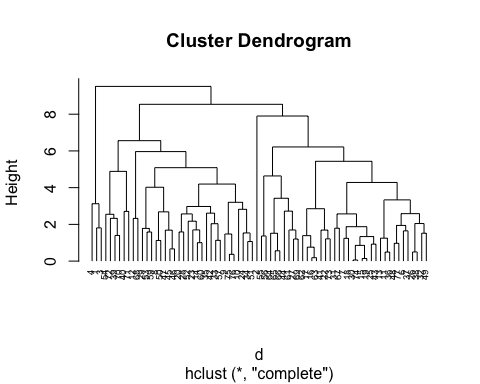
library(cluster)  
library(stats)  
  
  
#Task 1  
Cereals<- read.csv(file="~/Desktop/Cereals.csv", header = TRUE )  
df<-Cereals  
df1 <- df[,4:13]  
df1<- na.omit(df1)  
head(df1)

## calories protein fat sodium fiber carbo sugars potass vitamins shelf  
## 1 70 4 1 130 10.0 5.0 6 280 25 3  
## 2 120 3 5 15 2.0 8.0 8 135 0 3  
## 3 70 4 1 260 9.0 7.0 5 320 25 3  
## 4 50 4 0 140 14.0 8.0 0 330 25 3  
## 6 110 2 2 180 1.5 10.5 10 70 25 1  
## 7 110 2 0 125 1.0 11.0 14 30 25 2

df1 <- scale(df1)  
head(df)

## name mfr type calories protein fat sodium fiber carbo  
## 1 100%\_Bran N C 70 4 1 130 10.0 5.0  
## 2 100%\_Natural\_Bran Q C 120 3 5 15 2.0 8.0  
## 3 All-Bran K C 70 4 1 260 9.0 7.0  
## 4 All-Bran\_with\_Extra\_Fiber K C 50 4 0 140 14.0 8.0  
## 5 Almond\_Delight R C 110 2 2 200 1.0 14.0  
## 6 Apple\_Cinnamon\_Cheerios G C 110 2 2 180 1.5 10.5  
## sugars potass vitamins shelf weight cups rating  
## 1 6 280 25 3 1 0.33 68.40297  
## 2 8 135 0 3 1 1.00 33.98368  
## 3 5 320 25 3 1 0.33 59.42551  
## 4 0 330 25 3 1 0.50 93.70491  
## 5 8 NA 25 3 1 0.75 34.38484  
## 6 10 70 25 1 1 0.75 29.50954

d <-dist(df1,method = "euclidean")  
hc1 <- hclust(d, method = "complete")  
plot(hc1, cex = 0.6, hang = -1)



hc\_single<- agnes(df1, method = "single")  
hc\_complete <- agnes(df1, method = "complete")  
hc\_average <- agnes(df1, method = "average")  
hc\_ward <- agnes(df1, method = "ward")  
  
print(hc\_single$ac)

## [1] 0.6396208

print(hc\_complete$ac)

## [1] 0.8495565

print(hc\_average$ac)

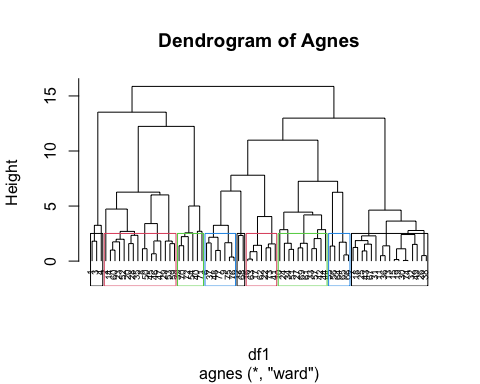
## [1] 0.7965659

print(hc\_ward$ac)

## [1] 0.9118102

#Ward would be the berst linkage method because it has the Agglomerative coefficient that is closest to 1.

pltree(hc\_ward, cex =0.6, hang = -1, main = "Dendrogram of Agnes")  
rect.hclust(hc\_ward, k=9, border =1:4)



#We used 9 clusters, because there were 9 variables to consider.

#Despite the 1st cluster of AllBran cereals being the “healthiest” overall. for an Elementary aged kid, I would chose the 4th cluster with Apple Cinnamon Cheerios, Honey Nut Cheerios, Multigrain Cheerios, Brans Check, Wheaties, Wheat Chex, and Wheaties Honey Gold. I would choose these because I think that they provide the most healthful benefit and are still popular with elementary age kids. ```

hc\_diana <-diana(df1)   
hc\_diana$dc

## [1] 0.8376705

pltree(hc\_diana, cex = 0.6, hang = -1, main = "Dendrogram of Diana")   
rect.hclust(hc\_diana, k =9, border = 1:4)

