Water Quality at Toronto Beaches*

An analysis of e-coli rates in toronto beaches 2007-2024

Jacob Gilbert

September 24, 2024

Introduction

Water contamination is important in Toronto, especially in the summer months as locals flock to the nearby beaches as the weather turns from snow to sun. E-coli testing is common at most public swimming locations as high levels highlight risk to those embarking into the water. While E-coli is unlikely to cause illness on it's own, If a water sample has high concentrations of E. coli, other more dangerous and infectious organisms may be present.(https://www.cleanlakesalliance.org/e-coli/). Thus the levels at public beaches can affect public health, and as such action should be taken on the results. To either close public beaches that exhibit unsafe levels, or continuation of previous action that has led to lower levels. Despite these public health measures, there is limited understanding of the specific patterns of E. coli contamination in Toronto's beaches. Most public advisories and beach management decisions rely on reactive measures rather than proactive analysis of historical trends. This paper aims to fill this gap by examining historical E. coli data from Open Data Toronto, focusing on two of the city's popular beaches: Sunnyside Beach and Marie Curtis Park East Beach. By analyzing data over several years, this study seeks to identify temporal patterns in contamination, such as variations by day of the week and season, to help answer the critical question: When are beachgoers at the highest risk of swimming in contaminated waters? Additionally, this analysis explores whether E. coli levels have shown improvement over time, informing on the effectiveness of current water management practices. Understanding the current and historic trends are paramount for city officials to understand if current water quality measures are sufficient. Or in the worse case, that interventions are needed in order to keep Toronto beaches safe for everyone. This paper is structured as follows: First, we present the dataset, describing its sources, key variables, and the data cleaning process. Next, we analyse patterns in E. coli levels, examining how contamination varies across different times and locations. We then discuss the implications of these findings for public health and beach

*Code and data are available at: https://github.com/JfpGilbert0/toronto-beaches-water-quality.

management, followed by a conclusion that summarises the key insights and recommendations for future action.

Referencing

In this paper, Python [] and it packages are utilized as the primary programming language for the following analysis and displaying the data. Key packages used include pandas [McKinney, 2010], which facilitated efficient data cleaning, manipulation, and summarization of the dataset, NumPy [Harris et al., 2020] was employed for numerical operations, Matplotlib [Hunter, 2007] and seaborn [Waskom, 2021] were critical for creating informative visualizations.

Data

Data Source

Opendatatoronto is a publicly available collection of wide ranges of data. This resource was used to obtain the e-coli levels of the water at two of Toronto's many beaches. The data set included coli levels from the SunnySide and Marie Curtis park east beaches, from the 3rd of July 2007 to September 5th 2024. The large date range allows us to extrapolate trends well and further the recency of the data allows us to generate a good understanding of the present. We have a total of 3084 data points from the two beaches, 1370 and 1714 data points respectively. Each detailing the collectiondate, beachname, sitename, geometric location (given by longitude and latitude) and the e coli levels.

(Summary stats)

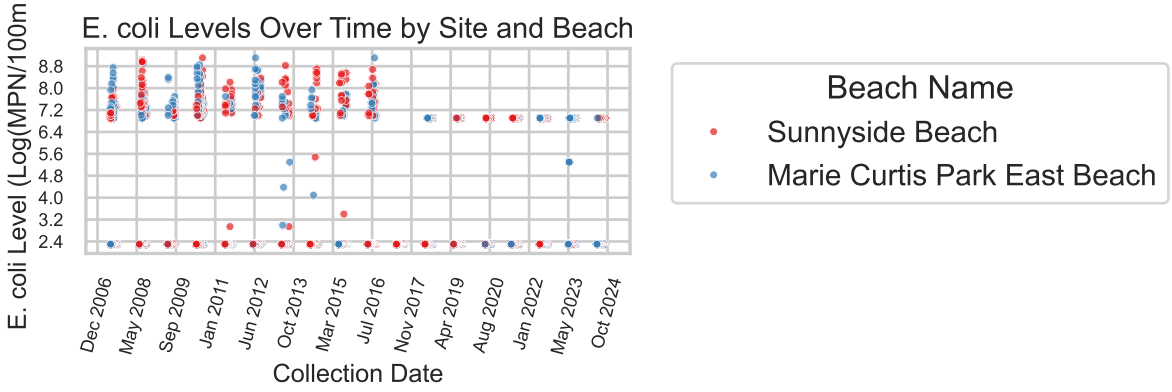
Measurement

E-coli is measured in MLN/100ml a measurement that reflects the ecoli particles in evey 100ml of lake water. For our data presentation log values are used to give a better visualisation of the data. To obtain an accurate representation of the beach as a whole 12 different locations here used to gather data across the 2 beaches as to remove bias that one testing location may result in. the result is that we obtain an unbiased representation of ecoli levels at each beach as a whole over more than a decade. Some missing data we encounter however is that although we have the collection date we do not observe the time that data was collected. Missing this data may lead to some bias as the time of day could impact the baseline levels of e-coli in the water. What this form of measurement does allow us to do is look at the differences in water quality at a date level, a figure() shows there is a good distribution of data from ach day of the week.

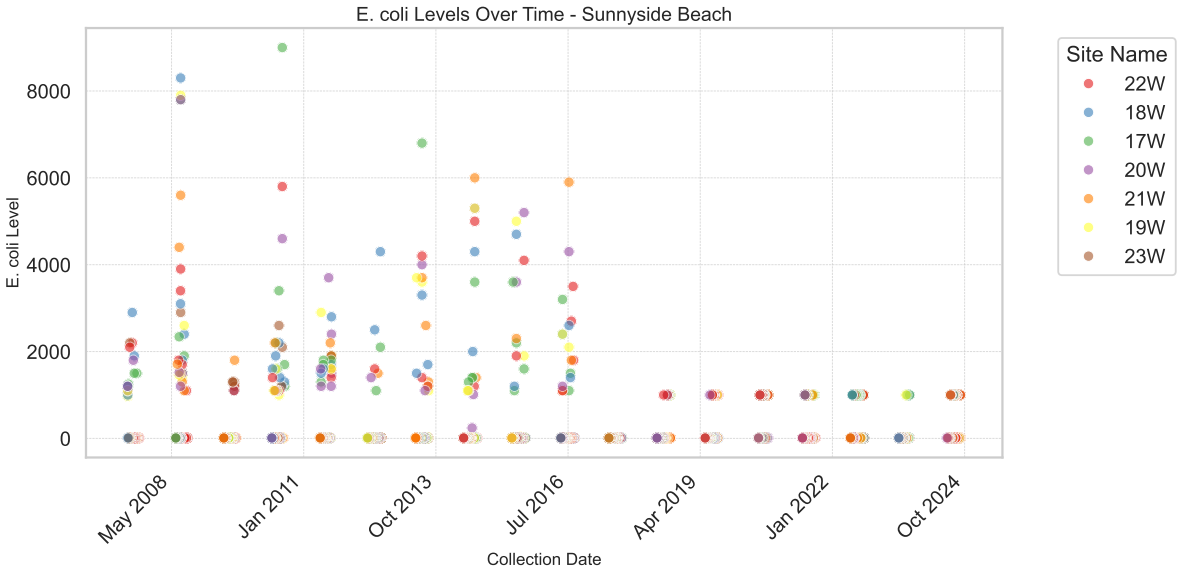
Methods

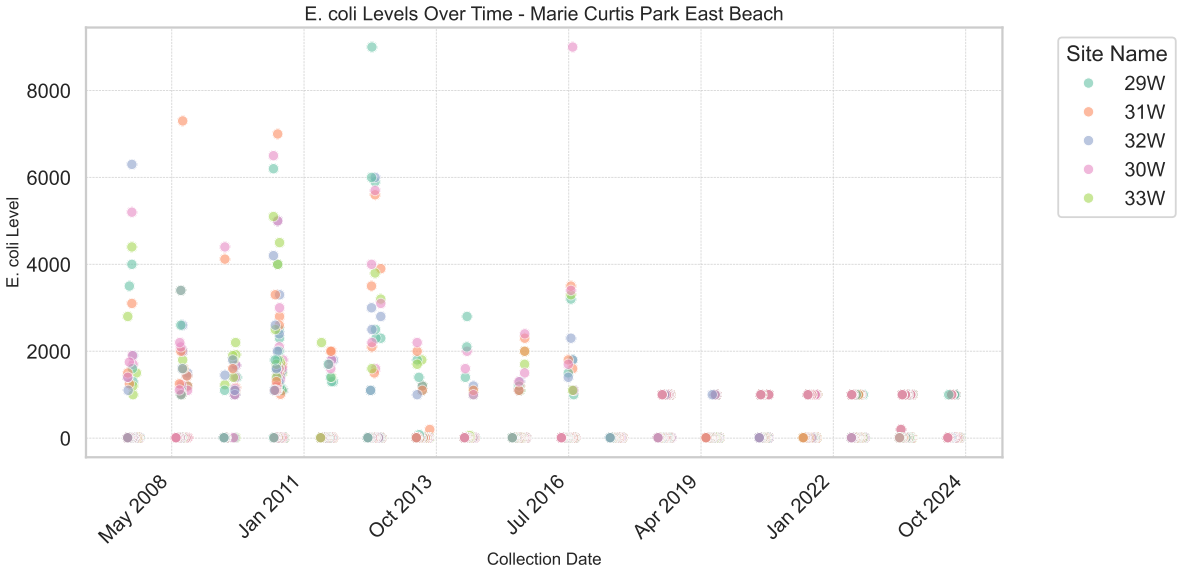
Before we look at the daily differences in levels we will look at the data as a whole, observing the distribution of higher and lower levels of ecoli in the water across time as well as by beach. Using canada.ca we observe that they define unsafe levels of ecoli in drinking water to ber 88ML/100ml. As such we will use this in our analysis to observe the frequency that "unsafe" levels are observed in the beach water. Finally we will look at these distributions over the course of a week to answer the question of when water quality is at its best, and as a lake swimmer when is it safest for me to swim at these beaches.

Results



Figure() displays our data as a whole, we see very widely distributed values of what could be considered high levels of e-coli from 2007 to 2016. We see the highest density of data at 10 MLN/100ml but on the higher end levels range from 1000 to 9000. Closer to the present we do not see as much variation on the higher end. In fact post 2017 the data groups either at the low level of 10 or at around 1000 almost exclusively. The variations observed appear to not vary by beach in figure.This is highlighted in the following figures where we see similar distributions of values when separating the data by beach. The following two figures alo highlight that the site locations appear in the data fairly randomly supporting that these different sites are not providing much bias.



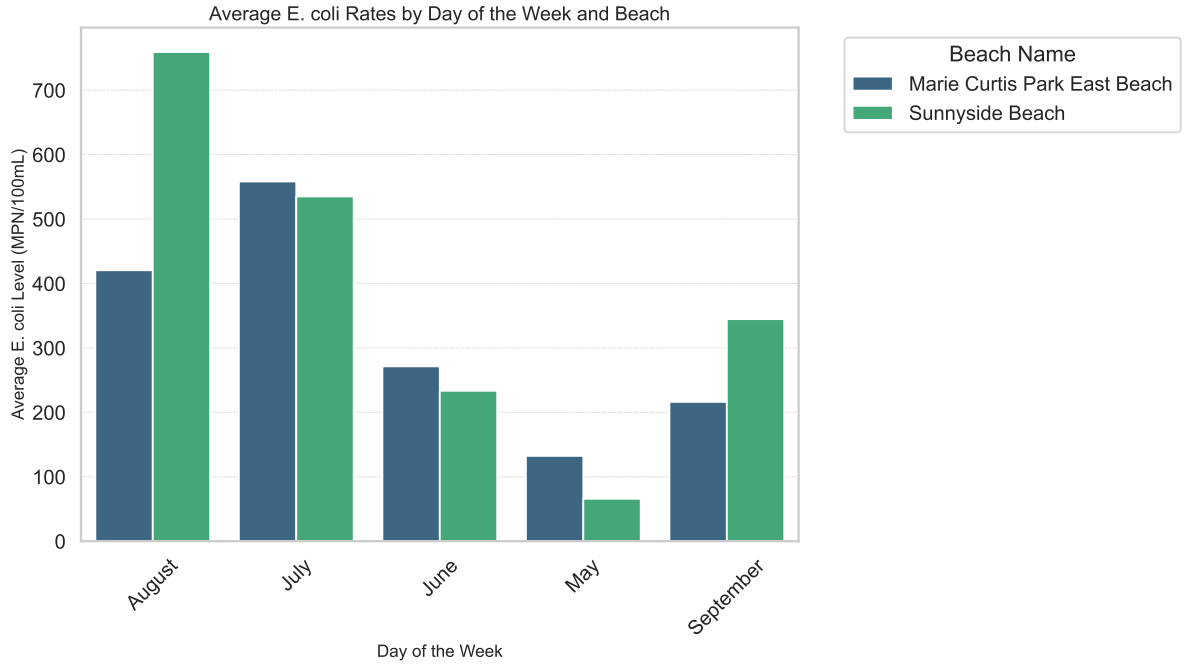


Monthly data

	month	n	average_ecoli	variance	unsafe_levels	percentage_unsafe
0	May	341	101.759531	1.673263e+05	21	6.16
1	June	1046	252.380497	5.981389e+05	146	13.96
2	July	910	548.439560	1.296202e+06	277	30.44
3	August	681	549.321586	1.058926e+06	224	32.89

	month	n	average_ecoli	variance	unsafe_levels	percentage_unsafe
4	September	106	270.754717	2.303556e+05	26	24.53

Figure 4 looks at the monthly data at each beach, as mentioned in the data section only th summer months provide data due to the cold weather in toronto limiting available data in the winter. This data can still provide useful data. May exhibits the lowest average E. coli levels, at 101.76 MPN/100mL, which is just above the unsafe threshold. With only 6.16% of samples being unsafe. The low variance suggests that water quality is fairly consistent during this early summer month. August and July show extremely high avrages, above 500MPN/100ml, with over 30% of daa collected here showing water quality above acceptable levels.



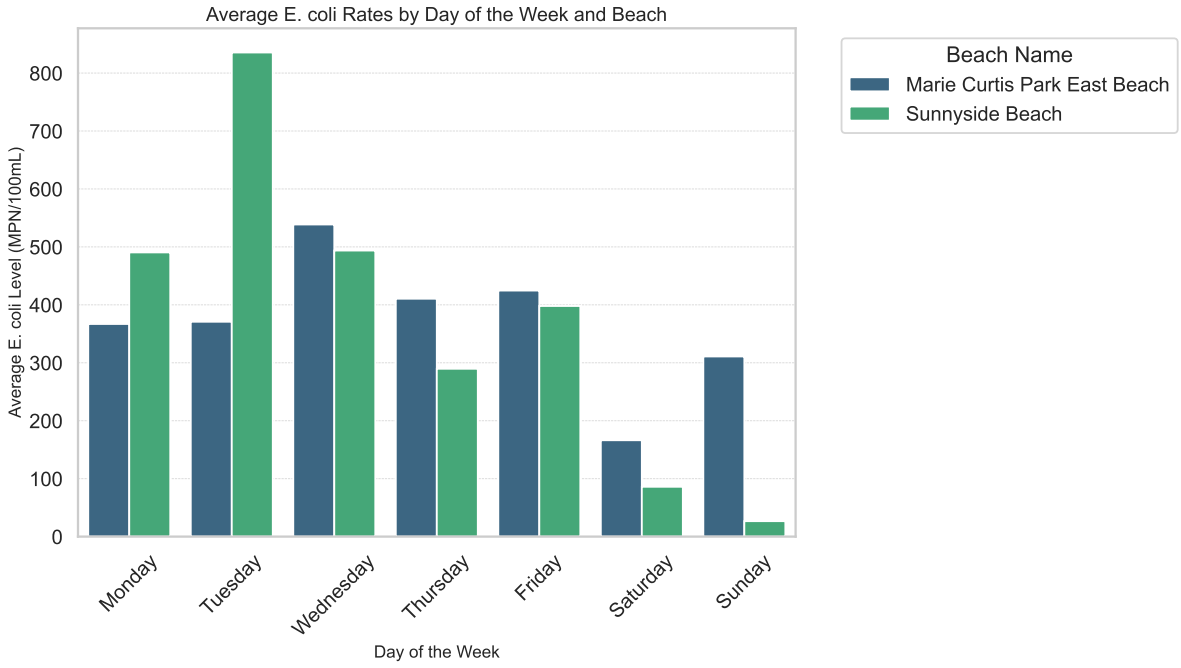
Weekly Data

	day	n	average_ecoli	variance	unsafe_levels	\
0	Monday	456	426.271930	7.933957e+05	118	
1	Tuesday	443	592.142212	1.457117e+06	135	
2	Wednesday	494	518.461538	1.463763e+06	136	
3	Thursday	488	357.540984	6.444026e+05	110	
4	Friday	467	413.511777	5.903443e+05	138	
5	Saturday	374	131.307487	2.182056e+05	31	
6	Sunday	362	196.243094	6.398706e+05	26	

```
   percentage_unsafe
0              25.88
1              30.47
2              27.53
3              22.54
4              29.55
5               8.29
6               7.18
```

	day	n	average_ecoli	variance	unsafe_levels	percentage_unsafe
0	Monday	456	426.271930	7.933957e+05	118	25.88
1	Tuesday	443	592.142212	1.457117e+06	135	30.47
2	Wednesday	494	518.461538	1.463763e+06	136	27.53
3	Thursday	488	357.540984	6.444026e+05	110	22.54
4	Friday	467	413.511777	5.903443e+05	138	29.55
5	Saturday	374	131.307487	2.182056e+05	31	8.29
6	Sunday	362	196.243094	6.398706e+05	26	7.18

Figure X summarises E. coli data by day of the week, including the total number of samples (n), average E. coli levels (average_ecoli), variance, the number of samples with unsafe levels (unsafe_levels), and the percentage of samples deemed unsafe (percentage_unsafe). The weekdays, Monday through Friday, see the highest average amounts of e-coli, as well as the higher probability that the water is deemed unsafe. Tuesday and Wednesday see average levels over 5x the 100 MPN/100ml unsafe , with relatively low variance highlighting these days as the worst for water quality over the course of a week. This point is further supported with both days having around 30% of their levels above the threshold. The weekend encounters the lowest levels by comparison, 134.91 MPN/100mL and 196.34 MPN/100mL on saturday and sunday respectively. with under 10% of tests having an unsafe result. This Is a positive sight as these days, it is assumed, encounter the most activity due to the 5 day work week. Figure() gives context to how the two beaches differ in their levels over the course of a week. Marie Curtis see's fairly consistent averages across the week, between 300 and 600 MPN/100ml. We still observe the lower average observed in the table here however it is far more significant at this location on Saturday compared to Sunday. As is clear from the graph we see significant differences in the average in the two locations. Sunnyside Beach shows extraordinarily high levels on Monday and Tuesday Compared to Marie Curtis. The rest of the weekdays we observe a fairly insignificant difference between the locations average level, but Sunnyside does appear to have a better average quality. The weekend we see the inverse to the start of the week, with Sunnyside having very low averages. In fact Sunday at Sunnyside is the only category here that has average levels below the safe threshold of 100 MPN/100ml.


Average E. coli Rates by Day of the Week and Beach
Average E. coli Level (MPN/100mL)
0
100
200
300
400
500
600
700
800
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Day of the Week
Beach Name
Marie Curtis Park East Beach
Sunnyside Beach

Conclusion