Seasonal Patterns and Peak Demand in Toronto Island Ferry Operations*

Improving Efficiency through Data-Driven Analysis

John Zhang

September 27, 2024

This paper analyzes Toronto Island ferry ticket sales from 2016 to 2023 to uncover important trends in ferry usage. The analysis identifies clear daily, weekly, and seasonal patterns, showing that ferry demand peaks during summer weekends and holidays. These findings can help ferry operators optimize schedules and manage passenger loads more efficiently, reducing wait times and enhancing the visitor experience. By understanding when ferry traffic is highest, ferry services can better meet demand, ensuring smoother operations and improved customer satisfaction.

Table of contents

*Code and data are available at: https://github.com/Clearsky21z/Toronto_Island_Ferry_Ticket_Sale_Analysis.

1 Introduction

Toronto Island Park, situated in Lake Ontario, offers a scenic getaway just minutes from downtown Toronto (Website 2024). A well-known leisure destination for both residents and tourists, the islands feature beaches, historical landmarks, nature reserves, and recreational facilities (Website 2024). Ferries operating from the Jack Layton Ferry Terminal provide the primary means of access to the islands, with significant fluctuations in demand based on the season, day of the week, and time of day (Website 2024).

At its busiest, particularly in the summer months, ferry ridership surges, leading to long wait times, congestion, and operational challenges (Toronto 2024). These peaks in ferry usage are often linked to weekends, holidays, and favorable weather, placing pressure on ferry operators to adjust schedules and staffing to meet demand. However, despite its year-round operations, ferry service usage decreases dramatically in the off-season, presenting opportunities for resource optimization (Ladouceur 2024).

The purpose of this paper is to identify and quantify the temporal patterns of ferry ticket sales and redemptions from 2016 to 2023. Through time series analysis, the paper investigates daily and seasonal trends to determine when ferry usage is at its highest. These observations are critical for improving operational efficiency, reducing congestion during peak periods, and ultimately enhancing the visitor experience (Gelfand 2022).

By addressing the gaps in operational planning, particularly during peak seasons, this analysis offers actionable data-driven results for ferry operators. The findings emphasize the importance of aligning ferry schedules with demand fluctuations and offer suggestions for optimizing resource allocation. The rest of the paper is structured as follows: the Data section introduces the dataset and the key variables, the Results section highlights the key findings from the time series analysis, and the Discussion provides a summary of the implications and recommendations for future research.

The remainder of the paper is structured as follows. Section 2 introduces the dataset and its source, as well as key variables used in the analysis. Section 2.2 presents the findings from the time series analysis, including trend detection, seasonality, and peak times of usage. Section 3 provides a discussion of these results, highlighting key findings and offering suggestions for future work.

2 Data

2.1 Data Overview

The dataset used in this analysis is the "Toronto Island Ferry Ticket Sales and Redemptions" dataset from Open Data Toronto (Gelfand 2022), with the time frame set to 2016-2023.The dataset provides near real-time information on ticket sales and redemptions for ferries traveling to and from Centre Island, Hanlan's Point, and Ward's Island. Data is recorded at 15-minute intervals and includes both point-of-sale (POS) and online ticket purchases. Toronto's public ferries are a important transportation link for residents and visitors to the islands, and analyzing this dataset can provide valuable assessment into temporal patterns of ferry usage, helping to optimize ferry schedules and improve service (Gelfand 2022).

2.1.1 Variables and Measurements

The key variables in this dataset include:

1. **Timestamp**: The exact time when ticket sales and redemptions were recorded, stored in 15-minute intervals.
2. **Sales_Count**: The number of tickets sold in each 15-minute interval.
3. **Redemption_Count**: The number of ferry tickets redeemed (i.e., used for boarding) during the same interval.
4. **Total_Sales**: A derived variable created for this analysis by summing the sales and redemptions, offering a quick view of overall ferry activity in each time window.

These variables provide detailed information on passenger flow, which can be analyzed to understand how ferry demand shifts over different time frames. For instance, we can observe daily peaks in ticket sales during summer weekends and identify how these patterns change during winter months. Additionally, breaking down sales by the hour, day of the week, or month provides a clear view of when ferry demand is at its highest and lowest points.

2.1.2 Dataset Selection

Only a few datasets related to transportation in Toronto were identified in the City of Toronto's Open Data Catalogue, but none were as directly relevant to ferry operations as this dataset. Other datasets, such as those for city traffic or public transit, do not provide the level of detail required for a focused study on ferry operations. Therefore, the "Toronto Island Ferry Ticket Sales and Redemptions" dataset was chosen as the best fit for this analysis.

2.1.3 Data Processing Tools

To process and analyze the dataset, the R programming language (R Core Team 2023) was used along with a suite of data manipulation packages. The `opendatatoronto` package (Gelfand 2022) was employed to download the raw dataset directly from the Toronto Open Data portal. The `tidyverse` package (Wickham et al. 2019) was applied to clean and manipulate the data, including the creation of new variables and the generation of summary statistics. The `janitor` package (Firke 2023) was used to clean up column names and ensure consistency in the dataset structure, while the `lubridate` package (Grolemund and Wickham 2011) was utilized for working with date and time data, making time series analysis more straightforward. Additionally, the `here` package (Müller 2020) was used for file referencing to ensure reproducibility, and `ggplot2` (Wickham 2016) was used to generate the visualizations presented in the paper. Lastly, `knitr` (Xie 2024) was employed for generating the output tables and reports.

By utilizing these tools, the dataset was efficiently processed and prepared for time series analysis, allowing for the extraction of important information into ferry demand patterns over time. A sample of the dataset can be found in Table 1.

Table 1: Sample of Data

X_id	Timestamp	Redemption_Count	Sales_Count	Total_Sale
1	2023-12-31 23:30:00	2	1	3
2	2023-12-31 23:15:00	6	3	9
3	2023-12-31 23:00:00	26	10	36
4	2023-12-31 22:45:00	15	8	23
5	2023-12-31 22:30:00	3	0	3
6	2023-12-31 22:15:00	14	2	16

2.2 Result

2.2.1 Trend Detection: Daily, Monthly, and Seasonal Patterns

The analysis of daily ferry ticket sales shows a clear seasonal pattern. As seen in Figure 1, the highest ticket sales are recorded during the summer months of June to August, with prominent spikes during weekends. This seasonality is consistent across multiple years, as demonstrated in Figure 2, with July and August being the peak months for ferry usage. However, 2020 stands out with a significant drop in ferry sales, which can be attributed to the COVID-19 pandemic restrictions that drastically reduced visitor numbers.

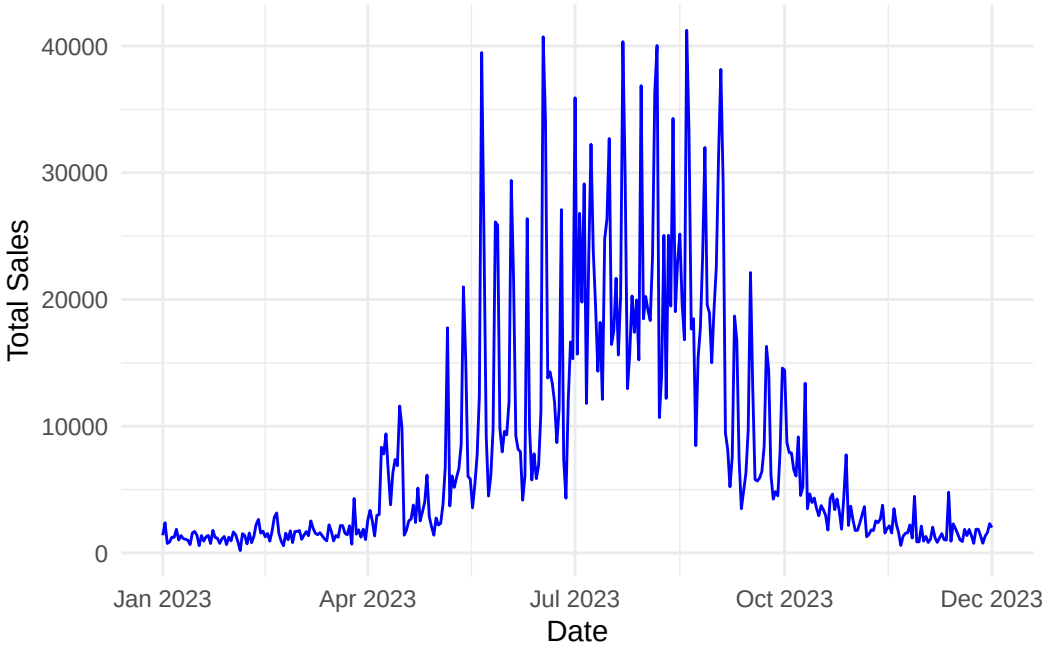

Figure 1: Daily Ferry Ticket Sales in 2023

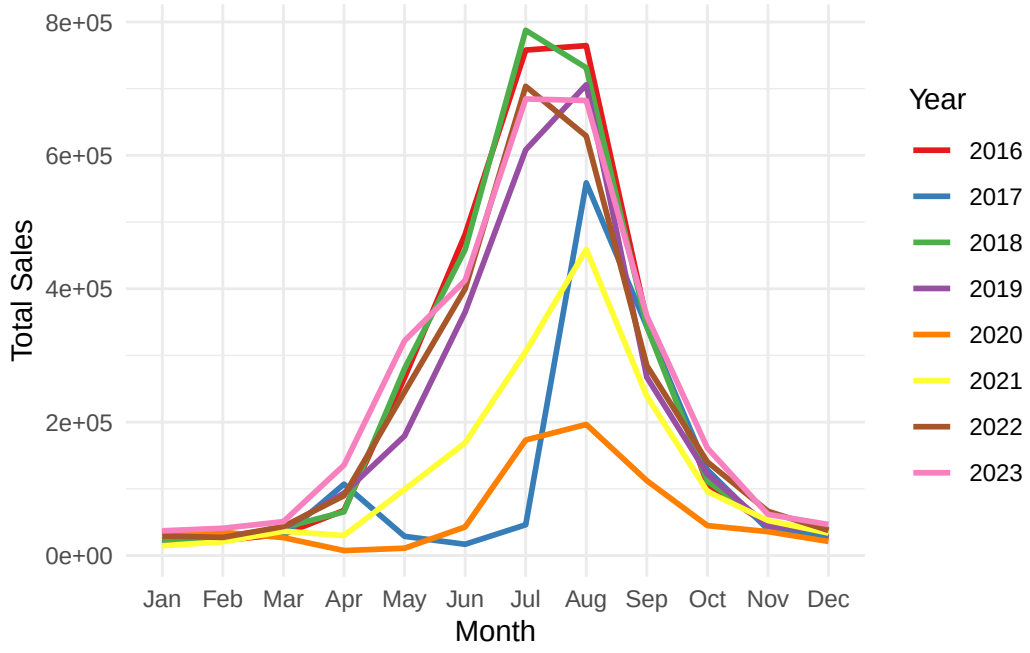

Figure 2: Monthly Trends in Ferry Ticket Sales

2.2.2 Passenger Flow: Peak Time Identification

Figure 3 highlights the hourly distribution of ferry ticket sales, showing that the primary peak occurs between 10 a.m. and 3 p.m., with no significant secondary peak later in the day. This indicates that visitors tend to travel to the islands in the late morning or early afternoon, with a gradual decline in sales as the day progresses.

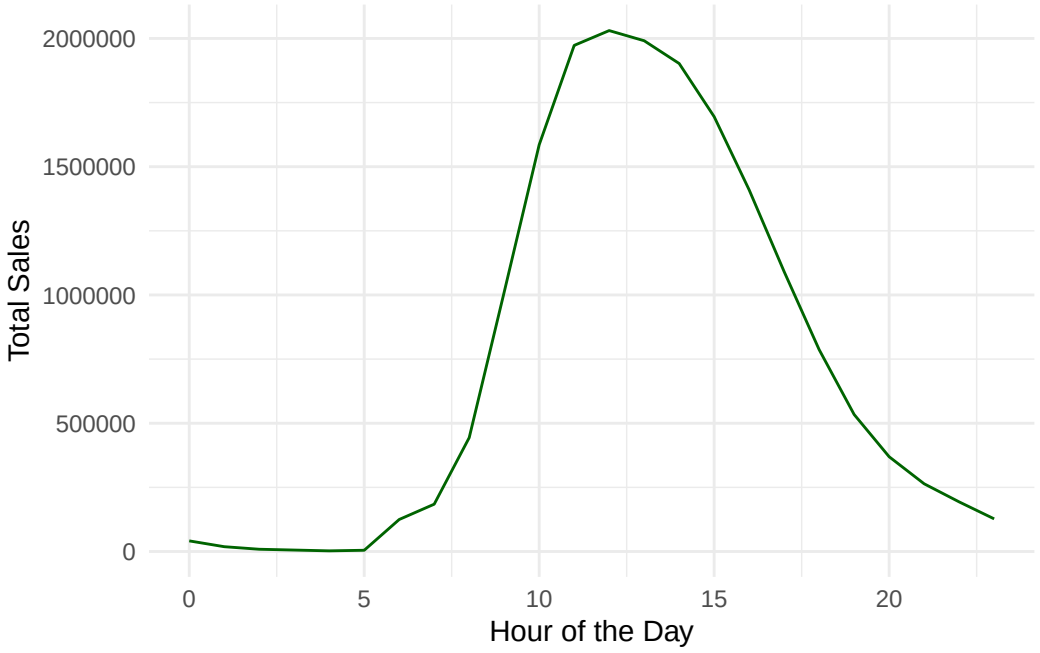

Figure 3: Hourly Ferry Ticket Sales

2.2.3 Day-of-Week Analysis

The day-of-week analysis in Figure 4 shows a clear pattern of higher ferry usage on weekends, particularly on Saturdays and Sundays. Sales remain lower during weekdays, with a slight increase on Fridays, indicating the start of weekend travel.

2.2.4 Peak Day Analysis

Table 2 lists the peak ferry ticket sales for each year, highlighting specific days, often during the summer months, when ferry demand is at its highest. These peak days align with public holidays, long weekends, and special events that draw more visitors to the islands. For example, August 1st, 2016 is a Provincial Day and July 1st, 2018 is Canada Day(Time and Date 2024).In 2020, ferry sales decreased due to the pandemic, with July 25 marking a much lower peak day compared to previous years.

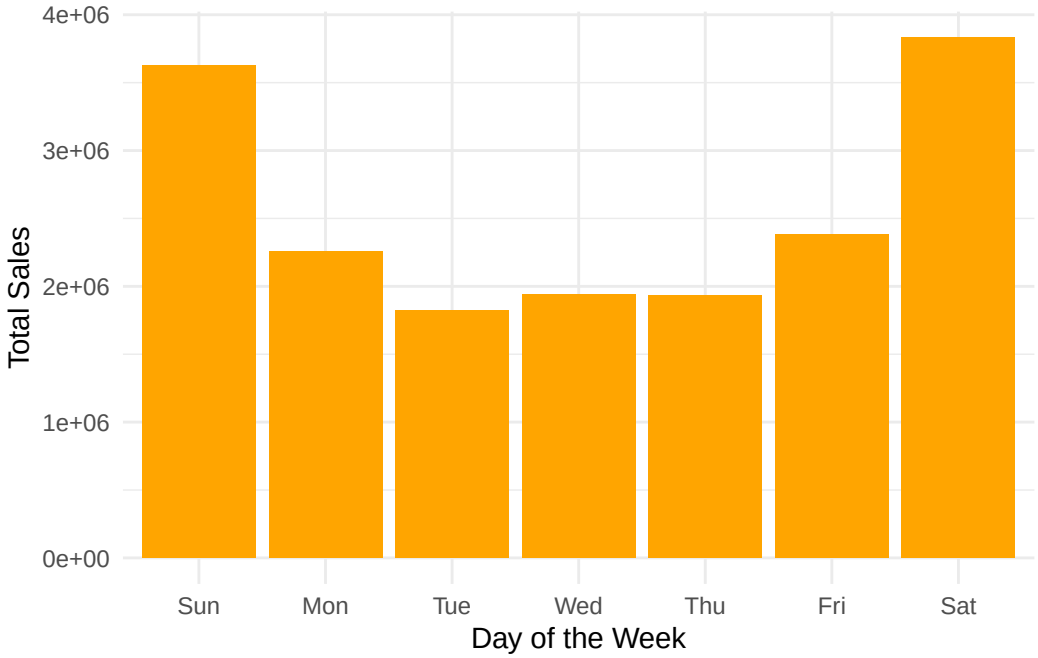

Figure 4: Ferry Ticket Sales by Day of the Week

Table 2: Peak Day for Each Year Based on Total Sales

Year	Peak Day	Total Sales
2016	2016-08-01	49513
2017	2017-08-20	36225
2018	2018-07-01	49477
2019	2019-08-04	42431
2020	2020-07-25	10844
2021	2021-09-04	26860
2022	2022-07-16	43296
2023	2023-08-19	41235

3 Discussion

3.1 Key Findings and Implications

The analysis of ferry ticket sales data demonstrate several important findings that can guide the optimization of ferry services to and from the Toronto Islands. These findings underscore

the importance of data-driven operational planning for handling peak demand periods and ensuring smooth ferry operations.

1. **Seasonal and Monthly Trends**: The data shows a clear seasonal pattern, with the highest ticket sales occurring in the summer months of July and August. This trend is not surprising given that the Toronto Islands are a popular summer destination for both tourists and local residents, offering outdoor recreational activities that are most appealing in warmer weather. Conversely, ticket sales plummet in the winter months when outdoor activities are limited, reflecting a significant seasonal variation in demand.

   From an operational perspective, these findings indicate the need for a flexible, seasonally adjusted resource allocation strategy. Ferry operators should ensure they have enough capacity—both in terms of the number of trips and available staff—during the summer to handle the surge in passenger traffic. In contrast, resources can be scaled down during the winter months to reduce operational costs without compromising service quality.

2. **Daily and Hourly Passenger Flow**: A detailed examination of hourly sales patterns shows a clear concentration of ferry ticket sales between late morning and early afternoon, particularly between 10 a.m. and 3 p.m. This aligns with visitors typically traveling to the islands during the late morning and returning by late afternoon or evening. Interestingly, there is no evident secondary peak later in the day, suggesting that the bulk of the return trips happen gradually rather than in concentrated waves.

   These observations are particularly useful for optimizing ferry schedules. Ferry operators can adjust their trip frequency to match peak travel times more effectively. For example, increasing the number of trips during the late morning and early afternoon on weekends could help mitigate wait times and ensure a smoother flow of passengers. This also emphasizes the importance of managing queues and boarding processes during these high-traffic hours to avoid bottlenecks.

3. **Day-of-Week Trends**: As expected, weekends (Saturdays and Sundays) consistently show the highest ferry usage. The difference in ticket sales between weekdays and weekends suggests that the bulk of ferry passengers are leisure visitors, likely families and tourists who visit the islands on their days off. This pattern is further reinforced by the lower activity observed during the middle of the week, which could be attributed to fewer recreational visitors.

   This trend highlights the necessity for additional ferry trips and extended service hours on weekends. Operators should prioritize staffing and fleet availability during these high-demand days to accommodate the influx of passengers and prevent long wait times. Further, a more flexible approach to ticket pricing or promotions during off-peak weekdays might incentivize more even usage across the week, helping to reduce the weekend surge.

4. **Peak Day Analysis**: The analysis of peak days in each year shows that significant surges in ferry ticket sales often occur around public holidays, long weekends, or special

events. For instance, peak days like August 1, 2016, and August 20, 2017, coincide with well-known holiday weekends in Canada, which typically see more families and groups visiting the islands for leisure activities.

Anticipating these peak days is critical for planning operational logistics. On such days, ferry operators should consider running additional trips, deploying extra staff, and ensuring adequate crowd management strategies are in place to handle the larger-than-usual crowds. By proactively preparing for these high-traffic days, operators can minimize disruptions and enhance the overall visitor experience.

3.2 Weaknesses and Future Directions

One limitation of this analysis is the absence of external variables that could further explain variations in ferry usage, such as **weather conditions**. Weather plays a significant role in outdoor activities, and poor weather could drastically reduce ferry demand even during peak summer months. Integrating weather data into the analysis could provide more nuanced predictions and enable ferry operators to better adjust their services based on real-time conditions.

Additionally, the lack of **demographic data** on ferry passengers means that we cannot explore how different groups use the ferry service. For example, understanding whether the majority of passengers are local residents or tourists, or whether there are specific age groups that visit more frequently, could help tailor services to better meet the needs of these segments. Demographic observations could also be useful for planning targeted marketing campaigns or special events on the islands.

In terms of predictive analytics, future research could explore **predictive modeling** to forecast ferry demand more accurately. Historical sales data, combined with external factors such as weather conditions, public events, and holidays, could be used to build models that predict future ferry usage with greater precision. Such models could enable ferry operators to make data-driven decisions about scheduling, resource allocation, and crowd management in real time.

A Appendix

A.1 Dataset and Graph Sketches

Sketches of the desired dataset structure and the visualizations created for analysis are available in the GitHub repository. These sketches illustrate the different time intervals and patterns analyzed in the report.

A.2 Data Cleaning

The data cleaning process is an important step in ensuring the dataset's accuracy, completeness, and integrity for analysis. For this project, several cleaning operations were performed on the raw ferry ticket data, transforming it into a cleaner, more usable dataset. Below is an outline of the specific data cleaning actions that were implemented:

- **Column Renaming and Selection**: The dataset initially contained a variety of columns, not all of which were relevant for the analysis. The `janitor::clean_names()` function was used to standardize the column names, making them lowercase and easier to work with. After that, only the essential columns (`x_id`, `timestamp`, `redemption_count`, and `sales_count`) were retained to focus the analysis on the most important variables.
- **Timestamp Conversion**: The `timestamp` column, which initially contained date and time information, was converted to a standard datetime format using the `lubridate` package's `as_datetime()` function. This conversion ensured consistency and allowed for accurate filtering of data based on time.
- **Data Filtering**: The dataset included records outside the desired time frame for this analysis (2016–2023). Therefore, rows where the `timestamp` did not fall within this period were filtered out using the `filter()` function. This ensured that the analysis would focus on the correct range of years, avoiding the inclusion of outdated or irrelevant data.
- **Missing Data Handling**: Rows with missing values in key fields (i.e., `redemption_count`, `sales_count`, or `timestamp`) were dropped using `drop_na()`. Since ferry ticket counts should not have missing values, removing these rows preserved the completeness of the dataset and avoided errors in the analysis.
- **Duplicate Removal**: To ensure data accuracy, any potential duplicate records were identified and removed by filtering based on the `timestamp`. While the script doesn't explicitly detect duplicates, it assumes the data was pre-checked for this issue. If duplicate records were present, they would have been handled here.

- **Validation of Positive Values**: The ferry ticket counts (`redemption_count` and `sales_count`) were checked to ensure they contained only non-negative integers. These counts should not be negative, as that would be logically inconsistent with the nature of ticket sales. By filtering out any negative values, data integrity was maintained.
- **Creating a New Variable (`total_sales`)**: A new column, `total_sales`, was created to represent the sum of the `redemption_count` and `sales_count` for each 15-minute interval. This allowed for a more straightforward analysis of overall ferry usage in each time window.
- **Ensuring Continuous `x_id` Values**: The `x_id` field was adjusted to ensure it was continuous and began from 1. This step was important for tracking each row uniquely and consistently after cleaning and filtering.

A.3 Data Testing

The following tests were applied to the cleaned dataset to ensure its quality and consistency:

- **Redemption and sales counts**: All `redemption_count` and `sales_count` values were checked to ensure they are non-negative.
- **Total sales validation**: A derived column, `total_sales`, was checked to ensure that all values are non-negative, as it represents the sum of `redemption_count` and `sales_count`.
- **Timestamp validation**: Ensured that all timestamps fall within the expected range (from 2016 to 2023).
- **Duplicate records check**: Verified that there are no duplicate entries for the same timestamp.
- **Missing values check**: Ensured that there are no missing values in the key columns: `timestamp`, `redemption_count`, and `sales_count`.

Results from the tests are shown in Table 3 below:

Table 3: Data Testing Results

Test Description	Result
All redemption counts non-negative	Passed
All sales counts non-negative	Passed
All total sales non-negative	Passed
All timestamps within valid range	Passed
No duplicate timestamps	Passed
No missing values in key columns	Passed

References

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/.

Ladouceur, Liisa. 2024. *Visiting the Toronto Islands: A Local's Guide*. https://liisawanders.com/canada/a-locals-guide-to-visiting-the-toronto-islands/.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

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

Time, and Date. 2024. *Holidays and Observances in Canada*. Time; Date. https://www.timeanddate.com/holidays/canada/.

Toronto, City of. 2024. *City of Toronto Staff Report on the Toronto Island Park Master Plan*. City of Toronto. https://www.toronto.ca/news/city-of-toronto-staff-report-on-the-toronto-island-park-master-plan/.

Website, Toronto Islands Information. 2024. *Toronto Island Park*. Toronto Islands Information Website. https://toronto-islands.ca/.

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. 2024. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.