# Have public transport delays in Toronto gotten worse over time? An analysis of the TTC street car delays*

Siddharth Arya

January 25, 2024

The Toronto Transit Commission (TTC) has been a core part of the commute of all residents of the city of Toronto for decades - keeping each part of the city connected. This paper looks at TTC-Streetcar-delay data in order to see detect the presence of any trend in said data. The paper founnd that there is a small positive correlation between the months elapsed since the start of 2021, and the total minutes of delayed experienced by the streetcars accross the Greater Toronto Area (GTA). Further analysis should be conducted to diagnose the reason behind this growth of delays in the streetcar network.

## Table of contents



*Code and data are available at: https://github.com/sid19arya/TTC-Delay-Analysis.git

# 1 Introduction

As the largest and most densely populated city in Canada, Toronto has an incredibly high amount of residents who require efficient transportation throughout the city. The Toronto Transit Comission (TTC) takes the role of managing this public transportation (streetcars, subways, buses) in order to fulfill the required task of transporting residents. While it provides a great service to the city, it responsibility also puts it at risk of a high amount of criticism.

As the COVID-19 pandemic died down, riders returned to the TTC, making the public transit around the city active once more. However, with their return came rampant criticism about the delays regarding the TTC (O'Neil 2022). In fact, articles claim confidently that the TTC is deteriorating as time passes and adding to the commutes of residents of the city (O'Neil 2022). The growing frustration pushes further news articles claiming that the TTC is failing to fulfill its duty and is 'broken' (Davis 2023).

However, these are empty claims yet to be strongly supported or disproved. This study looks more closely at open data regarding the delays of public transportation - specifically at streetcar delays, in order to reveal ether these claims and criticism have validity to them.

This paper is organized in the following sections: Data, Results, Discussion, Conclusion. First, the data section details the collection, cleaning and visualization of the datasets used - it describes the variables created and their use. Next, the results section takes the clean data and conducts analysis to assess the presence of a correlation between variables. Next, the discussion section addresses the implications of the found trend as well as limitations of the work of this paper, and suggests future works. Finally, the conclusion summarises the main discoveries of the paper.

# 2 Data

The data utilized in this paper is from OpenDataToronto Library, the Data sets used are entitled: "TTC-Streetcar-Delays". This dataset was last updated January 18, 2024 - and contains a detailed report of each streetcar delay from 2013 through to 2023. For the purposes of this paper, we will only examine data from 2021 on wards - as we are focused on post-pandemic trends in streetcar delays. The data was collected, sorted, analyzed and visualized through the help of the open source programming language R.(R Core Team 2022) Within R, the following libraryies also were used in order to aid in the process: tidyverse(Wickham et al. 2019), ggplot2(Wickham 2016), knitr(Xie 2014), kableExtra(Zhu 2021), opendatatoronto(Gelfand 2022), and here(Müller 2020)

Table 1: Sample of Cleaned TTC Delay Data

| Date | incidents | total_delay | months_elapsed |
|---|---|---|---|
| 01/21 | 934 | 9382 | 1 |
| 02/21 | 865 | 11819 | 2 |
| 03/21 | 1036 | 11708 | 3 |
| 04/21 | 1107 | 10847 | 4 |
| 05/21 | 1059 | 15955 | 5 |
| 06/21 | 955 | 11471 | 6 |

## 2.1 Features of the Data

From the source, three individual datasets were gathered, for the delays from the years 2021, 2022, and 2023. Each of these datasets have the same 10 variable columns, and the number of observations they include range from 14400 to 17700. Each observation contains information about the the time, location, reason, and cause of each and every streetcar delay. Specifically, the raw dataset included: Date, Line (the streetcars are split into lines across the city), Time, Day, Location (nearest station to delay),Incident (reason for delay),Min Delay (total amount of minutes the delay happened for), Min Gap, Bound (direction of streetcar travel),Vehicle (unique number of vehicle involved).

## 2.2 Cleaning the Data

For cleaning the data, we apply simple steps to organize and summarize information from the three similar datasets, and aggregate their information. First, we drop rows that lacked any information (some columns indicated 'N/A' perhaps due to human error in logging the incident). Next we changed the Date/time of the incidents into a format more appropriate for our use-case (in 'mm/yyyy' format). Finally, we aggregate over each of the incidents, and grouping by month, we create a count for the number of incidents in that month, as well as the total number of minutes delayed (summing the min_delay column from the initial data set). We then concatenate the three datasets together. Additionally, after concatenating we also added a months elapsed (from Jan 2021) column, this is because a numerical representation of passing time helps in visualizing the data, and calculating the correlation between factors.

A sample of the clean data is shown in table (Table 1).

## 2.3 Visualizing the data

From the clean data, we contain simple visual representations of the number of incidents, and total delay across months in figures 1(Figure 1) and 2(Figure 2) respectively.

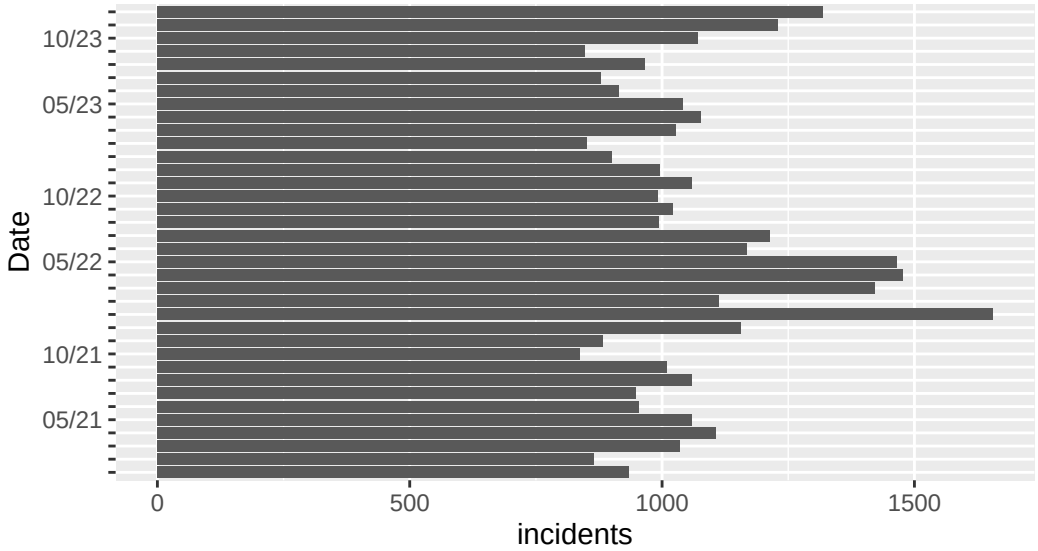


Figure 1: Number of Delay Incidents accross months

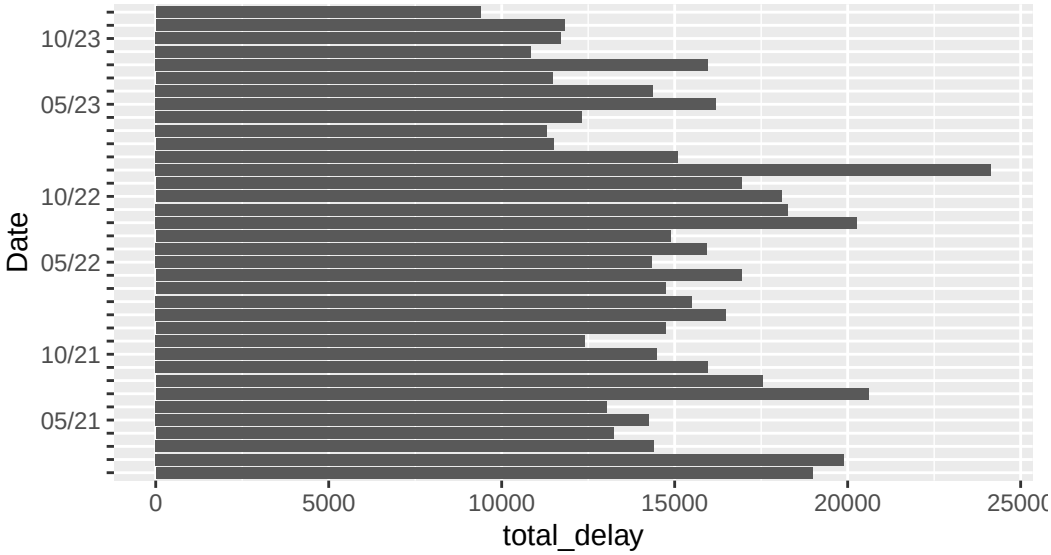


Figure 2: Total Delay (in minutes) accross months

# 3 Results

## 3.1 Presence of Correlation

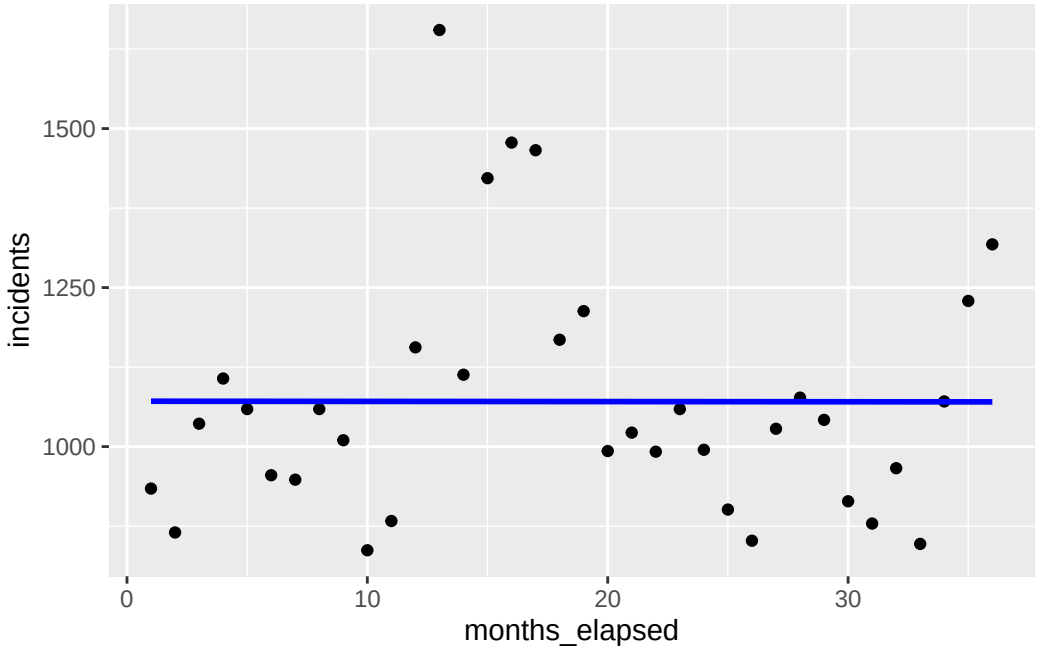


Figure 3: Incidents across months (elasped from Jan/2021), with line of best fit

Upon an initial rough visual inspection of the histograms, granting some amount of variance, neither the number of incidents nor the total delay seem to show notable growth. However, further investigation indicates differently. Figures 3(Figure 3) and 4(Figure 4) show the same data displayed as scatter plots, with a line of best fit to indicate any correlation. Although the number incidents are. This claim is given further validity when examine the correlation coefficient (R) values calculated for the two pairs of variables. The R values were -0.0016 and 0.3881 for the correlation of months elapsed with incident count and total delay respectively. The correlation coefficient of -0.0016 is negligible and therefore cannot be used to draw any indication that the number of streetcar delays incidents has increased the past 3 years. On the other hand, a correlation coefficient of 0.3881 between the total delay in minutes and months elapsed does seem to indicate a small but notably postie correlation between the two variables. Thus we see some support to the claim that the amount of delay faced by TTC riders has shown steady increase in the past 3 years.

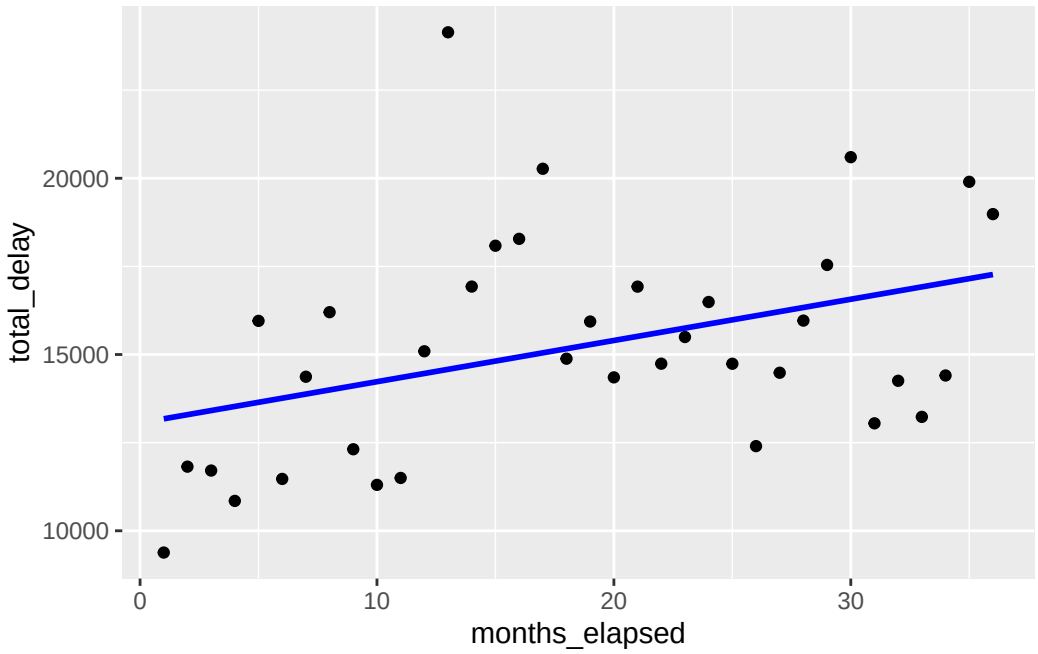


Figure 4: Total_delay (in minutes) across months (elasped from Jan/2021), with line of best fit

## 4 Discussion

The examiniation of the TTC-Streetcar-Delay Data set from 2021 to 2023 offers valuable insight into the trends of the transport systems in the city. The examination evaluated a detailed and thorough dataset and found an increase in the delay experienced by commuters over the past 3 years - specifically, a correlation coefficient of 0.3881 between the total minutes of delay per month, and months elapsed since January 2021.

This asserts that the criticism of the TTC which led to this investigation was grounded in some truth. That being said, the findings of this paper is not sufficient in levying criticism towards the TTC as there is no indication as to the reason of the growth in delays over time. Although a trend was found, the reason behind the trend is a topic to be taken up in order to build on the work of this paper.

Another note to consider is that while this paper shows an increase in minutes of delay over the past 3 years, it does not compare any of this to pre-pandemic data. As riders return to using TTC post-pandemic, the delays could simply be returning to their normal values, rather than indicating a deeper issue.

Finally, this analysis used specifically streetcar delay information - other public transportation in Toronto, like the subway and buses were not discussed, further research could evaluate wether this same trends holds with those modes too.

# 5 Conclusion

This paper investigated TTC delays spanning the years 2021, 2022, 2023 to see if a trend over time could be spotted. The data used was thorough in detailing each delay incident throughout these three years. Although there was no noticeable trend in the number of delay incidents over time, the total minutes of delay experienced by travelers did show a steady increase with time. However, this paper is still not enough to point towards the reason for this increase in delay over time. Future work should focus on understanding this increase in delays and examine multiple other features that could cause such a delay like number of daily commuters, TTC funding, etc.

## References


Davis, Stephen Spencer. 2023. "Who Broke the TTC? Inside Toronto's Public Transit Disaster." 2023. https://torontolife.com/deep-dives/who-broke-the-ttc-inside-torontos-public-transit-disaster/.

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

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://here.r-lib.org/.

O'Neil, Lauren. 2022. "Toronto's Commute Time Is the Longest in North America and Getting Worse." 2022. https://www.blogto.com/city/2022/12/toronto-longest-commute-time-north-america-getting-worse/.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

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.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. http://haozhu233.github.io/kableExtra/.