High Poll Engagement but sharply Declined in April 2020*

Hong Pan

29 January 2021

Abstract

The line chart and the table created help illustrate that the poll response rate has started to decline in 2019. Declining poll response rates may cause non-response bias, which misunderstanding the real feeling and thinking of people. Two bar charts created help show largest people have a question about Front Yard Parking. Let the government know where people are dissatisfied with the city.

Contents

1 Introduction

Low poll engagement may cause non-response bias, which would misunderstanding the real feeling and thinking of people.(*Assessing the Representativeness of Public Opinion Surveys* 2019) Public polls belong to human research surveys to help us know how a population thinks about any specific topic. The trained interviewers will ask questions to people chosen randomly from the population being measured. The interviewees' answers will be interpreted and help to make a final decision on the application applied.(*Opinion Poll* 2021) The main benefit for people from the public polls is that it gives regular people a chance to be heard instead of letting only media stars speak on behalf of all.(*What Is Public Opinion Polling and Why Is It Important?* 2007)

2 need to edit

After knowing the definition and importance of public polls, I want to look at how people's positively toward public polls by creating a variable called engagement rate in this paper. I construct a line chart to show the average engagement rate changes from 2015 to 2020. The graph shows that the engagement rate slightly rises from 2015 to 2019, but a sharp decrease from 2019 to 2020. Then, I create a table to illustrate the average engagement rate for each type of application, and I find that Permit Parking has the highest average

---

*Code and data are available at: https://github.com/HongTrista/folder.

engagement rate. To know more information in detail, I draw two histograms to display how many applicants for each type and the number of results, whether the application has passed. The graphs show that most people have a problem with Front Yard Parking, but it has the highest pass rate.

The remainder of the paper is structured as follows: Section 2 discusses the variables in the data set, Section 3 presents the figures and tables with their explanations,and finally Section 4 provides all the references.

3 Data

The analysis for this paper done using R (R Core Team 2020)and several packages, which listed as followed: 'ggplot2' helps to visualize the data in figures(Wickham 2016), 'tidyverse' used for data manipulation(Wickham et al. 2019), 'readr' used for import data set from URL by R(Wickham and Hester 2020), 'bookdown' used for cross-referencing figures and tables(Xie 2020),'ggpubr' used for combining graphs in one figure (Kassambara 2020), and 'zoo' used to edit the time variable.(Zeileis and Grothendieck 2005) The data set, Polls Conducted by the City, I used for analysis in the paper is from the City of Toronto Open Data Portal.(Office 2021)

The main goal of creating the dataset is to analyze the poll engagement in Toronto. The dataset is free to the public and starts on April 1, 2015. When the City Clerk's Office receives an application, a poll will distribute to all the people who may be influenced in the region mentioned in the application. People have the right to participate in the poll or not, the number of people who do not mail the poll back, the address is wrong, and any information will be recorded in the dataset. Only the application which meets the required response rate will be processed to the next step. The data will be updated after the polls have been closed and after the certification has been made. As we mentioned before, a low response rate may cause non-response bias, which means the real feeling and thinking of a specific event will be misunderstood. For example, the article analysis of health problem with non-response bias pointed out that non-response was determined to contribute to underestimated health risks. The strength of the data set I used in the paper is that it set the minimum response rate for each application, which may prevent non-response bias to some extent.(Turk 2019)

The data set after cleaning contains 6 variables with 922 conversations and no missing values. The explanations of each variable will be shown in the following:
. Application_topic: Type of each application, which is a categorical variable with levels of "Front Yard Parking", "Traffic Calming," "Boulevard Cafe", "Business Improvement Area," and "Commercial Boulevard Parking."
. Response_rate_met: Whether the number of ballots returned has met the required response rate, a categorical variable with values 'YES' and 'NO.'
. Real_total_number: The number of total voters on the final poll list, which is a numerical variable.
. Potential_total_number: The number of people residing within the poll boundary range, which is a numerical variable.
. Date: The date of close, a time categorical variable from June 5, 2015, to Dec 30, 2020.
. Final_result: The final result of the poll, which is a categorical variable with levels of "In favour', ' opposed,' and ' response rate not met.'

In order to analyze the real response rate for each application topic, I create a new variable, response_rate, based on the Real_total_number and Potential_total_number variables in the clean dataset.: . Response_rate: Real_total_number divided by Potential_total_number

Table 1 shows the total number of application for each different topic, we can see that from year 2015 to 2020, the number of application about the problem with Front Yard Parking occupied half of the total application number, which is 510 in 922. We can deduce that most people have the problems about the issue of front yard parking. The Figure 1 illustrate a line chart shows the trend of the total application number from year 2015 to 2020 in each quarter. The line chart shows a downward slope, which means all topics the city

organization cares about are gradually getting better because fewer and fewer application numbers are being received. This is a great sign for the city organization.

Table 1: Total number of applications per topic

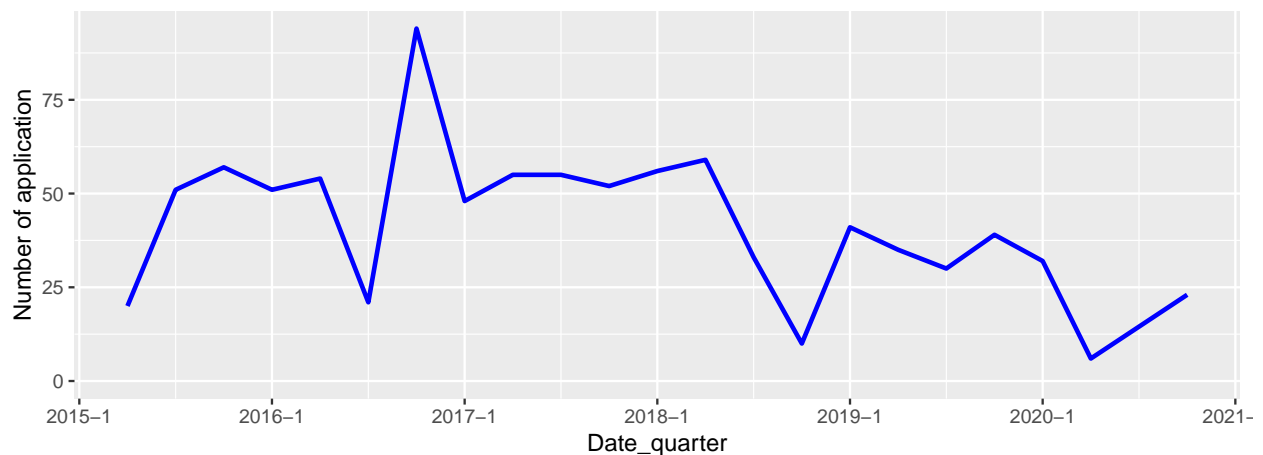| Application_topic	Number
Boulevard Cafe	113
Business Improvement Area	12
Commercial Boulevard Parking	13
Front Yard Parking	510
Permit Parking	73
Traffic Calming	201

Figure 1: Total number of application in each quarter

The time range for the data in the dataset is from 2015 to 2020 last day. To know how people's attitudes towards the public polls, I draw a line chart to show the response rate trend, how many people actually engaged in the City's polls. Because according to each application, there will be a list of people selected to participate in the public polls, however, some people will refuse to participate. I want to indirectly understand people's attitudes towards the public polls by knowing the response rate.

The line chart Figure @ref(fig: Figure2) shows that the response rate generally fluctuated around 0.8 from 2015 to April 2020. The relatively high engagement rate illustrates that people actively participate in the City's public polls in the past six years. The line sharply declined in April 2020 from 0.88 to 0.5 in October and bounced back to normal in November. As we all known, the covid-19 broke out in March in Toronto, which may be the reason influence the engagement sharply decline in April, and because the response rate quickly returned to normal levels in November. From this, it can be inferred that the drastic decline was not due to changes in people's attitudes towards polls, but more data is needed to support this conjecture.
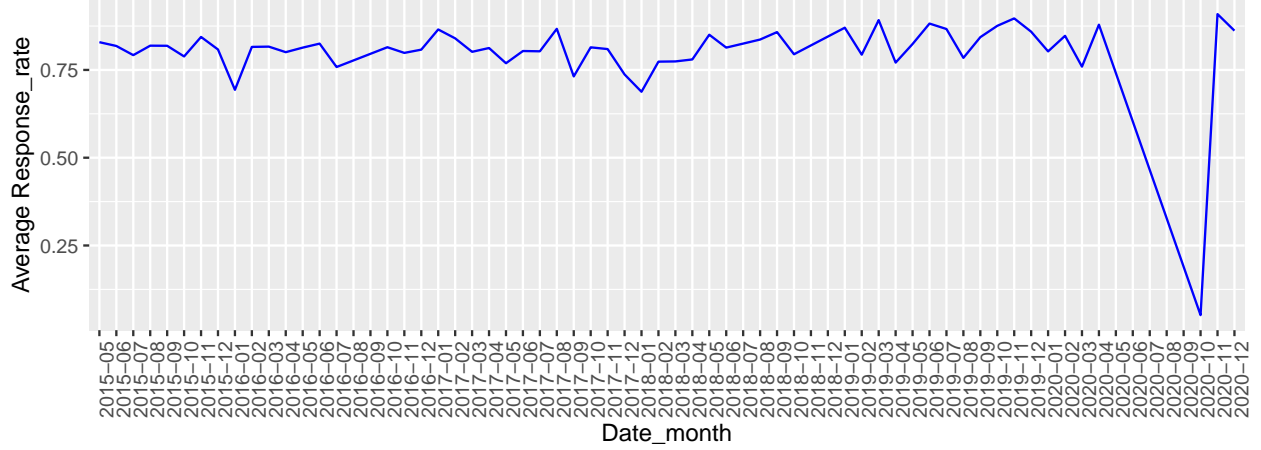
Figure 2: Average response rate per month

To know more about data, the graphs about different topics with whether the required response rate has been met or not, and with the final result situation for each topic been analyzed. From the left graph in Figure @ref(fig: Figure3), Traffic Calming has the highest number of applications that do not meet the required response rate among all types, which is around half of the total applications in Traffic Calming topic. Meanwhile, Front Yard Parking has the highest application number that meets the required response rate, which is around 95% of the total applications on the topic of Front Yard Parking.

The applications which do not meet the required response rate will be given the final result of "No", then, the graph about the final result for the applications which meet the required response rate has been plotted on the left in Figure @ref(fig: Figure3). The graph illustrates that Front Yard Parking has the highest pass rate by showing the highest proportion of result of "in favour," then, followed by Traffic Calming. Combined the information we explored out in Table 1, which shows that almost half of the applications are about Front Yard Parking, we can infer that although most people have problems with front yard parking, it is the easiest to pass topic.
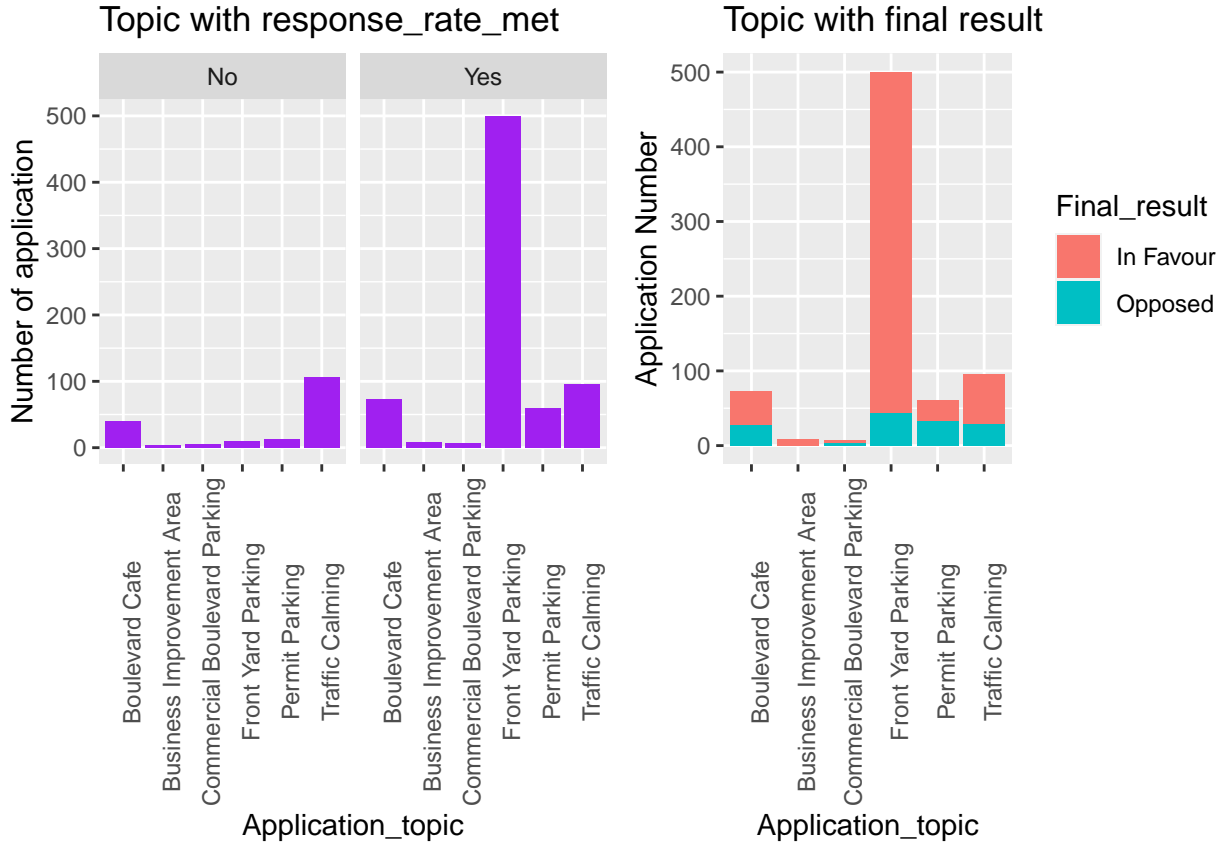
Figure 3: Combine two graphs to show information in detail

While the poll engagement behaves great in Toronto, shown by the paper's result. However, the result is not representative due to the limitations of the dataset using for analysis. The first point is that the application volume for each topic is not close to equal. For instance, almost half of the application is about front yard parking, which may cause the result unreliable. Also, the date of collecting data is inconsistent, which may hide the information behind. Finally, the required response rate for each application is different, which will influence the analysis. The information about how to generate the required response rate should be explained.

References

Assessing the Representativeness of Public Opinion Surveys. 2019. Pew Research Center. https://www.pewresearch.org/politics/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/.

Kassambara, Alboukadel. 2020. *Ggpubr: 'Ggplot2' Based Publication Ready Plots.* https://CRAN.R-project.org/package=ggpubr.

Office, City Clerk's. 2021. *Polls Conducted by the City.* Toronto, Canada. https://open.toronto.ca/dataset/polls-conducted-by-the-city/.

Opinion Poll. 2021. Wikipedia contributors. https://en.wikipedia.org/wiki/Opinion_poll.

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

Turk, Nunan, Heneghan. 2019. *Non-Response Bias.* https://catalogofbias.org/biases/non-response-bias/.

What Is Public Opinion Polling and Why Is It Important? 2007. Gallup. http://media.gallup.com/muslimwestfacts/pdf/pollingandhowtouseitr1dreveng.pdf.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, and Jim Hester. 2020. *Readr: Read Rectangular Text Data.* https://CRAN.R-project.org/package=readr.

Xie, Yihui. 2020. *Bookdown: Authoring Books and Technical Documents with R Markdown.* https://github.com/rstudio/bookdown.

Zeileis, Achim, and Gabor Grothendieck. 2005. "Zoo: S3 Infrastructure for Regular and Irregular Time Series." *Journal of Statistical Software* 14 (6): 1–27. https://doi.org/10.18637/jss.v014.i06.