f3

Auzzer\_Pang

6/11/2020

setwd("/Users/auzzer\_pang")  
data = read.csv("f3.csv")  
data = data[1:16,]

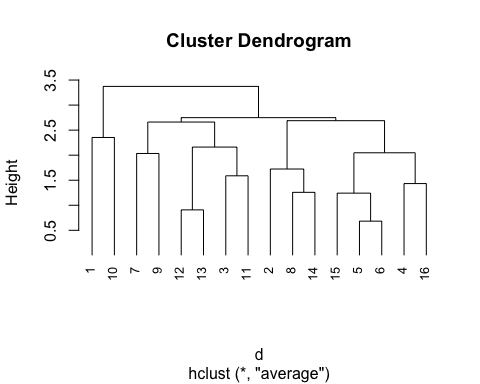
round(cor(data[,2:5]),3)

## x1 x2 x3 x4  
## x1 1.000 -0.314 0.216 0.216  
## x2 -0.314 1.000 -0.210 -0.054  
## x3 0.216 -0.210 1.000 -0.043  
## x4 0.216 -0.054 -0.043 1.000

#下面进行聚类分析  
data\_scaled = scale(data[,2:5])#标准化数据  
d = dist(data\_scaled)#求标准化欧式距离  
round(d,3)#保留三位小数

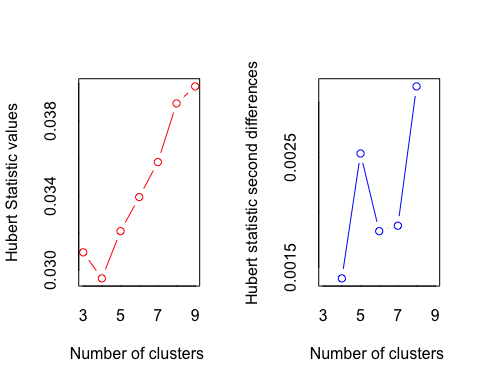
## 1 2 3 4 5 6 7 8 9 10 11 12  
## 2 3.177   
## 3 3.854 2.123   
## 4 3.757 2.485 3.388   
## 5 2.821 2.087 1.936 2.156   
## 6 2.810 1.517 1.801 1.835 0.683   
## 7 2.601 2.734 1.907 3.319 1.823 1.835   
## 8 2.691 1.985 2.877 2.414 2.532 1.975 2.236   
## 9 3.759 3.670 3.170 3.060 2.858 2.700 2.034 2.318   
## 10 2.353 3.668 3.872 2.720 2.140 2.428 2.634 3.081 2.945   
## 11 4.746 2.608 1.588 3.438 2.241 2.284 3.191 3.922 4.212 4.341   
## 12 4.832 3.116 1.812 3.493 2.912 2.653 2.499 3.006 2.439 4.396 2.696   
## 13 4.074 2.500 1.612 2.669 2.149 1.835 1.948 2.273 1.925 3.565 2.528 0.908  
## 14 3.587 1.463 2.497 2.663 2.805 2.141 2.745 1.258 3.008 4.022 3.414 2.623  
## 15 3.679 2.668 2.219 2.505 0.994 1.490 2.604 3.443 3.518 2.690 1.822 3.195  
## 16 3.683 3.265 3.521 1.433 1.845 2.006 3.142 3.159 3.017 1.877 3.463 3.762  
## 13 14 15  
## 2   
## 3   
## 4   
## 5   
## 6   
## 7   
## 8   
## 9   
## 10   
## 11   
## 12   
## 13   
## 14 2.102   
## 15 2.586 3.526   
## 16 2.955 3.653 1.941

#采用类平均法  
hc <- hclust(d,"average")  
plot(hc, hang=-1, cex=.8)

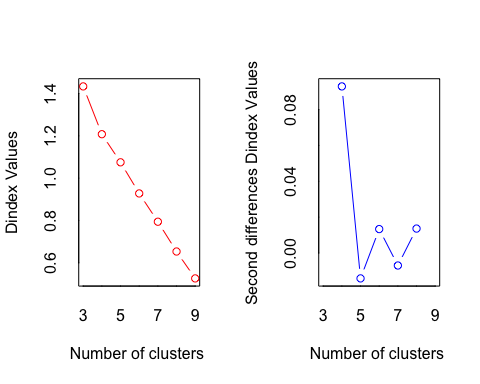


library(NbClust)  
#根据题目条件，我们认定分类个数在2到10之间为无法取到2和10  
#采用标准化欧式距离以类平均法进行聚类分析  
nc <- NbClust(data\_scaled, distance="euclidean",  
 min.nc=3, max.nc=9, method="average")

## Warning in pf(beale, pp, df2): 产生了NaNs  
  
## Warning in pf(beale, pp, df2): 产生了NaNs



## \*\*\* : The Hubert index is a graphical method of determining the number of clusters.  
## In the plot of Hubert index, we seek a significant knee that corresponds to a   
## significant increase of the value of the measure i.e the significant peak in Hubert  
## index second differences plot.   
##

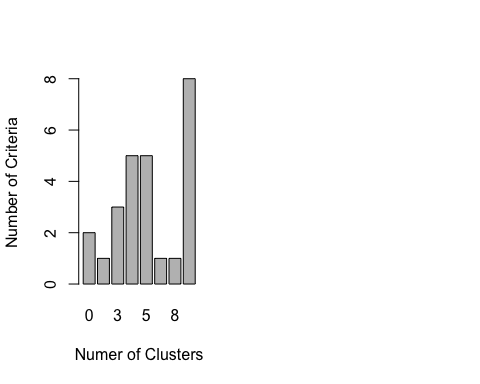


## \*\*\* : The D index is a graphical method of determining the number of clusters.   
## In the plot of D index, we seek a significant knee (the significant peak in Dindex  
## second differences plot) that corresponds to a significant increase of the value of  
## the measure.   
##   
## \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*   
## \* Among all indices:   
## \* 3 proposed 3 as the best number of clusters   
## \* 5 proposed 4 as the best number of clusters   
## \* 5 proposed 5 as the best number of clusters   
## \* 1 proposed 6 as the best number of clusters   
## \* 1 proposed 8 as the best number of clusters   
## \* 8 proposed 9 as the best number of clusters   
##   
## \*\*\*\*\* Conclusion \*\*\*\*\*   
##   
## \* According to the majority rule, the best number of clusters is 9   
##   
##   
## \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

table(nc$Best.n[1,])#找到最佳的聚类个数为9

##   
## 0 2 3 4 5 6 8 9   
## 2 1 3 5 5 1 1 8

barplot(table(nc$Best.n[1,]),  
xlab="Numer of Clusters", ylab="Number of Criteria")  
#将聚类结果以条形图画出来



#最终以图表的形式展示聚类结果  
plot(hc, hang=-1, cex=.8,)  
rect.hclust(hc, k=9)

