MULTIDIMENSIONAL SCALING: AN INTRODUCTION

William G. Jacoby
Michigan State University and ICPSR

Workshop in Methods
Indiana University
December 7, 2012

E-mail: *jacoby@msu.edu*
Web: *http://polisci.msu.edu/jacoby/iu/mds2012*

Begin with an Example

Assume we have information about the American electorate's perceptions of thirteen prominent political figures from the period of the 2004 presidential election.

Specifically, we have the perceived *dissimilarities* between all pairs of political figures.

- With 13 figures, there will be 78 distinct pairs of figures.
- Rank-order pairs of political figures, according to their dissimilarity (from least to most dissimilar).

For convenience, arrange the rank-ordered dissimilarity values into a square, symmetric matrix.

Matrix of Perceptual Dissimilarities among 2004 Political Figures

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

Clearly, there is too much information in this matrix to be comprehensible in its “raw” numeric form!

Instead, try “drawing a picture” of the information in the matrix.

Rules for Drawing the Picture

Each political figure is shown as a point.

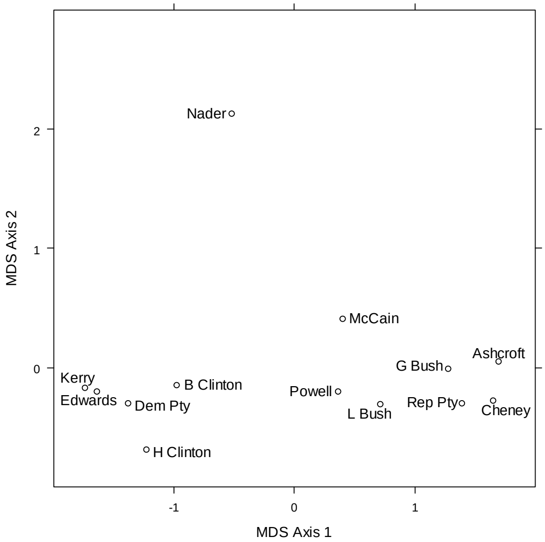
Points are located on the surface of the display medium.

- For example, the projection screen or a sheet of paper
- In any case, a two-dimensional space

Adjust the point locations so that the rank-order of the distances between pairs of points corresponds as closely as possible to the rank-ordered dissimilarities between pairs of political figures.

The process of constructing the picture from the dissimilarities is **multidimensional scaling**.

MDS Point Configuration of 2004 Political Figures



A General Definition of Multidimensional Scaling

A family of procedures for constructing a spatial model of objects, using information about the proximities between the objects.

Different “varieties” of MDS

- Classical multidimensional scaling (CMDS)
- Weighted multidimensional scaling (WMDS)
- Unfolding models (or “ideal points models”)
- Preference mapping (or “the vector model of preferences”)
- Correspondence analysis (CA)

Utility of MDS for Social Research

- Reducing dimensionality
- Modeling perceptions of survey respondents or experimental subjects
- Flexible with respect to input data
- Useful measurement tool
- Graphical output

The Map Analogy

A Familiar Task:

Starting with a map (a geometric model)

Can obtain the distances between locations (numeric data)

MDS “Reverses” the Preceding Task:

We begin with distances (numeric data)

Use that information to produce a map (geometric model)

More Formal Representation

Begin with a $k \times k$ symmetric 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}.

Greater proximity between objects i and j corresponds to *smaller* values of δ_{ij} and vice versa.

Therefore, the proximities are often called "dissimilarities."

Admittedly, this terminology is a bit confusing!

But, there is a reason for it . . .

More Formal Representation (Continued)

MDS 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 coordinate values, $\mathbf{X}$.

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 $\mathbf{X}$, using the Pythagorean formula:

$$d_{ij} = \left[(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 \ldots + (x_{im} - x_{jm})^2\right]^{0.5}$$

We want MDS to find $\mathbf{X}$ such that $d_{ij} \approx \delta_{ij}$ for all i and j.

More Formal Representation (Continued)

Stated a bit differently:

MDS uses the information contained in $\mathbf{\Delta}$ to find $\mathbf{X}$ such that the interpoint distances are functionally related to the pairwise dissimilarities.

For all pairs, i and j, with $i \neq j$:

$$d_{ij} = f(\delta_{ij})$$

The nature of the function, f, is determined by the type of MDS that is performed on the dissimilarities data.

Metric MDS

Metric multidimensional scaling requires that distances are related to dissimilarities by a linear function:

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

In the preceding formula, a and b are coefficients to be estimated, and e_{ij} is an error term associated with objects i and j.

Stated differently, metric MDS assumes that the input dissimilarities data are measured at the interval or ratio level.

Torgerson's Procedure for Metric MDS, I

Begin with $k \times k$ dissimilarities matrix, $\mathbf{\Delta}$

Assume the dissimilarities correspond to distances in m-dimensional space (except for random error)

The $k \times k$ matrix of distances between points is $\mathbf{D}$.

The $k \times k$ matrix, $\mathbf{E}$, contains random errors.

Hypothesis:

$$\mathbf{\Delta} = \mathbf{D} + \mathbf{E}$$

Objective (stated informally):

Find the $k \times m$ coordinate matrix, $\mathbf{X}$, such that entries in $\mathbf{E}$ are as close to zero as possible.

Torgeron's Procedure for Metric MDS, II

Create a "double-centered" version of $\mathbf{\Delta}$, designated $\mathbf{\Delta}^*$

Double-centering transforms $\mathbf{\Delta}$, so that the row sums, the column sums, and the overall sum of the cell entries in the matrix are all zero.

For dissimilarity δ_{ij}, the corresponding entry in the double-centered matrix is calculated as follows:

$$\delta_{ij}^* = -0.5(\delta_{ij}^2 - \delta_{i.}^2 - \delta_{.j}^2 + \delta_{..}^2)$$

$\mathbf{\Delta}^*$ can be factored to obtain the matrix of point coordinates:

$$\mathbf{\Delta}^* = \mathbf{X}\mathbf{X}'$$

The factoring process is somewhat analogous to obtaining a square root.

Torgeron's Procedure for Metric MDS, III

The factoring process is carried out by performing an eigendecomposition on $\mathbf{\Delta}^*$

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

$\mathbf{V}$ 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).

Create $\mathbf{X}$ from the first m eigenvectors ($\mathbf{V}_m$) and the first m eigenvalues ($\mathbf{\Lambda}_m^2$):

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

$\mathbf{X}$ contains point coordinates such that the interpoint distances have a least-squares fit to the entries in $\mathbf{\Delta}$.

Example, Using Ten American Cities

Driving distances between ten cities (in thousands of miles):

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 DC

The preceding $\boldsymbol{\Delta}$ is a good example for metric MDS:

We already know m (the dimensionality)

We already know the “shape” of the point configuration

Double-centered version of intercity distance matrix

The following matrix is the $\mathbf{\Delta}^*$ obtained when Torgerson's transformation is applied to all cells of the intercity distance matrix, $\mathbf{\Delta}$.

0.537	0.228	-0.348	0.199	-0.808	0.895	0.697	-1.005	-1.050	0.656
0.228	0.263	-0.174	-0.134	-0.594	0.234	0.585	-0.581	-0.315	0.488
-0.348	-0.174	0.236	-0.092	0.570	-0.563	-0.504	0.681	0.658	-0.463
0.199	-0.134	-0.092	0.352	0.029	0.516	-0.124	-0.163	-0.550	-0.033
-0.808	-0.594	0.570	0.029	1.594	-1.130	-1.499	1.751	1.399	-1.313
0.895	0.234	-0.563	0.516	-1.130	1.617	0.920	-1.542	-1.867	0.918
0.697	0.585	-0.504	-0.124	-1.499	0.920	1.416	-1.583	-1.130	1.222
-1.005	-0.581	0.681	-0.163	1.751	-1.542	-1.583	2.028	1.846	-1.432
-1.050	-0.315	0.658	-0.550	1.399	-1.867	-1.130	1.846	2.124	-1.115
0.656	0.488	-0.463	-0.033	-1.313	0.918	1.222	-1.432	-1.115	1.071

An eigendecomposition is carried out on the preceding $\mathbf{\Delta}^*$.

We expect a two-dimensional solution, so we use the first two eigenvectors and the first two eigenvalues.

Eigenvectors and eigenvalues from $\mathbf{\Delta}^*$

First two eigenvectors of double-centered data matrix, $\mathbf{\Delta}^*$:

-0.23217	-0.11011
-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

First two eigenvalues of double-centered data matrix, $\mathbf{\Delta}^*$:

9.58217	1.68664

Rescaled Eigenvectors

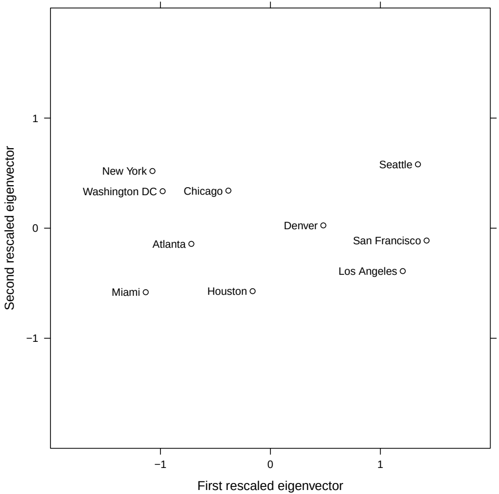
The eigenvectors are multiplied by the square roots of the corresponding eigenvalues to produce the matrix of two-dimensional point coordinates: $\mathbf{X} = \mathbf{V}_2 \mathbf{\Lambda}_2$

-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.

Point coordinates can be plotted in two-dimensional space

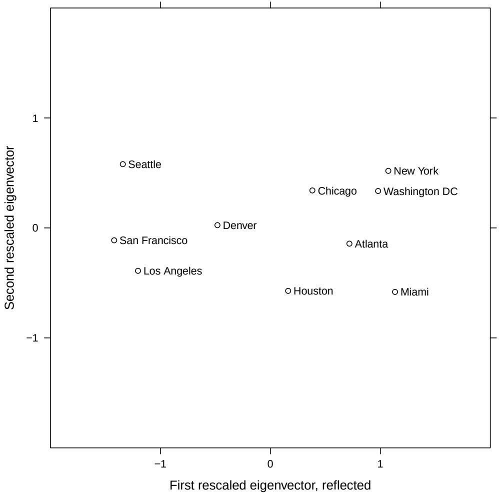
Point Configuration from Metric MDS

Graph of the rescaled eigenvectors:



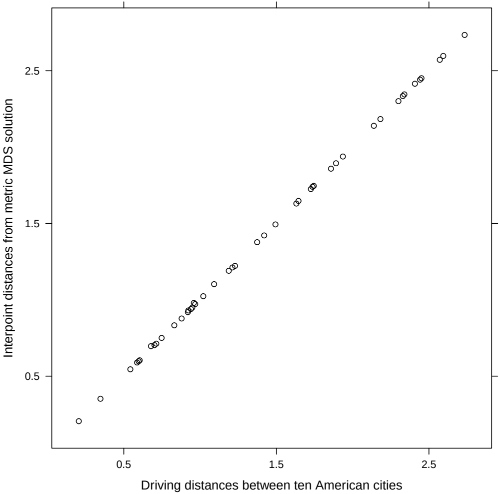
Revised Point Configuration from Metric MDS

Graph of the rescaled eigenvectors, first eigenvector reflected:



Scaled Distances as a Function of Dissimilarities

Graph of scaled interpoint distances versus input dissimilarities data (i.e., driving distances):



Model Fit, Graphical Assessment

Eigenvalues are related to the variance in the double-centered distances that is “explained” by each eigenvector.

The “Scree plot” shows the eigenvalues plotted in the order that they were factored from the dissimilarities matrix.

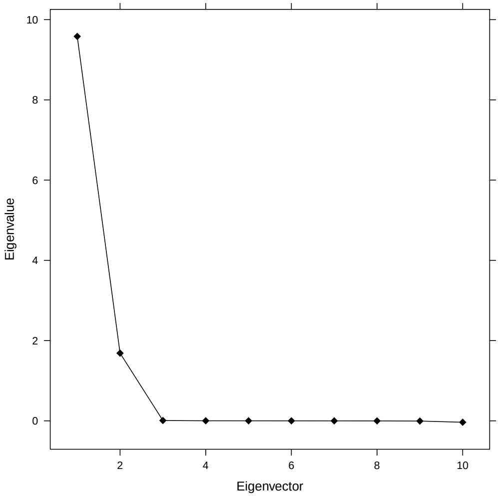
Useful visual representation of model fit

Logic:

- The “important” dimensions in a metric MDS solution should account for a large part of the variance in the dissimilarities data.
- Dimensions associated primarily with error should account for very little variance.

Scree Plot for Metric MDS Solution

Graph of eigenvalues versus order of extraction
in metric MDS of intercity driving distances:



Goodness of Fit Measure for Metric MDS

Eigenvalues measure variance associated with each dimension of the MDS solution

Sum of first m eigenvalues relative to sum of all q eigenvalues (usually $q = k$):

$$Fit = \frac{\sum_{i=1}^{m} \lambda_i^2}{\sum_{i=1}^{q} \lambda_i^2}$$

Here, first two eigenvalues are 9.58 and 1.69, and the sum of the eigenvalues is 11.32.

$$Fit = \frac{9.58 + 1.69}{11.32} = 0.996$$

A Conceptual Leap

So far, we have used physical distances as the input data.

If MDS works for physical distances, then it may also work for data that can be interpreted as "conceptual distances."

Many types of data can be interpreted as conceptual distances.

- Usually correspond to ideas like "proximity" and similarity (or dissimilarity)
- We will say more about this later . . .

Profile Dissimilarities

One important type of conceptual distance data:

Each of k objects has scores on each of v variables.

Each object's vector of scores is called its *profile*.

For each pair of objects, take the sum of squared differences across the v variables and, optionally, take the square root of the sum.

For objects i and j, each of which have scores on variables $x_1, x_2, \ldots x_v$, the profile dissimilarity is:

$$\delta_{ij} = \left[\sum_{l=1}^{v} (x_{il} - x_{jl})^2\right]^{0.5}$$

δ_{ij} is the profile dissimilarity between the two objects; it can be interpreted as the distance between them in v-dimensional space.

Substantive Example, Using Same Ten Cities

Socioeconomic characteristics of ten American cities:

	Climate, Terrain	Housing	Environ., Health	Crime	Transport-ation	Education	The Arts	Recreation	Economics
Atlanta	0.185	-1.338	-0.451	-0.609	0.817	-0.413	-0.700	-1.352	0.327
Chicago	-0.942	-0.350	0.977	-1.139	0.423	1.112	0.431	-0.201	-1.142
Denver	-0.899	-0.397	-0.820	-0.498	0.661	-0.100	-0.640	-0.611	1.451
Houston	-1.500	-0.789	-0.856	-0.239	-1.419	0.347	-0.470	-0.995	1.848
LA	1.356	0.774	0.636	0.652	-1.585	0.077	0.347	0.640	-0.995
Miami	-0.198	-0.596	-0.941	1.617	-1.078	-1.046	-0.879	0.911	-0.301
NYC	-0.174	0.580	2.135	1.692	0.945	-0.673	2.520	0.355	-0.966
SF	1.511	2.026	-0.156	-0.007	0.752	0.703	-0.264	1.142	-0.006
Seattle	0.879	-0.628	-0.718	-0.876	-0.231	-1.647	-0.568	1.297	-0.220
DC	-0.217	0.719	0.196	-0.591	0.715	1.642	0.225	-1.186	0.005

Table entries are standardized versions of scores assigned to cities in *Places Rated Almanac*, by Richard Boyer and David Savageau. The data are used here with the kind permission of the publisher, Rand McNally. For all but two of the above criteria, the higher the score, the better. For Housing and Crime, the lower the score the better.

Use this information to calculate the 10×10 matrix, $\boldsymbol{\Delta}$, of profile dissimilarities between the cities.

Profile Dissimilarities Matrix

Profile dissimilarities matrix, $\boldsymbol{\Delta}^*$:

Atlanta	0.000	3.438	2.036	3.394	4.630	3.930	5.555	4.615	3.330	3.168
Chicago	3.438	0.000	3.635	4.364	3.964	4.740	4.357	4.253	4.299	2.296
Denver	2.036	3.635	0.000	2.333	4.849	3.794	5.674	4.285	3.606	2.994
Houston	3.394	4.364	2.333	0.000	5.016	3.959	6.453	5.517	4.588	3.911
Los Angeles	4.630	3.964	4.849	5.016	0.000	3.361	4.181	3.180	3.611	4.039
Miami	3.930	4.740	3.794	3.959	3.361	0.000	5.234	4.471	2.959	4.906
New York	5.555	4.357	5.674	6.453	4.181	5.234	0.000	4.930	5.535	4.796
San Francisco	4.615	4.253	4.285	5.517	3.180	4.471	4.930	0.000	3.894	3.422
Seattle	3.330	4.299	3.606	4.588	3.611	2.959	5.535	3.894	0.000	4.744
Washington DC	3.168	2.296	2.994	3.911	4.039	4.906	4.796	3.422	4.744	0.000

Apply Torgerson’s formula to double-center this matrix.

Double-Centered Profile Dissimilarities

The $\mathbf{\Delta}^*$ matrix:

Atlanta	5.668	0.028	3.373	2.289	-4.232	-0.963	-4.318	-3.394	0.925	0.623
Chicago	0.028	6.211	-0.889	-1.201	-1.097	-4.205	1.890	-1.515	-2.498	3.277
Denver	3.373	-0.889	5.226	5.104	-5.490	-0.658	-5.212	-2.143	-0.251	0.940
Houston	2.289	-1.201	5.104	10.430	-3.712	1.303	-7.334	-5.579	-1.674	0.374
Los Angeles	-4.232	-1.097	-5.490	-3.712	7.310	1.931	3.191	3.021	0.771	-1.693
Miami	-0.963	-4.205	-0.658	1.303	1.931	7.851	-1.499	-1.644	3.183	-5.298
New York	-4.318	1.890	-5.212	-7.334	3.191	-1.499	16.554	0.548	-3.404	-0.416
San Francisco	-3.394	-1.515	-2.143	-5.579	3.021	-1.644	0.548	8.849	0.477	1.379
Seattle	0.925	-2.498	-0.251	-1.674	0.771	3.183	-3.404	0.477	7.275	-4.805
Washington DC	0.623	3.277	0.940	0.374	-1.693	-5.298	-0.416	1.379	-4.805	5.619

Use eigendecomposition to obtain metric MDS solution

But, an important preliminary question . . .

Assessing Dimensionality

What is appropriate dimensionality for the metric MDS solution?

Could be any number up to $k - 1$

Competing considerations:

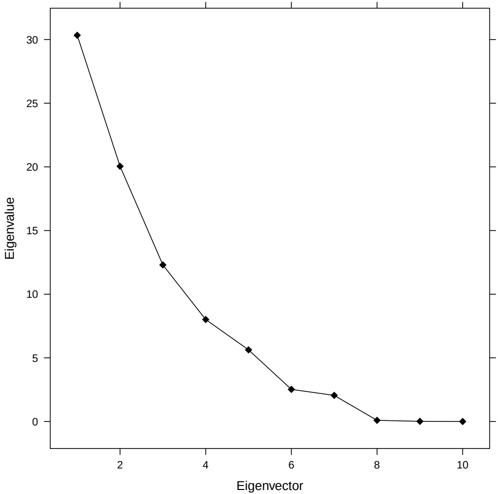
Enough dimensions to account for sufficient variance

Relatively few dimensions to facilitate interpretation

Try examining the scree plot.

Evidence from Scree Plot

Scree plot for eigendecomposition of $\mathbf{\Delta}^*$ matrix:



Interpreting Scree Plot Evidence

No obvious “elbow” in scree plot

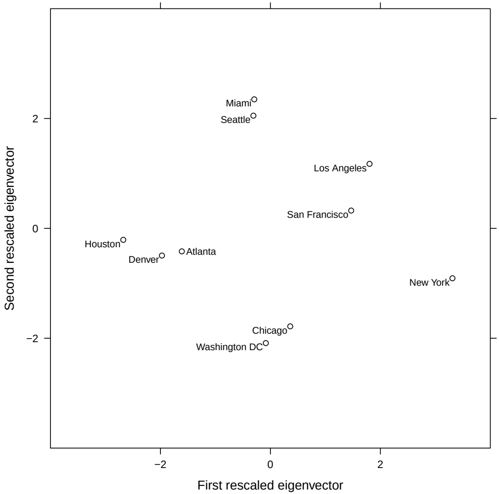
> No clear distinction between “important” and “unimportant” dimensions, in terms of variance explained.

Use parsimony for guidance

> Use $m = 2$ so we can visualize the MDS solution easily.

Metric MDS Solution

Two-dimensional point configuration obtained from metric MDS of double- centered matrix of profile dissimilarities.



Goodness of Fit

Calculate fit statistic for the metric MDS solution:

$$Fit = \frac{\sum_{i=1}^{m} \lambda_i^2}{\sum_{i=1}^{q} \lambda_i^2} = \frac{30.337 + 20.046}{81} = 0.622$$

The two-dimensional MDS solution “explains” about 60 percent of the variance in the profile dissimilarities

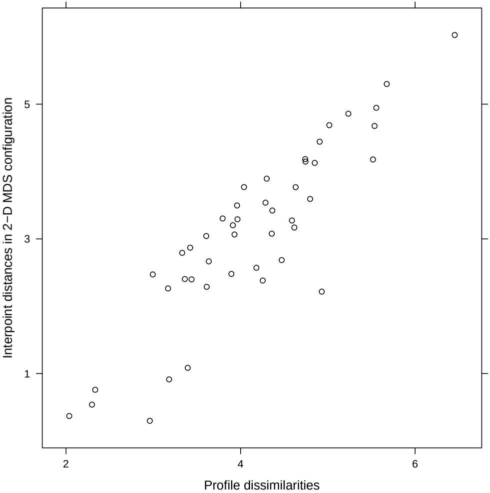
Remaining variance is error

We have assumed the error is random

Is it?

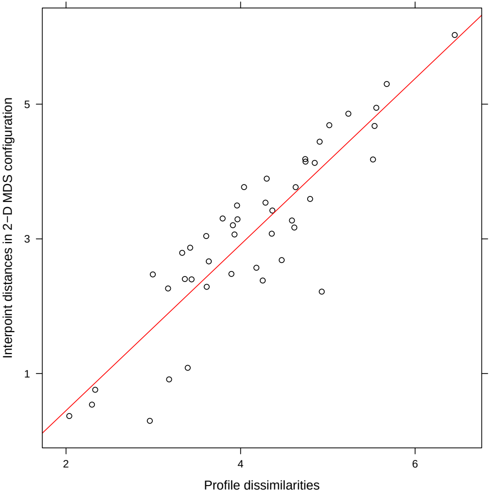
Assessing Errors in Metric MDS Solution

Shepard Diagram for Metric MDS of City Characteristics:



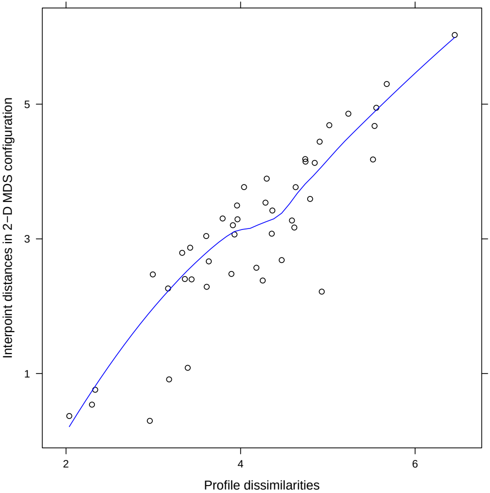
Assessing Errors in Metric MDS Solution

Shepard Diagram for Metric MDS of City Characteristics:



Assessing Errors in Metric MDS Solution

Shepard Diagram for Metric MDS of City Characteristics:



A New Concern

Distances in MDS solution do not appear to be a linear function of the dissimilarities.

Instead, distances seem to be monotonically related to dissimilarities.

Distances are monotonically related to dissimilarities if, for all i, j, and l, the following holds:

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

Distances Monotonic to Dissimilarities

Rather than assuming distances should be a linear function of the dissimilarities, perform MDS assuming that d_{ij}'s are a monotonic function of δ_{ij}'s:

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

In this expression, f^m means "a monotonic function" and e_{ij} is an error term.

But, do we really need to worry about this?

What if we simply treated ordinal dissimilarities data as interval-level?

To see what happens, we can analyze the *rank-order* of the driving distances between the ten American cities.

Again, a useful example because we know the "true" solution

Metric MDS of Ordinal Dissimilarities, I

Matrix of rank-ordered distances between ten American cities:

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

Apply Torgerson's double-centering transformation and perform metric MDS.

Metric MDS of Ordinal Dissimilarities, II

Double-centered version of rank-ordered distance matrix:

124.6	89.4	-147.6	91.4	-284.6	273.4	212.8	-266.8	-290.0	197.4
89.4	70.2	-17.3	-16.4	-214.3	43.6	195.1	-193.4	-119.2	162.2
-147.6	-17.3	64.2	21.1	202.6	-159.4	-132.0	187.5	147.2	-166.4
91.4	-16.4	21.1	122.0	4.1	128.0	-51.0	-47.6	-187.8	-64.0
-284.6	-214.3	202.6	4.1	462.2	-300.4	-450.4	512.5	382.2	-313.9
273.4	43.6	-159.4	128.0	-300.4	458.0	229.5	-455.6	-487.8	270.6
212.8	195.1	-131.9	-51.0	-450.4	229.5	401.0	-440.6	-303.8	339.5
-266.8	-193.4	187.5	-47.6	512.5	-455.6	-440.6	566.8	554.6	-417.6
-290.0	-119.2	147.2	-187.8	382.2	-487.8	-303.8	554.6	591.4	-286.8
197.4	162.2	-166.4	-64.0	-313.9	270.6	339.5	-417.6	-286.8	279.0

Perform eigendecomposition on double-centered matrix in order to obtain point coordinates.

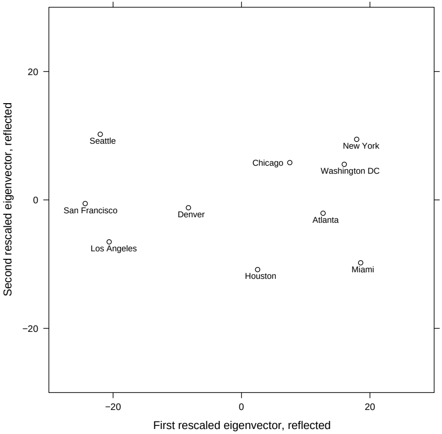
Metric MDS of Ordinal Dissimilarities, III

The point coordinate matrix, $\mathbf{X} = \mathbf{V}_2\boldsymbol{\Lambda}_2$, obtained from eigendecomposition of the double-centered, rank-ordered, distance matrix:

	Dim 1	Dim 2
Atlanta	12.699841	-2.0733232
Chicago	7.518241	5.8125445
Denver	-8.266138	-1.2161737
Houston	2.508042	-10.8456988
LA	-20.611218	-6.5285933
Miami	18.556184	-9.7896715
NYC	17.935017	9.4370549
SF	-24.342714	-0.5662753
Seattle	-21.999173	10.2237850
DC	16.001918	5.5463514

Metric MDS of Ordinal Dissimilarities, IV

Two-dimensional configuration obtained from metric MDS of rank-ordered intercity distances:



Metric MDS of Ordinal Dissimilarities, V

Goodness of fit for metric MDS solution obtained from ordinal dissimilarities data:

$$Fit = \frac{\sum_{i=1}^{m} \lambda_i^2}{\sum_{i=1}^{q} \lambda_i^2} = \frac{2715.84+520.32}{3258.34} = 0.9932$$

Point configuration still provides excellent fit, even though there appears to be less information conveyed in the data matrix.

Assessment

Metric MDS of ordinal data seems to work . . .

But, it is problematic

- It is “cheating” with respect to data characteristics.
- It imposes an implicit assumption about relative sizes of differences between dissimilarities
- Concept of “variance” is undefined for ordinal data
- Therefore, it is inappropriate to use the eigendecomposition, which maximizes variance explained by successive dimensions.

For these reasons, use a different strategy with ordinal dissimilarities data!

Nonmetric Multidimensional Scaling

General strategy:

Obtain scaling solution in a specified dimensionality

If fit is adequate, stop and report results

If fit is poor, try solution in higher dimensionality

General procedure:

Begin with initial configuration of points in space.

Move points around, as necessary, to make distances between points monotonic with dissimilarities.

Proceed Using an Example

Dissimilarities among four American presidential candidates: Santorum; Gingrich; Romney; Obama

This is a very small dataset!

- Useful for illustrating the steps of the scaling process
- Note that we would never perform a "real" nonmetric MDS with only four objects (which produce six dissimilarities).
- Too few metric constraints— the scaling solution (i.e., the point locations) could be changed without affecting the fit to the data.
- This problem is alleviated when more objects are included in the scaling analysis.

Hypothetical Data Matrix, $\boldsymbol{\Delta}$

Matrix of rank-ordered dissimilarities
among four presidential candidates:

	Santorum	Gingrich	Romney	Obama
Santorum	—	1	5	6
Gingrich	1	—	2	3
Romney	5	2	—	4
Obama	6	3	4	—

It is easy to demonstrate that these dissimilarities cannot be represented accurately with a unidimensional array of points.

So, try a two-dimensional solution . . .

Strategy for Nonmetric MDS

- Start with random configuration of points in two-dimensional space
  - We do not take this configuration seriously as a scaling solution; it just provides a “neutral starting position.”
  - Use Pythagorean formula to calculate distances between points in the random configuration
  - Distances should be monotonic to dissimilarities data, but they probably are not.
- Move points around to create a new configuration that is closer to the objective of a monotonic relationship between dissimilarities and distances
  - Want to be as efficient with movements as possible— no unnecessary movements.

Distances and Disparities

Create a set of "target distances" that can be compared to current actual distances in order to guide point movements

These target distances are called "disparities" in the MDS literature

- There is a disparity (target distance) associated with each distance between a pair of points.
- The disparity associated with distance d_{ij} is designated $\hat{d}_{ij}$.

The disparities guide the point movements

- If two points are too close together, then the disparity will be larger than the current distance (i.e., $\hat{d}_{ij} > d_{ij}$); if they are too far apart, then the disparity will be smaller than the current distance (i.e., $\hat{d}_{ij} < d_{ij}$)

Properties of Disparities

Disparities are characterized by two important properties:

1. Disparities are as similar to the actual distances as possible

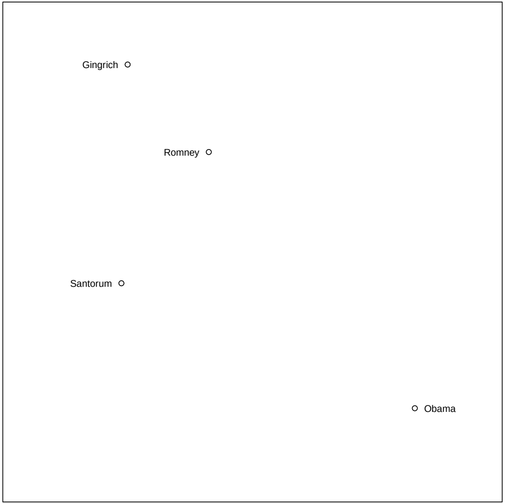
   That is, the correlation between the distances and the disparities ($r_{d_{ij}\hat{d}_{ij}}$) is maximized.

2. Disparities are *always* monotonic to dissimilarities, even if the associated distances are not.

   That is, if $\delta_{ij} < \delta_{il}$ then $\hat{d}_{ij} \leq \hat{d}_{il}$, even if $d_{ij} > d_{il}$

Random Starting Configuration

Point configuration obtained by generating random coordinates for the four presidential candidates



First Set of Distances and Disparities

Object pair, ordered by dissimilarity	Distance, d_{ij}	Disparity, $\hat{d}_{ij}$	Difference $d_{ij} - \hat{d}_{ij}$
Santorum-Gingrich	3.50	2.70	0.80
Gingrich-Romney	1.91	2.70	-0.79
Gingrich-Obama	7.17	4.98	2.19
Romney-Obama	5.26	4.98	0.28
Santorum-Romney	2.52	4.98	-2.46
Santorum-Obama	5.11	5.11	0.00

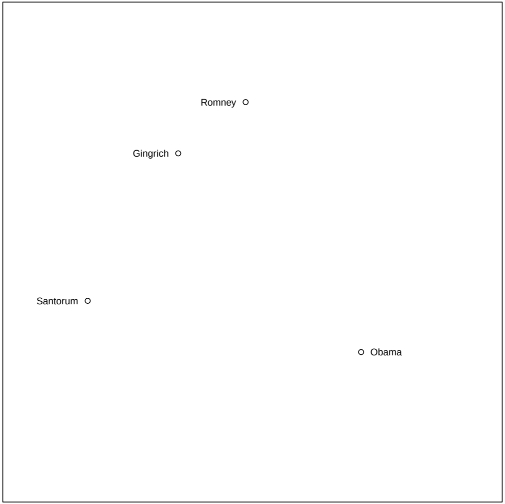
Disparities and Point Movements

Use difference between distances and disparities to guide point movements:

- For each pair, move the points along the line connecting them.
- If $d_{ij} - \hat{d}_{ij}$ is a positive value, current distance is larger than the target— so, move points representing *i* and *j* closer together.
- If $d_{ij} - \hat{d}_{ij}$ is a negative value, current distance is smaller than the target— so, move points representing *i* and *j* farther apart.
- Size of movement is determined by absolute value of $d_{ij} - \hat{d}_{ij}$
- For present purposes, does not really matter which pair of points is moved first.

Second Point Configuration

Point configuration obtained after making first set of moves, based upon previously-calculated disparities:

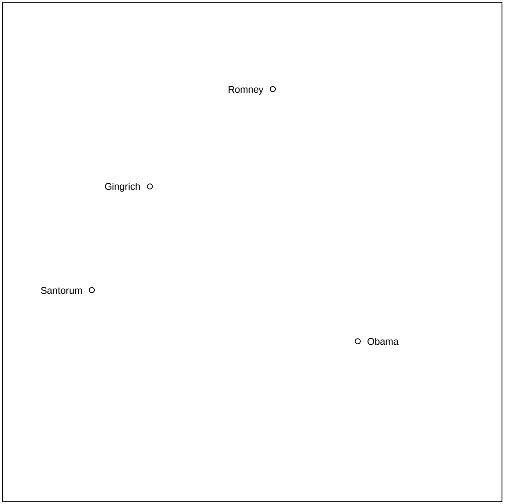


Second Set of Distances and Disparities

Object pair, ordered by dissimilarity	Distance, d_{ij}	Disparity, $\hat{d}_{ij}$	Difference $d_{ij} - \hat{d}_{ij}$
Santorum-Gingrich	2.77	2.065	0.705
Gingrich-Romney	1.36	2.065	-0.705
Gingrich-Obama	4.32	4.263	0.057
Romney-Obama	4.41	4.263	0.147
Santorum-Romney	4.06	4.263	-0.203
Santorum-Obama	4.46	4.460	0.000

Third Point Configuration

Point configuration obtained after making second set of moves, based upon second set of calculated disparities:



Third Set of Distances and Disparities

Object pair, ordered by dissimilarity	Distance, d_{ij}	Disparity, $\hat{d}_{ij}$	Difference $d_{ij} - \hat{d}_{ij}$
Santorum-Gingrich	1.90	1.90	0
Gingrich-Romney	2.51	2.51	0
Gingrich-Obama	4.15	4.15	0
Romney-Obama	4.26	4.26	0
Santorum-Romney	4.33	4.33	0
Santorum-Obama	4.34	4.34	0

Final Nonmetric MDS Solution

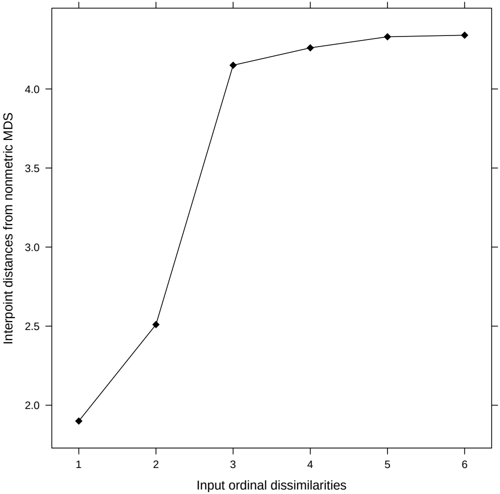
With the third set of points, the disparities are always equal to the actual interpoint distances

- The distances are monotonic to the ordinal dissimilarities
- There is no reason to carry out further point movements
- The objective of the nonmetric MDS has been achieved

This solution fits data *perfectly*

Graphical Depiction of Fit

Shepard diagram for simulated nonmetric MDS, showing scaled distances versus input ordinal dissimilarities:



Need for a Measure of Fit

Perfect fit in scaling solution really only due to artificial nature of problem

Problematic, because too few metric constraints in data to create sufficiently stable scaling solution

Using nonmetric MDS on “real” data

- More objects to be scaled
- More difficult to obtain a *perfect* solution
- But, often can reach a *very good* solution
- Need to develop a fit measure that provides a summary of degree to which a nonmetric MDS solution achieves its analytic objective

Fit Statistic for Nonmetric MDS

Logic:

Want distances to be monotonic to dissimilarities

Disparities are as close to distances as possible, but also monotonic to dissimilarities

So, develop a measure that summarizes how different the distances are from the disparities

Kruskal's Stress_1 coefficient:

$$Stress_1 = \left[\frac{\sum_{i \neq j}^{\#pairs} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i \neq j}^{\#pairs} d_{ij}^2} \right]^{0.5}$$

Alternative Fit Measures for Nonmetric MDS

Kruskal's Stress_2 coefficient:

$$Stress_2 = \left[\frac{\sum_{i \neq j}^{\#pairs} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i \neq j}^{\#pairs} (d_{ij} - \bar{d})^2} \right]^{0.5}$$

Correlation between scaled distances and disparities, $r_{d\hat{d}}$

Logic is that disparities are optimally-scaled versions of the input data in the nonmetric MDS.

Several other fit measures are available (e.g., S-Stress; Guttman's uncorrected correlation coefficient)

Procedure for a “Real” Nonmetric MDS Routine

The value of the Stress coefficient is a function of the point coordinates in the current configuration (which are used to calculate the distances and, indirectly, the disparities).

Therefore, can calculate the partial derivatives of Stress, relative to the coordinates:

$$\frac{\partial Stress_1}{\partial x_{ip}}$$

For $i = 1, 2, \ldots, k$ and $p = 1, 2, \ldots, m$

Partial derivatives show how Stress changes when point coordinates are changed by a minute amount

Therefore, change point coordinates in ways that make the partial derivatives the smallest possible negative values.

Steps in a Nonmetric MDS Routine

Step 1: Starting configuration

Step 2: Calculate fit for starting configuration

  If perfect fit, then terminate

Step 3: Calculate partial derivatives, move points

Step 4: Calculate disparities and fit for new configuration

  If perfect fit, then terminate

  If no change, or fit worsens, then terminate

  If fit is improving, go to Step 3 and repeat

Step 5: Terminate MDS routine and print results

Input Data for Nonmetric MDS Example

The following data matrix contains perceptual dissimilarities among 14 stimuli, calculated from the 1992 CPS National Election Study. The dissimilarities are obtained from the line-of-sight procedure and they are measured at the ordinal level:

0.00	84.00	47.00	3.00	82.00	73.00	66.00	36.00	81.00	20.00	22.00	54.00	75.00	2.00	George Bush
84.00	0.00	38.00	76.00	9.00	25.00	14.00	91.00	1.00	65.00	68.00	24.00	5.00	86.00	Bill Clinton
47.00	38.00	0.00	40.00	46.00	42.00	26.00	70.00	39.00	32.00	34.00	33.00	44.00	51.00	Ross Perot
3.00	76.00	40.00	0.00	80.00	60.00	56.00	50.00	74.00	18.00	11.00	49.00	79.00	6.00	Dan Quayle
82.00	9.00	46.00	80.00	0.00	29.00	21.00	90.00	12.00	67.00	72.00	41.00	16.00	83.00	Al Gore
73.00	25.00	42.00	60.00	29.00	0.00	13.00	87.00	19.00	64.00	57.00	30.00	31.00	69.00	Anita Hill
66.00	14.00	26.00	56.00	21.00	13.00	0.00	78.00	7.00	43.00	48.00	10.00	23.00	62.00	Thomas Foley
36.00	91.00	70.00	50.00	90.00	87.00	78.00	0.00	89.00	52.00	55.00	77.00	88.00	35.00	Barbara Bush
81.00	1.00	39.00	74.00	12.00	19.00	7.00	89.00	0.00	59.00	63.00	27.00	4.00	85.00	Hillary Clinton
20.00	65.00	32.00	18.00	67.00	64.00	43.00	52.00	59.00	0.00	8.00	37.00	61.00	17.00	Clarence Thomas
22.00	68.00	34.00	11.00	72.00	57.00	48.00	55.00	63.00	8.00	0.00	45.00	58.00	15.00	Pat Buchanon
54.00	24.00	33.00	49.00	41.00	30.00	10.00	77.00	27.00	37.00	45.00	0.00	28.00	53.00	Jesse Jackson
75.00	5.00	44.00	79.00	16.00	31.00	23.00	88.00	4.00	61.00	58.00	28.00	0.00	71.00	Democ. Party
2.00	86.00	51.00	6.00	83.00	69.00	62.00	35.00	85.00	17.00	15.00	53.00	71.00	0.00	Repub. Party

These dissimilarities are used as input to PROC MDS in the SAS statistical package. The steps in this procedure are virtually identical to those from other MDS routines (though details and output might differ a bit).

SAS Output from PROC MDS

```
                Nonmetric MDS of 1992 Candidates and Political Figures
                     Multidimensional Scaling:  Data=WORK.CAND92
           Shape=TRIANGLE Cond=MATRIX Level=ORDINAL UNTIE Coef=IDENTITY Dim=2
                                 Formula=1 Fit=1

      Mconverge=0.01 Gconverge=0.01 Maxiter=100 Over=2 Ridge=0.0001
```

Iteration	Type	Badness-of-Fit Criterion	Change in Criterion	Convergence Measures: Monotone	Convergence Measures: Gradient
0	Initial	0.108466	.	.	.
1	Monotone	0.072379	0.036086	0.064224	0.583970
2	Gau-New	0.057607	0.014772	.	.
3	Monotone	0.053911	0.003696	0.018435	0.299002
4	Gau-New	0.052272	0.001639	.	.
5	Monotone	0.047879	0.004393	0.019191	0.207558
6	Gau-New	0.047028	0.000851	.	.
7	Monotone	0.045217	0.001811	0.012829	0.155376
8	Gau-New	0.044913	0.000304	.	.
9	Monotone	0.044012	0.000901	0.008895	0.105132
10	Gau-New	0.043765	0.000247	.	0.016806
11	Gau-New	0.043758	0.000006436	.	0.005644

Convergence criteria are satisfied.

SAS Output from PROC MDS

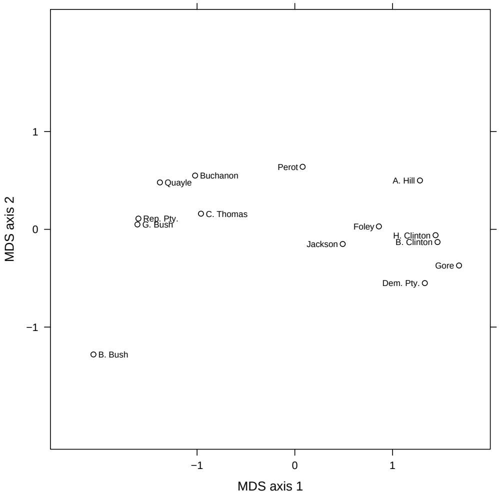
MATRIX	Number of Nonmissing Data	Weight	Badness-of-Fit Criterion	Distance Correlation	Uncorrected Distance Correlation
1	91	1	0.04	1.00	1.00

Configuration

	DIM1	DIM2
G_Bush	-1.61	0.05
B_Clint	1.46	-0.13
Perot	0.08	0.64
Quayle	-1.38	0.48
Gore	1.68	-0.37
A_Hill	1.28	0.50
Foley	0.86	0.03
B_Bush	-2.06	-1.28
H_Clint	1.44	-0.06
C_Thoma	-0.96	0.16
Buchano	-1.02	0.55
Jackson	0.49	-0.15
Dem_Pty	1.33	-0.55
Rep_Pty	-1.60	0.11

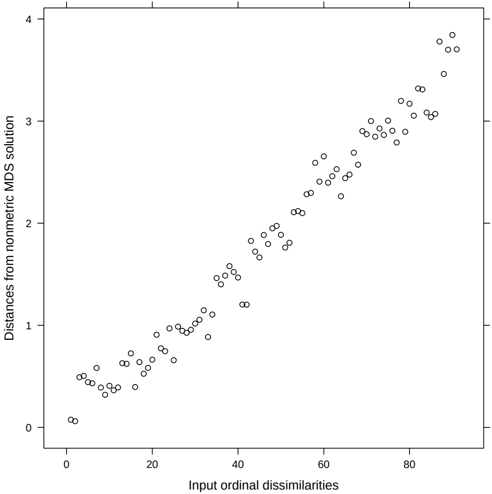
MDS Point Configuration

Two-dimensional point configuration obtained from nonmetric MDS of 1992 political figures



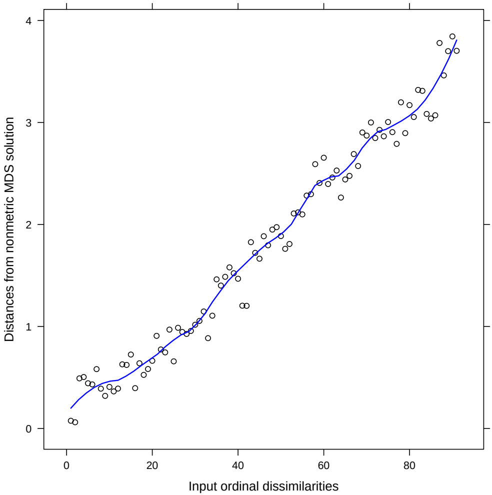
Shepard Diagram

Shepard diagram for two-dimensional point configuration from nonmetric MDS of 1992 political figures:

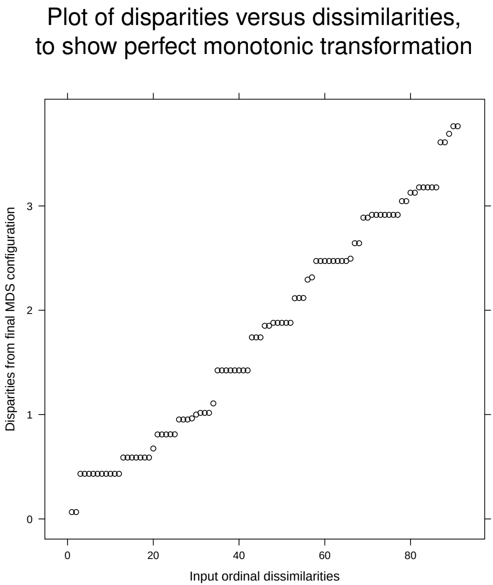


Shepard Diagram

Shepard diagram for two-dimensional point configuration from nonmetric MDS of 1992 political figures:



Dissimilarities and Disparities



Interpreting an MDS Solution

Metric and nonmetric versions of CMDS only determine relative distances between points in scaled *m*-space

The locations of the coordinate axes for the point configuration are completely arbitrary

- Final MDS point configuration usually rotated to a "varimax" orientation
- Point coordinates usually standardized to a mean of zero on each axis and a variance of 1.0 (or some other specified value)

Axes have no intrinsic substantive importance or interpretation!

Interpretation Strategies for MDS

Generally, try to look for two kinds of structure in an MDS solution:

- Interesting *directions* within the *m*-space
  - A direction would usually be “interesting” if points that fall at opposite sides of the space correspond to objects that are contrasting with respect to some substantive characteristic
- Interesting *groups* of points within the *m*-space
  - A grouping of points would be “interesting” if the objects corresponding to the grouped points are differentiated from the other objects in terms of some recognizable substantive characteristic

Of course, both kinds of structure can occur simultaneously, within a single MDS solution.

Some Cautions About Interpretation

Simplicity of underlying model and potential for graphical representation of scaling results both facilitate interpretation

For many purposes, simply "eyeballing" the point configuration is sufficient for interpretation

But, visual interpretation has potential limitations:

- It is much more difficult when $m > 2$, and almost impossible when $m > 3$.
- Highly subjective— we may see structured patterns that are not really there

For these reasons, it is useful to employ more systematic methods for interpreting an MDS solution

Embedding External Variables

Researcher often has prior hypotheses about dimensions that differentiate objects in MDS analysis

Useful to obtain “external” measures of the objects along these dimensions, separate from the data employed for the MDS

If point configuration really does conform to variability of the objects along the external criterion variable, then we can embed an axis representing that dimension within the MDS space

Simple regression procedure for doing so . . .

Embedding External Criterion Variables

Assume an external variable, Y, is available:

> Each of the k objects in the MDS have scores on the external variable, $y_1, y_2, \ldots, y_k$.

Regress Y on the MDS coordinate axes (Dim_1, Dim_2, ..., Dim_k):

$$y_i = \alpha + \beta_1 \, \mathrm{Dim}_{1i} + \beta_2 \, \mathrm{Dim}_{2i} + \ldots + \beta_k \, \mathrm{Dim}_{ki} + e_i$$

If regression equation fits well (i.e., R^2 is large), then Y is consistent with the spatial configuration of objects

Inserting External Variable into MDS Space

Possible interpretation:

> The external variable, Y, is a substantive source of variability in the point locations within the MDS configuration.

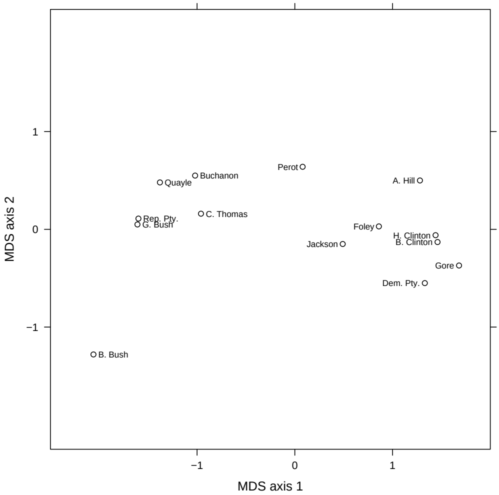
Can take ratio of regression coefficients to obtain slope of line representing external variable, relative to a pair of the MDS axes. For example:

$$\text{Slope}_Y = \hat{\beta}_2 / \hat{\beta}_1$$

Can locate a line with the preceding slope anywhere within the two-dimensional subspace of the MDS configuration; it is usually convenient to run the line through the origin of the space.

Example of Interpretation, 1992 Data

Two-dimensional point configuration obtained from nonmetric MDS of 1992 political figures (repeated)



Theoretical Predictions

Public perceptions of political figures are affected by two factors:

- Ideology
- Overall popularity

Operationalize the two variables:

- LC: A scale ranging from -100 to 100, with negative values indicating liberal positions, positive values indicating conservative positions
- AFF: A 0-100 scale, with larger values corresponding to greater popularity

Data Matrix with External Variables

D1:	D2:	AFFECT:	LC:	NAME:
-1.61	0.05	52	27	G. Bush
1.46	-0.13	56	-22	B. Clinton
0.08	0.64	45	0	Perot
-1.38	0.48	42	29	Quayle
1.68	-0.37	57	-18	Gore
1.28	0.50	49	-19	A. Hill
0.86	0.03	48	-9	Foley
-2.06	-1.28	67	12	B. Bush
1.44	-0.06	54	-17	H. Clinton
-0.96	0.16	45	15	C. Thomas
-1.02	0.55	42	19	Buchanon
0.49	-0.15	47	-16	Jackson
1.33	-0.55	59	-19	Dem. Pty.
-1.60	0.11	52	22	Rep. Pty.

OLS Estimates for Ideology, 1992 Data

Ideology equation:

$$\mathrm{LC}_i = 0.289 - 13.343\,\mathrm{Dim}_{1i} + 8.657\,\mathrm{Dim}_{2i} + e_i$$

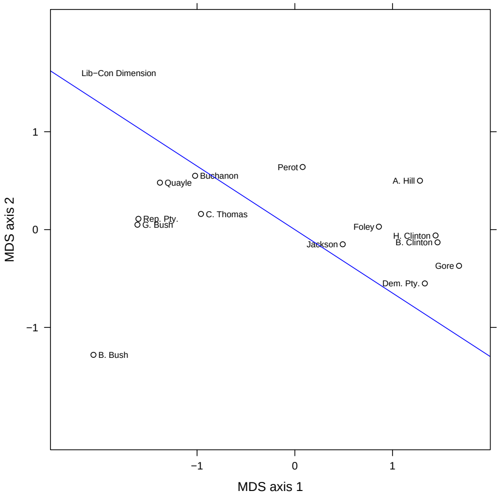
$R^2 = 0.940$

$$\mathrm{Slope}_{\mathrm{LC}} = \frac{8.657}{-13.343} = -0.649$$

Draw line representing ideology dimension into configuration

Inserting Ideology Dimension

Two-dimensional point configuration with
liberal-conservative dimension inserted



OLS Estimates for Popularity, 1992 Data

Popularity equation:

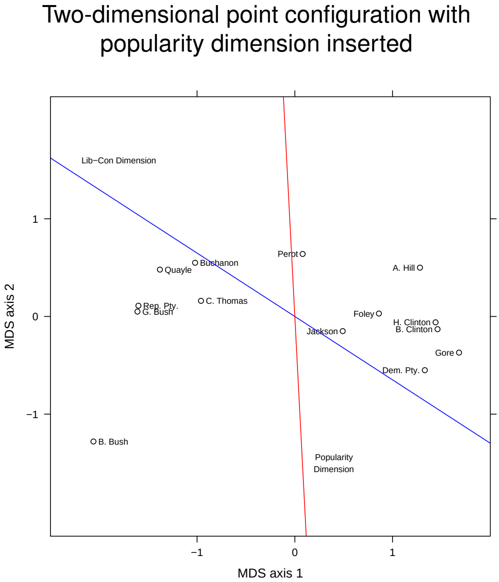
$$\mathrm{AFF}_i = 51.054 + 0.655\,\mathrm{Dim}_{1i} - 12.622\,\mathrm{Dim}_{2i} + e_i$$

$R^2 = 0.832$

$$\mathrm{Slope}_{\mathrm{AFF}} = \frac{-12.622}{0.655} = -19.270$$

Draw line representing popularity dimension into configuration

Adding Popularity Dimension



Finding Groups of Points in MDS Solution

Cluster analysis is a family of methods for creating taxonomies of objects

Divides objects into a set of (usually) mutually exclusive categories

Objects that are close to each other are placed into a common cluster; more distant objects are placed into different clusters.

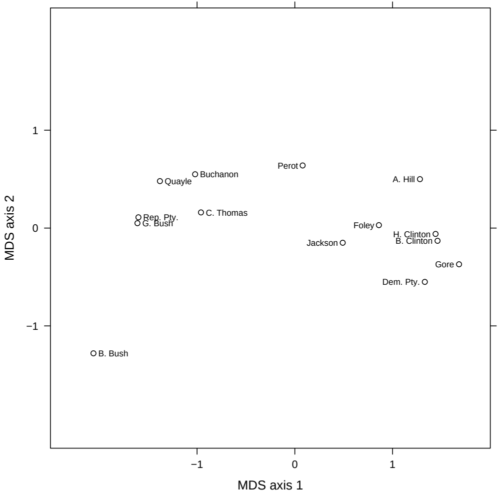
There are many varieties of cluster analysis

Hierarchical Cluster Analysis

- Begin with each object in a separate cluster
- Proceed through $k - 1$ steps, creating a new cluster on each step
  - On each step, join together the two closest clusters
  - The location of each new cluster is (usually) the mean location of the objects contained in the cluster
  - At the $k - 1$ step, all objects are joined in a common cluster
- Diagram called the “dendrogram” traces the steps of the clustering process.

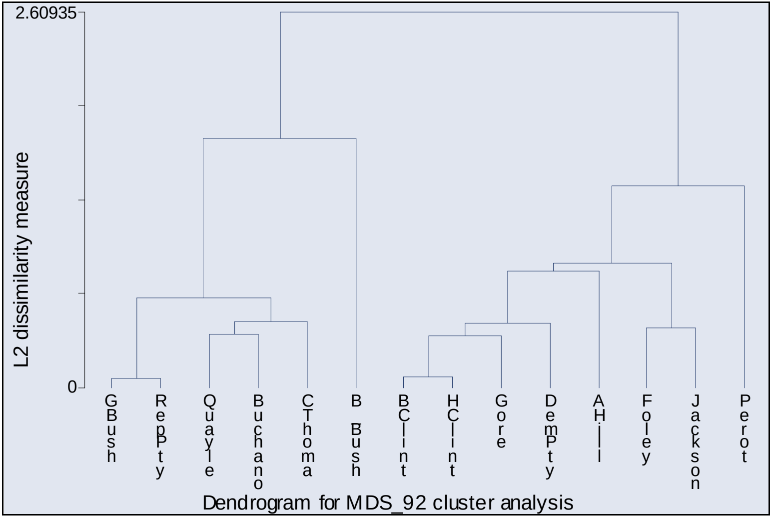
Example of Interpretation, 1992 Data

Two-dimensional point configuration obtained from nonmetric MDS of 1992 political figures (repeated . . . again)



Cluster Analysis of 1992 MDS Solution

Dendrogram from hierarchical cluster analysis:



More Caveats about Interpretation

Most MDS solutions are amenable to several different substantive solutions

> Objective strategies (i.e., external variables, cluster analysis) can show that scaling results are *consistent with* some particular interpretation
>
> Objective methods can never be used to find the single, “true” meaning of an MDS point configuration

This apparent uncertainty bothers some researchers . . .

MDS and Scientific Theories

In fact, the tentative nature of an MDS interpretation is no different from the general process of scientific theory construction and testing:

> The analyst demonstrates that the empirical data (MDS solution, in this case) are consistent with a specific story (i.e., a theory)
>
> The story becomes the accepted version of reality . . .
>
> Until someone demonstrates that another story (i.e., theory) provides a better description of the same data.

Data for MDS

Direct dissimilarity judgments about stimuli

Physical distances

Profile dissimilarities (sum-of-squared difference measures)

Confusion measures

Temporal change rates

LOS dissimilarities

Correlations (problematic)

Potential Problems with MDS

Too few stimuli in nonmetric MDS

Local minima

Degenerate solutions

Nonstandardized terminology

MDS usually regarded as data analytic procedure, not a statistical model