

# Forecasting the Canadian Federal Election 2025

## A Post-Stratified Regression Approach to Predicting Popular Voting Dynamics

Raima Zaman

22nd November 2023

### 1 Introduction

#### 1.1 Background and Motivation

Canada's Parliament consists of the monarch, along with two chambers: the House of Commons, representing the lower chamber, and the Senate, which stands as the upper chamber. In the Senate, the Governor General (the Queen's representative in Canada), appoints senators based on advice from the Prime Minister. The House of Commons on the other hand, is made up of Members of Parliament (MPs) elected by Canadian citizens during federal elections which take place at least one every four years [4]. Each Member of Parliament (MP) represents a specific geographical area; referred to as an electoral district, or riding.

Currently, there are 338 seats in the House of Commons as established by the 2013 Representation Order, each seat representing a specific riding. These seats are allocated across the nation's ten provinces and three territories; while each of the ten provinces have multiple ridings, each of the three territories have one representative riding [5]. The boundaries of these ridings are based off of the respective population distribution in each of these areas, ensuring that the population within each riding is relatively equal to provide a fair representation.

During the federal election, in each riding, any number of candidates may run for election either as an independent candidate or under a registered political party. Eligible citizens in each riding vote for their preferred candidate. Under this system, the candidate who receives the most votes (compared to all other candidates within the same riding) wins the election for that specific riding. This elected candidate hence wins a seat in the House of Commons and serves as the Member of Parliament (MP) for the respective riding [6].

Notably, if a single political party secures the majority of MPs, the leader of that party becomes the Prime Minister [7]. Thus, understanding citizens' preferences for political parties not only impacts their chosen candidate but also predicts the selection of Canada's future Prime Minister.

While several political parties are registered in Canada, since the federal election of 2021, only five abstain representation in the House of Commons. These are the parties we focus on in this paper; the Liberal Party of Canada, the Conservative Party of Canada, Bloc Quebecois, the New Democratic Party, and the Green Party of Canada.

The New Democratic Party, led by Jagmeet Singh, and the Green Party, led by Elizabeth May, both lean toward the centre-left to left-wing political positions. Contrarily, the Conservative Canadian Party, led by Pierre Poilievre, occupies a centre-right to right-wing stance. The Liberal Party, under Justin Trudeau's leadership, falls within the centre to centre-left position, while the Bloc Québécois, led by Yves-François Blanchet, holds a centre-left political stance [8].

## 1.2 Objectives

The Prime Minister is largely responsible for the implementation of governmental and economic policies in Canada. Additionally, they play a significant role in proposing legislation and guiding government bills through the legislative process [9]. Consequently, the Prime Minister plays a crucial role in steering the country's course, representing its interests both domestically and internationally.

As a result, predicting which party will secure the majority of votes is critical to determining the elected Prime Minister, and hence, shaping governmental, economic policies, and legislative agendas. This decision profoundly impacts the Canadian populace, directly influencing their lives and futures. Given these implications, our objective is to predict the party expected to secure the greatest number of votes in the forthcoming 2025 federal election in Canada.

To achieve this, we employ a post-stratification strategy, primarily building a regression model based on survey data. We adjust the weights according to our population from our census data to facilitate accurate inferences and predictions. This approach allows us to anticipate the potential outcomes of the 2025 Canadian Federal Election, thus providing valuable insights into Canada's political landscape.

The primary reason for using a post-stratification technique is to obtain representative results from unrepresentative samples. Previously, while election predictions had been based off of representative polling samples, obtaining these samples proved both costly and time-consuming. Further, as response rates have declined over the years, the statistical benefits of representative sampling processes (such as address-based, in-home interview, and random digit dialing sampling among several others), have declined. Thus, we implement a post-stratification strategy with a regression model that allow us to use non-representative yet extensive samples, to estimate voter choices by aggregating these estimates according to the target population's demographic structure.

To build our model and conduct our hypotheses, we've obtained information from the 2021 Canadian Election Study (CES) survey data, and for our census information, we refer to the General Social Survey (GSS) census data. The CES (Canadian Election Study) data refers to a survey conducted both during and after the 2021 federal election in Canada. It contains comprehensive information regarding voters' behavior, opinions, attitudes, and preferences during and since the electoral process. This survey encompasses various aspects, including but not limited to voter demographics, voting behaviour, and political attitudes. The General Social Survey (GSS) on the other hand, is a prominent survey conducted in Canada to gather information on various social trends and attitudes among the country's population. Conducted every 5 years, it covers a wide array of topics, such as demographics, education, employment, income, family dynamics, health, and more [10].

We thus intend to post-stratify and make predictions for each of the five political parties currently represented in the House of Commons. We expect demographic and socioeconomic differences in the populous to affect their voting choice.

In particular, we hypothesize that the Liberal Party typically attracts the votes of younger and middle-aged voters (those approximately in the 18-34 age range), individuals with higher education levels, ethnically diverse communities (particularly visible minorities), and generally those whose views align with progressive and inclusive policies. Similarly, we expect the New Democratic Party (NDP) to garner the votes of the younger populous (particularly those under the age of 35), individuals in the lower to middle-income brackets, voter bases located in diverse neighbourhoods, and generally those who prioritize social justice, environmental sustainability, and income equality. Particularly, we'd expect the NDP to attract the votes of labor union members as their policies highlight workers' rights and unions. While we'd expect the Green Party to attract the votes of younger individuals with higher education levels, we'd expect particularly those with strong environmental concerns and progressive views to align their votes with the Green Party. Alternately, we'd expect the Conservative Party to garner the votes of the older populous, those in higher-income brackets, and generally those whose views align with traditional values. Particularly, we'd expect a high likelihood of voting conservative among small business owners as a result of their policies advocating for lower taxes and reduced government regulation. Further, we'd expect a high ratio of votes for Bloc Quebecois to originate from french speakers, those either from Quebecois, or those who prioritize the interests of Quebec.

### 1.3 Outline

We organize this paper as follows. The Data section starts by introducing Canadian federal election related data and showcases some preliminary analysis, highlighting important descriptive summary statistics. The Methods section introduces notation, formulation, and our proposed modeling approach. The Results section summarizes our predictions for the upcoming Canadian federal election in 2025, and we end our paper by highlighting our final remarks in the Conclusion section.

## 2 Data

### 2.1 Canadian Census Data

For the purposes of our paper, we utilize the 2017 General Social Survey (GSS) dataset for Canada as our census data. The GSS is a comprehensive and nationally representative survey conducted by Statistics Canada, aimed at collecting social data on various aspects of Canadian society. The GSS collects data on a wide range of topics, including demographics, social attitudes, family dynamics, employment, education, and more. Typically, the survey involves thousands of Canadians across diverse age groups, ethnicities, and regions, making it representative of the entire country. The data collection process often includes face-to-face interviews or telephone surveys, employing both closed-ended and open-ended questions to gather comprehensive insights into societal trends and changes over time. The GSS’s longitudinal nature allows for the tracking of shifts in social norms and attitudes, making it a valuable resource for understanding the evolving landscape of Canadian society. Thus, it serves as a crucial tool for researchers, policymakers, and analysts by providing rich data on social trends, opinions, and behaviors across the Canadian population.

### 2.2 Canadian Federal Election Post-Survey Data

The Canadian Election Study (CES) 2021 dataset is a fundamental resource capturing comprehensive data about the political landscape, electoral behaviors, and attitudes of Canadian citizens during and after the federal election period in 2021. Conducted by a plethora of researchers from various Canadian universities, this survey-based study aimed to gather crucial information about voting patterns, political preferences, and the socio-political landscape at the time of and preceding the federal election in 2021. This extensive survey collected data through rigorous methods, including face-to-face interviews, telephone interviews, and online questionnaires, targeting a diverse sample of Canadian citizens across different regions, demographic groups, and socio-economic backgrounds. The survey questionnaire covered a wide spectrum of topics, including political attitudes, party preferences, trust in political institutions, satisfaction with government policies, perceptions about key political issues, and individual voting behaviors. Moreover, the CES 2021 dataset encompasses valuable information on demographic characteristics, socio-economic status, educational backgrounds, and geographical distributions of respondents, providing a more balanced view of the Canadian electorate during the federal election period.

### 2.3 Data Cleaning

In order to derive accurate and meaningful inferences, it’s imperative to preprocess our census data, as well as our survey data, to ensure they align with the analytical objectives for this paper. We first shift our focus to the cleaning process of our census data.

As we had highlighted earlier, we are interested in retaining the variables that represent key demographic and socioeconomic characteristics of the Canadian populous. Thus, we first concentrate on creating a subset of the census data, which includes the variables for age, sex, income level, education level, province, and language predominantly spoken at home.

It must be noted that our reduced census data set contains missing observations for the variable representing the education level of the populous, as well as for the variable representing the primary language spoken

at home by the populous. The education variable consists of 341 missing observations, representing 1.66% of the total observations. Since the categorical education variable doesn't account for individuals with a lack of schooling, we group these absent entries under a distinct label termed "No Schooling." The language variable consists of 448 missing observations, representing 2.17% of the total observations. As we are primarily interested in observing the distribution of the predominantly French speaking and English speaking populous, we designate these missing observations under a category labelled "Other". Furthermore, to concentrate solely on the eligible voting population, we filter out observations with ages below 18 years from the age variable, as they do not represent our intended demographic of interest.

After addressing these missing observations, our next step involves the creation of our final data set, where we incorporate our processed variables. Here, we do not modify the variables for income level, sex, or province. For all the remaining variables, we include alterations to the categories.

We initiated the data transformation process by categorizing the age variable, first transitioning it from a continuous to a discrete numeric structure. The segmentation of this variable involved establishing distinct age groups: individuals below 35, those aged between 35 to 54, and those aged 55 years or more. For the education variable, our focus was on enhancing interpretability. Consequently, we refined the labels within the distinctive education level categories to make them more easily understandable.

Shifting our attention to the language spoken at home variable, our primary aim was to emphasize the distribution of English and French speakers across Canada. Thus, we streamlined this variable to signify 'English' for individuals primarily speaking English at home, 'French' for those predominantly speaking French at home, and grouped observations of any other predominantly spoken language into a category labelled 'Other'.

Similarly, we preprocessed our survey data. Once more, we constricted the CES survey data to a reduced dataset containing only our variables of interest: gender, age, province, income, education level, language spoken at home, and an additional variable, voting choice. However, before stepping forth with the handling of our variables, we checked for missing observations and consequently found gaps in the voting choice and language spoken at home variables.

The voting choice variable consisted of 6391 missing observations, representing 30.5% of the total observations. Given that the CES survey contained options for the five parties currently represented in the House of Commons, along with an 'Other' and 'Unsure' category, we assumed that survey participants who chose to not answer must have either been indecisive, unlikely to vote, or unsure. Regardless, we categorized these missing observations under the 'Unsure' category and removed observations with this category from our data. The language spoken at home variable consisted of 5899 missing observations, representing 28.13% of the total observations. Recalling that we're only interested in the distribution of the primarily English and French speaking populous, we binned these missing observations under an 'Other' category. We rationalized this decision as the CES survey data assigns 'French' and 'English' as the first two options for the question representing language at home. We thus assumed participants who chose not to answer must predominantly speak another language at home, allowing for their categorization under 'Other'.

Following the mitigation of these missing values, the subsequent stage involved the pre-processing of our data, essentially adjusting category labels and funneling them into our cleaned data set. We began by converting how the observations for the voting choice variable were coded, altering them so that the numeric labels referred instead to the five distinct party names, along with the categories "Other", and "Unsure".

In the case of the gender variable, we ran into a difficult situation. While we require our survey data to match up to our census data to allow for post-stratification, we do not have any gender variable within our census data, instead participants were asked to answer their biological sex. Thus we modified the gender variable to represent the two distinct categories under the respective sex variable from our GSS census data. Given that the options for "non-binary" and "other" made up merely 0.43% and 0.16% of the total observations respectively, we opted to remove these observations entirely from our data set. Once more, we categorized our age variable, transitioning it from a discrete numeric structure to a categorical structure, establishing three distinct age groups: individuals below 35, those aged between 35 to 54, and those aged 55 or older. Further, we converted our variable for income from a discrete numerical structure to a categorical structure, setting different levels for income to allow for correspondence to the respective levels in our census data set. Unlike

our census, our survey contained the responses of participants from the provinces of Nuwanut, Yukon, and Northwest Territories. However, in order to match our census data, we removed these observations (they each made up a measly 0.02%, 0.13%, and 0.07% of the observations respectively) and simply changed the labels of the remaining 10 provinces to represent the respective province names instead of the allocated numeric values. Further, we changed the labels for the education variable from the existing numerical categories to categorical labels in order to enhance interpretability and understanding, and to increase conformity to our census data. Finally, we only retained the ‘English’ and ‘French’ responses according to their respective label names, consolidating all other responses under the ‘Other’ category.

Further, we created separate data sets for each categorization of the voting party choice variable. For our consequent analysis, we included a new variable under each of these data sets that portrayed a binary (0,1) value for each observation, respective to the data set. For example, for the data set we created for liberal party voters, we included a variable that presented as 1 if an observation favored the Liberal Party, and 0 if otherwise.

## 2.4 Key Variable Description

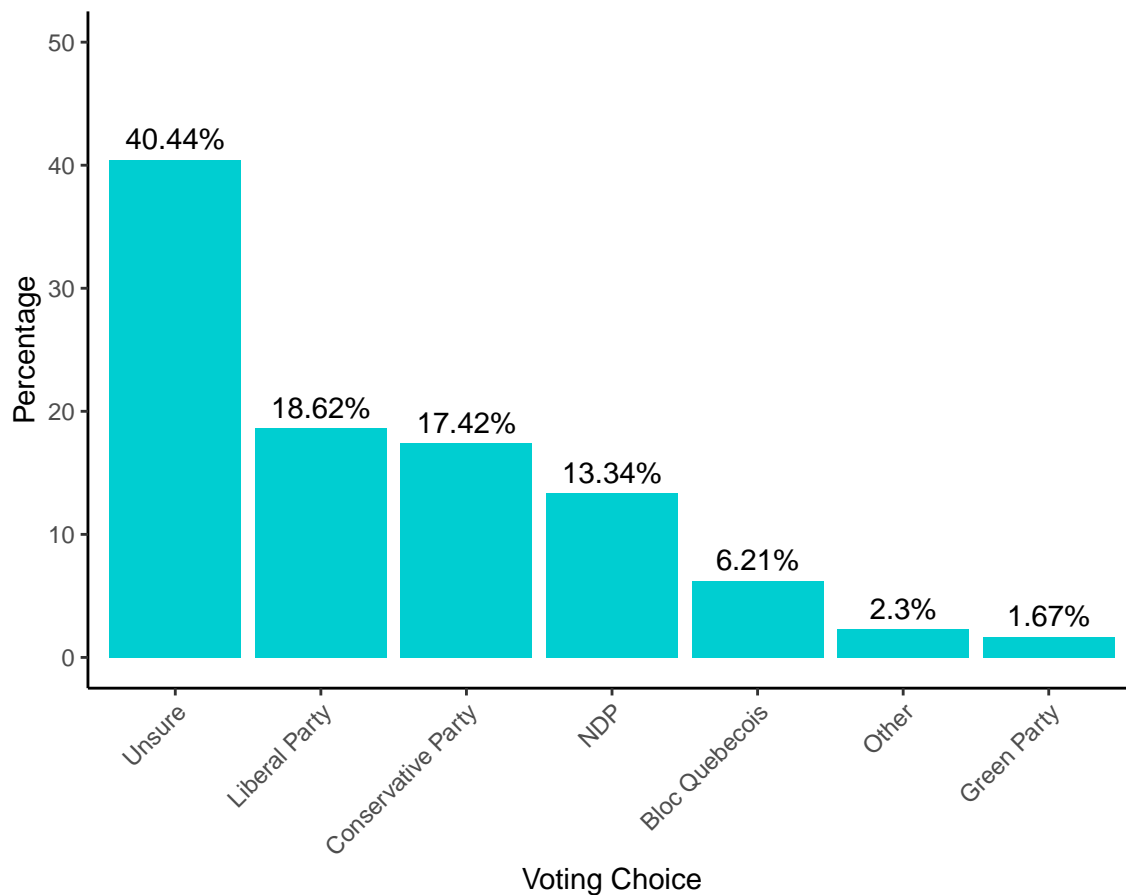
Before divulging into our model, it is of utmost importance to understand how each of our variables are formatted, what they represent, and how they are distributed in our data. We thus dive into a deep explanation of the designated variables in our cleaned and reduced census and survey data sets.

While both our datasets consist of the same variables, mapped relative to each other in order to allow for post-stratification, the cleaned survey data additionally consists of the variable representing a participant’s party vote choice. The predictor vote choice is a categorical variable, consisting of seven categories in total. Five of these categories represent the political parties currently represented in the House of Commons: Liberal Party, Conservative Party, NDP, Bloc Quebecois, and Green Party. The other two categories exist as representations for any other political party in Canada, and voters who are unsure of the political party they favor. In our analysis, we only focus on the five specific party categorizations.

```
##
##
## Table: 1: Distribution of Voting Choice in Survey Data
##
## |Vote Choice          | Percentage|
## |:-----|:-----:|
## |Unsure               |    40.44|
## |Liberal Party        |    18.62|
## |Conservative Party   |    17.42|
## |NDP                  |    13.34|
## |Bloc Quebecois       |     6.21|
## |Other                |     2.30|
## |Green Party          |     1.67|
```

As we had dumped roughly 30% of the missing data for this variable into the Unsure category, it is thus over-represented in our data. From Table 1, we can see that the Unsure category now represents 40.44% of of the total observations for vote choice, entailing 40.44% of survey respondents are unsure of the political party they are in favor of. However, as we are aware this category is unfairly over represented, we choose to ignore it, assuming that the true proportion of survey respondents unsure of their vote to be approximately 10% (40.44% of current representation - 30.5% of missing observations). We thus draw our attention to the distribution in the proportion of the remaining categories

Figure 1: Distribution of Voting Choice



From Figure 1, we can see that while the greatest majority of survey respondents favor the Liberal Party (18.62%), this majority is closely followed by survey respondents' favor for the Conservative Party (17.42%). The NDP holds the rank for the third highest percentage of survey respondents' favor (13.34%), while the Bloc Quebecois garners roughly half the support of the NDP (6.21%). Even the accumulation of votes for the various other political parties in Canada (2.3%), outnumber the percentage of favor for the Green Party (1.67%).

We continue our analysis of the remaining relevant variables with the aid of the plots under Figure 2, which highlight the distributions of these variables in our CES survey data.

Figure 2.1: Distribution of Sex in Survey Data

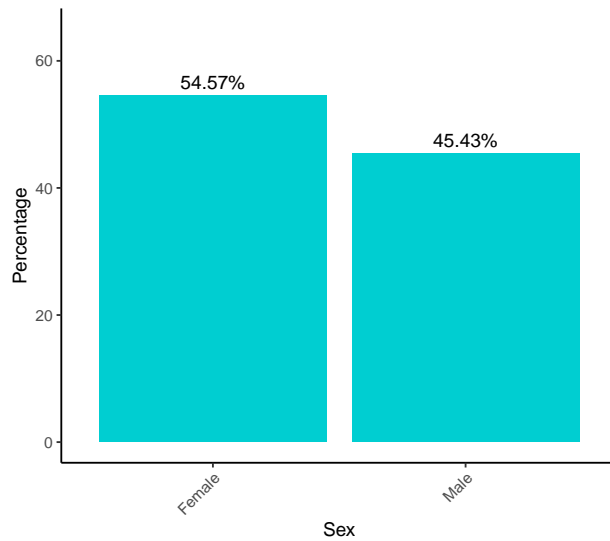


Figure 2.2: Distribution of Age in Survey Data

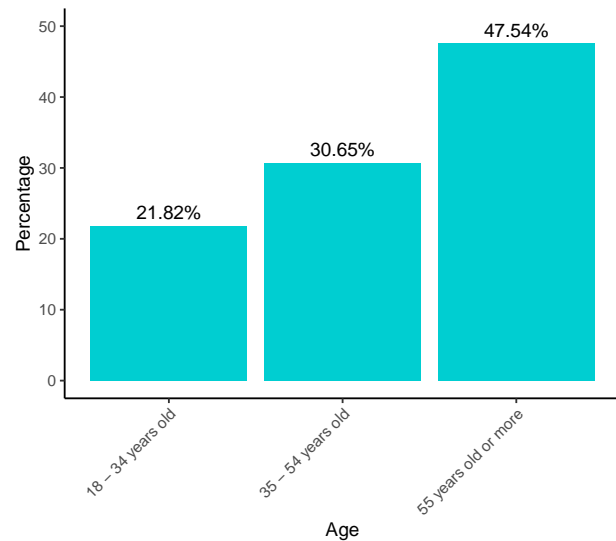


Figure 2.3: Distribution of Province in Survey Data

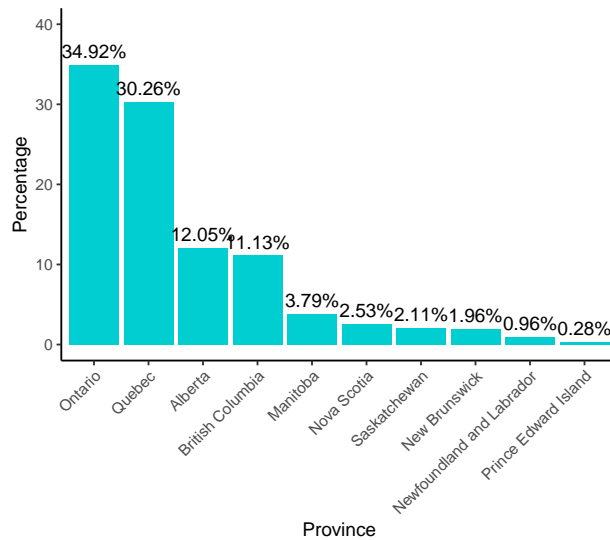


Figure 2.4: Distribution of Education Level in Survey Data

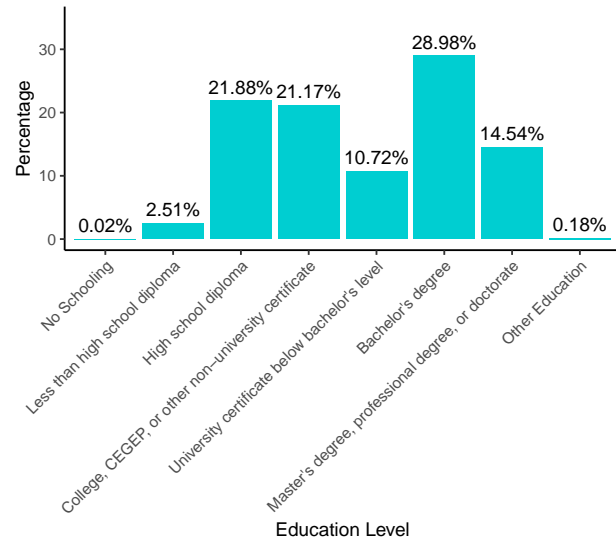


Figure 2.5: Distribution of Income Level in Survey Data

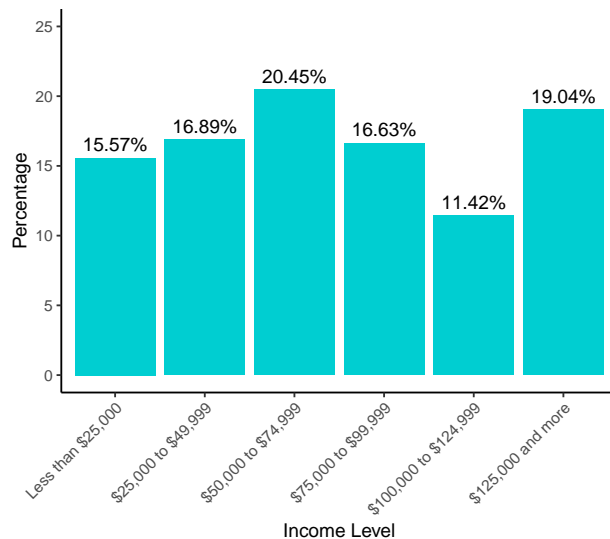
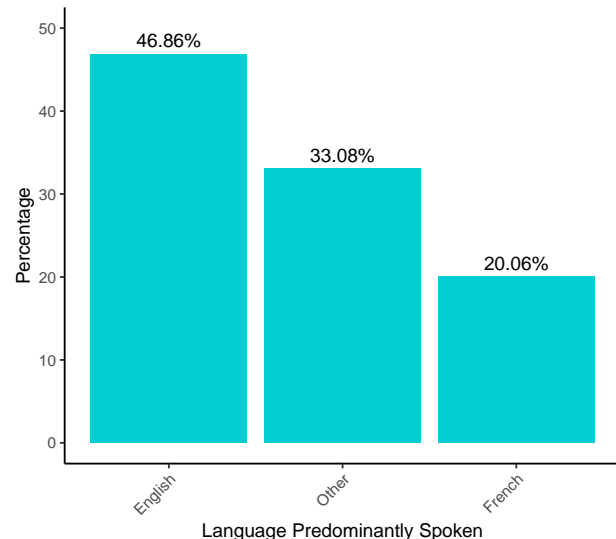


Figure 2.6: Distribution of Language Spoken in Survey Data



The cleaned sex variable is similarly coded as a categorical variable, holding two levels: Male and Female. From Figure 2.1, we can see that approximately 55% of survey respondents were male and 45% were female. Conversely, from Figure 4.1 in the Appendix, we found a similar proportion of male and female representation among the Canadian population. We can further observe this similarity in proportion in Figure 3.1, confirming a greater proportion of females overall.

The age variable is categorized into three distinct age groups: those between 18-34 years, indicating the youth, those between the ages 35-54 years, indicating the middle-aged, and those above the age of 54, indicating the elderly populous. From Figure 2.2, we can see that about 22% of survey respondents were youths, 31% were middle aged, and 47% were elderly. We observed a roughly similar distribution of age in our census data. From Figure 4.2 in our Appendix, we can explicitly see that roughly 19% were represented youths, 31% represented those middle aged, and a whopping 50% represented the elderly. While the proportion of those middle aged remain consistent, our census data appears to comprise of a slightly larger proportion of the elderly and a slightly smaller proportion of youths. From Figure 3.2, we can confirm that across both data sets, there exists a greater overall representation of the elderly population.

The province variable has been categorized according to the 10 different provinces of Canada, excluding the Territories. From Figure 2.3, we can see that a vast majority, about 65% of our survey respondents, are either residents of Ontario(34.92%) or Quebec(30.26%). Respondents from Alberta(12.05%) and British Columbia(11.13%), make up roughly 23% of the second tier majority. The remaining 12% of respondents in our survey are scattered across the other 6 provinces. Particularly, in order of highest representation to least in our survey data: we have Manitoba(3.79%), Nova Scotia (2.53%), Saskatchewan(2.11%), New Brunswick(1.96%), Newfoundland and Labrador(0.96%), and finally Prince Edward Island(0.28%).

Comparatively, we find from Figure 4.3 in our Appendix, that the differences in the distribution of population across the provinces are not as stark as we observed in our survey data. According to our census data, while Ontario still holds the majority of representation (27.18%), accompanied closely by Quebec(18.56%), it is followed not far off by British Columbia, with 12.27% representation, hence knocking Alberta(8.42%) out of its previously assumed third ranking majority position. The rest of the provinces hold a greater representation in the census data, roughly triple that of our survey data, and follow closely behind Alberta in rank; that is, in descending order, the representation goes as Nova Scotia(6.97%), New Brunswick(6.49%), Manitoba(5.8%), Saskatchewan(5.58%), Newfoundland and Labrador(5.31%), and Prince Edward island(3.42%). Figure 3.3 allows us to visualize this difference in distributions across both our data sets.

The education variable categorizes different degrees of education according to eight levels. The first level represents no schooling, the second, less than a high school diploma, the third, high school diploma, the fourth represents some college, CEGEP, or other non-university certificate, the fifth, university certificate below bachelor's degree, the sixth represents a bachelor's degree, followed by another level for master's degree, professional degree, or doctorate, and finally the last level categorizes other forms of attained education. From Figure 2.4, we can see that the majority of survey respondents have completed their bachelor's degree(28.98%). Further, respondents with a high school diploma, and those with a college, CEGEP or other non-university certificate form 22% of the total observations each. Roughly 15% have attained a some sort of master's degree, professional degree, or doctorate. Only 10.72% reach just shy of a bachelor's degree, while less than 3% either have not completed their high school degree, have not received any schooling, or have pursued other forms of education.

On the other hand, we find the general populous to be less educated on average compared to our survey respondents. In Figure 3.4, observe a greater proportion of survey respondents with higher educational attainments, indicative by a slight left skew. Whereas, we observe a slight right skew in educational attainments for our census data participants. Explicitly, from Figure 4.4 in our Appendix, we find a majority of the census populous to have an education level equivalent to a high school diploma(23.87%), followed closely by 22.63% possessing some level of college, CEGEP, or non-university certificate. Only 18.6% of the populous consists of individuals having completed a bachelor's degree, and only 9.13% have completed a doctorate, master's or professional degree. Whereas 13.13% of the population has less than a high school diploma, and 1.67% makeup those who received absolutely no schooling. Even 7.35% have completed some form of other education, and 3.62% possess a university certificate below the bachelor's level.



The income variable has similarly been categorized into six different income brackets: below 25,000 CAD, between 25,000 to 74,999 CAD, between 75,000 to 99,999 CAD, between 100,000 to 124,999 CAD, and above 125,000 CAD. From Figure 2.5, we can see that the distribution of income among survey respondents is roughly normal, with a peak at the highest income bracket (19.04%). Most survey respondents belong to the income bracket between 50,000 CAD to 74,999 CAD (20.45%). This makes sense as our survey respondents display higher education levels, stipulating the opportunity for higher income jobs. Our census data for the income bracket variable is similarly roughly normal around the 25,000 CAD to 49,000 CAD income bracket, with a peak at the highest income bracket. We can observe from Figure 4.5 in our Appendix that the majority of populous in our census data belong to the highest income bracket (22.44%), of \$125,000 or more. Figure 3.5 similarly displays this difference in income distribution, allowing us to visualize how respondents rates differ across income levels.

The language variable holds three categories, one to represent predominantly English speakers, one to represent predominantly French speakers, and finally one to represent predominant speakers in any other language. According to our survey data, from Figure 2.6, we can see 46.86% of survey respondents primarily speak English, while only 20.06% speak French. The Other category represents 33.08% of our survey respondents, indicating a diverse survey sample. Our census data on the other hand contains a vast majority of English speakers (73.57%), and 17.12% of French speakers, as we can see from Figures 3.6 and 4.6 in our Appendix. Consequently, only 9.31% of our populous predominantly speak in another language. We can undeniably view this difference in Figure 3.6 as the distribution of census data presents a right skew while the survey data presents rather a bimodal distribution, with high rates of English and other language speakers.

It is evident that while our survey encompasses extensive information about Canadians; it does not effectively reflect the true distribution in characteristics of the voting population in Canada. This is the underlying reason for the implementation of tools such as post-stratification.

Fig 3.1: Difference in Proportions of Sex

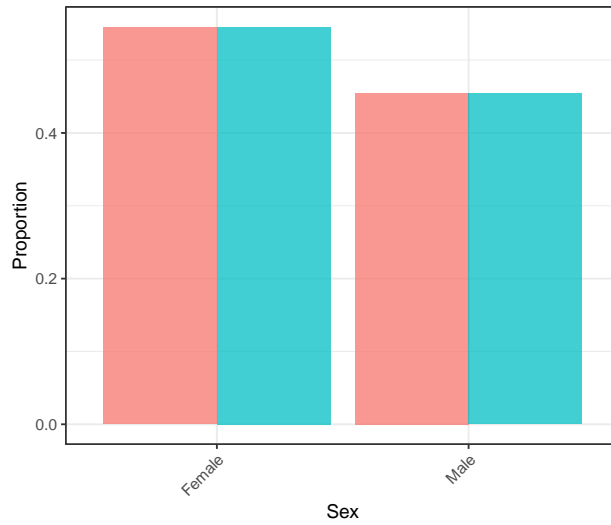


Fig 3.2: Difference in Proportions of Age

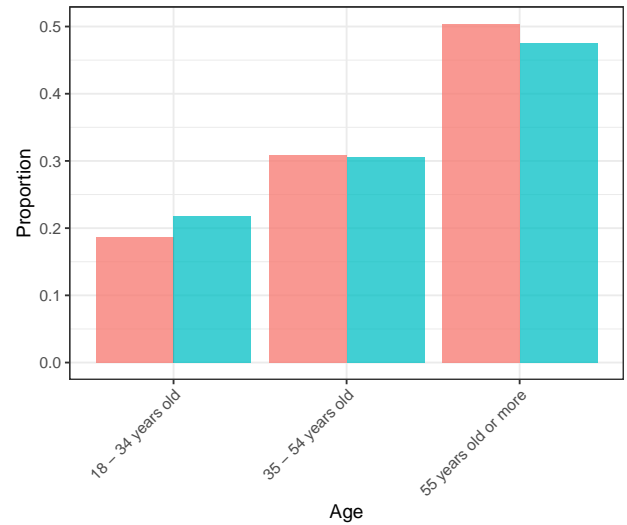


Fig 3.3: Difference in Proportions of Province Representation

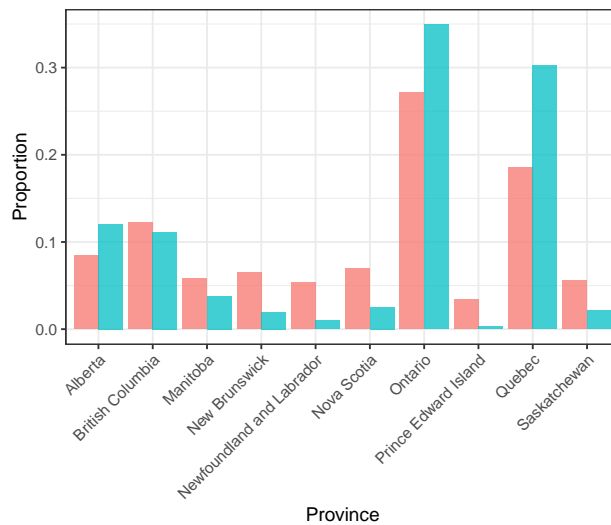


Fig 3.4: Difference in Proportions of Education Levels

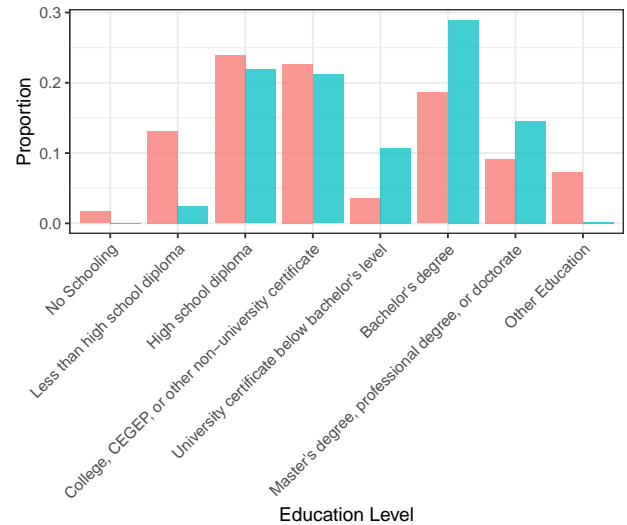


Fig 3.5: Difference in Proportions of Income Levels

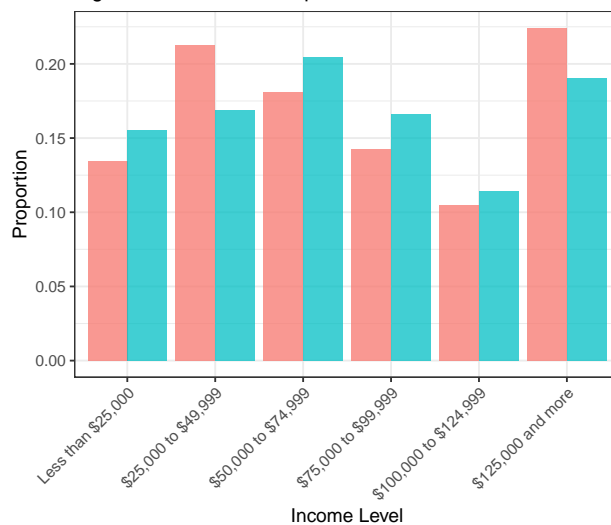
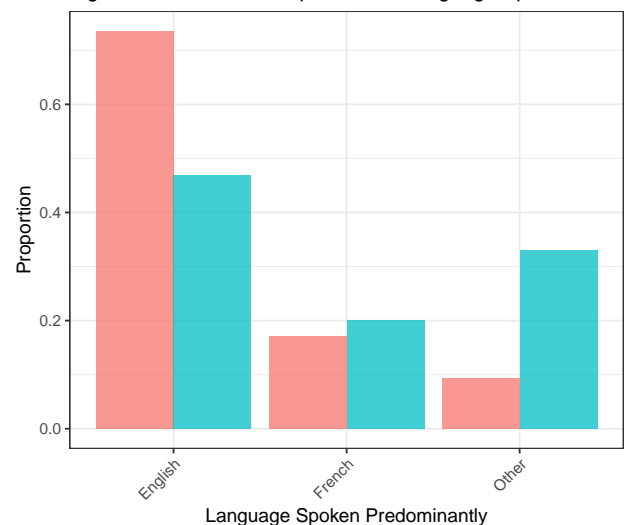


Fig 3.6: Difference in Proportions of Language Spoken



Data: ■ Census prop. ■ Survey prop.

## 2.5 Preliminary Analysis

Figure 5.1: Diversity in Party Favorability Across Sexes

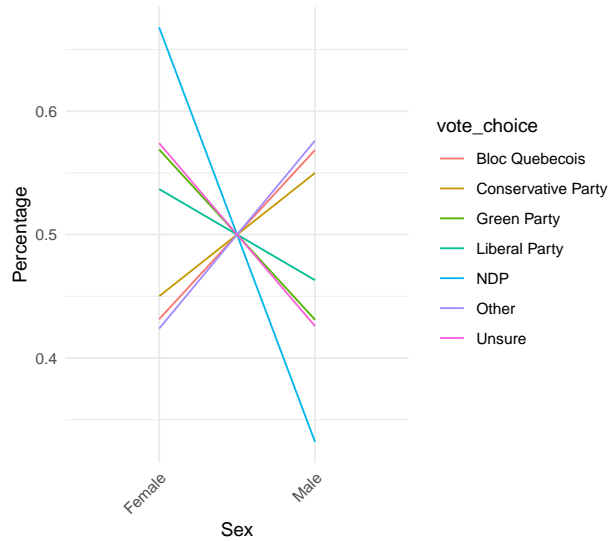


Figure 5.2: Diversity in Party Favorability Across Age Groups

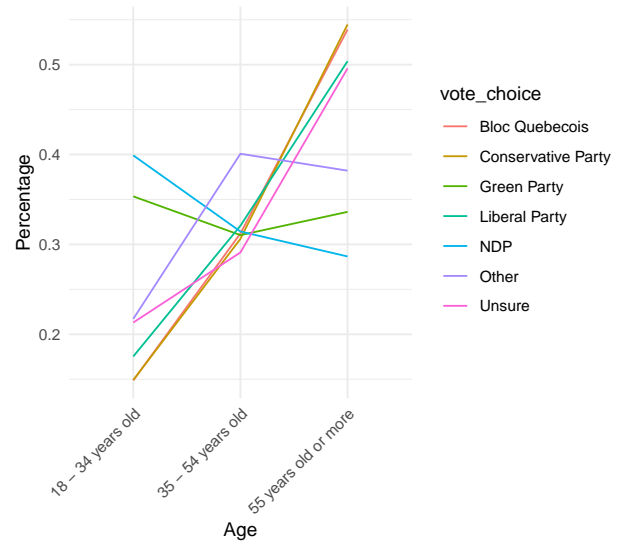


Figure 5.3: Diversity in Party Favorability Across Provinces

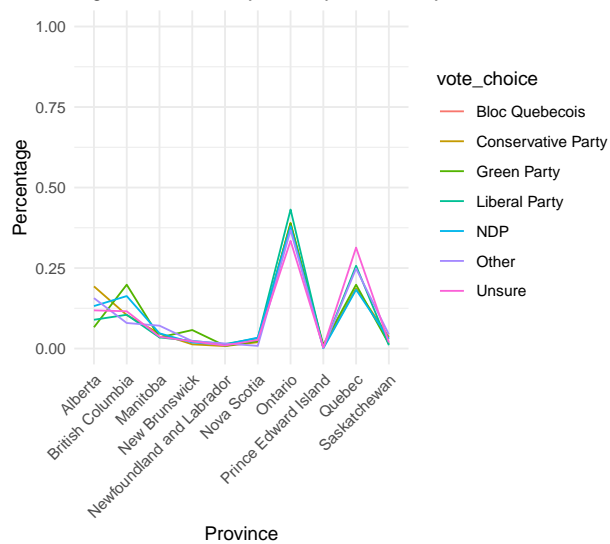


Figure 5.4: Diversity in Party Favorability Across Income Levels

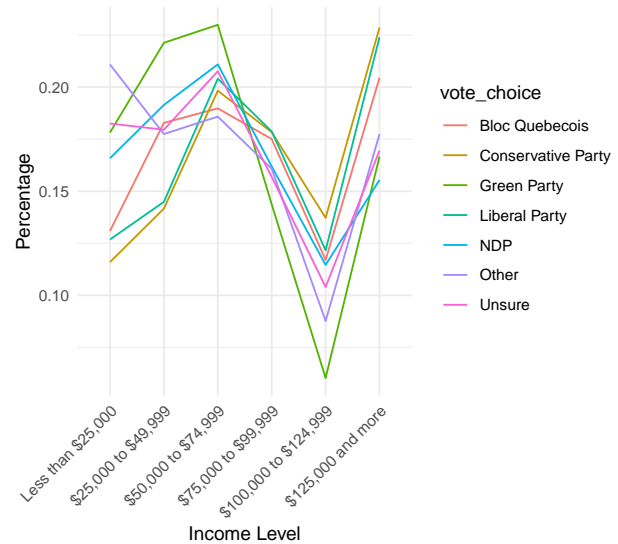


Figure 5.6: Diversity in Party Favorability Across Different Language Speakers

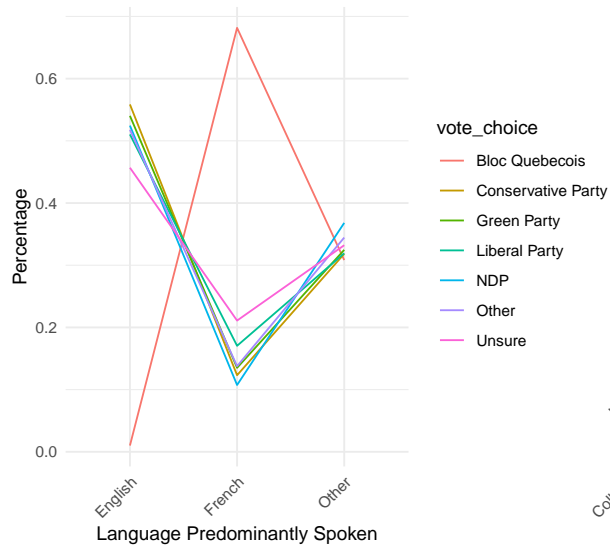
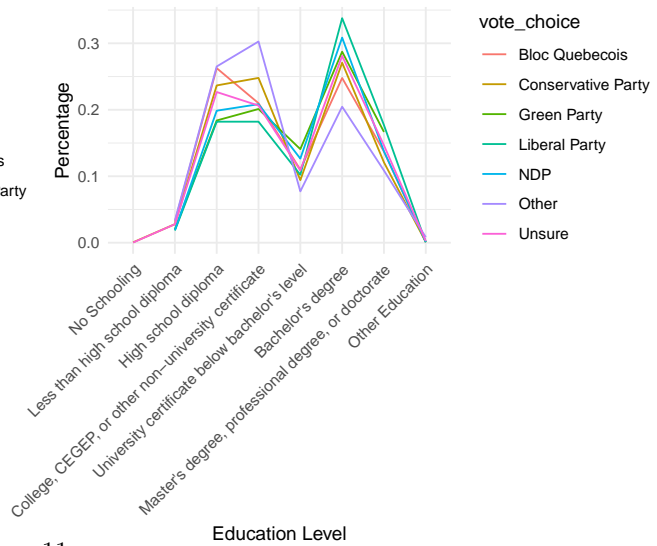


Figure 5.5: Diversity in Party Favorability Across Education Levels



Before we move on to creating our model, we can conduct some preliminary analysis on the distribution of voting party preference across the various demographic variables. With aid from the plethora of plots from Figures 5.1 to 5.6, we confirm some of our previously suggested hypotheses based off of our survey data.

From Figure 5.1, we can see that on average, males in our survey are either unsure, support some other political party, or favor the Conservative party. Females on the other hand appear to largely favor the NDP, followed by strong preference for Bloc Quebecois, the Liberal Party, and the Green Party.

From Figure 5.2, we observe the elderly populous to largely favor the Conservative Party, Bloc Quebecois, or to be unsure of their choice. This is followed by moderately strong preference among the greater than 55 years age group towards the Liberal Party. The younger age bracket appears to favor the NDP and Green Party, while the middle-aged bracket favor other existing political parties instead.

From Figure 5.3, we do not observe much of a preference for any political party according to province. We do, however, get a glimpse of the provinces of Ontario and Quebec signifying a majority of the survey population from the two significant peaks in our graph.

From Figure 5.4, we find that individuals in the highest income bracket tend to have the greatest favor for the Conservative party, while those in the lowest income bracket tend to have the greatest preference for other political parties not currently represented in the House of Commons. Generally, for individuals with income levels below approximately \$75,000, preference for the Green Party appears to be predominant. Particularly, we observe a significant rise and fall in Green Party favor as income levels rise.

In Figure 5.5, we observe two significant peaks. Individuals with higher education levels, on average tend to prefer the Liberal Party, followed by the NDP. Individuals with lower education levels on the other hand, appear to favor other existing political parties, followed by the Bloc Quebecois, and the Conservative party in particular.

Finally in Figure 5.6, we observe a distinctly high preference for Bloc Quebecois among predominantly French speakers, as we had previously hypothesized. For primarily English speakers, or those of other languages, there does not seem to be any particular preference for political party.

As a note, we've included Figures 5.1 to 5.6 in our Appendix for a more detailed look at each graph.

### 3 Methods

As we've seen, while expansive, our survey data is not entirely reflective of our target population, as represented by our census data. Thus we implement techniques such as logistic regression post stratification in order to ensure our sample is representative of our target population. We employ such techniques by fitting our hypothesized logistic model to our survey data and adjusting the weights of each cell respective to distributions in our census data.

#### 3.1 Model Specifics

In this paper, we employ binomial logistic models (under generalized linear models) to fit our survey data and predict estimated probabilities of political party favor using the underlying logit function. The logit function, also known as the log-odds, is the logarithm of odds. We rationalize the implementation of this particular model by confirming the assumptions associated with logistic regression modelling.

As mentioned in the Data section, we had created distinct data sets for our political parties of interest, where each consisted of a binary variable that presented as 1 for the party favor of the respective data set, and 0 otherwise. As the party favored by respondents is our outcome of interest, we can confirm the first assumption of logistic regression: the outcome is binary. The second assumption of logistic regression does not apply in our procedure. While linearity is required in the logit for continuous variables, all of our predictor variables are categorical. Hence we disregard this assumption and move on to the third: checking for absence of multicollinearity. Multicollinearity is the correlation between predictor variables in a model. We use tools such as the variance inflation factor (VIF), which measures the extent of multicollinearity between predictors

in a model and indicates potential cases of multicollinearity issues as values of either 5 or 10. In Tables 3.1 to 3.4 in our Appendix, we observe the variance inflated values of each of our models and consequently found no variance inflated value greater than 3, allowing us to conclude an absence of multicollinearity in our models. Finally, in Figures 6.1 to 6.4, we plot the Cook's Distance for each of our models to confirm the final logistic model assumption: lack of strongly influential outliers. While we do observe some outliers in our model, none of them exceed a Cook's Distance of 2, allowing us to deduce that while outliers may exist, they are not strongly influential. Hence, having confirmed the assumptions of logistic regression modelling, we can begin explaining the set-up of our models.

Utilizing the party specific data sets we created, we generated four different logistic regression models, with the binary variable representing voting preference as our outcome variable and each of our categorical variables, age, sex, income, education, province, and language spoken at home, representing our predictor variables. It must be noted that of the five political parties we chose to focus on, the Liberal Party, Conservative Party, New Democratic Party, Green Party, and Bloc Quebecois, we only build models for the first four. This is due to an error in perfect separation in the logistic regression model concerning the Bloc Quebecois party. Perfect separation arises as a result of a predictor variable, or a combination of predictor variables perfectly separating the outcome variable into distinct groups, thus leading to infinite parameter estimates [11]. It is likely we faced such an error due to the high proportions of Bloc Quebecois party preference by French speakers and Quebecois residents. Thus we exclude Bloc Quebecois from our analysis.

Moving forward, we notate our multiple binomial logistic regression models as follows:

$$\begin{aligned} \log\left(\frac{p_{\text{liberal}}}{1 - p_{\text{liberal}}}\right) = & \beta_0 + \beta_1 x_{\text{age}35-54\text{YearsOld}} + \beta_2 x_{\text{age}55\text{YearsOldOrMore}} + \beta_3 x_{\text{provinceBritishColumbia}} + \\ & \beta_4 x_{\text{provinceManitoba}} + \beta_5 x_{\text{provinceNewBrunswick}} + \beta_6 x_{\text{provinceNewfoundlandandLabrador}} + \\ & \beta_7 x_{\text{provinceNovaScotia}} + \beta_8 x_{\text{provinceOntario}} + \beta_9 x_{\text{provincePrinceEdwardIsland}} + \beta_{10} x_{\text{provinceQuebec}} + \\ & \beta_{11} x_{\text{provinceSaskatchewan}} + \beta_{12} x_{\text{sexMale}} + \beta_{13} x_{\text{educationCollege,CEGEP,orothernon-universitycertificate}} + \\ & \beta_{14} x_{\text{educationHighschooldiploma}} + \beta_{15} x_{\text{educationLessthanhighschooldiploma}} + \\ & \beta_{16} x_{\text{educationMaster'sdegree,professionaldegree,ordrctorate}} + \beta_{17} x_{\text{educationNoSchooling}} + \beta_{18} x_{\text{educationOtherEducation}} + \\ & \beta_{19} x_{\text{educationUniversitycertificatebelowbachelor'slevel}} + \beta_{20} x_{\text{income125,000CADandmore}} + \beta_{21} x_{\text{income25,000CADto49,999CAD}} + \\ & \beta_{22} x_{\text{income50,000CADto74,999CAD}} + \beta_{23} x_{\text{income75,000CADto99,999CAD}} + \beta_{24} x_{\text{incomeLessthan25,000CAD}} + \\ & \beta_{25} x_{\text{languageFrench}} + \beta_{26} x_{\text{languageOther}} \end{aligned}$$

Where  $p_{\text{liberal}}$  represents the probability of voting for the Liberal party,  $\beta_0$  represents the log-odds of voting for the liberal party for the reference levels of the categorical predictors,  $\beta_i$ , for  $i > 0$ , represents the change in log odds of voting liberal for every one unit increase in the respective predictors relative to the reference category for that specific predictor. So for example,  $\beta_{12}$  can be interpreted as the change in log odds of voting for the liberal party for males compared females (the reference category).

Tables 5.1 - 5.4 in the Results section showcase the summary for our multiple binomial logistic regression model estimates.

$$\begin{aligned} \log\left(\frac{p_{\text{conservative}}}{1 - p_{\text{conservative}}}\right) = & \beta_0 + \beta_1 x_{\text{age}35-54\text{YearsOld}} + \beta_2 x_{\text{age}55\text{YearsOldOrMore}} + \beta_3 x_{\text{provinceBritishColumbia}} + \\ & \beta_4 x_{\text{provinceManitoba}} + \beta_5 x_{\text{provinceNewBrunswick}} + \beta_6 x_{\text{provinceNewfoundlandandLabrador}} + \\ & \beta_7 x_{\text{provinceNovaScotia}} + \beta_8 x_{\text{provinceOntario}} + \beta_9 x_{\text{provincePrinceEdwardIsland}} + \beta_{10} x_{\text{provinceQuebec}} + \\ & \beta_{11} x_{\text{provinceSaskatchewan}} + \beta_{12} x_{\text{sexMale}} + \beta_{13} x_{\text{educationCollege,CEGEP,orothernon-universitycertificate}} + \end{aligned}$$

$$\begin{aligned}
& \beta_{14}x_{educationHighschoolDiploma} + \beta_{15}x_{educationLessthanhighschoolDiploma} + \\
& \beta_{16}x_{educationMaster'sdegree,professionaldegree,ordoctorate} + \beta_{17}x_{educationNoSchooling} + \beta_{18}x_{educationOtherEducation} + \\
& \beta_{19}x_{educationUniversitycertificatebelowbachelor'slevel} + \beta_{20}x_{income125,000CADandmore} + \beta_{21}x_{income25,000CADto49,999CAD} + \\
& \beta_{22}x_{income50,000CADto74,999CAD} + \beta_{23}x_{income75,000CADto99,999CAD} + \beta_{24}x_{incomeLessthan25,000CAD} + \\
& \beta_{25}x_{languageFrench} + \beta_{26}x_{languageOther}
\end{aligned}$$

Where  $p_{conservative}$  represents the probability of voting for the Conservative party, and the  $\beta_i$ s represent the same values as in the liberal logistic regression model, only respective to the conservative party in this case.

$$\begin{aligned}
\log\left(\frac{p_{ndp}}{1 - p_{ndp}}\right) = & \beta_0 + \beta_1x_{age35-54YearsOld} + \beta_2x_{age55YearsOldOrMore} + \beta_3x_{provinceBritishColumbia} + \\
& \beta_4x_{provinceManitoba} + \beta_5x_{provinceNewBrunswick} + \beta_6x_{provinceNewfoundlandandLabrador} + \\
& \beta_7x_{provinceNovaScotia} + \beta_8x_{provinceOntario} + \beta_9x_{provincePrinceEdwardIsland} + \beta_{10}x_{provinceQuebec} + \\
& \beta_{11}x_{provinceSaskatchewan} + \beta_{12}x_{sexMale} + \beta_{13}x_{educationCollege,CEGEP,orothernon-universitycertificate} + \\
& \beta_{14}x_{educationHighschoolDiploma} + \beta_{15}x_{educationLessthanhighschoolDiploma} + \\
& \beta_{16}x_{educationMaster'sdegree,professionaldegree,ordoctorate} + \beta_{17}x_{educationNoSchooling} + \beta_{18}x_{educationOtherEducation} + \\
& \beta_{19}x_{educationUniversitycertificatebelowbachelor'slevel} + \beta_{20}x_{income125,000CADandmore} + \beta_{21}x_{income25,000CADto49,999CAD} + \\
& \beta_{22}x_{income50,000CADto74,999CAD} + \beta_{23}x_{income75,000CADto99,999CAD} + \beta_{24}x_{incomeLessthan25,000CAD} + \\
& \beta_{25}x_{languageFrench} + \beta_{26}x_{languageOther}
\end{aligned}$$

Where  $p_{ndp}$  represents the probability of voting for the NDP party, and the  $\beta_i$ s represent the same values as in the liberal logistic regression model, only respective to the NDP party in this case.

$$\begin{aligned}
\log\left(\frac{p_{green}}{1 - p_{green}}\right) = & \beta_0 + \beta_1x_{age35-54YearsOld} + \beta_2x_{age55YearsOldOrMore} + \beta_3x_{provinceBritishColumbia} + \\
& \beta_4x_{provinceManitoba} + \beta_5x_{provinceNewBrunswick} + \beta_6x_{provinceNewfoundlandandLabrador} + \\
& \beta_7x_{provinceNovaScotia} + \beta_8x_{provinceOntario} + \beta_9x_{provincePrinceEdwardIsland} + \beta_{10}x_{provinceQuebec} + \\
& \beta_{11}x_{provinceSaskatchewan} + \beta_{12}x_{sexMale} + \beta_{13}x_{educationCollege,CEGEP,orothernon-universitycertificate} + \\
& \beta_{14}x_{educationHighschoolDiploma} + \beta_{15}x_{educationLessthanhighschoolDiploma} + \\
& \beta_{16}x_{educationMaster'sdegree,professionaldegree,ordoctorate} + \beta_{17}x_{educationNoSchooling} + \beta_{18}x_{educationOtherEducation} + \\
& \beta_{19}x_{educationUniversitycertificatebelowbachelor'slevel} + \beta_{20}x_{income125,000CADandmore} + \beta_{21}x_{income25,000CADto49,999CAD} + \\
& \beta_{22}x_{income50,000CADto74,999CAD} + \beta_{23}x_{income75,000CADto99,999CAD} + \beta_{24}x_{incomeLessthan25,000CAD} + \\
& \beta_{25}x_{languageFrench} + \beta_{26}x_{languageOther}
\end{aligned}$$

Where  $p_{green}$  represents the probability of voting for the Green party, and the  $\beta_i$ s represent the same values as in the liberal logistic regression model, only respective to the green party in this case.

### 3.2 Post-Stratification

Post-stratification is a statistical technique that occurs after data collection, used to adjust estimates obtained from a sample to match a known population’s characteristics, when the sample itself is not sufficiently representative of the target population. This technique makes for an adequate solution for non-probability based sampling.

Previously, we had divided both our population (GSS census data) and sample (CES survey data) into various strata based on key demographic features, particularly variables representing age, sex, educational attainment levels, income brackets, province of residence, and language predominantly spoken. Following, we obtained estimates for our response variables from our sample data for each cell with the help of our multiple binomial logistic regression model.

The next step in this process entailed the comparison of sample estimates with known population totals for each stratum, and the modification of the weights of these estimates for each stratum to accurately match the population totals. We examine the process of the rest of this procedure as conducted with our binomial regression model for estimating liberal party favor likelihood.

In particular, we first grouped our census data across all the demographic variables of interest, raking through and calculating the total census counts of each cell. Using this grouped data set, we used our previous binomial logistic model to predict the estimated response for each combination of the categories in our demographic variables, which we plugged back into our data set under a new column called ‘estimate’. For each row of predictor variable combinations, we multiply the total population counts for that particular combination to the respective estimates in order to obtain the predicted proportion of the response for each combination. Next we divide the sum of these predicted proportions by the total population counts in order to attain the overall estimated proportions of voting liberal across our entire data set. Taking the average of these proportions, we obtain our final estimate for the mean predicted proportion voting liberal for individuals across all the groups in our data set.

We similarly conducted this procedure for the census data using our models for the conservative, new democratic and green party. Thus allowing us to compute the expected probability of voting for each of these parties.

All analysis for this report was programmed using R version 4.0.2. [1][2][3]

## 4 Results

Table 5.1: Liberal Vote Predictor Model

```
## # A tibble: 27 x 5
##   term                estimate std.error statistic  p.value
##   <chr>              <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)        -1.33    0.0967   -13.8  3.14e-43
## 2 age35 - 54 years old    0.401   0.0582    6.90 5.10e-12
## 3 age55 years old or more  0.585   0.0561   10.4 1.59e-25
## 4 provinceBritish Columbia 0.389   0.0862    4.52 6.33e- 6
## 5 provinceManitoba        0.286   0.120     2.38 1.74e- 2
## 6 provinceNew Brunswick    0.938   0.153     6.15 7.79e-10
## 7 provinceNewfoundland and Labrador 1.12    0.198     5.68 1.35e- 8
## 8 provinceNova Scotia      0.908   0.134     6.79 1.14e-11
## 9 provinceOntario         0.717   0.0693   10.3 4.44e-25
## 10 provincePrince Edward Island 0.663   0.397     1.67 9.48e- 2
## # i 17 more rows
```

From Table 5.1, we observe the estimated response results for our logistic model on liberal party voting favor. The coefficient on the age35-54 years old category indicates individuals within the 35-54 age range

have a 0.401 higher log odds of voting for the liberal party compared to individuals within the 18-34 age range. More intuitively, we find an odds ratio of 0.82 and infer that individuals in the middle aged category have a slightly lower odds of voting for the liberal party compared to youths.

Surprisingly, we find an odds ratio of 1.18 for the over 55 years of age category, entailing that elders have slightly higher odds of voting for the liberal party compared to youths. Further, we find an odds ratio of 1.54 for the British Columbia province category, indicating residents of British Columbia are far more likely to vote liberal compared to residents of Alberta.

Table 5.2: Conservative Vote Predictor Model

```
## # A tibble: 27 x 5
##   term                estimate std.error statistic  p.value
##   <chr>              <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)        -0.692    0.0941    -7.35  1.99e-13
## 2 age35 - 54 years old  0.405    0.0625     6.48  9.28e-11
## 3 age55 years old or more  0.665    0.0599    11.1  1.09e-28
## 4 provinceBritish Columbia -0.799    0.0815    -9.80  1.16e-22
## 5 provinceManitoba      -0.499    0.112     -4.48  7.56e- 6
## 6 provinceNew Brunswick -1.13     0.177     -6.38  1.81e-10
## 7 provinceNewfoundland and Labrador -1.01    0.226     -4.47  7.69e- 6
## 8 provinceNova Scotia   -1.01     0.148     -6.86  7.13e-12
## 9 provinceOntario       -0.598    0.0623    -9.59  8.85e-22
## 10 provincePrince Edward Island -0.254    0.378     -0.672 5.02e- 1
## # i 17 more rows
```

Similarly, we can highlight another coefficient from the summary of our logistic model fit on predicting the probability of voting for the conservative party. For an odds ratio of 1.39 (log odds 0.665), we can infer that the elderly population is more likely to vote for the conservative party compared to youths. Likewise, we can make inferences using the rest of the coefficients in our logistic models.

Table 5.3: NDP Vote Predictor Model

```
## # A tibble: 27 x 5
##   term                estimate std.error statistic  p.value
##   <chr>              <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)        -0.234    0.102     -2.30  2.14e- 2
## 2 age35 - 54 years old -0.756    0.0576   -13.1  2.23e- 39
## 3 age55 years old or more -1.40     0.0595   -23.6  4.08e-123
## 4 provinceBritish Columbia  0.469    0.0876     5.36  8.37e- 8
## 5 provinceManitoba        0.181    0.125     1.45  1.47e- 1
## 6 provinceNew Brunswick   -0.0349   0.180     -0.194 8.46e- 1
## 7 provinceNewfoundland and Labrador  0.104    0.223     0.466 6.41e- 1
## 8 provinceNova Scotia      0.304    0.145     2.10  3.58e- 2
## 9 provinceOntario         -0.0908   0.0729    -1.25  2.13e- 1
## 10 provincePrince Edward Island -1.53     0.748     -2.04  4.09e- 2
## # i 17 more rows
```

Table 5.4: Green Vote Predictor Model

```
## # A tibble: 27 x 5
##   term                estimate std.error statistic  p.value
##   <chr>              <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)        -4.40     0.323   -13.6  3.56e-42
```



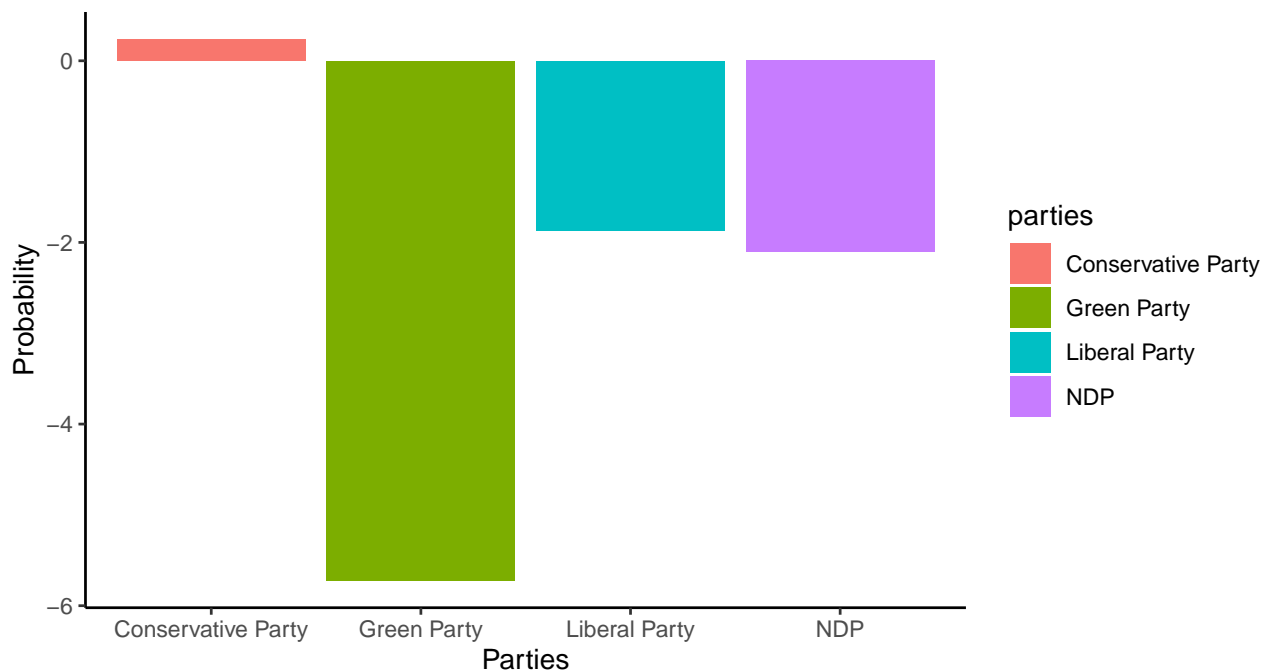
```
## 2 age35 - 54 years old          -0.427    0.140    -3.05  2.26e- 3
## 3 age55 years old or more      -0.872    0.142    -6.16  7.49e-10
## 4 provinceBritish Columbia     1.22     0.245     5.00  5.68e- 7
## 5 provinceManitoba              0.518    0.361     1.43  1.51e- 1
## 6 provinceNew Brunswick         1.83     0.322     5.69  1.30e- 8
## 7 provinceNewfoundland and Labrador 0.429    0.624     0.687 4.92e- 1
## 8 provinceNova Scotia           0.499    0.417     1.20  2.32e- 1
## 9 provinceOntario              0.664    0.228     2.91  3.60e- 3
## 10 provincePrince Edward Island  2.32     0.585     3.97  7.23e- 5
## # i 17 more rows
```

Thus from our logistic regression models, we obtain the predicted probabilities of voting for a specific political party among our survey respondents.

```
##
##
## Table: 6: Predicted Mean Proportion of Party Favor
##
## |Voting Party Choice |Mean Predicted Proportion |
## |:-----|:-----|
## |Liberal Party      |-1.8661                    |
## |Conservative Party |0.239                      |
## |NDP                |-2.1053                    |
## |Green Party        |-5.7241                    |
```

From our post stratification, we obtain the mean predicted proportion of voting for either political party. These estimates are more precise as they are weighed according to proportions in our target population. From our results, as displayed in Table 6, we observe a positive predicted estimation for the proportion of individuals voting for the Conservative party. The rest of our models granted negative predicted proportions which are not reasonable results.

Figure 6: Differences in Average Predicted Voting Proportion between Parties



Accordingly, in Figure 6, we can visualize the magnitude of these results. While the results of the remaining parties no longer bring logical results, we focus on the results for the conservative predicted proportion. In this context, we could conclude that our target population has the greatest possibility of voting for the conservative party. Essentially, this would make sense as we had previously seen in Figure 3.2, that our true target population consists of a greater proportion of an elderly population in comparison to our sample data. We had hypothesized earlier that older populations tend to vote more conservative [12]. Further, we had hypothesized a greater proportion of conservative voters having lower levels of educational attainment in comparison to liberal voters [13]. In Figure 3.4, we observed a greater proportion of our target population having lower levels of education compared to our sample population, i.e. our survey respondents.

## 5 Conclusions

As we had initially hypothesized, individuals from an older age bracket are more likely to share conservative, traditional views and thus should potentially have a higher voting potential for the Conservative Party. Further, individuals of lower educational attainment levels are more likely to favor traditional views in place of the progressive views put forth by the Liberal party and are thus more likely to vote for the liberal party. The inverse goes for the Liberal party. In Figures 5.1 through 5.6, we observed higher proportions of voting favour for the Liberal Party among younger individuals and those with higher education levels. However, the estimations in Figures 5.1 to 5.6 were not accurately representative of our target population. Thus while our analyses contained meaningful interpretations, they did not apply to our target population. This was due to a mis-representation of the target population in our sample. We observed a greater proportion of individuals in the older age groups and a greater proportion of individuals with lower levels of income in our population data. As a result, when we conducted our post-stratification regression, to infer precise estimates based off of our unrepresentative data, we obtained results that indicated that the party with the highest likelihood of attaining votes were the conservative party. Essentially according to our hypotheses, as we are aware of the over representation of older and less educated individuals in our target population, we are not surprised to obtain such results.

However, it must be stated that limitations exist to our model. Our survey data was conducted in 2021, while our census data was extracted from 2017. This discrepancy in time may be incurring some bias in our predictions. Further, characteristics change over time and our survey data may be based on outdated population characteristics as the predictions we are trying to make are for the 2025 federal election, approximately 4 years apart from our survey data, and a further 8 years away from our population data. It is highly likely that changing characteristics in our population are causing limitations and inaccuracies in our estimations.

Moving forth, we could work on strategies to mitigate such biases in our data. Further, we could refine the stratum we used for post-stratification if possible, ensuring more meaning in our strata. If anything, we should continue further research and attempt to use as much updated data as possible in order to strive for continuous improvement in our results and findings.

## 6 Bibliography

1. Grolemond, G. (2014, July 16) *Introduction to R Markdown*. RStudio. [https://rmarkdown.rstudio.com/articles\\_intro.html](https://rmarkdown.rstudio.com/articles_intro.html). (Last Accessed: April 4, 1991)
2. RStudio Team. (2020). *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.
3. Allaire, J.J., et. el. *References: Introduction to R Markdown*. RStudio. <https://rmarkdown.rstudio.com/docs/>. (Last Accessed: April 4, 1991)
4. Elections Canada. (2019). Elections Step by Step | Elections Canada's Civic Education. Election-sanddemocracy.ca; Elections Canada. <https://electionsanddemocracy.ca/canadas-elections/canadas-election-process/elections-step-step>
5. House of Commons. (2017). The Canadian Parliamentary System - Our Procedure - ProceduralInfo - House of Commons of Canada. [Www.ourcommons.ca](http://www.ourcommons.ca). [https://www.ourcommons.ca/procedure/our-procedure/ParliamentaryFramework/c\\_g\\_parliamentaryframework-e.html](https://www.ourcommons.ca/procedure/our-procedure/ParliamentaryFramework/c_g_parliamentaryframework-e.html)
6. Elections Canada. (2019). Elections Step by Step | Elections Canada's Civic Education. Election-sanddemocracy.ca; Elections Canada. <https://electionsanddemocracy.ca/canadas-elections/canadas-election-process/elections-step-step>
7. Christian, W., & Jansen, H. (2006, February 7). Party System | The Canadian Encyclopedia. [Thecanadianencyclopedia.ca](http://www.thecanadianencyclopedia.ca). <https://www.thecanadianencyclopedia.ca/en/article/party-system>
8. Parliament of Canada. (2016). Political Parties and Leaders. [Parl.ca](http://lop.parl.ca). [https://lop.parl.ca/sites/ParlInfo/default/en\\_CA/Parties/politicalPartiesLeaders](https://lop.parl.ca/sites/ParlInfo/default/en_CA/Parties/politicalPartiesLeaders)
9. Government of Canada, D. of J. (2020, September 30). How new laws and regulations are created. [Www.justice.gc.ca](http://www.justice.gc.ca). <https://www.justice.gc.ca/eng/laws-lois/index.html>
10. Government of Canada, S. C. (2016, December 20). General Social Survey - Family (GSS). [Www23.Statcan.gc.ca](http://www23.statcan.gc.ca). <https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=4501>
11. Ghosh, S. (2021, October 2). Class Separation cannot be overlooked in Logistic Regression. [Geek Culture](https://medium.com/geekculture/class-separation-cannot-be-overlooked-in-logistic-regression-f20e58b203eb). <https://medium.com/geekculture/class-separation-cannot-be-overlooked-in-logistic-regression-f20e58b203eb>
12. Geys, B., Heggedal, T.-R., & Sørensen, R. J. (2022). Age and vote choice: Is there a conservative shift among older voters? *Electoral Studies*, 78, 102485. <https://doi.org/10.1016/j.electstud.2022.102485>
13. Renfrew, M. (2023, August 16). Support for the Conservative Party of Canada decreases with education; increases with age, income. [Cult MTL](https://cultmtl.com). <https://cultmtl.com/2023/08/support-for-the-conservative-party-of-canada-decreases-with-education-increases-with-age-income/>

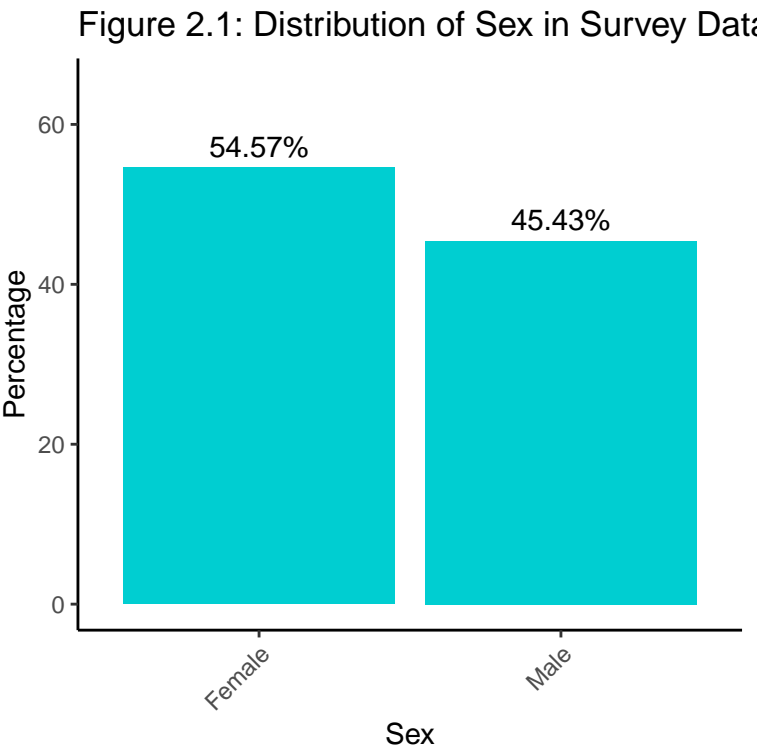
7 Appendix

7.1 Generative AI Statement

I did not have to use Generative AI tool(s) for this paper.

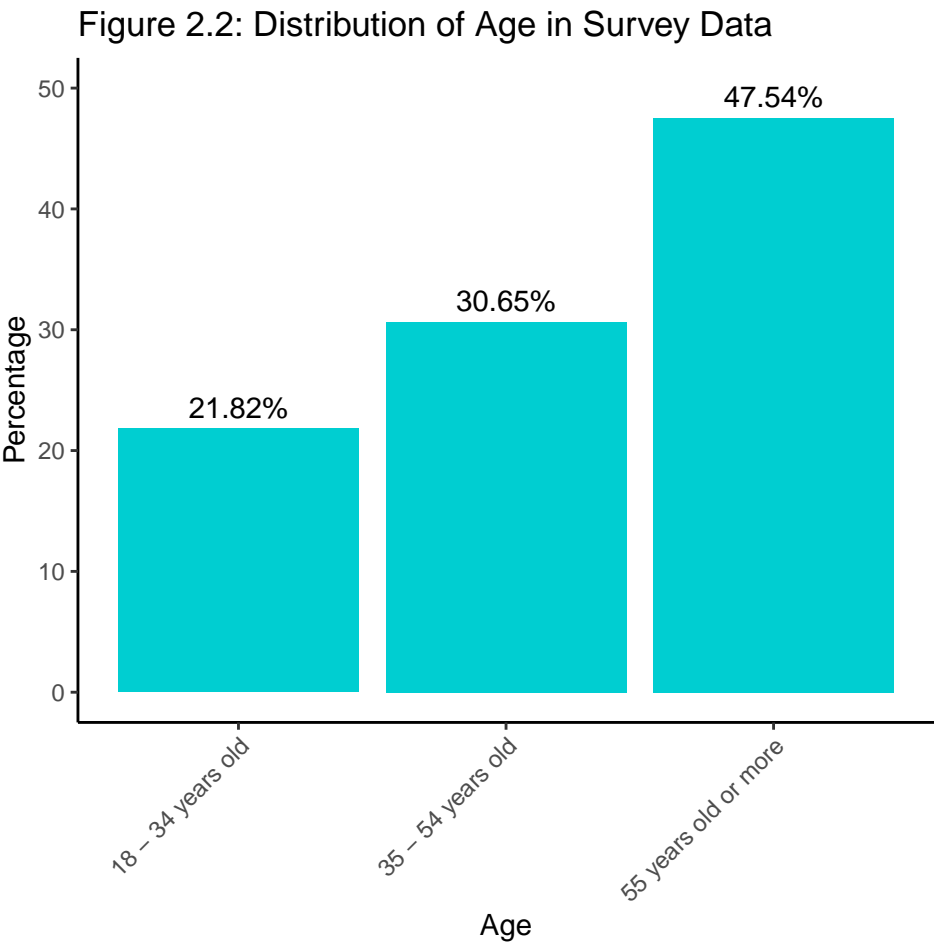
7.2 Supplementary Materials

Figure 2.1 & Table 2.1



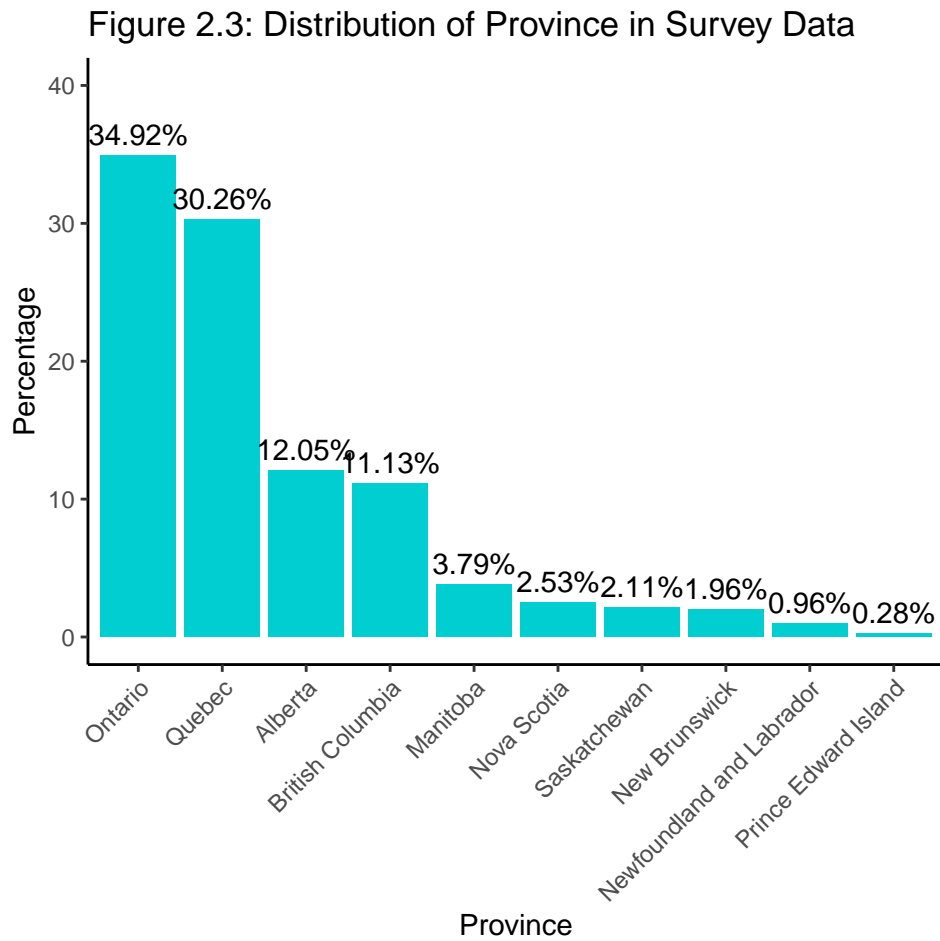
```
##
##
## Table: 2.1: Distribution of Sex in Survey Data
##
## |Sex      | Percentage|
## |:-----|:-----:|
## |Female  |      54.57|
## |Male    |      45.43|
```

Figure 2.2 & Table 2.2



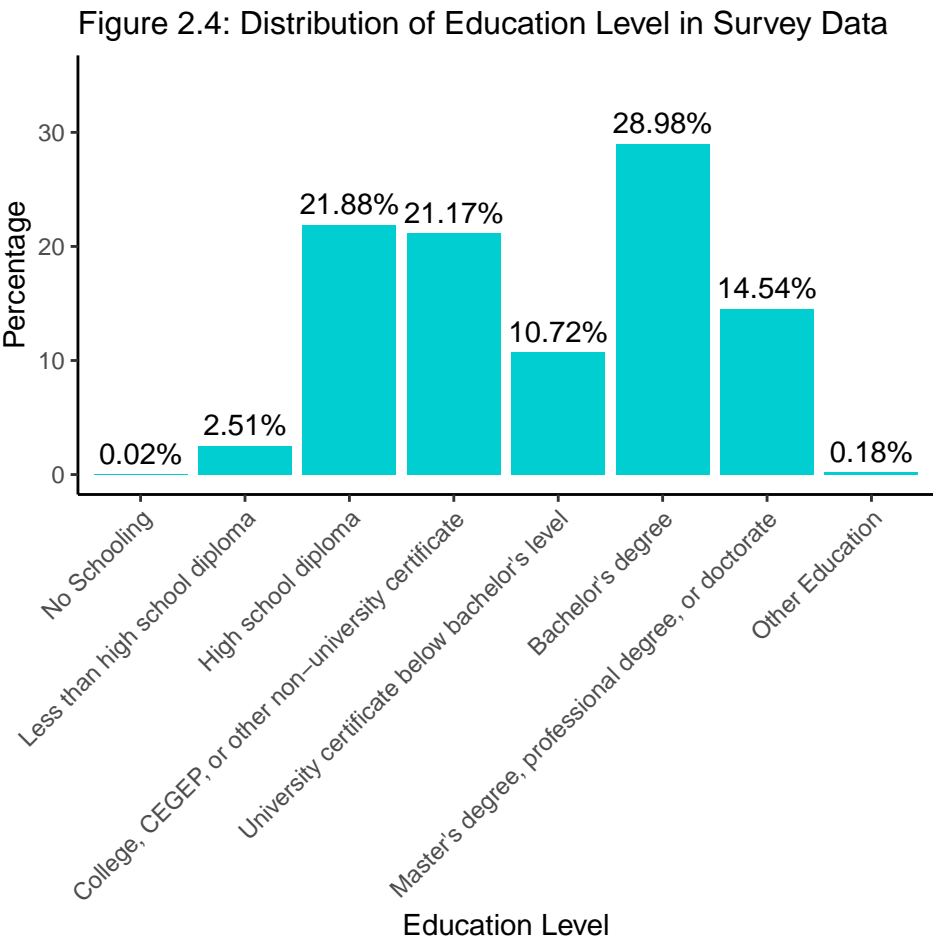
```
##
##
## Table: 2.2: Distribution of Age in Survey Data
##
## |Age                | Percentage|
## |:-----|:-----:|
## |55 years old or more |      47.54|
## |35 - 54 years old    |      30.65|
## |18 - 34 years old    |      21.82|
```

Figure 2.3 & Table 2.3



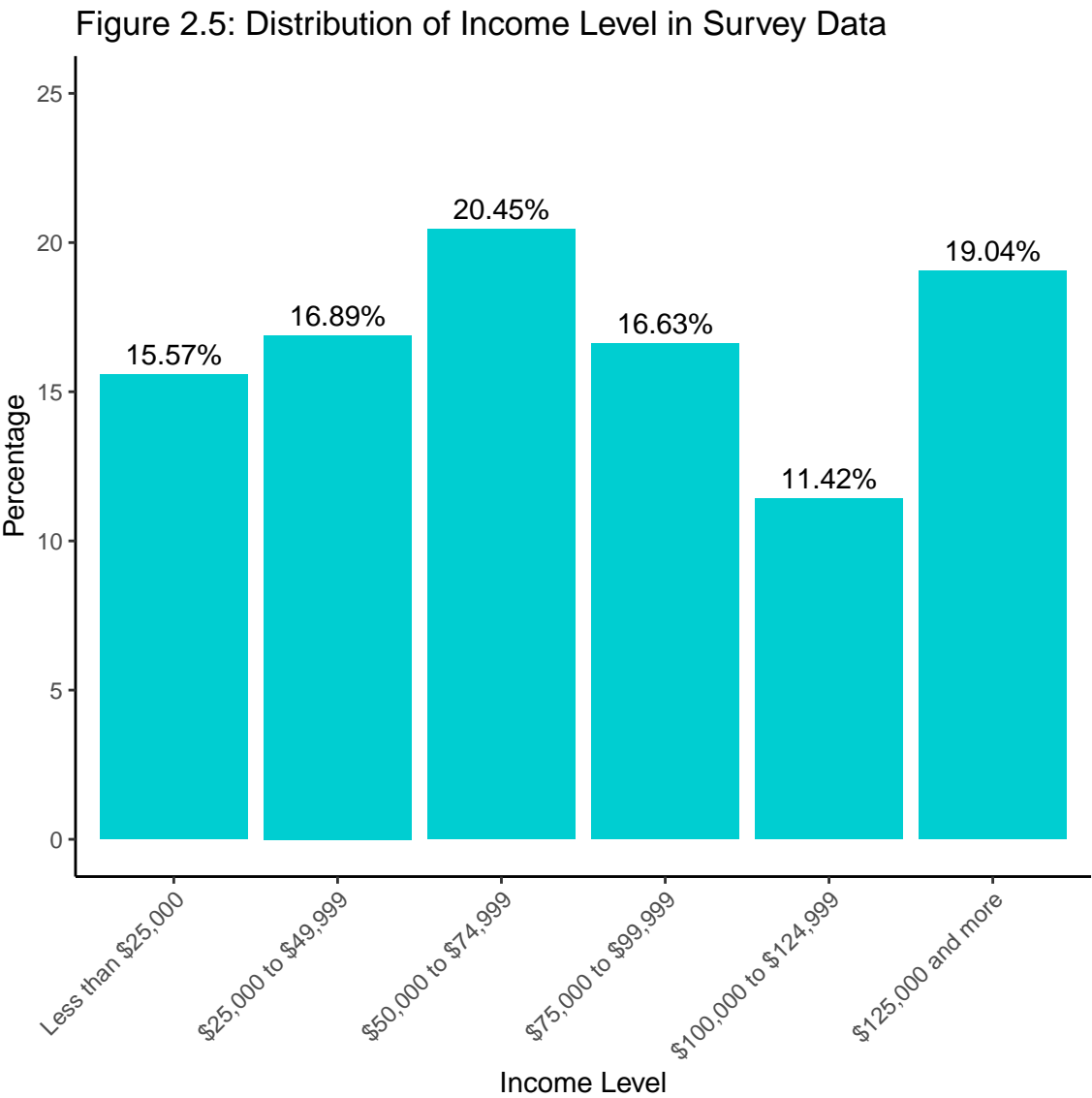
```
##
##
## Table: 2.3: Distribution of Province in Survey Data
##
## |Province| Percentage|
## |:-----|:-----:|
## |Ontario| 34.92|
## |Quebec| 30.26|
## |Alberta| 12.05|
## |British Columbia| 11.13|
## |Manitoba| 3.79|
## |Nova Scotia| 2.53|
## |Saskatchewan| 2.11|
## |New Brunswick| 1.96|
## |Newfoundland and Labrador| 0.96|
## |Prince Edward Island| 0.28|
```

Figure 2.4 & Table 2.4



```
##
##
## Table: 2.4: Distribution of Education Levels in Survey Data
##
## |Education Level| Percentage|
## |:-----:|:-----:|
## |Bachelor's degree| 28.98|
## |High school diploma| 21.88|
## |College, CEGEP, or other non-university certificate| 21.17|
## |Master's degree, professional degree, or doctorate| 14.54|
## |University certificate below bachelor's level| 10.72|
## |Less than high school diploma| 2.51|
## |Other Education| 0.18|
## |No Schooling| 0.02|
```

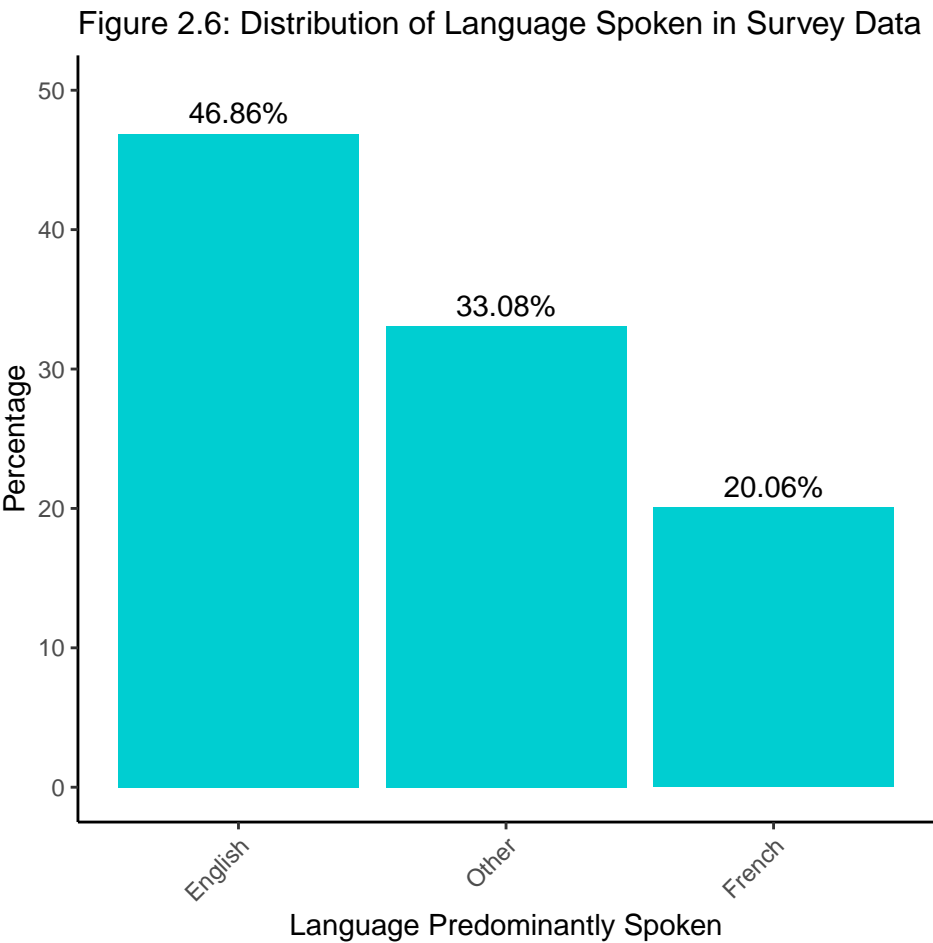
Figure 2.5 & Table 2.5



```
##
##
## Table: 2.5: Distribution of Income Levels in Survey Data
##
## |Income Level          | Percentage|
## |:-----:|:-----:|
## |$50,000 to $74,999   |    20.45|
## |$125,000 and more    |    19.04|
## |$25,000 to $49,999   |    16.89|
## |$75,000 to $99,999   |    16.63|
## |Less than $25,000    |    15.57|
## |$100,000 to $124,999|    11.42|
```



Figure 2.6 & Table 2.6



```
##
##
## Table: 2.6: Distribution of Language Predominantly Spoken in Survey Data
##
## |Language | Percentage|
## |:-----|-----:|
## |English  |      46.86|
## |Other    |      33.08|
## |French   |      20.06|
```

Figure 3.1

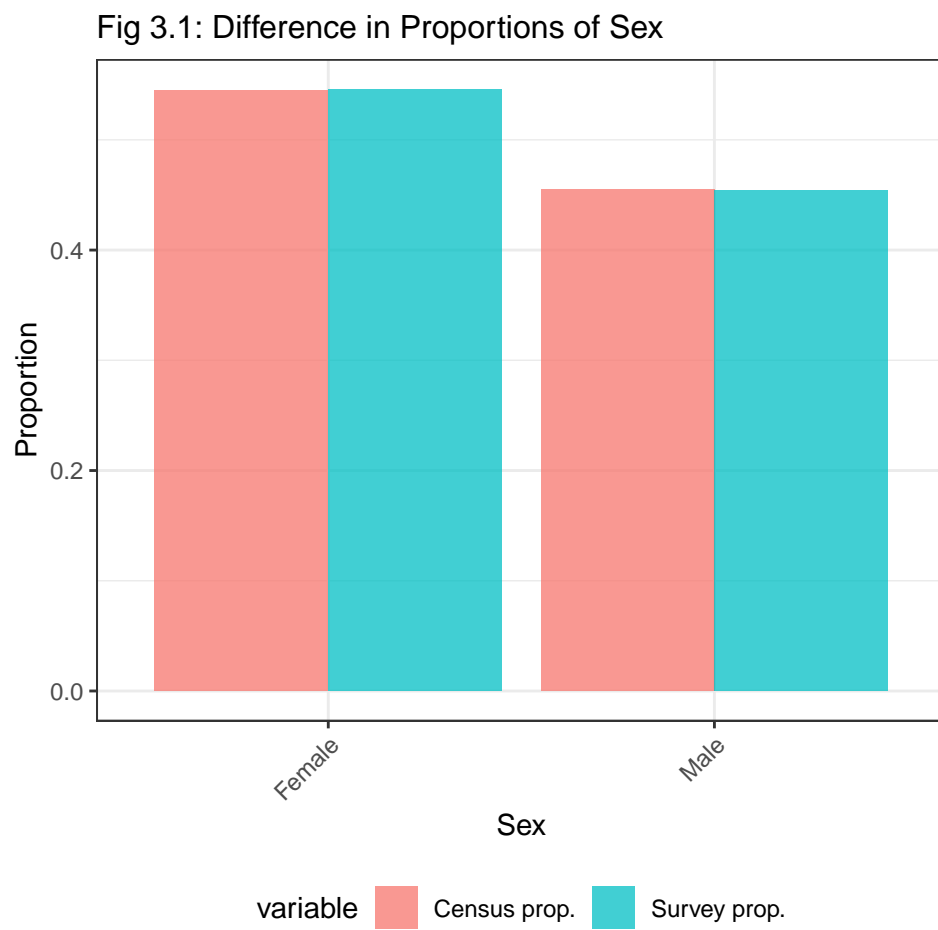


Figure 3.2

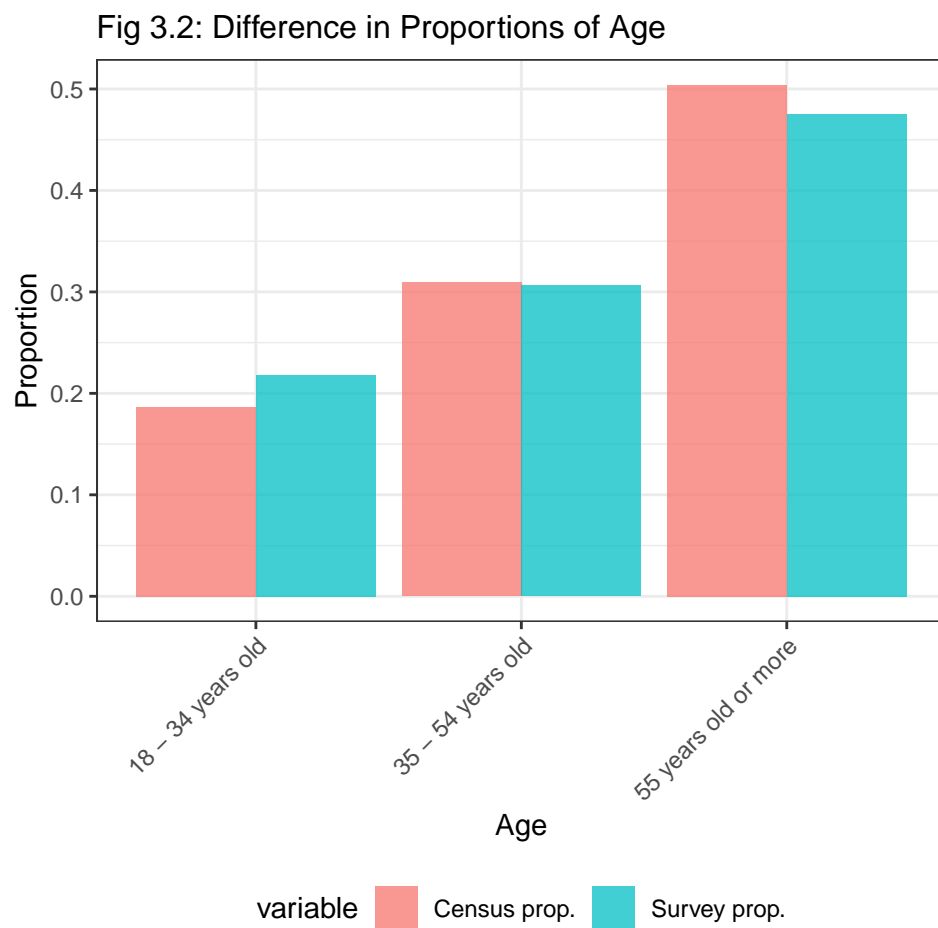


Figure 3.3

Fig 3.3: Difference in Proportions of Province Representation

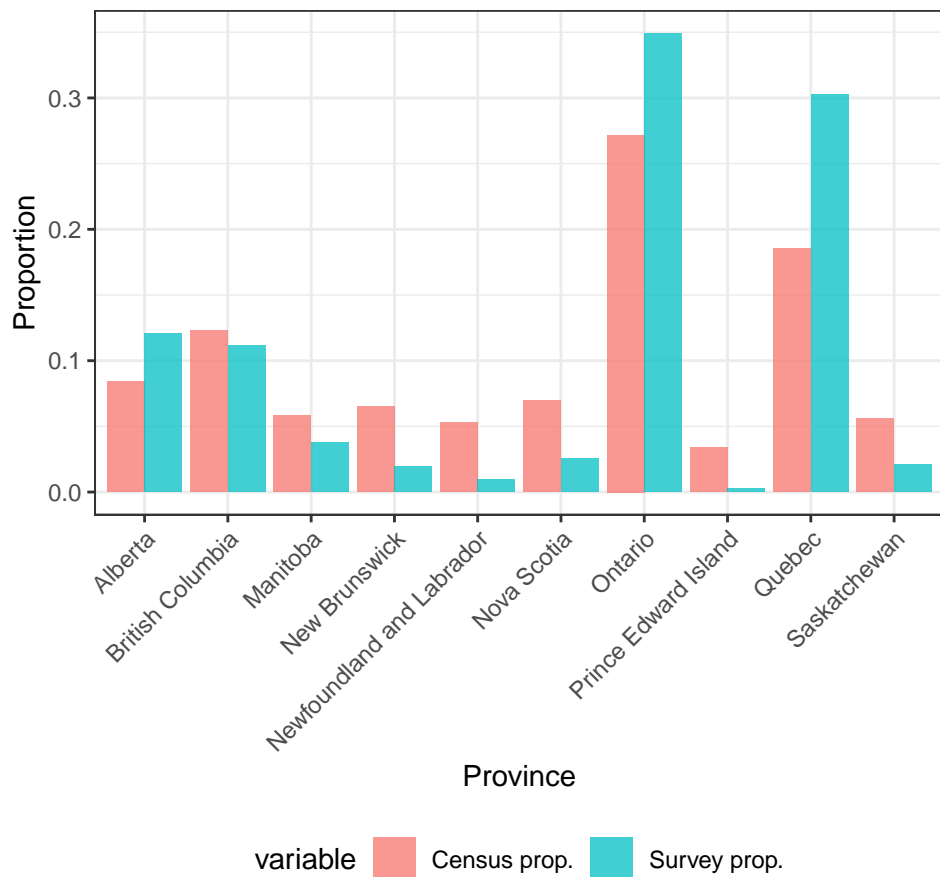


Figure 3.4

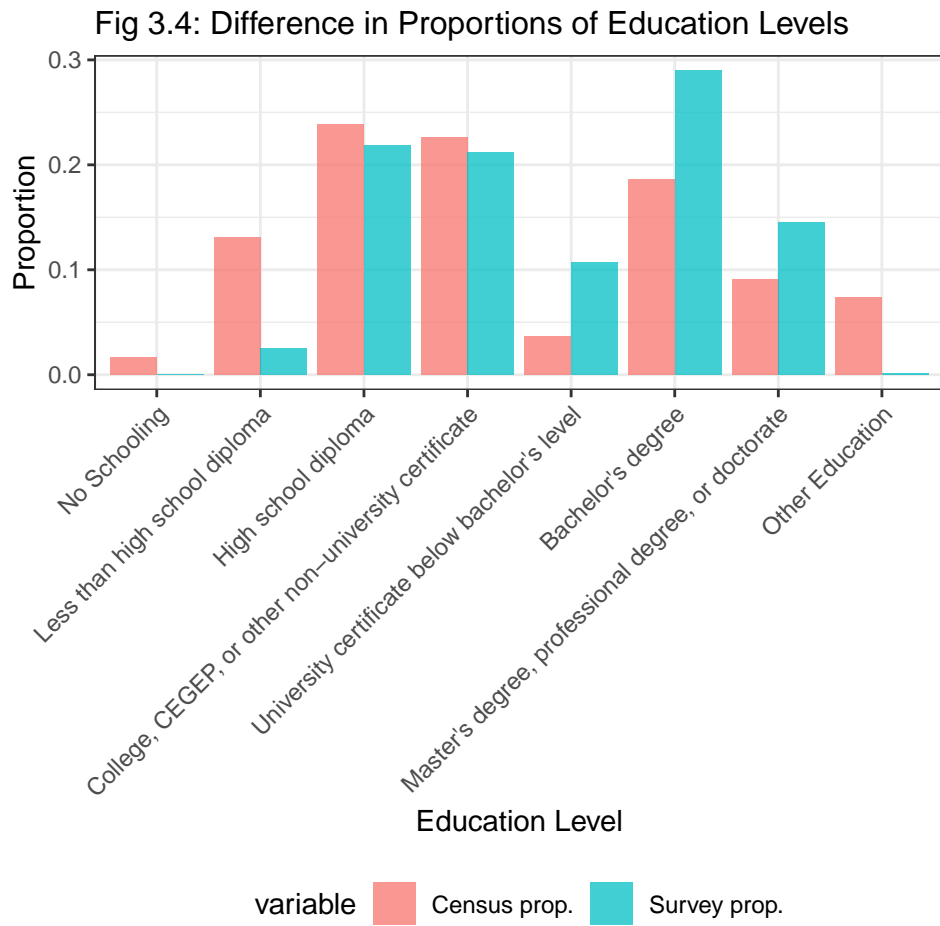


Figure 3.5

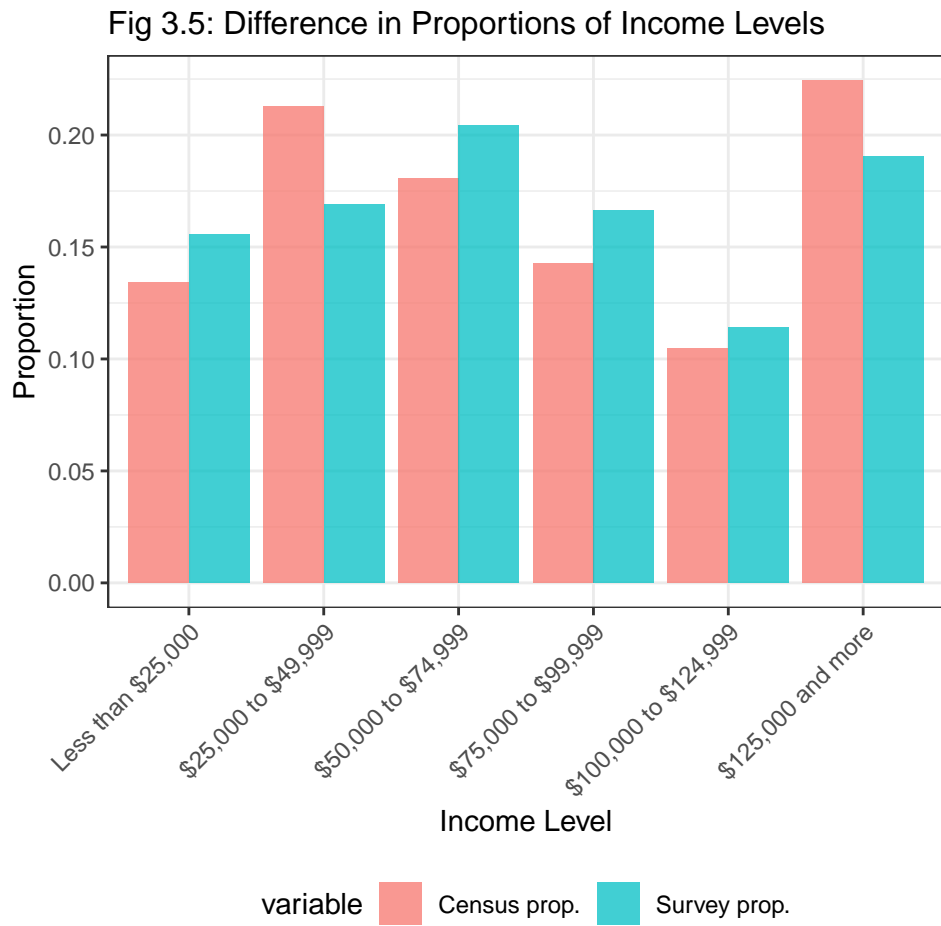


Figure 3.6

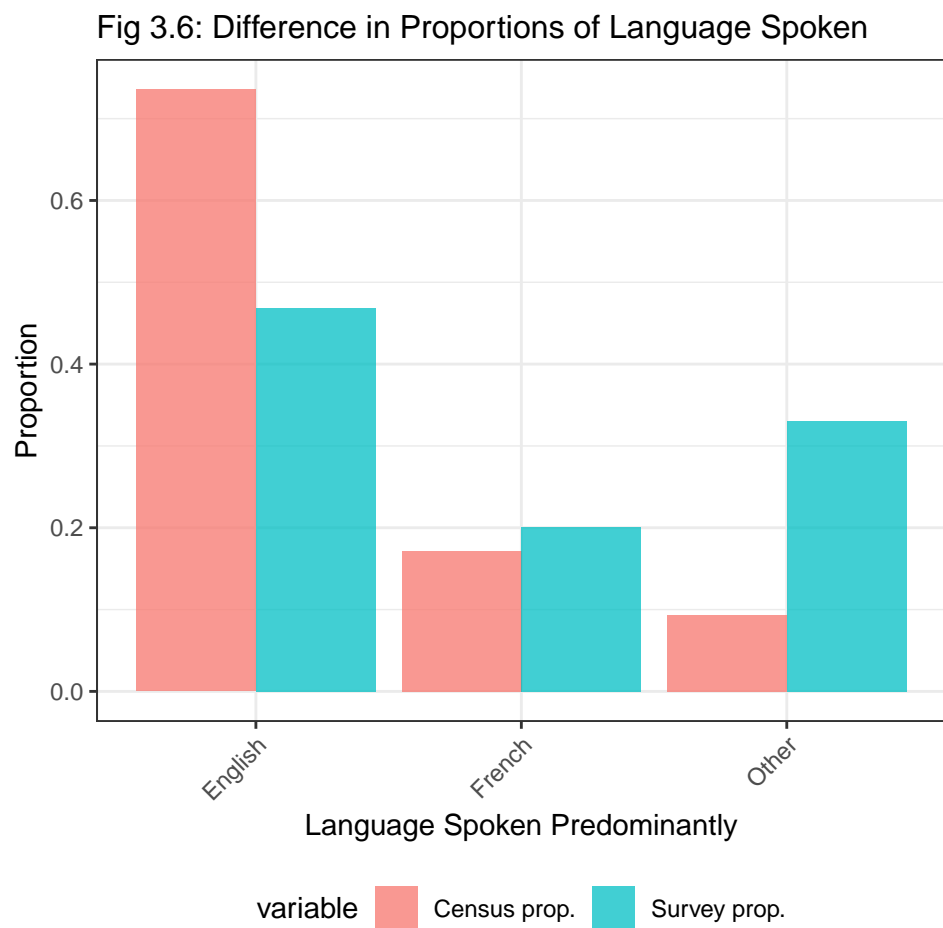
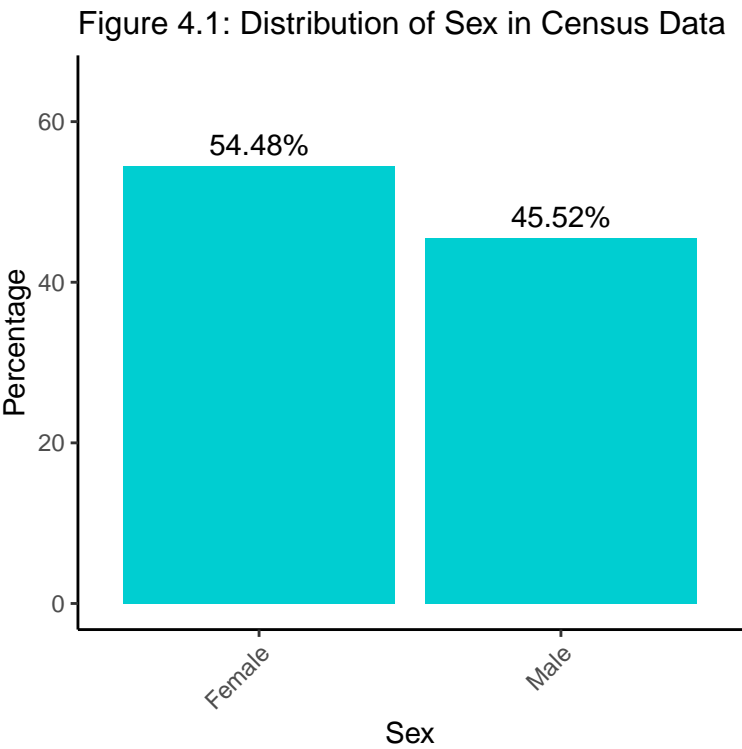


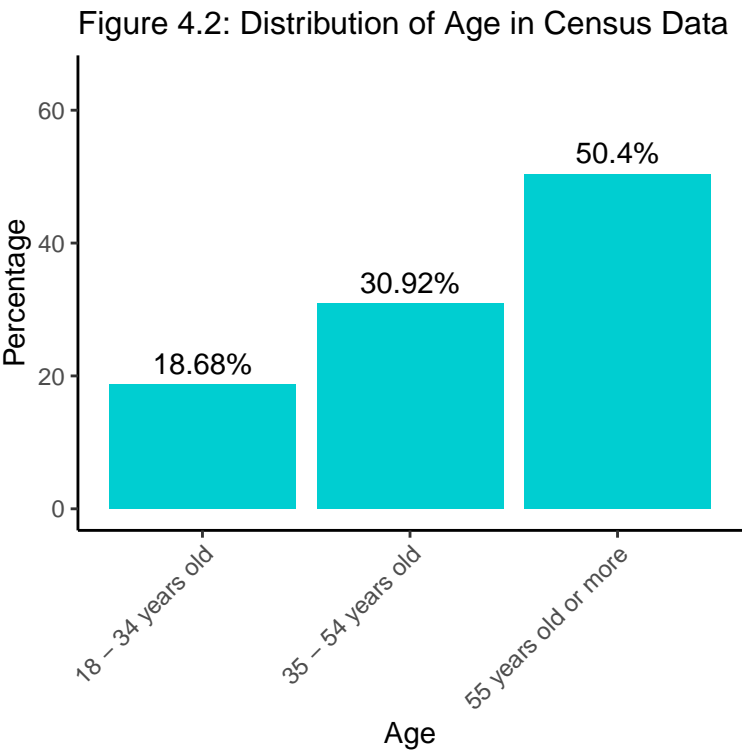
Figure 4.1 & Table 4.1



```
##
##
## Table: 4.1: Distribution of Sex in Census Data
##
## |Sex      | Percentage|
## |:-----|:-----:|
## |Female  |      54.48|
## |Male    |      45.52|
```



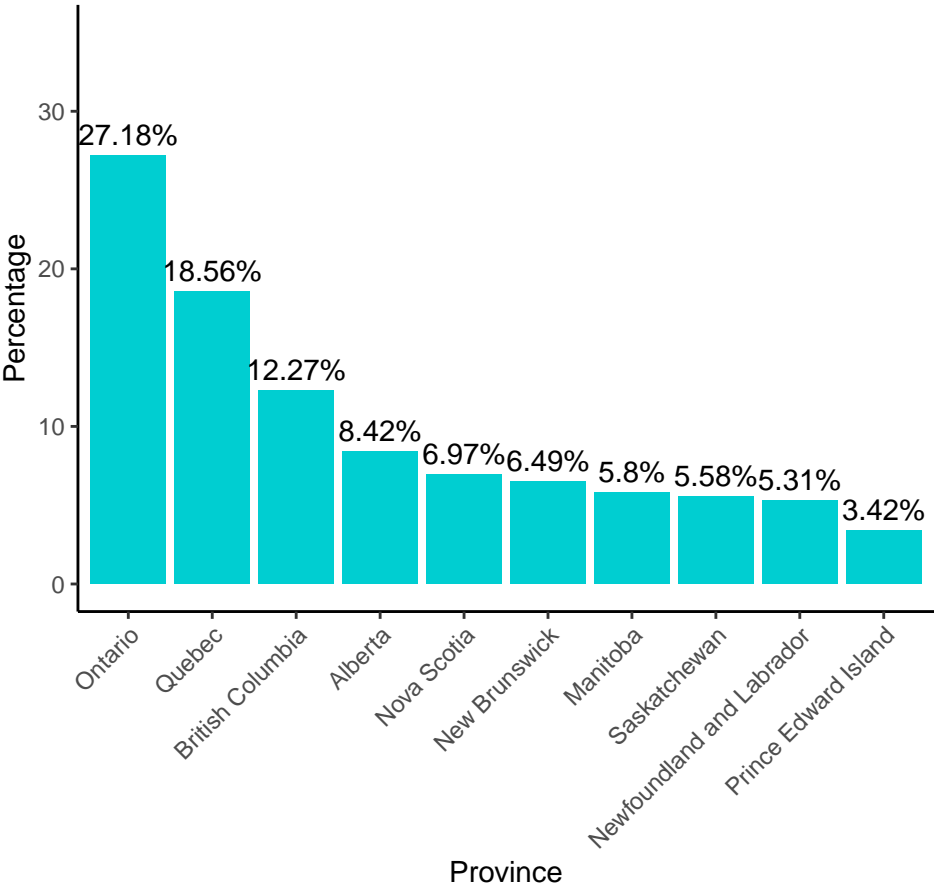
Figure 4.2 & Table 4.2



```
##
##
## Table: 4.2: Distribution of Age in Census Data
##
## |Age                | Percentage|
## |:-----|:-----:|
## |55 years old or more |      50.40|
## |35 - 54 years old    |      30.92|
## |18 - 34 years old    |      18.68|
```

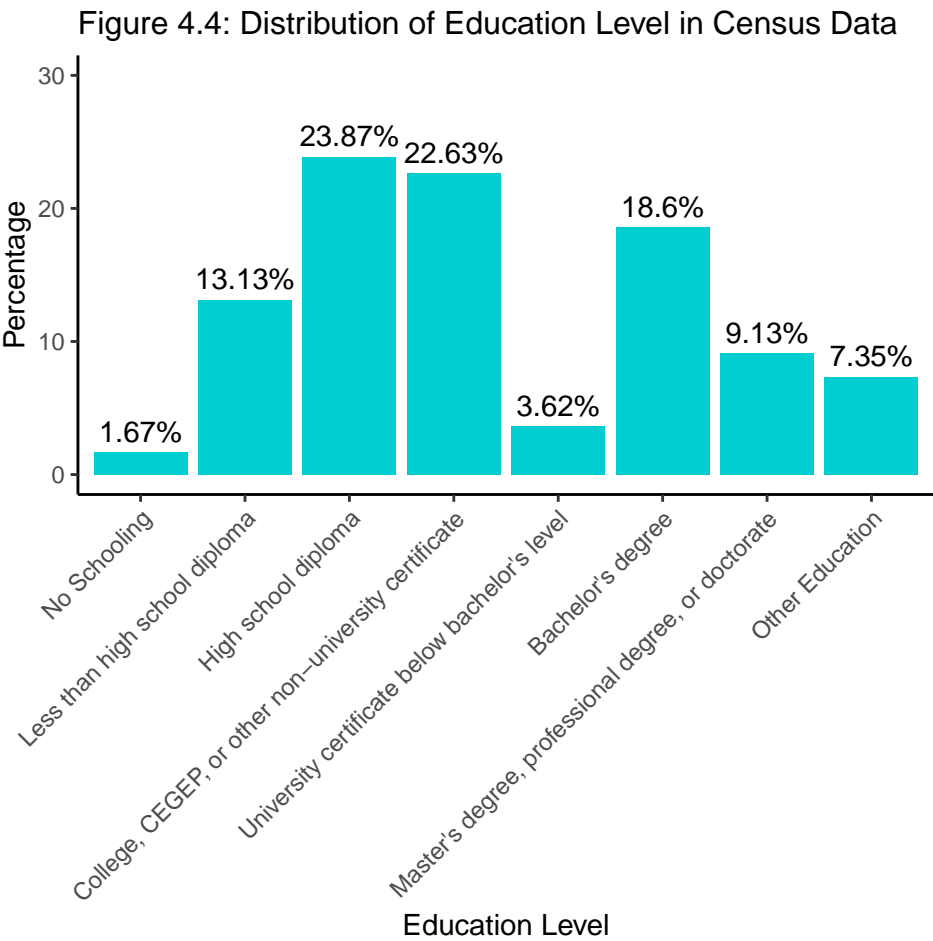
Figure 4.3 & Table 4.3

Figure 4.3: Distribution of Province in Census Data



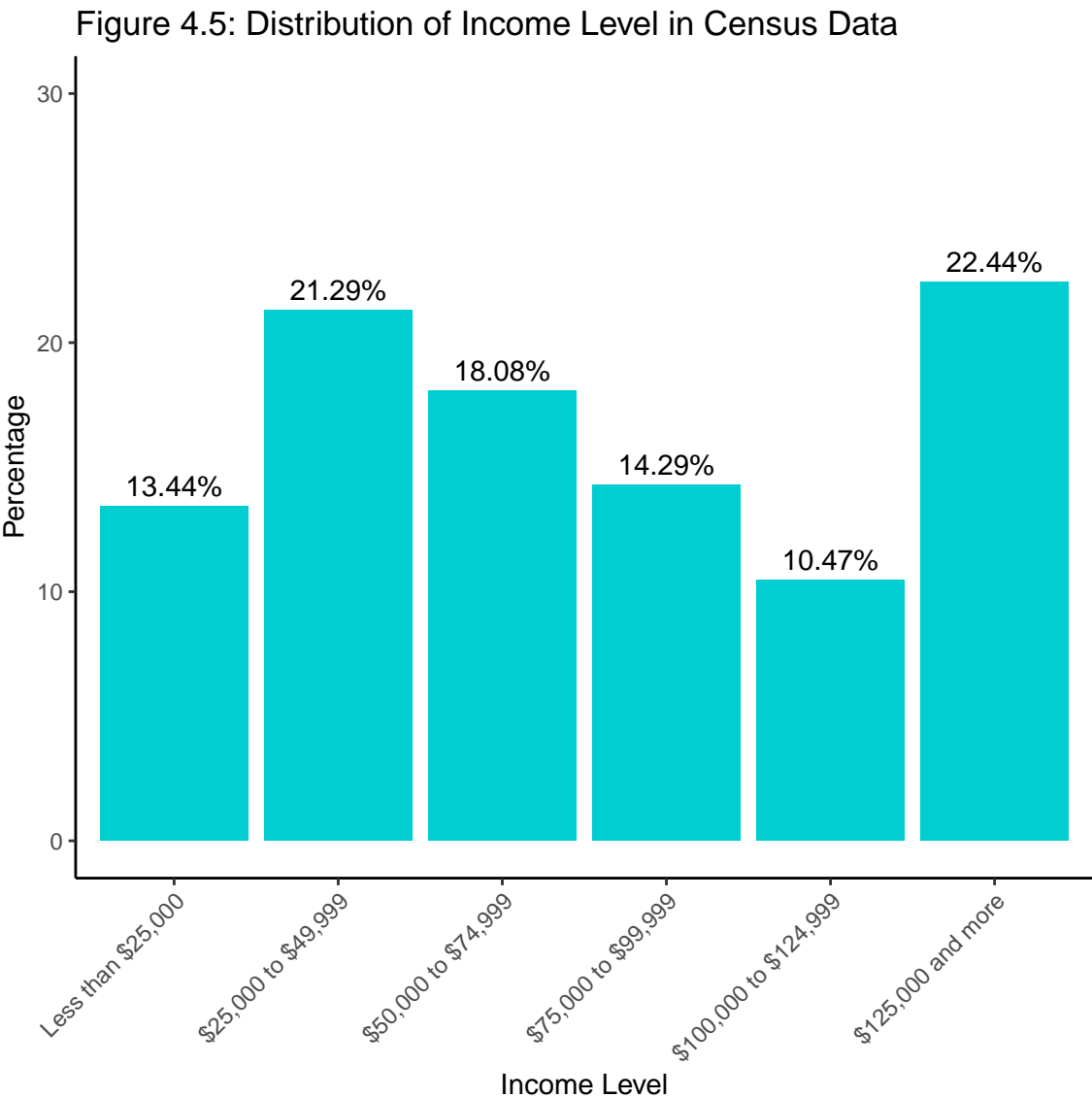
```
##
##
## Table: 4.3: Distribution of Province in Census Data
##
## |Province| | Percentage|
## |:-----| |-----:|
## |Ontario| | 27.18|
## |Quebec| | 18.56|
## |British Columbia| | 12.27|
## |Alberta| | 8.42|
## |Nova Scotia| | 6.97|
## |New Brunswick| | 6.49|
## |Manitoba| | 5.80|
## |Saskatchewan| | 5.58|
## |Newfoundland and Labrador| | 5.31|
## |Prince Edward Island| | 3.42|
```

Figure 4.4 & Table 4.4



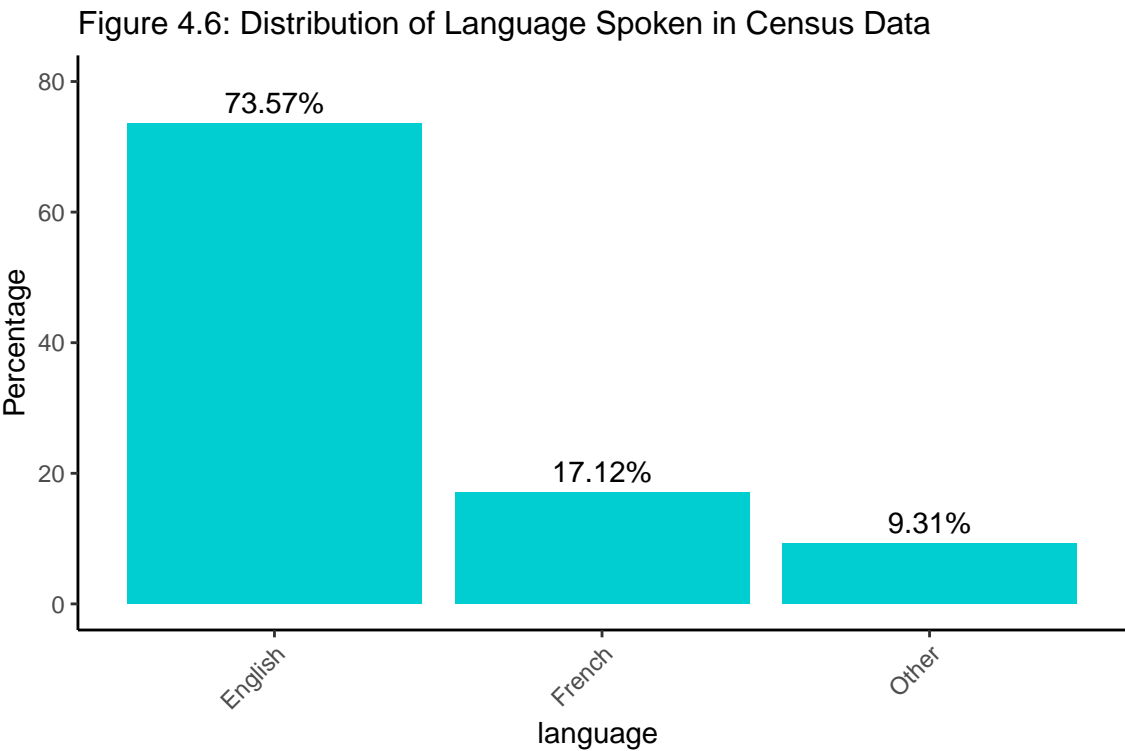
```
##
##
## Table: 4.4: Distribution of Education Levels in Census Data
##
## |Education Level| Percentage|
## |:-----:|:-----:|
## |High school diploma| 23.87|
## |College, CEGEP, or other non-university certificate| 22.63|
## |Bachelor's degree| 18.60|
## |Less than high school diploma| 13.13|
## |Master's degree, professional degree, or doctorate| 9.13|
## |Other Education| 7.35|
## |University certificate below bachelor's level| 3.62|
## |No Schooling| 1.67|
```

Figure 4.5 & Table 4.5



```
##
##
## Table: 4.5: Distribution of Income Levels in Census Data
##
## |Income Level          | Percentage|
## |:-----|:-----:|
## |$125,000 and more    |    22.44|
## |$25,000 to $49,999   |    21.29|
## |$50,000 to $74,999   |    18.08|
## |$75,000 to $99,999   |    14.29|
## |Less than $25,000     |    13.44|
## |$100,000 to $124,999 |    10.47|
```

Figure 4.6 & Table 4.6



```
##
##
## Table: 4.6: Distribution of Language Predominantly Spoken in Census Data
##
## |Language | Percentage|
## |:-----|-----:|
## |English  |      73.57|
## |French   |      17.12|
## |Other    |       9.31|
```

Figure 5.1

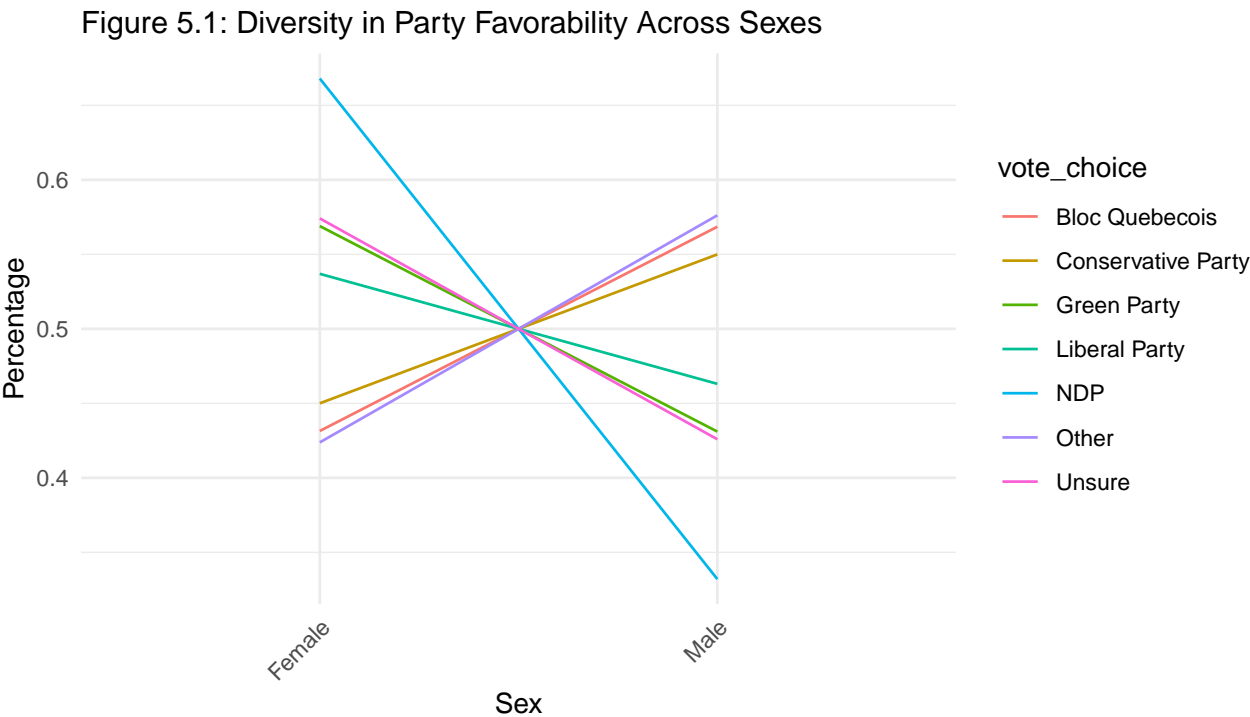


Figure 5.2

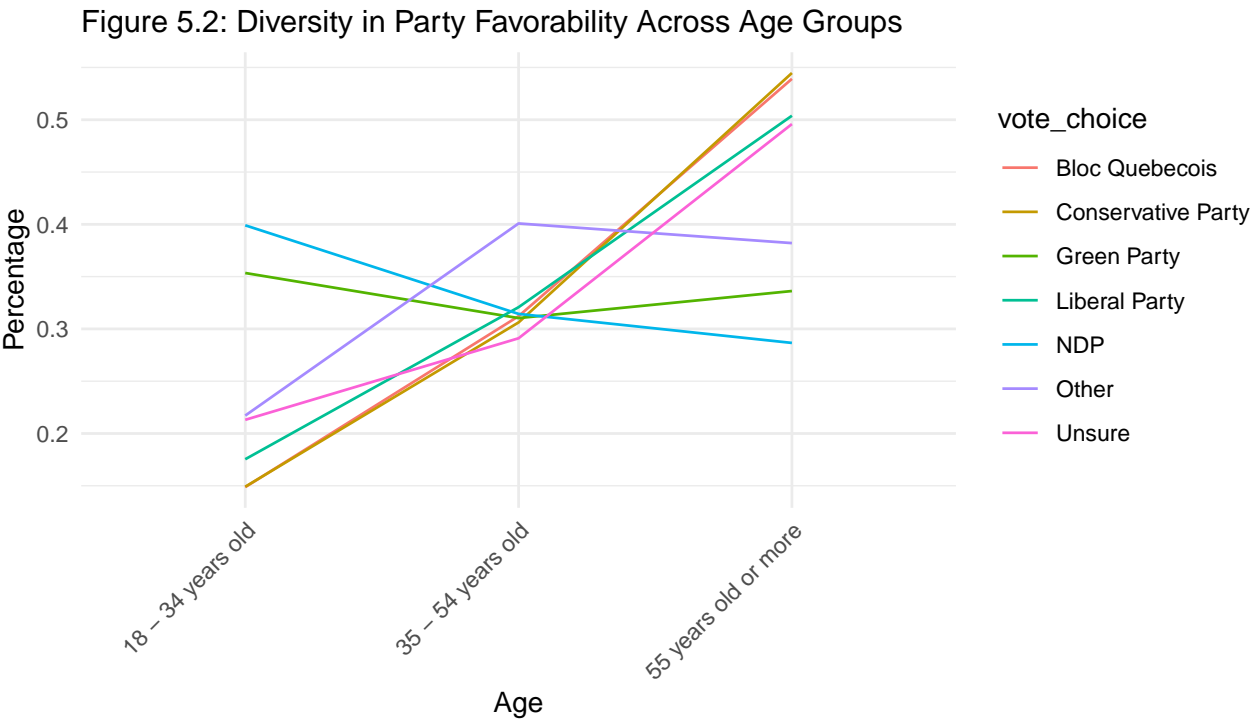


Figure 5.3

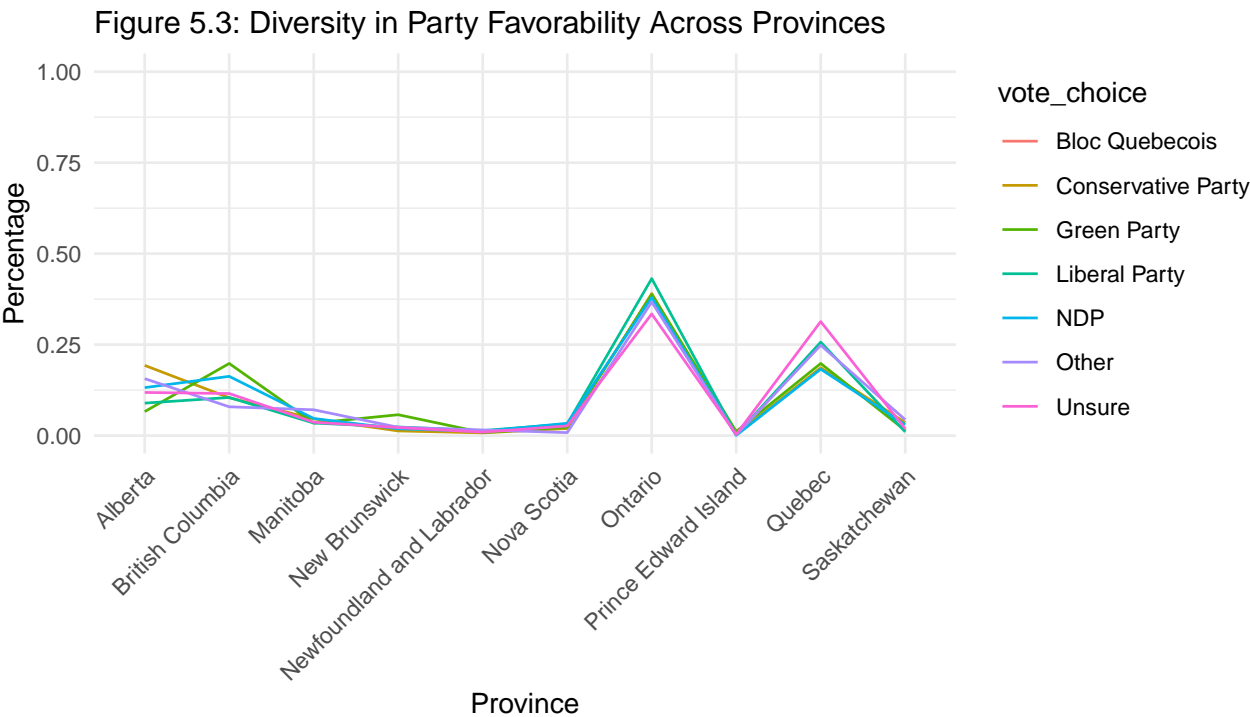


Figure 5.4

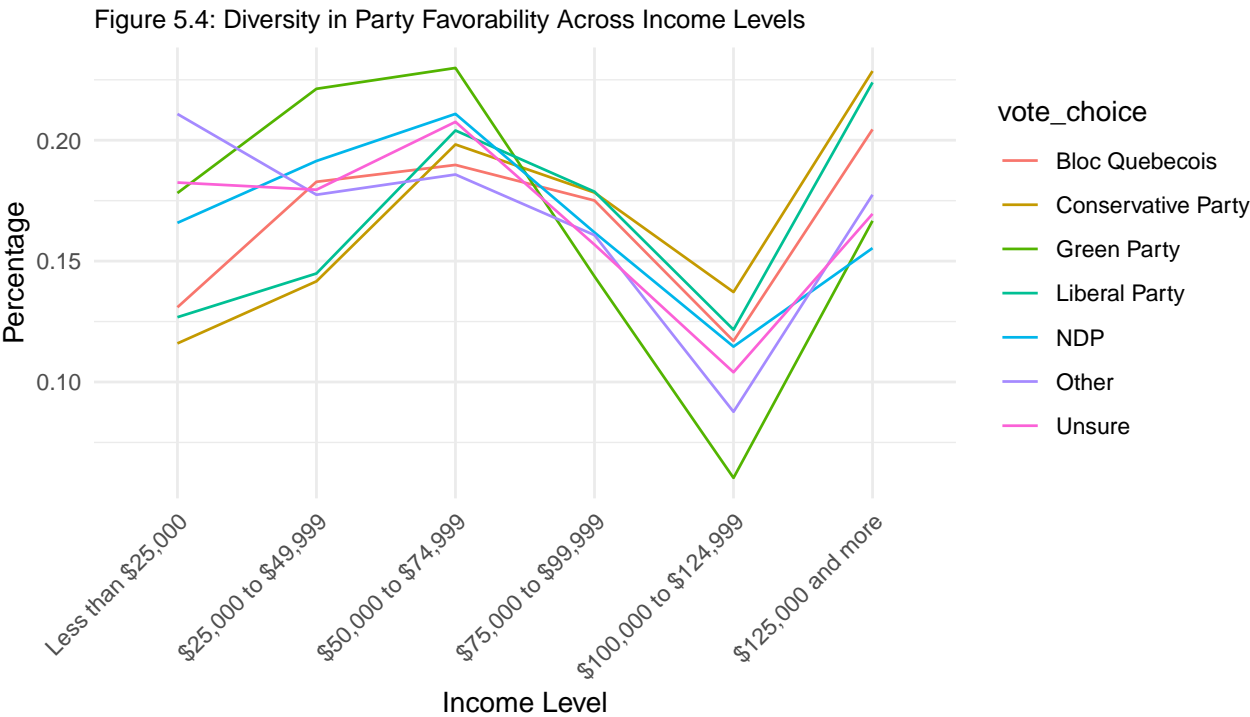


Figure 5.5

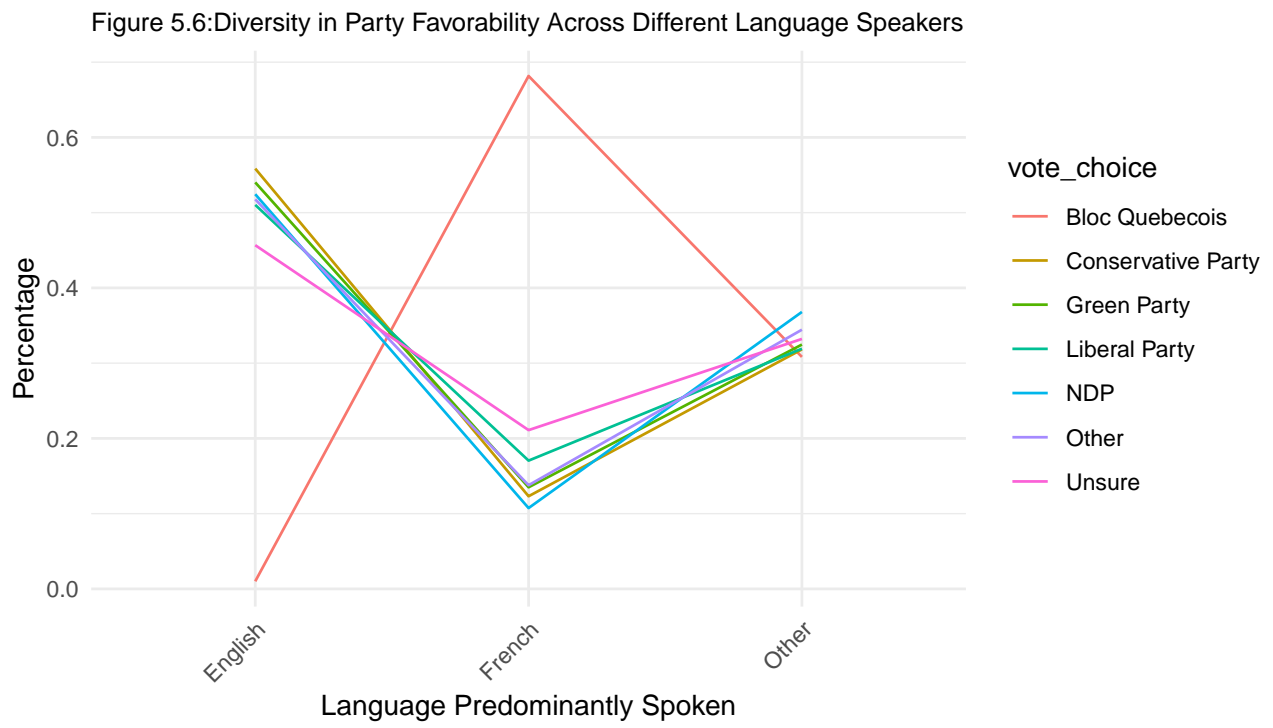




Figure 5.6

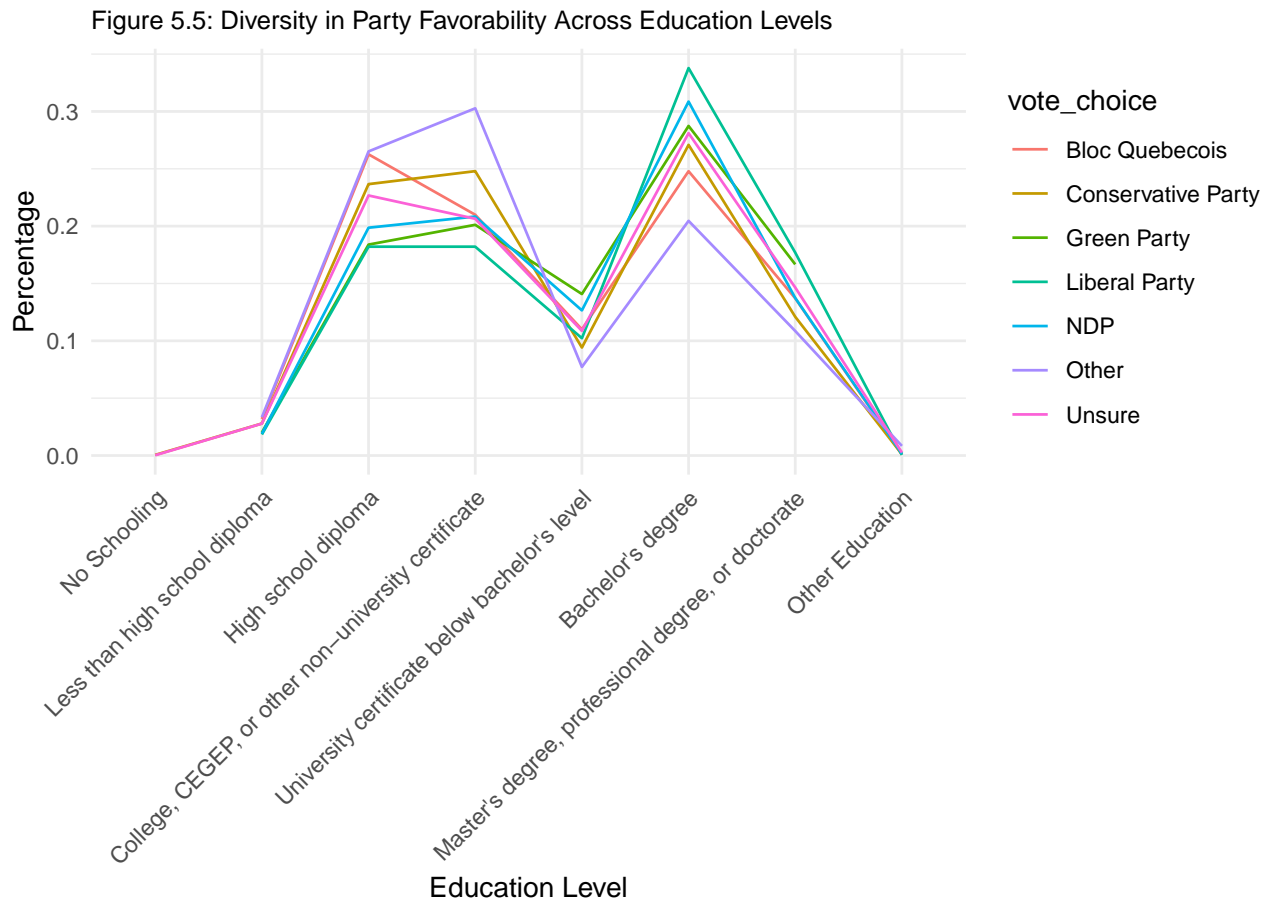


Table 3.1

##		GVIF	Df	GVIF <sup>1/(2*Df)</sup>
##	age	1.156951	2	1.037119
##	province	2.058342	9	1.040921
##	sex	1.065993	1	1.032469
##	education	1.165533	7	1.011001
##	income	1.151411	5	1.014199
##	language_home	2.099855	2	1.203781

Table 3.2

##		GVIF	Df	GVIF <sup>1/(2*Df)</sup>
##	age	1.145062	2	1.034445
##	province	2.040873	9	1.040428
##	sex	1.058725	1	1.028944
##	education	1.176313	7	1.011666
##	income	1.162596	5	1.015180
##	language_home	2.088777	2	1.202190

Table 3.3

##	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
## age	1.165945	2	1.039129
## province	1.900384	9	1.036314
## sex	1.041694	1	1.020634
## education	1.173818	7	1.011513
## income	1.143069	5	1.013461
## language_home	1.944754	2	1.180908

Table 3.4

##	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
## age	1.187035	2	1.043797
## province	1.918310	9	1.036854
## sex	1.082310	1	1.040341
## education	1.187305	7	1.012339
## income	1.152894	5	1.014329
## language_home	1.951782	2	1.181974

Figure 5.1

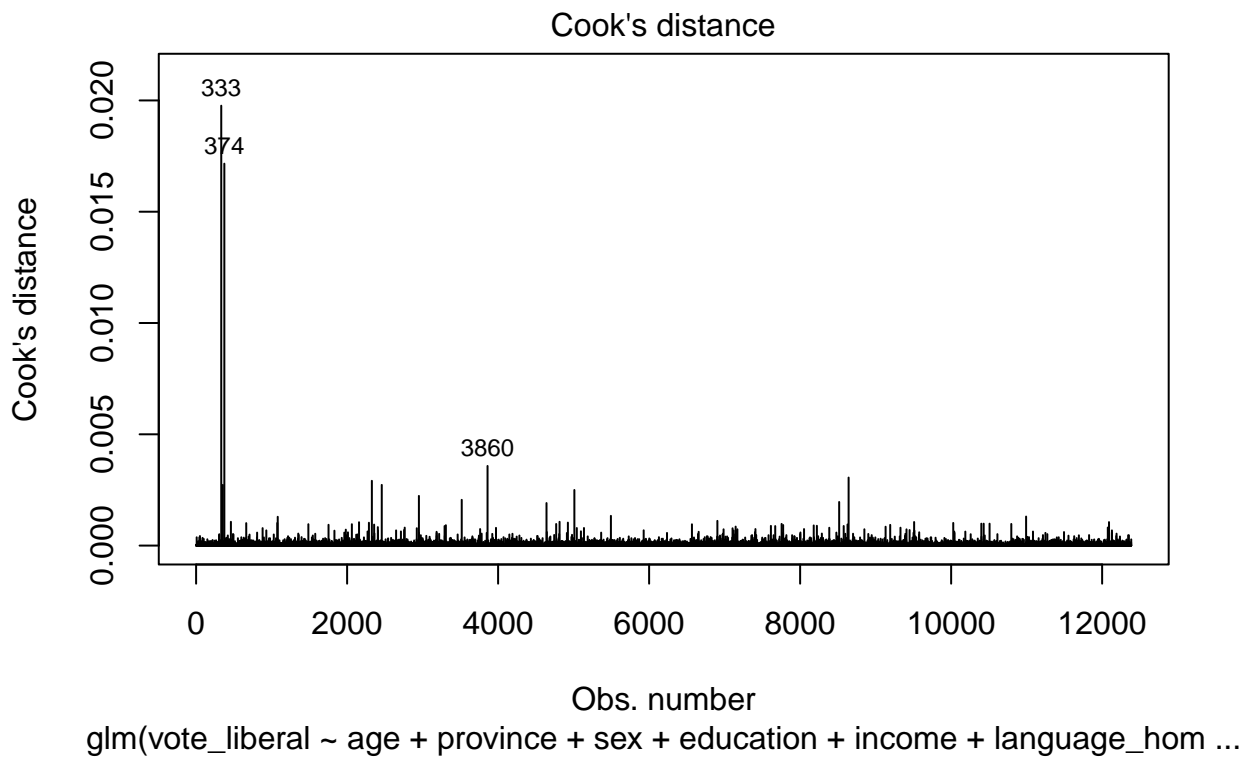


Figure 5.2

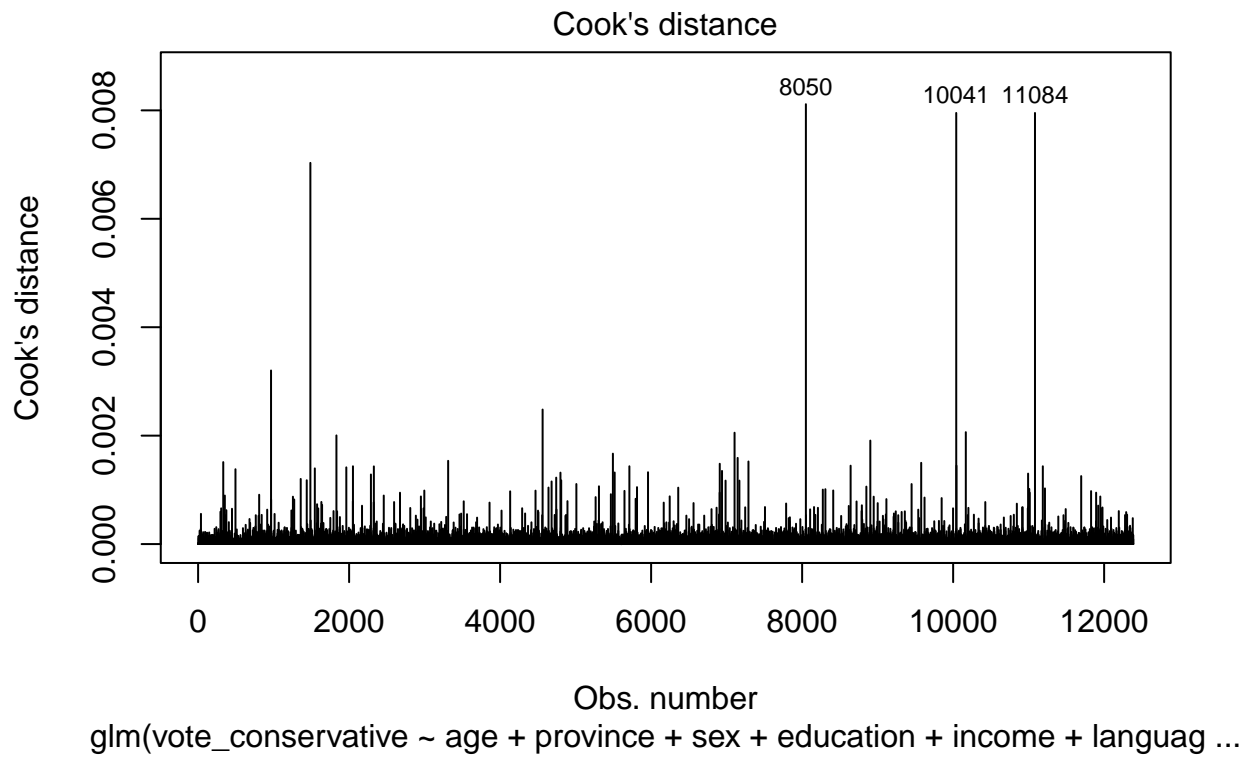


Figure 5.3

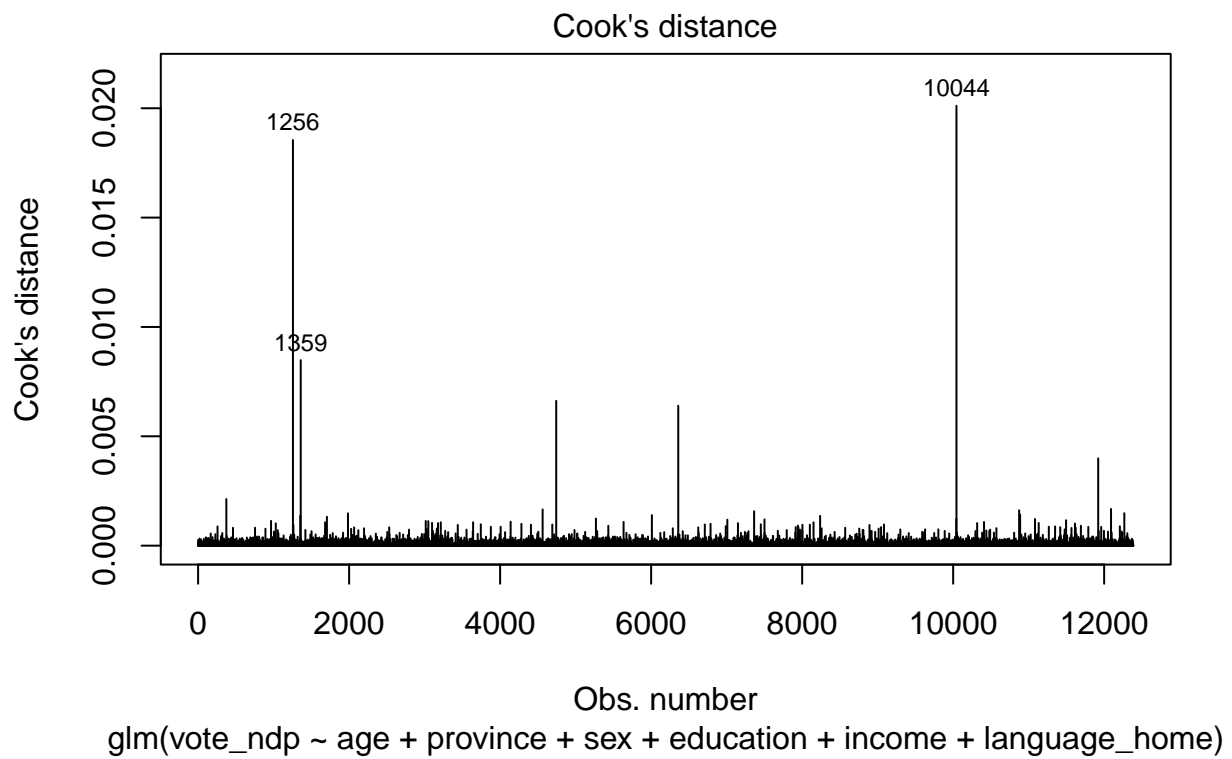


Figure 5.4

