# If All Eligible Canadians Had Voted in 2019, Could Justin Trudeau Still Win?*

Application of Multilevel Regression with Post-stratification (MRP) Framework

Minchen Cai

December 20th, 2020

**Abstract**

The Prime Minister Justin Trudeau and his Liberal Party barely won the last Canadian Federal Election, particularly the Liberals won the most seats but lost the popular vote to the Conservatives. Since the Conservative Party of Canada was expected to win in many polls, it is very interesting to explore the voting result if all eligible Canadians had voted. In this study, the multilevel (logistic) regression with post-stratification (MRP) model was trained with the real 2019 Canadian election voting intention data, then the universal voting result was predicted with Canadian national census data which is the dataset considered to be a unbiased representative of Canadian population. This work demonstrated that, if all eligible Canadians had voted in 2019, Justin Trudeau and his Liberal party might win the federal election by a larger margin.

**Keywords:** Justin Trudeau; Multilevel Regression with Post-Stratification (MRP); 2019 Canadian Federal Election; Voter Turnout; Representative Dataset; Canadian Census Data; The Liberal Party of Canada; The Conservative Party of Canada;

## Introduction

More than one year has passed since the last Canadian Federal Election, which was held on the October 21, 2019. In that election, the Liberal Party, led by incumbent Prime Minister Justin Trudeau, won 157 seats to form a minority government, but lost the popular votes to the Conservative Party of Canada, which won 121 seats. In many opinion polls, the Conservatives were expected to beat the Liberals by a narrow margin. In addition, it was reported that the Liberals were suffered from the continuing loss of the support rate for a long time since the 2015 federal election. After the election, defeated parties claimed recounts in three ridings. Interestingly, the election results also ignited calls for electoral reform, since the Liberals only won 33% of the popular votes but won 47% of all seats (Institute 2019). What's more, it is widely reported that the voter turnout was around 66%, which represents that the total ballots cast was 66% of total number of electors on list. The voter turnout decreased by 2.5% compared to the 2015 federal election. In contrast, the average voter turnout of federal elections from 1950s to 1980s was around 75%. In the modern advanced democracy world, it is a serious problem that many citizen fail to participate in voting, and those who vote are systematically unrepresentative of the eligible population (Fowler 2013). It has been demonstrated in previous studies that election results might be significantly altered if the voter turnout were increased (Fowler 2013). Therefore, there is no surprising that people nowadays might wonder whether the Conservatives could win the election if all eligible Canadians had voted.

For a long time, the representative polls, where individuals are randomly sampled from a certain population and are asked by who they would like to vote for, have been the standard method to forecast vote attitudes of that entire population. However, the pool of respondents from a random sampling might still be highly biased.

*Open Access Code at: https://github.com/minchencai/Final-Project

For instance, most political opinion polls provided misleading information for the 2016 United States federal election, where Hillary Clinton was favored in most polls but lost to Donald Trump eventually. Additionally, the representative polls are typically costly in both time and money and the response rates keep decreasing over past several decades (Wei Wang and Gelman 2014). Nowadays, as the rising of the Internet, it is increasingly popular and cost-effective to conduct highly non-representative online survey (Wei Wang and Gelman 2014). Overall, it has become more and more important that a political forecast tool is able to digest many non-representative survey datasets, and correct the biases between the sample and the population to be addressed (Ghitza and Gelman 2013). The multilevel regression with post-stratification (MRP) was specifically developed to derive precise vote choice estimates through the analysis of survey data in greater geographic, demographic and socioeconomic details compared with previous statistical tools (Ghitza and Gelman 2013; Wei Wang and Gelman 2014). The basic idea of MRP is to partition the entire population into thousands of subgroups by a multilevel regression model, and finally aggregate the vote choice estimates over different subgroups in accordance with the characteristics of the target population (Wei Wang and Gelman 2014).

Thus, the MRP framework was applied in this study to yield estimates with geographic, demographic and socioeconomic features. Based on the standard work-flow of the application of the MRP tool, this study consists of three stages: Firstly, the dataset from a election opinion poll and the Canadian Census dataset were collected and processed. Particularly, the vote intention was dichotomized to focus on the Liberals vs the Conservatives; Secondly, the multilevel logistic regression model was trained with the election poll data, and the best model was selected based on AIC (Akaike information criterion); Thirdly, the election forecast was generated using the census dataset and the results were demonstrated visually. As a result, this study showed that if all eligible Canadians had made their electoral preferences in the 2019 Canadian Federal Election, the Liberals might lead the popular votes by a narrow margin (50.32%) compared to the Conservatives (49.68%). Meanwhile, 60.9% of the Canadian population have the probability of greater than 0.5 in voting for the Liberal Party of Canada.

# Methodology

## Data

### Data source

In general, at least two datasets were required by the MRP framework. Firstly, a dataset containing the vote choice from a survey such as a election opinion poll was used to fit a multilevel logistic regression model. Based on properties of the MRP framework, this dataset can be a non-representative sample of the targeted population. Secondly, a representative dataset such as the census dataset was used to generate the vote choice estimates.

In terms of the first dataset, the election poll was obtained via a R package, `cesR` (Hodgetts and Alexander 2020). Particularly, the dataset gathered through a online sampling study, the 2019 Canadian Election Study, was used in this work (Stephenson et al. 2020; Stephenson and Loewen, n.d.). This survey had a two-wave panel, including the campaign period and post-election stages. The campaign period survey, which is the pre-election poll contains 37822 respondents with 620 variables, started from September 13th to October 21st 2019. The targeted population of this survey was the all Canadian citizens and permanent residents, aged 18 or older. Thus, all enrolled respondents were filtered by these criteria. The online sampling was conducted through a survey software Qualtrics, while the sampling frame was a list of panels on Qualtrics, which were labeled by the panel ID (Stephenson et al. 2020; Stephenson and Loewen, n.d.). And the sampling unit is eligible Canadian person, and the targets of this survey were stratified by region and balanced on gender and age within each region (Stephenson et al. 2020; Stephenson and Loewen, n.d.). In this work, the survey data was further preprocessed by filtering out all N/A and meaningless values, such as "Don't know" and "Prefer not to answer" in the variable eduction level and voting intention.

The second dataset was simulated by the census profile from the 2016 Canadian national census, which is the Census Program providing a statistical portrait of Canada every five years (Canada 2016). Since there was no census data available in 2019, the 2016 census data is the most updated one. In this study, it is assumed that the general Canadian population remained the same in 2019 compared to the population in 2016. The overall response rate was 98.4%, which was made up of responses from different census data collection methods (Canada 2016). Particularly, in 2016, Canadian households had many options to participate in the census, including responding online, completing and mailing back a paper questionnaire, contacting the Census Help Line or through telephone or in-person follow-up (Canada 2016). What's more, the Canadian household is required by law to complete the census questionnaire (Canada 2016). In addition, the wave methodology was applied to increase the coverage of the Census Program, which involves contacting non-respondent households at key times to remind them to participate in the census and persuade them to complete the questionnaire (Canada 2016). The sampling frame of the census was a enumerated list of 15.4 million dwellings in Canada. The sampling unit was a Canadian individual person, which means the observation was at the individual level and the entire population of Canada was enumerated as well. To further improve the efficiency of census information collection, 46000 collection units were set up according to the geographic location. The census contains different types of sampling methods. For instance, as a part of the general census, about 25% of the private households were received the long-form questionnaire to collect demographic and socioeconomic information of the Canadian population. In terms of this part of the census, the population is the entire Canadian private household. A stratified systematic sampling design was used to select 25% of the household from a list of private households (Canada 2016). Since the individual level census dataset was not publicly available, the 2016 Census Profile data, which is a subgroup level summary of the census data, was retrieved from Canadian Census Analyser (Canada 2016). The individual level dataset was simulated using this profile data through the `sample` function in `R` (R Core Team 2020).

Based on the previous literatures, variables such as age, gender, province location, education level and unemployment rate were selected as primary variables of interests (Rohan Alexander and Leslie 2019). The dependent variable is the vote choice in the 2019 Canadian Federal Election, and the vote was dichotomized where the Conservative Party was set as the reference level. At last, the features and subgroup levels of the above two datasets must be matched with each other.

### Data Description

The characteristics summary of these two datasets were demonstrated in Table 1 and Table 2. As shown in Table 1 and Table 2, there were 17303 observations in the survey data after preprocessing and 833296 observations in the simulated census dataset. The vote was binary, including "Liberal Party" and "Conservative Party", where the Conservatives was chosen as the reference level. 50.7% of them would like to vote for the Liberals, while the other 49.3% would like to vote for the conservatives. The age was binary as well, including "15-64" and "65-" to match with the census data. Particularly, the age 15 means this individual was aged 15 in 2016, which means aged 18 in 2019. The continuous age in the survey data was grouped in the same way according to the census data. The gender was not balanced in the survey data, where man was only 46.1% of the whole sample. As shown in Table 2, the gender distribution of the Canadian population in 2016 was 49% of man and 51% of woman.

Table. 1: Characteristics Summary: the Survey Data

| | Overall (N=17303) |
|---|---|
| **Liberals VS. Conservatives** | |
| Conservative Party | 8532 (49.3%) |
| Liberal Party | 8771 (50.7%) |
| **Age** | |
| 15-64 | 13078 (75.6%) |
| 65- | 4225 (24.4%) |
| **Gender** | |
| Man | 7984 (46.1%) |

| | Overall (N=17303) |
|---|---|
| Woman | 9319 (53.9%) |
| **Education** | |
| Under High School | 734 (4.2%) |
| High School | 4521 (26.1%) |
| Post High School | 12048 (69.6%) |
| **Employment** | |
| Unemployed | 639 (3.7%) |
| Employed | 16664 (96.3%) |
| **Provinces and Territories** | |
| Alberta | 2527 (14.6%) |
| British Columbia | 1847 (10.7%) |
| Manitoba | 823 (4.8%) |
| New Brunswick | 378 (2.2%) |
| Newfoundland and Labrador | 306 (1.8%) |
| Northwest Territories | 10 (0.1%) |
| Nova Scotia | 465 (2.7%) |
| Nunavut | 15 (0.1%) |
| Ontario | 7196 (41.6%) |
| Prince Edward Island | 63 (0.4%) |
| Quebec | 3012 (17.4%) |
| Saskatchewan | 647 (3.7%) |
| Yukon | 14 (0.1%) |

Table. 2: Characteristics Summary: the Census Data

| | Overall (N=833296) |
|---|---|
| **Age** | |
| 15-64 | 664997 (79.8%) |
| 65- | 168299 (20.2%) |
| **Gender** | |
| Man | 408726 (49.0%) |
| Woman | 424570 (51.0%) |
| **Education** | |
| Under High School | 152337 (18.3%) |
| High School | 303315 (36.4%) |
| Post High School | 377644 (45.3%) |
| **Employment** | |
| Unemployed | 47772 (5.7%) |
| Employed | 785524 (94.3%) |
| **Provinces and Territories** | |
| Alberta | 93943 (11.3%) |
| British Columbia | 112348 (13.5%) |
| Manitoba | 29258 (3.5%) |
| New Brunswick | 18271 (2.2%) |
| Newfoundland and Labrador | 12584 (1.5%) |
| Northwest Territories | 917 (0.1%) |
| Nova Scotia | 22401 (2.7%) |
| Nunavut | 681 (0.1%) |
| Ontario | 319554 (38.3%) |
| Prince Edward Island | 3325 (0.4%) |
| Quebec | 194408 (23.3%) |

| | Overall (N=833296) |
|---|---|
| Saskatchewan | 24813 (3.0%) |
| Yukon | 793 (0.1%) |

The differences in the gender between the two datasets were demonstrated in the Figure 1 as well. What's more, differences between these two datasets were also found in other geographical and socio-economical variables. For instance, the respondents from the poll were more educated compared to the general Canadian, with proportions of post high school level were 69.6% and 45.3%, respectively, and proportions of under high school level were 4.2% and 18.3%, respectively (Figure 1). In terms of the province, Alberta, British Columbia, Ontario and Quebec are the most important provinces, because these four provinces provide most seats in the Parliament. In the survey data, there were 14.6% of samples in Alberta, 10.7% in British Columbia, 41.6% in Ontario and 17.4% in Quebec. Meanwhile, there were 11.3% of the population in Alberta, 13.5% in British Columbia, 38.3% in Ontario and 23.3% in Quebec for the census data.

The voting intention collected in the election opinion poll was shown in the Figure 2 at the provinces and territories level. In general, the Conservatives dominated the Alberta (78%), tied with the Liberals in British Columbia, while the Liberals had a steady lead in Ontario (56%) and Quebec (67%). The vote choices in these four provinces are most important.

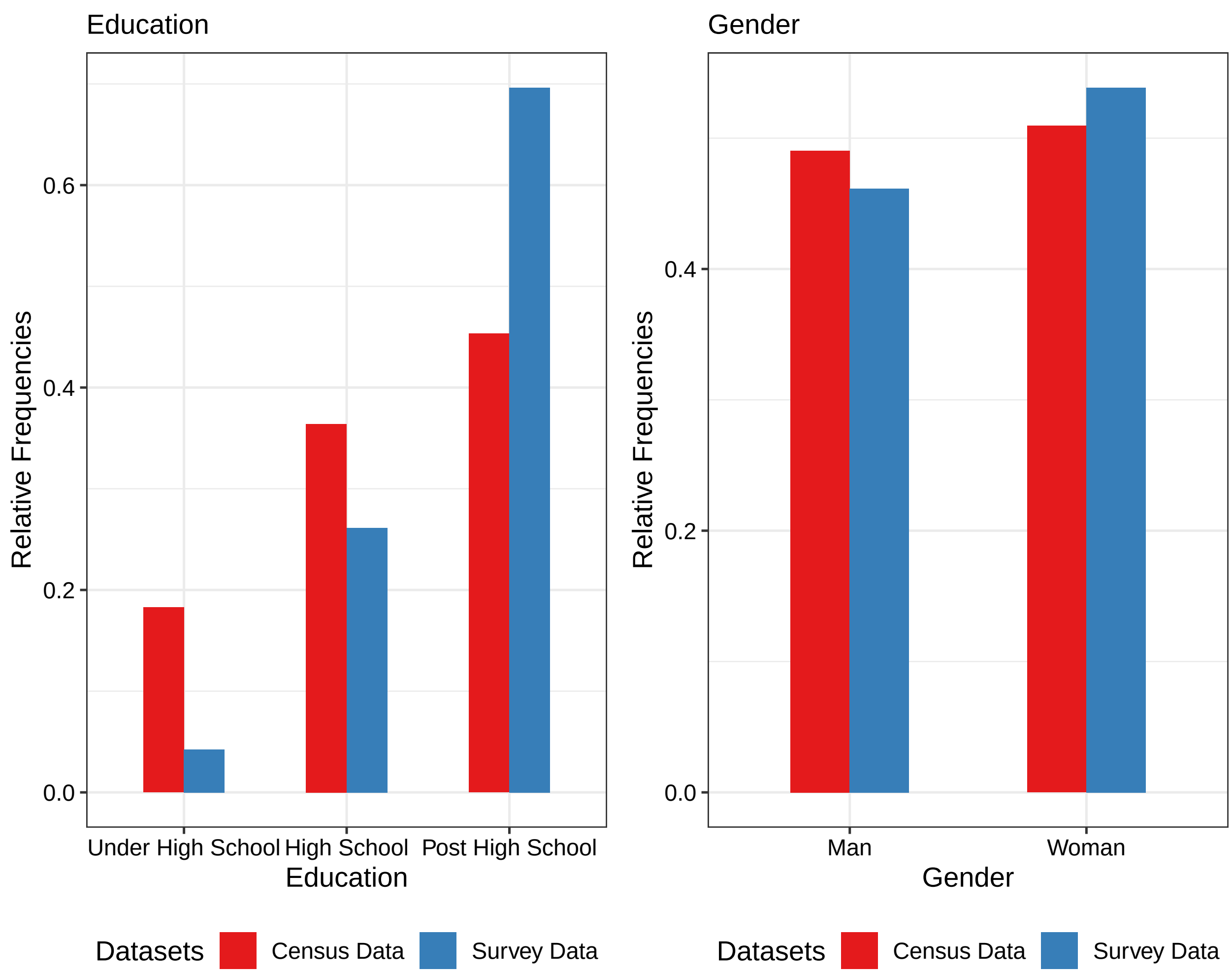


Figure 1: Comparison of Survey Data and Census Data

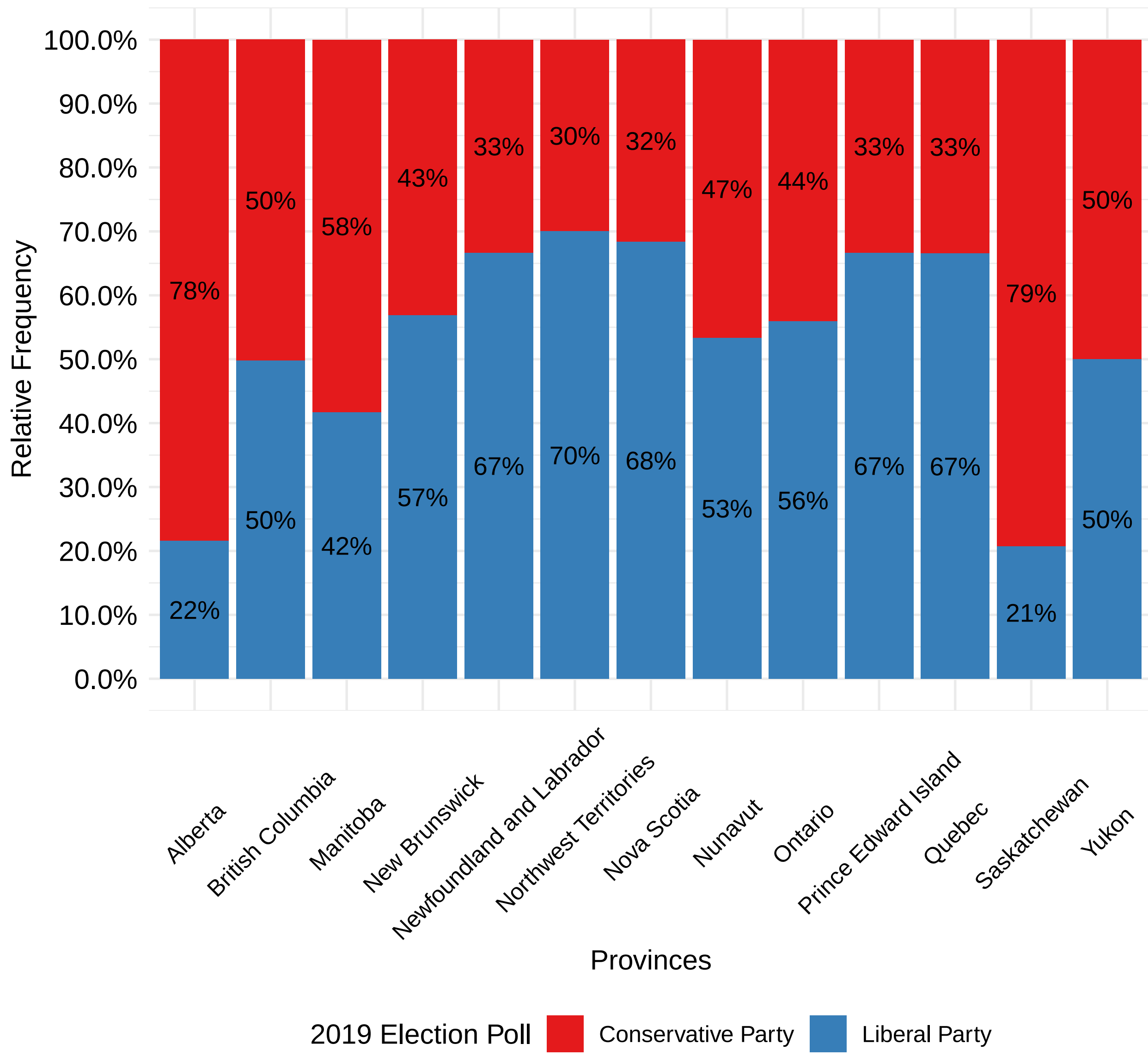


Figure 2: Vote Choice by Province: Opinion Poll

## Model

The core idea of the MRP tool is to divide the entire population into many small subgroups based on the combinations of socioeconomic, geographic and demographic variables, then estimates are made by post-stratification, which is the aggregation of estimates from subgroups by taking the average in proportion to the group size in the population (Ghitza and Gelman 2013; Wei Wang and Gelman 2014). In order to generate robust and accurate subgroup-level estimates, a multilevel logistic regression model fitted, instead of simply taking the average within each cell. The combination of multilevel logistic regression and post-stratification is the MRP framework used in this work.

First, a multilevel logistic regression was fit with categorical variables using the survey data. In details, categorical variables such as `Province`, `Gender`, age groups (`Age`), education levels (`Education`) and unemployment rate (`Employment`) were chosen as variables of interest. The dependent variable was `Vote`. Based on the previous literatures, those socioeconomic, geographic and demographic features were selected as primary variables of interests (Rohan Alexander and Leslie 2019; Wei Wang and Gelman 2014). AIC was used in the model selection and comparison, where smaller AIC represents a better model. A difference of AIC which is greater than 2 is considered as a significant difference between fittings of models. AIC is calculated from the log likelihood and the number of variables. Therefore, it is widely used to find the best model which achieves the desired level of fit with as few variables as possible.

The model selection procedure was started with the full model, which was shown here:

$$Vote \sim Bernoulli(\frac{1}{1 + exp(-(Intercept + b_i x_i))}) \tag{1}$$

where $x_i$ and $b_i$ represent the variables and corresponding coefficients set of age, gender, education, province and employment.

Then the reduced models were fitted and the final model was chosen as:

$$Vote \sim Bernoulli(\frac{1}{1 + exp(-(Intercept + b_j x_j))}) \tag{2}$$

where $x_j$ and $b_j$ represent the variables and corresponding coefficients set of age, gender, education and province.

All multilevel logistic regression model were fitted with `glm` function in R (R Core Team 2020). Equation (1) represents the full model, and Equation (2) represents the final model, where the employment was excluded.

With the selected trained model, the predictions were made with the simulated census data. Through post-stratification, the differences between the non-representative dataset and the target population were corrected. Finally, the probability of voting for the Liberals were predicted at the individual level and aggregated at different group levels.

All work were done in `R` (version 4.0.2) (R Core Team 2020) and `Rstudio` (version 1.3.1093). `Tidyverse` (version 1.3.0) and `ggplot2` was used for data wrangling and visualization (Wickham et al. 2019). R packages including `captioner`, `gridExtra`, `broom`, `magrittr`, `knitr`, `stargazer`, `RColorBrewer`, `labelled` and `arsenal` were used in this work (Alathea 2015; Hlavac 2018; Heinzen et al. 2020; Xie 2020; Wickham and Miller 2020; Auguie 2017; Robinson, Hayes, and Couch 2020; Bache and Wickham 2014; Larmarange 2020; Neuwirth 2014). Particularly, `arsenal` and `stargazer` were used to generate summary tables (Hlavac 2018; Heinzen et al. 2020). Code are available at: https://github.com/minchencai/Final-Project.

# Results

Table. 3: Mutilevel Logistic Regression Models

Table 3:

| | *Dependent variable:* | |
|---|---|---|
| | Vote | |
| | Full Model | No Employment |
| | (1) | (2) |
| Age: 65- | −0.001 | −0.006 |
| | (0.038) | (0.038) |
| Gender: Woman | 0.358*** | 0.358*** |
| | (0.033) | (0.033) |
| Education: High School | −0.039 | −0.044 |
| | (0.085) | (0.085) |
| Education: Post High School | 0.387*** | 0.380*** |
| | (0.081) | (0.081) |
| Employment: Employed | −0.124 | |
| | (0.087) | |
| Observations | 17,303 | 17,303 |
| Log Likelihood | −11,023.200 | −11,024.220 |
| Akaike Inf. Crit. | 22,082.410 | 22,082.440 |

*Note:* *p<0.1; **p<0.05; ***p<0.01

Province variable was included but not shown here.

The multilevel logistic regression models fit in this study with the survey data were summarized in Table 3. The full model was the model 1 shown in the Table 3. In the full model, the binary age variable and the binary employment variable were not significant in the full model. Thus, the reduced model without `Employment` was fit as model 2. Since `Age` is a basic variable which has been widely reported to be important in political context, `Age` was not excluded in the final model. Model 1 and model 2 had similar AICs (22082.4), while model 2 had one less variable. Thus, model 2 which contained age, gender, education and province was chosen as the final model.

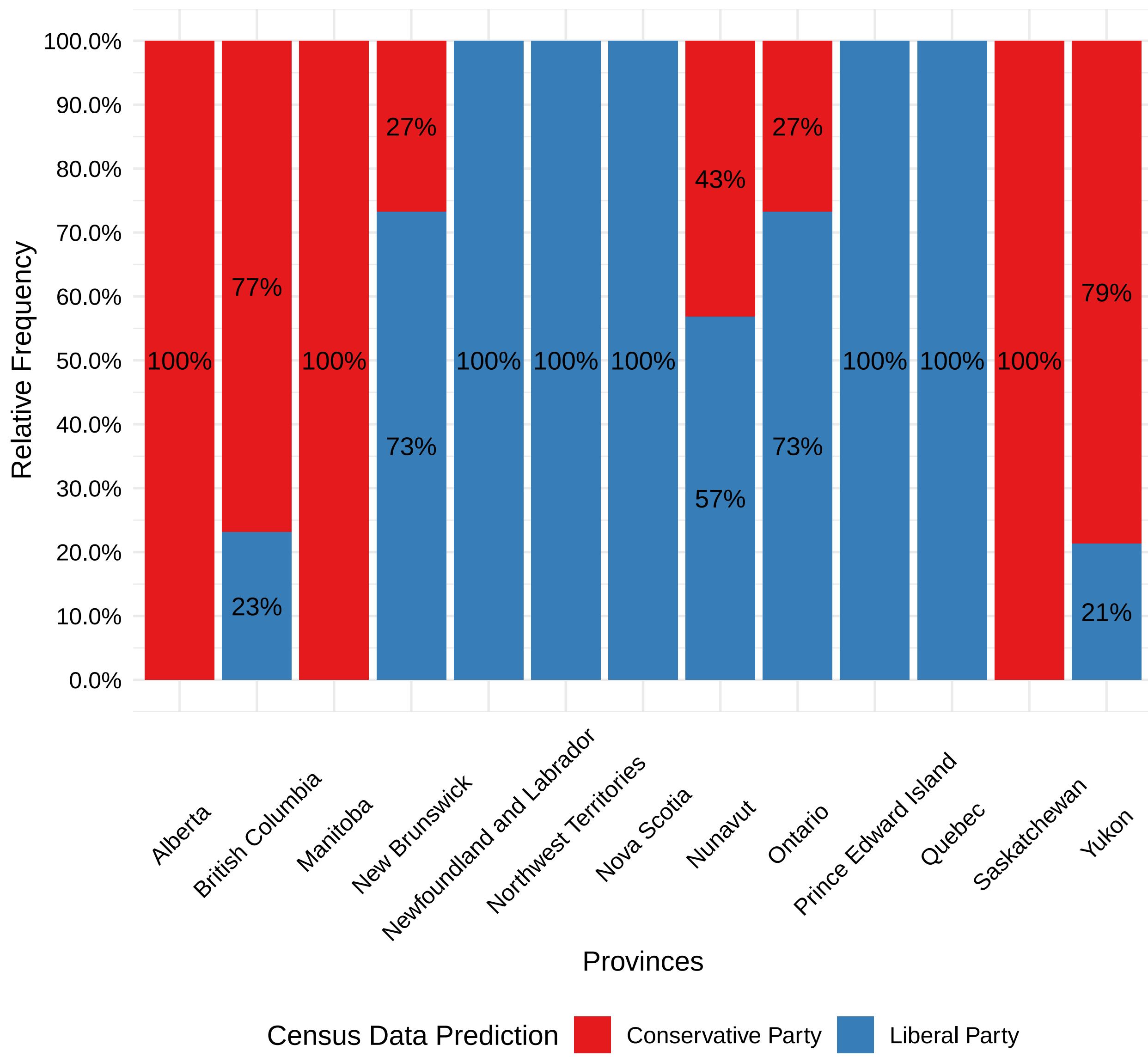


Figure 3: Election Prediction by Province with 2016 Canadian Census Data

The estimates were made by predicting the probability of voting for the Liberals using the 2016 Canadian census data. Then the threshold was chosen with the conservative probability 0.5, where estimates greater than 0.5 were grouped as the "Liberal Party", otherwise were grouped as the "Conservative Party". As shown in Figure 3, the vote choice prediction was visualized at the province level. The Liberals were predicted to dominate Ontario and Quebec while the Conservatives were predicted to dominate Alberta and British Columbia. The results here were similar to what happened in the 2019 election, where the Liberals lead against the Conservatives in Ontario and Quebec. The differences between the prediction and vote results in 2019 were the sizes of the lead of each party in corresponding province.

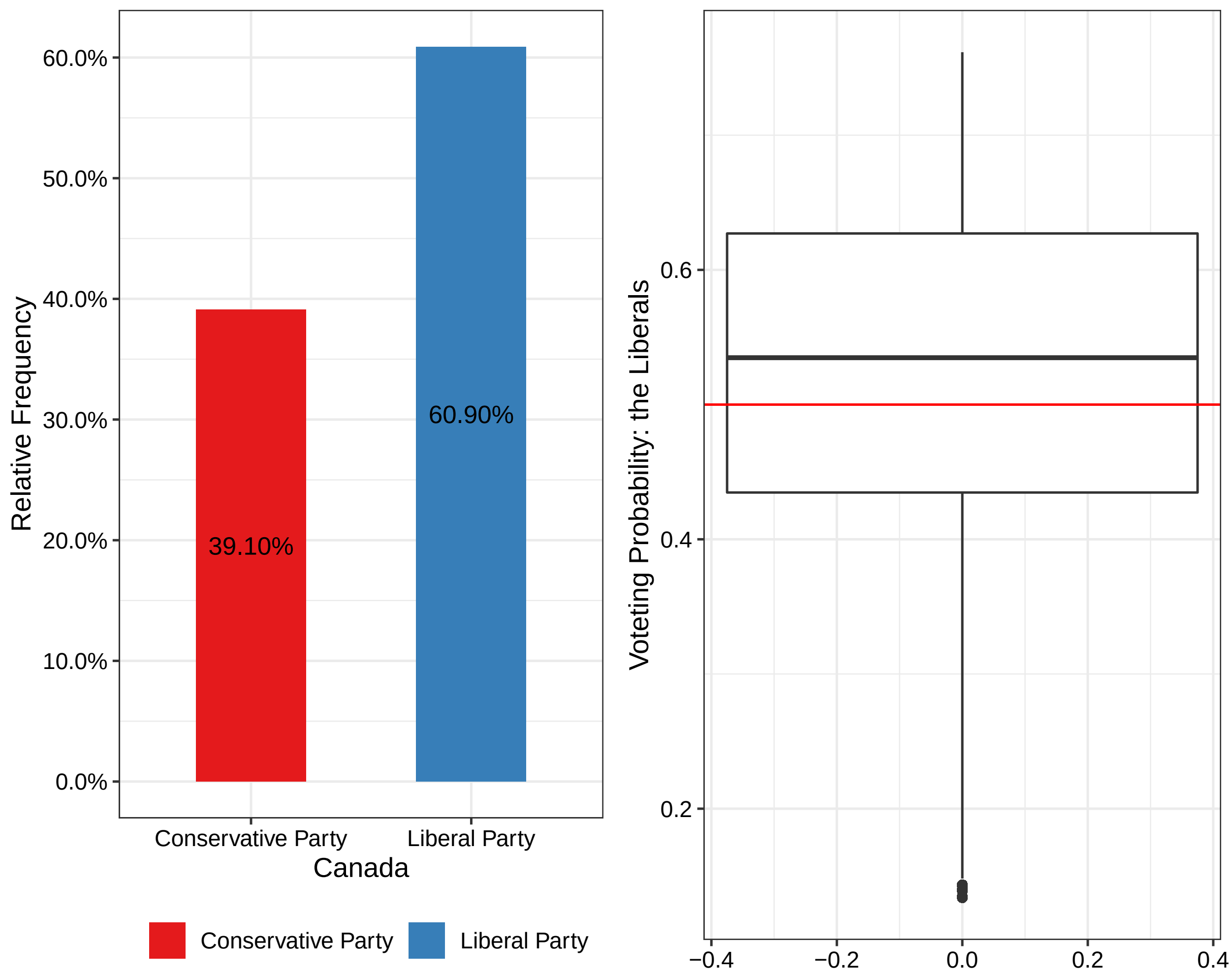


Figure 4: Overall Election Prediction with 2016 Canadian Census Data

Overall, as shown in the left panel of Figure 4, the Liberals could win 60.9% of the popular votes when compared to the Conservatives if all eligible Canadians had voted. As shown in the right panel of Figure 4, all estimated probabilities of voting for the Liberals were summarized. The median of the probabilities, 0.5348, is greater than 0.5, and the average, 0.5032 , is greater than 0.5 as well. This suggests the Liberals would win the popular votes against the Conservatives, where more than half of the entire population tend to vote for the Liberals.

# Discussion

Overall, the competition between the Liberals and the Conservatives in the 2019 Canadian Federal election was very close. To predict what the results of the popular votes would be if all eligible Canadian had voted, a simple logistic regression model was fit in this study as the final regression model, with only four categorical variables, age, gender, education and province. In general, the results generated from this simple model was reasonable when compared to the real 2019 Canadian election results.

The results of this study suggest that the Liberals might win more popular votes, if all eligible Canadians had voted. It has been reported that the voter turnout of the 2019 Canadian federal election was around 66% - 67%. Thus the people who voted in 2019 was not a good representation of the entire eligible population. It has been shown that the results from voluntary turnout are significantly different from the results of everyone turns out to vote (Fowler 2013). This might partially explain the disagreement between vote prediction from this study and real result of 2019. The results in this study also demonstrated that gender, province and education are important variables in election prediction (Wei Wang and Gelman 2014; Rohan Alexander and Leslie 2019). Detailed information of our final model fitting can be found in the **Appendix**.

Low and unequal voter turnout has been described as "democracy's unresolved dilemma" (Fowler 2013). Because the preferences of marginal voters are often different from those of regular voters, low voter turnout is a serious problem (Fowler 2013). This study provides a potential solution to discover the representative political opinions of the citizenry. All statistical modeling has two frames: the small world of the model itself and the large world we hope to deploy the model in (McElreath 2016). One of the benefits provided by MRP is that, it has the potential to be a unbiased represent of the targeted population, by the introducing of the post-stratification dataset. In this work, by using the census data, the results have the unique power to represent the Canadian citizenry, at voter turnout 100%, which usually can not be achieved by opinion polls.

One limitation of the MRP is that it requires two matched datasets. Due to the different designs in data collections and preparations, it is very difficult to find two exactly matched datasets. Even though the two datasets are generally matched in this work, certain levels of some categorical variables are not matched exactly. The survey data gathered the information of gender, which including man, woman and the others, while the census data recorded the sex and treated it as a binary feature. To match the datasets, this study failed to adopt a non-binary gender setting. Nowadays, a correct and adaptive way in coding and constructing sex or gender is becoming more and more important (Kennedy et al. 2020). A binary gender variable is a biased way in representing the population. Similarly, the age group was dichotomized in this work, because the age was not grouped as above 18 and below 18 in the 2016 census data. This definitely leads to the loss of power in the variable. Another limitation is that, there was no 2019 census data available. Moreover, the 2016 census data is only available at group level. In this study, the education and employment variables were simulated at the province level, while other variables were simulated at the country level. As a result, accurate small subgroup level estimates are not appropriate in this study.

In the future work, the accuracy of the prediction made in this study can be improved dramatically with a complete individual observation level census data. In this way, the geographical, demographical and socioeconomical features of each small subgroups are accurate and representative. The power of the MRP tool will be maximized with a precise census data, since the estimate of each small subgroup will be precise and robust (Ghitza and Gelman 2013). Another way to improve the prediction is to incorporate more survey data in the model training stage of MRP. The MRP tool is able to combine information from many different surveys by assigning different weights to corresponding datasets (Ghitza and Gelman 2013). One way to weight the survey data is based on the date of survey. Because the election opinion polls are more accurate as approaching the election day.

# Appendix

```r
######## full model ###########
tidy(Mod1)
```

```
## # A tibble: 18 x 5
##    term                              estimate std.error statistic   p.value
##    <chr>                                <dbl>     <dbl>     <dbl>     <dbl>
##  1 (Intercept)                      -1.63        0.122   -13.4    4.25e- 41
##  2 Age65-                           -0.000652    0.0379   -0.0172 9.86e-  1
##  3 GenderWoman                       0.358       0.0326   11.0    5.65e- 28
##  4 ProvinceBritish Columbia          1.31        0.0678   19.4    9.17e- 84
##  5 ProvinceManitoba                  0.968       0.0863   11.2    3.45e- 29
##  6 ProvinceNew Brunswick             1.56        0.115    13.5    1.12e- 41
##  7 ProvinceNewfoundland and Labrador 1.97        0.132    15.0    7.96e- 51
##  8 ProvinceNorthwest Territories     2.17        0.693     3.13   1.72e-  3
##  9 ProvinceNova Scotia               2.06        0.112    18.4    7.09e- 76
## 10 ProvinceNunavut                   1.42        0.522     2.72   6.56e-  3
## 11 ProvinceOntario                   1.53        0.0543   28.1    6.78e-174
## 12 ProvincePrince Edward Island      2.00        0.274     7.28   3.27e- 13
## 13 ProvinceQuebec                    2.02        0.0625   32.4    1.08e-229
## 14 ProvinceSaskatchewan             -0.0728      0.109    -0.668  5.04e-  1
## 15 ProvinceYukon                     1.21        0.541     2.24   2.54e-  2
## 16 EducationHigh School             -0.0390      0.0847   -0.461  6.45e-  1
## 17 EducationPost High School         0.387       0.0811    4.77   1.81e-  6
## 18 EmploymentEmployed               -0.124       0.0870   -1.42   1.54e-  1
```

```r
######## no employment ######
tidy(NonEmployment)
```

```
## # A tibble: 17 x 5
##    term                              estimate std.error statistic   p.value
##    <chr>                                <dbl>     <dbl>     <dbl>     <dbl>
##  1 (Intercept)                      -1.74      0.0936   -18.6    2.01e- 77
##  2 Age65-                           -0.00560   0.0377    -0.149  8.82e-  1
##  3 GenderWoman                       0.358     0.0326    11.0    4.53e- 28
##  4 ProvinceBritish Columbia          1.31      0.0677    19.4    1.56e- 83
##  5 ProvinceManitoba                  0.967     0.0863    11.2    3.92e- 29
##  6 ProvinceNew Brunswick             1.56      0.115     13.5    1.23e- 41
##  7 ProvinceNewfoundland and Labrador 1.97      0.132     15.0    9.35e- 51
##  8 ProvinceNorthwest Territories     2.17      0.693      3.12   1.78e-  3
##  9 ProvinceNova Scotia               2.06      0.112     18.4    1.00e- 75
## 10 ProvinceNunavut                   1.41      0.522      2.71   6.78e-  3
## 11 ProvinceOntario                   1.52      0.0543    28.1    1.28e-173
## 12 ProvincePrince Edward Island      1.99      0.274      7.26   3.85e- 13
## 13 ProvinceQuebec                    2.02      0.0625    32.3    2.25e-229
## 14 ProvinceSaskatchewan             -0.0737    0.109     -0.676  4.99e-  1
## 15 ProvinceYukon                     1.20      0.541      2.22   2.61e-  2
## 16 EducationHigh School             -0.0436    0.0847    -0.514  6.07e-  1
## 17 EducationPost High School         0.380     0.0809     4.70   2.64e-  6
```

# References


Alathea, Letaw. 2015. *Captioner: Numbers Figures and Creates Simple Captions.* https://CRAN.R-project.org/package=captioner.

Auguie, Baptiste. 2017. *GridExtra: Miscellaneous Functions for "Grid" Graphics.* https://CRAN.R-project.org/package=gridExtra.

Bache, Stefan Milton, and Hadley Wickham. 2014. *Magrittr: A Forward-Pipe Operator for R.* https://CRAN.R-project.org/package=magrittr.

Canada, Statistics. 2016. *2016 Census of Canada.* Retrieved%20from%20http://dc1.chass.utoronto.ca/census/index.html.

Fowler, Anthony George. 2013. "Five Studies on the Causes and Consequences of Voter Turnout." *Doctoral Dissertation.*

Ghitza, Yair, and Andrew Gelman. 2013. "Deep Interactions with Mrp: Election Turnout and Voting Patterns Among Small Electoral Subgroups." *American Journal of Political Science* 57 (3): 762–76. https://doi.org/https://doi.org/10.1111/ajps.12004.

Heinzen, Ethan, Jason Sinnwell, Elizabeth Atkinson, Tina Gunderson, and Gregory Dougherty. 2020. *Arsenal: An Arsenal of 'R' Functions for Large-Scale Statisticalsummaries.* https://CRAN.R-project.org/package=arsenal.

Hlavac, Marek. 2018. *Stargazer: Well-Formatted Regression and Summary Statistics Tables.* Bratislava, Slovakia: Central European Labour Studies Institute (CELSI). https://CRAN.R-project.org/package=stargazer.

Hodgetts, Paul A., and Rohan Alexander. 2020. *CesR: Access the Ces Datasets a Little Easier.*

Institute, Angus Reid. 2019. *Electoral Reform Revival? Support for Changing Voting Systems Skyrockets Post Election.* http://angusreid.org/electoral-reform-trend/.

Kennedy, Lauren, Katharine Khanna, Daniel Simpson, and Andrew Gelman. 2020. "Using Sex and Gender in Survey Adjustment." https://arxiv.org/abs/2009.14401.

Larmarange, Joseph. 2020. *Labelled: Manipulating Labelled Data.* https://CRAN.R-project.org/package=labelled.

McElreath, Richard. 2016. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan.* CRC Press/Taylor &amp, Francis Group.

Neuwirth, Erich. 2014. *RColorBrewer: ColorBrewer Palettes.* https://CRAN.R-project.org/package=RColorBrewer.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Robinson, David, Alex Hayes, and Simon Couch. 2020. *Broom: Convert Statistical Objects into Tidy Tibbles.* https://CRAN.R-project.org/package=broom.

Rohan Alexander, Patrick Dumount, and Patrick Leslie. 2019. "Forecasting Multi-District Elections." https://github.com/RohanAlexander/ForecastingMultiDistrictElections/blob/master/outputs/paper/paper.pdf.

Stephenson, Allison Harrel, Laura, and Peter Loewen. n.d. "Measuring Preferences and Behaviour in the 2019 Canadian Election Study." *Canadian Journal of Political Science.*

Stephenson, Laura B, Allison Harell, Daniel Rubenson, and Peter John Loewen. 2020. "2019 Canadian Election Study - Online Survey." https://doi.org/10.7910/DVN/DUS88V.

Wei Wang, Sharad Goel, David Rothschild, and Andrew Gelman. 2014. "Forecasting Elections with Non-Representative Polls." https://www.microsoft.com/en-us/research/wp-content/uploads/2016/04/forecasting-with-nonrepresentative-polls.pdf.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, and Evan Miller. 2020. *Haven: Import and Export 'Spss', 'Stata' and 'Sas' Files.* https://CRAN.R-project.org/package=haven.

Xie, Yihui. 2020. *Knitr: A General-Purpose Package for Dynamic Report Generation in R.*