Behind the Delays: Exploring Variations in Toronto's Transit Disruptions*

A Comparative Analysis of Delay Duration Across TTC Subway, Streetcar, and Bus Systems

Claire Chang

November 27, 2024

This paper uses a Bayesian Linear Regression Model to examine which Toronto Transit Commission (TTC) transit mode—subway, streetcar, or bus—experiences the highest frequency and duration of delays. We found that [summarize main finding here—e.g., buses experience significantly longer delays compared to subways and streetcars]. This finding is important because it highlights critical areas where transit improvements could alleviate the impact of delays on daily commuters. Overall, our study provides valuable insights into TTC transit reliability, informing efforts to enhance urban transportation efficiency and passenger experience.

Table of contents

*Code and data are available at: [https://github.com/ClaireUoft/Toronto_TTC_Transportation).

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

The remainder of this paper is structured as follows. Section 2 discusses the data used for this analysis, including key variables and sources, with particular attention to the quality metrics that affect polling accuracy. Section 3 outlines our modeling approach…, incorporating lessons learned from recent electoral cycles. Our predictions are under section of each model. Section 5 discusses the implications of our findings and suggests directions for future research. Finally, Section A evaluates methodology and survey copy.

2 Data

2.1 Overview

In this project, we used data from the opendatatoronto dataset created by (Gelfand 2022). This dataset provided bus, subway, streetcar records essential for our analysis. In this project, we used R(R Core Team 2023) and several R packages for data processing, analysis, and visualization. Specifically, tidyverse (Wickham et al. 2019), arrow(Richardson et al. 2024), here(Müller 2020), ggplot2(Wickham 2016), while dplyr(Wickham et al. 2023) was key for data manipulation tasks. For dynamic report generation, knitr(Xie 2023) and kableExtra (Zhu 2024) used, providing enhanced formatting for outputs. Together, these packages enabled efficient data cleaning, analysis, and visualization throughout the study.

2.2 Measurement

Some paragraphs about how we go from a phenomena in the world to an entry in the dataset.

Table 1: Preview of Cleaned Dataset for Bus, Subway, and Streetcar (One Example Each)

Date	Time	Day	Transit_mode	Line	Location	Incident	Min Delay
2024-01-01	02:08:00	Monday	Bus	89	KEELE AND GLENLAKE	Vision	10
2024-01-01	02:45:00	Monday	Streetcar	505	DUNDAS AND MCCAUL	Security	10
2024-01-01	02:00:00	Monday	Subway	1	SHEPPARD STATION	N/A	0

2.3 Outcome variables

2.3.1 Delay Duration (Min Delay)

The main outcome variable in this study is the duration of delays, measured in minutes (Min Delay). This variable quantifies the length of each delay experienced by the TTC's different transit modes (subway, streetcar, bus). The primary goal of the analysis is to determine which transit mode exhibits the highest average delay duration.

To visualize the distribution of delay durations, a histogram Figure 1 is presented to provide an overview of the range of delay durations for each transit mode. This figure highlights the spread and common ranges of delay times, allowing us to observe whether there are particularly long delays for any given mode.

Lastly, Figure 2 displays a boxplot comparing delay durations across subway, streetcar, and bus transit modes. The boxplot offers insights into the median delays and potential outliers that may suggest unusual or extreme events, while showing variability in delay times across the different transit modes.

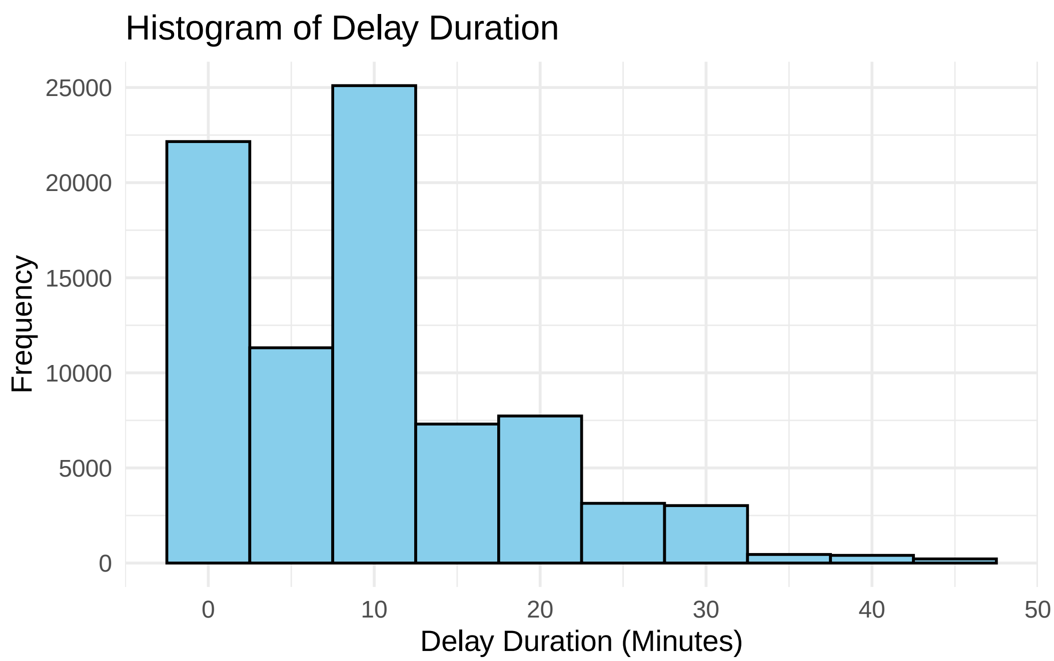

Figure 1: Histogram of Delay Duration (Filtered at 45 Minutes)

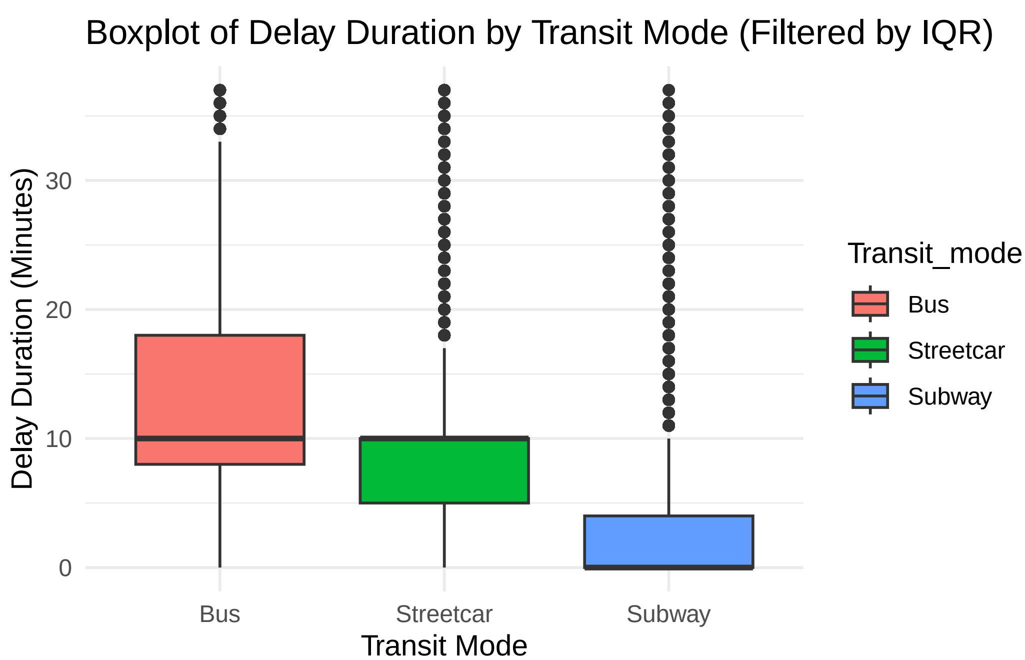

Figure 2: Boxplot of Delay Duration by Transit Mode (Filtered by IQR)

2.4 Predictor variables

2.4.1 Transit Mode

The main predictor variable is Transit Mode, which includes three categories: subway, streetcar, and bus. Each transit mode has distinct characteristics, service areas, and operational challenges that may contribute to variations in delay duration.

To examine this relationship, Figure 3 plots the average delay duration for each transit mode. This chart helps us visualize which transit mode experiences the longest delays on average and illustrates the extent to which each mode contributes to the overall delay picture.

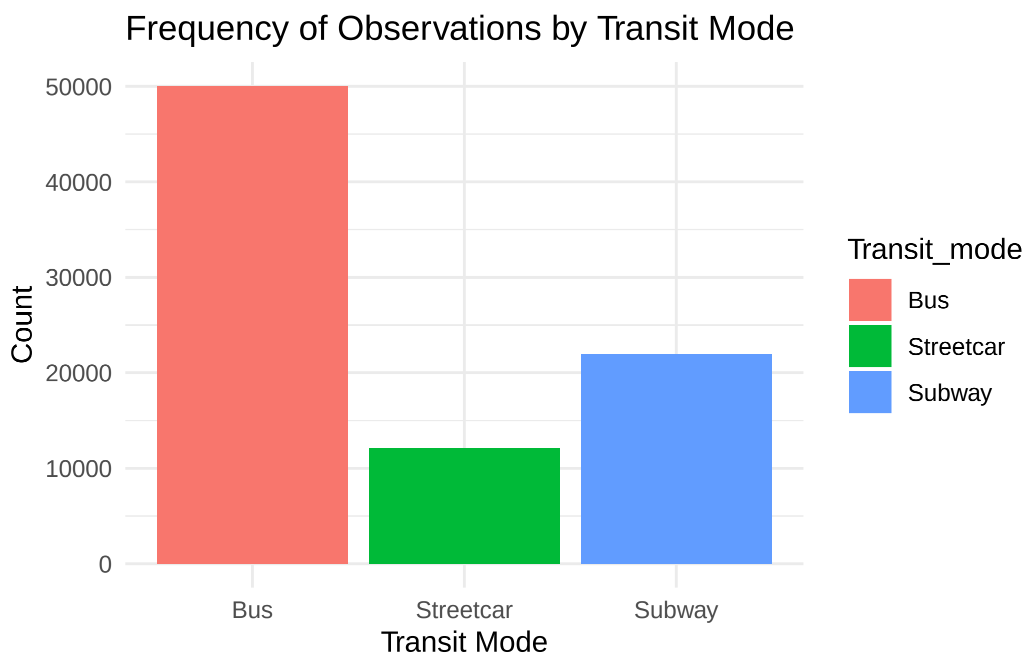

Figure 3: Average Delay Duration by Transit Mode

Table 2: Count of Observations by Transit Mode

Transit_mode	Count
Bus	50039
Subway	21979
Streetcar	12107

2.4.2 Time of Day

Another key predictor is Time of Day shown in Figure 4. This variable plays a crucial role in understanding how congestion and operational challenges fluctuate throughout the day. For example, delays might be more frequent and prolonged during peak hours (morning and evening rush periods) due to increased ridership.

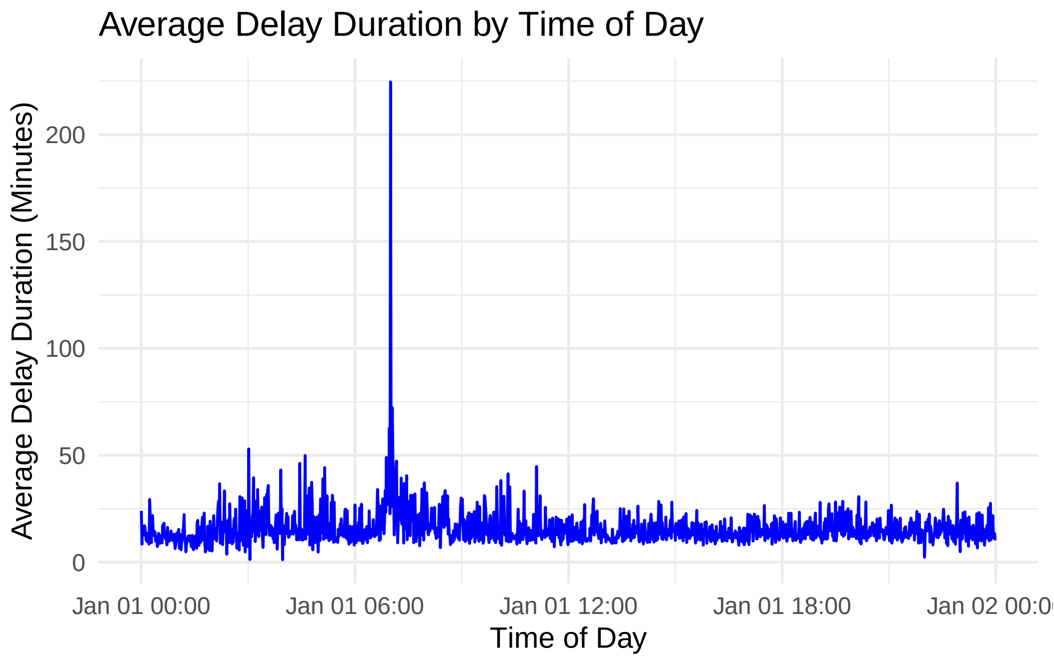

Figure 4: Average Delay by Time of Day

Figure 4 shows a line plot representing average delay duration across different hours of the day. This visualization allows us to determine the times at which each transit mode is most vulnerable to delays, and whether certain transit modes are disproportionately affected during rush hours.

2.4.3 Day of the Week

The Day of the Week is an important predictor that may influence delays due to varying passenger demand and maintenance schedules. Weekdays typically see different patterns compared to weekends, with increased rider demand and therefore greater potential for delays.

Figure 5 is a bar chart that shows the average delay duration for each day of the week, grouped by transit mode. It allows us to observe any weekly trends in delay duration, such as consistent

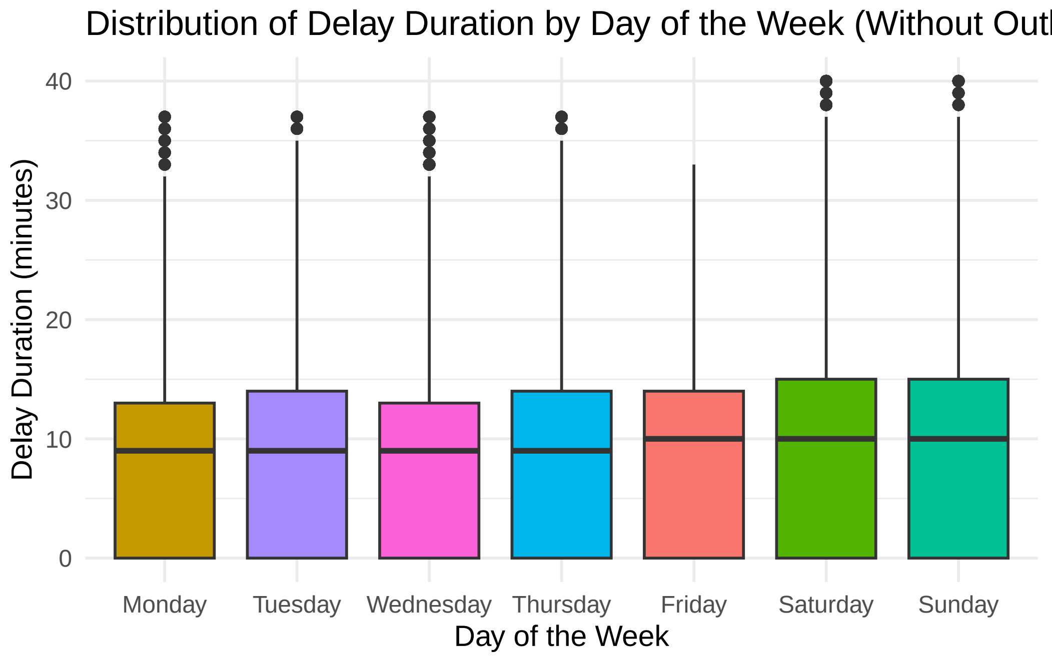

Figure 5: Delay Duration by Day of the Week

weekday delays or weekend-specific issues, giving more insight into the operational aspects of the different modes.

The predictor variables, including transit mode, time of day, and day of the week, help in assessing the specific conditions that affect the delay duration.

3 Model

The Bayesian linear regression model is designed to predict the duration of delays experienced by different TTC transit modes (subway, streetcar, or bus) based on various predictors such as transit mode, time of day, and day of the week. Background details and diagnostics are included in Appendix C.

3.1 Model Set-Up

Define y_i as the duration of the delay (in minutes) for a given transit event. Let β_i represent the transit mode (subway, streetcar, or bus), and γ_i and δ_i represent the time and day of the week, respectively. The generative model is specified as follows:

$$
\begin{aligned}
y_i|\mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) & (1)\\
\mu_i &= \alpha + \beta_i + \gamma_i + \delta_i & (2)\\
\alpha &\sim \text{Normal}(0, 2.5) & (3)\\
\beta &\sim \text{Normal}(0, 2.5) & (4)\\
\gamma &\sim \text{Normal}(0, 2.5) & (5)\\
\delta &\sim \text{Normal}(0, 2.5) & (6)\\
\sigma &\sim \text{Exponential}(1) & (7)
\end{aligned}
$$

Here:

- y_i is the observed delay duration for a specific transit event.
- α is the baseline average delay duration.
- β_i captures the effect of transit mode (subway, streetcar, or bus).
- γ_i captures the effect of the time of day (e.g., peak or off-peak).
- δ_i captures the effect of the day of the week.
- σ is the standard deviation, representing unexplained variability in y_i.

This Bayesian model allows for estimating the contributions of transit mode, time, and day to delay durations while accounting for uncertainty in the data.

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular…

We can use maths by including latex between dollar signs, for instance θ.

4 Results

Our results are summarized in **?@tbl-modelresults**.

5 Discussion

5.1 First discussion point

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

5.2 Second discussion point

Please don't use these as sub-heading labels - change them to be what your point actually is.

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

A Appendix

B Additional data details

C Model details

C.1 Posterior predictive check

References

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

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, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://CRAN.R-project.org/package=knitr.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.