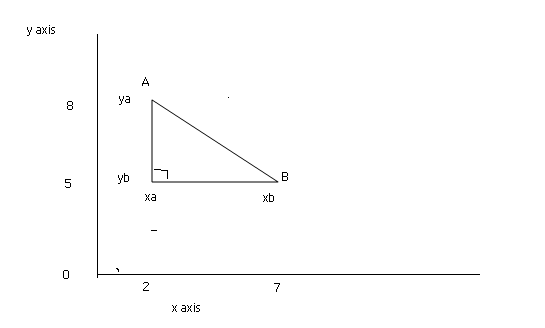
**Constructing perceptual maps with the aid of SOLVER in Office 2007**

Here we look at how one can use a standard spreadsheet such as Microsoft Excel © to produce a two dimensional and we also look at how such a map may be interpreted.

Manly (1986) outlines the procedure for classical multidimensional scaling indicating that it is first necessary to obtain a matrix of distances between objects that one wants to portray graphically. In fact one only needs the diagonal matrix (upper or lower) since the diagonal matrices are identical. The diagonal matrix of distances is the same as that we customarily scrutinize in the front or back of motoring map books to ascertain distances between towns or cities. Since we have decided on a two dimensional representation of the data we next have to set up a starting configuration for the *n* objects in the two dimensions (ie. Co-ordinates xn,, yn are arbitrarily selected for each object).The next step involves calculating the Euclidean distances between the objects. In simple terms this means that we assume that the distances between points in the representation that has been set up are in fact the hypotenuse side of a right angled triangle in which the distances along the two remaining sides can readily be assessed from inspection of the graph we have constructed. For example suppose we want to find the distance from point A to point B. Inspection of the graph shows us that the co-ordinates for A and B are (ya=8,xa=7 for A, and yb=5, xb=2 for B).

Figure 1. The Principles behind measuring Euclidean distance.



Since the triangle formed is a right angled one we can calculate AB by using Pythagorous’s Theorem which indicates that AB=√ (ya-yb)2 +(xb-xa)2  which in this case works out that the distance AB=3.Now from our original diagonal matrix of inter object distances we already know what that distance should be. This the task we now face is adjusting this first arbitrary configuration above until a point is reached where the distance AB coincides with we have in our original diagonal matrix. It is fairly obvious that this can be achieved in an almost infinite number of ways within the two dimensional vector space since in the first instance the length of the non-hypotenuse sides of the triangle can be adjusted so that the desired length of AB is achieved and the triangle itself can then be moved about to an infinite number of positions within the two dimensional vector space as long as the *adjacent* and *opposite*  sides of the triangle are maintained parallel to the vertical and horizontal axes. Of course one does not necessarily need to keep the *opposite* and *adjacent* sides parallel to the vertical axes as long as one maintains the position of the right angle opposite the line AB. However, for the purpose of the illustrations here we will assume that the adjacent and opposite sides will always be maintained parallel to the X and Y axes in two dimensional vector space since this facilitates calculations.

The complications arise when we add further objects into the graph and repeat the above exercise for each of the new data points we enter in relationship to other data points in the graph. It is a relatively simple matter to calculate the Euclidean distances in the arbitrarily constructed graph but the subsequent adjustment of the various data points so that inter-point distances coincide with those in the indicated original diagonal matrix of distances is a much more difficult and complex activity to carry out when there are more than two data points to consider. \effectively this means that however we arrange the data points within the graph there will always be a difference between the actual values in our original diagonal matrix and the inter-point distances reflected and measured in the graph. We might try thousands of different arrangements but still there will be errors and the best we can achieve is to minimize the cumulative errors in an arrangement and show that as the best representation we can make of the data provided. In other words it would be a trial and error or iterative process of finding the best cumulative error minimising arrangement. The errors would arise because the distances between points would in reality need to be represented in more than two dimensions in order for the cumulative errors to be zero.

Manly (op.cit) points out that the goodness of fit between the configuration distances and the disparities might be measured by a suitable statistic called he “stress formula 1”. That is:

STRESS 1 = {∑(dij – δ ij )2 /∑δ2ij} ½

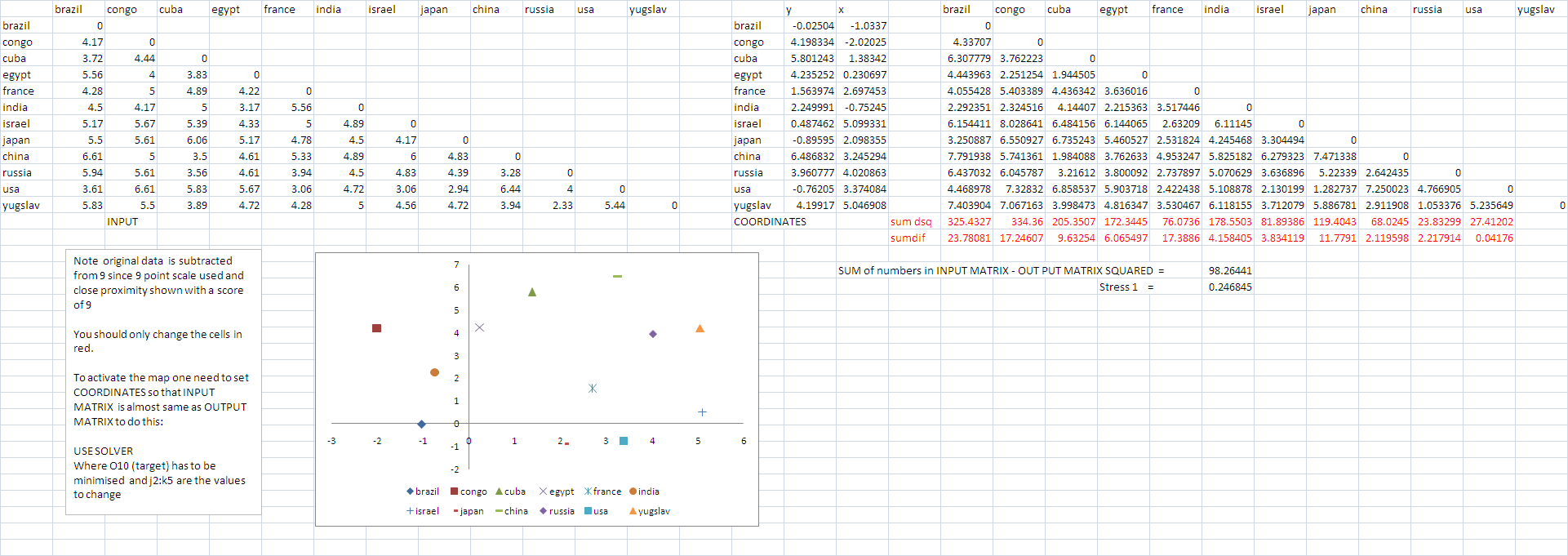
Where dij is the distance between objects i and j in the diagonal matrix and δ ij is the distance as shown in the graph. In other words a configuration in the graph has to be achieved which will minimize the value of STRESS 1.

Kruskal & Wish (1978) provide an excellent overview of the different approaches that can be taken and as is noted at the beginning of the paper Nakanishi and Cooper (op.cit.) provide a critical review and update. Here we will suggest an approach that makes use of the built in facility within Microsoft Excel to find solutions to problems without knowing the exact nature of the input data but knowing the solution sought and the various equations that are required to find the solution. Such an approach makes use of the built in SOLVER add-in which is part of the Microsoft package.

Kruskal and Wish (ibid, p. 30-2) illustrate how using the KYST (Kruskal JB, Young F W & Seery J B (1973)) MDS program one can produce a two dimensional map of perceptions nations. The data used in the original study (reflecting mean scores of 18 respondents’ perceptions of overall similarity between twelve nations on a scale ranging from 1 for “very different” to 9 for “ very familiar” ) was ordered as a diagonal matrix of 66 pairs and processed using KYST. A two dimensional perceptual map was produced and the axes rotated by 45 degrees to facilitate interpretation although the latter is not an essential procedure to conduct.

Figure 2 shows how, using the Excel package, the same analysis can be duplicated but the resultant map is not rotated though this could easily be done by simply superimposing rotated axes on a printout of the map. It should also be noted that because the original scale showed a high score of 9 showing close similarity and a low score of 1 showed dissimilarity the scores in the original data have had to be subtracted from 9 to provide appropriate data for illustration of the approach to MDS adopted here. Using the methodology adopted here similarity is always reflected by a low number and dissimilarity by a correspondingly higher figure.

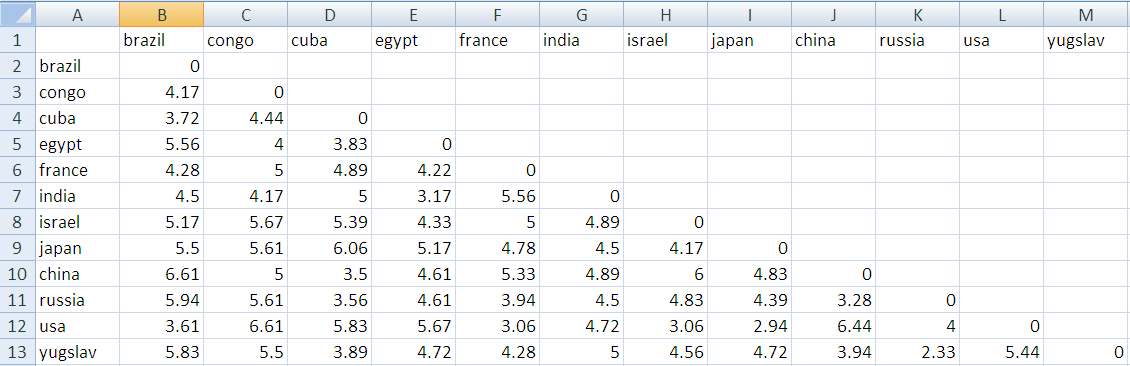
Figure 2 Data after manipulation by Excel formulae and use of Solver



Let us look at how the data is entered and processed in some detail.

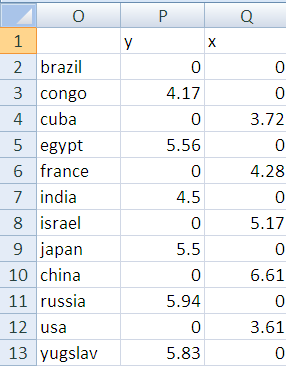
Step 1 is to enter the initial similarity data into the spreadsheet as a diagonal matrix. This is shown in figure 3.

Figure 3 Initial data entry of the 66 pairs of similarity data

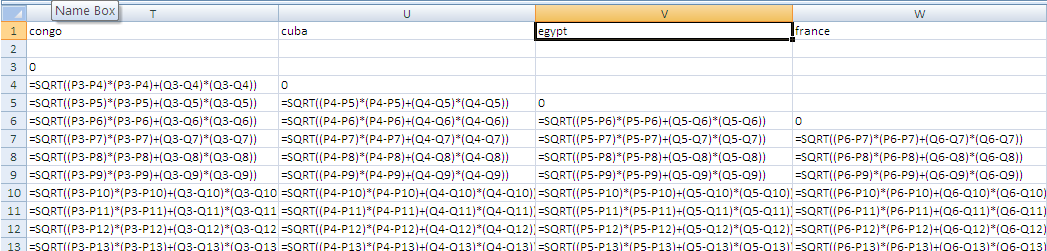


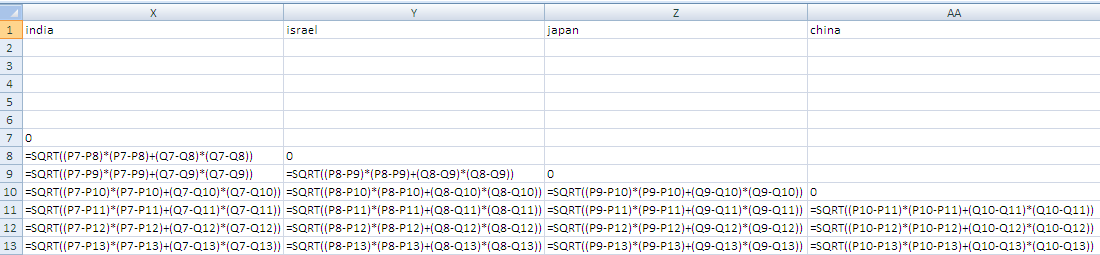
Step 2 involves setting up two columns which eventually will contain the actual co-ordinates for the perceptual map data. These columns with data for an arbitrary configuration are shown in figure 4. It may be noted that the data in column b in the figure 3 has been used to populate the initial arbitrary configuration.

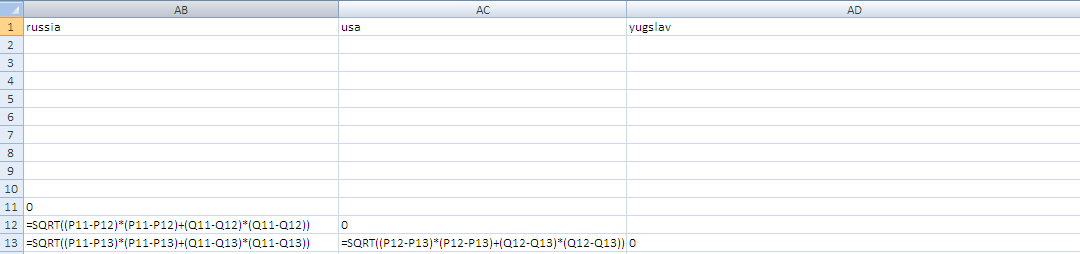
Figure 4 arbitrary initial configuration for the co-ordinates of the map.



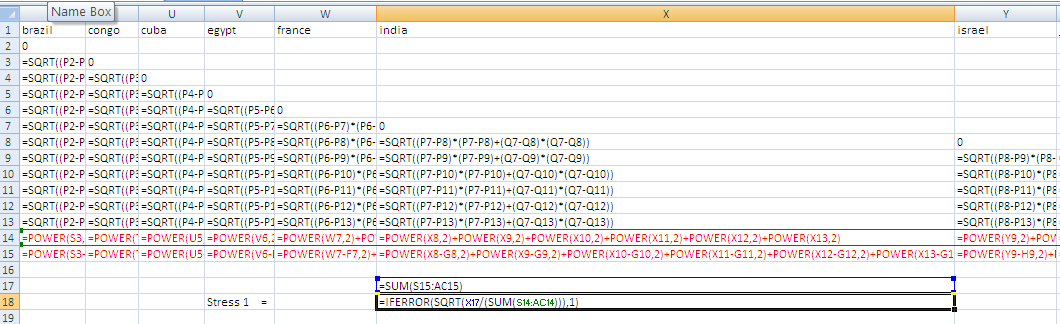
Step 3 involves entering the formulae which will enable the co-ordinates of the map to be calculated when using the standard SOLVER add-in. These formulae for the first four columns are shown in figure 5.







Step 4 involves entering the formulae which calculates the value for STRESS 1. An insight into how this is done is shown in figure 6.



In the case of calculating the stress values the formulae entered into some of the cells is long in some cases and there are other ways of managing the required calculations which may be more economical on the space available.

Step 7 is to activate SOLVER so that the coordinate values shown in figure 5 are optimised – i.e. so that the STRESS 1 value in cell X18 is minimised. Figure 7 and 8 show the values that need to be entered into SOLVER to do this.

Figure 7

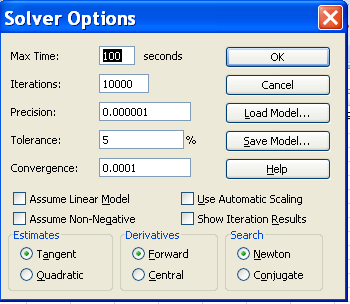
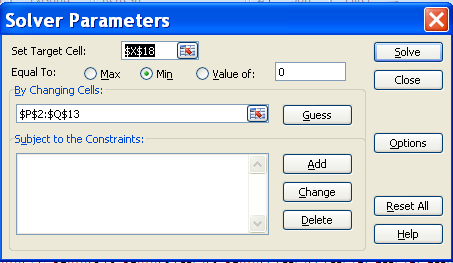
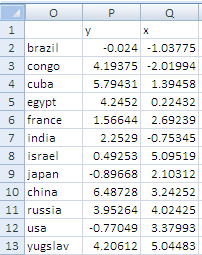


Figure 8



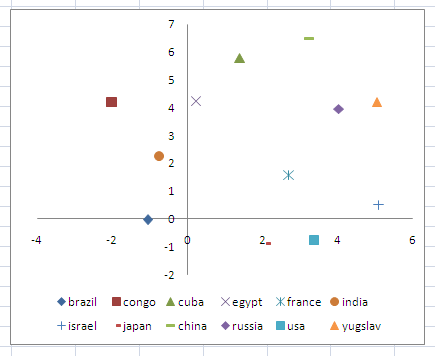
The above actions will lead to the co-ordinates for 12 countries to replace the numbers shown in figure 4 – viz.

Figure 9



These values can then be plotted as either a scatter graph (as shown in this paper) or alternatively as a bubble graph. The map for the above data is shown in figure 10.

Figure 10 Perceptual Map



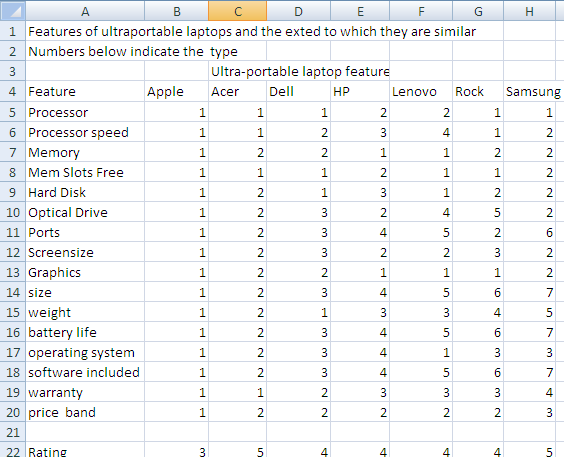
Using KYST Kruskal et al. indicated a STRESS 1 value of 0.19 was obtained whereas using this method the value was slightly higher at 0.24. STRESS 1 can be considerably reduced of course by showing the map in two or three dimensions. While the STRESS 1 value is higher than that obtained using the KYST program it does present a comparable perceptual map of the data and corresponds well with that obtained without rotating axes.

Another illustration

Having established the methodology and its comparability with the KYST program in terms of producing maps let us now turn for interpretation of perceptual maps of this time to a different data set. This is done because the above map and data set is well discussed by Kruskal and Wish (op.cit.) and is very much dated by the times and circumstance sunder which it was collected. First we will briefly review the data.

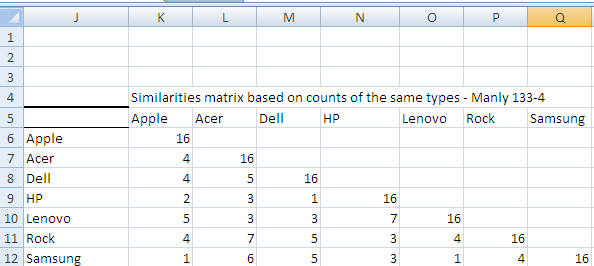
The data used is drawn from : "Group Test: ultra-portable”, Computer Buyer October 2007 p. 66 Issue 197, London: Dennis Publishing and we apply a variation on an approach suggested by Manly (op.cit. page 133-4. Manly suggests that in drawing up a diagonal similarity/dissimilarity matrix we simply count the number of cases for each paired comparison where the data is identical or very similar to produce a similarity matrix. In the Computer Shopper article a comparison is made between various ultra-portable laptop computers in terms of their various features. The aim of producing the perceptual map is to show how the products are positioned with respect to one another with respect to overall differences in the features they possess. Computer Shopper also provides feature ratings (p 59-60) for the machines and the map portrays this rating by the size of the bubble that is shown.

Step 1 involves entering the initial raw data in a form which can next be used to draw up a similarities matrix. This is shown in figure 11.



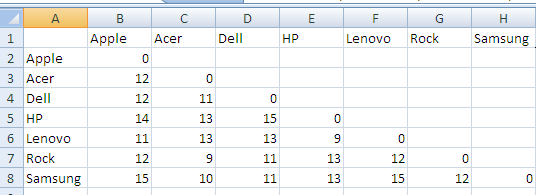
Next a similarities diagonal matrix is constructed by counting the number of cases for each pairing where the types indicated in figure 11 are identical. This is shown in figure 12.

Figure 12 Similarities matrix for computer features offered.



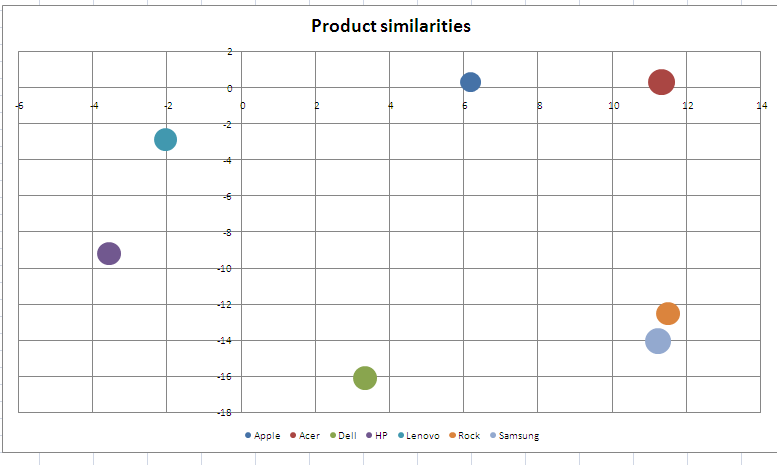
Next we change the matrix values so that they can be used to reflect distances between the paired comparisons by showing the number of dissimilarities rather than the similarities. To do this we subtract each of the above values except those comparing the product with itself away from the number of criteria on which they have been assessed (16). This results in the diagonal matrix shown below in figure 13:

Figure 13 the Input data diagonal matrix



The matrix is then processed in the same way as that shown in the Kruskal example shown previously and produces the following map (figure 14). The STRESS 1 value in this case is 0.1984

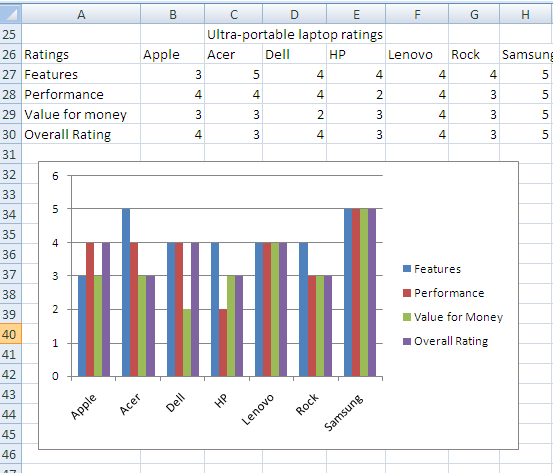
Figure 14 Perceptual map generated with respect to data comparing ultra-portable laptops on 16 criteria.



The map produced in figure 14 shows how closely the products are positioned with respect to one another in terms of the similarity and by the size of the bubble gives and evaluation of their overall rating by the magazine that is reviewing them.

Interestingly we might note that whereas Samsung is rated the highest of the machines and Rock is positioned very close to Samsung, Rock have the same feature rating in the view of the reviewers as Samsung. To explain this we need to return to the article and look at what other data is provided which can be used to help us understand how the products have been rated. Indeed referring back to figure 11 we find that while Rock and Samsung are identical in terms of only four features so that any variation in terms of feature rating must be attributed to variations in ratings of desirability across the remaining 12 features. The data does not provide any data regarding the relative importance of the various features so the mapping has to assign equal importance to the various data.

Figure 15 Ratings of the ultra-portable laptops



The article in Computer Buyer pp (pp 60-66) gives ratings for the machines on features, performance, value for money and an overall rating. These are reported in figure 15 above. The Samsung model is seen to be clearly rated on all criteria superior to the other models. The perceptual map shows how closely products are positioned in terms of having particular features and the relative desirability of these features. It does not, however, take account of the machines performance in terms of these features and the extent to which they provide value for money.

**References**

"Group Test: ultra-portable",Computer Buyer October 2007 pp.59- 66 Issue 197, London: Dennis Publishing

Eckart, C,& Young G (1936), “The Approximation of One Matrix by Another of Lower Rank,”

*Psychometrika*, 1, 211-218.

Kruskal JB & Wish M (1978), Multidimensional Scaling, Sage University Papers on Quantitative Applications in the Social Sciences, 07011. Sage Publications, Beverly Hills.

Kruskal JB, Young F W & Seery J B (1973), “Ho to use KYST, a very flexible program to do multidimensional scaling and unfolding”, Bell laboratories (unpublished).

Manly B F J (1986), Multivariate Statistical Methods: A Primer, London: Chapman & Hall.

Nakanishi M & Cooper L G (2003), Metric Unfolding Revisited: Straight Answers to Basic Questions, Department of Statistics, UCLA, Department of Statistics Papers, Paper 2003010112, posted at the eScholarship Repository, University of California. http://repositories.cdlib.org/uclastat/papers/2003010112

Torgerson, Warren S. (1958), *Theory and Method of Scaling*, New York: Wile