INDIVIDUAL ASSIGNMENT

K-Means Algorithm: Assume the following dataset is given:(2,2),(4,4),(5,5),(6,6), (8,8),(9,9),(0,4),(4,0). Given the following information:

* k=4 to cluster the dataset.
* Euclidean distance is used as the distance function to compute distances between centroids and objects in the dataset.
* The initial clusters C1, C2, C3, and C4 are as follows:

C1: {(2,2), (4,4), (6,6)}

C2: {(0,4), (4,0)}

C3: {(5,5), (9,9)}

C4: {(8,8}}

You have to perform the calculation and use the R code to:

1. Calculate what are their centroids for the given dataset?

C1 <- matrix( c ( 2, 2, 4, 4, 6, 6), nrow = 2, byrow = FALSE )  
C2 <- matrix( c( 0, 4, 4, 0), nrow = 2, byrow = FALSE )  
C3 <- matrix( c( 5, 5, 9, 9), nrow = 2, byrow = FALSE )

K1 <- matrix( ( (C1[1, 1:3])/3 , (C1[2, 1:3])/3 ), nrow = 2, byrow = FALSE )  
K2 <- matrix( ( (C2[1, 1:2])/2 , (C2[2, 1:2])/2 ), nrow = 2, byrow = FALSE )  
K3 <- matrix( ( (C3[1, 1:2])/2 , (C3[2, 1:2])/2 ), nrow = 2, byrow = FALSE )  
K3 <- matrix( (8,8), nrow = 2, byrow = FALSE )

1. Identify the new clusters centroid for the given dataset.

point <- matrix ( (2, 2, 4, 4, 6, 6, 0, 4, 4, 0, 5, 5, 9, 9, 8, 8), nrow = 2,   
 byrow = FALSE)

group <- matrix( (0, 1, 0, 0, 0, 1, 0, 0,  
 1, 0, 0, 1, 1, 0, 0, 0,  
 0, 0, 1, 0, 0, 0, 0, 0,  
 0, 0, 0, 0, 0, 0, 1, 1), nrow =4, byrow = TRUE)

K1 <- matrix( ( ( point[1,2] + point[1,6])/2,   
 ( point[2,2] + point[2,6])/2),  
 nrow =2, byrow = FALSE)  
K2 <- matrix( ( ( point[1,1] + point[1,4] + point[1,5])/3,  
 ( point[2,1] + point[2,4] + point[2,5])/3),  
 nrow =2, byrow = FALSE)  
K3 <- matrix( ( point[1,3], point[2,3]), nrow =2, byrow = FALSE)  
K4 <- matrix( ( point[1,7] + point[1,8])/2,   
 ( point[2,7] + point[2,8])/2),  
 nrow =2, byrow = FALSE)