

Predicting the overall popular vote for the next Canadian Federal Election using Socio-economic factors

STA304 - Group-4 Assignment 2

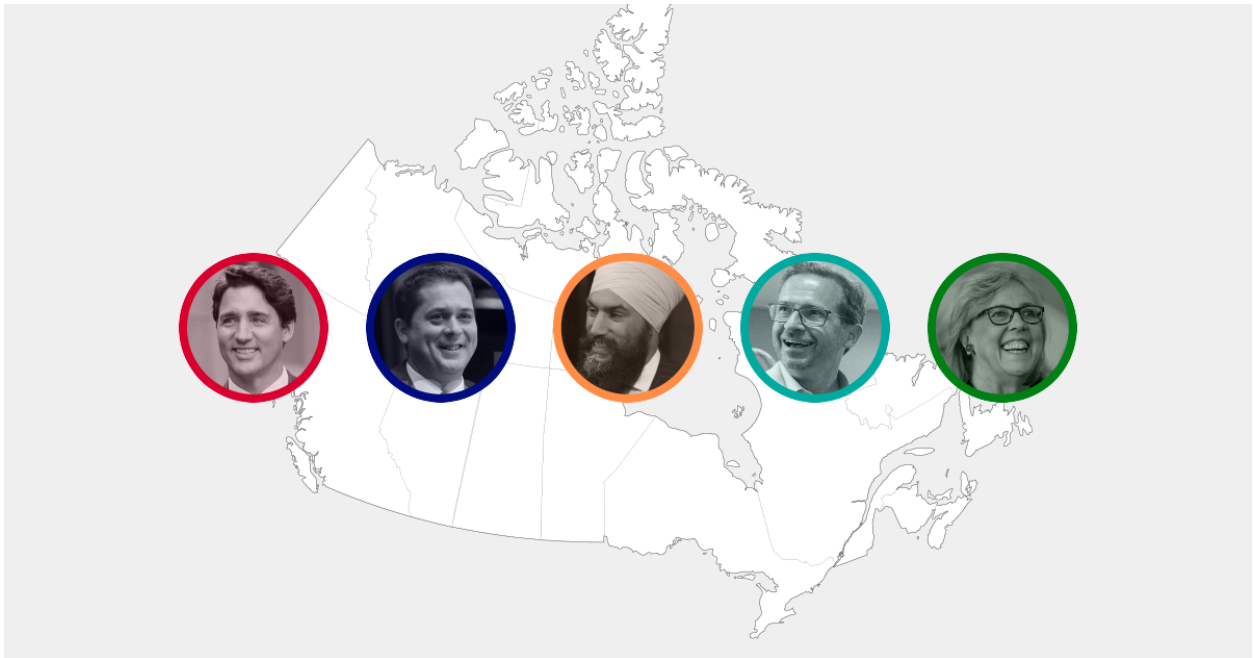
Uzair Mirza

Shruti Sood

Pranav Sethi

Krishna Selvakumar

May 28, 2021



[25]

Introduction

Goal

Canada has always been a vibrant democracy which is solely for the people, of the people and by the people. Canada's Egalitarian approach to Democracy has particularly made sure that during any form of decision making in the government at any level, the above statement is given the utmost importance. At the heart of Canadian Democracy, Canadian decision making lies in the Federal election during which a general vote is held to elect the people who will make decisions for Canada's future: the members of Parliament. Under the Canadian law, the federal elections are scheduled every four years during which election races are held in electoral districts to represent the people living in that area. There are a total of 338 electoral districts with the most being present in Ontario.

The goal of this study will be to **predict the overall popular vote of the next Federal elections** which are scheduled to take place on or before October 16 2023. This was done using regression models with post-stratification, both of which will be described in depth in later sections. The models were built using the Canadian Election Study of 2019 which was treated as Survey Data, and predictions were performed using General Social Survey from 2017 which was treated as census data.

Background

The first Canadian federal election was held in 1867, and after Bill C-16 was proclaimed into the law, federal elections have been regularly occurring every four years in most provinces[1][2]. Elections Canada hires nearly 250,000 people all over Canada during the federal elections making it the largest employer in Canada on election day [3].

Hypothesis

For the past three federal elections that were held in Canada, the top three parties have been Liberal, Conservative and New Democratic Party where the Liberal Party had the most popular vote for the last two elections [4]. Naturally one would be curious about the future election that is to take place in 2023 as the party with the most popular vote will be ruling our country for the next couple of years. This leads to our analysis of forecasting the popular vote for the upcoming federal election and we are expecting the overall popular vote to be Liberal based on patterns from the past election results.[4].

Significance

As touched upon earlier, the federal election lies at the heart of Canadian democracy and decision making. Once the votes are counted Canadians learn who their government will be[5]. Being such an important event for all residents of Canada and also to people outside Canada to a certain degree, it is of prime importance to have quality research done on the federal elections for the continued functioning of a healthy democracy. Research done on federal elections can not only reveal existing traits about the political preferences, but also future political preferences of Canadians, which is what this study is aimed at answering. Results of this study can be used by political parties to get an estimate of their standing in the future to plan upcoming campaigns and Canadians can use this study to get a greater understanding of the popular political parties.

Global Significance

In 2017, 96 out of 167 (57%) of countries in the world with populations of 500,000 or more were democracies of some kind. Since the mid-1970's, the world has seen an increase in the number of countries becoming democratic.[6]

Elections are one of the most important parts of a democratic state. In today's world, a direct-democracy where all political decisions are made by all the citizens of the country is impractical. Hence, democracy is conducted through representatives. These representatives are nominated through elections during which the citizens choose their leaders and hold them accountable for their performance in the office. Elections give the citizens an opportunity to have a say and express their opinions and feel a sense of belonging to a community and work for its betterment. Every vote is important. Credible Elections indirectly lead to a feeling of personal fulfillment in the citizens. Thus, the findings of this study can not only be studied in Canada but also in electoral democracies throughout the world to get a sense of understanding of the pros & cons of their own electoral system when compared to the Canadian electoral system. For instance, the findings of this study can be used by political analysts in the United States of America, which has historically been a two party system, to understand the effects a multiparty system has on the distribution of votes[7].

Terminology for the study

Regression is used to determine the relationship between variables. More specifically, it is used to analyze how certain explanatory variables influence a predictor variable. In our analysis, we will be using logistic regression as this statistical method is used when the predictor is a binary variable and in our case we are building multiple models to predict whether or not an individual will vote for a specific party or not in an effort to use these results to forecast the popular vote in the next federal election.

Post stratification is a process that involves grouping units with similar characteristics into bins and adjusting the weights of these bins as necessary in order to compare survey results to a census data or any other data set that already exists. We will be performing this method on the Canadian Election Study of 2019 data which will act as our survey results and compare it to the General Social Survey from 2017 data which will be used as our census data.

Drawbacks

The biggest potential drawback that needs to be considered is the use of data from 2017 & 2019 to predict the outcome of the 2023 federal election. This drawback will most likely have an even greater effect on the findings due to the fact that the data was collected a year before the start of Covid-19 pandemic which has caused unprecedented changes in government policies and spending that has had a significant impact on the political views of Canadians. Apart from that, another potential drawback could be false answers being put by the participants of the survey for the geography and demographic characteristics part of the survey which did not get detected during data cleaning because of them following the correct format but not being the accurate answer for that respective participant.

Data

Dataset description

As mentioned in the Intro, the models were built using the Canadian Election Study of 2019 which was treated as Survey Data, and predictions were performed using General Social Survey from 2017 which was treated as census data. The Canadian election study since 1965 has been a rich source of data on Canadians' political behavior and attitude. The 2019 iteration of the CES was conducted online with the target population being all Canadian citizens and permanent residents who are 18 years or older. [8] An online sample of the participants to be contacted was created from the members of Canadian general population using qualtrics, a survey software tool, with targets stratified by region which were 7 in total. Additionally, within each region, provincial quotas were split evenly. Lastly, the online sample was created with the several aims such as having a 50-50 male & female ratio, 80-20 French & English ratio in Quebec. The GSS program established in 1985 has been a rich source of data for monitoring changes in living conditions and providing information on social policy issues in Canada. The 2017 iteration of the GSS was conducted through the phone via computer assisted telephone interviews (CATI) with the target population being all people 15 years or older in Canada excluding residents of Yukon, Northwest Territories, Nunavut, and full-time residents of institutions. Each of the 10 Provinces were divided into strata using geographic areas and the survey frame was created using 2 elements: the list of telephone numbers in use that are available to Statistics Canada, and the address register which contained a list of all dwellings in each of the Provinces. Next, a simple random sample without replacement of records was performed in each stratum to select the household to be contacted. [9].

Data manipulation and cleaning

The journey of Data cleaning started by reducing the number of columns in GSS from 50 to 7 to only include the ones that are needed for the model building. These variables of interest are discussed in the section "description of important variables" which is right below. The reduction of columns was achieved through using the `select()` function. Next, it was decided to merge

column # 3: `place_birth_macro_region` which contained birthplace regions of participants born outside Canada

column # 4: `place_birth_province` which contained the Province the participants were born in.

The decision to merge these columns was made to reduce the number of parameters for our model and the ease of the reader as both the columns essentially reveal the same characteristic about the participant and each row for these columns had an "NA" whenever the other column didn't have an "NA". All the Inputs for both these columns were merged and inserted in a new column labeled `place_birth`. The inputs for these columns were merged using the `coalesce()` function and the new column was created using the `mutate()` function over the `coalesce` function i.e. `mutate(coalesce())`. Naturally, column # 3 and # 4 were then deleted using the `select()` function.

Lastly, all the NAs' were then dropped from our modified GSS using the `na.omit()` function.

Dataset Syncing

Since two different surveys were being used for the study, syncing had to be performed between the GSS and the CES data set to move forward. The syncing process began with re-categorizing answers from the column containing the education level of the participant in our modified GSS so that it matched the categories of the column containing the level of education of the participant in the CES data set. This was done using the `replace()` function. Next, all the decimal points from age in our modified GSS were dropped to match the format of age in CES. This was done through inserting the column containing age inside the `floor()` function. Lastly, all the columns in our modified GSS for which we couldn't find equivalent columns in CES were dropped using the `select()` function.

Description of Important Variables

As previously mentioned, we will be using the survey data to predict the popular vote among the country based on the individuals that are present in the census data. To conduct our analysis, we need to ensure that the variables in both the survey and census were in the same manner for obvious reasons. Initially we had begun with 81 variables in the data, however after narrowing down the variables we ended up giving importance to 6 main variables which include **sex**, **province**, **education level**, **political preference**, **family's income** and **age**.

Once again, our goal is to forecast the outcome of the next federal election by using political preference and province as our response variables which are two very important variables as it would be extremely challenging and nearly impossible to conduct a realistic analysis without knowing these details.

Ekos Research Associates is a Canadian research company that measured males are almost equally inclined to vote for Liberal as they are for Conservatives [10]. More specifically, 37% of males have intended to vote for Conservatives and 35% had intentions to vote Liberal [10]. This is to say that if the voting population consisted of just Males there would be a tie between the Liberal and Conservatives. On the other hand, Ekos reported that 47% of females were in favor of Liberals and only 23% were in favor of the Conservatives [10]. Thus, it can be seen that sex definitely has a huge impact on voting behavior as male and females differ between their political preferences.

The age of the electoral population matters a great deal as it has been hypothesized that younger people are more likely to vote for liberal in comparison to older people [11]. Accordingly, the prediction of the popular vote can be greatly influenced by age depending on the size of the different age groups. Similarly, the education level of the voters can impact the voter turnout as those with a university degree are 78% likely to vote whereas the voting rate for those with high school education or less is 60% [12].

Additionally, it has been recorded that families with higher income levels have a greater voter turnout in comparison to those who come from families with a lower income [13]. A possible reason for this is due to how costly an activity voting may be as it requires time, information, transportation, and other things that may be inconvenient for lower income families [13]. Furthermore, higher income could be associated with higher education and once again this would influence voting results as mentioned previously [13].

Data Drawbacks

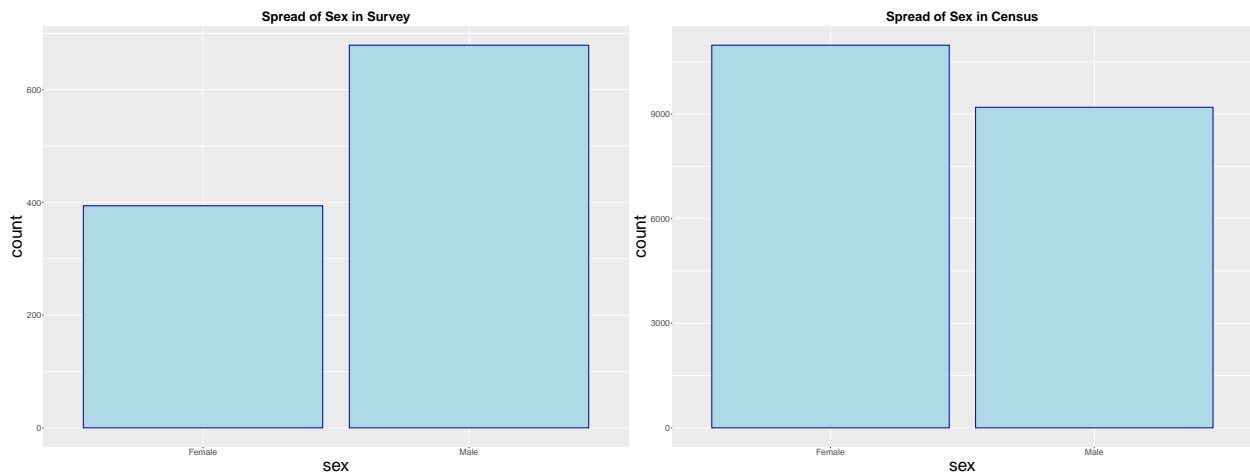
The biggest data limitations were encountered due to the fact that the two datasets being used were conducted by different organizations and for largely different goals. These differences were largely minimized due to our syncing of the two datasets but will still have a slight negative impact on our study. Apart from this, we encountered under representation of data for certain age groups and Provinces in our dataset which will lead to a degree of bias in our models. Same was the case when it comes to the income-level as that effects your views and also it has a high correlation with a person casting a vote or not[19].

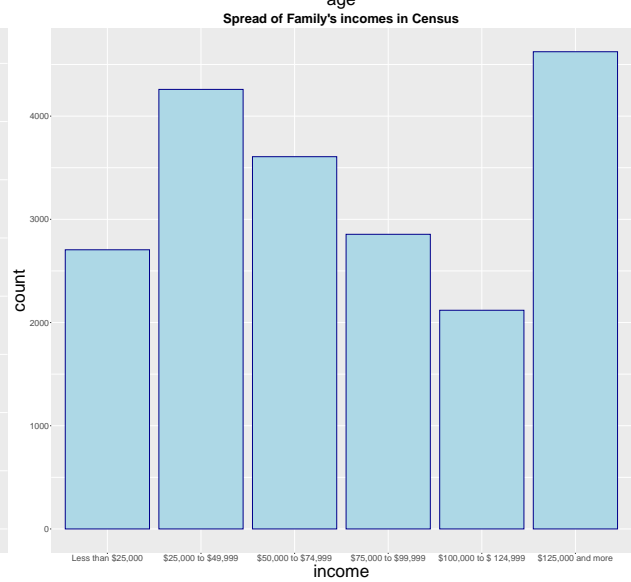
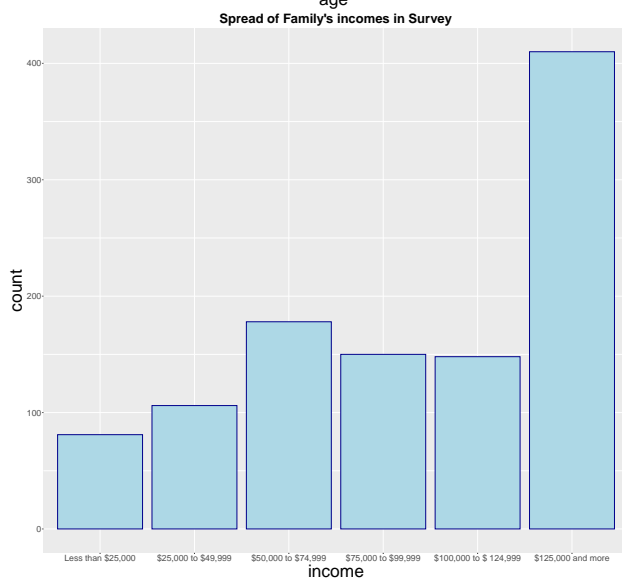
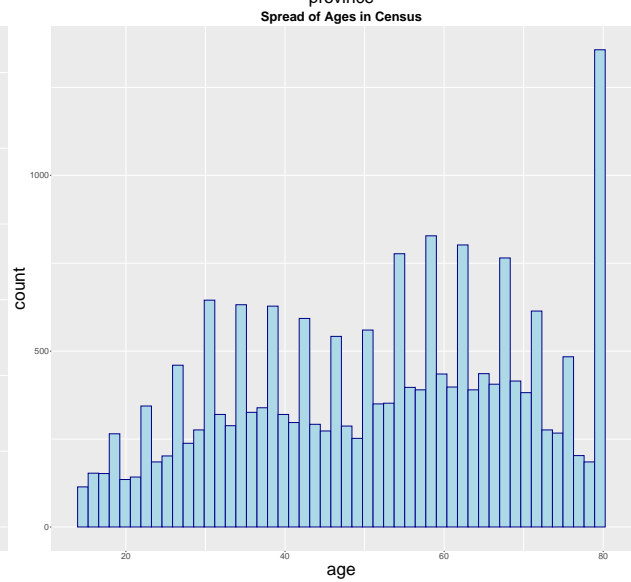
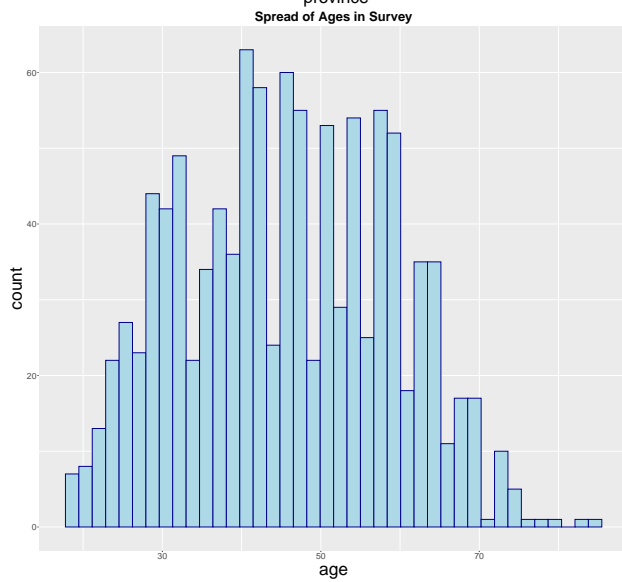
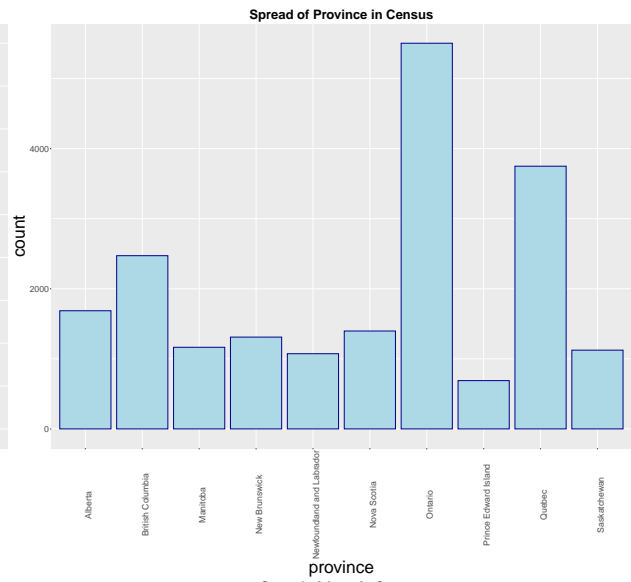
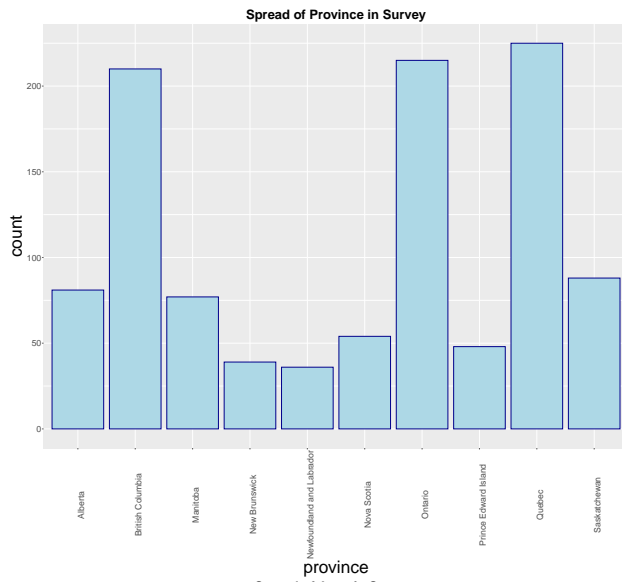
Another thing to consider was that we had information about Gender in one data set and Sex in the other, this not only is a source of error but might also discourage people from the LGBTQ+ group completing the survey.

Lastly, and most importantly we have no data about certain provinces(Northwest Territories, Yukon, Nunavut) and also we don't have enough entries from certain provinces to get much more reliable results. We are also missing questions and data to cross-check their responses.

Important Variable Plots

See the following page for the plot description.





Plot Description

Bar Chart 1 Spread of Sex in Survey:

There is a greater percentage of Male voters in the electoral population of the survey data. More precisely, around 63% of the population consists of Male voters.

Bar Chart 2 Spread of Sex in Census:

Unlike the survey data, the electoral population in the census data consists of more females than males as roughly 46% of the population consists of Male voters.

Bar Chart 3 Spread of Province in Survey:

The bar chart shows the current province the voters of the survey data reside in. The top three provinces that make up the electoral population are British Columbia, Ontario and Quebec as there are over 200 voters from each of these provinces which cover around 61% of the population. Meanwhile, 11% of the population consists of voters in New Brunswick, Newfoundland Labrador and Prince Edward Island with just under 50 voters from each of these provinces. The remaining 28% of the population resides in Alberta, Manitoba, Nova Scotia and Saskatchewan.

Bar Chart 4 Spread of Province in Census:

Similar to the bar chart about the spread of provinces in the survey, most of the population resides in British Columbia, Ontario and Quebec. More exactly, roughly 58% of the voters belong to these provinces. Unlike the survey data which had 20% of voters from British Columbia, the census data records only 12% of voters from this province. Manitoba, New Brunswick, Newfoundland and Labrador, Nova Scotia, and Saskatchewan all have 1000 to 1400 individuals from their province. Least amount of voters are from Prince Edward Island as expected since it is the smallest province in Canada.

Histogram 1 Spread of Ages in Survey:

The data spread closely represents a normal histogram. The mean is around 50 years. A very small amount of the old people (people over the age of 80) vote.

Histogram 2 Spread of Ages in Census:

This histogram is close to a left skewed distribution. Surprisingly, unlike the Spread of the ages in Survey, the max amount of the voters were from the older age group (80 years old).

Histogram 3 Spread of Family's incomes in Survey:

This histogram is close to a left skewed distribution. (FIX GRAPH). In this data, the majority of the population belongs to the \$125,000 and more group. The other groups are almost of the number of people in this group.

Histogram 4 Spread of Family's incomes in Census:

This histogram is close to a right skewed distribution. FIX GRAPH. In the census data too, the maximum amount of the voters have an income of \$125,000 and above. All other groups are decreasing, but there is not as much of a difference as there was in the previous graph.

Methods

Model Choice and Setup

The goal is to make the prediction for the upcoming election. There are a lot of factors which effect someone’s voting preference such as their area of residence, their place of birth, income levels, age, sex and many more. Now in order to account for and see how all these factors effect someone’s decision in choosing a party to vote for we will be using **Regression**. Regression is a statistical technique which measures the relationship as to how the changes in different quantities(independent variable) effect the quantity of interest(dependent variable).

In Canada the party which forms the government is the party which wins majority of the 338 available seats. These sectors are divided and belong to different provinces, the distribution of these seats across the provinces are as follows[16]:

Table 1: Seat distribution across provinces

province	Seats
British Columbia	42
Alberta	34
Saskatchewan	14
Manitoba	14
Ontario	121
Quebec	78
New Brunswick	10
Nova Scotia	11
Prince Edward Island	4
Newfoundland and Labrador	7
Ontario	121
Yukon	1
Northwest Territories	1
Nunavut	1
Total	338

Based on this table we can see how the distribution of the seats is not even across the provinces, hence we need to account for this. Furthermore another thing to note is how political preference has a really high correlation with which province a person belongs to.[14]

Having established how the distribution of seats are not even and how the voting preference differs across the provinces, to account for this we will be running our regression models for each province to get better results.

Another thing to note is that within each province voting preference shifts across different income levels [15]. To account for this we will be using **Multilevel Regression**. Multilevel regression differs from normal regression on the way that in normal regression not accounting for the different levels which effect our data would have an offset in our regression line. The intercept of the line needs to be adjusted to account for the different levels. **NOTE** here in our case “level” refers to the groups within each province, which is *family income level* which we expect to have similar effects on the variable of interest.

Lastly another important aspect about the past elections has been that they have been dominated by particular parties. These political parties also happen to dominate and have majority number of seats in each province, hence estimating and predicting whether they win their particular provinces will acts as an estimate as to whether they will win the federal election and form the government[17].

To make this estimate we are going to hypothesize and setup our model using the *survey data* and train it to estimate whether someone **will be voting for a particular party or not**. To do this we will be using **Multilevel Logistic Regression**, logistic regression is used when we are dealing with a **binary response variable**.

Post stratatification and Estimation

Now in our case we will be first divide our **CENSUS DATA** for each province then then into different **bins**. The concept of bins here is based on grouping the data into different groups and these are the variables which are present in our regression model. Now running this data set on the generated model for each popular party in each province will generate **log odds** of the particular group voting in favor of our hypothesized party.

Now in general the log odds generated from the regression model will look like:

$$\log\left(\frac{\hat{p}_{h_0}}{1 - \hat{p}_{h_0}}\right) = \beta_{0j} + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i \quad i \in [1, k], \quad j \in [1, m]$$

Here β_{0j} is the measure for the level is random intercept which accounts for the different income levels, as we have 6 income levels hence $m = 6$ in our case and β_i , i from $[1, k]$ are the k factors which effect our model.

Although as previously mentioned that there are many variables which effect someone's voting preference but due to the limitation of data we will be dividing our data set in different provinces and then for each province we will be making our model on **family income level**, **sex**, **education level** and **age group**. So the resulting β_i will be the following:

Table 2: β_i and the corresponding measure

β_i	Variable
β_0	Family income-level
β_1	Sex
β_2	Education level
β_3	Age group

From log Odds to Estimating Probability

Now after generating the model and running the model on our test data to get the log odds we would need to extract the probability of that particular person voting for in favour of the party. To do this we would need to follow these 4 steps:

STEP 1:

$$\log\left(\frac{\hat{p}_{h_0}}{1 - \hat{p}_{h_0}}\right) = \beta_{0j} + \sum_{i=1}^k \beta_i x_i \implies \frac{\hat{p}_{h_0}}{1 - \hat{p}_{h_0}} = \exp(\beta_{0j} + \sum_{i=1}^k \beta_i x_i)$$

STEP 2:

$$\hat{p}_{h_0} = (1 - \hat{p}_{h_0})(\exp(\beta_{0j} + \sum_{i=1}^k \beta_i x_i)) \implies \hat{p}_{h_0} = (\exp(\beta_{0j} + \sum_{i=1}^k \beta_i x_i) - \hat{p}_{h_0}(\exp(\beta_{0j} + \sum_{i=1}^k \beta_i x_i)))$$

STEP 3:

$$\hat{p}_{h_0} + \hat{p}_{h_0}(\exp(\beta_{0j} + \sum_{i=1}^k \beta_i x_i)) = \exp(\beta_{0j} + \sum_{i=1}^k \beta_i x_i) \implies \hat{p}_{h_0}(1 + (\exp(\beta_{0j} + \sum_{i=1}^k \beta_i x_i))) = \exp(\beta_{0j} + \sum_{i=1}^k \beta_i x_i)$$

STEP 4:

$$\hat{p}_{h_0}(1 + (\exp(\beta_{0j} + \sum_{i=1}^k \beta_i x_i))) = \exp(\beta_{0j} + \sum_{i=1}^k \beta_i x_i) \implies \hat{p}_{h_0} = \frac{\exp(\beta_{0j} + \sum_{i=1}^k \beta_i x_i)}{(1 + (\exp(\beta_{0j} + \sum_{i=1}^k \beta_i x_i)))}$$

So we have:

$$\hat{p}_{h_0} = \frac{\exp(\beta_{0j} + \sum_{i=1}^k \beta_i x_i)}{(1 + (\exp(\beta_{0j} + \sum_{i=1}^k \beta_i x_i)))}$$

Getting the final results, Post Stratification

Now note by grouping the data, some bins might have a higher weight that it more number of entries compared to the other bins. Hence we need to account for this, we do this using this general formula:

$$\hat{y}^{PS} = \frac{\sum_{i=1}^n N_i \hat{y}_i}{\sum_{i=1}^n N_i} \quad n \in [2, k], \quad \hat{y}^{PS}, \hat{y}_i \in [0, 1]$$

Here \hat{y}^{PS} in our case would be the **expected probability** of the total province sample voting in favor of the political party being tested now. N_i is the size of the bin and \hat{y}_i is the probability assigned to the bin, this is the same as \hat{p}_{h_0} when calculating the probability for bins instead of individuals. Finally k is the number of bins generated in our sample data.

Note the reason for the assumption for $n \in [2, k]$ is because we assume there to be more than 1 different group as without that condition, using this technique is of not much significance.

Results

As mentioned we are generating results for every province, so here we need to first outline the popular political parties in each province. Here is a table of the popular parties in each province[20],[21],[22]:

Table 3: Popular political parties	
province	Seats
Alberta	NDP (New Democratic Party, New Democrats, NDPers) Conservatives (Tory, PCs, Conservative Party of Canada) Liberal (Grits)
British Columbia	NDP (New Democratic Party, New Democrats, NDPers) Conservatives (Tory, PCs, Conservative Party of Canada) Liberal (Grits)
Ontario	NDP (New Democratic Party, New Democrats, NDPers) Conservatives (Tory, PCs, Conservative Party of Canada) Liberal (Grits)
Manitoba	NDP (New Democratic Party, New Democrats, NDPers) Conservatives (Tory, PCs, Conservative Party of Canada) Liberal (Grits)
Quebec	(Coalition Avenir Québec or the CAQ)* (Quebec Liberal Party)* Bloc Québécois (BQ, PQ, Bloc, Parti Québec
New Brunswick	Conservatives (Tory, PCs, Conservative Party of Canada) Liberal (Grits) NDP (New Democratic Party, New Democrats, NDPers)
Saskatchewan	Conservatives (Tory, PCs, Conservative Party of Canada) Liberal (Grits) NDP (New Democratic Party, New Democrats, NDPers)
Nova Scotia	Conservatives (Tory, PCs, Conservative Party of Canada) Liberal (Grits) NDP (New Democratic Party, New Democrats, NDPers)
New Brunswick	NDP (New Democratic Party, New Democrats, NDPers) Liberal (Grits) Conservatives (Tory, PCs, Conservative Party of Canada)
Prince Edward Island	Liberal (Grits) Conservatives (Tory, PCs, Conservative Party of Canada) NDP (New Democratic Party, New Democrats, NDPers)

*Note the given data was not available so we chose the overall popular parties in the region which are Liberals and Conservatives.

POST STRATIFICATION CHANGES

Note that to make the bins and have a healthy number of people in each bin we ended up grouping people in different age groups instead of leaving them as it is. Hence we ended up with grouping the different ages, the age groups were based on the standard age groups in Canada these are as follows [18]:

Table 4: Age group and corresponding category

Age group	Category
People under 18	Under 18
people between 18 and 24	18 to 24 years
people between 25 to 44	25 to 44 years
people between 45 and 64	45 to 64 years
people 65 and above	Over 65

The last thing we did was we had to drop two of our categories which were **feeling_life** and **religion_participation** the reason for doing so was to get a better much more reliable model.

R-Functions used in calculating results Lastly for our study the we used built in R functions for our calculations. To calculate log odds we used **predict(model, data.frame)**, the estimated probability we used a function we made named **est_p** as displayed below. Note the **input** “sum” are the log odds calculated using the predict() function.

```
est_p <- function(sum){
  return(exp(sum)/(1+(exp(sum))))
}
```

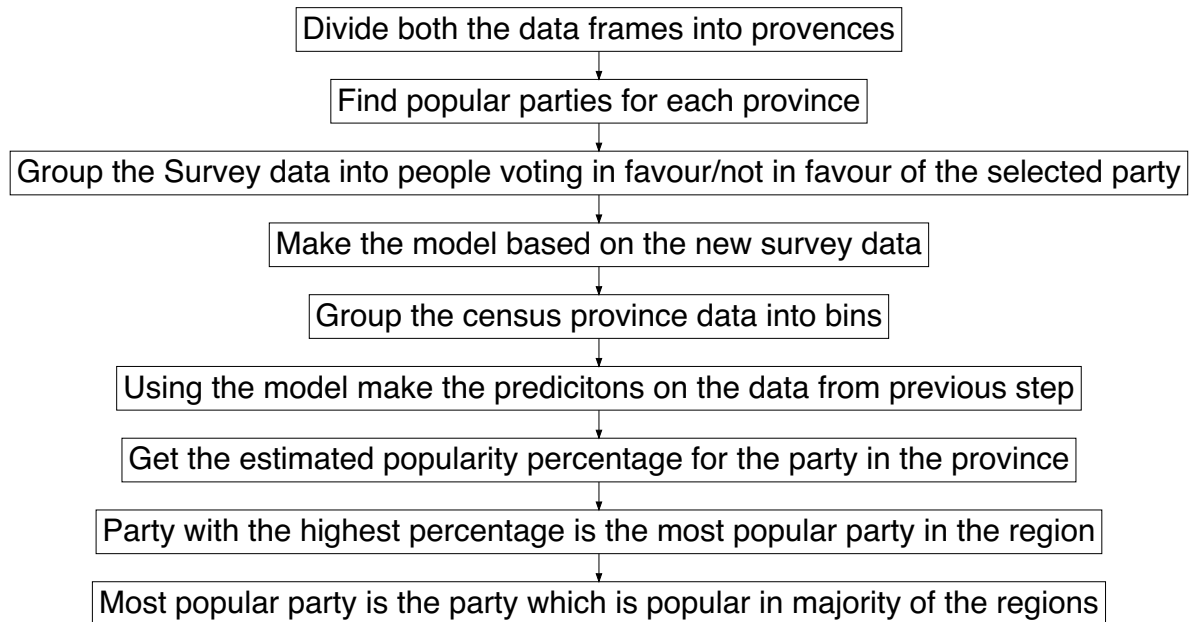
And finally the estimate for our hypothesis is done using this following lines of R-code displayed in the code chunk below.

```
Output <- data.frame %>% summarise(estimate = (sum(n*probability.bin)/N))
```

Here is a table detailing what everything means for in the above code chunk:

Table 5: R-code breakdown	
R-code	Breakdown
<i>n</i>	Bin size
<i>probability.bin</i>	Probability of the group
<i>N</i>	sum of all bin size
<i>sum(...)</i>	Sums the “...” within the brackets
<i>summarise(...)</i>	Generates a summary of the result
<i>data.frame</i>	Data frame containing the Estimated bin probability
<i>Output</i>	variable to store the Estimated Probability of Hypothesis

In summary here are the steps we will be using to calculate our results:



Results for Alberta

As mentioned earlier the top 3 popular parties in Alberta are **NDP (New Democratic Party, New Democrats, NDPers)**, **Conservatives (Tory, PCs, Conservative Party of Canada)** and **Liberal (Grits)**. Here are the results for the top 3 parties:

Table 6: Popular Political Vote in Alberta

Political Party	Party's Popularity
NDP (New Democratic Party, New Democrats, NDPers)	24.95%
Conservatives (Tory, PCs, Conservative Party of Canada)	67.76%
Liberal (Grits)	12.52%

Results for British Columbia

The top 3 popular political parties in British Columbia are **Conservatives (Tory, PCs, Conservative Party of Canada)**, **Liberal (Grits)** and **NDP (New Democratic Party, New Democrats, NDPers)**. Here are the results for the top 3 parties:

Table 7: Popular Political Vote in British Columbia

Political Party	Party's Popularity
NDP (New Democratic Party, New Democrats, NDPers)	23.10%
Conservatives (Tory, PCs, Conservative Party of Canada)	27.72%
Liberal (Grits)	26.75%

Results for Ontario

The top 3 popular political parties in Ontario are **Conservatives (Tory, PCs, Conservative Party of Canada)**, **Liberal (Grits)** and **NDP (New Democratic Party, New Democrats, NDPers)**. Here are the results for the top 3 parties:

Table 8: Popular Political Vote in Ontario

Political Party	Party's Popularity
NDP (New Democratic Party, New Democrats, NDPers)	14.77%
Conservatives (Tory, PCs, Conservative Party of Canada)	27.57%
Liberal (Grits)	26.76%

Results for Manitoba

The top 3 popular political parties in Ontario are **Conservatives (Tory, PCs, Conservative Party of Canada)**, **Liberal (Grits)** and **NDP (New Democratic Party, New Democrats, NDPers)**.

Here are the results for the top 3 parties:

Table 9: Popular Political Vote in Manitoba

Political Party	Party's Popularity
NDP (New Democratic Party, New Democrats, NDPers)	11.98%
Conservatives (Tory, PCs, Conservative Party of Canada)	31.06%
Liberal (Grits)	18.26%

Results for Quebec

The top 3 popular political parties in Ontario are **Coalition Avenir Québec or the CAQ, Quebec Liberal Party** and **Bloc Québécois (BQ, PQ, Bloc, Parti Québécois)** but as mentioned due to the limitation of the data we have we choose to go with the next popular parties and the final list of the of the parties as mentioned previously in Table3 are: **Conservatives (Tory, PCs, Conservative Party of Canada)**, **Liberal (Grits)** & **Bloc Québécois (BQ, PQ, Bloc, Parti Québécois)** and here are the results we found:

Table 10: Popular Political Vote in Quebec

Political Party	Party's Popularity
Bloc Québécois (BQ, PQ, Bloc, Parti Québécois)	24.48%
Conservatives (Tory, PCs, Conservative Party of Canada)	17.04%
Liberal (Grits)	39.63%

Results for New Brunswick

The top 3 popular political parties in New Brunswick are **Conservatives (Tory, PCs, Conservative Party of Canada)**, **Liberal (Grits)** and **Green Party (Greens)**. A thing to mention in making our estimate for this province is that the data we used to make the model for New Brunswick had a issue when it came to the age group, we had no data on the people between 18-24 years so our predictions are also missing the estimates from the people in that age group.

Here are the results for the top 3 parties:

Table 11: Popular Political Vote in New Brunswick

Political Party	Party's Popularity
Green Party (Greens)	25.52%
Conservatives (Tory, PCs, Conservative Party of Canada)	39.70%
Liberal (Grits)	22.97%

Results for Saskatchewan

The top 3 popular political parties in Saskatchewan are **Conservatives (Tory, PCs, Conservative Party of Canada)**, **Liberal (Grits)** and **NDP (New Democratic Party, New Democrats, NDPers)**.

Here are the results for the top 3 parties:

Table 12: Popular Political Vote in Saskatchewan

Political Party	Party's Popularity
NDP (New Democratic Party, New Democrats, NDPers)	11.98%
Conservatives (Tory, PCs, Conservative Party of Canada)	61.75%
Liberal (Grits)	10.16%

Results for Nova Scotia

The top 3 popular political parties in Nova Scotia are **Conservatives (Tory, PCs, Conservative Party of Canada)**, **Liberal (Grits)** and **NDP (New Democratic Party, New Democrats, NDPers)**.

Here are the results for the top 3 parties:

Table 13: Popular Political Vote in Nova Scotia

Political Party	Party's Popularity
NDP (New Democratic Party, New Democrats, NDPers)	15.59%
Conservatives (Tory, PCs, Conservative Party of Canada)	38.78%
Liberal (Grits)	34.827%

Results for Newfoundland and Labrador

The top 3 popular political parties in Newfoundland and Labrador are **Conservatives (Tory, PCs, Conservative Party of Canada)**, **Liberal (Grits)** and **NDP (New Democratic Party, New Democrats, NDPers)**. A thing to mention in making our estimate for this province is that the data we used to make the model for Newfoundland and Labrador had a issue when it came to the age group, we had no data on the people between 18-24 years so our predictions are also missing the estimates from the people in that age group. Furthermore due to the limitation of the data size available about Conservatives (Tory, PCs, Conservative Party of Canada) we were unable to make estimation for this party.

Here are the results for the top 2 parties:

Table 14: Popular Political Vote in Newfoundland and Labrador

Political Party	Party's Popularity
NDP (New Democratic Party, New Democrats, NDPers)	38.12%
Liberal (Grits)	33.71%

Results for Prince Edward Island

The top 3 popular political parties in Prince Edward Island are **Conservatives (Tory, PCs, Conservative Party of Canada)**, **Liberal (Grits)** and **NDP (New Democratic Party, New Democrats, NDPers)**. Here are the results for the top 3 parties:

Table 15: Popular Political Vote in Prince Edward Island	
Political Party	Party's Popularity
NDP (New Democratic Party, New Democrats, NDPers)	9.86%
Conservatives (Tory, PCs, Conservative Party of Canada)	20.73163%
Liberal (Grits)	41.70%

Summary, comments about results

Table 16: Provincial Seats along with Dominant Political Preference

Province	Seats	Dominant Party	Percentage Preference
British Columbia	42	Conservatives (Tory, PCs, Conservative Party of Canada)	27.72%
Alberta	34	Conservatives (Tory, PCs, Conservative Party of Canada)	67.76%
Saskatchewan	14	Conservatives (Tory, PCs, Conservative Party of Canada)	61.75%
Manitoba	14	Conservatives (Tory, PCs, Conservative Party of Canada)	31.06%
Ontario	121	Conservatives (Tory, PCs, Conservative Party of Canada)	27.57%
Quebec	78	Liberal (Grits)	39.63%
New Brunswick	10	Conservatives (Tory, PCs, Conservative Party of Canada)	39.70%
Nova Scotia	11	Conservatives (Tory, PCs, Conservative Party of Canada)	38.78%
Prince Edward Island	4	Liberal (Grits)	41.70%
Newfoundland and Labrador	7	NDP (New Democratic Party, New Democrats, NDPers)	38.12%
Yukon	1	NA	NA
Northwest Territories	1	NA	NA
Nunavut	1	NA	NA
Most Dominant Party		Conservatives (Tory, PCs, Conservative Party of Canada)	

Based on results displayed in Table 16 we can see that the most popular party on average is the Conservatives (Tory, PCs, Conservative Party of Canada) so we **expect** them to win. However there are still some things we need to take into account, firstly the mostly if we assume that all the people who choose Conservatives as their political preference they would still have about **~94** out the total 338. Note to have a majority a party would need above half of the total available seats which come to 170.

It is now a good time to also note that in some regions the party that was 2^{nd} when it comes to popularity was just behind the party that came 1^{st} in that region. Because now if the other party dominates the some other region with a higher seat count with a greater percentage difference then this political party would end having more seats. This is something we need to consider when making a prediction about the federal election and base our results. Furthermore governments are also formed on the basis of independent representatives joining or other parties getting together to form a majority government, note we don't have this information available to us. Lastly as mentioned we had no information about certain groups or about certain political parties to make our model so we were not able to make predictions about those groups or about those parties in our census data.

In summary however these results seem realistic when compared on a provincial level and also when we compare them to the most recent election from 2019 but there are still much that can be done to get better much more accurate prediction.[23]

Conclusions

Summary

Our main goal of the study was to predict the popular vote in Canada for the next federal election. Based on the hypothesis of our study we found the popular vote in each province and then using that made a conclusion on the overall popular vote in Canada.

To achieve our results we used multilevel-logistic regression to find the popular vote and based on the results generated we then compared the results to the number of seats in that province to get an estimate of the number of seats the party would secure. Now based on this we made a prediction about whether the party would win the federal election or not.

In our case we found that Conservatives (Tory, PCs, Conservative Party of Canada) were the overall popular party in the majority of the regions and based on their popularity percentage in each region we expect them to secure about 94 seats out of the 338 available seats. Note however this is not the required number of seats needed to form a government and we expect that if Conservatives (Tory, PCs, Conservative Party of Canada) secure this many seats they would still need support from other parties to reach the minimum number of 170 needed to form the government. When we compare this result to the last election which took place in 2019, we can see that firstly there was an increase in the number of seats secured by the Conservatives (Tory, PCs, Conservative Party of Canada). We can also see that there was a strong dominance by certain parties in certain provinces and also a strong competition between the top 2 parties; Conservatives (Tory, PCs, Conservative Party of Canada), and Liberals in some regions[23]. We can see that the predictions we made based on the data fall close to the election which took place in 2019, but as mentioned before we need to be careful and account for the possible biases when it comes to the quality of the data being collected such as biases involved in people filling the surveys and also the drawbacks of making inference using limited categories and insufficient amounts of data.

Looking forward

For improvement and making better inference, we need to consider ways of improving our data, methods and perhaps even our goals. As methods are derived and dependent on the data let us first discuss about possible improvements in data.

Firstly as mentioned, the main issue when it comes to data, that we used outdated data in the age of social media and online presence, opinions are influenced and can even change in any moment. So the best predictor or way to collect data would be to find an **ethical** way of collecting data online. This would also result in the most up to date form of data, furthermore we can also collect data at regular intervals to see the trend of how opinions and preferences change from time to time and also perhaps cross-check the data we are collecting.

Now if we don't have these means to collect data we should firstly use the most updated version of the data, secondly we should try to get information based on every region and perhaps add more variables that way we can add more variables to our model and get better results.

Lastly with a quality set of data we can also perhaps change our goal and the model to get a result of how many seats each party is expected to win in every province. That we we can have a good idea about the seats won by the other minor parties and how that influences the overall majority government formation.

Bibliography

1. 1867 Canadian federal election. (n.d.). DBpedia. https://dbpedia.org/page/1867_Canadian_federal_election
2. Bill C-16: An Act to amend the Canada Elections Act. (n.d.). Library Of Parliament. Retrieved May 30, 2021, from https://lop.parl.ca/sites/PublicWebsite/default/en_CA/ResearchPublications/LegislativeSummaries/391LS530E#fprovincial
3. Did you know? (n.d.). Elections Canada. Retrieved May 30, 2021, from <https://www.elections.ca/content.aspx?section=abo&dir=100/kno&document=index&lang=e>
4. Canadian Election Results by Party 1867 to 2019. (n.d.). Elections. <https://www.sfu.ca/~aheard/elections/1867-present.html>
5. Elections Step by Step. (n.d.). Elections Canada. <https://electionsanddemocracy.ca/canadas-elections/canadas-election-process/elections-step-step>
6. Despite global concerns about democracy, more than half of countries are democratic. (n.d.). Pew Research Center. <https://www.pewresearch.org/fact-tank/2019/05/14/more-than-half-of-countries-are-democratic/>
7. Functions of elections. (n.d.). Britannica. Retrieved May 30, 2021, from <https://www.britannica.com/topic/election-political-science/Functions-of-elections>
8. Welcome to the 2019 Canadian Election Study. (n.d.). Canadian Election Study. Retrieved May 30, 2021, from <http://www.ces-ec.ca/>
9. General Social Survey. (2017, September). https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/sdaweb/dli2/gss/gss31/gss31/more_doc/GSS31_User_Guide.pdf
10. Fournier, P. J. (2020, June 28). The biggest divide in Canadian politics? Men vs. Women. Macleans. <https://www.macleans.ca/politics/ottawa/the-biggest-divide-in-canadian-politics-men-vs-women/>
11. Holland, J. L. (2014, March 7). Age Gap? The Influence of Age on Voting Behavior and Political Preferences in the American Electorate. Washington State University. <https://research.libraries.wsu.edu/xmlui/handle/2376/4982>
12. Factors associated with voting. (2012, February 24). Statistics Canada. <https://www150.statcan.gc.ca/n1/pub/75-001-x/2012001/article/11629-eng.htm>
13. Akee, R. (2020, September 24). Voting and Income | Econofact. Econofact | Key Facts and Incisive Analysis to the National Debate on Economic and Social Policies. <https://econofact.org/voting-and-income>
14. Foladare, I. S. (1986, December). The Effect of Neighborhood on Voting Behavior on JSTOR. JSTOR. <https://www.jstor.org/stable/2146812?seq=1>
15. H.Y.A.E.M.M. (2014, January 8). The Relationship Between Income Level and Political Preferences for University Academic Staff: The Case of Hakkari. ScienceDirect. <https://www.sciencedirect.com/science/article/pii/S1877042813052051>
16. House of Commons Seat Allocation by Province – Elections Canada. (n.d.). Elections Canada. Retrieved May 31, 2021, from <https://www.elections.ca/content.aspx?section=res&dir=cir/red/allo&document=index&lang=e>
17. Wikipedia contributors. (2021, May 24). Politics of Canada. Wikipedia. [https://en.wikipedia.org/wiki/Politics_of_Canada#:~:text=The%20two%20dominant%20political%20parties,well%20as%20its%20numerous%20predecessors\).](https://en.wikipedia.org/wiki/Politics_of_Canada#:~:text=The%20two%20dominant%20political%20parties,well%20as%20its%20numerous%20predecessors).)

18. Statista. (2021, March 9). Canada - resident population, by gender and age group 2020. <https://www.statista.com/statistics/444858/canada-resident-population-by-gender-and-age-group/>
19. This is the link between voting in elections and income. (n.d.). World Economic Forum. Retrieved May 31, 2021, from <https://www.weforum.org/agenda/2018/07/low-voter-turnout-increasing-household-income-may-help>
20. Politics of Ontario. (n.d.). Politics Of Ontario. Retrieved May 31, 2021, from https://en.wikipedia.org/wiki/Politics_of_Ontario#:~:text=Ontario's%20current%20primary%20political%20parties,the%20Green%20Par
21. Politics of Manitoba. (n.d.). Wikipedia. Retrieved May 31, 2021, from https://en.wikipedia.org/wiki/Politics_of_Manitoba#:~:text=Manitoba's%20primary%20political%20parties%20are,Progressive%20Conservative%2
22. Politics of Canada. (n.d.). Wikipedia. Retrieved May 31, 2021, from https://en.wikipedia.org/wiki/Politics_of_Canada#:~:text=Five%20parties%20had%20representatives%20elected,the%20Green%20Party%20of%20
23. File:Canada 2019 Federal Election.svg. (n.d.). Wikipedia. Retrieved May 31, 2021, from https://en.wikipedia.org/wiki/File:Canada_2019_Federal_Election.svg
- *24. Add an Image to Your Report. (n.d.). Earth Lab. <https://www.earthdatascience.org/courses/earth-analytics/document-your-science/add-images-to-rmarkdown-report/>
25. Bloomberg.com Canada Election 2019: Live Results. (n.d.). [Illustration]. <https://www.bloomberg.com/graphics/2019-canada-election-results/img/2019-canada-election-results-facebook.png>

*citation in R chunks in the markdown