EBIO 3080 Lab Week 6

Kayden Adams, Yousef Al Obaidan, Dylan Oh

1 October 2020

Data Collection: Counting Bacteria

Hypothesis: The most antibiotic-resistant bacteria will be found in site 6 and the least in site 1, with other sites falling within this pattern, because the more urban areas that are upstream of a location, the more antibiotic resistant bacteria there will be due to run-off into water sources and accumulation downstream.

Classification of bacterial colonies in antibiotic-positive and -negative environments

Observed for six sites along the Boulder Creek watershed

		Total Number of Colonies			Number of ... Colonies in Antibiotic-Negative Plate			
SITE	**DISTANCE FROM SITE 1 (KM)**	ANTIBIOTIC-NEGATIVE PLATE	ANTIBIOTIC-POSITIVE PLATE	RELATIVE FREQUENCY OF ANTIBIOTIC RESISTANCE	WHITE	LIGHT YELLOW	DARK YELLOW	PINK
1	0	13	3	0.231	1	4	8	0
2	11	61	33	0.541	12	29	19	1
3	23	78	19	0.244	9	26	41	2
4	35	42	17	0.405	8	20	8	6
5	49	66	16	0.242	0	39	21	6
6	69	95	9	0.095	17	24	52	2

Relative Frequency of Antibiotic Resistance

This section will look at the relative frequency of antibiotic resistance (FABR) of bacterial plate colonies grown from bacteria collected at 6 sites across the Boulder Creek watershed in order to determine how the movement of potential antibiotics downstream affects the number of antibiotic-resistant bacteria that exist at various sites along the river.

Note: The plot shown below was produced using ggplot2; however, the code shown is for a general R plot. If the ggplot2 plots are not suitable for the rubric, the R plot can be forwarded to whoever is grading, with the proof that it was done before the due date shown in the code.

```
# Importing csv file
bacteria_data <- read.csv("sites.csv", header = TRUE, sep = ",",
                          fileEncoding = "UTF-8-BOM")

# FABR was calculated in the csv file, but the calculation would have been:
# bacteria_data$FABR <- bacteria_data$Total_AB_Pos/bacteria_data$Total_AB_Neg

# Calculating linear model parameters for FABR vs distance
linear_model <- lm(FABR ~ Distance, data = bacteria_data)

# Plotting
plot(bacteria_data$Distance, bacteria_data$FABR,
     xlab = "Distance from Site 1 (km)",
     ylab = "Relative Frequency of Antibiotic Resistance",
     main = "Relative Frequency of Antibiotic Resistance at Different Sites")
abline(linear_model, col = "blue")
```

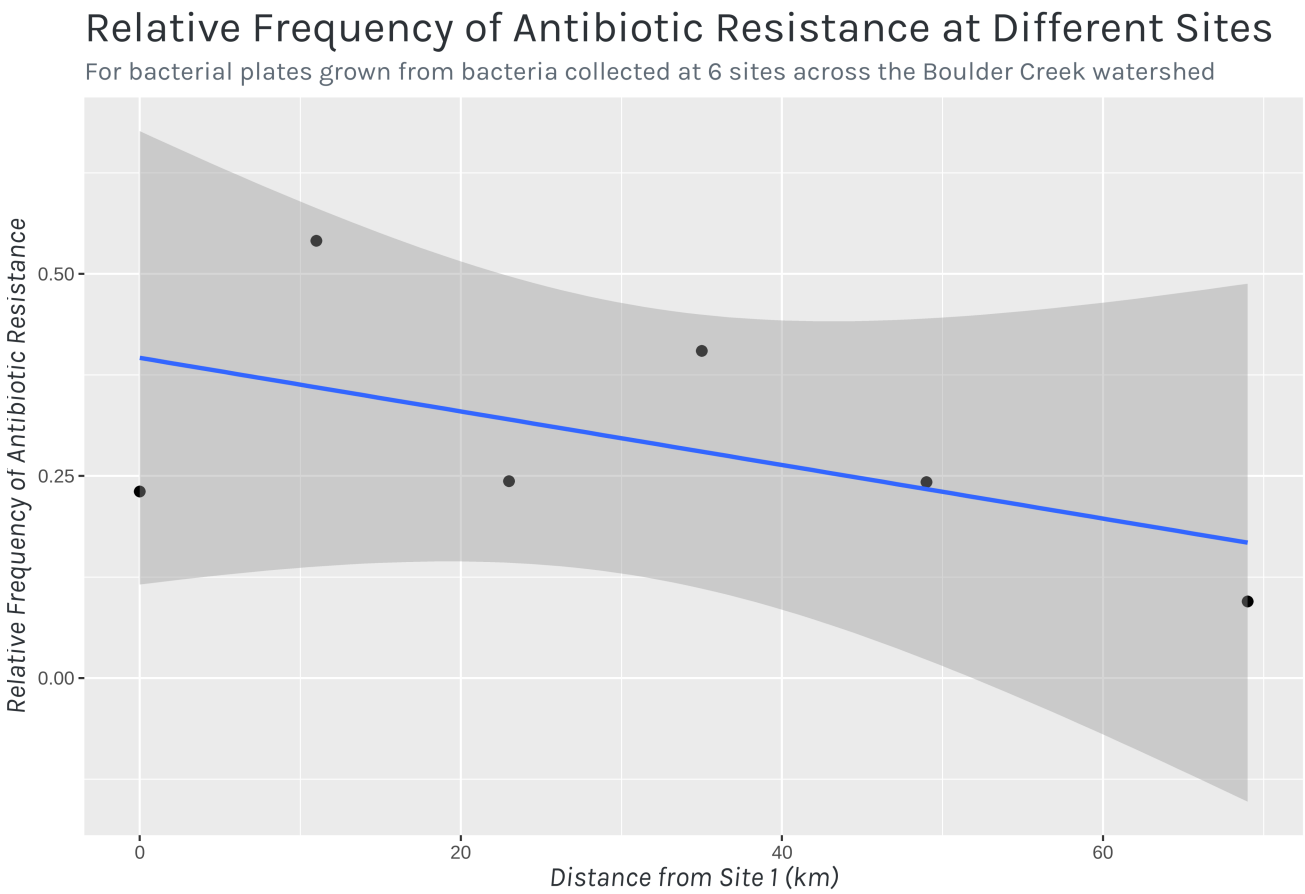

Figure 1. The ratio of experimental (AB+) colonies to control (AB-) colonies across different distances from site 1.

We observed that the relative frequency of antibiotic resistance generally seemed to decline further from site 1. This suggests that antibiotic resistance is less prevalent in bacteria found further downstream from site 1, because the data show that sites closer to site 1 have more colonies on the AB+ plate relative to the AB- plate than those further from site 1. According to the data that we have, it appears that the further from site 1 the bacteria are sampled from, the fewer bacteria found to be resistant to antibiotics. However, the hypothesis can neither be supported nor rejected because of the sampling design. The experiment that yielded these data only checked two petri dishes from each site, which is far too few samples to draw a strong conclusion from the data. Additionally, the **R^2 = 0.29** and **p = 0.27** reveal that the data used to construct the plot are statistically insignificant, and therefore the linear model is a very poor fit.

Evaluating Observation Data

```
# Importing csv and creating dataframe
class_data <- read.csv("Lab06_plate.csv", header = TRUE)
site5 <- subset(class_data, Sample_Site_No. == 5)[4:5]

# Statistics for the data (output not shown)
mean(site5$Total_No._Colonies_AB_Neg)
sd(site5$Total_No._Colonies_AB_Neg)
percentile <- length(which(site5$Total_No._Colonies_AB_Neg <= 66)) /
              length(site5$Sample_Site_No.) * 100
quantile(site5$Total_No._Colonies_AB_Neg, c(0, 0.25, 0.5, 0.75, 1))

# Plotting
boxplot(site5$Total_No._Colonies_AB_Neg, col = c("darkblue"),
        main = "Differences in Bacterial Colony Count in Class Data",
        xlab = "Site 5 Data",
        ylab = "Bacteria Colony Count")
```

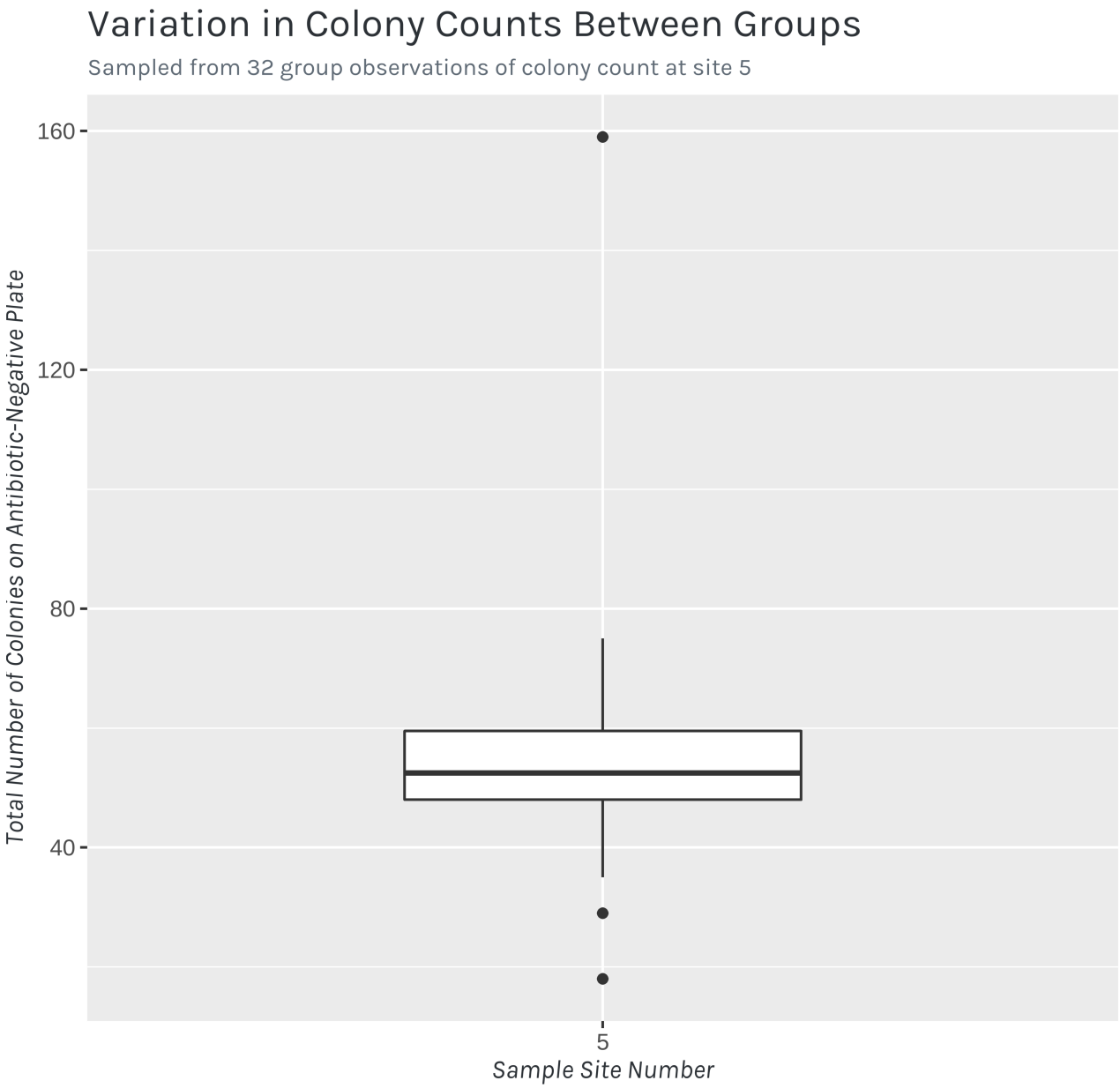

Figure 2. Boxplot of distribution of plate counts of bacterial colonies across the class data

The boxplot shows a reasonable amount of variation (**$\bar{x}$ = 55.66, σ = 22.42**). Our group's data for the total number of bacterial colonies on the antibiotic-negative plate at site 5 **(66** colonies counted**)** was in the **88th percentile (**87.5% rank**)**, meaning that we counted more colonies on the plate than average, but still fell within one standard deviation of the data. It is clear, however, that there is significant variation between group observations. Even within the standard deviation, there is a range of 44 colonies between all the observations, a number that exceeds the minimum observation in the data set. The outliers especially demonstrate the potentially extreme impact of human interpretation on the quality of the data. This exercise shows the importance not only of robust data collection **(**multiple samples for both antibiotic-positive and -negative plates**)** but also of accurate measurements **(**either by using tools like computers or by doing what we've done and incorporating many different counts**)**.

Quantile Data

0TH PERCENTILE	25TH PERCENTILE	50TH PERCENTILE	75TH PERCENTILE	100TH PERCENTILE
18	48	52.5	59.5	159

Group Member Contributions

Kayden: Majority of R coding and some of the write-up

Yousef: Most of the write-up

Dylan: Bacteria counting, plots/tables, some R code, and R Markdown report formatting