Annual Analysis of Delays in TTC Streetcars for the year 2023*

Is there a way to make commuter life easier?

Aakash Vaithyanathan

June 4, 2025

This paper aims to study the causes of delays in Toronto Transit Commissions (TTC) streetcar service by examining the relationship between various factors like hour of day, streetcar lines and seasons. Our study found that the night shift had the higest delays by average of 25 minutes with high variability in delays observed during the hours of 9 AM to 5 PM. Interestingly, we did not observe any seasonal variations in the overall delay in streetcar and one possible reason for this could be due to our limited dataset. We beleive the results from this paper can help in better designing policies and transit schedules to improve the overall commuting experience.

1 Introduction

Toronto Transit Commission (TTC) is the biggest and most used transit authority in Toronto. It was established on September 1, 1921 and since then supports various means of ground transportation like streetcar, shuttle bus and the subway system. Despite being the most used transit system, it does come with several challenges one of the most notable one being the delays experienced between the various stops. One such transportation system we focus on in this paper is streetcars.

The delays via streetcars make it especially difficult for commuters on a daily basis. One can image the inconveniences by these delays during the harsh winter weather conditions and the struggles commuters face. In 2019, streetcars averaged delays of about 13,790 minutes each month and by 2023, this number increased by 33% (Elliott (2023)). As a result, commuters are in need for improved policies and better transit schedules to address the difficulties they face everyday. Our study found that the night shift had the higest delays by average of 25 minutes

*Code and data are available at: https://github.com/aakash2002/study_of_ttc_streetcar_delays

with high variability in delays observed during the hours of 9 AM – 5 PM. Interestingly, we did not observe any seasonal variations in the overall delay in streetcar and one possible reason for this could be due to our limited dataset.

In this paper, we aim to dive deeper into the cause of delays in TTC, how these vary by the times of day and the various seasons and what incidents are the most occurring for the delays experienced. The starter code for this paper was provided by Professor Rohan (Alexander (2024)). We make use R (R Core Team (2023)) and its supported packages like tidyverse (Wickham et al. (2019)) and dplyr (Wickham et al. (2023)). For our dataset, we use the TTC streetcar delay data for the year 2023 by OpenData Toronto (Gelfand (2022)). The reported graphs in our study are generated using ggplot (R Core Team (2024)) and the tables are generated using the gt (Iannone et al. (2024)) package. The process to perform data cleaning was done using the Janitor (Firke (2023)) package.

The remainder of this paper is structured as follows. We first give a brief overview of the dataset under dataset description which talks about the various features, the assumptions we make and the measurements done to gather the data. We present some graphs on the results extracted from our collected data found under results. A discussion of our results can be found under discussion section with the relevant limitations in our study found under limitations section. We end our study with some proposed next steps for future version of this paper.

2 Dataset Description

2.1 Overview of dataset

The following data was made available from OpenData Toronto (Gelfand (2022)) for the year 2023. The data was extracted from the server and saved locally in a raw data folder using the instructions provided by the documentation. We note that this was the only available dataset with frequent updates and data dating back to 2014. Some of the features part of this dataset are date, time, incident, min delay, bound and location.

2.2 Assumption made in dataset

We observe that our dataset has several rows where the minimum delay time is over 900 minutes or are equal to 0. These values can heavily skew the results that we may observe. Therefore, one key assumption we make in our study is that we are only interested in at most a delay by 2 hours in streetcars. Therefore, during the data cleaning step, we will limit the rows of data to 120 minutes. As a result of this filtering, we only lose about 130 data and still retain about 10723 data entries to use in our study.

Table 1: First first five rows in the cleaned dataset with selected features

(a)

Year	Month	Hour	Incident Type	Min Delay (minutes)	Streetcar Line	Season
2023	1	2	Held By	15	505	Winter
2023	1	2	Cleaning - Unsanitary	10	504	Winter
2023	1	2	Held By	25	504	Winter
2023	1	5	Security	15	510	Winter
2023	1	8	Cleaning - Unsanitary	10	501	Winter

2.3 Data preprocessing and cleaning

The raw data is undergone a series of data cleaning steps to ensure adequate usability for the study. We first use the Janitor (Firke (2023)) package to clean the data columns removing any spaces between them and using a consistent variable name convention. Next, we proceed to split the date field into year and month fields and do the same with the time field extracting the hour of day from it. Following this, we create a new column to denote the season based on the month of the year from the data. This is used for studying any seasonal variation in delays that may exist in our data.

After selecting the appropriate columns in our dataset, we proceed to drop the rows with missing values in them. Out of the total 14413 rows of data, 2399 of them contain missing values. As a consequence of dropping these rows, we do not lose any valuable information in our dataset. Therefore, no transformations are necessary to populate any missing entries. Lastly, as highlighted in the assumption section, we filter our data to only include delays up to 120 minutes.

Table 1 shows the first five rows of our data and the selected features.

Table 2: Datatypes for the selected features from the cleaned dataset

(a)

Feature	Data Type
Year	int64
Month	int64
Hour	int64
Incident Type	object
Min Delay (minutes)	int64
Streetcar Line	int64
Season	object

2.4 Measurement of units for selected features

The data is collected and maintained by TTC and is updated monthly as reported by the website. Since the data is stored as a CSV, several numerical fields like the streetcar line or delay in minutes are saved as characters instead of numbers. Fields like date are saved as appropriate format of datetime type.

The dataset was last updated September 24, 2024 ensuring consistent updates to the collected data. OpenData Toronto also provides functionality and sufficient instructions for developers to easily query their data using scripts in Python (Python (2024)), R (R Core Team (2023)) and Node.js (nodeJS Team (2024)) instead of having to manually download them.

The website doesn't provided information on how the data was sourced but we can assume that the data is recorded by taking the difference in the reported arrival time via the TTC app and the actual time the streetcar arrives at various stops. The collected data represents common types like characters, numeric and datetime. Therefore, for the purposes of this study, the procedures used to collect the data and their types are appropriate for our use and we do not require any extra information to make any assumptions about them.

Table 2 describes the data types for the selected features in our study from the cleaned dataset.

3 Results

3.1 Average Delay By Streetcar Line

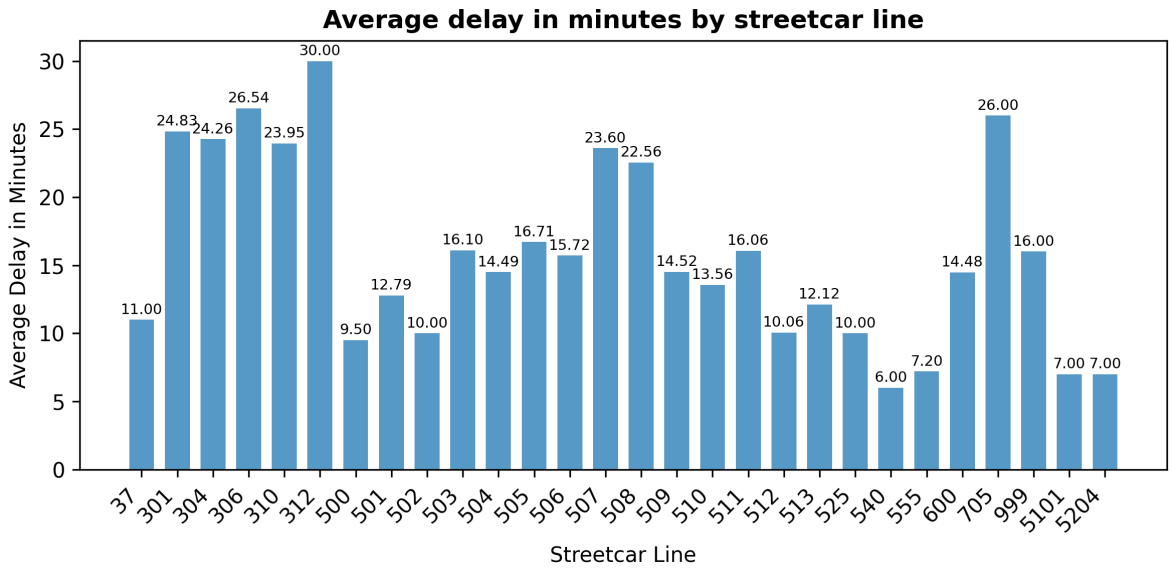

Figure 1: Average annual delay in minutes for each streetcar line in 2023

Figure 1 highlights the annual average delay in minutes for every streetcar line for the year 2023. The 300-series lines also called the Blue Network Routes (TTC Routes (2024)) are the night time streetcar lines that start at 12:00 AM. These average the most delays reaching at most 30 minutes for line 312.

The 500-series lines are the daytime lines that run from 8 AM. These streetcar average between 6 to 23.6 minutes with the longest delay experienced by line 507.

3.2 Average Delays During Dayshift

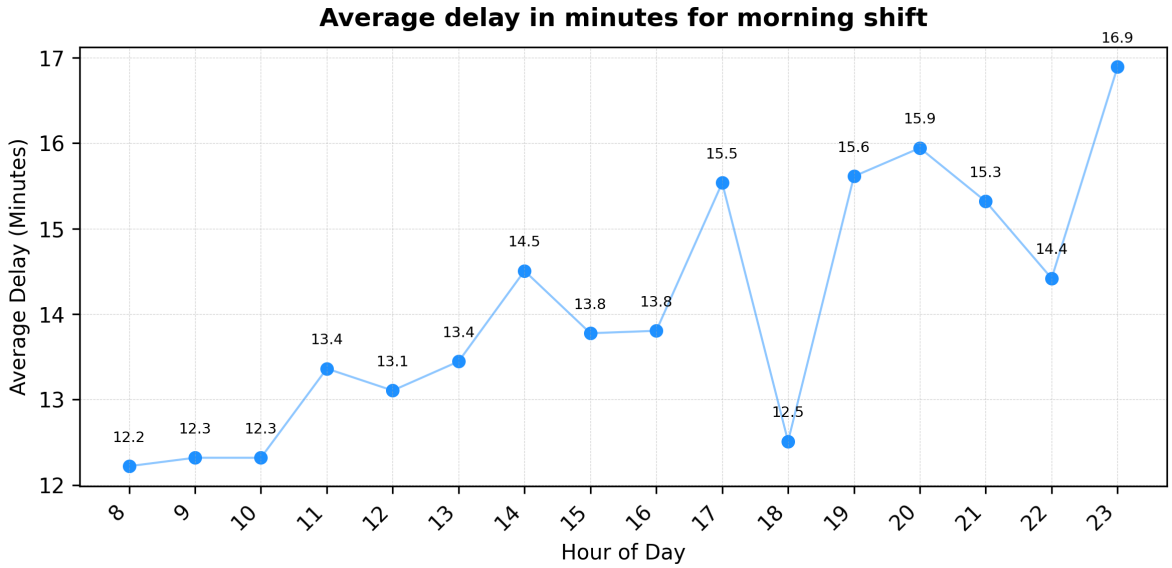

Figure 2: Average delay in minutes for dayshift in 2023

Figure 2 highlights the annual average delay in minutes experienced by the hour of day for streetcars that run during the day. We observe that there is high variability in the delays throughout the day with the minimum delay being about 12.2 minutes at 8 AM and the maximum being roughly 17 minutes at 11 PM.

3.3 Average Delays During Nightshift

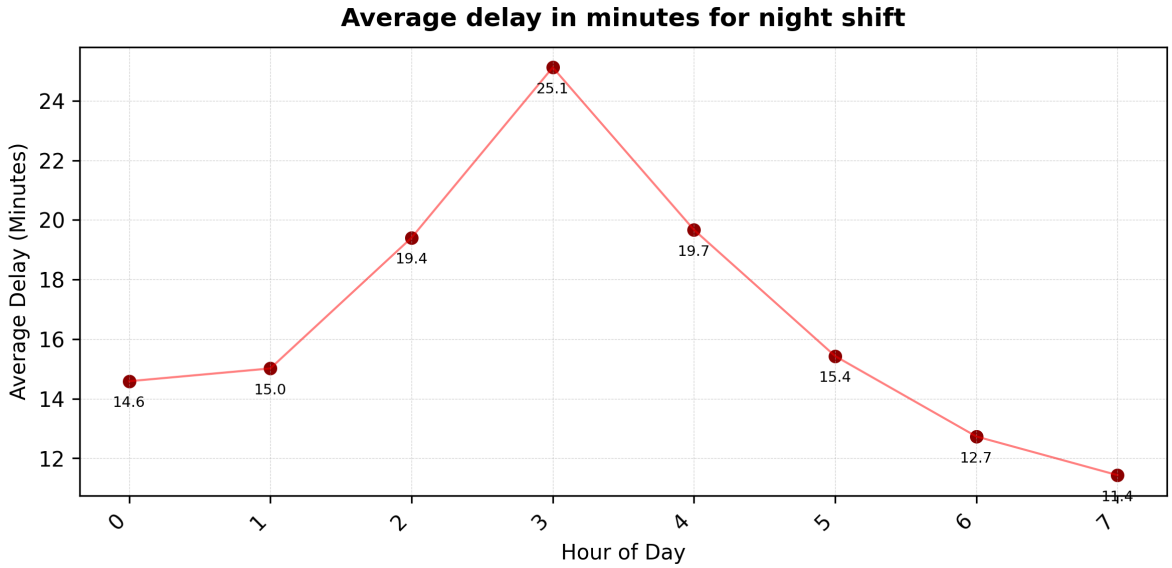

Figure 3: Average delay in minutes for nightshift in 2023

Figure 3 highlights the annual average delay in minutes experienced by the hour of day for streetcars that run during the night. We observe that there is maximum delay at 3 AM for roughly 25 minutes and the minimum delay being about 11.4 minutes at 7 AM.

3.4 Frequency of delays by incidents

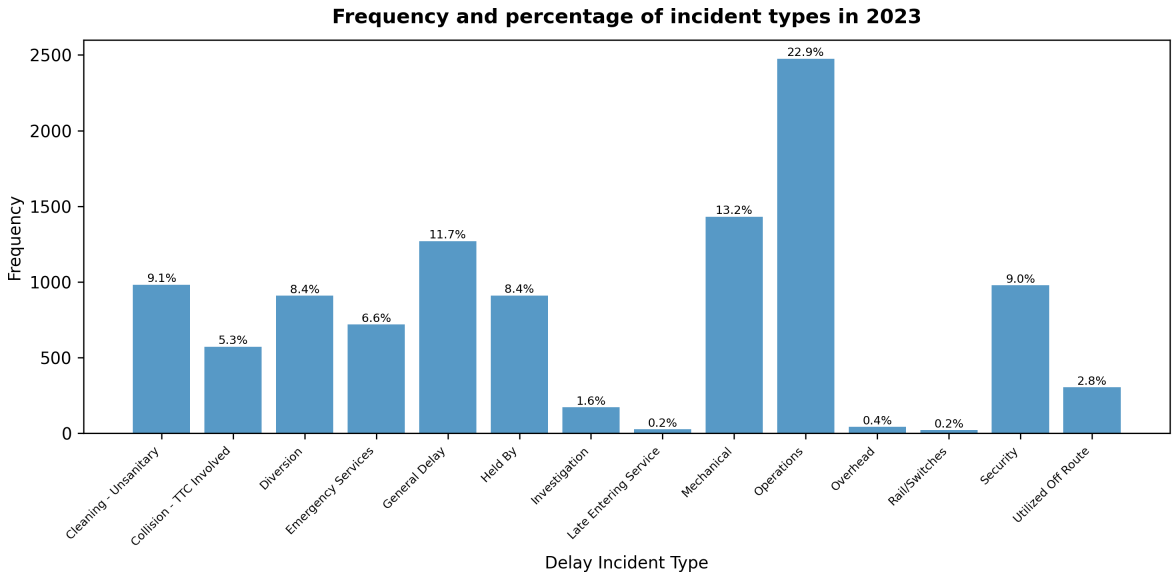

Figure 4: Frequency plot for cause of streetcar delay by incidents

Figure 4 highlights the total annual delays incurred based on incidents reported by TTC for the cause of said delay. We can that 23% of delay incidents were due to operational reasons amounting to a total of 2463 instances and rail switches were the least occurring with reported instances being only about 0.2% of all delays.

3.5 Delay by Seasons

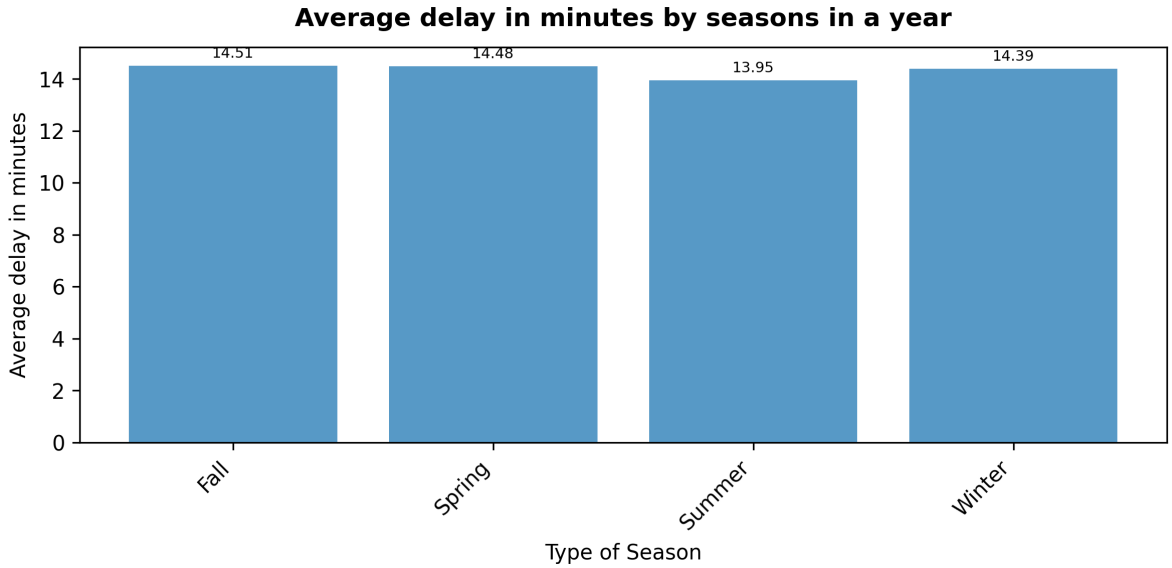

Figure 5: Average delay in minutes by seasons in 2023

Figure 5 highlights the annual average delay in minutes based on the seasons. We observe that were isn't as much variation between the streetcar delays between the fall, spring, summer and winter months with each averaging around 16 minutes delay.

4 Discussion

The above analysis provide valuable information regarding causes for TTC streetcar delays. Let us take a closer look into each of these.

4.1 How do delays vary by streetcars lines?

The results from Figure 1 shows that blue network routes (TTC Routes (2024)) streetcar incur the highest average delay of about 25 minutes. This result makes sense as compared to the day shift streetcar, there are far fewer night shift streetcar resulting in fewer drivers for various locations the streetcar may take.

4.2 Are night time streetcars busier than the day?

The results from Figure 2 showcases the variability in average delays by the hour during the morning shift working hours observed between 8 AM to 6 PM. We observe that the results are in-line with the expected behavior of longer average delays of around 18 minutes during peak hours like 8 AM or 5 PM which is usually when a large working population leave for and return from work. This can result in higher traffic hours on average causing an expected delay in streetcar arrivals. Additionally, we notice that throughout the day there are several points of interest for peak delays as seen at 11 AM or 1 PM incurring an average delay of 19 minutes. This observation could indicate the need for travel to various locations like grocery stores, perhaps to a friend's house or university students commuting for their class.

The results from Figure 3 show the average delays by the hour during the night shift working hours observed between 12:00 AM to 7 AM. These findings are what we might expect as there is fewer traffic delays during the night resulting in a more consistent decreasing delay pattern.

Despite daytime being busier than nighttime, why do streetcars at night report the highest reported delays? A possible reason for longer delays despite fewer traffic could be the fact there there are far fewer blue network routes (TTC Routes (2024)) streetcar that run resulting in fewer drivers that require to cover longer distances. The graph shows that as we get later into the night shift, we notice the delays dropping as low as 11 minutes.

4.3 What causes the delays in streetcars?

The results from Figure 4 showcase the total count of instances of delays by various incident reports in the year 2023. This graph shows that the most occurring incident was due to operational delays accounting for 23% of total reports with rail switches being the least amounting to just 0.2 %.

4.4 Is it true that winters making commuting especially challenging?

The results from Figure 5 show that on average there wasn't as much variability in delays between the various seasons in a year with summer month accounting for the longest reported average delays. We find it interesting that winter accounted to the least average delays as in 2015, 25 streetcars were knocked off-route due to harsh weather conditions (Toronto Sun (2015)). One reason for why this wasn't the case in 2023 is due to the fact that Toronto reported an average temperature above -5 degrees Celsius throughout the month of December and also reported a warmer winter season overall (Kevin Jiang (2024)).

4.5 Conclusion

By utilizing the results from Figure 1, and Figure 3, we can conclude that on average the night shift streetcars showcased the longest average delays with periods between 2 AM to 4 AM. These findings make sense as there are far fewer night shift streetcar that operate on average compared to the number of streetcars that operate during the day. Figure 2 reports that the day shift streetcars have their reported delays being high during peak hours like 8 AM, 11 AM, 1 PM, 5 PM or 8 PM. These results are in-line with the idea that hours between 9 AM to 5 AM are the general working hours in a day and likely can result in high traffic volumes for commuters.

In addition to studying the variation in delays by the working shift, their hours of day and by streetcar lines, the study also aims to examine the causes of delays and whether there are any seasonal variations that may attribute to potential delays in TTC streetcars. The findings from Figure 4 report that during the year 2023, operation delays were the most reported instances of delays of 2467. Furthermore, from Figure 5, the study finds that for the year 2023 there weren't as much variability in average delays of streetcars between the seasons. This result makes sense as during the winter season due to the harsh weather conditions, there is expected operational delays between streetcar arrivals as seen by the article in 2015 that shows that 25 streetcars were knocked off route (Toronto Sun (2015)).

The results from our study highlight important considerations on the types of delay incidents that occur in streetcars, what time of day between various working shifts are there highest reported delays and whether season has any impact on the cause of delays. By examining these results, we hope to be able to allocate resources to address the various reported incidents for delays. To address the issue of delays during peak hours, a potential solution might be to employ additional streetcars that run during these hours and by investing more resources in the night shift to reduce the long delays during late hours like 2 to 4 AM as these hours are usually dangerous to be out especially alone.

5 Limitations

Despite being able to make various conclusions as mentioned in the discussion section, our study does have certain limitations that need to be addressed as part of next steps.

One of the key limitation is due to the consequence of our dataset. It is important to highlight that the results and inferences made part of this study are for the year 2023 and cannot be generalized to the overall trend in delay patterns that we may observe. The results from Figure 5 is one such example as the result will likely have strong correlation to the observed weather data for the specific year. During the year 2023, Toronto reported a unusually warmer time resulting in less frequent and intense snowfall thus not negatively impacting streetcar track or traffic significantly. It is important to note that in order to make stronger inferences from

our data, we need to extend our dataset to account for past year's of TTC delay data collected from OpenData Toronto.

Another limitation is relating to the quality of the data that we may have. Every year TTC reports its Annual Service Plan (TTC Annual Plan (2024)) which highlights the track maintenance works and re-routes that may occur during the summer and fall months. As part of this, many routes are completely closed down and as such they switch over to alternative modes of transportation like using shuttle buses. When collecting data from various years, it is important to make sure that any lack of available data is mentioned in our study as data for these maintenance months may be missing.

6 Next Steps

In terms of the next steps, there are two ways we could extend our study to improve the findings from it.

The first is by addressing the above mentioned limitation of combining data from various years to provide a more richer quality of identified results. By combining the data from past years, we are better able to learn of any temporal relationships in the data which can lead to stronger conclusions.

The second way to improve our study is by making use of any available streetcar delay sentiment data to understand the recurring pain points and emotions that are expressed by commuters. This data when combined with existing streetcar delay data can provide richer contextual information about the types of incident delays and the sentiments expressed. Another interesting result we can hope to learn is how these sentiments vary between various seasons or times of day.

These findings can help TTC to better connect with the population and provide valuable information on how to improve policies and schedules that can help to improve the overall commuting experience for the residents of Toronto, the city they call home.

References

Alexander, Rohan. 2024. *Telling Stories with DataScience*. https://tellingstorieswithdata.com.

Elliott, Matt. 2023. "The Rise of TTC Delays, Minute by Minute — Toronto.cityhallwatcher.com." https://toronto.cityhallwatcher.com/p/chw257.

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.

Iannone, Richard, Joe Cheng, Barret Schloerke, Ellis Hughes, Alexandra Lauer, JooYoung Seo, Ken Brevoort, and Olivier Roy. 2024. *Gt: Easily Create Presentation-Ready Display Tables*. https://CRAN.R-project.org/package=gt.

Kevin Jiang. 2024. "For First Time in History, Toronto Temperatures Stayed Above -5 Celsius Degree All December." https://www.thestar.com/news/gta/for-first-time-in-history-toronto-temperatures-stayed-above--5-c-all-december/article_5e9f3194-ae3e-11ee-a035-5b280d75e673.html.

nodeJS Team. 2024. "Node.js — Run JavaScript Everywhere — Nodejs.org." https://nodejs.org/en.

Python. 2024. "Python.org." https://www.python.org/.

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

———. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Toronto Sun. 2015. "Cold Weather Knocks Old Streetcars Out of Service — Torontosun.com." https://torontosun.com/2015/01/07/cold-weather-knocks-old-streetcars-out-of-service.

TTC Annual Plan. 2024. "2024 Annual Service Plan." https://www.ttc.ca/about-the-ttc/projects-and-plans/2024-Annual-Service-Plan.

TTC Routes. 2024. "Routes & Schedules at Ttc.ca." https://www.ttc.ca/routes-and-schedules.

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.