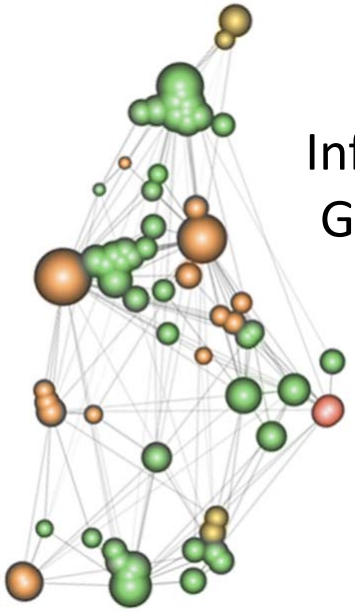


Informationsvisualisierung Geometrische Techniken [WS1011 | VO.04]

Theresia Gschwandtner
gschwandtner@ifs.tuwien.ac.at

22. 11. 2010

Content VO.04

- **Scatterplots**
- Parallel Coordinates
- Interactive Stardinates
- Influence Explorer + Prosection Matrix

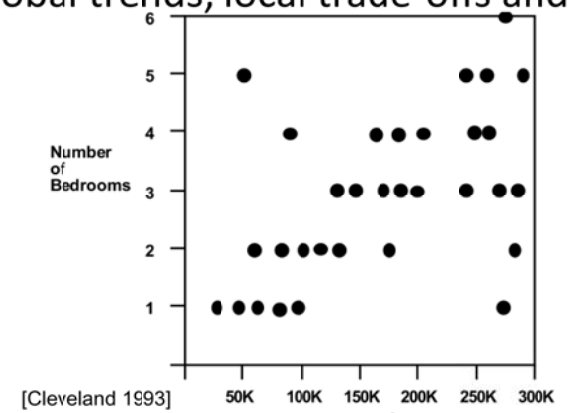
Scatterplots: Basic Idea

- Visualizes a Relation (Correlation) between two Variables X and Y e.g., weight and height
- Individual Data Points are Represented
  - in 2D
    where axes represent the variables
    X on the horizontal axis
    Y on the vertical axis
  - in 3D
  - in...

[Cleveland 1993]

Example: Scatterplot

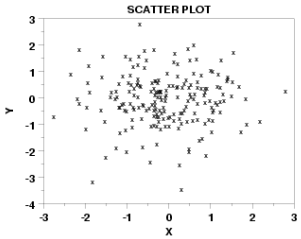
- House data:
  Price and Number of bedrooms
- User can identify global trends, local trade-offs and outliners.



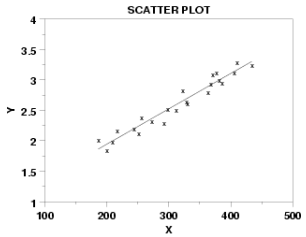
[Cleveland 1993]

Scatterplots: Relationship

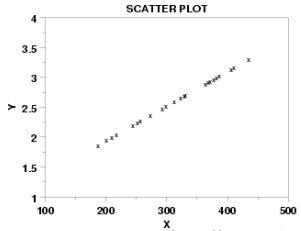
No relationship



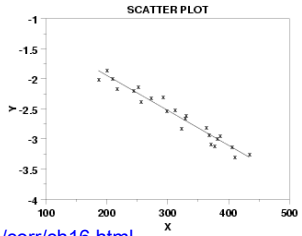
Strong linear (positive correlation)



Exact linear
(positive correlation)



Strong linear
(negative correlation)

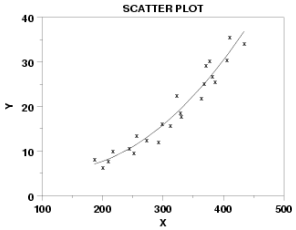


[Cleveland 1993]

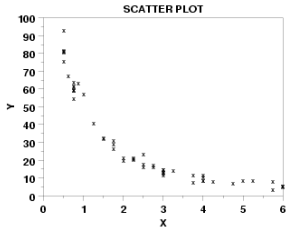
http://noppa5.pc.helsinki.fi/koe/flash/corr/ch16.html

Scatterplots: Relationship

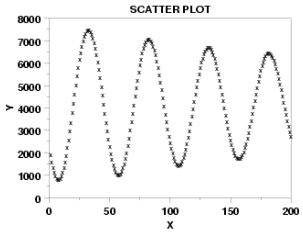
Quadratic relationship



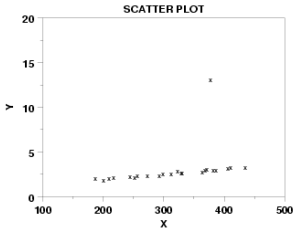
Exponential relationship



Sinusoidal relationship (damped)



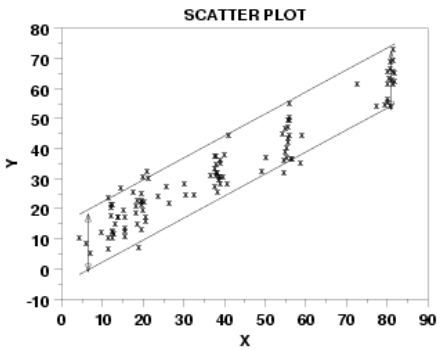
Outlier



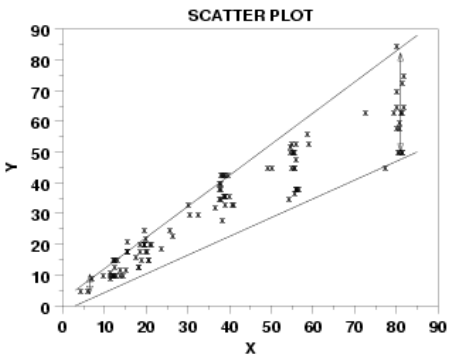
[Cleveland 1993]

Scatterplots: Relationship

Variation of Y
doesn't depend on X
(homoscedastic)



Variation of Y
does depend on X
(heteroscedastic)



[Cleveland 1993]

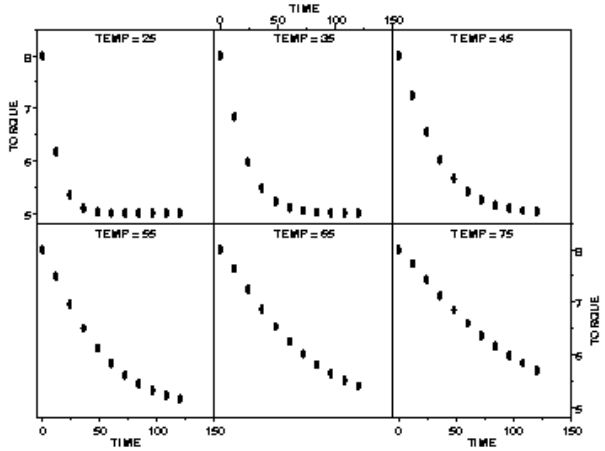
Scatterplots

More than two variables?

Scatterplot: Conditioning Plot

- One limitation of scatterplots is that it cannot show interaction effects with another variable
- Purpose: Check pairwise relationship between two variables conditional on a third variable

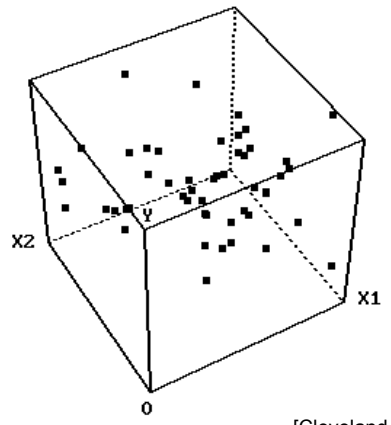
temperature:
torque (Verdrillung von Polymeren)
versus time
exponential - linear



[Cleveland 1993]

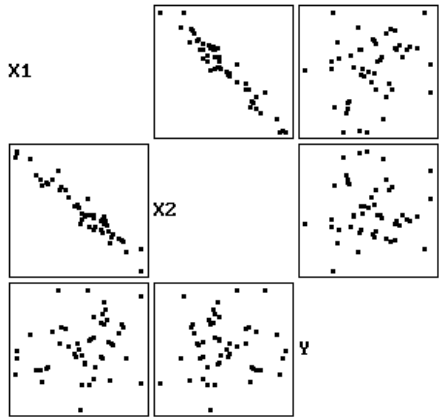
3 D Data in the Box

3 D Data Set of 50 Observations in the Box

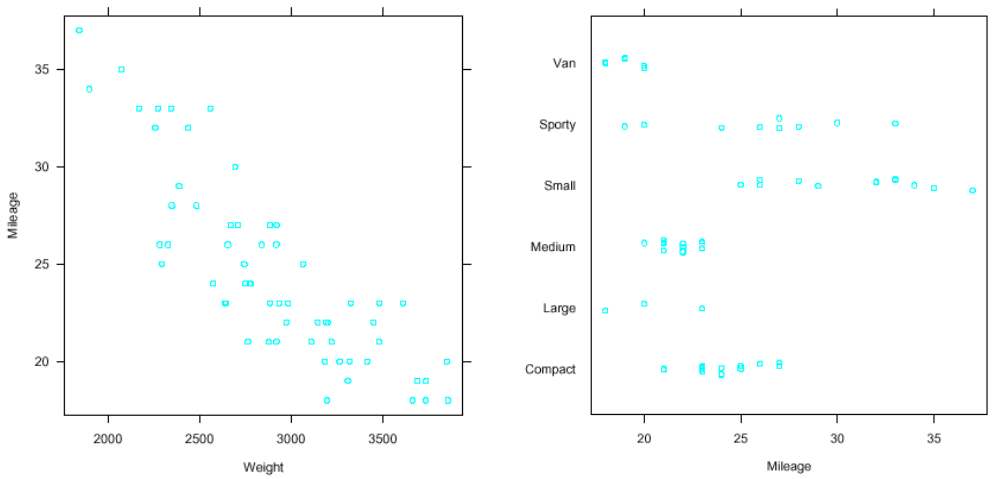


[Cleveland 1993]

Scatterplot Matrix of all pairwise Scatterplots



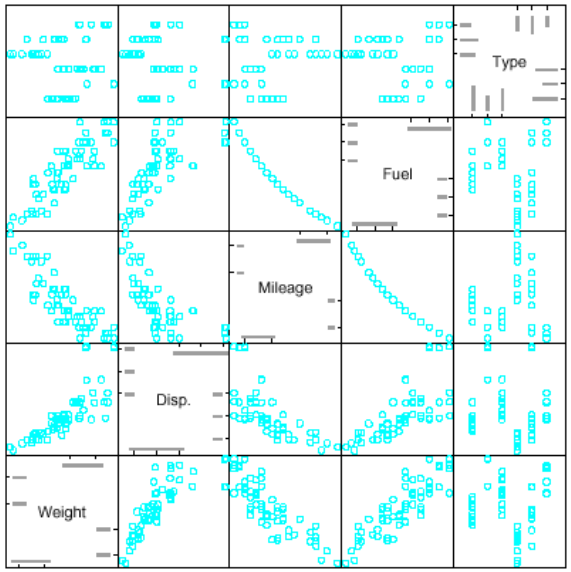
Example: Cars



[Becker & Cleveland, 1996]

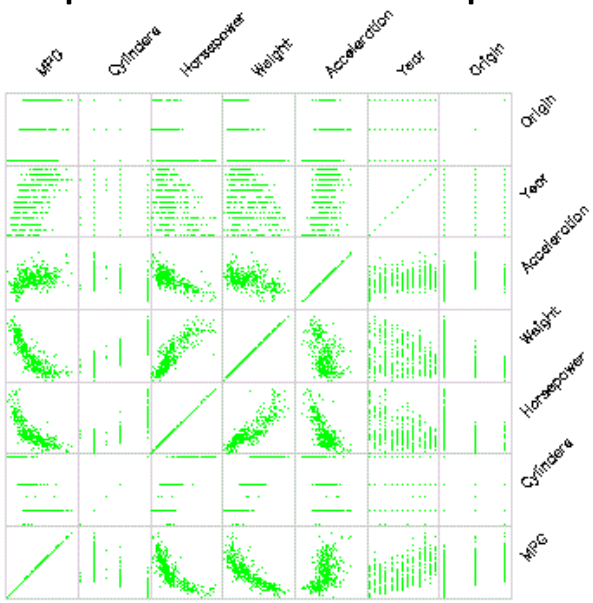
Example: m x m Scatterplots

- m x m scatterplots
  - Type (Compact, Large, Medium, Small, Sporty, Van)
  - Fuel
  - Mileage (Laufleistung)
  - Displacement (Hubraum)
  - Weight
- diagonal = same (m^2 - m)
- left -right the same (m^2 - m)/2

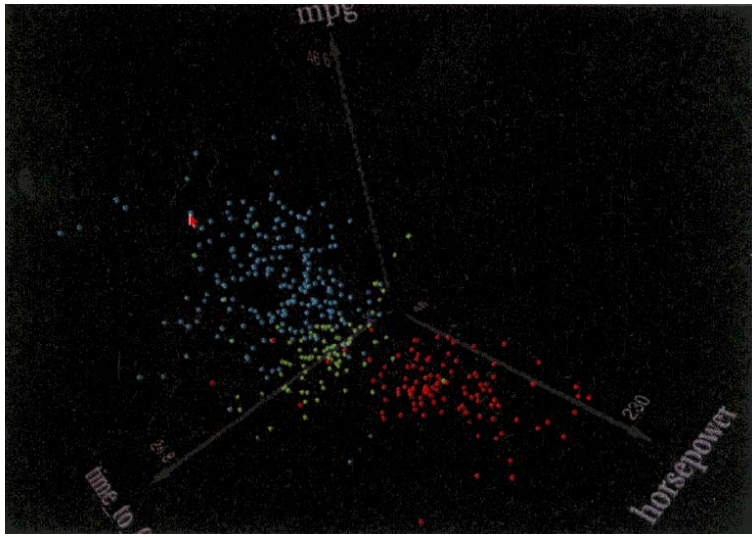


[Becker & Cleveland, 1996]

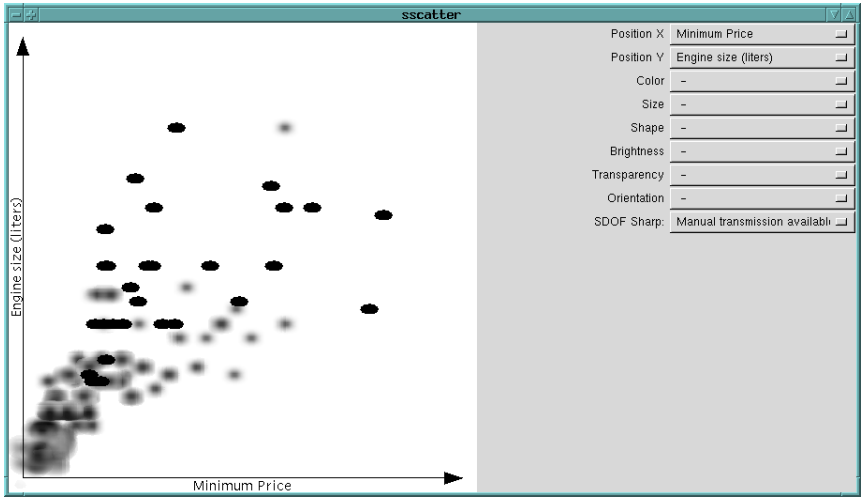
Example: Cars - Scatterplots



3 D Scatterplot plus Color



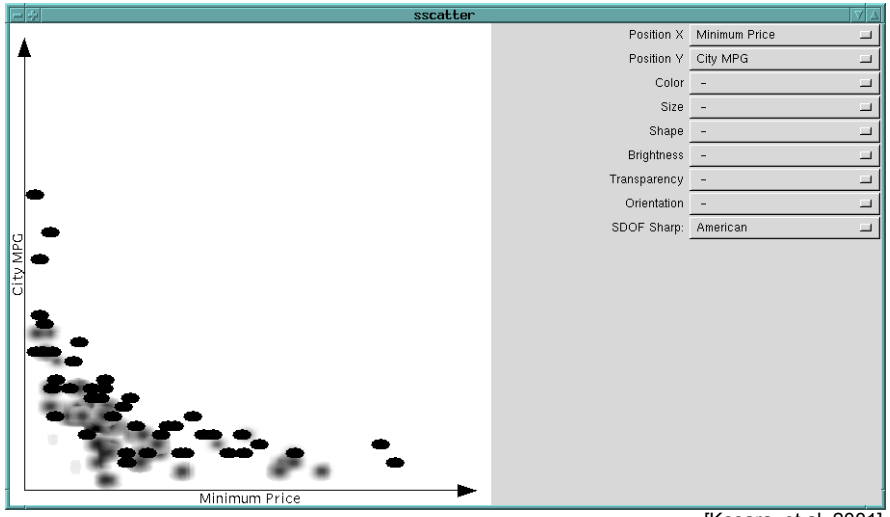
Scatterplot & SDOF (1)



[Kosara, et al. 2001]

- Depth of field (DOF) in photography: which depth range to depict sharply
- Semantic depth of field (SDOF): blur objects based on their relevance

Scatterplot & SDOF (2)



[Kosara, et al. 2001]

Summary Scatterplots

- Visualizes a relation between two variables:
  - Trends
  - Outliners
- More than 2 variables:
  - Conditioning Plot
  - 3D Scatterplot
  - Scatterplot Matrix
- Visual encoding of additional attributes:
  - Color
  - SDOF
  - Shape
  - Size

Content VO.04

- Scatterplots
- **Parallel Coordinates**
- Interactive Stardinates
- Influence Explorer + Prosection Matrix

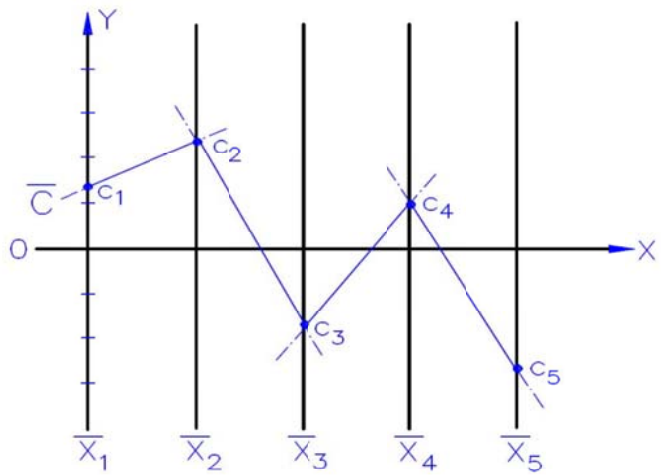
Parallel Coordinates: Basic Idea

- Assigns one Vertical Axis to each Variable
  - Evenly spaces these axes horizontally
  - Traditional Cartesian Coordinates
    All axes are mutually perpendicular
- Layout: k Parallel Axes
- Axes to [min, max]
  - Scaling individually for each variable
- Polygonal Line
  - Every data item corresponds to a polygonal line
  - Intersects each of the axes at the point corresponding to the value for the attribute

[Inselberg and Dimsdale, 1990]

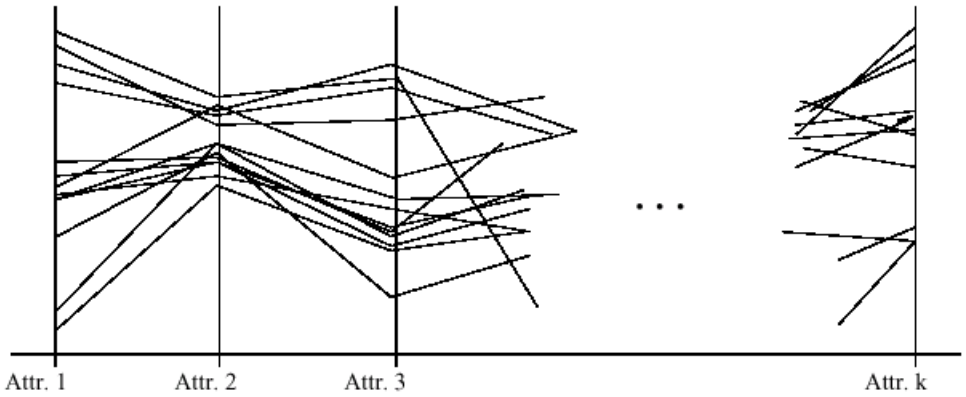
Parallel Coordinates

The point C = (c1, c2, c3, c4, c5) is represented by the polygonal line shown.



[http://www.math.tau.ac.il/~aiisreal/]

Parallel Coordinates



[Inselberg and Dimsdale, 1990]

A Detective Story

- The Dataset:

  Production data for 473 batches of a VLSI chip.

  16 process parameters:

  X1: The yield: % of produced chips that are useful

  X2: The quality of the produced chips (speed)

  X3 - X12: 10 types of defects (zero defects shown at top)

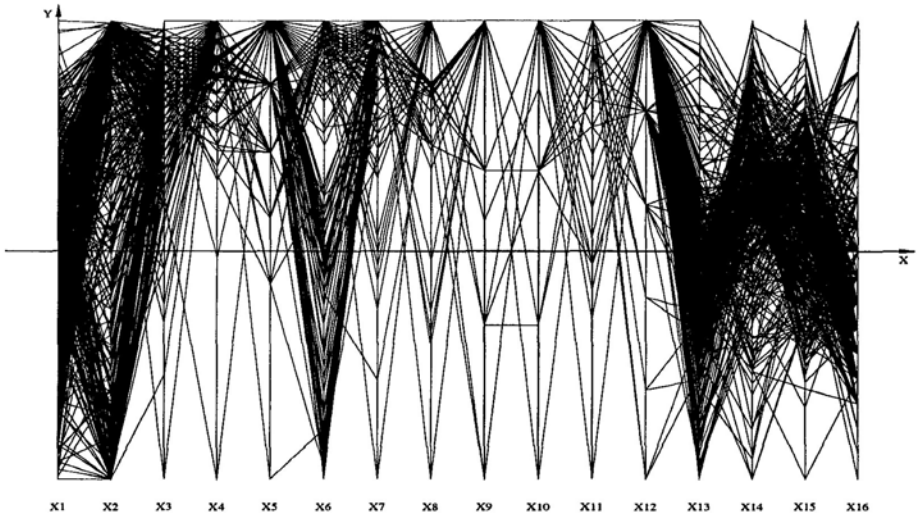
  X13 - X16: 4 physical parameters

- The Objective:

  Raise the yield (**X1**) and maintain high quality (**X2**)

[A. Inselberg, Multidimensional Detective, Proc. of IEEE Symposium on Information Visualization (InfoVis '97), 1997]

A Detective Story



[A. Inselberg: Multidimensional Detective, Proc. of IEEE Symposium on Information Visualization 1997]

Inselberg's 1st Principle: **Do not let the picture scare you!**

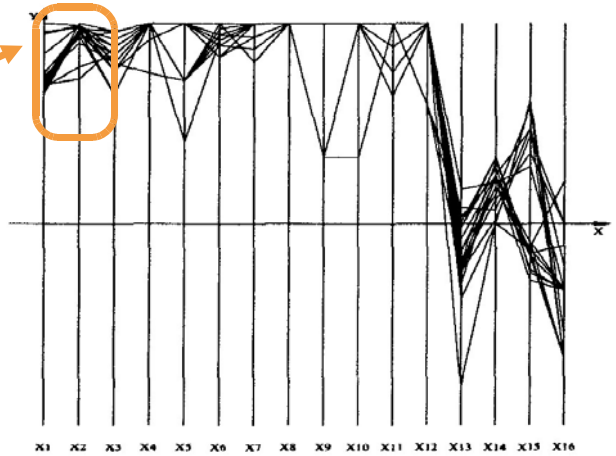
A Detective Story

- Inselberg's 2nd Principle: **Understand your objectives and use them to obtain visual cues!**
- Objectives:
  - Maximize yield (X1) and
  - Maximize quality (X2)
- Assumption: presence of defects (X3 - X12) hinder high yields and quality

  → Derived goal: **minimize defects!**

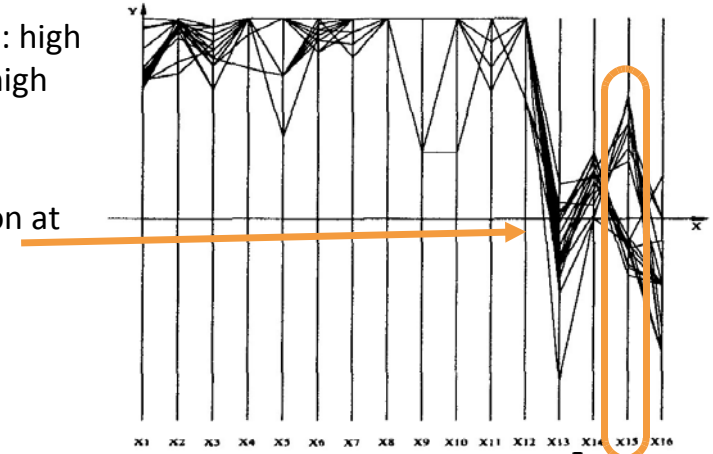
A Detective Story

- Inselberg's 3rd Principle: **Carefully scrutinize the picture!**
  - Selection: high yields + high quality



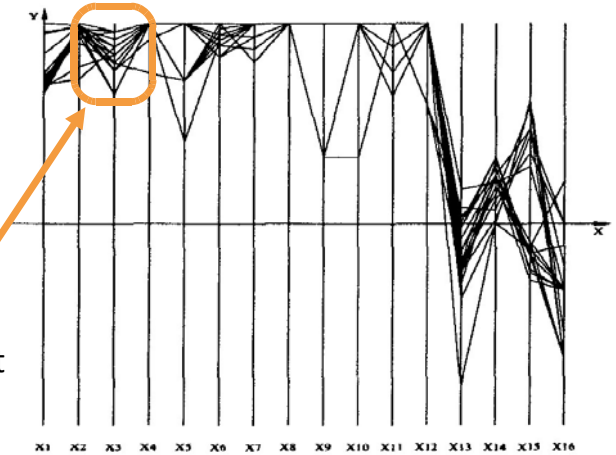
A Detective Story

- Inselberg's 3rd Principle: **Carefully scrutinize the picture!**
  - Selection: high yields + high quality
  - Notice: separation at X15



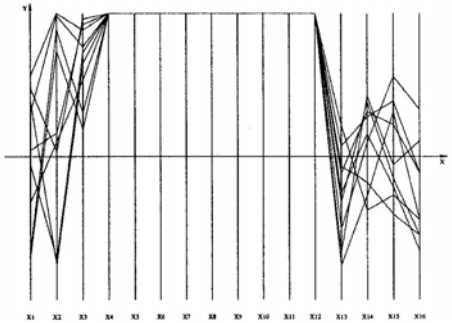
A Detective Story

- Inselberg's 3rd Principle: **Carefully scrutinize the picture!**
  - Selection: high yields + high quality
  - Notice: separation at X15
  - Notice: no batches without X3 defects

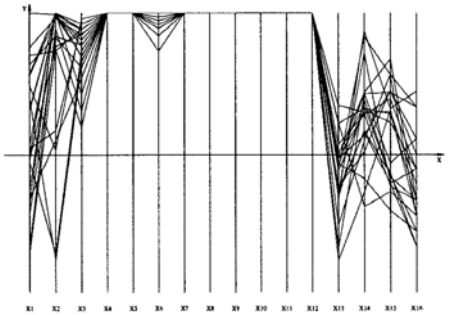


A Detective Story

No defects in **9** out of 10 defect types (only X3 defects) →**bad yields**



No defects in **8** out of 10 defect types (X3 and X6 defects) →**includes best batch**

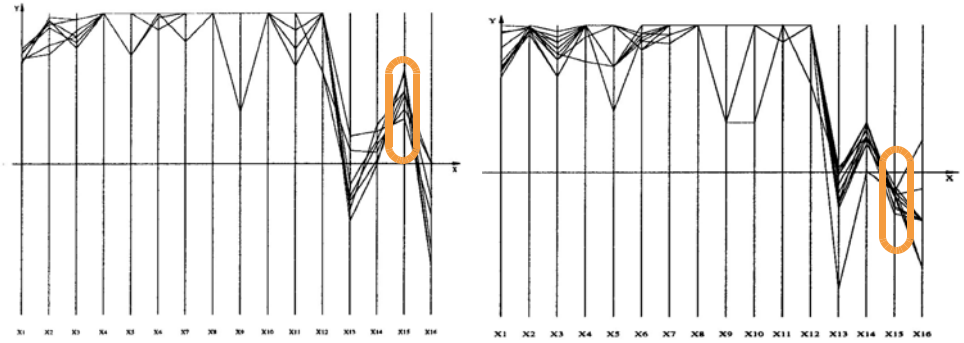


Inselberg's 4th Principle: **Don't believe your assumptions!**

A Detective Story

Looking now at X15 we see the separation is important

→ lower values of this property end up in the better yield batches

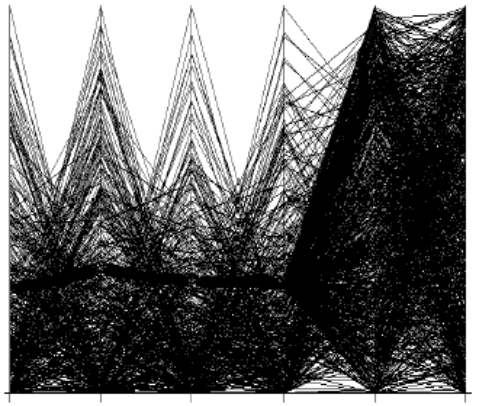


Exploration of the data → detection of characteristics of good batches

Inselberg's 5th Principle: **You can't be unlucky all the time!**

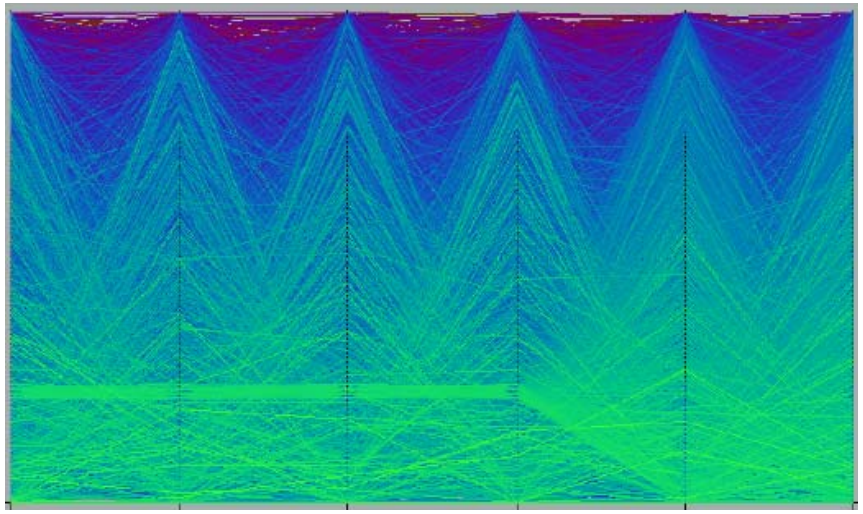
Problems with Parallel Coordinates

15.000 data items with noise



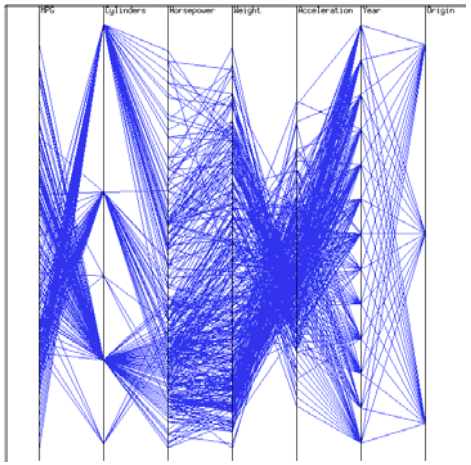
5% of the data (750 data items)

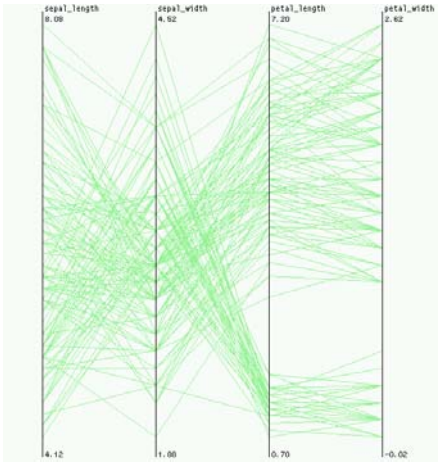
Color in Parallel Coordinates



Visualization of Correlation

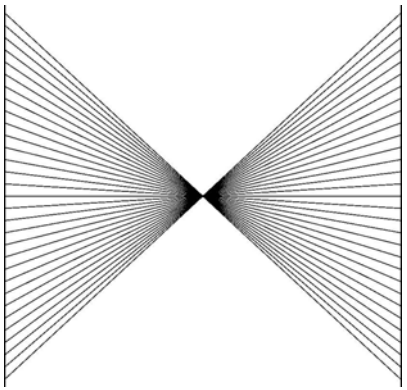
- Discover the Correlation



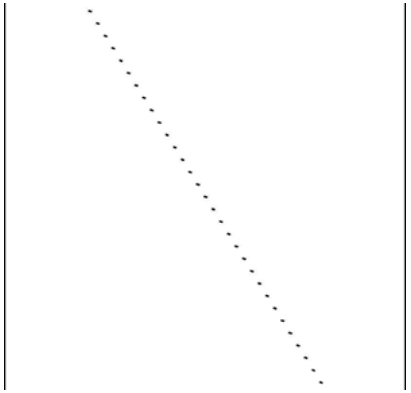


Correlation in Parallel Coordinates

Parallel Coordinates



Scatterplot

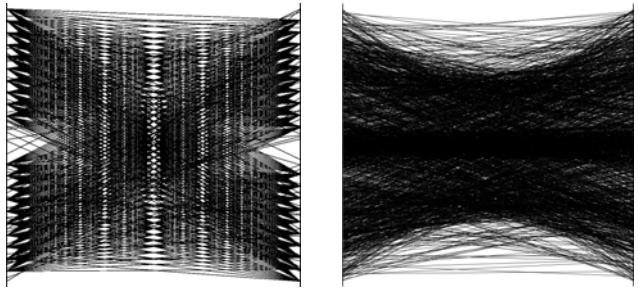


[Yuan et al., 2009: Scattering Points in Parallel Coordinates]

Problems with Parallel Coordinates

[Yuan et al., 2009: Scattering Points in Parallel Coordinates]

Correlations?

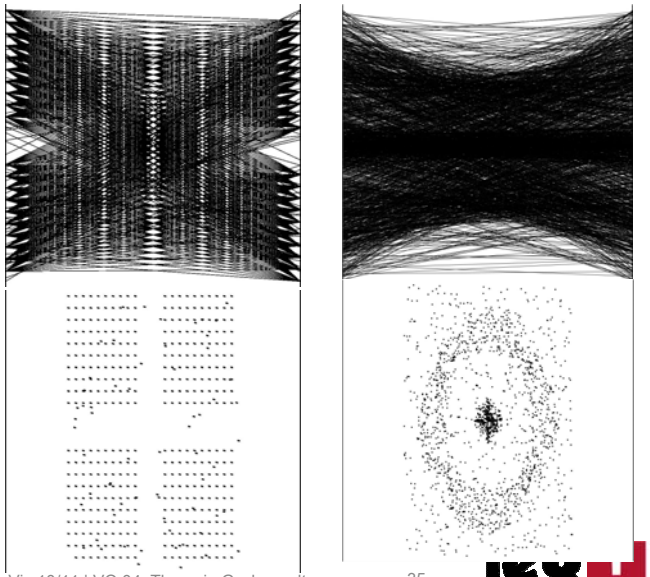


Problems with Parallel Coordinates

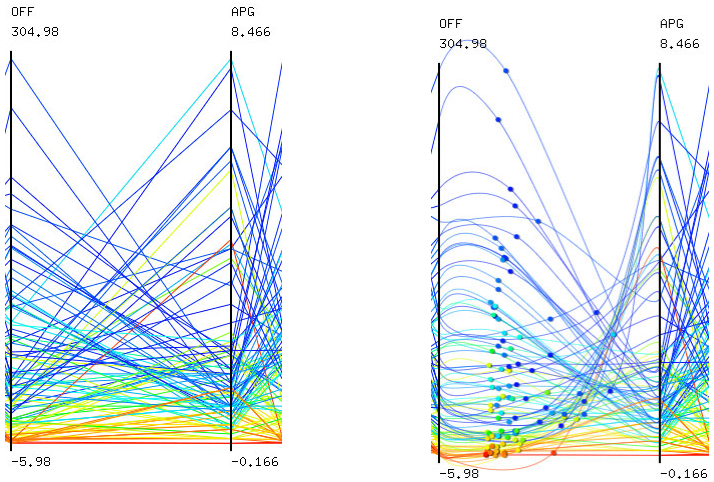
[Yuan et al., 2009: Scattering Points in Parallel Coordinates]

Correlations?

Scatterplots:



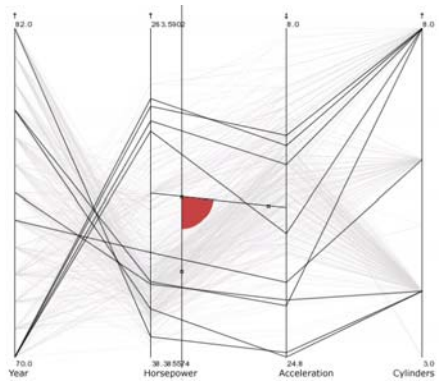
Scattering Points in Parallel Coordinates



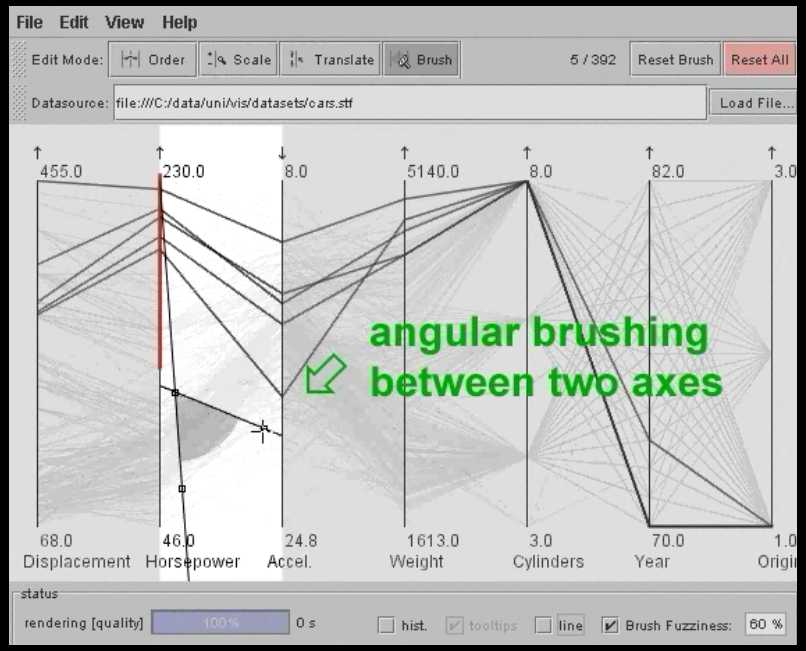
[Yuan et al., 2009: Scattering Points in Parallel Coordinates]

Common Extensions to Parallel Coordinates

- Brushing Extension
  - angular brushing
  - smooth brushing
- General Extension
  - histogram overlays
  - detail on demand
  - flexible layout (removing/adding axes, reordering, flipping,...)



[Hauser, Ledermann, Doleisch, 2002]



Angular Brushing of Extended Parallel Coordinates [Hauser, Ledermann, Doleisch, 2002]

→ Watch: http://old.vrvis.at/via/research/ang-brush/ (QuickTime, 10.2 MB, 00:04:30)

SpringView

- SpringView: Cooperation of Radviz and Parallel Coordinates for View Optimization and Clutter Reduction

[Bertini 2005] in: Proceedings of the Third International Conference on Coordinated & Multiple Views in Exploratory Visualization (CMV'05)

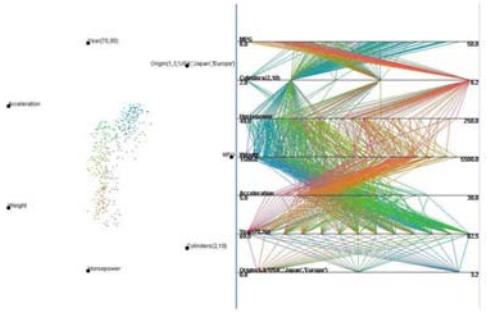


Figure 3. Color map based on radviz presentation, applied to parallel coordinates

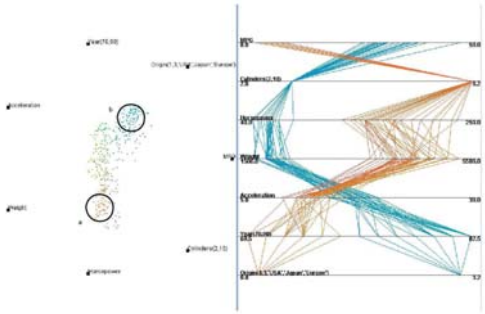
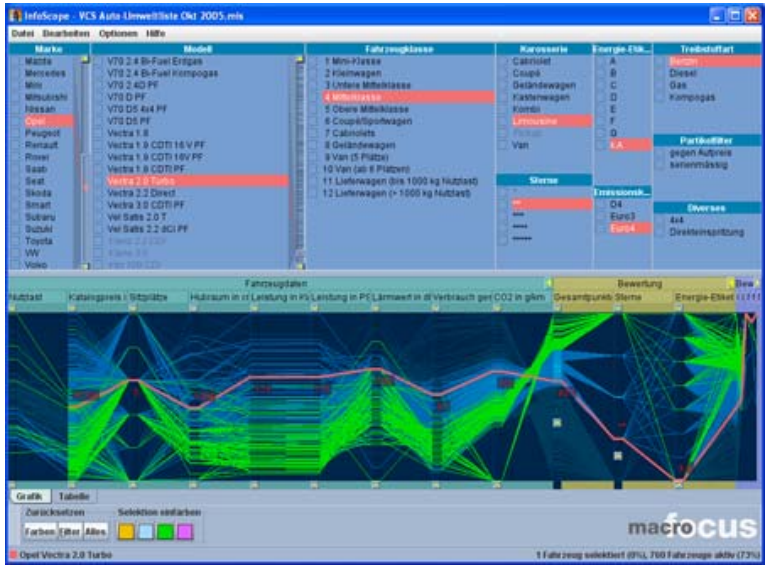


Figure 4. Selection of distinct clusters using coloring. a) powerful cars b) low power cars

Parallel Coordinates: Software

- XmdvTool (free)
  `http://davis.wpi.edu/~xmdv/vis_parcoord.html`
- InfoScope (free limited version)
  `http://www.macrofocus.com/public/products/infoscope.html`

InfoScope



→ Play around: http://www.macrofocus.com/public/products/infoscope.html

InfoScope

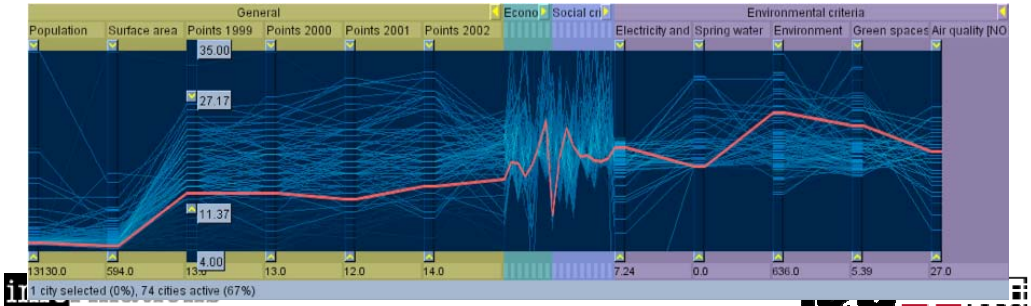
A highly interactive tool with the following key characteristics:

- Overview of global relationships between objects: Multiple views show the different aspects
- All views are highly interactive and tightly linked →interactions in one view are reflected in all the others
- Exploration: selection, details on demand (specific numeric values of attributes), dynamic queries by using range sliders
- Take advantage of the human perceptual abilities to support the identification of outliers, patterns, and anomalies

InfoScope: Multiple View Visualization

- **overview & detail**
  - Selcting an item in the overview navigates the detail view to the corresponding details.
  - Items are represented visually smaller in the overview. This provides context and allows direct access to details.

[Brodbeck & Girardin 2003]



Parallel Coordinates Benefits & Limitations

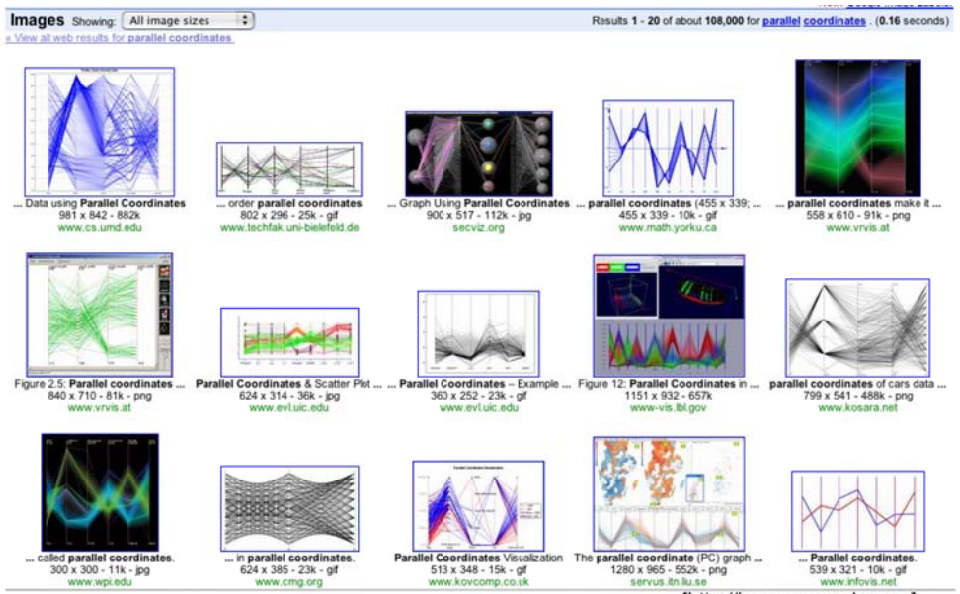
+ Represent data greater than three dimensions
+ Opportunities for human pattern recognition
+ Flexibility: each coordinate can be individually scaled
+ Zooming in or out: effectively brushing out or eliminating portions of the data set

- As the number of dimensions increases, the axes come closer to each other, making it more difficult to perceive patterns
- Huge sets of data
- Ordering of axes
- Representation of complex relations

Summary Parallel Coordinates

- Representation:
  - k-dimensional data items on k Parallel Axes
  - each data item (point in k-dimensional space) corresponds to a polygonal line
- Efficiency:
  - Visual exploration of complex multidimensional data
- Interactions:
  - Brushing (angular, smooth)
  - Detail on demand
  - Flexible layout (e.g., removing, reordering, flipping axes)
- Combinations with other visualization methods:
  - Histogram overlays
  - Scatterplots [Yuan et al., 2009: Scattering Points in Parallel Coordinates]
  - SpringView: combination of Parallel Coordinates and Radviz
  - InfoScope: multiple, interaktive, linked views (PC, maps, tables)

Parallel Coordinates Everywhere

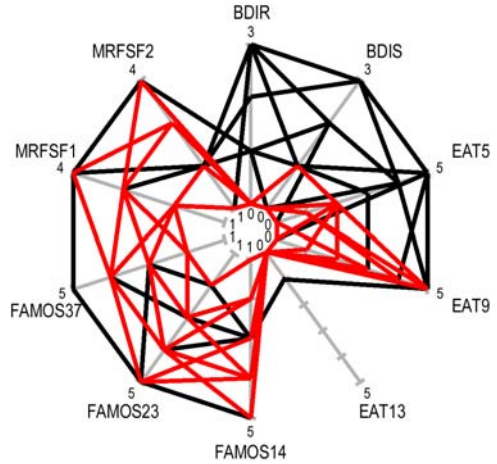


[http://images.google.com]

Content VO.04

- Scatterplots
- Parallel Coordinates
- **Interactive Stardinates**
- Influence Explorer + Prosection Matrix

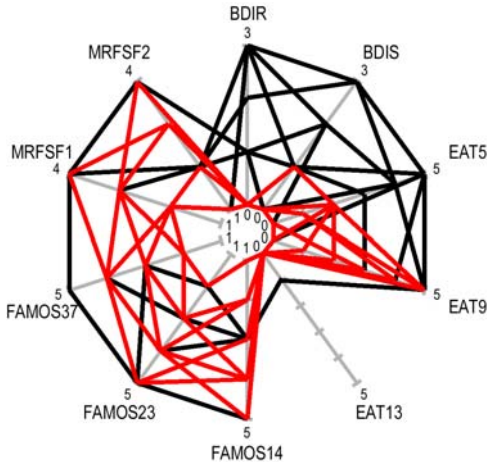
The Interactive Stardinates



[Lanzenberger, 2003]

The Interactive Stardinates

- ‘Stars and Coordinates‘
- Axes, scales, labels
- Data lines, data bundles
- Pre-attentive features:
  - Shape
  - Size
  - Relative position
  - Diversity and accumulation of lines



[Lanzenberger, 2003]

Different Levels of Detail

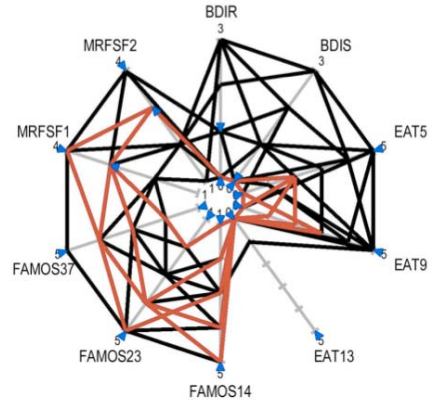
Query slider

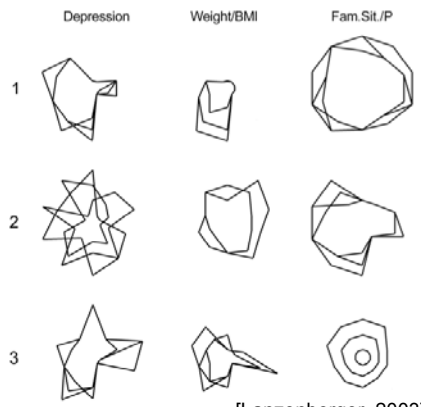
for Interactive exploration

➔ Create / investigate hypotheses

Small multiples

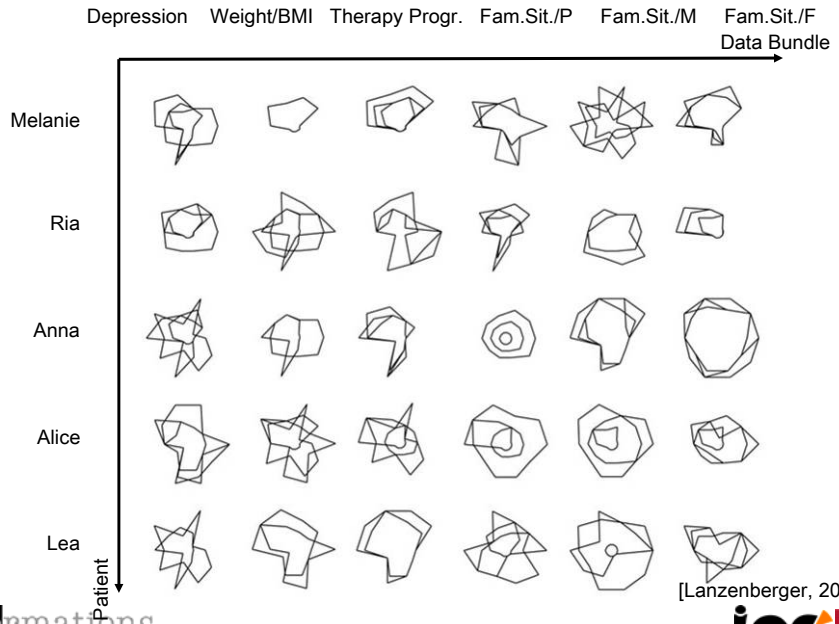
for overview visualization





[Lanzenberger, 2003]

The Interactive Stardinates



[Lanzenberger, 2003]

Interactions

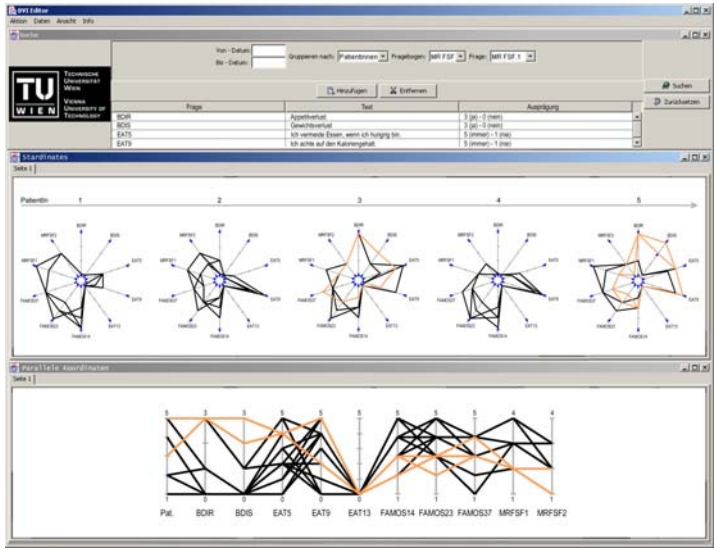
- Exploration, Analysis, and Understanding of Meaning

  A process of interactions on different levels of details
- Graphical Object Operation Scheme

  Manipulating:
  - the Axes
  - the Data Line
  - one Stardinate
  - all Stardinates at once

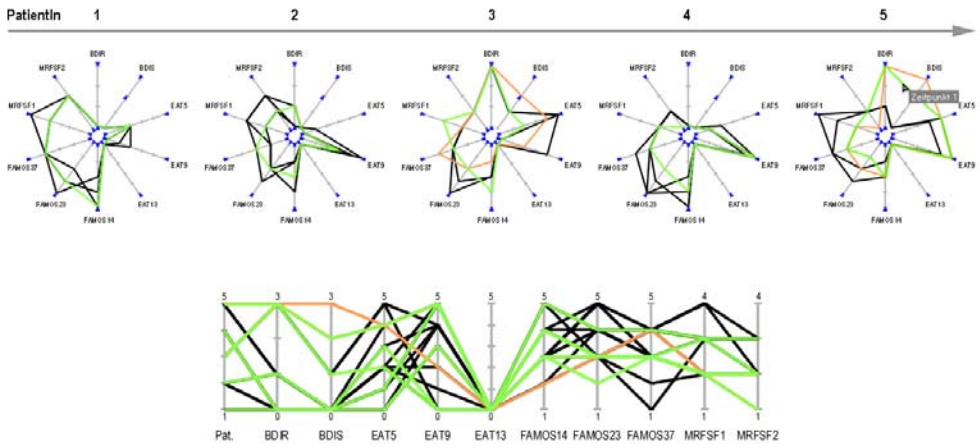
[Lanzenberger, 2003]

LinkStar



[Lanzenberger, 2003]

Linking & Brushing



[Lanzenberger, 2003]

Stardinates: Benefits & Limitations

+ Visualize overviews very effectively (differences, similarities)
+ Capable of visualizing data of high complexity
+ Human cognition of pre-attentive features: distinctive shapes, size, relative position, diversity and accumulation of lines
+ Decompose complexity
+ Allow for different levels of details with consistent visualization metaphors

- Data lines cover each other
- Limited number of dimensions
- Limited number of Stardinates
- Arrangement of axes

Summary Stardinates

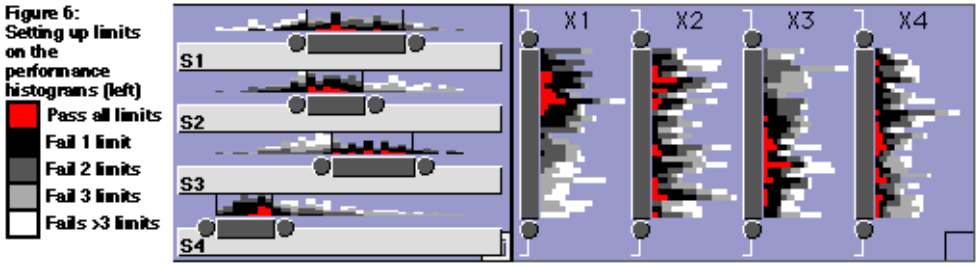
- Representation:
  - Similar to Parallel Coordinates but axes are arranged to form a star
- Efficiancy:
  - Overviews of data of high complexity
- Interactions:
  - Query slider
  - Brushing
  - Different levels of detail (detailed stardinate, small multiples)
- Combinations with other visualization methods:
  - LinkStar: combination of Stardinates and Parallel Coordinates

Content VO.04

- Scatterplots
- Parallel Coordinates
- Interactive Stardinates
- **Influence Explorer + Prosection Matrix**

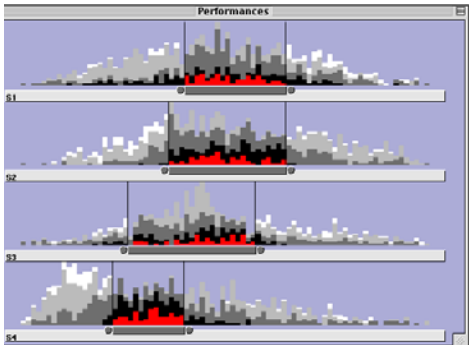
Influence Explorer

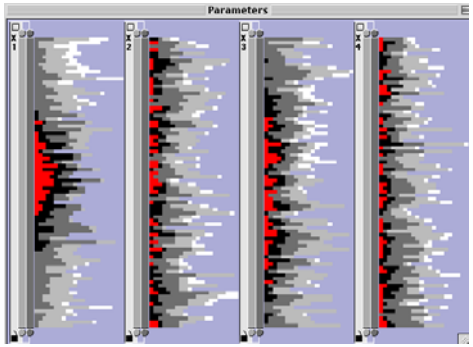
- Application Domain: Engineering Design (e.g., light bulb designs)
- Precalculated data: mathematical models of the relationship between parameters (physical nature) of an artifact and it's performances
- Histograms: Performances (e.g., bulb's brightness, lifetime) left, parameters (e.g., number of coils in filament, thickness of filament) right
- Mission: **Optimize** the artifact 's performance by specifying the tolerance range of the parameter variables



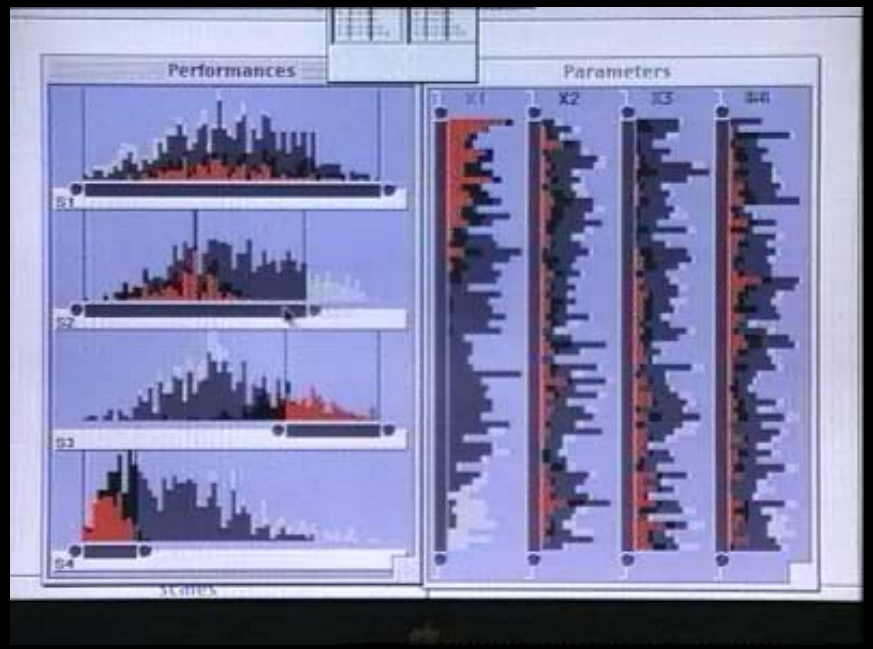
[Tweedie et al. 1995/1996]

Influence Explorer





[Tweedie et al. 1995/1996]



Influence Explorer [Tweedie et al. 1995/1996] Externalizing Abstract Mathematical Model

→ Watch: http://ieg.ifs.tuwien.ac.at/~gschwand/teaching/infovis_vo_ws09/ (QuickTime, 67 MB, 00:06:44)

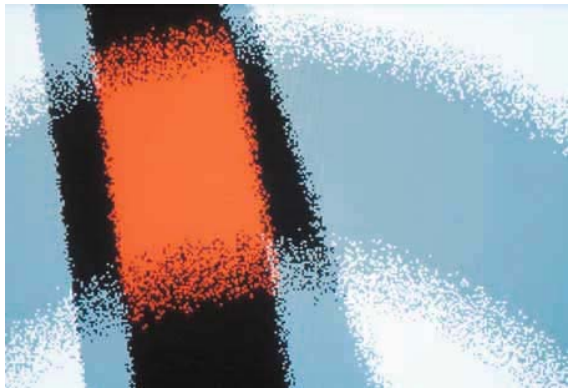
Influence Explorer

Exploration:

- Moving sliders (defining limits) -> identification of „trade-offs“
- Detail on demand per mouse-click
- Circle on the histogram scale indicates the mean of currently selected designs
- Connection of all values of the same design -> parallel coordinates
- Color coding of suitable designs

The Prosection Matrix

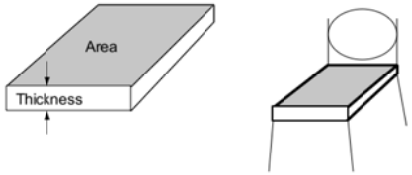
- Alternative perspective of the same precalculated data
- Prosection : a Projection of a section

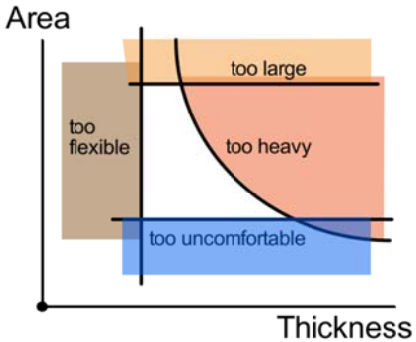


[Tweedie & Spence 1998; Spence 2001]

The Prosection Matrix

- Design of a chair seat
- A design is represented by a point in Area-Thickness space
- Various performance limits restrict the range of possible designs





[Tweedie & Spence 1998; Spence 2001]

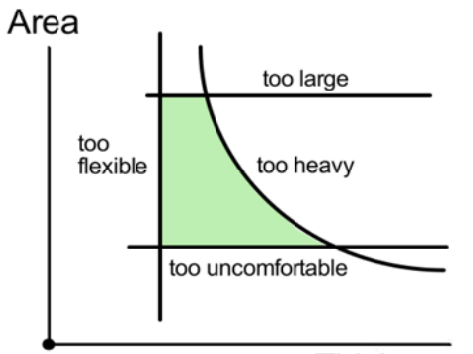
The Prosection Matrix

Problem:

- We don’t know where the green area is located
- Moreover, there are typically many parameters (not 2) and many performance limits

Solution?

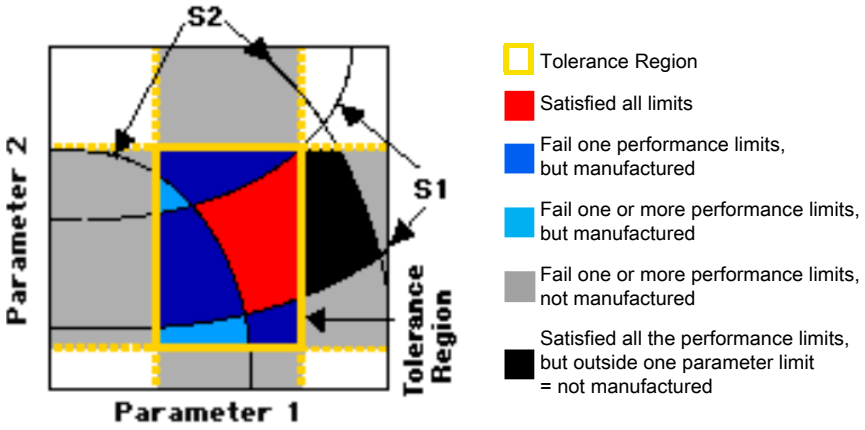
- Either iterative search (human, automated or mixed) or generation of data to visualise.



[Tweedie & Spence 1998; Spence 2001]

Color Coding

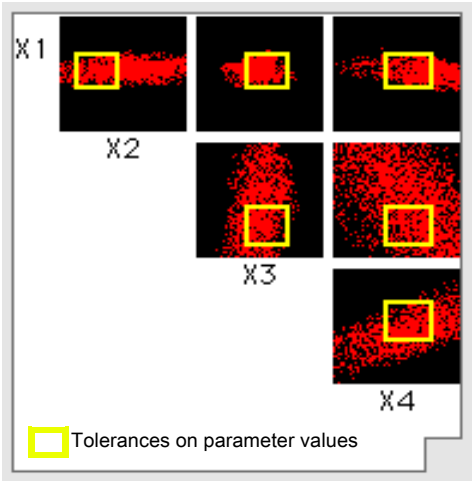
Parameter limits vs.
Performance limits



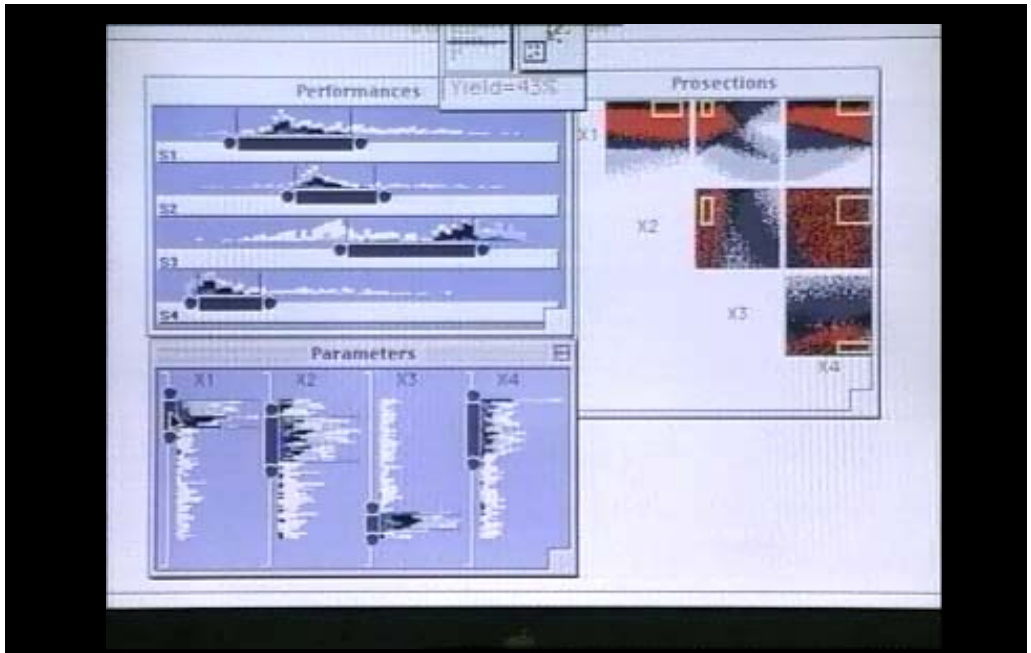
[Tweedie & Spence 1998; Spence 2001]

The Prosection Matrix

- A difficult cognitive problem is eased by a simple perceptual task
- Color Coding: customer's requirements on performance (different limits)
  - yes: red or green
  - no: black, dark gray, light gray, and white
- Yield enhancement: adjusting the tolerance-box (parameters) to the red points (useful performance)



[Tweedie & Spence 1998; Spence 2001]



Prosection Matrix [Tweedie & Spence 1998]
A Tool to Support the Interactive Exploration of Statistical Models and Data
→ Watch: http://ieg.ifs.tuwien.ac.at/~gschwand/teaching/infovis_vo_ws09/
(QuickTime, 28.6 MB, 00:02:53)

Summary Prosection Matrix

- Application Domain: Engineering Design
- Representation:
  - Set of scatterplots arranged in a matrix
  - m dimensional data sets: each scatterplot relates to a pair of parameters
- Color Coding: customer's requirements on performance
- Yield enhancement: fitting the parameter tolerance-box to the region of good performance with maximum overlap
- Problem: Color Coding (Red → Optimal Values)

Thanks to ...

- Silvia Miksch,
- Monika Lanzenberger,
- Markus Rester, and
- Marti Hearst

...for making nice slides of previous classes available