

# European Union: Membership Vote

*Guillem\_Amat*

*December 8th, 2019*

## Summary

This report investigates the association different socio-demographic and related variables have on the decision to vote whether to leave or remain in the European Union among individuals living across the main EU member states. It explores what the main variables are when predicting a remain vote and whether there are differences at the member state level. A Logistic Multilevel Model, where member states were allowed to have random intercepts, was fit to the data to answer these questions. *Age*, *Education Level*, *Total Household Income*, *Political Interest* and *European and Country Attachment* were found to be the most significant predictors while *Marital Status*, *Gender* and *Minority Membership* were found to be less important. EU Member States were found to have significantly different voting patterns, where Spain and Poland had clearly higher intercepts than the United Kingdom or the Czech Republic.

## Introduction

The European Union is union of a group of 28 member states that has institutions that govern common economic, social and security issues. Under the Common Internal Market, the EU guarantees the free movement of goods, services and people inside its borders. Since 2008, after the Eurozone Debt Crisis, there has been a growing wave of discontent in Europe that culminated in the decision of the United Kingdom to leave the bloc in a referendum in 2016. Populist and Eurosceptic parties that have an EU Membership Referendum in their political agenda have appeared in many countries.

In order to better understand the voters and to get a glimpse of the future of the European Union, this Report set out to answer two main questions:

- What predictors are useful in identifying individuals that would vote to leave the European Union?
- Are there differences at the country level?

## Data

The dataset for this report came from the European Research Infrastructure Consortium. The European Social Survey (ESS), is an initiative that aims to map the attitudes, behavioural patterns and beliefs of individuals across Europe. The dataset contained over 570 variables and 45.000 observations, of which there were approximately 3.000 per Country. The topics of the variables could be categorized into eight main groups that included Politics, Climate-Change, Socio-Demographic, Country, Human Values, Welfare Attitudes, Media and Social Trust and Well-Being.

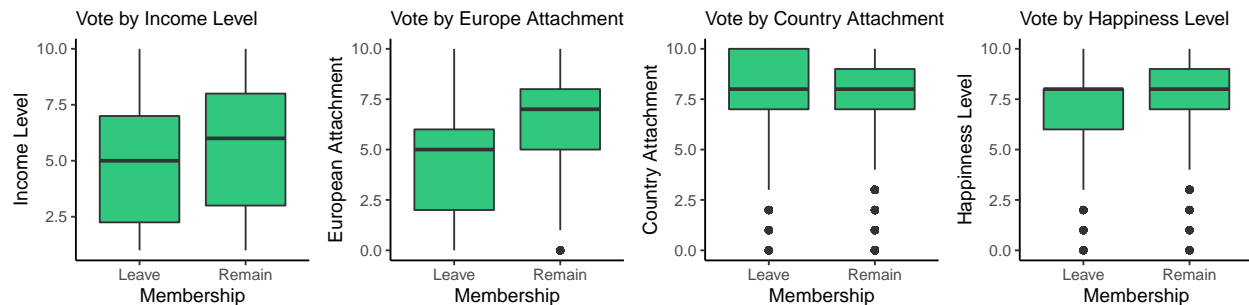
The first step in the analysis of the dataset was the selection of the variables that would be included in the model and tested for statistical significance. The process to do so was through the analysis of the documentation and the inspection of the variables. Variables that were country specific, that had over 30% of missing values or that were deemed not relevant to the prediction of a remain vote were automatically removed from the analysis. A second step in the variable selection process was the evaluation of the most interesting ones, which was a qualitative process, and the removal of variables that would be highly correlated i.e. Trust in Politicians and Trust in Political Parties.

The variables that were selected were the following:

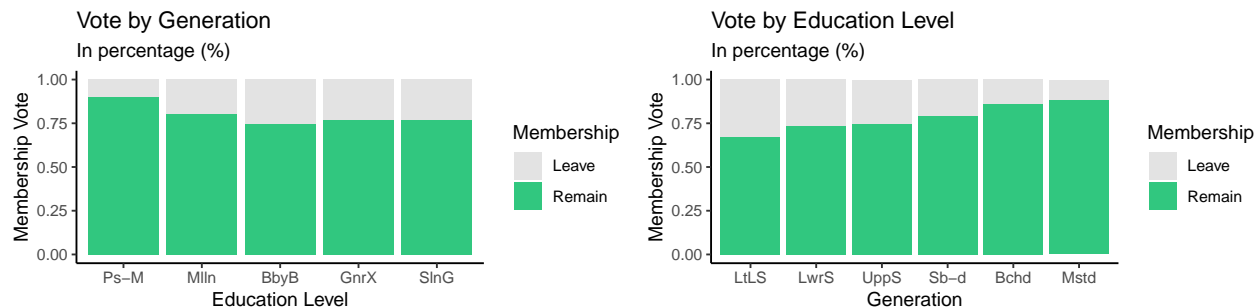
1. Country (factor): Country
2. Membership (binary): Would vote for [country] to remain member of European Union or leave
3. European\_Attachment (numeric): How emotionally attached to Europe
4. Country\_Attachment (numeric): How emotionally attached to [country]
5. Income\_Level (numeric): Household's total net income, all sources
6. Education\_Level (factor): Highest level of education
7. Generation (factor): Respondent's generation
8. Gender (factor): Respondent's gender
9. Minority (factor): Do you belong to a minority in [country]?
10. Marital\_Status (factor): Legal marital status
11. Happiness (numeric): How happy are you?
12. Political (numeric): How interested in politics are you?

After variables were selected, observations were filtered to only select countries where the Membership variable was applicable, that is member states of the European Union, including the United Kingdom. As a final step the dataset was filtered to only account for complete observations. For this analysis, it was assumed that data was missing completely at random for all variables, as there were no clear signs or indications that it was not the case. There was one exception to this though, as Income level seemed to miss not at random. It would be expected that income deciles among the population would be evenly distributed, but it was not the case in the dataset, as the top and bottom deciles represented a lower percentage of the population than would be expected.

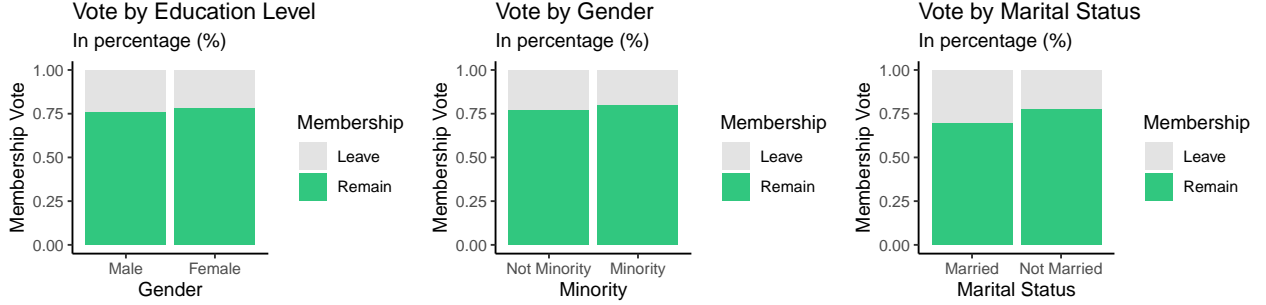
After having cleaned the dataset, the data was explored to identify any potential associations between the outcome variable and the predictors before fitting the data to the model.



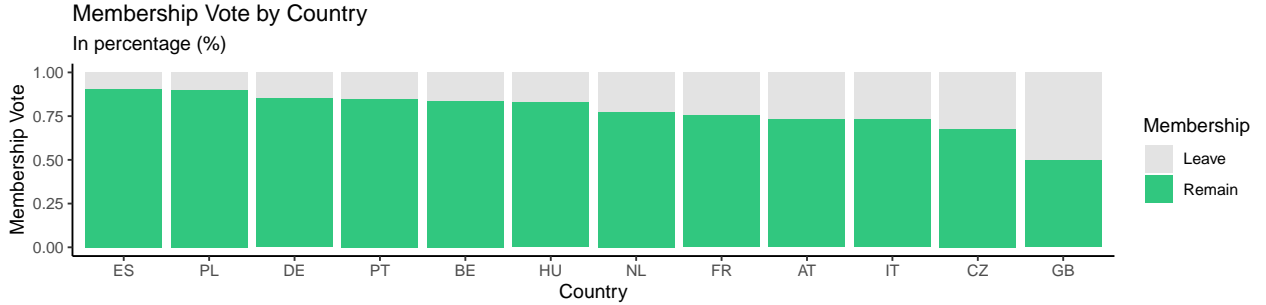
- There seems to be a significant positive relationship between Income Level, Attachment to Europe and Education Level. The wealthier, more educated and more attached people feel to Europe, the more likely it seems they are to vote to Remain in the European Union.
- It is hard to tell whether there is a relationship between someone's attachment to their own country and a vote in a hypothetical referendum to leave or remain in the EU. From the boxplot, medians seem to fall around the same value, although the Interquartile range is narrower for remainders.



- Younger generations of voters seem to be more likely to vote in favour of staying in the European Union than older ones. It seems support for the European Union plateaus with Baby Boomers, Generation X and the Silent Generation.
- There seems to be an association between more Education and a higher likelihood of a remain vote. Around 85% of the people with a degree would vote to remain in the EU.



- Females and Minorities seem to be slightly more in favour of staying in the EU. Married people would seem to be more likely to vote to leave the EU than their counterparts.



- At the Country Level, there are clear differences in the ratio of voters who would choose to remain and leave the European Union. By looking at aggregate figures, Spain and Poland are more supportive of the EU than the Czech Republic or the United Kingdom.

We will proceed to fit a model to the data to statistically test for significance.

## Model

A random-intercepts model was used to fit the data. The general formula for a logistic random-intercepts model is the following:

$$y_{ij}|x_{ij} \sim \text{Bernoulli}(\pi_{ij}); i = 1, \dots, n_j; j = 1, \dots, J;$$

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_{0j} + \tau_0 + \beta_1 x_{1ij};$$

$$\tau_{0j} \sim N(0, \sigma_0^2)$$

where  $i$  indexes observations,  $j$  indexes groups,  $\sigma^2$  is the within group (residual) variance, and  $\tau^2$  is the between group (intercept) variance.

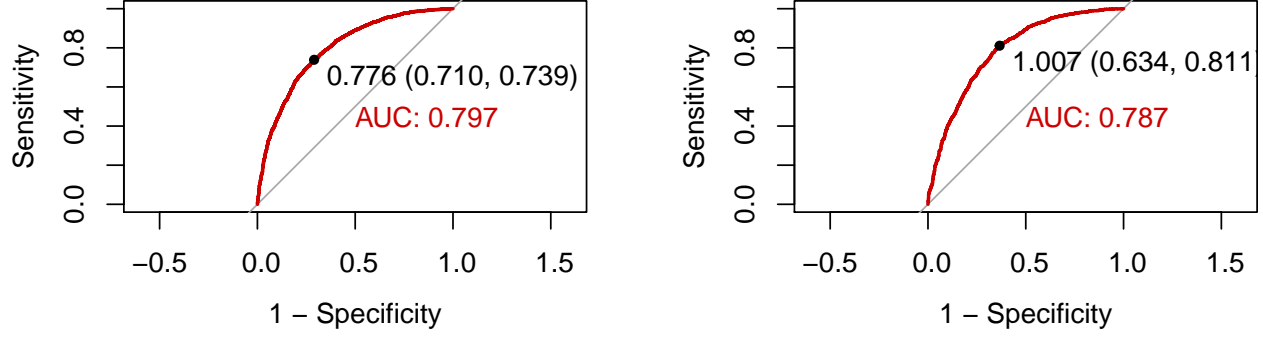


Figure 1: AUC results in Train(left) and Test(right) sets

The data was split between a Train (70%) and a Test (30%) set and a final model was fit with all the main variables. Countries were used as the level since we want to compare the results of a hypothetical referendum across member states. Therefore, it is most appropriate to generate the varying-intercepts model using Countries as the random effect variable. Most of the intuition and hypothesis that were developed by exploring the data were found to be true. Attachment to Europe, Education Level, Income Level, Political Interest and Happiness Level had a positive association with a remain vote. On the other hand Country Attachment had a negative association with a remain vote. At a lower significance level Females and Minorities seemed to be more likely to vote in favour of remaining in the European Union. Being Married did not seem to have an impact on a vote to remain or to leave the European Union. The final Model had a remarkably similar AUC between the Train (0.795) and Test (0.794) set, performing only marginally worse in the latter, suggesting there was very little overfitting. The Sensitivity and Specificity at the best threshold also hovered around 0.7 to 0.75 in both Test and Train data. If we combine the values of Accuracy, Sensitivity and Specificity for the model it seems that a model with all these variables would have a pretty good predictive ability.

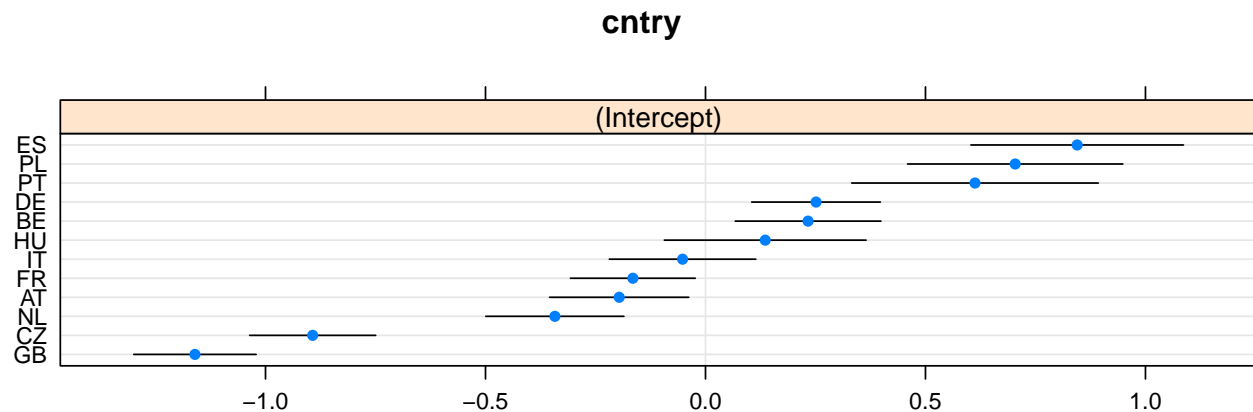
If we wanted to build a more accurate model to predict remain or leave vote based on the variables that have been analyzed we could include Interactions between Country Attachment and European Attachment, Income Level and Education Level and drop Minority. Multiple versions of the same model were built doing so and F-tests pointed to a better predictive ability of the latter. Nevertheless, this is not the aim of this report.

## Conclusions

The following table summarizes the Confidence Intervals of the odds of the predictors:

	2.5 %	97.5 %
(Intercept)	0.8276023	3.1916222
GenderMale	0.8261361	1.0120960
relevel(Generation, ref = "Post-Millennials")Baby Boomers	0.2173599	0.4240162
relevel(Generation, ref = "Post-Millennials")Generation X	0.2223760	0.4350384
relevel(Generation, ref = "Post-Millennials")Millennials	0.2846838	0.5643460
relevel(Generation, ref = "Post-Millennials")Silent Generation	0.3262927	0.6727356
MinorityNot Minority	0.5841119	1.0049308
Marital_StatusNot Married	0.7392148	1.3648160
Income_Level	1.0262889	1.0708222
European_Attachment	1.4394450	1.5091425
Country_Attachment	0.8470523	0.8923379
Happy	1.0377556	1.1000564
relevel(Political, ref = "Not at all Interested")Hardly Interested	1.0406590	1.3991869
relevel(Political, ref = "Not at all Interested")Quite Interested	1.1178288	1.5354261
relevel(Political, ref = "Not at all Interested")Very Interested	0.9103959	1.3585334
relevel(Education_Level, ref = "Less than Lower Secondary")Bachelor degree	1.6535126	2.8154255
relevel(Education_Level, ref = "Less than Lower Secondary")Lower Secondary	0.8208698	1.2345560
relevel(Education_Level, ref = "Less than Lower Secondary")Master degree	1.8615938	3.0834814
relevel(Education_Level, ref = "Less than Lower Secondary")Sub-degree	1.2003271	1.9205930
relevel(Education_Level, ref = "Less than Lower Secondary")Upper Secondary	0.9436310	1.3939609

A dotplot of the multilevel model can be used to identify voting pattern differences across member states. From the results of the dotplot shown below, it is clear that response results vary across countries, where Spain, Poland and Portugal have clearly higher intercepts than the rest, while the United Kingdom and the Czech Republic have significantly lower ones.



Even though we dropped uncomplete observations while data was missing not at random for Income and the Binned Plots of the response variable showed a certain pattern for the average of the residuals, many insights about the voting patterns of the people of Europe in a hypothetical Referendum could be generated. Specifically, the results showed that there is an increase in remain vote for wealthier, more educated people. Younger generations were clearly more pro-EU than their older counterparts.

From now onwards these are just my thoughts after seeing the results of the analysis. I feel more relieved after having worked on this topic. My conclusio after this project is that the current wave of euroscepticism will fade away as the economic situation in Europe gets better. While there is a large part of the population that would support to leave the EU, there is also solid support across all member state to remain. In the future, as new generations are born and the population gets increasingly better education, we should see increased support for the EU.