ 3 Keys
Volume

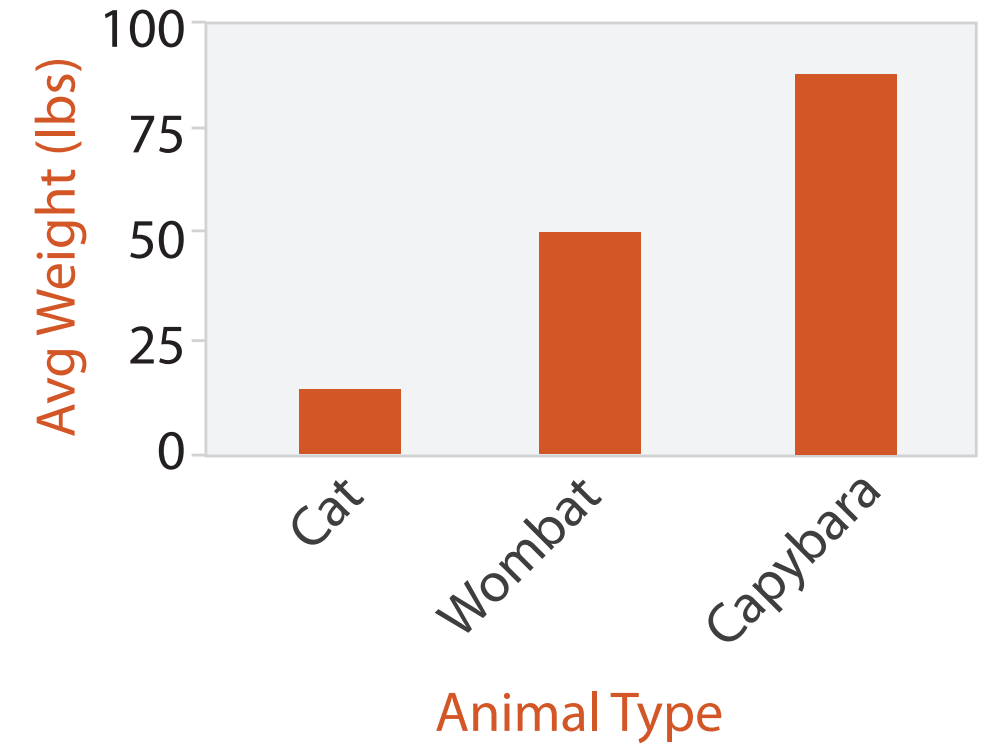
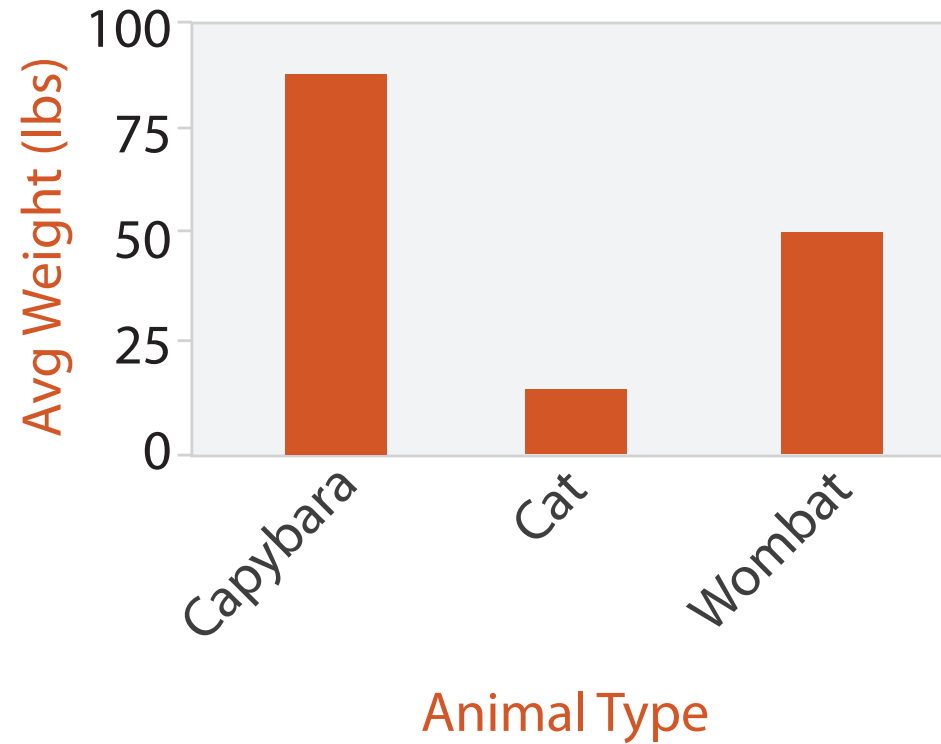


→ Many Keys
Recursive Subdivision



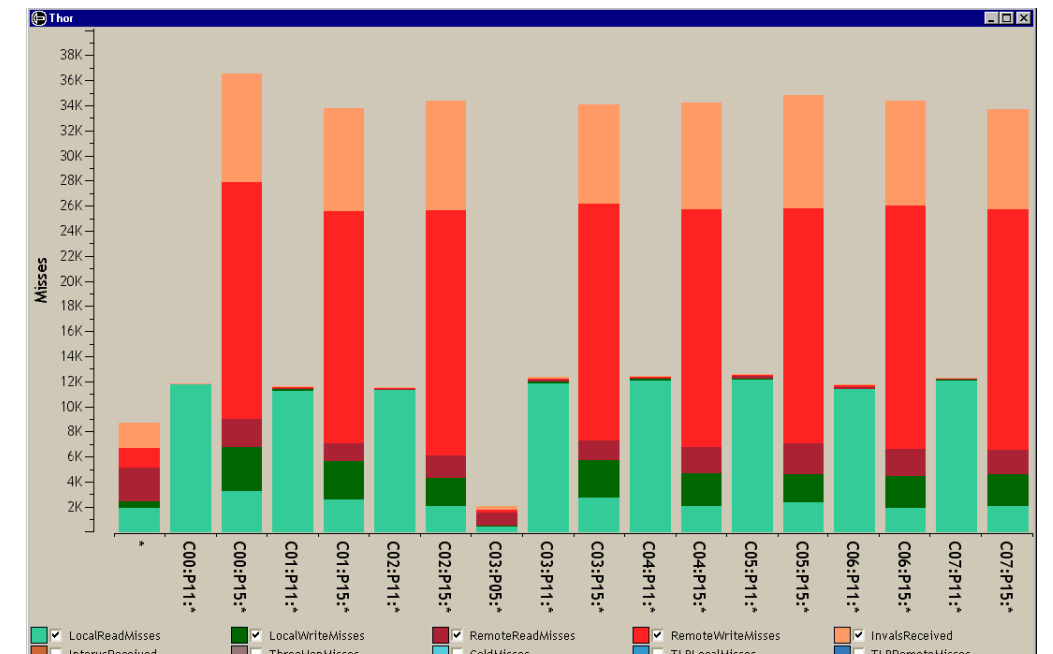
Idiom: bar chart

- 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



Idiom: stacked bar chart

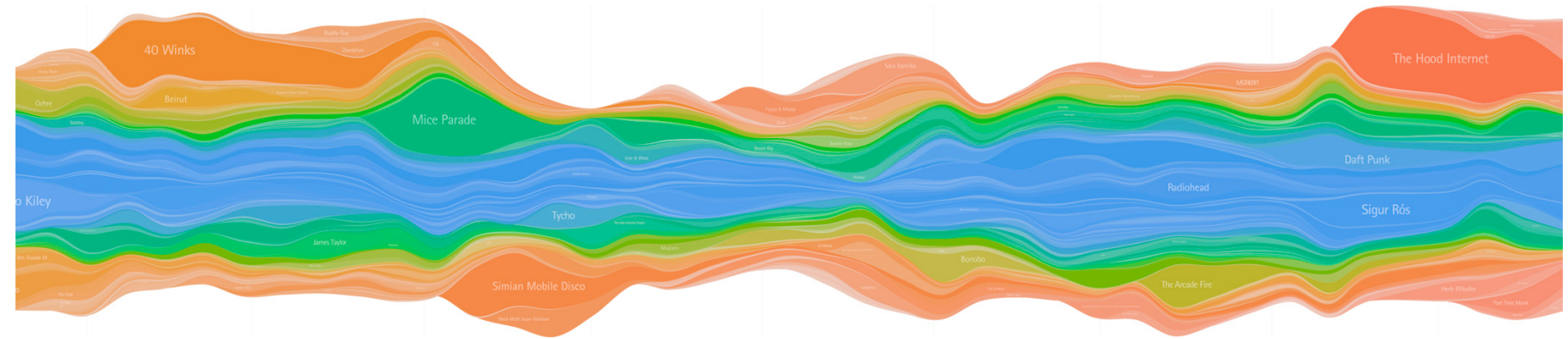
- one more key
 - data
 - 2 categ attrib, 1 quant attrib
 - mark: vertical stack of line marks
 - glyph: composite object, internal structure from multiple marks
 - channels
 - length [quant]
 - color hue [categ]
 - spatial regions: one per glyph [categ]
 - aligned: full glyph, lowest bar component
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



[Using Visualization to Understand the Behavior of Computer Systems. Bosch. Ph.D. thesis, Stanford Computer Science, 2001.]

Idiom: streamgraph

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



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

Idiom: **line chart**

- one key, one value

- data

- 2 quant attribs

- mark: points

- line connection marks between them

- channels

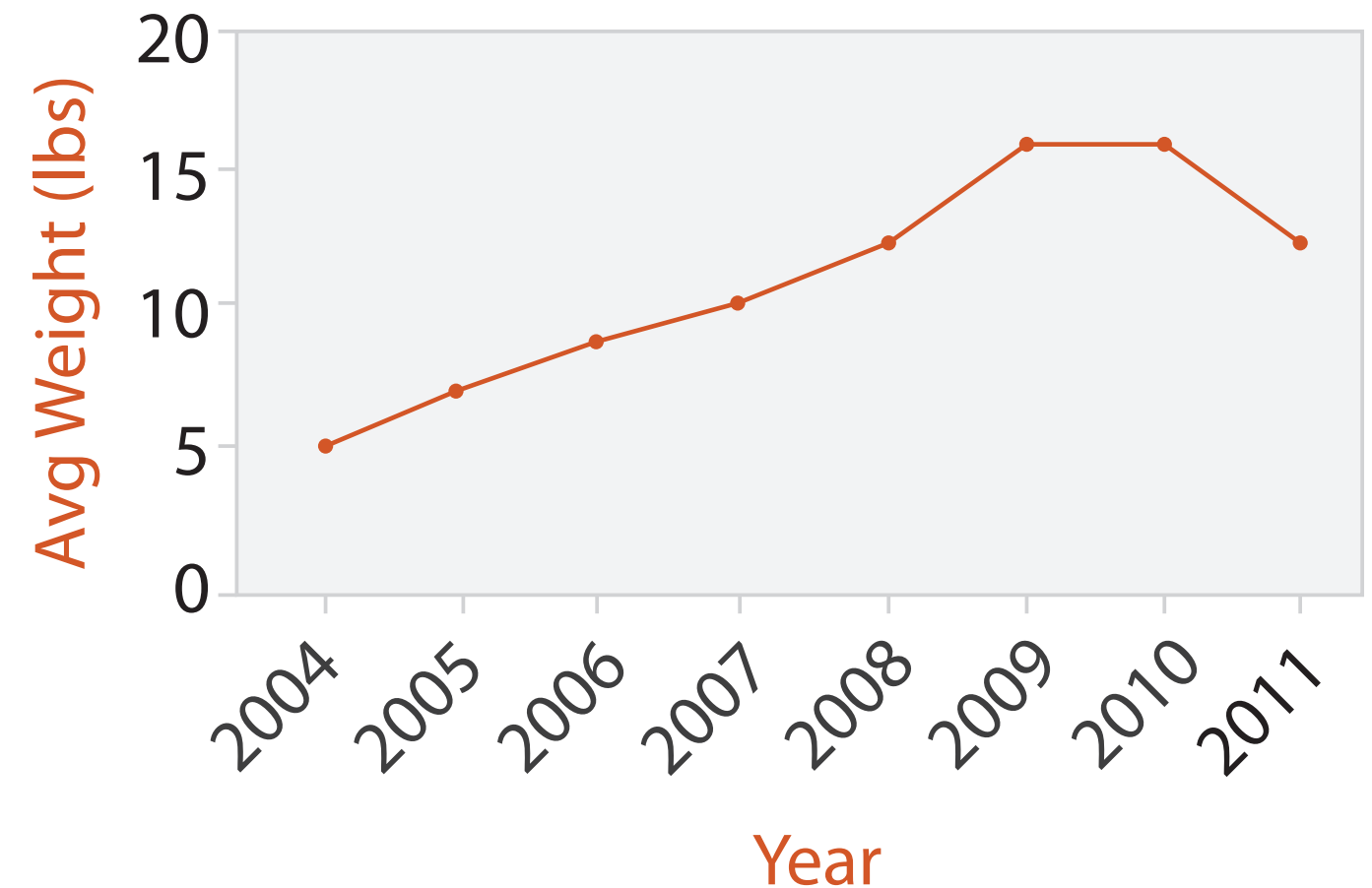
- aligned lengths to express quant value

- separated and ordered by key attrib into horizontal regions

- task

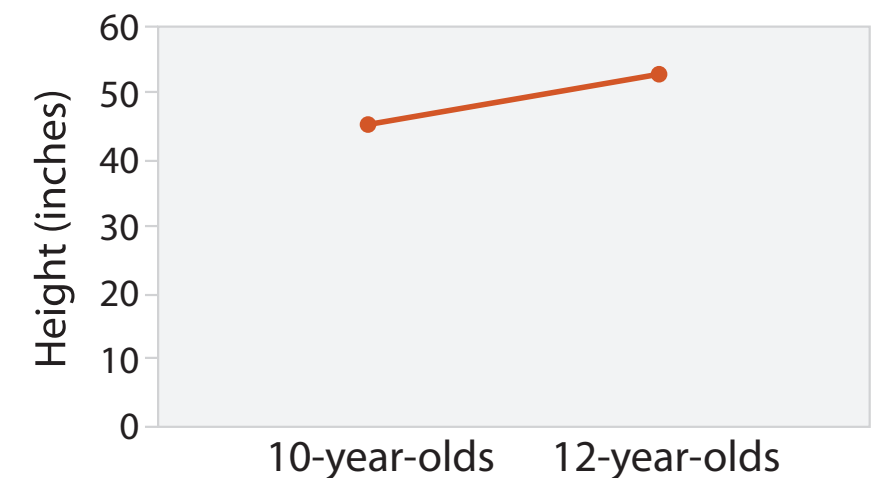
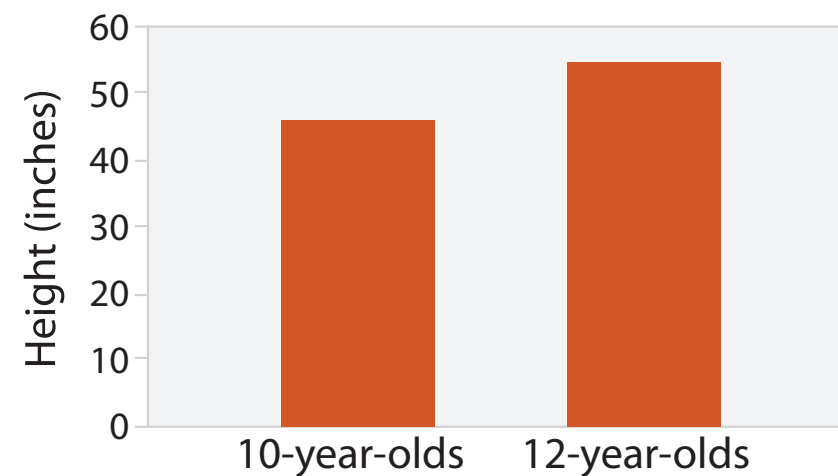
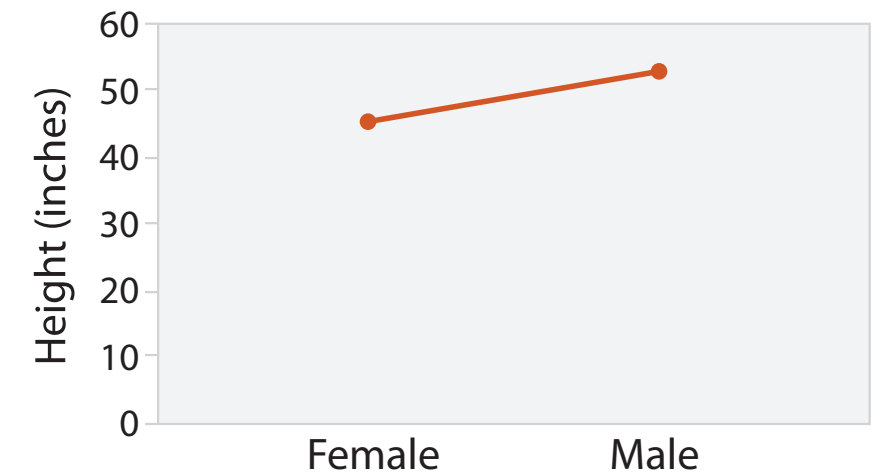
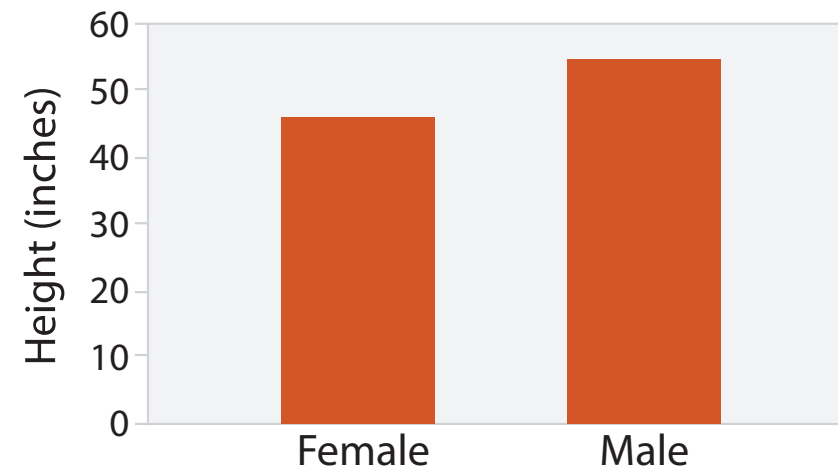
- find trend

- *connection marks emphasize ordering* of items along key axis by explicitly showing relationship between one item and the next



Choosing bar vs line charts

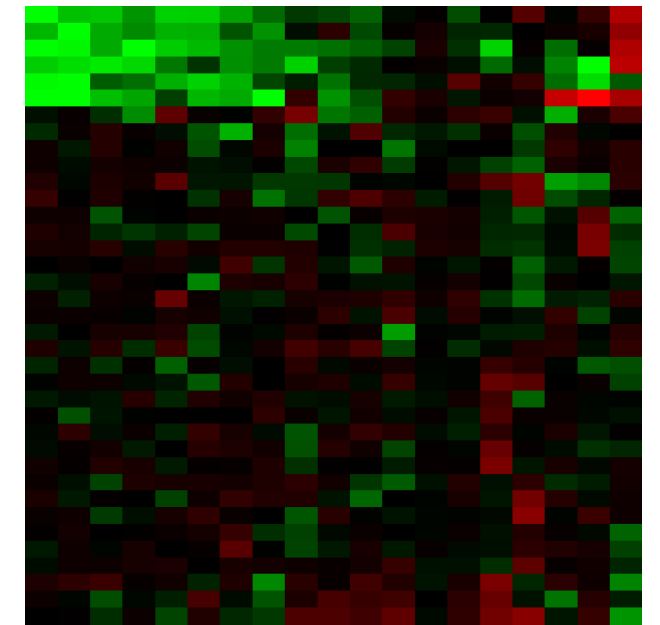
- depends on type of key attrib
 - bar charts if categorical
 - line charts if ordered
- do not use line charts for categorical key attribs
 - violates expressiveness principle
 - implication of trend so strong that it overrides semantics!
 - “The more male a person is, the taller he/she is”



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

Idiom: heatmap

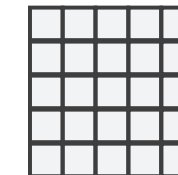
- two keys, one value
 - data
 - 2 categ attribs (gene, experimental condition)
 - 1 quant attrib (expression levels)
 - marks: area
 - separate and align in 2D matrix
 - indexed by 2 categorical attributes
 - channels
 - color by quant attrib
 - (ordered diverging colormap)
 - task
 - find clusters,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



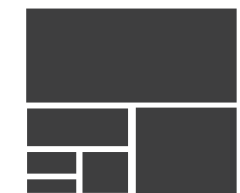
→ 1 Key
List



→ 2 Keys
Matrix

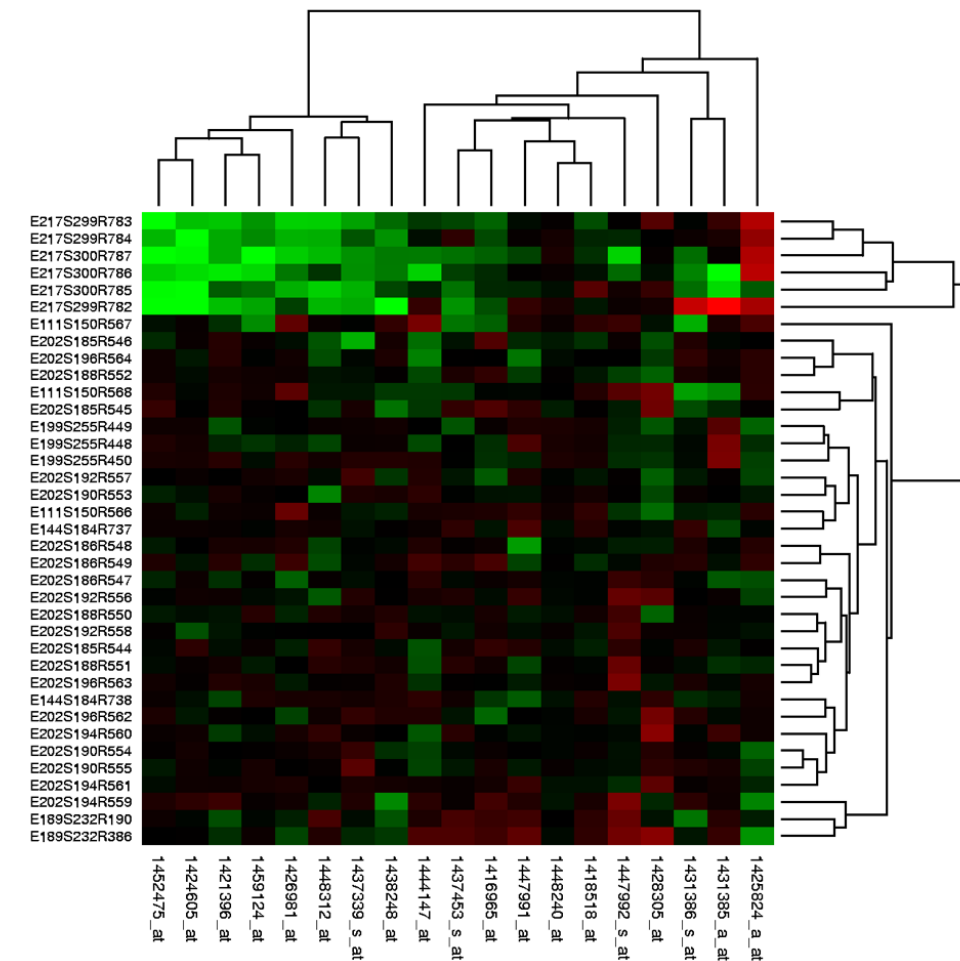


→ Many Keys
Recursive Subdivision



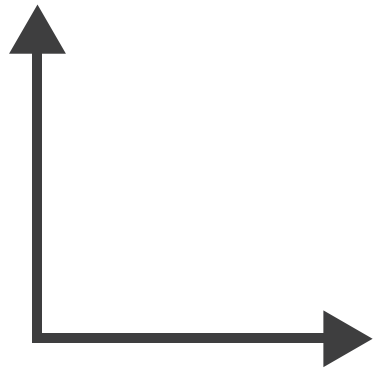
Idiom: cluster heatmap

- 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

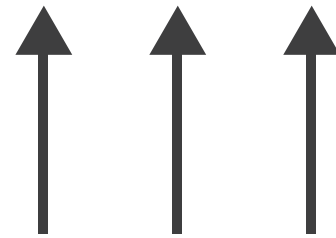


➔ Axis Orientation

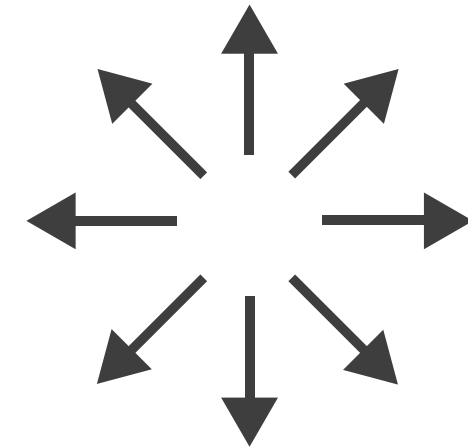
➔ Rectilinear



➔ Parallel



➔ Radial



Idioms: scatterplot matrix, parallel coordinates

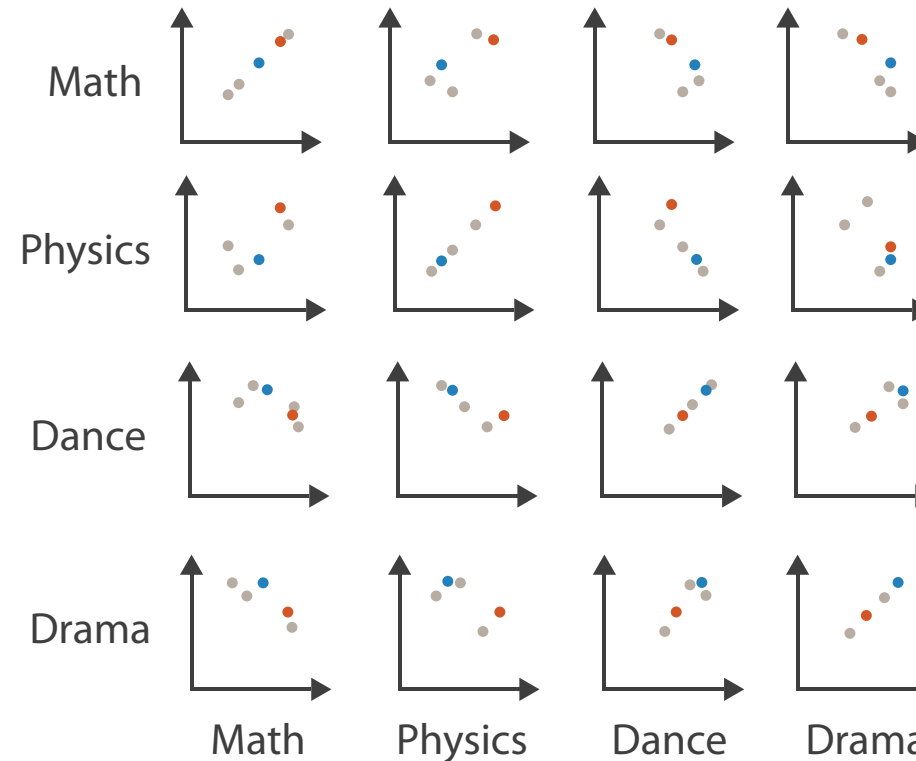
- scatterplot matrix (SPLOM)

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

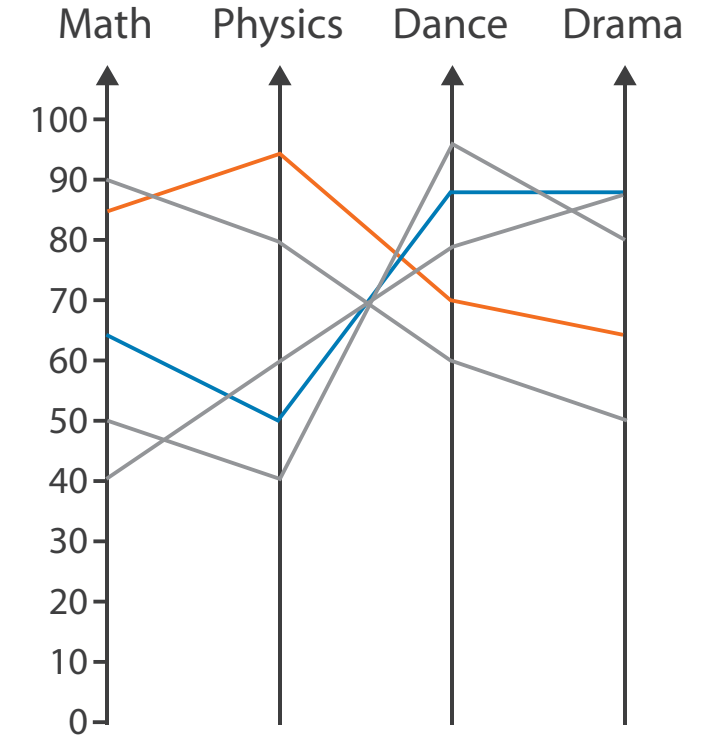
- parallel coordinates

- parallel axes, jagged line representing item
- rectilinear axes, item as point
 - axis ordering is major challenge (or opportunity!)
- scalability
 - dozens of attribs

Scatterplot Matrix



Parallel Coordinates

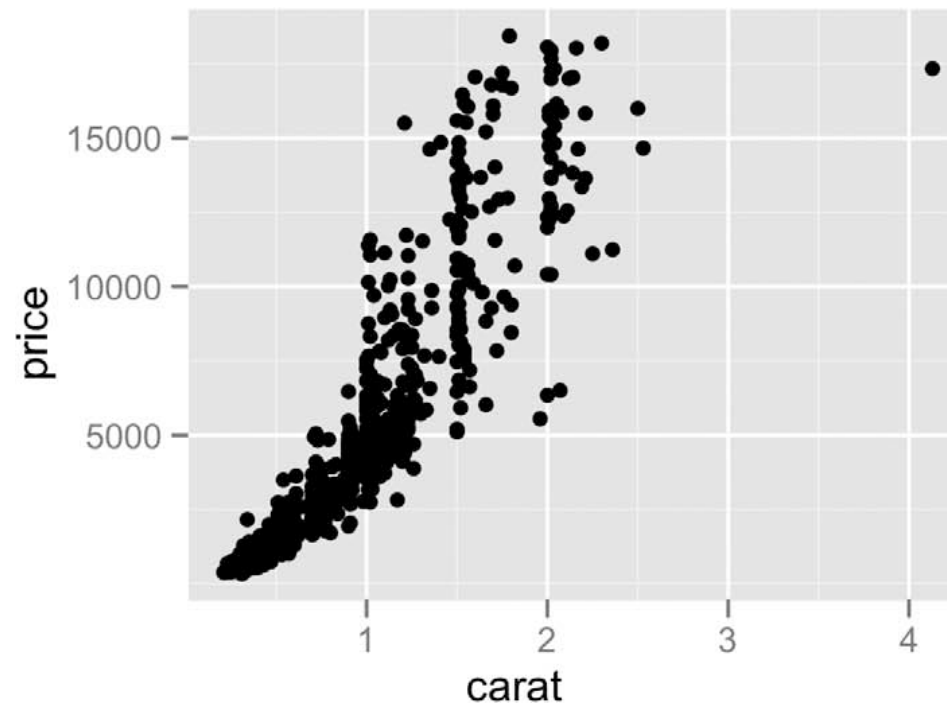


Table

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

Task: Correlation

- scatterplot matrix
 - positive correlation
 - diagonal low-to-high
 - negative correlation
 - diagonal high-to-low
 - uncorrelated
- parallel coordinates
 - positive correlation
 - parallel line segments
 - negative correlation
 - all segments cross at halfway point
 - uncorrelated
 - scattered crossings



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

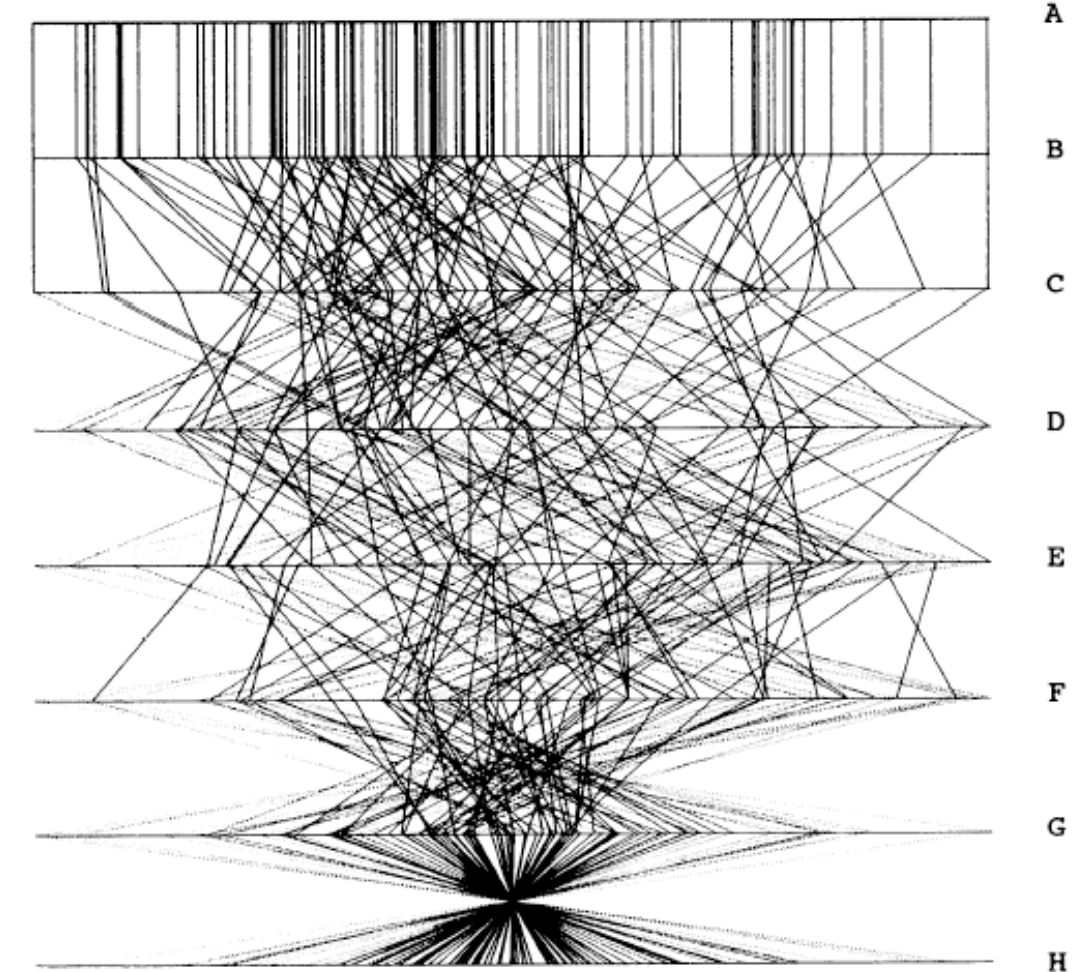
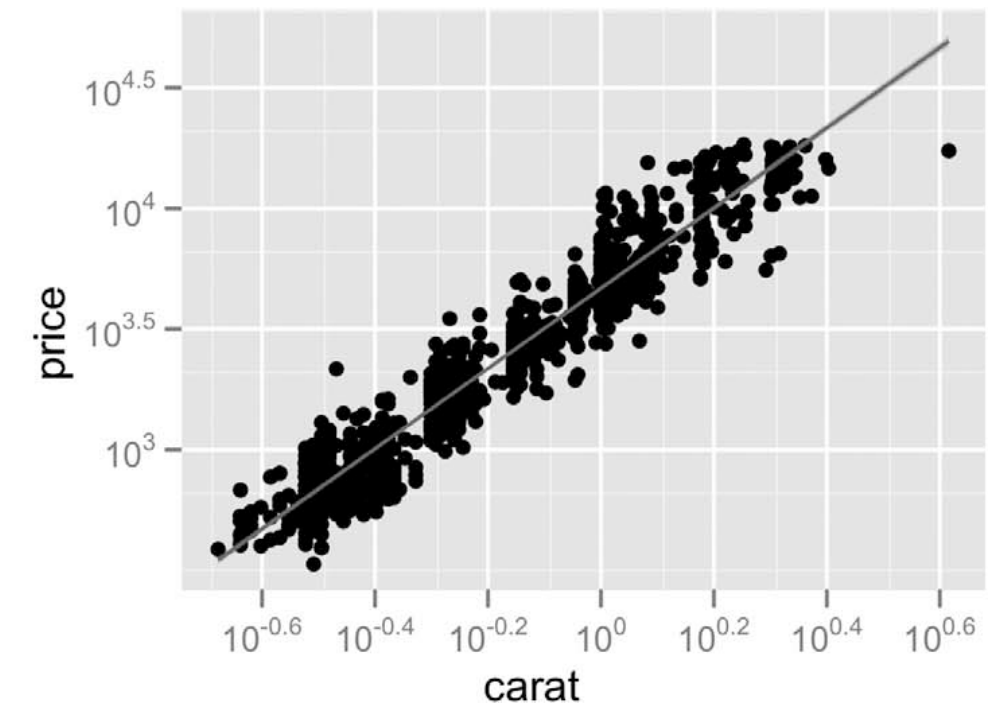
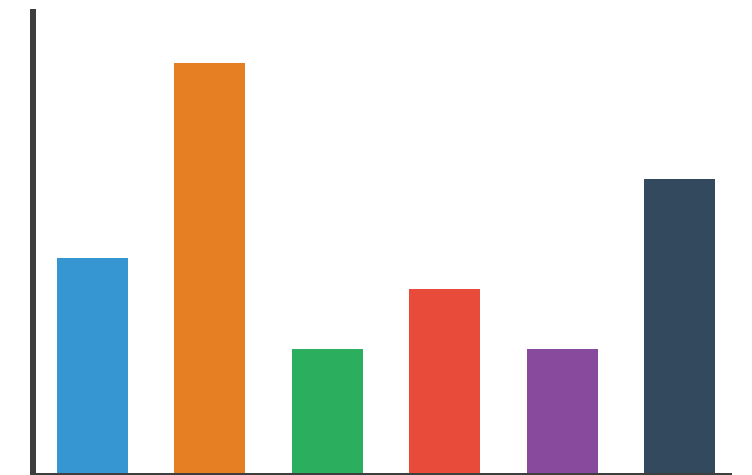
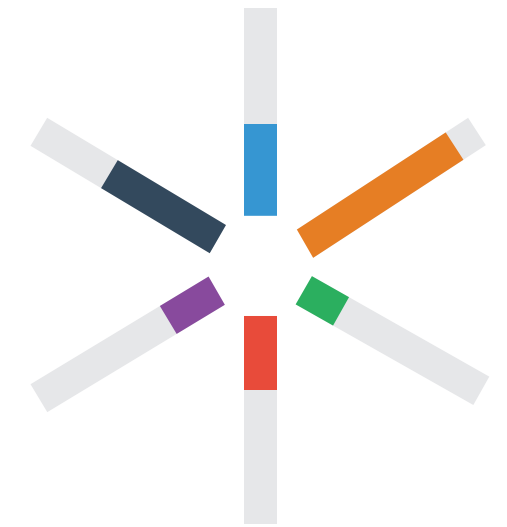


Figure 3. Parallel Coordinate Plot of Six-Dimensional Data Illustrating Correlations of $\rho = 1, .8, .2, 0, -.2, -.8, \text{ and } -1$.

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

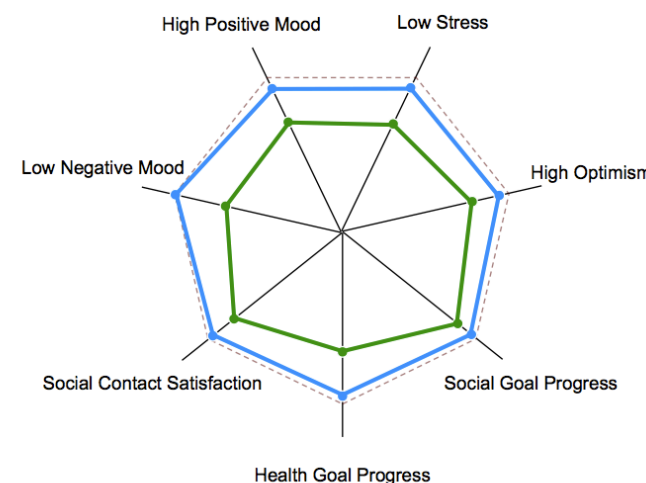
Idioms: radial bar chart, star plot

- radial bar chart
 - radial axes meet at central ring, line mark
- star plot
 - radial axes, meet at central point, line mark
- bar chart
 - rectilinear axes, aligned vertically
- accuracy
 - length unaligned with radial
 - less accurate than aligned with rectilinear



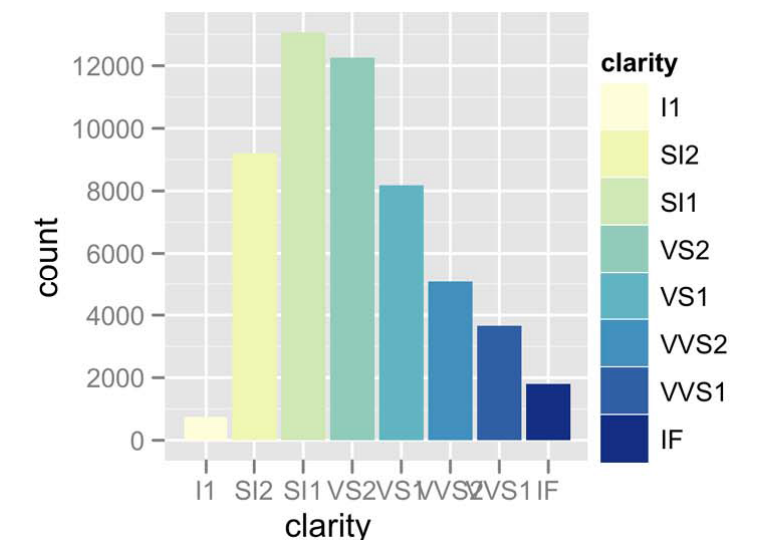
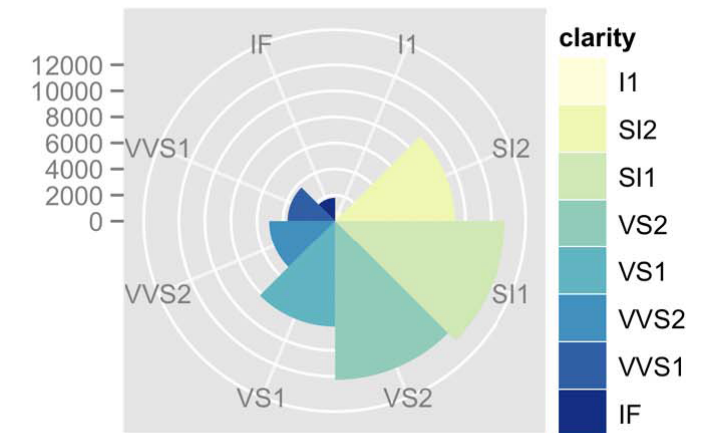
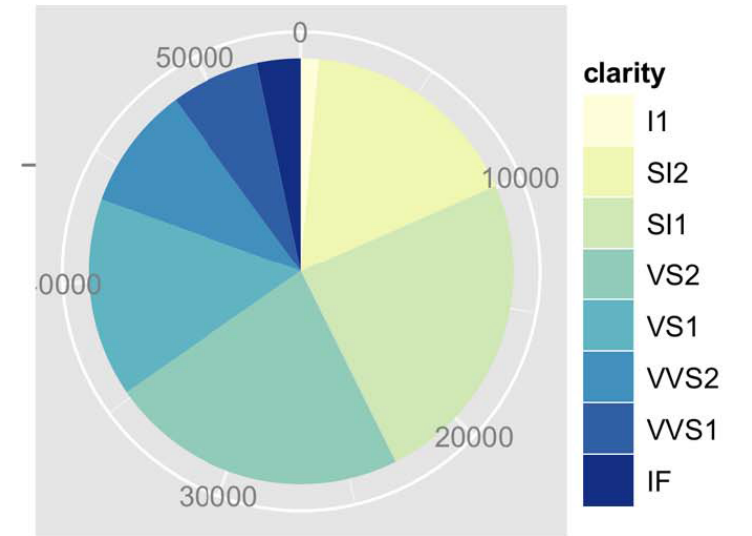
Idioms: radar chart

- radar plot
 - radial axes, meet at central point
- what is the mark?
- how does it effect expressiveness?



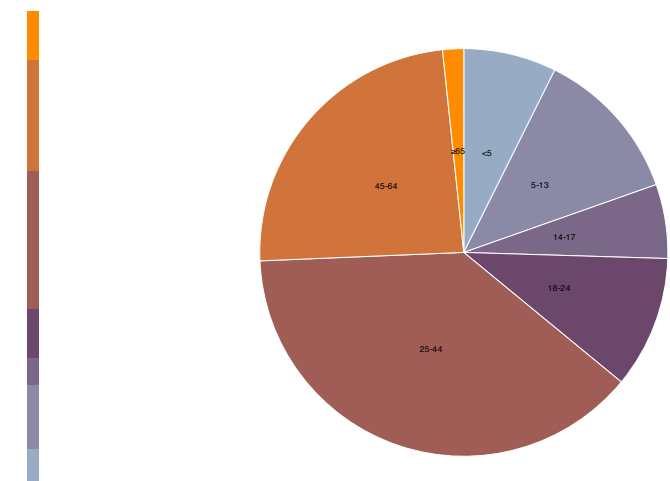
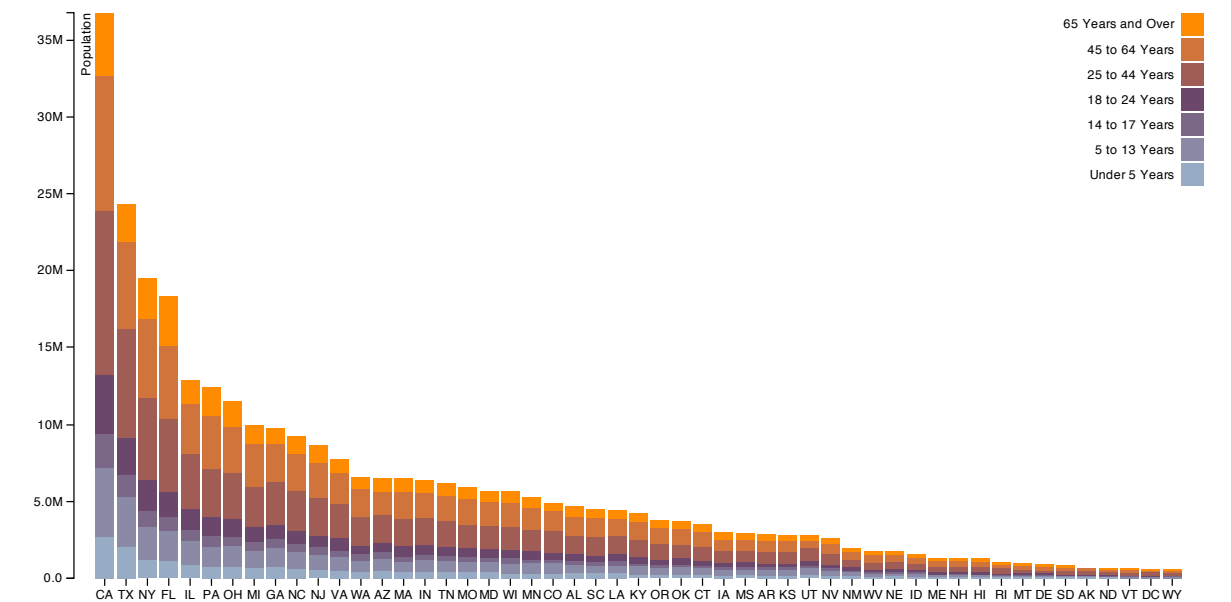
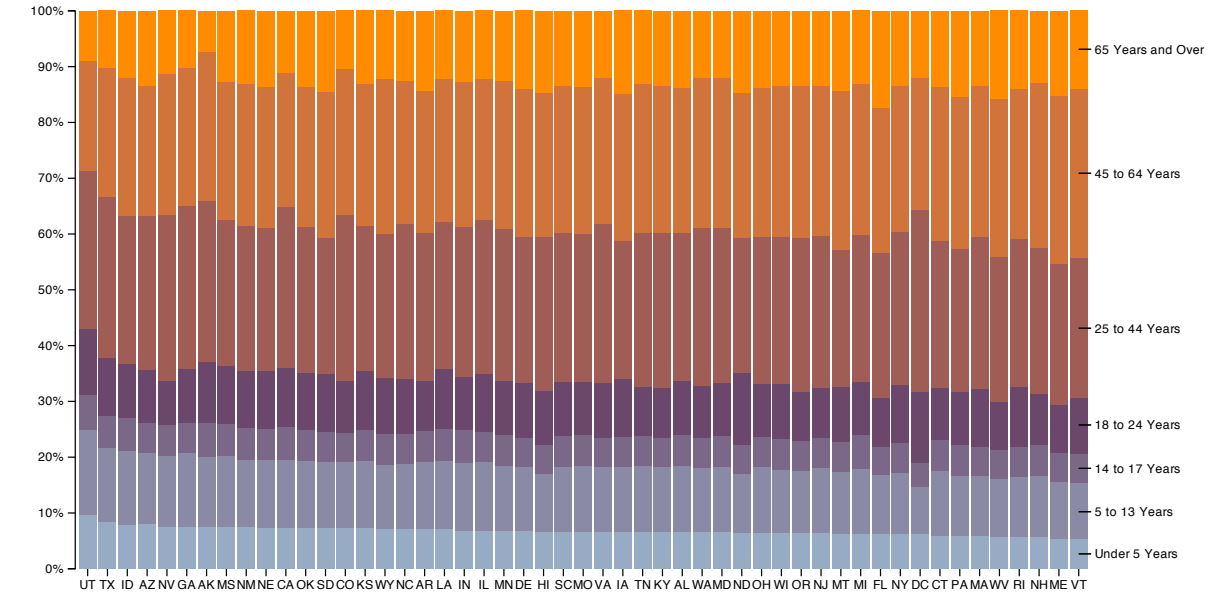
Idioms: pie chart, polar area chart

- pie chart
 - area marks with angle channel
 - accuracy: angle/area much less accurate than line length
- polar area chart
 - area marks with length channel
 - more direct analog to bar charts
- data
 - I categ key attrib, I quant value attrib
- task
 - part-to-whole judgements



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/3887235>,

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

<http://bl.ocks.org/mbostock/3886394>.

Further reading

- Visualization Analysis and Design. Munzner. AK Peters / CRC Press, Oct 2014.
– *Chap 7: Arrange Tables*
- Visualizing Data. Cleveland. Hobart Press, 1993.
- *A Brief History of Data Visualization*. Friendly. 2008.
<http://www.datavis.ca/milestones>