R steps for clustering demo in Data Mining class

- Install R (https://www.r-project.org) and Install RStudio (https://www.rstudio.com)
- Create R script for the analysis steps. Here is an example with steps to run clustering methods on data

```
# if R was not started from where the data file is located, navigate to the data file directory
setwd("~/Dropbox/teaching/DataMining/CenCode/R")
# load the data file
irisData<-read.table("iris.txt")
head(irisData)
plot(irisData)

# if we dont know where the file is, use this option to find and load the data file
anotherData <-read.table(file.choose(), header=TRUE, sep="\t")
head(anotherData)

res=kmeans(irisData, centers=3, iter.max = 10, nstart = 1, algorithm = c("Hartigan-Wong", "Lloyd", "Forgy",
"MacQueen"), trace=FALSE)
res=kmeans(irisData, centers=3, nstart=3, iter.max = 10, algorithm = "Lloyd")
res=kmeans(irisData, 3)
res                # "cluster" "centers" "totss" "withinss" "tot.withinss" "betweenss" "size" "iter" "ifault"
res$totss

plot(irisData[res$cluster==1,], col="red")
plot(irisData[res$cluster==2,], col="blue")
plot(irisData[res$cluster==3,], col="green")
plot(irisData, col=res$cluster)

resPam <- pam(irisData, 3)
resPam
resPam$medoids

resCla <- clara(irisData, 4)
resCla

resAVG <- agnes(irisData, diss=FALSE, metric="euclidean", stand=FALSE, method="average")
resAVG
pltree(resAVG)
cutree(resAVG, k=1:4)

resSL <- agnes(irisData, diss=FALSE, metric="euclidean", stand=FALSE, method="single")
resSL
pltree(resSL)
rect.hclust(resAVG, k=3, border="red")

install.packages("clValid")
library(clValid)

data <- read.table("sterol.txt", header=TRUE, sep="\t")   #read data
data <- t(data) # transpose matrix
internal <- clValid(data, 2:10, clMethods=c("hierarchical", "kmeans", "clara"), validation="internal")
plot(internal)

stability <- clValid(data, 2:10, clMethods=c("hierarchical", "kmeans", "clara"), validation="stability")
plot(stability)
```