MVA Homework 4 MCA and Clustering - Diego Garcia-Olano

1 Loading and preprocessing data

```
#1. Read the files ENQSCP.DAT and ENQSCP.DIC,
    the former containing the data and the second containing the dictionary.

dat <- read.table("assignment4/ENQSCP.DAT",header=FALSE)
dim(dat)    #316 x 53
summary(dat)
#We see missing data and realize a extra blank row at end exists, so we remove it.
dat <- dat[-316,]

#We manually transform .DIC file to a csv file that is easier to work with.
dic <- read.csv("assignment4/enqdic.csv",header=TRUE, stringsAsFactors=TRUE)

#We want to put everything into one dataframe so we get the columns from the dictionary
#and use them as the column names for the datafile.
tmpcolnames <- substring(names(dic),4)
tmpcolnames[c(10,18,41:45,47)] <- substring(tmpcolnames[c(10,18,41:45,47)],4)
tmpcolnames <- append(tmpcolnames,"id",0)
colnames(dat) <- tmpcolnames
summary(dat)

#Put back in column names to data matrix
for(i in 2:53){
  exclude.numeric.columns <- c(5,27,29,42,43,44,45,46,47,48,51,52)
  has.zero.answer <- c(15,20,21,28,32)     #these columns contains "0" answer so prepend

  if( !(i %in% exclude.numeric.columns)){
    dat[,i] <- as.factor(dat[,i]);

    if(i %in% has.zero.answer){
        lvls <- levels(dic[,i-1])[levels(dic[,i-1]) != ""]
        levels(dat[,i]) <- append(lvls,"not answered",0) }
    else{
      levels(dat[,i]) <- levels(dic[,i-1])[levels(dic[,i-1]) != ""] }
  }
}

#Make columns 42 through 48 factors as each only has 7 or 8 distinct values
for(i in 42:48){ dat[,i] <- as.factor(dat[,i]);}

#Same with Number of persons in a housing, and Number of children
dat$Number.of.persons.in.a.housing <- as.factor(dat$Number.of.persons.in.a.housing)
dat$Number.of.children <- as.factor(dat$Number.of.children)

#Age variable has 64 different levels so subdivide into more meaningful ones!
length(unique(dat$Age.of.respondent)) #64
table(dat$Age.of.respondent)
cumsum(table(dat$Age.of.respondent))

#There are 315 total observations, and we divided into 5 groups of similar sizes
agelvls = c("28 and under", "29 to 35", "36 to 49", "50 to 62", "63 and up")
```

```
dat$tmp.age <- as.character(dat$Age.of.respondent)
for(i in 1:315){
  age <- dat[i,5]
  #store results in tmp column
  if(age <= 28){ dat[i,54] <- agelvls[1];}
  if(age > 28 && age <= 35){ dat[i,54] <- agelvls[2];}
  if(age > 35 && age <= 49){ dat[i,54] <- agelvls[3];}
  if(age > 49 && age <= 62){ dat[i,54] <- agelvls[4];}
  if(age > 62){ dat[i,54] <- agelvls[5];} }
dat$Age.of.respondent = as.factor(dat$tmp.age)
dat <- dat[,-54]

#State.benefits variable has 74 levels so subdivide into more meaningful units
table(dat$State.benefits...average.monthly.amount);
cumsum(table(dat$State.benefits...average.monthly.amount)[-1])

#Most of the values (206 of 315) are 0 or 999999, so we label those ase "not answered"
#and split remaing 109 into 4 equal sized groups.
stblvls = c("not answered", "less than 600", "600 to 1000", "1000 to 2000", "2001 to 5100")
dat$tmp.stb <- as.character(dat$State.benefits...average.monthly.amount)
for(i in 1:315){
  stateb <- dat[i,51]
  if(stateb == 0 || stateb == 999999){ dat[i,54] <- stblvls[1];}
  if(stateb > 0 && stateb < 600){ dat[i,54] <- stblvls[2];}
  if(stateb >=  600 && stateb <= 1000){ dat[i,54] <- stblvls[3];}
  if(stateb > 1000 && stateb <= 2000){ dat[i,54] <- stblvls[4];}
  if(stateb > 2000 && stateb <= 5100){ dat[i,54] <- stblvls[5];}  }
dat$State.benefits...average.monthly.amount = as.factor(dat$tmp.stb)
dat <- dat[,-54]

#Salary.of.the.respondent has 54 levels so subdivide into more meaningful units
table(dat$Salary.of.the.respondent)
cumsum(table(dat$Salary.of.the.respondent)[-1])

#A big chunk of the values (110 of 315) again are 0 or 999999, so we label those ase "not answered"
#and split remaing 205 into 4 equal sized groups.
salblvls = c("not answered", "less than 3200", "3200 to 4900", "5000 to 8400", "8500 to 40000")
dat$tmp.sal <- as.character(dat$Salary.of.the.respondent)
for(i in 1:315){
  sal <- dat[i,52]
  if(sal == 0 || sal == 999999){ dat[i,54] <- salblvls[1];}
  if(sal > 0 && sal < 3200){ dat[i,54] <- salblvls[2];}
  if(sal >=  3200 && sal <= 4900){ dat[i,54] <- salblvls[3];}
  if(sal > 4900 && sal <= 8400){ dat[i,54] <- salblvls[4];}
  if(sal > 8400 && sal <= 40000){ dat[i,54] <- salblvls[5];} }
dat$Salary.of.the.respondent = as.factor(dat$tmp.sal)
dat <- dat[,-54]

#See appendix for how full dataset summary.
```

2   List of questions as active for analysis

```
#2. The idea is to group citizens according their opinions about everyday life. Hence specify the list
#of questions that you would consider as active for the analysis (the remaining will act as illustrative).
We consider the following subset of variables to represent opinions
survey.opinion.indexes = c(8:12,25,30:32,34:49,52)
names(dat.mca)[c(8:12,25,30:32,34:49,52)]
```

3 Perform the Multiple Correspondence Analysis and interpret the first obtained factors.

```
library(FactoMineR)
#first column is only for identification purposes
iden = dat[,1]
dat.mca = dat[,-1]
sop.dat = MCA(dat.mca, quali.sup=-survey.opinion.indexes, ncp=Inf )
#using opinions as active variables, and taking all components (instead of just 5)

plot(sop.dat$eig$eigenvalue,type="l")  #plot eigenvalues
abline(h=mean(sop.dat$eig$eigenvalue), lty=2, col="red")
sop.dat$eig[which(sop.dat$eig$eigenvalue > 0.01923077 ),] #77 out of 108 are significant factors
nd = 77
```
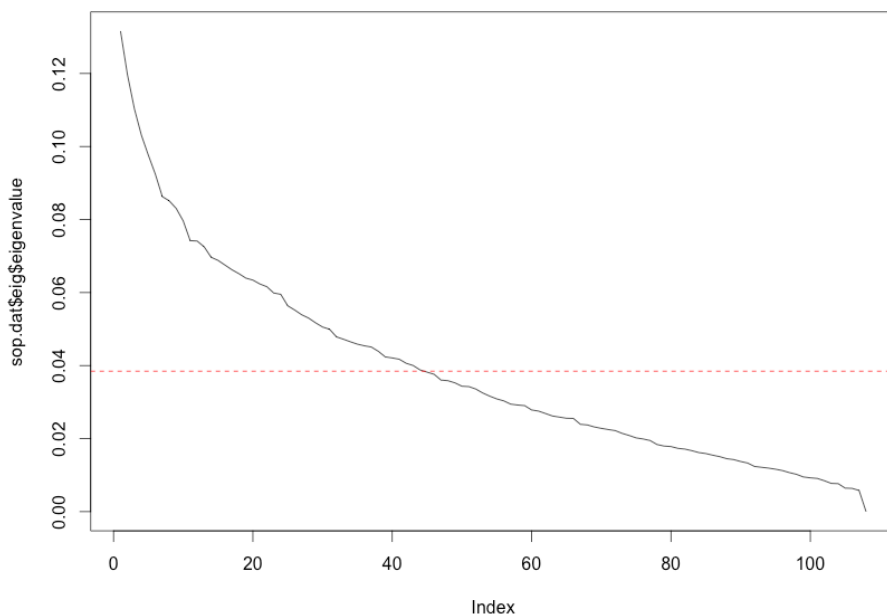


```
#Plotting the variable cloud we see:
plot(sop.dat,invisible=c("ind","quali.sup"),col.var='black',cex=.6,xlim=c(-4,4),ylim=c(-2,5))
```
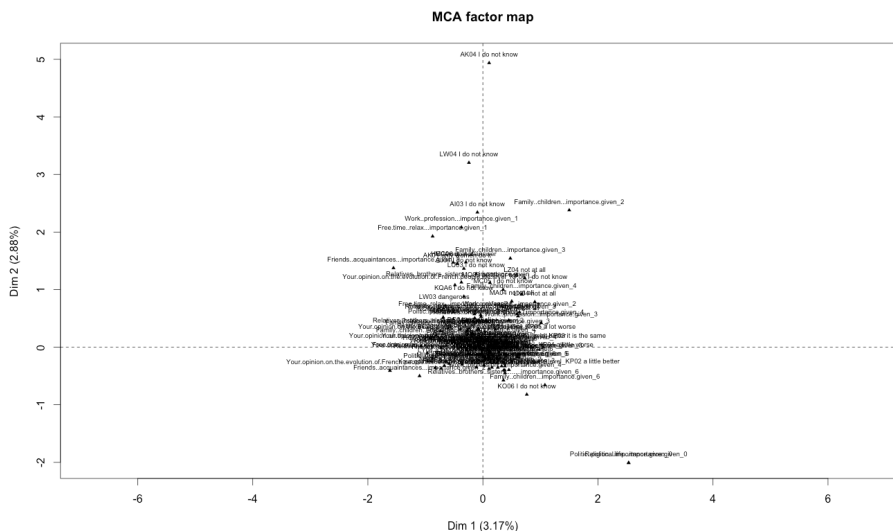
```
#Now we look at which variables which contribute the most to the forming of the 1st factors.
round(sort(sop.dat$var$contrib[,1] / sum(sop.dat$var$contrib[,1]),decreasing=TRUE),5)
round(sort(sop.dat$var$contrib[,2] / sum(sop.dat$var$contrib[,2]),decreasing=TRUE),5)

#Because of the large number of variables and modalities, we see that each of the first
#two factors contain a large number of modalities around 30 each, which contribute at
#least 1 percent and not more than 6.5 percent to the construction.  Just to have an idea
#here are the modalities which contribute at least 3 percent to factor 1:
LY01 a lot  ( Worried about the risk of a mugging )   0.06499
LZ01 a lot  ( Worried about the risk of a road accident )  0.05653
MB01 a lot  ( Worried about the risk of a nuclear plant accident) 0.05635
LX01 a lot  ( Worried about the risk of having a serious illness) 0.05610
MA01 a lot  ( Worried about the risk of unemployment) 0.05211
HG03 a little ( Satisfied with health ) 0.03620
AI02 -  no  - ( The family is the only place. where you feel well ) 0.03034

#It turns out that the group of people who are "worried alot" contribute a good deal
to the first factor, but still only around 30 percent of it, while the rest is split
between other modalities.

#For factor 2, those which contribute at least 3 percent are as follows:
LW04 I do not know ( Computer science diffusion is .. )   0.06290
Work..profession...importance.given_1     0.06190
LZ04 not at all  ( Worried about the risk of a road accident ) 0.05587
LU03 I do not know  ( Society needs change ) 0.04814
MA04 not at all  ( Worried about the risk of unemployment ) 0.04413
AK01 only women do it ( Housekeeping/care of children ) 0.03214

#The 2nd factor is a little less clear, but it seems to oppose people
who aren't worried, "don't know", give importance to work, and think
women should be in charge of housekeeping, though again these only
account for around 30 percent of the factor as well.
```

4   Perform a hierarchical clustering and consolitation

```
#  Perform a hierarchical clustering with the significant factors, decide the number
#  of final classes to obtain and perform a consolidation operation of the clustering.
Psi = sop.dat$ind$coord[,1:nd]
di <- dist(sop.dat$ind$coord, method = "euclidean")  #euclidean in
hc <- hclust(di, method = "ward")
barplot(hc$height)  #bar plot of heights
```
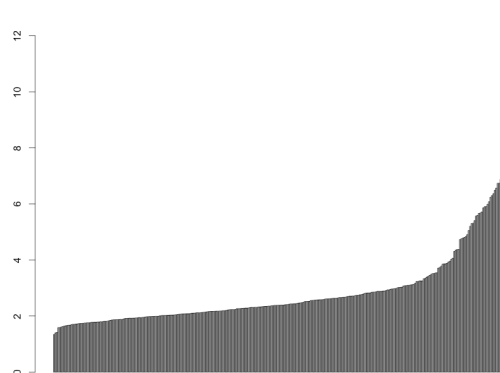

Figure 1: Dendogram created by HC

```
plot(hc,cex=.5)    #plot of dendogram
```

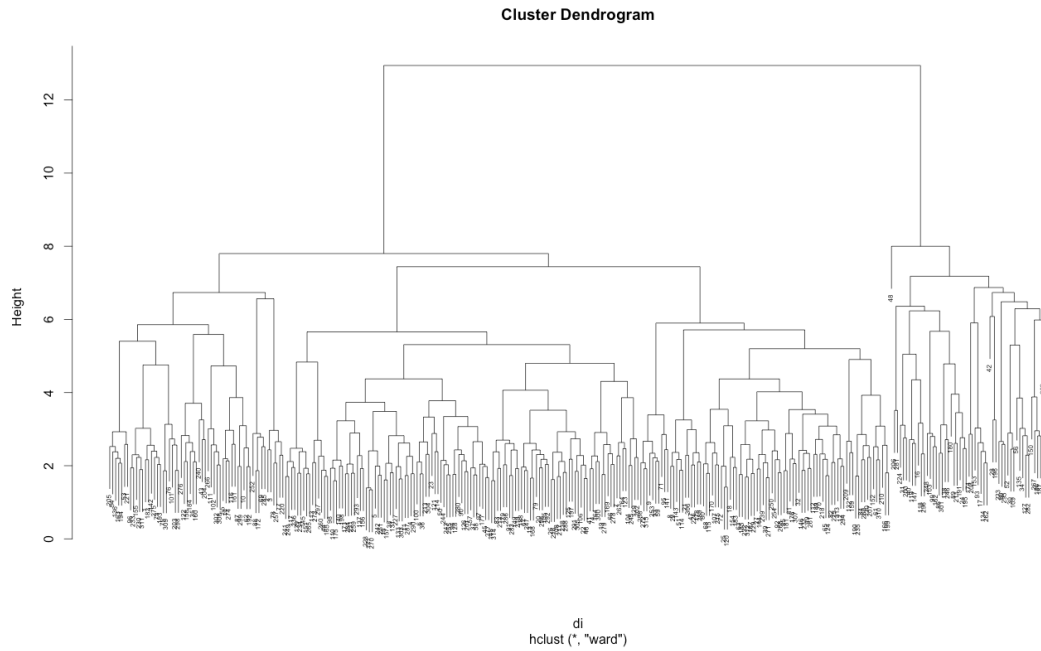**Cluster Dendrogram**

di
hclust (*, "ward")

Figure 2: barplot of heights of branches

```
#  Looking at the barplot, it appears that the last significant jump takes place just to the left
#  of the fifth bar to the left, thus we decide to perform k-means with 6 clusters.  This is
#  admittedly subjective ( as we could have just as easily chosen 3 clusters or ~37 clusters),
#  but unfortunately the tree doesn't provide a clearly best position where to cut.
#  Before hand we look at our cut, which isn't very well expressed in just the 1st 2 factors.
nc = 6
c1 <- cutree(hc,nc)
plot(Psi[,1],Psi[,2],type="n",main="")
text(Psi[,1],Psi[,2],col=c1, labels=rownames(sop.dat$ind$contrib),cex = .8)
abline(h=0,v=0,col="gray")
legend("topright",c("1","2","3","4","5","6"),pch=20,col=c(1:6))
```
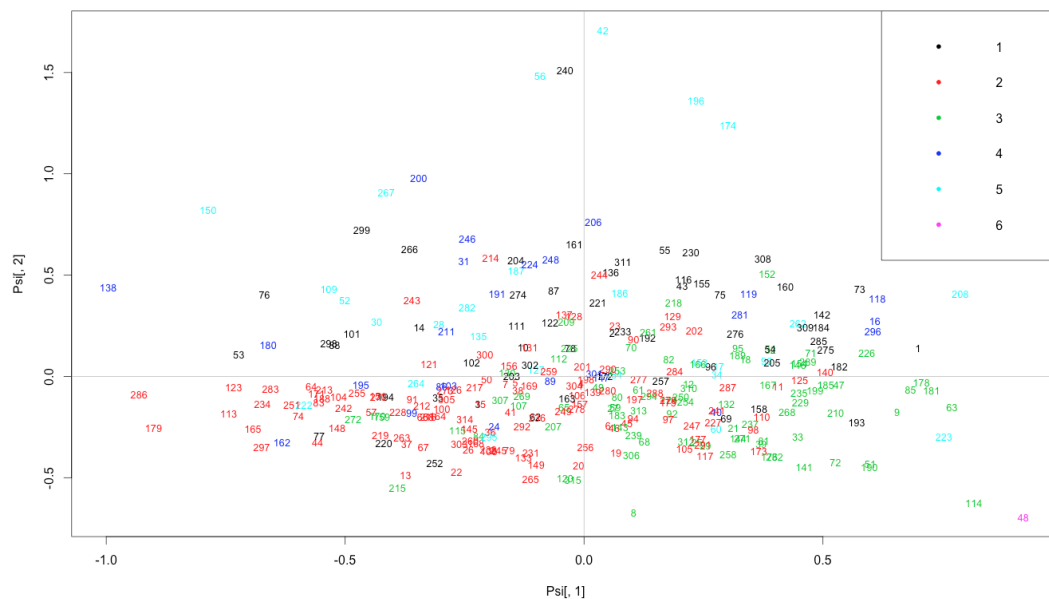
```
#We now proceed to apply k-means on our dataset with 6 clusters.
n = nrow(Psi)
end = nd + 1
cdg <- aggregate(as.data.frame(Psi),list(c1),mean)[,2:end]
Bss <- sum(rowSums(cdg^2)*as.numeric(table(c1)))
Tss <- sum(rowSums(Psi^2))

Tss /n                              #3.779619 to verify
sum(sop.dat$eig$eigenvalue[1:nd])   #3.779619
Ib4 <- 100*Bss/Tss
Ib4   #7.912371    #quality of clustering

#consolidate partition
k6 <- kmeans(Psi,centers=cdg)
Bss <- sum(rowSums(k6$centers^2)*k6$size)
Wss <- sum(k6$withinss)
Ib5 <- 100*Bss/(Bss+Wss)
Ib5   #9.886063
```

5 Interpret and name obtained clusters and represent them in the first factorial display.

```
#To get counts:
summary(dat.mca[which(dat.mca$cluster == 1),])
> length(which(dat.mca$cluster == 1))   # 62
> length(which(dat.mca$cluster == 2))   # 114
> length(which(dat.mca$cluster == 3))   # 109
> length(which(dat.mca$cluster == 4))   # 13
> length(which(dat.mca$cluster == 5))   # 16
> length(which(dat.mca$cluster == 6))   # 1

#To get descriptions
dat.mca$cluster = as.factor(k6$cluster)
res.catd <- catdes(dat.mca, num.var=53)

#This results in large number of modalities per cluster so we'll take the most significant of each:
res.catd$category$`1`
```

DIM 1	
Are.you.worried.about.the.risk.of.a.mugging=LY04 not at all	10.225553
Are.you.worried.about.the.risk.of.having.a.serious.illness=LX04 not at all	9.551310
Are.you.worried.about.the.risk.of.a.road.accident=LZ04 not at all	7.433065
Are.you.worried.about.the.risk.of.unemployment=MA04 not at all	6.543260
Are.you.worried.about.the.risk.of.a.nuclear.plant.accident=MB04 not at all	5.658656
Friends..acquaintances...importance.given=7	3.602776
Persons.like.me.often.felle.alone=MII4 not at all agree	3.469908
Work..profession...importance.given=1	3.106186
Region.where.the.respondent.lives=AA05 west	2.698312
Politic..political.life...importance.given=7	2.545766
Your.opinion.ont.he.life.conditions.in.the.future=KQA1 improving a lot	2.487467
Relatives..brothers..sisters........importance.given=7	2.423645
Housekeeping.works..take.care.of.children....=AK01 only women do it	2.380844
Your.opinion.ont.the.justice.running.in.1986=MC02 quite well	2.232824
Religion...importance.given=7	2.162689
Do.you.have.children=JZE2 - no -	2.114108
The.family.is.the.only.place.where.you.feel.well=AI02 - no -	2.075974
Family..children...importance.given=2	2.072307
Number.of.children=7	2.072307
Do.you.have.work.personal.life.problems=GC02 - no -	2.065735
Number.of.children=0	2.042316
Free.time..relax...importance.given=7	2.009600

```
Cluster 1:  The Slackers. (~ 20 percent)
People who are not worried at all,
who give a lot of importance to their friends, politics, family, religion, and freetime
who don't feel alone
```

who give little importance to their work
have no children for the most part (with the exception of the two who have 7 children)
who live in the west
who are optimistic about the future
and think only women should do housekeeping.

	Dim 2
Are.you.worried.about.the.risk.of.having.a.serious.illness=LX01 a lot	12.600006
Are.you.worried.about.the.risk.of.a.nuclear.plant.accident=MB01 a lot	9.786182
Are.you.worried.about.the.risk.of.unemployment=MA01 a lot	9.591996
Are.you.worried.about.the.risk.of.a.road.accident=LZ01 a lot	9.351238
Are.you.worried.about.the.risk.of.a.mugging=LY01 a lot	8.334892
Work..profession...importance.given=7	4.048816
Are.you.satisfied.of.your.health=HG03 a little	3.530734
The.family.is.the.only.place.where.you.feel.well=AI01 - yes -	3.427268
Your.opinion.ont.the.justice.running.in.1986=MC04 very bad	3.335872
Have.you.recently.had.headaches=HA01 - yes -	3.219800
Have.you.recently.been.depressed=HD01 - yes -	3.087563
Do.you.regularly.impose.restrictions=KD01 - yes -	2.932289
Politic..political.life...importance.given=5	2.913114
Your.opinion.ont.he.life.conditions.in.the.future=KQA4 a little worse	2.864585
Friends..acquaintances...importance.given=2	2.686184
Do.you.think.the.society.needs.to.change=LU01 - yes -	2.579280
Housekeeping.works..take.care.of.children....=AK03 men and women	2.406764
Persons.like.me.often.felle.alone=MII1 completely agree	2.405716
Do.you.have.a.video.tape=BK01 - yes -	2.341070
Have.you.recently.been.nervous=HC01 - yes -	2.309908
Do.you.watch.TV..=LT01 every day	2.189838
Relatives..brothers..sisters........importance.given=5	2.123731
Number.of.children=2	2.017645

Cluster 2: The Conservative workaholics (about ˜36 percent)
People who are worried a lot
who give high importance to work
are only a little satisfied with their health
only feel well at home
have a bad view of the justices running
have headaches, depression, are nervous
give midlevel importance to politics
think the future looks bleak
have low importance for friends
think housekeeping should be split between men and women
feel alone
have a VCR, watch TV every day
and have two children

	Dim 3
Are.you.worried.about.the.risk.of.having.a.serious.illness=LX02 enough	8.326336
Are.you.worried.about.the.risk.of.a.road.accident=LZ02 enough	6.051249
Are.you.worried.about.the.risk.of.a.mugging=LY03 a little	5.447379
Relatives..brothers..sisters........importance.given=6	5.322562
Politic..political.life...importance.given=3	5.031342
Are.you.worried.about.the.risk.of.a.nuclear.plant.accident=MB03 a little	4.180782
Housekeeping.works..take.care.of.children....=AK02 usually the women	4.066951
Are.you.worried.about.the.risk.of.unemployment=MA02 enough	3.982039
Friends..acquaintances...importance.given=6	3.968481
Are.you.worried.about.the.risk.of.unemployment=MA03 a little	3.530125
Politic..political.life...importance.given=6	3.502936
Have.you.recently.had.headaches=HA02 - no -	3.422204
Salary.of.the.respondent=8500 to 40000	3.352485
Are.you.worried.about.the.risk.of.a.mugging=LY02 enough	3.265975
Do.you.have.work.personal.life.problems=GC01 - yes -	3.165320
Are.you.worried.about.the.risk.of.a.nuclear.plant.accident=MB02 enough	2.967071
Are.you.worried.about.the.risk.of.having.a.serious.illness=LX03 a little	2.541544
Have.you.recently.been.nervous=HC02 - no -	2.508113
Have.you.recently.had.backaches=HB02 - no -	2.495812
Have.you.recently.been.depressed=HD02 - no -	2.405791
Work..profession...importance.given=5	2.365977
Region.where.the.respondent.lives=AA07 center east	2.351425
Free.time..relax...importance.given=5	2.322858
Do.you.have.a.piano=BL01 - yes -	2.295470
At.the.moment..do.you.have.a.professional.activity=GB01 yes, full time	2.247311
Your.opinion.ont.the.justice.running.in.1986=MC02 quite well	2.199001
Have.you.been.unemployed.during.the.last.twelve.months=GD02 - no -	2.124136
Educational.level.of.the.respondent=AH07 high school diploma	2.114954
Current.Situation.of.the.respondent=AD01 employed	2.009548
Are.you.satisfied.of.your.housing=BC01 a lot	1.980495
Persons.like.me.often.felle.alone=MII3 little agree	1.976339
Are.you.worried.about.the.risk.of.a.road.accident=LZ03 a little	1.976339

Cluster 3: The Normal Joes (about 35 percent)
People who are worried enough or a little
place high importance on brothers,sisters,friends,
have salaries between 8500 and 40000
aren't nervous, depressed or have backaches
live in the center east
have a piano
have a fulltime job
have a quite well view of justices running
have a high school education
are satisfied of their housing

	Dim 4
Free.time..relax...importance.given=2	7.203816
Relatives..brothers..sisters........importance.given=2	5.746018
Family..children...importance.given=3	3.160092
Your.opinion.ont.he.life.conditions.in.the.future=KQA5 a lot worse	2.880314
Are.you.satisfied.of.your.housing=BC03 a little	2.597034
Friends..acquaintances...importance.given=2	2.558357
Politic..political.life...importance.given=2	2.543743
Are.you.worried.about.the.risk.of.having.a.serious.illness=LX03 a little	2.435266
Work..profession...importance.given=2	2.427155
Housekeeping.works..take.care.of.children....=AK01 only women do it	2.313816
State.benefits...average.monthly.amount=not answered	2.116511
Persons.like.me.often.felle.alone=MII2 quite agree	1.973769

Cluster 4: The Depressed (about 4 percent)
People who give low importance to freetime/relaxing
give low importance to relatives,family,friends, poltics, and work
who think the future is going to be a lot worse
are only a little satisfied with their house
who are only a little worried about serious illness
who think only women should take care of the home
who didn't respond to how much benefits they get a month
and who feel alone

	Dim 5
The.computer.science.diffusion.is....=LW04 I do not know	5.811502
Friends..acquaintances...importance.given=1	4.585818
Free.time..relax...importance.given=1	4.585818
Educational.level.of.the.respondent=AH01 no qualifications	4.112913
Age.of.respondent=63 and up	3.586702
Have.you.been.unemployed.during.the.last.twelve.months=not answered	3.495895
Do.you.have.work.personal.life.problems=not answered	3.495895
Your.opinion.ont.the.justice.running.in.1986=MC06 do not answer	3.279357
Number.of.children=8	3.032366
The.family.is.the.only.place.where.you.feel.well=AI03 I do not know	3.032366
Relatives..brothers..sisters........importance.given=3	2.738061
Your.opinion.on.the.evolution.of.French.people.life.level=KP06 I do not know	2.738061
At.the.moment..do.you.have.a.professional.activity=GB03 no	2.723087
Politic..political.life...importance.given=1	2.544273
Are.you.satisfied.of.your.health=HG04 not at all	2.536012
Do.you.think.the.society.needs.to.change=LU03 I do not know	2.528932
Current.Situation.of.the.respondent=AD05 retired people	2.521209
Marital.status=AG05 widower/widow	2.505420
Are.you.satisfied.of.your.daily.life=BD01 a lot	2.481065
Has.the.respondent.been.interested.by.the.survey=MO03 a little or not	2.434399
Your.opinion.ont.he.life.conditions.in.the.future=KQA6 I do not know	2.241082
Have.you.recently.had.backaches=HB01 - yes -	2.230571
Work..profession...importance.given=1	2.155183
Relatives..brothers..sisters........importance.given=1	2.117373
Are.you.worried.about.the.risk.of.unemployment=MA04 not at all	2.007625
Urban.area.size..number.of.inhabitants.=AB01 less than 2000	2.003772

Cluster 5: The Elderly who don't care (about 5 percent)
People who don't know about computers
place no importance on friends, freetime, work, relatives
are not educated
are older than 63
who did not respond or responded to don't know a lot
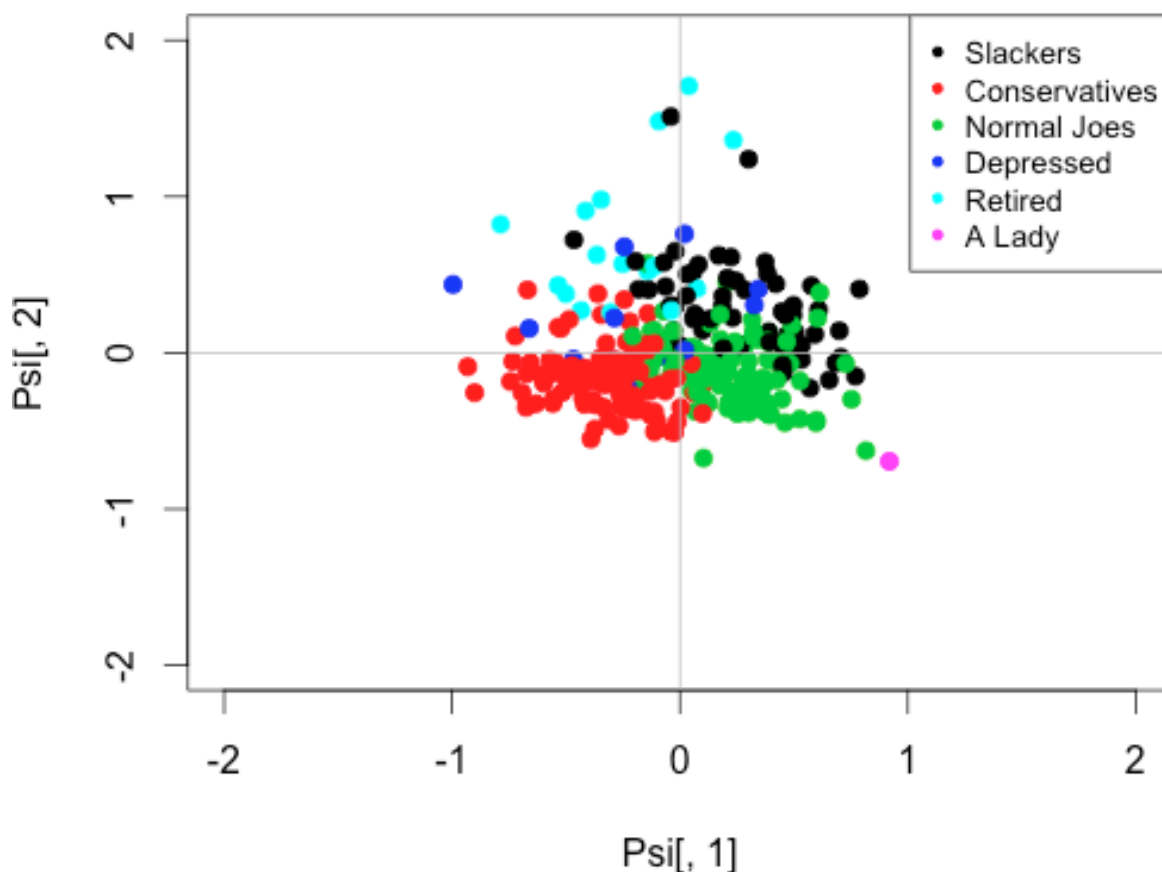are retired

	Dim 6
Politic..political.life...importance.given=0	2.950305
Religion...importance.given=0	2.950305
Your.opinion.ont.the.justice.running.in.1986=MC05 I do not know	2.073829

```
Cluster 6: A Lady Apart ( 1 person)
An under 28 year old, married employed women who lives in the east
and who is distinguishable solely because she gives no importance
to politics or religion and doesn't know about the justice running.
She represents a possible bad selection of cutting the hierarchical
tree, but is also seen on the edge of the plot in the next graphic.

Our representation of these clusters in the first factorial plane is as follows:
plot(Psi[,1],Psi[,2],type="p",main="",xlim=c(-2,2),ylim=c(-2,2),col=k6$cluster, pch=19)
abline(h=0,v=0,col="gray")
clusternames = c("Slackers","Conservatives","Normal Joes","Depressed","Retired","A Lady")
legend("topright",clusternames,pch=20,col=c(1:6),cex=.8)
```


6   Explain how you would assign new individuals to the defined clusters

```
# Without using a classifying technique, I would simply check to see
# how this new individual's responses compared to our six existing clusters
# and unless they were retired or depressed, they would go in one of the first 3
# clusters.  After that I'd base it on how much they "worry" in general,
# and use that for the most part to place them.  1 if they don't worry,
# 2 if they worry alot and 3 if they are in the midrange.
```

	Region where respondent lives	Urban area size number of inhabitants.	Sex of respondent	Age of respondent
1	AA01 Paris region :56	AB01 less than 2000 :84	AC01 male :138	28 and under:63
2	AA02 Paris basin :51	AB08 more than 200000 :68	AC02 female:177	29 to 35 :65
3	AA05 west :45	AB09 paris.agglo.paris :46		36 to 49 :71
4	AA06 south-west :38	AB07 100001 to 200000:28		50 to 62 :62
5	AA07 center east :36	AB05 20001 to 50000 :23		63 and up :54
6	AA08 mediterranean:36	AB02 2001 to 5000 :18		
7	(Other) :53	(Other) :48		

	Current Situation	Marital status	Educational level	The family is the only place. where you feel well
1	AD01 employed :172	AG01 single : 42	AH03 vocational training:56	AI01 - yes - :230
2	AD02 student : 12	AG02 married :223	AH01 no qualifications :54	AI02 - no - : 83
3	AD03 housewife w/o prof.: 52	AG03 cohabitation : 16	AH02 primary school :54	AI03 Don't know: 2
4	AD04 ill, disabled person : 2	AG04 separated divorced: 15	AH07 high school diploma:45	
5	AD05 retired people : 54	AG05 widower/widow : 19	AH08 more high school :43	
6	AD06 soldier : 23		AH04 GCSE diploma :28	
7	AD07 unemployed person : 0		(Other) :35	

	Opinion about marriage	Housekeeping/care of children	Satisfied of housing	Satisfied of your daily life
1	AJ01 indissoluble : 81	AK01 only women do it : 15	BC01 a lot :132	BD01 a lot : 85
2	AJ02 dissolved serious pb:108	AK02 usually the women: 85	BC02 enough :150	BD02 enough :180
3	AJ03 dissolved if agreem :114	AK03 men and women :214	BC03 a little : 22	BD03 a little : 37
4	AJ04 I do not know : 12	AK04 I do not know : 1	BC04 not at all: 11	BD04 not at all: 13

	Status of housing	The housing expenses	Have VCR	Have piano	Have second house
1	BE01 homeowner : 33	not answered : 1	BK01 - yes - : 51	BL01 - yes -: 34	BM01 - yes -: 42
2	BE02 owner :133	BF01 unimportant : 35	BK02 - no - :264	BL02 - no - :281	BM02 - no - :273
3	BE03 tenant :131	BF02 without big problem:154			
4	BE04 free housing: 17	BF03 a big problem : 99			
5	BE05 other : 1	BF04 a very big problem : 18			
6		BF05 do not face with : 1			
7		BF06 I do not know : 7			

	Professional activity	Work/personal problems	Unemployed.during last year	Recent headaches	Recent backaches
1	GB01 yes, full time :143	not answered :142	not answered :142	HA01 - yes -:115	HB01 - yes -:149
2	GB02 yes, part time : 29	GC01 - yes -: 76	GD01 - yes -: 17	HA02 - no - :200	HB02 - no - :166
3	GB03 no :111	GC02 - no - : 97	GD02 - no - :156		
4	GB04 never worked: 32				

	Recently nervous	Recently depressed	Satisfied with health	Num persons in house	Have children	Num children
1	HC01 - yes - :155	HD01 - yes -: 50	HG01 a lot : 80	2 :90	not answered : 3	2 :94
2	HC02 - no - :160	HD02 - no - :265	HG02 satisfied :191	4 :71	JZE1 - yes -:243	0 :70
3			HG03 a little : 34	3 :69	JZE2 - no - : 69	1 :67
4			HG04 not at all: 10	1 :38		3 :54
5				5 :34		5 :11
6				6 : 7		4 : 9
7				(Other): 6		(Other):10

Table 1: Data set combined and clean

	Impose restrictions	Evolution .. of personal life	.. of French life level	.. Future life	Watch TV..
1	KD01 yes :193	KO01 a lot better :17	not answered : 1	KQA1 improving a lot :21	LT01 daily: 176
2	KD02 no :122	KO02 a little better:71	KP01 a lot better : 7	KQA2 improving some:69	LT02 often: 76
3		KO03 it is the same :87	KP02 a little better: 50	KQA3 the same :95	LT03 not often: 5
4		KO04 a little worse :86	KP03 it is the same : 45	KQA4 a little worse :88	LT04 never: 9
5		KO05 a lot worse :51	KP04 a little worse :130	KQA5 a lot worse :29	
6		KO06 I do not know : 3	KP05 a lot worse : 66	KQA6 I do not know :13	
7			KP06 I do not know : 16		

	Society needs change	Computer science diffusion is ..	Worried about the risk of having a serious illness	Worried about the risk of a mugging	Worried about the risk of a road accident
1	LU01 - yes - :218	LW01 desirable :109	LX01 a lot :146	LY01 a lot :92	LZ01 a lot :115
2	LU02 - no - : 68	LW02 inevitable :170	LX02 enough : 76	LY02 enough :46	LZ02 enough : 98
3	LU03 don't know: 29	LW03 dangerous : 30	LX03 a little : 54	LY03 a little :94	LZ03 a little : 64
4		LW04 I do not know: 6	LX04 not at all: 39	LY04 not at all:83	LZ04 not at all: 38

	Worried about the risk of unemployment	Worried about the risk of a nuclear plant accident	Family importance	Work importance	Free time importance	Friends importance
1	MA01 a lot :125	MB01 a lot :89	1: 5	1: 14	7 :84	1: 4
2	MA02 enough : 71	MB02 enough :53	2: 2	2: 5	5 :79	2:12
3	MA03 a little : 51	MB03 a little :94	3: 2	3: 6	6 :64	3:21
4	MA04 not at all: 68	MB04 not at all:79	4: 8	4: 13	4 :56	4:54
5			5: 11	5: 48	3 :19	5:82
6			6: 16	6: 61	2 : 8	6:76
7			7:271	7:168	(Other): 5	7:66

	Relatives/brothers importance	Religion importance	Politics importance	Opinion on the justice running in 1986	People like me often fell alone
1	1: 6	1 :95	1 :83	MC01 very well : 5	MII1 completely agree: 32
2	2: 7	4 :52	4 :56	MC02 quite well : 99	MII2 quite agree : 68
3	3: 16	2 :41	2 :51	MC03 quite bad :113	MII3 little agree : 64
4	4: 28	3 :38	5 :49	MC04 very bad : 81	MII4 not at all agree:151
5	5: 65	5 :37	3 :45	MC05 I do not know: 12	MII5 I do not know : 0
6	6: 83	7 :31	6 :19	MC06 do not answer: 5	
7	7:110	(Other):21	(Other):12		

	State benefits average (monthly)	Salary of respondent	Has respondent been interested by survey
1	1000 to 2000 : 27	3200 to 4900 : 50	MO01 a lot :132
2	2001 to 5100 : 24	5000 to 8400 : 74	MO02 enough :163
3	600 to 1000 : 37	8500 to 40000 : 32	MO03 a little or not: 20
4	less than 600: 20	less than 3200: 49	
5	not answered :207	not answered :110	

Table 2: Data set continued: scale of 1 - 7 is from not at all to very much