This data originally came to us from Philip Hopke at Clarkson University.

This data consists of the percentage composition of 8 fatty acids (palmitic, palmitoleic, stearic, oleic, linoleic, eicosanoic, linolenic, eicosenoic) found in the lipid fraction of 572 Italian olive oils. (An analysis of this data is given in Forina et al, 1983). There are 9 collection areas, 4 from southern Italy (North and South Apulia, Calabria, Sicily), two from Sardinia (Inland and Coastal) and 3 from northern Italy (Umbria, East and West Liguria). The aim of the study on this data is to find combinations of the fatty acids which distinguish the oils from different regions. We describe an exploratory process to do this using the projection pursuit guided tour and other tools available in XGobi. This example is included purely to illustrate that graphical methods can often provide simple and readily interpretable results that are comparable to those given by more conventional methods.

Previous analyses of this data tested various techniques to find separation of regions and areas. One common thread between the studies is that principal component analysis is shown to be inadequate to the task. Other techniques used were discriminant analysis, nearest neighbour clustering, multidimensional scaling, decision trees (Forina et al, 1983). Batch projection pursuit using a 1-dimensional index iteratively to find a 3-dimensional solution was used by Glover and Hopke (1992).

We began by giving each {\it region} a different {\it glyph type}: plusses and crosses for southern Italy, squares for Sardinia and circles for northern Italy. {\it Areas} within each region are distinguished by {\it glyph size}, except for Umbria which is painted as a filled circle.

Various XGobi tools need to be used to understand this data. First individual dotplots reveal that {\it eicosenoic} acid separates the {\it southern Italian} oils clearly from all others. Projection pursuit with the Holes index on the northern Italian and Sardinian oils reveals that oleic and linoleic acids separate the northern oils from the Sardinian oils. If you search through the pairwise plots this can be seen as well, but it is nice to see that the projection pursuit guided tour finds the separation. To see a better separation of all the northern Italian oils and Sardianian use 3-d rotation on linoleic, oleic and linolenic acids. This gives as good a separation as the discriminant analysis coordinates.

Forina, M. and Armanino, C. and Lanteri, S. and Tiscornia, E. (1983)

Classification of olive oils from their fatty acid composition, In Martens, H. and Russwurm Jr., H. (ed), Food Research and Data Analysis, Applied Science Publishers, London, 189-214.

Glover, D. M. and Hopke, P. K. (1992) Exploration of Multivariate Chemical Data by Projection Pursuit, Chemometrics and Intelligent Laboratory Systems, Elsevier Science Publishers, Amsterdam.