# Investigation of the 2019 Canadian Election

Yixin Liang(1005549998)
Code and data are available at: https://github.com/YixinLiang/STA304

12/22/2020

## Contents

# 1 Abstract

The election is a significant political event for a country, which should be emphasized since the governing political party influences its future development and policies in the following four years. In this study, logistic regression model and post-stratification analysis are applied to the census data and survey data from the Canadian Election Study and General Social Survey. This study analyzes six predictors that would impact the voting rate: sex, province, income, age, education, whether the family has children. Based on the model, Liberals have a higher probability of winning the 2019 Canadian Election from both popular voting rate and electoral voting rate aspects.

# 2 Key Words

Post-stratification, Logistic Regression, Canadian Election, Liberal Party, Conservative Party

# 3 Introduction

With the development of technology, this is an era of information explosion. The importance of data is not the numbers but the underlying information it carries. Statistical analysis is crucial for explaining variations and making the prediction. In a highly informative society, statistical analysis is currently widely used in elections. Election predictions are essential to candidates and the general public. Candidates can analyze voters in different regions to plan more effective election arrangements. They can also design the election campaigns based on the analyzed voter preferences to obtain more votes to increase their winning rate. The statistical analysis is not limited to the prediction of results but also includes studying past polls, which can help achieve more future analysis improvements.

Logistic regression (LR) continues to be one of the most widely used methods in data mining in general and binary data classification in particular (Maalouf, 2011). Post-stratification is widely recognized as an effective method for obtaining more accurate estimates of population quantities in the context of survey sampling (Little, 1993). In this study, logistic regression and post-stratification analysis are used to estimate the 2019 Canadian Federal Election, assuming all residents voted. Individual-level data contains personal opinions on the election, and the census data is used to make predictions.

In the 2019 Canadian election, the Liberals and Conservatives are still the leading competitors. Canada's current applied election system is called "first past the post". Basically, in each electoral district, the candidate with the highest number of votes wins a seat in the House of Commons and represents that district as a Member of Parliament (Canadian Election, 2020). Notably, the absolute majority is not required under the "first past the post" process. The leader of the party that won most seats becomes Prime Minister.

Overall, this study uses two datasets to investigate the election result for the 2019 Canadian election. In the methodology section, the data source and the model used for conducting the analysis are introduced. The result section will demonstrate tables, graphs and associated outcomes. In the discussion section, the results, criticism, and references for this study will be presented.

# 4 Methodology

## 4.1 Data

### 4.1.1 Survey Data:

The survey data is obtained from the Canadian Election Study (CES), a large-scale online survey conducted during the Canadian federal election period. In this study, 'ces2019_web' is selected from the 'cesR' package that contains 37822 observations, which focuses on the Canadian Election in 2019. The non-sampling procedure used is the volunteer sampling. This dataset provides thorough and comprehensive information that contains respondents' attitudes to different political parties and includes the demographic variables. Selected demographic variables are used to separate the population into groups to apply the post-stratification method.

The **target population** includes all residents in Canada. The **frame population** is those who visit the website that posts the online survey. The **sample population** is people who answer the survey.

### 4.1.2 Census Data:

The census data is collected from the General Social Survey (GSS) on the family in 2017, conducted through telephone interviews. This dataset contains 20602 observations after cleaning[1], focusing on non-institutionalized persons who are at least 15-year-old and are from 10 provinces in Canada. The sampling procedure used is a stratified sampling that performs a simple random sampling without replacement within each stratum. Two types of variables are included in the dataset, core content and classification variables. Core content variables focus on social issues and classification variables can be used for separating population groups. Additionally, it faces the non-response problem due to its survey method.

The **target population** includes all non-institutionalized persons who are 15 years of age and older and live in 10 provinces of Canada. The **frame population** is the list of telephone numbers created in 2013 by Statistics Canada and the Address Register. The **sample population** is people who take the survey through phone calls.

### 4.1.3 Data Cleaning:

The variables in census data and survey data should match each other to conduct statistical analysis ad prediction. Therefore, data are reorganized according to six explanatory variables in both survey data and census data: education, income, age, gender, province, number of children in the family. In the survey data, one more variable is selected: people's vote choices. All the missing values are removed before data cleaning and modeling.

### 4.1.4 Dependent Variables:

Respondents' vote choice is used as the dependent variable in this study to apply the logistic model. Based on the original variable, two new variables are created, representing individuals' vote choice for Liberals and Conservatives, respectively.

### 4.1.5 Independent Variables:

- Education:

In the census data, individuals' education levels are distributed into seven groups. In the survey data, it is divided into eleven groups, expressing by the numbers. In this study, education levels are reorganized into four groups, which are "Primary Education", "Secondary Education", "Post-secondary Education" and "Post-graduation Education".

[1] Code for data cleaning is from Professor Samantha-Jo Caetano

- Household Income:

In the census data, the household income is separated into six income ranges. However, in the survey data, it is expressed as numbers. Notably, the maximum household income is 6.75e+60, which may be an outlier. Therefore, this study only focuses on the observations with less than 10000000 income to reduce the response error. Also, the household income is rearranged into three groups in both datasets. The "Higher income", "Medium Income" and "Lower Income" represent the household income that ranges from 0 to 49999, 50000 to 99999 and higher than 1000000 respectively.

- Age:

In both datasets, age is shown as numbers. Concentrated on this study's objective, observations that are younger than 18 years old are removed because they do not have voting rights (Elections New Brunswick, 2014). Additionally, observations are distributed into three age groups: "Youth", "Adult" and "Senior". This division uses 25 and 65 as the thresholds, based on the age grouping standard posted by Statistical Canada (Government of Canada, 2017).

- Gender:

In this study, gender is classified into two groups, female and male. To match the variables, responses with "others" in the survey data are removed.

- Province:

The census data only focuses on ten central provinces in Canada. However, 13 different regions are targeted in the survey data. Respondents currently living in Northwest Territories, Nunavut, and Yukon are removed to match the census data. Removing observations may not have a significant influence because these provinces only account for about 0.1% of the total results, respectively.

- Children:

This variable represents whether the respondents have children. In the census data, the corresponding variable, which measures how many children the respondents have, is reorganized into the binary variable during the data cleaning process.

#### 4.1.6 Cleaned Raw Dataset:

The cleaned raw survey data includes 21496 observations and eight variables. The cleaned raw census data has 19824 observations and six variables. The variables are categorical or binary. There is no missing value after cleaning the data, which is eligible for the following analysis and statistical modeling. The visualization of the cleaned data sets is shown by **Figure 1 (Cleaned Raw Survey data)** and **Figure 2 (Cleaned Raw Census data)** in Appendix 7.1 and Appendix 7.2.

#### 4.1.7 Strength and Weakness for data:

- Strength: Data set used in the study provides more comprehensive demographic variables, which helps conduct post-stratification and make a better prediction. Also, most questions in the questionnaires give the respondents a range of answers, which decreases misunderstanding problems.
- Weakness: Data collected in data sets are through the online survey and telephone survey, respectively. Therefore, the data used in the study may have non-response problems. Also, the questionnaires contain too many questions, which may lead to some incorrect and unreliable responses.

## 4.2 Model

### 4.2.1 Model Introduction:

This study investigates the 2019 Canadian Election results based on the Canadian Election Study and General Social Survey. The first step is modeling two binary response variables: people's vote for Liberals and Conservatives. The model is constructed using logistic regression in R. Logistic regression model is helpful to calculate the probabilities of binary variables. Secondly, the post-stratification method is involved. The model, generated based on the survey data, is used to calculate the possibility for each post-stratification cell that people will vote for Liberals or Conservatives.

### 4.2.2 Final Model:

The following mathematical notation and equation is the logistic regression model generating by R with six predictors, *sex*, *province*, *income*, *age*, *education* and *children*.

$$\log \frac{p_i}{1-p_i} = \beta_0 + \beta_{sex} X_{sex} + \beta_{province} X_{province} + \beta_{income} X_{income} + \beta_{age} X_{age} + \beta_{education} X_{education} + \beta_{children} X_{children}$$

$p_i$: the probability that an individual votes for Liberals or Conservatives in the 2019 Canadian Election

$\beta_0$: the intercept of the model; represents the log odds of an adult female who lives in Alberta, earns higher income, has post-graduation education and does not have any children.

$\beta_{sex}$: the slope for sex; represent the difference of preference between males and females.

$X_{sex}$: = 1 if the respondent is male.

$\beta_{province}$: the slope for province; represent the difference of preference between people lived in Alberta and other 9 provinces in Canada

$X_{province}$: = 1 if the respondent lives in the provinces except for Alberta.

$\beta_{income}$: the slope for income level; represent the difference of preference between people who earn higher income and other 2 income range.

$X_{income}$: = 1 if the respondent falls in a income range that is not higher than 100000.

$\beta_{age}$: the slope for age groups; represent the difference of preference between people who is an adult and others who are senior and youth.

$X_{age}$: = 1 if the respondent falls in an age range other than 25 - 64.

$\beta_{education}$: the slope for educational level; represent the difference of preference between people who have post-graduation education and other 3 educational levels.

$X_{education}$: = 1 if the respondents' educational level is lower than the post-graduation.

$\beta_{children}$: the slope for fertility status; represent the difference of preference between people who do not have children and those who have children.

$X_{children}$: = 1 if the respondent have at least one children.

### 4.2.3 Model Discussion:

Given the logistic model above, with the help of post-stratification analysis, it can effectively predict the possibility that Liberals or Conservatives wins the election. However, before calculating the probability, there is some discussion associated with the logistic model.

In the fitted model for Liberals, six independent variables are included. Notably, *sex* is not statistically significant because its P-value is greater than the significant level, 0.05. After applying BIC[2] to assess the variable selection, *sex* is removed from the final model. However, when constructing the Conservatives' logistic model, *sex* is statistically significant and is selected by the BIC criterion. Therefore, *sex* is kept in the final model for Liberals because it makes the study more consistent and helps further comparisons between Liberals and Conservatives. Additionally, some external references confirmed that sex identity actually influences individuals' vote preference (Hatemi et al., 2020). Therefore, the final logistic model still includes six variables.

### 4.2.4 Post-stratification:

Post-stratification is a method that divides the entire census data into demographic cells based on the selected variables. Ideally, the number of cells equals all the potential combinations of the chosen variables unless no observations exist for the particular cell.

Sex, province, income, age, education, and whether the family has children are used to form the post-stratification cells for further analysis. The census is divided into 1084 demographic cells. The dependent variable, vote rate, is estimated within each cell. The population-level estimate is calculated by weighting each cell according to the relative proportion of the population. Post-stratification is useful since it can efficiently reduce the bias resulting from non-probability sampling.

In Canada, different provinces may have their unique culture, history and political preference. As a country with a high immigration rate, different regions may attract various immigrants and residents, leading to a specific tendentious election choice. Also, there is a significant 'genderation' gap among both gender and age groups based on the data posted by Ipsos (Breen, 2019). Also, individuals' income, educational level and whether they have children may lead to different reactions and preferences to the political parties' potential policies. Therefore, these six variables are selected to divide the census data for making further predictions.

### 4.2.5 Model Diagnostics:

- Influential Observations

The influential points will significantly affect the fitted model; therefore, diagnostics is necessary. Cook's distance is used in the study to check the influential points. According to the criteria, points with a cook's distance that is greater than 0.5 should draw attention (Cook, 1977). Referring to the **Figure 3 (Cook's Distance Plot for Two Models)** in Appendix 7.3, there are no points with a value greater than 0.5 in the cook's distance plots for both Liberals and Conservatives.

- Multicollinearity

The multicollinearity is checked by using the Variance Inflation Factor. The multicollinearity problem reduces the statistical analysis's efficiency, which may even cause the coefficients to switch signs. Referring to **Table 3 (Multicollinearity Check for Liberals)** and **Table 4 (Multicollinearity Check for Conservatives)** in the Appendix 7.4 and 7.5, based on the calculated VIF for each predictor, they are around 1 and do not meet the VIF cut off. Therefore, there is no multicollinearity problem associated with the final model.

[2]Bayesian Information Criterion, a criteria to select model. Lower BIC indicates a better model.

#### 4.2.6 Electoral Vote:

Based on the Canadian Election process, the popular vote rate is a reference. The electoral vote rate, represented by the number of seats obtained in the House of Commons, is the final determination for the governing party. A party may win the popular rate but loses the electoral rate and thus fails in the election. Therefore, this study will calculate both popular rate and electoral rate. The electoral rate for each party is calculated by dividing the number of seats that are expected to win over total seats in Canada. The expected winning seats is calculated by multiplying the provincial vote rate by the seats for those provinces. This can help to predict the total seats that each party can obtain. Notably, three provinces, Northwest Territories, Nunavut, and Yukon, are not included in the study, which each has one seat.

# 5 Results:

## 5.1 Popular Vote Prediction

Referring to **Table 1 (Popular Vote Rate)**, Liberals expects to have a 30.18% popular support rate and Conservatives expects to have a 28.57% popular support rate.

Table 1: Predicted Popular Vote Rate (Liberals v.s. Conservatives)

| Liberals | Conservatives |
|---:|---:|
| 0.3018 | 0.2857 |

## 5.2 Electoral Vote Prediction

Referring to **Table 2 (Electoral Vote Rate)**, Liberals expects to have a 29.27% electoral vorte rate and Conservatives expects to have a 28.77% electoral vote rate.

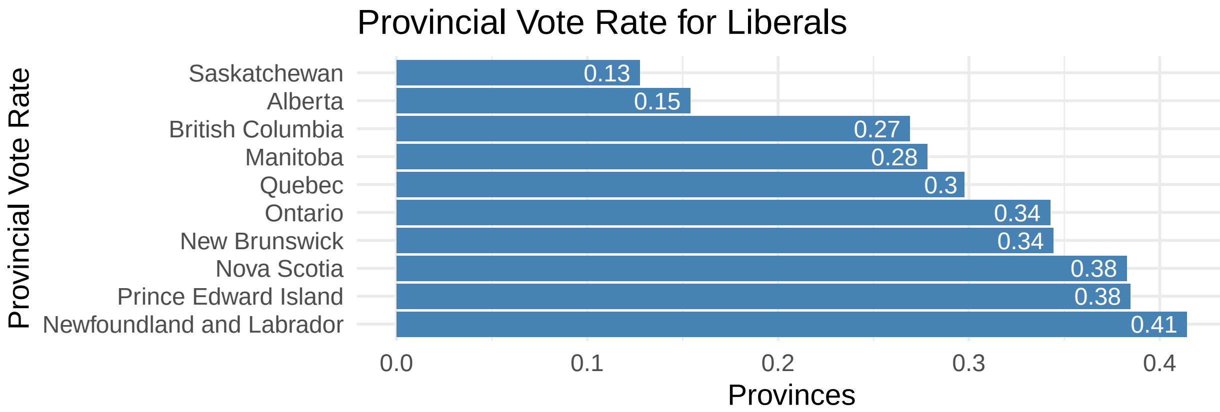
Table 2: Predicted Electoral Vote Rate (Liberals v.s. Conservatives)

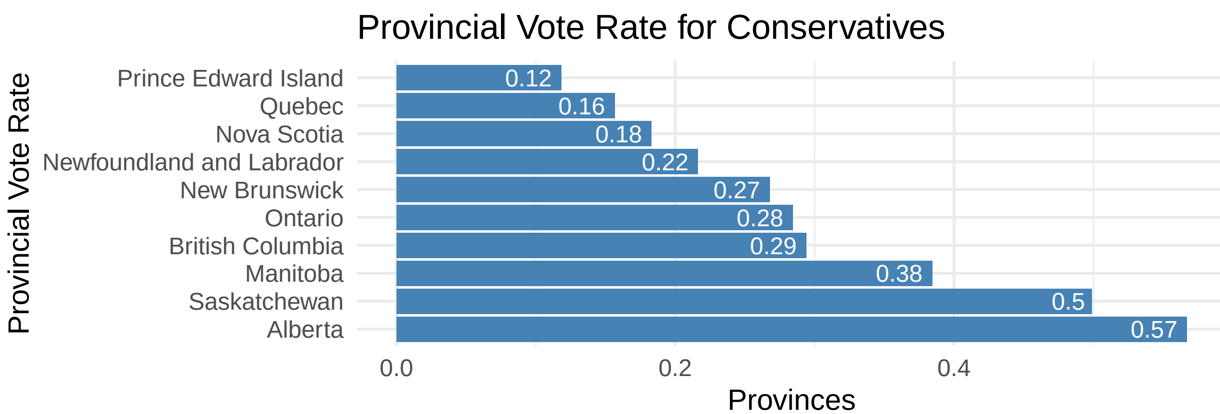
| Liberals | Conservatives |
|---:|---:|
| 0.2927 | 0.2877 |

## 5.3 Logistic Regression Models

According to the logistic regression, two fitted models are generated to predict Liberals and Conservatives' support rates, respectively. Referring to **Table 5 (Model for Liberals)** and **Table 6 (Model for Conservatives)** in the Appendix 7.6 and 7.7, most variables are statistically significant for the Liberals' model. However, some variables have relatively large P-value, such as gender, Saskatchewan province. According to the fitted model for Conservatives, almost all variables are statistically significant except for people who have primary education level. To predict the support rate for both political parties consistently, gender is kept in the final model, which has been sufficiently discussed before.

### 5.4 Provincial Vote Rate




The provincial vote rates for both parties are calculated by applying post-stratification analysis within each province. According to two bar graphs, Liberals face the lowest vote rate in Saskatchewan and Alberta provinces. Conservatives face the lowest supporters in Prince Edward Island and Quebec. Analyzing the estimated supporting rate helps the parties arrange their election campaign more efficiently to gain higher odds of winning.

# 6 Discussion

## 6.1 Summary

This study using two data sets, which contain the survey data and the census data. The survey data, Canadian Election Study, is obtained through the online survey and collects respondents' opinions on the 2019 Canadian Election. The census data, General Social Survey, uses the telephone interview to collect the responses, including solely demographic information.

To conduct the statistical analysis, six variables for survey data and eight variables for census data are selected and cleaned. This makes sure that the corresponding variables are consistent in both datasets. The logistic regression model, generated by using survey data, is used to assess each variable's impact on the dependent variable.

Based on the selected demographic characteristics, census data are divided into different cells to conduct further post-stratification analysis. The voting rate for both political parties is estimated within each cell. The population voting rate is calculated by considering the weight of each cell to the overall population.

Based on the results, Liberals are expected to gain a 30.18% popular support rate in the 2019 Canadian election if eligible residents all vote. Conservatives are expected to gain a 28.57% popular support rate, which is lower than the Liberals. The Liberals and the Conservatives are expected to gain a 29.27% electoral rate and a 28.77% electoral rate, respectively.

## 6.2 Conclusion

Referring to **Table 1 (Popular Vote Rate)** and **Table 2 (Electoral Vote Rate)**, Liberals are expected to have a 30.18% popular vote rate in the 2019 Canadian election, compared with the Conservatives' 28.57% popular vote rate. Therefore, based on the previous analysis and the database, Liberals are predicted to have a higher probability of winning the 2019 Canadian Election.

This electoral vote rate is consistent with the actual election that Liberals form the government. However, the popular vote rate is different from the actual result. In the 2019 Canadian Election, the Liberal has a 33.1% voting share, but the Conservative has a 34.4% voting share, which is slightly higher than the Liberal (CBC News, 2019).

Some potential issues may lead to the popular vote rate's inconsistency between the actual results and the study's prediction. Firstly, it may be due to low voter participation. Only 77% of Canadians reported voting in the 2019 federal election (Government of Canada, 2020). This may cause a difference in popular vote rate since the prediction in this study is based on the assumption that all eligible Canadians will vote. Secondly, it may be due to the criticism of the election process in Canada. There is no requirement for an absolute majority; therefore, the popular rate is not the election's primary determination. This may discourage voters' enthusiasm to vote, leading to a different result from the study's analysis.

These predictions can be used to compare with the actual results and provide some warnings and hints for both parties. For Conservatives, though the real results demonstrate that it wins the popular vote, it should be cautious and alert because the census prediction shows that Liberals have a slightly leading popular rate. Additionally, its failure of the election is due to fewer seats in the House of Commons, resulting from a bad election campaign. Conservatives' votes may be more clustered than the Liberals, whose vote spreads widely and evenly (Brean, 2019). On the other hand, Liberals can use the prediction for different provinces to help them make revisions and to improve the popular vote rate in the upcoming election.

### 6.3 Weaknesses

There is some weakness exist in the study. Firstly, some observations with missing values are removed before conducting the analysis. However, some useful information associated with the election prediction may be missed. For example, respondents who do not answer their gender as male or female are removed from the study, which may cause bias for the study results. Secondly, in this study, only 1084 demographic cells are created to conduct post-stratification analysis, which may not be sufficient to make an accurate estimation and thus constrains the prediction results. Thirdly, the overall study and prediction are made based on two datasets. The sampling method used for the survey and the respondents' incorrect information can lead to the bias of the prediction results. Finally, three provinces are not included in this study, which each has one seat in the House of Commons. The omission of these provinces may lead to the bias of the electoral rate prediction.

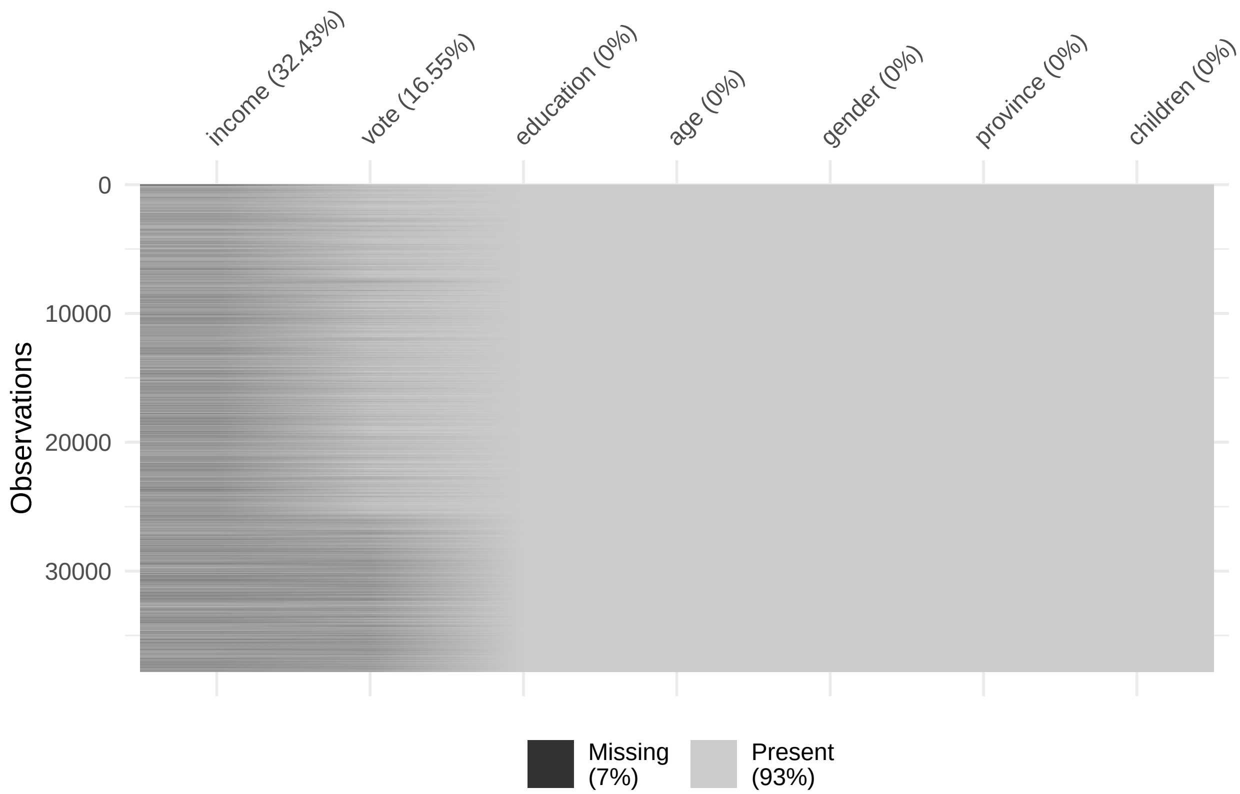
### 6.4 Next Steps

To improve the accuracy of predictions, the multilevel regression model can be considered as an alternative model. Using the BIC criteria, the model with the lowest BIC can be chosen to analyze further. In this study, three provinces are excluded from the electoral prediction. If the seats' difference is not large for Liberals and Conservatives, then these three provinces become essential. After getting enough census information about these provinces, a post-stratification analysis can be conducted. The prediction of the electoral rate will be more reliable.

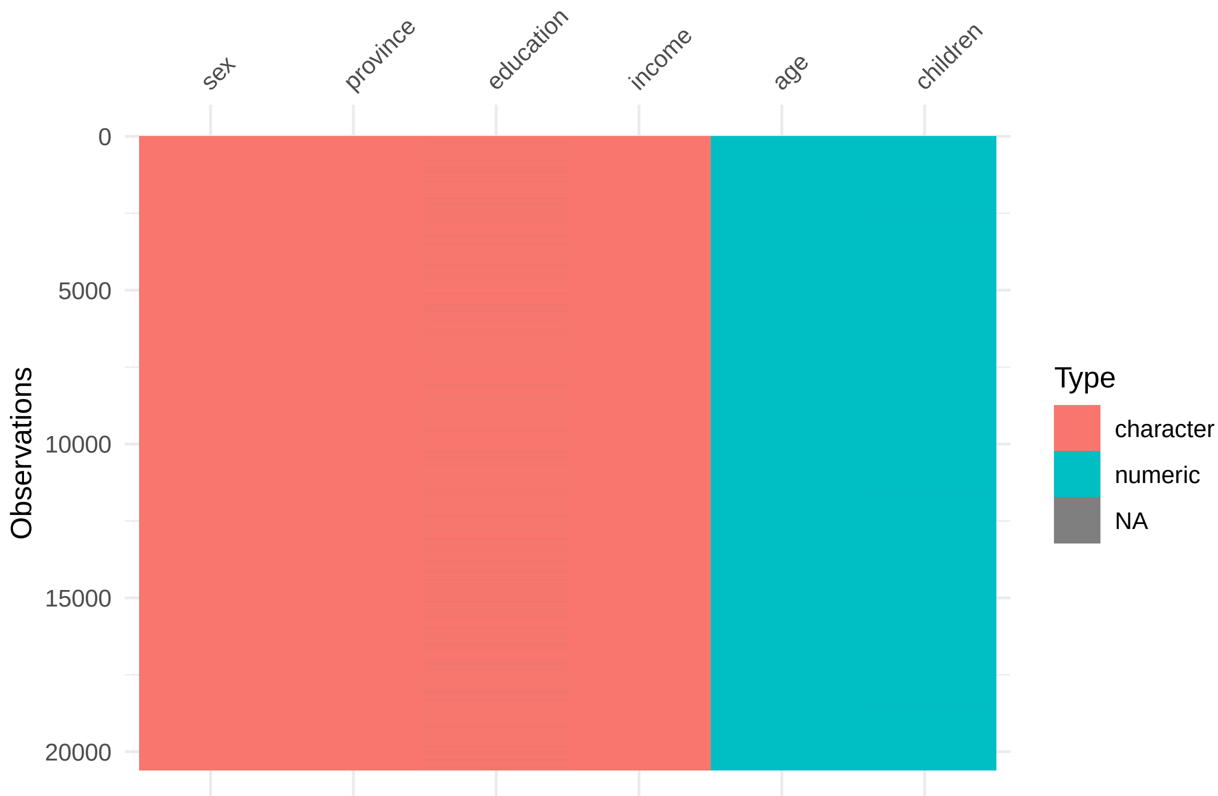
By comparing the actual results and the prediction results, the cause of the difference can be further investigated, which helps to improve the model accuracy and the future election prediction.

# 7 Appendix

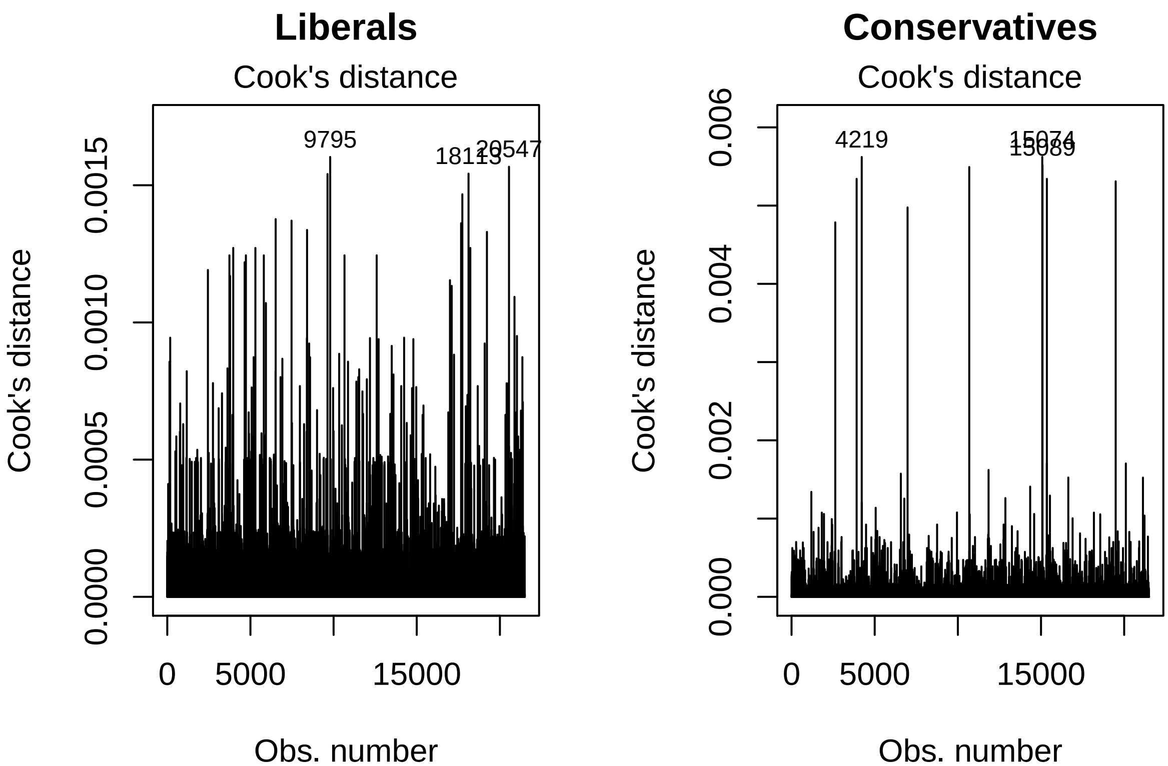
## 7.1 Figure 1: Cleaned Raw Survey data



### 7.2 Figure 2: Cleaned Raw Census data



### 7.3 Figure 3: Cook's Distance Plots for Two Models




### 7.4 Table 3: Multicollinearity Check for Liberals

Table 3: VIFs for 6 Predictors (Liberals)

| | GVIF | Df | GVIF^(1/(2*Df)) |
|---|---|---|---|
| sex | 1.049 | 1 | 1.024 |
| province | 1.025 | 9 | 1.001 |
| income_factor | 1.150 | 2 | 1.035 |
| age_factor | 1.154 | 2 | 1.037 |
| education | 1.095 | 3 | 1.015 |
| children | 1.110 | 1 | 1.054 |

## 7.5 Table 4: Multicollinearity Check for Conservatives

Table 4: VIFs for 6 Predictors (Conservatives)

| | GVIF | Df | GVIF^(1/(2*Df)) |
|---|---|---|---|
| sex | 1.055 | 1 | 1.027 |
| province | 1.033 | 9 | 1.002 |
| income_factor | 1.147 | 2 | 1.035 |
| age_factor | 1.149 | 2 | 1.035 |
| education | 1.108 | 3 | 1.017 |
| children | 1.100 | 1 | 1.049 |

## 7.6 Table 5: Model for Liberals

Table 5: Logit Regression for Liberals

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | -1.205 | 0.074 | -16.390 | 0.000 |
| sexMale | -0.027 | 0.031 | -0.874 | 0.382 |
| provinceBritish Columbia | 0.694 | 0.072 | 9.703 | 0.000 |
| provinceManitoba | 0.765 | 0.091 | 8.372 | 0.000 |
| provinceNew Brunswick | 1.080 | 0.110 | 9.778 | 0.000 |
| provinceNewfoundland and Labrador | 1.366 | 0.120 | 11.359 | 0.000 |
| provinceNova Scotia | 1.241 | 0.100 | 12.398 | 0.000 |
| provinceOntario | 1.036 | 0.060 | 17.373 | 0.000 |
| provincePrince Edward Island | 1.258 | 0.245 | 5.140 | 0.000 |
| provinceQuebec | 0.851 | 0.063 | 13.444 | 0.000 |
| provinceSaskatchewan | -0.193 | 0.121 | -1.599 | 0.110 |
| income_factorLower income | -0.211 | 0.042 | -5.000 | 0.000 |
| income_factorMedium income | -0.130 | 0.037 | -3.557 | 0.000 |
| age_factorSenior | 0.228 | 0.037 | 6.106 | 0.000 |
| age_factorYouth | -0.109 | 0.077 | -1.421 | 0.155 |
| educationPost-secondary Education | -0.258 | 0.044 | -5.826 | 0.000 |
| educationPrimary Education | -0.200 | 0.176 | -1.140 | 0.254 |
| educationSecondary Education | -0.558 | 0.058 | -9.669 | 0.000 |
| children | -0.202 | 0.033 | -6.132 | 0.000 |

## 7.7 Table 6: Model for Conservatives

Table 6: Logit Regression for Conservatives

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | -0.458 | 0.069 | -6.609 | 0.000 |
| sexMale | 0.516 | 0.033 | 15.830 | 0.000 |
| provinceBritish Columbia | -1.164 | 0.061 | -18.935 | 0.000 |
| provinceManitoba | -0.767 | 0.080 | -9.572 | 0.000 |
| provinceNew Brunswick | -1.292 | 0.113 | -11.448 | 0.000 |
| provinceNewfoundland and Labrador | -1.587 | 0.137 | -11.593 | 0.000 |
| provinceNova Scotia | -1.789 | 0.116 | -15.460 | 0.000 |
| provinceOntario | -1.204 | 0.048 | -25.091 | 0.000 |
| provincePrince Edward Island | -2.300 | 0.360 | -6.381 | 0.000 |
| provinceQuebec | -1.954 | 0.057 | -34.202 | 0.000 |
| provinceSaskatchewan | -0.312 | 0.085 | -3.683 | 0.000 |
| income_factorLower income | -0.395 | 0.045 | -8.837 | 0.000 |
| income_factorMedium income | -0.132 | 0.038 | -3.490 | 0.000 |
| age_factorSenior | 0.119 | 0.039 | 3.064 | 0.002 |
| age_factorYouth | -0.319 | 0.089 | -3.585 | 0.000 |
| educationPost-secondary Education | 0.352 | 0.051 | 6.916 | 0.000 |
| educationPrimary Education | 0.278 | 0.193 | 1.443 | 0.149 |
| educationSecondary Education | 0.648 | 0.062 | 10.517 | 0.000 |
| children | 0.355 | 0.036 | 9.983 | 0.000 |

# 8 References


1. Alboukadel Kassambara (2020). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. https://CRAN.R-project.org/package=ggpubr
2. Brean, J. (2019, October 22). All-time low share of popular vote is enough for Liberals to win power. Retrieved December 22,2020, from https://nationalpost.com/news/politics/election-2019/canadian-federal-election-2019-liberals-justin-trudeau-win
3. Breen, K. (2019, October 04). The 'genderation' gap: Political divisions exist between men, women, different age groups, polls show. Retrieved December 21, 2020, from https://globalnews.ca/news/5988160/genderation-gap-political-divisions-men-women/
4. Caetano, S., & Alexander, R. (2020, October 7). GSS Data Cleaning Code. Retrieved December 22, 2020, from https://q.utoronto.ca/courses/184060/files/9422740?module_item_id=1867317
5. Canadian Election. (2020). Elections Step by Step. Retrieved December 09, 2020, from https://electionsanddemocracy.ca/canadas-elections/canadas-election-process/elections-step-step
6. CBC News. (2019). Federal election 2019 live results. Retrieved December 22, 2020, from https://newsinteractives.cbc.ca/elections/federal/2019/results/
7. Cook, R. Dennis (February 1977). "Detection of Influential Observations in Linear Regression". Technometrics (American Statistical Association)).
8. Elections New Brunswick, C. (2014, December 11). Am I eligible to vote? Retrieved December 20, 2020, from https://www.electionsnb.ca/content/enb/en/voters/eligible.html
9. Elin Waring, Michael Quinn, Amelia McNamara, Eduardo Arino de la Rubia, Hao Zhu and Shannon Ellis (2020). skimr: Compact and Flexible Summaries of Data. R package version 2.1.2. https://CRAN.R-project.org/package=skimr
10. Government of Canada, S. (2017, May 08). Age Categories, Life Cycle Groupings. Retrieved December 20, 2020, from https://www.statcan.gc.ca/eng/concepts/definitions/age2
11. Government of Canada, S. (2020, February 26). Reasons for not voting in the federal election, October 21, 2019. Retrieved December 22, 2020, from https://www150.statcan.gc.ca/n1/daily-quotidien/200226/dq200226b-eng.htm
12. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
13. Hatemi, P., McDermott, R., Bailey, J., & Martin, N. (2012). The Different Effects of Gender and Sex on Vote Choice. Political Research Quarterly, 65(1), 76-92. Retrieved December 21, 2020, from http://www.jstor.org/stable/23209561
14. John Fox and Sanford Weisberg (2019). An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/
15. Little, R. J. A. (1993), "Post-Stratification: A Modeler's Perspective," Journal of the American Statistical Association, 88, 1001–1012.
16. Maalouf, M. (2011). Logistic regression in data analysis: An overview. International Journal of Data Analysis Techniques and Strategies, 3(3), 281. doi:10.1504/ijdats.2011.041335
17. Paul A. Hodgetts and Rohan Alexander (2020). cesR: Access the CES Datasets a Little Easier.. R package version 0.1.0.
18. Stephenson, Laura B; Harell, Allison; Rubenson, Daniel; Loewen, Peter John, 2020, '2019 Canadian Election Study - Online Survey', https://doi.org/10.7910/DVN/DUS88V, Harvard Dataverse, V1
19. Stephenson, Laura, Allison Harrel, Daniel Rubenson and Peter Loewen. Forthcoming. 'Measuring Preferences and Behaviour in the 2019 Canadian Election Study,' Canadian Journal of Political Science.

20. Technology, Advancing Knowledge through. Computing in the Humanities and Social Sciences. www.chass.utoronto.ca/.

21. Tierney N (2017). "visdat: Visualising Whole Data Frames." *JOSS*, *2*(16), 355. doi: 10.21105/joss.00355 (URL: https://doi.org/10.21105/joss.00355), <URL: http://dx.doi.org/10.21105/joss.00355>.

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