# Multivariate Verfahren

## 7.4. Multidimensional Scaling (MDS)

Hannah Kümpel

based on slides by Sabine Hoffmann

Institut für Statistik, LMU München

Sommersemester 2024

# Introducion

- **What is MDS?** A technique for visualizing the similarity or dissimilarity of data points in a low-dimensional space.

- **Key Features:**

  *Dimensionality Reduction:* Transforms complex, high-dimensional data into a simpler, visually interpretable form (2D or 3D).

  *Distance Preservation:* Aims to retain the original distances between data points as accurately as possible.

- **Why Use MDS?** for data visualization and exploratory data analysis, It helps in understanding patterns and relationships in data.

# The basic idea of MDS

- **Given**
  1. a set of $n$ objects
  2. the distances/dissimilarities $d_{ij}$ between them

- **We want to find** points in a lower dimensional space whose distances $\delta_{ij}$ are as close as possible to the $d_{ij}$.

- If all distances of the original objects are quantitative in nature, this in endeavour is more straightforward, since their distances are easily measured by **a metric** such as the euclidean distance $\longrightarrow$ *metric MDS*

- In other cases (quantitative distances), one needs to resort to more general dissimilarity measures to get the $d_{ij}$s $\longrightarrow$ *non-metric MDS*

# Metric MDS

# Metric MDS

The metric MDS goes back to Torgerson (1952, 1958) and can be divided into two models.

- distance model The objects $a_1, \ldots, a_n$ are transformed into distances $d_{i\ell},\ i, \ell = 1, \ldots, n$.

- spatial model The objects are represented by $n$ points $\mathbf{y}_1, \ldots, \mathbf{y}_n$ in $r$-dimensional space in such a way that the metric distances $d_p(i, \ell) = d_p(\mathbf{y}_i, \mathbf{y}_\ell)$ of the objects approximate the distances $d_{i\ell}$ specified by the distance model as closely as possible.

# Formal formulation of the problem

- Let $\mathbf{D} = (d_{i\ell})$ denote the matrix of original element-wise distances $d_{i\ell}$ and $\mathbf{\Delta} = (\delta_p(\mathbf{y}_i, \mathbf{y}_\ell))$ denote the corresponding matrix of element-wise distances for a lower-dimensional representation $\mathbf{Y} = (\mathbf{y}_1^\top, \ldots, , \mathbf{y}_n^\top)^\top \in \mathbf{R}^{nxr}$.

- We are looking for representation $\mathbf{Y}$ that minimizes the *cost or STRESS function*

$$\sum_{i \neq j} (d_{ij} - \delta_{ij})^2 \ .$$

- Clearly, this issue is **not unique**, because we could shift all points by a constant and obtain the same difference.

$\longrightarrow$ It can be is useful to assume centered points around the origin.

# Classical metric MDS (cMDS)

- The starting point is the matrix of squared Euclidean distances with elements

$$d_2(i,\ell) = d_2(\mathbf{y}_i, \mathbf{y}_\ell) = (\mathbf{y}_i - \mathbf{y}_\ell)^\top (\mathbf{y}_i - \mathbf{y}_\ell), \quad i,\ell = 1,\ldots,n\,.$$

- Instead of finding $\mathbf{Y}$, we can focus on finding the following matrix

$$B = \boldsymbol{Y}\boldsymbol{Y}^T$$

- for which it holds that

$$d_{i\ell}^2 \quad = \quad b_{ii} + b_{\ell\ell} - 2b_{i\ell}\,.$$

# Centering the distances first

- If we want to assume centering around the origin, which is the typical approach, we need to start with centering our distance matrix!

- For that, we use a centering matrix:

$$\mathbf{H} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top.$$

*Note:* we could replace $\frac{1}{n}$ with weights $w_i$, $\sum_{i=1}^n w_i = 1$ indicating the importance of each row of distances. Then, we would have $H = \mathbf{I}_n - \mathbf{1}_n(w_1, \ldots, w_n)^\top$.

- Then, we continue with

$$B = -\frac{1}{2}\boldsymbol{H}D\boldsymbol{H}.$$

# Eigenvalue decomposition in MDS

Either way, we consider the eigenvalue decomposition of the matrix $\mathbf{B}$:

$$\mathbf{B} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^\top,$$

where

- $\boldsymbol{P}$ is the matrix of the orthonormalized eigenvectors of $\mathbf{B}$ and
- $\mathbf{\Lambda}$ is the diagonal matrix of the eigenvalues of $\mathbf{B}$ ordered by size.

# Calculating the new representation in MDS

If one defines the eigenvectors for the $r$ positive eigenvalues of $\mathbf{B}$ $\mathbf{y}_1 = \sqrt{\lambda_1}\,\mathbf{p}_1, \ldots, \mathbf{y}_r = \sqrt{\lambda_r}\,\mathbf{p}_r$ and the matrix

$$\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_r)\,,$$

it holds that:

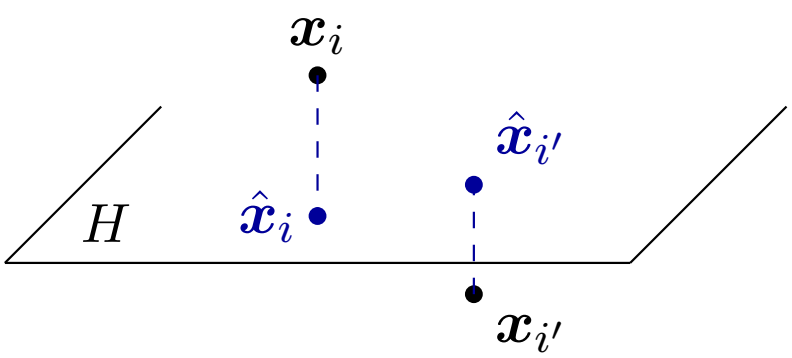
$$\mathbf{B} = \mathbf{P}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{P} = \mathbf{Y}\mathbf{Y}^\top\,.$$

$\rightarrow$ $\mathbf{Y} = (\mathbf{y}_1^\top, \ldots, , \mathbf{y}_n^\top)^\top$ is a representation of the objects and the lines of $\mathbf{Y}$ correspond to the coordinates.

**Note**

Classical MDS (i.e. MDS using the euclidean distance) yields the same results as PCA, see also Applied multivariate statistics. *However, MDS can also meaningfully be applied to distance matrices not generated under Euclidean distance measure where this no longer holds.*

# Equivalence of cMDS and PCA

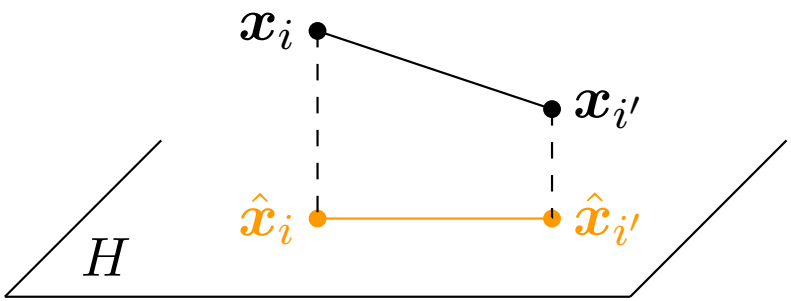



- **PCA:** Projection of the observations onto a subspace so that the maximum variance is retained

- Maximize

$$\boldsymbol{a}_p^\top \boldsymbol{S} \boldsymbol{a}_p\,, \quad p = 1, \ldots, m\,.$$

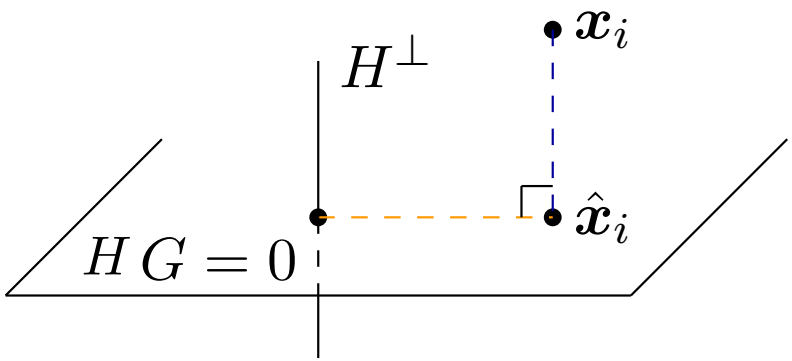
# Equivalence of cMDS and PCA




- **MDS:** Find a subspace so that the original distances are preserved by the projection if possible $d\left(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_{i'}\right)$
- Minimize

$$\sum_{i \neq i'} \left(d\left(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_{i'}\right) - d\left(\boldsymbol{x}_i, \boldsymbol{x}_{i'}\right)\right)^2$$

# Equivalence of cMDS and PCA




- Explained variance (by H):

$$I(H) = \frac{1}{n}\sum_{i=1}^{n} \|\hat{\boldsymbol{x}}_i\|^2$$

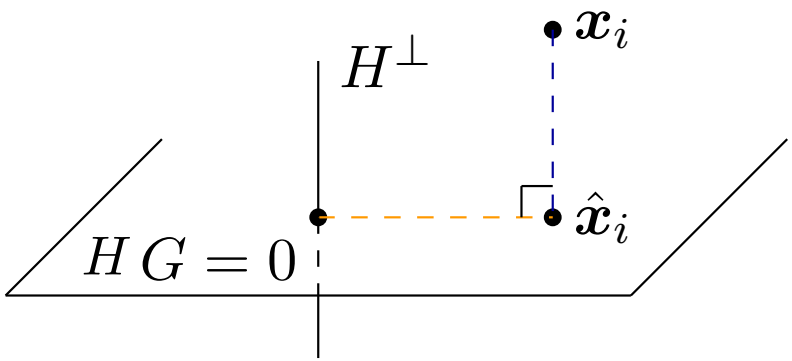
- Residual variance:

$$I(H^\perp) = \frac{1}{n}\sum_{i=1}^{n} \|\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i\|^2$$

- Total variance:

$$I_G = I_0 = \frac{1}{n}\sum_{i=1}^{n} \|\boldsymbol{x}_i\|^2$$

# Equivalence of cMDS and PCA




- Explained variance (by H):

$$I(H) = \frac{1}{n}\sum_{i=1}^{n} \|\hat{\boldsymbol{x}}_i\|^2$$

- Residual variance:

$$I(H^\perp) = \frac{1}{n}\sum_{i=1}^{n} \|\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i\|^2$$

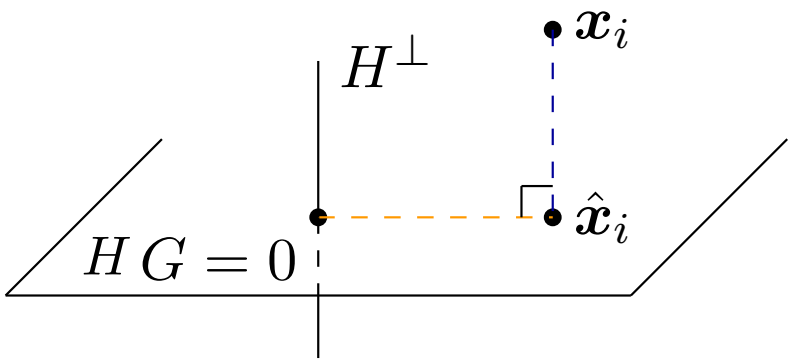
- According to Pythagoras' theorem:
  $\|\boldsymbol{x}_i\|^2 = \|\hat{\boldsymbol{x}}_i\|^2 + \|\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i\|^2$

- Total variance:

$$I_G = I_0 = \frac{1}{n}\sum_{i=1}^{n} \|\boldsymbol{x}_i\|^2$$

- Total variance = explained variance + residual variance

# Equivalence of cMDS and PCA




- Explained variance (by H):

$$I(H) = \frac{1}{n}\sum_{i=1}^{n} \|\hat{\boldsymbol{x}}_i\|^2$$

- Residual variance:

$$I(H^\perp) = \frac{1}{n}\sum_{i=1}^{n} \|\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i\|^2$$

$\Rightarrow$ Minimizing the distance criterion and maximizing the variance criterion leads to the same result for the euclidean distance!

- Total variance:

$$I_G = I_0 = \frac{1}{n}\sum_{i=1}^{n} \|\boldsymbol{x}_i\|^2$$

# Summary of cMDS

**Goal:** Determine a lower-dimensional representation from a distance matrix $D$

1. Calculate the centering matrix $H = \mathbf{I}_n - \mathbf{1}_n(w_1, \ldots, w_n)^\top$ (mostly, $w_i = \frac{1}{n}$).

2. Determine, for $\boldsymbol{D}$ denoting the matrix of squared euclidean distances

$$B = -\frac{1}{2}\boldsymbol{HDH}.$$

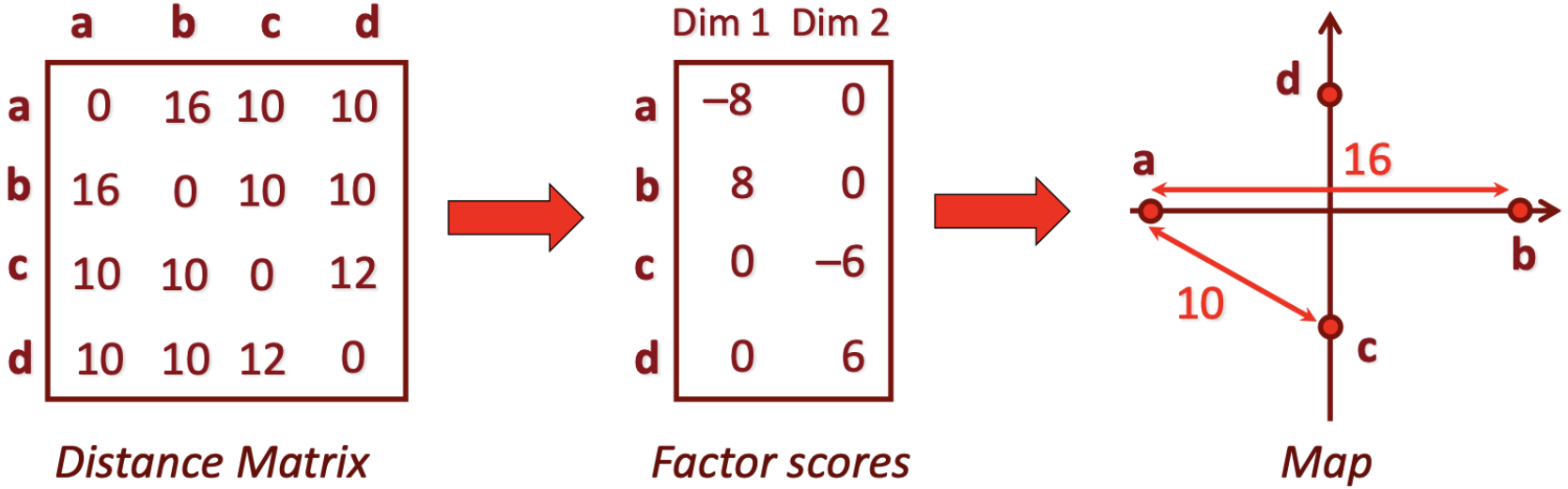
3. Carry out a spectral decomposition of $B$:

$$B = P\Lambda P^T$$

4. Calculate the points in the lower dimensional representation $\mathbf{Y}$.

# Small hands-on example of cMDS I

Figure 1: The Main Steps of Multidimensional Scaling: 1) Start with a distance matrix, 2) Transforms the distance matrix into a set of factor scores, and 3) Plot the observations using their factor scores. Here the distances on the map exactly recover the distances in the original matrix.




Source:https://personal.utdallas.edu/~herve/abdi-MDS-sage2022.pdf

# Small hands-on example of cMDS II

Here, we have

$$\mathbf{D}_{\text{Euclid}} = \begin{bmatrix} 0 & 16 & 10 & 10 \\ 16 & 0 & 10 & 10 \\ 10 & 10 & 0 & 12 \\ 10 & 10 & 12 & 0 \end{bmatrix}, \mathbf{D} = \begin{bmatrix} 0 & 256 & 100 & 100 \\ 256 & 0 & 100 & 100 \\ 100 & 100 & 0 & 144 \\ 100 & 100 & 144 & 0 \end{bmatrix}$$

and a "weights vector"

$$\mathbf{w}^\top = \begin{bmatrix} .25 & .25 & .25 & .25 \end{bmatrix}.$$

This results in the following centering matrix:

$$\boldsymbol{H} = \begin{bmatrix} .75 & -.25 & -.25 & -.25 \\ -.25 & .75 & -.25 & -.25 \\ -.15 & -.25 & .75 & -.25 \\ -.25 & -.25 & -.25 & .75 \end{bmatrix}.$$

# Small hands-on example of cMDS III

Next, we calculate

$$B = -\frac{1}{2}\boldsymbol{HDH} = \begin{bmatrix} 64 & -64 & 0 & 0 \\ -64 & 64 & 0 & 0 \\ 0 & 0 & 36 & -36 \\ 0 & 0 & -36 & 36 \end{bmatrix}$$

The eigen-decomposition of $B$ gives

$$B = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top \text{ with } \quad \mathbf{U} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 \\ 0 & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \text{ and } \boldsymbol{\Lambda} = \begin{bmatrix} 128 & 0 \\ 0 & 72 \end{bmatrix}.$$

# Small hands-on example of cMDS IV

Which gives us the following lower-dimensional representation:

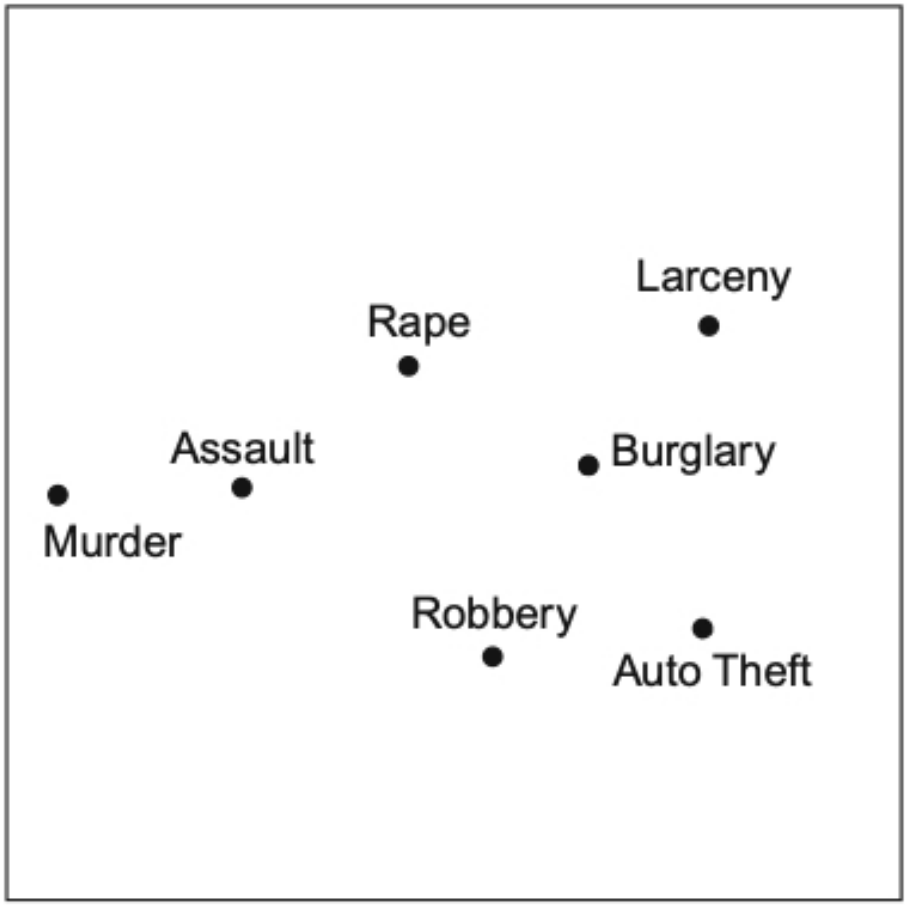
$$\mathbf{Y} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}} = \begin{bmatrix} -\sqrt{\frac{128}{2}} & 0 \\ \sqrt{\frac{128}{2}} & 0 \\ 0 & -\sqrt{\frac{72}{2}} \\ 0 & \sqrt{\frac{72}{2}} \end{bmatrix} = \begin{bmatrix} -\sqrt{64} & 0 \\ \sqrt{64} & 0 \\ 0 & -\sqrt{36} \\ 0 & \sqrt{36} \end{bmatrix}$$

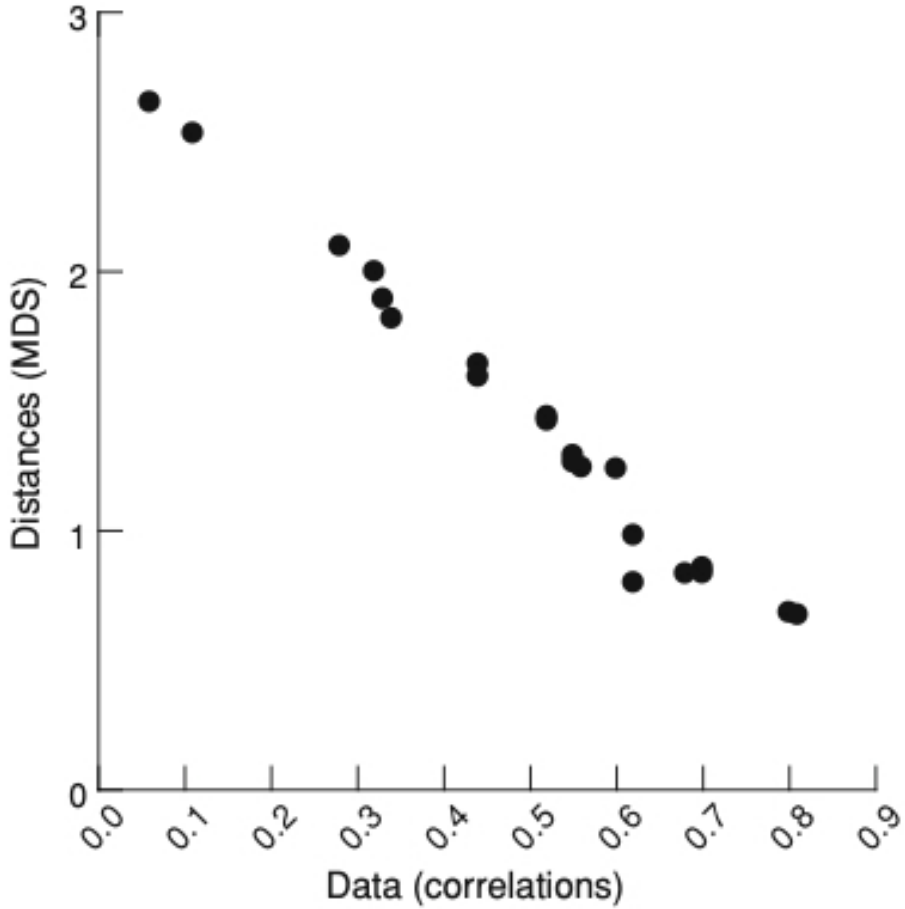
$$= \begin{bmatrix} -8 & 0 \\ 8 & 0 \\ 0 & -6 \\ 0 & 6 \end{bmatrix}.$$

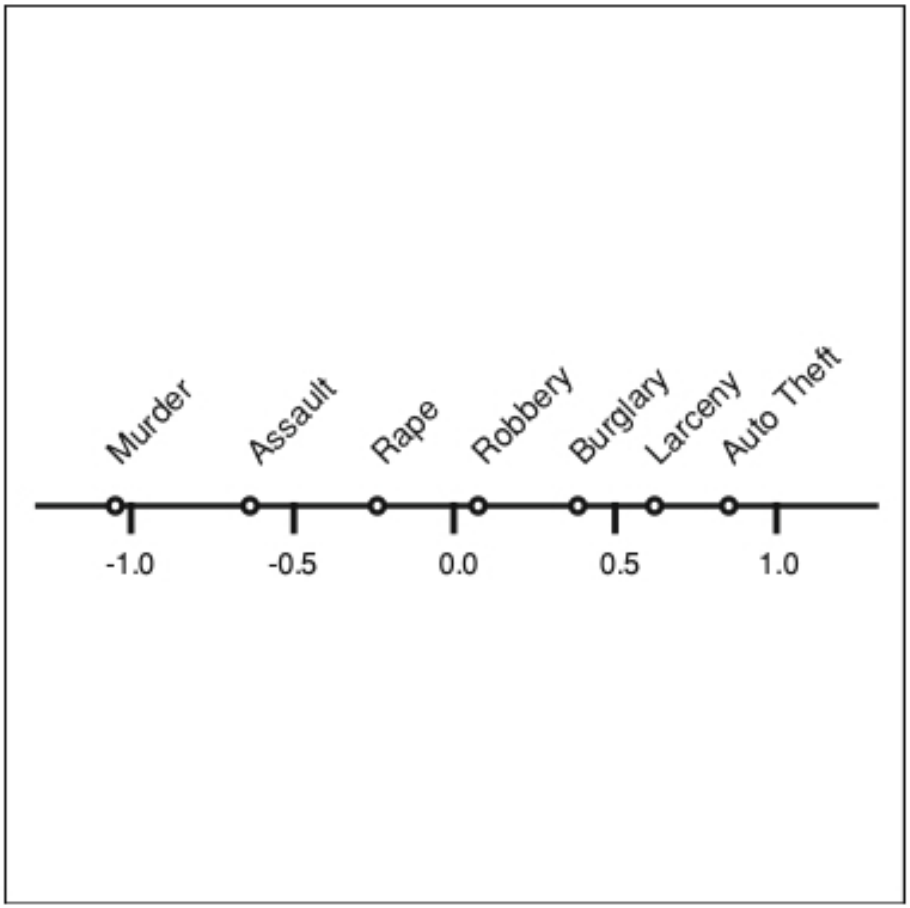
# Applied example: Criminality in the US

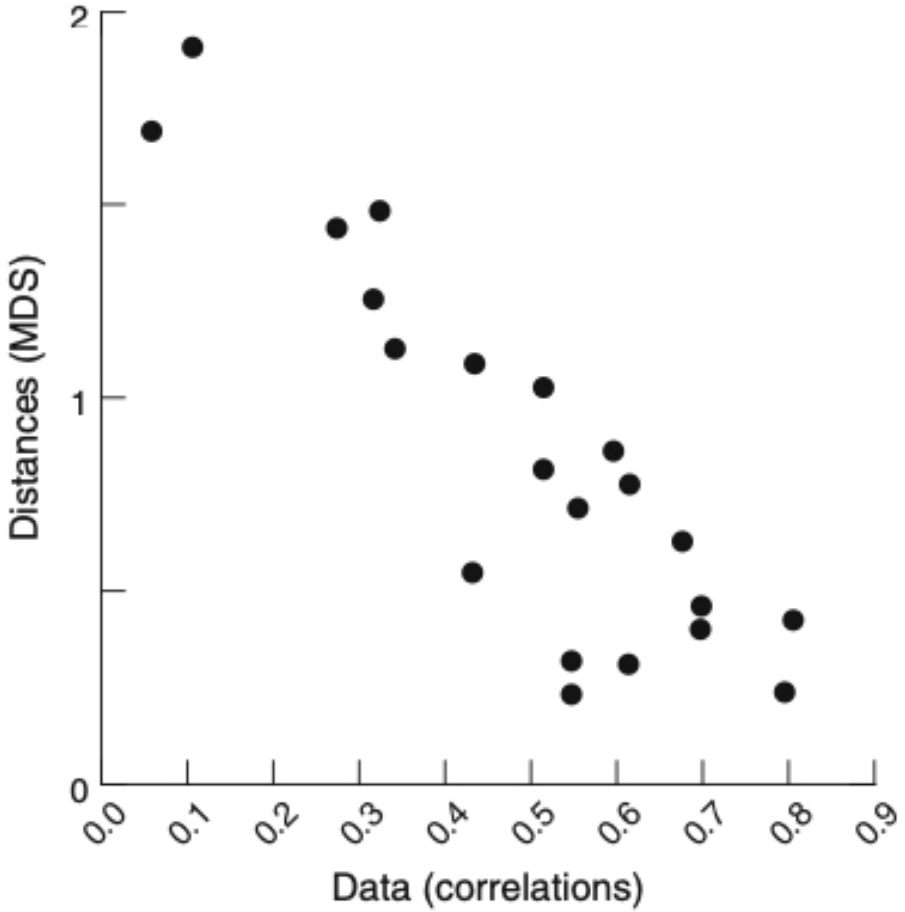
**Table 1.1** Correlations of crime rates over 50 U.S. states

| Crime | Murder | Rape | Robbery | Assault | Burglary | Larceny | Auto theft |
|---|---|---|---|---|---|---|---|
| Murder | 1.00 | 0.52 | 0.34 | 0.81 | 0.28 | 0.06 | 0.11 |
| Rape | 0.52 | 1.00 | 0.55 | 0.70 | 0.68 | 0.60 | 0.44 |
| Robbery | 0.34 | 0.55 | 1.00 | 0.56 | 0.62 | 0.44 | 0.62 |
| Assault | 0.81 | 0.70 | 0.56 | 1.00 | 0.52 | 0.32 | 0.33 |
| Burglary | 0.28 | 0.68 | 0.62 | 0.52 | 1.00 | 0.80 | 0.70 |
| Larceny | 0.06 | 0.60 | 0.44 | 0.32 | 0.80 | 1.00 | 0.55 |
| Auto theft | 0.11 | 0.44 | 0.62 | 0.33 | 0.70 | 0.55 | 1.00 |


Larceny
Rape
Assault
Burglary
Murder
Robbery
Auto Theft


Distances (MDS)
Data (correlations)
0
1
2
3
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9


Murder
Assault
Rape
Robbery
Burglary
Larceny
Auto Theft
-1.0
-0.5
0.0
0.5
1.0


Distances (MDS)
Data (correlations)
0
1
2
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

# Iterative approach to solutions

Especially not performing classical MDS, a typical solution is to iteratively solve the minimization of the *cost or STRESS* function by

1. Initializing a random lower-dimensional representation $\mathbf{Y}$
2. Improving on it until the STRESS function is smaller some constant $c$.

The improvement can be carried out via isotonic (or monotonic) regression, the technique of fitting a free-form line to a sequence of observations, *especially for non-metric MDS*.

# Non-metric MDS

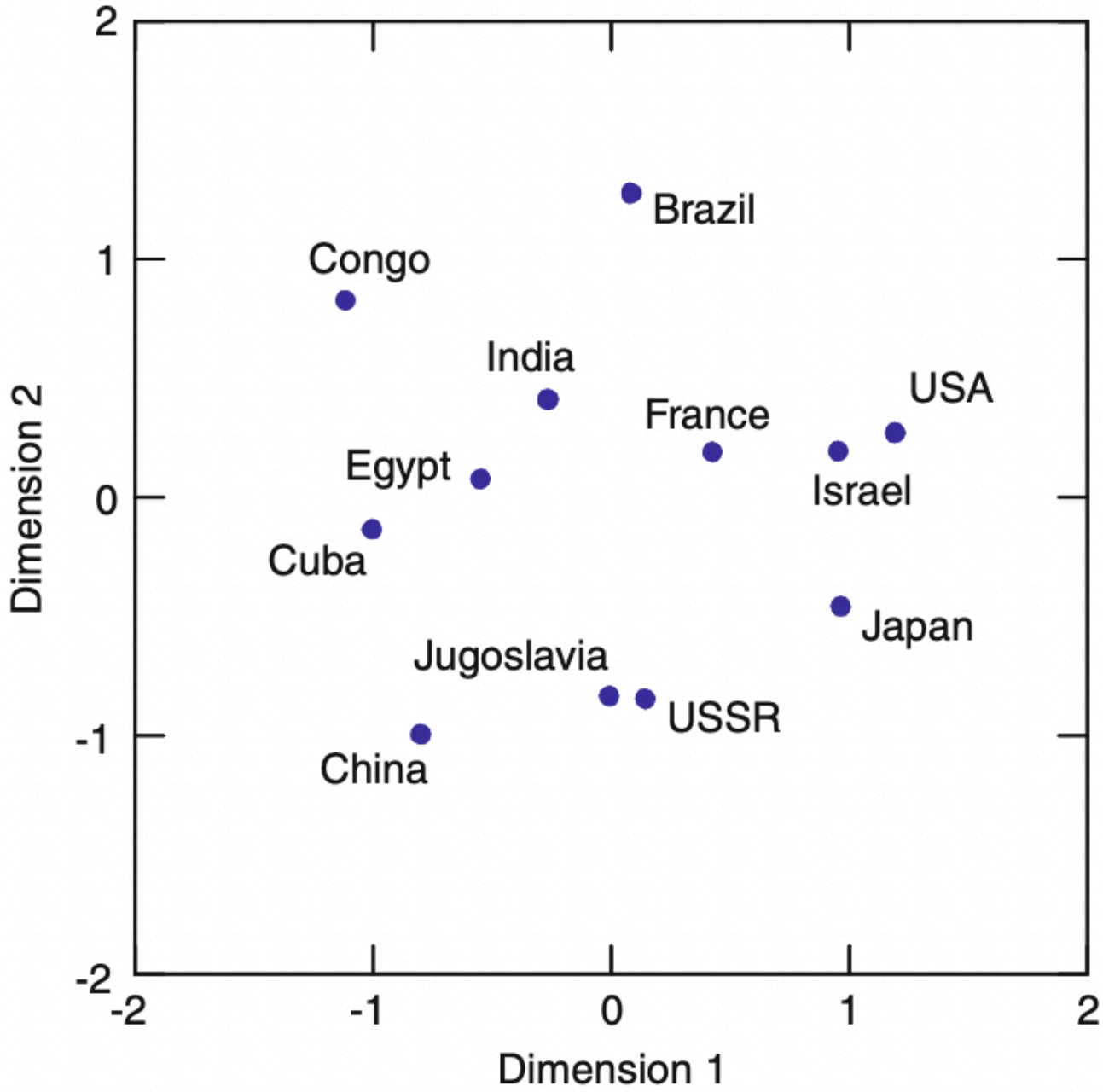
# Nichtmetrische MDS

Non-metric MDS methods go back to Shepard (1962) and only assume that there is a *monotonic* relationship between the similarity ranking of the object pairs and the object distances.

The monotonicity condition is as follows:

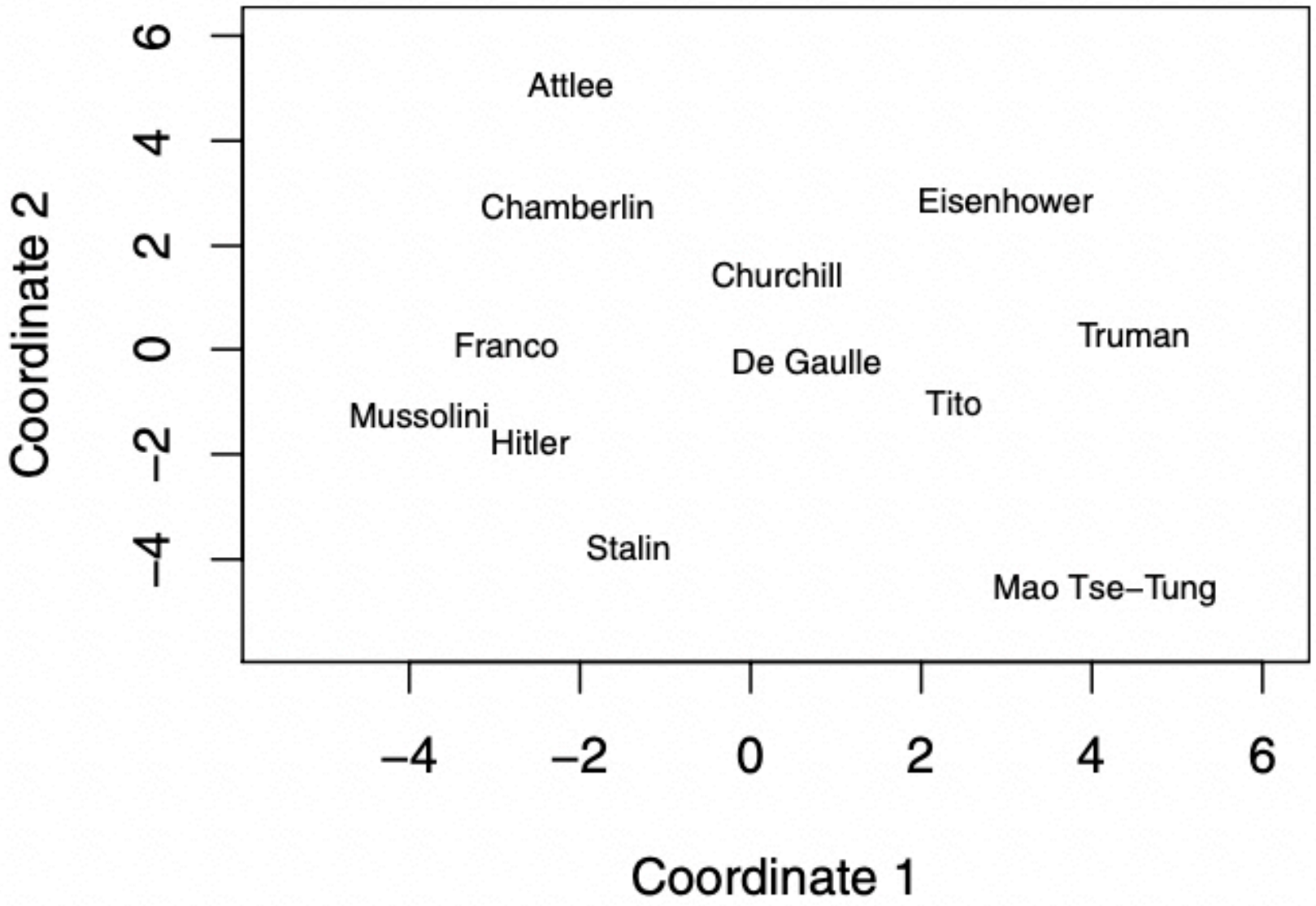
$$(i, \tilde{i}) \text{ more similar than } (k, \tilde{k}) \;\Rightarrow\; d_{i\tilde{i}} \;<\; d_{k\tilde{k}}$$

| Country | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Brazil | 1 | – | | | | | | | | | | | |
| Congo | 2 | 4.83 | – | | | | | | | | | | |
| Cuba | 3 | 5.28 | 4.56 | – | | | | | | | | | |
| Egypt | 4 | 3.44 | 5.00 | 5.17 | – | | | | | | | | |
| France | 5 | 4.72 | 4.00 | 4.11 | 4.78 | – | | | | | | | |
| India | 6 | 4.50 | 4.83 | 4.00 | 5.83 | 3.44 | – | | | | | | |
| Israel | 7 | 3.83 | 3.33 | 3.61 | 4.67 | 4.00 | 4.11 | – | | | | | |
| Japan | 8 | 3.50 | 3.39 | 2.94 | 3.83 | 4.22 | 4.50 | 4.83 | – | | | | |
| China | 9 | 2.39 | 4.00 | 5.50 | 4.39 | 3.67 | 4.11 | 3.00 | 4.17 | – | | | |
| USSR | 10 | 3.06 | 3.39 | 5.44 | 4.39 | 5.06 | 4.50 | 4.17 | 4.61 | 5.72 | – | | |
| USA | 11 | 5.39 | 2.39 | 3.17 | 3.33 | 5.94 | 4.28 | 5.94 | 6.06 | 2.56 | 5.00 | – | |
| Jugoslavia | 12 | 3.17 | 3.50 | 5.11 | 4.28 | 4.72 | 4.00 | 4.44 | 4.28 | 5.06 | 6.67 | 3.56 | – |


Brazil
Congo
India
USA
France
Egypt
Israel
Cuba
Japan
Jugoslavia
USSR
China
Dimension 2
Dimension 1
2
1
0
-1
-2
-2
-1
0
1
2

| | Htl | Mss | Chr | Esn | Stl | Att | Frn | DGl | MT- | Trm | Chm | Tit |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Hitler | 0 | | | | | | | | | | | |
| Mussolini | 3 | 0 | | | | | | | | | | |
| Churchill | 4 | 6 | 0 | | | | | | | | | |
| Eisenhower | 7 | 8 | 4 | 0 | | | | | | | | |
| Stalin | 3 | 5 | 6 | 8 | 0 | | | | | | | |
| Attlee | 8 | 9 | 3 | 9 | 8 | 0 | | | | | | |
| Franco | 3 | 2 | 5 | 7 | 6 | 7 | 0 | | | | | |
| De Gaulle | 4 | 4 | 3 | 5 | 6 | 5 | 4 | 0 | | | | |

| | Htl | Mss | Chr | Esn | Stl | Att | Frn | DGl | MT- | Trm | Chm | Tit |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mao Tse-Tung | 8 | 9 | 8 | 9 | 6 | 9 | 8 | 7 | 0 | | | |
| Truman | 9 | 9 | 5 | 4 | 7 | 8 | 8 | 4 | 4 | 0 | | |
| Chamberlin | 4 | 5 | 5 | 4 | 7 | 2 | 2 | 5 | 9 | 5 | 0 | |
| Tito | 7 | 8 | 2 | 4 | 7 | 8 | 3 | 2 | 4 | 5 | 7 | 0 |


Attlee
Chamberlin
Eisenhower
Churchill
Truman
Franco
De Gaulle
Tito
Mussolini
Hitler
Stalin
Mao Tse–Tung
Coordinate 2
6
4
2
0
−2
−4
−4
−2
0
2
4
6
Coordinate 1

# Solution of non-metric MDS

The aim is to determine a representation of the objects that fulfills the monotonicity condition in a space with the smallest possible dimensions.

Iteratively minimize the following STRESS function:

$$S(\mathbf{Y}) = \sqrt{\frac{\sum_{k<\ell} (d_{k\ell} - \hat{d}_{k\ell})^2}{\sum_{k<\ell} d_{k\ell}^2}}\,.$$

$\rightarrow$ The most widely used method is that of Kruskal (1964).

# Cool visualization

A really nice interactive visualization of MDS is given in the following blog-post:

Visualizing MNIST: An Exploration of Dimensionality Reduction