

RESEARCH NOTE/NOTE DE RECHERCHE

A New Dataset on the Demographics of Canadian Federal Election Candidates

Anna Johnson¹ , Erin Tolley^{2*}, Melanee Thomas³ and Marc André Bodet⁴

¹Department of Political Science, University of Toronto, Sidney Smith Hall, Room 3018, 100 St. George Street, Toronto, ON M5S 3G3, ²Department of Political Science, Carleton University, B640 Loeb Building, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, ³Department of Political Science, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4 and ⁴Department of Political Science, Université Laval, Pavillon Charles-De Koninck local 3449, 1030 avenue des Sciences-Humaines, Québec, QC G1V 0A6
*Corresponding author. E-mail: erin.tolley@carleton.ca.

Abstract

This research note reports on a new effort to track candidate diversity in Canadian elections. The dataset covers 4,516 candidates who ran in the 2008, 2011, 2015 and 2019 federal elections and includes novel data on their race, Indigenous background and age, alongside information on gender, occupation, prior electoral experience and electoral outcome. We outline the process for collecting and systematizing these data, which relied on genealogical methods and the principle of triangulation. The data can be used to track diversity among electoral candidates over time or merged with other sources to answer district-level questions about representational diversity, electoral dynamics, vote choice and political communications. The method could also be replicated and applied to other levels of government.

Résumé

Cette note de recherche rend compte d'un nouvel effort pour suivre la diversité des candidats aux élections canadiennes. L'ensemble de données couvre 4 516 candidates et candidats qui se sont présentés aux élections fédérales de 2008, 2011, 2015 et 2019, et comprend des données inédites sur leur race, leur origine autochtone et leur âge, ainsi que des informations sur le genre, la profession, l'expérience électorale antérieure et les résultats de l'élection. Nous décrivons le processus de collecte et de systématisation de ces données, qui s'est appuyé sur des méthodes généalogiques et le principe de triangulation. Les données peuvent être utilisées pour suivre la diversité des candidats aux élections dans le temps ou fusionnées avec d'autres sources pour répondre à des questions au niveau des districts sur la diversité de la représentation, la dynamique électorale, le choix du vote et les communications politiques. La méthode pourrait également être reproduite et appliquée à d'autres paliers de gouvernement.

Keywords: race; ethnicity; gender; representation; Canadian politics

Mots-clés: race; ethnicité; genre; représentation; politique canadienne

Scholars frequently use measures of representational diversity to assess the health of democracies and to understand the extent to which legislatures reflect the populations they govern (Cross, 2011). In Canada, the Library of Parliament provides information on federal election candidates' gender and occupation, and these data have recently been systematized for all elections since 1867 (Sevi, 2021).¹ We contribute to this effort by adding new data on candidates' race, Indigenous background and age, alongside data on their gender, occupation, prior electoral experience and electoral outcome. The dataset includes the 4,516 candidates who ran for the major federal parties in the 2008, 2011, 2015 and 2019 Canadian elections. This research note describes the main features of the dataset, explains the procedures for collecting and coding the information within it, and outlines some of the potential applications.

The dataset's most significant contribution is its inclusion of information on candidate race and Indigenous background, a focus of study where Canadian political science has typically lagged behind (Ladner, 2017; Thompson, 2008). Idle No More, Black Lives Matter and the rise of white supremacy all suggest we ignore matters of identity at our peril (Andrew, 2017). Political parties typically do not collect (or at least opt not to publicly share) individual-level data on candidate demographics, and this has left gaps in the knowledge base. Our new dataset is therefore an important step forward, providing new information on the representation of identities in Canadian federal politics, which can be used to develop single-case studies as well as comparative benchmarking.

Other Canadian research tracks candidates' gender and racial background but includes only aggregate numbers, such as the total number of women or racialized candidates, and this prevents researchers from linking these data to other variables (Black, 2017). Like others (Lucas et al., 2021; Sevi, 2021), we expand the possibilities by providing publicly accessible data at the district level for all candidates, and we add novel data on racial and Indigenous backgrounds. These data can be tracked as dependent variables across multiple elections or included as independent variables to understand the relationship between candidate background and other political phenomena.

Data Collection and Variables

Researchers have experimented with several methods for gathering data on candidate demographics, including surveys (Andrew et al., 2008; Black and Hicks, 2006), probability-based classifiers (Lucas et al., 2021; Sevi, 2021) and genealogical approaches based on publicly available biographical materials and surname and photographic analysis (Black, 2008). Many researchers combine two or more of these techniques. Surveys offer highly reliable information because they are based on self-identification, but they often have considerable missing data given low response rates. Non-responses are of particular concern because they may be systematically linked to candidate demographics (Walgrave and Joly, 2018), and this is why we do not use a survey-based approach. Probability-based classifiers based on candidate name offer the advantage of automation and are most commonly applied to the study of gender using the R package *genderizeR* (Wais, 2016), but they require sample data to train the classifier and generate reliable

inferences. In the United States, researchers have extrapolated candidates' racial identities using the Census Bureau's Surname List (Grumbach and Sahn, 2020), but this resource does not exist in Canada, so researchers must rely on other methods (Besco and Tolley, 2020). Given this situation, we developed a rigorous expert coding procedure based on genealogical methods, an approach that has been recognized as the "gold standard" in tracking candidate demographics (Shah and Davis, 2017).

Although researchers in a number of contexts rely on genealogical methods to infer or ascribe a candidate's background, there is the potential for error and bias. To guard against these limitations, our coding procedure was based on the principle of triangulation: all demographic observations were cross-referenced across multiple sources. We required positive verification in at least two sources, such as a photograph and a biographical reference, prior to making a determination about a candidate's background. In cases where two positive verifications could not be obtained, the data are coded as missing.

We began by compiling candidate lists using official returns from Elections Canada. These returns also include election year, province, and electoral district name and number; these are all included in the dataset. To this, we added variables on incumbency, gender, racial and Indigenous background, occupation, office-holding at other levels of government, age and electoral outcome. Data for variables not included in official returns were gathered from the Library of Parliament, Elections Canada, official party and candidate websites, social media and news media. Coders collected the data, which were then cleaned, verified and standardized by one of the authors. Coding took place in several rounds and at different timepoints, with the procedures adjusted and improved as new sources of candidate information were discovered. At least two separate coders examined each individual data point, and the approach to data generation was team-based and iterative. For this reason, we do not report conventional measures of intercoder reliability but instead provide a transparent account of the procedures and decision rules we used. We conservatively estimate that the data collection and coding took more than 1,200 hours. The supplementary appendix includes additional details on the coding procedures and sources we consulted; in this research note, we highlight the main features of the substantive variables.

Party. From 2008 through 2015, the dataset includes candidates for the Bloc Québécois, Conservative, Liberal and New Democratic parties. In 2019, we added candidates for the Green party.

Incumbency. Candidates who held a seat in the House of Commons immediately preceding the election were coded as incumbents; all others were coded as non-incumbents.

Gender. Candidates' gender identity or expression at the time of their candidacy was coded using biographical information, news reports, photographs and first names. In 2008, 2011 and 2015, candidates were coded as either male/man or female/woman. In 2019, when nine candidates publicly identified as non-binary—meaning they do not identify strictly as a man or a woman—we added a third code to this variable.

Racial or Indigenous background. Candidates were classified as racialized, Indigenous or white. We based these constructs on the definitions of "visible minority" and "Aboriginal" that are included in Canada's Census Dictionary (Statistics Canada, 2017). While we retain the content of these definitions, the labels

themselves have been criticized and are often not the ones used by the communities to which they refer (Hennig, 2019; United Nations, 2007; Vowel, 2016). The categorizing and labelling of demographic characteristics are not without contention, but consistent with best practices in the relevant scholarship; we use “racialized” and “Indigenous” rather than “visible minority” and “Aboriginal” (see, for example, Gonzalez-Sobrinio and Goss, 2019; Thompson, 2016; Vowel, 2016).

Racialized candidates are those who are not Indigenous and are “non-Caucasian in race or non-white in colour” (Statistics Canada, 2017). This category includes those who are South Asian, Chinese, Black, Filipino, Latin American, Arab, Southeast Asian, West Asian, Korean or Japanese and those who have mixed racialized backgrounds. Indigenous candidates are those who are First Nations (for example, Status Indian, Cree, Ojibway), Métis or Inuit. Candidates who are neither racialized nor Indigenous were coded as white. These classifications rely on both self-identification and ascription. Some candidates self-identify as racialized or Indigenous in their personal biographies or in media interviews, and we would code their background accordingly. In instances where existing sources provided no identification of a candidate’s racial or Indigenous background, this variable was treated ascriptively and coded based on photographs and surnames, using the principle of triangulation outlined above.² Although higher levels of disaggregation are desirable, this goal is impeded by a lack of data on subcategories of candidate identification (for example, whether a candidate identified as Indigenous is Métis or First Nations), as well as the near impossibility of reliably inferring a candidate’s specific racial background; to achieve this aim, expert coding using genealogical methods is generally not appropriate. Thus, with some reservations, we use macro-level categories (“racialized” and “Indigenous”).

Occupation. We coded candidates’ primary occupation prior to their entrance into politics. We relied extensively on the Library of Parliament’s list of elections and candidates, which provides every federal candidate’s occupation following each election. Elected representatives sometimes report “Parliamentarian” or “politician” as their primary occupation. To better reflect candidates’ occupational background at the time they were first elected, we therefore sought information from other sources, including candidate and party websites, interviews and media reports. Occupational data are reported in nine broad categories: agriculture; business; education; government and politics; physicians and doctors; other health care; journalism; law; and other.

Age. We include a variable on candidate age, reported by year of birth. The Library of Parliament does not include age in its candidate database. Year of birth is available for elected members of Parliament, as well as for some candidates who previously served in a provincial legislature, and some candidates report their age in their official biographies or in media interviews. However, many do not disclose this information, and we are missing data on age for approximately one-third of the candidates in the database.

Other levels. To capture electoral experience at other levels of government, we include a dichotomous variable indicating whether candidates previously held office at the provincial or municipal level, or if they have not.

Electoral outcome and competitiveness. In each election, the candidate who won the district is coded as elected; all others are coded as having been defeated. We also

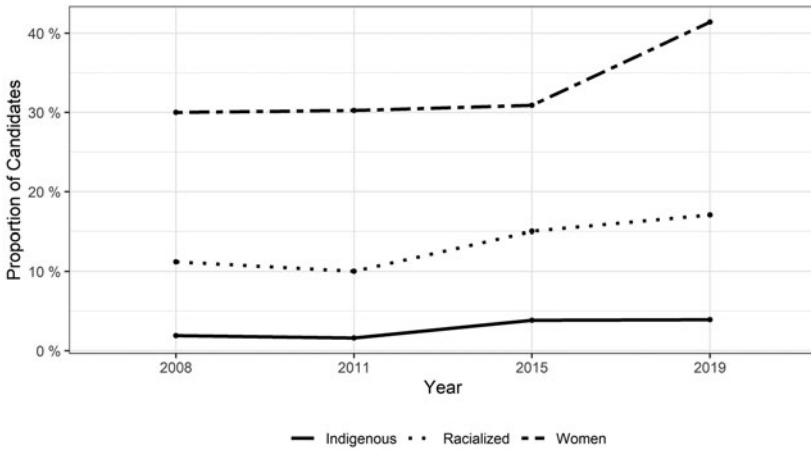


Figure 1 Women, Racialized and Indigenous Candidates in Canadian Elections

Note: For all elections, data capture information on candidates who ran for the Bloc Québécois, Conservative, Liberal and New Democratic parties, in addition to the Green party in 2019.

include a measure of competitiveness, which is the difference between a candidate's percentage of the vote and that of the victor. The information for both variables was derived from Elections Canada's official electoral returns.³

Applications

The dataset can be used, first, to understand and track diversity among electoral candidates. Figure 1 shows that since 2008, the proportion of candidates who are women, racialized or Indigenous has increased. The increase in women candidates between 2015 and 2019 is particularly notable and in contrast to the relative stagnation in this category in the preceding three election cycles. Other researchers can extend this analysis intersectionally to look at, for example, the presence of racialized women candidates in Canadian politics. Such an application would help to reveal whether racialized women are “doubly burdened” on account of their race and gender or, alternatively, whether they are “doubly advantaged,” with their race and gender providing party elites with an opportunity to tick two diversity boxes at once (Black and Erickson, 2006; Celis et al., 2014; Hughes, 2011). These questions have been posed in the comparative literature, but they have been less studied in Canada.

Given the centrality of political parties to candidate selection in Canada, the dataset also allows for comparisons across parties (Cross and Pruyers, 2019; Thomas and Bodet, 2013; Tolley, 2019). Figure 2 compares candidate diversity across the federal parties, with panels for the proportion of women, racialized and Indigenous candidates across the previous four federal elections. It shows the Bloc and Conservatives have consistently advanced the least racially diverse candidate slates and that they have generally nominated fewer women than the other parties. The uptick in racialized and Indigenous candidates in the 2019 election appears to have been led by the Liberals and the New Democrats, with both parties making large gains here, and the New Democratic party showing substantial

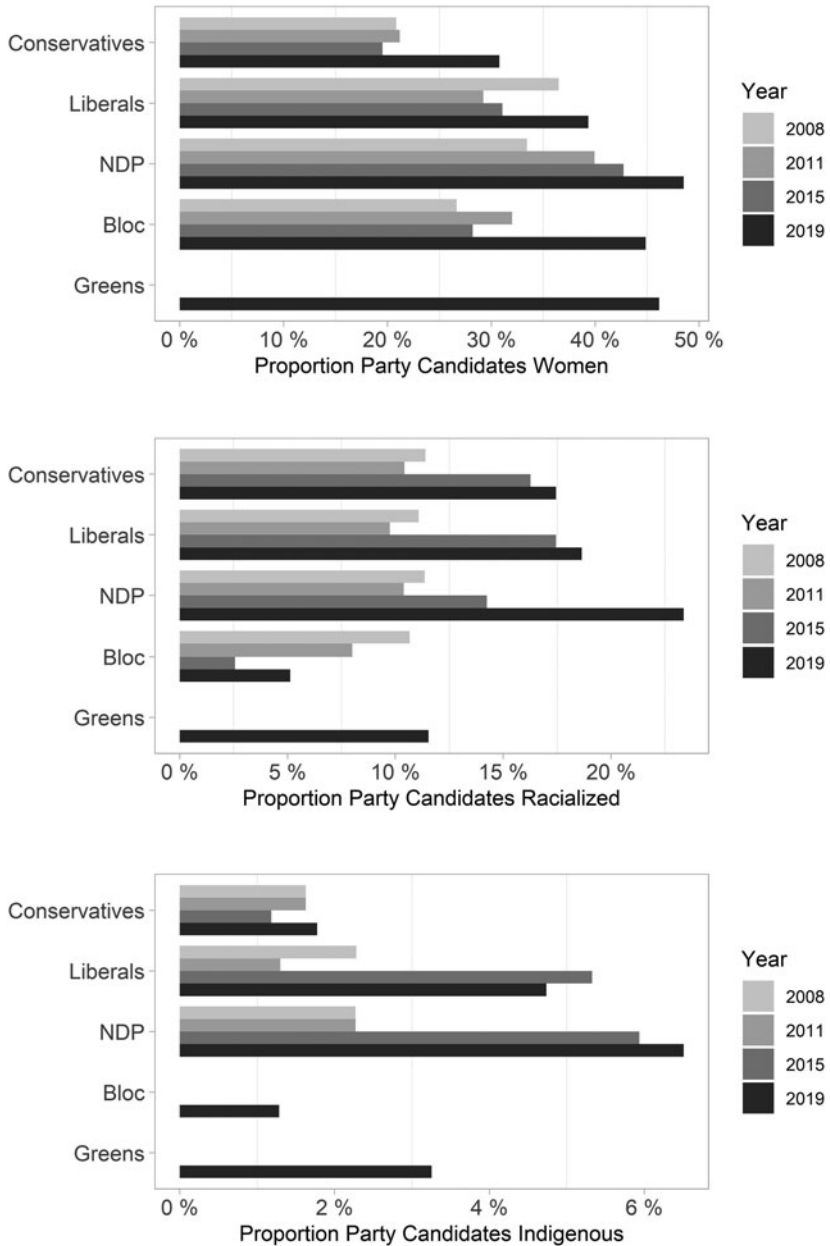


Figure 2 Candidate Diversity by Party

Note: For all elections, data capture information on candidates who ran for the Bloc Québécois, Conservative, Liberal and New Democratic parties, in addition to the Green party in 2019.

increases on both measures. Future research could use these data to probe parties’ strategies to recruit more diverse candidate slates and, in addition, advance new research on the competitive placement and outcome of these new recruits.

In addition to these descriptive possibilities, the dataset offers significant analytical leverage on a range of questions related to electoral dynamics, political donations, vote choice and media coverage. For example, by merging the candidate data with voter data from the Canadian Election Study, researchers could answer new questions on the relationship between candidate demographics, attitudinal measures and vote intention. What voter or district features are correlated with the election of more diverse candidates? Are more urban districts or those with more progressively minded voters more open to candidates who break the demographic mould? Researchers could also merge our candidate data with Elections Canada's district-level data on voter turnout, financial contributions and vote totals, in order to understand the relationship between these dependent variables and candidate demographics. Is voter turnout higher in districts with more diverse candidate slates? Do candidates with historically underrepresented backgrounds attract fewer donations? Candidate data could be merged with textual data, such as candidate mentions on Twitter or in election coverage, to track candidates' portrayal and self-presentation, or with a new digital record of Hansard to track demographic patterns in parliamentary communication (Beelen et al., 2017).

Finally, the dataset offers possibilities for comparative analysis. It could be used alongside candidate data collected in other countries (for example, Center for American Women and Politics, 2020) or with data on the gender of office-holders worldwide (for example, Inter-Parliamentary Union, 2019) to better understand the relationship between representational diversity, political institutions and public attitudes.

Contribution

Candidates are at the centre of Canadian electoral politics. Despite diversity being a key feature of the country's composition, researchers to date have had limited access to data on candidate demographics, and the time required to gather such data may have deterred additional inferential analysis. This dataset remedies the problem, providing a rigorous, reliable and comprehensive source of information on candidate race, Indigenous background and age, alongside data on gender, occupation, electoral experience and outcomes. On their own, these data provide a benchmark for understanding representational diversity in Canada, but they also open up avenues for better understanding the political salience of identities on questions that are key to the study of electoral politics. Finally, by detailing the range of sources that were consulted, the coding procedures we used and the principles adopted to make decisions about contentious categories, this research note helps to "pull back the curtain" on the creation of new sources of data. The methods we describe here could be applied to the collection of other candidate data, including at other levels of government. By documenting the data collection, this research note provides a basis for replication and, we hope, a move toward more institutionalized data collection on candidate diversity.

Supplementary material. To view supplementary material for this article, please visit <https://doi.org/10.1017/S0008423921000391>.

Data Availability Statement. The full dataset is available at: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/MI5XQ6>.

Acknowledgments. Funding for this project was provided by the University of Toronto and the Social Sciences and Humanities Research Council (grant #430-2017-00504). Kamille Leclair and Samuel Moir capably assisted with the collection and coding of candidate data.

Notes

1 At the time of writing, the Library of Parliament's "Elections and Candidates" section was available at https://lop.parl.ca/sites/ParlInfo/default/en_CA/ElectionsRidings/Elections, but we discovered the website is updated frequently and that the link is unstable.

2 During the 2019 election, at least four candidates faced allegations that they were falsely claiming Indigenous identity (Barrera and Deer, 2019). Those candidates who recanted their claims of Indigenous identity were not coded as Indigenous.

3 Peter Loewen, Daniel Rubenson and Md Mujahedul Islam generously shared data related to the official results of the 2019 general election.

References

- Andrew, Caroline. 2017. "Canada and Society." *Canadian Journal of Political Science* 50 (1): 329–43.
- Andrew, Caroline, John Biles, Myer Siemiatycki and Erin Tolley, eds. 2008. *Electing a Diverse Canada: The Electoral Representation of Immigrants, Minorities and Women*. Vancouver: UBC Press.
- Barrera, Jorge and Ka'nehsisio Deer. 2019. "4 Federal Candidates Accused of Indigenous Identity Appropriation by Halifax Academic." *CBC News*, October 10. <https://www.cbc.ca/news/indigenous/federal-candidates-claims-indigenous-identity-1.5314614>.
- Beelen, Kaspar, Timothy Alberdingk Thijm, Christopher Cochrane, Kees Halvemaan, Graeme Hirst, Michael Kimmins, Sander Lijbrink, Maarten Marx, Nona Naderi, Ludovic Rheault, Roman Polyanovsky and Tanya Whyte. 2017. "Digitization of the Canadian Parliamentary Debates." *Canadian Journal of Political Science* 50 (3): 849–64.
- Besco, Randy and Erin Tolley. 2020. "Ethnic Group Differences in Donations to Electoral Candidates." *Journal of Ethnic and Migration Studies*. FirstView at <https://doi.org/10.1080/1369183X.2020.1804339>.
- Black, Jerome H. 2008. "Ethnoracial Minorities in the 38th Parliament: Patterns of Change and Continuity." In *Electing a Diverse Canada: The Representation of Immigrants, Minorities and Women*, ed. Caroline Andrew, John Biles, Myer Siemiatycki and Erin Tolley. Vancouver: UBC Press.
- Black, Jerome H. 2017. "The 2015 Federal Election: More Visible Minority Candidates and MPs." *Canadian Parliamentary Review* 40 (1): 16–23.
- Black, Jerome H. and Lynda Erickson. 2006. "Ethno-racial Origins of Candidates and Electoral Performance: Evidence from Canada." *Party Politics* 12 (4): 541–61.
- Black, Jerome H. and Bruce M. Hicks. 2006. "Visible Minority Candidates in the 2004 Election." *Canadian Parliamentary Review* 29 (2): 26–31.
- Celis, Karen, Silvia Erzeel, Liza Mügge and Alyt Damstra. 2014. "Quotas and Intersectionality: Ethnicity and Gender in Candidate Selection." *International Political Science Review* 35 (1): 41–54.
- Center for American Women and Politics. 2020. 2020 Women Candidates for U.S. Congress and Statewide Elected Executive Office. <https://cawp.rutgers.edu/election-2020-women-candidates-us-congress-state-wide-elected-executive>
- Cross, William, ed. 2011. *Auditing Canadian Democracy*. Vancouver: UBC Press.
- Cross, William P. and Scott Pruyers. 2019. "The Local Determinants of Representation: Party Constituency Associations, Candidate Nomination and Gender." *Canadian Journal of Political Science* 52 (3): 557–74.
- Gonzalez-Sobrin, Bianca and Devon R. Goss. 2019. "Exploring the Mechanisms of Racialization beyond the Black–White Binary." *Ethnic and Racial Studies* 42 (4): 505–10.
- Grumbach, Jacob M. and Alexander Sahn. 2020. "Race and Representation in Campaign Finance." *American Political Science Review* 114 (1): 206–21.
- Hennig, Clare. 2019. "StatsCan Looks to Modernize Decades-Old Term 'Visible Minority' When Measuring Diversity." *CBC News*, May 8. <https://www.cbc.ca/news/canada/british-columbia/statcan-modernize-diversity-visible-minority-1.5128288>.
- Hughes, Melanie M. 2011. "Intersectionality, Quotas, and Minority Women's Political Representation Worldwide." *American Political Science Review* 105 (3): 604–20.
- Inter-Parliamentary Union. 2019. "Women in National Parliaments (as of 1 January 2019)." <http://www.ipu.org/wmmn-e/classif.htm> (April 2, 2019).

- Ladner, Kiera L. 2017. "Taking the Field: 50 Years of Indigenous Politics in the *CJPS*." *Canadian Journal of Political Science* 50 (1): 163–79.
- Lucas, Jack, Reed Merrill, Kelly Blidook, Sandra Breux, Laura Conrad, Gabriel Eidelman, Royce Koop, Daniella Marciano, Zack Taylor and Salomé Vallette. 2021. "Women's Municipal Electoral Performance: An Introduction to the Canadian Municipal Elections Database." *Canadian Journal of Political Science* 54 (1): 125–33.
- Sevi, Semra. 2021. "Who Runs? Canadian Federal and Ontario Provincial Candidates from 1867 to 2019." *Canadian Journal of Political Science*. Advance online publication. <https://doi.org/10.1017/S0008423920001213>.
- Shah, Paru R. and Nicholas R. Davis. 2017. "Comparing Three Methods of Measuring Race/Ethnicity." *Journal of Race, Ethnicity, and Politics* 2 (1): 124–39.
- Statistics Canada. 2017. *Dictionary, Census of Population, 2016*. Ottawa: Minister of Industry, <https://www150.statcan.gc.ca/n1/en/catalogue/98-301-X>.
- Thomas, Melanee and Marc André Bodet. 2013. "Sacrificial Lambs, Women Candidates, and District Competitiveness in Canada." *Electoral Studies* 32 (1): 153–66.
- Thompson, Debra. 2008. "Is Race Political?" *Canadian Journal of Political Science* 41 (3): 525–47.
- Thompson, Debra. 2016. *The Schematic State: Race, Transnationalism and the Politics of the Census*. Cambridge: Cambridge University Press.
- Tolley, Erin. 2019. "Who You Know: Local Party Presidents and Minority Candidate Emergence." *Electoral Studies* 58: 70–9.
- United Nations. 2007. *Report of the Committee on the Elimination of Racial Discrimination*. New York: United Nations.
- Vowel, Chelsea. 2016. *Indigenous Writes: A Guide to First Nations, Métis and Inuit Issues in Canada*. Winnipeg: HighWater Press.
- Wais, Kamil. 2016. "Gender Prediction Methods Based on First Names with GenderizeR." *The R Journal* 8 (1): 17–37.
- Walgrave, Stefaan and Jeroen K. Joly. 2018. "Surveying Individual Political Elites: A Comparative Three-Country Study." *Quality and Quantity* 52 (5): 2221–37.

Cite this article: Johnson, Anna, Erin Tolley, Melanee Thomas and Marc André Bodet. 2021. "A New Dataset on the Demographics of Canadian Federal Election Candidates." *Canadian Journal of Political Science* 54 (3): 717–725. <https://doi.org/10.1017/S0008423921000391>