CSE3020 - Data Visualization

Module 4 :Visual Analytics

Dr. K.P. Vijayakumar, VIT Chennai

Topics to be covered

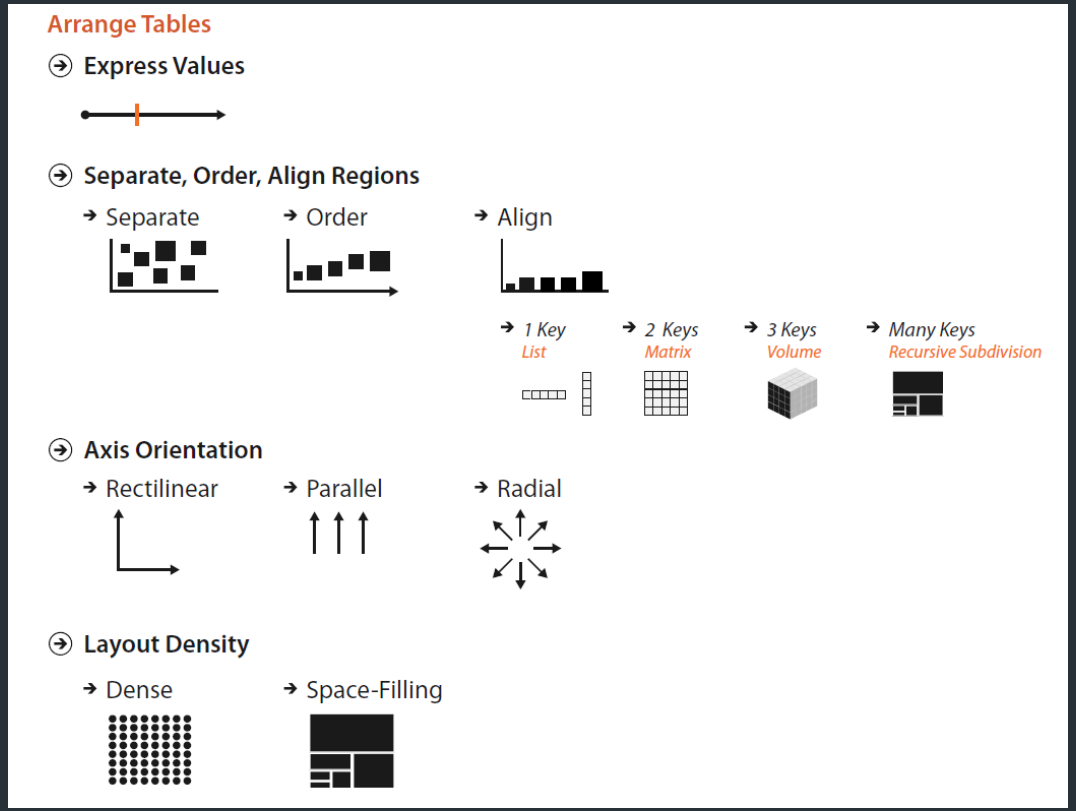
- Arrange Tables
- Arrange Geo Spatial Data
- Reduce Items and Attributes

Arrange Tables

- **Introduction**
  - Four visual encoding design choices for how to arrange the tabular data spatially
    - Express value
    - Separate, order and align regions

Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U

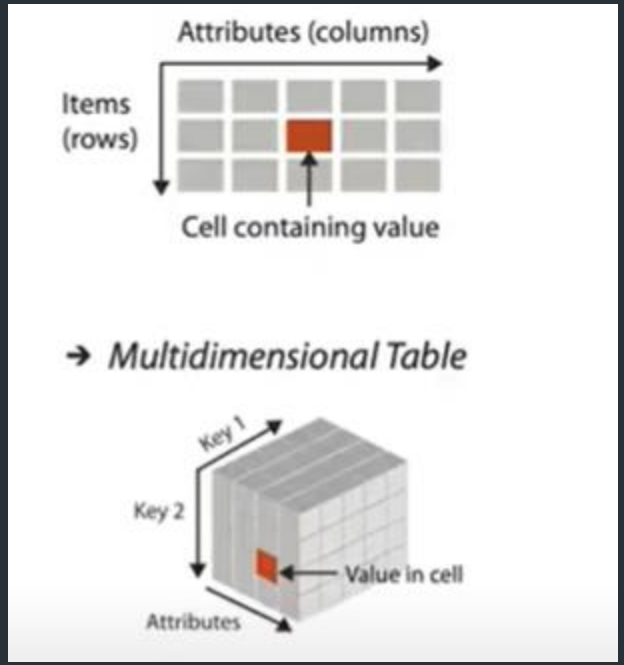
Arrange Tables



Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U

Arrange Tables

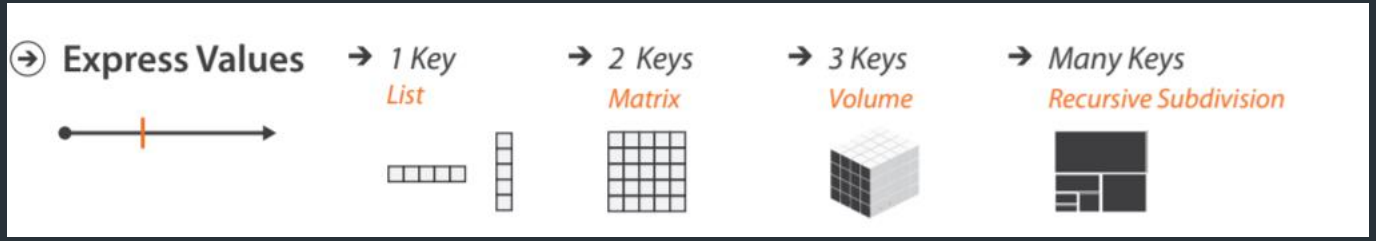
- **Arrange by Keys and Values**
- **Key**
  - Independent variable
  - Used as unique index to look up items
  - One key - simple table
  - Multiple keys- multi dimensional tables
  - Attribute – categorical, ordinal or quantitative
- **Value**
  - Dependent attribute
  - Value of a cell - 1,2,3 etc.



Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U

Arrange Tables

- **Express :Quantitative Value**

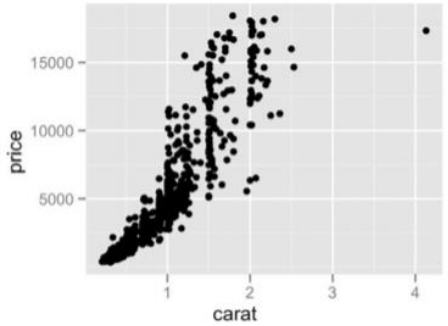


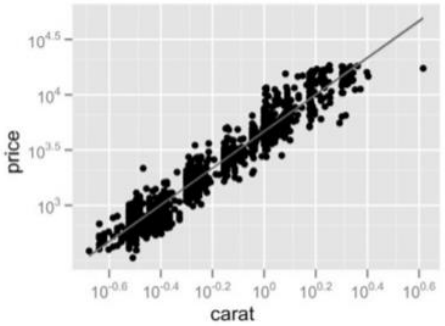
Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

Arrange Tables

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



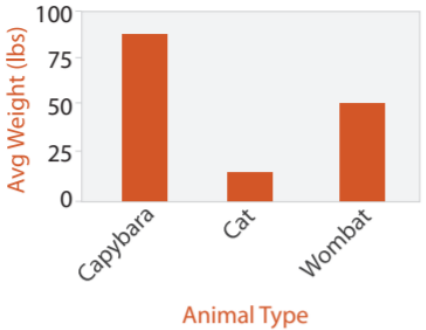


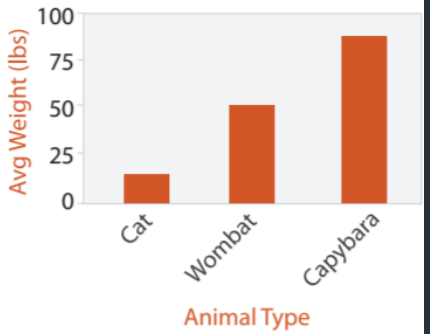
Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

Arrange Tables

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



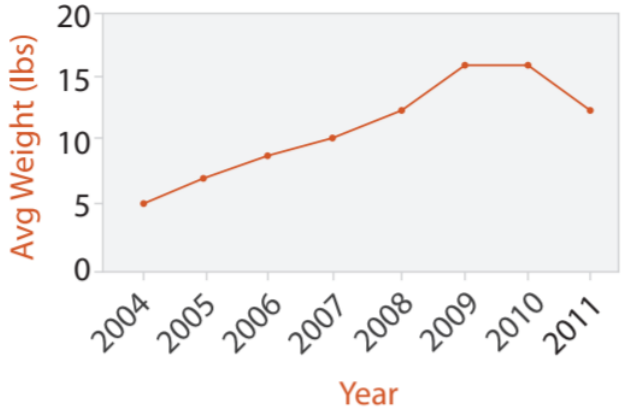


Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

Arrange Tables

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

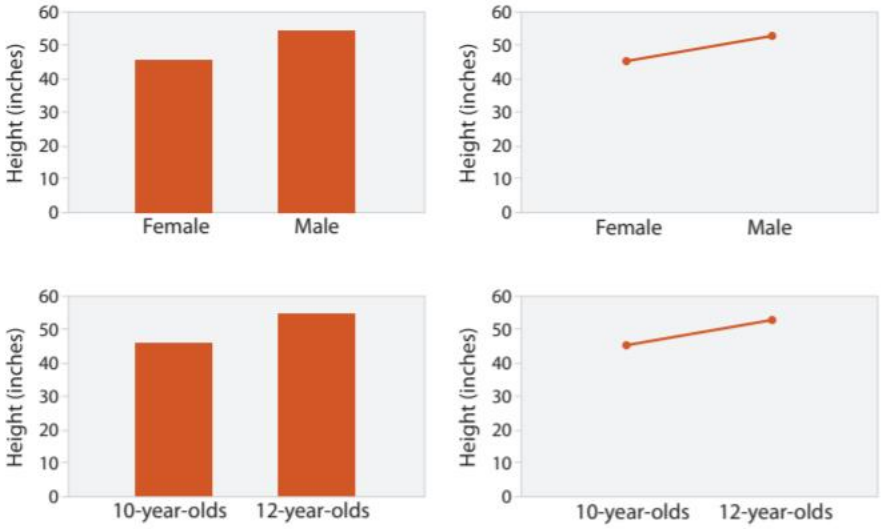


Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

Arrange Tables

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"

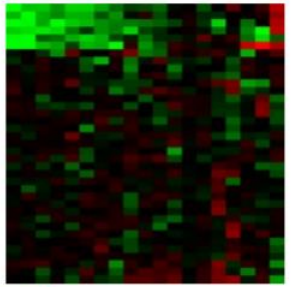


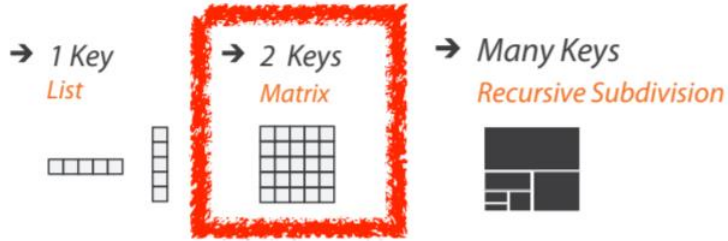
Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

Arrange Tables

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



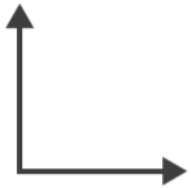


Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

Arrange Tables

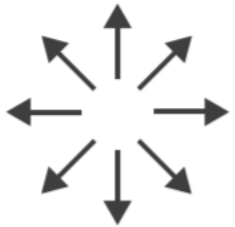
Axis Orientation

- Rectilinear



- Parallel

- Radial

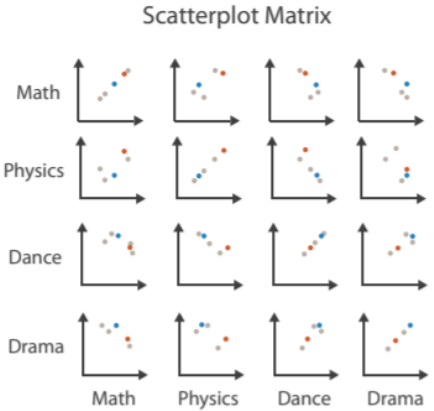


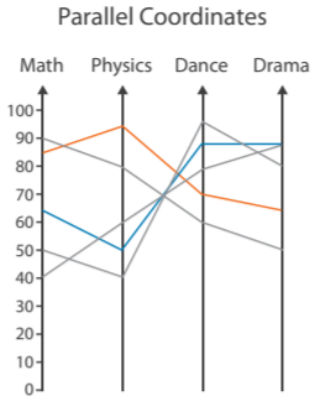
Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

Arrange Tables

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
  - scalability
    - dozens of attribs
    - hundreds of items





Table

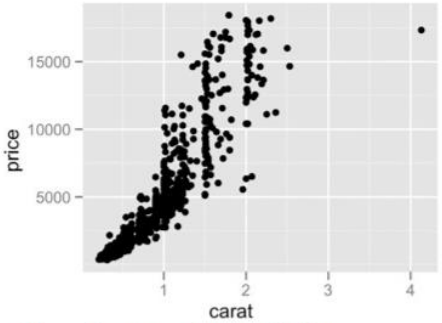
Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

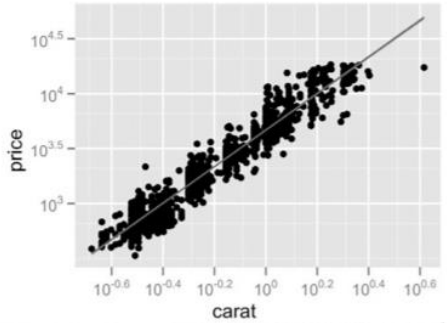
Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

Arrange Tables

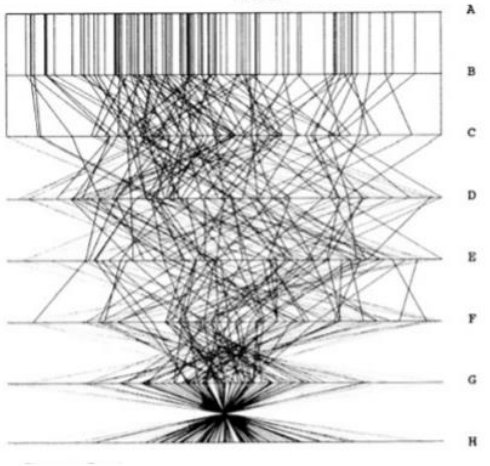
Task: Correlation

- 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.]



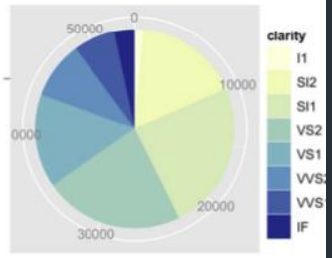
Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

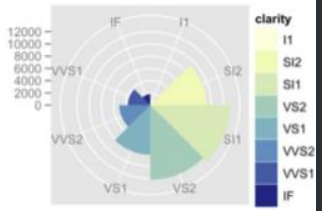
Arrange Tables

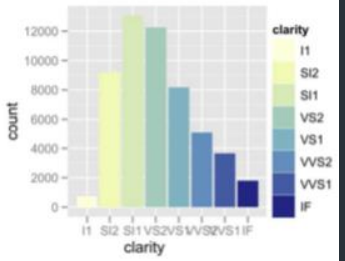
Idioms: **pie chart, polar area chart**

- pie chart
  - area marks with angle channel
  - accuracy: angle/area less accurate than line length
    - arclength also less accurate than line length
- polar area chart
  - area marks with length channel
  - more direct analog to bar charts

- data
  - 1 categ key attrib, 1 quant value attrib
- task
  - part-to-whole judgements





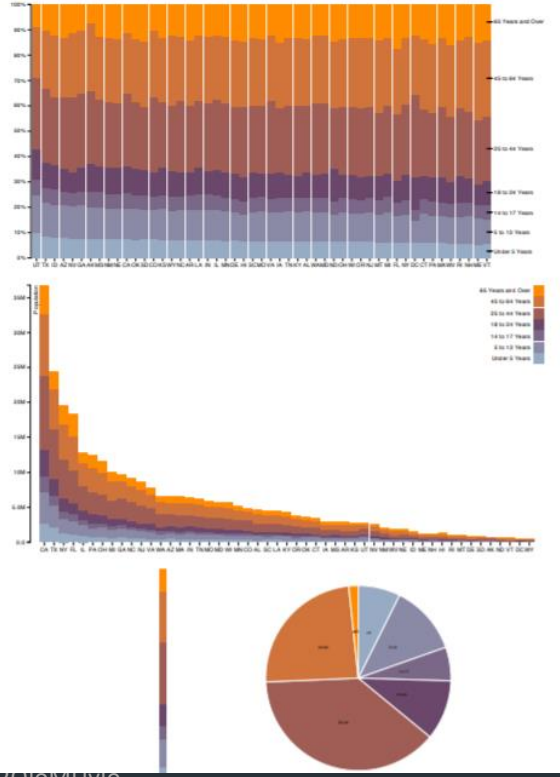


Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

Arrange Tables

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



Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QIoMUyIo

Arrange Tables

Orientation limitations

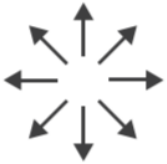
- rectilinear: scalability wrt #axes
  - 2 axes best
  - 3 problematic
    - more in afternoon
  - 4+ impossible
- parallel: unfamiliarity, training time
- radial: perceptual limits
  - angles lower precision than lengths
  - asymmetry between angle and length
    - can be exploited!

Axis Orientation

→ Rectilinear

→ Parallel

→ Radial

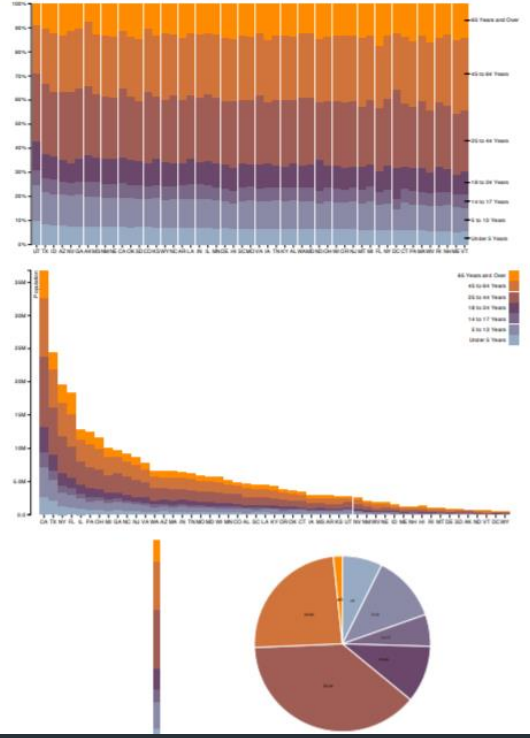


Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

Arrange Tables

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



Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

Arrange Spatial data

- Use Given
  - Geometry
    - *Geographic*
    - *Other Derived*

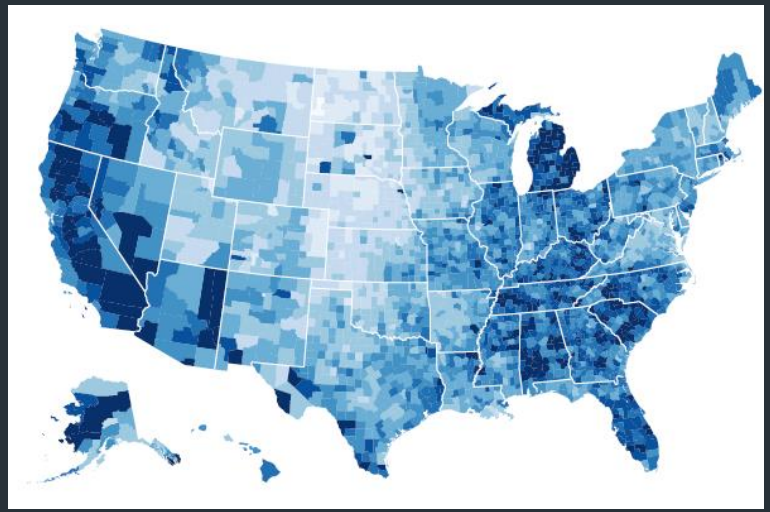
  - Spatial Fields
    - *Scalar Fields (one value per cell)*
      - *Isocontours*
      - *Direct Volume Rendering*

    - *Vector and Tensor Fields (many values per cell)*
      - *Flow Glyphs (local)*
      - *Geometric (sparse seeds)*
      - *Textures (dense seeds)*
      - *Features (globally derived)*

Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

Arrange Spatial data – Geometry

- Idiom: **choropleth map**
- ***use*** given spatial data
  - when central task is understanding spatial relationships
- data
  - geographic geometry data
  - table with 1 quantitative attribute per region
- encoding
  - use given geometry for area mark boundaries
  - Color: sequential segmented colormap



US unemployment rates from 2008 with a segmented sequential colormap

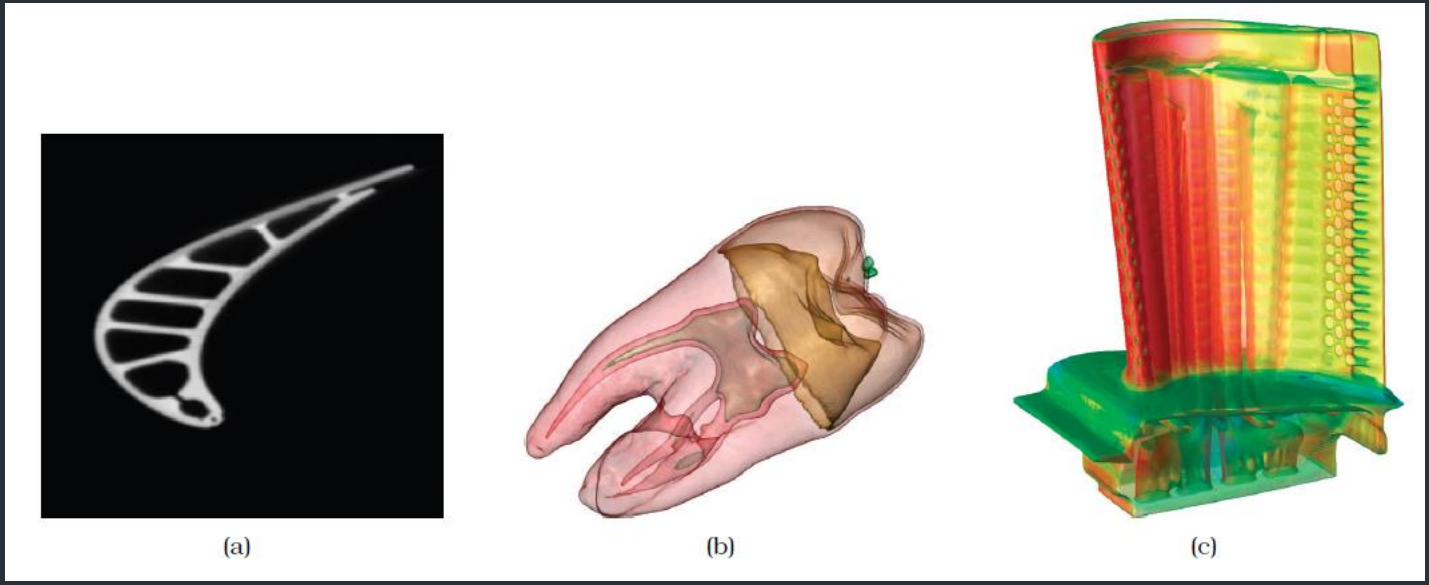
Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

Arrange Spatial data – Scalar Fields

- A scalar spatial field – one value
- Collected through medical imaging
  - computed tomography (CT) scans
  - magnetic resonance imaging (MRI) scans
- Idioms
  - Isocontours
  - Direct volume rendering
  - Slicing

Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

Arrange Spatial data – Scalar Fields



(a) A single 2D slice of a turbine blade dataset.
(b) Multiple semitransparent isosurfaces of a 3D tooth dataset.
(c) Direct volume rendering of the entire 3D turbine dataset

Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

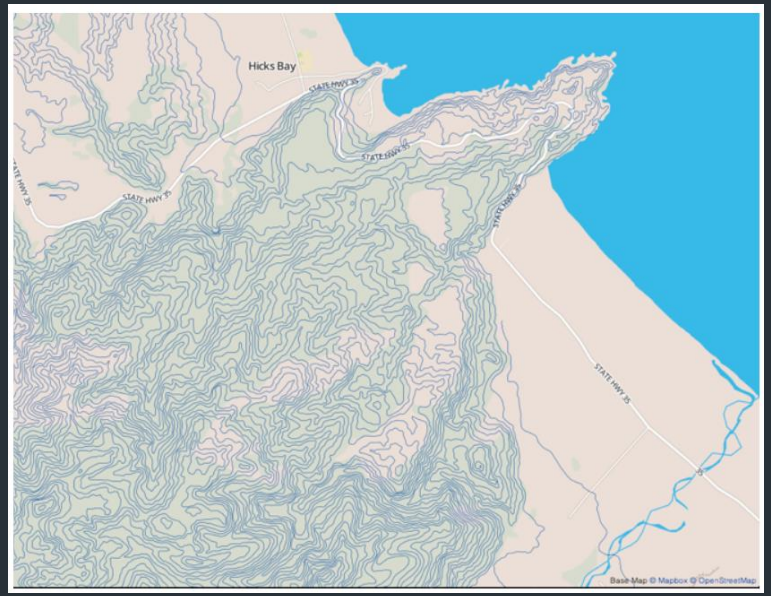
Arrange Spatial data – Scalar Fields

- Isolines
  - lines that represent the contours of a particular level of the scalar value, can be derived from a scalar spatial field.
  - Isolines occur
    - far apart in regions of slow change
    - close together in regions of fast change
    - Never occur overlaps

Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

Arrange Spatial data – Scalar Fields

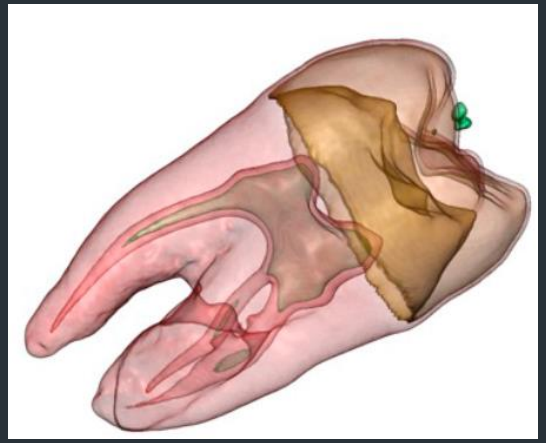
- Idiom: **topographic map**
- data
  - geographic geometry
  - scalar spatial field
    - 1 quant attribute per grid cell
- derived data
  - isoline geometry
    - isocontours computed for specific levels of scalar values



Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

Arrange Spatial data – Scalar Fields

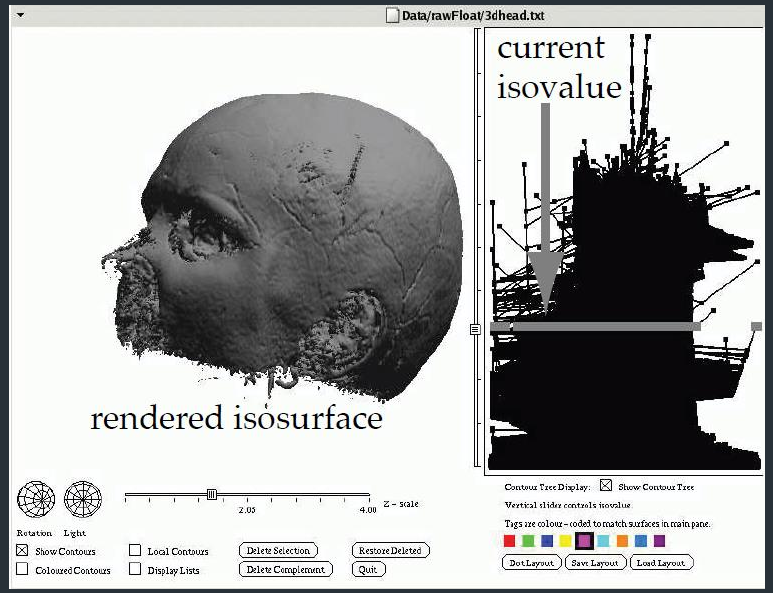
- Idioms: **isosurfaces**
- data
  - scalar spatial field
  - 1 quant attribute per grid cell
- task
  - shape understanding, spatial relationships
- isosurface
  - derived data: isocontours computed for specific levels of scalar values



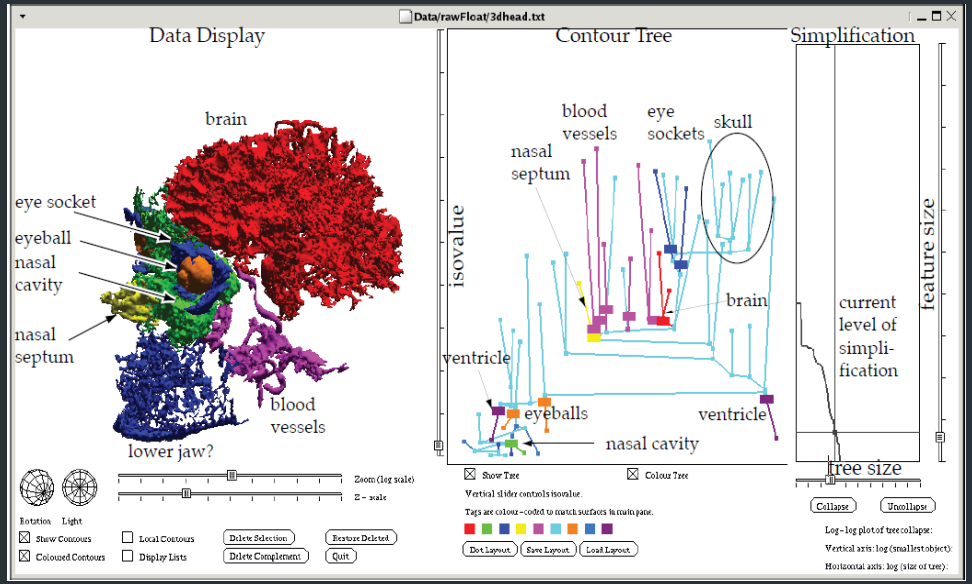
Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

Arrange Spatial data – Scalar Fields

- Idioms: **Flexible isosurfaces**
  - Uses one more level of user data



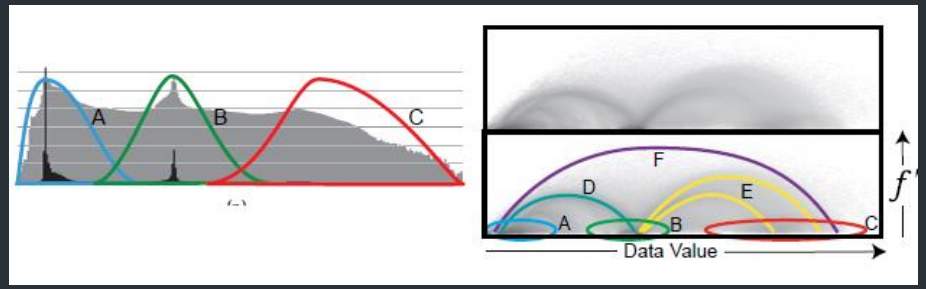
Full control Contour tree

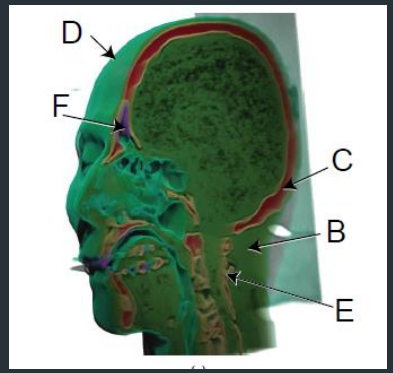


Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

Arrange Spatial data – Scalar Fields

- Idioms: **direct volume rendering**
- data
  - 3D spatial field
  - 1 derived quant value attribute



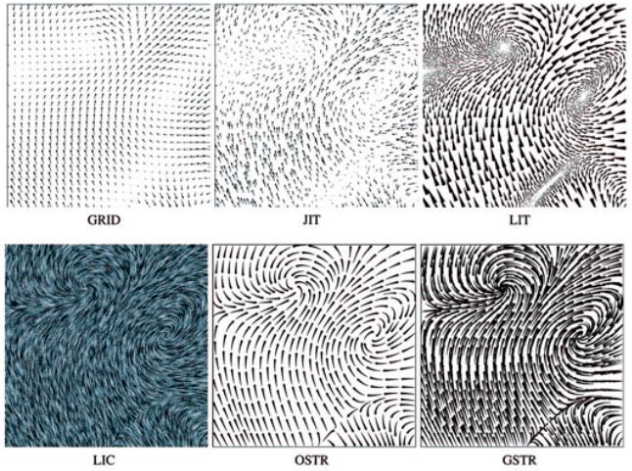


(A) air, (B) soft tissue, and (C) bone, (D) air–tissue, (E) tissue–bone, and (F) air–bone

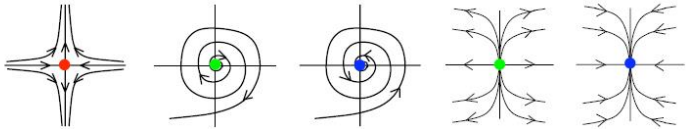
Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

Arrange Spatial data – Vector Fields

- data
  - many attribs per cell
- idiom families
  - flow glyphs
    - purely local
  - geometric flow
    - derived data from tracing particle trajectories
    - sparse set of seed points
  - texture flow
    - derived data, dense seeds
  - feature flow
    - global computation to detect features
      - encoded with one of methods above



[Comparing 2D vector field visualization methods: A user study. Laidlaw et al. IEEE Trans. Visualization and Computer Graphics (TVCG) 11:1 (2005), 59–70.]



[Topology tracking for the visualization of time-dependent two-dimensional flows. Tricoche, Wischgoll, Scheuermann, and Hagen. Computers & Graphics 26:2 (2002), 249–257.]

Source: Wikipedia, https://www.youtube.com/watch?v=ey8-X6sq4-U, https://www.youtube.com/watch?v=dg2QloMUyIo

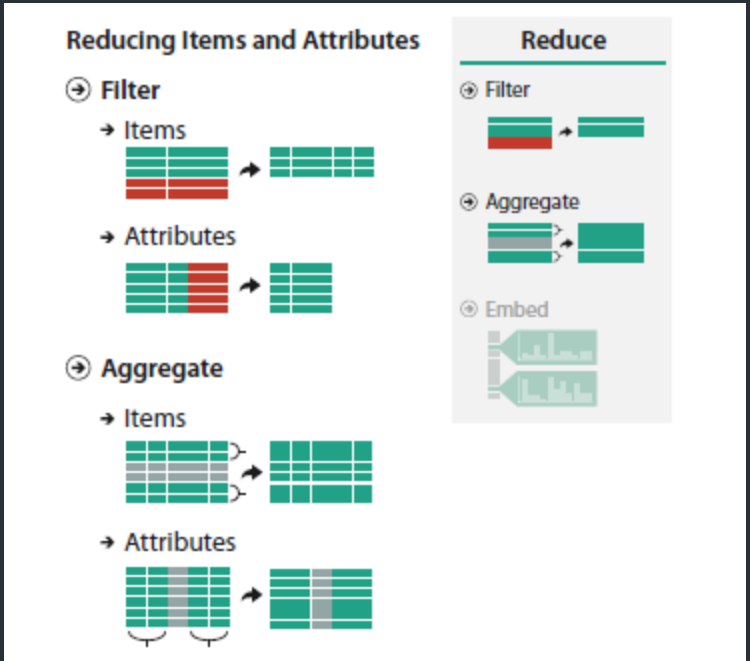
Arrange Spatial data – Vector Fields

- (a) arrow glyphs on a regular grid
- (b) arrow glyphs on a jittered grid.
- (c) triangular wedge glyphs inspired by oil painting strokes
- (d) dense texturebased Line Integral Convolution (LIC)
- (e) curved arrow glyphs with image-guided streamline seeding
- (f) curved arrow glyphs with regular grid streamline seeding

Source: Visualisation Analysis and Design, Tamara Munzner

Reduce Items and Attributes

- Design choices for reducing the amount of data items and attributes



Source: Visualisation Analysis and Design, Tamara Munzner

Reduce Items and Attributes

- Reduce/Increase: inverses
- Filter
  - straightforward and intuitive
    - to understand and compute
    - Select one or more ranges – min, max
  - out of sight, out of mind
    - It is difficult from user's viewpoint
  - Interactive Vis context
    - Dynamic queries
    - Controls
      - GUI user interface widgets such as sliders, buttons, combo boxes, and text fields
  - It is difficult from user's viewpoint

Source: Visualisation Analysis and Design, Tamara Munzner

Reduce Items and Attributes

- Types of Filter
  - Item Filtering
  - Attribute Filtering

Source: Visualisation Analysis and Design, Tamara Munzner

Reduce Items and Attributes

- Item Filtering
  - goal is to eliminate items
  - Fewer items, no. of attributes does not change
  - Example : - FilmFinder
    - Attributes – 9 Attributes
    - genre, year made, title, actors, actresses, directors, rating, popularity and length
    - Scatterplot
    - year made versus movie popularity

Source: Visualisation Analysis and Design, Tamara Munzner

Reduce Items and Attributes

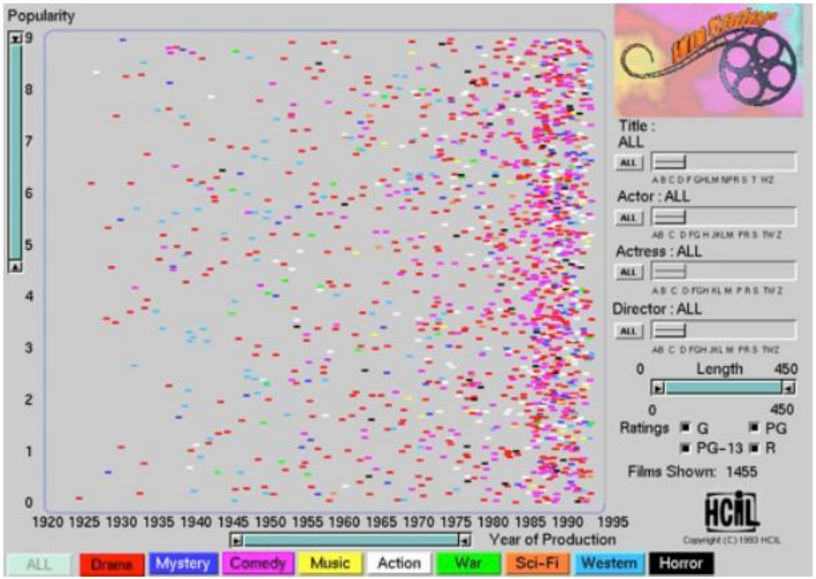
- Idiom : Dynamic Filtering

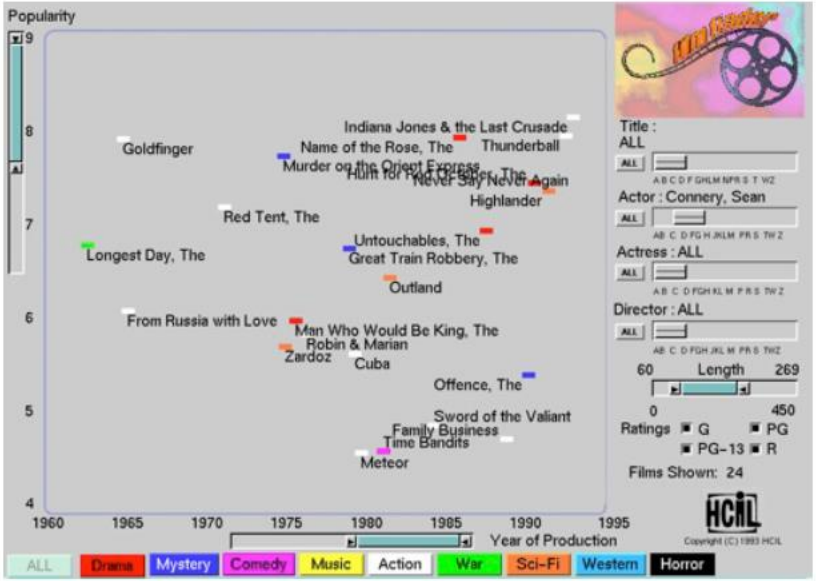
System	FilmFinder
What: Data	Table: nine value attributes.
How: Encode	Scatterplot; detail view with text/images.
How: Facet	Multiform, overview–detail.
How: Reduce	Item filtering.

Source: Visualisation Analysis and Design, Tamara Munzner

Reduce Items and Attributes

- Item Filtering - FilmFinder



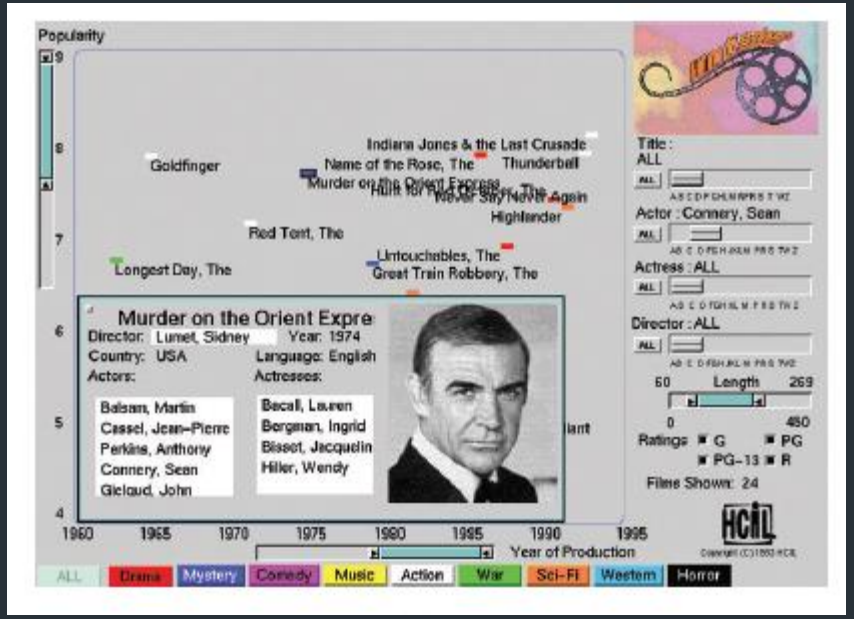


(a) Exploration begins with an overview of all movies in the dataset.
(b) Moving the actor slider to select Sean Connery filters out most of the other movies, leaving enough room to draw labels

Source: Visualisation Analysis and Design, Tamara Munzner

Reduce Items and Attributes

- Item Filtering - FilmFinder



(c) Clicking on the mark representing a movie brings up a detail view

Source: Visualisation Analysis and Design, Tamara Munzner

Reduce Items and Attributes

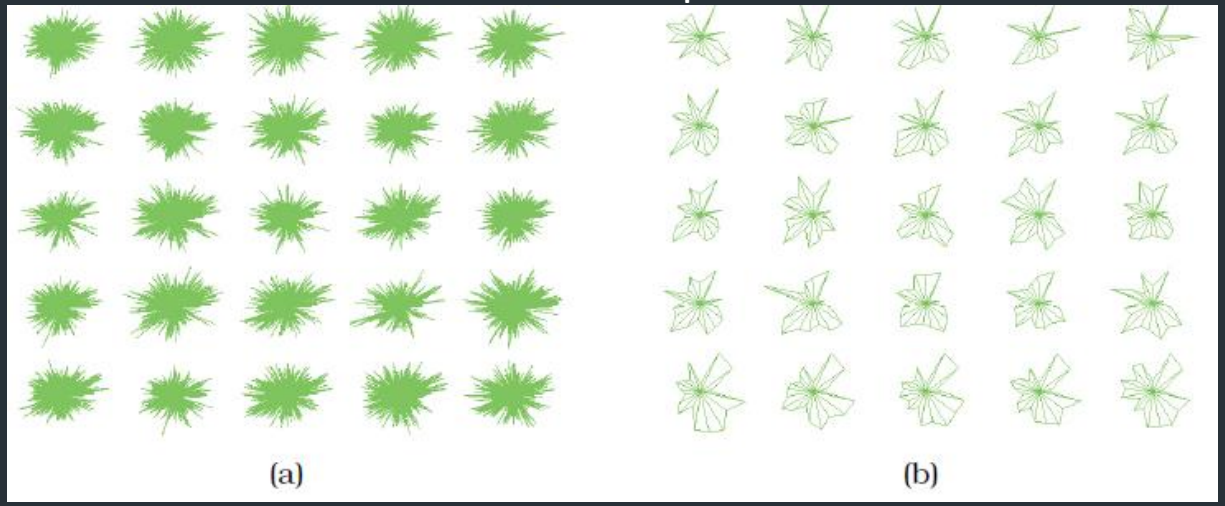
- Attribute Filtering
  - To eliminate attributes
  - **Example: DOSFA (**Dimensional Ordering, Spacing, and Filtering Approach)
  - Encoding ： Star Plot

System	DOSFA
What: Data	Table: many value attributes.
How: Encode	Star plots.
How: Facet	Small multiples with matrix alignment.
How: Reduce	Attribute filtering.

Source: Visualisation Analysis and Design, Tamara Munzner

Reduce Items and Attributes

- Attribute Filtering
  - Star Glyphs
  - Medical records – 215 dimensions and 298 points



(a) The full dataset is so dense that patterns cannot be seen.
(b) After ordering on similarity and filtering on both similarity and importance, the star glyphs show structure

Source: Visualisation Analysis and Design, Tamara Munzner

Reduce Items and Attributes

- Aggregation
  - Elements are merged, as opposed to eliminated completely with filtering
  - It involves the use of a derived attributes
    - Example : computing an average
      - Aggregation operators
        - Minimum, maximum, count, and sum
    - combine filter, aggregate

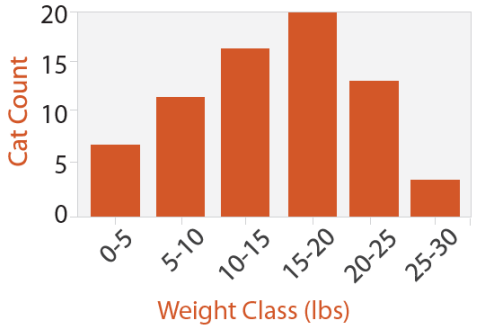
Source: Visualisation Analysis and Design, Tamara Munzner

Reduce Items and Attributes

- Aggregation - Idiom: Histogram

Idiom: **histogram**

- static item aggregation
- task: find distribution
- data: table
- derived data
  - new table: keys are bins, values are counts
- bin size crucial
  - pattern can change dramatically depending on discretization
  - opportunity for interaction: control bin size on the fly



Source: Visualisation Analysis and Design, Tamara Munzner

Reduce Items and Attributes

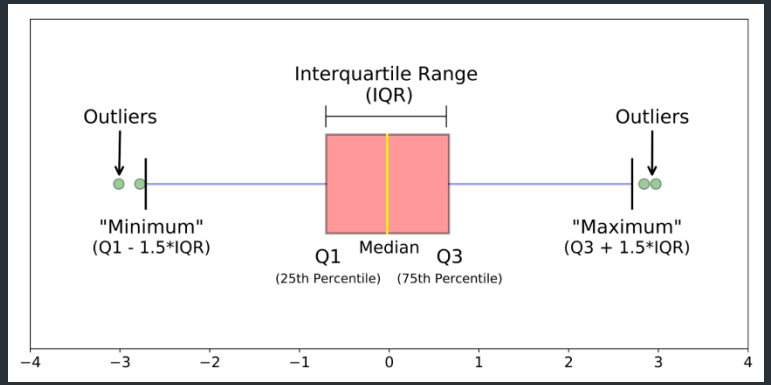
- Aggregation - Idiom: Histogram

Idiom	Histograms
What: Data	Table: one quantitative value attribute.
What: Derived	Derived table: one derived ordered key attribute (bin), one derived quantitative value attribute (item count per bin).
How: Encode	Rectilinear layout. Line mark with aligned position to express derived value attribute. Position: key attribute.

Source: Visualisation Analysis and Design, Tamara Munzner

Reduce Items and Attributes

- Aggregation - Idiom: Boxplot
- shows an aggregate statistical summary of all the values that occur within the distribution of a single quantitative attribute.
- uses five derived variables
  - the median (50% point),
  - The lower and upper quartiles
    - (25% and 75% points)
  - the upper and lower fences (chosen values near the extremes, beyond which points should be counted as outlier)



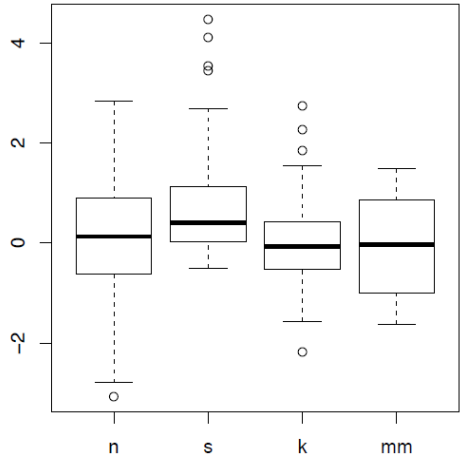
Source: Visualisation Analysis and Design, Tamara Munzner

Reduce Items and Attributes

- Aggregation

Idiom: **boxplot**

- static item aggregation
- task: find distribution
- data: table
- derived data
  - 5 quant attribs
    - median: central line
    - lower and upper quartile: boxes
    - lower upper fences: whiskers
      - values beyond which items are outliers
  - outliers beyond fence cutoffs explicitly shown



Source: Visualisation Analysis and Design, Tamara Munzner

Reduce Items and Attributes

- Aggregation - Idiom: Boxplot

Idiom	Boxplot Charts
What: Data	Table: many quantitative value attributes.
What: Derived	Five quantitative attributes for each original attribute, representing its distribution.
Why: Tasks	Characterize distribution; find outliers, extremes, averages; identify skew.
How: Encode	One glyph per original attribute expressing derived attribute values using vertical spatial position, with 1D list alignment of glyphs into separated with horizontal spatial position.
How: Reduce	Item aggregation.
Scale	Items: unlimited. Attributes: dozens.

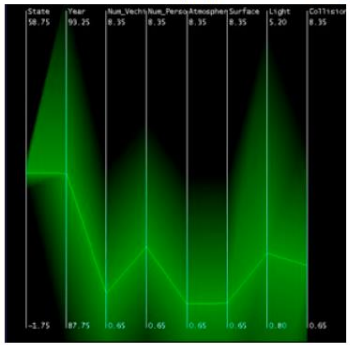
Source: Visualisation Analysis and Design, Tamara Munzner

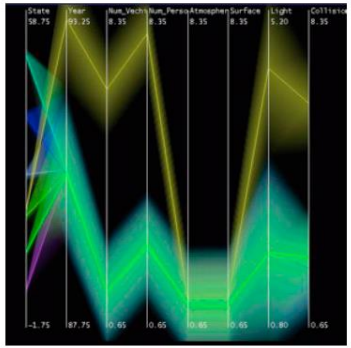
Reduce Items and Attributes

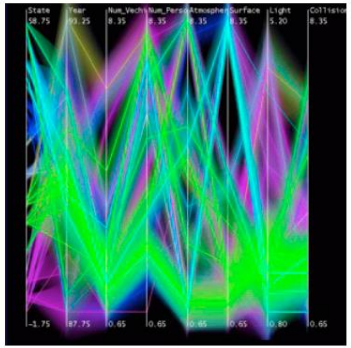
- Aggregation

Idiom: **Hierarchical parallel coordinates**

- dynamic item aggregation
- derived data: ***hierarchical clustering***
- encoding:
  - cluster band with variable transparency, line at mean, width by min/max values
  - color by proximity in hierarchy







Source: Visualisation Analysis and Design, Tamara Munzner

Reduce Items and Attributes

- Aggregation - Idiom : Hierarchical Parallel Coordinates

Idiom	Hierarchical Parallel Coordinates
What: Data	Table.
What: Derived	Cluster hierarchy atop original table of items. Five per-cluster attributes: count, mean, min, max, depth.
How: Encode	Parallel coordinates. Color clusters by proximity in hierarchy.
How: Reduce	Interactive item aggregation to change level of detail.
Scale	Items: 10,000–100,000. Clusters: one dozen.

Source: Visualisation Analysis and Design, Tamara Munzner

Reduce Items and Attributes

- Attribute Aggregation
  - a new attribute is synthesized to take the place of multiple original attributes
  - Simple approach
    - grouping them by similarity measure
  - Complex approach
    - Dimensionality Reduction

Source: Visualisation Analysis and Design, Tamara Munzner

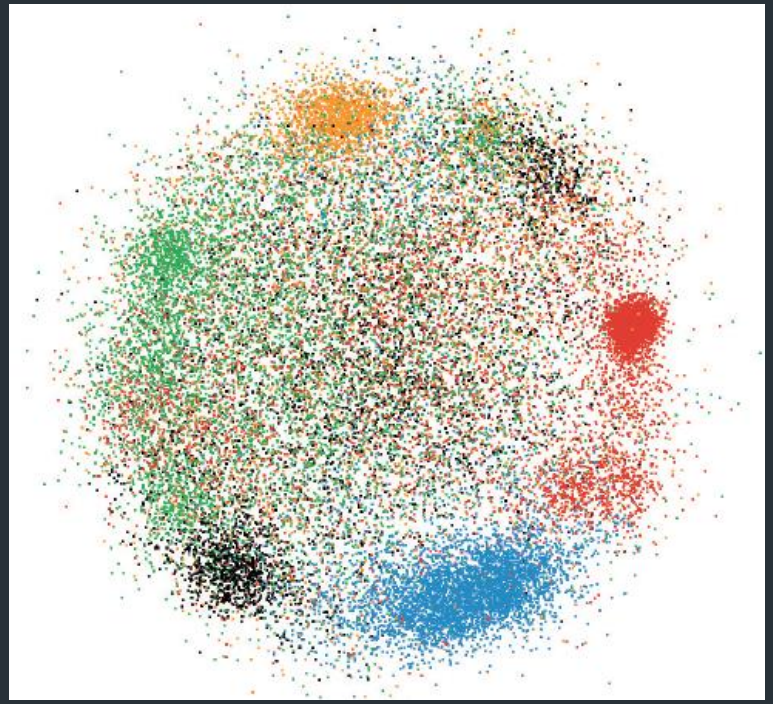
Reduce Items and Attributes

- Attribute Aggregation
  - Dimensionality Reduction for document collection
  - Document collections are not directly visualizeable, but they can be transformed into a dataset type that is: a derived high-dimensional table

Source: Visualisation Analysis and Design, Tamara Munzner

Reduce Items and Attributes

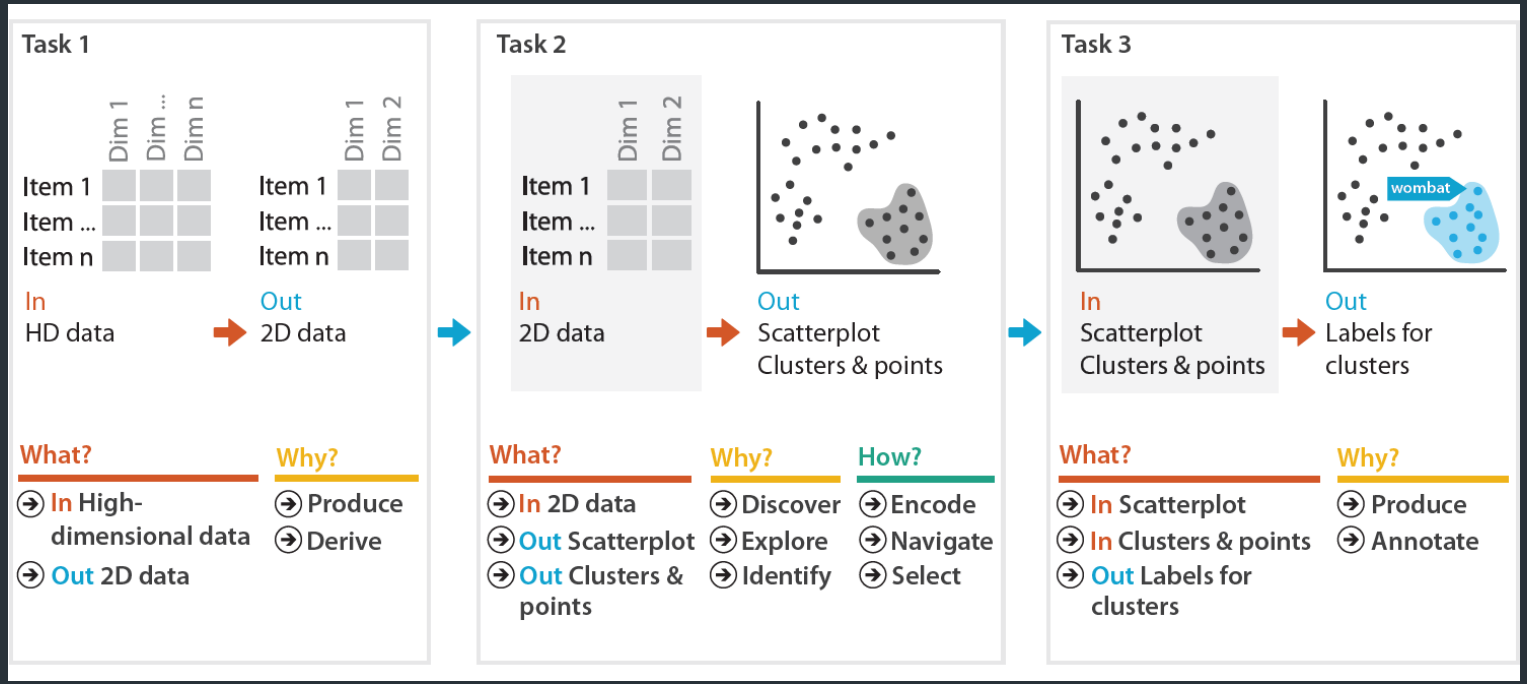
- Dimensionality Reduction for document collection



Source: Visualisation Analysis and Design, Tamara Munzner

Reduce Items and Attributes

- Idiom : Dimensionality Reduction for document collection



Source: Visualisation Analysis and Design, Tamara Munzner

Reduce Items and Attributes

- Idiom : Dimensionality Reduction for document collection

Idiom	Dimensionality Reduction for Document Collections
What: Data	Text document collection.
What: Derived	Table with 10,000 attributes.
What: Derived	Table with two attributes.
How: Encode	Scatterplot, colored by conjectured clustering.
How: Reduce	Attribute aggregation (dimensionality reduction) with MDS.
Scale	Original attributes: 10,000. Derived attributes: two. Items: 100,000.

Source: Visualisation Analysis and Design, Tamara Munzner