

The Impact of Socio-Economic and Demographic Factors on Voter Turnout in Toronto Wards*

An Analysis of the 2022 Toronto Municipal Election

Janel Gilani

April 15, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

Table of contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 1 |
| 2 | Data | 2 |
| 2.1 | Measurement | 2 |
| 2.2 | Ward Profiles, 2018 (25-Ward Model) | 3 |
| 2.3 | Election Voter Statistics | 4 |
| 2.4 | Voter Turnout and Ward's Socio-Economic and Demographic Factors | 6 |
| 2.5 | Relationship between Voter Turnout and Socio-Economic Factors | 8 |
| 2.6 | Relationship between Voter Turnout and Demographic Factors | 10 |
| 3 | Model | 11 |
| 3.1 | Model set-up | 12 |
| 3.1.1 | Socio-Economic models | 13 |
| 3.1.2 | Demographic models | 13 |
| 3.2 | Model justification | 14 |
| 4 | Results | 15 |
| 4.1 | Socio-Economic Models | 15 |
| 4.2 | Demographic Models | 15 |

*Code and data are available at: <https://github.com/JanelGilani/toronto-voter-turnout>

| | | |
|----------|--------------------------------------|-----------|
| 5 | Discussion | 15 |
| 5.1 | First discussion point | 15 |
| 5.2 | Second discussion point | 18 |
| 5.3 | Third discussion point | 18 |
| 5.4 | Weaknesses and next steps | 18 |
| | Appendix | 19 |
| A | Datasheet | 19 |
| B | Model Details | 19 |
| B.1 | Posterior Predictive Check | 19 |
| B.2 | Diagnostics | 20 |
| | References | 21 |

1 Introduction

Voting for the mayor of your ward is one of the most important decisions a Canadian citizen can make because it decides many of the decisions that will be made throughout that four year term as well as how their life will change in that period of time. This is why it is important to understand the many factors that influence the voter turnout and thus, ultimately affect the integrity and representativeness of the electoral process. As mandated by the Government of Ontario, residents of the City of Toronto went to the polls on October 24, 2022 to elect a mayor, councillors, and school board trustees. Then-sitting Mayor, John Tory, sought re-election, along with a number of incumbent city councillors. Coming out of the height of the COVID-19 pandemic, the 2022 election largely upheld the status quo and did not feature ambitious policy platforms. Voter turnout across the city fell to 30% - the lowest in the city's history since amalgamation in 1997, with turnout ranging on a ward-by-ward basis from 22% to 38% (Marshall 2023; Warren 2022).

While there is existing literature on most of these factors in isolation, there has been little research that has examined the combined effect of these explanatory factors on voter turnout, especially in Canadian cities like Toronto and this paper will contribute to the investigation of this phenomenon. Using the 2022 Toronto Municipal Election and Ward Profiles data from Open Data Toronto, this paper will be examining many factors that could have influenced voter turnout in the 2022 Toronto Municipal Election such as education, income, unemployment rate, population, and number of subdivisions within a ward. Our main focus is to conduct analysis on voter turnout as the response variable, and to see if our estimand, the effect of ward's socio-economic and demographic factors on the ward's voter turnout, has influenced voter turnout in the elections. We found that socio-economic factors such as income, unemployment rate, and education level have a significant impact on voter turnout, while demographic factors

such as population and number of subdivisions have a less significant impact. Using these understandings, we hope to highlight the potential implications these correlations may have on upcoming elections and the electoral process in Toronto.

Our paper begins with the Data section (Section 2) to visualize and further understand the measurement, source, methodology, and variables we are examining. Then, we introduce the Model (Section 3) used to understand the relationships in the data and report the findings in the Results section (Section 4). Finally, we include the Discussion (Section 5) of the findings, summarizing the takeaway and future of this research.

2 Data

Data analysis is performed in R (R Core Team 2020), and additional help is provided by libraries such as `dplyr` (`dplyr?`), `ggplot2` (`ggplot?`), `ggrepel` (`ggrepel?`), `tidyverse` (`thereferencecanbewhatever?`), `kableExtra` (`kableextra?`), `knitr` (`knitr?`), and `sf` (`pebesma2023spatial?`), `opendatatoronto` (`toronto?`), `readxl` (`excel?`), `here` (`here?`), `rstanarm` (`rstan?`), `arrow` (`arrow?`), `tidybayes` (`tidybayes?`), `modelsummary` (`modelsummary?`), `broom` (`broom?`), and `parameters` (`parameters?`). Data for this research comes from Open Data Toronto (`toronto?`), an open source data portal containing various topics of data for the city. For the data involved in this paper, we combine Elections - Voter Statistics (`PublicArt?`) and Ward Profiles (25-Ward Model) (`WardProfile?`)

2.1 Measurement

Our research question and estimand is divided into two parts: impact of socio-economic factors and demographic factors on voter turnout, which is our response variable. Starting off with socio-economic factors, we use education, income, and unemployment rate as the explanatory variables to represent these factors. Education is a key factor in determining voter turnout as it is associated with political knowledge, interest, and participation (`Verba?`). To measure education, we use the percentage of citizens with no certificate, diploma, or degree as an indicator of education level. Income and unemployment are also important factor in determining voter turnout as they are associated with financial security, political interest, and participation. Household income is not the sole indicator of wealth, but it is heavily related to wealth as both income and wealth are key indicators of financial security (`Schaeffer?`). We use the unemployment rate as an indicator of economic stability.

Moving on to demographic factors, we use population and number of subdivisions as the explanatory variables to represent these factors. Individual turnout by ward depended on a number of factors, including the accessibility of polling locations within subdivisions to cast a ballot. The City of Toronto (`ElectionDictionary?`) defines a ward as “a geographical area represented by a member of Council.” Subdivisions are defined by the City of Toronto

(**VoterStatistics?**) as “.... geographical area[s within a ward] designated by the City Clerk.” Previous studies concentrating on large cities in the United States, including Atlanta revealed that having polling locations in close proximity to a voter’s home bolsters turnout and even minor changes in placement of a polling location can have significant impact on a voter’s decision to cast a ballot (**Locationx3?**). Thus, we use the number of subdivisions as an indicator of accessibility to polling locations. Lastly, we use population as an indicator of the demographic size of the ward.

Election data is formed by collecting the relevant numbers observed during the phenomenon/process of voting, and ward profiles data is extracted from the 2022 Census data which is collected through a variety of methods such as self-reporting surveys, door-to-door enumeration, online/telephone surveys, administrative records, and sampling techniques.

2.2 Ward Profiles, 2018 (25-Ward Model)

Ward profiles such as income and population would be interesting factors to analyze alongside eligible voter turnouts. Therefore, the dataset for ward profiles (**torontoWardProfiles?**) based on 2022 census data has been included in analysis as well. This dataset is published by City Planning, and was last updated on January 3, 2024. The 25-ward model was chosen over the 47-ward model because it is matching the ward classification in the latest elections data. This dataset contains demographic, social, and economic information for each ward such as population, households, families, education, ethnocultural composition, spoken languages, income and housing costs. For our purpose of research, we are interested in population, number of citizens with no certificate, diploma, or degree, unemployment rate, and average household income of each ward. A sample of the cleaned dataset for the wards is shown below in **?@tbl-cleaned_ward**.

Ward profile data is stored in an Excel file with multiple tabs. The relevant data to be used for this paper’s analysis is included the first tab, **2022 Census One Variable**. As such, only data for this tab was downloaded for analysis. Further data cleaning was performed to transpose the data and only keeping information relevant to our research question (see **?@tbl-cleaned_ward_profile_data**).

The variables selected for analysis were:

- **Ward ID:** The unique identifier for each ward.
- **Population:** The total number of people living in the ward.
- **Uneducated Population (%):** The percentage of the population with no certificate, diploma, or degree. This was calculated by dividing the number of uneducated people by the total population and multiplying by 100.
- **Unemployment Rate (%):** The percentage of the population in labour market that is unemployed.
- **Income:** The average household income in the ward.

Table 1: Sample of Cleaned Toronto Ward Profile Data

| Ward ID | Population | Uneducated Population (%) | Unemployment Rate (%) | Income |
|---------|------------|---------------------------|-----------------------|--------|
| 1 | 115120 | 18.997568 | 16.5 | 95200 |
| 2 | 117200 | 11.053754 | 12.8 | 146600 |
| 3 | 139920 | 9.269583 | 11.8 | 127200 |
| 4 | 104715 | 9.072244 | 12.9 | 127200 |
| 5 | 115675 | 21.750594 | 16.4 | 88700 |

2.3 Election Voter Statistics

This dataset, published by the City Clerk’s Office (**VoterStatistics?**) outlines voter statistics on a ward-by-ward and entire city basis for the 2022 municipal election. For each subdivision within a ward, the data set shows the polling location name and address, number of additions and corrections to the voter’s list, number of eligible electors and number who voted, and rejected and declined ballots. This data set was last refreshed on February 7, 2023.

Upon analysis, columns such as rejected and declined ballots were deemed beyond the scope of this paper. We conducted the first step of basic data cleaning to eliminate additional columns and simplify the names of retained columns (see `?@tbl-cleaned_voter_statistics`).

The variables selected for analysis were:

- **Ward ID:** The unique identifier for each ward.
- **Eligible Voter Turnout (%):** The percentage of eligible voters who cast a ballot.
- **Number of Voters:** The total number of voters who cast a ballot.
- **Number of Subdivisions:** The number of subdivisions within a ward.

Table 2: Sample of Cleaned Elections Data

| Ward ID | Eligible Voter Turnout (%) | Number of Voters | Number of Subdivisions |
|---------|----------------------------|------------------|------------------------|
| 1 | 24 | 16962 | 55 |
| 2 | 30 | 26784 | 72 |
| 3 | 33 | 33544 | 86 |
| 4 | 38 | 32223 | 70 |
| 5 | 28 | 21807 | 65 |

To better visualize the voter turnout across Toronto wards, we have created a bar plot (see Figure 1) that displays the percentage of eligible voters who cast a ballot in each ward. The map of Toronto (see Figure 2) highlights the voter turnout across wards, providing a spatial

representation of the data. From Figure 1, we can observe that wards 4, 14, and 19 have the highest voter turnout, while wards 7, 1, and 10 have the lowest voter turnout. This trend confirms the intuition that more affluent and demographically developed areas such as Downtown Toronto have higher voter turnout. The map in Figure 2 further illustrates this trend, with darker shades in southern Toronto indicating higher voter turnout.

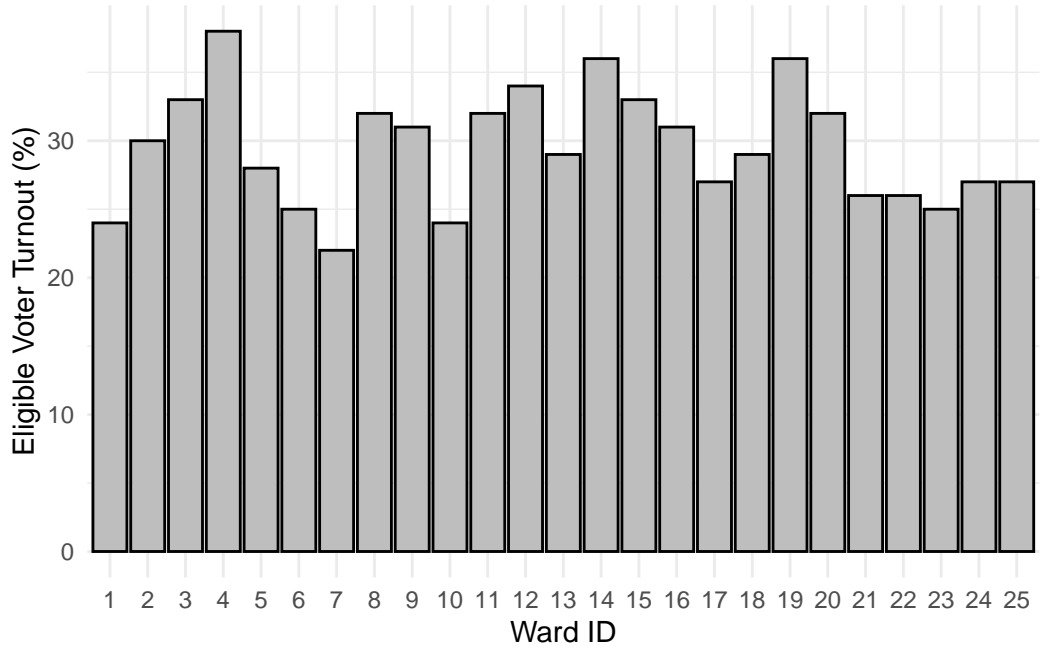


Figure 1: Voter Turnout (%) across Toronto Wards

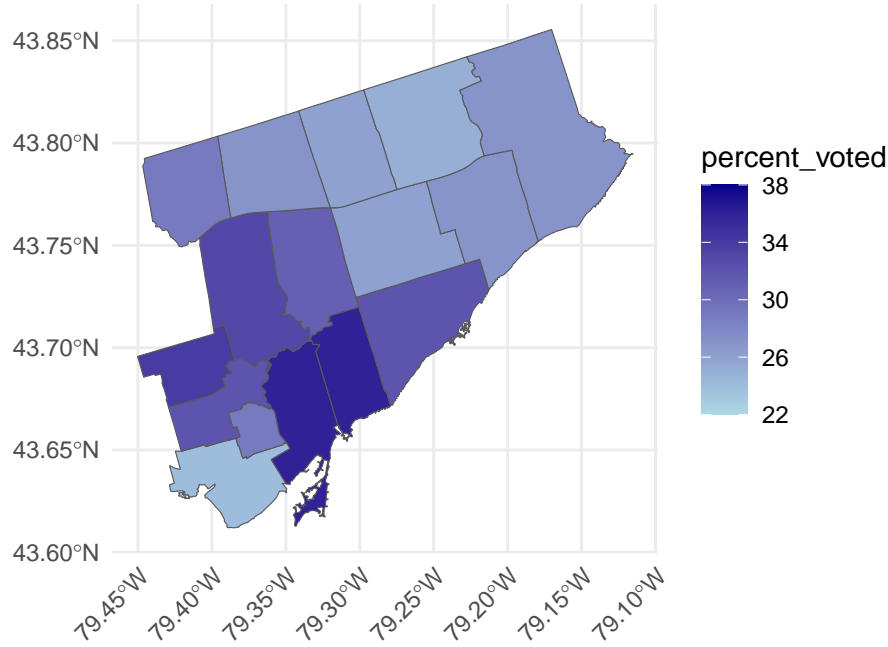


Figure 2: Map of Toronto highlighting the voter turnout across wards

2.4 Voter Turnout and Ward's Socio-Economic and Demographic Factors

As the goal of this research is to analyze the impact of socio-economic and demographic factors on voter turnout in the 2022 Toronto Municipal Election, we have combined the cleaned election data and ward profile data to create an analysis dataset. The analysis dataset includes the following variables: ward ID, ward name, population, number of subdivisions, percent of uneducated population, unemployment rate, income, voter turnout percentage, and number of voters. Below in Table 3 is a sample of the analysis data.

Table 3: Sample of Combined Ward Election, Income, Employment, and Education Data

| Ward ID | Ward Name | Pop. | Num. of SubDiv | Uneducated Pop. (%) | Unemployment Rate (%) | Income | Voter Turnout | Num. of Voters |
|---------|---------------------|--------|----------------|---------------------|-----------------------|--------|---------------|----------------|
| | | | | | | | (%) | |
| 1 | Etobicoke North | 115120 | 55 | 18.997568 | 16.5 | 95200 | 24 | 16962 |
| 2 | Etobicoke Centre | 117200 | 72 | 11.053754 | 12.8 | 146600 | 30 | 26784 |
| 3 | Etobicoke-Lakeshore | 139920 | 86 | 9.269583 | 11.8 | 127200 | 33 | 33544 |

Table 3: Sample of Combined Ward Election, Income, Employment, and Education Data

| Ward ID | Name | Pop. | Num. of SubDiv | Uneducated Pop. (%) | Unemployment Rate (%) | Income | Voter Turnout (%) | Num. of Voters |
|---------|--------------------|--------|----------------|---------------------|-----------------------|--------|-------------------|----------------|
| 4 | Parkdale-High Park | 104715 | 70 | 9.072244 | 12.9 | 127200 | 38 | 32223 |
| 5 | York South-Weston | 115675 | 65 | 21.750594 | 16.4 | 88700 | 28 | 21807 |

There are 25 wards in the City of Toronto, each with unique socio-economic and demographic characteristics as summarised in Table 4. Based on 2021 census data, the average population per ward is 110452, with standard deviation of 10594. The wards with the highest population are wards 3, 10 and 2. The wards with the lowest population are wards 23, 16 and 11. The average income per ward is \$120096, with standard deviation of \$33980.64 The wards with the highest income are wards 15, 8 and 11. The wards with the lowest income are wards 7, 5 and 13.

Moving on to unemployment rate, the average unemployment rate per ward is 14.13%, with standard deviation of 2.11%. The wards with the highest unemployment rate are wards 7, 23 and 1. The wards with the lowest unemployment rate are wards 10, 12 and 3 The average percentage of uneducated population per ward is 12.43%, with standard deviation of 4.93%. The wards with the highest percentage of uneducated population are wards 7, 5 and 1. The wards with the lowest percentage of uneducated population are wards 10, 11 and 13.

Voter turnout across the entire city during the 2022 municipal election fell to 30%, with Ward 4 (Parkdale-High Park) having the highest turnout at 38% and Ward 7 (Humber River-Black Creek) having the lowest at 22%. Ward 19 (Beaches-East York) and Ward 14 (Toronto Danforth) were tied for second highest turnout at 36% each. The third highest turnout at 34% was in Ward 12 (Toronto-St. Paul’s). The wards with the second lowest turnout were Ward 1 (Etobicoke North) and Ward 10 (Spadina-Fork York) at 24% each. Ward 6 (York Centre) and Ward 23 (Scarborough North) with 25% turnout respectively were tied for third lowest turnout.

During the 2022 municipal election, there were 1,535 subdivisions across Toronto’s 25 wards. Ward 10 (Spadina-Fort York) had the highest number of subdivisions at 94, while Ward 23 (Scarborough North) had the lowest number at 38. Ward 13 (Toronto Centre) had the second highest number of subdivisions at 90, with Ward 3 (Etobicoke-Lakeshore) coming in third highest with 83 subdivisions. Ward 7 (Humber River-Black Creek) had the second lowest number of subdivisions at 47, following by Ward 25 (Scarborough-Rouge Park) with 48.

Table 4: Summary Statistics

| Variable | Mean | Median | Std. Dev | Min | Max |
|---------------------------|-----------|-----------|----------|----------|-----------|
| Income | 120096.00 | 107300.00 | 33980.64 | 85700.00 | 224800.00 |
| Number of Subdivisions | 61.40 | 59.00 | 13.79 | 38.00 | 94.00 |
| Number of Voters | 22524.96 | 22522.00 | 5168.51 | 14616.00 | 33544.00 |
| Population | 110451.60 | 110095.00 | 10593.87 | 94025.00 | 139920.00 |
| Voter Turnout (%) | 29.48 | 29.00 | 4.22 | 22.00 | 38.00 |
| Uneducated Population (%) | 12.43 | 11.95 | 4.93 | 4.84 | 23.08 |
| Unemployment Rate (%) | 14.13 | 14.10 | 2.11 | 9.80 | 17.80 |

2.5 Relationship between Voter Turnout and Socio-Economic Factors

As one of our variables of interest, we are determined to examine the relationship between voter turnout and the ward's income. We expect income and voter turnout to be positively related because Higher-income individuals often have better access to resources, education, and stability, which can positively correlate with higher levels of political engagement and voter turnout. To visualize the relationship of interest, we plot ward's income with the voter turnout in Figure 3.

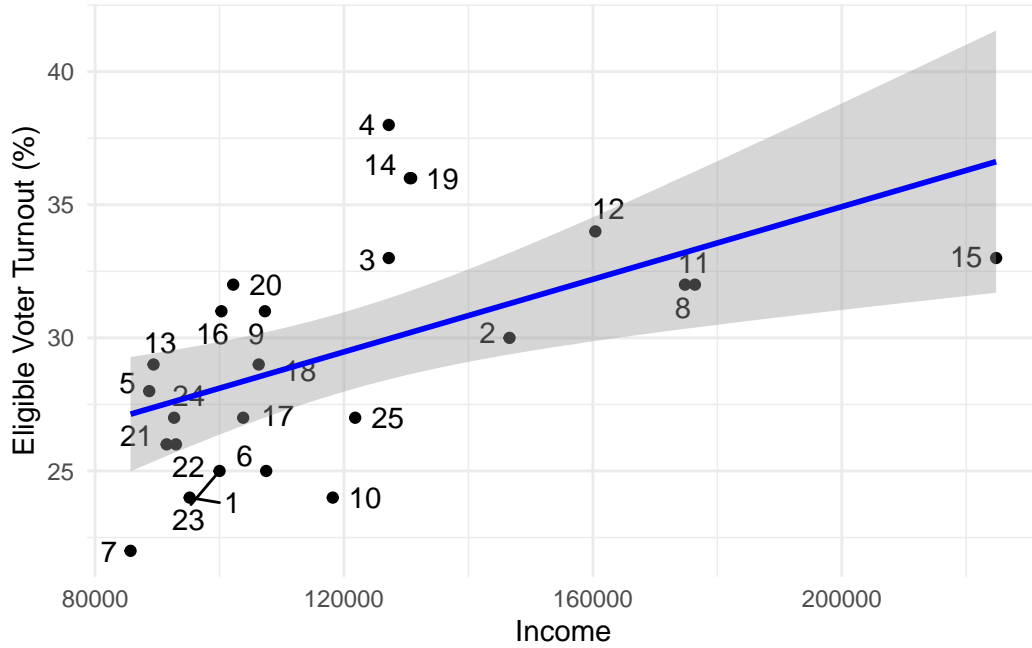


Figure 3: Correlation between Eligible Voter Turnout and Ward's Income

As expected, we see a moderate positive relationship between income of a ward and the voter turnout. Furthermore, Ward 4 is an outlier that has relatively low income and yet a high voter turnout. Intuitively, this aligns with our beliefs and confirms the trend between high income with higher voter turnout.

Another explanatory variable we are interested in is the employment rate of a region. In this case, we visualize the relationship between employment rate and the nvoter turnout per ward. We expect to see a negative relationship between these variables. Figure 4 displays the visualization.

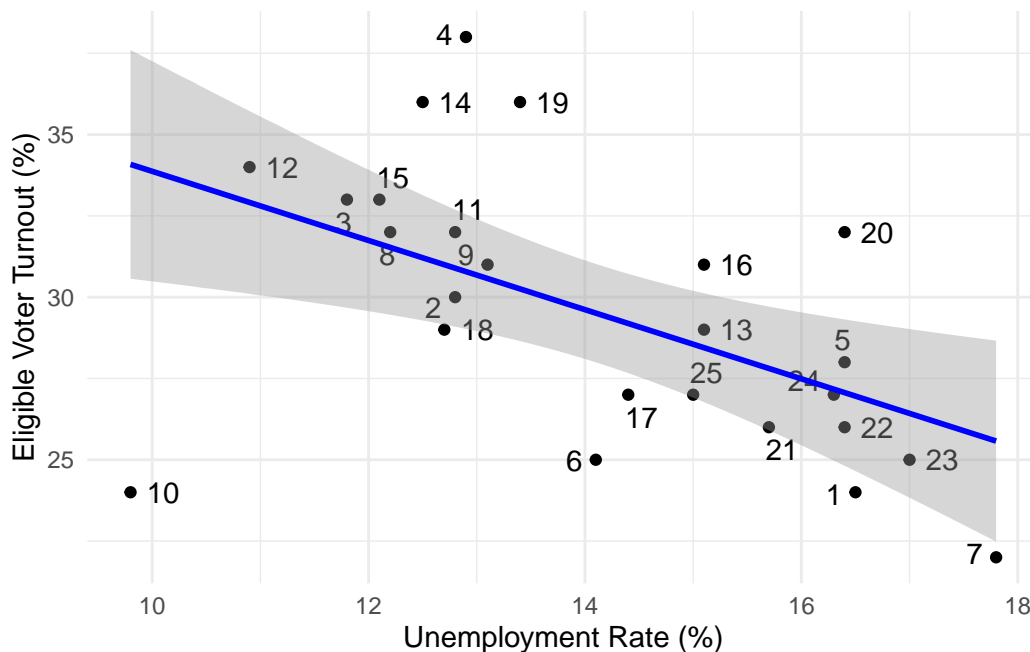


Figure 4: Correlation between Eligible Voter Turnout and Ward's Unemployment Rate

Through the plot, we see that there is a moderately negative relationship between the unemployment rate of a ward and the voter turnout. This is in line with our expectations as higher levels of employment often correlate with increased stability and engagement within a community. Employed individuals may feel more connected to their local area and thus be more likely to participate in civic activities such as voting. Ward 10 is an outlier that has a low unemployment rate and yet a low voter turnout. This observation leaves room for further research and investigation into the history and background behind Ward 10's voter turnout.

Lastly, we examine the relationship between the percentage of uneducated population in a ward and the voter turnout. We expect to see a negative relationship between these variables. Figure 5 displays the visualization.

The plot in Figure 5 shows a negative relationship between the percentage of uneducated

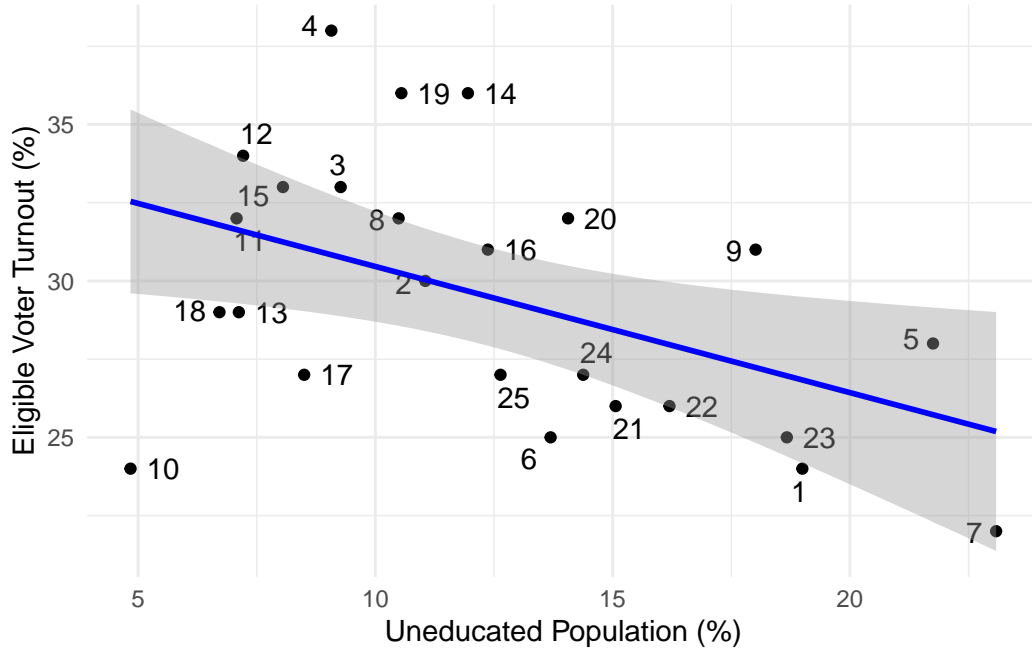


Figure 5: Correlation between Eligible Voter Turnout and Ward’s Level of Education

population in a ward and the voter turnout. This also aligns with our expectations as higher education levels typically correlate with greater awareness of civic duties, including voting, and a stronger sense of civic responsibility, which can lead to higher voter turnout. Ward 10 is again an outlier that has a low percentage of uneducated population and yet a low voter turnout.

2.6 Relationship between Voter Turnout and Demographic Factors

In addition to socio-economic factors, we are also interested in examining the relationship between voter turnout and demographic factors such as population and number of subdivisions. We expect to see a positive relationship between population and voter turnout as higher population density in a ward may indicate greater community engagement and political activity, leading to higher voter turnout. Additionally, densely populated areas often have more resources and infrastructure, making it easier for residents to access polling stations and participate in elections. To visualize this relationship, we plot the ward’s population with the voter turnout in Figure 6.

Through the plot, we see that there is no relationship between the population of a ward and the voter turnout. It is possible that the population of a ward is not a reflection of its full demographic nature. Thus, we would not need to worry about population acting as a confounding factor when creating our demographic models.

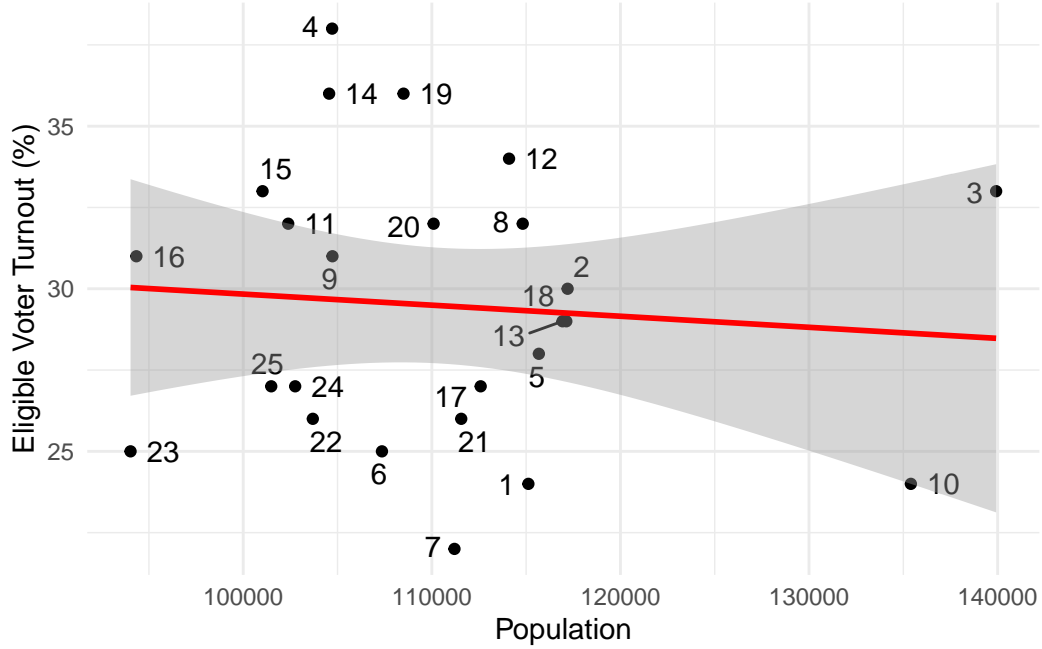


Figure 6: Correlation between Eligible Voter Turnout and Ward's Population

Next, we examine the relationship between the number of subdivisions in a ward and the voter turnout. We expect to see a positive relationship between these variables as a higher number of subdivisions may indicate greater accessibility to polling stations and increased voter turnout. Figure 7 displays the visualization.

Even though we see a slight positive correlation, the data reveals that having more subdivisions within wards does not automatically correlate to higher voter turnout, as illustrated by Figure 7. Ward 4 (Parkdale-High Park) which had the highest voter turnout at 38% has 70 subdivisions within the ward, which falls approximately in the middle of the number of subdivisions across all 25 wards. Ward 19 (Beaches-East York) and Ward 14 (Toronto Danforth) which both saw the second highest voter turnout at 36% had 57 and 52 subdivisions respectively, once again highlighting that more subdivisions does not necessarily correlate to higher voter turnout.

3 Model

Here we briefly describe the Bayesian analysis model used to investigate the relationship between the socio-economic and demographic factors and voter turnout. In particular, we divide our investigation into two parts: socio-economic models and demographic models. For each, we consider two types of models: multiple linear regression and Poisson regression models. We

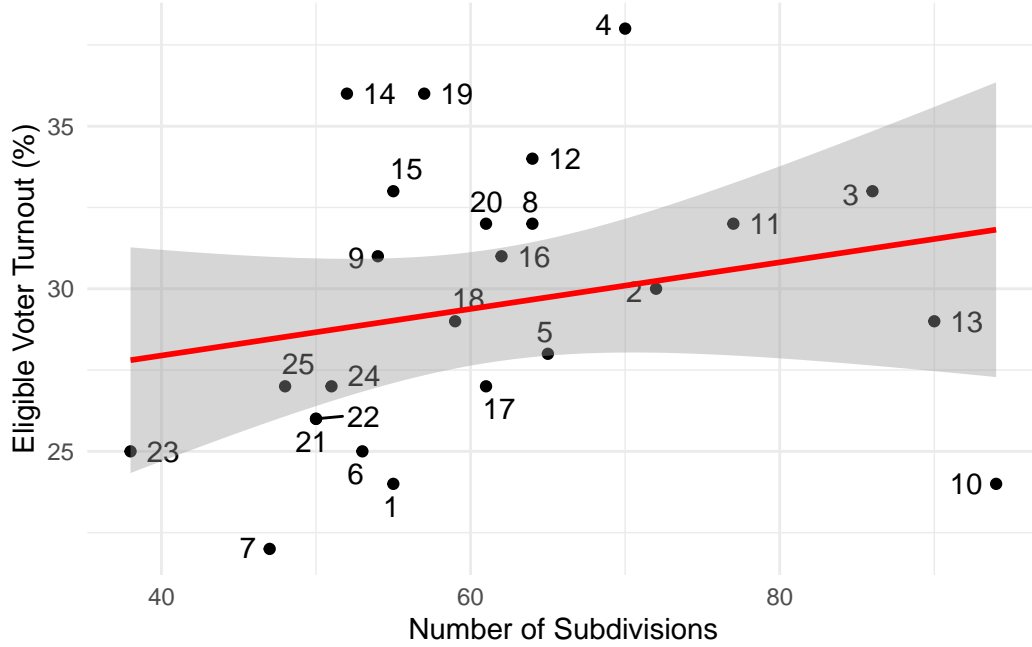


Figure 7: Correlation between Eligible Voter Turnout and Ward’s Number of Subdivisions

also further tested our results from Poission regression by employing negative binomial regression models to account for unequal mean and variance in the data (The negative binomial regression results are not displayed here but the models can be found in relevant files in the `models` folder).

Background details and diagnostics are included in Appendix [B](#).

3.1 Model set-up

From the Data section, we observed the relationships between voter turnout and socio-economic and demographic factors in isolation. Now, we aim to understand the relationships in a more comprehensive manner by adjusting for all factors simultaneously using the following models. We create and run the generalised linear models in R (R Core Team 2020) using the `rstanarm` package of (`rstan?`). Initially, we use the default priors from `rstanarm`, however, we allow `rstanarm` to improve the priors by scaling them based on the data. We allow auto-scaling and run both models with the updated priors specified above.

3.1.1 Socio-Economic models

Table 5 shows a model summary for socio-economic models, this will be discussed further in the results section.

3.1.1.1 Poisson Regression

Define y_i as the number of voters in the ward i . Then $income_i$ is the income of ward i , $uneducated_i$ is the uneducated population percentage of ward i , and $unemployment_i$ is the unemployment rate of ward i .

$$y_i | \lambda_i \sim \text{Poisson}(\lambda_i) \quad (1)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 \times uneducated_i + \beta_2 \times income_i + \beta_3 \times unemployment_i \quad (2)$$

$$\beta_0 \sim \text{Normal}(0, 2.5) \quad (3)$$

$$\beta_1 \sim \text{Normal}(0, 0.508) \quad (4)$$

$$\beta_2 \sim \text{Normal}(0, 0.074) \quad (5)$$

$$\beta_3 \sim \text{Normal}(0, 1.187) \quad (6)$$

3.1.1.2 Multiple Linear Regression

Define y_i as the voter turnout percentage in the ward i . Then $income_i$ is the income of ward i , $uneducated_i$ is the uneducated population percentage of ward i , and $unemployment_i$ is the unemployment rate of ward i .

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \quad (7)$$

$$\mu_i = \beta_0 + \beta_1 \times uneducated_i + \beta_2 \times income_i + \beta_3 \times unemployment_i \quad (8)$$

$$\beta_0 \sim \text{Normal}(0, 11) \quad (9)$$

$$\beta_1 \sim \text{Normal}(0, 2.14) \quad (10)$$

$$\beta_2 \sim \text{Normal}(0, 0.31) \quad (11)$$

$$\beta_3 \sim \text{Normal}(0, 5.01) \quad (12)$$

$$\sigma \sim \text{Exponential}(0.24) \quad (13)$$

3.1.2 Demographic models

Table 6 shows a model summary for demographic models, this will be discussed further in the results section.

3.1.2.1 Poisson Regression

Define y_i as the number of voters in the ward i . Then $population_i$ is the population of ward i , and $numsub_i$ is the number of subdivisions in ward i .

$$y_i | \lambda_i \sim \text{Poisson}(\lambda_i) \quad (14)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 \times numsub_i + \beta_2 \times population_i \quad (15)$$

$$\beta_0 \sim \text{Normal}(0, 2.5) \quad (16)$$

$$\beta_1 \sim \text{Normal}(0, 0.18125) \quad (17)$$

$$\beta_2 \sim \text{Normal}(0, 0.00024) \quad (18)$$

3.1.2.2 Multiple Linear Regression

Define y_i as the voter turnout percentage in the ward i . Then $population_i$ is the population of ward i , and $numsub_i$ is the number of subdivisions in ward i .

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \quad (19)$$

$$\mu_i = \beta_0 + \beta_1 \times numsub_i + \beta_2 \times population_i \quad (20)$$

$$\beta_0 \sim \text{Normal}(0, 11) \quad (21)$$

$$\beta_1 \sim \text{Normal}(0, 0.766) \quad (22)$$

$$\beta_2 \sim \text{Normal}(0, 0.0001) \quad (23)$$

$$\sigma \sim \text{Exponential}(0.24) \quad (24)$$

3.2 Model justification

We chose to use Bayesian analysis to investigate the relationship between socio-economic and demographic factors and voter turnout because it allows us to incorporate prior knowledge and uncertainty into our models. By using Bayesian analysis, we can estimate the posterior distribution of the parameters of interest, which provides a more comprehensive understanding of the relationships between the variables. Additionally, Bayesian analysis allows us to quantify the confidence levels in the estimates of our coefficients and make probabilistic statements about the relationships between the variables.

For our first model, we used a Poisson regression model to investigate the relationship between voter turnout and socio-economic factors. We chose this model because it is appropriate for count data, such as the number of voters, and allows us to model the relationship between the explanatory variables and the response variable. We also chose to use a multiple linear

regression model to investigate the relationship between voter turnout and socio-economic factors. We included income, unemployment rate, and the percentage of uneducated population as explanatory variables in the model. We chose this model because it allows us to examine the relationships between multiple variables and voter turnout simultaneously.

We then repeated the analysis for demographic factors, using population and the number of subdivisions as explanatory variables. We used the same models as for the socio-economic factors to investigate the relationships between these variables and voter turnout. We chose these models because they allow us to examine the relationships between the variables and voter turnout while controlling for other factors.

From Appendix Figure 8a, we see that the Poisson regression model is not a good fit for the observed data. To improve our model, we consider the multiple linear regression model with all the explanatory variables except using percentage of voter turnout instead of unnormalised voter count. From Appendix Figure 8b, we see that the multiple linear regression is an improved fit from the Poisson regression model. This could be because the key assumption that the mean and variance are equal is violated. From Table 4, we see that mean and variance are not equal.

Since an important assumption for the Poisson regression model does not hold, we also build a negative binomial model. We can relax the assumption of mean and variance as equal in negative binomial model. We got almost identical results from the negative binomial regression model as it is a close variant of the Poisson model with looser assumptions whose details can be found in the relevant files in the `models` folder.

4 Results

4.1 Socio-Economic Models

Hey.

4.2 Demographic Models

Hey.

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

Table 5: Socio-economic explanatory models of voter turnout based on education, employment and income

| | Multiple Linear Regression | Poisson Regression |
|--------------------|----------------------------|--------------------|
| (Intercept) | 31.28 | 11.03 |
| | [9.90, 52.02] | [11.00, 11.06] |
| percent_uneducated | −0.11 | 0.00 |
| | [−0.61, 0.39] | [0.00, 0.00] |
| income | 0.04 | 0.00 |
| | [−0.02, 0.10] | [0.00, 0.00] |
| unemployment_rate | −0.41 | −0.08 |
| | [−1.79, 0.97] | [−0.08, −0.07] |
| Num.Obs. | 25 | 25 |
| R2 | 0.354 | |
| R2 Adj. | 0.105 | |
| Log.Lik. | −66.902 | −7176.104 |
| ELPD | −71.8 | −7687.8 |
| ELPD s.e. | 4.8 | 1719.9 |
| LOOIC | 143.7 | 15 375.7 |
| LOOIC s.e. | 9.6 | 3439.7 |
| WAIC | 142.4 | 18 010.8 |
| RMSE | 3.33 | 3685.45 |

Table 6: Demographic explanatory models of voter turnout based on population and number of subdivisions

| | Multiple Linear Regression | Poisson Regression |
|-------------|----------------------------|--------------------|
| (Intercept) | 39.80 | 9.52 |
| | [12.40, 63.82] | [9.48, 9.56] |
| num_sub | 0.17 | 0.01 |
| | [−0.08, 0.39] | [0.01, 0.01] |
| population | 0.00 | 0.00 |
| | [0.00, 0.00] | [0.00, 0.00] |
| Num.Obs. | 25 | 25 |
| R2 | 0.185 | |
| R2 Adj. | −0.106 | |
| Log.Lik. | −69.574 | −9817.594 |
| ELPD | −73.3 | −10 373.2 |
| ELPD s.e. | 3.0 | 1962.9 |
| LOOIC | 146.5 | 20 746.4 |
| LOOIC s.e. | 6.0 | 3925.8 |
| WAIC | 146.1 | 23 902.6 |
| RMSE | 3.77 | 4268.57 |

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Datasheet

B Model Details

B.1 Posterior Predictive Check

In Figure 8a we implement a posterior predictive check on the Poisson regression model. This shows the comparison between the actual outcome variable (number of voters) with simulations from the posterior distribution. From the figure, we can see that the observed data has peaked while the posterior predictive distributions are more dispersed. This means that the model is not a good fit and does not replicate the observed distribution well.

In Figure 8b we implement a posterior predictive check on the multiple linear regression model. This shows the comparison between the actual outcome variable (voter turnout percentage) with simulations from the posterior distribution. Here the observed data and posterior predictions have some overlap. This model is a better fit than the multiple regression model.

In Figure 8c we compare the posterior with the prior. This shows how much the estimates of the coefficients of our variables: unemployment rate, uneducated population and income, have changed once data was taken into account.

Similar results are seen for the demographic models in Figure 9a, Figure 9b, and Figure 9c.

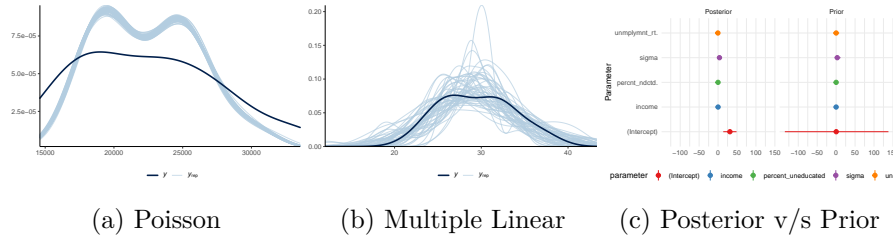


Figure 8: Examining how the socio-economic models fit, and is affected by, the data

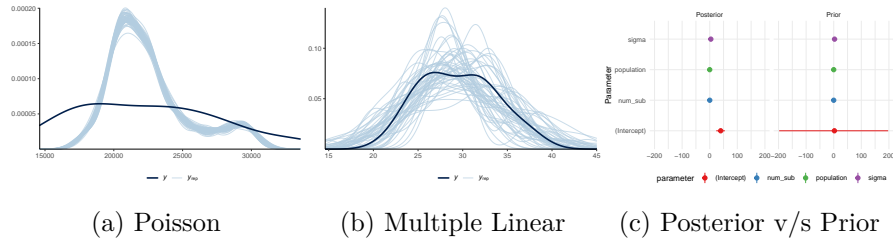


Figure 9: Examining how the demographic models fit, and is affected by, the data

B.2 Diagnostics

The Markov chain Monte Carlo sampling algorithm checks for signs that the algorithm has issues. We consider a trace plot and a Rhat plot for our socio-economic model. In Figure 10a, we see horizontal lines that bounce around and have overlap between the chains. In Figure 10b, we see that everything is close to 1. We do not see anything out of the ordinary in the trace plot or Rhat plot, indicating that the algorithm did not run into any issues.

We see the same results for the demographic model in Figure 11a and Figure 11b.

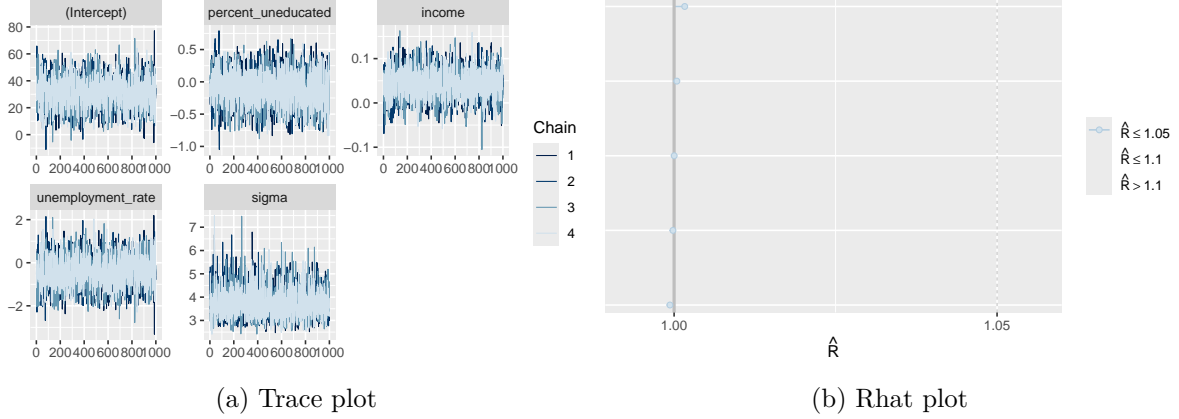


Figure 10: Checking the convergence of the MCMC algorithm for the socio-economic multiple linear regression model

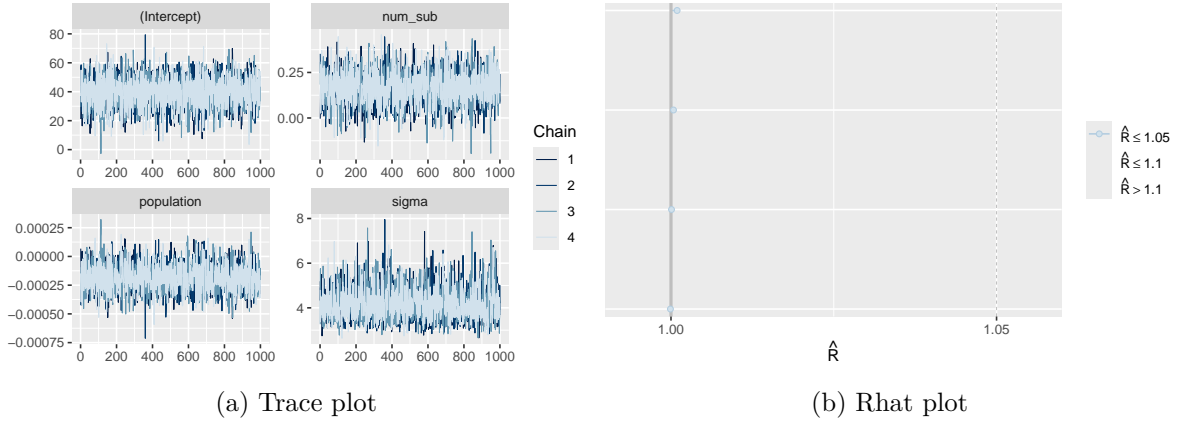


Figure 11: Checking the convergence of the MCMC algorithm for the demographic multiple linear regression model

References

- Marshall, Sean. 2023. “Election: Voter Turnout 2022.” <http://spacing.ca/toronto/2023/03/07/election-voter-turnout-in-2022/>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Warren, May. 2022. “Toronto 2022 Municipal Election Brings Dismal Voter Turnout.” <https://www.thestar.com/news/gta/2022/10/24/toronto-2022-municipal-election-brings-dismal-voter-turnout.html>.