Sailkapen-analisia
(*cluster analysis*)

Datu-analisiaren zenbait teknika

	Baieztatzailea Erabakitzailea *Gainbegiratua*	Miatzailea Deskribatzailea *Gainbegiratu gabea*
Lineala	Bariantz-analisia (ANOVA) Analisi bereizlea Erregresio-analisia	Faktore-analisia (osagai nagusiak (PCA), korrespondentziak, LSI, ...) Dimentsioanitzeko mailakatzea
Ez-lineala	Elkartze-arauak Segmentazioa (Erabakitze-zuhaitzak) (Sailkatze eta erregresio-zuhaitzak) K-auzokide gertuenak 'Support Vector Machine' (SVP)	Kontzeptuen sailkapena (COBWEB) Sailkapen automatikoa ('Cluster analysis') (Hierarkia, Partiketa ('k-means'),...)
	Neurona-sareak Probabilitate-sareak (bayestarrak)	

Sailkapen-analisia
(*cluster analysis*)

Sinonimoak

- ***Clustering*** (formen ezagutza, *pattern recognition*en arloan)
- *Salkapen automatikoa* (*datu-analisia*n)
- *Sailkapen gainbegiratu gabea* (*estatistika*n, *datu-analisia*n)
- *Ikasketa gainbegiratu gabea* (*adimen artifizial*ean)
- *Segmentazioa* (*marketing*ean, *seinaleen analisia*n)
- *Vector quantization* (*seinaleen analisia*n)
- *Ordenazioa* (*psikologia*n)
- *Taxonomia numerikoa* (*biologia*n)
- *Analisi tipologikoa*

Sailkapen-analisia
(*cluster analysis*)

Definizioa (ez zehatza)

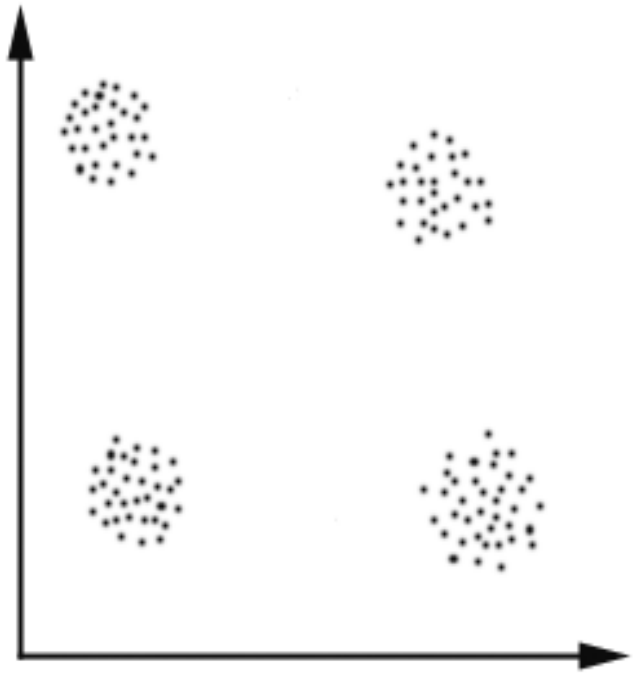
Sail edo *cluster* bat objektu-multzo bat da, zentzuren batean beraien artean antzekoak direnak, eta beste objektuekin alderatuta ezantzekoak.

Analisiaren **helburua**:

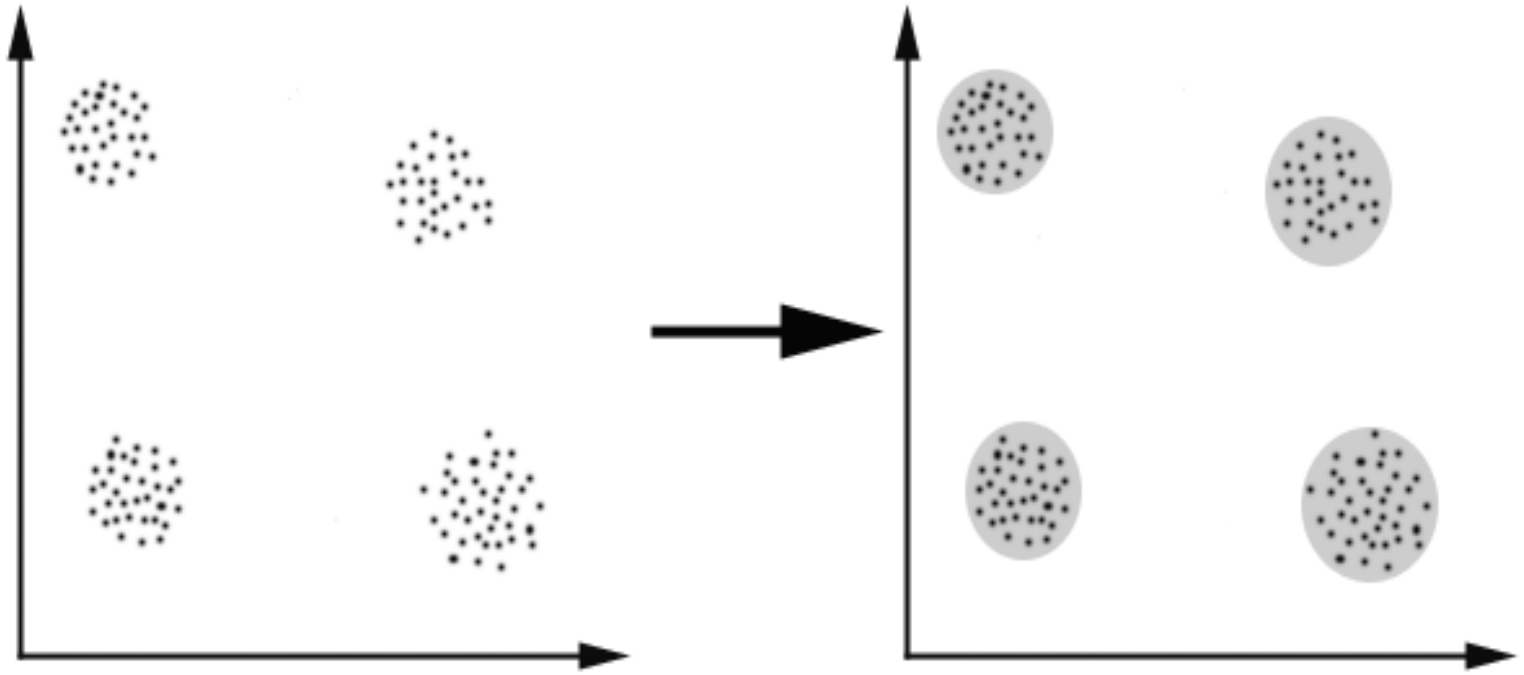
•Sail edo *cluster* horiek bilatu.

•Datuen multzoan *aldez aurretik* erezaguna den *egitura* bat bilatu.

Sailkapen-analisia
(*cluster analysis*)



Sailkapen-analisia
(*cluster analysis*)



Sailkapen-analisia
(*cluster analysis*)

‘Berezko irakasbide bat egiteko modu bakarra antzekoak diren gauzak elkarrekin jartzea, eta antzekoak ez direnak bereiztea da’. *Historia natural*, 1749.

Georges-Louis Leclerc, comte de Buffon (1707-1788)

Sailkapen-analisia
(*cluster analysis*)

Helburua

Konputagailu batean jarri gizakiok bi dimentsiotan sailak edo *cluster*rak **identifika**tzeko dugun gaitasuna, eta gero, gaitasun horiek dimentsio gehiagotara zabaldu.

Hasieran, zeregina erraza dela dirudi, baina praktikan zaila gertatzen da.

Sailkapen-analisia
(*cluster analysis*)

Helburua:
Sail edo *cluster* horiek bilatu.

Nola lortu helburua?

Sailkapen-analisia
(*cluster analysis*)

Helburua:
Sail edo *cluster* horiek bilatu.

Nola lortu helburua?

Abiapuntua
Datuak (erlazioak):
Objektuak aldagai edo ezaugarri anitzen bitartez daude deskribatuta.

Sailkapen-analisia
(*cluster analysis*)

Ω , objektuen multzoa

X_j , aldagaiak

x_{ij} , ω_i objektuaren balioa X_j aldagaian

Ω	X_1		X_j		X_p
ω_1	x_{11}		x_{1j}		x_{1p}
ω_i	x_{i1}		x_{ij}		x_{ip}
ω_n	x_{n1}		x_{nj}		x_{np}

Suitza, 1888: 47 herrialde frankofono

Zenbait adierazle sozioekonomiko (%) eta emankortasuna

X1 Fertility X2 Agriculture X3 Examination
X4 Education X5 Catholic X6 Infant.Mortality

	X1	X2	X3	X4	X5	X6
Courtelary	80.2	17.0	15	12	9.96	22.2
Delemont	83.1	45.1	6	9	84.84	22.2
Franches-Mnt	92.5	39.7	5	5	93.40	20.2
Moutier	85.8	36.5	12	7	33.77	20.3
Neuveville	76.9	43.5	17	15	5.16	20.6
Porrentruy	76.1	35.3	9	7	90.57	26.6
Broye	83.8	70.2	16	7	92.85	23.6
Glane	92.4	67.8	14	8	97.16	24.9
Gruyere	82.4	53.3	12	7	97.67	21.0
Sarine	82.9	45.2	16	13	91.38	24.4
Veveyse	87.1	64.5	14	6	98.61	24.5
Aigle	64.1	62.0	21	12	8.52	16.5
Aubonne	66.9	67.5	14	7	2.27	19.1
Avenches	68.9	60.7	19	12	4.43	22.7
Cossonay	61.7	69.3	22	5	2.82	18.7
Echallens	68.3	72.6	18	2	24.20	21.2
Grandson	71.7	34.0	17	8	3.30	20.0
Lausanne	55.7	19.4	26	28	12.11	20.2
La Vallee	54.3	15.2	31	20	2.15	10.8
Lavaux	65.1	73.0	19	9	2.84	20.0
Morges	65.5	59.8	22	10	5.23	18.0
Moudon	65.0	55.1	14	3	4.52	22.4
Nyone	56.6	50.9	22	12	15.14	16.7
Orbe	57.4	54.1	20	6	4.20	15.3
Oron	72.5	71.2	12	1	2.40	21.0
Payerne	74.2	58.1	14	8	5.23	23.8
Paysd'enhaut	72.0	63.5	6	3	2.56	18.0
Rolle	60.5	60.8	16	10	7.72	16.3
Vevey	58.3	26.8	25	19	18.46	20.9
Yverdon	65.4	49.5	15	8	6.10	22.5
Conthey	75.5	85.9	3	2	99.71	15.1
Entremont	69.3	84.9	7	6	99.68	19.8
Herens	77.3	89.7	5	2	100.00	18.3
Martigwy	70.5	78.2	12	6	98.96	19.4
Monthey	79.4	64.9	7	3	98.22	20.2
St Maurice	65.0	75.9	9	9	99.06	17.8
Sierre	92.2	84.6	3	3	99.46	16.3
Sion	79.3	63.1	13	13	96.83	18.1
Boudry	70.4	38.4	26	12	5.62	20.3
La Chauxdfnd	65.7	7.7	29	11	13.79	20.5
Le Locle	72.7	16.7	22	13	11.22	18.9
Neuchatel	64.4	17.6	35	32	16.92	23.0
Val de Ruz	77.6	37.6	15	7	4.97	20.0
ValdeTravers	67.6	18.7	25	7	8.65	19.5
V. De Geneve	35.0	1.2	37	53	42.34	18.0
Rive Droite	44.7	46.6	16	29	50.43	18.2
Rive Gauche	42.8	27.7	22	29	58.33	19.3

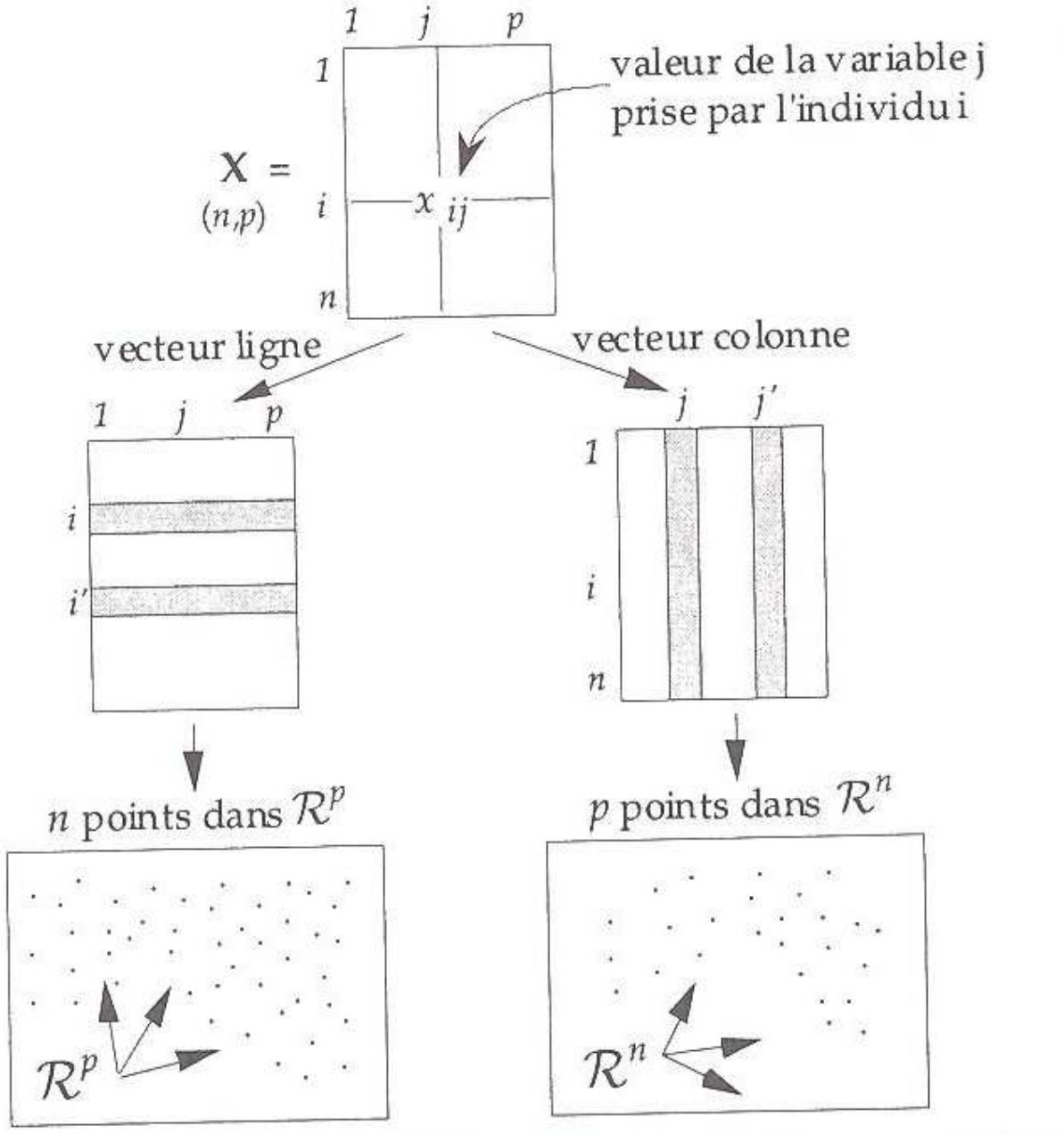
Sailkapen-analisia
(*cluster analysis*)

Helburua:
Sail edo *cluster* horiek bilatu.

Nola lortu helburua?

Objektuen **errepresentazio geometrikoa**:
- Objektu bakoitza puntu bat da dimentsio anitzezko espazio batean.
- Objektuen arteko ezantzekotasuna *distantzia* baten bitartez islatzen da.

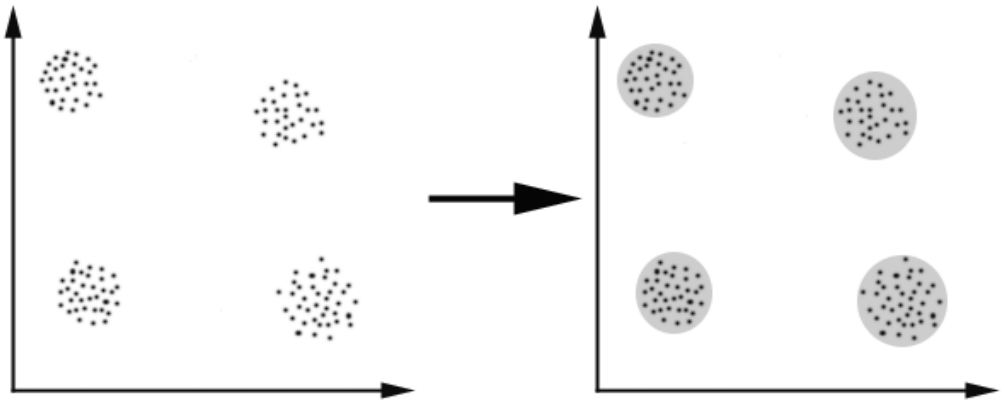
Errepresentazio geometrikoa



Sailkapen-analisia
(*cluster analysis*)

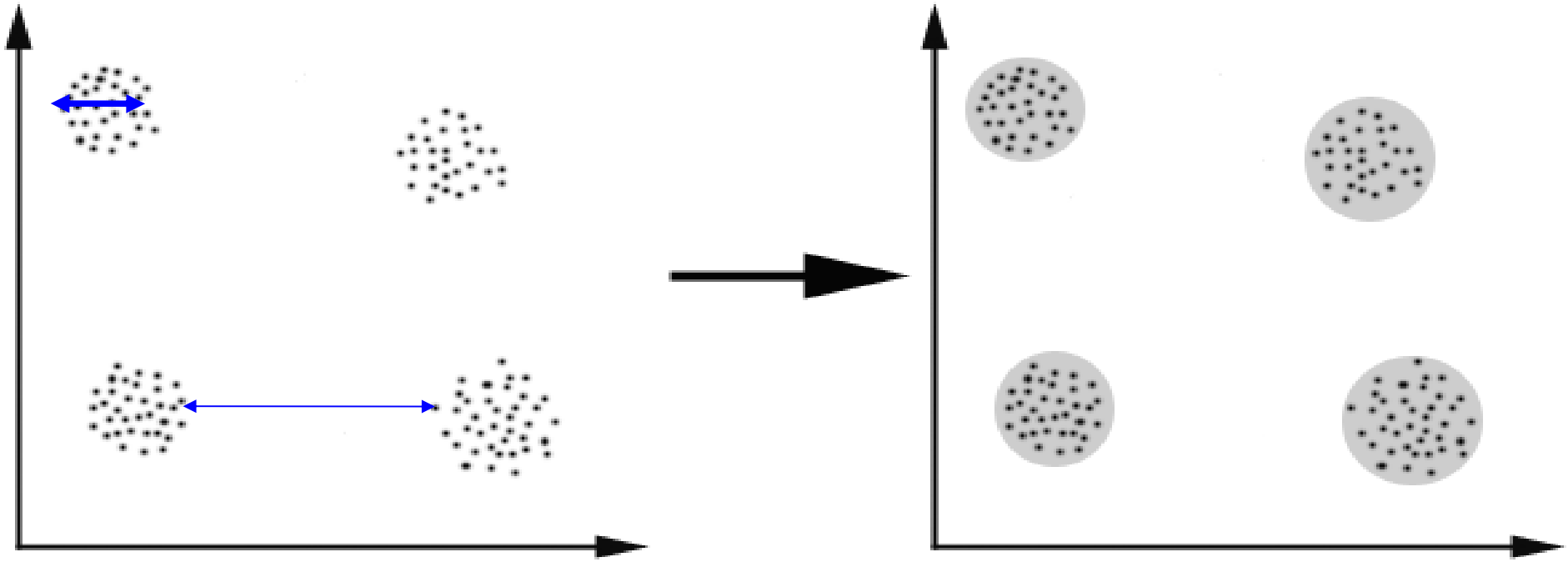
Objektuen **errepresentazio geometrikoa** :
Objektuen arteko *distantzia*: bi objektu are eta antzekoagoak izan, orduan eta gertuago espazioan.

Helburua:
Sail edo *cluster*rak bilatu. Sail bakoitzeko objektuak elkarren ondoan egongo dira eta beste sailetako objektuetik urruti.



Sailkapen-analisia
(*cluster analysis*)

Antzekotasuna edo distantzia



Datuak: Erlazioak

Ω , objektuen multzoa

X_j , aldagaiak

x_{ij} , ω_i objektuaren balioa X_j aldagaian

Ω	X_1		X_j		X_p
ω_1	x_{11}		x_{1j}		x_{1p}
ω_i	x_{i1}		x_{ij}		x_{ip}
ω_n	x_{n1}		x_{nj}		x_{np}

P^1	P^2		P^1
p_1^1	p_1^2		p_1^1
p_i^1	p_i^2		p_{il}
p_n^1	p_n^2		p_n^1

Miatze eta *aztertzearen* ondorioz **emaitzak**:

Zenbait erlazio berri

Suitza, 1888: 47 herrialde frankofono

Zenbait adierazle sozioekonomiko (%) eta emankortasuna

X1 Fertility X2 Agriculture X3 Examination **C Clustering**
X4 Education X5 Catholic X6 Infant.Mortality

	X1	X2	X3	X4	X5	X6	C
Courtelary	80.2	17.0	15	12	9.96	22.2	?
Delemont	83.1	45.1	6	9	84.84	22.2	?
Franches-Mnt	92.5	39.7	5	5	93.40	20.2	?
Moutier	85.8	36.5	12	7	33.77	20.3	?
Neuveville	76.9	43.5	17	15	5.16	20.6	?
Porrentruy	76.1	35.3	9	7	90.57	26.6	?
Broye	83.8	70.2	16	7	92.85	23.6	?
Glane	92.4	67.8	14	8	97.16	24.9	?
Gruyere	82.4	53.3	12	7	97.67	21.0	?
Sarine	82.9	45.2	16	13	91.38	24.4	?
Veveyse	87.1	64.5	14	6	98.61	24.5	?
Aigle	64.1	62.0	21	12	8.52	16.5	?
Aubonne	66.9	67.5	14	7	2.27	19.1	?
Avenches	68.9	60.7	19	12	4.43	22.7	?
Cossonay	61.7	69.3	22	5	2.82	18.7	?
Echallens	68.3	72.6	18	2	24.20	21.2	?
Grandson	71.7	34.0	17	8	3.30	20.0	?
Lausanne	55.7	19.4	26	28	12.11	20.2	?
La Vallee	54.3	15.2	31	20	2.15	10.8	?
Lavaux	65.1	73.0	19	9	2.84	20.0	?
Morges	65.5	59.8	22	10	5.23	18.0	?
Moudon	65.0	55.1	14	3	4.52	22.4	?
Nyone	56.6	50.9	22	12	15.14	16.7	?
Orbe	57.4	54.1	20	6	4.20	15.3	?
Oron	72.5	71.2	12	1	2.40	21.0	?
Payerne	74.2	58.1	14	8	5.23	23.8	?
Paysd'enhaut	72.0	63.5	6	3	2.56	18.0	?
Rolle	60.5	60.8	16	10	7.72	16.3	?
Vevey	58.3	26.8	25	19	18.46	20.9	?
Yverdon	65.4	49.5	15	8	6.10	22.5	?
Conthey	75.5	85.9	3	2	99.71	15.1	?
Entremont	69.3	84.9	7	6	99.68	19.8	?
Herens	77.3	89.7	5	2	100.00	18.3	?
Martigwy	70.5	78.2	12	6	98.96	19.4	?
Monthey	79.4	64.9	7	3	98.22	20.2	?
St Maurice	65.0	75.9	9	9	99.06	17.8	?
Sierre	92.2	84.6	3	3	99.46	16.3	?
Sion	79.3	63.1	13	13	96.83	18.1	?
Boudry	70.4	38.4	26	12	5.62	20.3	?
La Chauxdfnd	65.7	7.7	29	11	13.79	20.5	?
Le Locle	72.7	16.7	22	13	11.22	18.9	?
Neuchatel	64.4	17.6	35	32	16.92	23.0	?
Val de Ruz	77.6	37.6	15	7	4.97	20.0	?
ValdeTravers	67.6	18.7	25	7	8.65	19.5	?
V. De Geneve	35.0	1.2	37	53	42.34	18.0	?
Rive Droite	44.7	46.6	16	29	50.43	18.2	?
Rive Gauche	42.8	27.7	22	29	58.33	19.3	?

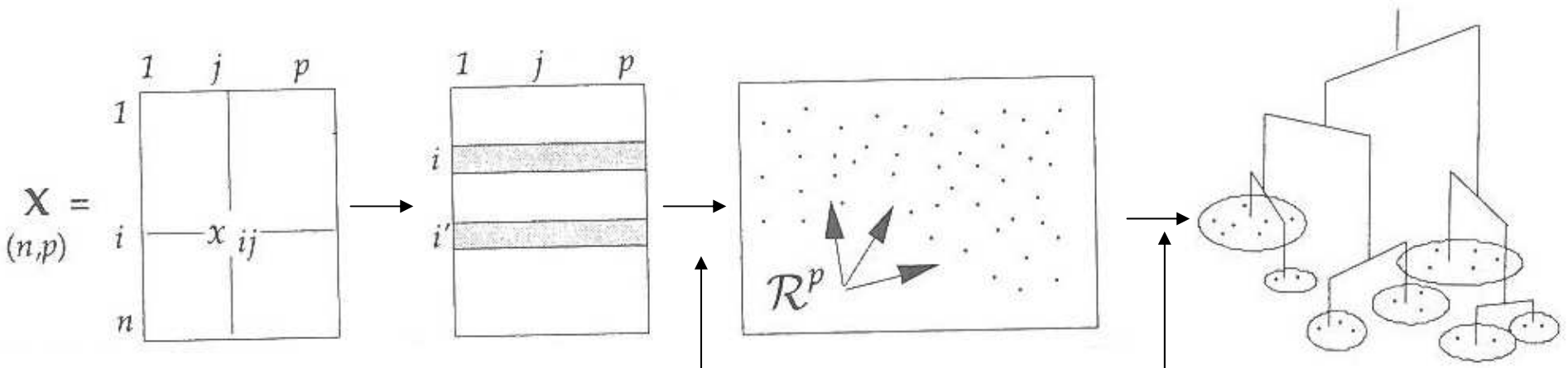
Suitza, 1888: 47 herrialde frankofono

Zenbait adierazle sozioekonomiko (%) eta emankortasuna

X1 Fertility X2 Agriculture X3 Examination
X4 Education X5 Catholic X6 Infant.Mortality

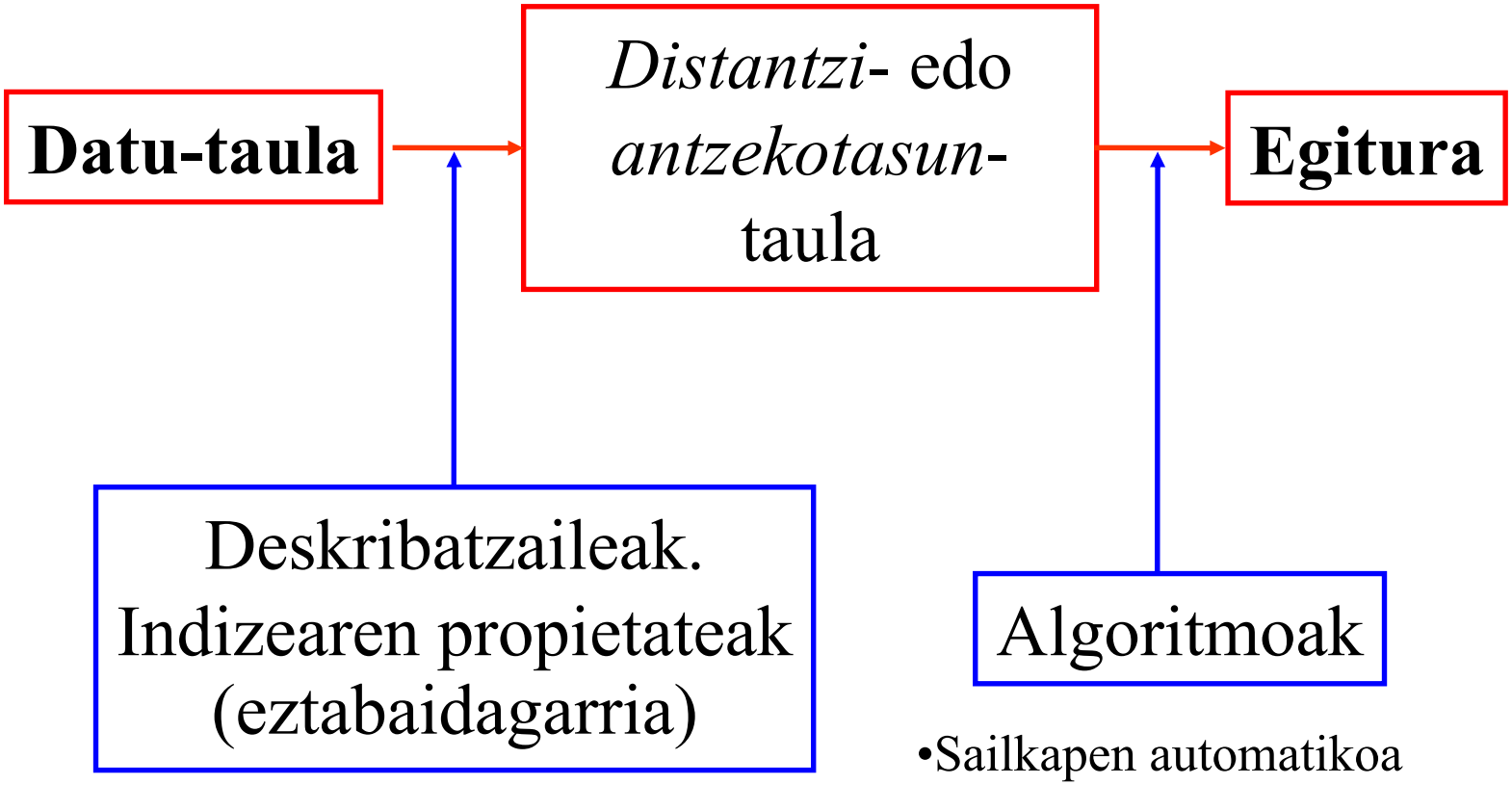
	X1	X2	X3	X4	X5	X6
Courtelary	80.2	17.0	15	12	9.96	22.2
Delemont	83.1	45.1	6	9	84.84	22.2
Franches-Mnt	92.5	39.7	5	5	93.40	20.2
Moutier	85.8	36.5	12	7	33.77	20.3
Neuveville	76.9	43.5	17	15	5.16	20.6
Porrentruy	76.1	35.3	9	7	90.57	26.6
Broye	83.8	70.2	16	7	92.85	23.6
Glane	92.4	67.8	14	8	97.16	24.9
Gruyere	82.4	53.3	12	7	97.67	21.0
Sarine	82.9	45.2	16	13	91.38	24.4
Veveyse	87.1	64.5	14	6	98.61	24.5
Aigle	64.1	62.0	21	12	8.52	16.5
Aubonne	66.9	67.5	14	7	2.27	19.1
Avenches	68.9	60.7	19	12	4.43	22.7
Cossonay	61.7	69.3	22	5	2.82	18.7
Echallens	68.3	72.6	18	2	24.20	21.2
Grandson	71.7	34.0	17	8	3.30	20.0
Lausanne	55.7	19.4	26	28	12.11	20.2
La Vallee	54.3	15.2	31	20	2.15	10.8
Lavaux	65.1	73.0	19	9	2.84	20.0
Morges	65.5	59.8	22	10	5.23	18.0
Moudon	65.0	55.1	14	3	4.52	22.4
Nyone	56.6	50.9	22	12	15.14	16.7
Orbe	57.4	54.1	20	6	4.20	15.3
Oron	72.5	71.2	12	1	2.40	21.0
Payerne	74.2	58.1	14	8	5.23	23.8
Paysd'enhaut	72.0	63.5	6	3	2.56	18.0
Rolle	60.5	60.8	16	10	7.72	16.3
Vevey	58.3	26.8	25	19	18.46	20.9
Yverdon	65.4	49.5	15	8	6.10	22.5
Conthey	75.5	85.9	3	2	99.71	15.1
Entremont	69.3	84.9	7	6	99.68	19.8
Herens	77.3	89.7	5	2	100.00	18.3
Martigwy	70.5	78.2	12	6	98.96	19.4
Monthey	79.4	64.9	7	3	98.22	20.2
St Maurice	65.0	75.9	9	9	99.06	17.8
Sierre	92.2	84.6	3	3	99.46	16.3
Sion	79.3	63.1	13	13	96.83	18.1
Boudry	70.4	38.4	26	12	5.62	20.3
La Chauxdfnd	65.7	7.7	29	11	13.79	20.5
Le Locle	72.7	16.7	22	13	11.22	18.9
Neuchatel	64.4	17.6	35	32	16.92	23.0
Val de Ruz	77.6	37.6	15	7	4.97	20.0
ValdeTravers	67.6	18.7	25	7	8.65	19.5
V. De Geneve	35.0	1.2	37	53	42.34	18.0
Rive Droite	44.7	46.6	16	29	50.43	18.2
Rive Gauche	42.8	27.7	22	29	58.33	19.3

Sailkapen-analisia
(*cluster analysis*)



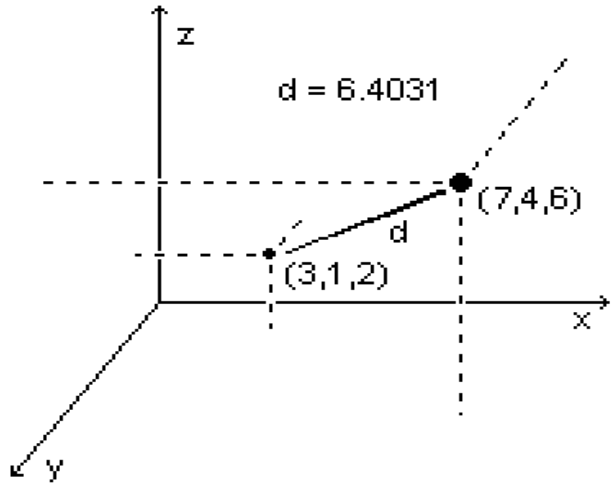
Analisi baten emaitzak hauen menpe daude:

- Eskuratutako datu-multzoan eta beren prozesamenduan.
- Objektuak beren artean alderatzeko erabili den distantzia edo antzekotasun-neurrian.
- Erabilitako algoritmoan.



Distantzia

Datuak errepresentatzen dituzten bektoreen osagaiak unitate berberaz baldin badaude adierazita (aldagai homogeneoak), *distantzia euklidestar*ra nahikoa izan daiteke sailkatze-lanak egiteko, bestela (aldagai heterogeneoak) ez.



$$d = \sqrt{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2}$$

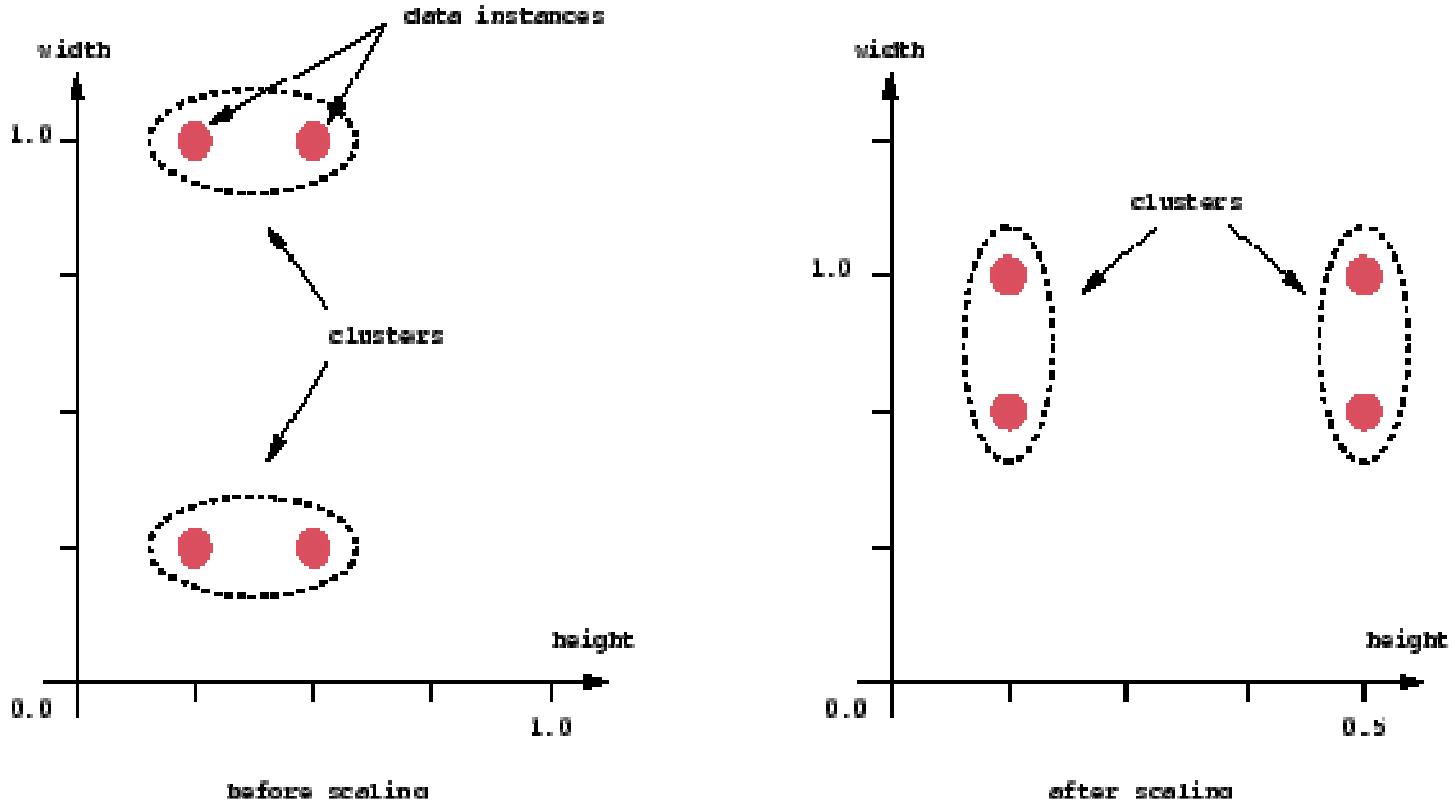
$$P = (p_1, p_2, \ldots, p_n) \qquad Q = (q_1, q_2, \ldots, q_n).$$

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}$$

Distantzia

Aldagaiak heterogeneoak badira ere, batzutan ez da komenigarri izaten distantzia euklidestarra zuzenean erabiltzea.

Aldagaien eskala ezberdinak sailkatze ezberdinak eman dezakete.



Distantzia

Arazoa neurri-bektore baten osagaiak formula batean konbinatzeko eran datza, gero *clustering*eko algoritmo batean erabilia izan dadin: nola aukeratu distantzia?

Erantzuna arazoaren arloaren ezagutzan aurkitzen da.

Distantzia bat aukeratzeko ez dago gida teorikorik.

Minkowski-ren distantzia orokorra:

$$d_p(x_i, x_j) = \left(\sum_{k=1}^{d} \left| x_{i,k} - x_{j,k} \right|^p \right)^{\frac{1}{p}}$$

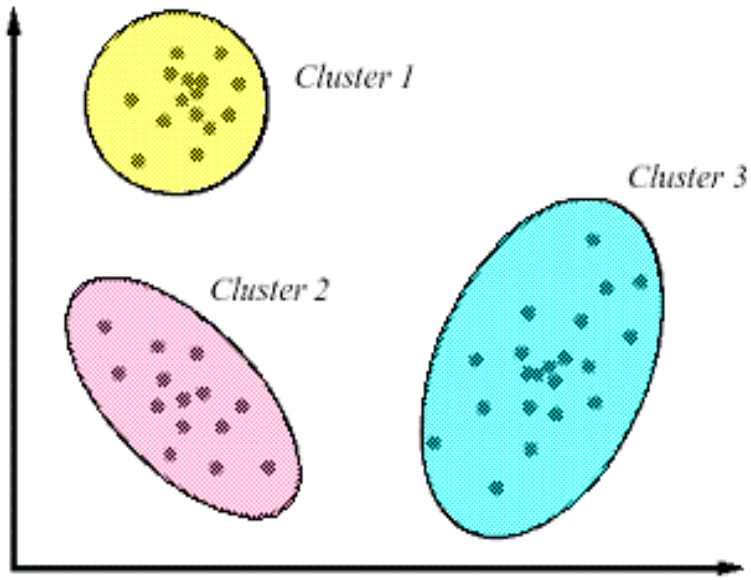
*p=2, Distantzia euklides*tarra.

p=1, Hiriko blokeen distantzia (*Manhattan* distantzia.

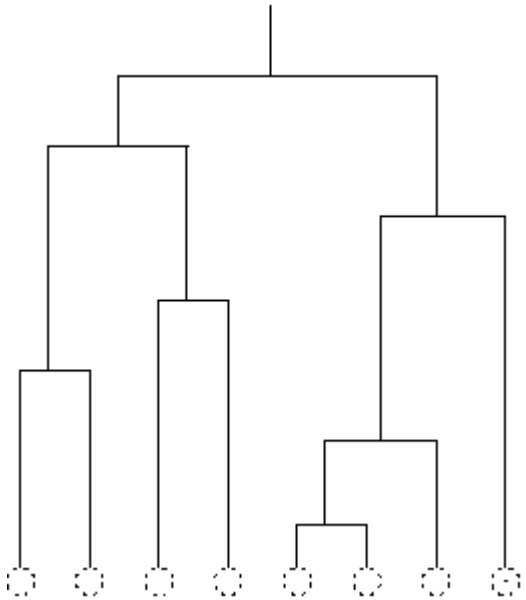
Algoritmoak

*Clustering*eko algoritmo gehienak bitan bana daitezke, ematen duten *egitura*ren arabera:

*Partiketa*rako algoritmoak



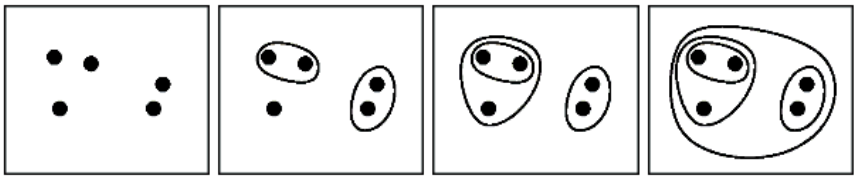
*Hierarkia*rako algoritmoak

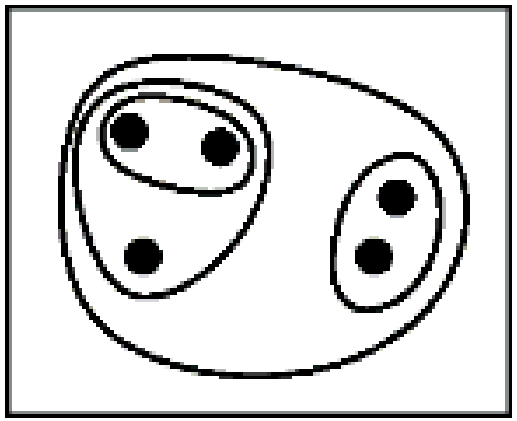


Algoritmoak

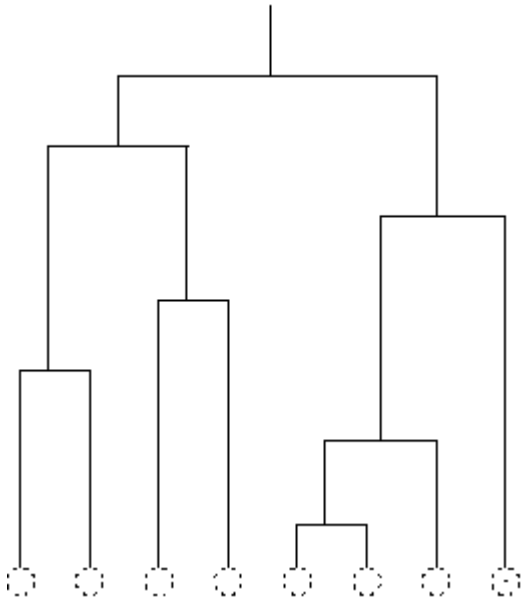
*Hierarkia*rako errepresentazioak:

Venn-en *diagramak*



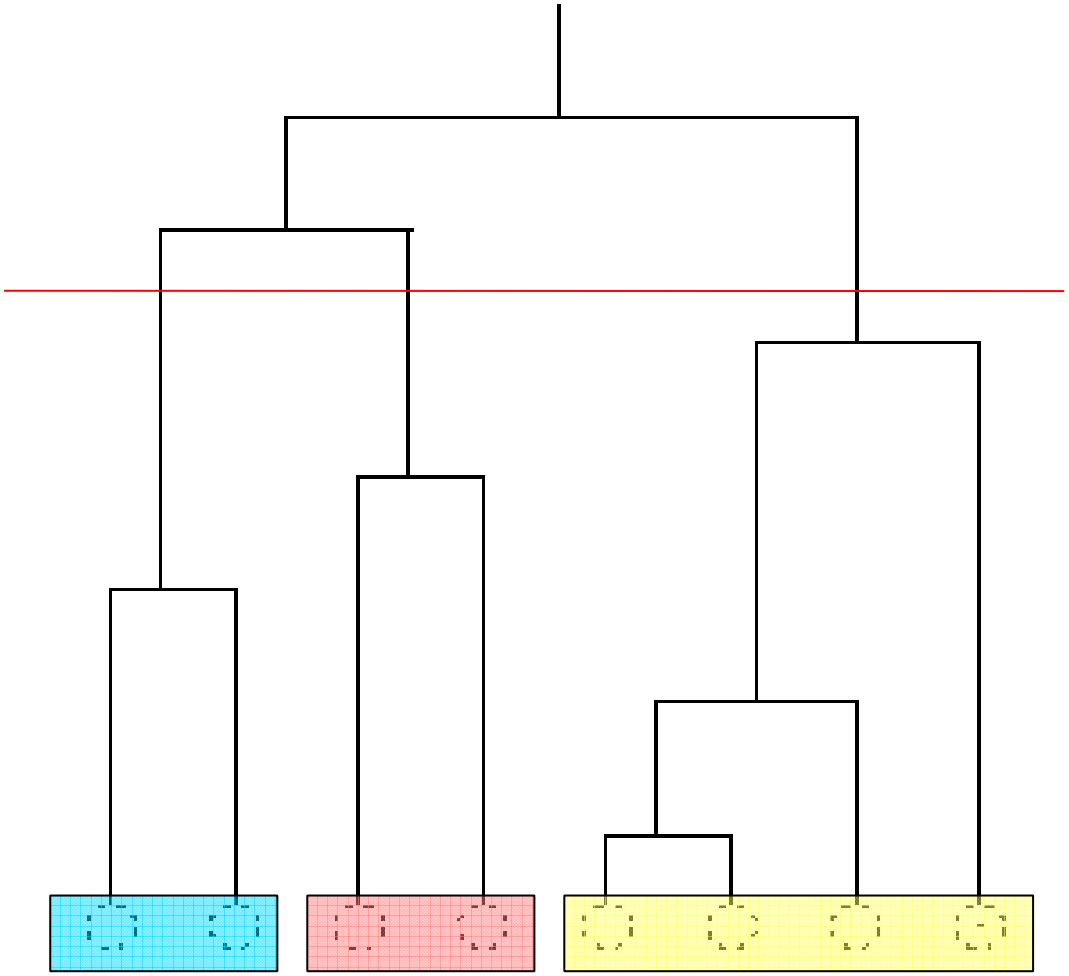


Dendrograma



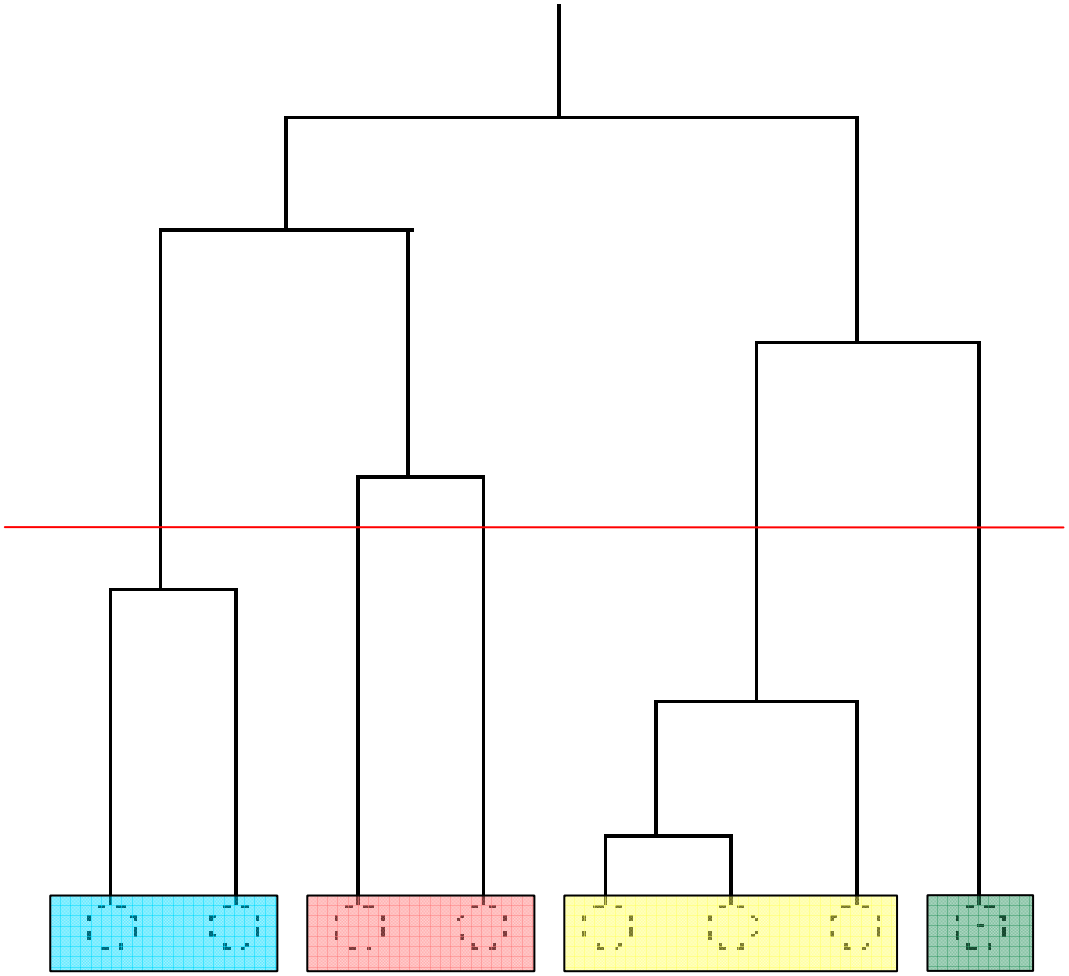
Algoritmoak

Hierarkia-rako algoritmoek *partiketa*-sorta bat ematen dute:



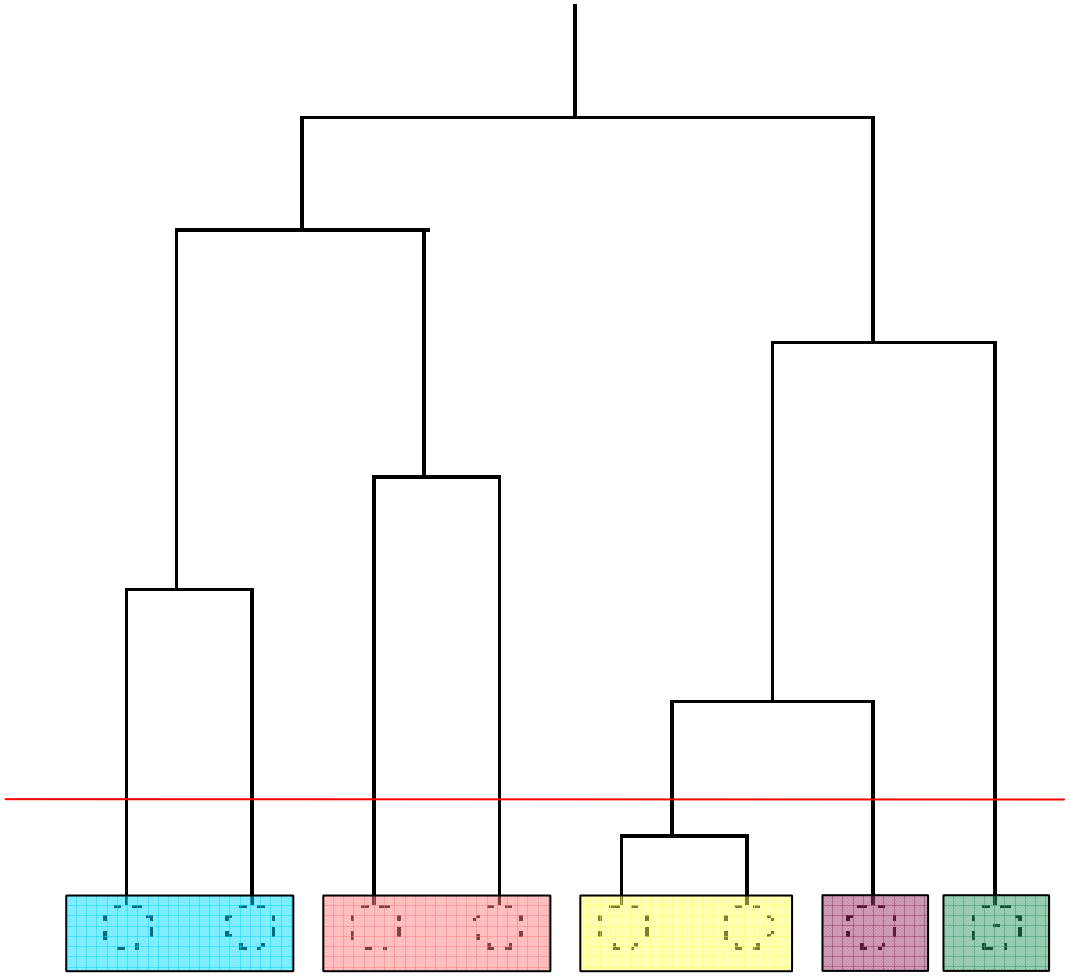
Algoritmoak

Hierarkia-rako algoritmoek *partiketa*-sorta bat ematen dute:



Algoritmoak

Hierarkia-rako algoritmoek *partiketa*-sorta bat ematen dute:



Partiketarako algoritmoa: k-means

Dimentsio anitzezko espazio batean errepresentatuta dauden *n* objektu emanez gero, eta eraiki behar diren *k* sail edo *cluster*-kopurua finkatuta, **helburua** objektuen beren saileko erdigunera (c_i, *means, batezbestekoak*) duten distantzia karratuen batura minimizatzea da.

$$\Sigma_i \Sigma_{x \in X_i} d^2(x, c_i)$$

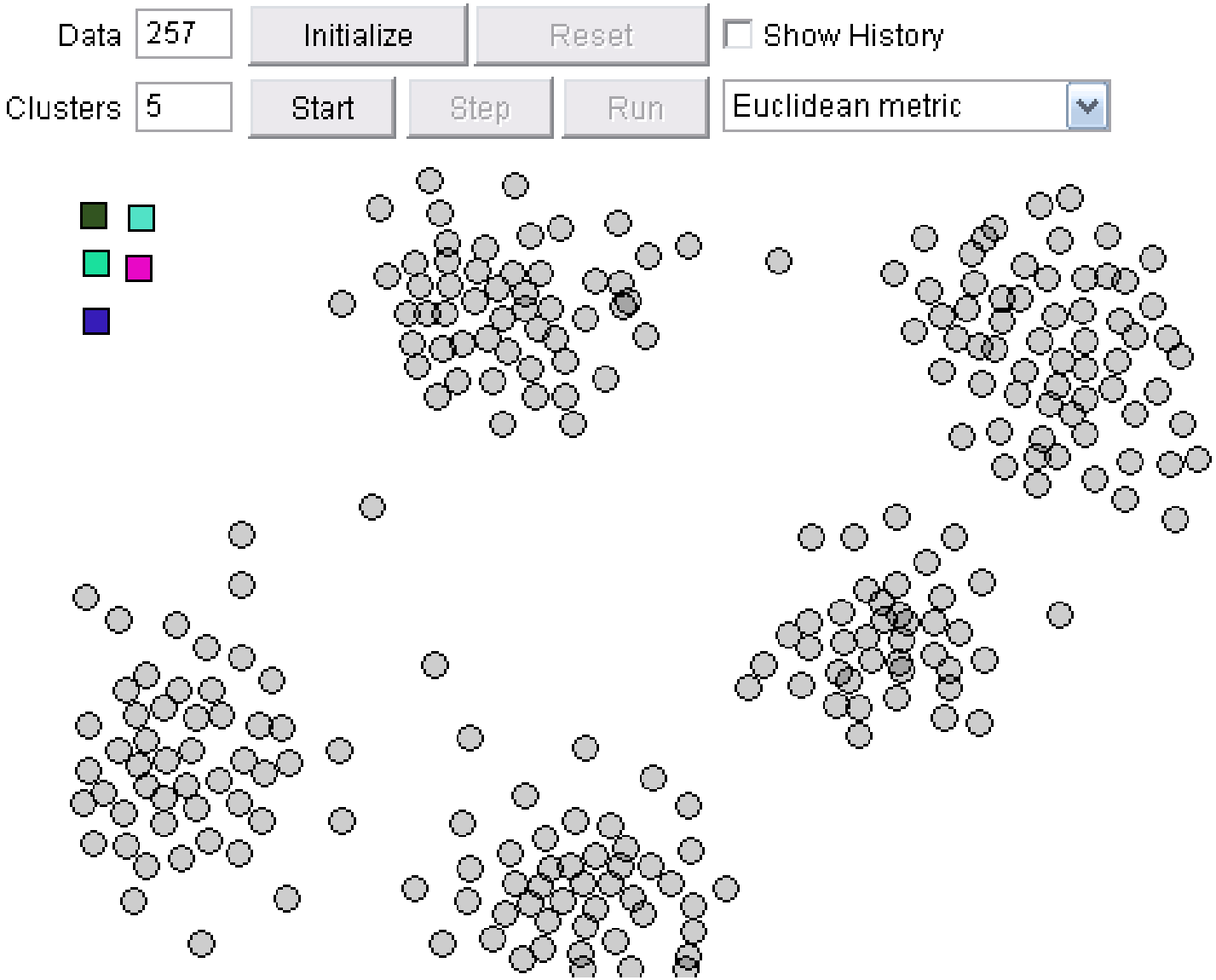
Partiketarako algoritmoa: k-means

1. Finkatu *cluster*-kopurua: *k*.
2. Aukeratu *k cluster*ren ‘haziak’.
3. Objektu bat gertuen daukan ‘hazia’ren sailekoa egiten da.
4. *k cluster*retan ‘hazi’ berriak lortzen dira (saileko batezbestekoak).
5. 3. eta 4. urratsetan aritu *cluster*rak egonkortu bitartean (‘haziak‘ ez dira aldatzen).

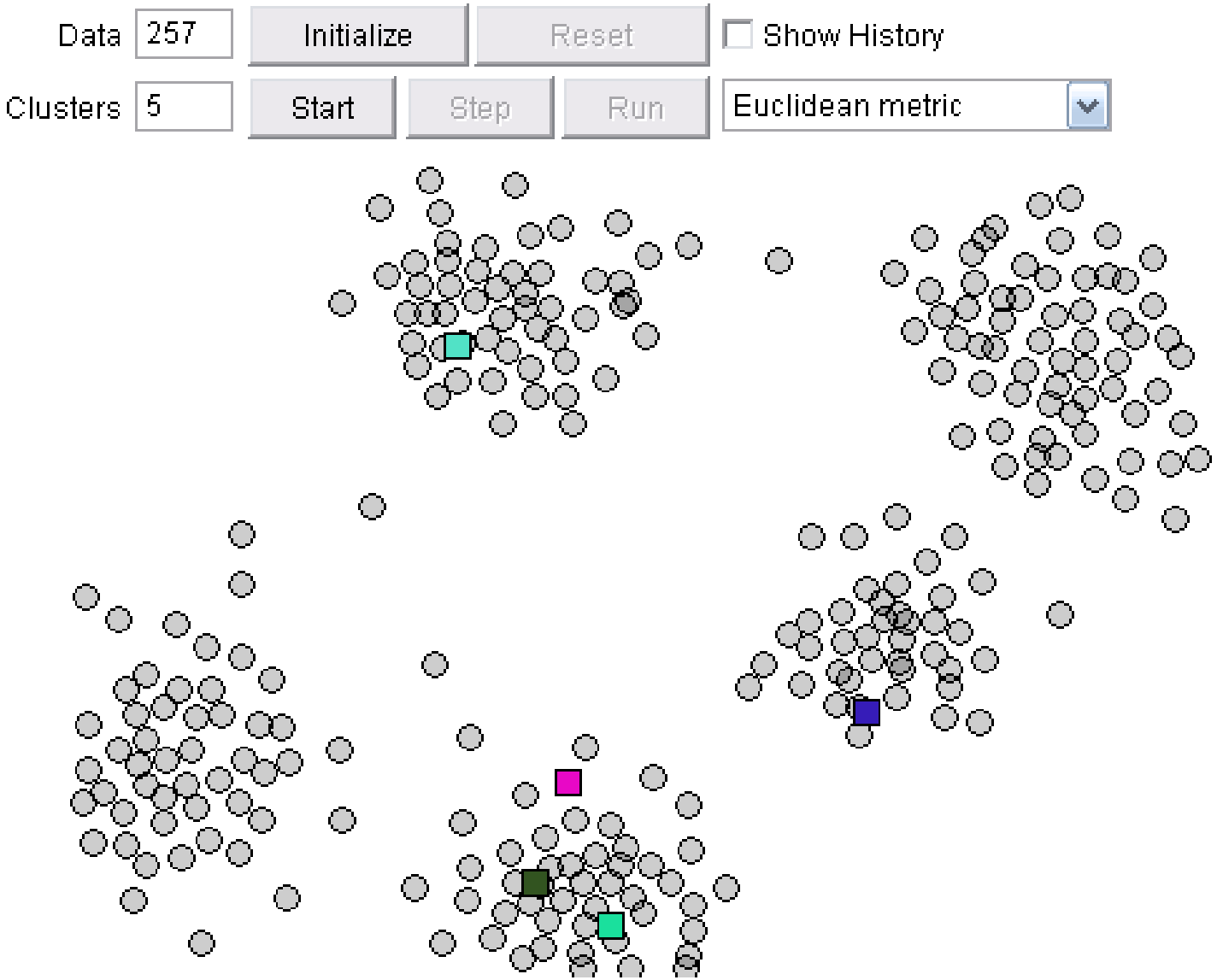
Partiketarako algoritmoa: k-means

- 1. urratsean ez dago argi zein izan behar duen *k*-ren balioa.
- 2.ean ‘haziak’ zoriz aukeratzen dira objektuen artetik, edo errepresentazio-espazioan, edo, informaziorik badago erabitzaileak erabakita.
- 3.ean *distantzia* definitu behar da: *distantzia euklidestar*ra da ohikoena.
- 3.ean, objektu guztiak bat-batean esleitu beharrean, gero *means* edo ‘hazi’ berriak kalkulatzearren, pausoz pauso berritu liteke objektu bakoitza *cluster* batean esleitzean.
- 4.ean ‘hazi’aren kontzeptua zabalagoa izan liteke: ‘nukleoa’; puntu-azpimultzo bat, zuzen bat, plano bat, eta abar.

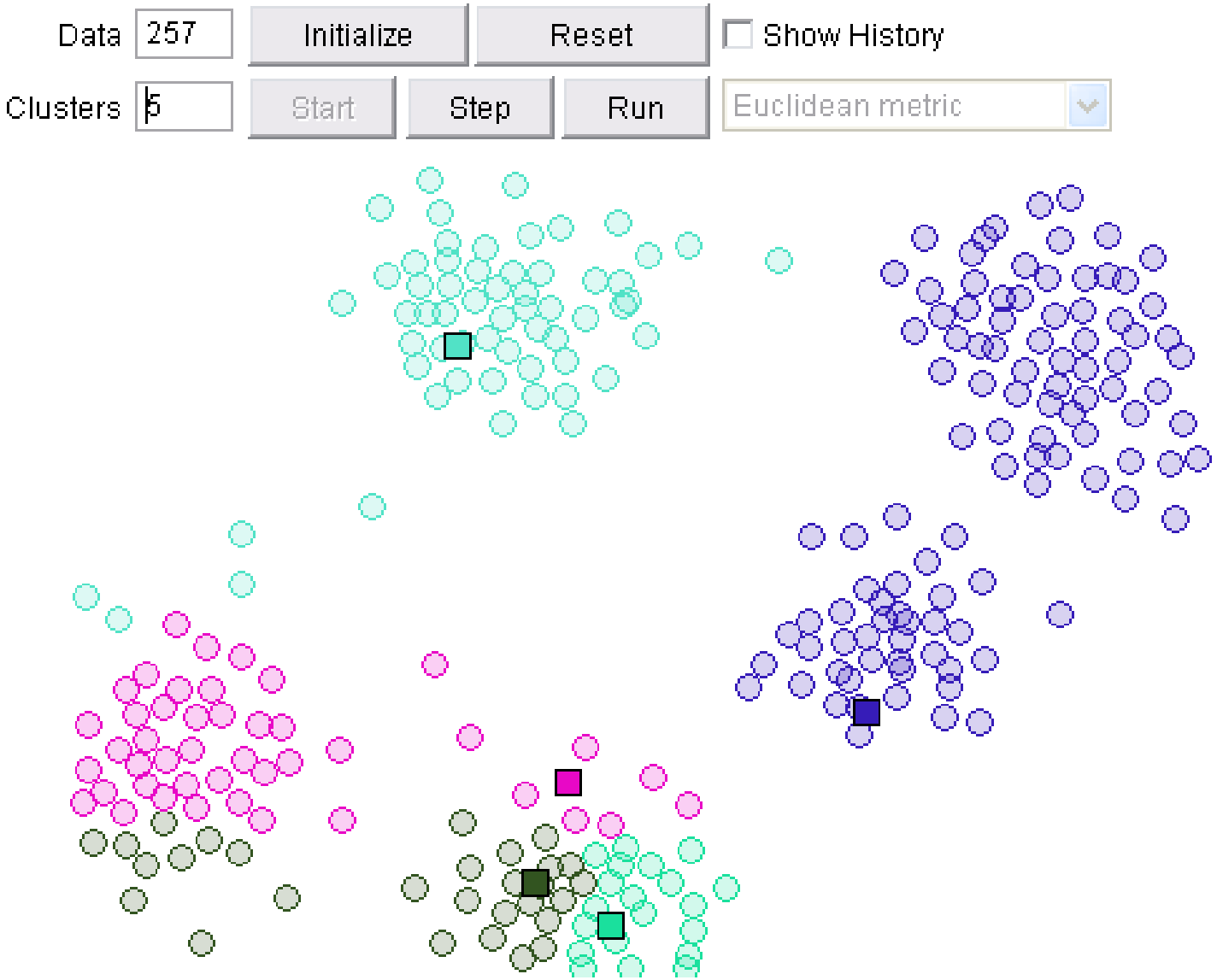
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



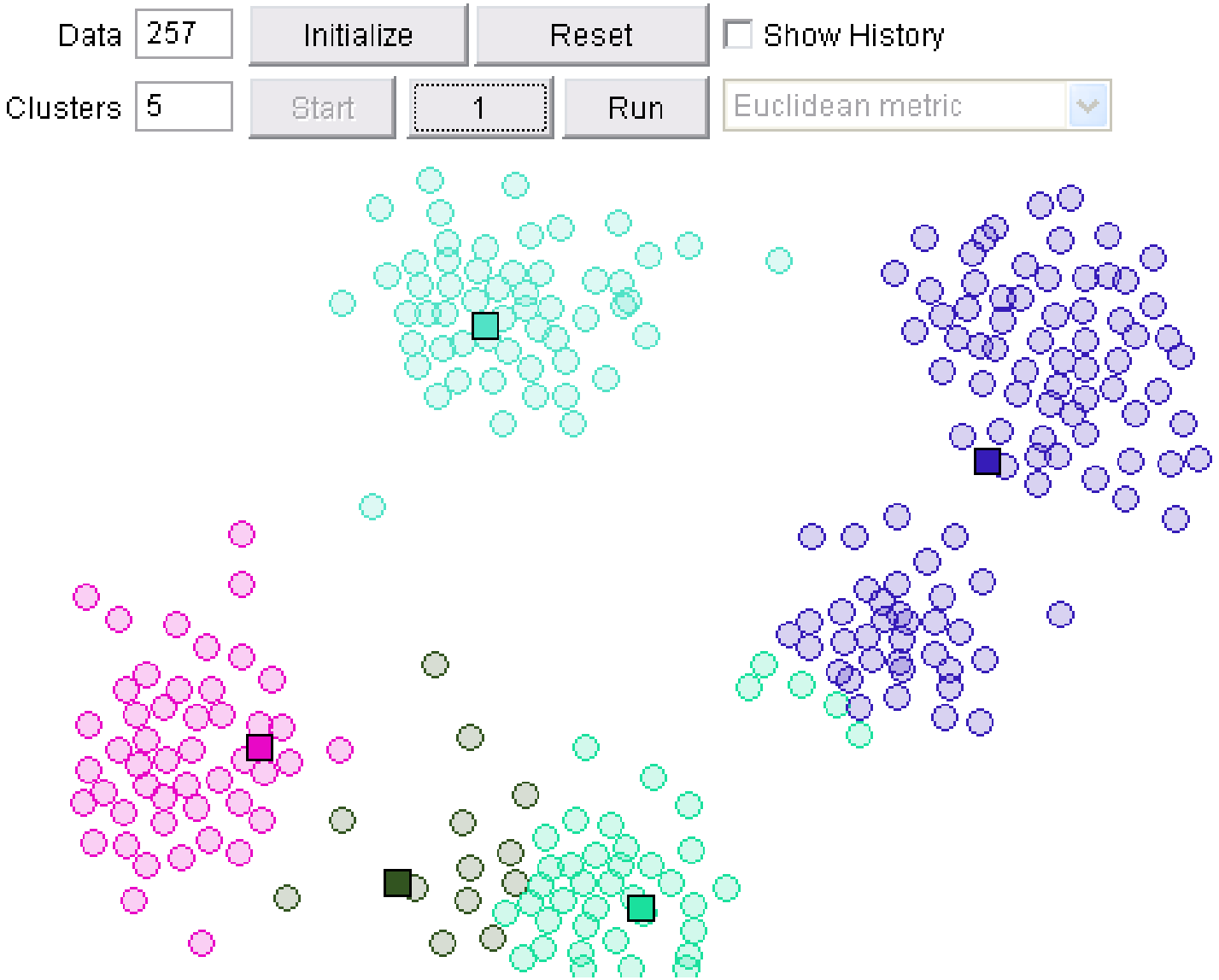
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



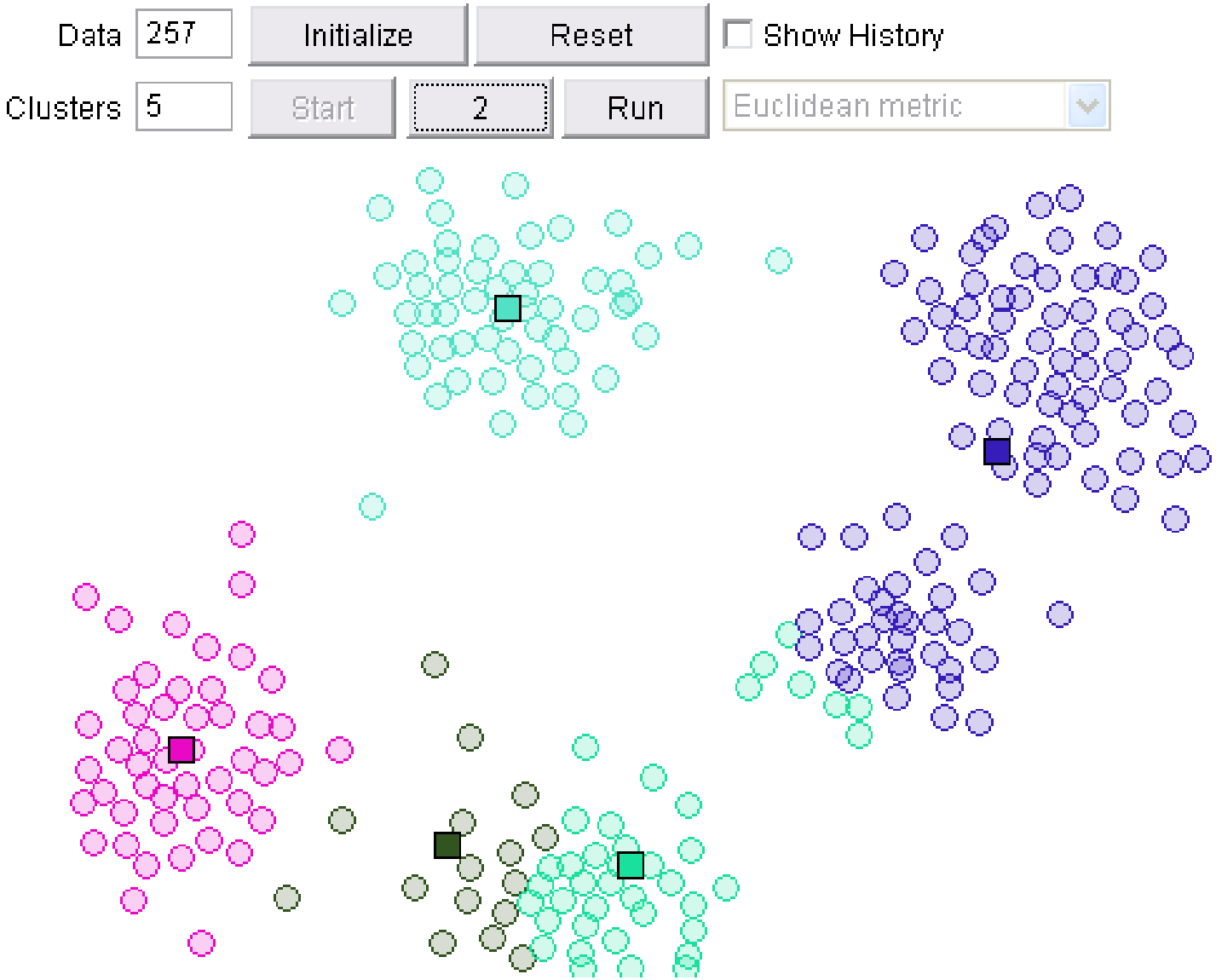
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



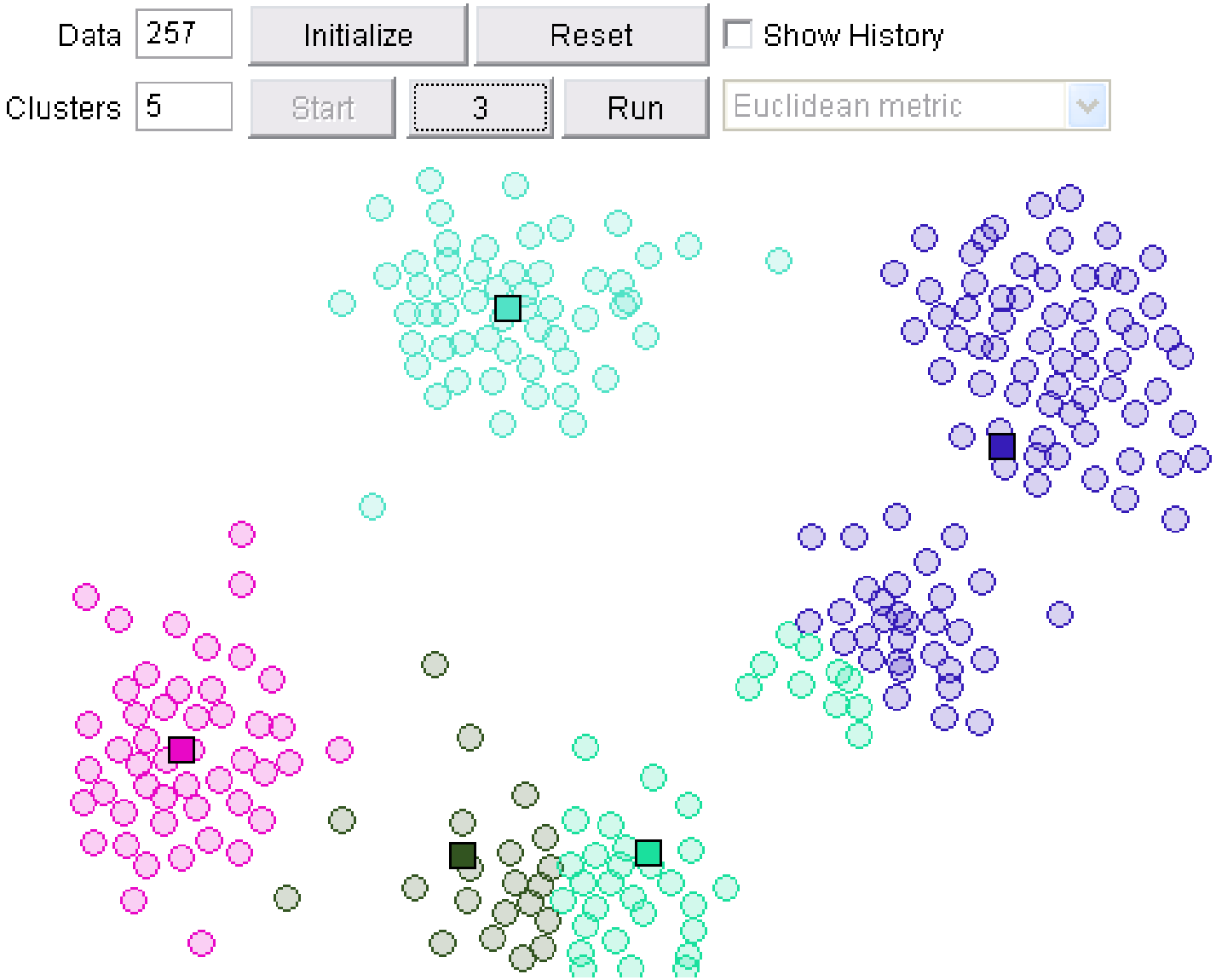
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



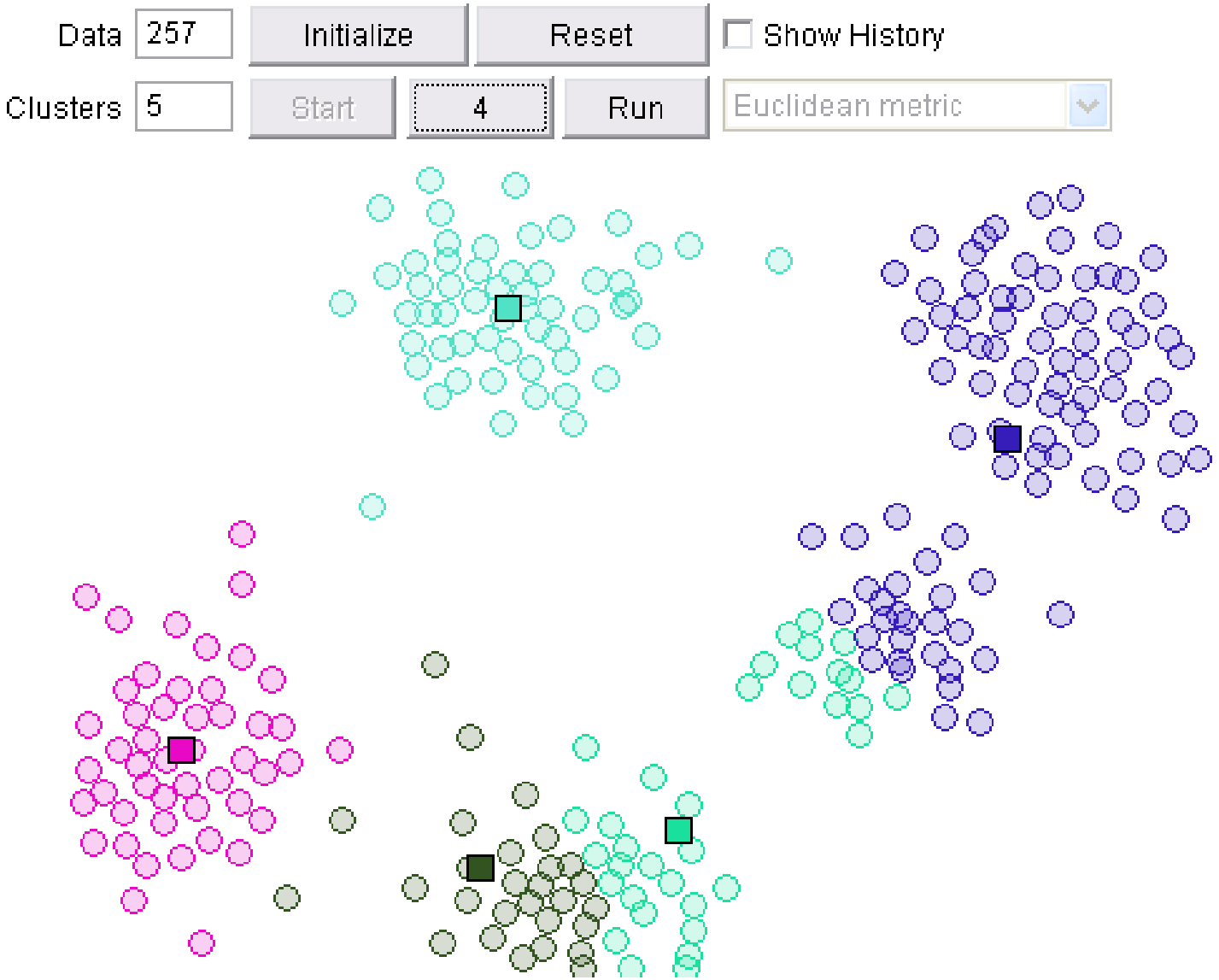
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



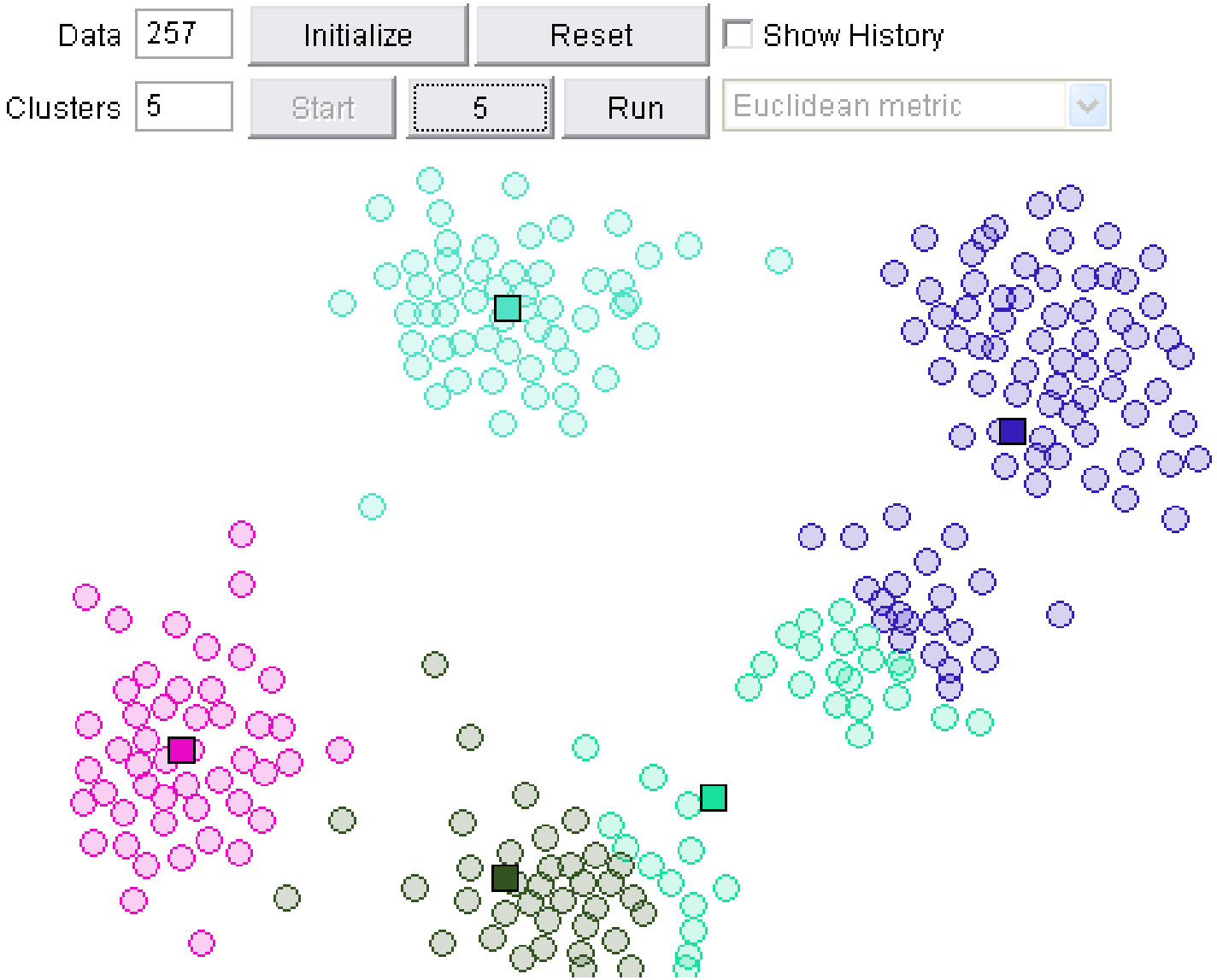
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



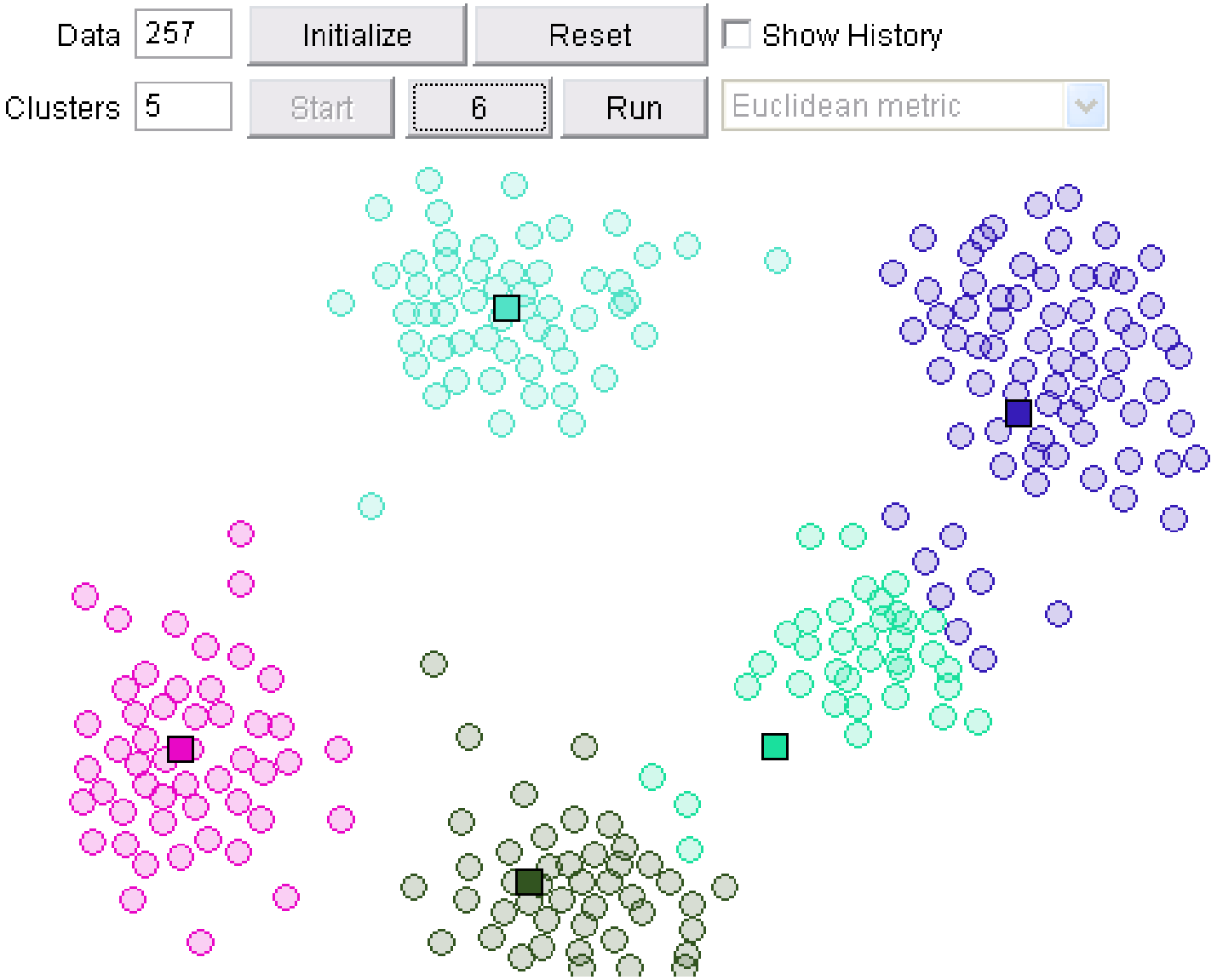
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



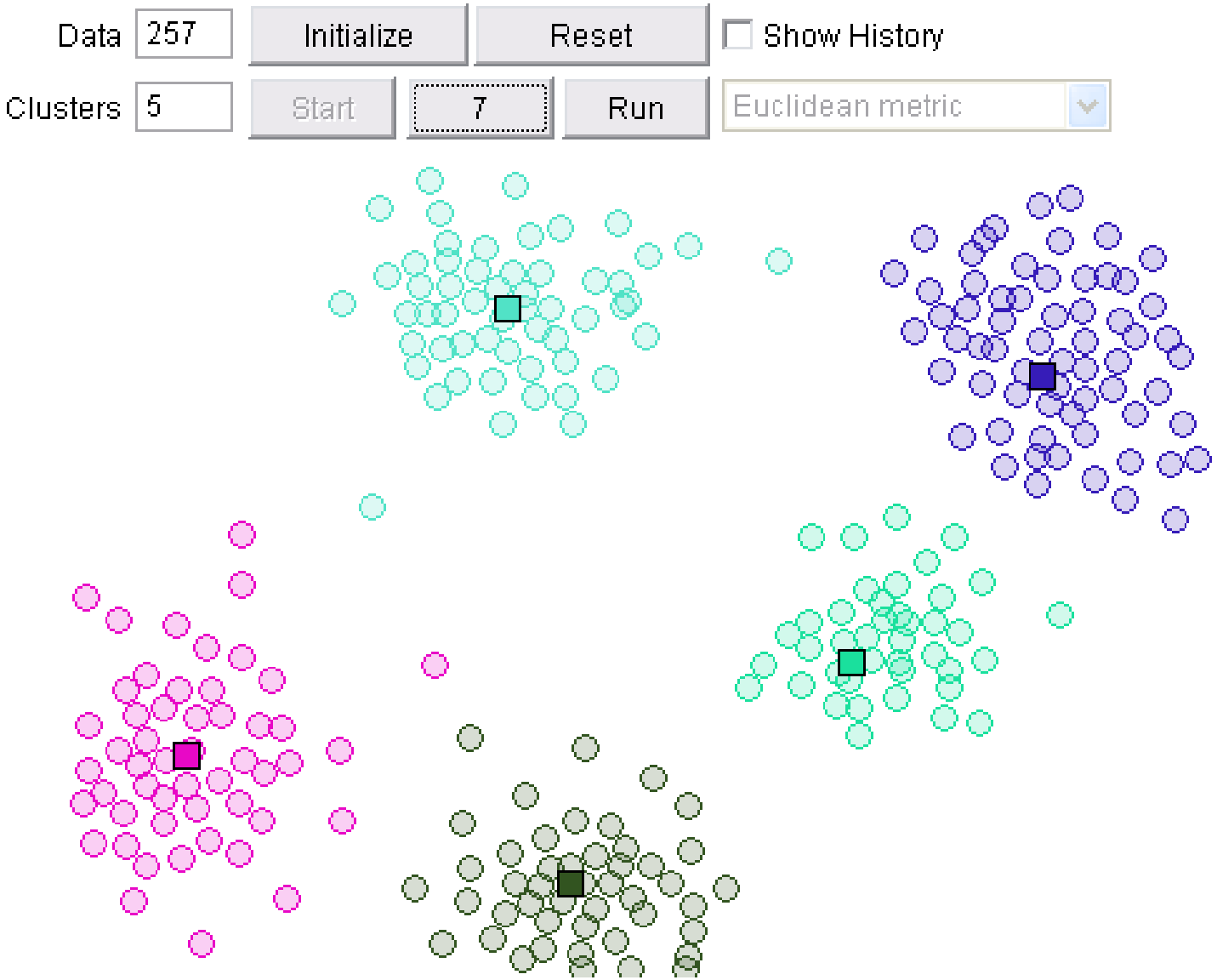
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



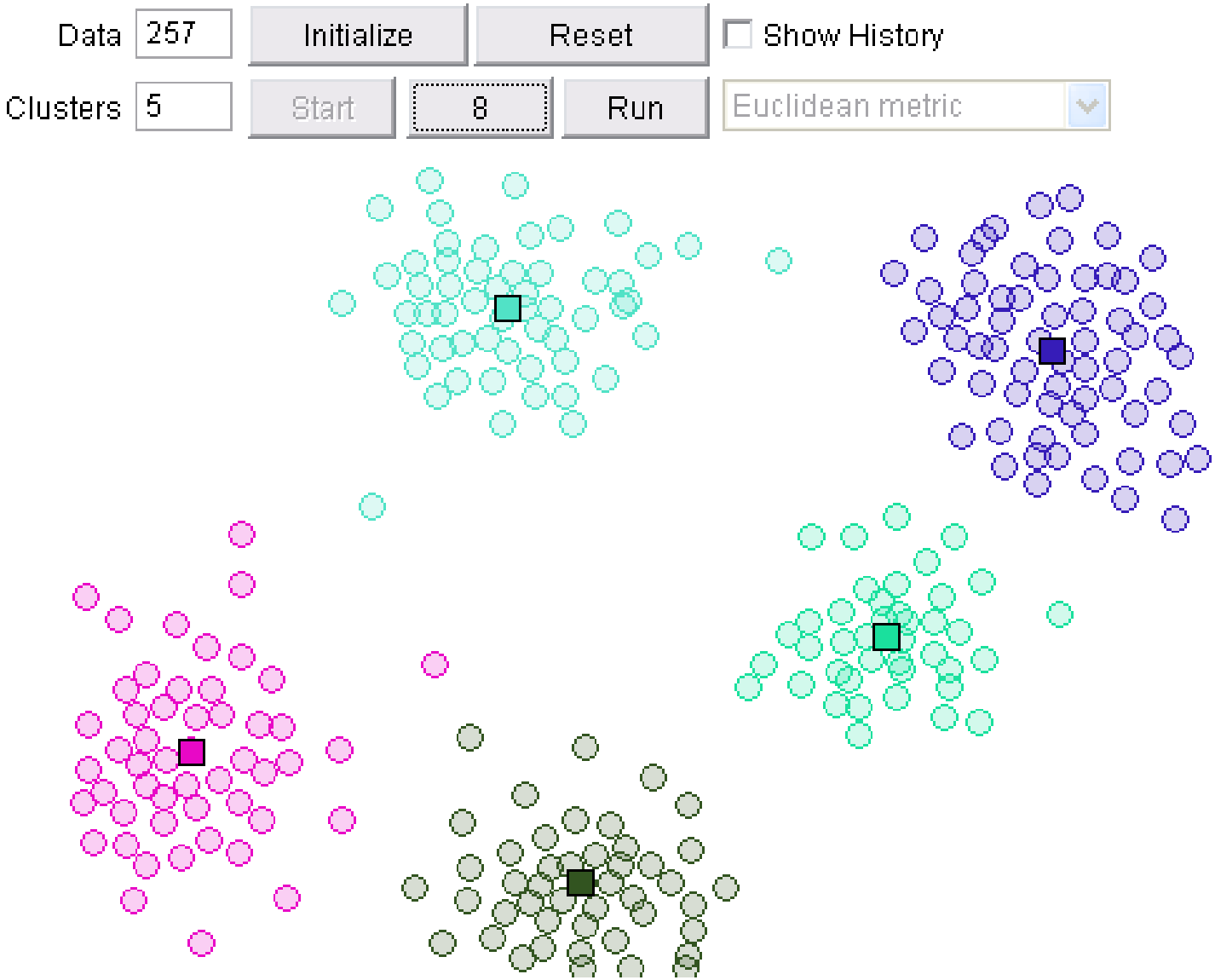
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



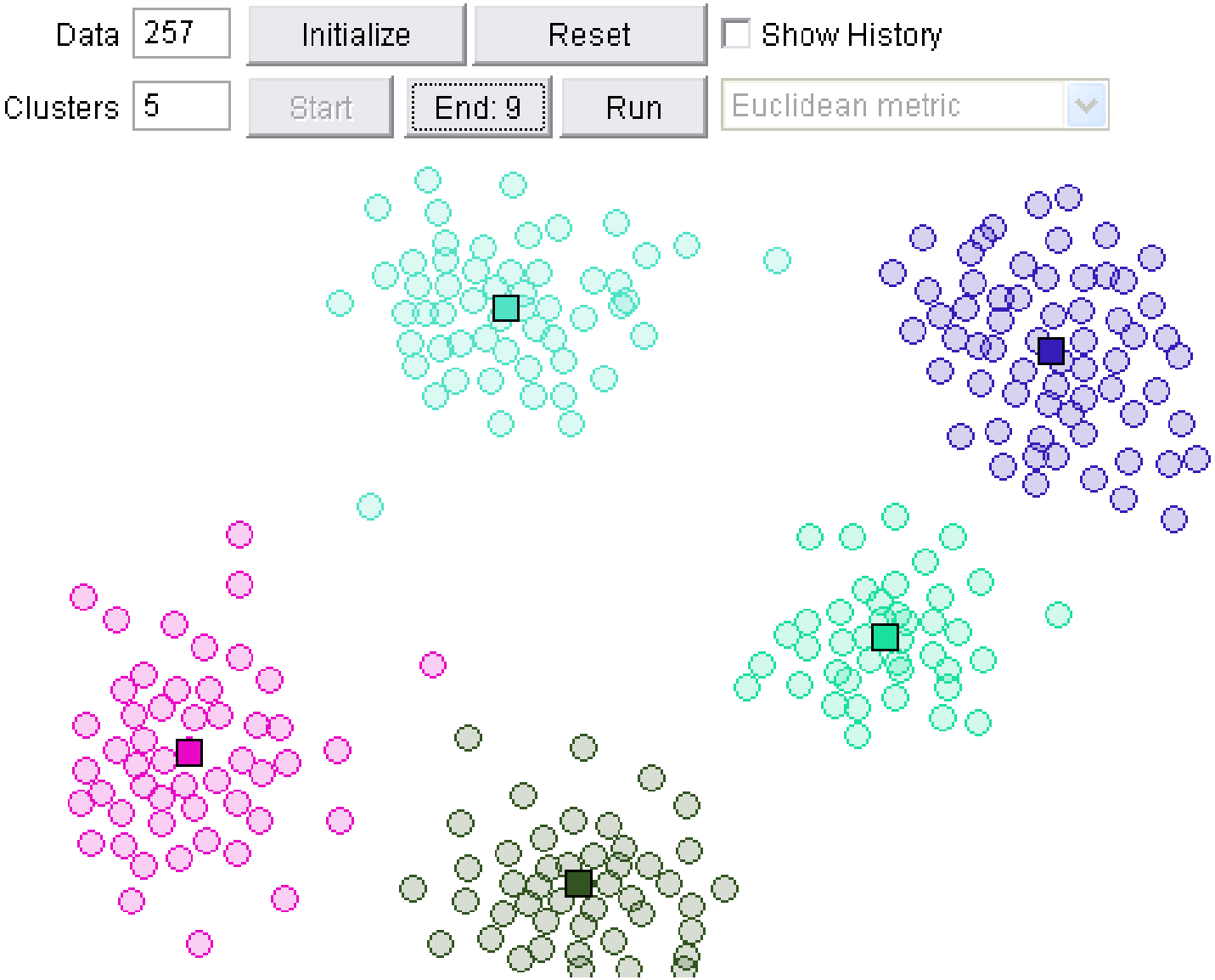
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



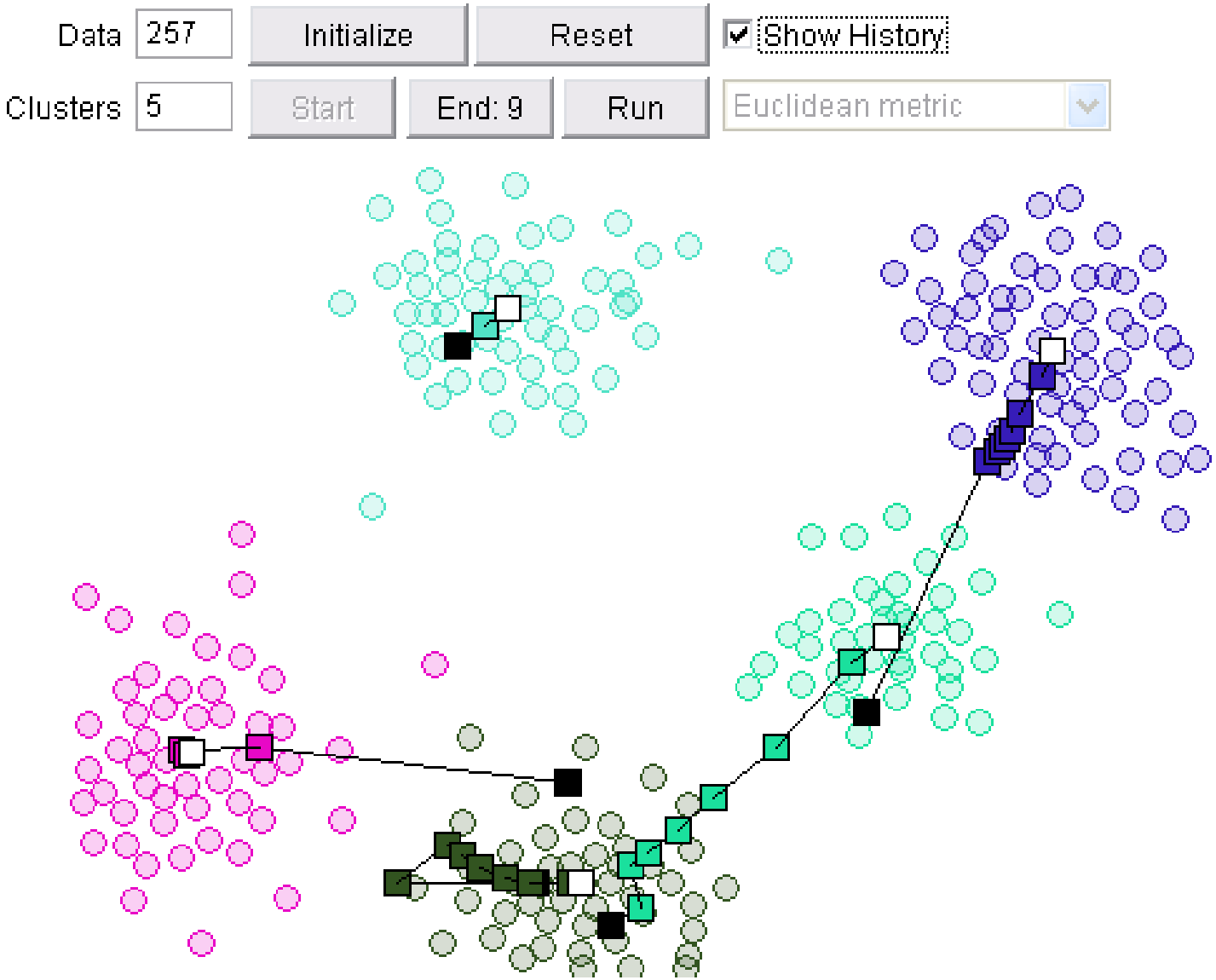
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



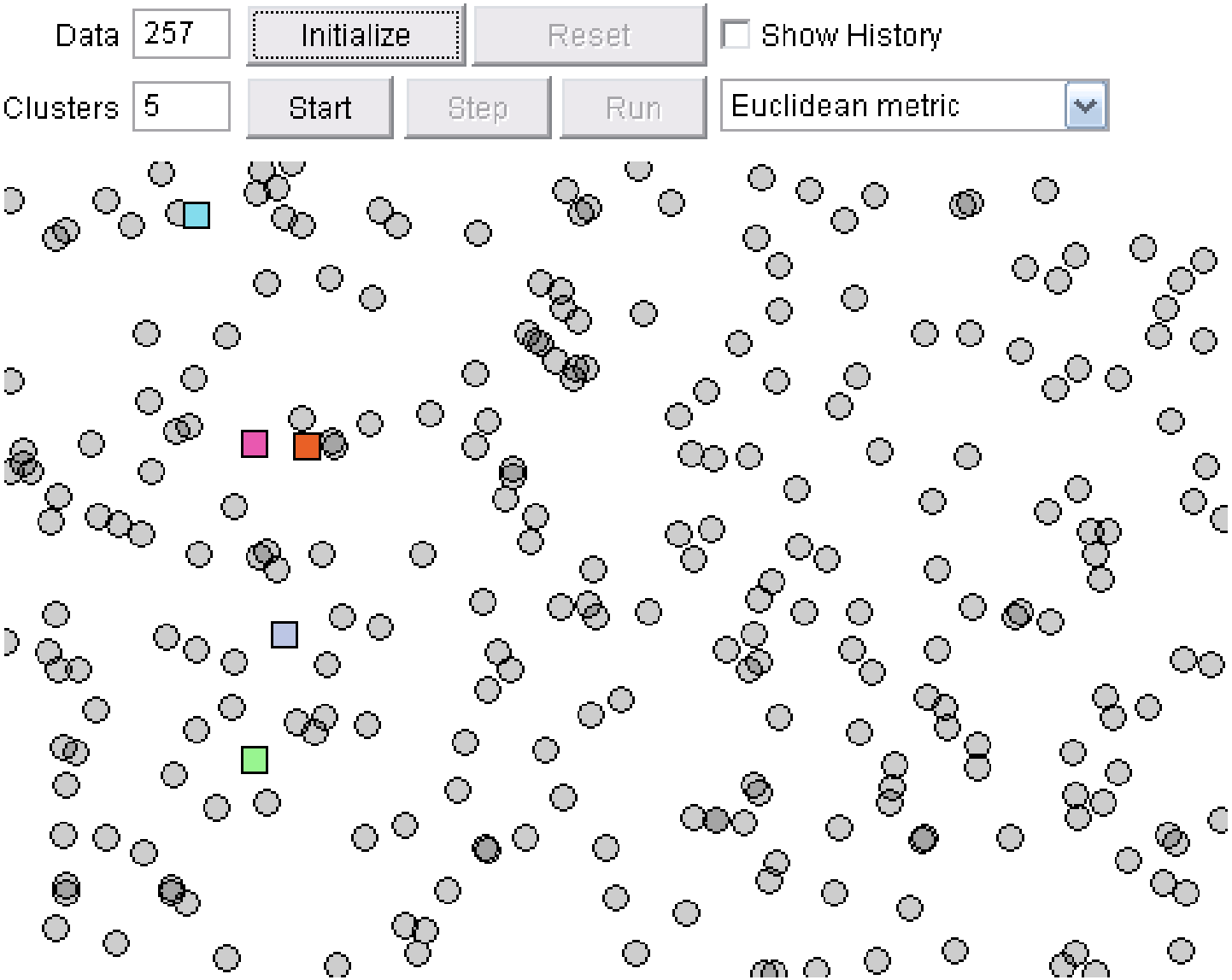
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



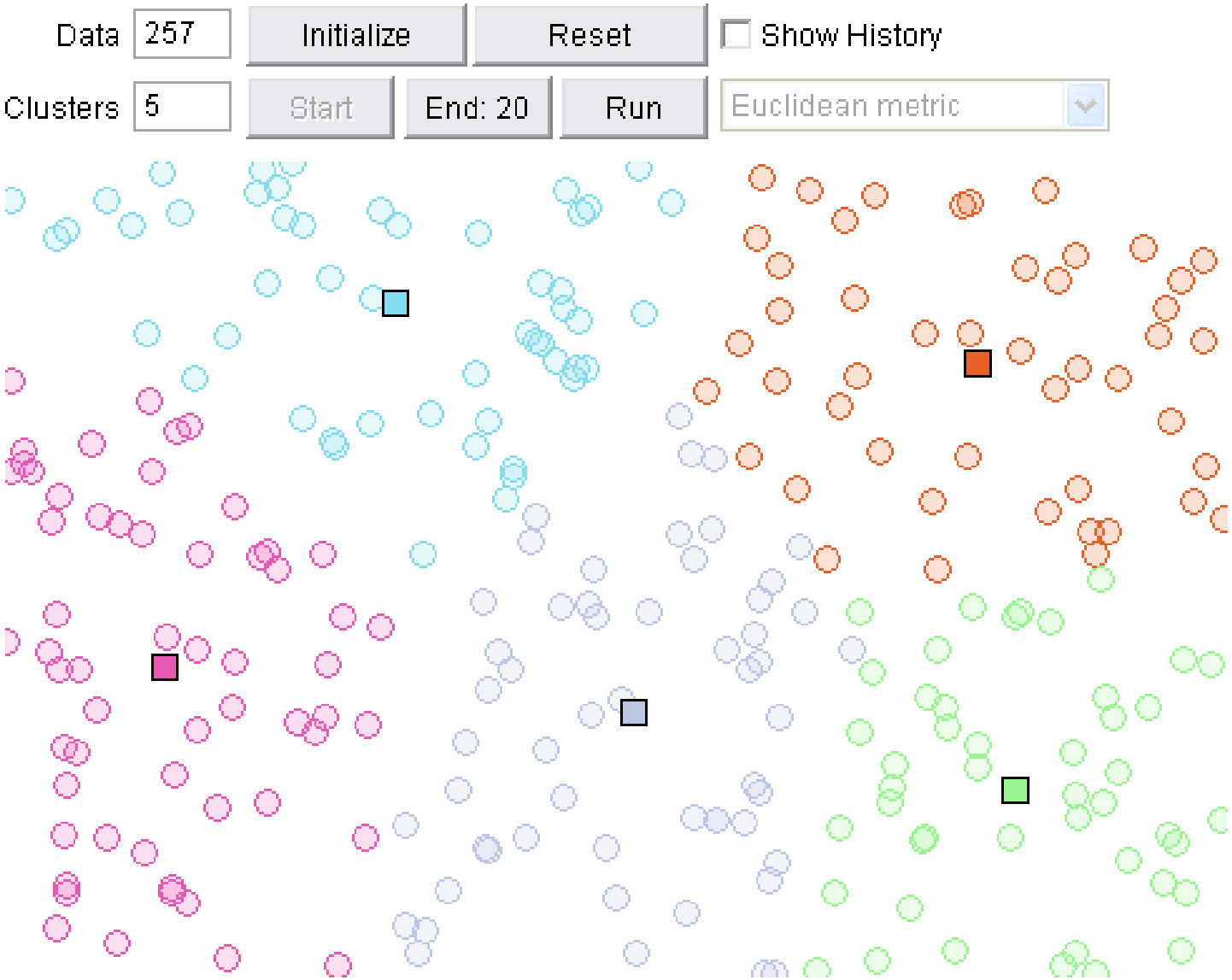
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



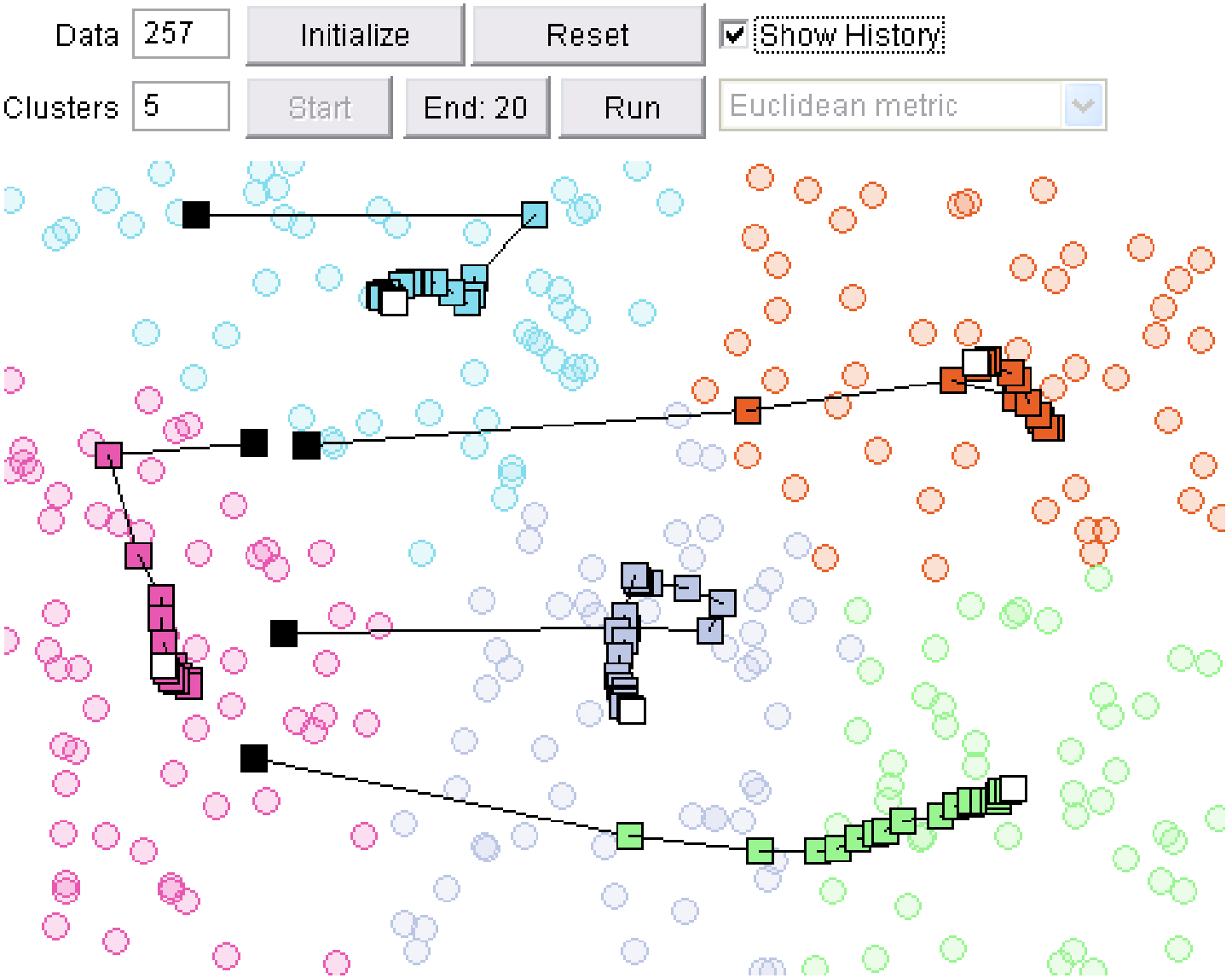
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



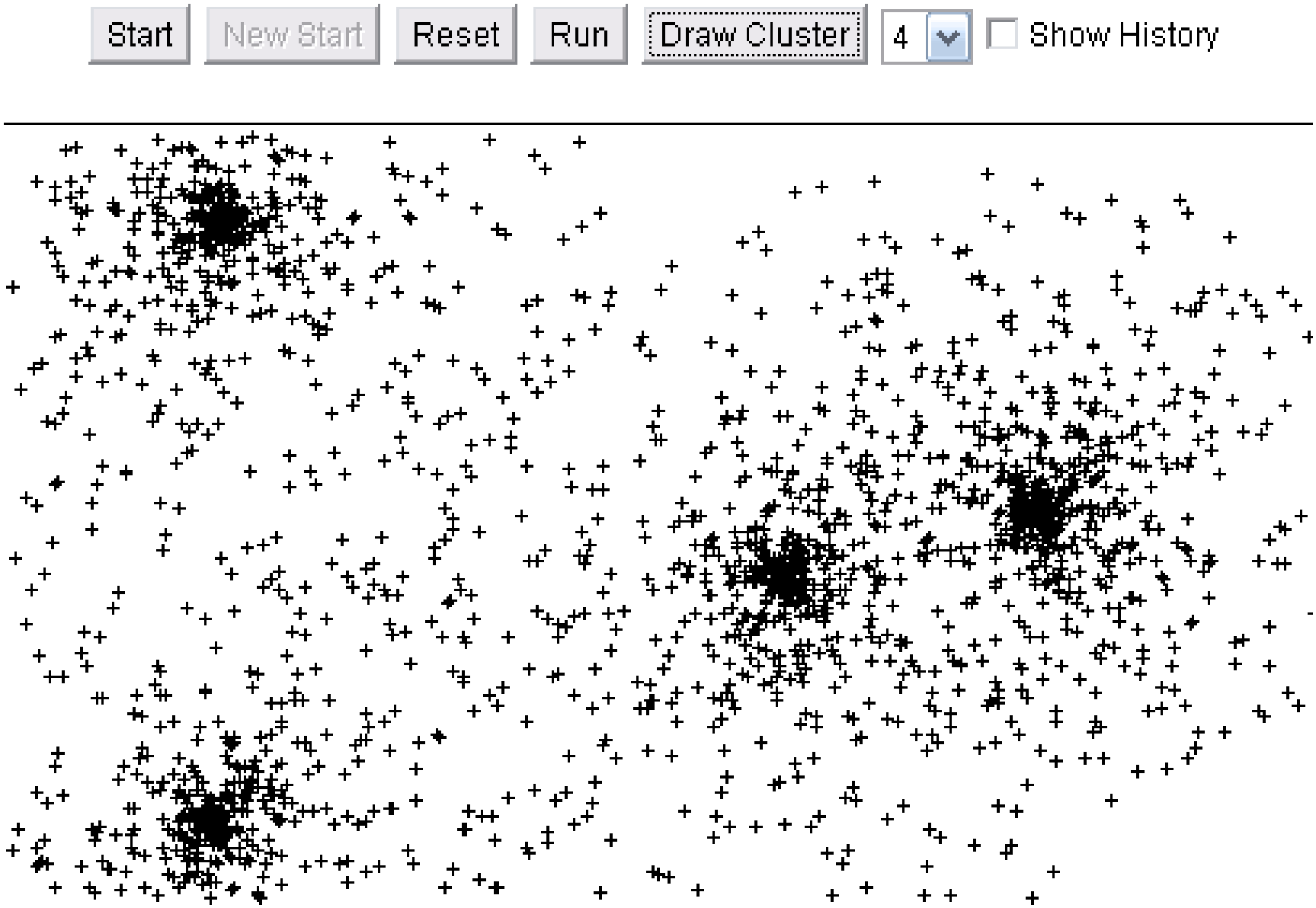
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



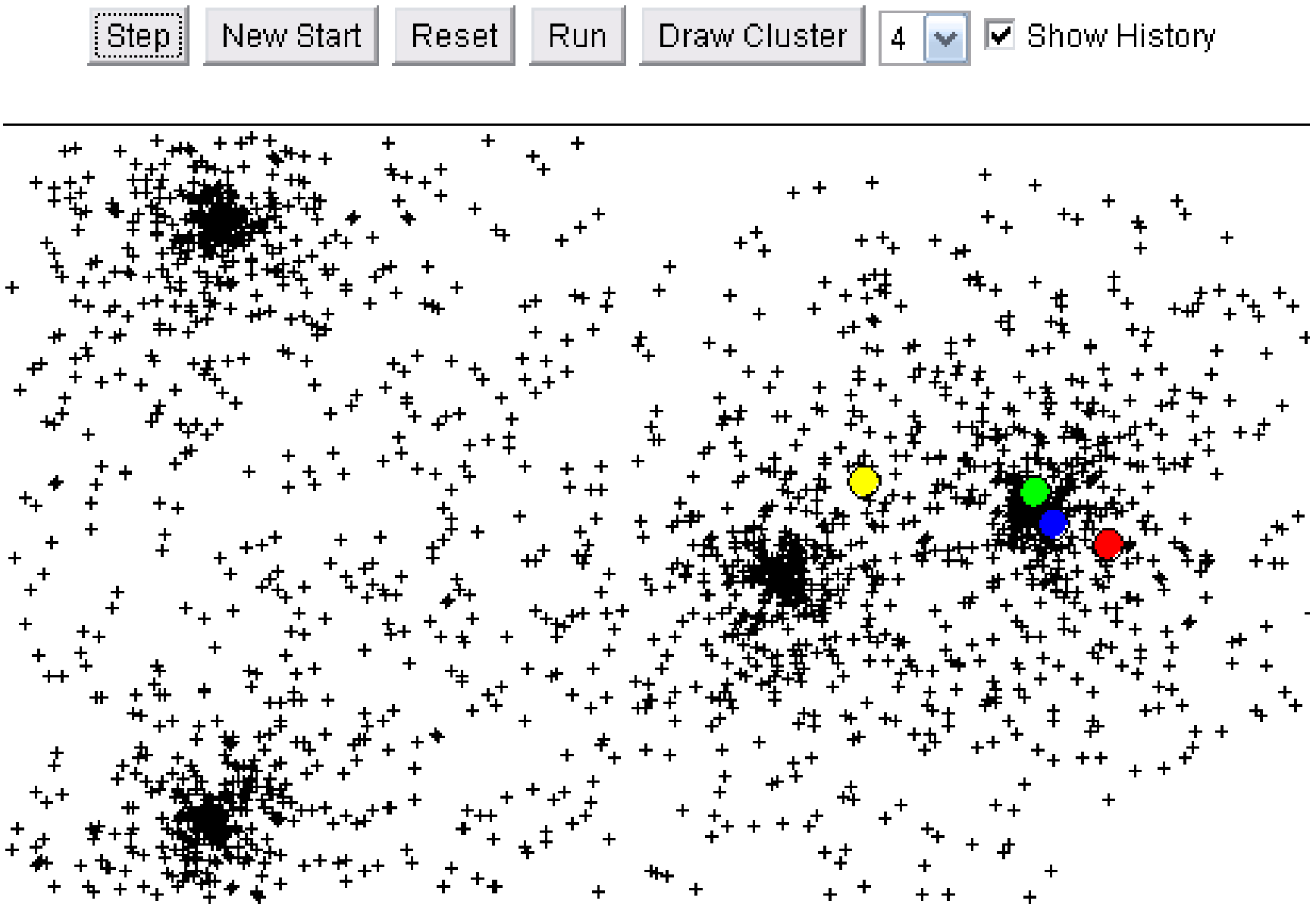
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



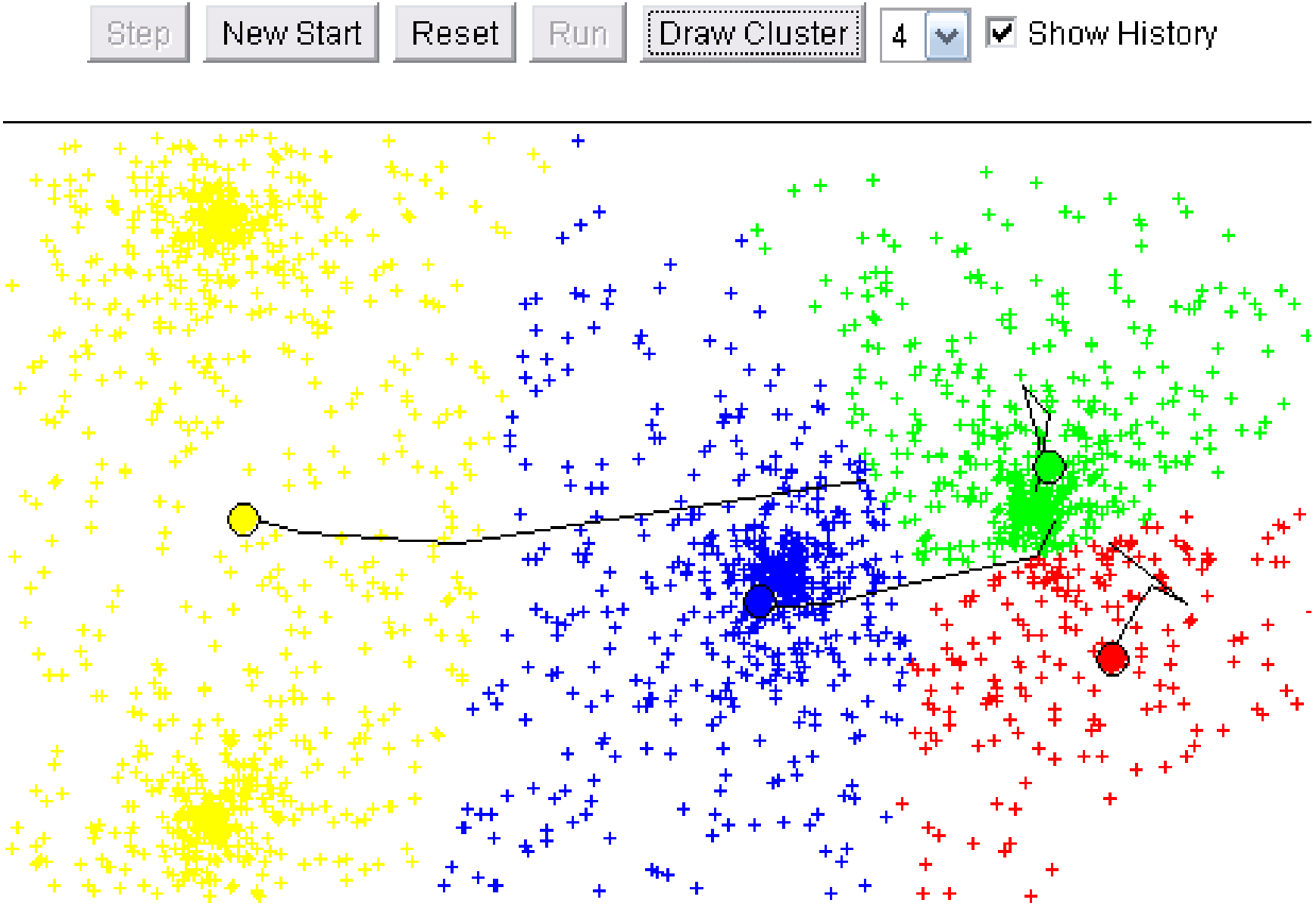
http://www.rob.cs.tu-bs.de/content/04-teaching/06-interactive/Kmeans/Kmeans.html



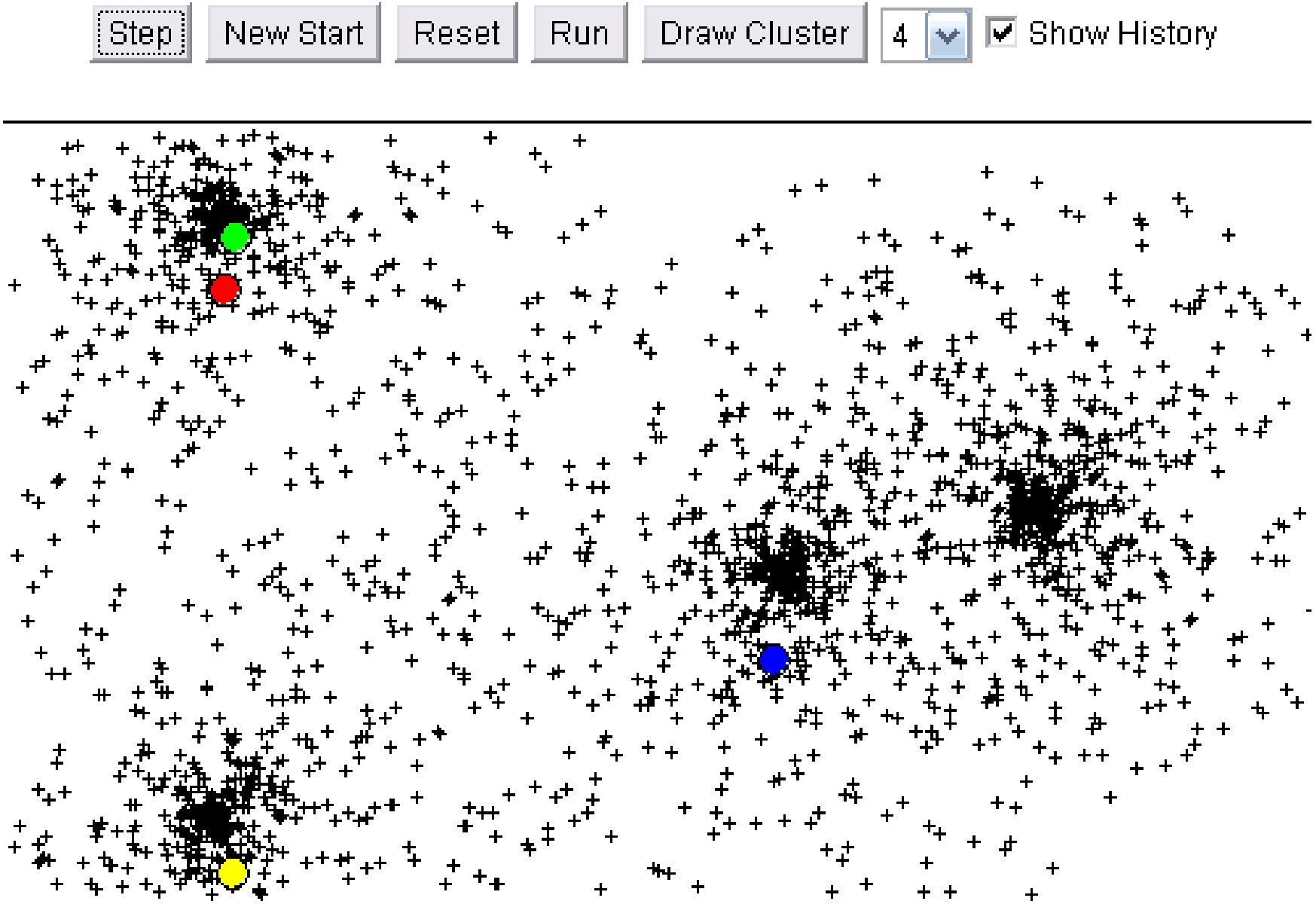
http://www.rob.cs.tu-bs.de/content/04-teaching/06-interactive/Kmeans/Kmeans.html



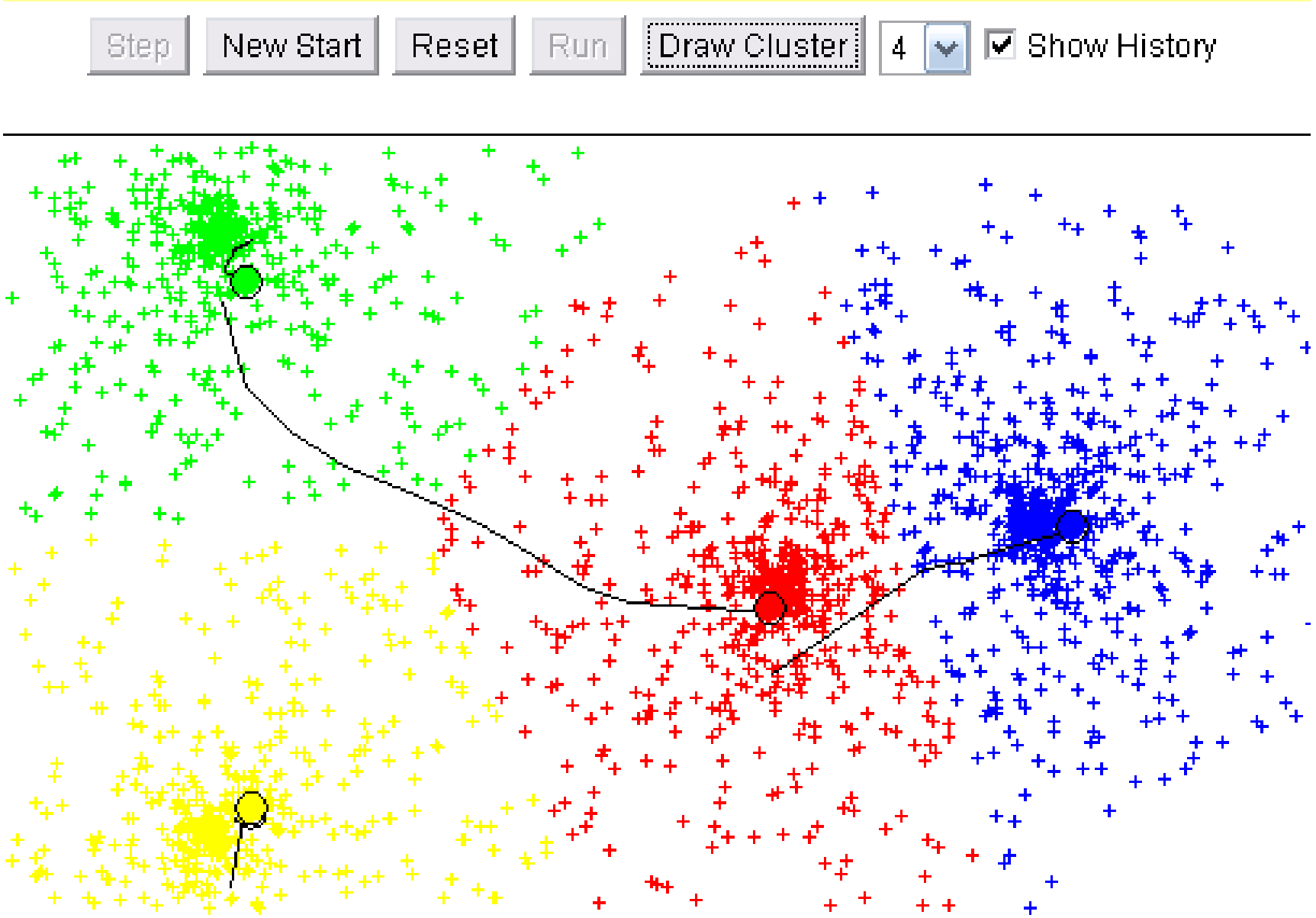
http://www.rob.cs.tu-bs.de/content/04-teaching/06-interactive/Kmeans/Kmeans.html



http://www.rob.cs.tu-bs.de/content/04-teaching/06-interactive/Kmeans/Kmeans.html



http://www.rob.cs.tu-bs.de/content/04-teaching/06-interactive/Kmeans/Kmeans.html



Partiketarako algoritmoa: k-means

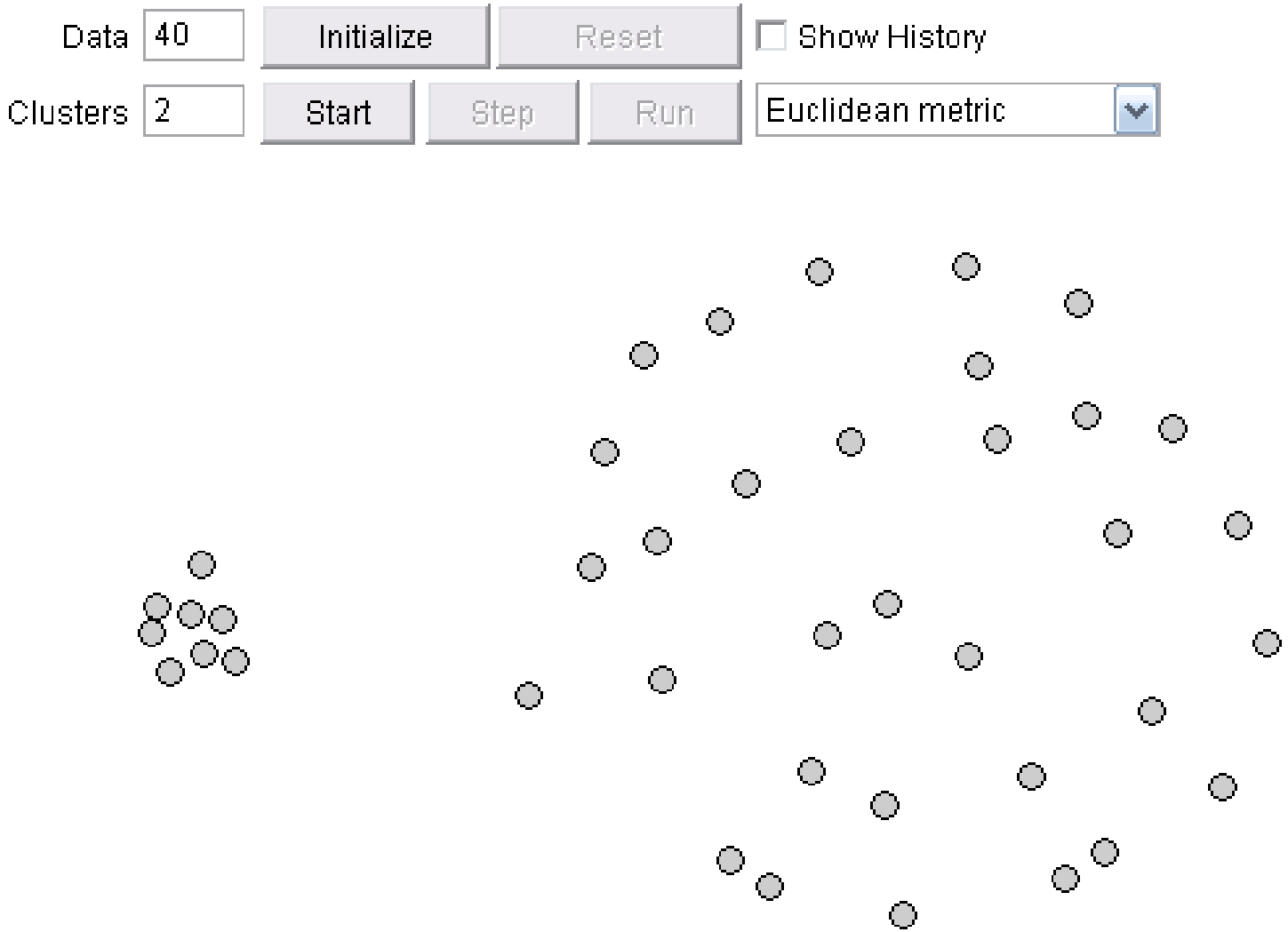
Emaitzaren kalitatea, eta beraz, soluzioarena, hasierako hazien menpe dago.

Partiketarako algoritmoa: k-means

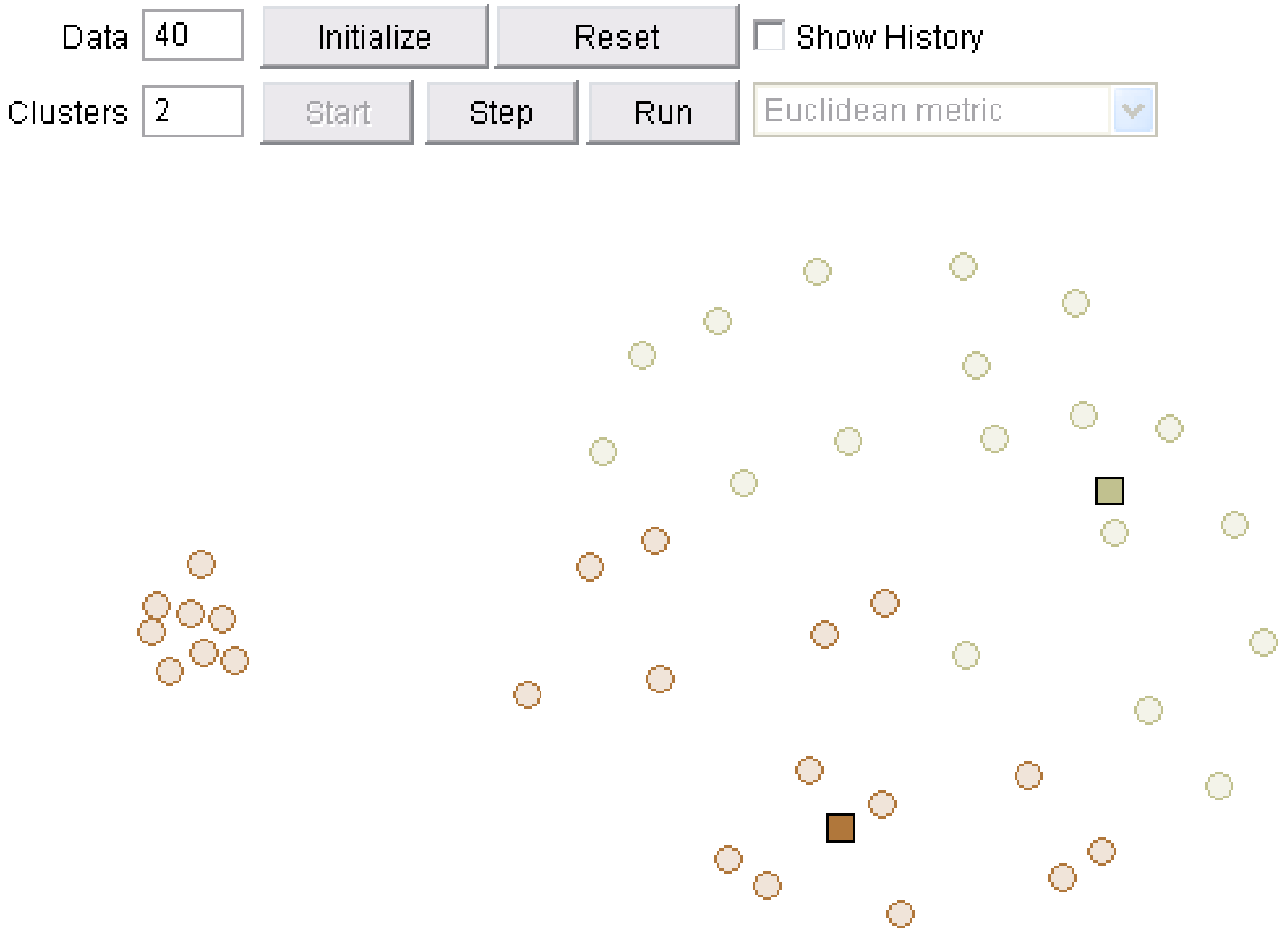
Irudi hauetan 2 sail edo cluster ikusten dugu.

Ea zein den algoritmoaren jokabidea

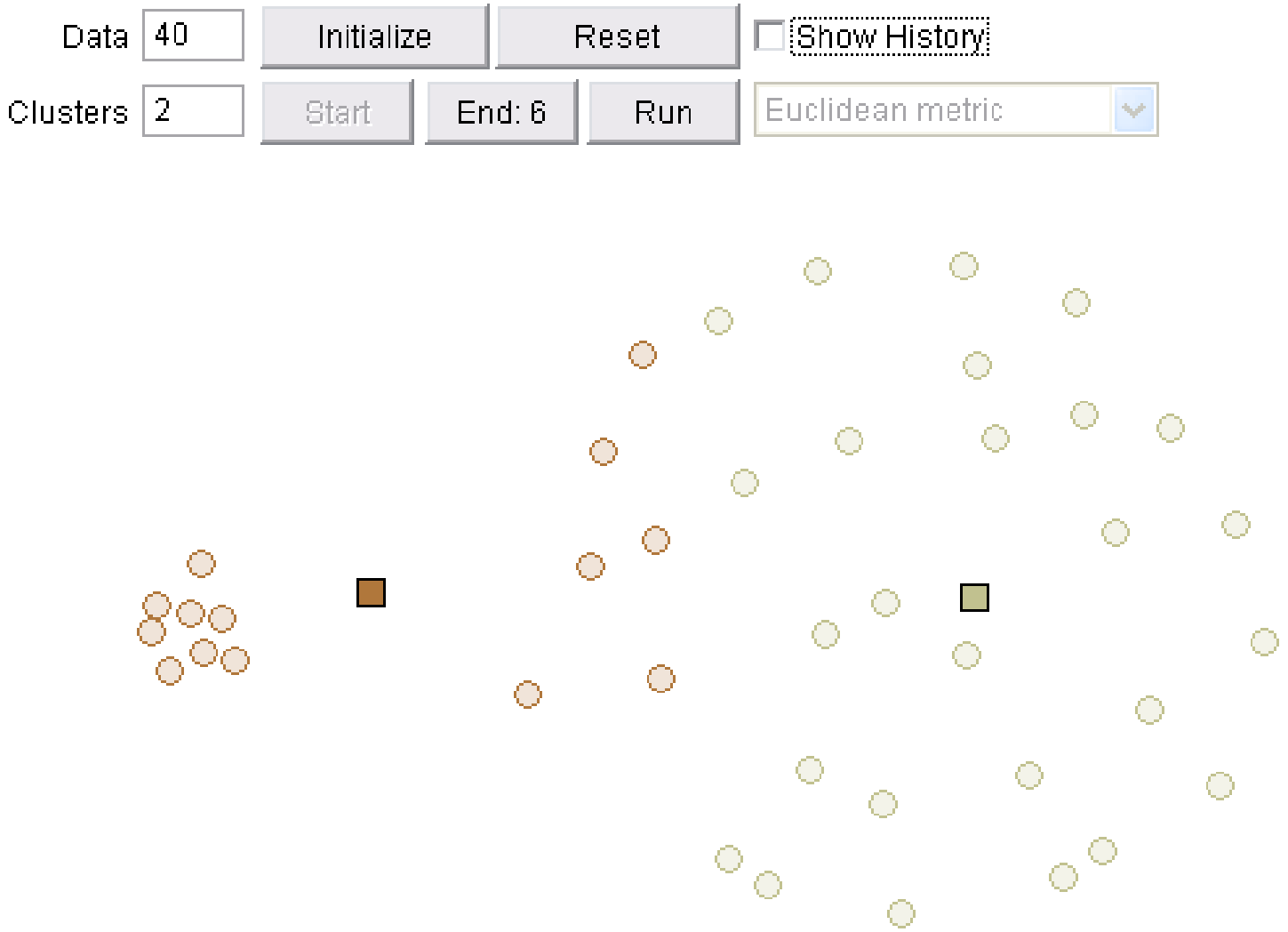
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



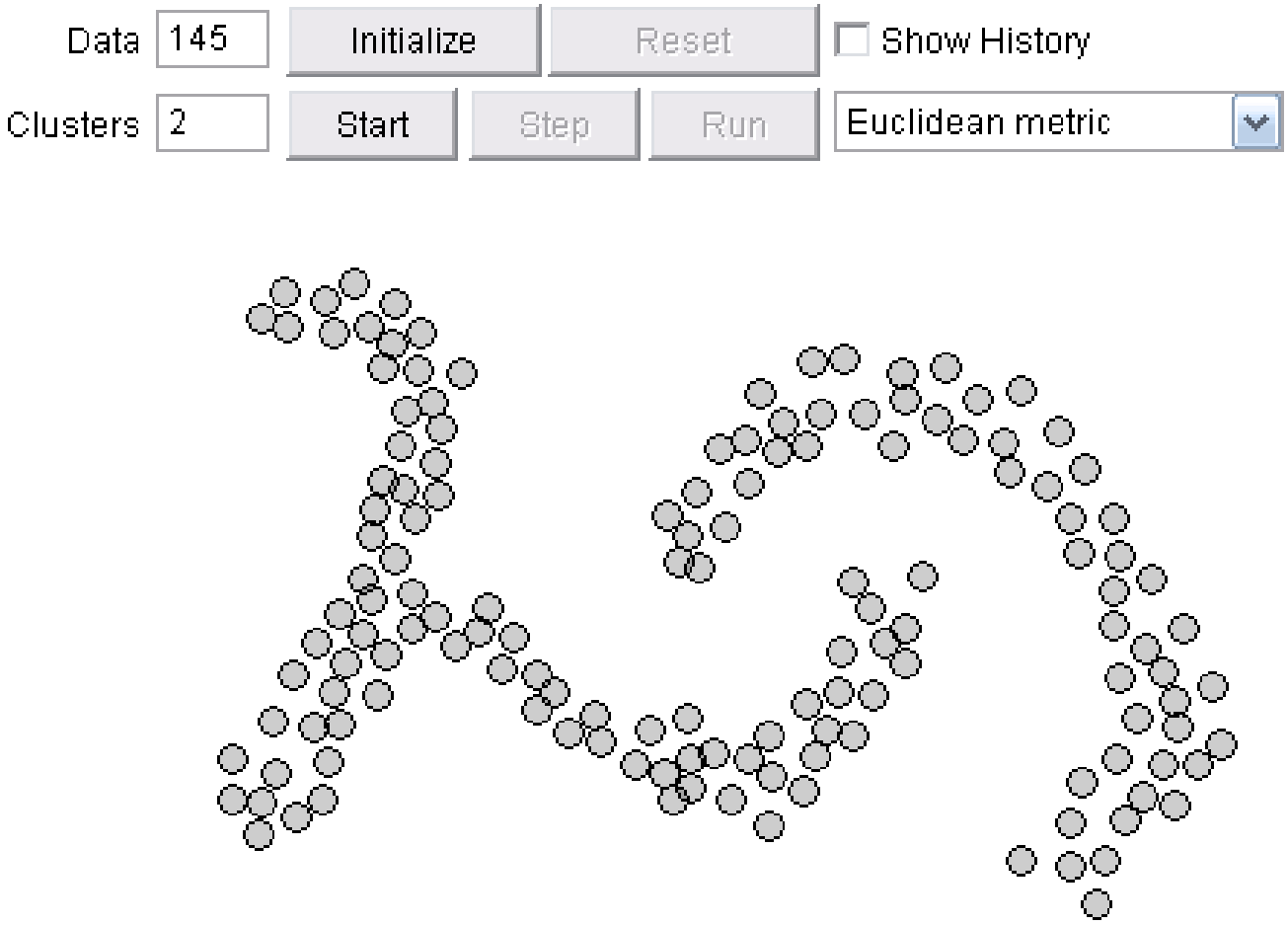
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



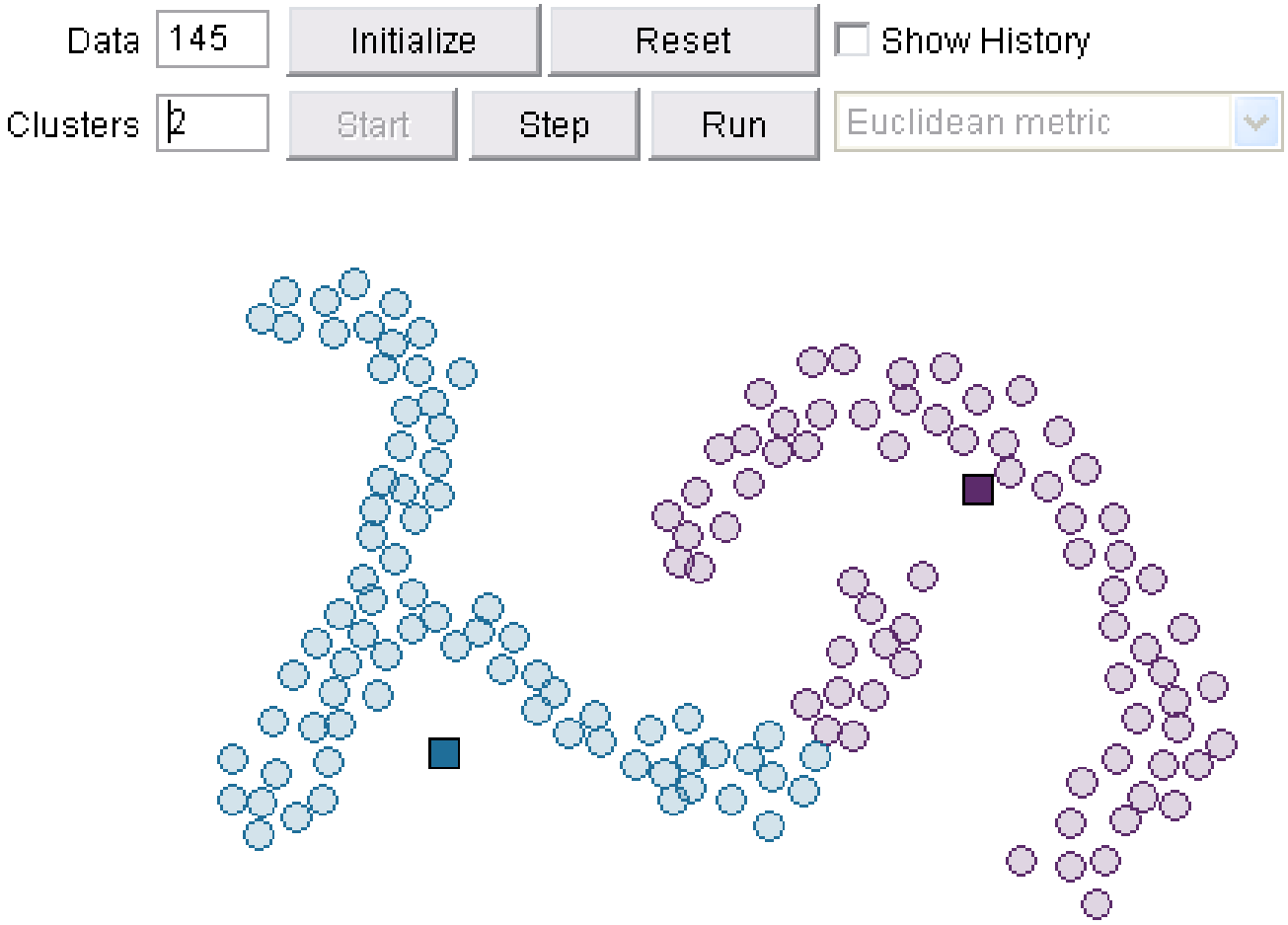
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



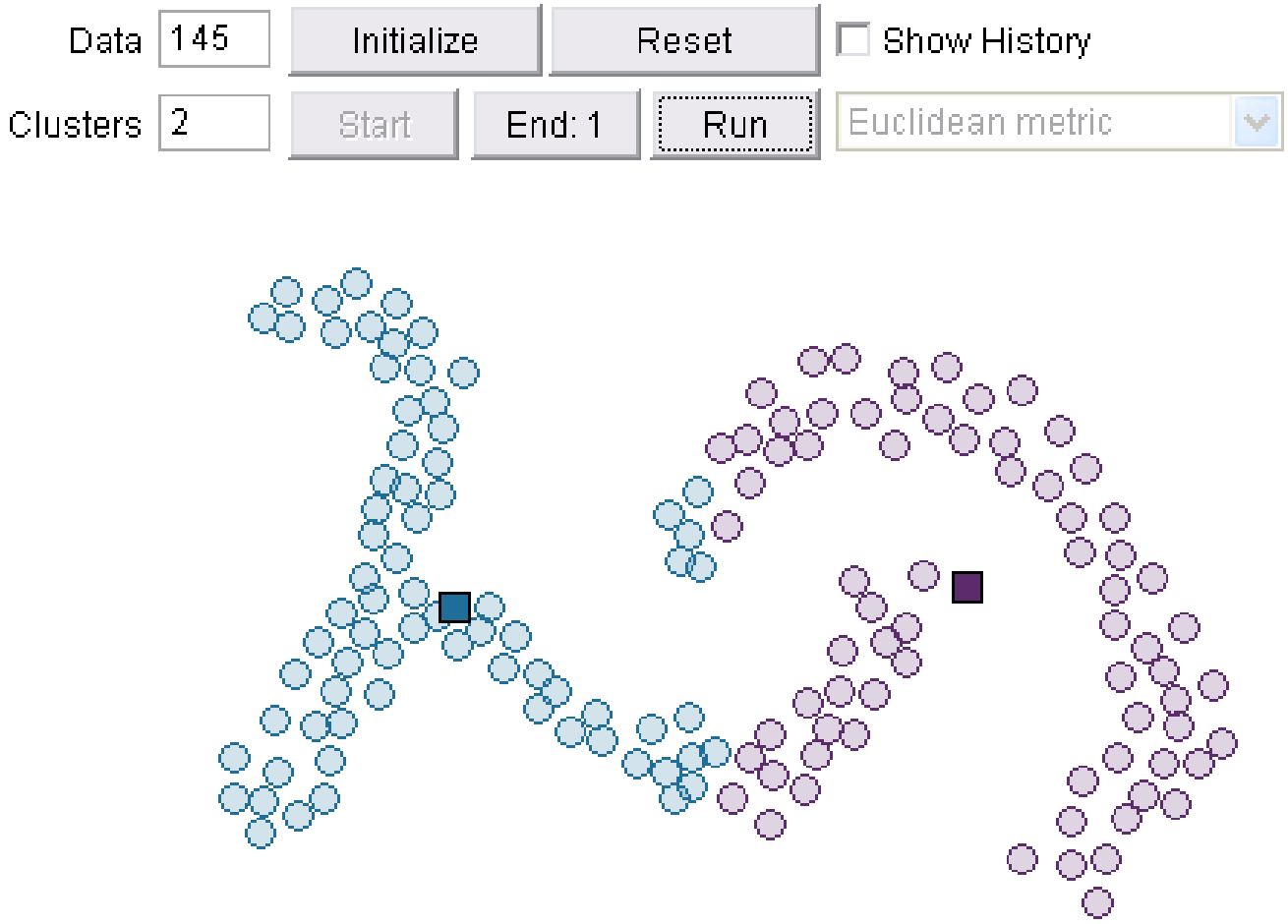
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html



Partiketarako algoritmoa: k-means

Irudi hauetan 2 sail edo cluster ikusten dugu.

*Clustering*a egiteko ohiko tekniken atzean dagoena zera da, *cluster*rak esferikoak eta trinkotasun berdinekoak direla.
*Cluster*rak ez direnean halakoak algoritmoek ez dute emaitza egokirik ematen.
Giza-senarekin bat etortzeko bestelako teknikak behar dira.

Hierarkiarako algoritmoa

Algoritmoa (zatikorra):

- Sail bakarra duen partiketa (Ω) hartu abiapuntutzat.
- Aurreko partiketaren bi sail *ezantzekoenak* ***zatitu***, hau da, partiketa berri bat osatu.
- Partiketa berria bakarkako sailez osatuta badago, bukatu, bestela itzuli 2. urratsera.

Konputazionalki garestia, ez bada bilaketa murrizten

Hierarkiarako algoritmoa

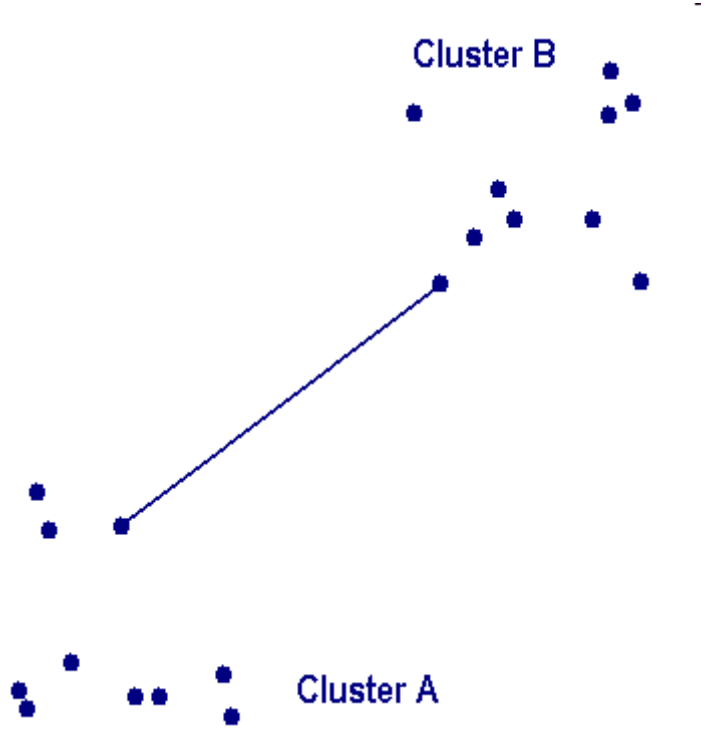
Algoritmoa (elkarkorra):

- Bakarkako partiketa edo sailkapena hartu abiapuntutzat (objektu bakoitzak sail bat osatzen du).
- Aurreko partiketaren bi sail gertuenak (*distantzia*) edo *antzekoenak* ***elkartu***, hau da, partiketa berri bat osatu.
- Partiketa berria sail bakar batez osatuta badago (Ω), bukatu, bestela itzuli 2. urratsera.

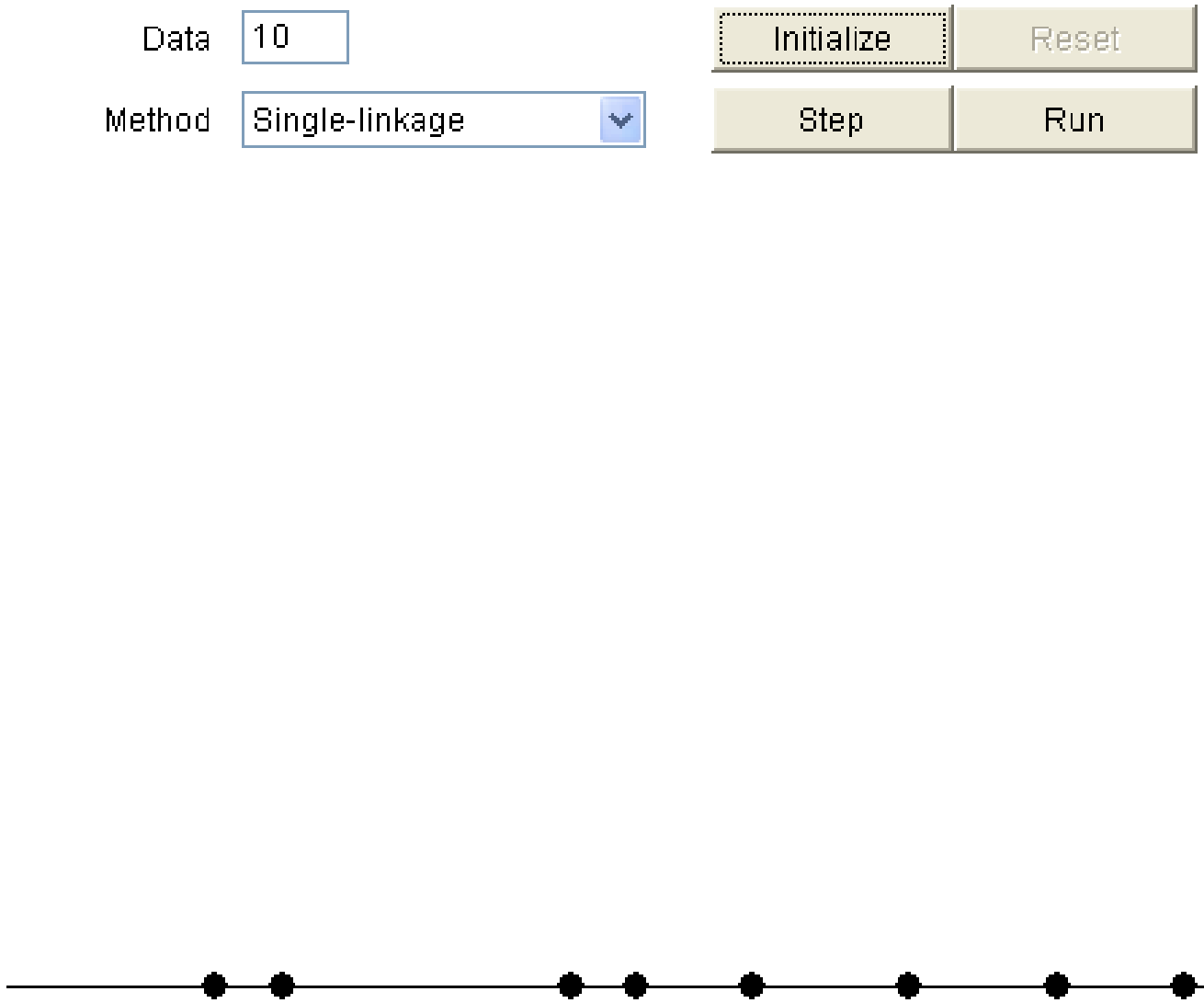
Hierarkiarako algoritmoa

Sailen arteko distantzia:

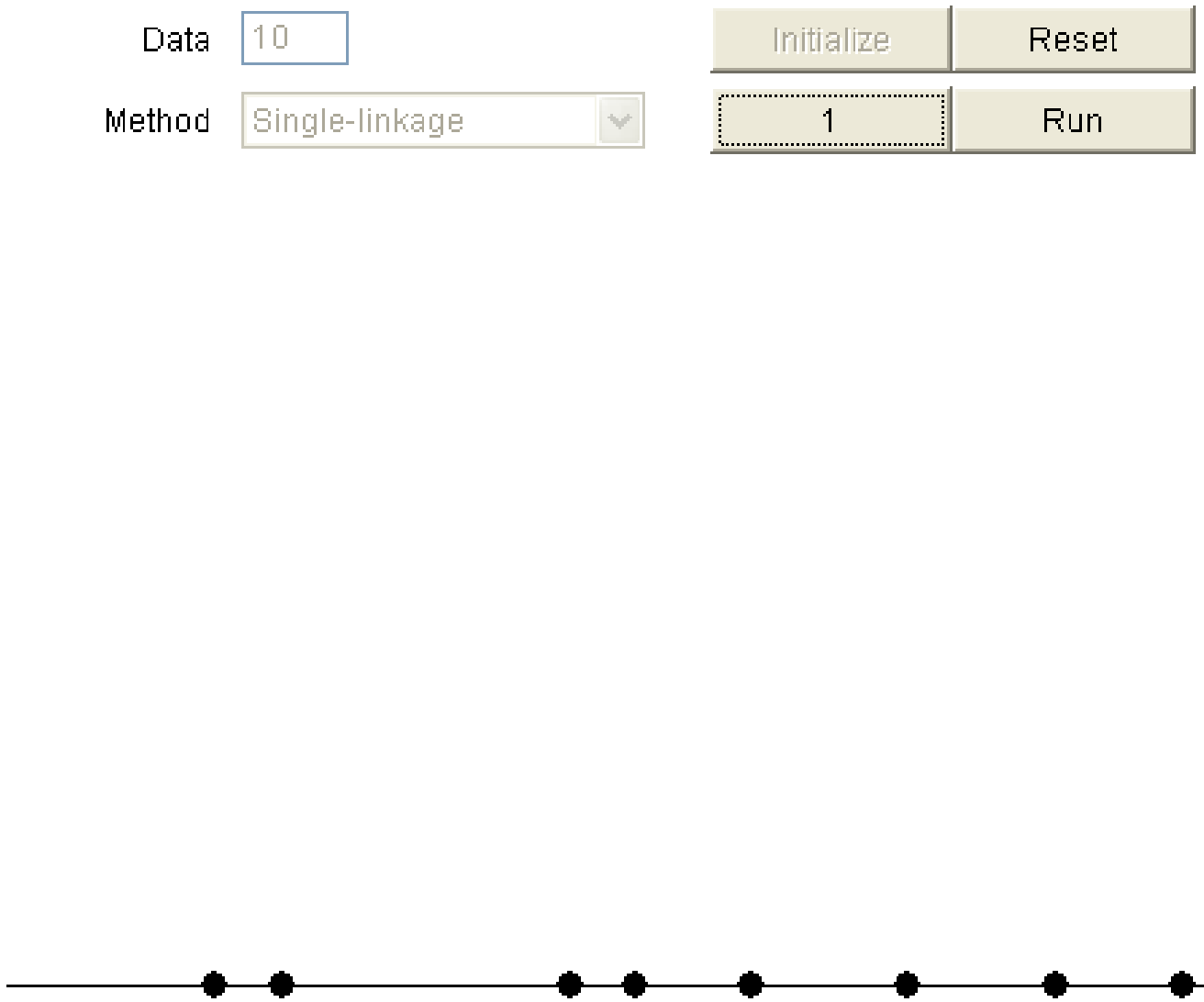
single-linkage



http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletH.html

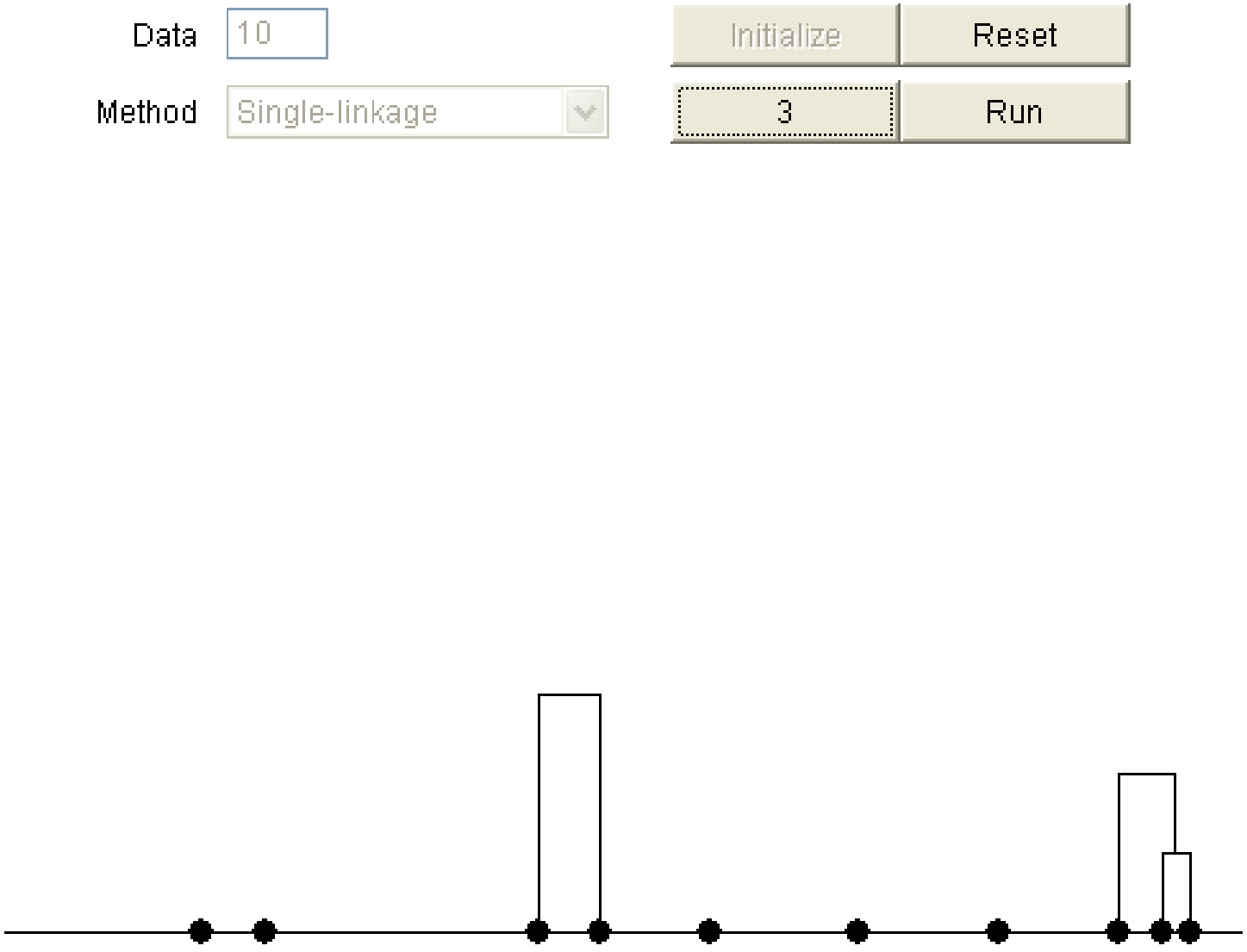


http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletH.html

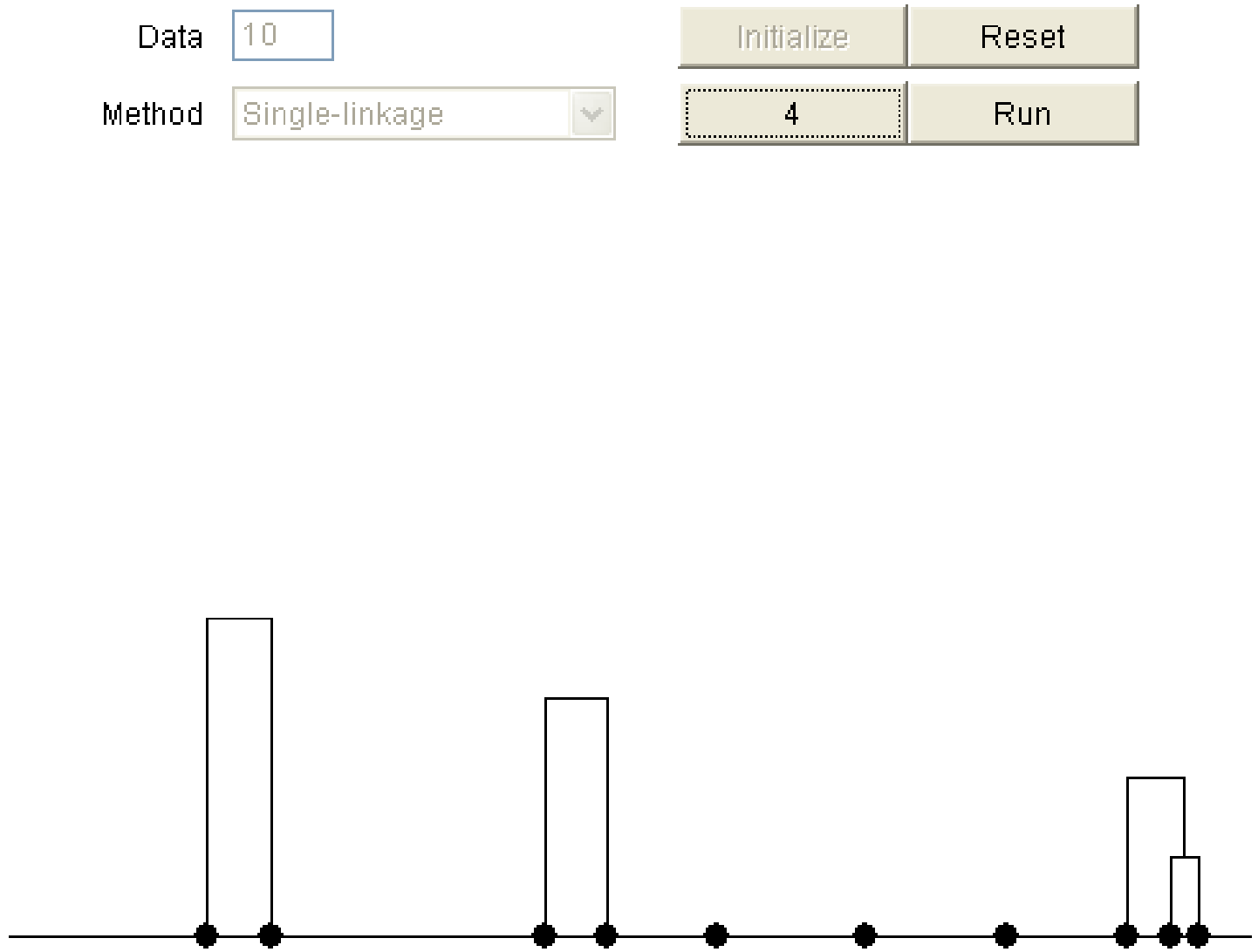


http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletH.html

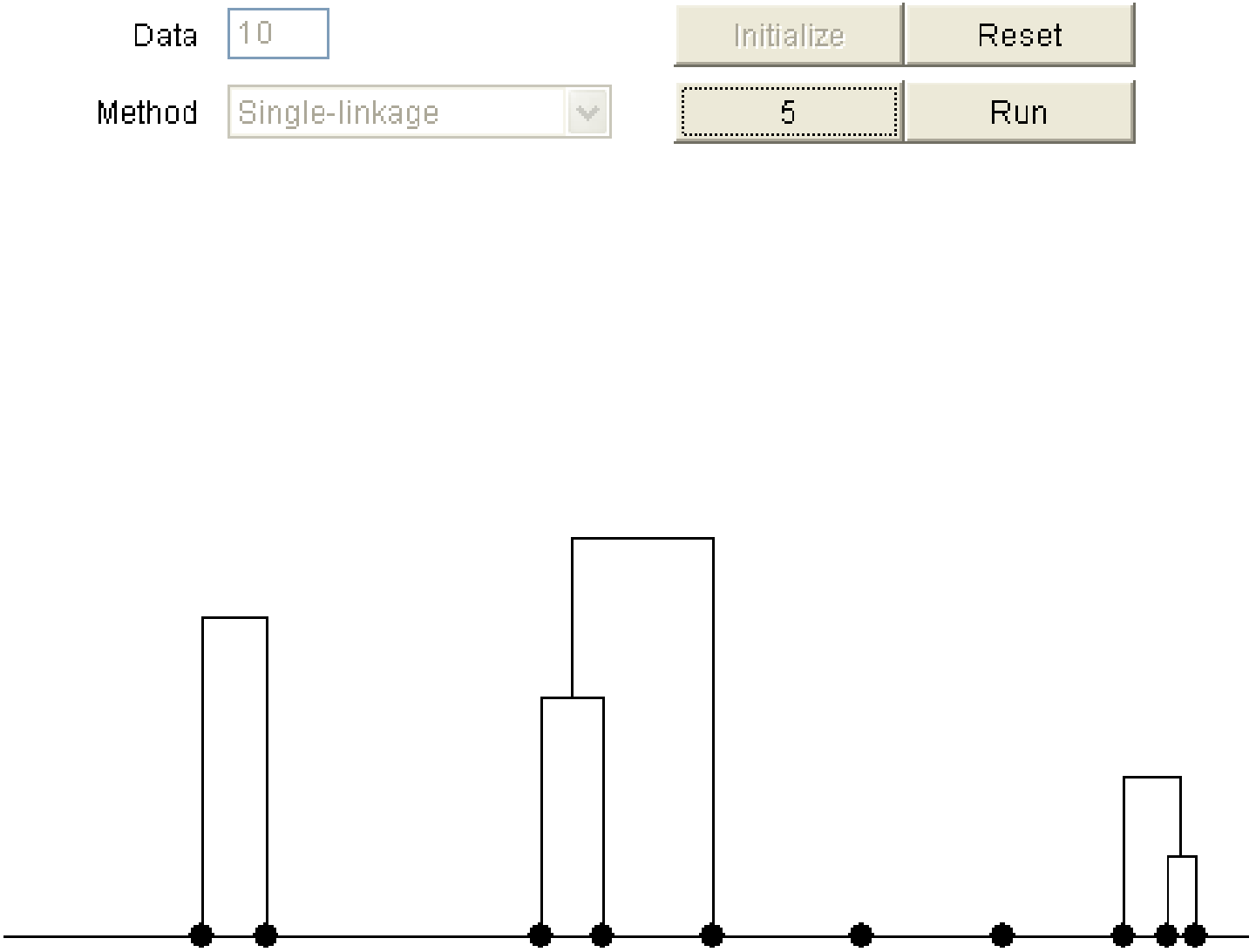
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletH.html



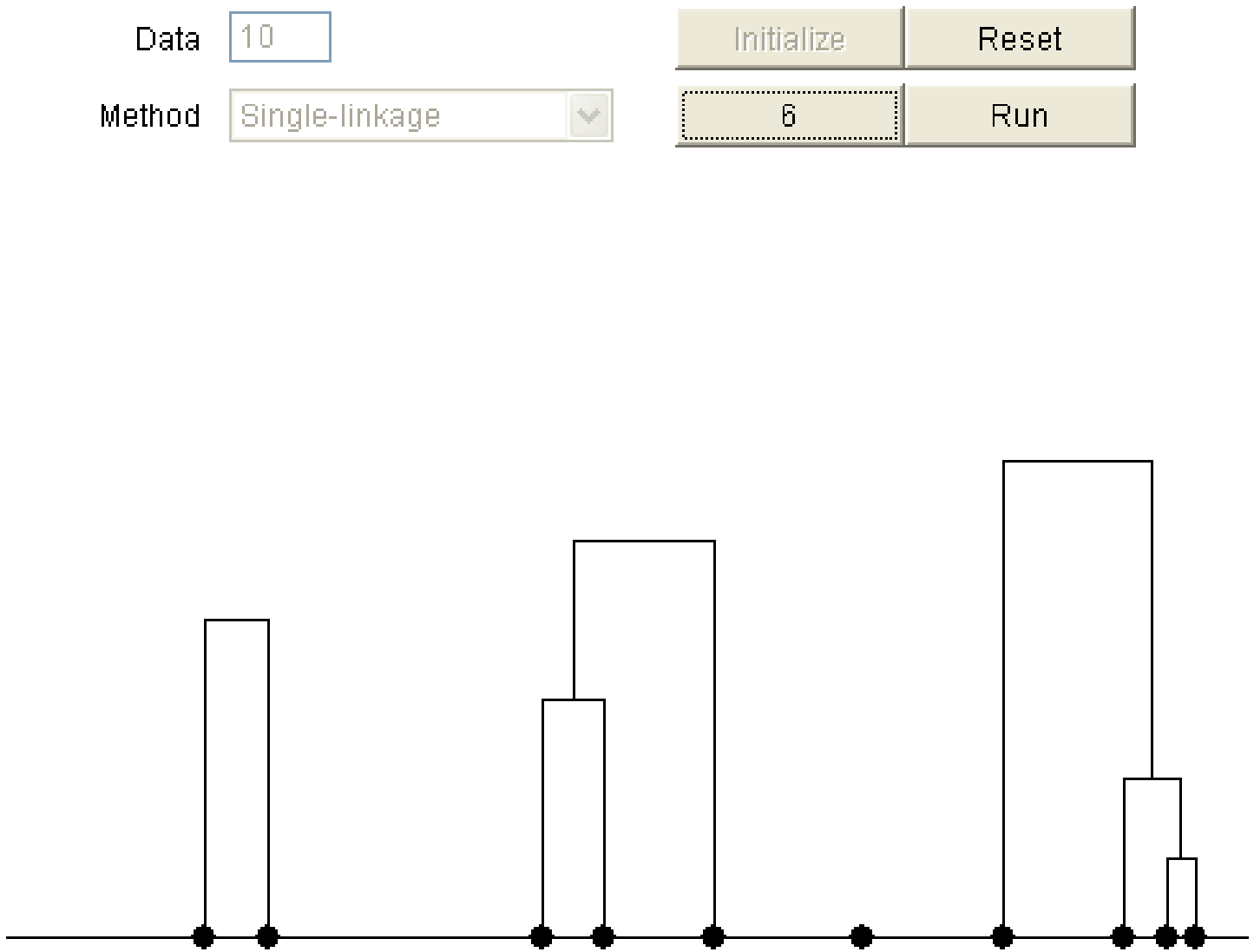
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletH.html



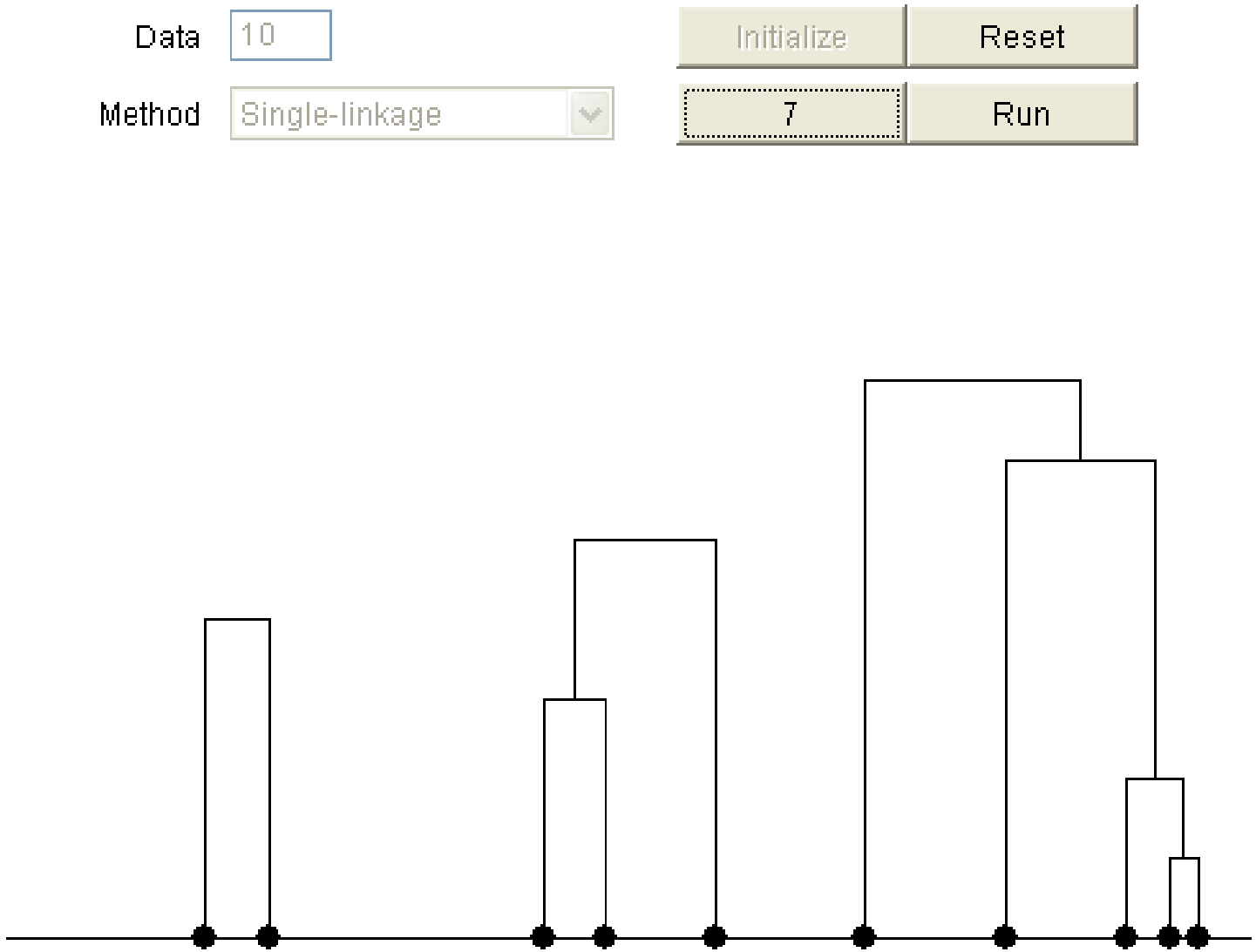
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletH.html



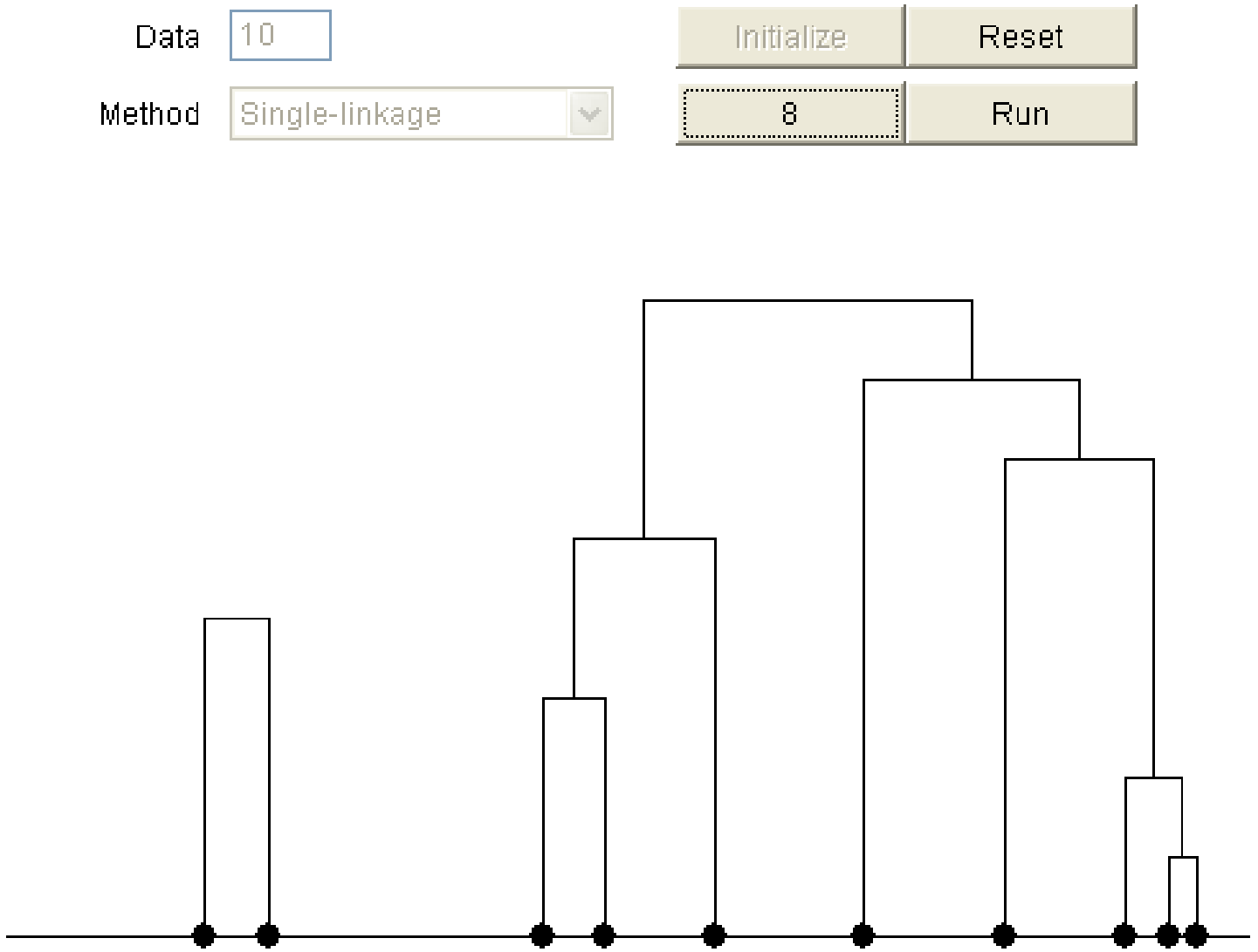
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletH.html



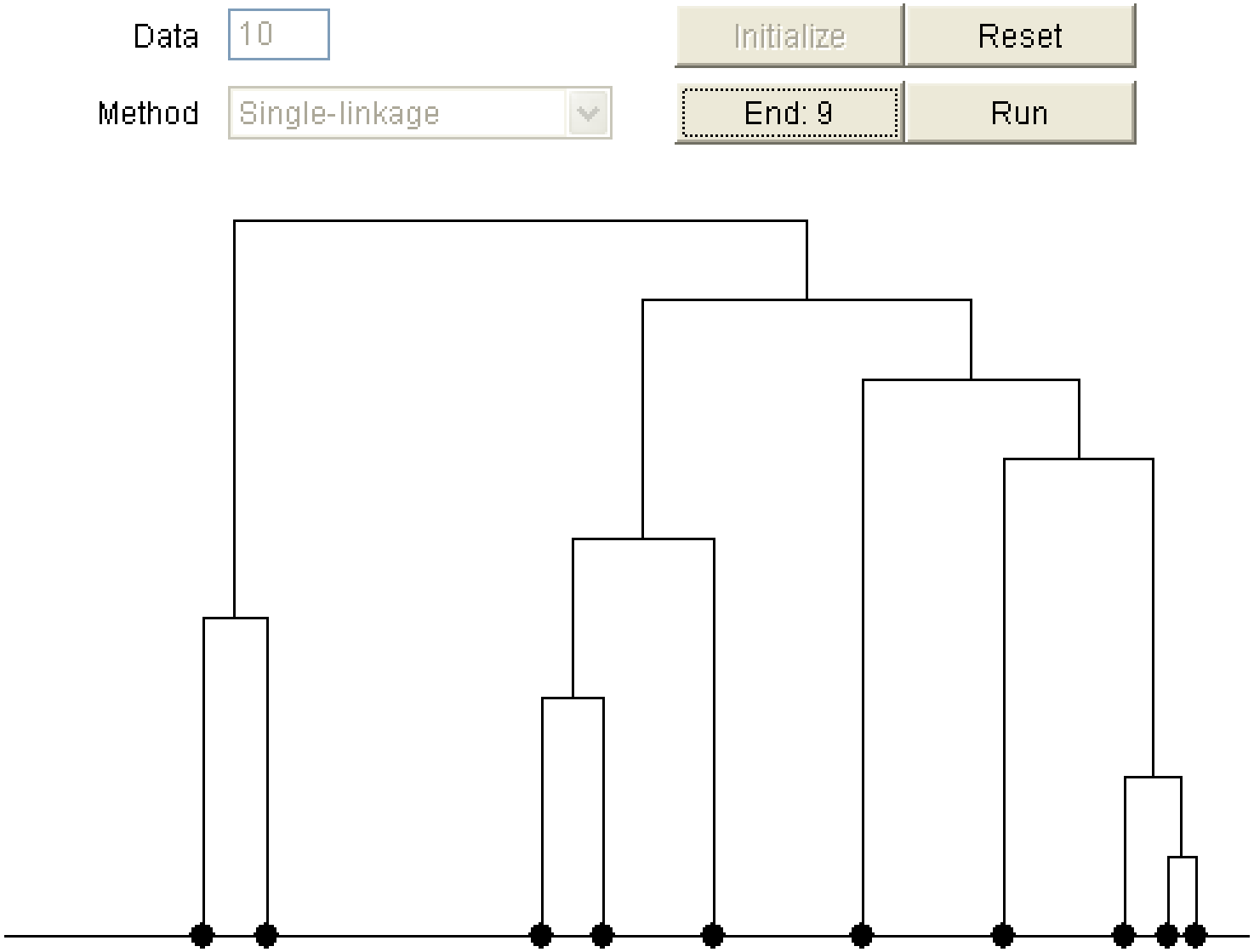
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletH.html



http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletH.html



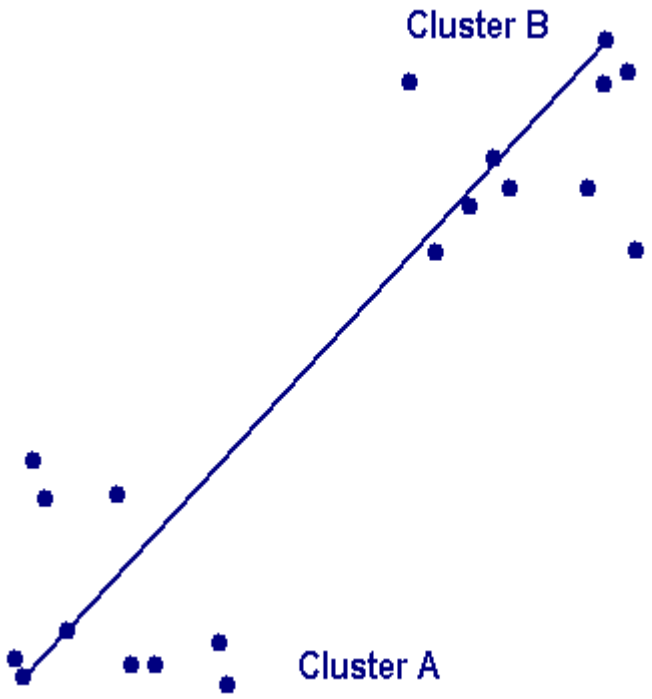
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletH.html



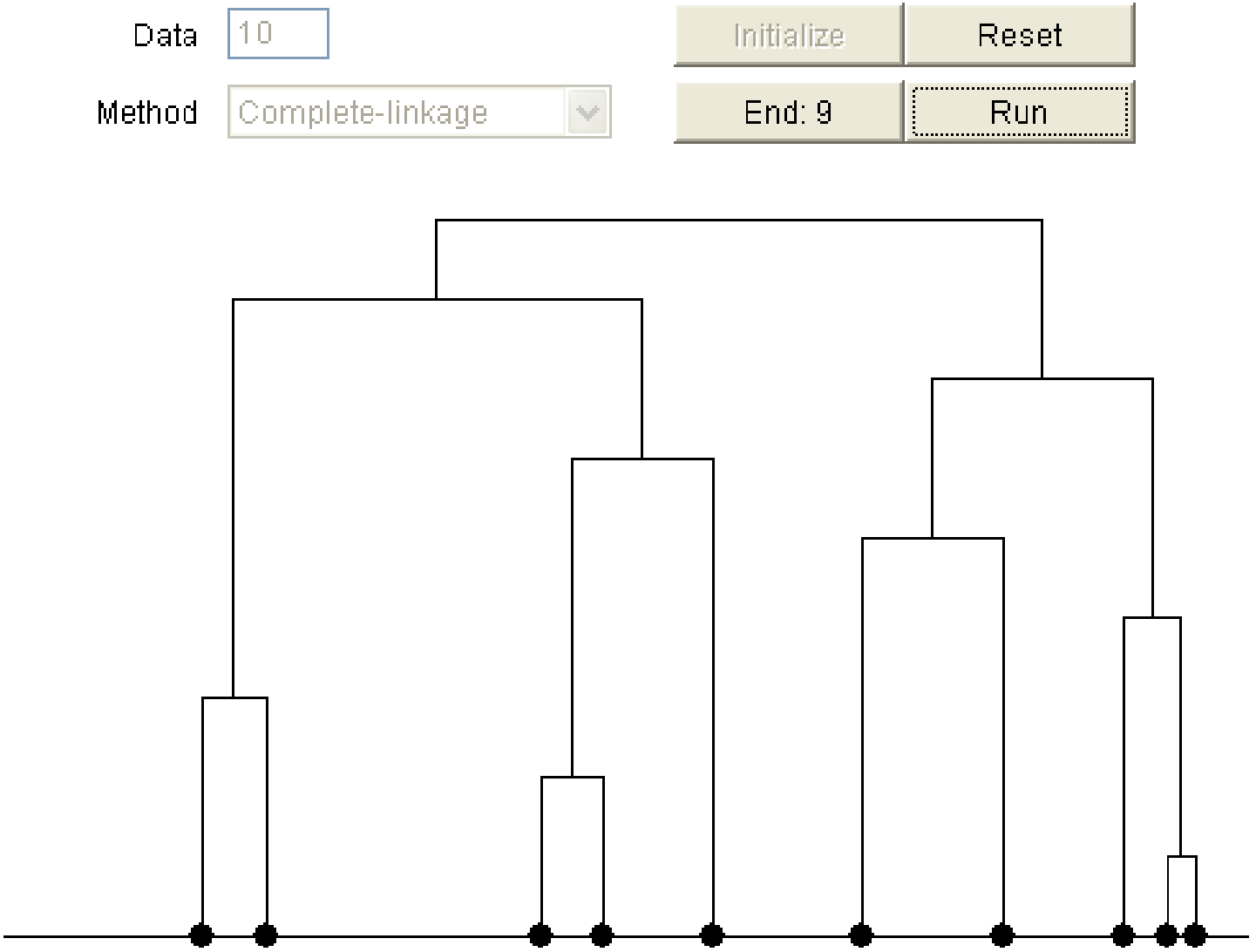
Hierarkiarako algoritmoa

Sailen arteko distantzia:

complete-linkage



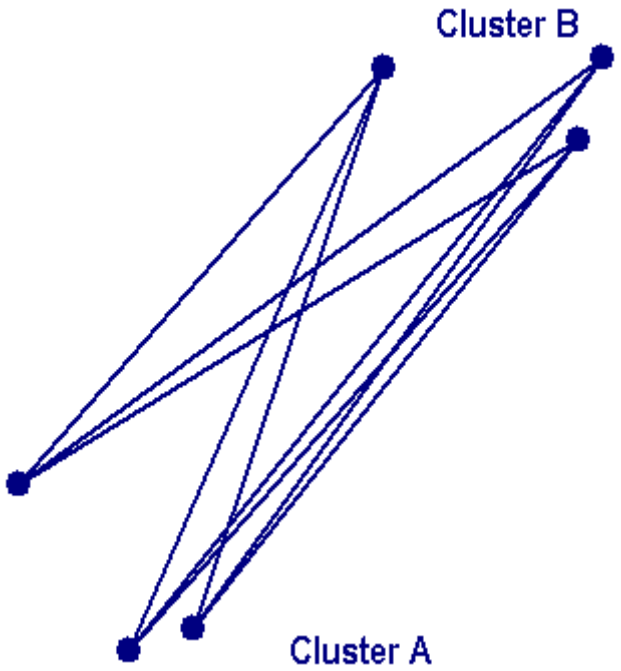
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletH.html



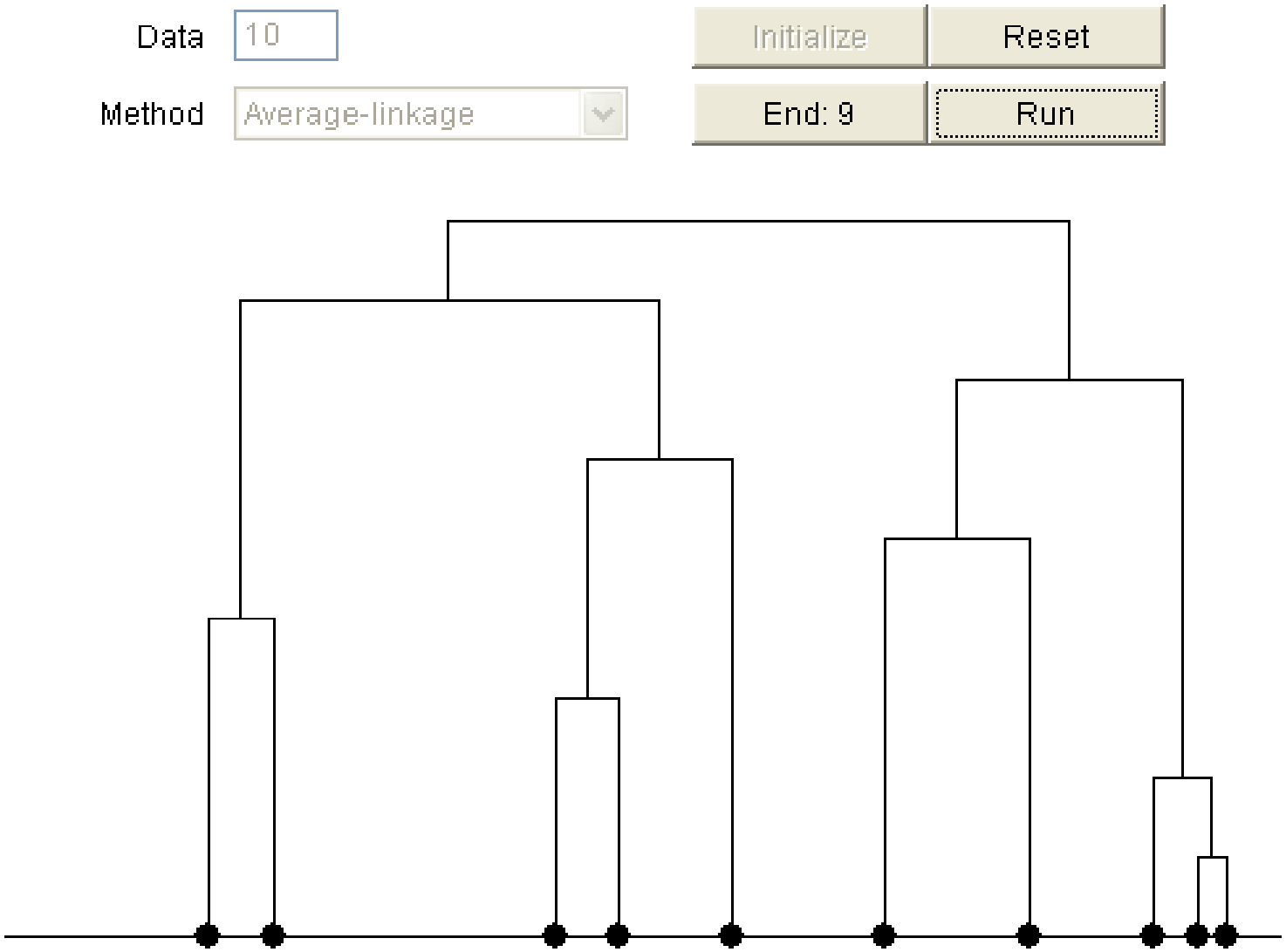
Hierarkiarako algoritmoa

Sailen arteko distantzia:

average-linkage

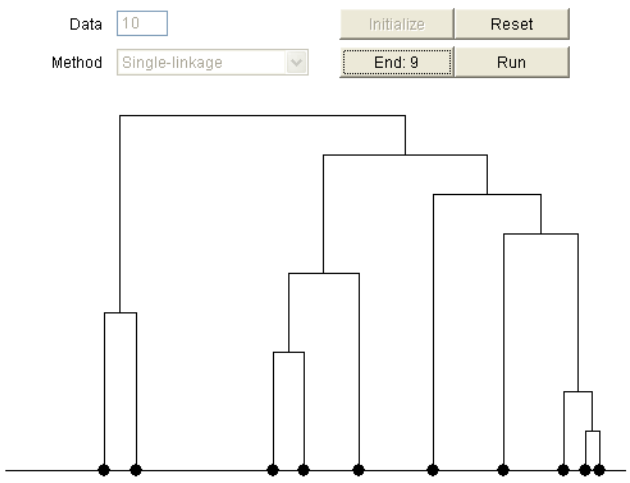


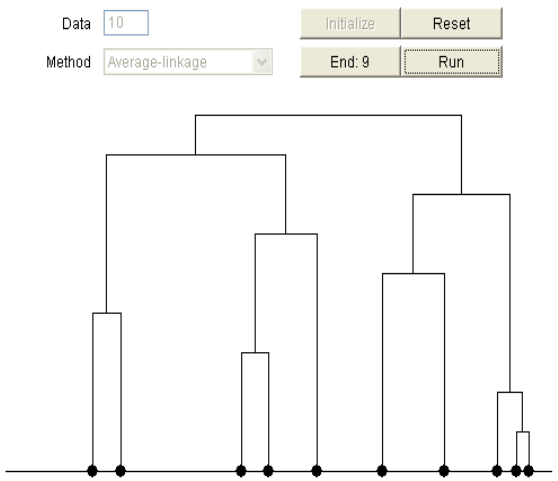
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletH.html

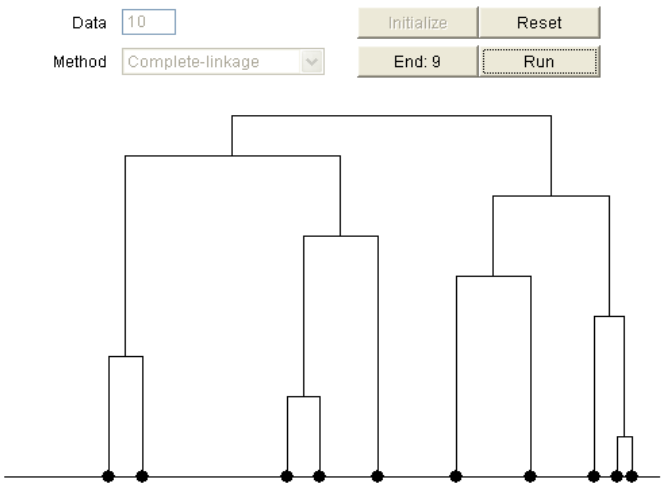


Hierarkiarako algoritmoa

Orokorrean, emaitzak ez dira berberak

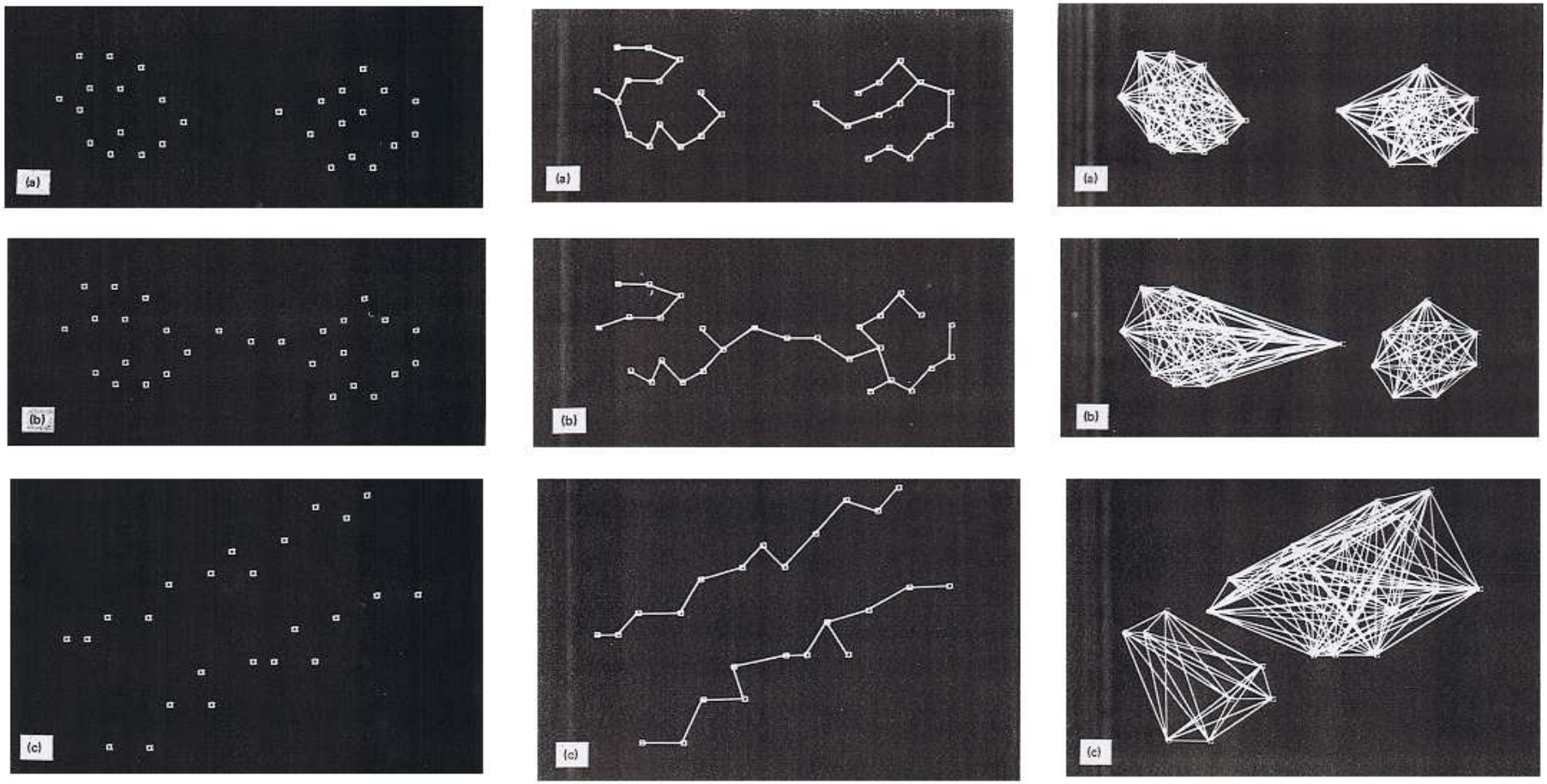






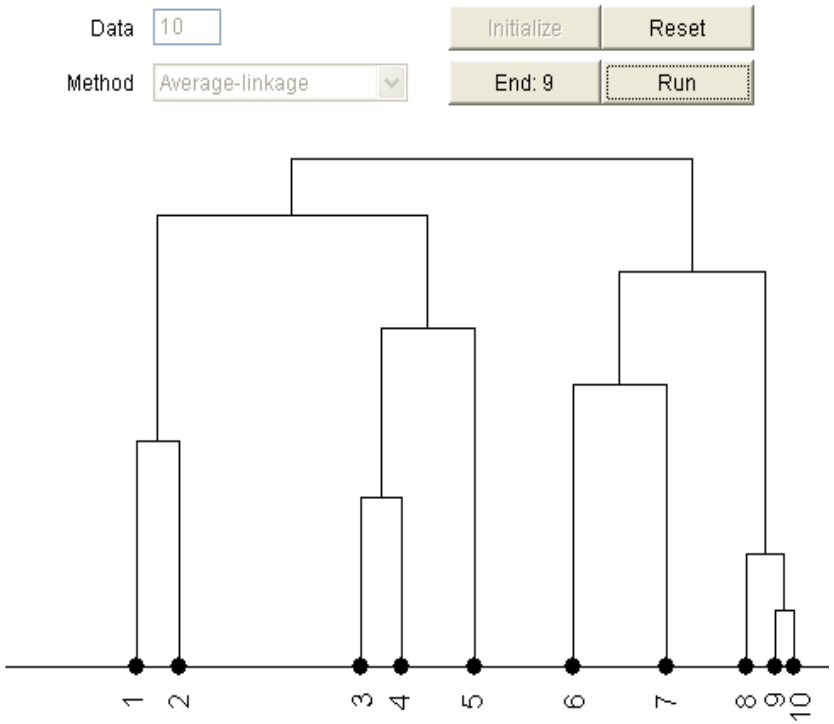
Hierarkiarako algoritmoa

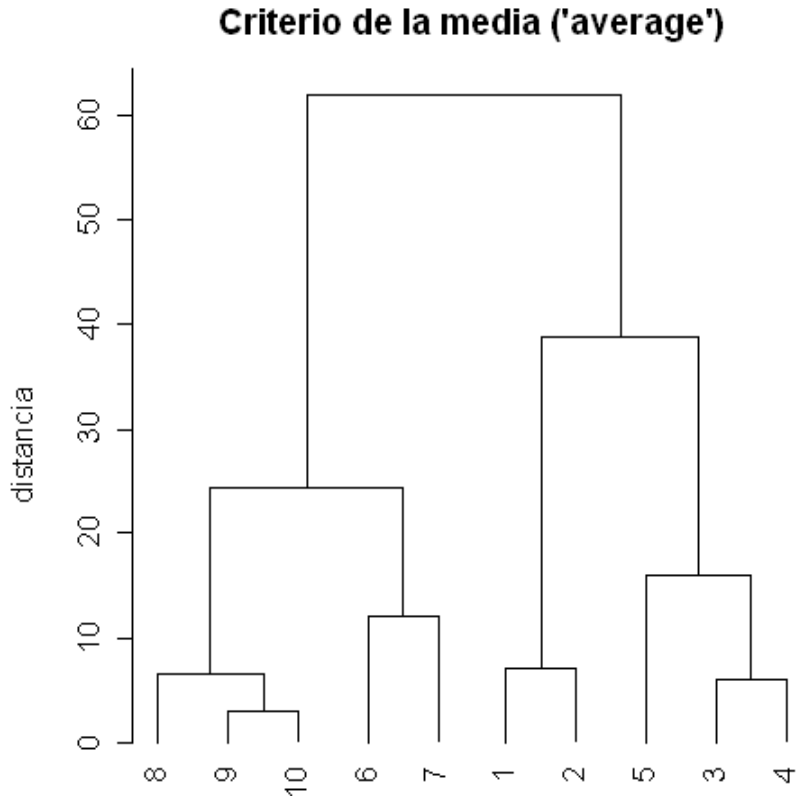
Orokorrean, emaitzak ez dira berberak



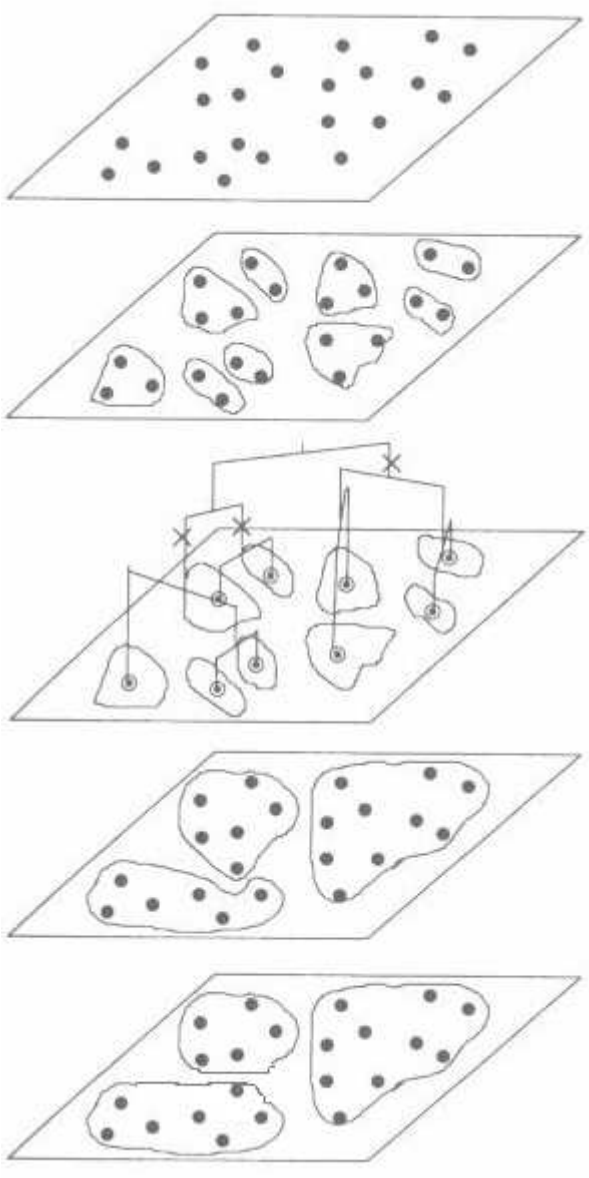
Hierarkiarako algoritmoa

Indexadun hierarkia





Datu-bolumen handia lantzeko sailakpen-estrategia



Datuak sailkapenaren aurretik

Hasierako partiketa azkarra: *k-means*

Partiketaren erdiguneen gaineko *hierarkia*

Dendrograma moztuz lortutako *partiketa*

Azkeneko partiketa: sendotu *berresleituz*

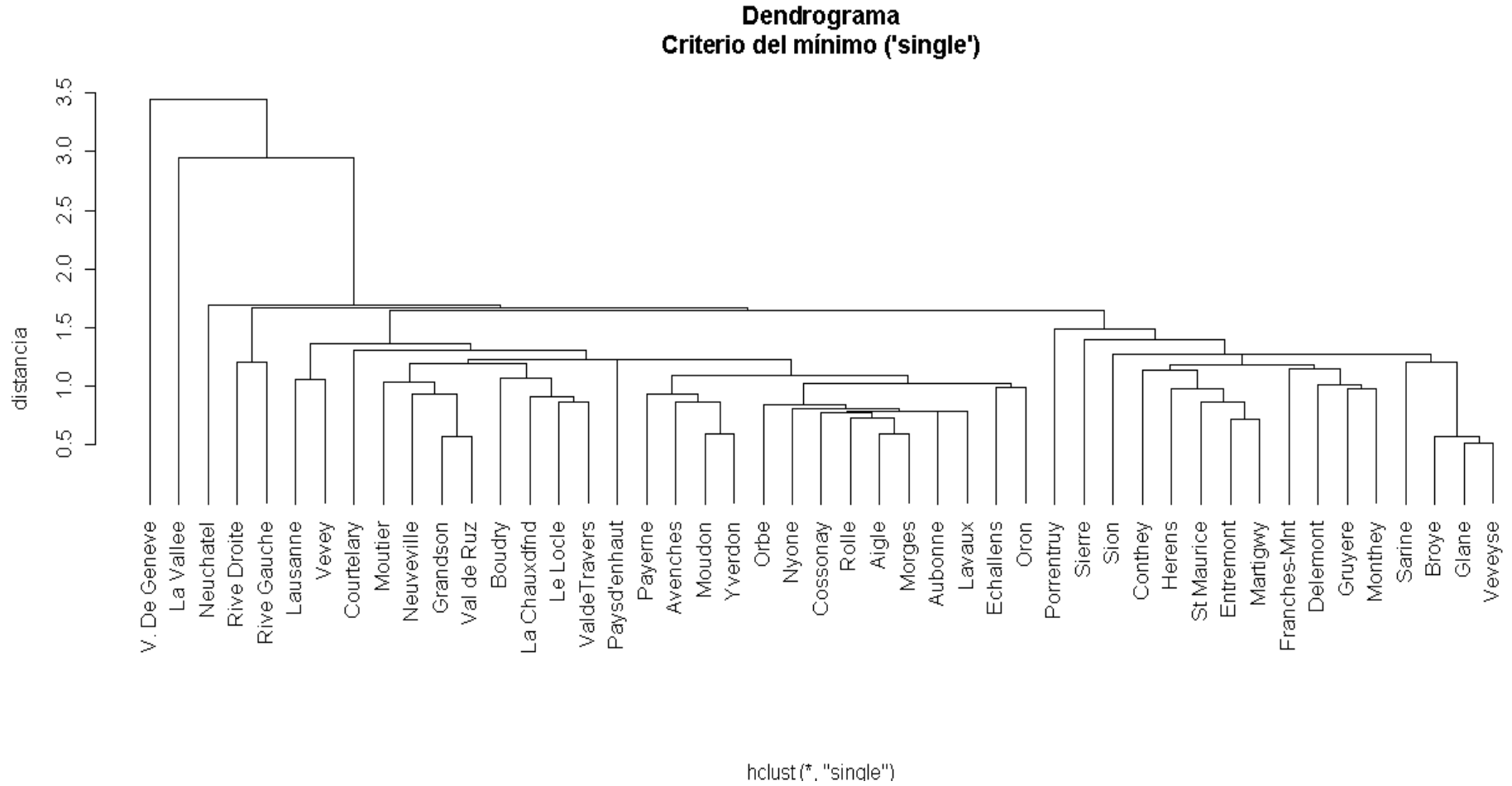
Suitza, 1888: 47 herrialde frankofono

Zenbait adierazle sozioekonomiko (%) eta emankortasuna

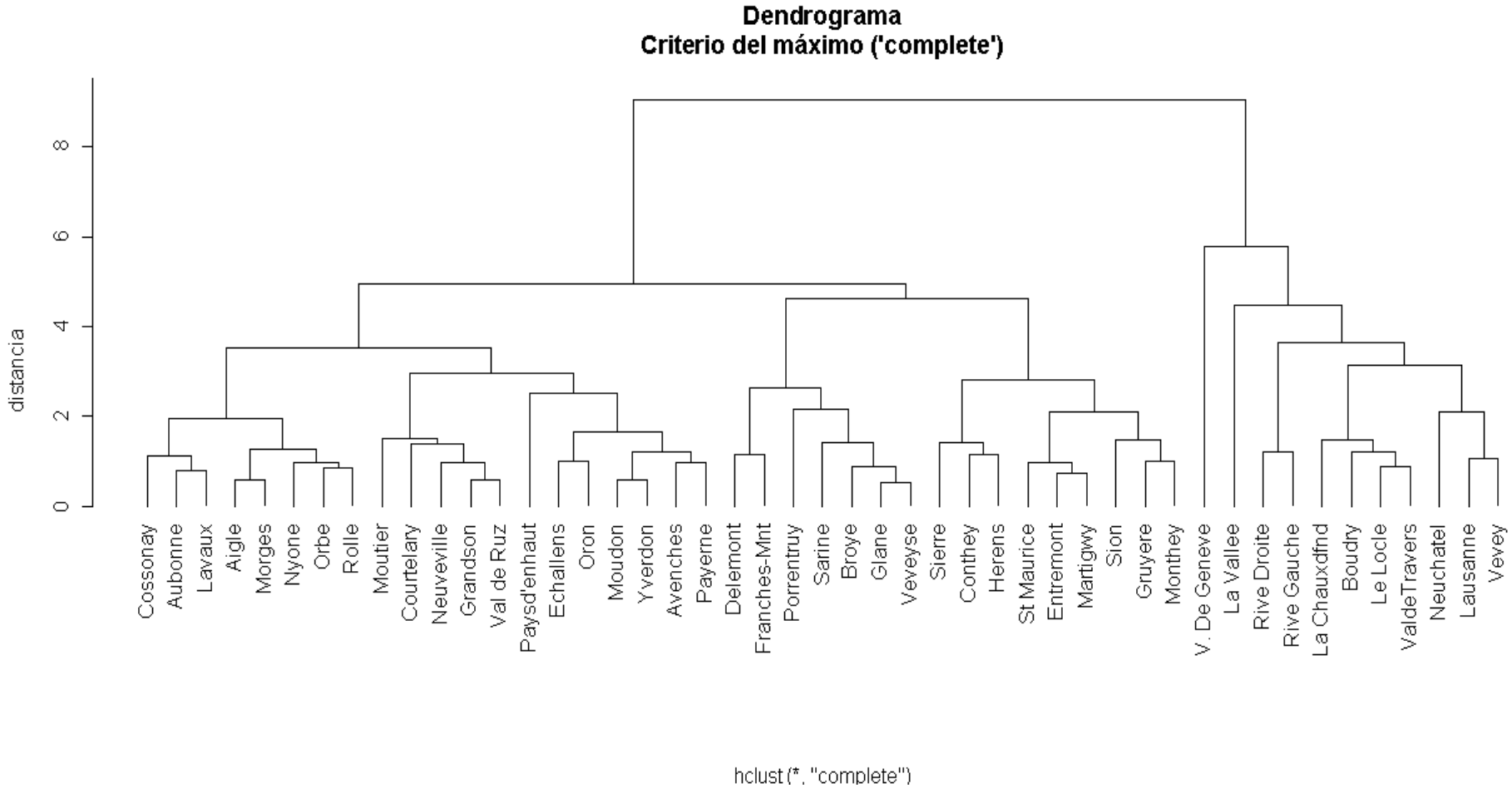
X1 Fertility X2 Agriculture X3 Examination
X4 Education X5 Catholic X6 Infant.Mortality

	X1	X2	X3	X4	X5	X6
Courtelary	80.2	17.0	15	12	9.96	22.2
Delemont	83.1	45.1	6	9	84.84	22.2
Franches-Mnt	92.5	39.7	5	5	93.40	20.2
Moutier	85.8	36.5	12	7	33.77	20.3
Neuveville	76.9	43.5	17	15	5.16	20.6
Porrentruy	76.1	35.3	9	7	90.57	26.6
Broye	83.8	70.2	16	7	92.85	23.6
Glane	92.4	67.8	14	8	97.16	24.9
Gruyere	82.4	53.3	12	7	97.67	21.0
Sarine	82.9	45.2	16	13	91.38	24.4
Veveyse	87.1	64.5	14	6	98.61	24.5
Aigle	64.1	62.0	21	12	8.52	16.5
Aubonne	66.9	67.5	14	7	2.27	19.1
Avenches	68.9	60.7	19	12	4.43	22.7
Cossonay	61.7	69.3	22	5	2.82	18.7
Echallens	68.3	72.6	18	2	24.20	21.2
Grandson	71.7	34.0	17	8	3.30	20.0
Lausanne	55.7	19.4	26	28	12.11	20.2
La Vallee	54.3	15.2	31	20	2.15	10.8
Lavaux	65.1	73.0	19	9	2.84	20.0
Morges	65.5	59.8	22	10	5.23	18.0
Moudon	65.0	55.1	14	3	4.52	22.4
Nyone	56.6	50.9	22	12	15.14	16.7
Orbe	57.4	54.1	20	6	4.20	15.3
Oron	72.5	71.2	12	1	2.40	21.0
Payerne	74.2	58.1	14	8	5.23	23.8
Paysd'enhaut	72.0	63.5	6	3	2.56	18.0
Rolle	60.5	60.8	16	10	7.72	16.3
Vevey	58.3	26.8	25	19	18.46	20.9
Yverdon	65.4	49.5	15	8	6.10	22.5
Conthey	75.5	85.9	3	2	99.71	15.1
Entremont	69.3	84.9	7	6	99.68	19.8
Herens	77.3	89.7	5	2	100.00	18.3
Martigwy	70.5	78.2	12	6	98.96	19.4
Monthey	79.4	64.9	7	3	98.22	20.2
St Maurice	65.0	75.9	9	9	99.06	17.8
Sierre	92.2	84.6	3	3	99.46	16.3
Sion	79.3	63.1	13	13	96.83	18.1
Boudry	70.4	38.4	26	12	5.62	20.3
La Chauxdfnd	65.7	7.7	29	11	13.79	20.5
Le Locle	72.7	16.7	22	13	11.22	18.9
Neuchatel	64.4	17.6	35	32	16.92	23.0
Val de Ruz	77.6	37.6	15	7	4.97	20.0
ValdeTravers	67.6	18.7	25	7	8.65	19.5
V. De Geneve	35.0	1.2	37	53	42.34	18.0
Rive Droite	44.7	46.6	16	29	50.43	18.2
Rive Gauche	42.8	27.7	22	29	58.33	19.3

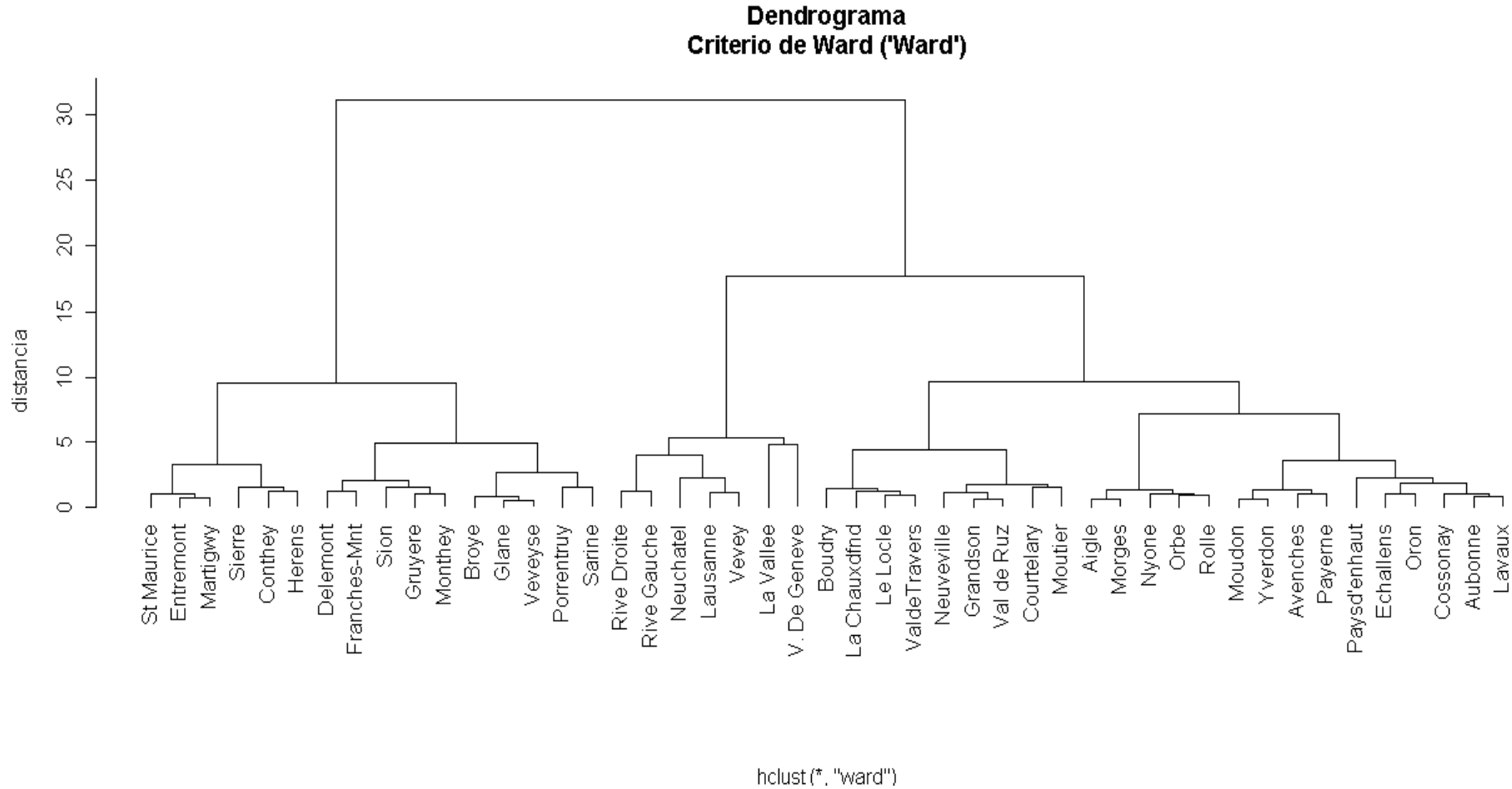
```
> dswiss <- dist(swiss)
> hsingle <- hclust(dswiss, method="single")
> plot(hsingle, hang=-1, xlab="", ylab="distancia",
      main="Dendrograma\nCriterio del mínimo ('single')")
```



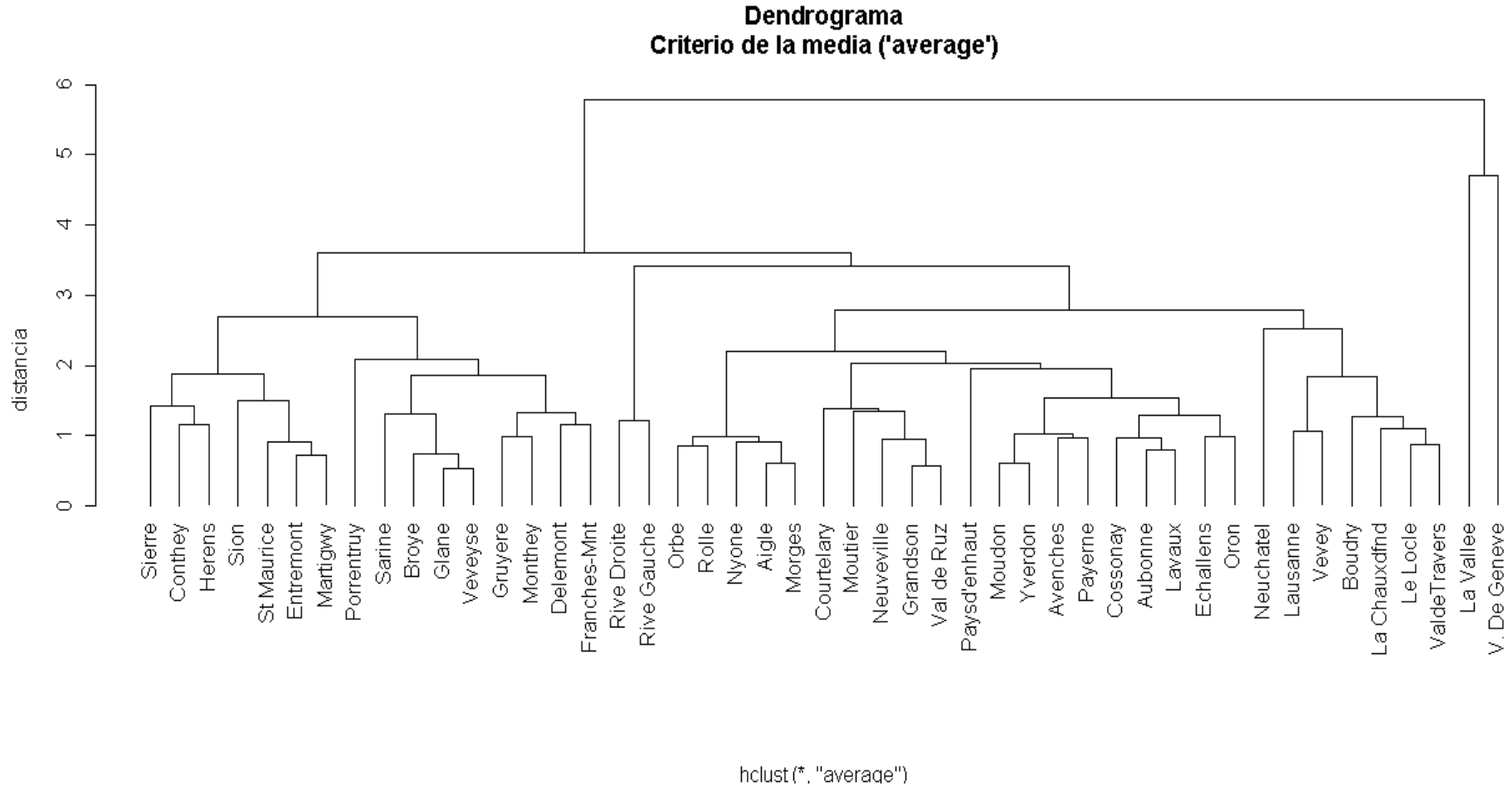
```
> dswiss <- dist(swiss)
> hcomplete <- hclust(dswiss, method="complete")
> plot(hcomplete, hang=-1, xlab="", ylab="distancia",
      main="Dendrograma\nCriterio del máximo ('complete')")
```



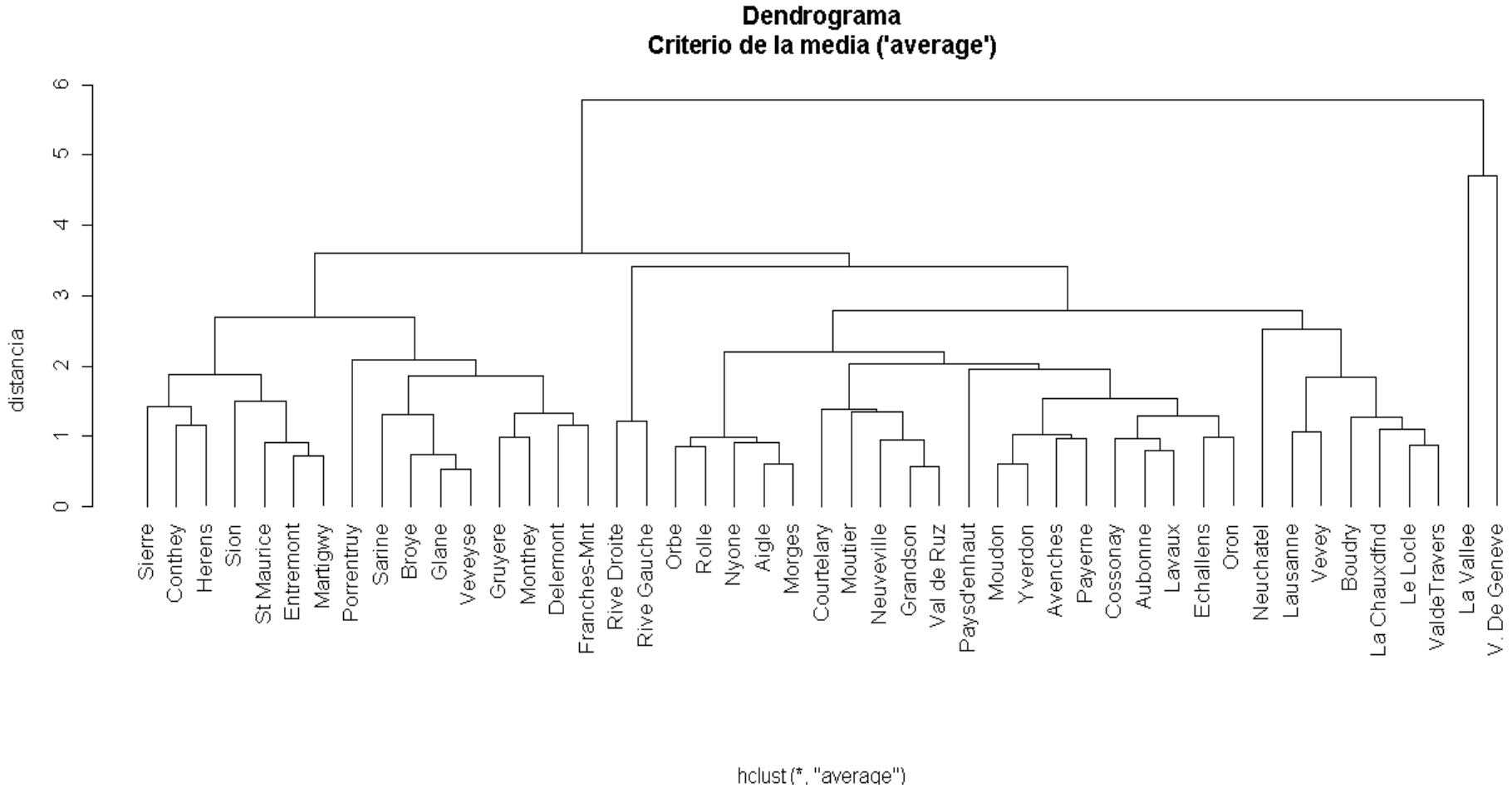
```
> dswiss <- dist(swiss)
> hward <- hclust(dswiss, method="ward")
> plot(hward, hang=-1, xlab="", ylab="distancia",
      main="Dendrograma\nCriterio de Ward ('ward')")
```



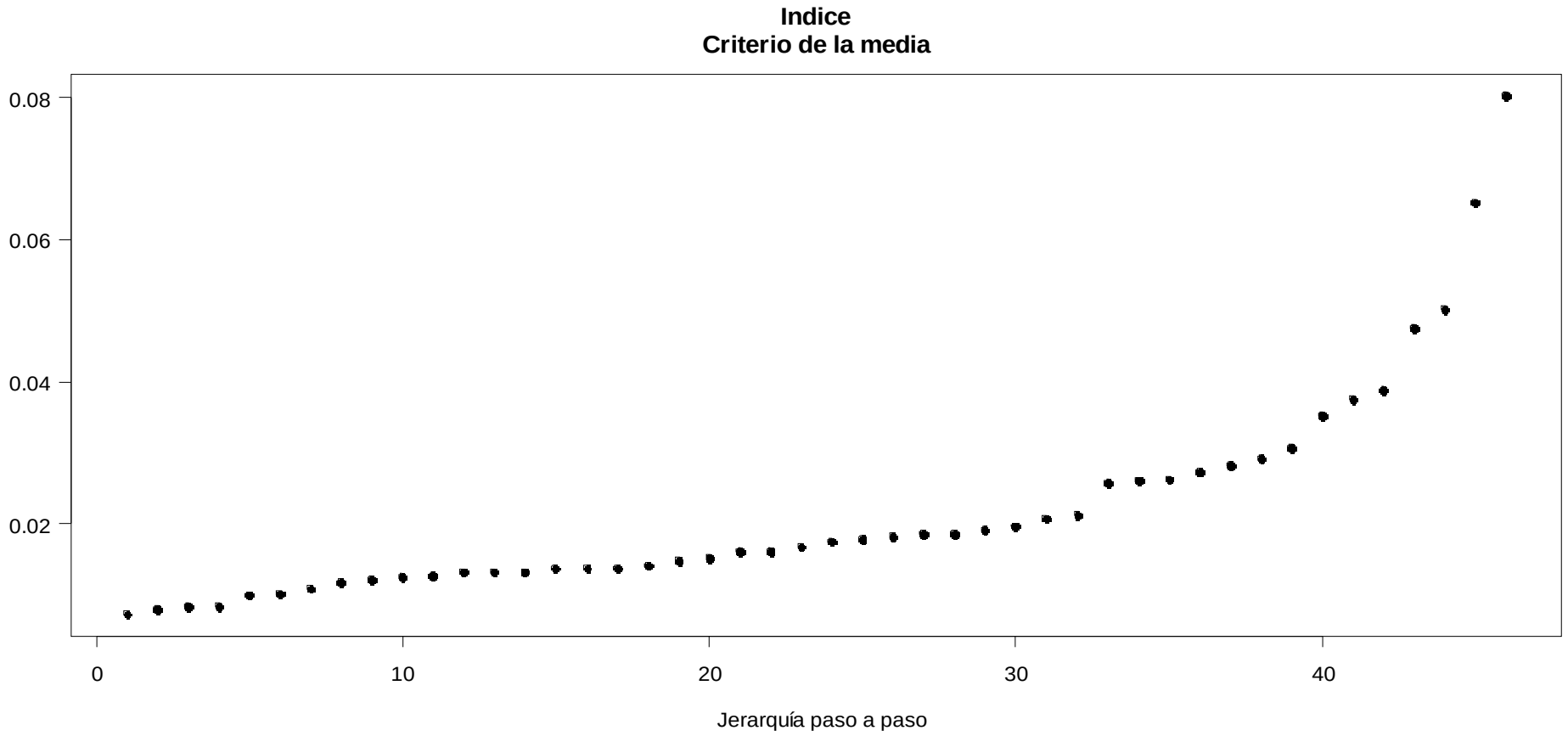
```
> dswiss <- dist(swiss)
> haverage <- hclust(dswiss, method="average")
> plot(haverage, hang=-1, xlab="", ylab="distancia",
      main="Dendrograma\nCriterio de la media ('average')")
```



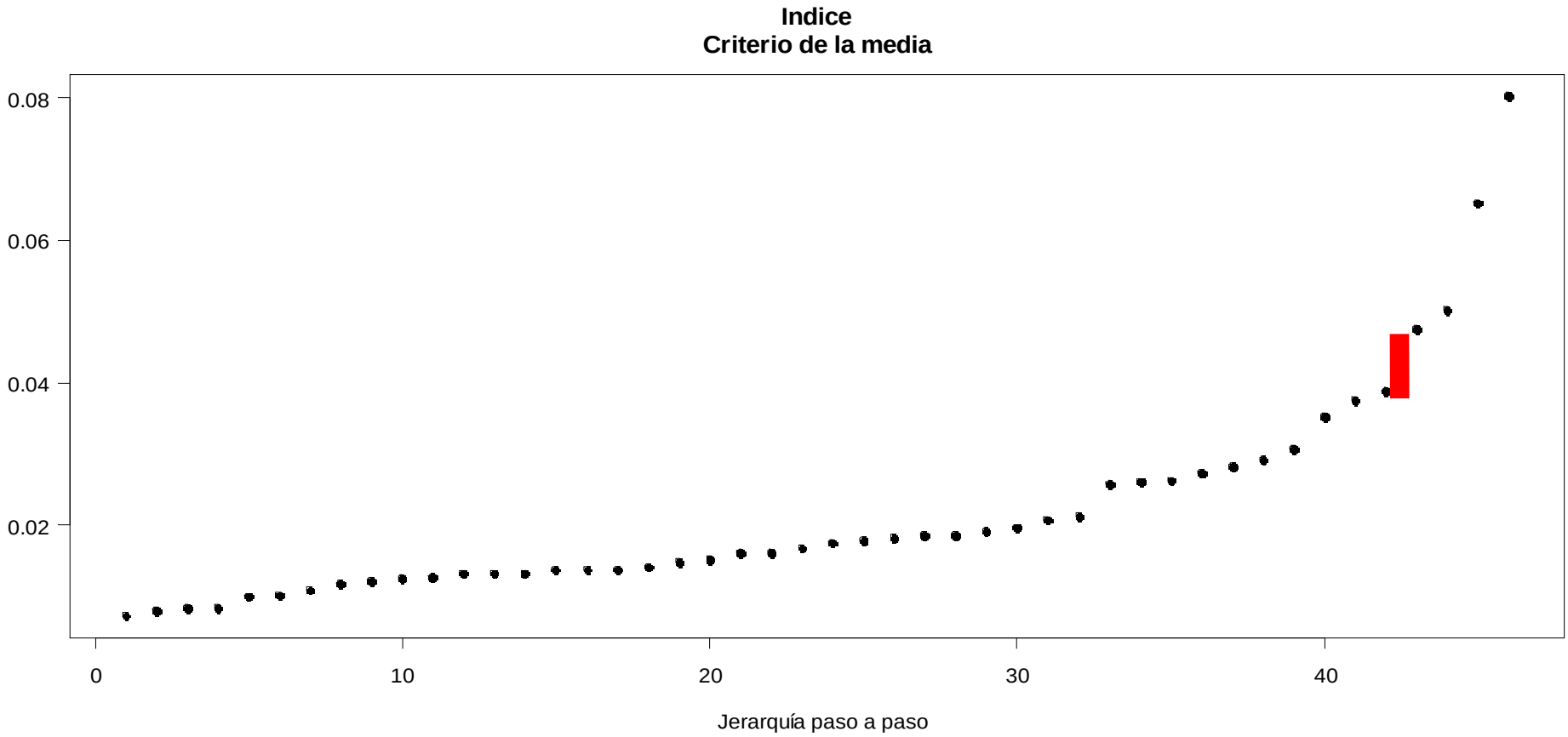
```
> dswiss <- dist(swiss)
> haverage <- hclust(dswiss, method="average")
> plot(haverage, hang=-1, xlab="", ylab="distancia",
      main="Dendrograma\nCriterio de la media ('average')")
```



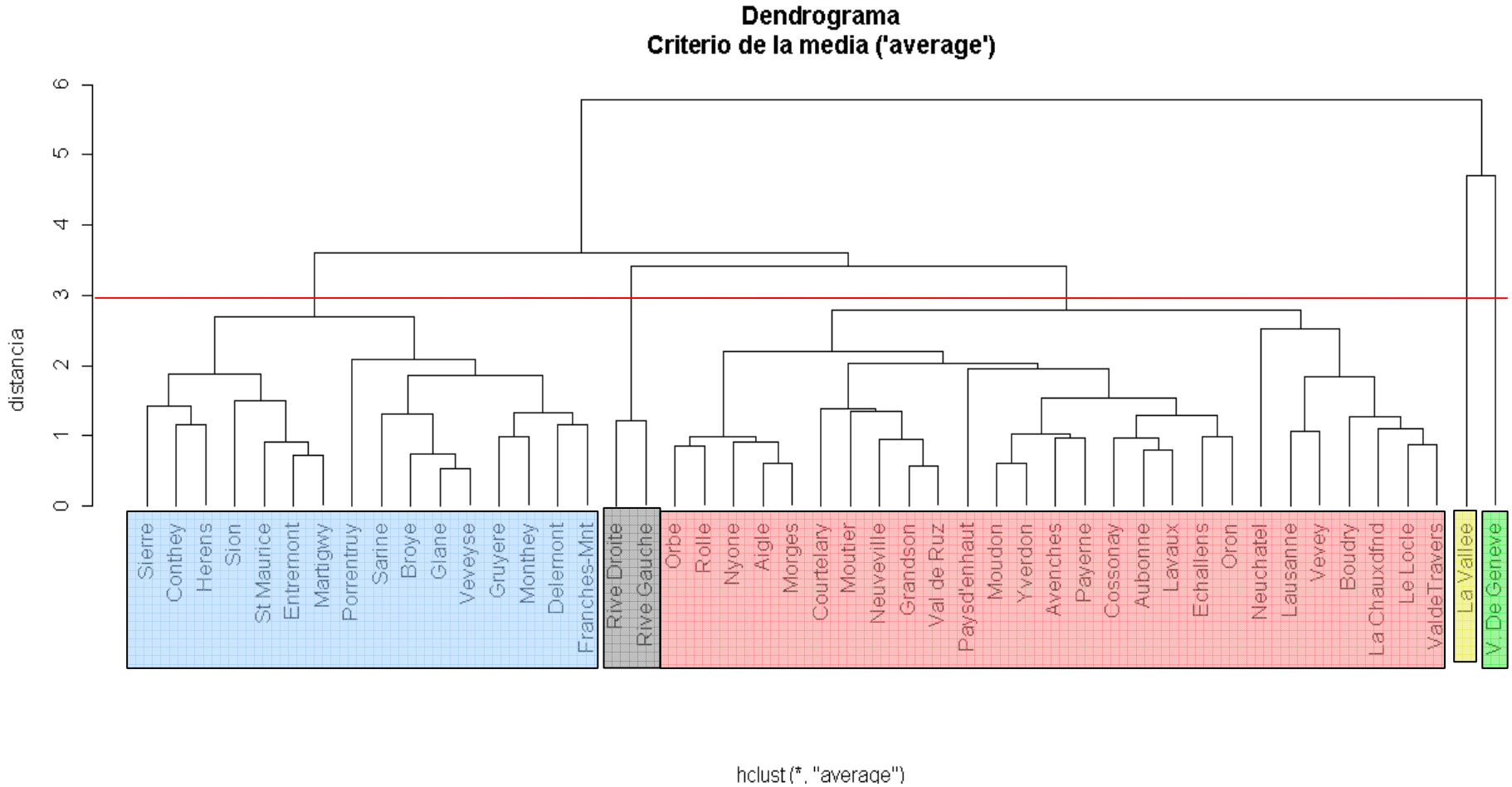
```
>  plot(haverage$height/sum(haverage$height), las=1, pch=19,
+     xlab="Jerarquía paso a paso", ylab="",
+     main="Indice\nCriterio de la media")
```



```
>  plot(haverage$height/sum(haverage$height), las=1, pch=19,
+     xlab="Jerarquía paso a paso", ylab="",
+     main="Indice\nCriterio de la media")
```



```
> dswiss <- dist(swiss)
> haverage <- hclust(dswiss, method="average")
> plot(haverage, hang=-1, xlab="", ylab="distancia",
      main="Dendrograma\nCriterio de la media ('average')")
```



```
> cuthaverage <- cutree(haverage, k=5)
> cuthaverage
> row.names(cutav) <- names(cuthaverage)
> cutav
```

Courtelary	1	Payerne	1
Delemont	2	Paysd'enhaut	1
Franches-Mnt	2	Rolle	1
Moutier	1	Vevey	1
Neuveville	1	Yverdon	1
Porrentruy	2	Conthey	2
Broye	2	Entremont	2
Glane	2	Herens	2
Gruyere	2	Martigwy	2
Sarine	2	Monthey	2
Veveyse	2	St Maurice	2
Aigle	1	Sierre	2
Aubonne	1	Sion	2
Avenches	1	Boudry	1
Cossonay	1	La Chauxdfnd	1
Echallens	1	Le Locle	1
Grandson	1	Neuchatel	1
Lausanne	1	Val de Ruz	1
La Vallee	3	ValdeTravers	1
Lavaux	1	V. De Geneve	4
Morges	1	Rive Droite	5
Moudon	1	Rive Gauche	5
Nyone	1		
Orbe	1		
Oron	1		

Balidazioa

```
> pca <- princomp(swiss, cor = TRUE, scores = TRUE)
```

Varianzas de las componentes construidas

```
> pca$sdev**2
    Comp.1    Comp.2    Comp.3
3.1997570 1.1883082 0.8476098
    Comp.4    Comp.5    Comp.6
0.4389287 0.2045337 0.1208626
> sum(pca$sdev**2)
[1] 6
```

Porcentajes de las *varianzas* de las componentes

```
> pca$sdev**2/sum(pca$sdev**2)
    Comp.1     Comp.2     Comp.3
0.53329283 0.19805137 0.14126830
    Comp.4     Comp.5     Comp.6
0.07315478 0.03408895 0.02014376
> sum((pca$sdev**2/sum(pca$sdev**2))[1:3])
[1] 0.8726125
```

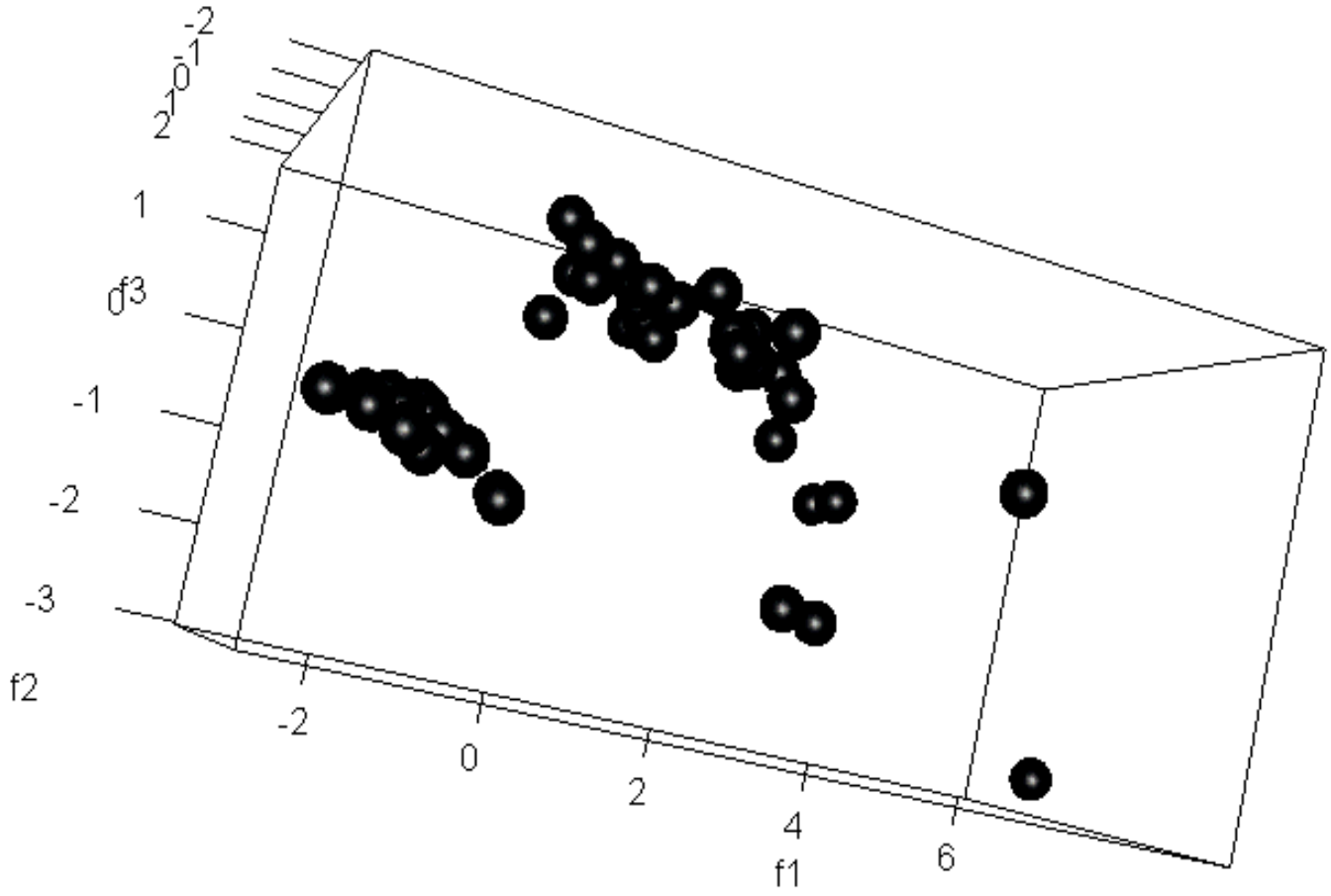
3 componentes acumulan el **87.26%** de la *varianza total*

```
> pca <- princomp(swiss, cor = TRUE, scores = TRUE)
> attributes(pca)
$names
[1] "sdev"     "loadings" "center"   "scale"
[2] "n.obs"    "scores"   "call"
$class
[1] "princomp"
> round(pca$scores, digits=2)
```

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6		Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Courtelary	0.36	1.40	0.86	0.91	-0.63	-0.28	Payerne	-0.49	0.83	1.10	-0.78	-0.28	-0.37
Delemont	-1.63	1.03	-0.55	0.51	-0.65	-0.09	Paysd'enhaut	-0.94	-1.19	0.98	0.13	-0.85	-0.40
Franches-Mnt	-2.10	0.75	-0.47	1.50	-0.41	-0.09	Rolle	0.56	-1.61	0.35	-0.11	-0.19	-0.13
Moutier	-0.75	0.60	0.58	1.07	-0.23	-0.31	Vevey	1.96	0.60	0.02	-0.20	-0.08	0.25
Neuveville	0.38	0.45	0.63	0.25	-0.07	-0.77	Yverdon	0.12	0.41	0.91	-0.68	-0.55	0.09
Porrentruy	-1.37	2.29	-0.35	-0.31	-0.84	0.71	Conthey	-2.41	-2.04	-1.05	0.44	-0.13	0.02
Broye	-1.72	1.13	-0.43	-0.47	0.83	0.08	Entremont	-1.96	-0.73	-1.05	-0.73	-0.13	0.22
Glane	-2.18	1.76	-0.40	-0.25	0.81	-0.27	Herens	-2.59	-1.13	-0.84	-0.19	0.08	0.04
Gruyere	-1.52	0.62	-0.71	0.47	0.20	0.23	Martigwy	-1.52	-0.61	-0.98	-0.45	0.32	0.39
Sarine	-0.96	1.90	-0.66	0.00	0.30	0.03	Monthey	-2.10	-0.05	-0.66	0.23	-0.15	0.31
Veveyse	-2.01	1.54	-0.41	-0.29	0.63	0.12	St Maurice	-1.24	-1.15	-1.34	-0.34	-0.16	0.29
Aigle	0.81	-1.36	0.34	-0.08	0.44	-0.28	Sierre	-3.01	-1.23	-0.80	0.97	0.31	-0.65
Aubonne	-0.18	-0.88	0.83	-0.50	-0.19	-0.29	Sion	-1.08	-0.33	-1.24	0.52	0.43	-0.27
Avenches	0.24	0.49	0.81	-0.96	0.12	-0.38	Boudry	1.17	0.38	0.79	0.13	0.61	0.02
Cossonay	0.42	-1.07	0.93	-0.74	0.54	0.25	La Chauxdfnd	1.99	0.93	0.72	0.78	0.29	0.86
Echallens	-0.60	-0.30	0.86	-0.90	0.41	0.32	Le Locle	1.31	0.41	0.57	1.14	-0.03	0.06
Grandson	0.47	0.18	0.97	0.47	-0.32	0.03	Neuchatel	3.08	1.92	-0.34	-0.26	0.80	-0.46
Lausanne	2.78	0.64	-0.41	-0.09	-0.25	-0.24	Val de Ruz	-0.01	0.22	1.04	0.65	-0.27	-0.21
La Vallee	3.43	-2.08	-0.24	1.71	0.51	0.24	ValdeTravers	1.36	0.34	0.95	0.79	0.17	0.73
Lavaux	0.15	-0.66	0.78	-0.96	0.32	-0.27	V. De Geneve	5.66	0.56	-2.66	-0.32	-0.04	-0.44
Morges	0.71	-0.89	0.63	-0.22	0.50	-0.12	Rive Droite	1.86	-0.71	-1.64	-0.83	-0.94	-0.23
Moudon	-0.25	0.16	1.19	-0.79	-0.53	0.34	Rive Gauche	2.55	0.01	-1.66	-0.59	-0.70	0.38
Nyone	1.29	-1.26	0.14	-0.11	0.15	0.26							
Orbe	0.95	-1.87	0.60	0.14	0.07	0.42							
Oron	-0.97	-0.41	1.32	-0.64	-0.18	-0.14							

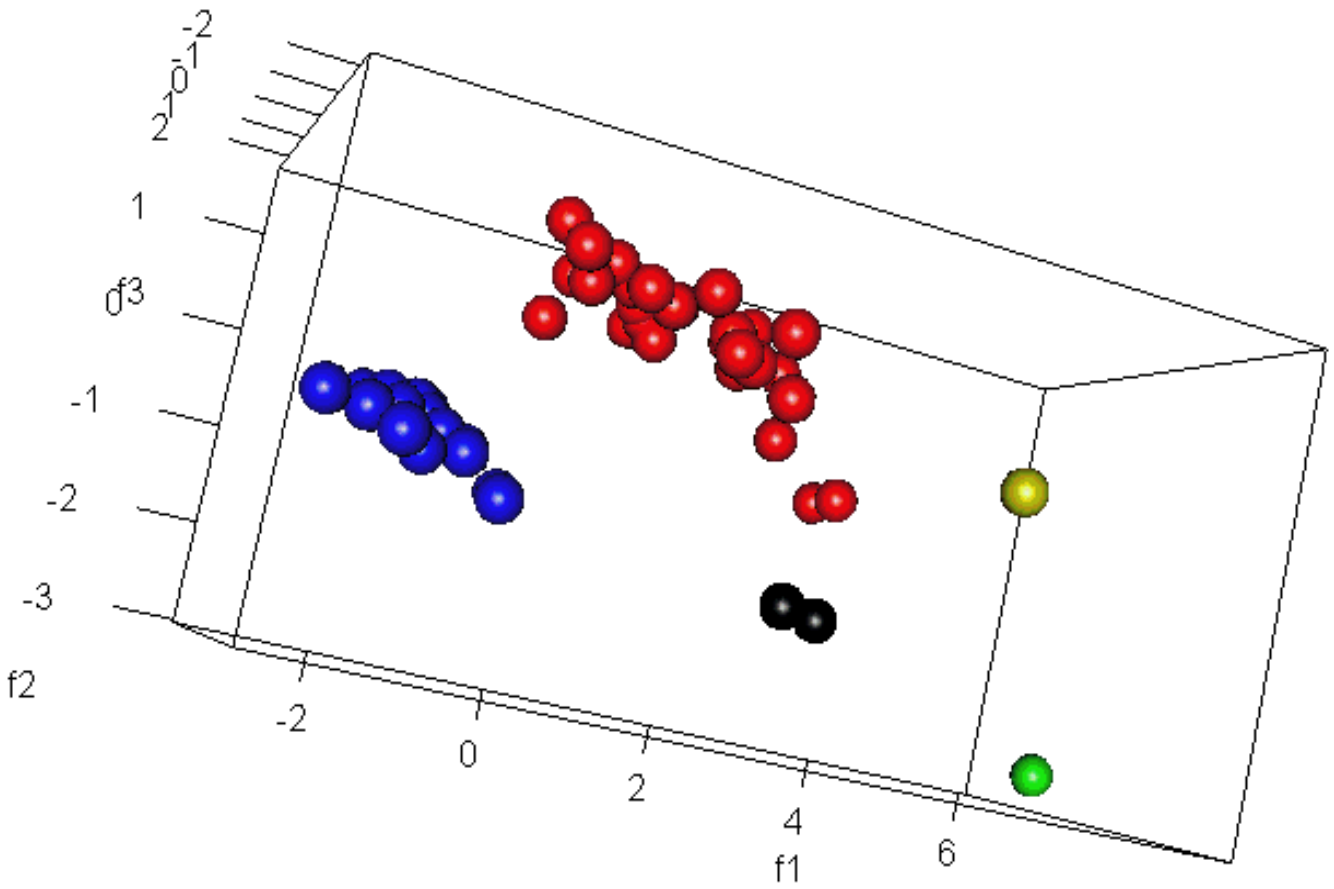
Balidazioa

```
>  plot3d(pca$scores[,1],pca$scores[,2],pca$scores[,3],
+      xlab="f1",ylab="f2",zlab="f3",
+      aspect=sd(pca$scores[,1:3]),
+      type = "s", size=1, col="black")
```



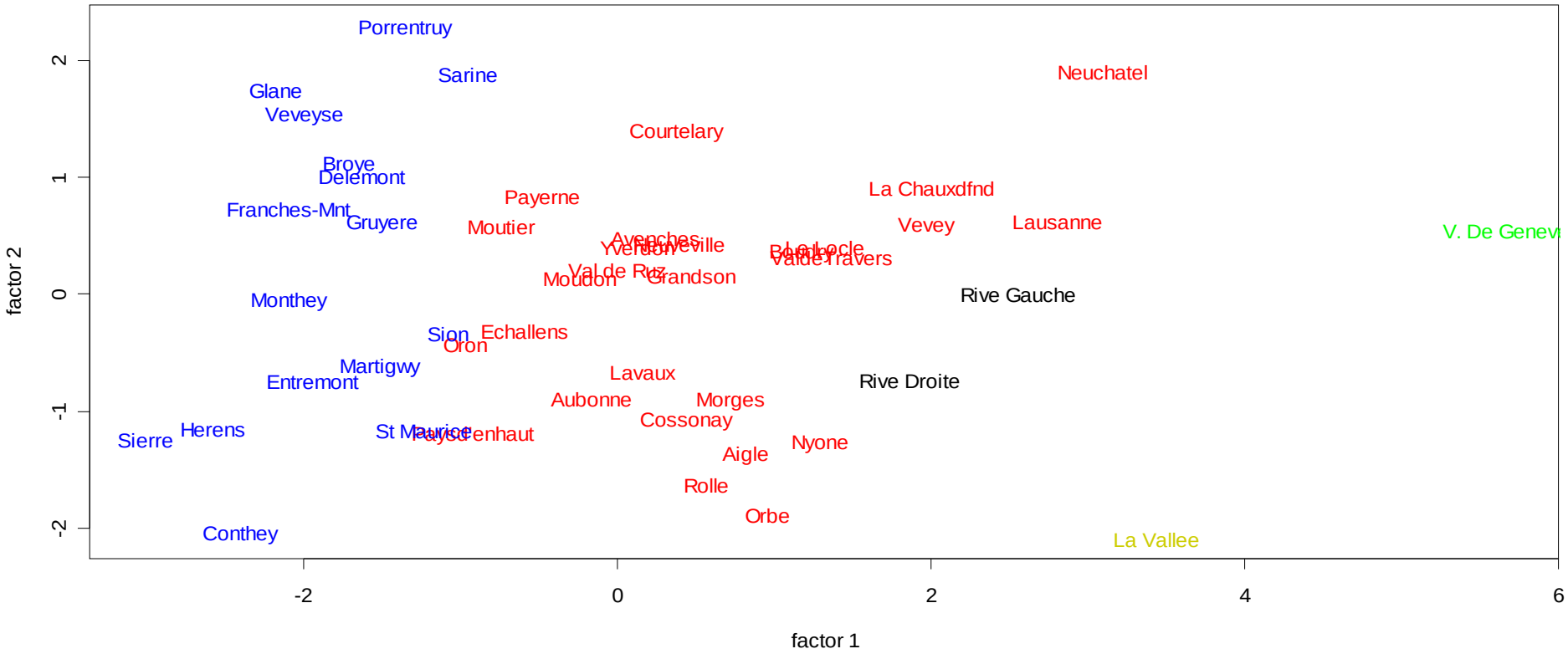
Balidazioa

```
> colores <- c("red","blue","yellow3","green","black")
> plot3d(pca$scores[,1],pca$scores[,2],pca$scores[,3],
+     xlab="f1",ylab="f2",zlab="f3",
+     aspect=sd(pca$scores[,1:3]),
+     type = "s", size=1, col=colores[cuthaverage])
```



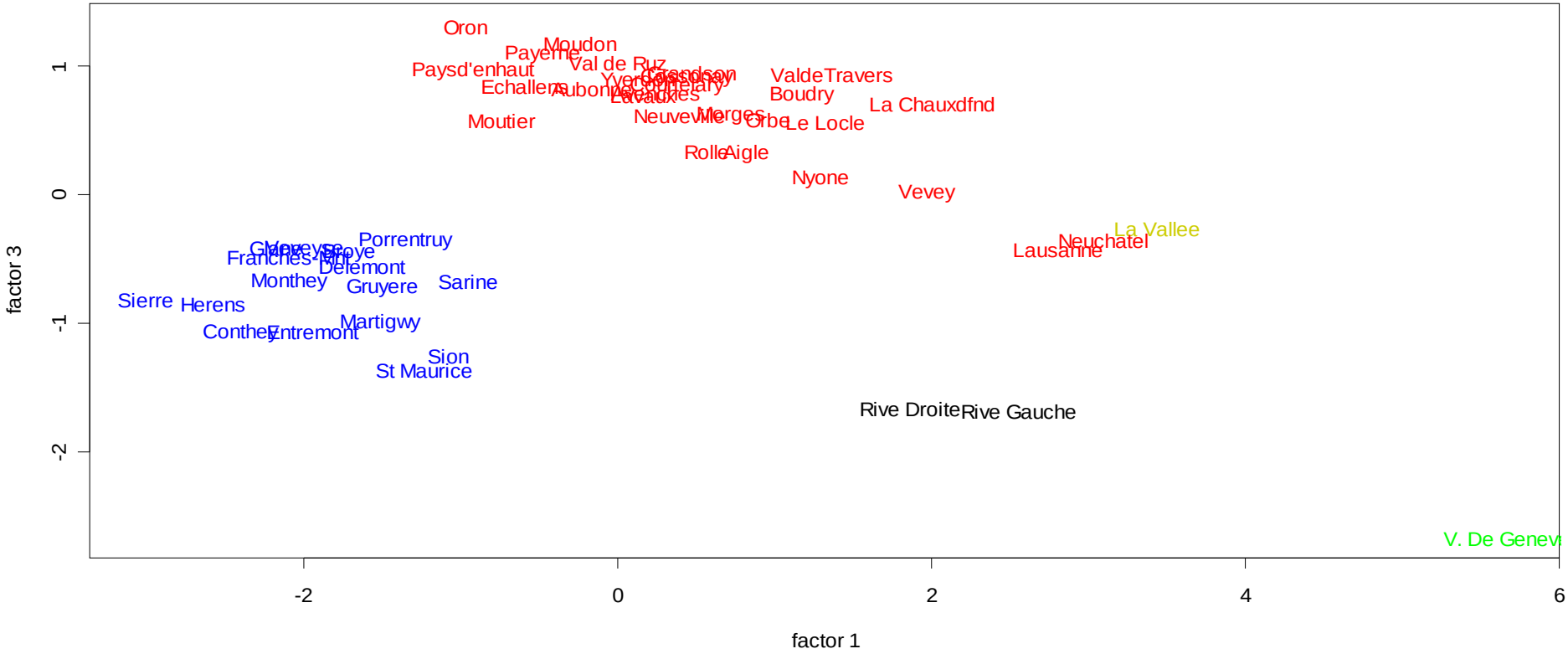
Balidazioa

```
> colores <- c("red","blue","green","cornsilk","black")
>  plot(pca$scores[,1],pca$scores[,2], type="n",
        xlab="factor 1", ylab="factor 2", main="")
> text(pca$scores[,1],pca$scores[,2], labels=row.names(swiss),
     col=colores[cuthaverage])
```



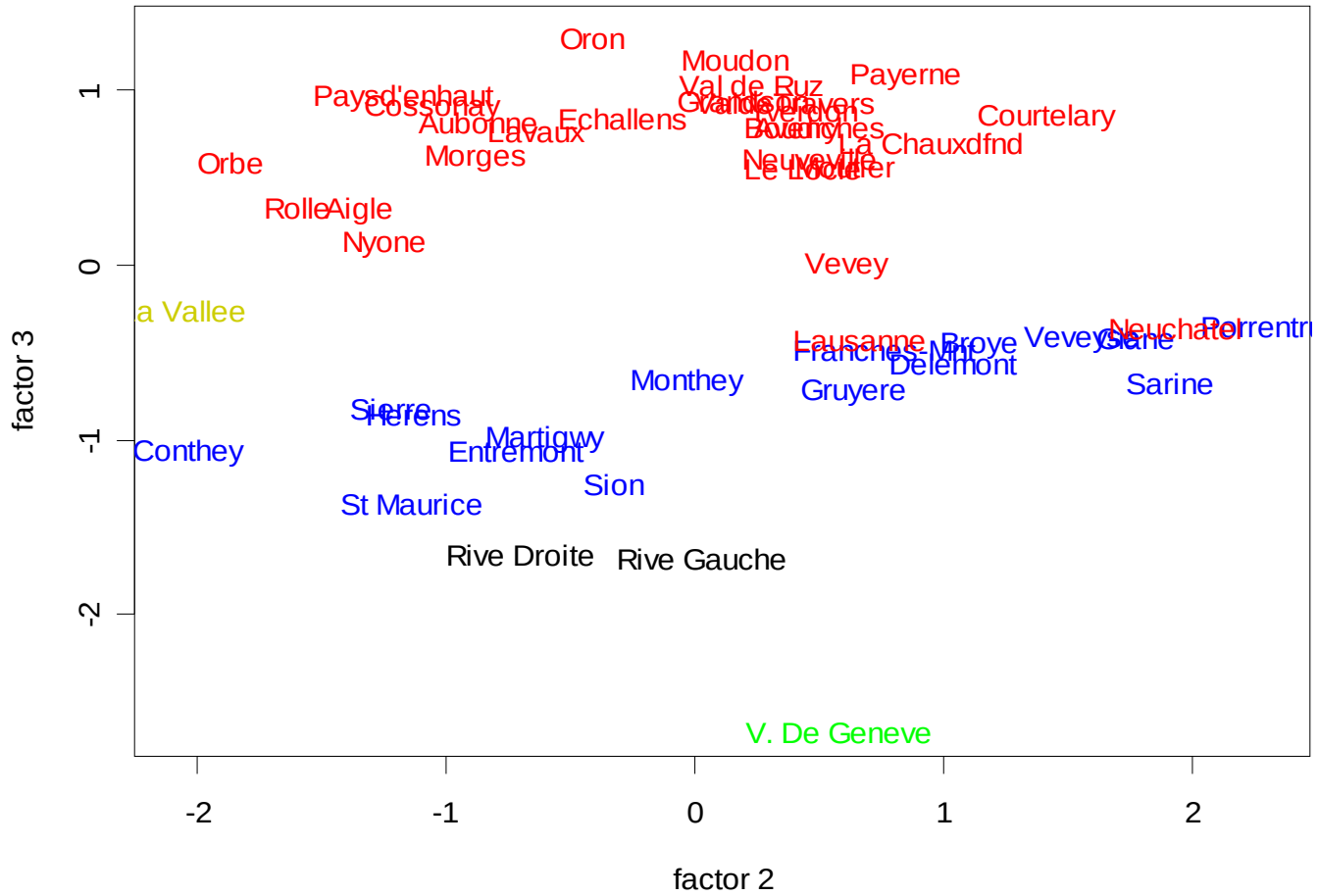
Balidazioa

```
> colores <- c("red","blue","green","cornsilk","black")
>  plot(pca$scores[,1],pca$scores[,3], type="n",
        xlab="factor 1", ylab="factor 3", main="")
> text(pca$scores[,1],pca$scores[,3], labels=row.names(swiss),
     col=colores[cuthaverage])
```



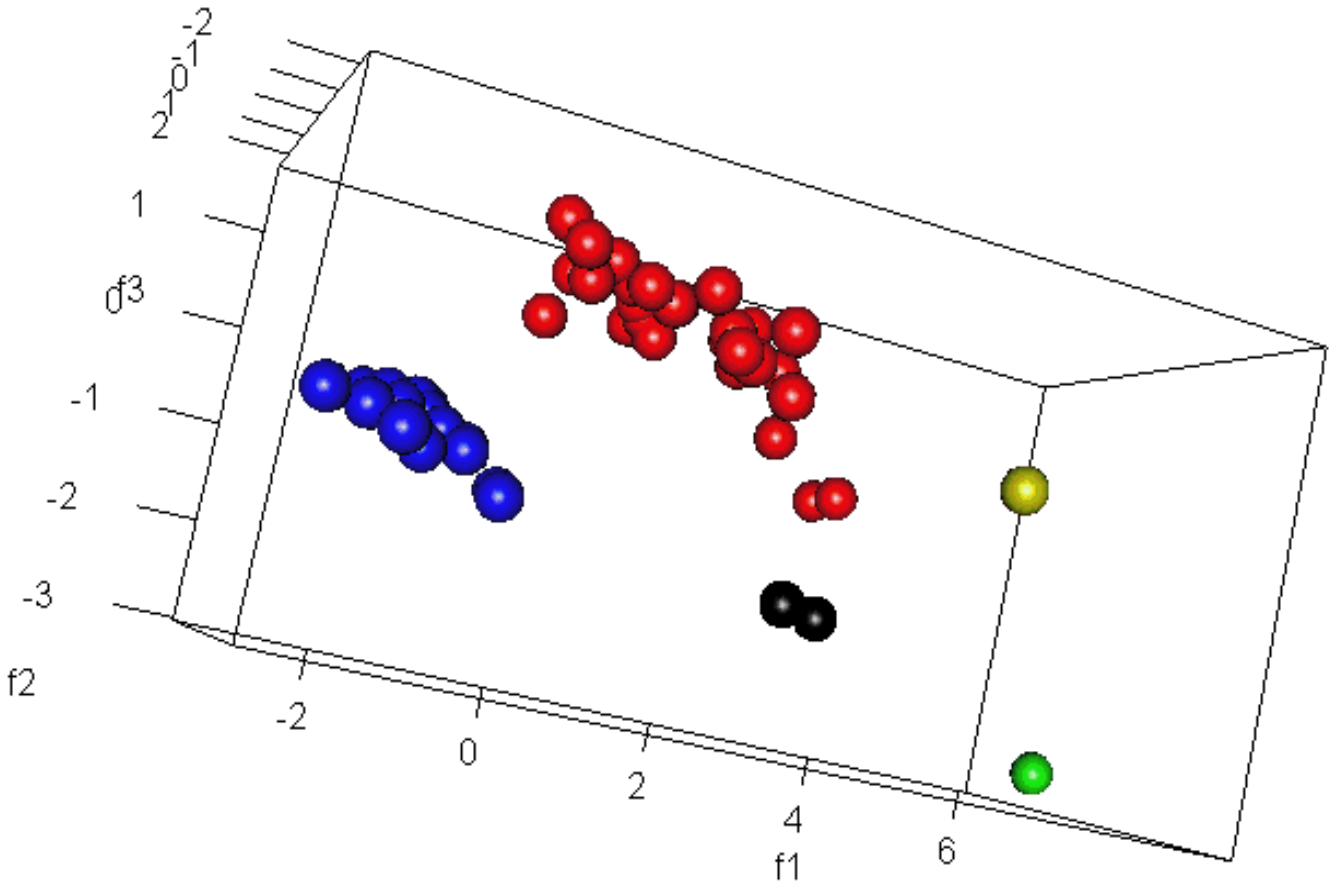
Balidazioa

```
> colores <- c("red","blue","green","cornsilk","black")
>  plot(pca$scores[,2],pca$scores[,3], type="n",
        xlab="factor 2", ylab="factor 3", main="")
> text(pca$scores[,2],pca$scores[,3], labels=row.names(swiss),
     col=colores[cuthaverage])
```



Balidazioa

```
> colores <- c("red","blue","yellow3","green","black")
> plot3d(pca$scores[,1],pca$scores[,2],pca$scores[,3],
+     xlab="f1",ylab="f2",zlab="f3",
+     aspect=sd(pca$scores[,1:3]),
+     type = "s", size=1, col=colores[cuthaverage])
```



Interpretazioa

Datu-taularen lerroak eta zutabeak berrordenatu **factor nagusia**ren arabera.

```
> order(pca$scores[,1])
 [1] 37 33 31  8  3 35 11 32  7  2 34  9  6 36 38 25
[17] 10 27  4 16 26 22 13 43 30 20 14  1  5 15 17 28
[27] 21 12 24 39 23 41 44 46 29 40 47 18 42 19 45
> order(pca$loadings[,1])
[1] 1 2 5 6 4 3
> cutav2 <- cutav[order(pca$scores[,1])]
> cutav2
 [1] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 1 1 1 1 1 1
[25] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 5 1 1 5 1 1 3 4
> round(cor(swiss,pca$scores),digits=3)
                  Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
Fertility         -0.817  0.351  0.160  0.355  0.173 -0.164
Agriculture       -0.759 -0.449 -0.035 -0.426  0.170 -0.107
Examination        0.912  0.136  0.084 -0.036  0.368  0.078
Education          0.813  0.195 -0.490 -0.065 -0.032 -0.237
Catholic          -0.626  0.159 -0.743  0.066  0.083  0.140
Infant.Mortality -0.268  0.884  0.147 -0.349 -0.047  0.026
```

Interpretazioa

Datu-taularen lerroak eta zutabeak berrordenatu **factor nagusia**ren arabera.

	Fertility	Agriculture	Catholic	Infant.Mortality	Education	Examination	cluster5
Sierre	92.2	84.6	99.46	16.3	3	3	2
Herens	77.3	89.7	100.00	18.3	2	5	2
Conthey	75.5	85.9	99.71	15.1	2	3	2
Glane	92.4	67.8	97.16	24.9	8	14	2
Franches-Mnt	92.5	39.7	93.40	20.2	5	5	2
Monthey	79.4	64.9	98.22	20.2	3	7	2
Veveyse	87.1	64.5	98.61	24.5	6	14	2
Entremont	69.3	84.9	99.68	19.8	6	7	2
Broye	83.8	70.2	92.85	23.6	7	16	2
Delemont	83.1	45.1	84.84	22.2	9	6	2
Martigwy	70.5	78.2	98.96	19.4	6	12	2
Gruyere	82.4	53.3	97.67	21.0	7	12	2
Porrentruy	76.1	35.3	90.57	26.6	7	9	2
St Maurice	65.0	75.9	99.06	17.8	9	9	2
Sion	79.3	63.1	96.83	18.1	13	13	2
Oron	72.5	71.2	2.40	21.0	1	12	1
Sarine	82.9	45.2	91.38	24.4	13	16	2

Interpretazioa

Datu-taularen lerroak eta zutabeak berrordenatu **factor nagusia**ren arabera.

	Fertility	Agriculture	Catholic	Infant.Mortality	Education	Examination	cluster5
Paysd'enhaut	72.0	63.5	2.56	18.0	3	6	1
Moutier	85.8	36.5	33.77	20.3	7	12	1
Echallens	68.3	72.6	24.20	21.2	2	18	1
Payerne	74.2	58.1	5.23	23.8	8	14	1
Moudon	65.0	55.1	4.52	22.4	3	14	1
Aubonne	66.9	67.5	2.27	19.1	7	14	1
Val de Ruz	77.6	37.6	4.97	20.0	7	15	1
Yverdon	65.4	49.5	6.10	22.5	8	15	1
Lavaux	65.1	73.0	2.84	20.0	9	19	1
Avenches	68.9	60.7	4.43	22.7	12	19	1
Courtelary	80.2	17.0	9.96	22.2	12	15	1
Neuveville	76.9	43.5	5.16	20.6	15	17	1
Cossonay	61.7	69.3	2.82	18.7	5	22	1
Grandson	71.7	34.0	3.30	20.0	8	17	1
Rolle	60.5	60.8	7.72	16.3	10	16	1
Morges	65.5	59.8	5.23	18.0	10	22	1
Aigle	64.1	62.0	8.52	16.5	12	21	1
Orbe	57.4	54.1	4.20	15.3	6	20	1
Boudry	70.4	38.4	5.62	20.3	12	26	1
Nyone	56.6	50.9	15.14	16.7	12	22	1
Le Locle	72.7	16.7	11.22	18.9	13	22	1
ValdeTravers	67.6	18.7	8.65	19.5	7	25	1
Rive Droite	44.7	46.6	50.43	18.2	29	16	5
Vevey	58.3	26.8	18.46	20.9	19	25	1
La Chauxdfnd	65.7	7.7	13.79	20.5	11	29	1
Rive Gauche	42.8	27.7	58.33	19.3	29	22	5
Lausanne	55.7	19.4	12.11	20.2	28	26	1
Neuchatel	64.4	17.6	16.92	23.0	32	35	1
La Vallee	54.3	15.2	2.15	10.8	20	31	3
V. De Geneve	35.0	1.2	42.34	18.0	53	37	4

Interpretazioa

Datu-taularen lerroak eta zutabeak berrordenatu **factor nagusia**ren arabera.

	Fertility	Agriculture	Catholic	Infant.Mortality	Education	Examination	cluster5
Sierre	92.2	84.6	99.46	16.3	3	3	2
Herens	77.3	89.7	100.00	18.3	2	5	2
Conthey	75.5	85.9	99.71	15.1	2	3	2
Glane	92.4	67.8	97.16	24.9	8	14	2
Franches-Mnt	92.5	39.7	93.40	20.2	5	5	2
Monthey	79.4	64.9	98.22	20.2	3	7	2
Veveyse	87.1	64.5	98.61	24.5	6	14	2
Entremont	69.3	84.9	99.68	19.8	6	7	2
Broye	83.8	70.2	92.85	23.6	7	16	2
Delemont	83.1	45.1	84.84	22.2	9	6	2
Martigwy	70.5	78.2	98.96	19.4	6	12	2
Gruyere	82.4	53.3	97.67	21.0	7	12	2
Porrentruy	76.1	35.3	90.57	26.6	7	9	2
St Maurice	65.0	75.9	99.06	17.8	9	9	2
Sion	79.3	63.1	96.83	18.1	13	13	2
Sarine	82.9	45.2	91.38	24.4	13	16	2

Interpretazioa

Datu-taularen lerroak eta zutabeak berrordenatu **factor nagusia**ren arabera.

	Fertility	Agriculture	Catholic	Infant.Mortality	Education	Examination	cluster5
Oron	72.5	71.2	2.40	21.0	1	12	1
Paysd'enhaut	72.0	63.5	2.56	18.0	3	6	1
Moutier	85.8	36.5	33.77	20.3	7	12	1
Echallens	68.3	72.6	24.20	21.2	2	18	1
Payerne	74.2	58.1	5.23	23.8	8	14	1
Moudon	65.0	55.1	4.52	22.4	3	14	1
Aubonne	66.9	67.5	2.27	19.1	7	14	1
Val de Ruz	77.6	37.6	4.97	20.0	7	15	1
Yverdon	65.4	49.5	6.10	22.5	8	15	1
Lavaux	65.1	73.0	2.84	20.0	9	19	1
Avenches	68.9	60.7	4.43	22.7	12	19	1
Courtelary	80.2	17.0	9.96	22.2	12	15	1
Neuveville	76.9	43.5	5.16	20.6	15	17	1
Cossonay	61.7	69.3	2.82	18.7	5	22	1
Grandson	71.7	34.0	3.30	20.0	8	17	1
Rolle	60.5	60.8	7.72	16.3	10	16	1
Morges	65.5	59.8	5.23	18.0	10	22	1
Aigle	64.1	62.0	8.52	16.5	12	21	1
Orbe	57.4	54.1	4.20	15.3	6	20	1
Boudry	70.4	38.4	5.62	20.3	12	26	1
Nyone	56.6	50.9	15.14	16.7	12	22	1
Le Locle	72.7	16.7	11.22	18.9	13	22	1
ValdeTravers	67.6	18.7	8.65	19.5	7	25	1
Vevey	58.3	26.8	18.46	20.9	19	25	1
La Chauxdfnd	65.7	7.7	13.79	20.5	11	29	1
Lausanne	55.7	19.4	12.11	20.2	28	26	1
Neuchatel	64.4	17.6	16.92	23.0	32	35	1
Rive Droite	44.7	46.6	50.43	18.2	29	16	5
Rive Gauche	42.8	27.7	58.33	19.3	29	22	5
La Vallee	54.3	15.2	2.15	10.8	20	31	3
V. De Geneve	35.0	1.2	42.34	18.0	53	37	4