Environmental Factors Affecting Water Quality at Toronto Beaches*

In each summer period (2010-2023)

Ziyuan Shen

September 26, 2024

This study investigated the effects of various environmental factors, including rainfall, wind speed, air and water temperature, and the presence of waterfowl, on turbidity levels at Toronto beaches from 2010 to 2023. By analyzing the data, the study found that rainfall was the most important predictor of increased turbidity, while other factors such as air temperature and the presence of waterfowl showed a weaker correlation. The results of the study highlight the importance of monitoring rainfall events to maintain water quality and ensure public health safety at recreational beaches. This study contributes to a broader understanding of how environmental conditions affect water clarity and provides insights for policymakers to enhance beach management strategies.

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, waterfowl populations and turbidity. The results of the study indicate that while rainfall is the main cause of increased turbidity, factors such as water temperature and the presence of waterfowl also play a role, albeit to a lesser extent. By identifying the conditions that contribute to poorer water quality, this study provides valuable insights for policymakers and beach management organizations, who can use these findings to develop more targeted interventions and recommendations. The methods used, the results obtained and their implications for public health and environmental management are described in detail in subsequent sections. And the following structure was 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. Also, I removed several outliers. Some of the outliers such as the water temperature was up to 220

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

degree centigrade, which is a very unusual value. So I set reasonable range values for water temperature, air temperature, and wind speed, and removed any unreasonable values.

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 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.

Thus, I selected eight key variables that were deemed most valuable to this analysis. These variables included date and beach name. 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

Summary statistics table

Table 2 summaries the numerical data. The average wind speed was 13.3km/h with a maximum of 50 km/h, which stirs up sediments and may increase water turbidity. The average air temperature was 19.9°C, which directly affects the comfort of beachgoers and water quality, since warm air leads to higher water temperatures and increased microbial activity. Water temperatures averaged 15.6°C, reaching a maximum of 27°C during the summer months. this is often associated with bacterial growth, leading to health warnings. The number of waterfowl is another important factor, with an average of 32.3 waterfowl, but sometimes the number spikes to 450. The presence of waterfowl can bring faecal pollution into the water which increases the risk of bacterial contamination and water quality warnings.

##Data Visualization {#sec3-data} ### Turbidity across different beach Figure 1 provides a clear visualization of the distribution of turbidity across various beach locations. Bluffer's Beach Park, Kew Balmy Beach, and Cherry Beach appear to have the highest overall turbidity levels, with numerous outliers reaching above 50 NTU and some exceeding 100 NTU. Hanlan's Point Beach, Ward's Island Beach, Gibraltar Point Beach, and Centre Island Beach exhibit much lower turbidity levels overall. Other beaches are showing more variability than low-turbidity beaches but less than the higher-turbidity ones.

2.2.1 Relationship between Air Temperature and Turbidity

Figure 2 shows a flat linear regression line, indicating that turbidity remains relatively unaffected by changes in air temperature.

2.2.2 Relationship between Water Temperature and Turbidity

Relationship between Water Temperature and Turbidity shows the similar trend to Figure 2. Both Figure 2 and Figure 3 shows that air temperature and water temperature are not a strong predictor of turbidity.

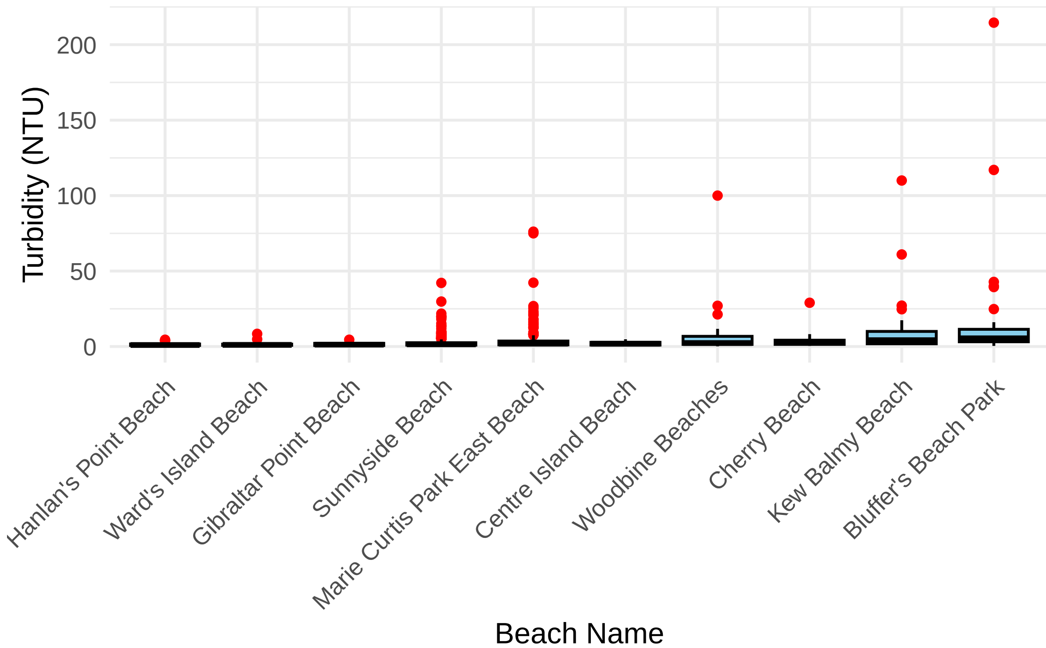

Figure 1: Distribution of Turbidity by Beach

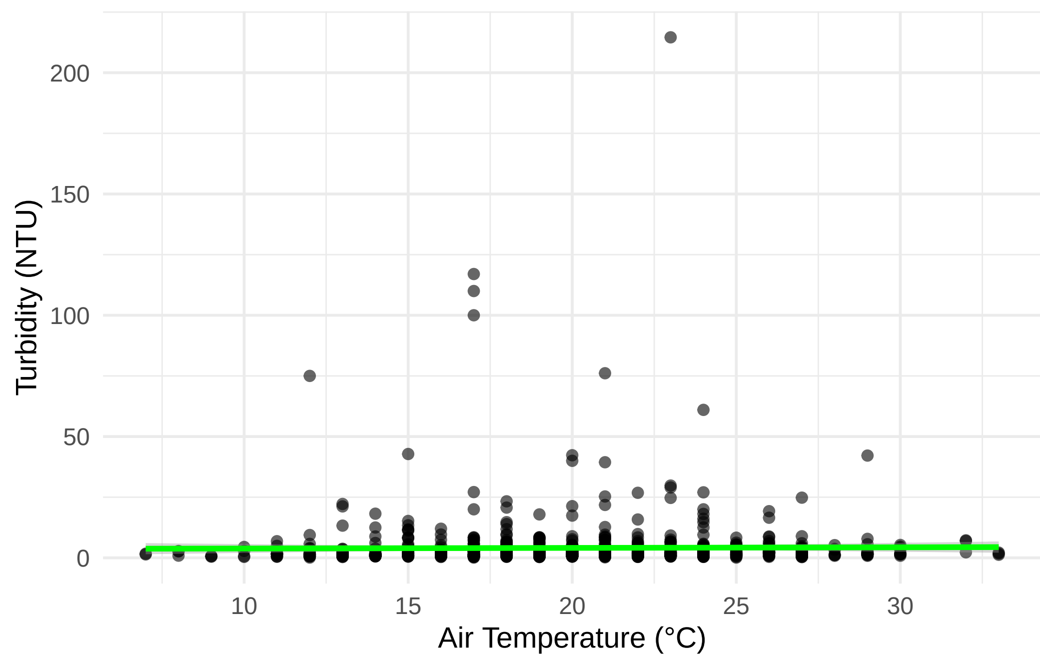

Figure 2: Relationship between Air Temperature and Turbidity

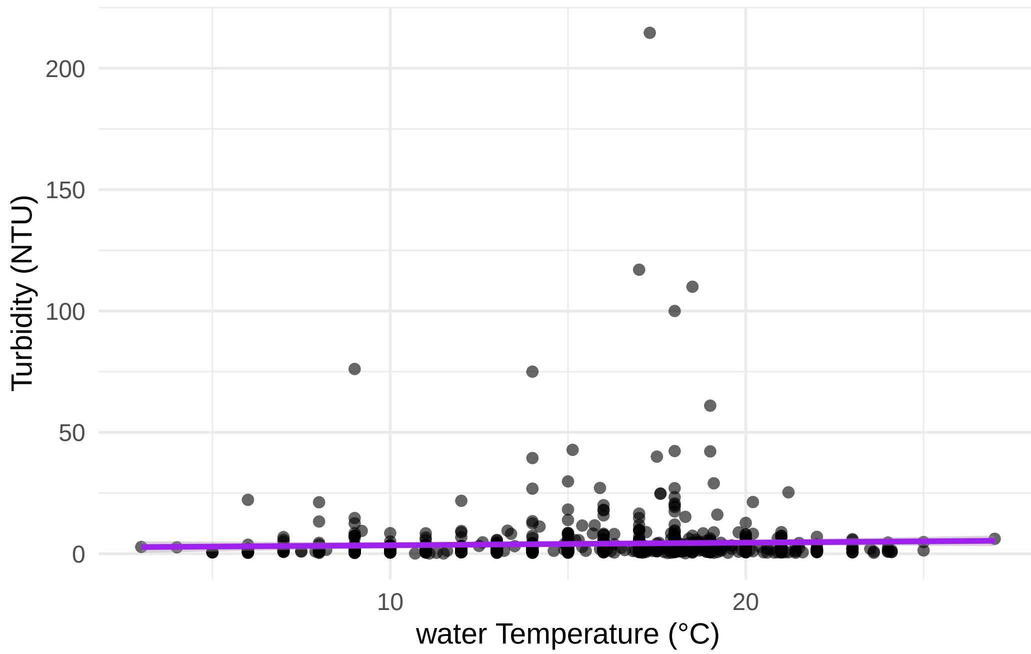

Figure 3: Relationship between Water Temperature and Turbidity

2.2.3 Relationship between Wind Speed and Turbidity

Figure 4 illustrates the relationship between wind speed (x-axis, measured in km/h) and turbidity (y-axis, measured in NTU). This scatter plot shows that there is no strong linear relationship between wind speed and turbidity.

2.2.4 Relationship between Waterfowl Count and Turbidity

Figure 5 is a scatter plot demonstrates the relationship between Waterfowl Count and Turbidity (measured in NTU). There is a slight upward trend, indicated by the blue linear regression line. This suggests that as the number of waterfowl increases, the turbidity also tends to rise, although the increase is relatively mild. Most of the data points are clustered around lower turbidity values (0-50 NTU), especially when waterfowl counts are below 100. This suggests that in many cases the presence of birds does not immediately lead to significant turbidity in the water column.

Since there are many small values and some large outliers, I try to consider using a logarithmic scale for both waterfowl counts and turbidity to spread out the data and make trends more apparent. As shown in Figure 6, it provides a better visualization of the data and suggests that while bird activity has some impact on water turbidity, it is not the only determinant.

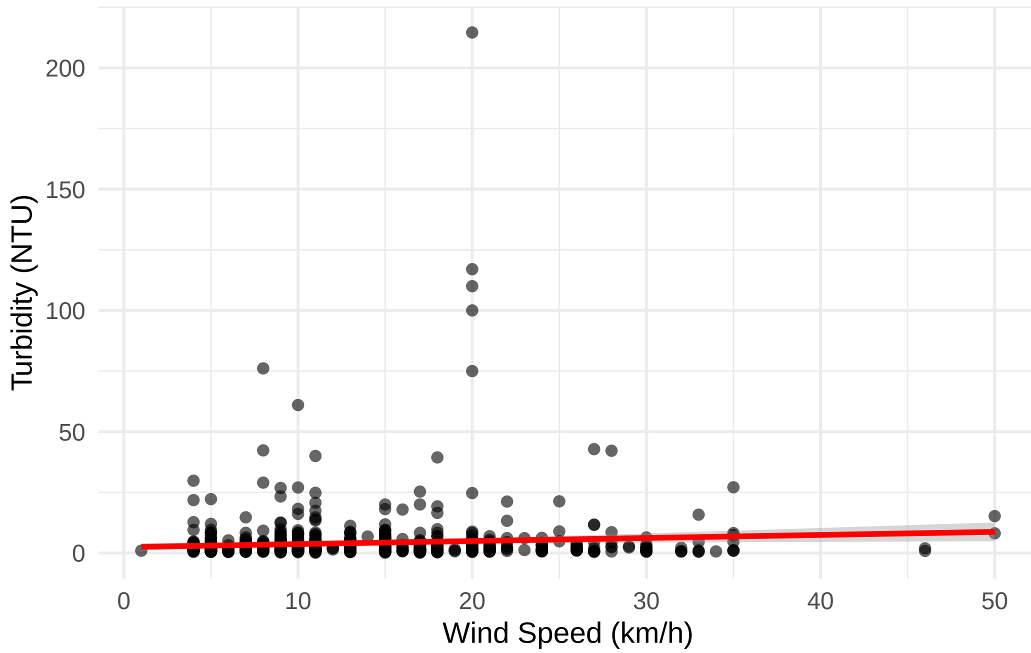

Figure 4: Relationship between Wind Speed and Turbidity

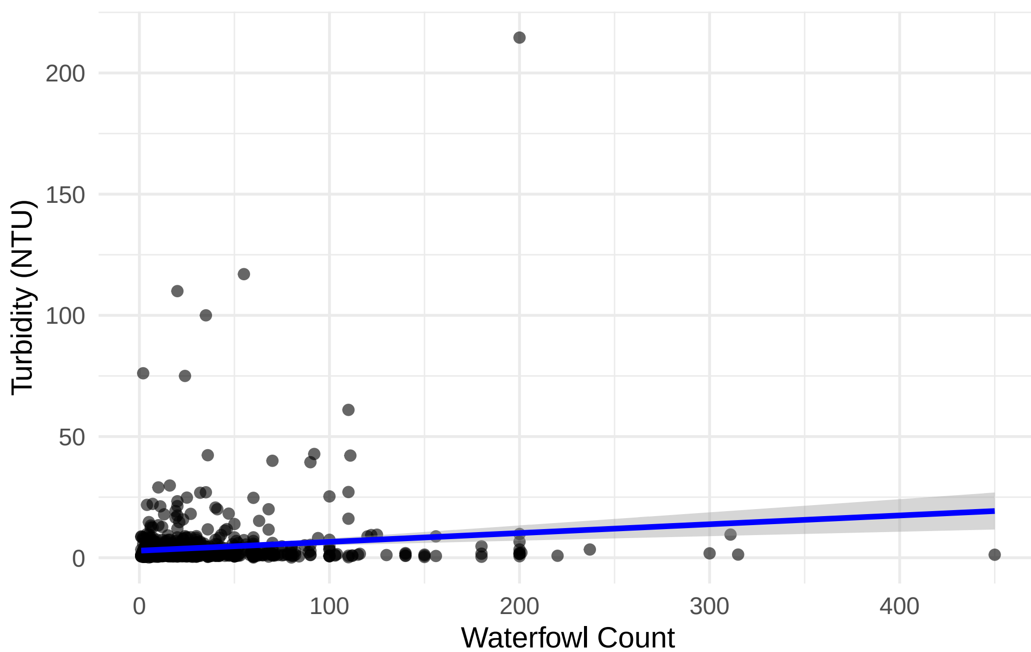

Figure 5: Relationship between Waterfowl Count and Turbidity

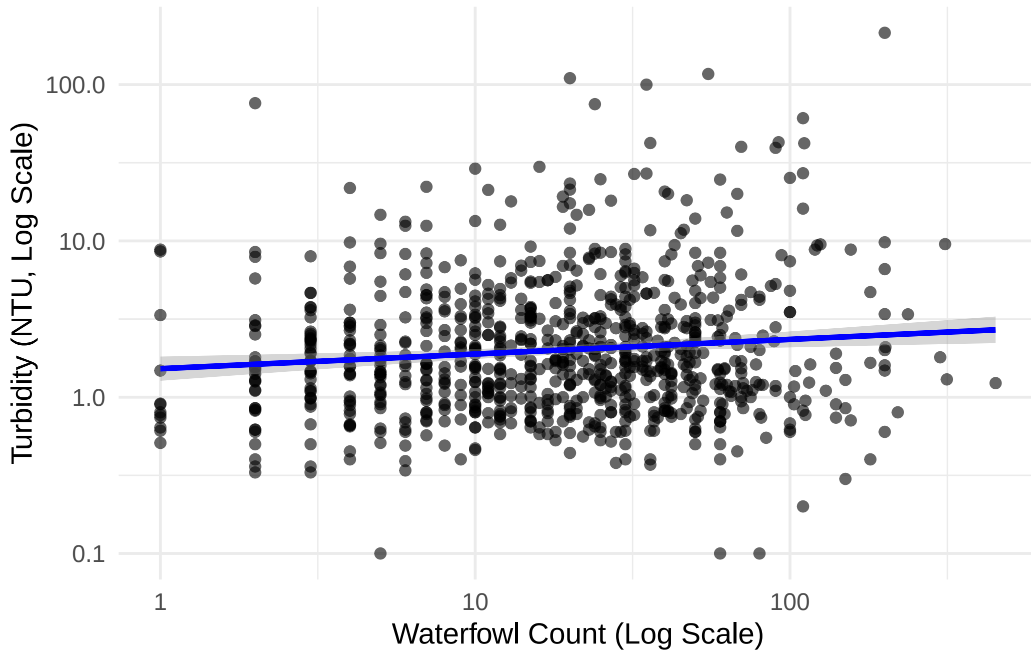

Figure 6: Relationship between Waterfowl Count and Turbidity

2.2.5 Turbidity Levels on Rainy vs. Non-Rainy Day

Figure 7 shows the distribution of turbidity levels (measured in NTU) on rainy versus non-rainy days. On rainy days, the change in turbidity increases relatively significantly. The occurrence of multiple outliers above 50 NTU indicates that rainfall is largely responsible for the increase in turbidity. The larger interquartile range (IQR) indicates that turbidity values are more scattered on rainy days compared to non-rainy days.

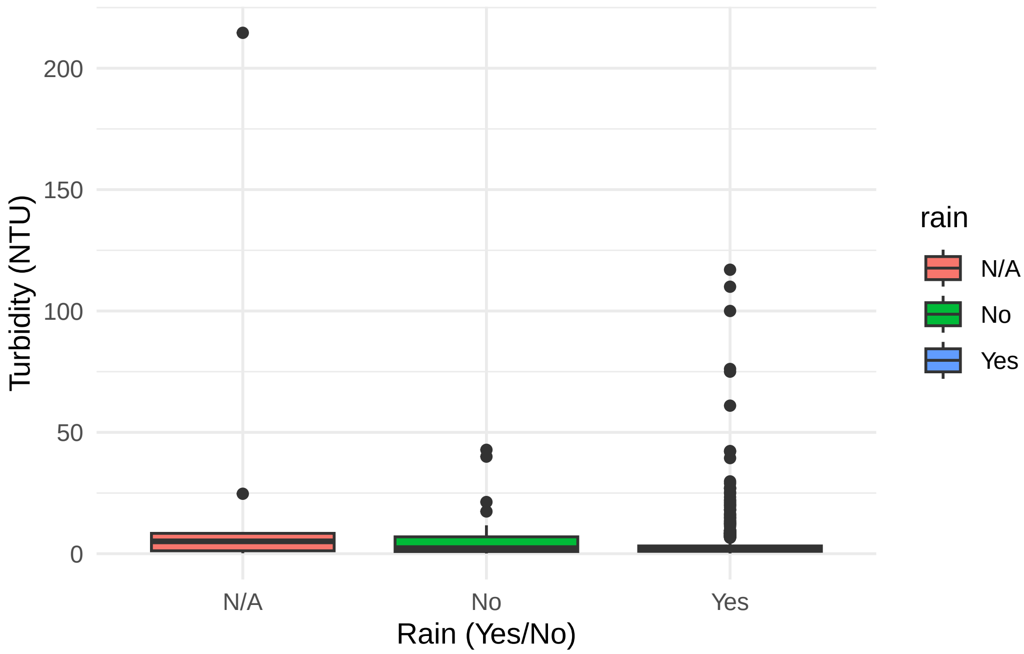

Figure 7: Turbidity Levels on Rainy vs. Non-Rainy Days

3 Discussion

3.1 Impacts of beach location on water quality

When analyzing the relationship between beach location and turbidity, Bluffer's Beach Park, Kew Balmy Beach and Cherry Beach were consistently higher compared to the other locations. These beaches not only had more outliers but also had larger interquartile ratios (IQRs), indicating large fluctuations in water quality over time. The high turbidity at these beaches may be due to their geographic location, proximity to sources of urban runoff, or local environmental conditions that increase the potential for pollution.

3.2 Impact of Rainfall on Water Quality

The above observations suggest that rainfall is an important factor in influencing turbidity levels in Toronto's beach waters. The study found that on rainy days, turbidity levels increased significantly, with several outliers exceeding 100 NTU, suggesting that rainfall may have washed pollutants into the water and reduced clarity. This finding underscores the role of rainfall in carrying sediment, pollutants, and other debris into water, leading to potential public health risks. The large interquartile range on rainy days suggests that water quality can be highly variable, further stressing the need for a robust water quality monitoring system following heavy rainfall.

3.3 The Role of Waterfowl in Turbidity

When examining the relationship between waterfowl and turbidity levels, the correlation between the two was weak, although a slight upward trend was observed. The data showed that in most cases, the presence of waterfowl did not result in an immediate spike in turbidity levels, suggesting that the impact of waterfowl is more cumulative or dependent on other environmental factors such as rainfall. However, the presence of large numbers of waterfowl does pose a pollution risk as they can carry faecal contaminants into the water and elevate bacterial levels. Therefore, I think more data and information on the impact of waterfowl can be sought for further study.

3.4 Wind speed and its effect on turbidity

It is often assumed that wind speed affects water quality by stirring up sediments at the bottom of the water. However, research results have shown no strong relationship between wind speed and turbidity. While lower wind speeds were associated with a wider range of turbidity, higher wind speeds did not necessarily result in higher turbidity. This suggests that while wind is important, it may not be the primary driver of water clarity at these beaches.

3.5 Temperature and Turbidity

It was found that both air and water temperatures had little effect on turbidity. The flat regression line in the temperature turbidity analysis showed that temperature fluctuations by themselves do not significantly affect changes in water clarity. This was a little surprising to me because higher temperatures tend to be associated with increased microbial activity, and increased microbial activity can affect water quality.

3.6 Data Limitations and future improvement

Observations in the dataset are obtained by city staff through a combination of both manual estimates and instrumental measurements. Manual data collection introduces potential bias and inconsistency because human judgement can vary and is more prone to observational errors. In addition, while some unusual outliers and missing data improves the completeness of the dataset, these filters may inadvertently omit important rare events or distort trends. For example, extreme weather events or unusual environmental conditions (e.g., large bacterial outbreaks) may have been filtered out during the data cleaning process.

Another limiting factor is the lack of detailed water quality measurements other than turbidity. While turbidity is a useful proxy for water clarity, it does not provide a complete picture of water contamination or health risks. Bacterial contamination may occur in relatively clear water, and the lack of a strong relationship between water temperature and turbidity suggests that turbidity alone may not adequately reflect microbial activity. So I expect that future research will provide a more complete understanding of public health risks by incorporating additional water quality indicators, such as bacterial counts or chemical contaminants.

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.