Bacterial Sources of Pollution in Urban Stormwater Impacted Bodies of Water

Megan Beaudry

2019-09-19

# Summary/Abstract

*Write a summary of your project.*

# Illustrating setup

*This section is only there to show how to insert results from other places in the project and how to cite figures and other references. Delete this whole section at some point.*

This paper (Leek & Peng, 2015) discusses types of analyses.

Figure 1 shows a result figure from the analysis.

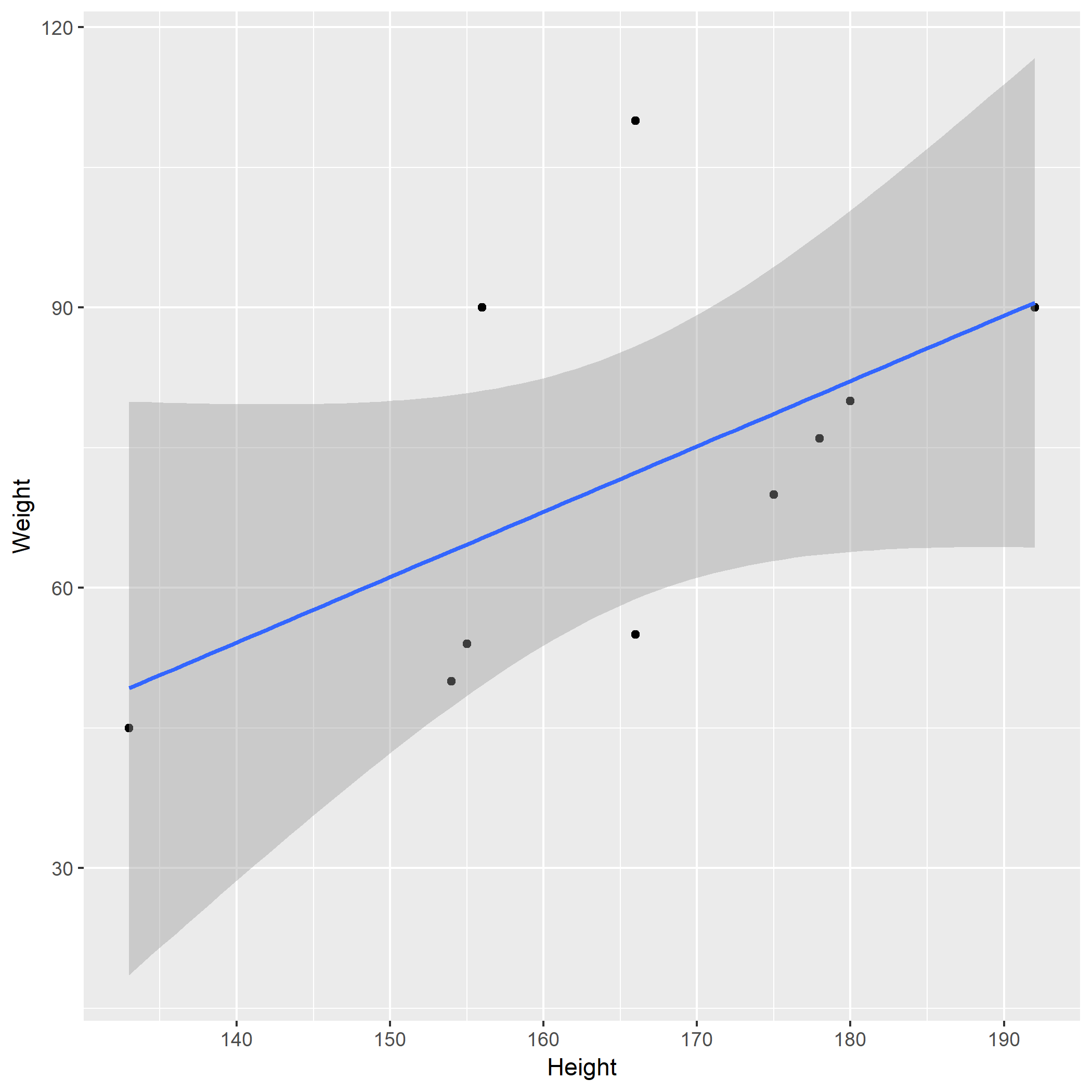


Figure 1: Analysis figure.

Table 1 shows a result table from the analysis.

Table 1: Result Table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| term | estimate | std.error | statistic | p.value |
| (Intercept) | -43.7883068 | 61.1150617 | -0.7164896 | 0.4940713 |
| Height | 0.6996272 | 0.3675692 | 1.9033889 | 0.0934786 |

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 the concentrations of A. butzleri decrease in colder weather? 5. Do we see a relationship between traditional water quality indicators (E. coli and Enterococcus) and A. butzleri? 6. Are certain sampling locations or sites more prone to A. butlzeri contamination? 7. What is the relationship between fecal contamination, water quality indicators and A. butlzeri? 8. 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': 561 obs. of 25 variables:  
## $ X : Factor w/ 55 levels "","10-Jul","12-Jul",..: 1 1 28 1 46 1 1 1 1 1 ...  
## $ X.1 : Factor w/ 22 levels "","Special investigation",..: 1 1 4 5 1 1 1 1 1 1 ...  
## $ X.2 : Factor w/ 20 levels "","Country Hillls",..: 1 1 20 19 5 11 5 11 5 17 ...  
## $ X.3 : Factor w/ 83 levels "","Dilution",..: 1 1 14 20 61 66 80 75 72 71 ...  
## $ X.4 : Factor w/ 552 levels "","13957","13958",..: 1 1 552 548 4 5 6 7 8 9 ...  
## $ X.5 : Factor w/ 3 levels "","10 mm","Rainfall": 1 1 3 1 1 1 1 1 1 1 ...  
## $ X.6 : Factor w/ 86 levels "","100470","10614",..: 85 84 82 86 86 86 86 86 86 86 ...  
## $ X.7 : Factor w/ 24 levels "","114105.6",..: 1 23 1 24 24 24 24 24 24 24 ...  
## $ X.8 : Factor w/ 11 levels "","11208","19989.6",..: 8 9 1 11 11 11 11 11 11 11 ...  
## $ X.9 : Factor w/ 26 levels "","10440","12178.2",..: 1 25 1 26 26 26 26 26 26 24 ...  
## $ X.10 : Factor w/ 11 levels "","2058","22119",..: 9 10 1 11 11 11 11 11 11 8 ...  
## $ X.11 : Factor w/ 12 levels "","#VALUE!","23046.6",..: 11 10 1 12 12 12 12 12 12 12 ...  
## $ X.12 : Factor w/ 8 levels "","170052","1800",..: 8 7 1 6 6 6 6 6 6 3 ...  
## $ X.13 : Factor w/ 104 levels "","100120.2",..: 102 101 1 104 104 104 104 104 103 103 ...  
## $ X.14 : Factor w/ 5 levels "","DNQ","InvA",..: 5 3 1 4 4 4 4 4 4 4 ...  
## $ X.15 : Factor w/ 10 levels "","19138.2","23373",..: 7 10 1 9 9 9 9 9 9 9 ...  
## $ X.16 : Factor w/ 404 levels "","0","10087",..: 399 397 398 402 54 30 150 111 40 271 ...  
## $ X.17 : Factor w/ 24 levels "","0","no","No",..: 9 8 1 1 4 5 4 5 5 5 ...  
## $ X.18 : Factor w/ 36 levels "","0","No","not tested",..: 10 9 1 1 3 5 3 5 5 5 ...  
## $ X.19 : Factor w/ 4 levels "","Culture Arco",..: 2 1 1 1 1 1 1 1 1 1 ...  
## $ X.20 : Factor w/ 6 levels "","both","no",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ QT.2000.MPN : Factor w/ 95 levels "","1011","1046",..: 94 95 1 1 60 74 53 46 42 13 ...  
## $ X.21 : Factor w/ 159 levels "","<1","1","10",..: 155 154 1 1 97 159 62 159 159 159 ...  
## $ QT.51.WELL.MPN: Factor w/ 67 levels "","<1","<10",..: 64 1 1 1 67 67 7 21 21 67 ...  
## $ X.22 : Factor w/ 70 levels ""," ","<10",">10 \*\*",..: 8 70 65 1 66 66 7 66 66 66 ...

# Methods and Results

*In most research papers, results and methods are separate. You can combine them here if you find it easier. You are also welcome to structure things such that those are separate sections.*

## Data aquisition

*As applicable, explain where and how you got the data. If you directly import the data from an online source, you can combine this section with the next.*

## Data import and cleaning

*Write code that reads in the file and cleans it so it’s ready for analysis. Since this will be fairly long code for most datasets, it might be a good idea to have it in one or several R scripts. If that is the case, explain here briefly what each file does. The files themselves should be commented well so everyone can follow along.*

## 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

Leek, J. T., & Peng, R. D. (2015). Statistics. What is the question? *Science (New York, N.Y.)*, *347*(6228), 1314–1315. <https://doi.org/10.1126/science.aaa6146>