Multidimensional Scaling in R: SMACOF

Patrick Mair
Department of Statistics and Mathematics
WU Vienna University of Economics and Business

Jan de Leeuw
Department of Statistics
University of California, Los Angeles (UCLA)

Eva Lienbacher
Institute for Retailing and Marketing
WU Vienna University of Economics and Business

Content

- Basics of MDS and SMACOF
- SMACOF implementation in R
- Symmetric SMACOF
- Spherical SMACOF
- Rectangular SMACOF
- 3-Way SMACOF

Multidimensional Scaling (MDS)

MDS: Family of data-analytic methods which represent distances between objects in a low-dimensional space.

- Torgerson (1952): Classical scaling approach introduced to Psychometrics.
- Shepard (1962): Non-metric MDS.
- Kruskal (1964): Stress, reduction of dimensions.
- de Leeuw (1977, 1988): SMACOF - Scaling by MAjorizing a COmplicated Function.

MDS Workflow

We have the following steps of analysis:

1. Input structure: Dissimilarity (distance) matrix.

2. Computation: Optimize target function (stress).

3. Output: Configurations in low-dimensional space.

4. Visualization: Configuration plot, goodness-of-fit plots.

Distance Matrix and Computation

```
R> islanddist
           Crete Euboea Lesbos Rhodes Kefalonia  Chios  Corfu Lemnos ...
Euboea    379.60
Lesbos    455.90 215.70
Rhodes    292.00 438.90 364.00
Kefalonia 521.90 307.70 518.70 702.90
Chios     367.30 177.00  88.62 304.00    483.70
Corfu     679.20 389.80 569.50 822.40    175.30 560.90
Lemnos    524.00 190.80 124.90 481.10    453.10 180.60 471.80
Samos     325.70 259.20 162.40 205.70    555.70  99.65 648.90 275.60
Naxos     210.00 206.70 246.30 246.30    457.30 157.50 578.30 318.50 ...
Zakynthos 477.60 295.30 510.10 670.20     49.53 467.40 224.70 455.70 ...
Thasos    607.50 247.30 219.20 578.00    451.20 276.80 433.20  96.93 ...

R> res.island <- smacofSym(islanddist)
```


Thasos
Lemnos
Lesbos
Chios
Samos
Eubosa
Corfu
Kefalonia
Zakynthos
Naxos
Rhodes
Crete

R Implementation: SMACOF package

The R Project for Statistical Computing

- R is an open source software environment for statistical computing and graphics
- `http://www.R-project.org`
- 1938 packages available

The smacof package

- CRAN: `http://CRAN.R-project.org`
- PsychoR: `http://r-forge.r-project.org/projects/psychor`.

Symmetric SMACOF

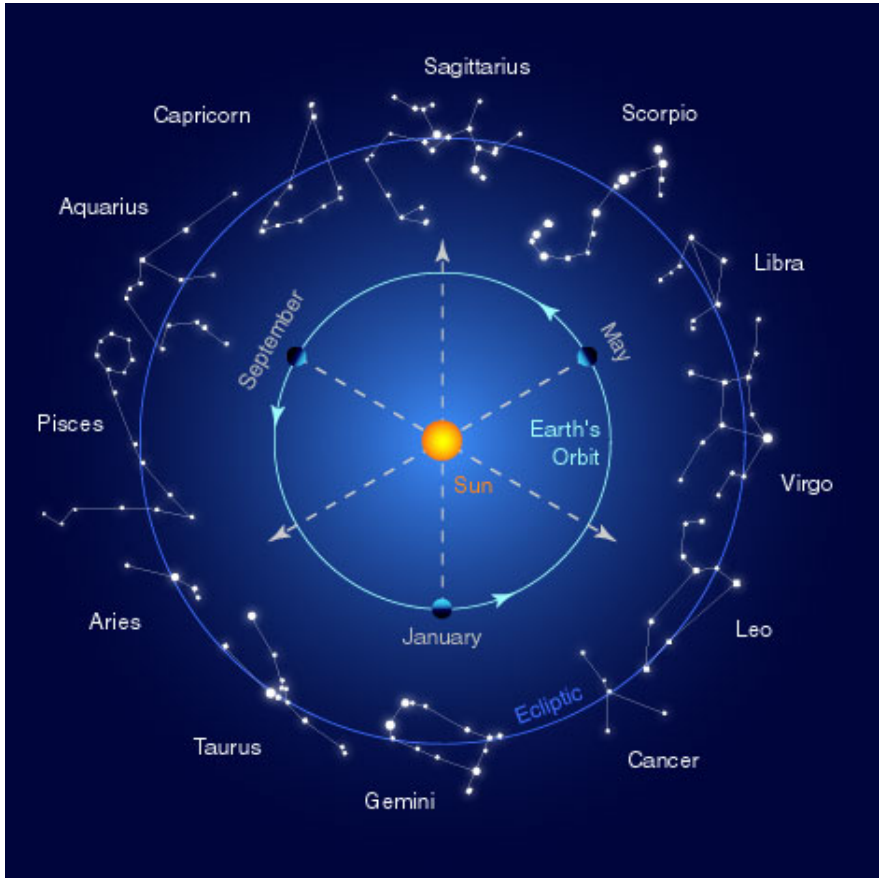
- Distance matrix Δ of dimension $n \times n$ with elements δ_{ij}.
- Problem to solve: Locate points (*configurations*) in a p-dimensional space such that the distances $d_{ij}(X)$ between the points approximate δ_{ij}.
- Configuration distances:

$$d_{ij}(X) = \sqrt{\sum_{s=1}^{p} (x_{is} - x_{js})^2}$$

- Minimize *stress* (*Majorization*; de Leeuw, 1977):

$$\sigma(X) = \sum_{i<j} w_{ij}(\delta_{ij} - d_{ij}(X))^2 \rightarrow \text{min!}$$

Example 1: Signs of the Zodiac



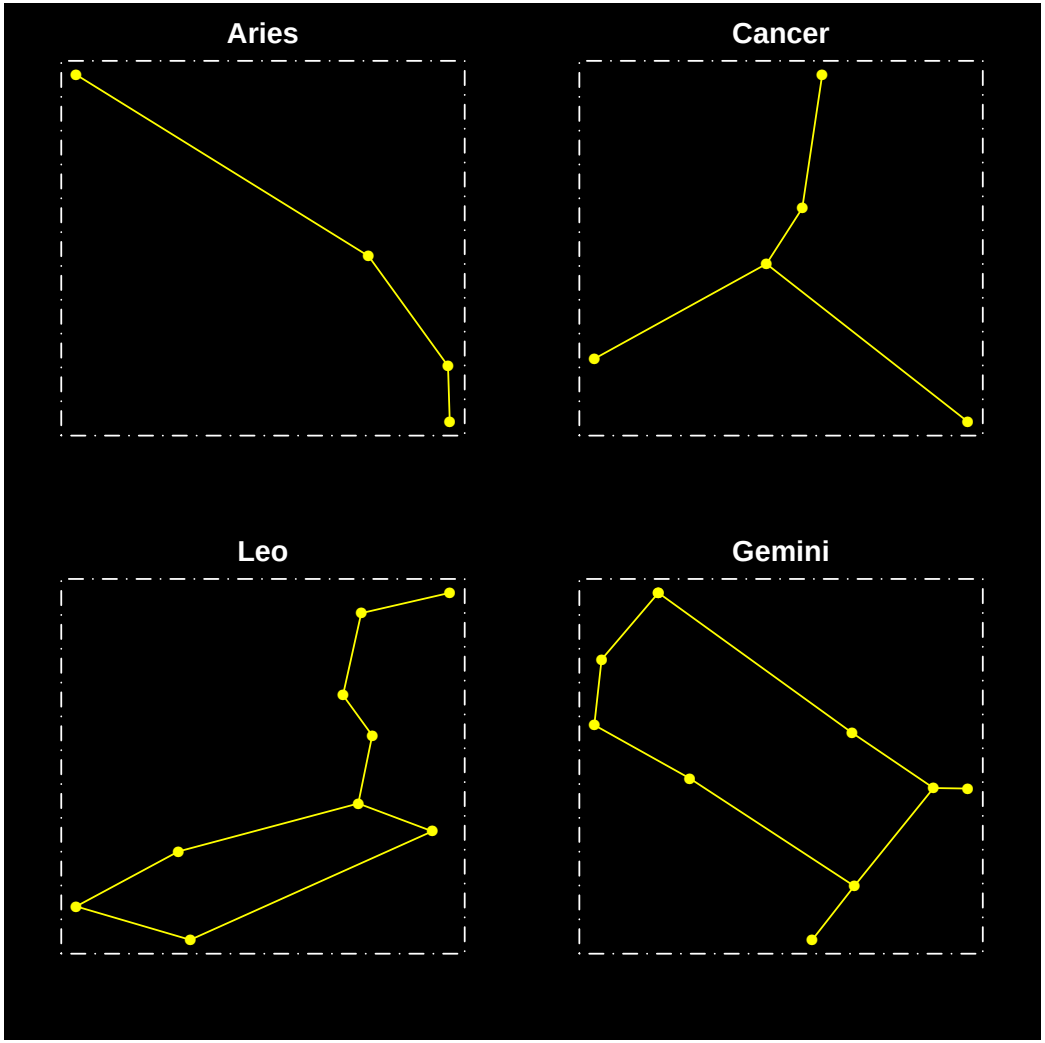
Example 1: Signs of the Zodiac

(Thanks to Paul Eigenthaler from the Institute of Astronomy, University of Vienna, for providing the distances.)

```
R> resall <- smacofSym(stardist, ndim = 2)
R> resall
Call: smacofSym(delta = stardist, ndim = 2)

Model: Symmetric SMACOF
Number of objects: 120

Metric stress: 7.817851e-05
Number of iterations: 818
```


Aries
Cancer
Leo
Gemini

Spherical SMACOF

Restrictions on the configurations (*weakly constrained MDS*).

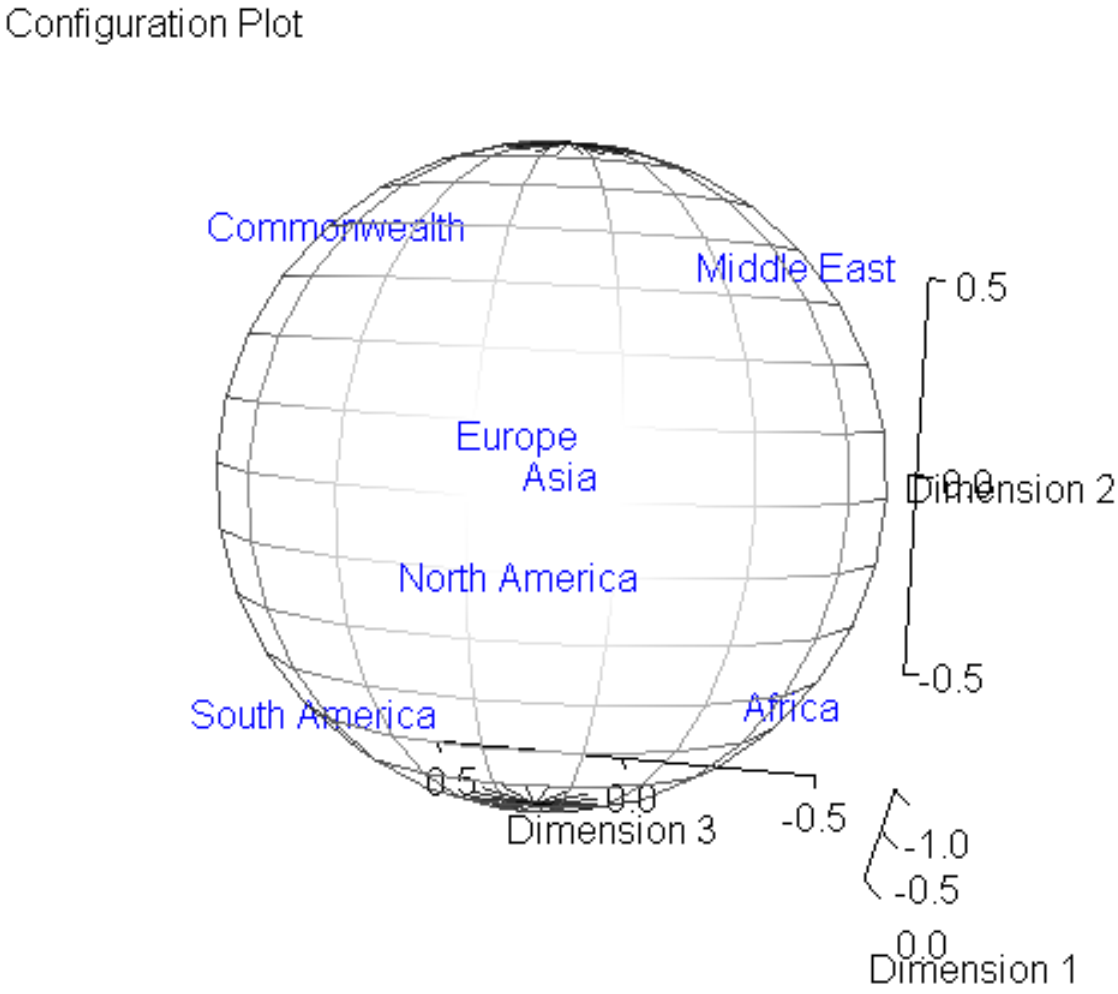
$$\mathbf{x}_i'\mathbf{\Lambda}\mathbf{x}_i + 2\mathbf{x}_i'\beta + \gamma = 0$$

- $\mathbb{R}^2$: circle, ellipse, hyperbola, parabola.
- $\mathbb{R}^3$: sphere, ellipsoid, hyperboloid, paraboloid, cylinder.
- Optimization: Primal and dual methods available.

Example 2: Trading Volume

```
R> tradedist
               North America South America Europe Commonwealth Africa ...
South America          867.3
Europe                 362.1         963.3
Commonwealth          1113.3        1136.6  672.8
Africa                1030.1        1121.0  834.1        1141.5
Middle East           1015.3        1135.8  888.1        1128.3 1111.3
Asia                    40.8         976.8    1.0        1009.9  977.0 ...

R> res.trade <- smacofSphere.dual(tradedist, ndim = 3, itmax = 2000)
R> plot3d(res.trade, sphere = TRUE)
```


Configuration Plot
Commonwealth
Middle East
Europe
Asia
North America
South America
Africa
0.5
0.0
-0.5
Dimension 2
0.5
0.0
-0.5
Dimension 3
-1.0
-0.5
0.0
Dimension 1

Rectangular SMACOF (Unfolding)

Rectangular $n_1 \times n_2$ preference matrix Δ.

- Stress becomes

$$\sigma(X) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} w_{ij}(\delta_{ij} - d_{ij}(X_1, X_2))^2 \rightarrow \text{min!}$$

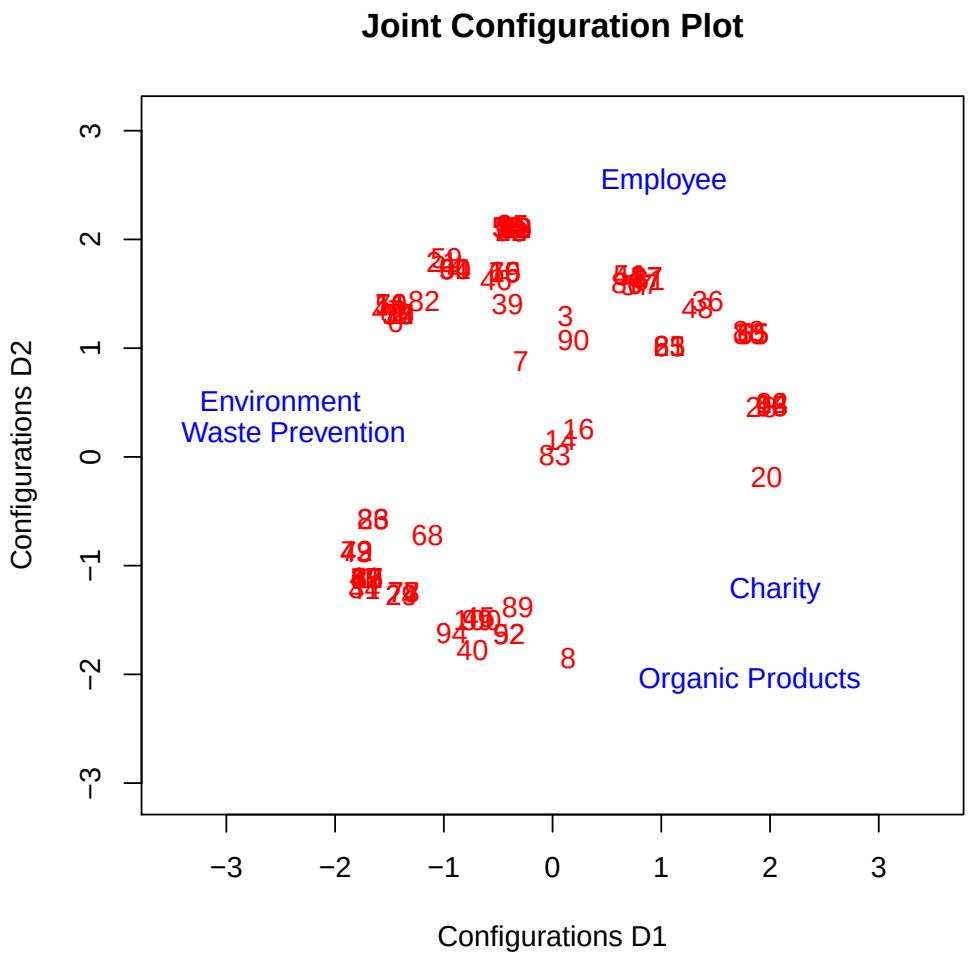
- Judge $n_1 \times p$ configuration matrix
- Object $n_2 \times p$ configuration matrix

Example 3: Company Rating

```
R> head(csr)
  Environment Waste Prevention Organic Products Charity Employee
1           1                2                4       3        5
2           2                1                5       4        3
3           1                5                3       4        2
4           3                1                5       4        2
5           2                3                4       5        1
6           2                1                4       5        3
```

```
R> res.csr <- smacofRect(csr)
```

```
R> plot(res.csr, xlim = c(-3, 3), joint = TRUE, asp = 1)
```


Joint Configuration Plot
Employee
Environment
Waste Prevention
Charity
Organic Products
Configurations D1
Configurations D2
−3
−2
−1
0
1
2
3

3-Way SMACOF

SMACOF for individual differences:

- $k = 1, \ldots, K$ separate symmetric distance matrices.
- Data cube, or, in R: List.
- Classical approach: INDSCAL (Carrol & Chang, 1970).

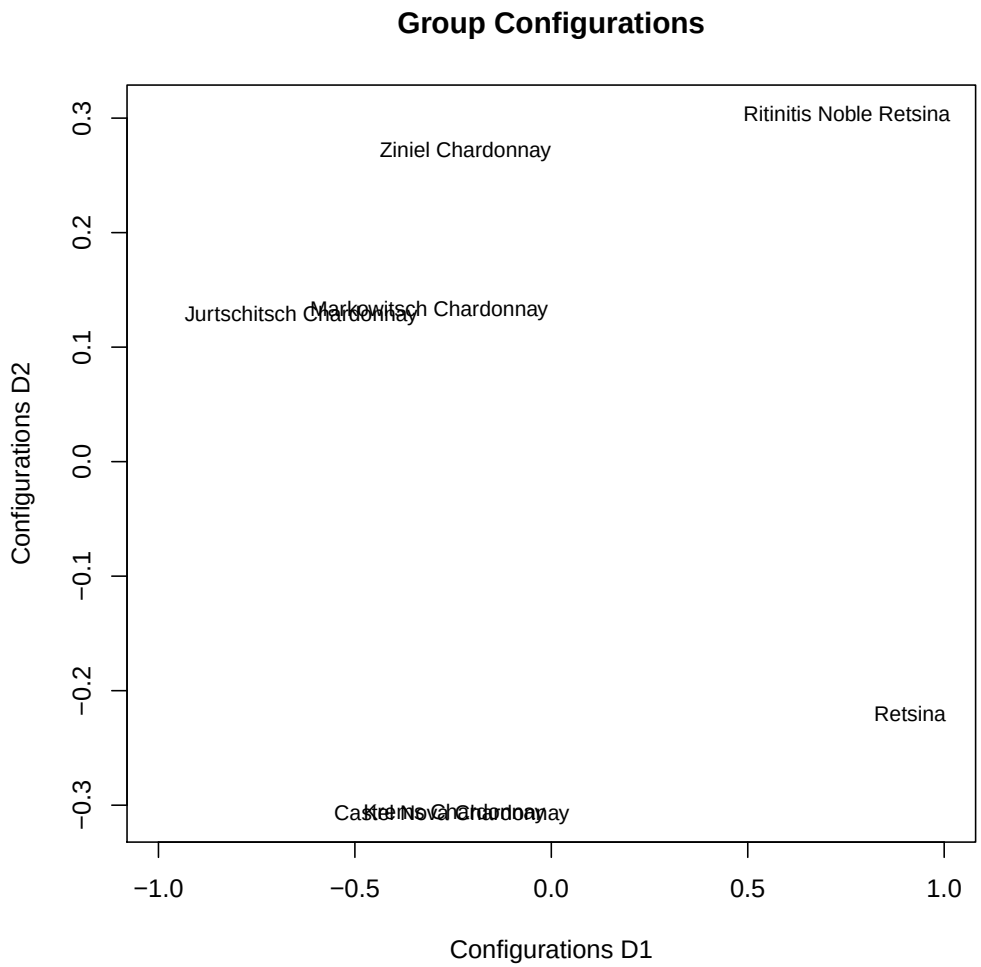
Example 4: Wine Tasting

- Ziniel Chardonnay
- Markowitsch Chardonnay
- Krems Chardonnay
- Castel Nova Chardonnay
- Ritinitis Noble Retsina
- Retsina

Criteria: color, smell, taste, fun, overall impression

```
R> reswine <- smacofIndDiff(winedat, metric = FALSE)
```

```
R> plot(reswine, xlim = c(-1, 1))
```


Group Configurations
Ritinitis Noble Retsina
Ziniel Chardonnay
Jurtschitsch Chardonnay
Markowitsch Chardonnay
Retsina
Configurations D2
Configurations D1
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−1.0
−0.5
0.0
0.5
1.0

Wine Tasting: Descriptives

	Price	Alcohol	Mean Rating
Jurtschitsch Chardonnay	14.99	13.00	2.00
Ziniel Chardonnay	7.00	12.00	2.60
Markowitsch Chardonnay	9.99	12.50	2.60
Ritinitis Noble Retsina	9.99	12.00	4.30
Retsina	2.99	11.50	4.60
Krems Chardonnay	5.99	12.50	2.70
Castel Nova Chardonnay	1.99	12.00	2.80

Additional Models and Options

Each SMACOF variant is implemented in a metric and non-metric way.

- If observed data are ordinal $\rightarrow$ distances will be ordinal as well $\rightarrow$ non-metric MDS.
- Various distance mesures (e.g. Euclidean, Jaccard, Minkowski, etc.), proxy package in R.
- Estimation: Additional isotone regression step (PAVA).

Additional models and options

Decomposition of the configurations (de Leeuw & Heiser, 1980):

- Linear decomposition $X = ZC$.
- SMACOF function `smacofConstraint()`.

More 3-way options:

- IDIOSCAL (Carrol & Wish, 1974)
- Various other decompositions of the weight matrix.

Goodness-of-fit examination: Shepard diagrams, Stress plots, Residual plots.

References

- de Leeuw, J. & Mair, P. (2009). Multidimensional Scaling using Majorization: SMACOF in R. Journal of Statistical Software, 31(3), p. 1-30. URL: `http://www.jstatsoft.org`

- Borg, I., & Groenen, P. J. F. (2005). Modern Multidimensional Scaling. New York: Springer.
- Cox, T. F., & Cox, M. A. A. (2001). Multidimensional Scaling (2nd edition). Boca Raton, FL: Chapman & Hall/CRC.

Links and Contact

PsychoR project:

- Website: `http://r-forge.r-project.org/projects/psychor`
- Next PsychoR topics: isotone optimization, exponential geometric models, homals with splines.

Patrick Mair
Department of Statistics and Mathematics
WU Vienna University of Economics
Augasse 2-6
1090 Vienna

Email: `patrick.mair@wu.ac.at`
Website: `http://statmath.wu.ac.at/~mair`