Multidimensional Scaling

Today . . .

Taking different sources into account: weighted multidimensional scaling.

More than one source can provide information about the similarity of objects in multidimensional scaling. Differences in the perceptual maps provided by multiple sources could be random, but might be linked to important systematic influences.

The multiple sources should be viewed broadly. They could be individuals, but could also be groups, different points in time, different settings, different methods, and so on.

The key is that more than one dissimilarity matrix is examined with the goal of understanding what is *common to all sources* and *what distinguishes them*.

The general goal is to find a common group space that best captures the perceptual map of the sources collectively, and, to identify how the dimensions of that group space are weighted by the individual sources to identify individual differences.

The individual weights can be correlated with source characteristics to help explain why different sources view the objects differently.

Object characteristics can be correlated with space coordinates to help explain why objects cluster the way they do and the nature of the dimensions uncovered.

In weighted MDS, we assume that all sources use the same number of dimensions when evaluating the objects, but they might weight the dimensions differently. The nature of these (configuration) weights defines the different forms of the analysis.

The individual weight matrix, **W**, transforms the coordinates of the group space, **G**, into the coordinates of the individual space, **S**: **S** = **GW**.

W can be an identity matrix (no individual differences), a diagonal matrix (INDSCAL), or a nondiagonal matrix (IDIOSCAL).

Data are from 13 surveys conducted between 1982 and 2011 that asked participants to rank U.S. presidents according to their overall quality. The rankings for presidents from Washington through Reagan are included here. The surveys were sponsored by a variety of entities: Wall Street Journal, C-SPAN, Siena College, The Times (British), Chicago Tribune, etc.

The sources ranged widely in their nature. Some were popular opinion polls, some were academics (political scientists, law professors), some were political journalists, some were a blend of various sources. Sample sizes varied widely.

The data can address a number of interesting questions:

- Are views about presidential leadership stable over time (nearly 30 years in this case) and source?
- Can the nature of presidential greatness be discerned from the common perceptual map?
- Can individual variation across the surveys be understood?

The data file contains columns of normalized ranks for the 13 different sources. These are converted into dissimilarity matrices.

	Name	Source_1	Source_2	Source_3	Source_4	Source_5	Source_6	Source_7	Source_8	Source_9	Source_10
Washington	Washington	0.050	0.00000	0.07317	0.00000	0.02439	0.02439	0.07143	0.05128	0.07895	0.07692
Adams	Adams	0.375	0.31579	0.26829	0.30769	0.29268	0.39024	0.38095	0.28205	0.23684	0.33333
Jefferson	Jefferson	0.150	0.07895	0.09756	0.07692	0.07317	0.14634	0.09524	0.07692	0.02632	0.05128
Madison	Madison	0.425	0.36842	0.19512	0.41026	0.34146	0.46341	0.11905	0.33333	0.21053	0.17949
Monroe	Monroe	0.325	0.39474	0.17073	0.38462	0.48780	0.31707	0.14286	0.30769	0.36842	0.25641
J_Q_Adams	J_Q_Adams	0.450	0.50000	0.39024	0.61538	0.36585	0.43902	0.42857	0.48718	0.42105	0.38462
Jackson	Jackson	0.300	0.13158	0.29268	0.23077	0.31707	0.29268	0.30952	0.20513	0.31579	0.20513
Van_Buren	Van_Buren	0.725	0.57895	0.56098	0.66667	0.95122	0.73171	0.52381	0.66667	0.52632	0.51282
Tyler	Tyler	0.875	0.86842	0.87805	0.87179	0.73171	0.82927	0.85714	0.92308	0.86842	0.82051
Polk	Polk	0.275	0.23684	0.24390	0.20513	0.19512	0.26829	0.26190	0.38462	0.28947	0.30769
Taylor	Taylor	0.675	0.78947	0.80488	0.82051	0.65854	0.68293	0.76190	0.82051	0.73684	0.84615
Fillmore	Fillmore	0.850	0.89474	0.90244	0.89744	0.78049	0.87805	0.88095	0.87179	0.81579	0.79487
Pierce	Pierce	0.950	0.94737	0.92683	0.94872	0.97561	0.95122	0.9[illegible] 0.88095	0.97436	0.89474	0.89744
Buchanan	Buchanan	1.000	1.00000	0.97561	1.00000	1.00000	1.00000	0.97619	1.00000	0.94737	0.94872
Lincoln	Lincoln	0.000	0.02632	0.02439	0.02564	0.00000	0.00000	0.04762	0.02564	0.05263	0.02564
A_Johnson	A_Johnson	0.975	0.92105	1.00000	0.92308	0.56098	0.97561	1.00000	0.89744	0.97368	0.97436
Grant	Grant	0.800	0.81579	0.82927	0.71795	0.41463	0.53659	0.59524	0.71795	0.92105	0.92308
Hayes	Hayes	0.625	0.55263	0.63415	0.58974	0.63415	0.78049	0.71429	0.74359	0.55263	0.56410
Arthur	Arthur	0.775	0.65789	0.70732	0.64103	0.51220	0.75610	0.57143	0.79487	0.60526	0.64103
Cleveland	Cleveland	0.400	0.28947	0.46341	0.28205	0.43902	0.48780	0.45238	0.51282	0.44737	0.41026
B_Harrison	B_Harrison	0.750	0.68421	0.75610	0.74359	0.70732	0.70732	0.78571	0.84615	0.78947	0.71795
McKinley	McKinley	0.350	0.34211	0.43902	0.33333	0.39024	0.36585	0.47619	0.41026	0.47368	0.46154
T_Roosevelt	T_Roosevelt	0.075	0.10526	0.04878	0.10256	0.09756	0.07317	0.02381	0.10256	0.10526	0.10256

```
Presidents_Dist_1 <- dist(Presidents[, 2], method = "euclidean", diag = TRUE)
Presidents_Dist_2 <- dist(Presidents[, 3], method = "euclidean", diag = TRUE)
Presidents_Dist_3 <- dist(Presidents[, 4], method = "euclidean", diag = TRUE)
Presidents_Dist_4 <- dist(Presidents[, 5], method = "euclidean", diag = TRUE)
Presidents_Dist_5 <- dist(Presidents[, 6], method = "euclidean", diag = TRUE)
```

```
Presidents_Dist <- list(Presidents_Dist_1, Presidents_Dist_2, Presidents_Dist_3,
    Presidents_Dist_4, Presidents_Dist_5, Presidents_Dist_6, Presidents_Dist_7,
    Presidents_Dist_8, Presidents_Dist_9, Presidents_Dist_10, Presidents_Dist_11,
    Presidents_Dist_12, Presidents_Dist_13)
```

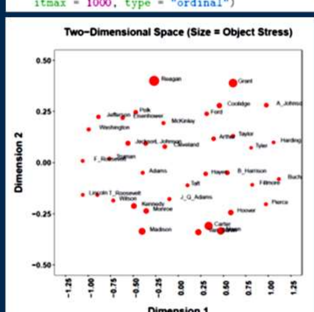
A list of distance matrices is then the input to the MDS function.

There is additional information about the sources that could prove useful in interpreting the MDS outcomes. This includes the nature of the sources (journalists, academics, the public), the year of the survey, and whether a Democrat or Republican was in the White House at the time of the survey.

Likewise, information about the objects (Presidents) could be important (age, height, etc.). Second, other surveys, in particular the Siena College President Survey, could be revealing. The Siena survey used here comes from 2010, includes historians and presidential experts, and asked the participants to rank the presidents on a number of dimensions.

- Background
- Imagination
- Integrity
- Intelligence
- Luck
- Willing to take risks
- Avoid crucial mistakes
- Court appointments
- Domestic accomplishments
- Executive appointments
- Foreign policy accomplishments
- Handling of U.S. economy
- Party leadership
- Relationship with Congress
- Ability to compromise
- Communication ability
- Executive ability
- Leadership ability
- Overall ability

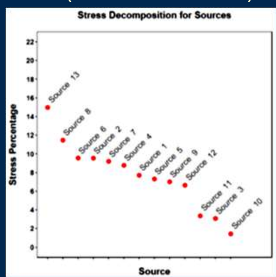
```
identity_fit <- smacofIndDiff(Presidents_Dist, ndim = 2, constraint = "identity",
    itmax = 1000, type = "ordinal")
```

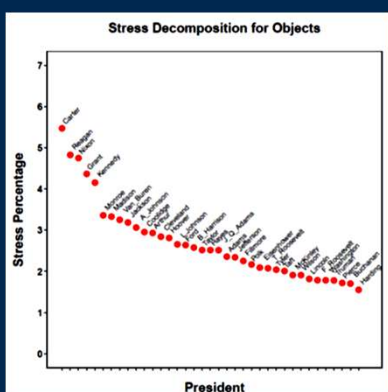


Initial examination assuming equal source weighting (W = I), gives a sense of the number of dimensions that may be necessary, the overall configuration of the objects, and provides a basis for determining if individual weighting will have any advantages.

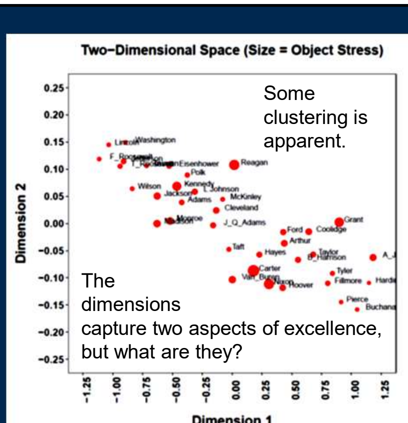
```
indscal_fit <- smacofIndDiff(Presidents_Dist, ndim = 2, constraint = "indscal",
    itmax = 1000, type = "ordinal")
```

Allowing individual differences reduces stress by 15% (from .192 to .162). Stress varies quite a bit by source. Source 13 was a public opinion poll conducted in 1996 (a Democrat was in the White House). Source 10 was a poll of academics conducted in 1990 (a Republican was in the White House).



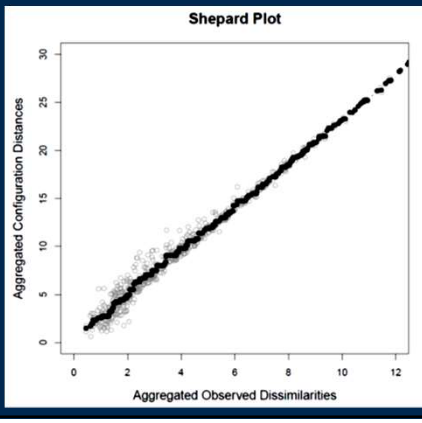


Stress also varies by object. Carter, Reagan, Nixon, Grant, and Kennedy were harder to fit (less agreement) than other presidents.

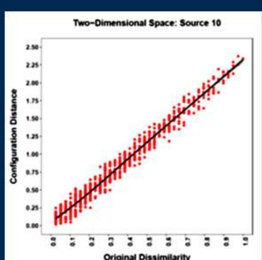


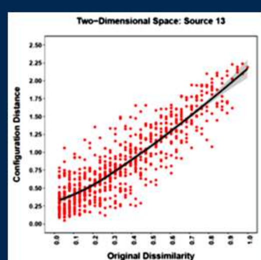
Some clustering is apparent.

The dimensions capture two aspects of excellence, but what are they?

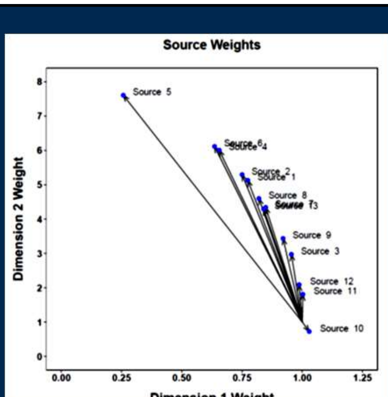


This solution provides a pretty good fit to the data.



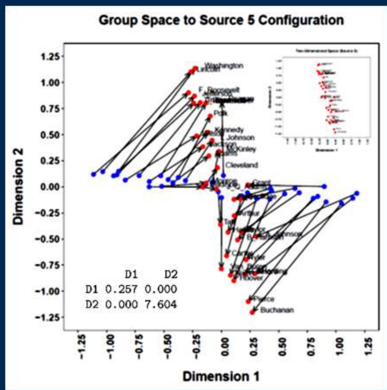


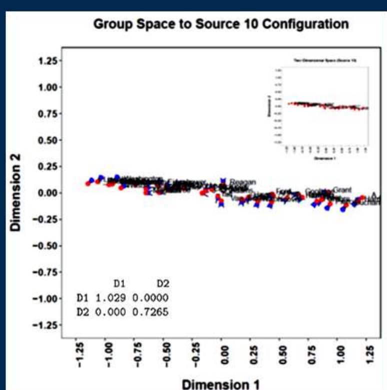
As was suggested by the stress values, the fit is better for some sources than for others. The restriction here is that the weight matrices are diagonal. That may be too restrictive.



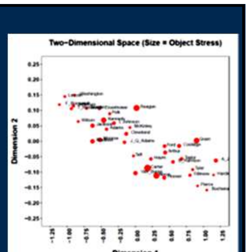
The weights indicate how much the group configuration needs to be stretched or shrunk to achieve a particular source configuration.

Weights of 1 indicate no adjustment is necessary.





	D1	D2
Height	-0.22	0.31
Age	0.01	0.03

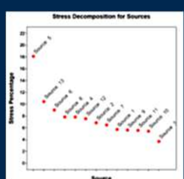


Physical stature is slightly related to greatness. Academics emphasized the first dimension over the second; journalists placed more emphasis on the second. Earlier surveys did too. Academic sources were easier to fit; earlier surveys were too.

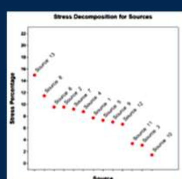
	Year	Academic	News	Dems	D1	D2	Stress
Year	1.00	-0.44	0.39	0.26	-0.56	0.63	0.38
Academic	-0.44	1.00	-0.62	-0.35	0.58	-0.65	-0.65
News	0.39	-0.62	1.00	-0.02	-0.76	0.78	0.20
Dems	0.26	-0.35	-0.02	1.00	0.17	0.02	0.49
D1	-0.56	0.58	-0.76	0.17	1.00	-0.92	-0.35
D2	0.63	-0.65	0.78	0.02	-0.92	1.00	0.58
Stress	0.38	-0.65	0.20	0.49	-0.35	0.58	1.00

We can relax the diagonal constraint in the weight matrices, which essentially allows individual source rotation of the dimensions (IDIOSCAL).

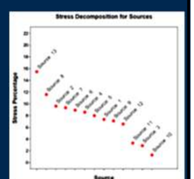
The overall stress for this approach is .160, a trivial improvement over the INDSCAL solution (.162). Ordinarily we would stop here. But, let's take a look at the additional information that IDIOSCAL provides.



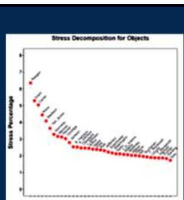
Identity Constraint



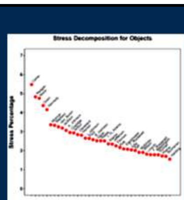
Diagonal Constraint



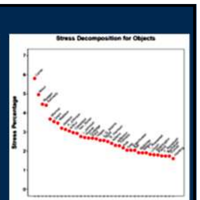
No Constraint



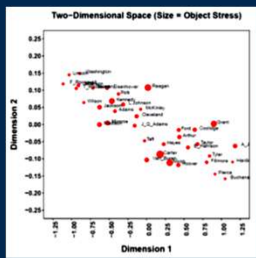
Identity Constraint

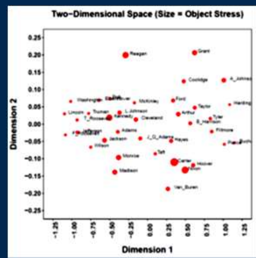


Diagonal Constraint



No Constraint

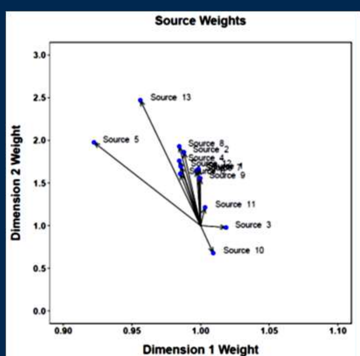




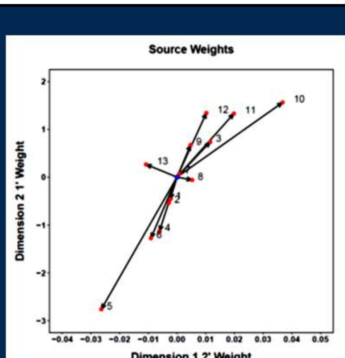
Diagonal Constraint

No Constraint

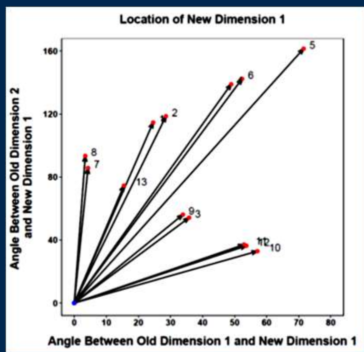
The group spaces look different because in the no-constraint model an additional rotation is allowed to achieve the individual solutions.



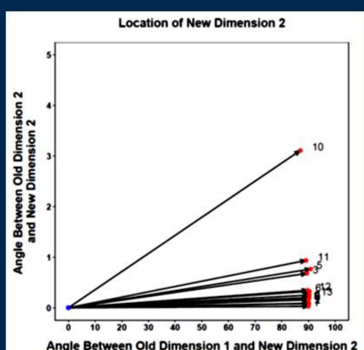
As in INDSCAL, the diagonal weights represent the degree to which the original axes are stretched or shrunk.



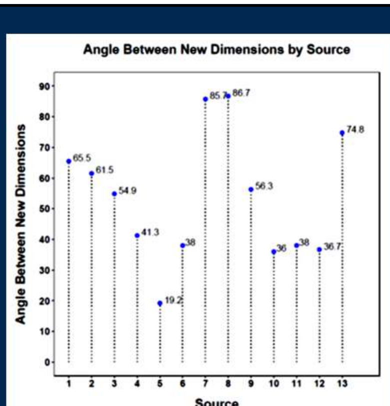
But, because the off-diagonal weights are not constrained to 0, the new coordinates represent linear combinations of the group space coordinates. The ordinate here is the contribution of Dimension 2 from group space to the new location of Dimension 1.

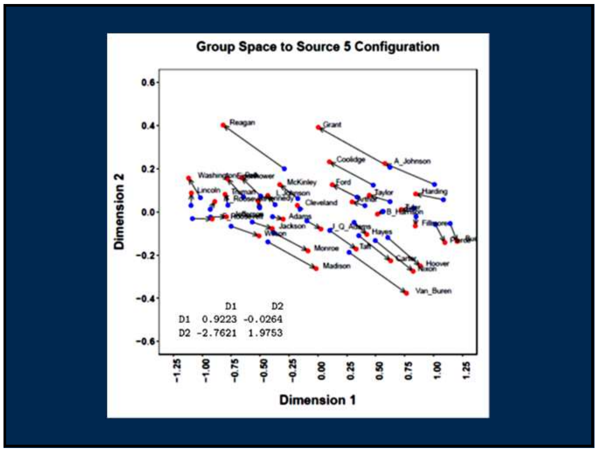


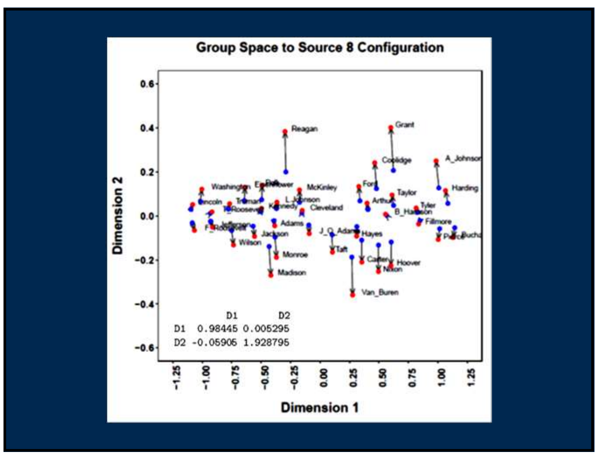
If the columns of the weight matrix are normalized, they become cosines of the angles between the group space coordinate system and the source coordinates.

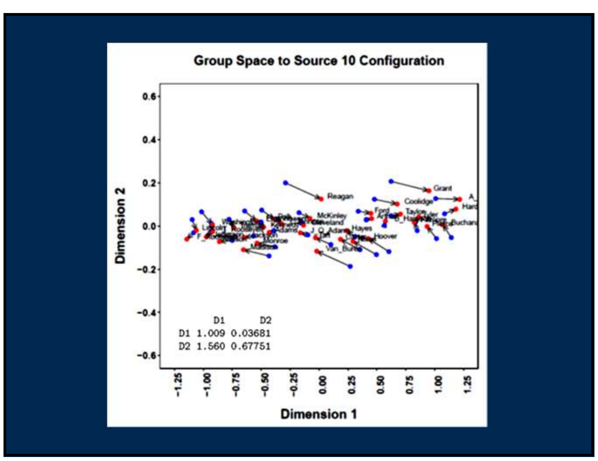


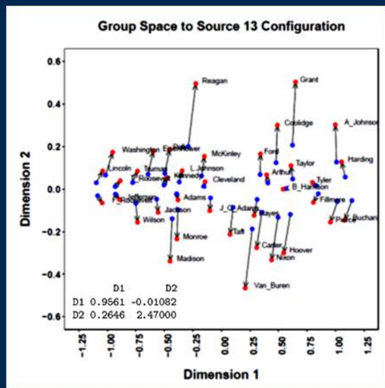
If the columns of the weight matrix are normalized, they become cosines of the angles between the group space coordinate system and the source coordinates.




Group Space to Source 5 Configuration
Dimension 2
Dimension 1
D1 D2
D1 0.9223 -0.0264
D2 -2.7621 1.9753


Group Space to Source 8 Configuration
Dimension 2
Dimension 1
D1 D2
D1 0.98445 0.005295
D2 -0.05906 1.928795


Group Space to Source 10 Configuration
Dimension 2
Dimension 1
D1 D2
D1 1.009 0.03681
D2 1.560 0.67751



Academic sources (compared to others) were more likely to expand Dimension 1 and contract Dimension 2, with no impact on the angle between the two dimensions. News sources had a decreased angle between the two dimensions. In other words, the multidimensional space for news sources was simpler than for other sources. The multidimensional space was more complex in surveys conducted while a Democrat was in the White House. The views of presidents have become a little more complex over the years.

	Year	Academic	News	Dems
Year	1.00	-0.44	0.39	0.26
Academic	-0.44	1.00	-0.62	-0.35
News	0.39	-0.62	1.00	-0.02
Dems	0.26	-0.35	-0.02	1.00
D_11	-0.33	0.62	-0.38	0.01
D_22	0.32	-0.73	0.28	0.41
D_12	-0.50	0.65	-0.62	-0.09
D_21	-0.63	0.60	-0.81	0.11
Stress	0.37	-0.64	0.17	0.47
Angle	0.19	0.03	-0.33	0.47

More sophisticated follow-up analyses are possible. For example, a multiple regression shows that the coordinates are not further predicted by the quadratic component of age, the height by age interaction, or the height by age^2 interaction. Principal components analysis of the Siena survey indicates it is dominated by a single component.

By now it should go without saying . . . cross-validate.

Next time . . .

Cluster analysis