

Increased Voter Turnout Would Change the Result of 2019 Canadian Federal Election*

Election Prediction with the Multilevel Logistic Regression with Post-Stratification (MRP)
Framework

Guangyu Du

December 18th, 2020

Abstract

About 66% of Canadian electors voted for the 2019 Canadian Federal Election, and this election was considered as a very difficult one to predict in terms of the competition between the Liberals and the Conservatives. Particularly, the Liberal Party of Canada which led by the Prime Minister Just Trudeau won the most seats by a narrow margin and formed a minority government, but lost the popular vote to the Conservative Party of Canada. In this work, the multilevel (logistic) regression with post-stratification (MRP) framework was trained firstly with the Canadian Election Survey 2019 data, then the prediction with post-stratification was made by the trained model with 2016 Canadian Census Profile data which is a simulated representative dataset of the whole Canadian population. This study demonstrated that, if all eligible Canadians had voted in this federal election, the Liberals would win the popular vote.

Keywords: Multilevel Regression with Post-Stratification (MRP); Liberal Party of Canada; 2019 Canadian Federal Election; Voter Turnout; Representative Sample; Prediction;

Introduction

The competition of the Liberals and the Conservatives during 2019 Canadian federal election was very intense. Even though Justin Trudeau barely won the most seats to form the minority government eventually, a lot of pre-campaign period and campaign period national polls provided results of voting intention against him. Nowadays, it is still very interesting to figure out why Justin Trudeau could beat the worst expectation against him and won again as the Prime Minister. What's more, since the campaign period national polls had very conflicting results at that time, it is crucial to figure out why political polls results were quite different from each other and how to improve further polling method based upon this. One possible answer to these two questions is the non-representative sampling datasets issue, which means the sample used in those polls might not be a good representative of the whole Canadian population who are eligible to vote, even the electors who voted in the 2019 election might not be inclusive enough (Wei Wang and Gelman 2014; Ghitza and Gelman 2013). The latter one is the issue related with the voter turnout, which is the measurement of civic participation in election. The election results might be significantly altered if the voter turnout were increased (Fowler 2013). In almost every election in democracy world, many citizen fail to participate in voting, and those who vote are systematically unrepresentative of the eligible population (Fowler 2013). Low and unequal turnout is a serious problem because election results fail to reflect the opinions of the citizenry (Fowler 2013). Low turnout pose a serious challenge to the accuracy of election opinion polls as well.

Although the standard solution is to use the representative polls and the ideal scenario for election is that the people who voted have similar political attitudes with non-voters. Several challenges are presented: the cost of representative polls has becoming enormous, the response rate keeps decreasing in the past several decades and

* Code are available at: <https://github.com/spicapica/stats/tree/main/Final>

the characteristics of voters could be different from non-voters such as there might be differences in age and gender distributions. Fortunately, a practical approach to generate unbiased results from non-representative sampling dataset is to use the multilevel regression with post-stratification (MRP) framework (Ghitza and Gelman 2013). It has been demonstrated that the MRP framework was able to generate accurate forecast compared to the results generated from the representative polls, correcting the known biases between sample and the target population (Wei Wang and Gelman 2014). What’s more, the MRP model can generate vote choice by subgroups defined by multiple demographic and geographic characteristics (Ghitza and Gelman 2013; Wei Wang and Gelman 2014).

In this work, the MRP framework was applied to revisit and simulate the results of 2019 Canadian Federal Election: Firstly, the multilevel logistic regression model was trained using the `glm` function in R, the response variable, `voting_intention`, was dichotomized to focus on the Liberals vs the Conservatives; Secondly, the vote intention of general Canadian population was generated by making prediction with the trained multilevel logistic regression model using Canadian census data. At the first stage, AIC (Akaike information criterion) was used in the backward model selection. Since the AIC taking both of the log likelihood and the number of estimated parameter into account, it is widely used in model selection to find the most parsimonious model, where the model achieves the desired level of fit with the least number of variables.

The results demonstrated that increased voter turnout would change the result of the 2019 Canadian Federal Election, where the Liberals would win the popular votes by 2.2% against the Conservatives, instead of losing the popular votes by 0.9%.

Methods

Data

Data Description

To set up the MRP tool, two datasets were used in this study, where the first dataset is usually the non-representative survey data with the voting intention variable to build the multilevel logistic regression model and the second dataset is the representative census dataset to represent the general population of all eligible Canadians. Both of the datasets were used with permission.

Particularly, the first dataset was obtained via a R package, `cesR` (Hodgetts and Alexander 2020). This package provides access to multiple election opinion polls, from which the dataset generated by the 2019 Canadian Election Study was used in this work (Stephenson et al. 2020; Stephenson and Loewen, n.d.). This is a online sampling study which composed of a two-wave panel, including the campaign period and post-election stages. The campaign period survey, which contains 37822 respondents with 620 variables, was used in this study. And this pre-election survey started from September 13th to the day of election October 21st, 2019. All enrolled respondents needed to be aged 18 or over, and Canadian citizens or permanent residents. Thus, the targeted population of this survey was the set of all Canadian citizens and permanent residents, aged 18 or older. The online sampling was conducted through a commercial survey software Qualtrics. The sampling unit of this survey is an eligible Canadian individual, 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.). The sampling frame was a list of panels on Qualtrics, which are identifiable by the panel ID. To improve the quality of the data, marginal values were successively weighted according to province, as well as gender, age group and education level (Stephenson et al. 2020; Stephenson and Loewen, n.d.). The survey data was preprocessed by filtering out ineligible observations by age and citizenship, and all N/A and meaningless values were removed.

The second dataset was the census profile from 2016 Canadian national census, which is the Census Program providing a statistical portrait of Canada every five years. This is the most updated census data, and it is assumed in this study that the general Canadian population remained the same in 2019 compared to the population in 2016. The overall response rate was 98.4% at the country level, which consists of responses

from three different types of collection, including the Internet, Self-response and the Long form. That is to say, 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. The sampling frame of the census was a enumerated list of 15.4 million dwellings in Canada. For all occupied dwelling, the corresponding households should complete the census questionnaire. 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. In the 2016 Census, there were roughly 46000 collection units in Canada to make the census collection more efficient. A collection unit is a small geographic unit. 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. Because the Canadian household is required by law to complete the census questionnaire, the wave methodology was designed to encourage people to participate earlier online. The census contains different types of sampling methods. For instance, about 25% of the long-form sample universe were received the long-form questionnaire to collect demographic and socioeconomic information of the Canadian population. In details, The population of the long-form sampling was the private households type of the entire Canadian household. With the sampling frame as a list of private households, the stratified systematic sampling design was used to select 25% of the population. In this study, the 2016 Census Profile data was retrieved (Canada 2016). The individual level observational dataset was simulated using random sampling with proportions calculated by 2016 census profile data.

Finally, the variables and subgroups within each variable from the above two datasets must be matched with each other. Thus, variables including age, gender, province of the person, education level, income level and unemployment rate were selected from both datasets. Particularly, in terms of the survey data, there were 12549 observations in the final dataset after data preprocessing; in terms of the census data, 1666941 observations were simulated after the removal of observations aged under 15, because the age of 15 in 2016 is the age of 18 in 2019. What' more, the census data was divided into age groups of 15 – 64 and 65 over, and the education, income and employment variables were only available for the age group of 15 – 64. In terms of the vote choice, the Conservatives was chosen as the reference level.

Data Exploration

The characteristics summary of these two datasets were demonstrated with Table 1 and Table 2. As shown in Table 1, 12549 observations were included in the survey dataset and 50.9% of them would like to vote for the Liberals, while the other 49.1% would like to vote for the conservatives. In both datasets, the age variable was dichotomized. According to the balanced design of the survey sampling, the gender distribution in the survey data is more balanced than the gender distribution of the 2016 census data, which was demonstrated in the Figure 1 as well. Moreover, the greater differences between the survey data and the census data were found in the **Education**, **Provinces** and **Income**.

Table. 1: Characteristics: Survey Data

	Overall (N=12549)
Vote Choice	
Conservative Party	6157 (49.1%)
Liberal Party	6392 (50.9%)
Age	
15-64	9415 (75.0%)
65-	3134 (25.0%)
Gender	
Man	6267 (49.9%)
Woman	6282 (50.1%)
Provinces and Territories	
Alberta	1791 (14.3%)
British Columbia	1357 (10.8%)

	Overall (N=12549)
Manitoba	606 (4.8%)
New Brunswick	292 (2.3%)
Newfoundland and Labrador	223 (1.8%)
Northwest Territories	8 (0.1%)
Nova Scotia	342 (2.7%)
Nunavut	12 (0.1%)
Ontario	5165 (41.2%)
Prince Edward Island	38 (0.3%)
Quebec	2244 (17.9%)
Saskatchewan	461 (3.7%)
Yukon	10 (0.1%)
Education	
Under High School	475 (3.8%)
High School	3152 (25.1%)
Post High School	8922 (71.1%)
Employment Status	
Unemployed	393 (3.1%)
Employed	12156 (96.9%)
Income	
-\$29999	1366 (10.9%)
\$30000-\$69999	3765 (30.0%)
\$70000-\$99999	3038 (24.2%)
\$100000-	4380 (34.9%)

Table. 2: Characteristics: Census Data

	Overall (N=1666941)
Age	
15-64	1328813 (79.7%)
65-	338128 (20.3%)
Gender	
Man	818988 (49.1%)
Woman	847953 (50.9%)
Provinces and Territories	
Alberta	186910 (11.2%)
British Columbia	225907 (13.6%)
Manitoba	58655 (3.5%)
New Brunswick	36149 (2.2%)
Newfoundland and Labrador	25156 (1.5%)
Northwest Territories	1862 (0.1%)
Nova Scotia	45025 (2.7%)
Nunavut	1426 (0.1%)
Ontario	638029 (38.3%)
Prince Edward Island	6898 (0.4%)
Quebec	388954 (23.3%)
Saskatchewan	50262 (3.0%)
Yukon	1708 (0.1%)
Education	
Under High School	305595 (18.3%)
High School	604400 (36.3%)
Post High School	756946 (45.4%)

	Overall (N=1666941)
Employment Status	
Unemployed	95120 (5.7%)
Employed	1571821 (94.3%)
Income	
-\$29999	749451 (45.0%)
\$30000-\$69999	599252 (35.9%)
\$70000-\$99999	181945 (10.9%)
\$100000-	136293 (8.2%)

In general, the respondents from the poll were more educated compared to the general Canadian, with proportions of post high school level were 71.1% and 45.4%, respectively, and proportions of under high school level were 3.8% and 18.3%, respectively. The **Income** recorded in the survey was the household level income, while the **Income** recorded in the census was the individual level income. This may explain the differences in **Income** between the survey data and the census data. The differences in the **Province** were more robust and profound. In the survey data, there were 14.3% in Alberta, 41.2% in Ontario and 17.9% in Quebec, while there were 11.2% in Alberta, 38.3% in Ontario and 23.3% in Quebec of the Canadian population. The differences in the socio-demographics between these two datasets were demonstrated in the Figure 1 as well.

The voting intention collected in the survey data were demonstrated in the Figure 2 at the provinces and territories level. The Conservatives dominate the Alberta, have a narrow lead in British Columbia, while the Liberals have a steady lead in Ontario and Quebec. These four provinces have the top four most seats in the Parliament. Thus, the vote choices in these four provinces are most important.

Model

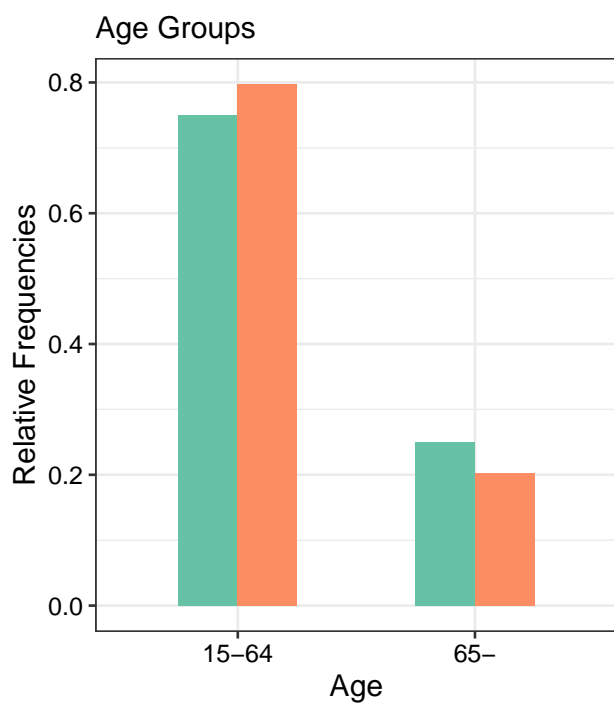
The MRP tool was developed to derive election forecasts in greater socioeconomic, geographic and demographic detail, what’s more, MRP framework is able to generate precise vote choice estimates at different levels and types of subgroups and yield accurate election forecasts using non-representative dataset compared to predictive model based on traditional representative polls (Ghitza and Gelman 2013).

In this study, the dependent variable vote was dichotomized as “Liberal Party” and “Conservative Party”. Therefore, to set up MRP, a multilevel logistic regression was fit with categorical variables. With those categorical variables, the population can be partitioned into cells, which is also called as subgroups, and each cell has its corresponding socioeconomic, geographic and demographic meaning. In details, categorical variables such as province and territory location of the voter gender, age groups (age), (province), education levels (education), income level (income) and unemployment rate (employment) were chosen as variables of interest. The dependent variable was **Vote**. Those variables were chosen based on the previous political election forecast papers and the availability of the retrieved census data set (Wei Wang and Gelman 2014; Rohan Alexander and Leslie 2019; Canada 2016).

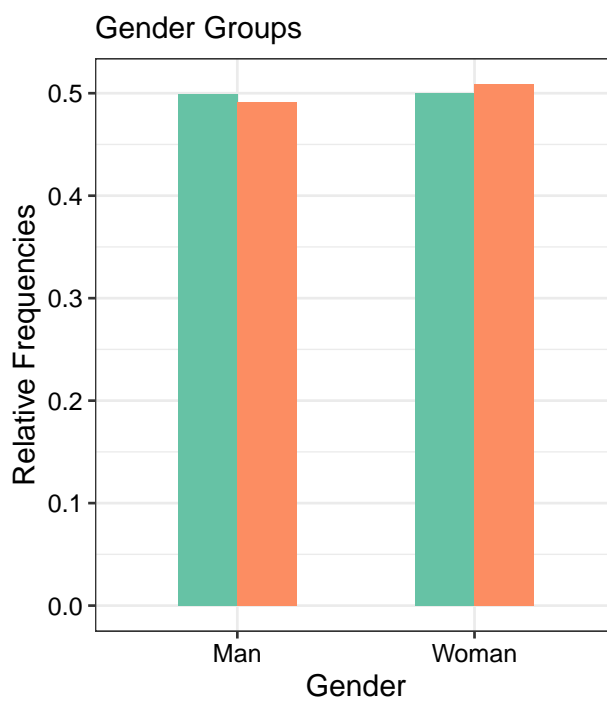
The backward model selection procedure was started with the complete model, and the goal was to identify the best model with the least number of variables and smallest AIC. The difference in AIC greater than two is considered as significant evidence of a better model. Models with similar AIC were selected based on the number of variables. And the complete model was shown here:

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

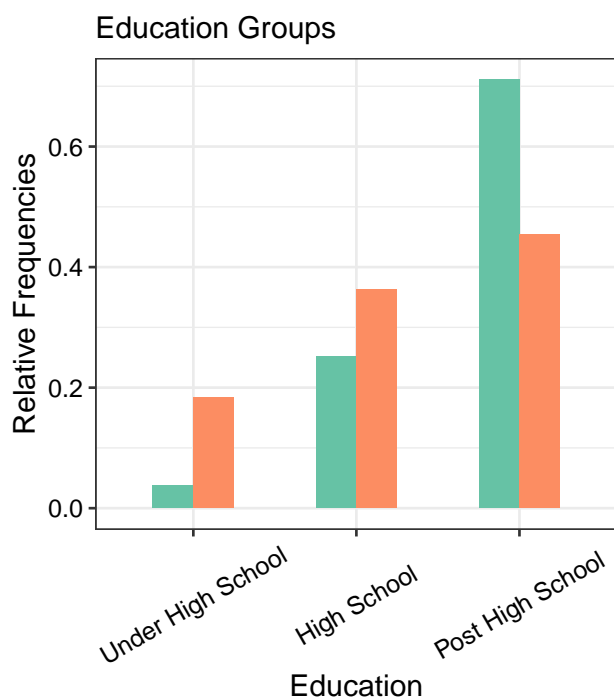
where x_i and b_i represent the variables and corresponding coefficients set of age, gender, education, province, income and employment.



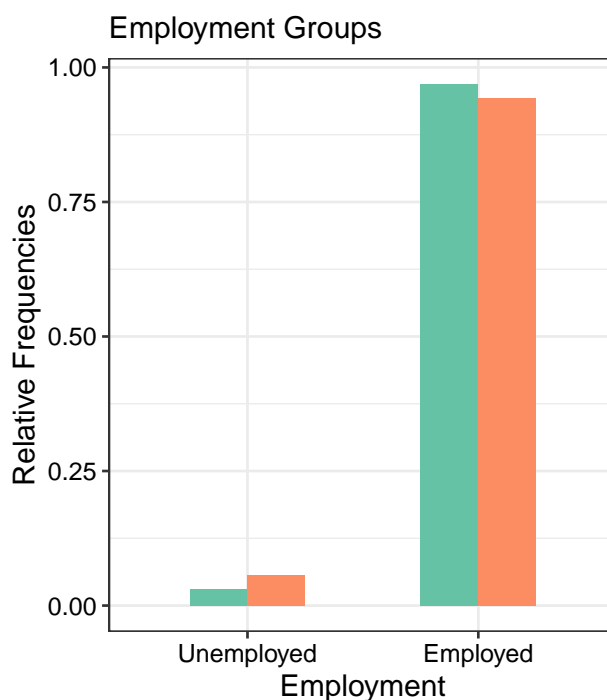
Label Survey Data Census Data



Label Survey Data Census Data



Label Survey Data Census Data



Label Survey Data Census Data

Figure 1: Comparison of Socio-demographics from Two Datasets

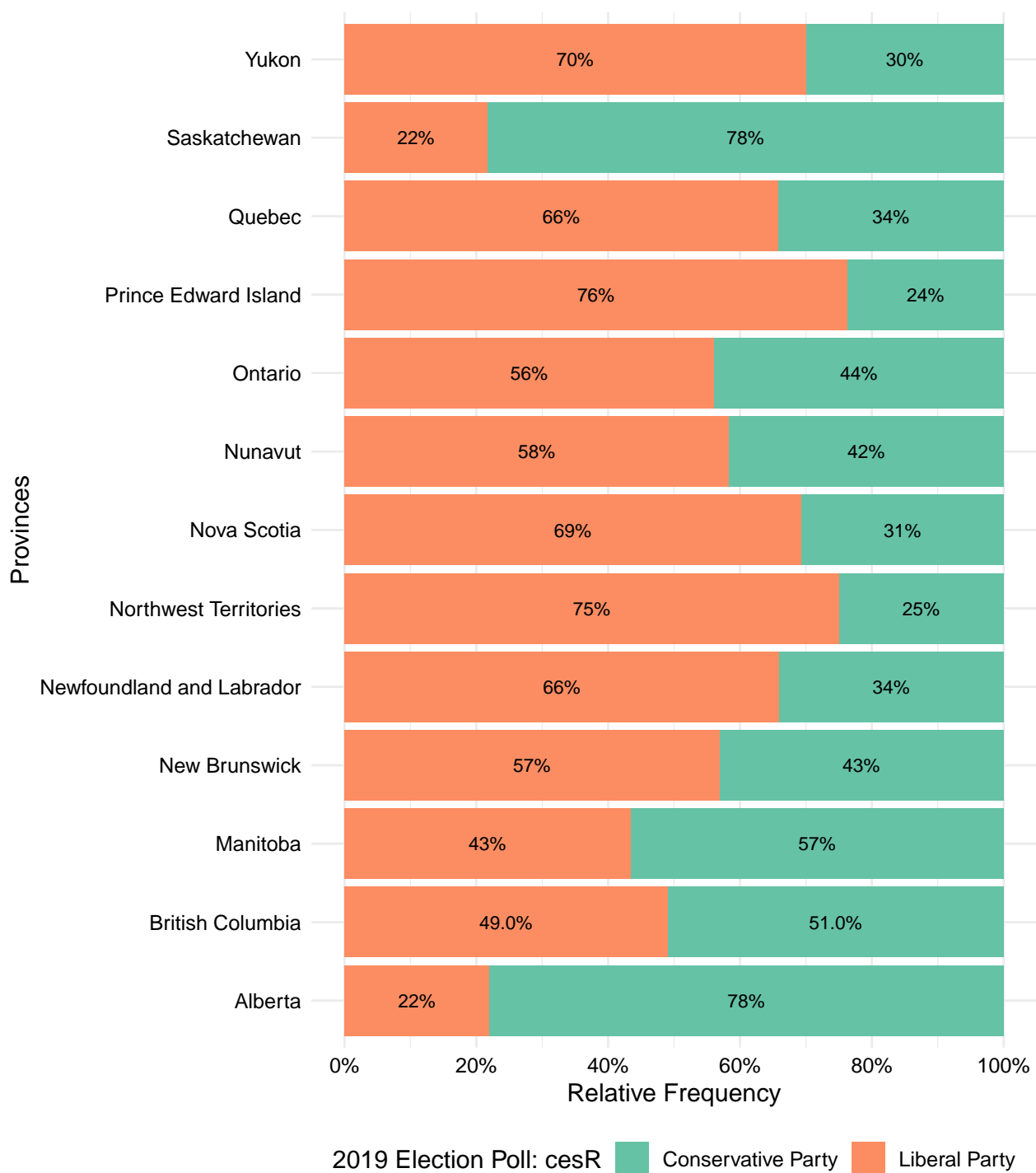


Figure 2: 2019 Election Opinion Poll by Province

With the model selection procedure, the reduced models were fitted and the final model was chosen as:

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

where x_j and b_j represent the variables and corresponding coefficients set of gender, education, income 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, which did not include employment status and age groups.

With the trained model, the post-stratification was carried out with the simulated census data. In other words, post-stratification is to make estimate by averaging over groups in proportion to their size in the population. Through post-stratification, the trained model is able to mimic the prediction as a representative poll. Therefore, the differences between the non-representative dataset and the target population were corrected. Eventually, the probability of voting for the Liberals were predicted at the individual level and aggregated at different subgroup levels or the population level.

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), ggplot2 and R-base were used for data wrangling and visualization (Wickham et al. 2019; R Core Team 2020). R package forcats (version 0.5.0) was also used during the data pre-processing stage (Wickham 2020). There are other packages also used such as captioner, gridExtra, broom, magrittr, gtable, stargazer, RColorBrewer, knitr, labelled and arsenal (Alathea 2015; Hlavac 2018; Heinzen et al. 2020; Xie 2020; Auguie 2017; Robinson, Hayes, and Couch 2020; Bache and Wickham 2014; Larmarange 2020; Neuwirth 2014). Code are available at: <https://github.com/spicapica/stats/tree/main/Final>.

Results

The multilevel logistic regression models fit in this study were summarized in Table 3. The backward model selection was started from the full model (1), then reduced models were compared based on AIC and number of variables. Since coefficients of the binary age variable and the binary employment status variable were not significant in the full model, the reduced models were fit by leaving those variables out. Eventually, the most reduced model (4) was chosen as final model with the smallest AIC and least number of variables, including gender, education, income and province, as shown in Table 3. All these models were trained with the survey data.

Table. 3: Multilevel Logistic Regression Models

Table. 4: Prediction Result: MRP Framework

	Overall (N=1666941)
Vote: MRP Prediction	
Mean (SD)	0.522 (0.159)
Range	0.119 - 0.861
Vote Choice	
Conservative Party	635725 (38.1%)
Liberal Party	1031216 (61.9%)

With the final multilevel logistic regression found in the first step of MRP, the post-stratification estimate of vote choice was made with the simulated census data. The result was aggregated at the country level, as shown in Table 4. Particularly, the group average probability of voting for the Liberals was 0.522, with standard deviation of 0.159. Compared to the vote choice of the initial survey data, which is 50.9% of the Liberal Party, the prediction made by the trained model with the general Canadian population was more

Table 3:

	<i>Dependent variable:</i>			
	Full Model	No Age	Vote	
			No Employment	Neither
	(1)	(2)	(3)	(4)
Age65-	-0.0002 (0.045)		0.001 (0.045)	
GenderWoman	0.346*** (0.038)	0.347*** (0.038)	0.346*** (0.038)	0.346*** (0.038)
EducationHigh School	0.093 (0.106)	0.093 (0.106)	0.094 (0.106)	0.094 (0.106)
EducationPost High School	0.581*** (0.103)	0.581*** (0.103)	0.583*** (0.103)	0.583*** (0.103)
EmploymentEmployed	0.039 (0.112)	0.039 (0.111)		
Income30000-69999	-0.288*** (0.068)	-0.288*** (0.068)	-0.286*** (0.068)	-0.286*** (0.068)
Income70000-99999	-0.290*** (0.071)	-0.290*** (0.071)	-0.287*** (0.071)	-0.287*** (0.071)
Income100000-	-0.354*** (0.069)	-0.354*** (0.069)	-0.352*** (0.069)	-0.352*** (0.068)
Observations	12,549	12,549	12,549	12,549
Log Likelihood	-7,998.919	-7,998.919	-7,998.979	-7,998.980
Akaike Inf. Crit.	16,039.840	16,037.840	16,037.960	16,035.960

Note:

*p<0.1; **p<0.05; ***p<0.01
Province was included but not shown.

in favor of the Liberal Party. If the probability was dichotomized with conservative probability 0.5, where greater than 0.5 means voting for the Liberals, the vote would be even more in favor of the Liberals. The result demonstrated that if all eligible Canadians had voted, the Liberal Party of Canada could win the popular votes by a narrow margin.

Additionally, the vote choice estimates were also averaged over subgroups by education or income levels. As shown in Figure 3, using 0.5 as the threshold, people with relatively low education and more wealth tend to vote for the conservatives, while the others would like to vote for the Liberals.

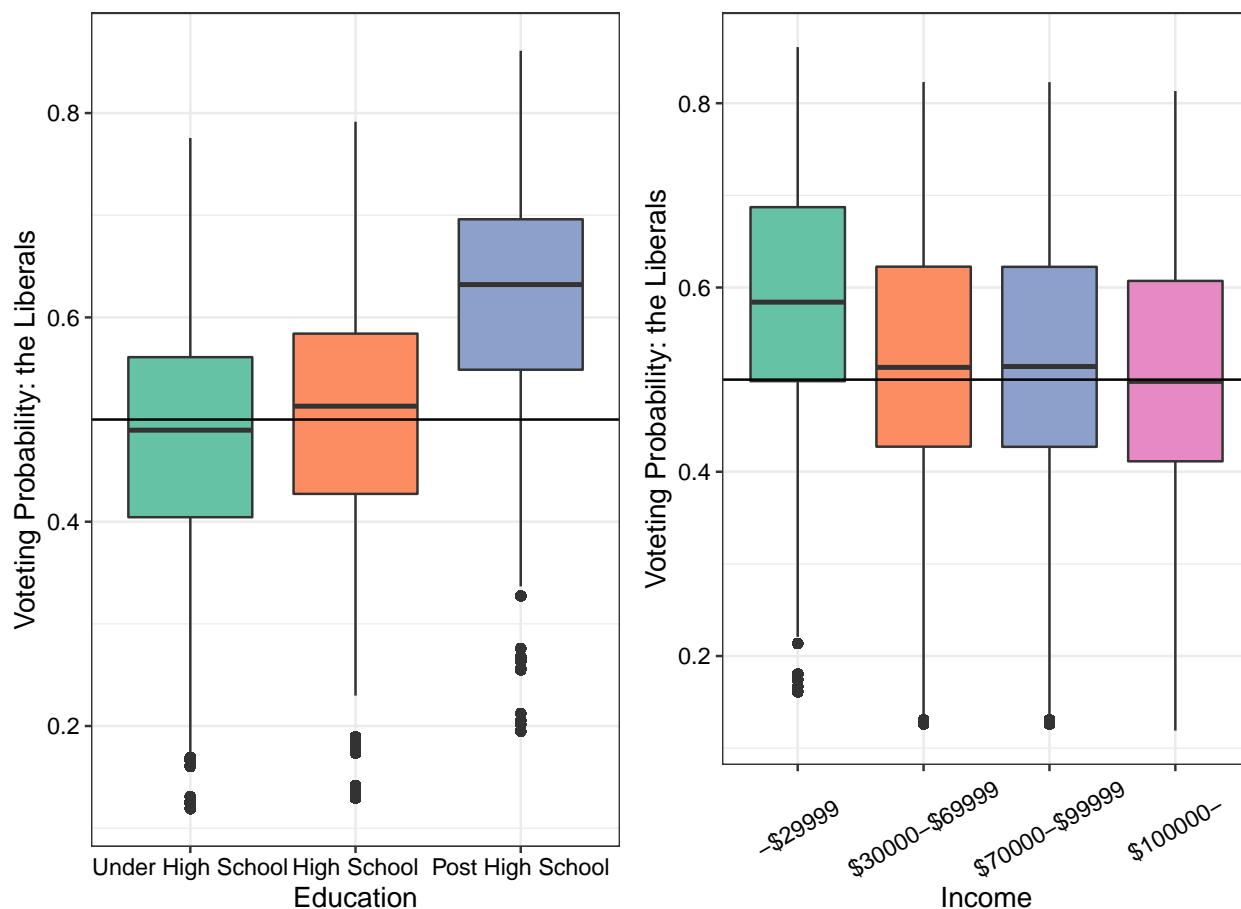


Figure 3: Votes Prediction by Socioeconomic Groups

Discussion

A simple multilevel logistic regression model was chosen as the final model in this study, with only four categorical variables, gender, education, income and province. However, the results generated from this simple model was generally comparable to most election opinion polls in 2019 and the final 2019 Canadian election results. Even though, the competition between the Liberals and the Conservatives was very close and intense in 2019. The results of this study suggested that the Liberals might have won more popular votes, if all eligible Canadians had voted. Particularly, the Liberals might have won 52.2% of all popular votes when compared to the Conservatives. In general, the population average probability of voting for the Liberals predicted by this study lies between the popular votes of 2016, 0.5531 and the popular votes of 2019 0.491, which is a reasonable prediction, because it has been well established that the Liberals had slowly lost support since 2016.

Through this work, the vote choices were estimated for small subgroups which were defined by categorical variables including gender, province, education and income levels. The correlation between the vote preferences and those variables were well documented (Wei Wang and Gelman 2014; Rohan Alexander and Leslie 2019). Incorporating correct and meaningful features in the model made the forecast precise and robust. The detailed information of all models fitted can be found in the **Appendix**.

The voter turnout of the 2019 Canadian federal election was around 66% - 67%, and the voter turnout decreased gradually for the past several federal elections. Low and unequal voter turnout has been described as “democracy’s unresolved dilemma”, which is a serious challenge to the democracy (Fowler 2013). It has been shown that the results from voluntary turnout are significantly different from the results of everyone turns out to vote, because the preferences of marginal voters are often different from those of regular voters (Fowler 2013). Thus the people who voted in 2019 was not a good representation of the entire eligible Canadian population. The biases between the voters and the eligible Canadian population can partially explain the disagreement between vote prediction from this study and real result of 2019. Moreover, this disagreement further proved that changes in voter turnout can alter the election results significantly. Therefore, it is very important to conduct a study to explore and demonstrate the potential election results under the condition that all Canadian citizens had voted in 2019.

The MRP framework is a well developed statistical tool to make vote choice estimates in the context of political topics. Multilevel modeling allows the information and estimates be partially pooled over common characteristics across different subgroups, and the post-stratification with the representative dataset corrects the biases in the election opinion polls. Additionally, the multilevel regression model can provide accurate and robust subgroup-level estimates, while avoiding those issues caused by taking empirical means.

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 derive unbiased estimates of the targeted population based on highly biased survey data, by the introduction of the post-stratification dataset. Typically, the post-stratification dataset such as the census data is a general purpose data and not restricted by certain research topic. Thus, the post-stratification data is always available and the same dataset can be applied to studies focused on different topics of the same population. By using the census data and the MRP framework, this study have the unique power to represent the Canadian citizenry, as voter turnout level of 100%, which usually can not be achieved by opinion polls.

Limitations of this study were mainly introduced by the lack of a proper and matched census dataset, since the MRP requires two matched datasets and datasets are expected to be riched in socioeconomic, geographic and demographic features. It is usually very difficult to find two exactly matched datasets. Moreover, the representative dataset of the targeted population may not exist. In addition, the MRP requires a very good representative dataset of the population, such as the census data, to maximize the prediction power of this tool.

In terms of this work, there was no census data of the general Canadian population in 2019. Therefore, in this work, it was assumed that the 2016 population is the same as the 2019 population, which may not be the case. This would reduce the power of this study as well. In terms of the mismatch, the gender variable in the survey dataset has three levels, including man, woman and the others, while the census data only treated it as a binary feature. To match the datasets, this study had to use a binary gender. However, a binary gender variable is a biased way in representing the population, since the diversity in gender is an important aspect of the population (Kennedy et al. 2020). Similarly, the age group was dichotomized in this work, because the age was grouped as 15-64 and 65- in the 2016 census data. This definitely leads to the loss of power in the variable. Thus the final model did not include age. Another mismatch is the income variable, while the household level income was collected in the survey data, the individual income was collected in the census data.

Another potential limitation of the MRP model is that only categorical variables such as socio-economic, demographic, and geographic features are suitable to build subgroups. Some important variables such as the satisfaction of the current economy was not included. These types of variables are typically not available in census data.

In the future work, a separated model can be developed to include variables mentioned in the Keys model, which contains continuous variables (Lichtman 2020). An popular idea is that the accuracy of the prediction can be improved through stacking different models together, where each model is in charge of different aspects of information, such as different sets of uncorrelated variables. Additionally, the MRP framework itself is able to take multiple survey data with different weights in the model building stage (Ghitza and Gelman 2013). As approaching the election day, the polls are more accurate compared to early polls. Thus, in the future work, multiple opinion polls should be included.

Many limitations of this study can be fixed if the individual level census data is available, such as the age would be categorized into at least five subgroups instead of only two subgroups. With the simulated dataset, it is not appropriate to use complex hierarchical logistic regression, while incorporating deeper interactions between variables. If a real census data was adopted in this study, subgroup level vote choice estimates, such as at province level, would definitely be more accurate and robust. Moreover, the simulated dataset does not have enough detailed subgroup information, such as the geographic location at the riding level. There are 338 ridings spread in the Canada, each one of the ten provinces has multiple ridings, and each one of the three territories is a riding itself. Each riding will decide a seat in the House of Commons. The party with the most seats in the House of Commons will win the election and form the government. Thus, without the riding level vote choice estimation, this study was limited in the resolution of the results. Particularly, this study is only able to make prediction of the popular votes but not able to predict which party can win the election. In addition, the Canadian federal election has more than two parties and the scenario was simplified in this work, which can be improved in the future.

Appendix

```
## # A tibble: 21 x 5
##   term                estimate std.error statistic  p.value
##   <chr>              <dbl>      <dbl>      <dbl>    <dbl>
## 1 (Intercept)        -1.64        0.156    -10.5    6.40e-26
## 2 Age65-            -0.000211    0.0447  -0.00472 9.96e- 1
## 3 GenderWoman         0.346        0.0384    9.01    1.99e-19
## 4 ProvinceBritish Columbia 1.26        0.0798   15.8    2.21e-56
## 5 ProvinceManitoba     1.01        0.101   10.0    9.68e-24
## 6 ProvinceNew Brunswick  1.51        0.133   11.4    4.70e-30
## 7 ProvinceNewfoundland and Labrador 1.92        0.154   12.5    1.30e-35
## 8 ProvinceNorthwest Territories 2.41        0.821    2.93    3.37e- 3
## 9 ProvinceNova Scotia   2.08        0.132   15.7    8.48e-56
## 10 ProvinceNunavut      1.60        0.591    2.71    6.71e- 3
## # ... with 11 more rows

## # A tibble: 20 x 5
##   term                estimate std.error statistic  p.value
##   <chr>              <dbl>      <dbl>      <dbl>    <dbl>
## 1 (Intercept)        -1.64        0.156    -10.5    5.73e- 26
## 2 GenderWoman         0.347        0.0380    9.11    8.26e- 20
## 3 ProvinceBritish Columbia 1.26        0.0797   15.8    1.69e- 56
## 4 ProvinceManitoba     1.01        0.101   10.0    9.68e- 24
## 5 ProvinceNew Brunswick  1.51        0.133   11.4    4.64e- 30
## 6 ProvinceNewfoundland and Labrador 1.92        0.154   12.5    1.29e- 35
## 7 ProvinceNorthwest Territories 2.41        0.821    2.93    3.37e- 3
## 8 ProvinceNova Scotia   2.08        0.132   15.7    8.39e- 56
## 9 ProvinceNunavut      1.60        0.591    2.71    6.71e- 3
## 10 ProvinceOntario     1.51        0.0642   23.4    1.78e-121
## 11 ProvincePrince Edward Island 2.50        0.390    6.42    1.39e- 10
## 12 ProvinceQuebec      1.95        0.0733   26.6    2.72e-156
## 13 ProvinceSaskatchewan -0.0426     0.128   -0.334    7.38e- 1
## 14 ProvinceYukon       1.97        0.695    2.84    4.52e- 3
## 15 EducationHigh School  0.0927     0.106    0.874    3.82e- 1
## 16 EducationPost High School 0.581      0.103    5.64    1.67e- 8
## 17 EmploymentEmployed    0.0387     0.111    0.348    7.28e- 1
## 18 Income$30000-$69999 -0.288      0.0680   -4.24    2.21e- 5
## 19 Income$70000-$99999 -0.290      0.0711   -4.07    4.65e- 5
## 20 Income$100000-      -0.354      0.0688   -5.15    2.62e- 7

## # A tibble: 20 x 5
##   term                estimate std.error statistic  p.value
##   <chr>              <dbl>      <dbl>      <dbl>    <dbl>
## 1 (Intercept)        -1.61        0.123   -13.0    8.23e- 39
## 2 Age65-             0.00124     0.0445    0.0279  9.78e- 1
## 3 GenderWoman         0.346        0.0384    9.01    2.00e- 19
## 4 ProvinceBritish Columbia 1.26        0.0797   15.8    1.77e- 56
## 5 ProvinceManitoba     1.01        0.101   10.1    9.00e- 24
## 6 ProvinceNew Brunswick  1.51        0.133   11.4    4.48e- 30
## 7 ProvinceNewfoundland and Labrador 1.92        0.154   12.5    1.26e- 35
## 8 ProvinceNorthwest Territories 2.41        0.821    2.93    3.34e- 3
## 9 ProvinceNova Scotia   2.08        0.132   15.7    7.15e- 56
## 10 ProvinceNunavut      1.60        0.591    2.71    6.65e- 3
## 11 ProvinceOntario     1.51        0.0642   23.5    1.05e-121
```

```
## 12 ProvincePrince Edward Island      2.50      0.389      6.42 1.35e- 10
## 13 ProvinceQuebec                    1.95      0.0733     26.7 1.44e-156
## 14 ProvinceSaskatchewan              -0.0424     0.128     -0.332 7.40e- 1
## 15 ProvinceYukon                    1.98      0.696      2.84 4.50e- 3
## 16 EducationHigh School              0.0940     0.106      0.887 3.75e- 1
## 17 EducationPost High School         0.583      0.103      5.66 1.49e- 8
## 18 Income$30000-$69999              -0.286     0.0678     -4.23 2.35e- 5
## 19 Income$70000-$99999              -0.287     0.0708     -4.06 5.01e- 5
## 20 Income$100000-                   -0.352     0.0687     -5.12 3.05e- 7

## # A tibble: 19 x 5
##   term                                estimate std.error statistic  p.value
##   <chr>                                <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)                       -1.61      0.122     -13.1 2.07e- 39
## 2 GenderWoman                        0.346      0.0380      9.11 8.59e- 20
## 3 ProvinceBritish Columbia          1.26      0.0796     15.9 1.28e- 56
## 4 ProvinceManitoba                  1.01      0.101     10.1 8.99e- 24
## 5 ProvinceNew Brunswick             1.51      0.133     11.4 4.45e- 30
## 6 ProvinceNewfoundland and Labrador 1.92      0.154     12.5 1.26e- 35
## 7 ProvinceNorthwest Territories      2.41      0.821      2.93 3.34e- 3
## 8 ProvinceNova Scotia               2.08      0.132     15.7 6.99e- 56
## 9 ProvinceNunavut                   1.60      0.591      2.71 6.66e- 3
## 10 ProvinceOntario                  1.51      0.0642     23.5 9.90e-122
## 11 ProvincePrince Edward Island      2.50      0.389      6.42 1.35e- 10
## 12 ProvinceQuebec                    1.95      0.0733     26.7 1.41e-156
## 13 ProvinceSaskatchewan              -0.0424     0.128     -0.332 7.40e- 1
## 14 ProvinceYukon                    1.98      0.695      2.84 4.49e- 3
## 15 Income$30000-$69999              -0.286     0.0677     -4.23 2.35e- 5
## 16 Income$70000-$99999              -0.287     0.0708     -4.06 4.95e- 5
## 17 Income$100000-                   -0.352     0.0684     -5.14 2.74e- 7
## 18 EducationHigh School              0.0940     0.106      0.887 3.75e- 1
## 19 EducationPost High School         0.583      0.103      5.66 1.49e- 8
```

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*.
- 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>.
- Lichtman, Allan. 2020. "The Keys to the White House: Forecast for 2020." <https://doi.org/10.1162/99608f92.baaa8f68>.
- McElreath, Richard. 2016. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press/Taylor & 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. 2020. *Forcats: Tools for Working with Categorical Variables (Factors)*. <https://CRAN.R-project.org/package=forcats>.

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>.

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