Appendix A: Empirical Appendix

August 20, 2015

1. SENSITIVITY ANALYSIS

I begin this appendix by conducting sensitivity analysis to confirm that the main results in the paper are robust to excluding or including different types of constituencies.

First, I excluded constituencies with only minor changes from the analysis reported in the main text to prevent an artificial 'deflation' of the RMSE and MAE. I determined which constituencies were only subject to 'minor' reorganisations by first calculating the 'effective number of (old) constituencies' (ENC) that constitute a new constituency. I use Laakso and Taagepera's 1979 to construct this measure based on the interpolated electorate. I exclude constituencies with a value below 1.2, which roughly corresponds to a new constituency for which approximately 90% of its population came from a single old constituency. This threshold is of course arbitrary and a visual inspection presented in Figure 1 suggests there is some correspondence between the ENC and MAE (as perhaps to be expected). However, whilst the smoothed local regression curves show an upward trend in some polities, it is worth noting that the magnitude is generally slight, and the predicted error is not especially large even at high levels of ENC.

Figure 2 further confirms the robustness of the results by varying the threshold of exclusion and reporting the MAE, pooling each country's results together. Whilst there is unsurprisingly an upwards slope, the largest jump (in all countries except the United States and New Zealand) occurs before the threshold used in the main paper. Even if a higher threshold were used—with a corresponding decrease in the number of observations, the results are still quite similar to the ones shown in the main paper; only rarely does the MAE exceed 3%.

Next, I examine the rate of correct prediction reported in the main paper; those results only included marginal constituencies, i.e. those where the majority of the winning party was below 10%. Figure 3 shows the rate of 'correct prediction' when this threshold is varied. As expected, including 'safer' seats causes the rate of correct prediction to rise—levelling off at around 90% for both methods and all countries when all seats are included. Interestingly, the DI method shows consistently better results insofar as it is less likely to incorrectly classify extremely marginal races.

Figure 1: Scatterplot of ENC and MAE

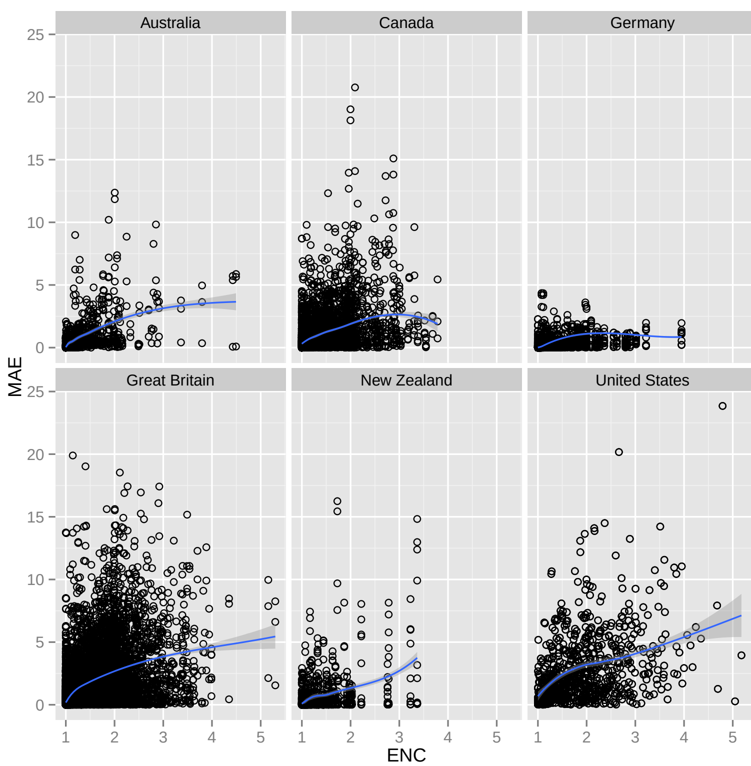


Figure 2: Sensitivity Analysis of Exclusion Threshold

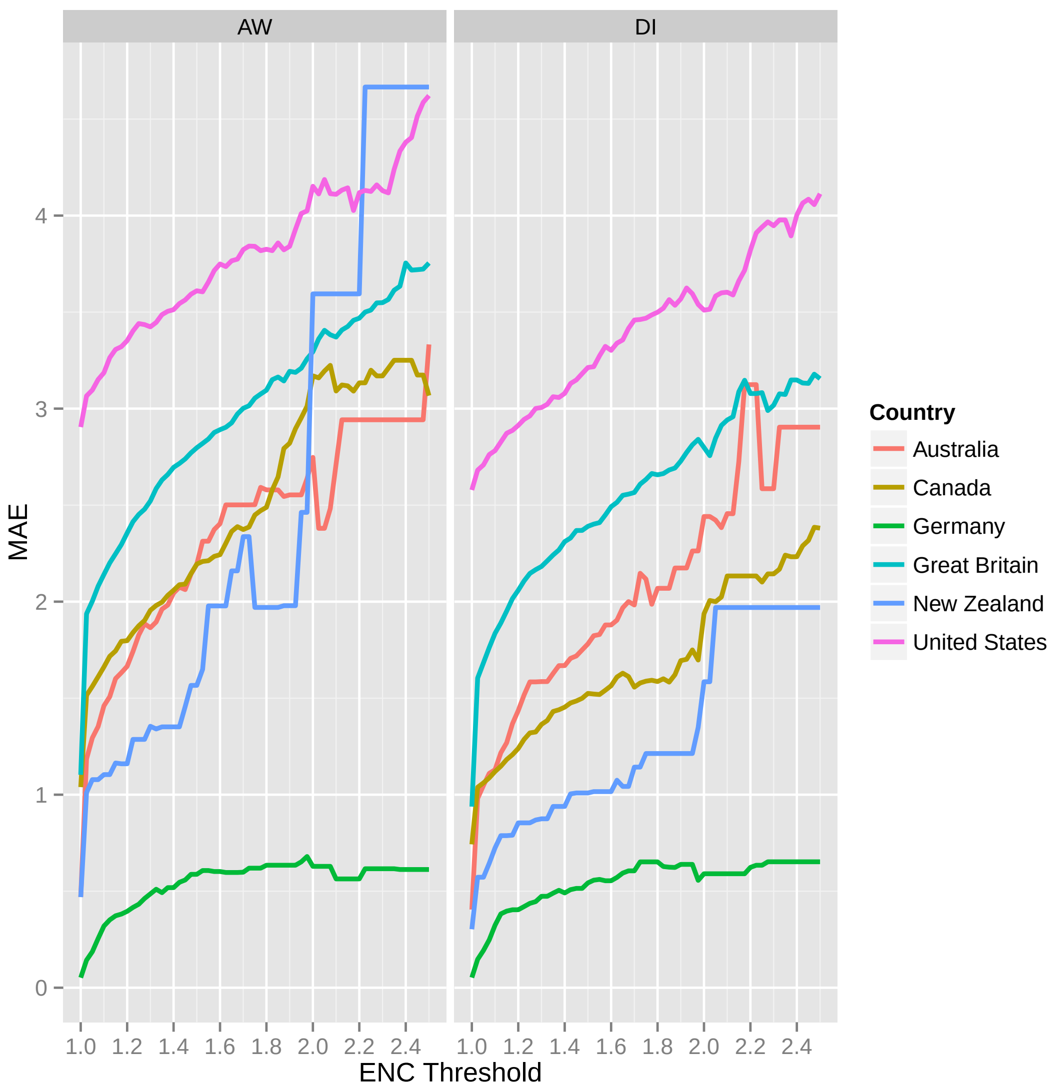
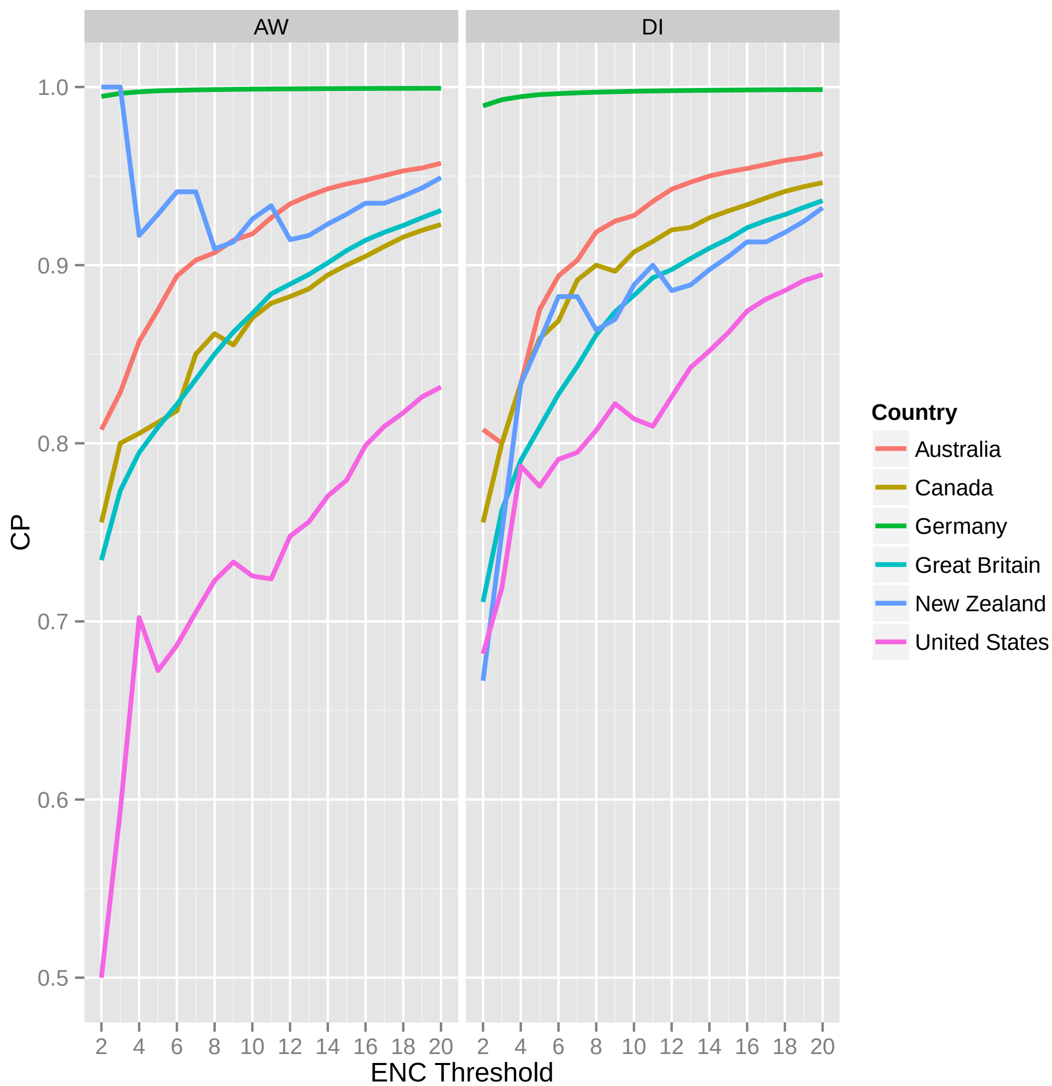


Figure 3: Sensitivity Analysis of Correct Prediction



2. DATA SOURCES AND IMPLEMENTATION DETAILS

This section details the sources of data for each country and details the decisions made to collapse the electorate data in a spatially usable form. In general, the data used in the analysis will not correspond exactly to the full election results insofar as certain types of information (e.g. advance voting, voting that took place overseas, etc.) must be systematically excluded. In the case where the data will not map exactly onto the official results, I use the subset of available data to generate both the 'old' and 'new' exact constituency totals, e.g. by spatial assignment or matching of sub-constituency identifiers. This strategy also implies that if one were to wish to use the interpolations as representing 'actual' electoral data, I would suggest using the interpolated method on the *actual* constituency-level vote totals which are readily available.

I also provide disaggregated RMSE and MAE for each party at each year in each country. This confirms that the error shown at the 'individual' party level is similar to that shown in the main analysis.

2.1. *Australia*

Electoral Shapefiles The electoral shapefiles come from the Australian Election Commission (AEC)[1] and the Australian Bureau of Statistics.[2] Australian states do not have a unified timetable for boundary changes and thus redistributions occur only in a subset of states between any two given elections.[3]

Between the 2007 and 2010 elections, redistributions occurred in five states: New South Wales, Queensland, Tasmania, Northern Territory, and Western Australia. Between the 2010 and 2013 elections, redistributions occurred in two states: Victoria and South Australia. The shapefiles reflect this.

Electoral Returns The AEC provides polling place specific electoral results for modern elections; I use the first-preference votes assigned to the major parties, i.e. the Coalition,[4] Labour, and the Greens. Crucially, the AEC also provides the coordinates of most polling places thereby permitting the assignment the results to constituencies in the old and new boundaries. There is two major and one minor caveat to this assignment; the first major caveat first notes that individuals are permitted to vote at any polling place inside their registered constituency. Thus, it cannot be known for certain whether a voter at polling place X would have been 'transferred' in the counterfactual election under the new boundaries. However, assuming that voters would tend to vote at the location nearest to them, this likely provides a reasonable estimate in the absence of better data.

The second major caveat notes that certain geographic locations (e.g. Sydney Town Hall) appear to host multiple polling places for *different* constituencies. For example, voters from many constituencies seem to have been able to vote at that location. Spatially, this would represent itself as a large cluster of points occupying the same geographic location (Sydney Town Hall) but being allocated in the 'actual' results to different constituencies. It is an intractable problem for the scope of this paper how to deal with those voters so I adopt the simplest solution and simply aggregate all points based on their spatial location, regardless of their 'actual' seat. Thus, my vote totals for Sydney are likely systematically different from the 'actual' result in Sydney because of the inclusion of all Sydney Town Hall polling booths in Sydney. A more minor caveat notes that this data must also exclude advance voting as well as other polling places for which no vote data is provided, e.g. some prisons. A related implication of the above discussion is that there is no polling both level 'electorate'; thus, when calculating the ENC, I rely on the interpolated turnout.

1. Specifically, `http://www.aec.gov.au/electorates/gis/index.htm` accessed 1 March 2015.
2. Specifically, `http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2923.0.30.0012006?OpenDocument` accessed 1 March 2015.
3. `http://www.aec.gov.au/Electorates/Redistributions/Dates.htm` accessed 1 March 2015.
4. I combine the totals of the various constituent parties together.

Population Data The population data came from the 2006 Australian census.[5] I used the TableBuilder Basic to get the ordinarily resident population at the level of Statistical Local Authority (SLAs).[6] All SLAs in the shapefile (1418) were exactly matched to the population data; population recorded as 'No Usual Address' or 'Offshore Areas & Migratory' could not be assigned.

Implementation Details The results reported in the main paper used rasters of resolution 6,000 by 6,000. This resulted in 99.7% of SLAs being successfully rasterised.

Table 1: Disaggregated Australian Results

Year	Party	Method	N	RMSE	MAE	Middle 90% of AE
2010	Coalition	DI	52	2.782	1.942	[0.096, 6.374]
	Labour	DI	52	2.830	2.132	[0.111, 5.789]
	Green	DI	52	1.011	0.577	[0.031, 2.237]
	Coalition	AW	48	3.188	2.250	[0.146, 5.345]
	Labour	AW	48	3.239	2.332	[0.275, 6.165]
	Green	AW	48	1.088	0.608	[0.071, 1.272]
2013	Coalition	DI	15	1.662	1.326	[0.128, 3.440]
	Labour	DI	15	1.443	1.200	[0.044, 3.195]
	Green	DI	15	0.710	0.591	[0.064, 1.239]
	Coalition	AW	16	2.413	1.863	[0.270, 6.043]
	Labour	AW	16	2.757	1.879	[0.020, 5.951]
	Green	AW	16	0.852	0.669	[0.121, 1.577]

Disaggregated Results

5. `http://abs.gov.au/websitedbs/censushome.nsf/home/historicaldata2006` accessed 1 March 2015.

6. Shapefiles for these units are located at `http://www.abs.gov.au/ausstats/abs@.nsf/DetailsPage/1259.0.30.0022006` access 1 March 2015. One could use more fine-grained population data if desired, however, a visual inspection suggests that boundary changes do not subdivide many SLAs and thus it is sufficiently granular.

2.2. *Canada*

Electoral Shapefiles Canadian shapefiles come from GeoGratis;[7] this contains the boundaries of polling stations, the point-locations of additional polling stations, and the shapefiles of the relevant federal election districts (2000, 2004, 2011, and 2013). Some polling stations are defined as ‘points’, e.g. single apartment blocks (400-class) or mobile polling stations (500-class), and are assigned to the (areal) polling station that contains them. Most mobile polling stations appear in at least two (areal) polling stations and thus cannot be assigned. Group 1 and Group 2 of ‘Special Voting Rules’ are not able to be assigned. Advance voters (600-class) are not assigned.[8] Small alterations to some shapefiles were required to ensure valid shapes and corresponding extents.

The assignment of polling stations to old and new constituency boundaries was complicated by the fact that not all polling districts perfectly aligned with the constituency boundaries. This may be due to slight inconsistencies in the creation of the shapefiles or to changes in the polling station boundaries between elections. Most discrepancies are slight and I deal with this by assigning polling stations to exactly one constituency (for each set of boundaries) based on the location of a central point of the polling station.[9]

Electoral Returns Elections Canada provided the poll-by-poll electoral data.[10] Formatting (performed in Python) was required to get the poll-by-poll data in a form to be merged with the shapefiles. The code used is available upon request. A small number of polling stations were unable to be merged; the vast majority of stations that were declared ‘void’ or for which there were no valid votes cast. The parties included are the Liberals, Conservatives (Canadian Alliance and Progressive Conservative in the 2000 election), Bloc Quebecois, and the National Democratic Party.

Population Data The population data used was simply the electorate reported in each polling station. Census data on a sufficiently detailed scale would likely perform similarly.

Implementation Details The results reported in the main paper used rasters of resolution 20,000 by 20,000. This resulted in 94% of polling station shapefiles being successfully rasterised; the remainder has their population data added to the closest available pixel. This under-performance (despite the much larger resolution than other interpolations) is likely due to the fact that some very large polling stations (e.g. those in the Northwest Territories and Nunavut) take up a large proportion of the raster. Appendix C shows that the estimate used here is very cautious over-estimate, as the correlation between the projections used in the main paper and ones using a much smaller raster are extremely high.

Disaggregated Results

7. `http://geogratis.gc.ca/` accessed 1 March 2015.

8. Such mapping is likely possible in some cases. To ensure constituency, I simply exclude all the relevant instances.

9. The command used is `gPointOnSurface` from the R package **rgeos**. Bivand and Rundel 2014. The point generated by this command is guaranteed to lie *inside* the polygon, unlike the ‘centroid’ which may be outside of the polygon for a sufficiently irregular shape. Comparing the two commands in R suggests that `gPointOnSurface` exactly recovers the correct assignment of polling stations for the old boundaries (where the ‘correct’ constituency is known) whilst assignment based on the centroid fails to do so. A more detailed description of the algorithm `gPointOnSurface` uses (or one highly similar) may be found here: `http://geos.osgeo.org/doxygen/classgeos_1_1algorithm_1_1InteriorPointArea.html` accessed 1 March 2015.

10. `http://www.elections.ca/` accessed 1 March 2015.

Table 2: Disaggregated Canadian Results

Year	Party	Method	N	RMSE	MAE	Middle 90% of AE
2004	Bloc Quebecois	DI	39	2.529	1.639	[0.158, 5.373]
	Canadian Alliance	DI	143	1.638	1.043	[0.049, 3.854]
	Liberal	DI	143	2.209	1.522	[0.116, 4.777]
	NDP	DI	143	0.810	0.434	[0.014, 1.560]
	Progressive Conservative	DI	143	0.923	0.587	[0.038, 2.061]
	Bloc Quebecois	AW	33	3.401	2.374	[0.279, 7.890]
	Canadian Alliance	AW	139	2.346	1.541	[0.072, 4.968]
	Liberal	AW	139	3.029	2.011	[0.105, 7.005]
	NDP	AW	139	1.620	0.865	[0.027, 3.420]
	Progressive Conservative	AW	139	2.062	1.144	[0.055, 3.922]
2013	Bloc Quebecois	DI	35	1.547	1.219	[0.107, 3.333]
	Conservative	DI	131	3.140	2.183	[0.119, 6.470]
	Liberal	DI	131	1.806	1.269	[0.096, 3.828]
	NDP	DI	131	2.421	1.652	[0.089, 4.419]
	Bloc Quebecois	AW	33	2.454	1.724	[0.057, 5.401]
	Conservative	AW	137	4.326	2.895	[0.089, 7.859]
	Liberal	AW	137	2.857	1.797	[0.081, 5.576]
	NDP	AW	137	3.345	2.234	[0.128, 5.850]

2.3. *Germany*

Electoral Shapefiles The shapefiles on German constituencies were downloaded from the *Bundeswahlleiter* (Federal Election Officer).[11]

Electoral Returns Constituency-level electoral returns were downloaded from the *Bundeswahlleiter*. That organisation also provides 'projections' as to what the immediately preceding election would have looked like on the new boundaries based on its detailed analysis of sub-constituency level returns.[12] I compare my projections against these results as they approximate the 'first-best' solution described in the main text.

The parties included are the CDU/CSU (treated as one entity), SPD, FDP, the Greens, and Die Linke. I include both erststimmen and zweitstimmen (constituency and list votes, respectively, in the MMP system).

Population Data The population data used came from the most recent German census (*ZENSUS 2011*). It provided shapefiles of municipalities (*gemeinden*).[13] I used the population data on the same level of disaggregation from the Federal Statistical Office (*Destatis*).[14] After cleaning, the population sub-units for which no data was available were unincorporated municipalities, presumably with zero or very low population.

Implementation Details The results reported in the main paper used rasters of resolution 6,000 by 6,000. This resulted in over 99% of municipalities being rasterised.

Disaggregated Results

11. `http://www.bundeswahlleiter.de/` accessed 1 March 2015.

12. An explanation, in German, is provided at the following location: `http://www.bundeswahlleiter.de/de/bundestagswahlen/BTW_BUND_13/ergebnis_2009_umgerechnet/` accessed 1 March 2015.

13. `https://www.zensus2011.de/EN/Media/Background_material/Background_material_node.html` accessed 1 March 2015. Some slight modifications because of administrative re-organisations were required to enable merging with the *Bundeswahlleiter*'s data.

14. `https://www.destatis.de/DE/ZahlenFakten/LaenderRegionen/Regionales/Gemeindeverzeichnis/Administrativ/AdministrativeUebersicht.html` access 1 March 2015.

Table 3: Disaggregated German Results

Year	Party	Method	N	RMSE	MAE	Middle 90% of AE
2005	CDU/CSU (Erststimmen)	DI	15	0.759	0.664	[0.046, 1.488]
	SPD (Erststimmen)	DI	15	0.675	0.534	[0.157, 1.338]
	Grune (Erststimmen)	DI	14	0.324	0.255	[0.027, 0.669]
	FDP (Erststimmen)	DI	15	0.164	0.122	[0.018, 0.290]
	Die Linke (Erststimmen)	DI	13	0.301	0.207	[0.010, 0.603]
	CDU/CSU (zweitstimmen)	DI	15	0.698	0.622	[0.106, 1.347]
	SPD (zweitstimmen)	DI	15	0.322	0.281	[0.052, 0.508]
	Grune (zweitstimmen)	DI	15	0.508	0.376	[0.059, 1.029]
	FDP (zweitstimmen)	DI	15	0.154	0.109	[0.015, 0.333]
	Die Linke (zweitstimmen)	DI	15	0.182	0.123	[0.006, 0.459]
	CDU/CSU (Erststimmen)	AW	14	0.851	0.583	[0.058, 1.914]
	SPD (Erststimmen)	AW	14	1.012	0.668	[0.124, 2.159]
	Grune (Erststimmen)	AW	13	0.416	0.264	[0.012, 0.864]
	FDP (Erststimmen)	AW	14	0.153	0.103	[0.003, 0.292]
	Die Linke (Erststimmen)	AW	11	0.418	0.250	[0.029, 0.929]
	CDU/CSU (zweitstimmen)	AW	14	0.756	0.526	[0.024, 1.654]
	SPD (zweitstimmen)	AW	14	0.367	0.295	[0.040, 0.668]
	Grune (zweitstimmen)	AW	14	0.715	0.392	[0.015, 1.684]
	FDP (zweitstimmen)	AW	14	0.199	0.128	[0.014, 0.470]
	Die Linke (zweitstimmen)	AW	14	0.119	0.091	[0.020, 0.267]
2009	CDU/CSU (Erststimmen)	DI	29	1.136	0.808	[0.032, 2.120]
	SPD (Erststimmen)	DI	29	0.753	0.556	[0.129, 1.798]
	Grune (Erststimmen)	DI	29	0.522	0.343	[0.047, 1.170]
	FDP (Erststimmen)	DI	29	0.519	0.289	[0.007, 1.218]
	Die Linke (Erststimmen)	DI	27	0.510	0.373	[0.028, 1.073]
	CDU/CSU (zweitstimmen)	DI	29	1.011	0.711	[0.039, 1.788]
	SPD (zweitstimmen)	DI	29	0.762	0.556	[0.077, 1.445]
	Grune (zweitstimmen)	DI	29	0.574	0.367	[0.036, 1.213]
	FDP (zweitstimmen)	DI	29	0.347	0.249	[0.025, 0.587]
	Die Linke (zweitstimmen)	DI	29	0.447	0.343	[0.020, 0.899]
	CDU/CSU (Erststimmen)	AW	29	1.138	0.825	[0.031, 2.321]
	SPD (Erststimmen)	AW	29	0.695	0.498	[0.082, 1.606]
	Grune (Erststimmen)	AW	29	0.547	0.372	[0.038, 1.195]
	FDP (Erststimmen)	AW	29	0.513	0.279	[0.011, 1.227]
	Die Linke (Erststimmen)	AW	27	0.466	0.335	[0.044, 0.969]
	CDU/CSU (zweitstimmen)	AW	29	0.981	0.722	[0.021, 1.722]
	SPD (zweitstimmen)	AW	29	0.754	0.573	[0.056, 1.348]
	Grune (zweitstimmen)	AW	29	0.594	0.393	[0.015, 1.288]
	FDP (zweitstimmen)	AW	29	0.342	0.234	[0.008, 0.662]
	Die Linke (zweitstimmen)	AW	29	0.399	0.301	[0.018, 0.775]
2013	CDU/CSU (Erststimmen)	DI	8	0.409	0.330	[0.050, 0.771]
	SPD (Erststimmen)	DI	8	0.269	0.218	[0.021, 0.478]
	Grune (Erststimmen)	DI	8	0.566	0.459	[0.053, 0.897]
	FDP (Erststimmen)	DI	8	0.191	0.143	[0.017, 0.342]
	Die Linke (Erststimmen)	DI	8	0.681	0.498	[0.097, 1.267]
	CDU/CSU (zweitstimmen)	DI	8	0.467	0.348	[0.019, 0.814]
	SPD (zweitstimmen)	DI	8	0.364	0.280	[0.038, 0.706]
	Grune (zweitstimmen)	DI	8	0.572	0.463	[0.066, 0.959]
	FDP (zweitstimmen)	DI	8	0.237	0.176	[0.038, 0.473]
	Die Linke (zweitstimmen)	DI	8	0.545	0.427	[0.140, 1.013]

CDU/CSU (Erststimmen)	AW	7	0.197	0.182	[0.076, 0.311]
SPD (Erststimmen)	AW	7	0.169	0.152	[0.067, 0.302]
Grune (Erststimmen)	AW	7	0.385	0.389	[0.047, 0.811]
FDP (Erststimmen)	AW	7	0.207	0.162	[0.020, 0.374]
Die Linke (Erststimmen)	AW	7	0.624	0.416	[0.059, 1.209]
CDU/CSU (zweitstimmen)	AW	7	0.314	0.241	[0.032, 0.602]
SPD (zweitstimmen)	AW	7	0.340	0.270	[0.014, 0.601]
Grune (zweitstimmen)	AW	7	0.434	0.447	[0.067, 1.019]
FDP (zweitstimmen)	AW	7	0.243	0.183	[0.018, 0.471]
Die Linke (zweitstimmen)	AW	7	0.486	0.349	[0.087, 1.002]

2.4. *Great Britain*

The British case differs from the others in that my projection method is compared against a different interpolation method, hereafter referred to as R&T's method, rather than the actual results.[15] Because this method is specific to the UK context, I describe it in some detail before providing information on the data used. The crux of their method is that they use local elections to estimate sub-constituency variation in party support and thus adjust their transfers of votes for the major parties. This is, however, a deceptively simple solution; I briefly try to describe their method using the available information, however, interested scholars should consult the books directly.

R&T's Interpolation Method They 'begin by building a matrix of local election results *as if each major party had contested every ward at the local elections* [emphasis added]'.[16] As they note, this runs into a few problems; first, not all wards hold elections every year and thus different years must be chosen depending on the local authority or region in question. Second, local elections in Britain are not simply first past the post; many include N-member districts where parties field up to N candidates and voters select N candidates, with the top N candidates being elected.[17] This requires the authors to standardise the results to get the 'effective' support for each party in each ward; this method has changed over time[18] and its modern iteration appears to require knowledge of the complete local electoral results in the relevant region to calculate the relevant weights.

A more serious problem is that not all (major) parties contest all wards in a given year and some wards are contested by only independent (non-party) candidates. The authors' solution is *ad hoc* based on their knowledge of local elections and involves some combination of the following:[19]

i 'see whether full contests were held in previous years anddraw conclusions from those results'

ii assigning shares for parties based on their performance in other wards in the area

iii '[make] use of [the relevant census year] ward level census data to examine the social characteristics of the ward and to draw conclusions from them about likely voting patterns'.

iv 'where independent or 'other' party candidates seem to have been given a free run by one of the major parties, their vote is given to that party as if they were its surrogate. Where independents or 'other' party candidates stand in addition to candidates from all the major parties, their vote is ignored'.

The above procedure is further complicated by periodic ward-level boundary changes that require further adjustment to use the methods above. The authors again use a variety of complicated methods to adjust for this problem.

When considering minor parties, the authors tend to lump these together but sometimes include them based on the number of local candidates fielded. For example, the United Kingdom Independence Party (UKIP) and the Greens appear in the 2010 notional results because they

15. BBC and ITN 1983; Rallings and Thrasher 1995; Denver, Rallings, and Thrasher 2004; Rallings and Thrasher 2007. The two authors do not appear to have contributed to the oldest published work (BBC and ITN 1983), however, they are clearly the natural heirs to this project and thus I refer to all four conversion guides as providing the R&T estimates.

One related point is to note that a small number of other scholars (Rossiter, Johnston, and Pattie 1997b, 1997a; Cornford, Dorling, and Tether 1995) have outlined other methods for interpolating results. Whilst not attempting to marginalise their contributions, I focus on R&T's method here and in the paper insofar as it is much more widely used and thus is the key benchmark.

16. Rallings and Thrasher 2007, p. 9.

17. This electoral system, sometimes referred to as 'block voting', was abolished for parliamentary elections after 1945 with the elimination of all remaining multi-member districts.

18. The most recent method is outlined in Ware et al. 2006. The changes in 1997 used the method described in Curtice, Payne, and Waller 1983. The changes in 1983 used the mean vote of a party's candidate in a district and adjusted for missing candidates separately. BBC and ITN 1983.

19. Rallings and Thrasher 2007, p. 10. Older books use similar methods of adjustment. The authors provide no information as to which methods are used in which wards or firm rules for which rule to apply, further complicating replication.

fielded a sizeable number of candidates in 2005; however, for other years and elections, disaggregation only occurs if a minor party crosses some unspecified threshold of support, e.g. British National Party results are listed separately for some constituencies but not others.

Once this local information is collated, the authors then calculate the total number of (local) votes cast for, say, the Conservatives in all wards in the old constituency. They note the proportion of those votes that come from any given ward and assume that the same proportion of Conservative votes at the previous parliamentary election came from that ward. Then, they create the new boundaries and move wards (and the corresponding vote totals) as necessary to get the 'notional results'.

Despite having the advantage over my method of including sub-constituency variation, clear concerns with this method are replicability, applying it to other variables,[20] and addressing boundary changes before the 1979/1983 change. Even setting aside the ambiguity over how the relevant local adjustments are to be calculated, it is prohibitively time-consuming to calculate the local results for historic elections. Moreover, whilst local election data exists for more modern elections, this is not available to the academic community in any sort of standardised form.[21]

Electoral Shapefiles Shapefiles for the elections come from EDINA and GBHGIS.[22]

Electoral Returns I interpolate the shares of the three major national parties (Labour, Conservative, and Liberal Democrats) as well as the nationalist parties (SNP and Plaid Cymru).[23] I exclude two constituencies for which the Speaker of the House of Commons ran (Cardiff West in 1979; Glasgow North East in 2005); R&T code the Speakers as members of their former party, however, my data codes them as a separate grouping. I also exclude two Social Democrat candidates in 1997 (Erith & Crawford and Greenwich & Woolwich); they ran under a separate designation but the Liberal Democrats did not field a candidate in that seat. R&T class these candidates as Liberal Democrats, but I code them according to their stated party affiliation. Clearly, the inclusion of these seats with a '0' for a major party would generate unduly large errors and are therefore excluded.

Population Data Population data is included at the ward level, as these are the building blocks for constituencies and thus are likely sufficiently detailed.[24] Around 10,000 wards exist in Great Britain at each interpolation.[25] The census population data ('usual residents', i.e. all present residents plus all absent residents) comes from MIMAS.[26]

Implementation Details The results reported in the main paper used rasters of resolution 6,000 by 6,000. This resulted in at least 99.9% of wards being successfully rasterised.

Disaggregated Results

20. Regarding electoral variables, they do not interpolate the shares of smaller parties due to their limited pattern of contestation. For example, one would be unable to model the effect of previous far-right and Eurosceptic support on UKIP's initial emergence in 1997 insofar as detailed interpolations of specific minor party support at the 1992 election are not included in their method.

21. The authors also publish (physical books of) local election results each year and thus they have access to this data. Parts of this has been released (Rallings, Thrasher, and Ware 2006); however, it does not give assignment of wards to constituencies. Whilst replicable, this is a somewhat arduous task.

22. This work is based on data provided through `www.VisionofBritain.org.uk` and uses historical material which is copyright of the Great Britain Historical GIS Project and the University of Portsmouth. I also relied on `http://census.edina.ac.uk/`, accessed 1 March 2015, for some shapefiles. Some merging and cleaning is required to combine shapefiles from various regions and years. For example, Milton Keynes was split in an interim boundary review in 1992 and thus the shapefiles used in 1983 and 1992 differ in that one crucial respect.

23. In the 1979 election, the Liberal Party is interpolated onto the 1983 boundaries.

24. Shapefiles come from EDINA.

25. The number drops to slightly below 9000 in the 2011 census.

26. `http://casweb.mimas.ac.uk/` accessed 1 March 2015 for the 1981, 1992, and 2001 censuses. 2011 census data, used when interpolating the 2005 boundaries onto the 2010 ones, comes from `http://infuse.mimas.ac.uk/` accessed 1 March 2015.

Table 4: Disaggregated British Results

Year	Party	Method	N	RMSE	MAE	Middle 90% of AE
1983	Conservative	DI	409	3.601	2.438	[0.109, 7.458]
	Labour	DI	409	3.735	2.562	[0.119, 8.089]
	Liberal/LibDem	DI	381	1.367	0.761	[0.025, 2.902]
	Plaid Cymru	DI	26	0.998	0.500	[0.041, 1.348]
	SNP	DI	53	3.084	2.040	[0.041, 5.869]
	Conservative	AW	448	3.738	2.694	[0.122, 7.777]
	Labour	AW	448	4.367	3.143	[0.151, 9.425]
	Liberal/LibDem	AW	419	2.198	1.263	[0.036, 4.420]
	Plaid Cymru	AW	28	2.754	1.376	[0.061, 4.646]
	SNP	AW	51	3.009	2.184	[0.134, 6.242]
1997	Conservative	DI	248	3.548	2.631	[0.167, 7.275]
	Labour	DI	248	3.705	2.854	[0.164, 7.768]
	Liberal/LibDem	DI	246	2.168	1.537	[0.108, 4.384]
	Plaid Cymru	DI	6	1.410	0.930	[0.199, 2.728]
	SNP	DI	37	1.928	1.395	[0.121, 2.913]
	Conservative	AW	263	3.525	2.702	[0.125, 6.874]
	Labour	AW	263	3.900	2.988	[0.200, 8.242]
	Liberal/LibDem	AW	261	2.416	1.722	[0.100, 4.912]
	Plaid Cymru	AW	7	1.389	1.174	[0.242, 2.241]
	SNP	AW	32	1.854	1.410	[0.129, 3.891]
2005	Conservative	DI	54	2.292	1.519	[0.033, 5.113]
	Labour	DI	54	3.929	2.497	[0.090, 7.375]
	Liberal/LibDem	DI	54	3.048	1.840	[0.068, 6.369]
	Plaid Cymru	DI	0	NaN	NaN	[NA, NA]
	SNP	DI	54	1.507	1.086	[0.076, 3.539]
	Conservative	AW	48	2.804	1.959	[0.242, 6.043]
	Labour	AW	48	4.667	3.196	[0.335, 10.719]
	Liberal/LibDem	AW	48	3.611	2.344	[0.144, 8.424]
	Plaid Cymru	AW	0	NaN	NaN	[NA, NA]
	SNP	AW	48	1.756	1.396	[0.361, 3.624]
2010	Conservative	DI	181	3.442	2.492	[0.116, 7.677]
	Labour	DI	181	3.028	2.217	[0.189, 6.494]
	Liberal/LibDem	DI	181	2.634	1.826	[0.112, 5.454]
	Plaid Cymru	DI	3	2.701	2.559	[1.917, 3.604]
	SNP	DI	0	NaN	NaN	[NA, NA]
	Conservative	AW	193	3.294	2.440	[0.182, 7.086]
	Labour	AW	193	3.347	2.389	[0.144, 7.543]
	Liberal/LibDem	AW	193	2.947	2.047	[0.137, 5.895]
	Plaid Cymru	AW	9	3.576	2.161	[0.155, 7.589]
	SNP	AW	0	NaN	NaN	[NA, NA]

2.5. *New Zealand*

A distinctive initial issue for New Zealand is how to treat the Maori electorates. I included them in the analysis as separate constituencies; however, it in fact turned out that none are subject to enough reorganisation to be included in the results. In separating votes between Maori and general electorates, I allocate votes listed in the official returns under the seven Maori constituencies to the Maori constituencies and analyse them separately. As with Germany, I poll both the constituency and list votes cast in the MMP electoral system.

Electoral Shapefiles As in Australia, the Electoral Commission provides shapefiles of the point locations of most polling booths used in recent elections.[27] Thus, it is easy to assign polling booths to the appropriate constituency (referred to 'electorates' in New Zealand). However, as ever, complications emerge for two primary reasons.[28] The first is the same in Australia; voters may choose to vote at any polling station in their election—the key implication here is that the effective number of constituencies (ENC) is similarly calculated using the valid votes cast in the constituency. The second is more complicated: voters may vote in polling stations *outside* their election (either by accident or intentionally). It is somewhat complicated how these votes are counted (as for some only the 'list' vote is counted), but the crucial complication for this analysis is that official results for some constituencies (e.g. Auckland Central) contain polling booths that are not geographically inside said constituency (e.g. two polling stations in Kingsland, normal general electorate of Mt. Albert). Whilst this complication does not apply to the majority of voters, the numbers are not trivial and thus I was hesitant to discard them entirely.

I solve this problem in the following fashion: I assign votes based on the geographic location of the *polling place* even if the votes themselves 'counted' in a different location as there is no way to geographically assign those votes. I use this assignment to construct the electoral results on the old boundaries that are then interpolated onto the new boundaries. The shapefiles for the electoral constituencies (General and Maori) were available from Statistics New Zealand.[29]

Electoral Returns Elections New Zealand provided the polling place-level returns.[30] Parties with parliamentary representation in the 2011 Parliament were included in the analysis;

Population Data The population data comes from the 2006 New Zealand census.[31] I use the 'usually resident population' data collected on 'area units'.[32]

Implementation Details The results reported in the main paper used rasters of resolution 6,000 by 6,000. This resulted in 99.5% of wards being successfully rasterised.

Disaggregated Results

27. `http://www.electionresults.govt.nz/` accessed 1 March 2015.
28. Common issues about the inability to assign advance voters and/or overseas voters apply to New Zealand and the solution is the same, i.e. exclusion. A more minor issue is that polling station vote breakdowns are not released if fewer than six individuals voted there; such voters are grouped into an aggregate total that cannot be assigned to a specific geographic location.
29. `http://www.stats.govt.nz/browse_for_stats/people_and_communities/Geographic-areas/digital-boundary-files.aspx` accessed 1 March 2015.
30. http://www.electionresults.govt.nz/ accessed 1 March 2015.
31. `http://www.stats.govt.nz/Census/2006CensusHomePage.aspx` accessed 1 March 2015.
32. Shapefiles downloaded from `http://www.stats.govt.nz/browse_for_stats/people_and_communities/Geographic-areas/digital-boundary-files.aspx` accessed 1 March 2015.

Table 5: Disaggregated New Zealand Results

Year	Party	Method	N	RMSE	MAE	Middle 90% of AE
2014	National (List)	DI	18	2.652	2.089	[0.326, 5.264]
	Labour (List)	DI	18	3.438	2.423	[0.242, 8.192]
	Green (List)	DI	18	1.192	0.760	[0.040, 2.361]
	ACT (List)	DI	18	0.116	0.087	[0.003, 0.239]
	Mana (List)	DI	18	0.027	0.021	[0.002, 0.052]
	United Future (List)	DI	18	0.036	0.028	[0.005, 0.061]
	Maori (List)	DI	18	0.036	0.030	[0.003, 0.061]
	National (Electorate)	DI	18	2.406	1.882	[0.549, 5.015]
	Labour (Electorate)	DI	18	3.046	2.322	[0.174, 6.211]
	Green (Electorate)	DI	18	0.710	0.529	[0.057, 1.189]
	ACT (Electorate)	DI	18	0.514	0.212	[0.002, 1.400]
	Mana (Electorate)	DI	9	0.034	0.028	[0.012, 0.064]
	United Future (Electorate)	DI	12	0.116	0.062	[0.003, 0.252]
	Maori (Electorate)	DI	1	0.000	0.002	[0.002, 0.002]
	National (List)	AW	19	3.701	2.758	[0.202, 9.713]
	Labour (List)	AW	19	4.258	2.983	[0.258, 8.636]
	Green (List)	AW	19	1.380	0.748	[0.035, 3.323]
	ACT (List)	AW	19	0.104	0.077	[0.005, 0.207]
	Mana (List)	AW	19	0.052	0.039	[0.003, 0.087]
	United Future (List)	AW	19	0.059	0.038	[0.007, 0.110]
	Maori (List)	AW	19	0.043	0.034	[0.003, 0.078]
	National (Electorate)	AW	19	4.665	3.003	[0.292, 12.774]
	Labour (Electorate)	AW	19	5.018	3.343	[0.123, 14.889]
	Green (Electorate)	AW	19	1.180	0.819	[0.033, 2.340]
	ACT (Electorate)	AW	19	0.524	0.277	[0.002, 1.364]
	Mana (Electorate)	AW	11	0.134	0.083	[0.005, 0.283]
	United Future (Electorate)	AW	13	0.812	0.338	[0.002, 1.527]
	Maori (Electorate)	AW	3	0.069	0.062	[0.032, 0.098]

2.6. *United States*

This analysis differs from the other countries insofar as it interpolates *presidential* (2008) rather than legislative electoral data. This is to ensure data consistency across states; the data source used (Harvard Elections Data Archive — HEDA) is reasonably consistent in reporting presidential data but only periodically reports House electoral data. As a general comment on the variables used for this analysis, my approach was to maximise the number of states included in the analysis (to test the method across as many sets of boundary changes as possible) even if it required relying on variables that were imperfect.

Electoral Shapefiles The electoral shapefiles used in this project came from two sources; the constituency-level shapefiles came from data gathered by Lewis et al.[33] I exclude all states with a single representative from the analysis. The sub-constituency level data and shapefiles came from the Harvard Election Data Archive.[34]

The shapefiles generally correspond to 'VTDs' (voting-tabulation districts), analogous to precincts or polling station shapefiles used in the other countries. I use the same 'point on surface' method as the Canadian case for allocating these VTDs to the 2010 and 2012 boundaries to get the 'actual' results for each election.

Electoral Returns Because of the general weakness of minor parties in the US (and partially due to data limitations), I calculate the two-party share of the vote for each party and use that in the interpolations. The next paragraph outlines the limitations of the HEDA data and notes the changes made.

First, the HEDA does not contain shapefiles with appropriate electoral data for the following states: Arkansas, Kentucky, Maine, Oregon, Rhode Island, Utah and West Virginia.[35] Second, a small number of HEDA files were corrupted and thus could not be loaded in R, ArcGIS or QGIS: Georgia, Washington, and Wisconsin. I reconstructed the files for Georgia and Washington were reconstructed based on downloading the relevant VTD (voting-tabulation district) shapefiles from the US Census and matching it to the HEDA data; Wisconsin could not be successfully repaired. Third, a number of states lacked data even on presidential elections. Because these three states (Connecticut, Massachusetts, and New York) contained comparably large numbers of representatives, I attempted to find alternative measures to avoid excluding them entirely from the analysis. For New York, I relied on US Senate (2010) data collapsed onto the 2010 House boundaries. For Massachusetts and Connecticut, no vote data *at all* was reported so I relied on the number of registered Republicans and Democrats as a proxy.

Population Data For many states the HEDA provided information on 'voting age population' in the VTDs that is used to construct the population shapefiles. Unfortunately, voting age population data was not provided for some states and thus the number of registered voters (or failing that the total population) was used.[36] Three states still lacked population data and thus, as a final resort, turnout at the VTD level was used as population data.[37]

Implementation Details The results reported in the main paper used rasters of resolution 6,000 by 6,000 and interpolated each state individually. This resulted in over 100% of VTDs being successfully rasterised in most states and only four states had under 99% rasterised: Nevada (96.7%), California (97.2%), Kansas (98.2%), Texas (98.8%). The remaining population in those states was assigned to the nearest pixel.

Disaggregated Results

33. Lewis et al. 2013; `http://cdmaps.polisci.ucla.edu/` accessed 1 March 2015.

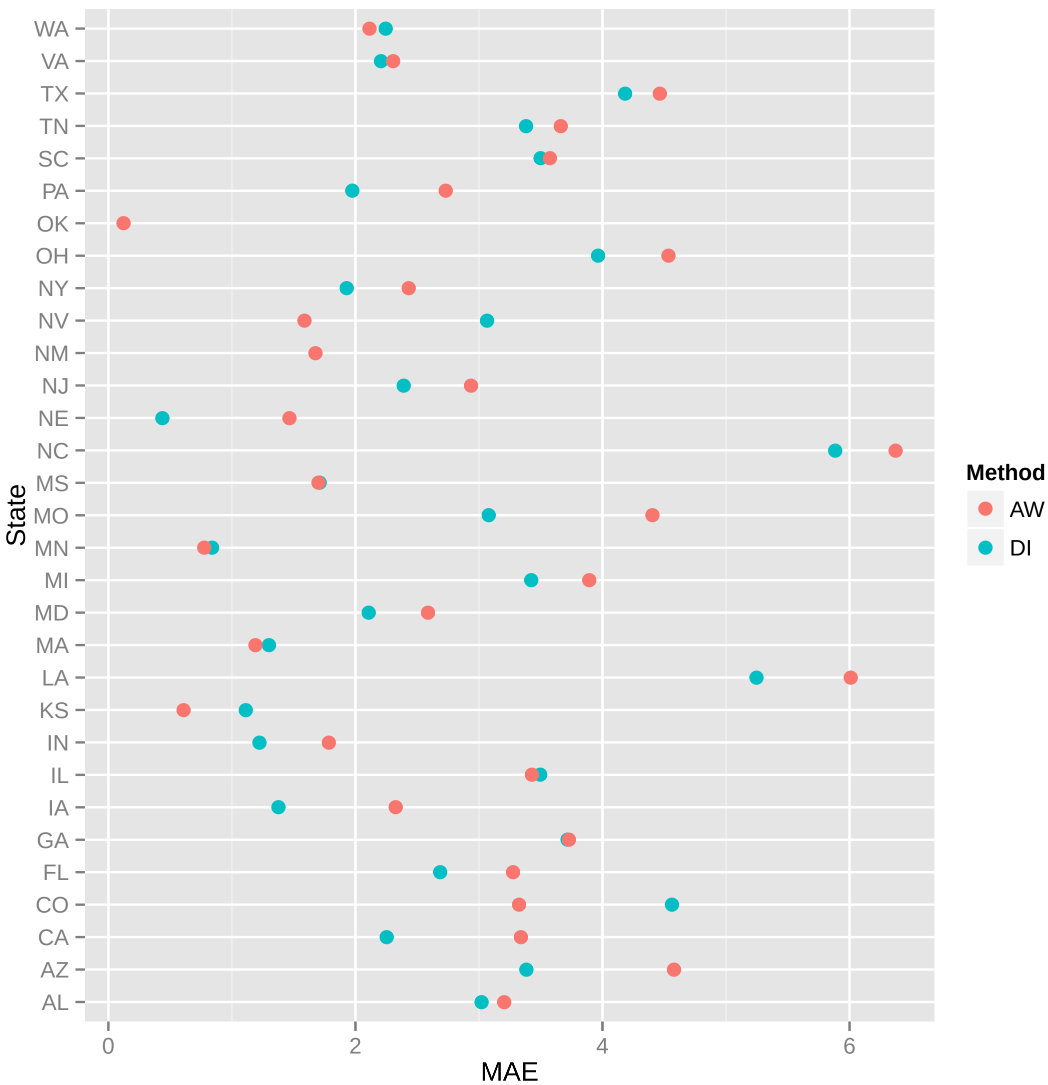
34. `http://projects.iq.harvard.edu/eda/` accessed 1 March 2015. Data on California was downloaded directly from `http://www.statewidedatabase.org/` accessed 1 March 2015.

35. No data exists for Montana, however, it is excluded regardless as it only has one Representative.

36. Registered voters were used in California, Georgia, Minnesota, North Carolina, Ohio, and Virginia. Total population was used in Illinois and total population (older than eighteen) was used in Wisconsin.

37. Maryland, Florida, and South Carolina.

Figure 4: MAE by State



References

BBC and ITN. 1983. *The BBC/ITN Guide to the New Parliamentary Constituencies.* Chichester: Parliamentary Research Services.

Bivand, Roger, and Colin Rundel. 2014. *rgeos: Interface to Geometry Engine — Open Source (GEOS).* R package version 0.3-8. `http://CRAN.R-project.org/package=rgeos`.

Cornford, James, Danny Dorling, and Bruce Tether. 1995. Historical Precedent and British Electoral Prospects. *Electoral Studies* 14 (2): 123–142.

Curtice, John, Clive Payne, and Robert Waller. 1983. Lessons of the 1982 English Local Elections. *Electoral Studies* 2 (1): 3–22.

Denver, David, Colin Rallings, and Michael Thrasher. 2004. *Media Guide to the New Scottish Westminster Parliamentary Constituencies.* Plymouth: Local Government Chronicle Elections Centre.

Laakso, Markku, and Rein Taagepera. 1979. Effective Number of Parties: A Measure with Applications to Western Europe. *Comparative Political Studies* 12 (3): 3–27.

Lewis, Jeffrey B., Brandon DeVine, Lincoln Pitcher, and Kenneth C. Martis. 2013. *Digital Boundary Definitions of United States Congression Districts, 1789-2012.* `http://cdmaps.polisci.ucla.edu`, 28 Nov 2014.

Rallings, Colin, and Michael Thrasher. 1995. *Media Guide to the New Parliamentary Constituencies.* Plymouth: Local Government Chronicle Elections Centre.

———. 2007. *Media Guide to the New Parliamentary Constituencies.* Plymouth: Local Government Chronicle Elections Centre.

Rallings, Colin, Michael Thrasher, and Lawrence Ware. 2006. *British Local Election Database, 1889-2003.* `http://dx.doi.org/10.5255/UKDA-SN-5319-1`, 28 Nov 2014 (accessed Sept. 1, 2014).

Rossiter, David, Ron Johnston, and Charles Pattie. 1997a. Estimating the Partisan Impact of Redistricting in Great Britain. *British Journal of Political Science* 27 (2): 319–331.

———. 1997b. Redistricting and Electoral Bias in Great Britain. *British Journal of Political Science* 27 (3): 466–472.

Ware, Lawrence, Galina Borisyuk, Colin Rallings, and Michael Thrasher. 2006. A New Algorithm for Estimating Voter Turnout When the Number of Ballot Papers is Unknown. *Electoral Studies* 25 (1): 59–71.