# Retrospective Report on 2019 Canadian Federal Election: if All Eligible Voters Participated

Ilke Sun

4 January 2021

## Retrospective Report on 2019 Canadian Federal Election: if All Eligible Voters Participated

**Ilke Sun**

**4 January 2021**

## Abstract

In this paper we analyze the 2019 Canadian Federal election using a multilevel regression with post-stratification approach. The estimates in this paper predict the results for the vote shares for the political parties in the nation and in the provinces if all citizens, who are eligible to vote, participated with a valid vote. We model seven different Bayesian logistic regressions that estimate the probability for an individual to support one of the seven different choices depending on their province, age group, sex, income and education. These choices include the six big parties in the 2019 election and other parties which have participated in the 43rd Federal Election. Then we post-stratify these estimates to the population and sub-populations. The population for this paper is the Canadian census and the sub-populations are considered to be the provinces. After comparing the real election results with the estimated results we cannot say that Liberal Party will definitely form a majority government. Nevertheless, Liberal Party is expected to obtain the larger share of votes in the nation. Also, if the turnout was 100% Conservatives (i.e. Conservative Party and Bloc Quebecois) would not be likely to gain as many seats as they did in the 2019 election.

**Keywords:**

Multilevel Regression with Post-stratification (MRP), 43rd Canadian Federal Election, Bayesian Multilevel Modelling, Observational Study

## 1 Introduction

Voter turnout is an important aspect of democracy. Democracy most effectively functions if all eligible citizens vote. In the last Canadian Federal Elections, the voter turnout was only 62% (Clarke, S., & Levett, C., 2019). Although, voter turnout has risen by 0.9 percentage points compared to previous federal elections it is still not at the desired rates (Lupick, 2019). Voter turnout is especially important in Canada's electoral system, first past the post (FPTP), in which the election results do not only depend on which party gets the majority of share of the votes.

Since the start of 2000s there has been an apparent decrease in percentage of eligible voters participating in the Canadian elections. Pammett and LeDuc (2003) aimed to understand this decline. They have found out that socioeconomic factors such as age, education, income, place of birth and mobility play an important role in citizen's voting behavior, moreover, this decline is majorly driven by age. Younger cohorts tend to

contribute much less in the electoral system (Pammett & LeDuc, 2003; Elections Canada, 2020). In addition to these factors, people living in the reserves also vote less. In the 2019 elections, only 51.8% of the citizens living in the reserves have participated. This percentage was 61.8 in the previous, 42nd Federal, election which was held on 19th of October 2015 (Elections Canada, 2020). Citizens' non-voting behavior, based on the 2000 survey, is also influenced by lack of interest in the election, negative attitudes towards politics, personal reasons and administrative reasons (Pammett & LeDuc, 2003).

In this paper we aim to analyze the possible changes in the results of 43rd Canadian federal election if the turnout was 100%, that is if all eligible voters have contributed with a valid vote. In order to obtain these retrospective results, we have employed multilevel regression with post-stratification (MRP). Hence, there are two datasets used in this paper: Canadian Election Study which was held in 2019 and General Social Survey which was done in 2017. The former is the survey data in order to understand the voter inclination and later is the demographics data which is used to post-stratify the estimates obtained from the survey. We aim to obtain the estimates by focusing on the socioeconomic characteristics of citizens and how they influence voting for each party. These variables are the province the citizen lives in, age group, income bracket, education level, and sex.

# 2 Methodology

Multilevel regression with post-stratification is a fairly new and popular approach in statistics (Alexander, 2019). It is used to make good predictions with non-representative samples. This approach is employed in this paper to model voting behavior from the survey data and then post-stratify these estimates to the census data to make predictions about citizen's probability of voting for each political party depending on their age, sex, income, education and province. Wang et al. (2015), has showed how a non-representative sample can make fairly accurate predictions for election results when post-stratified. Although, our survey data - CES 2019 - is representative of the Canadian population, in order to match the variables with the census data - GSS 2017 - we had to either remove or simulate data for certain provinces of Canada (i.e. Nunavut, Yukon and Northwest Territories) and certain demographic variables (i.e. sex and age). We have chosen not to simulate data since we have not acquired enough knowledge on how these variables influence voting behavior and in order to not raise any ethical concerns. Simulating data, in this case, may have cause this study to be biased and may influence our predictions in an undesired manner.

## 2.1 Data

The first dataset used in this paper was created through a phone survey which was done prior to the election in 2019. There were several questions asked in this survey and few of them are in the interest of this paper. These variables include province, age, gender, educational level, vote choice of the respondent and income of the household. In order to correctly post-stratify these results we needed to match the variables in the CES 2019 data with GSS 2017 data, which is the census dataset for this study. In addition to considerable amount of missing data in the CES 2019 dataset the discrepancies between the two datasets have influenced our decision for the choice of variables. The GSS dataset does not include the gender variable, instead it includes sex which was reported as binary variable (i.e. male or female). Although, gender may influence the voting behavior we have excluded non-binary respondents from the CES 2019 survey in order to not make any assumptions without sufficient evidence for these individuals' voting behavior. Moreover, the GSS 2017 dataset also does not include the three Canadian provinces: Yukon, Nunavut and Northwest Territories. Each of these provinces only have one seat in the parliament which makes a total of three seats. Although, three seats maybe the deciding factor between estabilishing a majority or a minority government, in order to accurately post-stratify we have also removed these respondents from the CES survey dataset.

In the survey data there are 18879 observations and 13 variables in total. 7 of these 13 variables indicate the respondent's support for the political parties. These 7 variables show if the respondent plans to vote for Liberal, Conservative, Bloc Québécois, New Democrat, Green, People's or another party. Hence, if one of these variables is equal to 1 rest them of them equal to 0 (e.g. $supportLP = 1, supportCP = 0, supportBQ = 0...$). The rest of the variables are province, age group, income bracket, education level, sex and age. In the model, which is going to be mentioned in the up coming section, we use age group instead of age which is made up of

7 different categories. Furthermore, there are 6 different income brackets and education levels. As seen from the table below there are only 10 provinces in the dataset which is due the deletion of the three provinces since those provinces were not present in the GSS 2017 data, hence, their estimates would not post-stratify.

In the census data there were 19841 observations and 5 variables, initially. These variables were province, age group, income bracket, education level and sex. After these variables are grouped and counted the concentrated dataset, which we have used in this study, is made up of 5040 observations and 6 variables. The added variables is n, this variables represents how many individuals have the same demographic indicators (i.e. province, sex, age, income and education) are in the census. Although, this variables is not included in the regression, it is used to proportion the weights of the estimates to make accurate predictions. The observations in this dataset correspond to number different cells made up from different combinations of demographic variables. Considering we have sex (2 categories), province (10 categories), education (6 categories), income (6 categories) and age group (7 categories) as our explanatory variables we must have 5040 ($2 \times 10 \times 6 \times 6 \times 7$) different cells.

Below we have five cross-tabulation tables for categorical auxiliary variables in our survey dataset and for some of the outcome variables (i.e. support for some party).

```
##                              SupportGP
## province                        0    1
##   Alberta                    2132   95
##   British Columbia           1836  308
##   Manitoba                    774   58
##   New Brunswick               349   83
##   Newfoundland and Labrador   295   10
##   Nova Scotia                 444   68
##   Ontario                    6741  682
##   Prince Edward Island         47   15
##   Quebec                     3928  335
##   Saskatchewan                644   35
```

```
##             SupportLP
## agegroup        0    1
##   ages18to24  643  272
##   ages25to34 2210 1018
##   ages35to44 2230 1148
##   ages45to54 2106 1073
##   ages55to64 2539 1338
##   ages65to74 2157 1245
##   ages75plus  567  333
```

```
##     SupportNDP
## male    0    1
##    0 8078 1925
##    1 7873 1003
```

```
##                          SupportCP
## income                       0    1
##   Less than $25,000       1408  457
##   $25,000 to $49,999      2417  975
##   $50,000 to $74,999      2781 1364
##   $75,000 to $99,999      2345 1194
##   $100,000 to $ 124,999   1458  783
##   $125,000 and more       2262 1435
```

```
##                            SupportBQ
## education                      0    1
```

```
##   No Highschool            736   62
##   Completed Highschool    4496  338
##   Some College            3794  195
##   Some University         1675  115
##   Bachelor's Degree       4727  220
##   Degree above Bachelor's 2390  131
```

## 2.2 Model

In this report, in order to make deductions regarding the possible results, we have written seven different equations that estimates person's probability of voting for each party depending on their province, sex, age group, income and education. These estimates are individual level estimates. In these equations only the outcome variable (i.e. probability of voting for a given party) differ. We model this regression with a random intercept model using Bayesian approach. The random intercept is generated by province differences. We expect that local and provincial factors to cause additional variation between different provinces, hence, in order to account for these group effects we use different intercepts for different states. Moreover, the model equations that estimate the citizen's probability of voting for the Liberal Party and Conservative Party are given below. There are 5 other equations that estimate for different parties.

$$log(\widehat{SupportLP}_{ik}/(1-\widehat{SupportLP}_{ik})) = B_0^{LP}+B_1^{LP}(male)+B_2^{LP}(agegroup)+B_3^{LP}(education)+B_4^{LP}(income)+s_k+e_{ik}$$

$$log(\widehat{SupportCP}_{ik}/(1-\widehat{SupportCP}_{ik})) = B_0^{CP}+B_1^{CP}(male)+B_2^{CP}(agegroup)+B_3^{CP}(education)+B_4^{CP}(income)+s_k+e_{ik}$$

where $B_0$ corresponds to average support for the party across provinces, $B_{1,2,3,4}$ are the coefficients associated with the explanatory variables, $s_k$ is the provinces mean deviations from the overall mean and $e_{ik}$ corresponds to the individual deviations from province means. These intercepts (i.e. $(B_0^{LP}+s_k), (B_0^{CP}+s_k)$...) are also simultaneously evaluated. Considering the random intercept model for Bayesian approach an example equation for these intercepts is

$$(B_0^{LP}+s_k) = r_{00}+r_{01}(province_k)+u_{0k}$$

In this model the outcome is the intercept in our model. $r_{00}$ is the intercept of the intercept and $r_{01}$ is the coefficient associated with the effect of the citizen's province on her voting behavior and $u_{0k}$ is the error term associated with this equation.

Using the model equations we obtain probability of support for one of the seven party options for the $i^{th}$ individual from the $k^{th}$ state. In addition, from the model equations we obtain different intercepts for different states. For example, the intercept for supporting New Democrat Party if the citizen is from Quebec is $B_0^{NDP}+s_{QB}$, whereas if the citizen is from British Columbia the intercept is $B_0^{NDP}+s_{BC}$.

After modeling these seven different equations we had to post-stratify these estimates to the census in order to obtain group level estimates, because only the individual level estimates given above would not predict the potential results. The main idea behind the MRP approach is partitioning the census into cells based on the combination of their demographics data which is done through weighing these estimates, appropriately, depending on the relative proportion in the population (Wang W. et al., 2015) . For example, there are 96 females in Quebec who are between the ages 65 and 74, who are in the lowest income bracket (i.e. Less than 25,000 CAD) and did not complete highschool, but there is only 1 female in Alberta who is younger than 24 and has competed highschool and earns between 50,000 to 74,999 CAD in the sample of the GSS data. Hence, the weight associated with the individual level estimate for the females in Quebec will be greater than the one associated with the female in Alberta since these weights are based on number of people belonging in same demographic combination. Thus, we must weight the probabilities that are obtained from the individual level estimates to obtain poststratification estimates. The postratification estimates are

$$\widehat{SupportLP}^{PS} = \frac{\sum_{j=1}^{J} N_j \widehat{SupportLP}_j}{\sum_{j=1}^{J} N_j}$$

$$\widehat{SupportCP}^{PS} = \frac{\sum_{j=1}^{J} N_j \widehat{SupportCP}_j}{\sum_{j=1}^{J} N_j}$$

for supporting the Liberal and Conservative parties. $\widehat{SupportLP}^{PS}$ is the postratified estimate of supporting the Liberal Party at the national level. In these equations $\widehat{SupportLP}_j$ and $\widehat{SupportCP}_j$ corresponds to estimate for party support in cell $j$. Where cells are the different combinations of demographics data. $N_j$ in these equations correspond to the size of the $j^{th}$ cell in the census. Again, there are 5 more poststratified estimates for other political parties. The important benefit of the post-stratification is that it enables us to not only make estimates for the national level (i.e. the whole population) but also for sub-populations (e.g. provincial level).

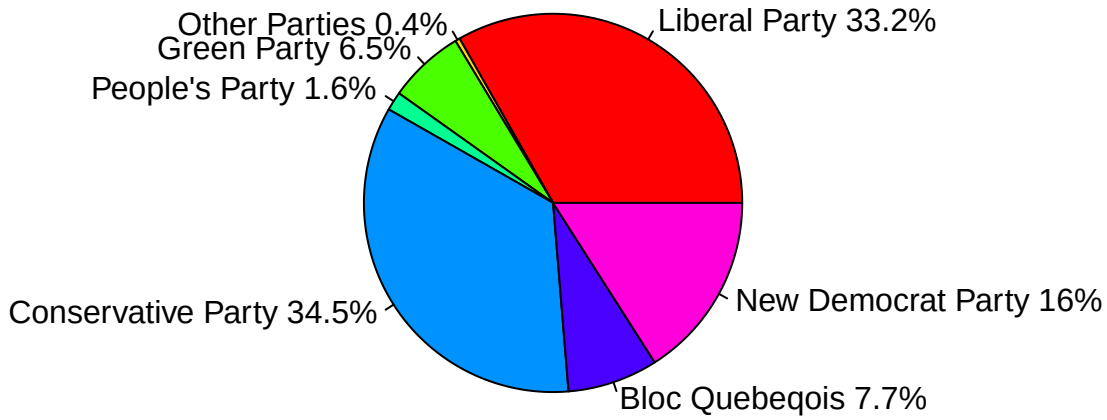
$$\widehat{SupportLP}_k^{PS} = \frac{\sum_{j \in J_k} N_j \widehat{SupportLP}_j}{\sum_{j \in J_k} N_j}$$

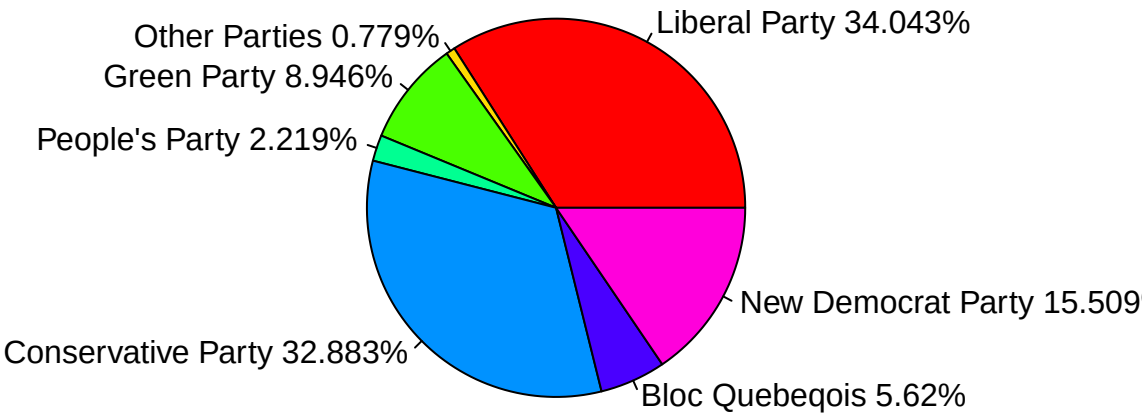
$$\widehat{SupportCP}_k^{PS} = \frac{\sum_{j \in J_k} N_j \widehat{SupportCP}_j}{\sum_{j \in J_k} N_j}$$

The equations above are the sub-population estimates for the Liberal and Conservative parties. $k$, again, denotes the province. Hence, these sub-populations are the 10 different provinces in our data. Now, we do not count the cells which do not belong to the sub-population, hence, the sigma notation differs from the national level estimates for each party. The reason for this change is sub-population weights may not include an individual from every cell they rather inlude the cells that are in correspond to demographic combinations for that province.

# 3 Results

## Vote Share of 2019 Canadian Federal Election (CBC News, 2019)






First, we will talk about the national vote share per party in Canada if all citizens participated in 2019 election. From our estimates we expect that if all eligible participants participated in voting, the national vote shares for Liberal, Green and People's parties would have grown whereas the vote shares for Bloc Quebeqois, Conservative and New Democrat parties would have shrunk. It is likely that if everyone participated vote share for other parties would have also increase. In the 2019 election only 142,347 votes were send for parties other then six biggest parties in Canada (CBC News, 2019). More than half of these votes indicated support for the independent candidate Jody Wilson-Raybould. Hence, her share of votes may have grew if all eligible voters participated but we cannot be sure of this because the CES 2019 questionnaire did not include her name specifically rather it was included in support for other parties. The discrepancies between the real vote share and the estimated vote share causes us to question if Liberal Party would be able to have a majority government? This is not only due to Liberal Party having a larger vote share in the estimations but also due to the fall in vote share of the three biggest rivals (i.e. Conservative, Bloc Quebecois). In these elections Liberals did much worse than 2015 elections. In 2019 Liberal Party has lost 20 seats, whereas Conservative Party has gained 26 and Bloc Quebecois has gained 22. Hence, using our models we estimate that Liberal Party would have gotten the majority share of votes but we cannot certainly suggest that they would be able to have a majority government.

In order to evaluate the potential results we must consider the provinces and districts within these provinces. However, since individuals' districts are not included in the 2017 GSS data our post-stratified estimates do not certainly predict which party would have won which seat. Our estimates rather predict the vote shares for parties in 10 different provinces.

For the provincial results we compare the estimates obtained from weighing our estimates accordingly for the sub-populations (i.e. provinces) to vote shares of the 2019 election results by province (Duffin, 2019). The vote shares that we have predicted have similarities and differences from the actual results. For example, there is not much surprise in Prairies. This region is dominated by Conservative Party supporters. Our model has predicted that in Alberta, province that has the fourth most seats in the parliament, 60.34% will support the Conservative Party. In the other two provinces in this region this percentages were approximately 40% in Manitoba and 52% in Saskatchewan. These three provinces have the highest votes shares for the Conservative party when compared with the other provinces. In the real elections Alberta had 1 seat for an NDP candidate and Manitoba had seven seats for liberal parties, moreover, Conservative Party has 64 out of 121 seats at the parliament from this region. From our analysis, however, we would expect that if all citizens in these provinces have voted the support for Conservative Party will fall for all three of the provinces in this region. The most drastic decrease from these states is from Saskatchewan where we expect approximately 12 percentage points decrease in the vote share for the Conservative Party if everyone has voted.

In all of the provinces besides Ontario, Quebec, New Brunswick and Prince Edward Island we have predicted

an increase in the vote share for the Liberal Party if turnout was 100%. This still does not make it certain that the party would have been able to form a majoritarian government. However, change in the support when turnout was 100% may have helped Liberal Party to win more districts and gain more seats in the parliament. In our estimates the Liberal Party had the highest share of votes in 6 provinces and Conservative Party was the highest in the other 4 but our model has estimated that votes shares for the Conservative Party would have declined in all provinces except Quebec. Although it is not a certainty, this may indicate that Conservatives would not be able to win 22 seats in the 2019 election.

From our estimates Bloc Quebecois Party did not appear to be in the top four choice of any province but Quebec which is also true in the real election results. Also, if everyone voted we predict that Bloc Quebecois will have only 26.62% of the votes in Quebec compared to 32.5%. We have also expected that in the provinces that have had representative from Green Party (i.e. New Brunswick and British Columbia) to have greater support for this party. We have also expected that support for Green Party would intensify by approximately 3 percentage points in Prince Edward Island.

# 4 Discussion

## 4.1 Summary

In our analysis we have reported what changes would occur in the election results in 2019 if all eligible citizens had participated a valid vote. In order to assess the changes we have modeled seven different multilevel random intercept models. In these models the explanatory variables were age group, income, education, sex and province. The outcome variables are the probability of supporting a given party. The reason why there are seven different model equations is that we have considered the support for the top six parties in the race and other parties. Moreover, other parties were also considered since an independent candidate has won a seat in the parliament in 2019. Furthermore, we have post-stratified estimates obtained by the model equations using a Bayesian approach. First we have post-stratified to the national level than to sub-populations which are considered as provinces for this report. The post-stratification is done through weighing the estimates and those weights depend on the demographic data available in the survey and the census data. For even better estimates and prediction of election results we should post-stratify to the districts within states so that we would be able to estimate the number of seats won by each party. Our results have indicated that national vote share for Liberal Party would likely to exceed Conservative Party's vote share. Moreover, Bloc Quebecois wouldn't be able to capture as much of vote share as they did in 2019 election. Furthermore, there is also an observed increase in Green Party's and other parties' vote shares; the later may indicate additional support for the independent candidate, Jody Wilson-Raybould. Observing the provincial estimates in most of the provinces Liberal Party's vote share would have increased. Also, in all provinces except Quebec we expect that the vote share of the Conservative Party will fall. Bloc Quebecois was in the top four parties per province only once and this was in Quebec which aligned with the actual election results. The support for Green Party, especially in the provinces that tend to vote for this party, would be likely to intensify if everyone that is eligible has participated.

## 4.2 Conclusion

From our analysis we can say that we expect vote shares for the Liberal Party to exceed Conservative Party if everyone has participated in the election. The national post-stratification estimates show that, but also observing the change in votes share in the provinces show if everyone participated vote shares for Liberal Party will increase in most states and vote share for the Conservative Party will fall in most states. However, vote shares for the Liberal Party have slight fell in the two biggest provinces of Canada (i.e. Ontario and Quebec) which may inhibit them from winning more seats. Our model has also predicted that Bloc Quebecois Party, which has made surprise increase in the seat count in this election, would have gotten less support in Quebec if everyone has participated. They had the 32.5% of the votes in Quebec which caused them to win 32 seats in total, all from Quebec. Our model has predicted that if everyone voted they would have gotten 26.6% of the votes from Quebec which may have caused them to win less seats. In our estimates vote shares for the New Democrat Party has fluctuated the most: in four provinces we expected their shares to

fall, in New Brunswick we expected it to relatively stay the same and in rest of the five states to slightly fall. However, our estimations have suggested that if everyone in Prince Edward Island has voted we would see that vote share for NDP almost doubles. The actual vote share from this state was 7.6% but we have predicted that it would be 15.74% if everyone has voted. Since the estimated vote share for the Liberal Party in Prince Edward Island is 43.42% it is unlikely that they will be able to have seat from this province.

By comparing the actual 2019 Canadian federal election results with our estimates we cannot confidently say that Liberal Party would be able to form a majority government. Our estimates suggest that the presidential race would also be close if every Canadian citizen has participated with a vote. However, our estimates suggest that Liberal Party would have had the greater vote share in the nation and Conservatives (i.e. Bloc Quebecois and Conservative Party) would not be able to gain a total of 48 seats because we predict that their vote shares would have fallen in most of provinces if everyone participated.

### 4.3 Weaknessess & Next Steps

Although MRP models make fairly accurate predictions using non-representative samples, missing data for non-binary individuals and for individuals from three territories of Canada (i.e. Yukon, Northern Territories, Nunavut) is a considerable weakness of this paper. In order to make more concrete predictions we must include these individuals in the data and should try to focus on gender instead of sex. Sex is a significant factor in most of our regressions on this paper but gender may also have an influence in the voting behavior. Hence, it should not be neglected. It is needless to say that we should also not neglect those three territories even though those places have much less population compared to other provinces. For the future, we may want to consider a census data that records gender and sample individuals from all of the provinces instead of specific ones. So that we can account for the effect of these different variables and create post-stratification estimates for three more provinces. Moreover, the GSS data, which is the census data, can be richer. When we clean the dataset for the desired interest variables we are left with 19841 observations in the census. Although, it is enough it can be much bigger. We would expect that as the census data enlarge weights associated with cells would resemble their true proportion in the population and this would help us to make even more accurate predictions.

In addition to missing variables and not including states, there is almost a two year difference in data collection time between the two dataset. The CES data was collected in 2019 prior to the elections whereas the GSS data was collected in 2017. Hence, there may be slight differences in the census in the two years. However, this is not a crucial weakness. It is unlikely that this 2-year difference created a bias in our results. We have considered to adjust the ages in the GSS data, however, this would change one of the variables even though it keeps all of the other variables the same. For example, a 17-year old which was not included in the census data we have used would have been included in the census and the variables besides age would have been kept the same. However, this individual may have completed highschool in this 2-year time period, or may have jumped a income bracket. Hence, we have decided to not manipulate age because we do not expect a drastic change in census demographics in these 2-years.

The most important step for the future is to consider the federal districts. The Canadian Elections do not depend on the vote shares in the provinces or the nation. Even though vote shares are likely to resemble the election results, they do not confidently predict the results. In the 2019 elections, Conservative Party had the most votes (i.e. highest vote share) but they were not able to win the election. Instead Liberal Party with 1.3 percentage points less vote share was able to form a minority government by exceeding 170 seats (i.e. more than half of the parliament)(CBC News, 2019). Hence, in order to predict the results we must predict the who are going to win the seats. And in order to predict the probability of winning these seats we should have different model than the one we have here. In this analysis province can be used an explanatory variable and distrcits can have intercepts of their own since local problems are often likely to influence voting behavior. In order to do that we must have data for which district the voter belongs to. Nevertheless, this was neither available in CES data nor the GSS data. Hence, future of this research may depend on post-stratifying to these sub-populations so that the model predicts the vote shares in the districts so that we can predict which party is going to win that district, thus, the seat in the parliament.

# References


Alexander (December 2019). Rohan Alexander: Getting started with MRP. Retrieved from https://rohanalexander.com/posts/2019-12-04-getting_started_with_mrp/

CBC News. (2019). Canada Votes 2019. Retrieved from Canadian Broadcasting Corporation News: https://newsinteractives.cbc.ca/elections/federal/2019/results/

Clarke, S., & Levett, C. (2019, October 23). Canada election 2019: full results. Retrieved from The Guardian: https://www.theguardian.com/world/2019/oct/22/canada-election-2019-full-results

Duffin, E. (2019, October 22). Canada election 2019: preliminary share of the popular vote per party by province. Retrieved from Statista: https://www.statista.com/statistics/1062264/canada-election-share-of-popular-vote-per-party-by-province/

Elections Canada. (2020). Voter Turnout at Federal Elections and Referendums . Retrieved from Elections Canada: https://www.elections.ca/content.aspx?section=ele&dir=turn&document=index&lang=e

Elections Canada. (2020). On-Reserve Voter Turnout – 43rd General Election. Elections Canada.

Lupick, T. (2019, October 22). Voter turnout for Canada's 2019 federal election was, well, meh. Retrieved from The Georgie Straight: https://www.straight.com/news/1316836/voter-turnout-canadas-2019-federal-election-was-well-meh

Pammett, J. H., & LeDuc, L. (March 2003). Explaining the Turnout Decline in Canadian Federal Elections: A New Survey of Non-voters. www.elections.ca.

Stephenson, Laura B., Allison Harell, Daniel Rubenson and Peter John Loewen. The 2019 Canadian Election Study – Online Collection. [dataset]

Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. International Journal of Forecasting.