Visualization, DD2257

Prof. Dr. Tino Weinkauf

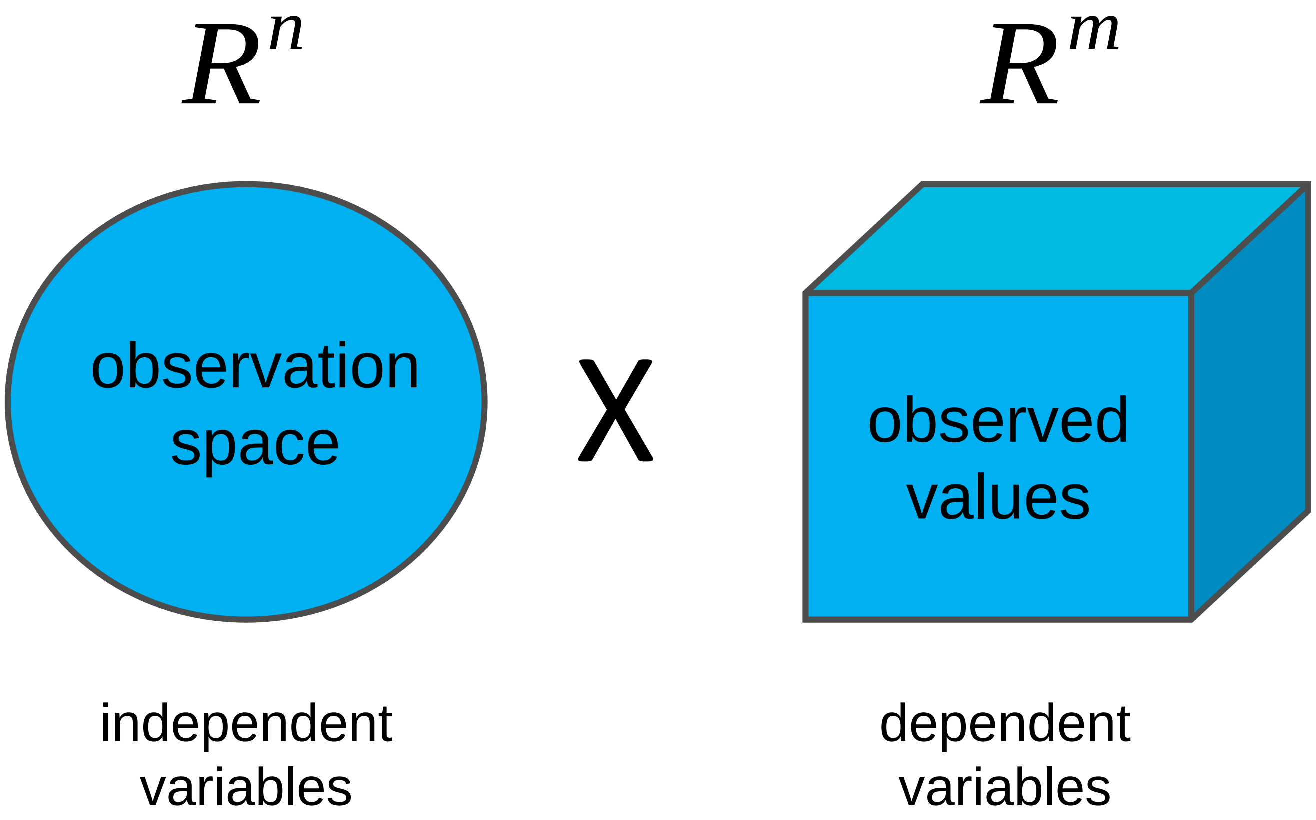
Visualization of Multiparameter Data

No specific prior on observation space

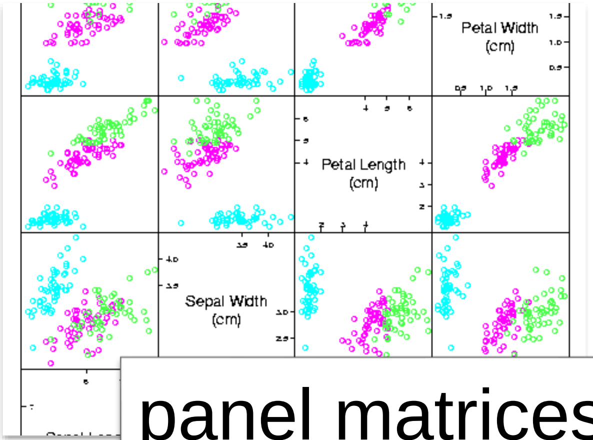
No prior on observation space

students, physical space, …

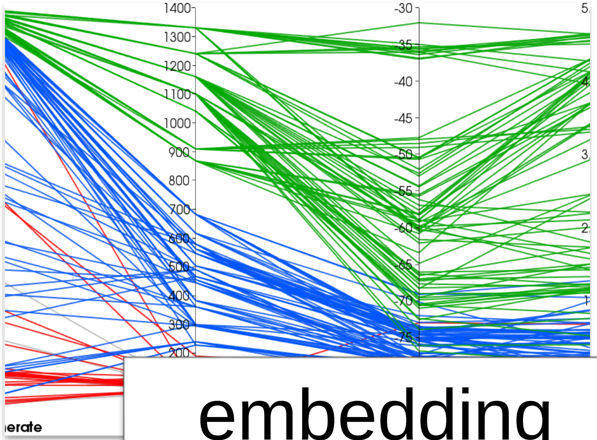
Still explore relationship between observation space and dependent variables.



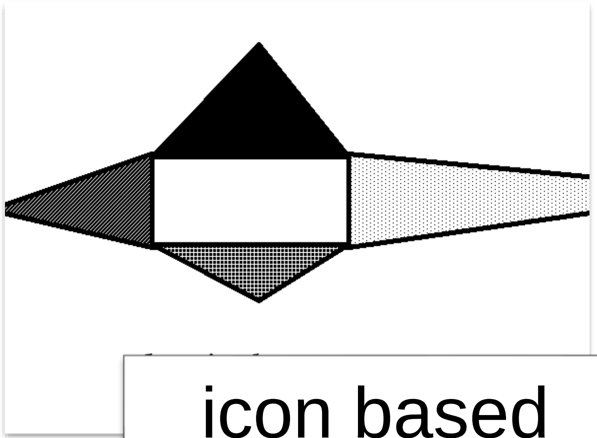
Classification of visualization methods for multiparameter data



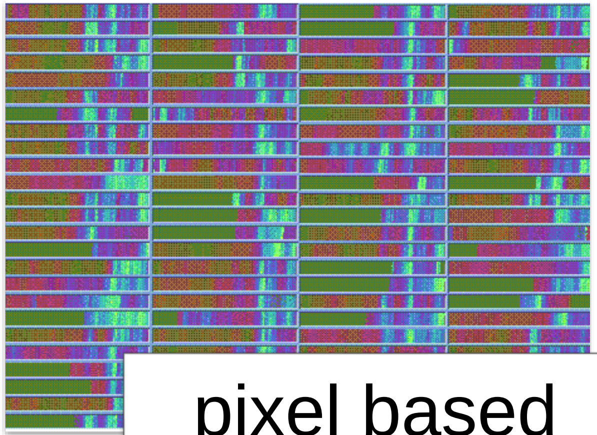
panel matrices



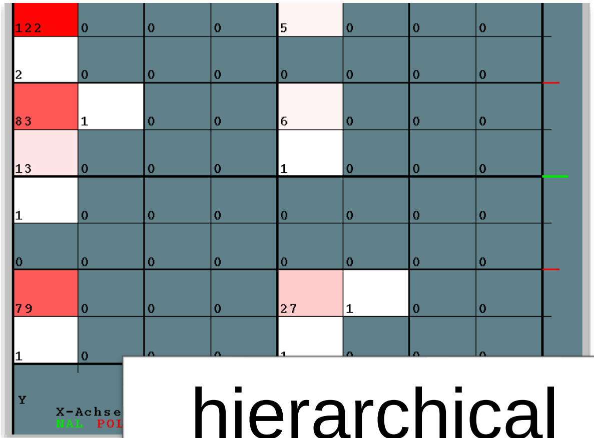
embedding



icon based



pixel based



hierarchical

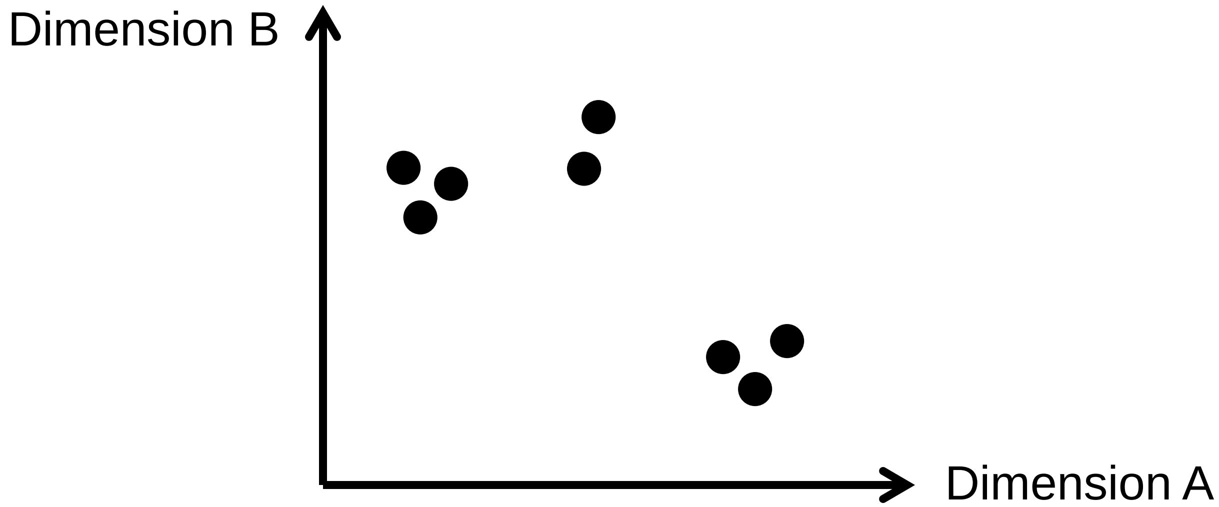
Panel Matrices

Matrix-like configuration of bivariate representations

- scatterplot matrices (Cleveland 93)
- prosection views (Furnas, Buja, 1994)
- hyper slice (van Wijk, J. J., van Liere R. D., 1993)

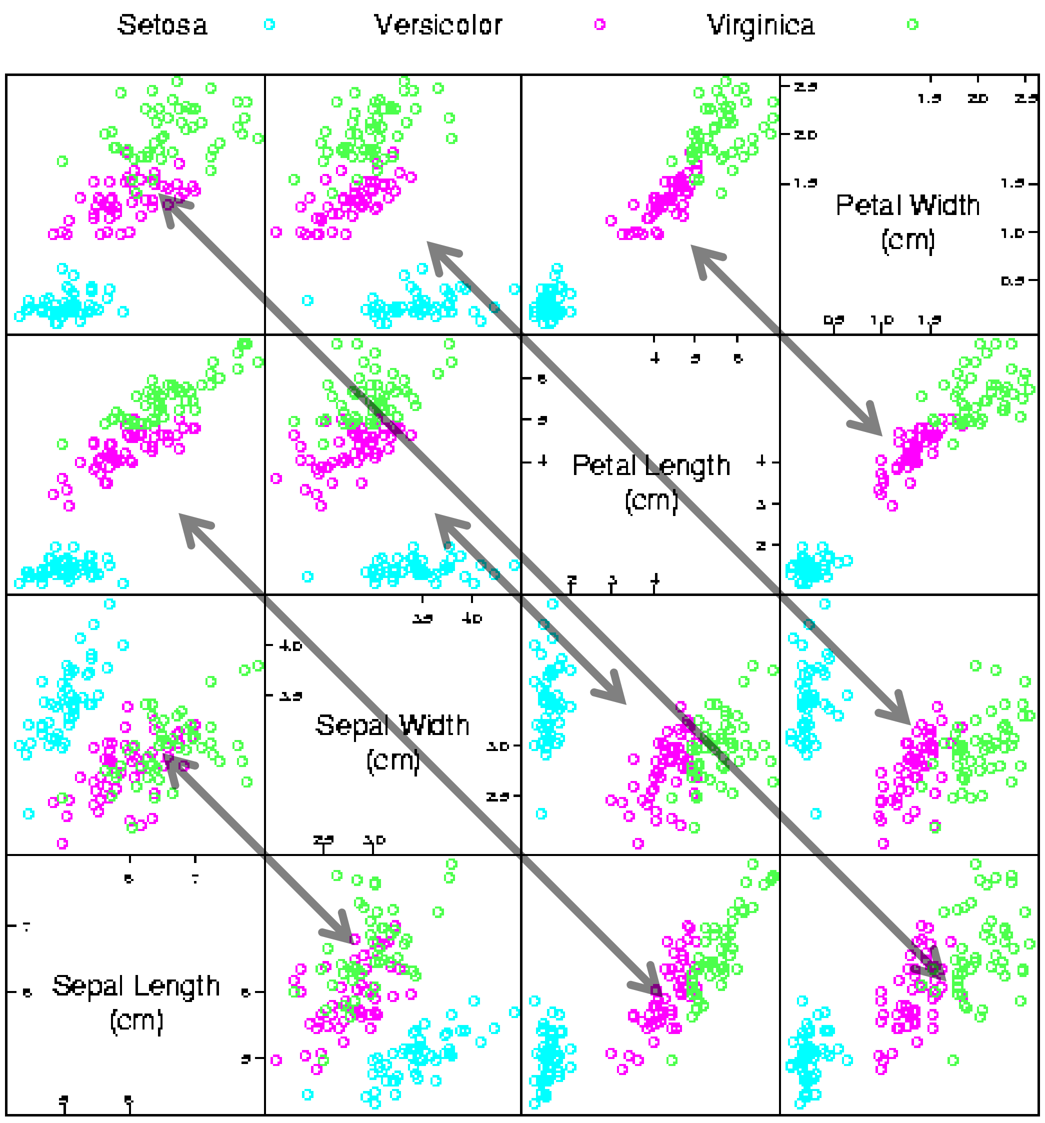
Scatterplot

Cylinders	Horsepower (PS)	Weight (kg)	Origin
4	110	1576	Germany
6	105	3535	Japan
8	150	4190	U.S.A.



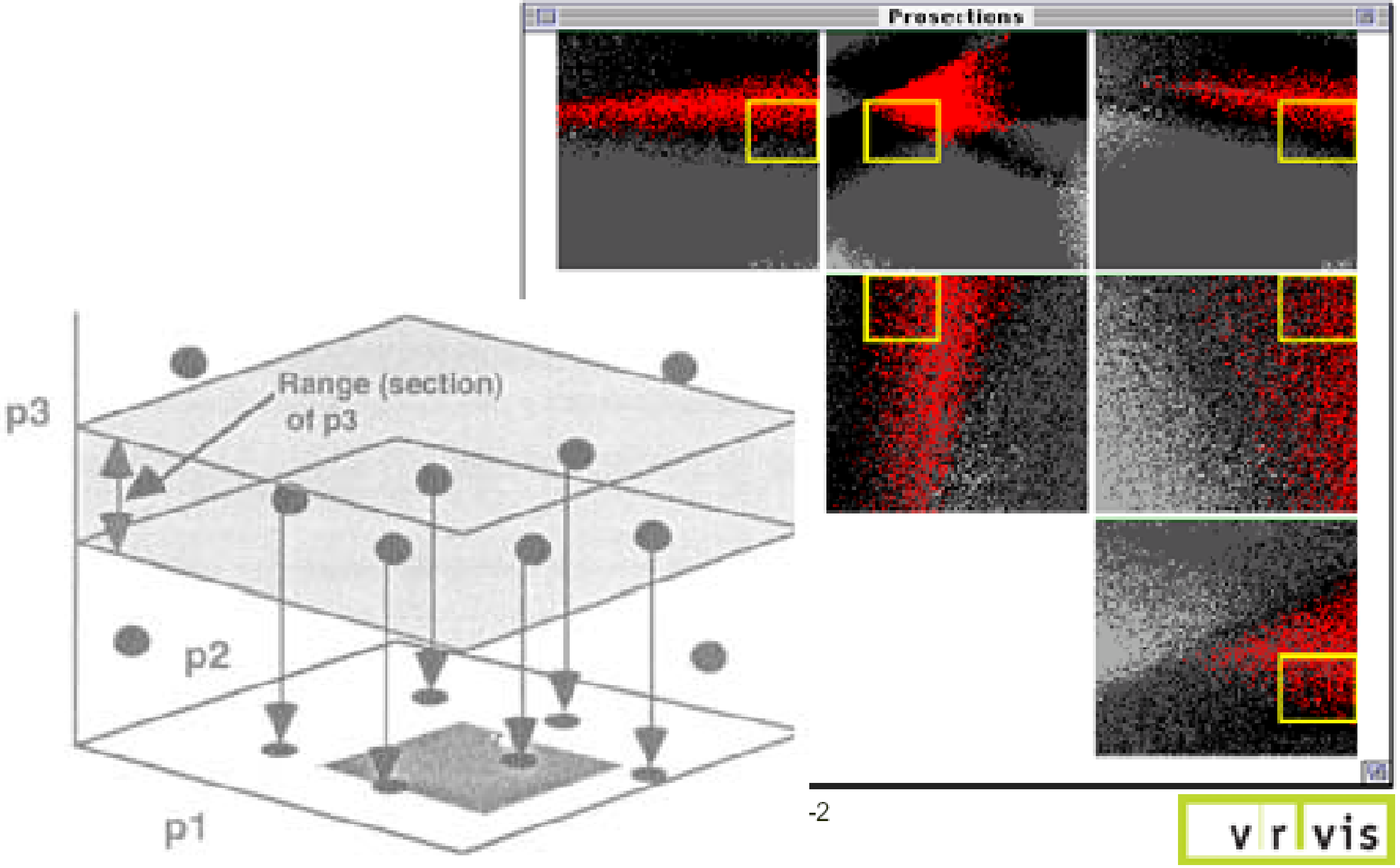
Scatterplot matrix, Cleveland 93

configuration of scatterplots in a matrix for showing different combinations of variables



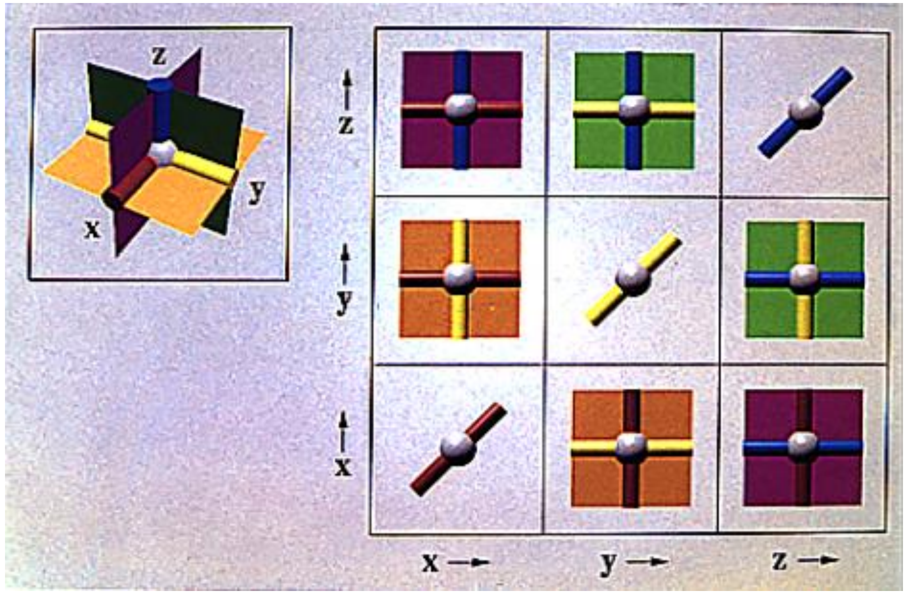
Prosection Views

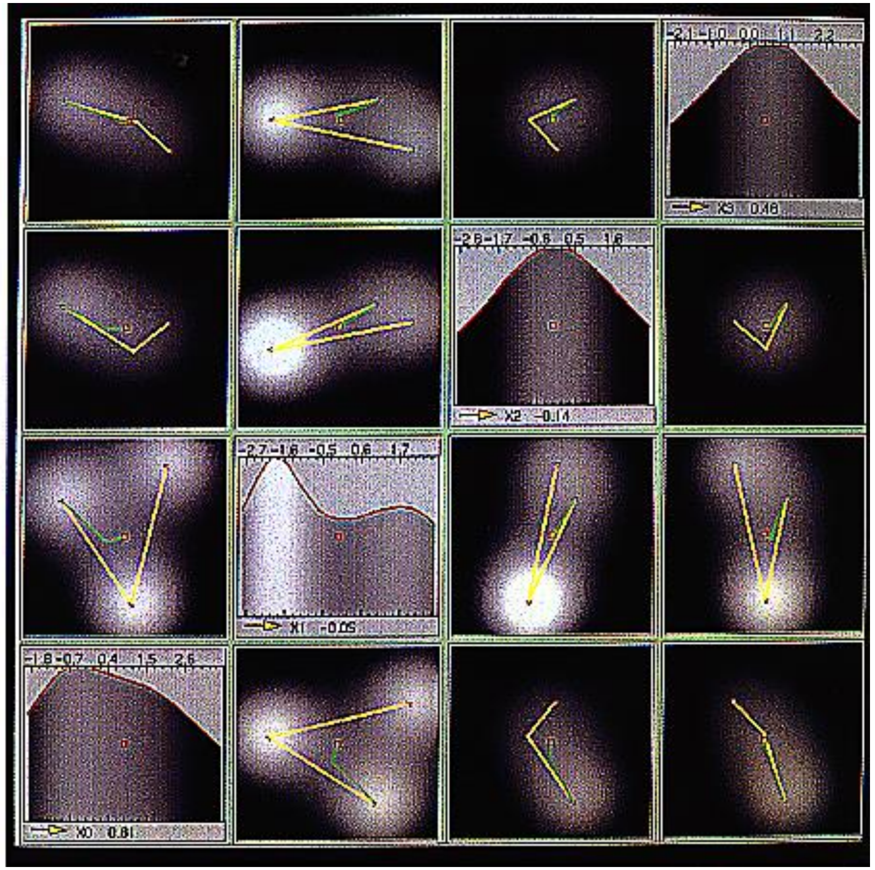
- matrix of all orthogonal projections for specified regions of the observation space (prosection = combination of selections and projections)



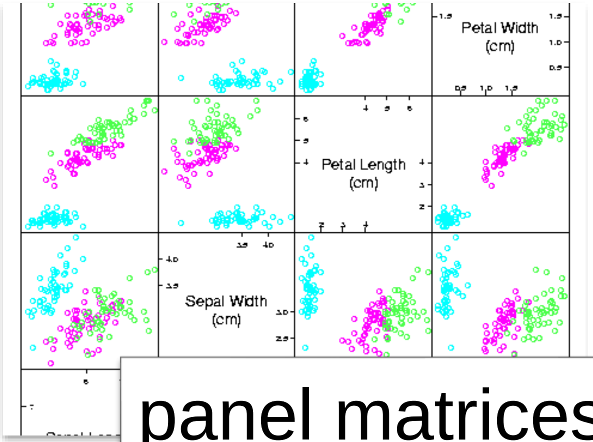
Hyper slice (van Wijk, J. J., van Liere R. D., 1993)

- matrix of n^2 slices through the n-dimensional observation space which intersect in a point of interest (current point). (navigation in observation space possible by moving the current point.)

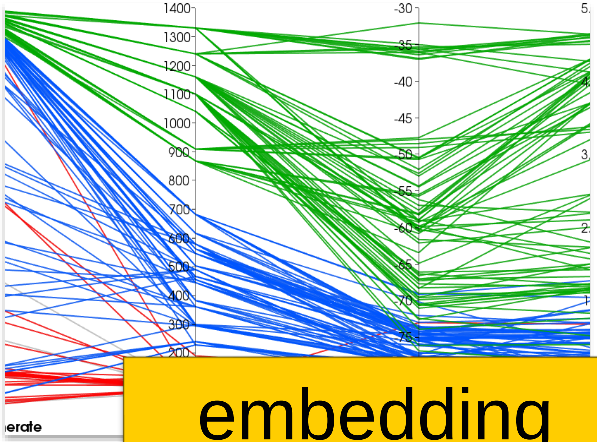




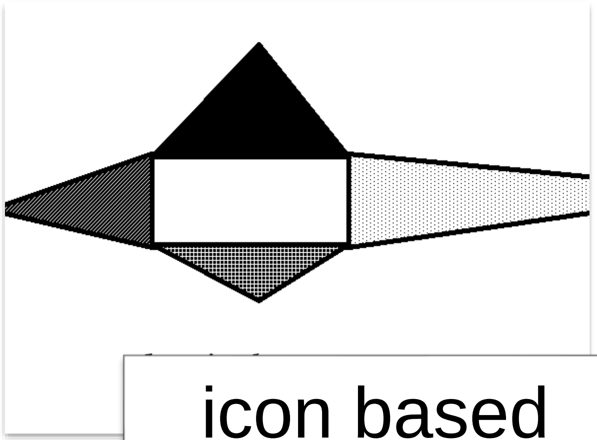
Classification of visualization methods for multiparameter data



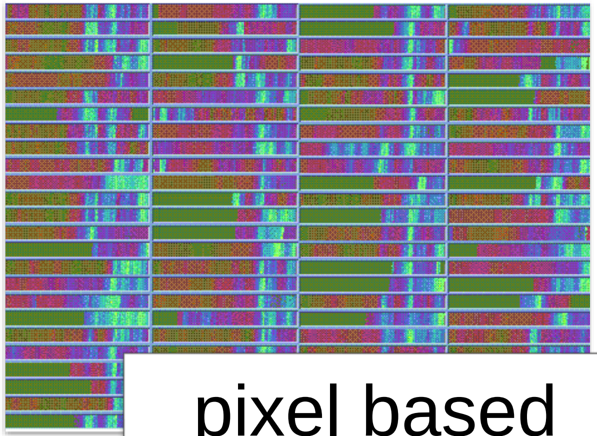
panel matrices



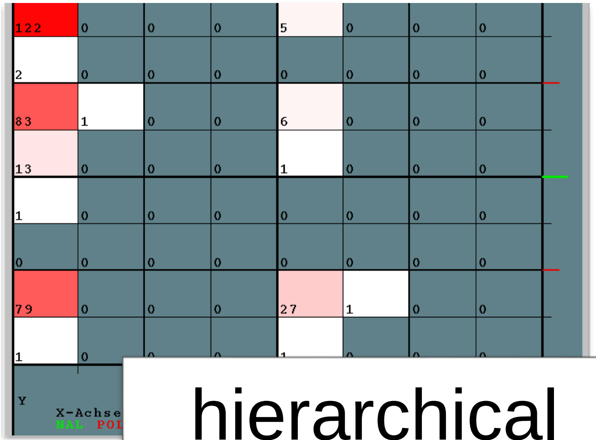
embedding



icon based



pixel based



hierarchical

Embedding high-dimensional observation cases in 2D or 3D

oriented towards visualization:

- parallel coordinates
- radar chart
- star coordinates

oriented towards data analysis:

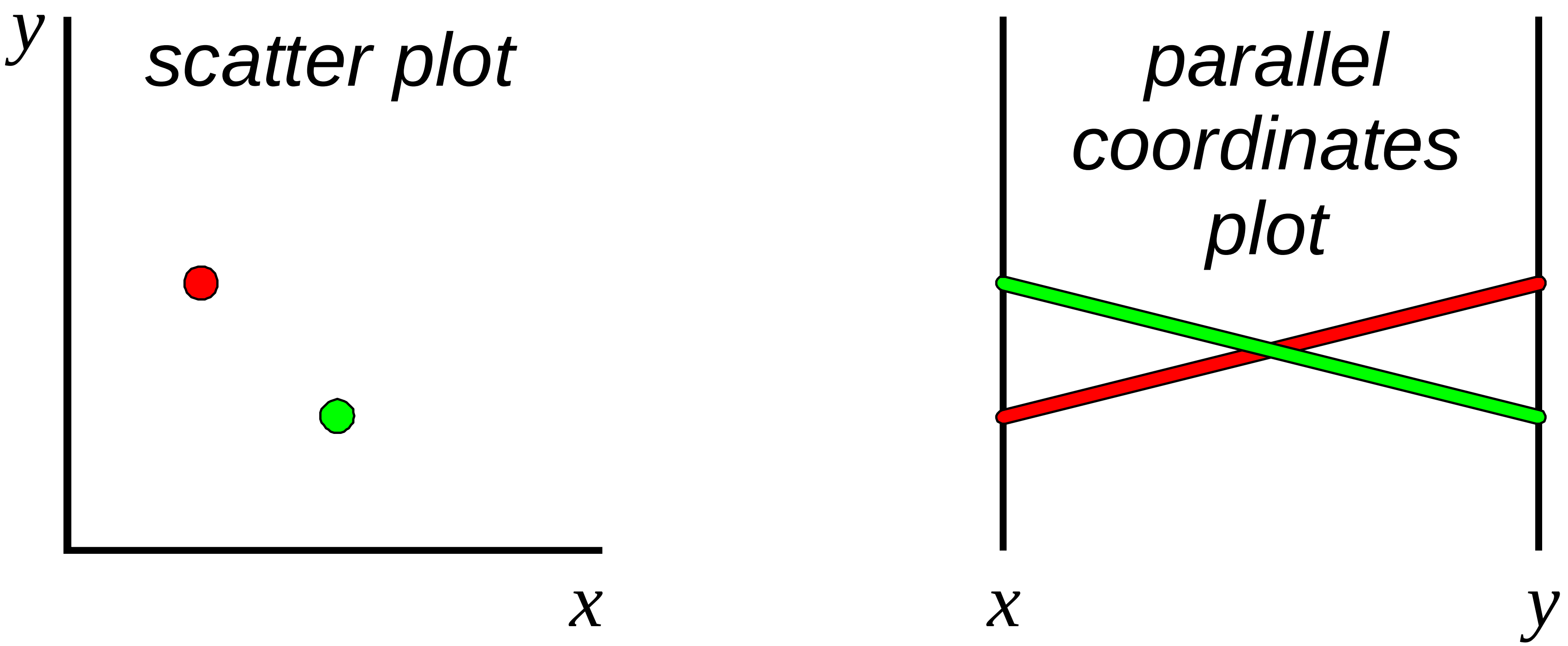
- principal component analysis
- multi-dimensional scaling
- t-SNE

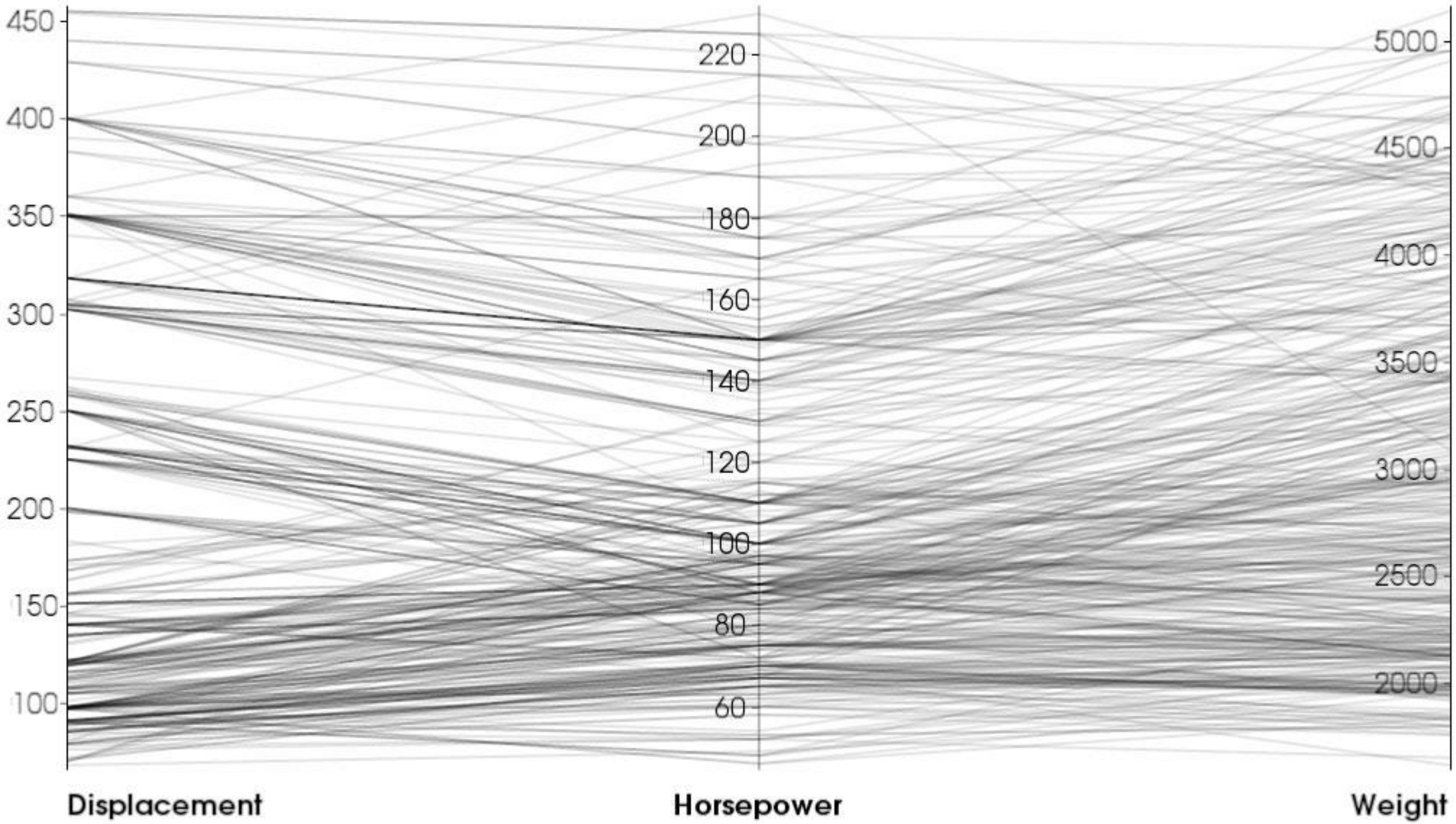
very important

Parallel coordinates (Inselberg, 1990)

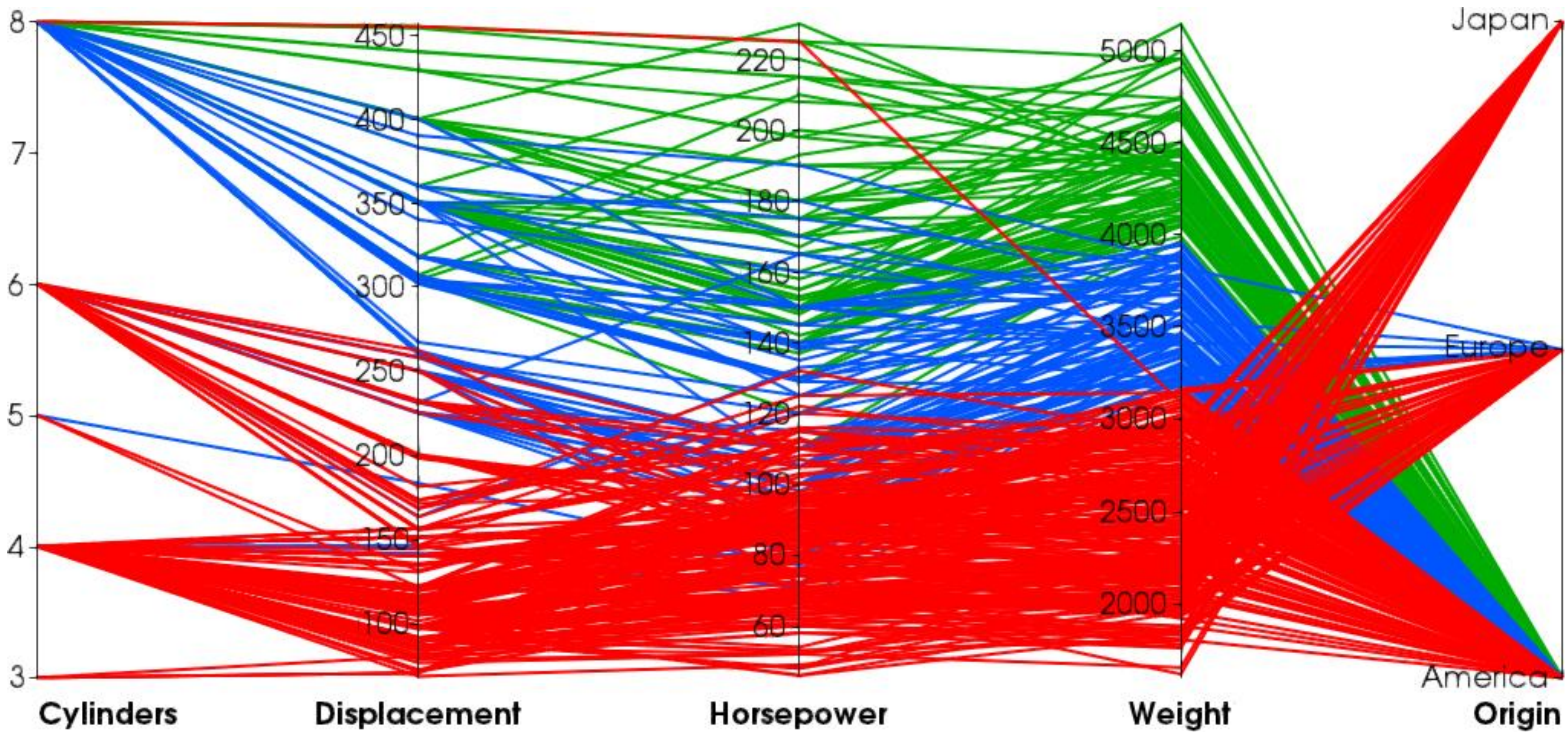
- n-dimensional parallel coordinate system in plane is defined by n parallel axes x_1, x_2, ..., x_n

- A point P with the coordinates (p_1, p_2, ..., p_n) in n-dimensional data set is represented by a line sequence in parallel coordinates. This line sequence is defined by the points P_1, P_2, ..., P_n on the axes x_1, x_2, ..., x_n of the parallel coordinate system. P_i lies on x_i and has the coordinate p_i ($i = 1,2,...,n$).


y
scatter plot
x
parallel
coordinates
plot
x
y



Parallel Coordinates Plot with 3 variables of the *Cars* data set

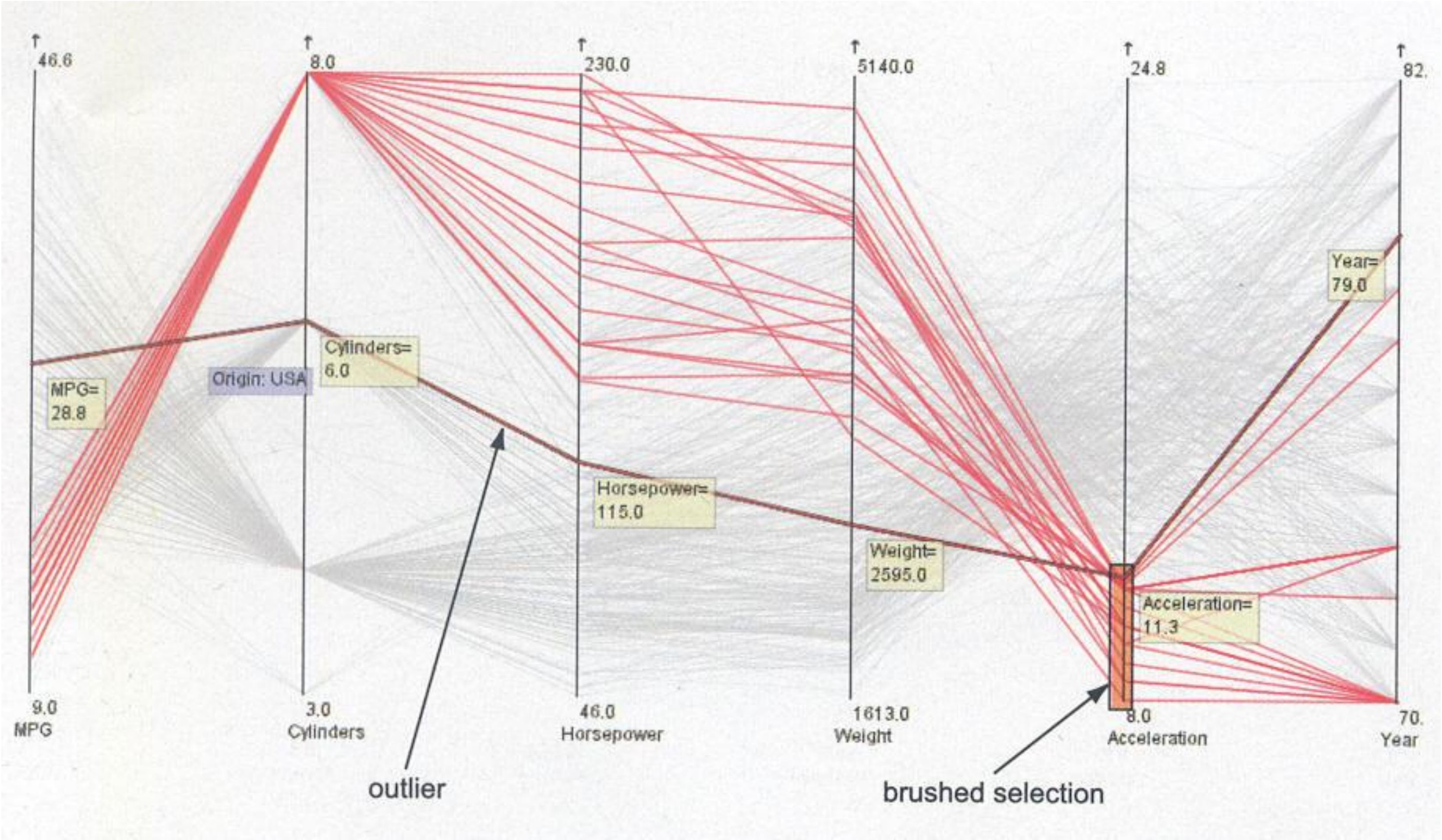


Parallel Coordinates Plot with 5 variables of the *Cars* data set very important

Parallel coordinates (Inselberg, 1990)

For the visual analysis, it is important to offer interaction tools such as:

- reducing the number of axes
- exchanging axes
- scaling axes
- reducing the number of line sequences
- emphasizing particular line sequences (e.g., using color).



Brushing selections to select subsets

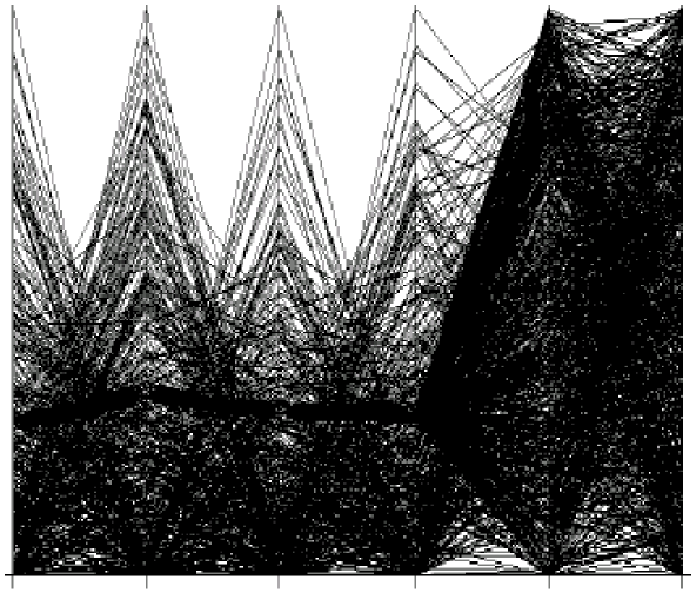
Car example (speed, cylinders, PS, weight, acceleration) from: [Telea, 2008]

Overplotting

possible remedies:

- interaction
- clustering & edge-bundling
- advanced rendering
- …

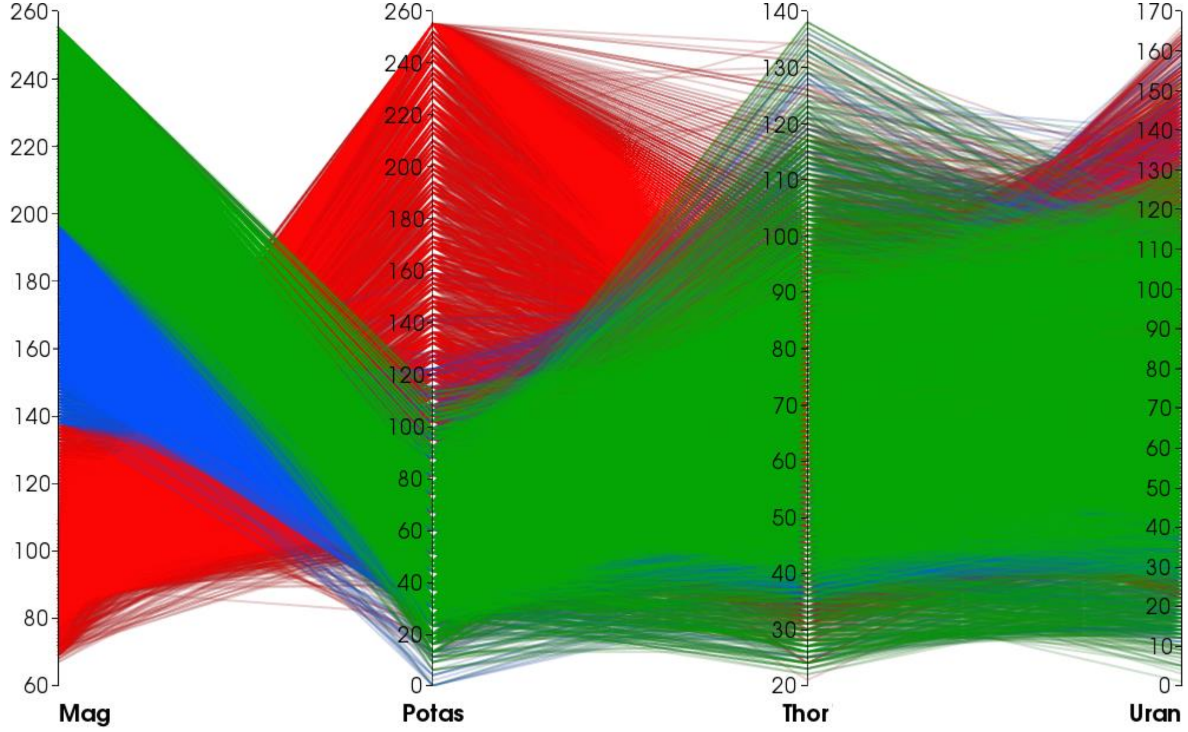
15.000 data items with noise



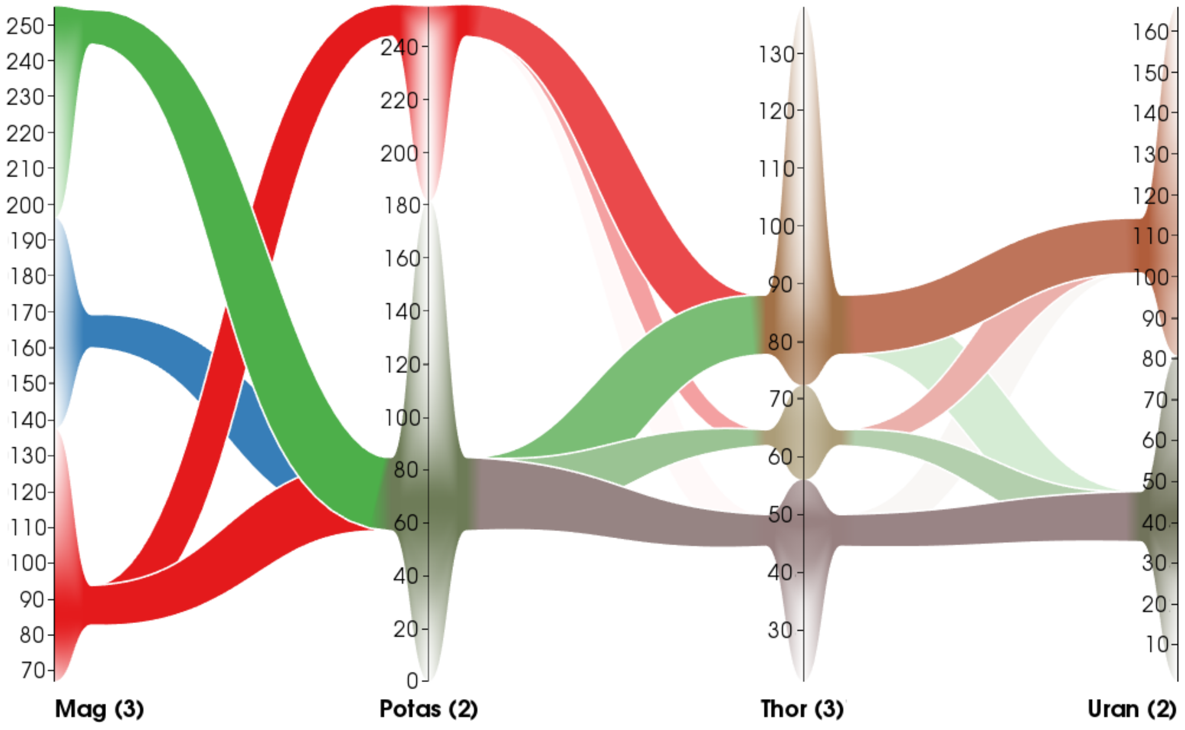
5% of the data (750 data items)

lines need space!

classic version with overplotting

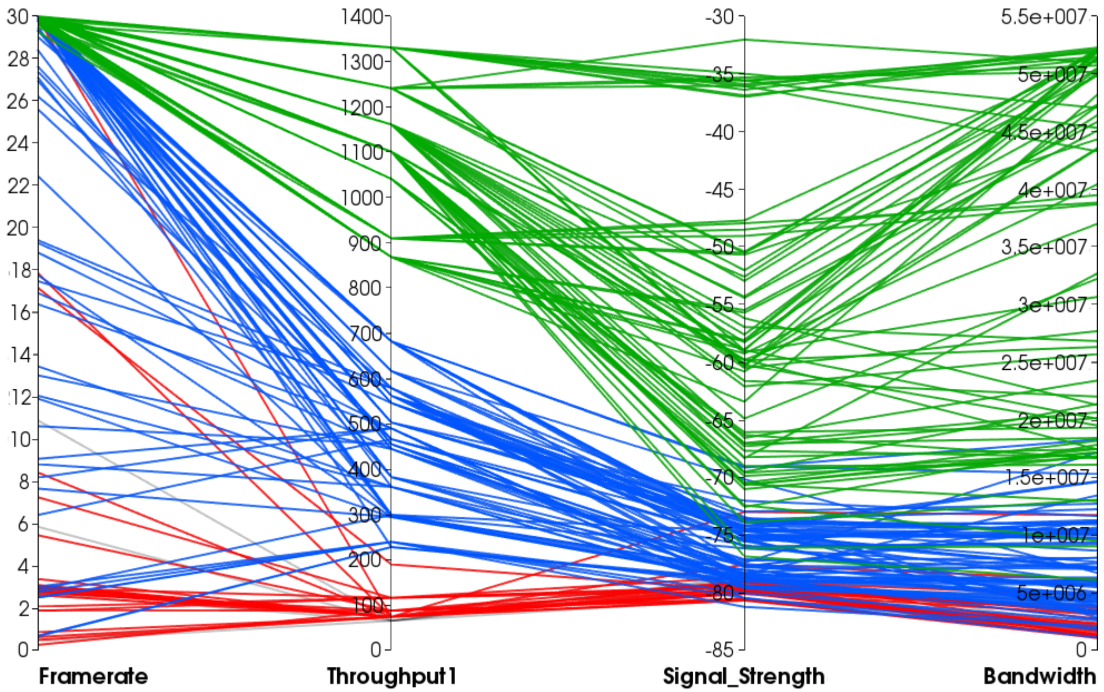


edge-bundled version with stripes

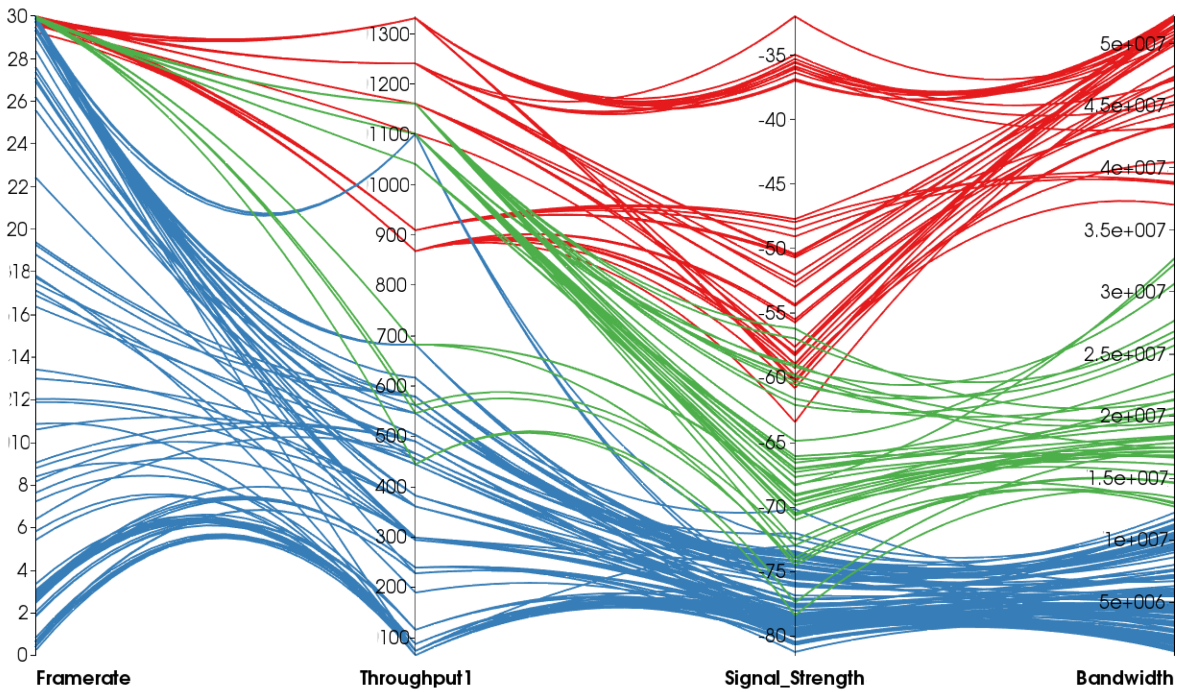


Palmas et al., PacificVis 2014

classic version

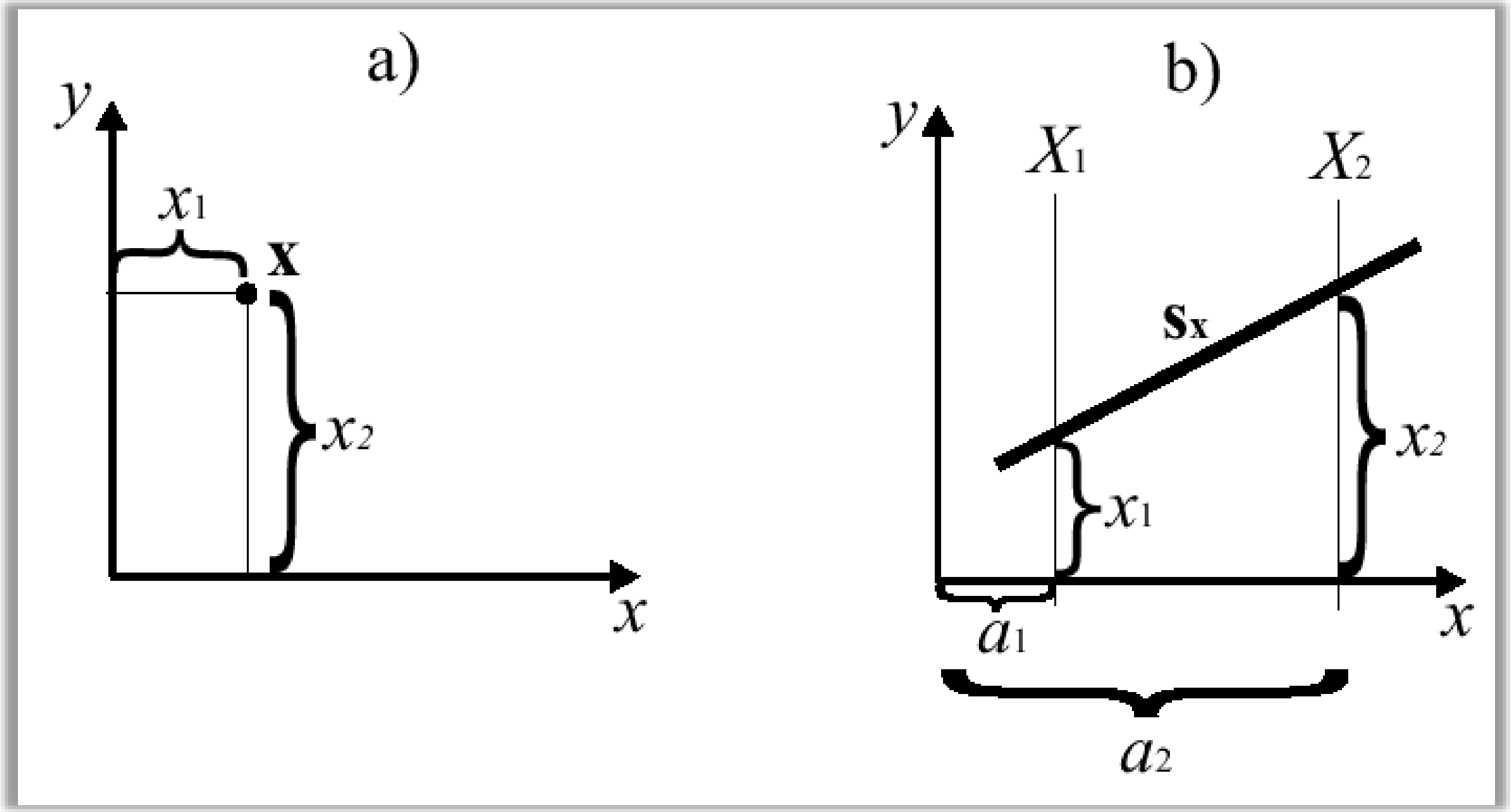


edge-bundled based on clusters



images from Palmas et al., PacificVis 2014; edge-bundling method according to McDonnell and Mueller, CGF 2008

Point-line duality for parallel coordinates I



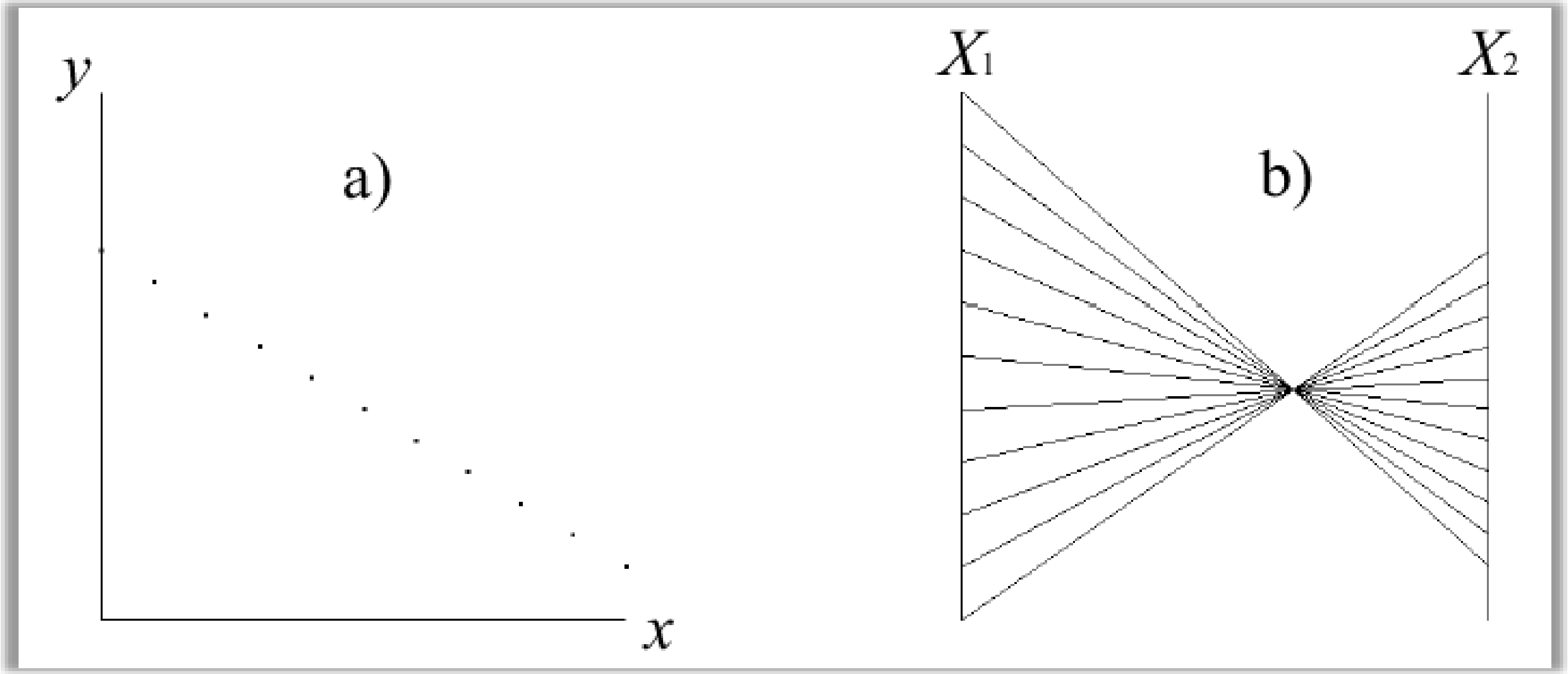
The point **x** in a) represented by the line $\mathbf{s_x}$ in parallel coordinates in b).

Point-line duality for parallel coordinates II

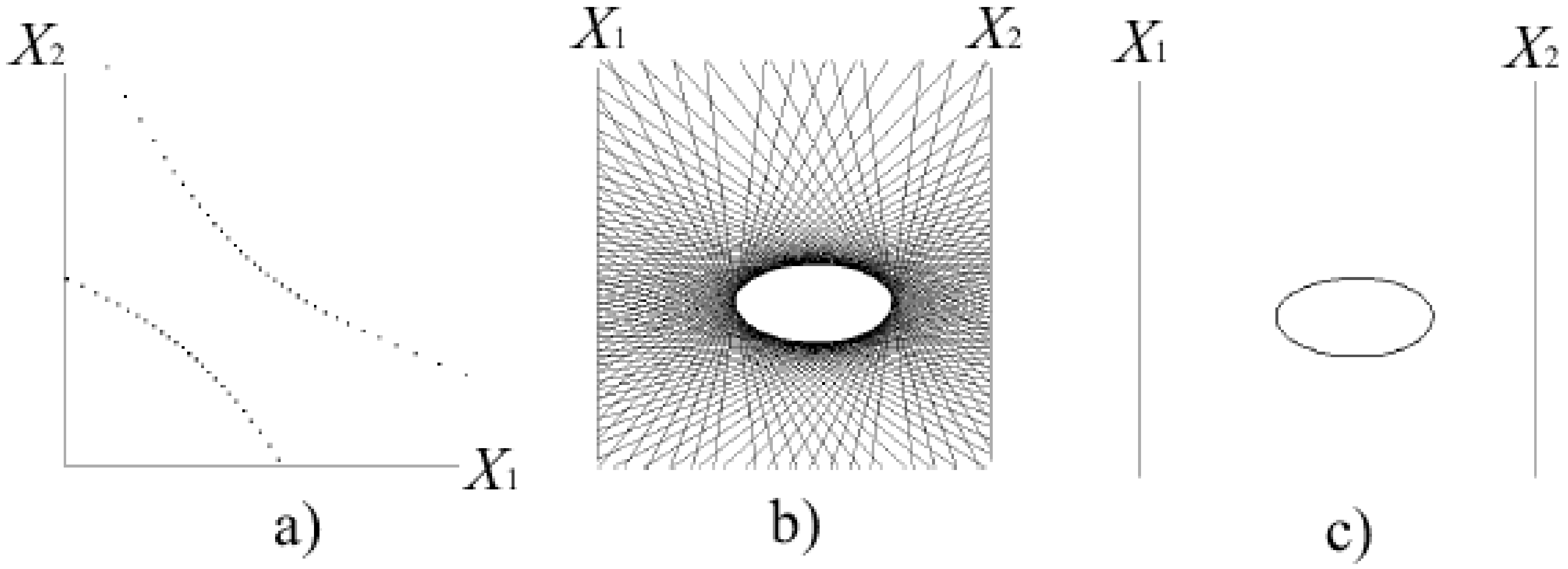
A straight line in the Cartesian coordinate system (a) is found as the intersection point of the lines in Parallel Coordinates (b). These lines in PC are defined by the sample points on the straight line.

The intersection is not necessarily between X_1 and X_2

An intersection point is not defined, if the slope of the straight line is 1, since then the lines in PC are parallel to each other.

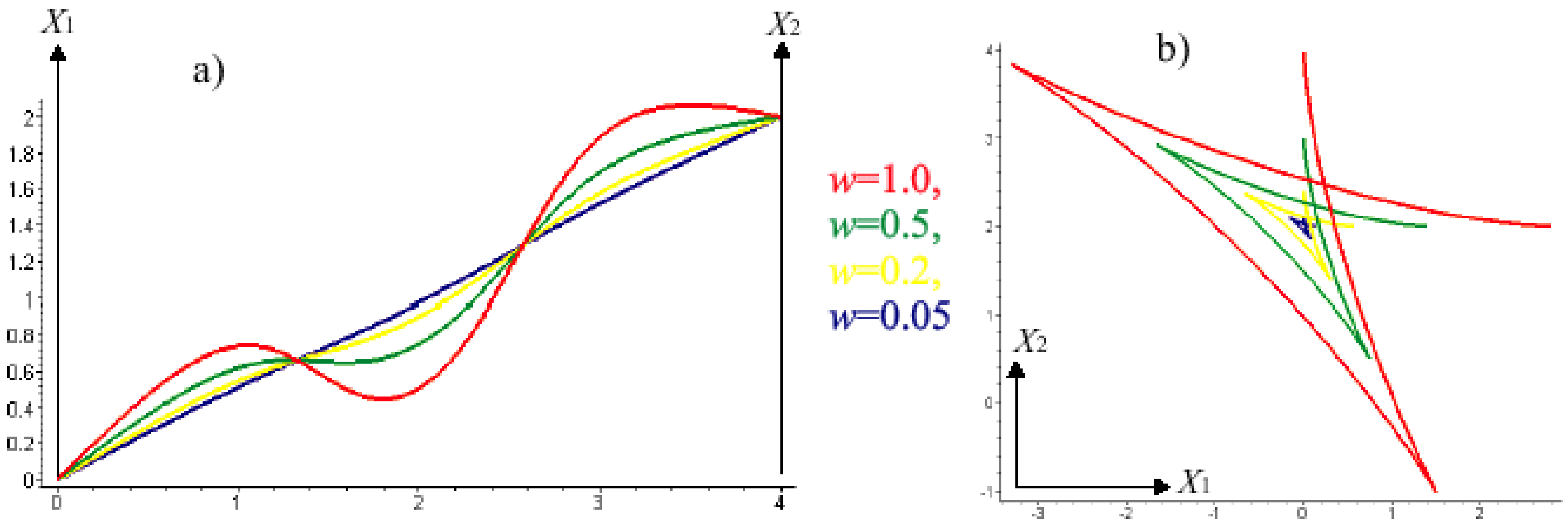


Conic-conic duality for parallel coordinates



Curve-curve duality for parallel coordinates

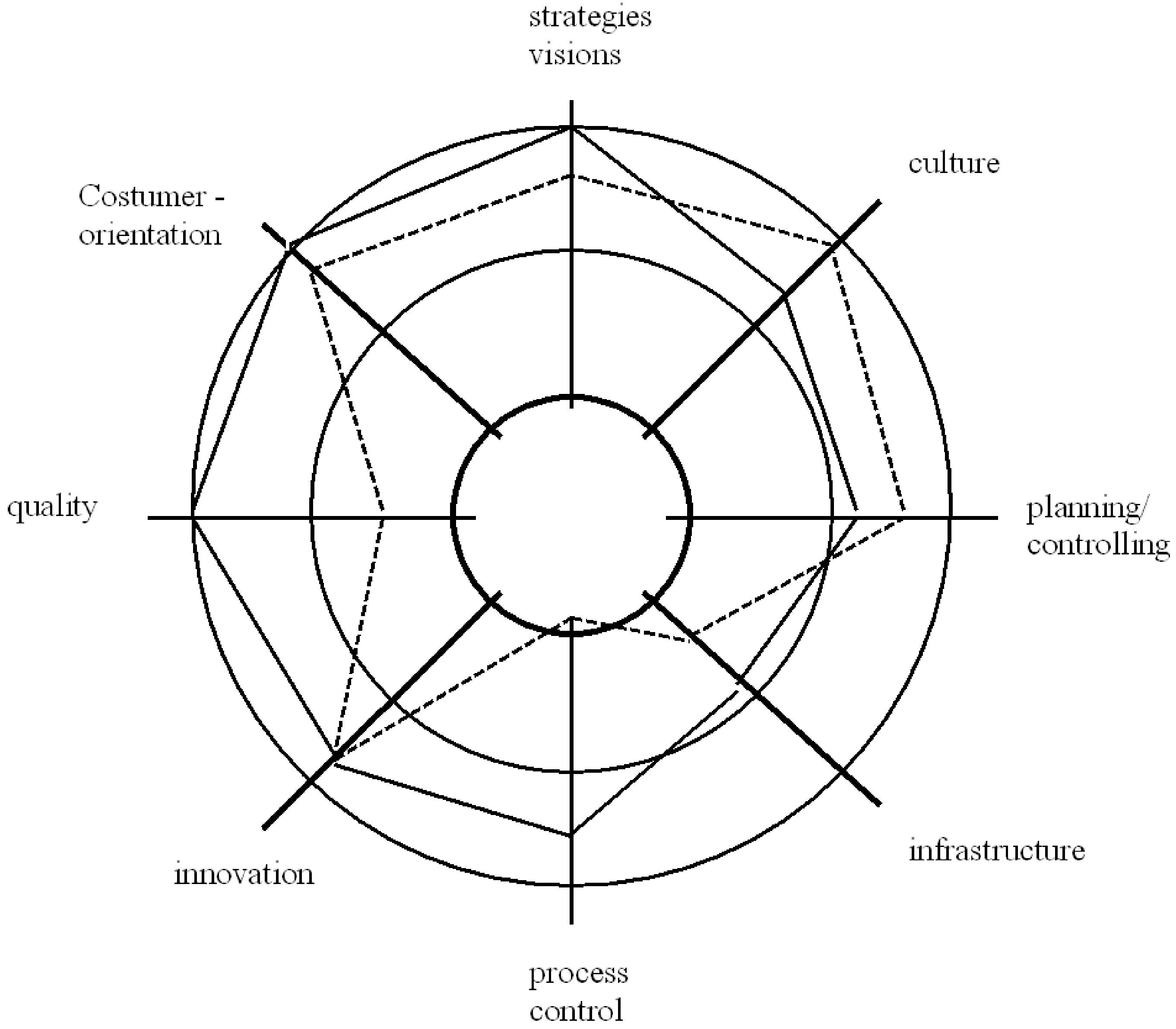
- Cusp <-> inflection point duality



Radar Chart

similar to parallel coordinates but circular arrangement of coordinate axes

also called Kiviatgraph



Properties of a company

Star Coordinates

circular arrangement of axes

not always the same angle or length

data value in each dimension becomes a vector

independently for each dimension

sum of vectors gives position of the observation case in the graph

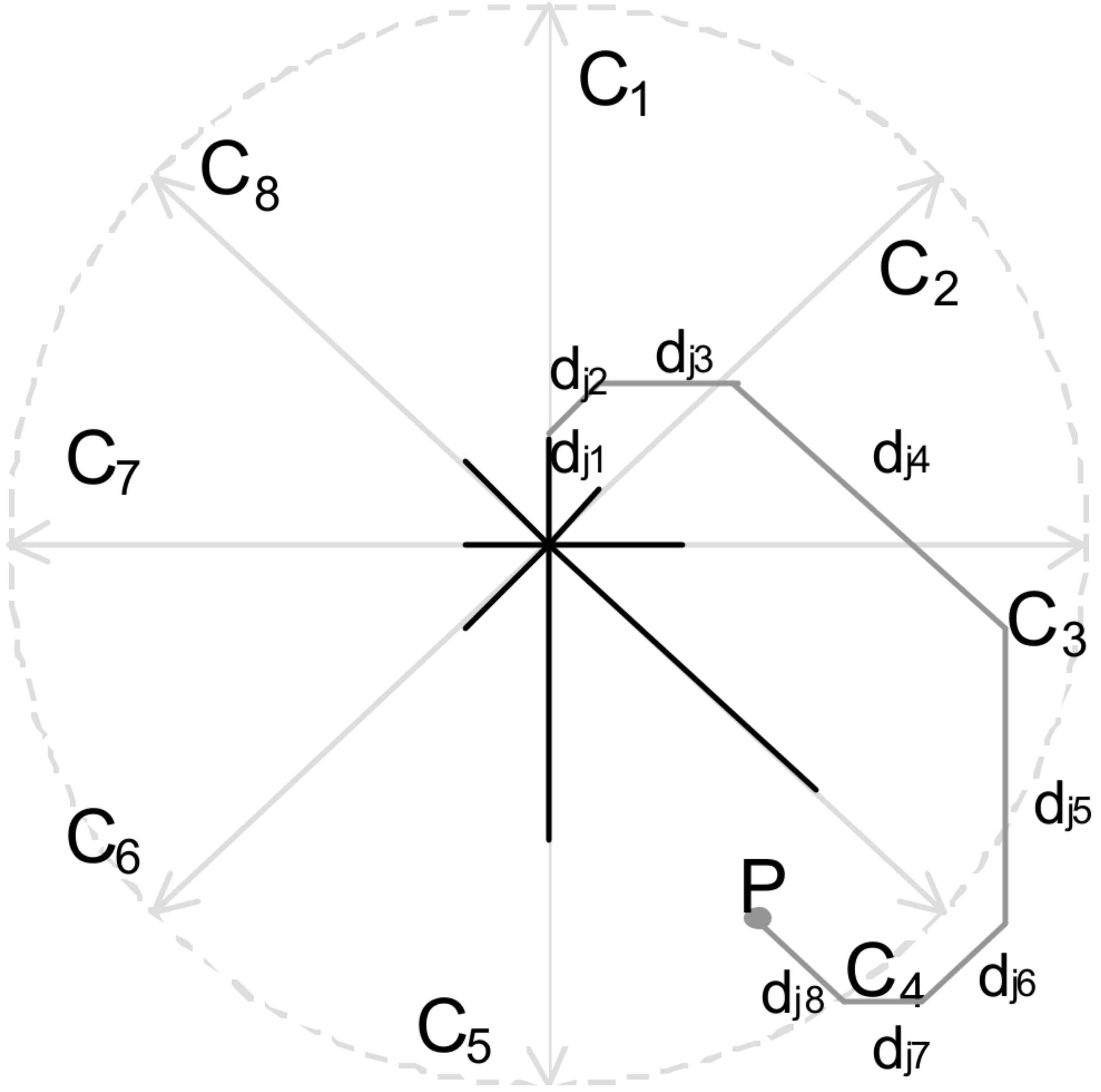


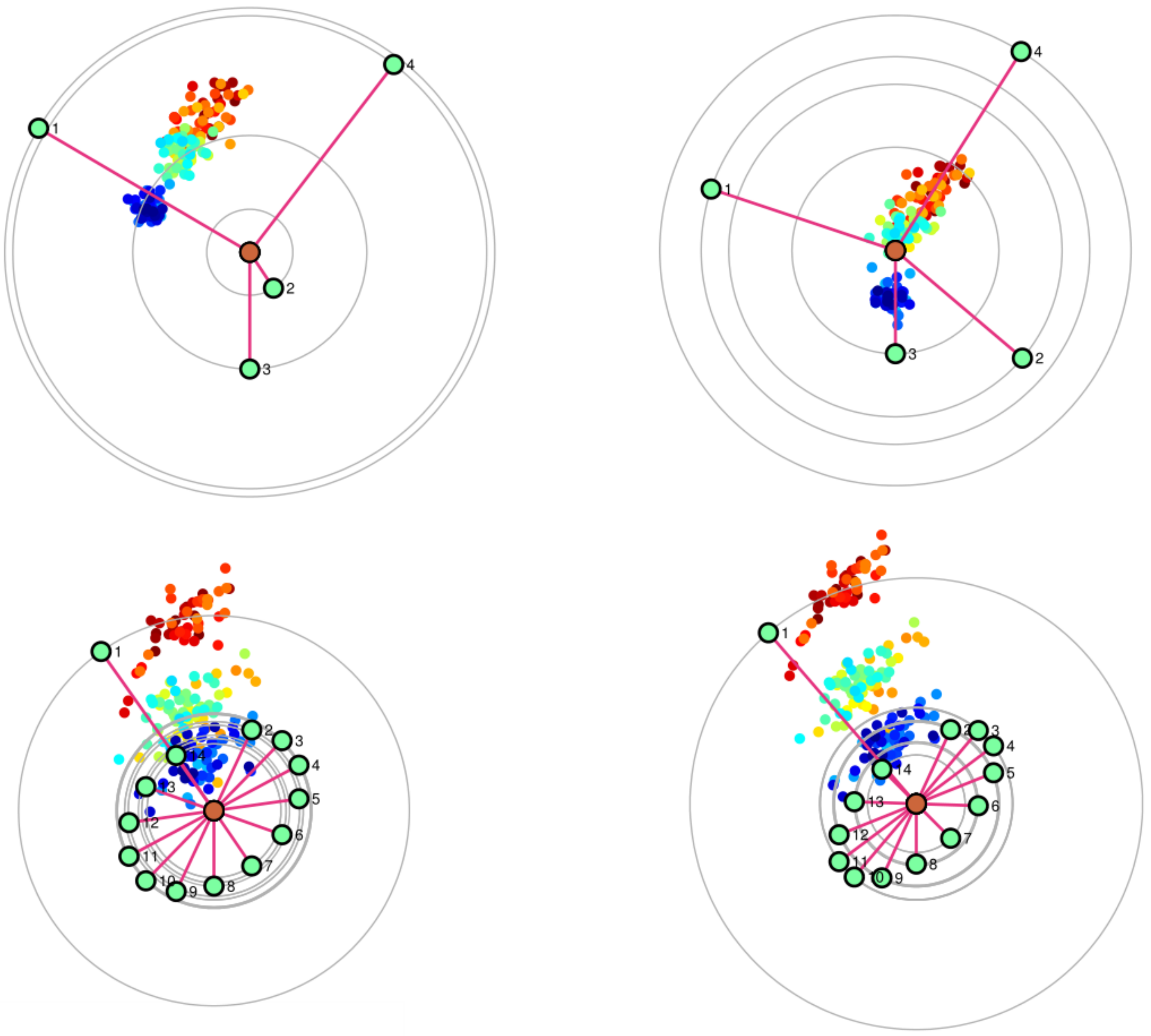
image from Kandogan, InfoVis 2000

Star Coordinates

different arrangement of axes leads to different projections

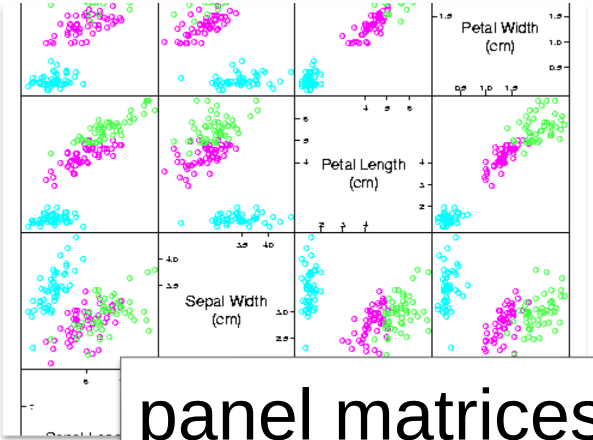
orthographic projection may be beneficial

different angle or length for coordinate axes

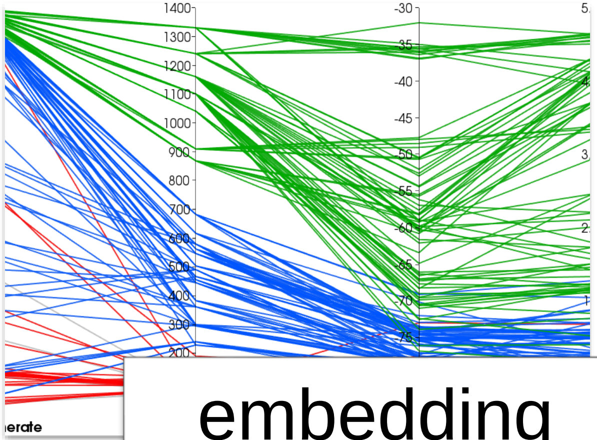


images from Lehmann and Theisel, InfoVis 2013

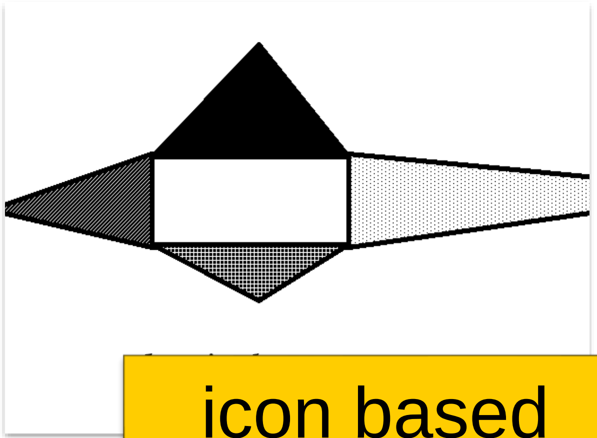
Classification of visualization methods for multiparameter data



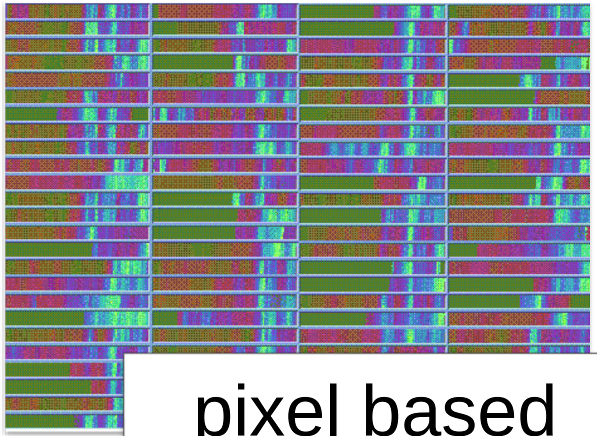
panel matrices



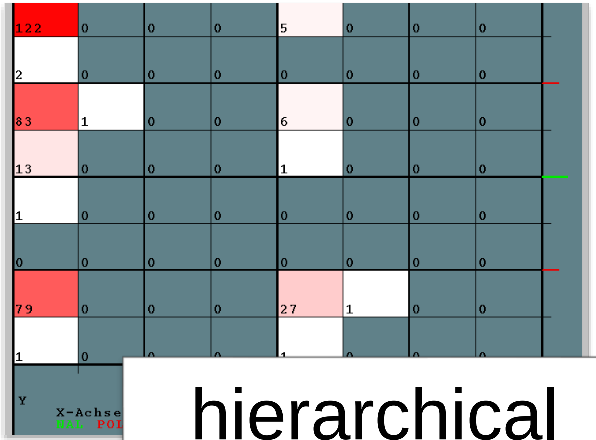
embedding



icon based



pixel based



hierarchical

Icon-based Techniques

icon = observation case

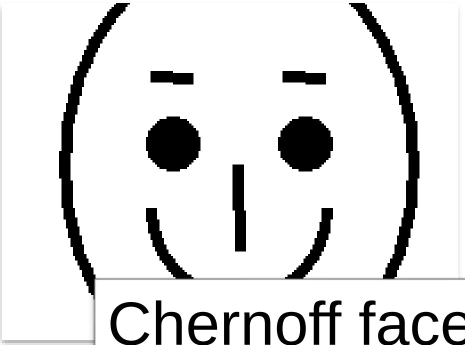
variables visualized as properties of icon

independent coding

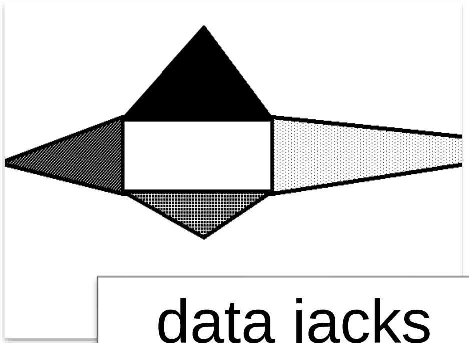
combinable, perceivable

often used on maps

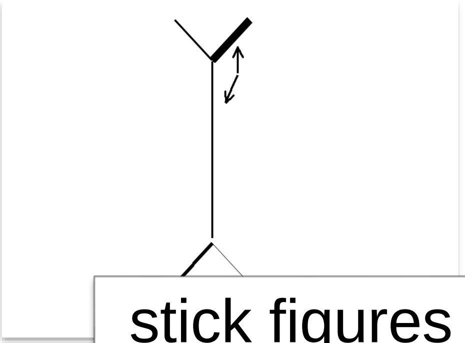
exact spatial placing



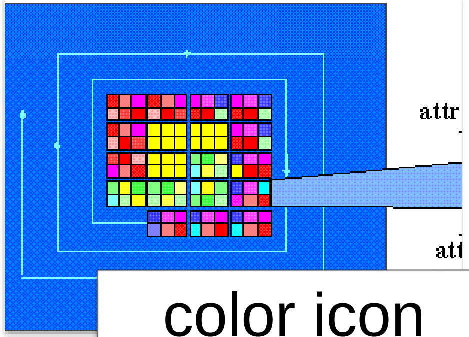
Chernoff face



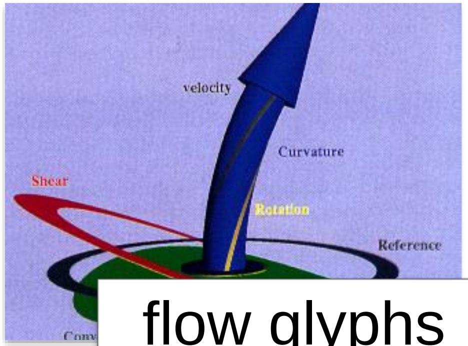
data jacks



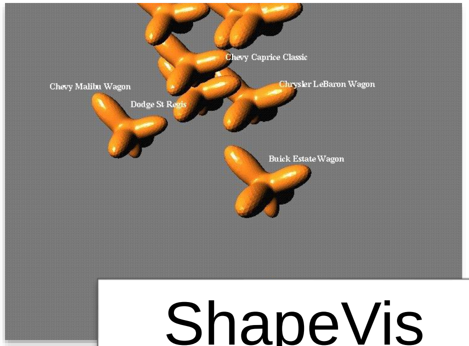
stick figures



color icon



flow glyphs



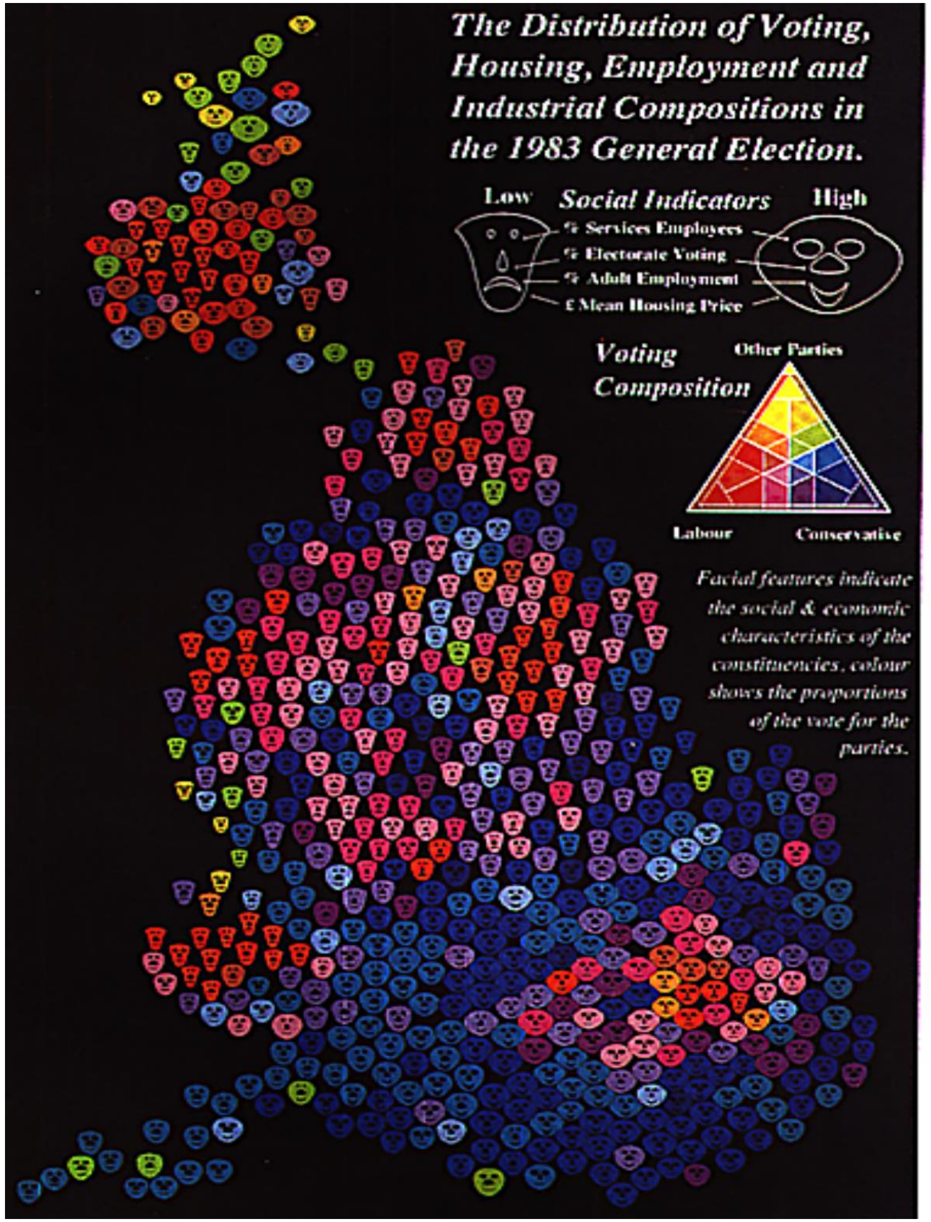
ShapeVis

- **Icon based techniques**
  - **idea**: visualization of data values as properties of icons / glyphs.
    ➔ One icon represents the values of one observation case.
  - There is no unique distinction between the names „icon“ and „glyph“ in the literature. In most cases they have the same meaning.

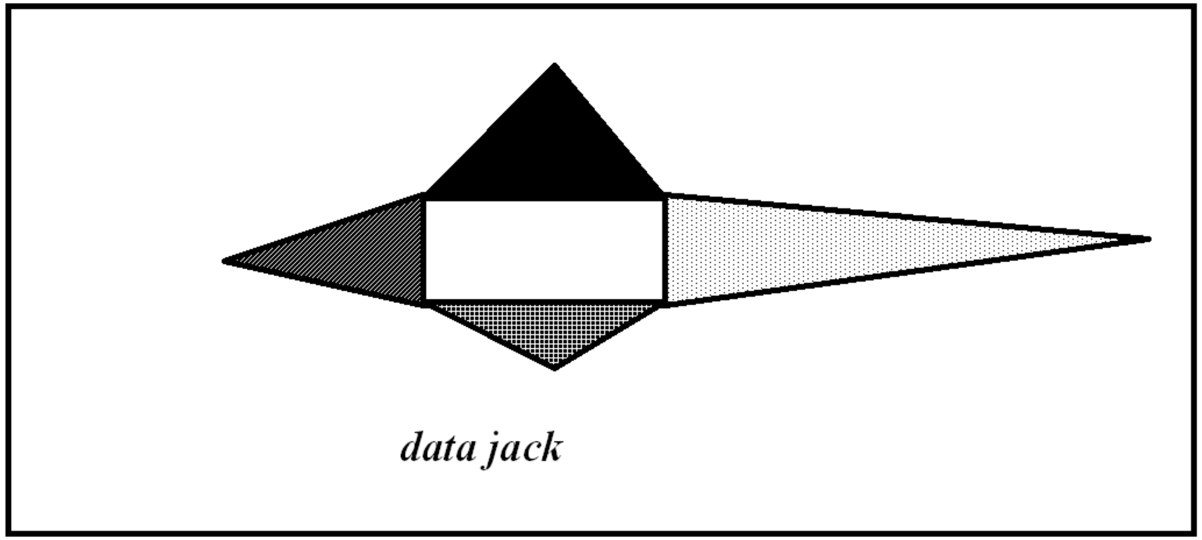
- **Glyphs/icons are geometric objects with the following properties**:
  - can be located at exact positions; these positions give spatial information.
  - encode values of physical variables:
    - in geometric characteristics like angle, length or shape,
    - in attributes of the representation like color or transparency.

- Icons are used for encoding more than 3 parameters.
- **These parameters should have the following properties:**
  - parameters in an icon should be easily combinable.
  - parameters should be separately perceivable.
  - icons should differ significantly if the coded parameters differ.

- The development of effective geometric codes is not trivial!
- One of the first (and most well-known) icons is the Chernoff face (Chernoff 1973).
- It codes 12 variables in the form of head, nose, mouth, eyes

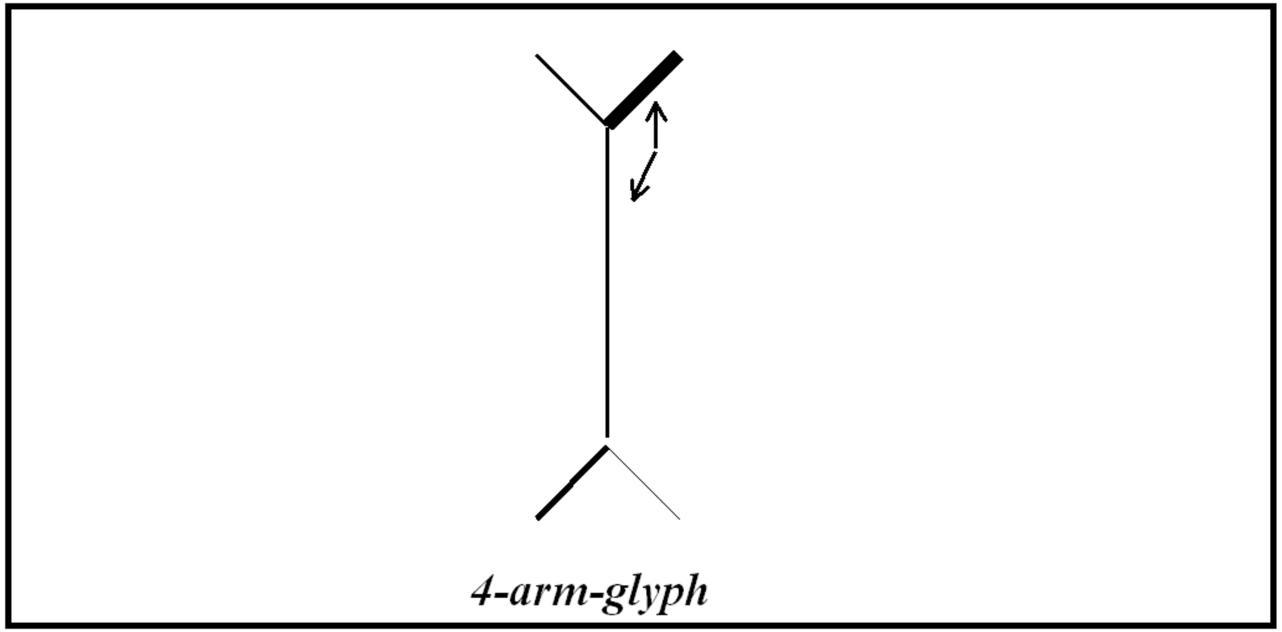

The Distribution of Voting, Housing, Employment and Industrial Compositions in the 1983 General Election.
Low
Social Indicators
High
% Services Employees
% Electorate Voting
% Adult Employment
£ Mean Housing Price
Voting Composition
Other Parties
Labour
Conservative
Facial features indicate the social & economic characteristics of the constituencies, colour shows the proportions of the vote for the parties.

- **Further examples**
  - data jacks (Cox, Ellson, Olano)
  - 3-dimensional figures consisting of 4 elements
  - data values are coded in color and length of these elements.



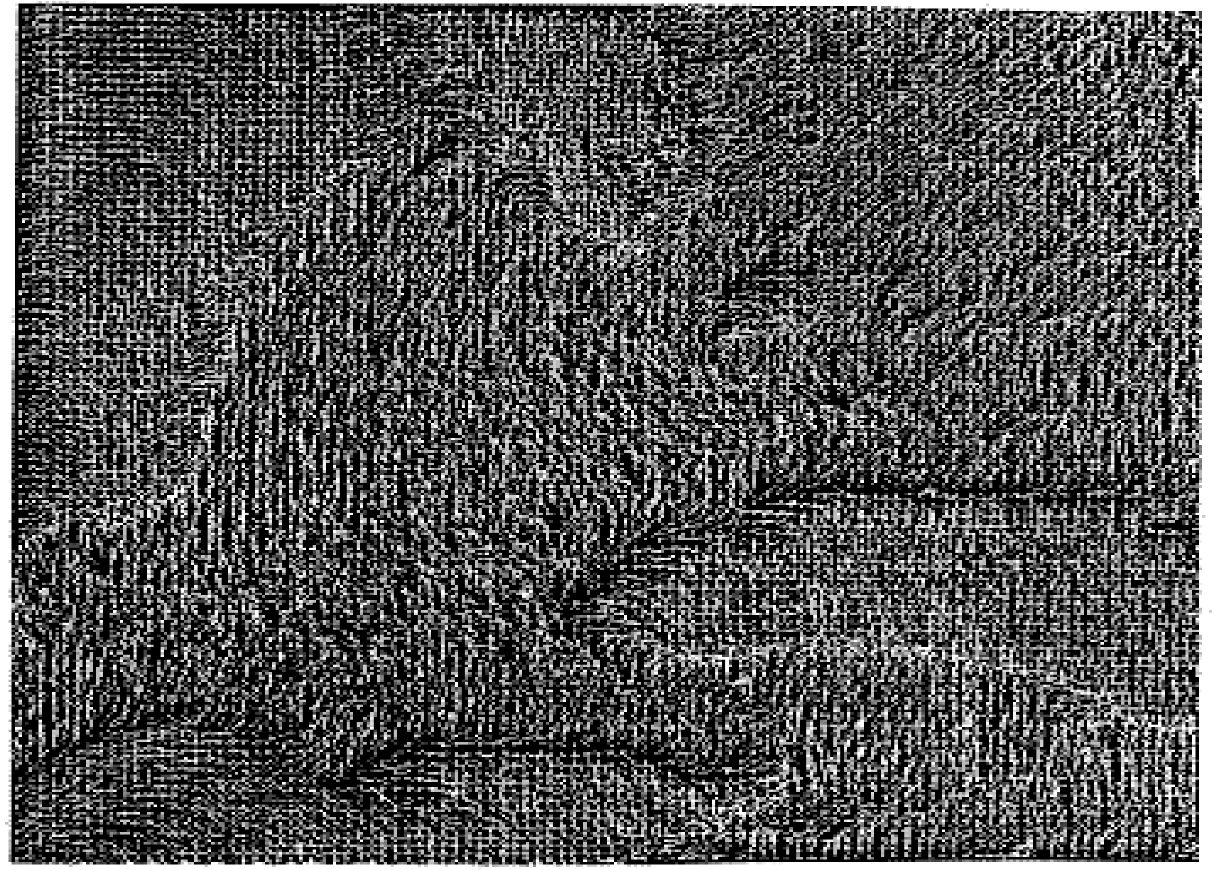
- **Further examples**
  - *m*-arm-glyph, stick figures (Pickett, 1970; Pickett & Grinstein, 1988)
  - 2-dimensional figure with *m* arms; data values are coded in
    - Length,
    - Width, and
    - Angle to main axis

    of the arms.



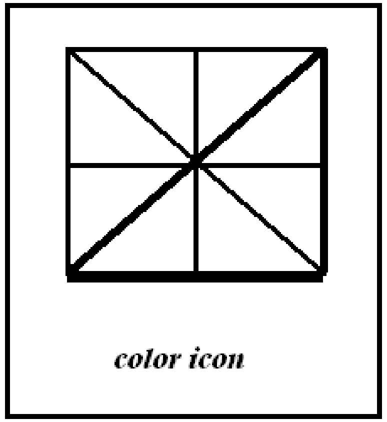
- **Stick figure example:**

Stick Figures (cont'd)

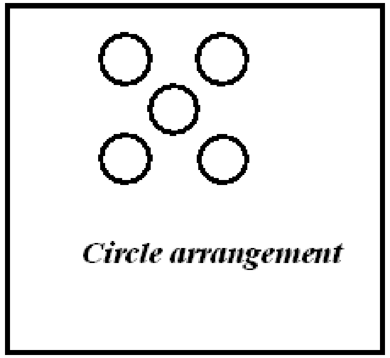


5-dim. image data from the great lake region

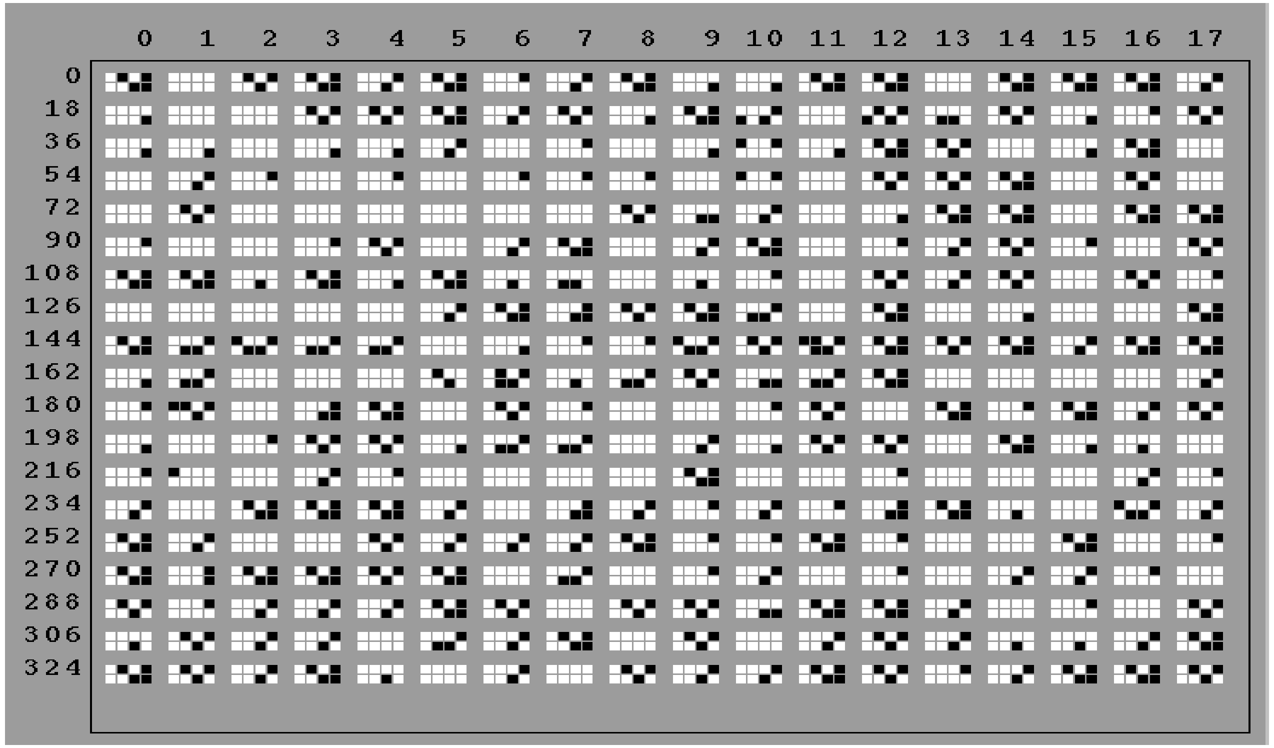
- **Color icon (Levkowitz)**
  - uses color, shape, and texture features;
  - data values are coded
    - in linear feature (edges, diagonal), or
    - in triangles.



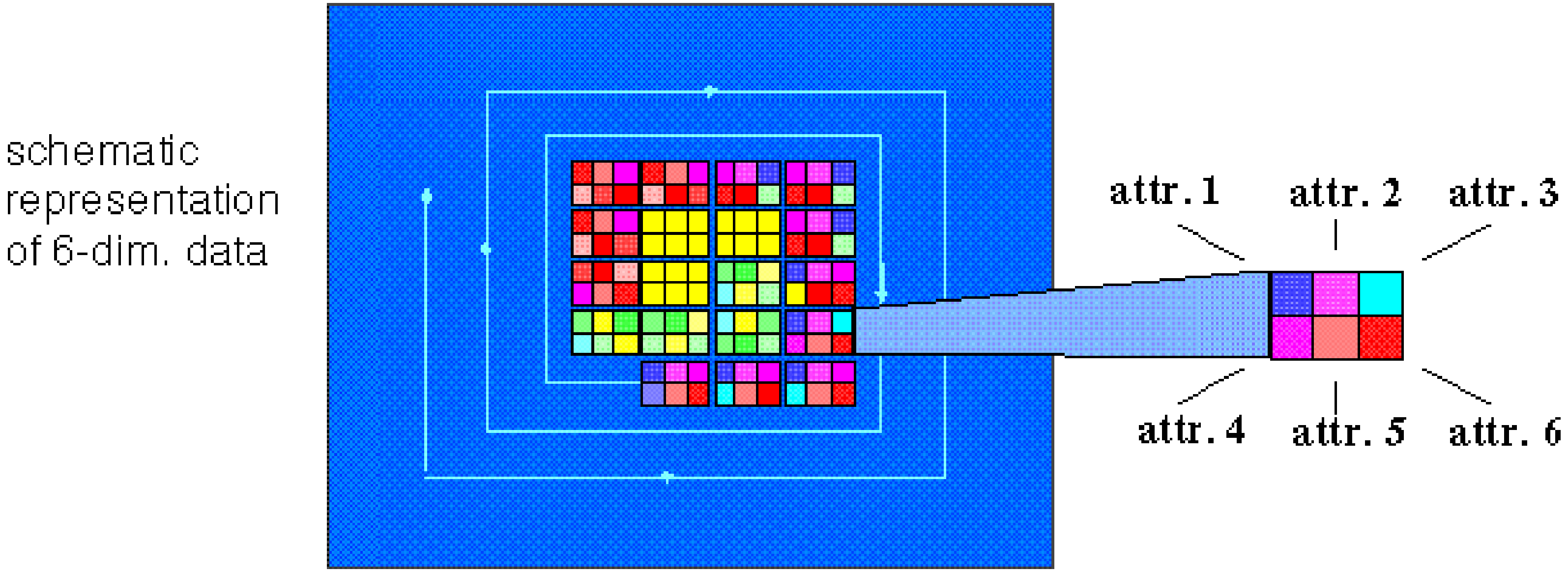
- **Circle arrangements**
  - codes data values in size and color of circles



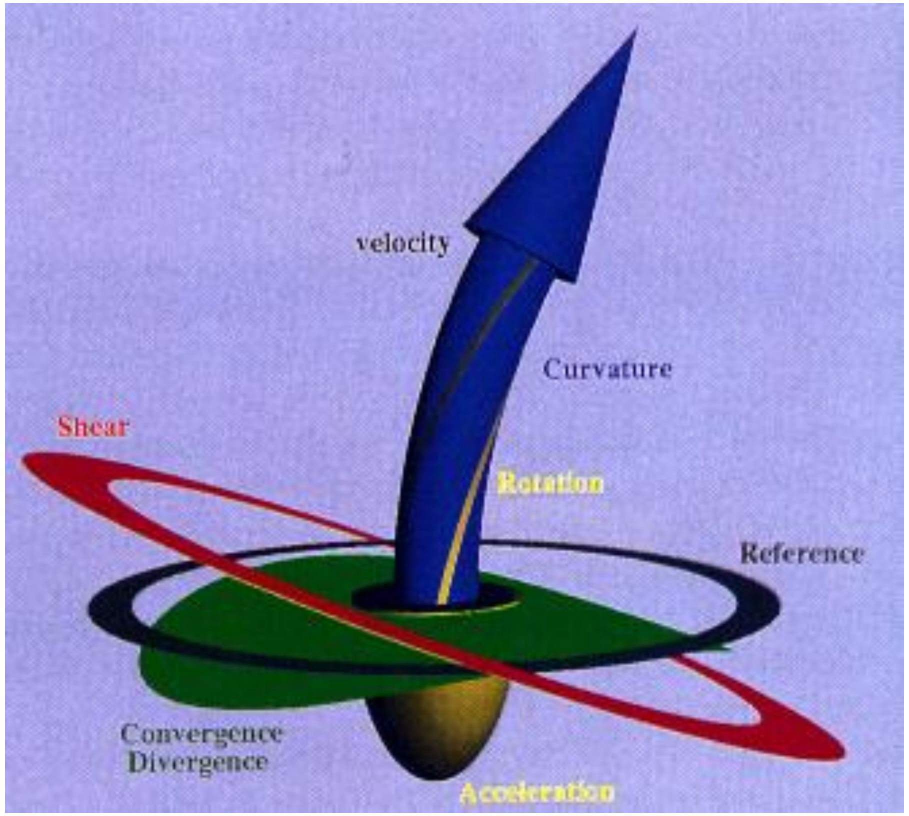
- **Shape coding (Beddow, 1990)**
  - Icon is a rectangle containing a regular grid. Each grid cell is used for one parameter.
  - The value of the parameter is coded in color.
  - The rectangular icons are usually placed row-wise. Hence, certain order criteria can be considered. (temporal or spatial correlations, ordering according to special data values).



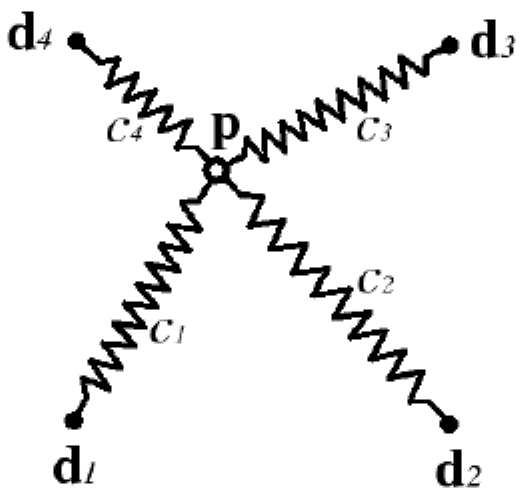
- **Spiral shaped arrangement of a car glyph**
  - The rectangular icons are usually placed row-wise.
  - ➔ Other configurations of the icons are possible. (spiral shaped configuration, grouping techniques, Keim, Kriegel, 1994)



- **Icons for flow visualization:**
- Place icon at selected locations and encode different values of the flow
- Seeding strategy necessary (usually interactive)
- Example: probe for local flow visualization [de Leeuw, van Wijk 93]



- **ShapeVis (Theisel/Kreuseler 98)**
  - Based on enhanced spring model
  - Classic spring model (Problems: ambiguity, insensitivity to coordinate scaling):
    - Every dimension is represented by a dimension point d_i (i=1..m) on a sphere
    - An observation case $O=(c_1,...,c_m)$ is represented by a point p
    - Springs from p to d_i with the stiffness c_i
      ➔ search for position of balance of p.



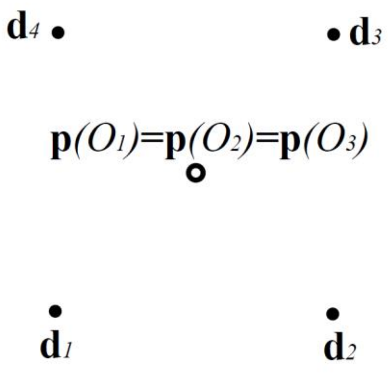
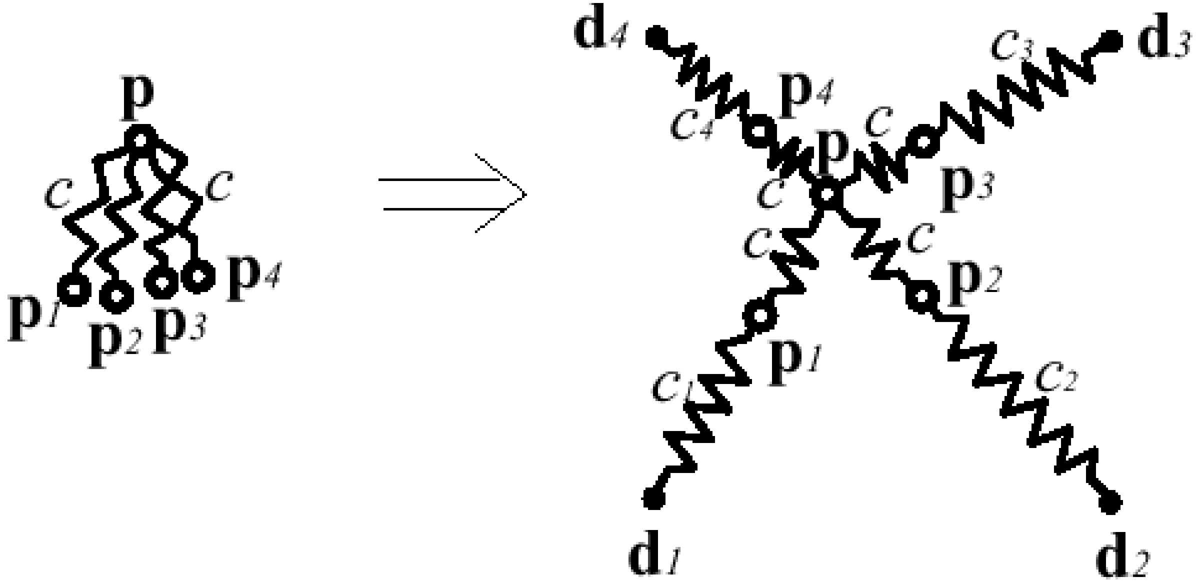
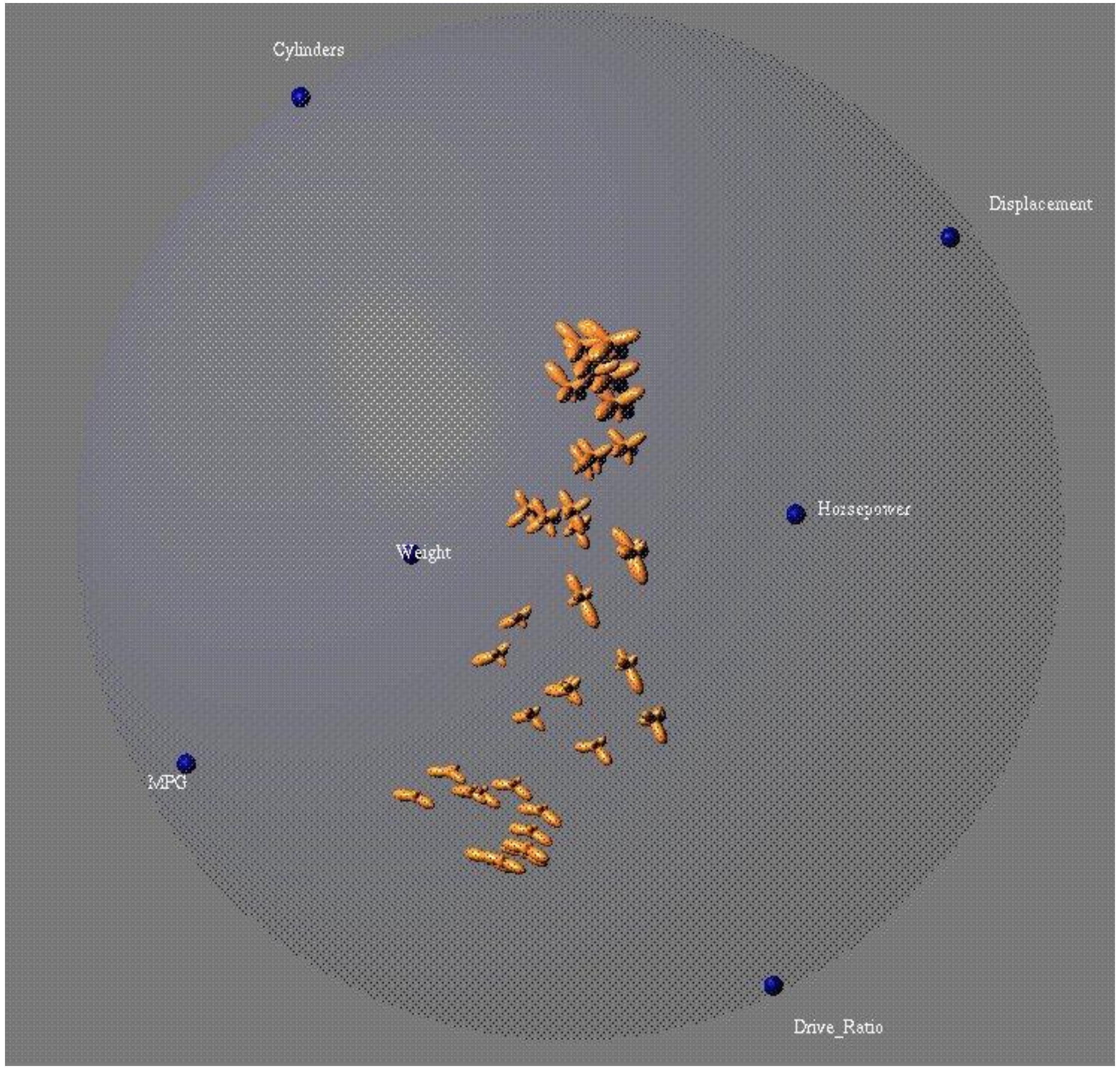


Figure 2: *Limitations of the classical spring model: the objects $O_1 = (1,2,1,2)$ and $O_2 = (2,1,2,1)$ collapse to one point (ambiguity). So do the objects O_1 and $O_3 = (2,4,2,4)$ (insensitivity to coordinate scalings).*

- **Solution:** observation case described by points $p, p_1, ..., p_m$:
- Then construct surfaces out of $p, p_1, ..., p_m$ in such a way that an observation case is uniquely described by its location, size and shape.

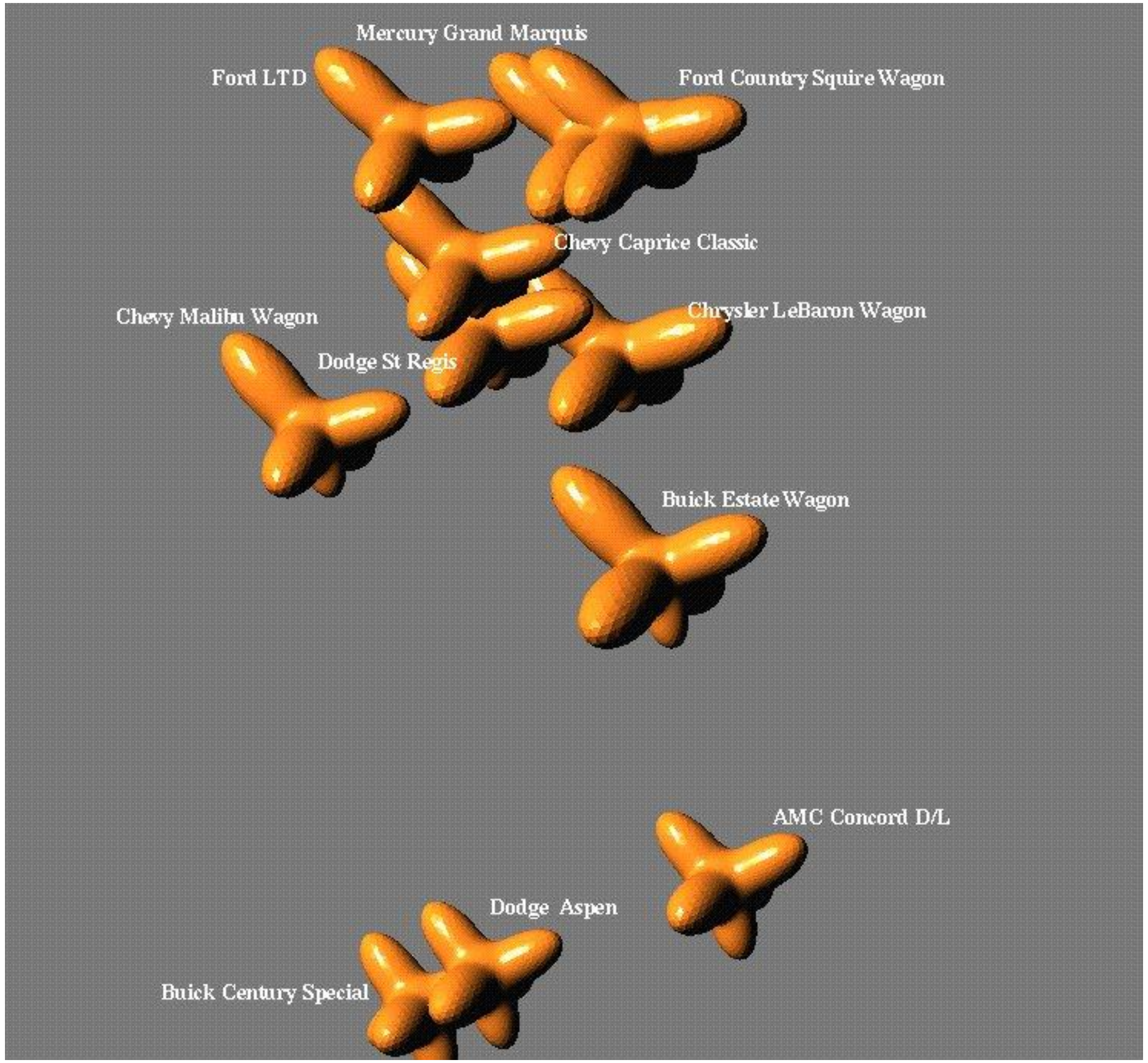


- **Example: car data set**



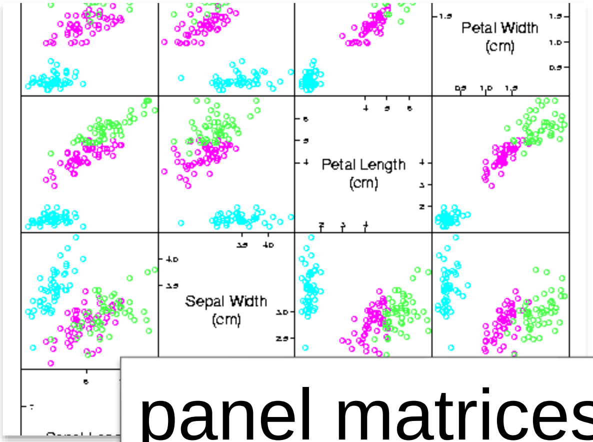
Overview

- **Example: car data set**

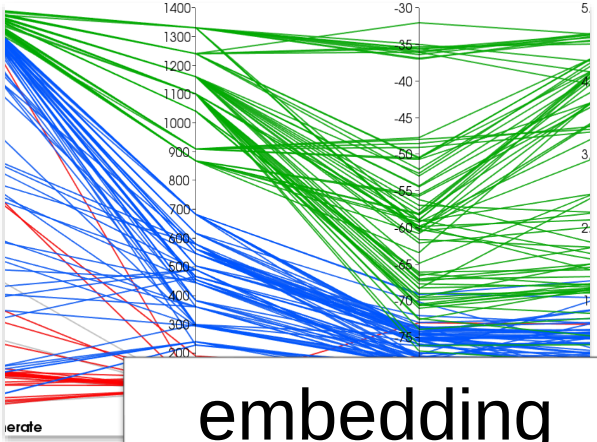


Zoom into a cluster

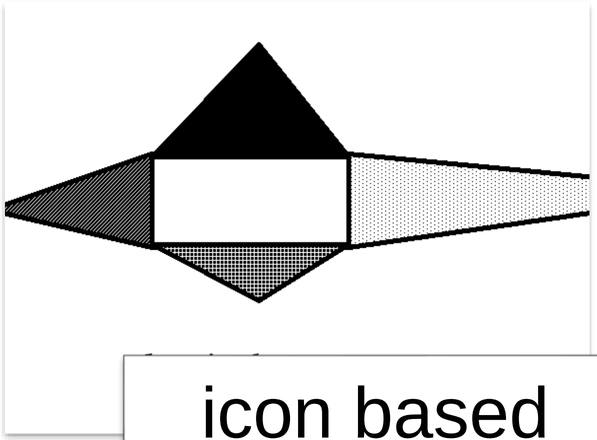
Classification of visualization methods for multiparameter data



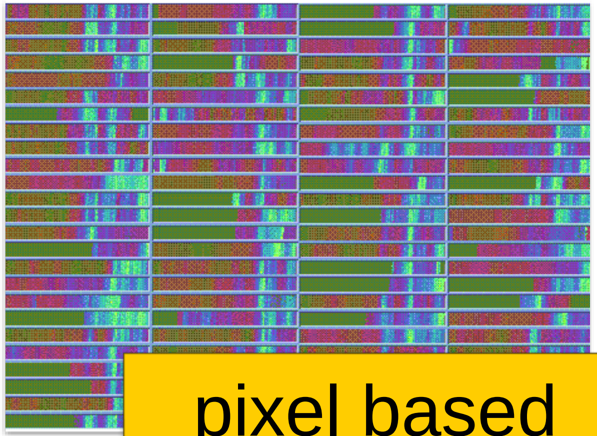
panel matrices



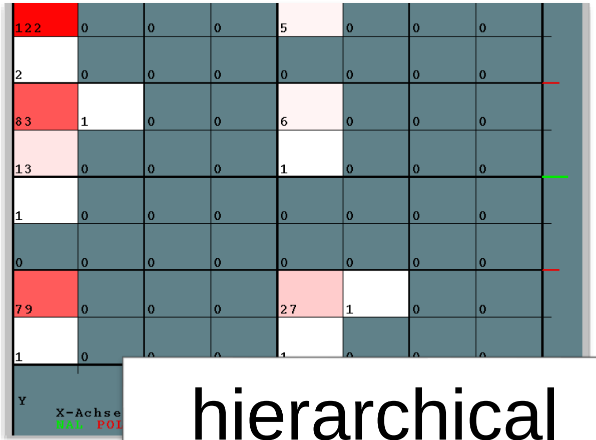
embedding



icon based



pixel based



hierarchical

Pixel-based Methods

one pixel for every data item

 # observation cases X # variables

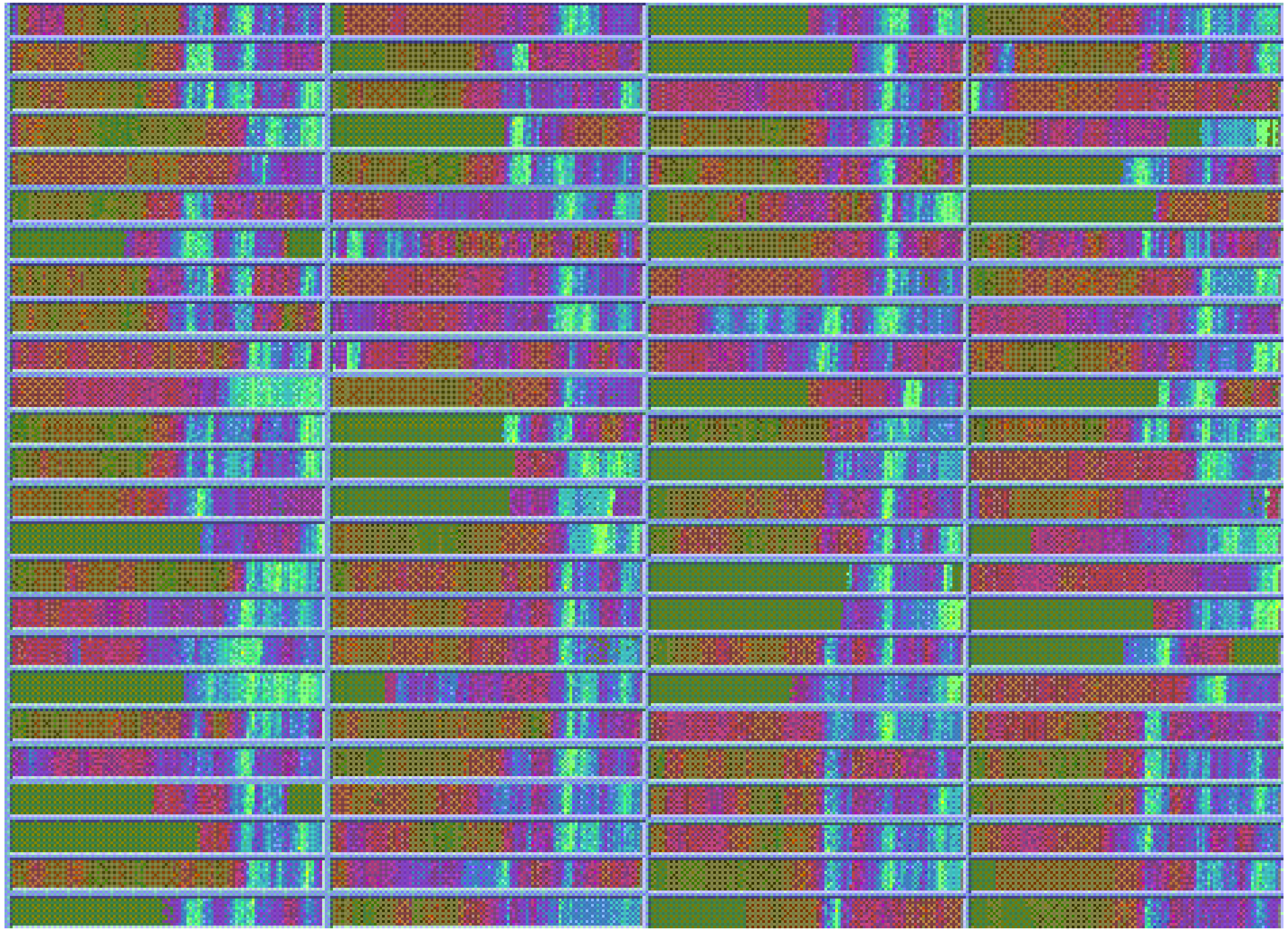
color coding

specific arrangement of pixels

application dependent

overview of data

details difficult to see



Stock prices of the FAZ index, Jan 74 – April 95

- **Pixel oriented techniques**
  - **Idea**:

    visualization of data values as color coded pixels; color coding of the data values is done in a separate window for each parameter.

    ➔ visualization of large data sets possible.

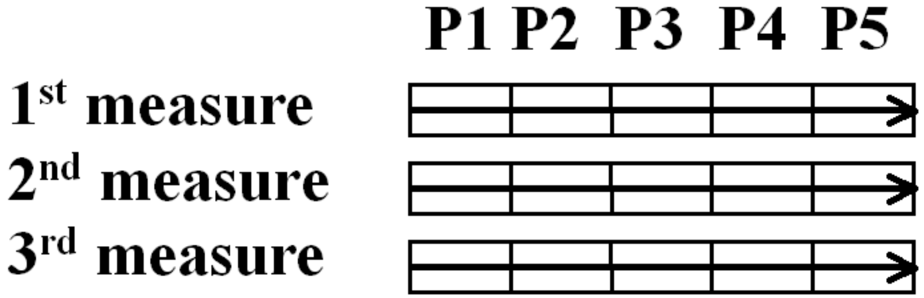
  - Arrangement of the pixels has to be specified:
    - Row-wise
    - Column-wise
    - Other patterns

- **2 step techniques:**
  - pixels organized in small groups
  - arrangement of the groups

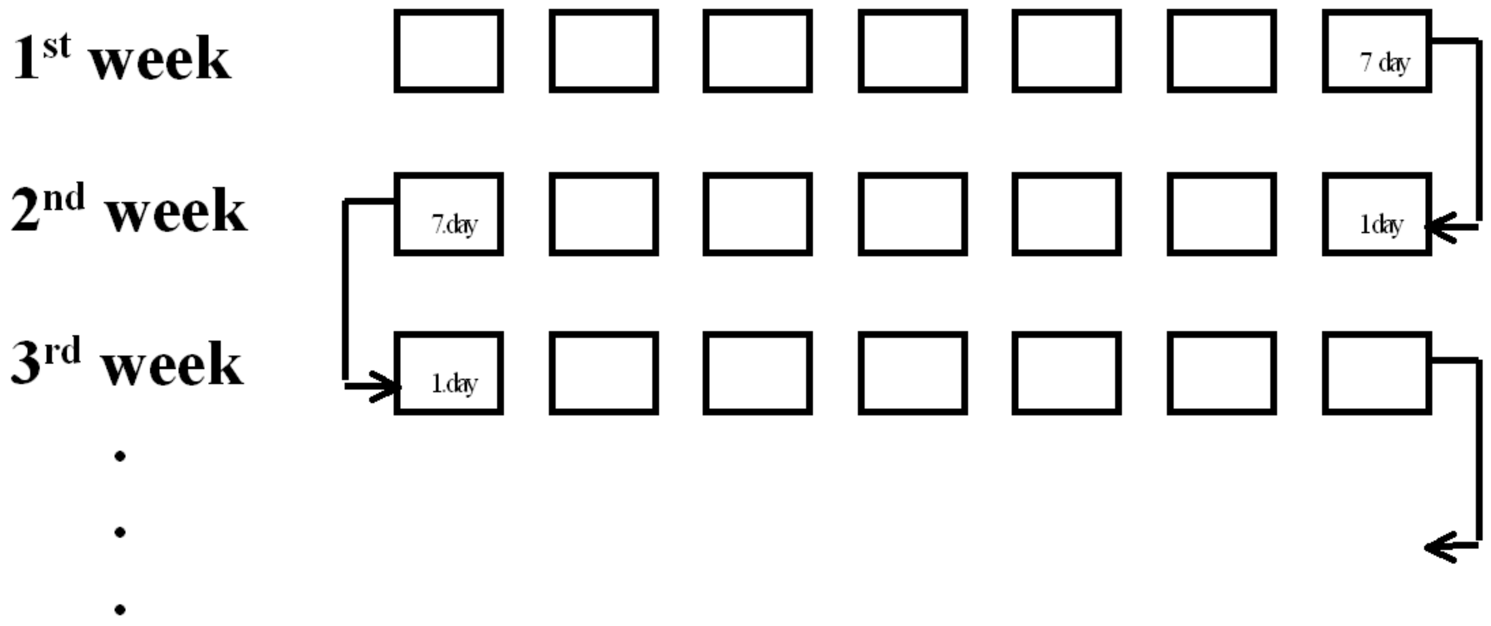
- **Recursive pattern technique (Keim, Kriegel, 1995)**
  - recursive arrangement of the groups

- **Example for recursive pattern technique:**
  A measurement station yields 5 parameters three times daily: air pressure, humidity, temperature, O_2- and CO_2 concentrations.

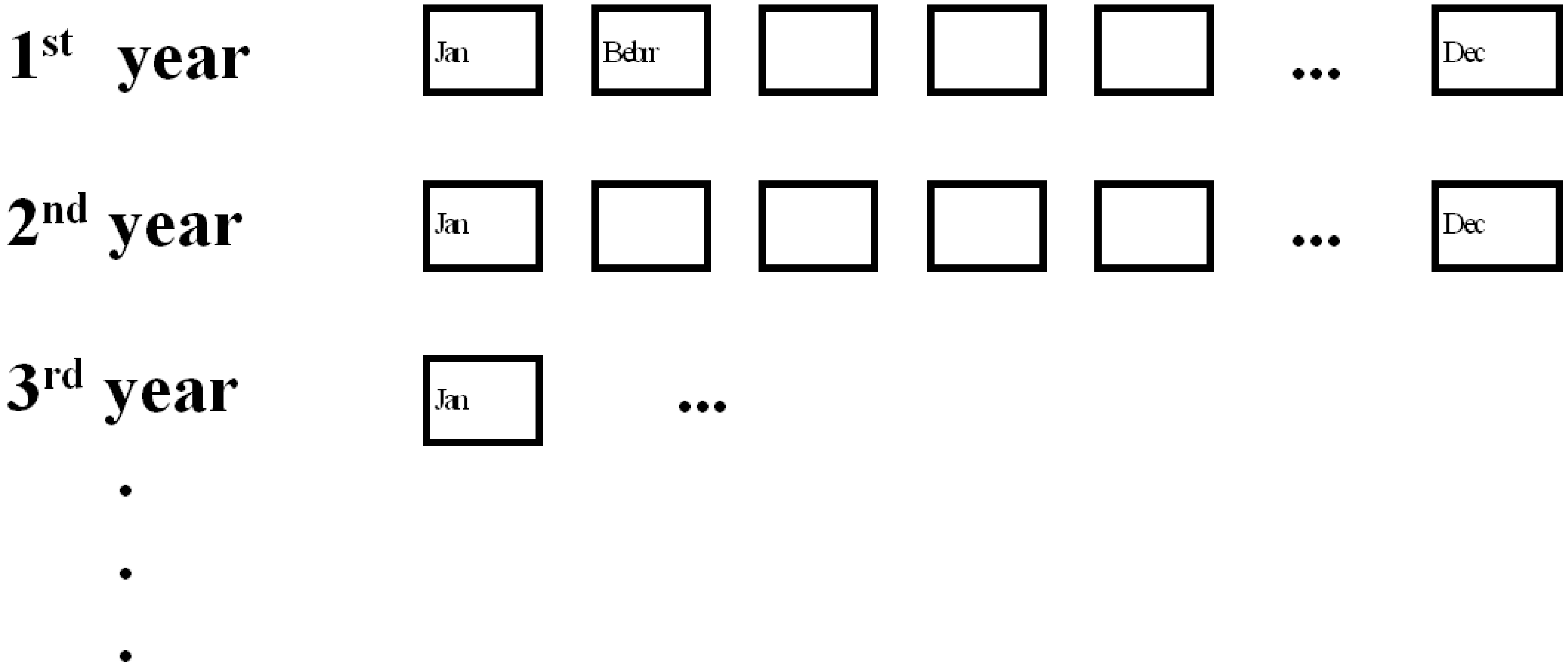
- **Pixel arrangement:**
  - 1st level: create a 3 * 5 pattern which codes the daily measures row-wise.



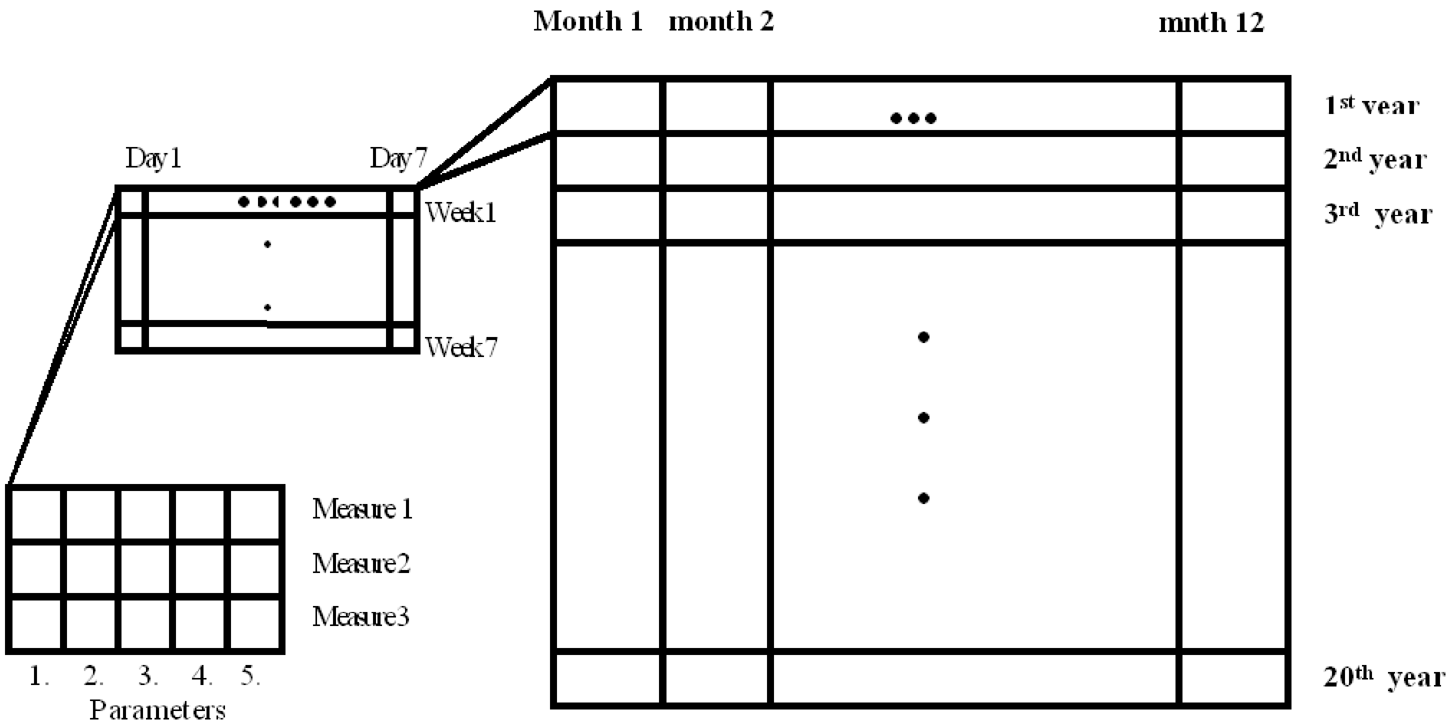
- **2nd level**: create a 5 * 7 pattern which codes measured values of a month in a weekly left-right-arrangement.



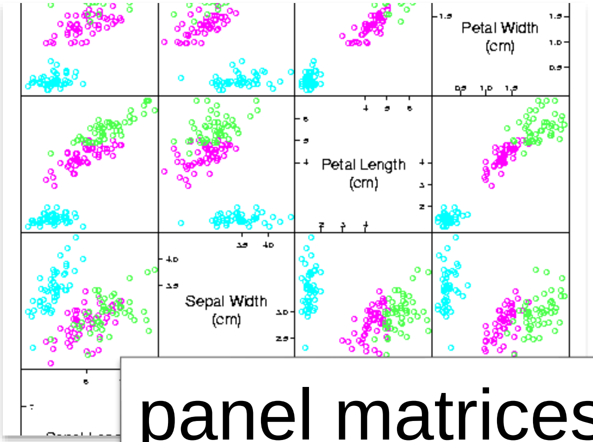
- **3rd level**: create a 20 * 12 Pattern with the measured values of 20 years; monthly in a left-right-arrangement



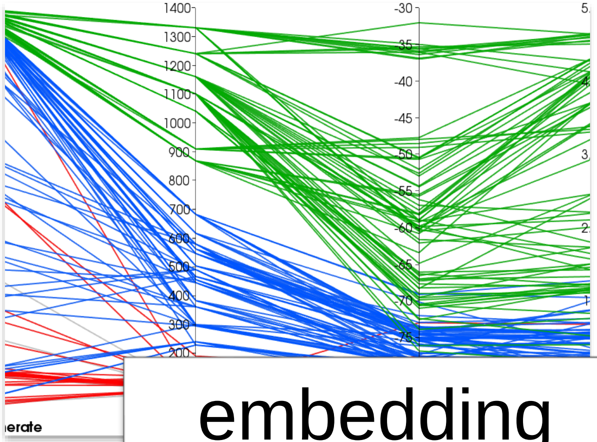
- **Result:** image of 300 * 420 pixels representing 126000 data values



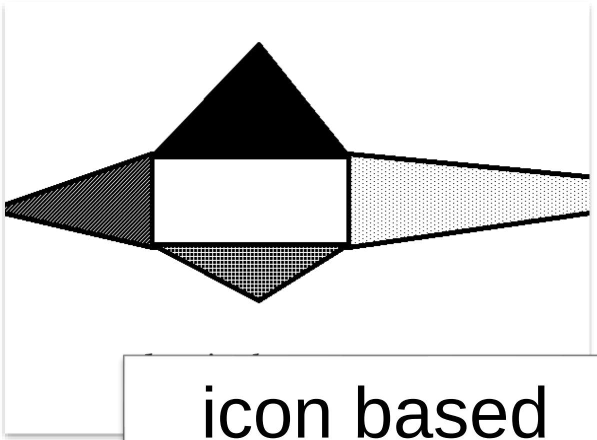
Classification of visualization methods for multiparameter data



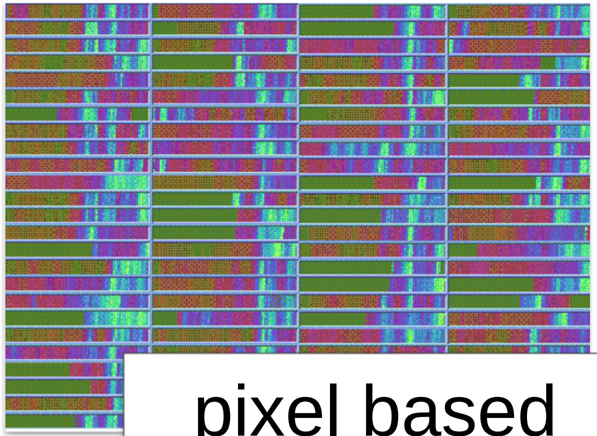
panel matrices



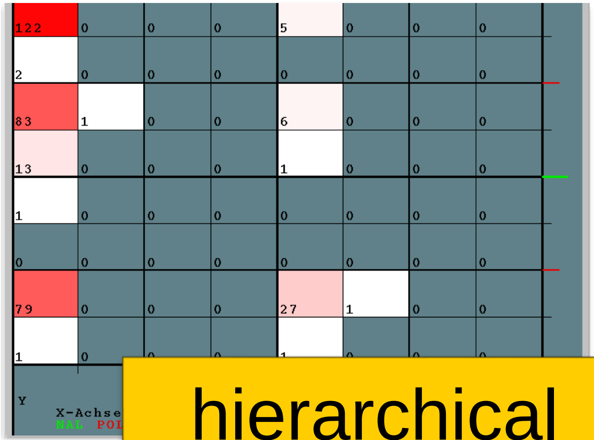
embedding



icon based



pixel based



hierarchical

Hierarchical Methods

stack dimensions into each other

only few methods in this group

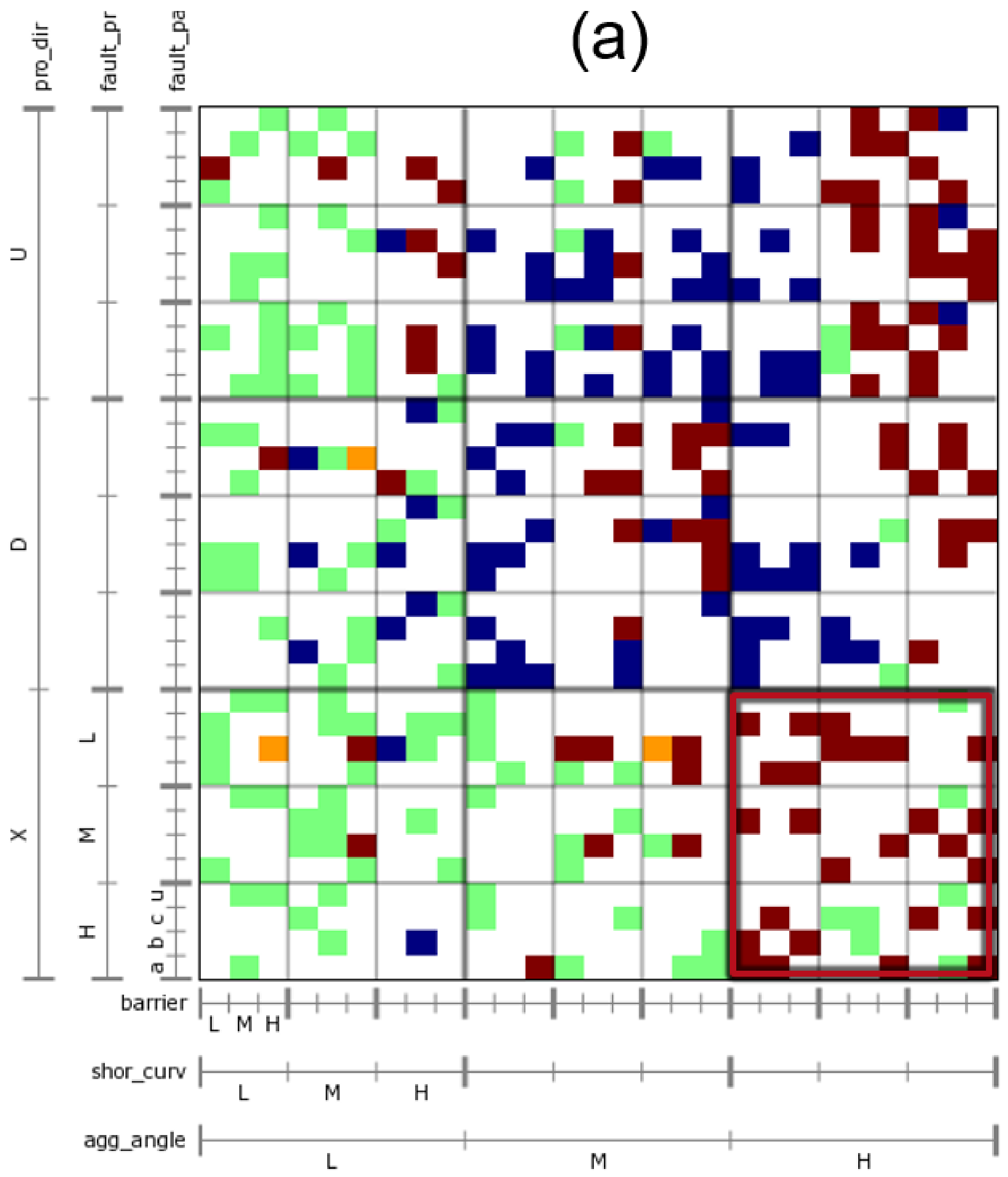


image from Jiang et al., Scenario Discovery Workflow for Robust Petroleum Reservoir Development under Uncertainty, Int. J. Uncertainty Quantification, 2016

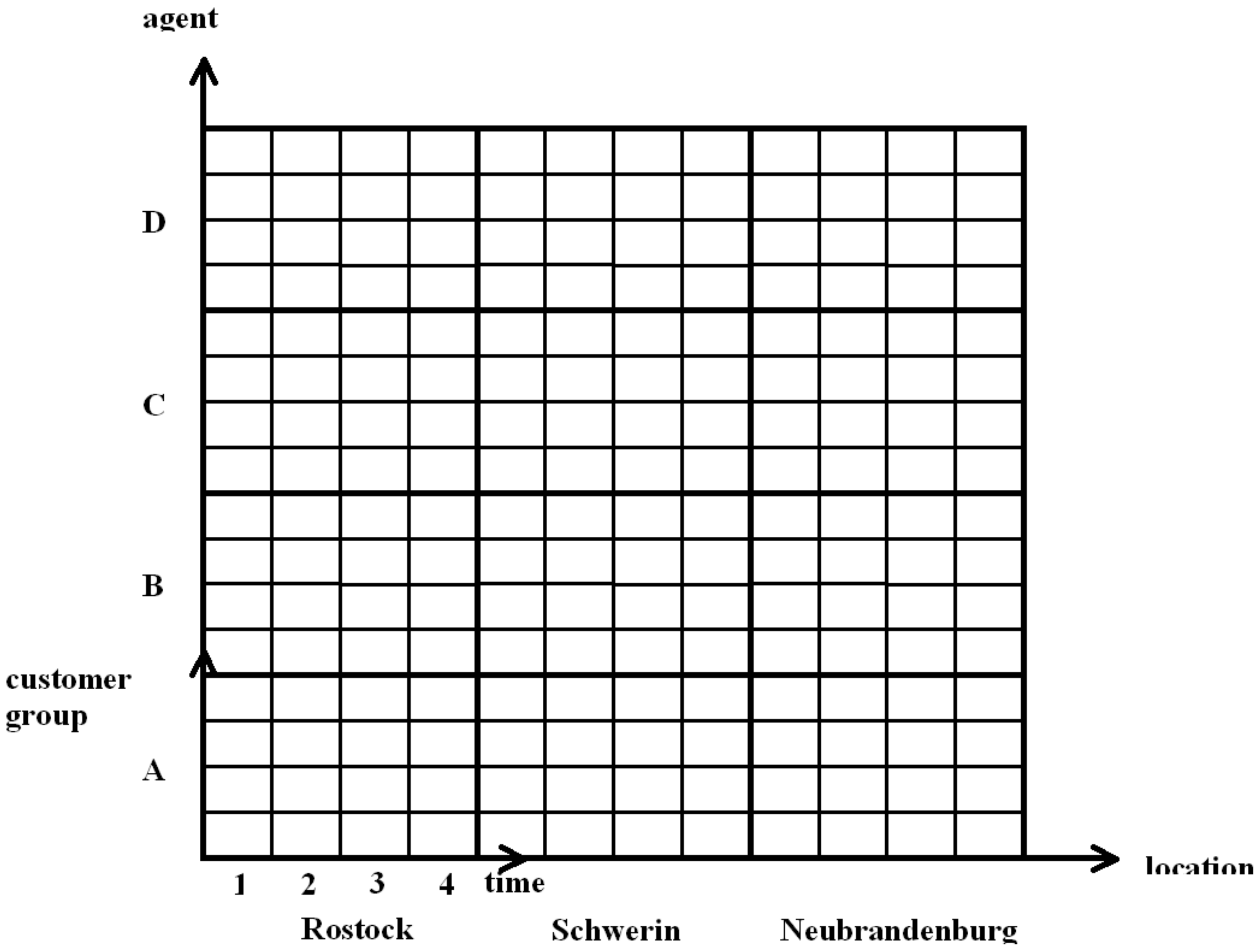
Dimensional Stacking

5-dimensional data set

- P_1=location, K_1=3
- P_2=time, K_2=4
- P_3=agent, K_3=4
- P_4=customer group, K_4=4
- P_5=sales, *quantitative*

Step 1: pair dimensions

Step 2: color code cells

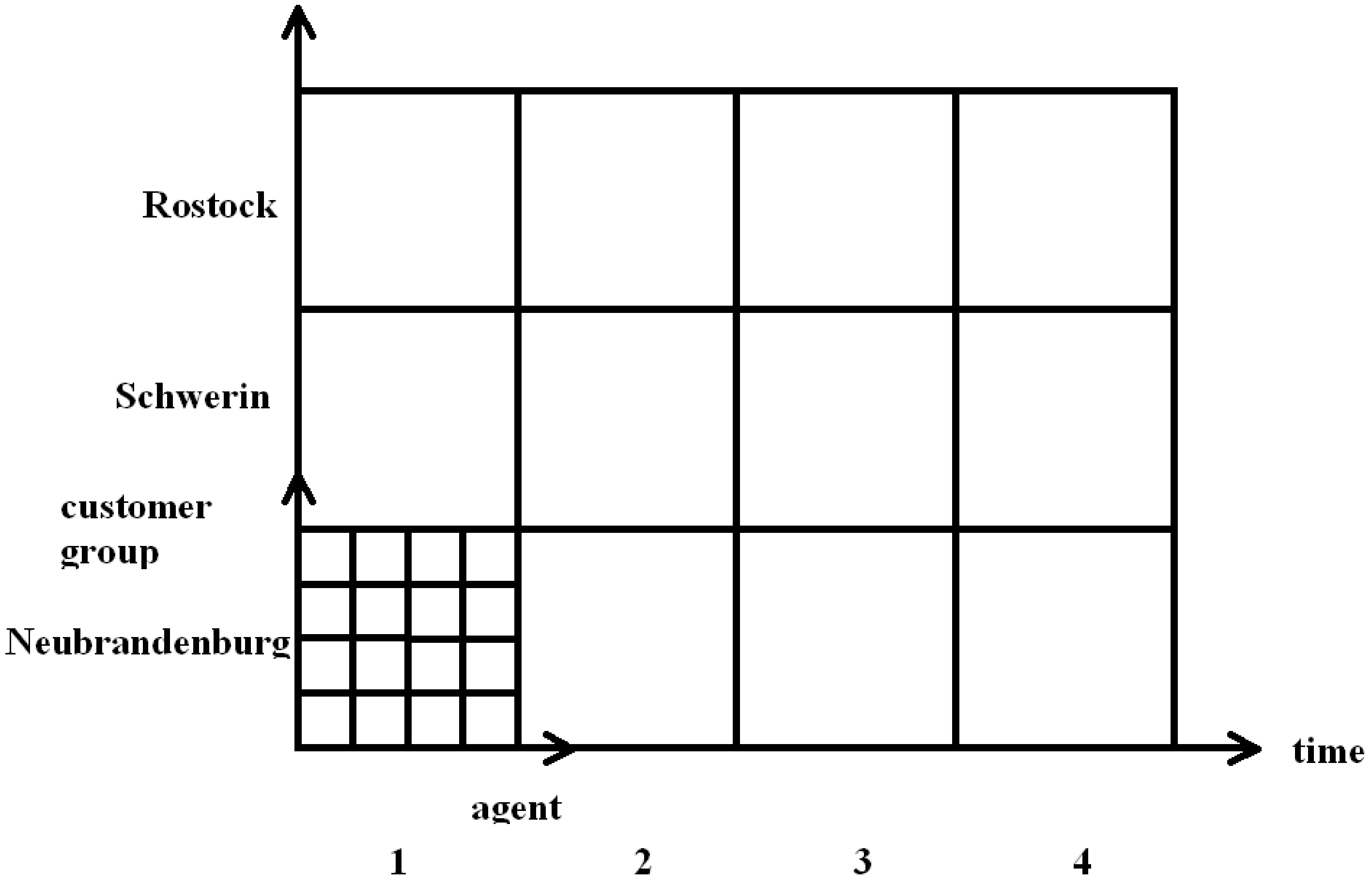


Dimensional Stacking method from Leblanc, Ward, Wittels, 1990

Dimensional Stacking

pairing is important

different pairings = different layouts



Dimensional Stacking method from Leblanc, Ward, Wittels, 1990

Dimensional stacking (Leblanc, Ward, Wittels, 1990)

starting point:

m parameters P_1 to P_m with *m* cardinal numbers K_1 to K_m
(cardinal number = number of classes in the domain of each parameter)

idea: define pairs of parameters and combine them

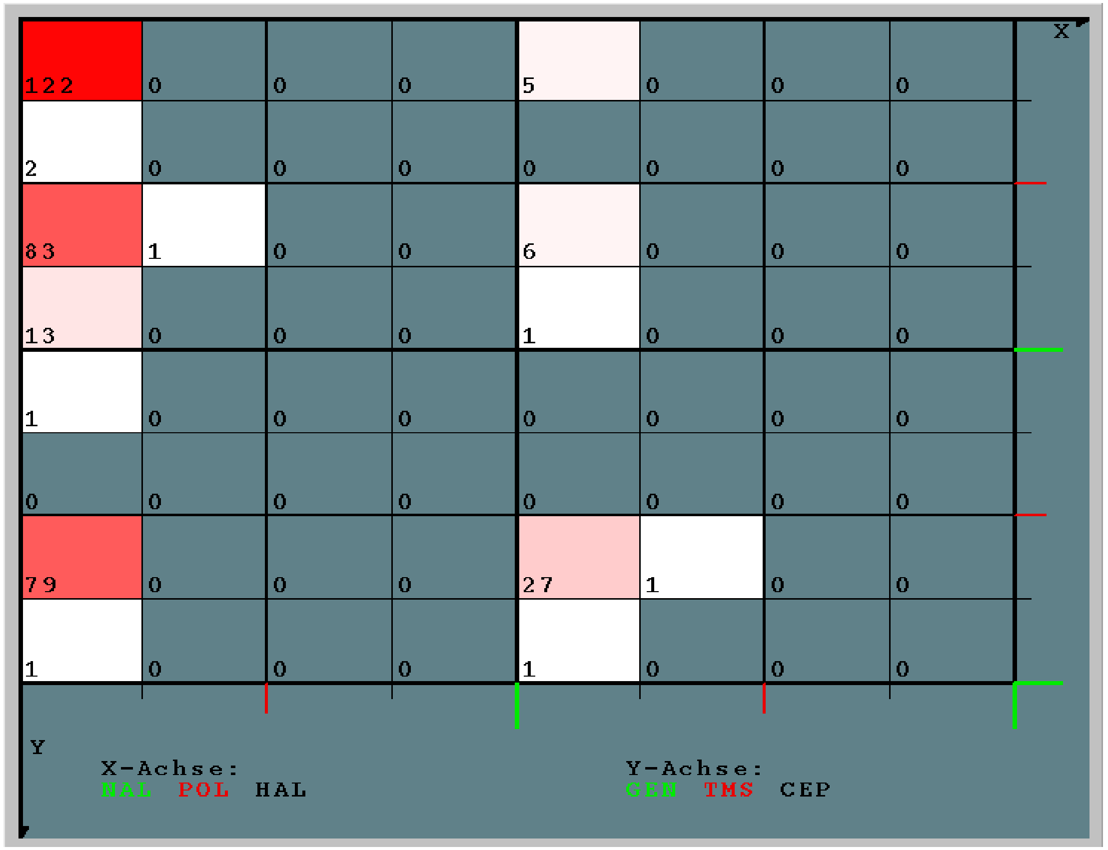
steps:

- choose the most important parameters P_i, P_j; define a 2D grid K_i x K_j;
- recursive subdivision of each grid cell using the next pairs of parameters;
- color coding of the final grid cells using:
- the value of a *dependent* parameter

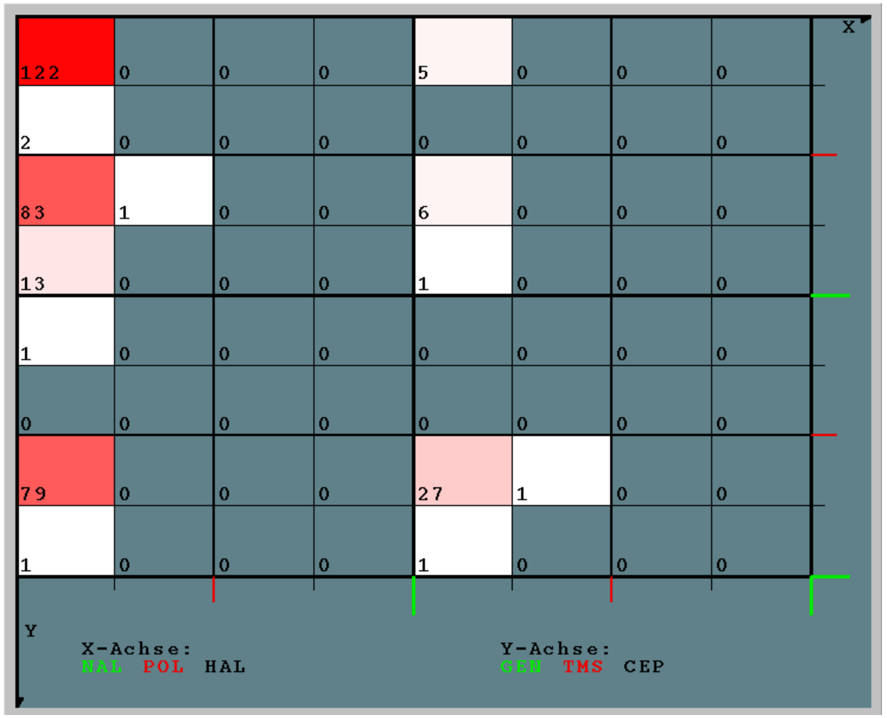
or

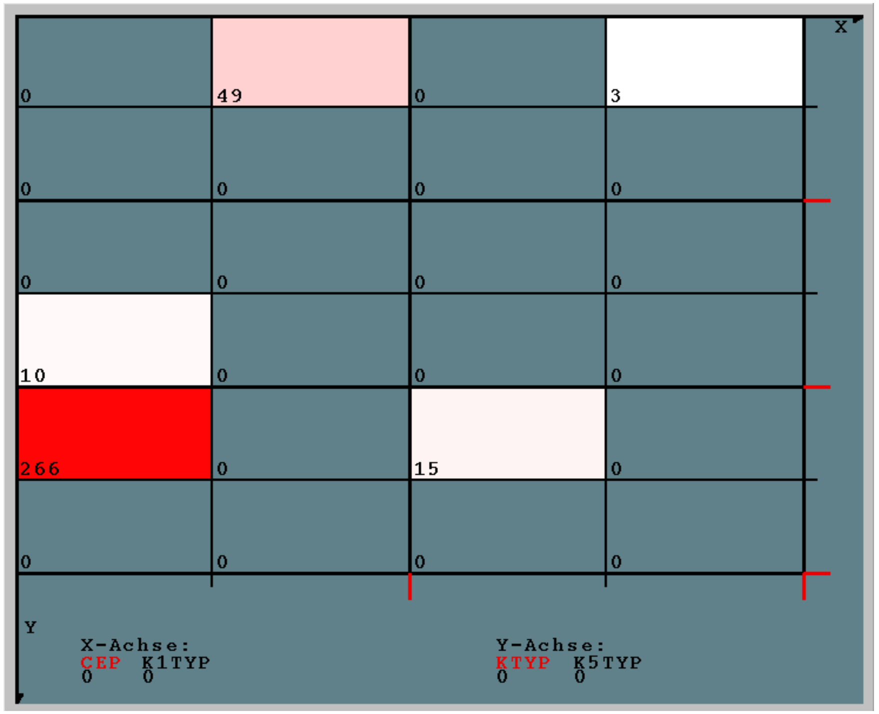
- the *frequency* of data in each grid cell

- **Example 2:**
  - microbiological data set
  - define pairs of parameters
  - color code: frequency



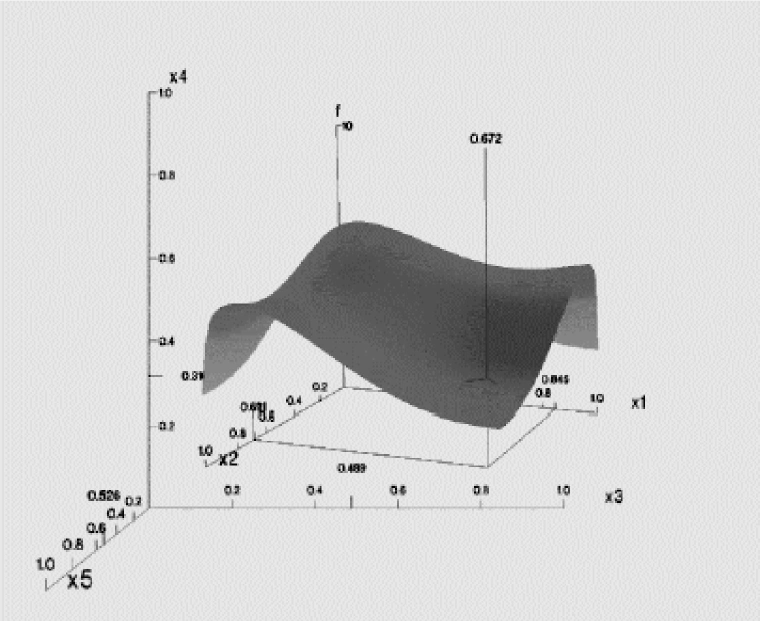
- **Problems:**
  - effective definition of the parameter pairs
  - effective partition
  - level of recursion



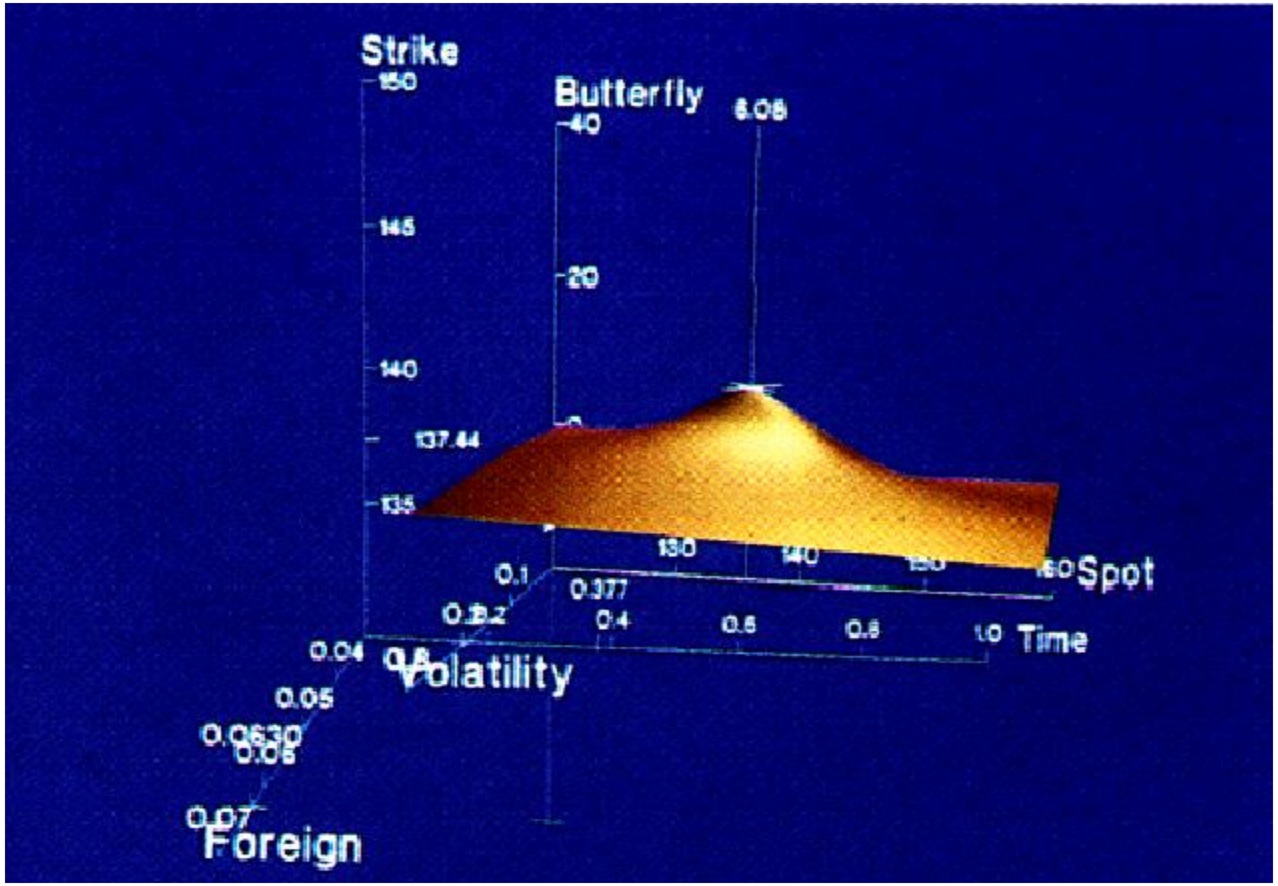


- **Worlds-within-Worlds** (Feiner, Beshers, 1990)

- **Approach similar to dimensional stacking:**
  - choose a triple of parameters; defines a 3D coordinate system;
  - choose a point in this coordinate system; define another coordinate system by choosing three more parameters at this point;
  - define triples of parameters and coordinate systems recursively;
  - visualize values in the last-defined coordinate system.



Worlds-within-Worlds (Feiner, Beshers, 1990)



Summary

- Multiparameter data: many variables, no specific prior on observation space
- Many methods exist, but only few general ones:
  - scatterplot matrices
  - parallel coordinates
  - star coordinates
- Classification of methods
  - panel matrices
  - embedding
  - icon based
  - pixel based
  - hierarchical