



Classification of Westminster Parliamentary constituencies using e-petition data

Stephen Clark^{1*} , Nik Lomax² and Michelle A Morris³

*Correspondence:

tra6sdc@leeds.ac.uk

¹ Consumer Data Research Centre,
Leeds Institute for Data Analytics
and School of Geography, University
of Leeds, Leeds, LS12 9JT, UK
Full list of author information is
available at the end of the article

Abstract

In a representative democracy it is important that politicians have knowledge of the desires, aspirations and concerns of their constituents. Opportunities to gauge these opinions are however limited and, in the era of novel data, thoughts turn to what alternative, secondary, data sources may be available to keep politicians informed about local concerns. One such source of data are signatories to electronic petitions (e-petitions). Such e-petitions have risen greatly in popularity over the past decade and allow members of the public to initiate and sign an e-petition online, with popular e-petitions resulting in media attention, a response from the government or ultimately a debate in parliament. These data are thus novel in their availability and have not yet been widely used for research purposes. In this article we will use the e-petition data to show how semantic classes of Westminster Parliamentary constituencies, fitted as Gaussian finite mixture models via EM algorithm, can be used to typify constituencies. We identify four classes: Domestic Liberals; International Liberals; Nostalgic Brits and Rural Concerns, and illustrate how they map onto electoral results. The findings and the utility of this approach to incorporate new e-petitions and adapt to changes in electoral geography are discussed.

Keywords: United Kingdom; Parliamentary Constituencies; classification; Gaussian finite mixture models; electronic petitions

1 Introduction

Knowledge of an area's characteristics is important in gaining an understanding of the needs of those who live in, work in or service the area. Whilst each area is unique, some areas will be very similar to others and some will be distinct. The classification or geodemographic segmentation of areas allows for those areas that are similar in nature to be grouped together as identifiable classes. These classes are usually established by using multi-variate data to characterise an area and then grouping together areas whose characteristics are broadly similar (Everitt et al. [1]). Given the nature of these data, there is the potential for these classes to be dispersed over space, with neighbouring areas belonging to different classes (Berry and Linoff [2]).

Classification can be applied at any level of geographic scale, from small neighbourhoods (Office for National Statistics [3]; Gale et al. [4]) through to municipalities (Office for National Statistics [5]). They can also be designed for general use or bespoke for a particular

topic, e.g. use of technology (Longley et al. [6]), retail (CallCredit [7]), leisure (CACI [8]), health (CACI [9]) or well-being (CACI [10]). Some are produced by commercial organisations using a combination of open source data (e.g. Census) and their own in-house survey data (CACI [11]; Experian [12]) whilst others are open and produced by government bodies or academics using public data (Classification of Workplace Zones for England and Wales [13]; Greater London Authority [14]).

In this study we form classifications of Westminster Parliamentary Constituencies (WPCs) which are characterised by the political sentiment of their constituents. Each of these 650 WPCs consists of around 70,000 electors who return a member to the House of Commons in the United Kingdom (UK) national Parliament, and their term lasts typically 5 years (Figure 7 provides a map of these WPCs). Only registered British nationals, Irish nationals and citizens of Commonwealth countries are allowed to vote in Westminster Parliamentary elections; citizens of other countries, including European Union citizens are not entitled to vote in such elections. Here the sentiment in each WPC is captured using the volume of signatories to e-petitions (Wright [15]) hosted on the UK Parliaments petitions web site. The UK Parliaments e-petitions system was established in 2006 (Miller [16]) and in its present form is administered by a Parliamentary committee. Anyone can set up such an e-petition, requiring only five co-signatories, and the government will respond to any e-petition that receives over 10,000 signatories. Those that receive over 100,000 signatories are additionally considered for debate in Parliament. A particularly successful e-petition can gain media attention, allowing its message to reach a wider audience. Individuals outside the UK are also allowed to sign e-petitions, but their participation is mapped to their country of residence and not a WPC (in reality around 98% of signatories are from the UK). This study will consider those e-petitions which have been debated during the current Parliament, from July 2015 to February 2017. Whilst UK e-petitions date back to 2006, these older e-petitions pre-date the formation of the previous government and data available before May 2015 is not geocoded. For this study there are 51 such e-petitions. The list of e-petitions, with the topic, number of signatories and the open and close dates are given in Table 1.

Information from such e-petitions are beginning to be widely used to monitor and understand the political discourse (Briassoulis [17]). This has included: examinations of how e-petitions have the potential to influence the interactions between politicians and policy makers (Bochel and Bochel [18]; Hough [19]; Dumas et al. [20]); the pattern of engagement with e-petitions and how to characterise this interaction (Huang et al. [21]; Puschmann et al. [22]); a textual analysis of the wording of the e-petition text (Hagen et al. [23]) to establish if certain topic influence the popularity of petitions; and how referendum outcomes correlate with the volume of e-petition signatories (Hanretty [24]).

The following section of this article will discuss the techniques used for identifying classes and section three will introduce the e-petitions data in more detail. Section four provides the results and diagnostics. The final section discusses the implications of the study and its wider utility.

2 Methodology

Various techniques exist to estimate classifications in n -dimensional space described using multi-dimensional data (Berry and Linoff [2]). Of the two broad approaches, partitioning and hierarchical classification, it is the former approach adopted in this article. The most

Table 1 List of e-petitions used in this study^a

ID	Topic	Signatories	Date Opened	Date Closed	Group
104334	To debate a vote of no confidence in Health Secretary the Right Hon Jeremy Hunt	231,136	20/07/2015	20/01/2016	L
104349	Make the production, sale and use of cannabis legal.	236,995	21/07/2015	21/01/2016	L
104796	Don't kill our bees! Immediately halt the use of Neonicotinoids on crops	99,909	24/07/2015	24/01/2016	L
105560	Fund more research into brain tumours, the biggest cancer killer of under-40s	120,129	03/08/2015	04/02/2016	M
105991	Accept more asylum seekers and increase support for refugee migrants in the UK.	450,287	13/08/2015	14/02/2016	L
106133	Make an allowance for up to 2 weeks term time leave from school for holiday.	127,199	14/08/2015	15/02/2016	C
106477	Stop allowing immigrants into the UK.	216,949	25/08/2015	26/02/2016	C
106651	Introduce a tax on sugary drinks in the UK to improve our children's health.	155,516	26/08/2015	27/02/2016	L
108072	Give the Meningitis B vaccine to ALL children, not just newborn babies.	823,348	14/09/2015	15/03/2016	C
108570	Free Sergeant Alexander Blackman	34,440	16/09/2015	17/03/2016	C
108782	The DDRB's proposals to change Junior Doctor's contracts CANNOT go ahead.	110,065	22/09/2015	23/03/2016	L
108944	Save British Steel making. Scunthorpe, Teesside, Port Talbot etc.	18,429	24/09/2015	25/03/2016	NA
109383	Stop the scathing cuts to the Police budget.	9,947	05/10/2015	06/04/2016	M
109649	Prevent the scrapping of the maintenance grant.	133,069	02/10/2015	03/04/2016	L
109702	Restrict the use of fireworks to reduce stress and fear in animals and pets	104,038	02/10/2015	03/04/2016	C
110776	Make fair transitional state pension arrangements for 1950's women	193,186	20/10/2015	21/04/2016	M
111731	Include expressive arts subjects in the Ebacc	102,499	09/11/2015	10/05/2016	L
112342	Stop the destructive 'building our future' office closure programme in HMRC	2,585	16/11/2015	17/05/2016	NA
113064	Vote no on military action in Syria against IS in response to the Paris attacks	227,745	20/11/2015	21/05/2016	L
113231	No UK airstrikes on Syria.	190,223	22/11/2015	23/05/2016	L
113491	Keep the NHS Bursary	162,568	24/11/2015	25/05/2016	L
114003	Block Donald J Trump from UK entry	586,930	08/12/2015	09/06/2016	L
114907	Don't ban Trump from the United Kingdom	46,622	09/12/2015	10/06/2016	M
115895	Scrap plans forcing self-employed & small business to do 4 tax returns yearly	114,504	16/12/2015	17/06/2016	L
116762	STOP CAMERON spending British taxpayers' money on Pro-EU Referendum leaflets	221,866	22/12/2015	23/06/2016	C
118060	Scrap the £35k threshold for non-EU citizens settling in the UK	114,636	12/01/2016	12/07/2016	L
120702	Urge the South Korean Government to end the brutal dog meat trade	102,131	09/02/2016	09/08/2016	M
121262	Jeremy Hunt to resume meaningful contract negotiations with the BMA.	120,680	11/02/2016	11/08/2016	L
123324	The 23 June should be designated as Independence Day, and celebrated annually.	21,292	03/03/2016	03/09/2016	C

Table 1 (Continued)

ID	Topic	Signatories	Date Opened	Date Closed	Group
125003	Ban driven grouse shooting	123,077	20/03/2016	20/09/2016	M
125333	Repeal the 1972 European Communities Act immediately on EU Referendum OUT vote.	32,943	23/03/2016	23/09/2016	C
125475	Debate in the House the Local Government Pension Scheme Investment Regulations	105,772	24/03/2016	24/09/2016	M
125692	Stop spending a fixed 0.7 per cent slice of our national wealth on Foreign Aid	235,979	24/03/2016	24/09/2016	C
129698	No more school penalty fines and bring back the 10 day authorised absence	204,790	09/05/2016	10/11/2016	C
131167	Stop retrospective changes to the student loans agreement	134,194	23/05/2016	24/11/2016	L
131215	EU Referendum Rules triggering a 2nd EU Referendum	4,150,260	25/05/2016	26/11/2016	L
132140	Free childcare when both parents are working. Not just those who are on benefits	133,921	10/06/2016	11/12/2016	C
133540	Let Parliament decide whether or not we remain a member of the European Union	25,729	29/06/2016	30/12/2016	L
133618	Invoke Article 50 of The Lisbon Treaty immediately.	128,238	30/06/2016	31/12/2016	C
133767	Ask the Government not to invoke article 50 of the Lisbon treaty.	27,151	30/06/2016	31/12/2016	L
154593	Not to allow freedom of movement as part of any deal with the EU after Brexit	21,537	01/07/2016	02/01/2017	C
162934	Force child cancer to the forefront of the NHS and government funding schemes	117,183	29/07/2016	30/01/2017	NA
164851	Protect grouse Moors and grouse shooting	25,322	15/08/2016	16/02/2017	M
165905	Shut down the domestic ivory market in the UK.	108,530	30/08/2016	01/03/2017	L
167596	Ban all non-recyclable/non-compostable packaging in the UK	101,172	21/09/2016		L
168127	Demand an end to the pay restraint imposed on agenda for change NHS staff.	106,156	03/10/2016		M
168524	Close all retail on boxing day, retail isn't needed on boxing day!	147,551	14/10/2016		C
168678	Give status to Police Dogs and Horses as 'Police Officers'	127,261	10/10/2016		C
171928	Prevent Donald Trump from making a State Visit to the United Kingdom.	1,862,036	29/11/2016		L
176138	Make it a specific criminal offence to attack any member of NHS Medical Staff.	117,298	22/12/2016		M
178844	Donald Trump should make a State Visit to the United Kingdom.	316,098	30/01/2017		C

often applied partitioning approach is k -means. To start the estimation process, for a given number of classes, initial class centres in n -dimensional space are defined. Then, at each iteration an assignment of data points to the nearest class centre takes place. After this assignment, the class centres are re-calculated (see (Wu et al. [25]) Figures 1 and 2 for an illustration of this process). As the algorithm proceeds these adjustments to the class

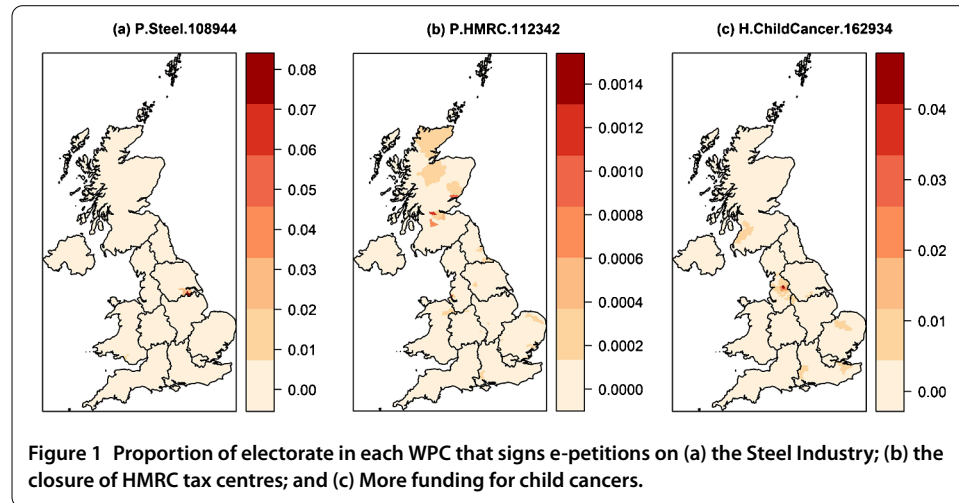


Figure 1 Proportion of electorate in each WPC that signs e-petitions on (a) the Steel Industry; (b) the closure of HMRC tax centres; and (c) More funding for child cancers.

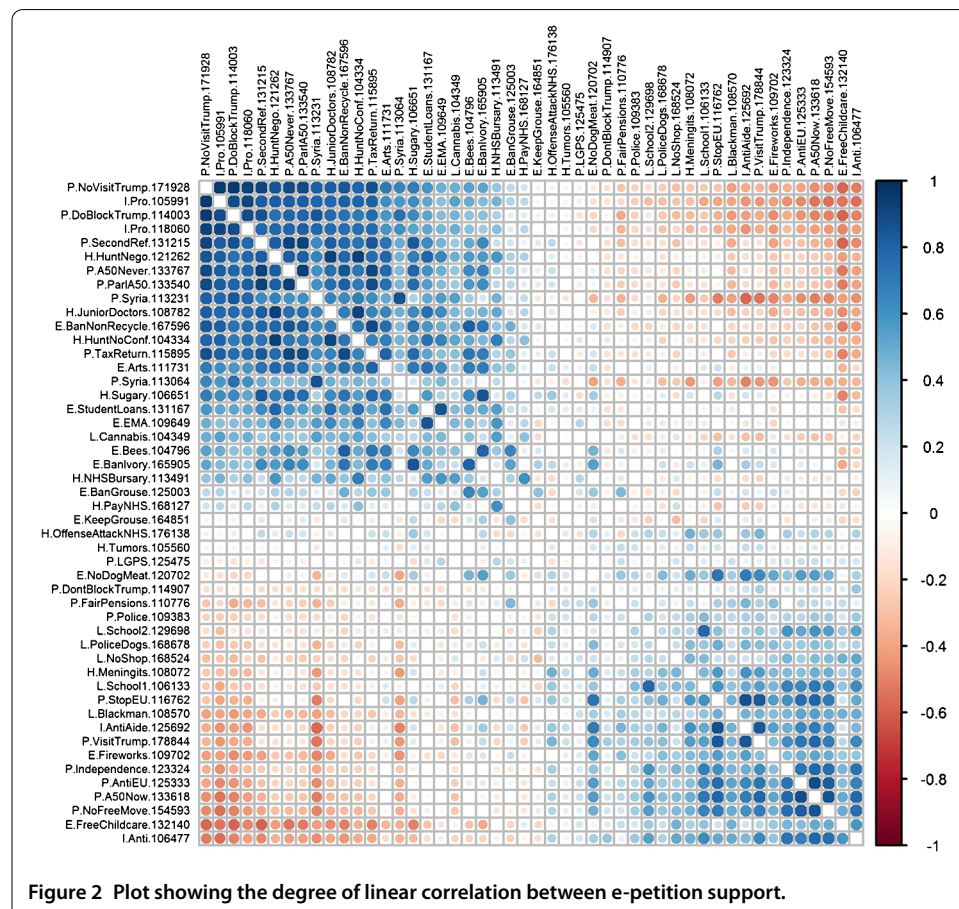


Figure 2 Plot showing the degree of linear correlation between e-petition support.

centres become smaller and the process stops when a level of convergence is achieved. There are no underlying distributional assumptions associated with this technique, it is entirely data-driven, but it is possible to make an ad-hoc assessment of the goodness of fit by measuring the within class sum of square differences from class means. However this measure is only suitable for judgements amongst alternative classification arrangements where the number of classes is the same, since this sum of squares will never increase as the number of classes increases. This approach makes the fitting of classes using k -means a highly subjective exercise. Other drawbacks are that the classes formed by k -means are 'spherical' in nature, having the same dimensions in each direction and also the classification outcome can be sensitive to the initial choice of class centres.

Here an alternative approach of Gaussian finite mixture model estimation via an Expectation-Maximisation algorithm is used. This approach assumes an underlying distributional model for the classes and is therefore able to measure goodness of fit amongst a range of alternative parameterisations of the model (Melnykov and Maitra [26]; Fraley and Raftery [27]). Each class is described by a multivariate Gaussian distribution with an n -dimensional mean vector (the equivalent of class centres) and a variance-covariance structure (a measure of within class spread). As each iteration progresses these distributions 'move' to create tighter clusters around a sub set of data points until a level of convergence is achieved. The goodness of fit of each parameterisation (the number of classes and the 'orientation' of the classes) can be assessed using a Bayesian Information Criteria (BIC) that 'penalises' parameterisations that have greater numbers of classes. Unlike with k -means, it is therefore possible to identify the best fitting parameterisation across a range of classes. Also the shape of the classes is not constrained to be 'spherical' as with k -means, they can be elongated and orientated on certain dimensions. The drawback to this approach is that it relies on an assumption that the shape of the classes can be best represented as Gaussian distributions. A further drawback is that the assignment of areas to classes is soft, each area has a finite probability of belonging to each class. In practice however this information on the fuzziness of the class membership can actually be informative, but if a hard class allocation is required, the decision can be taken to allocate the area to the most probable class.

The classification is performed in R using the `mclust` package (R Core Team [28]; Scrucca et al. [29]). As described above, most classification techniques are sensitive to the initial starting configuration. To help mitigate this the function `randomPairs()` is called to obtain a random hierarchical structure suitable for the initial classification partition.

```
Mclust(petitions_p_elect[,-(1:5)], initialization =
list(hcPairs =
randomPairs(petitions_p_elect[,-(1:5)], seed = 123)))
```

Where `petitions_p_elect` is an R dataframe of WPCs (rows) and e-petitions (columns) that expresses the number of e-petition signatories as a proportion of the WPC electorate. The first five columns are row descriptors for the WPC (namely, the Office for National Statistics code, the WPC name, the name of the MP, their political party and the 2015 electorate). This dataframe is available within the R Workspace supplied as Additional file 1.

3 Data

In this study use is made of the data on the UK Parliament's e-petitions web site (Houses of Parliament [30]). After an e-mail verification, the UK Parliament's Petitions committee collects the signatures for each e-petition and geo-locates them according to the location of the signatory. These data are provided as a JSON formatted file that provides the counts for each e-petition allocated to a country (if outside the UK) or to a WPC (if inside in the UK). The process of requiring a confirmation email and a postcode ensure that the veracity of the data is high when allocated to the area of residence (British Broadcasting Corporation [31]). The count of signatories is also updated frequently for those e-petitions that are live (Yasseri et al. [32]).

The e-petitions cover a variety of topics (see Table 1), with the most popular an e-petition on a second EU referendum which attracted over 4 million signatures. The themes range from health, immigration, education, foreign affairs and animal welfare. Often opposing e-petitions are launched, e.g. one to ban and another to promote a state visit by the United States' President, Donald Trump, or one to keep grouse shooting and another to ban it.

The JSON data are provided as raw counts but the size of the registered electorate in each WPC varies, from Na h-Eileanan an Iar (Western Isles) in Scotland with just under 22k electors up to the Isle of Wight with just under 110k electors. Thus a count of say, 500 signatories represents a greater level of concern in Na h-Eileanan an Iar than the Isle of Wight. To normalise the signatory counts, the number of signatories is divided by the size of the WPC electorate. The higher this proportion the more the politically active residents of the WPC (be they either registered voters or those who take the time to sign the e-petition) agree with the e-petition. The lack of a compatible divisor for those countries with signatories outside the UK means that they are excluded from this study, which just concentrates on the 650 WPCs.

When looking at the distribution of signatories for e-petitions amongst WPCs there are some that are concentrated in just a few WPCs. For example the e-petition number 108944 to '*Save British Steel making. Scunthorpe, Teesside, Port Talbot etc.*' has 36% of its signatories in just two WPCs that are heavily impacted by the proposed closure (Scunthorpe and Brigg and Goole) (Figure 1(a)) and e-petition 112342 which calls on the Government to '*Stop the destructive 'building our future' office closure programme in HMRC*' has 21% of its signatories located in just 10 WPCs (Figure 1(b)). E-petition 162934 to '*Force child cancer to the forefront of the NHS and government funding schemes*' is also concentrated in just a few WPCs (Figure 1(c)). In the context of this article, the geographical dominance of these three e-petitions makes them least informative and encourages the classification algorithm to form classes dominated by these single issues. Solutions including these highly concentrated e-petitions perform less well in terms of generalisability. Users wanting to replicate this methodology would be advised to identify and remove e-petitions which are 'greedy', where few WPCs consume a large proportion of the signatories. Following this reasoning, these three e-petitions are dropped from analysis, which leaves 48 e-petitions.

Figure 2 shows a correlation plot generated by the R package `corrplot` with the e-petitions re-ordered using the first principal component order, 'FPC', option (Wei and Simko [33]). Positive correlations are shown as blue whilst negative correlations are shown as red. Larger circles denote stronger correlations. Two distinct groupings of e-petitions form, where they have a positive relationship with those in the same group and a negative

relationship with those in the other group. These two groups of e-petitions are identified by their grouping label in Table 1. Group GL e-petitions are broadly left leaning, liberal whilst group GC are more conservative in their intent. The middle, group GM are those which don't correlate strongly with either group, or within themselves.

4 Results

The implementation of Gaussian finite mixture models used here estimates a range of possible models using a combination of class configurations (See Figure 2 and Table 3 of Scrucca et al. [29]) and numbers of components/classes and selecting the combination that gives the highest BIC goodness of fit. For these data, the algorithm estimates a model with four classes with Gaussians that are ellipsoidal, variable in size but with equal orientation (VVE). The BIC measure of goodness of fit for this model is 365,576.6 and the BICs for the competing models are shown in Figure 3 (some models cannot be estimated for some cluster class sizes). Since this method is a soft assignment of WPCs to classes, Figure 4 plots the certainty of the assignment to the most probable class. This shows that a very high proportion of the assignments are almost 100% certain and also in Figure 5 this certainty is mapped by WPC, illustrating that there is little or no spatial clustering in this measure.

The centres (in 48 dimensions) of these four classes are established and an index calculated that measures the level of support expressed by these centres relative to the mean

Figure 3 BIC for candidate classifications (see Scrucca et al. [29], Figure 2 and Table 3 for a description of the labels).

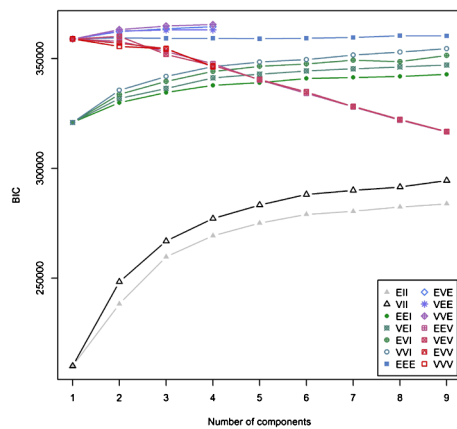
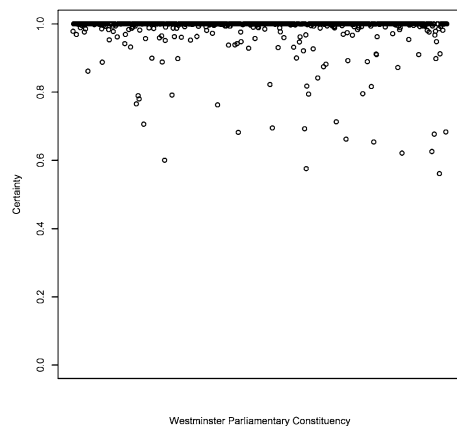
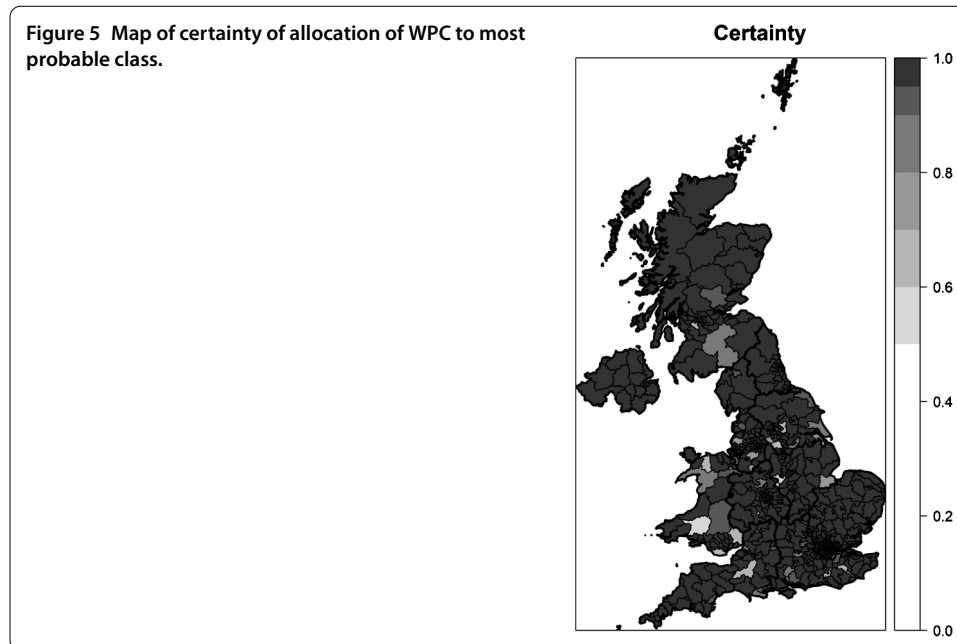


Figure 4 Certainty of allocation of WPC to most probable class.





level of support for that e-petition. This measures how important the e-petition is to the signatories who live in WPCs belonging to the class, an index of 2.0 indicates that this e-petition is twice as likely to be signed by people in these WPCs than people in all WPCs. Table 2 identifies the top 10 e-petitions that have the highest index of support in each class, along with a title and brief description.

4.1 Pen portraits of the classes

To better understand the nature of the classes that have been identified, it is instructive to look at those e-petitions that have a high value for this index in each class and see if there are any common sentiments. If such sentiments can be identified then a short title can be provided that captures this commonality. This has been done with the four classes identified.

- A. *Domestic Liberals (110 WPCs)*. These e-petitions are very anti-BREXIT, asking that article 50, to leave the EU is never triggered, and that if it is, the UK Parliament should make that decision and there should be a second referendum before leaving the EU. Support is also high for issues that the state should intervene on, i.e. a tax on sugary drinks, a ban on non-recyclable packaging, recognition for the arts in education, an end to trading in ivory and measures to protect the bee population. For a class that appears to support liberal causes there is an absence in the top ten of issues that take place outside the UK, e.g. the war in Syria and the status of Donald Trump.
- B. *International Liberals (115 WPCs)*. This class complements to some degree the Domestic Liberal class. Uniquely, it does have high support for those international causes (Syria, Trump) and is also pro-Immigration. There is some overlap with Domestic liberals in having some anti-BREXIT support and concern over environmental issues, but the importance of domestic state intervention is not as well supported.

Table 2 Index of support from each e-petition in each class^b

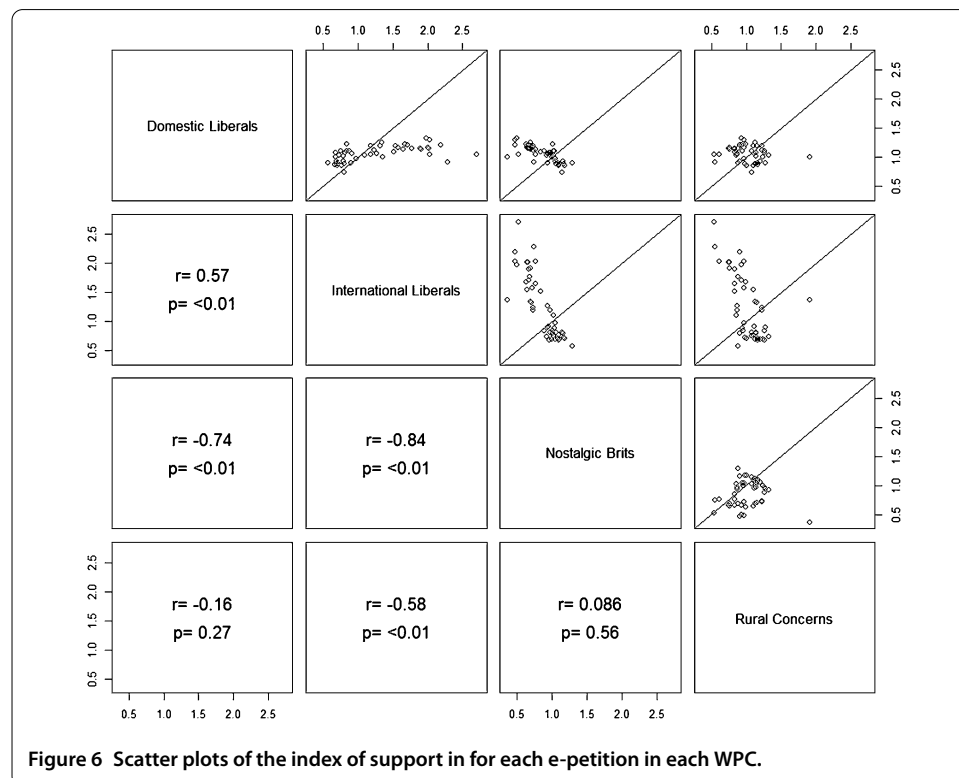
Title		Domestic Liberals		International Liberals		Nostalgic Brits		Rural Concerns		Index	
Rank		Anti-BREXIT, Green, support for state intervention		Pro-immigration, anti-intervention, anti-Trump		No EU, pro-police and army, parents in control not schools		Care about rural/animal issues, UK a priority and should be in control		Index	
1		PA50Never.133767	1.3279	I.Pro.118060	2.7079	P.Police.109383	1.2879	E.KeepGrouse.164851		1.9084	
2		E.BanNonRecycle.167596	1.2922	PSyria.113064	2.2815	E.FreeChildcare.132140	1.1763	P.StopEU.116762		1.3174	
3		H.Sugary.106651	1.2486	PParlAS0133540	2.1909	I.Anti.106477	1.1708	L.PoliceDogs.168678		1.2745	
4		PTaxReturn.115895	1.2287	E.BanNonRecycle.167596	2.0309	L.NoShop.168524	1.1611	H.Tumors.105560		1.2681	
5		H.OffenseAttackNHS.176138	1.2191	PSyria.113231	2.0253	LSchool2.129698	1.1413	I.AntiAide.125692		1.2612	
6		PParlAS0133540	1.2136	I.Pro.105991	2.0166	P.Independence.123324	1.1072	E.Fireworks.109702		1.2378	
7		PSecondRef.131215	1.2057	H.JuniorDoctors.108782	2.0083	LSchool1.106133	1.0999	E.BanIvory.165905		1.2304	
8		E.Bees.104796	1.1940	PA50Never.133767	1.9757	P.NoFreeMove.154593	1.0992	E.BanGrouse.125003		1.2277	
9		E.Arts.111731	1.1925	PDoBlockTrump.114003	1.9116	PA50Now.133618	1.0768	P.AntiEU125333		1.2049	
10		E.BanIvory.165905	1.1887	P.NoVisitTrump.171928	1.8905	L.Blackman.108570	1.0547	P.NoFreeMove.154593		1.1679	

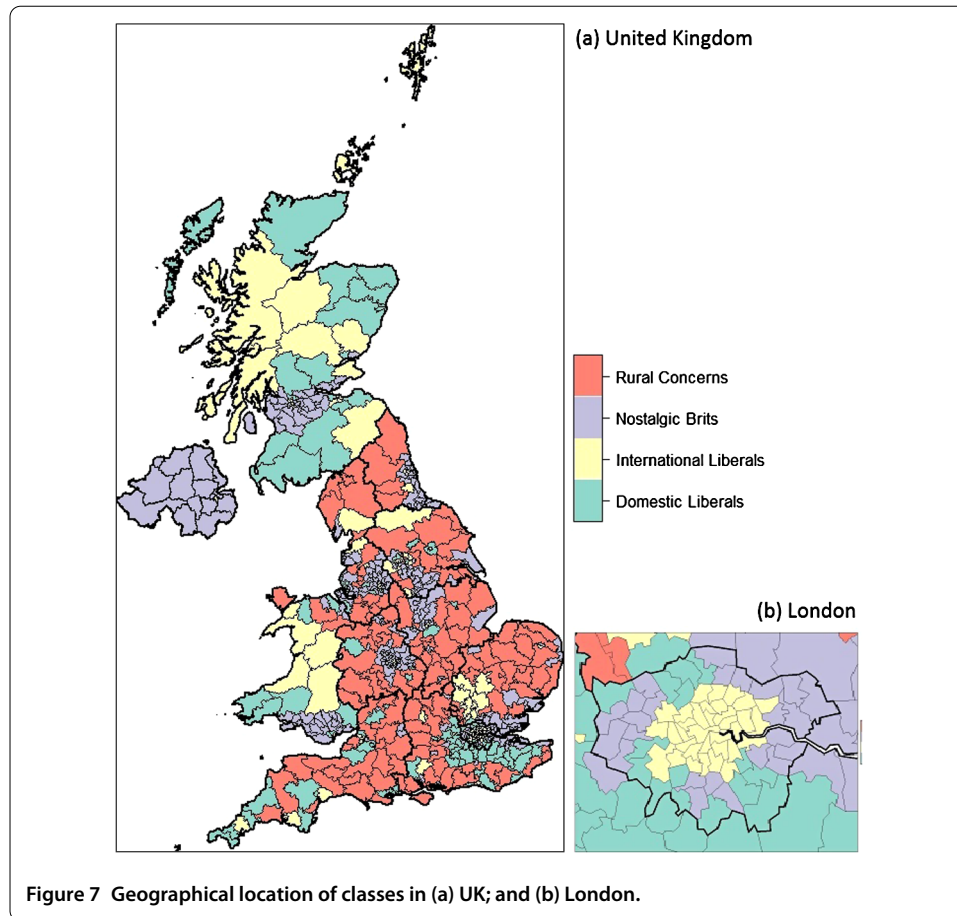
C. *Nostalgic Brits* (276 WPCs). This class gives high support to issues that may reflect a bygone age: Greater support for the police; less immigration; greater parental say over school attendance; better treatment of soldiers and less interference from the EU. These are very conservative issues.

D. *Rural Concerns* (149 WPCs). This is also a conservative class. E-petitions with a pro-BREXIT sentiment receive strong support and there is a desire to spend less money on international aid. There is, however, also support for some distinctive e-petitions. There is strong support for both banning and keeping grouse shooting, having more restrictions on the sale of fireworks (which can traumatised animals), and a ban on the ivory trade.

Figure 6 shows scatter plots of these support indices within each class. Only the Domestic Liberals and International Liberal classes show possible support for each other (with a 'middling' $r = 0.57$). The International Liberals and Nostalgic Brits are have the strongest relationship, all be it a negative one ($r = -0.84$). Rural Concerns do not significantly correlate with either Domestic Liberals or Nostalgic Brits and only correlate negatively with International Liberals ($r = -0.58$).

A map showing the spatial arrangement of these classes for the UK (Figure 7(a)) and London (Figure 7(b)) provides further context to the classes. The support for the Rural Concern class is evident in the larger rural WPCs of England and Wales, but there are none of these classes in Scotland or Northern Ireland. In Scotland and Northern Ireland, many issues around rural concerns have devolved responsibility away from the Westminster Parliament, to their respective Parliament and Assembly (e.g. agriculture, environment and land use planning) which would motivate few from Scotland or Northern Ireland to participate in such e-petitions. Liberals (both Domestic and International) are concentrated in





London and the surrounding Home Counties, with a significant number in the more rural parts of Scotland. Nostalgic Brits are to be found in the urban areas of England, Scotland and the Welsh Valleys.

5 Discussion

In this article we have used novel data in a way that allows new insights and provides empirical understanding of a complex political system. Use has been made of these data, namely e-petitions, to identify a set of classes that typify WPCs according to the political sentiment expressed by these e-petitions. Using the technique of Gaussian finite mixture model classification, four classes are identified and shown to be meaningful and coherent in terms of the aspects of the political discourse that they cover. Two liberal classes are identified that are concentrated in and around London, one conservative class to be found in the urban centres and a distinct class concerned with rural issues. Elsewhere, looking at individual survey data, Sanders [34] also identified four classes in the UK electorate, two of his classes are liberal (a left and a centre-right variant) that were in tune with a less-harsh foreign policy but approve of human rights interventions, and two authoritarian classes (with centre and right variants) that disapprove of the EU and have a negative assessment of the impact of immigration.

Further insight into the utility of these classes can be gained by seeing how these classes map onto other outcomes. Perhaps the most obvious politically, is the outcome of the

Table 3 Correspondence between 2017 General election results and class membership

Party	Domestic Liberals	International Liberals	Nostalgic Brits	Rural Concerns
Conservative	71	23	91	133
Labour	29	72	147	14
Scottish National Party	3	12	20	0
Liberal Democrats	6	3	1	2
Democratic Unionist Party	0	1	9	0
Sinn Féin	0	0	7	0
Plaid Cymru	1	3	0	0
Green	0	1	0	0
Independent	0	0	1	0
Total	110	115	276	149

Table 4 Correspondence between OAC and class membership

OAC	Domestic Liberals	International Liberals	Nostalgic Brits	Rural Concerns
Constrained City Dwellers	17%	16%	56%	11%
Cosmopolitans	20%	66%	10%	3%
Ethnicity Central	9%	74%	16%	1%
Hard-Pressed Living	15%	8%	57%	20%
Multicultural Metropolitans	15%	29%	51%	5%
Rural Residents	20%	14%	17%	48%
Suburbanites	19%	8%	48%	25%
Urbanites	24%	14%	36%	25%
Total	18%	20%	42%	21%

2017 General election for each WPC. Table 3 cross-tabulates the outcome of the 2017 General election with the class membership of the WPC. The Conservative party enjoys most support in the Rural Concerns class with little support in the International Liberals class. The Labour party is strongly represented in both the International Liberals and the Nostalgic Brits classes, which are perhaps the two most opposite classes, which suggest a broad range of issues that underlie the party's support (Goodwin and Heath [35]). Looking at voting intention, Sanders [34] also noted a large proportion of those intending to vote Labour categorised in the Centre-Right and Authoritarian Centre classes. The large support for the Scottish National Party in Scotland at the 2017 General election gives it a broad representation in three of the four classes.

A second more methodological comparison is with the official Output Area Classification (OAC) (Gale et al. [4]) which classifies small areas, of typically 125 households, using 2011 Census data. Table 4 show the percentage of the OAC classes that are also in each of the WPC classes. The majority of Constrained City Dwellers (56%), Hard-Pressed-Living (57%) and Multicultural Metropolitans (51%) OACs are to be found our Nostalgic Brits classes. The strong representation of a class termed Multicultural Metropolitan in our Nostalgic Brits class is a surprise. Less surprising is that OAC classes Cosmopolitans (66%) and Ethnicity Central (74%) are most likely to be in our International Liberals class. Finally, just under 50% of the Rural Residents OAC are in our Rural Concerns class (48%), which is to be expected.

A number of issues arise around the conduct of this study. Firstly, the signatories to these e-petitions are probably not a representative sample of the population in a WPC, they are those who feel some motivation to express their views on the topic of the e-petition. In terms of electoral outcome however, capturing the opinions of these people is important since, if they can be motivated to visit a web site, enter their details and then click on a

confirmatory email, they may be more likely to vote in an election. A second issue is that to progress the study it is necessary to normalise the number of signatories by the size of the electorate in each WPCs, however the pool of signatories is not necessarily the same as those registered to vote. Whilst anyone can sign an e-petition only those individuals aged over 18 and meeting nationality criteria can be registered to vote. However there may be some commonality between the two groups in that they are both politically engaged samples, either through their decision to sign an e-petition or register to vote. Thirdly, we do not know the motivation for the individual to sign an e-petition. For example, the two e-petitions to not intervene in the Syrian Conflict may be signed by the more liberal who were 'scared' by the recent interventions in Iraq and Afghanistan and have no wish to repeat that outcome. They may also be signed by those who are more conservative and think that British lives should not be deployed '*...because of a quarrel in a far away country between people of whom we know nothing*' (Chamberlain [36]). Given that these two e-petitions sit firmly in our International Liberals class, the first motivation appears to be the most likely. The fourth point is that the subject matters of these e-petitions are quite repetitious and polarising, with little representation of more nuanced middle ground positions on certain themes. There are some themes that are not even covered, e.g. transport, defence, the environment or social issues. Finally, there is the issue of apparent contradictions, exemplified by the outcome that the e-petition to ban grouse shooting and the separate e-petition to keep grouse shooting are strongly supported in the same class, Rural Concerns. This is not really a conundrum since in the local community where grouse shooting occurs, there may be strongly held views on both sides, thus both e-petitions attracting a lot of support on this locally contentious issue.

This approach to classifying areas can be applied more widely. As data from new e-petitions becomes available this exercise may be repeated to both test the stability of these classifications and reveal changing trends (Singleton et al. [37]). Thus it will be possible to get a contemporary picture of political sentiment around the country, removing the need to be reliant on out of date decennial Census data, geographically sparse and incomplete opinion poll data, or sporadic household survey data. Also if these new e-petitions capture a new dimension of the political discourse (e.g. transportation or international defence) then additional insight may be gained when they are incorporated. This classification technique also can be applied to a new political geography to inform policy makers and campaigners about the nature of the new geography. The UK is currently going through a process of re-drawing the WPC boundaries to reflect population changes and a mandated reduction in the number of MPs from 650 to 600 (Johnston et al. [38]; Johnston et al. [39]). To work with this new geography, all that is required is that the signatories be geo-relocated to the new constituencies and the classification carried out on the new 600 WPCs. The types of class that will emerge from such an exercise may be different to the four identified here but may provide some pointers as the political sentiment in these new WPCs. Also, rather than a re-shaping of an existing geography, a separate political geography may be of interest, possibly the constituencies in the devolved Scottish Parliament and Welsh and Northern Ireland Assemblies or electoral wards in Local Municipalities (Bochel and Bochel [18]), and classes can be identified using either the UK Parliaments e-petition data or data from be-spoke e-petition systems.

In this study we have shown that there is meaningful information in the responses to such e-petitions and that this can have an impact that shapes perceptions of the politi-

cal debate in each constituency. Sensible and savvy political parties and local politicians should exploit these data more fully to gauge and potentially tailor their message for the electorate (Hough [19]). Internationally, given the increased use of such systems in legislators around the world (Directorate-General for Internal Policies [40]), the methodology adopted here can be applied elsewhere.

Additional material

Additional file 1: R Workspace. (zip)

Acknowledgements

The authors would like to acknowledge the UK Parliaments Petitions committee for the open access to their data. They would also like to thank the reviewers who provided comments on an earlier draft of this article.

Funding

The work was funded by the ESRC Consumer Data Research Centre, Grant No ES/L011891/1.

Abbreviations

BREXIT, Britain exiting from the EU; DUP, Democratic Unionist Party; EM, Expectation-Maximisation; EU, European Union; HMRC, Her Majesty's Revenue and Customs; LD, Liberal Democrat; MP, Member of Parliament; NHS, National Health Service; OAC, Output Area Classification; PC, Plaid Cymru; SF, Sinn Féin; SNP, Scottish National Party; UK, United Kingdom of Great Britain and Northern Ireland; VVE, Ellipsoidal, variable volume, variable shape, equal orientation; WPC, Westminster Parliamentary constituencies.

Availability of data and materials

The R Workspace containing the data frame of raw counts and proportions of the electorate are available as Additional file 1.

Ethics approval and consent to participate

The data used in this study has been downloaded from the official UK Parliaments web site. It contains only counts in aggregate and no personal data is identifiable.

Competing interests

There are no competing interests.

Consent for publication

Not applicable.

Authors' contributions

The data preparation and analysis was conducted by Stephen Clark. The first draft was written by Stephen Clark. Michelle Morris and Nik Lomax made suggested changes and comments for subsequent drafts. All authors read and approved the final manuscript.

Author details

¹Consumer Data Research Centre, Leeds Institute for Data Analytics and School of Geography, University of Leeds, Leeds, LS12 9JT, UK. ²School of Geography and Leeds Institute for Data Analytics, University of Leeds, Leeds, LS12 9JT, UK.

³Leeds Institute for Data Analytics and School of Medicine, University of Leeds, Leeds, LS12 9JT, UK.

Endnotes

- ^a The Group identifies e-petitions that are broadly Left leaning (L), Conservative (C) or Middle of the road (M). NA are petitions not included in the analysis.
- ^b The initial letter indicates the general theme of the e-petition: Immigration, Education, Health, Politics, Libertarian; followed by a brief description; and finally the e-petition id, as given in Table 1.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 5 April 2017 Accepted: 20 July 2017 Published online: 08 August 2017

References

1. Everitt B, Landau S, Leese M (2001) Cluster analysis. 4th edn. Arnold, London
2. Berry MJA, Linoff G (1997) Data mining techniques for marketing, sales and customer support. Wiley, New York
3. Office for National Statistics (2015) Methodology note for the 2011 area classification for output areas. <https://www.ons.gov.uk/methodology/geography/geographicalproducts/areaclassifications/2011areaclassifications>

4. Gale CG, Singleton AD, Bates AG, Longley PA (2016) Creating the 2011 area classification for output areas (2011 OAC). *J Spat Inf Sci* 12
5. Office for National Statistics (2015) Methodology note for the 2011 area classification for local authorities. <https://www.ons.gov.uk/methodology/geography/geographicalproducts/areaclassifications/2011areaclassifications>
6. Longley PA, Webber D, Li C (2008) The UK geography of the e-society a national classification. *Environ Plan A* 40(2):362-382
7. CallCredit (2017) CAMEO UK. <https://www.cameodynamic.com/>
8. CACI (2017) ACORN family: social scene. <http://www.caci.co.uk/sites/default/files/resources/SocialSceneUG.pdf>
9. CACI (2006) HealthACORN user guide. <http://www.yhpho.org.uk/resource/item.aspx?RID=10140>
10. CACI (2017) ACORN family: wellbeing. http://www.caci.co.uk/sites/default/files/resources/Wellbeing_Acorn_User_Guide.pdf
11. CACI (2017) What is acorn? <http://acorn.caci.co.uk/>
12. Experian (2017) Experian Moosaic. <http://www.experian.co.uk/marketing-services/products/mosaic-uk.html>
13. Classification of Workplace Zones for England and Wales (2017) A classification of workplace zones from the 2011 census for England and Wales. <http://cowz.geodata.soton.ac.uk/>
14. Greater London authority (2017) London output area classification. <https://data.london.gov.uk/dataset/london-area-classification>
15. Wright S (2015) E-petitions. In: Coleman S, Freelon D (eds) *Handbook of digital politics*. Edward Elgar, Cheltenham
16. Miller L (2008) e-petitions at Westminster: the way forward for democracy? *Parliam Aff* 62(1):162-177
17. Briassoulis H (2010) Online petitions: new tools of secondary analysis? *Qual Res* 10(6):715-727
18. Bochel C, Bochel H (2016) 'Reaching in'? The potential for e-petitions in local government in the United Kingdom. *Inf Commun Soc* 20(5):683-699
19. Hough R (2012) Do legislative petitions systems enhance the relationship between Parliament and citizen? *J Legis Stud* 18(3-4):479-495
20. Dumas C, Harrison TM, Hagen L, Zhao X (2017) What do the people think?: E-petitioning and policy decision making. *Beyond bureaucracy*. Springer, Berlin, pp 187-207
21. Huang S-W, Suh MM, Hill BM, Hsieh G (2015) How activists are both born and made: an analysis of users on change.org. In: *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. ACM, New York, pp 211-220
22. Puschmann C, Bastos MT, Schmidt J-H (2016) Birds of a feather petition together? Characterizing e-petitioning through the lens of platform data. *Inf Commun Soc* 20(2):203-220
23. Hagen L, Harrison TM, Uzuner Ö, Fake T, Lamanna D, Kotfila C (2015) Introducing textual analysis tools for policy informatics: a case study of e-petitions. In: *Proceedings of the 16th annual international conference on digital government research*. ACM, New York, pp 10-19
24. Hanretty C (2017) Areal interpolation and the UK's referendum on EU membership. *Journal of Elections, Public Opinion and Parties*, 1-18
25. Wu X, Kumar V, Ross Quinlan J, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Yu PS, Zhou Z-H, Steinbach M, Hand DJ, Steinberg D (2007) Top 10 algorithms in data mining. *Knowl Inf Syst* 14(1):1-37
26. Melnykov V, Maitra R (2010) Finite mixture models and model-based clustering. *Stat Surv* 4(0):80-116
27. Fraley C, Raftery AE (2002) Model-based clustering, discriminant analysis, and density estimation. *J Am Stat Assoc* 97(458):611-631
28. R: a language and environment for statistical computing (2016) R foundation for statistical computing, Vienna, Austria
29. Scrucca L, Fop M, Murphy TB, Raftery AE (2016) mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *R J* 8(1):289-317
30. Houses of Parliament (2017) Petitions: UK Government and Parliament. <https://petition.parliament.uk/>
31. British Broadcasting Corporation (2017) Reality check: can we believe petition signature numbers? <http://www.bbc.co.uk/news/uk-politics-38828663>
32. Yasseri T, Hale SA, Margetts H (2013) Modeling the rise in internet-based petitions. *arXiv:1308.0239*
33. Wei T, Simko V (2016) R package 'corrplot': visualisation of a correlation matrix (Version 0.82). <https://github.com/taiyun/corrplot>
34. Sanders D (2016) The UK's changing party system: the prospects for a party realignment at Westminster. Unpublished
35. Goodwin M, Heath O (2016) The 2016 referendum, Brexit and the left behind: an aggregate-level Analysis of the result. *Polit Q* 87(3):323-332
36. Chamberlain N (1938) Quoted in 'Prime Minister on the issues'. *The Times*. 28th September
37. Singleton A, Pavlis M, Longley PA (2016) The stability of geodemographic cluster assignments over an intercensal period. *J Geogr Syst* 18(2):97-123
38. Johnston R, Pattie C, Manley D (2017) Britain's changed electoral map in and beyond 2015: the importance of geography. *Geogr J* 183(1):58-70
39. Johnston R, Pattie C, Rossiter D (2013) Manipulating territories: British political parties and new parliamentary constituencies. *Territ Politics Gov* 1(2):223-245
40. Directorate-General for Internal Policies (2015) The Right to Petition EU. European Parliament. [http://www.europarl.europa.eu/RegData/etudes/STUD/2015/519223/IPOL_STU\(2015\)519223_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2015/519223/IPOL_STU(2015)519223_EN.pdf)