Bacterial Sources of Pollution in Urban Stormwater Impacted Bodies of Water

Megan Beaudry

2019-11-07

#load clean data  
clean\_water\_quality\_data <- readRDS("../../data/processed\_data/clean\_water\_quality\_data.rds")

# Introduction

Although there is an abundance of literature on enteric bacterial pathogens in water systems, there is a lack of data on their presence in urban stormwater. Poor water quality within urban stormwater-impacted bodies of water represents a potentially important public health problem, as many of these water bodies are subjected to chronic issues with human and animal sources of fecal contamination, inferring the potential for zoonotic and anthropogenic enteric bacterial pathogens to be present. Pathogens as Arcobacter butzleri (Douidah, 2011) (Van Driessche & Houf, 2005) (Levican, 2013) [Hafliger2000] (Craun, 2005), and other enteric bacterial pathogens (e.g., Campylobacter spp., Salmonella spp., Shiga-toxin producing E. coli) (Moore, 2001) (Krometis, 2010) (Rangel, 2005) (Adams, 2016), have all been implicated in waterborne outbreaks throughout the world. Rain events mobilize and transport fecal pathogen in the environment thereby increasing the effects of non-point and point sources of contamination, which in turn can augment the risk to public health (Staley, 2018). Stormwater-impacted bodies of water can serve as reservoirs for transmission for enteric bacterial pathogens by the fecal-oral route through ingestion of contaminated water during recreational activities (e.g., swallowing water while swimming); by way of contaminated irrigation water on food that is then eaten; or by accidental ingestion that occurs during irrigation. An outbreak of HUS associated with a recreational water body in Connecticut, USA, occurred where STEC was detected in a storm drain that emptied onto the beach (McCarthy, 2001). In a study of stormwater discharges and gastrointestinal illness following wet weather in California, USA, (Soller, 2017) that wet weather exposure during surfing lead to higher than average illness rates due to human enteric viruses. In addition, they observed Campylobacter spp. above the method detection limit in over half of their samples from stormwater discharges. (Meng, 2018) identified Campylobacter spp. in stormwater constructed wetlands intended for reuse activities, and found the concentrations to be similar between wet and dry weather events. Furthermore, they found that log reduction targets for reuse activities were not being met (Meng, 2018). The Arcobacter genus was created in 1991 in an effort to accommodate aero-tolerant Campylobacter spp. (Vandenberg, 2004). Arcobacter spp. are gram-negative, curved bacteria (Van Driessche & Houf, 2005). Currently, there are 22 species (Van Driessche & Houf, 2005). Potential sources of Arcobacter spp. include humans, birds, and livestock. This genus has been characterized as a potential food or waterborne pathogen; and has been implicated in causing human disease, with such symptoms as bacteremia, diarrhea, and gastroenteritis from three species: A. butzleri, A. cryaerophilus, and A. skirrowii (Kayman, 2012). A. butzleri contains the ability to survive and grow in the environment. Further, it can survive at lower temperatures (15-30˚C) than Campylobacter spp.; and can grow in the presence of oxygen (Van Driessche & Houf, 2005) (Wesley, 2000). Arcobacter spp. are considered a zoonotic pathogen, and one study identified it as the fourth most frequent bacteria isolated from humans with acute enteric disease (Levican, 2013). The objective this study was to identify and determine the prevalence of the putative enteric bacterial pathogen Arcobacter butzleri in urban stormwater ponds located in western Canada, as there is currently limited information as to the critical role of enteric bacterial pathogens, in particular A. butzleri, in outbreaks associated with stormwater reuse. Furthermore, we sought to resolve the source of the enteric bacterial pathogens by using microbial source tracking techniques and analyzing environmental parameters.

Note that this cited reference will show up at the end of the document, the reference formatting is determined by the CSL file specified in the YAML header. Many more style files for almost any journal [are available](https://www.zotero.org/styles). You also specify the location of your bibtex reference file in the YAML. You can call your reference file anything you like, I just used the generic word references.bib but giving it a more descriptive name is probably better.

# Introduction (required for part 1)

## General Background Information

*Provide enough background on your topic that others can understand the why and how of your analysis* Water reuse provides countries with a strategy to address the growing demands on water resources due to climate change, industrial changes, and projected population growth. However, water reuse poses a variety of challenges. In particular, stormwater has the potential to be of low water quality, to be contaminated with human and animal feces, and thus enteric bacterial pathogens including Campylobacter spp., Salmonella spp., and pathogenic E. coli. In this project we aim to better understand the relationship between the presence of the putative enteric bacterial pathogen Arcobacter butzleri and other water quality indicators and risk factors, as the contamination of water with human waste, and thus microbial pathogens, poses a significant threat to public health. Furthermore, other studies have shown the presence of other enterical bacterial pathogen (i.e., Campylobacter spp, Salmonella spp.) but this study is the first of its kind to indicate the presence of A. butzleri.

## Description of data and data source

*Describe what the data is, what it contains, where it is from, etc.* This data was collected over the course of one sampling season (May-September 2017, when the water is not frozen) from select stormwater ponds in Alberta, Canada. Samples were taken biweekly over 21 weeks to catch any temporal trends that may occur. The focus of this project will be on one specific pathogen putative enteric bacterial pathogen, Arcobacter butzleri, and its relationship to other water quality indicators (i.e., E. coli and Enterococcus) and microbial sources of pollution tested (i.e., Human, Dog, Muskrat, Ruminant, Birds, and Canadian Goose). Water quality indicator data is from culture based methods. Microbial source tracking and pathogen data is from molecular based methods. We will find this data that many times there is no value, so “do not detect is assigned.” These are important to keep as even though nothing is detected we cannot technically assign these as a value of 0, because we cannot definately say that nothing is there.

I generated this data as part of my master of science degree. None of this data has been published. I am hoping to generate publishable figures in this class for a publication focusing on A. butzleri in urban stormwater ponds.

There are 5621 observations for 25 variables. ## Questions/Hypotheses to be addressed *State the research questions you plan to answer with this analysis* 1. Do we see a relationship between human fecal contamination (HF183 and HumM2) and A. butzleri? 2. Do we see a realtionship between bird fecal contaminaion (LeeSg and CGO1) and A. butzleri? 3. Do we see an increase in the levels of A. butzleri during times of rainfall? 4. Do we see a relationship between traditional water quality indicators (E. coli and Enterococcus) and A. butzleri? 5. Are certain sampling locations or sites more prone to A. butlzeri contamination? 6. Are other enteric pathogens found when A. butzleri is also found?

#Outcomes of interest There are several outcomes of interest for this project. Foremost, A. butzleri is a putative pathogen. Showing if it has relationship with human fecal material will help to establish its importance, as human fecal material is a risk factor for illness in recreational bodies of water. Further, water quality indicators are supposed to be indicators for pathogens such as A. butzleri. I would like to test how accurate that is.

#Brief overview of data

library(RCurl)

## Loading required package: bitops

data <- getURL("https://raw.githubusercontent.com/epid8060fall2019/MeganBeaudry-Project\_WQ/master/data/raw\_data/9-19-19%20raw%20data%20epid%20project.csv")  
dataWQ <- read.csv(text = data)  
str(dataWQ)

## 'data.frame': 0 obs. of 1 variable:  
## $ X404..Not.Found: logi

# Materials and Methods

Sampling: To determine what select enteric bacterial pathogens present, stormwater samples were collected from stormwater ponds southern Alberta, Canada. Semiweekly sampling began as soon as stormwater ponds were fully thawed (i.e., May 9th, 2017), and ended just before freezing (i.e., September 25th, 2017) at three stormwater ponds in Calgary, Alberta, Canada: McCall Lake, Country Hills Stormwater Facility, and Inverness Stormpond. At each pond, four (i.e., McCall Lake and Inverness) or five (i.e., Country Hills) locations were sampled 41 times. During sampling, 500 mL of water was collected into two sterile 250 mL Nalgene polyethylene bottles (Nalgene, Rochester, NY, USA) (Systems Plus, Baden, Ontario, Canada) by hand using sterile gloves. Following collection, were sent for pathogen analysis to the University of Alberta (Edmonton, AB, Canada) within 24 hours.

Culture Methods for Bacterial Water Quality Indicators: Traditional water quality indicators were analyzed using defined substrate culture methods (IDEXX, Westbrook, ME, USA) for E. coli and total coliform detection using Colilert Quantitray-2000®. All samples were analyzed according to guidelines from the manufacturer. The Quantitrays® (IDEXX, Westbrook, ME, USA) were sealed and incubated for 24 hours at 35˚C. Following incubation, most probable numbers (MPN) were determined by scoring the number of positive wells (Colilert® – yellow wells = total coliforms, yellow and fluorescent = E. coli) and transforming results into quantitative estimates using IDEXX MPN charts. If the Quantitrays® were positive for E. coli, they were set aside for further analysis. Enumeration of fecal coliform bacteria was performed by the Provincial Laboratory in Calgary, Alberta, Canada, in which water (10 mL) was filtered onto a 15x16 mm membrane that was placed on a Membrane Fecal Coliform (mFC) plate (Dalynn Biological, Calgary, AB, Canada), then incubated at 44.5 ˚C for 24 hours. Fecal coliform bacteria were enumerated according to standard operating procedures for the Provincial Lab of Public Health and in alignment with current practices of recreational water assessment in the province of Alberta, Canada.

Molecular Methods: DNA extraction. Immediately upon receipt of stormwater samples in the laboratory, 20 mL of water was filtered onto 0.4 micron polycarbonate filters by EPA method 1611 (Environmental Protection Agency, 2012) by the Provincial Laboratory in Calgary. Filters were then stored at -80˚C until they were shipped to Edmonton, Alberta, Canada on dry ice. Controls. In order to ensure amplification, plasmids or positive samples were used as positive controls. IAC was used as a control. DNA free water was used as a negative control, prepared in the same manner as water samples to ensure no carryover. qPCR Assays. qPCR assays included targets for bacterial indicators (i.e., Enterococcus spp.), microbial source tracking markers (i.e., human, cattle, seagull, goose, dog, and muskrat), as well bacterial pathogen-related genes (i.e., Campylobacter spp., Salmonella spp., and A. butzleri). Two markers were used for human contamination to ensure that the results were not due to cross-reaction with other fecal sources (e.g., dog, turkey, and chickens)(Green, 2014). Each marker has a different limit of detection as assayed previously in the laboratory. Each assay target species, target locus, and primer/probe names and their corresponding sequences are specified. Amplification was performed on an Applied Biosystems TaqMan 7500 fast real-time PCR system (Applied Biosystems, Foster City, CA, USA). The reaction mixture was made to a final volume of 20 µL. The calibrator control standards and plasmid controls were performed in triplicate. The negative controls, Enterococcus spp., and salmon DNA PCR assays were performed in duplicate. Sample assays were only performed once. All reactions were carried out in a MicroAmp Fast Optical 96-Well Reaction Plate (Applied Biosystems, Foster City, CA, USA). Samples were pulsed down at maximum speed in a centrifuge prior to thermal cycling. Thermal cycling conditions were 50°C for two minutes; 95°C for 30 seconds (holding); followed by 45 cycles of 95°C for three seconds; and 60°C for 30 seconds for all assays, except for the muskrat marker MuBac. The annealing temperature of MuBac was 57°C. The threshold cycle (CT) was set at 0.1 for Entero1 and Sketa, and 0.05 for all other targets. For further analysis, all CT values were entered into Microsoft Excel. DNA plasmid standards for each assay were developed previously in the laboratory. Briefly, DNA targets were PCR amplified, run on a 2% agarose gels and amplicons extracted by a QIAquick (Qiagen, Hilden, Germany) gel extraction kit. The products were cloned into pCR2.1-TOPO (Thermo Fisher Scientific, Waltham, MA, USA) per manufacturer’s instructions using TOP10 F’ E. coli competent cells. The plasmids were mini-prepped using QIAprep spin mini prep kit (Qiagen, Hilden, Germany). Plasmids were then quantitated using the Qubit 2.0 fluorimeter (Invitrogen Carlsbad, CA, USA) and diluted to 100,000 copies/µL stocks. Thus, stocks of plasmid DNA were aliquoted and stored at -80°C. During each qPCR analysis, 10-fold serial dilutions were made from the known concentration plasmid for each individual target. For further analysis, all CT values were entered into Microsoft Excel and volume corrected pior to data analysis using R statstical software.

## Data import and cleaning

The data was cleaned to format the variables (e.g., DNQ, not detects, >2419.6) in order to deal with limits of detection and detection limits . In the raw data, detectable but not quantifiable (DNQ) had not been assigned a value. See supplementary material - cleaning.

#Results *Your report should be more detailed than what would normally go into a peer reviewed paper. Include parts such as results from your exploratory analysis. Those can be part of the main document or if you prefer a separate supplementary document.* *—–* = TABLE/FIGURE NUMBER

## Water quality

A high-level descriptive overview of the bacteriological water quality in each of these ponds, and at each of the sites, is provided in *—–*, and is based on the percentage of samples violating water quality standards/guidelines, as evaluated against: the USEPA’s recreational water quality guideline for Enterococcus by molecular methods (Environmental Protection Agency, 2012); and Alberta’s former recreational water quality standards based on thermotolerant coliform concentrations (*—–*). A number of observations are worth noting from this high-level analysis.

knitr::include\_graphics("../../results/entero\_pond\_table.png")

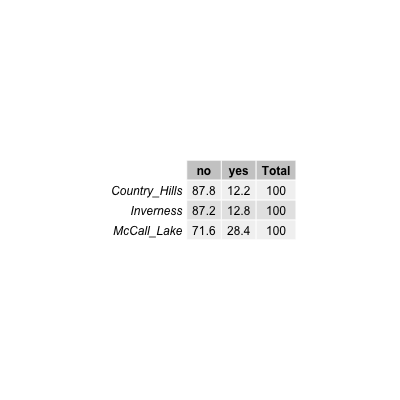


Figure 1: Frequency of occurrence of samples positive based on water quality indicator Enterococcus spp. in stormwater ponds in Calgary, Alberta.

knitr::include\_graphics("../../results/thermo\_pond\_table.png")

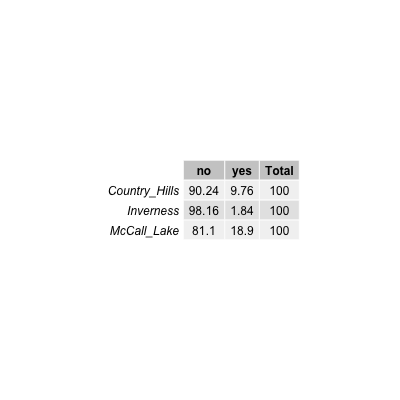


Figure 2: Frequency of occurrence of samples positive based on the water quality indicator thermotolerant coliforms of stormwater ponds in Calgary, Alberta.

Firstly, considerable spatial variation was observed with respect to the frequency of water quality failures among the urban stormwater ponds, with McCall Lake appearing to be the most contaminated of the three storm ponds. This result was true regardless of the bacterial water quality indicator chosen for analysis (i.e., Enterococcus or thermotolerant coliforms). Approximately 29% of all water samples taken at McCall Lake failed water quality guidelines for Enterococcus at the recommended STV or geomean values set out in the guidance documents (*\_\_\_\_*). Inverness Stormpond had the fewest water quality violations among the three ponds, also based on all bacteriological indicators examined, and therefore was considered to have the best water quality overall.

knitr::include\_graphics("../../results/entero\_y\_n\_site\_table.png")

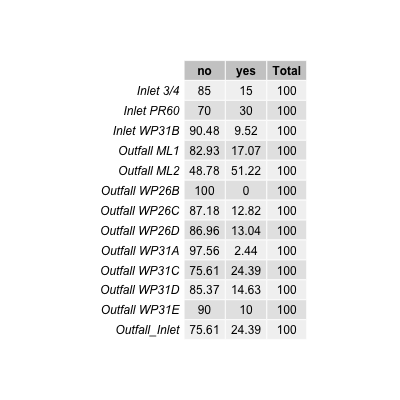


Figure 3: Frequency of occurrence of samples positive based on water quality indicator Enterococcus spp. broken down by sampling site in Calgary, Alberta.

knitr::include\_graphics("../../results/thermo\_y\_n\_site\_table.png")

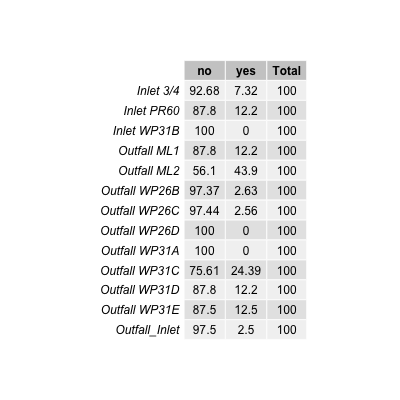
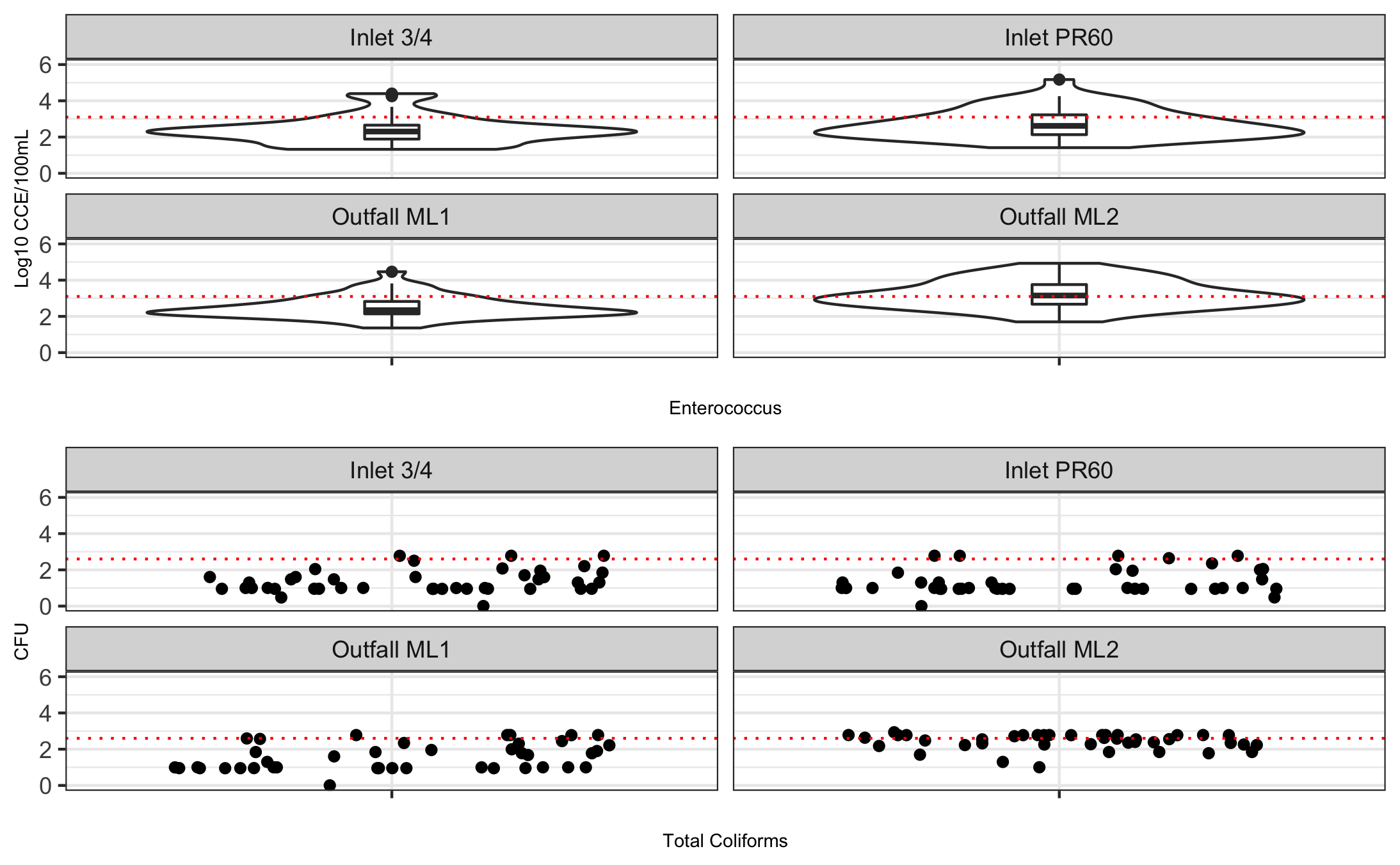


Figure 4: Frequency of occurrence of samples positive based on water quality indicator thermotolerant colifroms broken down by sampling site in Calgary, Alberta.

However, variation in bacteriological water quality was also observed among sampling sites within a single pond. The most contaminated site across all stormwater ponds examined was site ML2 at McCall Lake, with upwards of 51% of all samples failing the US EPAs Guidelines for Recreational Water Quality for Enteroccus spp. STV (\****). This site had the poorest water quality irrespective of the bacterial indicator used in the analysis. It is important to note, however, that ML2 was an above-grade outfall (*** and \_\_\_\*), thereby potentially explaining the more frequent bacteriological failures at this site as due to the fact that water samples were directly collected from the outfall and not after dilution into the pond.

Bacteriological water quality indicators were further assessed for all sampling sites (see univariate code), however due to the poor water quality at McCall Lake this assessment is included in the report.

knitr::include\_graphics("../../results/McCall\_WQ\_plots.png")

 Median levels of Enterococcus at the ML2 site approximated 3.1 log10 CCE/100 mL, whereas at all other sampling sites in McCall Lake (i.e., ML1, Inlet 3/4, and PR60), the median occurrence was almost an order of magnitude lower (~ 2.3 log10 CCE/100 mL) (*\_\_\_\_*). Incidentally, ML2 also had the largest overall interquartile variation in the concentration of Enterococcus during the study season (*—*). Concentrations of thermotolerant coliforms were also high at this site and followed a similar trend to that of Enterococcus (*—*).

It is important to note that in most cases for Enterococcus at sites other than ML2, there were several outliers in the data set (*\_\_\_\_*).Although outliers may reflect recent localized contamination events not necessarily reflective of overall water quality in the stormwater pond (e.g., aquatic birds in one area of the pond), their occurrence could also reflect the periods of peak contamination in stormwater ponds, and for which this effect may be contingent on temporal variables associated with water quality (e.g., first flush from storms, to be discussed later). Specifically, outliers for Enterococcus concentrations were represented by values higher than ~3.5 log10 for Inlet 3/4, ~3.75 log10 for ML1, and ~4 log10 for PR60.

This analysis revealed that stormwater ponds and stormwater-impacted urban rivers in southern Alberta, Canada, do not often meet existing guidelines or standards as laid out by USEPA, Health Canada, and Alberta Environments and Parks in terms of recreational water quality, surface water quality and/or irrigation water quality. For this reason, we sought to determine the potentiasl sources of poor water quality using microbial source tracking assays for humans, ruminants, dogs, gull, geese and muskrat.

Calgary stormwater ponds were mainly impacted by human and gull feces (*below*). The human specific markers, HF183 and HumM2, were detected at 42% and 16%, of samples, respectively in McCall Lake (*below*). On the lower end, they were dtected in 14% and 3% of samples at Inverness. The gull specific marker (i.e., LeeSg) was found in 15% of samples at McCall Lake (*below*). Of these, the more dominant source of fecal pollution was from humans (*below*). All other host-specific markers (i.e., dog, Canada geese, muskrat, and ruminants) were detected in <5% of pond samples (*below*).

knitr::include\_graphics("../../results/HF183\_y\_n\_pond\_table.png")

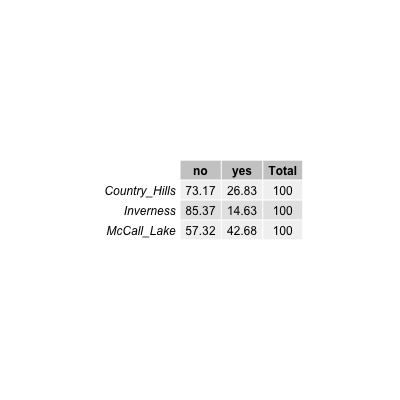


Figure 6: Frequency of occurrence of samples positive based on MST marker HF183 for human fecal contamination broken down by sampling site in Calgary, Alberta.

knitr::include\_graphics("../../results/HumM2\_y\_n\_pond\_table.png")

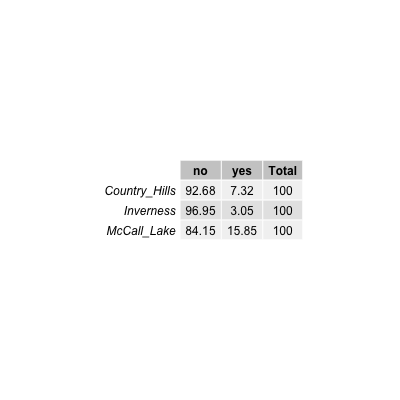


Figure 7: Frequency of occurrence of samples positive based on MST marker HuMm2 for human fecal contamination broken down by sampling site in Calgary, Alberta.

knitr::include\_graphics("../../results/CG01\_y\_n\_pond\_table.png")

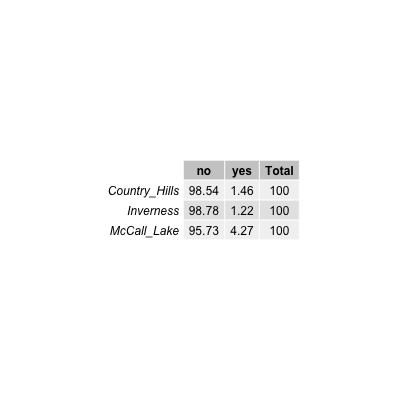


Figure 8: Frequency of occurrence of samples positive based on MST marker CG01 for human fecal contamination broken down by sampling site in Calgary, Alberta.

knitr::include\_graphics("../../results/LeeSg\_y\_n\_pond\_table.png")

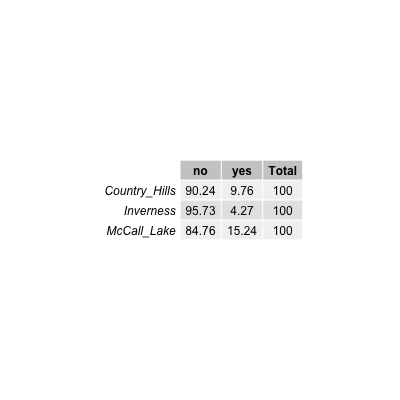


Figure 9: Frequency of occurrence of samples positive based on MST marker LeeSg for human fecal contamination broken down by sampling site in Calgary, Alberta.

knitr::include\_graphics("../../results/Dog3\_y\_n\_pond\_table.png")

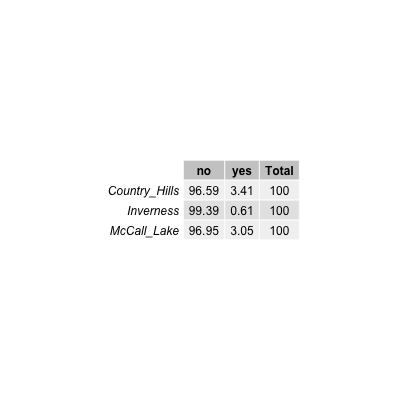


Figure 10: Frequency of occurrence of samples positive based on MST marker Dog3 for human fecal contamination broken down by sampling site in Calgary, Alberta.

knitr::include\_graphics("../../results/Rum2Bac\_y\_n\_pond\_table.png")

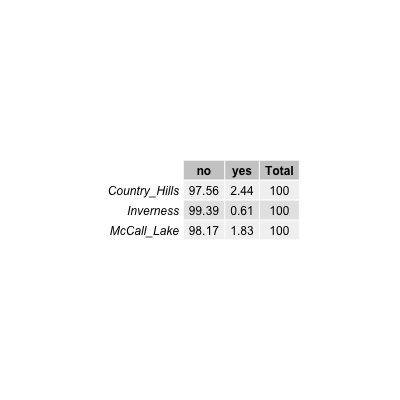


Figure 11: Frequency of occurrence of samples positive based on MST marker Rum2Bac for human fecal contamination broken down by sampling site in Calgary, Alberta.

knitr::include\_graphics("../../results/MuBac\_y\_n\_pond\_table.png")

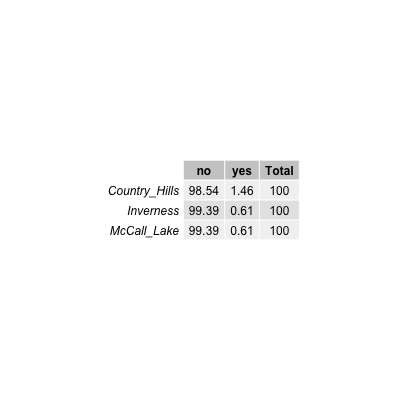


Figure 12: Frequency of occurrence of samples positive based on MST marker MuBac for human fecal contamination broken down by sampling site in Calgary, Alberta.

In congruence with finding that ML2 at McCall Lake was the most frequently contaminated site with human feces (*ABOVE TABLE*), this site also had the greatest median concentration of the human fecal marker HF183 (i.e., 4.2 log10 copies/100 mL) observed across all three stormwater ponds and sampling sites in these ponds (*BELOW FIGURE*). In comparison, all other McCall Lake sampling sites had a median concentration of HF183 at ~3.4 log10 copies/100 mL (i.e., close to the quantification limit of the assay (red dotted line)) (*BELOW FIGURE*). Specifically, at ML2, there was a single outlier in the data set for HF183, represented by a value of 6.0 log10 copies/100 mL (*BELOW FIGURE*). However, although ML2 represented the most consistently contaminated sampling site with human fecal contamination at McCall Lake, all other sites in this thesis appeared to be at risk for human fecal contamination (*BELOW FIGURE*).

knitr::include\_graphics("../../results/McCall\_bysite\_box\_HF183.png")

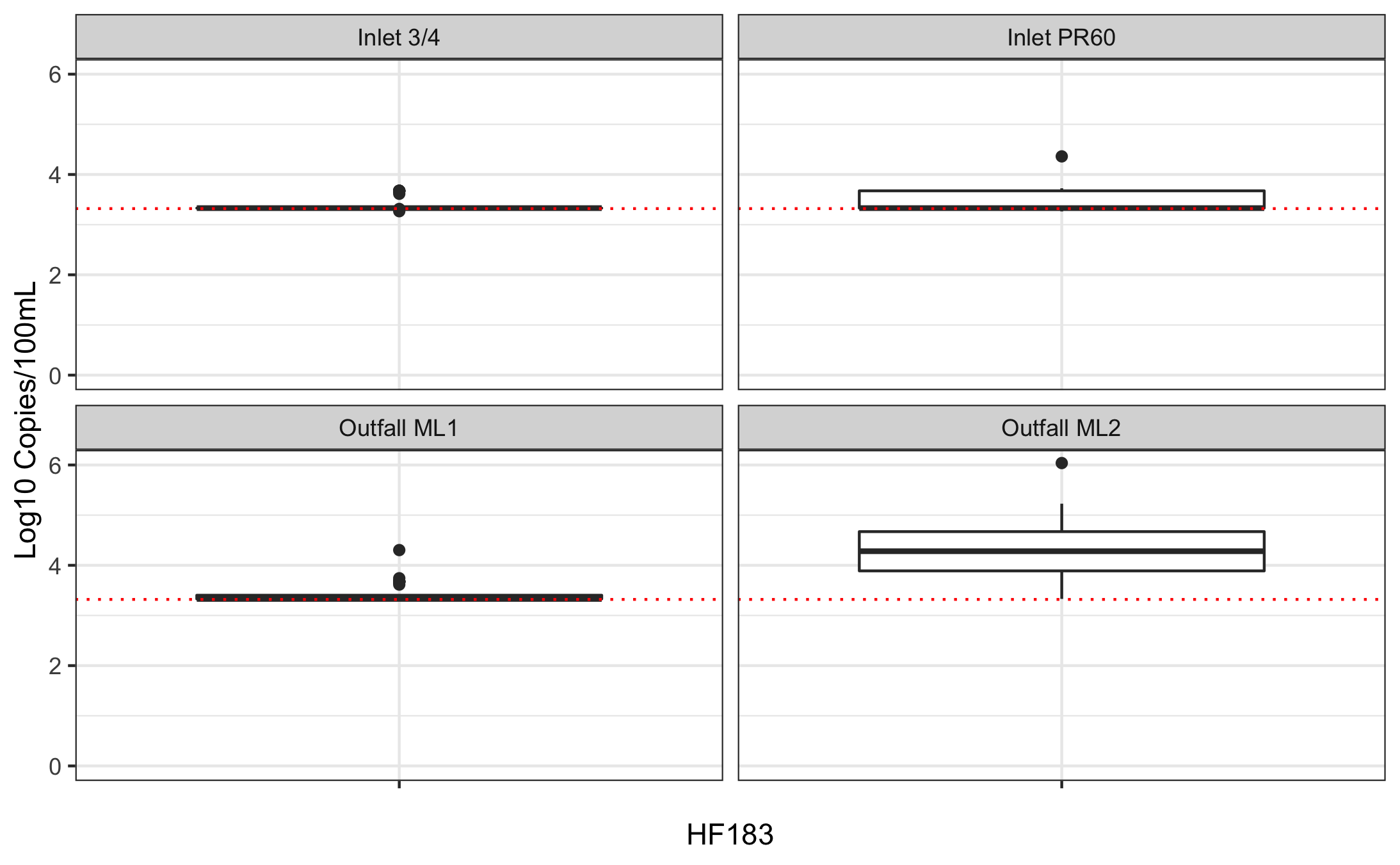


Figure 13: Box and Whisker Plot of HF183 levels by sampling site in McCall Lake (ML2 n=38, ML1 n=6, PR60 n=13, Inlet ¾ n= 5). The outer edges of the box represent the 25th and 75th percentiles (i.e., interquartile range), and the line within the box represents the median. The location of median indicates the skew of the data. The whiskers represent the interquartile range\*1.5. The outliers are determined by being greater or less than 1.5 times the upper of lower interquartile ranges as represented by circles.

knitr::include\_graphics("../../results/Country\_Hills\_bysite\_box\_HF183.png")

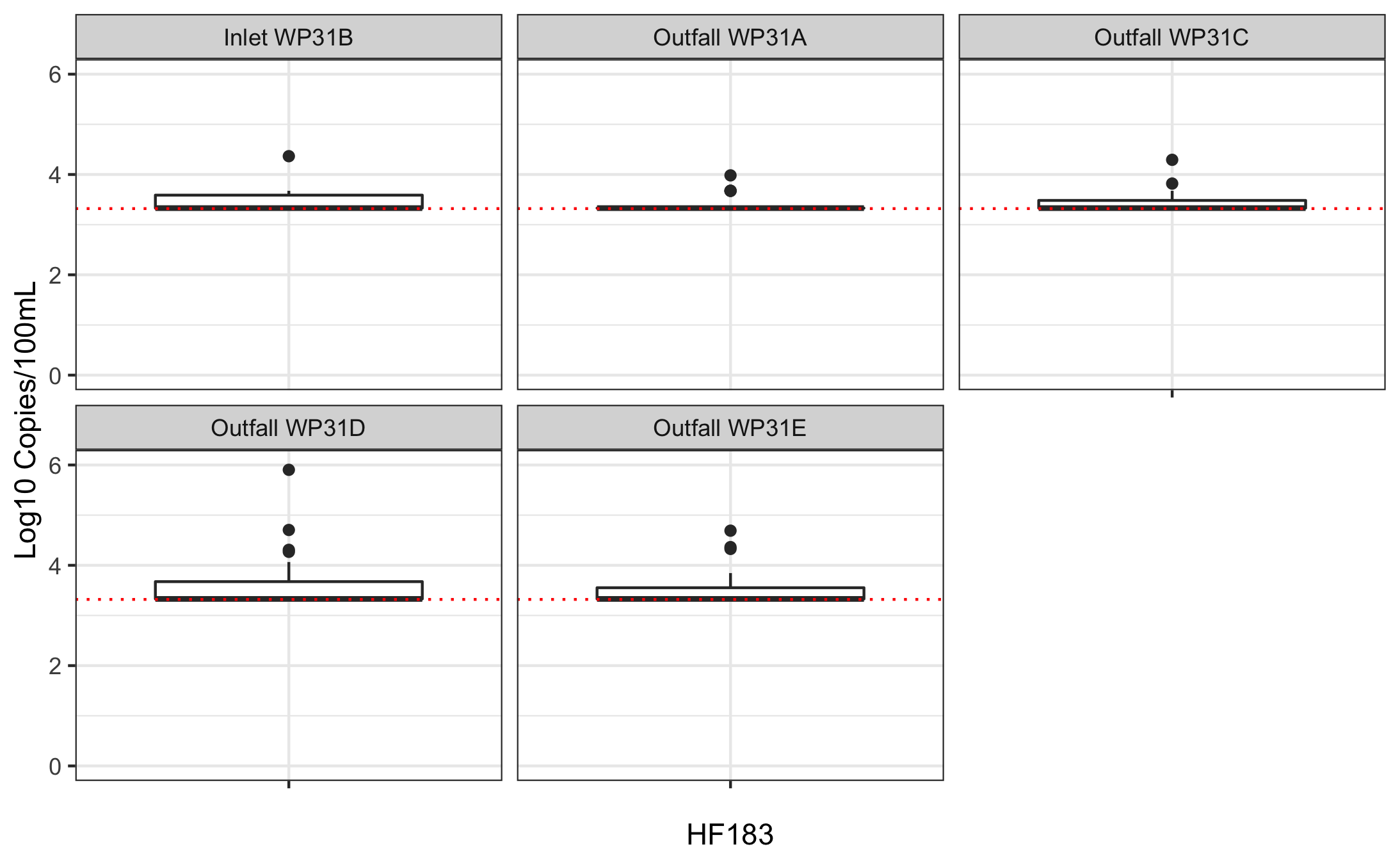


Figure 14: Box and Whisker Plot of HF183 levels by sampling site in Country Hills (ML2 n=38, ML1 n=6, PR60 n=13, Inlet ¾ n= 5). The outer edges of the box represent the 25th and 75th percentiles (i.e., interquartile range), and the line within the box represents the median. The location of median indicates the skew of the data. The whiskers represent the interquartile range\*1.5. The outliers are determined by being greater or less than 1.5 times the upper of lower interquartile ranges as represented by circles.

knitr::include\_graphics("../../results/Inverness\_bysite\_box\_HF183.png")

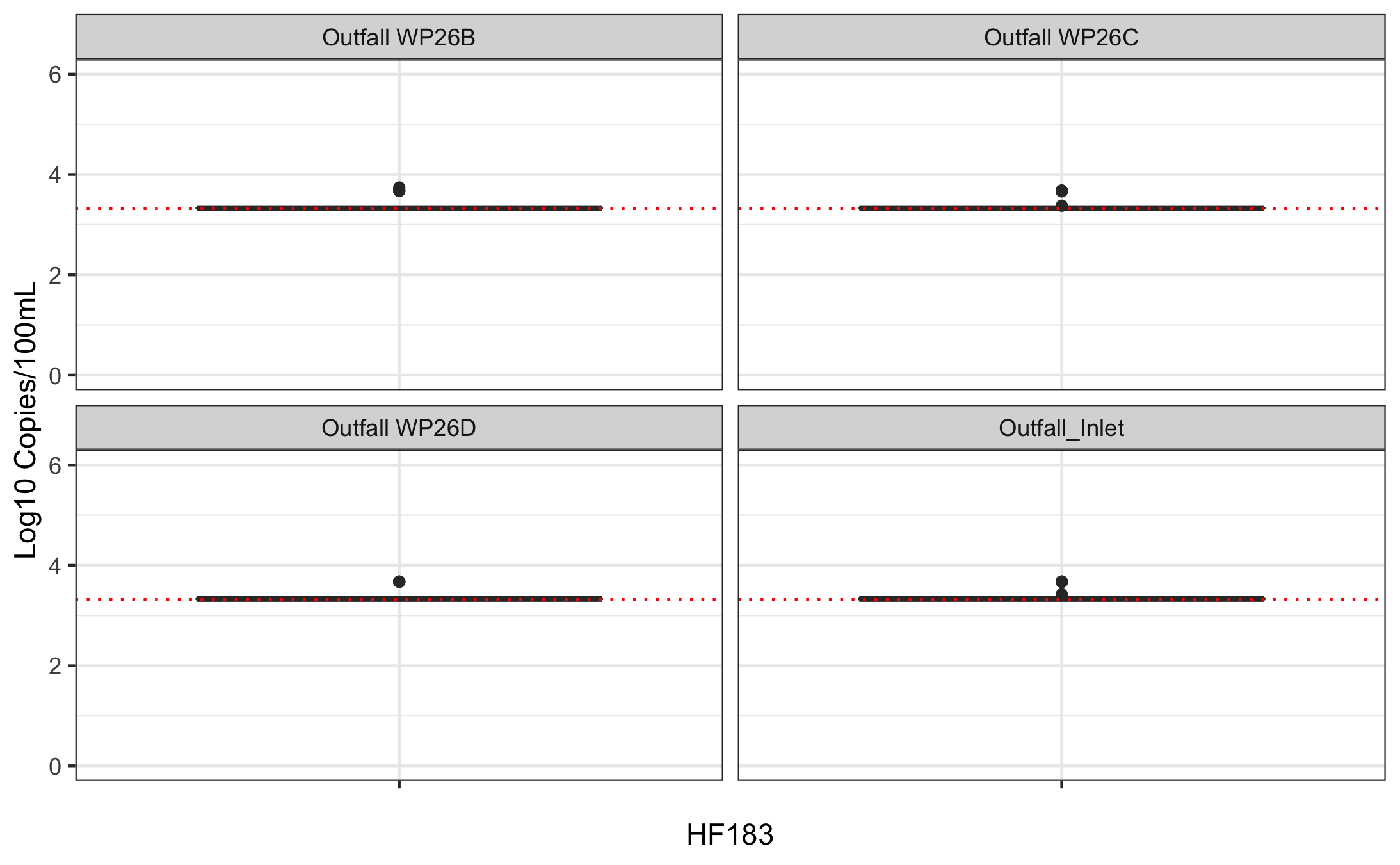


Figure 15: Box and Whisker Plot of HF183 levels by sampling site in Inverness (ML2 n=38, ML1 n=6, PR60 n=13, Inlet ¾ n= 5). The outer edges of the box represent the 25th and 75th percentiles (i.e., interquartile range), and the line within the box represents the median. The location of median indicates the skew of the data. The whiskers represent the interquartile range\*1.5. The outliers are determined by being greater or less than 1.5 times the upper of lower interquartile ranges as represented by circles.

Temporal fluctuations in human fecal pollution markers were noted between the stormwater ponds, and among the sampling sites within a stormwater pond (*Figure below*).

knitr::include\_graphics("../../results/HF183\_all\_arranged.png")

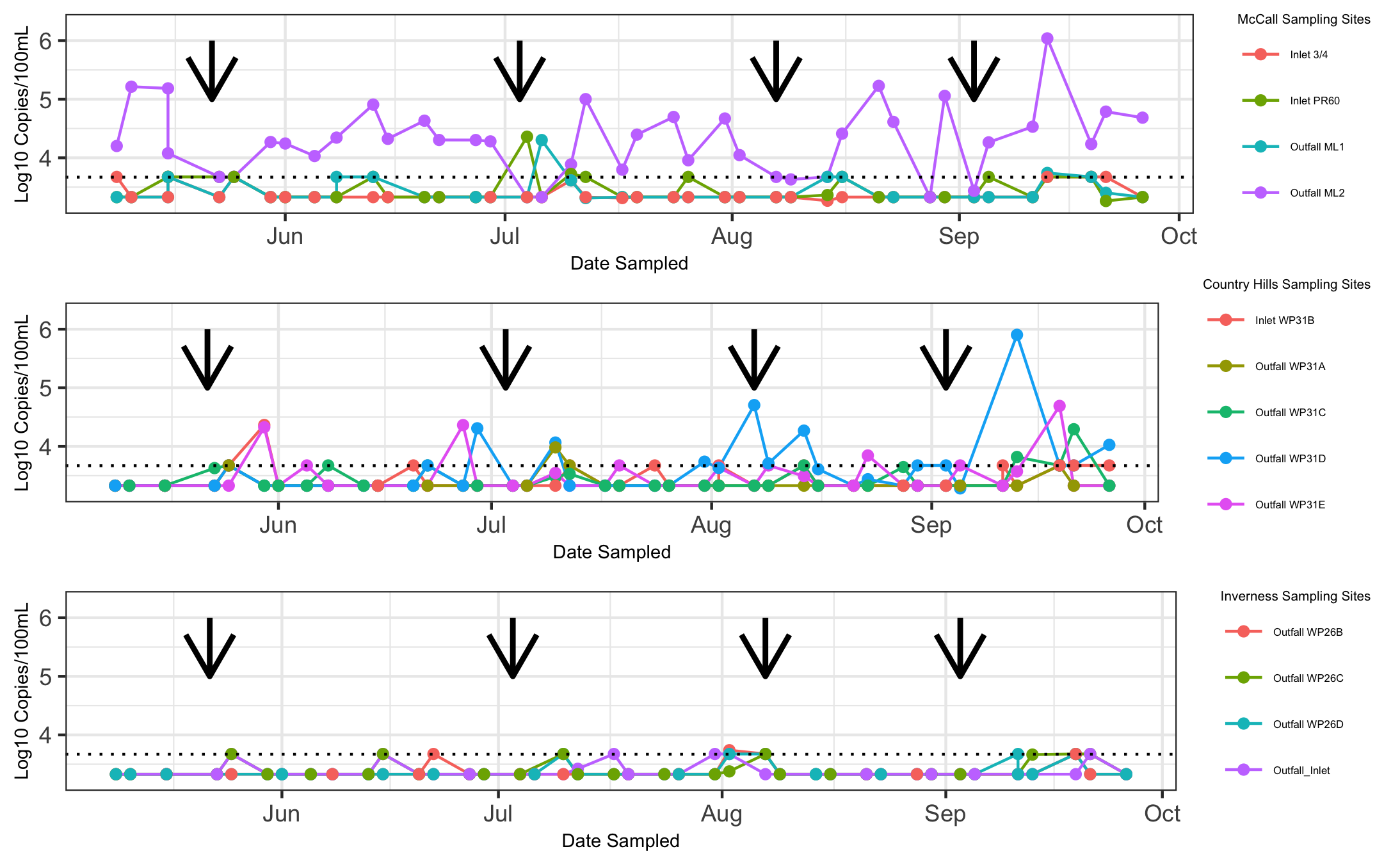


Figure 16: Temporal pattern of occurrence HF183 log10 concentrations at all sampling sites in McCall Lake (top), Country Hills (middle), and Inverness (botton) over the 21-week sampling season. The limit of quantification95 (LOQ95) as a black dotted line. The black arrows represent long holiday weekends.

Of all stormwater pond sampling sites, ML2 at McCall Lake experienced the most consistent temporal pattern of human fecal contamination throughout the sampling season. For example, within the 41 sampling dates, over the 21-week sampling season, there were only two sampling dates in which we did not detect HF183 at ML2 (i.e., July 4th and August 28th) (*FIGURE ABOVE*). However, there were other sampling dates when levels of HF183 decreased to a non-quantifiable level at ML2 (i.e., May 23rd, May 25th, August 8th, and August 14th). Interestingly, this pattern tended to occur after long weekends (i.e., holidays occurring on the following Mondays: May 22nd, July 3rd, August 7th, and September 3rd), and three of these long weekends corresponded to decreases in human fecal contamination markers on the following day of sampling (i.e., May 23rd, July 4th, and August 8th, which were Tuesdays). This suspicious temporal pattern of contamination suggested that the levels of human fecal contamination may have been related to industrial/commercial activities, as the levels of human fecal contamination decreased during times when industries/commercial premises may have been closed for the holidays.

Human fecal contamination at the sampling sites was often highly variable between sequential sampling dates. For example, at Inlet PR60 in McCall Lake, within a two-week span, biweekly HF183 values fluctuated between undetectable levels (i.e., June 29th and July 6th) and 4.3 log10 copies/100mL (i.e., July 4th) and 3.5 log10 copies/100mL (i.e., July 10th). This high variability in human fecal contamination markers over sequential sampling dates, elicits potential concerns regarding the sporadic nature of contamination and the stability of water quality in the urban stormwater ponds.

knitr::include\_graphics("../../results/HumM2\_at\_McCall.png")

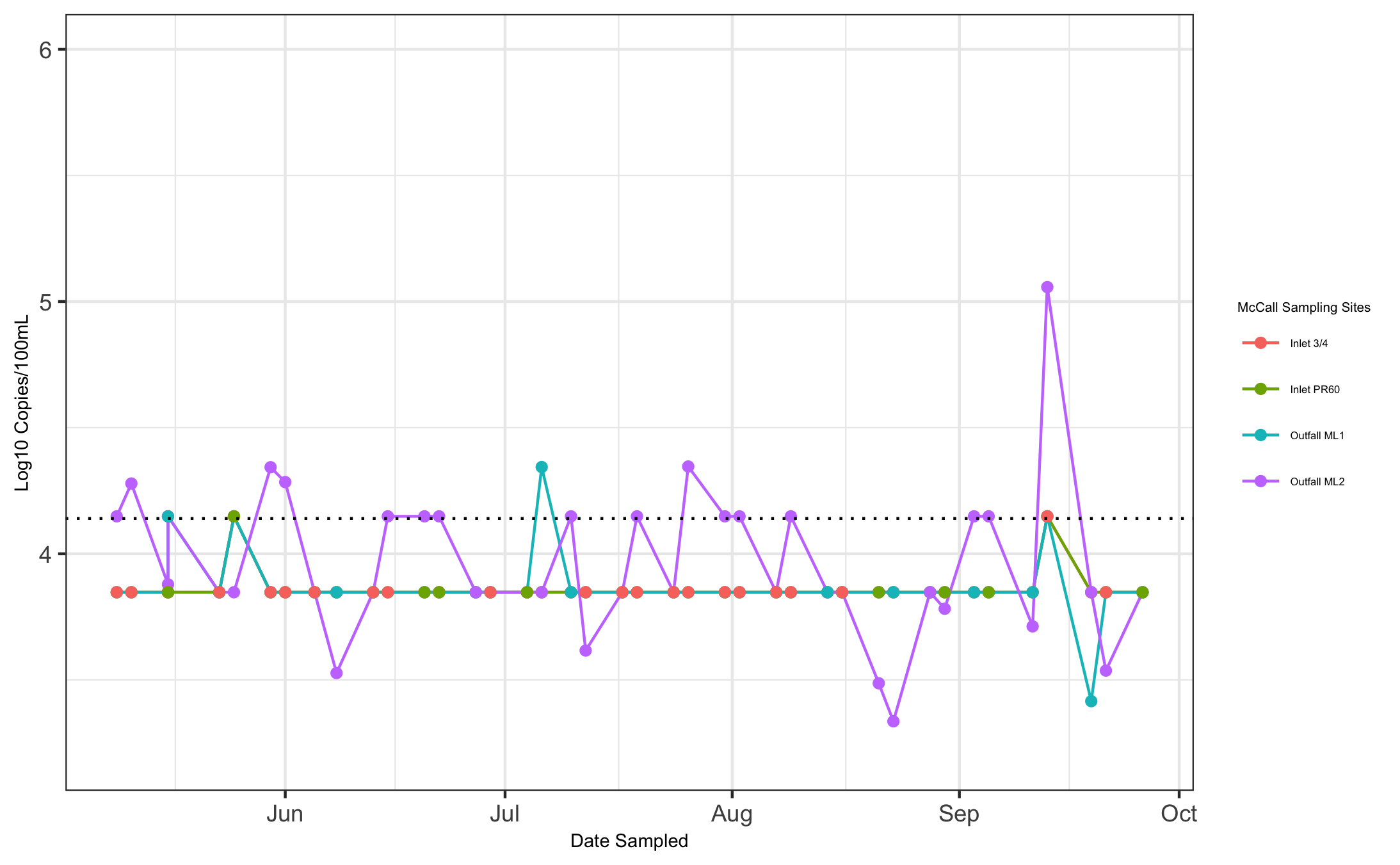


Figure 17: Temporal pattern of occurrence HuMm2 log10 concentrations at all sampling sites in McCall Lake over the 21-week sampling season. The limit of quantification95 (LOQ95) as a black dotted lines.

HumM2 was detected less frequently and at lower concentrations in all of the urban stormwater ponds tested. ML2 had the highest occurrence of HumM2 detections of all McCall Lake sampling sites, which corresponded with the findings with the human fecal contamination marker HF183 (*FIGURE ABOVE*). Furthermore, Inlet ¾ had the lowest occurrence of HF183 in McCall Lake, and was also tied for the lowest occurrence of HumM2 in McCall Lake.

In addition to seeing temporal and spatial trends in human fecal contamination, temporal and spatial trends in seagull contamination were also noted. In McCall Lake, seagull fecal contamination was considered to be a sporadic, highly variable, source of pollution. Seagull contamination was first noted in McCall Lake at the end of June, and tended to be episodic (*FIGURE BELOW*). For example, at ML2, seagull fecal contamination was detected on July 12th and at a level of 4.1 log10 copies/100 mL, and then it was not detected at quantifiable levels again until August 8th (i.e., 3.7 log10 copies/100 mL) (*FIGURE BELOW*). Although ML2 was most frequently positive for detection of seagull fecal contamination among all sampling sites, this pattern of sporadic, highly variable findings were also noted at the other McCall Lake sites (i.e., PR60, ML1, and Inlet ¾).

knitr::include\_graphics("../../results/LeeSg\_at\_McCall.png")

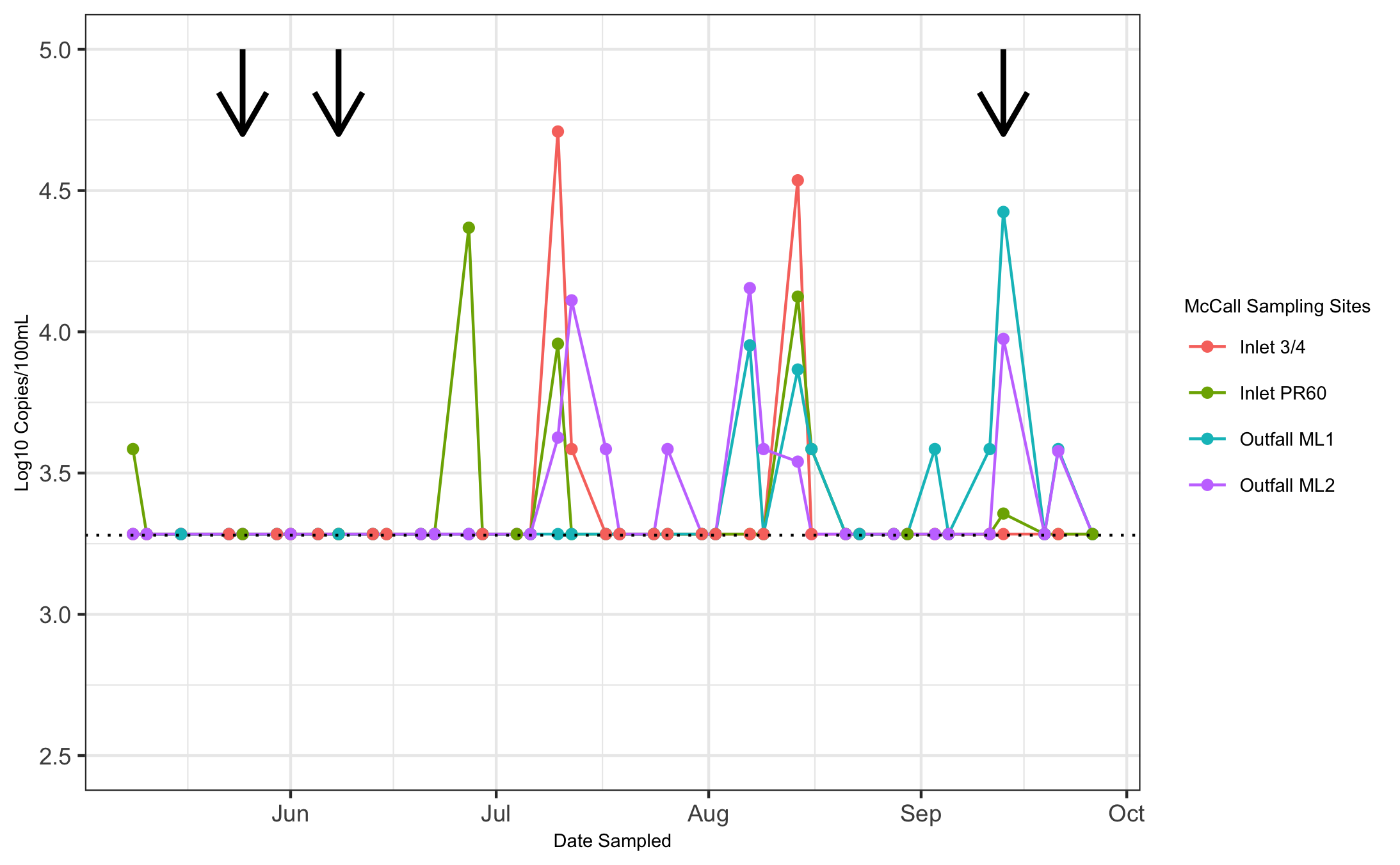


Figure 18: Temporal pattern of occurrence of LeeSg concentrations at all sampling sites in McCall Lake over the 21-week sampling season. The limit of quantification95 (LOQ95) as a blue dotted line. The black arrows represent greater than 10 mm of rain in the previous 72 hours.

Overall, there were patterns of similarity in seagull contamination regarding temporal trends. These patterns were similar across the sampling sites in McCall Lake, which was noted for two key reasons. Firstly, there were three instances where seagull fecal contamination occurred concurrently at three or more McCall Lake sites (i.e., July 10th, August 14th, and September 13th). Secondly, on the aforementioned sampling dates, the levels of seagull fecal contamination detected were all within one order of magnitude of each other. These patterns suggested that a potential environmental variable linked the contamination along the sampling sites at McCall Lake. One potential environmental variable examined was antecedent rainfall. Only three dates (i.e., May 25th, June 8th, and September 13th) had greater than 10 mm of rain. Seagull fecal contamination was detected on only one of the sampling dates (September 13th), though at three sampling sites on this date.

A high-level descriptive overview of the frequency of several enteric bacterial pathogens (i.e., A. butzleri, Campylobacter spp., Salmonella spp., and STEC) in each of the Calgary urban stormwater ponds (i.e., McCall Lake, Country Hills Stormwater Facility, and Inverness Stormpond), and at each sampling site within the ponds is provided in *—–*. The most frequently detected bacterial pathogen found in stormwater ponds was A. butzleri, detected in 36% of samples at McCall Lake, 24% of samples at Inverness and 18% of samples at Country Hills (*—–*). The second most common pathogen detected was Campylobacter spp. (4% at Country Hills) and Salmonella spp. (1% at McCall Lake) (*—–*).

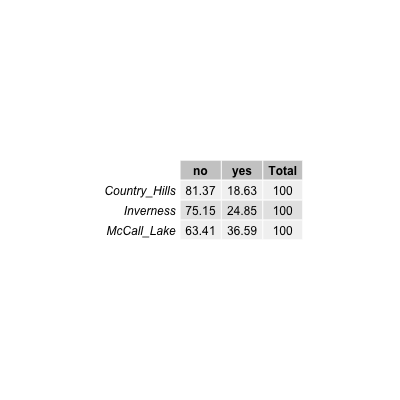


Figure 19: The frequency of occurence of the enteric bacterial pathogen A. butzleri in the Calgary, Alberta stormwater ponds.

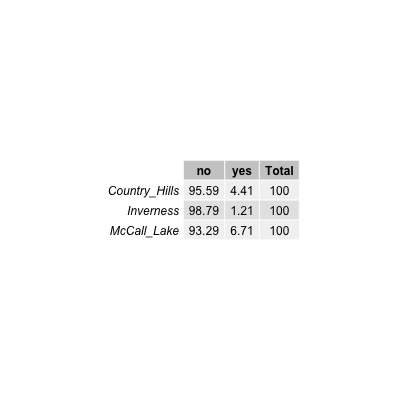


Figure 20: The frequency of occurence of the enteric bacterial pathogen Campylobacter spp. in the Calgary, Alberta stormwater ponds.

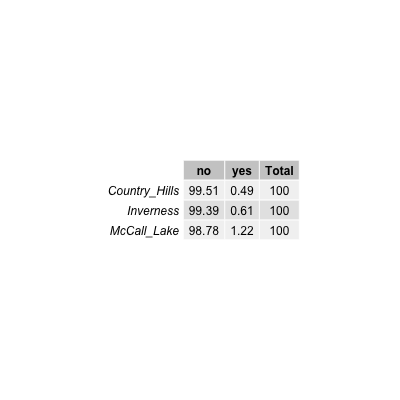
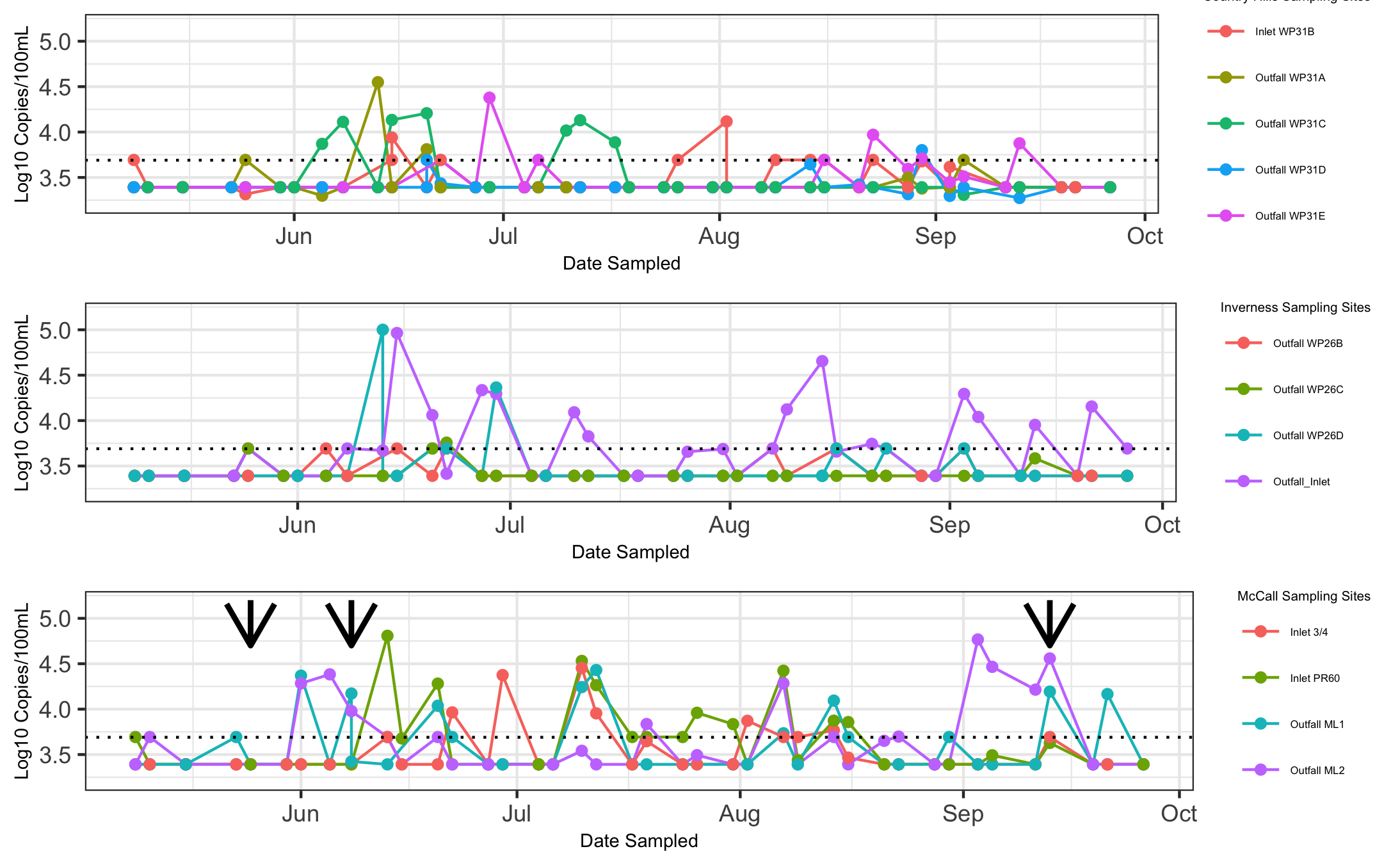


Figure 21: The frequency of occurence of the enteric bacterial pathogen A. butzleri in the Calgary, Alberta stormwater ponds.

To better understand temporal variation, we further examined patterns of occurrence based on molecular qPCR results. Notable temporal fluctuations in A. butzleri were observed between the urban stormwater ponds, and among the sampling sites within a pond (*BELOW*). We found that at Inlet ¾, in McCall Lake, considerable temporal fluctuations were detected in the levels of A. butzleri between sequential sampling dates. Within a two-week time period (i.e., four sequential sampling dates, June 20th – June 29th), the concentration of A. butzleri varied from being not detected (i.e., below the limit of quantification of 3.5 log10 copies/100 mL) on June 20th, then spiking to 3.9 log10 copies/100 mL on June 22nd, to be not detected on June 27th, and spiking again to 4.3 log10 copies/100 mL on June 29th.

We tracked environmental variables that could contribute to temporal fluctuations in A. butzleri concentrations (e.g., antecedent rainfall data, temperature, etc.). Of note, we recorded three sampling dates that had rainfall greater than 10 mm (i.e., May 25th, June 8th, and September 13th, Figure 5 1). We noted that A. butzleri was detected at all McCall Lake sampling sites on several sampling dates, July 10th, August 14th, August 16th, September 13th, of which September 13th had significant rainfall (Figure 5 1). However, on another rainfall date (i.e., May 25th) A. butzleri was not observed at any of the sampling sites, and on June 8th, A. butzleri concentrations reached detectable levels only at the outfalls (i.e., ML1 and ML2). `

knitr::include\_graphics("../../results/Arco\_all\_arranged.png")



Due to the prevalence and abundance of A. butzleri contamination, we sought to determine the potential sources of its contamination. Water samples were analyzed by identifying which microbial source tracking markers occurred most often with A. butzleri detections. We found that the most common source of pollution co-occurring with A. butzleri detection was human fecal pollution. The human marker HF183 was present in 43% of A. butzleri positive samples, while the human marker HumM2 was detected in 10% of A. butzleri positive samples (Table 5 6). The second most dominant source of fecal pollution was seagull (i.e., LeeSg), which corresponded to A. butzleri detection in 10% of A. butzleri stormwater positive samples. The only other markers found in conjunction with A. butzleri were for Canada geese (i.e., CGO1) and ruminants (i.e., Rum2Bac), which were detected in 2% and 1% of positive samples, respectively.

#co occurange figure

To further our understanding of contamination by stormwater pond, a random forest analysis was performed. Our outcome variable was the pond, and our predictors were all other variables excluding sampling site. This tree tells us if HF183 is greater than 3.9 log10 copies/100mL than the pond is McCall Lake. Based on our earlier analysis of how McCall Lake is the most conaminted sampling location by HF183, this is accurate. If HF183 is less than 3.9 log10 copies/100mL and E. coli is greater than or equal 0.87 CFU/100 mls than the pond is also contaminted with A. butzleri, otherwise the next predictor is the sampling date. If the sample has greater than 3.6 log10 copies/100mL of A. butzleri and Enterococcus levels higher than 2.5 log10 CCE/100ml than the sampling location is Country Hills. This is also reflective of the aforementioned data, as Country Hills is more contaminated than Inverness but less contaminated than McCall Lake. In addition, if A. butzleri is less than 3.6 log10 copies/100mL, than it may be contamined with seagull fecal material. This is also reflective of the aforementioned data, as we found A. butzleri to have a higher co-occurance with human fecal markers (i.e., HF183 and HumM2) than bird fecal markers (i.e., LeeSg or CG01). Following this point, our tree is further broken down by all the water quality indicators (i.e., Enterococcus, E. coli, thermotolerant coliforms). Moving back towards the top of tree we note that the third decision point is based off of date sampled, and then thermotolerant coliforms. All of these decision points lead to Inverness. This is reflective of the previous data presented as Inverness was the least contaminated of all sampling locations. Furthermore, we find that the following decision point is total coliforms. This is unsurprising, as total coliforms are found to be abundant in recreational waters and are no longer considered to be preditors of gastroitenstinal illness when assessing water quality. As such, the US EPA has moved away from using total coliforms. Additionally, it should be noted that this half of the tree soley relies on water quality indicators and does not create nodes based off microbial source tracking markers or pathogen specific qPCR markers. This is further reflective of the results that were previously presented, as Inverness has the best water quality of all stormwater ponds tested and did not consistently have high levels of any microbial source tracking marker tested or pathogen specific qPCR marker.

knitr::include\_graphics("../../results/pond\_tree.png")

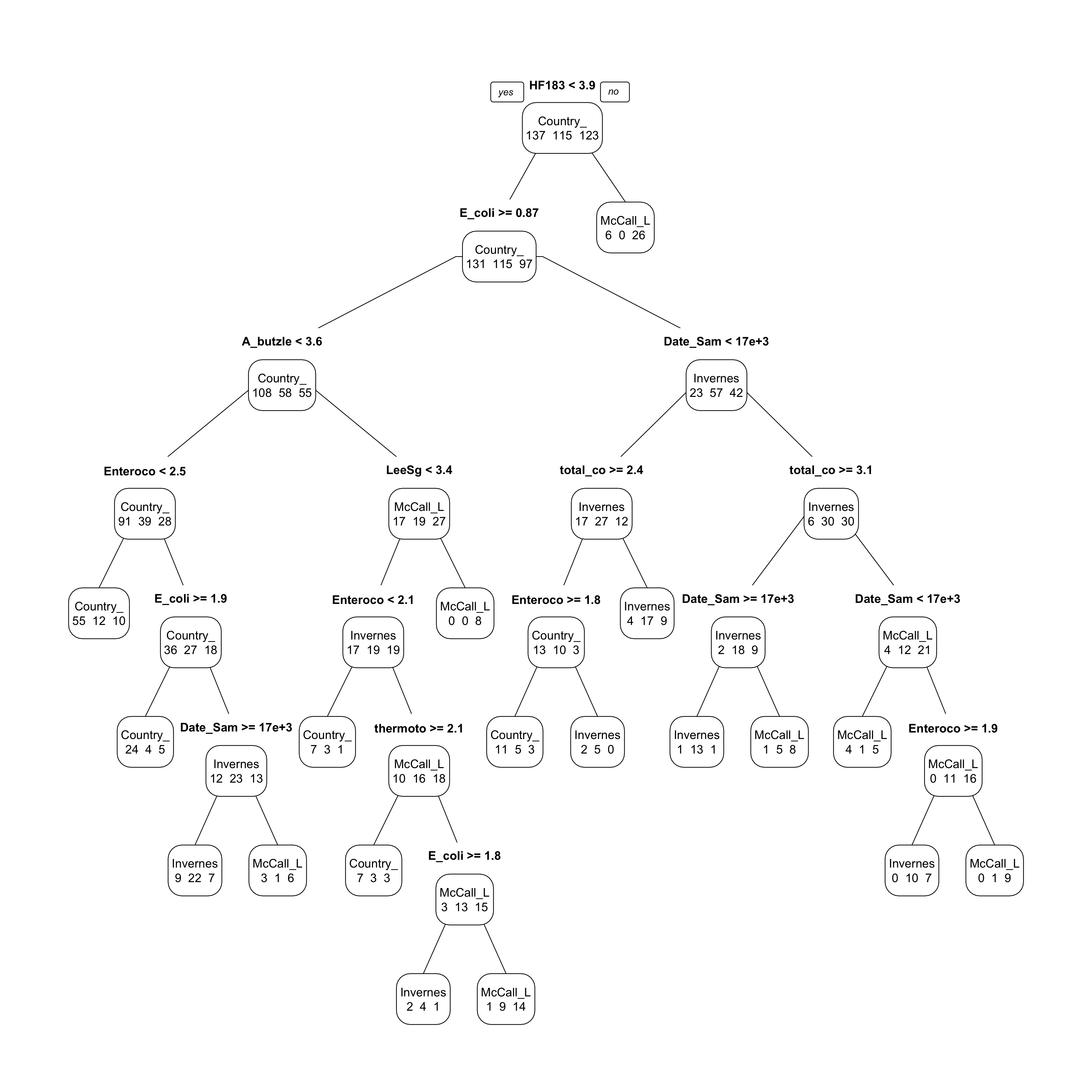


Figure 22: Random forest based on the outcome of pond (i.e., McCall Lake, Inverness, Country Hills.)

#Univariate Analysis *Use a combination of text/tables/figures to explore and describe your data. You should produce plots or tables or other summary quantities for most of your variables. You definitely need to do it for the important variables, i.e. if you have main exposure or outcome variables, those need to be explored. Depending on the total number of variables in your dataset, explore all or some of the others.*

## Bivariate analysis

*Create plots or tables and compute simple statistics (e.g. t-tests, simple regression model with 1 predictor, etc.) to look for associations between your outcome(s) and each individual predictor variable*

## Full analysis

*Use one or several suitable statistical/machine learning methods to analyze your data and to produce meaningful figures, tables, etc. This might again be code that is best placed in one or several separate R scripts that need to be well documented. You can then load the results produced by this code*

# Discussion

## Summary and Interpretation

*Summarize what you did, what you found and what it means.*

## Strengths and Limitations

*Discuss what you perceive as strengths and limitations of your analysis.*

## Conclusions

*What are the main take-home messages?*

*Include citations in your Rmd file using bibtex, the list of references will automatically be placed at the end*

# References

Adams, B., N. (2016). Shiga toxin producing escherichia coli o157, england and wales, 1983-2012. *Emerging Infectious Diseases*, *22*, 590–597.

Craun, C., G.F. (2005). Outbreaks associated with recreational water in the united states. *International Journal of Environmental Health Research*, 243–262.

Douidah, Z., L. (2011). Occurence of putative virulence genes in arcobacter species isolated from humans and animals. *Journal of Clinical Microbiology*, *3*, 735–741.

Green, H., H.C. (2014). Improved hf183 quantitative real-time pcr assay for characterization of human fecal pollution in ambient surface water samples. *Applied and Environmental Microbiology*, *80*, 3086–3094.

Kayman, A., T. (2012). Emerging pathogen arcobacter spp. In acute gastroenteritis: Molecular identification, antibiotic susceptibilities and genotyping of the isolated arcobacter. *Journal of Medical Microbiology*, *61*, 1439–1444.

Krometis, C., L. (2010). Comparison of the presence and partitioning behavior of indicator organisms and salmonella spp. In an urban watershed. *Water Research*, *8*, 44–59.

Levican, A., A. (2013). Adherence to and invasion of human intestinal cells by arcobacter species and their virulence genotypes. *Applied Environmental Microbiology*, *79*(16), 4951–4957.

McCarthy, &. G., N. (2001). Incidence of guillain-barré syndrome following infection with campylobacter jejuni. *American Journal of Epidemiology*, *153*, 610–614.

Meng, C., Z. (2018). Stormwater constructed wetlands: A source or sink for campylobacter spp. *Water Research*, *131*, 218–227.

Moore, C., J. (2001). Occurrence of campylobacter spp. In water in northern ireland: Implications for public health. *International Journal of Environmental Health Research*, *70*, 102–207.

Rangel, S., J. (2005). Epidemiology of escherichia coli o157:H7 outbreaks, united states, 1982-2002. *Emerging Infectious Diseases*, *11*, 603–609.

Soller, S., J. (2017). Incidence of gastrointestinal illness following wet weather recreational exposure: Harmonization of quantitative microbial risk assessment with an epidemiological investigation of surfers. *Water Research*, *121*, 280–289.

Staley, C., Z. R. (2018). Fecal source tracking and eDNA profiling in an urban creek following an extreme rain event. *Scientific Reports*, *8*, 14390.

Vandenberg, D., O. (2004). Arcobacter butzleri infections report diarrhea associated with abdominal pain. *Water Research*, *109*, 1053–1066.

Van Driessche, E., & Houf, K. (2005). Survival capacity in water of arcobacter species under different temperature conditions. *Vetinary Microbiology*, *105*, 149–154.

Wesley, W., I. (2000). Fecal shedding of campylobacter and arcobacter spp. In dairy cattle. *Applied and Enviornmental Microbiology*, *66*, 4241–4246.