

Water Quality at Toronto Beaches*

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

Jacob Gilbert

September 23, 2024

Introduction

Referencing

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 every 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

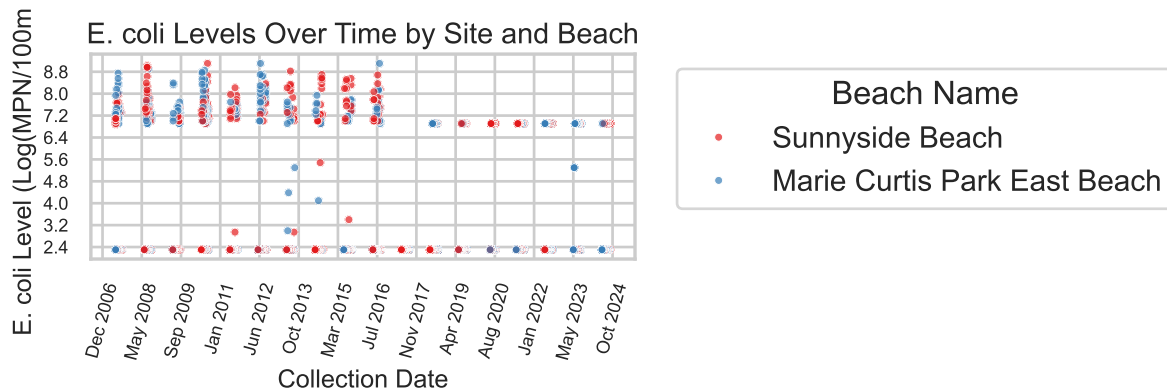
*Code and data are available at: <https://github.com/JfpGilbert0/toronto-beaches-water-quality>.

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

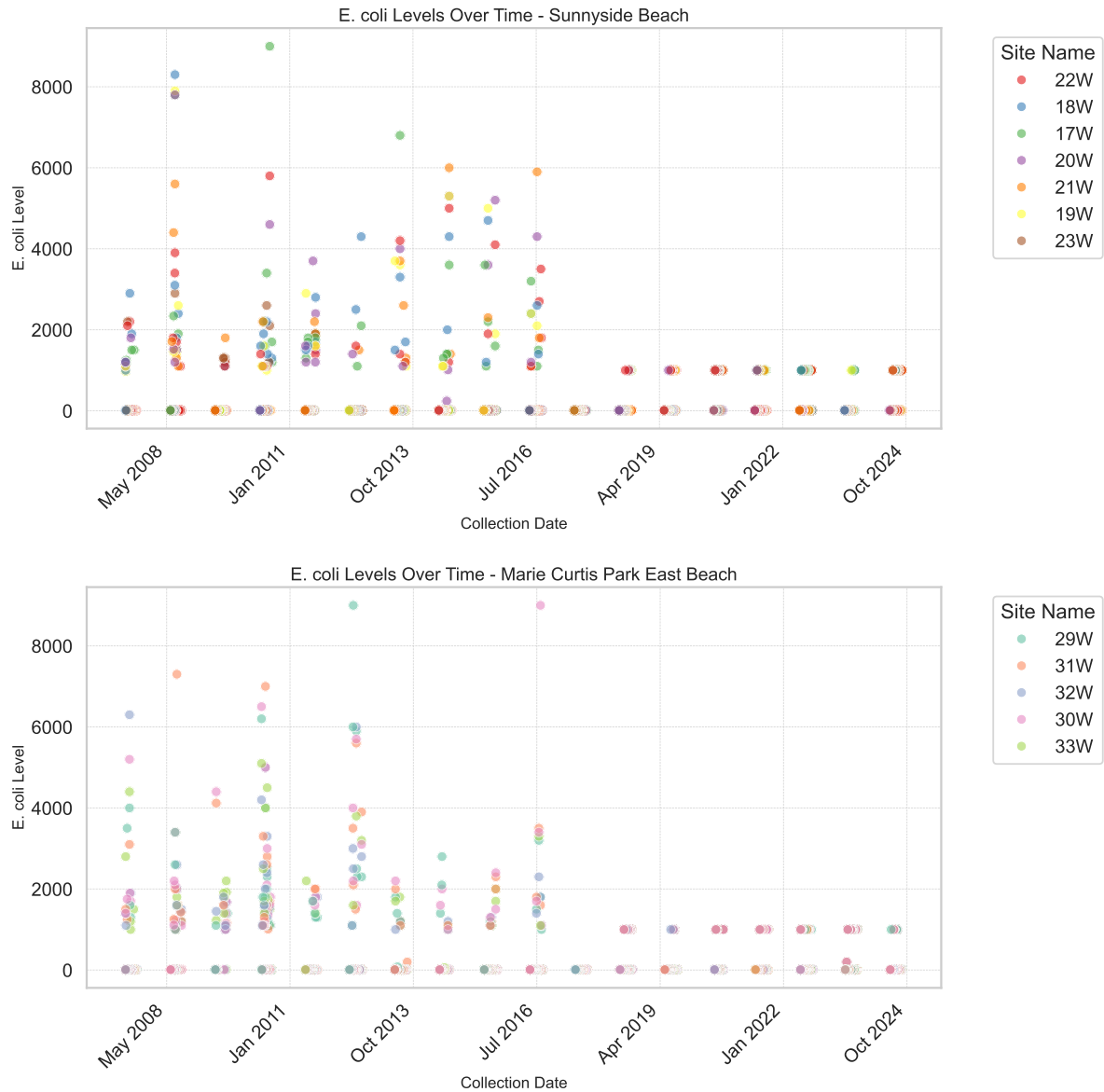
Before we look at the daily differences in levels we will look at the data as a whole, observing the distribution of higher and lower levels of ecoli in the water across time as well as by beach. Using canada.ca we observe that they define unsafe levels of ecoli in drinking water to ber 88ML/100ml. As 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



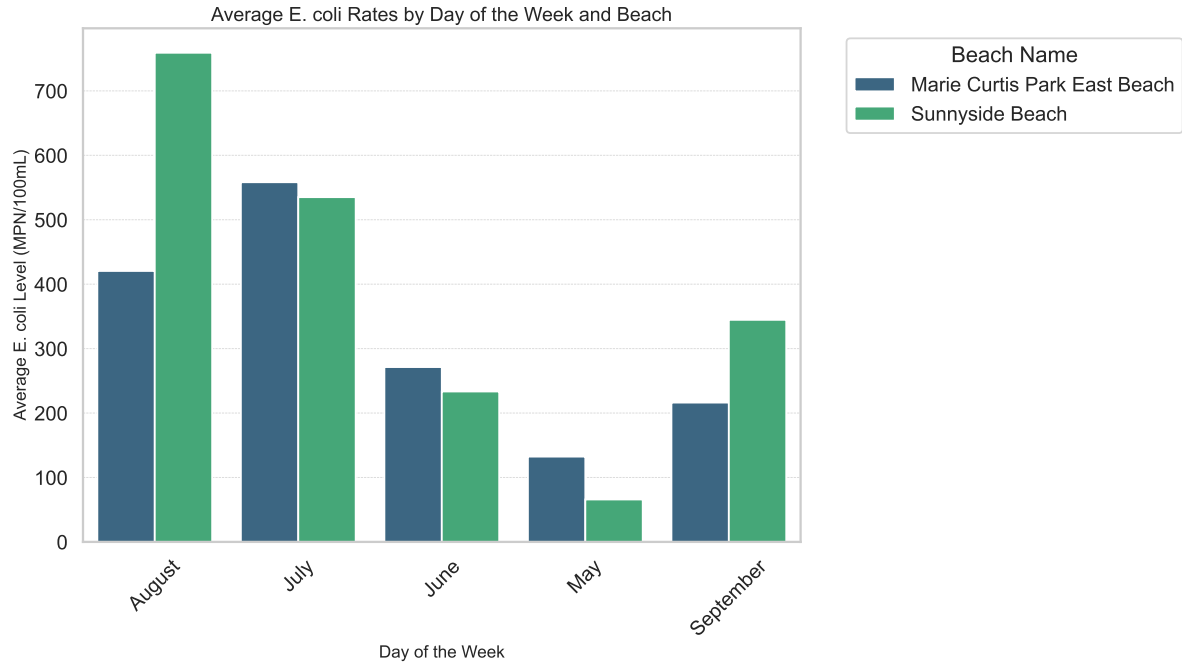
Figure() displays our data as a whole, we see very widely distributed values of what could be considered high levels of e-coli from 2007 to 2016. We see the highest density of data at 10 MLN/100ml but on the higher end levels range from 1000 to 9000. Closer to the present we do not see as much variation 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 also highlight that the

site locations appear in the data fairly randomly supporting that these different sites are not providing much bias.

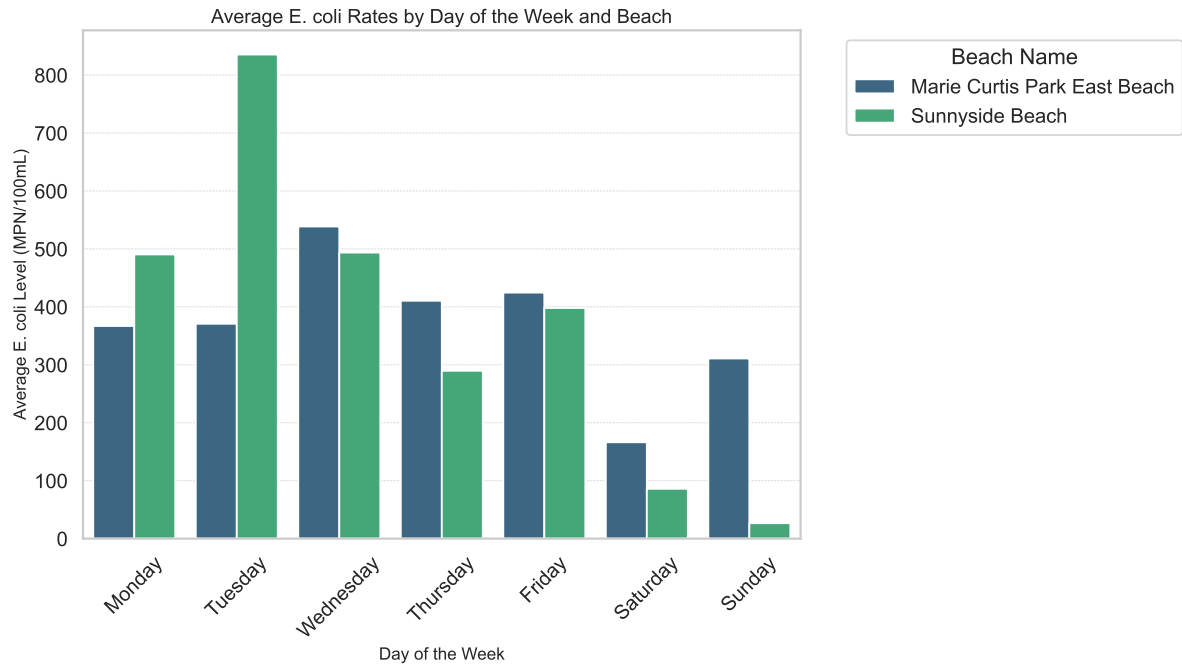


Monthly data

Figure 4 looks at the monthly data at each beach, as mentioned in the data section only the summer months provide data due to the cold weather in Toronto limiting available data in the winter. This data can still provide useful data.



Weekly Data



Conclusion