Cluster Analysis & Multidimensional Scaling

Looking for like approaches (and an introduction to Systat)
Cluster Analysis

- A way of grouping data
- May be by cases
  - You want to find who is like who else
  - How alike are the cases?
- May be by variables
  - Are some variables like other variables
  - If so you can reduce the number of variables you work with
  - Or you can verify that they are similar by the way folks have responded
Attributes

- Most cluster analysis is exclusive, that is, any variable or case cannot be in two clusters at the same time.

- Several kinds of clustering:
  - Hierarchical, additive and partitioned.

- Based on some kind of correlation of the data:
  - Some clustering techniques are swayed by having different scales while others are not. Stay tuned.
Data

- Uses a variable by case format
- Can also use a correlation matrix
- Data can be nominal, ordinal, interval or ratio but each should have a different way to join the clusters
Output

- Generally a tree, dendogram or icicle
- May show several user defined groups and how well each case (or variable) fits with it’s average or mean group
- Can be refined and localized
- Have face and relational reliability
- Works best with ~20 or less variables or cases
Looking at the PSP

How do people in the class cluster?
How do Profiles cluster?

Profiles

- Global <-> Local
- Alone <-> Collaboration
- Help <-> Persistence
- Innovation <-> Tried
- Plan <-> Serendipity
## Correlation matrix of the variables

<table>
<thead>
<tr>
<th></th>
<th>GBLCL</th>
<th>HNTPRSNC</th>
<th>INVTRNDTRU</th>
<th>PLNSRDPT</th>
<th>ALNOTHRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBLCL</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HNTPRSNC</td>
<td>0.012</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INVTRNDTRU</td>
<td>0.247</td>
<td>0.120</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLNSRDPT</td>
<td>0.049</td>
<td>-0.189</td>
<td>-0.554</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>ALNOTHRS</td>
<td>0.218</td>
<td>-0.774</td>
<td>-0.218</td>
<td>0.181</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Join command (Systat only)
Multidimensional Scaling

Multidimensional Scaling is a method to fit a set of points in space that best represents the dissimilarity between all the points.
Configuration

Dimension-2

Dimension-1

BOMBAY
BERLIN
PARIS
LONDON
MONTREAL
CHICAGO
NEW YORK
CAPETOWN
Configuration

Dimension-2

Dimension-1

BOMBAY

BERLIN

PARIS

LONDON

MONTREAL

CHICAGO

NEW YORK

CAPETOWN
Derived Stimulus Configuration

Euclidean distance model
Derived Stimulus Configuration

Euclidean distance model

Dimension 1

Dimension 2
Scatterplot of Linear Fit

Euclidean distance model