Analyzing how sociodemographic characteristics influence voting for the Liberal Party of Canada*

Decoding CES 2021 Vote Data by using A Multilevel Logistic Regression Model

Yichen ji

April 17, 2024

This research utilized the 2021 Canadian Election Study and a multilevel logistic regression to examine how sociodemographic factors affect voting for the Liberal Party of Canada. It found that age, wealth, and education level significantly increase the likelihood of voting liberal, while children number, gender, and province do not. These results highlight the impact of life experiences and advantages on political ideologies and decision-making. Understanding these voting behaviors allows the Liberal Party of Canada to refine their electoral strategies, improving campaign effectiveness and voter engagement.

Table of contents

*Code and data are available at: https://github.com/Selinayichenji/Canadian_election_prediction.git.

1 Introduction

Voting decisions are at the heart of democratic societies, as they shape the composition of governments and determine the policies that govern our lives (Waiphot Kulachai and Homyamyen 2023). To understand the main groups shaping Canada's political policies, this paper will focus on the demographic composition and behavioral patterns of the liberal voters historically take up the majority of the voting (University n.d).

Previous studies have often focused on broader demographic trends or isolated the effects of singular socioeconomic factors on voting behavior. However, there is a need for a more integrated analysis that considers how combinations of these factors interact to influence political orientation. This paper seeks to fill that gap by examining multiple variables simultaneously to provide a more comprehensive understanding of voter behavior.

Based on prior voting analyses, key factors such as age, education level, family status (represented by the number of children), and economic well-being (indicated by family income) significantly influence voting behaviors (S. Canada 2015). To provide a more comprehensive view, gender will also be included as an influential factor in this study. Additionally, considering that Canada's population is concentrated in specific provinces, provincial distinctions will be accounted for as a random effects variable in our model. Notably, this study excludes Quebec, Yukon, Northwest Territories, and Nunavut due to distinct electoral dynamics in these regions. Under this method, we assume the authenticity of our source, 2021 Canadian Election Study data, to build up the multilevel logistic model. The provinces will serve as our random effects variable, assigning different intercepts to each province. This approach reflects the unique characteristics of each province, thereby enhancing the accuracy of our analysis.

Whether the voting for particular parties or not will be recorded binary, as 1 or 0. The estimand is the log-odds ratio of the effect of each explanatory variable on the probability of voting for the liberal party of Canada.

The analysis reveals age, wealth, and education level have a positive relationship with voting liberal, while children number, gender, and province do not.

The remainder of this paper is structured as follows: the basic information contains data source, methodology, variables, measurements and the visualization of population structure and voting rate are in the Section 2. The equation and interpretation of model are presented in Section 3. The data and figure of coefficients are in the Section 4. And in Section 5, we discuss what we have learned, our understanding of the world, the limitations, and the next steps of our research. Section A contains pp check and r-hat check for the model and Section A.2 is listed all references in this paper.

2 Data

The Canadian Election Study (CES) is a world-class study of Canadian elections and related attitudes of key importance to the study and understanding of democratic processes and elections. In terms of global research stature, it is comparable to other national efforts such as the American National Election Studies and the British Election Study, among others.(G. of Canada 2013)

2.1 Measurement

The CES 2021 was a comprehensive study conducted online, engaging a cohort of 20,968 participants selected from across Canada via the Leger Opinion panel. This bilingual survey, accommodating both English and French speakers, was open to individuals who were at least 18 years old and held Canadian citizenship or permanent residency. The study was methodically organized into two phases: the campaign period survey (CPS), which ran from August 17 to September 19, 2021, and the post-election survey (PES) carried out from September 23 to October 4, 2021. The CPS was subdivided into three parts—'CPS', 'CPS Modules', and 'CPS Oversample'—and these were ultimately merged into a single dataset. Participants from the CPS were invited to join the PES after the election concluded, with 15,069 individuals responding, resulting in a 72% response rate for the latter phase of the study.

2.2 Analysis Tools

The data analysis performed by the statistical programming language `R` (R Core Team 2023).

Besides the programming tool, we also employ the following packages: `here`(Müller 2020), `ggplot2`(Wickham 2016), `dplyr` (Wickham et al. 2023), `tidyverse` (Wickham et al. 2019), `tidyr`(Wickham, Vaughan, and Girlich 2023), `kableExtra`(Zhu 2021), `arrow`(Richardson et al. 2023) and `lintr`(Hester et al. 2024).

The `rstanarm`(Goodrich et al. 2020) package was used to build the logistic multilevel linear regression model. `bayesplot`(Gabry and Mahr 2024) was ustilized for Posterior predictive check and drawing R-hat figure in order to check the convergence of the MCMC algorithm. And `modelsummary`(Arel-Bundock 2022) was used for obtaining the coefficients of our model.

2.3 Variables

- **Age**:five age groups: 18-29 years old, 30-44 years old, 45-59 years old, 60-74 years old, and above 74 years old.
- **vote_liberal/vote_conservative/vote_NDP**: the voting result on whether to vote for three different parties. (Liberal, Conservative, or NDP) The result of voting or not is recorded in binary.
- **Gender**: only include two types of gender, female and male, excluding non-binary or others
- **education_level**: including six different education levels from low to high: Less than high school, High school, Non-University, University certificate below the bachelor, Bachelor's degree, Above the bachelor.
- **family_income**: including six different groups, from low to high: ("Less than 25,000", "25,000 to 49,999", "50,000 to 74,999", "75,000 to 99,999", "100,000 to 124,999", "125,000 and more")
- **Province**: Ontario, Quebec, British Columbia, Alberta, Manitoba, New Brunswick, Newfoundland and Labrador, Nova Scotia, Prince Edward Island, Saskatchewan (excluding Quebec, Yukon, Northwest Territories, and Nunavut.) The corresponded short terms are ON, QC, BC, AB, MB, NB, NL, NS, PE, SK.

  children_number: the number of children within the family, which includes only six levels (0,1,2,3,4 and 4+ children)

2.4 Sample of cleaned data

Table 1: Sample of the Cleaned Data

age	vote_liberal	vote_conservative	vote_NDP	gender	education_level	family_income	province	children_number
18-29	0	0	1	Female	University certificate below the bachelor	$125,000 and more	British Columbia	0
45-59	0	0	1	Female	Above the bachelor	$125,000 and more	Ontario	2
45-59	0	0	1	Female	University certificate below the bachelor	$50,000 to $74,999	Ontario	1
60-74	0	1	0	Male	University certificate below the bachelor	$100,000 to $124,999	British Columbia	1
45-59	1	0	0	Male	Non-University	$50,000 to $74,999	Manitoba	0

2.5 Survey data summary

2.5.1 The Population Structure in sociodemographic characteristics

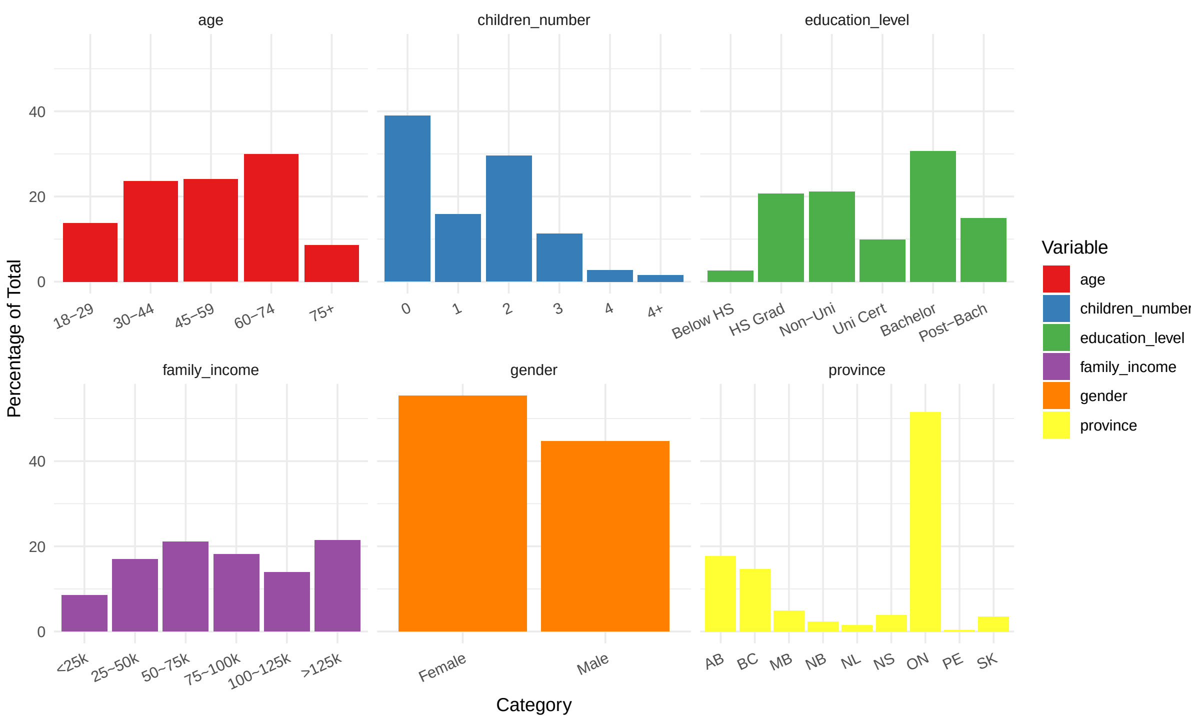

Figure 1: Population Percentage of different groups in a category

Figure 1 shows the proportion of the population within each subcategory to the population of the entire category after categorizing all voters according to six different demographic categories.

The demographic data indicates that the majority of the population falls within the age range of 60 to 74 years, accounting for approximately 30% of the sample.

Regarding family structure, 40% of individuals do not have children, which includes those who are unmarried, while a two-child family configuration is the next most prevalent, representing 30% of the sample.

In terms of educational attainment, the predominant qualification is a bachelor's degree. High school graduates and individuals without a university degree each make up about 20% of the population.

As for household income, the distribution appears relatively uniform across different brackets, with a notable concentration of individuals earning between 50k to 75k and those earning more than 125k, both groups constituting about 20%.

The gender composition of the population is slightly skewed towards women, who make up just over 50%, while men represent a marginally lower proportion, just above 40%.

Geographically, a significant portion of the electorate, around 50%, resides in Ontario, reflecting a centralization of voters in this province.

2.5.2 The vote rate for three main parties

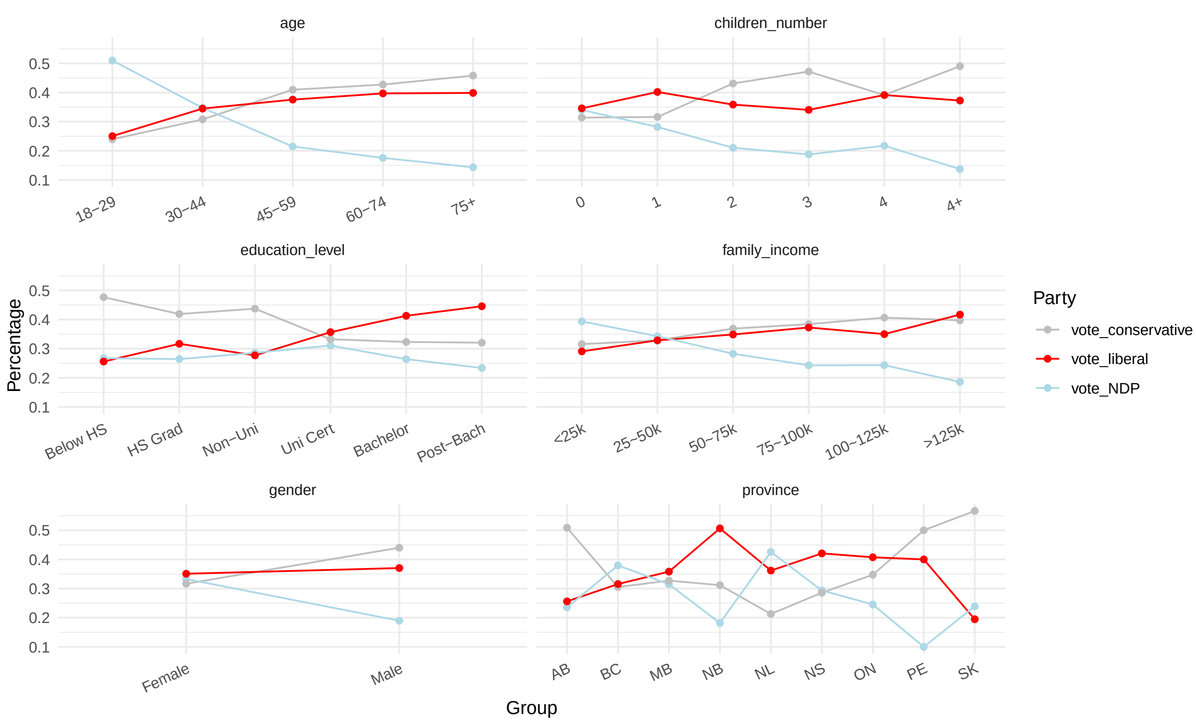

Figure 2: Percentage of Voting for 3 parties in different groups

Figure 2 displays the probability of voters voting for three main parties (the Liberal, the Conservative and the NDP) within each subcategory (e.g. '18-29') under six different demographic categories (e.g. 'age'), after classifying all voters according to these demographic categories.

Age: There appears to be a positive correlation between age and the percentage of votes for the Liberal party. As voters age, from the young demographic of 18-29 to those over 75, support for the Liberals seems to steadily rise from about 50% to just over 50%.

Education Level: From the lowest to the highest education levels, support for the Liberal party shows a relatively stable trend with a slight increase. Those with a bachelor's degree appear to be the most steadfast in their support for the Liberals, with a support rate above 40%.

Gender: Female voters show higher support for the Liberal party than males, indicating that gender may be a significant factor in support for the Liberals.

Number of Children: Families with more children seem to be more inclined to support the Liberal party, especially when there are four or more children, where the support rate peaks at about 45%.

Family Income: The Liberal party enjoys higher support in the middle-income brackets, while there is a decline in support among the higher income bracket (over 125k).

Province: Support for the Liberals is significantly higher in Ontario (ON) than in other provinces, possibly indicating the impact of regional factors on support for the Liberal party.

3 Model

3.1 Multilevel Logistic Regression Model

$$\begin{align}
y_i|\pi_i &\sim \text{Bern}(\pi_i) \tag{1}\\
\text{logit}(\pi_i) &= \beta_0 + \alpha^{\text{age}}_{a[i]} + \alpha^{\text{gender}}_{g[i]} + \alpha^{\text{education level}}_{e[i]} + \alpha^{\text{family income}}_{f[i]} + \alpha^{\text{children number}}_{c[i]} + \alpha^{\text{state}}_{s[i]} \tag{2}\\
\beta_0 &\sim \text{Normal}(0, 2.5) \tag{3}\\
\alpha^{\text{age}}_{a} &\sim \text{Normal}(0, 2.5) \text{ for } g = 1, 2, 3, 4, 5 \tag{4}\\
\alpha^{\text{gender}}_{g} &\sim \text{Normal}(0, 2.5) \text{ for } g = 1, 2 \tag{5}\\
\alpha^{\text{education level}}_{e} &\sim \text{Normal}(0, 2.5) \text{ for } g = 1, 2, 3, 4, 5, 6 \tag{6}\\
\alpha^{\text{family income}}_{f} &\sim \text{Normal}(0, 2.5) \text{ for } g = 1, 2, 3, 4, 5, 6 \tag{7}\\
\alpha^{\text{children number}}_{c} &\sim \text{Normal}(0, 2.5) \text{ for } g = 1, 2, ..., 6 \tag{8}\\
\alpha^{\text{state}}_{s} &\sim \text{Normal}(0, \sigma^2_{\text{state}}) \text{ for } s = 1, 2, ..., 9 \tag{9}\\
\sigma_{\text{state}} &\sim \text{Exponential}(1) \tag{10}
\end{align}$$

$$log(\frac{p_{ij}}{1-p_{ij}}) = \beta_{0j} + \beta_1 x_{ij\ age} + \beta_2 x_{ij\ gender} + \beta_3 x_{ij\ education\ level} + \beta_4 x_{ij\ family\ income} + \beta_5 x_{ij\ children\ number}$$

$$\beta_{0j} \sim Normal(\mu_{\beta_0}, \sigma^2_{\beta_0})$$

3.1.1 Model justification

In the model, j is the province and i is the individual voter indicator.

p represents the probability of the event of interest occurring, which is the voter's probability of voting for any specific party. p_{ij} means the voting of any specific party probability of ith voter in jth province.

Log Odds: The formula $log(\frac{p_{ij}}{1-p_{ij}})$ calculates the log odds of a voter choosing a specific party, effectively quantifying how likely they are to vote for the party versus not voting for it.

Intercept (β_0 and β_{0j}): The intercept, β_0, is the baseline log odds of voting for the party across all voters, assuming average conditions for all explanatory variables (age, gender, etc.). β_{0j} adjusts this intercept for each province, capturing local variations and assuming these effects are normally distributed.

Slopes ($\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$):

β_1: Change in log odds of voting for the party with each additional year of a voter's age.
β_2: Difference in log odds between male and female voters, adjusting for other factors.
β_3: Difference in log odds among education levels, controlling for other variables.
β_4: Difference in log odds among various family income levels.
β_5: Difference in log odds based on the number of children in the family.

Variables (x_{ij}): Represents categorical predictors such as age, gender, education, income, and number of children for voter i in province j, where each variable is coded as 0 or 1.

4 Results

4.1 Model Coefficients

Table 2: Summary Statistics of the Coefficients of the Linear Model

	(1)
(Intercept)	−1.871
age30-44	0.510
age45-59	0.755
age60-74	0.919
age75+	0.936
genderMale	−0.030
education_levelHigh school	0.229
education_levelNon-University	0.018
education_levelUniversity certificate below the bachelor	0.544
education_levelBachelor's degree	0.702
education_levelAbove the bachelor	0.734
family_income$25,000 to $49,999	0.136
family_income$50,000 to $74,999	0.228
family_income$75,000 to $99,999	0.301
family_income$100,000 to $124,999	0.166
family_income$125,000 and more	0.382
children_number1	0.084
children_number2	−0.168
children_number3	−0.248
children_number4	−0.084
children_number4+	−0.069
Sigma[province × (Intercept),(Intercept)]	0.186
Num.Obs.	3310
R2	0.060
R2 Marg.	0.042
ICC	0.1
Log.Lik.	−2066.231
ELPD	−2093.9
ELPD s.e.	20.0
LOOIC	4187.9
LOOIC s.e.	40.0
WAIC	4187.8
RMSE	0.47

Table 2 provides key coefficients to understand the relationship between variables and possibility of voting for the Liberal Party:

- **Intercept**: The model's intercept is estimated at approximately -1.871. indicating the log odds of voting for the Liberal Party when all other explanatory variables are at their reference levels, and the random effect of province is at its average level. A negative intercept suggests that without any other influences, the likelihood of not voting for the Liberal Party is higher than voting for it.
- **age groups**: The increasing coefficients indicates that, compared to the reference age group (usually the youngest group), the odds of voting for the Liberal Party increase with age.
- **gender**: The coefficient for Male is -0.030, suggesting a slightly lower likelihood for males to support the Liberal Party compared to females, assuming females are the reference category.
- **education levels**: The coefficients imply that individuals with higher educational qualifications, particularly those with a bachelor's degree (0.702) or higher (0.734), are more likely to vote for the Liberal Party compared to those without a high school diploma.
- **family income**: Higher-income families being more likely to vote for the Liberal party.
- **number of children**: The coefficients for number of children are negative, except for children_number1, meaning that households with children are less likely to vote for the Democratic Party compared to those without children.
- **random effect for province**: (Sigma[province × (Intercept),(Intercept)]) is valued at 0.186, indicating variability among intercepts across provinces, reflecting differences in voting behaviors among provinces.
- **Other statistical measures**: the number of observations (Num.Obs.) at 3310, the R^2 value at 0.060, and the Marginal R^2 (R^2 Marg.) at 0.042, suggesting that the model explains a modest amount of the variability overall. The model's log-likelihood (Log.Lik.) is -2066.231, which describes the probability of the model fitting the data. ELPD and LOOIC are criteria used for model selection and comparison, while the Root Mean Square Error (RMSE) is a measure of the model's predictive accuracy with a value of 0.47, where lower values indicate more accurate predictions.

Figure 3 is a statistical model's coefficient estimate graph, which demonstrates the impact of different variables on the probability of voting for the Liberal Party. Each row represents the coefficient of a variable in the model, with the point estimate of the coefficient represented by a solid black dot and its 90% credibility intervals (Credibility Intervals) represented by black lines. The point estimate of the coefficient (black dot) indicates the trend of the response variable (here, the probability of voting for the Liberal Party) when that variable increases by one unit. If the black dot is to the right of the zero line, it suggests that an increase in this variable may increase the probability of voting for the Liberal Party; if it is to the left, it may decrease the probability.

The width of the 90% credibility intervals (horizontal lines) reflects the uncertainty of the estimate. If the interval is narrow, it suggests a higher certainty about the impact of this

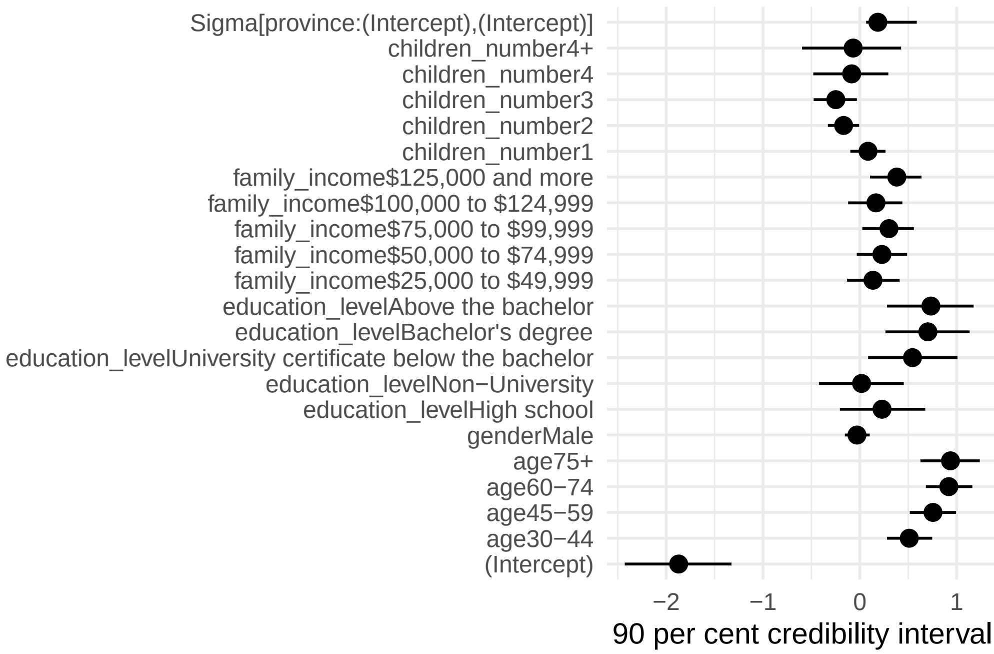

Figure 3

variable. If the interval is wide, it suggests a greater uncertainty in the estimate. If the 90% credibility interval does not contain zero, then we can consider that at the 90% confidence level, the variable has a significant impact on voting behavior. If the interval includes zero, we cannot confirm its impact at this level.

For instance, the credibility interval for 'genderMale' is quite narrow and does not contain zero. This suggests that the effect of being male on the probability of voting for the Liberal Party is estimated with higher certainty and is statistically significant at the 90% confidence level. The fact that the interval does not include zero implies that there is a discernible impact of gender on voting behavior. For Family Income (50,000 to 74,999), its interval is wider, suggesting that there's more uncertainty regarding the impact of this income level on the probability of voting for the Liberal Party. However, since the interval does not include zero, it suggests that there is likely a significant impact, but the exact magnitude of this effect is less certain.

Table 3: Coefficients of Liberal Party Model In Different Province

	(Intercept)
Alberta	-2.323063
British Columbia	-2.027230
Manitoba	-1.863658
New Brunswick	-1.382869
Newfoundland and Labrador	-1.776671
Nova Scotia	-1.573442
Ontario	-1.646282
Prince Edward Island	-1.782271
Saskatchewan	-2.441893

Table 3 shows the intercept for each province. The coefficients align with the trend illustrated in Figure 2.

5 Discussion

5.1 What is done in this paper?

In the paper, researchers conducted an in-depth analysis of Canadian voter behavior using a multilevel logistic regression model, specifically focusing on voters of the Liberal Party. They utilized data from the 2021 Canadian Election Study (CES), examining how factors such as age, education level, family income, gender, and the number of children influence voting support for the Liberal Party. Provinces were included as random effects variables in the model to account for political variations across regions.(exclude regions like Quebec, Yukon, Northwest Territories, and Nunavut due to their unique electoral dynamics.)

5.2 What we learn about the world?

Through this study, we have gained a nuanced understanding of the factors that influence voter support for the Liberal Party in Canada. As individuals age, their likelihood to support the Liberal Party increases, a trend particularly noticeable among older voters who have traditionally been Conservative supporters. This shift could be attributed to the Liberal Party's efforts in addressing issues pertinent to senior citizens, possibly influenced by recent events such as the COVID-19 vaccination rollout, where a significant percentage of senior Canadians have been vaccinated. The implication here is that policies that directly impact the well-being and concerns of older adults can be a powerful determinant of their voting preferences.

Furthermore, analysis of voting intentions suggests that gender plays a significant role among younger voters, where we see distinct patterns such as young men favoring the Conservative Party and young women showing stronger support for the NDP. Yet, for the Liberal Party, the support appears to be consistent across different segments, irrespective of age or gender. This consistency suggests that the Liberal Party's policies may have broad appeal across different demographics or that the party's stance on various issues is not polarizing enough to create a noticeable gender gap within its supporters.

Educational attainment is another influential factor, with those holding higher education levels likely resonating with the Liberal Party's emphasis on education, environmental protection, and technological innovation. These voters may view the Liberal Party as more aligned with their values and interests, particularly in policy areas that affect long-term societal progress.

Given these insights, it would be beneficial for policymakers and the Liberal Party to focus on developing and communicating policies that address the specific needs and concerns of these key voter groups. For the older demographic, continuing to provide and expand on healthcare, retirement insurance, and health protection will likely be vital. For the highly educated demographic, emphasizing the Liberal Party's commitment to advancing education and innovation could solidify support. Additionally, understanding the salient issues for different age and gender groups can help tailor communication and policy initiatives more effectively.

These observations are crucial for the Liberal Party if they are to maintain and strengthen their voter base, especially considering the evolving dynamics within the Canadian electorate.

5.3 Weaknesses and next steps

Although the model is sufficiently convergence in diagnostics section, the report relies on the assumption of complete trust in the authenticity of the data used; if responses are recorded in the survey or census showing fake information, it is impossible to eliminate the error caused by that misinformation. Meanwhile, accurate census data depicting the age distribution in the predicted year is essential. The percentage of different ages and the preference for different parties may vary throughout the years, leading to potential inaccuracies in our prediction. The exclusion of Quebec, Yukon, Northwest Territories, and Nunavut from our province variable, either due to distinct political parties within the province (Quebec) or missing data in our sources (Yukon et al.), limits the generalizability of our predictions of these regions. Furthermore, there are different economic conditions between the year of data collected and the year of the prediction. The variation of the economic situation could affect the outcomes of the winning party.

In addition, the relatively limited explanatory power of our model (R^2 value of 0.060) suggests that there are other factors that are not captured by the model that may be influencing voters' voting behavior. Future research could explore additional variables such as cultural orientation, social identity, and recent political events, all of which may have a significant impact on voter behavior.

Looking forward, analyzing the voting results within Quebec, Yukon, Northwest Territories, and Nunavut satisfies their unique provincial characteristic result better.

A Appendix

A.1 Posterior predictive check

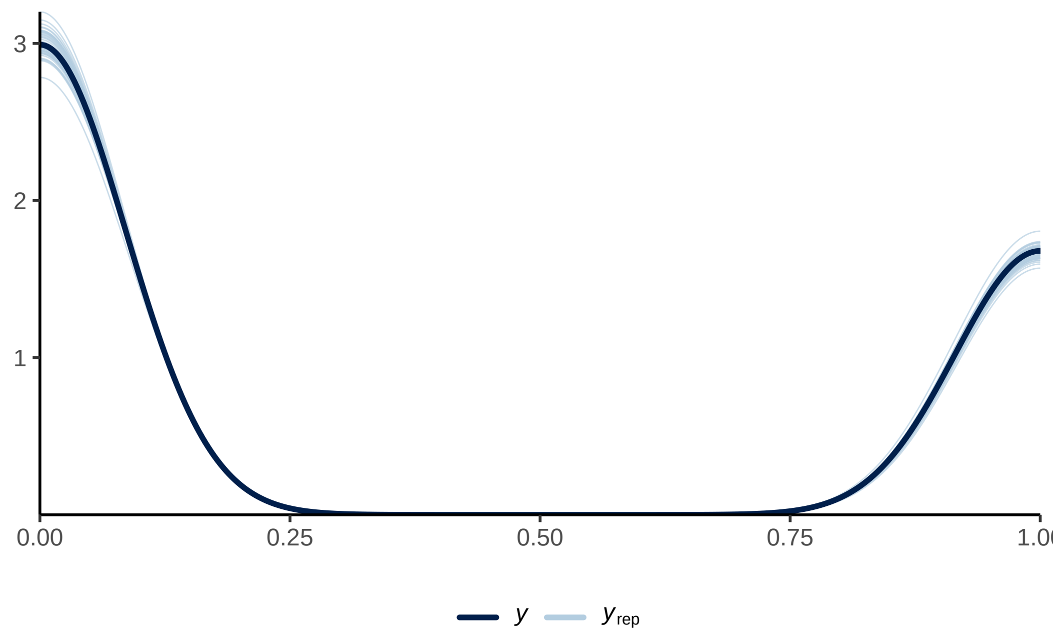

Figure 4: pp check

In Figure 4, we implement a posterior predictive check. This shows the simulated data from the posterior distribution (illustrated by the lighter lines) aligns closely with the distribution of the observed data (indicated by the solid line). The lighter lines, which represent the uncertainty in our predictions, nearly converge with the solid line. This convergence suggests that our model accurately fits the observed data, indicating that the selected priors and model structure effectively capture the variations in voting for the Liberal Party among interviewees.

A.2 Diagnostics

In Figure 5, the distribution of R-hat values is displayed, which is used in Bayesian statistics to assess convergence. Different R values are represented by circles of different colors. Blue circles represent R values that do not exceed 1.05, which is typically considered an acceptable level of convergence; light blue circles indicate R values between 1.05 and 1.1, which may suggest the need for more samples or further investigation; and dark blue circles represent R values greater than 1.1, which usually indicates insufficient convergence, requiring longer chains, more samples, or model improvements. This result suggests that there are no convergence issues in this particular Bayesian model estimation. All points are tightly concentrated around the standard line of 1.00, which is a very positive sign.

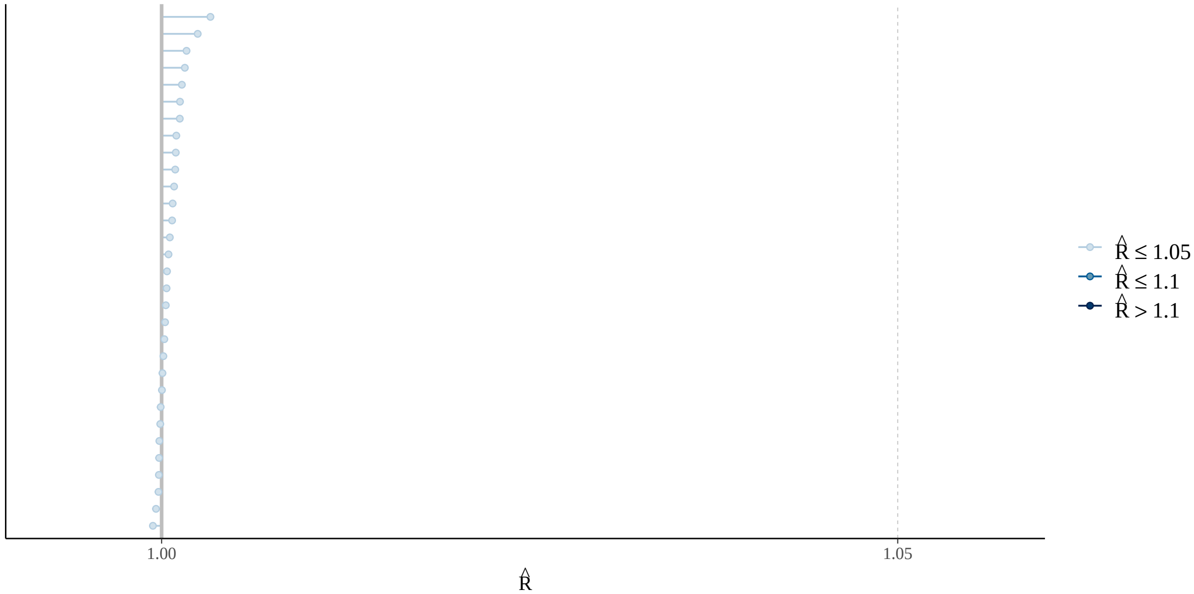

Figure 5: Checking the convergence of the MCMC algorithm

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Canada, Government of. 2013. *Canadian Election Study Data Set*. https://www.sshrc-crsh.gc.ca/funding-financement/programs-programmes/election_study_election_etude-eng.aspx.

Canada, Statistics. 2015. "Factors Associated with Voting." https://www150.statcan.gc.ca/n1/pub/75-001-x/2012001/article/11629-eng.htm.

Gabry, Jonah, and Tristan Mahr. 2024. "Bayesplot: Plotting for Bayesian Models." https://mc-stan.org/bayesplot/.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2020. *Rstanarm: Bayesian Applied Regression Modeling via Stan*. https://mc-stan.org/rstanarm.

Hester, Jim, Florent Angly, Russ Hyde, Michael Chirico, Kun Ren, Alexander Rosenstock, and Indrajeet Patil. 2024. *Lintr: A 'Linter' for r Code*. https://CRAN.R-project.org/package=lintr.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2023. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/packag e=arrow.

University, Simon Fraser. n.d. "Factors Associated with Voting," n.d. https://www.sfu.ca/~aheard/elections/1867-present.html.

Waiphot Kulachai, Unisa Lerdtomornsakul, and Patipol Homyamyen. 2023. "Factors Influencing Voting Decision: A Comprehensive Literature Review." *Social Sciences* 12 (9): 469. https://doi.org/10.3390/socsci12090469.

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

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

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Davis Vaughan, and Maximilian Girlich. 2023. *Tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.