Cluster Analysis of House of Commons Brexit Votes

The votes (V_n) are categorized into "*Aye_vote*", "*No_vote*" and "*Absent*", for each separate vote session (j) and these items were binarized to create two new variables (R_n) and (A_n):

$$R_{n,j} = \begin{cases} 0, & x = \text{"}No\_Vote\text{"} \\ 1, & x = \text{"}Aye\_Vote\text{"} \end{cases}$$

$$A_{n,j} = \begin{cases} 0, & x = \text{"}Absent\text{"} \\ 1, & x = \text{"}Not\ Absent\text{"} \end{cases}$$

The data could also be binarized by merging the categories, "*No_vote*" and "*Absent*", but this would mean that potentially important information on the reasons for remaining absent are lost. In other words, the factors that drive an MP to vote no are different from those that drive them to abstain. To combine these categories compromises the "*No_vote*" category by adding irrelevant data, while also disregarding the implications of abstaining from the vote.

The data were binarized to facilitate investigation of clustering using finite mixture modelling. Here we apply statistical models defined by $p(x|\tau, \theta, G)$ to assess the probability (τ_g) that each data point ($R_{n,j}$ and A_{nj}) came from a separate group (g):

$$p(x_n|\tau, \theta, G) = \sum_{g=1}^{G} \tau_g p(x_n|\theta_g).$$

Therefore, each vote has probability τ_n of belonging to group g, and data within group g are modelled as binomial with probability p_g

The data were modelled using a mixture model and the R function, "*multmix*" to assess if a grouped model has a better fit to the data. This function returns the EM algorithm output for the mixture of distributions in this example. The final mixing proportions (λ) were 0.5589945 and 0.4410055 for groups 1 and 2 respectively, suggesting that there is indeed clustering in the dataset, and a finite mixture model would be appropriate to accommodate this structuring.

This was further investigated by looping over values of g, to assess the number of clusters in the data. It is evident that dividing the data into 6 groups results in an increase in the markers of model fit, BIC and log likelihood, but there is little improvement gained by further division. Therefore, g=6 is the optimum grouping. These analyses show that the House of Commons politicians do fall into at least 6 important clusters when voting on Brexit.

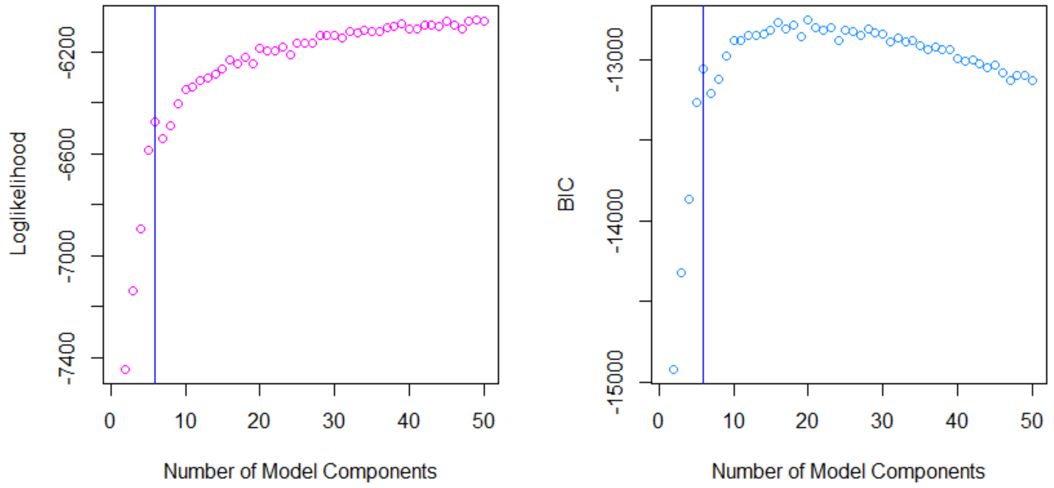

Figure 1 The BIC and log likelihood when data were clustered into groups or components ranging from 2-50, with 6 clusters providing the most useful division of the dataset.

There are at least 6 clusters found using the finite model in Q1, and next we further investigate the clusters using Latent Class Analysis (LCA). If $x_n = (x_{n1},x_{n2},....,x_{nm})$ represents a single MP's selection on one of the 14 votes, then the probability that X_n was an element of class g is:

$$p(x_n|\theta_g) = \prod_{m=1}^{M} p(x_{nm}|\theta_{gm}) = \prod_{m=1}^{M} \prod_{c=1}^{C_m} \theta_{gmc}^{\mathcal{I}(x_{nm}=c)},$$

Where $\{\theta_{gm1},..., \theta_{gm}C_m\}$ are the probabilities of observing the categories {"absent","aye_vote" and "no_vote"} in variable (vote) m = {V1,V2,.....,V14}. In this way, the parameters θ_g can be use to characterise the differences between the groups or classes selected to classify MPs based on their voting patterns.

LCA was applied to assess clustering among MPs in multi-way tables of the categorical vote variables. This model allows confounding between the variables can be explained by a single unobserved "latent" variable. The R function "poLCA" was used to fit a finite mixture model with th number of classes set to 6, as estimated using the model in Q1. The function *poLCA* uses EM and Newton-Raphson algorithms to maximize the log-likelihood of the latent class model.

A LCA model fitted to the vote data, with 6 latent classes was a good fit to the data, as evidenced by the BIC (9795) and the AIC (9024), in comparison to the null model and the model derived in Q1. A likelihood ratio test was significant:

$$\chi^2_{(173,637)} = 9382244,\ p < .0001$$

The 6 latent classes are summarised by the probability of voting yes, and of being absent for each vote (Figure 3 and 4). Further information on the clustering is given by the overall probability of belonging to class g, plotted with se in Figure 2. The probabilities show relatively low variation between groups, which indicates useful segregation of the MPs into clusters. Class 1 and 4 are the dominant categories, accounting for over 60% of the entire group of MPs.

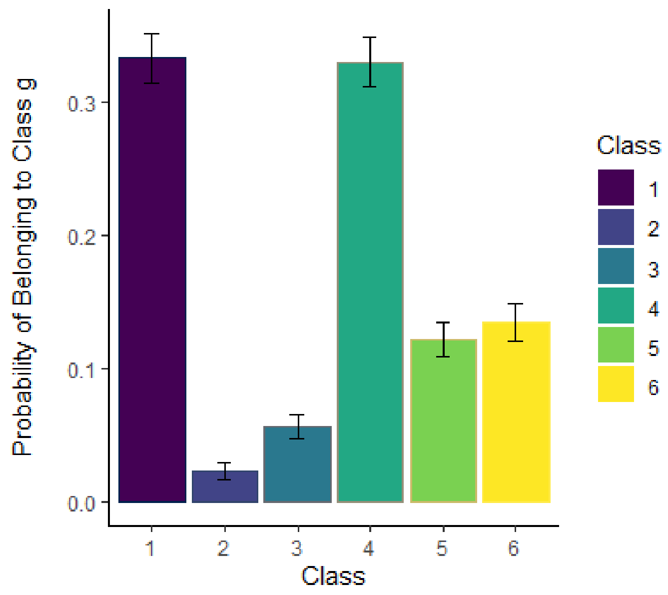

Figure 2 The probability and se of belonging to each of the 6 classes identified in these analyses.

The subject of each vote is given below, but it is not possible to make any conclusions about the importance of a no, yes or abstain option on any of these, in the absence of any information about their relevance to the Brexit vote. Vote 5, "*United Kingdom's withdrawal from the European Union*" is the only vote that has a clear meaning in this context.

List of Votes

V1	Nick Boles's motion D (Common Market 2.0)
V2	Mr Clarke's motion C (Customs Union)
V3	Peter Kyle's motion E (Confirmatory public vote)
V4	Joanna Cherry's motion G (Parliamentary Supremacy)
V5	United Kingdom's withdrawal from the European Union
V6	Mr Baron's motion B (No deal)
V7	Nick Boles's motion D (Common market 2.0)
V8	George Eustice's motion H (EFTA and EEA)
V9	Mr Clarke's motion J (Customs union)
V10	Jeremy Corbyn's motion K (Labour's alternative plan)
V11	Joanna Cherry's motion L (Revocation to avoid no deal)
V12	Margaret Beckett's motion M (Confirmatory public vote)
V13	Mr Fysh's motion O (Contingent preferential arrangements)
V14	Draft European Union (Withdrawal) Act 2018 (Exit Day) (Amendment) Regulations 2019

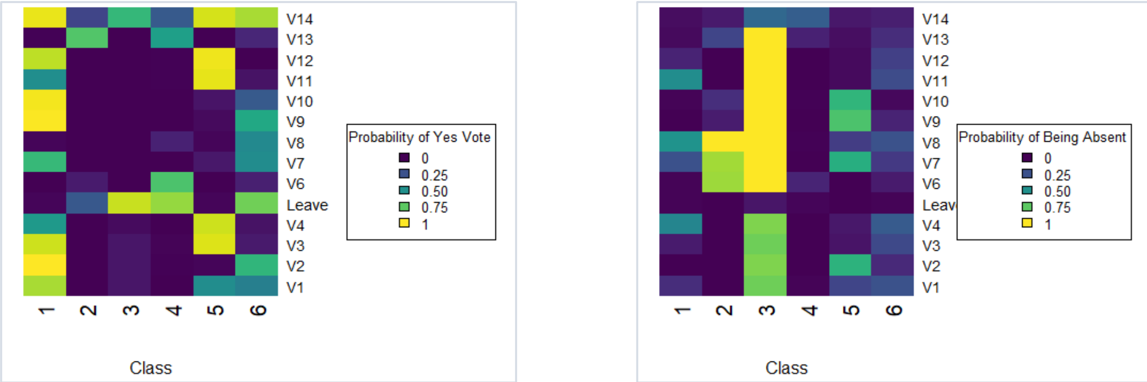

Figure 3 The probability of voting yes (left) and of being absent (right) for each of the 6 classes identified in these analyses.

Table 1 The probability of voting yes to each of the 14 questions for each of the 6 latent classes

Class	V1	V2	V3	V4	Leave	V6	V7	V8	V9	V10	V11	V12	V13	V14
1	0.87	1.00	0.92	0.54	0.01	0.00	0.67	0.01	1.00	0.98	0.49	0.90	0.00	0.97
2	0.00	0.00	0.00	0.00	0.27	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.73	0.20
3	0.06	0.06	0.06	0.03	0.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.67
4	0.00	0.01	0.01	0.00	0.84	0.72	0.00	0.09	0.00	0.00	0.00	0.00	0.56	0.28
5	0.49	0.01	0.95	0.92	0.01	0.00	0.06	0.01	0.03	0.05	0.96	0.97	0.00	0.94
6	0.43	0.65	0.07	0.05	0.78	0.09	0.48	0.47	0.60	0.28	0.05	0.00	0.11	0.87

Table 2 The probability of being absent for each of the 14 questions for each of the 6 latent classes

Class	V1	V2	V3	V4	Leave	V6	V7	V8	V9	V10	V11	V12	V13	V14
1	0.13	0.00	0.07	0.45	0.01	0.01	0.25	0.51	0.00	0.01	0.49	0.09	0.02	0.03
2	0.00	0.00	0.00	0.00	0.00	0.85	0.86	1.00	0.08	0.13	0.00	0.00	0.20	0.07
3	0.78	0.81	0.78	0.81	0.06	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.33
4	0.01	0.00	0.00	0.00	0.01	0.10	0.00	0.01	0.00	0.00	0.00	0.00	0.09	0.29
5	0.20	0.64	0.05	0.06	0.00	0.00	0.63	0.18	0.72	0.65	0.03	0.03	0.04	0.06
6	0.25	0.12	0.22	0.28	0.01	0.07	0.17	0.25	0.10	0.02	0.23	0.19	0.13	0.08

Recall that Class 1 and Class 4 were the two most dominant clusters, and it is evident that these represent a group of MPs that voted to leave (probability of voting yes to V5 = 0.01), and that voted to stay (probability of voting yes to V5 = 0.81). This suggests that the latent class model has segregated the MPs by their likelihood of voting for or against leaving the EU, and it seems likely that such clustering would represent diverse opinions and reflect their voting patterns on other issues too.

Class 1 voted mostly no to leave the UK, and then tended to vote yes to every other vote except V13. The converse is true for Class 3 and 4 who voted to leave the UK and then tended to vote no or were absent for to every subsequent vote.

Class 5 voted not to leave the UK, but then were equivocal about many of the other votes, almost like a less extreme version of Class 1.

Class 2 is interesting, as only 27% wanted to leave the EU, but then they voted no to almost every other vote, with their high acceptance (73%) of V13 being exceptional to all other classes. It seems odd that this group voted so decisively in all other votes except the one to leave the UK.

Class 3 are not worth considering, as their rate of abstaining is extremely high, although not for the leave the UK vote. Class 2 were notably absent from V6-V8. Otherwise, the rate of absence is less than one quarter for most votes. It is perhaps an indication of its significance that very few MPs in any Class abstained from V5, the Leave the UK vote.

In general, this latent class model seems to have identified realistic cluster of MPs from their patterns in voting. The model segregated the MPs into classes with different probability of wanting to leave the UK (yes answer to V5) and then highlighted different patterns of voting between the classes.

The classes identified can also be summarized by the variability in their probability of voting yes, or of being absent from each of the 14 votes (Table 3-4 and Figure4). The only notable feature is relatively high variability in the probability of being absent and of voting yes, in Class 2. Otherwise, the probabilities show relatively low variation within groups, which indicates that the model has efficiently segregated the MPs into clusters.

Table 3 The se in the probability of voting yes to each of the 14 questions for each of the 6 latent classes

Class	V1	V2	V3	V4	Leave	V6	V7	V8	V9	V10	V11	V12	V13	V14
1	0.03	0.00	0.04	0.06	0.02	0.00	0.04	0.01	0.01	0.01	0.06	0.05	0.00	0.01
2	0.00	0.00	0.00	0.00	0.10	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.63
3	0.03	0.03	0.03	0.02	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09
4	0.00	0.01	0.01	0.00	0.03	0.04	0.00	0.02	0.00	0.01	0.01	0.00	0.04	0.04
5	0.21	0.02	0.08	0.11	0.01	0.00	0.05	0.03	0.03	0.04	0.06	0.03	0.00	0.03
6	0.09	0.10	0.05	0.03	0.10	0.09	0.10	0.10	0.10	0.12	0.03	0.00	0.09	0.05

Table 4 The se in the probability of being absent for each of the 14 questions for each of the 6 latent classes

Class	V1	V2	V3	V4	Leave	V6	V7	V8	V9	V10	V11	V12	V13	V14
1	0.03	0.00	0.04	0.06	0.01	0.01	0.03	0.05	0.00	0.01	0.06	0.05	0.01	0.01
2	0.00	0.00	0.00	0.00	0.00	0.17	0.45	0.00	0.69	0.71	0.00	0.00	0.07	0.06
3	0.03	0.07	0.03	0.04	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09
4	0.01	0.00	0.00	0.00	0.01	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.02	0.03
5	0.09	0.17	0.08	0.09	0.00	0.00	0.17	0.08	0.14	0.14	0.04	0.03	0.02	0.03
6	0.07	0.05	0.08	0.08	0.02	0.04	0.07	0.08	0.05	0.02	0.07	0.08	0.05	0.04

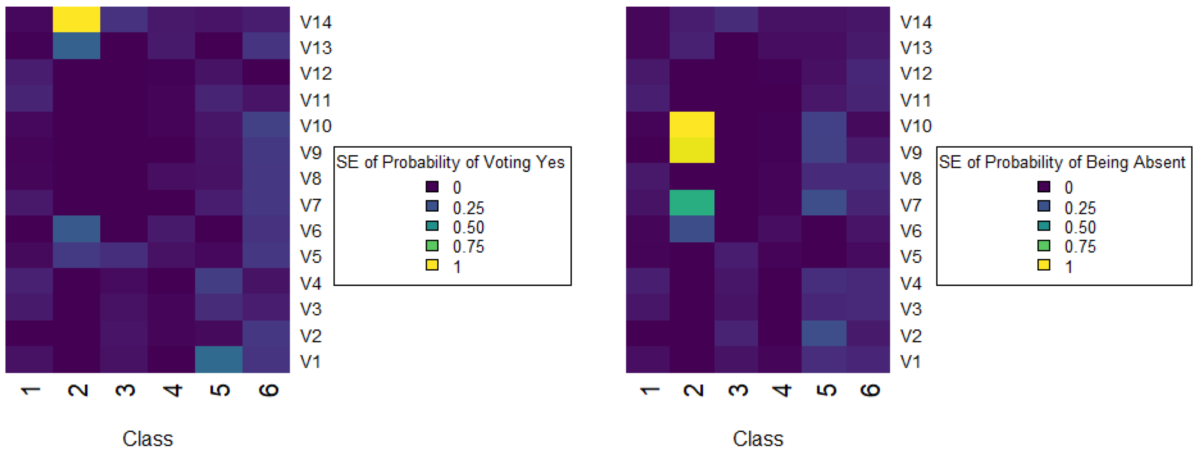

Figure 4 The probability of voting yes (left) and of being absent (right) for each of the 6 classes identified in these analyses.