

2019 Canadian Federal Election Result Analysis

Ruoxi Guan

Dec.21, 2020

Code and data supporting this analysis is available at:<https://github.com/ElsaGrx/ElsaGrx-2019-Canadian-Federal-Election-Result-Analysis>

Abstract

The outcome of the Canadian Federal Election usually has a huge impact on the country's policy, industry, economy and even every citizen, so numerous reports about forecasting the result of election will be published during each election period. In the 2019 Canadian Federal Election, which had 67% overall voter turnout, the Liberals had a narrow victory, despite the fact that they received the lowest percentage of popular votes of a governing party in Canadian history. A multilevel regression (a generalized linear mixed-effects regression model) with post-stratification analysis based on two datasets, "2019 CES Online Survey" and "Education Highlight Tables, 2016 Census", is performed in the report, supported by some graphs and tables. The final estimated outcome for the 2019 Canadian Federal election is that the Liberal Party would win seven seats in the House of Commons with 27.2% popular vote, indicating that Liberal party would be unlikely to win the 2019 Canadian Federal Election if "everyone" had voted.

Keywords

MRP, glmer, 2019 Canadian Federal Election, Prediction, Liberal Party, Different Outcome.

Introduction

Numerous predictions and opinions will be published during each federal election period. The outcome of the Canadian Federal Election may have a huge impact on the country's policy, industry, economy, even every citizen. By applying statistics to voter preferences and analyzing the statistical model, candidates can design new approaches to attract more voters. For example, the "simple seat project model" and "Poll Aggregation Methodology" were introduced by Rosenthal (Rosenthal, 2011) and Eric (Eric, ThreeHundredEight.com) when they tried to predict the election results. Due to a dramatically increasing demand for information about every aspect of the world, there are extensive pollings on important policy issues, such as the Federal Election. Although some of the surveys are not representative and accurate enough in the sense of having small samples or unprofessional design of the surveys, they are useful for fitting multilevel regression with post-stratification(MRP) - another excellent technique to predict the election results.

In the 2019 Canadian Federal Election, the Liberals had a narrow victory, despite the fact that they lost the popular vote. Compared to 184 seats with 39.47% popular vote that the Liberal Party won in the 2015 Canadian Federal Election, it only won 157 seats with 33.12% popular vote in 2019. The overall voter turnout in 2019 also decreased from 68.3% in 2015 to 67%. It would be interesting to investigate whether the 2019 election result will be different if all citizens vote, which can be done by applying MRP. In general,

MRP is a statistical technique to correct estimates when there are known differences between the target population and study population. Building upon ideas of Fay and Herriot, Gelman and T.Little initially developed MRP, which was further expanded by Park, Gelman, Bafumi, Lax, Phillips, Warshaw, Rodden, and Wang et al. (Multilevel regression with poststratification 2020).

All Canadian citizens who are at least 18 years old on election day and can provide acceptable proof of identity and address are eligible to register and vote in a federal election. The electoral system of Canada is referred to as a “single-member plurality” system. Representation in the House of Commons is based on geographical divisions called electoral districts, also known as ridings (Election Canada). In each of 338 electoral districts, the candidate wins a seat in the House of commons if the candidate has the highest number of votes in the district. The candidate who has the most seats in the House of commons wins. The goal of this project is to build a MRP model based on CES and 2016 census of population, identifying how the 2019 Canadian Federal Election would have been different if “everyone” had voted. The two datasets, “2019 CES Online Survey” and “Education Highlight Tables, 2016 Census”, are obtained from the “2019 Canadian Election Study” and “Statistics Canada, 2016 Census of Population”, respectively.

Two data sets (CES is the survey data and 2016 Census data is the Census) will be used to perform MRP and further predict the final results of 2019 Canadian Federal Election if “everyone” had voted. Data analysis and model building techniques will be included in the Methodology Section. The result section will include model analysis, tables, and graphs. Summary, conclusion, weaknesses, and next steps of this project will be in the Discussion. References and Appendix will be listed in the end.

Methodology

Data Cleaning Process

As mention above, the two datasets, “2019 CES Online Survey” and “Education Highlight Tables, 2016 Census”, are obtained from the “2019 Canadian Election Study” and “Statistics Canada, 2016 Census of Population”, respectively. CES will be the survey data and 2016 census data will be the census data. For the survey data (CES 2019 online survey), only those observations that are eligible to vote (Canadian citizens who are 18 years old or older) are selected. Since some observations will not vote or do not specify which party that they will vote, those observations are removed. “Cps19_votechoice” is the response variable, which has been converted into a binary variable (“Liberal Party” and “Other Party”) for simplicity. By setting “Other Party” as the reference, the further analysis aims to predict the probability of voting for Liberal Party. The census data (2016 census) is in the cartesian data frame, so measures are at the intersections of each group and number. By pivoting the data, the measure can be looked up by the corresponding group and number. Some duplicate observations are created due to pivoting, such as “both sexes” and “total education”, so those observations have been removed as well. N/A observations are removed in both data.

In order to ensure that variables’ names and categories in the cleaned survey data can be matched to those in the census data, further data cleaning processes are required. There are six age groups in the census data: “25 to 34”, “25 to 64”, “ 35 to 44”, “45 to 54”, “55 to 64”, “and all ages 15-plus”. Noticing that “24 to 64” and “all ages 15 plus” contain duplicate observations, a new age group can be created by subtracting 25 to 64 years old from all ages, so that the new age group contains the observations that are 15 to 24 years old or older then 64. After splitting the age into the same age group in survey data, age can be matched in two data. Sex and Province in two datasets are in the similar format, so that they can be matched by renaming and regrouping. For the education, some assumptions are introduced in order to match two datasets (Appendix #1).

Table 1: Baseline Characteristics of the Survey Data

	Liberal Party(N=8898)	Other Party(N=22437)	Overall(N=33644)
Age			
25 to 34	1394 (15.7%)	3689 (16.4%)	5434 (16.2%)
35 to 44	1533 (17.2%)	3933 (17.5%)	5768 (17.1%)
45 to 54	1443 (16.2%)	3903 (17.4%)	5679 (16.9%)
55 to 64	1820 (20.5%)	4685 (20.9%)	7057 (21.0%)
Other	2708 (30.4%)	6227 (27.8%)	9706 (28.8%)
Sex			
Female	5122 (57.6%)	13206 (58.9%)	19634 (58.4%)
Male	3776 (42.4%)	9231 (41.1%)	14010 (41.6%)
Education			
College, CEGEP or other non-university certificate or diploma (2016 counts)	2550 (28.7%)	7795 (34.7%)	11013 (32.7%)
Other	97 (1.1%)	252 (1.1%)	368 (1.1%)
Secondary (high) school diploma or equivalency certificate (2016 counts)	1328 (14.9%)	4625 (20.6%)	6307 (18.7%)
University certificate or diploma below bachelor level (2016 counts)	4923 (55.3%)	9765 (43.5%)	15956 (47.4%)
Which province or territory are you currently living in?			
Alberta	557 (6.3%)	3150 (14.0%)	4016 (11.9%)
British Columbia	925 (10.4%)	2618 (11.7%)	3824 (11.4%)
Manitoba	346 (3.9%)	1045 (4.7%)	1484 (4.4%)
New Brunswick	223 (2.5%)	490 (2.2%)	774 (2.3%)
Newfoundland and Labrador	209 (2.3%)	324 (1.4%)	559 (1.7%)
Northwest Territories	8 (0.1%)	13 (0.1%)	21 (0.1%)
Nova Scotia	322 (3.6%)	560 (2.5%)	912 (2.7%)
Nunavut	8 (0.1%)	12 (0.1%)	22 (0.1%)
Ontario	4077 (45.8%)	8166 (36.4%)	13033 (38.7%)
Prince Edward Island	42 (0.5%)	91 (0.4%)	149 (0.4%)
Quebec	2039 (22.9%)	4953 (22.1%)	7626 (22.7%)
Saskatchewan	135 (1.5%)	991 (4.4%)	1192 (3.5%)
Yukon	7 (0.1%)	24 (0.1%)	32 (0.1%)

Table 1

Age, Sex, Province and Education are the four variables that have been selected to fit the model. One the most important reason to choose those four variables is that they are the only four variables that are generated from the census data. Of 31,335 observations in the survey data, 8,898 observations will vote for the liberal party and 22,437 will not. Compared to males, more females register the vote, occupying 58.5% of the overall sample size. In addition, there is a positive correlation between the voter turnouts and voters' education levels. Most of the voters are from Ontario, Quebec, Alberta and British Columbia. Since the census data is originally in the cartesian data frame, measures are at the intersections of each group and number. After pivoting the data, each observation is weighted by their population. A baseline characteristics table for the census data does not provide extra information.

Model

Model Specifics

In this project, there is no assumption that parameters follow any distributions, so a frequentist approach is being used. The predictions are on the underlying truths of the experiment using only two datasets. Since MRP is using in this project and the response variable in the survey data is binary, fitting a generalized linear mixed-effects regression model (Using `glmer()` under package "lme4" in R) with a family of binomials

to predict the proportion of voters who will vote for the Liberal Party. Here, we set that if the predicted probability is larger than 0.3592, Liberal Party wins the election. The threshold is the average of 5 popular votes for the winner of the minority government. In other words, according to the historical data, a party will win a minority government on average if its popular vote is larger than 0.3592.

Age, Sex, Province, and Education are the four variables that have been selected to fit the model since the limitation of the census data. Before fitting the model, a self-defined cell with two variables, "Age" and "Sex", is created. Since "Age" has 5 categories and "Sex" has 2 categories, the population has been partitioned into 10 cells. By doing so, the response variable per cell of the census data can be estimated using the model that is built based on the survey data. Additionally, the values of intercept and coefficients of education and sex are expected to change as cells change. Thus, the model contains both random intercept and random coefficients. The generalized linear mixed-effects regression model is:

$$P(Y_i = \text{Vote for Liberal Party} \mid \text{cell}_j) = \text{logit}^{-1}(\alpha + \alpha_j + \beta_{j[i]}^{Education} + \beta_{j[i]}^{Sex} + \beta_{[i]}^{Age} + \beta_{[i]}^{Province})$$

Where $P(Y_i = \text{Vote for Liberal Party} \mid \text{cell}_j)$ represents the probability that respondents vote for the liberal Party, depending on the cell membership of the i^{th} respondent. α is the intercept baseline, and α_j is a random variable that follows $N(0, \sigma_\alpha^2)$, which can be represented by the difference between baseline and the intercept of each cell of the i^{th} respondent. The terms $\beta_{j[i]}^{Education}$ and $\beta_{j[i]}^{Sex}$ correspond to the varying coefficients associated with education and sex, which can be interpreted as the difference between the slope baseline and the coefficient of each cell of the i^{th} respondent. Here, the subscript $j[i]$ indicates the cell to which the i^{th} respondent belongs. For example, $\beta_{j[i]}^{Sex}$ takes values to form $\{\beta_{male}^{Sex}, \beta_{female}^{Sex}\}$ depending on the cell membership of the i^{th} respondent. The random coefficients $\beta_{j[i]}^{Sex}$ and $\beta_{j[i]}^{Education}$ follow $N(0, \sigma_{Sex}^2)$ and $N(0, \sigma_{Education}^2)$, respectively. $\beta_{[i]}^{Age}$ and $\beta_{[i]}^{Province}$ are the terms with constant slope that will not be affected as we change among j cells. The reference categories are "25 to 34" and "Alberta" for variables "Age" and "Province". The probability of an observation that is the i^{th} category to vote for the Liberal Party is $\beta_{[i]}^{Age}$ or $\beta_{[i]}^{Province}$ times the probability of observation in the corresponding reference category to vote for the Liberal Party, controlling for other covariates.

Model Checking

Besides the complexity, the accuracy of the model is also significant. Another model, model1, is built in order to do the model comparison. In model1, "Age" has a random coefficient and "Sex" has constant coefficient, keeping other variables the same as the previous model. The AIC is lower for the previous model, which means that it fits the data better in the sense of having fewer variables and higher accuracy. In addition, the AUC of the previous model is 0.6174 (see Appendix #2, Figure 2), which indicates that the chosen model can discriminate between voting for the liberal party and voting for another party 61.74% of the time. Thus, we can conclude that the original model fits the data better by comparing AIC and AUC of both models.

Results

Data

Table 2: Summary of Voting Status in each Province

Table 2: Summary of Vote %, by Party by Province

Party Standings	Total Votes	AB	BC	MB	NS	ON	QC	SK
Liberal Party	8949	15.07	26.02	24.66	36.41	33.26	29.13	12.02
Conservative Party	8713	53.87	26.44	35.04	17.02	26.28	14.74	46.24
ndp	4328	10.50	17.43	15.28	12.63	15.88	9.01	19.63
Bloc Quebecois	1404	0.00	0.00	0.00	0.00	0.00	19.99	0.00
Green Party	2456	4.04	12.62	7.53	11.16	7.78	6.61	4.69
People's Party	605	2.08	1.73	1.14	1.92	2.06	1.78	2.21

Table 2 summarizes the voter turnout according to each party and then each province. “AB” stands for Alberta; “BC” stands for British Columbia; “MB” stands for Manitoba; “NS” stands for “Nova Scotia”; “ON” stands for “Ontario”; “QC” stands for “Quebec”; “SK” stands for “Saskatchewan”. About half of voters in Alberta and Saskatchewan choose to vote Conservative party. On the other hand, Liberal Party has the highest voter turnout in Nova Scotia, Ontario, and Quebec. Table 1 indicates that Ontario and Quebec have most of the votes, so having a higher voter turnout in those two provinces is vital.

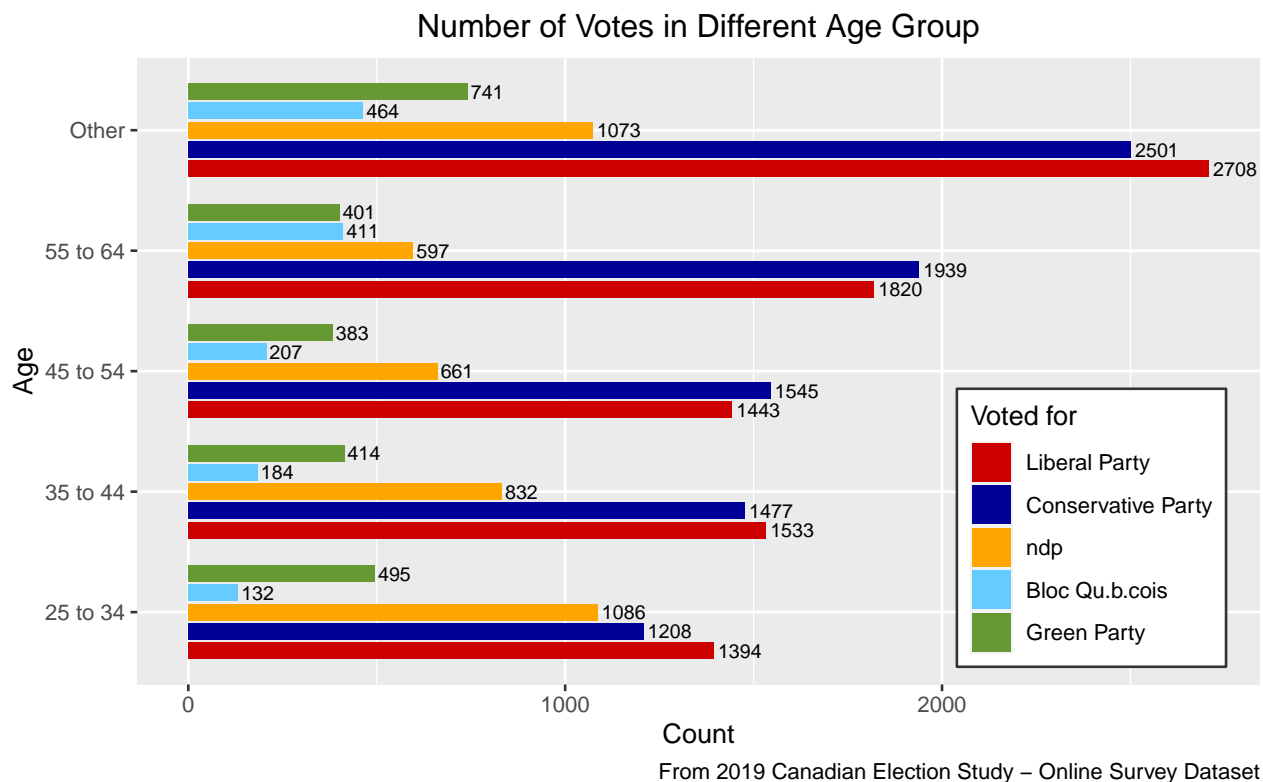


Figure 1: Number of Votes in Different Age Group

Figure 1 shows the results of votes based on different age groups. Most voters are younger than 25 or older than 65 and most of them vote for the Liberal party. The Liberal party also gains most of the supporters from age groups of 25 to 34 and 35 to 44. Also, there is a positive correlation between the age group and number of votes. The older age group has a higher total number of votes whereas the younger age group has a lower total number of votes. Overall, the total number of votes for Liberal is the largest among all parties in the survey dataset.

Model

Table 3: Summary of Voting Status in each Province

Table 3: Summary Table of Model

effect	group	term	estimate	std.error	statistic	p.value
fixed	NA	(Intercept)	-1.986	0.075	-26.643	0.000
fixed	NA	Age35 to 44	0.001	0.074	0.010	0.992
fixed	NA	Age45 to 54	0.029	0.075	0.390	0.697
fixed	NA	Age55 to 64	0.117	0.082	1.418	0.156
fixed	NA	AgeOther	0.202	0.067	2.995	0.003
fixed	NA	ProvinceBritish Columbia	0.668	0.060	11.104	0.000

Table 3 shows the first six observations of the summary table, which summarizes the model results. All the intercepts and coefficients for the chosen variables can be obtained from this table. The P-values for each category in different variables can also be obtained, which can be used to check the statistical significance of the variable, and further analysing whether it would be influential on the outcome.

Post-Stratification

If the proportion of voters who are willing to vote for the Liberal party is more significant than 35.92%, the multilevel regression model will predict the Liberal party will win the election. The post-stratification estimate $\hat{y}^{PS} = 0.272$. Since $0.272 < 0.3592$, the Liberal party will not be the winner.

Table 4: Predicted election results

Table 4: Predicted Election Result

Party	Total Votes
Liberal	7
Other	331

Table 4 shows the total votes of the Liberal party. The “Total_pop” variable in census data is a weighted variable, which means that each observation is differently weighted. Thus, the census data is adjusted by the “Total_pop” variable during the process of calculating the total number of votes. The above total votes are obtained by calculating the total votes in each province, then summing them up. However, in reality, the Canadian Federal Election is according to the electoral district as mentioned in the introduction, rather than by the province. Thus, the seats in a province can be occupied by all the parties, but here considering the seats in a province can only be occupied by one party. Under assumptions, of 338 seats, Liberal Party only wins seven seats in the House of Commons.

Discussion

Summary

The very first step of the prediction about the 2019 Canadian Federal Election is cleaning both the survey (2019 CES Online Survey) and the census (Education Highlight Tables, 2016 Census) datasets, ensuring all the selected variables can be matched in two datasets. After getting a general impression of the data by checking its baseline characteristics, a generalized linear mixed-effects regression model has been fitted, which has a cell variable that contains Age and Sex. Comparing models with AIC and ROC curve, Education and sex are the two random coefficients of the model. The post-stratification estimate is calculated in order to predict the popular vote of Liberal Party. A table of predicted total votes is also displayed. By combining all information gained from data, a final prediction is made.

Conclusion

The estimated proportion of voters in favour of voting for the Liberal Party is 0.272. This is based on the post-stratification analysis of the proportion of voters in favour of the Liberal Party modelled by a generalized linear mixed-effects regression model, which accounted for age, sex, province, education, and cell (sex and age). Since winners of the Federal Election will receive at least (since calculated by the minority government) 35.92% on average based on the historical data, Liberal Party is unlikely to win the election in the aspects of having a low popular vote. The prediction of the seats also provides the same answer, with only 7 seats in the House of commons, it is unlikely for the Liberal Party to win the election if “everyone” has voted.

Thus, the final prediction of the 2019 Canadian Federal Election is that Liberal Party will not win if everyone has voted, which is different to reality that Liberal Party won the election with 67% overall turnout rate. In reality, the liberals lost the popular vote but finally had a narrow victory. With the popular vote of 33.12%, the Liberals received the lowest percentage of the popular vote of a governing party in Canadian history. Therefore, it is reasonable that its position is unstable and loses the election if there are more voters.

Weakness

As described in the post-stratification section, the total votes is obtained by calculating the total votes in each province, then summing them up. However, in reality, the Canadian Federal Election is according to the electoral district as mentioned in the introduction, rather than by the province. Thus, the seats in a province can be occupied by all the parties, but here considering the seats in a province can only be occupied by one party. Moreover, the threshold of the predicted probability for the winner is calculated based on the historical data. However, past performances do not necessarily represent present or future. In addition, the Canadian census program provides a statistical portrait of the country every five years, so the most recent census data is the 2016 census of population. Since the census data was obtained from the 2016 census, not 2019, the final prediction might be affected by this issue.

Next Steps

According to the system of Canadian Federal Election, the final result strongly relies on the winner in each electoral district, which means that the electoral districts can be a good factor of predicting the election winner. This can be done by finding a larger survey dataset to support the model. Some other models such as multinomial regression models may also work.

Reference

- Elections Canada. (2013, January 14). Retrieved December 09, 2020, from <https://www.elections.ca/>
- Cite “2019 CES Online Survey”: Stephenson, Laura B; Harell, Allison; Rubenson, Daniel; Loewen, Peter John, 2020, “2019 Canadian Election Study - Online Survey”, <https://doi.org/10.7910/DVN/DUS88V>, Harvard Dataverse, V1
- Cite “2016 census data”: Statistics Canada. 2016, “Highest level of educational attainment (detailed) by sex and selected age groups”, Census of Population, Catalogue no. 98-402-X2016010
- Andrew Gelman, Thomas C. Little, 1997, “Post Stratification Into Many Categories Using Hierarchical Logistic Regression”. *Survey Methodology*, December 1997, Vol.23, No.2, pp. 127-135. Statistics Canada
- Multilevel regression with poststratification. (2020, October 14). Retrieved December 09, 2020, from https://en.wikipedia.org/wiki/Multilevel_regression_with_poststratification
- Eric. (1970, January 01). ThreeHundredEight.com. Retrieved December 09, 2020, from <http://www.threehundredeight.com/> List of State Electoral Votes For 2020. (n.d.). Retrieved December 09, 2020, from <https://state.1keydata.com/state-electoral-votes.php>
- Jeffrey S. Rosenthal. Was the conservative majority predictable? *The Canadian Journal of Statistics*, 39(4):721–733, 2011.
- Yuchen Cong, Jingwen Deng, Ruoxi Guan, Yuwei Sun, Nov.2, 2020, “2020 United States Presidential Election Prediction”
- Cite R : R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Cite “cesR” : Paul A. Hodgetts and Rohan Alexander (2020). cesR: Access the CES Datasets a Little Easier.. R package version 0.1.0.
- Cite “tidyverse”: Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>
- Cite “visdat” : Tierney N (2017). “visdat: Visualising Whole Data Frames.” *JOSS*, 2(16), 355. doi: 10.21105/joss.00355 (URL: <https://doi.org/10.21105/joss.00355>), <URL: <http://dx.doi.org/10.21105/joss.00355>>.
- Cite “skmr” : Elin Waring, Michael Quinn, Amelia McNamara, Eduardo Arino de la Rubia, Hao Zhu and Shannon Ellis (2020). skmr: Compact and Flexible Summaries of Data. R package version 2.1.2. <https://CRAN.R-project.org/package=skmr>
- Cite “labelled” : Joseph Larmarange (2020). labelled: Manipulating Labelled Data. R package version 2.7.0. <https://CRAN.R-project.org/package=labelled>
- Cite “dplyr” : Hadley Wickham, Romain Francois, Lionel Henry and Kirill Muller (2020). dplyr: A Grammar of Data Manipulation. R package version 1.0.2. <https://CRAN.R-project.org/package=dplyr>
- Cite “foreign” : R Core Team (2020). foreign: Read Data Stored by ‘Minitab’, ‘S’, ‘SAS’, ‘SPSS’, ‘Stata’, ‘Systat’, ‘Weka’, ‘dBase’, R package version 0.8-75. <https://CRAN.R-project.org/package=foreign>
- Cite “nnet”: Venables, W. N. & Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

- Cite “ggplot2” : H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- Cite “reshape2” :Hadley Wickham (2007). Reshaping Data with the reshape Package. *Journal of Statistical Software*, 21(12), 1-20. URL <http://www.jstatsoft.org/v21/i12/>.
- Cite “lme4”: Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.
- Cite “kableExtra”: Hao Zhu (2020). kableExtra: Construct Complex Table with ‘kable’ and Pipe Syntax. R package version 1.2.1. <https://CRAN.R-project.org/package=kableExtra>
- Cite “knitr”: Yihui Xie (2020). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.27.
- Cite “stringr”: Hadley Wickham (2019). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0. <https://CRAN.R-project.org/package=stringr>
- Cite “devtools”: Hadley Wickham, Jim Hester and Winston Chang (2020). devtools: Tools to Make Developing R Packages Easier. R package version 2.3.2. <https://CRAN.R-project.org/package=devtools>
- Cite “jtools”: Long JA (2020). jtools: Analysis and Presentation of Social Scientific Data. R package version 2.1.0, <URL: <https://cran.r-project.org/package=jtools>>.
- Cite “ broom.mixed” :Ben Bolker and David Robinson (2020). broom.mixed: Tidying Methods for Mixed Models. R package version 0.2.6. <https://CRAN.R-project.org/package=broom.mixed>
- Cite “table1”: Benjamin Rich (2020). table1: Tables of Descriptive Statistics in HTML. R package version 1.2.1. <https://CRAN.R-project.org/package=table1>
- Cite “pROC”: Xavier Robin, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frederique Lisacek, Jean-Charles Sanchez and Markus Muller (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12, p. 77. DOI: 10.1186/1471-2105-12-77 <http://www.biomedcentral.com/1471-2105/12/77/>
- Cite “rvest”: Hadley Wickham (2020). rvest: Easily Harvest (Scrape) Web Pages. R package version 0.3.6. <https://CRAN.R-project.org/package=rvest>

Appendix

1. Some assumptions for education:

- Assume Some secondary/ high school and Completed secondary/ high school as Secondary (high) school diploma or equivalency certificate (2016 counts)
- Assume Some technical, community college, CEGEP, College Classique and Completed technical, community college, CEGEP, College Classique as College, CEGEP or other non-university certificate or diploma (2016 counts)
- Assume Some university, Bachelor's degree, Master's degree and Professional degree or doctorate as University certificate or diploma below bachelor level (2016 counts)

2. List of Canadian federal electoral districts:

- https://en.wikipedia.org/wiki/List_of_Canadian_federal_electoral_districts

3. ROC curve and AUC:

Area under the curve: 0.6174

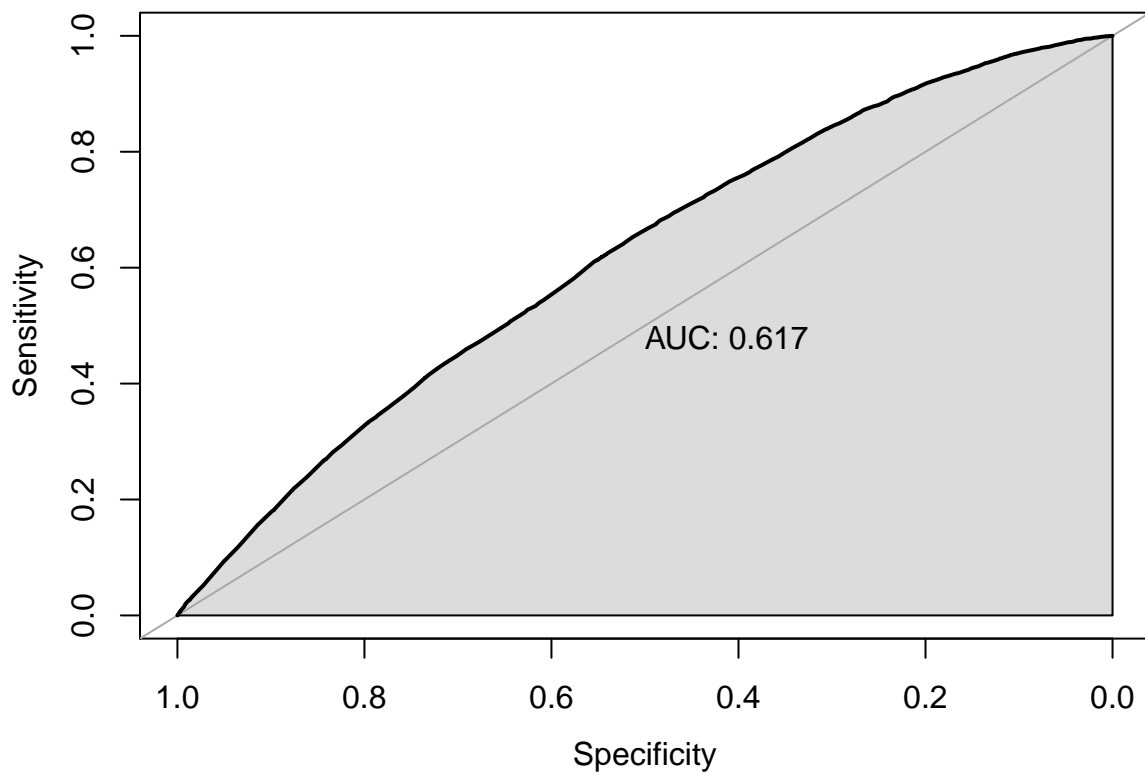


Figure 2: AUC of the Chosen Model