Chiropractic Clusters

Byron Dennis

April 29, 2017

chiro\_data<-read.csv("Chiro\_Data.csv")  
  
#set row names and remove the name column  
rownames(chiro\_data) = (chiro\_data[,1])  
chiro\_data <- chiro\_data[,-1]  
  
#remove untits, cases, and service Count  
chiro\_data <- chiro\_data[,-3]  
  
chiro\_data <- chiro\_data[,-3]  
  
chiro\_data <- chiro\_data[,-3]  
  
  
summary(chiro\_data)

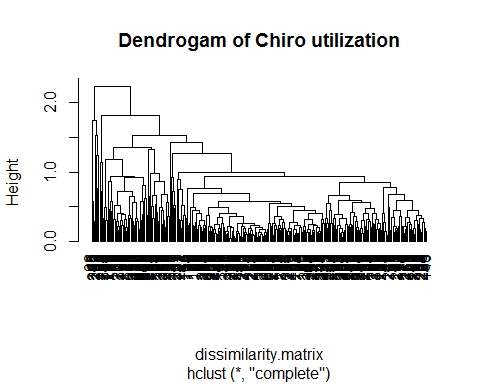
## Rnd.Prov.Cnt Mbr.Cnt Charged.per.Case Cases.per.Mbr   
## Min. : 1.000 Min. : 6.0 Min. : 43.38 Min. : 1.848   
## 1st Qu.: 1.000 1st Qu.: 47.0 1st Qu.: 82.41 1st Qu.: 4.409   
## Median : 1.000 Median : 81.0 Median :111.04 Median : 5.440   
## Mean : 1.373 Mean : 114.4 Mean :132.51 Mean : 6.619   
## 3rd Qu.: 1.000 3rd Qu.: 145.5 3rd Qu.:164.01 3rd Qu.: 7.543   
## Max. :11.000 Max. :1065.0 Max. :490.06 Max. :30.462   
## Units.per.Case Services.per.Case Acupuncture   
## Min. : 1.037 Min. :1.000 Min. :0.000000   
## 1st Qu.: 1.979 1st Qu.:1.740 1st Qu.:0.000000   
## Median : 2.741 Median :2.327 Median :0.000000   
## Mean : 3.113 Mean :2.503 Mean :0.006815   
## 3rd Qu.: 3.712 3rd Qu.:3.198 3rd Qu.:0.000000   
## Max. :17.090 Max. :6.050 Max. :0.280000   
## Chiropractic.Services Durable.Medical.Equipment  
## Min. :0.000 Min. :0.000000   
## 1st Qu.:1.080 1st Qu.:0.000000   
## Median :1.325 Median :0.000000   
## Mean :1.446 Mean :0.002877   
## 3rd Qu.:1.722 3rd Qu.:0.000000   
## Max. :3.910 Max. :0.090000   
## Physical.and.Speech.Therapy Outpatient.Surgery Other   
## Min. :0.0000 Min. :0.000000 Min. :0.00000   
## 1st Qu.:0.5275 1st Qu.:0.000000 1st Qu.:0.00000   
## Median :1.1600 Median :0.000000 Median :0.00000   
## Mean :1.3828 Mean :0.001575 Mean :0.01616   
## 3rd Qu.:1.9500 3rd Qu.:0.000000 3rd Qu.:0.00000   
## Max. :6.5700 Max. :0.270000 Max. :1.06000   
## Pathology.Lab X.Rays Office.Visits   
## Min. : 0.0000 Min. :0.00000 Min. :0.00000   
## 1st Qu.: 0.0000 1st Qu.:0.00000 1st Qu.:0.03000   
## Median : 0.0000 Median :0.00000 Median :0.06000   
## Mean : 0.1512 Mean :0.01873 Mean :0.08736   
## 3rd Qu.: 0.0000 3rd Qu.:0.03000 3rd Qu.:0.11250   
## Max. :12.1500 Max. :0.23000 Max. :0.52000

# Scale Data

normalize<- function(x){return((x-min(x))/(max(x)-min(x)))}  
chiro\_data\_n<-as.data.frame(lapply(chiro\_data[,1:15], normalize))  
summary(chiro\_data\_n)

## Rnd.Prov.Cnt Mbr.Cnt Charged.per.Case Cases.per.Mbr   
## Min. :0.00000 Min. :0.00000 Min. :0.00000 Min. :0.00000   
## 1st Qu.:0.00000 1st Qu.:0.03872 1st Qu.:0.08737 1st Qu.:0.08948   
## Median :0.00000 Median :0.07082 Median :0.15147 Median :0.12551   
## Mean :0.03733 Mean :0.10232 Mean :0.19955 Mean :0.16672   
## 3rd Qu.:0.00000 3rd Qu.:0.13173 3rd Qu.:0.27007 3rd Qu.:0.19902   
## Max. :1.00000 Max. :1.00000 Max. :1.00000 Max. :1.00000   
## Units.per.Case Services.per.Case Acupuncture   
## Min. :0.0000 Min. :0.0000 Min. :0.00000   
## 1st Qu.:0.0587 1st Qu.:0.1466 1st Qu.:0.00000   
## Median :0.1061 Median :0.2627 Median :0.00000   
## Mean :0.1293 Mean :0.2977 Mean :0.02434   
## 3rd Qu.:0.1666 3rd Qu.:0.4353 3rd Qu.:0.00000   
## Max. :1.0000 Max. :1.0000 Max. :1.00000   
## Chiropractic.Services Durable.Medical.Equipment  
## Min. :0.0000 Min. :0.00000   
## 1st Qu.:0.2762 1st Qu.:0.00000   
## Median :0.3389 Median :0.00000   
## Mean :0.3698 Mean :0.03196   
## 3rd Qu.:0.4405 3rd Qu.:0.00000   
## Max. :1.0000 Max. :1.00000   
## Physical.and.Speech.Therapy Outpatient.Surgery Other   
## Min. :0.00000 Min. :0.000000 Min. :0.00000   
## 1st Qu.:0.08029 1st Qu.:0.000000 1st Qu.:0.00000   
## Median :0.17656 Median :0.000000 Median :0.00000   
## Mean :0.21047 Mean :0.005835 Mean :0.01525   
## 3rd Qu.:0.29680 3rd Qu.:0.000000 3rd Qu.:0.00000   
## Max. :1.00000 Max. :1.000000 Max. :1.00000   
## Pathology.Lab X.Rays Office.Visits   
## Min. :0.00000 Min. :0.00000 Min. :0.00000   
## 1st Qu.:0.00000 1st Qu.:0.00000 1st Qu.:0.05769   
## Median :0.00000 Median :0.00000 Median :0.11538   
## Mean :0.01244 Mean :0.08145 Mean :0.16801   
## 3rd Qu.:0.00000 3rd Qu.:0.13043 3rd Qu.:0.21635   
## Max. :1.00000 Max. :1.00000 Max. :1.00000

dissimilarity.matrix <- dist(as.matrix(chiro\_data\_n, method="euclidean"))  
chiro\_hierarchical<- hclust(dissimilarity.matrix) #default agglomeration method is complete linkage.  
plot(chiro\_hierarchical, main = "Dendrogam of Chiro utilization", hang = -1) #three or four main clusters



# try three clusters

set.seed(225)  
chiro\_kmeans3 <- kmeans(chiro\_data\_n[,1:15], centers = 3)  
  
#add cluster assignment to original data  
chiro\_data\_n$Cluster3 <- chiro\_kmeans3$cluster  
  
#see how many chiro groups are assigned to each cluster  
chiro\_kmeans3$size

## [1] 35 66 191

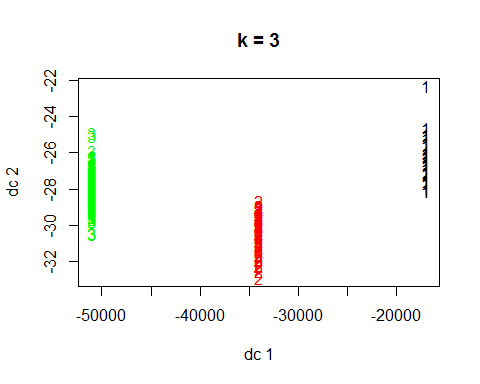
# View Characteristics  
t(round(chiro\_kmeans3$centers, digits = 2))

## 1 2 3  
## Rnd.Prov.Cnt 0.05 0.04 0.04  
## Mbr.Cnt 0.09 0.08 0.11  
## Charged.per.Case 0.32 0.39 0.11  
## Cases.per.Mbr 0.13 0.23 0.15  
## Units.per.Case 0.19 0.26 0.07  
## Services.per.Case 0.44 0.54 0.19  
## Acupuncture 0.07 0.02 0.02  
## Chiropractic.Services 0.42 0.47 0.33  
## Durable.Medical.Equipment 0.04 0.06 0.02  
## Physical.and.Speech.Therapy 0.31 0.41 0.12  
## Outpatient.Surgery 0.00 0.02 0.00  
## Other 0.04 0.00 0.01  
## Pathology.Lab 0.00 0.05 0.00  
## X.Rays 0.26 0.06 0.06  
## Office.Visits 0.42 0.14 0.13

#Visualize clusters  
library(fpc)

## Warning: package 'fpc' was built under R version 3.3.3

plotcluster(chiro\_data\_n, chiro\_data\_n$Cluster3, main="k = 3")



# try 4 clusters

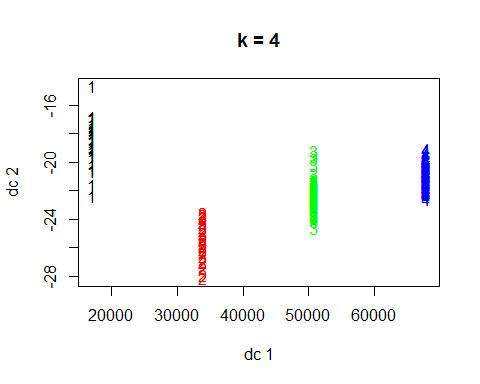
set.seed(225)  
chiro\_kmeans4 <- kmeans(chiro\_data\_n[,1:15], centers = 4)  
  
#add cluster assignment to original data  
chiro\_data\_n$Cluster4 <- chiro\_kmeans4$cluster  
  
#see how many chiro groups are assigned to each cluster  
chiro\_kmeans4$size

## [1] 32 25 95 140

# View Characteristics  
t(round(chiro\_kmeans4$centers, digits = 2))

## 1 2 3 4  
## Rnd.Prov.Cnt 0.05 0.01 0.04 0.04  
## Mbr.Cnt 0.10 0.06 0.09 0.12  
## Charged.per.Case 0.31 0.57 0.22 0.09  
## Cases.per.Mbr 0.13 0.21 0.22 0.13  
## Units.per.Case 0.19 0.39 0.15 0.06  
## Services.per.Case 0.45 0.64 0.39 0.13  
## Acupuncture 0.04 0.10 0.00 0.02  
## Chiropractic.Services 0.41 0.52 0.38 0.32  
## Durable.Medical.Equipment 0.03 0.14 0.03 0.01  
## Physical.and.Speech.Therapy 0.31 0.56 0.27 0.08  
## Outpatient.Surgery 0.00 0.02 0.01 0.00  
## Other 0.04 0.00 0.01 0.02  
## Pathology.Lab 0.00 0.12 0.01 0.00  
## X.Rays 0.26 0.10 0.05 0.06  
## Office.Visits 0.44 0.15 0.14 0.13

#Visualize clusters  
library(fpc)  
plotcluster(chiro\_data\_n, chiro\_data\_n$Cluster4, main="k = 4")



# check betweenness and withinness

#Separation  
clusters3\_betweenss<- chiro\_kmeans3$betweenss/chiro\_kmeans3$totss  
clusters4\_betweenss<- chiro\_kmeans4$betweenss/chiro\_kmeans4$totss  
  
  
betweenss.metric <- c(clusters3\_betweenss, clusters4\_betweenss)  
print(betweenss.metric) #Look for a ratio that is closer to 1.

## [1] 0.3324077 0.4039929

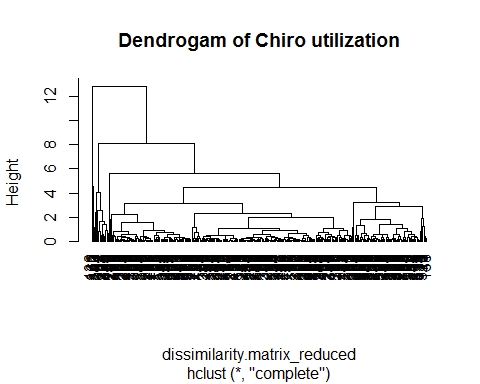
#cohesion  
clusters3\_withinss<- chiro\_kmeans3$tot.withinss/chiro\_kmeans3$totss  
clusters4\_withinss<- chiro\_kmeans4$tot.withinss/chiro\_kmeans4$totss  
  
totwithinss.metric <- c(clusters3\_withinss, clusters4\_withinss)  
print(totwithinss.metric) #Looking for a ratio that is closer to 0.

## [1] 0.6675923 0.5960071

# four clusters seems like the best option.

# try clusters on a subset of unscaled data to assist with interpretation

#create dataframe with which only inlucdes unit counts for different service types  
chiro\_data\_reduced<-chiro\_data[,7:15]  
  
  
dissimilarity.matrix\_reduced <- dist(as.matrix(chiro\_data\_reduced, method="euclidean"))  
chiro\_hierarchical\_reduced<- hclust(dissimilarity.matrix\_reduced) #default agglomeration method is complete linkage.  
plot(chiro\_hierarchical\_reduced, main = "Dendrogam of Chiro utilization", hang = -1) # Looks like 4-5 clusters now



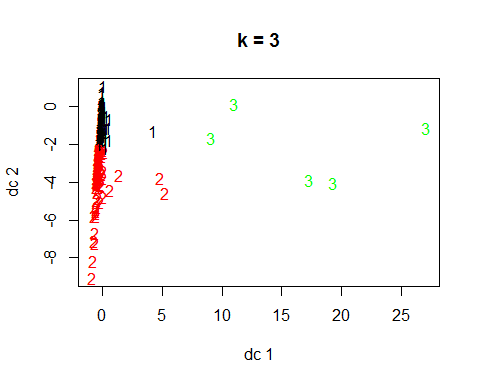
set.seed(225)  
chiro\_kmeans3\_red <- kmeans(chiro\_data\_reduced[1:9], centers = 3)  
  
#add cluster assignment to original data  
chiro\_data\_reduced$Cluster3\_red <- chiro\_kmeans3\_red$cluster  
  
#see how many chiro groups are assigned to each cluster  
chiro\_kmeans3\_red$size

## [1] 183 104 5

# View Characteristics  
clusters\_3\_red <- data.frame(t(round(chiro\_kmeans3\_red$centers, digits = 2)))  
clusters\_3\_red[order(-clusters\_3\_red$X3), ]

## X1 X2 X3  
## Pathology.Lab 0.01 0.05 7.47  
## Physical.and.Speech.Therapy 0.69 2.53 2.66  
## Chiropractic.Services 1.36 1.58 1.63  
## Office.Visits 0.08 0.10 0.06  
## Durable.Medical.Equipment 0.00 0.00 0.02  
## Acupuncture 0.01 0.01 0.00  
## Outpatient.Surgery 0.00 0.00 0.00  
## Other 0.02 0.02 0.00  
## X.Rays 0.02 0.02 0.00

#Visualize clusters  
library(fpc)  
plotcluster(chiro\_data\_reduced[1:9], chiro\_data\_reduced$Cluster3, main="k = 3")



# try clusters on a subset of unscaled data again but with 4 clusters

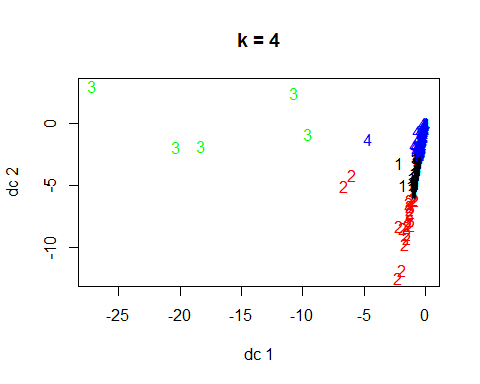
set.seed(225)  
chiro\_kmeans4\_red <- kmeans(chiro\_data\_reduced[1:9], centers = 4)  
  
#add cluster assignment to original data  
chiro\_data\_reduced$Cluster4\_red <- chiro\_kmeans4\_red$cluster  
  
#see how many chiro groups are assigned to each cluster  
chiro\_kmeans4\_red$size

## [1] 97 22 5 168

# View Characteristics  
clusters\_4\_red <- data.frame(t(round(chiro\_kmeans4\_red$centers, digits = 2)))  
clusters\_4\_red[order(-clusters\_4\_red$X3), ]

## X1 X2 X3 X4  
## Pathology.Lab 0.00 0.22 7.47 0.01  
## Physical.and.Speech.Therapy 2.03 4.03 2.66 0.62  
## Chiropractic.Services 1.54 1.76 1.63 1.35  
## Office.Visits 0.10 0.12 0.06 0.08  
## Durable.Medical.Equipment 0.00 0.00 0.02 0.00  
## Acupuncture 0.00 0.03 0.00 0.01  
## Outpatient.Surgery 0.00 0.01 0.00 0.00  
## Other 0.01 0.05 0.00 0.02  
## X.Rays 0.02 0.04 0.00 0.01

#Visualize clusters  
library(fpc)  
plotcluster(chiro\_data\_reduced[1:9], chiro\_data\_reduced$Cluster4, main="k = 4")



# Check Separation and Cohesion Again

#Separation  
clusters3\_betweenss\_red<- chiro\_kmeans3\_red$betweenss/chiro\_kmeans3\_red$totss  
clusters4\_betweenss\_red<- chiro\_kmeans4\_red$betweenss/chiro\_kmeans4\_red$totss  
  
  
betweenss.metric\_red <- c(clusters3\_betweenss\_red, clusters4\_betweenss\_red)  
print(betweenss.metric\_red) #Look for a ratio that is closer to 1.

## [1] 0.6476647 0.7364191

#cohesion  
clusters3\_withinss\_red<- chiro\_kmeans3\_red$tot.withinss/chiro\_kmeans3\_red$totss  
clusters4\_withinss\_red<- chiro\_kmeans4\_red$tot.withinss/chiro\_kmeans4\_red$totss  
  
totwithinss.metric\_red <- c(clusters3\_withinss\_red, clusters4\_withinss\_red)  
print(totwithinss.metric\_red) #Looking for a ratio that is closer to 0.

## [1] 0.3523353 0.2635809

# trying 5 clusters

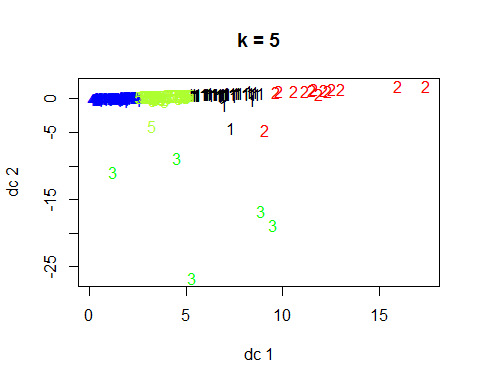
set.seed(225)  
chiro\_kmeans5\_red <- kmeans(chiro\_data\_reduced[1:9], centers = 5)  
  
#add cluster assignment to original data  
chiro\_data\_reduced$Cluster5\_red <- chiro\_kmeans5\_red$cluster  
  
#see how many chiro groups are assigned to each cluster  
chiro\_kmeans5\_red$size

## [1] 54 14 5 106 113

# View Characteristics  
clusters\_5\_red <- data.frame(t(round(chiro\_kmeans5\_red$centers, digits = 2)))  
clusters\_5\_red[order(-clusters\_5\_red$X3), ]

## X1 X2 X3 X4 X5  
## Pathology.Lab 0.04 0.18 7.47 0.00 0.02  
## Physical.and.Speech.Therapy 2.52 4.51 2.66 0.34 1.38  
## Chiropractic.Services 1.68 1.90 1.63 1.32 1.39  
## Office.Visits 0.10 0.11 0.06 0.07 0.10  
## Durable.Medical.Equipment 0.00 0.00 0.02 0.00 0.00  
## Acupuncture 0.00 0.05 0.00 0.01 0.01  
## Outpatient.Surgery 0.00 0.00 0.00 0.00 0.00  
## Other 0.00 0.08 0.00 0.02 0.01  
## X.Rays 0.03 0.04 0.00 0.01 0.02

#Visualize clusters  
library(fpc)  
plotcluster(chiro\_data\_reduced[1:9], chiro\_data\_reduced$Cluster5, main="k = 5")



# what about 6 clusters?

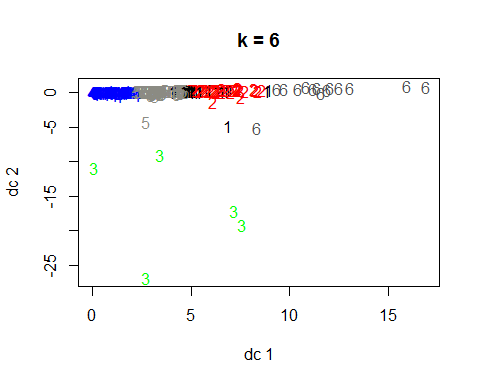
set.seed(225)  
chiro\_kmeans6\_red <- kmeans(chiro\_data\_reduced[1:9], centers = 6)  
  
#add cluster assignment to original data  
chiro\_data\_reduced$Cluster6\_red <- chiro\_kmeans6\_red$cluster  
  
#see how many chiro groups are assigned to each cluster  
chiro\_kmeans6\_red$size

## [1] 37 40 5 103 93 14

# View Characteristics  
clusters\_6\_red <- data.frame(t(round(chiro\_kmeans6\_red$centers, digits = 2)))  
clusters\_6\_red[order(-clusters\_6\_red$X3), ]

## X1 X2 X3 X4 X5 X6  
## Pathology.Lab 0.06 0.00 7.47 0.00 0.02 0.18  
## Physical.and.Speech.Therapy 2.15 2.46 2.66 0.32 1.25 4.51  
## Chiropractic.Services 0.98 1.94 1.63 1.32 1.48 1.90  
## Office.Visits 0.11 0.10 0.06 0.07 0.10 0.11  
## Durable.Medical.Equipment 0.00 0.01 0.02 0.00 0.00 0.00  
## Acupuncture 0.01 0.00 0.00 0.01 0.00 0.05  
## Outpatient.Surgery 0.01 0.00 0.00 0.00 0.00 0.00  
## Other 0.01 0.00 0.00 0.02 0.01 0.08  
## X.Rays 0.02 0.02 0.00 0.01 0.02 0.04

#Visualize clusters  
library(fpc)  
plotcluster(chiro\_data\_reduced[1:9], chiro\_data\_reduced$Cluster6, main="k = 6")



# Check Separation and Cohesion Final...going to exclude 6 clusters

#Separation  
clusters3\_betweenss\_red<- chiro\_kmeans3\_red$betweenss/chiro\_kmeans3\_red$totss  
clusters4\_betweenss\_red<- chiro\_kmeans4\_red$betweenss/chiro\_kmeans4\_red$totss  
clusters5\_betweenss\_red<- chiro\_kmeans5\_red$betweenss/chiro\_kmeans5\_red$totss  
  
betweenss.metric\_red <- c(clusters3\_betweenss\_red, clusters4\_betweenss\_red, clusters5\_betweenss\_red)  
print(betweenss.metric\_red) #Look for a ratio that is closer to 1.

## [1] 0.6476647 0.7364191 0.7786141

#cohesion  
clusters3\_withinss\_red<- chiro\_kmeans3\_red$tot.withinss/chiro\_kmeans3\_red$totss  
clusters4\_withinss\_red<- chiro\_kmeans4\_red$tot.withinss/chiro\_kmeans4\_red$totss  
clusters5\_withinss\_red<- chiro\_kmeans5\_red$tot.withinss/chiro\_kmeans5\_red$totss  
  
totwithinss.metric\_red <- c(clusters3\_withinss\_red, clusters4\_withinss\_red, clusters5\_withinss\_red)  
print(totwithinss.metric\_red) #Looking for a ratio that is closer to 0.

## [1] 0.3523353 0.2635809 0.2213859