Prediction of Individual Characteristic Shape Voting Preference for Liberal Party by Logistic Regression

Yilan Tan

December 5, 2021

Introduction

This report is a study on the prediction of the outcome of the election accurately. Many companies, government organizations, and agents built-up models, trying to find the relationship between voters' preference and other factors.

We used the survey data about people's preferences of voting and their characteristics to build our analysis. Then we use the census data from 2019 to complete the prediction.

Research Background

The winner of the Canada Elections in 2021 turned out to be Justin Trudeau (The Guardian,2021), and the election process was gripping every citizen's heart. In a democratic country like Canada, the U.K., or the U.S., that election is the most important political event in a power cycle. The change of party in power would lead to different policies in the economy, diplomacy, etc. The election will have a significant impact on an individual's life, the development of the country, and even the dynamics in global politics.

Based on the previous study, we have top reasons to believe that demographic variation plays a vital role in determining a voter's decision. For example, in 2006, Elisabeth Gidengil and her teammates pointed out that males and females presented significantly different preferences in voting for the right-wing and left-wing parties (Gidengil, Fournier and Everitt,2006). Meanwhile, according to Eric Guntermann, in the 2017 Catalan Election (Gunternmann,2017), variation in educational level led to significant differences in voting decisions. Inspired by those findings, we try to find what demographic factors affect voters' probability of voting for the Liberal Party. Therefore, in the research, we performed logistic regression by using predictors, including age, sex, education level, and living region, to figure out each variable's impact on the voting preference.

Furthermore, referring to the outcome in the regression model, we would perform a post-stratification analysis to predict the election in 2025. Specifically, the result of this research enables citizens to have a general idea about the potential party in power during the next cycle. It offers the Liberal Party quantitative support in strengthening her supporters and modifying her strategy to gain more votes in the 2025 election.

Data Sets

In the research, the first dataset we refer to is the GSS (General Social Survey: Families), conducted in 2017. The data was collected by using a computer-assisted telephone interview, with an overall response rate of 52.4%. The total number of usable observations is 20602 households, and the dataset includes essential variables, such as social-demographic information (age, education, sex), income level, etc.

The second dataset we use is the CES (Canadian Election Study) conducted in 2019. The data were collected with telephone interviews, with 4021 Canadian citizens aged over 18 completing the survey. There are 278 variables in the data, including sex, phone type, interest in the party, living region, and so on.

More detailed information about the above two datasets will be discussed in the later Data Part, including the general introduction, the methods of collecting data, and potential drawbacks based on their strategies.

Goal

We used the survey data about people's preferences of voting and their characteristics to build our analysis. We would like to see whether sex, education, and province will affect their final choice.

Hypothesis

We make four hypotheses based on our background research.

Hypothesis 1: We suppose that age will have a positive effect on our final result. As some report says, the Liberals are winning over the older(Fournier,2021). We want to justify whether a voter's age would impact the probability of their voting for a particular party in the 2017 election. With the population above 65 growing by 65% as expected, we intend to analyze the effect of age on voting decisions in an aging society.

Hypothesis 2: We suppose that sex will have a significant impact on people's voting choices. Based on Elisabeth Gidengil's findings, we want to test further whether sex leads to a significantly different voting preference for a specific party in the future election. Specifically, whether sex difference will increase or decrease voters' probability of voting for a particular political party to a significant level in the 2017 election.

Hypothesis 3:We suppose that education level will have a significant impact on people's voting choices. People with higher education levels may vote for Liberal Party more as some scholars found that the more educated people are more liberal(Jaschik,2016). Driven by Guntermann's research, we will test whether educational level plays an essential role in determining vote choice in Canada. Therefore, we will test whether the difference in education would affect the probability of voting for a particular party in the 2017 election.

Hypothesis 4: We suppose that the living province will significantly impact people's voting choices. Ontario and Atlantic Canada's LPCs are consistently at the top of the list at the regional level. Still, Quebec has recently widened the gap between liberals and Bloc Québécois, favoring liberals(Fournier,2021).In the 2017GSS, the survey stratified its target population according to the regions they live. Inspired by this sampling method, our team will test when holding other variables fixed whether people living in a particular area or province show significantly different probability in voting for a specific party.

Afterward, we will apply the outcome gained in logistic regression to predict the probability that Liberal Party will win the vote in the 2025 election. Meanwhile, we would perform post-stratification for other parties, including the Conservative and New Democratic Party, to compare better their probability of winning the election after all. Furthermore, we will discuss some improvable aspects and weaknesses in our research so that future related studies could produce even better and more accurate analyses.

Explanation of terminology that helps to understand the report:

1. The Liberal Party of Canada is Canada's longest-serving and oldest-acting Federalist Party. The party dominated Canada's federal politics for much of its history and has held power for almost 70 years in the 20th century. As a result, it is sometimes referred to as Canada's "natural ruling party."

2. The Conservative Party of Canada is a federal political party established in 2003 by several right-leaning parties, all of which possess a long history in the country for over one hundred years. When The Conservative Party was in power, it issued several policies and laws, including cutting taxes and meeting ends in the government budget.

3. The New Democratic Party is social-democratic in Canada, established in 1961 by the joint efforts of two political parties, called CCF and CLC. The political position, it's on the left side of the Liberal Party. As an official political party, the New Democratic Party has never possessed the largest share of seats at a national level.

4. The 2019 Canadian Federal Election (formally the 43rd Canadian Parliamentary Election) took place on October 21, 2019, and members of the House of Representatives were elected to the 43rd Parliament of Canada. Governor Julie Payette issued the ballot for the 2019 election on September 11, 2019.

Data

Data Summary and Collection Process

We used 2017 GSS Dataset about people's preferences of voting and their characteristics to build our analysis. Then we use 2019 CES Dataset to complete the prediction.

2017 GSS Dataset

1. Introduction to the 2017 GSS

GSS was originally established in 1995 and was responsible for collecting cross-sectional data by making telephone interviews with selected respondents in ten provinces in Canada. The main objectives of the program is to measure the changes in well-being of the whole society and to offer related information on social policy issues. The 2017 GSS was initiated on Feb 2nd, 2017 and was ended on Nov 30th, 2017. It put great emphasis on collecting comprehensive data of the selected citizen's household, such as the fertility intentions, health and subjective well-being. In the survey, the targeted population was all the people aged over 15 in Canada and selected respondents gave their feedback in telephone interviews.

2. Methods in 2017 GSS

The survey firstly divided the target population by dividing them into different geographic areas, called strata in stratified sampling. In the 2017 GSS, there were totally 27 strata. Afterwards, a simple random sample without replacement was conducted within each stratum to select the qualified respondents. The respondent would be contacted with a telephone interview and gave his feedback directly on the phone. However, due to this trait of 2017 GSS, there was a small proportion of households without landline or wireless telephones being excluded from the population. Although in 2017, the vast majority of household would be equipped with telephones, this sampling bias still exist and probably lead to distortion in data.

After all, with a response rate of 52.4%, the 2017 GSS had collected the household information from 20,602 qualified respondents.

3. Survey Content

The questionnaire covered 14 topics in different aspects of a household, including characteristics of respondents or their family members, the organization within the household and income level. Later, the questions about income were removed and as substitution, the survey added the data gained from respondents' fiscal files for reference. Under each section, the selected citizen was supposed to answer a series of questions both for verifying the qualification and offering crucial information.

4. Imputation

In order to guarantee the completeness of the data collected with telephone interviews, the 2017 GSS applied imputation to deal with the missing data. For example, for the missing value of birth month, the respondent's age would be automatically filled out with assigned solutions under different scenarios. This method has improved the completeness and reliability of original data, especially of those lacking in key values, but it also creates the concern that the imputed values were not accurate estimates of the true value, thus leading to distortions in the final dataset.

2019 CES Dataset

1. Introduction of 2019 CES

Similar to the 2017 GSS, 2019 CES also focus on the population who aged over 18 and live in ten provinces of Canada. The survey includes two phases in the whole process. During the election campaign, 4021 qualified Canadian Citizens were asked to complete the telephone interview for collecting their preference in different parties, while after the election, 72% of those respondents completed another telephone interview or online survey.

2. Methods in 2019 CES

In order to randomly select the qualified citizens, CES adopted the method called Modified Random Digital Dialing (RDD) and completed following process of surveys with computer assisted telephone interviewing in the first phase. As mentioned in the 2017 GSS, such method could contact respondents well but also exclude people without a landline or wireless telephone, creating the sampling bias.

3. Content of the Survey

In the survey, respondents were asked questions under various topics, such as their ratings towards the political leader of a certain party, and some specific questions only open to limited people. Due to the nature of the CES, there would be inevitably missing values in the dataset and the missing ones would be automatically plugged in with coding.

Table 1: Demo Survey Data

sex	age	province	education	vote_liberal
Male	56	Quebec	Bachelor's degree	0
Male	46	Quebec	University certificate or diploma below the bachelor's level	0
Male	25	Quebec	Bachelor's degree	1
Male	19	Quebec	University certificate or diploma below the bachelor's level	0
Male	35	Quebec	Above the bachelor's level	0
Male	80	Quebec	Less than high school diploma	0

Data cleaning

Step 1: Remove irrelevant observations

To study the relationship between the vote choice and the age, education, sex and province of the respondent, we choose the answer of question 2,question 3 ,question 4 and question 11 in the survey data for our analysis. Then we removed all the missing values in the data.

Step 2: Data transformation

Now we transform each of the variable.

For age: Q2: "In what year were you born?"

We firstly mutate a new variable 'age' to represent the age of the respondent by using 2019-the answer of the question 2(the birth year) as the survey was conducted in 2019.

By mutating this new variable, we can have a clearer sight of the age of the respondent. Using the year that the respondent was born in to measure the age of the building may misleading.

For sex: Q3: "with which sex do you identify?"

In surveys, there may be no appropriate way to reconcile sex and sex measurement. However, it is a significant systemic as well as individual issue. (Kennedy, L., Khanna, K., Simpson, D.P., & Gelman, A. ,2020). There are a number of ways to match the answer of sex identification in survey to sex in census. In this paper, for analyzing purposes, we transformed sex as a dummy variable, a variable can only be either 0 or 1. In our analysis, sex is 1 if the respondent is male or 0 otherwise.This may lead to inaccuracy of the final result and we will discuss the weakness in the following part.

For province: Q3: " In which province or territory are you currently living?"

We match the answer of the survey question with the answer of census data.As we have seen in the final result of survey. There are 3 provinces (Northwest Territories, Yukon, Nunavut) that do not have respondents. So we removed them from our dataset.

For education: Q61: " What is the highest level of education that you have completed?"

There are a lot of different answers to this question. For easy analysis, if the answer is 1="No schooling",2= "Some elementary school",3="Completed elementary school",4="Some secondary / high school",we will combine them as "Less than high school diploma". If the answer is 10="Master's degree",11="Professional degree or doctorate", we will combine them as "Above the bachelor's level".If the answer is 8="University certificate or diploma below the bachelor's level" ,we will match it to the answer in census "University certificate or diploma below the bachelor's level". If the answer is 6= "Some technical, community college, CEGEP, College Classi"or 7="Completed technical, community college, CEGEP, College C",we will combine them as "College, CEGEP or other non-university certificate".

For census, we match the answer from survey question to the answer of census.For example, if the answer is "College, CEGEP or other non-university certificate or di..." or "Trade certificate or diploma", we convert it to "College, CEGEP or other non-university certificate".

For the vote choice: Q11: " Which party will you likely to vote for?"

Final vote choice is a dependent variable. As we have discussed before, would like to see the odds of voting for Liberal Party/Conservative Party/New Democratic Party. So we mutate 3 new variables,for example,for predicting the odds of voting for Liberal Party,we mutate a new variable called"vote_liberal",this variable is binary. If the answer is 1=" Liberal (Grits)",then vote_liberal=1, otherwise=0.If the answer is 2="Conservatives",then vote_conservatives=1, otherwise=0. If the answer is 3=" NDP",then vote_NDP=1, otherwise=0.

Variable Description

1. Sex

Sex is a categorical variable that can take on two values: male and female. Approximately 56% of respondents in our data are male.

Table 2: Summary of sex

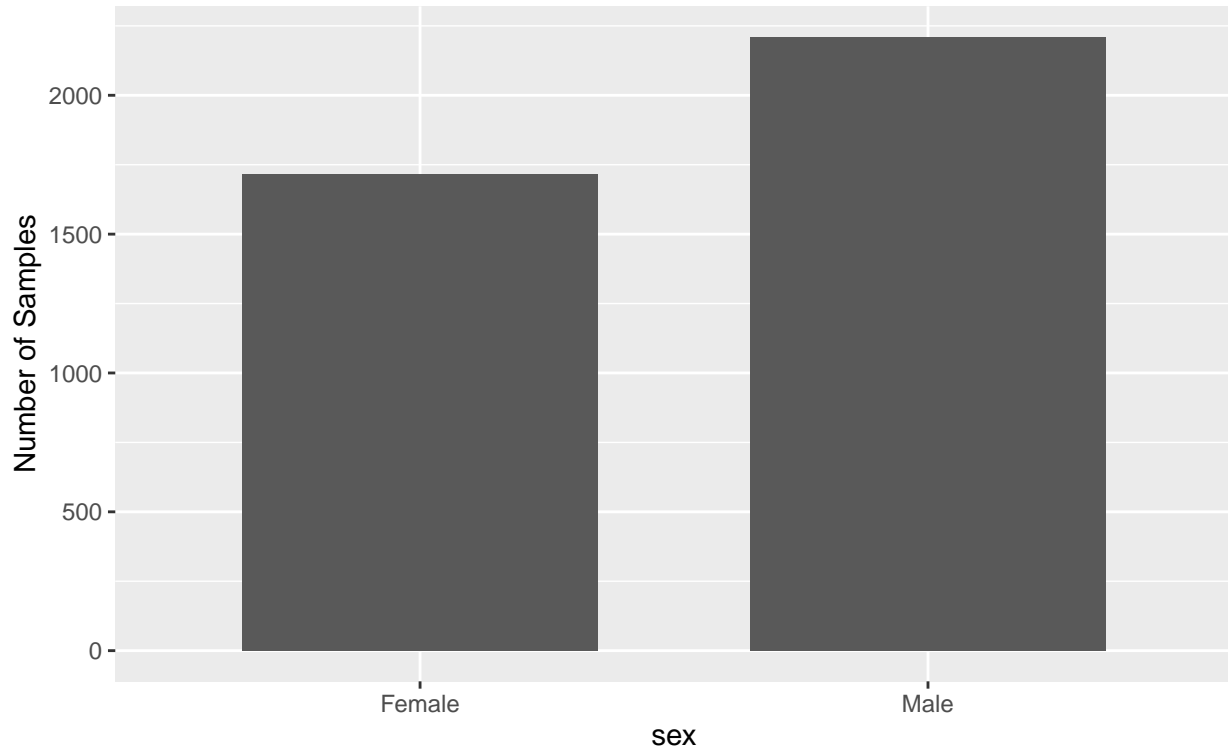| sex	Count
Female	1715
Male	2210

According to Graph One, the sex distribution is a little bit unbalanced among the respondents, as there are 2210 male respondents and 1715 female respondents.

Table 3: Overveiw of age

| min	Q1	median	Q3	max	mean	sd
18	38	51	64	100	50.89	16.83

Graph One
Bar Plot of Variable sex

2. Age

Age denotes the respondents' age at the time of the survey, which ranges from 18 to 100 years old. In particular, the median age is 51 years old, and 75% of all responses are under the age of 64.

According to Graph Two, the age distribution of GSS respondents can be seen. We used a centering approach on the variable of age to make comparing the age influence on conception rate easier.It's worth noting that the age distribution is a little bit right-skewed, indicating that there are a lot of young and middle-aged respondents but just a few seniors.

Graph Two
Distribution of Variable Age
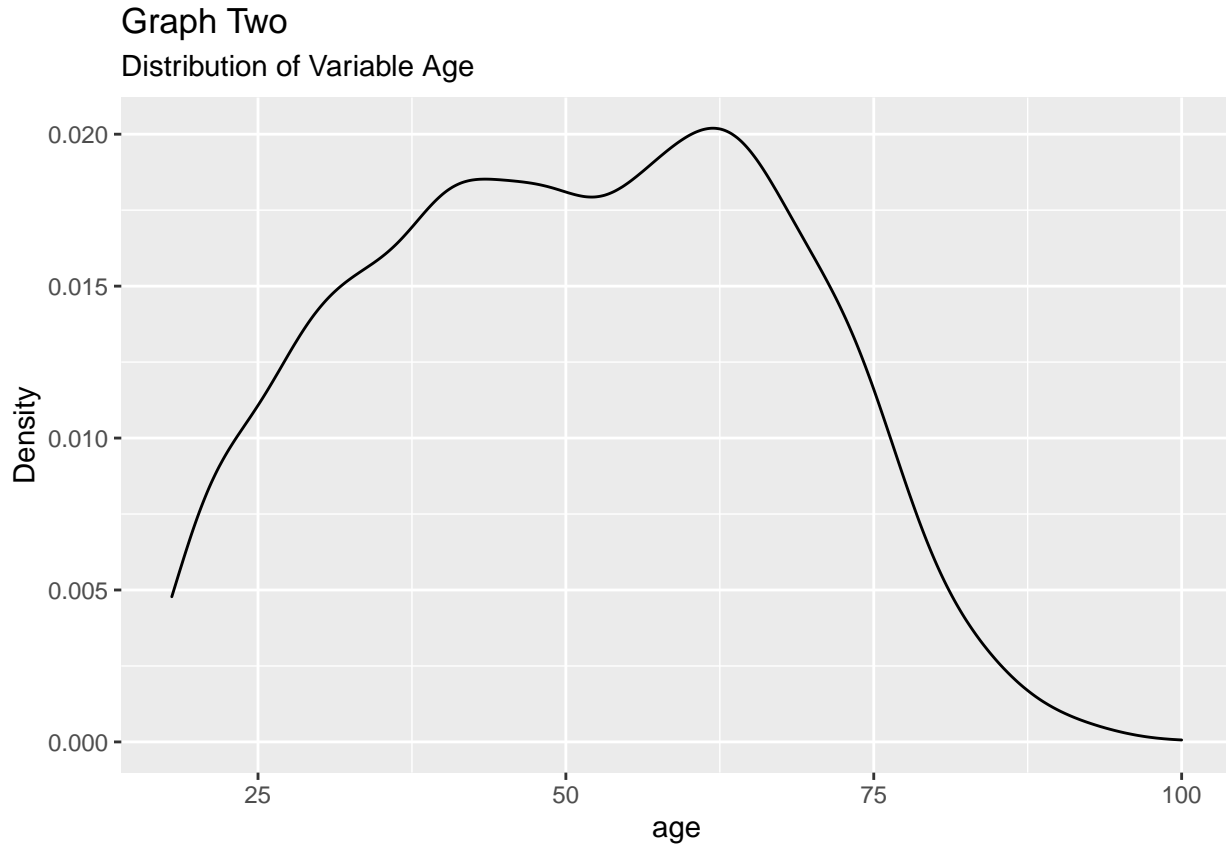
3. Province of the respondent

Province of the respondent indicates the province in which the respondent are currently living. With that said, it is a categorical variable that can take on ten values: "Newfoundland and Labrador", "Prince Edward Island", "Nova Scotia", "New Brunswick", "Quebec", "Ontario", "Manitoba", "Saskatchewan", "Alberta", "British Columbia".Table 4 is a summary table that details the number of respondents in each region of residence.

Table 4: Summary of Province

Province	Count
Alberta	279
British Columbia	782
Manitoba	255
New Brunswick	196
Newfoundland and Labrador	194
Nova Scotia	194
Ontario	794
Prince Edward Island	194
Quebec	779
Saskatchewan	258

According to Graph Three, we observed that the respondents from British Columbia, Ontario, Quebec account for the majority of our dataset,which is in line with the reality.

## Graph Three
Bar Plot of Variable Province
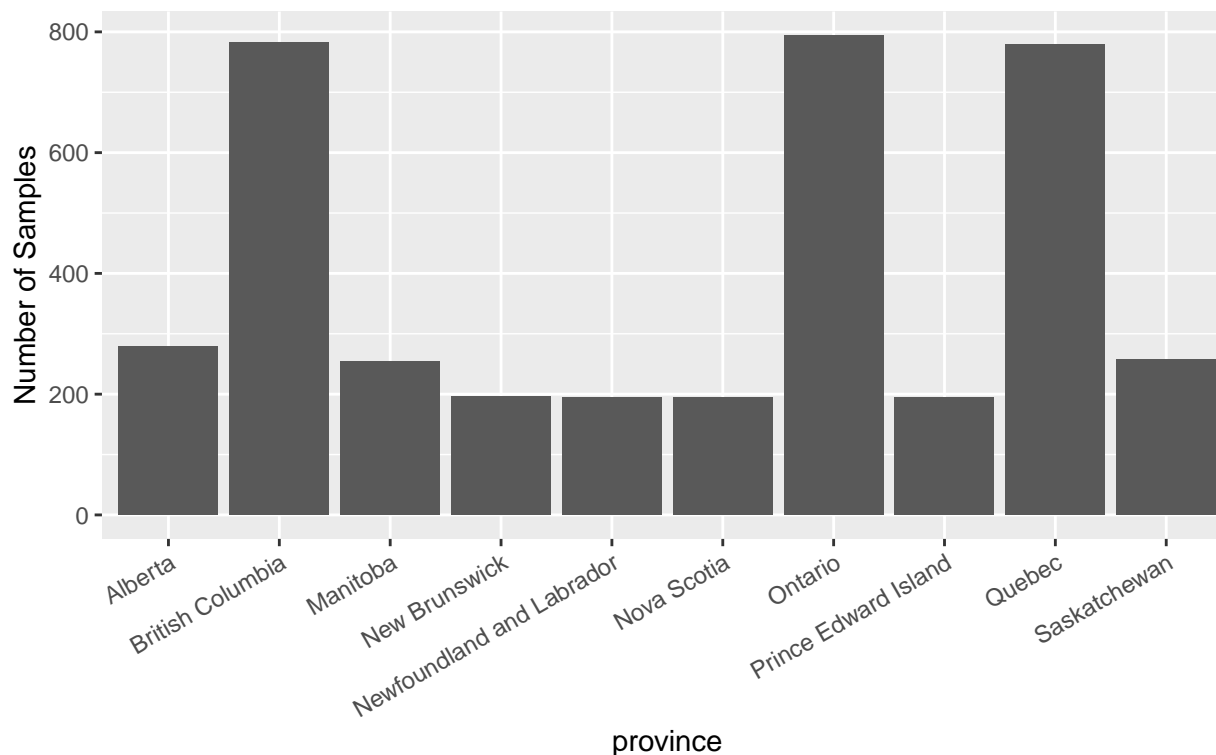
4. Education

Education indicates the highest level of degree that the respondent obtained. Therefore, it is a categorical variable that can take on 6 values:"Less than high school diploma", "High school diploma","College, CEGEP or other non-university certificate", "University certificate or diploma below the bachelor's level","Bachelor's degree","Above the bachelor's level". The number of responses for each degree of education is listed in the table.

Table 5: Summary of Education

Education	Count
Above the bachelor's level	626
Bachelor's degree	1077
College, CEGEP or other non-university certificate	1107
High school diploma	539
Less than high school diploma	230
University certificate or diploma below the bachelor's level	346

According to the Graph Four, respondents with a college diploma or other non-university credential made up the biggest sub-group in our sample, followed by those with a bachelor's degree and those with above the bachelor's level.

Graph Four
Bar Plot of Variable Education
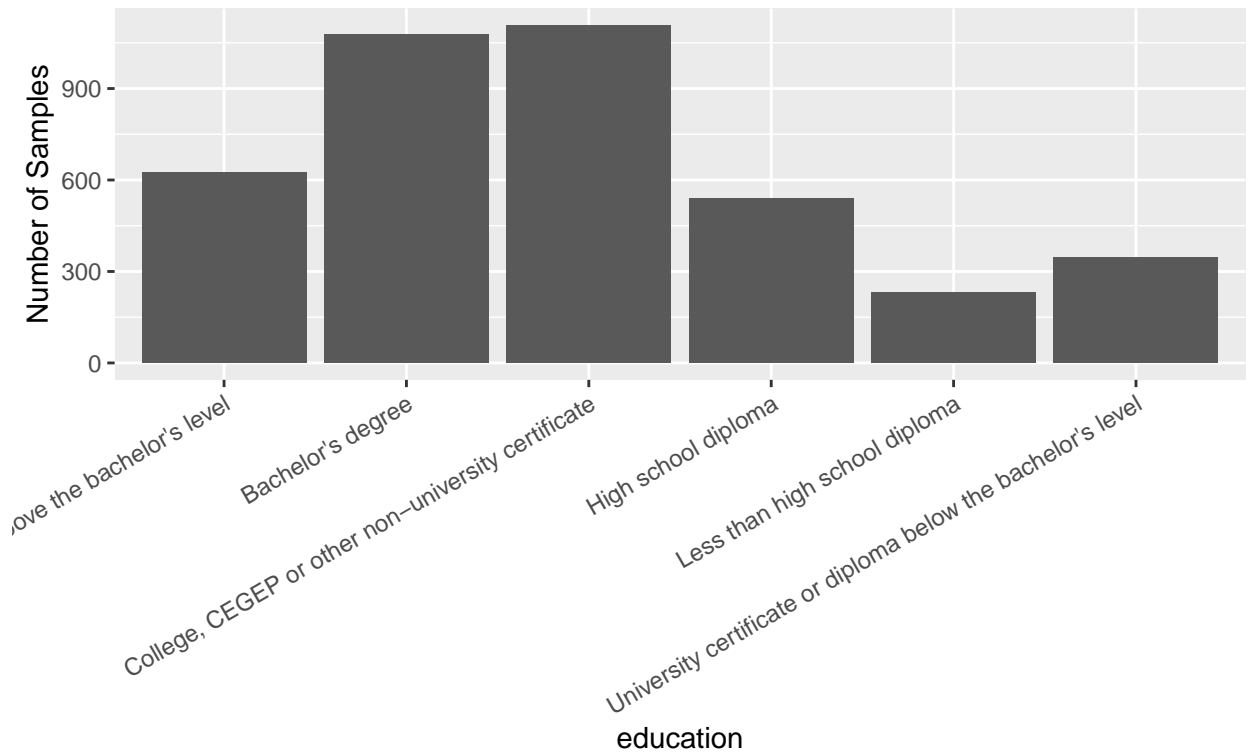
5.Vote for the Party

Final vote choice is a dependent variable. As we have discussed before, would like to see the odds of voting

for Liberal Party. We mutate a new variable called "vote_liberal",this variable is binary. If the answer is 1=" Liberal (Grits)", then vote_liberal=1, otherwise=0.From the Table 6,we can see in the total of 3925 respondents,there are 909 people who vote for Liberal Party, which accounts for 23.13% of the sample.

for Conservatives Party. We mutate a new variable called "vote_conservatives", this variable is binary. If the answer is 2=" Conservatives",then vote_conservatives=1, otherwise=0. From the Table 6,we can see in the total of 3925 respondents, there are 980 people who vote for Conservatives Party, which accounts for 24.9% of the sample.

for New Democratic Party. We mutate a new variable called "vote_NDP", this variable is binary. If the answer is 3=" NDP",then vote_NDP=1, otherwise=0. From the Table 6,we can see in the total of 3925 respondents, there are 405 people who vote for NDP, which accounts for 10.3% of the sample.

Methods

Model Specifics

A logistic regression model (GLM) shows a linear connection between a binary dependent variable and one or more covariates (also known as independent variables). The model's goal is to investigate the link between a binary dependent variable and a number of potentially affecting factors. We converted the variable of voting for each major party into a binary dependent variable via data cleaning, with 1 indicating they are intended to vote for this party and 0 suggesting that they do not. Given that we are seeking for how specific individual traits might impact respondents' intention to vote for one particular party (i.e., Liberal, Conservative, New Democrats), which is a binary answer, a logistic regression model is a suitable match.

We chose five variables, including one dependent variable: vote party (i.e., Liberal, Conservative, New

Democrats), and four independent variables: sex, age, province, and educational level, as mentioned in the preceding section. With this in mind, we may use a logistic model to see how individuals in various "groups" (sex, age, province, and educational level) respond to the question, "Which party are you going to vote for?" Take Liberal Party as an example. If we choose people who reside in Ontario as a reference group, the difference in log-odds between respondents who live in Ontario and those who live in British Columbia may be calculated using the variable coefficient that indicates whether the respondent lives in British Columbia. In the case of numeric variables, we can see how the log odds vary in response to changes in the numeric variable (basically, "age"). To summarise, a logistic regression model is an excellent option for exploring how individual traits influence voting.

The corresponding statistical model is shown below:

$$Y_i \sim Binomial(N_i, p_i)$$

$$log(\frac{p_i}{1 - p_i}) = logit(p_i) = X_i\beta$$

Where,

- N_i is the total sample size
- Y_i is the number of events in interest happening
- p_i is the probability of the respondent expressing the voting preference(for Liberal Party, Conservative Party, New Democratic Party)
- X_i is the vector of covariates (i.e., age, sex, the province they lived, education level)
- β represents the coefficient vector corresponding to each covariate

The model construction process:

Conduct a regression associating their voting preference regarding the specific Party to respondents' age, sex, province, and education level. As shown below, for individual i, let y_i equal 1 if he or she plans to vote for the Party and 0 otherwise.

$$Pr(y_i = 1) = \text{logit}^{-1}\left(\beta_0 + \beta^{age} + \beta^{educ}_{a[i]} + \beta^{province}_{b[i]} + \beta^{sex}_{c[i]}\right)$$

The log transformation is used to convert odds to log of odds. This is yet another monotone transformation. To put it another way, the higher the odds, the higher the log of odds, and vice versa.

Interpretation for the coefficient:

- β_0 represents the intercept of the model and is the log of odds of voting for the baseline group(for example, in Liberal Party's model, the baseline group is Alberta female respondents who hold an "Above the bachelor's level" degree)when the respondent is at a certain age.

- β^{age} represents the average difference in log odds of voting for the chosen Party (Liberal/NDP/Conservative) between different groups when for a particular sex, province and education.

- $\beta^{educ}_{a[i]}$ represents the average difference in log odds of voting for the chosen Party (Liberal/NDP/Conservative) between two groups(the iˆth group of education level and baseline group) for a certain age, sex and province. The notation $a[i]$ refers to the education-group a to which respondent i belongs.

- $\beta^{province}_{b[i]}$ represents the average difference in log odds of voting for the Party (Liberal/NDP/Conservative) between two groups (the iˆth group of province-level and Alberta province) for a certain age, sex, and education. The notation $b[i]$ refers to the province-group b to which respondent i belongs.

- $\beta_{c[i]}^{sex}$ represents the average difference in log odds of voting for the Party (Liberal/NDP/Conservative) between two groups (Male and Female) for a certain age, education, and province. The notation $c[i]$ refers to the education-group c to which respondent i belongs.

Model Assumptions

Now we list the assumptions for the logistic regression.And we will check them later.

1. Binary dependent variable

In our case, the vote choice(whether to vote Liberal Party) is coded with 1-"Yes,I will vote Liberal Party" and 0-"No, I will not Liberal Party." Therefore, this assumption is checked.

2. Independent observations

This criteria is met since each observation is made by a respondent who is above the age of 18 and lives in Canada, and the same respondent will not be surveyed twice.

3. Linearity to the logit for continuous predictor variables

The model assumes the log odds of the event in interest is linearly related to its predictors. As we only have one continuous variable,therefore, this assumption is checked.

4. No multi-collinearity among the independent variables

There should be no obvious correlations between the covariates of interests.

5. Large sample size

A large data collection with frequent measurements is required to run a logistic regression model. More than 20000 people participated in the 2017 GSS survey.

*Note:A linear relationship between the result variables and their predictors is not required for logistic regressions. As a result, the residuals do not need to follow a normal distribution; in other words, the data is not subject to variance homogeneity.

Model Check

1. Convergence in Logistic Regression

Likelihood maximization algorithm to converge would fail for estimating logistic regression model, which is indicated with non-existence under our context (Allison, 2008).

2. Multicolinearity Issues

Multicolinearity refers to the situation where the predictor variables in the logistic regression model are highly correlated. To assess this potential issue, variance inflation factors were computed using R, the result indicates none of the predictors in the model are highly correlated by the rule of thumb (VIF's >5).

3. Practical rational

In our case, the vote choice(whether to vote Liberal Party) is binary and we would like to use the survey data predict the possibility of the voting preference for Liberal Party in different groups (of sex, age and province) in the census data.So based on the definition of logistic regression,our choice is correct.And what's more,the previous data transformation and data cleaning process have matched the variable types in survey and census.In conclusion,choosing logistic regression can be justified.

Post-Stratification

Poststratification process:

Description of Poststratification process:

Table 6: predicted voting rate for Liberal Party of 2025

liberal_predict
0.2219591

Table 7: predicted voting rate for Conservative Party of 2025

conservative_predict
0.2568775

Firstly, data can be partitioned into thousands of demographic cells. Each cell is the combination of attributes that we are interested in. Then, we get estimated response variables for each cell. Finally, weighting each cell's relative proportion in the population when aggregating the cell-level estimates to create a population-level estimate.(STA304, 2021)

Usefulness of Poststratification process:

Post stratification is often used when a simple random sample does not reflect the distribution of some known variable in the population. In our report of Canadian Election Study, the survey data does not show the distribution of some variables in the census data which is the Canadian Election of Study of 2019. In this case, we place the observations into layers. We could essentially predict the result for each layer that is categorized based on those 4 independent variables that we are interested in. (Glasgow, 2005)

We chose 4 independent variables that we are interested in, which are sex, age, province, and educational level. We chose "age" because it is likely to influence voter outcome because people with older age were more likely to vote conclude from the past election, For instance, In 2011, with turnout rates increasing from 70% among 45- to 54-year-olds to a peak of 82% among those age 65 to 74.(Sharanjit & Sébastien, 2012). Then, we chose "Education" because the link between education and voting is also well known. For "sex", we chose it because voting rate differs among different sex.(Sharanjit & Sébastien, 2012) Regionally speaking, peopel who live in the area of Ontario and Canadian Atlantic are more willing to vote.(Sharanjit & Sébastien, 2012) We do not choose other variables since the variable levels dramatically differ among the survey data and census data , for instance, the question of "which party represents your views best?" is not in the census data.

In this report, we build logistics regression model by using survey data which is the Canadian Election Study of 2019 collected by phone call, we categorized model into different cells by age, sex, education level and province. After building the logistic regression model, we estimate the voting rates of Liberal Party, Conservative Party and New Democratic Party for each cell. Then, we found the weighted average of estimations from all possible combinations of attributes.(Wikipedia contributors, 2021) Then, we sum the weighted average of estimations of all possible combinations of age, sex, education and province and divided by the total number of voters to get population-level estimates for these 3 parties..

After computing the population-level estimate from table 7 above, we forecast that there will be 0.2219591 of the voting rate for Liberal Party of 2025. From table 8 above, we predict that there will be 0.2568775 of the voting rate for Conservative Party of 2025. From table 9, we estimate that there will be 0.09942015 of the voting rate for New Democratic Party of 2025. In this case, we could overcome the difference of the distribution of age, sex, education, province between survey data and population data. Moreover, we could effectively increase the influence of the group with larger population size toward the estimates' results.(STA304, 2021)

In order to estimate the proportion of voters who support Liberal Party, we derive mathematical model here.

Table 8: predicted voting rate for New Democratic Party of 2025

ND_predict
0.0994202

Table 9: Cells with highest predicted voting rate for Liberal Party of 2025

age	sex	education	province	estimate
80	Female	Above the bachelor's level	Ontario	0.4461590
79	Female	Above the bachelor's level	Ontario	0.4442960
78	Female	Above the bachelor's level	Ontario	0.4424346
77	Female	Above the bachelor's level	Ontario	0.4405747
75	Female	Above the bachelor's level	Ontario	0.4368601

Table 10: Cells with lowest predicted voting rate for Liberal Party of 2025

age	sex	education	province	estimate
16	Male	Less than high school diploma	Alberta	0.0502861
17	Male	Less than high school diploma	Alberta	0.0506475
18	Male	Less than high school diploma	Alberta	0.0510114
19	Male	Less than high school diploma	Alberta	0.0513778
20	Male	Less than high school diploma	Alberta	0.0517466

Where

$\widehat{y}_j$ is the estimated proportion of the voters who support a certain party in j^{th} cell.

N_j is the population size of the jth cell based off demographics.

$\hat{y}^{PS}$ is the population-level estimate which calculated by the sum of weighted average of each cell divides the total number of the population size. It represents the predicted probability of the event of voting rate for a certain party.

To predict how the entire population will vote, we derive the following mathematical formula:

$$\hat{y}^{PS} = \frac{\sum N_j \widehat{y}_j}{\sum N_j}$$

All analysis for this report was programmed using `R version 4.0.2`.

Results

Liberal Party

After extracting 5 tops estimated voting rate of cells of Liberal Party, from table 10, we see that females who live in Ontario with education level which is above bachelor associated with age 80, 79, 78, 77, 75 have estimated voting rate 0.4461590, 0.4442960, 0.4424346, 0.4405747, 0.4368601 respectively. Thus, we conclude that people who are likely to vote for Liberal party with nearly 50 percent voting rate among the specific cell turn out to be females who live in Ontario with age above 75 associated with education level which is above the bachelor's level. In practical sense, people with higher age and higher education level would more likely to vote. It seems reasonable since in the 2019 federal elections, 67% of voters voted in Canada. If you look at turnout by age group, you can see that there are far more older voters than younger voters.(Fournier,2021) At the regional level, the figures show the strong leadership of the voting rate in

Table 11: predicted voting rate for Liberal Party of 2025

liberal_predict
0.2219591

Table 12: Cells with highest predicted voting rate for Conservative Party of 2025

age	sex	education	province	estimate
80	Male	High school diploma	Alberta	0.7236634
79	Male	High school diploma	Alberta	0.7215314
78	Male	High school diploma	Alberta	0.7193893
80	Male	College, CEGEP or other non-university certificate	Alberta	0.7193337
77	Male	High school diploma	Alberta	0.7172371

Table 13: Cells with lowest predicted voting rate for Consertive Party of 2025

age	sex	education	province	estimate
26	Female	Above the bachelor's level	Quebec	0.0411213
28	Female	Above the bachelor's level	Quebec	0.0419683
29	Female	Above the bachelor's level	Quebec	0.0423981
30	Female	Above the bachelor's level	Quebec	0.0428320
31	Female	Above the bachelor's level	Quebec	0.0432702

Ontario and the Canadian Atlantic which makes sense that the people who live in Ontario would prefer to vote for Liberal Party.(Fournier,2021) Moving on to explore the education level, it turns out that the "positive" impact of education on voting is much stronger among young people.For example, among people aged 25 to 34, the difference in participation between those with at least a bachelor's degree and those with less than secondary education was 42 percentage points.Between college graduates and those with below-average education, There was a significant difference.(Sharanjit & Sébastien, 2012)

From table 11, we extracted 5 lowest estimated voting rate of cells among all cells, we see that males who live in Alberta with education level which is less than high school diploma associated with age 16, 17, 18, 19, 20 have estimated voting rate 0.05028607, 0.05064751, 0.05101141, 0.05137778, 0.05174663 respectively. Thus, we conclude that people who are unlikely to vote for Liberal party with approximately less than 5 percent voting rate among the specific cell turn out to be males who live in Alberta with age under 20 associated with education level which is less than high school diploma.This makes sense since voter turnout for those with higher education was 78%, while those with less than secondary education were less than 60%.(Sharanjit & Sébastien, 2012) In practical sense, people with lower educational level would unlikely to vote. Regionally speaking, people who live in Ontario or Canadian Atlantic would prefer to vote rather than people who live in Alberta.(Fournier, 2021) For population-level estimate, from table 12, we forecast that there will be 0.2219591 of the voting rate for Liberal Party of 2025.

Conservative Party

After extracting 5 tops estimated voting rate of cells, from table 13, we see that males who live in Alberta with education level of High school Diploma for age 77, 78, 79, 80 have predicted voting rate of 0.7172371, 0.7193893, 0.7215314, 0.7236634 and for age 80 with non-university certificate have voting rate of 0.719337. Thus, we conclude that people who are likely to vote for Conservative party with nearly 72 percent voting rate among the specific cell turn out to be males who live in Alberta with age above 75 associated with education level which is between College and High school. In practical sense, people with higher age. It seems reasonable since in the 2019 federal elections, 67% of voters voted in Canada. If you look at turnout

Table 14: predicted voting rate for Conservative Party of 2025

conservative_predict
0.2568775

Table 15: Cells with highest predicted voting rate for New Democratic Party of 2025

age	sex	education	province	estimate
19	Female	Above the bachelor's level	British Columbia	0.4203573
18	Female	High school diploma	Newfoundland and Labrador	0.4070896
19	Female	High school diploma	Newfoundland and Labrador	0.3991129
23	Female	University certificate or diploma below the bachelor's level	Newfoundland and Labrador	0.3914695
18	Female	High school diploma	British Columbia	0.3913631

Table 16: Cells with lowest predicted voting rate for New Democratic Party of 2025

age	sex	education	province	estimate
80	Male	Less than high school diploma	New Brunswick	0.0049648
79	Male	Less than high school diploma	New Brunswick	0.0051313
78	Male	Less than high school diploma	New Brunswick	0.0053033
77	Male	Less than high school diploma	New Brunswick	0.0054811
76	Male	Less than high school diploma	New Brunswick	0.0056648

by age group, you can see that there are far more older voters than younger voters.(Fournier,2021) At the regional level, the figures show the strong leadership of the voting rate in Alberta which makes sense that the people who live in Alberta would prefer to vote for Conservative as we have seen in the past few weeks, federal election campaigns are unpredictable. But one thing remains (almost) always true: Albertans vote for Conservatives.(Jack, 2021))

From table 14, we extracted 5 lowest estimated voting rate of cells among all cells, we see that males who live in Quebec with education level which is above the bachelor's level associated with age 26, 28, 29, 30, 31 have estimated voting rate 0.0411213, 0.0419683, 0.0423981, 0.0428320, 0.0432702 respectively. Thus, we conclude that people who are unlikely to vote for Conservative Party with approximately less than 5 percent voting rate among the specific cell turn out to be females who live in Quebec with age from 26 to 31 associated with education level which is above bachelor's level.This makes sense since voter turnout for those with higher education was 78%, while those with less than secondary education were less than 60%.(Sharanjit & Sébastien, 2012) In practical sense, people with lower educational level would unlikely to vote. Regionally speaking, people who live in Ontario or Canadian Atlantic would prefer to vote rather than people who live in Alberta.(Fournier, 2021) For population-level estimate, from table 15, we forecast that there will be 0.2568775 of the voting rate for Conservative Party of 2025.

New Democratic Party

By extracting 5 tops estimated voting rate of cells, from table 16, we see that females with age 18 who live in British Columbia, Newfoundland and Labrador have estimate voting rate of 0.3913631 and 0.4203573 for Conservative Party. Additionally, we notice that Females who live in Newfoundland and Labrador associated with age 19 and 23 have estimate voting rate of 0.3991129 and 0.3914695 for Conservative Party. What's more, we observed that for females with age 19 who live in British Columbia with above the bachelor's education level have estimate voting rate of 0.4203573. Thus, we conclude that people who are likely to vote for Conservative party with nearly 40 percent voting rate among the specific cell turn out to be females who live in Britsh Columbia or Newfoundland and Labrador with age from 19-23 associated with education level

Table 17: predicted voting rate for New Democratic Party of 2025

ND_predict
0.0994202

Table 18: Coefficient Table of Logistic Regression Model for Liberal Party

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.2701501	0.2625237	-8.6474088	0.0000000
age	0.0075426	0.0023500	3.2096127	0.0013291
sexMale	-0.0494219	0.0781253	-0.6325978	0.5269963
educationBachelor's degree	-0.1153694	0.1119870	-1.0302039	0.3029143
educationCollege, CEGEP or other non-university certificate	-0.6884149	0.1191198	-5.7791789	0.0000000
educationHigh school diploma	-0.5613032	0.1421561	-3.9484989	0.0000786
educationLess than high school diploma	-0.7395418	0.1999179	-3.6992267	0.0002163
educationUniversity certificate or diploma below the bachelor's level	-0.3091504	0.1566352	-1.9736967	0.0484162
provinceBritish Columbia	0.8254562	0.2293939	3.5984232	0.0003202
provinceManitoba	0.9286445	0.2617245	3.5481756	0.0003879
provinceNew Brunswick	1.0682151	0.2705990	3.9475953	0.0000789
provinceNewfoundland and Labrador	1.3732222	0.2657300	5.1677347	0.0000002
provinceNova Scotia	1.3028598	0.2662000	4.8942884	0.0000010
provinceOntario	1.4505421	0.2243193	6.4664173	0.0000000
provincePrince Edward Island	1.3528993	0.2646229	5.1125558	0.0000003
provinceQuebec	1.1235475	0.2271905	4.9453989	0.0000008
provinceSaskatchewan	0.2205188	0.2911725	0.7573476	0.4488416

which is between High school and Bachelor's level. It seems resonable since At the regional level, the figures show the strong leadership of British Columbia, which makes sense that the New Democratic Party of British Columbia is a social-democratic provincial political party in British Columbia.(Wikipedia, 2021)

From table 17, we extracted 5 lowest estimated voting rate of cells among all cells, we see that males who live in New Brunswick with education level which is less than high school diploma associated with age 76, 77, 78, 79, 80 have estimated voting rate 0.0056648, 0.0054811, 0.0053033, 0.0051313, 0.0049648 respectively. Thus, we conclude that people who are unlikely to vote for Conservative Party with approximately less than 0.5 percent voting rate among the specific cell turn out to be males who live in New Brunswick with age from 76 to 80 associated with education level which is less than high school diploma.This makes sense since voter turnout for those with higher education was 78%, while those with less than secondary education were less than 60%.(Sharanjit & Sébastien, 2012) In practical sense, people with lower educational level would unlikely to vote. Regionally speaking, people who live near British Columbia would prefer to vote rather than people who live in the area which is far away from British Columbia. Since New Democratic Party is a social-democratic provincial political party in British Columbia. For population-level estimate, from table 18, we forecast that there will be 0.09942015 of the voting rate for Conservative Party of 2025.

Combined with these 3 results, we could predict that among these 3 parities, Conservative Party would more likely to have a higher voting rate in 2025 since the estimate is higher than the estimates of Liberal party and New Democratic party.

Interpreting Logsitic Results

1. For Liberal Party:

According to coefficient table above, we could obtain the "baseline probability" from our voting preference for Liberal Party estimation model. The baseline probability is the estimated voting preference for Liberal Party probability of Alberta female respondents who held an "Above the bachelor's level" degree. We estimate this probability to be 9.36% . We are 95% confident that this probability would be between 5.72% and 14.55%.

The age effect on the odds of voting Liberal Party was statistically significant: aging post a positive influence on the odds of having children. As the respondents grow older, they are more likely to vote Liberal Party. For one year of age older than the average (i.e. 51-year-old), the respondent's log odds of voting for Liberal Party would increase by 0.75% as the respondent is one year older than the average.

Table 19: Coefficient Table of Logistic Regression Model for Conservatice Party

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.1747207	0.2101282	-5.590495	0.0000000
age	0.0106363	0.0023484	4.529261	0.0000059
sexMale	0.5984312	0.0815187	7.341026	0.0000000
educationBachelor's degree	0.3267861	0.1358628	2.405266	0.0161607
educationCollege, CEGEP or other non-university certificate	0.6665412	0.1318792	5.054180	0.0000004
educationHigh school diploma	0.6880892	0.1489442	4.619779	0.0000038
educationLess than high school diploma	0.6233362	0.1920152	3.246286	0.0011692
educationUniversity certificate or diploma below the bachelor's level	0.4083902	0.1706629	2.392964	0.0167129
provinceBritish Columbia	-1.4378712	0.1509460	-9.525731	0.0000000
provinceManitoba	-0.6956992	0.1801840	-3.861049	0.0001129
provinceNew Brunswick	-1.3092063	0.2081821	-6.288756	0.0000000
provinceNewfoundland and Labrador	-1.8595086	0.2296312	-8.097805	0.0000000
provinceNova Scotia	-1.4951321	0.2170161	-6.889499	0.0000000
provinceOntario	-1.3668883	0.1505973	-9.076447	0.0000000
provincePrince Edward Island	-1.5325408	0.2180410	-7.028682	0.0000000
provinceQuebec	-2.2510620	0.1689080	-13.327145	0.0000000
provinceSaskatchewan	-0.4052559	0.1777288	-2.280193	0.0225962

Table 20: Coefficient Table of Logistic Regression Model for New Democratic Party

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.6400860	0.3131118	-2.0442728	0.0409266
age	-0.0331530	0.0034162	-9.7047818	0.0000000
sexMale	-0.4060272	0.1089430	-3.7269709	0.0001938
educationBachelor's degree	-0.2790307	0.1680717	-1.6601882	0.0968766
educationCollege, CEGEP or other non-university certificate	-0.1929505	0.1663696	-1.1597703	0.2461423
educationHigh school diploma	-0.1534321	0.1970444	-0.7786677	0.4361755
educationLess than high school diploma	-0.5400736	0.3208750	-1.6831275	0.0923504
educationUniversity certificate or diploma below the bachelor's level	-0.0527964	0.2167117	-0.2436251	0.8075212
provinceBritish Columbia	0.9486861	0.2572630	3.6876126	0.0002264
provinceManitoba	0.4816882	0.3156686	1.5259303	0.1270272
provinceNew Brunswick	-1.0619809	0.5127734	-2.0710532	0.0383538
provinceNewfoundland and Labrador	1.0142622	0.3113761	3.2573540	0.0011246
provinceNova Scotia	0.5410326	0.3270241	1.6544117	0.0980439
provinceOntario	0.6287750	0.2616241	2.4033531	0.0162455
provincePrince Edward Island	-0.6751316	0.4535617	-1.4885109	0.1366162
provinceQuebec	-0.1413029	0.2773278	-0.5095158	0.6103907
provinceSaskatchewan	0.4767364	0.3092510	1.5415839	0.1231747

Now let's focus on the sex difference. Estimated odds are not considered statistically significant because the p-level is so small.(p<0.05).

For respondents living in various provinces, the odds of their voting preference would differ as well. Holding all the other variables constant,British Columbia residents have significantly 0.83 higher log odds of voting Liberal Party than the ones living in Alberta province. Our regression model also reveals that this increased log odds would be 0.93 in voting Liberal party odds when comparing Manitoba residents with Alberta residents. People living in New Brunswick's log odds of voting Liberal party would be 1.07 higher than Alberta residents.People living in Newfoundland and Labrador's log odds of voting Liberal party would be 1.37 higher than Alberta residents. People living in Nova Scotia's log odds of voting Liberal party would be 1.30 higher than Alberta residents. People living in Ontario's log odds of voting Liberal party would be 1.45 higher than Alberta residents. People living in Prince Edward Island's log odds of voting Liberal party would be 1.35 higher than Alberta residents.People living in Quebec's log odds of voting Liberal party would be 1.12 higher than Alberta residents.But for people living in Saskatchewan,the estimated log odds are not considered statistically significantly different Alberta residents compared to because the p-level is so small.(p<0.05).

Last, for the respondents that have different educational backgrounds, the odds of their voting choice on Liberal Party would vary. For university graduates who hold "College, CEGEP or other non-university certificate" degrees, they tend to have 0.69 times lower log odds of supporting Liberal Party than people who have above Bachelor's degree(i.e,Matser's degree or Professional degree or doctorate). While the respondents holding only high school diploma tend to have 0.56 lower log odds of voting for Liberal Party.The respondents holding less than high school diploma below the bachelor's level tend to have 0.74 lower log odds of voting for Liberal Party compared to the above Bachelor's degree group.The respondents holding University certificate or diploma below the bachelor's level tend to have 0.31 lower log odds of voting for Liberal Party compared to the above Bachelor's degree group.However, there is no enough statistical evidence supporting the hypothesis that the respondents with Bachelor's degree would have different odds in voting preference with the above Bachelor's degree group (p's >0.05). Large p-values indicate that people who hold a Bachelor's degree or above have a similar voting preference on Liberal Party.

2. For Conservative Party:

According to coefficient table above, we could obtain the "baseline probability" from our voting preference for Conservative Party estimation model. The baseline probability is the estimated voting preference for Conservative Party probability of Alberta female respondents who held an "Above the bachelor's level" degree. We estimate this probability to be 23.60%. We are 95% confident that this probability would be between 16.95% and 31.76%.

The age effect on the odds of voting Conservative Party was statistically significant: aging post a positive influence on the odds of having children. As the respondents grow older, they are more likely to vote for Conservative Party. For one year of age older than the average (i.e. 51-year-old), the respondent's log odds of voting for Conservative Party would increase by 0.01 as the respondent is one year older than the average.

Now let's focus on the sex difference. When the gender of the respondent is male, the log odds of voting for Conservative Party would increase approximately 0.60 in comparison with female respondents with the same age, education and province level. The estimated odds are considered to be statistically supported (p<0.05).

For respondents living in various provinces, the odds of their voting preference would differ as well. Holding all the other variables constant, British Columbia residents have significantly −1.44 lower log odds of voting Conservative Party than the ones living in Alberta province. Our regression model also reveals that there would be −0.70 in voting Conservative party's log odds when comparing Manitoba residents with Alberta residents. People living in New Brunswick's log odds of voting Conservative party would be 1.31 lower than Alberta residents. People living in Newfoundland and Labrador's log odds of voting Conservative party would be 1.86 lower than Alberta residents. People living in Nova Scotia's log odds of voting Conservative party would be 1.50 lower than Alberta residents. People living in Ontario's log odds of voting Conservative party would be 1.37 lower than Alberta residents. People living in Prince Edward Island's log odds of voting Conservative party would be 1.53 lower than Alberta residents.People living in Quebec's log odds of voting Conservative party would be 2.25 lower than Alberta residents.People living in Saskatchewans log odds of

voting Conservative party would be 0.41 lower than Alberta residents.

Last, for the respondents that have different educational backgrounds, the odds of their voting choice on Conservative Party would vary. For university graduates who hold "College, CEGEP or other non-university certificate" degrees, they tend to have 0.67 times higher log odds of supporting Conservative Party than people who have above Bachelor's degree(i.e,Matser's degree or Professional degree or doctorate). While the respondents holding only high school diploma tend to have 0.69 higher log odds of voting for Conservative Party. The respondents holding less than high school diploma tend to have 0.62 higher log odds of voting for Conservative Party compared to the above Bachelor's degree group. The respondents holding University certificate or diploma below the bachelor's diploma tend to have 0.41 higher log odds of voting for Conservative Party compared to the above Bachelor's degree group. For the respondents holding Bachelor's Degree tend to have 0.32 higher log odds of voting for Conservative Party.

3. For New Democratic Party:

According to coefficient table above, we could obtain the "baseline probability" from our voting preference for New Democratic Party estimation model. The baseline probability is the estimated voting preference for New Democratic Party probability of Alberta female respondents who held an "Above the bachelor's level" degree. We estimate this log odds to be 34.52% . We are 95% confident that this probability would be between 16.95% and 31.76%.

The age effect on the odds of voting New Democratic Party was statistically significant: aging post a positive influence on the odds of having children. As the respondents grow older, they are more likely to vote for New Democratic Party. For one year of age older than the average (i.e. 51-year-old), the respondent's log odds of voting for New Democratic Party would decrease by 0.03 as the respondent is one year older than the average.

Now let's focus on the sex difference. When the gender of the respondent is male, the log odds of voting for New Democratic Party would decrease approximately -0.41 in comparison with female respondents with the same age, education and province level. The estimated log odds are considered to be statistically supported (p<0.05).

For respondents living in various provinces, the odds of their voting preference would differ as well. Holding all the other variables constant,people living in New Brunswick's log odds of voting New Democratic Party would be 1.06 lower than Alberta residents.People living in Newfoundland and Labrador's log odds of voting New Democratic Party would be 1.01 higher than Alberta residents.People living in Nova Scotia's log odds of voting New Democratic Party party would be 0.63 higher than Alberta residents.People living in Ontario's log odds of voting New Democratic Party would be 20.31% higher than Alberta residents.People living in British Columbia's log odds of voting New Democratic Party would be 0.95 higher than Alberta residents.For the other provinces,the estimated odds are considered not to be statistically supported (p>0.05).

However, there is no enough statistical evidence supporting the hypothesis that the respondents have different educational backgrounds would have different odds in their voting choice on New Democratic Party(p's >0.05). Large p-values indicate that the educational backgrounds did not have a significant contribution in estimating the odds of voting for New Democratic Party.

Conclusions

At the beginning of our research, we made several hypotheses on whether demographic information of voters, including sex, age, educational level, and region living in, impact their voting decision to certain parties in the election. To test these hypotheses, we use the methods of logistic regression. The dependent variable is the odds of voting for parties, while the independent variables are sex, age, educational level, and regions of voters whose information was recorded in GSS.

From the Result Part, we found out the following conclusions:

1. When holding other variables fixed, older voters tend to vote for Liberal Party. Also, when voters grow older, they are more likely to vote for the Conservative Party when keeping other variables unchanged. Both of the results are of statistical significance.

2. When holding other variables fixed, we cannot find a difference in the probability of voting for the Liberal Party between males and females. However, we found a statistically significant difference between males and females when voting for Conservative Party. On average, male voters have a higher probability of voting for the Conservative Party.

3. When holding other variables fixed, voters of lower educational level tend not to vote for Liberal Party, compared with those who possess degrees above bachelor. In contrast, when holding other variables fixed, voters of lower educational level has increased odds of voting for the Conservative Party with statistical significance, compared with the voters who possess degrees above bachelor.

4. When holding other variables fixed, voters living in different regions (provinces) indeed have a significantly different likelihood of voting for the Liberal Party. Similarly, voters residing in the various areas present quite different odds of voting for the Conservative Party.

5. Based on the logistic regression model, we further applied the post-stratification method to predict the probability of each party's winning the future election. The likelihood of the Liberal Party's winning the vote is 22.20%, the possibility for Conservative Party is 25.69%, while the probability for New Democrat is 9.94%. Therefore, among these three parties who possess the most incredible popularity in the past few years, The Conservative Party has the highest probability of winning the election in 2025.

Based on the findings above, our team would make suggestions to Liberal Party so that the party's political leader could make better preparations for the future election in 2025 and gain more votes from other groups of people. Firstly, for the sex difference, our conclusion is opposite to the outcome found by Elisabeth Gidengil in 2006, so we consider a part of male or female voters is neither practical nor moral. Secondly, for the age difference in voting preference, Liberal Party could design slogans that contain messages more attractive and acceptable to the young in the campaign. For example, they could merge fashion and young-like elements into the campaign event and develop more effective broadcasts on social media, like ins or Twitter. Putting more emphasis on running the social media account enables the party to improve its friendliness with the young voters and observes the actual reaction of citizens towards social issues or policies. For the difference in educational level, we suggest that Liberal Party initiate more campaign events and address more public speeches in factories or low-income communities to get closer to voters with lower academic voters. More importantly, as the party in power, Liberal Party might need to consider improving social welfare to improve the living conditions of people with low educational levels. Last but not least, considering the different probability of voting for the Liberal Party based on voters' geographic traits, the party needs to spend more effort and capital in initiating political events and making advertisements in certain areas. But for the region problem, the party should also consider the historical factors that make certain provinces loyal to a party.

Though we perform qualified logistic regression based on demographic variables in our regression model, there are still weaknesses and limitations in both data analysis and political concerns in reality.

1. Omitted Variables

In the regression model, we mainly emphasize analyzing the impact of age, sex, and education on voting decisions. However, there must be other variables affecting voters' preference in the error terms, such as voters' satisfaction in the previous policies issued by the party in power, voters' race, regions voters live in, economic policies, etc. Predicting the future election referring to the three variables in the logistic model would likely bring about inaccurate outcomes.

2. Imputation in Original Data

As mentioned in Data Part, the GSS program applied the method of data imputing to guarantee that the crucial variables of selected citizens are completed if they were missed in the original data. The latter assigned values concerning the respondents' answers to other parts of the questionnaire. Although such a method is extremely helpful in maintaining the completeness of original data and creating great convenience for data analysts, it also leads to the potential risk of error information. There will be even more significant errors and distortion in our regression model if systematic bias in this imputed method.

3. Stratified Data Analysis

When describing the data, we pointed out that in the 2017 GSS, the survey divided the target population into 27 strata regarding their geological areas. However, in performing the regression model, we failed to stratify the data to be distorted compared with the original data.

4. Limitations in Economic Policy

Although our model found that older voters are more likely to vote for Liberal Party, the party in power could not issue the unfair policy to please its supporters. Specifically, the government should not impose a higher tax on young people to offer more generous pension and insurance to the old so that the party could possess higher support. Furthermore, such policies might lead to a chain reaction in the whole economy, hurting the supporters on the contrary. Under such circumstances, even though figuring out voters' preferences based on demographic data, the party in power could not freely adopt corresponding measures to attract supporters regarding the restrictions of economic policies.

5. Weakness of Variable

As we include a numerical variable-age in our model, the post-stratification result thus becomes less practical as it will make low counts and technically infeasible bins in the post-stratification process. To deal with this problem, we look at the result in a general way. As discussed in the Result Section, we look at the highest and lowest five estimated voting rate groups. So in further analysis, we can convert the continuous age into age groups so we can make predictions on different age groups.

Future Works

For related research in the future, we team gave the following suggestions and potential improvements:

1. Update DataBase: to predict the election in 2025, we built up the models based on the census data in 2017 and survey data in 2019, both of which are outdated, so more updated databases are necessarily required to perform a more accurate prediction. Moreover, in dealing with original data, we could imitate the methods applied in 2019 CES that referred to another database and merge the data in the personal fiscal file, which could better reflect the actual value.

2. Consider More Variables: in the weakness part, we mentioned that there must be omitted variables in the error terms. Therefore, future study needs to consider more related variables, such as the voters' race, to build up a more comprehensive model.

In this research, with the data support from the 2017GGS and 2019CES, we team succeeded in performing an empirical study to figure out the impact of demographic data on voters' probability of voting for the Liberal Party, the party in power right now in Canada. The output of the logistic regression demonstrates some relationship between those variables and based on that finding. We further apply them to predict the performance of the Liberal Party in the election of 2025. However, some limitations in the data study still need to be fixed and improved in future research. We hope our research could benefit prospective study in inspiring how to build up a comprehensive model in predicting the outcome of the political election, a crucial political event throughout the world.

Bibliography

1. "Canada Election 2021: Full Results." The Guardian, Guardian News and Media, 21 Sept. 2021,www.theguardian.com/world/2021/sep/21/canada-general-election-2021-full-results.

2. Gidengil, Elisabeth & Everitt, Joanna & Fournier, Patrick. (2006). sex and Vote Choice in the 2006 Canadian Election. https://www.researchgate.net/publication/253375392_sex_and_Vote_Choice_in_the_2006_Canadian_Election.

3. GUNTERMANN, ERIC. "Who Voted for Whom? Vote Choice in the 2017 Catalan Election." Making Electoral Democracy Work, January 5, 2018. https://electoraldemocracy.com/who-voted-for-whom-vote-choice-in-the-2017-catalan-election-2713/.

4. Fournier, P. J. (2021, April 25). 338Canada: The Liberals Are Winning Over Older—normally Conservative—voters - Macleans.ca. Macleans.ca. https://www.macleans.ca/politics/ottawa/338canad

a-the-liberals-are-winning-over-older-normally-conservative-voters/.

5. Study finds those with graduate education are far more liberal than peers. (n.d.). Retrieved November 5, 2021, from https://www.insidehighered.com/news/2016/04/27/study-finds-those-graduate-education-are-far-more-liberal-peers.

6. Philippe J. Fournier.(2021) "338Canada: The Liberals are winning over older—normally Conservative—voters" https://www.macleans.ca/politics/ottawa/338canada-the-liberals-are-winning-over-older-normally-conservative-voters/

7. Factors Associated With Voting. (n.d.). Factors associated with voting. https://www150.statcan.gc.ca /n1/pub/75-001-x/2012001/article/11629-eng.htm.

8. 2019 Canadian Federal Election - Wikipedia. (2020, June 22). 2019 Canadian federal election - Wikipedia. https://en.wikipedia.org/wiki/2019_Canadian_federal_election.

9. Liberal Party Of Canada - Wikipedia. (1861, January 1). Liberal Party of Canada - Wikipedia. https://en.wikipedia.org/wiki/Liberal_Party_of_Canada.

10. Gregory Jack, I. P. A. (2021, September 6). Commentary: Conservatives remain strong in Alberta, but the NDP is the emerging 'progressive' choice. Global News. Retrieved November 6, 2021, from https://globalnews.ca/news/8165875/conservatives-remain-strong-in-alberta-but-the-ndp-is-the-emerging-progressive-choice/.

11. Wikimedia Foundation. (2021, October 31). British Columbia New Democratic Party. Wikipedia. Retrieved November 6, 2021, from https://en.wikipedia.org/wiki/British_Columbia_New_Democratic_Party.