Arrange Tables

Keys and values

The distinction between key and value attributes is very relevant to visually encoding table data.

- A key - an independent attribute
  - used as a unique index to look up items in a table
  - categorical or ordinal
- a value is a dependent attribute: the value of a cell in a table
  - values can be categorical, ordinal, or quantitative

How to Arrange?

Arrange

→ Express

→ Separate

→ Order

→ Align

→ Use

How ? ? ?
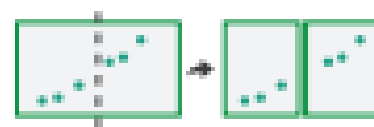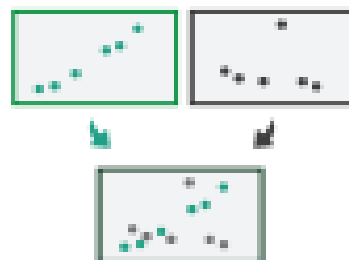
Manipulate	Facet	Reduce
Change	Juxtapose	Filter
Select	Partition	Aggregate
Navigate	Superimpose	Embed

Arrange Tables

Express Values

Separate, Order, Align Regions
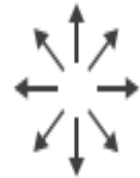→ Separate   → Order

→ Align

Axis Orientation
→ Rectilinear   → Parallel   → Radial
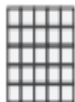
Layout Density
→ Dense   → Space-Filling

→ 1 Key   → 2 Keys   → 3 Keys   → Many Keys
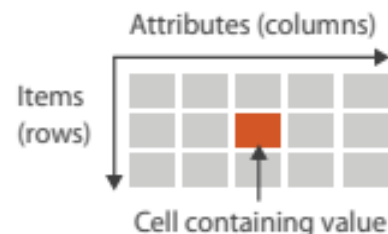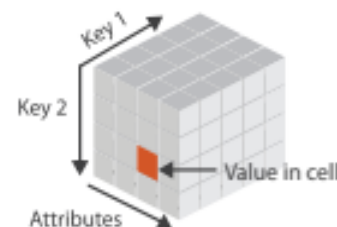List   Matrix   Volume   Recursive Subdivision

Keys and Values

→ Tables

- 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
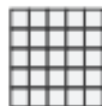- classify arrangements by key count
  - 0, 1, 2, many...

Attributes (columns)

Items (rows)

Cell containing value

→ *Multidimensional Table*

Key 1

Key 2

Value in cell

Attributes

⊙ **Express Values**

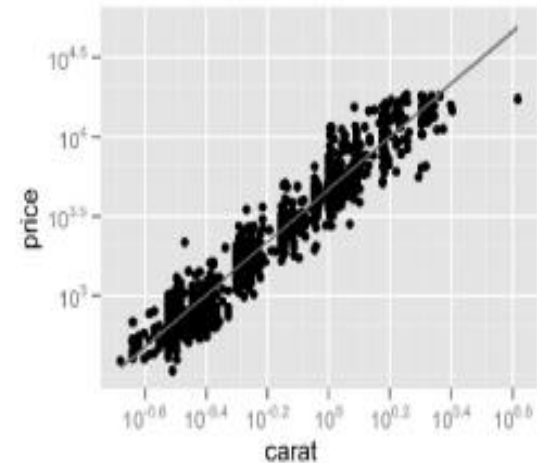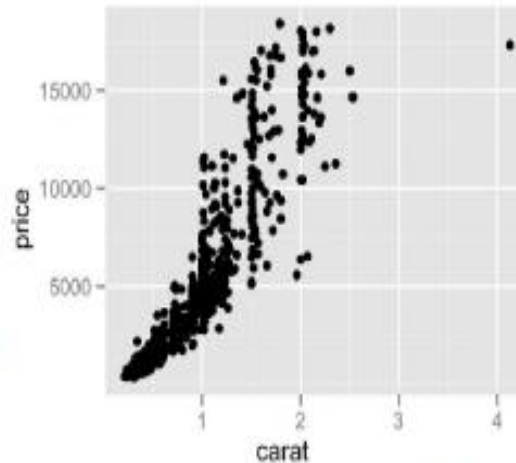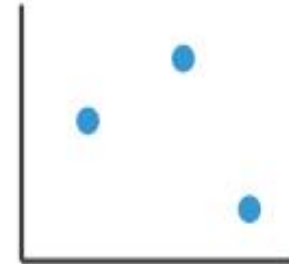|	→ 1 Key *List*	→ 2 Keys *Matrix*	→ 3 Keys *Volume*	→ Many Keys *Recursive Subdivision*

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
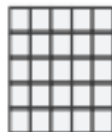- use ordered attribute to **order** and **align** regions

→ 1 Key
List

→ 2 Keys
Matrix

→ 3 Keys
Volume

→ Many Keys
Recursive Subdivision

Bar chart

- one key, one value
  - data
    - I categ attrib, I quant attrib
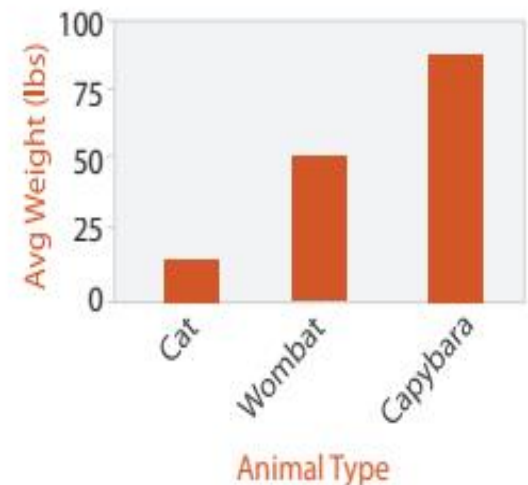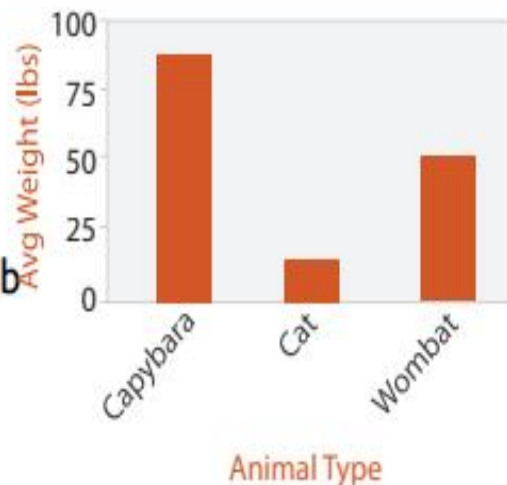 - mark: lines
  - channels
    - length to express quant value
    - spatial regions: one per mark
      - separated horizontally, aligned vertically
      - ordered by quant attrib
        - » by label (alphabetical), by length attrib (data-driven)
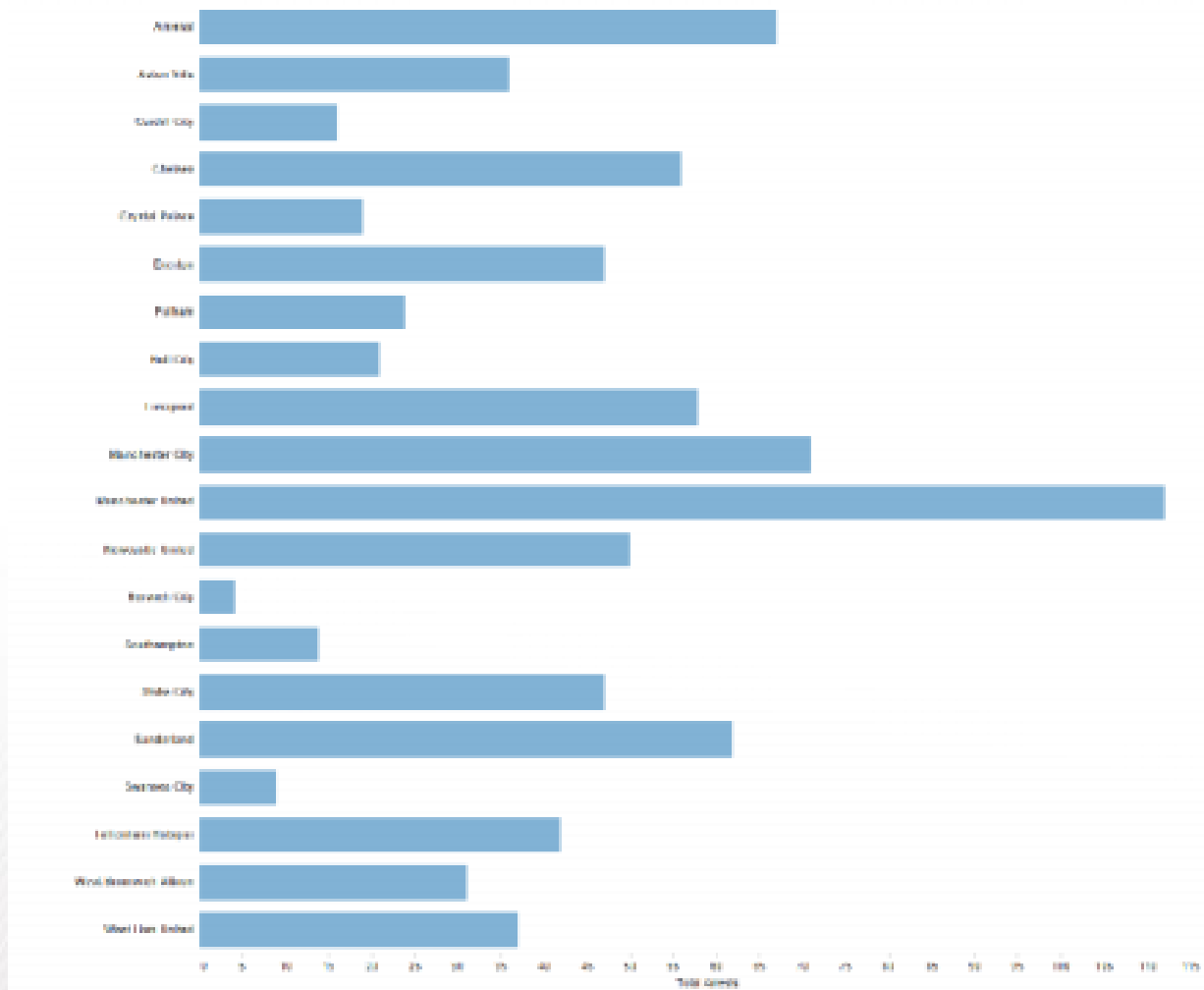 - task
    - compare, lookup values
  - scalability
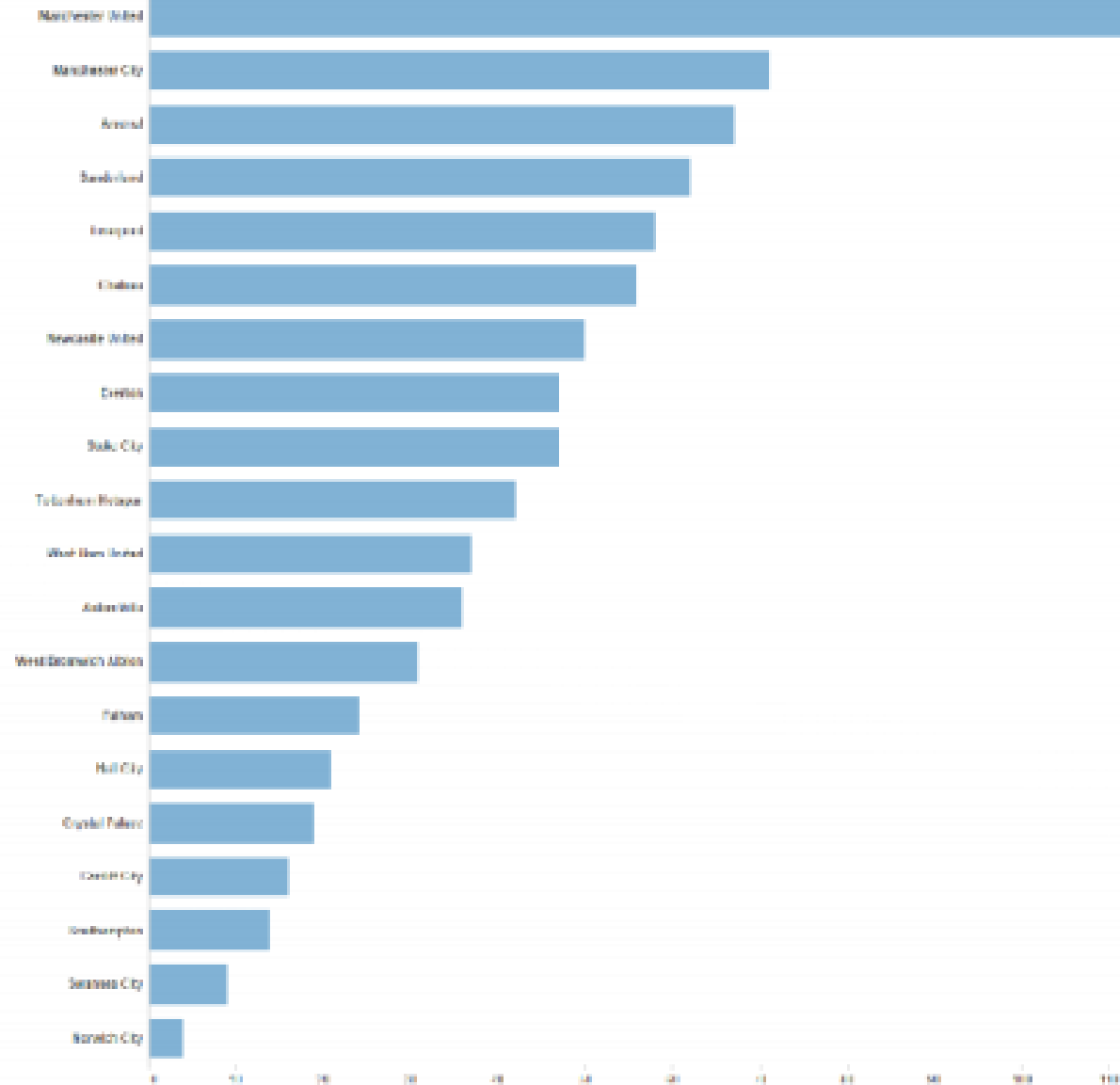   - dozens to hundreds of levels for key attrib

Separated and aligned but not ordered

LIMITATION: Hard to know rank. What's the 4th most? The 7th?

Separated aligned and ordered

Line chart / dot plot

- 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
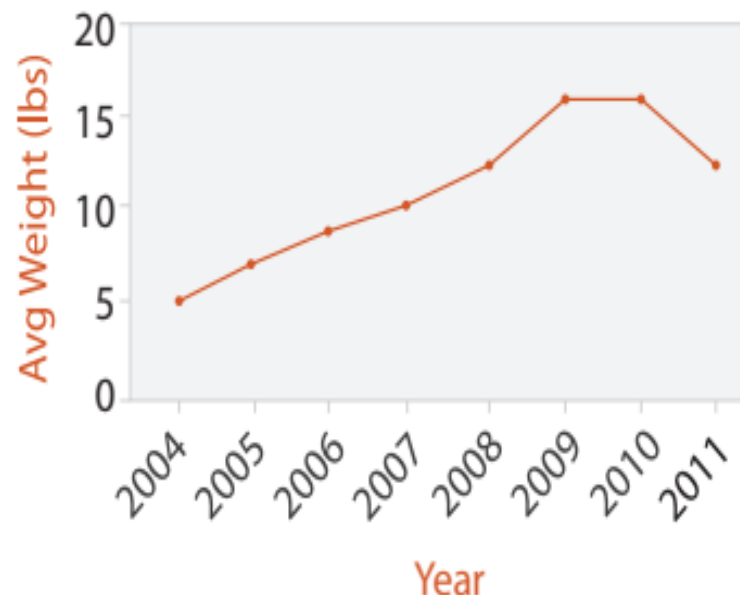 - task
    - find trend
      - connection marks emphasize ordering of items along key axis by explicitly showing relationship between one item and the next
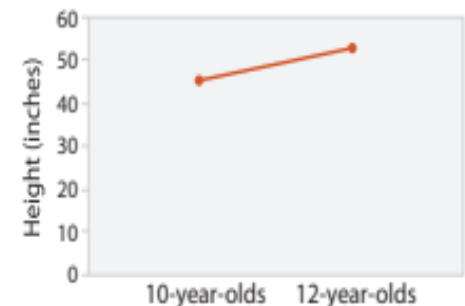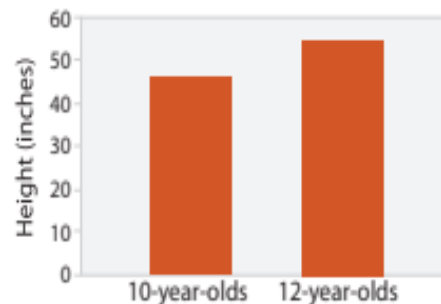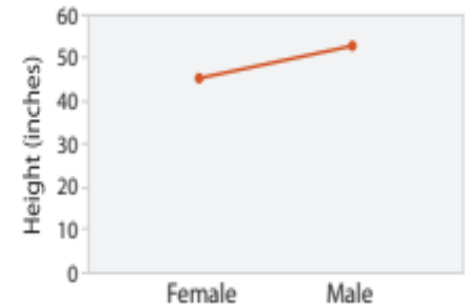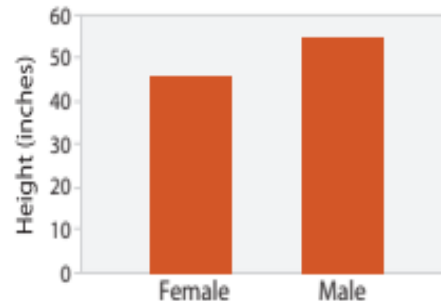 - scalability
    - hundreds of key levels, hundreds of value levels

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.]

Heatmaps

- 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
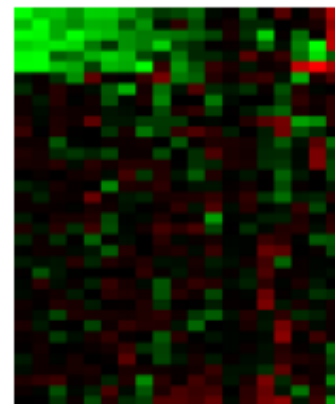 - channels
    - color by quant attrib
      - (ordered diverging colormap)
  - task
    - find clusters, outliers
  - scalability
    - 1M items, 100s of categ levels, ~10 quant attrib levels
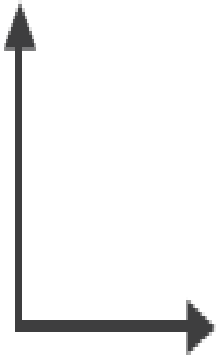

→ *1 Key*
List

→ *2 Keys*
Matrix
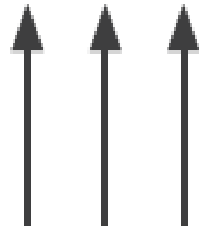
→ *Many Keys*
Recursive Subdivision
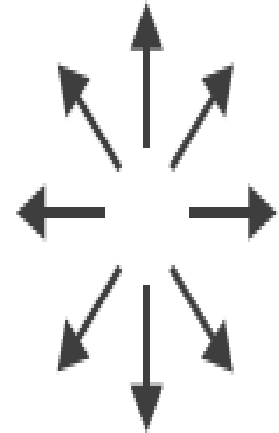
Axis Orientation

→ Rectilinear → Parallel → Radial

Scatterplot matrix, Parallel coordinates

- scatterplot matrix (SPLOM)
  - rectilinear axes, point mark
  - all possible pairs of axes
  - scalability
    - one dozen attribs
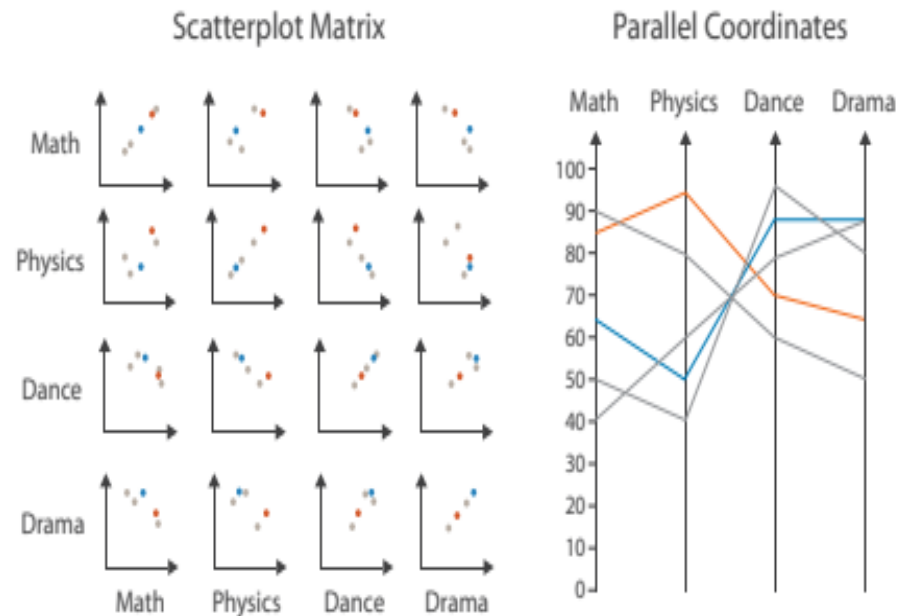   - dozens to hundreds of items
- parallel coordinates
  - parallel axes, jagged line representing item
  - rectilinear axes, item as point
    - axis ordering is major challenge
  - scalability
    - dozens of attribs
   - hundreds of items

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

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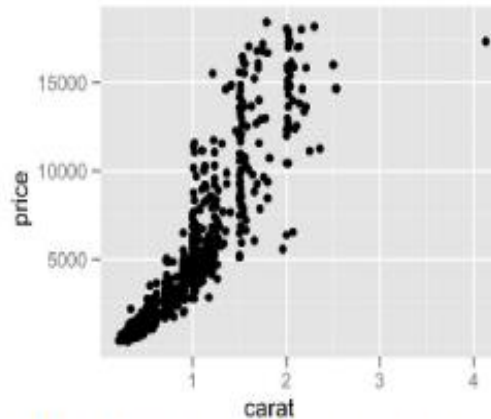
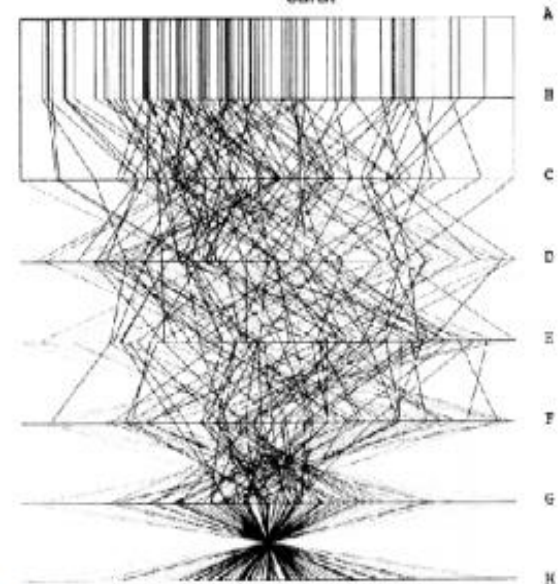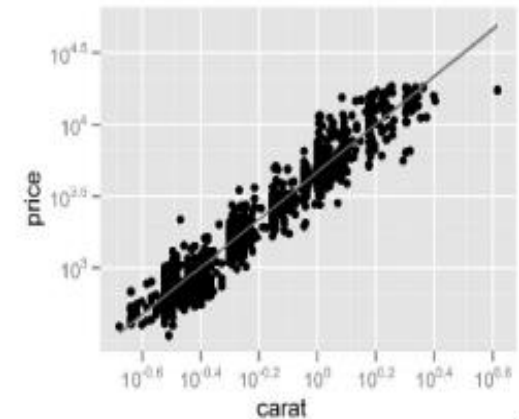[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]

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

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

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