```
x
1,2,3
4,5,4
3,6,4
8,5,4
3,2,7
```

$$D = \left[\begin{array}{c|cccc} & 2 & 3 & 4 & 5 \\ \hline 1 & 4.36 & 4.58 & 7.68 & 4.47 \\ 2 & & \underline{1.41} & 4.00 & 4.36 \\ 3 & & & 5.10 & 5.00 \\ 4 & & & & 6.56 \end{array}\right]$$

single linkage ("connected")

$$\left[\begin{array}{c|ccc} & 2,3 & 4 & 5 \\ \hline 1 & 4.36 & 7.68 & 4.47 \\ 2,3 & & \underline{4.00} & 4.36 \\ 4 & & & 6.56 \end{array}\right]$$

$$\left[\begin{array}{c|ccc} & 2,3,4 & 5 \\ \hline 1 & \underline{4.36} & 4.47 \\ 2,3,4 & & \underline{4.36} \end{array}\right]$$

$(5,1,(4,(2,3)))$

average

$$\left[\begin{array}{c|ccc} & 2,3 & 4 & 5 \\ \hline 1 & \underline{4.47} & 7.68 & 4.47 \\ 2,3 & & 4.55 & 4.68 \\ 4 & & & 6.56 \end{array}\right]$$

$$\left[\begin{array}{c|cc} & 4 & 5 \\ \hline 1,2,3 & 6.12 & \underline{4.58} \\ 4 & & 6.56 \end{array}\right]$$

$(4,(5,(1,(2,3))))$

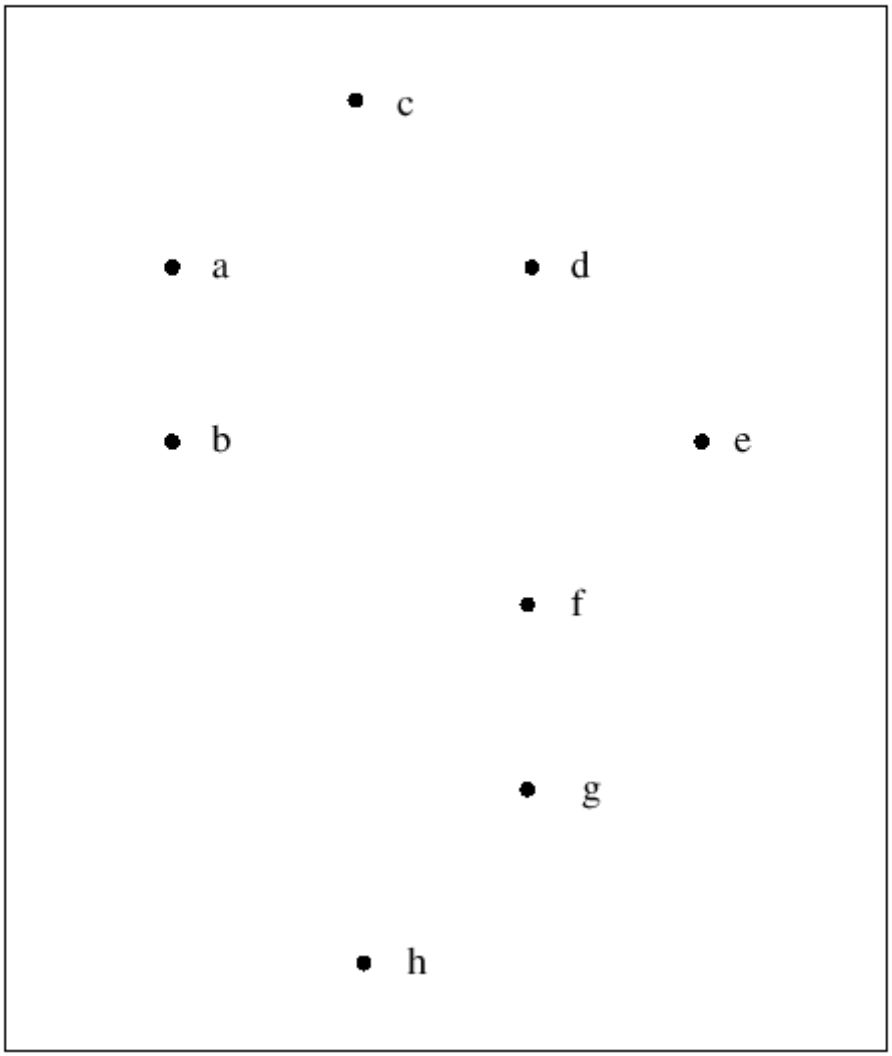
complete linkage ("compact")

$$\left[\begin{array}{c|ccc} & 2,3 & 4 & 5 \\ \hline 1 & 4.58 & 7.68 & \underline{4.47} \\ 2,3 & & 5.10 & 5.00 \\ 4 & & & 6.56 \end{array}\right]$$

$$\left[\begin{array}{c|cc} & 2,3 & 4 \\ \hline 1,5 & \underline{4.58} & 7.68 \\ 2,3 & & 5.10 \end{array}\right]$$

$(4,((2,3),(1,5)))$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $a$ | (0,4) | $c$ | (1,5) | $e$ | (3,3) | $g$ | (2,1) |
| $b$ | (0,3) | $d$ | (2,4) | $f$ | (2,2) | $h$ | (1,0) |




| $\delta$ | $b$ | $c$ | $d$ | $e$ | $f$ | $g$ | $h$ |
|---|---|---|---|---|---|---|---|
| $a$ | 1 | 2 | 4 | 10 | 8 | 13 | 17 |
| $b$ | | 5 | 5 | 9 | 5 | 8 | 10 |
| $c$ | | | 2 | 8 | 10 | 17 | 25 |
| $d$ | | | | 2 | 4 | 9 | 17 |
| $e$ | | | | | 2 | 5 | 13 |
| $f$ | | | | | | 1 | 5 |
| $g$ | | | | | | | 2 |

**Single Linkage:**

| | $c$ | $d$ | $e$ | $fg$ | $h$ |
|---|---|---|---|---|---|
| $ab$ | 2 | 4 | 9 | 5 | 10 |
| $c$ | | 2 | 8 | 10 | 25 |
| $d$ | | | 2 | 4 | 17 |
| $e$ | | | | 2 | 13 |
| $fg$ | | | | | 2 |

| $F_{sl}(\delta)$ | $b$ | $c$ | $d$ | $e$ | $f$ | $g$ | $h$ |
|---|---|---|---|---|---|---|---|
| $a$ | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| $b$ | | 2 | 2 | 2 | 2 | 2 | 2 |
| $c$ | | | 2 | 2 | 2 | 2 | 2 |
| $d$ | | | | 2 | 2 | 2 | 2 |
| $e$ | | | | | 2 | 2 | 2 |
| $f$ | | | | | | 1 | 2 |
| $g$ | | | | | | | 2 |

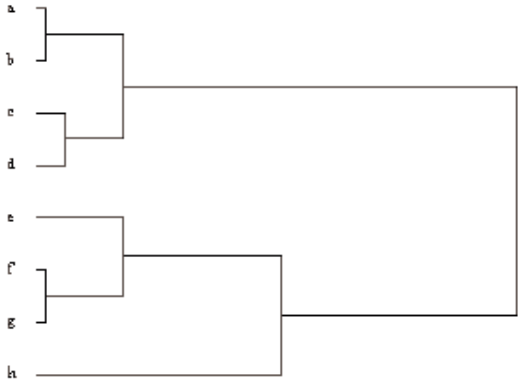
**Complete Linkage:**

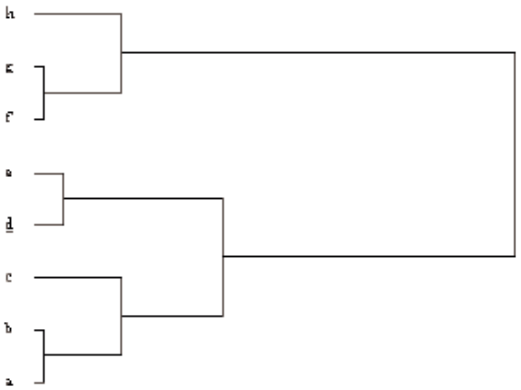
| | $c$ | $d$ | $e$ | $fg$ | $h$ |
|---|---|---|---|---|---|
| $ab$ | 5 | 5 | 10 | 13 | 17 |
| $c$ | | 2 | 8 | 17 | 25 |
| $d$ | | | 2 | 9 | 17 |
| $e$ | | | | 5 | 13 |
| $fg$ | | | | | 5 |

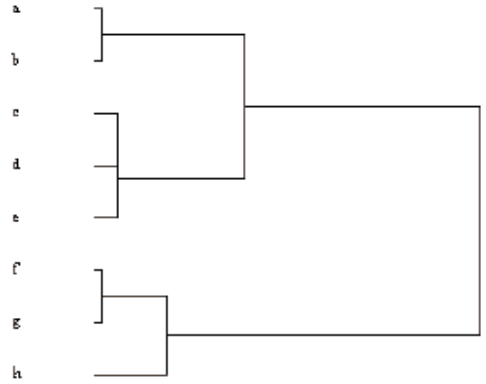
| | $cd$ | $e$ | $fg$ | $h$ |
|---|---|---|---|---|
| $ab$ | 5 | 10 | 13 | 17 |
| $cd$ | | 8 | 17 | 25 |
| $e$ | | | 5 | 13 |
| $fg$ | | | | 5 |

| | $efg$ | $h$ |
|---|---|---|
| $abcd$ | 17 | 25 |
| $efg$ | | 13 |

| $F_{cl}(\delta)$ | $b$ | $c$ | $d$ | $e$ | $f$ | $g$ | $h$ |
|---|---|---|---|---|---|---|---|
| $a$ | 1 | 5 | 5 | 25 | 25 | 25 | 25 |
| $b$ | | 5 | 5 | 25 | 25 | 25 | 25 |
| $c$ | | | 2 | 25 | 25 | 25 | 25 |
| $d$ | | | | 25 | 25 | 25 | 25 |
| $e$ | | | | | 5 | 5 | 13 |
| $f$ | | | | | | 1 | 13 |
| $g$ | | | | | | | 13 |



**UPGMA:** *unweighted pair group analysis*

| | $c$ | $d$ | $e$ | $fg$ | $h$ |
|---|---|---|---|---|---|
| $ab$ | 3.5 | 4.5 | 9.5 | 8.5 | 13.5 |
| $c$ | | 2 | 8 | 13.5 | 25 |
| $d$ | | | 2 | 6.5 | 17 |
| $e$ | | | | 3.5 | 13 |
| $fg$ | | | | | 3.5 |

| | $cd$ | $e$ | $fg$ | $h$ |
|---|---|---|---|---|
| $ab$ | 4 | 9.5 | 8.5 | 13.5 |
| $cd$ | | 5 | 10 | 21 |
| $e$ | | | 3.5 | 13 |
| $fg$ | | | | 3.5 |

| | $cd$ | $efg$ | $h$ |
|---|---|---|---|
| $ab$ | 4 | 8.83 | 13.5 |
| $cd$ | | 8.33 | 21 |
| $efg$ | | | 6.67 |

| | $efg$ | $h$ |
|---|---|---|
| $abcd$ | 8.58 | 17.25 |
| $efg$ | | 6.67 |

| $F_{upgma}(\delta)$ | $b$ | $c$ | $d$ | $e$ | $f$ | $g$ | $h$ |
|---|---|---|---|---|---|---|---|
| $a$ | 1 | 4 | 4 | 10.75 | 10.75 | 10.75 | 10.75 |
| $b$ | | 4 | 4 | 10.75 | 10.75 | 10.75 | 10.75 |
| $c$ | | | 2 | 10.75 | 10.75 | 10.75 | 10.75 |
| $d$ | | | | 10.75 | 10.75 | 10.75 | 10.75 |
| $e$ | | | | | 3.5 | 3.5 | 6.67 |
| $f$ | | | | | | 1 | 6.67 |
| $g$ | | | | | | | 6.67 |

**WPGMA:** *weighted average linkage*

| | $cd$ | $efg$ | $h$ |
|---|---|---|---|
| $ab$ | 4 | 9 | 13.5 |
| $cd$ | | 7.5 | 21 |
| $efg$ | | | 8.25 |

| | $efg$ | $h$ |
|---|---|---|
| $abcd$ | 8.25 | 17.25 |
| $efg$ | | 8.25 |

| | $h$ |
|---|---|
| $abcdefg$ | 12.75 |

Here now is the final DC.

| $F_{upgma}(\delta)$ | $b$ | $c$ | $d$ | $e$ | $f$ | $g$ | $h$ |
|---|---|---|---|---|---|---|---|
| $a$ | 1 | 4 | 4 | 8.25 | 8.25 | 8.25 | 12.75 |
| $b$ | | 4 | 4 | 8.25 | 8.25 | 8.25 | 12.75 |
| $c$ | | | 2 | 8.25 | 8.25 | 8.25 | 12.75 |
| $d$ | | | | 8.25 | 8.25 | 8.25 | 12.75 |
| $e$ | | | | | 3.5 | 3.5 | 12.75 |
| $f$ | | | | | | 1 | 12.75 |
| $g$ | | | | | | | 12.75 |

distance between two clusters: distance between their centroids

**UPGMC:** centroid method

(centroid of a merged cluster: centroid of all $x \in X$ and that of all $y \in Y$)

**WPGMC:** median method

(centroid of a merged cluster: centroid of the two centroids)

general clustering strategy

$$\delta(XY, Z) = \alpha_X \delta(X, Z) + \alpha_Y \delta(Y, Z) + \beta \delta(X, Y)$$

UPGMC : $\alpha_X = \dfrac{|X|}{|X| + |Y|}$, $\alpha_Y = \dfrac{|Y|}{|X| + |Y|}$, $\beta = \dfrac{-|X||Y|}{(|X| + |Y|)^2}$

WPGMC : $\alpha_X = \alpha_Y = 1/2$ with $\beta = -1/4$

**Warning:** It sometimes happens $\delta(XY, Z) < \min\{\delta(X, Z), \delta(Y, Z)\}$

$A(-.5, 0)$, $B(.5, 0)$, $C(0, K)$, $3/4 < K^2 < 1$

$\delta(A, B) = 1$, and $\delta(A, C) = \delta(B, C) = K^2 + 1/4 > 1$

$\delta(AB, C) = K^2 < 1$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $a$ | (0,4) | $c$ | (1,5) | $e$ | (3,3) | $g$ | (2,1) |
| $b$ | (0,3) | $d$ | (2,4) | $f$ | (2,2) | $h$ | (1,0) |

| $\delta$ | $b$ | $c$ | $d$ | $e$ | $f$ | $g$ | $h$ |
|---|---|---|---|---|---|---|---|
| $a$ | 1 | 2 | 4 | 10 | 8 | 13 | 17 |
| $b$ | | 5 | 5 | 9 | 5 | 8 | 10 |
| $c$ | | | 2 | 8 | 10 | 17 | 25 |
| $d$ | | | | 2 | 4 | 9 | 17 |
| $e$ | | | | | 2 | 5 | 13 |
| $f$ | | | | | | 1 | 5 |
| $g$ | | | | | | | 2 |

$2^{n-1} - 1$ distinct partitions with two classes.

127 partitions to consider.

**diameter** of $A$: $\max\{d(x, y) : x, y \in A\}$, and
**within cluster sum** of $A$: $\sum\{d(x, y) : x, y \in A\}$

| | $abcdef$ $gh$ | $abcde$ $fgh$ | $abcd$ $efgh$ | $acdefg$ $bgh$ | $acde$ $bfgh$ |
|---|---|---|---|---|---|
| sum of sums | 9 | 2 | 1 | 6 | 5 |
| max of sums | 24 | 6 | 1 | 9 | 2 |
| sum of diameters | 1 | 2 | 4 | 6 | 6 |
| max of diameters | 1 | 1 | 2 | 1 | 1 |
| within cluster DC | 4 | 2 | 1 | 7 | 3 |

different criteria produce different optimal bipartitions.
Computer heuristics need not produce the best results.