Polls for Traffic Calming Conducted by Toronto City

Yizhen Wang

06 Febrary 2022

Abstract

Polls conducted by City Office establish residents' opinions on various city topics and issues. This report will focus on the topic of traffic calming and analyze its poll results through polls dataset provided on Toronto Open Data Portal. The report has concluded that most residents tend to vote in favour of proposal for traffic calming, and the bias caused by missing ballots and potential voters are not common.

Introduction

With increasing size of cities, more and more urban issues such as lack of parking spots and traffic noises have appeared and influenced citizens' daily lives. Toronto City Office has began to conduct polls for residents to express their opinions on restricting these issues since 2015.
My purpose of this research report is to specifically analyze citizens' poll engagement towards traffic calming application. The originality of this goal comes from the fact that the urban noises has become one of the most common parts of negative externalities of the city expansion. Of all kinds of noises in cities, the sound of traffic has been the most serious one, which keeps making bad impacts on citizen's living. My interest in this area comes from living in the downtown area of Toronto. The rolling of car engines has affected my resting a lot, so I decided to do a research to find out how other citizens feel about the need of traffic calming. For this report, I will use the polls data to compare residents voting towards traffic calming proposals, and I will also analyze other factors that may affect poll results or create bias, such as missing ballots and potential voters. The dataset will be processed in R (R Core Team 2021) through packages tidyverse (Wickham et al. 2019) and dplyr (Wickham et al. 2021). Figures and tables will be produced and generated through packages ggplot2 (Wickham 2016), kableExtra (Zhu 2021) and knitr (Xie 2014).

Data

Data Source

Toronto's City Office has conducted over 1000 polls on behalf of City divisions, these polls are meant to establish the opinions of residents and businesses on various topics covered by a City by-law (*Polls Conducted by the City* 2022). The report is based on the data of poll results that published on the City of Toronto Open Data Portal, and the csv format of the data was loaded into my R Studio by the R package opendatatoronto (Gelfand 2020). The results contain several variables related to polls such as application topics, ballots distribution and casts, pass rate and so on. The collection of data starts from April, 2015, and the dataset was last refreshed on Feb 4th, 2022.[1]

Data Characteristic

The dataset mainly focuses on various polls' application topics and their important features as well as results. Some applications includes traffic calming, front yard parking and Boulevard Cafe, and this report will specifically analyze the voting towards traffic calming.

[1] https://github.com/KSaxophone/STA304-Paper1-code

Since the data collection has began from 2015 and continues to get updated, the dataset has covered a relatively long time span of polls engagement, which allows the data to be more inclusive and comprehensive. Also, the dataset only collects poll results in Toronto City, so the geographical range of people's voting would be more specific.
However, when polls are conducted, damaged or missing ballots will always exist and bring bias to the poll results. The consequences of ballots deficit may cause the final counts to be underestimated and influence the polls pass rate. What's more, there might be people who have opinions for the application topics but are not included in the polls, they can be seen as potential voters. The dataset has included this factor and counts the number of people residing within poll boundary range. This report will further analyze the numerical characteristic for potential voters on traffic calming polls and will also discuss the relationship between missing ballots and polls pass rate.

The population of the dataset is the residents in Toronto City, and the data of residents' engagement for various polls is collected through mailing. Every resident in the polling area is mailed a notice of the poll's information and deadline, a ballot, and a postage-paid return envelope (*Information Regarding Cafe, Parking & Traffic Polls* 2022). The problem of non-response may occur since some residents may miss the deadline to submit. If too many voters did not submit their ballots and non-response rate increased, the polls might not be able to meet the passing benchmarks and failed to proceed.

Specifically, for benchmark of traffic calming, the percentage of ballots mailed which must be returned is 50% plus 1, and the number of ballots submitted which must be marked in favor is 60% in order to proceed the proposal (*Information Regarding Cafe, Parking & Traffic Polls* 2022).

Data Processing

The whole dataset has 1054 observations and 25 variables, but I will only focus on the observations of traffic calming and several important variables in this report.
Crucial attributes include APPLICATION_FOR, which states the application topics; BALLOTS_CAST and BALLOTS_DISTRIBUTED, which respectively count the number of ballots returned and distributed; BALLOTS_IN_FAVOUR and BALLOTS_OPPOSED indicating number of ballots marked "in favour" and "opposed," POLL_RESULT, which shows final result of polls; POTENTIAL_VOTERS counting number of people who are potential voters and finally PASS_RATE counting number of returned ballots needed for a positive poll result (*Polls Conducted by the City* 2022).

In order to obtain the traffic calming data, first I used filter function to keep the application topics only include traffic calming, then I specifically selected the important variables mentioned above to exclude other less related factors. Finally, I checked missing values in the data and filtered them out(missing values only exist in pass rate variable). What's more, since the difference between ballots distributed counts and ballots cast counts indicates the missing ballots, thus I used mutate function to create a new variable called MISSING by subtracting BALLOTS_DISTRIBUTED and BALLOTS_CAST.
After the data cleaning process, there are 92 observations and 9 variables in my final dataset. A sample view of the dataset is displayed below.

```
## # A tibble: 6 x 9
##   APPLICATION_FOR BALLOTS_CAST BALLOTS_DISTRIB~ BALLOTS_IN_FAVO~ BALLOTS_OPPOSED
##   <chr>                  <int>            <int> <chr>            <chr>
## 1 Traffic Calming           24               43 19               5
## 2 Traffic Calming           55               78 47               8
## 3 Traffic Calming           54               90 43               10
## 4 Traffic Calming           82              138 63               13
## 5 Traffic Calming          134              271 99               33
## 6 Traffic Calming           68              138 44               17
## # ... with 4 more variables: POLL_RESULT <chr>, POTENTIAL_VOTERS <int>,
## #   PASS_RATE <chr>, MISSING <int>
```

Important Variables

Ballots In Favour and Opposite

First thing I want to analyze is to find our residents opinions distribution towards proposal of traffic calming. The variable poll results has recorded whether each poll of traffic calming is in favor or opposed. Thus, I have built both a table and a figure to demonstrate the distribution of ballots in favor and opposed. The graph was conducted by ggplot2(Wickham, 2016), the table was made through kableextra (Zhu 2021) and knitr (Xie 2014).

Table 1: Numeric Results for Ballots Results

POLL_RESULT	min	Q1	median	Q3	max	mean	sd
In Favour	11	38.0	55	101.0	297	73.53846	51.40558
Opposed	13	43.5	57	94.5	309	80.70370	64.88381

Table 1 has illustrated the numerical characteristics of poll results based on the number of ballots cast. Note that here I used ballots cast, which means ballots returned, rather than ballots distributed, in this case I could excluded the influence of missing ballots. From the table we can see that although the mean counts of polls opposed is a bit higher than polls in favour, the standard deviation of polls in favour is smaller than that of polls opposed. What's more, polls result in favour has a smaller difference of mean and median as well as a wider range of IQR, which indicates that the distribution of in favour poll results is more symmetric. These observations has shown that votes for "in favour" is more stable and less extreme than votes votes for "opposed."

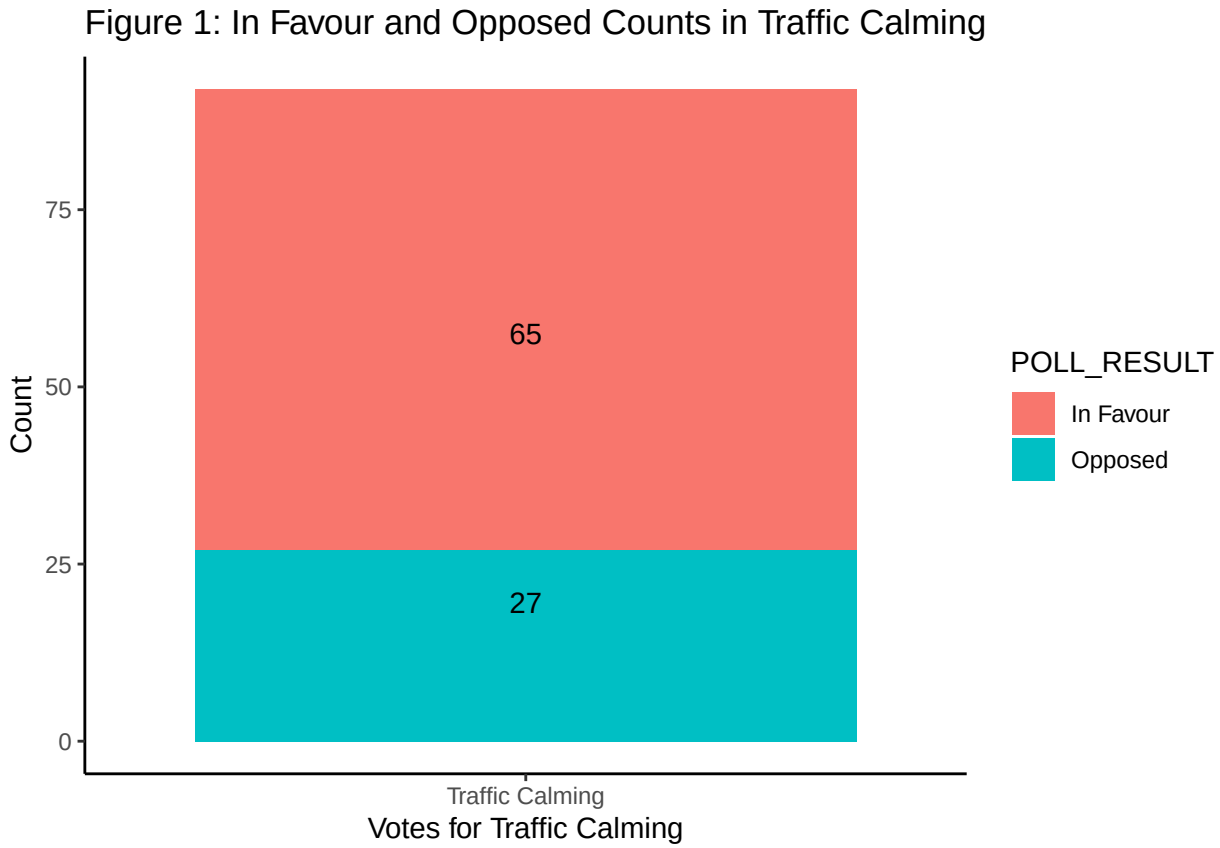

Figure 1: In Favour and Opposed Counts in Traffic Calming

Meanwhile, figure 1 has shown the "in favour" and "opposed" ballots counts in results of traffic calming polls.

Based on the graph, there are 65 counts of "in favour" and 27 counts of "opposed," which indicates that people who vote in favour for traffic calming take up 70% of the total. The result tells that traffic noises indeed have influence on residents' living and most people feel that a proposal for traffic calming is needed.

To sum up, most residents tend to vote for in favour towards traffic calming proposal.

Missing Ballots Counts and Pass Rate

Here I have used ggplot2 (Wickham 2016) to plot the scatter plot to observe the relationship between missing ballots counts and the pass rate of polls. The purpose of this section is to discover whether the number of missing ballots will have impact on polls pass rate.

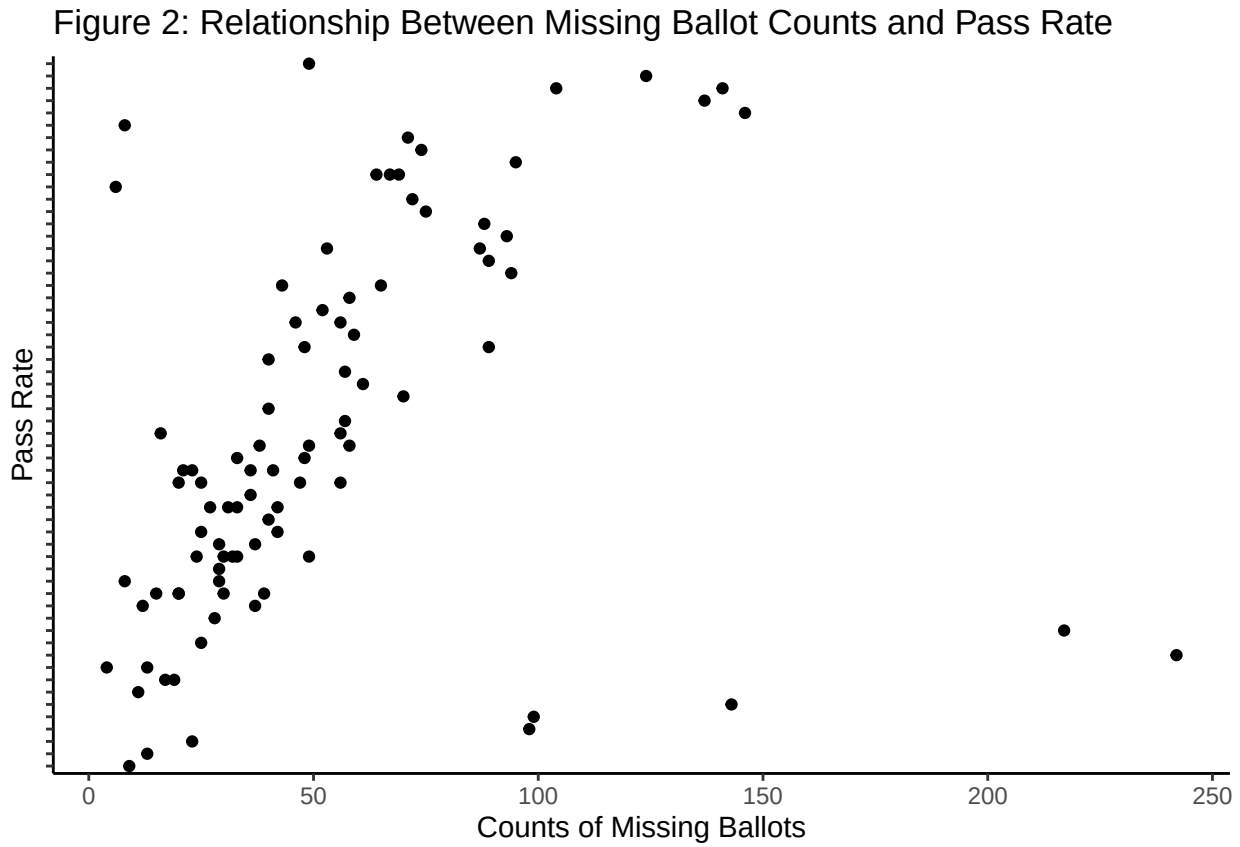

Figure 2: Relationship Between Missing Ballot Counts and Pass Rate

By plotting the counts of missing ballots as dependent value, we can see that the figure 2 scatter plot has a relatively linear trend. The pass rate and missing ballots obtain a mostly positive correlation. As number of missing ballots increases, the polls pass rate increases as well. However, there are several outliers appear around 100, 150 and 250 counts of missing ballots, the pass rate for these points are still quite low. Thus, the graph has proved that the quantity of missing ballots normally won't have too much impact on polls pass rate; while if the missing counts become very high, the final pass rate of polls will surely be underestimated at some level.

Potential Voters

Here I have used ggplot2 (Wickham 2016), kableExtra (Zhu 2021) and knitr (Xie 2014) to conduct the box-plot for potential voters and a table for ballots distributed. This factor has summarized the number of people residing within poll boundary range. The purpose of this section is to compare quantities of potential voters and ballots distributed.

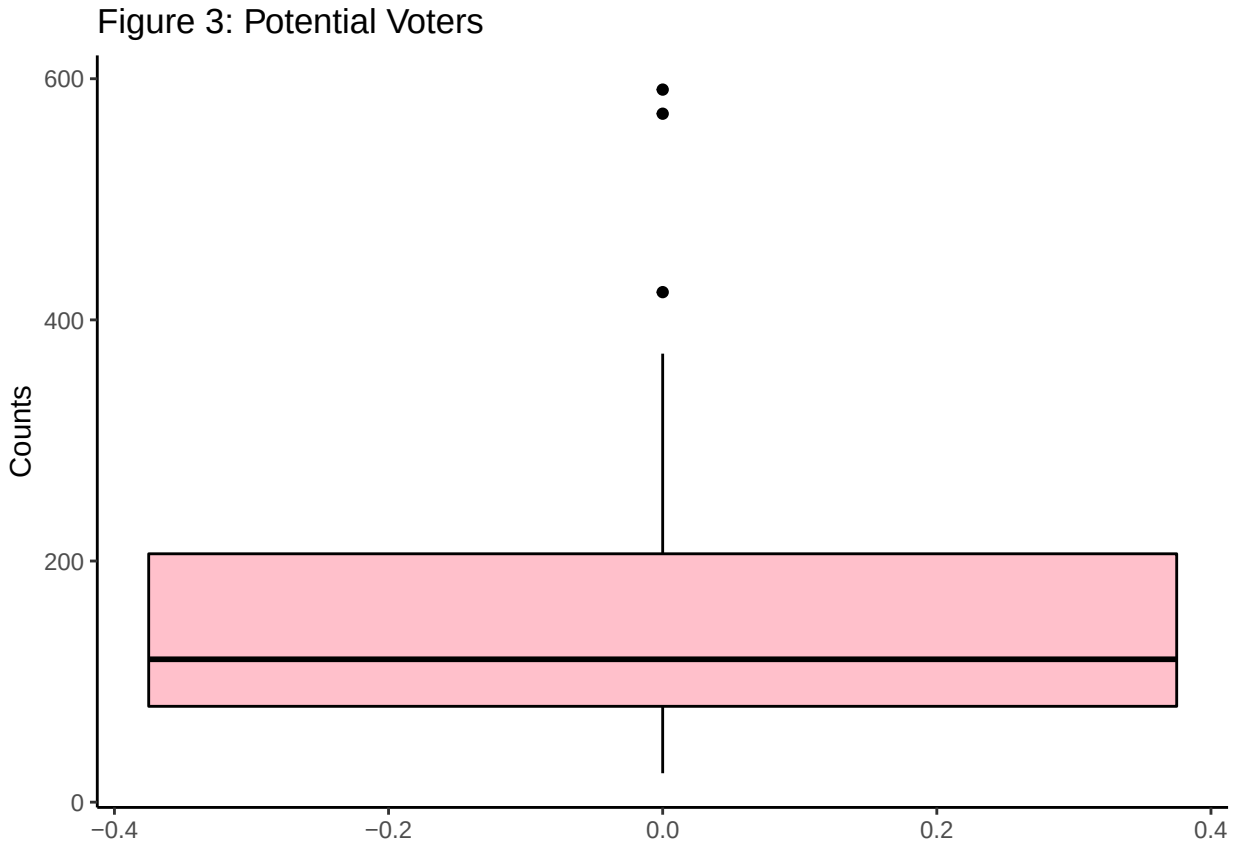

Table 2: Numeric Results for Ballots Distributed

min	Q1	median	Q3	max	mean	sd
17	69	97	186.25	539	129.163	95.0349

From the figure 3 we can see that quantity of potential voters has a range approximately from 80 to 200 and it is right skewed, its mean lies around 140. As for numerical results of ballots distribution in table 2, it has a IQR range from 70 to 190, and its mean is around 130. Compared to the box-plot information about potential voters, their means and IQR range are quite similar. Thus, the City Office has covered most potential voters in its polls conduct and the bias of missing qualified voters can be limited.

Reference

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal.*

Information Regarding Cafe, Parking & Traffic Polls. 2022. https://www.toronto.ca/city-government/planning-development/polls-regarding-changes-in-a-neighbourhood/information-regarding-cafe-parking-traffic-polls/.

Polls Conducted by the City. 2022. City Clerk's Office. https://open.toronto.ca/dataset/polls-conducted-by-the-city/.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

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, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*.