Election Poll for the Liberal Party Canada: the Aurora-Oak Ridges-Richmond Hill Electoral District

Xi Cheng, Shichao Feng and Zhitong Liu (Group 61)

October 6, 2020

Contents

1 Executive Summary

The Liberal Party only won 2019 federal election by a small margin and lost two important things compared to its 2015 winning: the popular vote and the majority government. To fully understand the changes in the last two elections, closely track most up-to-date support rate and win the next election for the Liberals, the famous swing riding Aurora-Oak Ridges-Richmond Hill in Toronto region was selected. It will serve as a good indicator for the next election. Online survey using random sampling was designed to target the whole population 18 years of age or older. The survey questions focusing on demographics, changes in the last two elections, economy policies, immigrants policies and trust in leaders were carefully designed. These issues are most cared ones which were claimed by local political leaders. By monthly polling updates, one can accurately estimate the overall opinions on those issues in this riding. Through a preliminary simulation study, we discovered the local demographic features such as respondents gender distribution could play an important role. Meanwhile, we also assured that the crossing the floor of Leona Alleslev, Member of Parliament, did damage the support rate of the Liberals decisively. In the future, we will definitely keep update results, questions and the data simulation methods as we are collecting more survey responses.

2 Introduction

In 2019, the Liberal Party which led by Justin Trudeau, won just 157 seats, compared to 184 seats in 2015. Although the Liberal Party was allowed to form a minority government, both the popular vote and the majority government were lost. Actually, the popular vote decreased by 6.35% to 33.12% in 2019 compared with 2015 (Zimonjic, 2019).
Toronto has a total of 25 Members of Parliament (MPs). Traditionally, Toronto has always been an vital source of support for the Liberal Party and the Liberals usually dominates the inner portion of the city. In the past two federal elections, the dominance of the Liberals was shaken in three electoral districts, Thornhill, Markham-Unionville and Aurora-Oak Ridges-Richmond Hill (AORRH). Out of those three ridings, the AORRH is the swing riding, where the Liberals was swept by the Conservatives in 2011, then won the next one by just 2.1% and eventually lost the last one by just 2.5% to the Conservatives again (CBC News, 2019). Interestingly, Leona Alleslev who was elected twice as the MP representing the riding of AORRH, crossed the floor to join the Conservative Party from the Liberal Party on September 17, 2018. The results of the last two election indicates that the Liberals has gained support from a significant amount of residents in this district. It is very likely that the Liberals might dominate this riding as well, through a better understanding of the resident profiles and a more accurate targeting.
Therefore, our campany, Petit Poll, has designed a online survey to target the 90,000 adult residents in the AORRH riding (Statistics Canada, 2016). The questions were developed to obtain basic demographics information of respondents, discover details during changes in votes of the last two elections, track the impact of the economy policies and the opinions on immigrants/refugees. Additionally, the impact of COVID-19 was also taken into considerations. In general, this is a probability sampling survey with a cross-section design. Moreover, monthly updates of this survey can build up a time-series analysis to track the opinions of the respondents on the Liberals. With this survey design, we were able to conduct preliminary simulation study utilizing random sampling and empirical probability distributions. The preliminary study showed that certain demographic features such as gender may contribute to the swing of this riding. Moreover, it was also shown that the lost of support rate in the Liberals as the crossing the floor of Leona Alleslev. All code will be hosted at https://github.com/Chelsea-Cheng99/STA304/tree/master/ProbSet2 to guarantee reproducibility.

3 Methodology

3.1 Survey methodology

This is a sample survey with a cross-sectional design. The target population for this survey is AORRH residents 18 years of age or older, roughly 90,000. The sampling frame used in this survey is the full list of registered address in this riding. The sample in this study is the respondent in this address. Only one respondent is allowed in an address within the same survey. This sampling frame provides a good probability based coverage of the population.
Incentives such as grocery store gift cards will be used to reduce non-response. Particularly, a letter with a five dollar gift card will be mailed to the randomly selected address (simple random sampling without replacement). The respondent will be directed to the online survey with the unique user name and password mailed in the letter. By completing the survey, the respondent will get another five dollar gift card. Hence, we will use probability sampling, which is any method of sampling that utilizes some form of random selection. To protect the privacy, no identifiable personal information will be recorded during the survey. The sample size (completed surveys) will be about 200 to get a good coverage of the population.
With simple random sampling without replacement, we will have good estimations of population parameters:

$$E(\bar{y}) = \mu_y$$

$$Var(\bar{y}) = (1 - n/N)(s_y^2/n)$$

The confidence interval for μ_y:

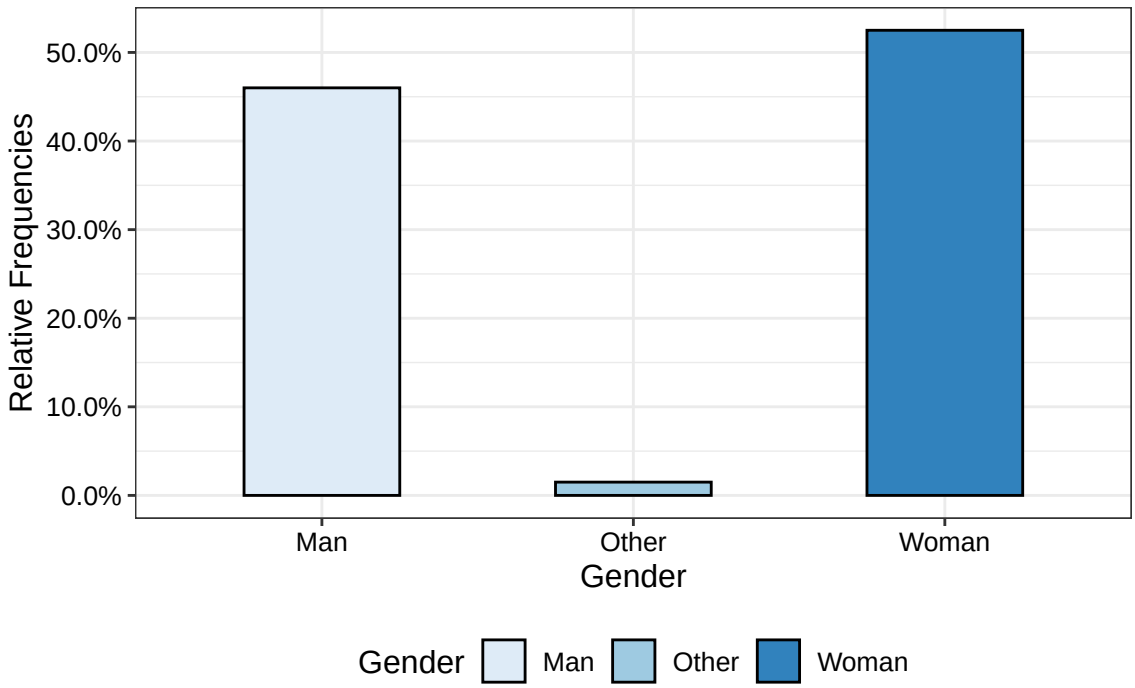
$$\bar{y} \pm z_{\alpha/2}\sqrt{(1 - n/N)(s_y^2/n)}$$

Where n is the sample size, N is the population size, $s_y^2 = \frac{1}{(n-1)}\Sigma_{i \in s}(y_i - \bar{y})^2$ and the significant level α is 0.05.
The questionnaire is relatively short and all questions are required to answer. All questions are provided with clear options, including the "other/ prefer not to say" option. The missing values can also replaced with imputation values such as sample mean during the post-survey analysis stage (Toepoel & Schonlau, 2017).
The total cost to conduct the first survey will be about 50,000 dollars, including background investigate, survey design and delivery, incentives, response collection, data analysis and report. Following survey will be 20,000 dollars each.

3.2 Simulation

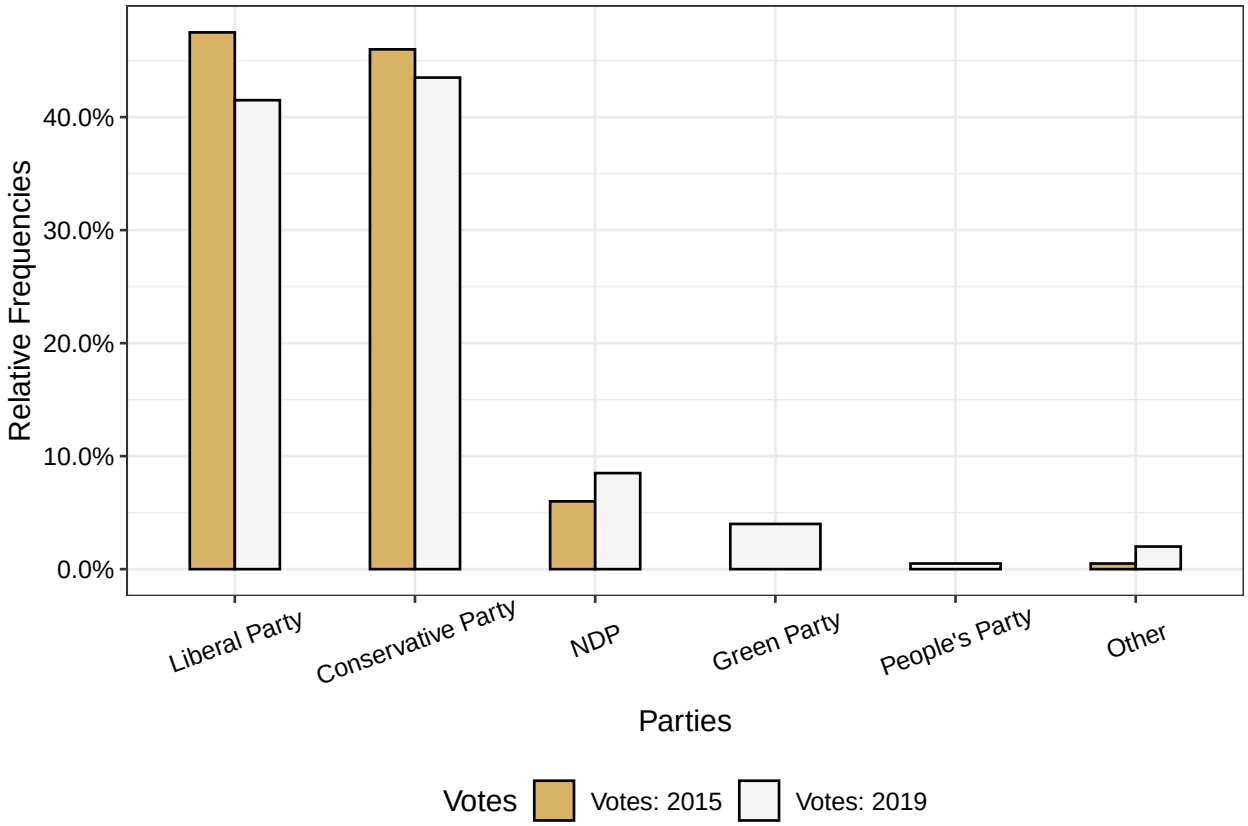
To test the performance of the survey, probability sampling methods were applied to generate the simulated survey results. Common probability sampling methods include simple random sampling, stratified random sampling, cluster sampling and systematic sampling. Simple random sampling technique was mainly used in this study. The results were then demonstrated descriptively. In terms of the simulation, several assumptions were made according to reference datasets: Census 2016 data (Statistics Canada, 2016) and 2019 Canadian Election Study (Stephenson et al., 2019). In general, the underlying distributions or probability mass functions were assumed for different variables. Meanwhile, independence was assumed between questions. The details will be found in the appendix.
The survey was designed to gain knowledge on five aspects: demographics, how did this riding swing, trust in leaders, impact of economy policies and immigrants policies during COVID-19. With the simulated data, we will focus on the demographics and the changes between the last two elections. Because the underlying probability distribution assumptions are closer to the real-world scenario in these two cases.
All work were done in R (version 4.0.2) (R Core Team 2020) and Rstudio (version 1.3.1093). Tidyverse (version 1.3.0) was used for data wrangling and visualization (Wickham et al., 2019). R package "forcats" (version 0.5.0) was also used for data pre-processing (Hadley Wickham, 2020).

4 Results



The simulated gender distribution data using the empirical probability from the 2019 Canadian Election Study

(Stephenson et al., 2019), which is 41% of man and 58% woman in that study. The data was generated as randomly sampling with replacement of the gender code with those probabilities. The Figure 1 shows that more than half of the respondents are woman. If this pattern could be consistent along monthly polling updates, it would indicate that, in this riding, women are more likely to participate in political activities, by 6.5% compared to men. Therefore, targeting the women group might be an efficient campaign strategy.



The simulated votes distribution data using the empirical probability from the 2019 and 2015 federal election results (CBC News 2015, 2019). The data was generated as randomly sampling with replacement of the party code with those probabilities. The detailed information can be found in the **Appendix**. The Figure 2 was generated to demonstrate, in AORRH riding, the lost in support rate of the Liberals in 2019 as Leona Alleslev crossed the floor to join the Conservative Party from the Liberal Party on September 17, 2018. The change in support rate might seem marginally, but it is significant enought to be decisive. Compared with the 2015 federal election, the support rate of the Liberals decreased by 6%, leading to the defeats in this riding.

5 Discussions

We believe that the swing riding is the most sensitive indicator of the federal election. Therefore, the Aurora-Oak Ridges-Richmond Hill riding was selected as the population of the survey. Due to the limits of the platform, we were only able to publish the first 10 questions in our survey. Questions on economy policies, immigrants and COVID-19 had to be removed. The questions were designed to focus on aspects about which the MP, Leona Alleslev most cares. Since we did not really have the data about the empirical distribution of answers to most of the survey questions, we simply used the random sample with replacement method with equal probability assumption. In future work, we will refine those empirical distributions by most up-to-date polling results. In terms of the underlying reasons of the swing, more sophisticated descriptive figures and tables can be generated to show how many individuals voted for the same party, how many voted for the Liberals then changed to the Conservatives as Leona Alleslev. In that case, it is a pre-post time series study. Because we assumed independence answers between questions, time-series analysis is not applicable in our

case.

6 Appendix

6.1 The survey

The first 10 questions can be found at https://chelsea942995612.typeform.com/to/yCpsK2tR.

The survey was designed to gain knowledge on five aspects: demographics, how did this riding swing, trust in leaders, impact of economy policies and immigrants policies during COVID-19.

6.2 The simulated results

The following code was used to generate the simulated data. All the code to repeat this work can be found at https://github.com/Chelsea-Cheng99/STA304/tree/master/ProbSet2.

```
# random seed to make reproducible
set.seed(666)
# simulate data
df <- tibble(Age = round(runif(n, min = 18, max = 98), 0), Gender = sample(c("Man",
    "Woman", "Other"), n, replace = TRUE, prob = c(0.41, 0.58,
    0.01)), Education = sample(c("Elementary", "High", "College",
    "University", "Master", "Doctorate", "Other"), n, replace = TRUE,
    prob = c(0.01, 0.2, 0.31, 0.34, 0.08, 0.04, 0.02)), Income = sample(c("No",
    "1 - 30000", "30001 - 60000", "60001 - 90000", "90001 - 110000",
    "110001 - 150000", "150001 - 200000", "200001 -", "Other"),
    n, replace = TRUE, prob = c(0.017, 0.157, 0.209, 0.159, 0.084,
        0.075, 0.038, 0.024, 0.236)), Votes_2015 = sample(c("Liberal Party",
    "Conservative Party", "NDP", "Green Party", "People's Party",
    "Other"), n, replace = TRUE, prob = c(0.47, 0.45, 0.06, 0.01,
    0, 0.01)), Votes_2019 = sample(c("Liberal Party", "Conservative Party",
    "NDP", "Green Party", "People's Party", "Other"), n, replace = TRUE,
    prob = c(0.422, 0.447, 0.071, 0.04, 0.01, 0.01)), MP = sample(c("Strongly agree",
    "Agree", "Neutral", "Disagree", "Strongly disagree", "Other"),
    n, replace = TRUE), JT = sample(c("Very Satisfied", "Satisfied",
    "Neutral", "Not Satisfied", "Strongly Not Satisfied"), n,
    replace = TRUE), Tax = sample(c("Very Satisfied", "Satisfied",
    "Neutral", "Not Satisfied", "Strongly Not Satisfied"), n,
    replace = TRUE), Econ = sample(c("Better", "Same", "Worse",
    "Other"), n, replace = TRUE), Finance = sample(c("Better",
    "Same", "Worse", "Other"), n, replace = TRUE), ImmiJobs = sample(c("Strongly agree",
    "agree", "Neutral", "disagree", "Strongly disagree"), n,
    replace = TRUE), Immi = sample(c("More", "Same", "Less",
    "Other"), n, replace = TRUE), Refu = sample(c("More", "Same",
    "Less", "Other"), n, replace = TRUE), COVID = sample(c("Very good",
    "good", "Neutral", "bad", "Very bad"), n, replace = TRUE),
    Prom = sample(c("Most of the time", "Some of the time", "Hardly ever",
        "Never", "Other"), n, replace = TRUE))
```

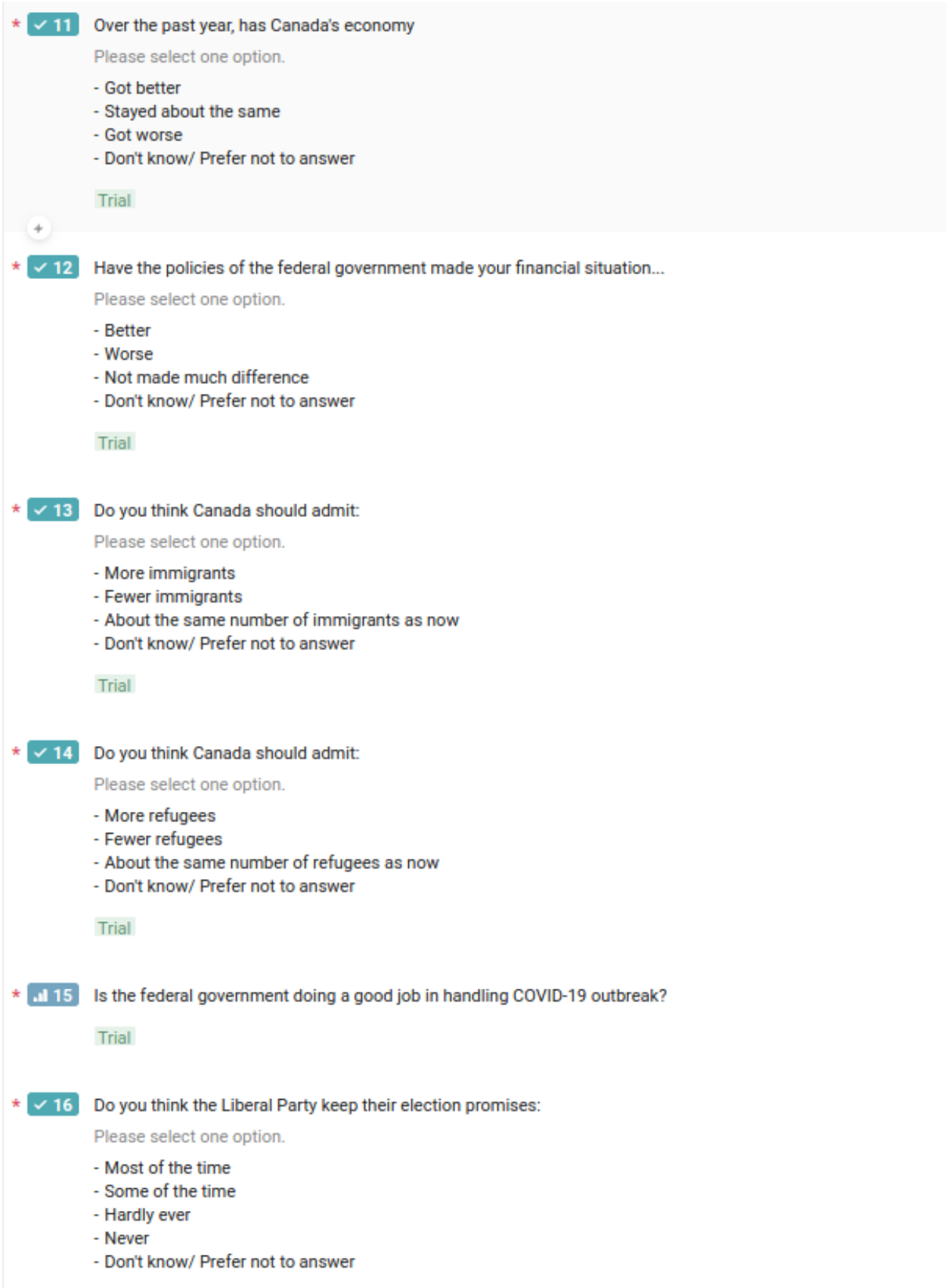


Figure 1: More survey questions which cannot be published on Typeform.com

6.3 References

CBC News. "Canada Votes 2019". https://newsinteractives.cbc.ca/elections/federal/2019/results/. 2019.

CBC News. "Canada Votes 2015". https://www.cbc.ca/news2/interactives/results-2015/. 2015.

R Core Team. 2020. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Statistics Canada. "Census Profile, 2016 Census". https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=FED&Code1=35003&Geo2=PR&Code2=35&SearchText=aurora&SearchType=Begins&SearchPR=01&B1=All&TABID=1&type=0. 2016.

Stephenson, Laura, Allison Harrel, Daniel Rubenson and Peter Loewen. Forthcoming. 'Measuring Preferences and Behaviour in the 2019 Canadian Election Study,' Canadian Journal of Political Science.

Vera Toepoel & Matthias Schonlau. Dealing with nonresponse: Strategies to increase participation and methods for postsurvey adjustments, Mathematical Population Studies, 24:2, 79-83, 2017.

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

Hadley Wickham (2020). forcats: Tools for Working with Categorical Variables (Factors). R package version 0.5.0. https://CRAN.R-project.org/package=forcats

Zimonjic, Peter. "Liberals take losses but win enough in Quebec and Ontario to form minority government". Canadian Broadcasting Corporation. Retrieved October 21, 2019.