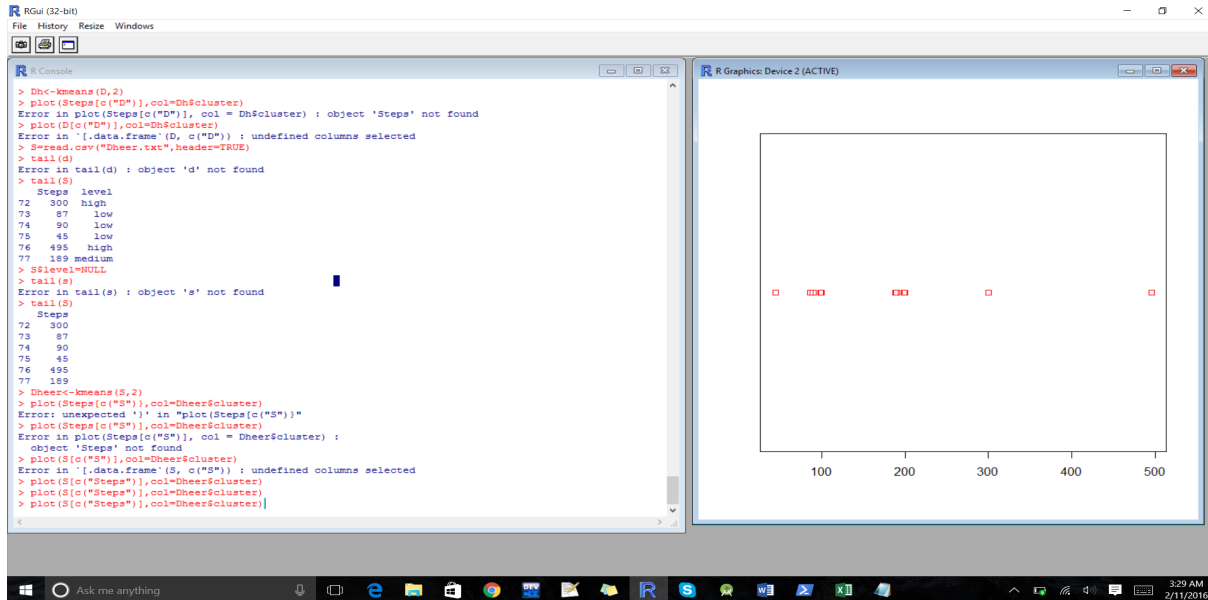# CS5542 Big Data Apps and Analytics
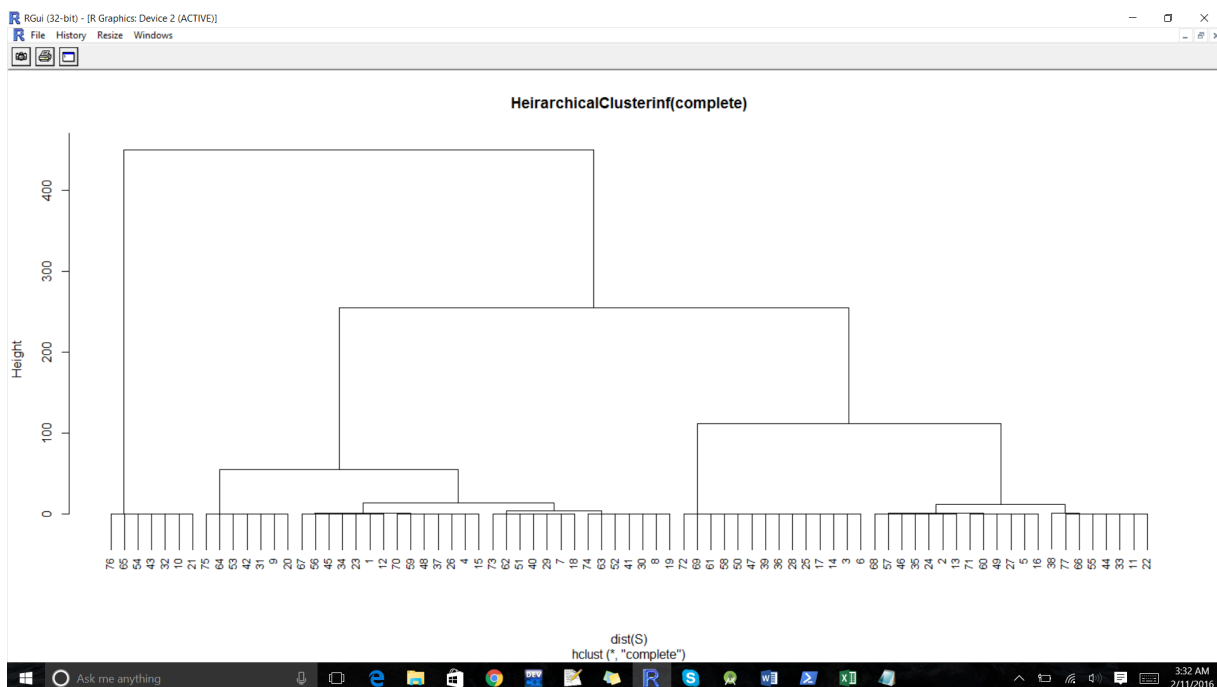
LAB ASSIGNMENT #3

1)R Project. Prepare a dataset related to your own project and perform k-Means, k-Medians, Expectation Maximisation (EM), Hierarchical Clustering.
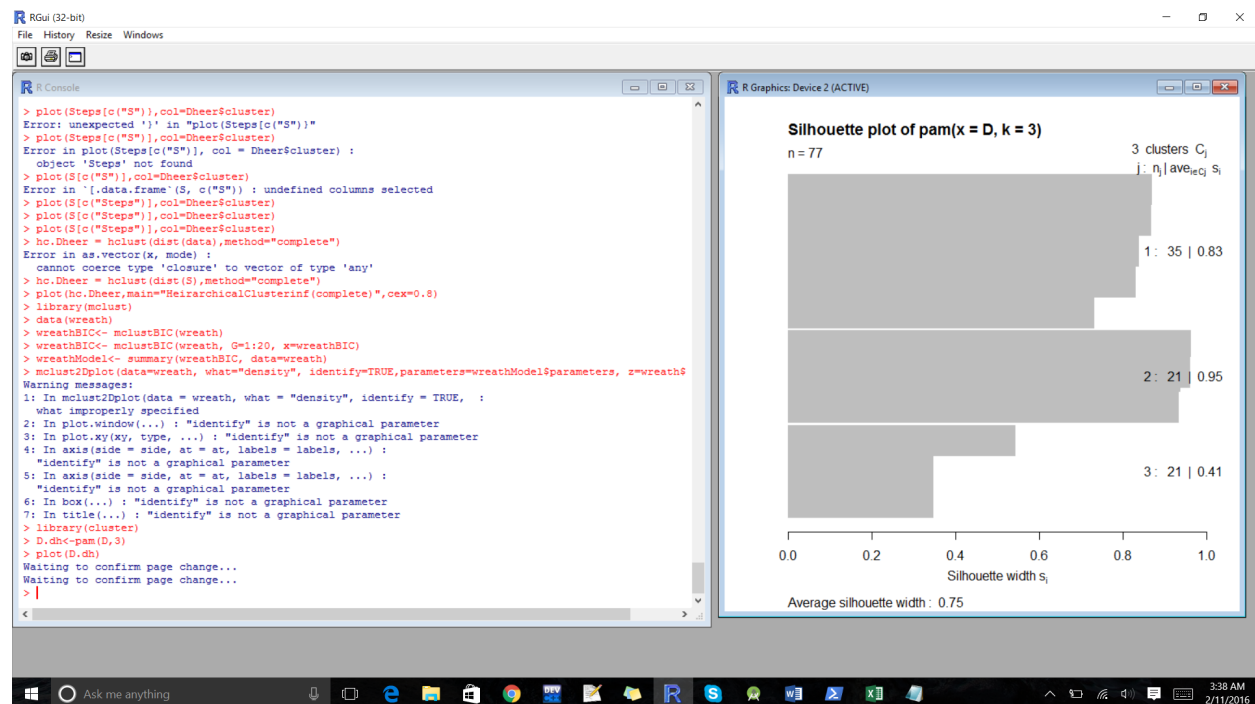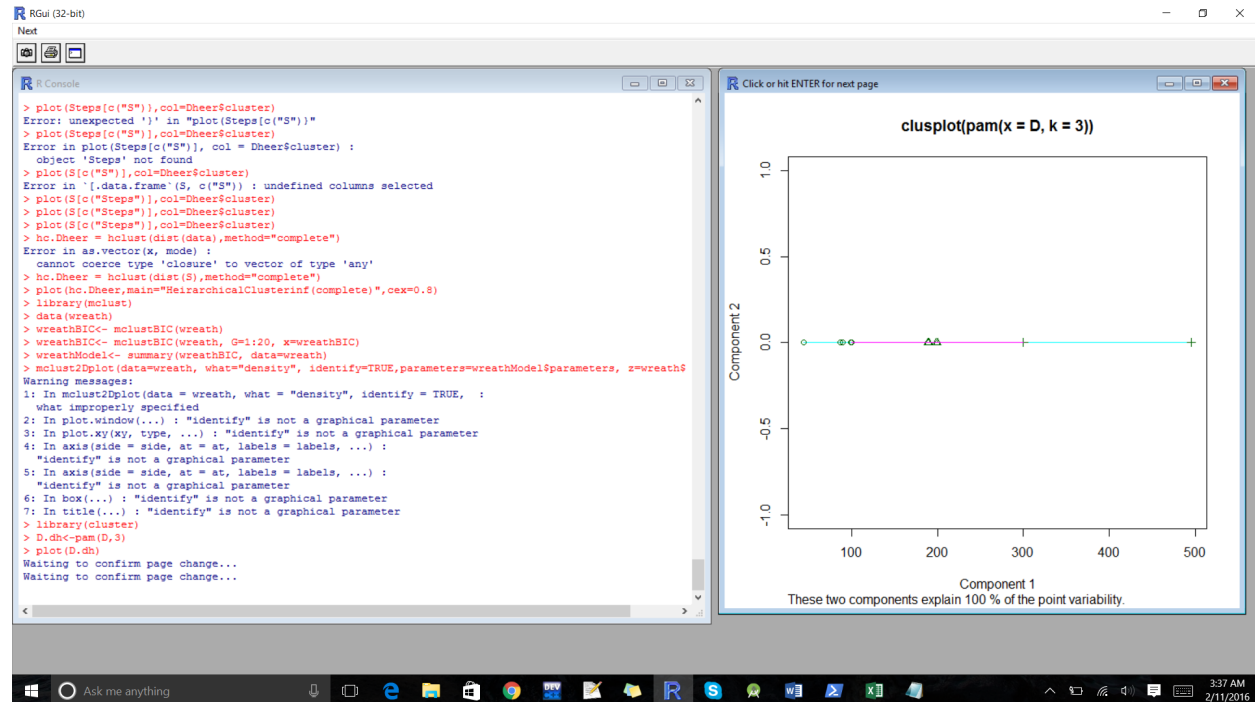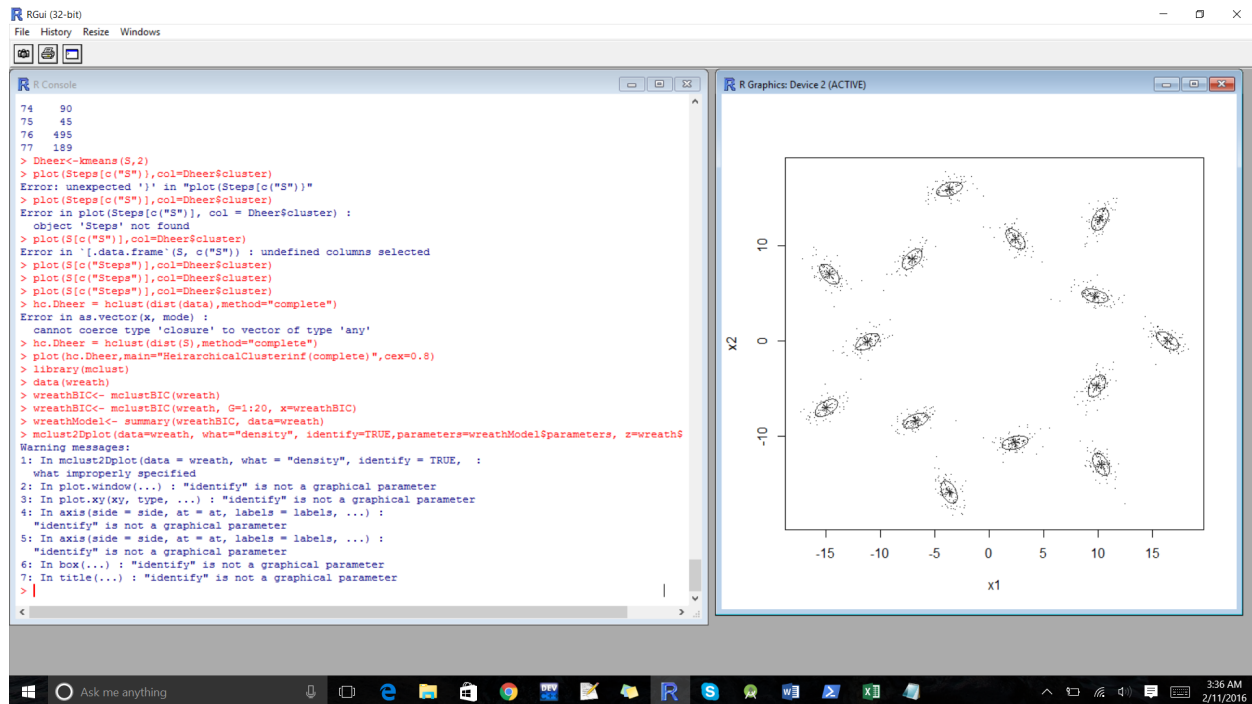
K Means

Hierarchical Clustering

K Medoids

```
> plot(Steps[c("S")),col=Dheer$cluster)
Error: unexpected ')' in "plot(Steps[c("S")}"
> plot(Steps[c("S")],col=Dheer$cluster)
Error in plot(Steps[c("S")], col = Dheer$cluster) :
  object 'Steps' not found
> plot(S[c("S")],col=Dheer$cluster)
Error in `[.data.frame`(S, c("S")) : undefined columns selected
> plot(S[c("Steps")],col=Dheer$cluster)
> plot(S[c("Steps")],col=Dheer$cluster)
> plot(S[c("Steps")],col=Dheer$cluster)
> hc.Dheer = hclust(dist(data),method="complete")
Error in as.vector(x, mode) :
  cannot coerce type 'closure' to vector of type 'any'
> hc.Dheer = hclust(dist(S),method="complete")
> plot(hc.Dheer,main="HeirarchicalClusterinf(complete)",cex=0.8)
> library(mclust)
> data(wreath)
> wreathBIC<- mclustBIC(wreath)
> wreathBIC<- mclustBIC(wreath, G=1:20, x=wreathBIC)
> wreathModel<- summary(wreathBIC, data=wreath)
> mclust2Dplot(data=wreath, what="density", identify=TRUE,parameters=wreathModel$parameters, z=wreath$
Warning messages:
1: In mclust2Dplot(data = wreath, what = "density", identify = TRUE,  :
  what improperly specified
2: In plot.window(...) : "identify" is not a graphical parameter
3: In plot.xy(xy, type, ...) : "identify" is not a graphical parameter
4: In axis(side = side, at = at, labels = labels, ...) :
  "identify" is not a graphical parameter
5: In axis(side = side, at = at, labels = labels, ...) :
  "identify" is not a graphical parameter
6: In box(...) : "identify" is not a graphical parameter
7: In title(...) : "identify" is not a graphical parameter
> library(cluster)
> D.dh<-pam(D,3)
> plot(D.dh)
Waiting to confirm page change...
Waiting to confirm page change...
```

clusplot(pam(x = D, k = 3))

Component 2 / Component 1

These two components explain 100 % of the point variability.

```
> plot(Steps[c("S")),col=Dheer$cluster)
Error: unexpected '}' in "plot(Steps[c("S")}"
> plot(Steps[c("S")],col=Dheer$cluster)
Error in plot(Steps[c("S")], col = Dheer$cluster) :
  object 'Steps' not found
> plot(S[c("S")],col=Dheer$cluster)
Error in `[.data.frame`(S, c("S")) : undefined columns selected
> plot(S[c("Steps")],col=Dheer$cluster)
> plot(S[c("Steps")],col=Dheer$cluster)
> plot(S[c("Steps")],col=Dheer$cluster)
> hc.Dheer = hclust(dist(data),method="complete")
Error in as.vector(x, mode) :
  cannot coerce type 'closure' to vector of type 'any'
> hc.Dheer = hclust(dist(S),method="complete")
> plot(hc.Dheer,main="HeirarchicalClusterinf(complete)",cex=0.8)
> library(mclust)
> data(wreath)
> wreathBIC<- mclustBIC(wreath)
> wreathBIC<- mclustBIC(wreath, G=1:20, x=wreathBIC)
> wreathModel<- summary(wreathBIC, data=wreath)
> mclust2Dplot(data=wreath, what="density", identify=TRUE,parameters=wreathModel$parameters, z=wreath$
Warning messages:
1: In mclust2Dplot(data = wreath, what = "density", identify = TRUE,  :
  what improperly specified
2: In plot.window(...) : "identify" is not a graphical parameter
3: In plot.xy(xy, type, ...) : "identify" is not a graphical parameter
4: In axis(side = side, at = at, labels = labels, ...) :
  "identify" is not a graphical parameter
5: In axis(side = side, at = at, labels = labels, ...) :
  "identify" is not a graphical parameter
6: In box(...) : "identify" is not a graphical parameter
7: In title(...) : "identify" is not a graphical parameter
> library(cluster)
> D.dh<-pam(D,3)
> plot(D.dh)
Waiting to confirm page change...
Waiting to confirm page change...
> |
```

Silhouette plot of pam(x = D, k = 3)

n = 77

3 clusters C_j

$j : n_j | ave_{i \in C_j} s_i$

1 : 35 | 0.83

2 : 21 | 0.95

3 : 21 | 0.41

Silhouette width s_i

Average silhouette width : 0.75

Expectation Maximisation (EM)

2)Watch Application: