

Author(s): Carl Berger, 2011

License: Unless otherwise noted, this material is made available under the terms of the **Attribution – Share Alike 3.0 license**

<http://creativecommons.org/licenses/by-sa/3.0/>

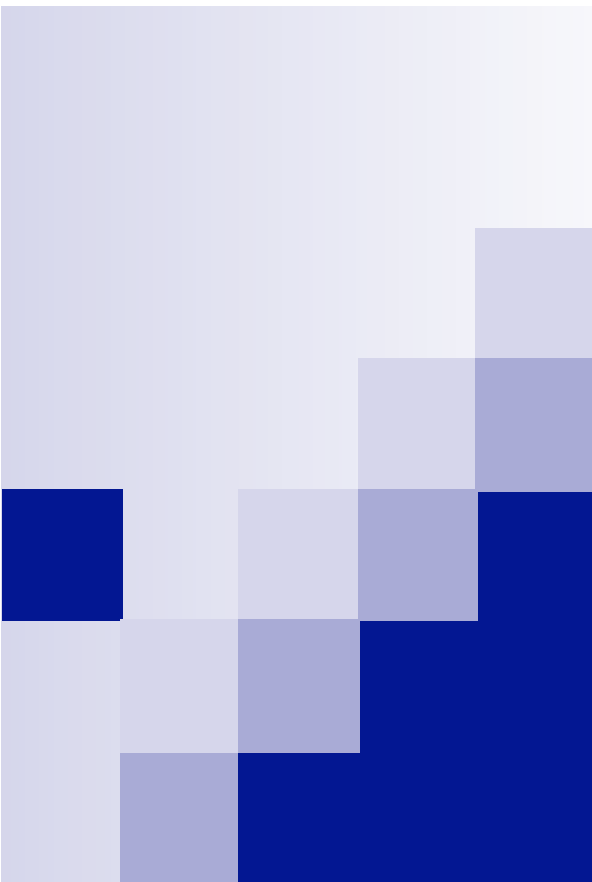
We have reviewed this material in accordance with U.S. Copyright Law **and have tried to maximize your ability to use, share, and adapt it.** The citation key on the following slide provides information about how you may share and adapt this material.

Copyright holders of content included in this material should contact **open.michigan@umich.edu** with any questions, corrections, or clarification regarding the use of content.

For more information about **how to cite** these materials visit <http://open.umich.edu/education/about/terms-of-use>.

Any **medical information** in this material is intended to inform and educate and is **not a tool for self-diagnosis** or a replacement for medical evaluation, advice, diagnosis or treatment by a healthcare professional. Please speak to your physician if you have questions about your medical condition.

Viewer discretion is advised: Some medical content is graphic and may not be suitable for all viewers.



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, dendrogram 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

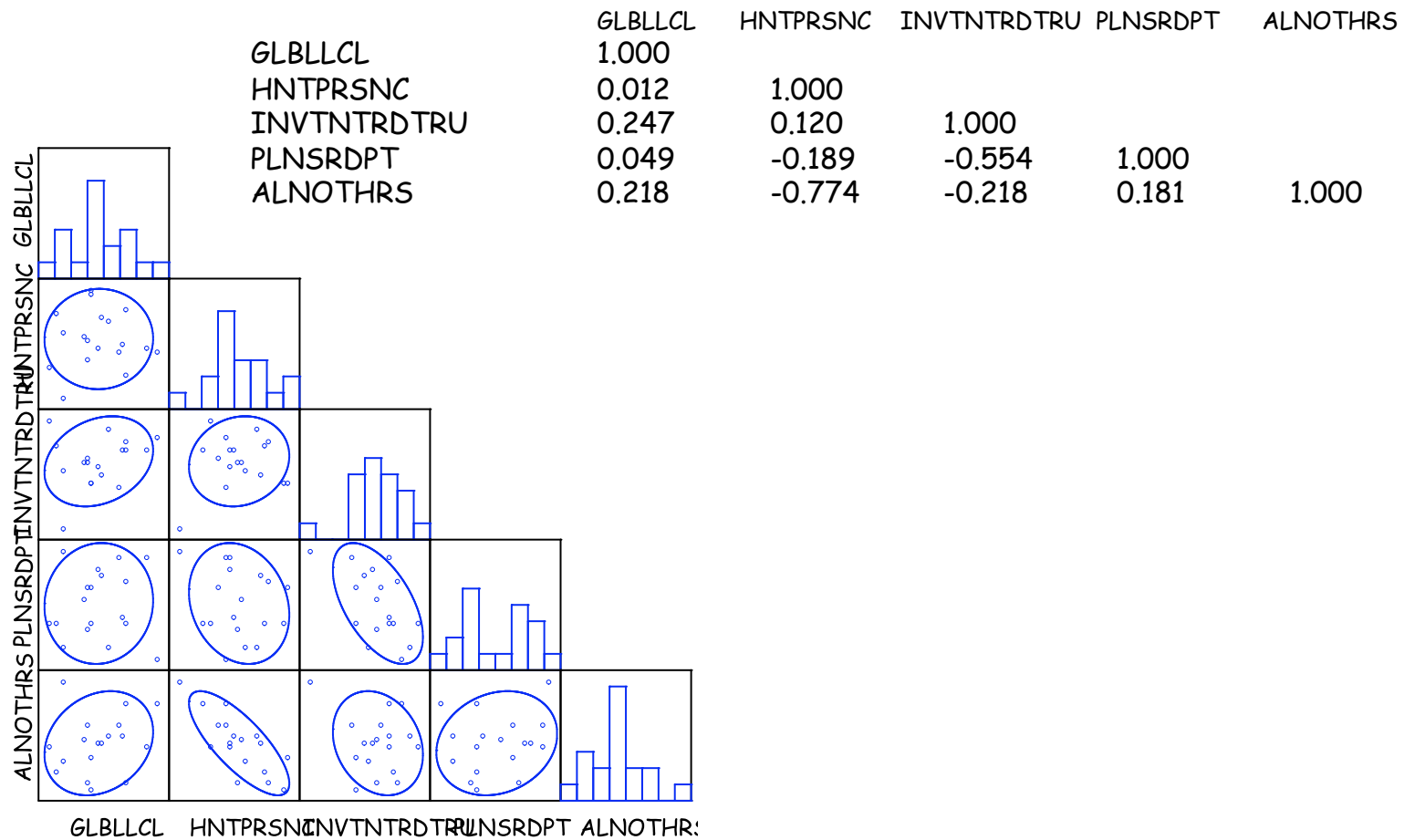
How do people in the class cluster?

How do Profiles cluster?

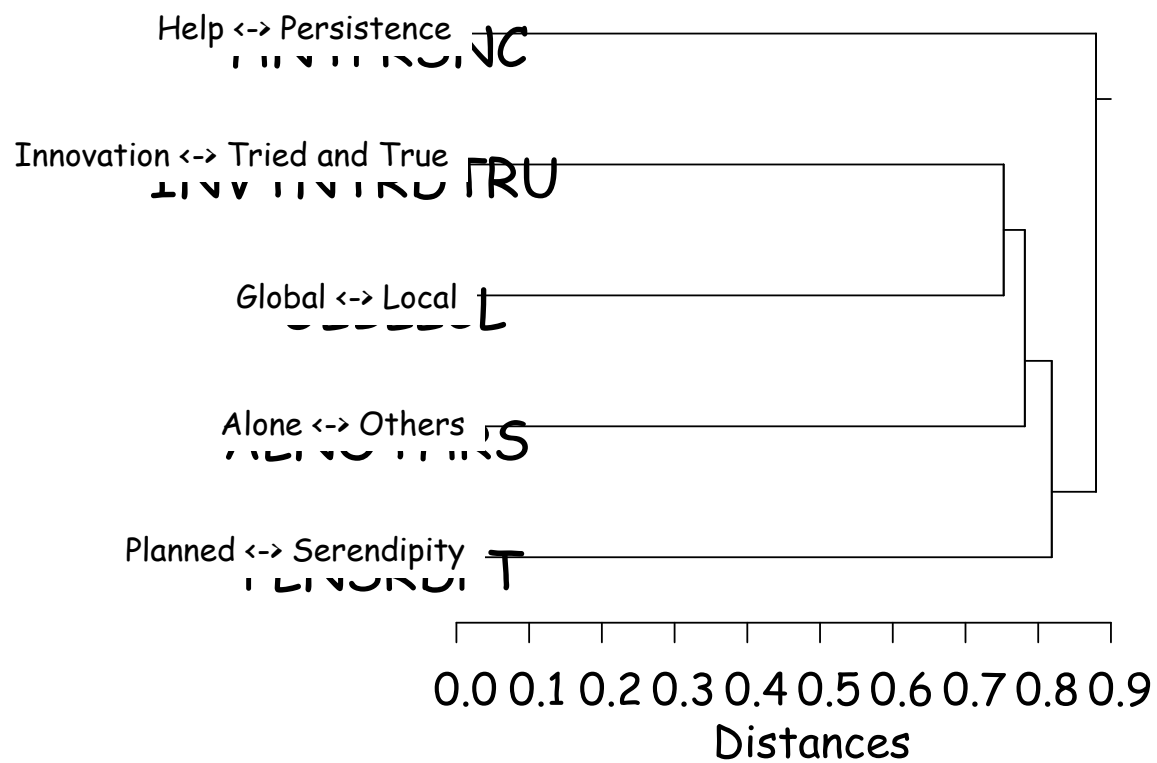
Profiles

- ☐ Global <-> Local
- ☐ Alone <-> Collaboration
- ☐ Help <-> Persistence
- ☐ Innovation <-> Tried
- ☐ Plan <-> Serendipity

Correlation matrix of the variables

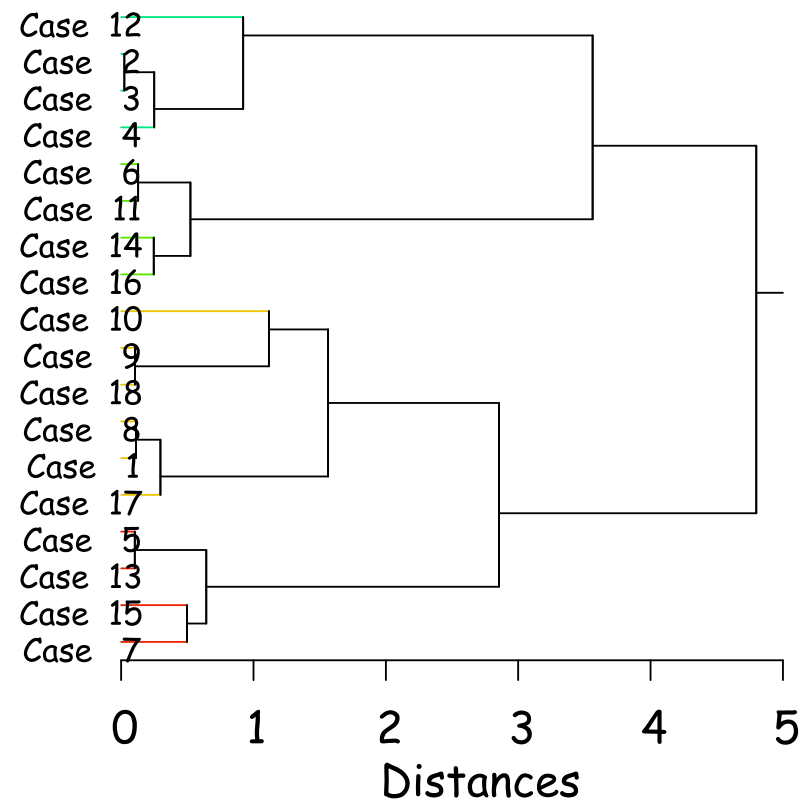


Cluster Tree

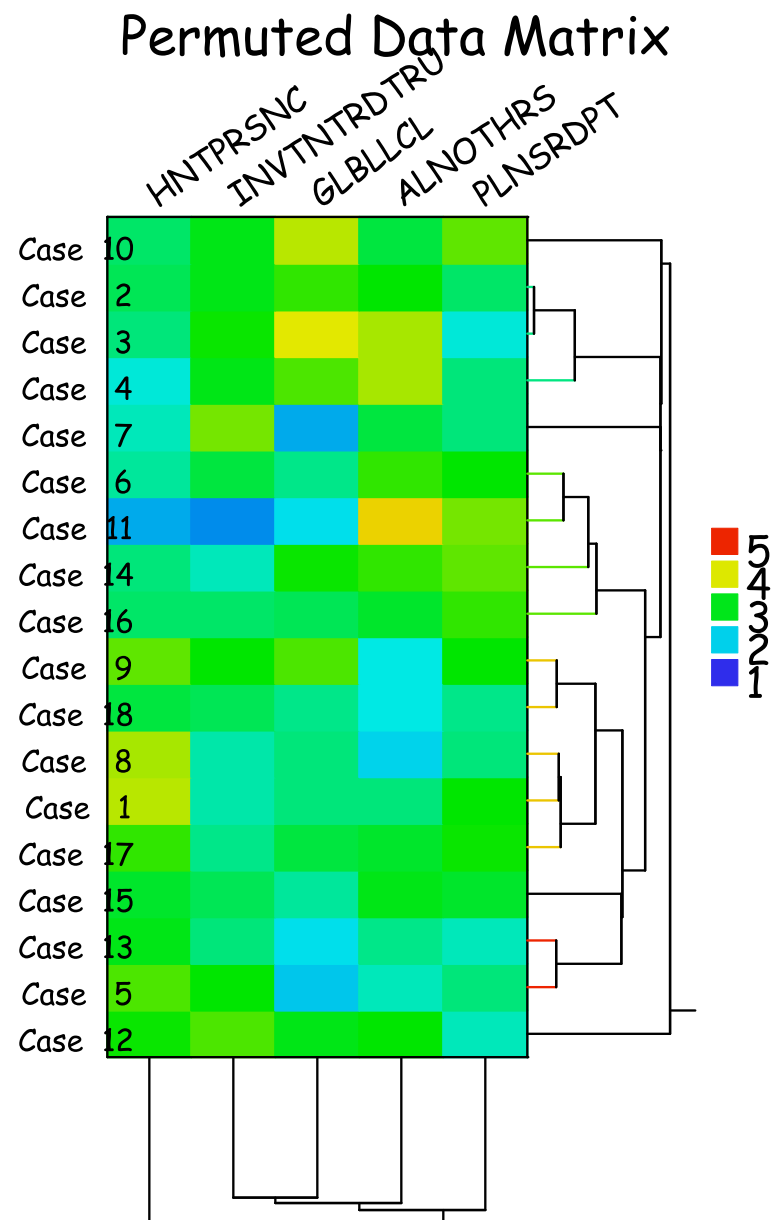




Cluster Tree



Join command (Systat only)

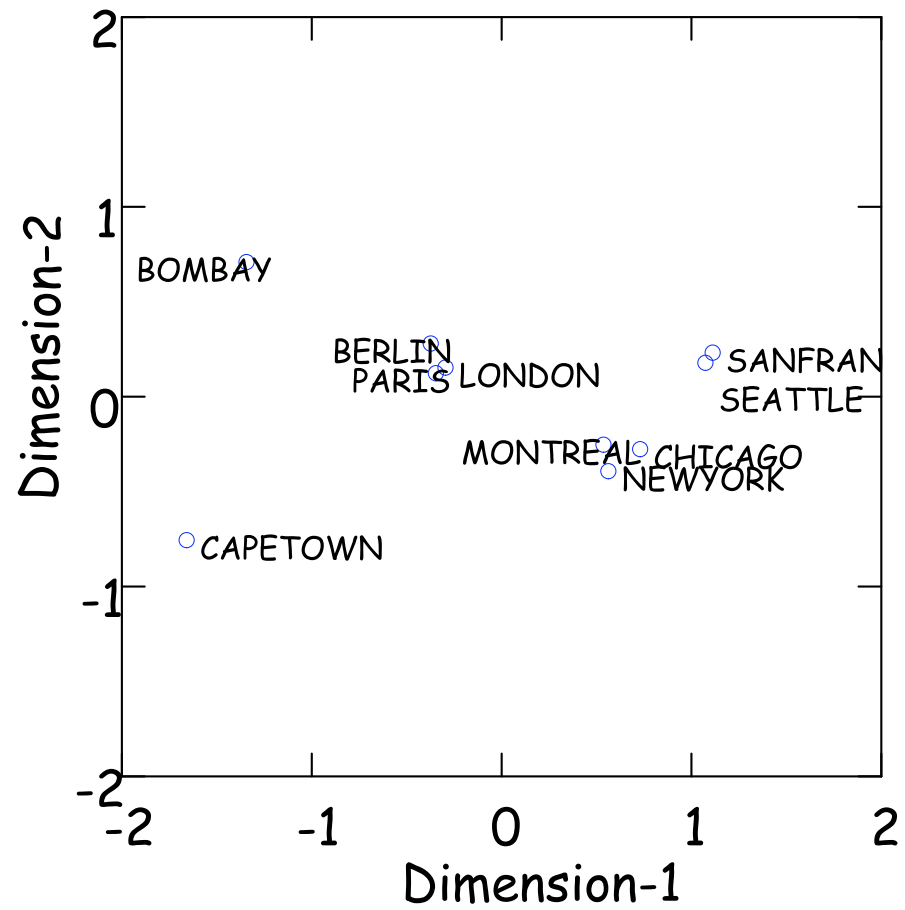




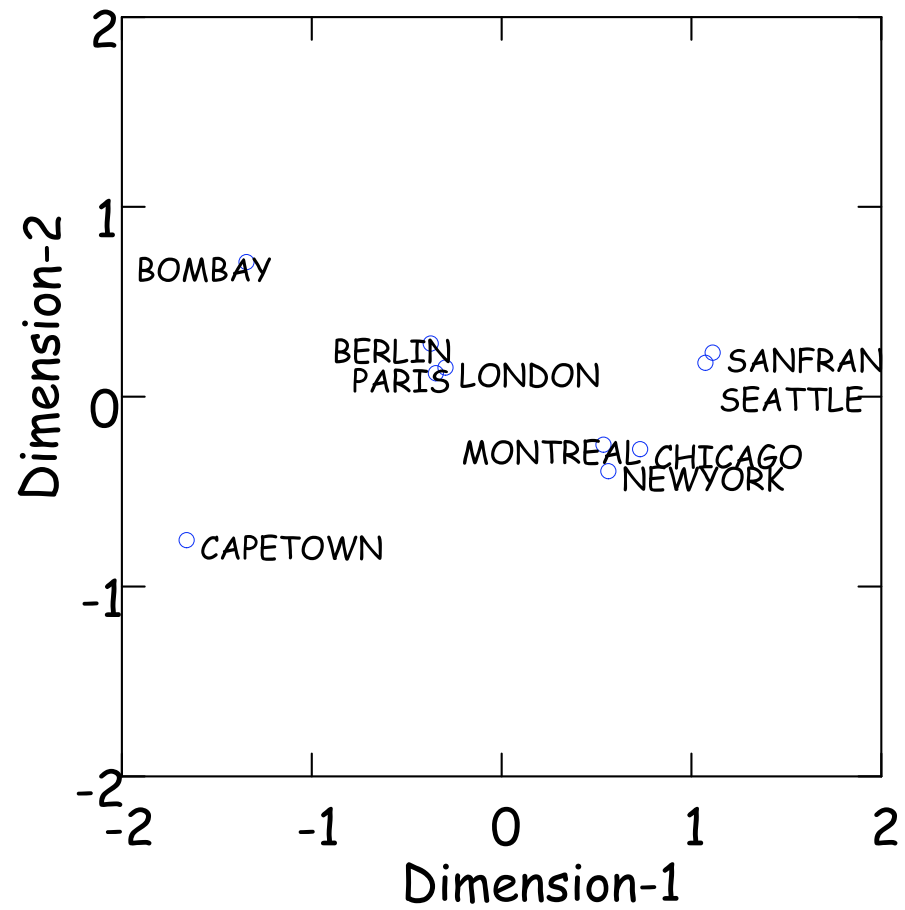
Multidimensional Scaling

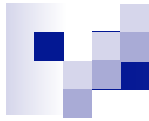
- Multidimensional Scaling is a method to fit a set of points in space that best represents the dissimilarity between all the points.

Configuration



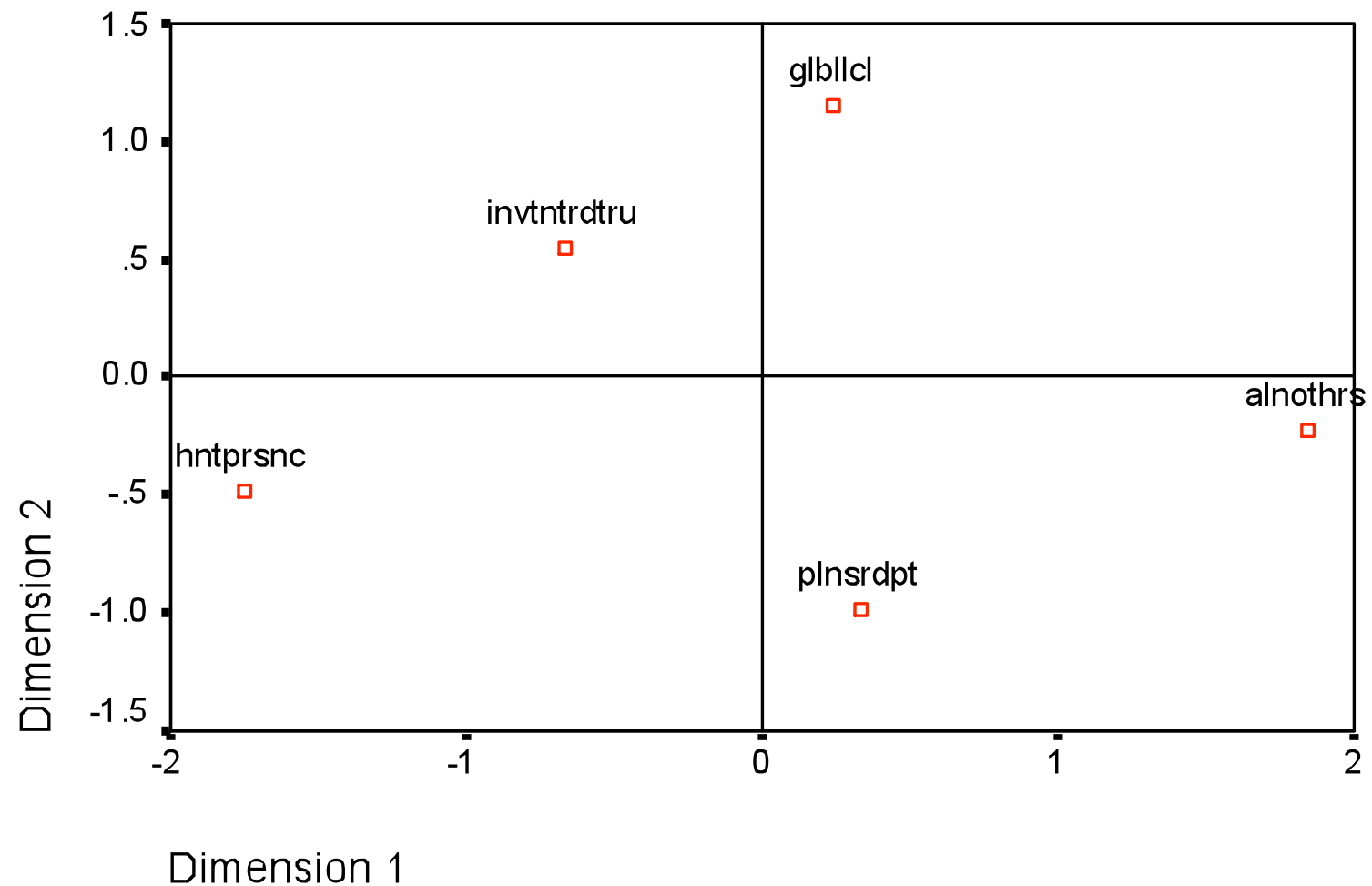
Configuration





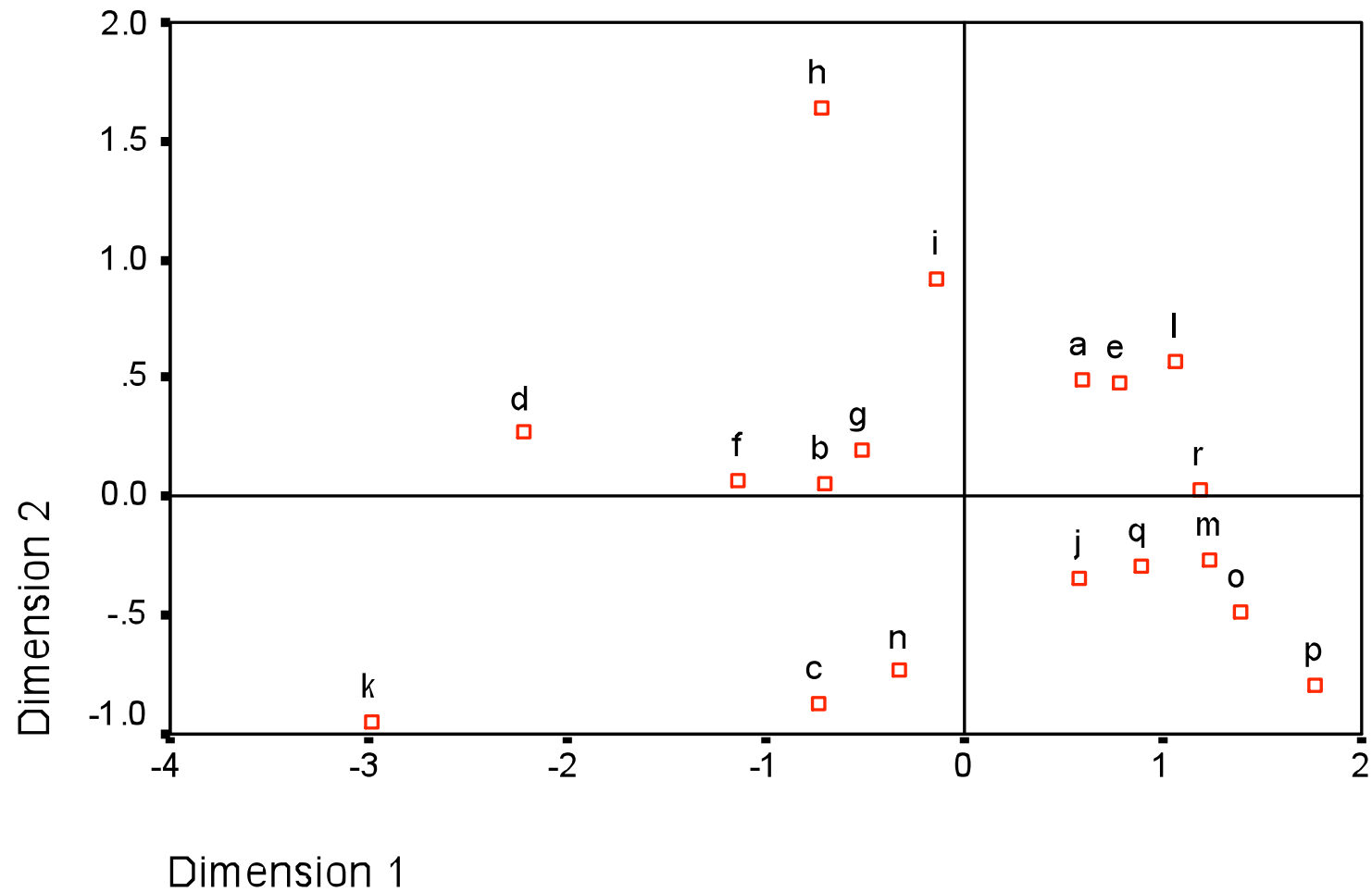
Derived Stimulus Configuration

Euclidean distance model



Derived Stimulus Configuration

Euclidean distance model



Scatterplot of Linear Fit

Euclidean distance model

