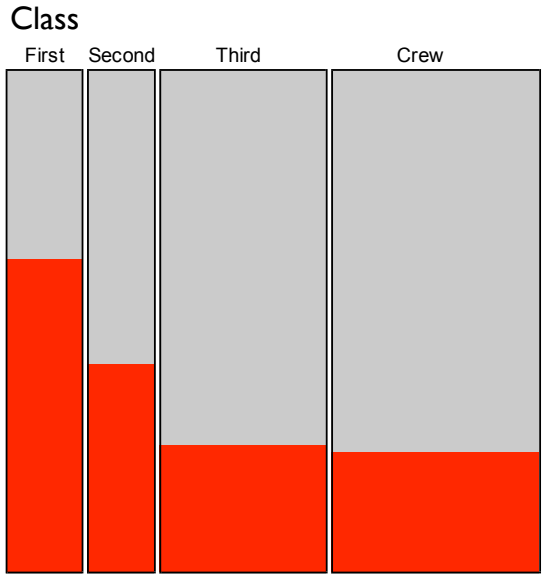
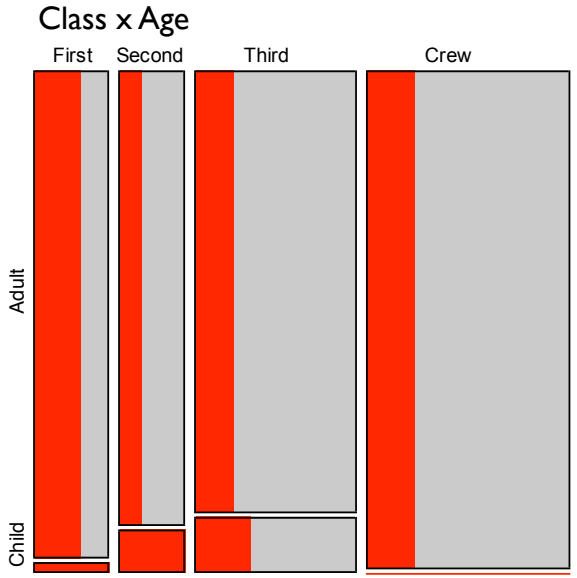
Chapter 4

Multidimensional Plots

Mosaic Plot: Construction

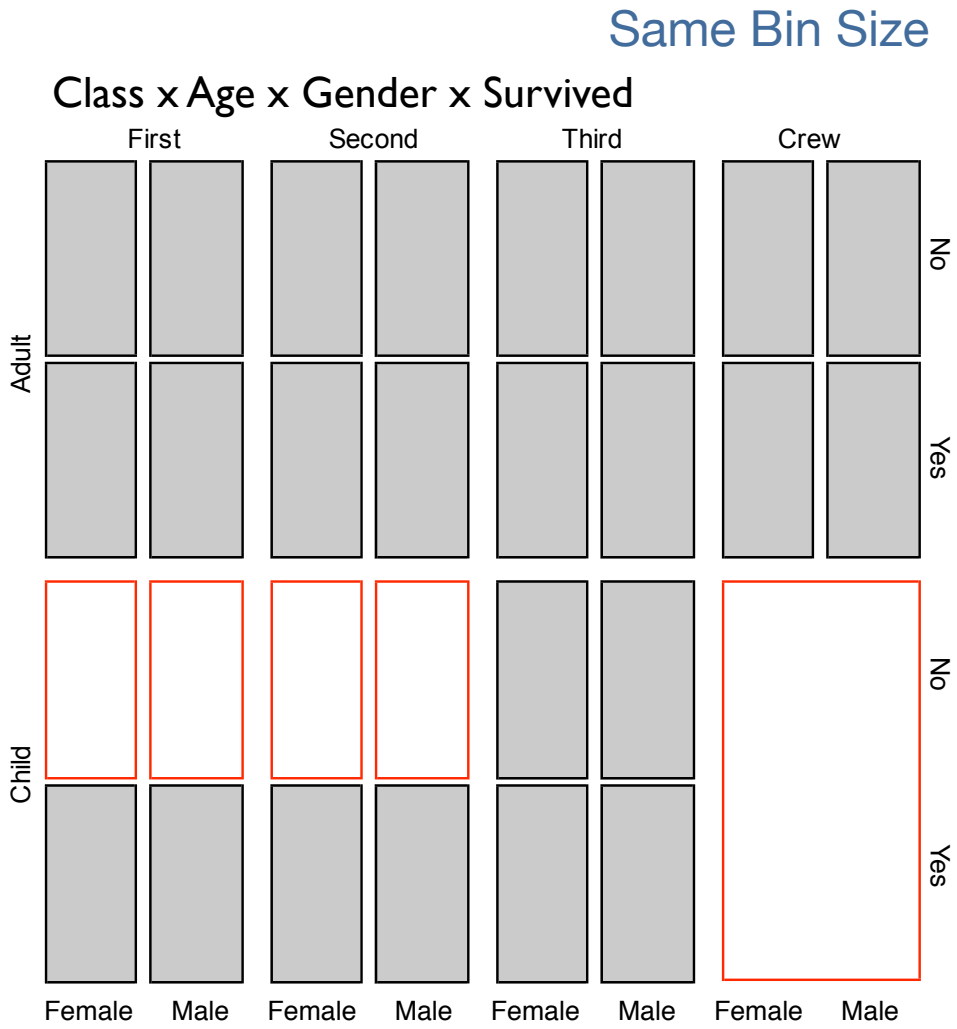
- Mosaic plots follow a strict recursive definition
- In one dimension it is nothing more than a simple spineplot
- Usually splitting alternates between x and y axis
- Gaps get smaller as the recursion progresses
- Highlighting adds just like an additional variable without gap





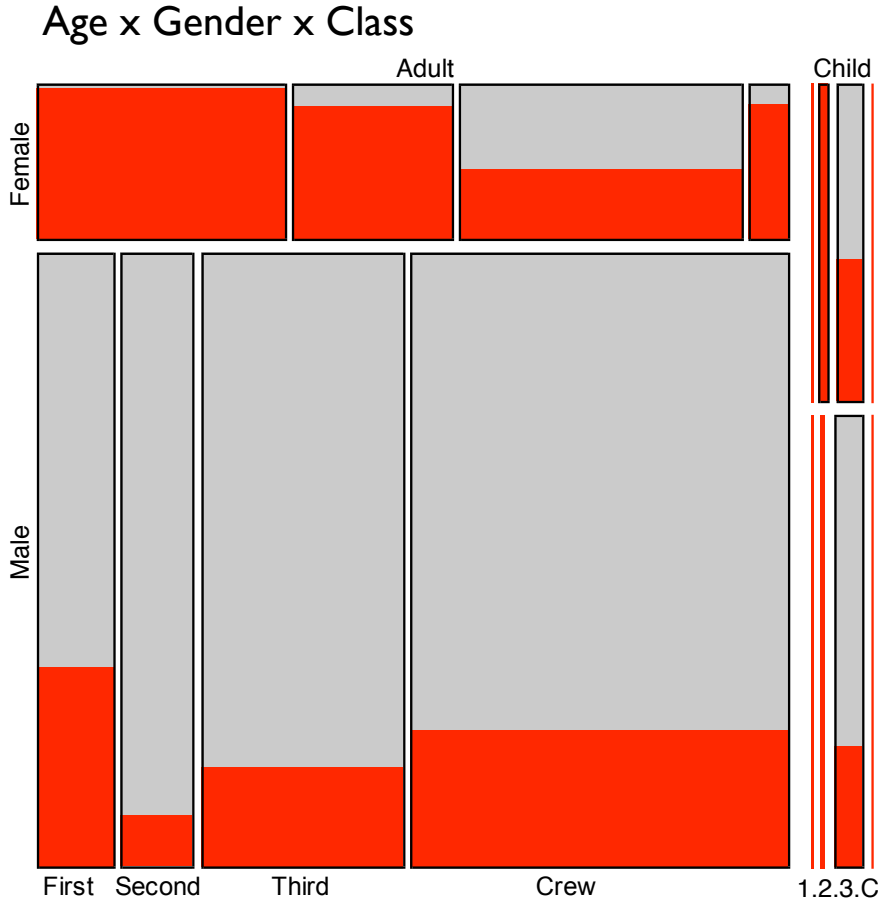
Mosaic Plot: Variations

- Two known variations which work well for few dimensions
  - multiple barchart ✔
  - fluctuation diagram ✔
- For many variables we face the "problem" of observing increasingly more empty cells
- In the same bin size view, we can focus on the distribution of the empty cells
- An optimized order may ease the interpretation of the empty cell structure



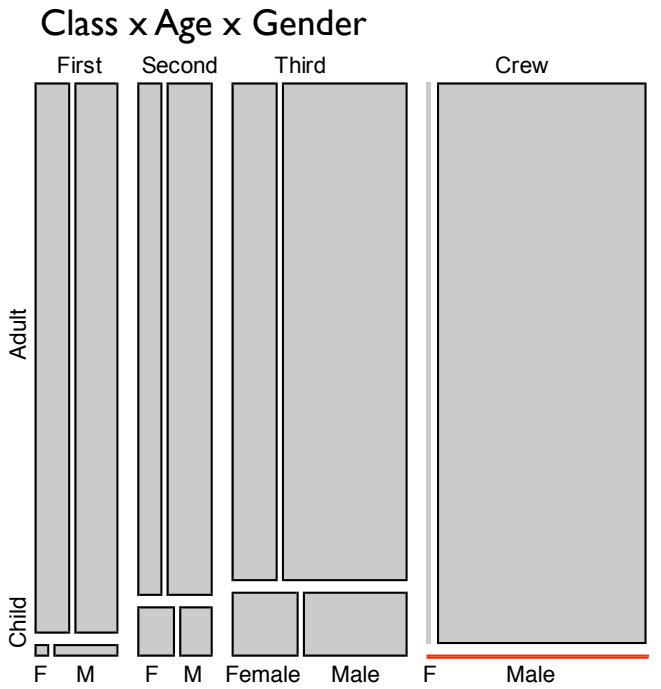
Mosaic Plot: Orderings

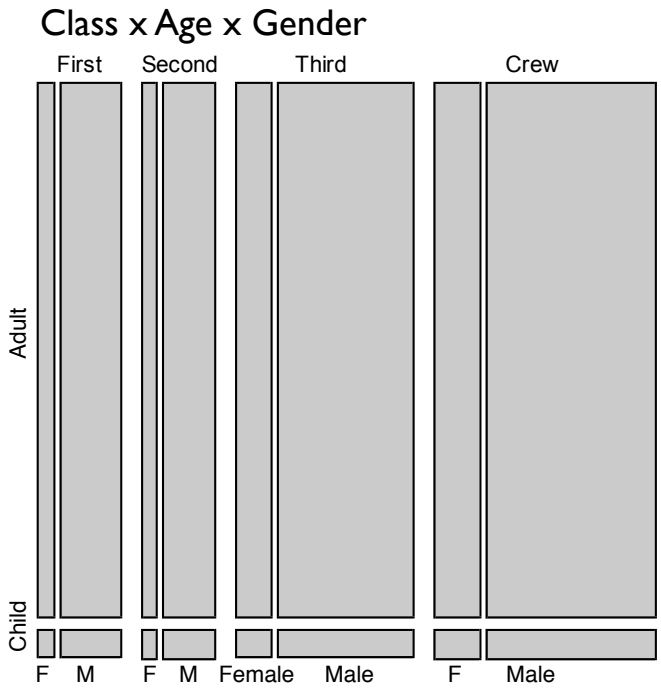
- The order of the variables has a great impact on the interpretation, not only in the same bin size view.
- In the “default view”, we compare survival rates conditioned on class, age and gender
- In this alternative view, we can focus on the comparison of the survival rates given age and gender
- The small survival rate for 2nd class males in now immediately visible



Mosaic Plot: Categorical Models

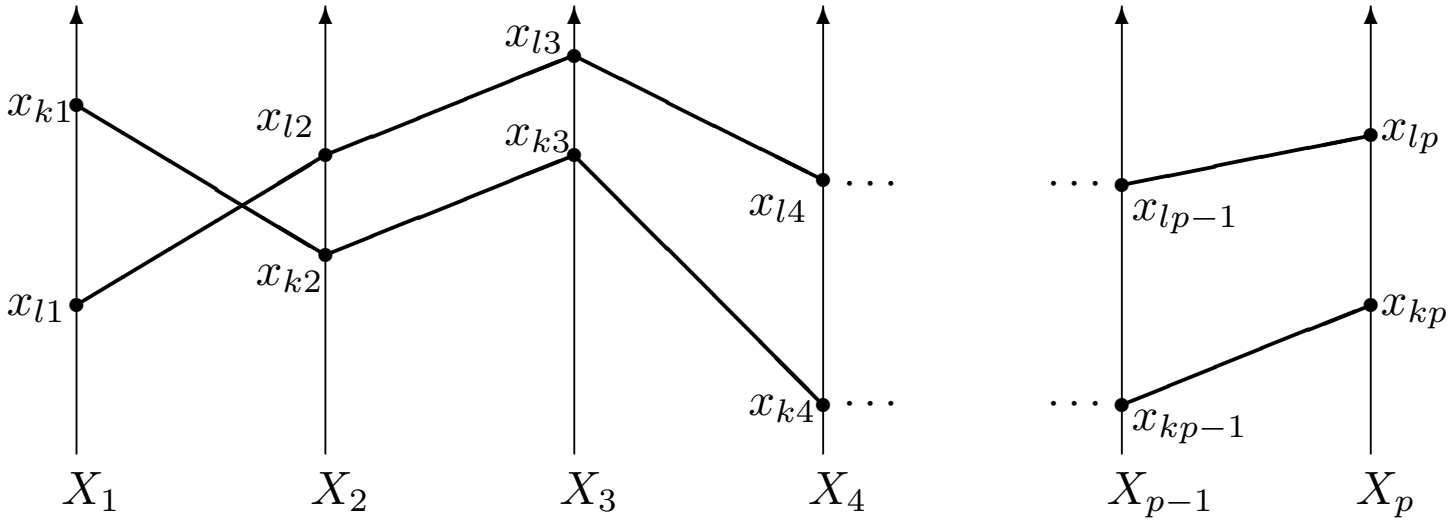
- A statistical model for categorical data can be easily visualized in a mosaic plot in the same way the raw data can be shown
- The most trivial model is actually the same bin size view
- The typical null model is usually the model of mutual independence
- In the independence model we find equal proportions shown by ‘straight crossings’
- Other – also more complex – models are relatively easy to spot, e.g., conditional independence
(again, order matters a lot)





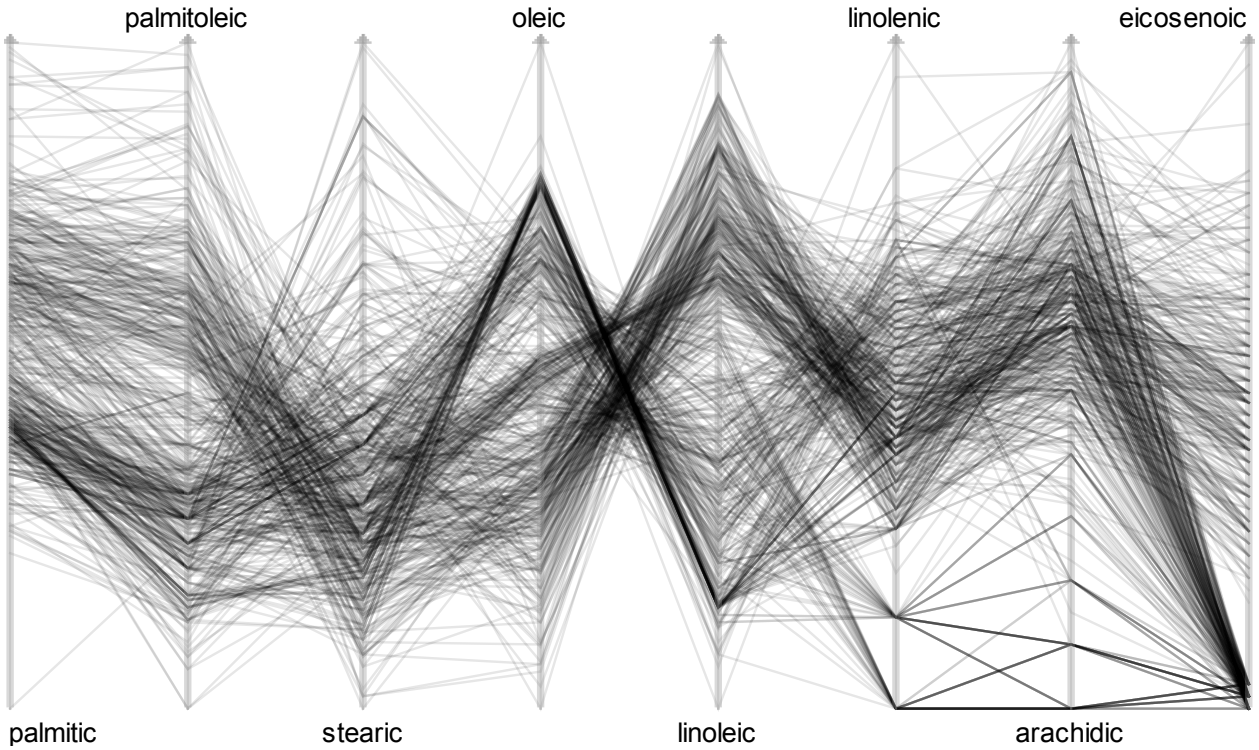
Parallel Coordinate Plots: Construction

- Parallel coordinates plot all axes in parallel and thus can escape the flatland of the 2-dimensional screen
- Each observation is plotted along each univariate coordinate axis
- In order to be able to follow an observation through all dimensions the points across the axes are joined by a polyline
- Parallel coordinate plots have a lot of interesting geometrical features
- In data analysis we search for patterns in the plot



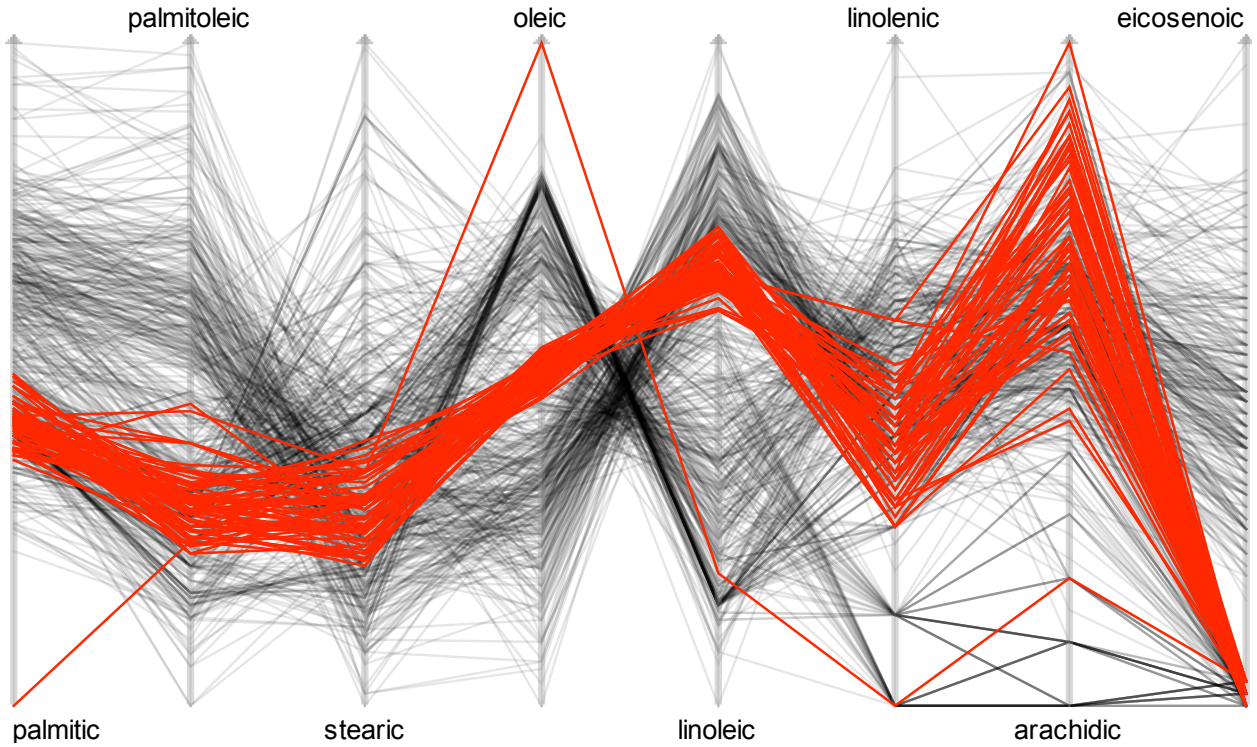
PCPs: Multivariate Overview

- The most general use of parallel coordinate plots is to get an efficient overview of many variables at a time
- We should not expect to see too much of multivariate features but rather exploit the benefit of seeing many distributions in one view along with some initial structural information
- Note:
  PCPs get cluttered quickly as we plot more and more data
  ➠ α-transparency is needed!



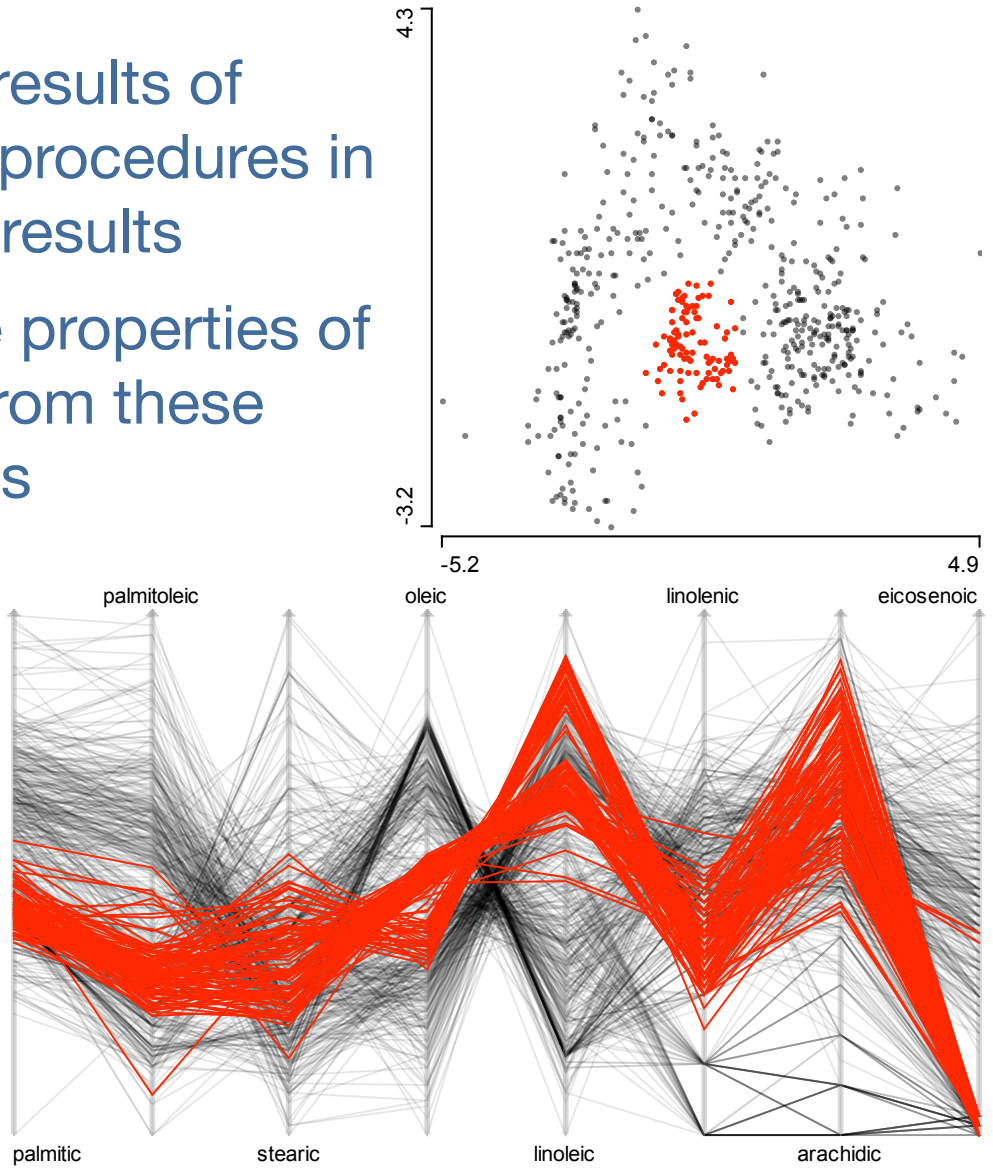
PCPs: Profiles

- With the ability to highlight single cases as well as whole sub-groups we can efficiently link these cases across all variables and see their properties for many variables simultaneously
- Example: Italian Olive Oils
  - A single outlier is selected for 'oleic' and turns out to be an outlier for two more variables
  - The group of 'Inland Sardinian' oils is selected which forms a very narrow and compact band through almost all variables



PCPs: Monitoring

- Often we want to interpret the results of statistical models or statistical procedures in order to iterate towards sound results
- In PCPs we can easily spot the properties of residuals or clusters resulting from these statistical procedures or models
- Example: MDS of fatty acids
  - The cluster in the center of the 2-dim output of the MDS is monitored in the PCP
  - It turns out that this cluster corresponds to all Sardinian oils which can be separated in the scatterplot of ‘oleic’ and ‘linoleic’



PCPs: Orderings

- PCPs suffer most from unsuitable defaults. The most obvious degree of freedom in a PCP is the ordering of the axes.
- As structures are best observed for adjacent axes the reordering of axes can make the difference
- To see all possible n! adjacencies we need only $\lfloor (n+1)/2 \rfloor$ plots
- Example: Order the axes by absolute correlation (there is no 'optimal' solution as each variable has only two neighbors)