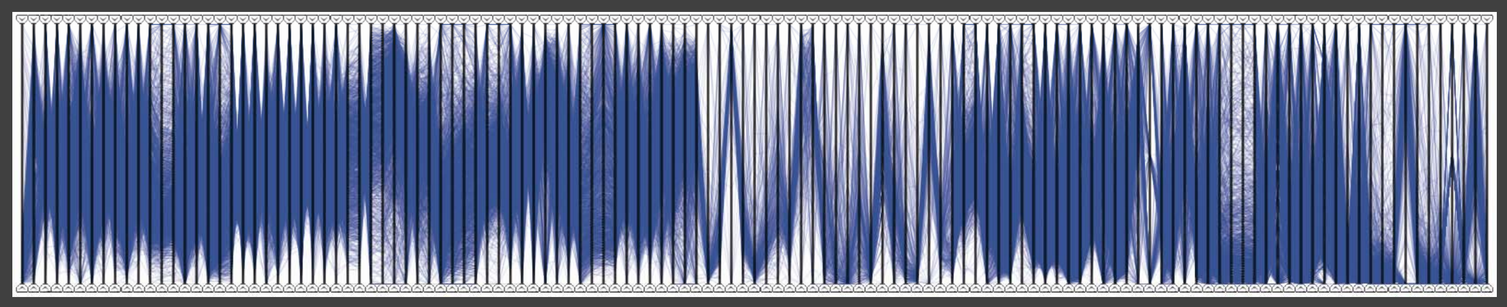
Interactive Dimensionality Reduction Through User-defined Combinations of Quality Metrics

Sara Johansson and Jimmy Johansson
C-Research, Linköping University

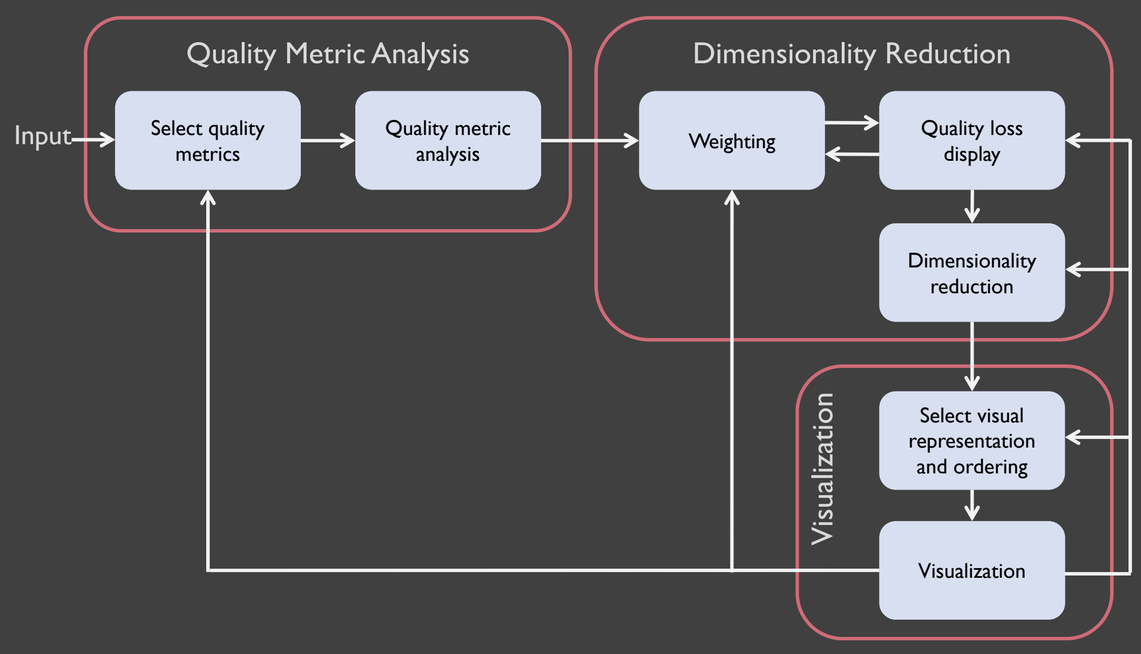
Background
Multivariate data sets including hundreds of variables are increasingly common in many application areas. Most multivariate visualization techniques are unable to display such data effectively, and a common approach is to employ dimensionality reduction prior to visualization. Most existing dimensionality reduction systems focus on preserving one or a few significant structures in data. For many analysis tasks, however, several types of structures can be of high importance.



The system
To overcome some of the issues involved in dimensionality reduction we present a dimensionality reduction system that enables:

- User defined combination of quality metrics for dimensionality reduction using weight functions to preserve as many important structures as possible.
- Automatic ordering of variables to enhance perception of diverse structures.
- Interactive quality guided reduction of variables through an interactive display facilitating investigation of trade-offs between loss of structure and the number of variables to keep.

Quality metric analysis
During quality metric analysis the full high dimensional data set is analysed and a quality value is extracted for each variable and each quality metric. The quality value is a measure of the variables involvement in a specific structure in the data set.

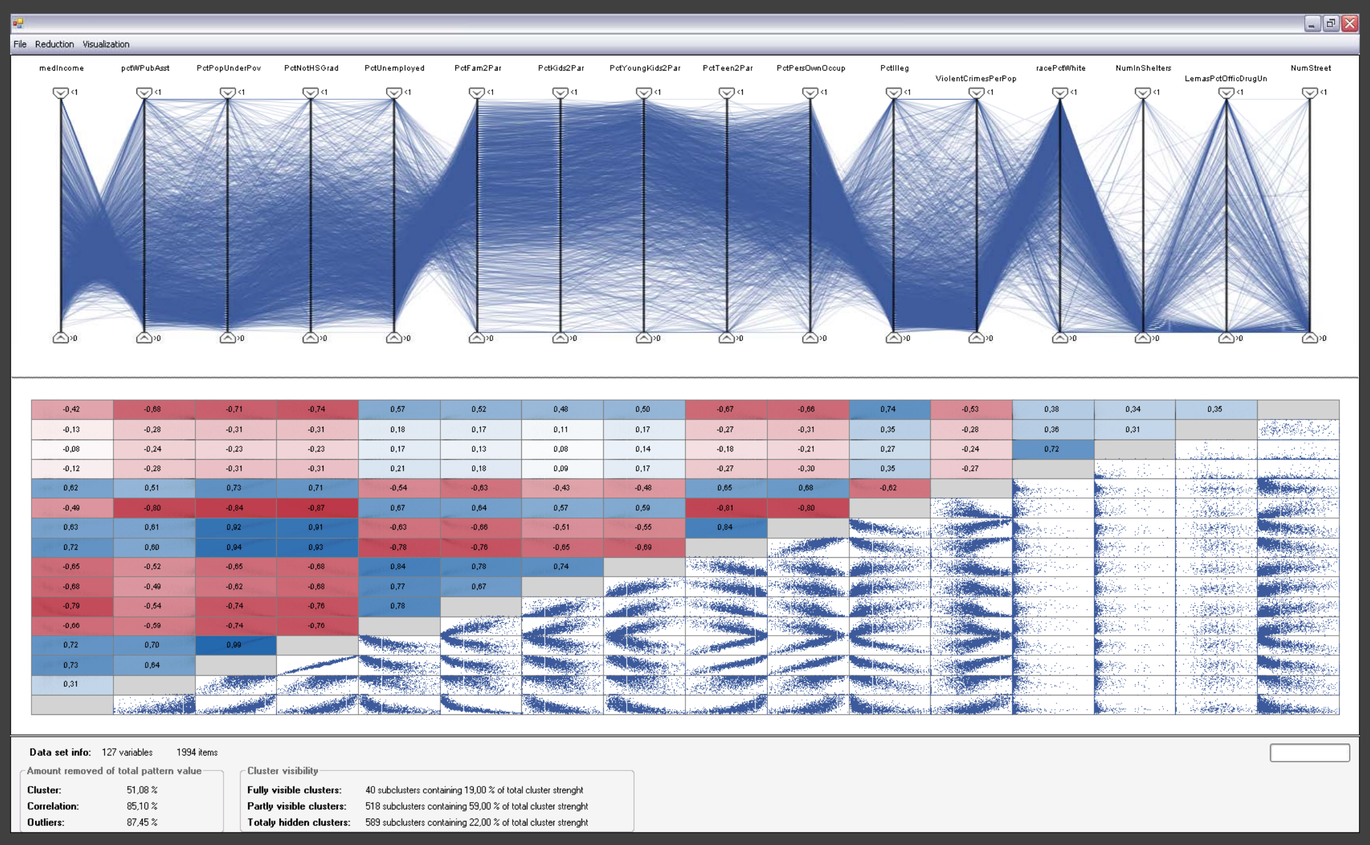


Quality loss display
The system offers an interactive display presenting the relationship between number of variables to keep and the loss of quality metrics, facilitating the decision on appropriate size for the reduced data set.

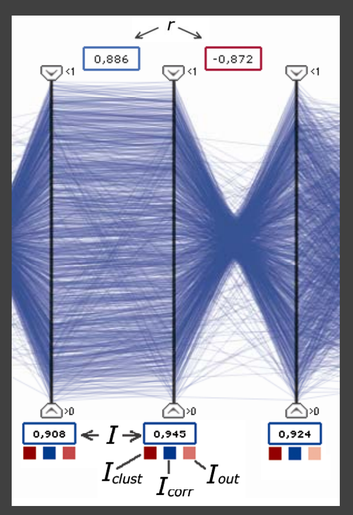
Dimensionality reduction
The relative importance of the individual metrics is defined by assigning weight values, *w*, and a global importance value, $I(x_j)$, is computed for each variable based on these.

$$I(x_j) = w_{corr} I_{corr}(x_j) + w_{out} I_{out}(x_j) + w_{clust} I_{clust}(x_j)$$

A reduction from an M-dimensional to a K-dimensional data set is performed by retaining the K variables with highest global importance value.



Through visual representations and automatic variable orderings, each enhancing the structures of an individual quality metric, the user can analyse the reduced data set and compare different variable orderings.



Visual aids provide information on correlation between adjacent axes (*r*), the global importance value (*I*) and the quality values (I_{clust} , I_{corr} and I_{out}) of individual variables.

CADICS

NORRKÖPING VISUALIZATION CENTER