Political Context Analysis

Data Mining

28/10/2021

Contents

Motivation and general description of the problem to be analyzed

The motivation for doing this project was primarily an urge to find easily accessible data. We thought public data published in Government websites would be a great place to find a massive amount of data with ease. We found on the Centre d'Estudis d'Opinió website a survey from 2020 on the political context in Catalonia. In this survey there are more than 50 questions about very different topics to contextualize the Catalan voter. These questions range from socioeconomical issues to ideology and national identity, so we thought having that amount of material would be a good way to start working on our data analysis project, because we could select whatever issues we found more interesting to draw our conclusions.

In newspapers we often find articles that analyze this sort of surveys, so now we had the opportunity to analyze the data ourselves with our own tools.

What we wanted to do with this data was to classify the surveyed in groups according to their answers. We had multiple questions that we wanted to find, but mainly we had three associations that we wanted to draw conclusions from: the relation between their ideology and voting party, voter participation and concern about the issues raised in the survey and opinion on the independence of Catalonia and national identity.

Data source presentation

Data collection process Our data comes from CEO (Centre d'Estudis d'Opinió). To get the data it's enough to access the two links above. In the first one you will find the data itself, after downloading the file "microdades_anonimitzades.csv", This data can be processed directly in RStudio without any additional steps. In the second one you will find the questionnaire from which the data have been obtained in pdf format.

We could have enlarged our database with older data (also from CEO) if we wanted to know how opinion has changed, but we decided against it as it isn't our main objective.

What are the data about The data responds to information about surveys conducted by the Centre d'Estudis d'Opinió to people residing in Catalonia. The aim is to extract specific information on the current opinion about the political, economic and social situation in Catalonia, as well as the attitude towards politics and electoral behavior.

A sample of 1200 respondents is provided, of which we have information on socio-economic and demographic data. These respondents were selected proportionally to be from the 4 provinces of Catalonia and from municipalities with different amounts of population and other different characteristics.

Formal description of Data structure and metadata

Complete Data Mining process performed

We started the process by getting the data from the Centre d'Estudis d'Opinió website. We downloaded a .csv file containing all the data from a 2021 survey on the political context in Catalonia. The first step after downloading the data, was to select the variables that we found more relevant and take out the columns from our matrix that were not important to our analysis. Every question of the survey was a variable, so we had to take a look at each one of them to see if they were useful for us. The second step, once we had selected the data we wanted to tread, was the preprocessing process. This was the most tedious part of the procedure, because we had a lot of variables and lots of them had missing values or the factors of the categorical variables were very long and difficult to read in our database. Each variable that had missing values was different to tread. We started with the categorical ones, which every one of them were very different. Mainly we grouped all the ambiguous responses into one group called unknown, although this is

not always the case because some "None of the above" responses were needed to take into account. This is shown in detail in the preprocessing section. Then we started treating the numerical variables. We had problems when applying the knn algorithm to our data using RStudio so, since we didn't have a lot of missings in these types of variables, we decided to use the function imputePCA. This function is normally used to prepare the data to perform the PCA process. Since this function worked and we had the approval of our professor, we decided to use it over knn which we had problems to execute. Having treated all the missing values, we decided to abbreviate some of the factors of our variables (categorical) because they were too long. Some of the questions of the survey had answers with long sentences, so we replaced them with one or two words instead so they would be easily readable in our database and, later on, in our graphs in the profiling section. Also, we decided to put an acronym of the corresponding variable before each factor of the database selected, because there were lots of answers to different questions that were yes, no, unknown, etc., so with that "prefix", the factors were more identifiable. When our data was clean, the next step was to start finding patterns. This was done in the clustering and PCA processes. On the first one, we created our dendrogram to find out how many clusters we had. We found out there were 4 groups, so during the profiling we contextualised and described each one of them, taking a selection of most relevant variables for our conclusions. Both procedures and their conclusions are explained in detail in the final sections of the project.

Preprocessing

Here we choose the variables with which we want to do the data analysis. We sort them by their numeric or qualitative nature.

In the input file we have imputed some missing values in select columns from values in the same column name plus **_LITERAL**, like for example: in the columns that refer to the vote in the *Congreso de los Diputados*, "PSOE" turns into "PSC", as we understand that that person wanted to vote for the Socialist Party, but when answering didn't take into account that, because of the nature of the Spanish electoral system, the "PSC" is the one that is voted for in Catalonia.

Then, we delete every **_LITERAL** column.

We replace the blank values with "NP" ("No Preguntat"), which will mean that the question wasn't even asked, probably because of previous answers, and that's why there's no answer.

In this chunk of code we calculate the percentage of missing values in each column and then in all the dataset:

```
## [1] "Percentage of missing values by variable > 0:"
```

```
##                            MUNICIPI                         SIT_ECO_CAT
##                           46.4166667                           1.6666667
##             SIT_ECO_CAT_RETROSPECTIVA            SIT_ECO_CAT_PROSPECTIVA
##                           1.6666667                           8.1666667
##                         SIT_ECO_ESP                      SIT_ECO_PERSONAL
##                           2.3333333                           0.3333333
##                         SIT_POL_CAT            SIT_POL_CAT_RETROSPECTIVA
##                           3.1666667                           2.7500000
##            SIT_POL_CAT_PROSPECTIVA                         SIT_POL_ESP
##                           9.1666667                           1.9166667
##                            CAT_ESP                            IDEOL_1_7
```

```
##                             1.7500000                             14.4166667
##                              CAT_0_10                              ESP_0_10
##                             4.9166667                             4.7500000
##                    MONARQUIA_REPUBLICA                        SIGNIFICA_ESP
##                            12.5833333                             5.5000000
##                  ACTITUD_INDEPENDENCIA                       VOT_DRET_DEURE
##                             6.4166667                             1.6666667
##                  ELECCIONS_IMPORTANCIA          SATISFACCIO_ELECCIONS_14F2021
##                             4.5000000                             4.6666667
##                         PART_PARLAMENT              REC_PARLAMENT_VOT_DECISIO
##                             1.0000000                            23.0000000
##                       REC_PARLAMENT_VOT                 REC_PARLAMENT_FERMESA
##                            14.4166667                             0.6666667
##          POST_PARLAMENT_DECISIO_RELACIO           POST_PARLAMENT_DECISIO_CRISI
##                             9.5000000                             9.6666667
##       POST_PARLAMENT_DECISIO_CORRUPCIO POST_PARLAMENT_DECISIO_CANVI_CLIMATIC
##                            10.4166667                             9.6666667
##            POST_PARLAMENT_DECISIO_COVID                      INT_PARLAMENT_VOT
##                             9.8333333                            27.3333333
##                PART_PARLAMENT_ANTERIORS            REC_PARLAMENT_VOT_ANTERIORS
##                             2.9166667                            32.5000000
##                          INT_CONGRES_VOT                           PART_CONGRES
##                            28.0833333                             2.5000000
##                          REC_CONGRES_VOT
##                            19.2500000
```

The dataset has 6.29% of missing values.

For one of the multichoice questions, we join the "NS" and "NC" values into "NP", a new column while delelting the two latter ones.

We need to parse the numeric variables, because the survey interviewers sometimes wrote more than the numeric value of the score. However, the numeric value is always in the first two characters, so we'll parse those as numeric.

Once everything else is treated, we'll convert again the blank spaces into "NP" and the *NA* to "Desconegut" or Uknown.

Now we fix some of the categories in the columns referring to political parties which wasn't fixed directly in the dataset, especially for "Catalunya en Comú Podem" and its member parties.

Next, we'll add a code representing the abbreviation of the variable name in each level of the categorical variables. This is because, later on, when we need to make a PCA and Profiling, we'll sometimes see the categories by themselves, and it'll be clearer and easier if we can immediately see which variable the categories belong to.

Here we'll impute the missing values in the numeric variables. In the course it was meant to do it with the KNN method, but we couldn't manage it, so we've used the method *imputePCA*, which is based in Principal Component Analysis.

```
## [1] "Before imputation: "
```

```
##       EDAT          CAT_0_10         ESP_0_10
##  Min.   :18.00   Min.   : 0.000   Min.   : 0.000
##  1st Qu.:39.00   1st Qu.: 5.000   1st Qu.: 0.000
##  Median :51.00   Median : 7.000   Median : 5.000
```

```
##  Mean   :51.54   Mean   : 6.366   Mean   : 3.996
##  3rd Qu.:65.00   3rd Qu.: 9.000   3rd Qu.: 6.000
##  Max.   :90.00   Max.   :10.000   Max.   :10.000
##                  NA's   :59       NA's   :57
##  SATISFACCIO_ELECCIONS_14F2021 POST_PARLAMENT_DECISIO_RELACIO
##  Min.   : 0.000                Min.   : 0.000
##  1st Qu.: 3.000                1st Qu.: 5.000
##  Median : 5.000                Median : 7.000
##  Mean   : 4.635                Mean   : 6.401
##  3rd Qu.: 7.000                3rd Qu.: 9.000
##  Max.   :10.000                Max.   :10.000
##  NA's   :56                    NA's   :114
##  POST_PARLAMENT_DECISIO_CRISI POST_PARLAMENT_DECISIO_CORRUPCIO
##  Min.   : 0.000               Min.   : 0.000
##  1st Qu.: 5.000               1st Qu.: 4.000
##  Median : 7.000               Median : 7.000
##  Mean   : 6.293               Mean   : 6.029
##  3rd Qu.: 8.000               3rd Qu.: 9.000
##  Max.   :10.000               Max.   :10.000
##  NA's   :116                  NA's   :125
##  POST_PARLAMENT_DECISIO_CANVI_CLIMATIC POST_PARLAMENT_DECISIO_COVID
##  Min.   : 0.000                        Min.   : 0.000
##  1st Qu.: 4.000                        1st Qu.: 4.000
##  Median : 6.000                        Median : 6.000
##  Mean   : 5.571                        Mean   : 5.557
##  3rd Qu.: 8.000                        3rd Qu.: 8.000
##  Max.   :10.000                        Max.   :10.000
##  NA's   :116                           NA's   :118

## [1] "After imputation: "

##       EDAT          CAT_0_10         ESP_0_10
##  Min.   :18.00   Min.   : 0.000   Min.   : 0.000
##  1st Qu.:39.00   1st Qu.: 5.000   1st Qu.: 1.000
##  Median :51.00   Median : 7.000   Median : 5.000
##  Mean   :51.54   Mean   : 6.338   Mean   : 4.018
##  3rd Qu.:65.00   3rd Qu.: 8.000   3rd Qu.: 6.000
##  Max.   :90.00   Max.   :10.000   Max.   :10.000
##  SATISFACCIO_ELECCIONS_14F2021 POST_PARLAMENT_DECISIO_RELACIO
##  Min.   : 0.00                 Min.   : 0.00
##  1st Qu.: 3.00                 1st Qu.: 5.00
##  Median : 5.00                 Median : 7.00
##  Mean   : 4.61                 Mean   : 6.36
##  3rd Qu.: 7.00                 3rd Qu.: 9.00
##  Max.   :10.00                 Max.   :10.00
##  POST_PARLAMENT_DECISIO_CRISI POST_PARLAMENT_DECISIO_CORRUPCIO
##  Min.   : 0.000               Min.   : 0.000
##  1st Qu.: 5.000               1st Qu.: 4.029
##  Median : 7.000               Median : 6.457
##  Mean   : 6.267               Mean   : 6.002
##  3rd Qu.: 8.000               3rd Qu.: 8.000
##  Max.   :10.000               Max.   :10.000
##  POST_PARLAMENT_DECISIO_CANVI_CLIMATIC POST_PARLAMENT_DECISIO_COVID
##  Min.   : 0.000                        Min.   : 0.000
```

```
##  1st Qu.: 4.000                      1st Qu.: 4.000
##  Median : 6.000                      Median : 6.000
##  Mean   : 5.533                      Mean   : 5.537
##  3rd Qu.: 8.000                      3rd Qu.: 8.000
##  Max.   :10.000                      Max.   :10.000
```

Once we've imputed, we check that there isn't any strange imputation and that the quantiles don't change too much. As we can see, it isn't the case, so we accept the imputation and the preprocessing as a whole.

Univariate Descriptive Analysis of our data

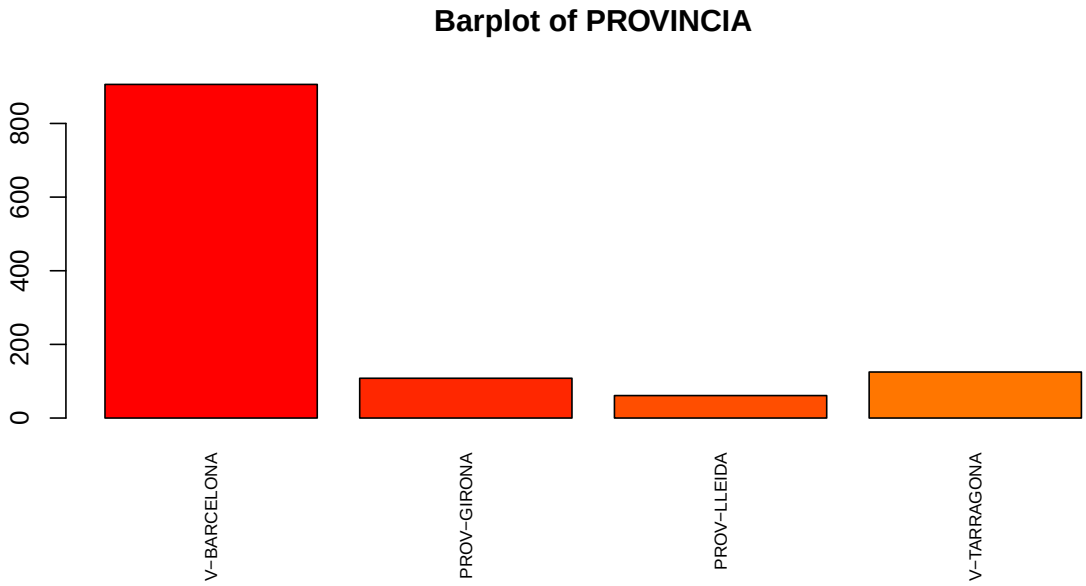


Table 1: Information about modalities of the Variable

Modalities	Frequency	Proportion
PROV-BARCELONA	906	0.7550000
PROV-GIRONA	108	0.0900000
PROV-LLEIDA	61	0.0508333
PROV-TARRAGONA	125	0.1041667

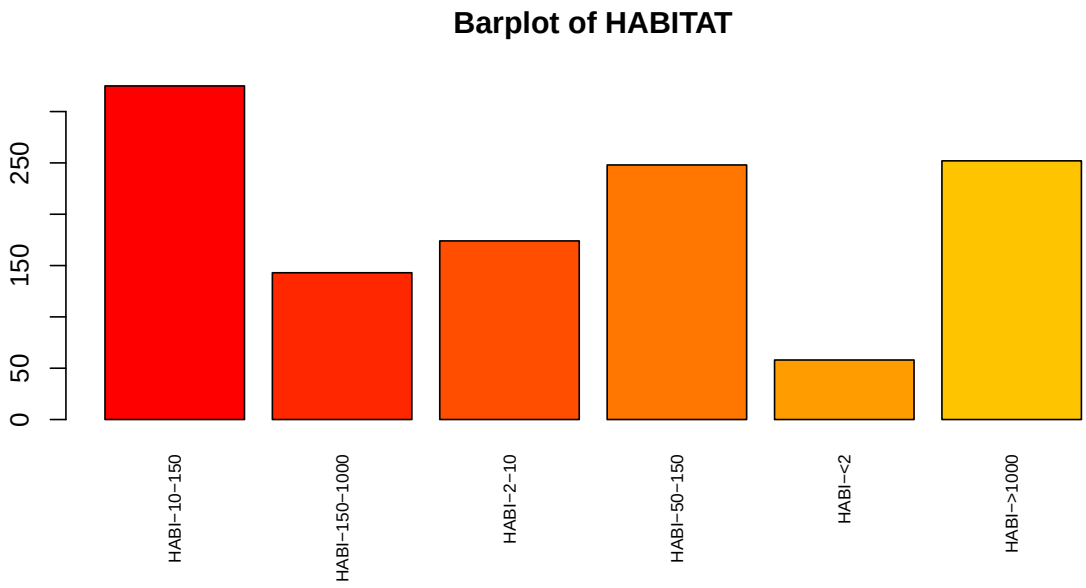


Table 2: Information about modalities of the Variable

Modalities	Frequency	Proportion
HABI-10-150	325	0.2708333
HABI-150-1000	143	0.1191667
HABI-2-10	174	0.1450000
HABI-50-150	248	0.2066667
HABI-<2	58	0.0483333
HABI->1000	252	0.2100000

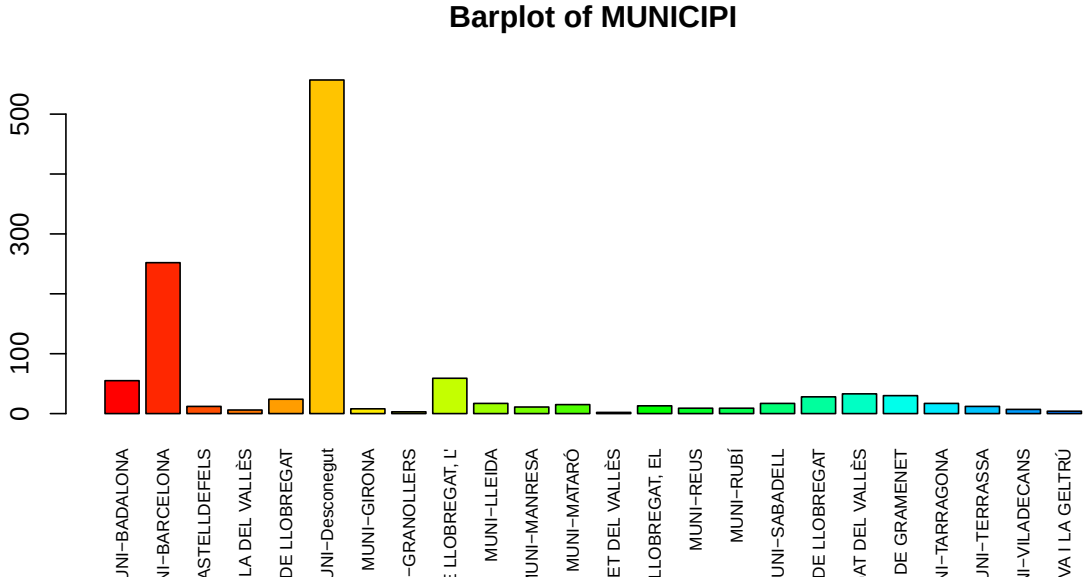


Table 3: Information about modalities of the Variable

Modalities	Frequency	Proportion
MUNI-BADALONA	55	0.0458333
MUNI-BARCELONA	252	0.2100000
MUNI-CASTELLDEFELS	12	0.0100000
MUNI-CERDANYOLA DEL VALLÈS	6	0.0050000
MUNI-CORNELLÀ DE LLOBREGAT	24	0.0200000
MUNI-Desconegut	557	0.4641667
MUNI-GIRONA	8	0.0066667
MUNI-GRANOLLERS	3	0.0025000
MUNI-HOSPITALET DE LLOBREGAT, L'	59	0.0491667
MUNI-LLEIDA	17	0.0141667
MUNI-MANRESA	11	0.0091667
MUNI-MATARÓ	15	0.0125000
MUNI-MOLLET DEL VALLÈS	2	0.0016667
MUNI-PRAT DE LLOBREGAT, EL	13	0.0108333
MUNI-REUS	9	0.0075000
MUNI-RUBÍ	9	0.0075000
MUNI-SABADELL	17	0.0141667
MUNI-SANT BOI DE LLOBREGAT	28	0.0233333
MUNI-SANT CUGAT DEL VALLÈS	33	0.0275000
MUNI-SANTA COLOMA DE GRAMENET	30	0.0250000
MUNI-TARRAGONA	17	0.0141667
MUNI-TERRASSA	12	0.0100000
MUNI-VILADECANS	7	0.0058333
MUNI-VILANOVA I LA GELTRÚ	4	0.0033333

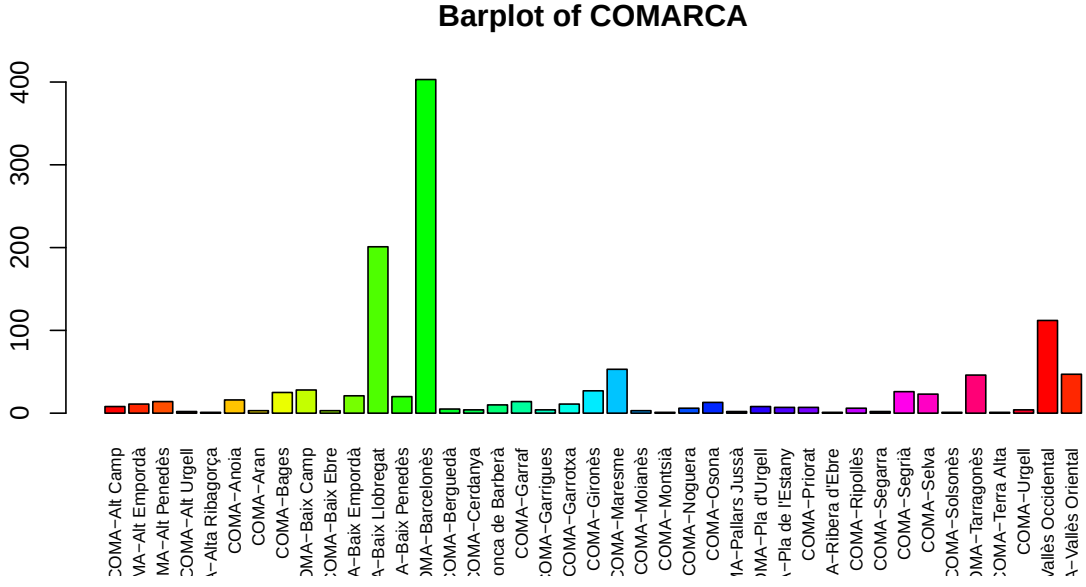


Table 4: Information about modalities of the Variable

Modalities	Frequency	Proportion
COMA-Alt Camp	8	0.0066667
COMA-Alt Empordà	11	0.0091667
COMA-Alt Penedès	14	0.0116667
COMA-Alt Urgell	2	0.0016667
COMA-Alta Ribagorça	1	0.0008333
COMA-Anoia	16	0.0133333
COMA-Aran	3	0.0025000
COMA-Bages	25	0.0208333
COMA-Baix Camp	28	0.0233333
COMA-Baix Ebre	3	0.0025000
COMA-Baix Empordà	21	0.0175000
COMA-Baix Llobregat	201	0.1675000
COMA-Baix Penedès	20	0.0166667
COMA-Barcelonès	403	0.3358333
COMA-Berguedà	5	0.0041667
COMA-Cerdanya	4	0.0033333
COMA-Conca de Barberà	10	0.0083333
COMA-Garraf	14	0.0116667
COMA-Garrigues	4	0.0033333
COMA-Garrotxa	11	0.0091667
COMA-Gironès	27	0.0225000
COMA-Maresme	53	0.0441667
COMA-Moianès	3	0.0025000
COMA-Montsià	1	0.0008333
COMA-Noguera	6	0.0050000
COMA-Osona	13	0.0108333
COMA-Pallars Jussà	2	0.0016667
COMA-Pla d'Urgell	8	0.0066667
COMA-Pla de l'Estany	7	0.0058333
COMA-Priorat	7	0.0058333
COMA-Ribera d'Ebre	1	0.0008333
COMA-Ripollès	6	0.0050000
COMA-Segarra	2	0.0016667
COMA-Segrià	26	0.0216667
COMA-Selva	23	0.0191667
COMA-Solsonès	1	0.0008333
COMA-Tarragonès	46	0.0383333
COMA-Terra Alta	1	0.0008333
COMA-Urgell	4	0.0033333
COMA-Vallès Occidental	112	0.0933333
COMA-Vallès Oriental	47	0.0391667

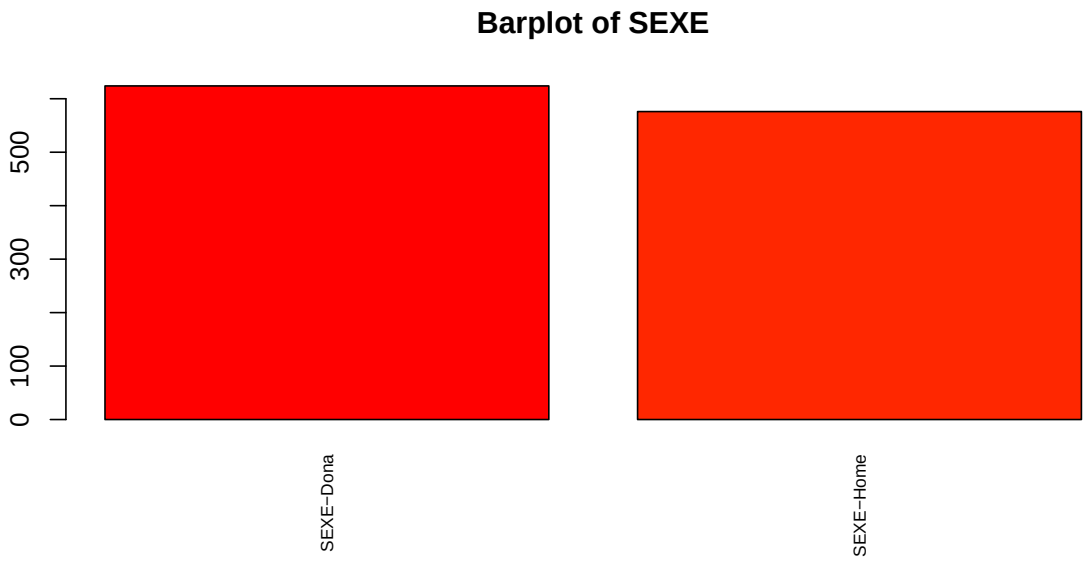


Table 5: Information about modalities of the Variable

Modalities	Frequency	Proportion
SEXE-Dona	624	0.52
SEXE-Home	576	0.48

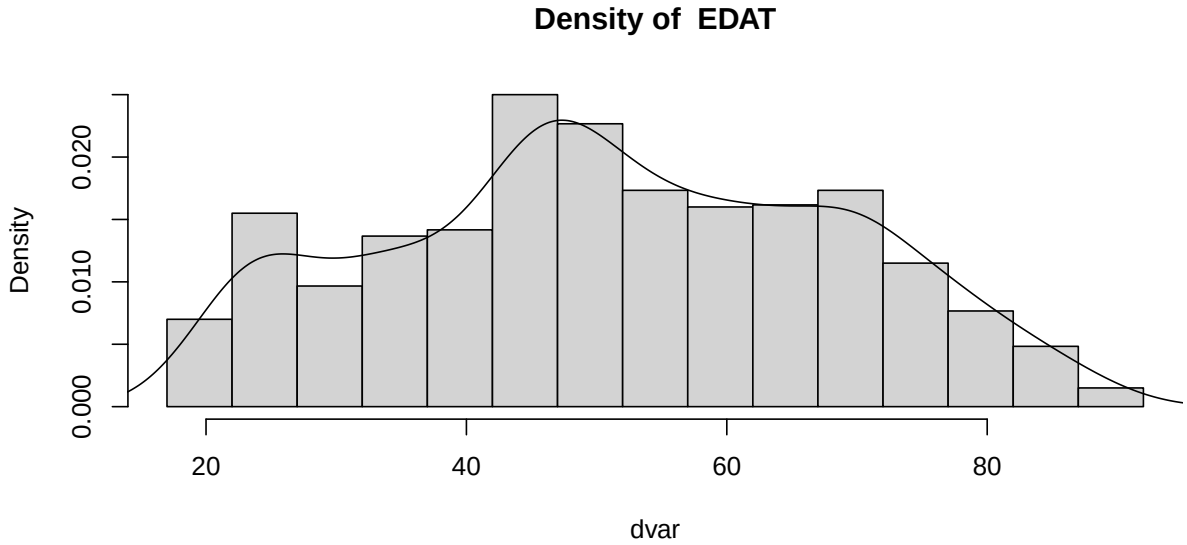


Table 6: Summary of the Variable

	variable	Min.	X1st.Qu.	Median	Mean	X3rd.Qu.	Max.
Min.	EDAT	18	39	51	51.5375	65	90

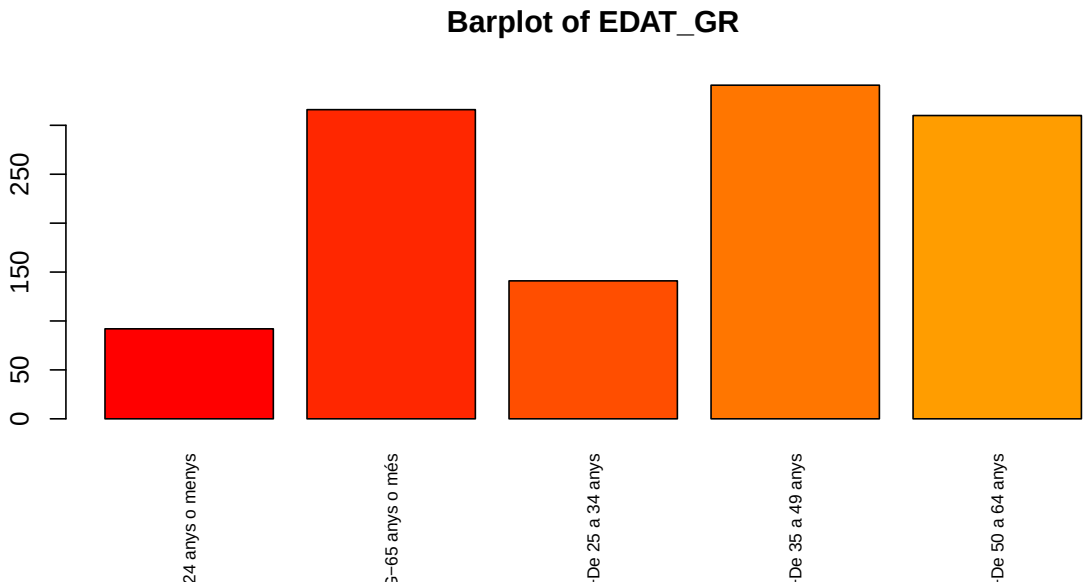


Table 7: Information about modalities of the Variable

Modalities	Frequency	Proportion
EDG-24 anys o menys	92	0.0766667
EDG-65 anys o més	316	0.2633333
EDG-De 25 a 34 anys	141	0.1175000
EDG-De 35 a 49 anys	341	0.2841667
EDG-De 50 a 64 anys	310	0.2583333

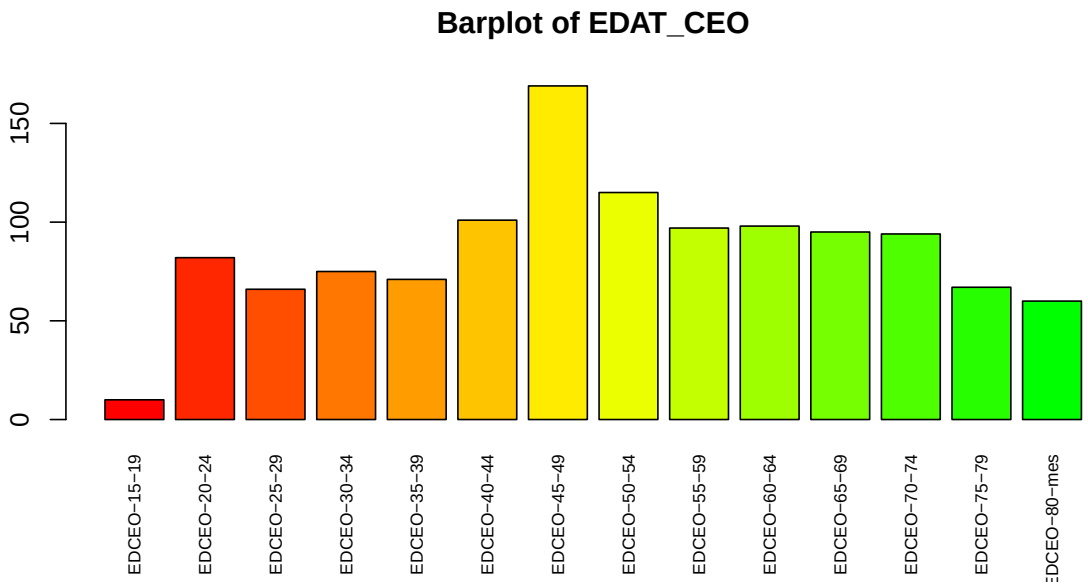


Table 8: Information about modalities of the Variable

Modalities	Frequency	Proportion
EDCEO-15-19	10	0.0083333
EDCEO-20-24	82	0.0683333
EDCEO-25-29	66	0.0550000
EDCEO-30-34	75	0.0625000
EDCEO-35-39	71	0.0591667
EDCEO-40-44	101	0.0841667
EDCEO-45-49	169	0.1408333
EDCEO-50-54	115	0.0958333
EDCEO-55-59	97	0.0808333
EDCEO-60-64	98	0.0816667
EDCEO-65-69	95	0.0791667
EDCEO-70-74	94	0.0783333
EDCEO-75-79	67	0.0558333
EDCEO-80-mes	60	0.0500000

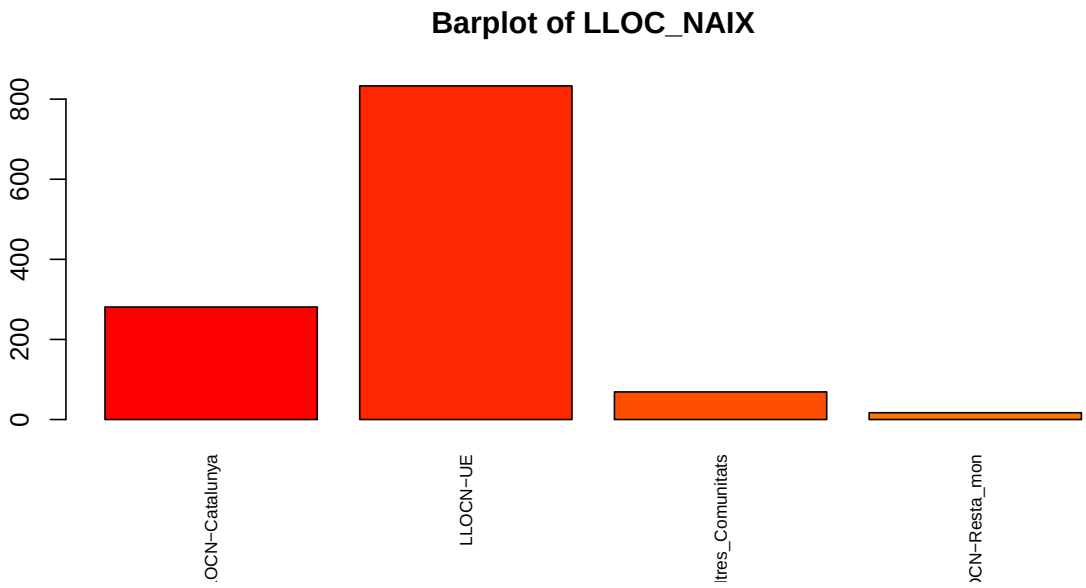


Table 9: Information about modalities of the Variable

Modalities	Frequency	Proportion
LLOCN-Catalunya	281	0.2341667
LLOCN-UE	833	0.6941667
LLOCN-Altres_Comunitats	69	0.0575000
LLOCN-Resta_mon	17	0.0141667

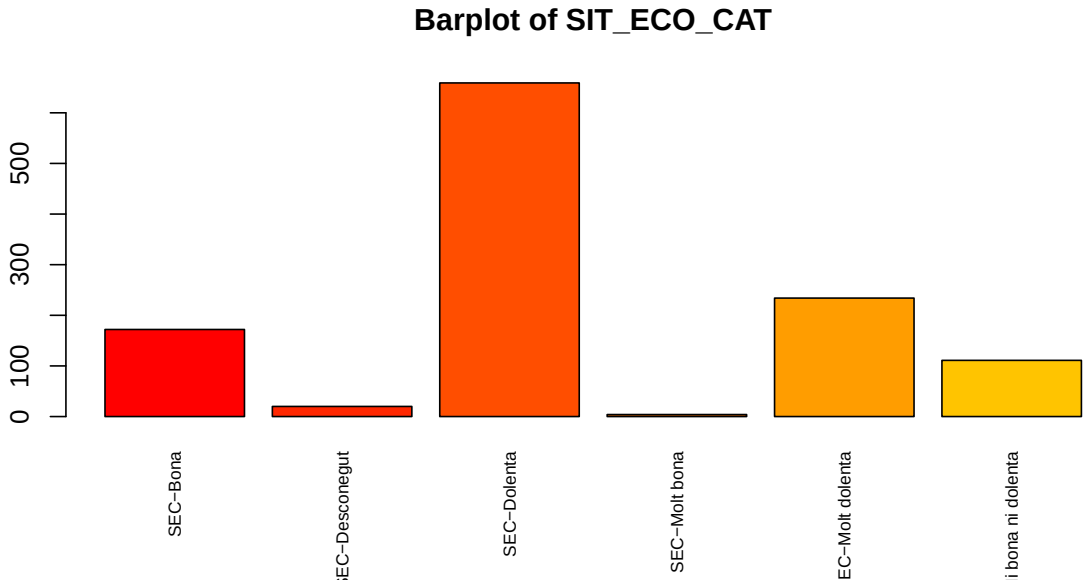


Table 10: Information about modalities of the Variable

Modalities	Frequency	Proportion
SEC-Bona	172	0.1433333
SEC-Desconegut	20	0.0166667
SEC-Dolenta	659	0.5491667
SEC-Molt bona	4	0.0033333
SEC-Molt dolenta	234	0.1950000
SEC-Ni bona ni dolenta	111	0.0925000

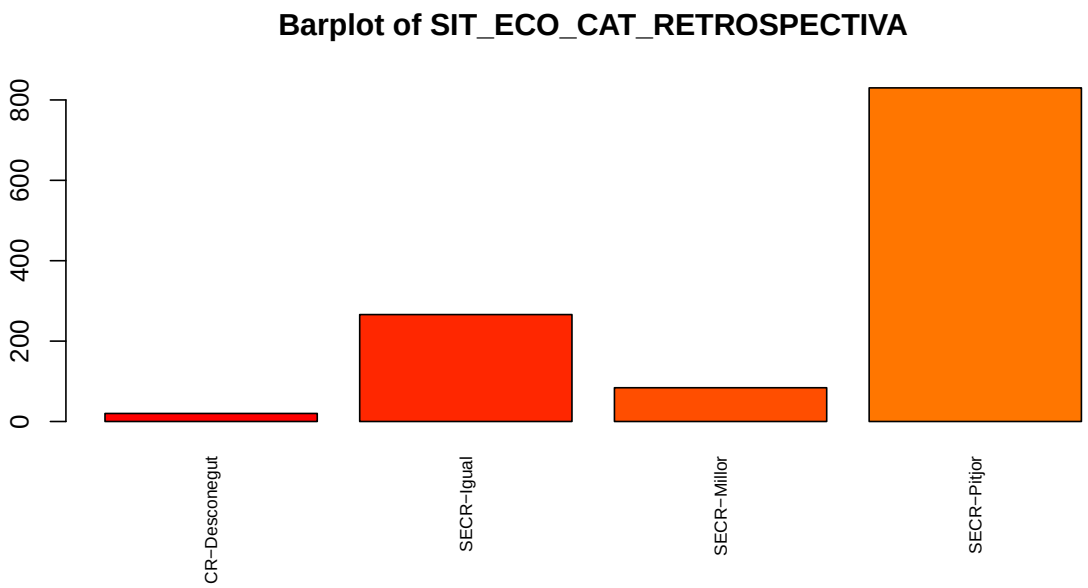


Table 11: Information about modalities of the Variable

Modalities	Frequency	Proportion
SECR-Desconegut	20	0.0166667
SECR-Igual	266	0.2216667
SECR-Millor	84	0.0700000
SECR-Pitjor	830	0.6916667

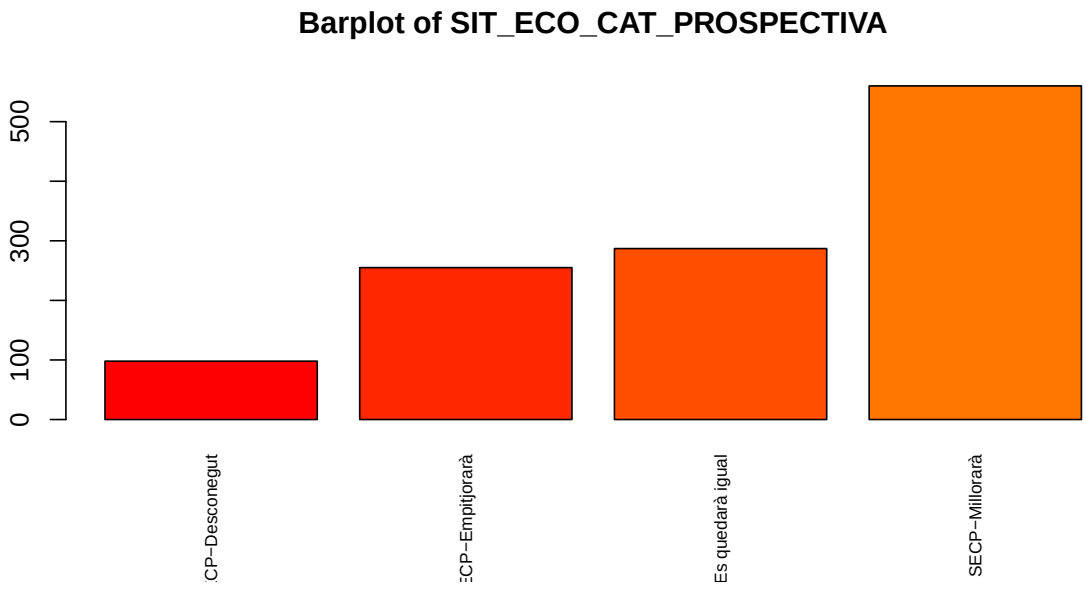


Table 12: Information about modalities of the Variable

Modalities	Frequency	Proportion
SECP-Desconegut	98	0.0816667
SECP-Empitjorarà	255	0.2125000
SECP-Es quedarà igual	287	0.2391667
SECP-Millorarà	560	0.4666667

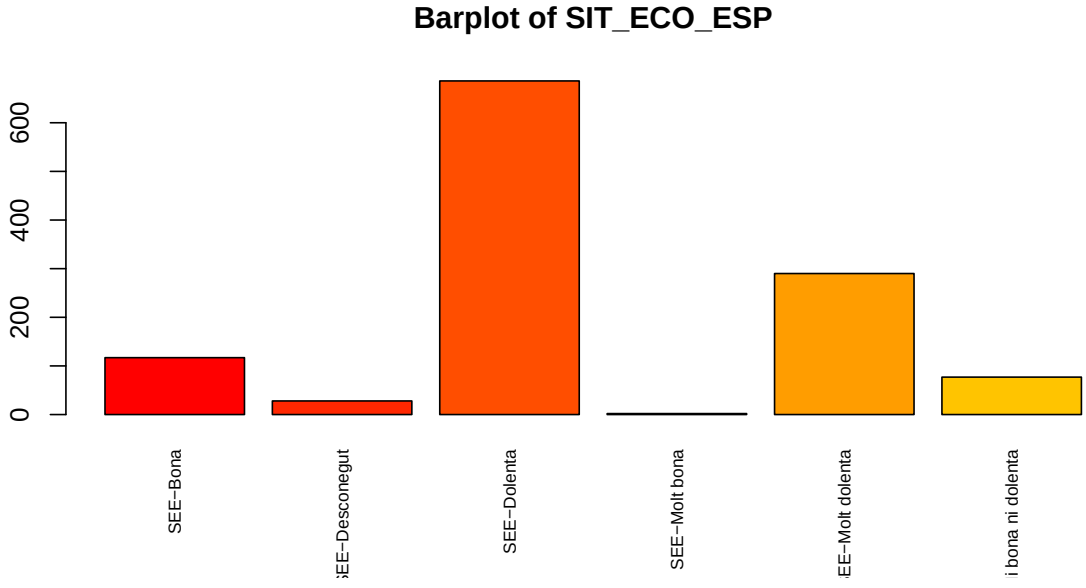


Table 13: Information about modalities of the Variable

Modalities	Frequency	Proportion
SEE-Bona	117	0.0975000
SEE-Desconegut	28	0.0233333
SEE-Dolenta	686	0.5716667
SEE-Molt bona	2	0.0016667
SEE-Molt dolenta	290	0.2416667
SEE-Ni bona ni dolenta	77	0.0641667

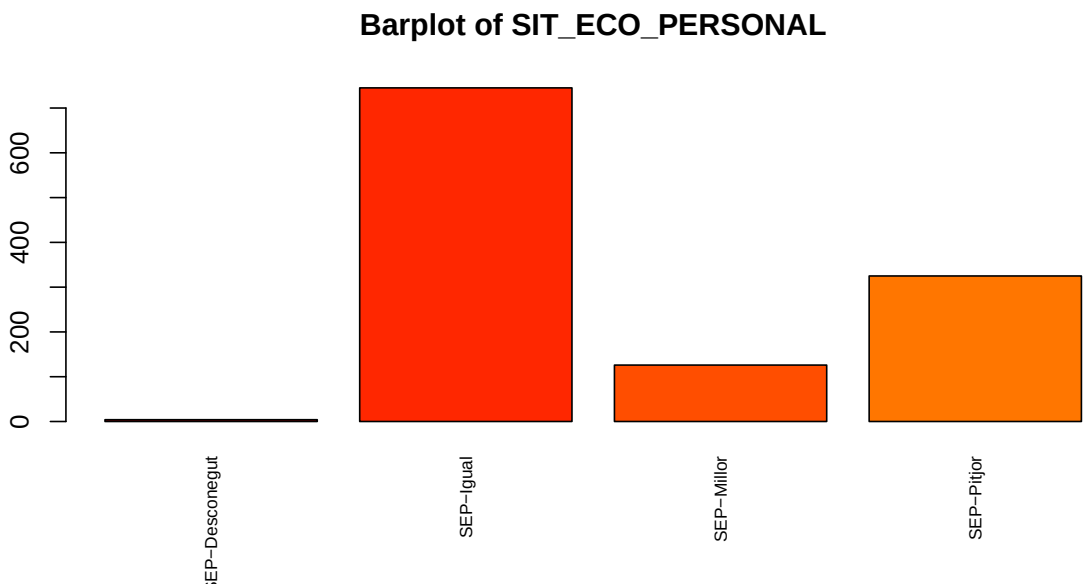


Table 14: Information about modalities of the Variable

Modalities	Frequency	Proportion
SEP-Desconegut	4	0.0033333
SEP-Igual	745	0.6208333
SEP-Millor	126	0.1050000
SEP-Pitjor	325	0.2708333

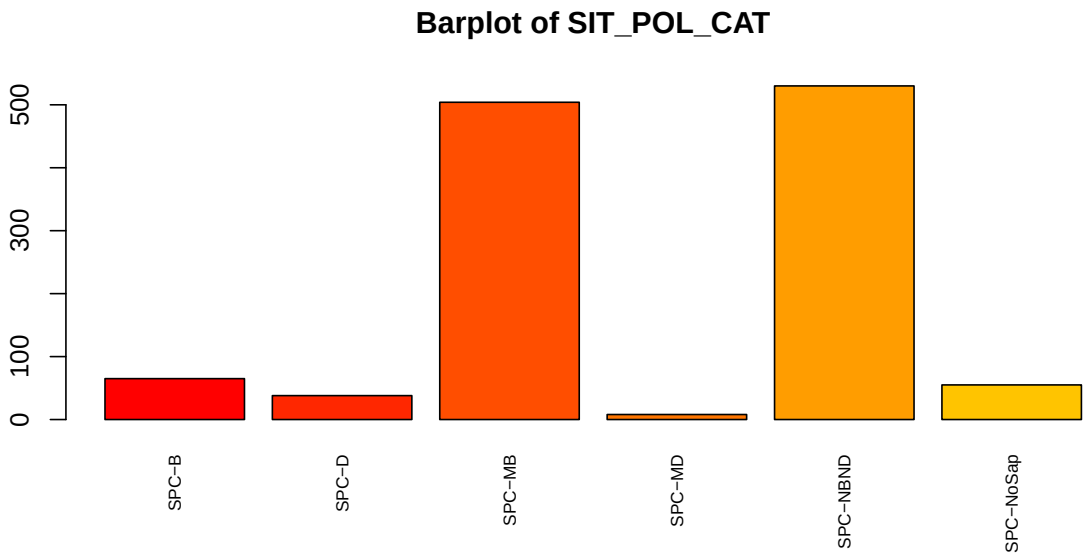


Table 15: Information about modalities of the Variable

Modalities	Frequency	Proportion
SPC-B	65	0.0541667
SPC-D	38	0.0316667
SPC-MB	504	0.4200000
SPC-MD	8	0.0066667
SPC-NBND	530	0.4416667
SPC-NoSap	55	0.0458333

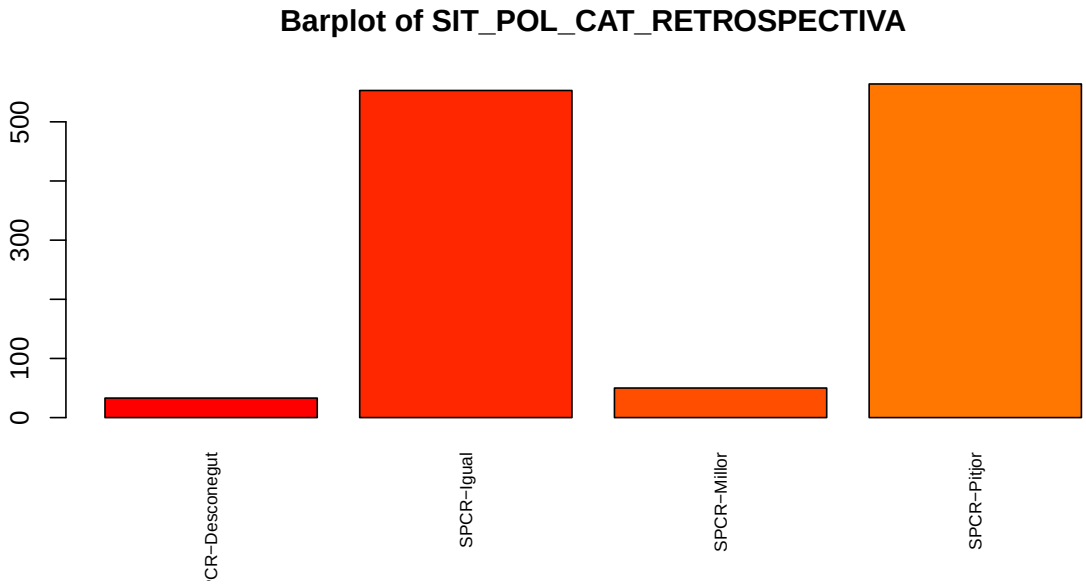


Table 16: Information about modalities of the Variable

Modalities	Frequency	Proportion
SPCR-Desconegut	33	0.0275000
SPCR-Igual	553	0.4608333
SPCR-Millor	50	0.0416667
SPCR-Pitjor	564	0.4700000

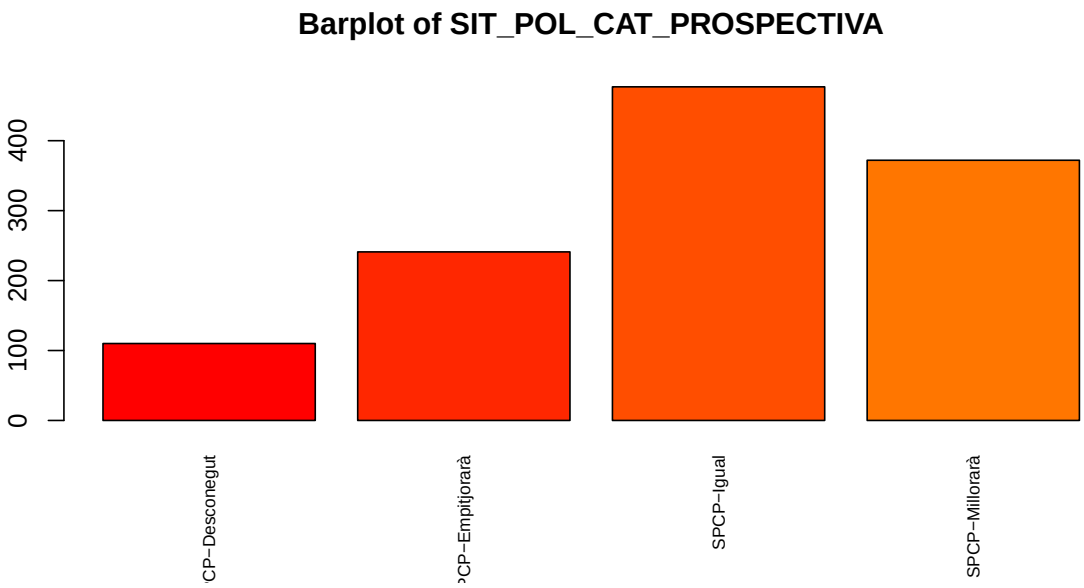


Table 17: Information about modalities of the Variable

Modalities	Frequency	Proportion
SPCP-Desconegut	110	0.0916667
SPCP-Empitjorarà	241	0.2008333
SPCP-Igual	477	0.3975000
SPCP-Millorarà	372	0.3100000

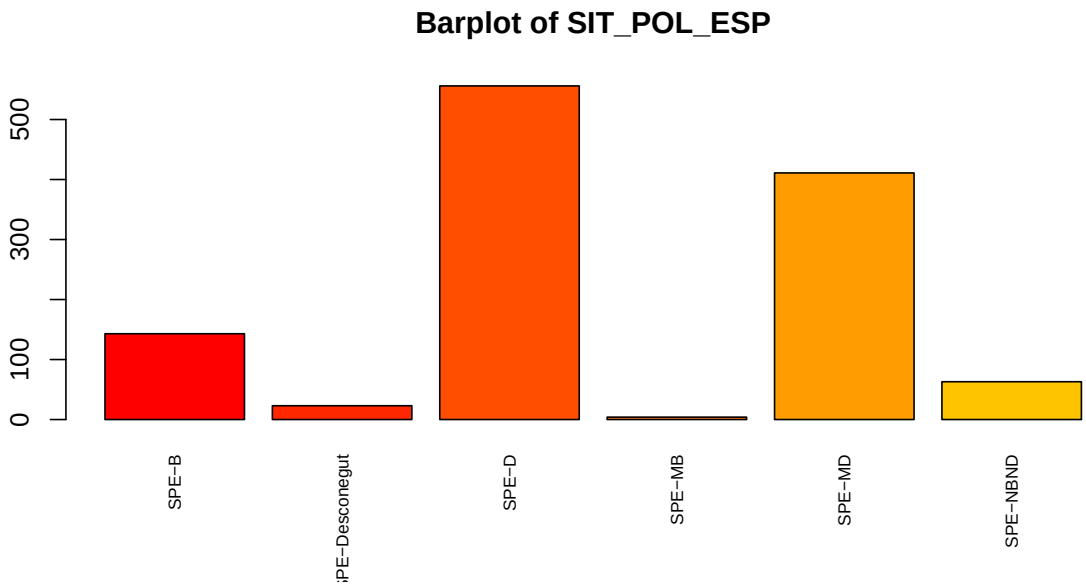


Table 18: Information about modalities of the Variable

Modalities	Frequency	Proportion
SPE-B	143	0.1191667
SPE-Desconegut	23	0.0191667
SPE-D	556	0.4633333
SPE-MB	4	0.0033333
SPE-MD	411	0.3425000
SPE-NBND	63	0.0525000

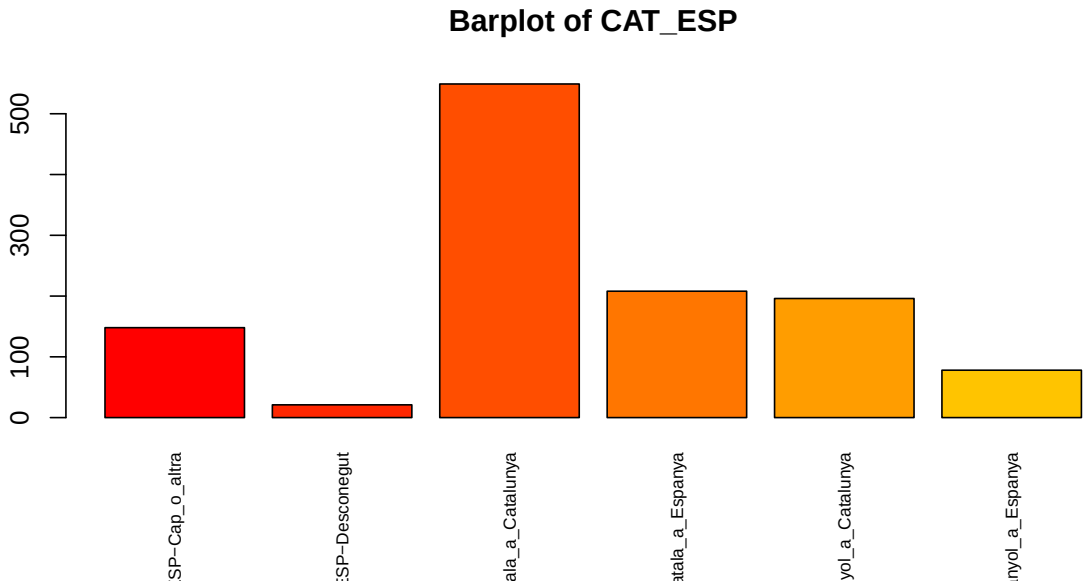


Table 19: Information about modalities of the Variable

Modalities	Frequency	Proportion
CATESP-Cap_o_altra	148	0.1233333
CATESP-Desconegut	21	0.0175000
CATESP-Catala_a_Catalunya	549	0.4575000
CATESP-Catala_a_Espanya	208	0.1733333
CATESP-Espanyol_a_Catalunya	196	0.1633333
CATESP-Espanyol_a_Espanya	78	0.0650000

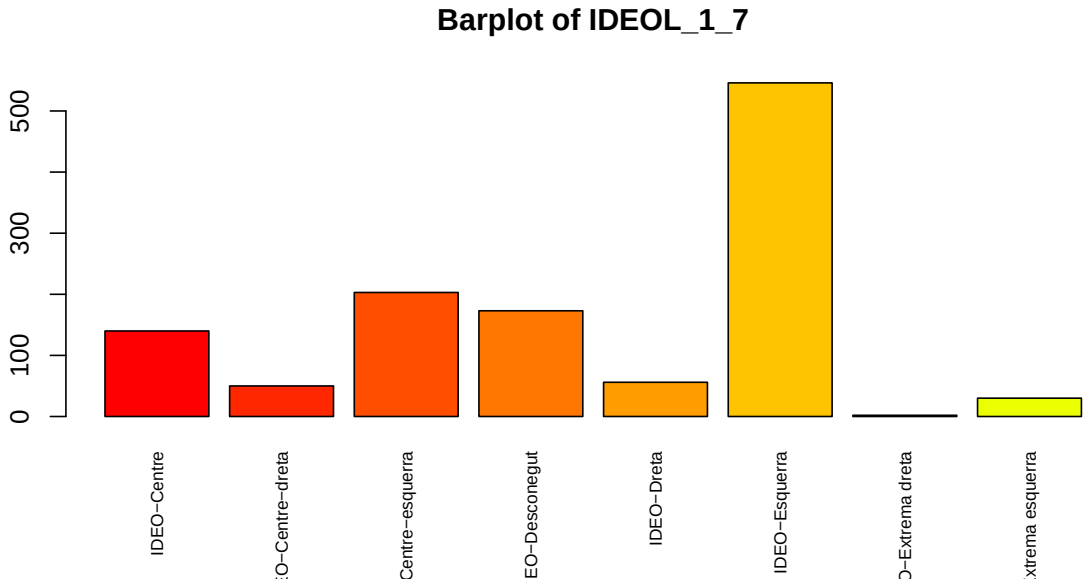


Table 20: Information about modalities of the Variable

Modalities	Frequency	Proportion
IDEO-Centre	140	0.1166667
IDEO-Centre-dreta	50	0.0416667
IDEO-Centre-esquerra	203	0.1691667
IDEO-Desconegut	173	0.1441667
IDEO-Dreta	56	0.0466667
IDEO-Esquerra	546	0.4550000
IDEO-Extrema dreta	2	0.0016667
IDEO-Extrema esquerra	30	0.0250000

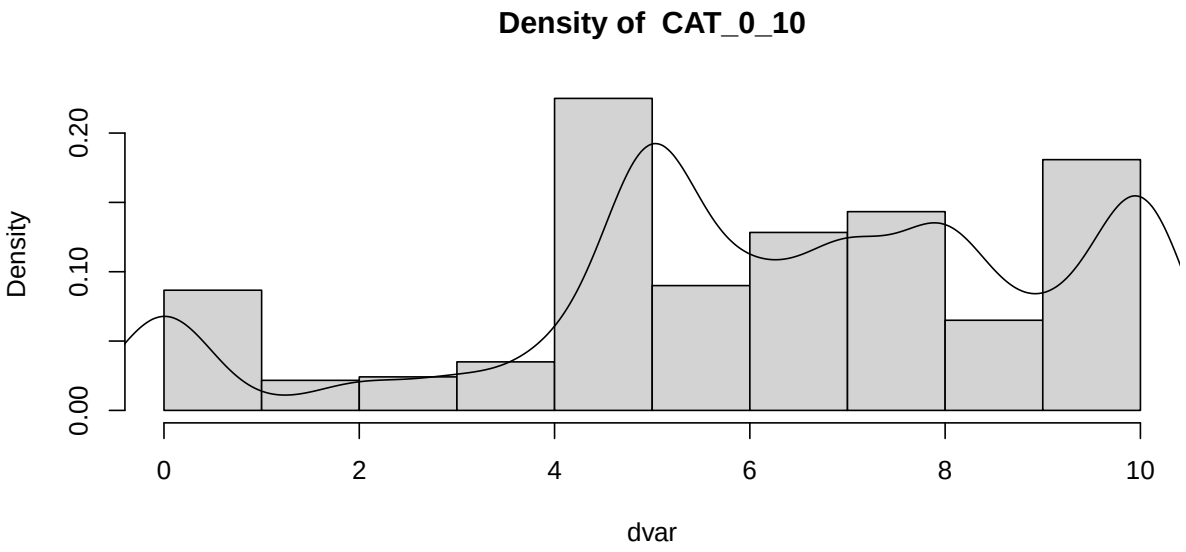


Table 21: Summary of the Variable

	variable	Min.	X1st.Qu.	Median	Mean	X3rd.Qu.	Max.
Min.	CAT_0_10	0	5	7	6.338329	8	10

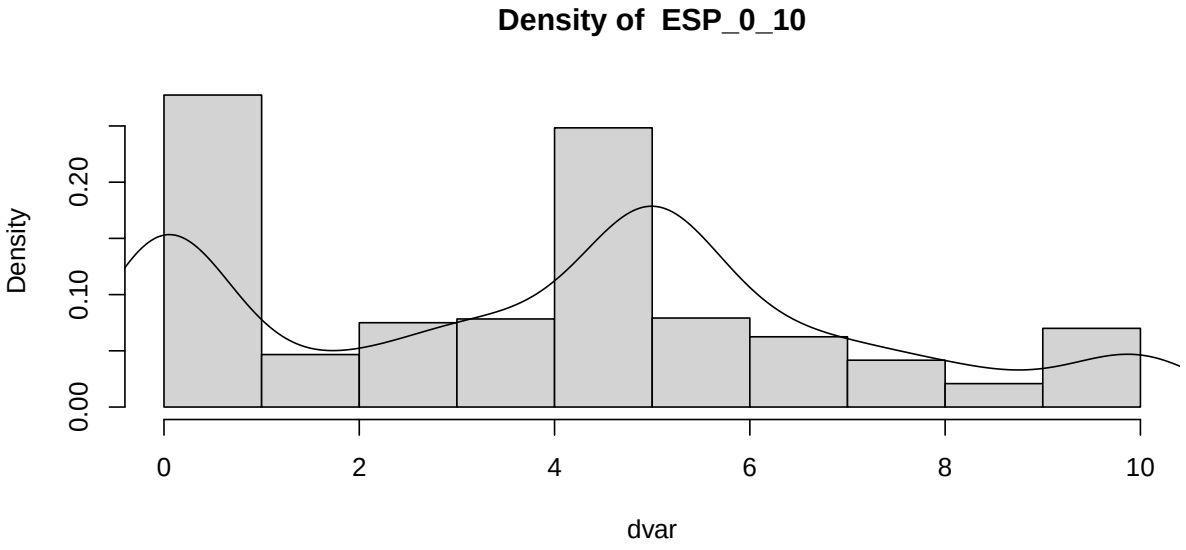


Table 22: Summary of the Variable

	variable	Min.	X1st.Qu.	Median	Mean	X3rd.Qu.	Max.
Min.	ESP_0_10	0	1	5	4.018124	6	10

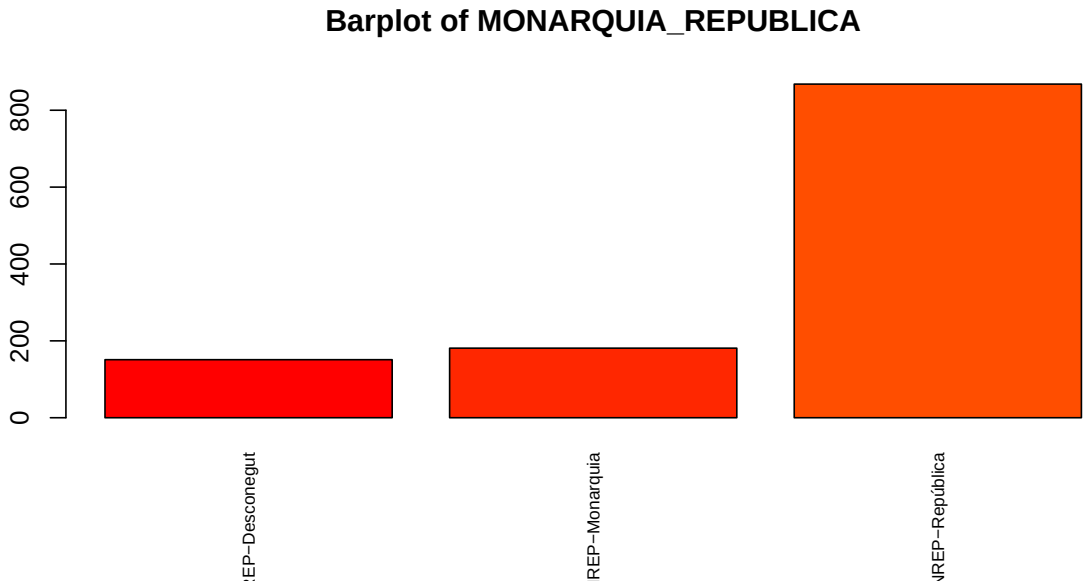


Table 23: Information about modalities of the Variable

Modalities	Frequency	Proportion
MONREP-Desconegut	151	0.1258333
MONREP-Monarquia	181	0.1508333
MONREP-República	868	0.7233333

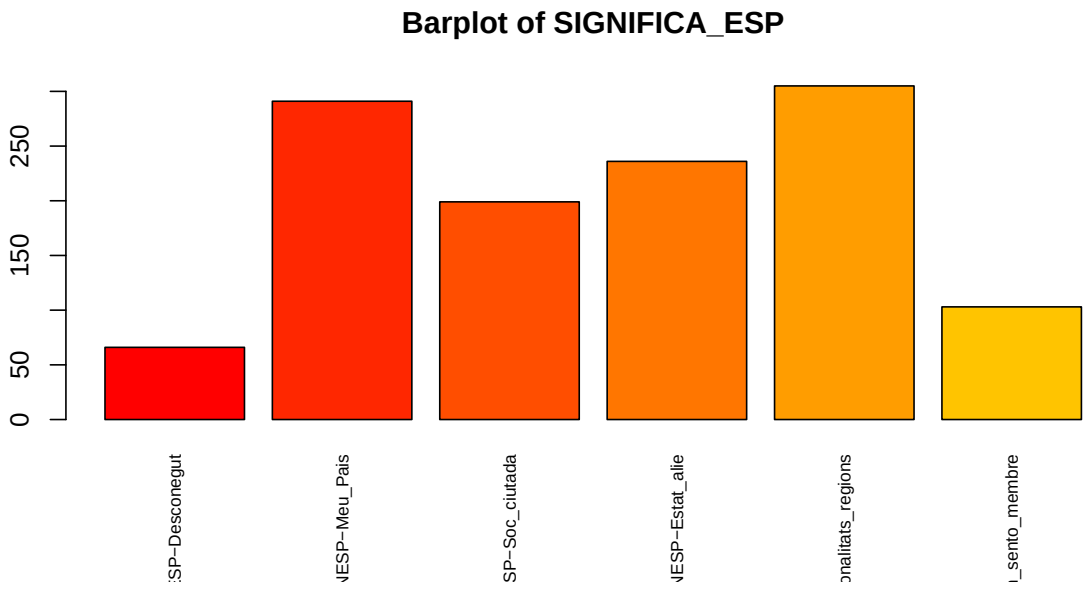


Table 24: Information about modalities of the Variable

Modalities	Frequency	Proportion
SIGNESP-Desconegut	66	0.0550000
SIGNESP-Meu_Pais	291	0.2425000
SIGNESP-Soc_ciutada	199	0.1658333
SIGNESP-Estat_alie	236	0.1966667
SIGNESP-Diverses_nacionalitats_regions	305	0.2541667
SIGNESP-Em_sento_membre	103	0.0858333

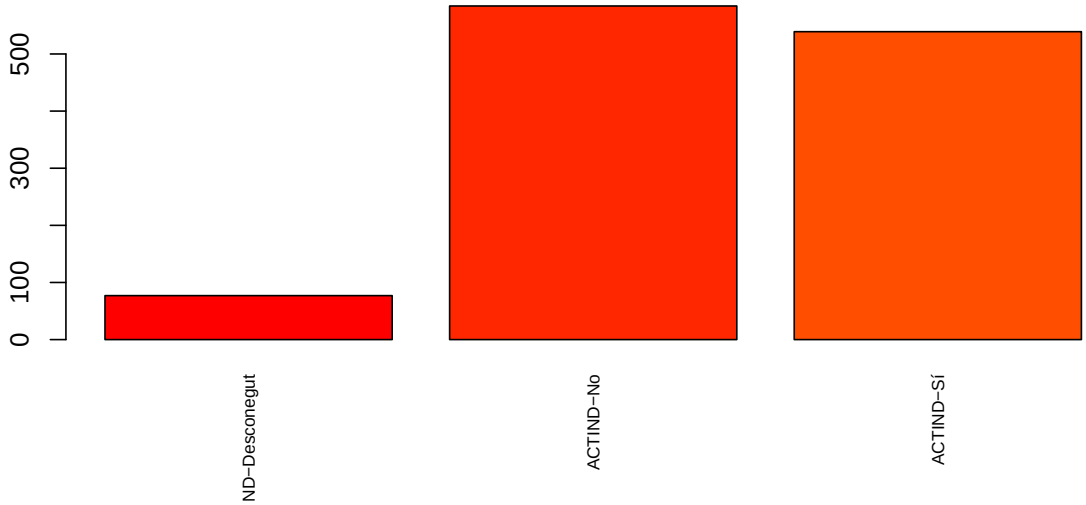


Table 25: Information about modalities of the Variable

Modalities	Frequency	Proportion
ACTIND-Desconegut	77	0.0641667
ACTIND-No	584	0.4866667
ACTIND-Sí	539	0.4491667

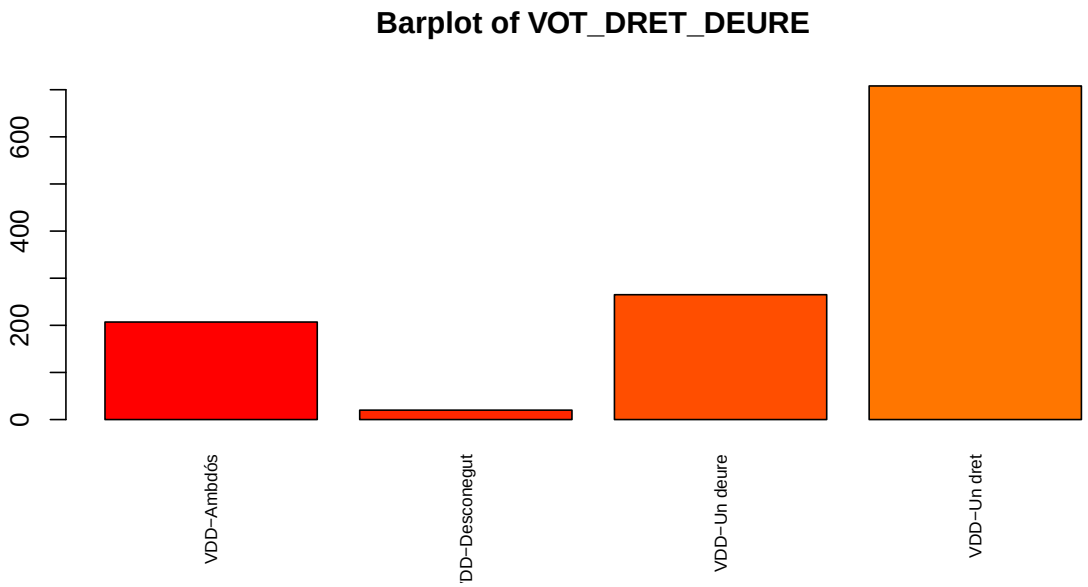


Table 26: Information about modalities of the Variable

Modalities	Frequency	Proportion
VDD-Ambdós	207	0.1725000
VDD-Desconegut	20	0.0166667
VDD-Un deure	265	0.2208333
VDD-Un dret	708	0.5900000

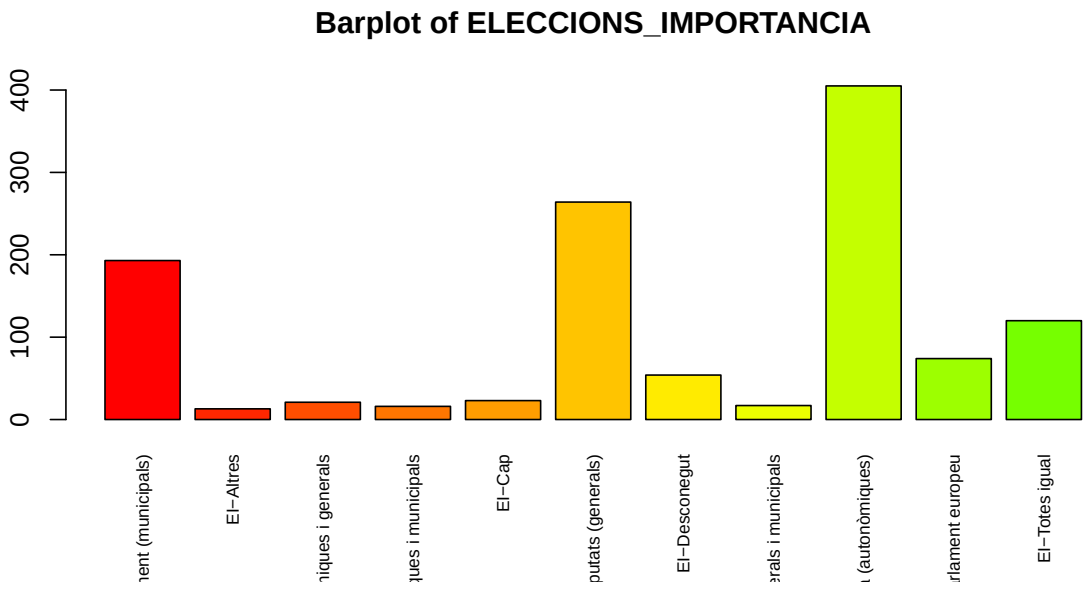


Table 27: Information about modalities of the Variable

Modalities	Frequency	Proportion
EI-Ajuntament (municipals)	193	0.1608333
EI-Altres	13	0.0108333
EI-Autonòmiques i generals	21	0.0175000
EI-Autonòmiques i municipals	16	0.0133333
EI-Cap	23	0.0191667
EI-Congrés dels diputats (generals)	264	0.2200000
EI-Desconegut	54	0.0450000
EI-Generals i municipals	17	0.0141667
EI-Parlament de Catalunya (autonòmiques)	405	0.3375000
EI-Parlament europeu	74	0.0616667
EI-Totes igual	120	0.1000000

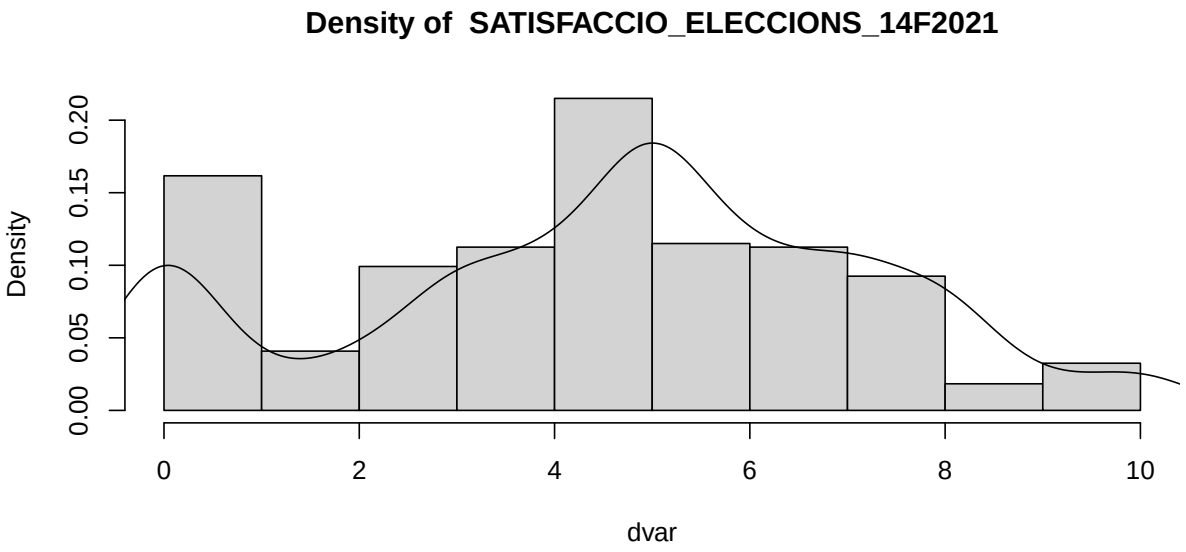


Table 28: Summary of the Variable

	variable	Min.	X1st.Qu.	Median	Mean	X3rd.Qu.	Max.
Min.	SATISFACCIO_ELECCIONS_14F2021	0	3	5	4.609912	7	10

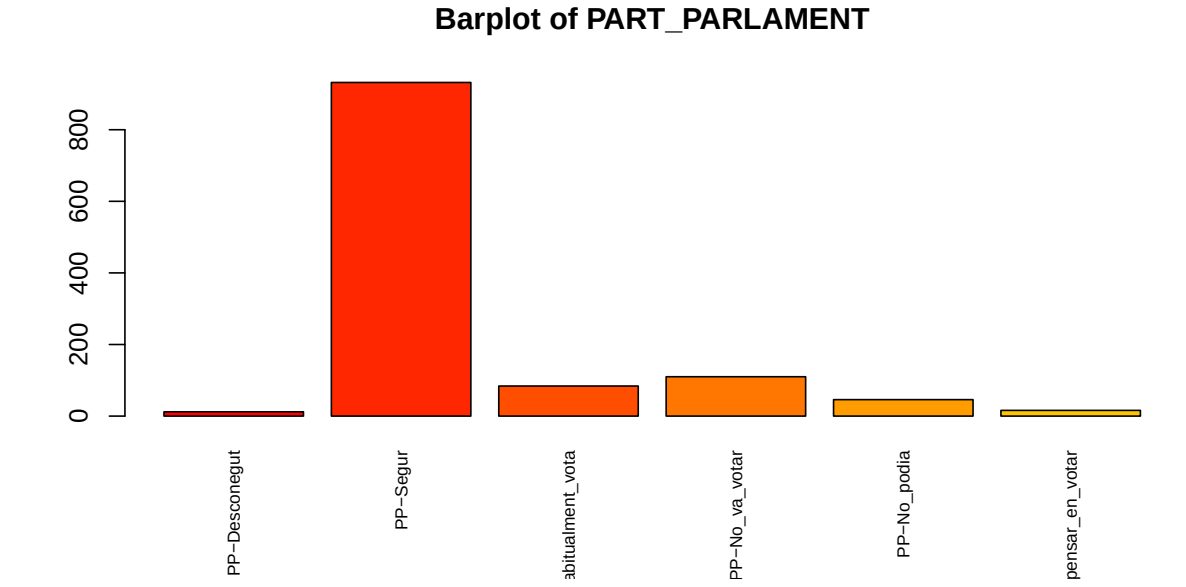


Table 29: Information about modalities of the Variable

Modalities	Frequency	Proportion
PP-Desconegut	12	0.0100000
PP-Segur	932	0.7766667
PP-No_pero_habitualment_vota	84	0.0700000
PP-No_va_votar	110	0.0916667
PP-No_podia	46	0.0383333
PP-No_pero_va_pensar_en_votar	16	0.0133333

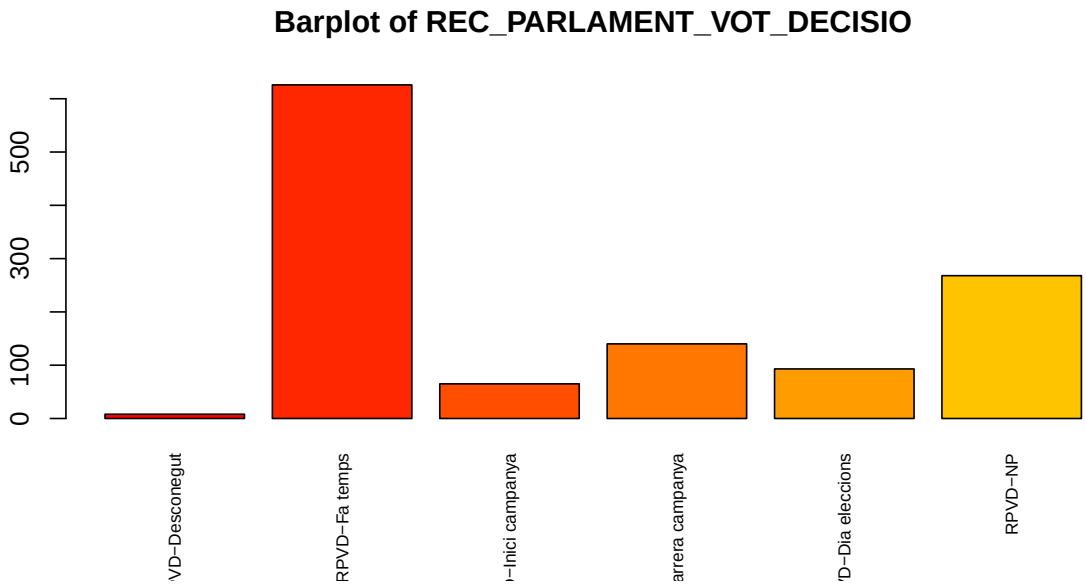


Table 30: Information about modalities of the Variable

Modalities	Frequency	Proportion
RPVD-Desconegut	8	0.0066667
RPVD-Fa temps	626	0.5216667
RPVD-Inici campanya	65	0.0541667
RPVD-Darrera campanya	140	0.1166667
RPVD-Dia eleccions	93	0.0775000
RPVD-NP	268	0.2233333

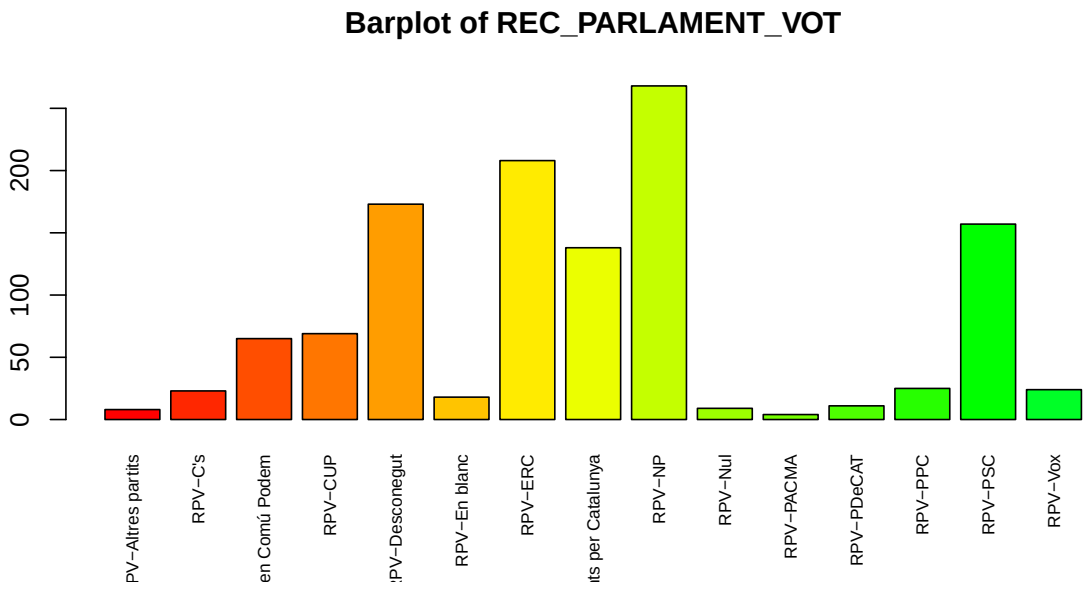


Table 31: Information about modalities of the Variable

Modalities	Frequency	Proportion
RPV-Altres partits	8	0.0066667
RPV-C's	23	0.0191667
RPV-Catalunya en Comú Podem	65	0.0541667
RPV-CUP	69	0.0575000
RPV-Desconegut	173	0.1441667
RPV-En blanc	18	0.0150000
RPV-ERC	208	0.1733333
RPV-Junts per Catalunya	138	0.1150000
RPV-NP	268	0.2233333
RPV-Nul	9	0.0075000
RPV-PACMA	4	0.0033333
RPV-PDeCAT	11	0.0091667
RPV-PPC	25	0.0208333
RPV-PSC	157	0.1308333
RPV-Vox	24	0.0200000

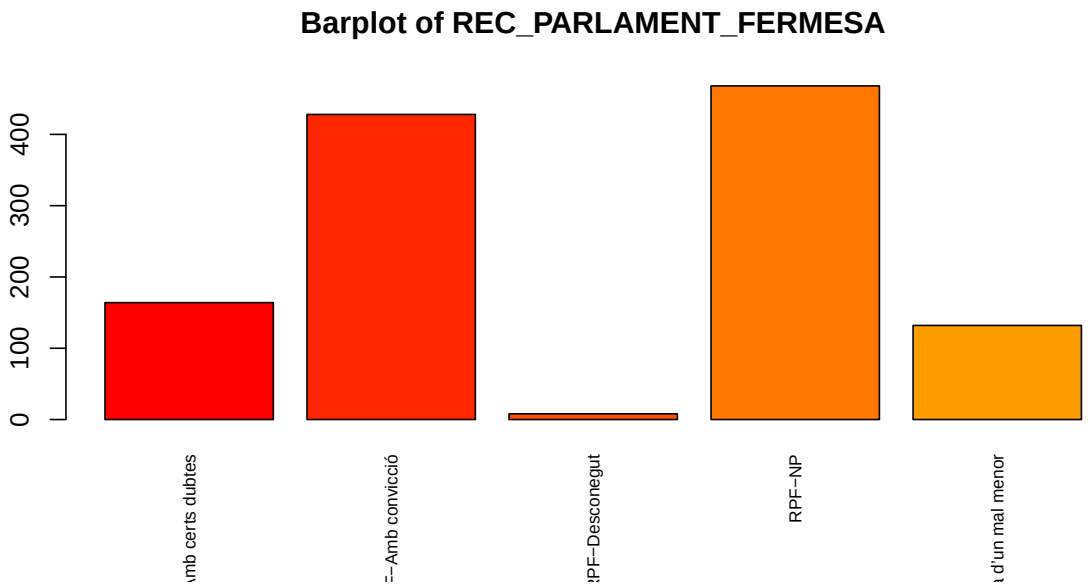


Table 32: Information about modalities of the Variable

Modalities	Frequency	Proportion
RPF-Amb certs dubtes	164	0.1366667
RPF-Amb convicció	428	0.3566667
RPF-Desconegut	8	0.0066667
RPF-NP	468	0.3900000
RPF-Perquè es tractava d'un mal menor	132	0.1100000

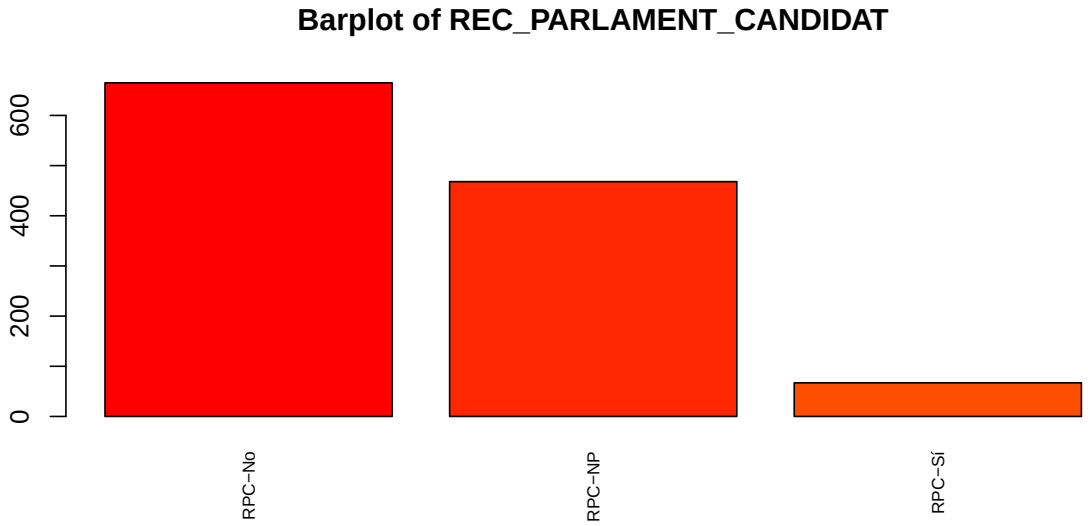


Table 33: Information about modalities of the Variable

Modalities	Frequency	Proportion
RPC-No	665	0.5541667
RPC-NP	468	0.3900000
RPC-Sí	67	0.0558333

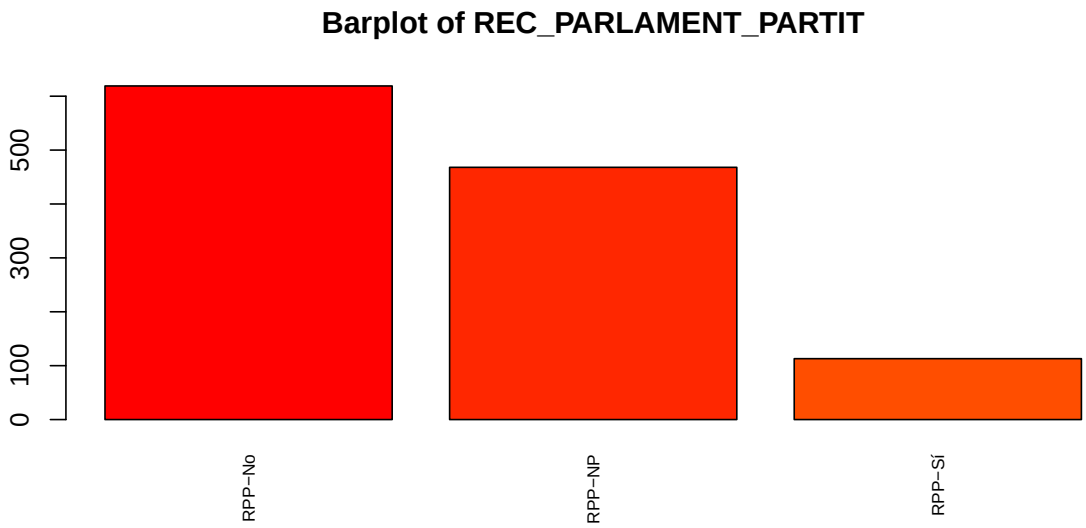


Table 34: Information about modalities of the Variable

Modalities	Frequency	Proportion
RPP-No	619	0.5158333
RPP-NP	468	0.3900000
RPP-Sí	113	0.0941667

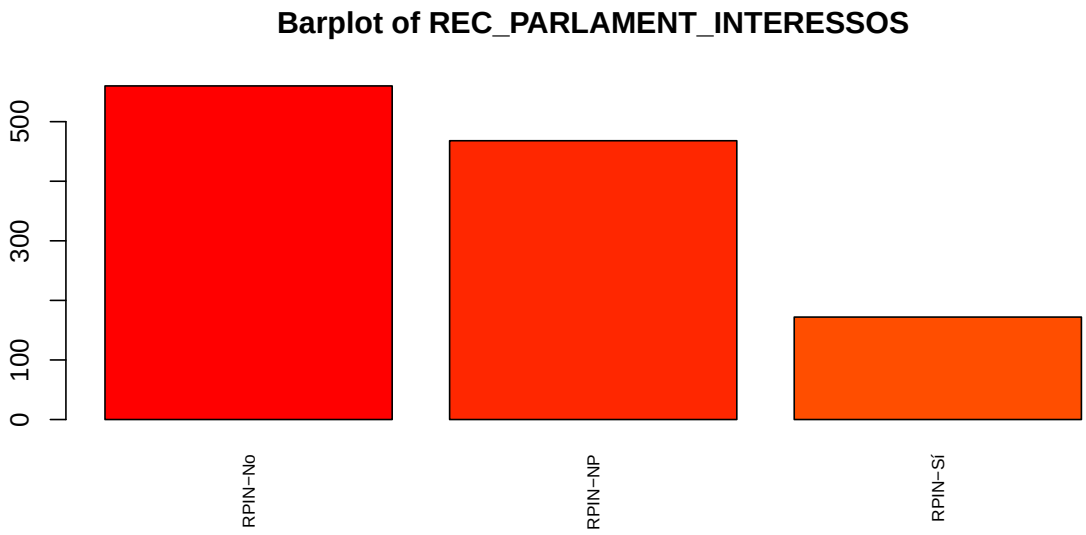


Table 35: Information about modalities of the Variable

Modalities	Frequency	Proportion
RPIN-No	560	0.4666667
RPIN-NP	468	0.3900000
RPIN-Sí	172	0.1433333

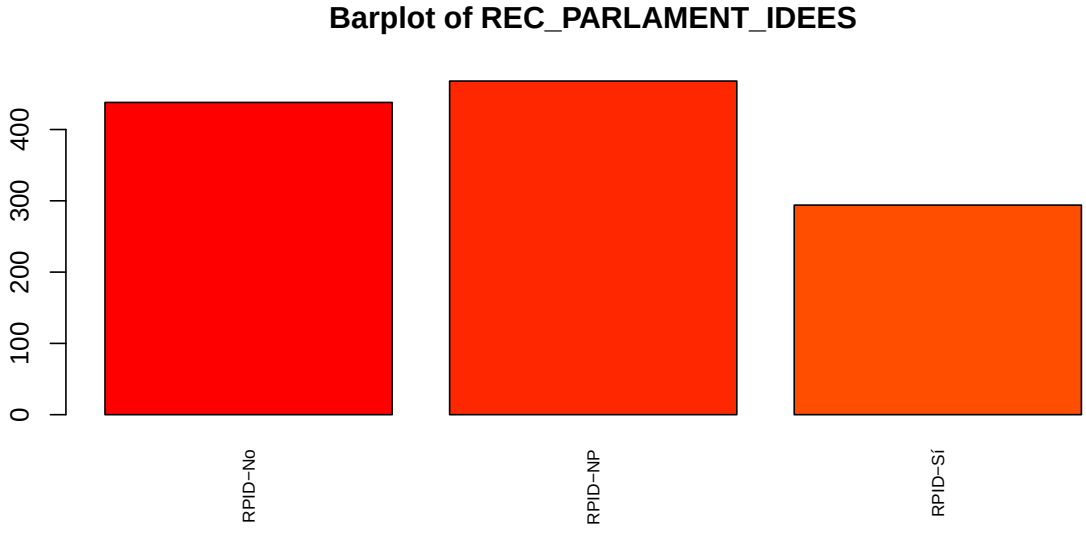


Table 36: Information about modalities of the Variable

Modalities	Frequency	Proportion
RPID-No	438	0.365
RPID-NP	468	0.390
RPID-Sí	294	0.245

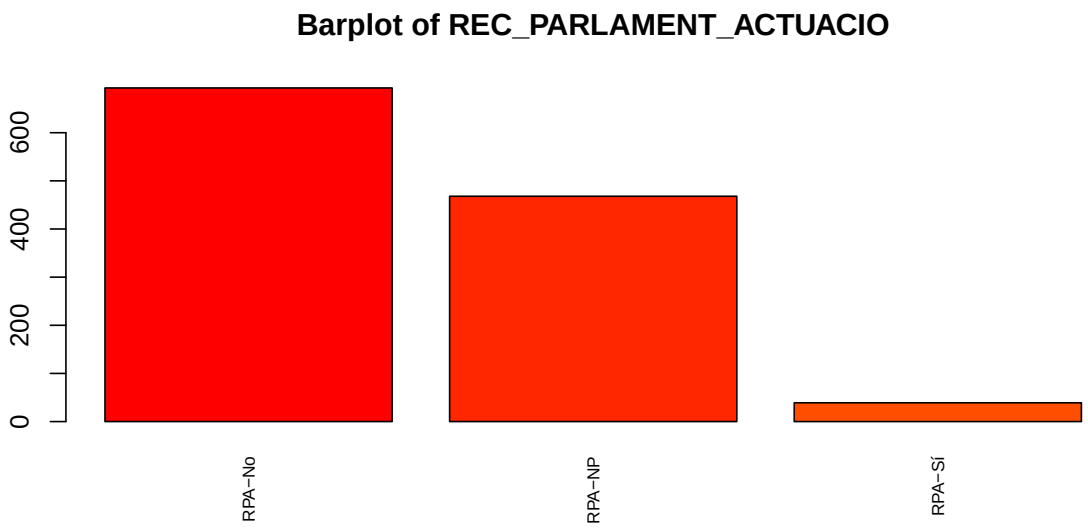


Table 37: Information about modalities of the Variable

Modalities	Frequency	Proportion
RPA-No	693	0.5775
RPA-NP	468	0.3900
RPA-Sí	39	0.0325

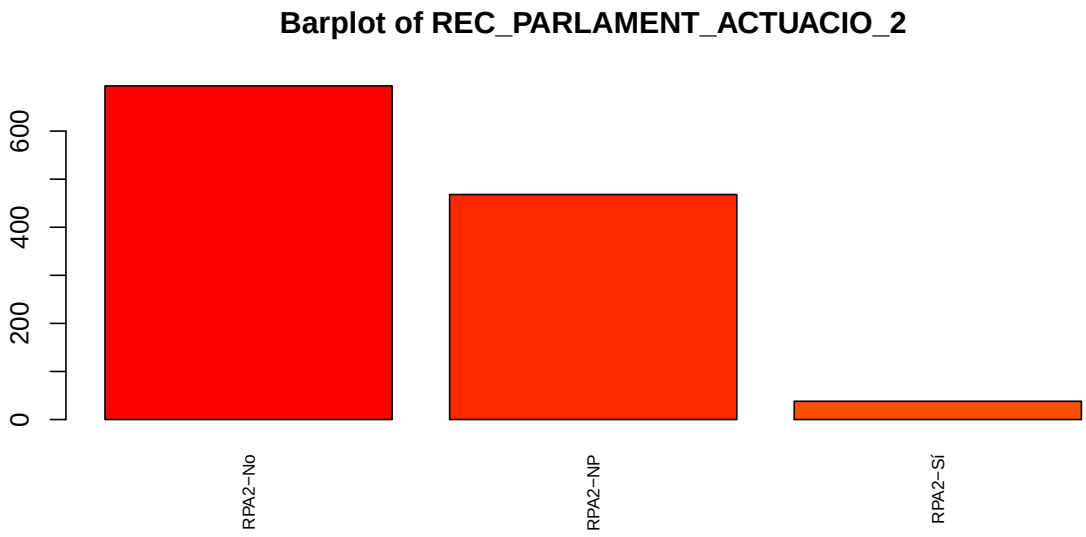


Table 38: Information about modalities of the Variable

Modalities	Frequency	Proportion
RPA2-No	694	0.5783333
RPA2-NP	468	0.3900000
RPA2-Sí	38	0.0316667

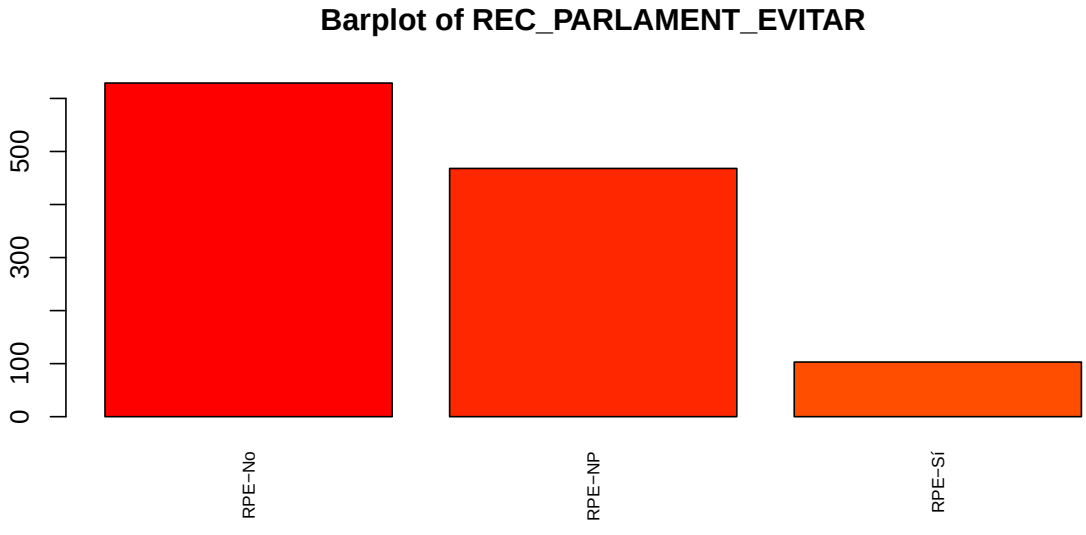


Table 39: Information about modalities of the Variable

Modalities	Frequency	Proportion
RPE-No	629	0.5241667
RPE-NP	468	0.3900000
RPE-Sí	103	0.0858333

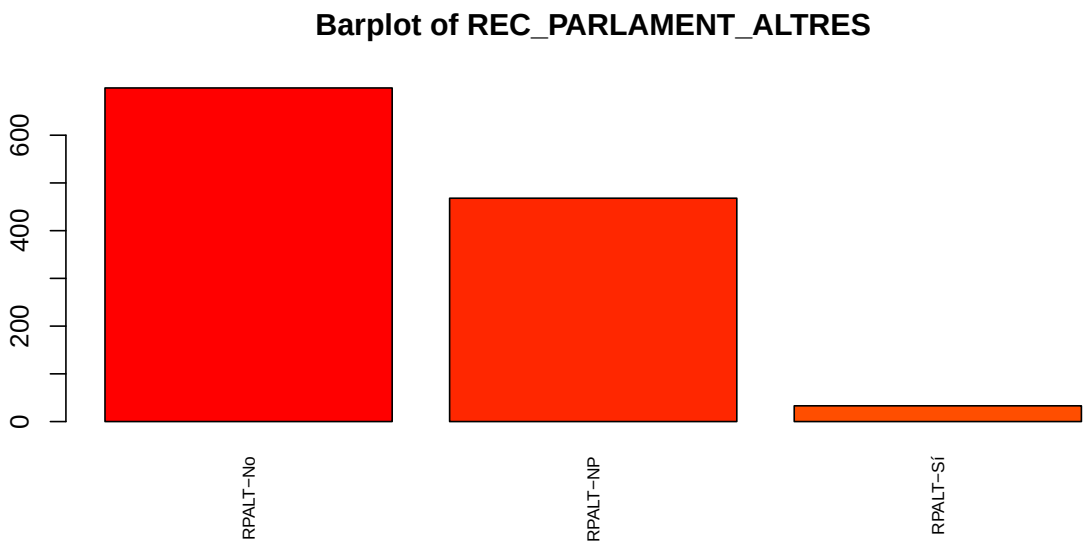


Table 40: Information about modalities of the Variable

Modalities	Frequency	Proportion
RPALT-No	699	0.5825
RPALT-NP	468	0.3900
RPALT-Sí	33	0.0275

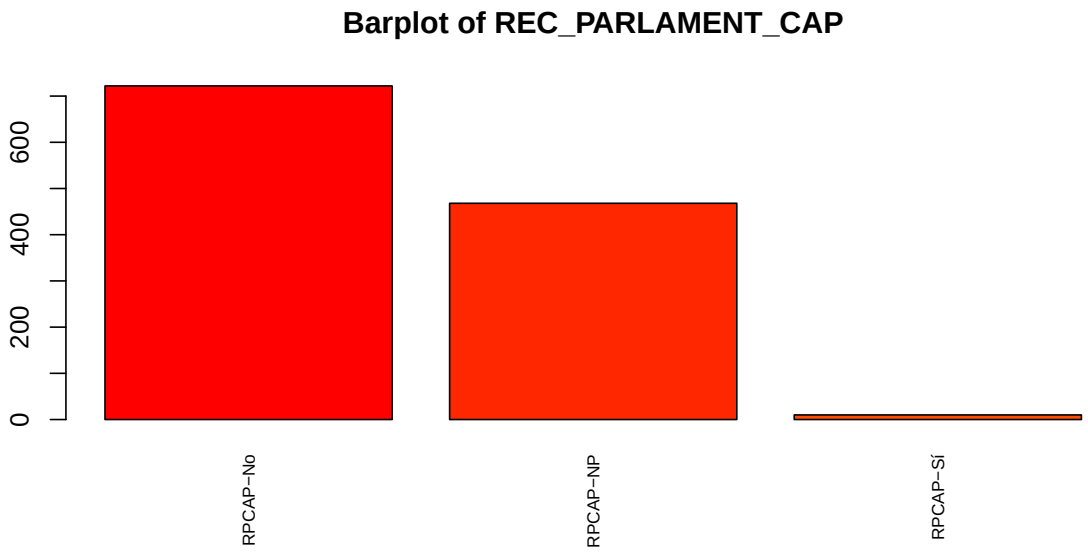


Table 41: Information about modalities of the Variable

Modalities	Frequency	Proportion
RPCAP-No	722	0.6016667
RPCAP-NP	468	0.3900000
RPCAP-Sí	10	0.0083333

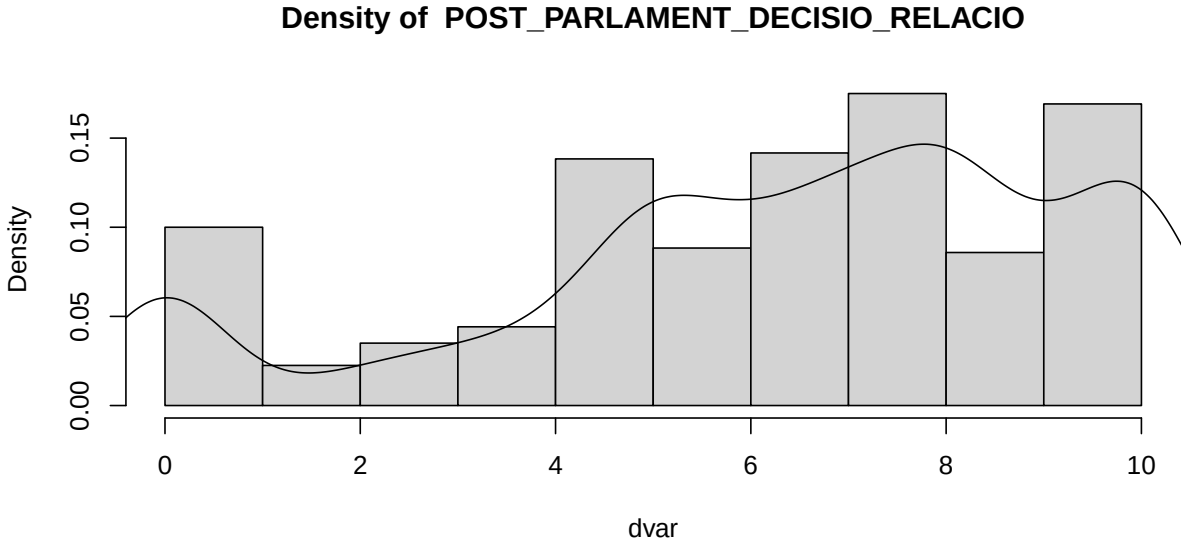


Table 42: Summary of the Variable

	variable	Min.	X1st.Qu.	Median	Mean	X3rd.Qu.	Max.
Min.	POST_PARLAMENT_DECISIO_RELACIO	0	5	7	6.360042	9	10

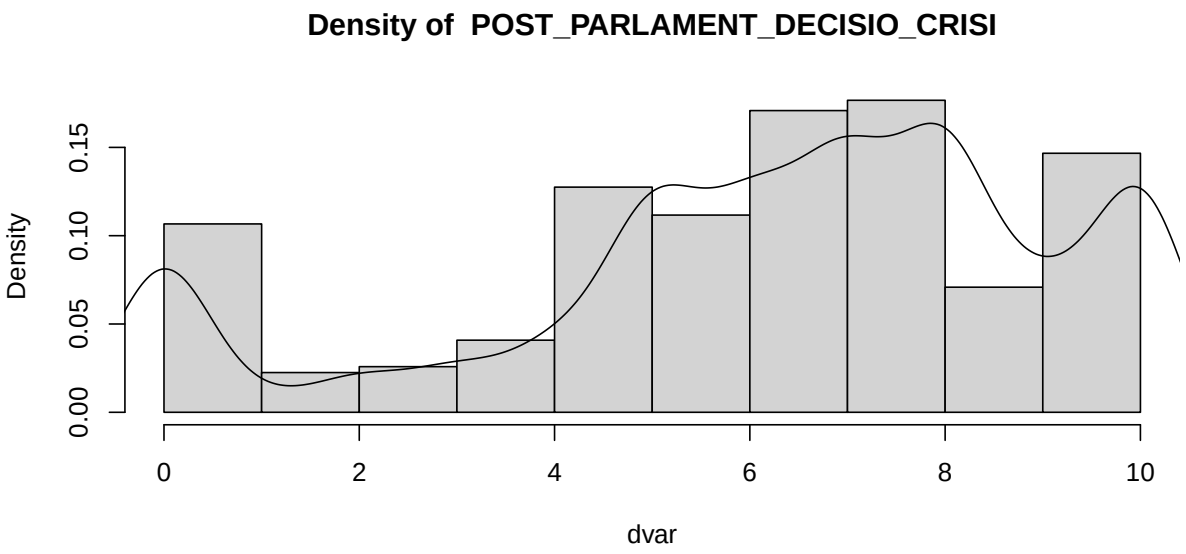


Table 43: Summary of the Variable

	variable	Min.	X1st.Qu.	Median	Mean	X3rd.Qu.	Max.
Min.	POST_PARLAMENT_DECISIO_CRISI	0	5	7	6.266579	8	10

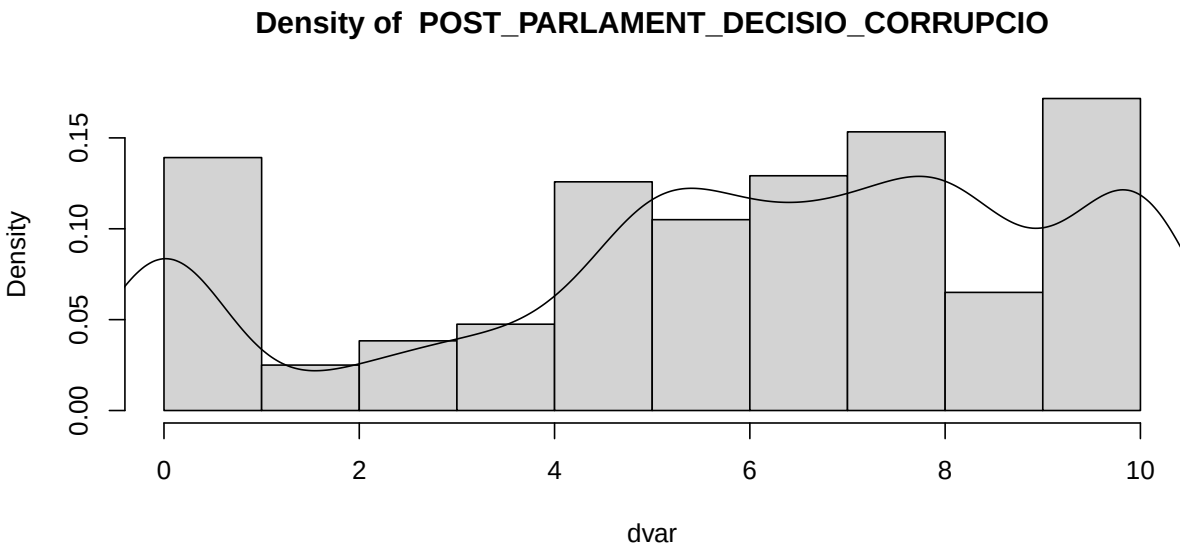


Table 44: Summary of the Variable

	variable	Min.	X1st.Qu.	Median	Mean	X3rd.Qu.	Max.
Min.	POST_PARLAMENT_DECISIO_CORRUPCIO	0	4.028619	6.456504	6.001658	8	10

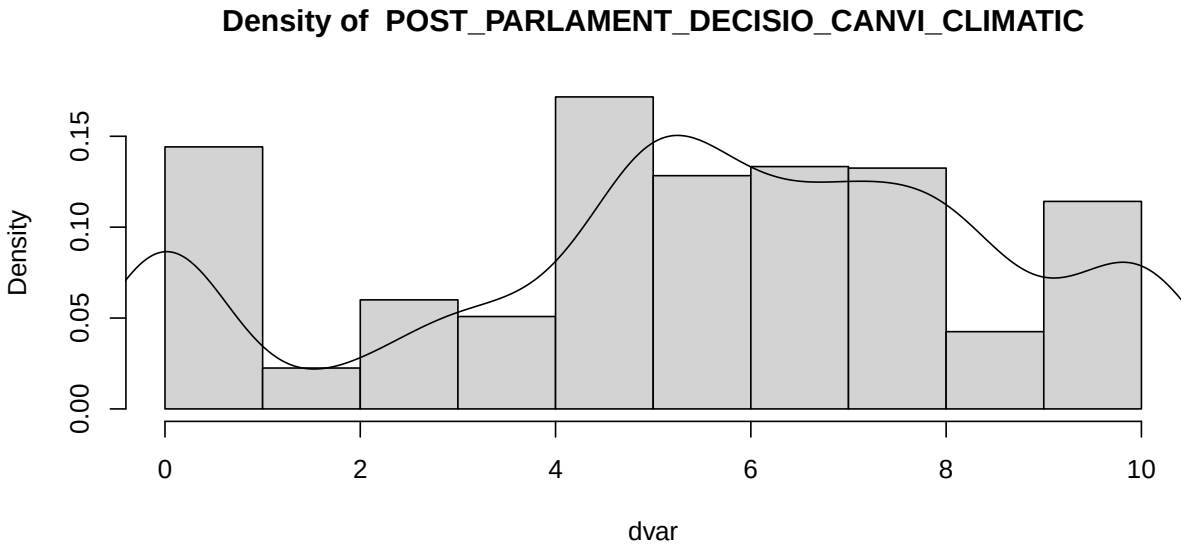


Table 45: Summary of the Variable

	variable	Min.	X1st.Qu.	Median	Mean	X3rd.Qu.	Max.
Min.	POST_PARLAMENT_DECISIO_CANVI_CLIMATIC	0	4	6	5.53274	8	10

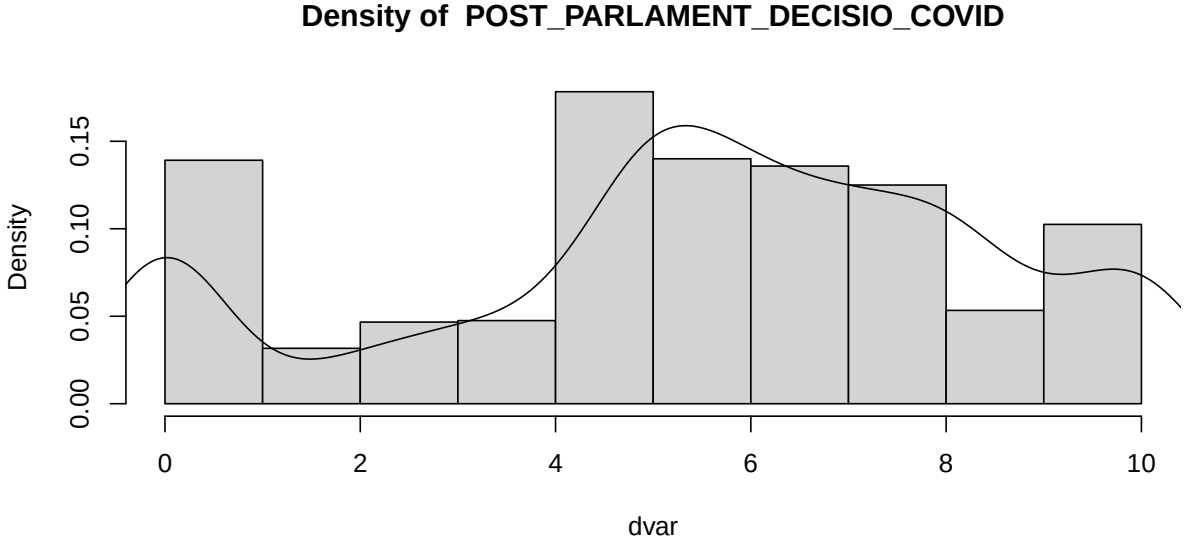


Table 46: Summary of the Variable

	variable	Min.	X1st.Qu.	Median	Mean	X3rd.Qu.	Max.
Min.	POST_PARLAMENT_DECISIO_COVID	0	4	6	5.536522	8	10

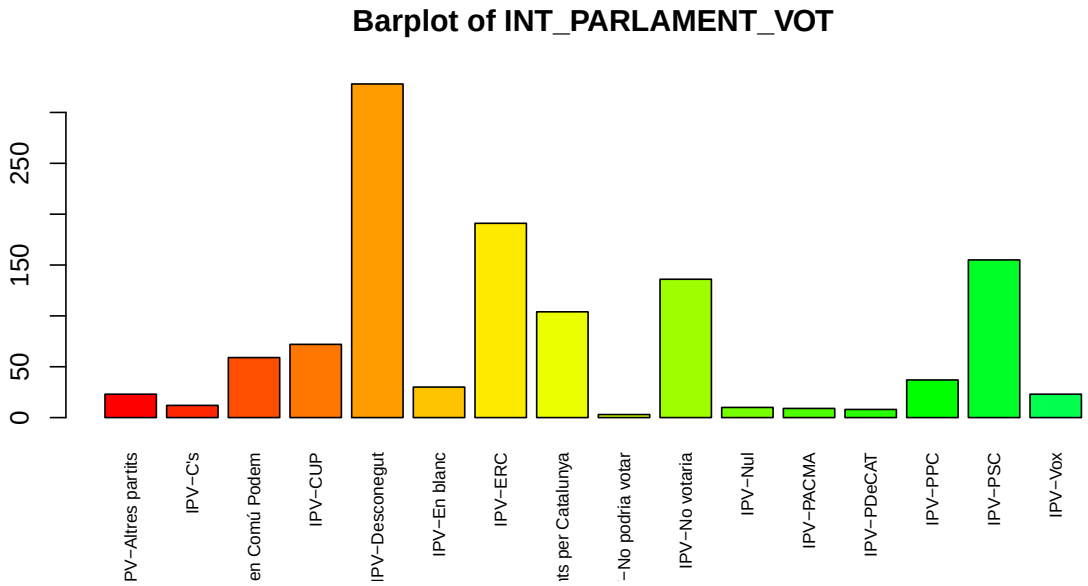


Table 47: Information about modalities of the Variable

Modalities	Frequency	Proportion
IPV-Altres partits	23	0.0191667
IPV-C's	12	0.0100000
IPV-Catalunya en Comú Podem	59	0.0491667
IPV-CUP	72	0.0600000
IPV-Desconegut	328	0.2733333
IPV-En blanc	30	0.0250000
IPV-ERC	191	0.1591667
IPV-Junts per Catalunya	104	0.0866667
IPV-No podria votar	3	0.0025000
IPV-No votaria	136	0.1133333
IPV-Nul	10	0.0083333
IPV-PACMA	9	0.0075000
IPV-PDeCAT	8	0.0066667
IPV-PPC	37	0.0308333
IPV-PSC	155	0.1291667
IPV-Vox	23	0.0191667

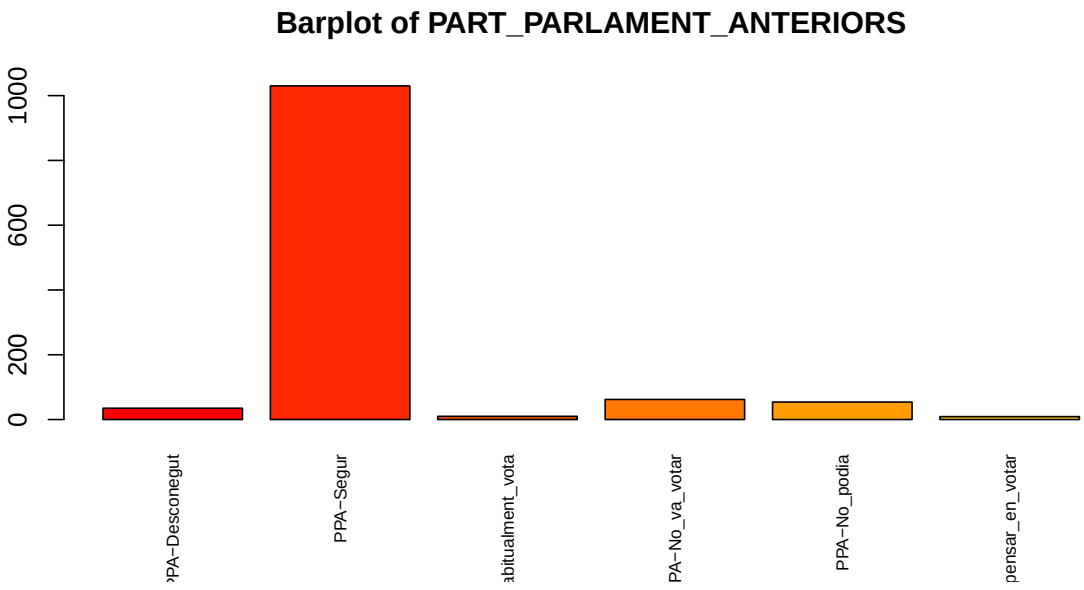


Table 48: Information about modalities of the Variable

Modalities	Frequency	Proportion
PPA-Desconegut	35	0.0291667
PPA-Segur	1030	0.8583333
PPA-No_pero_habitualment_vota	10	0.0083333
PPA-No_va_votar	62	0.0516667
PPA-No_podia	54	0.0450000
PPA-No_pero_va_pensar_en_votar	9	0.0075000

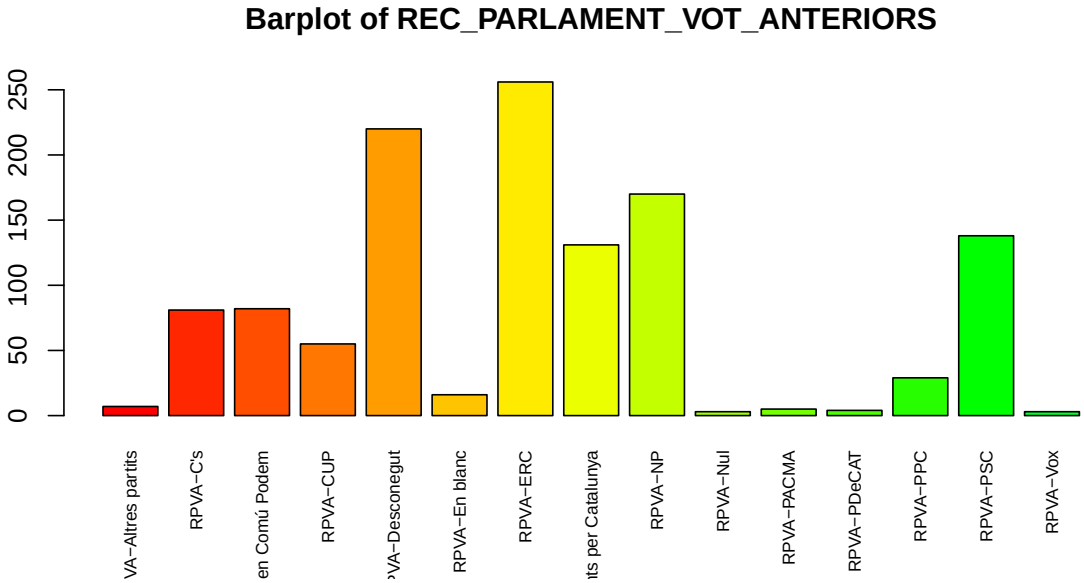


Table 49: Information about modalities of the Variable

Modalities	Frequency	Proportion
RPVA-Altres partits	7	0.0058333
RPVA-C's	81	0.0675000
RPVA-Catalunya en Comú Podem	82	0.0683333
RPVA-CUP	55	0.0458333
RPVA-Desconegut	220	0.1833333
RPVA-En blanc	16	0.0133333
RPVA-ERC	256	0.2133333
RPVA-Junts per Catalunya	131	0.1091667
RPVA-NP	170	0.1416667
RPVA-Nul	3	0.0025000
RPVA-PACMA	5	0.0041667
RPVA-PDeCAT	4	0.0033333
RPVA-PPC	29	0.0241667
RPVA-PSC	138	0.1150000
RPVA-Vox	3	0.0025000

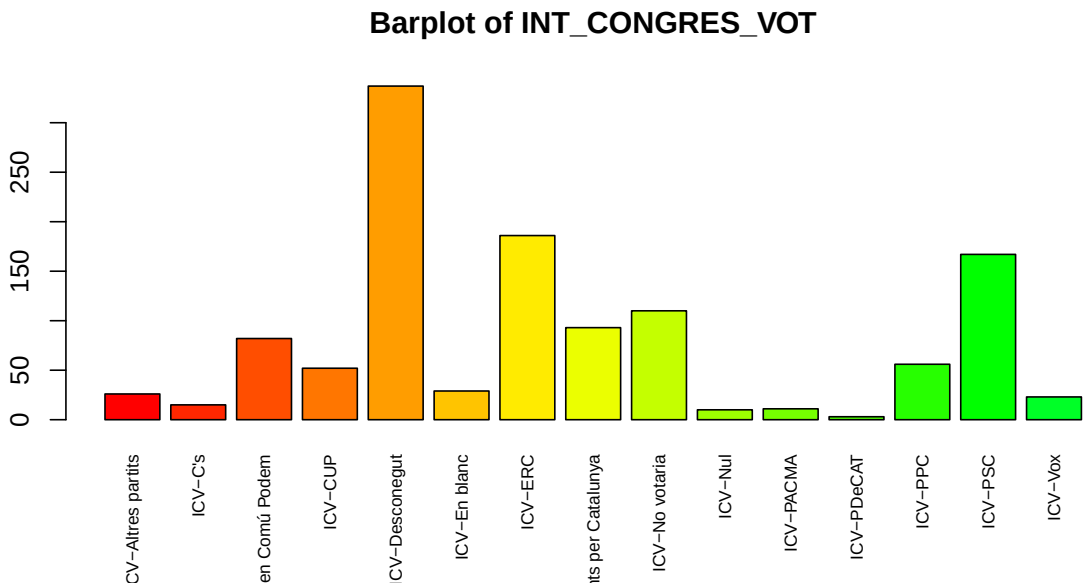


Table 50: Information about modalities of the Variable

Modalities	Frequency	Proportion
ICV-Altres partits	26	0.0216667
ICV-C's	15	0.0125000
ICV-Catalunya en Comú Podem	82	0.0683333
ICV-CUP	52	0.0433333
ICV-Desconegut	337	0.2808333
ICV-En blanc	29	0.0241667
ICV-ERC	186	0.1550000
ICV-Junts per Catalunya	93	0.0775000
ICV-No votaria	110	0.0916667
ICV-Nul	10	0.0083333
ICV-PACMA	11	0.0091667
ICV-PDeCAT	3	0.0025000
ICV-PPC	56	0.0466667
ICV-PSC	167	0.1391667
ICV-Vox	23	0.0191667

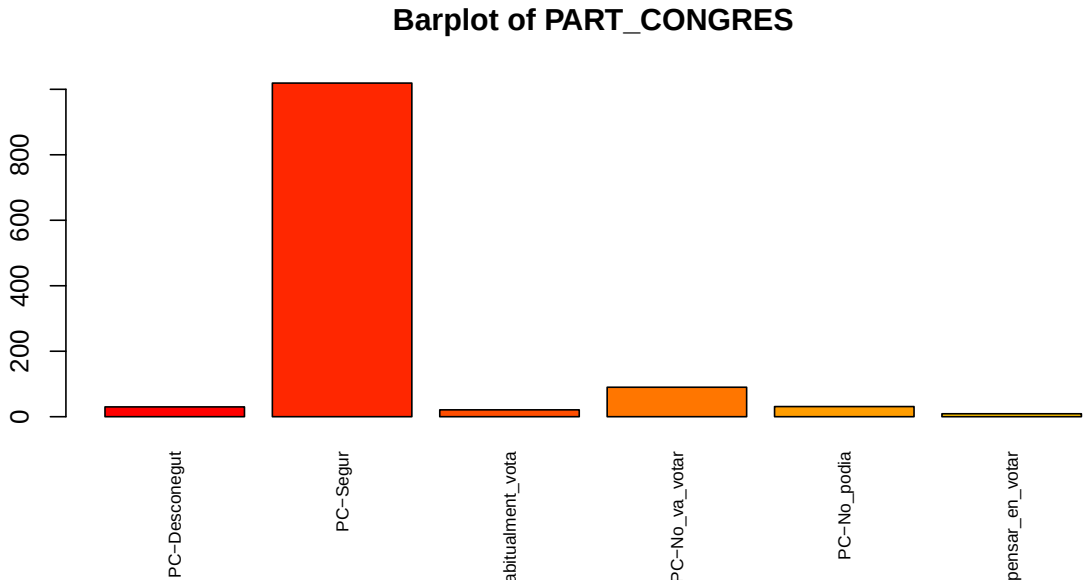


Table 51: Information about modalities of the Variable

Modalities	Frequency	Proportion
PC-Desconegut	30	0.0250000
PC-Segur	1019	0.8491667
PC-No_pero_habitualment_vota	21	0.0175000
PC-No_va_votar	90	0.0750000
PC-No_podia	31	0.0258333
PC-No_pero_va_pensar_en_votar	9	0.0075000

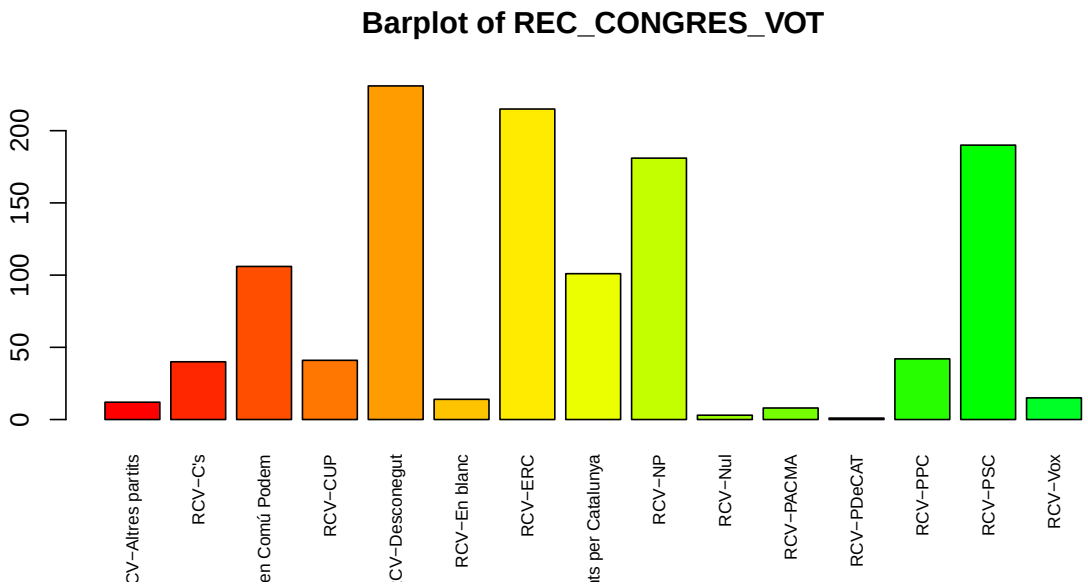


Table 52: Information about modalities of the Variable

Modalities	Frequency	Proportion
RCV-Altres partits	12	0.0100000
RCV-C's	40	0.0333333
RCV-Catalunya en Comú Podem	106	0.0883333
RCV-CUP	41	0.0341667
RCV-Desconegut	231	0.1925000
RCV-En blanc	14	0.0116667
RCV-ERC	215	0.1791667
RCV-Junts per Catalunya	101	0.0841667
RCV-NP	181	0.1508333
RCV-Nul	3	0.0025000
RCV-PACMA	8	0.0066667
RCV-PDeCAT	1	0.0008333
RCV-PPC	42	0.0350000
RCV-PSC	190	0.1583333
RCV-Vox	15	0.0125000

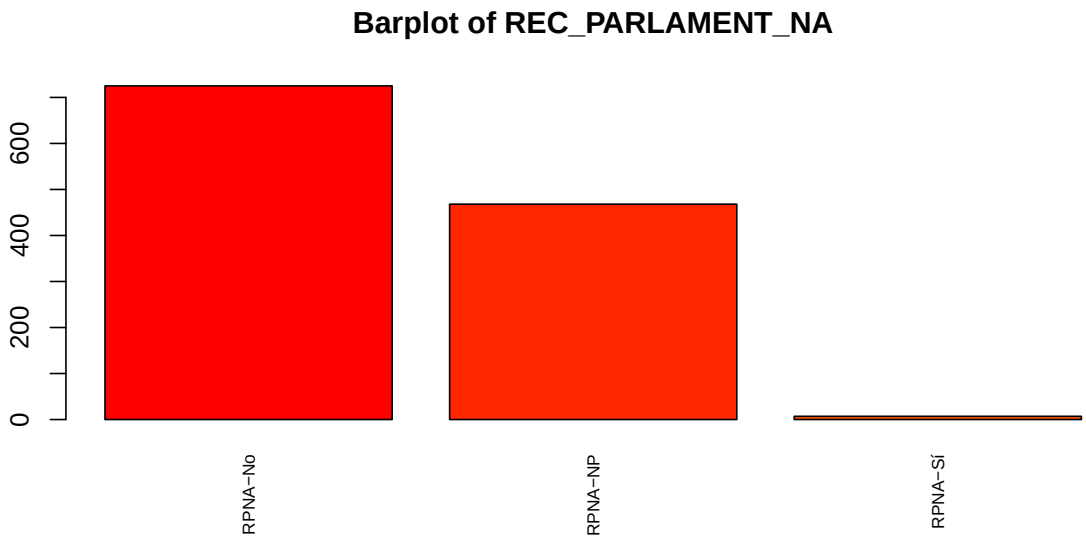


Table 53: Information about modalities of the Variable

Modalities	Frequency	Proportion
RPNA-No	725	0.6041667
RPNA-NP	468	0.3900000
RPNA-Sí	7	0.0058333

Metadata

Numerical Variable Metadata

Table 54: Metadata Numerical Variables

Variable_Name	Type	Range	Units	Missing_Code	Missing_Percent	Meaning	Collection_Method
EDAT	numerical	[18, 90]	years		0	respondent's age	Answer to question to the survey
CAT_0_10	numerical	[0, 10]			0	Answer Question 21	Answer ro question to the survey
ESP_0_10	numerical	[0, 10]			0	Answer Question 22	Answer ro question to the survey
SATISFACCIO_ELECCIONS_14F2021	numerical	[0, 10]			0	Answer Question 32	Answer to question to the survey
POST_PARLAMENT_DECISIO_CRISI	numerical	[0, 10]			0	Answer Question 38 b	Answer to question to the survey
POST_PARLAMENT_DECISIO_CORRUPCIO	numerical	[0, 10]			0	Answer Question 38 c	Answer to question to the survey
POST_PARLAMENT_DECISIO_CANVI_CLIMATIC	numerical	[0, 10]			0	Answer Question 38 d	Answer to question to the survey
POST_PARLAMENT_DECISIO_COVID	numerical	[0, 10]			0	Answer Question 38 e	Answer to question to the survey

Categorical Variables Metadata

Table 55: Metadata Categorical Variables

Variable_Name	Type	Modalities	Missing_Code	Missing_Percent	Meaning & Collection_Method
PROVINCIA	categorical	PROV-BARCELONA PROV-GIRONA PROV-LLEIDA PROV-TARRAGONA		0	respondent's province
HABITAT	categorical	HABI-<2 HABI->1000 HABI-10-150 HABI-150-1000 HABI-2-10 HABI-50-150		0	respondent's habitat
MUNICIPI	categorical	MUNI-BADALONA MUNI-BARCELONA MUNI-CASTELLDEFELS MUNI-CERDANYOLA DEL VALLÈS MUNI-CORNELLÀ DE LLOBREGAT MUNI-Desconegut MUNI-GIRONA MUNI-GRANOLLERS MUNI-HOSPITALET DE LLOBREGAT, L' MUNI-LLEIDA MUNI-MANRESA MUNI-MATARÓ MUNI-MOLLET DEL VALLÈS MUNI-PRAT DE LLOBREGAT, EL MUNI-REUS MUNI-RUBÍ MUNI-SABADELL MUNI-SANT BOI DE LLOBREGAT MUNI-SANT CUGAT DEL VALLÈS MUNI-SANTA COLOMA DE GRAMENET MUNI-TARRAGONA MUNI-TERRASSA MUNI-VILADECANS MUNI-VILANOVA I LA GELTRÚ	MUNI-Desconegut	0.46	respondent's municipally
COMARCA	categorical	COMA-Alt Camp COMA-Alt Empordà COMA-Alt Penedès COMA-Alt Urgell COMA-Alta Ribagorça COMA-Anoia		0	respondent's region

Table 55: Metadata Categorical Variables *(continued)*

Variable_Name	Type	Modalities	Missing_Code	Missing_Percent	Meaning & Collection_Method
		COMA-Aran			
		COMA-Bages COMA-Baix Camp COMA-Baix Ebre COMA-Baix Empordà COMA-Baix Llobregat			
		COMA-Baix Penedès COMA-Barcelonès COMA-Berguedà COMA-Cerdanya COMA-Conca de Barberà			
		COMA-Garraf COMA-Garrigues COMA-Garrotxa COMA-Gironès COMA-Maresme			
		COMA-Moianès COMA-Montsià COMA-Noguera COMA-Osona COMA-Pallars Jussà			
		COMA-Pla d'Urgell COMA-Pla de l'Estany COMA-Priorat COMA-Ribera d'Ebre COMA-Ripollès			
		COMA-Segarra COMA-Segrià COMA-Selva COMA-Solsonès COMA-Tarragonès			
		COMA-Terra Alta COMA-Urgell COMA-Vallès Occidental COMA-Vallès Oriental			
SEXE	binary			0	respondent's gender
		SEXE-Dona SEXE-Home			Answer to question to the survey
EDAT_GR	categorical			0	respondent's group age
		EDG-24 anys o menys			Assigned to a group depending of var EDAT in 15 years interval
		EDG-65 anys o més			
		EDG-De 25 a 34 anys EDG-De 35 a 49 anys EDG-De 50 a 64 anys			
EDAT_CEO	categorical			0	respondent's group age
		EDCEO-15-19			Assigned to a group depending of var EDAT in 5 years interval

Table 55: Metadata Categorical Variables *(continued)*

Variable_Name	Type	Modalities	Missing_Code	Missing_Percent	Meaning & Collection_Method
		EDCEO-20-24 EDCEO-25-29 EDCEO-30-34 EDCEO-35-39 EDCEO-40-44			
		EDCEO-45-49 EDCEO-50-54 EDCEO-55-59 EDCEO-60-64 EDCEO-65-69			
		EDCEO-70-74 EDCEO-75-79 EDCEO-80-mes			
LLOC_NAIX	categorical		LLOCN-Altres_Comunitats	0.06	respondent's place of birth
		LLOCN-Altres_Comunitats			Answer to question to the survey
		LLOCN-Catalunya LLOCN-Resta_mon LLOCN-UE			
SIT_ECO_CAT	categorical		SEC-Desconegut	0.02	Answer Question 1
		SEC-Bona			Answer to question to the survey
		SEC-Desconegut SEC-Dolenta SEC-Molt bona SEC-Molt dolenta SEC-Ni bona ni dolenta			
SIT_ECO_CAT_RETROSPECTIVA	categorical		SECR-Desconegut	0.02	Answer Question 2
		SECR-Desconegut SECR-Igual SECR-Millor SECR-Pitjor			Answer to question to the survey
SIT_ECO_CAT_PROSPECTIVA	categorical		SECP-Desconegut	0.08	Answer Question 3
		SECP-Desconegut SECP-Empitjorarà SECP-Es quedarà igual SECP-Millorarà			Answer to question to the survey
SIT_ECO_ESP	categorical		SEE-Desconegut	0.02	Answer Question 4
		SEE-Bona SEE-Desconegut SEE-Dolenta SEE-Molt bona			Answer to question to the survey
		SEE-Molt dolenta SEE-Ni bona ni dolenta			
SIT_ECO_PERSONAL	categorical		SEP-Desconegut	0.00	Answer Question 5
		SEP-Desconegut SEP-Igual			Answer to question to the survey
		SEP-Millor SEP-Pitjor			

Table 55: Metadata Categorical Variables *(continued)*

Variable_Name	Type	Modalities	Missing_Code	Missing_Percent	Meaning & Collection_Method
SIT_POL_CAT	categorical	SPC-B SPC-D SPC-MB SPC-MD SPC-NBND SPC-NoSap		0	Answer Question 6 Answer to question to the survey
SIT_POL_CAT_RETROSPECTIVA	categorical	SPCR-Desconegut SPCR-Igual SPCR-Millor SPCR-Pitjor	SPCR-Desconegut	0.03	Answer Question 7 Answer to question to the survey
SIT_POL_CAT_PROSPECTIVA	categorical	SPCP-Desconegut SPCP-Empitjorarà SPCP-Igual SPCP-Millorarà	SPCP-Desconegut	0.09	Answer Question 8 Answer to question to the survey
SIT_POL_ESP	categorical	SPE-B SPE-D SPE-Desconegut SPE-MB SPE-MD SPE-NBND	SPE-Desconegut	0.02	Answer Question 9 Answer to question to the survey
CAT_ESP	categorical	CATESP-Cap_o_altra CATESP-Catala_a_Catalunya CATESP-Catala_a_Espanya CATESP-Desconegut CATESP-Espanyol_a_Catalunya CATESP-Espanyol_a_Espanya	CATESP-Desconegut	0.02	Answer Question 10 Answer to question to the survey
IDEOL_1_7	categorical	IDEO-Centre IDEO-Centre-dreta IDEO-Centre-esquerra IDEO-Desconegut IDEO-Dreta IDEO-Esquerra IDEO-Extrema dreta IDEO-Extrema esquerra	IDEO-Desconegut	0.14	Answer Question 20 Answer ro question to the survey
MONARQUIA_REPUBLICA	binary	MONREP-Desconegut MONREP-Monarquia MONREP-República	MONREP-Desconegut	0.13	Answer Question 23 Answer ro question to the survey
SIGNIFICA_ESP	categorical	SIGNESP-Desconegut SIGNESP-Diverses_nacionalitats_regions SIGNESP-Em_sento_membre	SIGNESP-Desconegut	0.06	Answer Question 24 Answer ro question to the survey

Table 55: Metadata Categorical Variables *(continued)*

Variable_Name	Type	Modalities	Missing_Code	Missing_Percent	Meaning & Collection_Method
		SIGNESP-Estat_alie SIGNESP-Meu_Pais SIGNESP-Soc_ciutada			
ACTITUD_INDEPENDENCIA	binary	ACTIND-Desconegut	ACTIND-Desconegut	0.06	Answer Question 28 Answer ro question to the survey
		ACTIND-No ACTIND-Sí			
VOT_DRET_DEURE	categorical	VDD-Ambdós VDD-Desconegut	VDD-Desconegut	0.02	Answer Question 30 Answer to question to the survey
		VDD-Un deure VDD-Un dret			
ELECCIONS_IMPORTANCIA	categorical	EI-Ajuntament (municipals) EI-Altres	EI-Desconegut	0.04	Answer Question 31 Answer to question to the survey
		EI-Autonòmiques i generals EI-Autonòmiques i municipals EI-Cap EI-Congrés dels diputats (generals) EI-Desconegut			
		EI-Generals i municipals EI-Parlament de Catalunya (autonòmiques) EI-Parlament europeu EI-Totes igual			
PART_PARLAMENT	categorical		PP-Desconegut	0.01	Answer Question 33
		PP-Desconegut PP-No_pero_habitualment_vota PP-No_pero_va_pensar_en_votar PP-No_podia PP-No_va_votar			Answer to question to the survey
		PP-Segur			
REC_PARLAMENT_VOT_DECIDIT	categorical	RPVD-Darrera campanya RPVD-Desconegut RPVD-Dia eleccions	RPVD-NP	0.22	Answer Question 34 Only asked if PART_PARLAMENT = Vote
		RPVD-Fa temps RPVD-Inici campanya RPVD-NP			
REC_PARLAMENT_VOT	categorical	RPV-Altres partits	RPV-NP	0.22	Answer Question 35 Only asked if PART_PARLAMENT = Vote
		RPV-C's RPV-Catalunya en Comú Podem RPV-CUP RPV-Desconegut RPV-En blanc			

Table 55: Metadata Categorical Variables *(continued)*

Variable_Name	Type	Modalities	Missing_Code	Missing_Percent	Meaning & Collection_Method
		RPV-ERC RPV-Junts per Catalunya RPV-NP RPV-Nul RPV-PACMA RPV-PDeCAT RPV-PPC RPV-PSC RPV-Vox			
REC_PARLAMENT_FERMESA	Categorical		RPF-NP	0.39	Answer Question 36
		RPF-Amb certs dubtes RPF-Amb convicció RPF-Desconegut RPF-NP RPF-Perquè es tractava d'un mal menor			Answer to question to the survey
REC_PARLAMENT_CANDIDAT	Categorical		RPC-NP	0.39	Answer Question 37
		RPC-No			The enquestee selected this option as one of the two options
		RPC-NP RPC-Sí			
REC_PARLAMENT_PARTIT	binary		RPP-NP	0.39	Answer Question 37
		RPP-No			The enquestee selected this option as one of the two options
		RPP-NP RPP-Sí			
REC_PARLAMENT_INTERESSOS	binary		RPIN-NP	0.39	Answer Question 37
		RPIN-No			The enquestee selected this option as one of the two options
		RPIN-NP RPIN-Sí			
REC_PARLAMENT_IDEES	binary		RPID-NP	0.39	Answer Question 37
		RPID-No			The enquestee selected this option as one of the two options
		RPID-NP			
		RPID-Sí			
REC_PARLAMENT_ACTUACIO	binary		RPA-NP	0.39	Answer Question 37
		RPA-No			The enquestee selected this option as one of the two options
		RPA-NP RPA-Sí			
REC_PARLAMENT_ACTUACIO_2	binary		RPA2-NP	0.39	Answer Question 37
		RPA2-No			The enquestee selected this option as one of the two options
		RPA2-NP RPA2-Sí			
REC_PARLAMENT_EVITAR	binary		RPE-NP	0.39	Answer Question 37
		RPE-No			The enquestee selected this option as one of the two options
		RPE-NP			

Table 55: Metadata Categorical Variables *(continued)*

Variable_Name	Type	Modalities	Missing_Code	Missing_Percent	Meaning & Collection_Method
REC_PARLAMENT_ALTRES	binary	RPE-Sí RPALT-No RPALT-NP RPALT-Sí	RPALT-NP	0.39	Answer Question 37 The enquestee selected this option as one of the two options
REC_PARLAMENT_CAP	binary	RPCAP-No RPCAP-NP RPCAP-Sí	RPCAP-NP	0.39	Answer Question 37 The enquestee selected this option as one of the two options
POST_PARLAMENT_DECISIO_RELACIO	binary			0	Answer Question 38 a
INT_PARLAMENT_VOT	categorical	IPV-Altres partits IPV-C's IPV-Catalunya en Comú Podem IPV-CUP IPV-Desconegut IPV-En blanc IPV-ERC IPV-Junts per Catalunya IPV-No podria votar IPV-No votaria IPV-Nul IPV-PACMA IPV-PDeCAT IPV-PPC IPV-PSC IPV-Vox	IPV-Desconegut	0.27	Answer Question 39 Answer to question to the survey
PART_PARLAMENT_ANTERIORS	categorical	PPA-Desconegut PPA-No_pero_habitualment_vota PPA-No_pero_va_pensar_en_votar PPA-No_podia PPA-No_va_votar PPA-Segur	PPA-Desconegut	0.03	Answer Question 40 Only asked if PART_PARLAMENT_ANTERIORS = Vote
REC_PARLAMENT_VOT_ANTERIORS	categorical	RPVA-Altres partits RPVA-C's RPVA-Catalunya en Comú Podem RPVA-CUP RPVA-Desconegut RPVA-En blanc RPVA-ERC RPVA-Junts per Catalunya	RPVA-NP	0.14	Answer Question 41 Only asked if PART_PARLAMENT_ANTERIORS = Vote AND Selected = Altres partits as answer

Table 55: Metadata Categorical Variables *(continued)*

Variable_Name	Type	Modalities	Missing_Code	Missing_Percent	Meaning & Collection_Method
		RPVA-NP RPVA-Nul RPVA-PACMA RPVA-PDeCAT RPVA-PPC			
		RPVA-PSC RPVA-Vox			
INT_CONGRES_VOT	categorical		ICV-Desconegut	0.28	Answer Question 42 Answer to question to the survey
		ICV-Altres partits ICV-C's			
		ICV-Catalunya en Comú Podem ICV-CUP ICV-Desconegut ICV-En blanc ICV-ERC			
		ICV-Junts per Catalunya ICV-No votaria ICV-Nul ICV-PACMA ICV-PDeCAT			
		ICV-PPC ICV-PSC ICV-Vox			
PART_CONGRES	categorical		PC-Desconegut	0.02	Answer Question 43 Answer to question to the survey
		PC-Desconegut			
		PC-No_pero_habitualment_vota PC-No_pero_va_pensar_en_votar PC-No_podia PC-No_va_votar PC-Segur			
REC_CONGRES_VOT	categorical		RCV-NP	0.15	Answer Question 44 Only asked if in PART_CONGRES Selected = Altres partits as answer
		RCV-Altres partits			
		RCV-C's RCV-Catalunya en Comú Podem RCV-CUP			
		RCV-Desconegut RCV-En blanc RCV-ERC RCV-Junts per Catalunya RCV-NP			
		RCV-Nul RCV-PACMA RCV-PDeCAT RCV-PPC RCV-PSC			
		RCV-Vox			
REC_PARLAMENT_NA	categorical		RPNA-NP	0.39	Answer Question 37

Table 55: Metadata Categorical Variables *(continued)*

Variable_Name	Type	Modalities	Missing_Code	Missing_Percent	Meaning & Collection_Method
		RPNA-No			The enquestee selected this option as one of the two options
		RPNA-NP			
		RPNA-Sí			

PCA

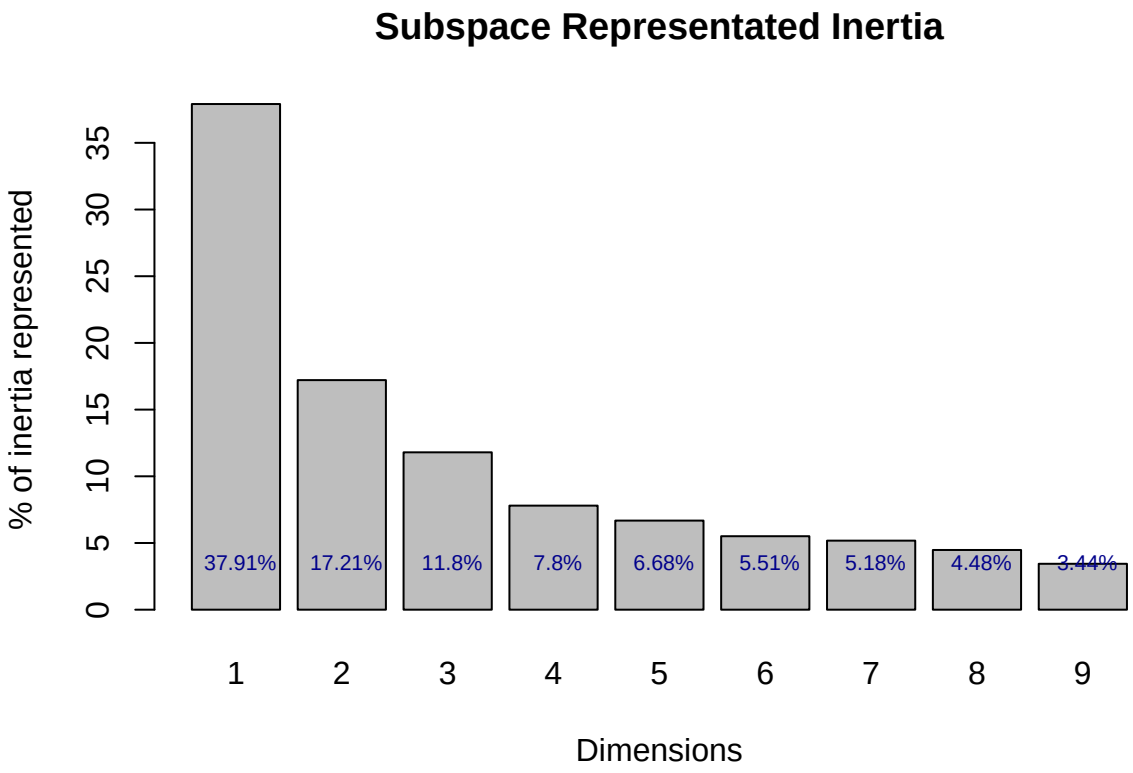
In this section we're going to execute a PCA and select and analyze only one factorial map, because we have a lot of categorical variables in this dataset and we're going to analyze quite a bit of them. That's because, when we tried to plot fewer variables in multiple factorial maps, the extra information that they gave was very little. Then, we decided to focus on more variables in the best factorial map we could find.

We start the process by getting the numeric variables of the dataframe.

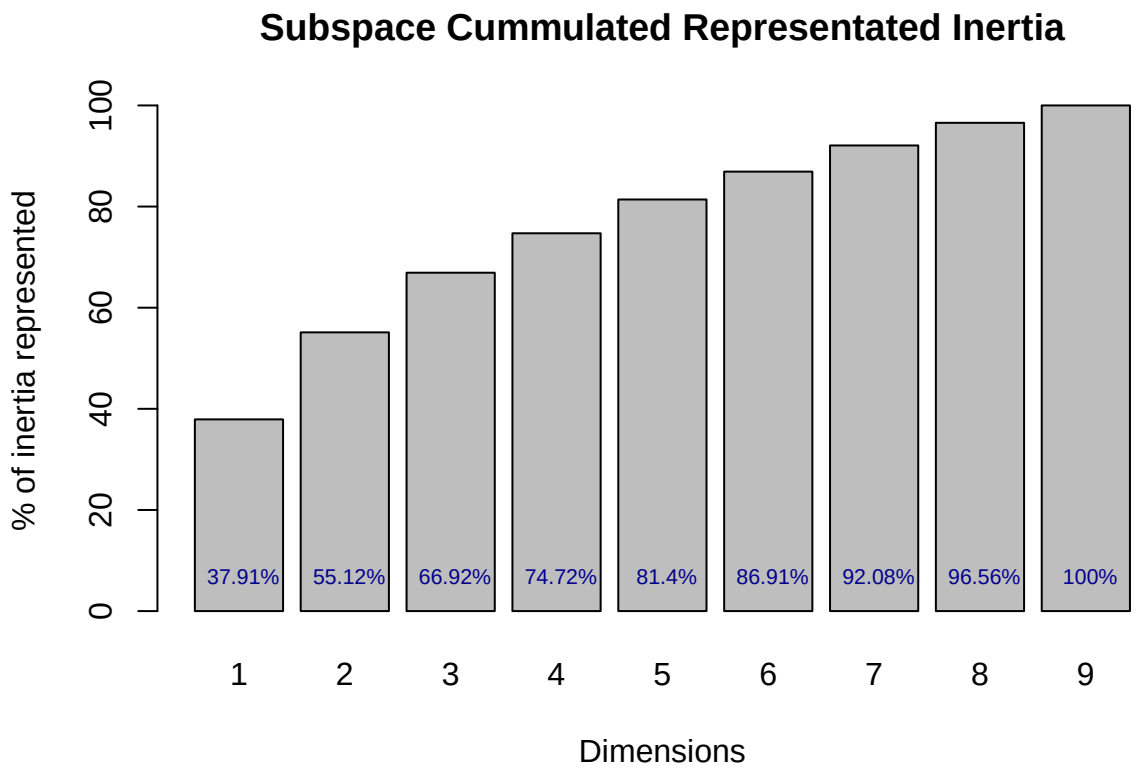
We execute the PCA function.

Subspace Inertia

We graph the inertia represented by each dimension of the PCA, ordered decreasingly.



We now graph the cummulated inertia.

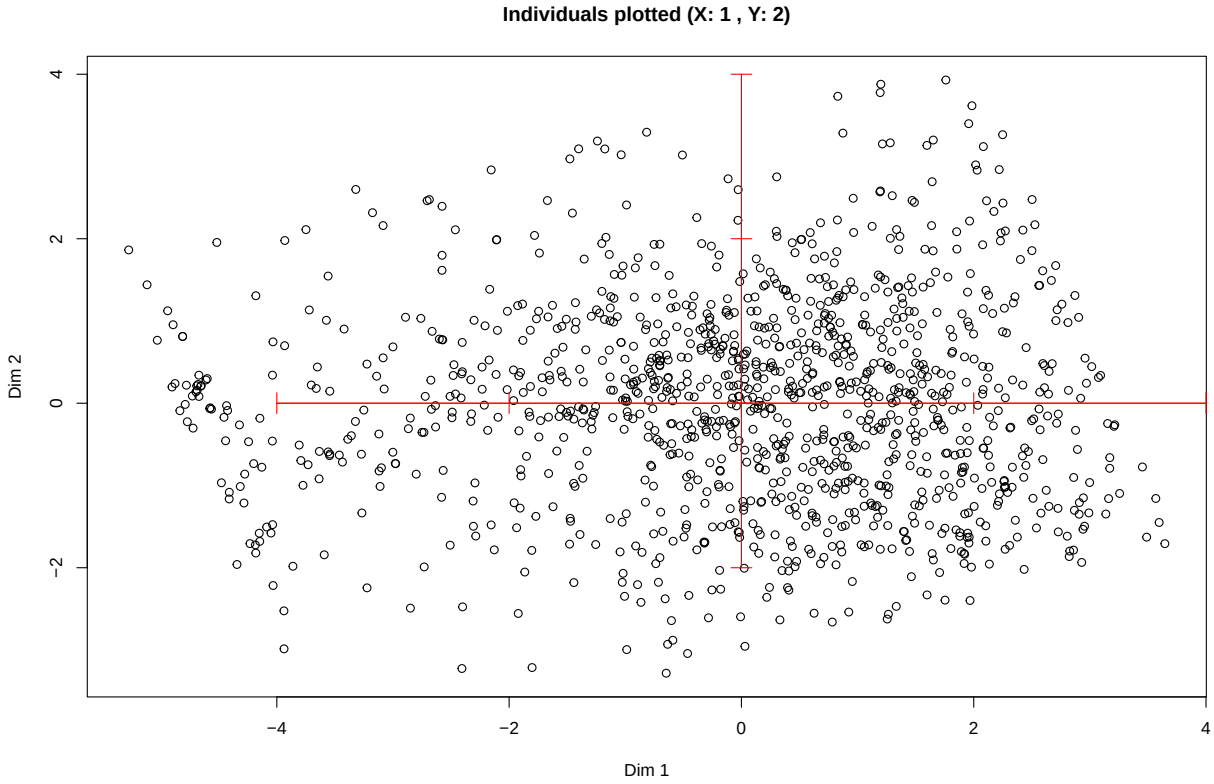


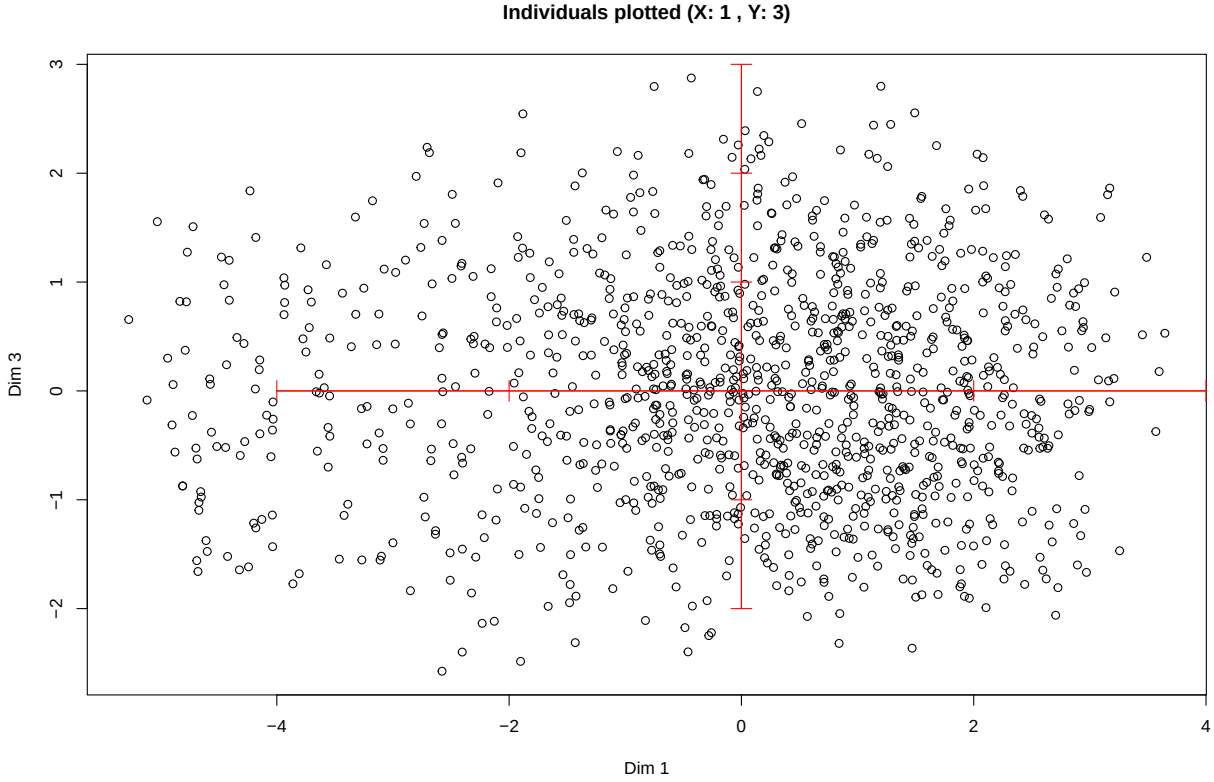
Now we select the number of dimensions we want to keep when getting information from the PCA. The criteria we'll use is to keep the number of dimensions which represent at least 80% of the total inertia.

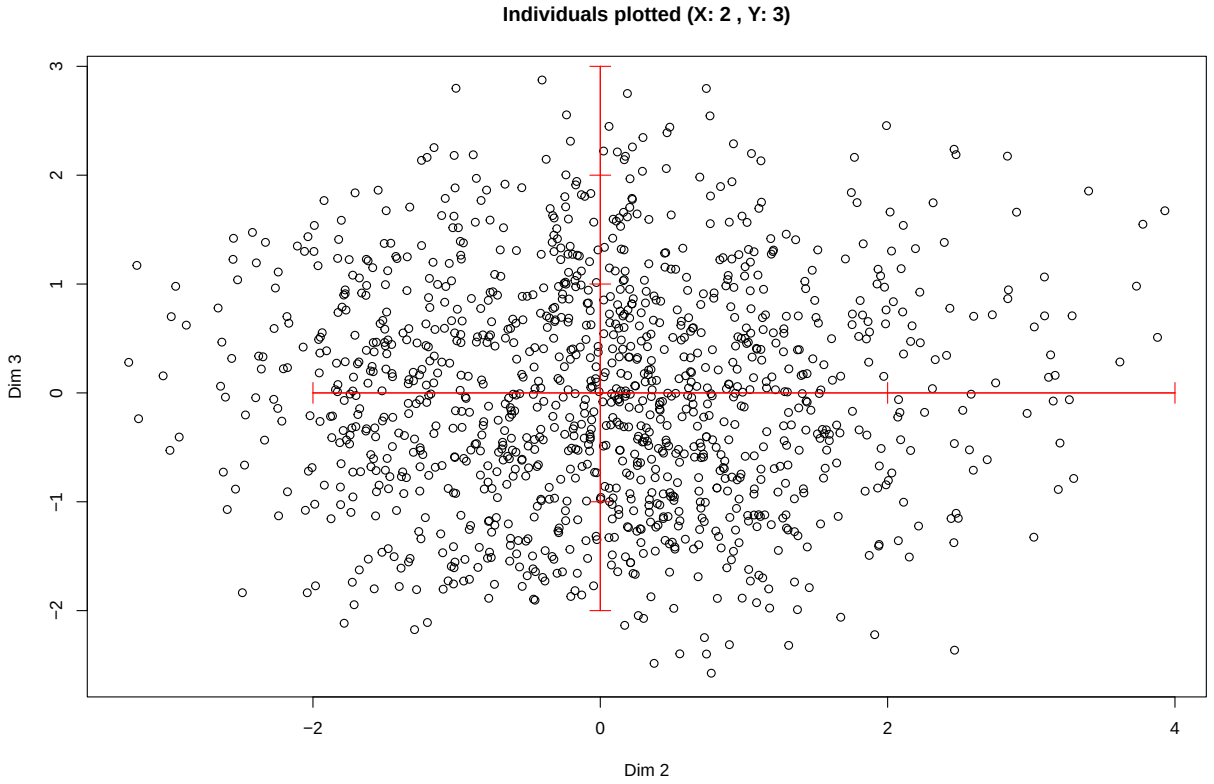
We'll take 5, which represents 81.4% of the total inertia.

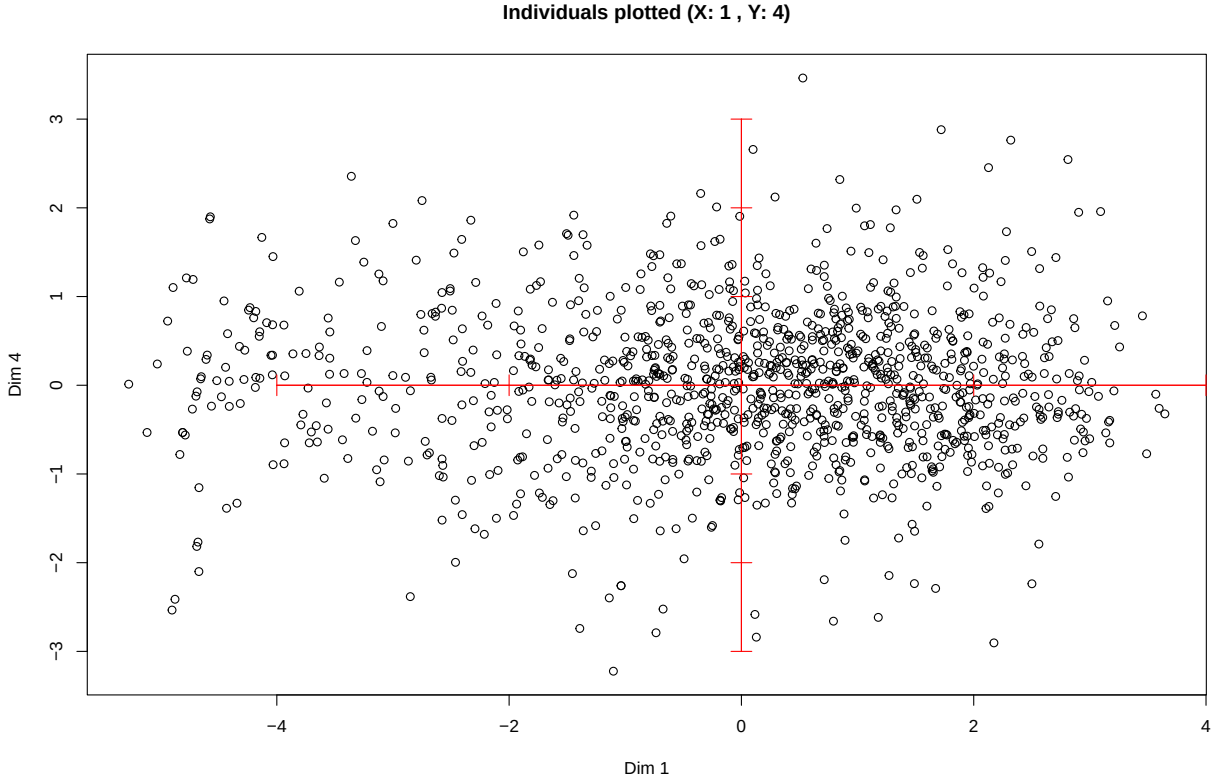
Choosing subspace

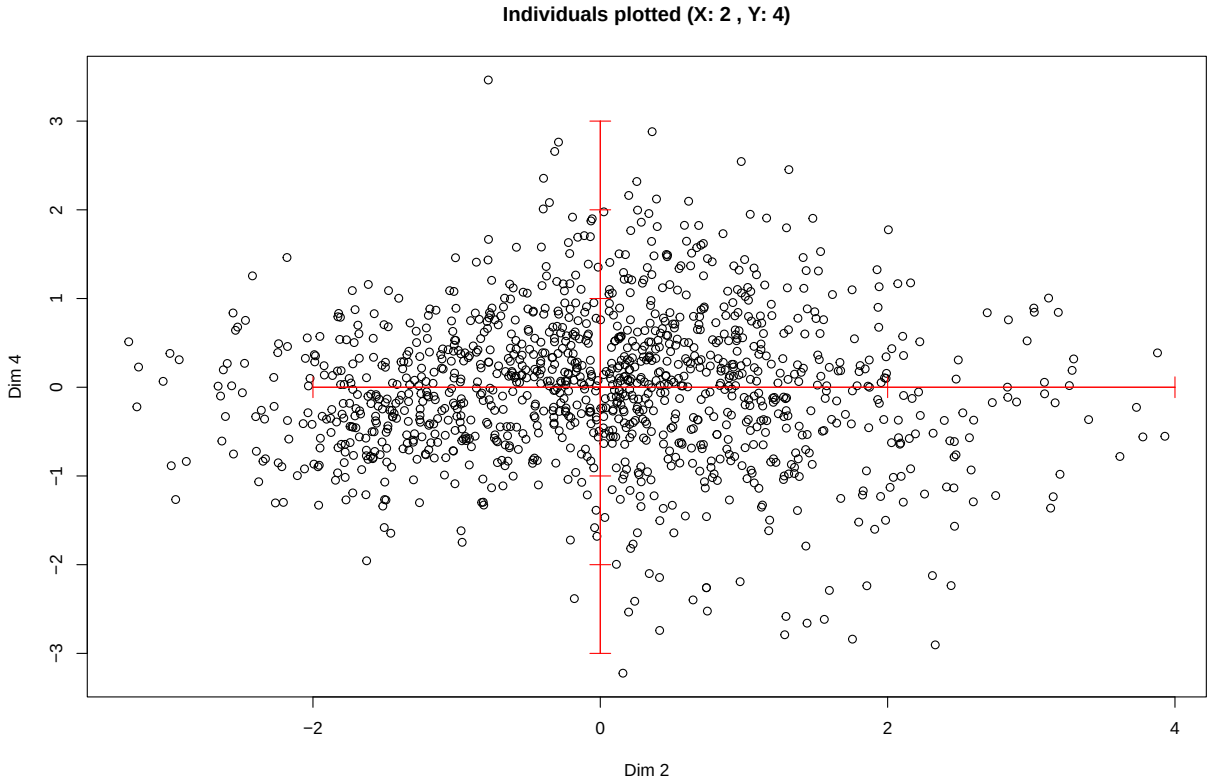
To choose which subspace we're gonna use, we'll plot the individuals in each possible subspace from the components we kept.

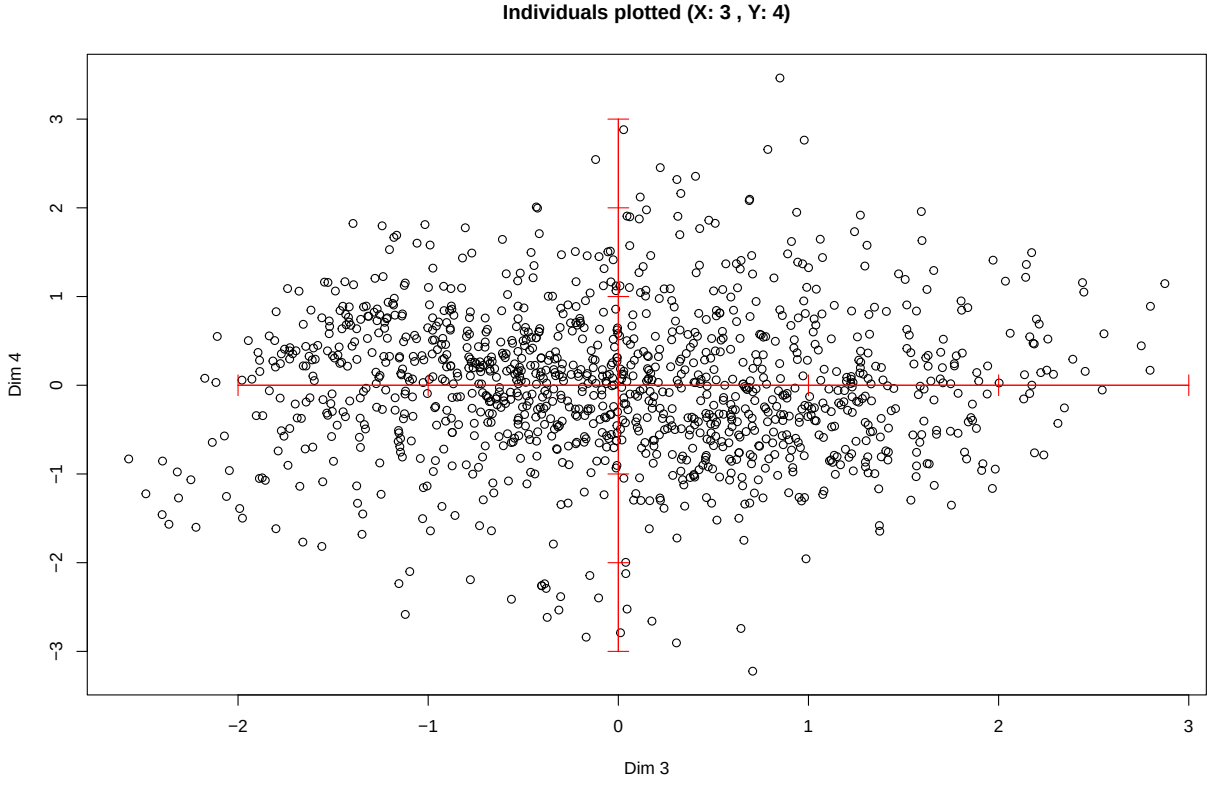

Individuals plotted (X: 1 , Y: 2)
Dim 2
Dim 1
4
2
0
−2
−4
−2
0
2
4

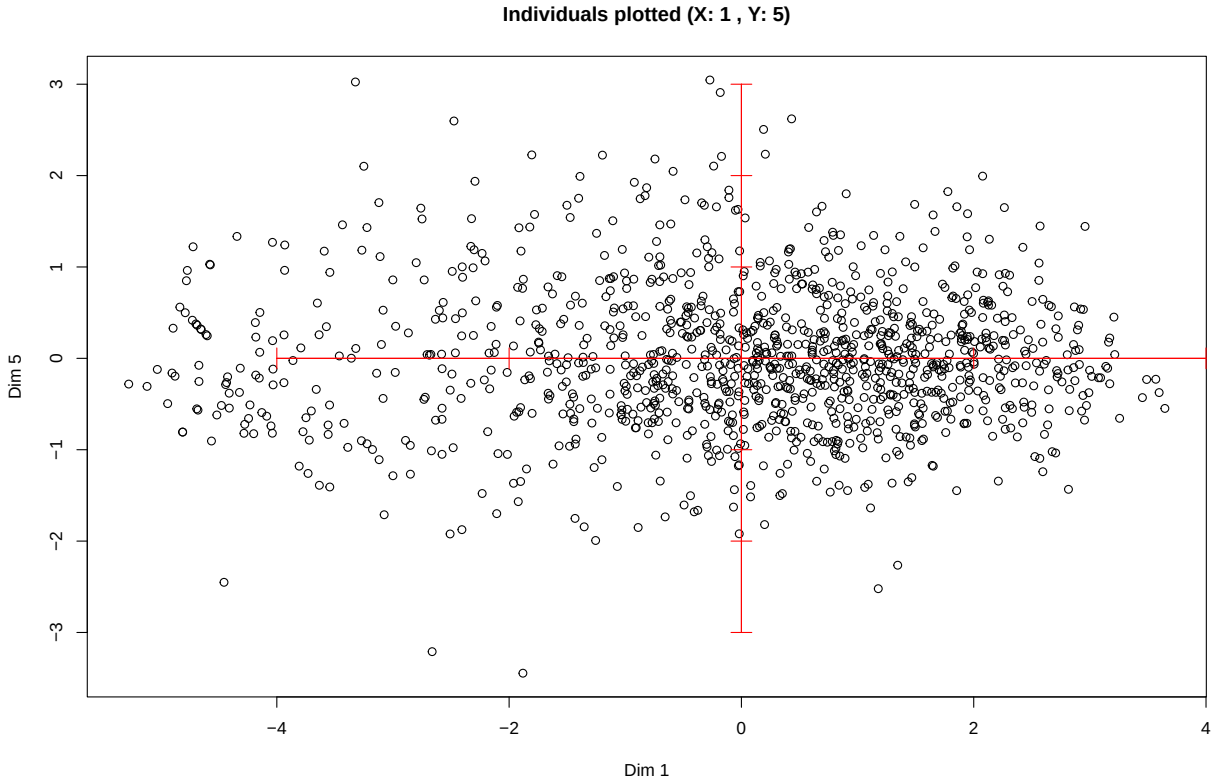

Individuals plotted (X: 1 , Y: 3)
Dim 3
Dim 1
3
2
1
0
−1
−2
−4
−2
0
2
4

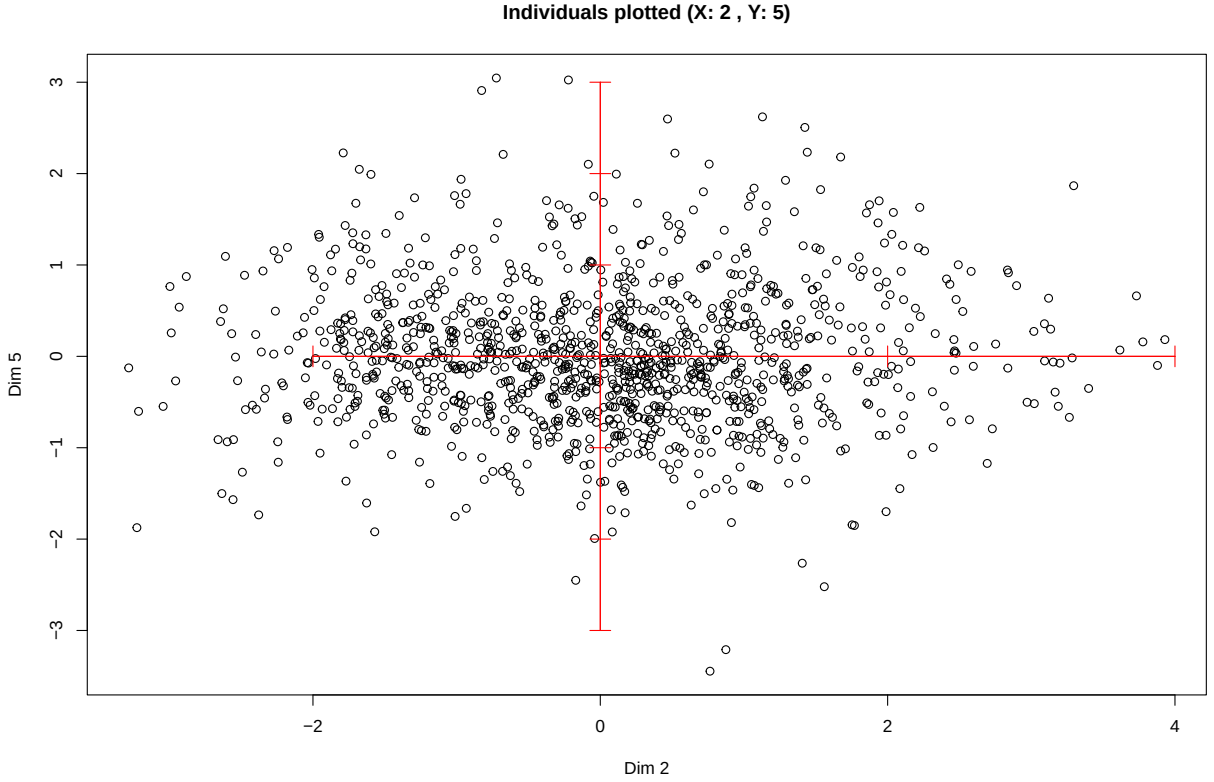

Individuals plotted (X: 2 , Y: 3)
Dim 3
Dim 2
3
2
1
0
−1
−2
−2
0
2
4

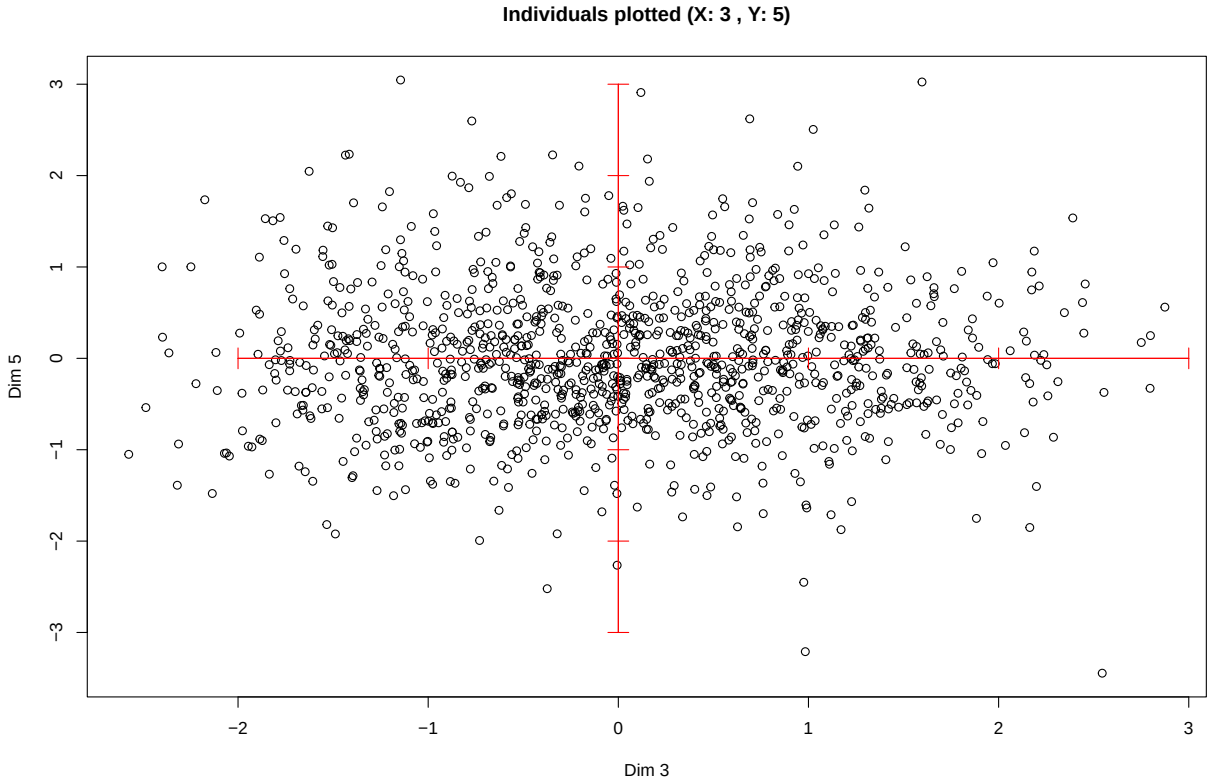

Individuals plotted (X: 1 , Y: 4)
Dim 4
Dim 1
3
2
1
0
−1
−2
−3
−4
−2
0
2
4

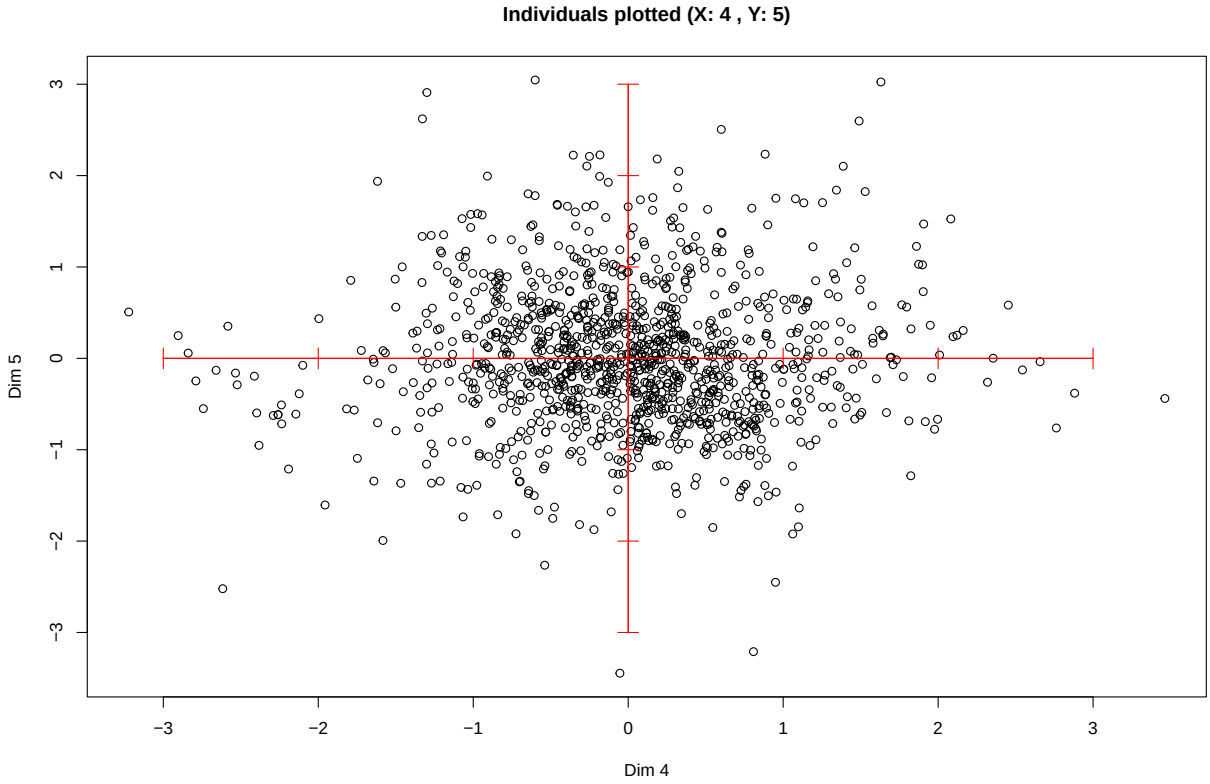

Individuals plotted (X: 2 , Y: 4)
Dim 4
Dim 2
3
2
1
0
−1
−2
−3
−2
0
2
4


Individuals plotted (X: 3 , Y: 4)
Dim 4
Dim 3
3
2
1
0
−1
−2
−3
−2
−1
0
1
2
3


Individuals plotted (X: 1 , Y: 5)
Dim 5
Dim 1
3
2
1
0
−1
−2
−3
−4
−2
0
2
4

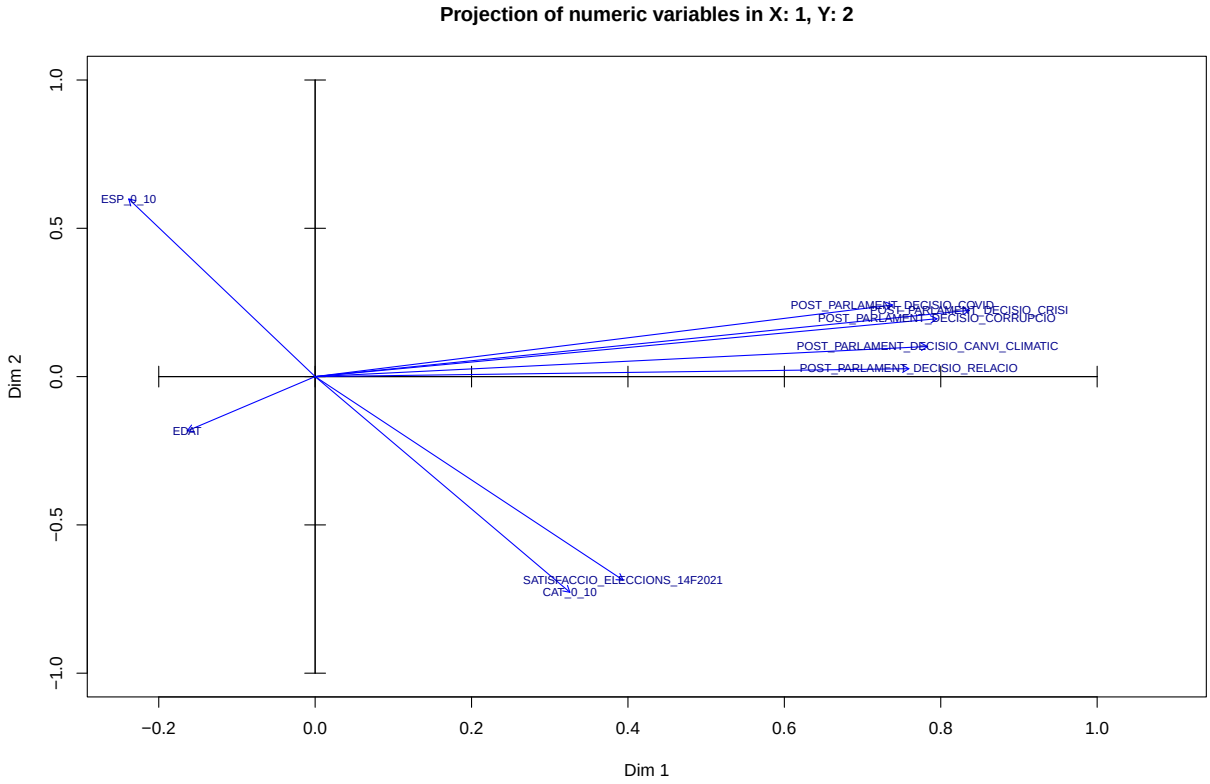

Individuals plotted (X: 2 , Y: 5)
Dim 5
Dim 2
3
2
1
0
−1
−2
−3
−2
0
2
4

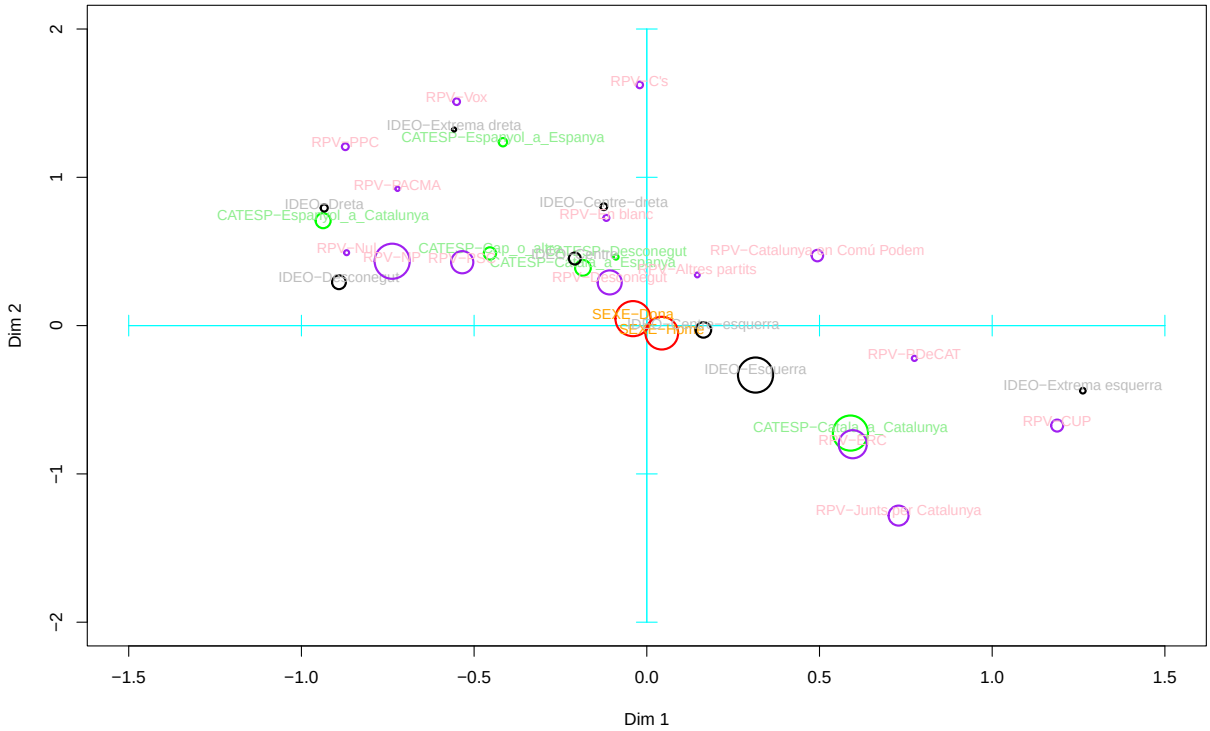


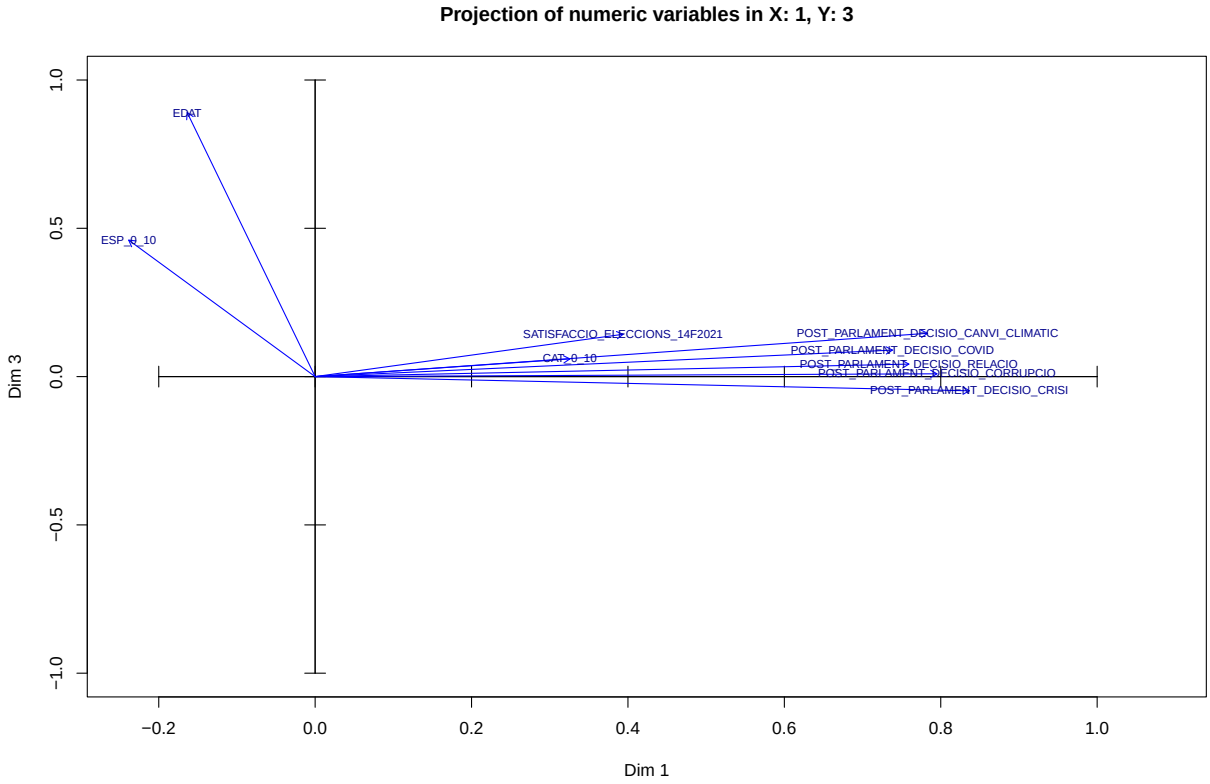


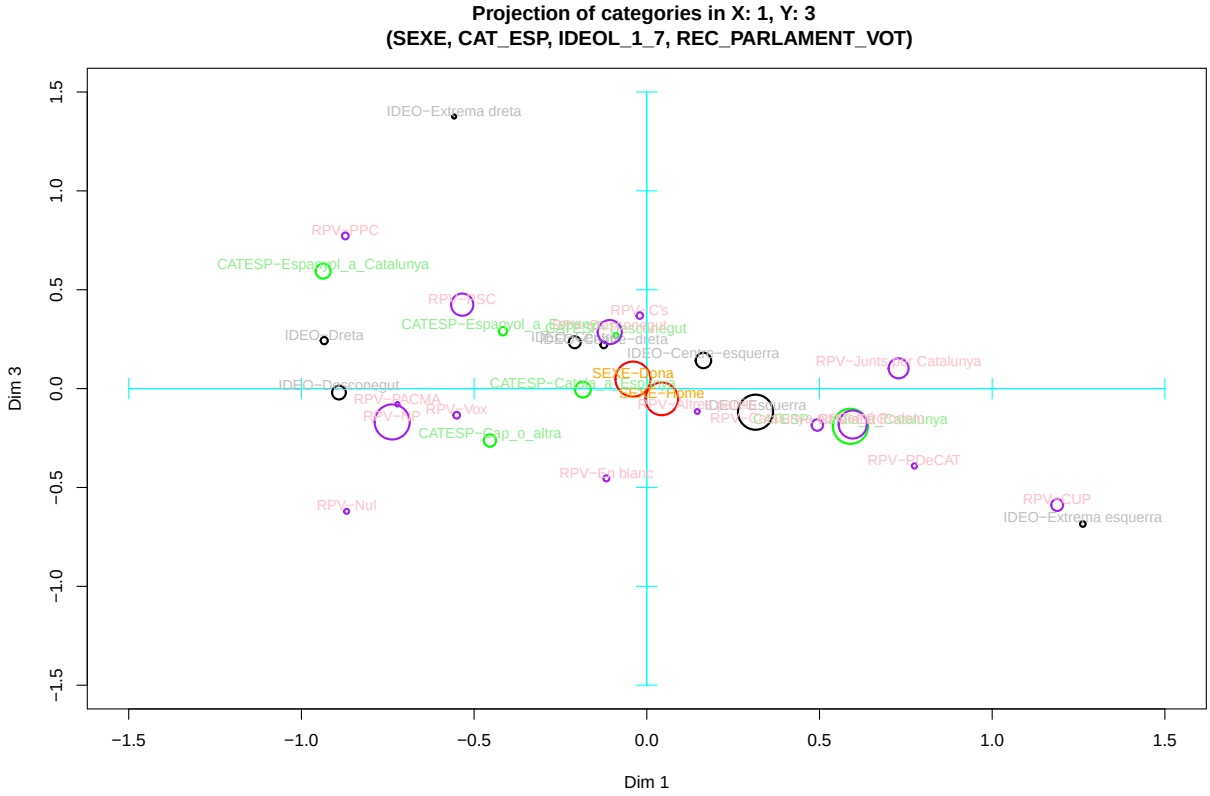
By these different plots, we can see that some have distributions too close to the center, which means that

they represent smaller variability. Taking in to account the ammount of inertia each dimension represents, we are left with 2 possible subspaces: (1,2) and (1,3), with the former representing a slightly more inertia.


Projection of numeric variables in X: 1, Y: 2
ESP_0_10
EDAT
POST_PARLAMENT_DECISIO_COVID
POST_PARLAMENT_DECISIO_CRISI
POST_PARLAMENT_DECISIO_CORRUPCIO
POST_PARLAMENT_DECISIO_CANVI_CLIMATIC
POST_PARLAMENT_DECISIO_RELACIO
SATISFACCIO_ELECCIONS_14F2021
CAT_0_10
Dim 1
Dim 2


Projection of categories in X: 1, Y: 2
(SEXE, CAT_ESP, IDEOL_1_7, REC_PARLAMENT_VOT)
RPV-C's
RPV-Vox
IDEO-Extrema dreta
RPV-PPC
CATESP-Espanyol_a_Espanya
RPV-PACMA
IDEO-Dreta
CATESP-Espanyol_a_Catalunya
IDEO-Centre-dreta
RPV-En blanc
RPV-Nul
RPV-NP
CATESP-Cap_o_altra
CATESP-Desconegut
RPV-Catalunya en Comú Podem
RPV-Altres partits
IDEO-Desconegut
SEXE-Dona
SEXE-Home
IDEO-Centre-esquerra
RPV-PDeCAT
IDEO-Esquerra
IDEO-Extrema esquerra
RPV-CUP
CATESP-Catalunya_Catalunya
RPV-ERC
RPV-Junts per Catalunya
Dim 1
Dim 2





We select the subspace formed by axis 1 and axis 2 to analyze further because we can see that the vectors

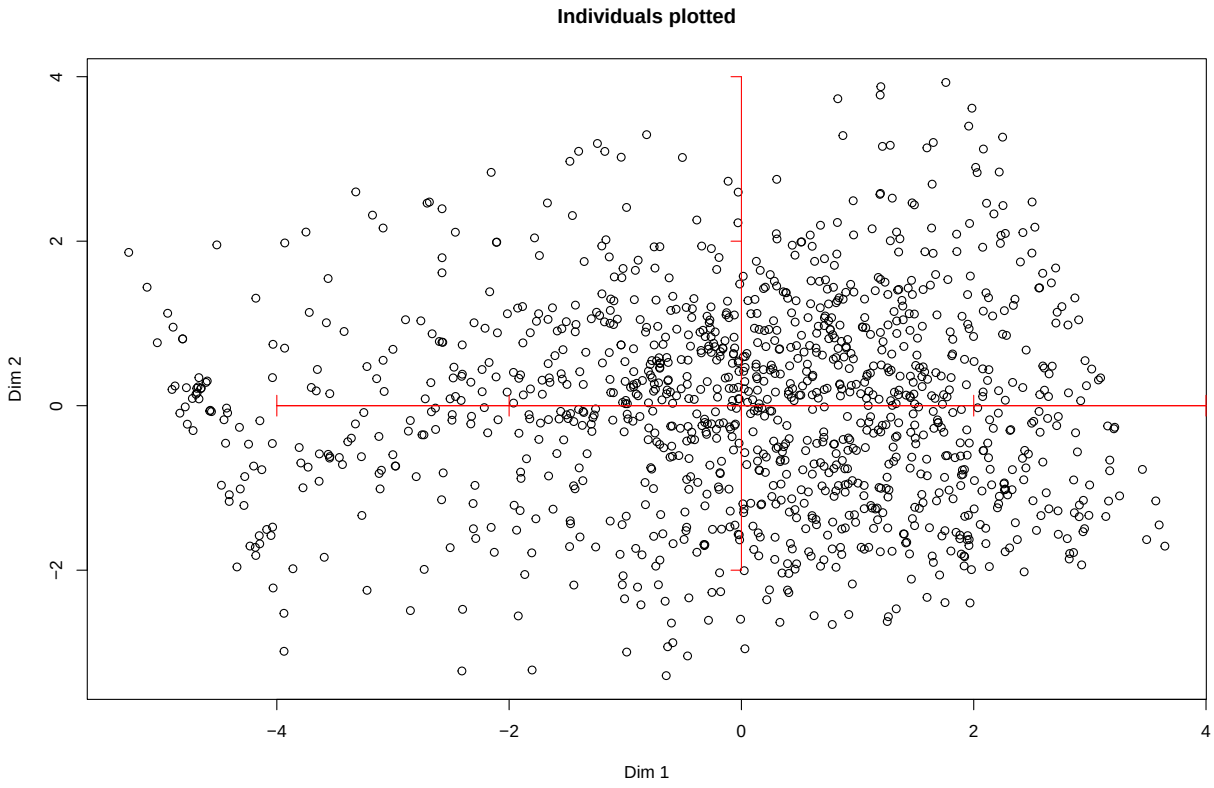
representing the variables are further apart from each other, which means that the axis 2 represents a larger amount of variability, and they are longer, which means that the contribution is also higher.

We have also plotted some categorical variables which we'll analyze to see if they are spread nicely and it seems to be that the choice we made is acceptable.

Factorial map visualization

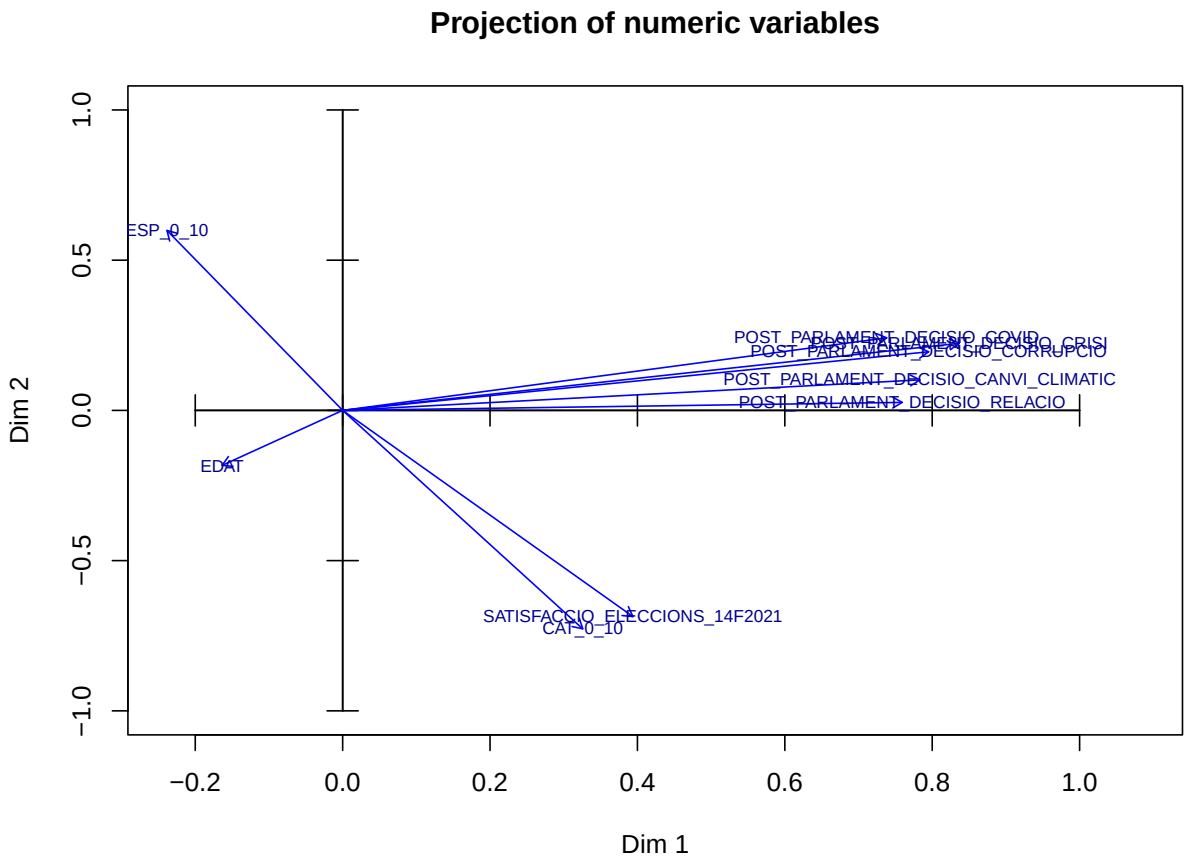
Individual projection

So, here we'll represent the required plots with the selected subspace:



Numerical variables projection and interpretation

The numerical values plotted in the chosen plane are the following:



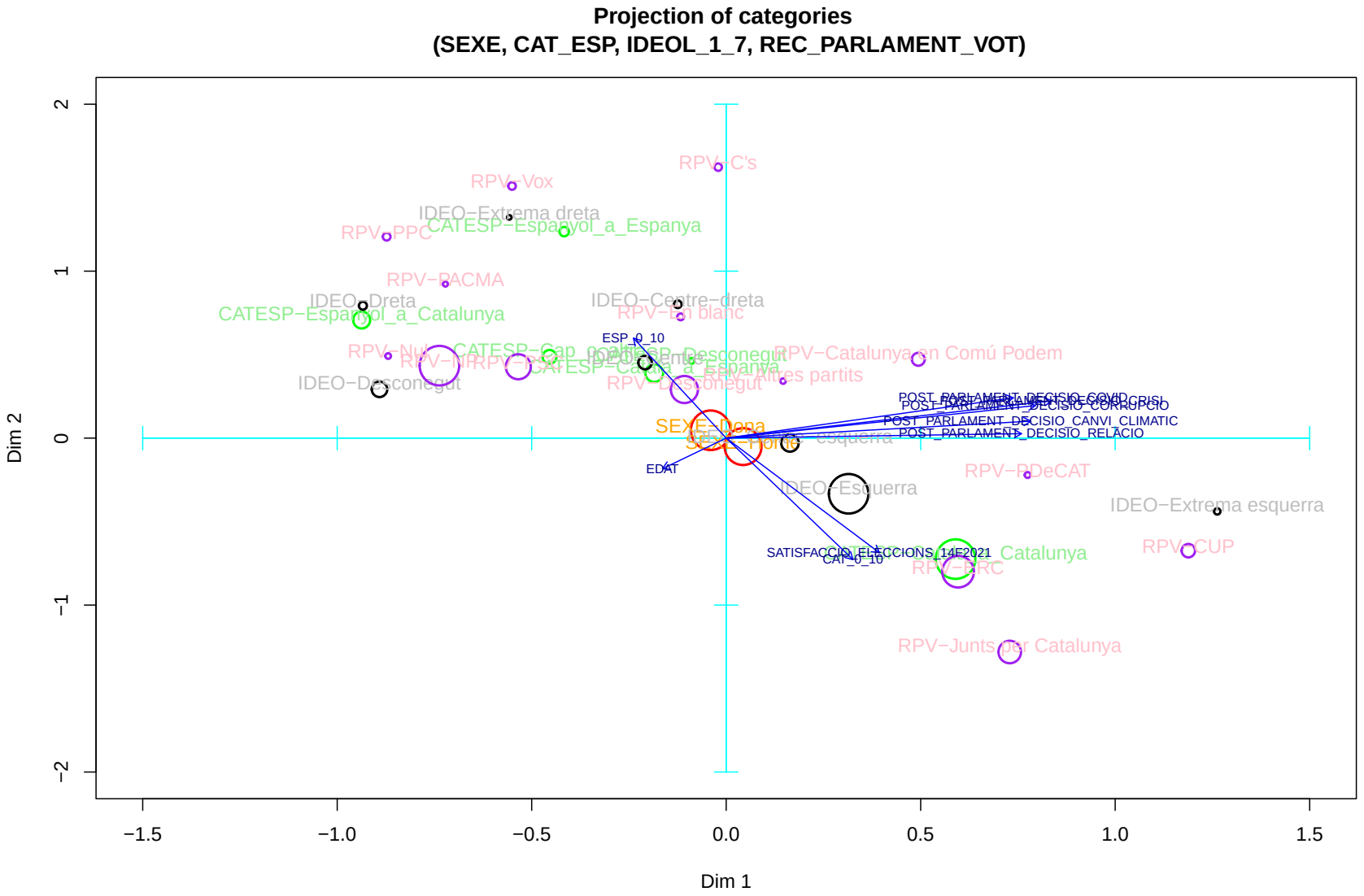
Dimension 1 seems to be associated mainly with the scores that the interviewees gave to various decisions by the *Parlament*, which are also contributing a lot the the subspace. The rest of the variables are also represented in the first component, although not as much, especially age, which is considerably shorter than any other of variables.

Dimension 2 is not as cut and dry, but basically represents the feelings towards Spanish or Catalan sentiment (which are inversely related) and the satisfaction from the elections for the "Parlament de Catalunya" of February 14th.

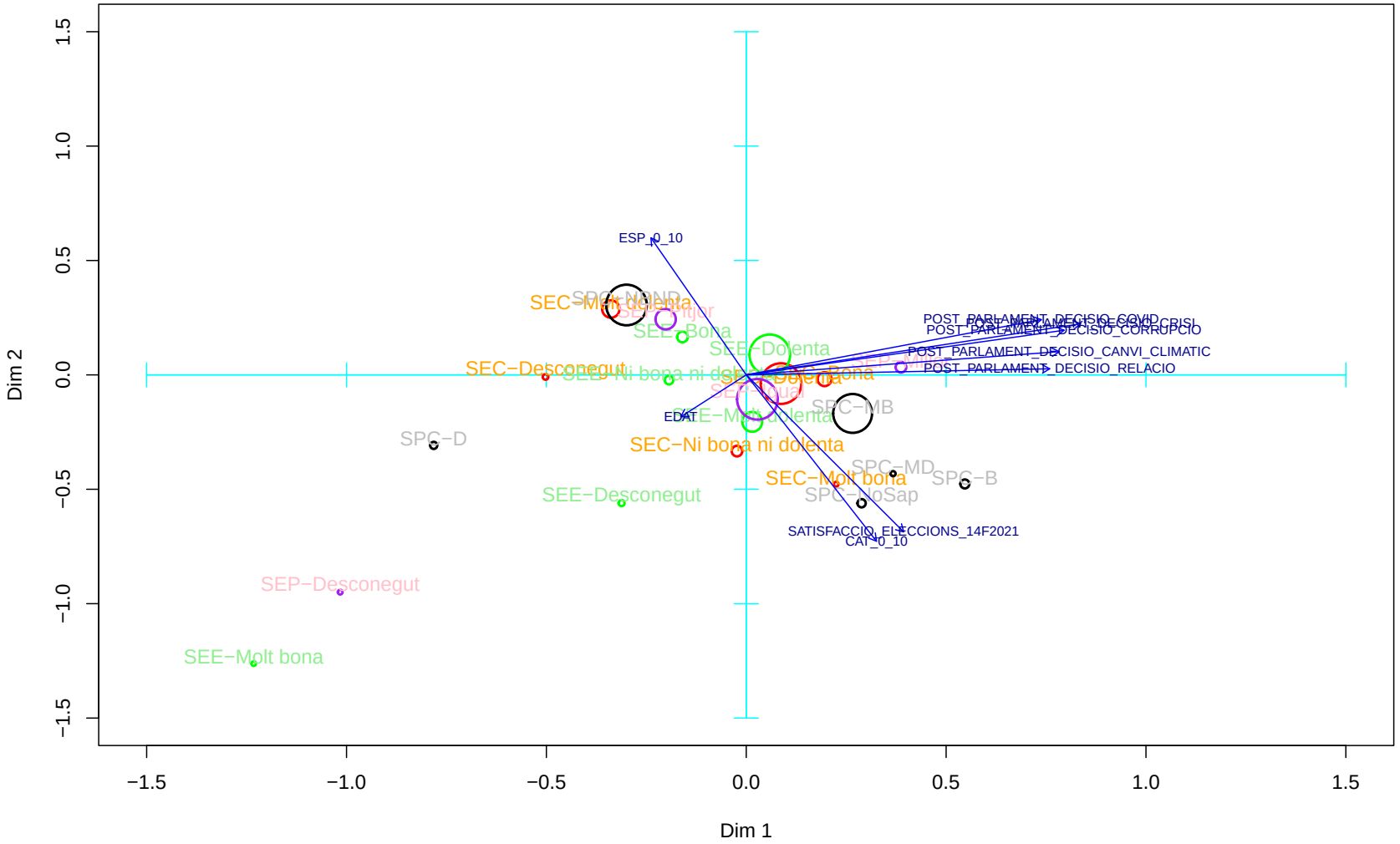
Qualitative variables projection and interpretation

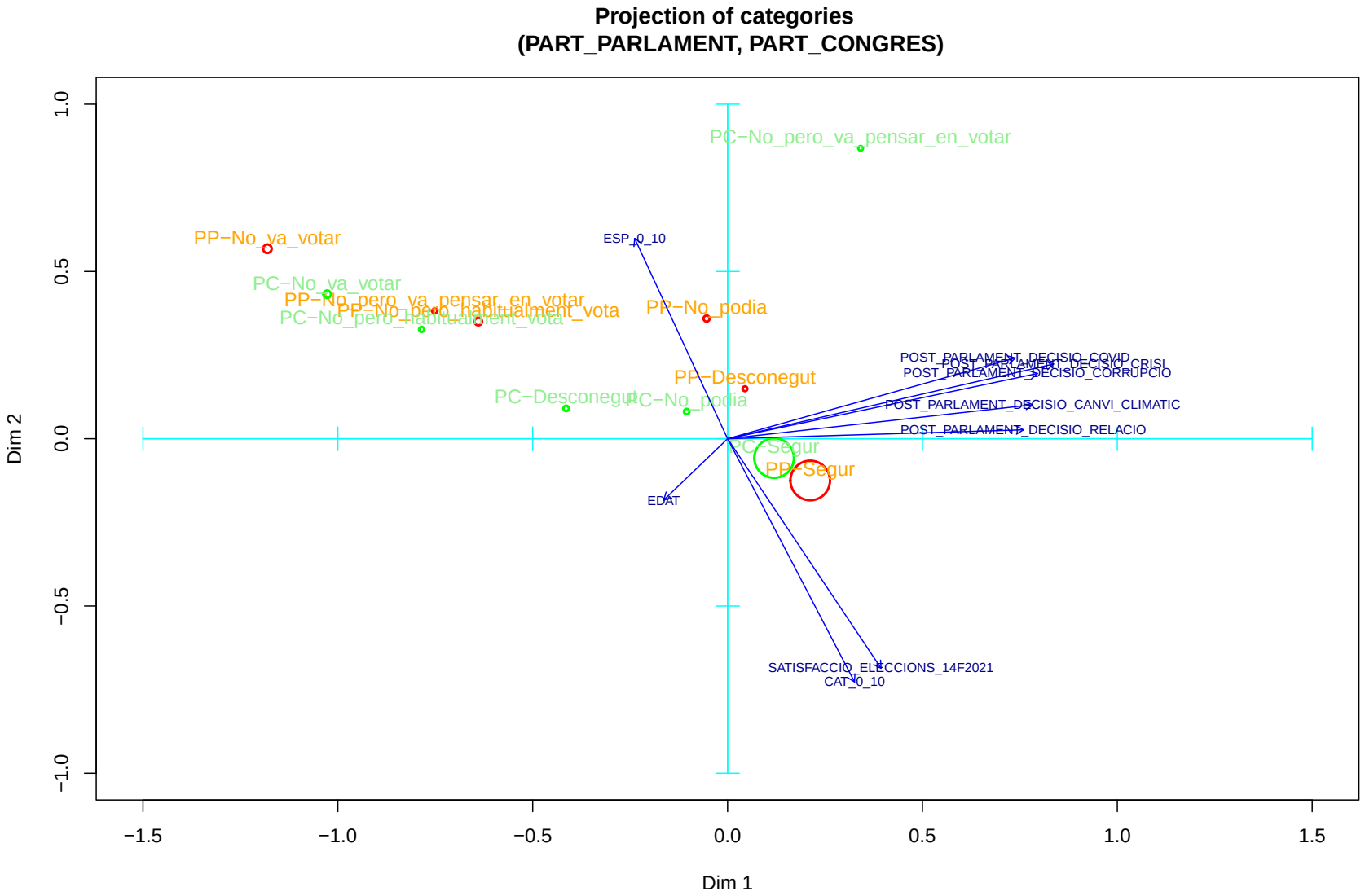
Once the axis are described with the numerical variables from which they were generated, we can analyze the categorical variables.

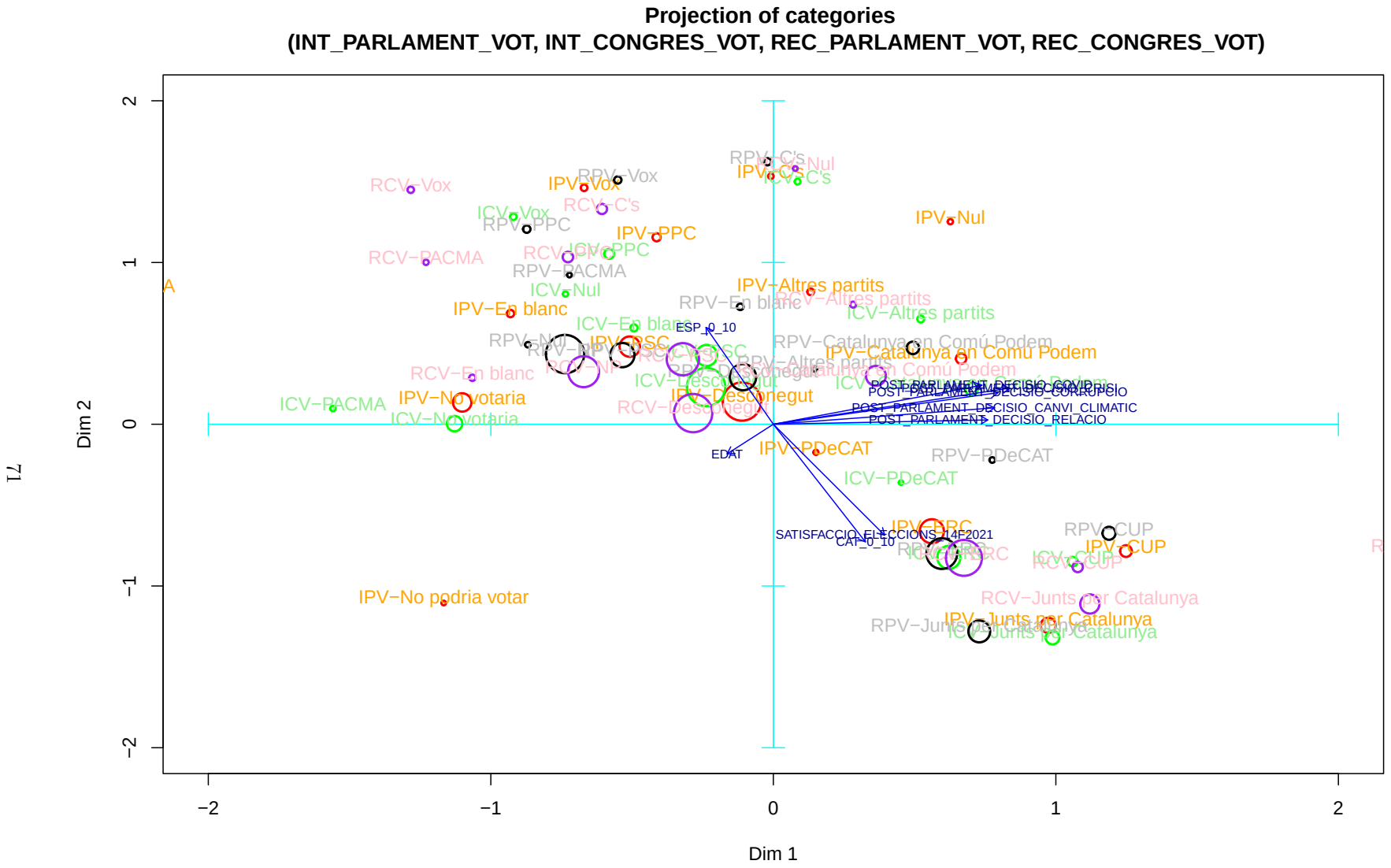
We'll not analyze every categorical variable, because there 44 of those and it would be benyond the scope of the assignment. Then, we have selected 14 of them and we'll analyze them in groups because they are connected in some way.


Projection of categories
(SEXE, CAT_ESP, IDEOL_1_7, REC_PARLAMENT_VOT)
Dim 2
Dim 1
2
1
0
−1
−2
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
RPV−C's
RPV−Vox
IDEO−Extrema dreta
CATESP−Espanyol_a_Espanya
RPV−PPC
RPV−PACMA
IDEO−Dreta
IDEO−Centre−dreta
CATESP−Espanyol_a_Catalunya
RPV−En blanc
ESP_0_10
RPV−Catalunya en Comú Podem
IDEO−Desconegut
RPV−Altres partits
POST_PARLAMENT_DECISIO_COVID
POST_PARLAMENT_DECISIO_CORRUPCIO
POST_PARLAMENT_DECISIO_CANVI_CLIMATIC
POST_PARLAMENT_DECISIO_RELACIO
SEXE−Dona
SEXE−Home
EDAT
RPV−ERC
IDEO−Esquerra
IDEO−Extrema esquerra
RPV−CUP
SATISFACCIO_ELECCIONS_14E2021
CAT_0_10
RPV−Junts per Catalunya

Projection of categories
(SIT_ECO_CAT, SIT_ECO_ESP, SIT_POL_CAT, SIT_ECO_PERSONAL)




Projection of categories
(PART_PARLAMENT, PART_CONGRES)
PC−No_pero_va_pensar_en_votar
PP−No_va_votar
ESP_0_10
PC−No_va_votar
PP−No_pero_va_pensar_en_votar
PP−No_pero_habitualment_vota
PC−No_pero_habitualment_vota
PP−No_podia
POST_PARLAMENT_DECISIO_COVID
POST_PARLAMENT_DECISIO_CRISI
POST_PARLAMENT_DECISIO_CORRUPCIO
PP−Desconegut
PC−Desconegut
PC−No_podia
POST_PARLAMENT_DECISIO_CANVI_CLIMATIC
POST_PARLAMENT_DECISIO_RELACIO
PC−Segur
PP−Segur
EDAT
SATISFACCIO_ELECCIONS_14F2021
CAT_0_10
Dim 2
Dim 1
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5


Projection of categories
(INT_PARLAMENT_VOT, INT_CONGRES_VOT, REC_PARLAMENT_VOT, REC_CONGRES_VOT)
Dim 2
Dim 1
RCV−Vox
IPV−Vox
RPV−Vox
ICV−Vox
RCV−C's
RPV−C's
IPV−C's
ICV−C's
RPV−PPC
IPV−PPC
ICV−PPC
RCV−PACMA
RPV−PACMA
IPV−Nul
ICV−Nul
IPV−Altres partits
RPV−Altres partits
ICV−Altres partits
IPV−En blanc
RPV−En blanc
ICV−En blanc
ESP_0_10
RPV−Catalunya en Comú Podem
IPV−Catalunya en Comú Podem
ICV−Catalunya en Comú Podem
RCV−En blanc
ICV−PACMA
IPV−No votaria
ICV−No votaria
IPV−Desconegut
RCV−Desconegut
POST_PARLAMENT_DECISIO_CORRUPCIO
POST_PARLAMENT_DECISIO_CANVI_CLIMATIC
POST_PARLAMENT_DECISIO_RELACIO
EDAT
IPV−PDeCAT
RPV−PDeCAT
ICV−PDeCAT
SATISFACCIO_ELECCIONS_14F2021
CAT_0_10
IPV−ERC
RPV−CUP
IPV−CUP
ICV−CUP
RCV−Junts per Catalunya
IPV−Junts per Catalunya
RPV−Junts per Catalunya
ICV−Junts per Catalunya
IPV−No podria votar

SEXE, CAT_ESP, IDEOL_1_7, REC_PARLAMENT_VOT

Here we can see that both of the categories of the variable **SEXE** seat next to the origin point, which means that the their contribution to the subspace is very little. It seems like that men lean more to the left and satisfaction with the previous election while women the opposite, but it's so slight it's barely significative.

As for the variable **CAT_ESP**, there's a clear separation between the people that feel they're Catalans in Catalonia and the other categories, especially the ones that declare themselves as Spanish. The political ideology generally goes from negative to positive in the first axis and slightly from positive to negative in the second axis as we go more to the left. This would suggest that left-leaning people are more satisfied with the "Parlament"'s decision, the February 14th election and, curiously, those that feel more catalan.

The political parties votes occupy spaces near their ideologies, except maybe "Junts per Catalunya" which would be occupying a left-leaning space, which is not necessarily adequate.

SIT_ECO_CAT, SIT_ECO_ESP, SIT_POL_CAT, SIT_ECO_PERSONAL

This plot suggests that people that think that the Spanish economic situation is good also think the Catalan one is very bad and it will get worse. Howver, they say their personal economy is neither good or bad. The categories of people that think that the situation (either political or economic) is bad, are very close to the origin point, which means that they're not distinctive or representative in this subspace.

People that think that their personal economic situation is very good or good are more likely to be satisfied with the February 14th election and believe the "Parlament" decisions were good.

PART_PARLAMENT, PART_CONGRES

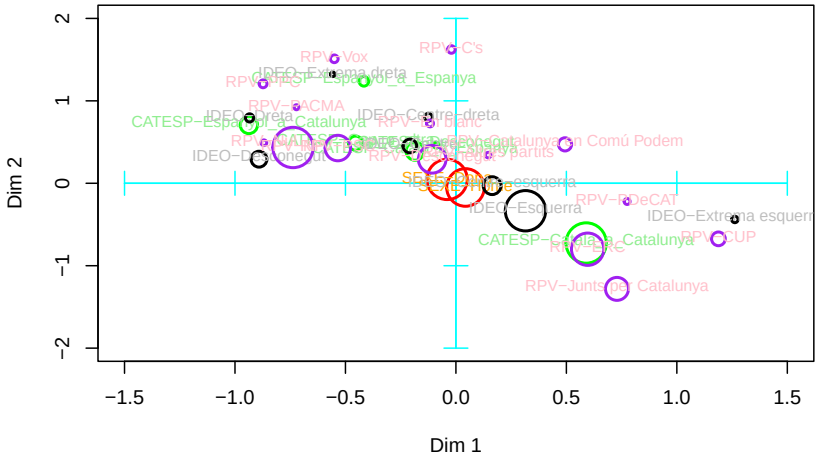
The participation in both elections ("Parlament" and "Congreso") is very associated. The categories of people that didn't vote, regardless of the reason, are very associated and it's more likely that they weren't satisfied with the February 14th election. The opposite happens for the ones that voted for sure, although they are larger categories and closer to the center, which means that it isn't as strongly correlated.

INT_PARLAMENT_VOT, INT_CONGRES_VOT, REC_PARLAMENT_VOT, REC_CONGRES_VOT

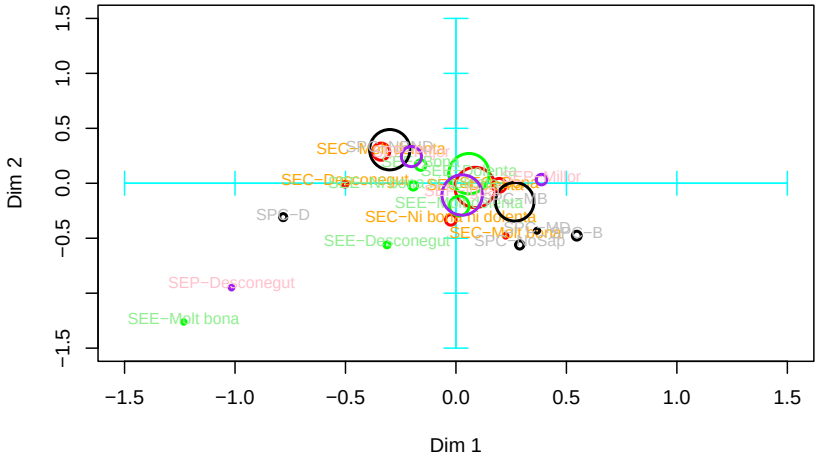
In this plot we can see, for the most part, exactly what we expected: each category representing the vote to a party in previous and present "Parlament" and "Congreso" elections are very closely related, with 3 curious facts:

- The profile of "C's"'s votes in previous "Congreso" elections is different from the rest of possibilities, much more close to "Vox", mainly, and the "PPC".
- The "PACMA" seems to be a party close to the right and very unsatisfied with the "Parlament" decision. However, this might be simply a result from a small sample size.
- The "PDeCAT" is in the space in which we'd expect it to be, but it's very spread. It could be, again, caused by the small sample size.

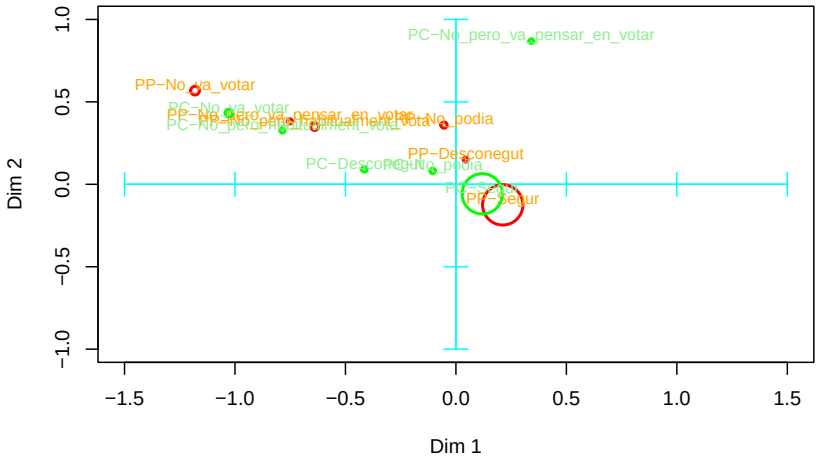
Projection of categories
(SEXE, CAT_ESP, IDEOL_1_7, REC_PARLAMENT_VOT)



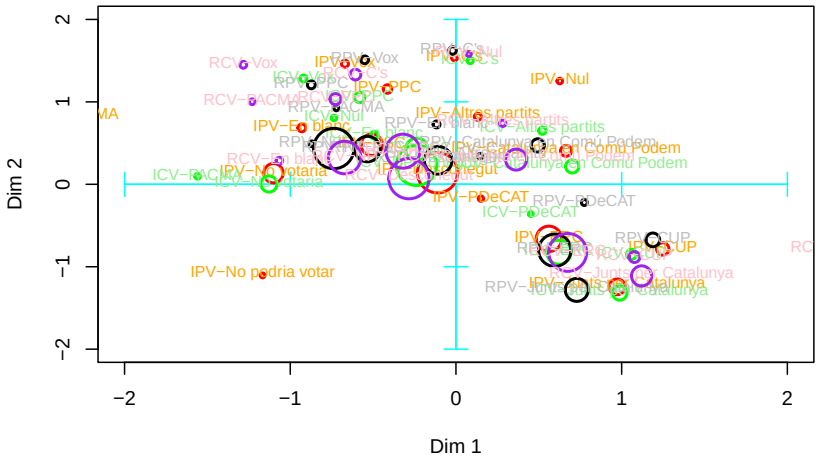
Projection of categories
(SIT_ECO_CAT, SIT_ECO_ESP, SIT_POL_CAT, SIT_ECO_PERSONAL)



Projection of categories
(PART_PARLAMENT, PART_CONGRES)



Projection of categories
'ARLAMENT_VOT, INT_CONGRES_VOT, REC_PARLAMENT_VOT, REC_CONG



Conclusions

Looking at all the variable groups analyzed in the previous section, we can interpret them all together and finding associations between them. This will be easier if we do it by quadrants, as it's an easy partition to do, which corresponds to direct or inverse relations of the categories to the axis or principal component.

In the upper left quadrant, we see a strong presence of spanish nationalism right-leaning parties votes in previous and future elections in both the "Parlament" and "Congreso". As expected, these are closely related to the people that feel "less catalan". Also, the ideologies present there are center or right-leaning. Lastly, there's a high association with people who didn't vote and a ligther one with people that think the spanish economic is good but the Catalan one is bad. They also say their personal economy has worsened and, surprisingly, the political situational in Catalonia is neither good or bad.

In the upper right quadrant there isn't much, besides the votes to "Catalunya en Comú Podem" and "Other parties", which seems to be logical, considering they are ambivalent regarding the independence of Catalonia.

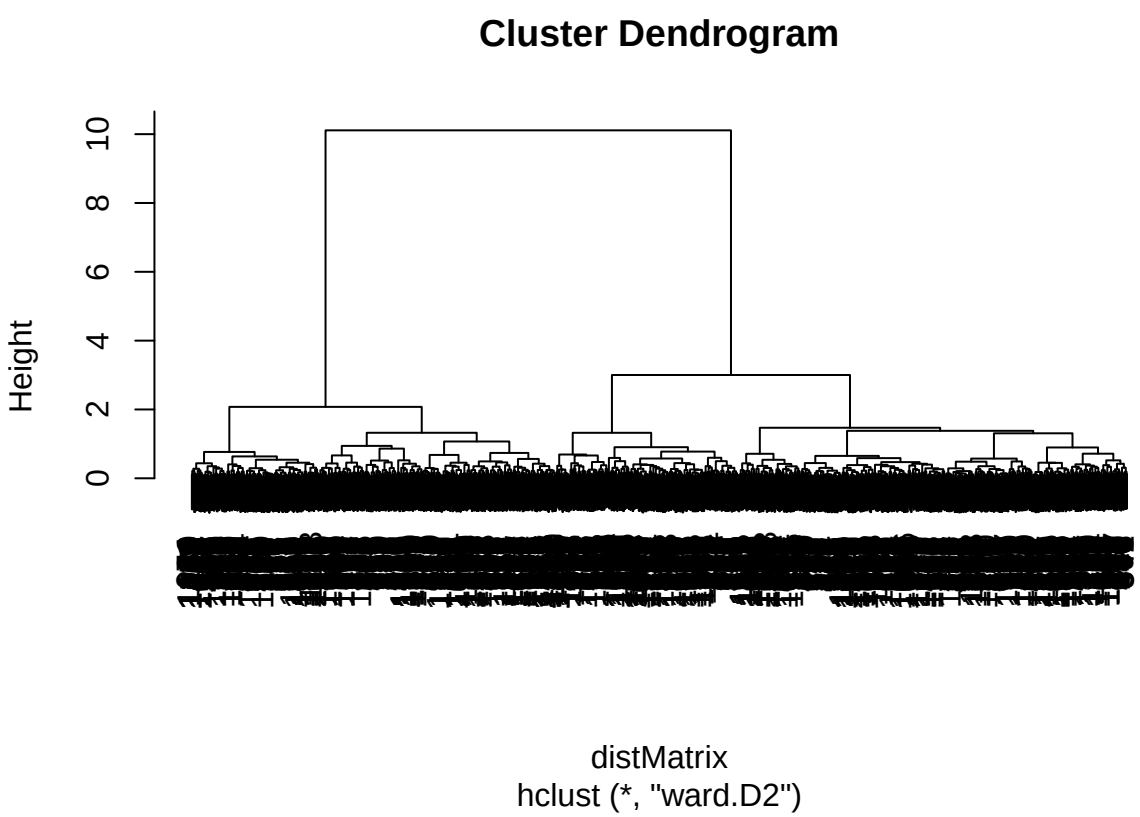
In the lower left side, we only see a couple of Unknown answers to economic situation question answers. Also of people who think that the spanish economic situation is very good, but that is probably due to the fact that only 2 people answered that way.

Lastly, in the lower right quadrant we see almost all the opposite answers from the upper left quadrant. There we see catalan nationalistic parties. There's also a left-leaning tendency, with the possible exception of "Junts per Catalunya". The answers about spanish and catalan economic situation are diametrically opposed.

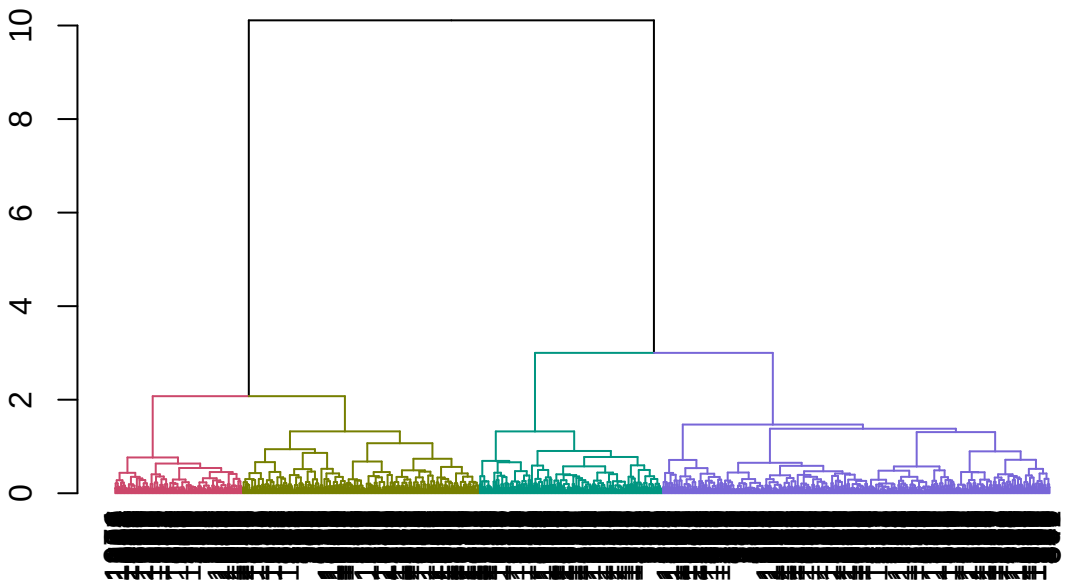
Clustering and profiling

Dendrogram

This is the dendrogram we obtained from our data, observing it we decided to cut our dendrogram at height 2, so we have 4 different clusters



Here we see the 4 different classes more clearly. We were also thinking of only doing 3 different classes because when we started with the profiling we saw that classes 1 and 4 were very similar, later we will explain why we finally decided to stay with four classes.



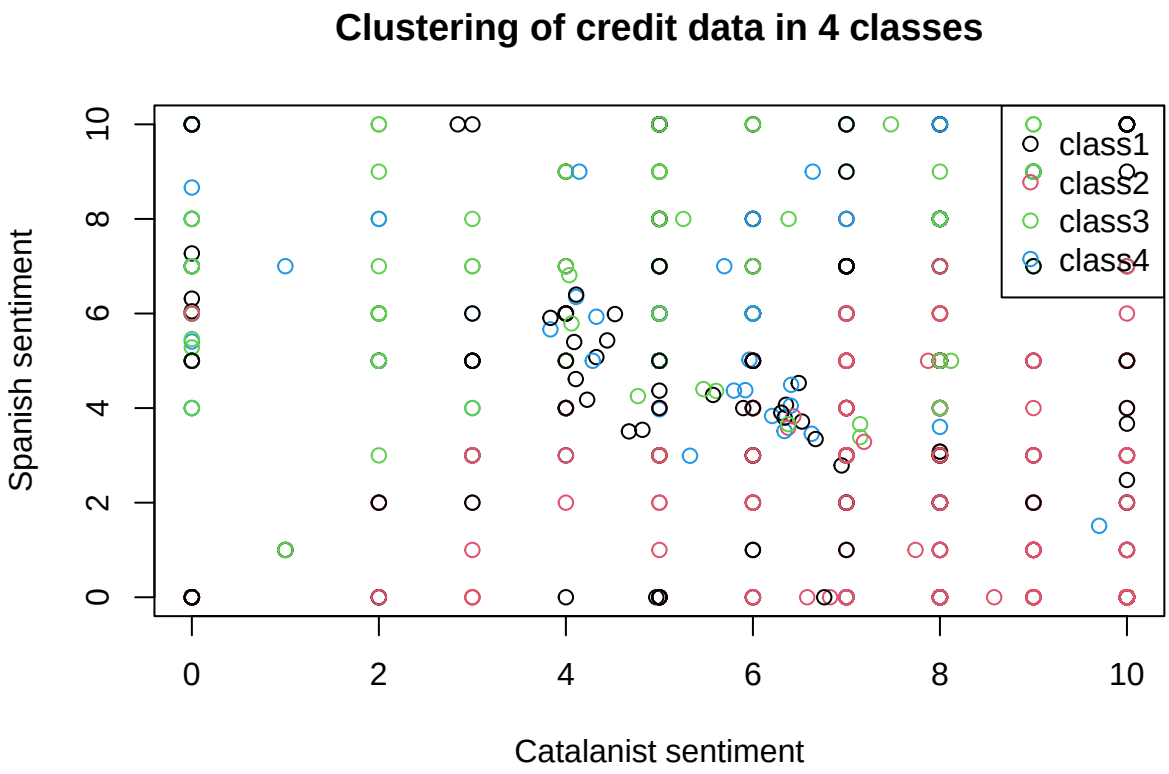
Here we see the number of persons in each of the classes

```
## c2
##   1   2   3   4
## 304 497 235 164
```

We can begin to identify the different classes in this scatter plot, where we can see class 3 in green is for a person with a Spanish sentiment. And the other way around for class 2 being more identified with a Catalan sentiment.

It's important to say that the answers given were Integers from 0 to 10, but we can also see decimals on this plot, that because those were missing values we imputed.

That brings us to classes 1 and 2 that as we can see are distributed all over the plot, the interesting thing is that we can see that these classes have a lot of missing values (later we will see that probably this is correlated to this classes being where the persons are more uncertain or don't care as much about politics).



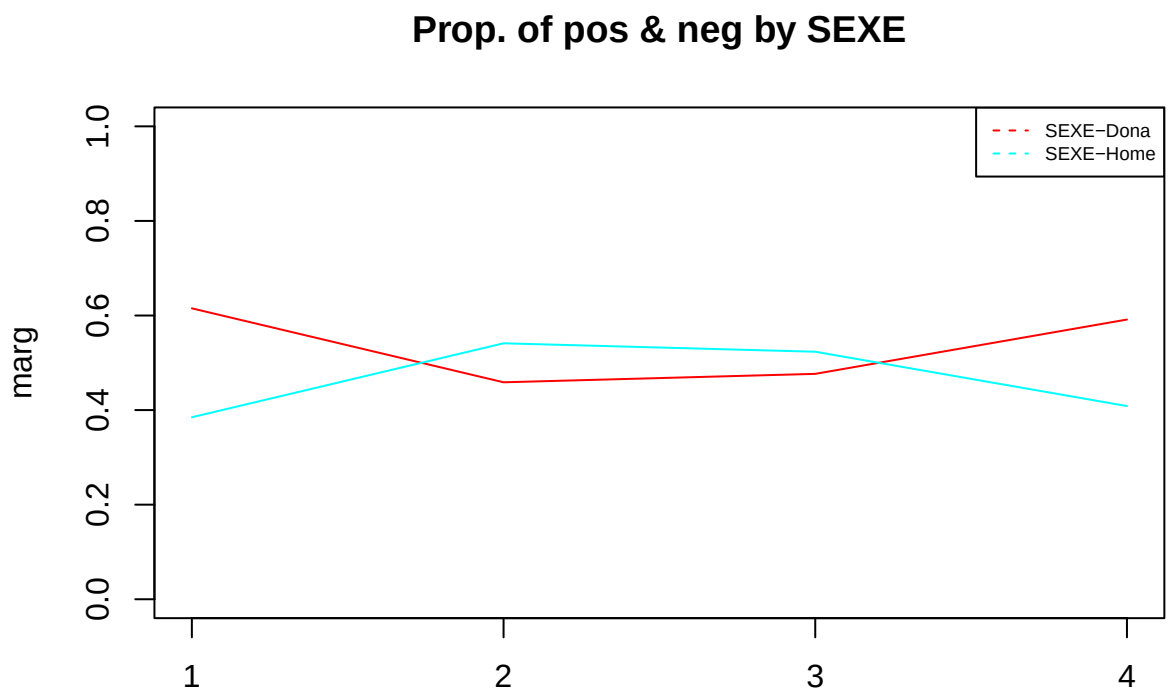
Profiling preparation

Once we figured out that we had 4 clusters in our dendrogram, we started the profiling section where we examine each selected variable and we obtain information about each cluster. Before starting, we defined 3 main questions that we wanted to know about each group: Ideology, electoral participation and opinion on the independence of Catalonia. Of course, we had a lot more questions that we wanted figure out, but we thought these 3 would be the ones that defined the groups the most. Of the multiple variables that we had in our database, we selected the following, which were the ones that, when looking at their graphs, were the most interesting to draw conclusions.

Dataprofiling of variables

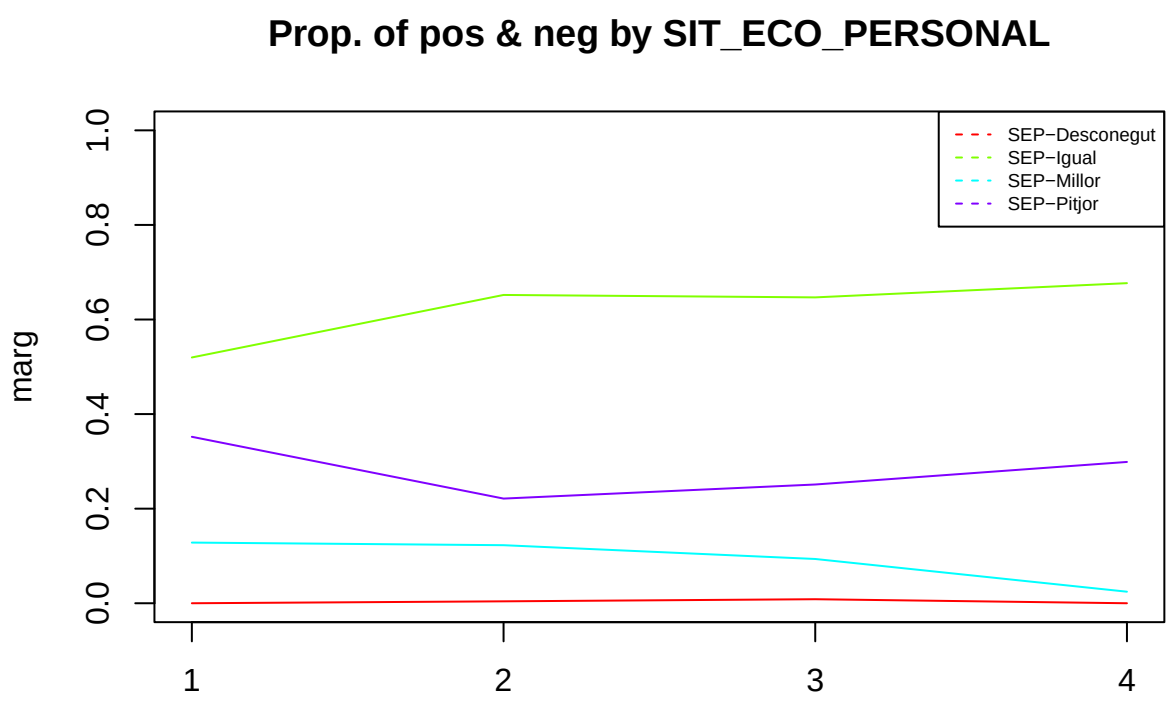
SEXE

First we have the categorical variable SEXE, and here we found out that men tend to be slightly more predominant in the clusters 2 and 3, while women prevail a bit more in the clusters 4 and 1. Later on we will see that groups 1 and 2 are a lot more participatory and caring about the issues questioned in the CEO's survey. However, groups 4 and 1 either don't vote, vote blank or their vote is unknown. We see difference in electoral participation, but we don't see a lot of differences in ideology or other sociopolitical issues between men an women. Even if they are participatory or not they all tend to be similar in their ideological groups.



SIT_ECO_PERSONAL

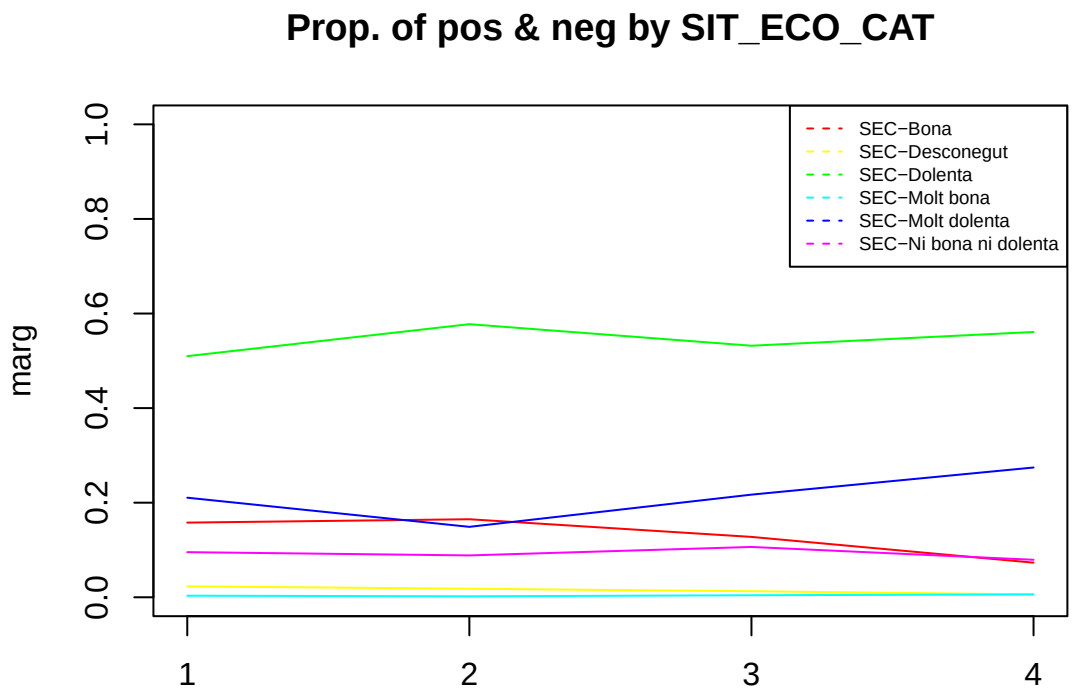
We have defined the surveyed by sex, now we wanted to see their economical situation. We have 5 questions related to that. Some define their opinion on the economical situation in Catalonia (currently, prospectively and retrospectively) and others in Spain. But the ones that we were more curious about were the ones that talked about their personal economical position, specifically their current one and if they think it has improved or not since the last election. What we saw with these graphs is that the tendency of all groups, except the second one, is that the economical situation is the same or it has gotten worse. Cluster 1 tends to have a little a bit more of the option that says it is worse, although the option that says it is the same also prevails just like the other clusters. We will see in other plots that the second group contains people who hardly vote and who care less about the questions in the survey.



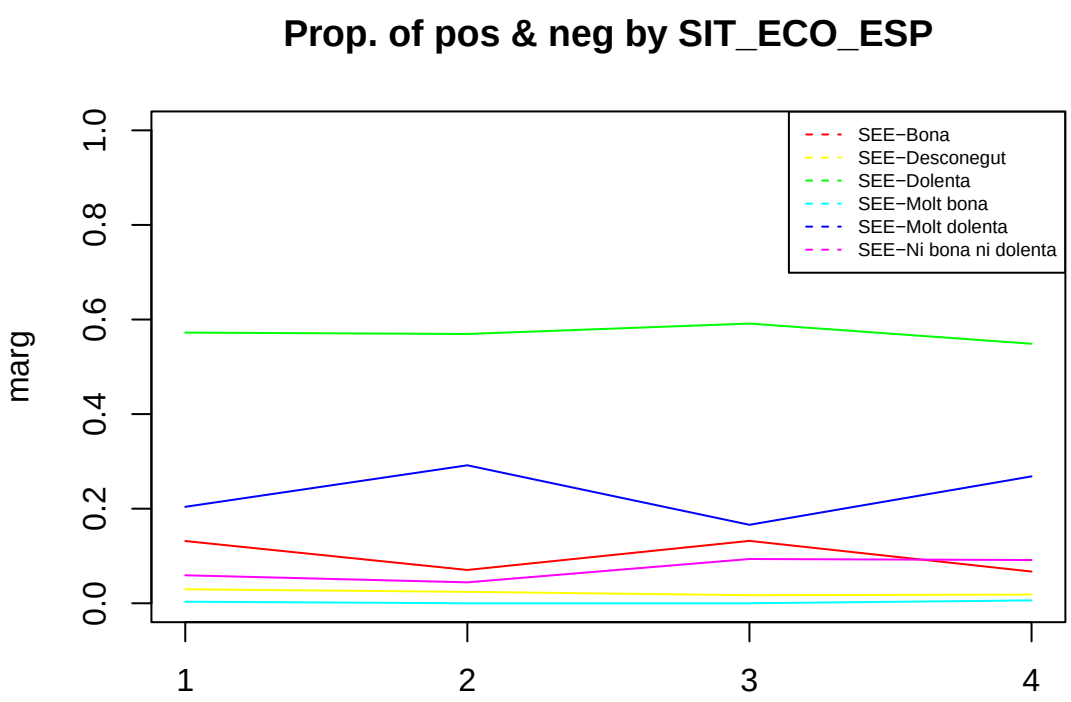
SIT_ECO_CAT and SIT_ECO_ESP

Now we are going to see what do the surveyed think about the economical situation in Catalonia and in Spain. Here we are starting to see some ideological and identity (Catalan or Spanish) differences between groups. The possible responses are: good, bad, very good, very bad, neither good nor bad or unknown. On both of the plots the tendecy in every cluster is that the economical situation is bad. Although we start to see differences in the extreme responses in the 2 plots: on the Catalan economical situation plot we see that cluster 2 has less "very bad" responses and more "good" responses than the other clusters, and cluster 3 has a lot more "very bad" and less "good" responses than the other groups. And if we look at the Spain's economical situation plot the options seem to reverse (cluster 2 has more "very bad" and less "good" response, and vice versa with cluster 3). Here we can clearly see the polarization between the 2 most participatory groups (2 and 3). We see that cluster 2 tends to have a more positive response regarding Catalonia's situation and a more negative one regarding Spain's, while cluster 3 is the other way around. Cluster 4 and 1 have a more neutral or unknown response about this issue, so we count them as less participatory.

Catalonia's current economical situation:

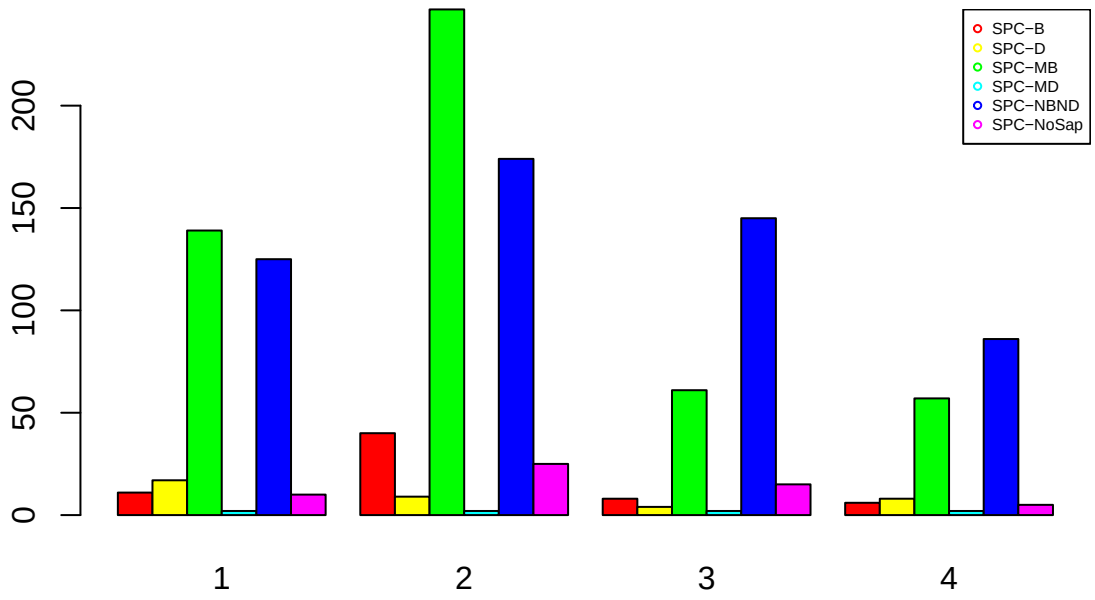


Spain's current economical situation:



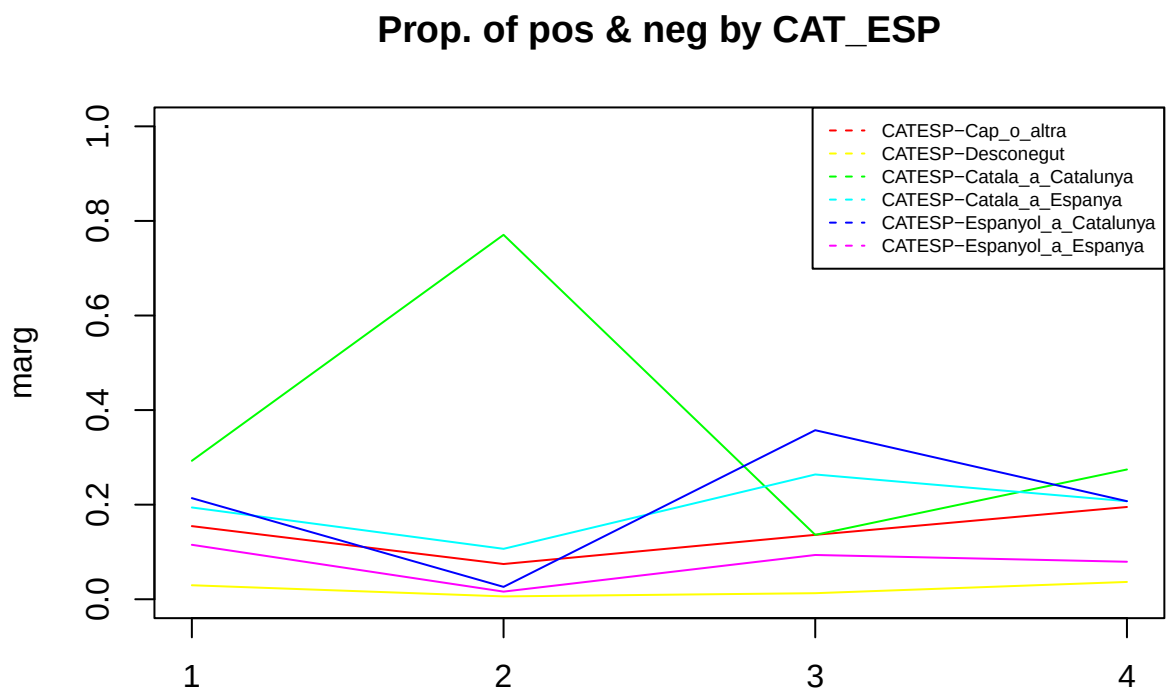
SIT_POL_CAT

Now we wanted to see what do they think about the current political situation in Catalonia, since the surveyed are all Catalan residents. We are continuing to see the same polarizing tendency between cluster 2 and 3, but we have some surprises with cluster 1. Cluster 2, the most favorable to Catalonia, has the least "very bad" and the most "very good" responses than the other groups and vice versa with cluster 3. Again, we see the main differences with the more extreme options, either good or bad. Cluster 1, this time is the one that has the most "bad" responses. It surprised us to see that the least participatory cluster had such a bad opinion on Catalan politics. This could be interpreted in many ways, like for example "voter frustration", although this conclusion is out of our data scope.



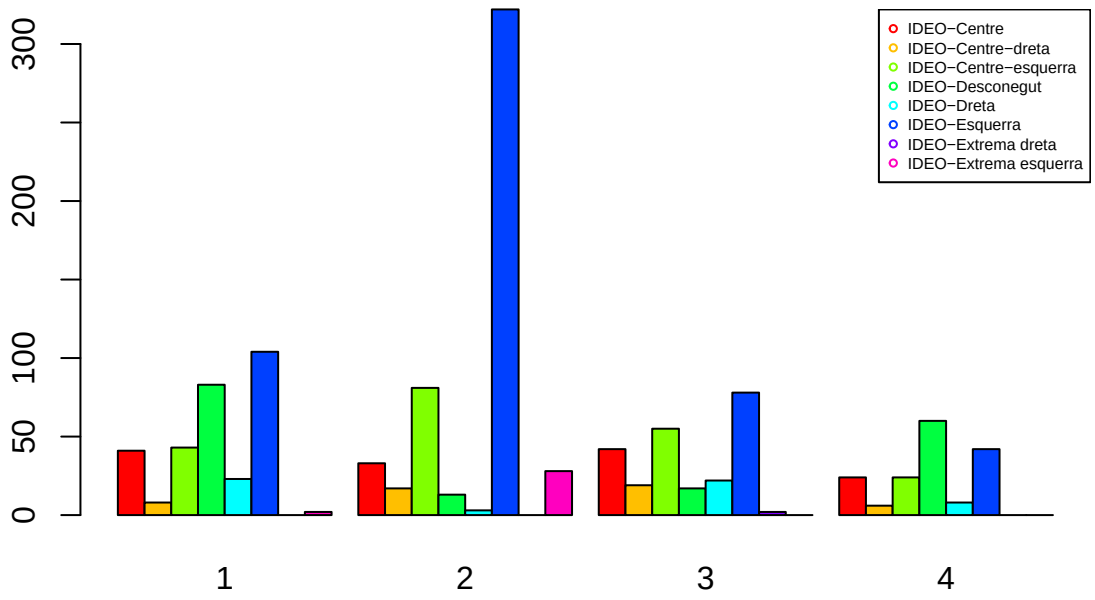
CAT_ESP

In this next categorical variable we are going to see the national identity of the surveyed. It is asked if they feel a Catalan residing in Catalonia, a Catalan residing in Spain, a Spanish residing in Catalonia, a Spanish residing in Spain, neither or another identity or unknown. In the following plot, we clearly see a group that really stands out, which is cluster 2. It has a lot of "Catalan residing in Catalonia" responses, way more than the other clusters, and the least "Catalan residing in Spain" responses, although it is not the one with less "Spanish residing in Catalonia" (that would be group 3). Cluster 3 is the other way around with those responses. Cluster 4 and 1 have a lot of "neither or another identity" or "unknown" (second group) responses. Here we still can't say group 2 only prevails a more pro-Catalan identity (maybe pro-independence) group, because it has quite a lot "Catalan residing in Spain" responses. This is going to be more refined later on with the voting party question.



IDEOL_1_7

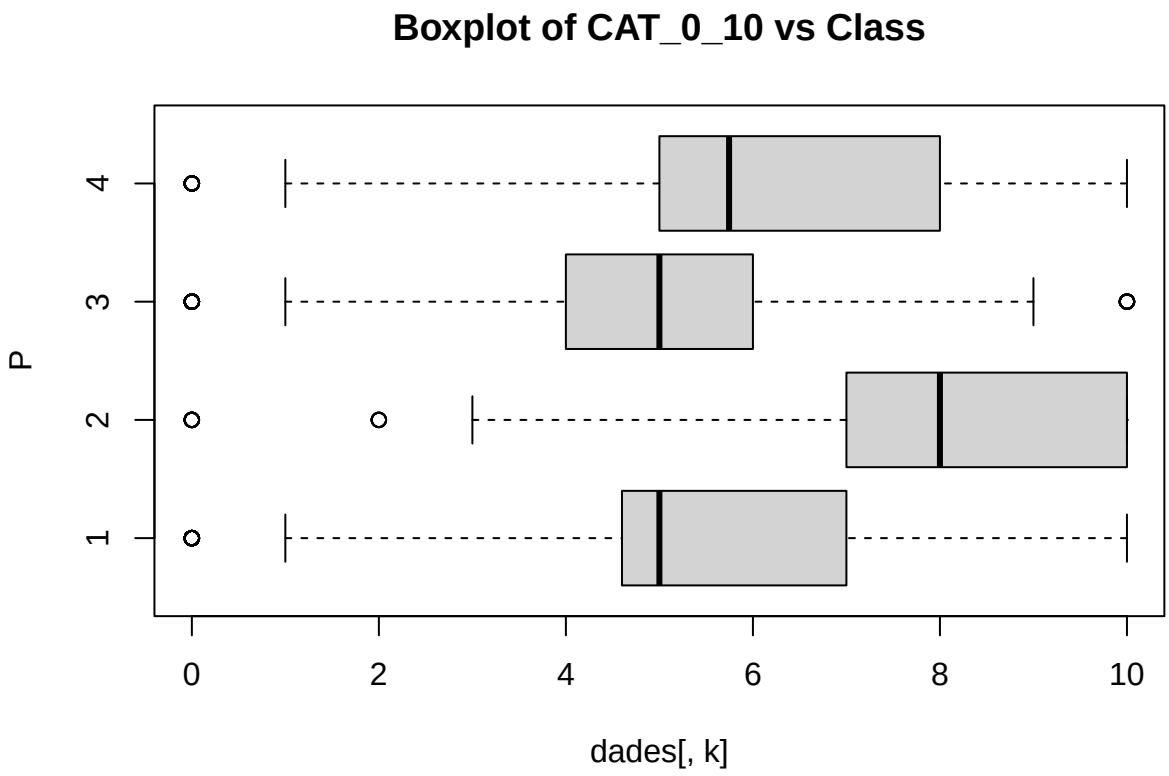
In this next question we see the ideology of the surveyed. This is defined between the following options: center, center-right, center-left, right, left, far-right, far-left or unknown. In the data that we gathered for this questions we clearly see an overall leftist ideology, although some groups more than the others. The one that stands out as the most leftist is cluster 2, with a far more left, center-left and far-left ideology, and the least right and far-right than the other groups. Cluster 3, although it has a majority left, is the one, along with cluster 1 in some instances, with the most center and center-right ideology and also it has quite a lot right-wing ideology. Cluster 4 and 1 on the other hand have lots of "unknown" options and their ideology tends to be less extremist than 2 and 3. The second group, although, is slightly more right-wing than the first one (more center-right and right).



At this point, we see that in every question cluster 4 and 1 are very similar at every issue. We considered unifying both clusters in one, however, we decided not to do it after seeing some variables that we are going to see next. This variables represent plots regarding voter participation, what did they vote in the past elections or intend to vote in the following ones.

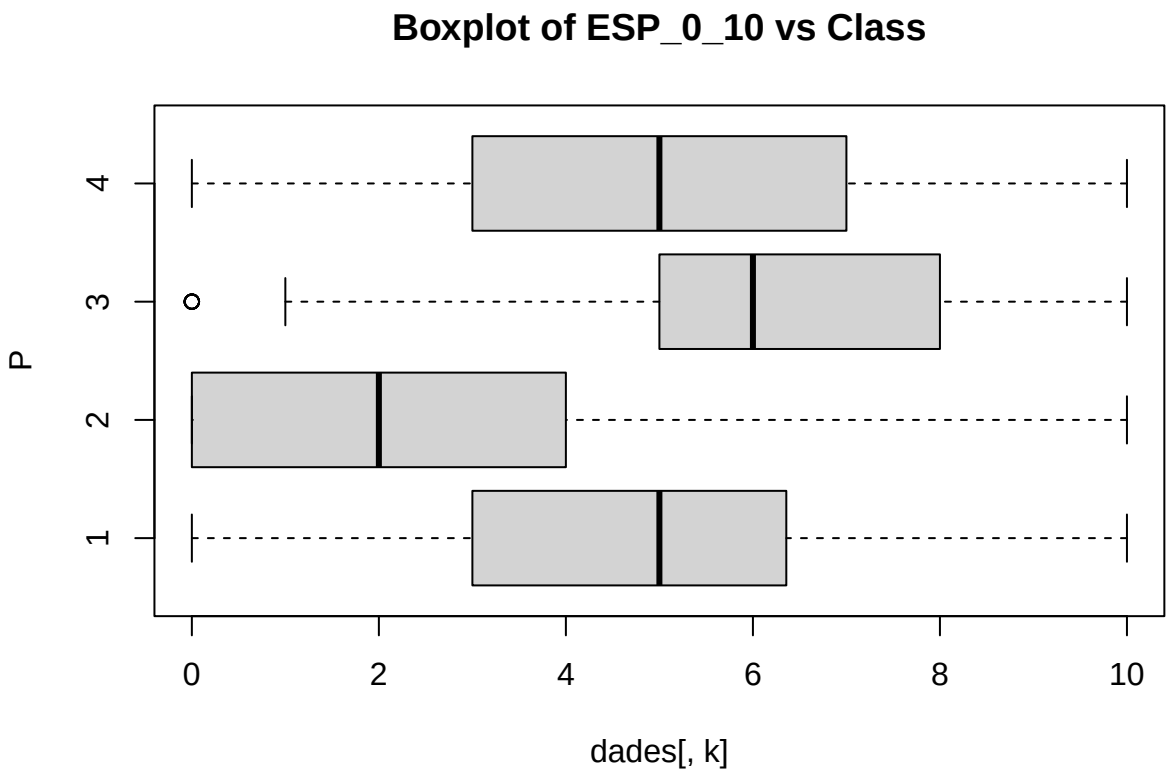
CAT_0_10

The most interesting thing we can extract from this boxplot is how much class 2 is on the right in respect to the others indicating the tendency we already observed of it being the class with a stronger Catalan sentiment.



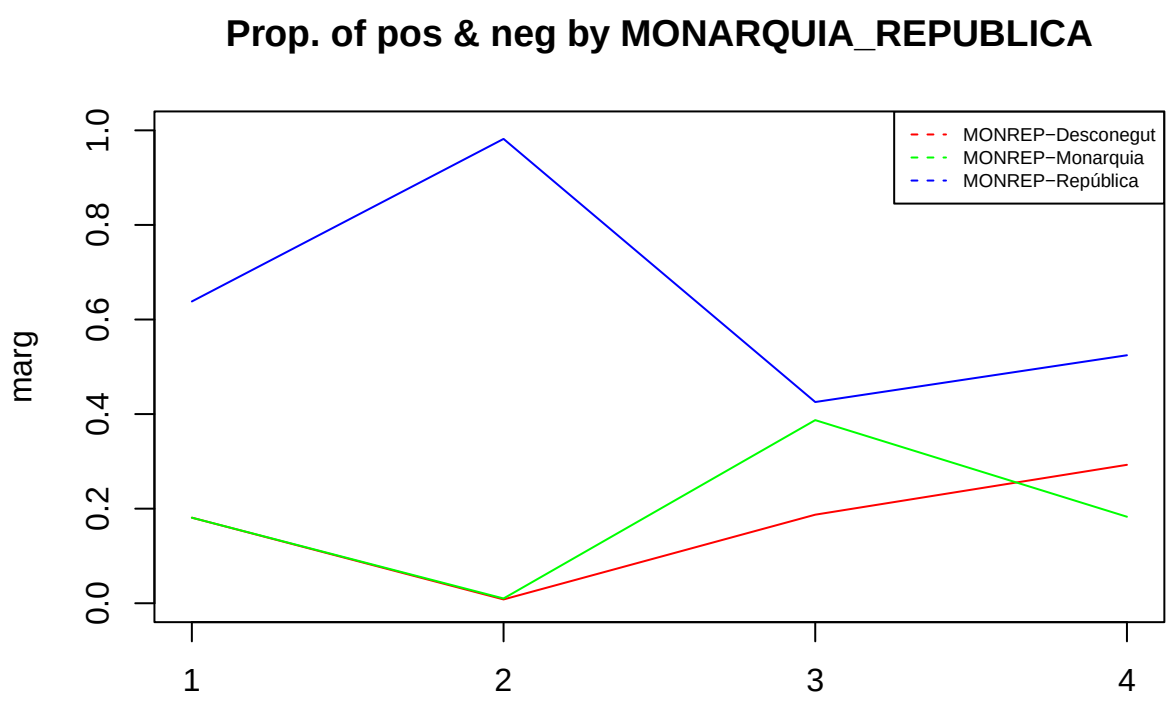
ESP_0_10

Here we can see the same but the other way around, being class 3 the one with a stronger Spanish sentiment. What we found interesting about these two plots is that class 3 is more on the Spanish side it stays around the middle when asked about Catalan sentiment. But with class 2 this doesn't happen instead when asked about Spanish sentiment the answers given were very low, this may be because people from class 2 are more extremist, or have a stronger sentiment than those from class 3.



MONARQUIA_REPUBLICA

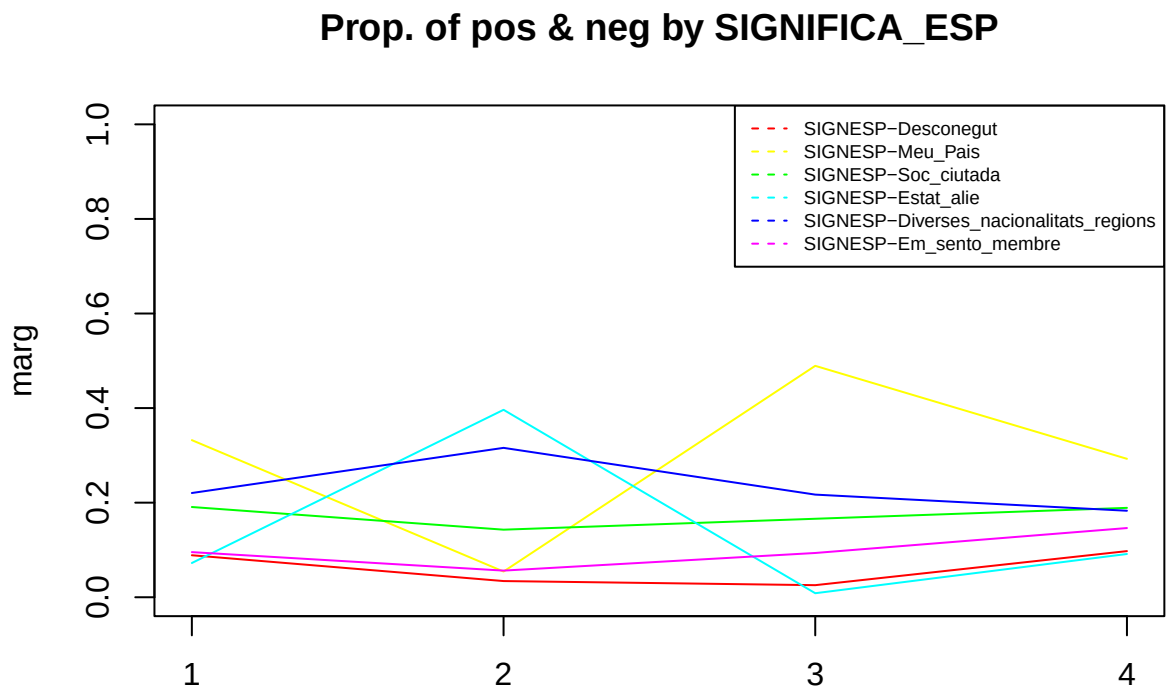
Here we see that class 3 is the one that wants a republic the most, and class 4 being the one that wants a monarchy the most, this is to be expected, seeing the results we've got until now, being class 3 on the left side of the political spectrum and class 4 on the right side.



SIGNIFICA_ESP

In this plot we see the sentiment of the people of different classes. Both class 1 and 3 are the ones that feel stronger about Spain being their country. Were class 2 considers Spain is an alien state. This alings with what we've seen before.

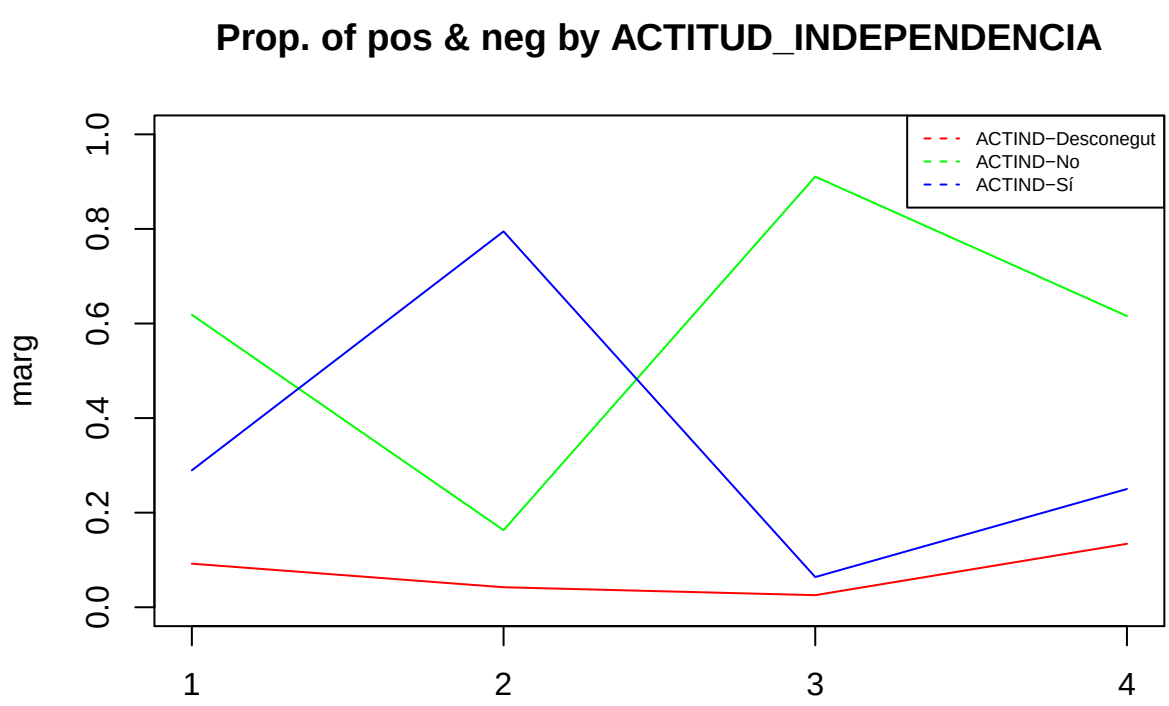
```
## [1] "Categories="
## [2] "SIGNESP-Desconegut"
## [3] "SIGNESP-Meu_Pais"
## [4] "SIGNESP-Soc_ciutada"
## [5] "SIGNESP-Estat_alie"
## [6] "SIGNESP-Diverses_nacionalitats_regions"
## [7] "SIGNESP-Em_sento_membre"
```



ACTITUD_INDEPENDENCIA

In this plot we see the what we should expect, class 2 is the one with most people that want independence, and class 3 the one with most people that don't want it. What's interesting is that in classes 1 an 4 (which usually don't have as much opinion) the predominant attitude is non-independence.

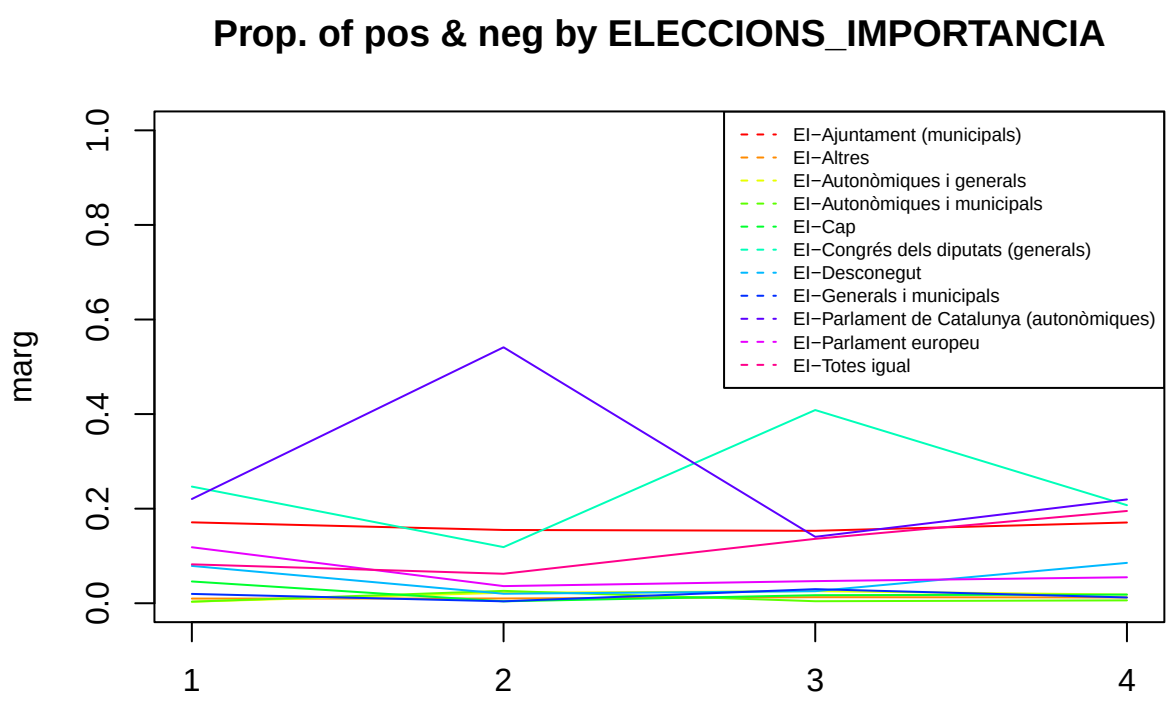
```
## [1] "Categories="       "ACTIND-Desconegut" "ACTIND-No"
## [4] "ACTIND-Sí"
```



ELECCIONS_IMPORTANCIA

Again what's interesting on this plot is that class 2 feels Parliament of Catalonia elections are more important. And the other way around for class 3 where they feel general elections are more important.

```
##  [1] "Categories="
##  [2] "EI-Ajuntament (municipals)"
##  [3] "EI-Altres"
##  [4] "EI-Autonòmiques i generals"
##  [5] "EI-Autonòmiques i municipals"
##  [6] "EI-Cap"
##  [7] "EI-Congrés dels diputats (generals)"
##  [8] "EI-Desconegut"
##  [9] "EI-Generals i municipals"
## [10] "EI-Parlament de Catalunya (autonòmiques)"
## [11] "EI-Parlament europeu"
## [12] "EI-Totes igual"
```

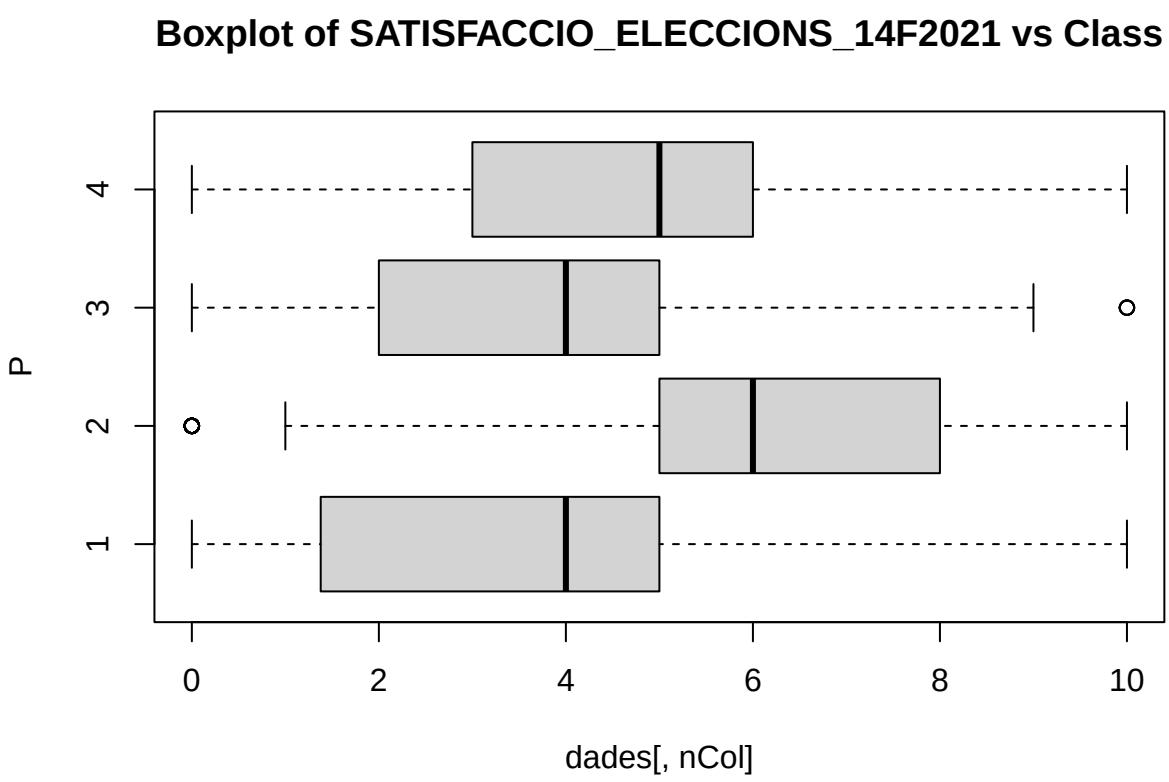


SATISFACCIO_ELECCIONS_14F2021

Here we can see how the cluster 2 is satisfied with the results of the last elections. This is because in this elections win the majority independent parties , and this cluster is formed for independent people.

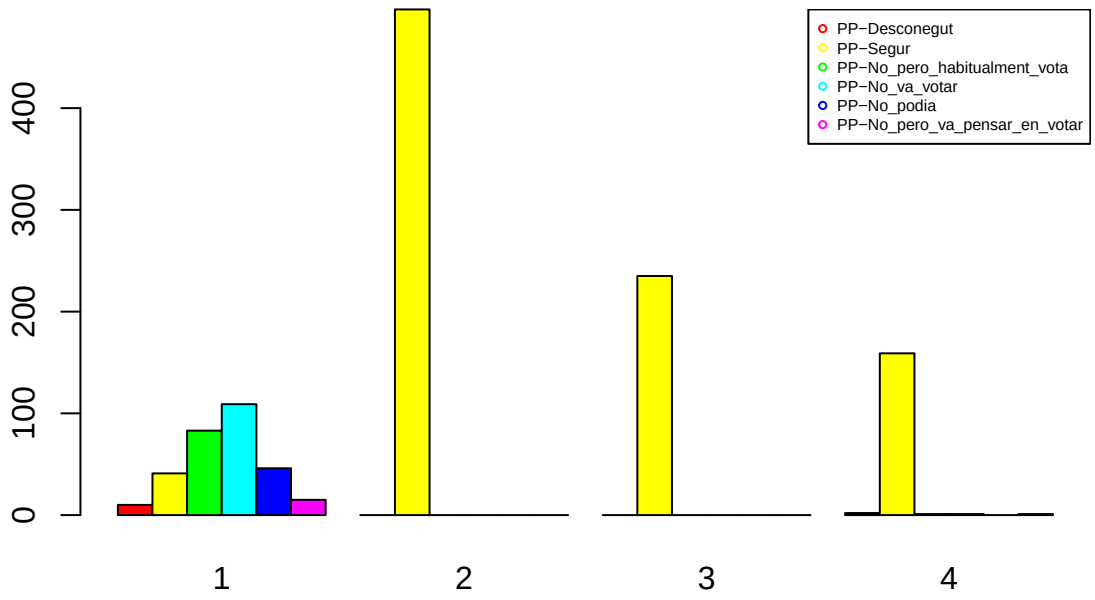
The cluster 4 is a little bit satisfied because they don't vote but if they have to, they will vote left and independent parties.

The other two clusters they are a little bit unsatisfied because they are from center, right and center-left ideology.



PART_PARLAMENT

Here we can see the difference between cluster 1 and cluster 4. Both clusters don't care about politics but the cluster 1 goes vote and the cluster 2 don't go to vote (for different reasons).



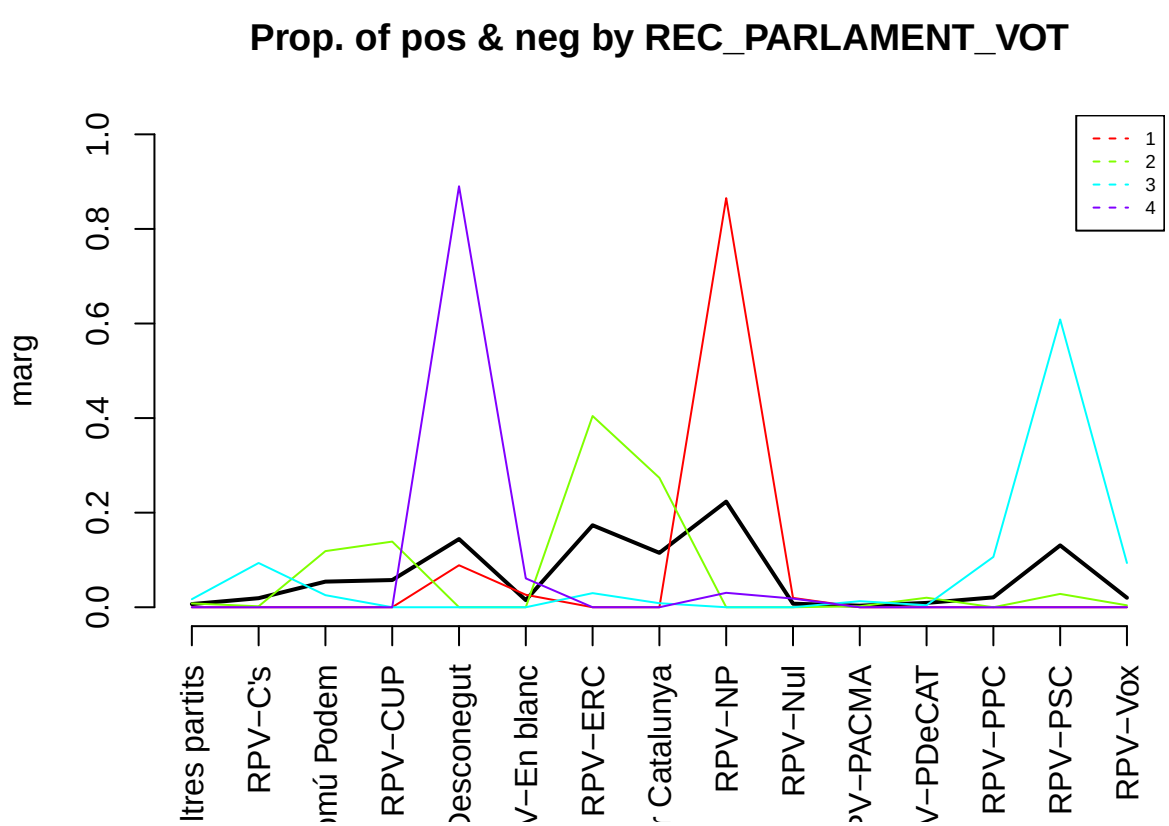
REC_PARLAMENT_VOT

In this plot we can see, and confirm some conclusions. The cluster 2 vote independent and left parties.

In the cluster 3 we see the voters of right, center and left-center parties (non-independent parties).

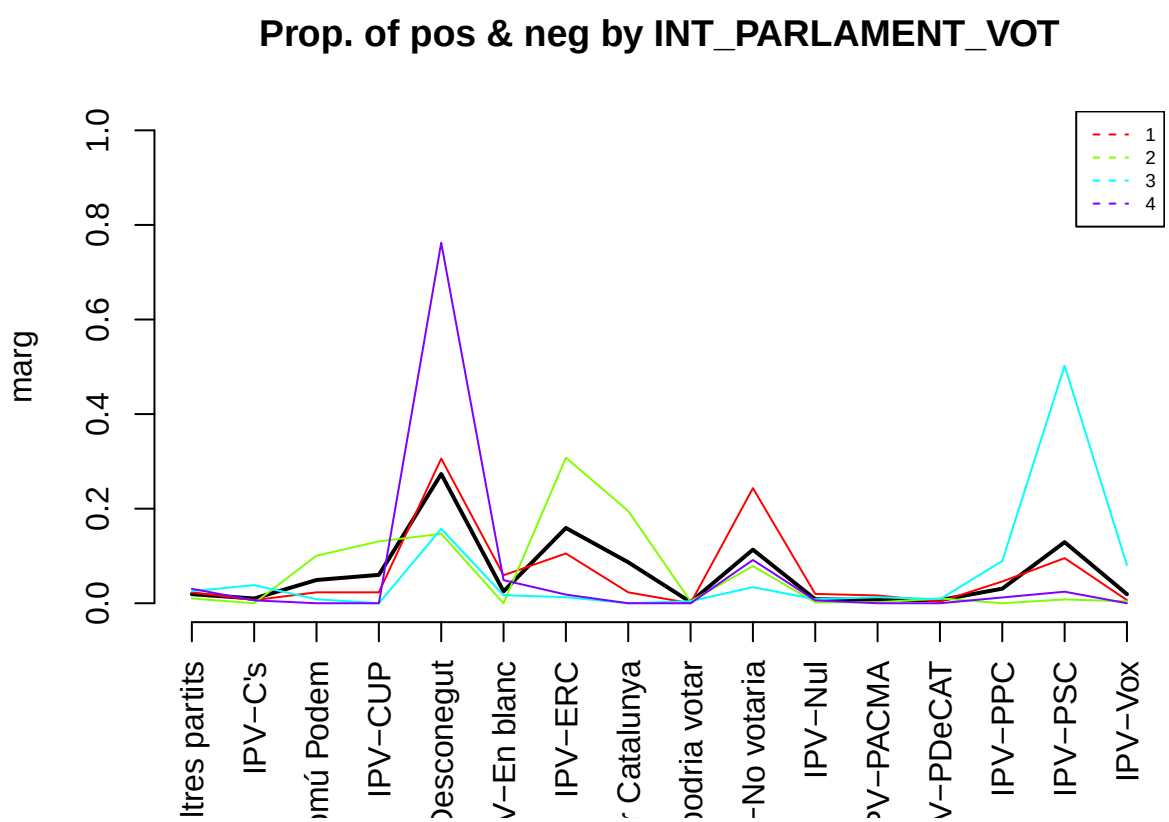
In the cluster 1 we see the people that don't go to vote.

In the cluster 4 we see the people who vote in blank and the people that don't want to say their vote. Maybe because they vote in blank, or they are extremists or they don't what to say it.



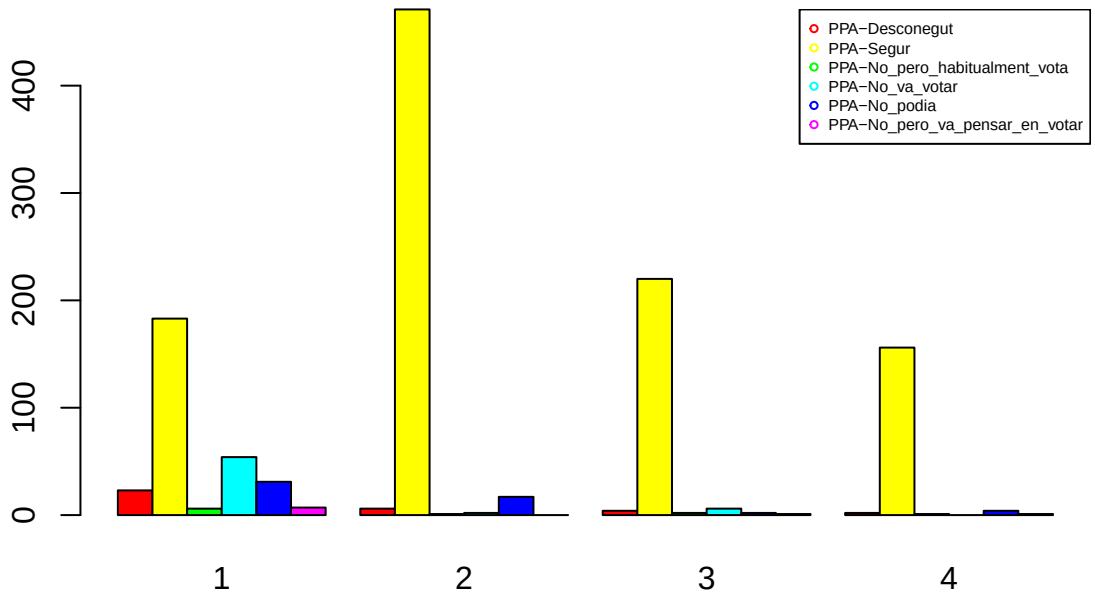
INT_PARLAMENT_VOT

In this plot we see that the intention of the vote is quite similar with no many differences. Here we can define the cluster 1 that they will not vote in the next elections but also if they have to vote, they will vote more left parties (people that don't vote maybe have mostly left ideology).



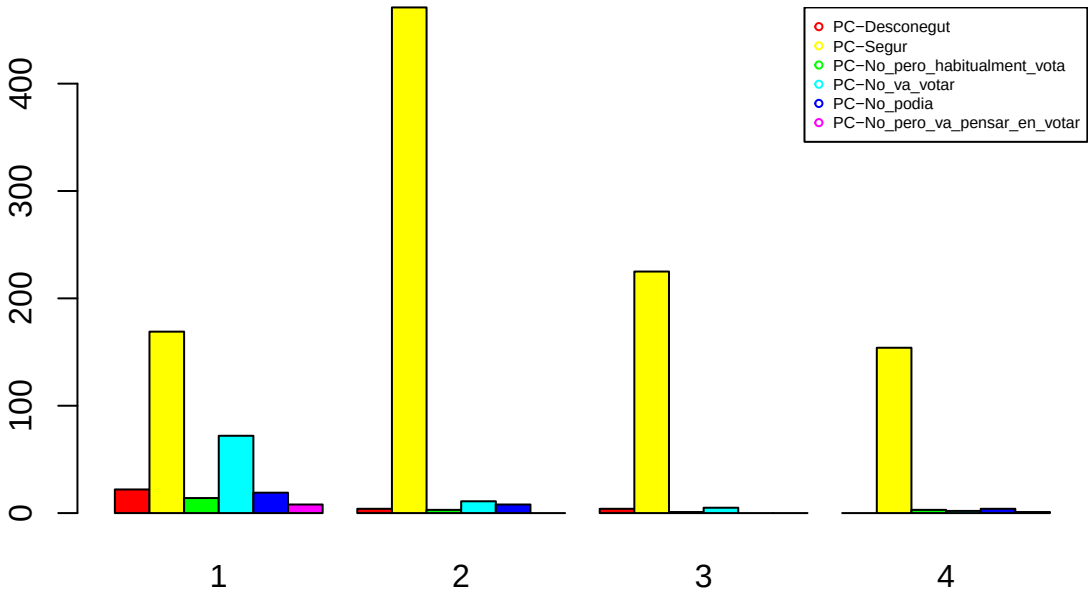
PART_PARLAMENT_ANTERIOR

Here we can see the majority of people vote in the last autonomic elections. Even the cluster 1, that in 14F elections didn't vote.



PART_CONGRES

Here we can see the majority of people participate in the national elections. Cluster 1 is the one where don't vote or don't want to say if they vote.


PC−Desconegut
PC−Segur
PC−No_pero_habitualment_vota
PC−No_va_votar
PC−No_podia
PC−No_pero_va_pensar_en_votar
0
100
200
300
400
1
2
3
4

PCA & Clustering, coincidences and divergences

Through our analysis of political context data, we can conclude after doing the clustering and PCA processes that there are lots of coincidences and some divergences between them.

We see on both procedures that there are 2 clear ideological tendencies, and these ones tend to be associated with certain ideas. On the PCA plots, for example, we see on the upper left quadrant a more right-wing ideological tendency and this is associated with a high Spanish identity and less Catalan. We can get to the same conclusion with the graphs showing the data per cluster. Cluster number 4 gathers a group of people who have a stronger Spanish identity, while the parties they vote tend to have a more right-wing ideology and are way less favorable to the independence of Catalonia.

On the opposite quadrant to the more right wing tendency we have the parties with a more left-leaning ideology. They are also a lot more favorable to the independence of Catalonia, so that is why Junts per Catalunya (a center-right independentist party) is also included in this region of the graph. This quadrant's tendencies can be compared with cluster 3, which gathers surveyed people that answered with a more left-wing ideology, and their identity is way more Catalan than Spanish.

Some divergences that we found is that on the upper-right quadrant, where there is a low participation in the elections and the predominant parties are En Comú Podem and Other parties. On the other hand, in the graphs by clusters, parties in the En Comú Podem candidacy and others are gathered in the third cluster along with the left-leaning parties, while people who didn't vote or have little care of the issues answered in the survey are grouped in the second cluster.

To sum it up, most of the conclusions that we draw during the cluster analysis are similar to the ones we found in the PCA, although in this last process we saw things that are represented better than what we saw during the profiling.

Profiling conclusions

By looking at our dendrogram, we determined that we have 4 clusters. Two first clusters are formed by people who in the last elections didn't vote for any political party (Cluster 1 and 4). Also we can say that they are formed by people who don't care about politics or they don't want to talk about their political preferences.

Cluster 2 and cluster 3 are similar because they have political participation but their ideology is quite different.

Cluster 2 is defined for voting left and independent political parties. They are satisfied with the results of the last elections because there was an independent majority. They defend the republic and they have a catalan nationalist feeling.

Cluster 3 is defined for voting center and left political parties (also includes the votants of PSC). They are satisfied with monarchy, they don't want independence and they have more Spanish feelings than Catalan.

Working Plan

During the making of this project we had some difficulties and obstacles that have arisen. We had problems that we didn't consider at the beginning. One of them being that some commands on the script didn't work on our data, so we had to find alternatives along with our professor. We could avoid a lot of scheduling risk by assigning 2 or 3 people on a task so we could help each other and finish the task effectively. Also, we had a lot of questions regarding the functionality of RStudio and the data mining methods that we were not familiar with, so we asked our professor, via email or in-person, and we could get on with the task.

Membres:	Alex	Andreu	Lluc	Raúl	Guillermo	Joan	Entrega
Task 1	x	x					
Task 2	x						
Task 3						x	
Task 4			x				
Task 5	x						
Task 6		x					
Task 7				x			
Task 8			x			x	
Task 9					x		
Task 10				x	x		
Task 10.5		x					
Task 11	x	x	x	x	x	x	
Task 12		x					
Task 13				x	x		
Task 14							
Task 15	x				x		
Task 16		x	x			x	
Task 17		x	x			x	
Task 18	x	x	x	x	x	x	
Task 19				x	x		
Task 20	x						
Task 21	x						

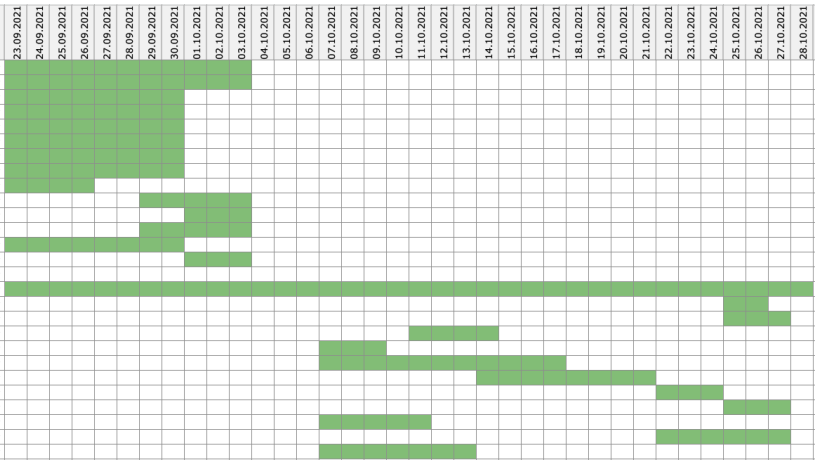
Task description:

1. Recode factors: shorten names and include an abbreviation of the variable name in the factor name. Remember to make tables that specify recoding.
2. Recognize numerical variables as such. A script is required.
3. Investigate variables if blanks are NA or have been coded as NP (Not Asked).
4. In REC_PARLAMENT_VOT_ANTERIORS there's people that respond "Other Parties" ("Altres Partits") and put the equivalent in a national or municipal in variable REC_PARLAMENT_VOT_ANTERIORS_LITE in a party that was as an option in REC_PARLAMENT_VOT_ANTERIORS.
5. The same with REC_CONGRES_VOT i REC_CONGRES_VOT_LITERALS.
6. Look for more.
7. Determine whether blank binary variables are NP ("Not Asked") or NA (Not Applicable) depending on the variable.
8. Do Gantt chart, assignment grid and risk plan.
9. Basic initial univariate descriptive statistics of raw variables:
10. Explain the pre-processes done (preferably, do them and explain them more or less at the same time) and the decisions taken.
11. Finish the Basic initial univariate descriptive statistics for new or modified variables.
12. Metadata file that describes the selection of the variables considered for the analysis.
13. Working plan.

14. Make the document structure of the deliverable.
15. Write non-technical sections and justify
16. Formal description of Data structure and metadata
17. Complete Data Mining process performed.
18. PCA analysis for numerical variables.
19. Hierarchical Clustering on original data.
20. Profiling of clusters.
21. Global discussion and general conclusions of the whole work. Analyze coincidences and divergences between ACP, AMC, Clustering.
22. Fix univariate descriptive analysis
23. Finish knn preprocessing (or find another formula)
24. Adapt party votes to national candidacy (e.g. REC_PARLAMENT_VOT → RPV-Barcelona en Comú should be En comú podem)

Gantt Chart

Nombre de la tarea	Fecha de inicio	Fecha de finalización	Asignado	Estado
D3	23.09.2021	03.10.2021	Everyone	Acabat
Tasca 1	23.09.2021	03.10.2021	Andreu/Àlex	Acabat
Tasca 2	23.09.2021	30.09.2021	Àlex	Acabat
Tasca 3	23.09.2021	30.09.2021	Joan	Acabat
Tasca 4	23.09.2021	30.09.2021	Lluc	Acabat
Subtasca 4.1	23.09.2021	30.09.2021	Lluc	Acabat
Subtarea 4.2	23.09.2021	30.09.2021	Lluc	Acabat
Tasca 5	23.09.2021	30.09.2021	Àlex	Acabat
Tasca 6	23.09.2021	26.09.2021	Andreu	Acabat
Tasca 7	29.09.2021	03.10.2021	Raúl	Acabat
Tasca 8	01.10.2021	03.10.2021	Joan/Lluc	Acabat
Tasca 9	29.09.2021	03.10.2021	Guillermo	Acabat
Tasca 10	23.09.2021	30.09.2021	Guillermo/Raúl	Acabat
Tasca 10.5	01.10.2021	03.10.2021	Andreu	Acabat
Entrega final	23.09.2021	28.10.2021	Everyone	Acabat
Tasca 11	25.10.2021	26.01.2021	Everyone	Acabat
Tasca 12	25.10.2021	27.10.2021	Andreu	Acabat
Tasca 13	11.10.2021	14.10.2021	Guillermo/Raúl	Acabat
Tasca 14	07.10.2021	09.10.2021	Andreu	Acabat
Tasca 15	07.10.2021	17.01.2020	Guillermo/Àlex	Acabat
Tasca 16	14.10.2021	21.01.2020	Andreu/Joan/Lluc	Acabat
Tasca 17	22.10.2021	24.10.2021	Andreu/Joan/Lluc	Acabat
Tasca 18	25.09.2021	27.01.2020	Everyone	Acabat
Tasca 19	07.10.2021	11.01.2020	Guillermo/Raúl	Acabat
Tasca 20	22.10.2021	27.10.2021	Àlex	Acabat
Tasca 21	07.10.2021	13.10.2021	Àlex	Acabat



Risk assessment

Risk	How to prevent	How to manage
Membre del grup marxa	S'assigna a 2 persones per cada tasca almenys	Es reassigna tasca a altres membres
Ningú sap com avançar una tasca	Consultar a la classe de lab anterior	Enviar correu a Xavier
Torna el confinament	Estar sempre en contacte	Treballar per videotrucada
S'apropa la data d'examens i s'ha de fer una entrega	Tenir una setmana abans feta l'entrega	Assignar feina a company amb menys assignatures
Membre del grup no ve a una sessió de laboratori	Estar sempre en contacte	Comunicar el que s'ha realitzat i el que se li ha assignat