

Internet Usage and Voting Trends in Canada*

Mahak Jain

27 April 2022

Abstract

Understanding the demographics of voting is essential for every country to include policies for a larger voter turnout. This paper understands the relationship between the characteristics of an individual such as social media usage and internet usage, voting participation and to understand the relationship between them. These findings help implement policies to ensure better voter participation and also understand trends in voter turnout. Keywords: voting, internet, social media, volunteer, gender, toronto residents

1 Introduction

Internet consumption is a worldwide phenomenon. Globally, the number of Internet users increased from only 413 million in 2000 to over 3.4 billion in 2016 alone. The usage and accessibility of internet consumption depend on several factors, differing through socioeconomic factors and demographically. Internet usage has become a prominent part of everyone's daily life, from an accessory to a necessity. In 2019, Canada had an estimated 34.56 million internet users. Over 96.5% of Canadians have access to the internet in their households.

There exists a long tradition of research documents and demographic and psychological determinants of political participation, there exists evidence to suggest that changes in communication technology may play an important role in influencing electoral behavior Mcneal and Tolbert (2003). Web-based and theoretical studies often claim that Internet use can mobilize political participation, while survey-based studies generally conclude that Internet use will normalize political participation Hirzalla, Zoonen, and Ridder (2011).

Voter turnout in Canada has declined in the last several decades in Canada, now being the lowest in the western world (Kushner and Siegel (2006)). According to research by Milner, 'countries that have the most politically knowledgeable voters have the highest voter turnout and that when information levels are high, individuals are more likely to vote' (Kushner and Siegel (2006)). The Internet may enhance voter information about candidates and elections, and in turn, stimulate increased participation Mcneal and Tolbert (2003). Though many European jurisdictions have established well-developed online voting models, Canada is quickly emerging as an important research case. To date, there have been more instances of remote Internet voting in local Canadian elections than in any other country. There have been more than two million remote Internet voting opportunities in over 90 local Canadian elections. This paper explores internet usage and political engagement with voter participation and turnout in federal, municipal, and provincial voting.

The Internet may enhance voter information about candidates and elections, and in turn, stimulate increased participation Mcneal and Tolbert (2003). Though many European jurisdictions have established well-developed online voting models, Canada is quickly emerging as an important research case. To date, there have been more instances of remote Internet voting in local Canadian elections than in any other country. There have been more than two million remote Internet voting opportunities in over 90 local Canadian elections. This paper explores internet usage and political engagement with voter participation and turnout in federal, municipal, and provincial voting.

This report is divided into three sections. The first section outlines the data that has been used to support the later analysis. Next, the second section discusses the trends observed from the data. Finally, the last

*Code and data are available at: https://github.com/MahakJain74/finalpaper_canadianvoting.

section explores the progress that has been made in this report, the weakness of the analysis, and the future directions. The data and code that support this report can be found in the Github repository:

2 Data

The cleaning, analysis, and visualization of the dataset in this paper were produced with the R statistical programming language R Core Team (2020) and in R markdown file format. Features from several packages were used for the data. The haven Wickham and Miller (2021), janitor Firke (2021), and tidyverse Wickham et al. (2019) package is used for data cleaning and manipulation. Analysis and figures were created using ggplot Wickham (2016). Other packages such as tidyr Wickham and Girlich (2022), patchwork Pedersen (2020), sf Pebesma (2018), dplyr Wickham et al. (2022).

2.1 Data Source and Features

The General Social Survey, Cycle 27, Social Identity was conducted in 2013. It is a cross-sectional survey sample. This paper uses General Social Survey, Cycle 27, Social Identity microdata focusing on two aspects: social networks and civic participation. A few key time-use metrics gathered are voter participation in federal elections, voter participation in municipal elections, voter participation in provincial elections, social networking accounts, internet usage, internet usage to buy goods and services, and internet usage for electronic banking. Not all variables are related to time-use as the survey is also interested in other metrics as well such as the respondent's sex, income, marital status, number of children, and other variables that are relevant to social policy. There exists a large variation for the respondents, for example, for most of the variables there exists "Don't know" or "Valid skip" to account for uncertainty from the respondents. Due to this, accurate information can be collected by respondents. To ensure data accuracy, the respondents who have access to their telephones are contacted for an interview to ensure receiving relevant information. The survey sample doesn't include respondents proportionate to their provinces and doesn't account for territories in Canada, therefore it is not representative of the Canadian population. The location of the respondents based on the original survey, General Social Survey: Social Identity, 2013, is shown in (Figure 1), made using the province's longitude and latitude.

2.2 Potential Ethical and Bias Concerns

There exists a bias in the Canadian General Social Survey and this paper due for several reasons. In the 2013 Canadian General Survey: Social Identity, all provinces were included except Yukon, Nunavut, and Northwest territories, leading to sampling errors. Exclusion of these territories, specifically the ones housing marginalized communities and only targeting respondents with access to a telephone leads to exclusion in the data frame and other policies formed based on the information. There exist questions about gender in the survey. Respondents are only given two binary options, with other options such as "Don't know" or "Refusal". Therefore in a way, eliminating the existence of other groups. Additionally, this also does not give an opportunity for those people who have intersectional identities to feel comfortable and accepted by the political system in Canada.

Moreover, there exists non-response bias and voluntary bias. Exclusions based on socioeconomic standpoint as equal opportunity is not given to all Canadians answering the survey as they can only access it if they have a stable telephone connection. There may exist voluntary bias due to respondents choosing not to answer certain survey questions due to privacy or reduction of invasivity. There exist non-sampling errors in the survey, which occurs when respondents give a 'wrong' answer to the questions, due to it being more desirable.

2.3 Data Terminology

2.3.1 Voting

The (Figure 2) is made using Wickham (2016). It shows the voter participation in the last federal election by age amongst the respondents. As respondents age, voter participation gradually increases, and by age 55-54, it starts to decrease. In this, the higher number of no voter participation is for the ages 15-24 but

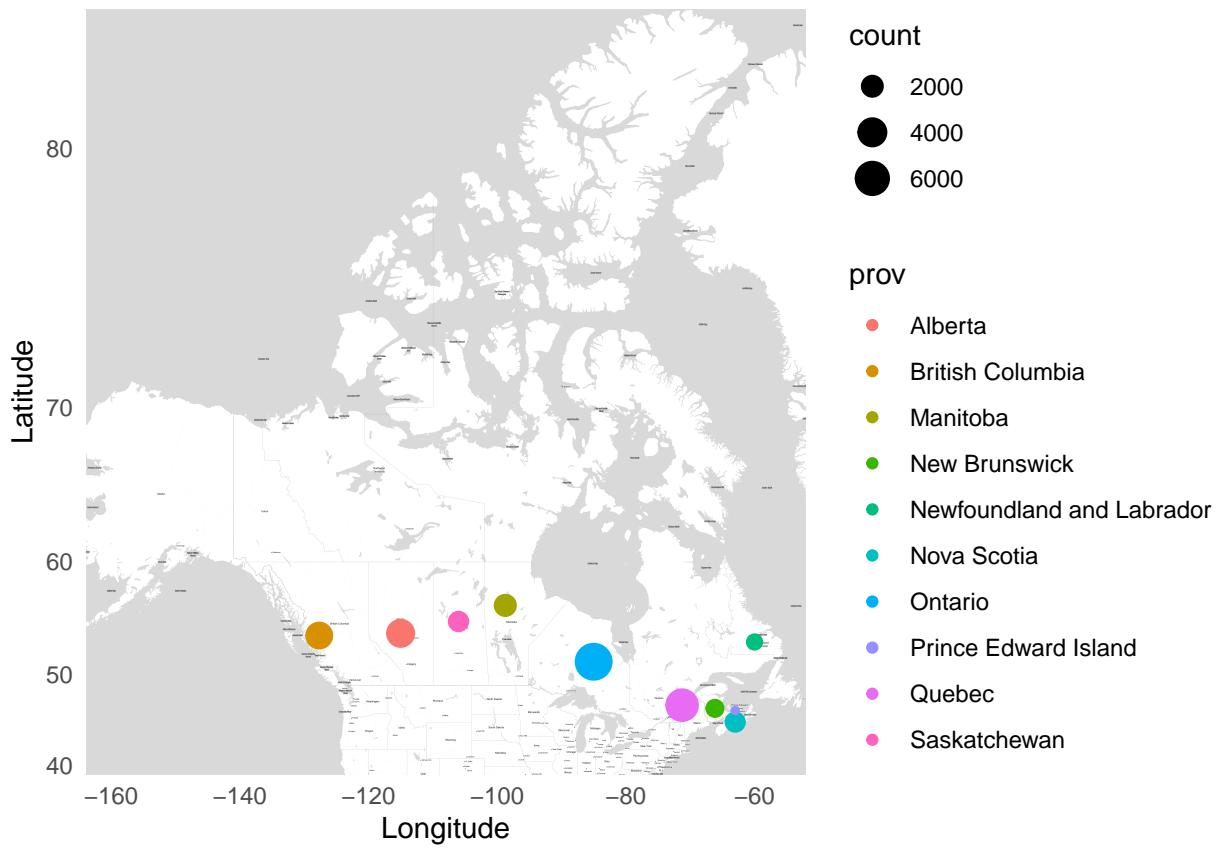


Figure 1: A map showing the approximate location of survey respondents

this could be due to non-eligibility, whereas for the ages 25 and above, there still exists a high number of respondents who did not vote in the last federal elections. The number of respondents who did not vote in the last federal election gradually decreases as respondents' age increases. Here it is important to note that there may exist cases of ineligibility above the required age of 18. This is explored in (Figure 3). (Figure 3) showcases the respondent's eligibility to participate in federal elections. As expected, respondents between 15 and 24 years of age have the highest cases of ineligibility. The cases of ineligibility are high for respondents between the age of 25-34 and 35-44. This could be due to several reasons concerning immigration. The cases of respondent ineligibility gradually decrease after the age group 35-44.

After considering the cases of non-voter participation and voter ineligibility, there still exists eligible voters who did not participate in federal elections. (Figure 4) explores the reasons for non-participation amongst voters in the federal elections. The main reason behind non-participation amongst respondents is due to interest. There is a high percentage of respondents who did not vote in the last election because they were not interested in voting. The second main reason for non-participation is due to scheduling issues. The participants were unable to make out time in their schedules to vote. This could also be because of the short time frame and prior commitments. Other primary reasons for non-participation were because respondents weren't informed on issues, they were out of town, they didn't like any candidates or campaigns, and felt as though their vote does not make a difference.

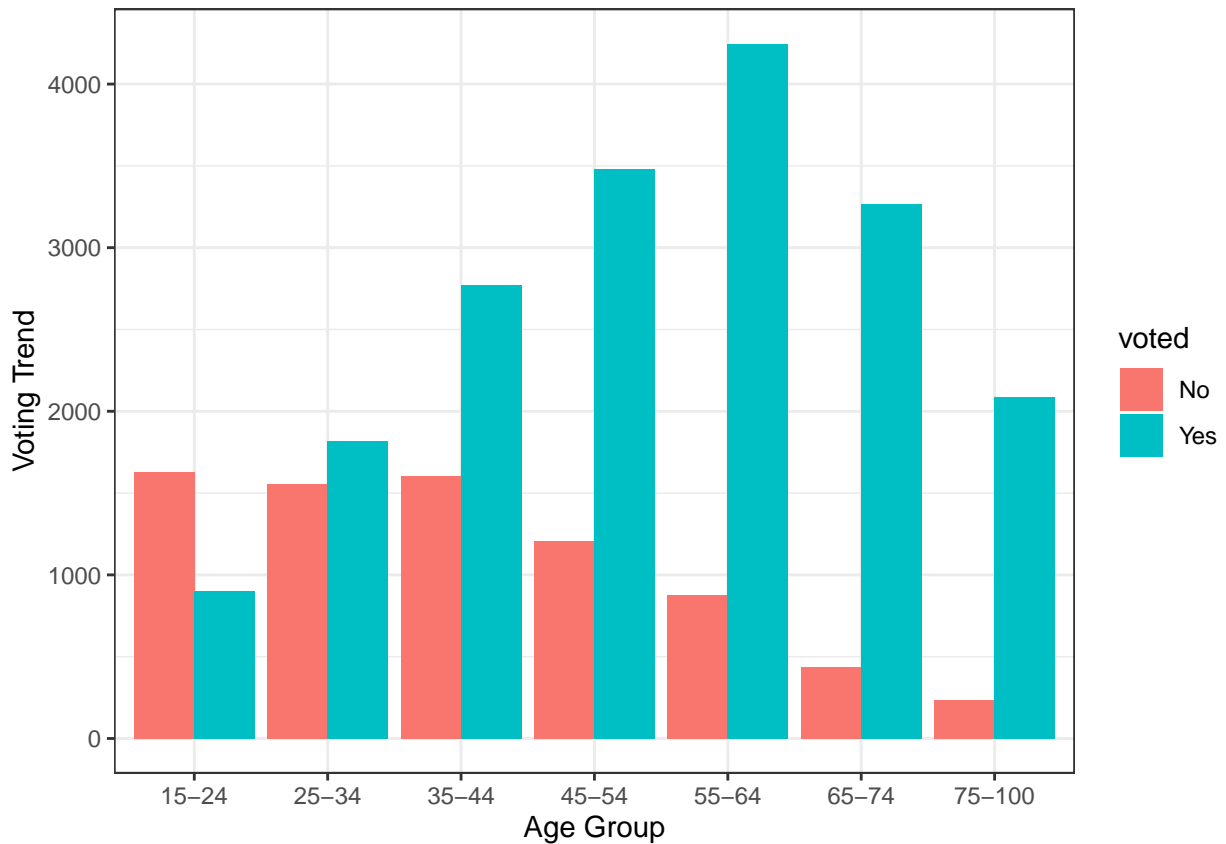


Figure 2: Voter Participation for Last Federal Elections amongst Respondents

2.3.2 Internet Usage

(Figure 5) showcases whether respondents have a social networking account according to their age. As expected, the age group 15-24 had a very high ratio of having a social media account. All ages, in general, showcase a high percentage of having a social media account. Here it is essential to note that (Figure 5) doesn't account for the total population of respondents between age groups. (Figure 6) explores the frequency

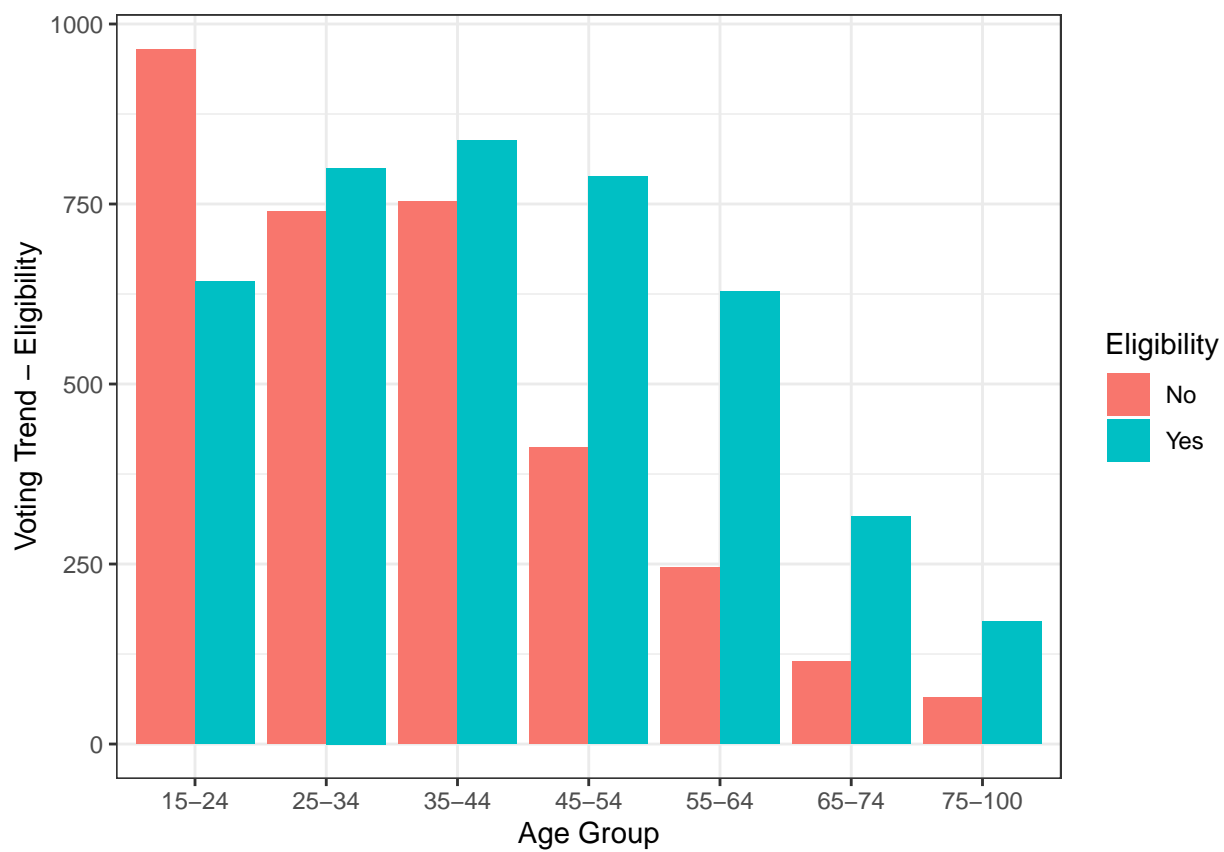


Figure 3: Voter Eligibility for Last Federal Elections amongst Respondents

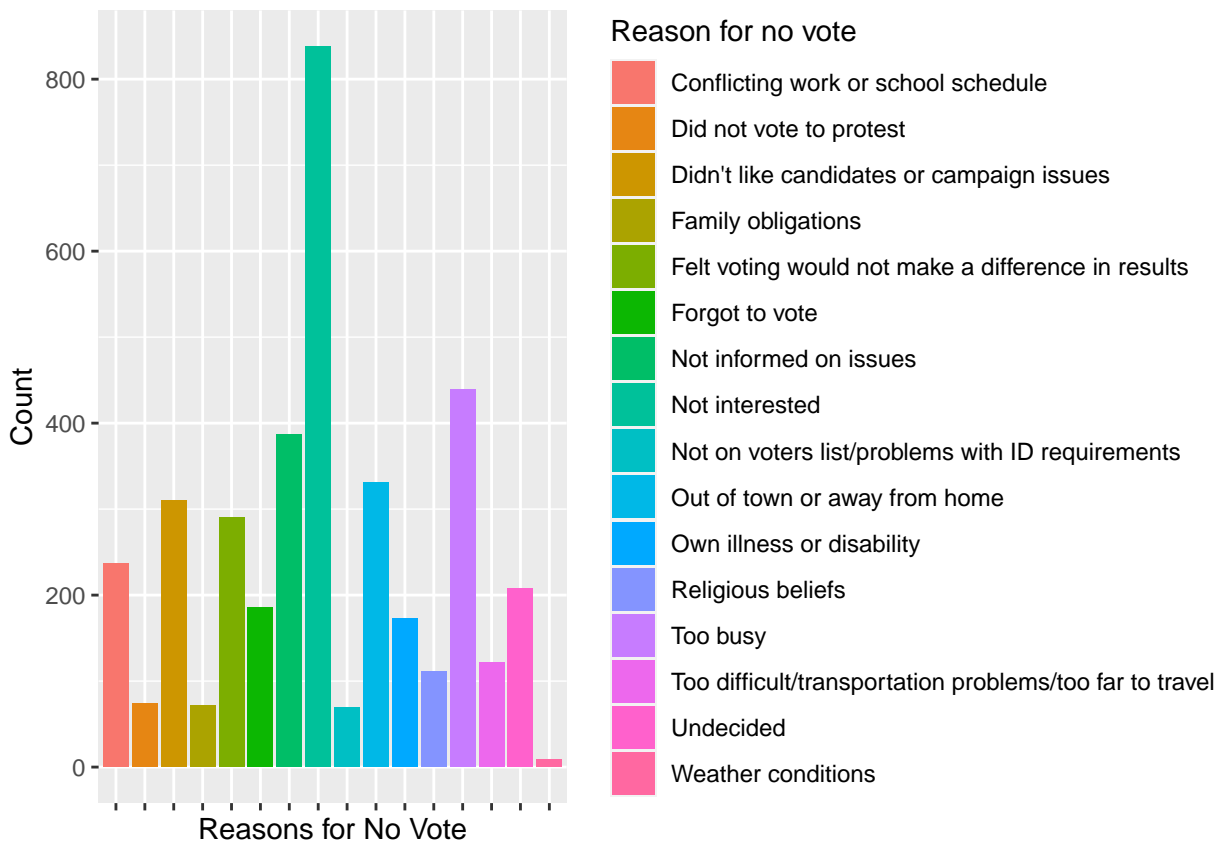


Figure 4: Why did Respondent not Vote Last Federal Elections?

of access to social media networking accounts amongst respondents. As expected, social media is accessed several times a day for all age groups, followed by accessing social media about once a day. There is a consistent high frequency of access to social media accounts for all age groups except for age group 75 and over.

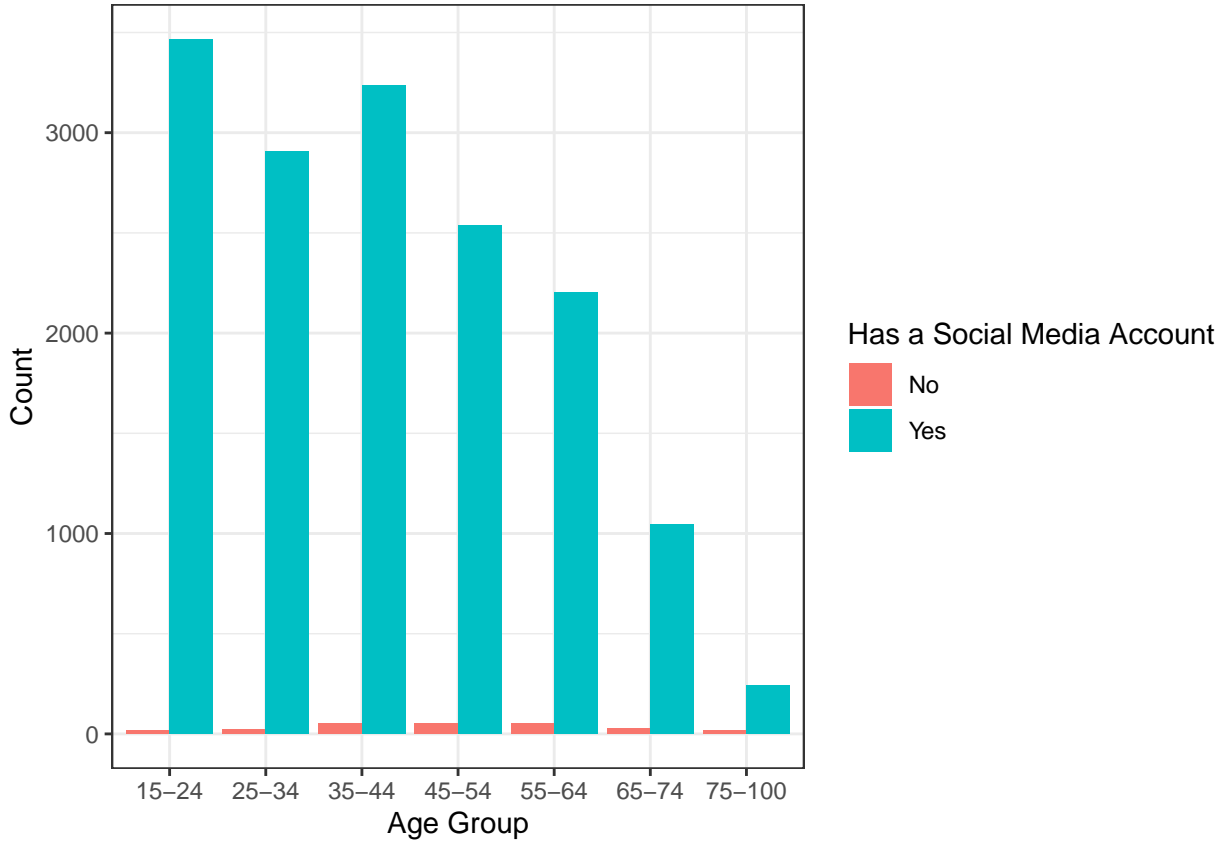


Figure 5: Social networking account - amongst Respondents

(Figure 7) explores the internet usage to buy goods and services in a month amongst the respondents. As expected, the age group 15-24 has a high percentage to purchase goods and services on the internet and the percentage gradually increases till the age group of 35-44. There also exists a high percentage of respondents between the ages of 15-24 who never use the internet to purchase goods and services. Similarly, (Figure @ref(fig:elecbanl_intuse)) explores the usage of electronic banking on the internet usage for all age groups amongst the respondents. For the ages group 15-24, there exists a large percentage of respondents who use the internet for electronic banking at least once a week, followed by never using the internet, followed by using it once a month, and then once a day. These trends exist similarly for mostly all age groups, with a higher percentage of respondents using the internet once a week, except for the age group 65 and over.

2.4 Model Methodology

This paper aims to understand the effect of internet usage on voter participation in the federal election. Since voter participation is a binary outcome, logistic regression is best suited for the model. The logistic model helps estimate P , the probability of the outcome variable Y to be successful given the inputs. The survey, General Social Survey: Social Identity, 2013 consisted of multinomial variables, of which variables related to internet usage were selected for the model. No continuous variables were included in the model, so the precision is not affected.

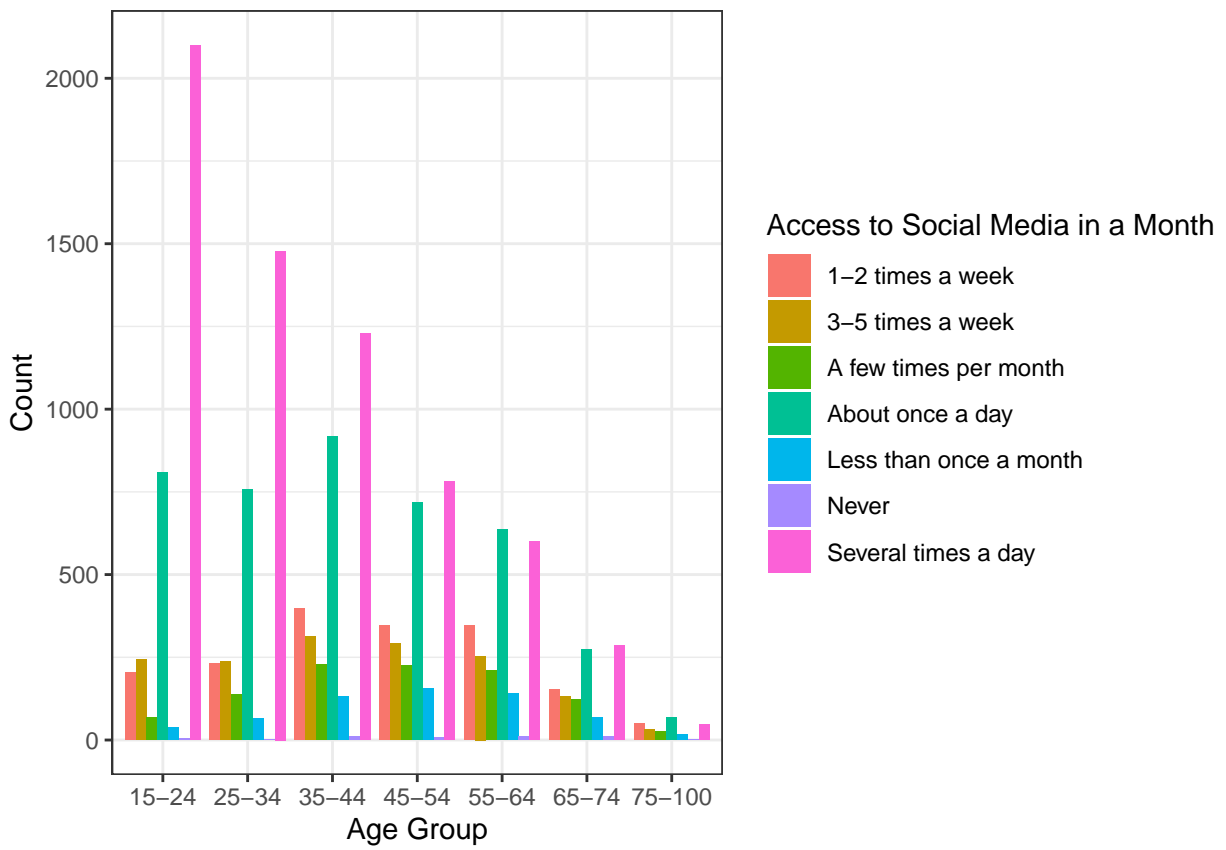


Figure 6: Social networking account - Frequency of access amongst Respondents

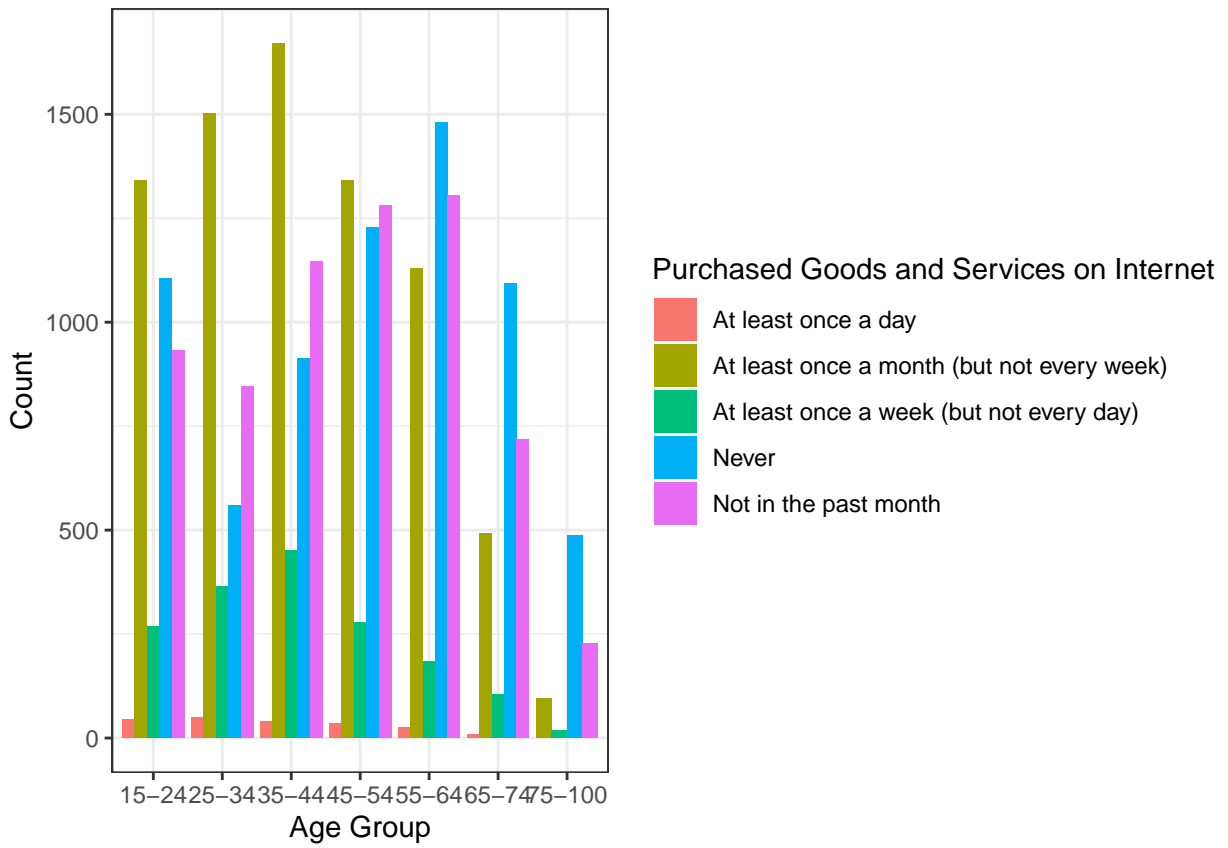


Figure 7: Respondent internet usage - buying goods and services

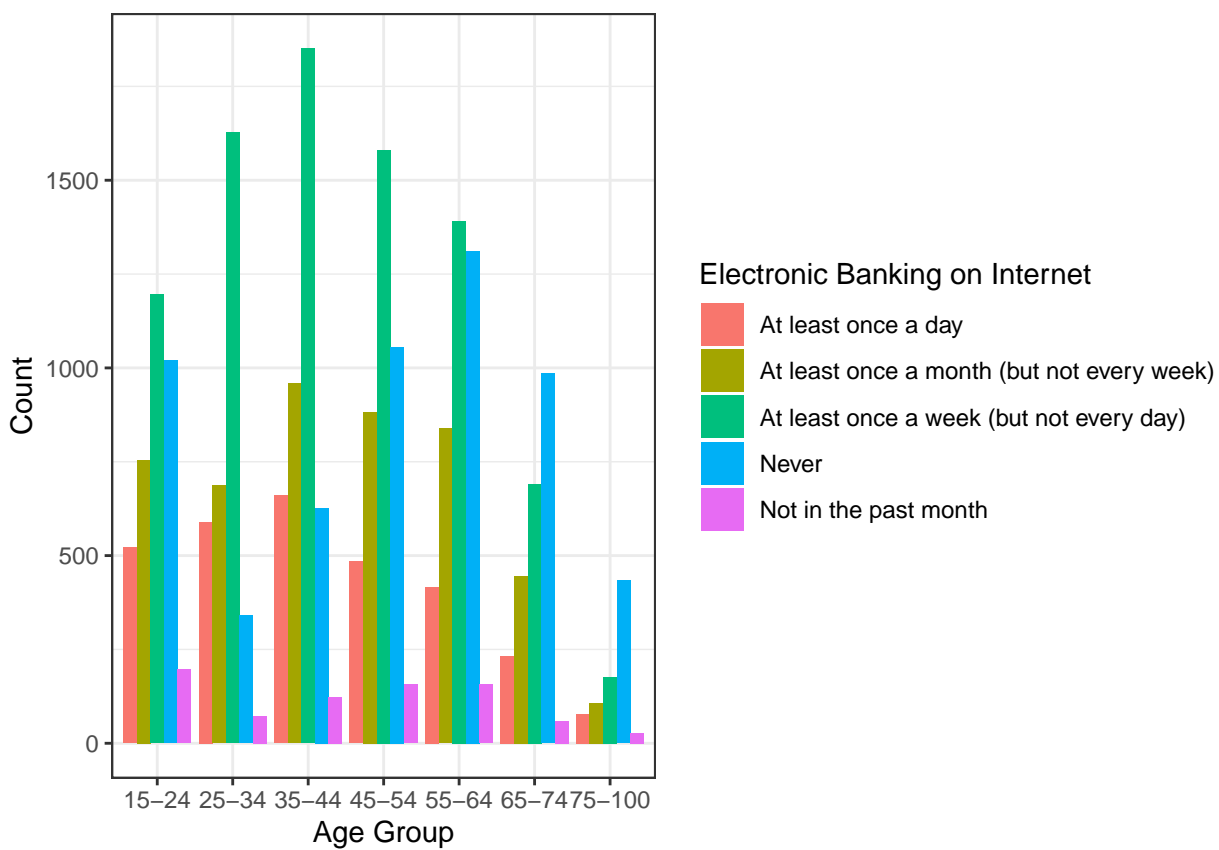


Figure 8: (#fig:elecbank_intuse)Respondent internet usage - Electronic Banking

$$P(Y = 1|X_k) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4)}$$

A binomial logistic regression is performed. Most variables with additional answers, such as “Not stated”, and “Refusal” are removed as these two variables created a broader category. Additional variables relating to the specific province were created for the model. The original variable stated the province was divided into categorical variables of provinces: Alberta, Manitoba, British Columbia, New Brunswick, Newfoundland and Labrador, Nova Scotia, Ontario, Prince Edward Islands, Quebec, and Saskatchewan. They contain 14,382 observations.

It is essential to note that all data models are imperfect, but are useful in ascertaining information. In this model, survey samples were not randomly collected thus due to the design, it does not account for causality and only accounts for correlation. This model operates on observational data, therefore it can leave a false level of confidence when drawing claims, due to which there exists human error and bias in the model and the results.

3 Results

3.1 Findings

The results give some insight about the respondents of the General Social Survey: Social Identify relationship between internet usage and voter participation. Using the package `gtsummary` Sjoberg et al. (2021) on the regression model, the results of the regression table are shown.

This table has the following findings. The usage of social media in terms of the respondent owning a social media account had a statistically significant relationship with voter participation. The tendency to buy goods and services online has a statistically significant relationship with voter participation but electronic usage of banking does not have a significant relationship with voter participation.

According to trends in society and research evidence, it is suggested that changes in communication technology play an important role in influencing electoral behavior (Tolbert and Mcneal (2003)). It has been well documented that socio-economic characteristics such as age, gender, ethnicity, and attitudinal factors affect political interest. There has been some evidence that suggests that changes in communication technology such as newspapers, television, and now social media play an important role in electoral behavior (Tolbert and Mcneal (2003)). Therefore, it does not come as a surprise that there exists a positive correlation between voter participation and owning a social media account.

The emergence of the internet and its capabilities in the twenty-first century comes with its own benefits, one of which is ease. It provides the ease of everything by having to press one button. It provides numerous entertainment opportunities to its users as well as reduced the process of acquiring information (Falck, Gold, and Heblich (n.d.)). Similarly, minute everyday tasks can now be done digitally. There exists a statistically significant positive relationship between the internet and the purchase of goods and services online, and a negative, statistically insignificant relationship between the internet and electronic banking usage. There exists no identification of the causal relationship between the internet and political behavior.

3.1.1 The Internet and Voting in Canada

Voter turnout in Canada has declined in the last several decades in Canada, now being the lowest in the western world (). According to the Library of Parliament of Canada, Canada’s youngest voter age group has turned out to be the lowest voter turnout rate amongst all age groups for federal elections and this disengagement from the electoral participation by Canadian youths has acted significantly on the downward drag of the turnout of the general population (Barnes (n.d.)). Canadian youngest eligible voters suggest that simply aging and turning into another age cohort would not increase their propensity to vote, as has been the case with the past generations. Therefore, the youngest non-voters would age, and replace old voters

who have shown a higher propensity to vote. The overall voter turnout in Canada has not only reached its all-time low but will further continue to drop (Barnes (n.d.)).

The emergence of the Internet as the new mass medium of the twenty-first century has changed the mass-media market substantially. Information can now be distributed at high speed, low cost, and more extensively, bringing more egalitarian access to the production and consumption of information (Falck, Gold, and Heblich (n.d.)). That being said, as shown in (Figure 5), there still exist several individuals in Canada with no access to a social media account. Nevertheless, social media accounts and the internet reduce a significant burden on humankind.

Earlier, according to a meta-analysis on the internet and voting, the results have been mixed (Tolbert and Mcneal (2003)). As per voting is concerned, several times in history has the internet made its ties with voting. In 2016, Barack Obama used the internet to keep in touch with his voters, which was the driving reason for his success in his presidential campaign. According to Milner, the main reason for a lower voter turnout may be due to a lack of information amongst voters, which can be mitigated by the internet, as the internet would help spread information and in turn increase voter turnout. (Figure 4) gives an overview of the reasons voters decided not to vote according to the subset of the data frame. Even though observers responded that the main reason may be the lack of interest, not being informed on issues was the second main concern leading to no voter turnout. It is essential to remember marginal groups such as Aboriginal people have not been included in the survey. Therefore, as ethnicity is an important consideration in voting, the lack of inclusion may also lead to non-cohesive information.

3.1.2 Future Direction

The role of social media and politics has continued to grow exponentially in recent times. Social media has a significant impact on politics in today's world. Internet and Voter participation may be related but the internet can help reduce the ramifications of a lower voter turnout. Even though the internet can not promise the certainty of a citizen voting, it can help encourage the citizen to vote by increasing voter education and political awareness. In a modern democracy such as Canada, social media and the internet can be used by the Government of Canada to involve citizens in the decision-making process of their society. Social Media is where most people consume their political information nowadays. The Government of Canada can use the internet to disseminate useful information to the youth and the active users of the internet in the form of voter education by forestalling incidences of election violence and electoral manipulations (Toba Paul (2019)). Social Media is the best avenue to reach out to the majority of the voters, especially the youth, who have the lowest turnout rate in Canada.

Motivation

1. *For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.*
 - Understanding the demographics of voting is essential for every country to include policies for a larger voter turnout. This paper understands the relationship between the characteristics of an individual such as age, gender, social media usage, internet usage, political engagement, trust in society, and discrimination to predict voting participation and to understand the relationship between them. These findings help implement policies to ensure better voter participation and also understand trends in voter turnout.
2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*
 - General Social Survey
3. *Any other comments?*
 - no

Composition

1. *What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*
 - Respondent information: age, education, marital status, residence, province, number of sexual partners. Internet usage and voting participation
2. *How many instances are there in total (of each type, if appropriate)?*
 - One dataframe containing several variables from the General Social Survey, 2013
3. *Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).*
 - It is a subset of the General Social Survey, 2013
4. *What data does each instance consist of? "Raw" data (for example, unprocessed text or images) or features? In either case, please provide a description.*
 - It contains a dataframe from questionnaires
5. *Does the dataset contain data that might be considered confidential. If so, please provide a description.*
 - There is no confidential data, the dataset is publicly available.
6. *Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.*
 - TBD
7. *Does the dataset identify any sub-populations (for example, by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.*
 - The dataset comprises men and women aged 15 years or above.
8. *Is it possible to identify individuals (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset? If so, please describe how.* -No
9. *Does the dataset contain data that might be considered sensitive in any way (for example, data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.*
 - Sensitive information may include but is not limited to: Number of sexual partners for men and women before 18, The Percent of women aged 15-19 and 20-24 married before 18. Percent of men aged 15-19 and 20-24 married before 18.

Collection process

1. *How was the data associated with each instance acquired? Was the data directly observable (for*

example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/derived from other data (for example, part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.

- The data was taken from the General Social Survey, 2013.
2. *What mechanisms or procedures were used to collect the data (for example, hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?*
 - Household questionnaires, in-person interviews, and other types of model questionnaires were employed.
 3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?*
 - The enumeration areas are drawn from census files, and samples of households were selected from the areas.
 4. *Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)?*
 - TBD
 5. *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (for example, recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.*
 - TBD
 6. *Were any ethical review processes conducted (for example, by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.*
 - TBD
 7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?*
 - TBD
 8. *Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.*
 - TBD
 9. *Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.*
 - TBD
 10. *If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).*
 - TBD
 11. *Has an analysis of the potential impact of the dataset and its use on data subjects (for example, a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.*
 - TBD
 12. *Any other comments?*
 - TBD

Preprocessing/ cleaning/ labeling

1. *Was any preprocessing/cleaning/labeling of the data done*
 - the data was originally obtained from CHASS. Then into a usable data frame in R using the libraries at r
2. *Is the software that was used to preprocess/clean/label the data available?*

- R software is available at <https://www.R-project.org/>

Uses 1. *Has the dataset been used for any tasks already?* -The dataset has not been used for other tasks yet. 2. *Is there a repository that links to any or all papers or systems that use the dataset?* - 3. *What (other) tasks could the dataset be used for?* The dataset would be used for understanding behavior to voting amongst different socio-economic demographic 4. *Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?* - The cleaning process is very specific to the dataset required but it can be collected.

Distribution

1. *Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created?* -No. The data set is for personal use only.
2. *How will the dataset will be distributed (e.g., tarball on website, API, GitHub)?* -The dataset can be found in the Github repository mentioned in the report. 3. *When will the dataset be distributed?* -The dataset has already been distributed in April 2022. 4. *Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)?* -The report is under the MIT license.
3. *Have any third parties imposed IP-based or other restrictions on the data associated with the instances?*
 - There are no IP-based or other restrictions on the data. 6. *Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?* -No.
7. *Any other comments?* -None.

Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?* -The of the report will be maintaining the dataset. The author is Mahak Jain
2. *How can the owner/curator/manager of the dataset be contacted (e.g., email address)?* -They can be contacted via the email addresses listed in the README file on Github.
3. *Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)?* -No dataset updates are scheduled yet.
4. *If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were the individuals in question told that their data would be retained for a fixed period of time and then deleted)?* -The data set was collected by the GSS. There are no application limits on the retention of the data since it was collected voluntarily.
5. *Will older versions of the dataset continue to be supported/hosted/maintained?* -No. Only the current data is available for now and future.
6. *If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?* -The extensions, augmentations and contributions from other users are not supported.

References

- Barnes, Andrea. n.d. "Youth Voter Turnout in Canada : Trends and Issues." <https://lop.parl.ca/content/lop/TeachersInstitute/YouthVoterTurnoutinCanada.pdf>.
- Falck, Oliver, Robert Gold, and Stephan Heblich. n.d. "E-Lections: Voting Behavior and the Internet." *American Economic Review*. <https://www.aeaweb.org/articles?id=10.1257%2Faer.104.7.2238>.
- Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Hirzalla, Fadi, Liesbet van Zoonen, and Jan de Ridder. 2011. "Internet Use and Political Participation: Reflections on the Mobilization/Normalization Controversy." *The Information Society* 27 (1): 1–15. <https://doi.org/10.1080/01972243.2011.534360>.
- Kushner, Joseph, and David Siegel. 2006. "Why Do Municipal Electors Not Vote?" *Canadian Journal of Urban Research* 15 (2): 264–77. <http://www.jstor.org/stable/26192462>.

- Mcneal, Ramona S, and Caroline J Tolbert. 2003. *Unraveling the Effects of the Internet on Political Participation?* <https://journals.sagepub.com/doi/abs/10.1177/106591290305600206?source=mfc&rs=s=1>.
- Pebesma, Edzer. 2018. “Simple Features for R: Standardized Support for Spatial Vector Data.” *The R Journal* 10 (1): 439–46. <https://doi.org/10.32614/RJ-2018-009>.
- Pedersen, Thomas Lin. 2020. *Patchwork: The Composer of Plots*. <https://CRAN.R-project.org/package=patchwork>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Sjoberg, Daniel D., Karissa Whiting, Michael Curry, Jessica A. Lavery, and Joseph Larmarange. 2021. “Reproducible Summary Tables with the Gtsummary Package.” *The R Journal* 13 (1): 570–80. <https://doi.org/10.32614/RJ-2021-053>.
- Toba Paul, Ayeni. 2019. “The Role of Social Media in Voter Education in Nigeria.” *Scientific Journal of Research & Reviews* 1 (3). <https://doi.org/10.33552/sjrr.2019.01.000511>.
- Tolbert, Caroline J, and Ramona S Mcneal. 2003. “Unraveling the Effects of the Internet on Political Participation?” <https://journals.sagepub.com/doi/full/10.1177/106591290305600206>.
- 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. 2022. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Wickham, Hadley, and Maximilian Girlich. 2022. *Tidyr: Tidy Messy Data*. <https://CRAN.R-project.org/package=tidyr>.
- Wickham, Hadley, and Evan Miller. 2021. *Haven: Import and Export 'Spss', 'Stata' and 'Sas' Files*. <https://CRAN.R-project.org/package=haven>.