

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

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

Claire Chang

November 26, 2024

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

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.

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

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) used, providing enhanced formatting for outputs. Together, these packages enabled efficient data cleaning, analysis, and visualization throughout the study.

2.2 Overview text

2.3 Measurement

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

2.4 Outcome variables

Add graphs, tables and text. Use sub-sub-headings for each outcome variable or update the subheading to be singular.

Talk way more about it.

2.5 Predictor variables

Add graphs, tables and text.

Use sub-sub-headings for each outcome variable and feel free to combine a few into one if they go together naturally.

3 Model

The goal of our modelling strategy is twofold. Firstly,...

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in [Appendix C](#).

3.1 Model set-up

3.2 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:

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \quad (1)$$

$$\mu_i = \alpha + \beta_i + \gamma_i + \delta_i \quad (2)$$

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

$$\beta \sim \text{Normal}(0, 2.5) \quad (4)$$

$$\gamma \sim \text{Normal}(0, 2.5) \quad (5)$$

$$\delta \sim \text{Normal}(0, 2.5) \quad (6)$$

$$\sigma \sim \text{Exponential}(1) \quad (7)$$

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`.

3.2.1 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 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, Dragoş 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 Golemund, 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>.