

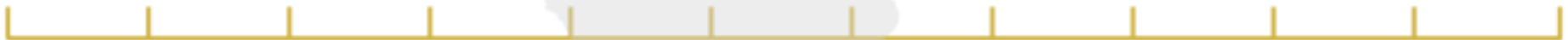
PROMiDAT
IBEROAMERICANO

Programa Iberoamericano de
Formación en Minería de Datos

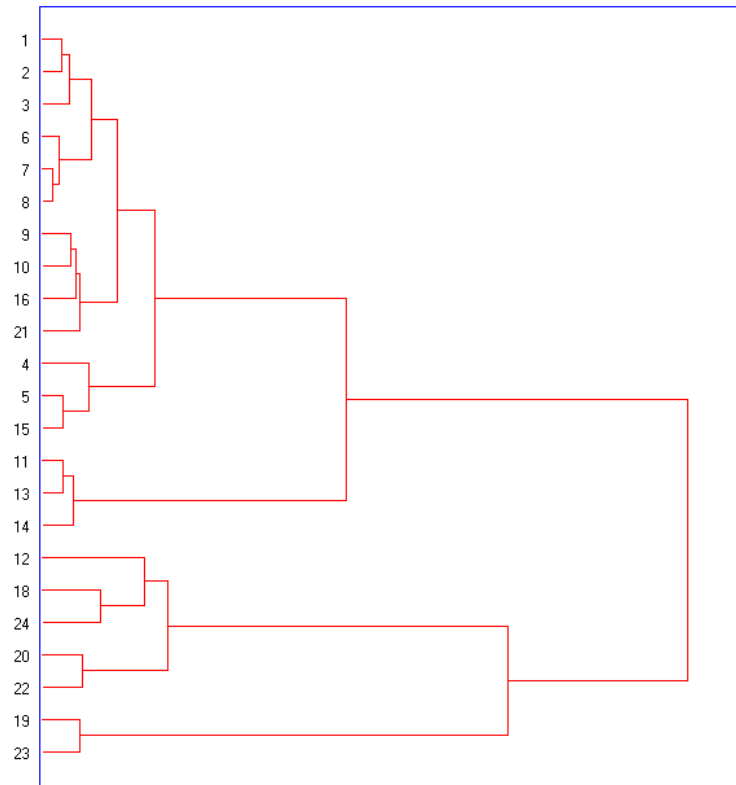


Clasificación Jerárquica Ascendente

[Una introducción]



Clasificación Jerárquica



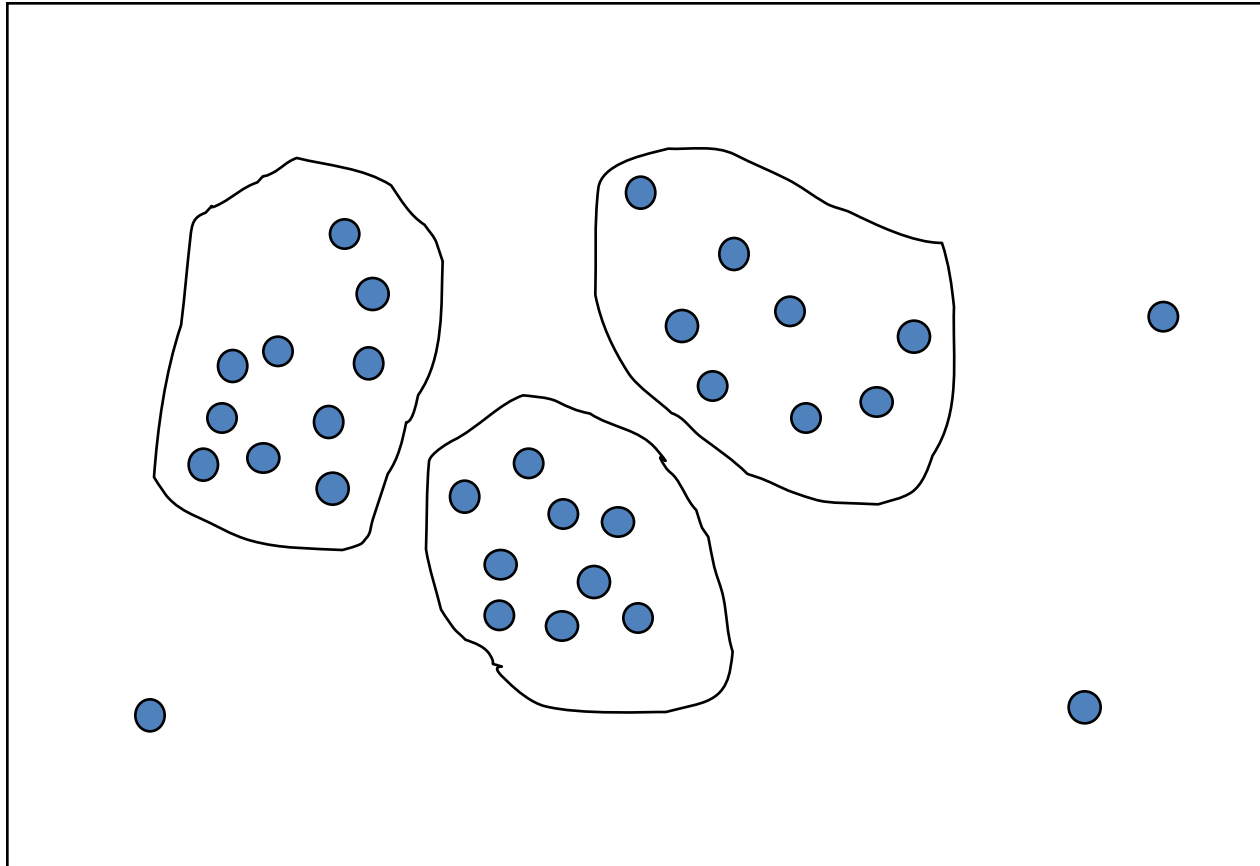
Clasificación Automática

- “La clasificación automática tiene por objetivo reconocer grupos de individuos homogéneos, de tal forma que los grupos queden bien separados y bien diferenciados.”
- “Estos individuos pueden estar descritos por una tabla de datos de individuos por variables, con variables cuantitativas o cualitativas, o por una tabla de proximidades.”

Tareas de la Minería de Datos

- **“Clustering”**: (clasificación no supervisada, aprendizaje no supervizado): Es similar a la clasificación (discriminación), excepto que los grupos no son predefinidos. El objetivo es particionar o segmentar un conjunto de datos o individuos en grupos que pueden ser disjuntos o no. Los grupos se forman basados en la similaridad de los datos o individuos en ciertas variables. Como los grupos no son dados a priori el experto debe dar una interpretación de los grupos que se forman.
- **Métodos**:
 - Clasificación Jerárquica (grupos disjuntos).
 - Nubes Dinámicas o k-means (grupos disjuntos).
 - Clasificación Piramidal (grupos NO disjuntos).

Cluster Analysis



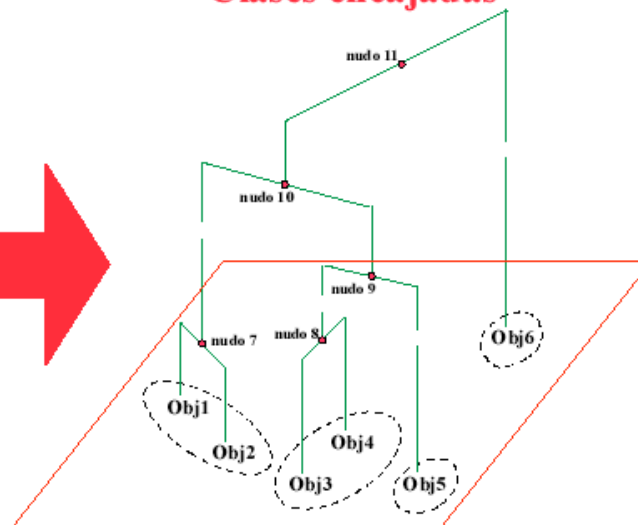
Clasificación Jerárquica

Tabla T(n,p)

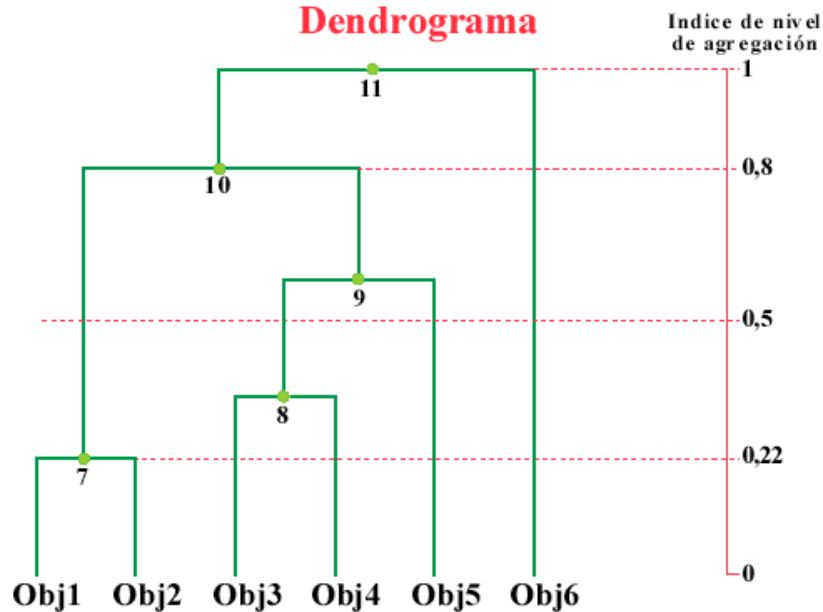
	Var.1	...	Var. j	...	Var. p
Obj.1					
Obj.2					
Obj.3			x_{ij}		
Obj.4					
Obj.5					
Obj.6					

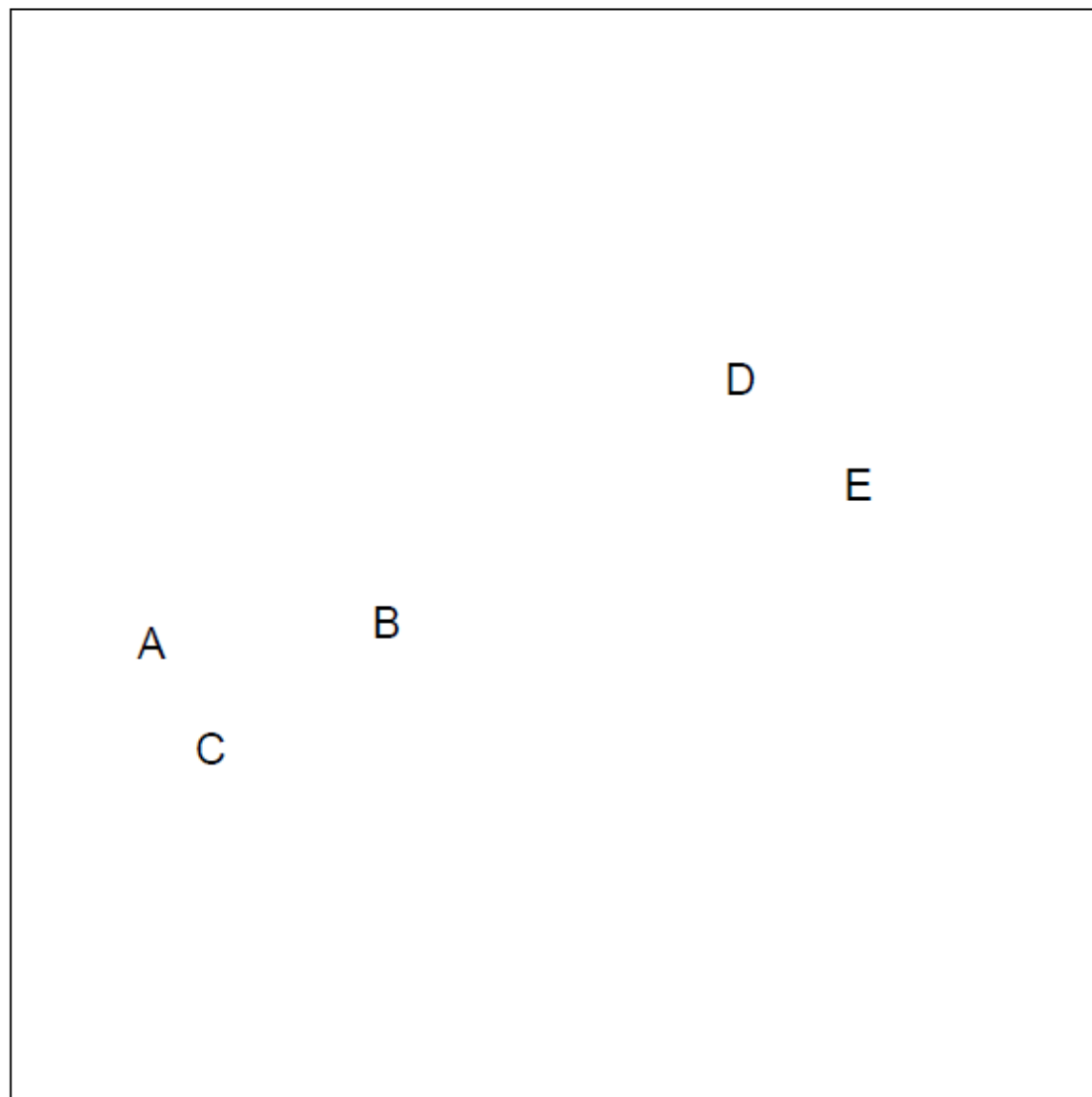


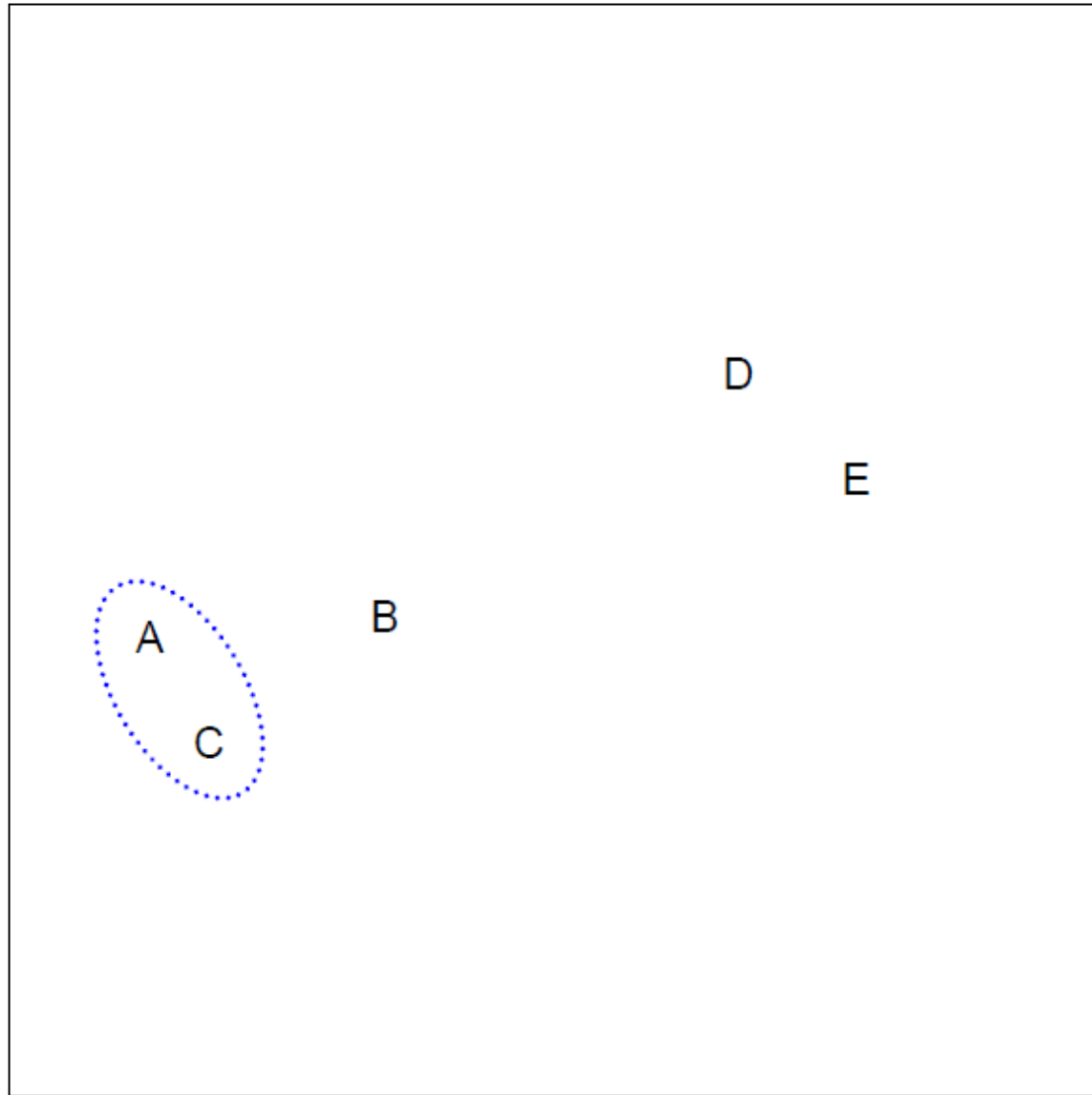
Clases encajadas

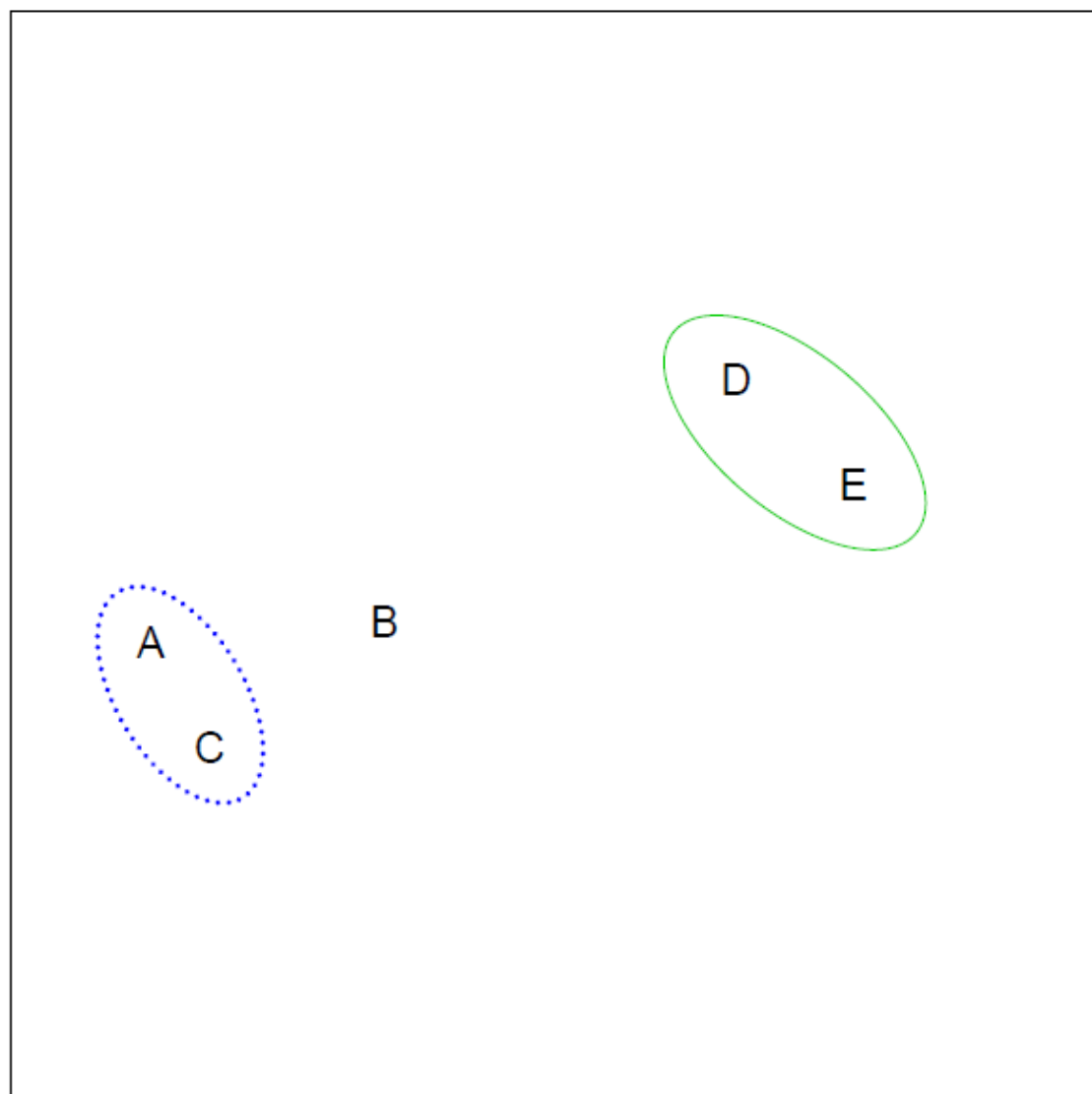


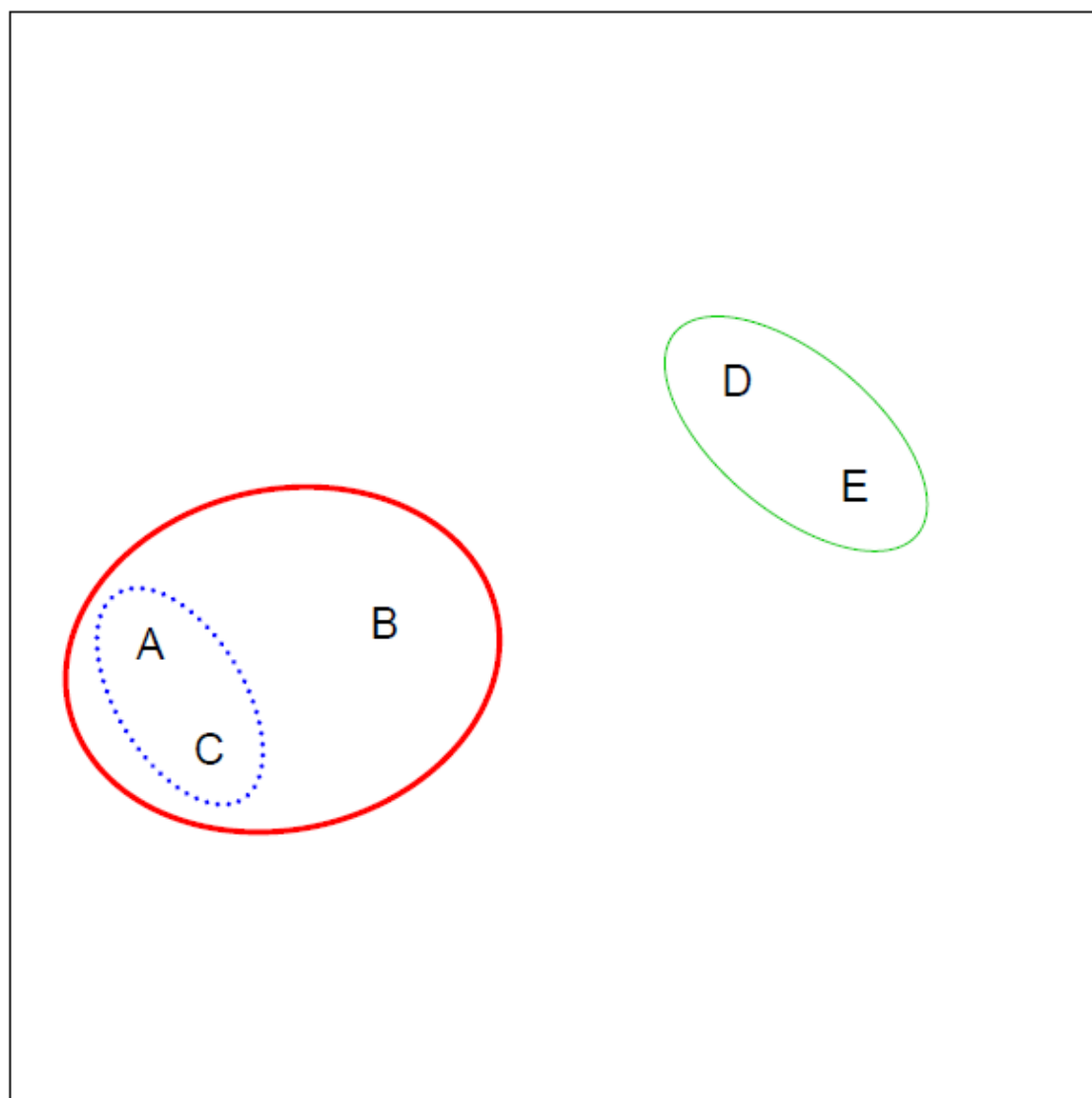
Dendrograma

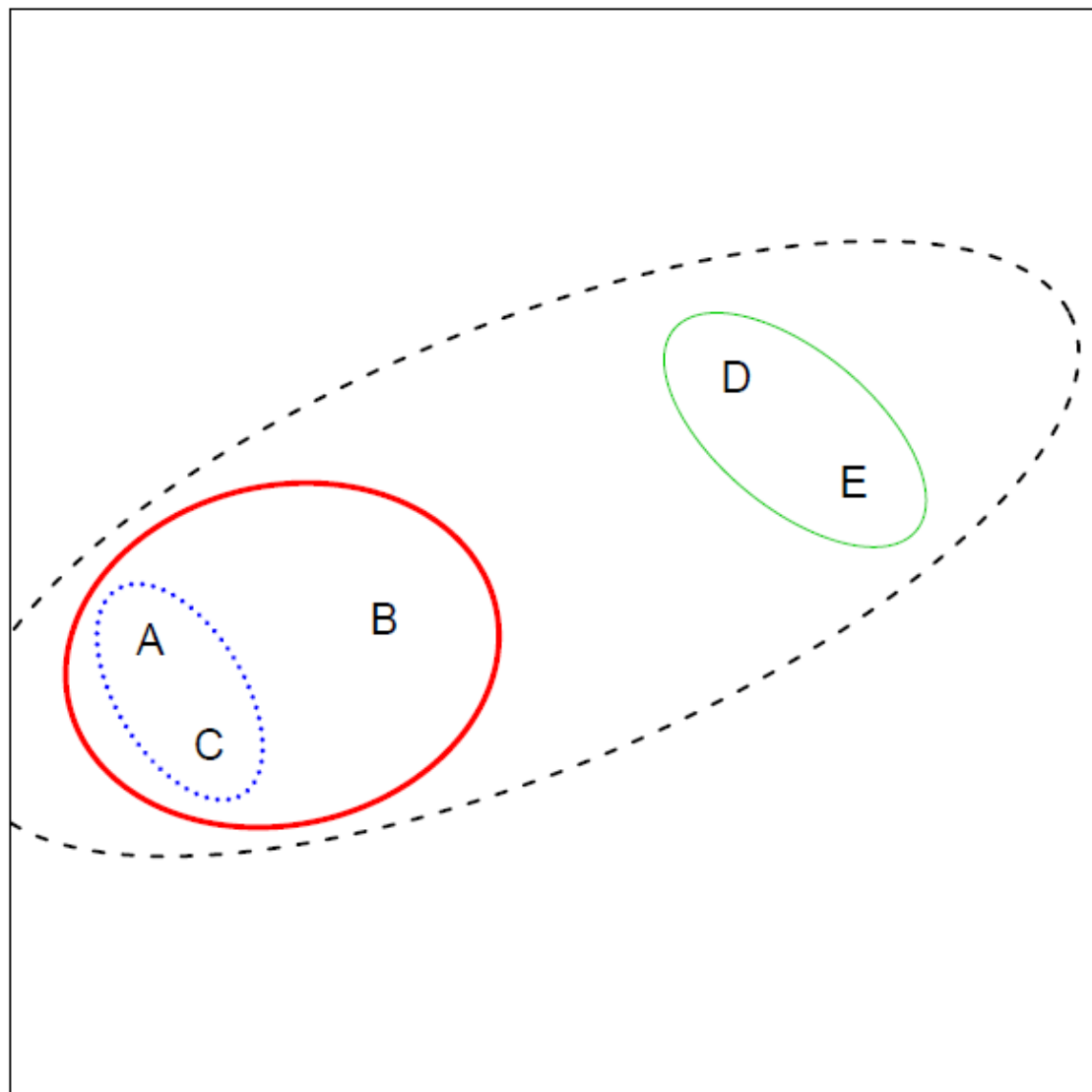


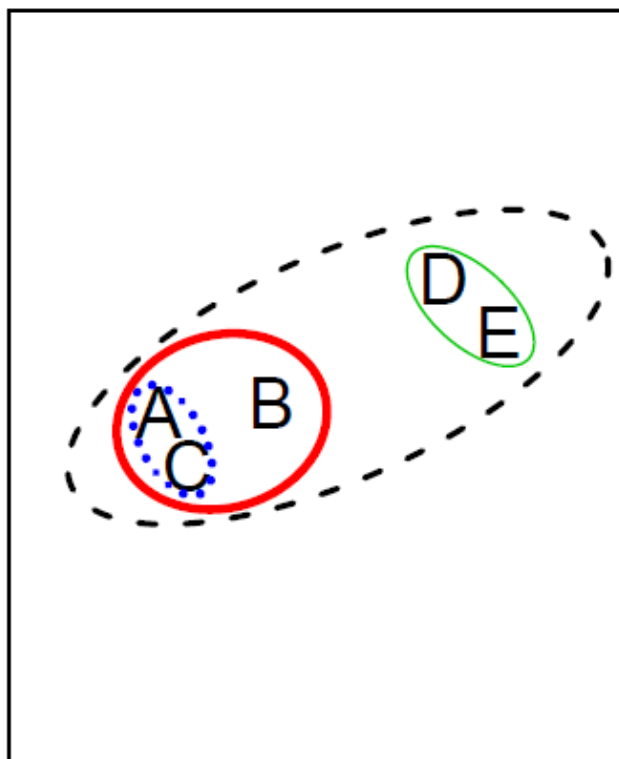




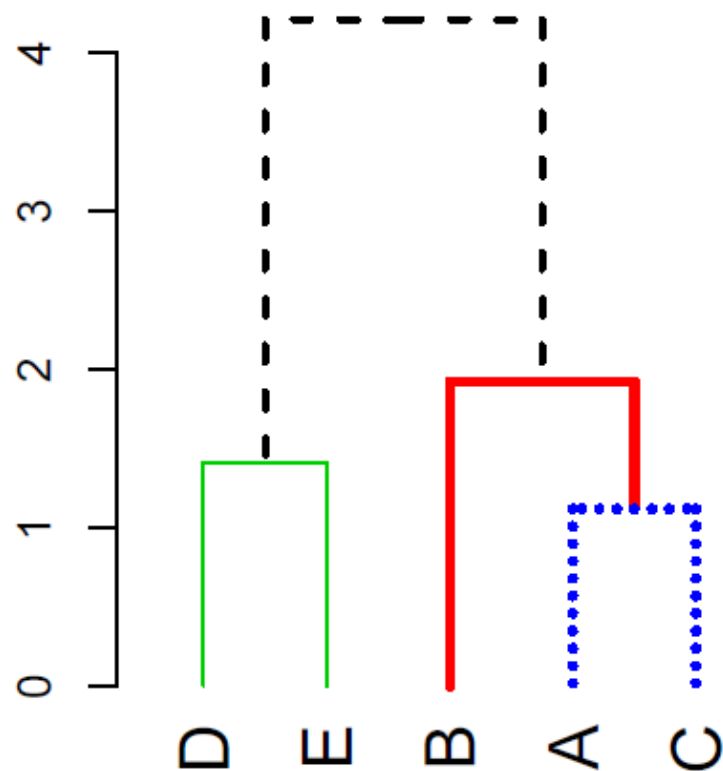








Dendrogram



2. Definiciones básicas

Sea X la matriz de datos cuyas n filas o p columnas, forman el conjunto del cual se busca una buena partición. Supondremos que X es una matriz de n individuos por p variables continuas, una tabla de contingencia, o alguna otra forma de datos asimilables a los anteriores.

Tabla de Datos

$$\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}$$

Ejemplo: Tabla Notas Escolares

	Matemáticas	Ciencias	Español	Historia	EdFísica
Lucía	7.0	6.5	9.2	8.6	8.0
Pedro	7.5	9.4	7.3	7.0	7.0
Inés	7.6	9.2	8.0	8.0	7.5
Luis	5.0	6.5	6.5	7.0	9.0
Andrés	6.0	6.0	7.8	8.9	7.3
Ana	7.8	9.6	7.7	8.0	6.5
Carlos	6.3	6.4	8.2	9.0	7.2
José	7.9	9.7	7.5	8.0	6.0
Sonia	6.0	6.0	6.5	5.5	8.7
María	6.8	7.2	8.7	9.0	7.0

2.1. *Disimilitudes y agregaciones*

Con el propósito de encontrar una clasificación de las filas o de las columnas de X , el primer problema a resolver es cómo cuantificar la similitud entre esos objetos o entre grupos de objetos.

2.2. *Indices de disimilitud*

Un índice de disimilitud entre un conjunto de objetos I (filas de la tabla de datos) es una función d tal que

$$d : I \times I \longrightarrow [0, +\infty[$$

y

$$d(x, y) = d(y, x) \text{ para todo } x, y \in I.$$

Ejemplo: Distancias Notas Escolares

Matriz de Distancias										
	Lucía	Pedro	Inés	Luis	Andrés	Ana	Carlos	José	Sonía	María
Lucía	0	3,98	3,11	3,85	1,947	3,89	1,517	4,28	4,32	1,39
Pedro		0	1,34	4,39	4,214	1,24	3,91	1,51	4,43	3,36
Inés			0	4,42	3,7	1,135782	3,265	1,69	4,77	2,53
Luis				0	3,072	1,89	3,439	5,45	1,89	4,07
Andrés					0	4,2	0,656	4,46	3,9	1,73
Ana						0	3,772	0,56	5,36	3
Carlos							0	4,05	4,2	1,09
José								0	5,64	3,3
Sonía									0	4,7
María										0

También llamamos a esta función, por abuso de lenguaje, una distancia. La distancia seleccionada depende, en general, de la naturaleza de los datos. Así, en nuestro caso tendremos tres clases de distancias:

- *Distancia euclídea clásica*: supongamos que $x_i = (x_{i1}, \dots, x_{ip})$ y $x_s = (x_{s1}, \dots, x_{sp})$ son dos filas cualesquiera de la matriz X , entonces la distancia euclídea diagonal es

$$d(x_i, x_s) = \sqrt{\sum_{j=1}^p (x_{ij} - x_{sj})^2}$$

Este tipo de disimilitud se usa comunmente cuando las variables observadas son continuas.

❑ **Ejemplo en Excel – Notas Escolares**

ver NotasEscolaresExcelCJ.xlsx

Tabla de Datos

	Matemáticas	Ciencias	Español	Historia	EdFísica
Lucía	7	6.5	9.2	8.6	8
Pedro	7.5	9.4	7.3	7	7
Inés	7.6	9.2	8	8	7.5
Luis	5	6.5	6.5	7	9
Andrés	6	6	7.8	8.9	7.3
Ana	7.8	9.6	7.7	8	6.5
Carlos	6.3	6.4	8.2	9	7.2
José	7.9	9.7	7.5	8	6
Sonía	6	6	6.5	5.5	8.7
María	6.8	7.2	8.7	9	7

Distancia Lucía-Pedro

0.25	8.41	3.61	2.56	1	3.9787
------	------	------	------	---	--------

Matriz de Distancias

Una tabla o matriz de distancias: es aquella que se calcula a partir de una tabla de datos *individuos-variables* y que en la entrada (i,j) tiene la distancia calculada entre el individuo i -ésimo (fila i) y el individuo j -ésimo (fila j), denotada $d(x_i, x_j)$.

❑ Ejemplo en Excel – Notas Escolares

ver EjemploEstudiantesCJ.xlsx

Matriz de Distancias

	Lucía	Pedro	Inés	Luis	Andrés	Ana	Carlos	José	Sonía	María
Lucía	0	3.98	3.11	3.85	1.947	3.89	1.517	4.28	4.32	1.39
Pedro		0	4.39	4.39	4.214	1.24	3.91	1.51	4.43	3.36
Inés			0	4.42	3.7	1.14	3.265	1.69	4.77	2.53
Luis				0	3.072	1.89	3.439	5.45	1.89	4.07
Andrés					0	4.2	0.656	4.46	3.9	1.73
Ana						0	3.772	0.56	5.36	3
Carlos							0	4.05	4.2	1.09
José								0	5.64	3.3
Sonía									0	4.7
María										0

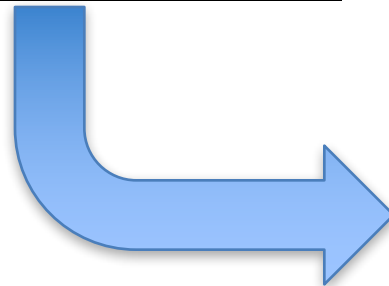
- *Distancia euclídea de las varianzas*: cuando las variables tienen varianzas muy desiguales, la magnitud del término $(x_{ij} - x_{sj})^2$ puede depender de la varianza σ_j^2 de la variable x^j , haciendo depender la distancia entre filas, de la estructura de varianzas más que de la estructura de correlaciones. Para corregir este efecto se usa la fórmula

$$d(x_i, x_s) = \sqrt{\sum_{j=1}^p \frac{1}{\sigma_j^2} (x_{ij} - x_{sj})^2}$$

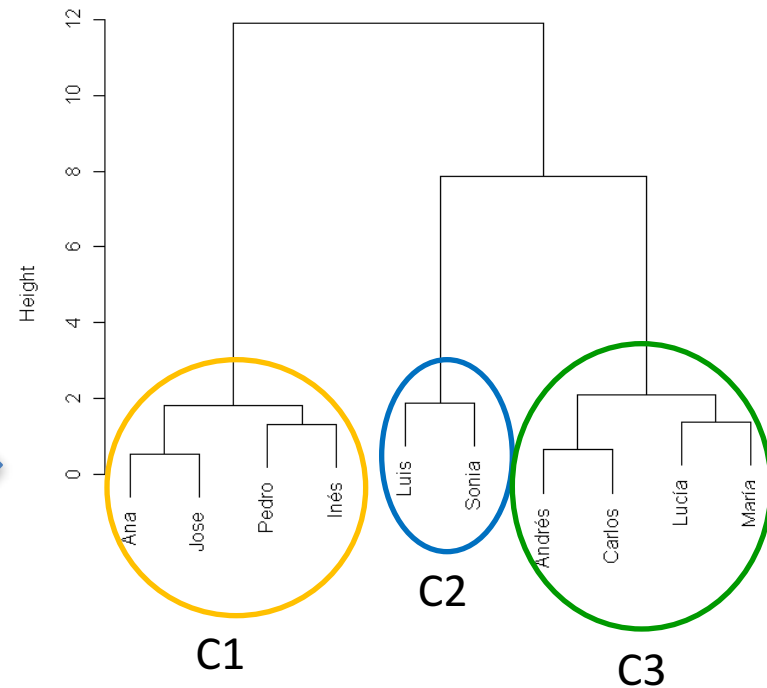
Obsérvese que lo anterior equivale a dividir cada columna x^j por su desviación estándar σ_j y usar la distancia euclídea clásica sobre los datos así transformados.

¿Cómo se construye el árbol?

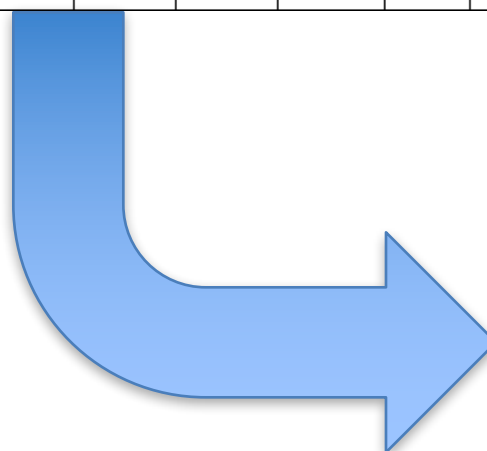
Análisis de los Clústeres					
	Matemáticas	Ciencias	Español	Historia	EdFísica
Lucía	7	6,5	9,2	8,6	8
Pedro	7,5	9,4	7,3	7	7
Inés	7,6	9,2	8	8	7,5
Luis	5	6,5	6,5	7	9
Andrés	6	6	7,8	8,9	7,3
Ana	7,8	9,6	7,7	8	6,5
Carlos	6,3	6,4	8,2	9	7,2
José	7,9	9,7	7,5	8	6
Sonía	6	6	6,5	5,5	8,7
María	6,8	7,2	8,7	9	7



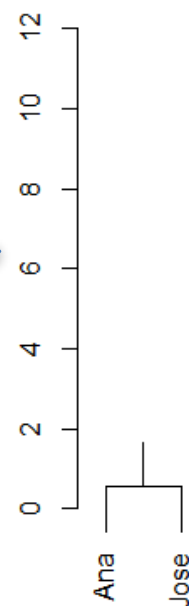
Cluster Dendrogram for Solution HClust.3



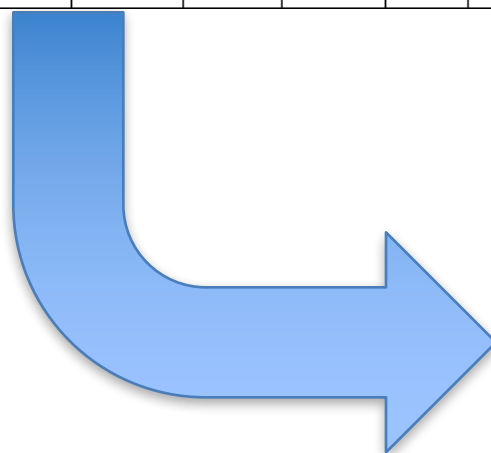
Matriz de Distancias										
	Lucía	Pedro	Inés	Luis	Andrés	Ana	Carlos	José	Sonía	María
Lucía	0	3,98	3,11	3,85	1,95	3,89	1,517	4,28	4,32	1,39
Pedro		0	1,34	4,39	4,21	1,24	3,91	1,51	4,43	3,36
Inés			0	4,42	3,7	1,14	3,265	1,69	4,77	2,53
Luis				0	3,07	1,89	3,439	5,45	1,89	4,07
Andrés					0	4,2	0,656	4,46	3,9	1,73
Ana						0	3,772	0,56	5,36	3
Carlos							0	4,05	4,2	1,09
José								0	5,64	3,3
Sonía									0	4,7
María										0



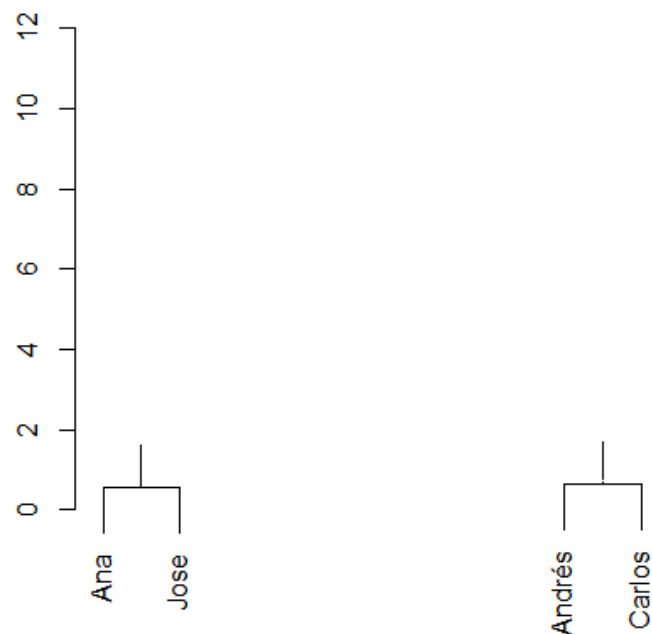
Cluster Dendrogram for Solution HClust.1



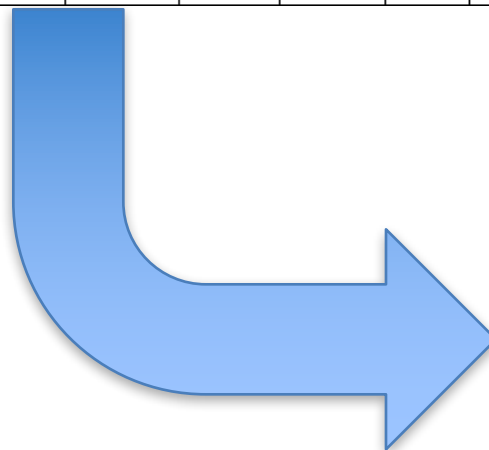
Matriz de Distancias										
	Lucía	Pedro	Inés	Luis	Andrés	Ana	Carlos	José	Sonía	María
Lucía	0	3,98	3,11	3,85	1,947	3,89	1,517	4,28	4,32	1,39
Pedro		0	1,34	4,39	4,214	1,24	3,91	1,51	4,43	3,36
Inés			0	4,42	3,7	1,14	3,265	1,69	4,77	2,53
Luis				0	3,072	1,89	3,439	5,45	1,89	4,07
Andrés					0	4,2	0,656	4,46	3,9	1,73
Ana						0	3,772	0,56	5,36	3
Carlos							0	4,05	4,2	1,09
José								0	5,64	3,3
Sonía									0	4,7
María										0



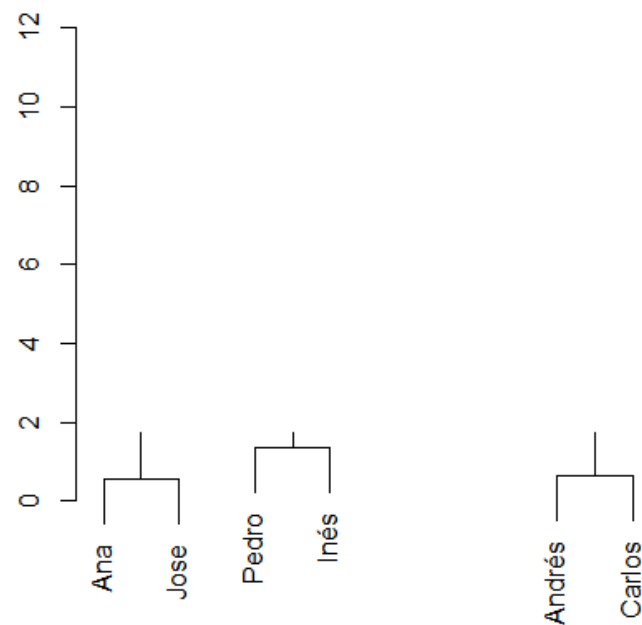
Cluster Dendrogram for Solution HClust.1



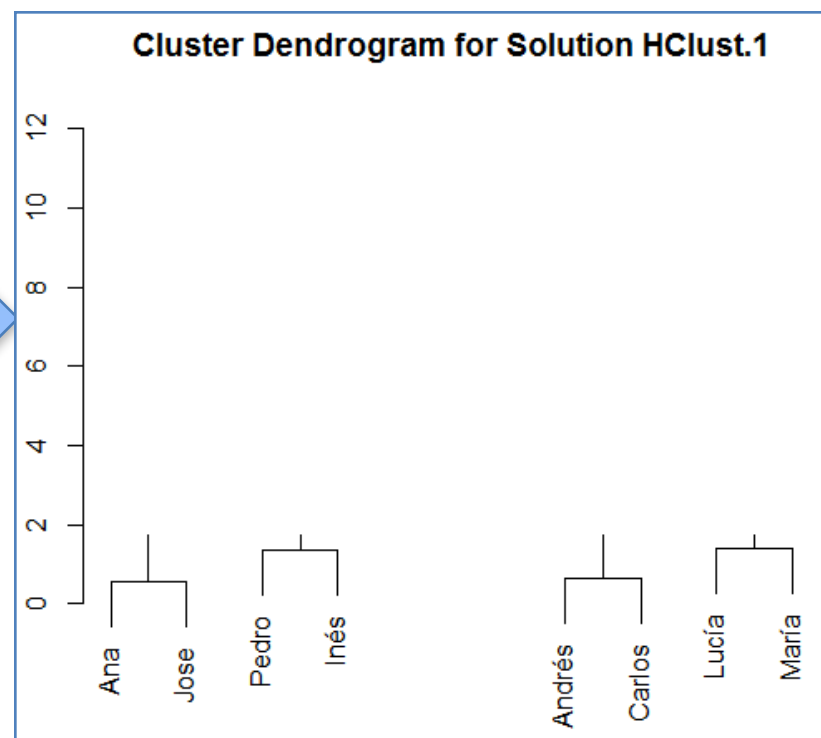
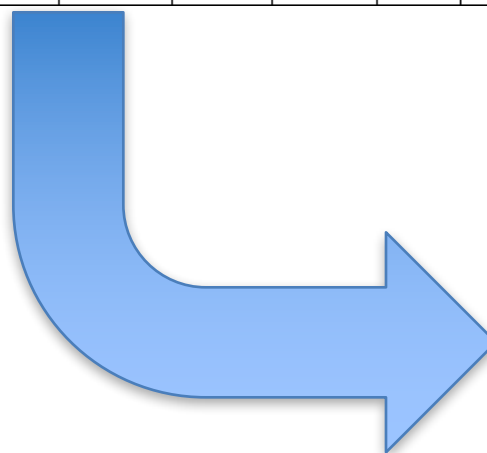
Matriz de Distancias										
	Lucía	Pedro	Inés	Luis	Andrés	Ana	Carlos	José	Sonía	María
Lucía	0	3,98	3,11	3,85	1,947	3,89	1,517	4,28	4,32	1,39
Pedro		0	1,34	4,39	4,214	1,24	3,91	1,51	4,43	3,36
Inés			0	4,42	3,7	1,14	3,265	1,69	4,77	2,53
Luis				0	3,072	1,89	3,439	5,45	1,89	4,07
Andrés					0	4,2	0,656	4,46	3,9	1,73
Ana						0	3,772	0,56	5,36	3
Carlos							0	4,05	4,2	1,09
José								0	5,64	3,3
Sonía									0	4,7
María										0



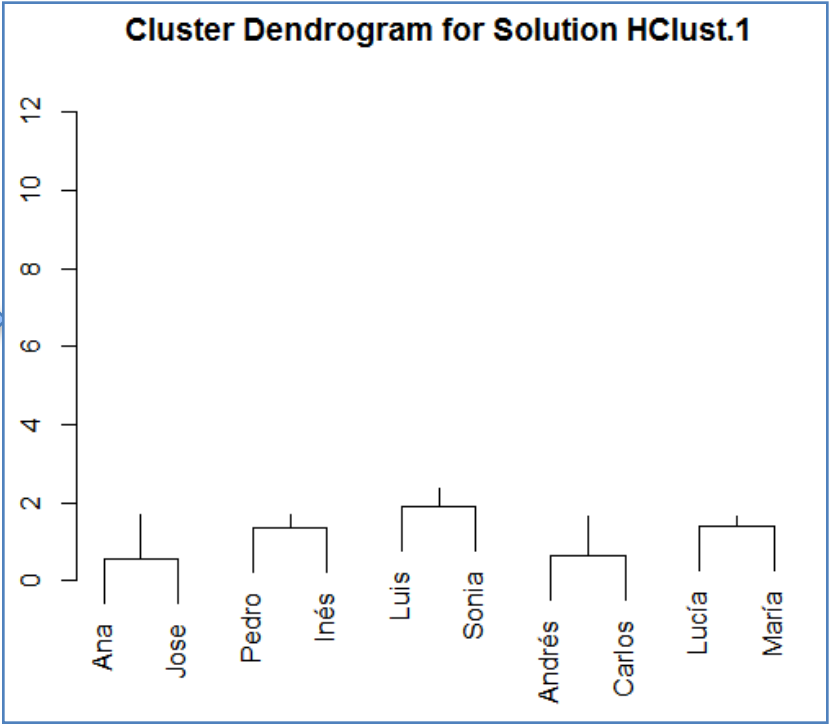
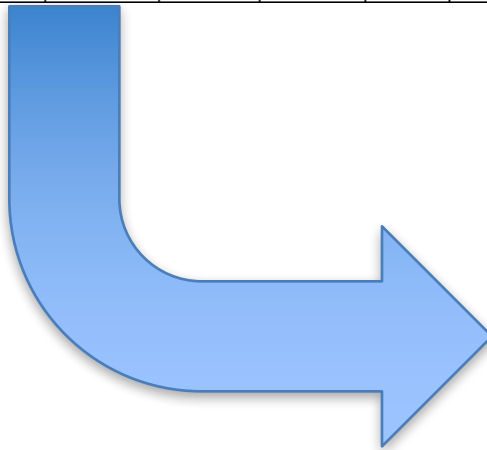
Cluster Dendrogram for Solution HClust.1



Matriz de Distancias										
	Lucía	Pedro	Inés	Luis	Andrés	Ana	Carlos	José	Sonía	María
Lucía	0	3,98	3,11	3,85	1,947	3,89	1,517	4,28	4,32	1,39
Pedro		0	1,34	4,39	4,214	1,24	3,91	1,51	4,43	3,36
Inés			0	4,42	3,7	1,14	3,265	1,69	4,77	2,53
Luis				0	3,072	1,89	3,439	5,45	1,89	4,07
Andrés					0	4,2	0,656	4,46	3,9	1,73
Ana						0	3,772	0,56	5,36	3
Carlos							0	4,05	4,2	1,09
José								0	5,64	3,3
Sonía									0	4,7
María										0



Matriz de Distancias										
	Lucía	Pedro	Inés	Luis	Andrés	Ana	Carlos	José	Sonía	María
Lucía	0	3,98	3,11	3,85	1,947	3,89	1,517	4,28	4,32	1,39
Pedro		0	1,34	4,39	4,214	1,24	3,91	1,51	4,43	3,36
Inés			0	4,42	3,7	1,14	3,265	1,69	4,77	2,53
Luis				0	3,072	1,89	3,439	5,45	1,89	4,07
Andrés					0	4,2	0,656	4,46	3,9	1,73
Ana						0	3,772	0,56	5,36	3
Carlos							0	4,05	4,2	1,09
José								0	5,64	3,3
Sonía									0	4,7
María										0



Agregaciones

Los métodos de clasificación automática usan generalmente una noción de proximidad entre *grupos* de elementos, para medir la separación entre las clases que se buscan. Para ello, se introduce el concepto de *agregación*, que no es más que una disimilitud entre grupos de individuos: sean $A, B \subset \Omega$, entonces la agregación entre A y B es:

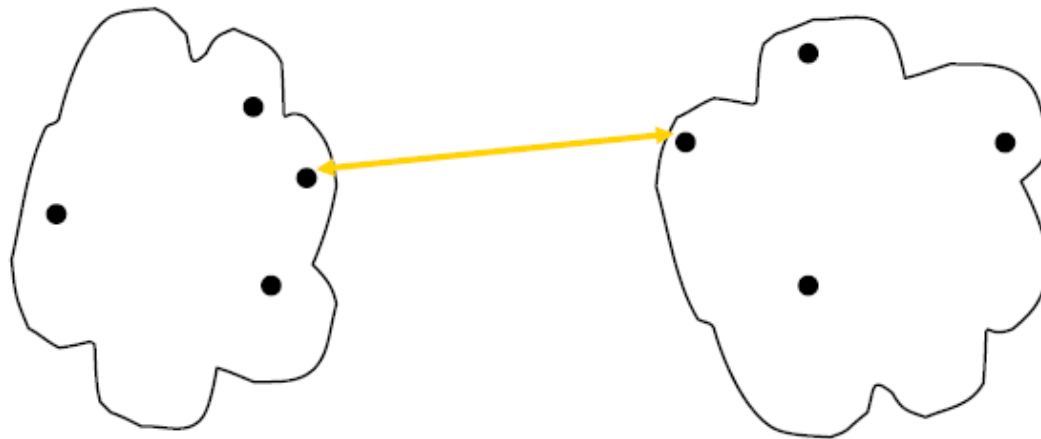
$$\delta(A, B)$$

tal que δ es una disimilitud en el conjunto de partes $\mathcal{P}(\Omega)$:

- i) $\delta(A, A) = 0$ para todo $A \in \mathcal{P}(\Omega)$
- ii) $\delta(A, B) = \delta(B, A)$ para todo $A, B \in \mathcal{P}(\Omega)$

Agregación del salto mínimo o del vecino más cercano:

$$\delta_{\min}(A, B) = \min\{d(\mathbf{x}_i, \mathbf{x}_j) | \mathbf{x}_i \in A, \mathbf{x}_j \in B\}$$

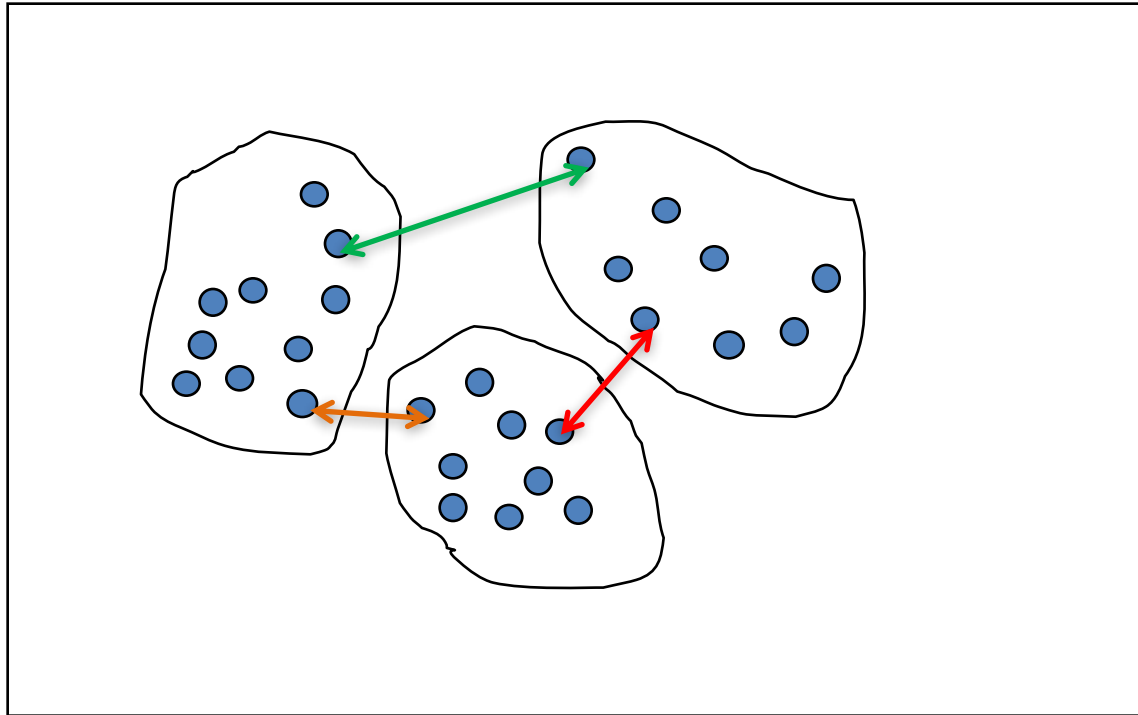


Single linkage uses the minimum distance.

Ejemplos

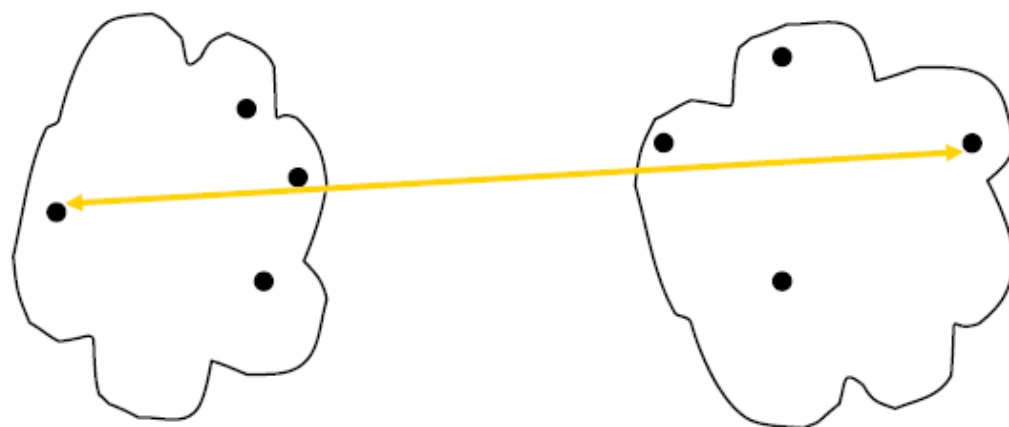
2. *Agregación del salto mínimo:*

$$\delta_{\min}(x, y) = \min \{d(h, k) \mid h \in x \text{ y } k \in y\}.$$



Agregación del salto máximo:

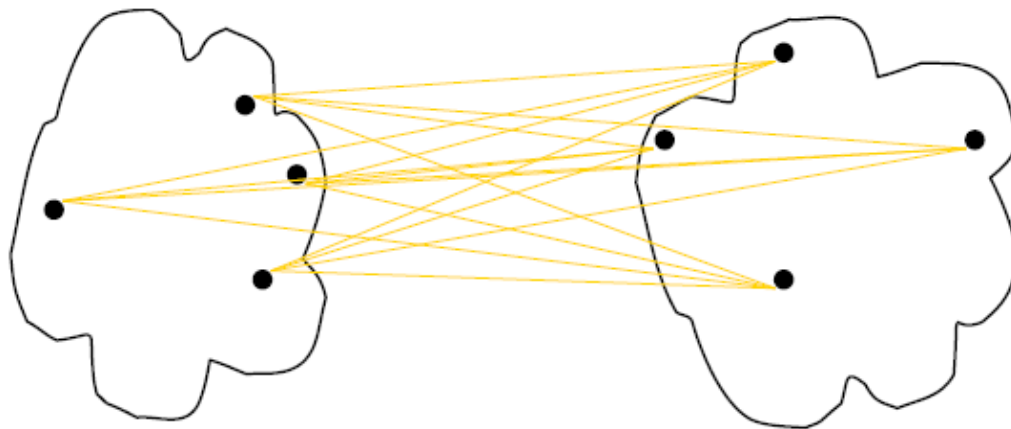
$$\delta_{\max}(A, B) = \max\{d(\mathbf{x}_i, \mathbf{x}_j) | \mathbf{x}_i \in A, \mathbf{x}_j \in B\}$$



Complete linkage uses the maximum distance.

Agregación del salto promedio:

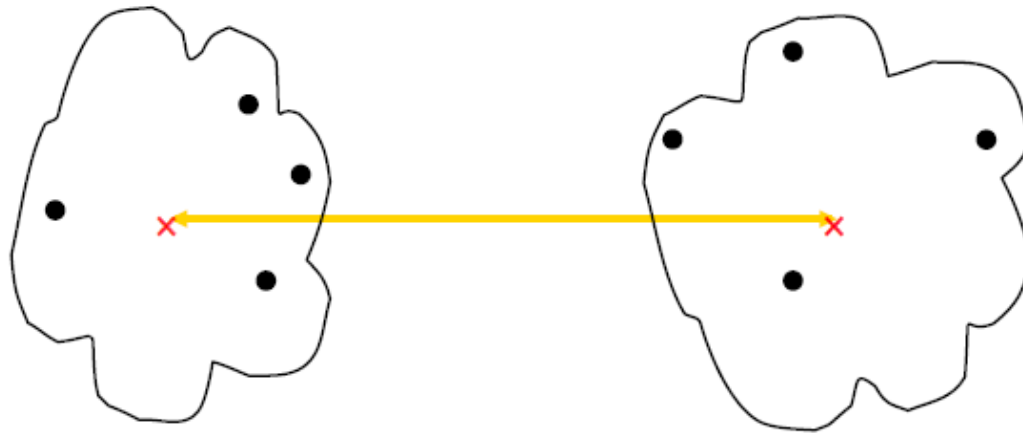
$$\delta_{\text{prom}}(A, B) = \frac{1}{|A| \times |B|} \sum_{\substack{\mathbf{x}_i \in A \\ \mathbf{x}_j \in B}} d(\mathbf{x}_i, \mathbf{x}_j)$$



Group average linkage uses the average distance between groups.

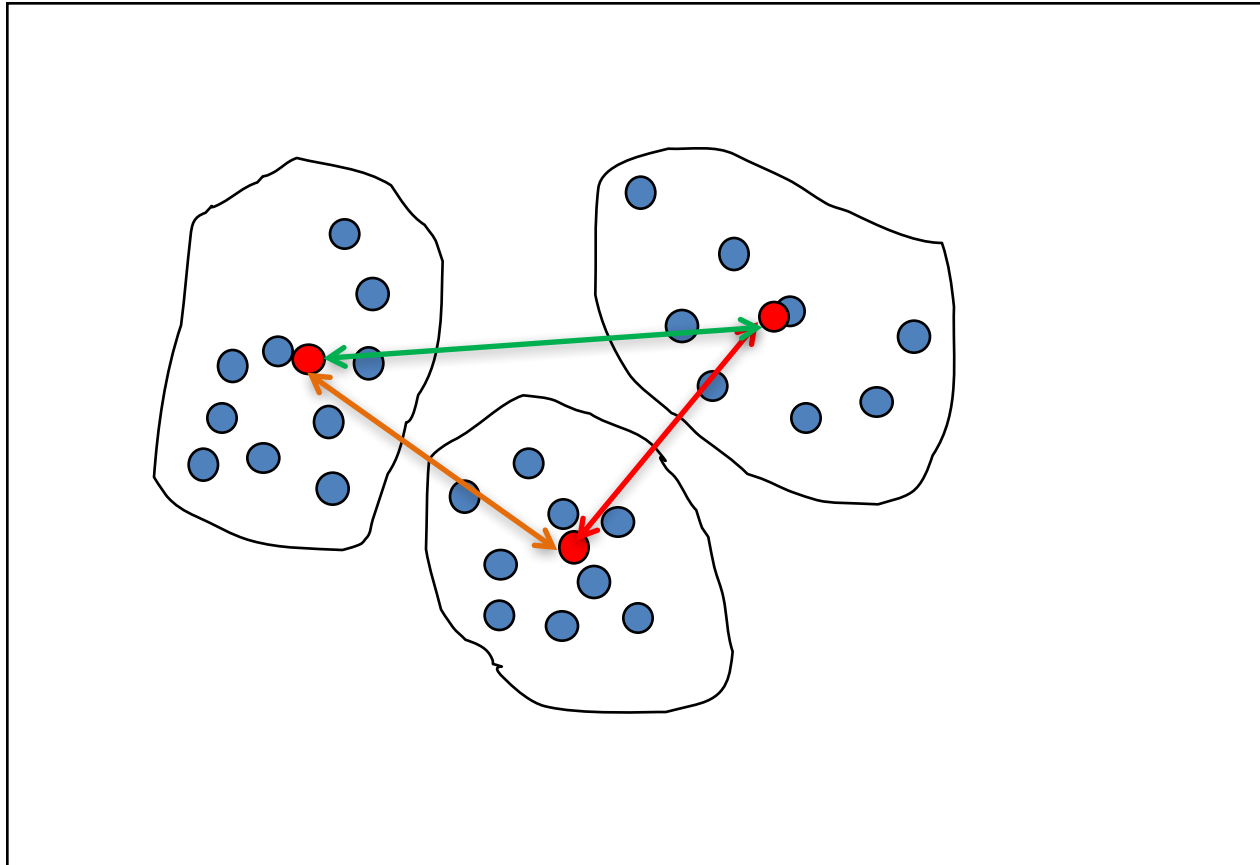
Agregación de Ward:

$$\delta_{\text{ward}}(A, B) = \frac{|A||B|}{|A| + |B|} \|\mathbf{g}_A - \mathbf{g}_B\|^2$$

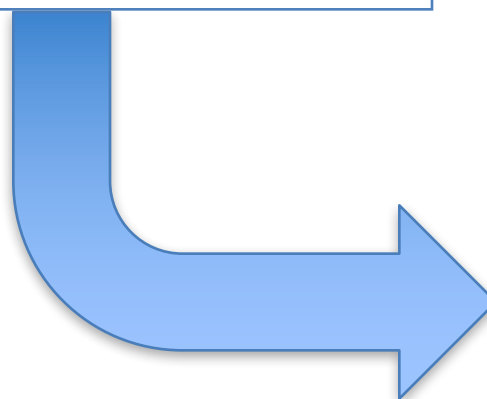
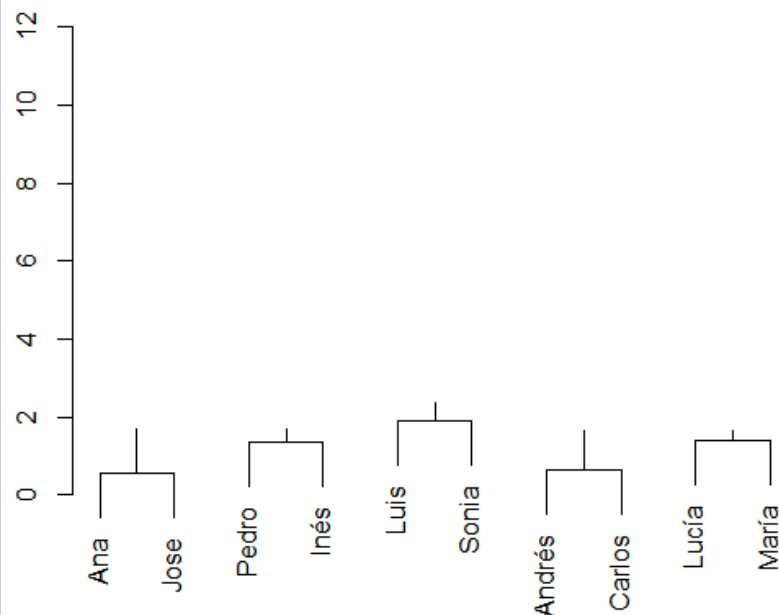


Centroid uses the distance between the centroids of the clusters
(presumes one can compute centroids...)

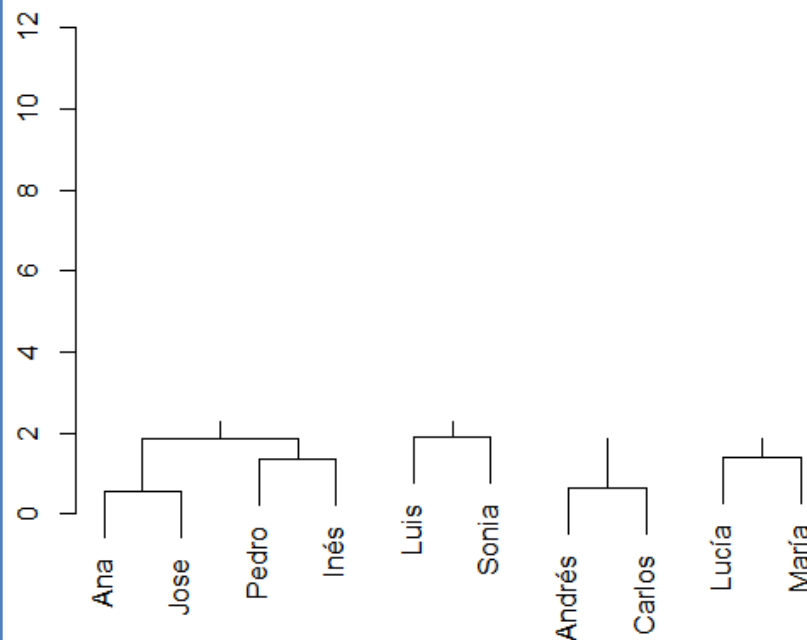
Agregación de Ward



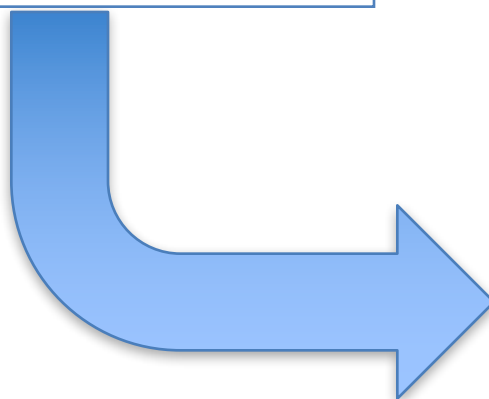
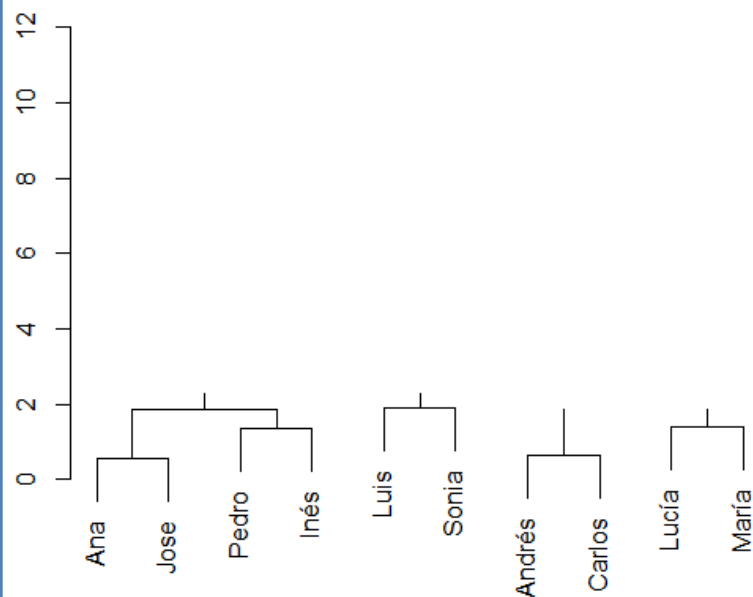
Cluster Dendrogram for Solution HClust.1



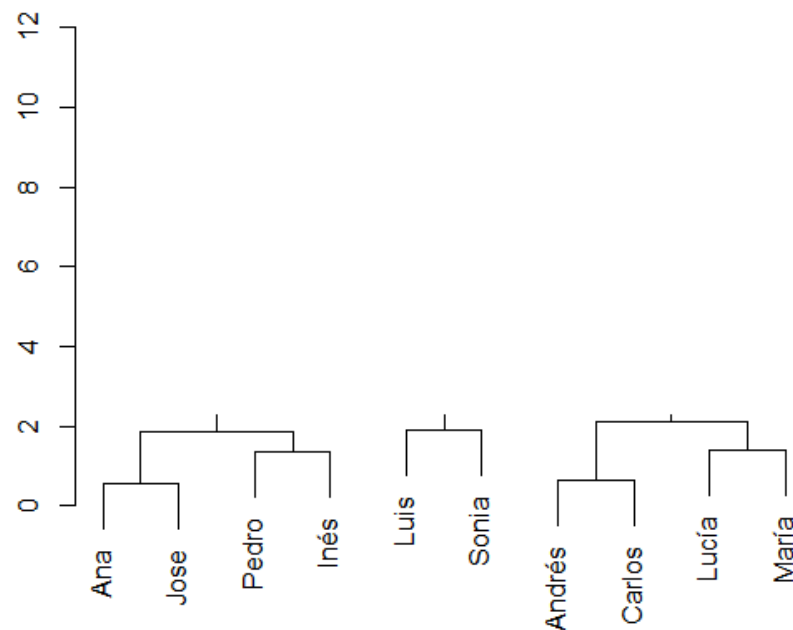
Cluster Dendrogram for Solution HClust.1



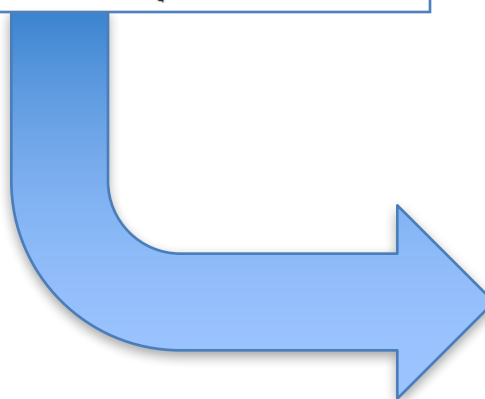
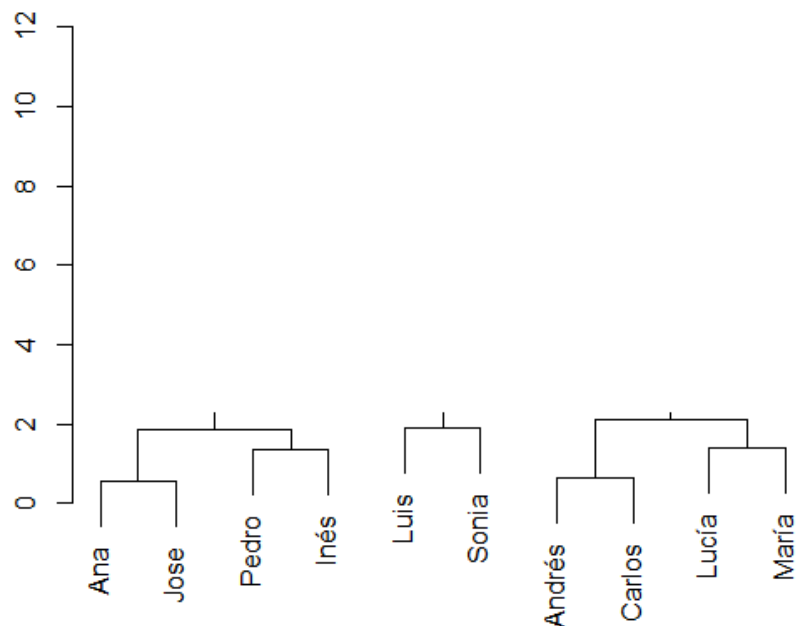
Cluster Dendrogram for Solution HClust.1



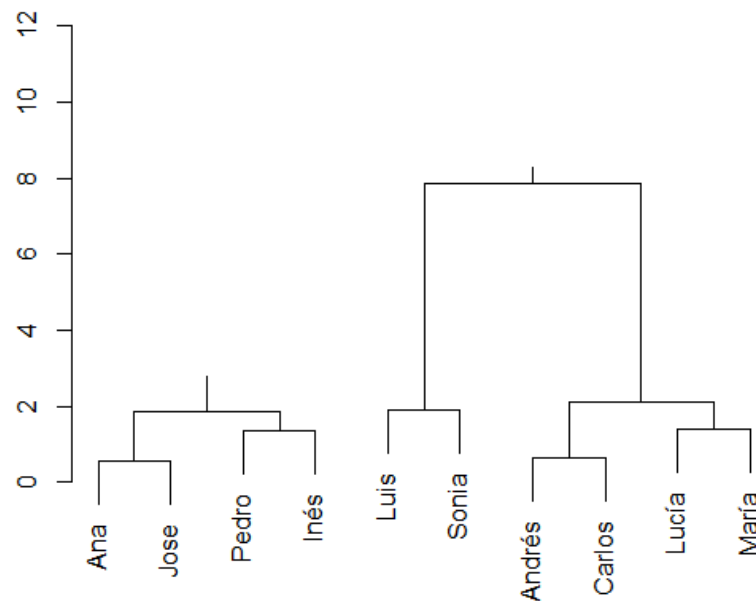
Cluster Dendrogram for Solution HClust.1



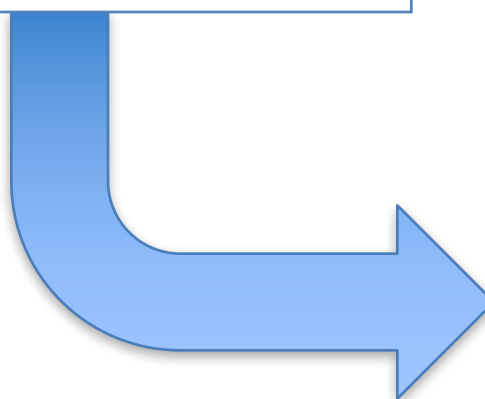
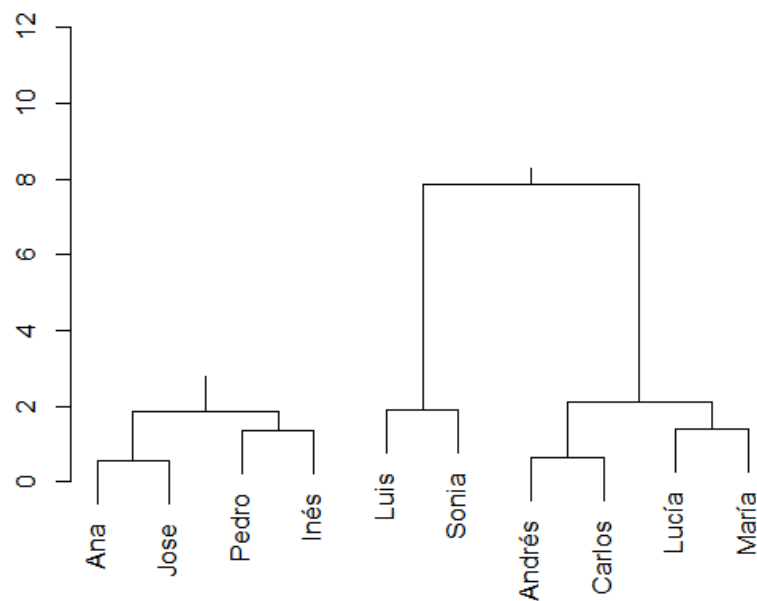
Cluster Dendrogram for Solution HClust.1



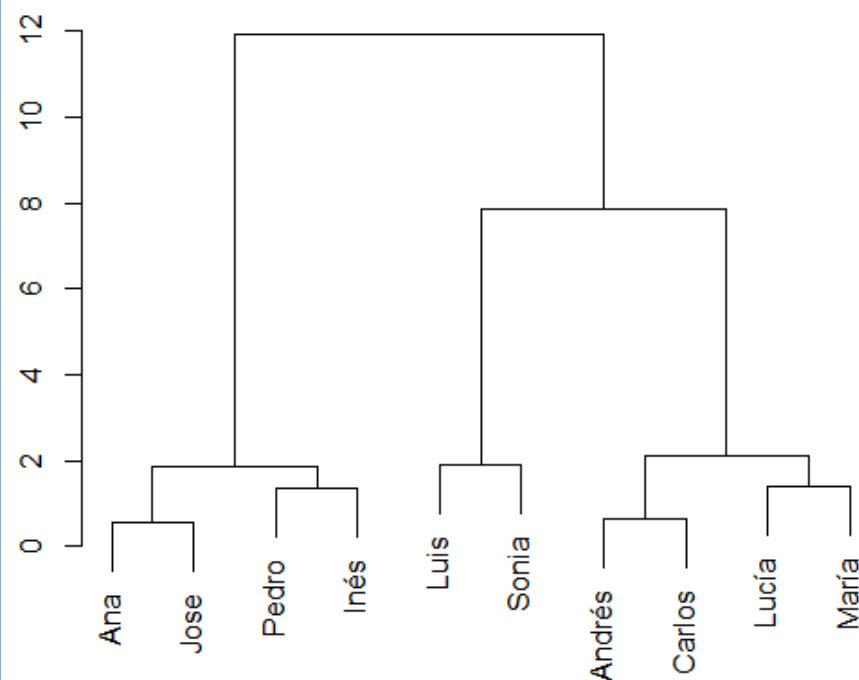
Cluster Dendrogram for Solution HClust.1



Cluster Dendrogram for Solution HClust.1



Cluster Dendrogram for Solution HClust.1



Ejemplo completo “a pie”

Ejercicio 1. Importaciones Centroamérica - México, 1979-1988

Considérese la tabla de datos siguiente, la cual contiene las importaciones hechas por los países centroamericanos, provenientes de México, entre 1979 y 1988.

Año	Costa Rica	El Salvador	Guatemala	Honduras	Nicaragua	Panamá
1979	44.4	27.2	45.6	20	6	14.1
1980	75.5	11.8	58.9	22.6	17.8	14.4
1981	110.7	50.6	128.3	17.2	119.4	118.5
1982	80.3	70.6	102.2	15.2	154.9	146.1
1983	81.6	82.3	89	35.1	169.4	127.1
1984	76.4	97.4	185	51	75.5	129
1985	32	89.5	195.3	31.1	33.4	110.2
1986	55.5	63.1	66.3	24.4	9.7	66.7
1987	74.3	72.6	76.3	28.1	11.2	110.7
1988	84.5	76.2	80.1	29.5	11.8	110.2

Ejemplo completo “a pie”

1.3. Construcción paulatina de la jerarquía binaria

A continuación se muestran las matrices de distancias sucesivas realizadas para el cálculo de la jerarquía binaria utilizando distancia euclídea y agregación de salto mínimo. En cada matriz se resalta en color salmón el menor valor de la distancia entre dos elementos y, en amarillo y verde, los valores de las distancias para los dos individuos (o agregación de individuos) cuya distancia es menor y que serán utilizadas para escoger, tal como se solicita en este caso, el menor de los valores.

	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988
1979	0	39,08	188,55	214,46	214,48	210,66	191,19	68,12	115,42	120,66
1980		0	169,56	203,84	204,34	201,24	189,81	76,75	115,54	118,06
1981			0	63,49	79,31	98,8	140,9	147,7	128,06	124,35
1982				0	35,78	124,37	166,32	171,56	150,93	150,03
1983					0	136,18	180,58	175,59	160,17	159
1984						0	68,28	156,91	132,02	128,03
1985							0	142,79	129,38	129,13
1986								0	49,96	55,91
1987									0	11,58
1988										0

Tabla 3: Importaciones CA - México, Tabla inicial de distancias

Ejemplo completo “a pie”

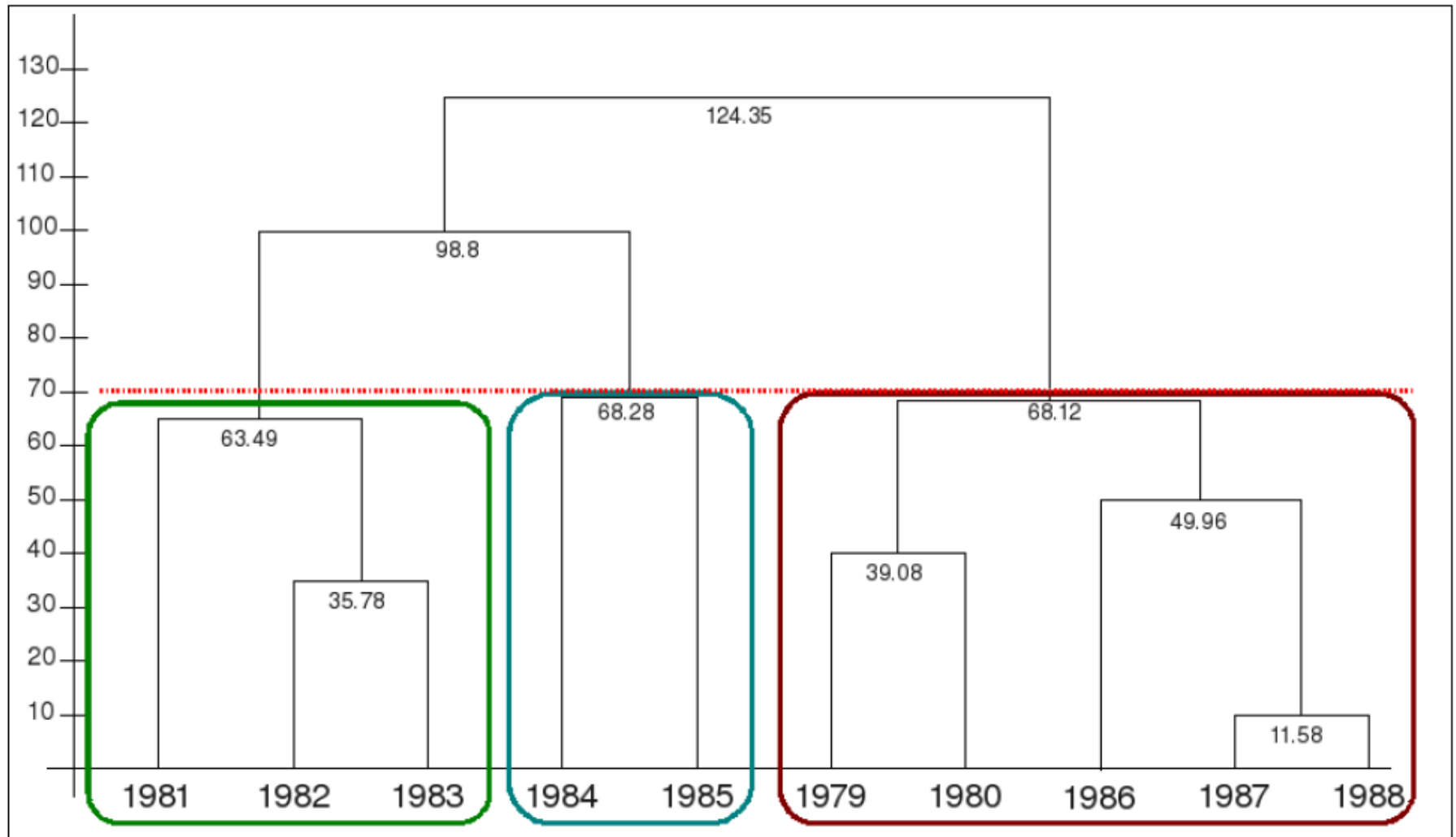


Ilustración 4: Importaciones CA - México, dendrograma realizado con agrupación de valor mínimo

Ejemplo completo “a pie”

	1979	1980	1981	1982	1983	1984	1985	1986	1987-1988
1979	0	39,08	188,55	214,46	214,48	210,66	191,19	68,12	115,42
1980		0	169,56	203,84	204,34	201,24	189,81	76,75	115,54
1981			0	63,49	79,31	98,8	140,9	147,7	124,35
1982				0	35,78	124,37	166,32	171,56	150,03
1983					0	136,18	180,58	175,59	159
1984						0	68,28	156,91	128,03
1985							0	142,79	129,13
1986								0	49,96
1987-1988									0

Tabla 4: Importaciones CA – México, cálculo de jerarquía binaria, paso 1

	1979	1980	1981	1984	1985	1986	1987-1988	1982-1983
1979	0	39,08	188,55	210,66	191,19	68,12	115,42	214,46
1980		0	169,56	201,24	189,81	76,75	115,54	203,84
1981			0	98,8	140,9	147,7	124,35	63,49
1984				0	68,28	156,91	128,03	124,37
1985					0	142,79	129,13	166,32
1986						0	49,96	171,56
1987-1988							0	150,03
1982-1983								0

Tabla 5: Importaciones CA – México, cálculo de jerarquía binaria, paso 2

Ejemplo completo “a pie”

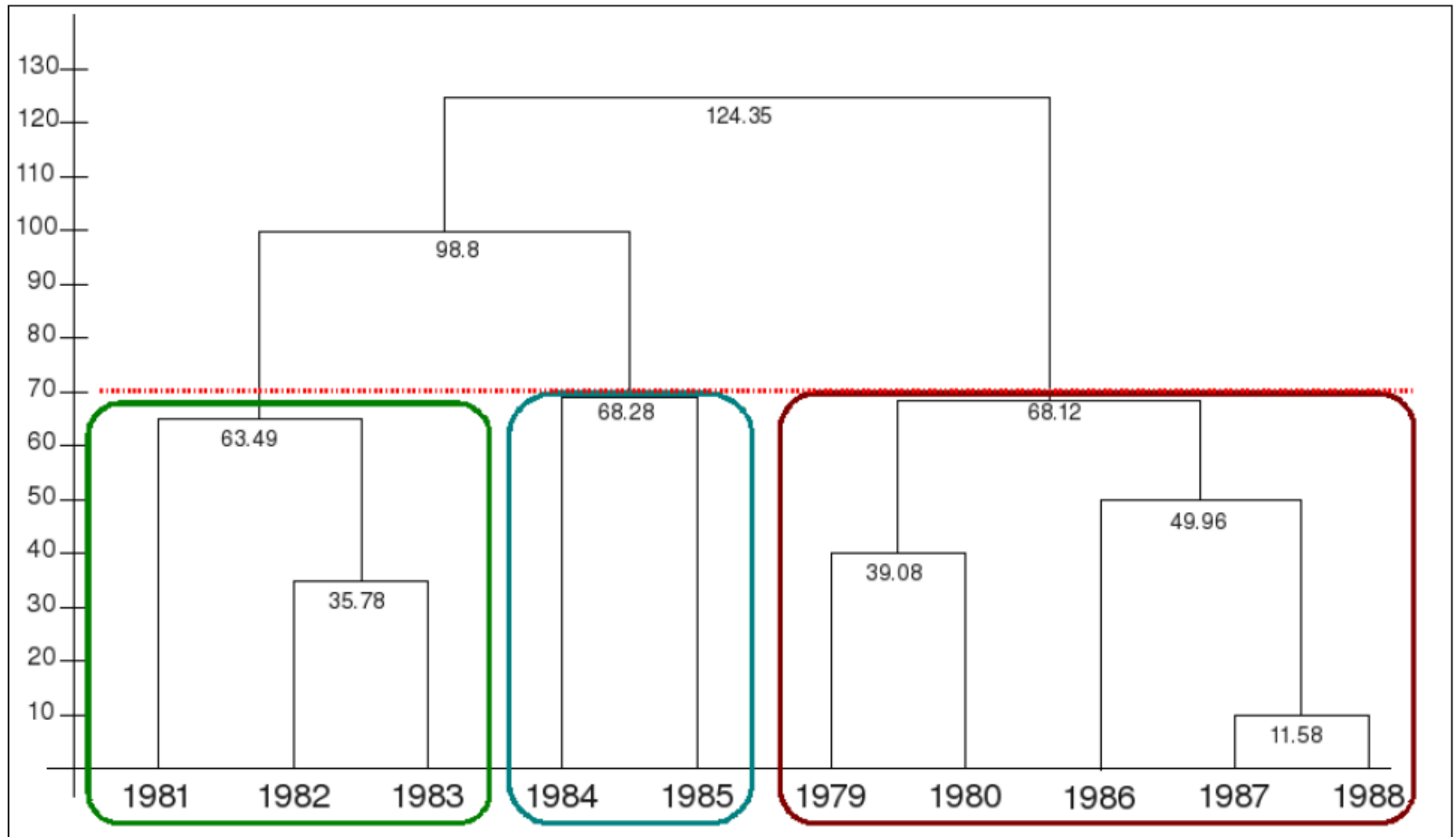


Ilustración 4: Importaciones CA - México, dendrograma realizado con agrupación de valor mínimo

Ejemplo completo “a pie”

	1981	1984	1985	1986	1987-1988	1982-1983	1979-1980
1981	0	98,8	140,9	147,7	124,35	63,49	169,56
1984		0	68,28	156,91	128,03	124,37	201,24
1985			0	142,79	129,13	166,32	189,81
1986				0	49,96	171,56	68,12
1987-1988					0	150,03	115,42
1982-1983						0	203,84
1979-1980							0

Tabla 6: Importaciones CA - México, cálculo de jerarquía binaria, paso 3

	1981	1984	1985	1982-1983	1979-1980	1986-(1987-1988)
1981	0	98,8	140,9	63,49	169,56	124,35
1984		0	68,28	124,37	201,24	128,03
1985			0	166,32	189,81	129,13
1982-1983				0	203,84	150,03
1979-1980					0	68,12
1986-(1987-1988)						0

Tabla 7: Importaciones CA - México, cálculo de jerarquía binaria, paso 4

Ejemplo completo “a pie”

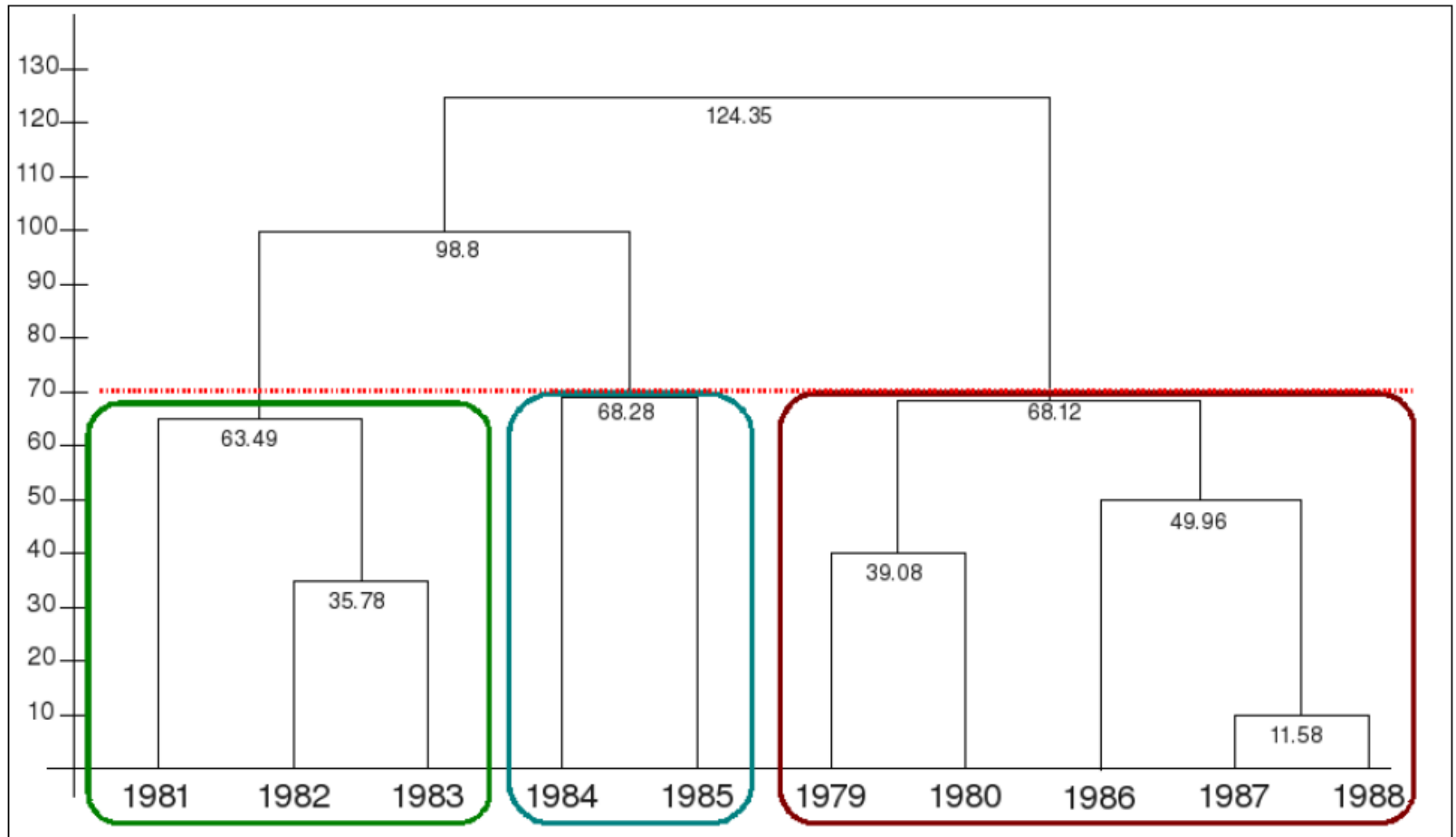


Ilustración 4: Importaciones CA - México, dendrograma realizado con agrupación de valor mínimo

Ejemplo completo “a pie”

	1984	1985	1979-1980	1986-(1987-1988)	1981-(1982-1983)
1984	0	68,28	201,24	128,03	98,8
1985		0	189,81	129,13	140,9
1979-1980			0	68,12	169,56
1986-(1987-1988)				0	124,35
1981-(1982-1983)					0

Tabla 8: Importaciones CA – México, cálculo de jerarquía binaria, paso 5

	1984	1985	1981-(1982-1983)	(1979-1980)-(1986-(1987-1988))
1984	0	68,28	98,8	128,03
1985		0	140,9	129,13
1981-(1982-1983)			0	124,35
(1979-1980)-(1986-(1987-1988))				0

Tabla 9: Importaciones CA – México, cálculo de jerarquía binaria, paso 6

Ejemplo completo “a pie”

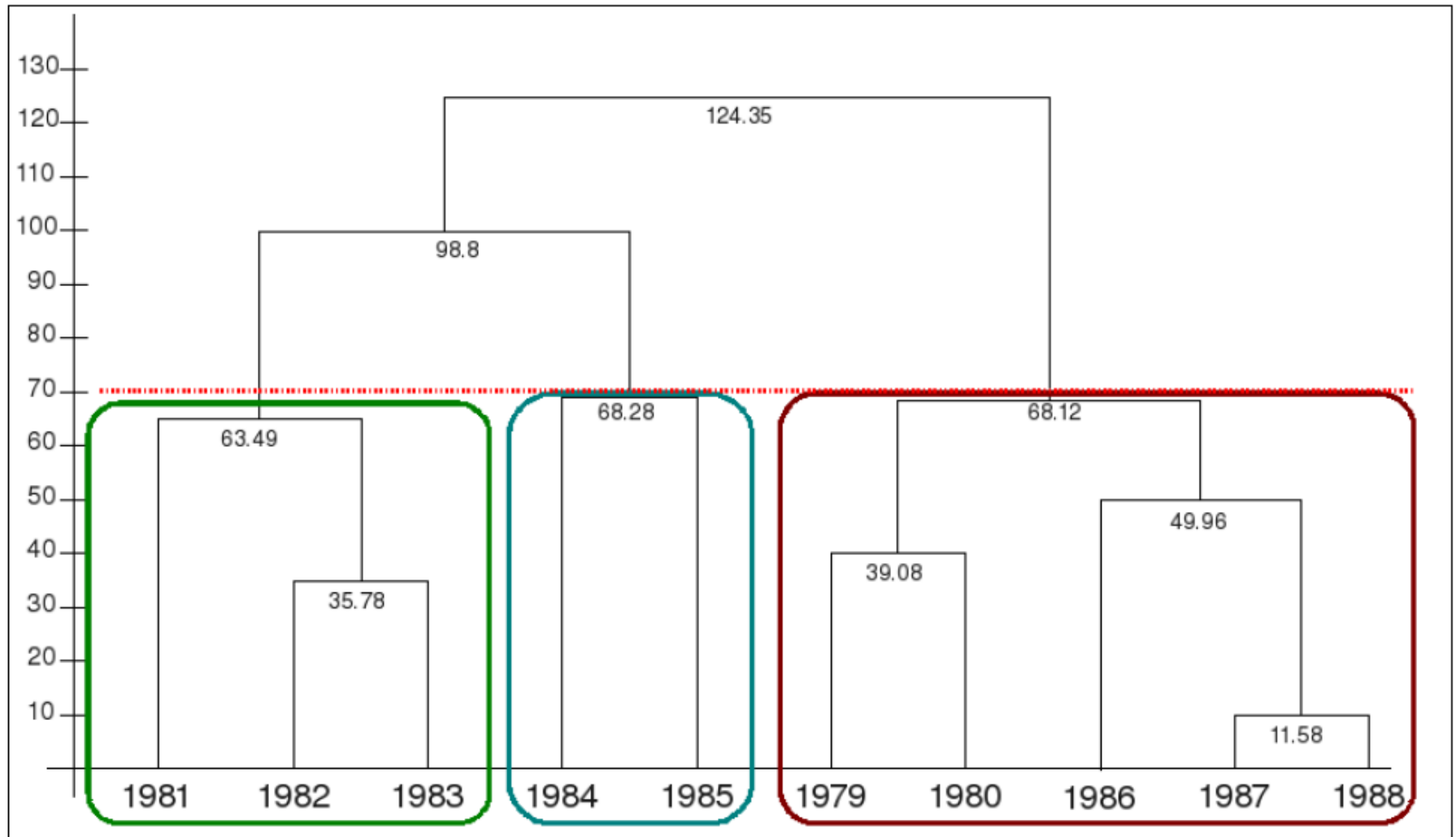


Ilustración 4: Importaciones CA - México, dendrograma realizado con agrupación de valor mínimo

Ejemplo completo “a pie”

	1981-(1982-1983)	(1979-1980)-(1986-(1987-1988))	1984-1985
1981-(1982-1983)	0	124,35	98,8
(1979-1980)-(1986-(1987-1988))		0	128,03
1984-1985			0

Tabla 10: Importaciones CA – México, cálculo de jerarquía binaria, paso 7

	(1979-1980)-(1986-(1987-1988))	(1981-(1982-1983))-(1984-1985)
(1979-1980)-(1986-(1987-1988))	0	124,35
(1981-(1982-1983))-(1984-1985)		0

Tabla 11: Importaciones CA – México, cálculo de jerarquía binaria, paso 8

Ejemplo completo “a pie”

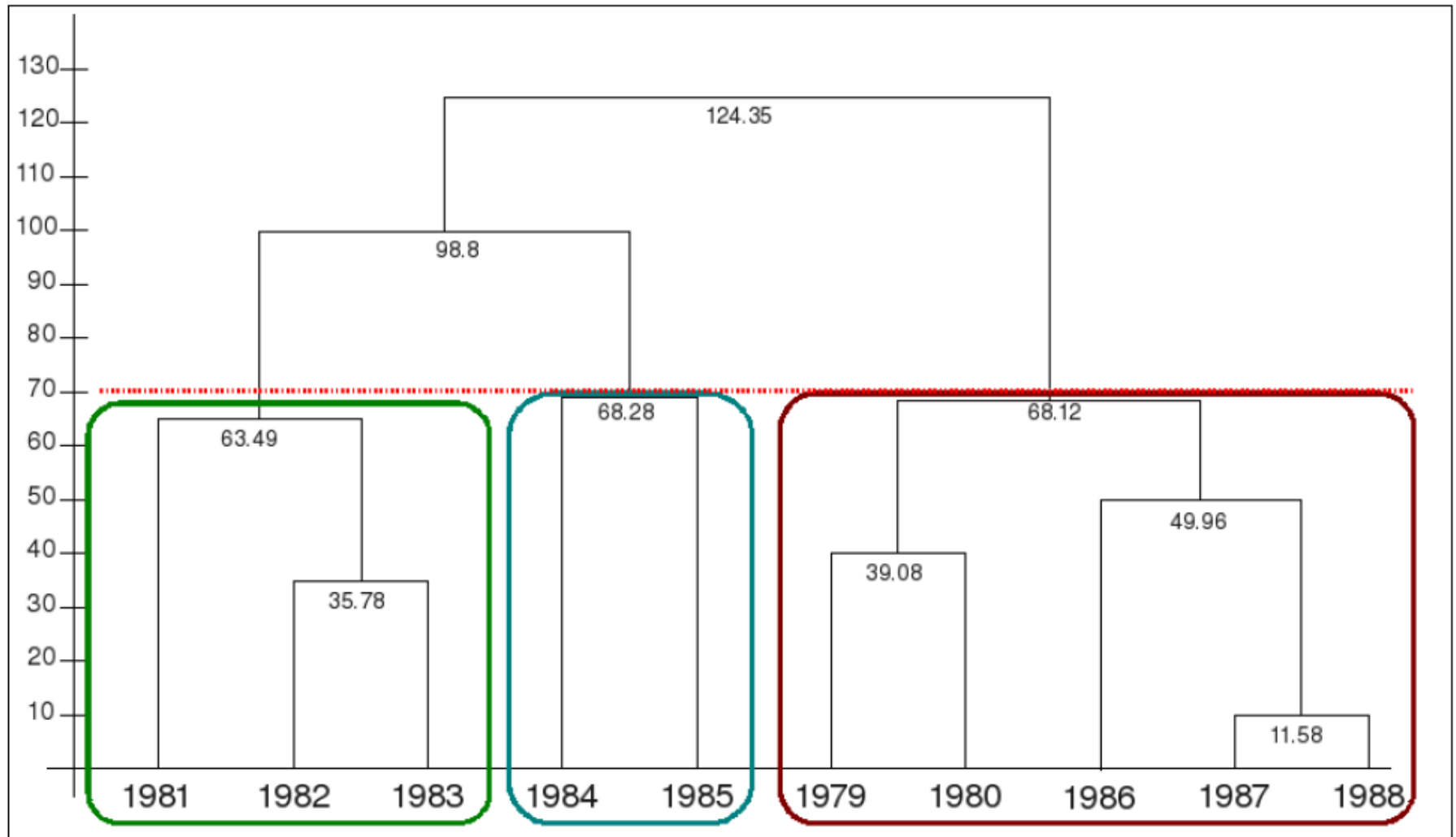


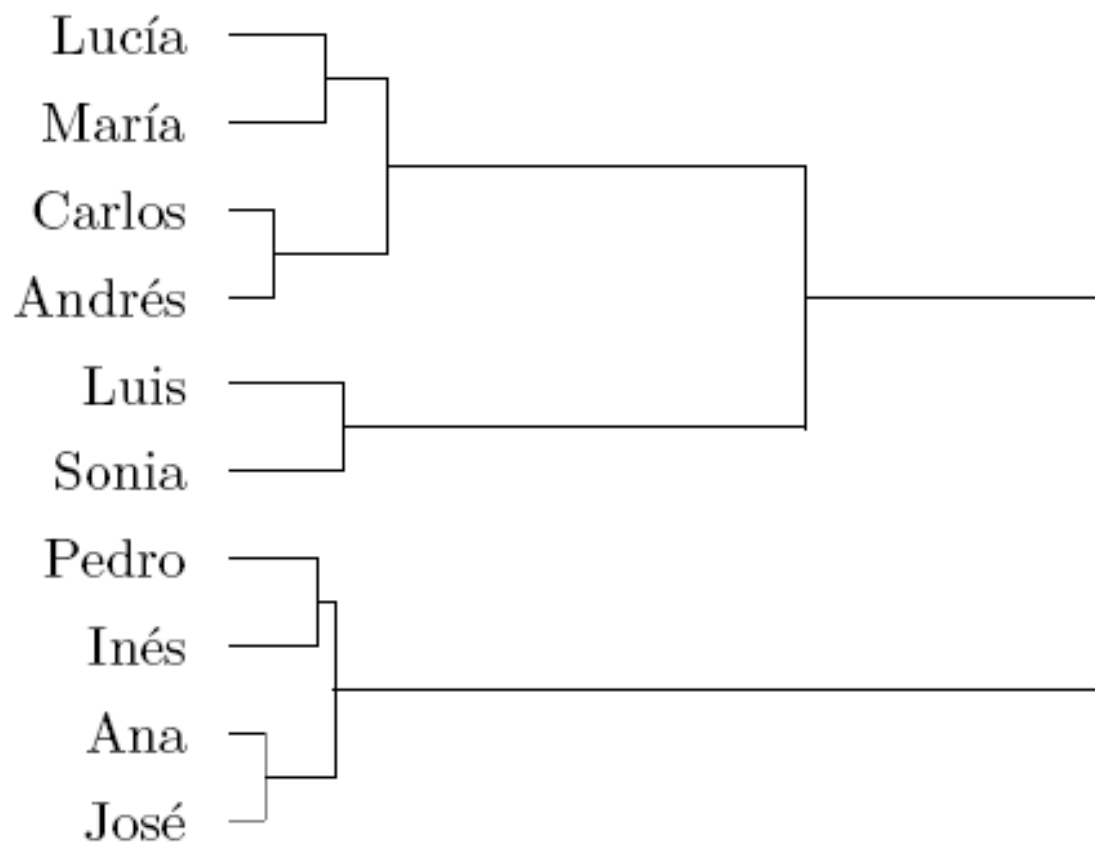
Ilustración 4: Importaciones CA - México, dendrograma realizado con agrupación de valor mínimo

Algoritmo

Para construir una jerarquía binaria se utiliza el *algoritmo general* que resumimos así: Sea $I = \{1, \dots, n\}$ el conjunto del cual se busca una buena partición.

1. *Inicialización*: el procedimiento empieza con las clases que se reducen a un solo elemento, es decir con la partición $P_h = \{\{1\}, \dots, \{n\}\}$, con $h = 0$.
2. *Formación de nuevos nodos*: se fusionan los dos nodos de P_h más cercanos en el sentido de la agregación δ . Es decir, si x y y son estos dos nodos entonces $\delta(x, y) = \min\{\delta(l, k) \mid l, k \in P_h\}$.
3. *Actualización de P_h* : sea $h \leftarrow h + 1$ y $P_h \leftarrow [P_h \cup \{x \cup y\}] - \{x, y\}$.
4. *Test*: Si $h < n - 2$ regresar a 2. En otro caso, hacer la última fusión y terminar.

La clasificación jerárquica usando la agregación de Ward con la distancia Euclídea, da como resultado:



Interpretación

Análisis de los Clústeres

	Matemáticas	Ciencias	Español	Historia	EdFísica
Lucía	7	6.5	9.2	8.6	8
Pedro	7.5	9.4	7.3	7	7
Inés	7.6	9.2	8	8	7.5
Luis	5	6.5	6.5	7	9
Andrés	6	6	7.8	8.9	7.3
Ana	7.8	9.6	7.7	8	6.5
Carlos	6.3	6.4	8.2	9	7.2
José	7.9	9.7	7.5	8	6
Sonía	6	6	6.5	5.5	8.7
María	6.8	7.2	8.7	9	7

Centro Gravedad C1={Pedro,Inés,Ana,José}

Matemáticas	Ciencias	Español	Historia	EdFísica
7.7	9.475	7.625	7.75	6.75

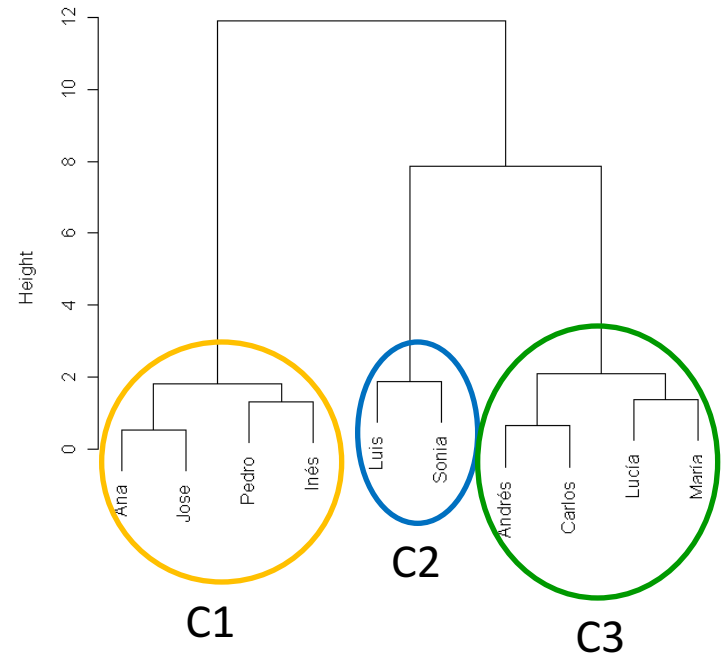
Centro Gravedad C2={Luis,Sonía}

Matemáticas	Ciencias	Español	Historia	EdFísica
5.5	6.25	6.5	6.25	8.85

Centro Gravedad C3={Lucía,Andrés,Carlos,María}

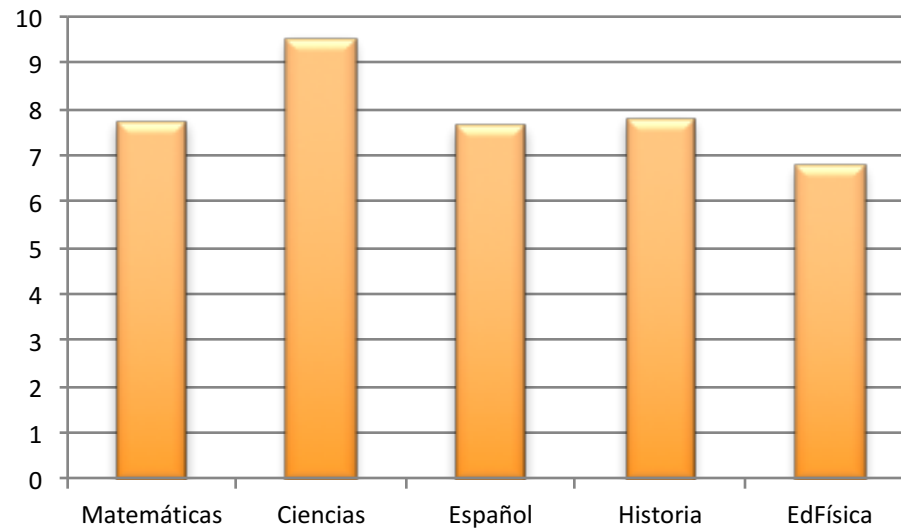
Matemáticas	Ciencias	Español	Historia	EdFísica
6.525	6.525	8.475	8.875	7.375

Cluster Dendrogram for Solution HClust.3



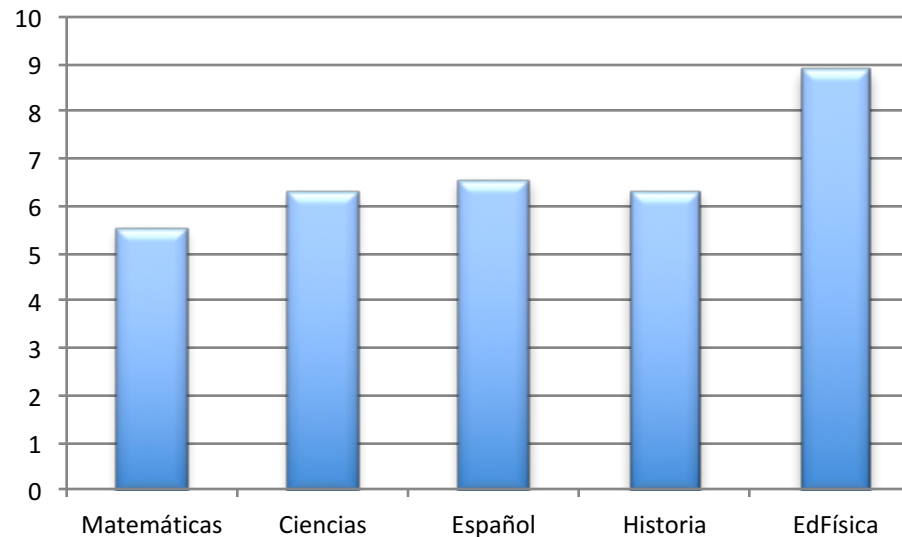
Interpretación Horizontal

Centro Gravedad $C1=\{\text{Pedro, Inés, Ana, José}\}$: Son los estudiantes buenos en Ciencias, Matemáticas y promedio en las demás materias.



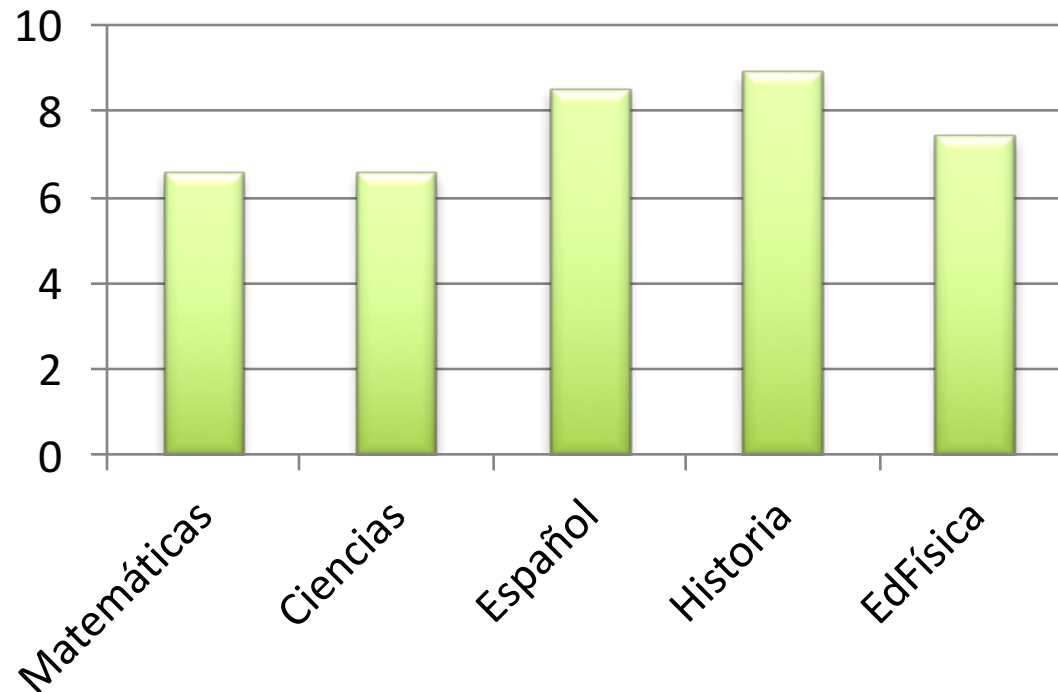
Interpretación Horizontal

Centro Gravedad C2={Luis,Sonia}: Son los estudiantes buenos en Educación Física y estudiantes de regulares a malos en las demás materias.



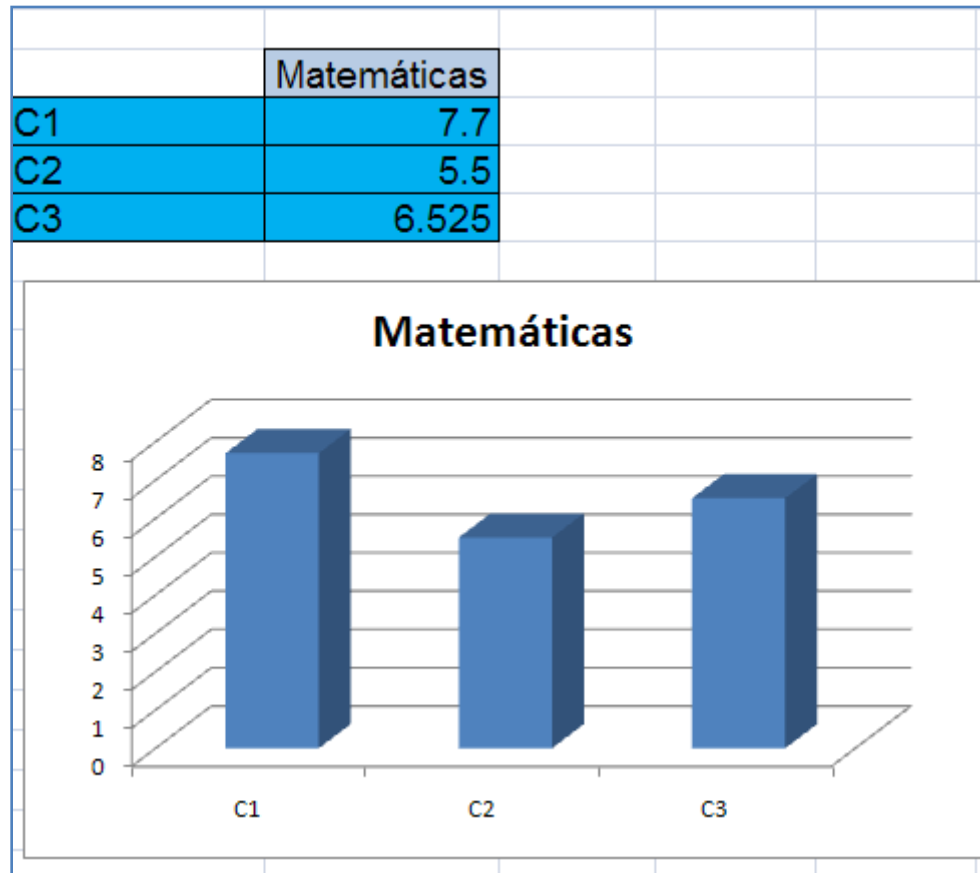
Interpretación Horizontal

Centro Gravedad C3={Lucía,Andrés,Carlos,María}. Son los estudiantes buenos en letras, es decir, Español e Historia, además son estudiantes promedio en las demás materias.



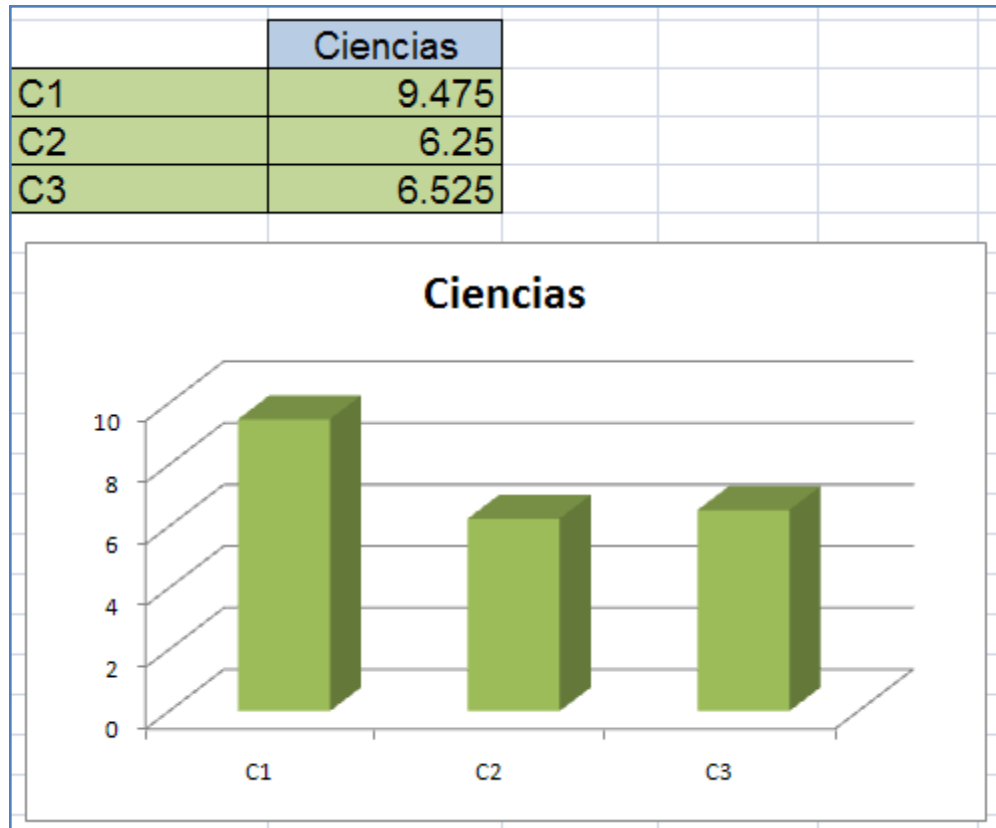
Interpretación Vertical

El clúster C1 es el mejor en Matemática



Interpretación Vertical

El clúster C1 es el mejor en Ciencias



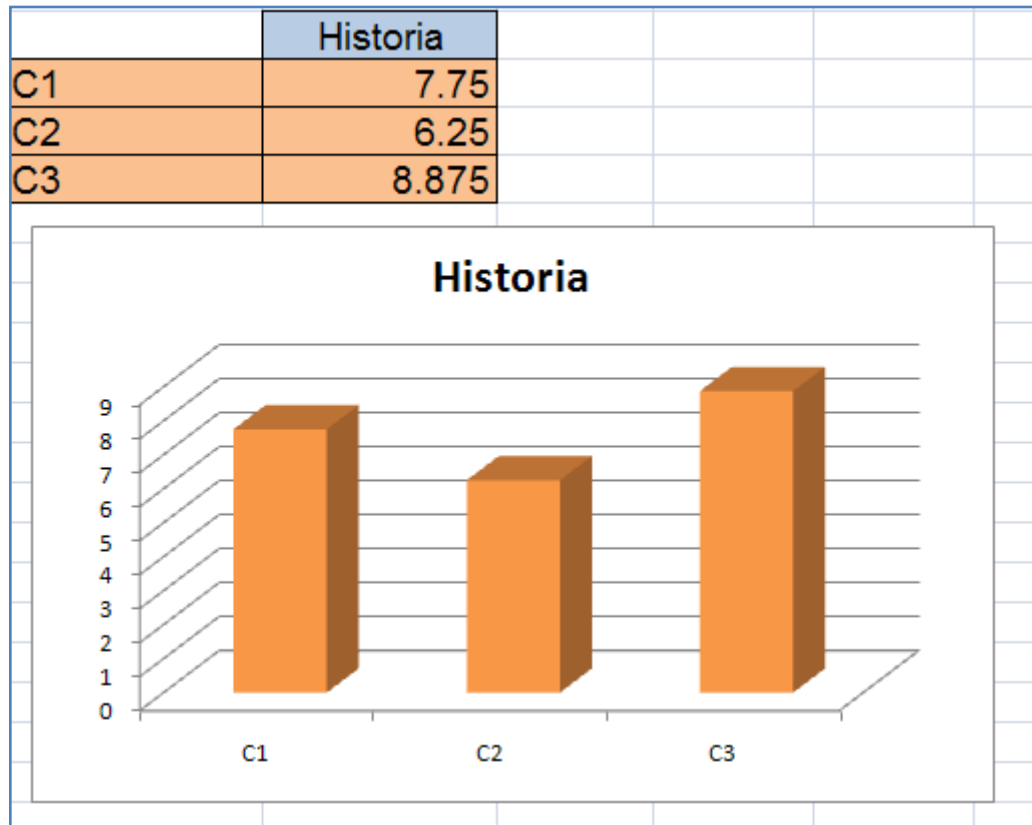
Interpretación Vertical

El clúster C3 es el mejor en Español



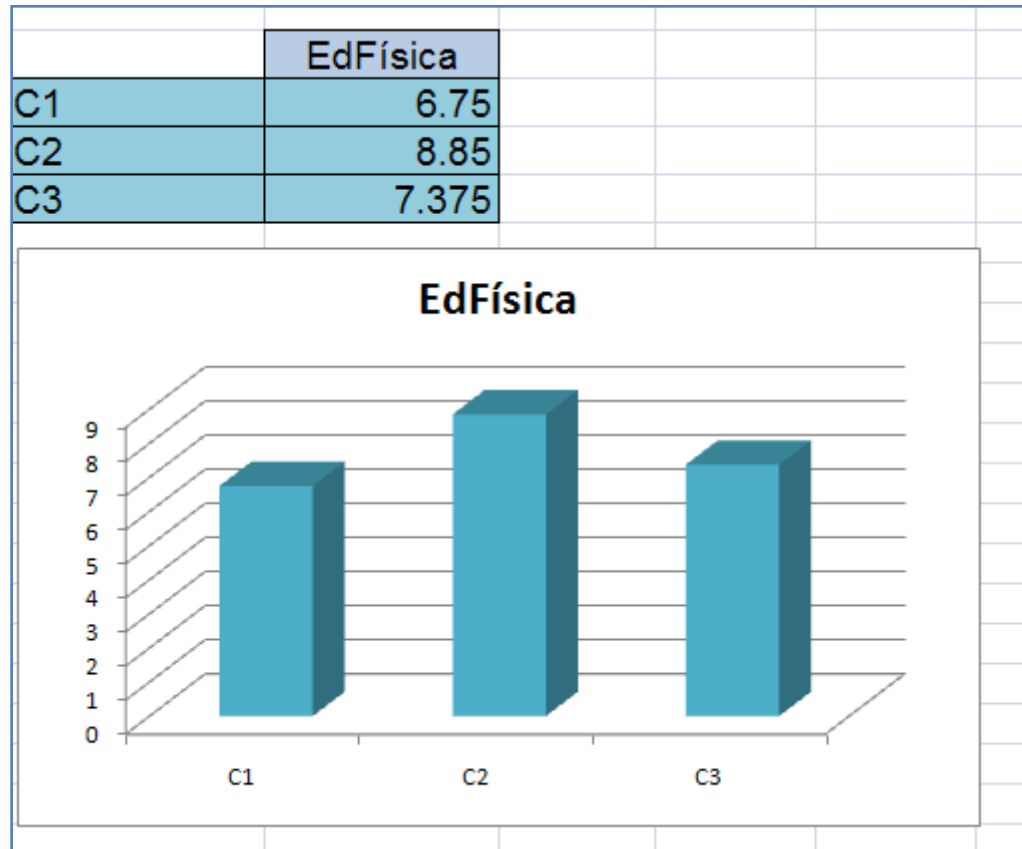
Interpretación Vertical

El clúster C3 es el mejor en Historia

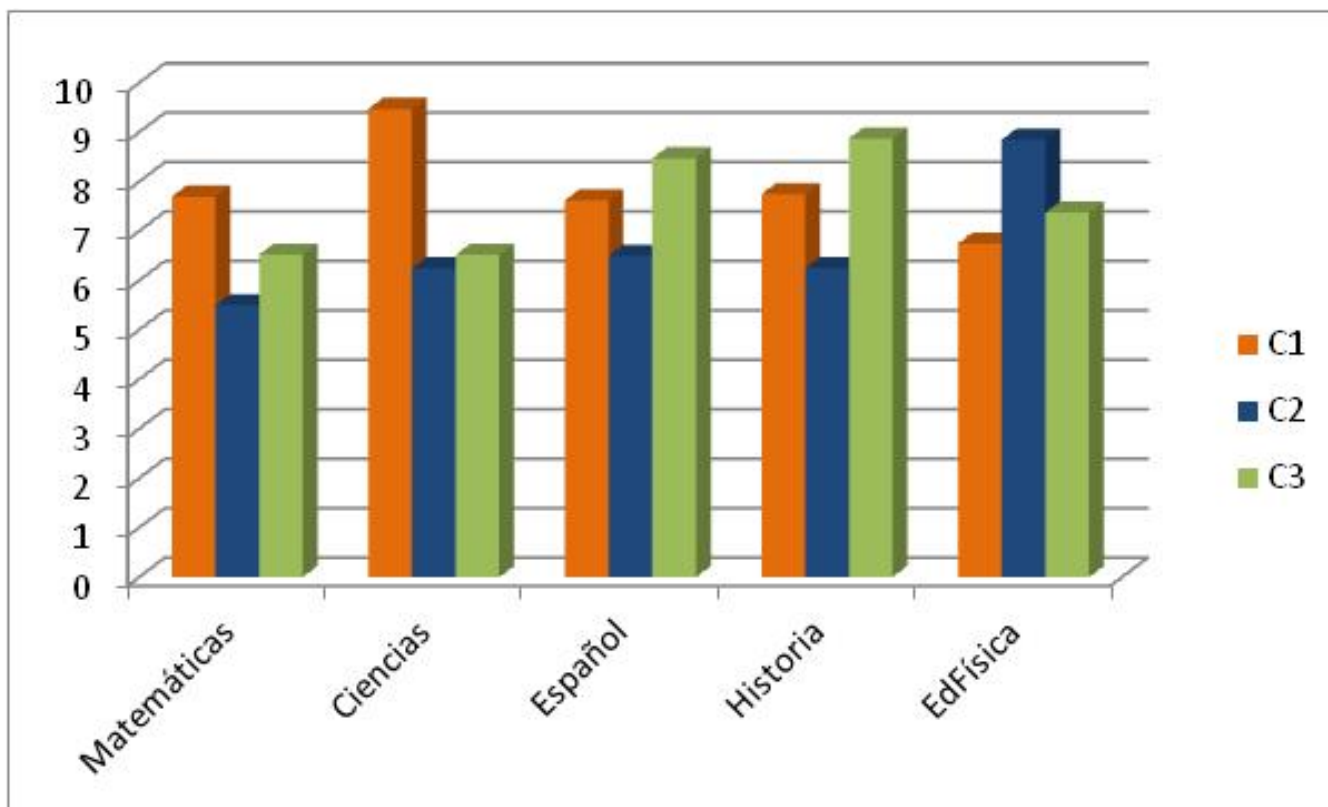


Interpretación Vertical

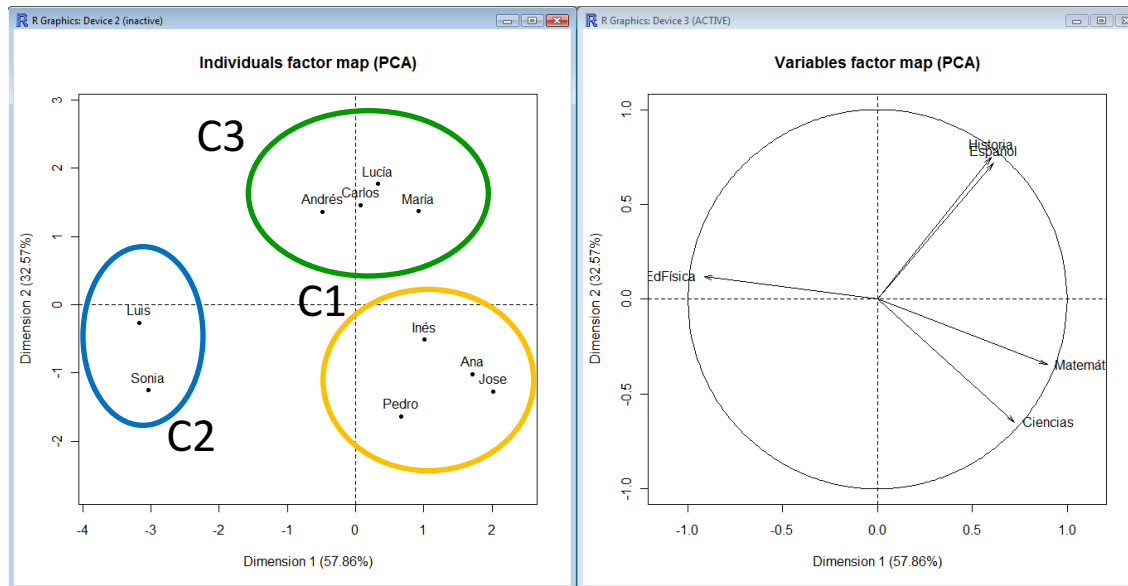
El clúster C2 es el mejor en Educación Física



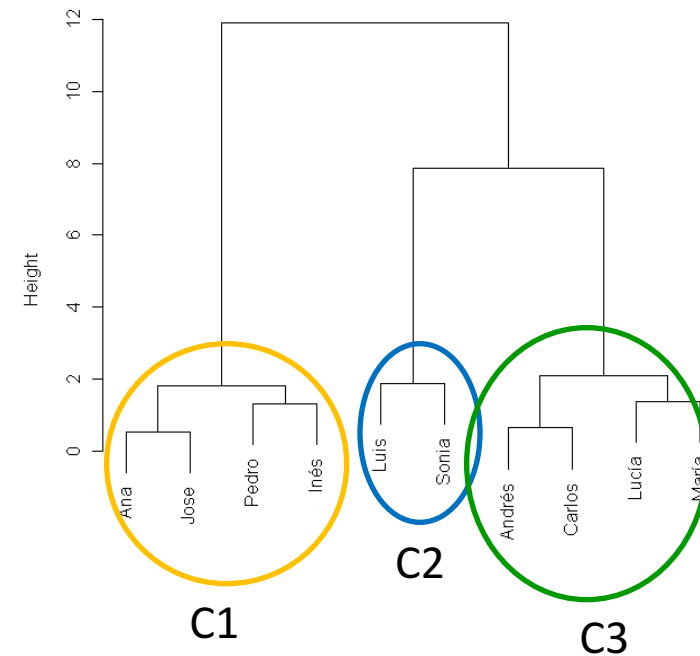
Interpretación Horizontal-Vertical



Comparación entre ACP y CJ



Cluster Dendrogram for Solution HClust.3





PROMiDAT

IBEROAMERICANO

Programa Iberoamericano de
Formación en Minería de Datos

Gracias....