

The Impact of Voter Turnout on the 2019 Canadian Federal Election: A Logistic Regression & Poststratification Analysis

Amber Kao

2020/12/22

Voter turnout is an important issue to be investigated because votes directly decide the governing body of Canada. In this analysis, logistic regression, poststratification, and propensity score matching are conducted with Canadian Election Study (CES) and General Social Survey (GSS) data to investigate the predictability of vote choice with the demographic factors age, gender, province, education level, language knowledge, and religious affiliation. All factors were found to correlate significantly with vote choice, the Conservative Party was predicted to gain more of the popular vote if all eligible voters had voted in the 2019 Canadian federal election, and religious affiliation was found to have a causal relationship with voting for the Conservative Party but not the Liberal Party.

Code and data supporting this analysis is available at: https://github.com/amber-kao/2019_canadian_election

Keywords

Logistic Regression, Poststratification, Canadian Election, Voter Turnout, Propensity Score Matching

Introduction

A democracy, as defined by the Merriam-Webster dictionary (n.d.), is both a “government by the people” and a “rule of the majority.” These two concepts may seem analogous, but when individuals decline to exercise their right to vote, the vote outcome may not be truly representative of the majority. To ensure that the governing body enacts policies to the best interest of the majority, it is imperative that every eligible individual votes. Eligible voters are Canadian citizens at least 18 years of age. If politicians are elected by voters that are not truly representative of the people with the right to vote, then the policies enacted may not be in the interest of the majority. Therefore, the following analysis provides insight into the potentially large impact of low voter turnout on major elections in a democratic country and demonstrates the vital importance of exercising the right to vote.

Canada is a constitutional monarchy ruled by the Queen, but since the Queen does not have governing power and the representatives in the governing body, the House of Commons, are elected by citizens, Canada is essentially a democratic country (Elections Canada, 2020b). There are 338 electoral districts in total, also known as ridings, and the representative receiving the most votes in a riding wins a seat in the House of Commons. The leader of the party that obtains the most seats becomes the Prime Minister. In the 2019

Canadian federal election, the Liberal Party won 157 out of 338 total seats with 33.1% of the popular vote, and its leader, Justin Trudeau, remained Prime Minister for a second term. However, since 170 seats are needed for any party to reach a majority government, Justin Trudeau now leads a Liberal minority government (Elections Canada, 2020c; Zimonjic, 2019). The Conservative Party, the party with the next highest number of seats, won 121 seats, but received 34.4% of the popular vote (CBC, n.d.). In the 2019 election, voter turnout was 18,350,359 out of 27,373,058 eligible voters, which was approximately 67% (Elections Canada, 2020a). This voter turnout percentage is surprising considering Canada's relatively long history of democracy, with its first election held in 1867 (Elections Canada, 2020a). It also raises the questions—had the remaining 33% of voters also voted, which party would have obtained the most votes, and would Justin Trudeau have retained his position as Prime Minister?

In the following analysis, logistic regression, poststratification, and propensity score matching models are constructed in R using data from the 2019 Canadian Election Study (CES) (Hodgetts, 2020; Stephenson, 2020a; Stephenson, 2020b) and the 2017 Canadian General Social Survey (GSS) (CHASS, n.d.; Statistics Canada, 2020a) to predict the popular vote percentage for the Liberal Party and the Conservative Party during the 2019 Canadian election, had every eligible individual voted. In other words, which party would have received the most votes, had voter turnout been 100%?

In the Methodology section, I introduce the CES and GSS data sets and their base characteristics. I also describe the methodology behind the logistic regression models, poststratification analysis, and propensity score matching calculations. In the Results section, I present the major results of the statistical analyses performed. In the Discussion section, I discuss the implications and broader impacts of the results and draw conclusions. I also reflect on the weaknesses and limitations of the analytical methods and propose future directions for research. Logistic regression models showed that age, gender, province of residence, education level, knowledge of English and/or French, and religious affiliation are all significant factors in influencing voter preference for either the Liberal or the Conservative Party. Poststratification analysis showed that if everyone had voted in the 2019 election, the Conservative Party would have received more votes compared to the Liberal Party. Propensity score matching determined that religious affiliation is an indicator for vote preference for the Conservative Party, but not for the Liberal Party.

Methodology

Data

Data from the 2019 online CES was used for logistic regression analysis and propensity score matching (Hodgetts, 2020; Stephenson, 2020a; Stephenson, 2020b). The data was obtained using the `cesR` package (Hodgetts, 2020). The CES was conducted online through the Qualtrics platform, within the two months leading up to the election, and was available in both English and French (Stephenson, 2020a; Stephenson, 2020b). The target population was Canadian citizens who were at least 18 years old, the frame population was those with

access to the online survey, and sample population consisted of the 37,822 respondents of the survey. The survey contained more than 600 questions regarding respondents' demographic background and political opinions, most of which had multiple choice answers and very few of which allowed open-ended answers. Low quality survey responses were removed, and non-response problems were avoided with the option to choose "Don't know/prefer not to answer" for most of the questions. A strength of this data set is its wealth of information regarding respondents' political opinions, and a weakness is the unnecessary separation of some responses into multiple variables.

The year of birth, gender, province/territory, education level, knowledge of English and/or French, religious affiliation, and vote choice variables were chosen from the CES data set for analysis. Data cleaning procedures included excluding non-citizens, transforming the year of birth data into five age groups, excluding genders other than female or male, excluding observations from the three territories, re-labelling education levels, combining the English knowledge and French knowledge variables into one, and transforming both religious affiliation and vote choice into binary variables for analysis. These data manipulations were mostly performed to match with the GSS data, and 36,120 observations were retained for analysis.

Data from the 2017 GSS was used as census data for poststratification analysis (CHASS, n.d.; Statistics Canada, 2020a). The data was obtained from CHASS, then cleaned initially with R code written by Rohan Alexander and Sam Caetano (CHASS, n.d.; Alexander & Caetano, 2020). The GSS was conducted through telephone interviews, from February through November 2017, and was available in both English and French (Statistics Canada, 2020a). The target population consisted of the 30,633,177 Canadian residents of the 10 provinces who were at least 15 years old, the frame population consisted of those whose phone numbers were available to Statistics Canada, and the sample population consisted of the 20,602 individuals who responded to the phone interview (Statistics Canada, 2020b). The survey contained questions relating to the demographics, lifestyle, and familial relationships of the surveyed individuals and their families. Individuals were able to decline to participate in the survey or withhold answers, which were recorded as missing responses. A strength of this data set is its even distribution of samples across different demographics, and a weakness of this data is the relatively small sample size compared to the CES data. Another weakness of the GSS data is that it only records sex instead of gender, which does not match with the CES data that records gender. Since we are interested in gender as a social construct instead of the biological characteristics, recording sex instead of gender in the census results in misrepresentation of the experiences of individuals whose sex and gender may be different.

The age, sex, province, education level, knowledge of English and/or French, and religious affiliation variables were chosen from the GSS data set for analysis, to match with the variables chosen for the CES data. Data cleaning procedures included transforming age data into five age groups, changing the sex variable to gender, re-labelling education levels, and transforming religious affiliation into a binary variable for analysis. These manipulations were also mostly performed to match with the CES data. The data was then grouped into all the possible combinations of demographic bins and the number of individuals in each bin was counted. Out of 5,600 possible bin combinations, only 2,061 contained at least one

person. Table 1 shows the final variables and categories chosen from the two data sets that were used for analysis.

Table 1: Key characteristics of the CES and GSS data

Age	Language Knowledge	Religious Affiliation	Gender	Vote Choice ¹
18-29	English only	Yes (1) ²	Female	1 (Liberal/Conservative) ³
30-44	French Only	No (0)	Male	0 (Other)
45-59	Both English and French			
60-74	Neither English nor French			
75+				
Province	Education Level			
Alberta	Less than high school			
British Columbia	High school diploma or its equivalent			
Manitoba	College, CEGEP, or other technical or non-university certificate			
New Brunswick	University below the bachelor's level			
Newfoundland and Labrador	Bachelor's degree			
Nova Scotia	Degree above the bachelor's level			
Ontario	Don't know/ Prefer not to answer			
Prince Edward Island				
Quebec				

¹ All the variables except for vote choice were in both data sets. Only the CES data contained the vote choice variable.

² "Yes" was used in logistic regression, while 1 was used in propensity score matching.

³ Models used either Liberal as 1 or Conservative as 1.

Model

All the following statistical analyses were performed with R programming software (R Core Team, 2020; RStudio Team, 2020). Two logistic regression models were constructed using the CES data to predict the percentages of the popular vote that the Liberal and the Conservative Party each would have received if all eligible voters had voted. The predictor variables are age group, gender, province, education level, knowledge of English and/or French, and religious affiliation, which are all categorical or binary variables, and the response variable is vote choice, which is a binary response variable. Age was originally numerical in both the CES and GSS data sets, but I chose to form age groups to aid demographic bin partitioning in the poststratification analysis that will be described below. In one model, votes for the Liberal Party were recorded as 1 and all other choices were recorded as 0, and in the other model, votes for the Conservative Party were recorded as 1 and all other choices were recorded as 0. The logistic regression equation for modeling the Liberal vote percentage is as follows:

$$y_{Liberal} = \log\left(\frac{p_{Lib}}{1 - p_{Lib}}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{24} x_{24}$$

$p_{Liberal}$ represents the predicted proportion of votes that the Liberal Party would have received, and $y_{Liberal}$ is the log odds of the proportion. β_0 is the regression estimate for the reference category, which is female individuals from Alberta who are between 18 and 29 years of age, have a bachelor's degree, have understood both English and French since childhood, and do not have any religious affiliation. The other β estimates are in comparison to this reference category because it is the fixed baseline intercept of the regression model. For every unit change in the reference category, each β represents the corresponding change in the respective non-reference category, when all other variables are held constant.

β_1 to β_4 are the regression estimates of the age groups 30-44, 45-59, 60-74, and 75+, respectively. β_5 is the regression estimate for the male gender. β_6 to β_{14} are the regression estimates for the provinces British Columbia, Manitoba, New Brunswick, Newfoundland and Labrador, Nova Scotia, Ontario, Prince Edward Island, Quebec, and Saskatchewan, respectively. β_{15} to β_{20} are the regression estimates for the different levels of education—College/CEGEP/other technical or non-university certificate, degree above the bachelor's level, don't know/prefer not to answer, high school diploma or its equivalent, less than high school, and university below the bachelor's level, respectively. β_{21} to β_{23} are the regression estimates for knowing English only, French only, and neither English nor French, respectively. β_{24} is the regression estimate for having a religious affiliation.

The logistic regression equation for modeling the Conservative vote percentage is as follows:

$$y_{Conservative} = \log\left(\frac{p_{Con}}{1 - p_{Con}}\right) = \beta_{25} + \beta_{26} x_{25} + \beta_{27} x_{26} + \dots + \beta_{49} x_{48}$$

$y_{Conservative}$ is the log odds of the proportion of votes that the Conservative Party is predicted to receive, and $p_{Conservative}$ represents that proportion. β_{25} is the regression estimate for the reference category, which is the same as that of the Liberal model. The remaining β values, β_{26} to β_{49} , represent the same variables and categories in the same order as the Liberal model as well.

Next, poststratification analysis was conducted using the GSS census data as the second part of estimating the popular vote proportions of the Liberal Party and the Conservative party. Since the CES data may not be demographically representative of the Canadian population, poststratification analysis provides a method of correcting for errors caused by statistical analysis with non-representative samples. Ideally, the census data is a large data sample, and one drawback of the GSS data is that it has fewer data points compared to the CES data. For the poststratification calculation, the census data is first partitioned into cells (or bins) of every possible demographic combination of the variables selected, and the number of observations in each cell is counted. The logistic regression models mentioned above are used to estimate a response for each cell, and the predicted responses are weighted by the number of individuals per cell before being summed into an overall predicted response and then divided by the population size. The variables used for dividing demographic bins are five age groups, two genders, ten provinces, seven education levels, four types of language knowledge, and two modes of religious affiliation. Out of the 5,600 possible demographic bins, only 2,061 contain at least one data point. The formula used for the poststratification calculations is as follows:

$$\hat{y}^{PS} = \frac{\sum N_j \hat{y}_j}{\sum N_j}$$

\hat{y}^{PS} is the overall estimated response. \hat{y}_j is the prediction in the j^{th} cell and N_j is the number of observations per cell. The same calculation process is performed to predict both the Liberal votes and the Conservative votes.

Propensity score matching (PSM) is conducted with the CES data to investigate whether there is a direct causal relationship between a treatment and an outcome, as opposed to both the treatment and the outcome being influenced by other confounding instrumental variables. In this analysis, PSM is performed to determine whether religious affiliation, has a causal relationship with Liberal or Conservative vote preference, with age group, gender, province, education level, and language knowledge as possible instrumental variables that may be correlated with both the treatment, religious affiliation, and the outcome, vote preference. The R code for PSM analysis was adapted from code written by Rohan Alexander (2020).

In PSM, a logistic regression model is constructed with the treatment, religious affiliation, as the response variable, and the hypothesized instrumental variables as the predictors. A forecast response is then added to the data. Matches of observations with treatment and without treatment are created, and the matched pairs are selected to be as similar to each other as possible, only differing in whether they have treatment. Then, another logistic regression model is constructed with the outcome of interest as the binary response variable and the other variables, including the treatment, as predictors. This logistic

regression model then indicates whether the other variables are correlated with both the treatment and the outcome. In this analysis, PSM was performed for both the Liberal and Conservative models. The reference category for both PSM models is the same as the previous logistic regression models.

Results

Table 2: Comparison of selected significant variable categories between the Liberal and Conservative models

Category	Likelihood of voting Liberal	Likelihood of voting Conservative
Ages 60-74	1.2881127	1.4413608
Ages 75 +	1.3168903	1.8378807
British Columbia	1.8744655	0.2941236
Newfoundland and Labrador	3.420157	0.1697833
Ontario	2.6478238	0.2842154
Saskatchewan	0.7824056	0.6701715
College/CEGEP/other...	0.6470099	1.230232
High school diploma...	0.3032904	1.0027987
Less than high school	0.4900208	1.1440554
Religion - yes	1.0672502	1.7839145

Complete tables with descriptive statistics for the Liberal and Conservative logistic regression estimates (Tables 3 and 4) can be found in the Appendix.

The logistic regression analysis of the correlation between age, gender, province, education level, language knowledge, religious affiliation, and vote preference for the Liberal Party shows that in comparison to the reference category, every variable has at least one category with a significant regression estimate, except for gender. In this analysis, significance is determined by $p < 0.05$. Since the β values are the log odds of the proportions, to interpret the regression estimates, the β estimates must be converted by $\exp(\beta)$ to find the change in the category based on a unit change in the reference category. For example, in Table 2, the regression estimate for ages 60-74 in the Liberal model is 0.2531781, which indicates that for every unit change in the reference category, ages 60-74 is likely to change by $\exp(\beta_3) = 1.2881127$.

For the regression model for Liberal vote preference, the following categories are significant: the intercept, ages 60-74, ages 75+, all the provinces, college/CEGEP/other technical or non-university certificate, don't know/prefer not to answer, high school diploma or its equivalent, less than high school, university below the bachelor's level, French only, and existence of religious affiliation. For the regression model for

Conservative vote preference, the following categories are significant: the intercept, all the age categories, male gender, all the provinces, college/CEGEP/other technical or non-university certificate, degree above the bachelor's level, high school diploma or its equivalent, less than high school, English only, neither English nor French, and existence of religious affiliation.

From the calculated proportions shown in Table 2, elder people seem to be more likely than the reference category to vote both Liberal and Conservative. It is interesting to note that most of the provinces have 2- to 3-fold increased likelihood of voting Liberal compared to the reference category, with Saskatchewan being the exception, though Saskatchewan is still more likely to vote Liberal than to vote Conservative. Most of the provinces also have a greatly decreased likelihood of voting Conservative compared to the reference category. As seen in Table 4 (Appendix), male voters are more likely to vote Conservative, while their likelihood of voting Liberal is not significant compared to the reference category (Table 3). Another interesting result is that people with college/CEGEP/other technical or non-university certificate, high school diploma, or less than high school education are more likely to vote Conservative compared to the reference category and are less likely to vote Liberal compared to the reference category.

$$\hat{p}_{Liberal} = 0.2445936$$

$$\hat{p}_{Conservative} = 0.2989589$$

The poststratification analysis of the logistic regression models predicted that if all eligible individuals had voted in the 2019 election, the Liberal Party would have received 24.46% of the popular vote, and the Conservative Party would have received 29.90% of the popular vote. The two logistic regression models accounted for age groups, gender, province, knowledge of English and/or French, and religious affiliation.

In the propensity score matching (PSM) analysis, 27,728 observations were matched out of 36,120 observations. As seen in Tables 5 and 6 (Appendix), religious affiliation was found to be non-significant for the Liberal Party model, since its p-value is greater than 0.05. However, in the Conservative model, religious affiliation was found to have a regression estimate of 0.584767 at significance level $p < 0.001$, which indicates that it has a causal relationship with vote preference for the Conservative Party, but not for the Liberal Party.

Discussion

Summary

In this analysis, CES data and GSS data were used to construct logistic regression, poststratification, and propensity score matching models to examine the relationship between age, gender, province of residence, education level, knowledge of English and/or French, and religious affiliation and whether an individual would vote for the Liberal Party or the Conservative Party. Two logistic regression models demonstrated that most of the variables have a significant impact on voter preference for either party. Poststratification analysis with census data and demographic bin partitioning revealed that if voter turnout

were 100% in the 2019 federal election, the Liberal Party would have received 24.46% of the popular vote, which is less than the 29.90% of the popular vote that the Conservative Party would have received. Propensity score matching determined that while religious affiliation has a causal relationship in indicating voting preference for the Conservative Party, it has no significant causal relationship with voting preference for the Liberal Party.

Conclusions

The Conservative Party is predicted to receive 5.44% more votes than the Liberal Party, in contrast to receiving 1.3% more votes than the Liberal Party in the real election (CBC, n.d.). This analysis does not estimate the votes in each riding or the riding distribution among the provinces because the original data does not include the postal codes of the respondents. Additionally, weighting the popular vote percentages to determine the total number of seats won could be inaccurate, because in the election, the Liberal party won 46.4% of the seats with just 33.1% of the votes, and the Conservative party won 35.8% of the seats with 34.4% of the votes (CBC, n.d.). This phenomenon could be a result of wider spread of Liberal voters across more ridings, compared to more concentrated Conservative voters in certain areas.

From this analysis, it seems that many of the non-voters are likely to support the Conservative Party, since the Conservative votes are predicted to increase when they are included in the prediction. This could be due to a misrepresentation of the demographic by the GSS data, since in addition to only 2,061 out of 5,600 possible demographic bins containing at least one individual, 1,558 out of the 2,061 cells have 10 or fewer data points. An accurate poststratification analysis relies on a large amount of data both in the entire census and in each demographic cell, so that the predicted response of each cell is more likely to be accurate. Since the predicted response for many of the bins in this analysis may be skewed because of the small sample size, the poststratification analysis may not yield as accurate of a result as it should. An improvement to the poststratification analysis would be to select other census data with a larger sample size, or statistically correct for the small sample size.

The regression estimate of having a religious affiliation in the Conservative PSM model was 0.584767, as seen in Table 6 (Appendix), and this indicates that compared to the reference category, people with a religious affiliation are 1.7945728 times more likely to vote for the Conservative Party, with $p < 0.001$. The significant p-value indicates that there is a causal relationship between the treatment and the outcome of interest, and that the two are not simply correlated. Since the p-value for religious affiliation in the Liberal PSM model was not significant, religious affiliation and vote preference for the Liberal Party may simply be correlated. This result is valuable for studying vote choice, as we get insight into the factors that influence these choices and their relation to each other.

Weaknesses & Next Steps

There are some weaknesses of both the data and the analytical methods used in this analysis. In the CES data, 291 observations had gender recorded as Other, and since the GSS data did not record any observations that were not female or male, they were removed

from analysis. This is a drawback because nonbinary individuals are not represented in this analysis. Also, the census only records sex instead of gender, and this study assumed that the gender of the individuals matched their sex. Assuming gender from sex has a risk of misgendering individuals, which is unethical, therefore an improvement would be to estimate the appropriate gender proportions in the population to minimize the mean square error (Kennedy et al., 2020).

In the data cleaning process, some aspects of the data were excluded in favour of matching the CES and GSS data to aid statistical analysis. For example, the specific religious affiliations were recorded in the CES data, which would have been interesting to analyze to investigate the differences between people with different religions, but the GSS data only recorded the presence or absence of religious affiliation. The GSS also did not record observations from the three territories, and a few observations in the CES data that were from the three territories were excluded. Another weakness of the GSS data, which has been mentioned above, is the small sample size relative to the CES data. An improvement would be to select a different census data set for poststratification analysis.

A limitation for both the CES and GSS data sets is the non-response problem, especially in the GSS data set, where the respondents were approximately half of those contacted for a survey (Statistics Canada, 2020a). The non-response limitation in the CES is the option of answering “don’t know/prefer not to answer,” which is not as detrimental as missing values, but still reduces the accuracy of analysis.

A future direction of research includes conducting poststratification analysis with a more comprehensive census data set and predicting the outcome in each riding to better estimate the outcome of the election. However, this is difficult to achieve because voters’ postal codes are personal information that is kept confidential. Other future directions for research would include elucidation of the causal factors that predict vote preference for the Conservative Party in addition to religious affiliation, and further investigation to identify causal factors that predict vote preference for the Liberal Party.

References:

Alexander, R & Caetano, S. (2020, October 7). *gss_cleaning.R*. [STA304 Lecture material].

Alexander, R. (2020, November 5). *Difference in differences*. Telling Stories With Data. Retrieved December 20, 2020, from https://www.tellingstorieswithdata.com/06-03-matching_and_differences.html

Canada Broadcasting Corporation (CBC). (n.d.). *Federal election 2019 live results*. <https://newsinteractives.cbc.ca/elections/federal/2019/results/>

Computing in the Humanities and Social Sciences (CHASS). (n.d.). *CHASS Data Centre*. Retrieved October 19, 2020, from <http://www.chass.utoronto.ca/>

Elections Canada. (2020a, December 2). *Voter Turnout at Federal Elections and Referendums*. <https://elections.ca/content.aspx?section=ele&dir=turn&document=index&lang=e>

Elections Canada. (2020b, May 12). *The Electoral System of Canada*. <https://www.elections.ca/content.aspx?section=res&dir=ces&document=part1&lang=e>

Elections Canada. (2020c, October 26). *Map of Canada - the 43rd Parliament (PDF format - 34 MB)* https://www.elections.ca/res/cir/maps2/images/ERMap_43.pdf

Gelman, A. & Su, Y-S. (2020). *arm: Data Analysis Using Regression and Multilevel/Hierarchical Models*. R package version 1.11-2. <https://CRAN.R-project.org/package=arm>

Hodgetts, P. A. & Alexander, R. (2020). *cesR: Access the CES Datasets a Little Easier*. R package version 0.1.0. <https://hodgettsp.github.io/cesR/>

Kennedy, L., Khanna, K., Simpson, D. & Gelman, A. (2020). *Using sex and gender in survey adjustment*. [stat.AP]. <https://arxiv.org/abs/2009.14401v1>

Merriam-Webster. (n.d.). *Democracy*. In *Merriam-Webster.com dictionary*. Retrieved December 9, 2020, from <https://www.merriam-webster.com/dictionary/democracy>

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Robinson, D., Hayes, A., & Couch, S. (2020). *broom: Convert Statistical Objects into Tidy Tibbles*. R package version 0.7.1. <https://CRAN.R-project.org/package=broom>

RStudio Team (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC, Boston, MA.

Statistics Canada. (2020a, November 13). *General Social Survey - Family (GSS)*. Retrieved December 22, <https://www.statcan.gc.ca/eng/survey/household/4501>

Statistics Canada. (2020b, September 29). *Population estimates on July 1st, by age and sex*. Retrieved December 22, 2020, from <https://doi.org/10.25318/1710000501-eng>

Stephenson, L. B., Harell, A., Rubenson, D. & Loewen, P. J. (2020a). 2019 Canadian Election Study - Online Survey. *Harvard Dataverse*, V1. <https://doi.org/10.7910/DVN/DUS88V>

Stephenson, L. B., Harell, A., Rubenson, D. & Loewen, P. J. (2020b). Measuring Preferences and Behaviour in the 2019 Canadian Election Study. *Canadian Journal of Political Science*. <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DUS88V>

Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>

Wickham, H., François, R., Henry, L. & Müller, K. (2020). *dplyr: A Grammar of Data Manipulation*. R package version 1.0.0. <https://CRAN.R-project.org/package=dplyr>

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>

Wu, C., & Thompson, M. E. (2020). Basic Concepts in Survey Sampling. In: Sampling Theory and Practice. *ICSA Book Series in Statistics* (pp. 3-15). Springer, Cham. http://doi-org-443.webvpn.fjmu.edu.cn/10.1007/978-3-030-44246-0_1

Xie, Y. (2020). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.30.

Zimonjic, P. (2019, October 22). Liberals take losses but win enough in Quebec and Ontario to form minority government. *Canadian Broadcasting Corporation (CBC)*. <https://www.cbc.ca/news/politics/federal-election-results-2019-cbc-leaders-1.5329485>

Appendix

Table 3: Prediction of votes in favour of the Liberal Party - regression estimates and descriptive statistics

term	estimate	std.error	statistic	p.value
(Intercept)	-1.729	0.067	-25.679	0.000
age30-44	0.050	0.042	1.187	0.235
age45-59	0.067	0.042	1.578	0.115
age60-74	0.253	0.042	5.973	0.000
age75+	0.275	0.067	4.125	0.000
genderMale	-0.003	0.026	-0.128	0.898
provinceBritish Columbia	0.628	0.059	10.607	0.000
provinceManitoba	0.599	0.076	7.845	0.000
provinceNew Brunswick	0.947	0.092	10.303	0.000
provinceNewfoundland and Labrador	1.230	0.097	12.637	0.000
provinceNova Scotia	1.168	0.082	14.162	0.000
provinceOntario	0.974	0.049	19.677	0.000
provincePrince Edward Island	0.899	0.187	4.809	0.000
provinceQuebec	1.104	0.063	17.590	0.000
provinceSaskatchewan	-0.245	0.102	-2.402	0.016
educationCollege, CEGEP, or other technical or non-university certificate	-0.435	0.033	-13.196	0.000
educationDegree above the bachelor's level	0.044	0.041	1.093	0.275
educationDon't know/ Prefer not to answer	-1.193	0.338	-3.526	0.000
educationHigh school diploma or its equivalent	-0.566	0.042	-13.487	0.000
educationLess than high school	-0.713	0.065	-10.891	0.000
educationUniversity below the bachelor's level	-0.176	0.046	-3.836	0.000
languageEnglish only	-0.031	0.038	-0.805	0.421
languageFrench only	-0.550	0.050	-10.898	0.000
languageNeither English nor French	-0.041	0.061	-0.677	0.499
religionYes	0.065	0.027	2.451	0.014

Table 4: Prediction of votes in favour of the Conservative Party - regression estimates and descriptive statistics

term	estimate	std.error	statistic	p.value
(Intercept)	-1.024	0.068	-15.161	0.000
age30-44	0.327	0.048	6.798	0.000

age45-59	0.390	0.048	8.147	0.000
age60-74	0.366	0.048	7.595	0.000
age75+	0.609	0.071	8.562	0.000
genderMale	0.512	0.027	18.692	0.000
provinceBritish Columbia	-1.224	0.052	-23.633	0.000
provinceManitoba	-0.845	0.067	-12.643	0.000
provinceNew Brunswick	-1.422	0.097	-14.603	0.000
provinceNewfoundland and Labrador	-1.773	0.116	-15.222	0.000
provinceNova Scotia	-1.854	0.097	-19.059	0.000
provinceOntario	-1.258	0.040	-31.276	0.000
provincePrince Edward Island	-1.938	0.244	-7.946	0.000
provinceQuebec	-1.800	0.064	-28.111	0.000
provinceSaskatchewan	-0.400	0.070	-5.708	0.000
educationCollege, CEGEP, or other technical or non-university certificate	0.207	0.036	5.789	0.000
educationDegree above the bachelor's level	-0.175	0.049	-3.579	0.000
educationDon't know/ Prefer not to answer	0.003	0.309	0.009	0.993
educationHigh school diploma or its equivalent	0.282	0.044	6.487	0.000
educationLess than high school	0.135	0.066	2.024	0.043
educationUniversity below the bachelor's level	-0.066	0.053	-1.253	0.210
languageEnglish only	0.271	0.044	6.238	0.000
languageFrench only	-0.104	0.064	-1.620	0.105
languageNeither English nor French	0.358	0.067	5.324	0.000
religionYes	0.579	0.030	19.584	0.000

Table 5: Propensity score matching of the effect of religious affiliation on vote preference for the Liberal Party - regression estimates and descriptive statistics

term	estimate	std.error	statistic	p.value
(Intercept)	-1.744	0.072	-24.057	0.000
age30-44	0.040	0.042	0.963	0.336
age45-59	0.041	0.044	0.935	0.350
age60-74	0.202	0.053	3.805	0.000
age75+	0.263	0.103	2.558	0.011
genderMale	0.031	0.030	1.043	0.297
provinceBritish Columbia	0.623	0.062	10.118	0.000
provinceManitoba	0.603	0.084	7.206	0.000
provinceNew Brunswick	0.850	0.115	7.367	0.000

provinceNewfoundland and Labrador	0.897	0.143	6.282	0.000
provinceNova Scotia	1.069	0.096	11.097	0.000
provinceOntario	0.965	0.053	18.100	0.000
provincePrince Edward Island	0.876	0.210	4.172	0.000
provinceQuebec	1.068	0.070	15.237	0.000
provinceSaskatchewan	-0.184	0.119	-1.539	0.124
educationCollege, CEGEP, or other technical or non-university certificate	-0.418	0.037	-11.160	0.000
educationDegree above the bachelor's level	0.047	0.045	1.041	0.298
educationDon't know/ Prefer not to answer	-1.134	0.356	-3.181	0.001
educationHigh school diploma or its equivalent	-0.559	0.052	-10.837	0.000
educationLess than high school	-0.758	0.075	-10.039	0.000
educationUniversity below the bachelor's level	-0.202	0.052	-3.867	0.000
languageEnglish only	0.013	0.042	0.301	0.763
languageFrench only	-0.460	0.060	-7.676	0.000
languageNeither English nor French	-0.054	0.072	-0.748	0.454
religion	0.054	0.029	1.853	0.064

Table 6: Propensity score matching of the effect of religious affiliation on vote preference for the Conservative Party - regression estimates and descriptive statistics

term	estimate	std.error	statistic	p.value
(Intercept)	-0.976	0.072	-13.473	0.000
age30-44	0.315	0.048	6.540	0.000
age45-59	0.392	0.049	7.958	0.000
age60-74	0.355	0.059	6.029	0.000
age75+	0.713	0.107	6.644	0.000
genderMale	0.517	0.032	16.272	0.000
provinceBritish Columbia	-1.228	0.054	-22.735	0.000
provinceManitoba	-0.818	0.074	-11.018	0.000
provinceNew Brunswick	-1.359	0.126	-10.767	0.000
provinceNewfoundland and Labrador	-1.562	0.175	-8.923	0.000
provinceNova Scotia	-1.842	0.119	-15.505	0.000
provinceOntario	-1.291	0.044	-29.171	0.000
provincePrince Edward Island	-1.821	0.271	-6.731	0.000
provinceQuebec	-1.748	0.074	-23.612	0.000
provinceSaskatchewan	-0.389	0.084	-4.662	0.000
educationCollege, CEGEP, or other technical or	0.175	0.040	4.339	0.000

non-university certificate				
educationDegree above the bachelor's level	-0.169	0.054	-3.149	0.002
educationDon't know/ Prefer not to answer	-0.032	0.320	-0.099	0.921
educationHigh school diploma or its equivalent	0.234	0.053	4.430	0.000
educationLess than high school	0.081	0.076	1.067	0.286
educationUniversity below the bachelor's level	-0.097	0.060	-1.620	0.105
languageEnglish only	0.237	0.048	4.886	0.000
languageFrench only	-0.094	0.077	-1.221	0.222
languageNeither English nor French	0.370	0.080	4.639	0.000
religion	0.585	0.032	18.379	0.000