Decoding Delays: Analyzing Toronto's Streetcar System*

Investigating Patterns and Trends in TTC Streetcar Delays (2022-2024)

Daniel Du

2024-09-24

This paper analyzes delay data from Toronto's TTC streetcar system between 2022 and 2024 to assess patterns in delay frequency and length. The study found that while the frequency of delays has decreased, the average delay length has increased, with delays showing no consistent patterns based on month, day of the week, or hour. Understanding streetcar delay trends helps determine methods of improving public transportation in Toronto.

1 Introduction

Public transportation systems are vital components of urban infrastructure, providing an essential service that connects people to their communities, reduces traffic congestion, and promotes sustainability. Streetcars were once a common mode of transportation across North American cities, but have largely disappeared due to car-centric urban planning Britannica (2024). Today, Toronto stands alone as the continent's only major city still utilizing a streetcar network, with the Toronto Transit Commission (TTC) running 11 daytime and 5 nighttime lines. These are often criticized for being slow and unreliable, with frequent delays frustrating commuters while contributing to broader congestion, undermining the positives of public transit Micallef (2024). Addressing these inefficiencies could significantly improve accessibility, sustainability, and the overall quality of life in Canada's largest city.

Despite the central role of the streetcar, which serves approximately 230,000 daily commuters, there is a lack of comprehensive analysis quantifying the impact of streetcar delays on transit efficiency Arsenych (2022). This paper explores the frequency, duration, and patterns of these disruptions and aims to highlight patterns. Understanding where, when, and why these delays

*Code and data are available at https://github.com/danield424/TTC-Streetcar-Delays occur is central to identifying the areas most in need of improvement and optimizing this key component of Toronto's public transportation system.

This study analyzes public streetcar delay data from 2022 to 2024 to assess the extent and impact of delays in Toronto's streetcar network. After examining delay patterns across different years, months, days, and hours, and comparing average delay lengths and frequencies across various service lines, the analysis reveals a uniform increase in average delay times, despite a reduction in delay frequency. These findings highlight systemic inefficiencies in the streetcar network, showing the need for further research and investment to improve transit performance. The findings from this study support movements to enhance the efficiency and reliability of Toronto's public transportation system.

The paper is structured as follows: Section 2 introduces the data and methodology used; Section 3 presents the key findings; Section 4 discusses the real-world implications of these results, and Section 5 discusses the limitations of the dataset and potential avenues for future research.

2 Data

To investigate streetcar delays, delay logs from 2022, 2023, and 2024 (up until September) is sourced from the Opendatatoronto R package, which provides data from the City of Toronto Open Data Catalogue Gelfand (2022). These datasets, intended to monitor and improve public transit performance, are regularly updated and were current as of September 1, 2024. Although similar datasets exist, this specific dataset was chosen due to its official nature, reliability, and size.

The data includes the date and time, line number, and delay length of over 40,000 instances of streetcar delays since 2022. It is unclear how the data was gathered; reliability is ensured as it comes directly from the TTC, but it is possible there were missed or incorrect measurements due to limitations. Delay times are measured in minutes, and time is measured to the second.

The data was thoroughly cleaned to ensure accuracy and prepare it for analysis; the process is detailed in Appendix~A. In Section 2.1 and Section 2.2 we explore the dataset variables before conducting further analysis in Section 3.

2.1 Overview of Data

Figure 1 provides a sample of 6 data observations. In total, there are 36701 observations of 5 variables: `day` contains the day of the week, `date` contains the timestamp of the delay, `service_type` was constructed and assigns the service type of the line in `line`. `min_delay` is the delay time in minutes; this is the main variable of interest. We will analyze `min_delay` in relation to the 4 other variables. There are no additional relationships between the variables aside from the obvious `day - date` and aforementioned `service_type` construction.

day	date	service_type	line	min_delay
Saturday	2022-01-01 07:21:00	Regular	504	30
Saturday	2022-01-01 08:22:00	Regular	501	16
Saturday	2022-01-01 08:28:00	Regular	504	18
Saturday	2022-01-01 08:34:00	Regular	510	30
Saturday	2022-01-01 08:39:00	Night	301	5
Saturday	2022-01-01 08:42:00	Regular	501	10

Figure 1: Summary of Streetcar Delay Data 2022-2024

2.2 Delay by Year and Service Type

The number of delay observations per year and for each service type are below. Note the low number of delays in 2024 is due to the data ending in August. The vast majority of observations come from the regular streetcar service, which is intuitive since the regular service has the most lines (9 total) and runs the most regularly.

year	n
2022	16416
2023	12198
2024	8087

Figure 2: Number of Delay Observations by Year

service_type	n
Night	613
Reduced	395
Regular	35693

Figure 3: Number of Delay Observations by Service Type

We can calculate summary statistics for the delay length, and then filter them by year and service type:

Table 4: Min. Delay (Delay Length) Summary

Statistic	Value
Min.	3.00000
1st Qu.	8.00000
Median	10.00000
Mean	13.69287
3rd Qu.	12.00000
Max.	119.00000

Table 5: Min. Delay Summary by Service Type

service_type	minimum_delay	median_delay	mean_delay	maximum_delay	sd_delay
Night	3	20	24.1	116	16.6
Reduced	3	10	21.0	109	20.2
Regular	3	10	13.4	119	14.1

Table 6: Min. Delay Summary by Year

year	minimum_delay	median_delay	mean_delay	maximum_delay	sd_delay
2022	3	9	12.2	119	13.1
2023	3	10	14.5	119	14.7
2024	3	10	15.5	119	15.9

The minimum delay of 3 minutes and maximum of 119 minutes are due to constraints placed when cleaning the data, as discussed in Appendix~A. The median delay is twice as long for the night service compared to the two others. The median delay times over the 3 years are similar, although the yearly mean delay rose by more than 3 minutes between 2022 and 2024.

Following the introduction of the variables, we can explore the results of our data analysis.

3 Results

3.1 Analysis of Average Delay Time

One of the key factors of interest is the average delay time for different variable values, such as streetcar lines, types, and dates. Examining average delays allows us to identify patterns and areas where delays are most pronounced. This highlights specific points of inefficiency and potential targets for improvement. In Figure 4 we begin by looking at the average delay time for each line.

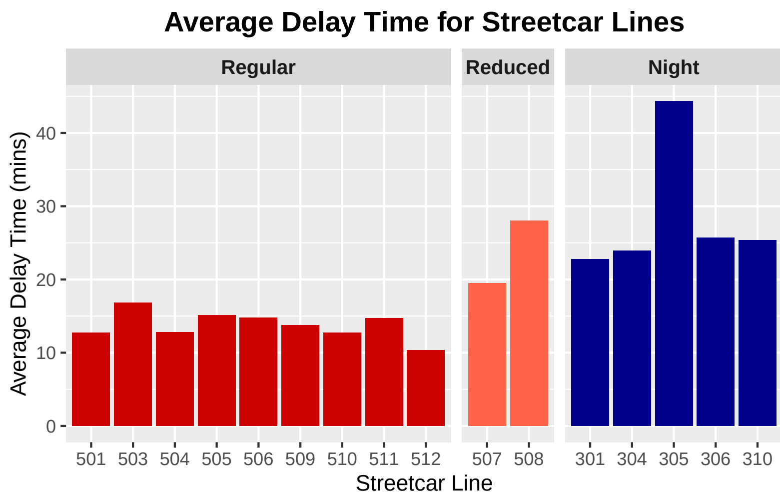

Figure 4: Average delay time for streetcar lines, sorted by service type.

Average delay times vary significantly across streetcar lines and service types. Regular service lines have relatively consistent delay times, with lines 503 and 505 experiencing slightly higher averages than others. Night service lines, particularly line 305, exhibit the highest average delays. Reduced service lines also show higher delay times compared to regular service lines. In Figure 5 we see if this is the case for each year.

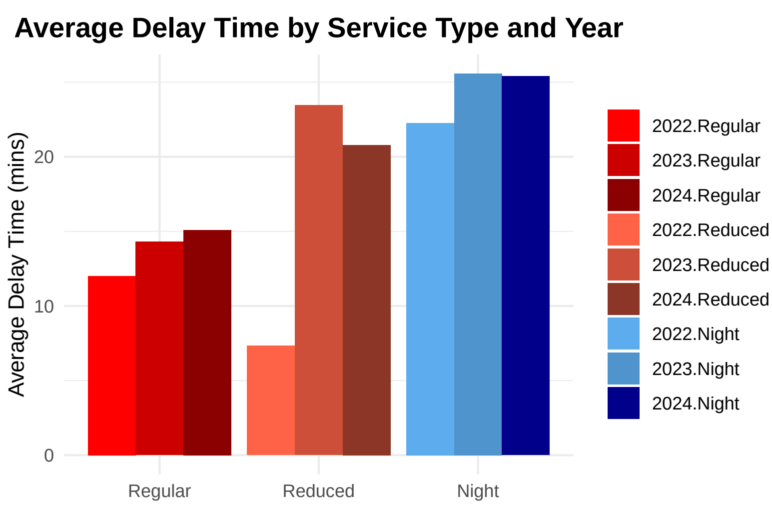

Figure 5: Average delay time for streetcars by service type, per year.

Delay times have increased over the years across all service types, with the most significant increase seen in reduced service lines, particularly from 2022-2023. In years 2023 and 2024, both reduced and night services had a significantly higher average delay length.

We can now look at specific lines for each type individually, starting with regular service in Figure 6 and Figure 7.

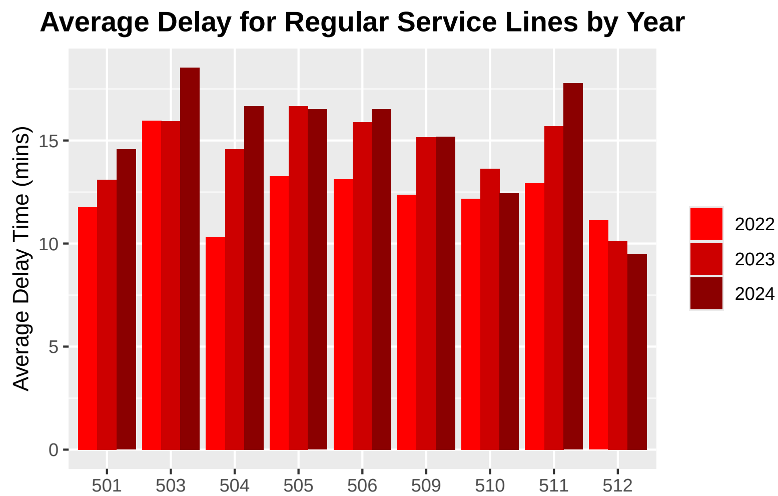

Figure 6: Average delay time for regular service streetcars, per year.

Across almost all regular lines, average delay times have increased over time, indicating a worsening in service reliability. Line 503 and 511 are the regular lines with the most concerning delay time and upward trend, while 512 is the only line with a downward trend.

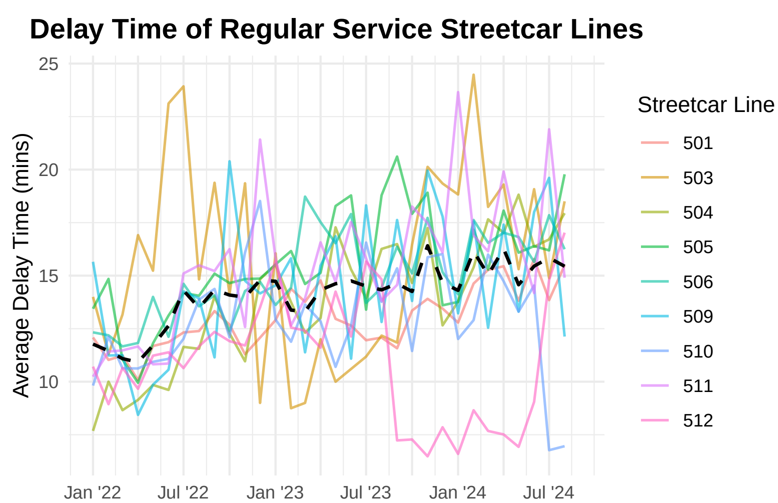

Figure 7: Average delay time for regular service streetcars over time.

In Figure 7 we examine each streetcar line's growth individually over time. Growth is not linear and many lines share similar patterns of delay increases and decreases, particularly during common periods. For instance, peaks and troughs in delays across several lines around mid-2023 suggest external factors impacting multiple lines simultaneously.

Most lines align with the overall average delay trend, dotted in black. Line 503's high peaks and variance stand out compared to the rest, suggesting that it has unique factors or circumstances influencing its delays. The 512, previously shown to be the only line with a decreasing average delay time over the years, appears to have had a temporary period of delay times under 10 minutes starting in late 2023, before reverting back to being near the average.

Similarly, Figure 8, Figure 9, and Figure 10 analyze reduced and night services over time.

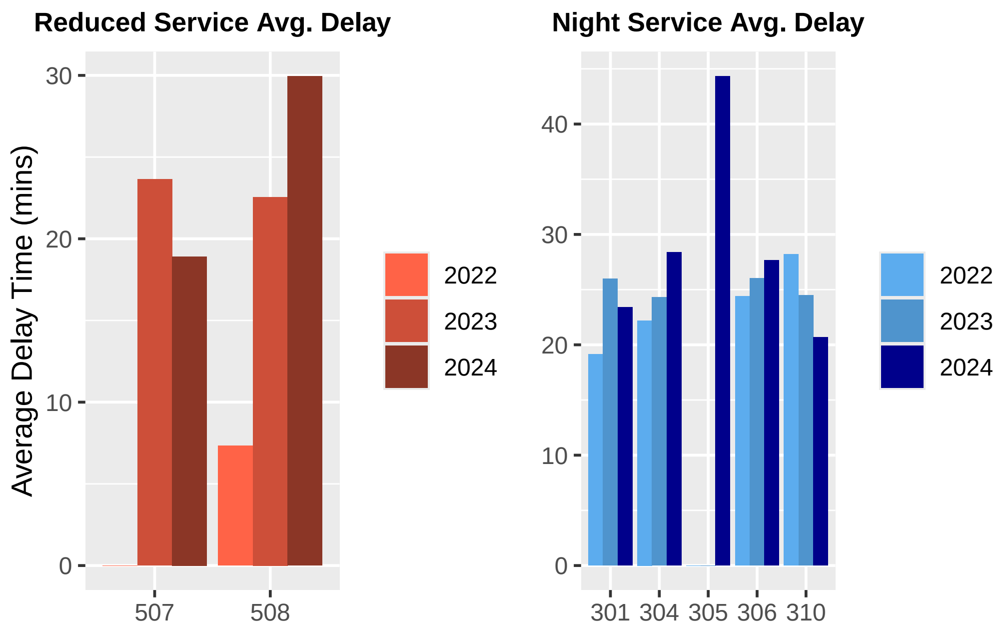

Figure 8: Average delay time for reduced and night service streetcars, per year.

We can now see that there are no observations of the reduced service 507 in 2022 as it was yet to be routed. Therefore, the increase in average delay time for the reduced lines can be attributed entirely to the 508. Similarly, the night service 305 was not routed until 2024, and has by far the lengthiest average delays.

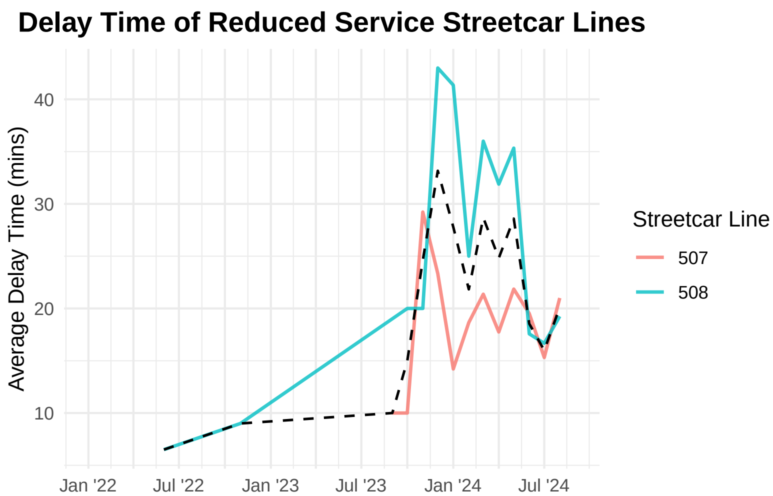

Figure 9: Average delay time for reduced service streetcars over time.

Since the 507 line was in service, the average delay time for the two reduced lines follow similar trends, indicating a correlation between the two. Both peaked late in 2023 and have decreased since.

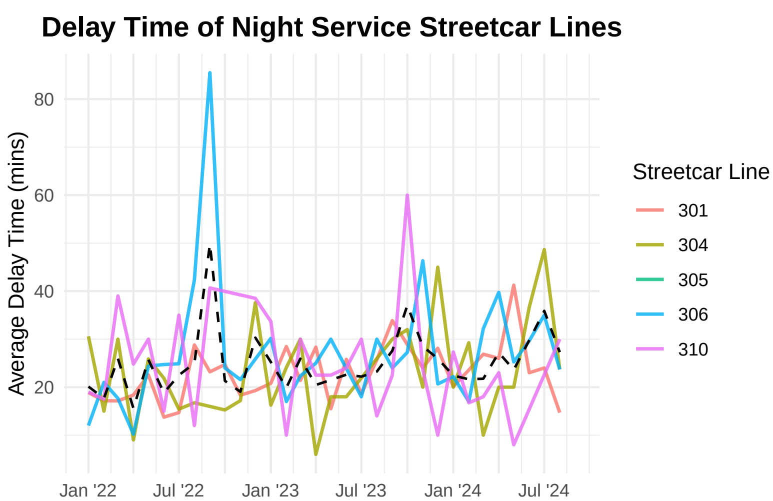

Figure 10: Average delay time for night service streetcars over time.

There is no visible correlation between night service lines, with most of them peaking randomly. The overall average delay time is often follows peaks of lengthy delays. The average delay time has plenty of variance for each line, except the 301 which is fairly consistent.

While analysis of the average delay is useful and identifies general trends, looking at the frequency of delays gives us a clearer picture of the data and each line's efficiency.

3.2 Analysis of Delay Frequency

Before we proceed with delay frequency, we have to deal with missing values in 2024 from September onwards, which we do in Appendix~B. We accomplish this by only using data from January to August when making year-by-year comparisons. Since the total observations for each service type are different, we will look at each service type individually.

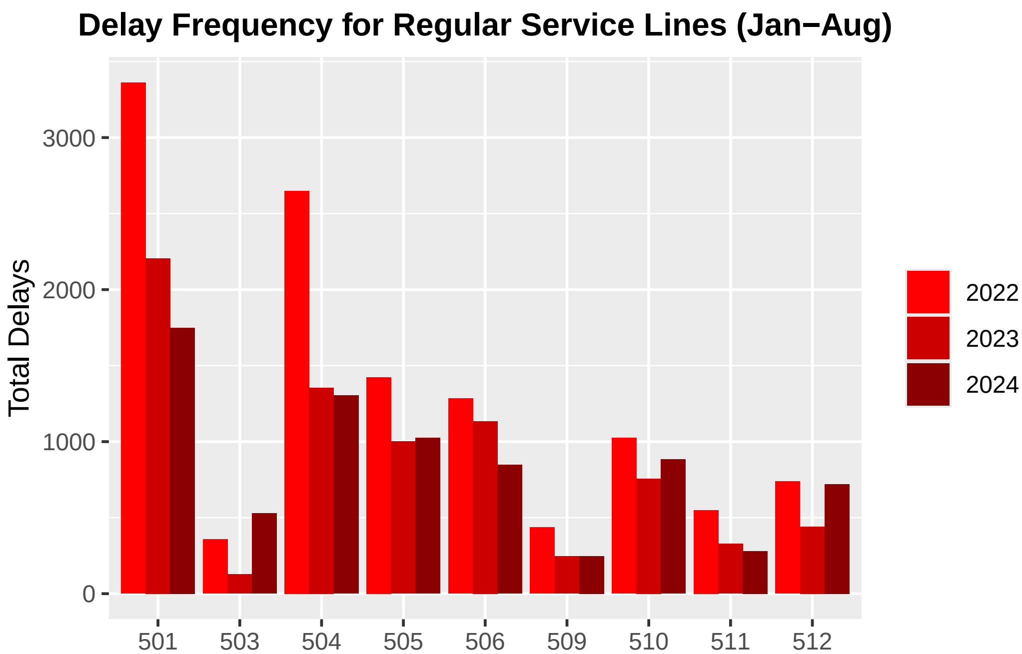

Figure 11: Total delays in January-August for regular service lines by year.

Figure 11 shows the delay frequency by line from January to August of the 3 years. Since the dataset does not provide information on successful trips, a high delay frequency can either mean operational issues or a high service volume, so looking at the value of a frequency alone is not insightful. However, we can see an overall downward trend in delay frequency over the years.

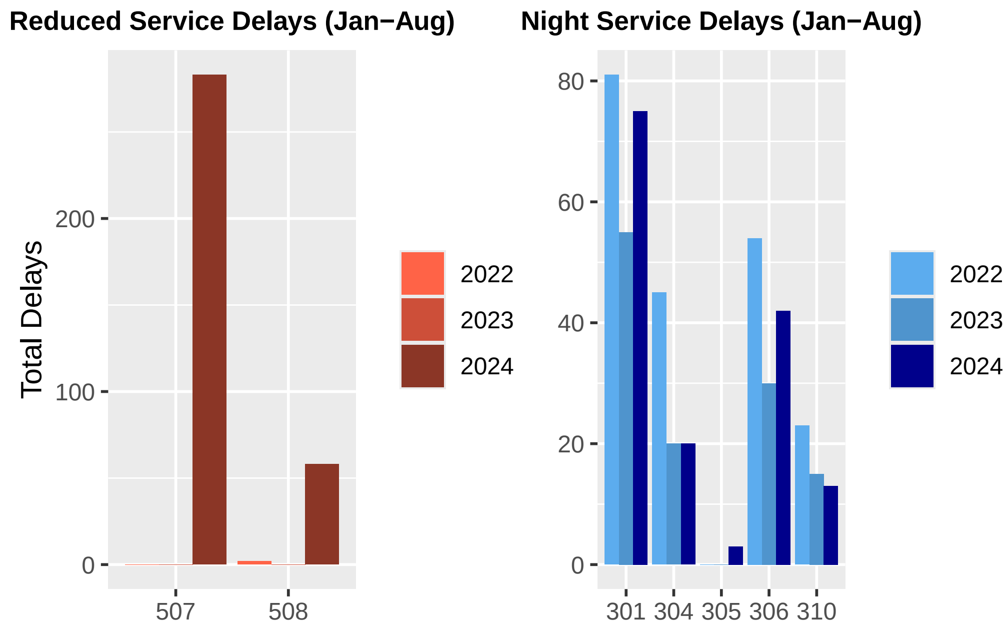

Figure 12: Total delays in January-August for reduced and night service lines by year.

Not many conclusions can be drawn from the reduced service graph given the sample size. The night service graph has a bit more information, with overall delays decreasing significantly from 2022 to 2023 before increasing again in 2024. The tiny sample size for the 305 indicates that its high average delay in Figure 8 may not necessarily be an indicator of poor performance, but rather that more data is necessary to determine a trend.

Up until now, we have mainly looked at patterns in streetcar performance over time, but there are also elements to be found in researching more specific time frames.

3.3 Specific Time Periods

For further data exploration, we look at the hour, day of the week, month, and year. Details for acquiring this data are in Appendix~C. Since the sample size for reduced and night service lines are much smaller than regular service lines, little information is gained from analyzing their specific time periods, so we will look at regular service data.

Figure 13 and Figure 14 visualize the average delay time and delay frequency for each month. Figure 15 and Figure 16 explore the same for each day of the week. Figure 17, Figure 18, and Figure 19 analyze the hourly breakdown.

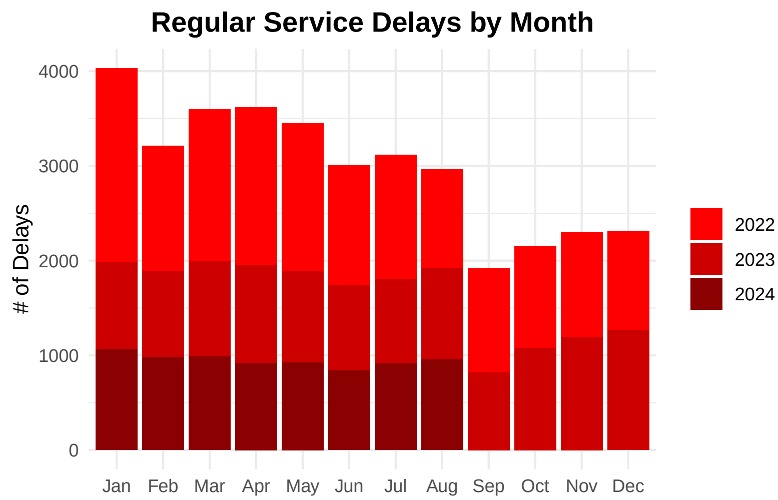

Figure 13: Total number of regular service delays by month.

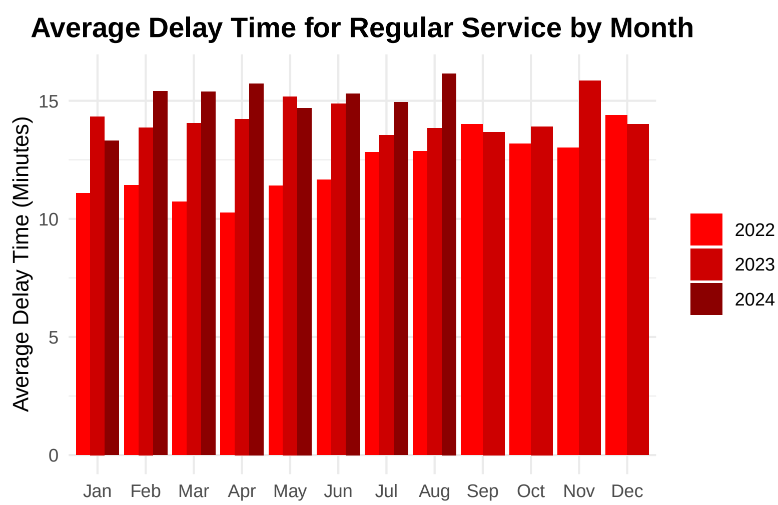

Figure 14: Average length of regular service delays by month.

The number of service delays seem to be lowest in the summer and highest in the winter, which suggests the weather may have a large affect on streetcar punctuality. The average delay time appears to be lowest at the start of the year and gradually increases as the year goes on.

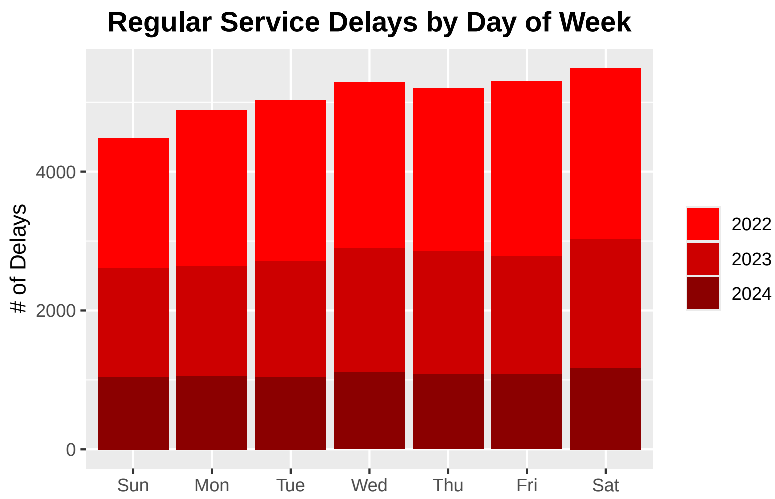

Figure 15: Total number of service delays by day of week.

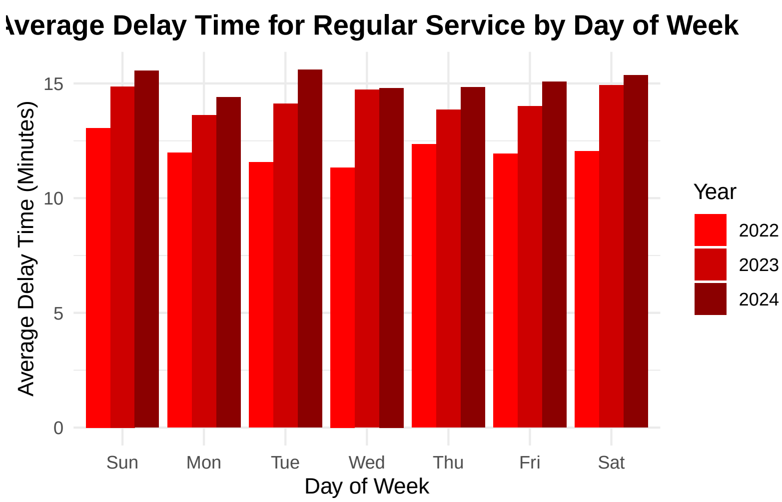

Figure 16: Average length of regular service delays by day of week.

In 2022, the number of service delays were lowest on Sunday and increased as the week went along, but this trend has disappeared; 2024 had roughly a uniform number of delays for each day of the week. The same is true for the average delay time, which seems to have no correlation with the day of the week.

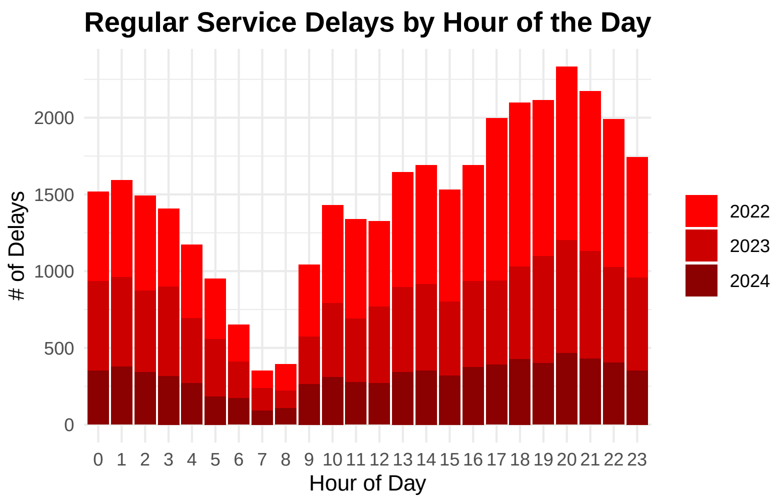

Figure 17: Total number of regular service delays by hour of the day.

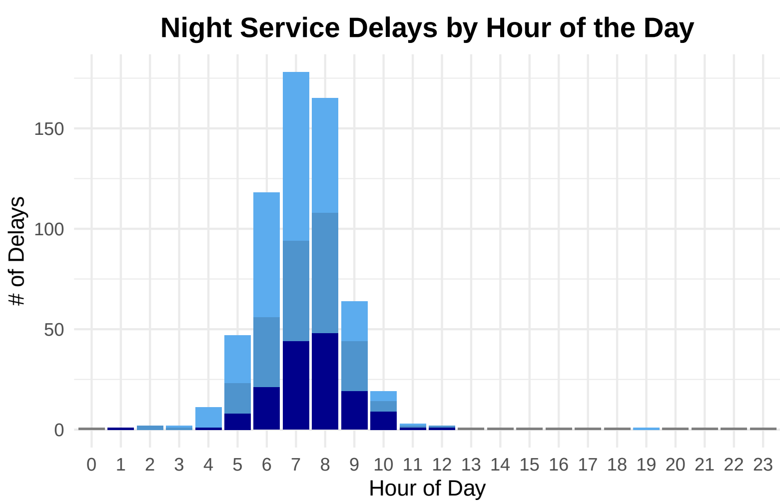

Figure 18: Total number of night service delays by hour of the day.

The number of delays for the regular service dips between 4 and 10 am, but this seems to entirely be because of reduced service volume at those hours. The night service delay total corroborates this, as the total peaks at the same time the regular service dips.

The delay peak around 6-9pm for regular service coincides with a rush hour, but this peak lessens year on year from 2022 to 2024.

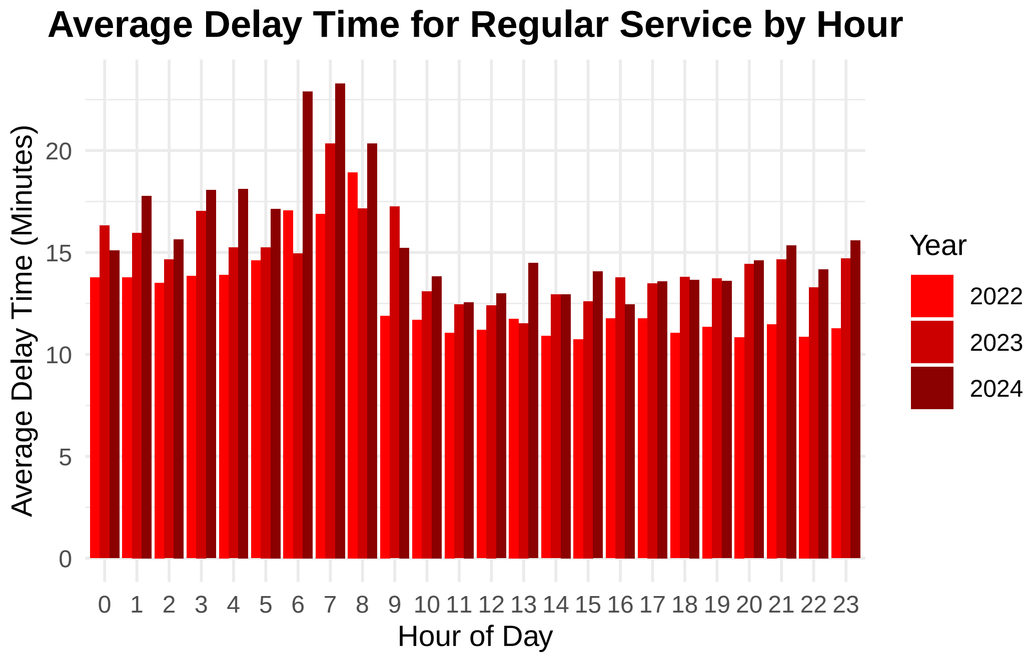

Figure 19: Average length of regular service delays by hour.

The average delay time has a clear peak around 6 to 8 am. Aside from this peak, delay time is uniform throughout.

4 Discussion

4.1 Trend of Increasing Average Delay Lengths

The analysis shows a general upward trend in average delay times each year across all service types and specific time intervals. This consistent increase could be the result of rowing traffic congestion or operational issues The exact reason for this increase is unclear; it could be from systemic issues within the streetcar network, infrastructure inefficiencies, or just a rise in population. Whatever the case, addressing these delays should be paramount; if there is no simple solution, there are ways to mitigate the problem, such as enhancing real-time communication with commuters through online platforms like Twitter or the TTCwatch app to manage expectations.

4.2 Decreasing Frequency of Delays

There is a general trend of the frequency of delays decreasing over the observed period. However, this decrease could be misleading. Since the dataset does not have total service volume, it is unclear if the reduced frequency of delays is due to operational improvements or reduced service levels. If fewer streetcars are operating due to maintenance, route cuts, or lack of demand, the lower frequency may not represent an actual improvement in streetcar reliability. Further analysis with complete service data would be helpful.

4.3 Independence of Delay Patterns from Specific Times

Delay lengths and frequencies appear to be mostly independent of the month, day of the week, or hour of the day, suggesting that there are no outlier areas where delays are a big issue, and any issue with delays cannot be stopped easily with targeted short-term fixes. This could mean that delays are driven by broader systemic factors; addressing these will require even more comprehensive research and long-term strategizing.

5 Limitations and Future Steps

- Lack of data on total trip count: The absence of data on the total number of streetcar trips for each line makes it challenging to determine whether variance in delay frequency is due to actual changes in delay patterns, or changes in the total volume of services. Future research should incorporate data on total streetcar trips to provide a more accurate context and allow for more robust conclusions about service reliability.
- Missing future data: The absence of data from September to December 2024 limits the ability to fully assess seasonal trends, particularly during the autumn months, which may impact overall year-to-year comparisons. As time goes on, more detailed data will be available, increasing our understanding of delays and whether existing trends will continue to hold.
- Lack of location data: Not having data on the specific locations of lines and stops limits the ability to explore spatial patterns and relationships between different streetcar lines, which would increase knowledge on the interactions between lines and service types. Collecting more detailed data, including line coordinates, intersections, and distances would provide a clearer picture of how delays are related.

Appendix

A Data Cleaning

This included constructing a variable to determine the type of service - regular, reduced, night - of each specific streetcar line, and filtering observations to only include streetcar lines currently in service Commission (2024). Additionally, delays under 2 minutes were removed and considered negligible, as were delays over 2 hours, as these were likely to be announced stoppages. Other information in the dataset such as station location, incident type, bound, and vehicle number were not discussed in this analysis. Data was cleaned using R Core Team (2023) and packages Wickham et al. (2019), Wickham et al. (2023), and Firke (2023). Data was analyzed with Wickham (2016), Grolemund and Wickham (2011), Wickham et al. (2019), Wickham et al. (2023), Pedersen (2024), and Xie (2014).

B Incomplete Data

One possible consideration would be to scale the 2024 data to account for a whole year's worth of observations. However, this is an unsafe estimate as it assumes delays are evenly distributed throughout the year, which is unlikely considering seasonal changes, unexpected occurrences, and planned events like construction. For best accuracy, when we directly compare between years, we will use data from January to August, which is present in all 3 years of the dataset.

C Time Data

We get the year, month, and hour from the `date` variable, and use the existing `day` variable to assign an ordinal variable `day_of_week`.

References

Arsenych, Alex. 2022. "Queen and King Streetcars Show Toronto Why the Subway Is the Better Way: It Works." *Toronto Star*. https://www.cp24.com/news/the-ttc-says-ridership-reached-a-post-pandemic-high-1.7041849.

Britannica. 2024. "Streetcar." *Encyclopaedia Britannica*. https://www.britannica.com/technology/streetcar.

Commission, Toronto Transit. 2024. "Streetcar Routes." https://www.ttc.ca/routes-and-schedules/listroutes/streetcar.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

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

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

Micallef, Shawn. 2024. "The TTC Says Ridership Reached a Post-Pandemic High." *Toronto Star*. https://www.thestar.com/opinion/contributors/queen-and-king-streetcars-show-toronto-why-the-subway-is-the-better-way-it-works/article_c8a54ecc-6d68-54f2-84fd-ccd5c41dd4ed.html.

Pedersen, Thomas Lin. 2024. *Patchwork: The Composer of Plots*. https://CRAN.R-project.org/package=patchwork.

R Core Team. 2023. *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.

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. 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.