National Centre for Social Research

Dividing Lines Voter Classification: Technical Note

Authors: Lovisa Moller Vallgarda, Daniel Fisher, Ekaterina Khriakova

JULY 2024

Contents

1. Background

1.1 Introduction

The National Centre for Social Research (NatCen) launched a classification of British voters on Wednesday 5 June. This 'Dividing Lines' classification aimed to assess the underlying structure of ideological divisions within the British electorate. It is intended to provide information on, and support further analysis of, voter behaviour and voter coalitions. The classification helped inform the BBC's Undercover Voters project[1] during the 2024 General election campaign. Based on who voters are and what values they uphold, the research identified the six key types of voters in Britain today.

This technical note sets out the data selection, data transformations and cluster algorithm used in the research. For a description of other aspects of the work, please visit https://natcen.ac.uk/dividing-lines.

1.2 Project contributors

The research project that this technical note is related to was led by Lovisa Moller Vallgarda, Director of Analysis, and Sir Prof John Curtice, Senior Research Fellow. Project contributors included Dr Daniel Fisher, NatCen Associate, and Ekaterina Khriakova, Researcher, who led on data modelling and data preparation respectively. Other project contributors were Gianfranco Addario, Research Director, Sam Beardsworth, Senior Data Scientist, Joe Crowley, Senior Researcher, Ian Montagu, Senior Researcher, Bernard Steen, Research Director, and Charles Wilson, Researcher.

2. Data selection

2.1 British Social Attitudes (BSA) 2023

This analysis uses data from the most recent wave of the British Social Attitudes (BSA) survey. BSA survey is carried out by the National Centre for Social Research (NatCen) and has been tracking public attitudes for over 40 years. It uses a gold standard random probability sampling design to reduce the risk of bias and produce robust estimates of the views of British adults. A total of 5,578 interviews with a representative, random sample of adults in Britain was conducted between 12 September and 31 October 2023 for BSA 2023.

2.2 Questions used in the classification

Over 7,000 survey question combinations were considered, based on hypotheses generated by the group of project contributors. We considered letting a larger number of survey questions drive the grouping, and then explored whether we could remove some and still achieve an equally powerful outcome. Other things being equal, a smaller number of survey questions were preferred as this makes introducing these questions in other surveys less taxing on the survey respondents.

[1] BBC Newscast, 02 Jun 2024, 'Electioncast: The UK Undercover Voters', https://www.bbc.co.uk/sounds/play/p0j1skx6

The questions included in NatCen's left-right and libertarian-authoritarian scales featured in several hypotheses. Many of the questions in these scales, introduced by our predecessors in 1986, also feature on the list below as they were found to be powerful predictors of other political views. In contrast, the redistribution scale questions did for example perform less well in this regard are not included below. A broad range of other both well-established and more recent BSA questions were also considered. The table below shows the 12 survey questions used to create the final classification.

Table 1 Questions used in the classification

Question	Response options
SupParty: Generally speaking, do you think of yourself as a supporter of any political party?	1 Yes 2 No 8 Don't know 9 Prefer not to answer
Politics: How much interest do you generally have in what is going on in politics?	1 A great deal 2 Quite a lot 3 Some 4 Not very much 5 None at all 8 Don't know 9 Prefer not to answer
GTNext: How would you feel if someone from the Roma, Gypsy or Traveller community moved in to a house or flat next door to you?	1 Very happy 2 Happy 3 Neither happy nor unhappy 4 Unhappy 5 Very unhappy 8 Don't know 9 Prefer not to answer
Redistrb: Government should redistribute income from the better-off to those who are less well off	1 Agree strongly 2 Agree 3 Neither agree nor disagree 4 Disagree 5 Disagree strongly 8 Don't know 9 Prefer not to answer
BigBusnN: Big business benefits owners at the expense of workers	1 Agree strongly 2 Agree 3 Neither agree nor disagree 4 Disagree 5 Disagree strongly 8 Don't know 9 Prefer not to answer
Wealth: Ordinary working people do not get their fair share of the nation's wealth	1 Agree strongly 2 Agree 3 Neither agree nor disagree 4 Disagree 5 Disagree strongly

	8 Don't know 9 Prefer not to answer
RichLaw: There is one law for the rich and one for the poor	1 Agree strongly 2 Agree 3 Neither agree nor disagree 4 Disagree 5 Disagree strongly 8 Don't know 9 Prefer not to answer
TradVals: Young people today don't have enough respect for traditional British values	1 Agree strongly 2 Agree 3 Neither agree nor disagree 4 Disagree 5 Disagree strongly 8 Don't know 9 Prefer not to answer
StifSent: People who break the law should be given stiffer sentences	1 Agree strongly 2 Agree 3 Neither agree nor disagree 4 Disagree 5 Disagree strongly 8 Don't know 9 Prefer not to answer
DeathApp: For some crimes, the death penalty is the most appropriate sentence	1 Agree strongly 2 Agree 3 Neither agree nor disagree 4 Disagree 5 Disagree strongly 8 Don't know 9 Prefer not to answer
Wonglaw: The law should always be obeyed, even if a particular law is wrong	1 Agree strongly 2 Agree 3 Neither agree nor disagree 4 Disagree 5 Disagree strongly 8 Don't know 9 Prefer not to answer
Censor: Censorship of films and magazines is necessary to uphold moral standards	1 Agree strongly 2 Agree 3 Neither agree nor disagree 4 Disagree 5 Disagree strongly 8 Don't know 9 Prefer not to answer

Source: National Centre for Social Research (2024). British Social Attitudes Survey 2023. Survey data scheduled for publication via the UK Data Service. DOI: http://doi.org/10.5255/UKDA-Series-200006

3. Addressing non-substantive answers

3.1 Overview

Table 2 outlines the number of survey respondents who provided a substantive answer for the 12 questions. The response other than “Don’t know” or “Prefer not to answer” was recorded for more than 99% of the sample for 11 of the 12 questions. The exception was GTNext, where the response rate was 98.0%. The substantive response rate ranges from 5,465 to 5,560, which is 98.0% to 99.8% of the sample. This is likely related to the survey design, where the options “Don’t know” or “Prefer not to answer” can be spontaneous answers from the respondent, but the question prompt does not mention them.

3.2 Handling “Don’t know” responses

To segment the data, we considered pairwise distances between all respondents, where survey respondents were considered closer or further apart depending on their answers to the 12 questions. Responses of 'Don't know' can be handled in several ways: 'Don't know' responses can be treated as missing data and subsequently imputed; respondents with one or more 'Don't know' answers can be excluded from the analysis (complete case analysis); or they can be assigned a specific distance from the other response options. All approaches were explored.

We considered multiple imputation with predictive mean matching using the mice package (version 3.16.0) in R. This process generated 10 separate copies of the dataset to handle uncertainty in the imputed values. These 10 datasets were combined and used as input for the clustering algorithm, which then segmented the merged dataset. We also ran a complete case analysis. Finally, we settled on a specific distance assignment. This assignment generated a stronger cluster solution than the imputation, which was highly similar in its assignment to the complete case analysis and where the respondents who were affected by the specific distance assignment were assigned to clusters which they broadly agreed with on related survey questions.

As detailed in table 2, less than 1% of the sample selected the “Don’t know” option for 11 of the 12 questions. The exception was GTNext, where this percentage was slightly higher but still comparatively low at 1.5%. When assessing patterns in the data between questions on a similar topic, the “Don’t know” response was identified as most similar to a “Neither agree nor disagree” answer based on the respondents’ response patterns on related questions. We expect that this reflects these respondents’ lack of strong opinions, or willingness to share a strong opinion, on the topic. Both “Don’t’ know” responses and “Neither agree nor disagree” responses were therefore treated as the midpoint of the scale when the distance matrix was computed.

For the questions that do not use a strongly agree to strongly disagree scale (5-Point Likert Scale), the approach was adapted as follows. For the GTNext question, both “Don’t’ know” responses and “Neither happy nor unhappy” responses were treated as the midpoint of the scale. For the SupParty question, “Don’t know” responses were treated as in between “Yes” and “No”. Finally, for the Politics question, “Don’t know” could either be treated as equal to “Some” or “Not very much”. After some reflection and assessments of patterns in the data, we ran both analyses. Not a single survey respondent changed cluster assignment depending on this decision.

Table 2 Substantive answers

Question	Percentage of people who selected "Don't know"	Percentage of people who selected "Prefer not to answer"	Number of people who provided a substantive answer	Percentage of people provided a substantive answer
SupParty: Generally speaking, do you think of yourself as a supporter of any political party?	0.14	0.18	5560	99.7
Politics: How much interest do you generally have in what is going on in politics?	0.04	0.13	5569	99.8
GTNext: How would you feel if someone from the Roma, Gypsy or Traveller community moved in to a house or flat next door to you?	1.45	0.57	5465	98.0
Redistrb: Government should redistribute income from the better-off to those who are less well off	0.43	0.3	5537	99.3
BigBusnN: Big business benefits owners at the expense of workers	0.54	0.27	5533	99.2
Wealth: Ordinary working people do not get their fair share of the nation's wealth	0.39	0.3	5539	99.3
RichLaw: There is one law for the rich and one for the poor	0.43	0.27	5539	99.3
TradVals: Young people today don't have enough respect for traditional British values	0.29	0.23	5549	99.5
StifSent: People who break the law should be given stiffer sentences	0.29	0.22	5550	99.5
DeathApp: For some crimes, the death penalty is the most appropriate sentence	0.34	0.16	5550	99.5
Wonglaw: The law should always be obeyed, even if a particular law is wrong	0.29	0.29	5546	99.4
Censor: Censorship of films and magazines is necessary to uphold moral standards	0.3	0.2	5550	99.5

Source: National Centre for Social Research (2024). British Social Attitudes Survey 2023. Survey data scheduled for publication via the UK Data Service. DOI: http://doi.org/10.5255/UKDA-Series-200006

4. Cluster analysis

4.1 Approach

There are a series of approaches available to examines respondents' responses across a series of survey questions to identify groups with similar views. For this project, we utilised various extensions of the k-medoids clustering algorithm, a machine learning technique used to detect distinct groups in multidimensional data. The final cluster solution uses the k-medoids pam clustering algorithm from sklearn in Python, with Manhattan distance, 500 maximum iterations when fitting, and heuristic initialisation.

The k-medoids clustering method is sensitive to variable scaling. To equalise the impact of each survey question, the responses to each question was first standardised by subtracting their means and dividing by their standard deviations.

This clustering method can also be influenced by initialisation. In this project we noticed a marginal impact of the initialisation chosen which indicates the presence of more than one local cluster optimum. This occurs because k-medoids relies on iterative processes that start from an initial set of points (medoids) and then refine these points to minimise the distances within clusters. To address this, multiple initialisations were used selectively to assess whether stronger cluster solutions were possible to achieve.

4.2 Selection

The project team explored over 7,000 combinations of survey questions and evaluated more than 30,000 cluster solutions. All survey question combinations, along with their cluster solutions, UMAP, and respective silhouette scores, were documented.

The strength of the clusters obtained guided our decision-making process, but it was also crucial for us and the broader research project group that the selected cluster solution was conceptually meaningful. Therefore, other cluster solutions that performed well in terms of relative cluster strength were evaluated to determine whether they could provide meaningful distinctions between the groups on issues not directly covered by the selected survey questions.

The six voter types that the research identified and the views that unite them and set them apart is available via https://natcen.ac.uk/dividing-lines.


National Centre
for Social Research