Analysis of Toronto's Beaches Water Quality From 2007 to 2023*

Samantha Barfoot

January 24, 2024

During the summer months, the Toronto beaches are popular destinations to cool off at the hot weather. However, there is yearly issue of water contamination from bacteria such as E. coli, which puts the city's residents at risk. Using data obtained by the City of Toronto's Open Data Catalogue, this paper examines the E. coli trends at the Sunnyside and Marie Curtis beaches between 2007 and 2023. Additionally, it explores the implications and causes of this health and safety issue, revealing that the number of unsafe swimming days throughout the summer is hugely impacted.

1 Introduction

Between June and September of each year, the City of Toronto samples the city's beaches to test for bacteria such as E. coli. High levels of these bacteria can cause adverse health risks, which can result in illnesses such as strokes, kidney failure, or seizures (Public Health Agency of Canada 2017). Therefore, the City of Toronto performs daily testing to ensure that E. coli levels do not exceed more than 100 Colony Forming Units (CFU) per 100ml of water ("Toronto Marks 5 Swimming Beaches as Unsafe Because of E. Coli Bacteria Levels" 2021). If E. coli levels exceed this level, then the beach is deemed unsafe, and people are discouraged from swimming. E. coli levels depend on various environmental factors such as heavy rain, birds, strong wind, and waves. Figure 1 shows the locations of several sources of water runoff that can carry E. coli into Lake Ontario; for example, in Combined Sewer Overflows (CSO), heavy rainfall mixes with raw sewage and overflows into the lake ("Toronto Marks 5 Swimming Beaches as Unsafe Because of E. Coli Bacteria Levels" 2021).

Therefore, it is important to regularly monitor the bacteria levels of the water at the Toronto beaches. Additionally, it is important to examine the trends related to this water contamination. This paper explores how contamination and E. coli levels change throughout the

*Code and data are available at: https://github.com/SamanthaBarfoot/paper-water-quality-analysis.git open beach season and how it impacts when beaches are open or closed. This was done by examining the frequency of open beach days between 2007 and 2023. Specifically, we compare how often beaches were open on the same day of the season (aggregated over all the sampling years). We also compared the E. coli levels on the same day of the season (aggregated over the sampling years). It was found that E. coli levels at the Sunnyside and Marie Curtis beaches (see Figure 1 for locations) increase greatly throughout the summer season, resulting in unsafe contamination levels. This is important because once E. coli levels exceed the threshold, they are considered unsafe to swim. E. coli-contaminated water can cause adverse health risks, putting the city's residents in danger. As the summer season progresses, temperatures rise but with high E. coli levels, people in the city cannot always safely use the beaches to cool off ("4 Toronto Beaches Marked Unsafe for Swimming Amid Heat Wave" 2023).

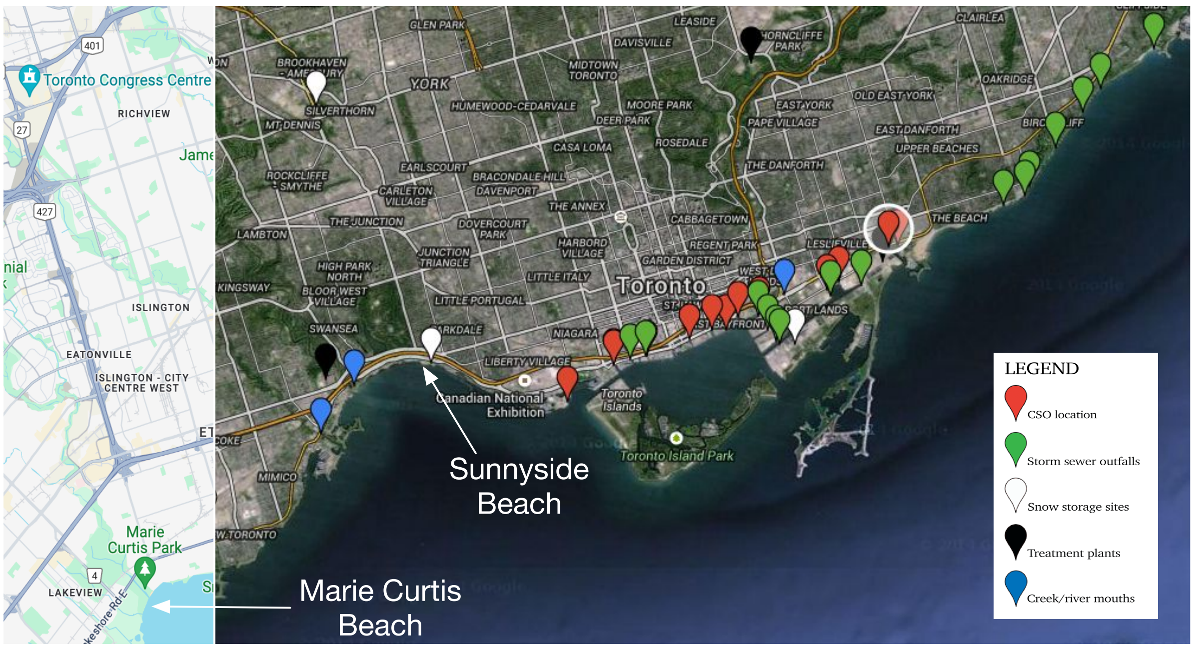

Figure 1: Map of Toronto waterfront with Sunnyside and Marie Curtis beaches indicated. Additionally, sources of water runoff are shown (image credit: swimdrinkfish.ca)

Section 2 of this paper covers the broader context of the Toronto Beaches Water Quality dataset, examines the variables used, and high-level aspects of the data cleaning. It also discusses observations made from the dataset, trends, and compares the two beaches, Sunnyside and Marie Curtis. Section 3 of this paper further explores the observations made in Section 2 about the Toronto Beaches Water Quality dataset and the implications and causes of these trends.

The graphs in this paper were created in R Studio with R as well as the analysis (R Core Team 2022). To conduct the analysis tidyverse (Wickham et al. 2019), janitor (Firke 2023), opendatatoronto (Gelfand 2022), MASS (Venables and Ripley 2002), tibble (Müller and Wickham

2023), readr (Wickham, Hester, and Bryan 2024), dplyr (Wickham et al. 2023) packages were all used. To create all the figures ggplot2 (Wickham 2016) was used.

2 Data

The dataset used in this report, Toronto Beaches Water Quality, was obtained from the City of Toronto's Open Data Catalogue. The data were last refreshed on the 11th of September 2023 and the full dataset is published by Toronto Public Health. The data were collected from two beaches, Sunnyside Beach and Marie Curtis Beach, and several sub-locations along each beach. The data span 16 years (2007 and 2023) between May and September. The data were collected by the Parks, Forestry, & Recreation Department through the Beach Water Sampling Program to reduce water-borne related illness incidents in the city (City of Toronto 2023).

The dataset analyzed in this paper contained 20,526 entries across eight variables used to describe the water samples. This included entry number, beach ID, beach name, beach site name, collection date, E. coli level, comments, and GPS coordinates. To better understand this information, data cleaning was performed to focus on three main variables: day within the sample collection season (May 1st to September 30th), E. coli levels, and number of days the beach was open. Instead of analyzing all 20,526 separate dataset entries, the individual days of the collection season, 153 in total, were examined in aggregate over the 16 years. From this, the E. coli levels were averaged for each day of the season. E. coli levels are represented in CFUs. The variables of safe and unsafe swimming days were created by comparing the E. coli levels to the safe swimming threshold of 100 CFU per 100ml of water ("4 Toronto Beaches Marked Unsafe for Swimming Amid Heat Wave" 2023). If the E. coli levels exceeded this threshold, the beach would be considered unsafe to swim. Data cleaning also involved separating the data into two different datasets: one for Sunnyside Beach and one for Marie Curtis Beach as they are in very different areas and have their own particular trends. At each beach there were multiple different sample sites, and we aggregated these sites and their data. Additionally, it involved the cleaning of certain bad data points that exceeded any realistic values.

2.1 Sunnyside Beaches

Examining the cleaned Sunnyside beach dataset, several observations can be made. First, on average, there are relatively few days in the season where swimming is considered safe as the E. coli safety threshold is often crossed (see Figure 2). The highest average E. coli level recorded was on the 70th day of the season at 962.88 CFU per 100ml. The lowest, value was on the 20th day of the season at 10 CFU per 100ml. Second, the E. coli levels have a moderate, positive, linear relationship (see Figure 2). As the season progresses, the E. coli levels tend to increase. Therefore, we can also see that as the day of the season increases, the frequency of safe E. coli levels decreases and thus there are fewer safe swimming days (see Figure 3). This can be seen in a negative, linear relationship (see Figure 3).

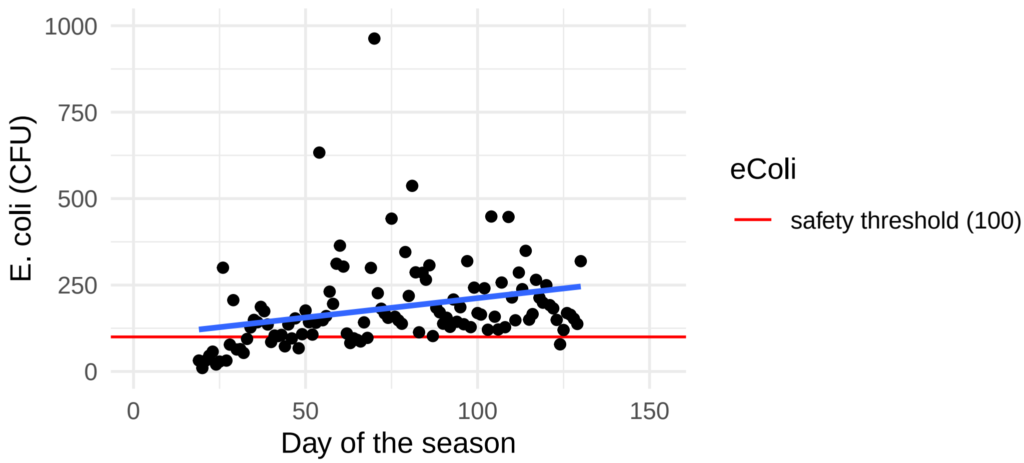

Figure 2: E. coli levels in CFU over the sampeling season. The safe E. coli level threshold is also shown (red) as well as the line of best fit (blue).

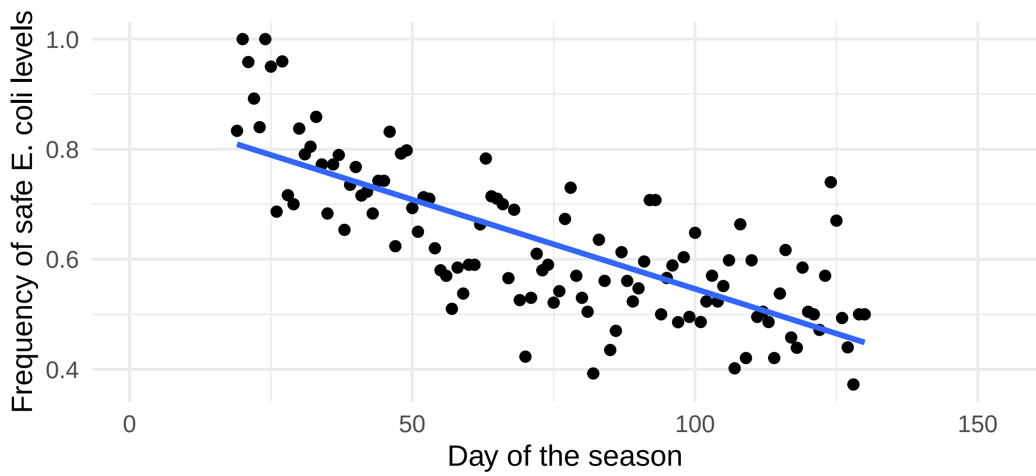

Figure 3: Frequency of safe E. coli levels over the sampeling season. The frequency indicates how often the given beach is considered open and safe to use. Blue is the line fo best fit.

2.2 Marie Curtis Beaches

Similar observations can be made about the Marie Curtis dataset. As the collection season progresses, the E. coli levels tend to increase in a moderately linear behavior (see Figure 4).

However, the data are more spread out. The highest average E. coli level recorded was on the 72nd day of the season at 570.75 CFU per 100ml. The lowest, value was on the 19th day of the season at 10 CFU per 100ml (see Figure 4). Additionally, there are very few days in the season where the E. coli levels are on average below the safe threshold (see Figure 4). Therefore, we can also see that as the season progresses, the frequency of safe swimming days decreases linearly (see Figure 5).

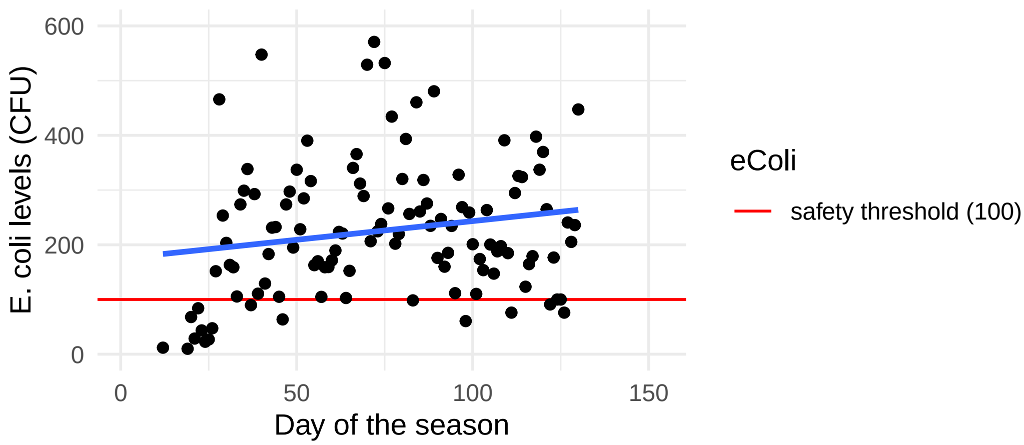

Figure 4: E. coli levels in CFU over the sampeling season. The safe E. coli level threshold is also shown (red) as well as a line of best fit (blue).

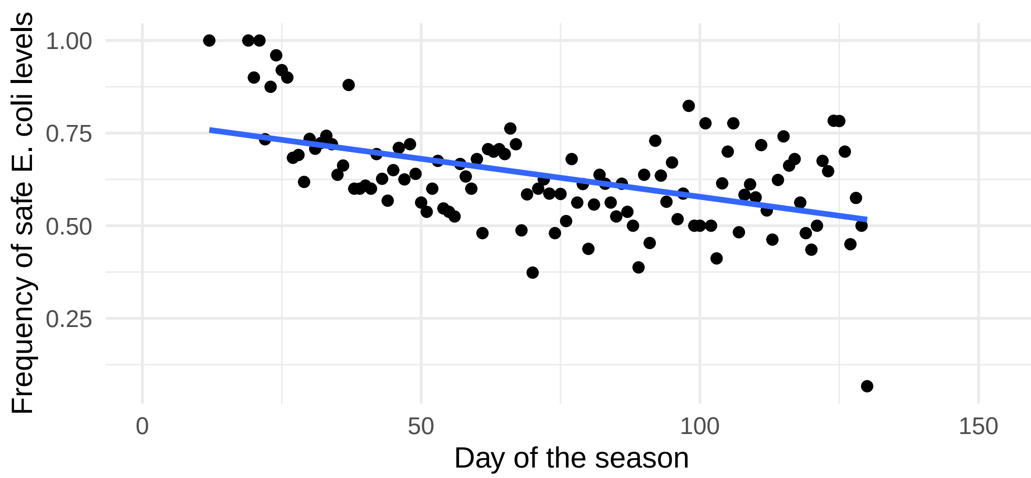

Figure 5: Frequency of safe E. coli levels over the sampeling season. The frequency indicates how often the given beach is considered open and safe to use. Blue is a line of best fit.

Comparing the Marie Curtis and Sunnyside beach data we can see that there are some differences. One of these is that the majority of the E. coli levels at the Sunnyside beach are below or around 250 CFU (see Figure 2) Whereas the Marie Curtis beach levels are more spread out. It appears that most of the Marie Curtis E. coli levels are below 400 CFU (see Figure 4). Therefore, Marie Curtis Beach tends to have high E. coli levels. However, both beaches have similar spreads of data when looking at the frequency of safe beach days. The majority of the data points fit between 0.4 and 1.0 for both beaches (see Figure 3 and Figure 5).

3 Discussion

The reason for this decline in safe beach days and the rise of E. coli levels throughout the summer could be due to a variety of different reasons. E. coli and other bacteria levels tend to increase after storms and heavy rainfall, which often occur during the spring and summer seasons which overlap with the collection season. Storms and heavy rainfall result in stormwater runoff as precipitation washes over land, carrying bacteria downstream to other tributaries such as Lake Ontario. It picks up substances such as animal and bird waste, garbage, chemicals, and other pollutants that make their way to the Toronto beaches (City of Toronto 2017b). Additionally, heavy rainfall can cause sewage pipes to overflow and enter the stormwater runoff as the underground sewage and stormwater share the same sewer and pipelines (City of Toronto 2017b). Therefore, heavy rainfall and storms can cause E. coli pollution at Toronto's beaches as it can be contained in stormwater runoff (Mittelstaedt 2004). Which, even if located far away from the waterfront, can still make its way though the different tributaries and into Lake Ontario and along Toronto's beaches.

While the Sunnyside and Marie Curtis beaches were the only two included in this dataset, the observation that there are more often than not unsafe contamination levels at these two very different locations is very important with regards to the rest of the Toronto shoreline. If these two beaches, almost 11km apart, both have high E. coli levels then what about the rest of Toronto's waterfront? Looking at Figure 1, many runoff points are actually far away from the analyzed beaches. While other Toronto beaches are also monitored for E. coli levels, people use more than just those designated beaches for recreational use. While it is only recommended that people swim inside designated swim areas (City of Toronto 2017a) people nonetheless use areas such as the city's inner harbour for activities such as kayaking and sailing. Areas such as the Toronto Harbour Front are not monitored for E. coli (Cribb 2017). This puts recreational users of these areas at risk for E. coli infection and other health risks. While not directly swimming, if someone falls into the water or practices flipping they are put in an unsafe position (Cribb 2017).

Therefore, this brings up the question of how much of Toronto's shoreline should be tested for bacteria contamination. Additionally, does more need to be done to prevent this contamination in the first place, and what can be done? Runoff from the sewage pipes that mix with the stormwater in the sewers appears to play a big role in contamination (Mittelstaedt 2004). However, it would be an expensive, and time-consuming endeavor to rework this system. In the meantime, we must look for smaller solutions. If prevention is not achievable then perhaps awareness of the dangers of recreational use in non-tested areas, especially after heavy rainfall could be emphasized.

References

"4 Toronto Beaches Marked Unsafe for Swimming Amid Heat Wave." 2023. *CBC News*. https://www.cbc.ca/news/canada/toronto/heat-wave-swimming-toronto-beaches-1.6919750.

City of Toronto. 2017a. "Beaches." *City of Toronto*. https://www.toronto.ca/explore-enjoy/parks-gardens-beaches/beaches/.

———. 2017b. "Why You Should Not Swim During or 48 Hours After Heavy Rain." *City of Toronto*. https://www.toronto.ca/services-payments/water-environment/protecting-improving-lake-river-stream-water-quality/where-water-pollution-comes-from/.

———. 2023. "Toronto Beaches Water Quality." Open Data Catalogue. https://open.toronto.ca/dataset/.

Cribb, Robert. 2017. "The Dirty Secret of Toronto's Bacteria-Laced Harbour." *Toronto Star*. https://www.thestar.com/news/gta/the-dirty-secret-of-toronto-s-bacteria-laced-harbour/article_d1b174cf-2dfa-5c4f-b372-4936d03931d6.html.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/janitor.

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

Mittelstaedt, Martin. 2004. "Worse Than You Think." *The Globe and Mail*, July. https://www.theglobeandmail.com/news/national/worse-than-you-think/article18268391/.

Müller, Kirill, and Hadley Wickham. 2023. *Tibble: Simple Data Frames*.

Public Health Agency of Canada. 2017. "Risks of E. Coli (Escherichia Coli) Infection." Education and awareness. https://www.canada.ca/en/public-health/services/diseases/e-coli/risks-e-coli.html.

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

"Toronto Marks 5 Swimming Beaches as Unsafe Because of E. Coli Bacteria Levels." 2021. *CBC*. https://www.cbc.ca/news/canada/toronto/swimming-beaches-toronto-safety-1.6089401.

Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with s*. Fourth. New York: Springer. https://www.stats.ox.ac.uk/pub/MASS4/.

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.

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

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. https://readr.tidyverse.org.