CAMBRIDGE
UNIVERSITY PRESS

RESEARCH NOTE/NOTE DE RECHERCHE

Who Runs? Canadian Federal and Ontario Provincial Candidates from 1867 to 2019

Semra Sevi* iD

Département de science politique, Pavillon Lionel-Groulx, 3150 rue Jean-Brillant, Montréal, QC H3T 1N8
*Corresponding author. Email: semra.sevi@umontreal.ca

Abstract
Who runs and is elected is one of the most fundamental questions in political science as it pertains to the issue of descriptive representation. Despite the importance of this issue, until recently there were no longitudinal datasets on candidates in Canadian elections. This article presents two novel datasets including information on all candidates who ran in Canadian federal and Ontario provincial elections from 1867 to 2019. I present how these data were collected and how they can be used to gain new insights. I expect these data will be a valuable resource to Canadian political scientists for both research and teaching purposes.

Résumé
La question de savoir qui se présente et qui est élu est l'une des questions les plus fonda-mentales de la science politique en ce qui concerne l'enjeu de la représentation descriptive. Malgré son importance, il n'existait jusqu'à récemment aucun ensemble de données longitudinales sur les candidats aux élections canadiennes. Cet article présente deux nouveaux ensembles de données comprenant des informations sur tous les candidats qui se sont présentés aux élections fédérales canadiennes et aux élections provinciales de l'Ontario de 1867 à 2019. Je présente la manière dont ces données ont été recueillies et comment elles peuvent être utilisées pour obtenir de nouvelles informations. Je pense que ces données constitueront une ressource précieuse pour les politologues canadiens, tant aux fins de la recherche que de l'enseignement.

Introduction

In this article, I present two new datasets that contain individual-level data on all candidates between 1867 and 2019 in both the Canadian federal and Ontario pro-vincial elections. In total, the data provide information on 44,462 candidates at the federal level and 15,529 in the Ontario provincial elections.[1] While there are

existing studies on political elites in Canada, until now there were no longitudinal datasets covering all candidates going back to 1867.

These data will allow researchers to investigate a number of important topics, including: the share of female candidates over an extended period of time; which occupations do better in politics (Sevi, Blais and Mayer, 2020); if women get fewer votes than men (Sevi, Arel-Bundock and Blais, 2018; Sevi, Blais and Arel-Bundock, 2021); how long politicians tend to stay in politics; the advantage gained by incumbency; how well independents do; the consequences for politicians who are elected under a party banner and then switch their party affiliation and run again either at the same level of government (Sevi, Yoshinaka and Blais, 2018) or across two different levels (provincial and federal); the progressive ambition of politicians across different levels of elections; and if by-elections are more favourable to smaller parties and/or independent candidates.

Data Collection

The original sources for these datasets are the Library of Parliament of Canada and Elections Ontario. The federal data come from the Library of Parliament and consist of the names of all the candidates, the date of the election, the number of ballots cast, the occupations of the candidates, and the name of the constituency, province and party affiliation. I manually recorded these data twice between 2014 and 2017.[2]

I then completed an extensive cleaning of the data, using candidate websites, historical newspaper archives[3] and other journalist summaries of the candidates; in doing so, I also added variables that include the gender for all candidates, birth year for elected politicians[4] and whether the candidates were acclaimed or switched parties after being elected. I also included the parliament number, and I calculated the percentage of the vote obtained by the candidates, as well as whether the candidate was elected or not. Moreover, I assigned unique IDs to all candidates and matched the IDs of the same individuals over time. Unique IDs help mitigate several problems: first, many candidates' names were not spelled consistently across different elections in the Library of Parliament database—for example: the same name could be given as John A., J. A., Sir John, or J.; furthermore, sometimes different candidates have the same name, or—in earlier elections—the same candidate may have run in different ridings in different years, and sometimes even in the same year.

To assign unique IDs, I looked up every candidate's profile in the Library of Parliament and used alternative biographical information to triangulate their identity. Because I manually recorded the unique IDs, I was also able to create an incumbent variable that indicates whether the candidate ran in the previous election. The names of parties are also spelled differently across different elections, so a similar treatment was necessary—for example: Liberal, Liberal Party of Canada, Liberal Progressive, Opposition, Opposition/Laurier Liberals, and so on. I give researchers the option to use either the 155 unique party names or the categories I created that put together similar parties but also parties that are named differently across different elections. All the data were independently checked at another time to ensure accuracy.

Researchers may reasonably express concerns about the quality of my federal dataset, given that it is gathered manually. To address these concerns, I re-collected all the variables after six months of not touching the dataset and merged these with my initial data collection. This second step was an opportunity to verify my initial data collection.

Ontario is the province in Canada where the constituencies and parties are most similar to those at the federal level; therefore, in 2019, I collected similar data from 1867 to 2018 for the Ontario provincial elections. The original PDF documents that were retrieved from Elections Ontario contained the candidates' name, party, constituency, date of the election and the number of ballots cast. To this dataset, I added gender by first making use of the R package, genderizerR (Wais, 2016),[5] which infers the gender of candidates by analyzing first names. GenderizerR is based on the genderize.io API, which is a web scraping tool (http://genderize.io). GenderizerR provides a likely gender and probability score for each candidate. I kept all the probabilities. I then verified each entry on two different occasions. I kept all the entries and not simply probabilities that are close to 100 per cent. (Both the genderizeR probabilities and my manual check are retained in the Ontario dataset. I made a total of 2,376 corrections to the gender variable.[6])

Applications

The data presented in this paper allow political scientists to better understand political phenomena in Canada. They can be used to replicate existing studies on political elites in Canada and verify whether patterns hold over an extended period. These studies cover questions such as minority representation (Black and Erickson, 2006; Black, 2013), women's representation (Blais and Gidengil, 1991; Hunter and Denton, 1984; Thomas and Bodet, 2013; Tremblay and Trimble, 2006; Tolley, 2011) and incumbency advantages (Kendall and Rekkas, 2015). Much of the existing literature on elites in Canada has examined a few elections or a few variables because longitudinally rich data were previously not collected or available. My data are the first to contain information on all candidates in all the elections since 1867 and therefore make available a unique tool for researchers focusing on political elites in Canada. As an example, there is a large body of literature on the inclusion of women in politics (Blais and Gidengil, 1991; Stockemer, 2017; Trimble and Arscott, 2003; Trimble et al., 2013; Tolley, 2011); however, my data are the first to offer longitudinal data on the number of female candidates and their success compared to their male counterparts.

Figure 1 highlights four applications of my federal dataset that are of particular relevance. Given that the federal dataset is unique not only in terms of the variables collected but also in its longitudinal nature, it can be used for all types of exploratory and descriptive questions. First, in the upper left corner, we see that the mean number of candidates per constituency has increased over time, but since 1997 it has decreased, with an uptick in the 2019 election. Second, the share of incumbents has also decreased over time. This decrease can partially be explained by the increase in the number of candidates and the increasing competitiveness of elections, but it is worth exploring further in future studies. Third, since 1921, the share of elected women in Parliament has increased from .004 in 1921, when the
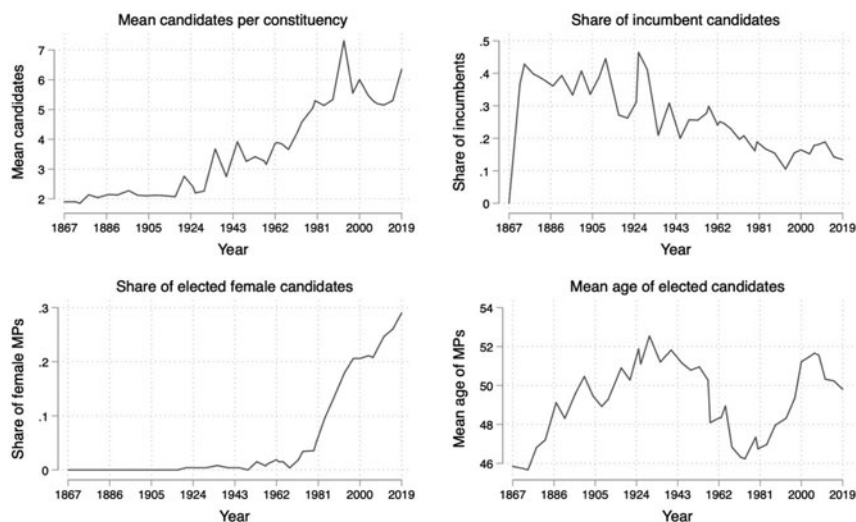
Figure 1. Four Applications of the Federal Dataset.

first woman was elected to Parliament, to .29 in 2019. Fourth, the federal dataset will be useful for researchers interested in the age of Members of Parliament (MPs) over time. The mean age of elected MPs varies between 45 and 53. There is a notable decline and rebound in the mean age, reaching a minimum in the 1970s, which is worth exploring further.

These data are flexible in that it is possible to create many more variables or to merge them with other existing data. The data could be merged with existing datasets using candidate and constituency names. For example, a researcher could merge the federal dataset with data from the Canadian Census at the level of federal electoral districts to examine the characteristics of MPs' constituencies compared to MPs themselves. It could also be combined with data on nomination contests to study how parties select and/or appoint candidates. Or it could be merged with administrative data to study campaign contributions across gender (Tolley et al., 2020). Comparative analyses are also possible given that similar datasets exist for other countries (see Klarner, 2018; Kollman et al., 2019; Yoshinaka, 2016). Finally, the data can be used as a teaching tool in introductory courses in Canadian politics, as well as in research methods.

Conclusion

I have introduced two new datasets that are the largest available data on candidates in Canadian and Ontario elections, respectively. My data cover the period 1867–2019 and contain detailed information on candidates in both federal elections and Ontario provincial elections. These are unique data that can be used by researchers as well as nonacademic stakeholders such as journalists to address substantive research questions about gender, incumbency, the careers of candidates over time, and so on. I created uniform datasets with standardized information

about all candidates who run in elections in Canada federally and in the province of Ontario.

So far, these data have been used for five peer-reviewed manuscripts (Sevi, Yoshinaka and Blais, 2018; Sevi, Arel-Bundock and Blais, 2018; Sevi, Blais and Mayer, 2020; Tolley, Besco and Sevi, 2020; Sevi, Blais and Arel-Bundock, 2021); they have also been used in a number of presentations at academic conferences, in reports by think tanks,[7] by journalists,[8] and in many papers currently online or in the pipeline. As such, these data have been scrutinized by different researchers.

The data are available on the Harvard Dataverse here: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ABFNSQ

Notes

1 The data include all candidates at both levels, irrespective of party affiliation, including individuals who ran and lost.

2 I added the 2019 data in 2020.

3 I used all the national newspapers, which include the *Globe and Mail*, the *National Post*, the *Star* and *Maclean's*, as well as community newspapers in each riding.

4 The birth year of all candidates is not consistently available; therefore, I decided to collect this data only for elected MPs in the federal dataset.

5 A different approach is used here to record the genders for the provincial data because by 2019 the genderizeR package was widely used by researchers and included Canadian names, which was not the case previously.

6 Note that in earlier elections—in 1867, for example—it was common for the first name of a candidate to be recorded by first initial; in these cases, genderizeR cannot pick up the gender. But since we know that the first female candidate in Ontario did not appear on the ballot until 1902, I was able to manually make everyone before this date a male and then look up the candidates starting in 1902. I did not collect age for the Ontario dataset because such information is very hard to compile for elected parliamentarians at the provincial level.

7 For example: https://www.samaracanada.com/docs/default-source/reports/party-favours-by-the-samara-centre-for-democracy.pdf.

8 For example: https://policyoptions.irpp.org/magazines/september-2018/research-shows-that-mps-who-cross-the-floor-lose-votes/ and https://www.thestar.com/politics/federal/2019/10/22/as-an-independent-mp-can-jody-wilson-raybould-make-a-difference.html.

References

Black, Jerome H. 2013. "Racial Diversity in the 2011 Federal Election: Visible Minority Candidates and MPs." *Canadian Parliamentary Review* **36** (3): 21–26.

Black, Jerome H. and Lynda Erickson. 2006. "Ethno-Racial Origins of Candidates and Electoral Performance: Evidence from Canada." *Party Politics* **12** (4): 541–61.

Blais, André and Elisabeth Gidengil. 1991. *Making Representative Democracy Work: The Views of Canadians.* Vol. 17 of the *Research Studies of the Royal Commission on Electoral Reform and Party Financing.* Toronto: Dundurn Press.

Hunter, Alfred A. and Margaret A. Denton. 1984. "Do Female Candidates 'Lose Votes'? The Experience of Female Candidates in the 1979 and 1980 Canadian General Elections." *Canadian Review of Sociology* 21 (4): 395–406.

Kendall, Chad and Marie Rekkas. 2015. "Incumbency Advantages in the Canadian Parliament." *Canadian Journal of Economics* 45 (4): 1560–85.

Klarner, Carl. 2018. "State Legislative Election Returns, 1967–2016." https://doi.org/10.7910/DVN/3WZFK9, Harvard Dataverse, V3.

Kollman, Ken, Allen Hicken, Daniele Caramani, David Backer and David Lublin. 2019. "Constituency-Level Elections Archive (CLEA)." Produced and distributed by Center for Political Studies, University of Michigan, Ann Arbor, MI. http://www.electiondataarchive.org.

Sevi, Semra, Vincent Arel-Bundock and André Blais. 2018. "Do Women Get Fewer Votes? No." *Canadian Journal of Political Science* 52 (1): 201–10.

Sevi, Semra, André Blais and Vincent Arel-Bundock. 2021. "Do Women Get Fewer Votes in Ontario Provincial Elections?" In *Women, Power, and Political Representation: Canadian and Comparative Perspectives*, ed. Roosmarijn de Geus, Erin Tolley, Elizabeth Goodyear-Grant and Peter Loewen. Toronto: University of Toronto Press.

Sevi, Semra, André Blais and Danielle Mayer. 2020. "Do Lawyers Get More Votes?" *American Review of Canadian Studies* 50 (2): 216–28.

Sevi, Semra, Antoine Yoshinaka and André Blais. 2018. "Legislative Party Switching and the Changing Nature of the Canadian Party System, 1867–2015." *Canadian Journal of Political Science* 51 (3): 665–95.

Stockemer, Daniel. 2017. "The Proportion of Women in Legislatures and Cabinets: What Is the Empirical Link?" *Polity* 49 (3): 434–60.

Thomas, Melanee and Marc André Bodet. 2013. "Sacrificial Lambs, Women Candidates, and District Competitiveness in Canada." *Electoral Studies* 32 (1): 153–66.

Tolley, Erin. 2011. "Do Women 'Do Better' in Municipal Politics? Electoral Representation across Three Levels of Government." *Canadian Journal of Political Science* 44 (3): 573–94.

Tolley, Erin, Randy Besco and Semra Sevi. 2020. "Who Controls the Purse Strings? A Longitudinal Study of Gender and Donations in Canadian Politics." *Politics & Gender*, First View, November 16: 1–29. https://doi.org/10.1017/S1743923X20000276.

Tremblay, Manon and Linda Trimble. 2006. "Still Different after All These Years? A Comparison of Female and Male Canadian MPs in the Twentieth Century." *Journal of Legislative Studies* 10 (1): 97–122.

Trimble, Linda and Jane Arscott. 2003. *Still Counting: Women in Politics across Canada.* Toronto: University of Toronto Press.

Trimble, Linda, Jane Arscott and Manon Tremblay. 2013. *Stalled: The Representation of Women in Canadian Governments.* Vancouver: UBC Press.

Wais, Kamil. 2016. "Gender Prediction Methods Based on First Names with genderizeR." *The R Journal* 8 (1): 17–37.

Yoshinaka, Antoine. 2016. *Crossing the Aisle: Party Switching by U.S. Legislators in the Postwar Era.* New York: Cambridge University Press.