HW#8 (May 6th)

You may work in groups of two. Both names must appear on the turned-in assignment. Note that you may not work with someone who was in your group previously in this class.

1) The following table lists the <u>distances</u> between each of 5 observations in 3 dimensions:

	Obs1	Obs2	Obs3	Obs4	Obs5
Obs1	0	2	5	3	6
Obs2	2	0	0	1	7
Obs3	5	0	0	4	8
Obs4	3	1	4	0	9
Obs5	6	7	8	9	0

All computations and Dendrogram sketches should be done by hand:

    (a) Draw a Dendrogram for Complete-Linkeage.
    (b) Draw a Dendrogram for Average-Linkeage.

For question 2 refer back to the cps_income2 dataset and instructions provided in HW7.

2) State Rstudio code which could be used to accomplish the following:
    (a) Plot a dendrogram for clustering this data using Manhattan Distance and Average-Linkeage.
    (b) Determine which observations are clustered together when using 2 clusters with Manhattan Distance and Average-Linkeage.
    (c) Print the income levels of the observations belonging in the first cluster of (b).

3) Consider the 'cereal.csv' dataset located in the 0.Data folder on Canvas. Use hierarchical clustering with single linkage (where the distance between observations is computed using Euclidean Distance) to compute the following:

    (a) Misclassification rate for the Clustering Algorithm in predicting whether a cereal has a rating above or below 50.
    (b) FPR and TPR for the Clustering Algorithm in predicting whether a cereal has a rating above 50.