Forecasting the Next Federal Election: A Logistic Regression Model

STA304 - Fall 2023 -Assignment 2

GROUP 90: Hanzhong Li, Xiaodong Lu, Jiaxuan Tian, Ruining Liu

November 22nd, 2023

Part 1: Introduction

Research Goal

As a method for the public to choose a political leader, election is an important component in the pursuit of democracy and public well-being. Predicting the result of an election can shed light on future directions of socioeconomic development, diplomatic affairs, or more down-to-earth aspects such as the public's lifestyles and quality of life.

The goal of this report is to predict the overall popular vote of the next Canadian federal election. This report will showcase a logistic regression model to forecast the winning probabilities of the three most popular parties in Canada (Liberal, Conservative, and NDP) and compare them to conclude which party is most likely to win the next Canadian election. From our model, we aim to show a relationship between our selected predictors and the outcome of the party voted.

By looking at federal elections for the past decade, we notice that the past elections have favoured the Liberal Party (CTV News, 2021). Our hypothesis, therefore, is that for the next Canadian federal election, the Liberal Party is more likely to win.

Importance of the Analysis

The importance of this analysis applies to both the candidates and the public. This report, providing information for potential outcomes, will give the candidates opportunities to understand the likelihood of winning in different regions or demographics, adjust their political strategies, align to the public's preferences and values, and efficiently allocate resources for campaign and media coverage respectively. Regarding the public, had the prediction in this report been valid, the citizens would have the opportunity to anticipate certain policies to take place, and be able to plan ahead for those changes. Moreover, in diplomatic terms, a prediction of the future governing party could also reshape and potentially improve international relations based on the expected election outcomes.

Part 2: Data

Two datasets are incorporated in this report. The first one is the 2021 Canadian Election Study (CES) dataset, "a rich source of data on Canadians' political behaviour and attitudes, measuring preferences on key political issues" (CES, 2021). The 2021 CES dataset (Survey data) will be used to fit the logistic regression model. The second is Canada's General Social Survey (GSS) dataset, a source of "a series of independent, annual, cross-sectional surveys" (Statistics Canada, 2017). The GSS dataset (Census data) will be used to compute our forecast and draw conclusions on the voting result.

Variables (Before Cleaning)

In the regression model, we selected 6 variables to be predictors of the voting outcome: Age, Province, Citizenship Status, Family Income, Marital status, and Sex. The outcome variable is the voting result in the Survey Data.

Census Data (20,602 observations):

1. Age (numerical): numerical value of the voter's age

2. Province (categorical, non-binary): the province of where the voter is from, "Alberta", "British Columbia", "Manitoba", "New Brunswick", "Newfoundland and Labrador", "Nova Scotia", "Ontario", "Prince Edward Island", "Quebec", and "Saskatchewan", respectively

3. Citizenship Status (categorical, non-binary): the citizenship of the voter, "By birth" (Canadian citizen), "By naturalization" (Permanent Resident), "Don't know", and "NA", respectively

4. Family Income (categorical, non-binary): the family income, as an interval, of the voter

5. If Married (categorical, Yes/No): whether the voter has been married

6. Sex (categorical, Yes/No): whether the voter is male (we assume the answer "No" to be female)

Survey Data (20,968 observations):

1. Age (numerical): numerical value of the voter's age

2. Region (categorical, non-binary): the province of where the voter is from, "Atlantic", "NewBrunswick", "Newfoundland", "NovaScotia", "PrinceEdwardIsland", "Ontario", "Quebec", "Territories" (Northwest Territories, Nunavut, Yukon), "West" (Alberta, British Columbia, Manitoba, Saskatchewan), respectively

3. Citizenship Status (categorical, non-binary): "Canadian Citizen", "Permanent Resident", and "Other", respectively

4. Family Income (numerical): numerical value of the family income

5. Marital status (categorical, non-binary): "married", "living with a partner", "divorced", "separated", "widowed", "never married", and "don't know or prefer not to say", respectively

6. Gender (categorical, non-binary): "men", "women", "Non-Binary", and "Another gender", respectively

7. Voting Outcome (categorical, non-binary): "Liberal Party", "Conservative Party", "NDP", "Bloc Québécois", "Green Party", "Another party", "Don't know/ Prefer not to answer", respectively

8. Consent (categorical, Yes/No): whether the survey taker consents to participate

Cleaning process

Considering that imputation (replacing missing data with substituted values, such as with the mean) may cause more deviations in the analysis, we applied deletion to all unclear and missing data.

We will showcase the complete cleaning procedure in steps.

1. Census data: select the 6 chosen variables

2. Census data: delete all observations with missing values

3. Census data: filtered Age: bigger than or equal to 18, eligible for voting

4. Census data: filtered Citizenship Status that has the answer "Don't know"

5. Census data: Province: added a column to map census province to survey data region

6. Survey data: select the 6 variables, plus the variables of voting outcome (one of the three parties) and consent (whether participant consent to take part in the survey for using their information)

7. Survey data: delete all observations with missing values

8. Survey data: left only with those who voted for the three parties of interest in this report

9. Survey data: filtered Age: bigger than or equal to 18, eligible for voting

10. Survey data: filtered citizenship status: left only those are Canadian citizens

11. Survey data: filtered to leave only those who gave consent to study

12. Survey data: created a new column of marital status to map with the census dataset

13. Survey data: only has gender variable, filtered with only answers "Male" and "Female"

14. Survey data: created new column mapping numerical income value to corresponding intervals in Census data (Note in mapping procedure: in Census data, for the interval "$100,000 to $ 124,999", there is a space between "$" and "124,999")

Variables: After Cleaning

Census (19,006 observations) & Survey (3,261 observations):

1. Age

2. Province

3. Citizenship Status (not used)

4. Family Income

5. If Married

6. Sex

7. Voting Outcome (only in Survey data)

See the Appendix for more plots of the distributions of Survey data variables.

Proportion of Missingness in Variables

Table 1: Census Variables - Proportion of Missingness

age	province	citizenship	income_family	ever_married	is_male
0	0	0.0554801	0	0.0002427	0

From Table 1, we can see that there are no NAs for variables age, province, family income, and sex, which means for these variables, everyone who participated provided valid information. However, there is a percentage of missing value of 5.55% for citizenship and 0.0243% for marital status. This means among all the

participants, 5.55% of them provided invalid or no information on their citizenship, and 0.0243% for the variable of marital status.

Table 2: Survey Variables - Proportion of Missingness

vote_advance	age2	genderid2	income2	marital2	region2	citizenship2	consent2
0.7841473	0	0	0	0	0	0	0

From Table 2, we can find that there are no NAs for variables age, gender, family income, marital status, region, and consent, which means for these variables, everyone who participated in the voting provided valid information. However, there is a 78.415% percentage of missingness for the voting outcome, which means among all the participants, 78.415% of them provided invalid or no information on their voting outcome. This is a significantly high percentage, but considering that the survey is not aimed at voting and many participants are not eligible to vote, this loss of data is understandable. We will proceed with the cleaned Survey data that only counts the observations that have valid voting outcomes recorded.

Numerical Summaries

Table 3: Census Data - Numerical Summary

	MINIMUM	MAXIMUM	MEAN	MEDIAN	1ST QUANTILE	3RD QUANTILE
AGE	18	80	53.66749	55.7	39.4	67.4

Table 3 gives a numerical summary of the Census data. As age is the only numerical variable chosen, the table includes only the numerical summary of the age variable. We see that the minimum age is 18, and the maximum age is 80, indicating a large spread. The mean age is 53.66749, median at 55.7. The first quantile, or the youngest 25% bar of the Census data, is at 39.4, whereas the third quartile (where 75% of the participants' ages fall below) is at 67.4. From these statistics, it is reasonable to postulate that most of the participants are middle-aged people.

Table 4: Survey Data - Numerical Summary

	MINIMUM	MAXIMUM	MEAN	MEDIAN	1ST QUANTILE	3RD QUANTILE
AGE	18	90	55.84759	60	42	70
INCOME NUMBER	-99	1400000	87709.37197	75000	45000	120000

Table 4 presents a numerical summary of the Survey data. Similarly, the table only includes numerical variables age and income. We can see that the minimum age is 18, maximum is 90. The mean age is 55.85, and the median is 60. The interquartile range is from 42 to 70. Overall, we can roughly reach the same conclusion with the Census data: most of the participants are likely middle-aged people. In terms of income, the range is roughly 1,400,000. The distribution has a mean around 90,000, with a median of 75,000. The interquartile is around 75,000.

Part 3: Method

The method of logistic regression models the probability of a discrete outcome given input variables, or predictors (Edgar, 2017). The regression model performs classification on a binary outcome. In the context

of this report, the outcome of our logistic regression models is whether or not the voter votes for the party of interest. We plan to build three logistic models for predicting the probability that a voter would vote for the three parties of interest, respectively.

Model Assumptions

Unlike many other regression models, a logistic regression model does not require a strict linear relationship between the dependent and independent variables. Also, it does not require the error terms (residuals) to be normally distributed, and constant variance is not required.

Some requirements for logistic regression are:

1. A binary response variable;

2. Little or no collinearity among the independent variables, meaning that variables should not be too highly correlated with each other. (Statistics Solutions, 2021)

Assumption Justifications

1. The logistic regression models we built each have two outcomes, voting for the party of interest or not voting for it.

2. The selection process of the predictor variables is done with discretion, avoiding obvious correlations and aiming for a comprehensive grasp of factors that could potentially influence voting outcomes. There are both numerical and categorical variables in our selection of predictors. We will attempt to justify the absence of multicollinearity with analysis for these variables separately. Details are attached in the Appendix.

Building the Model: Procedure

Our desired response variable, voting result, has three outcomes with respect to three parties, "Liberal", "Conservative", and "NDP". Our response variable will have a value between 0 and 1, representing the winning probability.

Therefore, the three outcomes will correspond to three individual linear regression models, each predicting the winning probability of one party of interest. Here, model 1 corresponds to the Liberal Party, model 2 corresponds to the Conservative Party, and model 3 corresponds to NDP.

To successfully build these models, we started off with extra cleaning. We added a new column (cleaned, slightly modified response variable) in the survey dataset with binary outcomes, 1 and 0. Then, it is easier to carry out logistic regression model fitting. For example, when the party of interest is "Liberal", all voters who voted for "Conservative" and "NDP" will be recorded as a 0, whereas only the ones who voted for "Liberal" will be recorded as a 1.

Among our 5 variables (the variable citizenship serves as a filter, thus not a variable in our models), only age is a numerical variable. For the rest categorical variables, sex and marital status have binary outcomes, where income and region have multiple levels. To incorporate these two variables into our model, we divide them by setting up different levels in each factor.

For the variable income, we set the category "Below 25000" as the base case. For the variable region, we set the category "Atlantic" as the base case. Then, we construct the rest level with the regression coefficient, beta.

This is the basic model that can predict whether the Party of interest will get voted:

$$log(\frac{P_{Party}}{1 - P_{Party}}) = \beta_0 + \beta_1 x_{age} + \beta_2 x_{gender} + \beta_3 x_{income25,000-49,999} + \beta_4 x_{income50,000-74,999} + \beta_5 x_{income75,000-99,999}$$

$$+\beta_6 x_{income100,000-124,999} + \beta_7 x_{income125,000-more} + \beta_8 x_{Region\_Ontario} + \beta_9 x_{Region\_Quebec}$$

$$+\beta_{10} x_{Region\_Territories} + \beta_{11} x_{Region\_West} + \beta_{12} x_{ever\_married}$$

Explanation of the terms:

- P: the winning probability of the party of interest

- Beta0: the intercept of log(p/(1-p)) when all predictors are 0

- Beta1: all other variables kept unchanged, the change in voting outcome when variable age increases by one unit

- Beta2: all other variables kept unchanged, the change in voting outcome when variable sex changes by one unit

- Beta3-7: categorized income by its levels, where different income levels will influence the voting outcomes differently

- Beta8-11: categorized region by its levels, where levels refer to different geographical regions that will influence the voting outcomes differently

- Beta12: all other variables kept unchanged, the change in voting outcome when variable marital status changes by one unit

Note: the outcome we are referring to here, namely log(p/(1-p)), is not our desired winning probability. We need further decomposition to calculate the final winning probability.

Post-Stratification

Poststratification is a method for changing the proportions of predictions to improve their representation of the actual model. Enhancing accuracy is the aim of poststratification. For instance, it is a fact that disparate datasets, such as the one we will use to conduct this election, will have varying amounts of the same variable. Using data from the old community will give the age distribution more weight in the elderly population; on the other hand, looking at the new community, which indicates there are more young people there, will give us more information about the younger population. It is well known that young people's opinions can diverge from those of elder people. Consequently, in order to obtain a fair result for the parameter of interest, we must employ poststratification to alter the subgroup's weight.

Steps of Poststratification

1. Find the initial weight of each category - look at the data point in our survey data.

2. Identify our subgroups according to their characters, like age, gender, and other factors.

3. Calculate the proportion of our population by using the census data.

4. Adjust our weight by the formula $W_{i,new} = W_{i,old} * \frac{P_j}{S_j}$, where $W_{i,new}$ is the new weight, $W_{i,old}$ is the old weight, P_j is the proportion of subgroup j in the population, and S_j is the proportion of subgroup in the sample.

Formula

A typical post-stratified estimator of the population mean is

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

There are three terms in the expression of y^(ps):

1. N_j: The population size of each stratum j respectively

2. y_j: the sample mean for each stratum j respectively

3. $\hat{y}^{PS}$: the post-stratified estimate of the population parameter. To get the post-stratified estimate result, we need to multiply the population total with the sample mean for stratum j respectively, sum them up, and divide the total population.

Application to the Model

To incorporate the stratification model into our model, namely the Census data:

1. Define subgroups based on the variables in our Census data.

2. Calculate each proportion for each subgroup.

3. Adjust the weight in our survey. (Adjust the weight for each response in the survey data until the weight distribution of the subgroup in the sample matches the distribution in the Census data.)

4. Reanalyze the adjusted data.

5. Report the final result of our data.

Part 4: Results

Summary equations for the three logistic models are (base cases are not explicitly shown):

The prediction model for the Liberal Party:

$$log(\frac{P_{Liberal}}{1 - P_{Liberal}}) = -1.127816 + 0.016678 * x_{age} + 0.406333 * x_{gender}$$
$$+0.057531 * x_{income25,000-49,999} + 0.080921 * x_{income50,000-74,999} + 0.239490 * x_{income75,000-99,999}$$
$$+0.241901 * x_{income100,000-124,999} + 0.283310 * x_{income125,000-more}$$
$$-0.290126 * x_{Region\_Ontario} - 0.134239 * x_{Region\_Quebec} + 0.919561 * x_{Region\_Territories}$$
$$-1.084825 * x_{Region\_West} - 0.154896 * x_{ever\_married}$$

The prediction model for the Conservative Party:

$$log(\frac{P_{Conservative}}{1 - P_{Conservative}}) = -1.955730 + 0.012767 * x_{age} - 0.524153 * x_{gender}$$
$$-0.112467 * x_{income25,000-49,999} + 0.038412 * x_{income50,000-74,999} + 0.021898 * x_{income75,000-99,999}$$
$$+0.005490 * x_{income100,000-124,999} + 0.291110 * x_{income125,000-more}$$
$$+0.402795 * x_{Region\_Ontario} + 0.445900 * x_{Region\_Quebec} - 12.923094 * x_{Region\_Territories}$$
$$+0.884793 * x_{Region\_West} + 0.374587 * x_{ever\_married}$$

The prediction model for NDP:

$$log(\frac{P_{NDP}}{1-P_{NDP}}) = 0.959714 - 0.034119x_{age} + 0.155485x_{gender}$$

$$+0.074350x_{income25,000-49,999} - 0.135154x_{income50,000-74,999} - 0.313374x_{income75,000-99,999}$$

$$-0.273302x_{income100,000-124,999} - 0.731712x_{income125,000-more}$$

$$-0.074148x_{Region\_Ontario} - 0.272721x_{Region\_Quebec}$$

$$+0.955516x_{Region\_Territories} + 0.351198x_{Region\_West} - 0.189232x_{ever\_married}$$

Table 5: Predicted Probabilities

	Prediction Liberal	Prediction Conservative	Prediction NDP
prediction value	0.396987391720344	0.338718729231059	0.260014707691479

Based on our model:

The predicted probability of the Liberal Party getting voted is 0.396987.

The predicted probability of the Conservative Party getting voted is 0.338719.

The predicted probability of the NDP getting voted is 0.260015.

Thus, according to the 3 models that were built and the poststratification processes, the Liberal Party has the biggest probability of winning. This leads to reaching our original goal of predicting which party will win among the three. The sum of the three prediction values is 1, which is in line with the basic assumptions of winning probability in this model.

Table 6: Beta Values For Model 1

	beta_0	beta_1	beta_2	beta_3	beta_4	beta_5	beta_6	beta_7	beta_8	beta_9	beta_10	beta_11	beta_12
Value	-1.13	0.02	0.41	0.06	0.08	0.24	0.24	0.28	-0.29	-0.13	0.92	-1.09	-0.16

Table 7: Beta Values For Model 2

	beta_0	beta_1	beta_2	beta_3	beta_4	beta_5	beta_6	beta_7	beta_8	beta_9	beta_10	beta_11	beta_12
Value	-1.96	0.01	-0.52	-0.11	0.04	0.02	0.01	0.29	0.4	0.45	2.92	0.89	0.38

Table 8: Beta Values For Model 3

	beta_0	beta_1	beta_2	beta_3	beta_4	beta_5	beta_6	beta_7	beta_8	beta_9	beta_10	beta_11	beta_12
Value	0.96	-0.03	0.16	0.07	-0.14	-0.31	-0.27	-0.73	-0.07	-0.27	0.96	0.35	-0.19

Table 6: Interpretation of the Beta Values

The log-odds of the result when all other predictors are maintained at zero are represented by the intercept (beta_0). As can be seen in Models 1 and 2, a negative intercept indicates that if all predictors are zero, the log odds of the result occurring are less than even.

Positive beta coefficients show that the log odds of the result occurring grow along with the predictor. As an illustration, Model 1's beta_2, beta_5, beta_6, and beta_9 all have positive values, suggesting that a rise in these predictor variables is linked to a higher log-odds of the Liberal party-related outcome.

Negative beta coefficients imply that the log odds of the outcome occurring will drop with an increase in the predictor. For example, Model 1's beta_7 and beta_11 are negative, suggesting that higher values of these variables correspond to lower log odds of the Liberal Party-related outcome.

Beta Coefficients Near Zero suggest a nonexistent correlation with the result. Little or no influence of the predictor on the log odds of the outcome is suggested by small values, such as beta_1 in Models 1 and 2.

Part 5: Conclusion

Through research beforehand, we are confident enough to assume that the age, sex, household income, marital status, and geographical location of Canadian citizens are all important factors that influence voter decisions. (Uppal, 2012) Thus, we chose these five factors as the main variables in this modelling.

Since the target outcome we are predicting is probability, we chose the most appropriate logistic regression. To complete the model and ensure its accuracy, we modeled the different parties independently and practiced poststratification on them.

After completing our model, we obtained the probability of the three parties being elected in the next federal election, which adds up to a result just equal to 1. By comparing the prediction probability value of three main parties – Liberal, Conservative, and NDP – we can observe that the probability of Liberal Party winning in the next federal election is 0.396987, which is the highest among these three. The Liberal Party gets the highest probability, which is consistent with the facts of recent years, so our model has some guaranteed accuracy.

The two key results worth focusing on throughout our report are:

1. the values of all betas in the three models

2. the probability of the three parties being elected

Drawbacks & Limitations

There are almost 70% N/As in the original dataset. Also, throughout our cleaning process, we transformed one numerical variable (income) into categorical (income) to carry on with the Census dataset. This could potentially decrease the accuracy, as it modifies a piece of data from a specific number to an interval.

For Future Analysis

If possible, a more ideal logistic regression model could incorporate more variables than ours. Our model has more categorical variables than numerical ones, which could be less accurate had there been more valid numerical data to fit the model with.
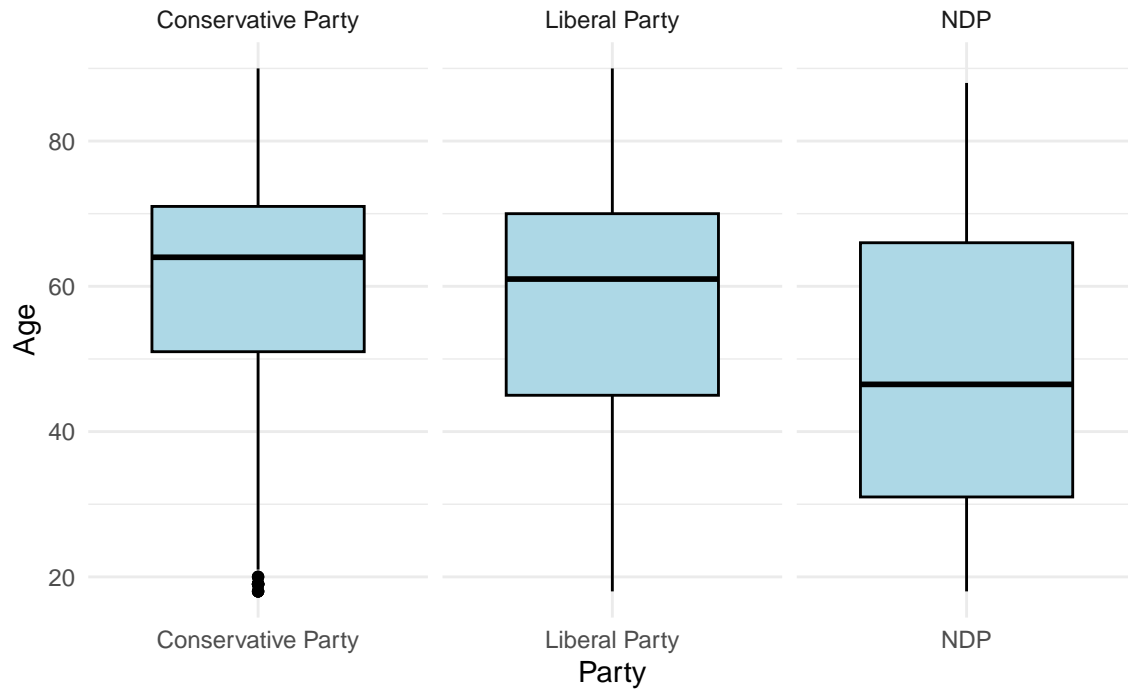
Bibliography

1. Catano, S. (2023, October). Multilevel Regression & Poststratification (MRPs). STA304 Lecture. Toronto.

2. Edgar, T. W., & Manz, D. O. (2017, April 28). Chapter 4 - Exploratory Study. Research Methods for Cyber Security.https://www.sciencedirect.com/science/article/pii/B9780128053492000042

3. Government of Canada, S. C. (2017, February 27). The General Social Survey: An overview. Government of Canada, Statistics Canada. https://www150.statcan.gc.ca/n1/pub/89f0115x/89f0115x2013001-eng.htm

4. Hahn, P. (2021, August 20). Interactive: How Canadians voted in the past 7 federal elections. CTVNews. https://www.ctvnews.ca/politics/federal-election-2021/interactive-how-canadians-voted-in-the-past-7-federal-elections-1.5553874?cache=

5. Kuha, J. (2022, May 12). Why are polls important during elections?: LSE research. Why are polls important during elections? | LSE Research. https://www.lse.ac.uk/research/research-for-the-world/impact/the-politics-of-polling-why-are-polls-important-during-elections

6. Solutions, S. (2021, August 11). Assumptions of logistic regression. Statistics Solutions. https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/assumptions-of-logistic-regression/

7. Uppal, S., & LaRochelle-Côté, S. (2012, February 24). Factors associated with voting. https://www150.statcan.gc.ca/n1/en/pub/75-001-x/2012001/article/11629-eng.pdf?st=IJGfTbZ-
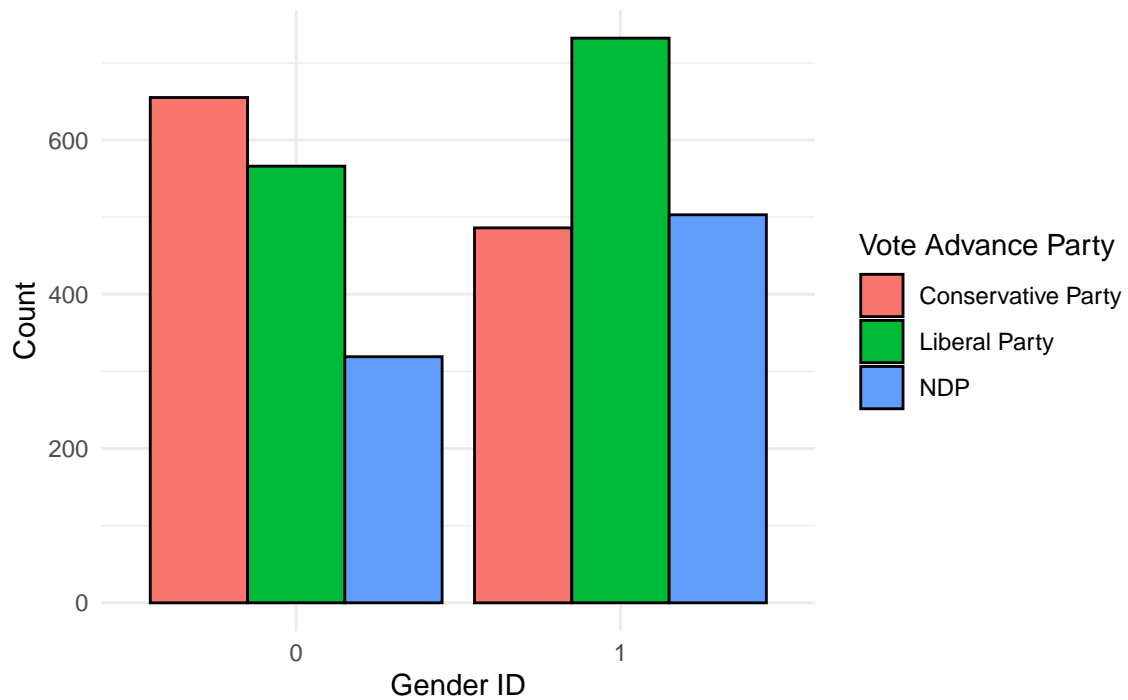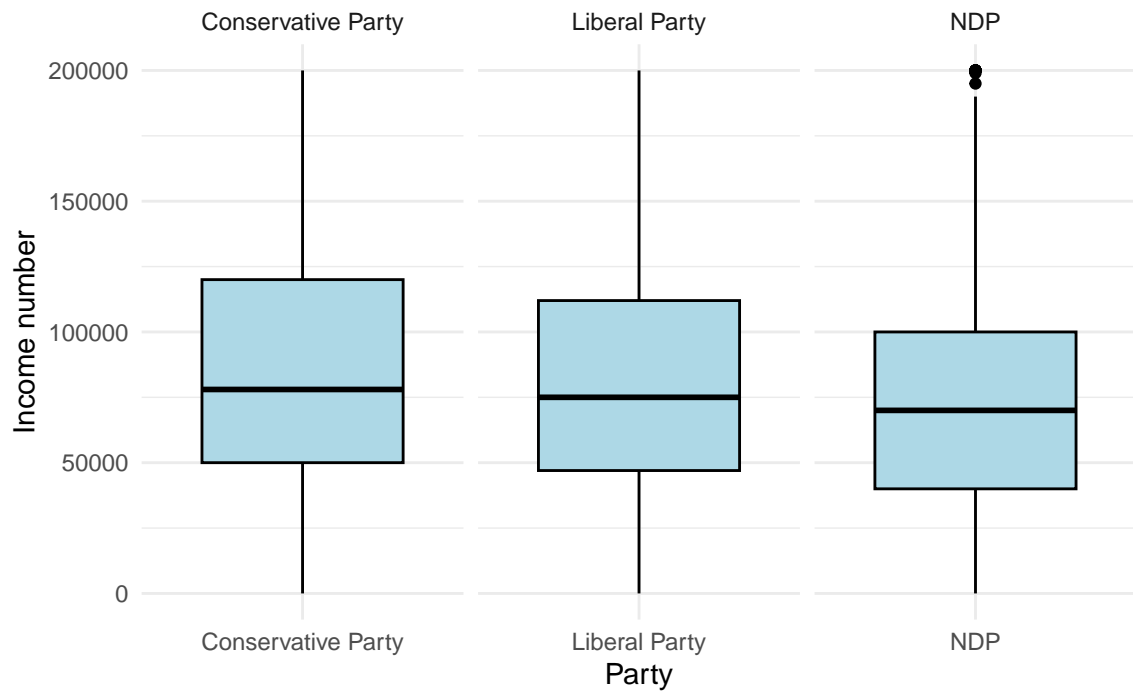
Appendix

Survey Data Graphs

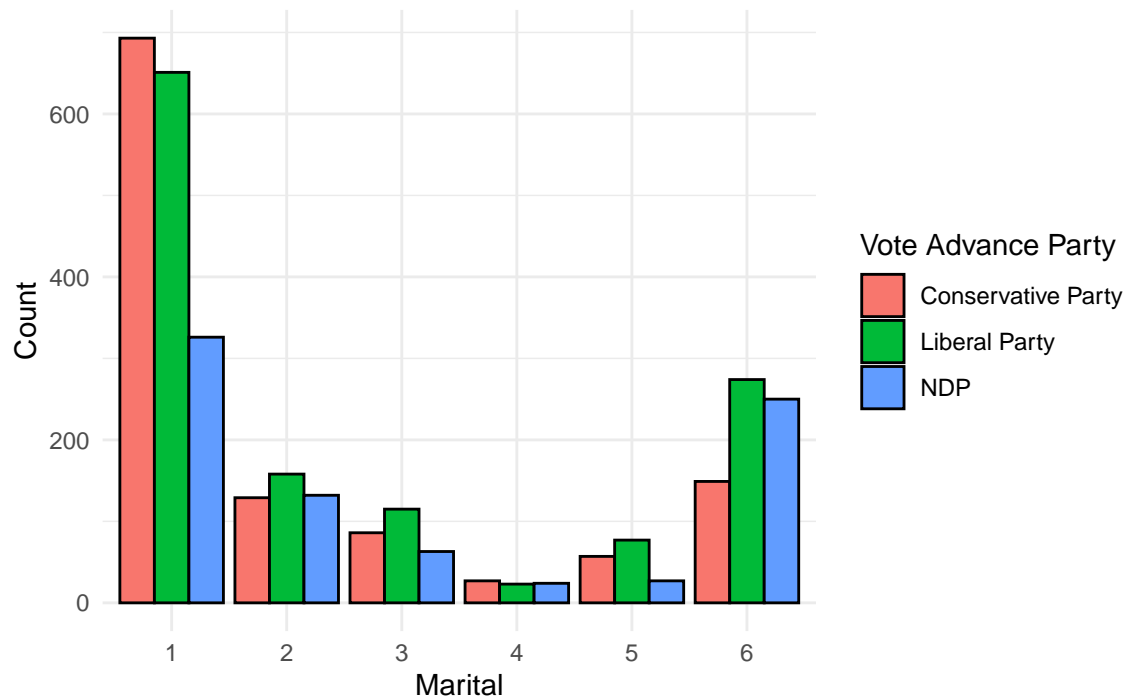## Plot 1: Distribution of Age By Voted Parties

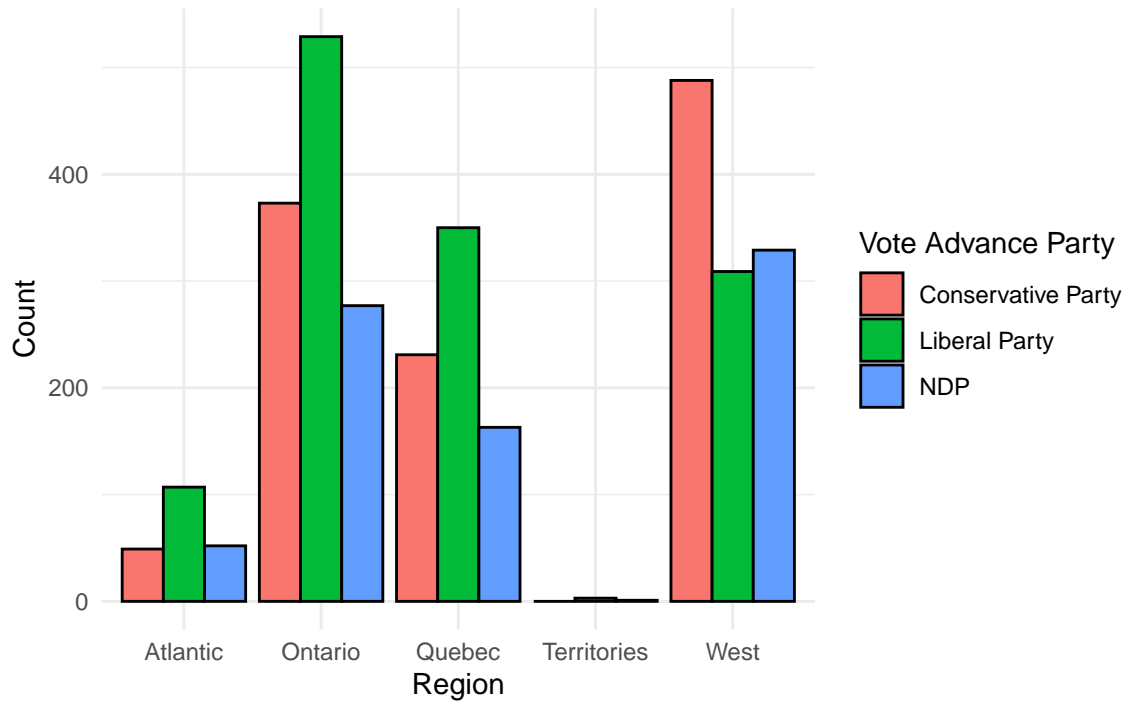## Plot 2: Distribution of Votes By Voted Parties and Sex
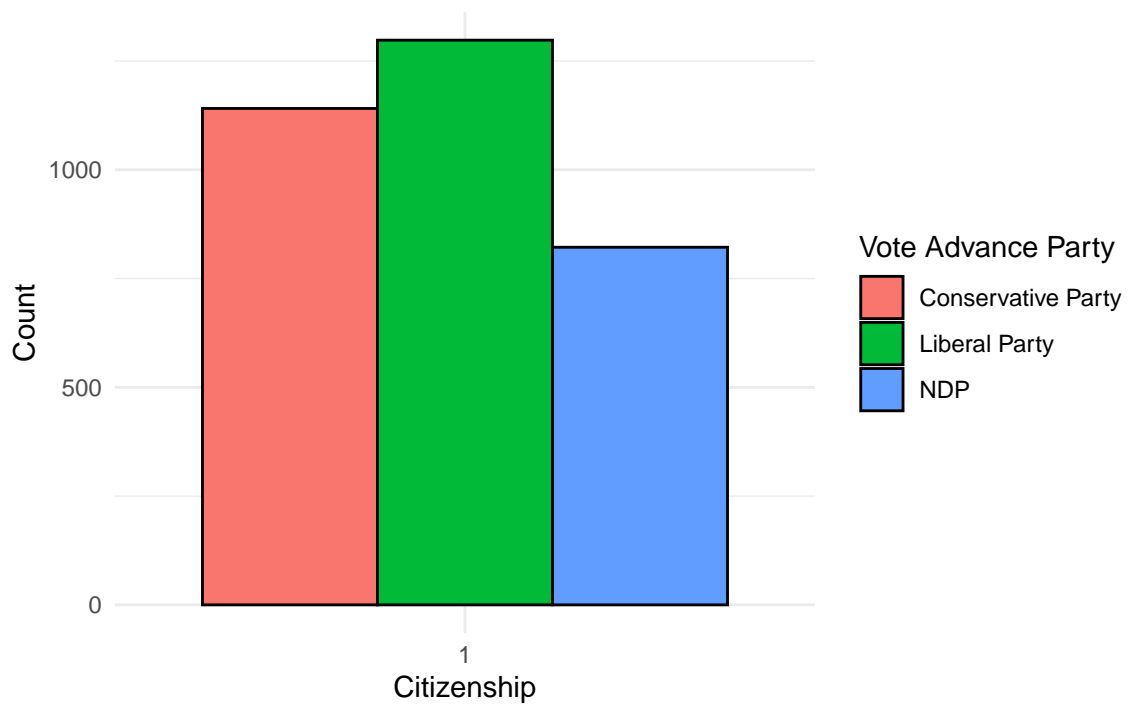
Plot 3: Distribution of Income By Voted Parties

Plot 4: Count of Votes By Voting Parties and Marital Status

Plot 5: Count of Votes By Voting Parties and Regions

Vote Advance Party
- Conservative Party
- Liberal Party
- NDP

Plot 6: Count of Votes By Voting Parties

Vote Advance Party
- Conservative Party
- Liberal Party
- NDP

Condition 2: Correlation Matrix

For the numerical variables age and family income in Survey data, we created a correlation plot, showing little correlation between these two numerical variables.
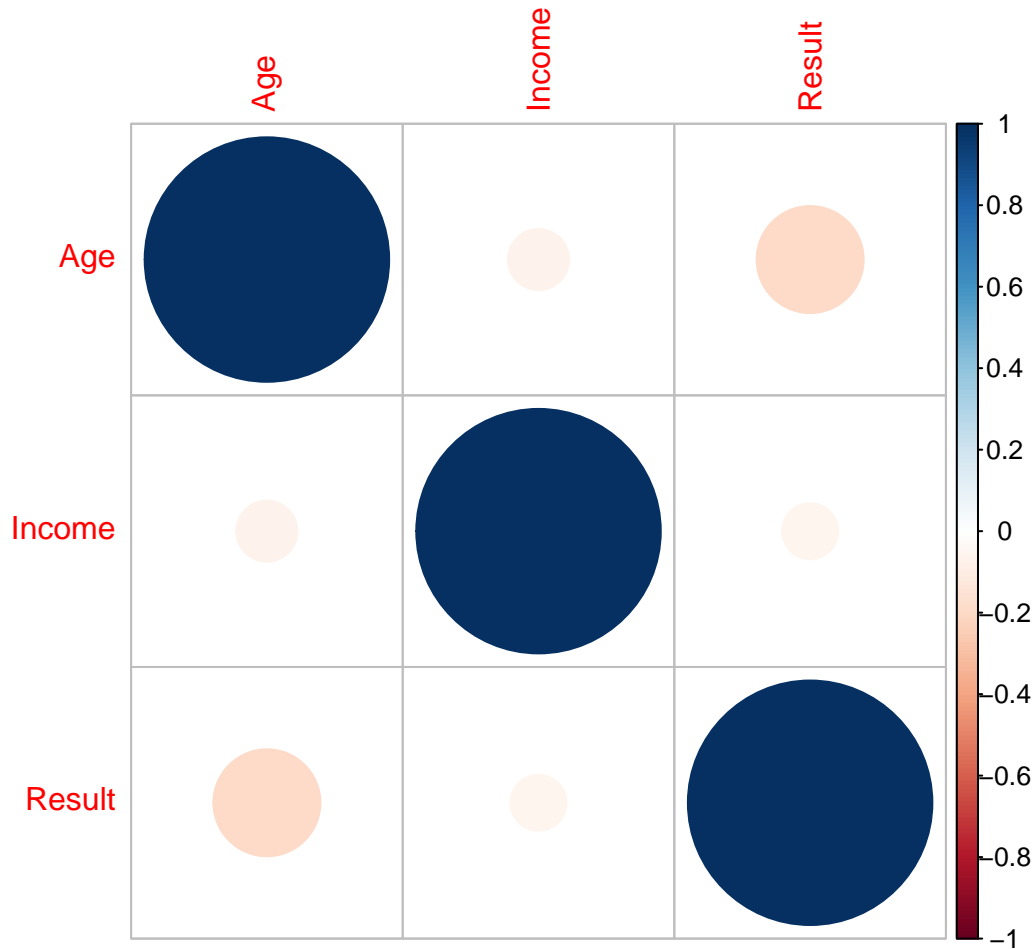
Condition 2: Categorical Variables

For the categorical variables, we will attempt to demonstrate non-multicollinearity by looking at some of the marginal distributions of population proportion of the categorical variables in Survey data.

1. Sex and Marital Status

2. Sex and Family Income

Table 9: Proportion when interaction the variable gender and variable ever married.

	Male	Female
No	0.1901257	0.3376265
Yes	0.1447409	0.3275069

Table 10: Proportion when interaction the variable gender and variable income family.

	Male	Female
Less than $25,000	0.0877032	0.1082490
$25,000 to $49,999	0.0944496	0.0947562
$50,000 to $74,999	0.0000000	0.0000000
$75,000 to $99,999	0.0877032	0.0665440
$100,000 to $124,999	0.1008893	0.0892364
$125,000 and above	0.0000000	0.0000000

We see that in both cases, the marginal distributions of male and female with respect to age are relatively similar, thus suggesting that either the variation in marital status or income depends on variation in sex.

Generative AI Statement

We used the following generative artificial intelligence (AI) tool: ChatGPT Version available on November 22nd, 2023. We used the tool only in the introduction section of this assignment, and we gave it the following prompt of: Why is the prediction of the election important?

And it gave me the response of: The prediction of elections is important for several reasons, and accurate predictions can have significant implications for various stakeholders, including the public, political candidates, policymakers, and the media. Here are some key reasons why election prediction is considered important: Informed Decision-Making: Accurate election predictions provide voters with valuable information to make informed decisions. Voters can better understand the dynamics of the political landscape and potential outcomes, allowing them to align their preferences with candidates or parties that align with their values and policy preferences. Campaign Strategy: Political candidates and parties use election predictions to shape their campaign strategies. Understanding the likelihood of winning in different regions or demographics allows candidates to allocate resources effectively, target key constituencies, and tailor their messages to maximize their chances of success. Resource Allocation: Political campaigns often involve significant financial and human resources. Accurate predictions help campaigns allocate their resources efficiently by focusing efforts on competitive races or swing districts. This can lead to more cost-effective campaigning. Media Coverage: The media plays a crucial role in informing the public about election dynamics. Predictions help guide media coverage, allowing journalists to focus on races that are expected to be competitive or have strategic importance. This enhances the quality of political reporting and analysis. Public Confidence: Accurate election predictions contribute to public confidence in the electoral process. When predictions align closely with actual election results, it reinforces the perception that the democratic system is functioning properly and that the outcomes reflect the will of the people. Policy Planning: Policymakers and government officials may use election predictions to anticipate potential changes in political leadership. This foresight can inform policy planning and help ensure a smooth transition of power, as officials prepare for potential policy shifts based on the expected election outcomes. Investor and Market Confidence: Election outcomes can influence economic and market conditions. Investors and businesses often seek to understand potential policy changes that may result from elections, and accurate predictions contribute to overall market confidence and stability. International Relations: Election predictions have implications beyond national borders. Other countries, international organizations, and foreign governments may monitor election predictions to anticipate changes in diplomatic relations, international agreements, and global geopolitical dynamics. Despite the importance of election predictions, it's essential to note that predictions are inherently uncertain, and unexpected events or shifts in public sentiment can influence actual outcomes. Consequently, predictions should be interpreted with an awareness of the inherent uncertainties in the electoral process.

We then started our brainstorm from those directions given and wrote up a part of our introduction paragraph.