Assignment 4

Part 1. (50 marks)

The dataset Cereals.csv includes nutritional information for 77 breakfast cereals.

Apply hierarchical clustering to the data using Euclidean distance. Compare the dendrograms from single linkage and complete linkage. Comment on the structure of the clusters and on their stability. Choose one of the methods. How many clusters would you use? Use k-means clustering with the number of clusters that you found above. Does the same picture emerge?

The elementary public schools would like to choose a set of cereals to include in their daily cafeterias. Every day a different cereal is offered, but all cereals should support a healthy diet. For this goal, you are requested to find a cluster of "healthy cereals."

Part 2. (20 marks)

The distance between pairs of five items are as follows

	a	***b***	***c***	***d***	***e***
a	0	4	6	1	6
b	4	0	9	7	3
c	6	9	0	10	5
d	1	7	10	0	8
e	6	3	5	8	0

1. Cluster the five items using the average linkage hierarchical procedure and draw the dendrogram. (10)
2. Cluster the five items using the single linkage hierarchical procedure and draw the dendrogram. (10)