APPENDIX

The Development of the Urban-Rural Cleavage in Anglo-American Democracies

Zack Taylor, University of Western Ontario, zack.taylor@uwo.ca
Jack Lucas, University of Calgary
David A. Armstrong II, University of Western Ontario
Ryan Bakker, University of Essex

Appendix A: Data Sources

We initially attempted to assemble election results from the Comparative Legislative Elections Database (CLEA) but found it lacking in several respects: inadequate temporal coverage, the presence of errors and missing data, and inconsistent handling of results in multi-member districts.

United States

For the United States, House of Representatives general election results for the entire period were obtained from the CQ Press Voting and Elections Collection (CQ Press 2022), which, unlike other data sources we explored, include vote shares for each candidate, including third-party and endorsed candidates endorsed by multiple parties. These data are restricted to subscribers. As a result, they are not posted on the replication data repository.

District population counts for the 78th through 105th Congresses (1942–1996) were obtained from Adler (n.d.) and from the National Historical Geographical Information System (NHGIS) (Manson et al. 2021) for more recent Congresses.

Land areas of districts were calculated from digital boundary files created by Lewis et al. (2013) up to the 114th Congress elected in 2014, and by the NHGIS for the 115th and 116th Congresses.[1] We also appended Census Bureau standard region codes.

United Kingdom

Constituency-level party vote shares for the United Kingdom are available from the House of Commons Library (Watson, Uberoi, and Loft 2020). In single-member constituencies, the party receiving the highest vote share was coded as the winner. To correctly identify winning candidates in two-member constituencies where parties each ran multiple candidates, we appended candidate-level election results transcribed from Craig (1977).

To calculate land area, digital boundary files for the 1918 through 1973 periodic reviews were obtained from the *A Vision of Britain Through Time* project (Southall 2017) and for the 1983 through 2010 reviews from the UK Data Service (2018). As the 1974 boundaries are not publicly available, we redrew them in GIS with reference to available printed and digital sheet maps (Boundary Commission for Scotland 2011; Ordnance Survey 1974). Unfortunately, historical digital constituency boundaries are not consistently available for Northern Ireland, and the historical university constituencies are non-territorial in nature (Ratcliffe 2015). Northern Ireland and university constituency results were thus dropped from the analysis, leaving regular constituencies for Great Britain: England, Wales, and Scotland.

The British Census does not regularly disseminate the populations of parliamentary constituencies. This is unsurprising as legislation requires constituencies to have similar numbers of enrolled electors, not residents. However, the election results data from the House of

[1] We used the *sf* package in R for all land area calculations.

Commons Library contain counts of registered electors in each constituency at each general election. We calculated elector density as a proxy for population density.

Canada

Canadian candidate-level election results are available from Sevi (2019) based on earlier digitization work by the Library of Parliament (n.d.).

District land area was calculated in GIS using digital boundary files. Statistics Canada's federal electoral district digital boundary files begin with the 1987 boundary set. For earlier years we adapted boundaries created by freelance cartographer J.P. Kirby (Taylor et al. 2023). To ensure a more accurate density calculation, we subtracted internal water bodies from the gross district area to produce a net land area.

Digital census population counts are available starting in 1961 for the 1952 boundary set. Earlier district population counts were transcribed from tables in published volumes of the decennial census.

Temporal coverage

To enable "apples to apples" comparison of elector density values across time in Britain, we begin our period of analysis in 1928, the first year in which women were fully enfranchised. (While British property-owning or university-graduate women over the age of 30 gained the right to vote in 1918, equal suffrage to men was not extended until 1928.) Although we use population density in Canada, for consistency we apply the same criterion to Canada, where suffrage was extended in 1920.

Finally, while the 19th Amendment extended full voting rights to American women in 1920, we lack the district-level population data required to calculate population density prior to

1944. Using district-level registered elector counts instead of populations offered a potential alternative as in Great Britain, but this information is not available. Our available data thus include Britain from 1928, Canada from 1920, and the United States from 1944.

Data structure

Each country's election results source data are formatted differently. We reformatted them so that each row represents an individual candidate and marshalled available information into columns with harmonized names. The table's columns are described in **Table 1**.

Table 1: Description of table columns

Variable	Description
case	Country (CAN, GBR, USA)
ryear	Boundary set year
eyear	Election year
edate	Election date
parl	Parliament or Congress number (USA and Canada only)
seats	Number of seats in Parliament or Congress
gov	Party in government
govgrp	Party grouping #1 – See appendix B
govgrp2	Party grouping #2 – See appendix B
distid	Unique district identifier code (USA and Canada only)
distsystem	Electoral system (SMP)
distmag	District magnitude
distname	District Name
distnameflat	Flattened version of district name
region	Region (census division in USA, country in UK, census region in Canada)
division	Subregion (census region in USA, district grouping in GBR, not present in Canada)
stateprov	State or province
electors	Number of electors (GBR only)
pop	Population (Canada and GBR only)
pty	Candidate's party
ptygrp1	Candidate's party grouping #1 – See appendix B (used in Canada and GBR)
ptygrp2	Candidate's party grouping #2 – See appendix B (used in USA)
candname	Candidate name (USA and Canada only)
candid	Unique candidate identifier (USA and Canada only)
candinc	Incumbent = 1, challenger = 0 (USA and Canada only)
win	Elected = 1, not elected = 0
acclaim	Acclaimed = 1, not acclaimed = 0
votes	Number of votes received
pct	Percentage of votes received
cast_valid	Number of valid votes cast

cast_spoil	Ignore
cast_total	Ignore
cast_ballots	Ignore
turnout	Turnout (cast_total / electors; GBR only)
areatotal	Gross land area in square kilometers
arealand	Land area net of water features in square kilometers (Canada only)
density	Population or elector density based on land area if available; if not, based on gross area
logdensity	Log-transformed density

Appendix B: Compositional Regression Models

Method

A *composition* is an array of data whose elements add up to a constant sum. These are most prevalent where the elements of the composition sum to 1 (or 100) as the vote shares do here, but that need not be the case. Because the values in the composition (e.g., vote shares for parties running in a single district in an election) sum to one (or 100 in percentage terms), the relevant information is the relative magnitude of shares across parties. For example, we may be interested in the ratio of vote shares for Conservative versus Liberal candidates in Canada. As Tomz, Tucker, and Wittenberg (2002) suggest, OLS is not the optimal model to estimate relationships among independent variables and compositional dependent variables. In broad strokes, we follow the strategy they set out.

When analyzing compositional data, the data must first be transformed to leverage the relative vote share information, rather than the absolute magnitudes of the original vote shares. Once transformed, they can be analyzed using classical multivariate techniques (multivariate meaning multiple dependent variables) like seemingly unrelated regression (SUR). Aitchison (1982) provides the seminal technical introduction to the topic. In the analyses below, we use the additive log-ratio (ALR) transformation as implemented in the `compositions` package in R (van

den Boogaart, Tolosana-Delgado, and Bren 2022). The ALR transformation maps the D-length compositions onto a D-1 length Euclidian vector such that:

$$ALR(x_d) = log\left(\frac{x_d}{x_D}\right) \forall\, d = \{1, \ldots, D-1\}$$

The general idea is to transform the M-dimensional matrix of vote-shares into an $D-1$ dimensional matrix that captures the relevant relative information using the ALR transformation. Next, we estimate a SUR on the $D-1$ dimensional ALR matrix. We then generate predictions for the SUR model under two conditions – one that where all observations take on the lowest log-density value in the region and one where each observation takes on the highest log-density value in the region. Those predictions are then transformed back into vote shares using the following formula.

$$\text{Vote Share for Party } d = ALR^{-1}(x_d) = \frac{e^{ALR(x_d)}}{1 + \sum_{d=1}^{D-1} e^{ALR(x_d)}} \forall\, d = \{1, \ldots, D-1\}$$

The vote share for party D is just $ALR^{-1}(x_d) = \frac{1}{1+\sum_{d}^{D-1} e^{ALR(x_d)}}$ or equivalently $1 - \sum_{d=1}^{D-1} ALR^{-1}(x_d)$, which is just one minus the sum of all other vote shares.

Finally, we use first differences to estimate the effect of log-density. One way of doing this would be to induce the maximal change in log-density for all observations. However, we take a slightly different approach. In recognizing that region and urbanity are related,[2] we induce the maximal *within-region* change in log-density. This prevents us from generating effects that are due to extrapolation beyond the values observed in the data. Looking at the 2021 Canadian election, the full range of the log-density variable is (–2.29, 9.74). However, in the Atlantic

[2] We find that region explains between 10% and 19% of the variation in urbanity in Canada, between 26% and 33% of the variation in urbanity in the UK and between 5% and 21% of the variation in urbanity in the US (depending on election year).

provinces, we only observe values in the range (–2.29, 6.90). In Ontario, we observe values in the range (–1.56, 9.74). To suggest that we should evaluate the same maximal change in all regions would be to over-estimate even the maximal effect of urbanity. To generate estimates of sampling variability, we use a parametric bootstrap of the coefficients and summarise those with raw percentile confidence intervals.

Inconsistency in party coverage over time

One of the problems identified by Tomz, Tucker, and Wittenberg (2002) is that parties do not always run candidates in every district. In the formulae above, we assume that there is a party that gets non-zero vote-share in each district, but there are a small number of elections in each country where the winning candidate ran unopposed. There are also cases where small or regional parties did not field candidates in all districts. Independent candidates are also important in some district elections. To address these issues, we first eliminate uncompetitive elections (those where a single candidate ran for office). We then create party profiles for the districts where we identify the combination of parties that run candidates in each district. We eliminate profiles that have fewer than 10 observations in a particular election. With fewer than 10 observations, we cannot reasonably estimate a regression on log-density and region.

Canada

Of the 8,513 district-years in Canada, all but 8 are competitive. Building candidate profiles for each district, we find that there are 65 profiles that do not meet the sample-size requirement of 10. In total, they account for 545 district-years that are excluded from the analysis. This leaves 7,950 or 94% of the competitive district-years included in the analysis. Roughly 50% of the excluded districts occur between 1921 and 1953. Aside from that period, the only election years

of note are 1993 and 1997 where 30 and 35 districts were excluded, respectively, as well as 17 in 1988, 18 in 2000, 16 in 2008 and 20 in 2011. Otherwise, the number of excluded districts after 1953 was 15 or fewer in each year. As for region, 33% of the excluded district-years are from Quebec, roughly 24% from both Ontario and the Prairies, 11% from the Atlantic provinces and 8% from BC. Figure B1 shows the district urbanity scores from the included and excluded district-years. The excluded districts tend to be more rural, though the two distributions are not so different as to cause concern.

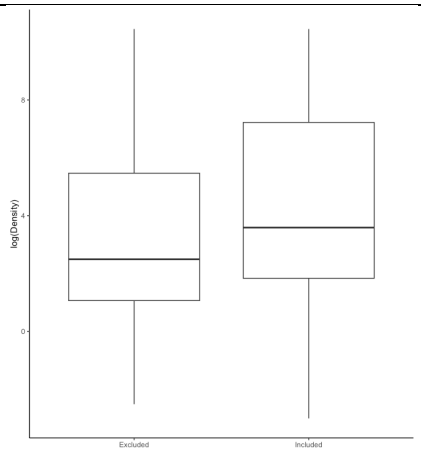


Figure B1: District Urbanity Scores for Included and Excluded District Years for Canada

Party	Included	Excluded
Conservative	7,899 (95%)	398
Liberal	7,893 (94%)	465
CCF-NDP	6,491 (96%)	287
Third	5,487 (95%)	285
Independent	1,758 (86%)	283
Reform-Alliance	745 (93%)	52
Bloc	701 (92%)	58
Progressive	208 (66%)	105

Figure B2: Number of District-years Parties are Included in and Excluded from Analysis

As for which parties are included and excluded, the table in Figure B2 gives the number of district-years in which each party fielded candidates by whether they were included or excluded. Most parties are well-represented in the analysis with more than 90% of their district-years included. The biggest exception is the Progressive party where only about two-thirds of its candidates are included in the analysis.

Great Britain

There are 14,864 district-years in the UK and all but 135 are competitive. Using the 8 party labels in the "ptygrp1" variable, we lose 491 district-years. This leaves 14,498 or 98% of competitive riding-years in the analysis. As in Canada, roughly 50% of the excluded districts occur between 1929 and 1951 and roughly 70% between 1929 and 1970. After 1970, the only notable years are 2017 and 2019 where we lose 20 and 36 districts, respectively, due to profiles with samples that are too small. Perhaps not surprisingly, Scotland and Wales are hardest hit – 37% and 24% of the excluded district-years, respectively. All other regions account for less than 10% of the excluded district-years. As with Canada, the excluded district-years tend to be more rural than the included ones, but there is a great deal of variation in both groups (see Figure B3).

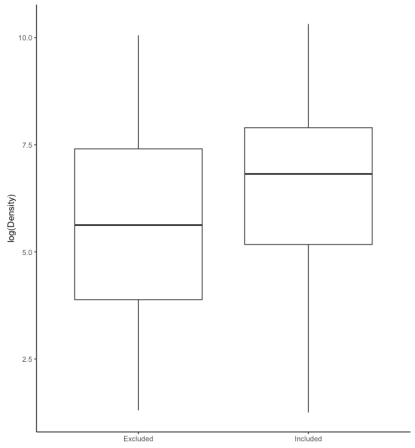


Figure B3: District Urbanity Scores for Included and Excluded District Years for Great Britain

Party	Included	Excluded
Conservative	14,320 (98%)	307
Labour	14,299 (98%)	348
Liberal	10,787 (98%)	202
Other	6,309 (97%)	185
English Nationalist	1,529 (99%)	21
Scottish Nationalist	1,092 (93%)	82
Welsh Nationalist	294 (69%)	135
Nationalist	252 (92%)	23

Figure B4: Number of District-years Parties are Included in and Excluded from Analysis

Figure B4 shows the parties included in and exclude from the analysis. All parties except the Welsh Nationalist party have more than 90% of their candidates included. For the Welsh Nationalists, like the Progressives in Canada, that percentage is closer to two-thirds.

USA

In the USA, there are 16,968 district-years of which 1,500 are not competitive. Using the "ptygrp2" variable, we lose 1749 district-years due to small profile sample sizes or uncompetitive elections. This leaves us with 15,152 districts, 97% of all competitive district-years. As with Canada and Great Britain, the district-years we lose are predominantly in the early period with 50% of the lost district-years occurring between 1944 and 1960. Starting around 1970, the missing district-years are spread more or less evenly across time. 72% of the excluded district-years come from the South – many of these are the result of non-competitive elections in the early part of the analysis. Figure B5 shows the box plots of density by inclusion in the analysis. Again, the excluded district-years are more rural, on average, but there is still considerable variation among the included districts.

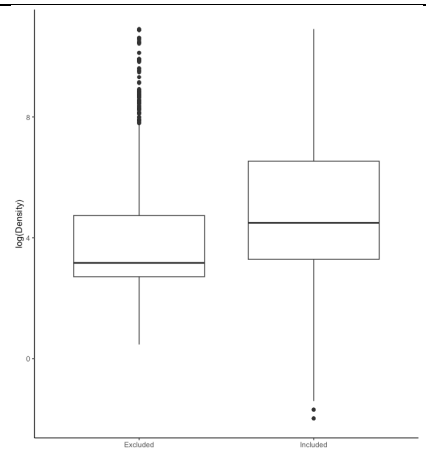


Figure B5: District Urbanity Scores for Included and Excluded District Years for the USA

Party	Included	Excluded
Democrat	14,985 (92%)	1,220
Republican	14,598 (97%)	403
Minor	7,049 (97%)	251

Figure B6: Number of District-years Parties are Included in and Excluded from Analysis

Figure B6 shows the parties by inclusion status in the USA. Here, all parties are well-represented in the analysis. The biggest exception is the Democratic party, which ran unopposed in many Southern districts in the early part of the analysis.

Grouping of parties

For consistency across time, parties and independent candidates were coded into groups. Generally small and transitory parties, and also independents without any recorded affiliation with a party, were grouped into a "minor" category for the purposes of the vote share analysis. These collectively received minimal vote share in any given election.

Canada
The major parties, some of which have changed names over time, are grouped as below. The residual parties and independents, which number over 100, are grouped together for the purposes of the vote share analysis.

Group	**Parties**
Bloc Québécois	Bloc Québécois
CCF-NDP	Cooperative Commonwealth Federation, New Democratic Party
Conservative	Conservative, National Government (WWI), Progressive Conservative, Liberal Conservative Coalition
Liberal	Liberal, Liberal Progressive
Progressive	Progressive, Progressive Canadian Party
Reform_Alliance	Reform Party, Canadian Alliance Party

Great Britain
As Northern Ireland is not included in the analysis, Northern Irish nationalist and unionist parties are not included in this list.

Group	**Parties**
Conservative	Conservative, Constitutionalist (Anti-Labor group 1923-24), Independent Conservative, National Independent, National, National Liberal (Lloyd George)
Labour	Labour, Independent Labour, National Labour
Liberal	Liberal, Independent Liberal, Social Democrats, Liberal Democrats
English nationalist	UKIP
Nationalist	Brexit Party
Scottish nationalist	Scottish Nationalist Party
Welsh nationalist	Plaid Cymru
Minor	Commonwealth Movement, Communist, Green, Independent, Workers Party

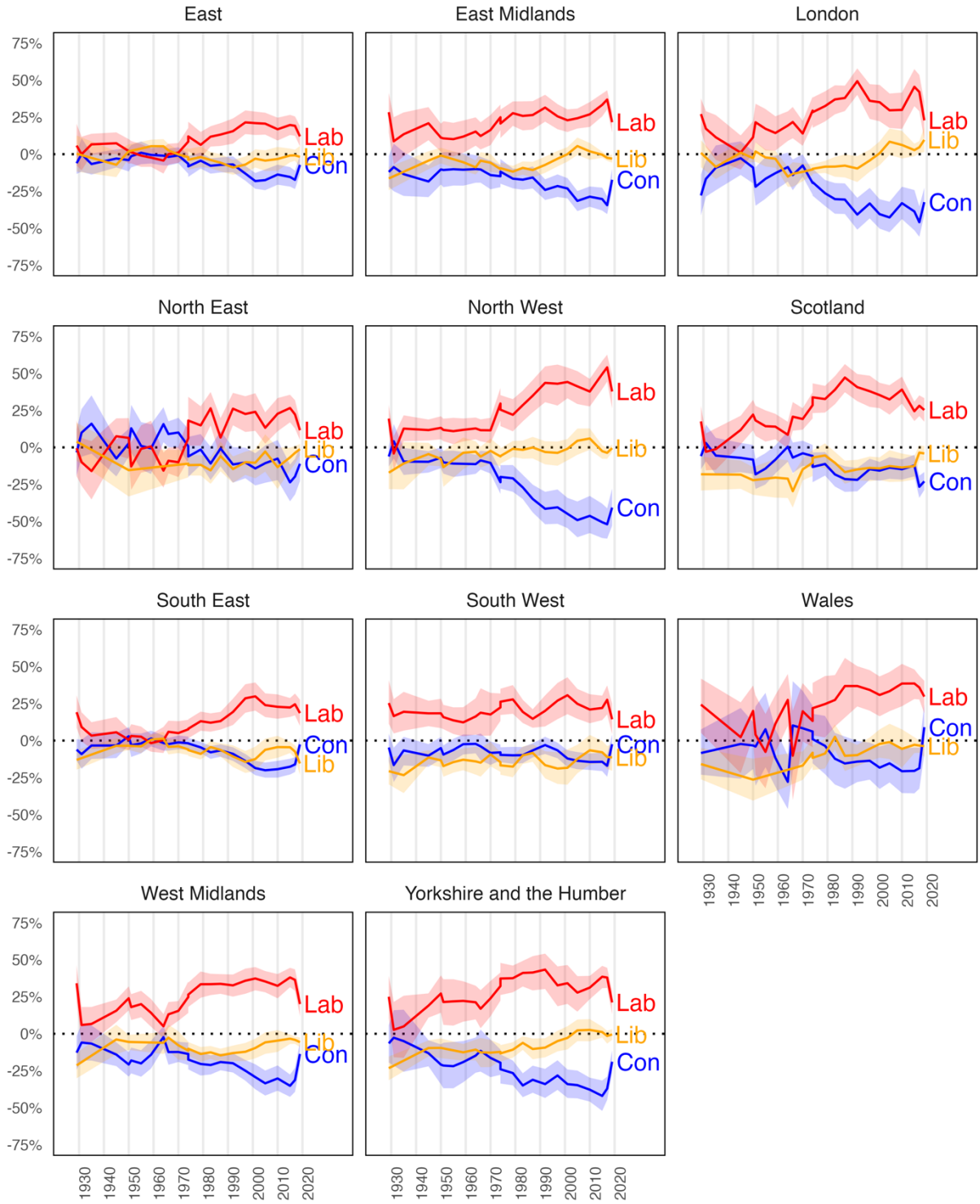
USA
It is common in the United States for fringe parties to endorse the candidate of one of the two major parties. Where endorsements are recorded in the source dataset, the endorsing party is grouped with the major party. Dozens of other parties have run candidates in elections, rarely

winning or attracting significant vote shares; these are grouped into a “minor” category for the purposes of the vote share analysis.

Group	**Parties**
Democrat	Democrat, Democrat and Progressive, Democrat Farmer-Labor, Democrat-Independent Progressive, Democrat-Republican, Independent Democrat, Labor, Social Democrat, Socialist, Socialist Equality
Republican	Republican, Conservative, Independent Republican, Libertarian, Republican-Democrat, Trump Conservative

Appendix C: Vote Share Analysis, United Kingdom by Region

Figure C1: Expected Shift in Party Vote Share, Lowest Density to Highest Density District, United Kingdom by Region



References

Adler, E. Scott. n.d. "Congressional District Data File, [congressional term]." University of Colorado.

Aitchison, J. 1982. "The Statistical Analysis of Compositional Data." *Journal of the Royal Statistical Society. Series B (Methodological)* 44 (2):139-177.

Boundary Commission for Scotland. 2011. "2nd Review Constituency Maps." https://www.bcomm-scotland.independent.gov.uk/?q=2nd-review-constituency-maps.

CQ Press. 2022. "Voting and Elections Collection." https://library.cqpress.com/elections/.

Craig, F.W.S. 1977. *British Parliamentary Election Results 1918-49*. London, UK: Palgrave Macmillan.

Lewis, Jeffrey B., Brandon DeVine, Lincoln Pritcher, and Kenneth C. Martis. 2013. "Digital Boundary Definitions of United States Congressional Districts, 1789-2012. Version 1.5.", Last Modified 22 Apr. 2020, accessed 3 June 2020. https://cdmaps.polisci.ucla.edu.

Library of Parliament. n.d. "Elections and Candidates." Government of Canada, accessed 28 May 2022. https://lop.parl.ca/sites/ParlInfo/default/en_CA/ElectionsRidings/Elections?permalink=662.

Manson, S., J. Schroeder, D. Van Riper, T. Kugler, and S. Ruggles. 2021. IPUMS National Historical Geographic Information System: Version 16.0 [dataset]. edited by IPUMS. Minneapolis, MN.

Ordnance Survey. 1974. Administrative areas diagram, showing the boundaries of the county, districts and civil parishes, or equivalent areas and the parliamentary constituencies [cartographic material]. Southampton, UK.

Ratcliffe, Mike. 2015. "The history of university representation." WONKHE, Last Modified 28 Jan. 2015. https://wonkhe.com/blogs/the-history-of-university-representation/.

Sevi, Semra. 2019. "Canadian Federal Candidates: 1867-2017." Harvard Dataverse. https://doi.org/10.7910/DVN/ABFNSQ.

Southall, Humphrey R. 2017. "A Vision of Britain Through Time / Data Access." University of Portsmouth. https://www.visionofbritain.org.uk/data/.

Taylor, Zack, Jack Lucas, J.P. Kirby, and Christopher Macdonald Hewitt. 2023. "Canada's Federal Electoral Districts, 1867–2021: New Digital Boundary Files and a Comparative Investigation of District Compactness." *Canadian Journal of Political Science* 56 (2):451–467. doi: https://doi.org/10.1017/S0008423923000185.

Tomz, Michael, Joshua A. Tucker, and Jason Wittenberg. 2002. "An Easy and Accurate Regression Model for Multiparty Electoral Data." *Political Analysis* 10 (1):66-83. doi: 10.1093/pan/10.1.66.

UK Data Service. 2018. "Census Support / Boundary Data Selector." https://borders.ukdataservice.ac.uk/bds.html.

van den Boogaart, K. Gerald, Raimon Tolosana-Delgado, and Matevz Bren. 2022. compositions (R package), version 2.0.4. http://www.stat.boogaart.de/compositions/.

Watson, Christopher, Elise Uberoi, and Philip Loft. 2020. "Research Briefing: General Election Results from 1918 to 2019." House of Commons Library, UK Parliament, Last Modified 17 Apr. 2020. https://commonslibrary.parliament.uk/research-briefings/cbp-8647/#fullreport.