Hw 5.

2019150445 신백주

Q1.

a)

$$D = \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 0 & & & & \\ 13 & 0 & & & \\ 32 & 5 & 0 & & \\ 50 & 13 & 2 & 0 & \\ 20 & 5 & 4 & 10 & 0 \end{bmatrix} \end{array}$$

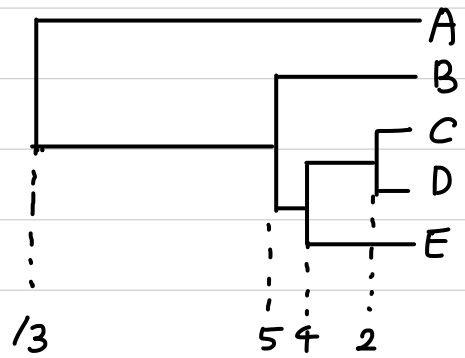
b)

① Since d_{CD} is the Smallest entry in the matrix D,
C and D are merged first.

② $d_{(CD)E}$ is the Smallest distance between A, B, (CD), E.

③ $d_{(CDE)B}$ is the Smallest distance between A, B, (CDE)

④ Combine A & (BCDE).

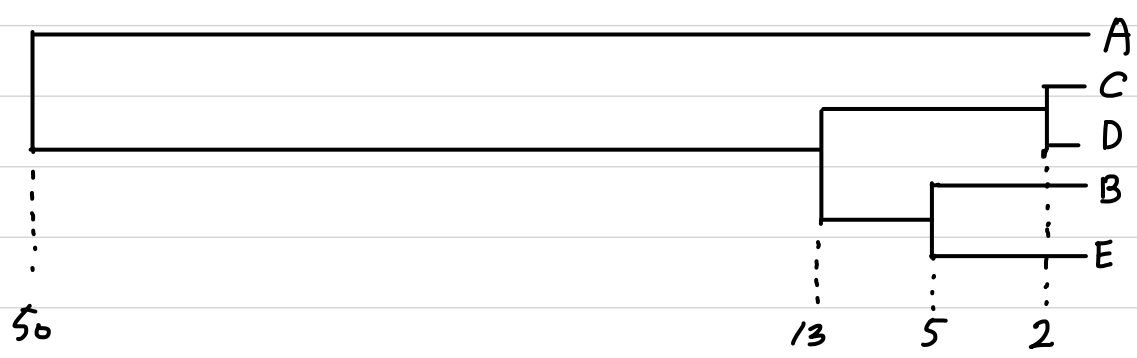


C) ① Since d_{CD} is the Smallest entry in the matrix D,
C and D are merged first.

② d_{BE} is the Smallest distance between A, B, (CD), E.

③ $d_{(CD)(BE)}$ is the Smallest distance between A, B, (CDE)

④ Combine A & (BCDE).



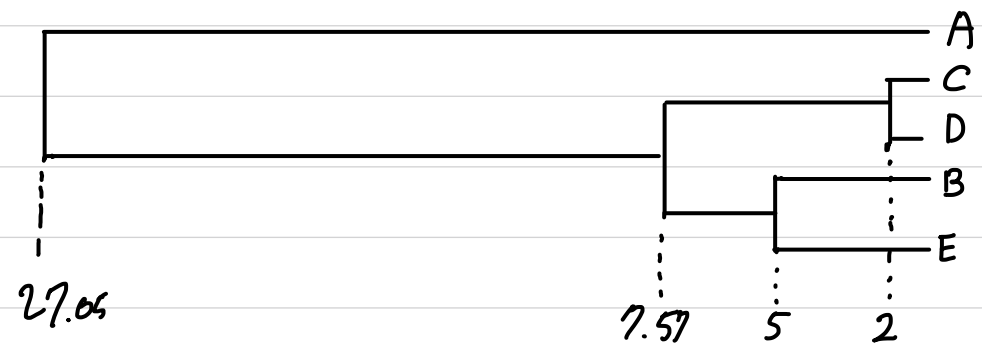
d)

① Since d_{CD} is the smallest entry in the matrix D.
C and D are merged first.

② d_{BE} is the smallest distance between A, B, (CD), E.

③ $d_{(CD)(BE)}$ is the smallest distance between A, B, (CDE)

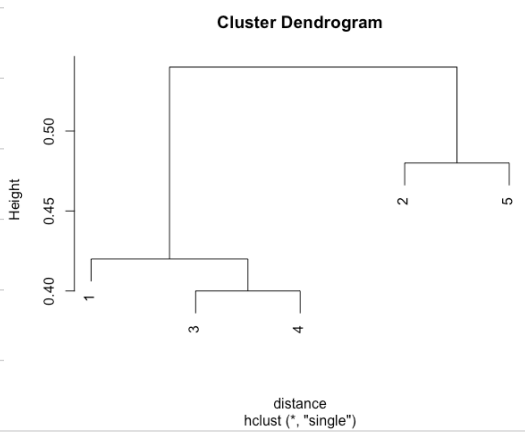
④ Combine A & (BCDE).



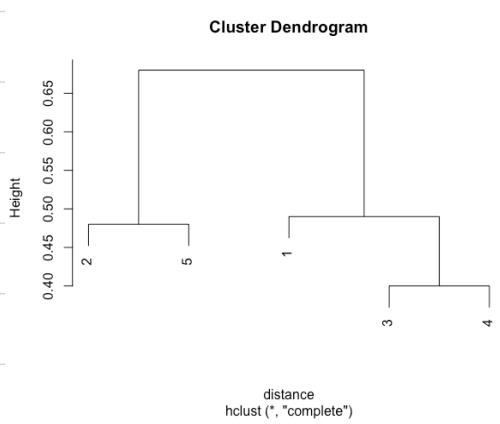
e) For (b) at the second step (CD)&E are merged.
But for (c) & (d) at the second step, B&E are merged.

Q2.

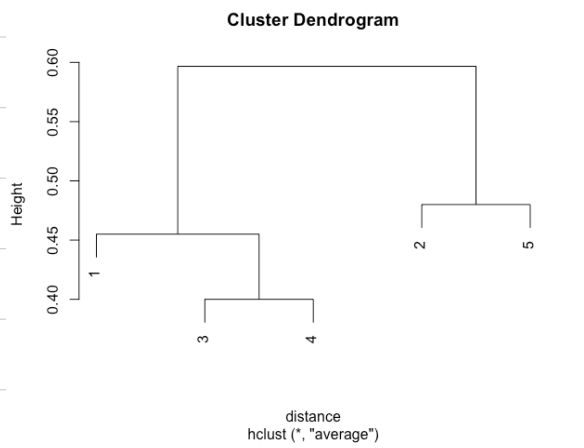
a)



b)



C)



d) All results from (a),(b),(c) are same except distances.

Q4. a) ① Coordinates of centroids

Cluster	X_1	X_2
(AB)	3	3.5
(CD)	1	1

② Squared distances.

$d^2_{A(AB)} = 2^2 + 0.5^2 = 4.25$

$d^2_{A(CD)} = 4^2 + 3^2 = 25$

$\rightarrow d^2_{A(AB)} < d^2_{A(CD)}$

$d^2_{B(AB)} = 2^2 + 2^2 = 8$

$d^2_{B(CD)} = 0^2 + 2^2 = 4$

$\rightarrow d^2_{B(AB)} > d^2_{B(CD)}$

$\therefore$ B is reassigned to cluster (CD)

③ Coordinates of centroids

Cluster	X_1	X_2
(A)	5	4
(BCD)	1	$5/3$

④ $d^2_{AA} < d^2_{A(BCD)}$

$d^2_{BA} = 17 > d^2_{B(BCD)} = 4/9$

$d^2_{CA} = 45 > d^2_{C(BCD)} = 40/9$

$d^2_{DA} = 13 > d^2_{D(BCD)} = 40/9$

$\therefore$ (A), (BCD) is final clusters

b)

① Coordinates of Centroids

Cluster	X_1	X_2
(AD)	4	$5/2$
(BC)	0	2

② Squared distances.

$d^2_{A(AD)} = 1^2 + \frac{3}{2}^2 = 13/4$

$d^2_{A(BC)} = 5^2 + 2^2 = 29$

$\rightarrow d^2_{A(AD)} < d^2_{A(BC)}$

$d^2_{B(AD)} = 3^2 + \frac{1}{2}^2 = 37/4$

$d^2_{B(BC)} = 1^2 + 1^2 = 2$

$\rightarrow d^2_{B(AD)} > d^2_{B(BC)}$

$d^2_{C(AD)} = 5^2 + \frac{3}{2}^2 = 109/4$

$d^2_{C(BC)} = 1^2 + 1^2 = 2$

$\rightarrow d^2_{C(AD)} > d^2_{C(BC)}$

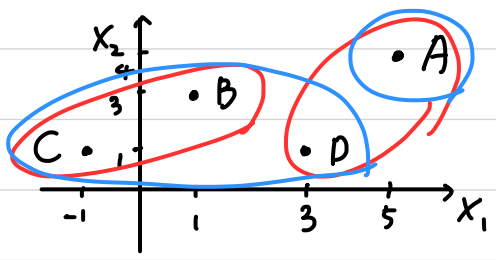
$d^2_{D(AD)} = 1^2 + \frac{3}{2}^2 = 13/4$

$d^2_{D(BC)} = 3^2 + 1^2 = 10$

$\rightarrow d^2_{D(AD)} < d^2_{D(BC)}$

∴ (AD), (BC) are two Clusters.

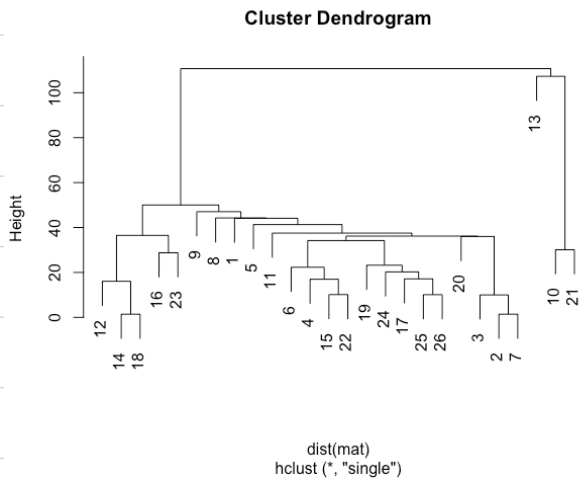
Not same.



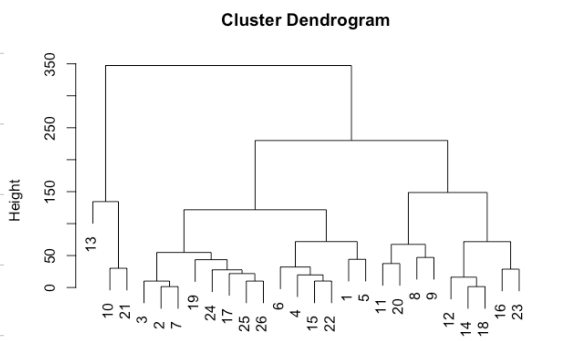
Results are depend on where you assign the initial values.

Q5.

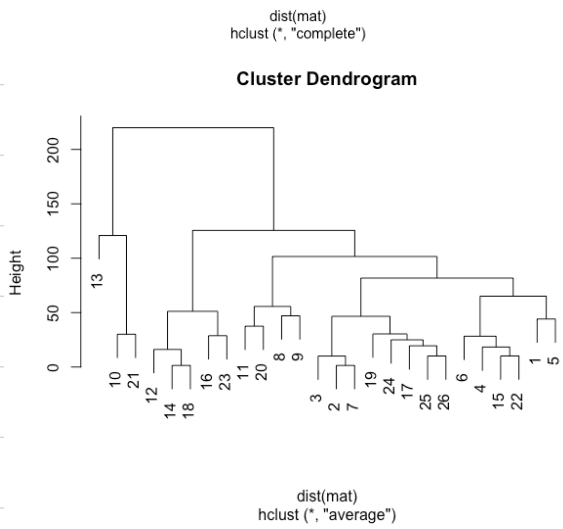
a)



(Single)



(Complete)



(Average)

Clusters are differ depend on Single, Complete, Group.

b)

(1, 4, 5, 6, 15, 17, 19, 22, 24, 25, 26)

(2, 3, 7, 8, 9, 11, 12, 14, 16, 18, 20, 23)

(10, 13, 21)

It is similar with Complete linkage of (a)