Assignment HC: Hierarchical Clustering

1. Using the compiled (mean) Olympic winter sports proximity ratings, apply a hierarchical clustering method, and draw the resulting dendrogram. You have several options:

   a. do single-link clustering, by hand

   b. do average-link clustering (by hand, SPSS, or R)

In one or two sentences, interpret the resulting dendrogram (tree).

```
# R code to read data:
winter<-read.csv("C:/Users/X/olympic dissim data.csv", header=TRUE)
winter
x<-as.matrix(winter)
sports<-x[,1]
sports
x<-x[,2:13]
y<-matrix(as.numeric(x),12,12)
rownames(y)<-sports
colnames(y)<-sports
spdist<-as.dist(y)
```

2. Using the posted multivariate data set on crimes for a sample of American cities from Hartigan, 1975 (file "*Hartigan crimes data_MLT.txt*", or see below), apply Ward's method (using SPSS CLUSTER, or R) to obtain a hierarchical clustering of the cities. In one or two sentences, interpret the resulting dendrogram (tree).

OPTIONAL: Standardize the variables before re-running Ward's method in SPSS or before computing distances and re-running hclust in R.   For a standardization method, use either conversion to z-scores, or write code to normalize by the range. Re-run Ward's method and compare the solution to the one using unstandardized variables.

----

city	MurderMan.	Rape	Robbery	Assault	Burglary	Larceny	AutoTheft
Atlanta	16.5	24.8	106	147	1112	905	494
Boston	4.2	13.3	122	90	982	669	954
Chicago	11.6	24.7	340	242	808	609	645
Dallas	18.1	34.2	184	293	1668	901	602

Denver	6.9	41.5	173	191	1534	1368	780
Detroit	13.0	35.7	477	220	1566	1183	788
Hartford	2.5	8.8	68	103	1017	724	468
Honolulu	3.6	12.7	42	28	1457	1102	637
Houston	16.8	26.6	289	186	1509	787	697
Kansas City	10.8	43.2	255	226	1494	955	765
Los Angeles	9.7	51.8	286	355	1902	1386	862
New Orleans	10.3	39.7	266	283	1056	1036	776
New York	9.4	19.4	522	267	1674	1392	848
Portland	5.0	23.0	157	144	1530	1281	488
Tucson	5.1	22.9	85	148	1206	756	483
Washington	12.5	27.6	524	217	1496	1003	739