Computational Social Science Exam - 28.06.2021

Luisa Lo Presti

INTRODUCTION

In recent years there have been a great discussion about the level of equality and inclusion in our society. In sociology, in particular in the field of gender studies, but also in demography, several authors referred to the so-called gender revolution as “incomplete” (Esping-Andersen, 2009), “stalled” (England, 2010) or “unfinished” (Gerson, 2009). Therefore, in modern societies there is something lacking that does not allow to achieve full equality. In the context of this paper, the different steps of the gender revolution theory will be explored and, in light of this theorical framework, the level of equality achieved in two different European societies will be analyzed. This process will be carried out using data from the European Value Study (EVS), a longitudinal survey repeated every nine years in Europe. The EVS is performed since 1981 on representative samples of the population of each country and the last data collection, which is the one considered in this paper, has been made in 2017. A country-based latent class analysis will be performed on a subset of these questions in order to better understand the perception of the role of women in society, family, work and education.

PREVIOUS STUDIES AND RESEARCH QUESTION

The gender revolution theory, which focuses on three kinds of balances, has been developed in order to explain the economic and demographic consequences of the greater equality achieved by women after their entry into the labor market. Indeed, the theory recognizes an improvement in the rights of women: greater equality has been achieved in the public sphere thanks to widespread higher education and the possibility to enter the job market. However, this step forward has not concerned all aspects of life: first of all, public equality is not completed in sectors such as political representation and equal pay. Moreover, on the private sphere women have to face most of the family workload on their own. Research showed that women spend way more time in doing housework and taking care of their children than men, even when they are the breadwinner. Some scholars believe that the traditional roles that still rule in the private sphere are a consequence of the incoherence between the new gained role of women in society and the strong social norms that view women as mothers and men as workers. This particular topic has been covered by a specific question in the EVS and will be analyzed in a while. All in all, if we take into account the history toward gender equality, we will notice three macro-phases: in the first one, the lack of equality confine women to the role of mothers and wives. This is considered as a stable balance because the same beliefs and views are shared in the public and private sphere. In the following phase, women acquire more rights in the public sphere, which are not met by increasing equality in the private one: this is an unstable balance because it leads women to deal with full-times jobs and almost all housework, making it very hard to find a compound between the two things. According to most demographic studies, there is a high correlation between this unbalance and the decreasing fertility rate: having children becomes a sort of opportunity-cost factor to consider for women. Finally, a new stable balance will be reached when equality will be achieved in the private sphere, generating exchangeable working and parental roles within the household. This last phase is to consider a future achievement, even though some countries, such as the Scandinavian ones, are farther ahead in reaching this goal than others. Noteworthy steps to be taken to move in this direction are public policies in support of interchangeable parental roles, equal parental leaves, and educational work. Considering this framework, what the following analysis aim to do is to recognize this evolution pattern in the selected European countries and to reflect on the population values concerning the role of women in society, family, and life in general. Gender equality can be thought as a sort of latent variable buildup of several aspects (mainly concerning political representation and equality, job market and working life, family life, and education): by looking at relevant questions in the EVS 2017, the aim is to identify the main groups of people sharing similar views and defining what kind of achievements and gaps are present and reinforced by people's beliefs. Using latent class analysis (LCA), three groups of people will be outlined depending on their opinions and views on selected statements: based on their beliefs, it will be possible to define profiles of families and households characterized by different levels of gender equality and to contextualize these classes in the above-mentioned theory.

METHODOLOGY

Dataset

The EVS is a cross-national and longitudinal survey carried out in an increasing number of countries all over Europe and it concerns beliefs, ideas, values, and opinions of the citizens about several topics, which can be classified in six main areas: life, family, work, religion, politics, and society. The project started in 1981 and the last data collection took place in 2017 and it led to the confirmation of the profound changes that are occurring, with different speeds and characteristics, in all modern societies. It goes without saying that most cultural and social changes seem to depend upon several socio-economic and historical factors, which are clearly different in the various European states. Considering the last survey available, the analysis will focus mainly on questions taken from the "family" macro-area. Nowadays, the traditional model (male breadwinner – female housekeeper) is less and less pervasive, yet most of the European population seems to be conservative in values and ideas concerning family life. Europeans believe that both a father and a mother are needed for a child to be happy and generally the best and preferred choice is following the traditional family pattern. Given the research topic, only certain questions within the family demands will be analysed, while few statements regarding the role of women will be taken from other areas. In details, the questions that have been selected from the EVS 2017 are the following:

Q25) Do you strongly agree, agree, disagree, or strongly disagree?
v72) When a mother works for pay, the children suffer.
v73) A job is alright but what most women really want is a home and children.
v74) All in all, family life suffers when the woman has a full-time job.
v75) A man's job is to earn money; a woman's job is to look after the home and family.
v76) On the whole, men make better political leaders than women do.
v77) A university education is more important for a boy than for a girl.
v78) On the whole, men make better business executives than women do.

Q26) Do you strongly agree, agree, neither agree not disagree, disagree, or strongly disagree?
v81) When jobs are scarce, men have more right to a job than women.

Q27) Do you strongly agree, agree, neither agree not disagree, disagree, or strongly disagree?
v83) It is a duty towards society to have children.

Q39) Many things are desirable, but not all of them are essential characteristics of democracy. Please tell me for each of the following things how essential you think it is as a characteristic of democracy. Use this scale where 1 means "not at all an essential characteristic of democracy" and 10 means it definitely is "an essential characteristic of democracy".
v141) Women have the same rights as men.

The answers to the first questions are encoded in a Likert scale, thanks to which respondents specify their level of agreement or disagreement to a statement on a symmetric agree-disagree scale. The idea of this scale is to capture how strong is the feeling, idea, or opinion toward some given aspect of life. When studying a latent phenomenon, it is common to deal with a set of partially correlated statements, which show that there is consistency in the questions chosen, thus their ability to describe a multifaced aspect of life is validated. For this reason, correlations have been considered at the beginning of the analysis. The typical Likert scale involves five levels of agreements (strongly agree, agree, neither agree nor disagree, disagree, and strongly disagree), however variations are common as well: in the first question (Q25), respondents had only four levels available to answer, taking away the neutral position. The first questions deal with the role of women in business, politics, education, but mostly in family and in relation to children. Indeed, the studies of the EVS proved that European families are "children-centered". The last question was selected from a different macro-area of the EVS 2017, but it was retained noteworthy, even if it deals with the idea of democracy. Indeed, people tend to associate democracy with the idea of justice and a positive regime, and this is confirmed empirically by the fact that even dictatorships tend to legitimize their regimes through a semblance of democracy or through the choice of words in speeches. Therefore, asking if it is democratic for women to have the same rights as men is indirectly related to questioning the desirability and justice of the statement, which could be used as a basic factor involved in the political emancipation.

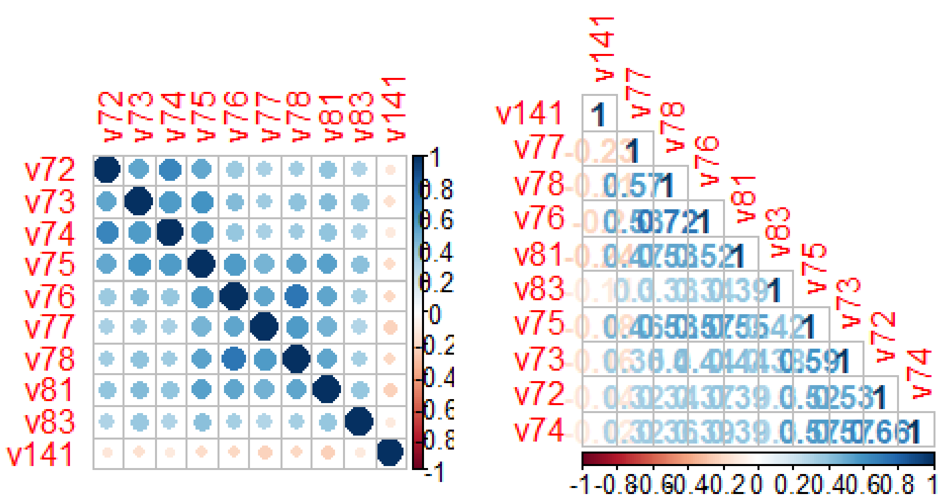
PCA and LDA

The aim of this project is to define latent classes based on people beliefs: groups would describe one of the three levels of emancipation in the context of the gender revolution theory and outline which is the most common opinions pattern in the selected societies. Firstly, since Europe is very diverse and a single analysis on a European base would be less meaningful, two countries, Hungary and Sweden, were selected. The selection was carried out evaluating cultural, historical and social differences, in order to explore two very different realities. Before implementing the LCA, for each of the selected country a fast principal component analysis (PCA) was computed. The reason for this is that it helps finding the variables with highest variance, thus the statements that obtained very different answers; this was useful since it gave a first understanding of the most divisive and interesting factors to take into account later in the LCA. PCA can be used in this way since it is a dimensionality reduction technique which creates principal components made of variables and their loading factors, with the aim of maximizing the amount of variance explained in the first component. As mentioned, this paper will use LCA to outline different profiles of people based on their beliefs and ideas concerning the role of women in various aspects of life. The groups that will be found will be useful to understand how much the emancipation of women permeate the beliefs of people. Indeed, most people would say they support gender equality, but they may not recognize certain ideas they have undermine this process: for this reason, using multiple questions as indicators is appropriate. LCA provides a powerful, yet very flexible, approach to the analysis of categorical data. Through LCA, latent categorical variables (by definition not directly observable) are measured by means of several observed categorical indicators. LCA is a type of model-based cluster analysis founded on the idea of soft or probabilistic clustering; this means that, for each observation, it computes probabilities of belonging to each class and then assigns it to the class with the highest probability, following the expectation-maximization algorithm. LCA aims to identify patterns of responses to create classes of individuals sharing similar and peculiar characteristics. This means that it minimizes the difference (variance) within the class and maximize the difference among classes. It is a person-oriented approach designed for segmentation and profiling and it has been preferred over traditional clustering methods, such as k-means or hierarchical clustering, for several reasons. Firstly, it works better with categorical variables, and it suits surveys analysis; then, being it model-based, it is not distance dependent; finally, it is less sensible to scale and, belonging to soft clustering methods, it is very flexible, allowing membership to multiple clusters with different probabilities.

ANALYSIS AND INTERPRETATION OF RESULTS

```
#Load the dataset and subset it to get only the variables needed for this analysis
.

library(haven)

path = file.path("C:/Users/luisa/Documents/ComputationalSocialScience/EuropeanValu
esSurvey_2017/ZA7500_v4-0-0.sav")
dataset = read_sav(path)

dfr<- subset(dataset, select = c(v72, v73, v74, v75, v76, v77, v78, v81, v83, v141
, c_abrv))

#Pre-processing step:
#remove NAs and codify all variables so that the answers appear in the same scale
1-4.

anyNA(dfr)

## [1] TRUE

#remove rows with NA
dfr <- na.omit(dfr)

#For the purpose of this analysis we will consider "neither agree nor disagree" as
#NA and DK (don't know), which were already removed from the dataframe.
dfr <- dfr[dfr$v81 != 3, ]
dfr <- dfr[dfr$v83 != 3, ]
dfr$v81[dfr$v81 == 4] <- 3
dfr$v81[dfr$v81 == 5] <- 4
dfr$v83[dfr$v83 == 4] <- 3
dfr$v83[dfr$v83 == 5] <- 4
```

```
#v141 is codified on a scale from 1 to 10, showing different level of agreement or
disagreement.
#The answers will be re-coded on a scale 1-4, as all the previous ones.
dfr$v141 <- ifelse(dfr$v141 <= 2, "strongly disagree",
                   ifelse(dfr$v141>2 & dfr$v141<=5, "disagree",
                          ifelse(dfr$v141>5 & dfr$v141<9, "agree", "strongly agree
")))

dfr$v141 <- ifelse(dfr$v141 == "strongly disagree", 4,
                   ifelse(dfr$v141 == "disagree", 3,
                          ifelse(dfr$v141 == "agree", 2, 1)))

#EXPLORATORY DATA ANALYSIS AND VISUALIZATION

#CORRELATION PLOT:
#Create a correlation matrix and plot it to explore the data.

library("corrplot")

## corrplot 0.84 loaded

df.only.num <- subset(dfr, select = -c(c_abrv))
cor.matrix <- cor(df.only.num, use="complete", method = "pearson")
par(mfrow=c(1,2))
corrplot(cor.matrix, addrect = 2)
corrplot(cor.matrix, method = "number", order = "AOE", type="lower")
```

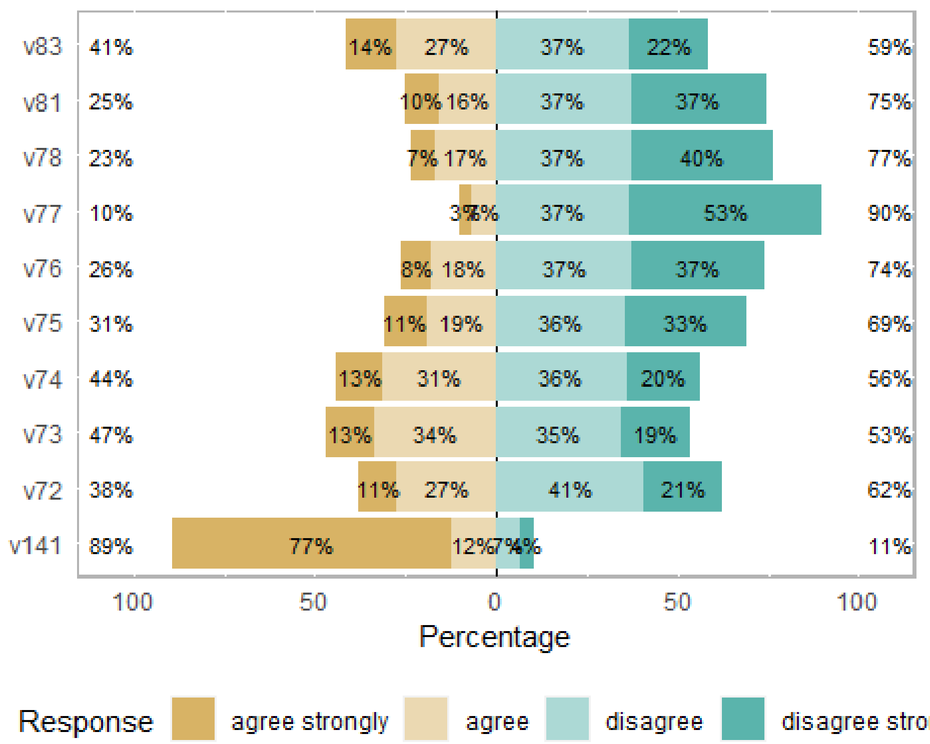


The correlation plots suggest that there is a pretty high correlation among all variables but v141. This is probably due to the fact that this statement relates to a really basic principle that is considered to be well-established, namely the principle of equal rights between women and men in a democracy, thus not much divergence between different groups of people is found. Basically, it means that there is almost no relation between the opinions concerning the other statements and this one. Nevertheless, since the purpose of the analysis is to analyze the steps toward gender equality, this variable was kept as an indicator of one of the first steps, which, in the framework of the gender revolution theory, can be conduct at the very beginning of the first phase.

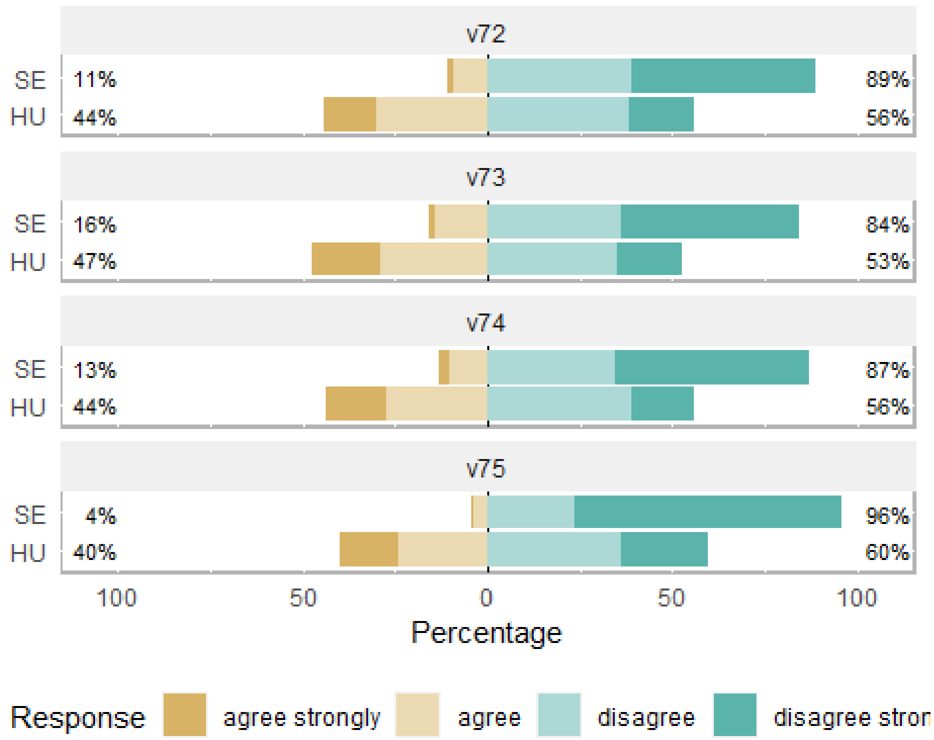
As a first exploration of the data, one may want to visualize the answers graphically to have a first clear understanding of the beliefs in Europe and in the particular countries selected.

```
#Visualize the levels of agreement and disagreement in Europe as a whole and then
the different response patterns in the selected countries.

library("dplyr")

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library("likert")

## Warning: package 'likert' was built under R version 4.0.5

## Loading required package: ggplot2

## Loading required package: xtable

##
## Attaching package: 'likert'

## The following object is masked from 'package:dplyr':
##
##     recode

library("plyr")

## ------------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dp
lyr:
## library(plyr); library(dplyr)

## ------------------------------------------------------------------------------
```

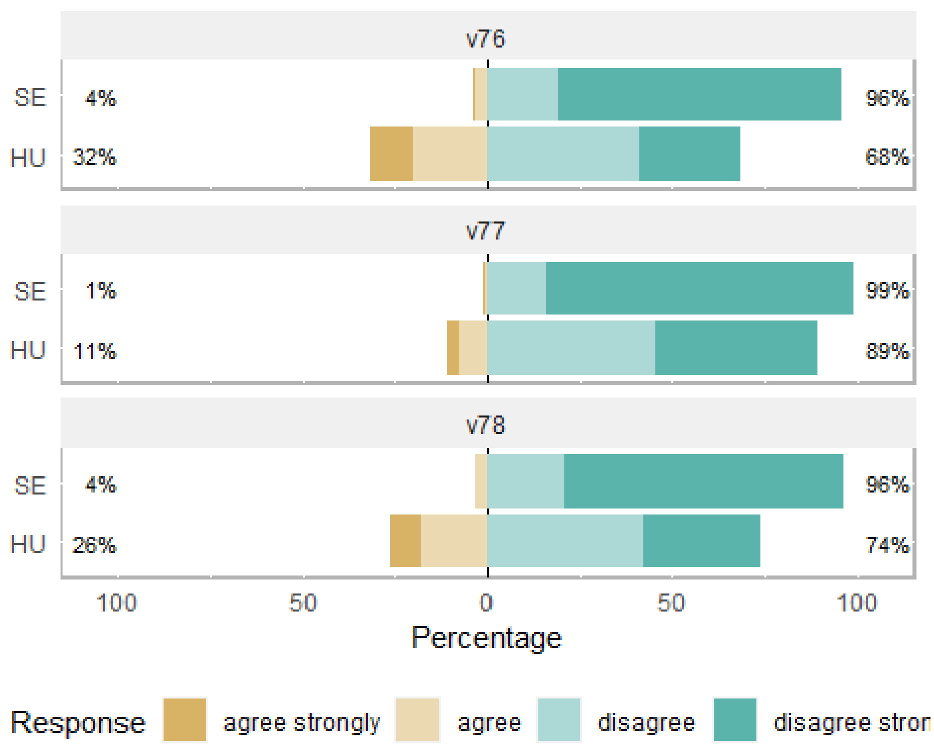
```
##
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

#Answers collected in all European countries are involved in this visualization.
lev <- c("agree strongly", "agree", "disagree", "disagree strongly")
survey <- dfr %>%
  dplyr::mutate_if(is.character, factor) %>%
  dplyr::mutate_if(is.numeric, factor, levels = 1:4, labels = lev) %>%
  na.omit() %>%
  as.data.frame()

plot(likert(survey[,1:10]), plot.percents = TRUE, ordered = FALSE, wrap = 60)
```



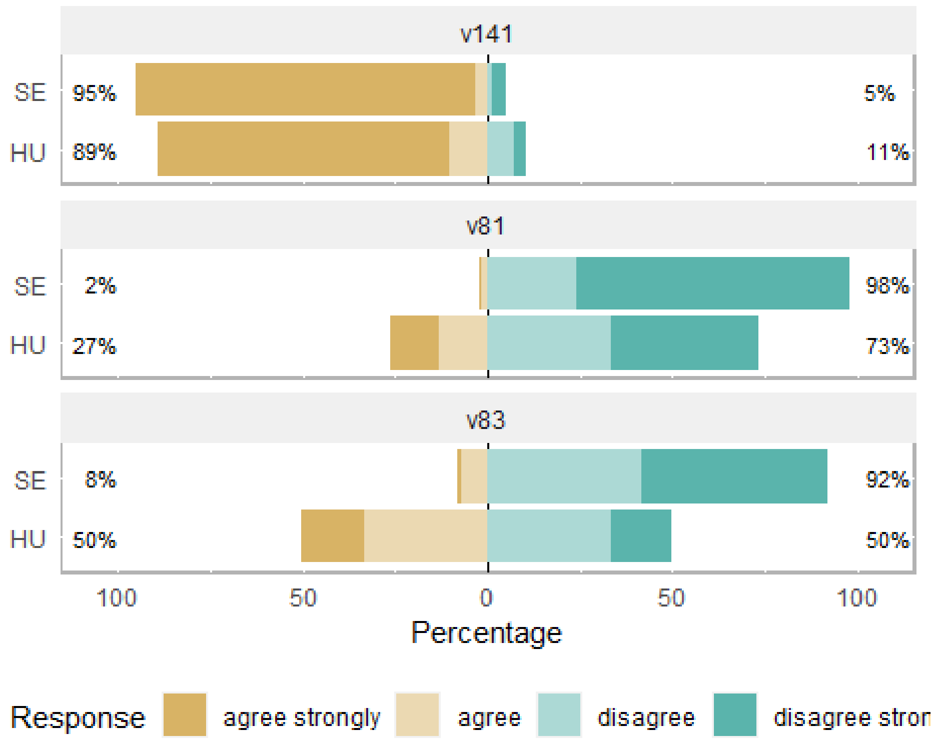

```
#Answers on a country base:
df.cntr <- dfr[dfr$c_abrv == 'HU' | dfr$c_abrv == 'SE',]

survey.cntr <- df.cntr %>%
  dplyr::mutate_if(is.character, factor) %>%
  dplyr::mutate_if(is.numeric, factor, levels = 1:4, labels = lev) %>%
  na.omit() %>%
  as.data.frame()

plot(likert(survey.cntr[,1:4], grouping = survey.cntr$c_abrv))
```




```
plot(likert(survey.cntr[,5:7], grouping = survey.cntr$c_abrv))
```




```
plot(likert(survey.cntr[,8:10], grouping = survey.cntr$c_abrv))
```



The plots above show different patterns of answers in the two countries: I will not go into details in the description of the plots, since a more detailed analysis will be done later using the LCA; however, clear differences among opinions in the two countries are shown.

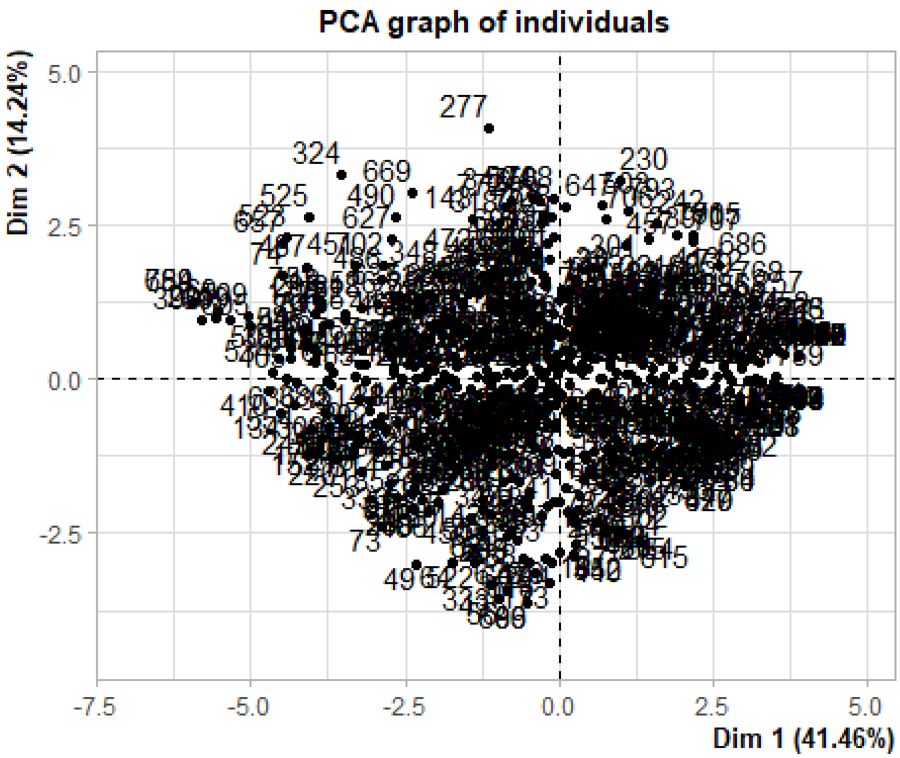
After this initial exploration, it is possible to precede in the actual analysis.

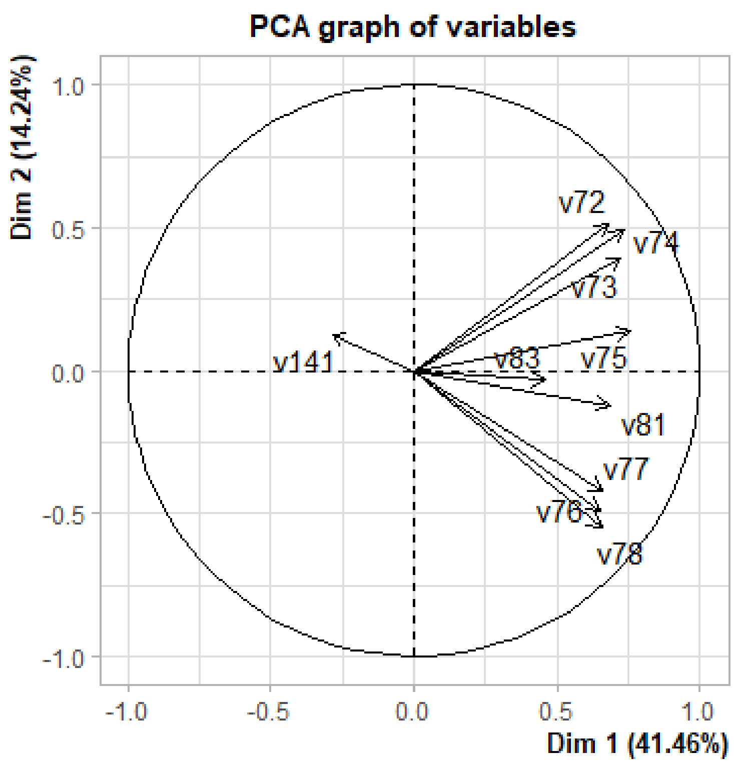
```
#Create single dataframes for each country to perform LCA:
hu <- dfr[dfr$c_abrv == 'HU',]
se <- dfr[dfr$c_abrv == 'SE',]
```

Firstly, let's take into consideration Hungary. Before looking into LCA, we may want to understand which are the most important variables among the one chosen. Indeed, there may be variables, which value do not variate much between different categories of people. Thanks to PCA, it is possible to identify variables that explain the highest variation, meaning statements that are divisive and therefore cause different groups of people to have diverging opinions toward these affirmations. These statements are the most interesting to analyze and the one that actually characterize the different profiles of people.

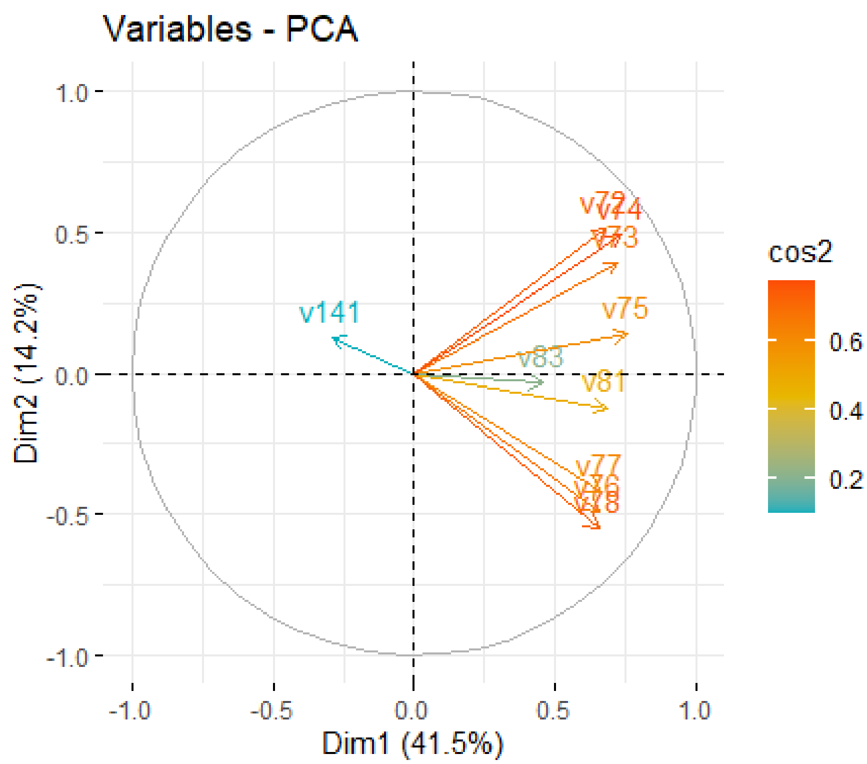
```
#HUNGARY

#PCA

library(FactoMineR)

## Warning: package 'FactoMineR' was built under R version 4.0.5

library(factoextra)
```

```
## Warning: package 'factoextra' was built under R version 4.0.5

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl
/ve3WBa

hu <- subset(hu, select = - c(c_abrv))
PCanalysis <- PCA(hu)
```



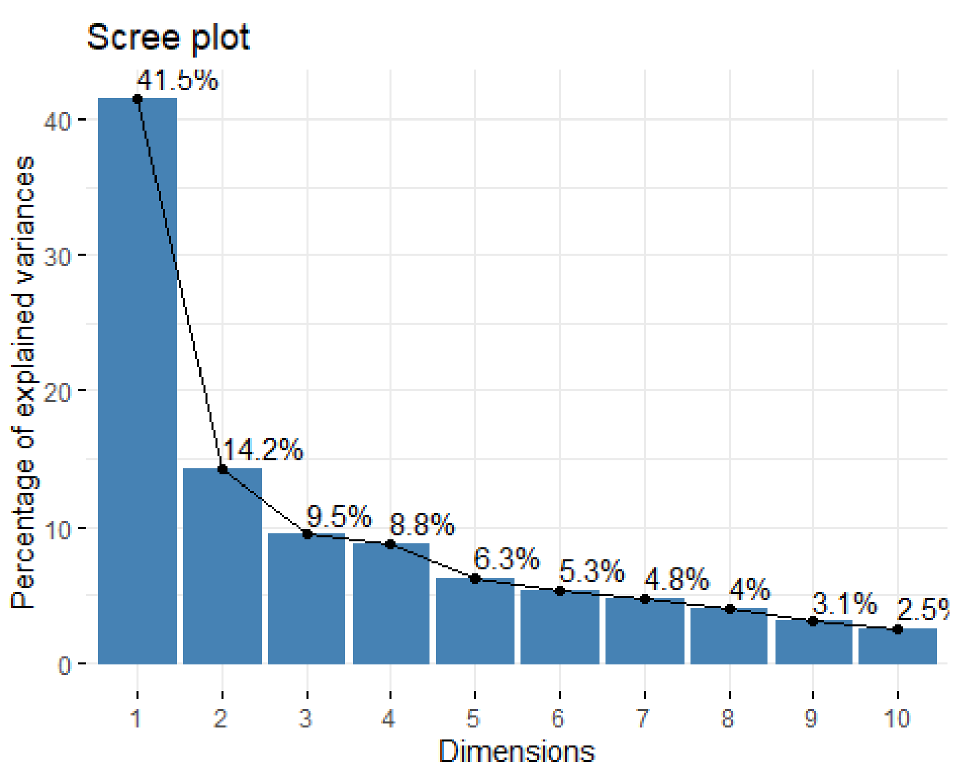



```
#Follows a different visualization of the same plot: it outlines better the import
ance of variables, coloring in red the most important ones and in green/blue the o
ne contributing the less to explain the variance.
fviz_pca_var(PCanalysis, col.var='cos2', gradient.cols=c('#00afbb', '#e7b800', '#f
c4e07'))
```



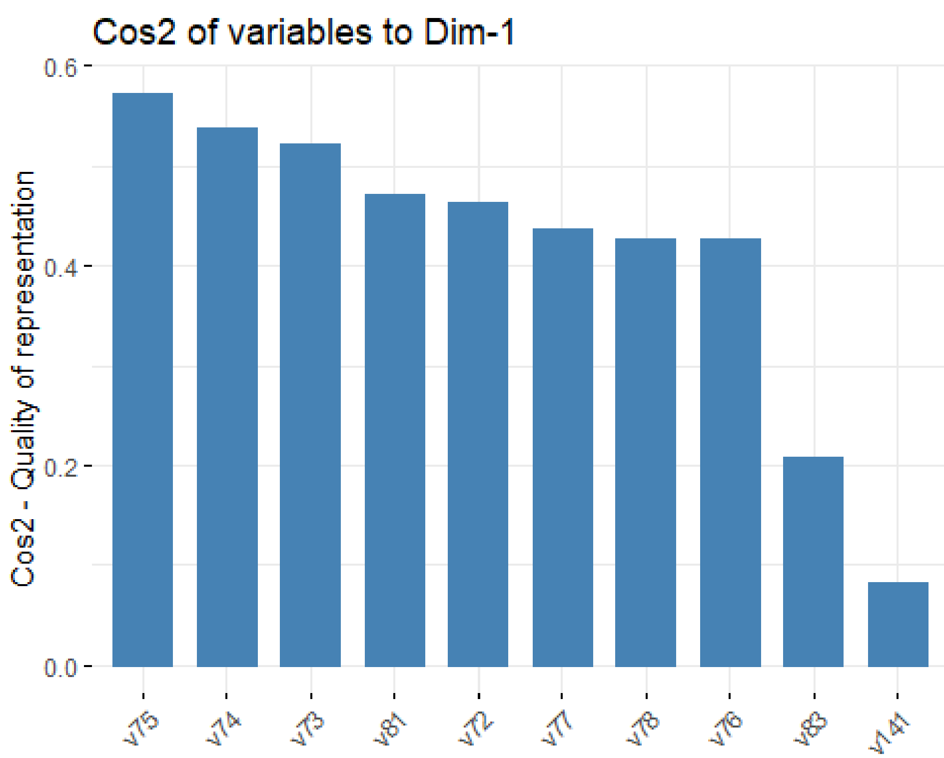

```
summary(PCanalysis)

##
## Call:
## PCA(X = hu)
##
##
## Eigenvalues
##                        Dim.1   Dim.2   Dim.3   Dim.4   Dim.5   Dim.6   Dim.7
## Variance               4.146   1.424   0.946   0.884   0.630   0.534   0.479
## % of var.             41.463  14.239   9.459   8.842   6.303   5.344   4.790
## Cumulative % of var.  41.463  55.701  65.160  74.003  80.305  85.649  90.440
##                        Dim.8   Dim.9  Dim.10
## Variance               0.398   0.306   0.252
## % of var.              3.977   3.065   2.519
## Cumulative % of var.  94.416  97.481 100.000
##
## Individuals (the 10 first)
##              Dist    Dim.1    ctr   cos2    Dim.2    ctr   cos2    Dim.3    ctr
## 1        |  2.796 | -2.555  0.198  0.835 |  0.010  0.000  0.000 |  0.342  0.016
## 2        |  3.867 |  2.940  0.262  0.578 |  0.266  0.006  0.005 |  1.187  0.187
## 3        |  2.765 |  0.295  0.003  0.011 | -0.464  0.019  0.028 | -0.182  0.004
## 4        |  3.853 |  0.152  0.001  0.002 |  1.309  0.151  0.115 |  2.914  1.130
## 5        |  3.682 |  3.630  0.400  0.972 |  0.388  0.013  0.011 |  0.156  0.003
## 6        |  2.726 |  2.537  0.195  0.866 | -0.831  0.061  0.093 | -0.022  0.000
## 7        |  2.951 | -1.017  0.031  0.119 |  0.037  0.000  0.000 | -1.328  0.235
## 8        |  3.966 |  3.863  0.453  0.949 |  0.358  0.011  0.008 | -0.004  0.000
## 9        |  2.703 |  1.830  0.102  0.458 |  0.184  0.003  0.005 | -0.441  0.026
```

```
## 10    |  3.260 |  1.884  0.108  0.334 |  0.586  0.030  0.032 | -0.295  0.012
##         cos2
## 1      0.015 |
## 2      0.094 |
## 3      0.004 |
## 4      0.572 |
## 5      0.002 |
## 6      0.000 |
## 7      0.203 |
## 8      0.000 |
## 9      0.027 |
## 10     0.008 |
##
## Variables
##           Dim.1    ctr   cos2    Dim.2    ctr   cos2    Dim.3    ctr   cos2
## v72   |  0.680 11.168  0.463 |  0.512 18.402  0.262 |  0.062  0.410  0.004 |
## v73   |  0.722 12.570  0.521 |  0.391 10.729  0.153 | -0.002  0.000  0.000 |
## v74   |  0.733 12.965  0.538 |  0.490 16.830  0.240 |  0.104  1.133  0.011 |
## v75   |  0.756 13.797  0.572 |  0.137  1.319  0.019 |  0.012  0.015  0.000 |
## v76   |  0.653 10.299  0.427 | -0.491 16.909  0.241 |  0.211  4.698  0.044 |
## v77   |  0.661 10.532  0.437 | -0.420 12.371  0.176 |  0.023  0.056  0.001 |
## v78   |  0.654 10.313  0.428 | -0.549 21.204  0.302 |  0.185  3.619  0.034 |
## v81   |  0.686 11.345  0.470 | -0.122  1.049  0.015 | -0.105  1.169  0.011 |
## v83   |  0.456  5.025  0.208 | -0.035  0.086  0.001 | -0.149  2.360  0.022 |
## v141  | -0.287  1.986  0.082 |  0.125  1.101  0.016 |  0.905 86.539  0.819 |

#Scree plot: a bar chart of the eigenvalues, which represent the variances.
fviz_screeplot(PCanalysis, addlabels=TRUE, ncp=10)
```



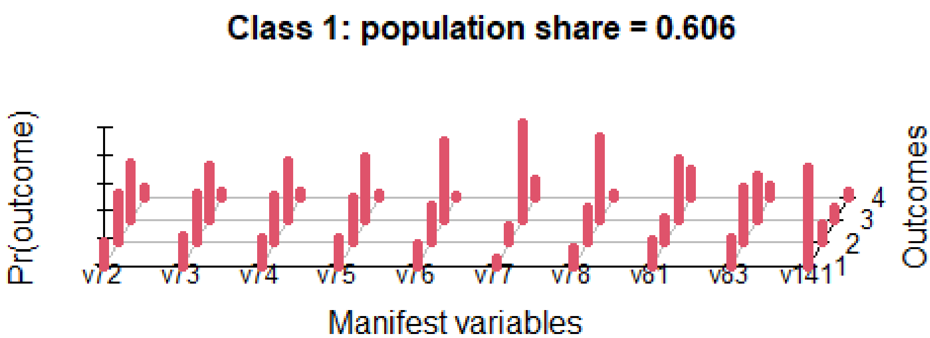
Both graphically and in the summary, it is possible to notice that with two principal components we can explain around 55.7% of the total variance. Looking at the summary, we can also identify the most important variables contributing to each component, in particular to the first one, which by definition is the one that expresses the greatest amount of variance. Variables with loading higher than 0.3 (in absolute value) are considered to have a significant contribution; the highest the loading, the more the variable is relevant. The only variable not meeting the 0.3 threshold is v141, which has loading equal to -0.28. Graphically, we can also note that it points in a completely different direction, belonging to a group by itself. This is not surprising since it was selected from a different category of statements concerning democracy and we expect, also in light of theory and on the democratic contexts we are studying, wide agreement with the necessity for women and men to have equal rights in this political framework. All variables seems to be significant but the aim is to find the most divisive aspects: it is possible to notice that the less varying response pattern appears in v141 and v83, while the statements collecting very different opinions are v73, v74, and v75, which concern the relation between women, family life and children. This seems to be coherent with the theory of the three balances, since it outlines a context in which the opinion about the role of women in the private sphere is still quite diverse, while there is some kind of agreement on the public role. The contribution of the variables to the first component can also be visualized as follows.

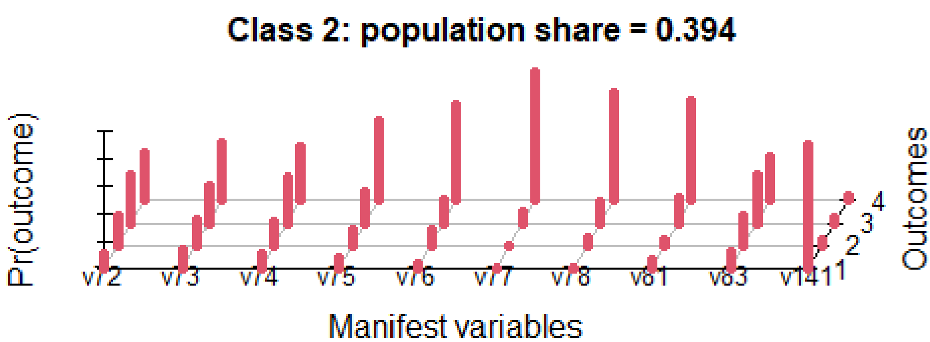
```
fviz_cos2(PCanalysis, choice='var', axes=1)
```




```
#LCA

library(poLCA)

## Warning: package 'poLCA' was built under R version 4.0.5
```

```
## Loading required package: scatterplot3d

## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked _by_ '.GlobalEnv':
##
##     survey

## The following object is masked from 'package:dplyr':
##
##     select

f <- cbind(v72, v73, v74, v75, v76, v77, v78, v81, v83, v141)~1
lca1 <- poLCA(f, hu, nclass = 2, graphs = TRUE)
```






```
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $v72
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.1702 0.3525 0.4077 0.0695
## class 2:  0.1022 0.2178 0.3434 0.3366
##
## $v73
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.2132 0.3604 0.3952 0.0312
```

```
## class 2:  0.1318 0.1897 0.2734 0.4051
##
## $v74
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.1993 0.3373 0.4294 0.0340
## class 2:  0.1117 0.1805 0.3260 0.3817
##
## $v75
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.2059 0.3227 0.4487 0.0227
## class 2:  0.0857 0.1232 0.2230 0.5681
##
## $v76
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.1611 0.2671 0.5661 0.0056
## class 2:  0.0364 0.1096 0.1735 0.6805
##
## $v77
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0526 0.1246 0.6951 0.1277
## class 2:  0.0054 0.0000 0.0774 0.9172
##
## $v78
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.1261 0.2552 0.5926 0.0261
## class 2:  0.0136 0.0641 0.1552 0.7671
##
## $v81
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.1839 0.1871 0.4309 0.1981
## class 2:  0.0589 0.0443 0.1834 0.7134
##
## $v83
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.2000 0.4021 0.3200 0.0779
## class 2:  0.1235 0.2245 0.3474 0.3047
##
## $v141
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.7177 0.1439 0.0945 0.0439
## class 2:  0.8977 0.0468 0.0303 0.0251
##
## Estimated class population shares
##  0.6057 0.3943
##
## Predicted class memberships (by modal posterior prob.)
##  0.6038 0.3962
##
## =========================================================
## Fit for 2 latent classes:
## =========================================================
```

```
## number of observations: 795
## number of estimated parameters: 61
## residual degrees of freedom: 734
## maximum log-likelihood: -8820.475
##
## AIC(2): 17762.95
## BIC(2): 18048.33
## G^2(2): 7530.083 (Likelihood ratio/deviance statistic)
## X^2(2): 8530383 (Chi-square goodness of fit)
##
```

The default for LCA is to use two classes: it is possible to see that 60.6% of the Hungarian population would belong to the first class, which outlines disagreement (even though not strong) with most of the statements. Most people belonging to this group disagree with the fact that children suffer when their mother works (v72); nevertheless, a quite high share of this group agrees: indeed, the probability of agreeing with this statement, given the belonging to class one, is around 35% and the probability of agreeing strongly is more than 17%.

Considering strong and moderate agreement and disagreement together, around 52% of people belonging to this profile actually agree with this statement, while in class two almost 68% of people disagree. This answers could be consider particularly related to the outcome of the agreement with statement v74: in class one, 53% of people agree that family life suffers if a woman works full-time, while in class two the majority (almost 71%) disagrees.

In class one, most people (around 57%) believe that what women really desire is home and children (v73: "A job is alright but what most women really want is a home and children."), while work is not an essential part of their lives; while in class two around 68% of people disagree. As far as concerns statement v75 ("A man's job is to earn money; a woman's job is to look after the home and family."), which propose a more traditional and conservative type of family, almost 53% of people belonging to class one (60,8% of the population) agree with this statement, while less than 21% of people in group two are found in agreement.

Altogether, this suggests that the majority of the Hungarian population is still very conservative and has as reference a traditional type of family with male-breadwinner and female-housekeeper. This could be a severe indicator that women having a job is seen as a secondary source of income, probably less important, with the possibility to give it up, since their main role within the family is the cure of the house and children. This undermines the reaching of equal and exchangeable parental roles and make it hard for women to enter the job market, since the greatest amount of their time is absorbed by home and children. This also shows that, in time of crisis, the one to suffer the most will be women: since their job is perceived as less relevant, it is more likely that they will lose it if a choice need to be taken. This is also shown by the answers given at question v81 ("When jobs are scarce, men have more right to a job than women."), where 37% of people in the first class agree, while the disagreement is almost total in the second class, and at question v78 ("On the whole, men make better business executives than women do."), where 39% of people belonging to the first class seems to agree against a complete disagreement in class two.

Other statements seem to be not particularly divisive, even thought the same pattern can be noticed in the two groups. Noteworthy is the fact that, despite the agreement with the statements that see men as better workers, people belonging to the first class perceive university education as important for both males and females.

The output also shows the predicted class memberships, which assign observations to each latent class using posterior probabilities. The estimated class population shares being similar to the predicted class memberships is an indicator of good model fit, even though it cannot be used to assess model fit on its own. Indeed, to do this either AIC, BIC, G^2 or X^2 are used. The most common choice to compare models is the BIC and it is preferred over the AIC since it introduces highest penalties on complex models, resulting in simpler model fits, with less latent classes.

```
#To find the optimal number of classes, we look for the one that minimizes the BIC
.
bic.vector <- c()
for (c in 2:5) {
  lca2 <- poLCA(f, hu, nclass = c, graphs = FALSE)
  bic.vector <- append(bic.vector, lca2$bic)
}

## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $v72
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.1022 0.2178 0.3434 0.3366
## class 2:  0.1702 0.3525 0.4077 0.0695
##
## $v73
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.1318 0.1897 0.2734 0.4051
## class 2:  0.2132 0.3604 0.3952 0.0312
##
## $v74
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.1117 0.1805 0.3260 0.3817
## class 2:  0.1993 0.3373 0.4294 0.0340
##
## $v75
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0857 0.1232 0.2230 0.5681
## class 2:  0.2059 0.3227 0.4487 0.0227
##
## $v76
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0364 0.1096 0.1735 0.6805
## class 2:  0.1611 0.2671 0.5661 0.0056
##
## $v77
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0054 0.0000 0.0774 0.9172
## class 2:  0.0526 0.1246 0.6951 0.1277
##
```

```
## $v78
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0136 0.0641 0.1552 0.7671
## class 2:  0.1261 0.2552 0.5926 0.0261
##
## $v81
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0589 0.0443 0.1834 0.7134
## class 2:  0.1839 0.1871 0.4309 0.1981
##
## $v83
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.1235 0.2245 0.3474 0.3047
## class 2:  0.2000 0.4021 0.3200 0.0779
##
## $v141
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.8977 0.0468 0.0303 0.0251
## class 2:  0.7177 0.1439 0.0945 0.0439
##
## Estimated class population shares
##  0.3943 0.6057
##
## Predicted class memberships (by modal posterior prob.)
##  0.3962 0.6038
##
## =========================================================
## Fit for 2 latent classes:
## =========================================================
## number of observations: 795
## number of estimated parameters: 61
## residual degrees of freedom: 734
## maximum log-likelihood: -8820.475
##
## AIC(2): 17762.95
## BIC(2): 18048.33
## G^2(2): 7530.083 (Likelihood ratio/deviance statistic)
## X^2(2): 8530380 (Chi-square goodness of fit)
##
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $v72
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0327 0.1803 0.3960 0.3910
## class 2:  0.4453 0.4456 0.0587 0.0504
## class 3:  0.0198 0.2909 0.5975 0.0917
##
## $v73
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0589 0.1491 0.3021 0.4898
```

```
## class 2:  0.5220 0.3762 0.0891 0.0127
## class 3:  0.0394 0.3483 0.5626 0.0497
##
## $v74
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0199 0.1563 0.3653 0.4584
## class 2:  0.5238 0.3669 0.0791 0.0303
## class 3:  0.0282 0.3054 0.6231 0.0433
##
## $v75
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0199 0.0884 0.2189 0.6728
## class 2:  0.5129 0.3085 0.1325 0.0461
## class 3:  0.0202 0.3214 0.6290 0.0294
##
## $v76
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0180 0.1006 0.1607 0.7208
## class 2:  0.3011 0.2690 0.2795 0.1504
## class 3:  0.0539 0.2425 0.7004 0.0032
##
## $v77
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0036 0.0000 0.0771 0.9193
## class 2:  0.1169 0.1861 0.3321 0.3649
## class 3:  0.0000 0.0576 0.8293 0.1131
##
## $v78
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0084 0.0635 0.1534 0.7747
## class 2:  0.2310 0.2465 0.3052 0.2172
## class 3:  0.0353 0.2248 0.7100 0.0298
##
## $v81
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0339 0.0196 0.1708 0.7756
## class 2:  0.3417 0.2679 0.1873 0.2031
## class 3:  0.0692 0.1226 0.5630 0.2452
##
## $v83
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0954 0.2076 0.3699 0.3272
## class 2:  0.2985 0.4310 0.1664 0.1041
## class 3:  0.1386 0.3609 0.4148 0.0857
##
## $v141
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.9080 0.0526 0.0199 0.0196
## class 2:  0.6838 0.1378 0.1179 0.0605
## class 3:  0.7680 0.1250 0.0740 0.0330
##
```

```
## Estimated class population shares
##  0.3165 0.2809 0.4026
##
## Predicted class memberships (by modal posterior prob.)
##  0.3119 0.283 0.405
##
## =========================================================
## Fit for 3 latent classes:
## =========================================================
## number of observations: 795
## number of estimated parameters: 92
## residual degrees of freedom: 703
## maximum log-likelihood: -8367.685
##
## AIC(3): 16919.37
## BIC(3): 17349.78
## G^2(3): 6624.503 (Likelihood ratio/deviance statistic)
## X^2(3): 3663462 (Chi-square goodness of fit)
##
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $v72
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.1018 0.4525 0.4458
## class 2:  0.3148 0.4528 0.0882 0.1442
## class 3:  0.3985 0.5111 0.0904 0.0000
## class 4:  0.0129 0.2413 0.6406 0.1051
##
## $v73
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0218 0.0938 0.3221 0.5622
## class 2:  0.3834 0.3945 0.1237 0.0985
## class 3:  0.4584 0.4426 0.0990 0.0000
## class 4:  0.0347 0.2969 0.6144 0.0539
##
## $v74
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0657 0.3864 0.5479
## class 2:  0.3534 0.3964 0.1741 0.0760
## class 3:  0.4488 0.4691 0.0765 0.0055
## class 4:  0.0240 0.2512 0.6751 0.0496
##
## $v75
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0097 0.0660 0.1793 0.7449
## class 2:  0.3069 0.2275 0.2425 0.2231
## class 3:  0.4650 0.4002 0.1348 0.0000
## class 4:  0.0105 0.2807 0.6752 0.0337
##
## $v76
```

```
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0850 0.1758 0.7392
## class 2:  0.2035 0.1634 0.1319 0.5012
## class 3:  0.2734 0.3419 0.3788 0.0060
## class 4:  0.0517 0.2238 0.7190 0.0056
##
## $v77
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0000 0.0914 0.9086
## class 2:  0.0261 0.0469 0.0000 0.9270
## class 3:  0.1342 0.2242 0.5658 0.0758
## class 4:  0.0000 0.0500 0.8325 0.1175
##
## $v78
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0676 0.1609 0.7714
## class 2:  0.1061 0.0706 0.1338 0.6895
## class 3:  0.2343 0.3355 0.4302 0.0000
## class 4:  0.0352 0.2114 0.7214 0.0320
##
## $v81
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0205 0.0145 0.1490 0.8160
## class 2:  0.1747 0.1130 0.2197 0.4926
## class 3:  0.3803 0.3238 0.2308 0.0651
## class 4:  0.0469 0.1023 0.5751 0.2757
##
## $v83
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0611 0.2134 0.3784 0.3472
## class 2:  0.2944 0.2322 0.2489 0.2245
## class 3:  0.2540 0.5410 0.1711 0.0340
## class 4:  0.1400 0.3315 0.4310 0.0975
##
## $v141
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.9127 0.0498 0.0204 0.0170
## class 2:  0.8575 0.0426 0.0638 0.0361
## class 3:  0.6232 0.1833 0.1307 0.0629
## class 4:  0.7726 0.1250 0.0682 0.0342
##
## Estimated class population shares
##  0.2549 0.1594 0.2221 0.3635
##
## Predicted class memberships (by modal posterior prob.)
##  0.2579 0.1535 0.2226 0.366
##
## =========================================================
## Fit for 4 latent classes:
## =========================================================
## number of observations: 795
```

```
## number of estimated parameters: 123
## residual degrees of freedom: 672
## maximum log-likelihood: -8176.134
##
## AIC(4): 16598.27
## BIC(4): 17173.7
## G^2(4): 6241.401 (Likelihood ratio/deviance statistic)
## X^2(4): 2641742 (Chi-square goodness of fit)
##
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $v72
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0728 0.7305 0.1967 0.0000
## class 2:  0.2882 0.5169 0.1280 0.0668
## class 3:  0.5977 0.3108 0.0257 0.0659
## class 4:  0.0136 0.1343 0.7264 0.1256
## class 5:  0.0000 0.0741 0.4388 0.4871
##
## $v73
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0963 0.7396 0.1641 0.0000
## class 2:  0.3260 0.4253 0.1783 0.0705
## class 3:  0.7180 0.1718 0.0948 0.0154
## class 4:  0.0322 0.2113 0.6871 0.0693
## class 5:  0.0235 0.0826 0.2924 0.6015
##
## $v74
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0982 0.7641 0.1377 0.0000
## class 2:  0.3120 0.4312 0.2421 0.0147
## class 3:  0.6862 0.1877 0.0718 0.0543
## class 4:  0.0176 0.1508 0.7688 0.0627
## class 5:  0.0000 0.0567 0.3526 0.5907
##
## $v75
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.1881 0.6242 0.1877 0.0000
## class 2:  0.2255 0.2409 0.3023 0.2313
## class 3:  0.6180 0.2121 0.1389 0.0310
## class 4:  0.0098 0.2036 0.7391 0.0476
## class 5:  0.0170 0.0585 0.1594 0.7651
##
## $v76
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0842 0.3484 0.5591 0.0083
## class 2:  0.0989 0.1392 0.1560 0.6059
## class 3:  0.4311 0.2979 0.2711 0.0000
## class 4:  0.0623 0.2158 0.7123 0.0096
## class 5:  0.0156 0.0813 0.1707 0.7325
```

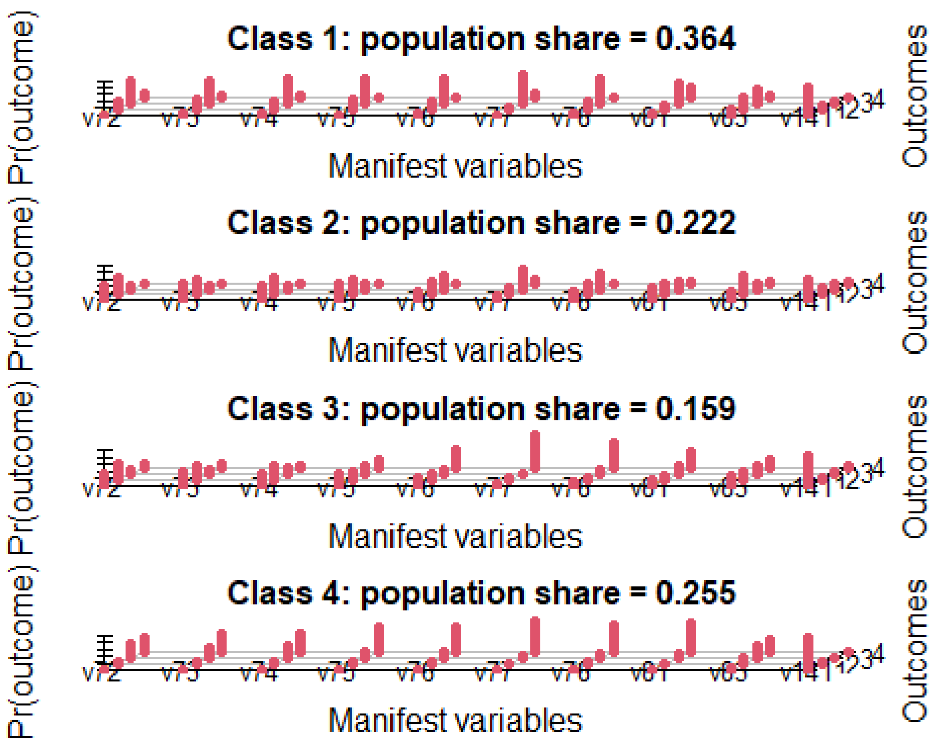
```
##
## $v77
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.1390 0.7977 0.0633
## class 2:  0.0117 0.0000 0.0075 0.9808
## class 3:  0.2292 0.2643 0.3525 0.1540
## class 4:  0.0000 0.0514 0.8180 0.1307
## class 5:  0.0000 0.0000 0.0877 0.9123
##
## $v78
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0291 0.3314 0.6395 0.0000
## class 2:  0.0388 0.0358 0.1515 0.7739
## class 3:  0.3935 0.2871 0.2857 0.0336
## class 4:  0.0440 0.2115 0.7099 0.0346
## class 5:  0.0101 0.0646 0.1503 0.7750
##
## $v81
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.2052 0.3194 0.4133 0.0622
## class 2:  0.1210 0.1023 0.2607 0.5160
## class 3:  0.4586 0.2630 0.1389 0.1395
## class 4:  0.0435 0.0783 0.5672 0.3110
## class 5:  0.0218 0.0105 0.1412 0.8266
##
## $v83
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.1107 0.6004 0.2641 0.0248
## class 2:  0.2322 0.2463 0.2844 0.2372
## class 3:  0.4072 0.3909 0.1342 0.0677
## class 4:  0.1462 0.3004 0.4450 0.1083
## class 5:  0.0621 0.2099 0.3748 0.3532
##
## $v141
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.6990 0.1690 0.1023 0.0297
## class 2:  0.8794 0.0316 0.0499 0.0391
## class 3:  0.5962 0.1711 0.1524 0.0803
## class 4:  0.7849 0.1188 0.0623 0.0339
## class 5:  0.9097 0.0542 0.0189 0.0171
##
## Estimated class population shares
##  0.1629 0.1506 0.1405 0.3055 0.2405
##
## Predicted class memberships (by modal posterior prob.)
##  0.1635 0.1497 0.1358 0.3044 0.2465
##
## =========================================================
## Fit for 5 latent classes:
## =========================================================
## number of observations: 795
```

```
## number of estimated parameters: 154
## residual degrees of freedom: 641
## maximum log-likelihood: -8043.337
##
## AIC(5): 16394.67
## BIC(5): 17115.14
## G^2(5): 5975.807 (Likelihood ratio/deviance statistic)
## X^2(5): 2271995 (Chi-square goodness of fit)
##
```

```
bic.hu <- data.frame(class = c(2:5), bic = bic.vector)
bic.hu[which.min(bic.hu$bic), ]
```

```
##   class      bic
## 4     5 17115.14
```

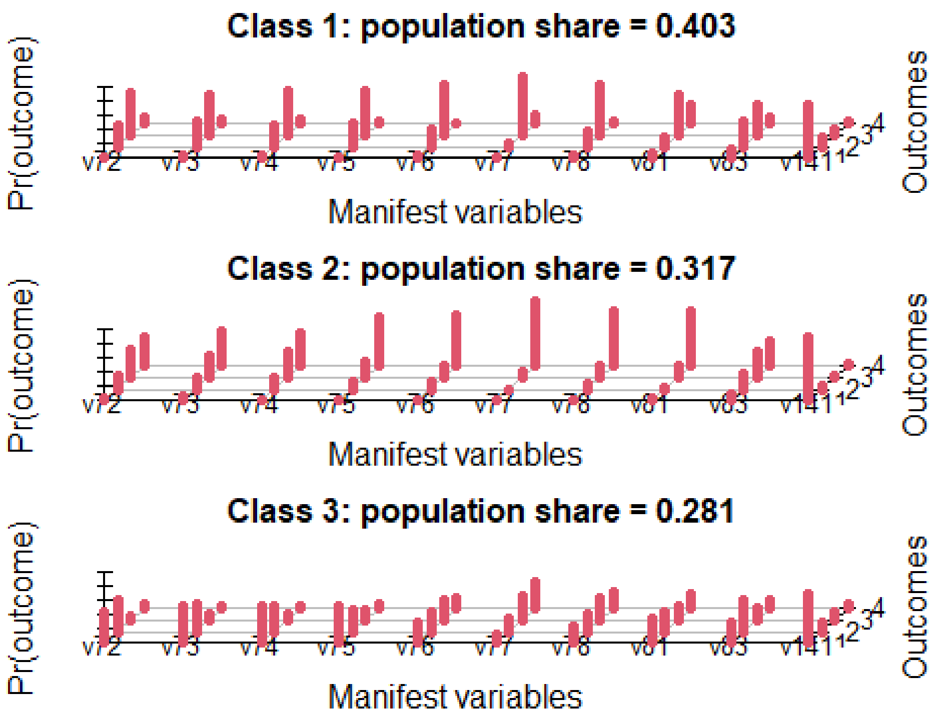
```
lca.hu.4 <- poLCA(f, hu, nclass = 4, graphs = TRUE)
```




```
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $v72
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0129 0.2413 0.6406 0.1051
## class 2:  0.3985 0.5111 0.0904 0.0000
## class 3:  0.3148 0.4528 0.0882 0.1442
## class 4:  0.0000 0.1018 0.4525 0.4458
##
## $v73
```

```
##             Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0347 0.2969 0.6144 0.0539
## class 2:  0.4584 0.4426 0.0990 0.0000
## class 3:  0.3834 0.3945 0.1237 0.0985
## class 4:  0.0218 0.0938 0.3221 0.5622
##
## $v74
##             Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0240 0.2512 0.6751 0.0496
## class 2:  0.4488 0.4691 0.0765 0.0055
## class 3:  0.3534 0.3964 0.1741 0.0760
## class 4:  0.0000 0.0657 0.3864 0.5479
##
## $v75
##             Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0105 0.2807 0.6752 0.0337
## class 2:  0.4650 0.4002 0.1348 0.0000
## class 3:  0.3069 0.2275 0.2425 0.2231
## class 4:  0.0097 0.0660 0.1793 0.7449
##
## $v76
##             Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0517 0.2238 0.7190 0.0056
## class 2:  0.2734 0.3419 0.3788 0.0060
## class 3:  0.2035 0.1634 0.1319 0.5012
## class 4:  0.0000 0.0850 0.1758 0.7392
##
## $v77
##             Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0500 0.8325 0.1175
## class 2:  0.1342 0.2242 0.5658 0.0758
## class 3:  0.0261 0.0469 0.0000 0.9270
## class 4:  0.0000 0.0000 0.0914 0.9086
##
## $v78
##             Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0352 0.2114 0.7214 0.0320
## class 2:  0.2343 0.3355 0.4302 0.0000
## class 3:  0.1061 0.0706 0.1338 0.6895
## class 4:  0.0000 0.0676 0.1609 0.7714
##
## $v81
##             Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0469 0.1023 0.5751 0.2757
## class 2:  0.3803 0.3238 0.2308 0.0651
## class 3:  0.1747 0.1130 0.2197 0.4926
## class 4:  0.0205 0.0145 0.1490 0.8160
##
## $v83
##             Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.1400 0.3315 0.4310 0.0975
```

```
## class 2:  0.2540 0.5410 0.1711 0.0340
## class 3:  0.2944 0.2322 0.2489 0.2245
## class 4:  0.0611 0.2134 0.3784 0.3472
##
## $v141
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.7726 0.1250 0.0682 0.0342
## class 2:  0.6232 0.1833 0.1307 0.0629
## class 3:  0.8575 0.0426 0.0638 0.0361
## class 4:  0.9127 0.0498 0.0204 0.0170
##
## Estimated class population shares
##  0.3635 0.2221 0.1594 0.2549
##
## Predicted class memberships (by modal posterior prob.)
##  0.366 0.2226 0.1535 0.2579
##
## =========================================================
## Fit for 4 latent classes:
## =========================================================
## number of observations: 795
## number of estimated parameters: 123
## residual degrees of freedom: 672
## maximum log-likelihood: -8176.134
##
## AIC(4): 16598.27
## BIC(4): 17173.7
## G^2(4): 6241.401 (Likelihood ratio/deviance statistic)
## X^2(4): 2641742 (Chi-square goodness of fit)
##

#Four classes seems to be the optimal number; however, given the absence of very b
ig groups (the percentage of people belonging to certain classes is quite low) and
the theoretical framework we are considering (the three balances context), three c
lasses may work as well.

lca.hu.3 <- poLCA(f, hu, nclass = 3, graphs = TRUE)
```




```
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $v72
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0198 0.2910 0.5975 0.0917
## class 2:  0.0327 0.1803 0.3960 0.3910
## class 3:  0.4453 0.4456 0.0587 0.0504
##
## $v73
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0394 0.3483 0.5626 0.0497
## class 2:  0.0589 0.1491 0.3021 0.4898
## class 3:  0.5220 0.3762 0.0891 0.0127
##
## $v74
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0282 0.3054 0.6231 0.0433
## class 2:  0.0199 0.1563 0.3653 0.4584
## class 3:  0.5238 0.3669 0.0791 0.0303
##
## $v75
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0202 0.3214 0.6290 0.0294
## class 2:  0.0199 0.0884 0.2189 0.6728
## class 3:  0.5129 0.3085 0.1325 0.0461
##
## $v76
```

```
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0539 0.2425 0.7004 0.0032
## class 2:  0.0180 0.1006 0.1607 0.7208
## class 3:  0.3011 0.2690 0.2795 0.1504
##
## $v77
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0576 0.8293 0.1131
## class 2:  0.0036 0.0000 0.0771 0.9193
## class 3:  0.1169 0.1861 0.3321 0.3649
##
## $v78
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0353 0.2248 0.7100 0.0298
## class 2:  0.0084 0.0635 0.1534 0.7747
## class 3:  0.2310 0.2465 0.3052 0.2172
##
## $v81
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0692 0.1226 0.5630 0.2452
## class 2:  0.0339 0.0196 0.1708 0.7756
## class 3:  0.3417 0.2679 0.1873 0.2031
##
## $v83
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.1386 0.3609 0.4148 0.0857
## class 2:  0.0954 0.2076 0.3699 0.3272
## class 3:  0.2985 0.4310 0.1664 0.1041
##
## $v141
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.7680 0.1250 0.0740 0.0330
## class 2:  0.9080 0.0526 0.0199 0.0196
## class 3:  0.6838 0.1378 0.1179 0.0605
##
## Estimated class population shares
##  0.4026 0.3165 0.2809
##
## Predicted class memberships (by modal posterior prob.)
##  0.405 0.3119 0.283
##
## =========================================================
## Fit for 3 latent classes:
## =========================================================
## number of observations: 795
## number of estimated parameters: 92
## residual degrees of freedom: 703
## maximum log-likelihood: -8367.685
##
## AIC(3): 16919.37
## BIC(3): 17349.78
```

```
## G^2(3): 6624.503 (Likelihood ratio/deviance statistic)
## X^2(3): 3663435 (Chi-square goodness of fit)
##
```

It is possible to notice how the population is well differentiated in three groups of big sizes, which suggest a widespread presence of all the three kind of balances, understood as family models, almost in equal measure in the Hungarian society.

Class one, which involves 28.1% of the population, is composed for more than 89% of people that believes children suffer when the mother works and almost 89% that believes that family life suffers in general. Almost 90% of people agree with the fact that what women really want is home and children and almost 61% would give priority to men in the job market if jobs are scarce. This last statement, find correspondence in the beliefs that men are better business executives for almost 48% of this class respondents. It seems quite evident that this group of people is loyal to the traditional family model, which we could associate to the first level of the gender revolution theory, with the risk of scarce gender equality, both in public and private sphere.

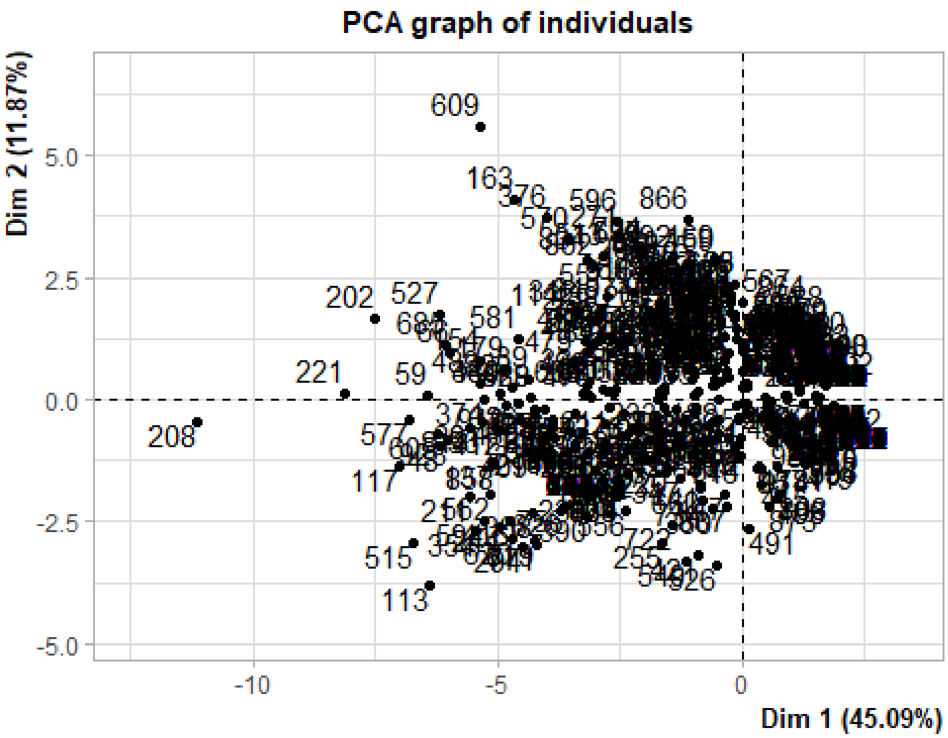
Class two, composed by 31.7% of the population, manifest significant levels of disagreement with statements related to family life. Between 70 and 95% of the people belonging to this profile, express disagreement, mostly strong disagreement, with all the statements analyzed above. This seems to be an indicator of households where taking care of home and children is not believed to be a specific duty of women only.

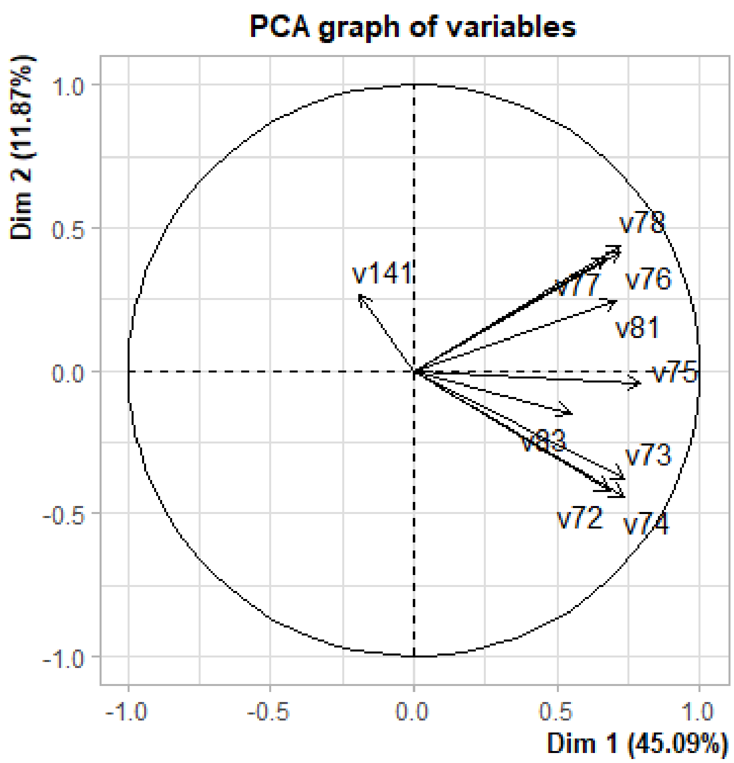
Class three, which characterizes more than 40% of the population, shows a moderate disagreement with most questions, leaving space for a minor but significant part of people in this group to show moderate agreement. Only few people (18%) believe that men have the precedence in working if jobs are scarce (v81), however around 34% is loyal to traditional roles (v75), and think children and family life suffer when mothers works (v72 and v74) and, for 39% of them a woman's desires are linked to home and children. This may be seen as a transitional part of society, associating these opinions with the ones found in the second stage of the gender revolution theory, where equality in the job market and education (public sphere) is recognized, but housework are also still considered mostly a woman's duty.

The three groups are quite different, nevertheless none questions the need of democratic equal right. Interesting is the role of education: while in class two and three no one agrees with the statement "A university education is more important for a boy than for a girl." (v77), more than 30% of people in class one believes this is right. The right to education for everyone seems to be a well-established principle for most, but the difference between class one and the other two is quite strong.

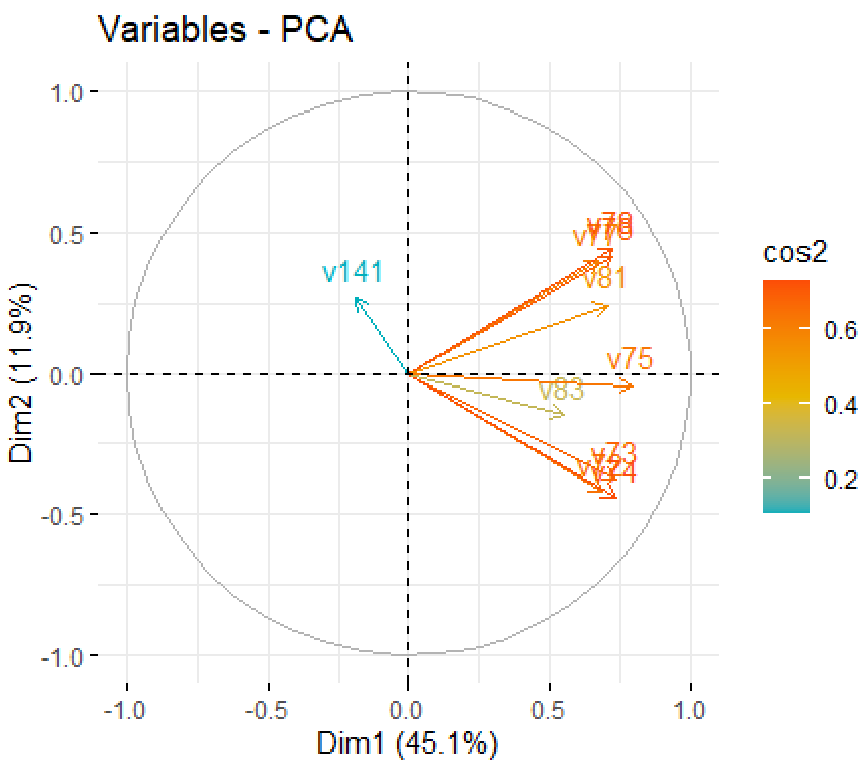
Let's analyze the situation in Sweden and see if and how the values change.

```
#SWEDEN

#PCA

#Changing country, we can once again look at the PCA to see if we find common patt
ern in which are the most divisive questions.
se <- subset(se, select = - c(c_abrv))
PCanalysis2 <- PCA(se)
```





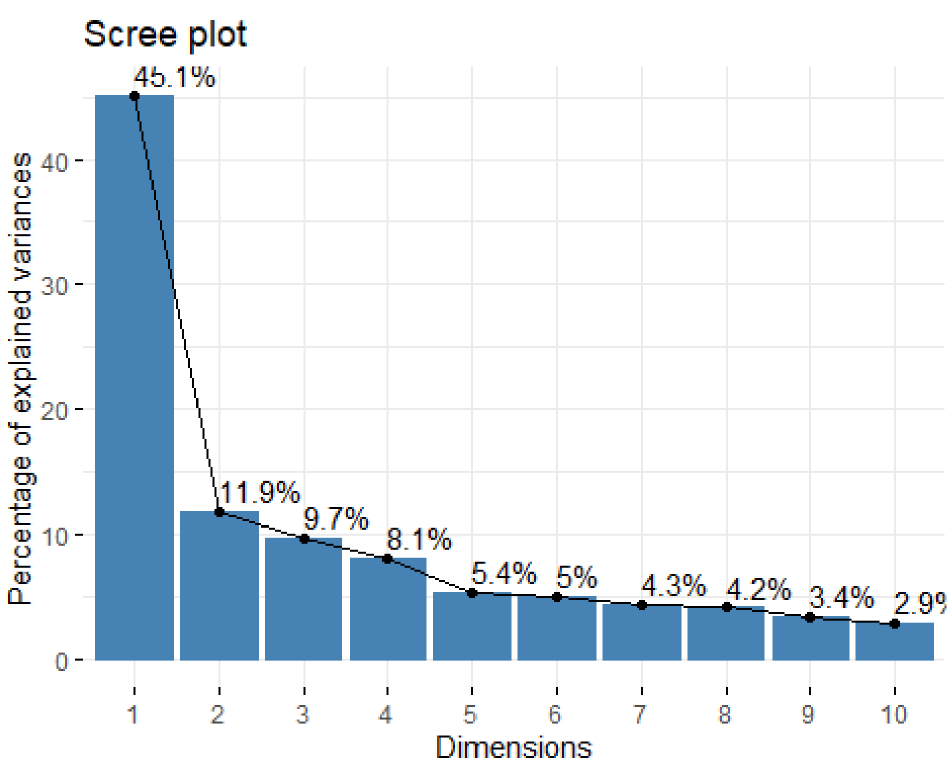

```
fviz_pca_var(PCanalysis2, col.var='cos2', gradient.cols=c('#00afbb', '#e7b800', '#fc4e07'))
```



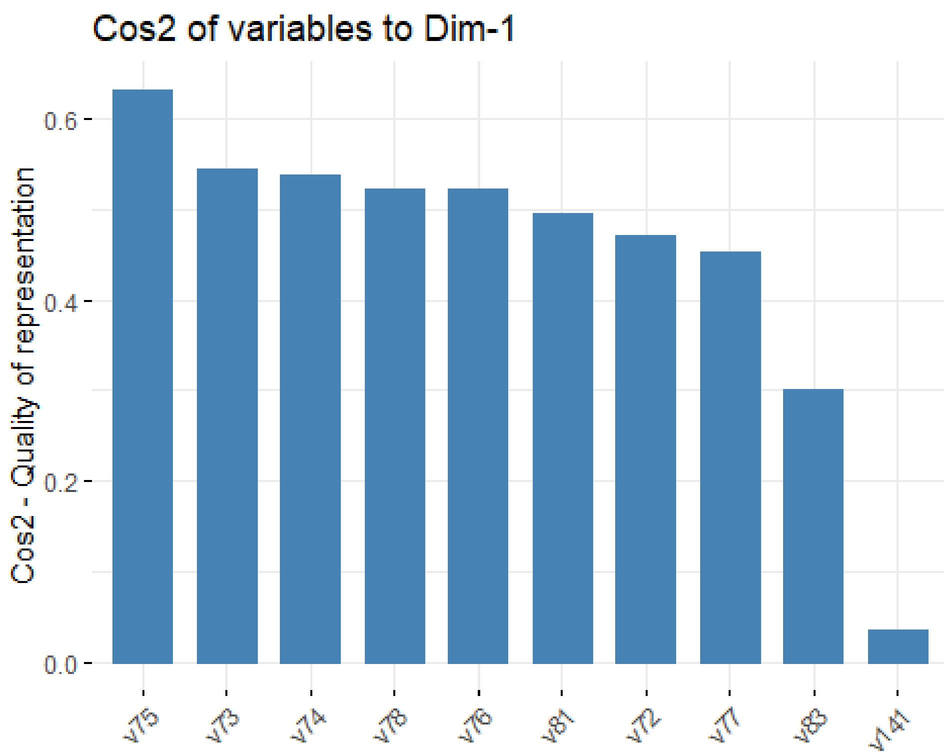

```
summary(PCanalysis2)

##
## Call:
## PCA(X = se)
##
##
## Eigenvalues
##                        Dim.1   Dim.2   Dim.3   Dim.4   Dim.5   Dim.6   Dim.7
## Variance               4.509   1.187   0.975   0.811   0.535   0.503   0.433
## % of var.             45.093  11.870   9.748   8.111   5.355   5.034   4.326
## Cumulative % of var.  45.093  56.963  66.711  74.822  80.177  85.211  89.537
##                        Dim.8   Dim.9  Dim.10
## Variance               0.415   0.337   0.294
## % of var.              4.151   3.369   2.944
## Cumulative % of var.  93.687  97.056 100.000
##
## Individuals (the 10 first)
##           Dist    Dim.1    ctr   cos2    Dim.2    ctr   cos2    Dim.3    ctr
## 1     |  2.099 |  2.005  0.096  0.912 | -0.500  0.023  0.057 | -0.052  0.000
## 2     |  2.452 |  0.729  0.013  0.088 |  0.592  0.032  0.058 | -0.118  0.002
## 3     |  4.027 | -2.077  0.103  0.266 |  2.971  0.801  0.544 | -0.904  0.090
## 4     |  3.452 | -2.545  0.155  0.543 | -1.983  0.357  0.330 |  0.044  0.000
## 5     |  2.585 |  0.377  0.003  0.021 | -0.372  0.013  0.021 | -0.347  0.013
## 6     |  2.099 |  2.005  0.096  0.912 | -0.500  0.023  0.057 | -0.052  0.000
## 7     |  1.999 |  1.626  0.063  0.661 | -0.300  0.008  0.022 |  0.234  0.006
## 8     |  2.387 |  1.484  0.053  0.386 |  0.092  0.001  0.001 |  1.721  0.328
## 9     |  2.099 |  2.005  0.096  0.912 | -0.500  0.023  0.057 | -0.052  0.000
```

```
## 10    |  2.722 |  0.023  0.000  0.000 |  1.149  0.120  0.178 | -0.899  0.089
##         cos2
## 1      0.001 |
## 2      0.002 |
## 3      0.050 |
## 4      0.000 |
## 5      0.018 |
## 6      0.001 |
## 7      0.014 |
## 8      0.520 |
## 9      0.001 |
## 10     0.109 |
##
## Variables
##          Dim.1    ctr   cos2    Dim.2    ctr   cos2    Dim.3    ctr   cos2
## v72  |  0.686 10.428  0.470 | -0.419 14.808  0.176 |  0.161  2.669  0.026 |
## v73  |  0.737 12.047  0.543 | -0.376 11.899  0.141 |  0.135  1.857  0.018 |
## v74  |  0.732 11.897  0.536 | -0.440 16.335  0.194 |  0.173  3.058  0.030 |
## v75  |  0.794 13.989  0.631 | -0.045  0.170  0.002 |  0.097  0.962  0.009 |
## v76  |  0.723 11.577  0.522 |  0.415 14.498  0.172 | -0.037  0.139  0.001 |
## v77  |  0.672 10.024  0.452 |  0.396 13.241  0.157 | -0.046  0.217  0.002 |
## v78  |  0.723 11.590  0.523 |  0.439 16.206  0.192 | -0.080  0.659  0.006 |
## v81  |  0.704 10.983  0.495 |  0.243  4.985  0.059 | -0.027  0.073  0.001 |
## v83  |  0.549  6.677  0.301 | -0.149  1.863  0.022 | -0.193  3.814  0.037 |
## v141 | -0.188  0.787  0.036 |  0.267  5.995  0.071 |  0.919 86.552  0.844 |
```

```
#Scree plot:
fviz_screeplot(PCanalysis2, addlabels=TRUE, ncp=10)
```

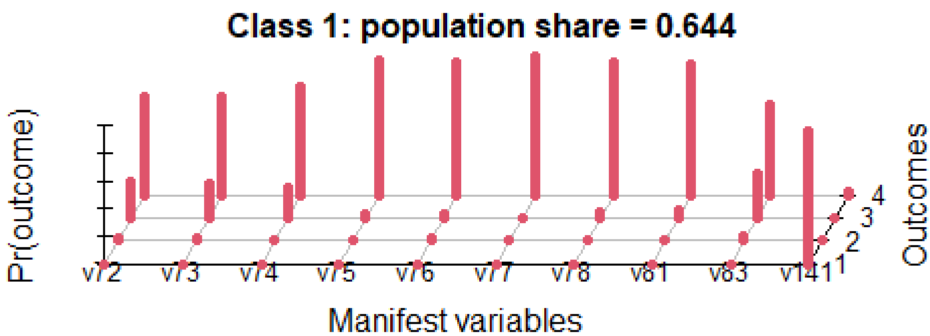


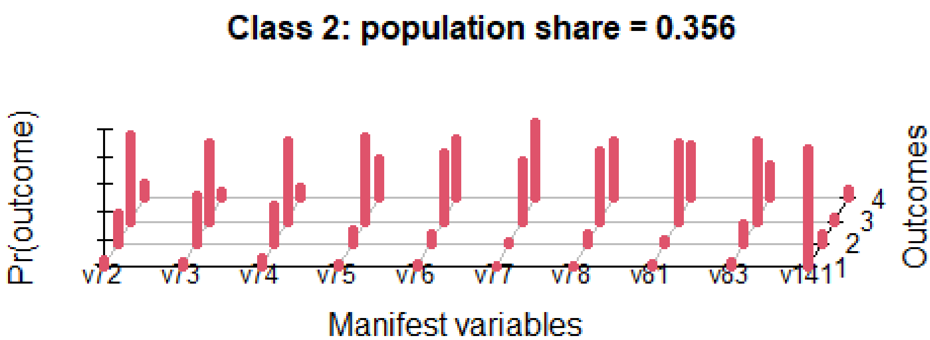
```r
#Contributions to the first component:
fviz_cos2(PCanalysis2, choice='var', axes=1)
```



All statements but v141 seems to be quite important in their contributions to explain the total variance and with two principal component we can explain almost 57% of it. As before, we expected v83 and v141 to be the least divisive and important aspects, while the most relevant ones appears to be v73, v74, and v75.

```r
#LCA

f <- cbind(v72, v73, v74, v75, v76, v77, v78, v81, v83, v141)~1
lca1 <- poLCA(f, se, nclass = 2, graphs = TRUE)
```






```
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $v72
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0027 0.0223 0.2596 0.7153
## class 2:  0.0435 0.2261 0.6202 0.1102
##
## $v73
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0040 0.0284 0.2538 0.7138
## class 2:  0.0413 0.3483 0.5581 0.0523
##
## $v74
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0024 0.0074 0.2156 0.7746
## class 2:  0.0684 0.2773 0.5818 0.0724
##
## $v75
##            Pr(1) Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0017 0.000 0.0285 0.9699
## class 2:  0.0182 0.103 0.6026 0.2762
##
## $v76
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0149 0.0295 0.9556
## class 2:  0.0151 0.0729 0.4887 0.4233
##
```

```
## $v77
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0017 0.0060 0.9924
## class 2:  0.0091 0.0212 0.4314 0.5384
##
## $v78
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0100 0.0385 0.9515
## class 2:  0.0061 0.0758 0.5057 0.4125
##
## $v81
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0040 0.0582 0.9378
## class 2:  0.0151 0.0352 0.5670 0.3827
##
## $v83
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0019 0.0294 0.3183 0.6504
## class 2:  0.0360 0.1436 0.5837 0.2368
##
## $v141
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.9589 0.0088 0.0023 0.0300
## class 2:  0.8503 0.0719 0.0231 0.0546
##
## Estimated class population shares
##  0.6442 0.3558
##
## Predicted class memberships (by modal posterior prob.)
##  0.6466 0.3534
##
## =========================================================
## Fit for 2 latent classes:
## =========================================================
## number of observations: 928
## number of estimated parameters: 61
## residual degrees of freedom: 867
## maximum log-likelihood: -5857.662
##
## AIC(2): 11837.32
## BIC(2): 12132.14
## G^2(2): 3206.284 (Likelihood ratio/deviance statistic)
## X^2(2): 1.170009e+12 (Chi-square goodness of fit)
##

bic.vector <- c()
for (c in 1:5) {
  lca2 <- poLCA(f, se, nclass = c, graphs = FALSE)
  bic.vector <- append(bic.vector, lca2$bic)
}
```

```
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $v72
##            Pr(1)  Pr(2)  Pr(3) Pr(4)
## class 1:  0.0172 0.0948 0.3879   0.5
##
## $v73
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0172 0.1422 0.3621 0.4784
##
## $v74
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0259 0.1034 0.3459 0.5248
##
## $v75
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0075 0.0366 0.2328 0.7231
##
## $v76
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0054 0.0356 0.1929 0.7662
##
## $v77
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0032 0.0086 0.1573 0.8308
##
## $v78
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0022 0.0334 0.2047 0.7597
##
## $v81
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0054 0.0151 0.2392 0.7403
##
## $v83
##           Pr(1) Pr(2)  Pr(3)  Pr(4)
## class 1:  0.014  0.07 0.4127 0.5032
##
## $v141
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.9203 0.0312 0.0097 0.0388
##
## Estimated class population shares
##  1
##
## Predicted class memberships (by modal posterior prob.)
##  1
##
## =========================================================
## Fit for 1 latent classes:
```

```
## =============================================================
## number of observations: 928
## number of estimated parameters: 30
## residual degrees of freedom: 898
## maximum log-likelihood: -7093.37
##
## AIC(1): 14246.74
## BIC(1): 14391.73
## G^2(1): 5677.701 (Likelihood ratio/deviance statistic)
## X^2(1): 4.96268e+15 (Chi-square goodness of fit)
##
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $v72
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0435 0.2261 0.6202 0.1102
## class 2:  0.0027 0.0223 0.2596 0.7153
##
## $v73
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0413 0.3483 0.5581 0.0523
## class 2:  0.0040 0.0284 0.2538 0.7138
##
## $v74
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0684 0.2773 0.5818 0.0724
## class 2:  0.0024 0.0074 0.2156 0.7746
##
## $v75
##            Pr(1) Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0182 0.103 0.6026 0.2762
## class 2:  0.0017 0.000 0.0285 0.9699
##
## $v76
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0151 0.0729 0.4887 0.4233
## class 2:  0.0000 0.0149 0.0295 0.9556
##
## $v77
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0091 0.0212 0.4314 0.5384
## class 2:  0.0000 0.0017 0.0060 0.9924
##
## $v78
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0061 0.0758 0.5057 0.4125
## class 2:  0.0000 0.0100 0.0385 0.9515
##
## $v81
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
```

```
## class 1:  0.0151 0.0352 0.5670 0.3827
## class 2:  0.0000 0.0040 0.0582 0.9378
##
## $v83
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0360 0.1436 0.5837 0.2368
## class 2:  0.0019 0.0294 0.3183 0.6504
##
## $v141
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.8503 0.0719 0.0231 0.0546
## class 2:  0.9589 0.0088 0.0023 0.0300
##
## Estimated class population shares
##  0.3558 0.6442
##
## Predicted class memberships (by modal posterior prob.)
##  0.3534 0.6466
##
## =========================================================
## Fit for 2 latent classes:
## =========================================================
## number of observations: 928
## number of estimated parameters: 61
## residual degrees of freedom: 867
## maximum log-likelihood: -5857.662
##
## AIC(2): 11837.32
## BIC(2): 12132.14
## G^2(2): 3206.284 (Likelihood ratio/deviance statistic)
## X^2(2): 1.170009e+12 (Chi-square goodness of fit)
##
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $v72
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0248 0.1924 0.6235 0.1593
## class 2:  0.0560 0.2081 0.6320 0.1038
## class 3:  0.0000 0.0057 0.1829 0.8114
##
## $v73
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0216 0.2465 0.6171 0.1148
## class 2:  0.0492 0.3547 0.5657 0.0304
## class 3:  0.0040 0.0156 0.1609 0.8195
##
## $v74
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0547 0.2030 0.6275 0.1147
## class 2:  0.0579 0.2495 0.6166 0.0760
```

```
## class 3:  0.0000 0.0020 0.1080 0.8899
##
## $v75
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0038 0.0219 0.3780 0.5964
## class 2:  0.0293 0.1652 0.6602 0.1452
## class 3:  0.0020 0.0000 0.0111 0.9869
##
## $v76
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0000 0.1528 0.8472
## class 2:  0.0291 0.1388 0.7396 0.0925
## class 3:  0.0000 0.0184 0.0255 0.9561
##
## $v77
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0000 0.0794 0.9206
## class 2:  0.0175 0.0408 0.7246 0.2172
## class 3:  0.0000 0.0020 0.0023 0.9957
##
## $v78
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0153 0.1483 0.8364
## class 2:  0.0116 0.1249 0.7816 0.0819
## class 3:  0.0000 0.0113 0.0354 0.9533
##
## $v81
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0039 0.0025 0.2715 0.7221
## class 2:  0.0233 0.0643 0.7480 0.1643
## class 3:  0.0000 0.0046 0.0477 0.9477
##
## $v83
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0230 0.0947 0.4957 0.3866
## class 2:  0.0355 0.1574 0.6668 0.1403
## class 3:  0.0020 0.0273 0.2827 0.6880
##
## $v141
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.8900 0.0358 0.0143 0.0598
## class 2:  0.8449 0.0865 0.0251 0.0436
## class 3:  0.9617 0.0099 0.0020 0.0263
##
## Estimated class population shares
##  0.2767 0.1851 0.5382
##
## Predicted class memberships (by modal posterior prob.)
##  0.278 0.1789 0.5431
##
## =========================================================
```

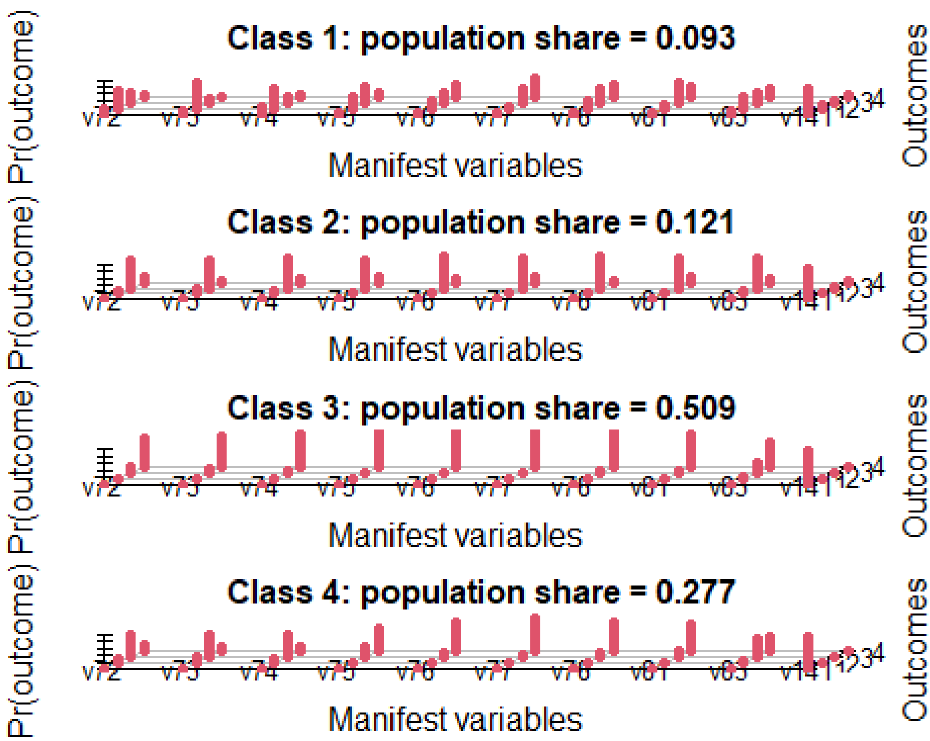
```
## Fit for 3 latent classes:
## =============================================================
## number of observations: 928
## number of estimated parameters: 92
## residual degrees of freedom: 836
## maximum log-likelihood: -5623.011
##
## AIC(3): 11430.02
## BIC(3): 11874.66
## G^2(3): 2736.983 (Likelihood ratio/deviance statistic)
## X^2(3): 88426287807 (Chi-square goodness of fit)
##
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $v72
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0144 0.1220 0.6599 0.2037
## class 2:  0.1410 0.5045 0.2785 0.0760
## class 3:  0.0000 0.0905 0.7806 0.1289
## class 4:  0.0000 0.0053 0.1672 0.8275
##
## $v73
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0050 0.1664 0.6813 0.1473
## class 2:  0.0958 0.7538 0.1316 0.0188
## class 3:  0.0379 0.1490 0.7768 0.0363
## class 4:  0.0044 0.0143 0.1329 0.8483
##
## $v74
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0246 0.1329 0.6795 0.1629
## class 2:  0.2025 0.5602 0.1869 0.0504
## class 3:  0.0000 0.1075 0.8206 0.0718
## class 4:  0.0000 0.0020 0.0814 0.9166
##
## $v75
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0000 0.3252 0.6748
## class 2:  0.0687 0.3480 0.4359 0.1474
## class 3:  0.0000 0.0319 0.8075 0.1605
## class 4:  0.0021 0.0000 0.0082 0.9897
##
## $v76
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0093 0.1550 0.8358
## class 2:  0.0572 0.2370 0.3069 0.3989
## class 3:  0.0000 0.0157 0.9101 0.0742
## class 4:  0.0000 0.0173 0.0216 0.9611
##
## $v77
```

```
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0044 0.0610 0.9346
## class 2:  0.0257 0.0476 0.3945 0.5322
## class 3:  0.0067 0.0152 0.8439 0.1342
## class 4:  0.0000 0.0021 0.0022 0.9956
##
## $v78
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0186 0.1374 0.8440
## class 2:  0.0229 0.1975 0.4198 0.3598
## class 3:  0.0000 0.0335 0.9183 0.0481
## class 4:  0.0000 0.0111 0.0316 0.9574
##
## $v81
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0038 0.2232 0.7730
## class 2:  0.0458 0.1041 0.5488 0.3013
## class 3:  0.0089 0.0170 0.8469 0.1272
## class 4:  0.0000 0.0043 0.0460 0.9498
##
## $v83
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0668 0.5245 0.4087
## class 2:  0.1312 0.3297 0.2677 0.2714
## class 3:  0.0046 0.0562 0.8464 0.0928
## class 4:  0.0022 0.0271 0.2757 0.6951
##
## $v141
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.9144 0.0241 0.0055 0.0561
## class 2:  0.7487 0.1381 0.0637 0.0495
## class 3:  0.8924 0.0519 0.0092 0.0465
## class 4:  0.9618 0.0105 0.0021 0.0256
##
## Estimated class population shares
##  0.2759 0.0942 0.1211 0.5088
##
## Predicted class memberships (by modal posterior prob.)
##  0.2834 0.0862 0.1196 0.5108
##
## =========================================================
## Fit for 4 latent classes:
## =========================================================
## number of observations: 928
## number of estimated parameters: 123
## residual degrees of freedom: 805
## maximum log-likelihood: -5456.315
##
## AIC(4): 11158.63
## BIC(4): 11753.09
## G^2(4): 2403.59 (Likelihood ratio/deviance statistic)
```

```
## X^2(4): 137587473 (Chi-square goodness of fit)
##
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $v72
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0352 0.4503 0.4662 0.0482
## class 2:  0.0136 0.1177 0.6790 0.1897
## class 3:  0.2702 0.5732 0.0000 0.1567
## class 4:  0.0000 0.0387 0.8254 0.1359
## class 5:  0.0000 0.0057 0.1693 0.8250
##
## $v73
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0700 0.6883 0.2417 0.0000
## class 2:  0.0060 0.1777 0.6834 0.1329
## class 3:  0.1290 0.7947 0.0000 0.0763
## class 4:  0.0319 0.0413 0.8781 0.0487
## class 5:  0.0043 0.0144 0.1383 0.8430
##
## $v74
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0607 0.4985 0.3898 0.0510
## class 2:  0.0229 0.1410 0.6738 0.1622
## class 3:  0.3746 0.5533 0.0000 0.0721
## class 4:  0.0000 0.0530 0.8943 0.0528
## class 5:  0.0000 0.0021 0.0883 0.9097
##
## $v75
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.3585 0.5627 0.0788
## class 2:  0.0000 0.0031 0.3455 0.6515
## class 3:  0.1597 0.2463 0.3448 0.2491
## class 4:  0.0000 0.0000 0.8139 0.1861
## class 5:  0.0021 0.0000 0.0085 0.9894
##
## $v76
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.2541 0.5785 0.1675
## class 2:  0.0039 0.0105 0.1654 0.8202
## class 3:  0.1065 0.1389 0.1117 0.6429
## class 4:  0.0000 0.0000 0.9320 0.0680
## class 5:  0.0000 0.0170 0.0218 0.9612
##
## $v77
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0449 0.0399 0.6794 0.2358
## class 2:  0.0000 0.0067 0.0610 0.9323
## class 3:  0.0000 0.0405 0.1405 0.8190
## class 4:  0.0000 0.0122 0.8796 0.1083
```

```
## class 5:  0.0000 0.0021 0.0023 0.9956
##
## $v78
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.2333 0.6514 0.1152
## class 2:  0.0000 0.0199 0.1483 0.8318
## class 3:  0.0532 0.1354 0.2271 0.5842
## class 4:  0.0000 0.0000 0.9493 0.0507
## class 5:  0.0000 0.0109 0.0315 0.9576
##
## $v81
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0151 0.1052 0.7895 0.0903
## class 2:  0.0000 0.0037 0.2413 0.7550
## class 3:  0.0798 0.0800 0.2976 0.5426
## class 4:  0.0111 0.0108 0.8304 0.1476
## class 5:  0.0000 0.0043 0.0460 0.9498
##
## $v83
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0520 0.3302 0.4658 0.1520
## class 2:  0.0000 0.0724 0.5307 0.3969
## class 3:  0.2270 0.2790 0.1240 0.3700
## class 4:  0.0000 0.0116 0.8844 0.1040
## class 5:  0.0021 0.0269 0.2764 0.6946
##
## $v141
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.8358 0.1489 0.0000 0.0153
## class 2:  0.9155 0.0259 0.0077 0.0509
## class 3:  0.6523 0.0910 0.1330 0.1237
## class 4:  0.8887 0.0452 0.0115 0.0546
## class 5:  0.9616 0.0103 0.0021 0.0259
##
## Estimated class population shares
##  0.072 0.2765 0.0405 0.0963 0.5148
##
## Predicted class memberships (by modal posterior prob.)
##  0.0679 0.2672 0.0431 0.0959 0.5259
##
## =========================================================
## Fit for 5 latent classes:
## =========================================================
## number of observations: 928
## number of estimated parameters: 154
## residual degrees of freedom: 774
## maximum log-likelihood: -5390.91
##
## AIC(5): 11089.82
## BIC(5): 11834.11
## G^2(5): 2272.78 (Likelihood ratio/deviance statistic)
```

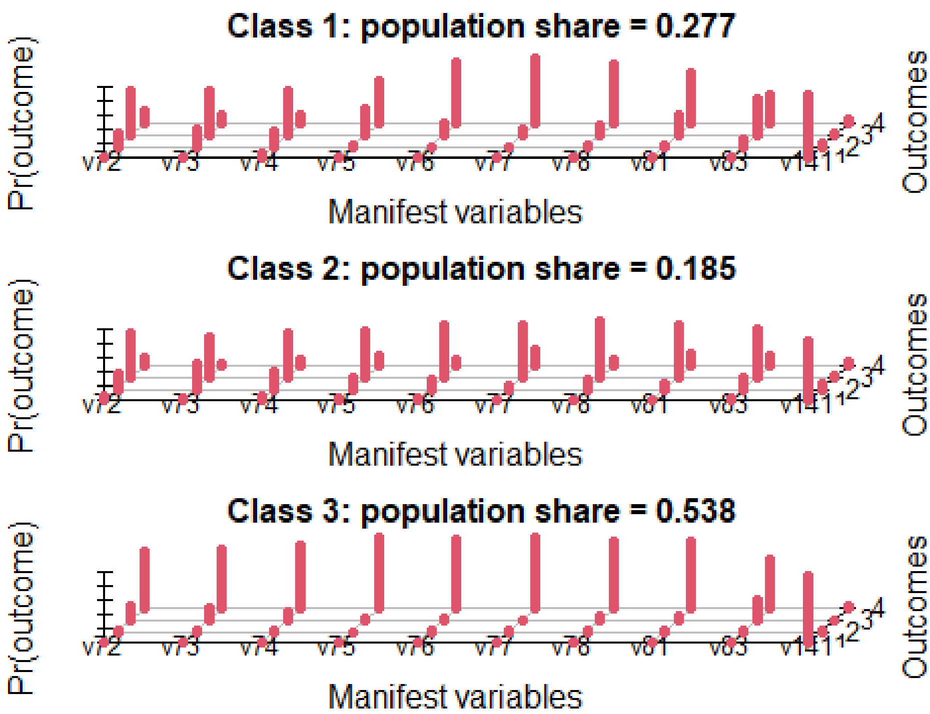
```
## X^2(5): 8558008 (Chi-square goodness of fit)
##

bic <- data.frame(class = c(1:5), bic = bic.vector)
bic[which.min(bic$bic), ]

##   class      bic
## 4     4 11753.09

lca.se <- poLCA(f, se, nclass = 4, graphs = TRUE)
```




```
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $v72
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.1418 0.5053 0.2830 0.0700
## class 2:  0.0000 0.0884 0.7823 0.1293
## class 3:  0.0000 0.0053 0.1676 0.8271
## class 4:  0.0147 0.1246 0.6559 0.2048
##
## $v73
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0968 0.7517 0.1334 0.0181
## class 2:  0.0379 0.1442 0.7808 0.0371
## class 3:  0.0044 0.0141 0.1333 0.8482
## class 4:  0.0051 0.1724 0.6766 0.1459
##
## $v74
```

```
##             Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.2030 0.5533 0.1933 0.0505
## class 2:  0.0000 0.1047 0.8224 0.0729
## class 3:  0.0000 0.0019 0.0817 0.9164
## class 4:  0.0253 0.1387 0.6748 0.1613
##
## $v75
##             Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0696 0.3547 0.4404 0.1353
## class 2:  0.0000 0.0304 0.8068 0.1627
## class 3:  0.0021 0.0000 0.0083 0.9896
## class 4:  0.0000 0.0000 0.3253 0.6747
##
## $v76
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.058 0.2423 0.3157 0.3840
## class 2:  0.000 0.0139 0.9103 0.0757
## class 3:  0.000 0.0173 0.0218 0.9609
## class 4:  0.000 0.0093 0.1532 0.8375
##
## $v77
##             Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0267 0.0483 0.4066 0.5184
## class 2:  0.0062 0.0155 0.8423 0.1360
## class 3:  0.0000 0.0021 0.0022 0.9956
## class 4:  0.0000 0.0043 0.0600 0.9357
##
## $v78
##             Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0232 0.2024 0.4278 0.3466
## class 2:  0.0000 0.0317 0.9185 0.0498
## class 3:  0.0000 0.0111 0.0317 0.9572
## class 4:  0.0000 0.0185 0.1366 0.8449
##
## $v81
##             Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0464 0.1061 0.5605 0.2870
## class 2:  0.0089 0.0165 0.8449 0.1296
## class 3:  0.0000 0.0043 0.0461 0.9497
## class 4:  0.0000 0.0038 0.2223 0.7739
##
## $v83
##             Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.1268 0.3316 0.2707 0.2709
## class 2:  0.0042 0.0544 0.8476 0.0938
## class 3:  0.0021 0.0270 0.2758 0.6950
## class 4:  0.0023 0.0682 0.5222 0.4074
##
## $v141
##             Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.7465 0.1395 0.0645 0.0495
```

```
## class 2:  0.8931 0.0517 0.0092 0.0460
## class 3:  0.9618 0.0104 0.0021 0.0256
## class 4:  0.9140 0.0243 0.0054 0.0563
##
## Estimated class population shares
##  0.0929 0.1209 0.5092 0.2771
##
## Predicted class memberships (by modal posterior prob.)
##  0.0862 0.1175 0.5065 0.2899
##
## =========================================================
## Fit for 4 latent classes:
## =========================================================
## number of observations: 928
## number of estimated parameters: 123
## residual degrees of freedom: 805
## maximum log-likelihood: -5456.164
##
## AIC(4): 11158.33
## BIC(4): 11752.79
## G^2(4): 2403.288 (Likelihood ratio/deviance statistic)
## X^2(4): 128311812 (Chi-square goodness of fit)
##

#As before, even if the optimal value of the BIC suggests choosing four classes, w
e prefer considering three for interpretation purposes. Moreover, when using four
classes, there are small groups involving only around 9% and 12% of the population
, which may be not so relevant for this analysis.

lca.se <- poLCA(f, se, nclass = 3, graphs = TRUE)
```




```
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $v72
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0248 0.1924 0.6235 0.1593
## class 2:  0.0560 0.2081 0.6320 0.1038
## class 3:  0.0000 0.0057 0.1829 0.8114
##
## $v73
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0216 0.2465 0.6171 0.1148
## class 2:  0.0492 0.3547 0.5657 0.0304
## class 3:  0.0040 0.0156 0.1609 0.8195
##
## $v74
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0547 0.2030 0.6275 0.1147
## class 2:  0.0579 0.2495 0.6166 0.0760
## class 3:  0.0000 0.0020 0.1080 0.8899
##
## $v75
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0038 0.0219 0.3780 0.5964
## class 2:  0.0293 0.1652 0.6602 0.1452
## class 3:  0.0020 0.0000 0.0111 0.9869
##
## $v76
```

```
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0000 0.1528 0.8472
## class 2:  0.0291 0.1388 0.7396 0.0925
## class 3:  0.0000 0.0184 0.0255 0.9561
##
## $v77
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0000 0.0794 0.9206
## class 2:  0.0175 0.0408 0.7246 0.2172
## class 3:  0.0000 0.0020 0.0023 0.9957
##
## $v78
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.0153 0.1483 0.8364
## class 2:  0.0116 0.1249 0.7816 0.0819
## class 3:  0.0000 0.0113 0.0354 0.9533
##
## $v81
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0039 0.0025 0.2715 0.7221
## class 2:  0.0233 0.0643 0.7480 0.1643
## class 3:  0.0000 0.0046 0.0477 0.9477
##
## $v83
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0230 0.0947 0.4957 0.3866
## class 2:  0.0355 0.1574 0.6668 0.1403
## class 3:  0.0020 0.0273 0.2827 0.6880
##
## $v141
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.8900 0.0358 0.0143 0.0598
## class 2:  0.8449 0.0865 0.0251 0.0436
## class 3:  0.9617 0.0099 0.0020 0.0263
##
## Estimated class population shares
##  0.2767 0.1851 0.5382
##
## Predicted class memberships (by modal posterior prob.)
##  0.278 0.1789 0.5431
##
## =========================================================
## Fit for 3 latent classes:
## =========================================================
## number of observations: 928
## number of estimated parameters: 92
## residual degrees of freedom: 836
## maximum log-likelihood: -5623.011
##
## AIC(3): 11430.02
## BIC(3): 11874.66
```

```
## G^2(3): 2736.983 (Likelihood ratio/deviance statistic)
## X^2(3): 88426287834 (Chi-square goodness of fit)
##
```

The profiles of the Swedish population seems to be totally different from the Hungarian one.

Considering class one, which represents more than 53% of the population, and taking aside the statement v141 concerning equal rights between men and women (almost everyone agrees), respondents show a strong and total disagreement with all statements. This seems quite a huge step toward a new balance characterized by equality both in the public and private sphere, with exchangeable parental roles, both at home and in the job market.

Class two, corresponding to around 28% of respondents, shows strong and moderate disagreement with all questions as well: the difference lays in the intensity of the feeling, in particular toward statements like v72, v73, v74, and v83.

The last class, which characterizes the profiles of 18.4% of respondents, sees as dominant position the moderate disagreement one, for all statements (but v141). Nevertheless, around 16% believes men make better political leaders (v76), 40% that women desire children and house above all (v73), more than 30% that family life suffers if a woman works full-time (v74), and more than 26% that a child suffers (v72).

The outcome of the analysis of Swedish society is not surprising: as all Scandinavian countries, Sweden is one of the states that made the greatest progresses toward gender equality, reducing the pay gap, introducing comparable parental leaves and generous family and educational policies that aim to make the roles of both men and women equal in every aspect of everyday life. It is possible to consider Swedish society as in a transitional state, moving in the direction of a new balance, granting women equality on public and private sphere.

CONCLUSION: advantages and disadvantages of LCA

Latent class analysis is a person-oriented approach which captures very well the diversity of people even when classifying them into groups. Indeed, it assigns probabilities of having all kind of opinions given the belonging to a certain class. This is a great advantage of the method since it allows to go into details and stratify the latent groups. Thanks to this, it was possible to outline the stages toward a no longer "incomplete" revolution in various segments of the population, and to understand how commonly spread are beliefs of the first, second and third phase in each latent class. The exploratory LCA worked well because there were lot of data and few missing values, which were removed in advance, even though there is the possibility for LCA in the "poLCA" package to remove them automatically. Nevertheless, some critical aspects and disadvantages need to be considered. First of all, no previous pilot study was done to establish whether the chosen statements measured exactly the emancipation of women. Indeed, the original goal of the questions was slightly different: understanding family life and models in Europe. LCA needs a strong theorical base in order to be applied correctly and using the right indicators is also an essential part of this process: finding that the chosen indicators (which however were selected based on previous studies and not randomly) do not exactly represent the concept would undermine all results. Independently from this aspect, another disadvantage typical of all surveys is the social desirability bias, which imply that the subject may answer differently from their beliefs and give a response which is more socially acceptable and well-seen. In Likert scales, respondents may also want to avoid extreme positions, which introduce the so-called central tendency bias. Another element of disturb in the analysis might be the manual rescaling done at the beginning so that all the answers were encoded in a Likert scale from one to four, taking out respondents which showed neutral positions in questions v81 and v83. This was done due to the need of having all statements in the same scale, however there is the risk this may have distorted the sample. An important factor that could not be taken into account was the stratification by socio-economic and educational levels, which would have provided a more complete view of the situation.