

An Analysis of TTC Streetcar Delay Times*

Daniel Du

2024-09-24

This paper analyzes TTC streetcar delay data to understand frequencies and lengths of delays. Conclusions... Discussion point...

1 Introduction

One of the most important areas of any city's infrastructure is its public transportation system. Streetcars, once a common mode of transport in North America, <https://www.britannica.com/technology/streetcar>, dwindled in usage as cities became increasingly car-centric. Today, Toronto is the only city that still uses streetcars, with the TTC (Toronto Transit Commission) currently operating 11 day lines and 4 night lines. However, the streetcar system's inefficiency and delay frequency hinder its effectiveness and increase overall traffic. An improvement in this area would go a long way towards increasing the accessibility and sustainability of Canada's largest city.

The TTC streetcar system is relatively archaic and unreliable, which is a problem since 230,000 commuters rely on it daily <https://www.cp24.com/news/the-ttc-says-ridership-reached-a-post-pandemic-high-1.7041849>. Nearly every Torontonians has an anecdote about a lengthy streetcar delay inconveniencing them in a time of need, causing many to rely on alternative transit modes 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. Assessing the total impact and effect caused by streetcar delays should help give us a better idea of where and when its efficiencies are most pronounced.

This paper examines streetcar delay data to determine the extent of streetcar delay issues. In Section 2, we analyze public streetcar delay data from 2023 to look for trends and observations within delays. In (**Discussion?**), we discuss our findings and their real-world relevance.

*Code and data are available at <https://github.com/daniel424/TTC-Streetcar-Delays>

2 Data

Our data is sourced from the (`opendatatoronto`) R package, which provides a spreadsheet of all streetcar delays in 2023. Data was cleaned using R Core Team (2023) and packages (`tidyverse`), (`dplyr`), (`lubridate`), and (`janitor`). Data is analyzed with (`ggplot2`), (`ggpubr`), (`gridExtra`) and (`knitr`). We filtered relevant variables for our analysis, which include day of week, date and time, the line number, line type, and delay time for each observation. We also removed extreme outlier delays, which were likely to be pre-planned and made clear to commuters ahead of time. In total, the cleaned data has 12396 observations.

In (2.1) we explore the basics of the data before further analysis.

2.1 2.1 Exploring Data

Here is a sample of the data:

Sunday	2023-01-01T07:40:00Z	Regular	505	15
Sunday	2023-01-01T07:52:00Z	Regular	504	10
Sunday	2023-01-01T07:59:00Z	Regular	504	25
Sunday	2023-01-01T10:38:00Z	Regular	510	15
Sunday	2023-01-01T11:35:00Z	Regular	506	10
Sunday	2023-01-01T13:52:00Z	Regular	501	10
Sunday	2023-01-01T13:53:00Z	Regular	510	14
Sunday	2023-01-01T14:00:00Z	Regular	506	10
Sunday	2023-01-01T14:37:00Z	Regular	506	10
Sunday	2023-01-01T14:58:00Z	Regular	506	10
Sunday	2023-01-01T15:07:00Z	Regular	505	7
Sunday	2023-01-01T15:10:00Z	Regular	506	10
Sunday	2023-01-01T15:12:00Z	Regular	512	7
Sunday	2023-01-01T16:22:00Z	Regular	506	10

Figure 1: 2023 Streetcar Delays

To provide a bit more insight we can look into the summary statistics of delay times, and filter for type of service, line, and date. For example, we can see from (`table-2`) that average delay times are higher for the night service lines, and on Saturday compared to most other days of the week. Of course, this is very surface level and we analyze the data further in (2.2).

[1] "Minutes Delay Summary Statistics"

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.00	8.00	10.00	14.99	14.00	148.00

[1] "Min Delay Summary Statistics (Night Service Lines)"

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.00	15.00	22.50	25.84	30.00	126.00

[1] "Min Delay Summary Statistics (Saturday)"

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.00	8.00	10.00	15.95	15.00	145.00

[1] "Min Delay Summary Statistics (Jan-Mar 2023)"

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.00	8.00	10.00	14.84	15.00	148.00

2.2 2.2 Analysing Data

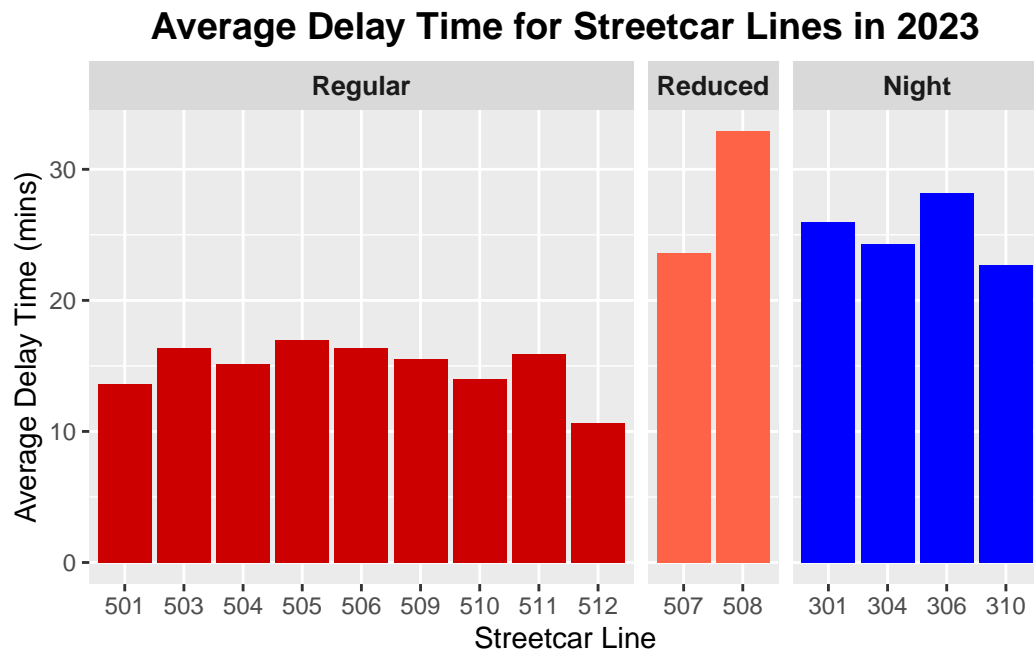


Figure 2: average delay

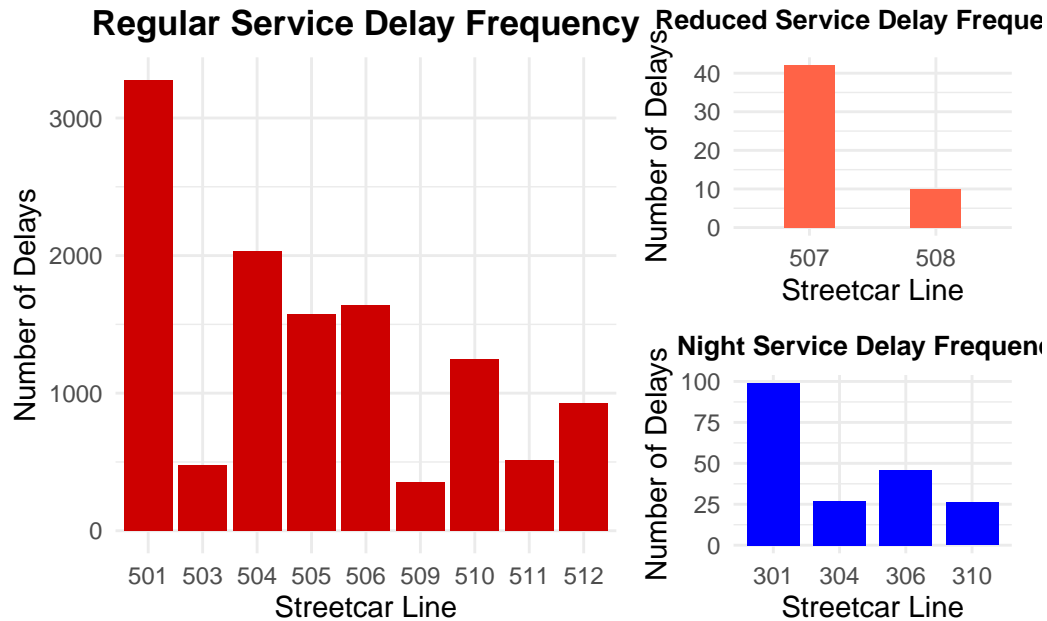


Figure 3: delay frequency

From (2.2.1?) and (2.2.2?), we can see the average delay time is greater for the reduced and night services compared to the regular. From the regular streetcars, the streetcar line with the highest average delay time is the 505, although there is very little variance between all 11 lines. There is a much greater variance in the delay frequency - the 501 and 301 (which run on the same line) have a much higher delay frequency than the rest of the lines.

```
# Extract hour and day of the week
cleaned_data <- cleaned_data %>%
  mutate(hour = hour(date), # Extract the hour from datetime
         day_of_week = wday(date, label = TRUE) # Extract the day of the week as a label
  )

regular_data <- cleaned_data %>% filter(service_type == "Regular")
reduced_data <- cleaned_data %>% filter(service_type == "Reduced")
night_data <- cleaned_data %>% filter(service_type == "Night")

# Function to create plots for day of week
create_delay_plot <- function(data, colour) {
  plot_day <- ggplot(data, aes(x = day_of_week, fill = service_type)) +
    geom_bar(fill = colour) +
    labs(title = "Delays by Day of Week",
```

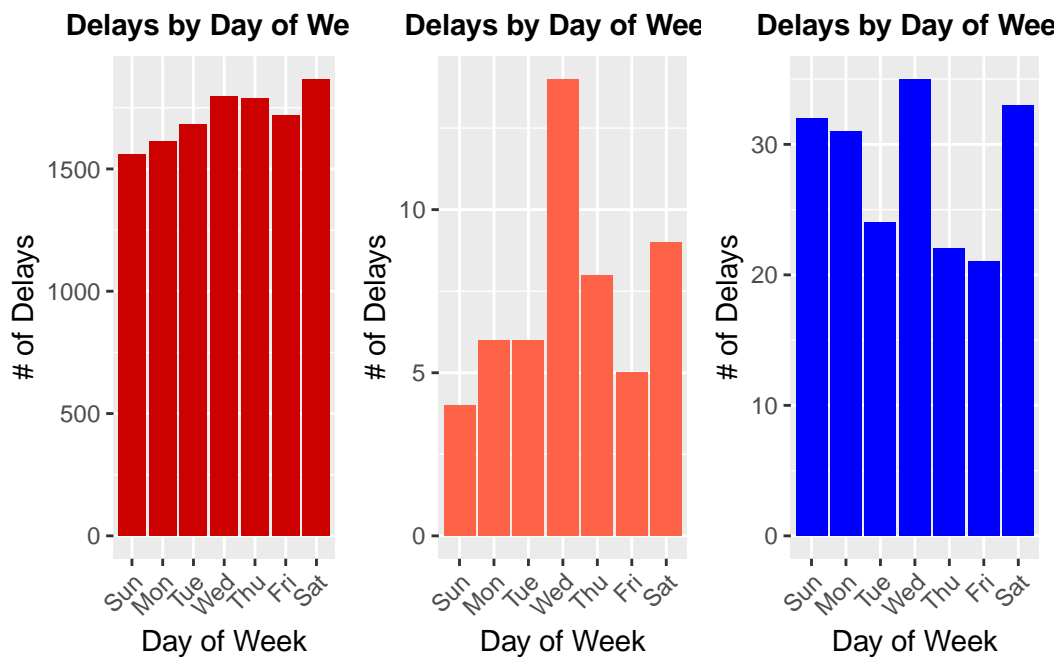
```

    x = "Day of Week",
    y = "# of Delays"
  ) +
  theme(plot.title = element_text(hjust = 0.5, size = 11, face = "bold"),
        axis.text.x = element_text(angle = 45, hjust = 1))
}

# Create plots for each service type
regular_plot <- create_delay_plot(regular_data, "red3")
reduced_plot <- create_delay_plot(reduced_data, "tomato")
night_plot <- create_delay_plot(night_data, "blue")

grid.arrange(
  regular_plot,
  reduced_plot,
  night_plot,
  ncol = 3
)

```



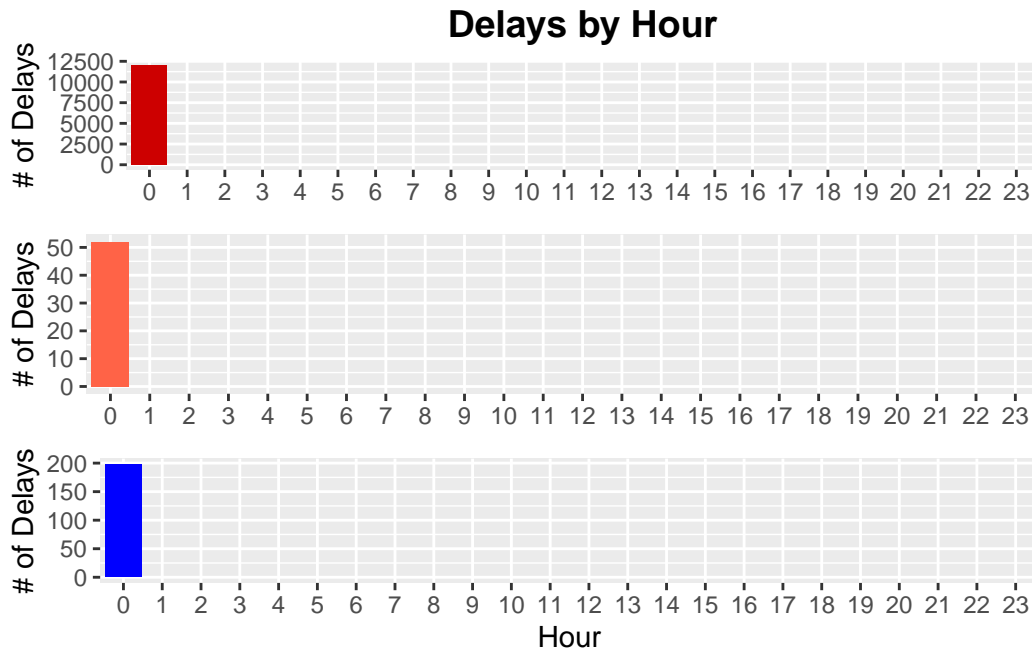
From @ the graphs, we can see that delays are most common on Wednesday among all 3 streetcar services.

```

regular_plot_hour <- ggplot(regular_data, aes(x = as.factor(hour), fill = service_type)) +
  scale_x_discrete(drop = FALSE, limits = as.character(0:23)) +
  geom_bar(fill = "red3", width=0.95) +
  labs(title = "Delays by Hour",
       x = NULL,
       y = "# of Delays"
  ) +
  theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
        )
reduced_plot_hour <- ggplot(reduced_data, aes(x = as.factor(hour), fill = service_type)) +
  scale_x_discrete(drop = FALSE, limits = as.character(0:23)) +
  geom_bar(fill = "tomato", width=0.95) +
  labs(title = NULL,
       x = NULL,
       y = "# of Delays"
  )
night_plot_hour <- ggplot(night_data, aes(x = as.factor(hour), fill = service_type)) +
  scale_x_discrete(drop = FALSE, limits = as.character(0:23)) +
  geom_bar(fill = "blue", width=0.95) +
  labs(title = NULL,
       x = "Hour",
       y = "# of Delays"
  )

grid.arrange(regular_plot_hour, reduced_plot_hour, night_plot_hour, ncol = 1)

```



From @ the graph, we can see that the regular streetcar service has delays around 3-4 pm, although the distribution of delays is fairly uniform (with the exception of early morning hours when there is less service).

3 Discussion

3.1 First discussion point

Looking at the average delay lengths, it is somewhat surprising that night service streetcars experience significant delays compared to the regular service, as one would expect less traffic to result in smoother operations. Possible reasons for this include a drop in operational support for drivers during night hours, as there are fewer staff available to manage and respond to issues and maintenance activities. Additionally, overnight road maintenance is often scheduled during these hours and may affect night service. Coupled with the reduced frequency of night service, the lengthy delays appear to be an area of concern, as passengers already have fewer options and longer wait times during the times when night service is operating. This pattern highlights the need for targeted strategies to address delays during night operations.

3.2 Second discussion point

Looking at the delay frequency by line, the frequency of delays along the 501 and 301 that operate along Queen St are by far the highest compared to their counterparts. While it is possible that its operations were less efficient due to rerouting, https://www.thestar.com/news/gta/queen-streetcar-is-about-to-be-rerouted-for-more-than-four-years-here-s-where/article_d61f5cf0-1380-50cb-b73d-33283035943e.html this does not fully account for the sheer amount of delays. This shows the need either for increased service in this area, or better scheduling and planning to increase commuter's confidence in the 501 and 301 arrival time.

todo: get comparison from 2024 data.

3.3 Third discussion point

Regular streetcar delays are most common around 3-4pm, which is outside of peak hours. The data seems to indicate that delays are less frequent before noon, which is a positive sign.

3.4 Weaknesses and next steps

Data does not include total trips, so it is difficult to determine if variance is due to delay variance or total trip amount variance.

Weaknesses and next steps should also be included.

Appendix

A Additional data details

References

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