

Will the Liberal Party Win the Next General Election in Toronto Centre?

Yu Yuan

October 7, 2020

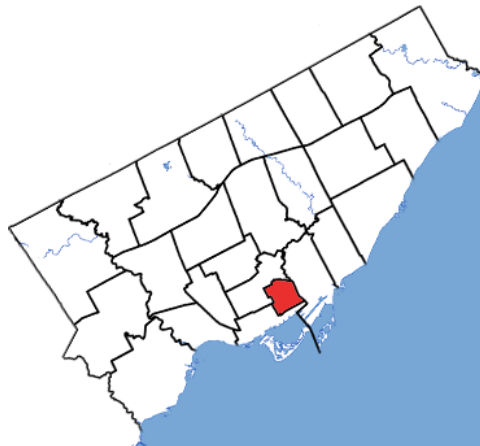
Abstract

Political polling surveys are prevailing and conducive for making campaign strategy. In this report, we apply the method of simple random sampling without replacement (SRSWOR) and conduct a straightforward online survey for Canadian citizens living in the electoral district of Toronto Centre. We find that the proportion of registration in older people is high than the young generation, and the Liberal Party tend to keep this seat of Federal parliament in the next general election.

Introduction

The fundamental basis of Canadian election is a Federal parliament system, including the sovereign, an upper house and a lower house, which is original from the United Kingdom. This democratic legislative system stipulates that the candidate with the most votes in each electoral district, more commonly referred as a riding, will win that constituency and become a member of parliament representing that riding. There are 338 ridings in Canada, because the population of various ridings is different, for instance, a densely populated riding could have more than 100,000 voters while a small riding might only have 5,000 voters. Consequently, the key for winning the general election as the Prime Minister of Canada is number of ridings rather than the overall percentage of votes.

Figure 1. Toronto Centre in relation to other Toronto ridings (2015 boundaries) [1]



As a Canadian polling company, we focus on a specific federal riding called Toronto Centre that covers the heart of Downtown Toronto. From 2015 to 2020, the riding of Toronto Centre was represented by Bill Morneau, a former finance minister from the Liberal Party, as member of Parliament, he recently resigned because of unpleasant rumors and scandals. For last decades, most of the votes of the Liberal Party rely on the strong support of voters in major cities such as Toronto, Vancouver and Montreal, but it is difficult to control the remote and poor areas of the Atlantic coast, where most of the voters support the Conservative Party. Keep tracking and analysing the polling trend in Toronto Centre is significant for the Liberal Party, considering its political and economic roles in the general election.

In this report, we initially introduce our survey methodology, such as target population and data collection. Then we apply R to summarize and visualize the basic statistical features of our data. The polling evidence shows that the chance of the Liberal Party to win the next general election is promisingly greater than others'.

Link for our online survey: <https://www.surveymonkey.com/r/3896BZH>

GitHub repo for R programming and Data file: <https://github.com/cristalyu/Political-Polling-in-Toronto-Centre->

Methodology

Target population, sample size and sample frame

A Canadian citizen is qualified to vote if he or she is at least 18 years old when registering to vote, or will be 18 years old on the referendum day. Therefore, our targeted population for polling survey is Canadian citizens who have been resident in the area of Toronto Centre at least for six months and older than 18 years old. According to the Ontario Past Election Results [2], the average of total valid votes for last three general elections in the riding of Toronto Centre is approximately 45,000. For the determination of sample size, we utilize the following estimation process.

Suppose \hat{p} denotes the proportion of Toronto Centre population who supports a particular candidate. Then, \hat{p} is approximately normal distribution with $\mu = n\hat{p}$ and $\sigma^2 = \hat{p}(1 - \hat{p})/n$, if our sample size n is sufficiently large. Then the 95% confidence interval for the estimated support proportion is

$$\left(\hat{p} - Z_{0.05/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \quad \hat{p} + Z_{0.05/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}\right)$$

where $Z_{0.05/2}$ is a score from the standard normal distribution. For a 95% confidence level, $Z_{0.05/2} = 1.96$. Suppose that we would like to control the width of this confidence interval within W , that is,

$$2 \times Z_{0.05/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = W$$

By solving n , we have

$$n = \frac{4Z^2\hat{p}(1 - \hat{p})}{W^2}$$

Since the maximum value for $\hat{p}(1 - \hat{p})$ is 0.25 when $\hat{p} = 0.5$, and the scenario of picking a candidate to vote is generally a 50/50 case in reality. Then the theoretical sample size we need for survey is

$$n = \frac{4 \times 1.96^2 \times 0.5 \times 0.5}{0.02^2} = 9604$$

if we control the width of the 95% confidence interval within 2%.

For the sampling frame, we prefer to use the online method of survey on social media, rather than transitional distribution by telephone frame. Although the mode of telephone interview is the most common, it might not be the most cost-effective way to collect data since few people actually answering calls from unfamiliar numbers, which would diminish the effectiveness in the digital era.

Sampling method

The sampling method we apply is simple random sampling without replacement (SRSWOR). Basically, it is a method of randomly selecting n unites out of the N unites one by one such that at any stage of selection, any one of the remaining units have the same probability of being selected, i.e., $1/N$. Unlike the method of simple random sampling with replacement (SRSWR), once an element is selected as a sample unit, will not be replaced in the population pool by SRSWOR [3]. Suppose n unities are selected by SRSWOR, the total number of possible samples are $C_N^n = \binom{N}{n}$, so the probability of selecting any one of these samples is $\frac{1}{C_N^n} = \frac{1}{\binom{N}{n}}$.

Reach respondents

As it has been mentioned above, the basic sampling frame is based on online method of survey distribution because the rate of respondents by telephone surveys might be unpleasant since few people actually answering calls from unfamiliar numbers. Changes in technology and consumer behavior that allow us for faster, better results and more opportunities to reach our target targeted population, for instance, a random device engagement inside mobile apps, where consumers are engaged and can be easily incentivized non-monetarily; survey links on social media, like Facebook, Instagram, Twitter, etc., could be relatively inexpensive to target down to our sampling population. For improving the rate of online respondents, an appealing form of questionnaire and reasonable credits for accomplishment are inessential for respondents with high quality.

Invalid respondents and non-respondents

Generally, there are four types of inattentive survey respondents: speeders, straightliners, slackers and survey bots [4]. For identifying these invalid results, we could mark all respondents with duplicate IP address and select a threshold for an impossibly fast response time on common sense (e.g. at least 1 second per question). For non-respondents, normally referred as missing values, we could conduct a process of data cleaning. For instance, the

method of mean substitution is a method that the missing data value is replaced by the mean value of a variable that for same variable [5], however, the mean substitution method may lead to biased estimation because the distribution of missing value is not random. Alternatively, we can consider to adopt the direct method of deletion while it may also result in a bias when the percentage of missing value is unreasonable great. Nevertheless, King et al. stated that the percentage of incomplete records in the political sciences exceeded 50% on average, with some studies having over 90% incomplete records [6].

Survey Cost and Privacy

According to the Google Survey Help, the average price for a 2-10 questions survey is approximately 3.96 CAD. Our survey only contains 5 straightforward questions, then the total cost for survey will be 19,000CAD, if our sample size is exactly following the number we computed above. To protect the privacy of respondents, our online survey will maintain anonymity and confidentiality, and design without any pressure on individuals to participant.

Survey Results and Discussion

Our online survey has a simple layout consisting of 5 straightforward questions (Attached in Appendix). With accessible R resource, we having following numerical and graphic outcomes from 10,000 valid responding data (Simulated).

First of all, we notice that the most frequent surveyed respondents are in the age range of 18-30 years old, taking 34.90% of the total respondents. This scenario is consistent with the reality considering that the older people are not familiar and convenient with the procedure of online survey. Additionally, we observe that there is no obvious difference of gender within the various ranges of age, in Figure 2. The percentage of voting registration is somehow unpleasant, since only 34.30% respondents have been registered, 51.80% surveyed individuals with non-registration claim, and remaining 13.90% people are hesitating for voting. From the Figure 3, we find out that the proportion of registration is lower in young generations (18-30 years old) than the ones in older people groups, however, the number of registered young generations is still the largest due to the large sampling base in that group.

Table 1. Frequency Table of Age

Age	Count	Proportion
<18	501	5.01%
18-30	3,487	34.90%
31-49	2,958	29.60%
50-69	2,542	25.40%
70 or older	512	5.12%

For evaluating the current performance by the Liberal Party, the sampling respondents provided left-skewed distribution of rating, indicating that people in Toronto Centre are living in harmony with the Liberal Party. In fact, most of the surveyed individuals comment on the performance as “Neither satisfied nor dissatisfied” and “satisfied”, the average score of performance is 3.26 (standard deviation = 1.26). For approval rating in register people, the polling results in both Table 2 and Figure 5 show that the Liberal Party is leading the Conservative Party by approximately 20 percentage points. Therefore, the Liberal Party is more likely to keep its seat in Federal parliament in next generation.

Figure 2. The distribution of Age groups on Gender

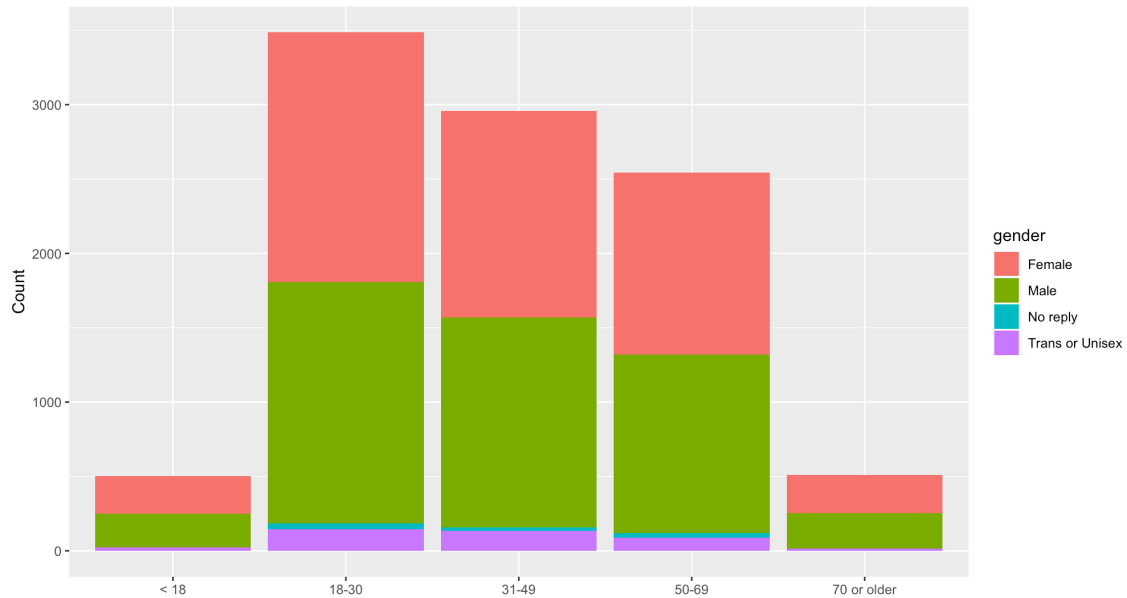


Figure 3. The distribution of Age groups on Registration status

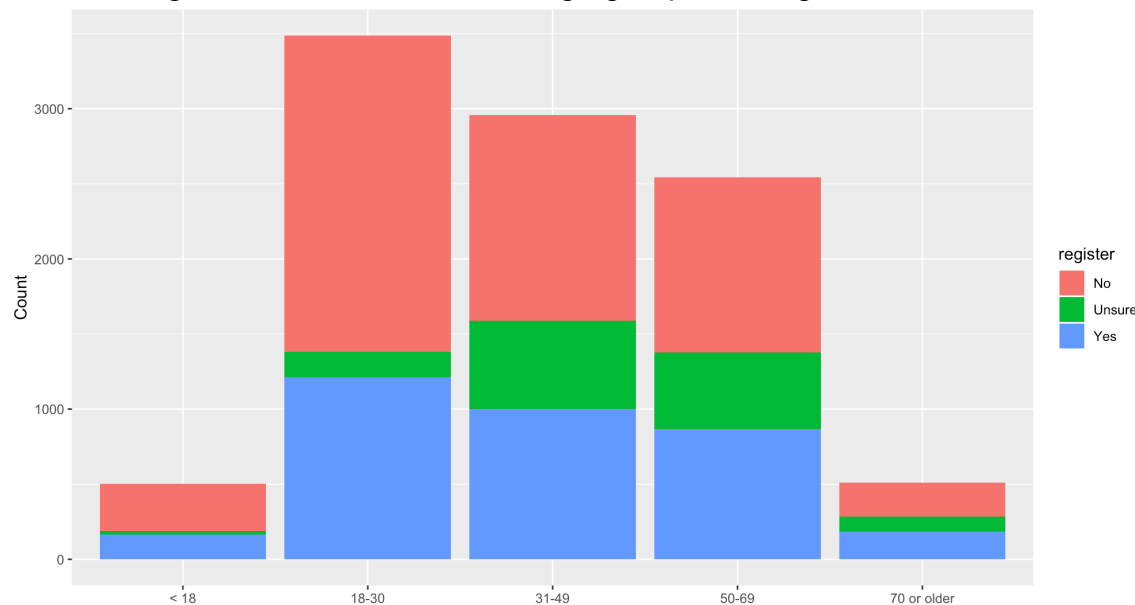


Figure 4. The distribution of Rate on the Liberal Party's performance

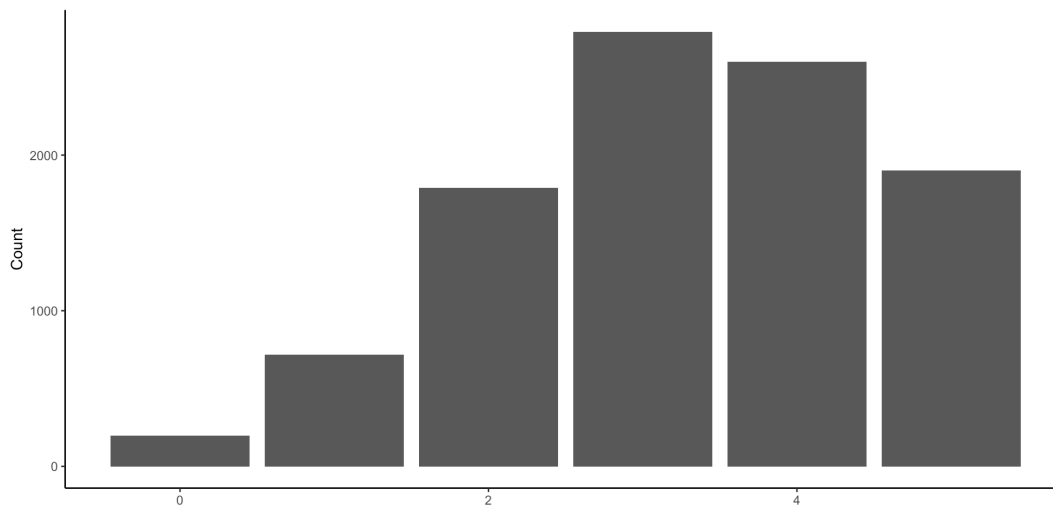


Table 2. The frequency table of Party preference in registered people

Party	Count	Proportion
Liberal	1,577	46.00%
Conservative	860	25.10%
NDP	341	9.95%
Green	172	5.02%
Other	304	8.87%
Unsure	174	5.08%

Weakness and future work

The sample size we select is based on a 95% confidence interval controlled by the marginal error equals to 1%, the sampling size is 9,604, we should be doubt about that whether this number of sampling is overwhelmingly large or just sufficient, regarding the total number of valid voters in Toronto Centre is 45,000 averagely. The cost and validation of our respondents also need to be thoroughly considered and test, due to the limited campaign fund. Actually, the riding of Toronto Centre is leading by the Liberal Party in the last decades, whether we should cut down our budget on polling survey in this electoral district need to another comprehensive assessment. For the future work, we might focus on the relationship between ratings of the Liberal Party and the willingness of voting for it.

Reference

- [1] By Ajbutler - Own work, CC0, <https://commons.wikimedia.org/w/index.php?curid=42759643>.
- [2] M Trivedi, Unit 2 simple random sampling - eGyanKosh, 2017, <http://egyankosh.ac.in/bitstream/123456789/20693/1/Unit-2.pdf>.
- [3] Ontario Past Election Results, <https://www.elections.on.ca/en/resource-centre/elections-results.html>.
- [4] Aaron Moss, PhD, How to Identify and Handle Invalid Responses to Online Surveys, <https://www.cloudresearch.com/resources/guides/ultimate-guide-to-survey-data-quality/how-to-identify-handle-invalid-survey-responses/>.
- [5] Kang H. 2003, The prevention and handling of the missing data. Korean journal of anesthesiology, 64(5), 402–406. <https://doi.org/10.4097/kjae.2013.64.5.402>
- [6] King, G., J. Honaker, A. Joseph, and K. Scheve. 2001. Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation. American Political Science Review 95 (1): 49–69.
- [7] Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, <https://doi.org/10.21105/joss.01686>

Appendix

Survey link: <https://www.surveymonkey.com/r/3896BZH>

Online Political Polling in Toronto Centre

1. What is your gender?

- ☐ Female
- ☐ Male
- ☐ Transgender or Unisex
- ☐ Prefer not to answer

2. What is your age?

- ☐ < 18
- ☐ 18 to 30
- ☐ 31 to 49
- ☐ 50 to 69
- ☐ 70 or older

3. Are you a registered voter?

- ☐ Yes
- ☐ No
- ☐ Unsure / Don't know

4. How would you comment on the current performance of the Liberal Party in Toronto Centre?

Very dissatisfied	Dissatisfied	Neither satisfied nor dissatisfied	Satisfied	Very satisfied
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

5. Which of the following party are you planning to vote in the next general election?

- | | |
|---------------------------------------------------|----------------------------------------------|
| <input type="checkbox"/> The Liberal Party | <input type="checkbox"/> The Green Party |
| <input type="checkbox"/> The Conservative Party | <input type="checkbox"/> Other Parties |
| <input type="checkbox"/> The New Democratic Party | <input type="checkbox"/> Unsure / Don't know |