FORECASTING THE 2019 CANADIAN FEDERAL ELECTION IF EVERYONE HAD VOTED

Jiayi YANG

December 22, 2020

Abstract

In this paper, we aim to forecast the result 2019 Canadian Federal election if everyone participated in the vote.In order to achieve the goal, we firstly develop a logistic regression model using the data given by Canadian Election Survey(CES). To improve the precision on prediciton, we then use Multilevel regression with postratification (MRP) using the observations from Stat Canada 2016 Education Census. The result suggests that some social-demographic variables have large impacts on individual choicses in the Canadian Federal Election.

Keywords: Public Opinion; 2019 Canadian Federal Election; Bayesian Approach; Multilevel Regression with Poststratification

1 Introduction

The 2019 Canadian Federal Election was hold on October 21, 2019. It was in order to elect members of the House of Commons to the 43rd Canadian Parliament. As a result, the leader of the Liberal Party, Justin Trudeau, won 184 seats finally and continued to be the Prime Minister of Canada ("2019 Canadian Federal Election" 2020). Although Justin Trudeau received 39.4% of the last election, it is worth noting that in the previous national Popular vote he only received less than 35% ("2019 Canadian Federal Election" 2020). The Liberals have the lowest national support among all the parties. So, we wondered if all the Canadian citizens with age over 18 could vote in the last election, how would the results of the election change. In this study, we attempt to determine the possibility that Justin Trudeau would still be elected when the eligible voters were expanded to be all the Canadian citizens who meet the requirements.

To accomplish this goal, we do analysis based on the data from Canadian Election Survey (CES) and Stat Canada 2016 Education Census. We will firstly build a logistic regression by Bayesian approach. And then, we will also utilize the multilevel regression model with post-stratification. We analyzed individual-level survey data and post-stratified census data to analyze the effects caused by key demographic variables on the voting decision and predict the probability of Justin Trudeau still winning the election. Our prediction suggests that,

as the voter has been expanded to all the citizen, the probability that Trudeau will still be elected is x%.

In section two, we present the datasets we analysis. It includes the source and the summary of these two datasets, we also interpret the meaning of them. In section three, we show the models we built, and use AIC method choose the best prediction model out of them. Followed by this, in section four, we describe and conclude the result we find by the model. Finally, in section five, we give a conclusion of our study and do discussion on the weakness of our study and consider what else can we do to improve this study.

2 Data

2.1 CES Survey Data
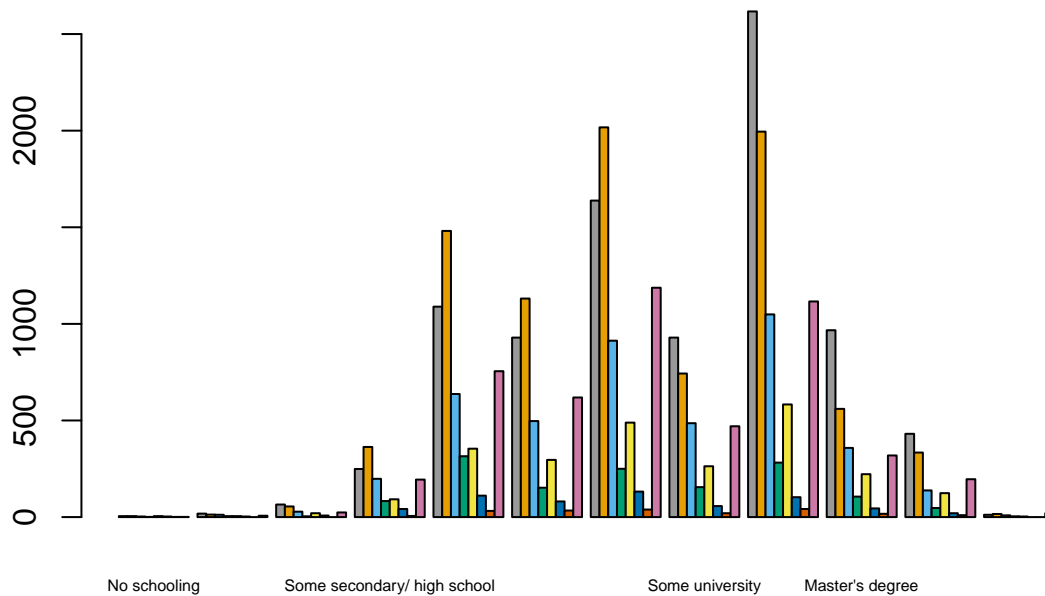
Figure 1. The Preference of Voting in 2019 By Gender
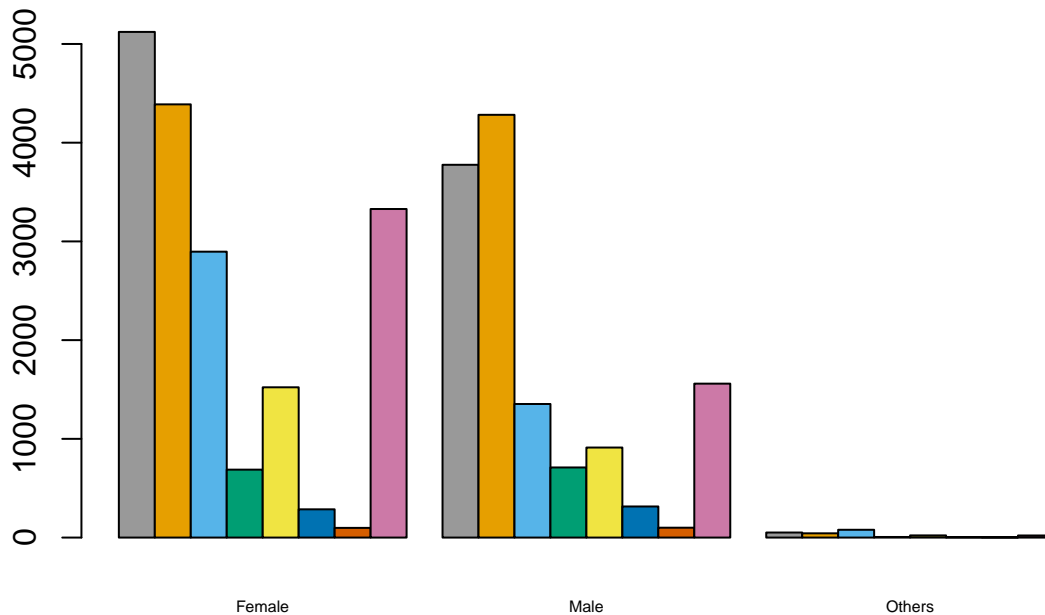
Gender

The trend shows that the CES data demonstrates that female more likely tend to vote for the Liberal Party. And the majority of male and female voters would probability planned to vote the Green party not the Liberal Party.
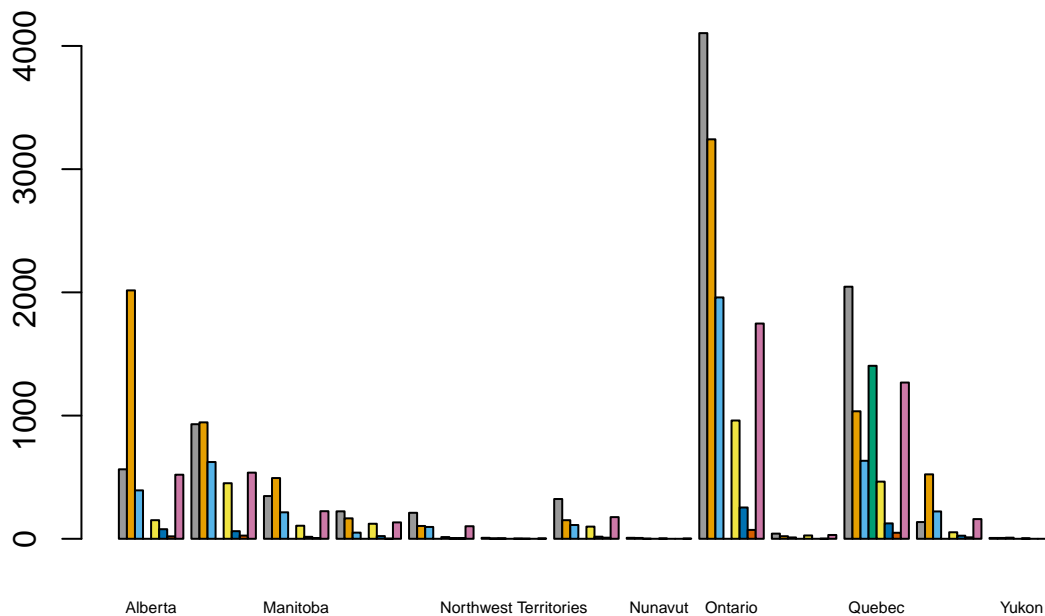
Figure 2. The Preference of Voting in 2019 By Education

Education   For educa-
tion level of respondent, it did not show a relationship with voting intentions for the 2019
Canadian Federal election. Across all education levels, the Liberal party do not have the
most number of voting.

Figure 3. The Preference of Voting in 2019 By Province

Province   Similar to ed-
ucation, there is no apparent pattern. The Green Party holds the most number of votes. If
all Canadian citizens could attend to vote, the Liberal party and Justin Trudeau will not win

any more.

2.2 GSS Census Data

As we want to correctly estimate the election result when all the Canadian citizens could participate to vote, we used post-stratification via adjusting the weights between groups by using an additional dataset. The census data we used was the 2017 General Social Survey (GSS) on the family. It is a nationwide survey which was designed to social-demographic variables of Canadians and capture social-social demographic and social-economic trends.Its target population includes all non-institutionalized person with age over 15, living in the 10 chosen provinces inside Canada. So, it could be a reliable census data which support to be a study population. However, the sample size for this database is really small which only has around 4894 observations.

3 Model Development

3.1 Logistic regression model using Bayesian approach

ddd

3.2 Multilevel regression with post-stratification (MRP)

4 Results

In conclusion, the result we find is .

5 Discussion

5.1 Conclusion

5.2 Weakness

5.3 Next step

In the next step, we will.

Appendix

All code used to generate this results can be found here: https://github.com/erikaaa233/Final Wickham et al. (2019) R Core Team (2020) Wickham (2016) Hodgetts and Alexander (2020)

References

"2019 Canadian Federal Election." 2020. https://en.wikipedia.org/wiki/2019_Canadian_federal_election.

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

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

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

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