lab3q6

March 20, 2017

In [9]:
```
## question 6
library(dplyr)
data = read.csv("datassss.csv")
datas <- data

datas <- datas[, colMeans(is.na(datas)) <= 0.65]
conames <- c("int_date","age","zipcode","weight","standwt","ï..psraid")

datas[,!colnames(datas) %in% conames] = lapply(datas[,!colnames(datas) %in%

s <- colnames(datas)[colSums(is.na(datas)) > 0]

#Mode of the data
Mode <- function(x, na.rm = TRUE) {
  if(na.rm){
    x = x[!is.na(x)]
  }
  ux <- unique(x)
  ux[which.max(tabulate(match(x, ux)))]
}
md <- apply(datas[,s], 2, Mode)
for(var in s)
{
  datas[is.na(datas[,var]),var] <- Mode(datas[,var],na.rm = T)
}


newdatas <- select(datas,sample:q1,game1:game4,age,race,marital,par,educ2,i
newdatas <- select(newdatas, -(sample:form))
newdatas <- select(newdatas, -q1,-marital)

###K-means

kmeans_model.3 = kmeans(newdatas, centers = 3)

kmeans_model.5 = kmeans(newdatas, centers = 5)
```