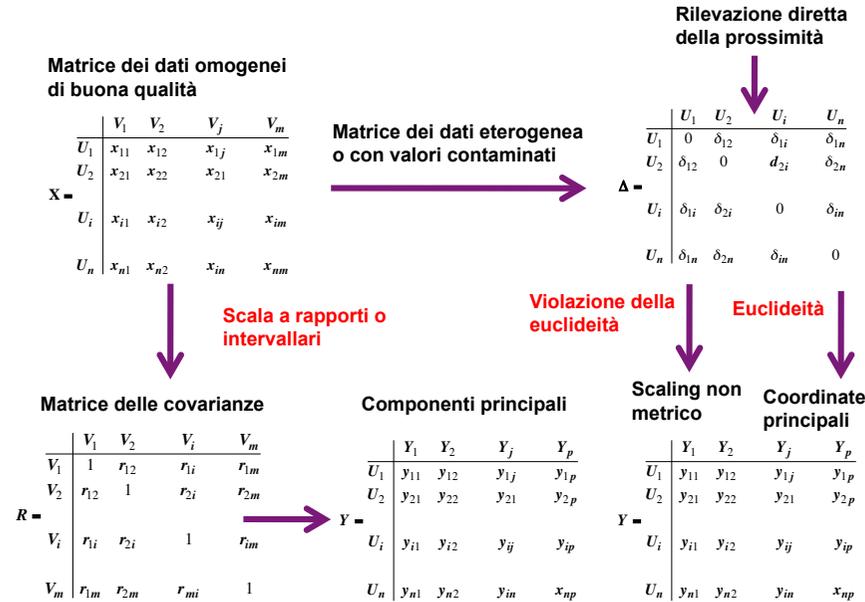


# Transizione



# Scaling delle distance (distance scaling)

Qui la base di partenza è la matrice delle dissimilarità/distanze

$$\Delta = \begin{matrix} & U_1 & U_2 & U_i & U_n \\ U_1 & 0 & \delta_{12} & \delta_{1i} & \delta_{1n} \\ U_2 & \delta_{12} & 0 & \delta_{2i} & \delta_{2n} \\ \vdots & \vdots & \vdots & 0 & \vdots \\ U_i & \delta_{1i} & \delta_{2i} & 0 & \delta_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ U_n & \delta_{1n} & \delta_{2n} & \delta_{in} & 0 \end{matrix}$$

Le dissimilarità  $\delta_{ij}$  sono rappresentate con punti le cui interdistanze  $d_{ij}$  approssimano le dissimilarità originali.

**Unità molto simili corrispondono a punti ravvicinati.**

**Punti molto distanti corrispondono a unità dissimili**

Le distanze dovrebbero essere almeno concordi in senso ordinale con le dissimilarità osservate fra le unità.

$$d_{ij} \leq d_{rs} \text{ se } \delta_{ij} \leq \delta_{rs}$$

La rappresentazione geometrica delle dissimilarità dovrebbe avvenire con distanze riferite a poche dimensioni: due o, al massimo, tre.

# Mappe & distanze

# Cenno storico



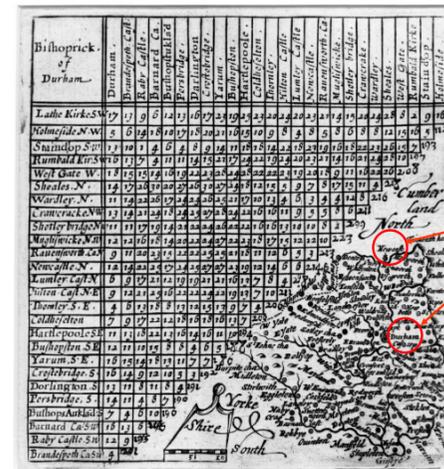
*Aveva un bavero color zafferano e la marsina color ciclamino veniva a piedi da Lodi a Milano per incontrare la bella Gigogin.*

Per andare da Milano a Lodi in macchina si impiegano 43 min.

Napoleone ci mise quasi sette ore a cavallo a passo d'uomo

**Lo scaling multidimensionale opera nella direzione opposta. Sono note le distanze e si deve arrivare ad una mappa.**

In questa procedura si privilegiano gli ordinamenti tra le dissimilarità originali e quindi la tecnica è robusta rispetto alla contaminazione dei dati e ai valori remoti



Newcastle Durham

Map of Durham county

- Cartographer: Jacob van Langren
- Date 1633

**Tuttavia, molte introduzioni allo scaling presentano come esempio iniziale una matrice delle interdistanze tra località diverse**

## Mappe & distanze/3

	ATL	BOS	ORD	DCA	DEN	LAX	MIA	JFK	SEA	SFO	MSY
ATL	0	934	585	542	1209	1942	605	751	2181	2139	424
BOS	934	0	853	392	1769	2601	1252	183	2492	2700	1356
ORD	585	853	0	598	918	1748	1187	720	1736	1857	830
DCA	542	392	598	0	1493	2305	922	209	2328	2442	964
DEN	1209	1769	918	1493	0	836	1723	1636	1023	951	1079
LAX	1942	2601	1748	2305	836	0	2345	2461	957	341	1679
MIA	605	1252	1187	922	1723	2345	0	1092	2733	2594	669
JFK	751	183	720	209	1636	2461	1092	0	2412	2577	1173
SEA	2181	2492	1736	2328	1023	957	2733	2412	0	681	2101
SFO	2139	2700	1857	2442	951	341	2594	2577	681	0	1925
MSY	424	1356	830	964	1079	1679	669	1173	2101	1925	0

Airline distances between 11 US cities

Data set "cities", package "psych"



Le soluzioni dello scaling metrico sono spesso sottoposte a rotazione per riprodurre la spazialità dei dati

## Metodo di scaling

La funzione che lega dissimilarità originali e distanze approssimanti determina il metodo di scaling che si intende adoperare

### SCALING METRICO (Procedura di Torgerson, 1950)

$$\delta_{ij} = \beta_0 + \beta_1 d_{ij} + e_{ij} \begin{cases} \delta_{ij} = \text{osservate} \\ d_{ij} = \text{teoriche} \end{cases}$$

Dove  $\beta_0$  e  $\beta_1$  sono due coefficienti incogniti che caratterizzano la relazione.

Se la matrice delle distanze è euclidea allora  $\beta_0 = 0$  e  $\beta_1 = 1$ ; se solo pochi autovalori sono negativi rimanendo piccoli in assoluto allora  $\beta_0$  e  $\beta_1$  forniscono la correzione.

### SCALING NON METRICO (Procedura di Shepard-Kruskal)

$$\delta_{ij} = f(d_{ij}) + e_{ij} \quad \text{dove} \begin{cases} \delta_{ij} & \text{distanza originale} \\ d_{ij} & \text{distanza approssimata} \\ e_{ij} & \text{errore} \end{cases}$$

La funzione  $f$  deve essere monotona (la derivata prima non cambia di segno).

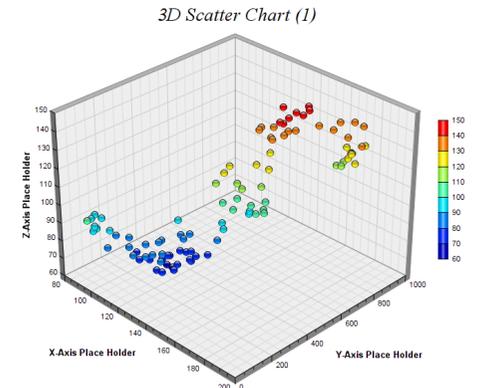
## Mappe & distanze/4

### Dissimilarità osservate

$$\Delta = \begin{matrix} & U_1 & U_2 & \dots & U_i & \dots & U_n \\ U_1 & 0 & \delta_{12} & \dots & \delta_{1i} & \dots & \delta_{1n} \\ U_2 & \delta_{12} & 0 & \dots & \delta_{2i} & \dots & \delta_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ U_i & \delta_{1i} & \delta_{2i} & \dots & 0 & \dots & \delta_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ U_n & \delta_{1n} & \delta_{2n} & \dots & \delta_{in} & \dots & 0 \end{matrix}$$

### Pseudo-coordinate

$$Y = \begin{matrix} & Y_1 & Y_2 & \dots & Y_j & \dots & Y_p \\ U_1 & y_{11} & y_{12} & \dots & y_{1j} & \dots & y_{1p} \\ U_2 & y_{21} & y_{22} & \dots & y_{2j} & \dots & y_{2p} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ U_i & y_{i1} & y_{i2} & \dots & y_{ij} & \dots & y_{ip} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ U_n & y_{n1} & y_{n2} & \dots & y_{in} & \dots & y_{np} \end{matrix}$$



Rappresentazione geometrica con le nuove coordinate

## Pietre miliari

P. J.F. Groenen,  
I. Borg

Years	Main author(s)	Topic
<i>Past</i>		
1958, 1966	Torgerson, Gower	Classical MDS
1962	Shepard	First MDS heuristic
1964	Kruskal	Least-squares MDS through Stress with transformations
1964	Guttman	Facet theory and regional interpretations in MDS
1969, 1970	Horan, Carroll	Three-way MDS models (INDSCAL, IDIOSCAL)
1977-	De Leeuw and others	The majorization algorithm for MDS
<i>Present</i>		
1986-1998	Meulman	Distance-based MVA through MDS
1994	Buja	Constant dissimilarities
1978, 1995-	Various	Local minimum problem
1998	Buja	Smart use of weights in MDS
<i>Future</i>		
1999-	Heiser, Meulman, Busing	Modern MDS software: Proxscal in SPSS (PASW)
2000	Tenenbaum, et al.	Large scale MDS ISOMAP heuristic
2002	Buja, Swayne, Cook	Dynamic MDS in GGvis (part of GGobi)
2003	Groenen	Dynamic MDS visualization through iMDS
2005-	Groenen, Trosset, Kagié	Large scale MDS through Stress
2002	Deneux, Groenen, Winsberg, Diday	Symbolic MDS of interval dissimilarities
2006	Groenen, Winsberg	Symbolic MDS of histograms
2009	De Leeuw, Mair	SMACOF package in R

## Esempio

Matrice dei punteggi assegnati a 14 zone nel comprensorio di Thurio

Matrice dei dati 14 x 9

	Clima	Abitazioni	Salute	Crimine	Trasporti	Istruzione	Cultura	Spettacolo	Ricchezza
Rondena	0.185	-0.942	-0.899	-1.500	1.356	-0.198	-0.174	1.511	0.879
Righino	-0.217	-1.338	-0.350	-0.397	-0.789	0.774	-0.596	0.580	2.026
Fontanelle	-0.628	0.719	-0.451	0.977	-0.820	-0.856	0.636	-0.941	2.135
Bosco_Mezzano	-0.156	-0.718	0.196	-0.609	-1.139	-0.498	-0.239	0.652	1.617
Andrano	1.692	-0.007	-0.876	-0.591	0.817	0.423	0.661	-1.419	-1.585
Sornieto	-1.078	0.945	0.752	-0.231	0.715	-0.413	1.112	-0.100	0.347
Cozzo_D_Este	0.077	-1.046	-0.673	0.703	-1.647	1.642	-0.700	0.431	-0.640
Santa_Bruna	-0.470	0.347	-0.879	2.520	-0.264	-0.568	0.225	-1.352	-0.201
Pascino	-0.611	-0.995	0.640	0.911	0.355	1.142	1.297	-1.186	0.327
Lago_d_Ora	-1.142	1.451	1.848	-0.995	-0.301	-0.966	-0.006	-0.220	0.005
Torniella	-1.542	1.951	2.348	-1.495	-0.801	-1.466	-0.506	-0.720	0.505
Centrano	-0.111	-0.495	1.140	1.411	0.855	1.642	1.797	-0.686	0.827
San_Macloedio	-0.085	-1.142	-1.099	-1.700	1.156	-0.398	-0.374	1.311	0.679
Selluzzi	-0.102	-0.338	-0.175	-0.196	-0.389	0.372	-0.282	0.290	1.013

I valori in tabella esprimono valori ottenuti dalle zone in base alle diverse variabili.

Gli andamenti non sono tutti concordi: abitazioni e crimine con valori bassi indicano una posizione migliore delle zone con valori alti

## Scaling metrico delle distanze

Supponiamo che i dati siano contaminati. Adoperiamo una tecnica diversa: lo scaling metrico.

Trasformazione delle informazioni contenute nella matrice dei dati in dissimilarità usando la distanza di Manhattan con variabili standardizzate.

$$\delta_{ij} = \sum_{r=1}^9 \frac{|x_{ir} - x_{jr}|}{\sigma_r} \Rightarrow D$$

Matrice delle dissimilarità 14 x 14

E' ovvio che si tratta di una matrice euclidea

	Rondena	Righino	Fontanelle	Bosco_Mezzano	Andrano	Sornieto	Cozzo_D_Este	Santa_Bruna	Pascino	Lago_d_Ora	Torniella	Centrano	San_Macloedio	Selluzzi
Rondena	0.00	8.24	12.91	7.07	11.41	10.77	10.64	12.04	12.15	12.40	14.74	12.34	1.97	6.67
Righino	8.24	0.00	8.69	3.97	13.77	11.85	6.60	12.01	10.37	12.29	13.54	12.01	7.77	3.98
Fontanelle	12.91	8.69	0.00	7.88	12.68	8.24	12.39	6.37	8.92	9.44	10.91	11.06	12.92	8.20
Bosco_Mezzano	7.07	3.97	7.88	0.00	13.63	9.46	8.28	10.45	10.95	9.38	11.29	11.89	6.91	3.84
Andrano	11.41	13.77	12.68	13.63	0.00	11.16	12.84	9.87	10.76	14.24	17.77	12.21	12.14	10.54
Sornieto	10.77	11.85	8.24	9.46	11.16	0.00	13.97	9.28	6.86	5.75	9.64	8.56	10.38	8.33
Cozzo_D_Este	10.64	6.60	12.39	8.28	12.84	13.97	0.00	10.94	10.19	14.00	16.22	11.51	10.37	7.07
Santa_Bruna	12.04	12.01	6.37	10.45	9.87	9.28	10.94	0.00	8.66	9.41	12.94	11.22	11.66	8.65
Pascino	12.15	10.37	8.92	10.95	10.76	6.86	10.19	8.66	0.00	11.55	14.44	4.70	12.07	8.77
Lago_d_Ora	12.40	12.29	9.44	9.38	14.24	5.75	14.00	9.41	11.55	0.00	4.57	13.34	12.02	8.93
Torniella	14.74	13.54	10.91	11.29	17.77	9.64	16.22	12.94	14.44	4.57	0.00	15.95	13.84	11.61
Centrano	12.34	12.01	11.06	11.89	12.21	8.56	11.51	11.22	4.70	13.34	15.95	0.00	12.66	9.17
San_Macloedio	1.97	7.77	12.92	6.91	12.14	10.38	10.37	11.66	12.07	12.02	13.84	12.66	0.00	6.81
Selluzzi	6.67	3.98	8.20	3.84	10.54	8.23	7.07	8.65	8.77	8.93	11.61	9.17	6.81	0.00

Se i dati contengono valori anomali è preferibile evitare la distanza euclidea e la distanza di Mahalanobis.

## Analisi delle componenti principali

I dati sono su scala a rapporti e quindi, in assenza di informazioni sulla loro qualità, dovremmo applicare la tecnica delle componenti principali.

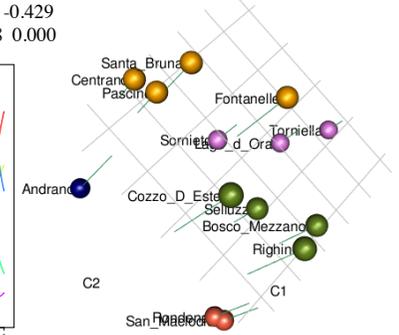
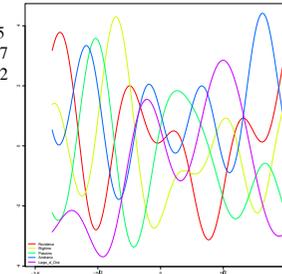
	PC1	PC2	PC3	Varimax	PC1	PC2	PC3
Clima	-0.493	0.000	0.000	Clima	0.530	0.000	0.000
Abitazioni	0.530	0.000	0.000	Abitazioni	0.000	0.469	0.000
Salute	0.474	0.000	0.000	Salute	0.000	0.493	0.664
Crimine	0.000	0.469	0.493	Crimine	0.000	0.000	0.417
Trasporti	0.000	0.000	-0.664	Trasporti	0.000	-0.277	0.000
Istruzione	-0.413	0.000	0.253	Istruzione	0.000	-0.401	0.259
Cultura	0.000	0.535	0.000	Cultura	-0.559	0.000	0.000
Spettacolo	0.000	-0.548	0.000	Spettacolo	0.447	0.000	-0.429
Ricchezza	0.000	-0.297	0.387	Ricchezza	0.292	-0.468	0.000

Sono stati ricostruiti 5 gruppi  
Unità leader

Rondena	-1.315	-1.918	-1.230
Righino	-0.877	-1.549	1.676
Pascino	-0.224	1.885	0.801
Andrano	-1.378	1.611	-1.640
Lago d'Or	2.760	-0.398	-0.658

Eigv	22.9	15.7	12.5
p.Var.	49.3	23.2	14.7
cum.p. Var	49.3	72.5	87.2

C1= Movida  
C2= Benessere  
C3= Sicurezza



## Scaling metrico delle distanze/2

Lo scaling metrico prevede il doppio centramento della matrice delle distanze

$$\tilde{B} = CHC$$

Dove

$$C = \left( I - \frac{1}{n} uu^t \right) \quad e \quad H = -\frac{1}{2} D^2$$

Seguita dalla scomposizione in valori singolari

$$\tilde{B} = \sum_{r=1}^p \lambda_r v_r v_r^t = VL^t V, \quad \text{con } V^t V = I_n, \quad L = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$$

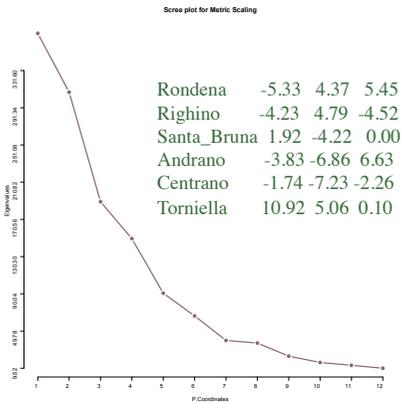
P=2 o p=3

Si passa poi alla determinazione delle coordinate principali

$$VL^{0.5} = \tilde{X}$$

## Esempio: comprensorio di Thurio

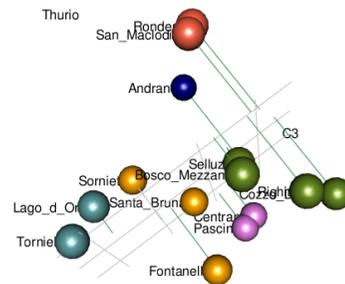
Lo scarto maggiore si nota a  $p=3$ , ma nel complesso le prime tre coordinate principali spiegano i 2/3 della variabilità originale



Percentuale cumulata di variabilità spiegata dalle Princ. Coord.

[1]	28.317	51.780	66.224	77.627
[5]	84.535	89.570	92.588	95.387
[9]	97.109	98.308	99.275	100.000
[13]	100.000	100.000		

Sono stati ricostruiti 6 gruppi



La configurazione dei punti in 3d è diversa da quella ottenuta con l'ACP.

Si tratta di due modi diversi di analizzare i dati. Entrambi possono essere utili

## Stress

Ci serve subito una misura indicativa di quanto la soluzione ottenuta si discosti da quella originale

$$S_p(X) = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n [\delta_{ij} - d_{ij}(X)]^2}{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \delta_{ij}^2}$$

$p$  è il numero di dimensioni o pseudo-coordinate.

Il compito delle  $\delta_{ij}$  è di dare un ordinamento per le  $d_{ij}(X)$

Lo STRESS dipende solo dalle distanze stimate  $d_{ij}(X)$  (le dissimilarità sono fisse) e quindi è invariante rispetto a rotazioni e traslazioni perché tali sono le sue componenti.

Minore è il valore dello stress maggiore è la qualità della approssimazione.

Valori prossimi all'unità indicano che le dissimilarità stimate (disparità) sono vicine allo zero e quindi non danno alcuna idea della configurazione spaziale delle unità.

Lo stress è nullo se la matrice ricostruita è uguale a quella originale.

## Distanze vere/Distanze presunte

L'intero  $p$  indica il numero di pseudo-coordinate. La matrice delle distanze ricostruite in base alle coordinate principali è ottenuta da

$$\tilde{D}^2 = \tilde{b}u^t + u\tilde{b}^t - 2\tilde{B}$$

Se  $p < 9$  le distanze ricostruite non possono essere esatte, ma solo delle approssimazioni.

Adottiamo le definizioni

$$\begin{cases} \delta_{ij} = \text{dissimilarità originale} \\ d_{ij}(X) = \text{distanza approssimata} \end{cases}$$

La  $X$  in parentesi ci ricorda la dipendenza delle distanze approssimate calcolate sulle pseudo-coordinate.

## Valutazione dello stress

Lo stress varia tra zero ed uno. Gli intervalli maggiormente condivisi sono i seguenti:

Stress values	goodness-of-fit
Stress > 0.3	any positioning of points
0.2 < Stress ≤ 0.3	deficient (for N < 50)
0.1 < Stress ≤ 0.2	unsatisfactory
0.05 < Stress ≤ 0.1	satisfactory
0.01 < Stress ≤ 0.05	good
0.00 < Stress ≤ 0.01	very good
Stress = 0.00	perfect

## SStress

Takane-Young\_De Leew (1977) hanno proposto una misura di adattamento ritenuta più consona alla idea dello scaling delle distanze

$$SS_p(X) = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n [\delta_{ij}^2 - \phi_{ij}(X)]^2}{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \delta_{ij}^4}$$

La potenza 2 delle dissimilarità originali potrebbe essere sostituita con un altro esponente.

Dove  $\phi_{ij}$  è la approssimazione del quadrato della distanza e non la approssimazione della distanza al quadrato

Stress e Sstress non sono equivalenti. Il secondo privilegia le distanze grandi, sia quelle note che quelle approssimate.

Sarà necessario accertare che il software consenta questa opzione.

## Scree plot/2

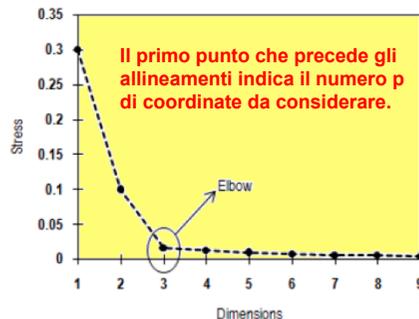
Il grafico scree porta a valutare lo scarto tra valori dello stress per un numero di dimensioni crescenti:  $S_k(X) - S_{k+1}(X)$

Il metodo consiste nel trovare il punto in cui la inclinazione dei segmenti diventa quasi costante in modo da presentarsi praticamente come facenti parte di una sola retta.

Questo vuol dire che i contributi che possono pervenire da queste dimensioni sono dei disturbi che poco possono contribuire a migliorare la soluzione.

Il difetto del metodo è che in molti casi o non ci sono punti di svolta ben delineati oppure ve ne sono diversi.

In questi casi lo scree plot è di scarsa utilità.

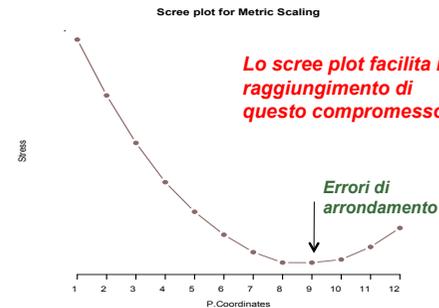


## Scree plot

A parità di condizioni lo STRESS diminuisce all'aumentare del numero di dimensioni su cui sono calcolate le disparità.

Quindi per migliorare la soluzione basterebbe aumentare il numero di coordinate.

Tuttavia, il nostro obiettivo è ottenere il migliore adattamento tra distanze originali e disparità con poche dimensioni perché più facili da interpretare.



## Esempio introduttivo: scaling non metrico

Comici italiani: varie forme di umorismo

Matrice delle dissimilarità

	Albanese	Bisio	Brignano	Cacioppo	Crozza	Zalone
Albanese	0	8	20	10	25	15
Bisio	8	0	12	16	2	28
Brignano	20	12	0	40	22	6
Cacioppo	10	16	40	0	5	4
Crozza	25	2	22	5	0	30
Zalone	15	28	6	4	30	0

> is.euclid(D)  
[1] FALSE

Invece di tentare lo scaling metrico previo aggiustamento alla euclideanità della matrice delle dissimilarità, ipotizziamo una rappresentazione nel piano iniziale, da migliorare progressivamente.

## Idea dello scaling non metrico

E' possibile seguire una sequenza iterativa di soluzioni che riscalano gradualmente la matrice delle dissimilarità originali.

Si fissa il numero di dimensioni p a partire da valori piccoli: 1,2, 3, ... , (n-1)

Per il numero p fissato di coordinate principali si cerca la migliore matrice delle distanze (nello spazio a p dimensioni) della matrice originale delle dissimilarità

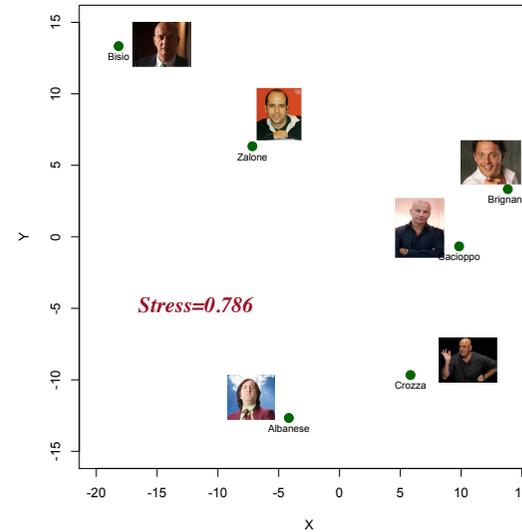
Si parte da una configurazione iniziale dei punti scelta in modo opportuno (anche a caso)

$$X_{ij}, \quad i = 1,2,\dots,n; \quad j = 1,2,\dots,p$$

E si perviene a quella finale con dei passaggi che via via riducono lo scarto tra le distanze geometriche e le dissimilarità tra le unità.

Tra una iterazione e la successiva i punti della configurazione si spostano in base ad un criterio orientato al miglioramento della approssimazione. Ad esempio in accordo con una riduzione dello stress.

## Esempio introduttivo/2



Generiamo a caso 6 coppie di valori tra 1 e 40 e calcoliamo le distanze euclidee tra i punti così ottenuti. Queste distanze sono la nostra prima approssimazione delle dissimilarità

Confronti	Oss	App	scarto	Dir
Bri-Cac	40	5.66	34.34	1
Cac-Cro	5	9.85	4.85	-1
Alb-Cro	25	10.44	14.56	1
Bis-Zal	28	13.04	14.96	1
Bri-Cro	22	15.26	6.74	1
Cac-Zal	4	18.38	14.38	-1
Alb-Cac	10	18.44	8.44	-1
Alb-Zal	15	19.24	4.24	-1
Cro_Zal	30	20.62	9.38	1
Bri-Zal	6	21.21	15.21	-1
Alb-Bri	20	24.08	4.08	-1
Alb-Bis	8	29.53	21.53	-1
Bis-Cac	16	31.30	15.30	-1
Bis-Cro	2	33.24	31.24	-1
Bis-Bri	12	33.53	21.53	-1

Gli scarti tra le coordinate e le dissimilarità sono grandi. Per ridurli dobbiamo spostare i punti: Bisio e Albanese debbono essere più vicini, ma Bisio deve anche essere allontanato da Zalone

## Smacof (Scaling by majorizing a convex function)

Lo stress può essere riscritto come:

*Sviluppo del quadrato*

$$S(X) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n [\delta_{ij} - d_{ij}(X)]^2 = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \delta_{ij}^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij}^2(X) - 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \delta_{ij} d_{ij}(X)$$

$$= \tau + \eta^2(X) - 2\theta(X) \text{ dove } \begin{cases} \eta^2(X) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij}^2(X) \\ \theta(X) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \delta_{ij} d_{ij}(X) \\ \tau = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \delta_{ij}^2 \end{cases}$$

La somma delle dissimilarità osservate al quadrato è prefissata e non dipende dalle pseudo-coordinate X. Il denominatore può essere accantonato.

$\eta^2(X)$  è una funzione convessa delle coordinate X

## Smacof /2

Il secondo addendo è legato alla traccia della matrice dei prodotti incrociati

$$d_{ij}^2(X) = (x_i - x_j)'(x_i - x_j) = \text{Traccia}[X^t A_{ij} X] \text{ dove } A_{ij} = (e_i - e_j)(e_i - e_j)'$$

$A_{ij}$  è data dal prodotto esterno della differenza tra due vettori coordinate cioè vettori di zeri tranne uno nella posizione indicata.

$$e_3 - e_5 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}; (e_3 - e_5)(e_3 - e_5)' = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Notare che le matrici  $A_{ij}$  hanno somma zero sulle righe e sulle colonne

$$d_{13}^2(X) = \sum_{a=1}^m x_a'(e_1 - e_3)(e_1 - e_3)'x_a$$

La matrice  $A_{ij}$  ha elementi nulli tranne che in  $a_{ii}=a_{jj}=1$  e  $a_{ij}=-1$

$$\begin{aligned} &= \sum_{a=1}^m x_a' \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} x_a = \sum_{a=1}^m x_a' A_{13} x_a \\ &= \text{tr } X' A_{13} X. \end{aligned}$$

## Smacof /2b

A questo punto abbiamo

$$d_{ij}^2(\mathbf{X}) = Tr(\mathbf{X}'\mathbf{A}_{ij}\mathbf{X})$$

Sommando rispetto a tutti gli  $i < j$  otteniamo

$$\eta^2(X) = Traccia[\mathbf{X}'\mathbf{V}\mathbf{X}] \text{ con } \mathbf{V} = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbf{A}_{ij} \quad v_{ij} = \begin{cases} -1 & \text{se } i \neq j \\ (n-1) & \text{se } i = j \end{cases}$$

Poiché  $\eta^2(X)$  è quadratica rispetto ad  $X$  si può agevolmente ottimizzare.

$\mathbf{V}$  è somma di matrici doppio centrate (cioè a somma di colonna e di riga zero) e quindi anche  $\mathbf{V}$  sarà tale.

La matrice  $\mathbf{V}$  non ha rango pieno e la matrice inversa tradizionale non esiste.

## Smacof 3/b

Se dividiamo entrambi lati della disuguaglianza per  $d_{ij}(Y)$  avremo

$$\frac{\sum_{i=1}^n (x_{ih} - x_{jh})(y_{ih} - y_{jh})}{d_{ij}(Y)} \leq d_{ij}(X) \Rightarrow -d_{ij}(X) \leq \frac{\sum_{i=1}^n (x_{ih} - x_{jh})(y_{ih} - y_{jh})}{d_{ij}(Y)}$$

Procedendo come prima possiamo scrivere

$$(x_i - x_j)'(y_i - y_j) = Traccia[\mathbf{X}'\mathbf{A}_{ij}\mathbf{Y}]$$

Combinando le relazioni

$$-d_{ij}(X) \leq \frac{Traccia[\mathbf{X}'\mathbf{A}_{ij}\mathbf{Y}]}{d_{ij}(Y)} \Rightarrow -d_{ij}(X)\delta_{ij} \leq \frac{\delta_{ij}Traccia[\mathbf{X}'\mathbf{A}_{ij}\mathbf{Y}]}{d_{ij}(Y)}$$

## Smacof 3

Il terzo addendo è definito come  $\theta(X) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \delta_{ij} d_{ij}(X)$

Ricordiamo la disuguaglianza di Cauchy-Schwarz

*L'uguaglianza si ha se*

$$\sum_{h=1}^n p_h q_h \leq \left( \sum_{h=1}^n p_h^2 \right)^{0.5} \left( \sum_{h=1}^n q_h^2 \right)^{0.5} \quad q_h = c p_h$$

Sostituiamo  $p$  e  $q$  con due scarti uno tra le variabili  $X$  (già note) ed uno tra le variabili  $Y$  (ancora da trovare)

$$p_h = (x_{ih} - x_{jh}); \quad q_h = (y_{ih} - y_{jh})$$

$$\sum_{h=1}^n (x_{ih} - x_{jh})(y_{ih} - y_{jh}) \leq \left( \sum_{h=1}^n (x_{ih} - x_{jh})^2 \right)^{0.5} \left( \sum_{h=1}^n (y_{ih} - y_{jh})^2 \right)^{0.5} = d_{ij}(X) d_{ij}(Y)$$

Con uguaglianza se  $\mathbf{X}=\mathbf{Y}$ .

## Smacof 3/c

Sommando rispetto a  $i < j$  si ottiene

$$-\theta(X) = -\sum_{i=1}^{n-1} \sum_{j=i+1}^n \delta_{ij} d_{ij}(X) \leq -Tr \left[ \mathbf{X}' \left( \sum_{i < j} b_{ij} \mathbf{A}_{ij} \right) \mathbf{Y} \right] = -Tr[\mathbf{X}'\mathbf{B}(Y)\mathbf{Y}]$$

Dove la matrice  $\mathbf{B}(Y)$  è formata da

$$b_{ij} = \begin{cases} -\frac{\delta_{ij}}{d_{ij}(Y)} & \text{se } i \neq j \text{ e } d_{ij}(Y) \neq 0 \\ 0 & \text{altrimenti} \\ -\sum_{j=1, j \neq i}^n b_{ij} & \text{se } i = j \end{cases}$$

La relazione importante è che

$$-\theta(X) = -Tr[\mathbf{X}'\mathbf{B}(Y)\mathbf{X}] \leq Tr[\mathbf{X}'\mathbf{B}(Y)\mathbf{Y}]$$

C'è quindi modo di ridurre  $\theta(x)$  se si pone  $\mathbf{Y}$  al posto di  $\mathbf{X}$

# Algoritmo di De Leuw (smacof)

# Comici. Smacof

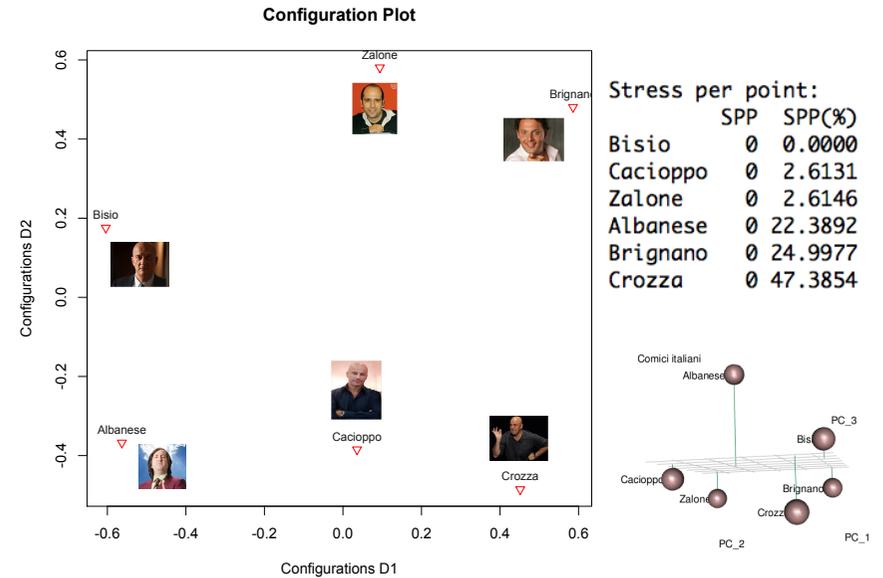
La minimizzazione dello stress diventa uno schema iterativo:

- 1) Si determina una configurazione iniziale dei punti  $X_0$  nel numero di dimensioni già prefissato (2,3,...).  
*Di solito si parte da una generazione di punti casuali. Usare lo scaling metrico come base di partenza si può, ma è dispendioso.*
- 2) Si valuta lo stress della configurazione. Se è soddisfacente lo schema si ferma altrimenti si continua purché non si sia superato il numero massimo consentito di iterazioni.
- 3) Si modifica la configurazione con la trasformazione Guttman

$$X_{r+1} = V^+ [B(X_r)] X_r, \quad V^+ = \left(\frac{1}{n}\right) \left[ I - \left(\frac{1}{n}\right) uu^t \right]$$

Per poi tornare al punto 2.

Lo schema di De Leuw è tale che ad ogni iterazione lo stress diminuisce ovvero diminuisce una funzione che è superiore allo stress (maggiorazione). La soluzione finale non è necessariamente unica ed occorre effettuare più tentativi con diverse configurazioni di partenza.



## Distanze approssimate

Perché ricostruire le dissimilarità in base alle distanze euclidee? Non potrebbe essere utile una delle metriche di Minowski?

$$d_{ij}(X) = \left[ \sum_{r=1}^p |x_{ir} - x_{jr}|^\alpha \right]^{1/\alpha} \quad \alpha \geq 1$$

In particolare la city-block ( $\alpha=1$ ) ha mostrato di essere la candidata ideale quando si vuole ridurre l'enfasi sulle dissimilarità più grandi

Una prima ragione per privilegiare la distanza euclidea ( $\alpha=2$ ) è la relazione che sussiste tra tale norma al quadrato ed il prodotto scalare tra vettori.

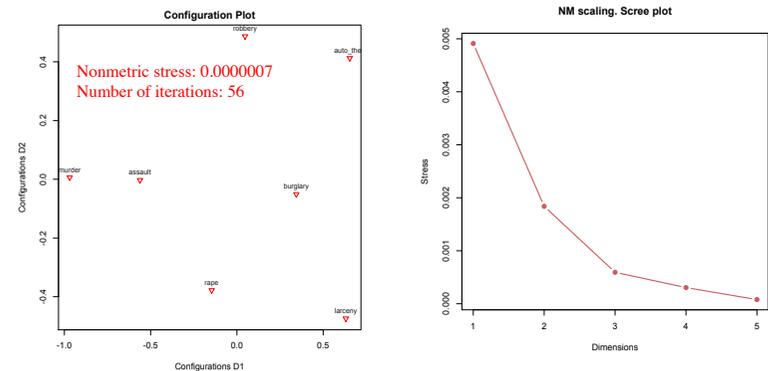
Le distanze euclidee sono invarianti rispetto alle rotazioni ortogonali. Lo stesso non si può dire delle altre metriche.

Infine, la distanza euclidea costituisce una buona approssimazione delle altre distanze.

## Applicazione\_1: crimine

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

$$D = \sqrt{2(1-R)}$$



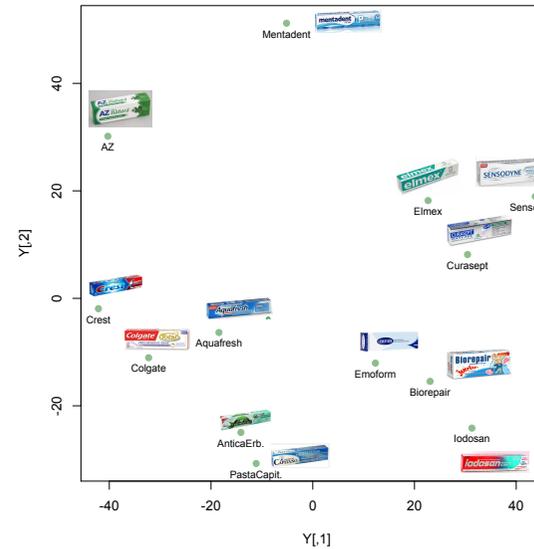
## Applicazione 4: dentifrici

L'elicitazione del gradimento di un prodotto o di un servizio è una scienza a parte. Nel caso dei dentifrici abbiamo trovato la tabella che segue

	AnticaErb.	Aquafresh	AZ	Biorepair	Colgate	Crest	Curasept	Elmex	Emoform	Iodosan	Mentadent	PastaCapit.	Sensodyne
AnticaErb.	31												
Aquafresh	23	31											
AZ	21	22	31										
Biorepair	30	28	9	31									
Colgate	25	27	18	15	31								
Crest	28	26	26	12	27	31							
Curasept	17	17	8	28	10	6	31						
Elmex	23	18	13	29	9	4	24	31					
Emoform	30	28	2	26	5	21	21	14	31				
Iodosan	29	21	1	25	8	2	17	17	18	31			
Mentadent	18	25	30	11	7	16	14	29	25	11	31		
PastaCapit.	28	23	25	19	13	19	12	26	26	21	1	31	
Sensodyne	3	5	6	27	7	8	28	26	22	29	15	6	31
AnticaErb.	0												
Aquafresh	8	0											
AZ	10	9	0										
Biorepair	1	3	22	0									
Colgate	6	4	13	16	0								
Crest	3	5	5	19	4	0							
Curasept	14	14	23	3	21	25	0						
Elmex	8	13	18	2	22	27	7	0					
Emoform	1	3	29	5	26	10	10	17	0				
Iodosan	2	10	30	6	23	29	14	14	13	0			
Mentadent	13	6	1	20	24	15	17	2	6	20	0		
PastaCapit.	3	8	6	12	18	12	19	5	5	10	30	0	
Sensodyne	28	26	25	4	24	23	3	5	9	2	16	25	0

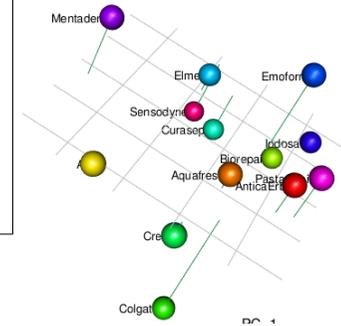
Dalla graduazione delle preferenze medie si passa alle dissimilarità come complemento a 31.

## Dentifrici: scaling metrico

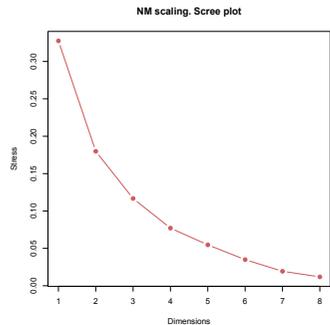


Le prime due Coord P. spiegano il 34% e quindi la mappa percettiva non è esplicativa.

L'aggiunta della 3ª porta al 46% che è ancora poco soddisfacente

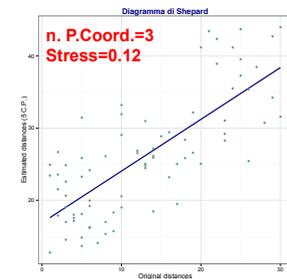
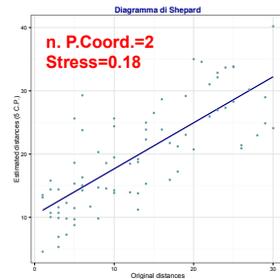


## Dentifrici: scaling non metrico

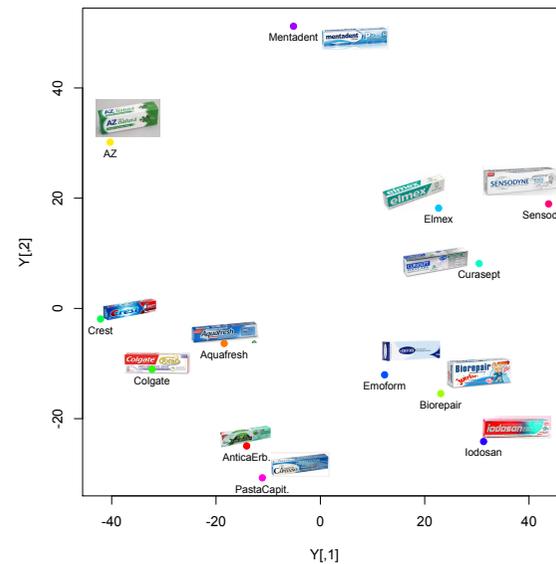


Il numero di P.Coord. necessario per una rappresentazione accurata è superiore a 3. Quindi, mappe percettive di dubbia efficacia.

Le rappresentazioni grafiche per il 4D sono possibili con 3D e forme diverse o colori diversi, ma con 4 coordinati lo stress sarebbe ancora a 0.08



## Dentifrici: scaling non metrico



Il risultato è identico a quello ottenuto con lo scaling non metrico, ma almeno non si è intervenuti per "curare" la mancanza di euclideanità nella matrice delle dissimilarità.

La disposizione sulle ascisse sembra alludere al luogo d'acquisto. Sulle ordinate verosimilmente c'è un affioramento del costo dei prodotti.

## Dipendenza dalla configurazione iniziale

Se la prima disposizione dei punti è ottenuta generando casualmente le loro coordinate si corre il rischio che la soluzione finale non sia quella ottima.

Le trasformazioni Guttman convergono, ma non necessariamente verso la configurazione migliore.

E' possibile avviare le schema Guttman dalla soluzione fornita dallo scaling metrico, magari corretta con la costante additiva ovvero ottenuta con un numero di dimensioni corrispondenti agli autovalori positivi.

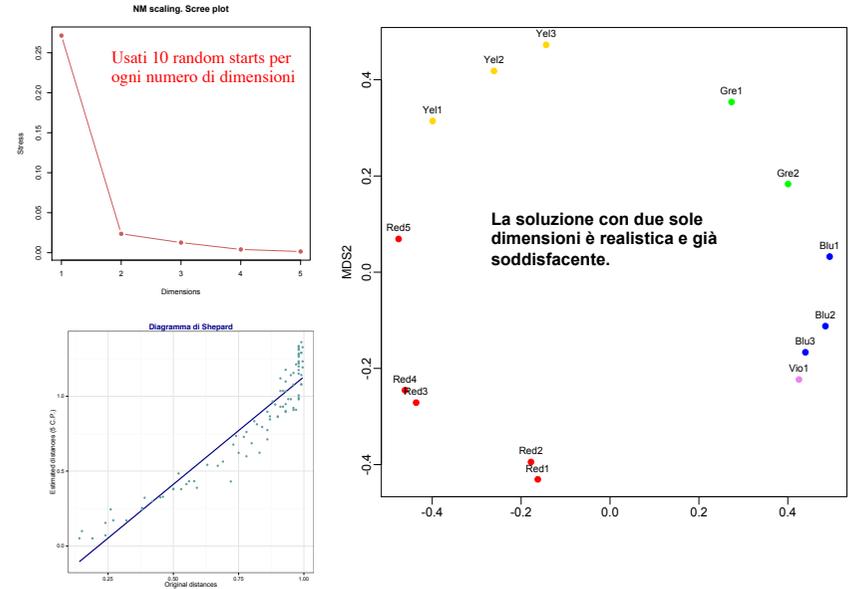
E' onerosa e l'efficacia non è garantita.

Un'altra strategia è di ripetere il ciclo iterativo un certo numero di volte:

**100, 1000, 2000**

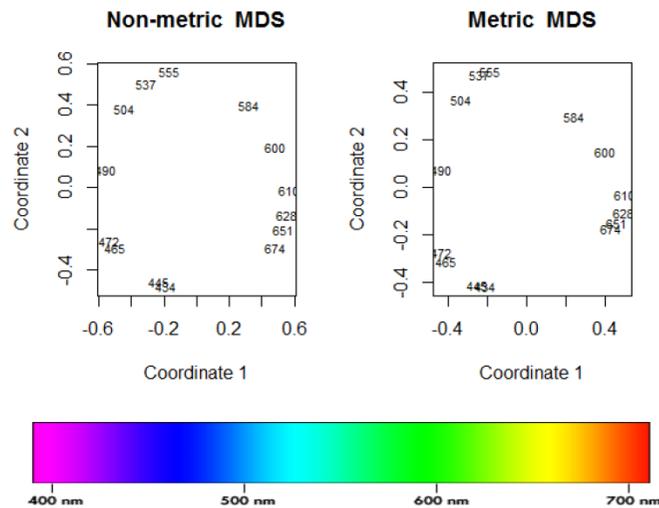
Tenuto conto del numero di unità e delle risorse di calcolo disponibili.

## Esempio: Ekman dat (colors)

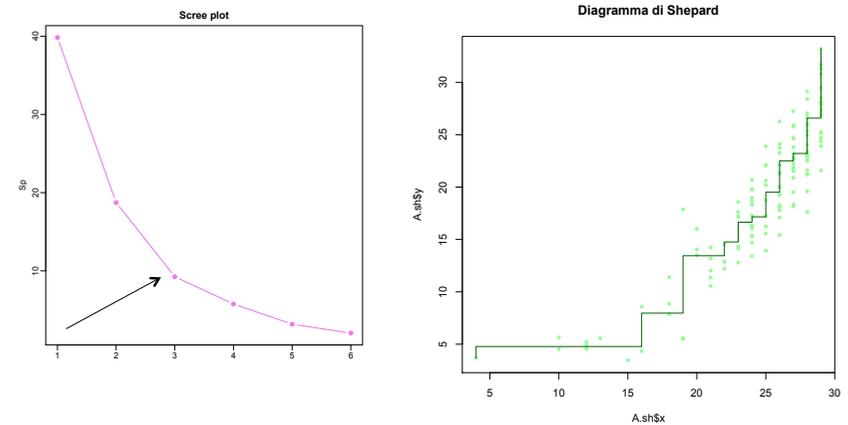


## Esempio: Ekman dat (colors)/2

MDS reproduces the well-known two-dimensional *color circle*.



## Applicazione\_5: Sounds (isoMDS)

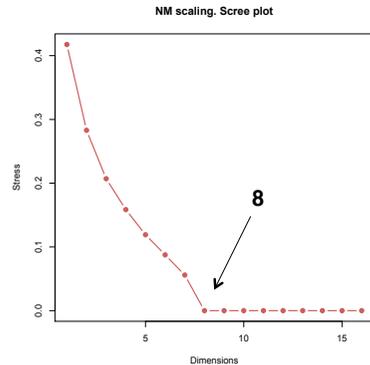


Questo algoritmo di MDS consente di stabilire l'esponente della metrica di Minkowski usata per ricostruire le distanze.

Il numero di dimensioni sembra essere tre.



## Confronto MS/NMS large data set/3



Con lo scaling non metrico si ottiene

K=3, ARI=0.15 e K=9, ARI=0.29

Sui dati completi originali si ha invece

K=3, ARI=0.71 e K=9, ARI=0.63

In questo caso, l'uso delle pseudo-variabili ha disperso informazioni importanti.

Dai dati non si è potuto estrarre tutto quello che serviva.

Lo scaling metrico e non metrico offrono sempre e comunque un risposta ai quesiti loro posti. Si tratta di valutare l'utilità di questa risposta.

## Sammon Mapping

E' una tecnica che converte un vettore in uno spazio m-dimensionale in un piano cartesiano oppure in 3D.

L'idea guida è riorganizzare tutte le osservazioni in modo che le distanze nel piano siano quanto più possibile prossime alle distanze originali.

La funzione obiettivo da minimizzare proposta da Sammon (1969) è

$$S = \left(\frac{1}{\tau}\right) \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{[\delta_{ij} - d_{ij}]^2}{\delta_{ij}}$$

*Disparità o distanze approssimate*

$$\tau = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \delta_{ij}^2, \quad d_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2}$$

La procedura è di tipo non lineare e riesce a conservare abbastanza bene la struttura delle interdistanze nella matrice originale.

## Sammon Mapping/2



Si parte da una configurazione iniziale (di solito random) nel piano (p=2) o nello spazio (p=3) e si calcolano le distanze euclidee dei punti virtuali (disparità)



Si valuta la funzione obiettivo  $SE_2(X)$  e si controlla che lo scarto tra le dissimilarità originali non abbia raggiunto uno livello minimo accettabile rispetto alle disparità



In caso contrario si aggiorna la configurazione

$$x_{ij}^{(q+1)} = x_{ij}^{(q)} - \alpha \Delta_{ij}^{(q)} \quad \text{con } 0.3 \leq \alpha \leq 0.4$$

$$\Delta_{ij}^{(q)} = \frac{\frac{\partial S^{(q)}}{\partial x_{ij}^{(q)}}}{\frac{\partial^2 S^{(q)}}{\partial [x_{ij}^{(q)}]^2}} \leftarrow \text{Questo è il gradiente della funzione obiettivo}$$

## Sammon Mapping/3

$$\begin{cases} \frac{\partial S^{(q)}}{\partial x_{ij}^{(q)}} = \frac{-2}{\tau} \sum_{r=1, s \neq i}^n \left( \frac{\delta_{ij} - d_{ij}^{(q)}}{\delta_{ij} d_{ij}^{(q)}} \right) [x_{ij}^{(q)} - x_{rj}^{(q)}] \\ \frac{\partial^2 S^{(q)}}{\partial [x_{ij}^{(q)}]^2} = \frac{-2}{\tau} \sum_{r=1, s \neq i}^n \frac{1}{\delta_{ij} d_{ij}^{(q)}} \left[ \left( \delta_{ij} - d_{ij}^{(q)} \right) - \frac{(x_{ij}^{(q)} - x_{rj}^{(q)})^2}{d_{ij}^{(q)}} \left( 1 + \frac{\delta_{ij} - d_{ij}^{(q)}}{d_{ij}^{(q)}} \right) \right] \end{cases}$$

La convergenza della procedura è lenta e comunque sensibile al numero magico  $\alpha$ . Può essere applicata con tranquillità solo quando il numero di unità non è troppo grande.

Non sono ammessi gli zeri fuori diagonale né nella matrice originale che in quella approssimante.

D'altra parte le mappe di Sammon sono strettamente legate alla procedura di MDS

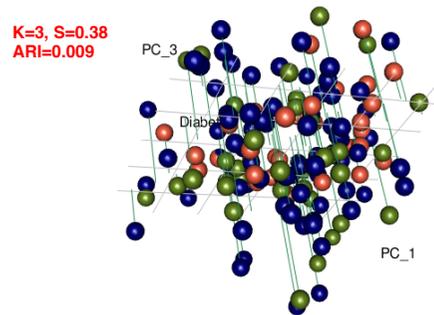
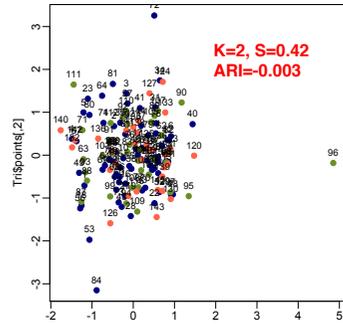
## Esempio: diabetes

Partiamo da un data set nella usuale forma di matrice dei dati: 145 unità e cinque variabili divisi in  $k=3$  cluster.

La dissimilarità è misurata con la distanza di Mahalanobis che è euclidea per costruzione.

Tentiamo di riscalarare la matrice in due o tre dimensioni.

La configurazione iniziale per l'avvio di Sammon è ottenuta dalla comando runifpointx del pacchetto spatstat.

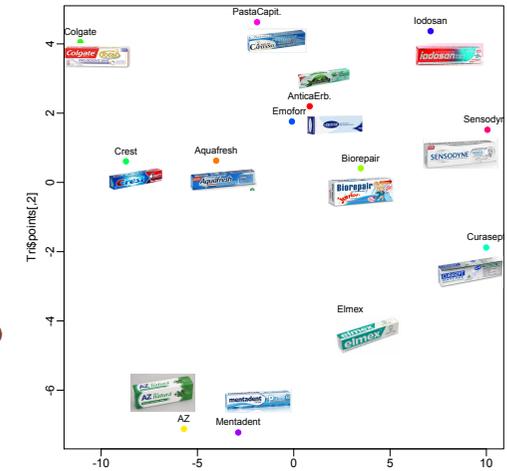
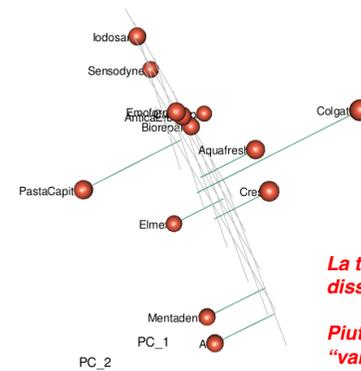


## Sammon Mapping/4

Qualcosa è cambiato rispetto agli scaling già fatti.

La circolarità della formazione è tipica del metodo Sammon.

Il 3D rivela l'isolamento del colgate e della pasta del capitano



*La tecnica non ha come obiettivo la mappatura delle dissimilarità originali.*

*Piuttosto si tenta di ottenere un nuovo data set con meno "variabili" di quelle originali, ma con struttura simile.*