Homework 4

1. Given the distance(dissimilarity)table for 6 data objects as shown below. Assuming O2 and O4 are the medoids of the 2 clusters initially. After the objects are distributed to the two clusters, we randomly selected O1 to replace O4 as the new medoid for cluster 2, what's the total cost for this replacement? Should this replacement be carried out?

	Obj1	**Obj2**	**Obj3**	**Obj4**	**Obj5**	**Obj6**
Obj1	--					
Obj2	0.94	--				
Obj3	0.36	0.91	--			
Obj4	0.19	0.75	0.39	--		
Obj5	1.15	1.38	0.99	1.3	--	
Obj6	1.38	2.16	1.22	1.5	1.2	--

2. Perform UPGMA average link hierarchical clustering on 6 objects. The distance/dissimilarity values between pairwise objects are shown in the table below. Show the intermediate distance tables computed as well as the intermediate clustering hierarchies constructed. If we are to recommend K clusters for the data, what will be a reasonable distance threshold to use to obtain the clusters from the dendrogram?

	Obj1	Obj2	Obj3	Obj4	Obj5	Obj6
Obj1	0					
Obj2	0.5	0				
Obj3	4	3.5	0			
Obj4	1.5	2	2.5	0		
Obj5	4.5	1	3	5	0	
Obj6	6	2.5	6.5	4	3.5	0