**CLUSTER ANALYSIS**

#factor variables

col\_names <- names(pref)

pref[,col\_names] <- lapply(pref[,col\_names] , factor)

str(pref)

library(cluster)

dist<-daisy(pref, metric="gower") #distances

dist1<-as.matrix(dist)

res.hc<- hclust(d=dist, method=“single”) #cluster

#or

res.hc<- hclust(d=dist, method=“complete”)

plot(res.hc, hang=-1)

plot(cut(as.dendrogram(res.hc), h=0.5)$lower[[1]])

plot(cut(as.dendrogram(res.hc), h=0.5)$lower[[2]])

plot(cut(as.dendrogram(res.hc), h=0.5)$lower[[3]])

#circular dendrogram

library(dendextend)

library(circlize)

hc <- as.dendrogram(hclust(dist))

circlize\_dendrogram(hc, labels\_track\_height = NA,dend\_track\_height = 0.5)

rect.hclust(res.hc, k=4, border="red")

seg.hc.segment <- cutree(res.hc, k=4)

table(seg.hc.segment)

seg.summ <- function(data, groups) {aggregate(data, list(groups), function(x)mean(as.numeric(x))) }

seg.summ(pref, seg.hc.segment)

#

#creating a subset

data1<-pref[,c("v1","v2","v3","v4","v8\_Fonzies","v8\_Puff","v9\_ape","v9\_trav","v11\_price", "v\_choice","v11\_brand","v13","v15")]

…