

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 all residents of the city for decades - keeping each part of the city connected. That being said, the TTC is under constant scrutiny and criticism, especially in recent years regarding delays worsening - inconveniencing people's commutes. This paper examines the TTC Delays, using data from OpenDataToronto, and disproves the notion that delays have been getting increasingly worse over the past few years.

Table of contents

1	Introduction	2
2	Data	2
2.1	Features of the Data	2
2.2	Cleaning the Data	3
2.3	Visualizing the data	3
3	Results	3
4	Discussion	5
5	Conclusion	5
	References	6

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

1 Introduction

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. Pseudo-studies claim confidently that the TTC is deteriorating as time passes (sources) and adding to the commutes of residents of the city.

However, these are empty claims yet to be strongly supported or disproved. This study looks more closely at TTC data - specifically at streetcar delays, to reveal whether these claims are truly founded or not.

This paper is organized in the following sections: Data, Results, Conclusion...

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 onwards - as we are focused on post-pandemic trends in streetcar delays. The data was collected, sorted and analyzed through the help of ... (Cite all the packages).(R Core Team 2022)

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 time, location, reason, and cause of each and every streetcar delay.

The data in its raw form includes all these values ((**tbl_raw?**))

Table 1: Sample of uncleaned TTC Delay Data

Date	Line	Time	Day	Location	Incident	Min Delay	N
2021-01-01	501	03:15:00	Friday	QUEEN AND MCCAUL	Operations	19	
2021-01-01	504	03:37:00	Friday	BROADVIEW AND QUEEN	Operations	15	
2021-01-01	504	04:00:00	Friday	BROADVIEW STATION	Cleaning	15	
2021-01-01	504	04:03:00	Friday	DUNDAS WEST STATION	Cleaning	15	
2021-01-01	506	05:37:00	Friday	MAIN STATION	Mechanical	10	
2021-01-01	555	06:00:00	Friday	TORONTO TRANSIT COMMIS	General Delay	0	

Table 2: 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

2.2 Cleaning the Data

For cleaning the data, we apply simple steps to organize and summarise 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 (‘mm/yyyy’). 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 dataset). 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 ((**tabl-clean?**)).

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((**incidents-hist?**)) and 2((**delay-hist?**)) respectively.

3 Results

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((**inc_plot?**)) and 4((**delay_plot?**)) 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

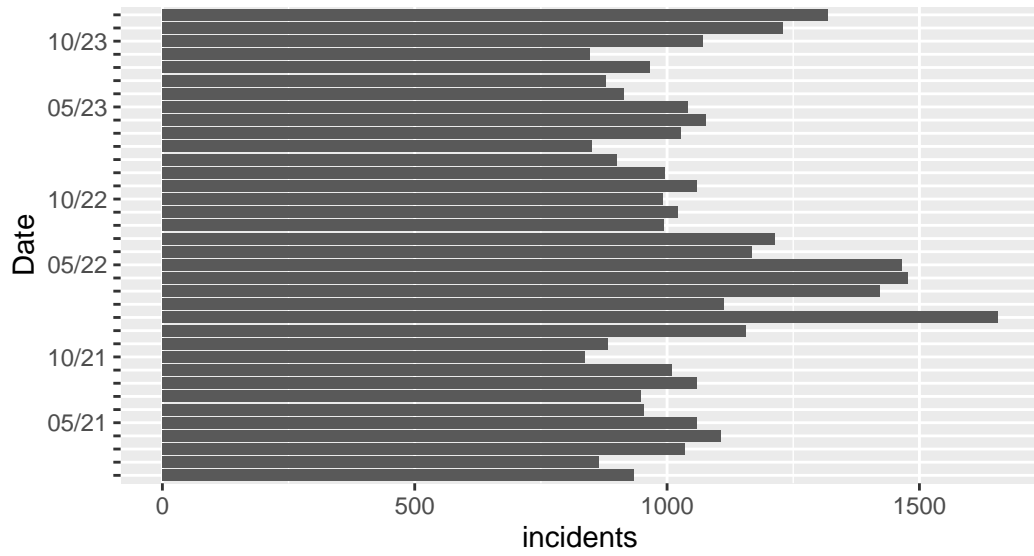


Figure 1: Number of Delay Incidents accross months

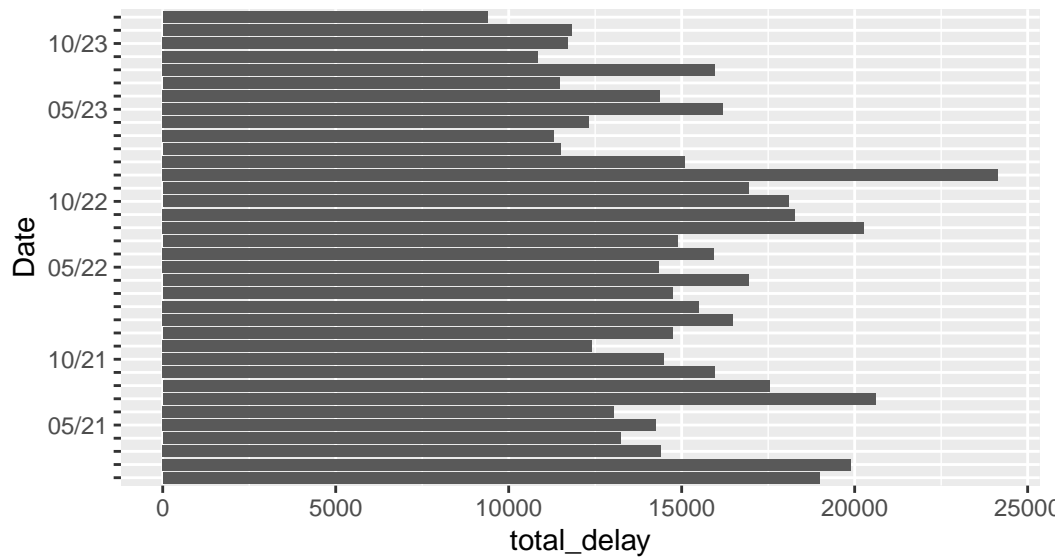


Figure 2: Total Delay (in minutes) accross months

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 positive 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.

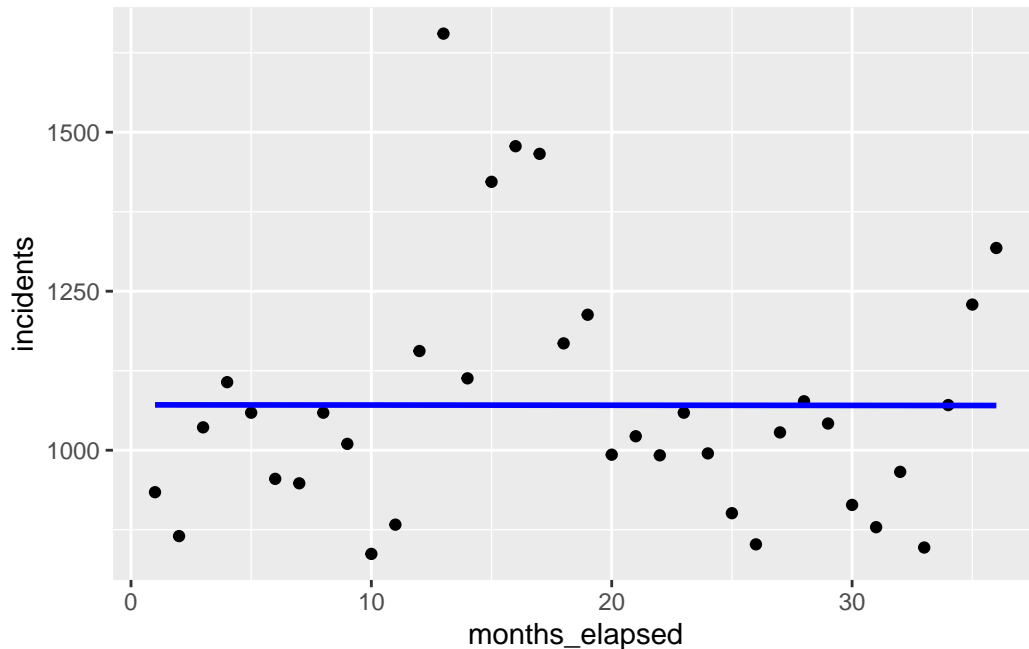


Figure 3: Incidents across months (elapsed from Jan/2021), with line of best fit

```
[1] "Incidents_Correlation (R value): -0.00155711284854218"
```

```
[1] "Delay_Correlation (R value): 0.388129231071415"
```

4 Discussion

- mainly that correlation can be explained for many reasons. Criticism show that the ttc is just bad, further investigation might observe many people have just been returning more and more, and more use of the ttc causes more delay incidents or whatnot.

5 Conclusion

This paper investigated TTC delays spanning the years 2021, 2022, 2023 to see if a trend over time could be spotted. The paper concluded that there was no noticeable

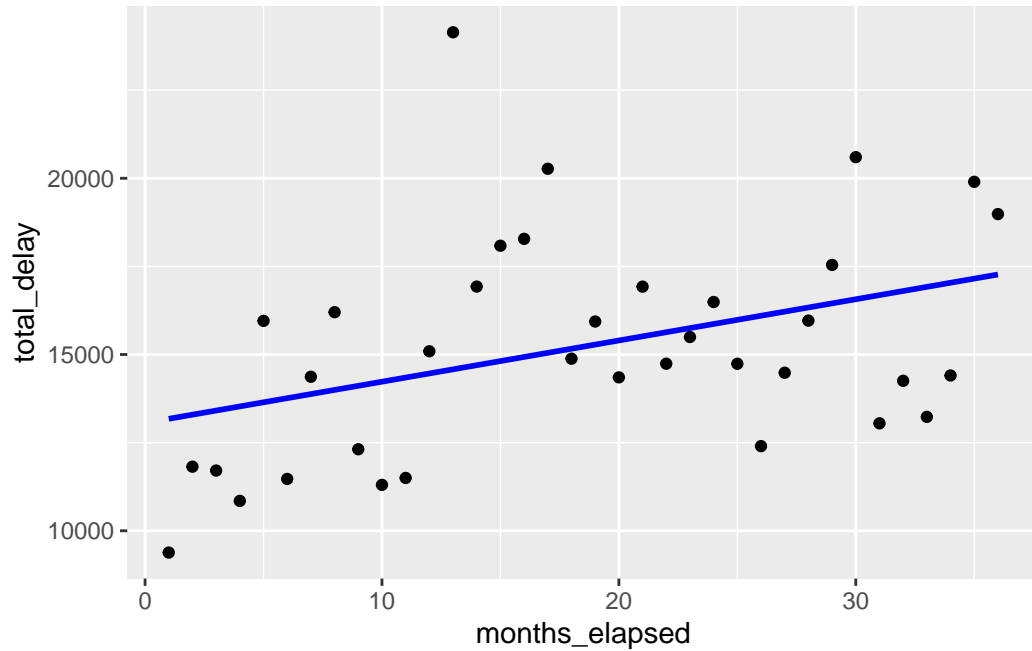


Figure 4: Total_delay (in minutes) across months (elapsed from Jan/2021), with line of best fit

References

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