Water Contamination and Quality: a Look Into 2 Toronto Beaches*

Michelle Ji

September 19, 2024

Water quality based on E. Coli contamination levels of two Toronto beaches were examined and analyzed in this report from Gelfand (2022), spanning from 2007 to 2024. The contamination levels of E. Coli are based on the number of colonies per 100 mL of sample water. Investigation reveals one particular beach, Marie Curtis Park East Beach, had higher average yearly levels of contamination through the 17 years of data collection. This highlights the need for additional public health measures to be directed towards this beach in an effort to keep citizens safe and healthy.

1 Introduction

Water contamination continues to be a prevalent problem for human health, as contaminated water consists of fecal bacteria that can result in various diseases and death. It is estimated to cause 1.8 million deaths each year Ishii and Sadowsky (2008). Contamination of larger water bodies, such as Lake Ontario, can result from sewage or agriculture runoff leakage, posing a serious issue to the ecosystem and human health. As a result, drinking, ground, and recreational water are continuously monitored for levels of indicator bacteria, which are used to determine the presence of more dangerous pathogens in water. Two common types of indicator bacteria are E. Coli and Enterococci. While these are milder strains of bacteria and often do not cause severe disease, their presence is often linked to more pathogenic bacteria that can cause severe disease in humans Price and Wildeboer (2017). For example, some more pathogenic strains of bacteria that replicate in water bodies include Shiga toxin-producing E. Coli (STEC), which can cause bloody diarrhea and potentially fatal human diseases, such as hemolytic uremic syndrome and hemorrhagic colitis Ishii and Sadowsky (2008).

*Code and data are available at: https://github.com/michj3/toronto-beaches-water

In Toronto, water contamination at beaches is monitored and data is collected daily. The provincial safety standard for E. coli in Ontario, Canada, is 100 colonies per 100 mL of water. If the amount of E. coli surpasses this threshold, Toronto Water and Safety will flag that water body and advise against swimming. Currently, Toronto is conducting projects and investing funds into enclosing a portion of frequently contaminated beaches and providing water that meets the provincial recreational water quality standard. Other courses of action include the relocation of beaches where the water quality is consistently better and safer for citizens to swim in "Action Required: Toronto Beaches Plan" (2009).

To examine the impact of water contamination of E. coli on two Toronto beaches, Sunnyside Beach and Marie Curtis Park East Beach, this paper is broken down into various sections, including Data, Results, Discussion, and Conclusion. This paper uses data made available by Gelfand (2022) related to the contamination of the waters of the two beaches listed above. The Data section explores the data, highlighting certain key aspects that may be alarming to public health safety. Using this data, an investigation is conducted into the relationship between the average contamination level and the relative beach, average yearly trends of contamination, the distribution of E. coli levels by each beach, and if there is a relationship between the relative date and the contamination levels. The Results section introduces observations and potential conclusions from the data analysis. The Discussion section ties the previous sections to potential action courses and insights. Finally, the Conclusion section summarizes the key findings of this paper.

2 Data

The dataset was obtained through the City of Toronto's OpenData Library Gelfand (2022) and is titled "Toronto Beaches Water Quality". As a part of the city's commitment to clean water, Toronto Public Health, Toronto Water, the Marine Police Unit, and Parks and Recreation Department have a joint effort to measure and monitor the city's water of nearby beaches. Every year from June to August, peak beach season, Toronto's Forestry and Recreational division collects water samples from Toronto's supervised beaches daily to measure E. Coli levels. The Forestry and Recreational division collects samples and has its own lab to measure contamination levels. When levels are unsafe, flags will be posted for civilian notice. The dataset includes daily water measurements of E. Coli levels from 2007 to 2024.

In the City of Toronto's OpenData Library, there was one other dataset that detailed similar content to the dataset used in this paper, called "Toronto Beaches Observations". However, the only variable of interest was "turbidity", a unit of measurement for water contamination. However, that dataset was not chosen since water contamination levels were not as clearly defined and would be harder to interpret. Other variables in that dataset were not of interest and therefore, that dataset was not used.

The variables used in the dataset include observation, beach number, beach name, site location, collection date, and E. Coli levels. The "observation" variable is to help keep track of the

number of observations there are in the dataset. Since two beaches are of interest in this dataset, "Beach Number" denotes Marie Curtis Park East Beach as "beach 1", and Sunnyside Beach as "beach 2". This helps differentiate the beaches. The "Beach Name" variable differentiates the beaches by name, as listed above. "Site Number" is used by Toronto officials to refer to specific areas within each beach. "Collection date" refers to the specific date on which the sample was collected, which will range from June to August between 2007 and 2024. Finally, "E. Coli levels (# colonies/100mL)" is a variable that refers to the level of contamination of the water sample. Table 1 details the first few observations of the cleaned dataset, in which I removed unnecessary variables.

Table 1: First few observations of cleaned data set

Observation	Beach Number	Beach Name	Site Location	Collection Date	E. Coli Levels (# colonies/100mL
1	1	Marie Curtis Park East Beach	29W	2024-09-08	0
2	1	Marie Curtis Park East Beach	33W	2024-09-08	0
3	1	Marie Curtis Park East Beach	32W	2024-09-08	0
4	1	Marie Curtis Park East Beach	31W	2024-09-08	0
5	1	Marie Curtis Park East Beach	30W	2024-09-08	0
6	2	Sunnyside Beach	23W	2024-09-08	0
7	2	Sunnyside Beach	18W	2024-09-08	70
8	2	Sunnyside Beach	17W	2024-09-08	60
9	2	Sunnyside Beach	20W	2024-09-08	130
10	2	Sunnyside Beach	21W	2024-09-08	120

For further analysis, the cleaned dataset was separated into two different datasets by beach name, for simplicity purposes. Data was analyzed through the R programming software R Core Team (2023) and packages such as tidyverse Wickham et al. (2019), ggplot2 Wickham (2016), knitr Xie (2014), and dplyr Wickham et al. (2023) were used to help download, clean, simulate, analyze, and test the data.

3 Results

During the data collection between 2007 and 2024, Marie Curtis Park East Beach had the highest average contamination levels when compared to Sunnyside Beach, which is detailed

in Figure 1. Figure 1 illustrates that during the 17-year collection period, Marie Curtis Park East Beach had an average of 220.5 colonies per 100 mL of sample water, whereas Sunnyside Beach had an average of 174.765 colonies per 100 mL of sample water.

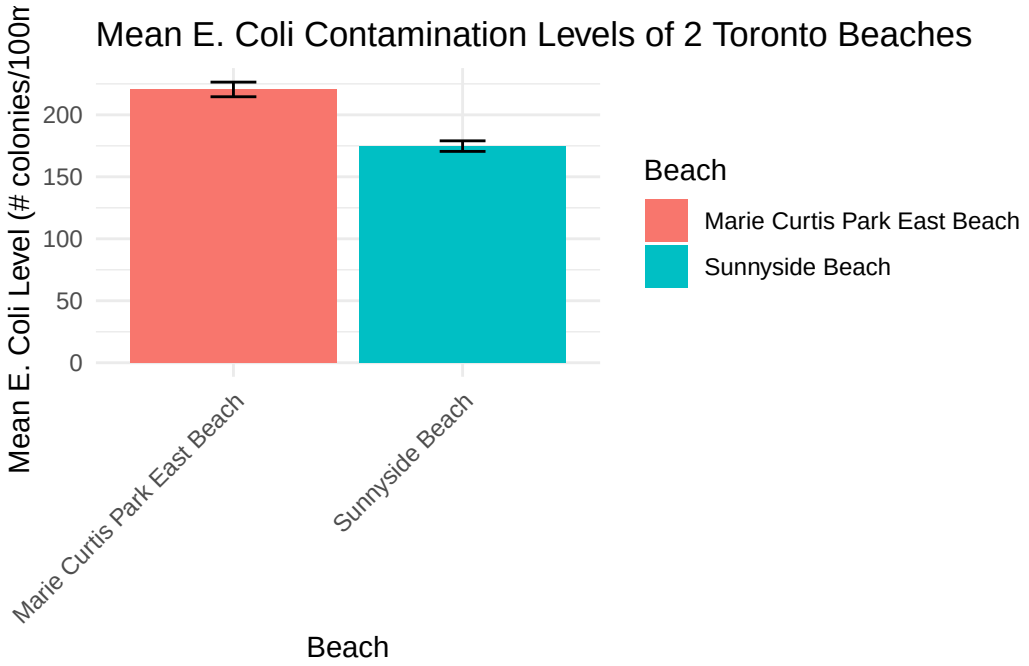

Figure 1: Average number of E. Coli colonies for each beach

Table 2 is a table showing the mean and standard deviation contamination levels of E. Coli for the two beaches. Further analysis also shows that the median contamination level of Marie Curtis Park East Beach is also higher than Sunnyside Beach, with a median contamination level of 3600 colonies per 100 mL of water compared to 3350 colonies per 100 mL of water respectively.

Table 2: Summary table of beach contamination level data

Beach	Mean	SE
Marie Curtis Park East Beach	220.4877	5.902066
Sunnyside Beach	174.7646	4.262949

Figure 2 illustrates the yearly average contamination levels between the two beaches on a closer level, looking at the yearly average per year. The red line represents Marie Curtis Park East Beach and the blue line represents Sunnyside Beach. The graph illustrates that from 2007 to 2010, there was an average upward trend of contamination levels for both beaches, with a very concerning peak for Marie Curtis Park East Beach in 2010. Following 2010, there

was a general decrease in E. Coli levels and the average seems to have remained steady in the upper 200s for the number of colonies per 100 mL of sample water. It is apparent that Marie Curtis Park East Beach almost consistently has higher levels of contamination than Sunnyside Beach.

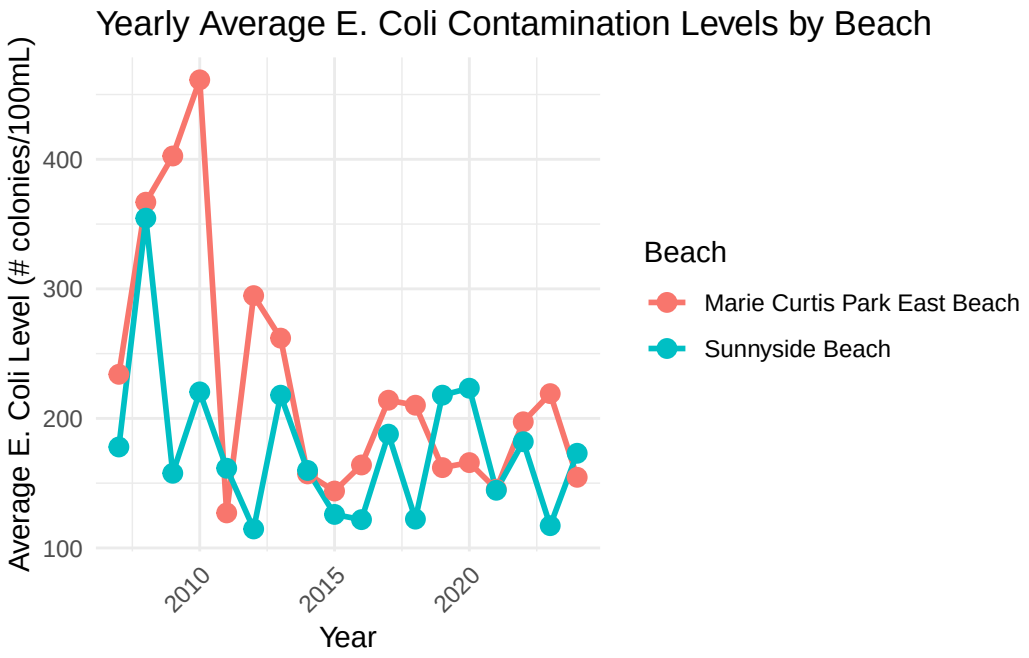

Figure 2: Yearly average amount of E. Coli contamination for 2 Toronto beaches

Figure 3 highlights the distribution of E. Coli levels for both beaches. The graph shows how often a specific level of contamination of E. Coli was recorded and is broken down into seven intervals. The main level of concern is from 0-100 colonies per 100 mL of water sample and anything above it. 100 colonies of E. Coli per 100 mL of water sample is the threshold at which Ontario deems the water unsafe for recreational use and should be flagged and closed.

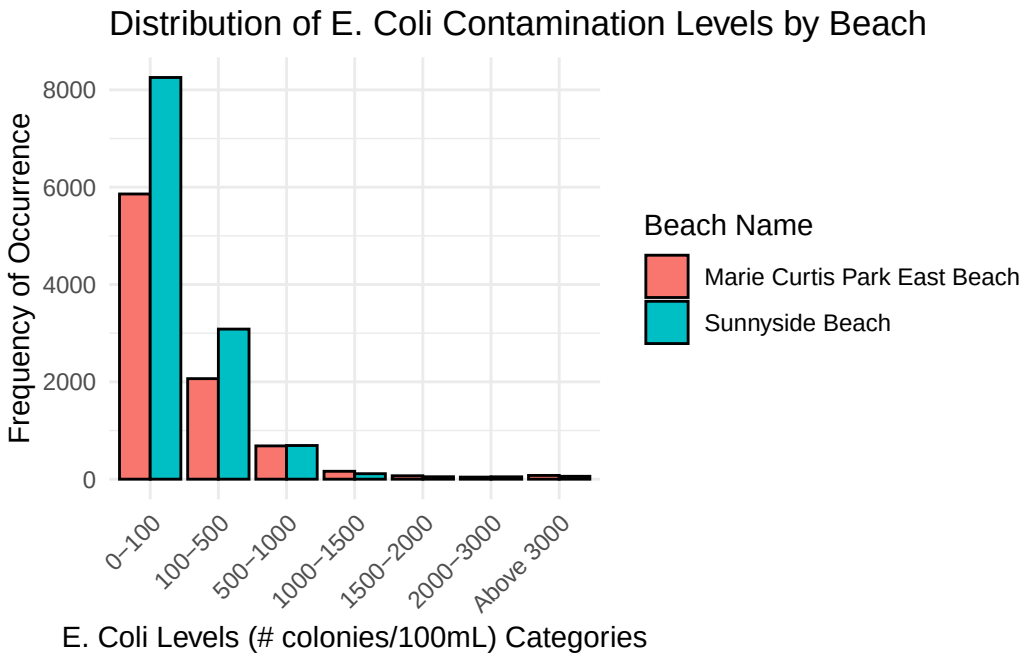

Figure 3: Distribution of contamination levels

Figure 4 is a depiction of a possible relationship between the E. Coli contamination levels of the two beaches and the date the sample was collected. The goal was to determine whether a specific date or month was correlated with higher contamination trends. However, from the figure, there is no apparent trend or relationship between time and contamination levels of the two Toronto beaches.

4 Discussion

Overall, the data reveals that Marie Curtis Park East Beach consistently had higher levels of contamination, measured through the number of colonies of E. Coli per 100mL of sample water, illustrated by Figure 1, Table 2, and Figure 2. Additionally, from Figure 3, the majority of days measured by the City of Toronto is between 0 to 100 colonies per 100 mL of water, which is under the safe threshold deemed by Ontario, so it is safe to swim in. However, in extremely unsafe intervals, such as above 1000 colonies per 100 mL of sample water, Marie Curtis Park East Beach has more recorded instances than Sunnyside Beach, suggesting it is likely much more hazardous for recreational use. It's important to keep this in mind when participating in activities in these waters; make sure to check contamination levels beforehand, as swimming in highly contaminated waters can lead to illness and other harmful health consequences.

Moreover, both beaches are frequently contaminated with chemical spillovers from the neighboring creeks, such as Etobicoke and Humber Creek. It is not just potential pathogenic

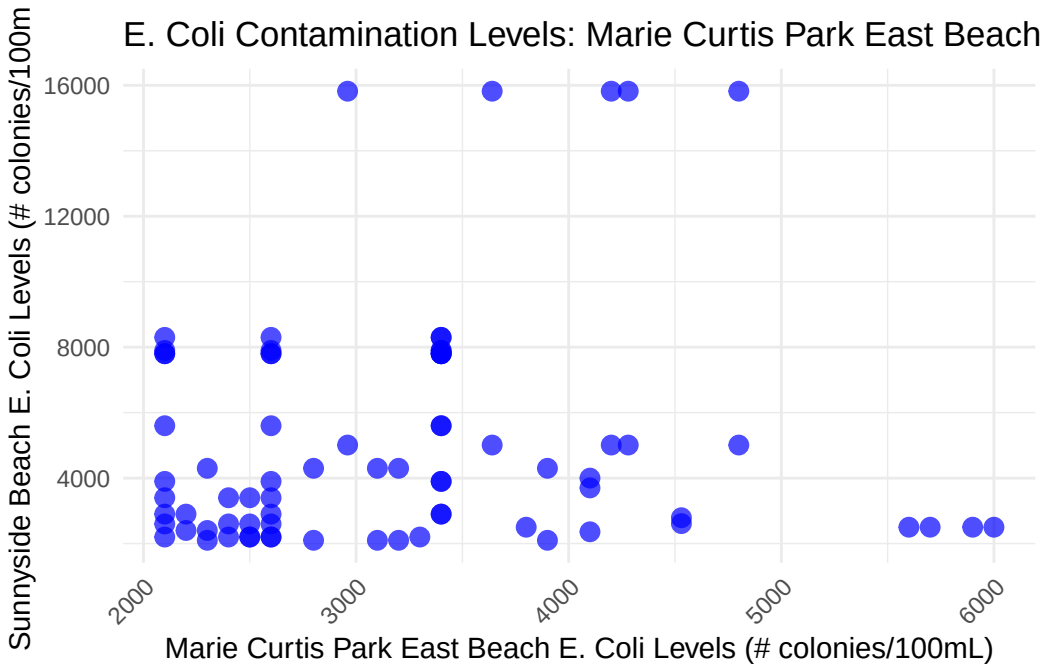

Figure 4: Possible correlation between date and contamination level

bacteria that are in these waters, but dangerous chemicals too. A potential explanation for Marie Curtis Park East Beach's consistent high contamination over Sunnyside Beach is that Etobicoke Creek is frequently the victim of pollution. Within the past two years, there have been multiple major pollution events affecting Etobicoke Creek, with major runoff into the water of Marie Curtis Park East Beach, which could explain the upward trend in Figure 6 for years after 2020 and before 2024. For example, in September 2023, Brenntag Canada, a chemical company, reported a fire in one of its industrial buildings near Etobicoke Creek and neighboring water, resulting in chemical and oil spillage Swadden (2023). The contamination was not cleaned up promptly and caused much of the ecosystem to collapse, struggling to recover. The chemical spillage potentially creates a favorable environment for additional bacteria to grow, causing E. Coli and other bacteria to increase in average number of colonies present.

While the dataset and paper provide better insight into the contamination levels of Toronto beaches, limitations do exist. First, the data collected is only from June to August of each year. While these months are more crucial because of the beach season, a more thorough analysis can be done if year-round data were available. This could potentially give more insight into trends concerning time; for example, if specific months had higher levels of contamination than others. This would also help with future policies directed at keeping Toronto's water clean. Another limitation is that both beaches in the dataset are geographically located near Etobicoke Creek. While Marie Curtis Park East Beach is a lot closer to the creek, this may

influence the sampling data. For a broader scope of Toronto beach water quality, it can be useful to collect data from beaches that are evenly spaced out.

5 Conclusion

The results of this investigation can help local and provincial governments, as well as the general public, understand and address gaps in overall trends of water contamination, providing important insights into location-specific sanitation. Additionally, the findings can help better inform policymakers by highlighting whether contamination levels are alarming, providing a basis to decrease health risks of its citizens.

References

"Action Required: Toronto Beaches Plan." 2009. https://www.toronto.ca/legdocs/mmis/2009/ex/bgrd/backgroundfile-18572.pdf.

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

Ishii, Satoshi, and Michael J. Sadowsky. 2008. "Escherichia Coli in the Environment: Implications for Water Quality and Human Health." *Microbes and Environments* 23 (2). https://doi.org/10.1264/jsme2.23.101.

Price, Robert G., and Dirk Wildeboer. 2017. "E. Coli as an Indicator of Contamination and Health Risk in Environmental Waters." https://doi.org/10.5772/67330.

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

Swadden, Patrick. 2023. "Residents Frustrated with Cleanup of Etobicoke Creeks 1 Month After Industrial Fire." https://www.cbc.ca/news/canada/toronto/mimico-creek-cleanup-one-month-later-1.6961229.

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, and Kirill Müller. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2014. *Knitr: A Comprehensive Tool for Reproducible Research in R*. Edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.