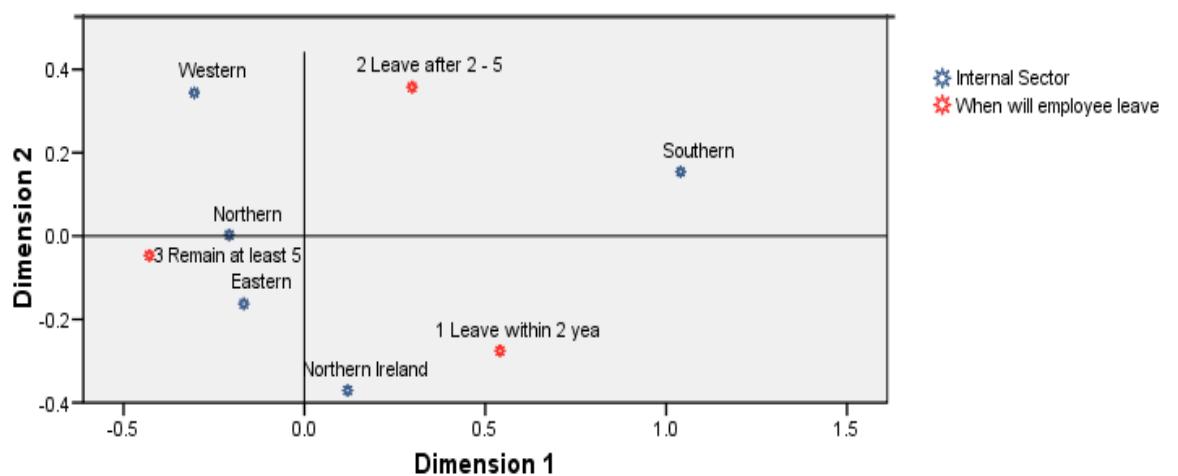


## **IBM SPSS Categories**

- IBM SPSS Categories provides a number of algorithms based on a family of techniques called *optimal scaling*. Optimal scaling attempts to *quantify* the category groups of categorical fields i.e. assign numerical values to the categories as if they existed on a scale.
- There are already a number of methods for analysing categorical data as well as lots of modifications to modelling techniques usually employed in the analysis of numerical data. However, these approaches often do not perform well for datasets that feature for example:
  - Too few observations
  - Too many variables
  - Too many values per variable
- By quantifying categories, optimal scaling techniques avoid many of these problems and can be used as excellent exploratory tools when modelling multivariate categorical data.
- The analytical procedures available within IBM SPSS Categories are as follows:
  - Categorical Regression: This technique is used to predict the values of a nominal dependent variable from a group of categorical independent variables.
  - Categorical Principal Components Analysis: This form of PCA can be used to account for patterns of variation in a single set of mixed variables.
  - Nonlinear Canonical Correlation Analysis: This procedure is used to assess the extent to which two or more sets of variables of mixed types are correlated.
  - Correspondence Analysis: This technique is used to analyse two-way crosstabs or data that can be expressed as a two-way table, such as brand preference or demographic choice data.
  - Multiple Correspondence Analysis: This procedure can be used to analyse the relationships between a number of nominal categorical variables .
  - Multidimensional Scaling: This technique can be used to analyse proximity data to find a least-squares representation of a single set of objects in a low-dimensional space.
  - Multidimensional Unfolding: This technique can be used to analyse proximity data to find a least-squares representation of two sets of objects in a low-dimensional space.



- Using Correspondence analysis to quantify the two categorical fields '*Internal Sector*' and '*When will employee leave*' helps us to explore the relationship between the two variables as if they were continuous fields in a scatterplot