library(ISLR)

# use NCI60 data set

nci.labs=NCI60$labs

nci.data=NCI60$data

sd.data=scale(nci.data)

# perform hierarchical clustering and plot them

hc.complete1=hclust(dist(sd.data), method="complete")

hc.average1=hclust(dist(sd.data), method="average")

hc.single1=hclust(dist(sd.data), method="single")

plot(hc.complete1,main="Complete Linkage", cex=.9)

plot(hc.average1, main="Average Linkage", cex=.9)

plot(hc.single1, main="Single Linkage", cex=.9)

# cutting dendrogram to yield 5 clusters

hc.cut1 = cutree(hc.complete1,5)

table(nci.labs, hc.cut1)

# kmeans VS hierarchical clustering

set.seed(3)

km.out1=kmeans(sd.data,5,nstart=20)

km.out1

table(km.out1$cluster,hc.cut1)

# correlated distance VS euclidean distance

dd1=as.dist(1-cor(t(sd.data)))

hc\_cor = hclust(dd1, method="complete")

hc.cut\_cor = cutree(hc\_cor,5)

# complete with correlated distance VS complete with euclidean distance

table(hc.cut\_cor, hc.cut1)

# kmeans VS complete linkage with correlated distance

table(hc.cut\_cor, km.out1$cluster)

Nstart = 1

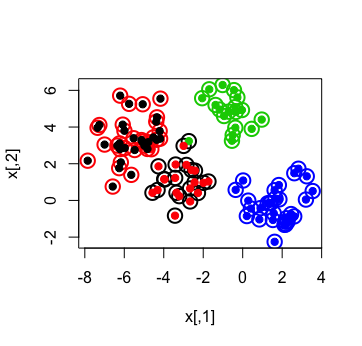
Within cluster sum of squares by cluster:

[1] 28.947802 21.049518 3.889402 244.503221

(between\_SS / total\_SS = 79.2 %)

Nstart = 20

Within cluster sum of squares by cluster:

[1] 30.82790 71.98228 21.04952 54.48008

(between\_SS / total\_SS = 87.6 %)

