MULTIDIMENSIONAL SCALING: AN INTRODUCTION

William G. Jacoby
Michigan State University
jacoby@msu.edu

David J. Ciuk
Reed College
ciukd@reed.edu

October 2014

The term, "multidimensional scaling" (often abbreviated "MDS") refers to a family of procedures used to construct geometric models of sets of objects, using information on proximities between these objects. The term "objects," in this context, refers to whatever the researcher is examining. For a psychologist, the objects might be individual test questions drawn from an item pool. For a market researcher, the objects might be consumer products (e.g., brands of cereal). For a sociologist, the objects might be different occupations. For a political scientist, the objects might be candidates for public office. Of course, the list of possible object sets goes on endlessly.

The term "proximities" refers to any kind of symmetric relationship defined for pairs of the objects. Among the most comprehensible of these symmetric relationships is similarity—information about how similar or dissimilar each pair of objects are to each other. For the moment, we will assume that this information is available for every distinct pair of objects that can be created from the full set of objects.

In the "geometric model" each object is represented as a point within a space, and the proximities between the objects are shown as distances between the points. Specifically, more similar pairs of objects are shown as points separated by smaller distances. Conversely, more dissimilar pairs of objects are shown in the model as points separated by larger distances. Thus, the results from a multidimensional scaling analysis consist of a spatial array of points, where the distances between the points correspond (as closely as possible) to the proximities between the objects. The results from the MDS analysis are most useful when the dimensionality of the space containing the scaled points is small (no more than two or, at most, three) such that the researcher can look directly at the "cloud" of points produced in the solution.

In sum, multidimensional scaling generates a fairly intuitive and readily comprehensible representation of proximities information. But, why would a researcher want to create such a model? The typical answer is that he or she is trying to determine whether there is any underlying systematic structure that exists across the object set. Do the points that lie close

to each other within the scaled space represent objects that share common characteristics? Do points that are widely separated from each other correspond to objects that differ from each other in some identifiable way? Answering these kinds of question provides insights about the nature of variability across the objects. That is, it helps the researcher understand *why* some objects are more similar, and others are less similar, to other objects. MDS facilitates this process because the spatial configuration of points generated by the scaling analysis is usually easier to comprehend than the original, numeric, information about the inter-object proximities.

Multidimensional scaling possesses several important strengths that make it a very useful tool for social and behavioral research:

- MDS is an important data reduction strategy. This can be useful for exploring large datasets, but it is also very relevant for testing substantive theories whose elements can be represented in spatial form.
- In several fields, MDS can be useful for modeling the perceptual structures that underlie survey respondents' or experimental subjects' preference responses toward interesting stimuli. This was probably the main motivation behind the original development of multidimensional scaling methods. The general idea is that better understanding of the ways people think about a set of objects will provide important insights about their reasons for liking and disliking the respective objects.
- Multidimensional scaling methods are extremely flexible with respect to the input data. Many different types of information can be used to create proximities data. At the same time, MDS methods exist for varying levels of measurement, from ratio through interval and ordinal, down even to nominal-level data (although the latter can be extremely tricky). MDS can also be used to combine proximities information from multiple sources, in several different ways.
- Because MDS output provides at least interval-level information about the distances between objects, it can be used as a tool to improve measurement. While this may not be surprising if the input data are measured at the interval or ratio level, it is also the case when ordinal-level proximities are analyzed. The general idea is that a successful MDS provides insights about how objects differ from each other; it is a relatively small additional step to quantify those differences from the MDS output. But, once this is done, the resultant numeric values "measure" the scaled objects with respect to each other.
- The main output from an MDS analysis is graphical in nature. This is relevant because it provides a convenient way to represent complex information in a relatively compre-

hensible manner. While parsimony is valued in scientific communities, this feature of MDS is especially valuable when analytic results must be presented to lay audiences.

MULTIDIMENSIONAL SCALING: A GENERAL FORMULATION

In order to understand multidimensional scaling, it is useful to think in terms of a map. All of us who travel are familiar with using a map. But, the map itself is a geometric model of the "proximities" (i.e., physical distances) between points representing a set of objects (i.e., we will assume they are cities, but they could represent other geographic features). By measuring the size of the interval between two points on the map, and comparing that to the map's scale, we can determine the distances between the two cities represented by the points.

Multidimensional scaling "reverses" the preceding task. We start with the intercity distances, perhaps conveniently arrayed into a triangular table. These comprise our numeric proximities data. We use the information in these proximities to produce the map, or the geometric model in which the actual distances between cities are represented as (much smaller) distances between points on a two-dimensional surface. That is, basically, all there is to it!

Let us next move to a slightly more formal representation of multidimensional scaling. Begin with a $k \times k$ square matrix, $\mathbf{\Delta}$, of "proximities" among k objects. The proximity between the object represented by the i^{th} row and the object represented by the j^{th} column is shown by the cell entry, δ_{ij}. In the vast majority of situations, $\mathbf{\Delta}$ is a symmetric matrix, implying that $\delta_{ij} = \delta_{ji}$.[1]

Greater proximity between objects i and j corresponds to a *smaller* value of δ_{ij}, and vice versa. Therefore, the proximities are often called "dissimilarities." Admittedly, this terminology is a bit confusing! But, it leads to a more straightforward characterization of the scaling problem.

Multidimensional scaling tries to find a set of k points in m-dimensional space such that the distances between pairs of points approximate the dissimilarities between pairs of objects. More specifically, MDS uses the information in $\mathbf{\Delta}$ to find a $k \times m$ matrix of point coordinates,

$\mathbf{X}$. The distance between the points representing objects i and j, d_{ij}, is calculated from the entries in the i^{th} and j^{th} rows of $\mathbf{X}$, usually using the familiar Pythagorean formula:

$$d_{ij} = \left[(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \ldots + (x_{im} - x_{jm})^2\right]^{1/2} \tag{1}$$

We want MDS to find $\mathbf{X}$ such that $d_{ij} \approx \delta_{ij}$ for all pairs of objects, i and j. Stated a bit differently, multidimensional scaling uses the proximities contained in $\mathbf{\Delta}$ to find the point coordinates (contained in $\mathbf{X}$) such that the interpoint distances are functionally related to the pairwise dissimilarities. So, for all pairs, composed of objects i and j, with $i \neq j$:

$$d_{ij} = f(\delta_{ij}) + e_{ij} \tag{2}$$

In expression 2, the f represents a function, or a specific rule for transforming dissimilarities into distances. The nature of the function (i.e., the actual contents of the tranformation rule) is determined by the type of MDS that is performed on the dissimilarities data. The e_{ij} term represents an "error" that also contributes to the distance; its presence in this functional relationship implies that the scaled distances do not need to be a deterministic function of the input dissimilarities. For the moment, let us assume that the number of dimensions, m, is known prior to the analysis.

In a multidimensional scaling analysis, the researcher begins by specifying the nature of the function, f, that maps from dissimilarities to distances. This is tantamount to selecting a specific type of MDS, as we will see. Once the type of analysis is selected, the scaling procedure, itself, calculates the values of the parameters associated with f. That, in turn, enables us to move from the numeric information in the $\mathbf{\Delta}$ matrix of dissimilarities to the $\mathbf{X}$ matrix of point coordinates. Finally, we use the information in $\mathbf{X}$ to plot the points, and generate a graphical representation of the model. If we were to calculate the distances between all pairs of points in the graphical display, they could be placed into a square matrix of order k that we could designate as $\mathbf{D}$. There is a one-to-one correspondence between the

elements in $\mathbf{\Delta}$ and the elements in $\mathbf{D}$. And, MDS always includes a fit measure which summarizes in a single numeric value how close the correspondence is between the elements of these two matrices. Alternatively, the fit measure shows how closely the transformation from dissimilarities to distances comes to being a deterministic function, in which the error terms are all equal to zero.

So-called Classical Multidimensional Scaling (or CMDS) uses information from a single set of pairwise dissimilarities among objects to generate a spatial configuration of points representing those objects. This general category actually subsumes a number of specific scaling methods that vary mainly according to the assumptions they make about the nature of the input data. The major distinction is between *metric* MDS and *nonmetric* MDS. The former, metric MDS, assumes that the input data are measured at the interval or ratio level. Nonmetric MDS only assumes that the input data provide an ordering of the dissimilarities between the objects; hence they are only measured at the ordinal level. There are two very important caveats that must be raised about the distinction between metric and nonmetric MDS. First, the output from the MDS is *always* metric, in the sense that it produces a map in which the distances between the object points are measured at the interval or ratio levels, regardless of the assumptions about the metric or nonmetric nature of the input data. Second, it always is the researcher that decides whether the input data are measured at the ordinal, interval, or ratio level; there is nothing intrinsic to any dataset that determines the measurement level.

METRIC MULTIDIMENSIONAL SCALING

Metric MDS requires that the distances in the scaling solution are related to the dissimilarities by a parametric function, usually linear in form. For example:

$$d_{ij} = a + b\ \delta_{ij} + e_{ij} \tag{3}$$

To reiterate, δ_{ij} is the dissimilarity between the i^{th} and j^{th} objects, while d_{ij} is the distance

between the points representing objects i and j in the multidimensional scaling solution. The a and b terms are coefficients to be estimated, and e_{ij} is an error term that is associated with objects i and j. If the a coefficient is constrained to be zero, then the input data are assumed to be ratio-level. If a is permitted to be a nonzero value, then the input data are interval-level.[2]

A well-known scaling procedure, developed by Warren S. Torgerson (1958), begins with the k by k dissimilarities matrix, $\mathbf{\Delta}$. We assume that the dissimilarities are *equal to* the distances between the points for the k objects in m-dimensional space, except for random error. The k by k matrix of interpoint distances is $\mathbf{D}$, and the k by k matrix, $\mathbf{E}$ contains random errors. The main hypothesis of the MDS analysis is that $\mathbf{\Delta} = \mathbf{D} + \mathbf{E}$. Stated informally, the objective of the analysis is to find the $k \times m$ coordinate matrix, $\mathbf{X}$, such that the entries in $\mathbf{E}$ are as close to zero as possible.

Torgerson's procedure creates a "double-centered" version of $\mathbf{\Delta}$, designated $\mathbf{\Delta}^*$. It does so by squaring all the entries in the original $\mathbf{\Delta}$, and then transforming the resultant squared matrix so that the row sums, the column sums, and the overall sum of the cell entries in the final double-centered matrix are all zero. For dissimilarity δ_{ij}, Torgerson proved that the corresponding entry in the double-centered matrix can be calculated as follows:

$$\delta^*_{ij} = -0.5(\delta^2_{ij} - \delta^2_{i\cdot} - \delta^2_{\cdot j} + \delta^2_{\cdot\cdot}) \tag{4}$$

Where $\delta^2_{i\cdot}$ is the mean of the entries in the i^{th} row of the squared dissimilarities matrix, $\delta^2_{\cdot j}$ is the mean of the entries in the j^{th} column of the squared dissimilarities matrix, and $\delta^2_{\cdot\cdot}$ is the mean of all entries in the squared dissimilarities matrix. The double-centered matrix, $\mathbf{\Delta}^*$, can be factored to obtained the point coordinates for the scaling solution. The factoring process is carried out by performing an eigendecomposition on $\mathbf{\Delta}^*$:

$$\mathbf{\Delta}^* = \mathbf{V}\mathbf{\Lambda}^2\mathbf{V}' \tag{5}$$

In equation 5, $\mathbf{V}$ is the $k \times q$ matrix of eigenvectors, $\mathbf{\Lambda}^2$ is the $q \times q$ diagonal matrix of eigenvalues, and q is the rank of $\mathbf{\Delta}^*$ (usually equal to k). Next, we create $\mathbf{X}$ from the first m eigenvectors ($\mathbf{V}_m$) and the first m eigenvalues ($\mathbf{\Lambda}_m^2$):

$$\mathbf{X} = \mathbf{V}_m \mathbf{\Lambda}_m \tag{6}$$

And, the matrix of interpoint distances is created from $\mathbf{X}$ as follows:

$$\mathbf{D} = \mathbf{X}\mathbf{X}' \tag{7}$$

Torgerson proved that the $\mathbf{X}$ matrix contains point coordinates such that the distances between the points (i.e., the entries in $\mathbf{D}$) have a least-squares fit to the entries in $\mathbf{\Delta}$. In this sense, the metric MDS procedure produces the best-fitting set of points for the dissimilarities within the specified dimensionality of the scaling solution.

Substantive Example of Metric MDS: Constructing a Map

As the first substantive example of MDS, we will build upon the map analogy described earlier. Table 1 shows the driving distances (in thousands of miles) between ten American cities. Thus, the table can be interpreted as a dissimilarities matrix for the cities, in which "dissimilarity" is operationalized as physical distance. We will apply metric MDS to this information in order to generate a map in which cities are shown as points and the distances between the points are, to the greatest extent possible, exactly proportional to the dissimilarities. Of course, this task is a bit trivial in terms of its substantive importance (after all, perfectly adequate maps of the U.S. are already available). But, it provides an excellent example for metric multidimensional scaling because: (1) we already know the dimensionality of the solution; and (2) we already know the true "shape" of the point configuration. Thus, we are in a good position to evaluate how well the metric MDS routine actually works.

As a preliminary caveat, note that Table 1 contains 100 cells. But, there are really only 45 interesting pieces of information, or entries in the input data for the MDS. First, the entries

in the main diagonal are all zero (because an object is not at all dissimilar to itself, by definition). Second, the matrix is perfectly symmetric, so the upper-right triangle contains the same entries as the lower-left triangle. With 10 objects (cities, in this case), there are 45 distinct pairs, each of which is associated with a dissimilarity. Those 45 dissimilarities comprise the actual data that are analyzed in the metric MDS.

Since this is the first example, we will show the results from all of the steps. Table 2 shows the double-centered version of the dissimilarities matrix— this is obtained by squaring the entries from Table 1, calculating the row, column, and grand means of the new matrix, and applying the transformation from equation 4 to all of the cells. We will obtain a two-dimensional MDS solution (thus ignoring any potential effects due to the curvature of the earth). So, Table 3 gives the first two eigenvectors and eigenvalues for the double-centered dissimilarities matrix. Each eigenvector is multiplied by the square root of the corresponding eigenvalue to produce the "rescaled" eigenvectors shown in Table 4. These are also the point coordinates for the cities, so each row of this last matrix is labelled accordingly.

The point coordinates can be plotted in two-dimensional space, and the results are shown in the first panel of Figure 1. Obviously, we can see that the relative positions of the points approximate those of cities in the U.S. But the horizontal direction looks odd, because eastern cities fall to the left, and western cities fall to the right. While there is nothing intrinsically wrong with this depiction, it contradicts common practice. Fortunately, an eigendecomposition is invariant under a reflection of the eigenvalues. So, we can multiply the elements in the first eigenvalue by -1, thereby reversing the positions of the points on the first (i.e., horizontal) coordinate axis. This reflected set of points is shown in the second panel of Figure 1, and the similarity to a typical map of the U. S. should be obvious.

Here, the MDS solution *seems* to reproduce a regular map very well. But, "eyeballing" the scaling results is not enough to provide a rigorous assessment of how well the geometric model represents the numeric data in the dissimilarities matrix. One way of evaluating the correspondence between the input data and the output scaled interpoint distances is to

construct a scatterplot showing the relationship between the two. Since we are conducting a metric MDS, a good scaling solution would be indicated by a linear array of points. Since we are assuming ratio-level data, the linear trend should pass through the origin and if the coordinate axes are scaled to equal the input data values (as they are in this example) the slope of the linear trend should be 1.0. Figure 2 shows such a scatterplot for the metric MDS of the intercity distances. There are 45 points in the graph, one for each unique pair of cities. The horizontal axis corresponds to the intercity driving distances (the input data), and the vertical axis shows the interpoint distances in the scaling solution. The gray line shown behind the array of points is a line with slope of one and a zero intercept. So, this diagram confirms that the scaling solution represents the input data *very* well. Note that a graphical display plotting the scaled distances against the input dissimilarities data is called "Shepard diagram" after one of the early pioneers of multidimensional scaling, Roger B. Shepard.

The Shepard diagram provides a visual representation of the scaling solution's adequacy as a representation of the input data. But, it is also useful to have a numeric measure of fit. In order to obtain such a measure, we will use the eigenvalues that were produced during the process of calculating the coordinate matrix, $\mathbf{X}$. The sum of all the eigenvalues is equal to the sum of squares in the dissimilarities matrix. And, each eigenvalue gives the sum of squares associated with the corresponding eigenvector (i.e., the first eigenvalue is associated with the first eigenvector, the second eigenvalue is associated with the second eigenvector, and so on). We used the first two eigenvectors to form the scaling solution, so the sum of the first two eigenvalues gives the sum of squares associated with (or "explained by") that solution. Here, the sum of all ten eigenvalues is 11.32; the first two eigenvalues are 9.58 and 1.69, respectively. So, the fit measure is $(9.58 + 1.69)/11.32 = 0.996$. From this value, we can see that the two-dimensional MDS solution accounts for 99.6% of the total variance in the dissimilarities matrix.[3] Obviously, the geometric model produced by the metric MDS strategy fits the data extremely well— almost perfectly. Thus, metric MDS works!

DATA FOR MULTIDIMENSIONAL SCALING

In the previous example of metric MDS, we used physical distances as the input data. So, perhaps it is not so surprising that the output distances in the scaling solution worked so well. But, if multidimensional scaling works for physical distances, then it may also work for data that can be interpreted as "conceptual distances." And, there are many types of data that can be interpreted this way.

Even though we usually call the input data "dissimilarities," MDS can actually handle any kind of data that can be interpreted as a distance function.[4] And, many kinds of information can be interpreted as distances. This leads to one of the strongest features of MDS— its ability to analyze many substantively-different kinds of data.

One obvious type of data for MDS consists of direct dissimilarity judgments. For example, a single individual might indicate his or her perceptions of the similarities of k objects. Or, a researcher might have a set of subjects sort pairs of objects according to their perceived pairwise similarities. Alternatively, the analyst could have subjects or respondents rate the similarity of object pairs on some pre-defined scale and then take the mean of the individual similarity ratings. In any case, the *similarity* judgments are easily converted to dissimilarities by subtracting the "raw" values from some constant, thereby reflecting the data values.

Profile dissimilarity measures assume that each of the k objects to be scaled possesses scores on a common set of characteristics; let v represent the number of these characteristics. The profile dissimilarity between any pair of objects is obtained by taking the sum of the squared differences in the scores of the two objects, across the full set of characteristics. Often, the square root of the sum of squared differences is used; in that case, the profile dissimilarities are the distances between the objects in the v-dimensional space defined by the characteristics. Profile dissimilarities are particularly useful because they convert multivariate data (i.e., a rectangular matrix in which each object is characterized by a vector of scores) to dissimilarities (i.e., a square matrix summarizing the differences of each pair of objects across their respective vectors of scores). Of course, the profiles themselves could

be constructed from an arbitrarily large number (say, v) of scores for the respective objects. An MDS of these data could thus be employed as a strategy to summarize hypervariate v-dimensional data in a space of m dimensions, where m is assumed to be quite small, perhaps only 2 or 3.

Still another kind of information that might be interpreted as dissimilarities are "confusions data." So, if the researcher has information on the degree to which one stimulus is mistaken for another stimulus, it might be interpreted as a representation of the similarity between those stimuli— the assumption, of course, is that more similar objects are more easily confused for each other. Following the same kind of logic, measures of temporal stability could also be interpreted as dissimilarities. For example, assume that a set of objects are sorted into k different categories at each of two time points. The proportion of objects that move from one category to the another over the time interval could be interpreted as the similarity of the two categories. Note that in this latter case, it is the *categories* that are scaled, not the objects that are sorted into the respective categories.

There are a variety of theory-based measures of spatial separation that can be used as input dissimilarities for MDS. For example, the number of times two objects display the same behaviors can be interpreted as the similarity between the objects. The degree to which a pair of objects share common characteristics can be viewed the same way. And, the line-of-sight (LOS) dissimilarity measure developed by Rabinowitz (1976) creates a pairwise dissimilarity measure among a set of k objects using n subjects rating scale scores for those objects.

Finally, correlation coefficients are sometimes used as similarity measures and input for MDS. In fact, this is usually problematic, because the geometric interpretation of a correlation coefficient consists of the angular separation between two variable vectors, and not the spatial distance between two points. By imposing certain assumptions, the correlation coefficient can be transformed into a dissimilarity measure. But, those assumptions are fairly stringent, so it is generally best to seek an alternative, more appropriate, type of dissimilarity

to use as input for multidimensional scaling.

So, the flexibility in the kinds of input data that can be used comprise an important strength of MDS as an analytic strategy. But, they also raise a new concern: In many such cases, we would be hesitant to assume that our dissimilarities comprise interval- or ratio-level data. Instead, we may only have ordinal-level dissimilarities among the objects we seek to scale. In that case, we cannot talk about a specific parametric function (e.g., linear) relating the input dissimilarity values to the scaled interpoint distances. Instead, we could try the following:

$$d_{ij} = f^m(\delta_{ij}) + e_{ij} \tag{8}$$

In expression 8, f^m means "a monotonic function" and e_{ij} is an error term. A monotonic relationship exists if, for all subsets of three objects (say, i, j, and l), the following holds:

$$\delta_{ij} < \delta_{il} \Longrightarrow d_{ij} \leq d_{il} \tag{9}$$

Here we are making a far less stringent assumption about the "translation" from dissimilarity to distance. Now, it is only the *ordering* of the distances that needs to be consistent with the values of the input dissimilarities.[5]

Do we really need to worry about this? Why not simply treat ordinal data as if it were interval- or ratio level? This strategy is used often for other statistical analyses in the social and behavioral sciences (e.g., ordinal dependent variables in regression equations). And, it is often the case that metric MDS analysis of ordinal dissimilarities seems to generate interpretable results. Nevertheless, doing so remains a highly problematic strategy. For one thing, it is "cheating" with respect to the data characteristics. Treating ordinal data as if they were interval-level (or higher) imposes an implicit but extremely stringent assumption about the relative sizes of the differences between the dissimilarities. In addition, the concept of "variance" is undefined for ordinal data. Therefore, it is inappropriate to use the eigendecomposition, which maximizes the variance explained by successive dimensions in the

scaling solution. For these reasons, it is better to use an entirely different strategy for MDS with ordinal dissimilarities data.

NONMETRIC MULTIDIMENSIONAL SCALING

Nonmetric MDS is used to construct a geometric model of dissimilarities data, relying strictly on the ordinal properties, rather than the actual numeric values, of the data values. The general strategy is different from that used to obtain the solution in metric MDS. With the latter, one estimates each dimension of the scaling solution successively; that is, new dimensions are added to the previously-estimated dimensions until adequate goodness of fit is achieved. With nonmetric MDS, a complete scaling solution is obtained in a given dimensionality; that is, all of the dimensions are estimated simultaneously rather one at a time, successively. If the goodness of fit for the resultant point configuration is adequate, there is no need to go any further; simply stop and report the results of the analysis. But, if the fit of the scaled interpoint distances to the input ordinal dissimilarities is poor, then discard the results and try a solution in the next higher dimensionality (again, estimating all of the dimensions in the new solution at the same time). Continue this process until an adequate fit is obtained.

Within a specific dimensionality, the scaling procedure relies on an iterative process. We start with some initial configuration of k points representing the objects to be scaled. The interpoint distances in this initial configuration are not expected to reflect the dissimilarities between the objects very accurately. So we move the points around within the space, as necessary, in order to make the interpoint distances monotonic with the order of the dissimilarities. This usually takes several (and often many) sets of point movements. We (or, more realistically, the scaling software) keep moving the points until one of two things occurs: (1) the interpoint distances are perfectly monotone with respect to the dissimilarities; or (2) the movements stop enhancing the consistency between the dissimilarities and the distances. In real MDS applications, the second condition is more common than the first. In other words,

the scaled configuration of points does not produce distances that are a *perfectly* monotonic function of the dissimilarities. But, hopefully, they are nearly so.

In order to illustrate the general idea underlying (if not the exact procedure used in) nonmetric MDS, we will once again scale the ten U. S. cities we used in the previous metric MDS example. Table 5 shows a new dissimilarities matrix for the ten cities. But now, the dissimilarities are created from the respective cities' economic, social, and cultural characteristics.[6] And, the matrix only rank-orders the pairs of cities in terms of their dissimilarities with respect to these characteristics. Thus, there is no question that this information comprises strictly ordinal data. Again, the objective is to find a configuration of points in m-dimensional space such that the ordering of the interpoint distances matches as closely as possible the ordering of the pairs of cities in Table 5.

We start by specifying m, the dimensionality in which we will seek a scaling solution. For now, we will set $m = 2$. If we can produce a set of points in two-dimensional space such that the relative sizes of the interpoint distances are consistent with the rank-ordered dissimilarities, then we will be able to show the results easily in a graphical display. We will only move to a higher dimensionality if the two-dimensional solution fails to produce an adequate fit to the data.

We next need to generate a configuration of points, so we have the "raw material" for the point-moving process that will produce the scaling solution. This initial configuration is just to get us started; we do not expect that the interpoint distances will reflect accurately the ordered dissimilarities. The starting configuration is obtained by randomly positioning ten points in two-dimensional space. Figure 3 shows such a starting configuration for the ten cities. Clearly, this configuration is *not* an accurate representation of the dissimilarities. For example, the least dissimilar pair of cities is Denver and Atlanta, scored "1" in Table 5. Yet, the points representing these cities are widely separated in the space, with Atlanta near the bottom of the point configuration and Denver near the top. At the opposite extreme, the third-most dissimilar pair is New York City and Atlanta, scored "43" in Table 5. But, the

points for these cities are adjacent to each other at the bottom of the point configuration.

The inadequacy of the scaling solution is illustrated in Figure 4, where interpoint distances from Figure 3 are plotted against the corresponding ordinal dissimilarity values; again, this type of graph is usually called a Shepard diagram. If the distances were consistent with the dissimilarities, then the plotted points would form a monotonic array from lower-left to upper-right. Obviously, that is not the case here. So, it is definitely necessary to move the points from their current randomly-determined locations to make the distances more consistent with the dissimilarities. In order to do this, we will generate a new set of 45 values—one for each interpoint distance— called *disparities*. The disparity for objects i and j is designated $\hat{d}_{ij}$. The disparities have two properties:

1. The disparities are as similar as possible to the distances, in the least-squares sense.

2. The disparities are weakly monotonic to the rank-ordered dissimilarities; that is, as the rank of the distances increases, the values of the disparities cannot decrease (though they need not increase).

Stated differently, the disparities comprise a set of values that is maximally correlated with the distances, but still monotonic to the original ordinal input data for the MDS. A simple procedure for calculating the disparities was developed by Joseph Kruskal (1964b).

The first panel of Figure 5 shows the Shepard diagram from Figure 4, with the disparities (plotted as "x's") superimposed over the points determined by the distances and dissimilarities (plotted as open circles). Notice that, unlike the distances, the disparity points *do* form a weak monotonic array: As we move from left to right along the horizontal axis, the array of x's always moves upward or to the right; it never reverses direction on itself. The second panel of Figure 5 connects each disparity to its corresponding distance with a gray line segment. This emphasizes how the former are "targets" for the latter. If the disparity point falls below the distance point, then the points in the MDS solution for that object pair need to be moved closer to each other. Conversely, if the disparity point falls above the

distance point, then the points in the MDS solution for that object pair need to be moved farther apart. The relative size of the point movement is indicated by the length of the line segment; shorter segments represent point-pairs in the MDS configuration that do not need to be moved very much, while longer segments indicate that a larger movement is necessary.

The separation between the distances and the disparities also enables us to construct a fit measure for the MDS solution. If the configuration of points in the MDS solution *were* monotonic to the dissimilarities, then the disparities would be equal to the distances themselves. The greater the departure from monotonicity in the MDS solution, the greater the discrepancy between the distances and disparities. Thus, the sum of squared differences between the two summarizes the degree to which the scaling solution departs from the stated objective of distances that are monotone with the dissimilarities. We also need to normalize our fit measure, to remove its dependence on the measurement units in the scaling solution (which are arbitrary). There are several ways to do this, but one simple method is to divide the sum of squares by the sum of the squared distances. This leads to a fit measure called Kruskal's Stress (an acronym for "standardized residual sum of squares"), defined as follows:

$$Stress_1 = \left[\frac{\sum^{\#pairs} (d_{ij} - \hat{d}_{ij})^2}{\sum^{\#pairs} d_{ij}^2} \right]^{0.5} \tag{10}$$

The subscript indicates that this is the first of two Stress formulas Kruskal developed. The other, $Stress_2$, is identical except that the denominator contains the corrected sum of squares for the distances. In some widely-quoted guidelines, Kruskal (1964a) suggests that a Stress_1 value of 0.1 is "fair," while 0.05 is "good," and anything 0.025 or below is "excellent." Kruskal's stress is a *badness-of-fit* measure, in that values closer to zero are better. The Stress value for the bad point configuration from Figure 3 is quite a bit larger than zero, at 0.148.

While we want Stress values closer to zero, it is a bit difficult to interpret the specific values for any given scaling solution. Therefore, many analysts rely on an alternative fit

measure, the squared correlation between the distances and the disparities. Here, the idea is that the disparities are monotonic to the dissimilarities. The higher the correlation between the distances and the disparities, the greater the extent to which the current point configuration meets the scaling objective (i.e., distances which are monotonic to the dissimilarities). The squared correlation of the distances and disparities for MDS solution in Figure 3 is very small, at 0.076. Of course, this is not surprising, since the points are randomly-located.

Getting back to the scaling procedure, we use the disparities to move the points. We do this by creating a new "dissimilarities" matrix, in which the δ_{ij}'s are replaced by the $\hat{d}_{ij}$'s. Then, we carry out a metric MDS on the new matrix, generating a new configuration of ten city-points in two-dimensional space. This latter MDS effectively moves the points from their initial random locations to a new configuration in which (hopefully) the interpoint distances are more nearly monotonic to the original rank-ordered dissimilarities. In order to save space, we do not show the new point configuration. But, the two panels of Figure 6 show the Shepard diagram for this new MDS solution— though it is important to emphasize that the values on the horizontal axis are the original rank-ordered profile dissimilarities, and *not* the disparities that were actually used as the input to the metric MDS that was just carried out. From the first panel, we can see that the point cloud is more clearly oriented from lower-left to upper-right in the plotting region, suggesting that the interpoint distances in the MDS solution are coming closer to monotonicity with the input dissimilarities. The second panel adds the disparities for the latest MDS point configuration. From the vertical gray line segments, we can see that the discrepancies between the MDS interpoint distances and the disparities tend to be quite a bit smaller than they were in the initial, random, configuration of city-points. This is confirmed by the two fit measures. The second point configuration produces a Stress value of 0.079, and a squared correlation of 0.624; clearly, we are moving toward a better-fitting model.

We now create a new matrix containing the latest set of disparities (i.e., from Figure 6), and repeat the process. Each time we carry out an MDS on the disparities constitutes an

iteration. We will not show the intervening steps, to save space. But, on each iteration, the fit improves, the point cloud in the Shepard diagram provides a closer approximation to monotonicity, and the discrepancies between the disparities and the distances get smaller. After eight such iterations, the movements of the points in the MDS solution become extremely tiny, and the fit of the point configuration to the data barely changes from the last iteration. Figure 7 shows the two versions of the Shepard diagram for the MDS solution after the eighth iteration. At this stage, the MDS configuration provides an excellent fit to the ordinal dissimilarities: Stress is very close to zero, at 0.014, and the squared correlation is very large, at 0.922.

Figure 8 shows the nonmetric MDS solution, itself. Without going into a great deal of detail, this scaling solution does seem to make sense in substantive terms. Consider the cities that fall into the various groupings or clusters of points. On the right side of the plot, the points for three sun-belt cities (Houston, Denver, and Atlanta) are relatively close to each other. Near the bottom, two large metropolitan areas that are well-known for their cosmopolitan environments (Chicago and Washington DC) form another cluster. The two California cities are separated from the other points in the upper-left part of the plotting region. At the top of the region, two coastal cities at opposite corners of the country (Miami and Seattle) fall close to each other. And, finally, New York City is off by itself in the lower-left part of the space; the isolation of this point is not at all unreasonable, given the unique aspects of that particular city.

The preceding example gives an intuitive sense of the general strategy underlying nonmetric multidimensional scaling. But, even though it works quite well, this approach is not the way most "real" nonmetric MDS software works. Instead, they capitalize on the fact that the Stress coefficient is a function of the point coordinates in the current MDS configuration (which are used to calculate the distances and, indirectly, the disparities). Accordingly, the partial derivatives of Stress, relative to the coordinates, can be calculated. The partial derivatives show how Stress changes when the point coordinates are changed by

an infinitesimal amount. Therefore, they help us achieve a scaling solution by showing how to change the point coordinates in the ways that make the partial derivatives the smallest possible negative values (thereby decreasing the discrepancy between distances and disparities). After the points are moved, the scaling software recalculates the disparities, evaluates the fit of the new configuration, and repeats the process as necessary. So, it is an iterative process, as in the preceding intuitive example. But, the MDS software focuses more directly on the process of changing the point coordinates and, hence, moving the points, rather than carrying out a complete metric MDS on each iteration.

The steps in a typical nonmetric MDS routine proceed as follows:

1. Generate a starting configuration of object points. The points could be located randomly, as in the previous example. But, most modern MDS software uses a "rational" starting configuration, obtained by performing a metric MDS on the ordinal data.
2. Calculate the disparities for the starting configuration

   Use the distances from the starting configuration and the disparities to calculate Stress

   If Stress is zero (unlikely with a random start, but possible with a rational start), then terminate the routine.

   Otherwise, proceed to the next step
3. Calculate the partial derivatives of Stress relative to the coordinates, and move the points in order to produce the largest possible decrease in the Stress value.
4. Use the new distances between the just-moved points to calculate disparities and Stress for the new object point configuration.

   If Stress is zero, then the distances are monotonic to the dissimilarities so terminate the routine.

   If Stress has not changed since the last iteration or it has gotten larger, indicating a worse fit than on the previous iteration (unlikely with modern software) then terminate the routine.

   If Stress is smaller than it was on the previous iteration, the MDS solution is improving, so go back to step 3 and repeat.
5. Terminate the MDS routine and print out the results:

   An iteration history, showing how Stress has changed as the object points are moved from the starting configuration to the final configuration.

The coordinates for the object points in the final MDS solution

The fit statistics for the final MDS solution (usually Stress and the squared correlation between the interpoint distances and the disparities)

Substantive Example of Nonmetric Multidimensional Scaling

In order to illustrate the utility of multidimensional scaling for dealing with substantive research problems, let us move on to a more realistic substantive example. The problem is drawn from the field of political psychology, and the specific research question is, what do citizens "see" when they think about a field of candidates for public office? Stated a bit differently, what is the nature of the cognitive structure that the electorate brings to bear on a set of electoral stimuli?

Assume that we have information about the American electorate's perceptions of thirteen prominent political figures from the period of the 2004 American presidential election. The figures are: George W. Bush; John Kerry; Ralph Nader; Richard Cheney; John Edwards; Laura Bush; Hillary Clinton; Bill Clinton; Colin Powell; John Ashcroft; John McCain; the Democratic Party; and the Republican Party. The information comes from a high-quality public opinion survey (the 2004 American National Election Study) and it is the perceived dissimilarities between all pairs of these political figures. With 13 figures, there will be 78 distinct pairs of figures. And, the perceptions of the individual survey respondents are aggregated to form one summary dissimilarity score for each pair of political figures. With such perceptual information, we probably would be hesitant to attribute metric properties to the specific dissimilarity scores; instead, we will only assume that the latter represent ordinal-level measurement. Therefore, we rank-order the pairs of political figures according to their perceived dissimilarity, assigning integer values from one (for the least dissimilar or most similar pair) to 78 (for the most dissimilar or least similar pair).[7]

As with the previous smaller examples, we arrange the rank-ordered dissimilarity values into the square, symmetric matrix shown in Table 6. The zeroes in the main diagonal indicate that each figure is not at all dissimilar to him/her/itself. In the off-diagonal cells, the value

"1" appears in the fifth row, second column (and the second row, fifth column, since the matrix is symmetric), indicating that Kerry and Edwards are the most similar pair of figures. At the opposite extreme, the "78" in the cell at the fourth row and second column shows that Kerry and Cheney are the most dissimilar pair. Noninteger values, such as the "74.5" in the cell at the fifth row, fourth column, and also in the cell at the thirteenth row and second column, indicate tied degrees of dissimilarity.

The results from a nonmetric MDS analysis of the data in Table 6 are shown in Figure 9. The point configuration in this figure is a nearly perfect representation of the information that was contained in Table 6. The Stress value is tiny, at 0.04 and the squared correlation between the disparities and the scaled interpoint distances is very high, at 0.993. Thus, the rank-order of the distances between pairs of points corresponds extremely closely, to the rank-ordered dissimilarities between pairs of political figures. Of course, Figure 9 is not the *only* point configuration that would provide such an excellent depiction of the data. But, any other sets of points that provide equally good representation of the dissimilarities would be almost identical to this one. For example, the orientation of the points could differ (e.g., what is now the horizontal axis could be vertical, and vice versa), the points could be reflected on one or both of the axes, and the overall size of the point cloud could be magnified or shrunk. But, these kinds of differences do not affect the *relative* distances between the points. If we were to hold the overall orientation and size of the point cloud fixed, then it is impossible to move any of the points very much at all, without degrading the exact correspondence between pairwise dissimilarities and pairwise distances.

The contents of this point configuration should make a great deal of sense to anyone familiar with American electoral politics. Looking first at the lower half of the plotting region, the points representing Democratic figures are grouped hear the left side of the space (i.e., Kerry, Edwards, the Clintons, and the Democratic Party). In contrast, the points closer to the right side represent Republicans (Ashcroft, Cheney, George Bush, and so on). The point representing Nader, an independent candidate in 2004, lies separated by a great

distance from both of the two partisan sets of points. So, these groupings correspond to partisan differences between the political figures. But, there is more: Looking at the array of points from left to right (i.e., ignoring their differences in the vertical direction), the ordering corresponds closely to common understandings of ideology, with the clearest left-leaning figures (Kerry and Edwards) positioned fittingly enough at the left side of the space, the most extreme right-wing figures (Ashcroft and Cheney) over at the opposite extreme, and relatively moderate figures (e.g., Colin Powell) located near the center. For now, it is sufficient to emphasize that interesting substantive features are immediately obvious in the scaled points, but they would have been much more difficult to discern in the original 13 by 13 matrix of dissimilarity scores. This, in itself, demonstrates the power of a successful multidimensional scaling analysis.

INTERPRETATION STRATEGIES FOR MULTIDIMENSIONAL SCALING

Of course, MDS (like any other data analysis strategy) is usually used to provide substantive insights about the objects contained in the data being analyzed. One important interpretational aspect of MDS involves the dimensionality of the scaling solution. Specifically, how many dimensions should be used in any given analysis? It is impossible to provide a general answer to this question. There is always a trade-off between two considerations: First, the scaling solution needs enough dimensions to provide an accurate representation of the dissimilarities (i.e., a low Stress value and a high correlation between the distances and the input dissimilarities or disparities). Second, the scaling solution needs a relatively small number of dimensions to facilitate interpretation. In practical terms, the latter criterion usually wins out— MDS solutions with more than two, or at most three, dimensions are quite rare.

Once the MDS solution is obtained, it is important to emphasize that the analysis only determines the relative distances between the points in the scaling solution. The locations of the coordinate axes, and their orientation relative to the scaled points, are completely arbitrary. Regardless of the software used to perform the analysis, the final MDS solution is

usually rotated to a varimax orientation. That is, the location of the point "cloud" relative to the coordinate axes is set in a way that maximizes the variance of the point coordinates along each of the rectangular axes. The numeric values of the point coordinates, themselves, are usually standardized in some way. For example, they may be set so the mean coordinate on each axis is zero, and the variance of the point coordinates across all of the axes is equal to one (or to some other arbitrary value, such as the sum of squares for the entries in the dissimilarities matrix). Regardless of the rotation or numeric values, the coordinate axes in the MDS solution are simply a device to "hang the points within the m-dimensional space. They have no intrinsic substantive importance or interpretation!

The simplicity of the underlying geometric model and the graphical nature of the main output comprise big advantages for interpreting MDS solutions. Typically, the analyst would look for two kinds of features: (1) interesting "directions" within the space, which may correspond to the substantive dimensions underlying the processes that produced the dissimilarities in the first place; and (2) distinct groups or clusters of points, which may reveal how the data source differentiates the objects being scaled. Simply "eyeballing" the MDS results is often sufficient for identifying these kinds of features. In fact, that is exactly what we did with the configuration of 2004 political figures in the preceding example. Similarly, the first metric MDS example (reproducing the map from the intercity distances) implicitly emphasized directions (i.e. north-south and east-west) while interpreting the solution. And, the second substantive example (the metric and nonmetric analyses of the profile dissimilarities between the cities) emphasized groupings of the city-points. So, visual inspection is often enough in itself. But, the subjectivity inherent in simple visual interpretation of MDS results makes it desirable to use more systematic (and hopefully, "objective") interpretation strategies.

Embedding External Information

One interpretation strategy uses regression analysis to "embed" additional information into the MDS point configuration. The researcher often has prior hypotheses about the

substantive dimensions that differentiate the stimulus objects in an MDS analysis. If so, then it is useful to obtain external measures of these dimensions; that is, variables that are obtained separately from the dissimilarities used to create the MDS solution. Each object in the MDS analysis is assumed to have a score on this external dimension. Then, it is easy to incorporate the external information into the scaling solution by regressing the external measure on the point coordinates. If the point coordinates are highly correlated with the external variable, then the estimated regression coefficients can be used to draw a new axis within the space, corresponding to the external measure. This strategy is useful for determining whether the MDS results conform to prior substantive expectations.

We can illustrate the use of an external measure with the MDS of 2004 political figures. Table 7 shows the point coordinates for the MDS configuration from Figure 11, along with a third variable, a measure each figure's liberal-conservative ideology. The latter was obtained by having a panel of experts locate each of the 13 figures along a numeric continuum ranging from -100 for "maximally liberal" or left-leaning, to +100 for "maximally conservative" or right-leaning. The 13 political figures' scores on this ideology variable are regressed on their MDS point coordinates, and the OLS estimates are as follows:

$$Ideol_i = -3.574 + 16.963 Dim_{1i} + -2.130 Dim_{2i} + e_i \tag{11}$$

The R^2 for equation 11 is 0.918; this excellent fit confirms that the spatial configuration of points is highly consistent with the ideological positions of the political figures. Hence, that may well be one of the substantive criteria that the survey respondents used when they evaluated the dissimilarities among the figures to produce the input data for the MDS.

In order to locate an ideology axis within the MDS configuration itself, we take the ratio of the regression coefficients for the second MDS axis relative to the first MDS axis. Thus, $-2.130/16.963 = -0.126$. Now just insert a line with slope -0.126 into the MDS configuration. The perpendicular projections from the scaled points onto this line correspond

to the ideology scores for the respective figures. The exact location of the line does not really matter (since the projections are invariant, as long as the line's slope remains at -0.126), but it is convenient to have this line pass through the origin.[8]

Figure 10 shows the nonmetric MDS solution with the ideology axis drawn in. The nearly horizontal orientation of this objectively-determined line confirms the more informal interpretation based on simple visual inspection. That is, the left-right positioning of the points corresponds to ideological distinctions among the political figures, with more liberal candidates toward the left side of the space, and more conservative figures toward the right. But, the MDS solution really goes beyond simple intuition because the metric distances between the points along the fitted dimension provide an interval-level estimate of the respective figures' liberal-conservative ideology. And, since the MDS solution was, itself, produced without any reference to ideology at all, we can take these results as empirical evidence supporting the hypothesis that the American public has ideology in mind when it looked at the "political field" in the 2004 presidential election.

Cluster Analysis

Cluster analysis provides an objective strategy for identifying groups of stimulus points within the MDS solution. The analyst can then determine whether the clusters correspond to substantively-interesting groups of stimulus objects. There are *many* varieties of cluster analysis. They all work by joining closer objects together into groups or "clusters." Here, the "objects" are the points in the MDS solution, and "closeness" is based upon the interpoint distances in the scaled space.

Hierarchical clustering methods are probably the most common variant of this general methodology. The analysis begins by regarding each object point as a separate "cluster." In the first step, the routine joins together the two closest clusters to form a new cluster; from that step on, the constituent objects are no longer regarded as separate. The location of this new cluster is some summary of the locations of the original two stimulus points (e.g., perhaps the mean of their coordinates along each axis of the space). The routine proceeds

through k such steps, on each one joining together the two most similar (i.e., closest) clusters to form a new cluster. It continues until all k objects are together in a single cluster. A tree diagram called a "dendrogram" traces out the process of joining clusters, and is usually considered the main output from a cluster analysis.

Figure 11 shows the dendrogram from a cluster analysis of the MDS configuration for the 2004 political figures. The diagram starts at the bottom of the display, where the figures form 13 distinct clusters. Then, it is easy to see the two partisan groupings develop as we move upward through the diagram. And, it is interesting to observe the intra-party distinctions— for example, Laura Bush, Colin Powell, and John McCain are relatively distinct from the other Republican figures, forming a separate cluster of their own within the overall Republican group. Similarly, Bill and Hillary Clinton are separate from the 2004 Democratic candidates (Kerry and Edwards) and the Democratic party, itself. Such subgroups actually make quite a bit of sense from a substantive perspective. Cluster analysis helps accentuate these kinds of distinctions and, in so doing, helps the analyst understand the variability within the scaled point locations.

A Caveat Regarding MDS Interpretation

It is important to be aware that most MDS solutions are amenable to several different substantive interpretations. So, for example, does the horizontal direction in Figures 11 and 12 correspond to partisanship or ideology? In fact, the positions of the points are fairly consistent with either of these two interpretations. Objective strategies like those discussed in the previous two sections can be used to show that the scaling results are consistent with some particular interpretation of the space. But, objective methods can never be used to find the single "true" meaning of the MDS-produced point configuration. While this uncertainty bothers some researchers, it is really no different from the general scientific approach of theory construction and revision through empirical hypothesis-testing.

A VARIETY OF MULTIDIMENSIONAL SCALING MODELS

Up to this point, the discussion has focused on classical multidimensional scaling, or

CMDS, in which one square matrix of proximities data is used to produce a single geometric map of points representing the objects. This is both the simplest MDS model and the most common type of MDS used in actual research settings. But, there are a variety of additional MDS variants and models that adapt the general methodology to specific data analysis contexts.

First, there are several models that generalize MDS to situations where there is more than one square matrix of proximities among the k objects. For example, the researcher may have perceived dissimilarities among a set of stimuli for several (say, n) different experimental subjects. Or, there may be several matrices of profile dissimilarities for a common set of objects, collected at n different time points. In any case, there are n distinct square, symmetric, matrices of dissimilarities among the same set of k objects. So, the question is how can the information from the n different data sources be combined in an MDS analysis?

The answer to the preceding question depends upon how the researcher wants to treat the multiple data sources. If the differences between the data sources are uninteresting, so that the n matrices are regarded as simple repeated measurements of the same dissimilarity structure, then replicated MDS (or RMDS) may be appropriate. In the RMDS model, the differences across the n dissimilarity matrices are "smoothed out" to still produce a single configuration of k points in m-dimensional space. In technical terms, the function relating each of the n sets of input dissimilarities to the single set of output distances between the k scaled points is allowed to vary across the n data sources. If the analysis is a metric RMDS, then each data source's dissimilarities are related to the distances by a linear function; but, the slope and intercept of the linear relationship between dissimilarities and distances differs across the n data sources. Similarly, a nonmetric RMDS would have a different monotonic function connecting each of the n data sources' dissimilarities to the final set of interpoint distances. Note that the goodness of fit for the scaling solution can vary across the data sources depending upon how well the common MDS point configuration represents the dissimilarities data associated with each source. Typically, the individual fit measures for

the respective data sources are averaged to produce a global fit measure for the entire scaling solution. Note that RMDS can be useful for handling a problem known as "differential item functioning" where a common measurement instrument is used in different ways by different individuals (or other data sources). But it usually still is the case that the differences across the matrices are treated as errors that are not interesting from a substantive perspective.

If the differences across the n matrices *are* substantively important, then the weighted multidimensional scaling (or WMDS) model may be appropriate. This model still generates a single m-dimensional point configuration for the k objects. But, each data source is associated with its own set of m weights (one for each dimension). A larger weight value for a given dimension "stretches out" the corresponding coordinate axis relative to the other axes in the scaling solution. And, a smaller weight "shrinks in" an axis, relative to the others. Thus, each data sources is associated with a "distorted" version of the common point configuration, obtained by stretching and shrinking the m coordinate axes according to its own set of weights. Note that WMDS differs from CMDS in that the axes of the space should be interpretable in substantive terms— they cannot be rotated without degrading the model fit.

To give some sense about substantive examples of RMDS and WMDS, we will return to the context of the American electorate's perceived dissimilarities between political candidates and other figures. But now, let us assume that we have stratified a random sample of the electorate into subgroups based upon their levels of political knowledge (presumably measured through answers to factual questions on the survey used to obtain the information about the dissimilarities). WMDS analyses of such data from the 1980's and 1990's typically revealed systematic differences in the salience or importance of the ideological dimension; that is, higher levels of political knowledge corresponded to larger weights on the dimension that separates liberal from conservative figures, and vice versa (e.g., Jacoby 1986). Interestingly, however, similar analyses of data from 2000 and 2004 produce miniscule differences in the weights; more and less knowledgeable citizens seem to show relatively equal reliance on

the liberal-conservative continuum in their judgments about candidates and political figures (2004). In the latter case, a single point configuration generated by RMDS is sufficient to represent the perceived dissimilarities of all the knowledge strata in the electorate.

The WMDS model actually has some variants of its own. For example, General Euclidean Model MDS (sometimes called GEMSCAL) allows for differential rotations, as well as stretching and shrinking of coordinates, across the data sources. Using this approach allows for even greater idiosyncracies across the data sources than the simpler WMDS, which only allows for differential importance of a common set of dimensions. For example, assume we have dissimilarities in the voting records for members of a legislative body (e.g., the percentage of times each pair of members votes differently from each other on bills), across several legislative sessions. Partisan voting coalitions are usually stable over long periods of time, but punctuated by brief intense periods of realignment when old voting blocs break down and new coalitions form. If this is the case, applying the GEMSCAL model to the legislative dissimilarities data across sessions could show how the underlying dimensions of legislative voting exhibit stability and change over time (e.g., Easterling 1987).

Finally, Asymmetric MDS (sometimes called ASYMSCAL) uses a single matrix of proximities data but it allows for asymmetric data values; that is, it anticipates the situation where $\delta_{ij} \neq \delta_{ji}$. For example, such a model might be useful for analyzing sociometric data on interpersonal interactions (Collins 1987). Basically, the ASYMSCAL approach generates a single point configuration for the k objects, but also estimates two k by m weight matrices—one for the rows of the matrix and one for the columns. Thus, there is quite a bit of flexibility in the kind of geometric representation that can be constructed using MDS. But, it should be noted that actual empirical applications of the GEMSCAL and ASYMSCAL models are quite rare.

A second generalization of MDS applies to non-square data matrices. The unfolding model is useful for preferential choice data, in which n "subjects" each give their preference orders for k objects. An unfolding analysis seeks to recover an m-dimensional space con-

taining n points for the respective subjects and k points for the objects. The relative point locations are determined by the preferential choices expressed by the subjects for the objects. Speaking very informally, if a subject indicates greater preference for an object, then the distance between the subject's point and the object point is smaller, and vice versa. The dimensions of the space recovered by the unfolding analysis should provide insights about the judgmental criteria that the subjects used to make their choices among the objects. And, as in CMDS, the relative positions of the points within the two sets contained in the space depict the variability across the subjects and objects, respectively. Unfolding models are particularly useful for market research, in which the objective is to explicate consumers' preferential choices across a set of products. A variant of the unfolding model is sometimes called MDPREF, for "Multidimensional Analysis of Preferences." Here, the k objects are still shown as points within a space, located so that objects that receive similar patterns of preferences from the subjects are represented by points that are close together within the space, and vice versa. But, unlike the unfolding model (which represents subjects as points), the MDPREF model shows each subject as a unit-length vector emanating from the origin of the space. Each subject's vector is positioned so that ordering of the perpendicular projections from the object points onto that vector correspond to the subject's preference ordering for the objects. In this manner, variability across the subjects is represented by the angular separation between the vectors. The MDPREF model has been used recently to study individual choices across a set of core values (Jacoby 2014).

Finally, correspondence analysis (abbreviated CA) is usually regarded as a different methodology than MDS. But, CA can be viewed as a strategy for scaling the rows and columns of a crosstabulation table. Basically, each row and column from a table is shown as a point in a low-dimensioned space. Rows that have similar distributions of observations across their cells (i.e., across the columns of the table) have points that are located closer together in the space, and similarly for the columns. The relationships between the rows and columns are depicted in a slightly more complicated manner: For a given row,

the point for that row can be regarded as a vector emanating from the origin. And, the perpendicular projections from the column points onto that row's vector will correspond to the distribution of observations across the cells in that row (i.e., more heavily populated cells for that row project onto the vector at a point farther away from the origin). In this manner, CA can be used to produce a geometric representation of the numeric information from a crosstabulation, which may prove very useful for interpreting large or complex tables.

SOFTWARE FOR MULTIDIMENSIONAL SCALING

Until fairly recently, multidimensional scaling required stand-alone, special-purpose software. But, a tangible indicator of MDS's increasing popularity is the fact that all of the major statistical software packages now contain routines to estimate these models. For example, SPSS has two MDS routines, `ALSCAL` and `PROXSCAL`. Both of these procedures are very flexible and capable of performing many different varieties of MDS (e.g., metric and nonmetric CMDS, RMDS, WMDS). They can also calculate profile dissimilarities from multivariate data prior to performing the scaling analysis itself. The SAS system has long included `PROC MDS`, a very flexible procedure that can perform many varieties of MDS. `PROC MDS` generally is modeled after `ALSCAL` although its estimation algorithm is a bit different. Although less well-known than the preceding two packages, SYSTAT is a venerable and very powerful statistical software package. SYSTAT contains a very flexible and easy-to-use MDS routine that can perform most varieties of MDS. SYSTAT also contains an excellent graphics system which integrates well with the output from its MDS routine.

For many years, STATA did not include multidimensional scaling among its formidable array of statistical procedures. That changed with Version 9.0, when STATA introduced the `mds` and `mdsmat` procedures. The two procedures carry out identical analyses but make different assumptions about the nature of the input data. `mds` assumes multivariate input data that must be converted to profile dissimilarities prior to the scaling analysis; `mds` carries out this conversion. `mdsmat` assumes that the data are already contained in a dissimilarities matrix. The `mds` and `mdsmat` procedures can perform metric and nonmetric CMDS. Users

will find that both of these procedures are well-integrated with STATA's overall system of model estimation and post-estimation commands, and its extensive graphics system.

The R statistical computing environment contains several options for MDS. The Base R installation only includes the function, `cmdscale`, which performs metric MDS. The MASS package (distributed with all R installations) includes functions `isoMDS`, `sammon`, and `Shepard`, all of which perform nonmetric CMDS. But, most researchers using R probably will prefer to use the `smacof` package which provides functions for estimating metric and nonmetric CMDS, WMDS, RMDS, and GEMSCAL models, using a unified analytic approach. The `smacof` package includes generic plotting methods for the MDS models it estimates, but users will often find it more useful to take the `smacof` model estimates and use them as input to R's unparalleled graphics functions and packages to produce geometric representations of MDS models.

Finally, `ggvis` is an MDS module within the `ggobi` software package. `ggobi` is a program for visualizing high-dimensional data. It is freely available on the web, at `http://ggobi.org`. The `ggobi` system (including `ggvis`) can be integrated with the R statistical computing environment, via the `rggobi` package. `ggvis`, itself, can estimate metric and nonmetric CMDS models. It is an interactive MDS program in that the user can manipulate directly the scaled configuration of points. This can be very useful for evaluating the robustness of an MDS solution.

CONCLUSION

In conclusion, multidimensional scaling is a relatively new methodological approach, having been in existence for less than 70 years. As such, it is fairly straightforward to specify a set of useful references for this methodological strategy. Recent general texts on MDS include Cox and Cox (2001), Borg and Groenen (2005), and Armstrong, Bakker, Carroll, Hare, Poole, and Rosenthal (2014). Seminal contributions to the field include Torgerson (1958) for metric MDS, Shepard (1962a; 1962b) and Kruskal (1964a; 1964b) for nonmetric MDS, Carroll and Chang (1970) for weighted MDS. Young and Hamer (1987) give a theoretical

overview of MDS that integrates a wide variety of specific models. Work aimed at developing the statistical properties of MDS includes Ramsay (1977), Takane and Carroll (1981), Brady (1985), Oh and Raftery (2001), Bakker and Poole (2013), and Jacoby and Armstrong (2014). Coombs (1964) provides the original development of unfolding, while Carroll (1972) provides further discussion of its relationship to MDS. Greenacre (1984) is a standard reference for correspondence analysis.

MDS is not necessarily viewed as a "mainstream" analytic strategy for most social and behavioral scientists. It is certainly not encountered as frequently as several other scaling or data reduction techniques, such as principal components or factor analysis. Nevertheless, there are signs that the use of MDS is becoming more widespread. For one thing, MDS applications are appearing across a widening variety of academic disciplines. The methodology was originally developed by psychologists who were interested in constructing empirical representations of human cognitive structures. That application has receded somewhat, but MDS has proven invaluable in a variety of other contexts. For example, anthropologists have used it to compare characteristics across cultures (Herrmann and Raybeck 1981), and archaeologists have employed MDS for seriation of artifacts (Drennan 1976) and also to develop classification schemes for objects recovered from site excavations (Hodson, Sneath, Doran 1966). Market researchers have long employed MDS to assess consumers' perceptions of products (Carroll and Green 1997). And, political scientists have found the methodology useful to operationalize the formal spatial model of voting (Enelow and Hinich 1984) and to represent legislative roll-call votes (Hoadley 1980; Poole 2005). More recently, MDS has been used in conjunction with network analysis as a tool for laying out graph models of interacting units (Di Battista, Eades, Tamassia, Tollis 1998). And, statisticians have found MDS to be a useful tool for overcoming the "curse of dimensionality" that is encountered when one attempts to visualize hypervariate data (Buja and Swayne 2002). Moving on into the future, researchers should routinely consider MDS whenever they find it useful to produce geometric representations of dissimilarities data.

NOTES

1. A square matrix can be subdivided into two triangular arrays. One such triangle consists of the cells below and to the left of the main diagonal (which, itself, consists of the cells running from the upper-left corner to the lower-right corner) and the other triangle consists of the cells above and to the right of the main diagonal. If the matrix is symmetric, then the two triangles contain the same information; one is a transposed version of the other. Strictly speaking, MDS only requires one of these triangles (since the other is redundant). But, the metric MDS procedure discussed below requires the full, square, matrix form with both triangles, as well as the main diagonal.

2. This distinction about the value of the a coefficient stems from the definitions of the respective levels of measurement. With interval-level measurement, the numeric values reflect the amount of difference between the objects being measured (this is shown by the value of the b coefficient). But, the zero point is arbitrary. And, the latter feature is represented by a nonzero value for the a coefficient. At the ratio level of measurement, the numeric values still reflect differences between the objects but, now, the origin of the scale is fixed (usually at the position that represents complete absence of the property being measured); hence, the a coefficient must be constrained to zero.

3. This use of eigenvalues to measure the "variance explained" in the MDS solution is an exact parallel to the use of summed eigenvalues to represent the amount of variance explained in a multivariate dataset by a subset of components in principal components analysis. And, just as we do here, the summed eigenvalues for the subset of components can be divided by the sum of all the eigenvalues (which is equal to the sum of the variances of the observed variables) to provide the proportion of variance explained by those components. Similarly, in some variants of common factor analysis, the sum of the eigenvalues for the retained set of factors gives the amount of variance in the observed variables that is "explained by" the retained factors. But, the interpretation as a proportion of variance explained can be tricky. If the summed eigenvalues for the retained factors are divided by the summed eigenvalues for the full set of variables, then the resulting numeric value gives the proportion of *common* variance in the observed variables that is explained by the factors. In order to obtain the proportion of *total* variance in the observed variables that is explained by the factors, the summed eigenvalues for the retained factors must be divided by the summed variances for the full set of observed variables.

4. Assume that D is some function that takes two arguments (say, a and b). D is a distance function if the following four properties hold for all possible subsets of objects, a, b, and c, in a given set of objects:

    1. $D(a, b) \geq 0$ (Non-negativity)
    2. $D(a, a) = 0$ (Identity)
    3. $D(a, b) = D(b, a)$ (Symmetry)

4. $D(a, b) + D(a, c) \geq D(b, c)$ (Triangle inequality)

Any type of data that possesses the preceding four characteristics (or can be transformed to possess them) can be treated as a distance, and used as input to an MDS analysis.

5. The monotonic relationship between scaled distances and input dissimilarities is analogous to the monotonic item response functions that are used in Mokken scaling (Sijtsma and Molenaar 2002; van Schuur 2011), a form of nonmetric item response theory (IRT). In metric IRT, the analyst specifies a particular parametric function (usually a sigmoid or "S-shaped" function such as the logistic or the cumulative normal) that relates a subject's position along an unobserved dimension to the probability that person will give a positive response on a given test item. This enables estimation of interval-level values for the items and for the subjects. In nonmetric IRT, the researcher only specifies that the item response functions for the items are monotonic with the respect to the underlying dimension (i.e., the exact shape is not known or assumed) and that the functions for separate items never overlap each other. The cost associated with the weaker assumptions about the item response functions used in nonmetric IRT is that Mokken scaling only produces an ordinal-level depiction of the items and the subjects (although, in practice, good-fitting Mokken scales can often be treated as interval-level data with few harmful consequences). Nonmetric MDS does not suffer from a similar limitation: If the analysis is successful, then the scaling solution produces interval-level estimates of the distances between the scaled points, despite the fact that the input data are only assumed to be ordinal-level (and, therefore, monotonically related to the distances). The reason for the difference is that a properly-conducted nonmetric MDS results in a very sizable number of metric constraints on the relative locations of the points. Therefore, it is impossible to move the points in a nonmetric MDS solution very much without violating the monotonicity assumption and degrading the fit of the scaling solution.

6. The values in the matrix are created by calculating the profile dissimilarities between the cities across the nine measures of social, economic, and cultural characteristics presented in the *Places Rated Almanac* (Savageau and D'Agostino 2000). The table just gives the rank-order of the actual profile dissimilarity values.

7. The dissimilarities are obtained by applying the line-of-sight procedure (Rabinowitz 1976) to survey data on respondents' feeling thermometer ratings of the thirteen political figures from the Center for Political Studies' 2004 American National Election Study. A more complete analysis of these data is provided in Jacoby (2009).

8. This strategy for embedding external information generalizes very easily for MDS solutions in higher dimensionalities. For example, in a three-dimensional solution the external variable is regressed on the point coordinates for the three coordinate axes in order to define a two-dimensional plane within the three-dimensional space. In still higher dimensionalities, the regression would produce a hyper-surface of one fewer dimensions than the dimensionality of the MDS space. From a practical perspective, higher-dimensioned MDS solutions are rarely a problem for fitted external criteria. In most cases, the user looks "into" the dimensional space by looking at two-dimensional subsets of the space (e.g., in a three-dimensional solu-

tion, we would inspect plots of dimension 1 versus dimension 2, dimension 1 versus dimension 3, and dimension 2 versus dimension 3). And, in each such subspace, we could simply locate a line for the external variable by taking the ratio of the regression coefficients associated with those two coordinate axes.

REFERENCES

Armstrong II, David A.; Ryan Bakker; Royce Carroll; Christopher Hare; Keith T. Poole; Howard Rosenthal. (2014) *Analyzing Spatial Models of Choice and Judgment with R.* Boca Raton, FL: CRC Press.

Bakker, Ryan and Keith T. Poole. (2013) "Bayesian Metric Multidimensional Scaling." *Political Analysis* 21: 125-140.

Borg, Ingwer and Patrick Groenen (2005) *Modern Multidimensional Scaling: Theory and Applications (Second Edition).* Springer-Verlag.

Buja, A. and D.F. Swayne. (2002) "Visualization Methodology for Multidimensional Scaling." *Journal of Classification* 19: 7-43.

Carroll, J.D. (1972) "Individual Differences and Multidimensional Scaling." In Shepard, Roger N., A. Kimball Romney, Sara Beth Nerlove (Editors), *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences (Volume I, Theory).* Seminar Press: pp. 105-155.

Carroll, J.D. and J.J. Chang (1970) "An Analysis of Individual Differences in Multidimensional Scaling via an N-Way Generalization of 'Eckart-Young' Decomposition." *Psychometrika* 35: 283-319.

Carroll, J. Douglas and Paul E. Green. (1997) "Psychometric Methods in Marketing Research: Part II, Multidimensional Scaling." *Journal of Marketing Research* 34: 193-204.

Collins, Linda M. (1987) "Sociometry: Deriving Sociograms via Asymmetric Multidimensional Scaling." in Forrest W. Young and Robert M. Hamer (1987) *Multidimensional Scaling: History, Theory, and Applications.* Lawrence Erlbaum.

Coombs, Clyde H. (1964) *A Theory of Data.* John Wiley (Reprinted 1976, Mathesis Press).

Cox, Trevor F. and Michael A. A. Cox (2001) *Multidimensional Scaling (Second Edition).* Chapman and Hall.

Di Battista, Giuseppe; Peter Eades; Roberto Tamassia; Ioannis G. Tollis. (1998) *Graph Drawing: Algorithms for the Visualization of Graphs.* Indianapolis, IN: Prentice-Hall.

Drennan, Robert D. (1976) "A Refinement of Chronological Seriation Using Nonmetric Multidimensional Scaling." *American Antiquity* 41: 290-302.

Easterling, Douglas V. (1987) "Political Science: Using the Generalized Euclidean Model to Study Ideological Shifts in the U.S. Senate." in Forrest W. Young and Robert M. Hamer (1987) *Multidimensional Scaling: History, Theory, and Applications*. Lawrence Erlbaum.

Enelow, James M. and Melvin Hinich. (1984) *The Spatial Theory of Voting*. Cambridge University Press.

Greenacre, Michael J. (1984) *Theory and Applications of Correspondence Analysis*. London, UK: Academic Press.

Herrmann, Douglas J. and Douglas Raybeck. (1981) "Similarities and Differences in Meaning in Six Cultures." *Journal of Cross-Cultural Psychology* 12: 194-206.

Hoadley, John F. (1980) "The Emergence of Political Parties in Congress, 1789-1803." *American Political Science Review* 74: 757-779.

Hodson, F. R.; P. H. A. Sneath; J. E. Doran. (1966) "Some Experiments in the Numerical Analysis of Archaeological Data." *Biometrika* 53: 311-324.

Jacoby, William G. (1986) "Levels of Conceptualization and Reliance on the Liberal- Conservative Continuum." *Journal of Politics* 48: 423-432.

Jacoby, William G. (2004) "Ideology in the 2000 Election: A Study in Ambivalence." In Herbert F. Weisberg and Clyde Wilcox (Editors) *Models of Voting in Presidential Elections: The 2000 Election*. Stanford, CA: Stanford University Press.

Jacoby, William G. (2009) "Ideology and Vote Choice in the 2004 Election." *Electoral Studies* 28: 584-594.

Jacoby, William G. (2014) "Is There a Culture War? Conflicting Value Structures in American Public Opinion." *American Political Science Review* (Forthcoming). doi: `http://dx.doi.org.proxy2.cl.msu.edu/10.1017/S0003055414000380`.

Jacoby, William G. and David A. Armstrong II. (2014) "Bootstrap Confidence Regions for Multidimensional Scaling Solutions." *American Journal of Political Science* 58:

Kruskal, Joseph B. (1964a) "Multidimensional Scaling by Optimizing Goodness of Fit to a Non-Metric Hypothesis." *Psychometrika* 29: 1-27.

Kruskal, Joseph B. (1964b) "Non-Metric Multidimensional Scaling: A Numerical Method." *Psychometrika* 29: 115-129.

Oh, Man-Suk and Adrian E. Raftery. (2001) "Bayesian Multidimensional Scaling and Choice of Dimension." *Journal of the American Statistical Association* 96: 1031-1044.

Poole, Keith T. (2005) *Spatial Models of Parliamentary Voting.* Cambridge, UK: Cambridge University Press.

Rabinowitz, George B. (1976) "A Procedure for Ordering Object Pairs Consistent with the Multidimensional Unfolding Model." *Psychometrika* 41: 349-373.

Ramsay, J. O. (1977) Maximum Likelihood Estimation in Multidimensional Scaling." *Psychometrika* 42: 241-266.

Savageau, David and Ralph D'Agostino. (2000) *Places Rated Almanac.* Foster City, CA: IDG Books.

Shepard, Roger N. (1962a) "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function, I." *Psychometrika* 27: 125-140.

Shepard, Roger N. (1962b) "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function, II." *Psychometrika* 27: 219-246.

Sijtsma, Klaas and Ivo W. Molenaar (2002) *Introduction to Nonparametric Item Response Theory.* Thousand Oaks, CA: Sage.

Takane, Yoshio and J. D. Carroll. (1981) "Nonmetric Maximum Likelihood Multidimensional Scaling from Directional Rankings of Similarities." *Psychometrika* 46: 389-405.

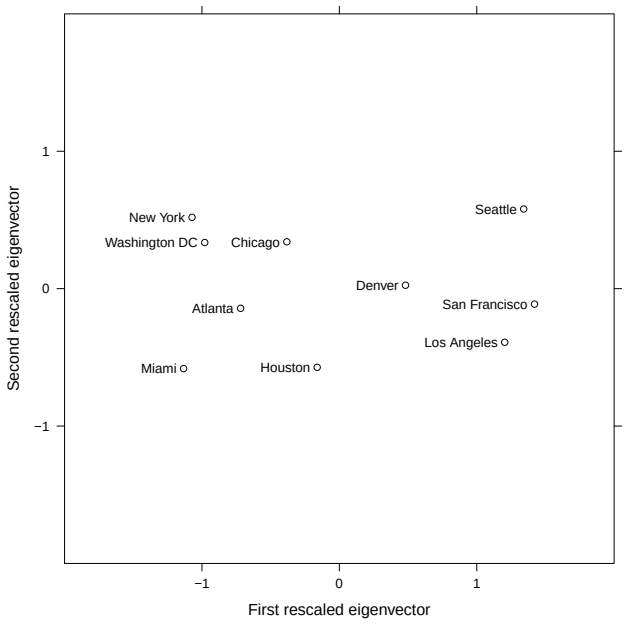
Torgerson, Warren S. (1958) *Theory and Methods of Scaling.* John Wiley.

van Schuur, Wijbrandt H. (2011) *Ordinal Item Response Theory: Mokken Scale Analysis.* Thousand Oaks, CA: Sage.

Young, Forrest W. and Robert M. Hamer (1987) *Multidimensional Scaling: History, Theory, and Applications.* Lawrence Erlbaum.

Figure 1: Point configuration obtained from metric multidimensional scaling of intercity driving distances.

A. Point configuration obtained directly from rescaled eigenvalues.



B. Point configuration with horizontal axis reflected (i.e., coordinates are multiplied by -1).

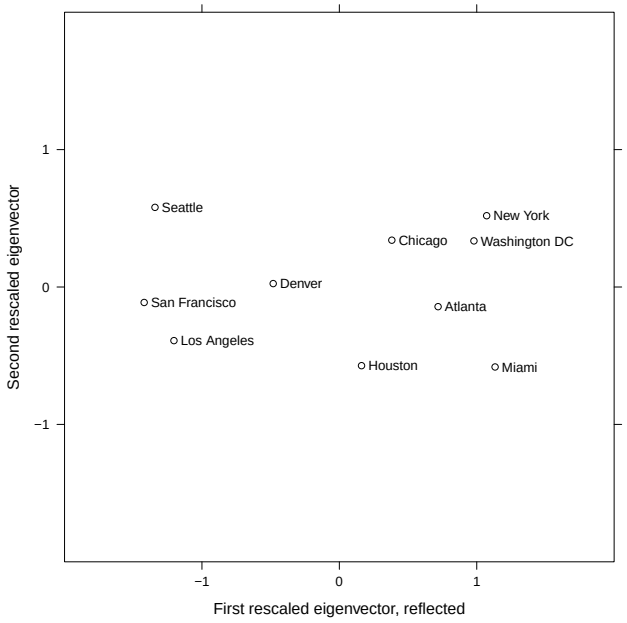


Figure 2: Shepard diagram showing interpoint distances from metric MDS as a function of input dissimilarities (intercity driving distances). Gray reference line has a slope of one and an intercept of zero.

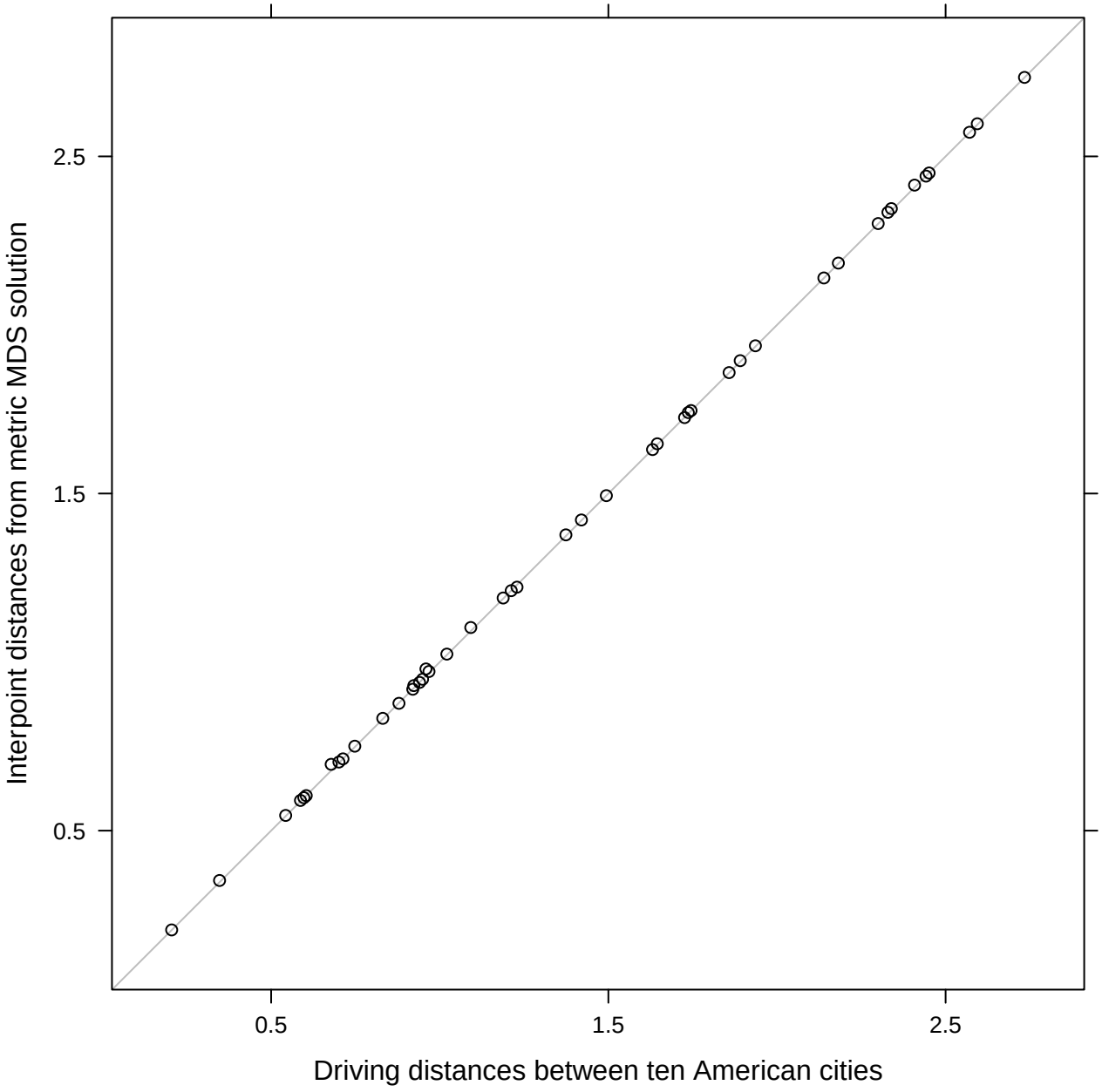


Figure 3: Randomly-located points for ten American cities, to be used as starting configuration for nonmetric MDS of cities' social, economic, and cultural characteristics.

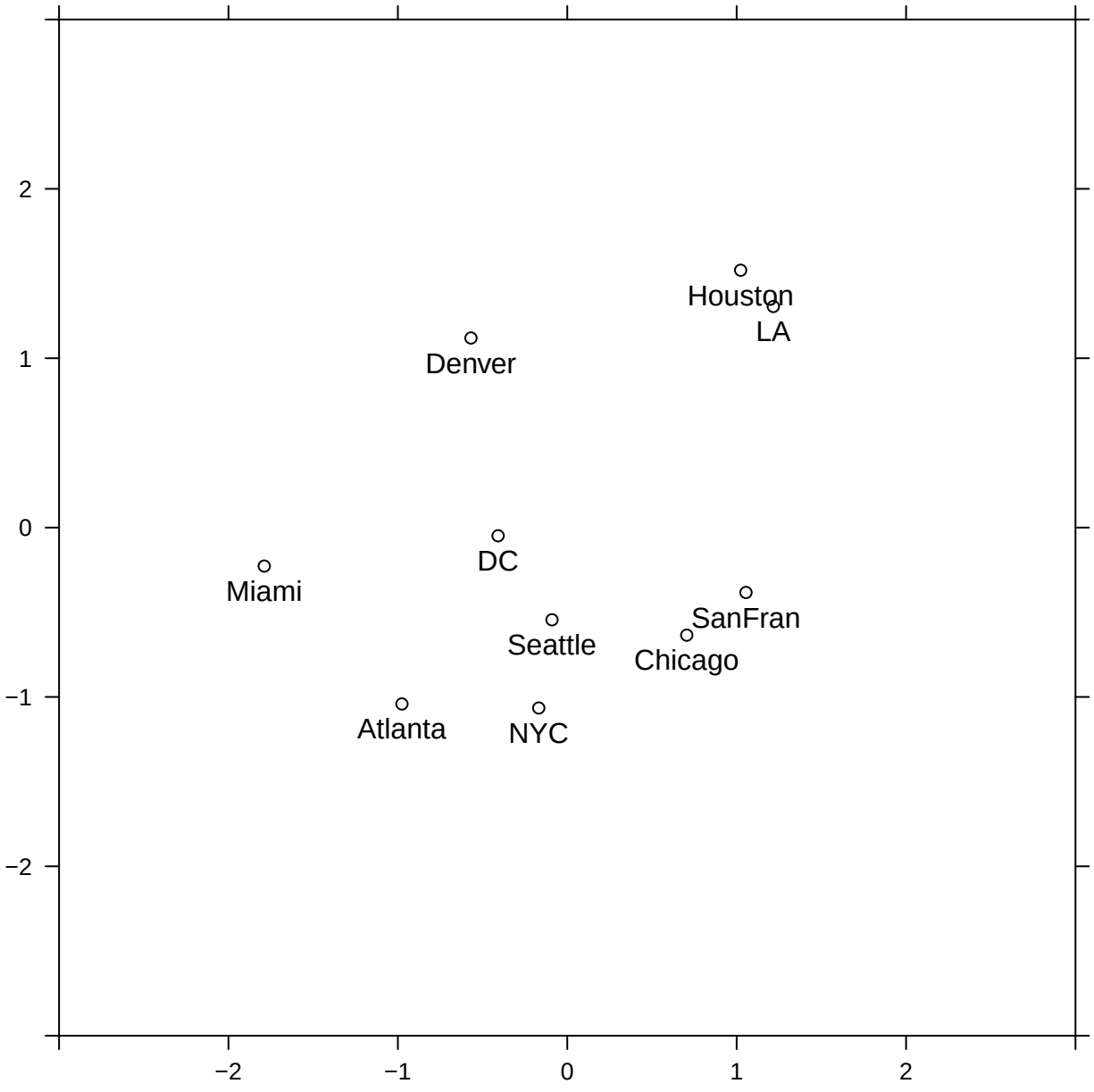


Figure 4: Shepard diagram showing distances between pairs of city points (from random starting configuration) versus input ordinal dissimilarities between pairs of cities.

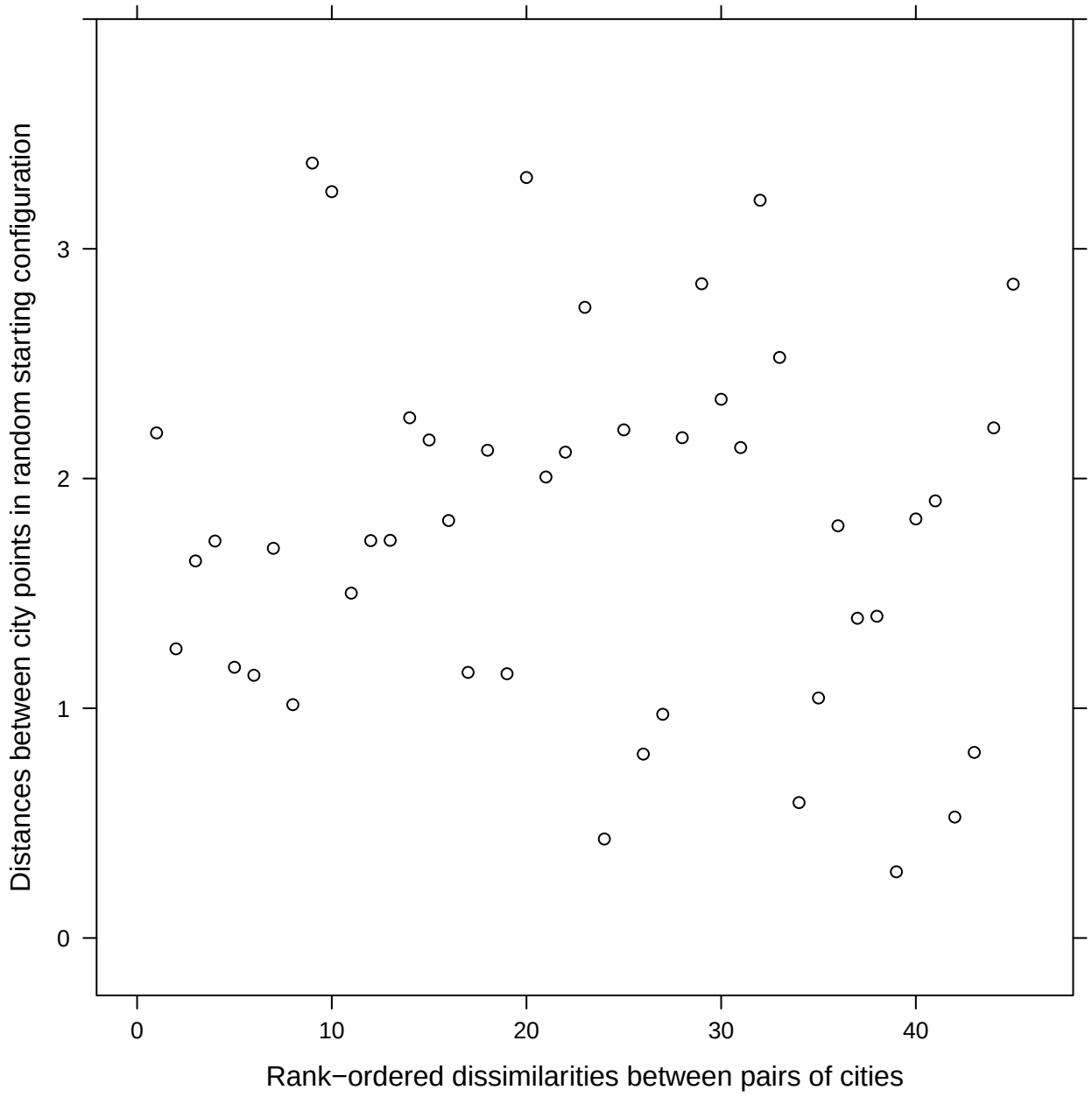
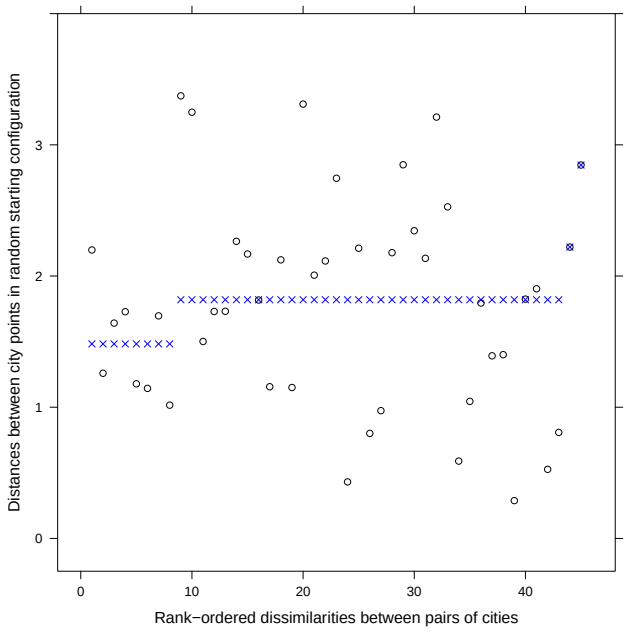


Figure 5: Shepard diagram for random starting configuration in nonmetric MDS of dissimilarities between cities, showing disparities for the scaled distances.

A. Shepard diagram, including disparities (plotted as x's).



B. Shepard diagram, with line segments connecting disparities to corresponding distances.

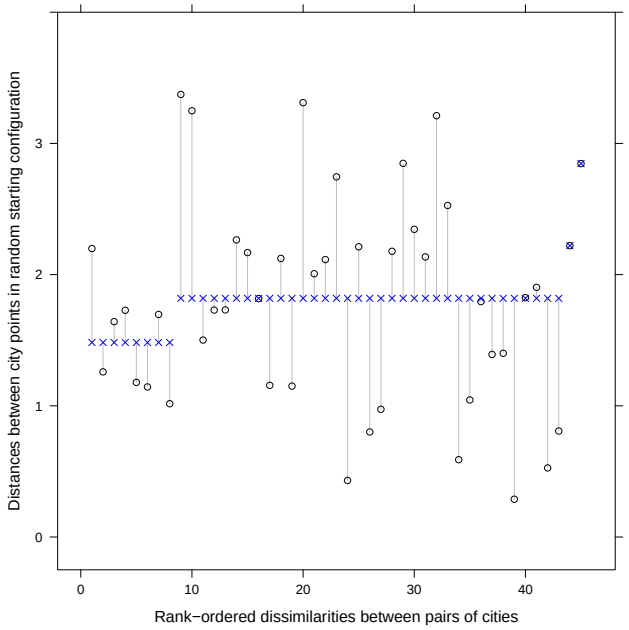
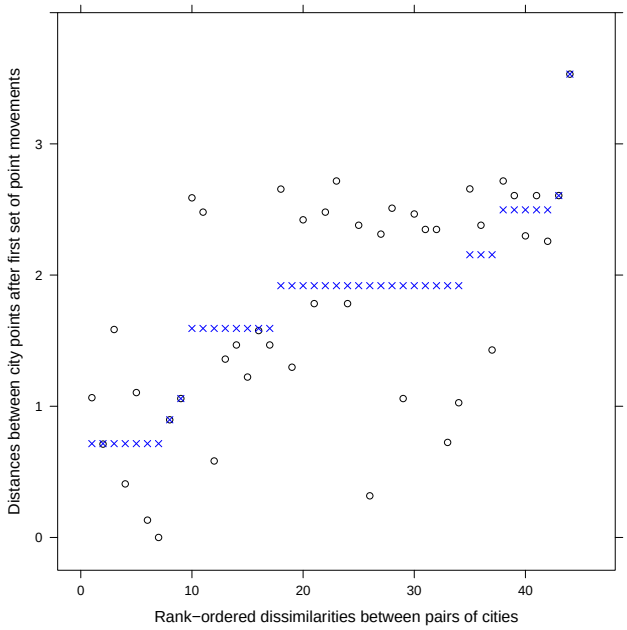


Figure 6: Shepard diagram for interpoint distances after the first set of point movements (obtained by performing metric MDS on the disparities from the random starting configuration).

A. Shepard diagram, including disparities (plotted as x's).



B. Shepard diagram, with line segments connecting disparities to corresponding distances.

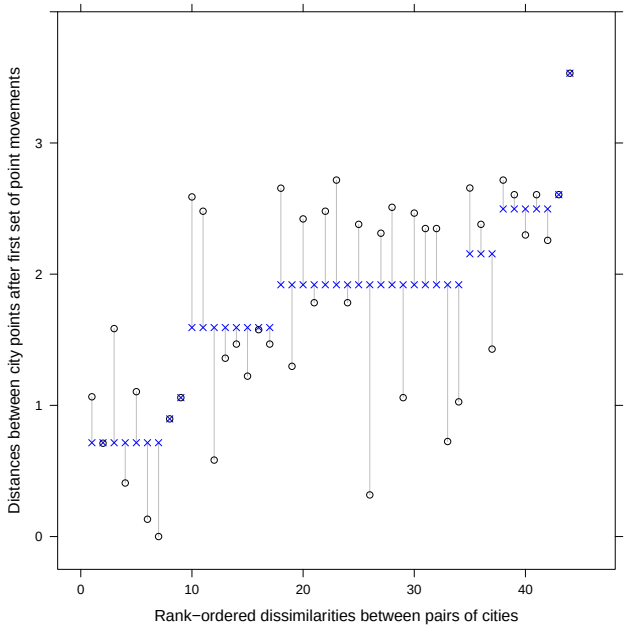
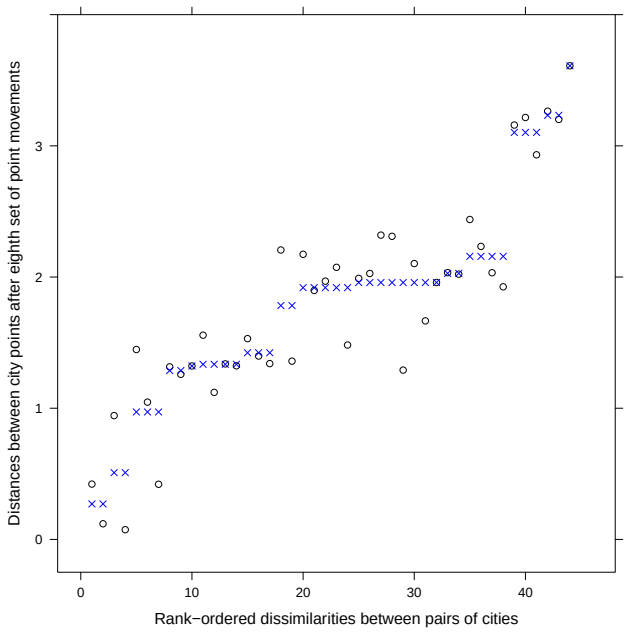


Figure 7: Shepard diagram for interpoint distances after the eighth set of point movements.

A. Shepard diagram, including disparities (plotted as x's).



B. Shepard diagram, with line segments connecting disparities to corresponding distances.

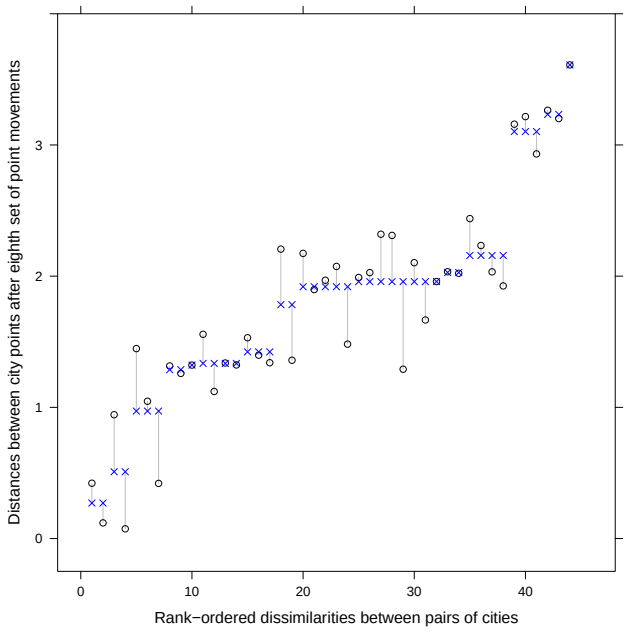


Figure 8: Point configuration obtained from nonmetric multidimensional scaling of social, economic, and cultural dissimilarities among ten American cities.

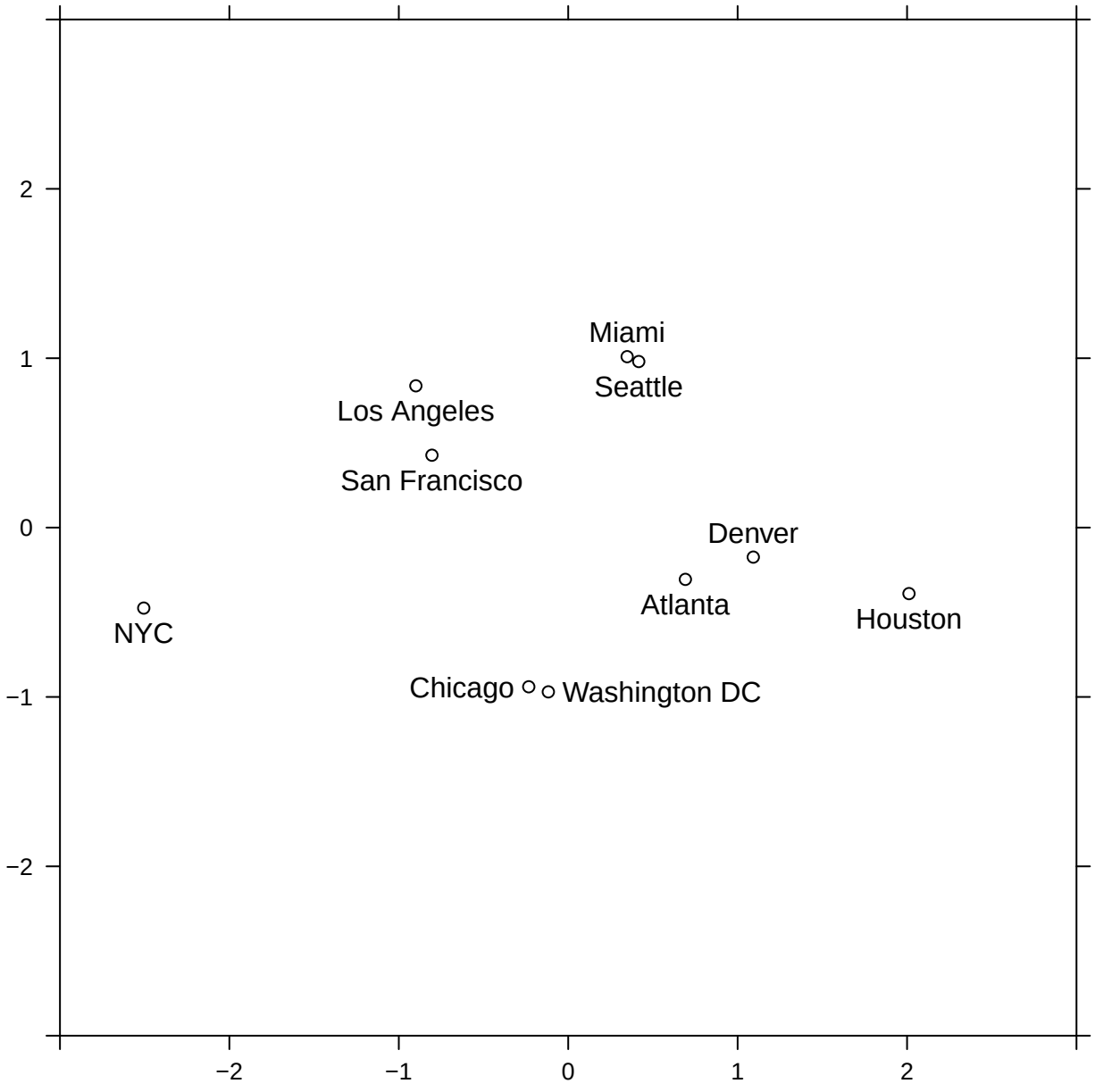


Figure 9: Point configuration obtained from nonmetric multidimensional scaling of political figures from the 2004 U.S. Presidential election.

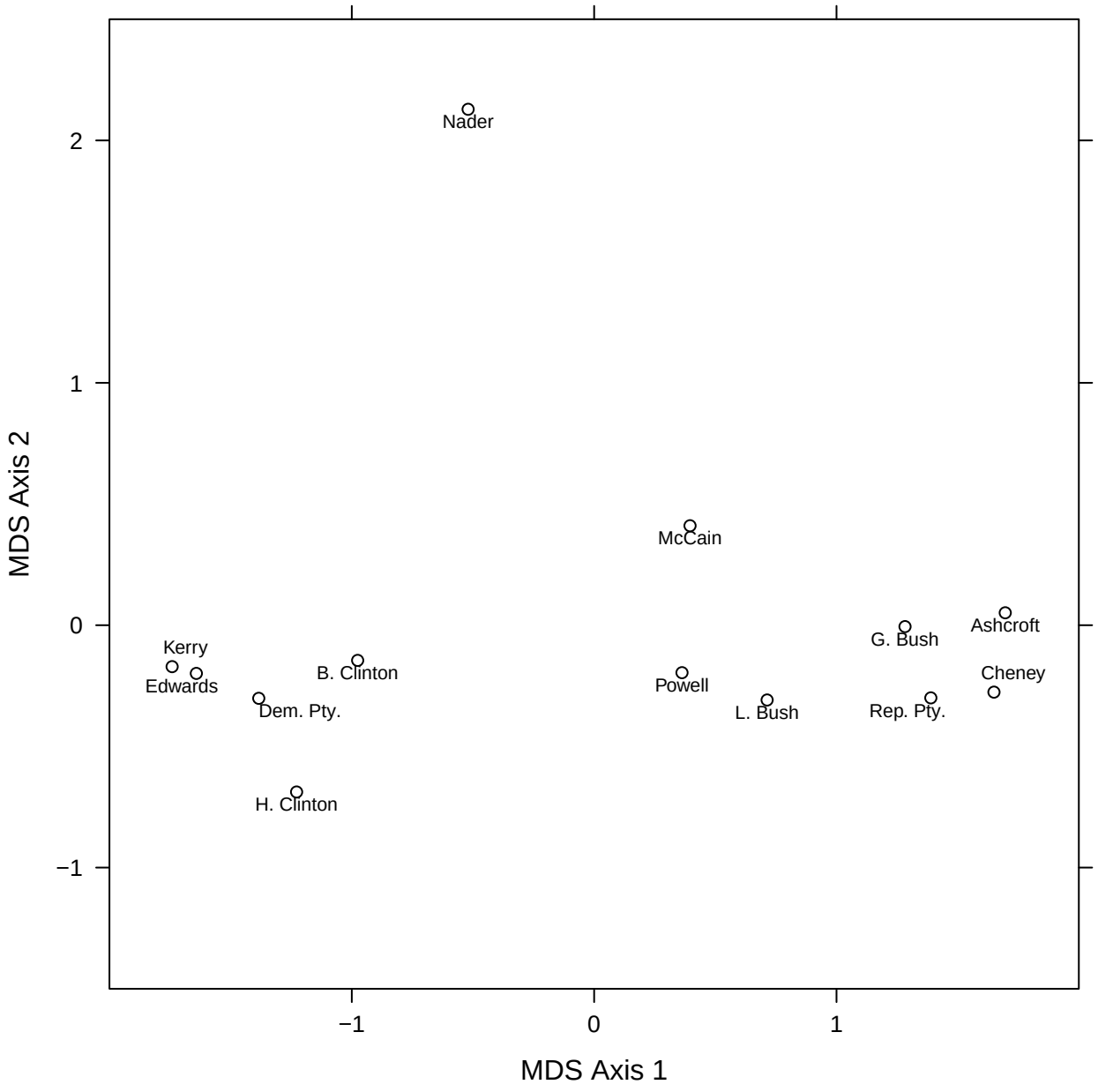


Figure 10: Nonmetric MDS solution for 2004 political figures, with ideological dimension embedded within the point configuration.

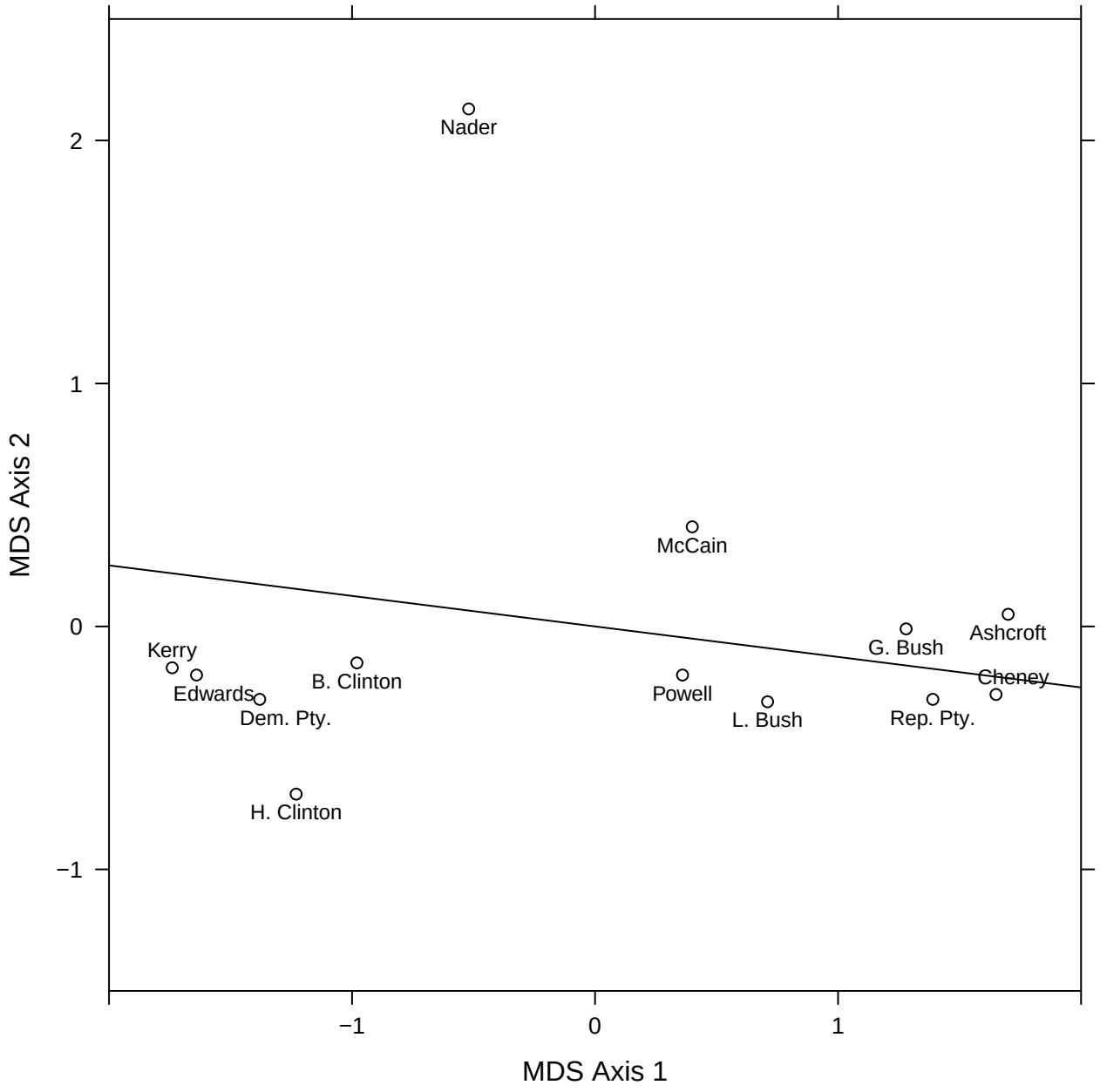


Figure 11: Dendrogram from cluster analysis of point configuration from nonmetric MDS of 2004 political figures.

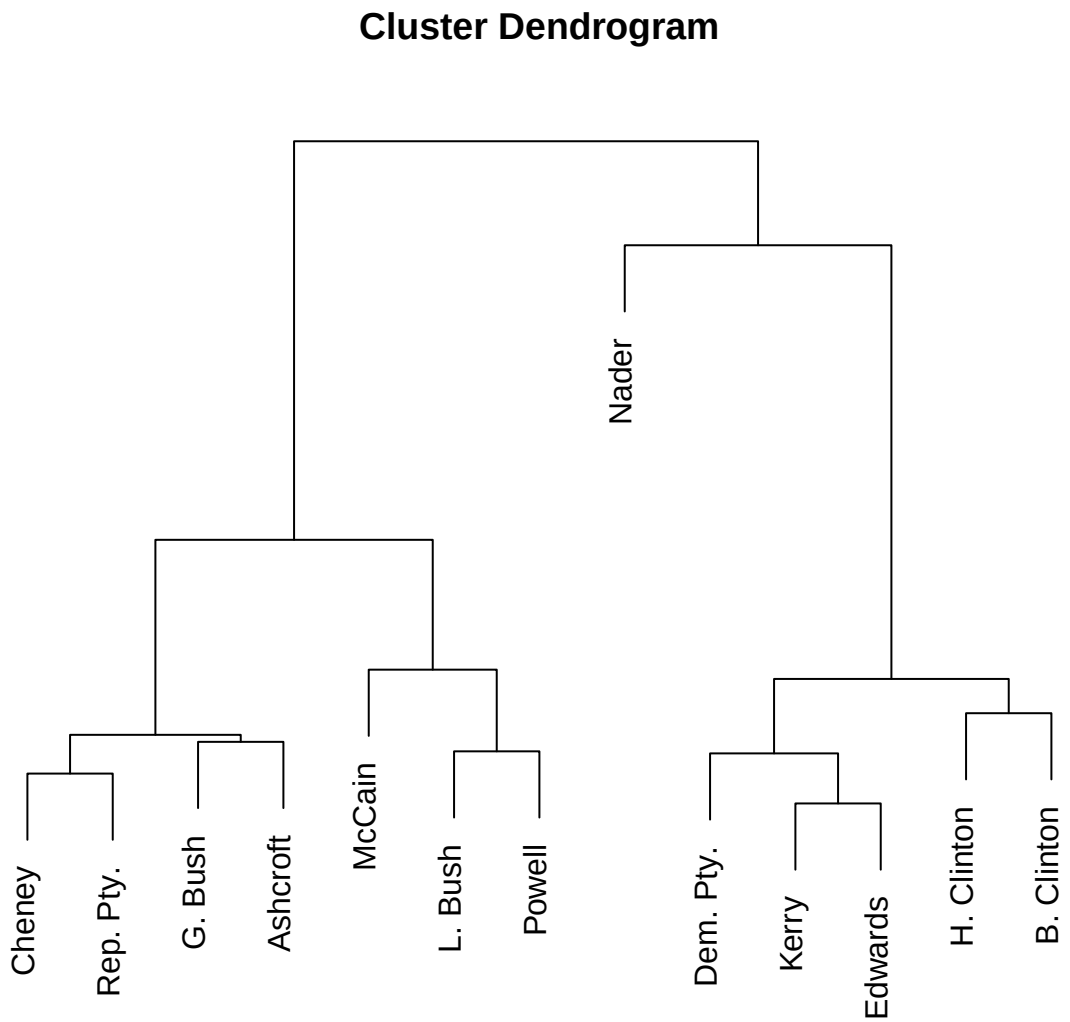


Table 1: Driving distances between ten U.S. cities (in thousands of miles).

Atlanta	Chicago	Denver	Houston	Los Angeles	Miami	New York	San Francisco	Seattle	Washington D. C.	
0	0.587	1.212	0.701	1.936	0.604	0.748	2.139	2.182	0.543	Atlanta
0.587	0	0.920	0.940	1.745	1.188	0.713	1.858	1.737	0.597	Chicago
1.212	0.920	0	0.879	0.831	1.726	1.631	0.949	1.021	1.494	Denver
0.701	0.940	0.879	0	1.374	0.968	1.420	1.645	1.891	1.220	Houston
1.936	1.745	0.831	1.374	0	2.339	2.451	0.347	0.959	2.300	Los Angeles
0.604	1.188	1.726	0.968	2.339	0	1.092	2.594	2.734	0.923	Miami
0.748	0.713	1.631	1.420	2.451	1.092	0	2.571	2.408	0.205	New York
2.139	1.858	0.949	1.645	0.347	2.594	2.571	0	0.678	2.442	San Francisco
2.182	1.737	1.021	1.891	0.959	2.734	2.408	0.678	0	2.329	Seattle
0.543	0.597	1.494	1.229	2.300	0.923	0.205	2.442	2.329	0	Washington D.C.

Table 2: Double-centered version of driving distances between ten U.S. cities.

Atlanta	Chicago	Denver	Houston	Los Angeles	Miami	New York	San Francisco	Seattle	Washington D. C.	
0.537	0.228	-0.348	0.199	-0.808	0.895	0.697	-1.005	-1.050	0.656	Atlanta
0.228	0.263	-0.174	-0.134	-0.594	0.234	0.585	-0.581	-0.315	0.488	Chicago
-0.348	-0.174	0.236	-0.092	0.570	-0.563	-0.504	0.681	0.658	-0.463	Denver
0.199	-0.134	-0.092	0.352	0.029	0.516	-0.124	-0.163	-0.550	-0.033	Houston
-0.808	-0.594	0.570	0.029	1.594	-1.130	-1.499	1.751	1.399	-1.313	Los Angeles
0.895	0.234	-0.563	0.516	-1.130	1.617	0.920	-1.542	-1.867	0.918	Miami
0.697	0.585	-0.504	-0.124	-1.499	0.920	1.416	-1.583	-1.130	1.222	New York
-1.005	-0.581	0.681	-0.163	1.751	-1.542	-1.583	2.028	1.846	-1.432	San Francisco
-1.050	-0.315	0.658	-0.550	1.399	-1.867	-1.130	1.846	2.124	-1.115	Seattle
0.656	0.488	-0.463	-0.033	-1.313	0.918	1.222	-1.432	-1.115	1.071	Washington D.C.

Table 3: First two eigenvectors and eigenvalues from the double-centered matrix of driving distances between ten U.S. cities.

First two eigenvectors of double-centered data matrix, $\mathbf{\Delta}^*$:		First two eigenvalues of double-centered data matrix, $\mathbf{\Delta}^*$:	
-0.23217	-0.11011	9.58217	1.68664
-0.12340	0.26253		
0.15554	0.01929		
-0.05216	-0.44079		
0.38889	-0.30037		
-0.36618	-0.44802		
-0.34640	0.39964		
0.45892	-0.08658		
0.43346	0.44649		
-0.31645	0.25843		

Table 4: Point coordinates for metric MDS of driving distances between ten U. S. cities. Columns in table are obtained by multiplying each eigenvector from Table 3 by the square root of its associated eigenvalue.

First MDS point coordinate	Second MDS point coordinate	
-0.71867	-0.14300	Atlanta
-0.38197	0.34095	Chicago
0.48149	0.02505	Denver
-0.16147	-0.57246	Houston
1.20382	-0.39009	Los Angeles
-1.13352	-0.58185	Miami
-1.07228	0.51901	New York
1.42058	-0.11244	San Francisco
1.34179	0.57986	Seattle
-0.97958	0.33562	Washington D.C.

Table 5: Rank-ordered dissimilarities of ten U.S. cities, based upon their social, economic, and cultural characteristics.

Atlanta	Chicago	Denver	Houston	Los Angeles	Miami	New York	San Francisco	Seattle	Washington D. C.	
0	12	1	10	32	19	43	31	8	6	Atlanta
12	0	15	28	21	33	27	24	26	2	Chicago
1	15	0	3	36	16	44	25	13	5	Denver
10	28	3	0	39	20	45	41	30	18	Houston
32	21	36	39	0	9	23	7	14	22	Los Angeles
19	33	16	20	9	0	40	29	4	37	Miami
43	27	44	45	23	40	0	38	42	35	New York
31	24	25	41	7	29	38	0	17	11	San Francisco
8	26	13	30	14	4	42	17	0	34	Seattle
6	2	5	18	22	37	35	11	34	0	Washington D.C.

Note: Table entries are rank-ordered versions of profile dissimilarities calculated across the nine social, economic, and cultural scores assigned to the cities in the *Places Rated Almanac*, by Richard Boyer and David Savageau. The data are used here with the kind permission of the publisher, Rand McNally.

Table 6: Perceptual dissimilarities between political figures from the 2004 U.S. presidential election.

G.W. Bush	John Kerry	Ralph Nader	Dick Cheney	John Edwards	Laura Bush	Hillary Clinton	Bill Clinton	Colin Powell	John Ashcroft	John McCain	Democ. Party	Repub. Party	
0.0	73.0	62	8.0	68.0	20.0	51.5	41.0	24.0	7	25.5	50	5.0	G.W. Bush
73.0	0.0	56	78.0	1.0	54.0	15.0	17.0	47.0	77	37.0	2	74.5	Kerry
62.0	56.0	0	72.0	59.0	53.0	60.0	49.0	58.0	70	39.0	57	71.0	Nader
8.0	78.0	72	0.0	74.5	25.5	65.0	51.5	29.0	12	30.0	66	4.0	Cheney
68.0	1.0	59	74.5	0.0	44.0	14.0	16.0	46.0	76	38.0	3	69.0	Edwards
20.0	54.0	53	25.5	44.0	0.0	42.0	34.0	9.5	23	22.0	45	18.0	L. Bush
51.5	15.0	60	65.0	14.0	42.0	0.0	19.0	32.0	67	40.0	13	55.0	H. Clinton
41.0	17.0	49	51.5	16.0	34.0	19.0	0.0	31.0	61	36.0	11	48.0	B. Clinton
24.0	47.0	58	29.0	46.0	9.5	32.0	31.0	0.0	28	9.5	35	21.0	Powell
7.0	77.0	70	12.0	76.0	23.0	67.0	61.0	28.0	0	33.0	63	6.0	Ashcroft
25.5	37.0	39	30.0	38.0	22.0	40.0	36.0	9.5	33	0.0	43	27.0	McCain
50.0	2.0	57	66.0	3.0	45.0	13.0	11.0	35.0	63	43.0	0	64.0	Dem. Pty.
5.0	74.5	71	4.0	69.0	18.0	55.0	48.0	21.0	6	27.0	64	0.0	Rep. Pty.

Note: Table entries are line-of-sight dissimilarities (Rabinowitz 1976) calculated from feeling thermometer ratings in the 2004 CPS American National Election Study.

Table 7: Nonmetric MDS point coordinates and ideology scores for political figures from the 2004 U.S. Presidential election.

	First MDS Coordinate	Second MDS Coordinate	Ideology Score
George W. Bush	0.615	-0.022	36.359
John Kerry	-0.830	-0.079	-21.135
Ralph Nader	-0.234	1.030	-0.010
Dick Cheney	0.778	-0.157	28.010
John Edwards	-0.785	-0.087	-20.614
Laura Bush	0.356	-0.132	20.288
Hillary Clinton	-0.600	-0.311	-26.450
Bill Clinton	-0.484	-0.075	-24.178
Colin Powell	0.190	-0.086	12.810
John Ashcroft	0.809	-0.005	24.423
John McCain	0.198	0.200	4.499
Dem. Pty.	-0.673	-0.135	-16.563
Rep. Pty.	0.661	-0.140	28.985

Note: MDS coordinates are obtained from nonmetric multidimensional scaling analysis of data in Table 6. Ideology variable consists of mean scores assigned by experts on a scale from 100 for Maximally liberal to +100 for maximally conservative.