Exploratory markdown

Hannah Haskell

4/15/2021

Libraries

`tidyverse`

`lubridate`

`modelr`

`broom`

`knitr`

Data

A single data file will be uploaded for this exploratory analysis. The BIO539_mock_data includes physiological measurements of the oysters during the exposure trials as well as size measurements (shell length), moralities and the number of algae cells cleared over time during a post exposure clearance rate trial. All data to be analyzed has been made up for the sake of this assignment.

Basic Data Exploration

Rows: n = 48

Columns: n = 23

Column Names: Species, Tank (replicate #), Treatment, Fiber concentration, Animal ID #, CHO mj/mg, LiP mj/mg, PRT mj/mg, Ea mj/mg, Ec mj/mg, CEA mj/mg/min, Shell width (mm), X13, X14, Treatment a, Fiber concentration a, Initial no.s, End of expt. mortality, X19, Treatment b, Fiber concentration b, Time, Algal cell concentration

After looking over the column names it appears that the CEA data seems a bit disorganized. It looks like there are some spaces left between columns that should be removed and multiple different types of data included in the single file. I think it is worth while to assign a few variables to help separate and organize all of the data for subsequent analyses.

New variables created to help organize initial data:

Cleaned_CEA_data <- contains only the CEA data (i.e carbohydrates, lipids, proteins, energy consumed, energy available, cellular energy allocated (mj/mg))

Mortality <- contains mortality data from the duration of the experiment

Growth <- contains data on oyster growth throughout the experiment (i.e. shell lengths) for each of the experimental treatments

To begin the exploratory analysis, I am curious to see if there was any noticeable growth among the oysters over the duration of the experiment.

Treatment	Mean Shell Length (mm)
[Fiber 1]	54.7500
[Fiber 1] + PCB	54.6375
[Fiber 2]	54.8375
[Fiber 2] + PCB	48.7625
Baseline	46.7125
Control	54.6500

```
treatments <- subset (CEA_data, select = c(3,4)) %>%
  rename("Fiber_concentration" = "Fiber concentration") %>%
  group_by(Treatment) %>%
  summarise(Fiber_concentration = mean(Fiber_concentration)) %>%
  arrange(Fiber_concentration) %>%
  rename("Fibers/mL seawater" = "Fiber_concentration")

kable(treatments)
```

Treatment	Fibers/mL seawater
Baseline	0
Control	0
[Fiber 1]	1
[Fiber 1] + PCB	1
[Fiber 2]	10
[Fiber 2] + PCB	10

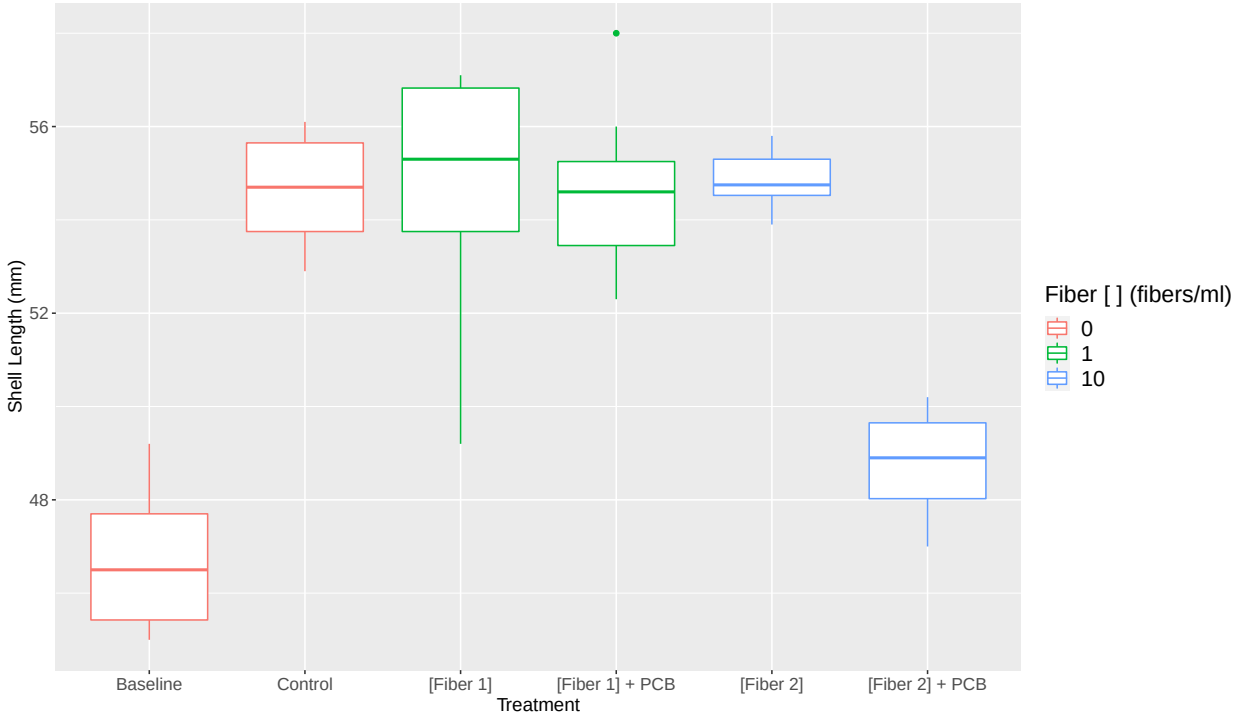


Figure 1. Shell length of oysters (n = 8 per treatment). Baseline measurements represent shell length at the beginning of the experiment. All other lengths were measured 3 months after the baseline measurements, at the end of the experiment. Colored outlines indicate the actual fiber concentration of each treatment (red = 0 fibers/ml, green = 1 fiber/ml, blue = 10 fibers/ml) Overall there appears to have been consistent growth among treatment groups with the exception of the [Fiber 2] + PCB group which had less growth overall.

I am now curious to see if the differences in growth between the baseline and the various treatments are actually significant, especially the [Fiber 2] + PCB group. To determine this I will assess for normality and then preform an ANOVA as well as a Tukey significance test.

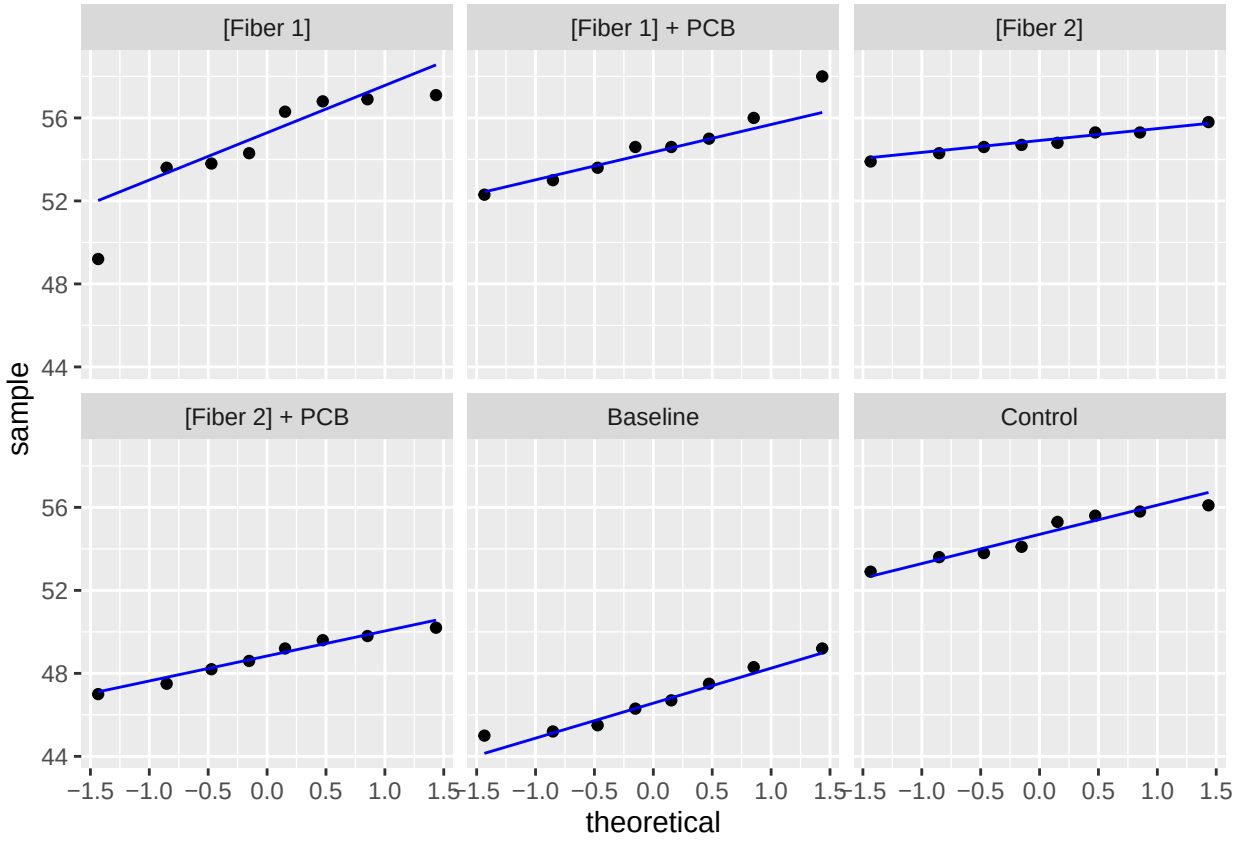


Figure 2. Q-Q plot to assess the normality of the growth data among each of the treatment groups. Each treatment group appears to follow a normal distribution with the exception of the [Fiber 1] treatment group.

Contrast	p value
Baseline-[Fiber 1]	0.0000000
Baseline-[Fiber 1] + PCB	0.0000000
Baseline-[Fiber 2]	0.0000000
Baseline-[Fiber 2] + PCB	0.1385803
Control-Baseline	0.0000000

Since analysis of the Q-Q plots appears to show that the majority of the growth data is normally distributed (with the exception of [Fiber 1]) I plan to assume normality and proceed with the ANOVA analysis for the sake of this project.

Results from the ANOVA test indicate that there is a significant difference in growth among the experimental treatments ($p = 4.409318 \times 10^{-15}$).

A Tukey test was then preformed on the growth data to determine where this significant difference is within the data. Results from the Tukey test indicate that differences between all treatments and the baseline are significant ($p < 0.05$ for each treatment) with the exception of [Fiber 2] + PCB group ($p = 0.139$).

These results tell me that there was significantly less growth overall in the oysters exposed to the [Fiber 2] + PCB treatment compared to each of the other experimental treatments. This may likely have to do with the increased presence of PCBs in the tank as a result of the higher concentration of the PCB laced fibers.

After analyzing growth, I then wanted to look into the mortality of oysters within each of the experimental treatments. For this analysis, a simple bar chart will likely be sufficient to visualize differences between treatments.

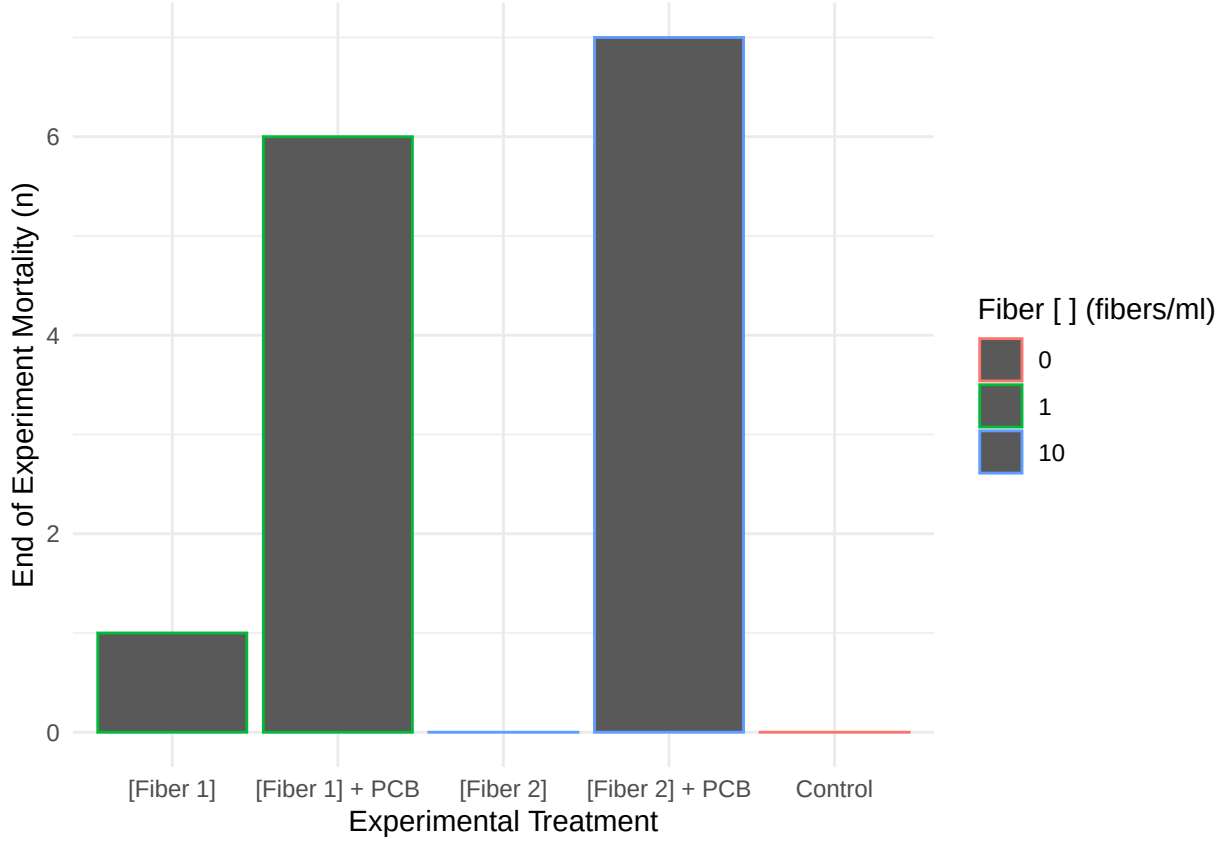


Figure 3. End of experiment mortality for each of the 5 experimental treatments. Experimental treatments that contained fibers laced with PCBs experienced higher moralities than those exposed to fibers only. There were no moralities within the control group over the duration of the experiment.

#Because this data is made up of single counts (1 number for each treatment group) I am unable to determine whether or not the differences between mortality are statisticaly different from one another via statistican significance testing.

When oysters were exposed to PCBs the number of moralities increased regardless of the fiber concentration. That being said, the [Fiber 2] + PCB had the highest number of moralities ($n = 7$)

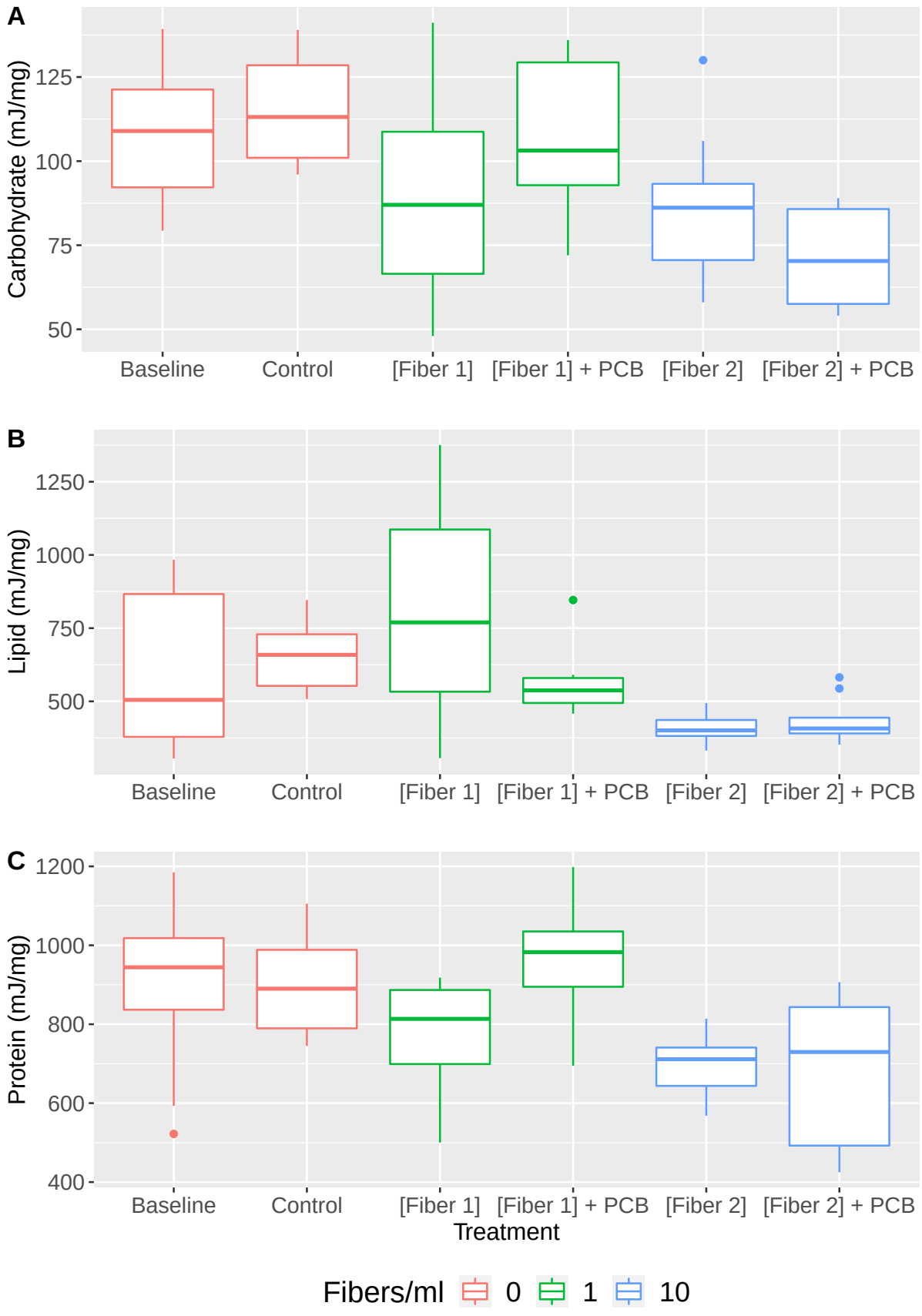
Now that I have assessed what growth and moralities looked like across treatments I am now curious to see what the cellular energy allocation data looks like. I will first examine how the individual physiological measurements such as lipids, carbohydrates, and proteins all vary between treatments and how they compare to the baseline values.

Figure 4. Total energy from cellular carbohydrate content (mJ/mg) extracted from whole body *C. virginica* tissue samples.

Figure 5. Total energy from cellular lipid content (mJ/mg) extracted from whole body *C. virginica* tissue samples.

Figure 6. Total energy from cellular protein content (mJ/mg) extracted from whole body *C. virginica* tissue samples.

Next, I am curious to see if the differences in physiological measurements between the baseline and the various treatments are significant. To determine this I will assess for normality and then preform an ANOVA as well as a Tukey significance test.


A
Carbohydrate (mJ/mg)
125
100
75
50
Baseline
Control
[Fiber 1]
[Fiber 1] + PCB
[Fiber 2]
[Fiber 2] + PCB
B
Lipid (mJ/mg)
1250
1000
750
500
Baseline
Control
[Fiber 1]
[Fiber 1] + PCB
[Fiber 2]
[Fiber 2] + PCB
C
Protein (mJ/mg)
1200
1000
800
600
400
Baseline
Control
[Fiber 1]
[Fiber 1] + PCB
[Fiber 2]
[Fiber 2] + PCB
Treatment
Fibers/ml
0
1
10