首先加载数据集并且进行一些数据处理，把数值型变量转换成因子型

house\_price<-read.csv(file.choose())

#概览数据集，并且进行变量类型转换

summary(house\_price)

class(house\_price[,7])

house\_price[,7]<-as.factor(house\_price[,7])

class(house\_price[,9])

house\_price[,9]<-as.factor(house\_price[,9])

class(house\_price[,10])

house\_price[,10]<-as.factor(house\_price[,10])

class(house\_price[,12])

house\_price[,12]<-as.factor(house\_price[,12])

class(house\_price[,13])

house\_price[,13]<-as.factor(house\_price[,13])

class(house\_price[,14])

house\_price[,14]<-as.factor(house\_price[,14])

house\_price2<-house\_price[,-17]

summary(house\_price2)

house\_price3<-house\_price2[,1:14]

#设置随机种子

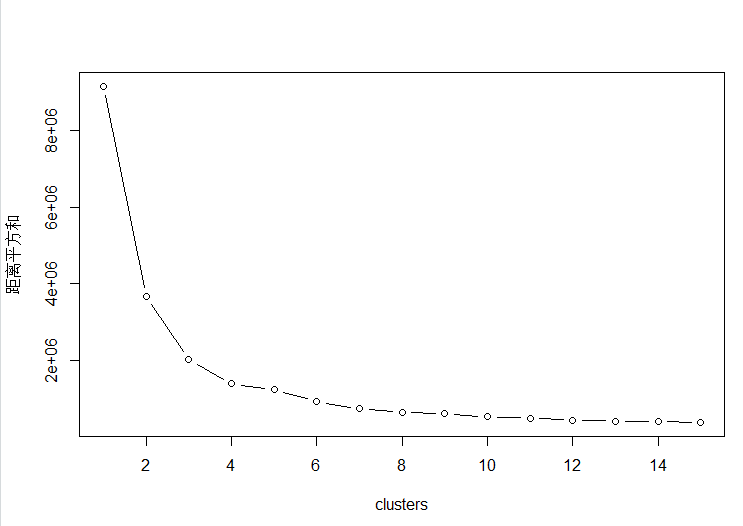
set.seed(5032)

#绘制拐点图

a<-(nrow(house\_price3)-1)\*sum(apply(house\_price3,2,var))

for(i in 2:15) a[i]<-sum(kmeans(house\_price3,centers=i)$withinss)

plot(1:15,a,type="b",xlab="clusters",ylab="距离平方和")



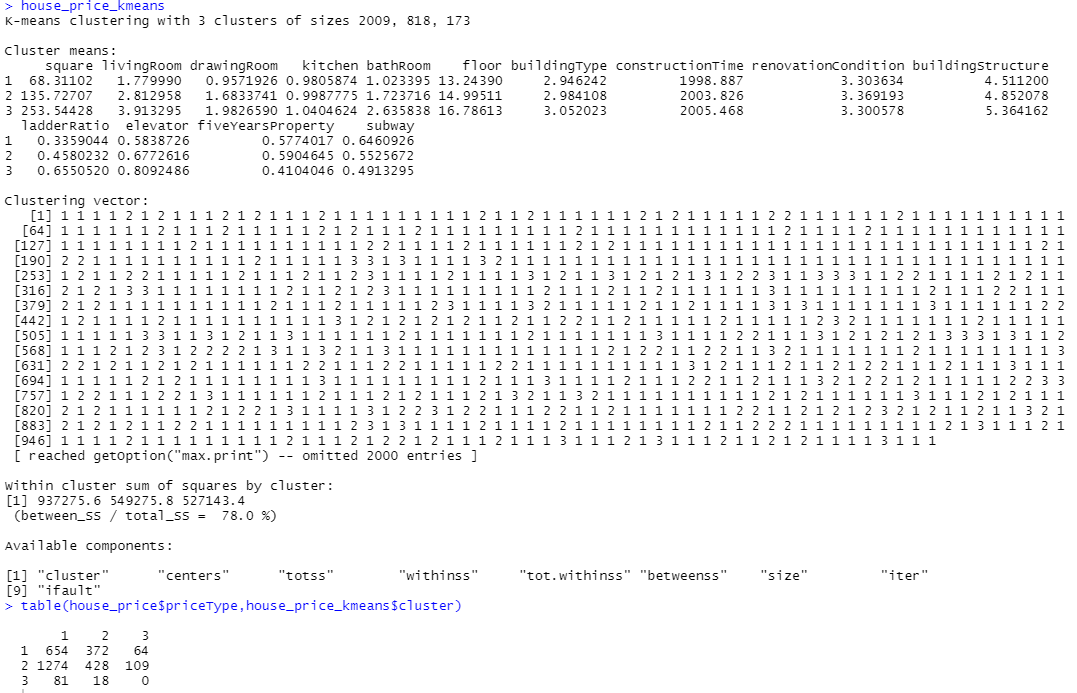
由拐点图可知，聚类时簇应选择为3，此时效果最好

#用k-means方法聚类

house\_price\_kmeans<-kmeans(house\_price3,3)

house\_price\_kmeans

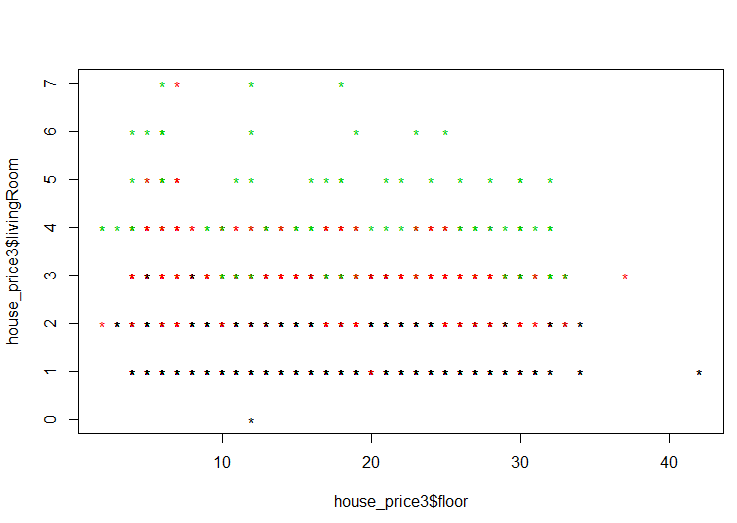
table(house\_price$priceType,house\_price\_kmeans$cluster)



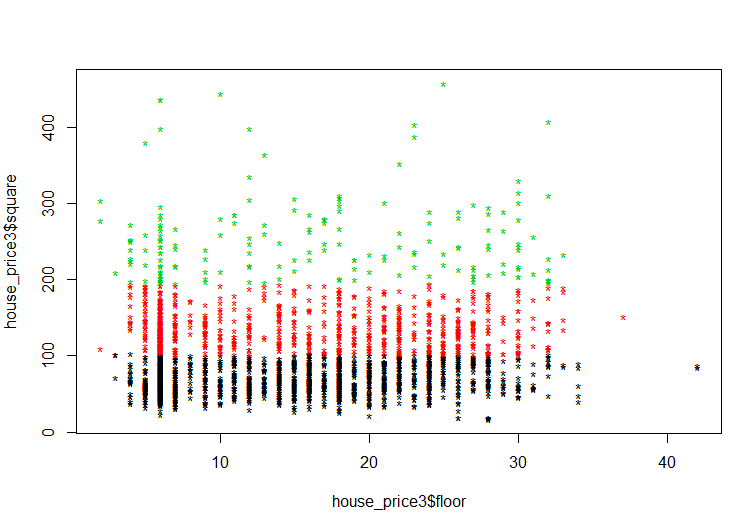
#下边我们将分类以及中心点打印出来

plot(house\_price3$floor,house\_price3$livingRoom,col=house\_price\_kmeans$cluster,pch="\*")

points(house\_price\_kmeans$centers,pch="X",cex=1.5,col=4)



plot(house\_price3$floor,house\_price3$square,col=house\_price\_kmeans$cluster,pch="\*")

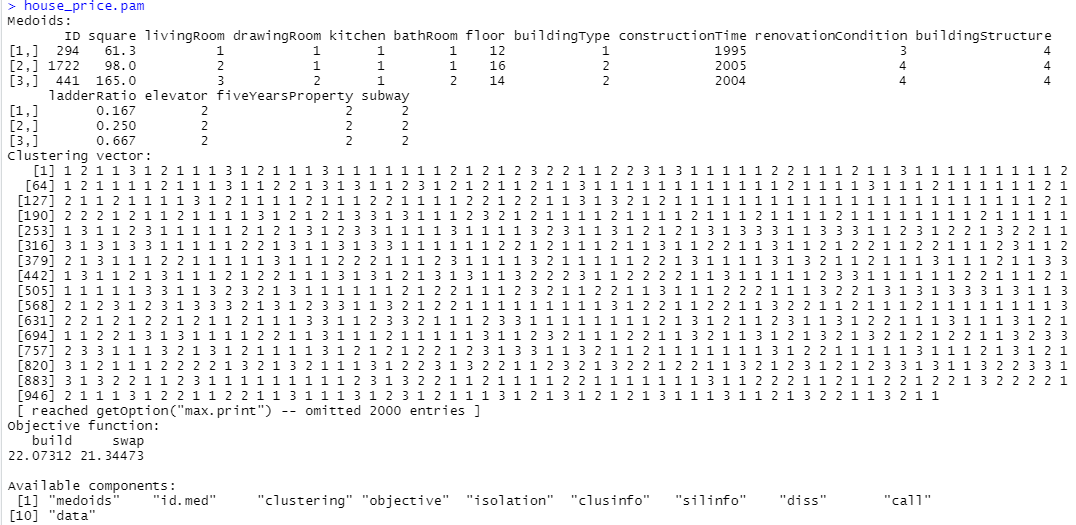


library(cluster)

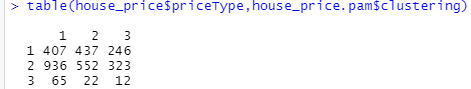
set.seed(6598)

house\_price.pam<-pam(house\_price3,3)

house\_price.pam



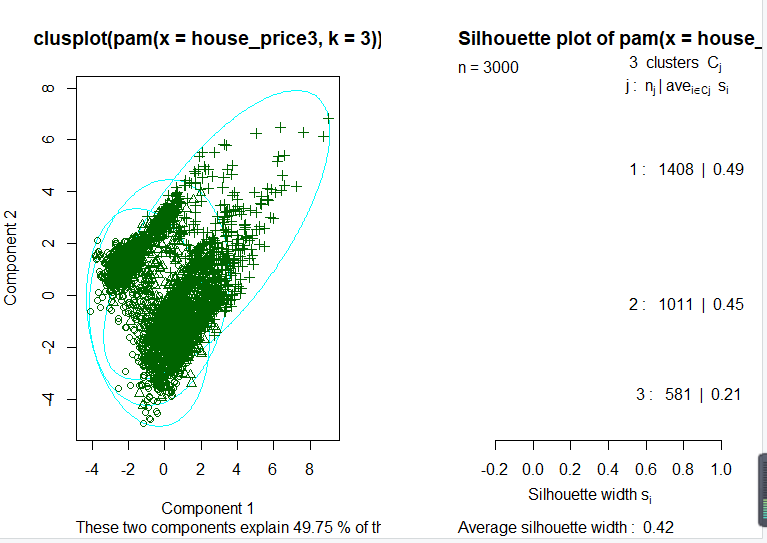
table(house\_price$priceType,house\_price.pam$clustering)



layout(matrix(c(1,2),1,2)) #每页显示两个图

plot(house\_price.pam)

layout(matrix(1))



house\_price.clara<-clara(house\_price3,3)

house\_price.clara

table(house\_price$priceType,house\_price.clara$clustering)

layout(matrix(c(1,2),1,2)) #每页显示两个图

plot(house\_price.clara)

