Multivariate Analysis
Exercise 2

Denaldo Lapi, Samy Chouiti, Francesco Aristei

April 24, 2022

Contents

1 Introduction

The goal of this report is to describe our solutions and interpretations of the three problems of Exercise 2 of the Multivariate Analysis course. We solved three problems using the R software.

The problems are related to the following methods:

- Multiple Factor Analysis
- Correspondence Analysis for Textual Data Analysis
- Association Rules

1.1 R scripts and packages

We built 3 R notebooks (*.Rmd* files), one per each problem, in which we describe in details all the steps of our analysis. For each problem we follow the same notebook structure:

- Data loading and exploration: we load the data of interest and we investigate it
- Data preprocessing: apply some preprocessing steps to the loaded data in order to make it compatible for the problem to solve and in order to improve the quality of the dataset
- Execution of the multivariate analysis method: we apply the algorithm of interest among the three mentioned before
- Results interpretation and plots: describe the results obtained after applying the algorithm

In particular, we mainly used 2 R packages to solve the excercise:

- *FactoMinerR*: used to apply both MFA and CA
- *Arules*: used for association rules mining

We also used other R packages in order to obtain better visualizations, such as *factoextra* which allows to do plots about CA and MCA results.

1.2 Some notes about the report

We would like to remark that the 3 provided notebooks contain a detailed analysis of each one of the problems, with all the steps that we followed to execute and interpret each analyzed method. In this report we'll describe only the most meaningful steps and results we obtained.
We also added an appendix in the end of the report containing the most important plots, so that this document could be able to provide a self-contained analysis of the 3 problems. However, we strongly suggest to read the R notebooks for a more detailed and clear explanation of all the solutions.

2 Multiple Factor Analysis (MFA)

The first problem required us to apply MFA to the orange dataset, included in the *missMDA* package. The dataset consists of sensory description of 12 orange juices by 8 attributes. Some values are missing. Rows represent the different orange juices, columns represent the attributes. We had to consider two groups of quantitative variables, with the first five attributes for group 1 and the last three attributes for group 2.

2.1 Preprocessing

As first step we started doing some preprocessing on the dataset. First we checked the presence of NA values. We found out some missing values, so we proceeded in removing them from the dataset, specifically, we substituted the NA values with the mean of the column. Now that we have treated the missing values, it is necessary to choose the group in which we want to divide the columns (variables) of the dataset. Therefore, we divided it in two groups of quantitative variables:

- Color.intensity, Odor.intensity, Attack.intensity, Sweet and Acid
- Bitter, Pulp, Typicity

The first group gives us an idea about the intensity of the orange juice (regarding color, odor and attack) and a description of their acidity and sweetness. While the second is used to give a description of bitterness, the amount of pulp, and the typicity of each orange juice.

Given that we are performing MFA on the dataset, it is necessary to scale to unit variance each quantitative variable. The *MFA()* method provided by *FactoMiner* does it for us, it is in fact sufficient to specify the parameter: *type = 's'* when applying it.

2.2 MFA Implementation

Once the preprocessing phase is completed, we applied the MFA algorithm.

2.2.1 Groups of Variables

Here, we analyzed the groups of variables, interpreting the graph representing them. We used the function *get_mfa_var()* to extract the results. This function returns a list containing the coordinates, the cos2 and the contribution of each group. From the plot: (Figure 1) we can observe that the groups give a similar representation of the individuals, in fact, they are close to each other. Specifically, we see that for both, the value on the first dimension is close to 1 (around 0.8) which means that the Lg (Lg(Kj, v1) : projected inertia of all the variables of j-th group on v1) value with the first dimension of the MFA is high for both the groups, so they contribute similarly to the first dimension. For the second dimension we have that the Taste group has a greater value, which means that it contributes more to it. To visualize better the contribution of each group to the dimensions we used two bar plots: (Figure 2) , (Figure 3)

Another useful plot to study in order to understand the relationship between each group of variables and the first two dimensions is the plot of the partial axes: (Figure 4)

The graph of partial axes shows the relationship between the principal axes of the MFA and the ones obtained from analyzing each group. From the graph obtained after applying the MFA, we observe that the first dimension of the Intensity group is very much related to the first dimension of the global MFA, while it is pointing in the opposite direction with respect to the second dimension of the global MFA, which means that in this dimension the description provided by such group of variables, diverge from the one of the global MFA. For the Taste group instead, we have that the first dimension, it is pointing in opposite direction with respect to the first dimension of the global MFA, so for this dimension, it has a description which will be different from the one of the global MFA. While it is highly correlated in the second dimension with the global MFA. The graphical results is confirmed by the $coord variable obtained in the res$partial.axes. In fact, Dim1.Intensity and Dim1 have a value of 0.89, highlighting that they point in the same direction, while Dim1.Taste and Dim1 have a negative value of -0.89. At the same time Dim2.Intensity and Dim2 have a value of -0.74 while Dim2.Taste and Dim2 are highly correlated, with a value of 0.89.

2.2.2 Variables

In this section, we studied the similarities between the different attributes (variables). First we observed that in the data set given, all the variables are of quantitative type. The first plot we studied was the one representing the correlation between variables and dimensions. Briefly, the graph of variables (correlation circle) shows the relationship between variables, the quality of the representation of variables, as well as, the correlation between variables and the dimensions. Positive correlated variables are grouped together, whereas negative ones are positioned on opposite sides of the plot origin (opposed quadrants). The distance between variable points and the origin measures the quality of the variable on the factor map. Variable points that are away from the origin are well represented on the factor map. For a given dimension, the most correlated variables to the dimension are the one close to it in the graph. We decided to analyze the plot coloring the variables according to three criteria:

- The group to which they belong: (Figure 5)
- Their quality of representation on the factor map: (Figure 6)
- Their contribution to the dimensions: (Figure 7)

In this case we observe that variables like Color.Intensity, Acid and Attack.Intensity are highly correlated, and are the ones who are more correlated to the first dimension. Then the first dimension represents essentially the color intensity and the acidity of the orange juices showing a relationship between the color of the orange and the acidity of it. The second dimension instead is represented mainly by the Odor.Intensity attribute and the bitterness and pulp attributes. Moreover the Bitter and Odor.Intensity attributes are close in the graph, which may indicate a relationship between the intensity of the odor and the bitterness of the orange juice. We can also observe how opposite attributes are coherently represented in the plot. For example, Sweet and Bitter, which represent opposite tastes, are in opposite quadrants.

2.2.3 Individuals

Finally we plotted the individuals (orange juices), analyzing their relationship with the groups of variables and the similarities between each other. From the plot: (Figure 8)

Individuals with similar profiles are close. The first dimension, opposes wines 12 and 4 against for example 3, 2 and 11. As explained before, the first dimension is more associated with the Color.Intensity and the acidity of the juice. Therefore, the orange juices with number 12 and 4 are the one having stronger color and acid taste, while the 3, 2 and 11 have high typicity, being in the same direction of the Typicity attribute. The second dimension is mostly correlated with the orange juice number 5. The second dimension describes the juices having the strongest odor and a bitter taste. The orange juices 7 and 8 for example, are far from where the second dimension is pointing, which may indicate that they are the one having less odor and bitter taste, but a more sweet taste. Regarding the sweetness, orange juices 10 and 1 lay in the direction of the sweet attribute, which may indicate that they are the most sweet orange juices.

After having considered the individuals as seen by every group, it may be useful to inspect the graph of the individuals for single groups. The results for individuals obtained from the analysis performed with a single group are named partial individuals. In other words, an individual considered from the point of view of a single group is called partial individual.

In the default fviz_mfa_ind() plot, for a given individual, the point corresponds to the mean individual or the center of gravity of the partial points of the individual. That is, the individual viewed by all groups of variables.

For a given individual, there are as many partial points as groups of variables.

The graph of partial individuals represents each juice viewed by each group and its barycenter. Here the plot of the partial points of all individuals:

(Figure 9)

We can inspect each orange juice as seen by each group. For example we can see that orange juice n.4 as described before, has high values of acidity and color intensity, and this characteristics are mostly assigned by the Intensity group of variables, while the Taste group tends to give a description of it less related to such attributes. Another example is for instance orange juice n.11 to which the Taste group assigns high values in the second dimension, describing it as a bitter juice, with a strong odor, while the Intensity group tends to give an opposite opinion with regards to these attributes. For the other points the reasoning proceeds in the same way. Generally, we observe that the descriptions given by the two groups, are not so aligned with the mean description given by the global MFA, in fact the partial points, tend to be distant from the mean points. Moreover, the partial points tend to be often distant from each other, which mean that generally, the two group of variables have different opinions about the attributes to describe the orange juices. The only orange juices in which the partial descriptions made by the two groups of variables are similar to the mean one are the number 2, 5 and 12.

3 Correspondence Analysis (CA) for textual data analysis

The second problem required as to apply the Correspondence Analysis technique for textual data to the "words_english" frequency table, given to us as a *.txt* file.

3.1 Dataset

As a first phase we just explored the frequency table to get some precise insights about its main structure.

The table is related to a survey conducted by a railway company to know the opinion and satisfaction of their passengers concerning high-quality night rail service. Passengers were asked to rate their satisfaction about 14 different aspects related to comfort (general, cabin, bed, seat), cleanliness (common areas, cabin, toilet), staff (welcome attention, trip attention, language skills) and others (cabin room, air conditioning, punctuality, general aspects). Each aspect was scored on a 11 point Likert scale from 0 (very bad) to 10 (excellent). Additionally, an open-ended question was added to the questionnaire asking for the aspects that should be improved. This question required free and spontaneous answers in English. Respondents $\times$ words frequency table was built following the classical preprocessing steps. Stop words were used and lemmatization from plural to singular form was performed. Only the words used at least 5 times among all the answers were kept. Thus, 60 distinct words and 829 occurrences were kept for 274 respondents.

Therefore, the analyzed frequency table is composed by 274 rows (respondents) and 60 columns (words).

3.2 CA and results

We applied CA to the dataset by using the *CA* function of the *FactoMineR* package. All the detailed steps of the followed procedure are described in the attached notebook "CA_textual_data.Rmd".

We then analyzed in details all the outputs of the performed CA and we gave our interpretations by answering the required questions.

We report below only the most representative parts of the R notebook, i.e. the ones useful to answer the list of the required questions and the way we answered them.

3.2.1 Which percentage of variability is explained by the first two dimensions?

The first aspect we typically look at in FA methods are eigenvalues, i.e. we try to understand how the newly created axes/dimensions are able to capture the deviation from independence, i.e. how the overall inertia of our cloud of points is explained by the new axes.

Besides printing the list of eigenvalues and their explained variance and cumulative variance, we graphically visualized the variance explained by each dimension by means of a Scree Plot, shown in figure 10.

From the obtained graph and from the displayed inertia values, we easily concluded that the first 2 dimensions explain 6.78% of the variability of the entire cloud of points. Since the performed CA builds 59 dimensions, as expected the inertia is spread along all the components.

This means that in order to do a complete analysis, we'll need to take into consideration also the other axes, besides the first 2: for instance we can see that the variance captured by the 3rd and 4th dimensions is around 6.2%, which is still a pretty high value, compared to the one of the 1st factorial plane.

3.2.2 Similarities between words

After having performed the full biplot of the CA analysis (simultaneous representation of rows and columns), we then analyzed separately the words, i.e. the column points of our frequency table.

In order to interpret the similarities among words, we first plotted them on the first factorial plane, which groups together words with similar profiles. In order to have a clear visualization, we plotted only the 30 words most contributing to the construction of the first two axes, as we can see in the figure 11 in the appendix. We know that the distance between the points and the origin measures the quality of the column points (i.e. words) on the factor map. Column points that are away from the origin are well represented on the factor map. What we can see, in this case, is that many points are very closed to the origin and so we'll need to look also at the other axes.

In particular, this plot shows how the first factorial plane captures the variance in the column points, and it also shows the relationships between the words.
What we can see is:

- the 1st axes separates (on the right) words regarding the staff of the railway company from the rest of the words. Indeed on the right of the plot we have a cluster of words such as "staff", "english", "speaking", "crew", "speak", which are all related to the staff and, in particular, to their language skills.

- while on the left side of the axes we have words related to other aspects of the quality of the night-rail service of the company, such as words related to "comfort" ("seats", "cabins", "beds"), and also to "cleanliness" ("cleanliness", "toilets").

- the 2nd axes clearly separates the word "less" (with very high coordinate value) from the rest of the words. Also the word "night" is pretty far from the others. For what regards the other words, almost all of them have very small coordinate value in the 2nd axes.

- We can identify other groups of similar words. For instance, we can spot some words related to food, which includes words such as "need", "food", "dining", grouped together according to the 1st dimension. Also the words "larger", "bigger", seats", "space" clearly identify a group related to a dimensional aspect. Another interesting similarity is the one among the words "prices", "ticket".

- However, the groups we identified around the origin of the axes are not clearly understandable, since the points are very closed to the origin. The only clearly identifiable groups of similar words in this 1st factorial plane are the one including words related to the staff, and the one regarding the "food" aspect.

  That's why we thought to be useful to study word similarities also by looking at the biplot with the 3rd and 4th axes, which still captures more than 6% of the overall variance. The plot is reported in the figure 12 in the appendix section.

The main relationships we spotted in this case are:

- The 3rd axes creates clusters of similar words depending on the quality aspect of the railway service they are related to: for instance, we can clearly see the words "conditioning", "air", cold" visualized very close to each other (very similar coordinate value on the 3rd axes), and we know they are related to the "air conditioning" aspect.

  Also words related to "comfort" are close to each other: "size", "space", "rooms". In particular, notice that all these words are related to dimension features.

In the lower part of the 3rd axes (with negative coordinates) we have words related to economical aspects, in particular the words "prices" and "ticket".

- By focusing on the 4th axes, we can notice other clusters of words, such as the words "cleanliness", "clean", "toilets", all related to the "cleanliness" aspect.
- A very interesting property of this plot, related to the grouping of the 'similar' words, is related to the separation between the words "sleeping" and "bathrooms" which are in opposite sides w.r.t. the 4th axes: this indicates a strong distance among the 2 words which, as we may expect, do not usually appear together.

3.2.3 Similarities between respondents

In order to interpret the similarities among respondents, we first plotted them on the first factorial plane, which groups together respondents with similar profile, i.e. suggesting similar improvements to the railway company. However, this is not enough to capture similarities among them, since we should add to the plot also the words/columns: indeed this allows to group participants depending on the words they use inside their answers; in that way we can find similar participants, according to the words they used in their answers to the open-ended question.

In order to do that, we visualized side by side the 2 biplots (i.e. the one of the participants and the one of the words) by considering only the most contributing points (we could have plotted the biplot with the simultaneous representation of rows and columns, but it was not so clear due to the high number of overlapping).

The plot is shown in the figure 13 in the appendix section.

The main similarities among participants we spotted by looking at the 1st factorial plane are (pay attention to the different scales of the axes between the 2 plots):

- Participants 2788 and 2963 are closely related to each other in the position of the word "less", this means that their answers include that word.
- In the right side of the 1st axes, we have a group of participants who suggested to the company improvements related to the "staff": indeed these row points (such as 2776, 3194, 3020, 2961,...) are positioned in the same zone where are located also words related to the "staff" aspect.
- We can identify, for instance, a group of users (composed by 3107, 3272, 3233) which suggested improvements related to the "food" aspect.
- Participants 2682, 2899 used the words "night" and "dinner" in their free answers.

As we did for the relationships among the words, we repeated the same analysis by visualizing the plot of the 3rd and 4th dimensions, shown in figure 14. The main similarities we spotted in this case are:

- We have a group of participants with a negative coordinate in the 4th dimension that suggested improvements for what regards the "air conditioning" aspect: these users are, for instance, 2838, 3238, 3235, 3296, 3259, 3233).
- Participants 3202, 3190, 3180 are in the plot area corresponding to words related to the "cleanliness" aspect, this indicates that in their answers they suggest to the company to improve this aspect
- In the top of the 3rd axis, we have a group of participants that are related to the words "bathrooms", "door" (the participants with a coordinate value above 2)

3.2.4 What are the words best represented in the first factorial plane?

In under to understand how good a word is represented in the new dimensions obtained after applying CA, we need to analyze the so-called quality of representation of the columns in the new dimensions: indeed, the best represented words in the first factorial plane are those with the highest values of the squared cosine (cos2).

We know that the values of the cos2 of each row or column are comprised between 0 and 1: therefore the sum of the cos2 for row/column points over all the CA dimensions is equal to one. In our case, if a word is well represented by the first two dimensions, the sum of the cos2 is closed to one.

In order to understand the best represented words, besides visualizing the *cos2* for all the column variables, we built a bar plot showing the best represented columns (i.e. words) in the first factorial plane, composed by the first 2 dimensions. The plot is illustrated if the figure 15. We then easily concluded that the best represented words are, in order: "less", "english", "staff", "speak", "crew", "speaking", and so on.

Basically, these are the words related to the "staff" aspect (except for "less") and correspond to the points that in the 1st factorial plane are more distant from the origin of the axes.

4 Association rules mining

4.1 Introduction

The purpose of association rule mining is to analyse relations between elements subsets in a transaction set. In any given transaction with a variety of items, association rules are meant to discover the rules that determine how or why certain items are connected [1]. To so, it relies on different metrics: Support, Confidence and Lift; which all three will be used and interpretation will be made in order to understand their purpose and influence on the data.

4.2 Exploration of the transaction set

The transaction set that we are using here is containing 20082 transactions with 858 distinct items, with a length ranging from 1 (single-item transaction) to 145 with a mean of 29.16 items per transaction. The most frequent item is *1104010101* that can be found in more than half of transactions (0.51 of presence ratio).

Any transaction is represented as the following[2]:

Id	Itemset					
357	1101010101	1104020101	1206010101	1301010101	1301120101	1302060401

4.3 Support: finding most frequent item-sets

4.3.1 Introduction

In order to find the most frequent item of the transactions set, we can either count them then sort them or use the support metric that will be used again. The support metric, in general, gives the probability of co-occurrence of a list of items:

$$Supp(I_k) = \frac{T|I_k \subseteq T|}{Card(T)}$$

4.3.2 Application to our data

Because we have been asked to apply certain filters to the transaction set (support=0.001, confidence[3]=0.1 and maxlen[4]= 5), we could have expected the most frequent item-sets to be different with and without the filtering which was not the case as can be seen in the table.

4.3.3 Interpretation

The interpretation of the support is pretty straightforward having the formula: for example, the most frequent item (*1104010101*) has a support of 0.51 which means that 51% of transaction contains, which is more than half of them. Although we were supposed to study only the top 10 items, if we dig further, we can see that the 12th most frequent item-set is containing 2 items (*1301120101,1301120200* with a support of 0.26) which is actually composed of the 2nd and 5th

[1] Wikipedia contributors. (2022, February 4). Association rule learning. In Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Association$_r ule_l earningoldid$ = 1069959346

[2] First row of the *Transaction.txt* file used for the assignment. We noted that transaction id's were relatively sparse, thus we concluded that we had not the full original dataset.

[3] Detailed in next subsection

[4] Used to keep only transaction containing 5 or less items

	items	support
1	1104010101	0.51
2	1301120101	0.46
3	1104020101	0.44
4	1103010101	0.42
5	1301120200	0.37
6	1104040101	0.37
7	1101010101	0.30
8	1301010101	0.28
9	1103010103	0.27
10	1301110301	0.27
11	1101060101	0.27
12	1102060101	0.25

(a) Without filtering, single-item item-set

items	support
1104010101	0.51
1301120101	0.46
1104020101	0.44
1103010101	0.42
1301120200	0.37
1104040101	0.37
1101010101	0.30
1301010101	0.28
1103010103	0.27
1301110301	0.27
1101060101	0.27
1301120101,1301120200	0.26

(b) With filtering

Table 1: Item-set sorted by support

most frequent (single-item) item-sets. This can also be due to the fact that the a) table was obtained used basic computation of the presence of each element when the b) table was a result of the apriori analysis using proper support computation methods.

Besides that, the 10 most frequent item-sets are single-item sets, which means that they are more likely to appear than any other item-set (single or multiple).

4.4 Confidence: LHS and RHS presence ratio

4.4.1 Introduction

When the support metric used only the transaction set, the confidence will be using rules determined by the computation [5] using our previously presented filters. The confidence measure the ratio of presence of both LHS and RHS compared to the LHS alone, using the following formula:

$$Conf(LHS \rightarrow RHS) = \frac{Supp(LHS, RHS)}{Supp(LHS)}$$

4.4.2 Application to our data

	lhs		rhs	support	confidence
1	{1104050403, 1205010202}	=>	{1201070203}	0.00	1.00
2	{1104050403, 1205040101}	=>	{1206020501}	0.00	1.00
3	{1101130101, 1104050403}	=>	{1206020701}	0.00	1.00

Table 2: Top 3 rules sorted by confidence

[5] With the *Arules* R package.

4.4.3 Interpretation

As a consequence of the support formula, the fact that those 3 rules 2 have a confidence of 1 means that the LHS is never present in transactions without the RHS. But interestingly, none of the items of the above rules are present in the item table sorted by support 1, which can be confirmed by the very low support score of those rules. What can be interpreted here is that the items that those rules are composed of are **very rare** in the transactions set but when they are present, they always do with the rest of elements of the corresponding rule.

4.5 Lift: prediction score compared to random choices

4.5.1 Introduction

The last metric to study is the lift criteria which quantifies how much better the rule is than a random prediction and can be computed using the following formula:

$$Lift(LHS \rightarrow RHS) = \frac{Supp(LHS, RHS)}{Supp(LHS)Supp(RHS)}$$

A lift higher than 1 means that the rule is indeed reliable to make a prediction when a lift lower than 1 means that a random prediction is better than the rule, which makes it non-relevant.

4.5.2 Application to our data

	lhs		rhs	confidence	lift
1	{1302140800}	=>	{1302140700}	0.51	153.20
2	{1302140700}	=>	{1302140800}	0.34	153.20
3	{1203140500, 1205020300, 1206010301}	=>	{1201020404}	0.91	143.25

Table 3: Top 3 rules sorted by lift

4.5.3 Interpretation

We can see that those three rules have a very high lift (around 150) which means that we can reliably predict the right-hand side element (for example *1302140700* for the first rule) using this rule.

Interestingly, the third rule is made of 3 left-hand side elements (to interpret as being very likely to induce the presence of the right-hand side element in a transaction) and have a high confidence (0.91) which means that the RHS is very rare to be found without the LHS (from the definition of confidence), which confirm the reliability of this rules to predict presence of RHS, along with the lift score.

Also, we can see that the 2 first rules are composed of the same LHS and RHS but with reversed order (1302140700 and 1302140800). Although having the same lift score, which makes sense as the Lift function of LHS and RHS is symmetric, they have different confidence score. Because, the first rule have a better confidence than the second, we can interpret that the support of 1302140800 is higher than *1302140700* and thus that the former is more present in transaction sets than the latter.

5 Conclusions

5.1 Factor analysis methods

This second assignment allowed us to deepen our knowledge of factor analysis methods for dealing with multivariate data. In particular, we understood how those methods can be applied to analyze textual data and data organized as groups of variables.

During this exercise we had the possibility to improve our R skills and to deepen our knowledge of the *FactoMineR* package which provide a very simple and intuitive way to apply the various factor analysis methods.

In conclusion, we are very satisfied with what we learnt during this assignment, especially beacuse we know that principal component methods are a very useful preprocessing steps in data analysis, they are also important to get some useful insights about the relationships and associations among different variables, items, individuals in multivariate datasets.

5.2 Association rules mining

This assignment was offering an interesting dataset to perform association rules mining on a transaction set. We were able to make interpretations based on support (for item-sets) and confidence and lift (for rules). However, missing the real-world meaning of this data limits the range of conclusion that we can draw from our analysis although the interpretation we made above.

6 Appendix

6.1 MFA

Main plots of the MFA analysis.

Figure 1: Representation of the groups in the first two dimensions.

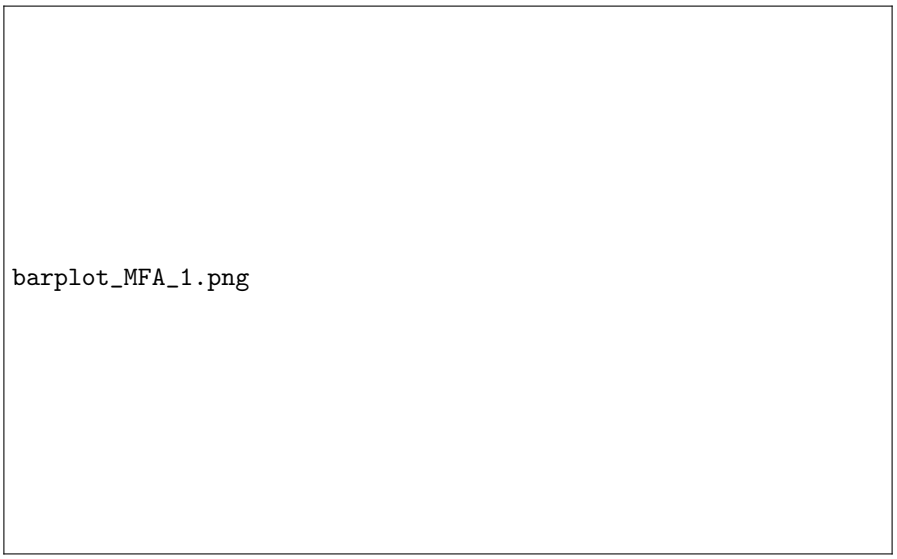

Figure 2: Barplot representing the contributions to the first dimension.

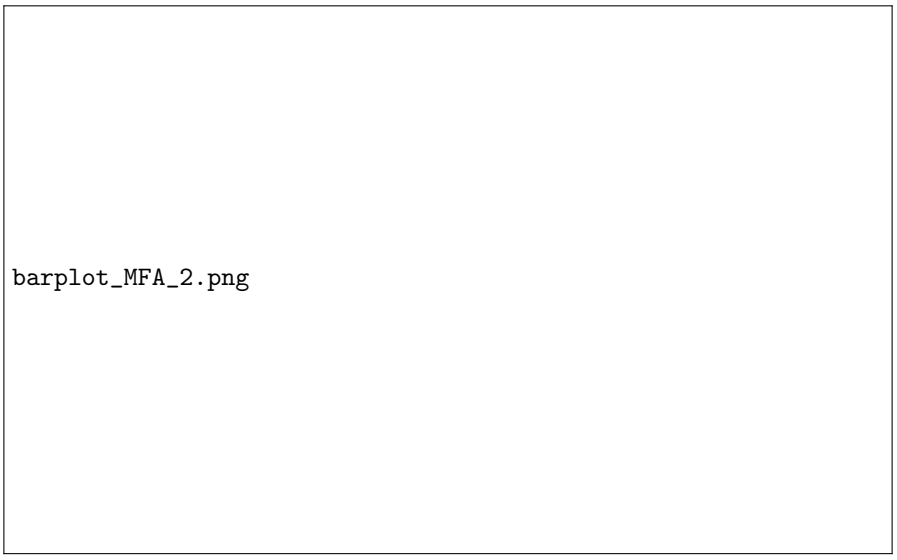

Figure 3: Barplot representing the contributions to the second dimension.

Figure 4: Partial axes plot.

Figure 5: Correlation circle with variables colored by groups.

Figure 6: Correlation circle with variables colored by contribution.

Figure 7: Correlation circle with variables colored by cos2.

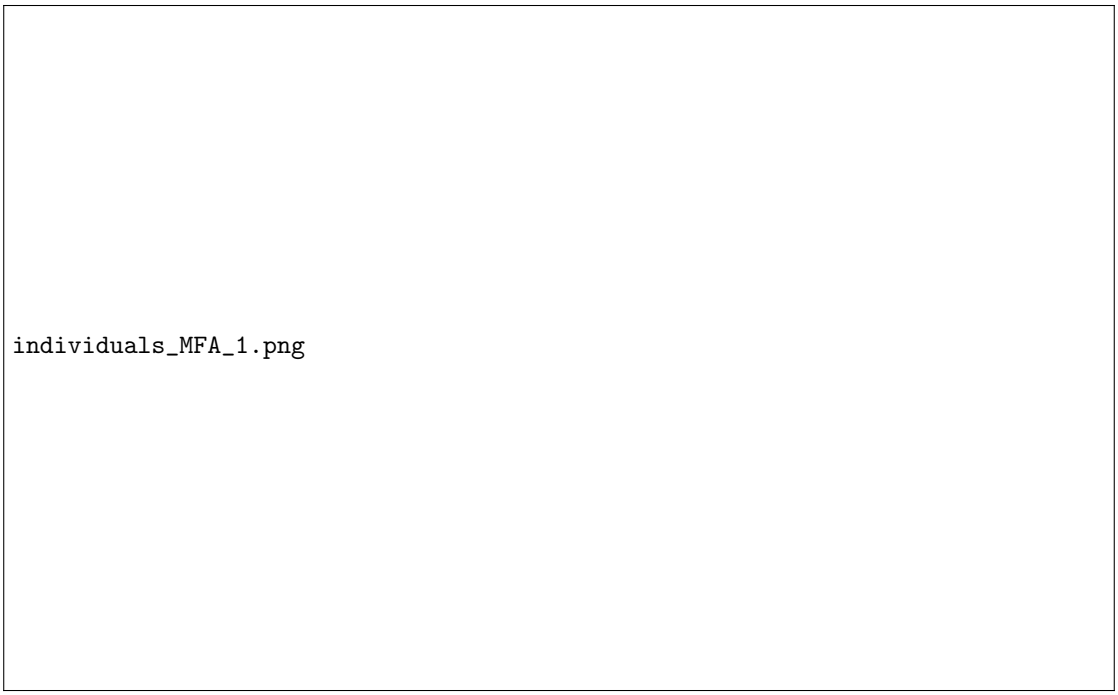

Figure 8: Plot of the individuals in the first two dimensions.

Figure 9: Plot of the points as seen by each group of variables.

6.2 CA for textual data analysis

Main plots of the CA analysis.

Figure 10: CA Scree Plot

Figure 11: Most contributing words to the first factorial plane

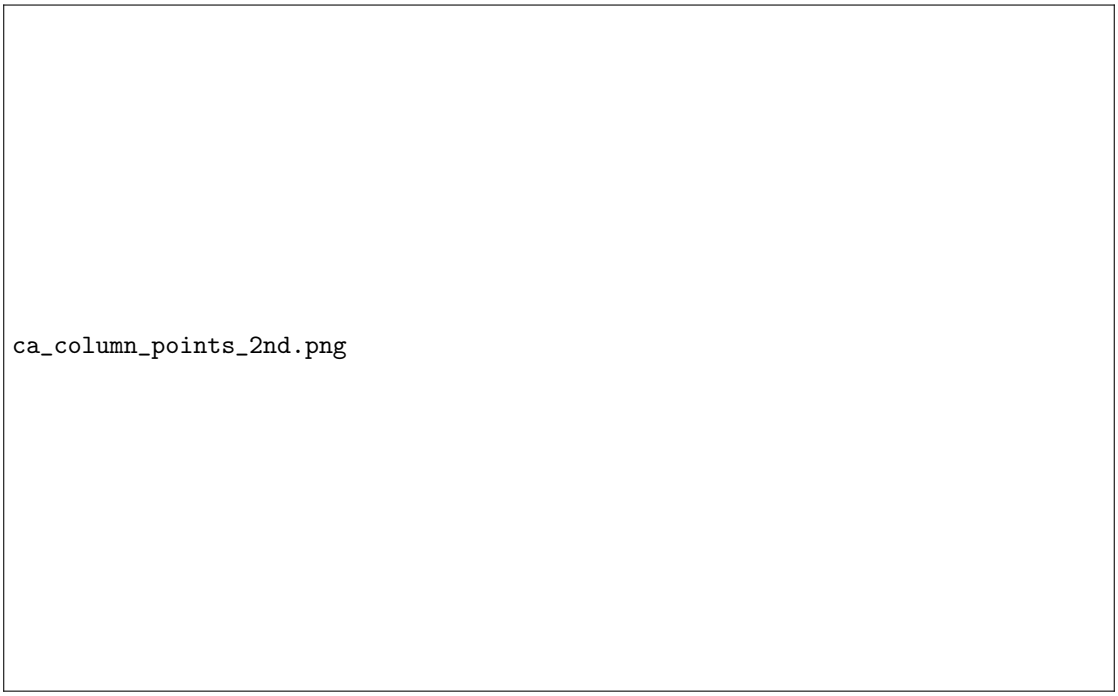

Figure 12: Most contributing words to the 3rd and 4th dimensions

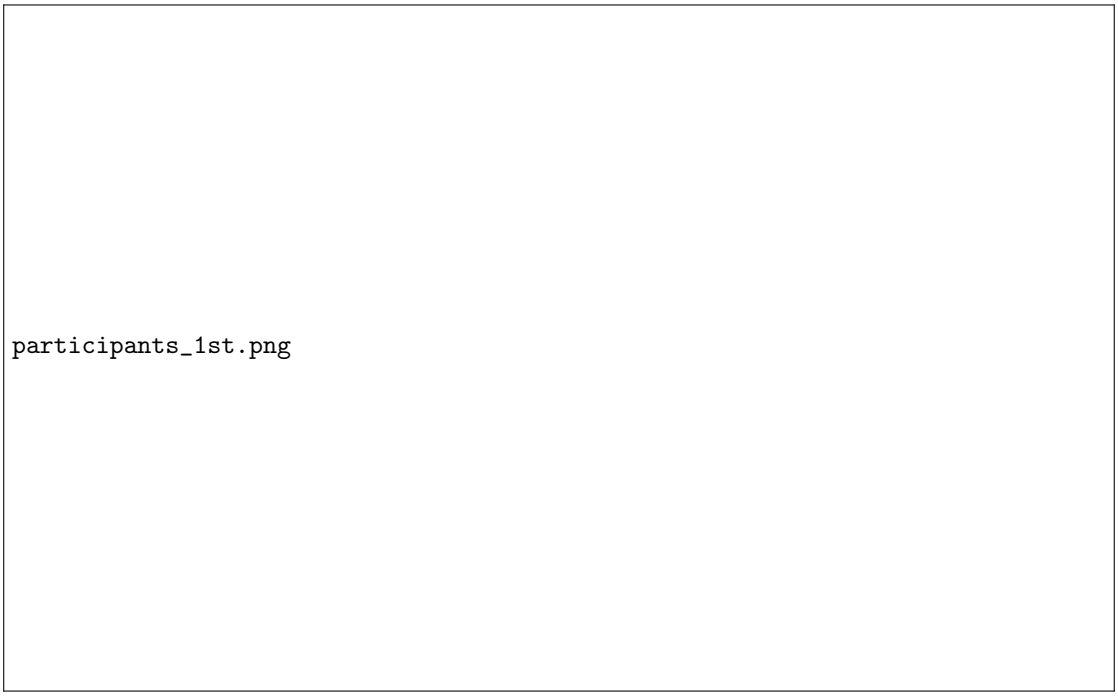

Figure 13: Words and participants in the 1st factorial plane

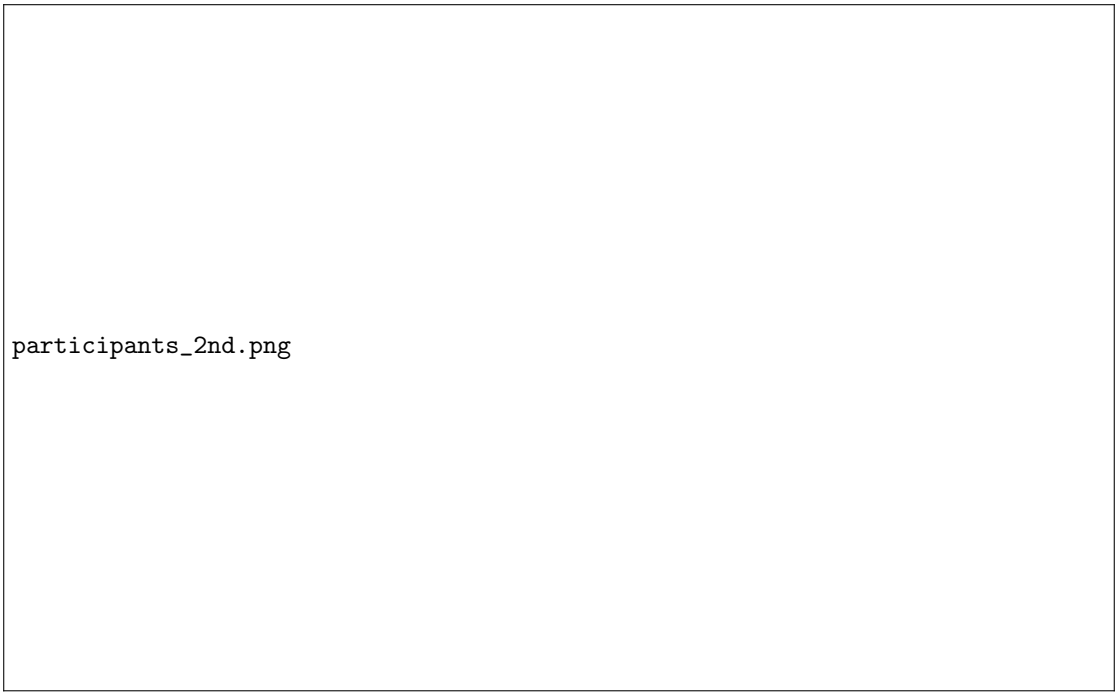

Figure 14: Words and participants in the 3rd and 4th dimensions

Figure 15: Quality of representation of words on the 1st factorial plane