Surveying Toronto residents' opinions on various applications covered by the City by-law:City Clerk's Office needs to improve the polling mechanism and further explore possibilities*

Xincheng Zhang

22 January 2024

Abstract

As the number of residents continues to grow in Toronto, the poll is an effective method in the process of collecting and analyzing citizens' opinions on various project applications at the City Clerk's Office. This survey is mainly based on a data set from April 1, 2015, to January 20, 2024 (the last refresh), through discovering various variables such as application project name, total number of votes, etc. of Toronto citizens. Analyzing the enthusiasm, results, and hidden possibilities and giving relevant departments such as the City Clerk's Office several results and guiding suggestions for making polls in the future.

Contents

*Code and data are available at https://github.com/zxc0707/Paper1.git

1 Introduction

Toronto is located on the northwest coast of Lake Ontario, Canada. It is one of the most modern and prosperous cities in Canada. Since this city hosts a large number of native residents and residents who immigrated here from all over the world, effective public opinion polling has become a critical social activity that the city needs to pay attention to and implement. Specifically, residents' parking areas, safety areas, improvement areas, and traffic sound issues have become the focus of this research report. I think these applications deserve the spotlight. For instance, the reason why the parking topic can be regarded as a focus also comes from the increasingly strict parking enforcement in urban Toronto, Ontario, especially for commercial vehicles, which gradually take the lead in ticketing and towing.(Rosenfield et al. 2016) Besides, the issue of traffic calming is also worthy of the City Clerk's Office to collect public opinions from Toronto and conduct governance. Various studies in distinctive countries have shown that pedestrians and drivers can benefit from traffic calming methods such as slowing down vehicle traffic and shortening crossing distances.(Huang and Cynecki 2000)

Data were collected and organized based on the concerns of Toronto residents to conclude the decisions of each application type and provide several reasonable suggestions to The City Clerk's Office.

Data from About Polls conducted by the City, compiled by The City Clerk's Office, provide residents' attitudes towards a total of 11 different application project types in various areas of Toronto, divided into three categories: support, opposition, and neutral attitudes.

With the help of statistical data analysis, It provides The City Clerk's Office with some results on public interest for each application type. In the data section, I will explain the basic information of the data set I used in this report in the "About Polls conducted by the City" section, and use code and text to explain the production of each statistical chart and what useful information it provides. Next, I will elaborate on my thoughts and insights on other possibilities in the "Uncertain poll" section. Finally, the "Discussion" section summarizes the contents of the entire report and gives reasonable suggestions to the City Clerk's Office.

This statistical report examines Toronto citizens' attitudes toward 11 hot-spot applications in a citizen survey that opened voting for 60 days or less from April 1, 2015, to January 20, 2024. I used code to generate 4 statistical charts with visual meaning based on the given data features and content. Because of the importance of public opinion polls to the city of Toronto, I used Figure 1 to study the sensitivity of Toronto residents to the length of time polls are open, and to identify the number of days polls are open that are more likely to arouse Toronto residents' enthusiasm for voting. This will help The City Clerk's Office design more beneficial activities when conducting citizen surveys every month in the future. From Figure 2, I can clearly see how citizens' support for these 11 application projects is distributed. However, through observation, I found that the number of support and opposition for some projects is relatively equal, such as Commercial Boulevard Parking and Traffic calming island. Because there are still some uncertain votes, I have doubts about the final results of these application projects. Therefore, I created new variables and derived the comparison between the number of uncertain votes and the total number of distributed votes for each project in Figure 3, and discussed the support and opposition for these application projects that may have the results reversed in Figure 4, which compare the number of projects with similar votes and uncertain votes, try to explore whether these uncertain votes have the ability to reverse the final decision of The City Clerk's Office. The results of the data show that among these 11 types of application projects, 2 applications have the possibility of opposite results. This shows that when inconclusive votes are confirmed by some method, in the case of majority opposition, The City Clerk's Office will present the opposite situation based on the results of the citizen survey.

2 Data

2.1 About Polls conducted by the City

To investigate Toronto residents' attitudes toward 11 different application programs, I obtained the "About Polls conducted by the City" data set from the open data Toronto website. This data set is a table type provided by the City Clerk's Office and updated daily, including citizens' attitudes of 11 focus application

types from April 1, 2015, to January 20, 2024 (the last refresh). The database reflects Toronto citizens' support for various projects through three types of vote content: support, opposition, and neutral attitude (No mark to identify in "favor" or "opposed"). In addition, in addition to the three variables of ballot content, there is also an indeterminate ballot marked as "SPOILED", which refers to ballots received that were not clearly marked as either "in favor" or "opposed". In order to avoid errors of judgment in the final decision made by The City Clerk's Office. I will also use 2 constructed variables "DURATION" and "DIFFERENCES" in the data set to perform further analysis on some items that have the potential to be reversed.

The data set contains 1,235 citizen surveys (rows) and 25 different variables (columns) about specific application projects. Each line represents a poll and contains various information, such as application type name, application address, total number of votes, etc.

Table 1: Selecting first 10 rows of part of required columns from the residents'opinions dataset

BALLOTS_SPOILED	BALLOTS_DISTRIBUTED	OPEN_DATE	CLOSE_DATE
1	34	2015-05-01	2015-05-31
0	36	2015-05-04	2015-06-03
3	97	2015-05-12	2015-06-11
5	334	2015-05-06	2015-06-04
1	106	2015-05-06	2015-06-05
1	235	2015-05-13	2015-06-11
2	235	2015-05-11	2015-06-09
5	109	2015-05-28	2015-06-26
30	340	2015-09-14	2015-10-14
1	80	2015-05-25	2015-06-23

Table 1 shows the first 10 rows of the resident opinion data set. It should be noted that I used more than these 4 columns in this report. The data information represented by each column is explained in detail below by listing them one by one.

This statistical report focuses on 11 of these 25 variables: vote start time, vote end time, vote type name, number of votes in favor, number of votes in oppose, number of neutral votes, number of uncertain votes, total number of votes distributed, differences between support and oppose, final voting list of the total number of voters, election duration.For time variables, it is time format data. For text variables, it is data in text form. For quantitative variables, measurement is in the form of a numeric variable.

By using R (R Core Team 2020), and R packages "tidyverse" (Wickham et al. 2019), "dplyr" (Wickham et al. 2023), "janitor" (Firke 2021), "lubridate" (Grolemund and Wickham 2023) and "knitr" (Xie 2022).

The variable "BALLOTS_SPOILED" represents the Number of ballots received and were not clearly marked as either "in favor" or "opposed".The variable "BALLOTS_DISTRIBUTED" represents the Number of ballots distributed.The variable "BALLOTS_IN_FAVOUR" represents the Number of ballots received and marked "in favor' '.The variable"BALLOTS_OPPOSED" represents the Number of ballots received and marked "opposed".The variable "BALLOTS_BLANK" represents the Number of ballots received with no mark to identify in "favor" or "opposed".The variable "BALLOTS_SPOILED" represents the Number of ballots received and were not clearly marked as either "in favor" or "opposed".The variable "OPEN_DATE" represents the Date the poll is open to public.The variable " CLOSE_DATE" represents the Date the poll has been closed.The variable "APPLICATION_FOR" represents the Type of application.The variable "DURATION" represents the duration of this poll and the variable "DIFFERENCES" represents the differences in support and opposition of this poll.

There is also a variable in the original database called BALLOTS_RECEIVED_BY_VOTERS, which represents the Number of ballots returned to City Clerk's Office. The reason why this variable does not replace BALLOTS_DISTRIBUTED (Number of ballots distributed) is because when the ballots are sent out

and filled out by citizens, some errors such as lost or omitted ballots that are returned to the City Clerk's Office should not be allowed to lose the total statistics. This is unfair to citizens and will affect the final judgment.

The poll is an important method in Toronto to collect citizens' thoughts on different types of application projects. I want to study citizens' sensitivity to voting duration by creating a scatter plot first. In other words, the relationship between duration and participation. It effectively gave the City Clerk's Office a persuasive suggestion to adjust the length of future polls.

Figure 1 is composed of the voting duration on the x-axis and the number of total voters on the final poll list on the y-axis. This chart mainly shows the sensitivity of Toronto citizens in making judgments about the length of time to submit their ballots. To be more specific, from the available data, a 30-day voting duration is the most encouraging for citizens to actively participate in voting. As for the comparison of length of time, it can be seen that when the voting duration is less than 30 days, there is less participation than when the voting duration is 60 days. This means that for The City Clerk's Office in Toronto, a 30-day voting period can ensure the highest level of citizen participation. The second is 60 days, and the worst effect is within 30 days for the citizen in Toronto.

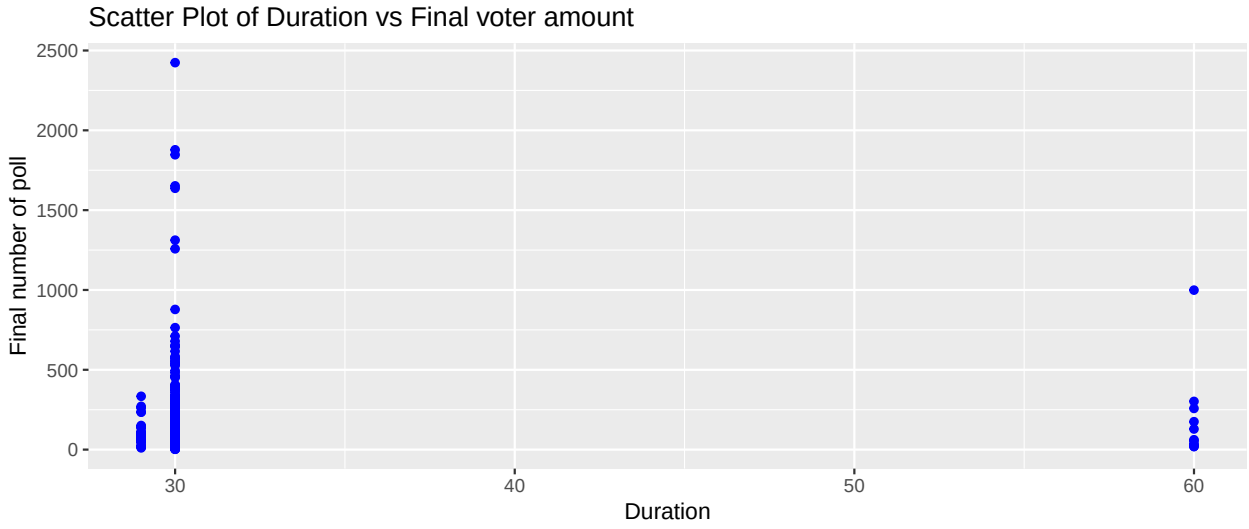

Figure 1: The impact of voting duration on citizen participation starts from April 1, 2015

Figure 2 shows the voting results of Toronto citizens for different application types. There are three types of voting results: support, opposition, and neutral (no mark to identify in "favor" or "opposed"), which correspond to three different colors in the histogram: green, blue, and red. In the x-axis, there are 11 different application types: Appeal-Front Yard Parking/Front Yard Parking/Permit Parking/Traffic Calming/Commercial Boulevard Parking/Boulevard Cafe/Permit Parking Removal/Proposed Business Improvement Area/Traffic Calming island/Traffic Calming Safety zone/Business Improvement Area.

By observing the histogram, It can be clearly observed that for the Front Yard Parking application project, citizens showed very different voting results. The number of votes supporting this application project was nearly 6 times that of those opposing it. This voting results not only reflect citizens' demand for front yard parking but also reflect citizens' enthusiasm for such topics. This can also be concluded through the statistical results of the Appeal-Front Yard Parking project. For the Commercial Boulevard Parking and Proposed Business Improvement Area, citizens did not respond with strong feedback.

In addition to clearly marked ballots, The City Clerk's Office also received many samples of unmarked ballots. I think this also influences the overall research topic and deserves to be discussed.

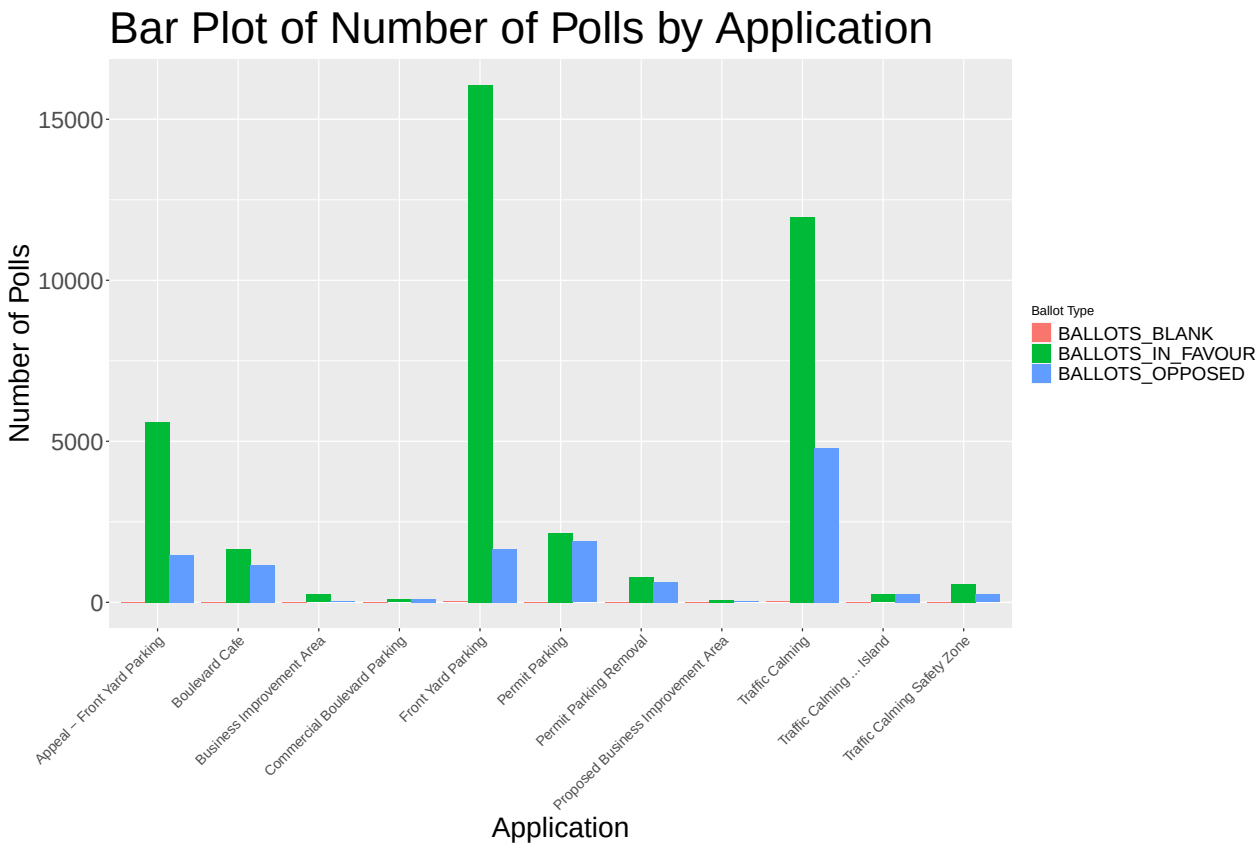

Figure 2: Levels of citizen support/opposition/neutrality for different types of applications

2.2 Uncertain poll

When browsing the entire database, besides the three choices mentioned above: support, opposition, and neutrality, there is also a sample of votes marked as "SPOILED", which is a vote that does not clearly indicate one's own thoughts. I think we also need to divide this kind of votes into consideration ranges and make corresponding charts to analyze and conjecture a summary of all possibilities compared with clearly labeled samples. Figure 3 shows us a comparative histogram of the total number of ballots issued by The City Clerk's Office (BALLOTS_DISTRIBUTED) and the number of indeterminate ballots (BALLOTS_SPOILED) for various application types.

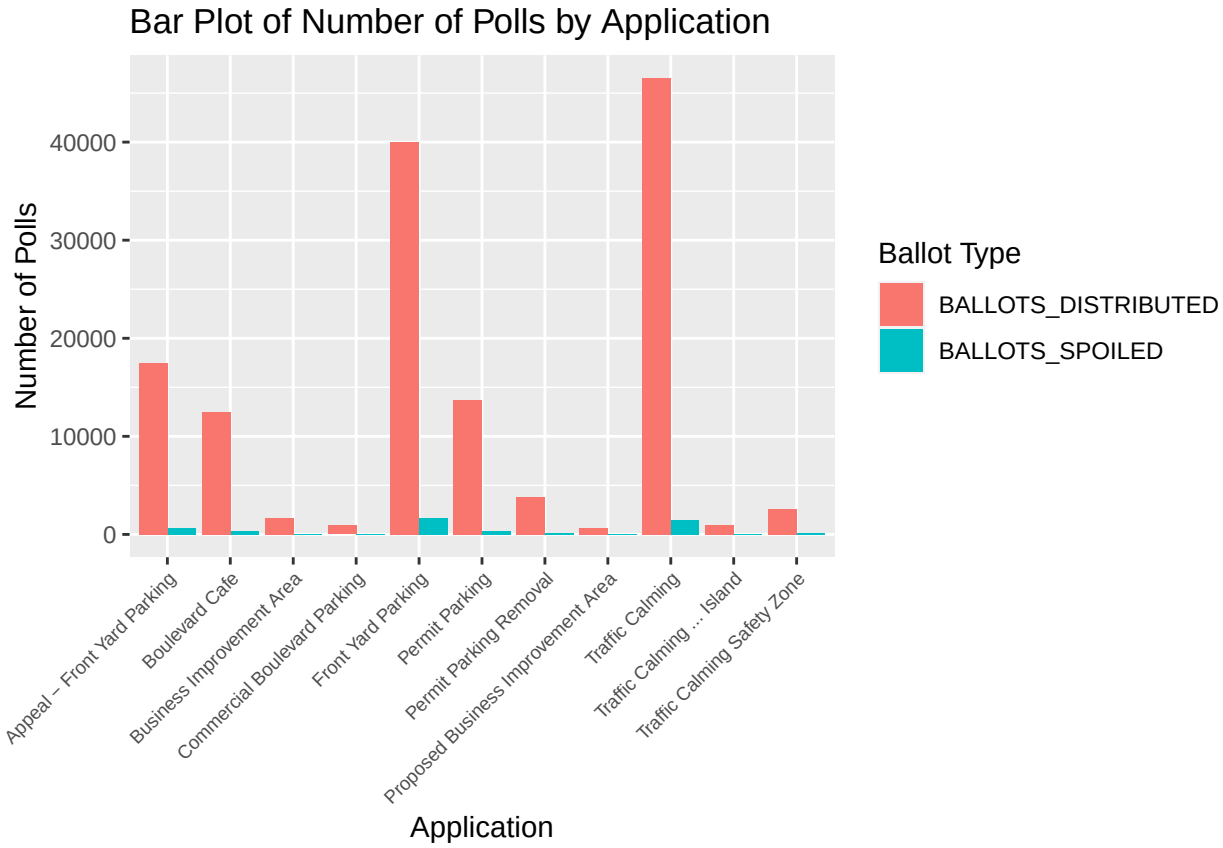

Figure 3: Total distributed electoral votes vs. indeterminate votes

We can clearly observe that as the total number of released votes increases, the number of uncertain votes increases. What I'm interested in is whether these uncertain votes counted in the whole 11 areas have the possibility of reversing the final research results. Therefore, based on the results in Figure 2, I concluded that among these 11 areas, only 3 areas do not have this possibility: Appeal-Front yard parking/Front yard parking/Traffic calming. Because for these three areas, their results are very different, there is not a relatively large difference in the number of people supporting and opposing in other areas, which means that when these uncertain votes are determined. The final decision will be changed which requires further analysis.

Figure 4 is to compare the difference between support votes and opposition votes (pink) and the number of uncertain votes (grey). Through the chart, we found that the two areas of Commercial Boulevard Parking and Boulevard Cafe are evenly distributed. This shows that when the results of uncertain votes are confirmed through some methods, the conclusions of data research on these two areas will be overturned.

In addition, the reason why uncertain ballots are marked as "SPOILED" should definitely be considered within the scope of consideration. Specifically, the reason why the ballot is not confirmed is because the handwriting is confusing, unreadable, or the ballot is damaged. This situation will further determine the

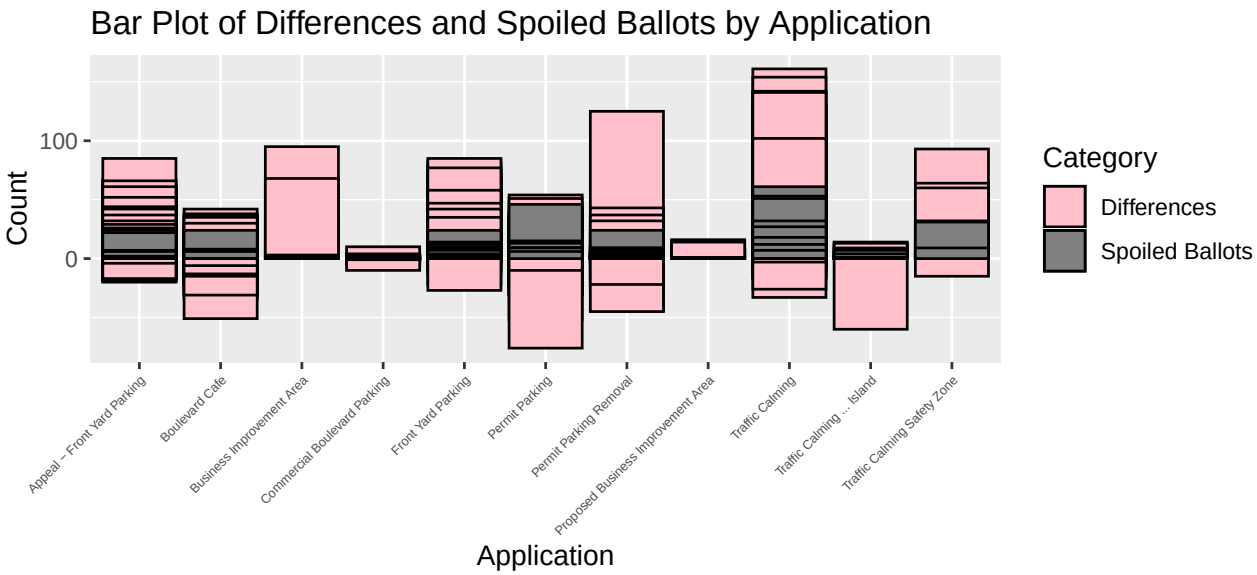

Figure 4: Differences in support and opposition compared with uncertain votes

control of the variable of uncertain votes on the overall results.

However, since there is no summary of the above-mentioned possibilities in the original database, we can only discuss and summarize based on the existing data.

3 Discussion

Based on the whole report, several conclusions can be drawn from the analysis of the data set that extracted 9 variables from the original database of open data Toronto and added two new variables by using code.

Firstly, the duration of the poll has a certain impact on citizen participation in Toronto. The relation can be concluded from the scatter plot in Figure 1, which implies the City Clerk's Office needs to try to control future polls to 30 days to attract more residents to participate in voting or to effectively increase citizen participation in the poll.

Secondly, from Figure 2 we can clearly observe that citizens' opinions on the two application projects, Front yard parking and Traffic calming, showed clear voting results. The City Clerk's Office can begin to gradually consider getting these two applications up and running. For the remaining nine categories of projects, public opinion collection needs to be continued.

Thirdly, Figure 3 shows us the total number of votes distributed by the City Clerk's Office for each project. Since the two items mentioned above do not require continuous public opinion adoption in the future, the excess votes can be averaged or allocated to the other 9 items based on adjustment preferences. As the total number of allocated votes increases, the overall voting results of the people will be more obvious. After this, the same analytic approach can be used to identify those projects that citizens clearly support and are considering for implementation. On the contrary, the applications are opposed by citizens and replaced with new proposals.

Lastly, due to the existence of votes marked as "SPOILED", by observing Figure 4, we can see that there are risks in judging the two projects of Commercial Boulevard Parking and Boulevard Cafe if only based on existing data. This means that if the majority of the uncertain votes are determined to be objectionable through some method (for example, re-election, home visits to ask for opinions, etc.), the City Clerk's Office will make errors in the judgment of these two results and cause dissatisfaction among citizens. My suggestion is to distribute more votes to these two items in subsequent poll settings and increase the overall number of

votes so that citizens' opinions become obvious trends or differences for subsequent processing of applications.

In addition, there appear to be different views on the effectiveness of this type of polling in civic activities depending on the voting items. For example, for presidential election, the error range of poll election is mostly maintained within 6%. On the contrary, polls about social and economic issues did not show such a high accuracy rate.(Field and Connelly 1942) The prospect of polling as a citizen activity is also a huge challenge for the City Clerk's Office, because limited resources and incompatible rewards will lead to citizens not paying enough attention to the activity of polling and lacking enthusiasm, as mentioned in the article.(Parry, Kisida, and Langley 2008)

References

Field, Harry H, and Gordon M Connelly. 1942. "Testing Polls in Official Election Booths." *The Public Opinion Quarterly* 6 (4): 610–16.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/janitor.

Grolemund, Garrett, and Hadley Wickham. 2023. *Lubridate: Make Dealing with Dates a Little Easier*. https://CRAN.R-project.org/package=lubridate.

Huang, Herman F, and Michael J Cynecki. 2000. "Effects of Traffic Calming Measures on Pedestrian and Motorist Behavior." *Transportation Research Record* 1705 (1): 26–31.

Parry, Janine A, Brian Kisida, and Ronald E Langley. 2008. "The State of State Polls: Old Challenges, New Opportunities." *State Politics & Policy Quarterly* 8 (2): 198–216.

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

Rosenfield, Adam, James Lamers, Mehdi Nourinejad, and Matthew J Roorda. 2016. "Investigation of Commercial Vehicle Parking Permits in Toronto, Ontario, Canada." *Transportation Research Record* 2547 (1): 11–18.

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. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2022. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.