1.

The three clusters have center points of A: (0, 1), B: (2, –2), and C: (0, 2)

Calculating the Euclidian distance between the clusters

x	y	Dist A	Dist B	Dist C	New Cluster
2	-1	2.828427	1	3.605551	B
-1	2	1.414214	5	1	C
-2	1	2	5	2.236068	A
1	2	1.414214	4.123106	1	C
4	0	4.123106	2.828427	4.472136	B
4	-1	4.472136	2.236068	5	A
0	-2	3	2	4	B
0	-5	6	3.605551	7	B
-1	0	1.414214	3.605551	2.236068	A
3	8	7.615773	10.04988	6.708204	C
-2	0	2.236068	4.472136	2.828427	A
0	0	1	2.828427	2	A

Cluster C has the fewest data points, with three.

The answer is ~~C~~ D

2.

The means for cluster 1 are 2.3333 and 3.

The variation for cluster 1 is

$$(1-2.3333)^2+(3-3)^2+(0-2.3333)^2+(4-3)^2+(6-2.3333)^2+(2-3)^2=22.6667$$

The means for cluster 2 are 3 and 4.

The variation for cluster 2 is

$$(5-3)^2+(2-4)^2+(1-3)^2+(6-4)^2=16$$

The total within-cluster variation is $2(22.6667+16)=77.33$

The answer is C

3.

I and II are both false.

III is true, as there is no exact method to determine the value of k.

The answer is C.

4.

$Dist(x_1, x_4) = \sqrt{(-1-5)^2 + (0-10)^2} = 11.66$

$Dist(x_1, x_4) = \sqrt{(1-5)^2 + (1-10)^2} = 9.85$

Since complete linkage is used, the maximum distance is chosen, which rounds to 11.7.

The answer is E.

5.

I is false, as all points are assigned to a cluster

II is true, the genogram can be cut at different heights to obtain different numbers of clusters

III is true, as this method has high variance

The answer is D

6.

I is false, as the number of clusters is not pre-specified for hierarchical clustering.

II is false, as both methods are sensitive to outliers, as outliers are still forced into a cluster regardless of method.

III is true, only k-means clustering requires random assignments.

The answer is C

7.

D is applicable to k-means clustering and not hierarchical clustering, as k-means clustering starts with data points randomly assigned to a cluster.

The answer is D

8.

I and II are true, as the rows and columns can be reversed

III is true, as clustering is classified as an unsupervised learning method.

The answer is E, as all three are true.

9.

I is false, as the initial selection of clusters can change the final outcome.

II is true, as hierarchical clustering does not depend on a random selection.

III is false, the two methods are different can result in different assignments.

The answer is B

10.

I is true, cutting the genogram at a lower light can only increase the number of clusters.

II is false, the data does not need to be plotted before using k-means clustering.

III is true, as hierarchical clustering can miss cases where clusters are not nested.

The answer is C

11.

D is not true, and does not refer to clustering.