Public Health Risk in Bear Creek

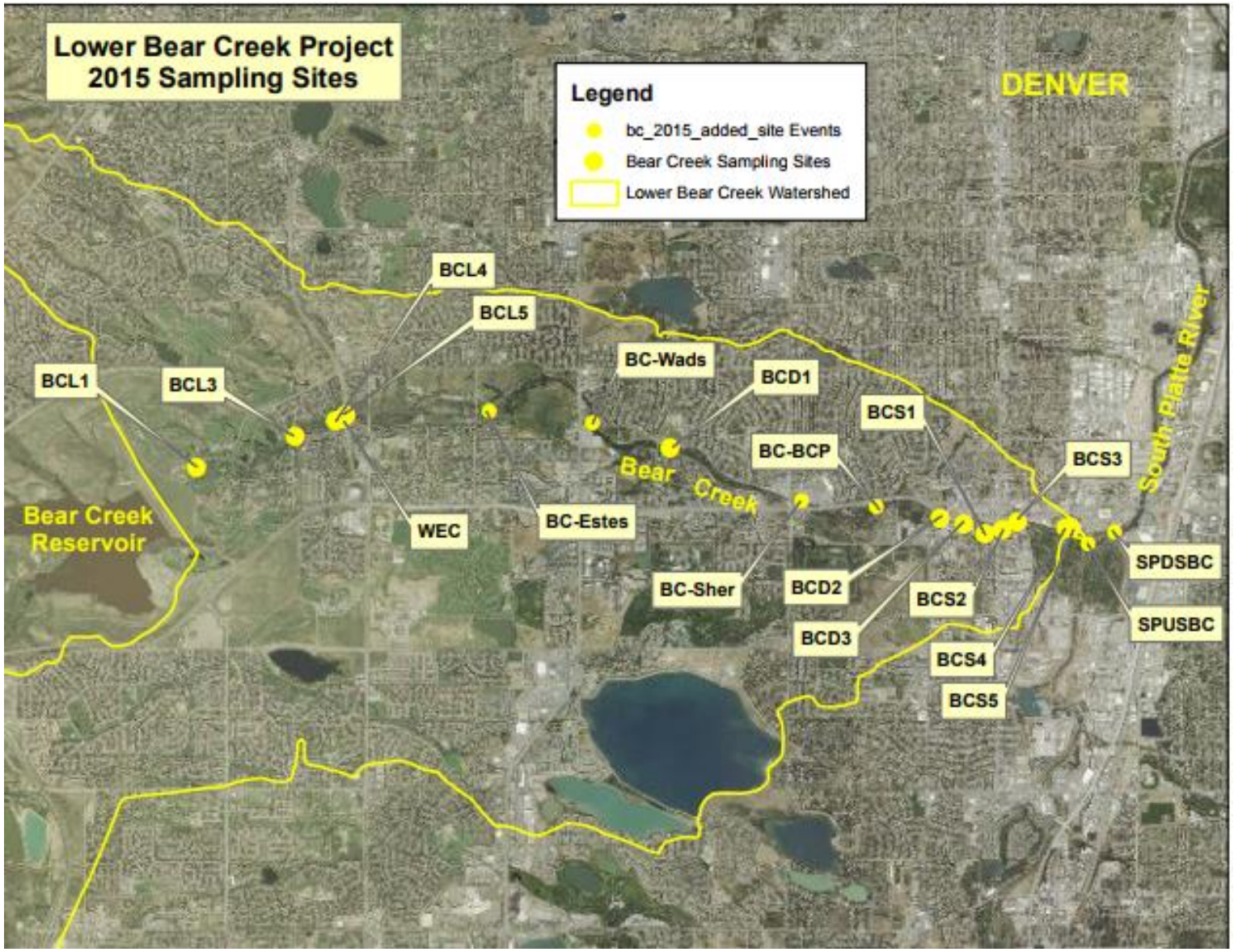
A Time Series Analysis Approach to Estimating E. coli Levels in Bear Creek Watershed: Implications for Further Study.

Christopher Campbell, Ahern Nelson, and Keenan O'Brian,

with Faculty Advisor Dr. Elizabeth Ribble

MSU DENVER

Introduction to Project

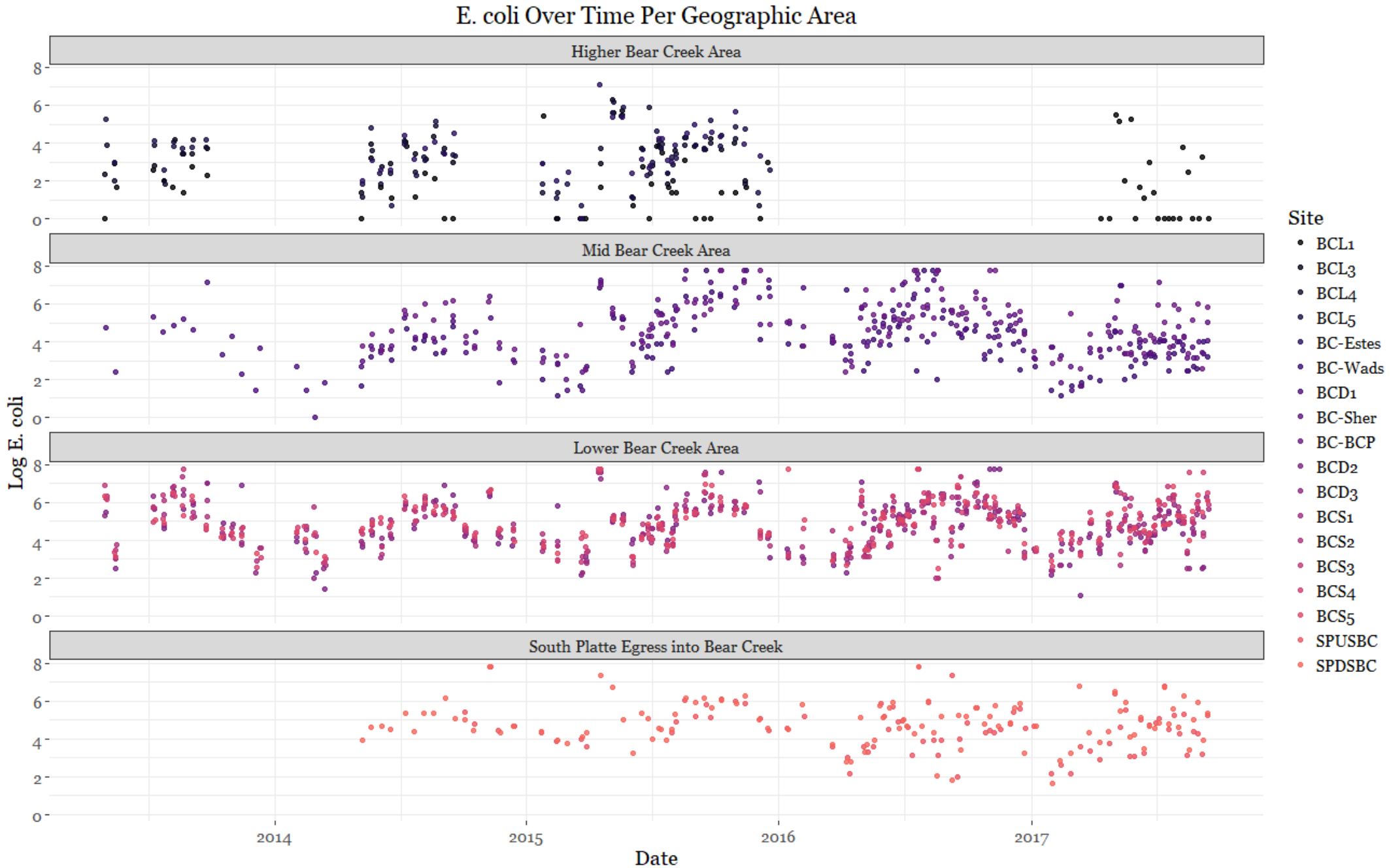


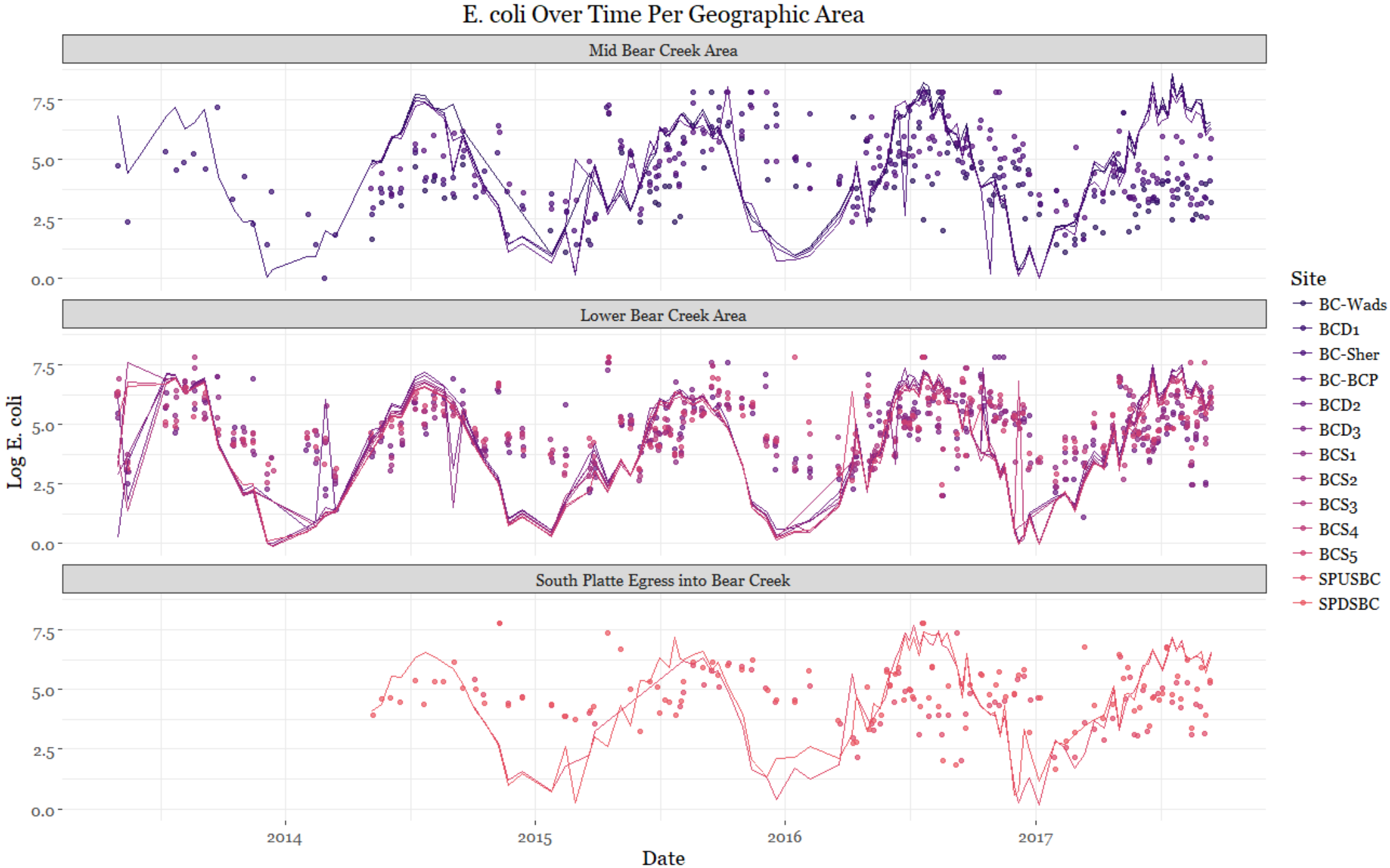
Sampling Sites

Aims of Project

1. Understanding risk posed to humans by bacteria similar to E. coli
2. Determine possible predictors of E. coli
3. Understand variation we see in the data

Exploratory Data Analysis

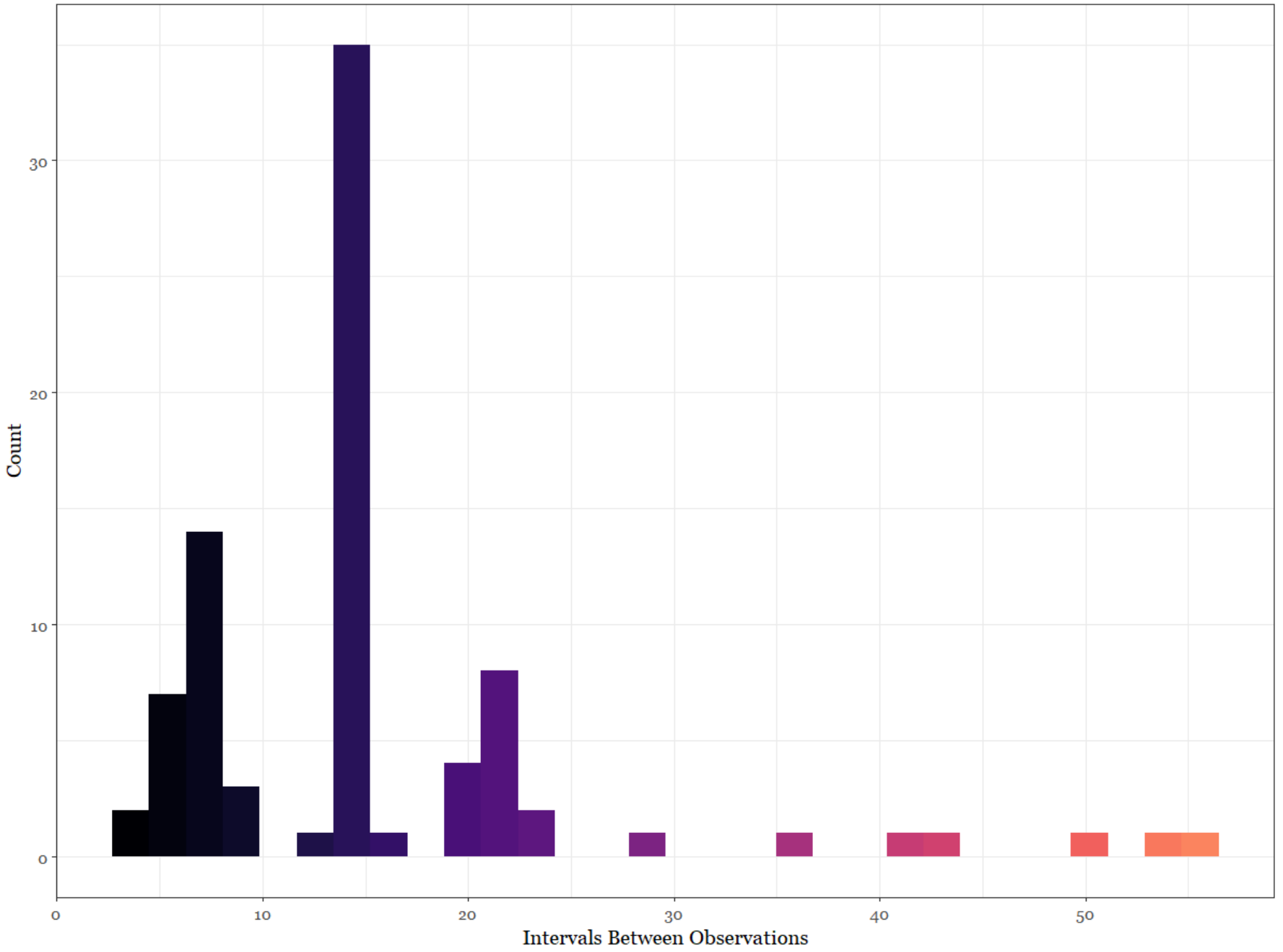

E. coli Over Time Per Geographic Area
Higher Bear Creek Area
Mid Bear Creek Area
Lower Bear Creek Area
South Platte Egress into Bear Creek
Log E. coli
0
2
4
6
8
2014
2015
2016
2017
Date
Site
BCL1
BCL3
BCL4
BCL5
BC-Estes
BC-Wads
BCD1
BC-Sher
BC-BCP
BCD2
BCD3
BCS1
BCS2
BCS3
BCS4
BCS5
SPUSBC
SPDSBC

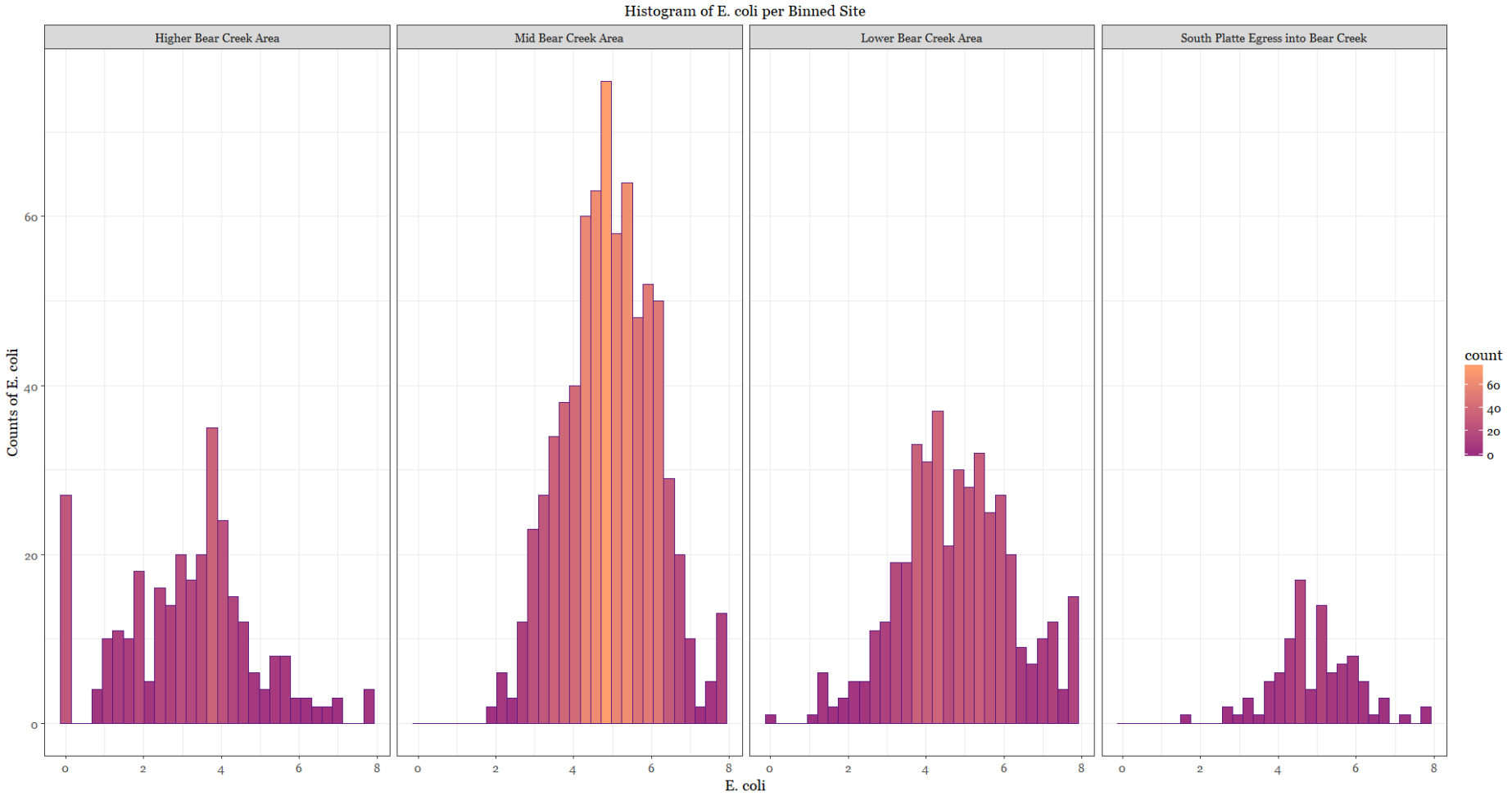

E. coli Over Time Per Geographic Area
Mid Bear Creek Area
Lower Bear Creek Area
South Platte Egress into Bear Creek
Log E. coli
7.5
5.0
2.5
0.0
2014
2015
2016
2017
Date
Site
BC-Wads
BCD1
BC-Sher
BC-BCP
BCD2
BCD3
BCS1
BCS2
BCS3
BCS4
BCS5
SPUSBC
SPDSBC

Some Control Issues

- Data collected over sporadic intervals…

- ….Can bin almost all data bi-weekly.





Data Ceiling

Past Modeling Attempts and Independence

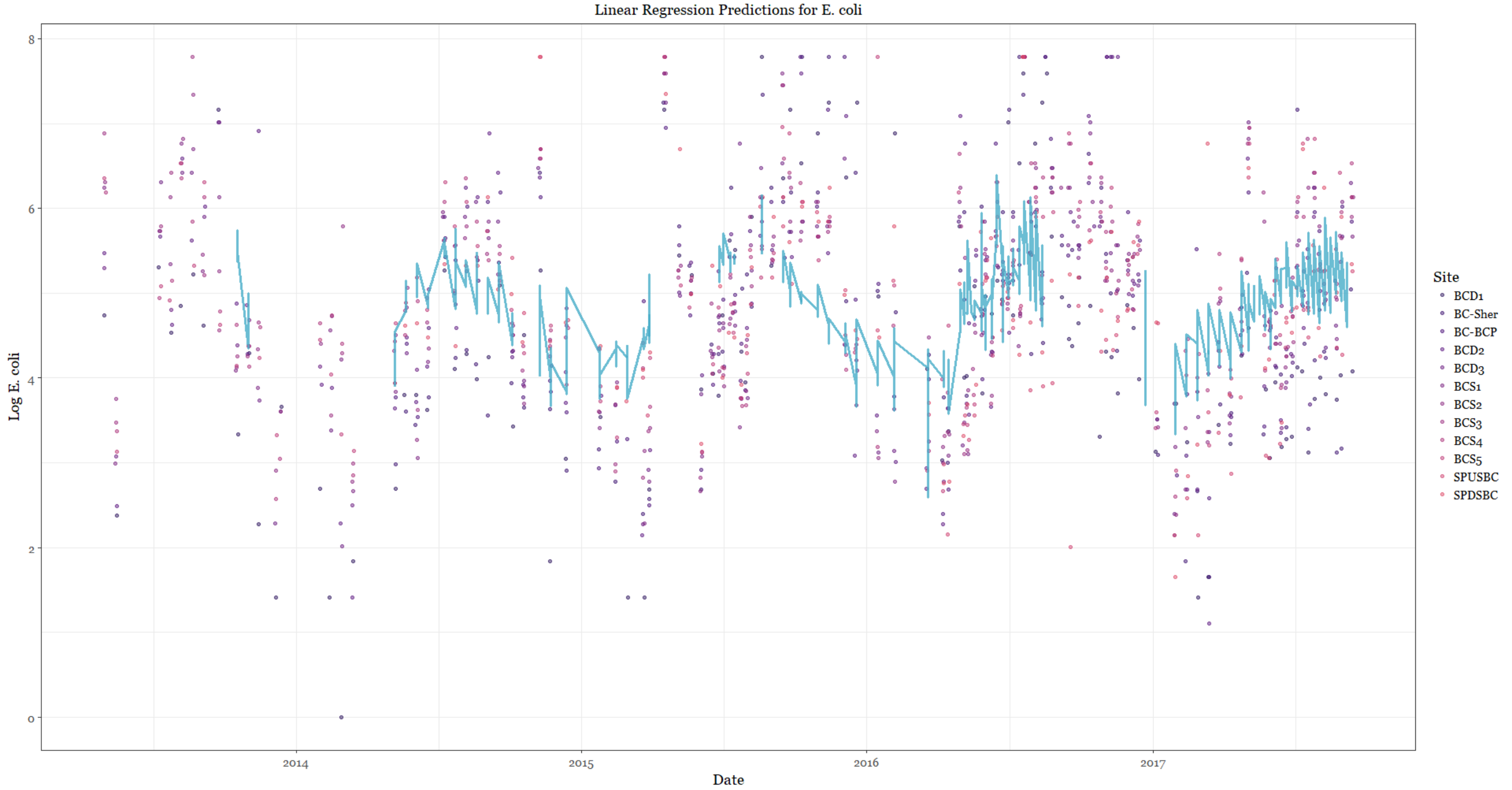
- Statistical tests rely on the assumption of independence of the data. This includes ANOVA and T-Tests.
- Also, regressions will not produce accurate t-values when the data has autocorrelation.

We may want to try regression…

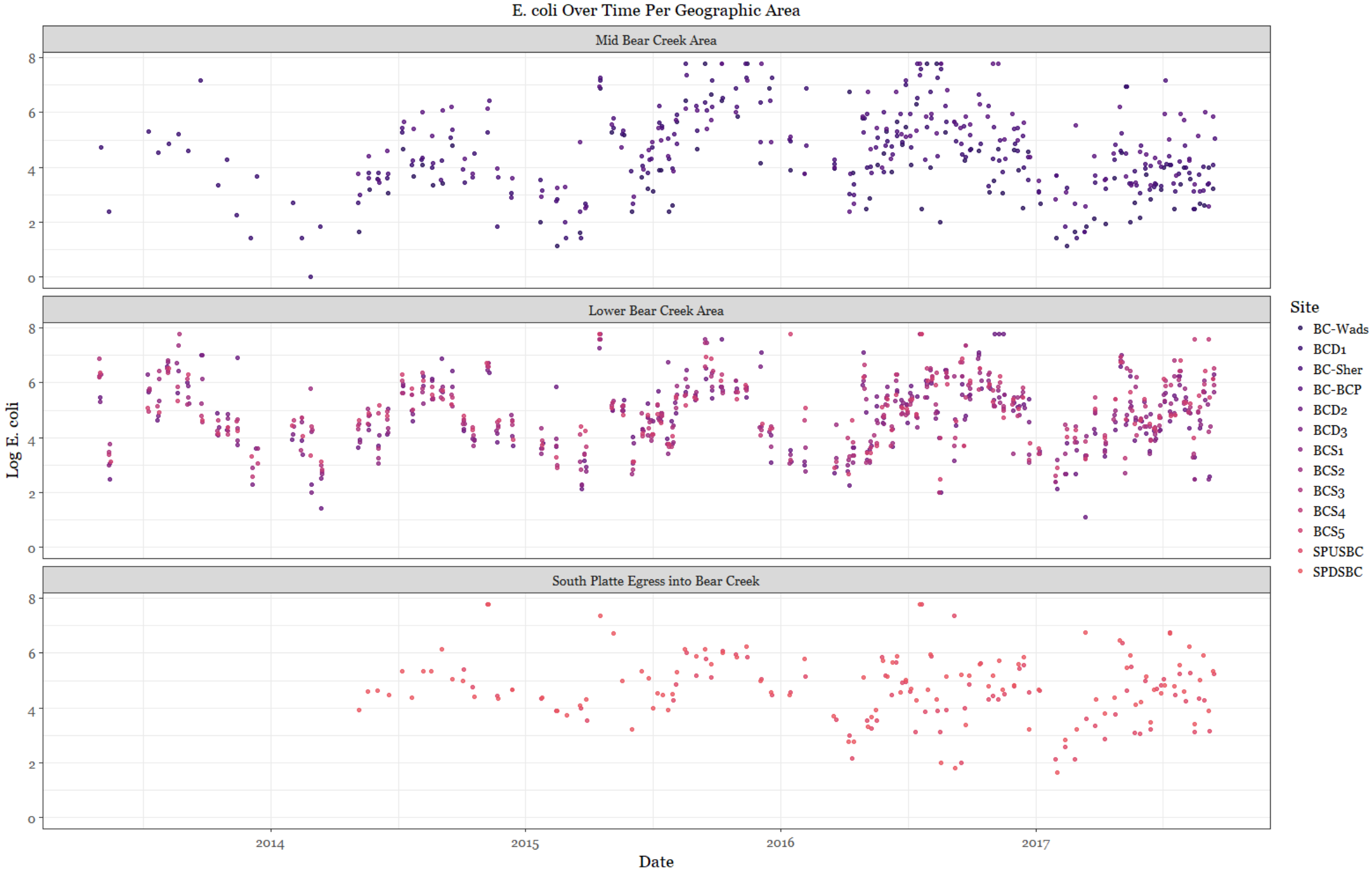
```
Call:
lm(formula = logEColi ~ tempC + pH + turbidity)

Residuals:
    Min      1Q  Median      3Q     Max
-3.0745 -0.7365 -0.0902  0.6455  3.8253

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.285571   0.826594  12.443  < 2e-16 ***
tempC        0.065491   0.006478  10.109  < 2e-16 ***
pH          -0.837798   0.105711  -7.925 7.21e-15 ***
turbidity    0.027899   0.006466   4.315 1.79e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

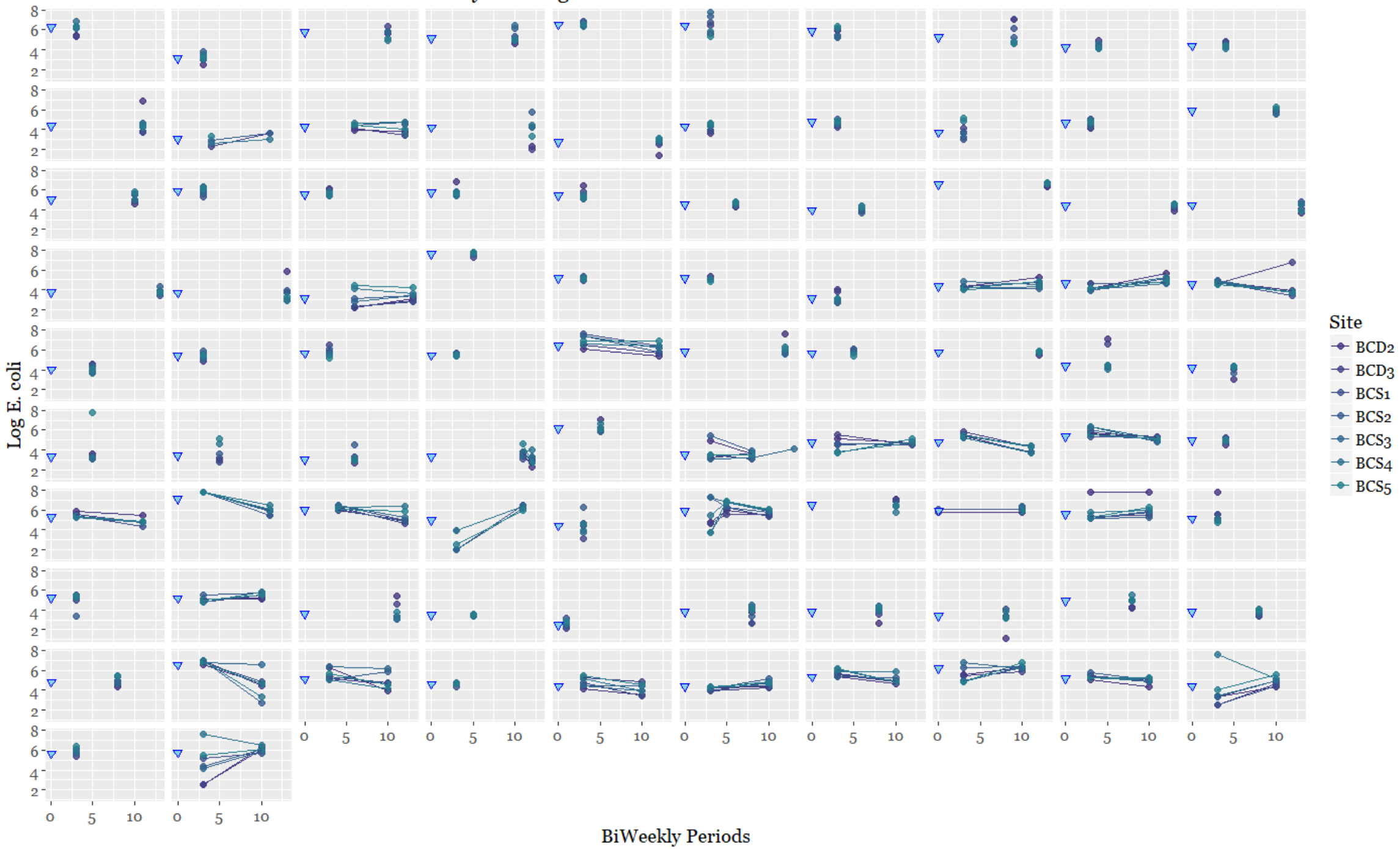
This regression has highly significant coefficients,

however….

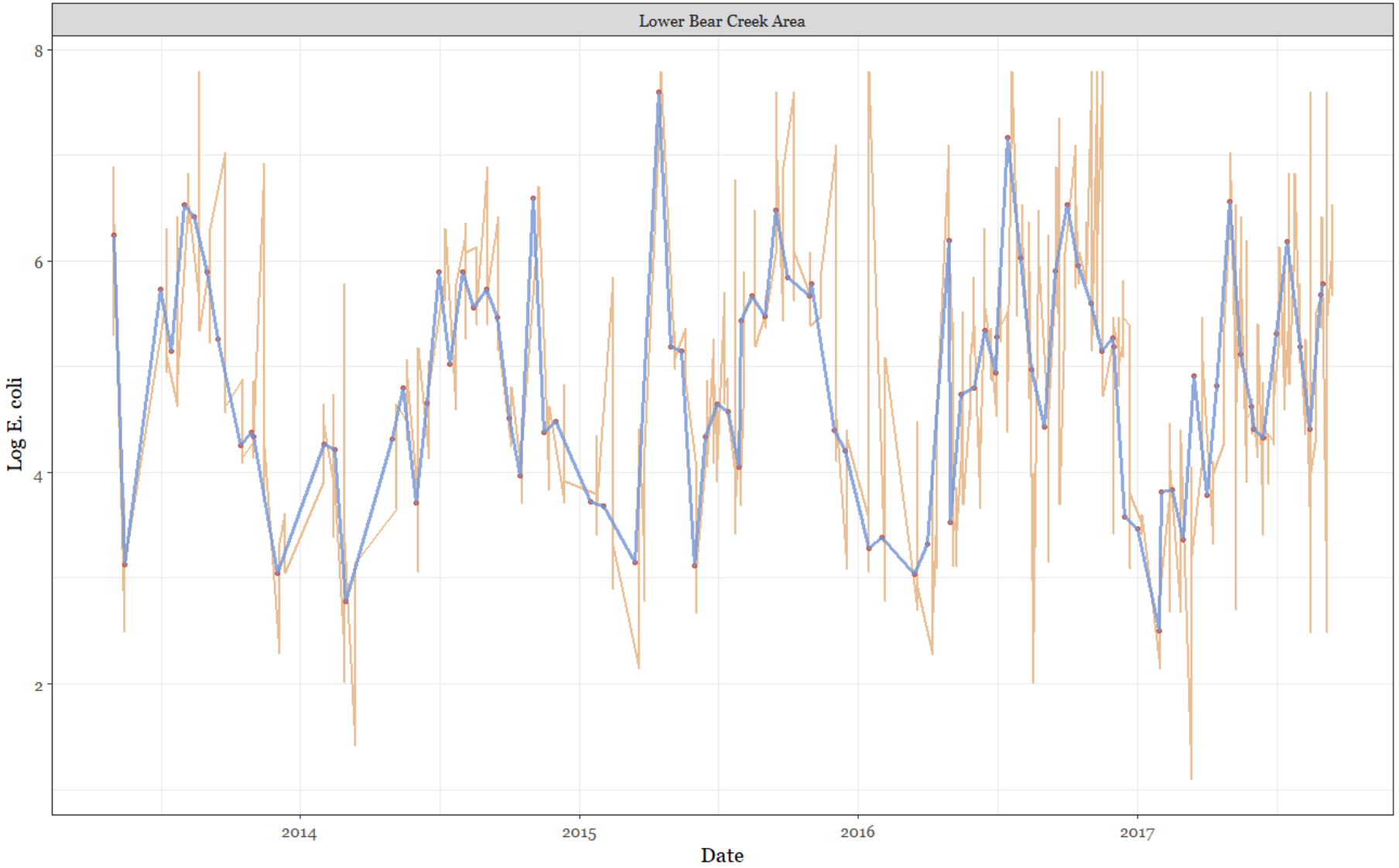

Linear Regression Predictions for E. coli
Log E. coli
8
6
4
2
0
2014
2015
2016
2017
Date
Site
BCD1
BC-Sher
BC-BCP
BCD2
BCD3
BCS1
BCS2
BCS3
BCS4
BCS5
SPUSBC
SPDSBC

Time Series Analysis


E. coli Over Time Per Geographic Area
Mid Bear Creek Area
Lower Bear Creek Area
South Platte Egress into Bear Creek
Log E. coli
0
2
4
6
8
2014
2015
2016
2017
Date
Site
BC-Wads
BCD1
BC-Sher
BC-BCP
BCD2
BCD3
BCS1
BCS2
BCS3
BCS4
BCS5
SPUSBC
SPDSBC

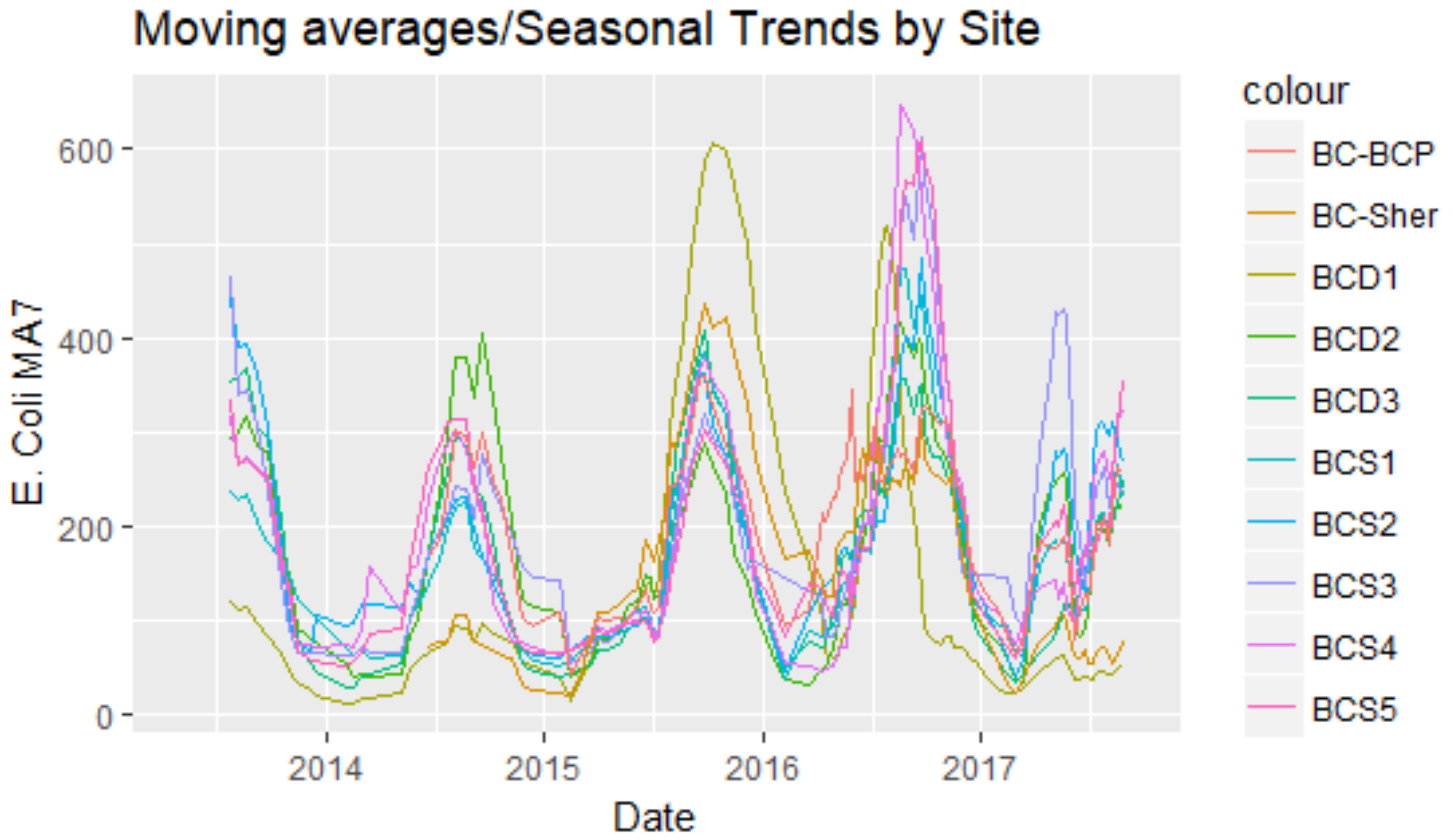

BiWeekly Readings of E. coli in the Lower Bear Creek Area
Log E. coli
BiWeekly Periods
Site
BCD2
BCD3
BCS1
BCS2
BCS3
BCS4
BCS5

Median as a Binning Statistic

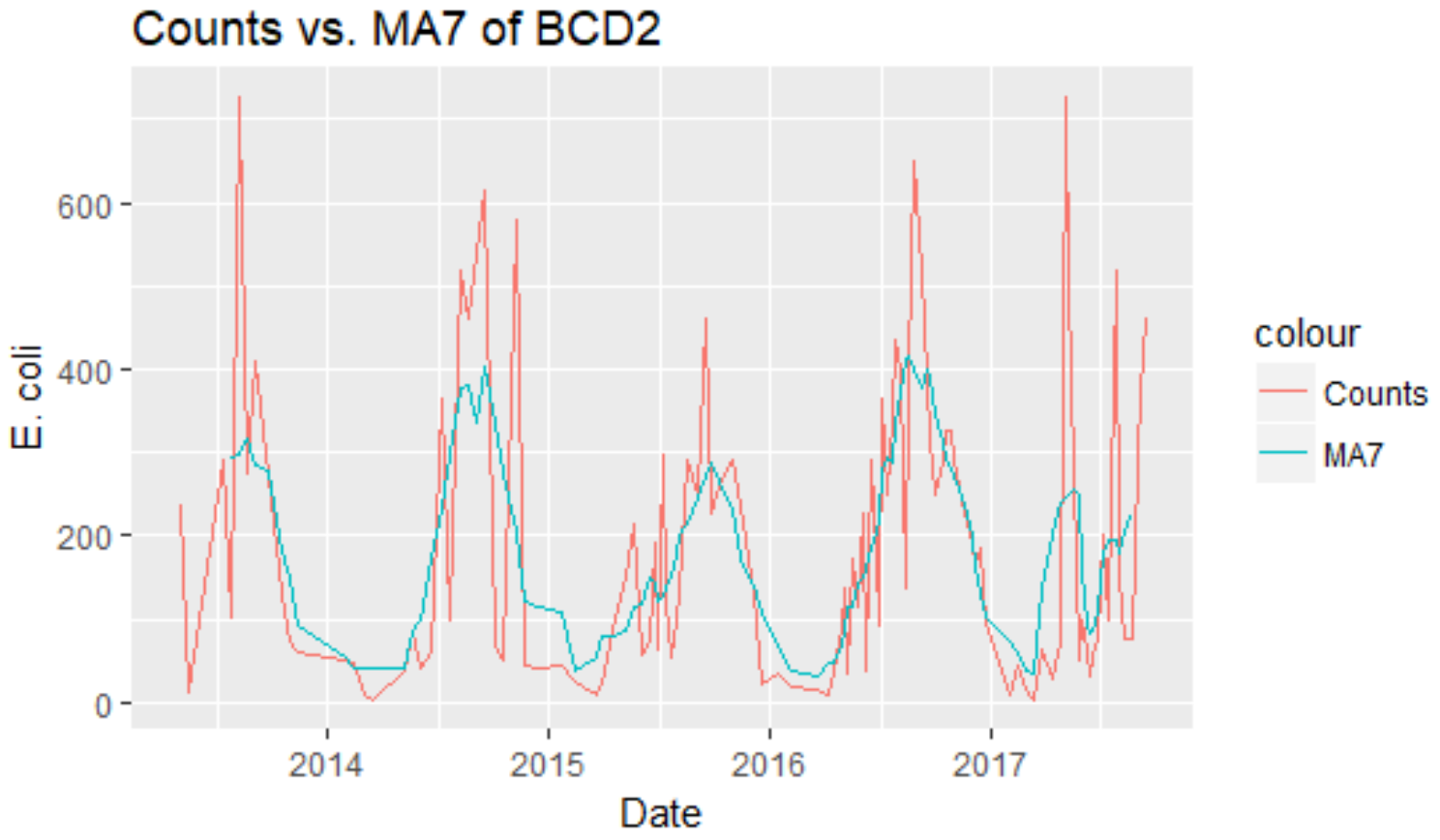


Time Series Model Specification

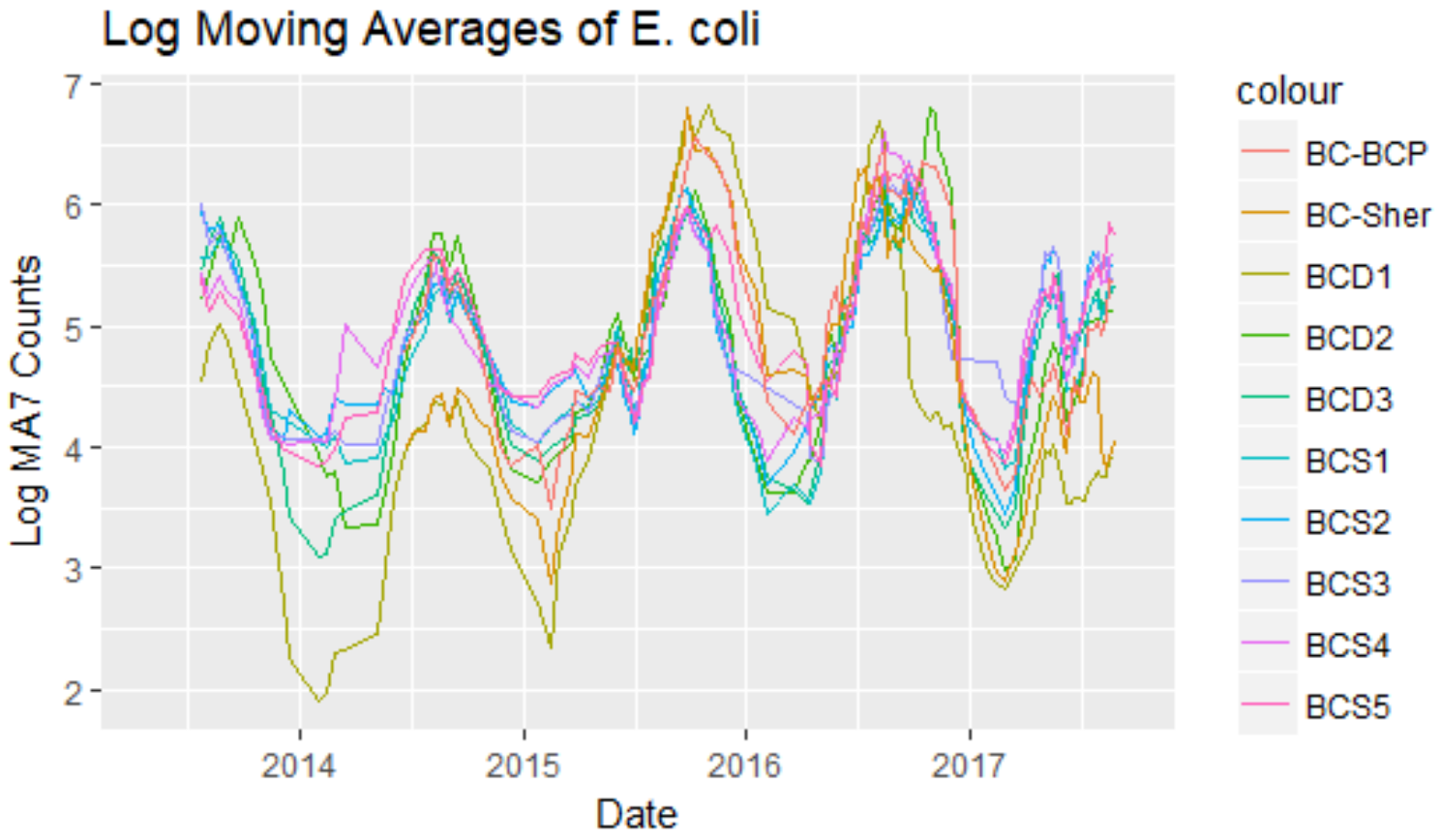
Moving Averages of Individual Sites demonstrates clear seasonal trends



Example: Moving Average smooths trends



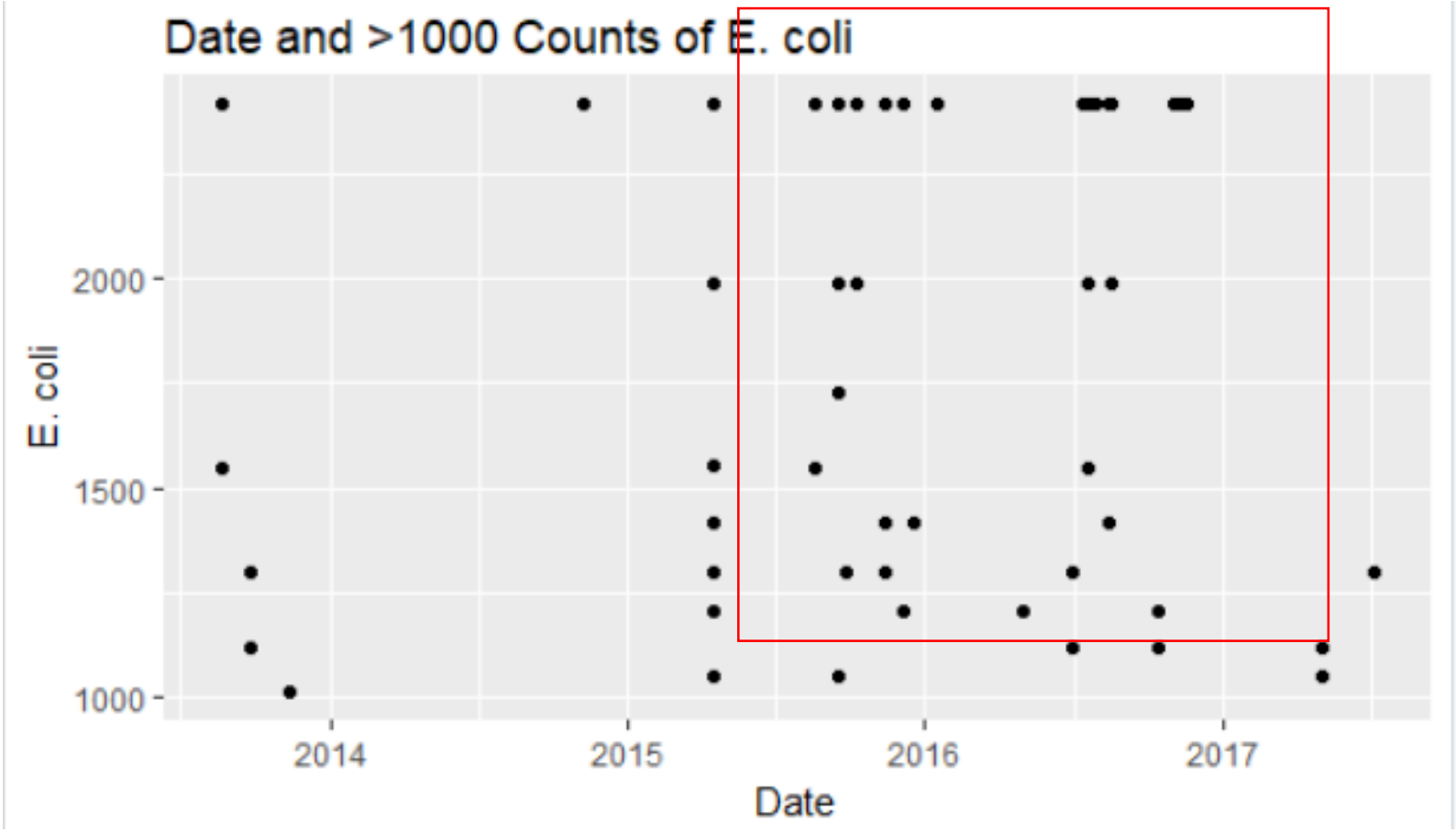
Log Transformation: Possible upward trend?



Possible Upward Trend in Data?

- Observed trend for 2014-2016, data should be monitored to see if trend persists in future years
- Trend wasn’t apparent from 2013 to 2014 or 2016 to 2017
- Upon inspection, it appears that 2015 and 2016 had a uniquely high number of high counts compared to other years
- Sewage Break incidents?

Clusters of high counts for 2015 and 2016

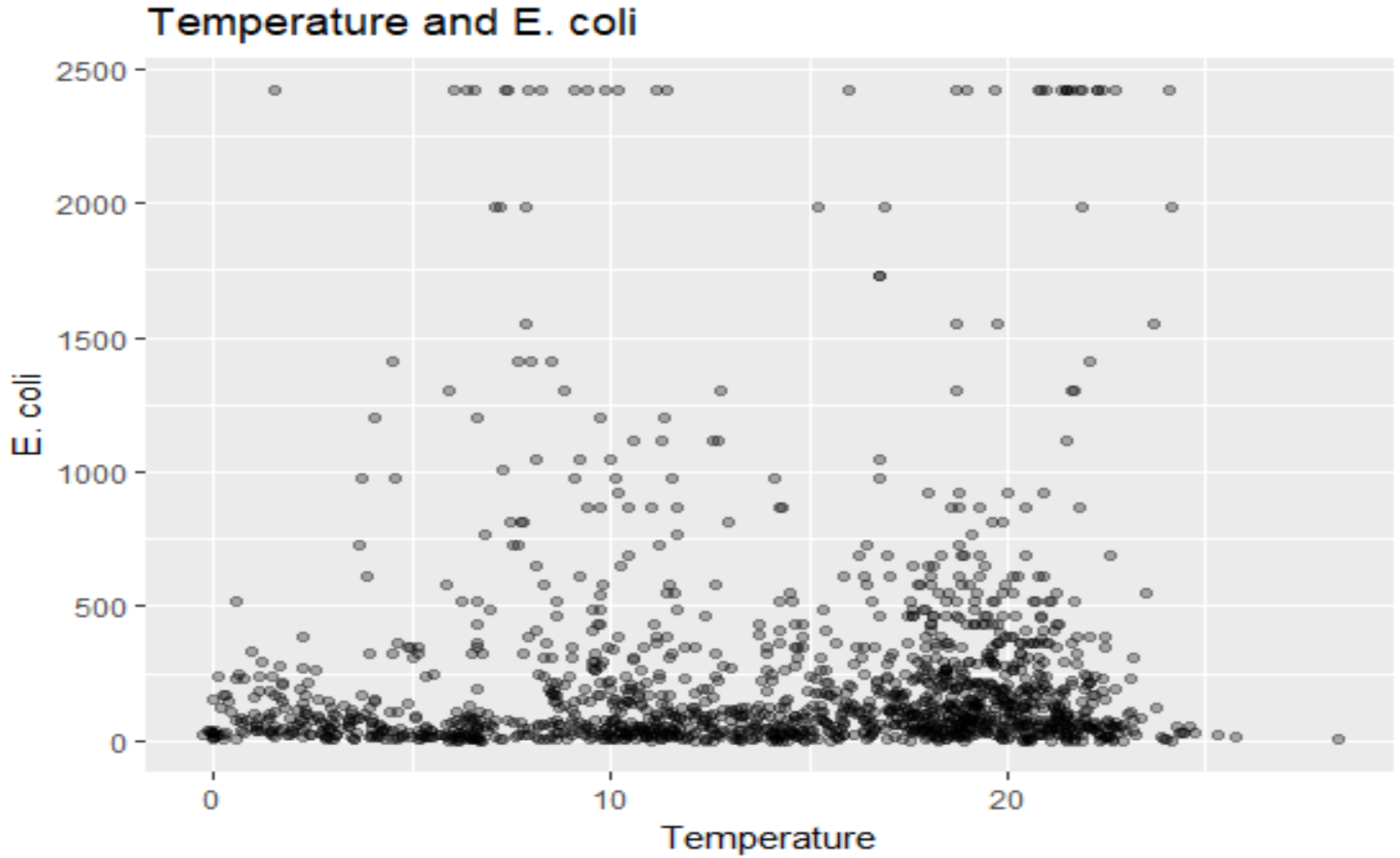


Activation Temperature effects

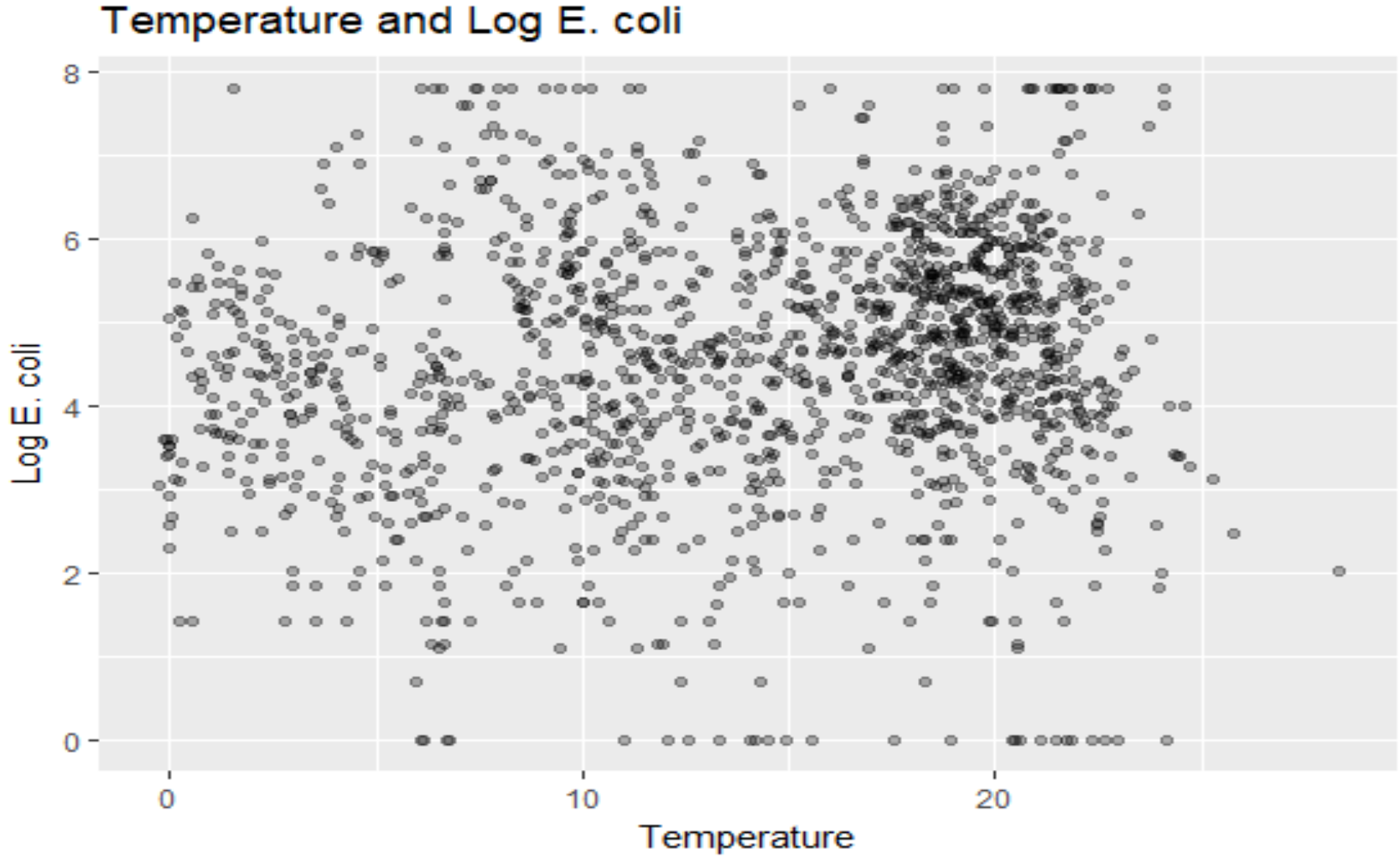
- Activation temperature range of E. coli falls between 15°C and 45°C
- T-test analysis (with log transformation) reveals a significant difference of e. coli counts for observations in activation range vs. not, though not a particularly large difference…

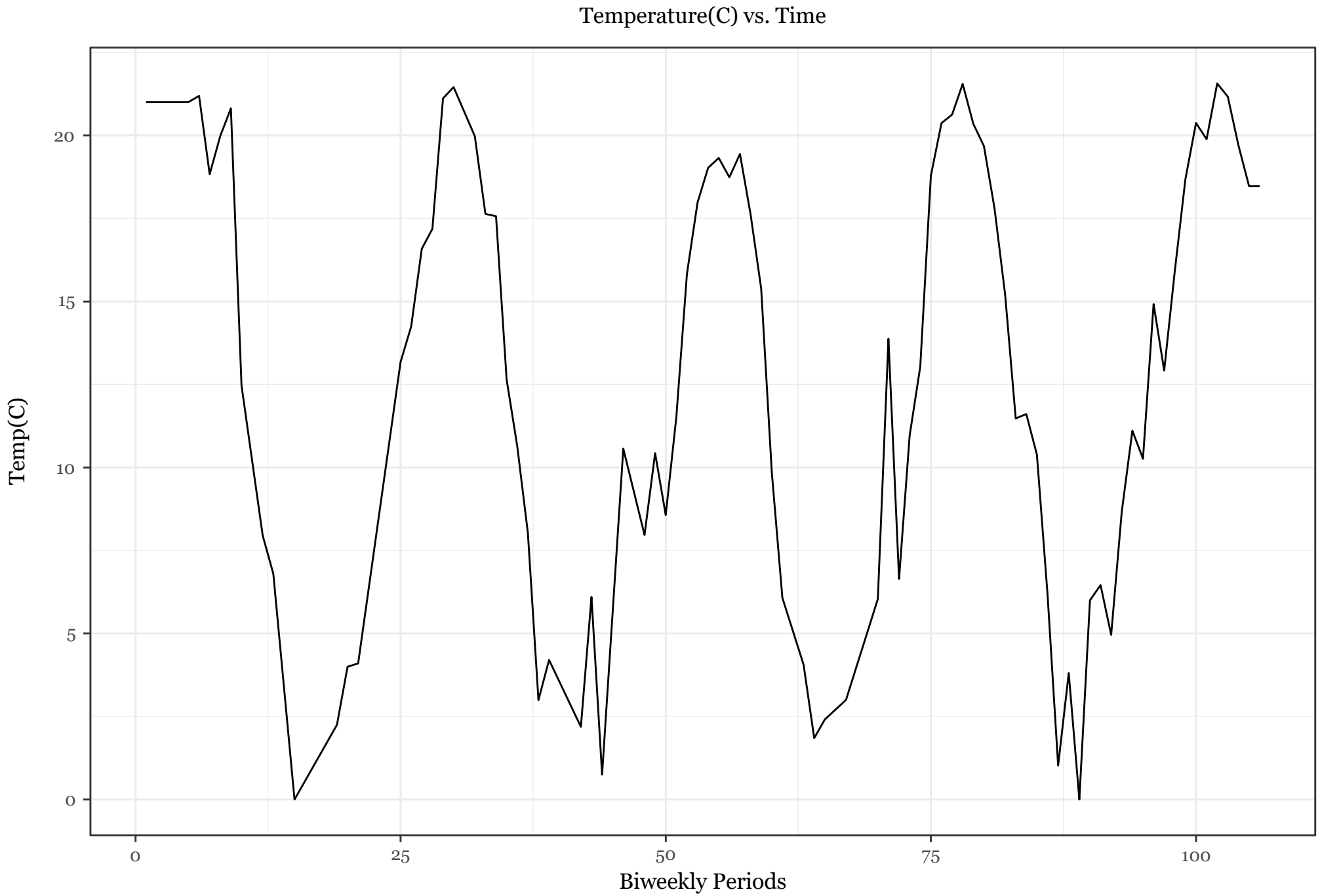
```
        Welch Two Sample t-test

data:  activation_log$logE.coli and nonactivation_log$logE.coli
t = 5.9528, df = 1466.8, p-value = 3.292e-09
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.3060827 0.6069460
sample estimates:
mean of x mean of y
 4.775455  4.318941
```

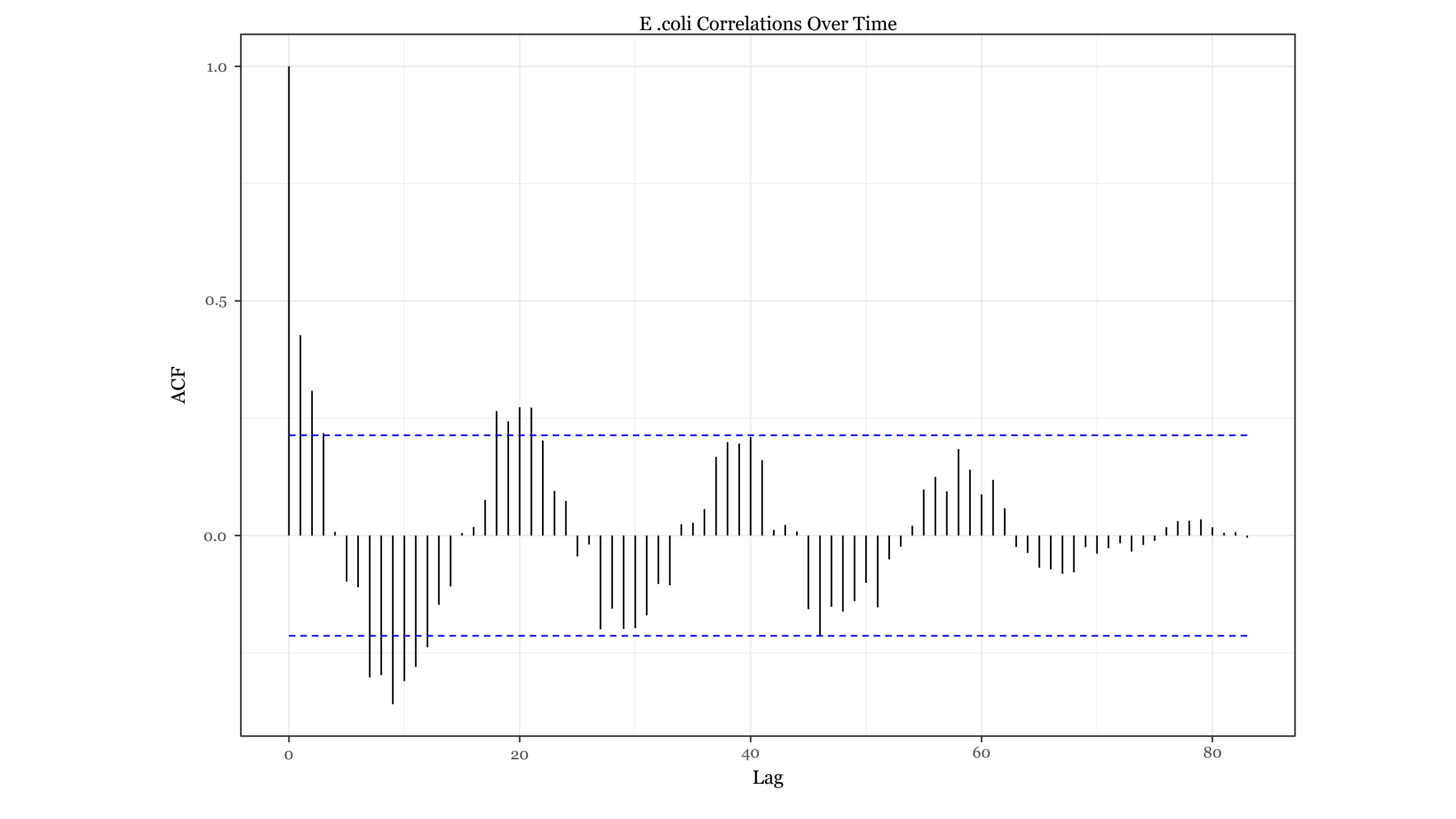
Activation Temperature effects

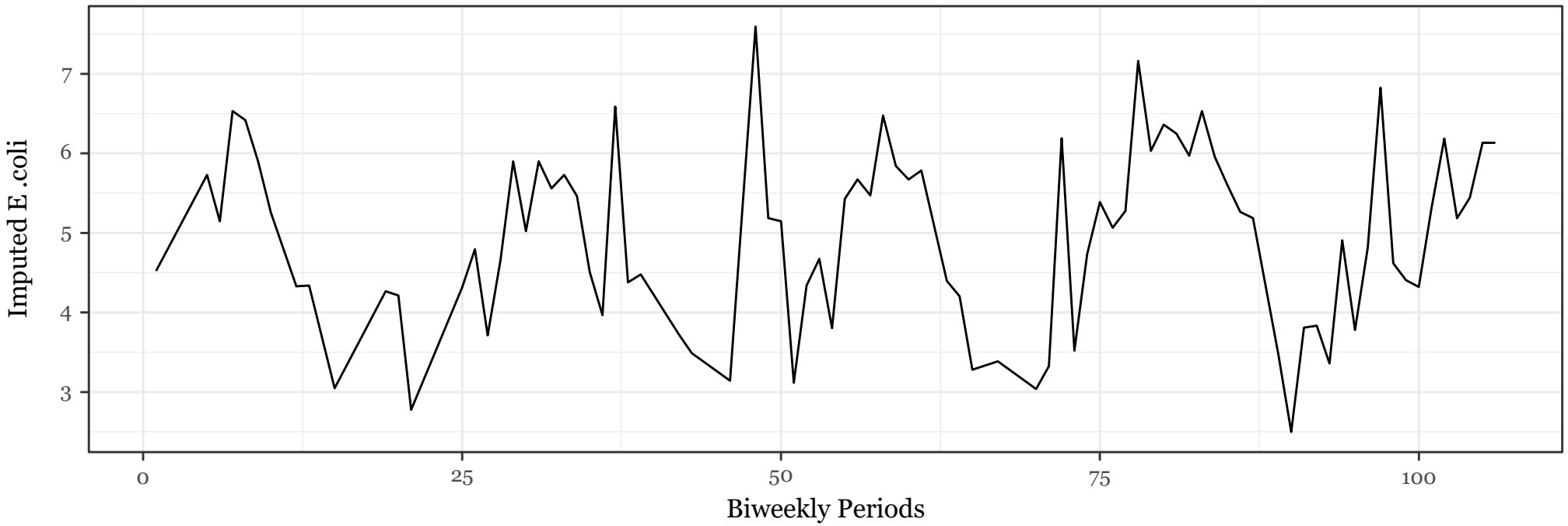


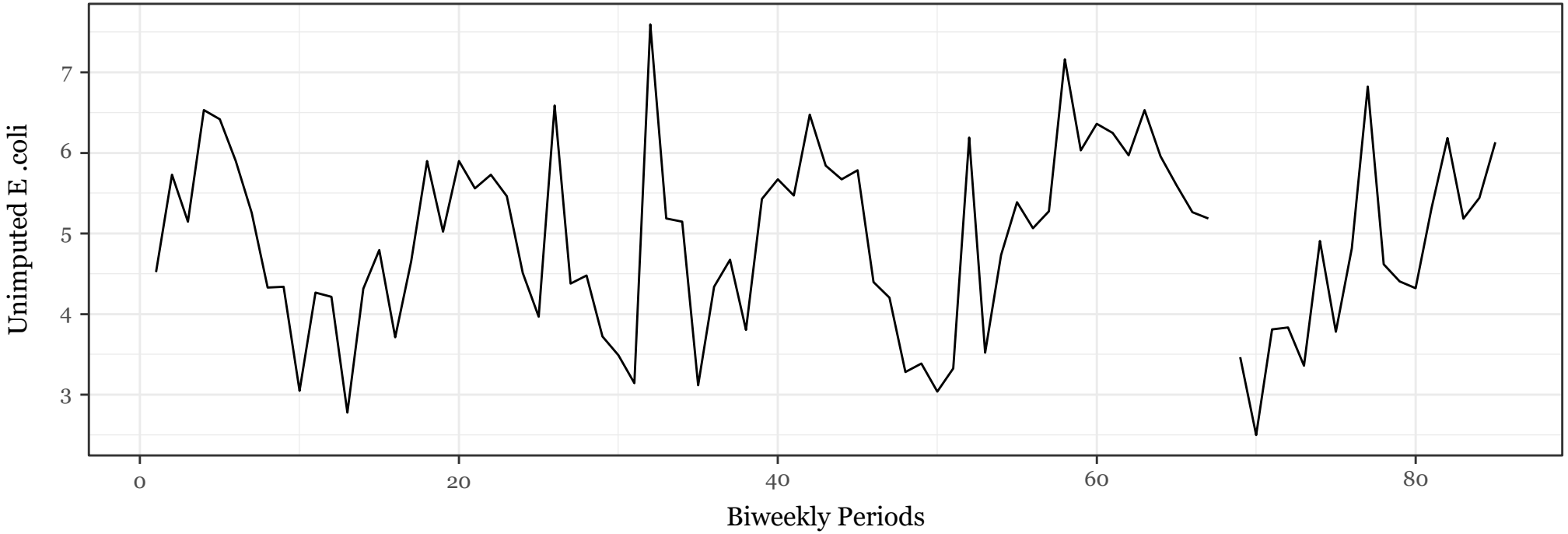
Activation Temperature effects

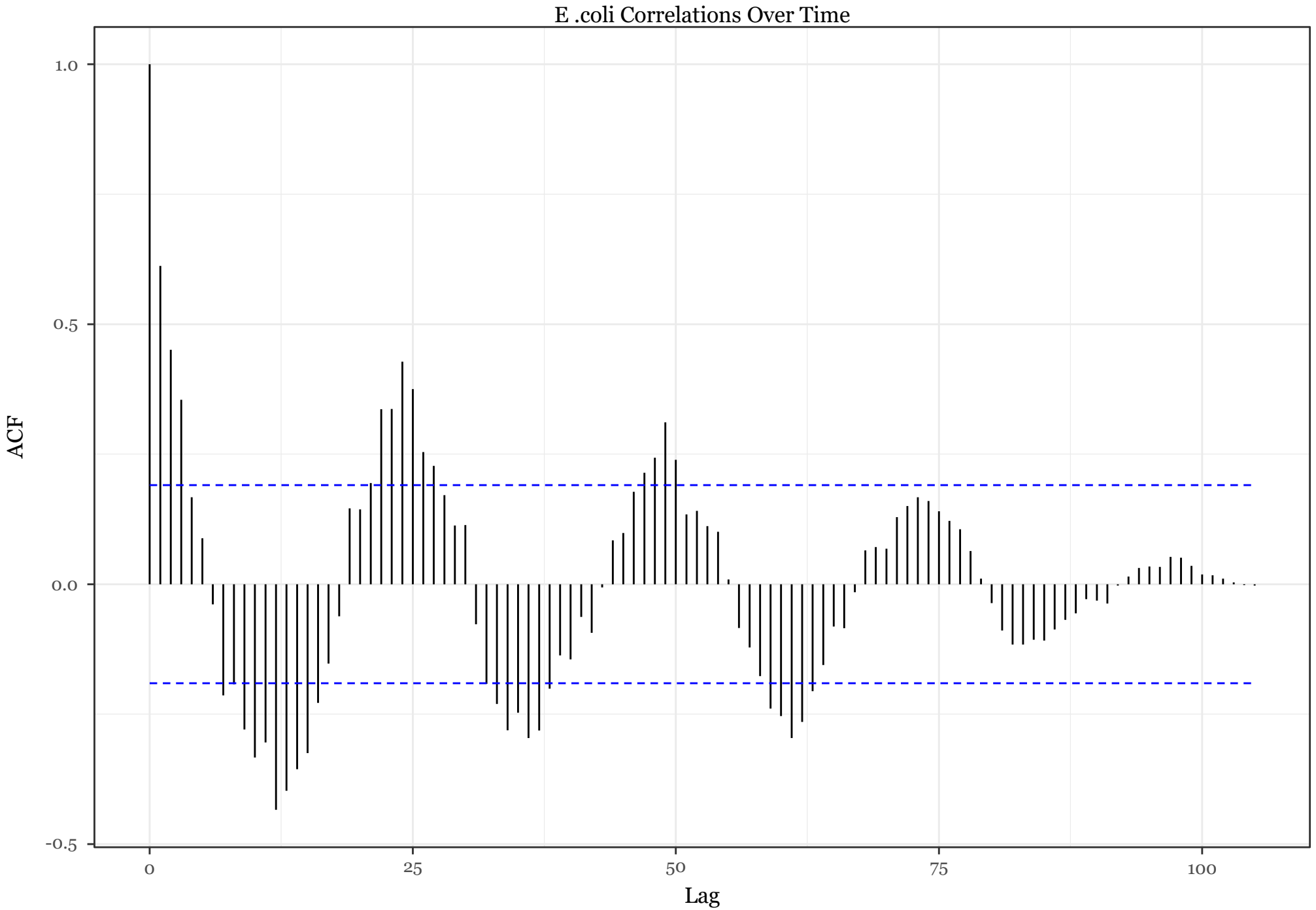



Temperature(C) vs. Time
Temp(C)
20
15
10
5
0
0
25
50
75
100
Biweekly Periods

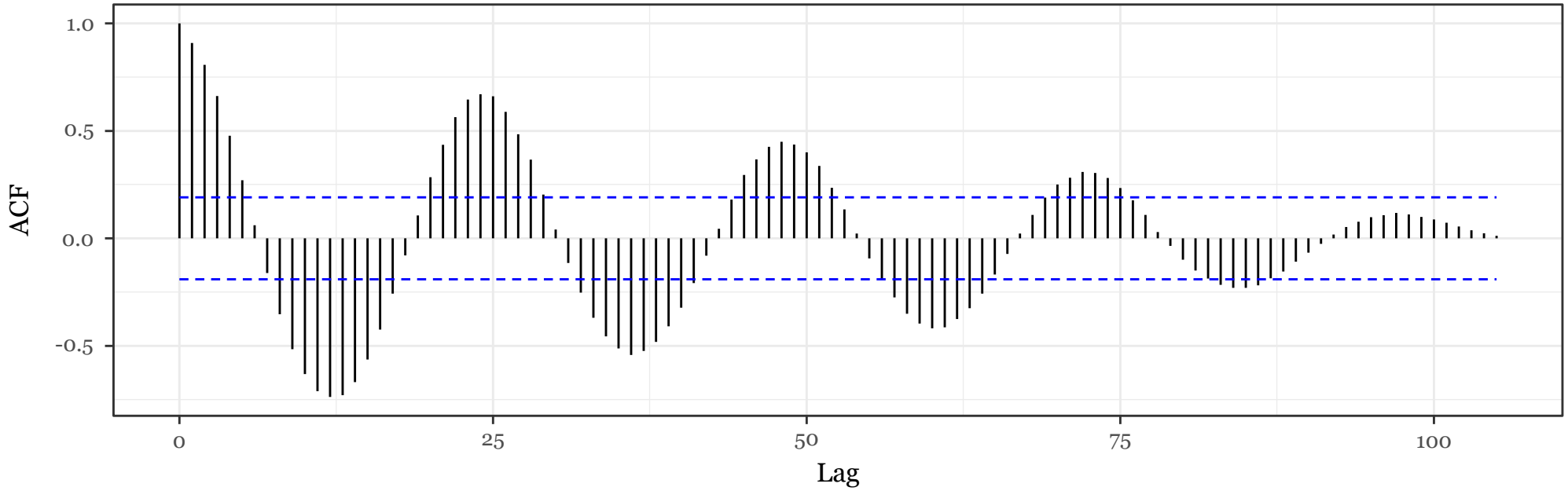

E .coli Correlations Over Time
1.0
0.5
0.0
ACF
0
20
40
60
80
Lag


Imputed E .coli
7
6
5
4
3
0
25
50
75
100
Biweekly Periods

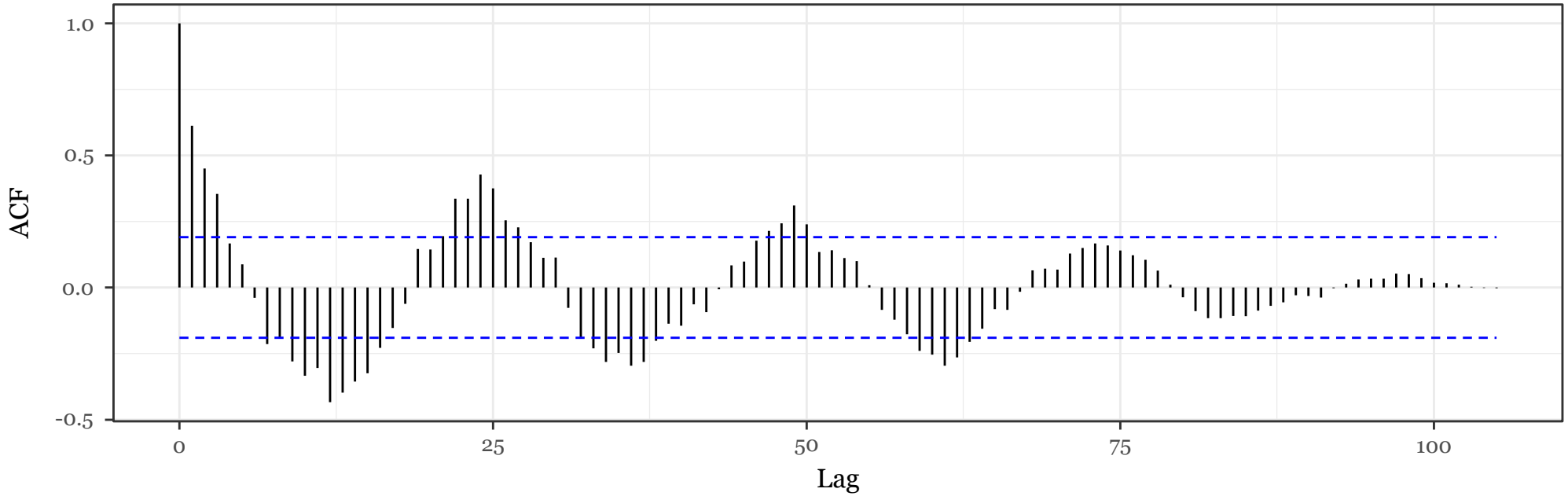

Unimputed E .coli
7
6
5
4
3
0
20
40
60
80
Biweekly Periods

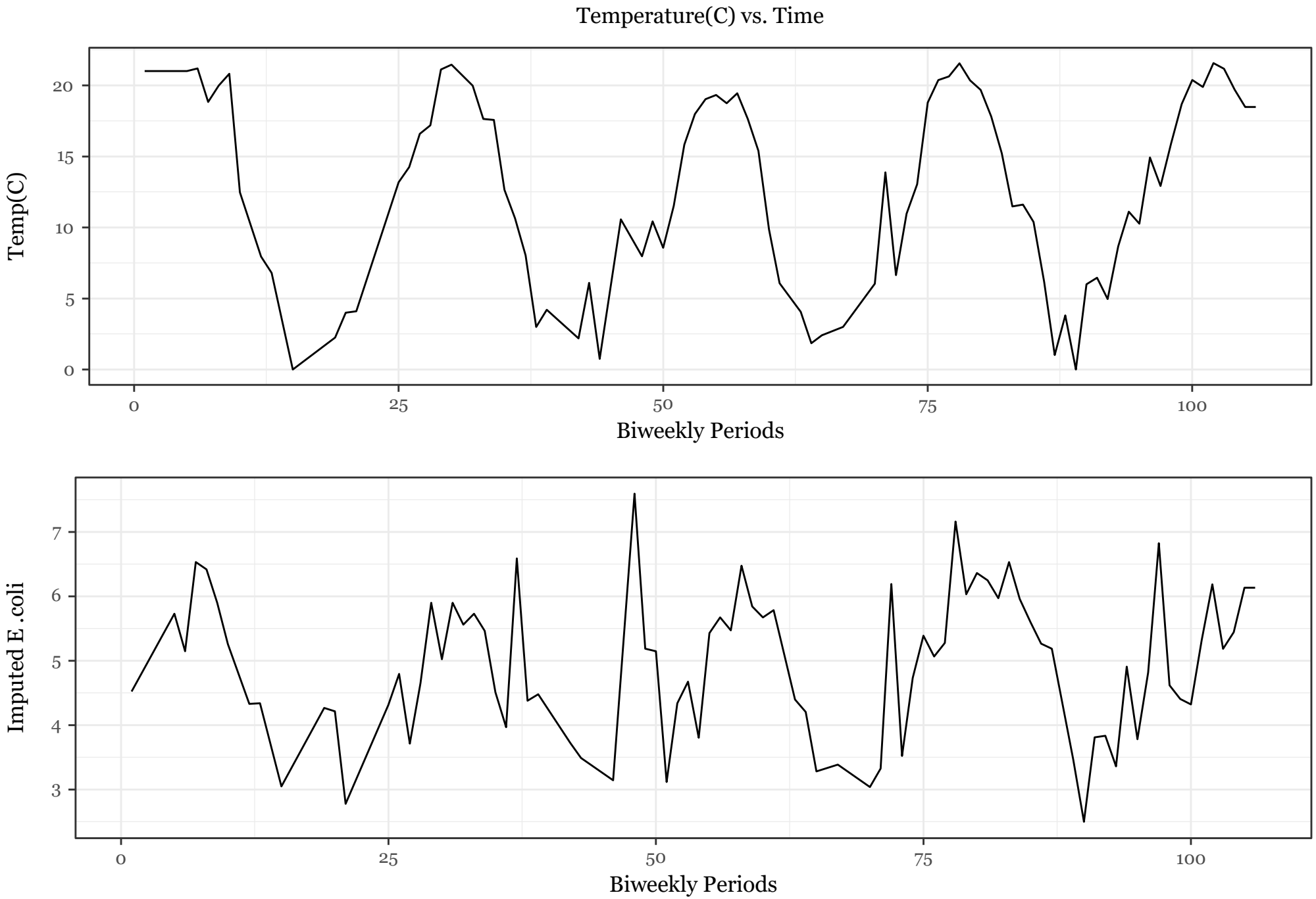

E .coli Correlations Over Time
1.0
0.5
0.0
-0.5
ACF
0
25
50
75
100
Lag

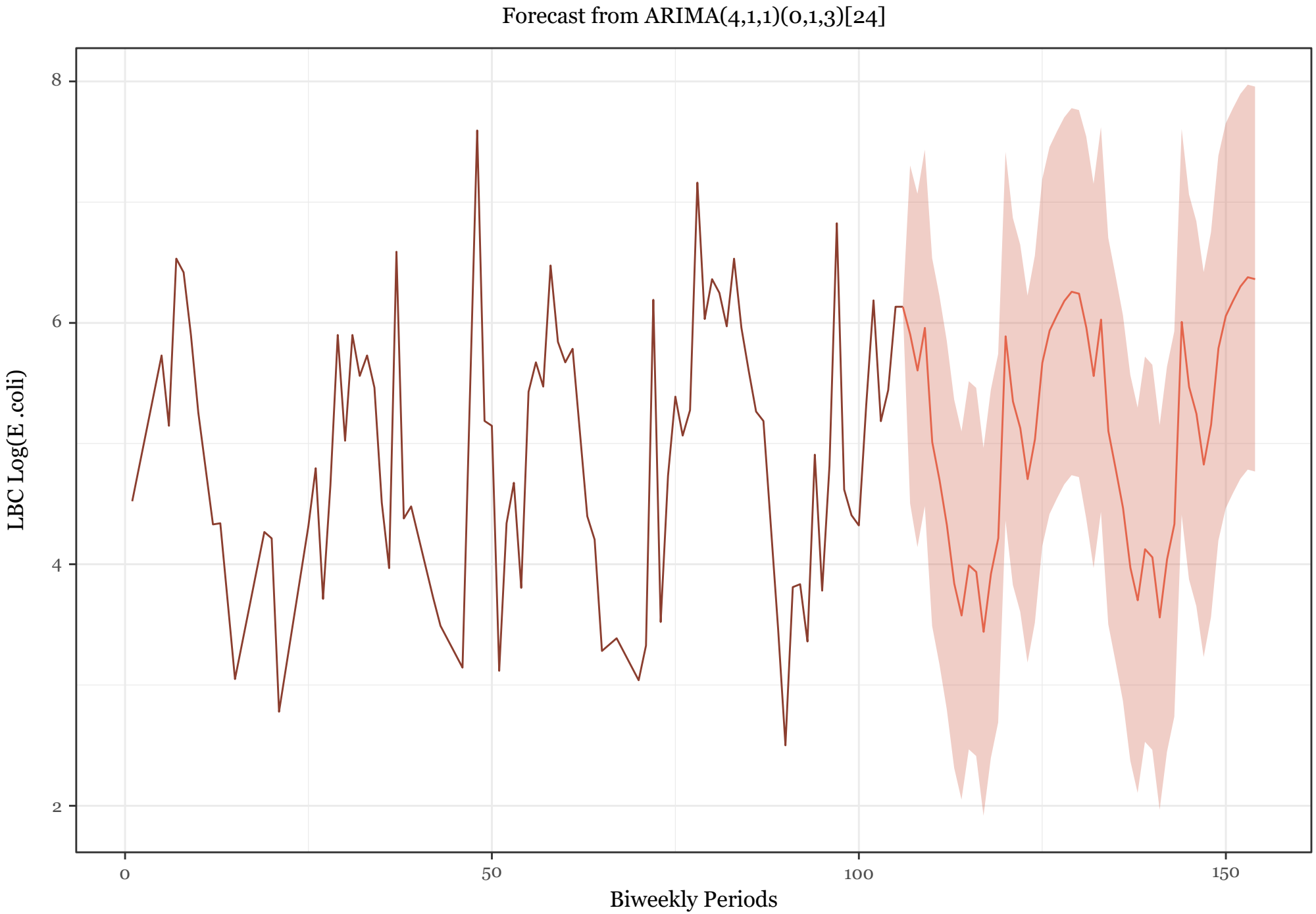
Temperature Correlations Over Time

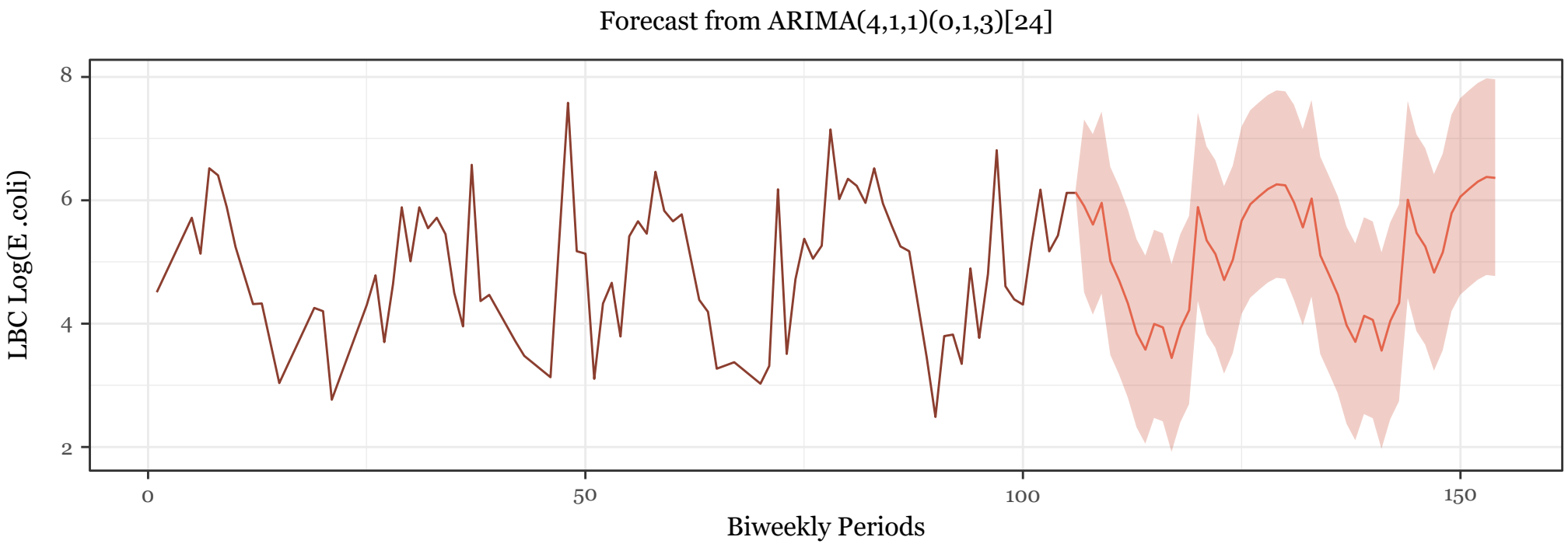


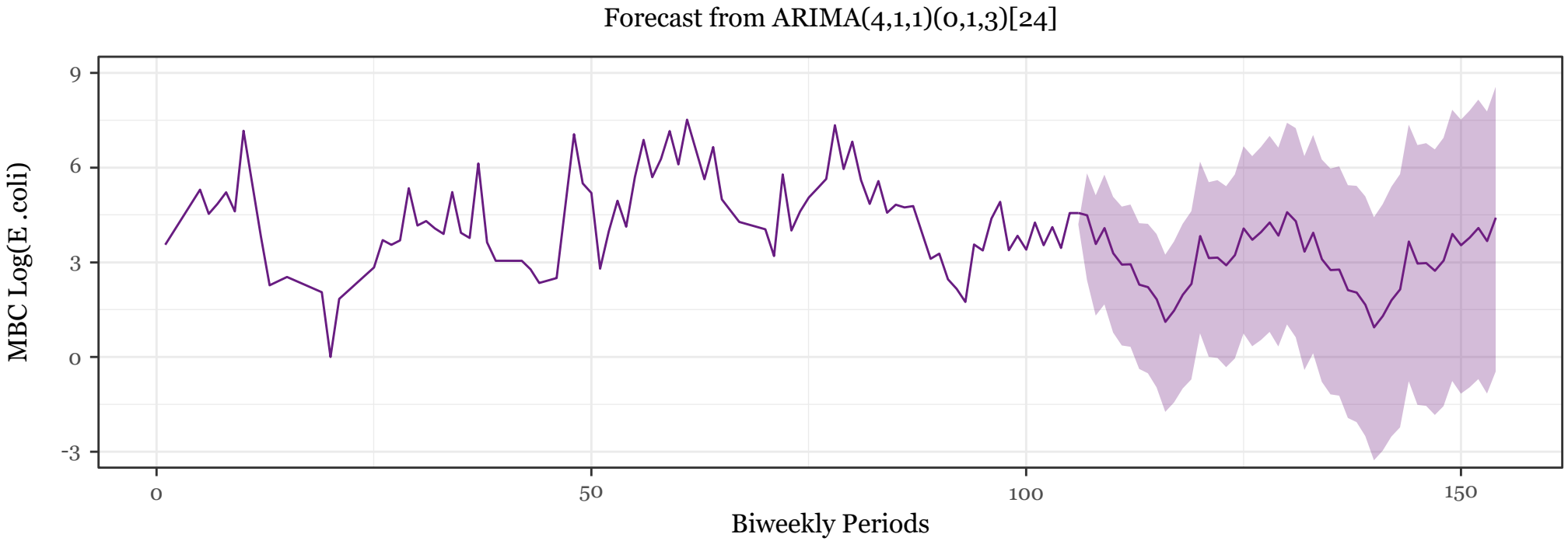
E .coli Correlations Over Time

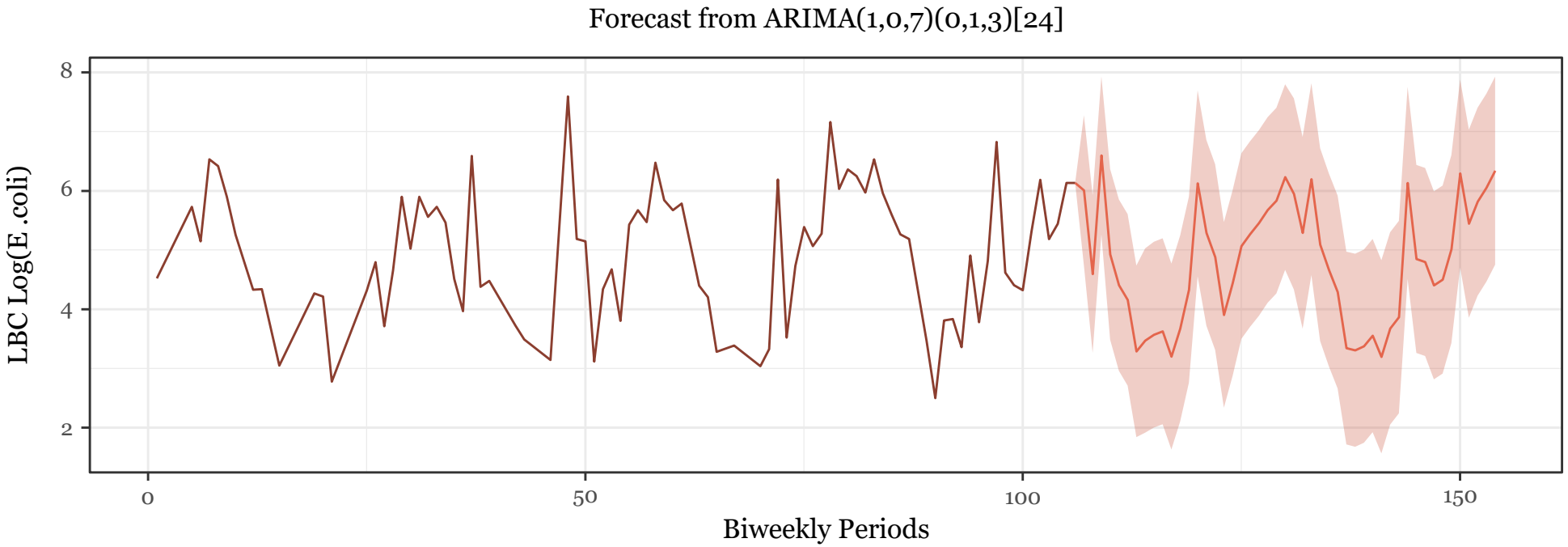


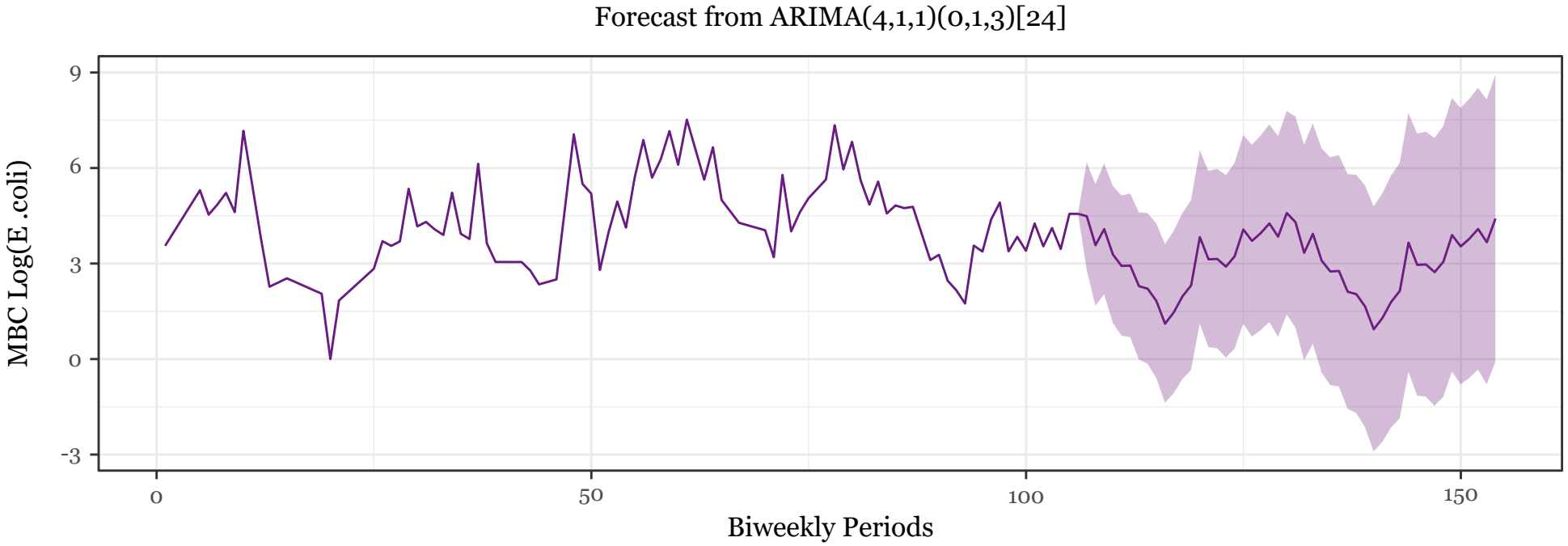

Temperature(C) vs. Time
Temp(C)
0
5
10
15
20
0
25
50
75
100
Biweekly Periods
Imputed E .coli
3
4
5
6
7
0
25
50
75
100
Biweekly Periods


Forecast from ARIMA(4,1,1)(0,1,3)[24]
LBC Log(E .coli)
8
6
4
2
0
50
100
150
Biweekly Periods


Forecast from ARIMA(4,1,1)(0,1,3)[24]
LBC Log(E .coli)
8
6
4
2
0
50
100
150
Biweekly Periods


Forecast from ARIMA(4,1,1)(0,1,3)[24]
MBC Log(E .coli)
9
6
3
0
-3
0
50
100
150
Biweekly Periods


Forecast from ARIMA(1,0,7)(0,1,3)[24]
LBC Log(E .coli)
8
6
4
2
0
50
100
150
Biweekly Periods


Forecast from ARIMA(4,1,1)(0,1,3)[24]
MBC Log(E .coli)
9
6
3
0
-3
0
50
100
150
Biweekly Periods

Some Implications

Final Thoughts/Inferences

- Seasonality appears to be largest predictor of E. coli levels
- Weather data should be monitored in closer time intervals in future to observe runoff effects
- Confounds: What caused occasional random high counts of E. coli in cold temperatures, particularly in 2015 and 2016?

Can we answer the questions we wanted to answer?

- Dataset may not be amenable to finding causes of E. coli in stream
- E. coli may or may not be related to human activity

Data Swamped by Seasonality?

- Difficulties with inferences about:
  - E. coli and human sources
  - Non-seasonal predictors of E. coli
- Focus could be on local effects
  - Monitor local behavior using control site with little human presence

References

- Hogg, R. V., & Craig, A. T. (1995). *Introduction to mathematical statistics.(7th edition)* (pp. 269-278). Upper Saddle River, New Jersey: Prentice Hall.