Homework 6

1. (10 pts) Consider the matrix of distance

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ 11 & 2 & 0 & \\ 5 & 3 & 4 & 0 \end{bmatrix} \end{array}$$

   Without using R or other software, cluster the four items using each of the following procedures
   a. Single linkage hierarchical procedure.
   b. Complete linkage hierarchical procedure.
   c. Draw the dendrograms and compare the results in a and b.

2. (10 pts) The weekly rates of return for five stocks (JP Morgan, Citibank, Wells Fargo, Royal Dutch Shell, and ExxonMobil) listed on the New York Stock Exchange were determined for the period January 2004 through December 2005. The weekly rates of return are defined as (current week closing price-previous week closing prices)/ previous week closing price), adjusted for stock splits and dividends. The observations in 103 successive weeks appear to be independently distributed, but the rates of return across stocks are correlated, because as one expects, stocks tend to move together in response to general economic conditions.
   Use the data named stock-price on canvas, cluster **the stocks** using the single linkage and complete linkage hierarchical procedures. Draw the dendrograms and compare the results.

3. (10 pts) Suppose we measure two variables X1 and X2 for each of four items A, B, C, and D. The data are as follows:

	Observations	
Item	x_1	x_2
A	5	4
B	1	−2
C	−1	1
D	3	1

   Use the k-means clustering technique to divide the items into k=2 clusters, start with the initial groups (AB) and (CD). Please show all the details. Do not use statistical software.

4. (10 pts) Ex. 6.1 Apply k-means to the **crime rate** data (named as crime on canvas) after standardizing each variable by its standard deviation. Compare the results with those given in the text found by standardizing by a variable's range.