Effect of Turnout Rate on the Canadian Federal Election

Ying Xiong(1004795885)
Code and data supporting this analysis is available at: https://github.com/xiongy15/STA304-FINAL

12/22/2020

Contents

1 Abstract

The sliding voter turnout has been a significant issue faced by Canada's electoral system. (Heard, 2019) The democratic election result should fairly represent Canadians' opinions on the development of the country. A low turnout implies a lack of legitimacy for the system, which reduces the vote's credibility. (Morden & Urban, 2017) It leads to incongruent results in the popular and electoral votes. In this study, I explore the dataset of "General Social Survey on Family 2017" and "Canadian Election Study - Online 2019" to develop a logistics regression model for estimating the probability of voting for a specific party. I apply the post-stratification technique to the model and generate a simulated popular vote. Notably, I analyze five predictors (age, sex, household income, education level, living geographic area) that may affect the respondent's intention to vote. From the result, I can conclude that the popular vote will be an effective indicator of the winner when the turnout rate is high.

2 Keywords

Canadian Federal Election, Popular Vote, Turnout, Post-stratification, Liberal Party, Conservative Party

3 Introduction

The 2019 Canadian Federal Election ended with Justin Trudeau (Liberal Party) 's continuation as the Prime minister. A much different result was expected from the popular vote, where the Conservative Party headed by Andrew Scheer was 1.2% ahead. (Global News, 2019) Generally, leading in public opinion polls indicates the party will win the election as the polls promptly reflect the voters' thoughts. It functions as a non-official indicator. (Watters, 2019) Notably, the turnout rate in the 2019 election collected by Elections Canada was around 67%. (Heard, 2019) In other words, only two-thirds of the eligible voters participated in the election and expressed their views. It primarily weakens the credibility of the popular vote result because fewer people were represented.

This study aims to analyze the importance of turnout rates on popular votes. I will use a logistics regression model with the post-stratification technique to predict the election result of the popular vote with a 100% turnout and make comparisons. From the model, I will explore how the demographics related to the probability of voting for specific parties. The model can provide a guideline to political parties on their supporters' general information to improve the campaigns for the next elections. The model and predictions will be constructed through two data sets, the 2019 Canadian Election Study online survey and the 2017 General Social Survey. This report will focus on five elements: age, sex, household income, education level, and living geographic area.

The report will provide a reflective analysis through different sections. In the Data section, I will describe the raw data with the cleaning process. The Methodology section will contain the final model with diagnostics and the post-stratification technique used to perform the predictions. I will present the simulated prediction results with in-depth discussions and analysis in the result section and discussion section. Finally, I will make inferences in conclusion, along with the weakness and future steps. Supplemental information will be included in the Appendix.

4 Data

4.1 General Social Survey (GSS) on the Family, 2017

The 2017 GSS data is retrieved from the CHASS data center and conducted from Feb.2 to Nov.30, 2017. It gathers information on Canadians' demographics, well-being, and living conditions, collecting through computer-assisted telephone interviews. The survey's target population is all full-time residents of institutions that are at least 15 years old and live in the 10 Canadian provinces. The frame population contains people on the valid phone number lists available to Statistics Canada and the Address Register. People who answer the calls are the sample population. The survey applies stratified sampling and simple random sampling without replacement. It divides strata based on provinces and treats Census Metropolitan Areas[1] as separate strata. Respondents are randomly selected from each eligible household within each stratum. The non-responses are adjusted for the weight and later dropped from the dataset. In the raw data, there are 20602 observations with 461 variables.

- Strength

The data includes various information, including basics like demographics and detailed records of living conditions. Along with regular collections, it can be used to analyze social trends and emerging policy issues over time. The responses in the data are specific due to the well-design of the survey questions. The stratified sampling method ensures the inclusion of all the strata to reduce the non-represented and underrepresented population.

- Weakness

The data excludes residents from the territories, making it less representative when others want to investigate all Canadian residents. It also subjects to both sampling and non-sampling errors. Non-responses occur when respondents misunderstand the questions or refused to answer, resulting in non-sampling errors. Since the total survey takes a long time, many useless variables are filled with too many NAs (Not Available).

4.2 Canadian Election Study (CES) - Online Survey, 2019

The 2019 CES Online data is collected by Advanis Inc and distributed by the Canadian Opinion Research Archive. It is conducted from Sep.13 to Nov.11, 2019, through an online survey. The survey gathers information on Canadians' attitudes and thoughts during and after the 2019 Canadian Federal election. The target population is all Canadian citizens and permanent residents that are at least 18 years old. The frame is people who are accessible to the Internet and see the survey. The sample population is people who take the survey. The online sample applies a two-wave panel with a modified rolling-cross section. The non-probability survey is distributed on all kind of online platforms and wait for respondents. In other words, participants volunteer to take the survey and share their opinions. There will be no non-responses since it only includes people who take the survey voluntarily. In the raw data, there are 37822 observations with 634 variables.

- Strength

The data has a large sample size from online platforms, reducing the non-represented and underrepresented population. The survey includes various essential political topics and gains opinions from respondents. It makes the dataset useful in accounting for variables that affect election from different aspects. Researchers can explore from different angles and develop well-grounded analysis.

- Weakness

The survey could be biased based on researcher judgment since it is non-probability sampling. Its validity cannot be measured through sampling error factors. For questions without options to choose, there are enormous different long answers, which could be useless for analysis.

4.3 Data Cleaning

I correspond the survey data variables to the census data variables through recategorizing to apply the post-stratification technique. Irrelevant observations are also removed at this phase. Significant changes are done in variables about age, income, and citizenship and discussed below. The visualizations of cleaned data sets are included in Appendix 8.2 (Figures 1 & 2).

- Age

Initially, the age variables are numerical in both data. I classify them into 6 groups that contain all the eligible ages. I used 25 as the cut-off for the youngest group because most people will be around 25 when they finish graduate studies. Notably, the census data GSS was conducted in 2017, while the survey data

[1]Based on 2011 Census geography. These are places with a highly dense population such as Toronto.

CES was in 2019. Since 18[2] is the minimum age to be eligible to vote in Canada, I retain observations with ages ranging from 16 to 25 in the census and 18 to 25 in the survey. I also set a retired group with a lower boundary of 65[3], the retirement age in Canada. The remaining ones are separated with a 10-year difference.

<age_25, age_46_55, age_56_65, age_26_35, age_66, age_36_45>.

- Income

Initially, the census data records income in 6 levels. There is also a categorical variable about income in the survey data, *cps19_income_cat*. However, groups are divided with different boundaries. Specifics are included in Appendix 8.1. Thus, I turn to the numeric variable *cps19_income_number* and reclassify the income groups into 3 larger ones. The average income of $500000 (Dodge, 2020) and the top 0.1% income of $750000 (Data & Insights, 2020) are used as the cut-offs.^[The Canadian average wage in 2020 is above $54,630 per year for a full-time worker.

<Above Avg, Below Avg, Rich>.

- Citizenship

To be eligible to vote in Canada, one has to be at least 18 years old and identified as a Canadian citizen. (Government of Canada, 2020) Thus, I removed observations identified as permanent residents or residents without citizenship from both data sets.

5 Methodology

I aim to predict the vote result of the 2019 Canadian federal election with the full turnout. To simulate the outcome, I apply the post-stratification technique on the census data, General Social Survey 2017. I estimate the probability of each post-stratification cell voting for the Liberal Party with my model. The model is generated from the survey data Canadian Election Survey 2019.

5.1 Model Specifics

This model is for the Liberal Party. The following mathematical notation is the logistic regression model I create in R with five predictors, *province*, *age_group*, *edu_group*, *income_group*, and *sex*.

$$\log \frac{p_i}{1-p_i} = \beta_0 + \beta_{province} X_{province} + \beta_{age} X_{age} + \beta_{edu} X_{edu} + \beta_{income} X_{income} + \beta_{sex} X_{sex}$$

p_i: the probability of the respondent voting for Liberal Party in the 2019 Canadian federal election

p_i: the probability of the respondent voting for the Liberal Party in the 2019 Canadian federal election

β_0: the intercept value; represents the log odds of a female living in Alberta who is smaller than 25 years old and has an above-average income with post-graduation education level. This information would be the reference groups for each variable in the following discussion.

$\beta_{province}$: the slope values for provinces; represent the difference in preferences between people in Alberta and the other 9 provinces

$X_{province}$: = 1 if the respondent lives in provinces except for Alberta

β_{age}: the slope values for age groups; represent the difference in preferences between people aged below 25 and the other 5 age groups

X_{age}: = 1 if the respondent's age belongs to the age group except for below 25

β_{edu}: the slope values for education level; represent the difference in preferences between people with post-graduation education level and the other 3 levels

[2]https://www.electionsnb.ca/content/enb/en/voters/eligible.html
[3]https://www.canada.ca/en/services/benefits/publicpensions/cpp.html

X_{edu}: = 1 if the respondent has an education level except for post-graduation

β_{income}: the slope values for income groups; represent the difference in preferences between people with above-average income and the other 2 income groups

X_{income}: = 1 if the respondent has an income above the average of Canadians

β_{sex}: the slope values for sex; represent the difference in preferences between females and males

X_{sex}: = 1 if the respondent is identified as a male

Given each variable's information, I can effectively predict a person's probability of voting for the Liberal Party. The model variables separate the data into different categories that allow quick classification of new respondent's information and output a relatively precise prediction.

5.2 Model Selection

To simulate the electoral votes with the full turnout, I employ a logistic regression model to predict the probability of a person voting for the Liberal Party. The model applies the frequentist method because vote preferences may change from year to year due to external factors such as politics. There is no applicable prior information for the Bayesian inference. Since the response variable *vote_liberal* is binary (i.e., whether the respondent will vote for the Liberal Party or not), a logistic regression model is the best to fit the data. More importantly, one of this report's purposes is to develop a model that can accurately predict the election results so that it can be used in future elections for reference. Logistic models can provide the probabilities directly through p_i.

5.3 Analysis on Variables

I employ 5 predictors in my final prediction model, *province*, *age_group*, *edu_group*, *income_group*, and *sex*. To avoid potential errors that may mislead the probability, I choose to use the existing variables that are matchable to the census data and have a small number of NAs. These variables are also chosen based on my interest and for simplicity.

All the variables are categorical. Notably, the *age_group* and *income_group* are created through numeric variable *cps19_age* and *cps19_income_number*. The cleaning process is in the data − data cleaning section.

- I separate ages into different groups to match the census data. The survey provides the ages in integers. In the GSS data, ages are presented in numbers with decimals, making it hard to define for each age. I need to create age groups for the post-stratification cells while it is impossible for me to add decimals to the survey age data. Thus, I have to create age groups to match. To save time and effort, I divide them with a 10-year difference, representing a generation. However, this will lead to larger variation since different people within one age range will be considered identical and applied with the same coefficient.
- I separate the income into different groups to match the census data. The GSS data presents incomes in groups, while the CES data has numeric values. To save time and match, I reclassify the income groups into 3 big ones. Similar to the age groups, this will result in larger variation as the group definition is broader.

The model above is the one with comparatively small AIC[4] in terms of variable inclusion among trials and runs.

- Since most betas for the variable *age_group* have large p-values, I generate another model without it. Referring to Table 2 (AIC comparison for Liberal Party's Models (Age)) in Appendix 8.3, it turns out that the AIC of the model would increase if I exclude the age group variable.
- The *sex* variable reduces the AIC for the Liberal Party model by a minimal amount. However, I consider this factor as an essential aspect to include. One reason is that it enlarges the AIC value for the Conservative Party model. Another reason is that a conspicuous gender gap in polls is discovered

[4]Akaike's information criterion. A method used to select models. The smaller, the better.

for the Liberal Party in the latest federal poll created by Abacus Data. According to the data, women's willingness to voting for the Liberal party is 5% higher than men's. (Abacus Data, 2020) Thus, I will keep the *sex* variable. Specifics are shown in Appendix 8.3 (Table 3 & 4)

The other three variables are all statistically significant, and I retain them as predictors without further tests.

5.4 Model Diagnostics

1. Linearity Assumption

Since I do not include any numeric variables, I do not need to check the linear relationship between continuous predictor variables and the log odds.

2. Multicollinearity

I check the multicollinearity between the predictors to reduce the effect on the inference. If there is a moderate correlation between independent variables, the model's inference will deviate from the actual relationship and result in wrong interpretations. The model will then be useless for further analysis if I do not remove one of the correlated variables. I applied the Variance Inflation Factor to check. As shown in Appendix 8.4 (Table 5 & 6), the VIFs for each predictor is around 1, while the typical cut-off for VIF is 5 or 10. Thus, I don't need to remove variables.

3. Influential Observations

Refer to Appendix 8.5 (Table 7). There are 1983 leverage points in the survey data, which means they are extreme in their x-values and may deviate from the regression line. Then I applied the Cook's Distance to find influential observations, and there are 0. Thus, I don't need to remove any observations.

5.5 Alternative Model

One may argue that a multiple linear regression model can reproduce the work above if I change the response variable to *cps19_lr_scale_bef_1*. [5] The model will result in a numeric answer with decimals implying the respondent's political stand. However, it is hard to define the position of each political party on the spectrum. According to Appendix 8.6 (Table 8), people tend to describe the liberal party from 2 to 6 and the conservative party from 6 to 9. If I get a 6, there will be a dispute on whether this person will be more likely to vote for liberal or conservative. In other words, the cut-off number for each party is blurry and would need an in-depth investigation, which requires much more effort and information.

5.6 Post-Stratification Technique

I employ the post-stratification technique with the logistic model to predict the outcome of the 2019 Canadian Federal Election with the full turnout. The technique applies to the census data and partitions it into different cells according to variables. Cells are all possible combinations, and each stands for people with identical information. Then the model will predict the vote probability for each subgroup and then output a final election result. The post-stratification technique can effectively reduce bias from non-responses and underrepresented groups and thus have a lower variance.

Same as the model, I use *province*, *age_group*, *edu_group*, *income_group*, and *sex* as variables to create the post-stratification cells. The cleaned census data consists of 1121 different cells.

In Canada, provinces use the same plurality voting system[6] while they can still change as they will without request permissions. There are leading parties in each province and may differ from each other, which adds to the necessity to analyze. For age groups, people of varying generations usually have different opinions towards politics. It applies to the gender gap, as well. As mentioned above in the variable discussion, the latest federal poll collected by Abacus Data shows that the Liberal party is welcomed by citizens from provinces

[5] This numeric variable contains people's self-evaluation on the political position. The survey asked the respondents to answer in the range from 0(most left) to 10(most right), "Where would you place yourself on the political spectrum?"

[6] https://www.fairvote.org/plurality_majority_systems

like British Columbia, the youngest eligible generation and females. (Abacus Data, 2020) There is also a strong relationship between the education level and voting preferences. (Taube, 2019) Besides, a party's policies toward the economy and employment may affect people's support rate with different income levels.

6 Results

Table 1: Predicted Popular Vote for Liberal v.s. Conservative Party

Liberal	Conservative
0.2927	0.2593

- Referring to **Table 1(Predicted Populat Vote for Liberal v.s. Conservative Party)**, I estimate that the Liberal party may receive 29.3% of the total votes, and the Conservative party may receive 25.9% of the total votes.

The results are generated from the post-stratification analysis with the logistic regression model. For each post-stratification cell, the logistic model I developed outputs a probability of voting for the Liberal Party. The final result is the sum of probabilities weighted by their weight calculated by dividing each cell size by the total population. The process is done in both the Liberal Party model and the Conservative Party model. I only make predictions for these two parties because they are the major competitors in the 2019 Canadian Federal Election.

Referring to Appendix 8.7(Table 9: Model for Liberal Party), the model results present large p-values for the sex variable and most age groups, which implies the variables are statistically insignificant. Nevertheless, I retain these variables due to the emerging gender gap and varying thoughts from generations discussed above. These are also basic demographics, which adds to the necessity to keep them. In Appendix 8.8(Table 10: Model for Conservative Party), all variables are considered statistically significant. This indicates that the model perfectly fits the data when predicting votes for the Conservative party. Therefore, my prediction from the model assumes the Liberal Party will defeat the Conservative Party by about 3% in the popular vote in the 2019 Federal Election.

7 Discussion

7.1 Summary

The report is conducted on the basis of the survey data 2019 Canadian Election Study Online Survey from Canadian Opinion Research Archive and the 2017 General Social Survey on the Family from the CHASS data center as the census data. The CES data includes respondents' demographics with their attitudes on the 2019 Canadian Federal Election vote. The GSS data consists of respondent's demographics and living conditions. To adjust for the year of GSS data, I include all observations with age starting from 16 in 2017 because they will 18 in 2019 and become eligible to vote.

To match the variable of model and post-stratification cells, I clean and create 5 variables (*province*, *age_group*, *edu_group*, *income_group*, and *sex*) in both datasets. Using the CES survey data, these explanatory predictors are employed in my logistic regression model to predict the probability of voting for one specific political party. I apply the post-stratification technique to the GSS data and partition it into cells based on the above variables. For each cell, I estimate the probability of voting by inputting the information into the model. The final vote prediction result is the sum of probabilities weighted by cell size over total observation size. I estimate the proportion of voters in favor of voting for the Liberal Party is 29.3% and 25.9% for the Conservative Party, implying a win for the former.

7.2 Conlusion

According to the model predictions, I estimate 29.3% of the total popular vote for the Liberal Party and 25.9% for the Conservative Party. The result implies that the Liberal Party will be favored by voters in the 2019 Canadian Federal Election by about 3% higher when the turnout rate is full.

Notably, the winner of election in Canada is decided by the party with more seats. Thus, the popular vote will be an indication of the winner instead of the real outcome. In reality, the final result of the election was the winning of the Liberal Party. Simultaneously, the Conservative popular vote was 1% higher than that of Liberal, and the turnout was about 67%. (Heard, 2019) The result of my prediction verifies that a party with a higher popular vote is usually the election winner. Occasions like the 2019 election do exist but are very rare. There were only 3 similar elections in history. (Watters, 2019) For the theory to be realized, there is a vital assumption of full or high turnout rate. With a low turnout, the credibility of popular vote outcome plunges, and it becomes vague when predicting the winner of the election. Therefore, it is crucial to enhance the turnout rate to improve the popular vote's usefulness as an indicator. Besides, a higher turnout rate indicates a better representation of the voters' thoughts and contributes to the development of a country.

In addition to the popular vote result, political parties could use the model results to clarify their supporters' demographics and recognize people they should attract in the next election to win higher seats. They can adjust their policy proposals by analyzing their target voters to achieve more effective campaigns and get votes from those citizens.

7.3 Weakness

1. The variable selection process is full of subjectivity. I choose the variables that I think will have an impact on their political standing and preferences. Other factors may affect the probability as well and add accuracy to the model if I include them. For example, ethnicity could be a vital factor to discuss.
2. The census data is collected in 2017, which is 2 years away from the 2019 Canadian Federal Election. Factors such as education level and income could vary vastly. Thus, less representative inputs may output deviated probabilities and affect the precision of the model.
3. During the data cleaning, I removed all the observations with NA, consisting of about one-third of the raw data. The model could have a different result if I can attain complete information.
4. The survey data use non-probability sampling through online platforms. Although the sampling size is big, there will still be less representative of the entire target population. It limits the sample to people who can access the Internet, which may underrepresent populations such as the old and the poor.

7.4 Next Steps

1. I can gather more updated census data to create detailed and complete post-stratification cells and develop a more well-grounded prediction result.
2. A multilevel regression model could be employed for alternative models since data includes a group-level variable like provinces.
3. I can employ other variable selection techniques and do trials and runs to find the best model. Measures like adjusted R^2 and BIC (Bayesian Information Criterion) can also be used to compare models. The residual sum of squares (RSS) can be applied when choosing among models with the same number of predictors. To avoid overfitting, cross validations will be helpful when there are too many independent variables.
4. I can simulate the probabilities of voting for each political party within each state and calculate the party's potential seats. The party with more seats will be the winner, which corresponds to the Canadian electoral system. Then I can compare the results to enhance the completeness of the report.

8 Appendix

8.1 Variable Data Comparison

- Survey Data:

<Don't know/ Prefer not to answer, $110,001 to $150,000, NA, $90,001 to $110,000, More than $200,000, $150,001 to $200,000, $60,001 to $90,000, $1 to $30,000, $30,001 to $60,000, No income>.

- Census Data:

<$25,000 to $49,999, Less than $25,000, $50,000 to $74,999, $125,000 and more, $75,000 to $99,999, $100,000 to $ 124,999>.

8.2 Data - Cleaned Raw Dataset

Figure 1: Cleaned Census Data

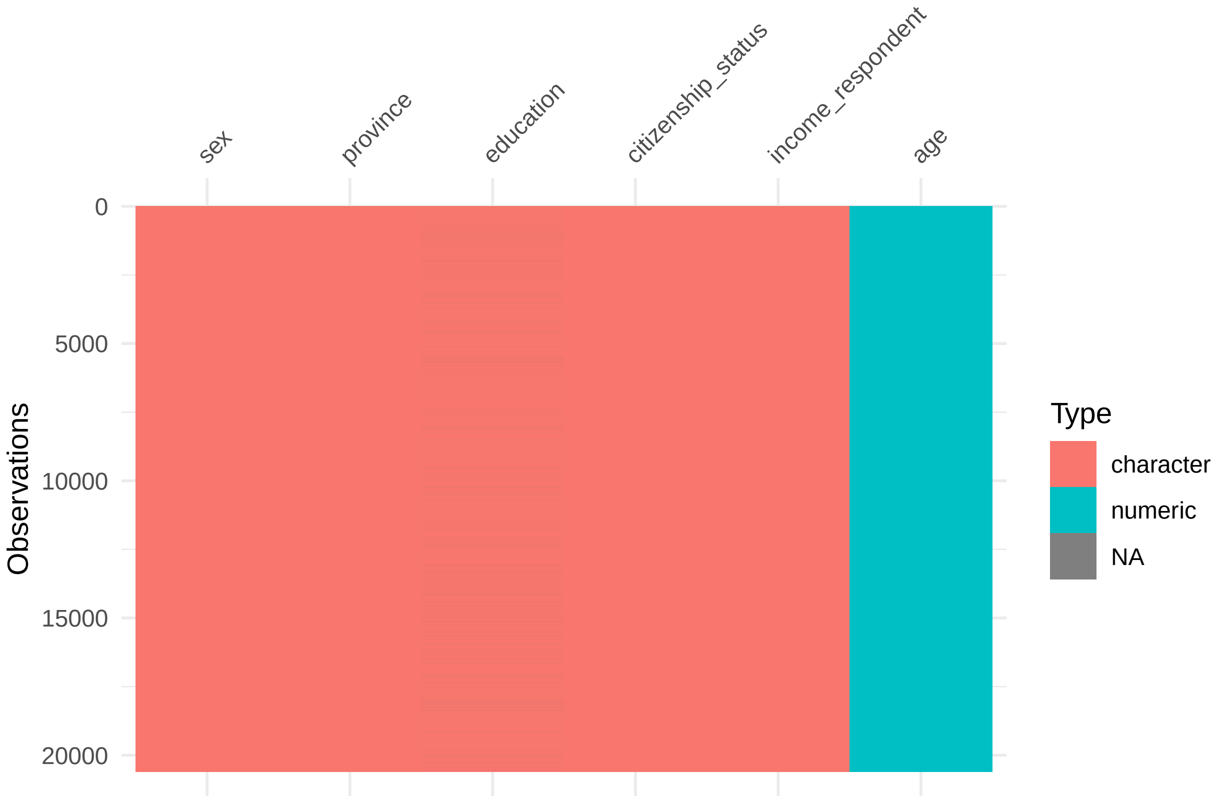
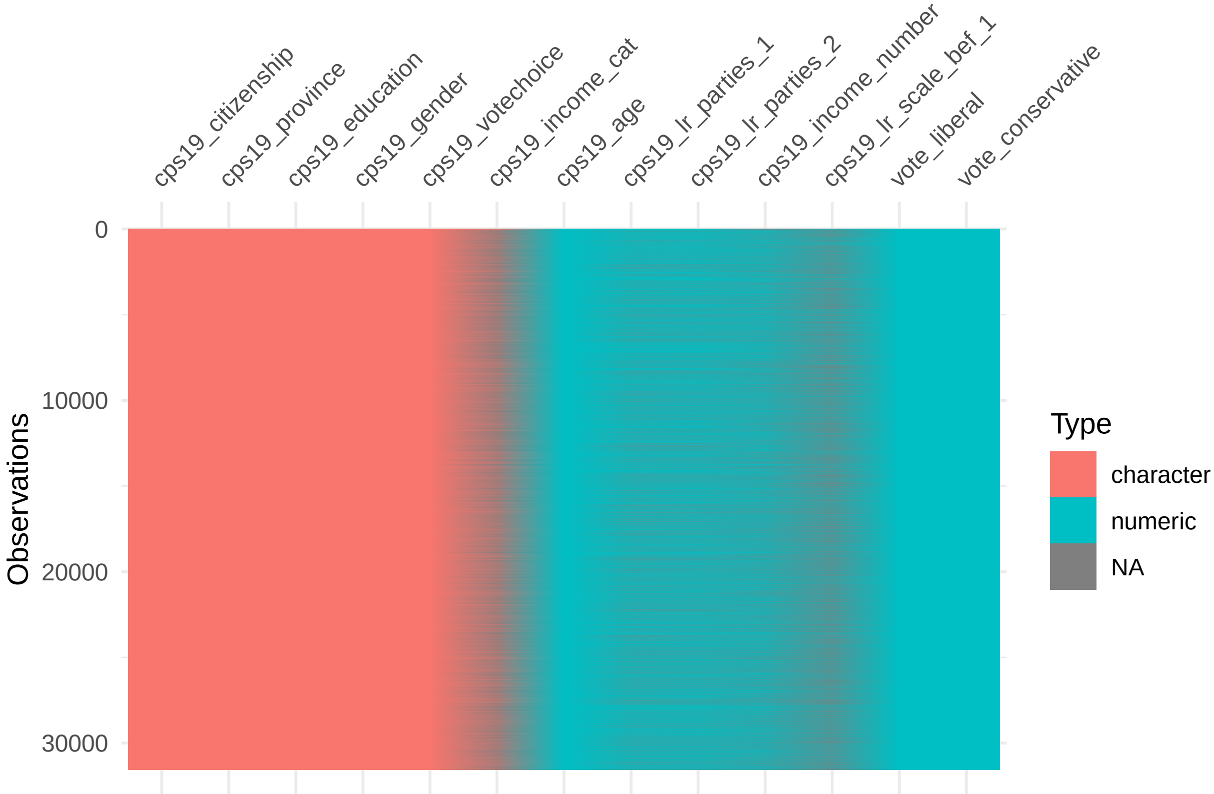


Figure 2: Cleaned Survey Data



8.3 Model - Variable Selection

Table 2: AIC comparison for Liberal Party's Models (Age)

With_age	Without_age
25583.5	25605.3

Table 3: AIC comparison for Liberal Party's Models (Sex)

With_sex	Without_sex
25583.5	25581.8

Table 4: AIC comparison for Conservative Party's Models (Sex)

With_sex	Without_sex
23896.4	24115.8

8.4 Model - Multicollinearity

Table 5: VIFs for Liberal Party's Model

	GVIF	Df	GVIF^(1/(2*Df))
province	1.029	9	1.002
age_group	1.102	5	1.010
edu_group	1.119	3	1.019
sex	1.055	1	1.027
income_group	1.131	2	1.031

Table 6: VIFs for Conservative Party's Model

	GVIF	Df	GVIF^(1/(2*Df))
province	1.038	9	1.002
age_group	1.098	5	1.009
edu_group	1.123	3	1.019
sex	1.055	1	1.027
income_group	1.126	2	1.030

8.5 Model - Influential Observations

Table 7: Leverage Points & Influential Obeservations

Leverage_pt	Influential_pt
1983	0

8.6 Alternative Model

Table 8: Political Position of Each Party (0 Left - 10 Right)

	Liberal_Party	Conservative_Party
0%	0	0
25%	2	6
50%	4	8
75%	6	9
100%	10	10

8.7 Model for Liberal Party

Table 9: Summary of Logit Regression for Liberal Party

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.469	0.105	-13.941	0.000
provinceBritish Columbia	0.700	0.071	9.793	0.000
provinceManitoba	0.762	0.091	8.357	0.000
provinceNew Brunswick	1.067	0.110	9.660	0.000

	Estimate	Std. Error	z value	Pr(>|z|)
provinceNewfoundland and Labrador	1.352	0.120	11.246	0.000
provinceNova Scotia	1.238	0.100	12.388	0.000
provinceOntario	1.033	0.060	17.334	0.000
provincePrince Edward Island	1.233	0.245	5.034	0.000
provinceQuebec	0.844	0.063	13.349	0.000
provinceSaskatchewan	-0.208	0.121	-1.727	0.084
age_groupage_26_35	-0.043	0.082	-0.525	0.599
age_groupage_36_45	-0.002	0.082	-0.020	0.984
age_groupage_46_55	-0.052	0.082	-0.636	0.525
age_groupage_56_65	0.027	0.080	0.339	0.735
age_groupage_66	0.183	0.079	2.298	0.022
edu_groupPost-Secondary	-0.240	0.045	-5.274	0.000
edu_groupPrimary	-0.674	0.092	-7.287	0.000
edu_groupSecondary	-0.421	0.049	-8.578	0.000
sexMale	-0.018	0.031	-0.590	0.555
income_groupBelow Avg	-0.029	0.044	-0.659	0.510
income_groupRich	0.124	0.040	3.132	0.002

8.8 Model for Conservative Party

Table 10: Summary of Logit Regression for Conservative Party

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.847	0.110	-7.695	0.000
provinceBritish Columbia	-1.181	0.061	-19.276	0.000
provinceManitoba	-0.772	0.080	-9.661	0.000
provinceNew Brunswick	-1.267	0.113	-11.246	0.000
provinceNewfoundland and Labrador	-1.563	0.137	-11.444	0.000
provinceNova Scotia	-1.791	0.115	-15.508	0.000
provinceOntario	-1.204	0.048	-25.163	0.000
provincePrince Edward Island	-2.289	0.360	-6.352	0.000
provinceQuebec	-1.953	0.057	-34.326	0.000
provinceSaskatchewan	-0.308	0.084	-3.663	0.000
age_groupage_26_35	0.331	0.094	3.523	0.000
age_groupage_36_45	0.453	0.093	4.867	0.000
age_groupage_46_55	0.562	0.093	6.042	0.000
age_groupage_56_65	0.602	0.091	6.604	0.000
age_groupage_66	0.676	0.091	7.453	0.000
edu_groupPost-Secondary	0.315	0.052	6.039	0.000
edu_groupPrimary	0.590	0.090	6.532	0.000
edu_groupSecondary	0.498	0.055	9.079	0.000
sexMale	0.482	0.032	14.833	0.000
income_groupBelow Avg	-0.274	0.047	-5.886	0.000
income_groupRich	0.113	0.041	2.761	0.006

9 Reference

1. Abacus Data. (2020, June 23). Abacus Data: Liberals up by 11 as Conservatives stuck below 30%. Retrieved December 22, 2020, from https://abacusdata.ca/canadian-politics-poll-vote-june-2020-abacus-data/

2. Caetano, S. & Alexander, R. (2020, October 7). GSS Data Cleaning Code. Retrieved December 22, 2020, from https://www.tellingstorieswithdata.com/

3. Data & Insights. (2020, September 06). Rich, Richer, Richest: A Profile of High-Income Canadians. Retrieved December 22, 2020, from https://themeasureofaplan.com/high-income-canadians/

4. Dodge, P. (2020, September 02). Average Canadian Salary In 2020. Retrieved December 22, 2020, from https://www.jobillico.com/blog/en/the-average-canadian-salary-in-2020/

5. Employment and Social Development Canada. (2020, April 27). CPP retirement pension: Overview. Retrieved December 22, 2020, from https://www.canada.ca/en/services/benefits/publicpensions/cpp.html

6. FairVote.org. (n.d.). Plurality-Majority Systems. Retrieved December 22, 2020, from https://www.fairvote.org/plurality_majority_systems

7. Global News. (2019, October 24). Canada election 2019: Results from the federal election. Retrieved December 22, 2020, from https://globalnews.ca/news/6023150/live-canada-election-results-2019-real-time-results-in-the-federal-election/

8. Government of Canada. (2019, June 13). Canada Elections Act. Retrieved December 22, 2020, from https://laws-lois.justice.gc.ca/eng/acts/e-2.01/fulltext.html

9. Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2020). dplyr: A Grammar of Data Manipulation. R package version 1.0.2. https://CRAN.R-project.org/package=dplyr

10. Heard, A. (2019). Historical Voter Turnout in Canadian Federal Elections & Referenda, 1867-2019. Retrieved December 22, 2020, from https://www.sfu.ca/~aheard/elections/historical-turnout.html

11. John Fox and Sanford Weisberg (2019). An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

12. Kassambara, & U, M. (2018, November 03). Logistic Regression Assumptions and Diagnostics in R. Retrieved December 22, 2020, from http://www.sthda.com/english/articles/36-classification-methods-essentials/148-logistic-regressionassumptions- and-diagnostics-in-r/.

13. Morden, M., & Urban, M. C. (2017, February 7). Low voter turnout tends to produce bad government, so how do we get more Canadians to the polls? Retrieved December 22, 2020, from https://munkschool.utoronto.ca/mowatcentre/low-voter-turnout-tends-to-produce-bad-government-so-how-do-we-get-more-canadians-to-the-polls/

14. 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

15. Stephenson, Laura, Allison Harrel, Daniel Rubenson and Peter Loewen. Forthcoming. 'Measuring Preferences and Behaviour in the 2019 Canadian Election Study,' Canadian Journal of Political Science.

16. Taube, M. (2020, October 17). Conservatives "the party of the uneducated?" Retrieved December 22, 2020, from https://troymedia.com/politicslaw/conservatives-party-uneducated/

17. Technology. (n.d.). Advancing Knowledge through. Computing in the Humanities and Social Sciences. Retrieved December 22, 2020, from http://www.chass.utoronto.ca/.

18. Tierney N (2017). "visdat: Visualising Whole Data Frames." *JOSS*, *2*(16), 355. doi: 10.21105/joss.00355 (URL: https://doi.org/10.21105/joss.00355), <URL: http://dx.doi.org/10.21105/joss.00355>.

19. Watters, H. (2019, October 23). How often does the party with the most seats lose the popular vote? | CBC News. Retrieved December 22, 2020, from https://www.cbc.ca/news/politics/post-election-question-answers-1.5330943

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

21. Yihui Xie (2020). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.30.

22. Yihui Xie (2015) Dynamic Documents with R and knitr. 2nd edition. Chapman and Hall/CRC. ISBN 978-1498716963

23. Yihui Xie (2014) knitr: A Comprehensive Tool for Reproducible Research in R. In Victoria Stodden, Friedrich Leisch and Roger D. Peng, editors, Implementing Reproducible Computational Research. Chapman and Hall/CRC. ISBN 978-1466561595