How the 2019 Canadian Federal Election would have been different if 'everyone' had voted*

Cao Boyu

22 December 2020

Abstract

In this paper, we develop two models with the purpose of forecasting how the 2019 Canadian Federal Election would have been different if 'everyone' had voted. Multilevel regression with postratification (MRP) and Stacked regression with postratification models where constructed using the 2019 Canadian Election Study survey and postsratified using the 2017 General Social Survey. We concluded that if "everyone" votes, there is a high probability that there is no obvious difference.

Keywords: Forecasting; 2019 Canadian Federal Election; Multilevel Regression with Poststratification; No different

1 Introduction

The 2019 Canadian federal election (formally the 43rd Canadian general election) was held on October 21, 2019, to elect members of the House of Commons to the 43rd Canadian Parliament. The Liberal Party, led by incumbent Prime Minister Justin Trudeau, won 157 seats to form a minority government and lost the majority they had won in the 2015 election. The Liberals lost the popular vote to the Conservatives, which marks only the second time in Canadian history that a governing party formed a government while receiving less than 35 per cent of the national popular vote.The Conservative Party, led by Andrew Scheer, won 121 seats and remained the Official Opposition. The Bloc Québécois, under Yves-François Blanchet, won 32 seats to regain official party status and became the third party for the first time since 2008. The New Democratic Party, led by Jagmeet Singh, won 24 seats, its worst result since 2004. The Green Party, led by Elizabeth May, saw its best election results with three seats and for the first time received over one million votes.

However, the gap between the Conservative Party and the Liberal Party is not very large. When voters vote, they may be affected by other conditions that may cause some voters to absent or abstain. The final result may still be somewhat different from the real public

*Code and data are available at: https://github.com/Xeon0312/final_report/

opinion. This time, we use multilevel regression with postratification (MRP) and Stacked regression with postratification models to simulate the answer if 'everyone' had voted. For the data part, we are going to use 2019 Canadian Election Study survey and postsratified using the 2016 Canadian Education Highlight census dataset. Details will be in the Model section.

After a series of simulation experiments, we found that although the proportion of votes among the parties has changed, the overall result is the same as before. The Liberal Party is still elected by a narrow margin. This reflects that the public opinion of Canadian voters has been truly reflected and has not been affected by other factors.

2 Data

2019 CES

The 2019 Canadian Election Study was conducted to gather the attitudes and opinions of Canadians during and after the 2019 federal election. It continues the tradition of Canadian Election Studies started in 1965. There are 2 data files for the 2019 CES - online survey and phone survey. We only used online survey this time. The primary mandate of the Canadian Election Study is to provide a thorough account of the election, to underline the main reasons why people vote the way they do, to indicate what does and does not change during the campaign and from one election to another, and to highlight similarities Campaign period survey start from 2019-09-13 to 2019-10-21. The population is all people did the web sruvey during the period.

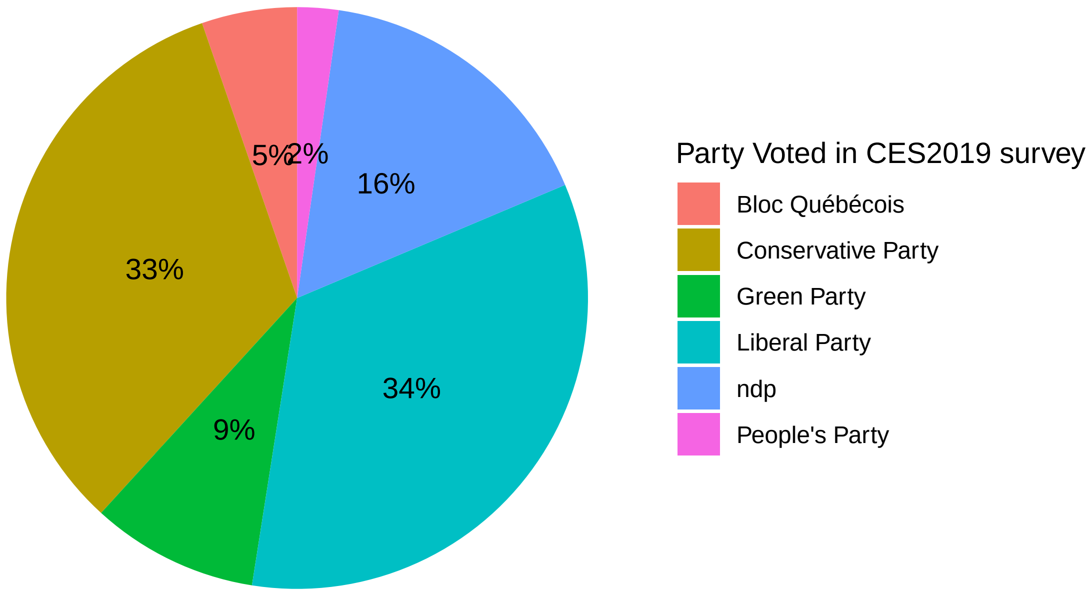

Figure 1: Party Voted in CES2019 survey

As can be seen in Figure 1, The Liberal Party and the Conservative Party are the two most mainstream parties with support ratings of 34% and 33% respectively. NDP and Green Party ranked third and fourth with 16% and 9% respectively

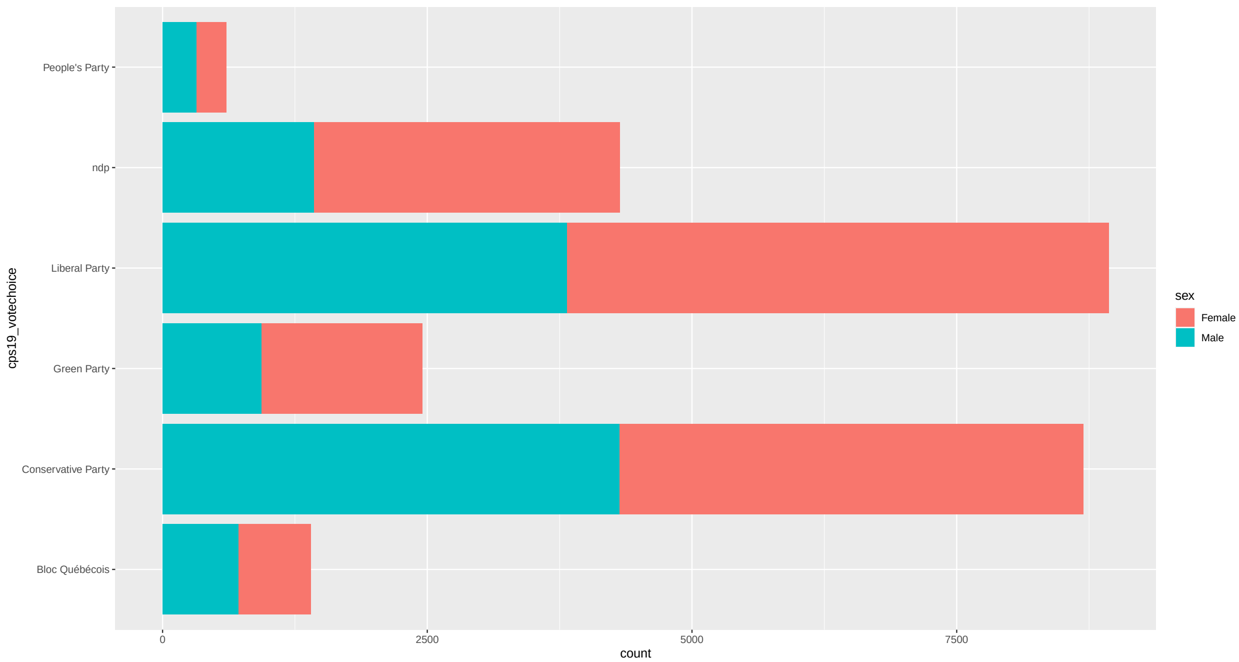

Figure 2: How Respondents Plan to Vote in 2019 By Gender

It can be seen from the Figure 2 that the proportion of men and women in the voting population is similar, with slightly more women. Women accounted for the majority of votes obtained by NDP.

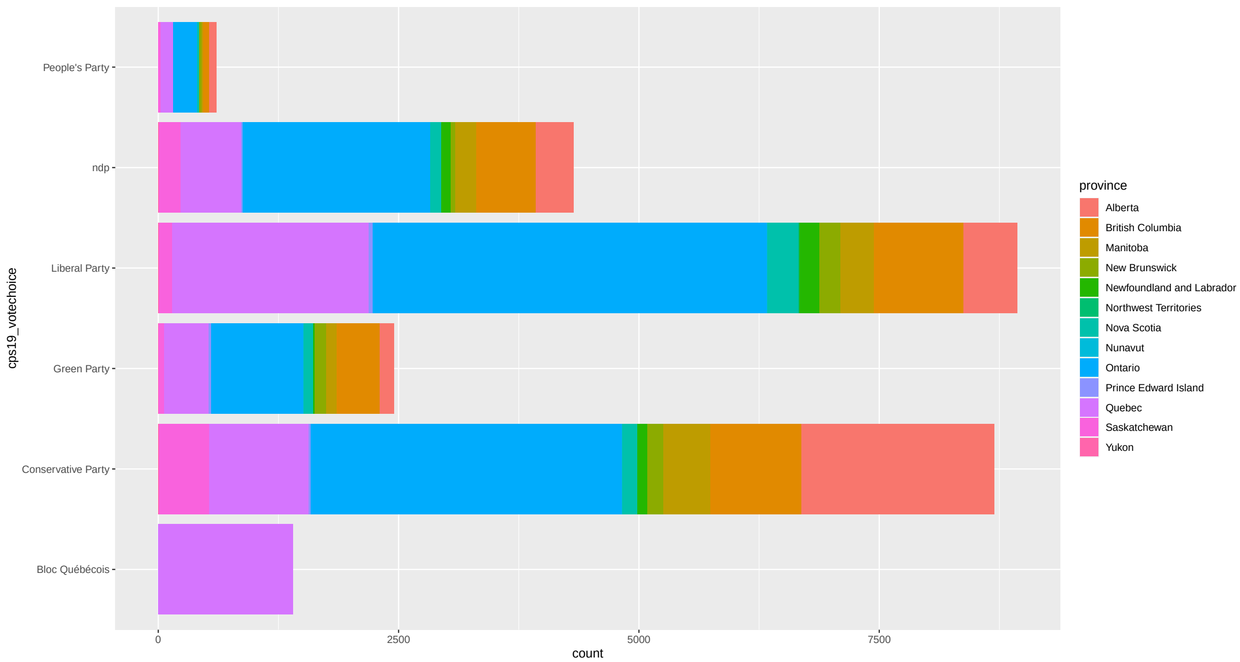

Figure 3: How Respondents Plan to Vote in 2019 By Province

It can be concluded from the Figure 3 that the proportion of people in Ontario who voted is higher. Interestingly, the votes of the Bloc Quebecois party only come from Quebec.

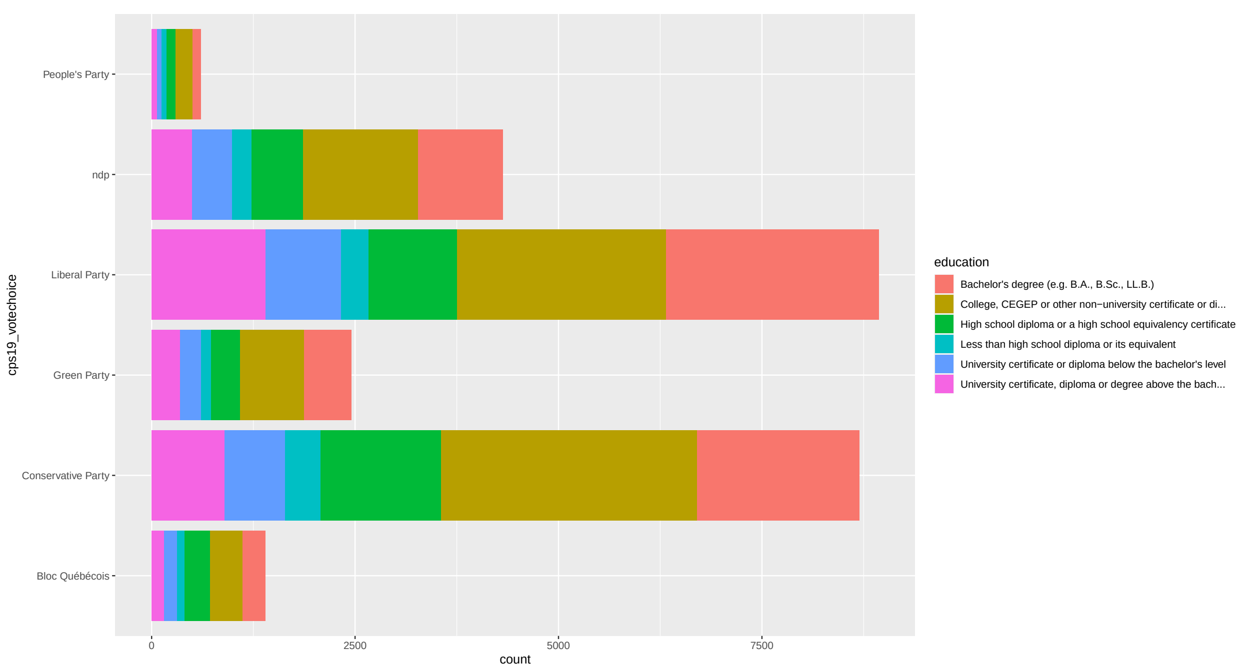

Figure 4: How Respondents Plan to Vote in 2019 By education

The data obtained from the Figure 4 shows that most of the voting population has higher education than high school. Those who voted for the Conservative Party had an average education level higher than those who voted for the Liberal Party.

GSS Data

The dataset we used for our analysis came from responses to the “2017 General Social Survey (GSS): Families Cycle 31”. The 2017 GSS, conducted from February 2nd to November 30th, 2017, is a sample survey with a cross-sectional design. The target population includes all non-institutionalized persons 15 years of age and older, living in the 10 provinces of Canada. The survey uses a new frame, created in 2013, that combines telephone numbers (landline and cellular) with Statistics Canada’s Address Register, and collects data via telephone. Data are subject to both sampling and non-sampling errors. Each record in the survey frame was assigned to a stratum within its province. A simple random sample without replacement of records was next performed in each stratum. The target sample size (i.e. the desired number of respondents) for the 2017 GSS was 20,000 while the actual number of respondents was 20,602. The survey frame was created using two different components: 1. Lists of telephone numbers in use (both landline and cellular) available to Statistics Canada from various sources (telephone companies, Census of population, etc.); 2. The Address Register (AR): List of all dwellings within the ten provinces. The overall response rate for the 2017 GSS was 52.4%.

The target population for the 2017 GSS included all persons 15 years of age and older in Canada, excluding: 1. Residents of the Yukon, Northwest Territories, and Nunavut; and 2. Full-time residents of institutions. That will cause some data missing.

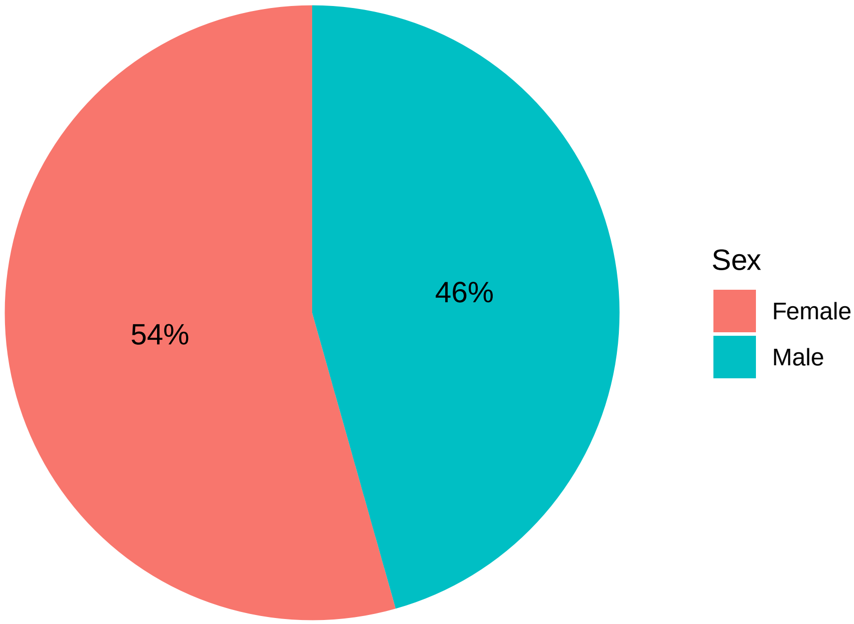

Figure 5: GSS By Sex

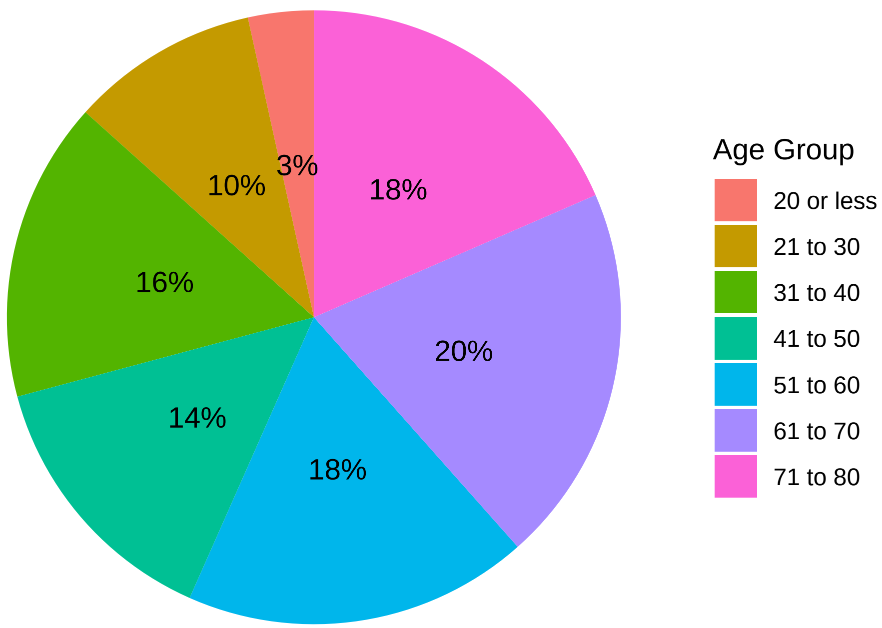

Figure 6: GSS By Age group

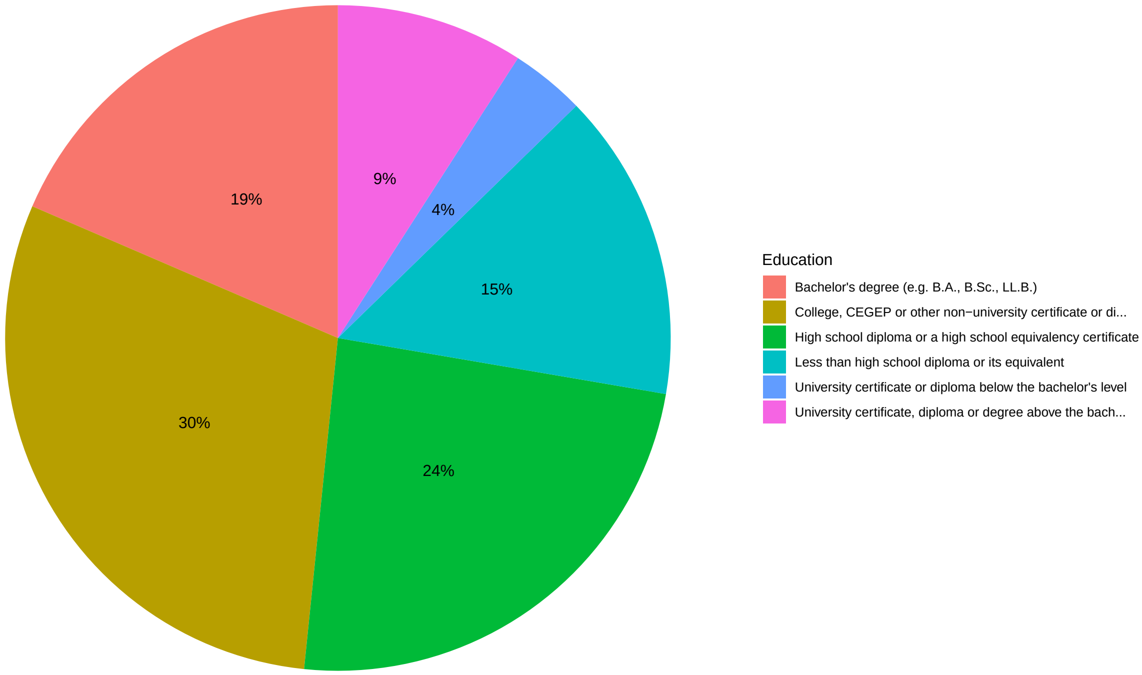

Figure 7: GSS By Education

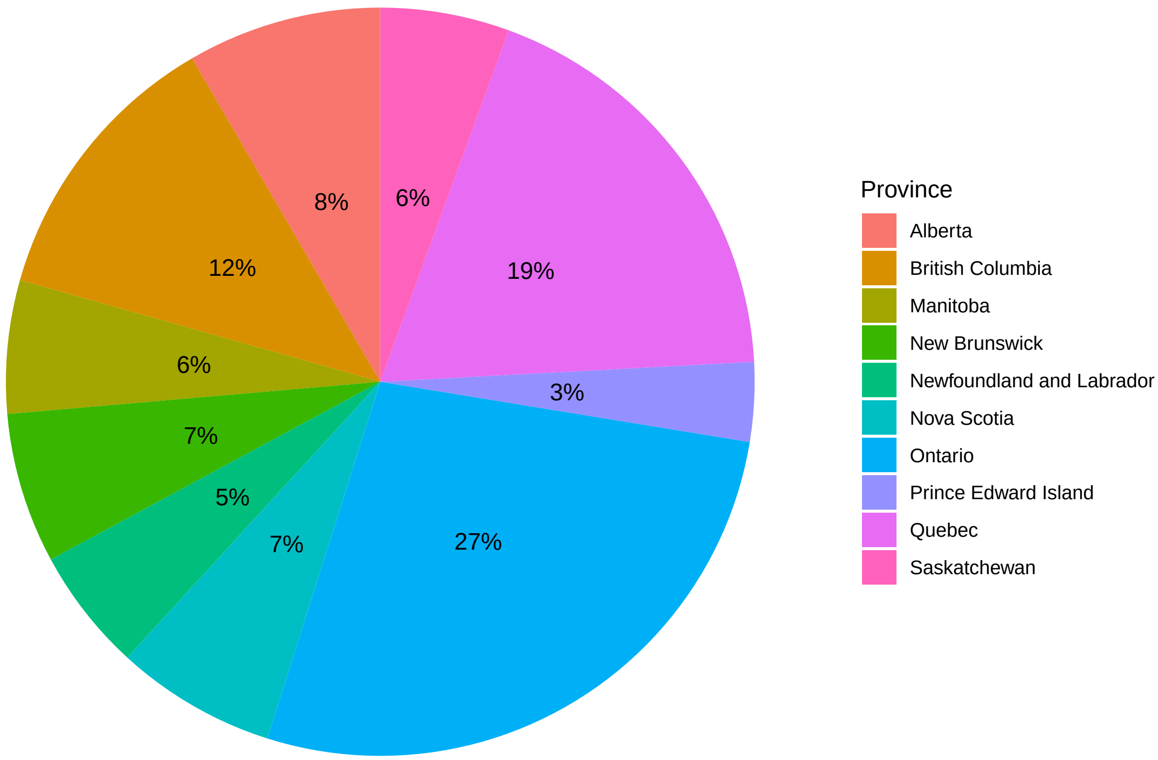

Figure 8: GSS By Province

3 Model

A Bayesian multilevel logistic model with post-stratification was used to predict the plurality for each state in the 2019 Canadian Federal Election. We used a logistic model as they can associate binary and continuous independent variables with a binary categorical response variable. Unlike the previous assignment, the choice was made between two parties. However, what we have to do in this homework is to make a distribution forecast among the six parties. So this time I made 6 models to predict whether they will vote for the party. Since the decision of whether they will vote for the party given is condensed to a binary response variable, the logistic model is well suited for our goal of predicting the winner. Linear models were unsuited for our goals as they require the response variable to be a continuous variable such as height. Our response variable was either 0 (representing not to vote for this party) or 1(representing a vote for this party), which is binary in nature. Through the logical model, we can express the probability "p" that an individual will vote for the selected party as a logarithmic function of various dependent variables.

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k$$

In our case, we will consider the following variables for the model: gender, age, education level, and province. We use them because they can match in our two data. We decided to make the model multi-layered in order to group voters by province. Since voting preferences and demographic groups vary greatly in each province, we think it is appropriate to analyze multiple states separately. The most important thing is that there are three provinces in our GSS data that are missing, and we need to predict separately by province.

$$P(\beta_0, \beta_1, \ldots, \beta_k | y, X_1, ..., X_k) \propto P(y, X_1, ..., X_k | \beta_0, \beta_1, \ldots, \beta_k) * P(\beta_0, \beta_1, \ldots, \beta_k)$$

We used "brm" to fit Bayesian generalized multivariate multilevel models. Then we used "add_predicted_draws" to add draws from the posterior fit. We performed 4000 simulations for each person in the GSS according to the model of the corresponding party. In each simulation, 0 or 1 was obtained, which respectively represented "will vote for the corresponding party" and "will not vote for the corresponding party". Average the simulation results to get the probability that this person will vote for the corresponding party.

$$E(\text{votes}) \approx \frac{\sum \text{draws} * \text{vote}}{\text{Number of draw}}.$$

4 Results

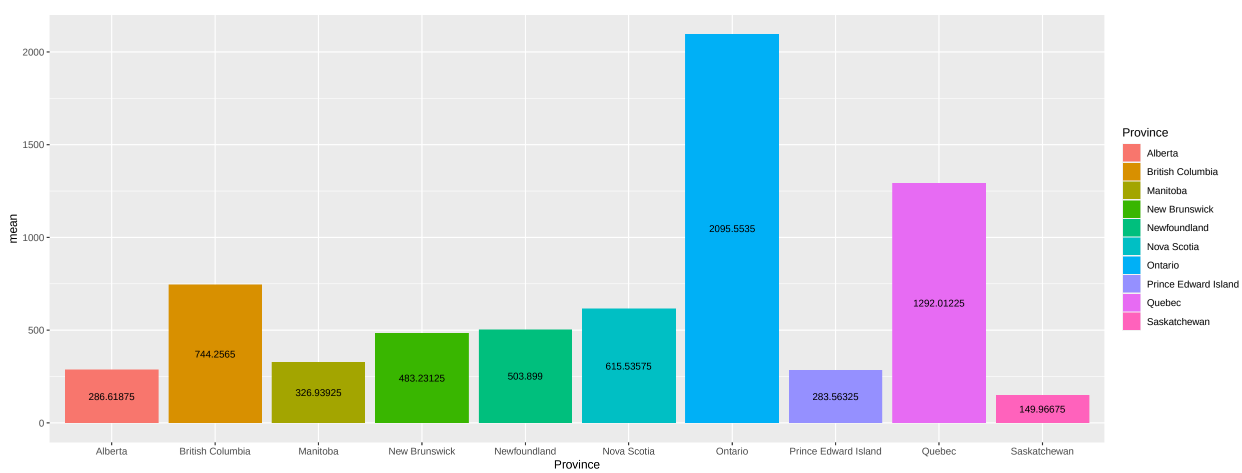

Figure 9: Result of Liberal Patry

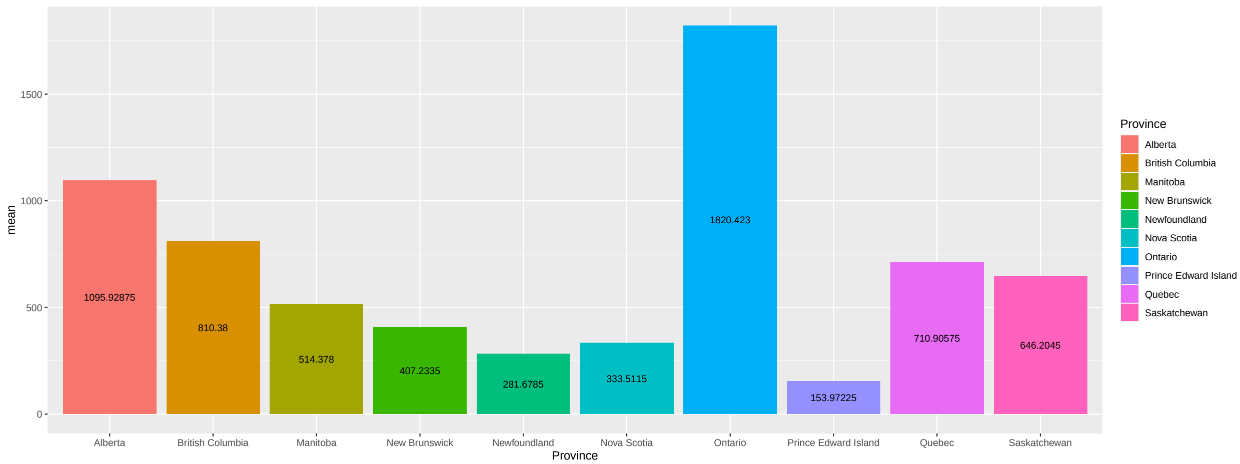

Figure 10: Result of Conservative Patry

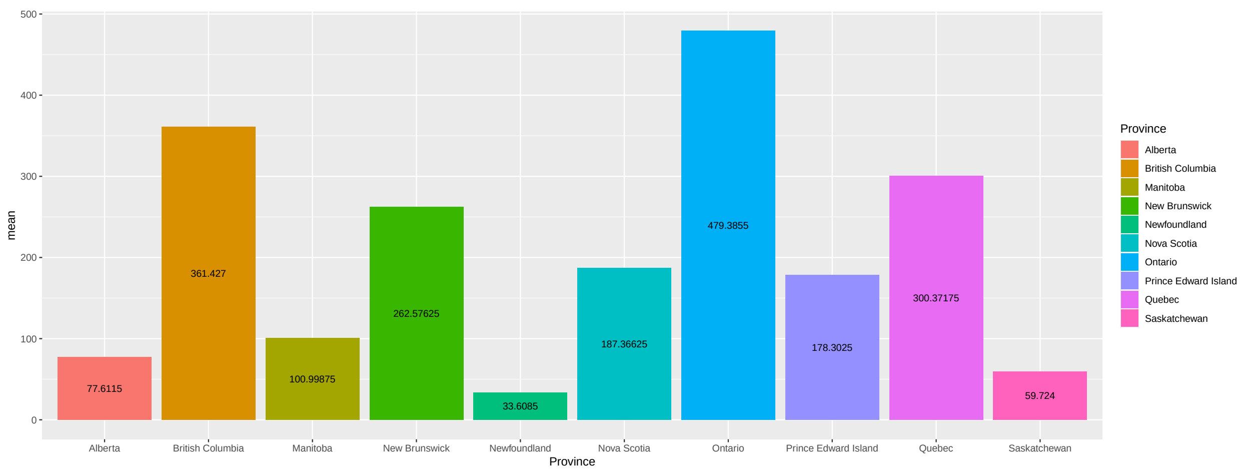

Figure 11: Result of Green Patry

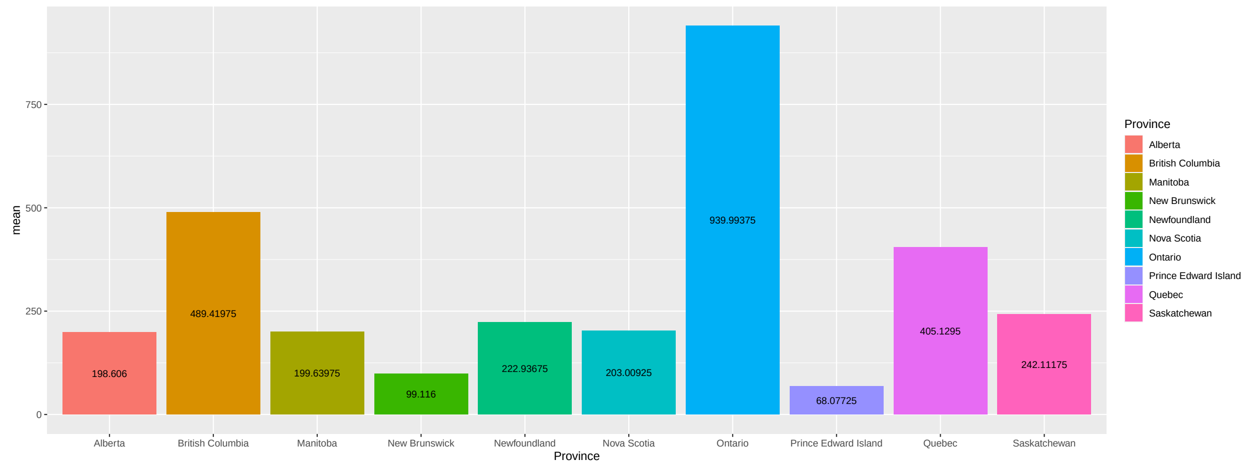

Figure 12: Result of ndp Patry

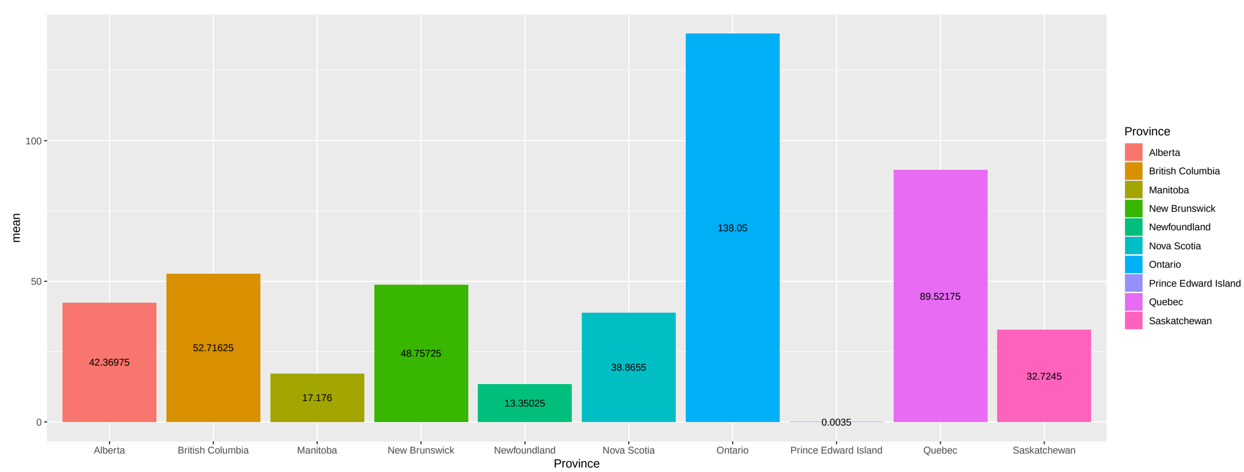

Figure 13: Result of People's Patry

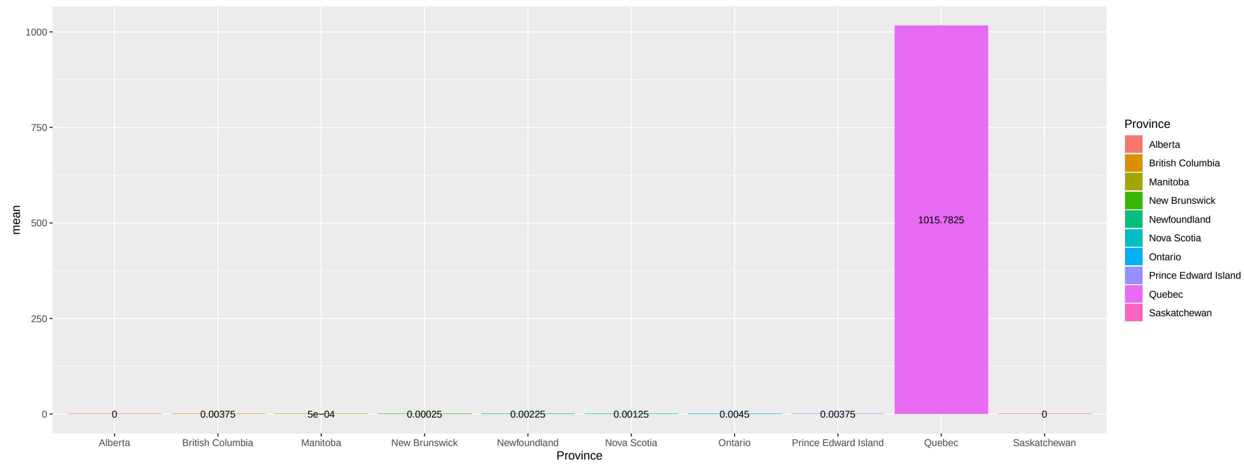

Figure 14: Result of Bloc Québécois

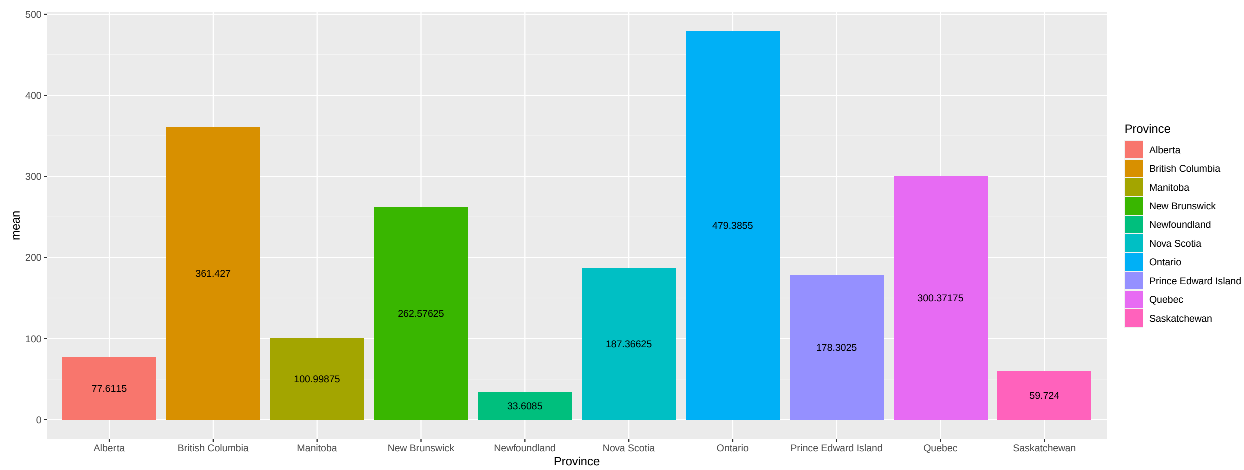

Figure 15: Result of Green Patry

As can be seen from Figure 9 to Figure 15.

Alberta voted 1095.92 for the Conservative Party, far more than the number of votes Alberta voted for other parties. Voting in British Columbia is more scattered, with the most people voting for the Conservative Party with 810.38 votes, followed by the Liberal Party with 744.26 votes. Manitoba and Saskatchewan have the highest approval ratings for the Conservatives, which won 514.38 and 333.51 respectively.

New Brunswick, Newfoundland, Nova Scotia, Ontario, and Prince Edward Island have the most support for the Liberal Party. Among them, Ontario contributed 2095.5535 votes to the Liberal Party, Nova Scotia was 615.53575, Newfoundland was 503.899, New Brunswick was 483.23125, and Prince Edward Island voted 283.5632.

It is worth mentioning that the Quebec Party, all of their votes are from Quebec, which is 1015.7825. However, the Liberal Party received 1,292.01225 votes, making it the most supported party in Quebec by a narrow margin.

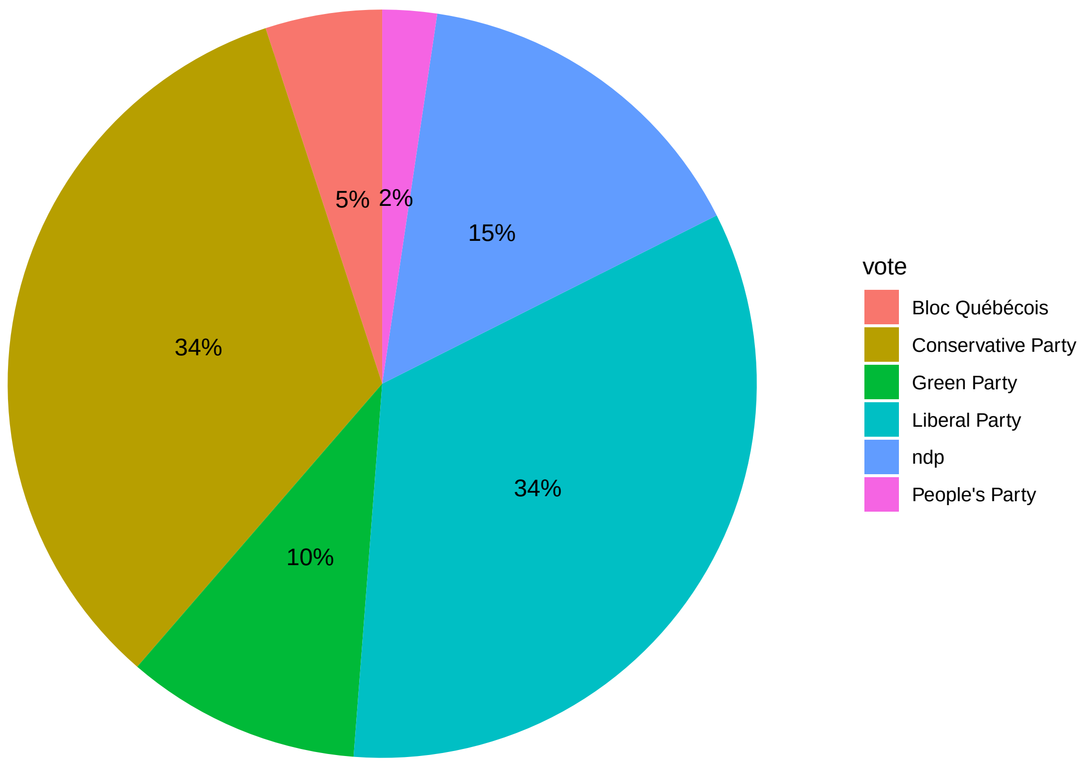

Figure 16: Over all predict

5 Discussion

Appendix

References

Bürkner, P (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software, 80(1), 1-28. doi:10.18637/jss.v080.i01

Bürkner, P (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. The R Journal, 10(1), 395-411. doi:10.32614/RJ-2018-017

Gelman, A., & Kennedy, L. (2020). "Know your population and know your model: Using model-based regression and post-stratification to generalize findings beyond the observed sample." Retrieved from: https://arxiv.org/pdf/1906.11323.pdf

JJ Allaire et al. (2020). rmarkdown: Dynamic Documents for R. R package version 2.3. Retrieved from: https://rmarkdown.rstudio.com.

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from: https://www.R-project.org/.

Robinson, D., Hayes, A., & Couch, S. (2020). broom: Convert Statistical Objects into Tidy Tibbles. R Package version 0.7.0 Retrieved from: https://CRAN.R-project.org/package=broom.

Rudis, B. (2020). statebins: Create United States Uniform Cartogram Heatmaps. R package version 1.4.0. Retrieved from: https://CRAN.R-project.org/package=statebins.

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

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag. New York.

Wickham, H., & Miller, E. (2020). haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files. R package version 2.3.1. Retrieved from: https://CRAN.R-project.org/package=haven.

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

Wu, C., & Thompson, M. (2020). Sampling Theory and Practice, Springer.