Gastone Riccardo Tolli
ID: 33506151

Date: 20/04/2018

IS71069B Geometric Data Analysis

Assignment 2

Introduction and Objectives

The dataset used is the European Social Survey (ESS) round 6 from 2012. The survey is taken in 36 European countries and it is a collection of social statistics used to monitor and interpret public attitudes and values. The Sampling method used was partly repetitive cross section. Some of the columns had been removed because of too many missing values.

The objective of the project is to investigate some of the questions in the survey that are part of how people see themselves, namely the master question H inside of the ESS round 6. Conclusions will be drawn if any is found from a multiple correspondence analysis on the data and its multiple visualizations.

Data preparation and analysis

Preliminary data preparation

The software used for the analysis is R (version 3.4.3) used through RStudio.
The main library used is FactoMiner and factoextra which has been downloaded with the following command:

```
> install.packages("FactoMineR")
> install.packages("factoextra")
```

and loaded with:

```
> library(FactoMineR)
> library(factoextra)
```

Loading the dataset:

```
> load("C:/Users/rikka/ESS3.RData")
```

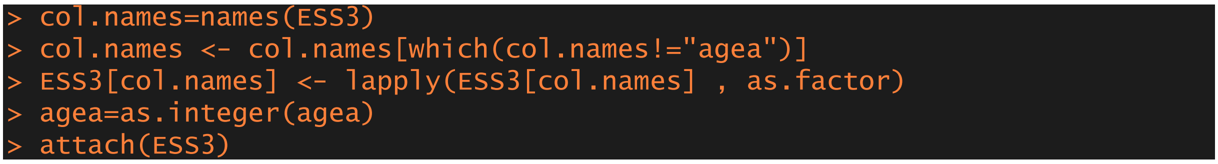
dimensions of dataframe:

```
> dim(ESS3)
[1] 52177    50
```

The dataset contains 52177 observations and 50 variables, each observation represent a person and the variables are:

```
> names(ESS3)
 [1] "cntry"   "gndr"    "agea"    "eisced"  "ipcrtiv" "imprich" "ipeqopt"
 [8] "ipshabt" "impsafe" "impdiff" "ipfrule" "ipudrst" "ipmodst" "ipgdtim"
[15] "impfree" "iphlppl" "ipsuces" "ipstrgv" "ipadvnt" "ipbhprp" "iprspot"
[22] "iplylfr" "impenv"  "imptrad" "impfun"  "wkvlorg" "optftr"  "pstvms"
[29] "flrms"   "fltdpr"  "flteeff" "slprl"   "wrhpp"   "fltlnl"  "enjlf"
[36] "fltsd"   "cldgng"  "enrglot" "fltanx"  "fltpcfl" "dclvlf"  "lchshcp"
[43] "accdng"  "wrbknrm" "pplahlp" "trtrsp"  "dngval"  "nhpftr"  "lfwrs"
[50] "flclpla"
```

Converting all the columns into factor, beside agea which is converted to integer.



```
> col.names=names(ESS3)
> col.names <- col.names[which(col.names!="agea")]
> ESS3[col.names] <- lapply(ESS3[col.names] , as.factor)
> agea=as.integer(agea)
> attach(ESS3)
```

Variable description

I will describe every variable below and their possible values because the variable names are not self-explanatory. The variable description are taken from a different source (http://www.europeansocialsurvey.org/docs/round6/survey/ESS6_appendix_a7_e02_2.pdf) than prof.Fionn Murtagh's document because many of these variable are not described in the document on his website.

I will provide here a summary list with the definition for each variable in order to quickly understand the meaning of the variables (the definitions are taken from the above link):

contextual variables
cntry = country
gndr = gender
agea = age
eisced = education level

First set of questions:
For the variables below the following question applies = "Now I will briefly describe some people. Please listen to each description and tell me how much each person is or is not like you. Use this card for your answer".

1 Very much like me
2 Like me
3 Somewhat like me
4 A little like me
5 Not like me
6 Not like me at all
7 Refusal
8 Don't know
9 No answer

ipcrtiv = Important to think new ideas and being creative
imprich = Important to be rich, have money and expensive things
ipeqopt = Important that people are treated equally and have equal opportunities
ipshabt = Important to show abilities and be admired
impsafe = Important to live in secure and safe surroundings
impdiff = Important to try new and different things in life
ipfrule = Important to do what is told and follow rules
ipudrst = Important to understand different people
ipmodst = Important to be humble and modest, not draw attention
ipgdtim = Important to have a good time

impfree = Important to make own decisions and be free
iphlppl = Important to help people and care for others well-being
ipsuces = Important to be successful and that people recognize achievements
ipstrgv = Important that government is strong and ensures safety
ipadvnt = Important to seek adventures and have an exciting life
ipbhprp = Important to behave properly
iprspot = Important to get respect from others
iplylfr = Important to be loyal to friends and devote to people close
impenv = Important to care for nature and environment
imptrad = Important to follow traditions and customs
impfun = Important to seek fun and things that give pleasure

Second set of questions:
The next variables have different value encodings, therefore they are difficult to group, I will only provide their definition**.**

wkvlorg = Involved in work for voluntary or charitable organizations, how often past 12 months
optftr= Always optimistic about my future
pstvms = In general feel very positive about myself
flrms = At times feel as if I am a failure
fltdpr = Felt depressed, how often past week
flteeff = Felt everything did as effort, how often past week
slprl = Sleep was restless, how often past week
wrhpp = Were happy, how often past week
fltlnl = Felt lonely, how often past week
enjlf = Enjoyed life, how often past week
fltsd = Felt sad, how often past week
cldgng = Could not get going, how often past week
enrglot = Had lot of energy, how often past week
fltanx = Felt anxious, how often past week
fltpcfl = Felt calm and peaceful, how often past week
dclvlf = Free to decide how to live my life
lchshcp = Little chance to show how capable I am
accdng = Feel accomplishment from what I do
wrbknrm = When things go wrong in my life it takes a long time to get back to normal
pplahlp = Feel people in local area help one another
trtrsp = Feel people treat you with respect
dngval = Feel what I do in life is valuable and worthwhile
nhpftr = Hard to be hopeful about the future of the world
lfwrs = For most people in country life is getting worse
flclpla = Feel close to the people in local area

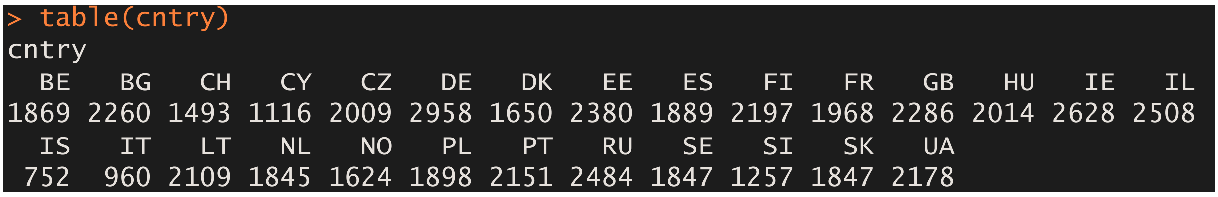
In depth description for the contextual variables

cntry= country

Possible values =
AT Austria
BE Belgium

BG Bulgaria
CH Switzerland
CY Cyprus
CZ Czech Republic
DE Germany
DK Denmark
EE Estonia
ES Spain
FI Finland
FR France
GB United Kingdom
GR Greece
HR Croatia
HU Hungary
IE Ireland
IL Israel
IS Iceland
IT Italy
LT Lithuania
LU Luxembourg
NL Netherlands
NO Norway
PL Poland
PT Portugal
RU Russia
SE Sweden
SI Slovenia
SK Slovakia
TR Turkey
UA Ukraine

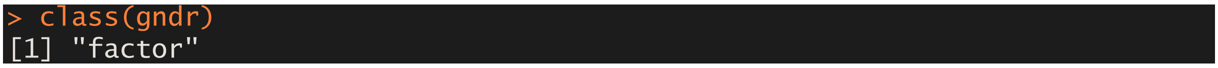
Table showing how many observations per each variable:



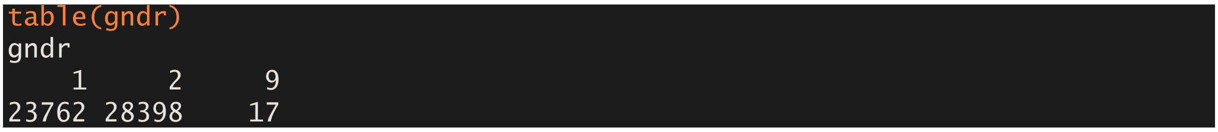
```
> table(cntry)
cntry
  BE   BG   CH   CY   CZ   DE   DK   EE   ES   FI   FR   GB   HU   IE   IL 
1869 2260 1493 1116 2009 2958 1650 2380 1889 2197 1968 2286 2014 2628 2508 
  IS   IT   LT   NL   NO   PL   PT   RU   SE   SI   SK   UA 
 752  960 2109 1845 1624 1898 2151 2484 1847 1257 1847 2178 
```

gndr = gender
possible values= 1 for male, 2 for female,9 for no answer

Turning the column gndr into a factor



```
> class(gndr)
[1] "factor"
```



```
table(gndr)
gndr
    1     2     9 
23762 28398    17 
```

Gastone Riccardo Tolli
ID: 33506151

Date: 20/04/2018

Agea= age, 999 is for no answer
Visualizing age as a histogram since a table is not much use

```
> hist(agea)
```

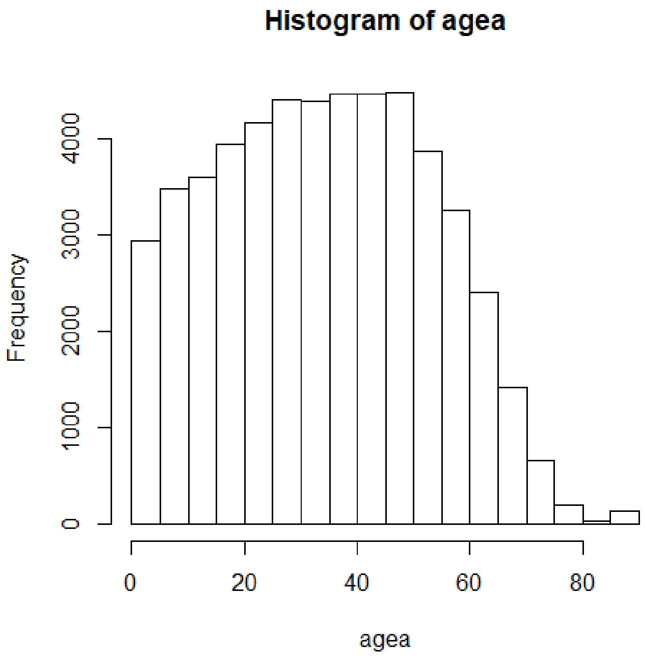
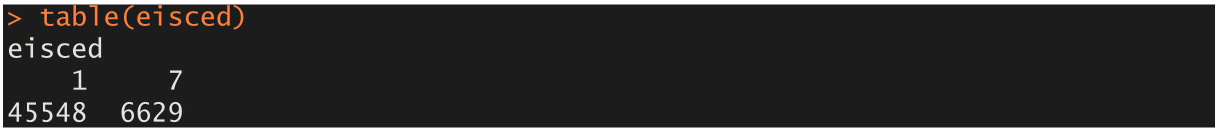


Fig.1

eisced= highest level of education, 1 for less than lower secondary and 7 for higher tertiary education (>= MA level)

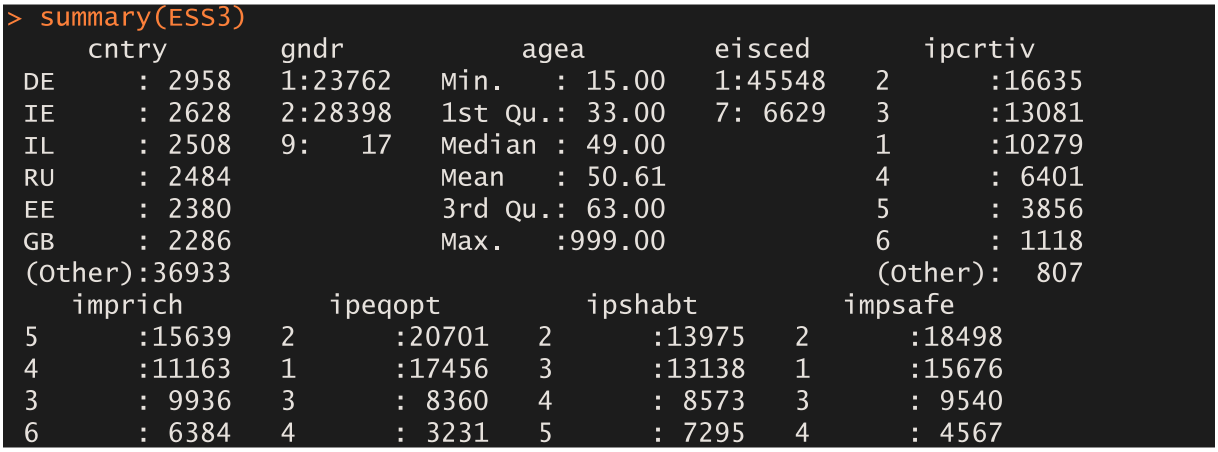


```
> table(eisced)
eisced
    1     7 
45548  6629 
```

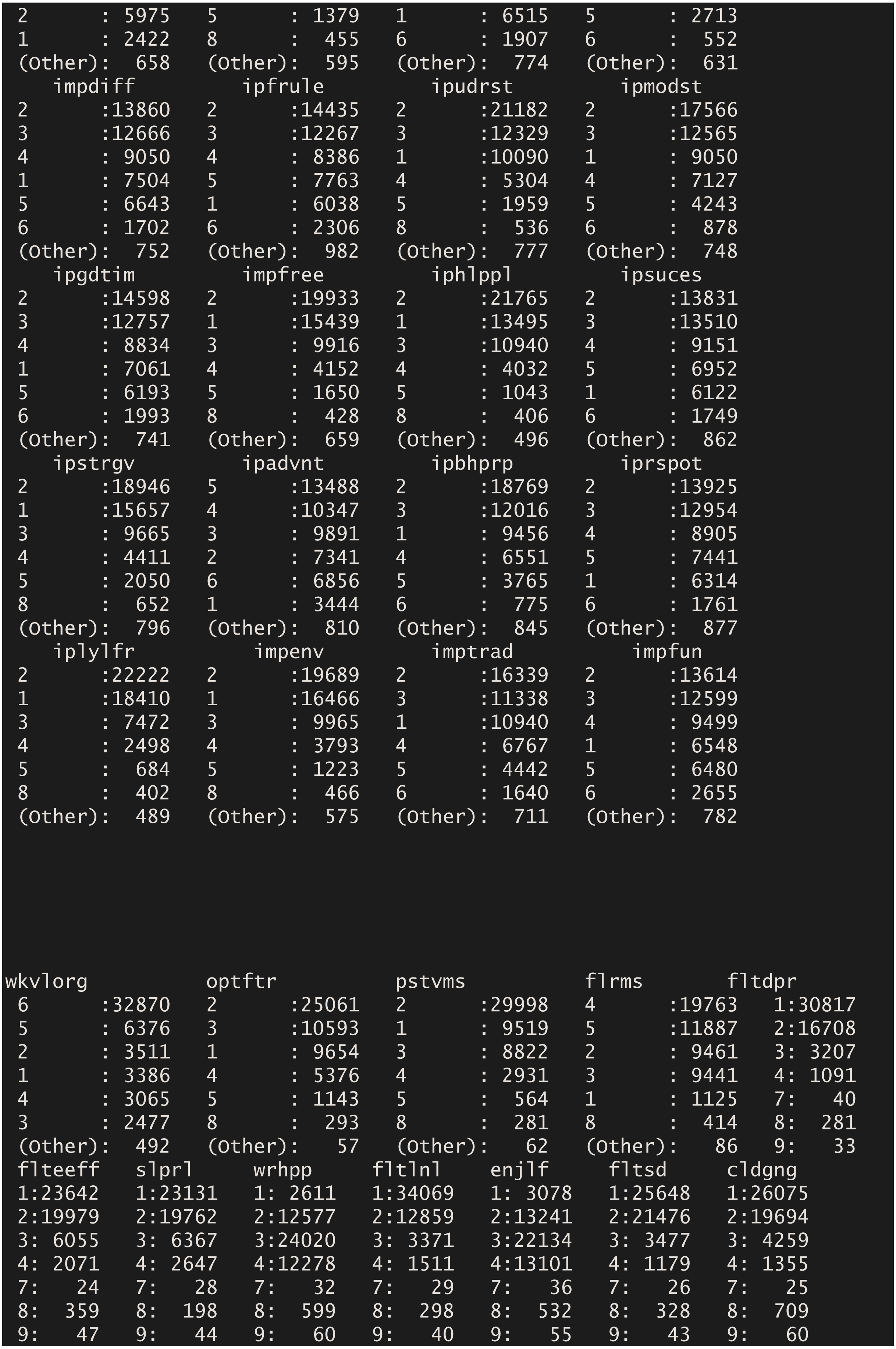
The rest of the questions all follow their own value coding therefore they cannot be easily grouped.

Detailed description of the questions can be found at this address http://www.europeansocialsurvey.org/docs/round6/survey/ESS6_appendix_a7_e02_2.pdf

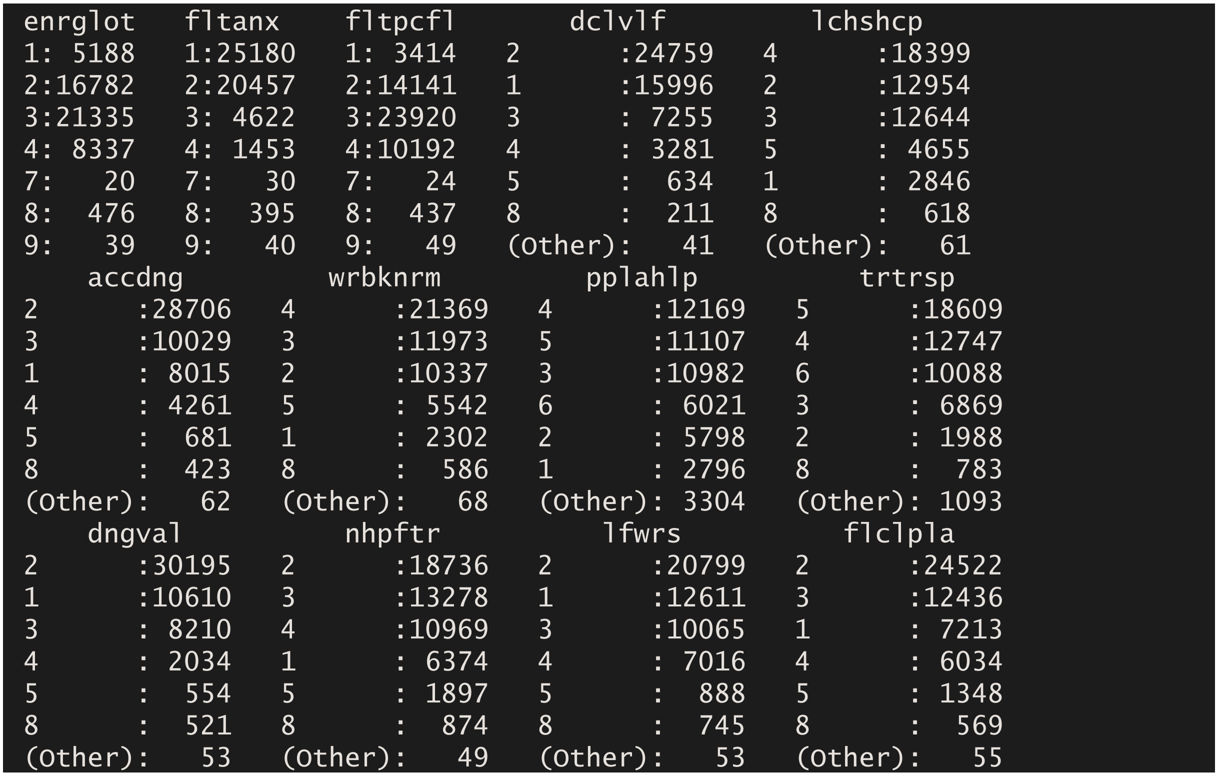
Overview of the dataset:



```
> summary(ESS3)
     cntry          gndr           agea          eisced        ipcrtiv     
 DE     : 2958   1:23762   Min.   : 15.00   1:45548   2      :16635  
 IE     : 2628   2:28398   1st Qu.: 33.00   7: 6629   3      :13081  
 IL     : 2508   9:   17   Median : 49.00             1      :10279  
 RU     : 2484             Mean   : 50.61             4      : 6401  
 EE     : 2380             3rd Qu.: 63.00             5      : 3856  
 GB     : 2286             Max.   :999.00             6      : 1118  
 (Other):36933                                        (Other):  807  
    imprich         ipeqopt          ipshabt          impsafe     
 5      :15639   2      :20701   2      :13975   2      :18498  
 4      :11163   1      :17456   3      :13138   1      :15676  
 3      : 9936   3      : 8360   4      : 8573   3      : 9540  
 6      : 6384   4      : 3231   5      : 7295   4      : 4567  
```



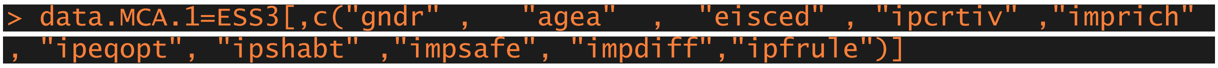
```
 2      : 5975   5      : 1379   1      : 6515   5      : 2713
 1      : 2422   8      :  455   6      : 1907   6      :  552
 (Other):  658   (Other):  595   (Other):  774   (Other):  631
    impdiff         ipfrule         ipudrst         ipmodst
 2      :13860   2      :14435   2      :21182   2      :17566
 3      :12666   3      :12267   3      :12329   3      :12565
 4      : 9050   4      : 8386   1      :10090   1      : 9050
 1      : 7504   5      : 7763   4      : 5304   4      : 7127
 5      : 6643   1      : 6038   5      : 1959   5      : 4243
 6      : 1702   6      : 2306   8      :  536   6      :  878
 (Other):  752   (Other):  982   (Other):  777   (Other):  748
    ipgdtim         impfree         iphlppl         ipsuces
 2      :14598   2      :19933   2      :21765   2      :13831
 3      :12757   1      :15439   1      :13495   3      :13510
 4      : 8834   3      : 9916   3      :10940   4      : 9151
 1      : 7061   4      : 4152   4      : 4032   5      : 6952
 5      : 6193   5      : 1650   5      : 1043   1      : 6122
 6      : 1993   8      :  428   8      :  406   6      : 1749
 (Other):  741   (Other):  659   (Other):  496   (Other):  862
    ipstrgv         ipadvnt         ipbhprp         iprspot
 2      :18946   5      :13488   2      :18769   2      :13925
 1      :15657   4      :10347   3      :12016   3      :12954
 3      : 9665   3      : 9891   1      : 9456   4      : 8905
 4      : 4411   2      : 7341   4      : 6551   5      : 7441
 5      : 2050   6      : 6856   5      : 3765   1      : 6314
 8      :  652   1      : 3444   6      :  775   6      : 1761
 (Other):  796   (Other):  810   (Other):  845   (Other):  877
    iplylfr          impenv         imptrad          impfun
 2      :22222   2      :19689   2      :16339   2      :13614
 1      :18410   1      :16466   3      :11338   3      :12599
 3      : 7472   3      : 9965   1      :10940   4      : 9499
 4      : 2498   4      : 3793   4      : 6767   1      : 6548
 5      :  684   5      : 1223   5      : 4442   5      : 6480
 8      :  402   8      :  466   6      : 1640   6      : 2655
 (Other):  489   (Other):  575   (Other):  711   (Other):  782




wkvlorg          optftr           pstvms           flrms       fltdpr
 6      :32870   2      :25061   2      :29998   4      :19763   1:30817
 5      : 6376   3      :10593   1      : 9519   5      :11887   2:16708
 2      : 3511   1      : 9654   3      : 8822   2      : 9461   3: 3207
 1      : 3386   4      : 5376   4      : 2931   3      : 9441   4: 1091
 4      : 3065   5      : 1143   5      :  564   1      : 1125   7:   40
 3      : 2477   8      :  293   8      :  281   8      :  414   8:  281
 (Other):  492   (Other):   57   (Other):   62   (Other):   86   9:   33
 flteeff   slprl     wrhpp     fltlnl    enjlf     fltsd     cldgng
 1:23642   1:23131   1: 2611   1:34069   1: 3078   1:25648   1:26075
 2:19979   2:19762   2:12577   2:12859   2:13241   2:21476   2:19694
 3: 6055   3: 6367   3:24020   3: 3371   3:22134   3: 3477   3: 4259
 4: 2071   4: 2647   4:12278   4: 1511   4:13101   4: 1179   4: 1355
 7:   24   7:   28   7:   32   7:   29   7:   36   7:   26   7:   25
 8:  359   8:  198   8:  599   8:  298   8:  532   8:  328   8:  709
 9:   47   9:   44   9:   60   9:   40   9:   55   9:   43   9:   60
```



```
 enrglot    fltanx     fltpcfl       dclvlf          lchshcp
 1: 5188    1:25180    1: 3414    2       :24759    4       :18399
 2:16782    2:20457    2:14141    1       :15996    2       :12954
 3:21335    3: 4622    3:23920    3       : 7255    3       :12644
 4: 8337    4: 1453    4:10192    4       : 3281    5       : 4655
 7:   20    7:   30    7:   24    5       :  634    1       : 2846
 8:  476    8:  395    8:  437    8       :  211    8       :  618
 9:   39    9:   40    9:   49    (Other):   41    (Other):   61
     accdng          wrbknrm          pplahlp          trtrsp
 2       :28706    4       :21369    4       :12169    5       :18609
 3       :10029    3       :11973    5       :11107    4       :12747
 1       : 8015    2       :10337    3       :10982    6       :10088
 4       : 4261    5       : 5542    6       : 6021    3       : 6869
 5       :  681    1       : 2302    2       : 5798    2       : 1988
 8       :  423    8       :  586    1       : 2796    8       :  783
 (Other):   62    (Other):   68    (Other): 3304    (Other): 1093
     dngval           nhpftr           lfwrs            flclpla
 2       :30195    2       :18736    2       :20799    2       :24522
 1       :10610    3       :13278    1       :12611    3       :12436
 3       : 8210    4       :10969    3       :10065    1       : 7213
 4       : 2034    1       : 6374    4       : 7016    4       : 6034
 5       :  554    5       : 1897    5       :  888    5       : 1348
 8       :  521    8       :  874    8       :  745    8       :  569
 (Other):   53    (Other):   49    (Other):   53    (Other):   55
```

Multiple Correspondence Analysis

I am going to perform an MCA on the contextual variables and the first set of questions. Creating new dataset with the columns used for this analysis.
As during the analysis, the graphs were incredibly overcrowded I am removing some variables. Namely cntry from the contextual variables and all the variables after ipfrule.

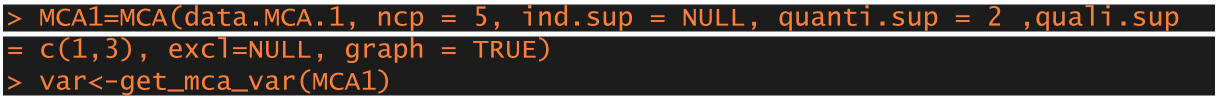


```
> data.MCA.1=ESS3[,c("gndr" ,   "agea"   ,  "eisced" , "ipcrtiv" ,"imprich"
, "ipeqopt", "ipshabt" ,"impsafe", "impdiff","ipfrule")]
```

These variables have information about how much respondents identify with the specific question asked (each variable is a question).

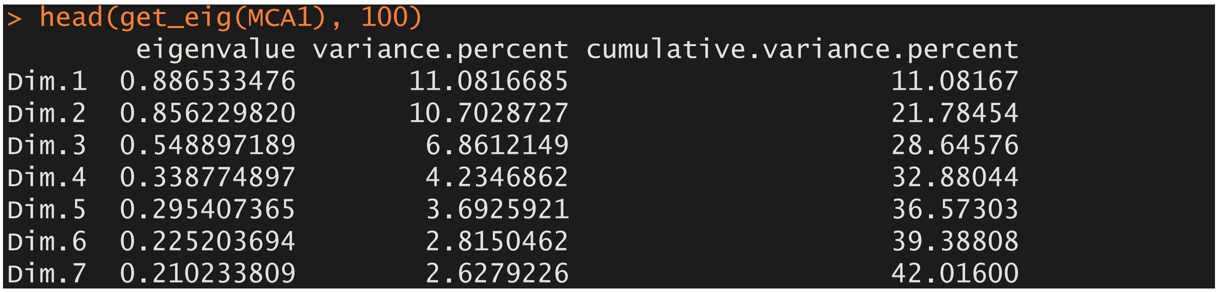
The variables cntry, gndr, agea and eisced are considered supplementary.

Carrying out the MCA

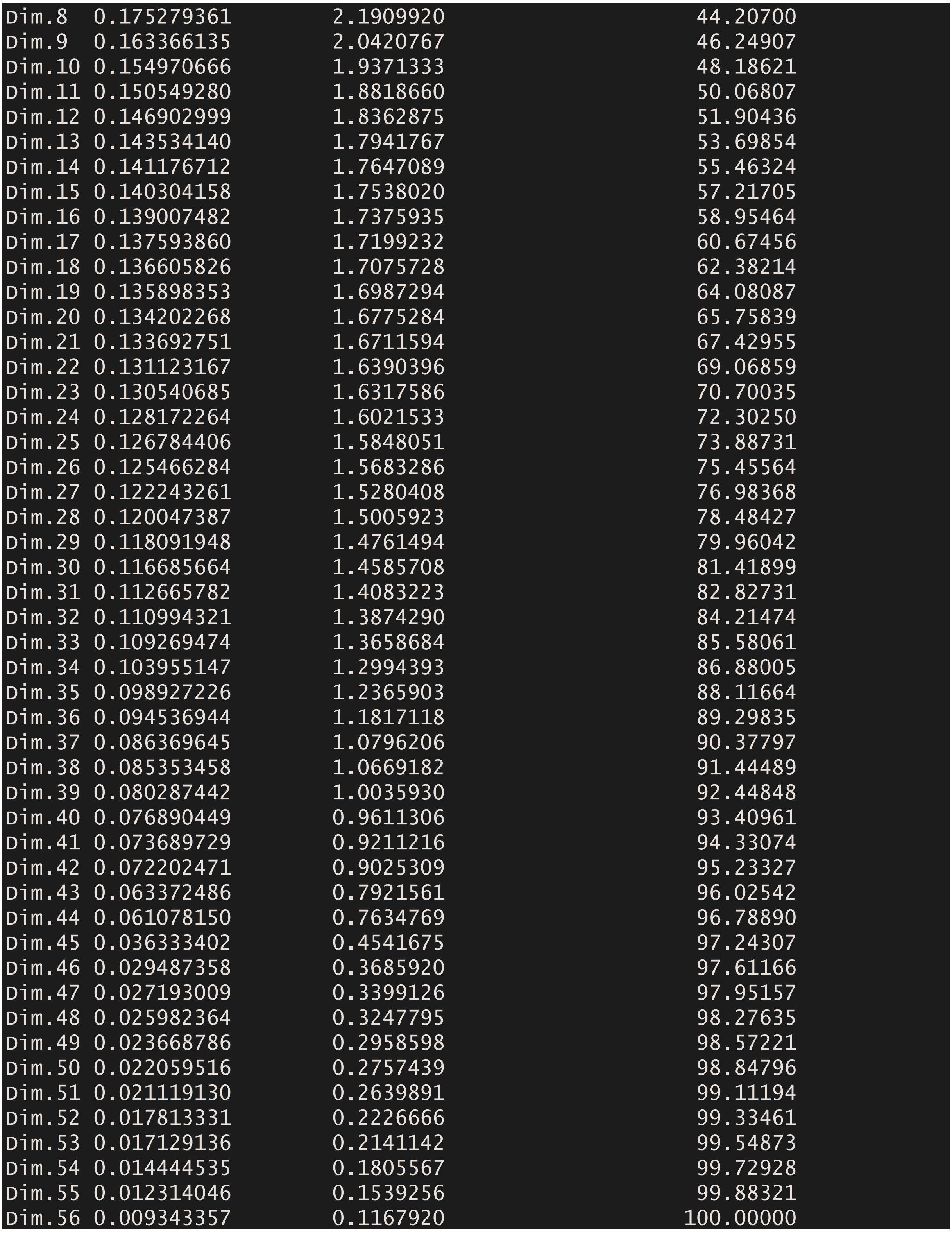


```
> MCA1=MCA(data.MCA.1, ncp = 5, ind.sup = NULL, quanti.sup = 2 ,quali.sup
= c(1,3), excl=NULL, graph = TRUE)
> var<-get_mca_var(MCA1)
```

Let us look at the explained variance and eigenvalues for the top 25 dimensions



```
> head(get_eig(MCA1), 100)
        eigenvalue variance.percent cumulative.variance.percent
Dim.1  0.886533476       11.0816685                    11.08167
Dim.2  0.856229820       10.7028727                    21.78454
Dim.3  0.548897189        6.8612149                    28.64576
Dim.4  0.338774897        4.2346862                    32.88044
Dim.5  0.295407365        3.6925921                    36.57303
Dim.6  0.225203694        2.8150462                    39.38808
Dim.7  0.210233809        2.6279226                    42.01600
```



```
Dim.8  0.175279361        2.1909920                    44.20700
Dim.9  0.163366135        2.0420767                    46.24907
Dim.10 0.154970666        1.9371333                    48.18621
Dim.11 0.150549280        1.8818660                    50.06807
Dim.12 0.146902999        1.8362875                    51.90436
Dim.13 0.143534140        1.7941767                    53.69854
Dim.14 0.141176712        1.7647089                    55.46324
Dim.15 0.140304158        1.7538020                    57.21705
Dim.16 0.139007482        1.7375935                    58.95464
Dim.17 0.137593860        1.7199232                    60.67456
Dim.18 0.136605826        1.7075728                    62.38214
Dim.19 0.135898353        1.6987294                    64.08087
Dim.20 0.134202268        1.6775284                    65.75839
Dim.21 0.133692751        1.6711594                    67.42955
Dim.22 0.131123167        1.6390396                    69.06859
Dim.23 0.130540685        1.6317586                    70.70035
Dim.24 0.128172264        1.6021533                    72.30250
Dim.25 0.126784406        1.5848051                    73.88731
Dim.26 0.125466284        1.5683286                    75.45564
Dim.27 0.122243261        1.5280408                    76.98368
Dim.28 0.120047387        1.5005923                    78.48427
Dim.29 0.118091948        1.4761494                    79.96042
Dim.30 0.116685664        1.4585708                    81.41899
Dim.31 0.112665782        1.4083223                    82.82731
Dim.32 0.110994321        1.3874290                    84.21474
Dim.33 0.109269474        1.3658684                    85.58061
Dim.34 0.103955147        1.2994393                    86.88005
Dim.35 0.098927226        1.2365903                    88.11664
Dim.36 0.094536944        1.1817118                    89.29835
Dim.37 0.086369645        1.0796206                    90.37797
Dim.38 0.085353458        1.0669182                    91.44489
Dim.39 0.080287442        1.0035930                    92.44848
Dim.40 0.076890449        0.9611306                    93.40961
Dim.41 0.073689729        0.9211216                    94.33074
Dim.42 0.072202471        0.9025309                    95.23327
Dim.43 0.063372486        0.7921561                    96.02542
Dim.44 0.061078150        0.7634769                    96.78890
Dim.45 0.036333402        0.4541675                    97.24307
Dim.46 0.029487358        0.3685920                    97.61166
Dim.47 0.027193009        0.3399126                    97.95157
Dim.48 0.025982364        0.3247795                    98.27635
Dim.49 0.023668786        0.2958598                    98.57221
Dim.50 0.022059516        0.2757439                    98.84796
Dim.51 0.021119130        0.2639891                    99.11194
Dim.52 0.017813331        0.2226666                    99.33461
Dim.53 0.017129136        0.2141142                    99.54873
Dim.54 0.014444535        0.1805567                    99.72928
Dim.55 0.012314046        0.1539256                    99.88321
Dim.56 0.009343357        0.1167920                   100.00000
```

The cumulative variance percent reaches 100 % at the 56th dimension. It seems like there are no overwhelmingly dominant axes. There are no dimensions that explain a lot of variance. The top 10 dimensions explain 48.2 % of the variance, and the top 5 dimensions 36.7 % while top 2 they both explain roughly 11 % each.

Now I will plot the explained variance

```
fviz_eig(MCA1, choice="variance", labels=TRUE)
```

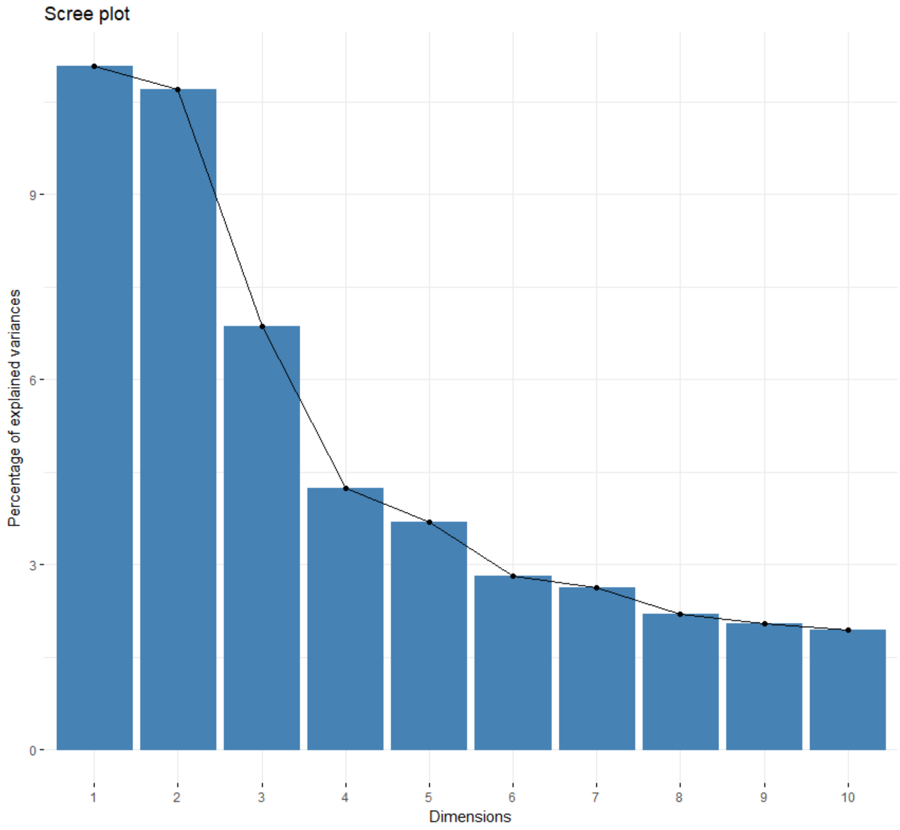
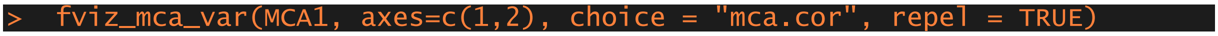


Fig. 2

The following charts is the squared correlation of coefficient of variables with regards to the first 2 dimensions.



```
> fviz_mca_var(MCA1, axes=c(1,2), choice = "mca.cor", repel = TRUE)
```

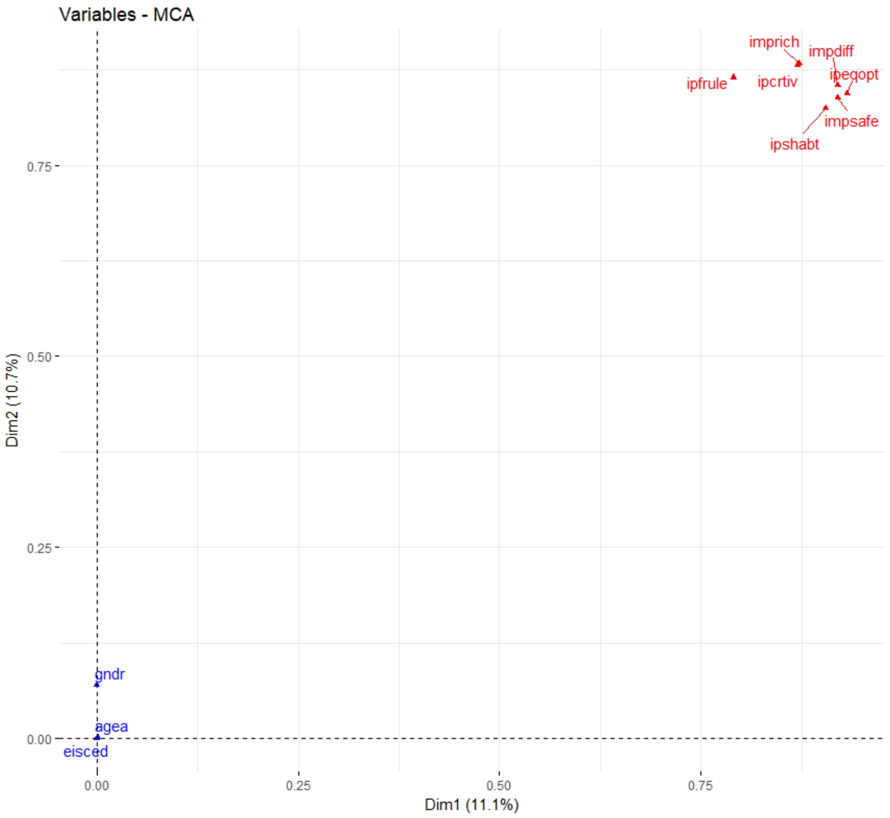
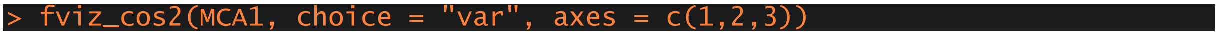


Fig.3

It is possible to see how there is a very low correlation between the supplementary variables in blue and the first two dimensions. While all the variables selected from the first question have a high correlation.

The next visualization is the squared cosine. It shows the quality of representation of variable categories on the first 2 axes which are the ones that explain more variance before it starts declining too much.



```
> fviz_cos2(MCA1, choice = "var", axes = c(1,2,3))
```

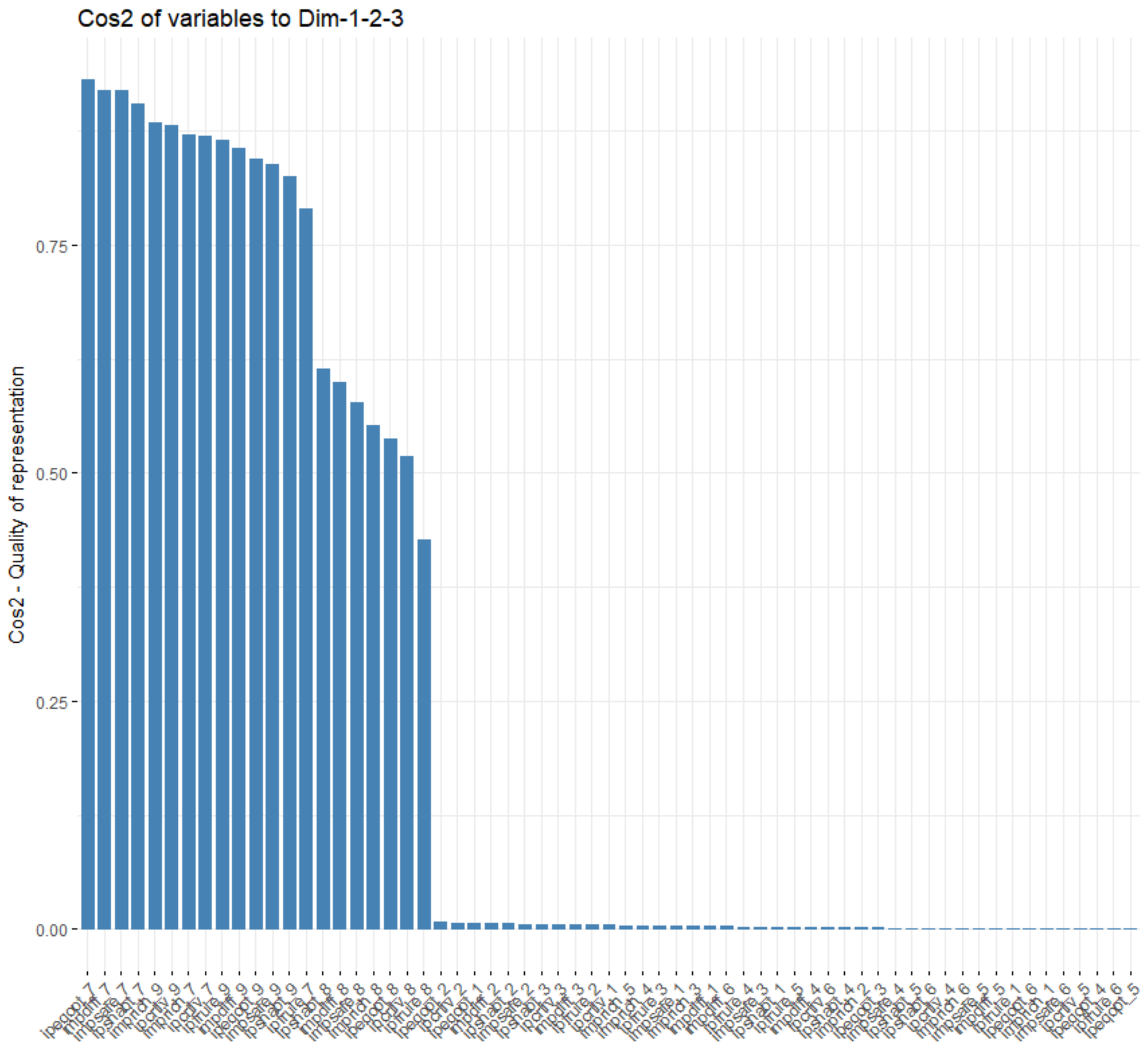


Fig.4

What can be understood from this is that the coefficients for some variables are drastically higher than the rest of the variables.

Another way of visualizing it:

```
> fviz_mca_var(MCA1, axes=c(1,2), col.var = "cos2", repel = TRUE)
```

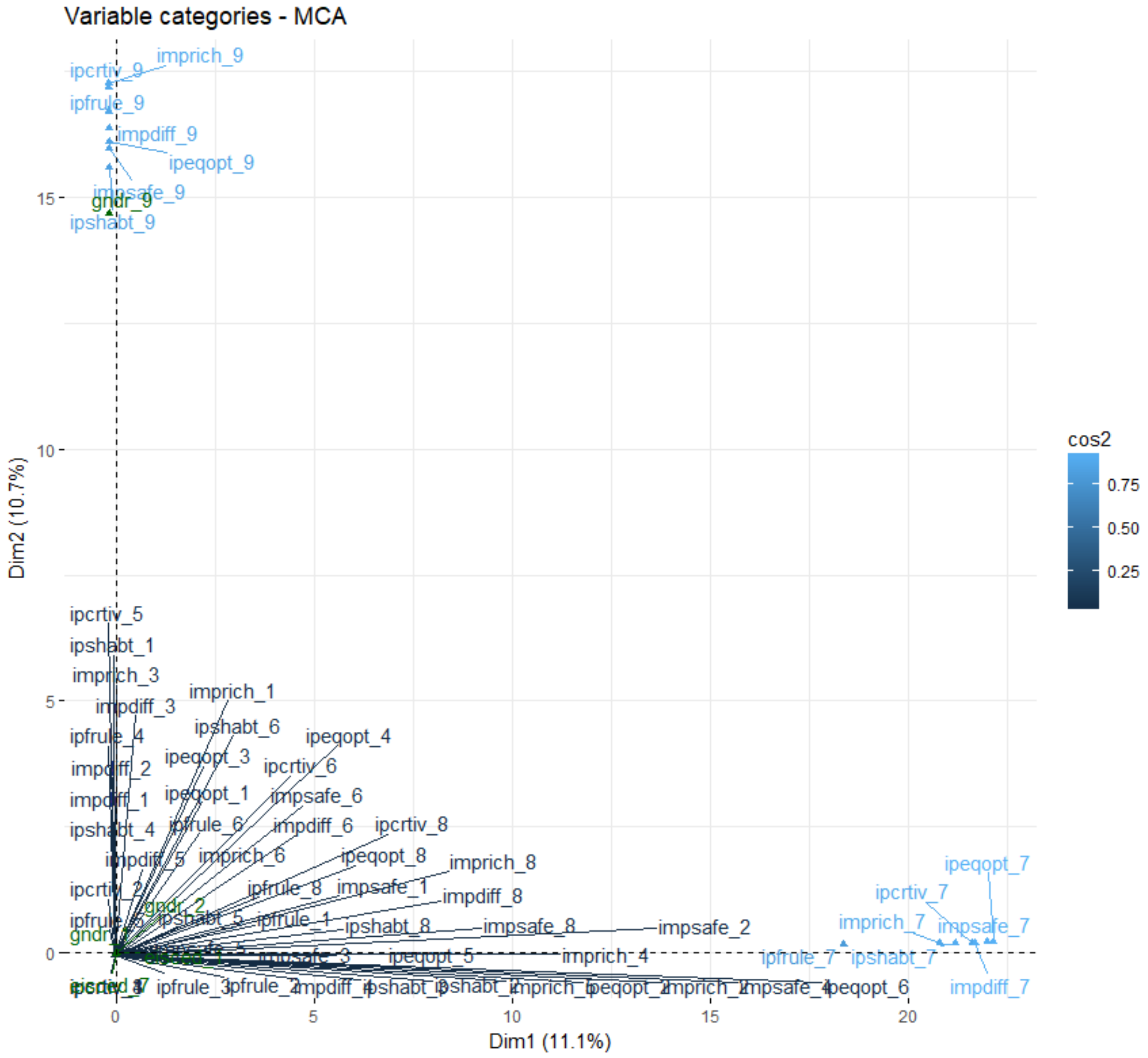


Fig.5

Now I will compute the point clouds of individuals.

```
> fviz_mca_ind(MCA1, col.ind = "cos2", axes=c(1,2), geom=c("point"), repel
=TRUE)
```

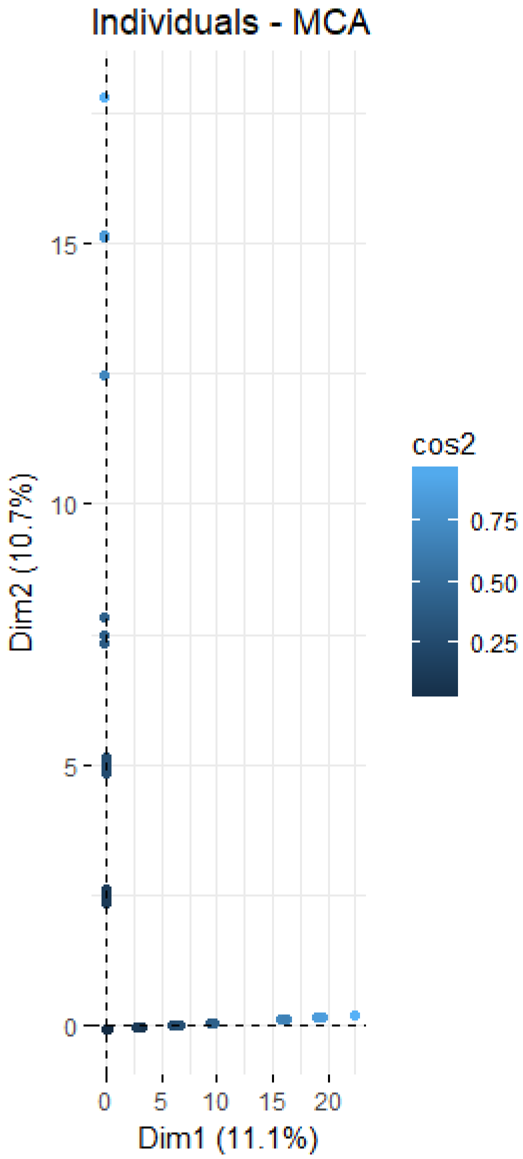


Fig.6

As we can see the points follow very closely the axis, with lower squared cosine values being closer to the origin and higher squared cosine values being further. The points are also fairly spaced between each other, the furthest ones could be considered outliers. Specific groups do not seem evident.

I am going to explore the coordinates of the variable categories on dimension 1 and 2

```
> fviz_mca_var(MCA1, axes=c(1,2), geom=c("point", "text"), repel=TRUE
```

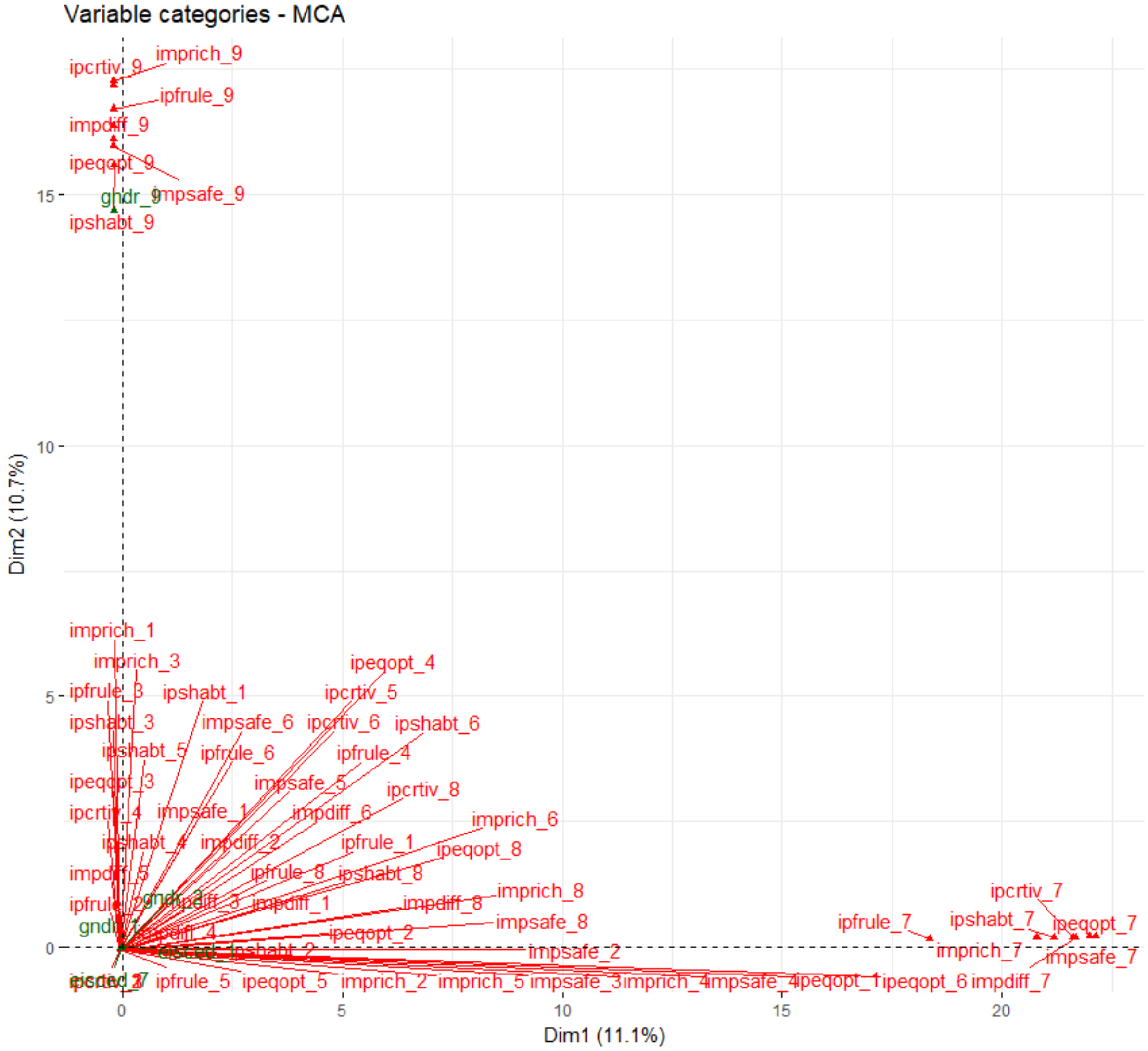


Fig.7

It is possible to see how all the response 9 (no answer) and all the response 7 (refusal) have their own cluster, while all the other responses are squished close to the origin.

Clustering

The next analysis to carry out is the hierarchical clustering of the individuals with 4 clusters. As I ran into memory allocation issues I am randomly slicing the dataset to only contain 10 % of the observations.

```
> res.hcpc<-HCPC(MCA1, nb.clust=4, graph = FALSE)
Error: cannot allocate vector of size 10.1 Gb
```

therefore, reducing the entries in the dataframe:

```
> set.seed(123)
> data.cluster= sample.int(n = nrow(data.MCA.1), size = floor(.10*nrow(dat
a.MCA.1)), replace = F)
> data.MCA.2<- data.MCA.1[data.cluster, ]
> dim(data.MCA.2)
[1] 5217   10
> MCA2=MCA(data.MCA.2, ncp = 5, ind.sup = NULL, quanti.sup = 2 ,quali.sup
= c(1,3), excl=NULL, graph = TRUE)
> var2<-get_mca_var(MCA1)
```

Now, 4-clusters hierarchical clustering

```
> MCA3.hcpc<-HCPC(MCA2, nb.clust=4, graph = FALSE)
```

and the plot:

```
> plot.HCPC(MCA3.hcpc, axes=c(1,2), choice="map")
```

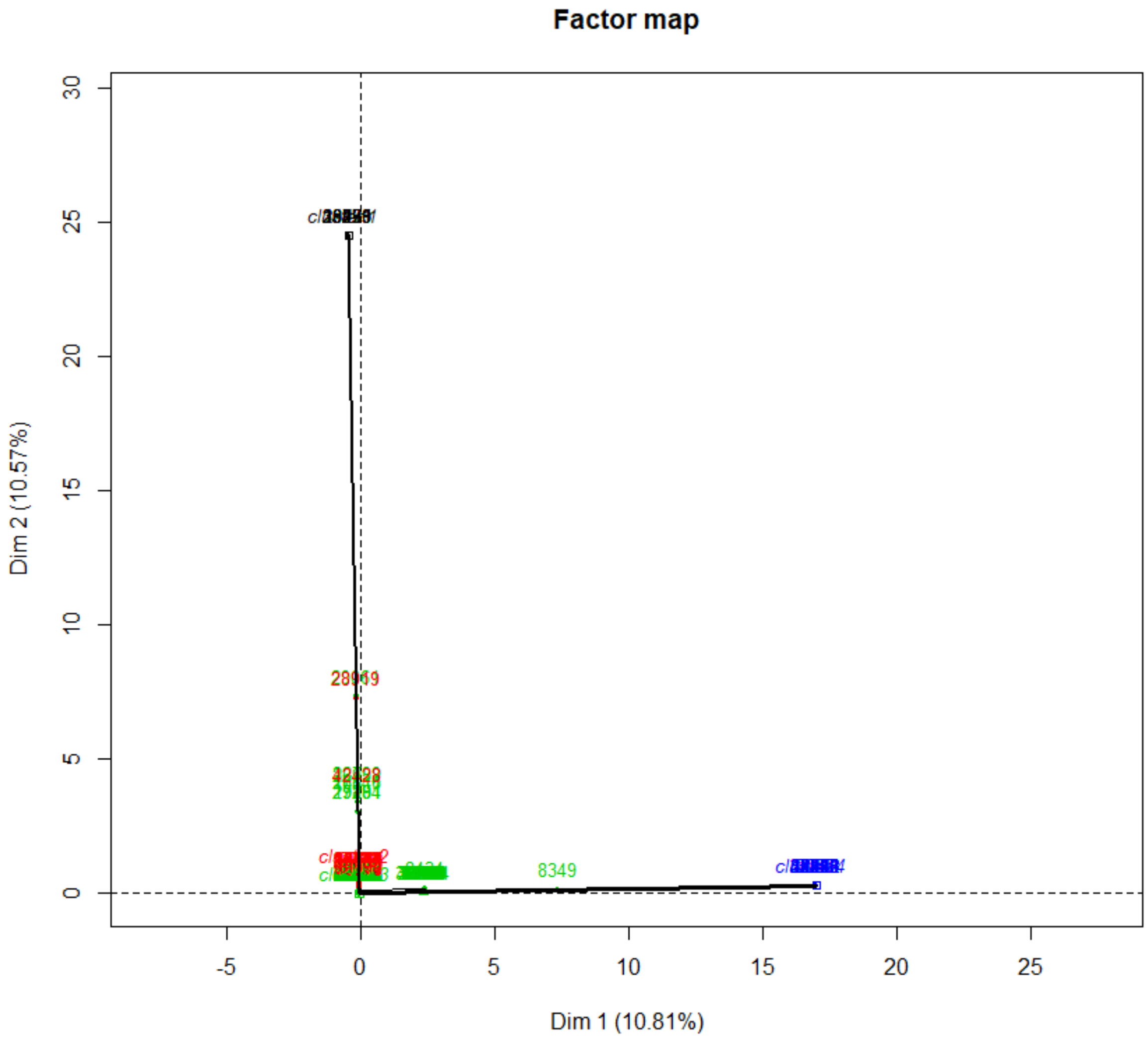
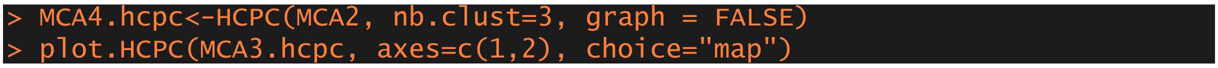


Fig.8

Clusters very sharply separated beside a few outsiders.

Trying again with three clusters because the green and red groups do not seem well separated.



```
> MCA4.hcpc<-HCPC(MCA2, nb.clust=3, graph = FALSE)
> plot.HCPC(MCA3.hcpc, axes=c(1,2), choice="map")
```

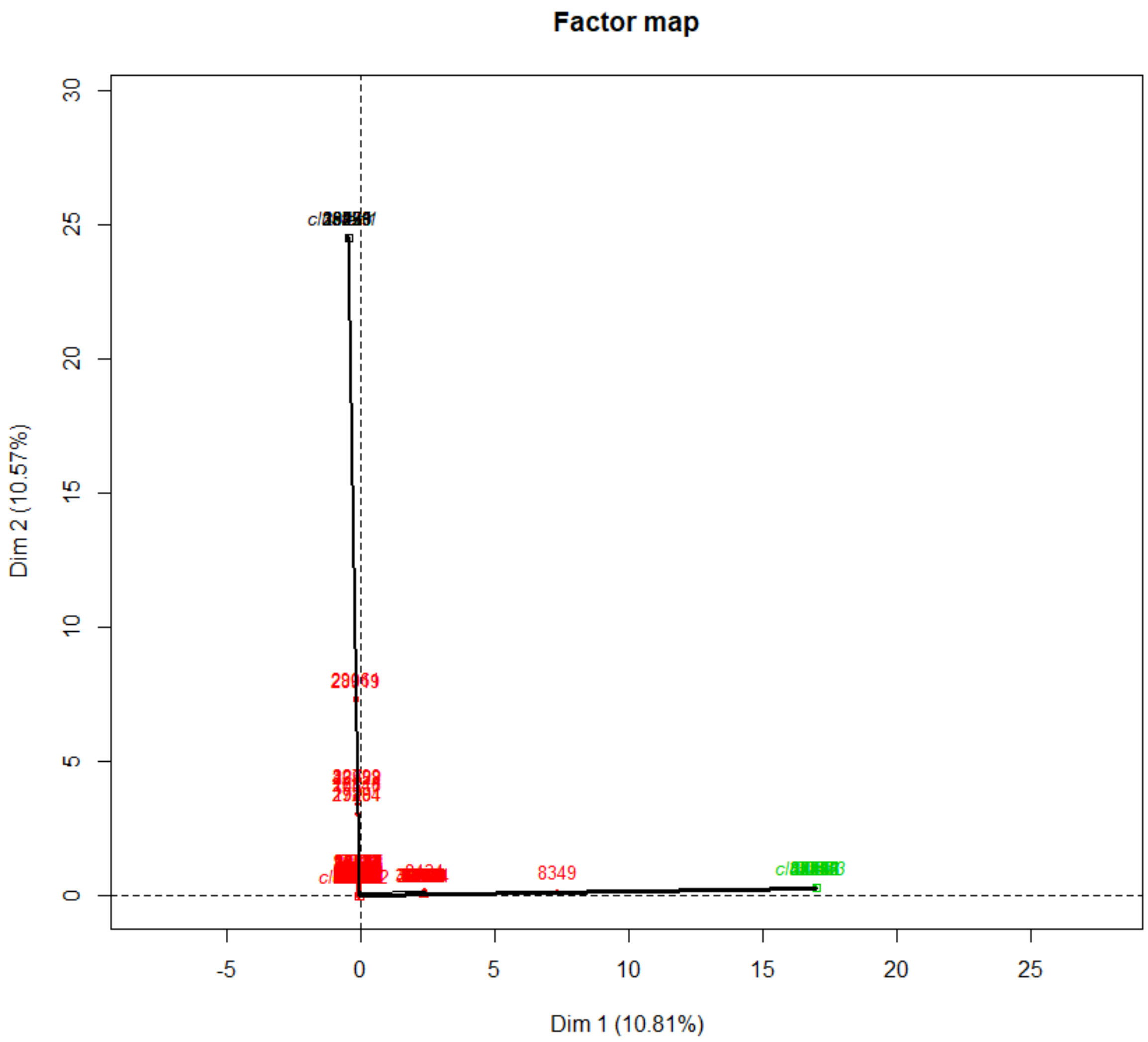


Fig.9

As expected the three clusters are very well defined now and this confirms the clear separations of the group. The red cluster has a few outsiders but green and black are very concentrated

Conclusion

From Fig.7 it is possible to conclude that people that didn’t want to respond to some questions also didn’t want to respond to other questions; which suggests that response rate is more about the person and personality rather than the question itself. We see two clusters, ‘no answer’ and ‘refusal’. More qualitative research is necessary to investigate the reasons behind this trend. This observation is reflected in other plots, especially Fig.5. Moreover, the clear separation of the data is also discernible from the two clustering visualisations.

Another observation could be the fact that 'no answer' in the contextual variable gender is connected with no answer with the other questions. There may be some social science insight in this trend, which could be investigated with more quantitative research based on in-depth qualitative interviews or focus groups.

One more possible reason for the non-response is the Hawthorne effect: since we are dealing with obtrusive data (the researcher had contact with the respondent) the behaviour of the respondent could be altered. Maybe some of the question alienated the subject into not responding multiple questions.

A limitation of the sampling method of this research could be the fact that repeated cross-sectional sampling measures different respondents at different time frames, maybe a longitudinal study with the same people could be examined as well to see if the trends are the same.

Future perspectives

Creating new ways for social research to be carried out anonymously and seamlessly. Maybe directly asking questions is not the best way to gather insightful data. The new technological and constantly connected era we are living in allows for analysing web data and the 'digital trace' from social media, Internet of Things and general internet usage.

We could go as far as saying that an in-person survey lacks the objectivity of quantitative data and it lacks the depth of qualitative data.

Nevertheless, the ESS is a massive repeated survey that help with insights into the perspective on normal European citizen on socio-political aspects of Europe.

The new era mentioned above means that social scientists and data scientists can together analyse the incredible amount of data (including Big data) at our disposal in this historical time, which was very hard to come by in the past. It is up to researcher, of multiple fields, to harness the power of this widely available data and the widely available technologies that allow us to analyse it (Ackland,2013) (Egger, Stuetzer and Welker, 2018).

It is a period where anyone could make their contribution to long-standing problems of any field and build new applications/models. Areas like business, marketing, sales, social sciences, sanitary system among many are about to be revolutionized by this new digital society. Understanding social behavioural phenomena is going to lead us into the application of tailored products and services to every citizen (Kotler Kartajaya and Setiawan, 2017). The digital society already provides a lot of data which can be extremely expanded by the concept of smart city aided by the internet of things. A smart city would supplement all sort of quality (potentially unobtrusive) social data to make the quality of life skyrocket but without compromising too much of the individuals' privacy.

Bibliography

Ackland, R., 2013. Web social science: concepts, data and tools for social scientists in the digital age. London: Sage

Egger, M., Stuetzer, C.M. and Welker, M., eds. 2018. Computational Social Science in the Age of Big Data Concepts, Methodologies, Tools, and Applications. Köln, GERMANY: Herbert von Halem Verlag

Kotler, P., Kartajaya, H. and Setiawan, I., 2017. Marketing 4.0 moving from traditional to digital. Hoboken, New Jersey: Wiley

Murtagh F. and Heck A. (1987), Multivariate Data Analysis. D. Reidel Publishing Company

Variable description taken from:
http://www.europeansocialsurvey.org/docs/round6/survey/ESS6_appendix_a7_e02_2.pdf