Toronto Beaches Water Quality*

Haoyu Wang

January 23, 2024

This paper examines E.coli levels in Toronto's public beaches, using data from the Opentoronto website. We analyze how weather conditions affect water quality and the occurrence of water-borne illnesses. Our findings reveal trends in beach water safety and the impact of public health warnings. This study offers insights into managing urban beach water quality for public safety.

1 Introduction

The Beach Water Sampling Program for the City of Toronto is a co-operative effort between Toronto Public Health, Toronto Water, the Marine Police Unit, Parks and Recreation Department. It is implemented in accordance with the requirements of the Ministry of Health Beach Management Protocol (January 01, 1998) in order to reduce the incidence of water-borne illness in the population.Every year between June and September (Labour Day), the City of Toronto's Parks, Forestry & Recreation division collects daily water samples from Toronto's supervised public beaches to be tested for E.coli bacteria. Toronto Public Health (TPH) measures E.coli levels to determine the beach water quality (and lifeguards monitor the safety conditions) for public swimming. When E.coli levels are unsafe, TPH posts warning signs against swimming.

Swimming is not recommended during and after storms, floods, or heavy rainfall. Cloudy water can be an indicator of high levels of bacteria that may pose a risk to human health. Conditions are based upon E. coli counts in beach water samples taken over the past 24 hours.

This paper explores the intricacies of the Beach Water Sampling Program and its impact. In Section 2, by analyzing E. coli data from Toronto's beaches, I aim to identify patterns, evaluate the program's effectiveness in ensuring safe swimming conditions, and discuss its broader implications for public health and environmental management in urban settings. The collection of graphical analyses presents a detailed overview of water quality at Toronto's

*Code and data are available at: https://github.com/groundUofT888/Toronto-Beaches-Water-Quality.git.

beaches, as measured by E. coli levels in 2023. The box plot(Figure 2) comparison elucidates the spread and outliers of E. coli counts(Figure 1), offering a snapshot of variability between different beaches. The time series line plot traces the fluctuating E. coli levels over the summer months, revealing specific dates when the counts peaked, potentially signaling health risks. Meanwhile, the scatter plot with a linear regression(Figure 3) provides a broader perspective on the trend of water quality throughout the season, highlighting whether there is a general increase or decrease in contamination over time. Together, these visual insights are crucial for public health monitoring and guiding safe recreational water use(Section 2).

2 Data

This report focuses on four of these variables which are: beach_id, collect_date,site_name, Coil, comment and geometry.The data is downloaded from opentdataoronto website(Gelfand 2022). I analysis the data by using R(R Core Team 2022), and R packages , "dplyr"(Wickham et al. 2023) and kable()(Xie 2014), a dataset(Table 1) is shown below.

The table below presents the initial observations from the 2023 dataset on E. coli measurements at Toronto beaches. It showcases a snapshot of the data collected on September 10, 2023, from various sampling sites at Marie Curtis Park East Beach and Sunnyside Beach. As a result, although the recorded E. coli counts are not available (marked as 'NA'), the information provides a framework for understanding the geographical distribution of the testing sites and the frequency of data collection. This preliminary view into the dataset underscores the comprehensive nature of the Beach Water Sampling Program and sets the stage for a more detailed analysis of water quality trends over the swimming season.

Table 1: First Few Observations of 2023 E. coli Data

X_id	beachId	beachName	siteName	collectionDate	eColi	comments	geometry
1	1	Marie Curtis Park East Beach	29W	2023-09-10	NA	None	c(-79.539696, 43.58592)
2	1	Marie Curtis Park East Beach	32W	2023-09-10	NA	None	c(-79.54056, 43.58511)
3	1	Marie Curtis Park East Beach	33W	2023-09-10	NA	None	c(-79.540886, 43.58486)
4	1	Marie Curtis Park East Beach	31W	2023-09-10	NA	None	c(-79.54039, 43.585304)
5	1	Marie Curtis Park East Beach	30W	2023-09-10	NA	None	c(-79.540054, 43.58561)
6	2	Sunnyside Beach	23W	2023-09-10	NA	None	c(-79.461525, 43.63625)

Figure 1(Figure 1) is a time series tracking E. coli levels at Marie Curtis Park East Beach and Sunnyside Beach throughout the summer of 2023. The data reveals a trend of fluctuating E. coli concentrations, with occasional sharp increases that suggest transient contamination events, possibly after rainfall or other environmental inputs.

The line for Marie Curtis Park East Beach, shown in red, demonstrates several significant peaks, which indicate days with notably high E. coli levels. These peaks may be of particular interest to public health officials as they potentially coincide with public advisories and highlight the beach's vulnerability to pollution. On the other hand, Sunnyside Beach, depicted in blue, shows a relatively steady state with less dramatic fluctuations in E. coli counts, hinting at different environmental conditions or the effectiveness of local sanitation measures.

While most days record E. coli levels within expected ranges, the graph draws attention to outlier events that could be critical for evaluating beach management practices. It underscores the importance of consistent water testing and could guide further investigation into the causes behind these sporadic elevations in bacterial counts.

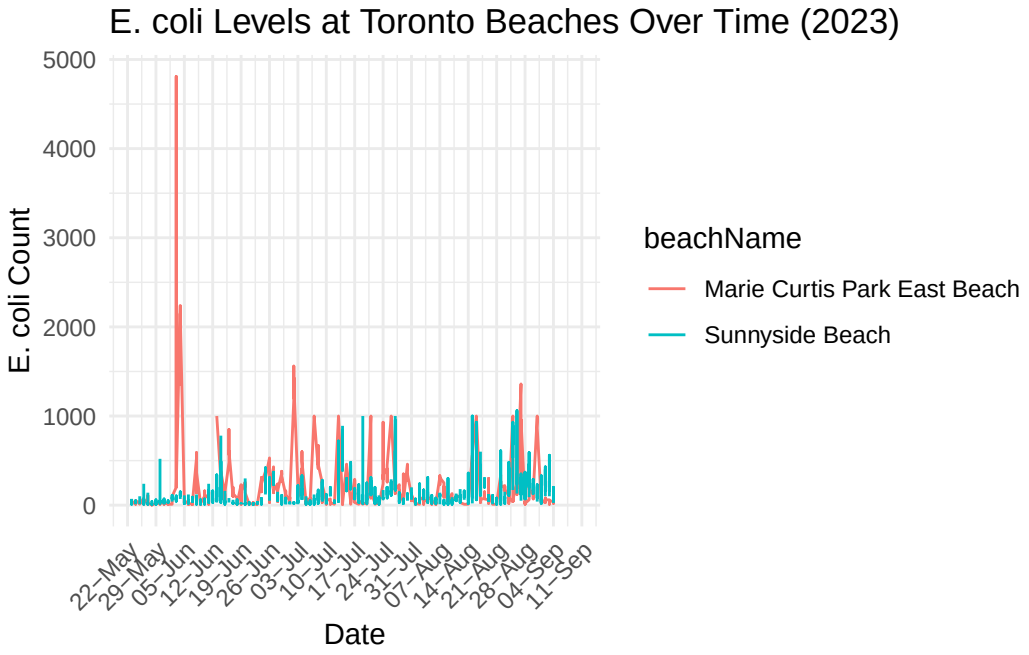

Figure 1: ecoilcount by date

And also the image (Figure 2)shows a box plot graphic that represents the distribution of E. coli bacteria levels found at two different beaches in Toronto for the year 2023. This type of plot is helpful to see the range of E. coli counts, where most of the values lie (which is seen in the "box" part of the plot), and to identify any outliers, which are E. coli counts that are unusually high or low compared to the rest of the data.

Dots above or below the boxes represent the outliers, which in this context could indicate days

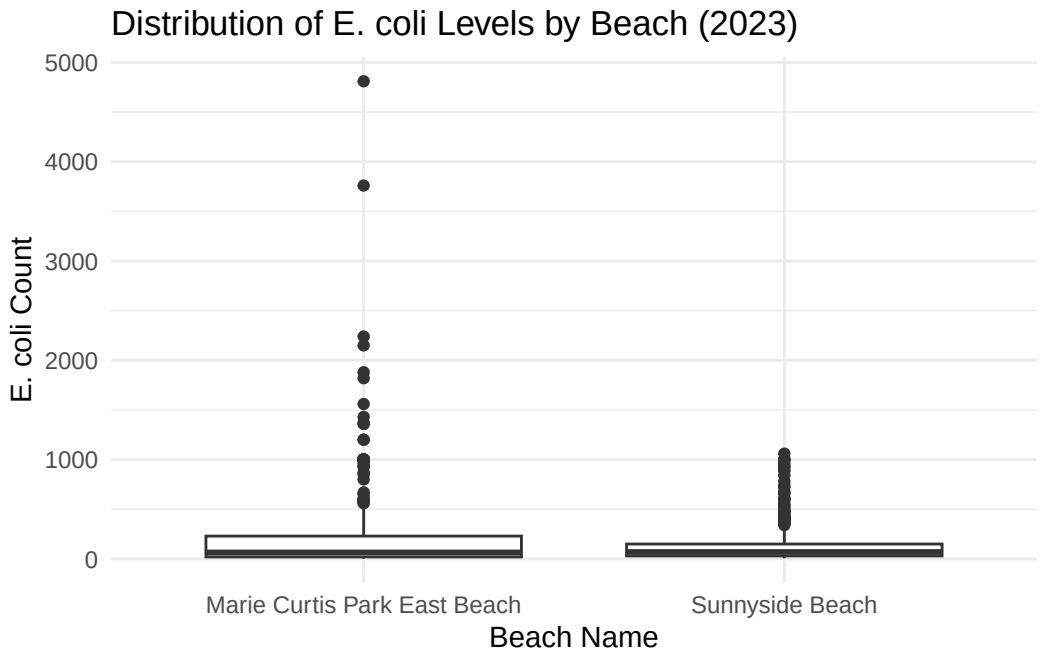

Figure 2: ecoilcount of beaches

when the E. coli levels were exceptionally high or low. These outliers represent times when the water quality was significantly different from usual, potentially indicate pollution events or other environmental changes.

(Figure 3)Each dot on the plot is a specific measurement of E. coli levels on a given date. The linear regression line that runs through the data points provides a visual representation of the overall trend. The line in this graph(Figure 3) is going up, it suggests that E. coli levels are slowly increasing over time. This type of analysis is important for understanding long-term changes in water quality and can be useful for public health officials to assess whether water quality is improving or if there are periods of higher risk that need attention.

3 Discussion

The analysis of E. coli levels at Toronto's beaches in 2023 reveals a complex picture of water quality over the swimming season. The data, represented in the time series plot, shows variability in E. coli counts at Marie Curtis Park East Beach and Sunnyside Beach, with several spikes indicating days where levels exceeded the typical range. These peaks are of particular interest as they might correlate with environmental factors such as heavy rainfall or other pollution events. While the overall trend suggests that the water quality at Sunnyside Beach is generally better with fewer high-count incidents, Marie Curtis Park East Beach exhibits more frequent fluctuations in E. coli levels. It's essential for public health officials to investigate

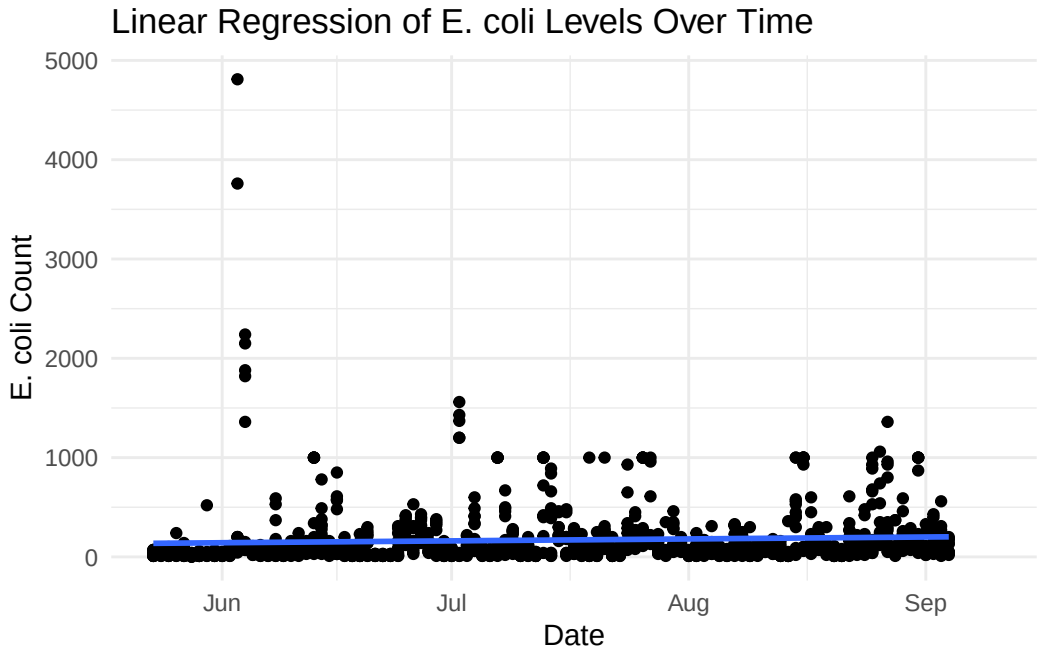

Figure 3: E. coli Levels Over Time

these occurrences to determine their causes and address any potential sources of contamination. Moreover, these findings underscore the importance of regular water testing and public notification systems, as they play a crucial role in safeguarding the health of beach visitors. Future initiatives could focus on enhancing the predictive capabilities of water quality models by integrating meteorological data, which could enable preemptive action to maintain the safety and enjoyment of Toronto's valued recreational water sites.

References

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://sharlagelfand.github.io/opendatatoronto/.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.