Part III

Geometrically-transformed displays

Outline

Geometrically-transformed display

A class of Visual mapping transformation:

- for table data
- mostly data independent
- maps one object to a set of points (and maybe lines) in 2D or in 3D
- examples:
  - scatter plot matrix
  - Andrews' curves
  - parallel coordinates

Outline

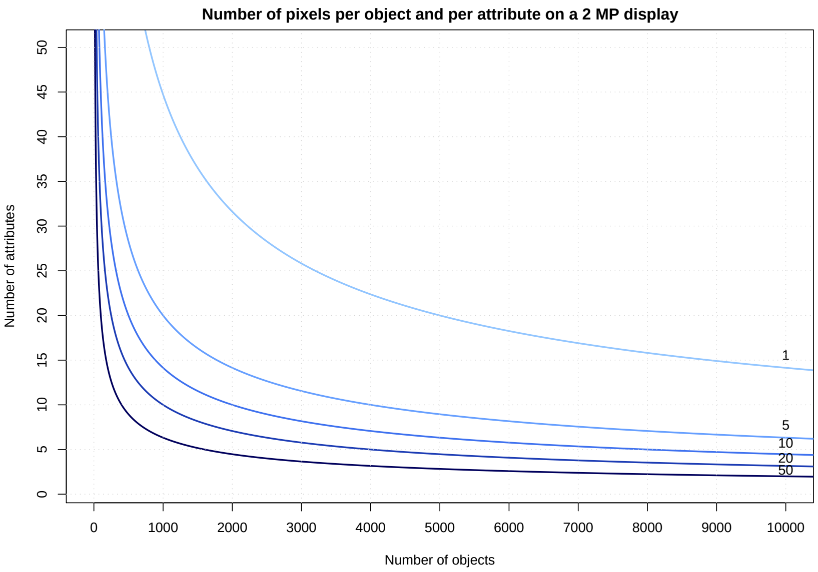
Scatter plot matrix

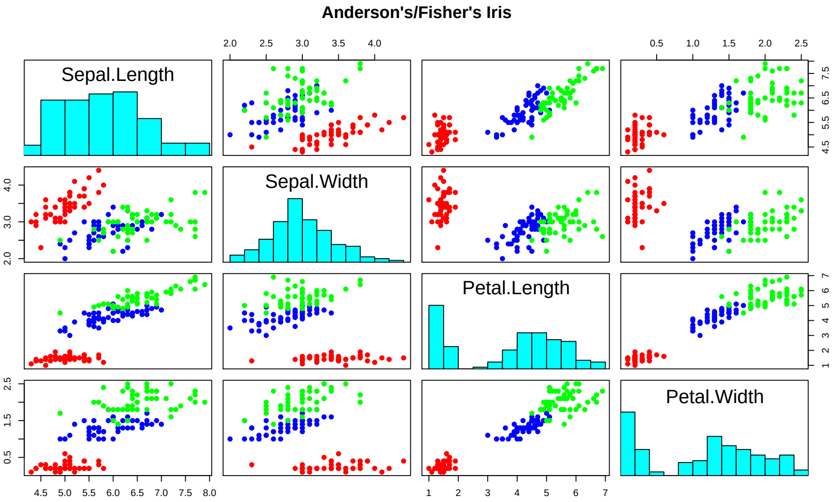
- A matrix of scatter plots

X_1 alone	(X_2, X_1)	(X_3, X_1)
(X_1, X_2)	X_2 alone	(X_3, X_2)
(X_1, X_3)	(X_2, X_3)	X_3 alone

- Pros:
  - almost no learning curve
  - inherits scatter plot extensions
- Cons:
  - worsen scatter plot overlapping problem: p variables $\Rightarrow p^2$ scatter plots $\Rightarrow p^2 n$ values (n objects)
  - links between plots are difficult to understand

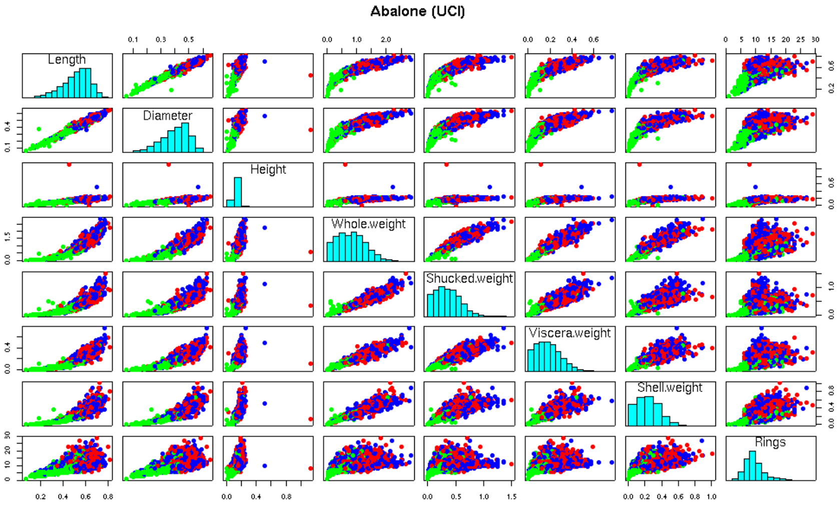
Scalability





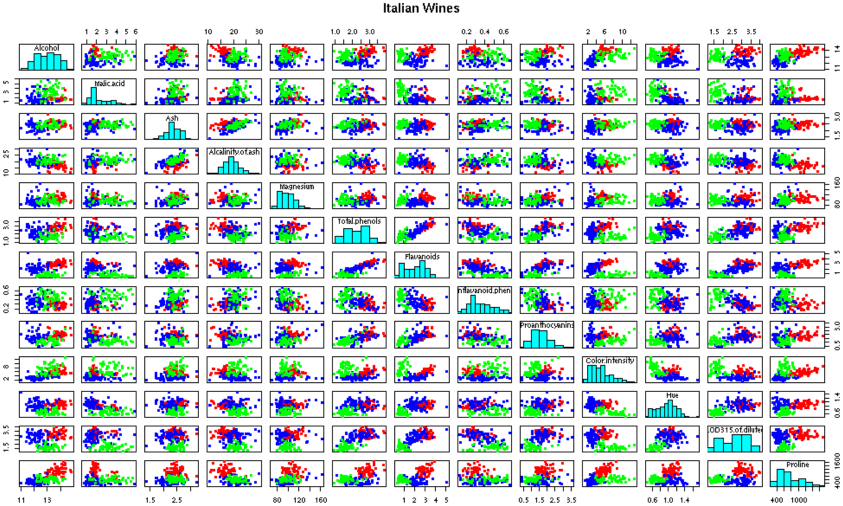
4+1 variables, 150 objects

Scalability



8+1 variables, 4177 objects

Scalability

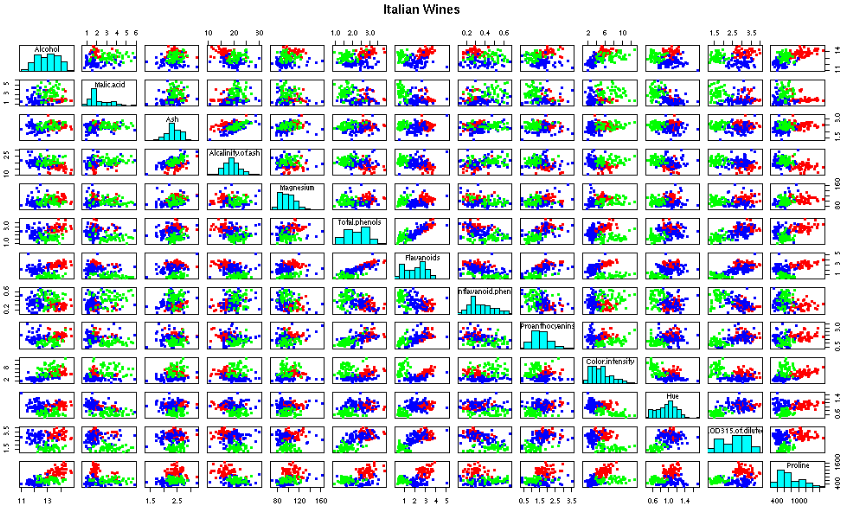


13+1 variables, 178 objects

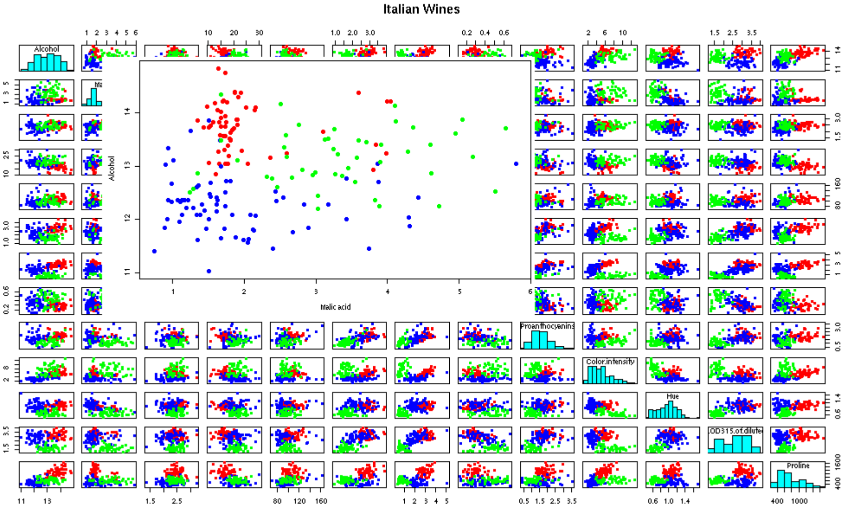
Scalability via interaction

- User choice:
  - variable selection
  - variable ordering
- Instant zoom (popup)
- Linking and brushing:
  - brushing:
    - selection of a subset of the objects
    - display of the selection (hue based)
  - linking:
    - brush in one view
    - display the results in all views
    - views don't have to be scatter plots
  - In previous slides: class information!

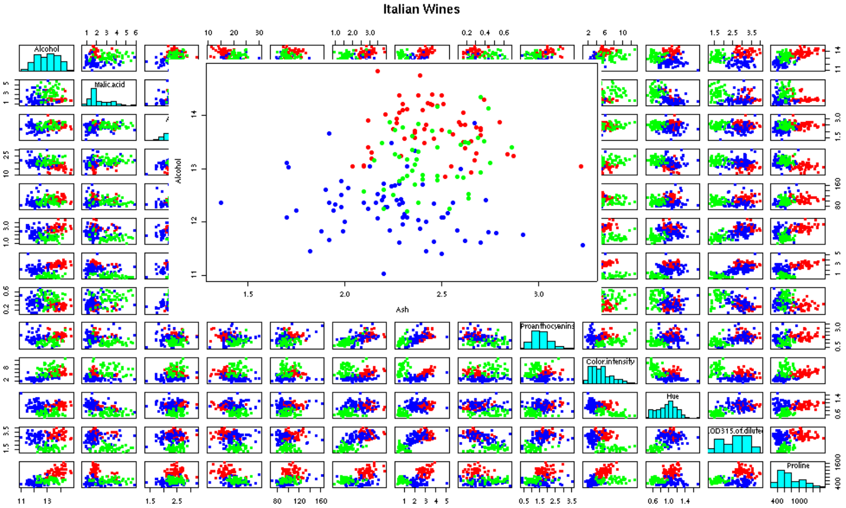
Zooming popup



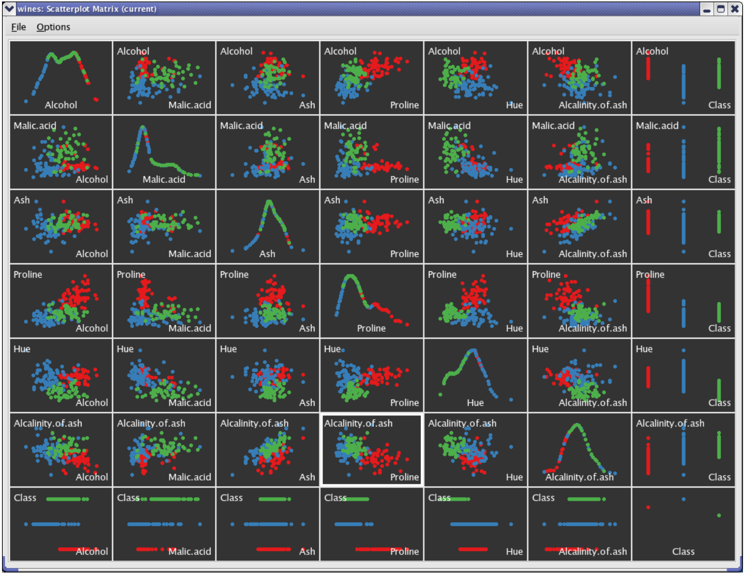
Zooming popup



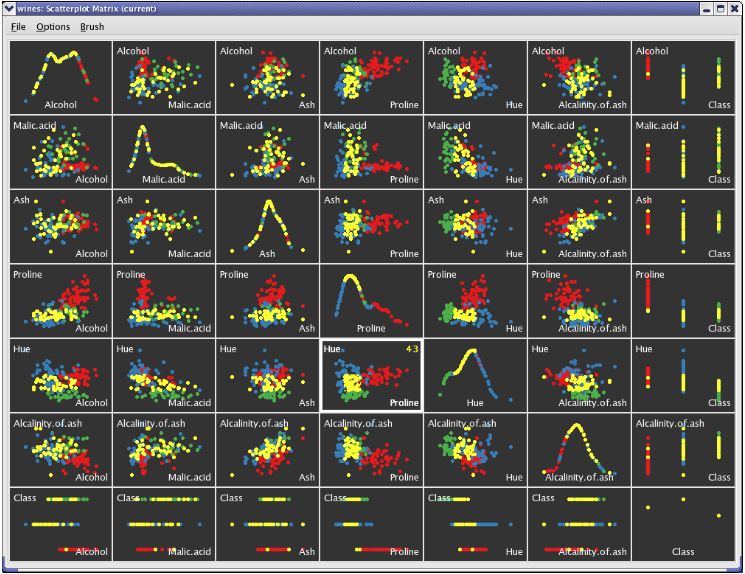
Zooming popup



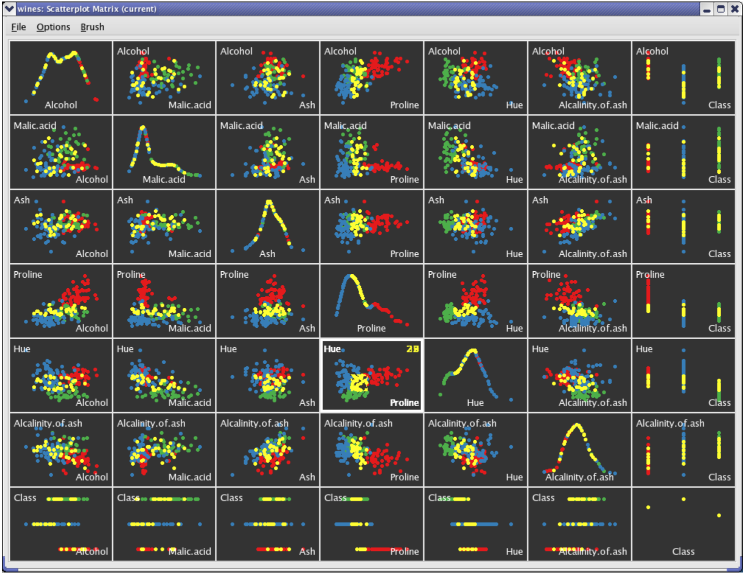
Linking and brushing (Ggobi)



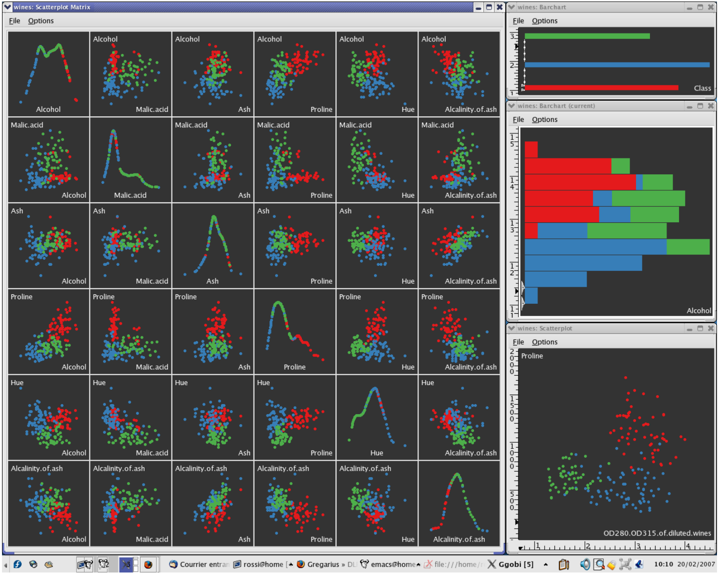
Linking and brushing (Ggobi)



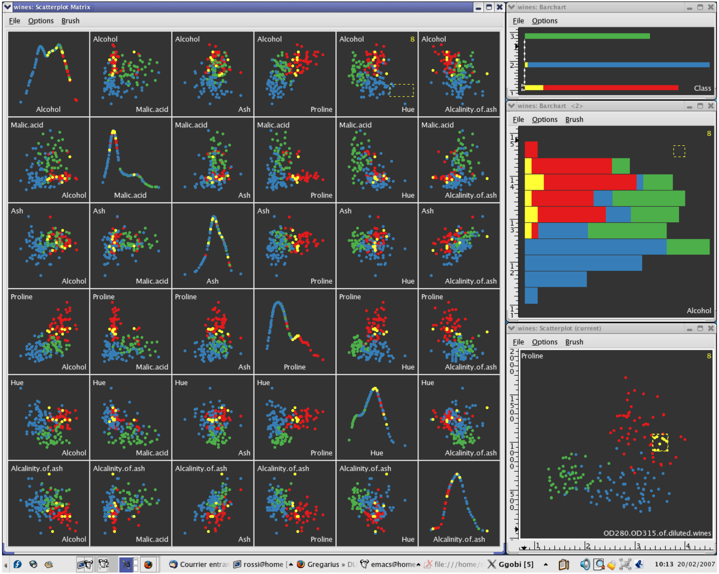
Linking and brushing (Ggobi)



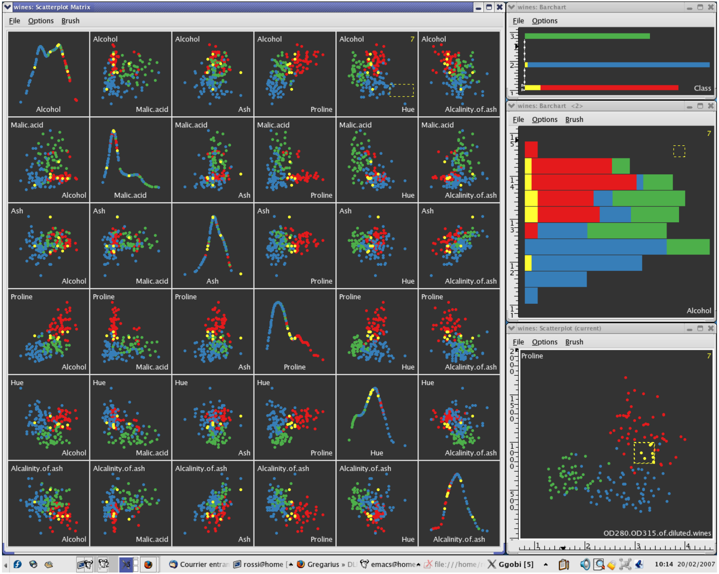
Linking and brushing (Ggobi)



Linking and brushing (Ggobi)



Linking and brushing (Ggobi)

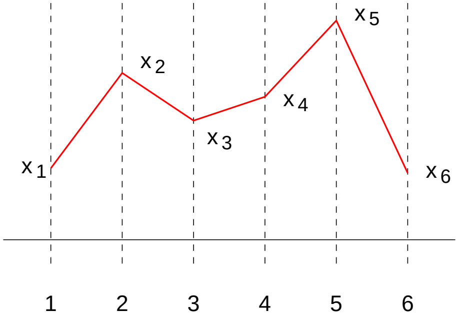


Outline

Parallel Coordinates

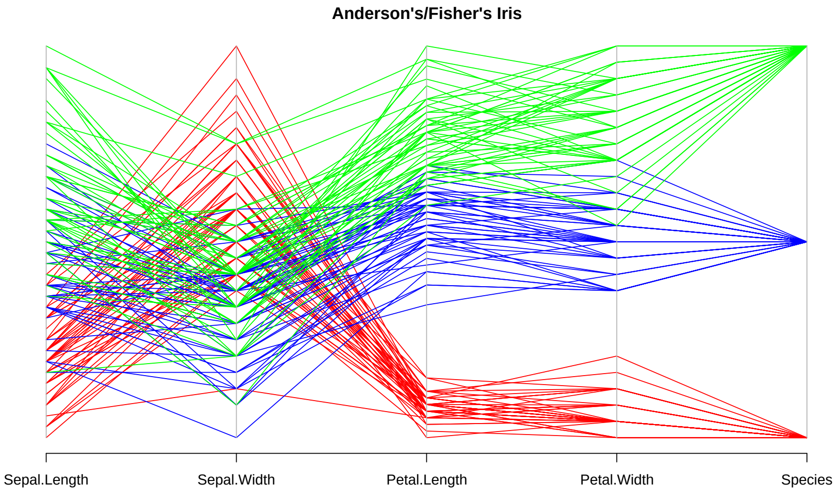
Proposed in 1985 by A. Inselberg

- A variable $=$ a vertical axis
- An object $=$ a piecewise linear curve
- $(x_1, \ldots, x_p)$ is mapped to the polyline that joins $(1, x_1)$, $(2, x_2)$, $\ldots$, (p, x_p)



Iris dataset

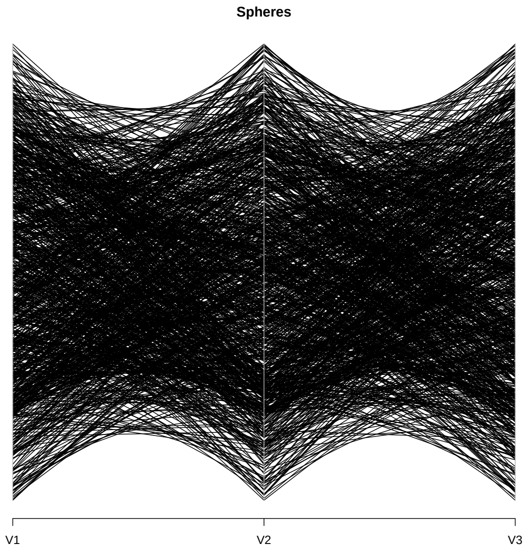
Iris dataset



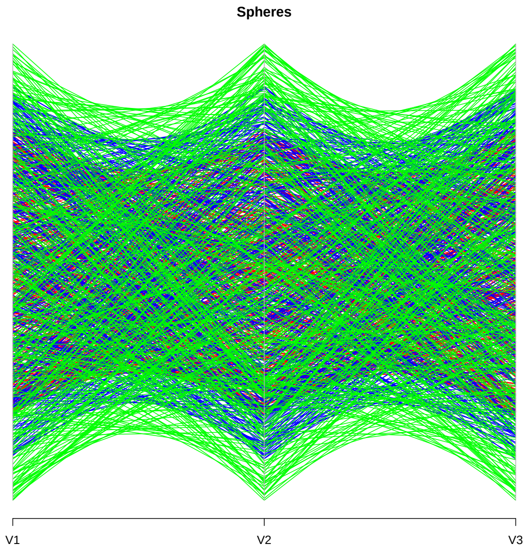
Parallel Coordinates

- Pros:
  - reasonable variable scalability
  - powerful after proper training
- Cons:
  - learning curve
  - major overlapping problem
  - variable order
- Reducing the problems:
  - rendering (transparency)
  - interactivity (brushing)
  - simplification (data clustering)
  - variable ordering (user guided and/or automated)

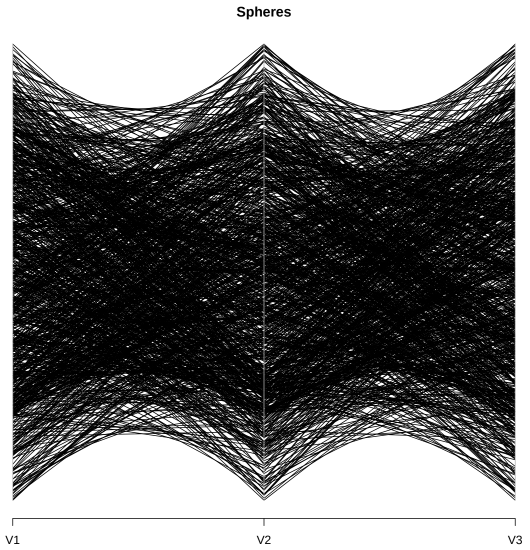
Overlapping



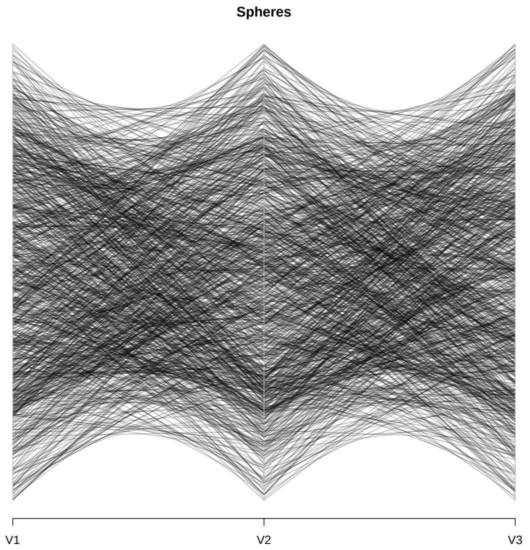
Overlapping



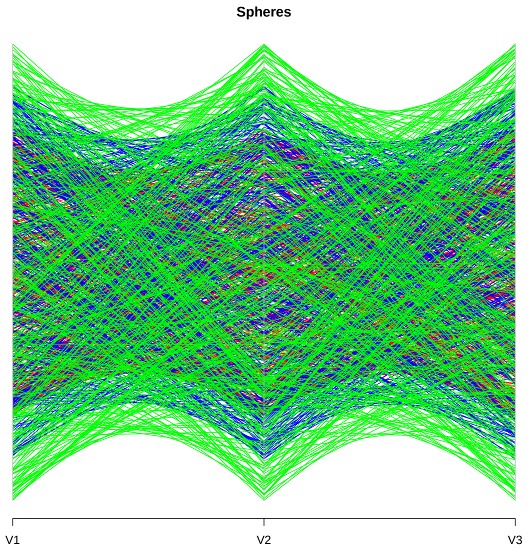
Transparency



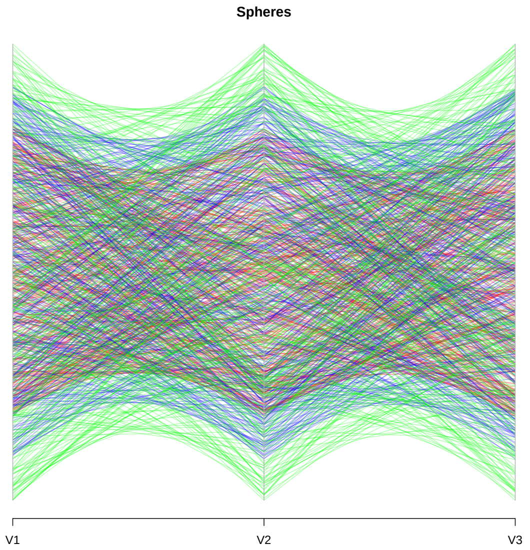
Transparency



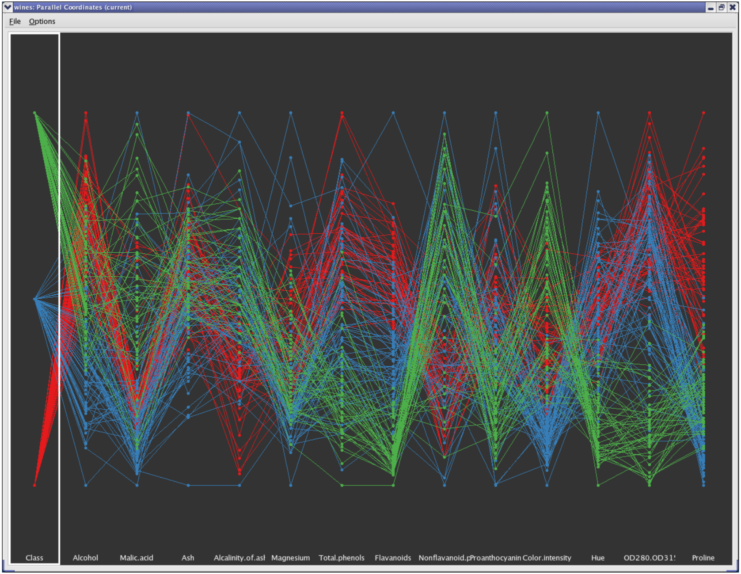
Transparency



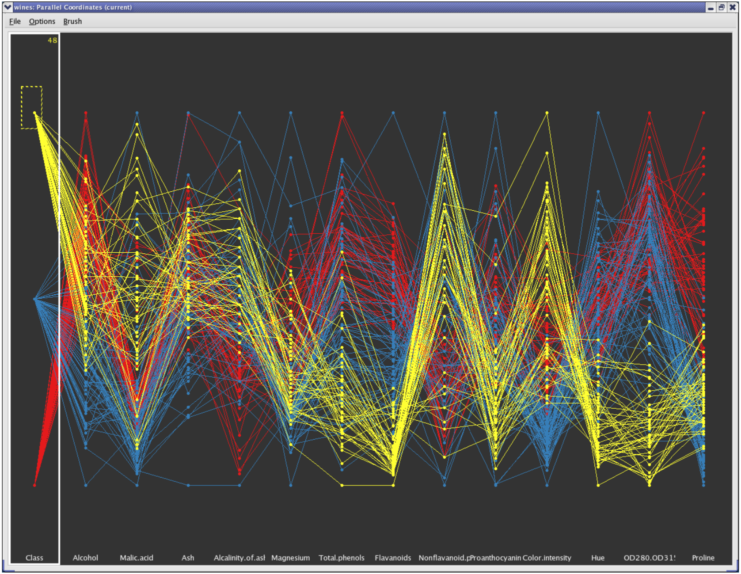
Transparency



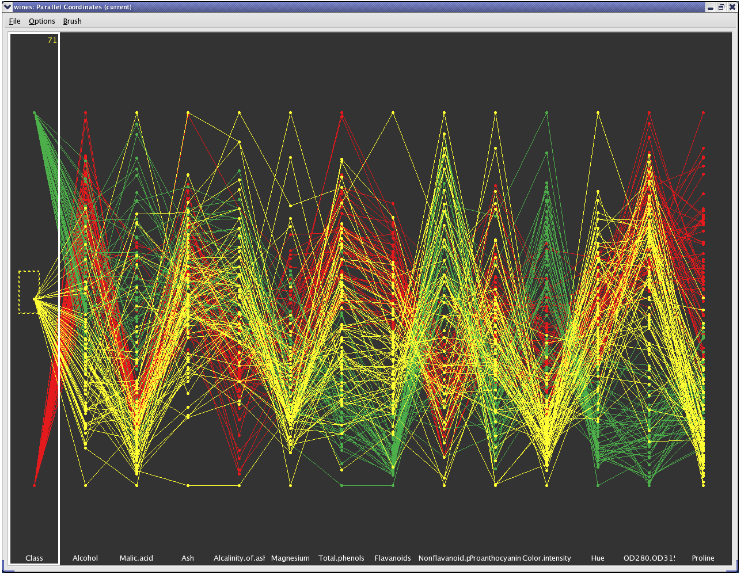
Brushing


wines: Parallel Coordinates (current)
File Options
Class
Alcohol
Malic.acid
Ash
Alcalinity.of.ash
Magnesium
Total.phenols
Flavanoids
Nonflavanoid.p
Proanthocyanin
Color.intensity
Hue
OD280.OD31!
Proline

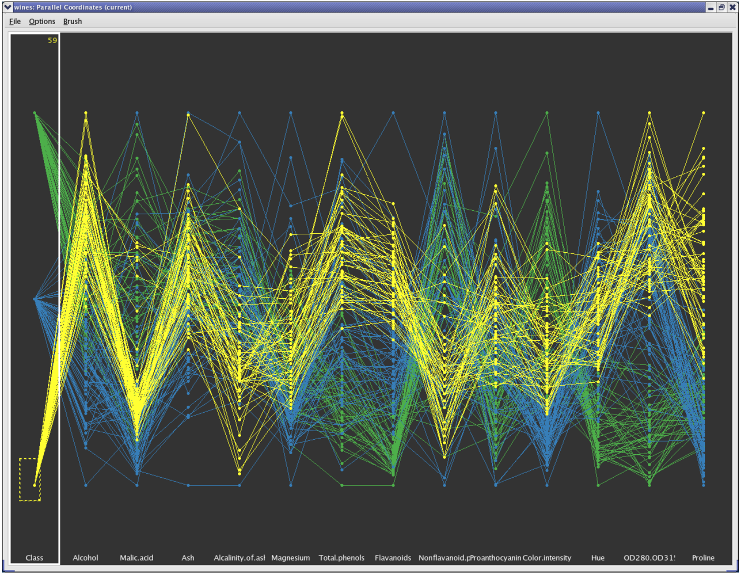
Brushing


wines: Parallel Coordinates (current)
File Options Brush
48
Class
Alcohol
Malic.acid
Ash
Alcalinity.of.ash
Magnesium
Total.phenols
Flavanoids
Nonflavanoid.p
Proanthocyanin
Color.intensity
Hue
OD280.OD315
Proline

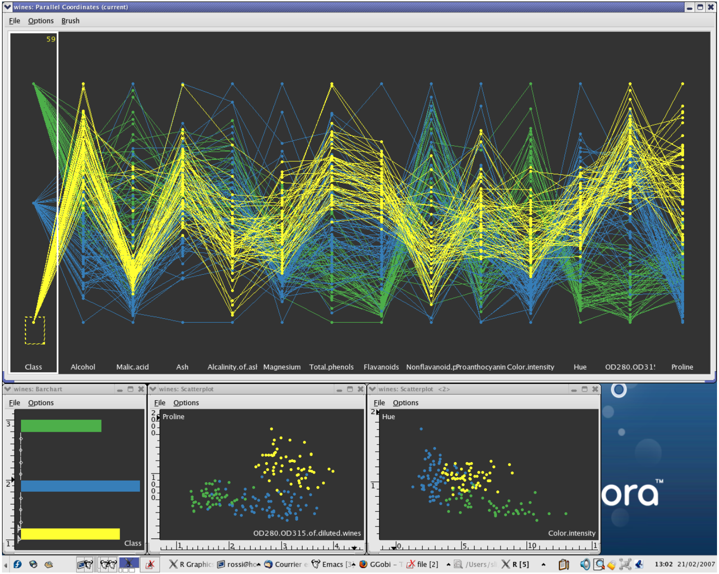
Brushing


wines: Parallel Coordinates (current)
File Options Brush
71
Class
Alcohol
Malic.acid
Ash
Alcalinity.of.ash
Magnesium
Total.phenols
Flavanoids
Nonflavanoid.p
Proanthocyanin
Color.intensity
Hue
OD280.OD315
Proline

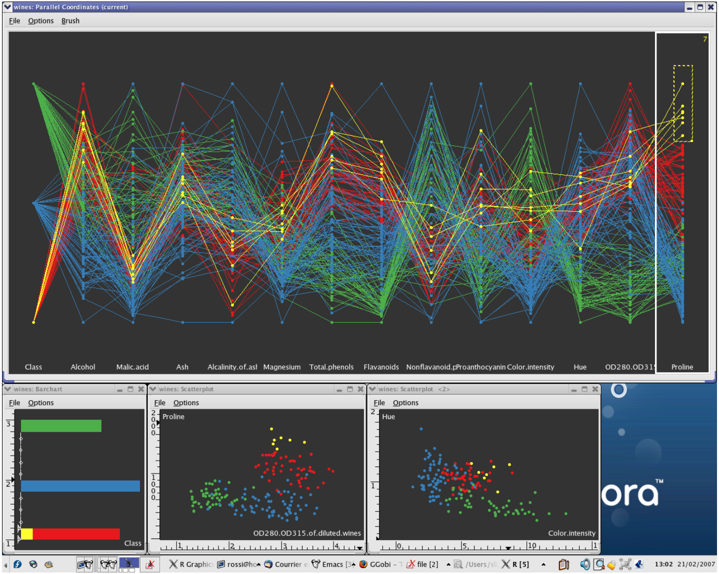
Brushing


wines: Parallel Coordinates (current)
File
Options
Brush
59
Class
Alcohol
Malic.acid
Ash
Alcalinity.of.ash
Magnesium
Total.phenols
Flavanoids
Nonflavanoid.p
Proanthocyanin
Color.intensity
Hue
OD280.OD31!
Proline

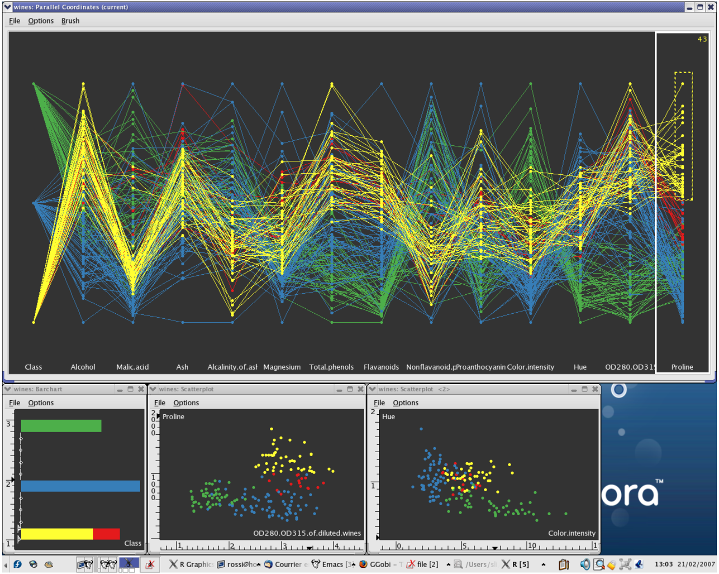
Brushing and linking



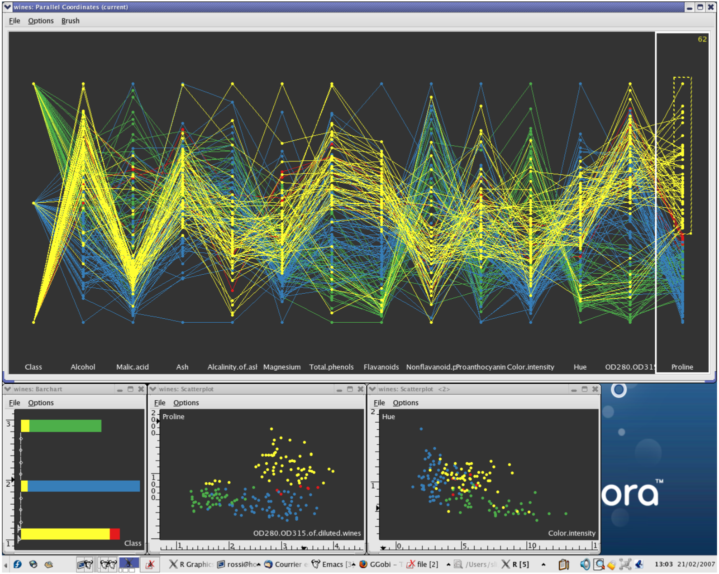
Brushing and linking



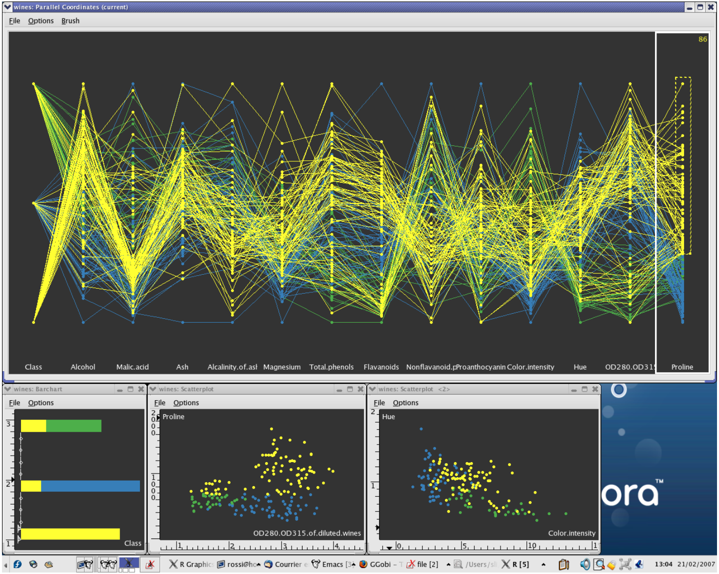
Brushing and linking



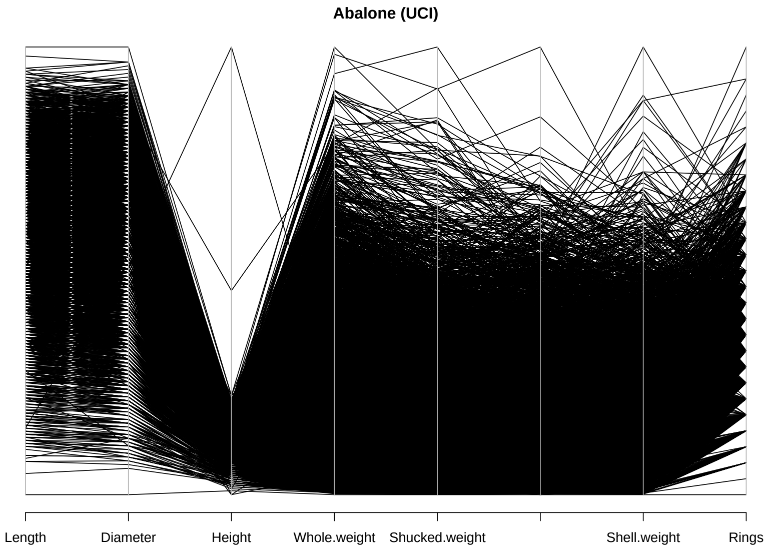
Brushing and linking



Brushing and linking

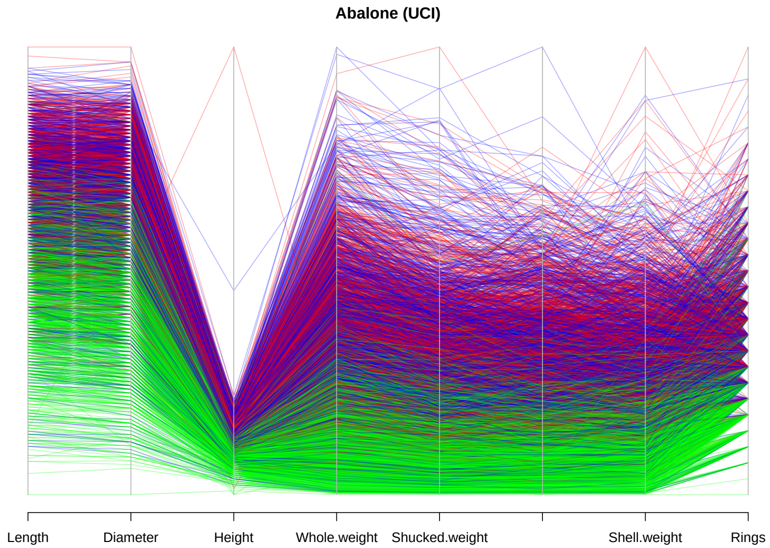


Clustering



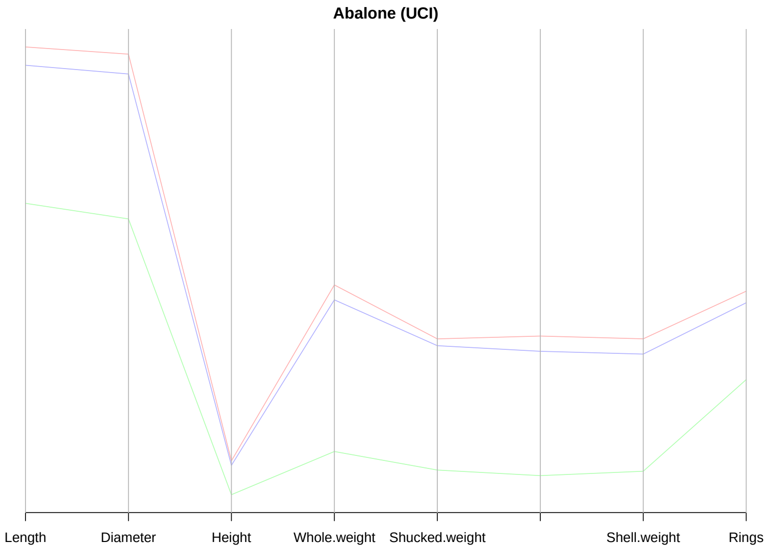
Abalone (UCI)
Length
Diameter
Height
Whole.weight
Shucked.weight
Shell.weight
Rings

Clustering

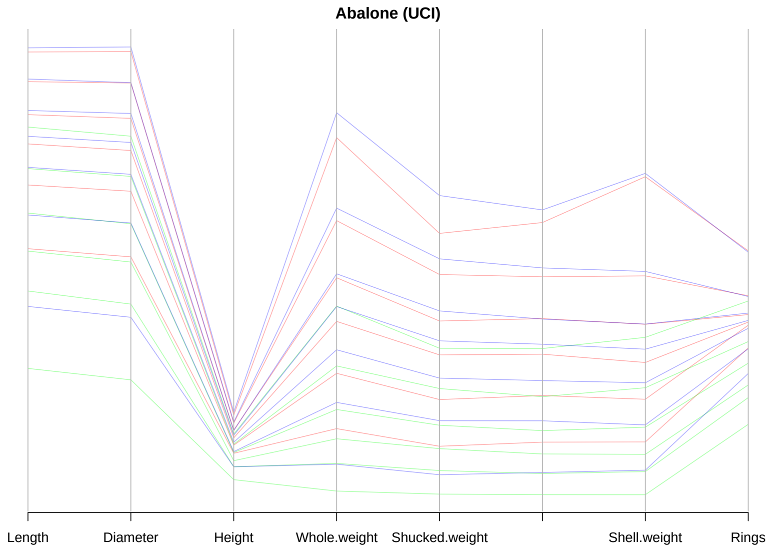


Abalone (UCI)
Length
Diameter
Height
Whole.weight
Shucked.weight
Shell.weight
Rings

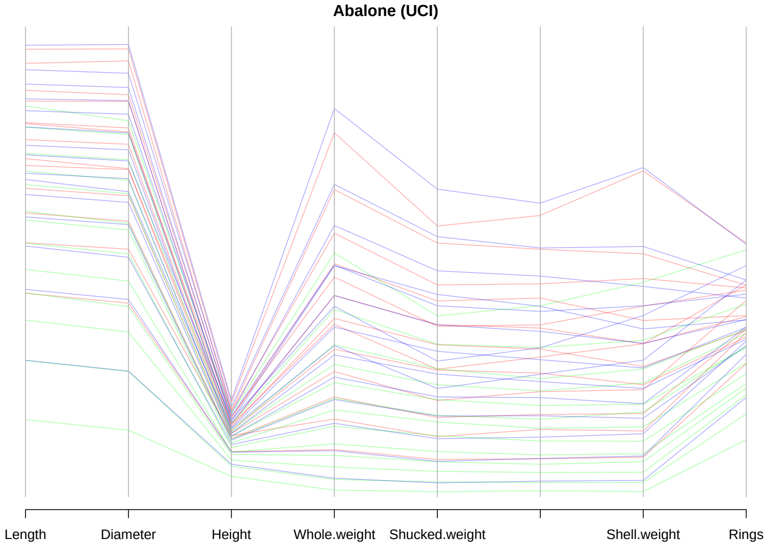
Clustering


Abalone (UCI)
Length
Diameter
Height
Whole.weight
Shucked.weight
Shell.weight
Rings

Clustering

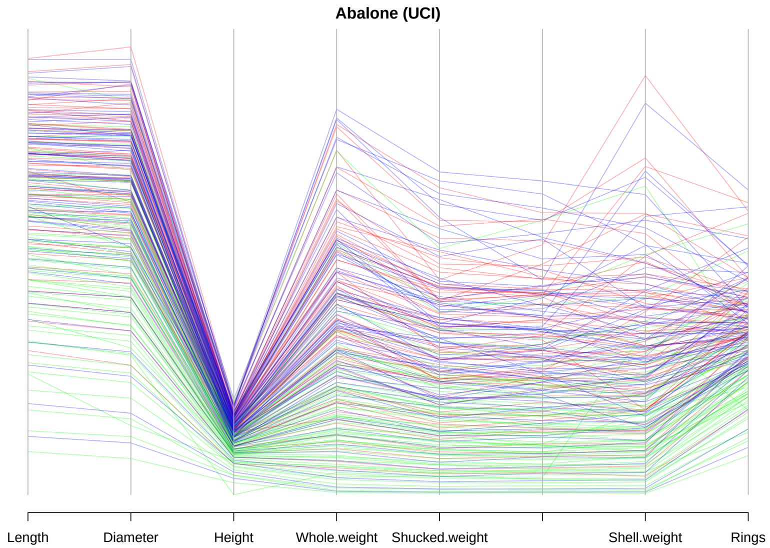

Abalone (UCI)
Length
Diameter
Height
Whole.weight
Shucked.weight
Shell.weight
Rings

Clustering



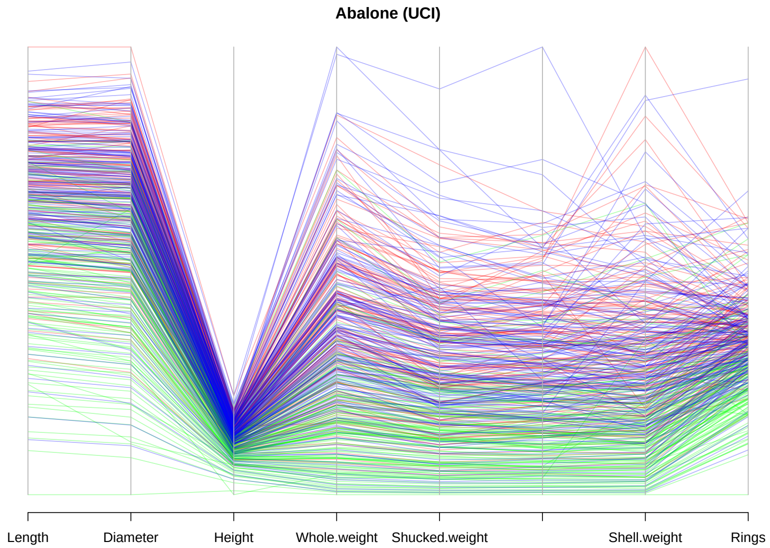
Abalone (UCI)
Length
Diameter
Height
Whole.weight
Shucked.weight
Shell.weight
Rings

Clustering

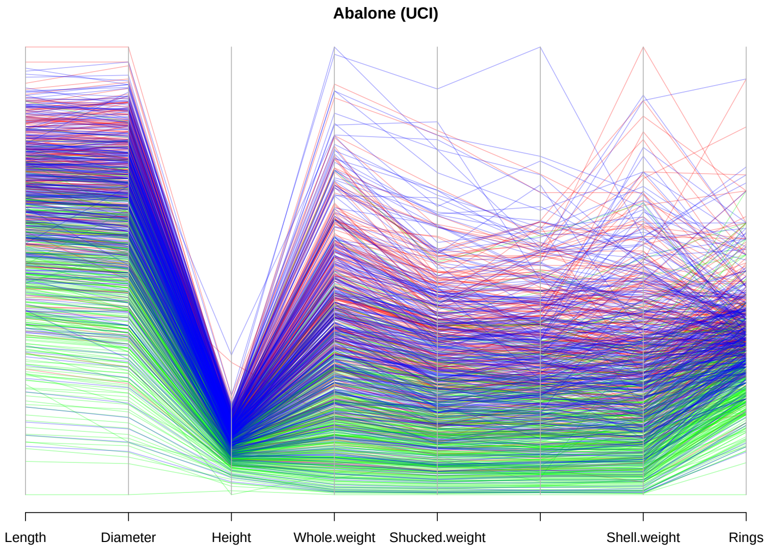


Abalone (UCI)
Length
Diameter
Height
Whole.weight
Shucked.weight
Shell.weight
Rings

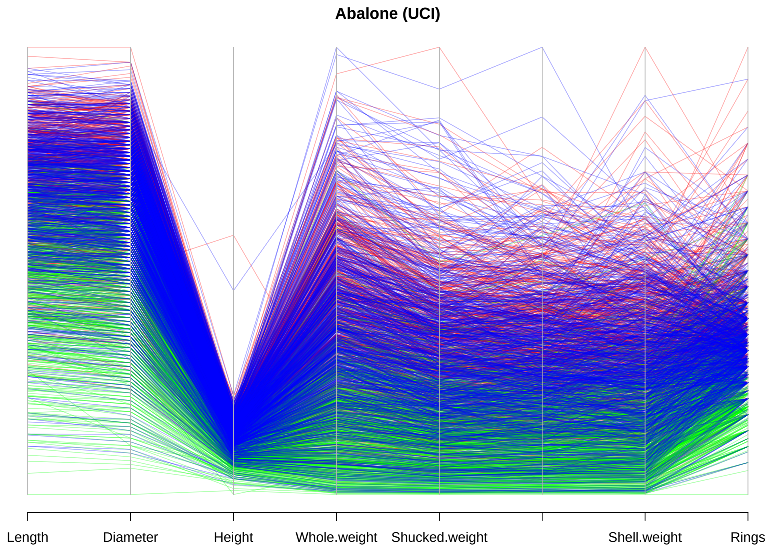
Clustering


Abalone (UCI)
Length
Diameter
Height
Whole.weight
Shucked.weight
Shell.weight
Rings

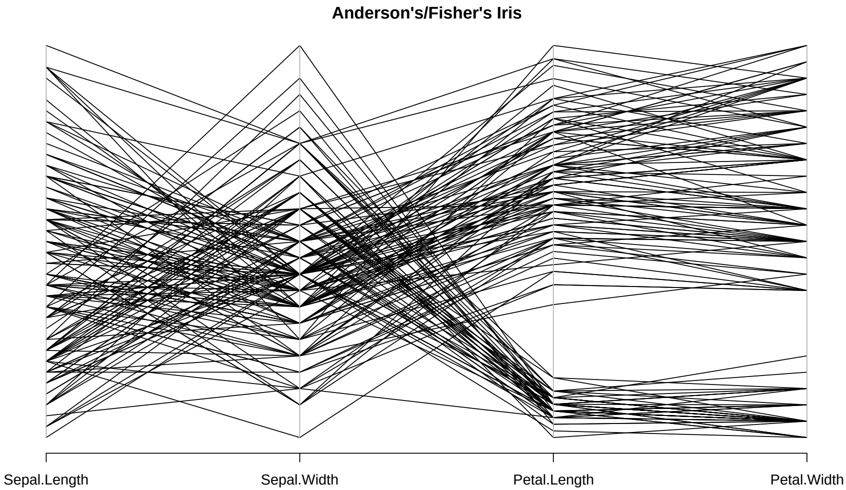
Clustering


Abalone (UCI)
Length
Diameter
Height
Whole.weight
Shucked.weight
Shell.weight
Rings

Clustering

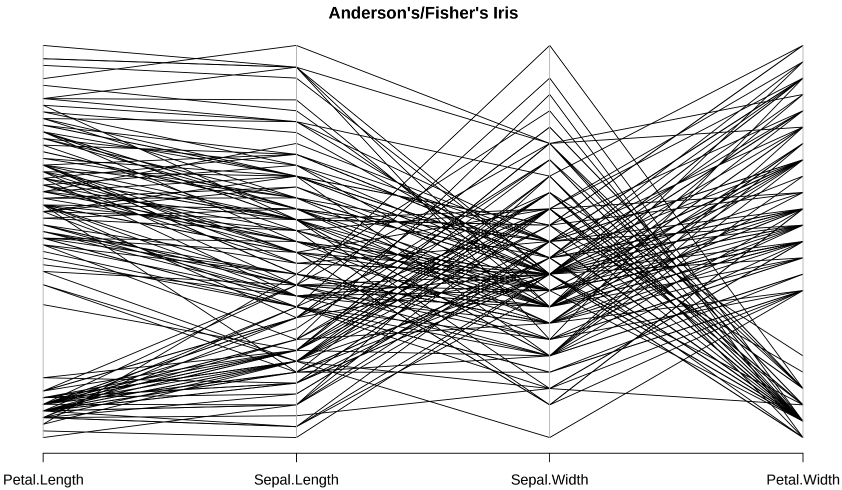

Abalone (UCI)
Length
Diameter
Height
Whole.weight
Shucked.weight
Shell.weight
Rings

Variable order



Are petal variables correlated?

Variable order



Are petal variables correlated?

Automated ordering

- resemblance measure between variables (e.g., correlation): $s(i, j)$
- variable permutation $\sigma : \{1, \ldots, p\} \to \{1, \ldots, p\}$
- optimal σ with respect to:

$$Q(\sigma) = \sum_{k=1}^{p-1} s\left(\sigma(k), \sigma(k+1)\right)$$

- NP-complete (Ankerst, Berchtold & Keim, 1998) $\Rightarrow$ heuristics:
  - traveling salesman analogy: optimization methods (e.g., Ant colony)
  - variable clustering
  - etc.
- an instance of the seriation problem: finding an optimal order among objects

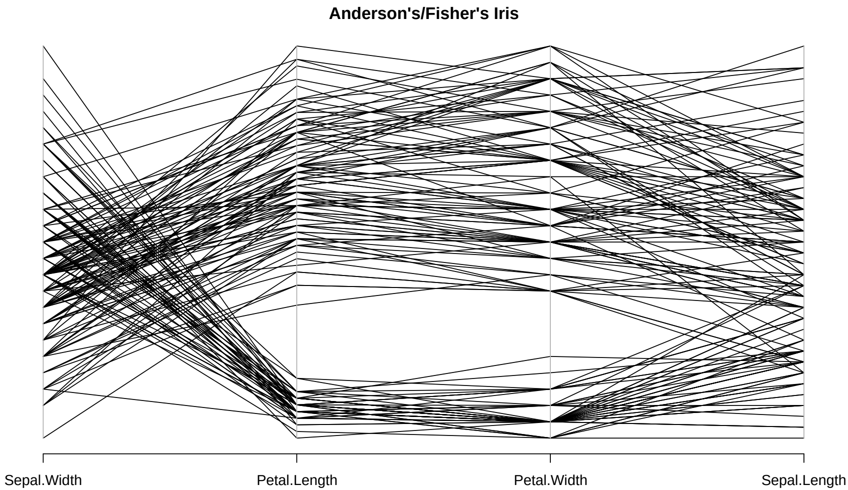
Projection based approach

Projection to 1D:

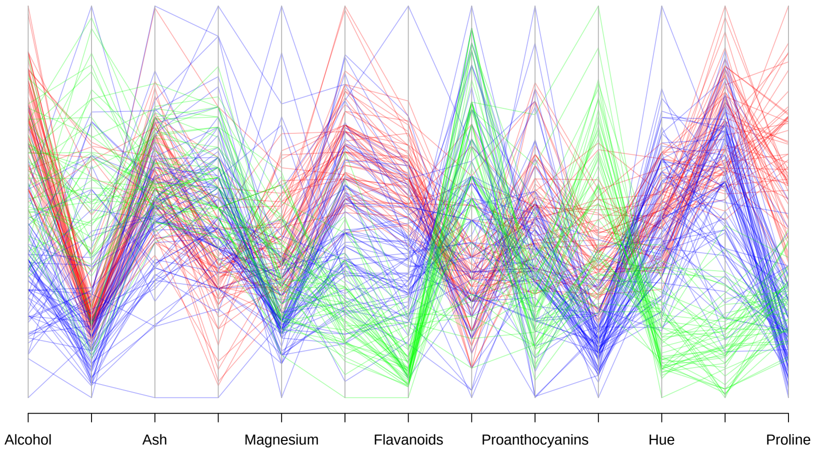
- $d(i,j)$ dissimilarity between variables
- map a variable i to a real number y_i
- choose $(y_i)_{1 \leq i \leq p}$ such that $(y_i - y_j)^2 \simeq d(i,j)^2$:
  - many possibilities
  - see Prof. Lee's lecture
- order variables according to $(y_i)_{1 \leq i \leq p}$
- this is *exactly* a projection problem

Ordering by Multi Dimensional Scaling

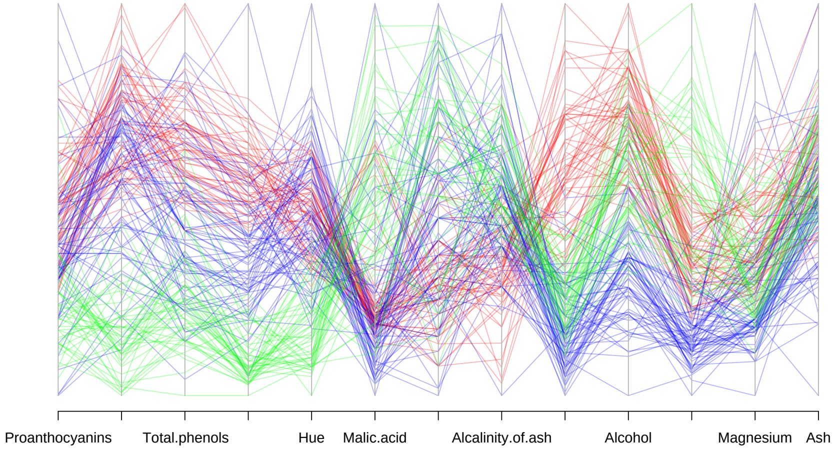
Ordering by Multi Dimensional Scaling



Ordering by Multi Dimensional Scaling



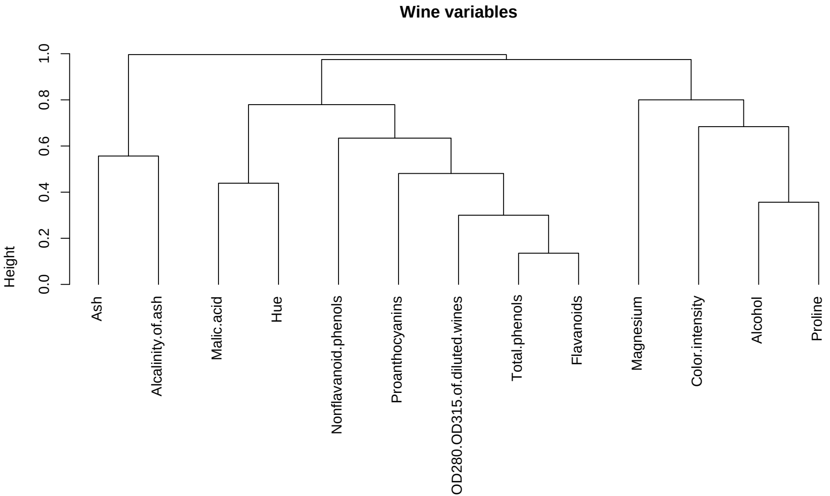
Ordering by Multi Dimensional Scaling



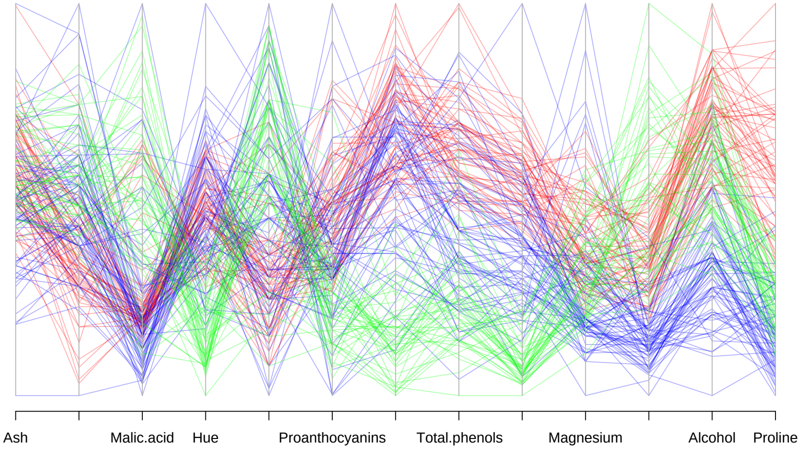
Clustering based approach

- Cluster variables according to resemblance measure
- Use clustering results to produce an order:
  - One dimensional Self Organizing Map
  - Hierarchical clustering:
    - with some heuristics
    - with optimal leaf ordering ($O(p^4)$, Bar-Joseph, Gifford and Jaakkola, 2001)
- Can be used to reduce the number of variables

Hierarchical clustering with heuristics



Hierarchical clustering with heuristics



Hierarchical clustering with heuristics