Clustering Activity

The questions below will refer the to the following pairwise distance matrix:

	A	B	C	D	E	F	G	H
A	0.0	15.5	12.8	10.6	14.1	2.1	13.3	1.0
B	15.5	0.0	5.3	4.9	2.4	16.5	15.8	14.6
C	12.8	5.3	0.0	4.3	6.2	13.2	10.6	11.8
D	10.6	4.9	4.3	0.0	3.8	11.6	13.2	9.7
E	14.1	2.4	6.2	3.8	0.0	15.3	16.4	13.2
F	2.1	16.5	13.2	11.6	15.3	0.0	12.0	2.3
G	13.3	15.8	10.6	13.2	16.4	12.0	0.0	12.6
H	1.0	14.6	11.8	9.7	13.2	2.3	12.6	0.0

1. **Agglomerative Hierarchical Clustering**

   Show the dendrogram that results from performing single-link agglomerative hierarchical clustering on the data above. (In *single-link* hierarchical clustering the distance between two clusters is based on minimum distance between any two points in those clusters.)

2. **DBSCAN**

   Answer the following questions, assuming that the DBSCAN clustering algorithm has been executed with $Eps = 5$, $MinPts = 3$.

   (a) Which points are *core points*?

   (b) Which points are *border points*?

   (c) Which points are *noise points*?

   (d) How many clusters are there, and which points belong to which clusters?

3. **Working with Distance Matrices**

   Which of the following algorithms can we perform using only the distance matrix above, without having access to the original data points?

   (a) Principal Components Analysis

   (b) Multidimensional Scaling

   (c) Isomap

   (d) K-Means Clustering

4. **K-Means vs. DBSCAN**

   For each pair of clusterings below, determine which was performed by K-Means and which was performed by DBSCAN.