# AC53013 Kmeans Investigation.

During the intensive week mark presented the following R Code:  
  
iris

newiris <- iris

newiris$Species <- NULL

newiris

kc <- kmeans(newiris, 3)

kc

table(iris$Species, kc$cluster)

plot(newiris[c("Sepal.Length", "Sepal.Width")], col=kc$cluster)

points(kc$centers[,c("Sepal.Length", "Sepal.Width")], col=1:3, pch=8, cex=2)

Which produced the following plot:  
  


Mark pointed out that each time this is run you will get slightly different outputs. Andy suggested that you could run this multiple times and then aggregate the runs to create a plot of how many times each point was in each cluster, this could be used to create a 3D plot that shows some points are frequently in one cluster and some points aren’t.

However, examining the data and the plots it was discovered that there is a difficulty in this simple idea that means it may not work. However, it’s clear that there is some information to be gained from this idea, or something similar.

In this assignment, carefully examine the plots from multiple runs and come to a conclusion on how KMeans is working on this data. Use your ideas to create a visualization that shows this behavior and shows how the points are distributed through the clusters. You should extend the R code provide to create this visualization.

Your report should be no more that 2-4 pages and should include the visualization or visualizations. You should include the code as an appendix.

You will need to make sure that the assignment is properly referenced.