Environmental Factors Affecting Water Quality at Toronto Beaches*

Investigation into Seasonal Trends(2010-2023)

Ziyuan Shen

September 23, 2024

This paper investigates the environmental factors that influence water quality at Toronto beaches using summer data collected from 2010 to 2023. The research shows key variables such as rainfall, water temperature and waterfowl populations that significantly impact water quality safety warnings and public health. The study found that rainfall and water temperature fluctuations were strongly associated with beach warnings, emphasizing the necessity for improved safety measures. This research will help improve management strategies for Toronto's beaches and ensure a safer recreational experience for residents and visitors.

Table of contents

*Code and data are available at: https://github.com/Serena-SHEN1011/Toronto-Beach-Data

1 Introduction

Toronto's beaches are important recreational destinations for residents and visitors, especially during the summer period. However, the safety and quality of these beaches are largely influenced by everyday environmental conditions such as wind speed, air and water temperatures, and the presence of wildlife. These factors not only affect beach availability, but also have significant public health implications. For example, a study in PLOS ONE (Young et al. (2023)) investigated microbial contamination in recreational waters. Understanding these environmental factors is critical for public health officials to issue appropriate recommendations and ensure that beachgoers are aware of potential risks.

Despite ongoing efforts to monitor and manage beach conditions, there has been limited comprehensive analysis of how these environmental variables interact and affect overall beach safety. For example, a study published in Springer explored the correlation between environmental factors and recreational water illnesses, pointing to the need for better predictive models and interventions (Saleem et al. (2023)). Similarly, research in MDPI found that rainfall, water turbidity and the presence of waterfowl were often associated with increased bacterial contamination of beach waters, which further emphasises the importance of detailed analysis of beach conditions (Sanchez et al. (2021)). However, the specific relationships between wind speed and temperature changes on Toronto beaches remain under-explored.

There are over 18,000 daily observations of beach conditions in this dataset. In the following section, I chose eight key environmental variables to examine in depth: date, beach name, wind speed, air temperature, water temperature, and waterfowl populations. By using these variables, we explore how environmental factors affect beach safety and availability. Specifically, we examined the relationship between wind speed and water quality, bird populations and water pollution risk, and how temperature variations affect the overall beach experience. This analysis will help to better plan and communicate the safety and attractiveness of Toronto's beaches. Structure is inspired mainly by Wickham et al. (2019a).

2 Data

2.1 Data Overview

To investigate water quality and beach conditions in Toronto, I found the 'Toronto Beach Observations' dataset from the Open Data Toronto portal through the library opendatatoronto (Gelfand (2022)). The dataset contains more than 18,000 observations of daily beach conditions and includes a number of various environmental variables, including air temperature, water temperature, wind speed, and waterfowl population, which play a significant role in determining beach water safety. The analysis focuses on the key variables affecting water turbidity, which is a proxy for water pollution risk. All variables are explained and described in (Section A.1) In my this report, I will focuses on 8 key variables: date, beach name, wind's speed, air temperature, wether rain or not, water temperature, waterfowl, water turbidity.

I used R(R Core Team (2023)) for all statistical computing and visualizing and associated software packages tidyverse(Wickham et al. (2019b)) to extensively for data wrangling, visualization, and manipulation. I also used dplyr(Wickham et al. (2023)) or the efficient processing of data, including filtering, summarising and selecting variables. Additionally, I leveraged kableExtra (Zhu (2024)) package to improve the tables in R Markdown documents and used (Müller (2020)) package to simplify file path management. It provides a more reliable and reproducible way to construct file paths by using the project root directory as a reference point. Similar methods have been used to assess environmental impacts on public spaces, for example, a ScienceDirect(Saleem, Edge, and Schellhorn (2022)) study examined the impact of environmental stressors on the availability of urban green space. My analyses explored how environmental factors affect the safety and availability of beaches over time, with a particular focus on relationships such as the effect of wind speed on water turbidity and the impact of bird populations on pollution risk. This exhaustive approach provides valuable insights for public health officials and beach managers, helping to enhance the decision-making process regarding beach proposals.

2.2 Cleaned Data

During the data cleaning process, I took several important steps to ensure that the dataset was suitable for analysis. First, all rows containing missing values (NA) were removed. This decision was made to avoid any potential biases or inaccuracies that might result from incomplete data, as missing values could skew the results of the statistical analyses. By removing these rows, the dataset remained consistent and could be used for further analysis without introducing unnecessary complexity in dealing with missing data.

At the same time, several variables were removed because they were either redundant or irrelevant to the focus of this analysis. Wind direction was excluded because it has less impact on water conditions than wind speed. Similarly, rainfall was replaced by a simplified binary

Table 1

Date	Beach	Wind (km/h)	Air Temp (°C)	Rain	Water Temp (°C)	Fowl Count	Turbidity (NTU)
2010-06-04	Hanlan's Point Beach	10	22	Yes	17.5	30	1.70
2010-06-06	Hanlan's Point Beach	20	12	Yes	13.0	18	4.00
2010-06-16	Hanlan's Point Beach	26	16	Yes	17.5	36	1.00
2010-06-22	Hanlan's Point Beach	15	22	Yes	17.1	17	0.70
2010-06-24	Hanlan's Point Beach	7	20	No	19.3	12	4.50
2019-07-17	Sunnyside Beach	4	20	Yes	18.0	3	1.70
2021-08-12	Sunnyside Beach	15	22	Yes	18.0	32	1.56
2024-06-06	Sunnyside Beach	8	17	Yes	20.0	12	1.15
2019-07-30	Marie Curtis Park East Beach	9	18	Yes	11.0	12	2.38
2019-07-30	Sunnyside Beach	9	18	Yes	14.0	50	2.88

Displaying the first and last 5 rows of seven key variables

rainfall variable that indicates whether or not it rained on a given day. This binary approach was chosen to simplify the analysis without compromising the central idea that rainfall affects water quality. The role of waves was removed. This is because, although wave intensity affects beach conditions, its overall effect is not as important to the current discussion of weather, waterfowl, and water quality.

In addition, I have carefully selected nine key variables that were deemed most valuable to this analysis. These variables included date (the specific date on which observations were recorded) and beach name (identifying the different Toronto beaches on which environmental data were collected). Wind speed and air temperature were listed as key weather-related variables, as these factors have a significant impact on beach conditions and overall water quality. Water temperature was also considered critical because of its impact on both the beachgoer experience and biological activity in the water. Waterfowl population and turbidity were chosen because they are directly related to water quality. Waterfowl can cause pollution, while turbidity measures water clarity and is often used as an indicator of pollution. Finally, the presence or non-presence of rain was included to capture binary rainfall scenarios, as rainfall can affect water clarity and carry pollutants into the beach environment.

I cleaned and prepared a subset of the data as shown in Table 1 below. I showed the first and last 5 rows of the key variables in this table.

Table 2

Variable	Mean	SD	Min	Median	Max
WindSpeed	13.3	7.2	1.0	11.0	50.0
AirTemp	19.9	4.5	7.0	20.0	33.0
WaterTemp	15.6	4.2	3.0	16.0	27.0
FowlCount	32.3	41.0	1.0	20.0	450.0
Turbidity	4.1	11.3	0.1	1.7	214.6

2.3 Data Visualization

Air temperature is a key indicator of beach conditions. Figure 1 visualizes the distribution of air temperature at Toronto beaches and the count of observations for different temperature ranges. The x-axis represents air temperature in degrees Celsius (°C), while the y-axis represents the count of observations. The distribution is approximately normal, centered around 20°C, indicating that most observations were recorded in this temperature range. There are fewer observations in the lower and higher temperature ranges, with temperatures ranging from around 5°C to nearly 35°C. The bar heights reflect the number of observations within each 2°C bin, showing that the majority of temperatures fall between 15°C and 25°C, typical of summer conditions. The steelblue color used for the bars provides a clear contrast against the black outlines, making the data easily distinguishable. This histogram effectively summarizes the air temperature data, revealing that Toronto beaches experience mild to warm temperatures during the observation period.

Wind speed affects both the comfort of beachgoers and water quality. Figure 2 illustrates the average wind speed at different Toronto beaches. The x-axis represents the average wind speed in kilometers per hour (km/h), while the y-axis lists the names of the beaches. The chart is horizontally oriented for clarity, with the wind speed values increasing from left to right.The lightblue bars represent the average wind speed at each beach, and the bars are arranged in descending order of wind speed. The chart highlights the variability in wind conditions across different beaches, suggesting that some beaches are more exposed to wind, potentially affecting people's experiences and water conditions.

Water temperature varies between beaches, which may influence swimmers' comfort and number of waterfowl in the water. Figure 3 shows the average water temperature at different Toronto beaches, with the x-axis representing the average water temperature in degrees Celsius (°C) and the y-axis listing the beach names. The chart is color-coded to distinguish each beach, and the bars are arranged in descending order of average water temperature.

Figure 4 shows the total waterfowl count on rainy versus non-rainy days. The x-axis represents the presence or absence of rain, with three categories: "N/A" (unknown rain status), "No" (non-rainy days), and "Yes" (rainy days). The y-axis represents the total number of waterfowl observed. The chart indicates a noticeable difference in waterfowl presence based on weather conditions, with more waterfowl observed on non-rainy days compared to rainy ones. We can

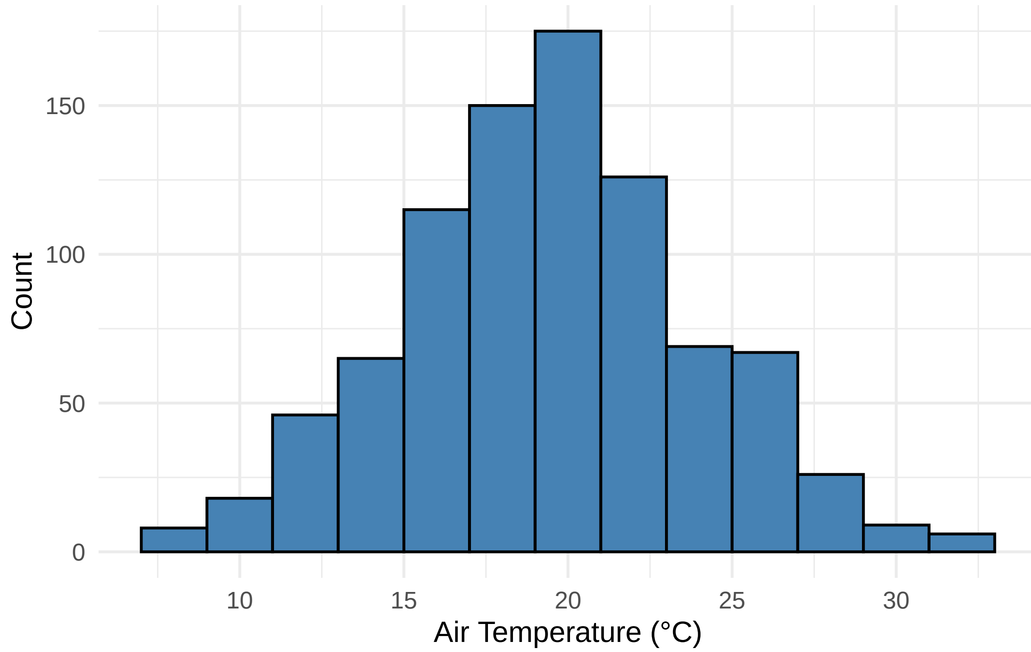


Figure 1: Air Temperature Distribution

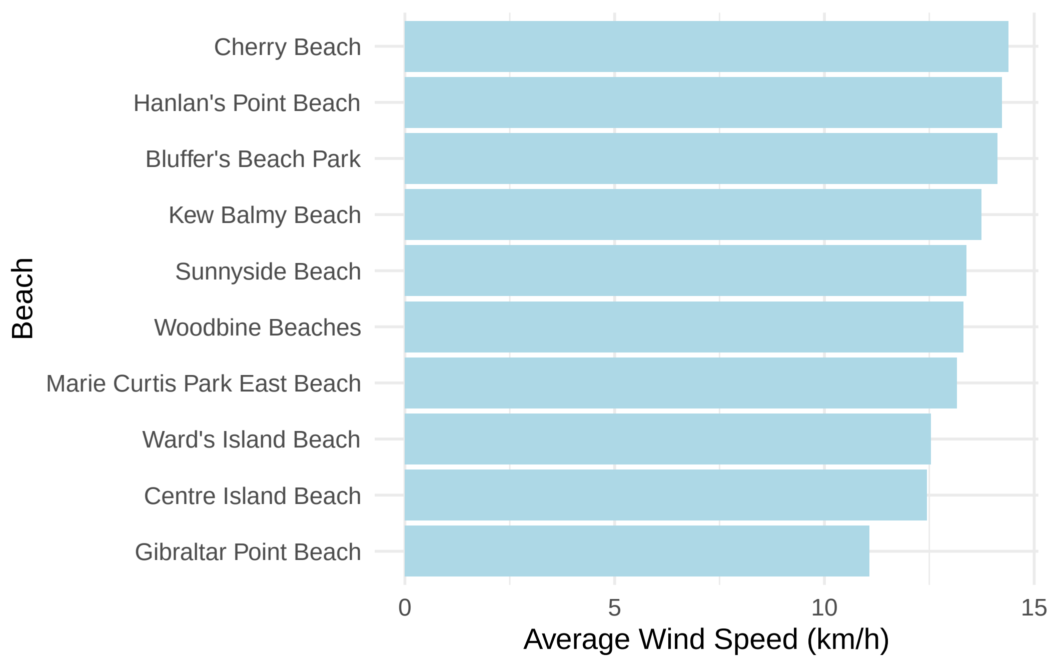
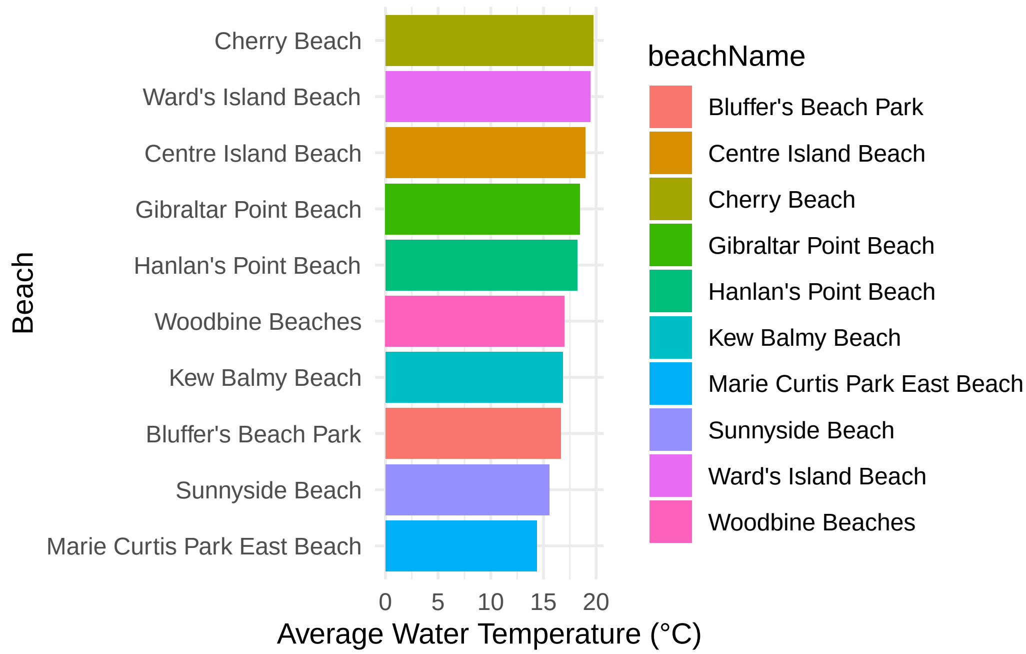


Figure 2: Average Wind Speed by Beach



(a) Across different Toronto beaches

Figure 3: Average water temperature by beach

see that on non-rainy days (labeled "No"), the highest number of waterfowl were observed, with a count exceeding 60,000. On rainy days (labeled "Yes"), the total waterfowl count is significantly lower, around 40,000. The "N/A" category, representing days with unknown rain status, shows very few waterfowl, with a total count close to zero.

Figure 5 illustrates the relationship between wind speed (x-axis, measured in km/h) and turbidity (y-axis, measured in NTU). Each point represents an observation, with wind speed along the horizontal axis and turbidity along the vertical axis. This scatter plot shows that there is no strong linear relationship between wind speed and turbidity. However, lower wind speeds tend to be associated with a wide range of turbidity levels, whereas higher wind speeds appear to be less common and do not show higher turbidity levels. Most data points are clustered in areas with lower wind speeds (below 20 km/h) and turbidity levels below 50 NTU. Some outliers show higher turbidity levels above 150 NTU, especially at lower wind speeds. At wind speeds above 20 km/h, the data points become scattered, with very few data points over 40 km/h.

Rainfall significantly impacts water turbidity, as rain can introduce pollutants and sediments into the water. Figure 6 shows the distribution of turbidity levels (measured in NTU) on rainy versus non-rainy days. The x-axis represents whether rain was present ("Yes"), absent ("No"), or unknown ("N/A"), while the y-axis represents turbidity levels. The box plot suggests that most observations, whether rainy or not, tend to have lower turbidity levels. Non-rainy days

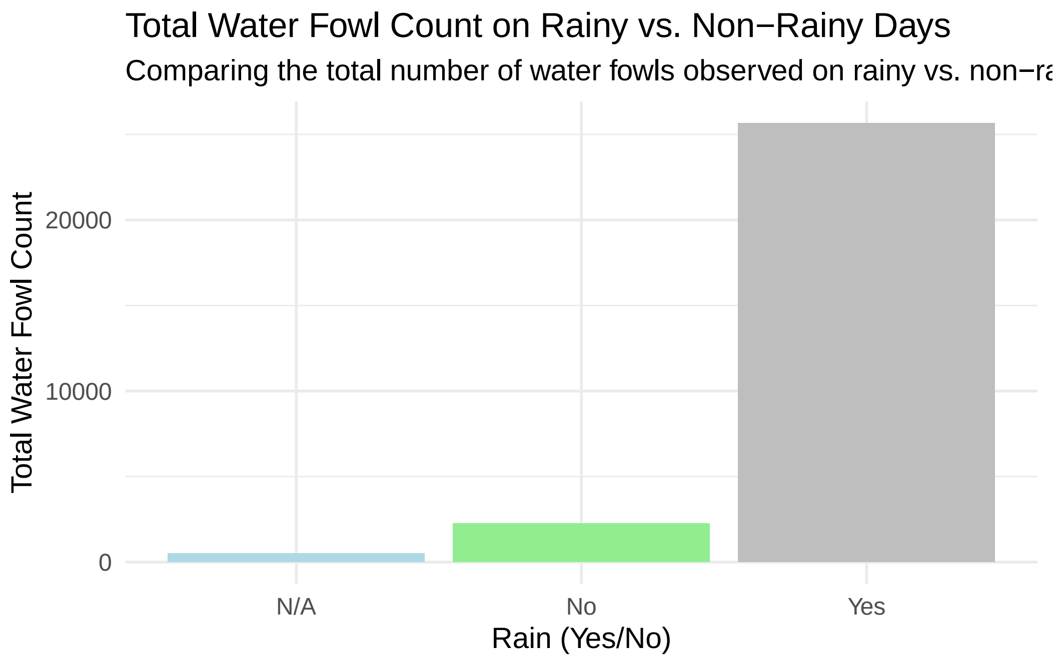
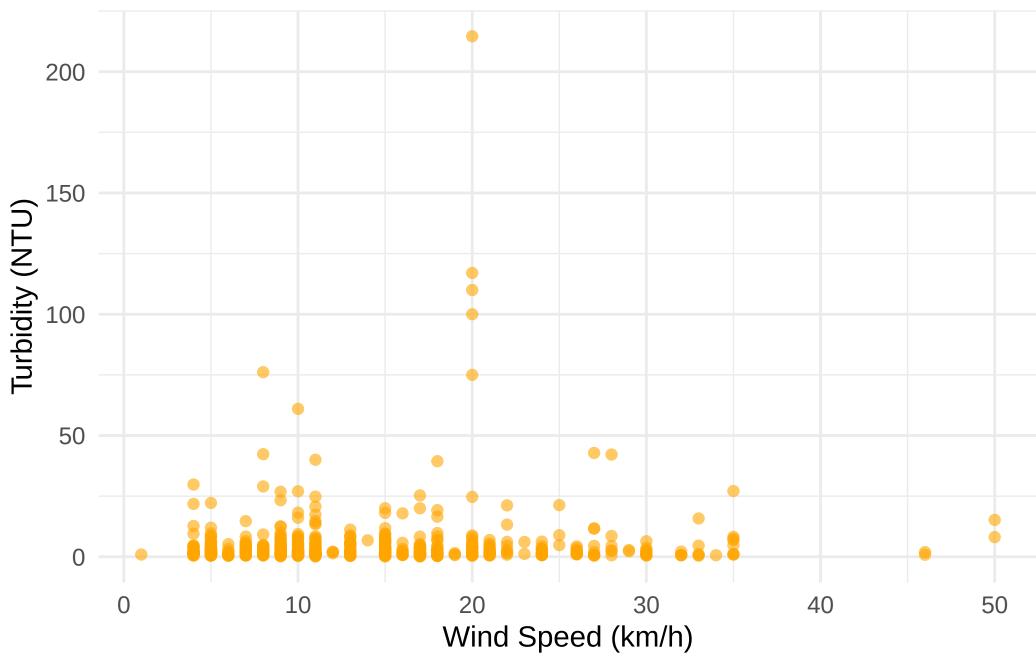


Figure 4: Total water fowl count on rainy vs non-rainy days



(a) Analyzing the relationship between wind speed and water turbidity

Figure 5: Wind speed vs turbidity

exhibit a slightly broader spread of turbidity levels, but the majority of observations have relatively low turbidity (close to 0 NTU), with several outliers reaching beyond 200 NTU. Turbidity levels were also mostly low in Rainy (Yes) and N/A(Unknown), but there were outliers over 150 NTU.

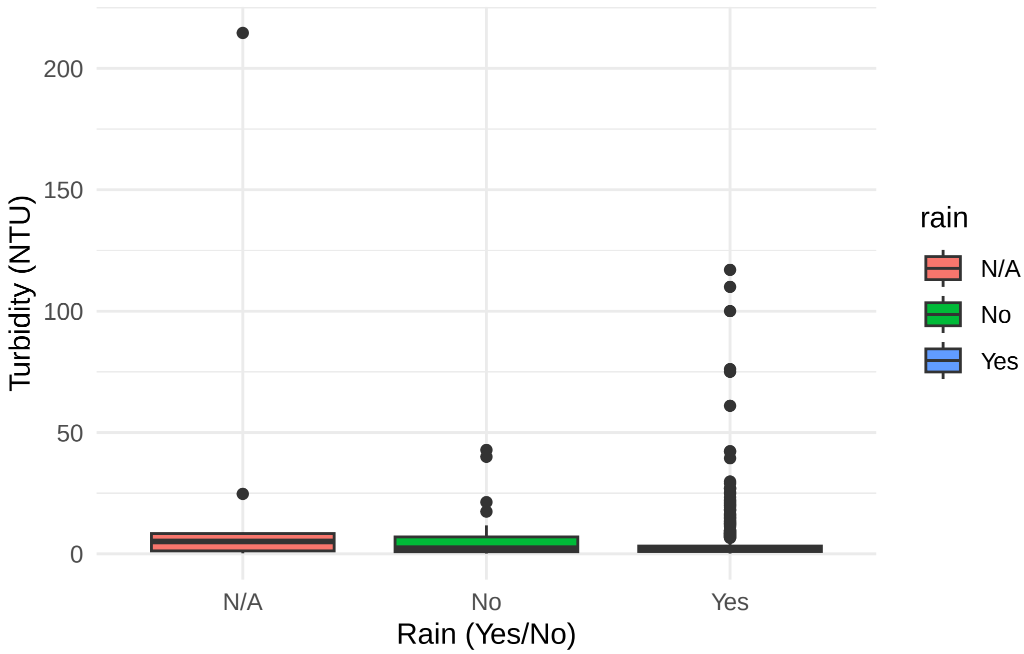


Figure 6: Turbidity Levels on Rainy vs. Non-Rainy Days

3 Discussion

3.1 Data Collection and Limitations

Observations in the dataset are obtained by city staff through a combination of both manual estimates and instrumental measurements. While the use of automatic tools such as thermometers and turbidimeters can provide more objective and consistent data, manual observations such as waterfowl counts and qualitative assessments of water clarity introduce an element of subjectivity. In addition, removing outliers and missing data entries during the clean-up process addresses errors that can occur during data collection, especially extreme or unreasonable values (e.g., water temperatures above 200°C).

In summary, we measured real-world temperature, speed of wind, rainfall, turbidity, and waterfowl activity using a variety of instruments and estimation techniques, resulting in quantifiable data entries in the dataset. These measurements form the basis for analysing environmental conditions at Toronto Beach, and transform natural phenomena into structured data that can be analysed to reveal trends, patterns and relationships between different variables.

4 Appendix

4.1 Variable description

Table 3

Column	Description
_id	Unique row identifier for Open Data database
dataCollectionDate	Date observations were collected
beachName	Name of beach where observations were collected
windSpeed	Wind speed measured in km/h
windDirection	Wind directions - East, North-East, North, North-West, West, South-West, South, South-E
airTemp	Air temperature in degrees Celsius
rain	Presence of rain - Yes, No
rainAmount	Amount of rainfall in mm in the last 24 hours
waterTemp	Water temperature in degrees Celsius
waterFowl	Count of water fowl seen
waveAction	Observations on wave action - high, low, moderate, none
waterClarity	Visual state of water's clarity - free form text
turbidity	Turbidity of water measured in Nephelometric Turbidity Units (NTUs)

Description of variables in the dataset

References

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

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/.

Saleem, Faizan, Thomas A Edge, and Herb E Schellhorn. 2022. "Validation of qPCR Method for Enterococci Quantification at Toronto Beaches: Application for Rapid Recreational Water Monitoring." *Journal of Great Lakes Research* 48 (3): 707–16.

Saleem, Faizan, Herb E Schellhorn, Albert Simhon, and Thomas A Edge. 2023. "Same-Day Enterococcus qPCR Results of Recreational Water Quality at Two Toronto Beaches Provide Added Public Health Protection and Reduced Beach Days Lost." *Canadian Journal of Public Health* 114 (4): 676–87.

Sanchez, Johanna, Jordan Tustin, Cole Heasley, Mahesh Patel, Jeremy Kelly, Anthony Habjan, Ryan Waterhouse, and Ian Young. 2021. "Region-Specific Associations Between Environmental Factors and Escherichia Coli in Freshwater Beaches in Toronto and Niagara Region, Canada." *International Journal of Environmental Research and Public Health* 18 (23): 12841.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019a. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

———, et al. 2019b. "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.

Young, Ian, J Johanna Sanchez, Binyam Negussie Desta, Cole Heasley, and Jordan Tustin. 2023. "Recreational Water Exposures and Illness Outcomes at a Freshwater Beach in Toronto, Canada: A Prospective Cohort Pilot Study." *PLoS One* 18 (6): e0286584.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.