

Initial Data Analysis

Joan Moreaux and Meredith Miller

07/11/2021

```
library(performance)
library(DHARMA)
library(mgcv)
library(fitdistrplus)
library(goft)
library(gamlss)
library(FSA)
library(fGarch)
library(LambertW)
library(ordinal)
library(cowplot)
library(tidyverse)
```

```
health_data <- read.csv("Big_Data_Green.csv")
heat_data <- read.csv("Heat_Data.csv")
```

Health Data

Includes feeding time, PAM measurements, base measurements, and symbiont density

```
health_data <- health_data %>%
  mutate(Fv_Fm_1 = as.numeric(Fv_Fm_1), PAM_avg = ((Fv_Fm_1 + Fv_Fm_2 + Fv_Fm_3)/3)) %>%
  mutate(Base_Diameter_mm = (Base_Width + Base_Length + Base_Diagonal)/3) %>%
  mutate(Green_Density = (Green_Cells/1)*(1/0.1)*(1/0.001)*(0.5/Weight_Tentacle_g)) %>%
  mutate(Dino_Density = (Dino_Cells/1)*(1/0.1)*(1/0.001)*(0.5/Weight_Tentacle_g)) %>%
  mutate(Bucket = as.factor(Bucket), Treatment = as.factor(Treatment), Event = as.factor(Event),
         Species_ID = as.factor(Species_ID), Field_Site = as.factor(Field_Site)) %>%
  filter(Species_ID != "G32B" & Species_ID != "G33B")
```

PAM data

```
shapiro.test(health_data$PAM_avg) # p-value < 0.05, distribution is not normal
```

```
##
## Shapiro-Wilk normality test
##
## data: health_data$PAM_avg
## W = 0.84407, p-value < 2.2e-16
```

```
bartlett.test(PAM_avg ~ Treatment, data = health_data) # p-value < 0.05, does not meet equal variance a
```

```
##  
## Bartlett test of homogeneity of variances  
##  
## data: PAM_avg by Treatment  
## Bartlett's K-squared = 65.813, df = 2, p-value = 5.116e-15
```

Checking distribution: If we get a p-value < 0.5, then they are significantly different and we should not test. Since we cannot use any that we learned in class, we are using fitDist to check many distributions at once. fitDist showed that the SHASH distribution fits the best with our data, so we can use histDist to visualize it.

histDist Graph

```
exp_test(health_data$PAM_avg) #p-value < 2.2e-16, cannot use
```

```
##  
## Test for exponentiality based on a transformation to uniformity  
##  
## data: health_data$PAM_avg  
## T = -16.067, p-value < 2.2e-16
```

```
gamma_test(health_data$PAM_avg) #p-value < 2.2e-16, cannot use
```

```
##  
## Test of fit for the Gamma distribution  
##  
## data: health_data$PAM_avg  
## V = -14.501, p-value < 2.2e-16
```

```
lnorm_test(health_data$PAM_avg) #p-value < 2.2e-16, cannot use
```

```
##  
## Test for the lognormal distribution based on a transformation to  
## normality  
##  
## data: health_data$PAM_avg  
## p-value < 2.2e-16
```

```
normal_test(health_data$PAM_avg) #p-value = 5.414e-08, cannot use
```

```
##  
## Correlation test for normality  
##  
## data: health_data$PAM_avg  
## R = 0.96843, p-value = 3.468e-08  
## alternative hypothesis: health_data$PAM_avg does not follow a normal distribution.
```

```
weibull_test(health_data$PAM_avg)      #p-value < 2.2e-16, cannot use
```

```
##  
## Test for the Weibull distribution  
##  
## data:  health_data$PAM_avg  
## p-value < 2.2e-16
```

```
#Look for coral resiliancy Fv/Fm
```

```
PAM_visual <- fitDist(PAM_avg, data = health_data, type = "realAll", try.gamlss = T)
```

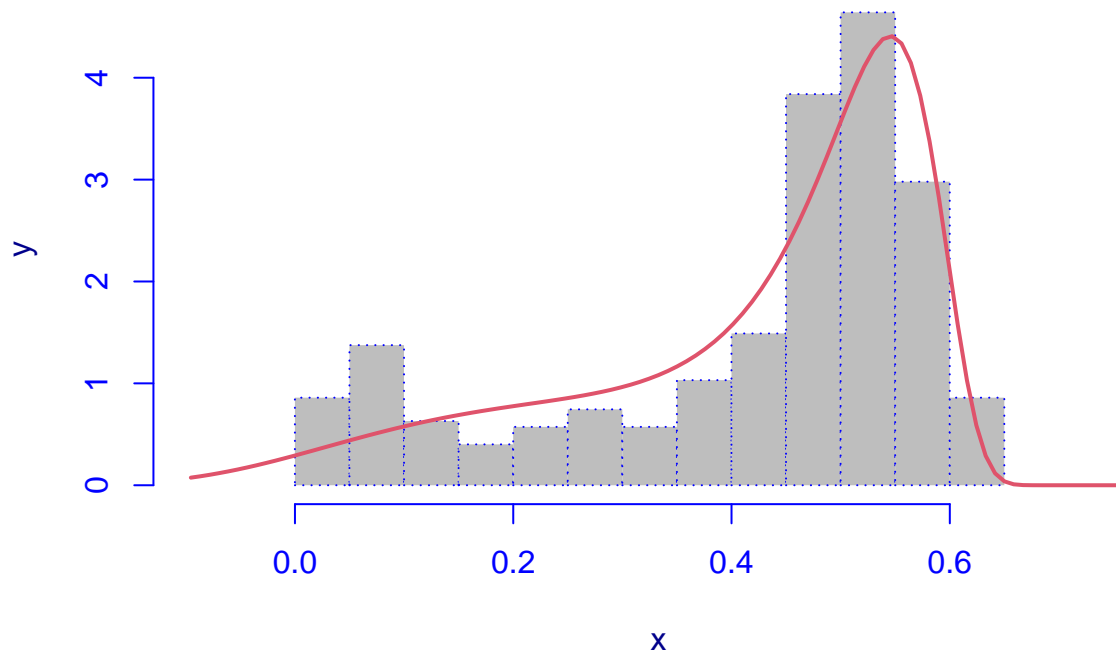
```
## |
```

```
# Family:  c("SHASH", "Sinh-Arcsinh"), Fitting method: "nlminb" with 12% error  
# Call:   gamlssML(formula = y, family = DIST[i])
```

```
#Visualizing fistDist
```

```
histDist(y, family = SHASH, density = FALSE,  
         nbins = 10, xlab = "x", ylab = "y", data = PAM_visual,  
         col.hist = "gray", border.hist = "blue",  
         fg.hist = rainbow(12)[9])
```

The y and the fitted SHASH distribution



```
##
## Family: c("SHASH", "Sinh-Arcsinh")
## Fitting method: "nlminb"
##
## Call:  gamlssML(formula = y, family = "SHASH", data = PAM_visual)
##
## Mu Coefficients:
## [1]  0.4756
## Sigma Coefficients:
## [1]  8.707
## Nu Coefficients:
## [1]  9.802
## Tau Coefficients:
## [1] 11.16
##
## Degrees of Freedom for the fit: 4 Residual Deg. of Freedom   345
## Global Deviance:      -467.534
##           AIC:        -459.534
##           SBC:        -444.114
```

#Modelling fitDist

```
SHASH_data <- health_data %>%
  select(-c(Event, Field_Site, Acclimation_Period, Base_Width, Base_Length, Base_Diagonal, Base_Diameter))
drop_na(PAM_avg) %>%
```

```

drop_na(Event_True) %>%
drop_na(Time_Point) %>%
filter(Species_ID != "G32B" & Species_ID != "G33B")

SHASH_data$orderTreatment = ordered(SHASH_data$Treatment, levels = c("Control", "25C", "30C"))

SHASH_model <- gamlss(formula = PAM_avg ~ Time_Point*Treatment + random(Species_ID), family = SHASH(), data = SHASH_data)

## GAMLSS-RS iteration 1: Global Deviance = -977.4286
## GAMLSS-RS iteration 2: Global Deviance = -985.8937
## GAMLSS-RS iteration 3: Global Deviance = -986.6611
## GAMLSS-RS iteration 4: Global Deviance = -986.9623
## GAMLSS-RS iteration 5: Global Deviance = -987.11
## GAMLSS-RS iteration 6: Global Deviance = -987.3525
## GAMLSS-RS iteration 7: Global Deviance = -987.5675
## GAMLSS-RS iteration 8: Global Deviance = -987.6751
## GAMLSS-RS iteration 9: Global Deviance = -987.7842
## GAMLSS-RS iteration 10: Global Deviance = -987.9023
## GAMLSS-RS iteration 11: Global Deviance = -987.9715
## GAMLSS-RS iteration 12: Global Deviance = -988.0976
## GAMLSS-RS iteration 13: Global Deviance = -988.2519
## GAMLSS-RS iteration 14: Global Deviance = -988.3885
## GAMLSS-RS iteration 15: Global Deviance = -988.4503
## GAMLSS-RS iteration 16: Global Deviance = -988.5316
## GAMLSS-RS iteration 17: Global Deviance = -988.5784
## GAMLSS-RS iteration 18: Global Deviance = -988.636
## GAMLSS-RS iteration 19: Global Deviance = -988.6987
## GAMLSS-RS iteration 20: Global Deviance = -988.7595
## GAMLSS-RS iteration 21: Global Deviance = -988.8494
## GAMLSS-RS iteration 22: Global Deviance = -988.9322
## GAMLSS-RS iteration 23: Global Deviance = -988.9752
## GAMLSS-RS iteration 24: Global Deviance = -989.0269
## GAMLSS-RS iteration 25: Global Deviance = -989.0594
## GAMLSS-RS iteration 26: Global Deviance = -989.1339
## GAMLSS-RS iteration 27: Global Deviance = -989.1797
## GAMLSS-RS iteration 28: Global Deviance = -989.2367
## GAMLSS-RS iteration 29: Global Deviance = -989.2373

summary(SHASH_model)

## *****
## Family:  c("SHASH", "Sinh-Arcsinh")
##
## Call:  gamlss(formula = PAM_avg ~ Time_Point * Treatment +
##      random(Species_ID), family = SHASH(), data = SHASH_data,
##      control = gamlss.control(n.cyc = 200))
##
## Fitting method: RS()
##
## -----
## Mu link function:  identity
## Mu Coefficients:

```

```

##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.467240   0.007285  64.136   <2e-16 ***
## Time_Point       0.002956   0.001270   2.328   0.0206 *
## Treatment25C     -0.158368   0.009339 -16.958   <2e-16 ***
## Treatment30C     -0.017513   0.009355  -1.872   0.0623 .
## Time_Point:Treatment25C -0.001190   0.001675  -0.711   0.4780
## Time_Point:Treatment30C -0.002611   0.001736  -1.504   0.1337
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -----
## Sigma link function: log
## Sigma Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.3197      0.1858  -23.25   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -----
## Nu link function: log
## Nu Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.70173     0.08273  -8.483 1.33e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -----
## Tau link function: log
## Tau Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.78083     0.07791  -10.02   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -----
## NOTE: Additive smoothing terms exist in the formulas:
## i) Std. Error for smoothers are for the linear effect only.
## ii) Std. Error for the linear terms maybe are not accurate.
## -----
## No. of observations in the fit: 323
## Degrees of Freedom for the fit: 45.69826
##      Residual Deg. of Freedom: 277.3017
##              at cycle: 29
##
## Global Deviance: -989.2373
##      AIC: -897.8408
##      SBC: -725.2086
## *****

```

```

#               Estimate Std. Error t value Pr(>|t|)
#(Intercept)      0.467240   0.007285  64.136   <2e-16 ***
#Time_Point       0.002956   0.001270   2.328   0.0206 *
#Treatment25C     -0.158368   0.009339 -16.958   <2e-16 ***
#Treatment30C     -0.017513   0.009355  -1.872   0.0623 .

```

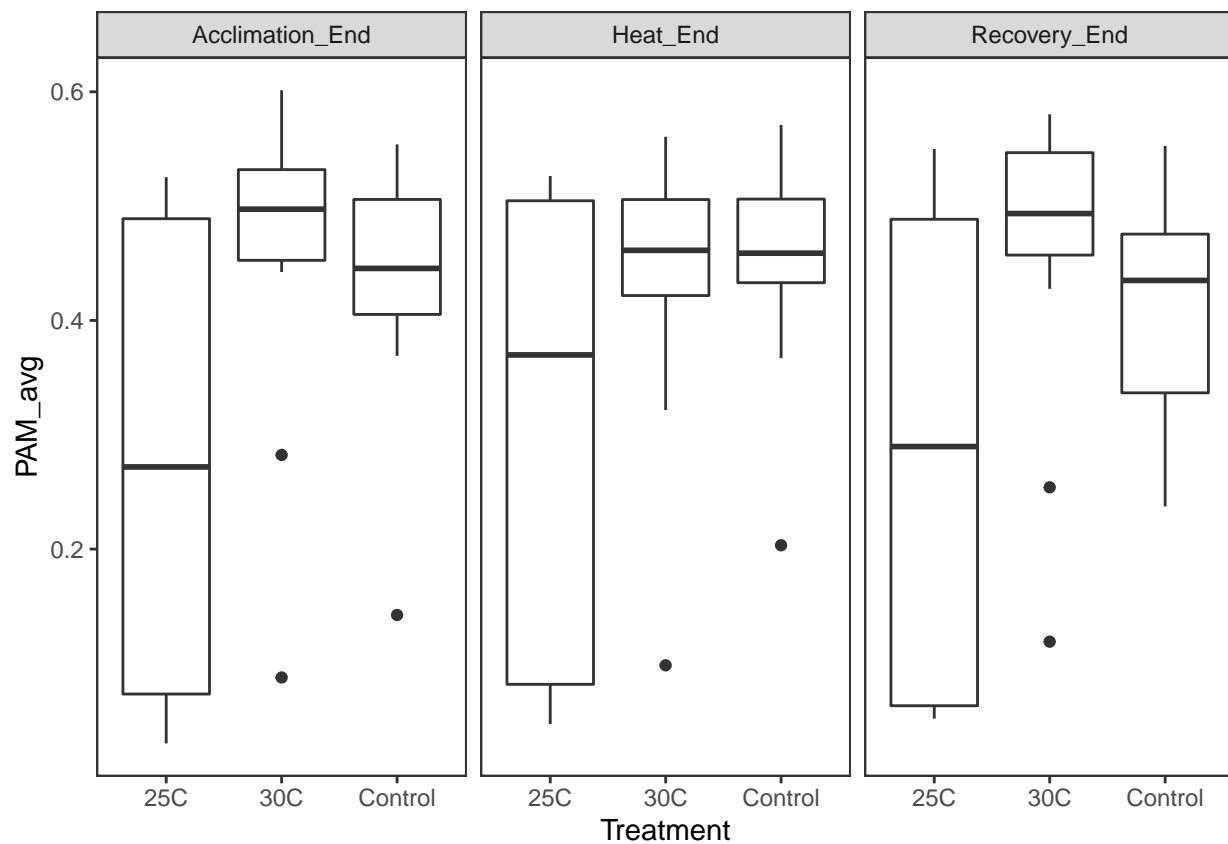
```
#Time_Point:Treatment25C -0.001190 0.001675 -0.711 0.4780
#Time_Point:Treatment30C -0.002611 0.001736 -1.504 0.1337
```

#Time point and 25C is significant, 30C is not
#plot avg fu/fm over all time points (this was significantly different over time)

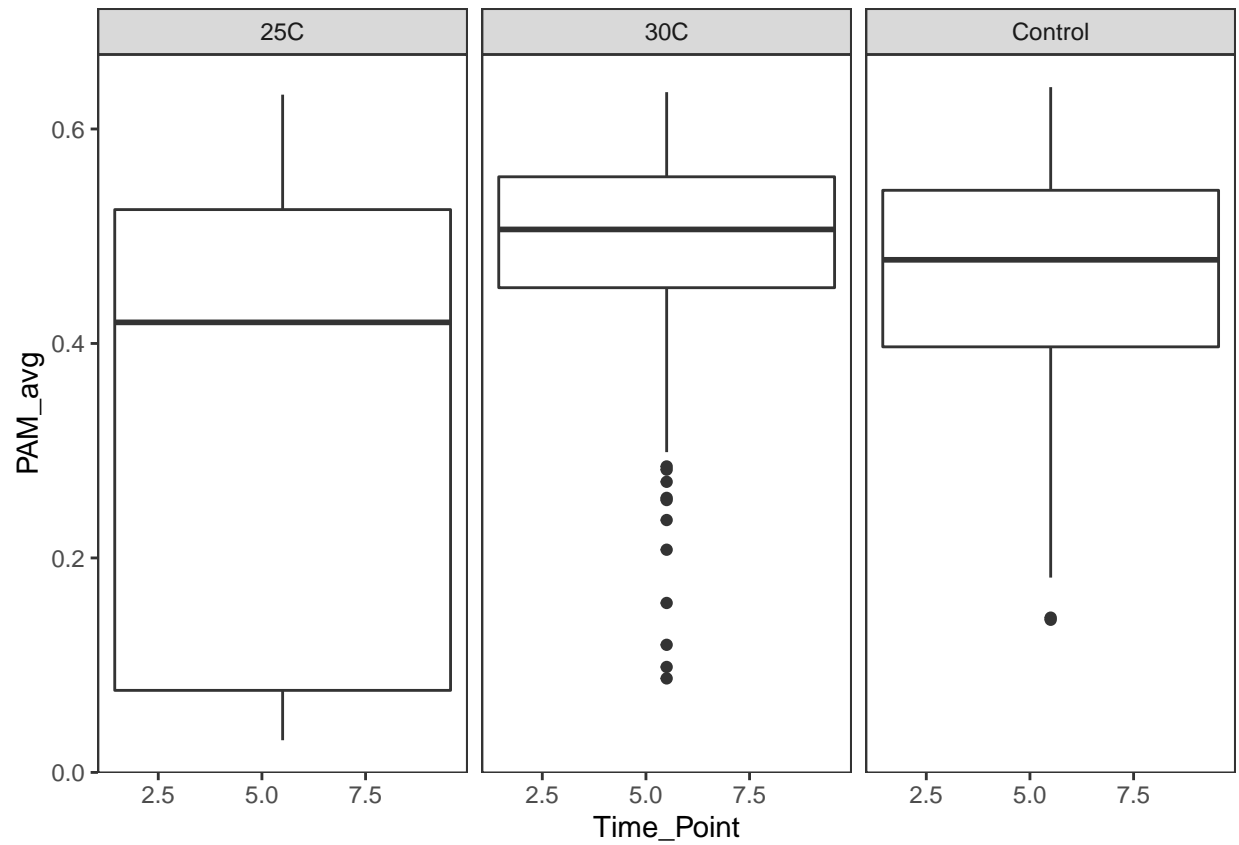
PAM Graph We see no significant differences between treatments at different times of the experiment.

```
PAM_data_plot <- health_data %>%
  filter(Event_True == "Acclimation_End" | Event_True == "Heat_End" | Event_True == "Recovery_End")

ggplot(PAM_data_plot, aes(x = Treatment, y = PAM_avg)) +
  geom_boxplot() +
  facet_grid(. ~ Event_True) +
  theme_test()
```

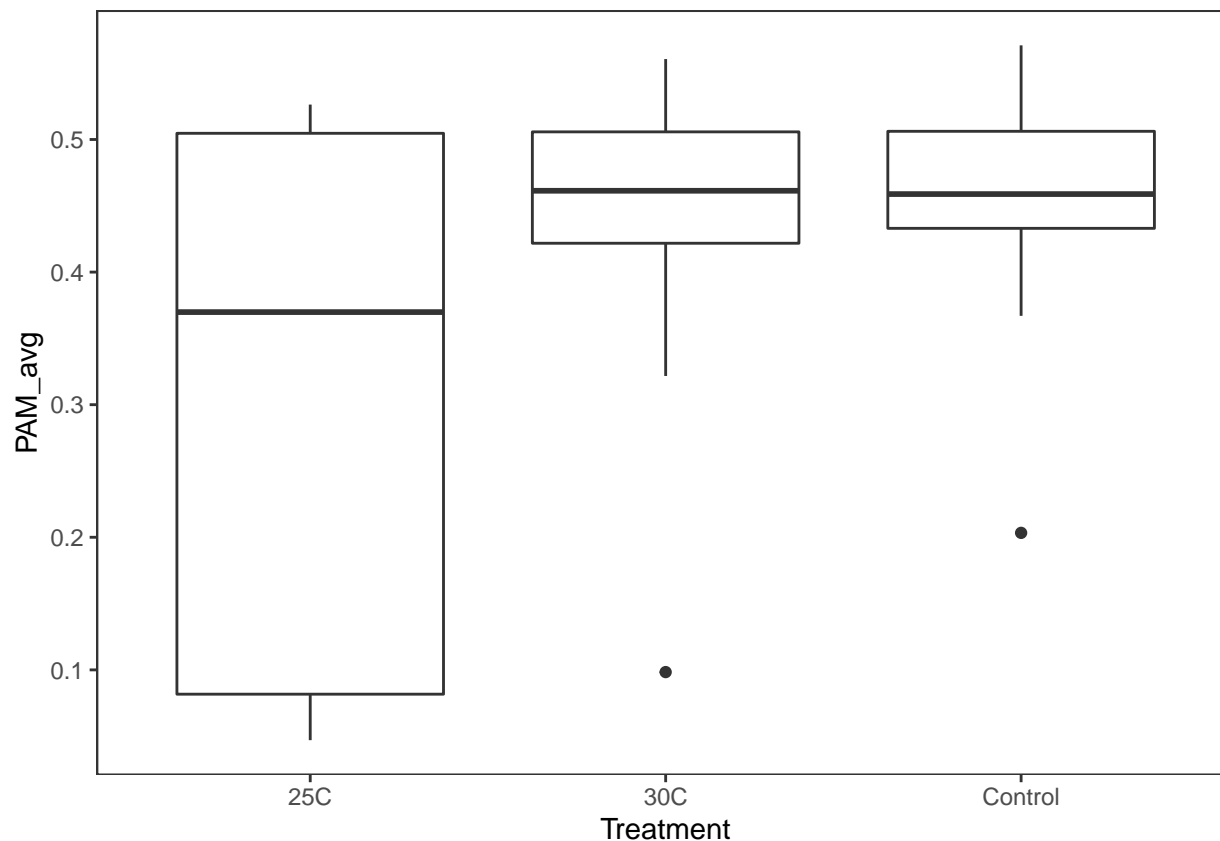


```
ggplot(SHASH_data, aes(x = Time_Point, y = PAM_avg)) +
  geom_boxplot() +
  facet_grid(. ~ Treatment) +
  theme_test()
```



```
PAM_data_end <- health_data %>%  
  filter(Event_True == "Heat_End") %>%  
  drop_na(PAM_avg)
```

```
ggplot(PAM_data_end, aes(x = Treatment, y = PAM_avg)) +  
  geom_boxplot() +  
  theme_test()
```

Symbiont Data

```
filtered_dates <- health_data %>%
  filter(Date == "11/5/2021" | Date == "11/9/2021" | Date == "11/13/2021")

shapiro.test(filtered_dates$Green_Density)      #p-value < 0.05, not normal distribution
```

```
##
## Shapiro-Wilk normality test
##
## data: filtered_dates$Green_Density
## W = 0.72047, p-value = 1.303e-07
```

```
bartlett.test(Green_Density ~ Treatment, filtered_dates)      #p-value < 0.05, does not meet assumption
```

```
##
## Bartlett test of homogeneity of variances
##
## data: Green_Density by Treatment
## Bartlett's K-squared = 8.9514, df = 2, p-value = 0.01138
```

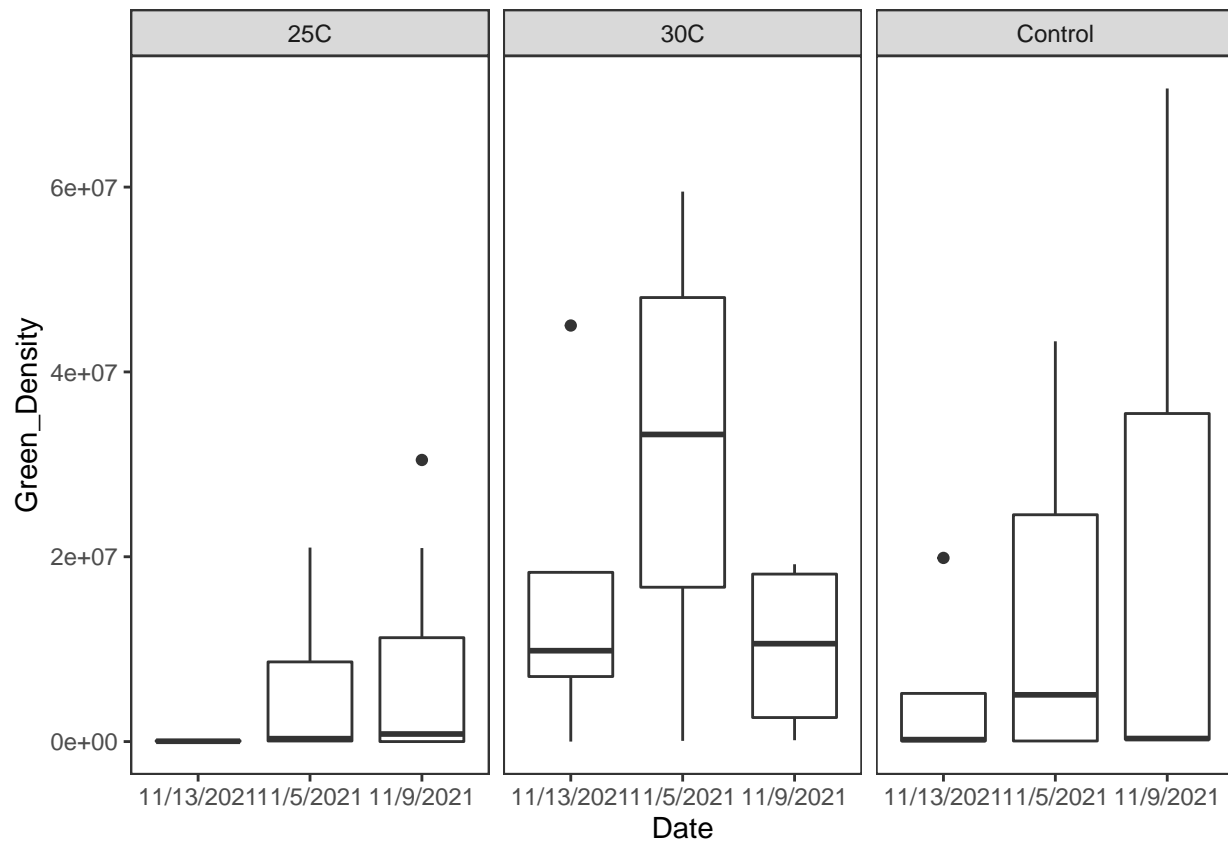
```
shapiro.test(filtered_dates$Dino_Density)      #p-value < 0.05, not normal distribution
```

```
##
## Shapiro-Wilk normality test
##
## data: filtered_dates$Dino_Density
## W = 0.59253, p-value = 1.379e-09
```

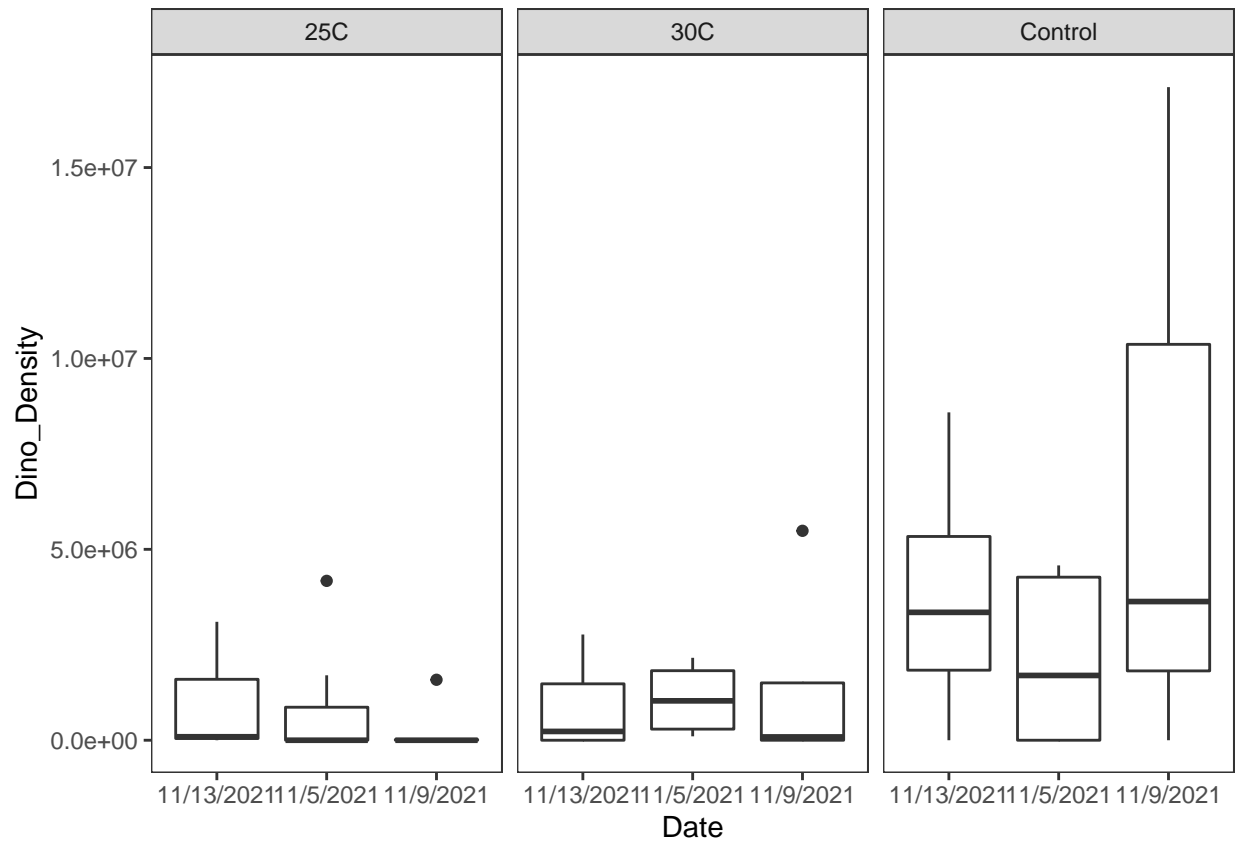
```
bartlett.test(Dino_Density ~ Treatment, filtered_dates)      #p-value < 0.05, does not meet assumption
```

```
##
## Bartlett test of homogeneity of variances
##
## data: Dino_Density by Treatment
## Bartlett's K-squared = 27.188, df = 2, p-value = 1.248e-06
```

```
ggplot(filtered_dates, aes(x = Date, y = Green_Density)) +
  geom_boxplot() +
  facet_grid(. ~ Treatment) +
  theme_test()
```



```
ggplot(filtered_dates, aes(x = Date, y = Dino_Density)) +
  geom_boxplot() +
  facet_grid(. ~ Treatment) +
  theme_test()
```



Size Data

```
shapiro.test(filtered_dates$Base_Diameter_mm) # p-value = 0.001, distribution is not normal
```

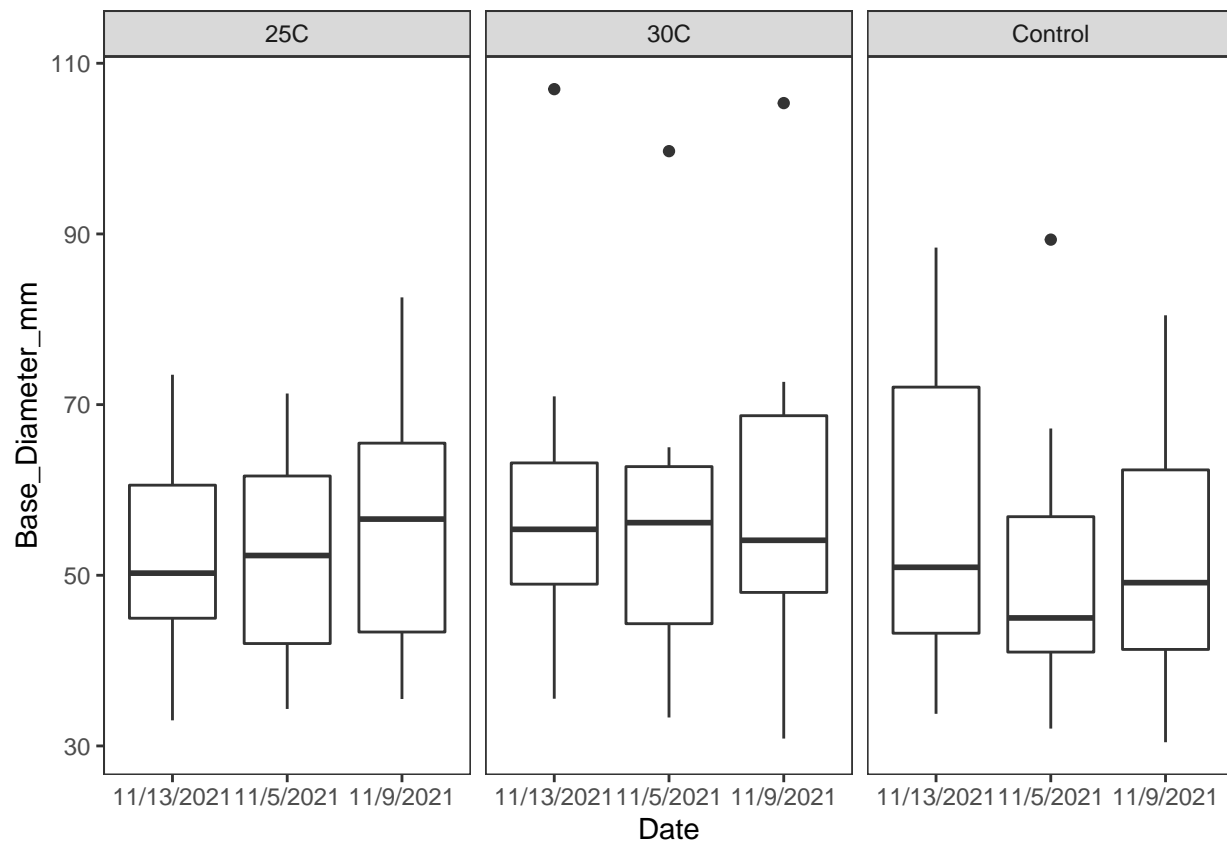
```
##
## Shapiro-Wilk normality test
##
## data: filtered_dates$Base_Diameter_mm
## W = 0.94517, p-value = 0.0002429
```

```
bartlett.test(Base_Diameter_mm ~ Treatment, data = filtered_dates) # p-value = 0.30, yay equal variance
```

```
##
## Bartlett test of homogeneity of variances
##
## data: Base_Diameter_mm by Treatment
## Bartlett's K-squared = 3.8061, df = 2, p-value = 0.1491
```

Size Graph

```
ggplot(filtered_dates, aes(x = Date, y = Base_Diameter_mm)) +
  geom_boxplot() +
  facet_grid(. ~ Treatment) +
  theme_test()
```



Feeding Data

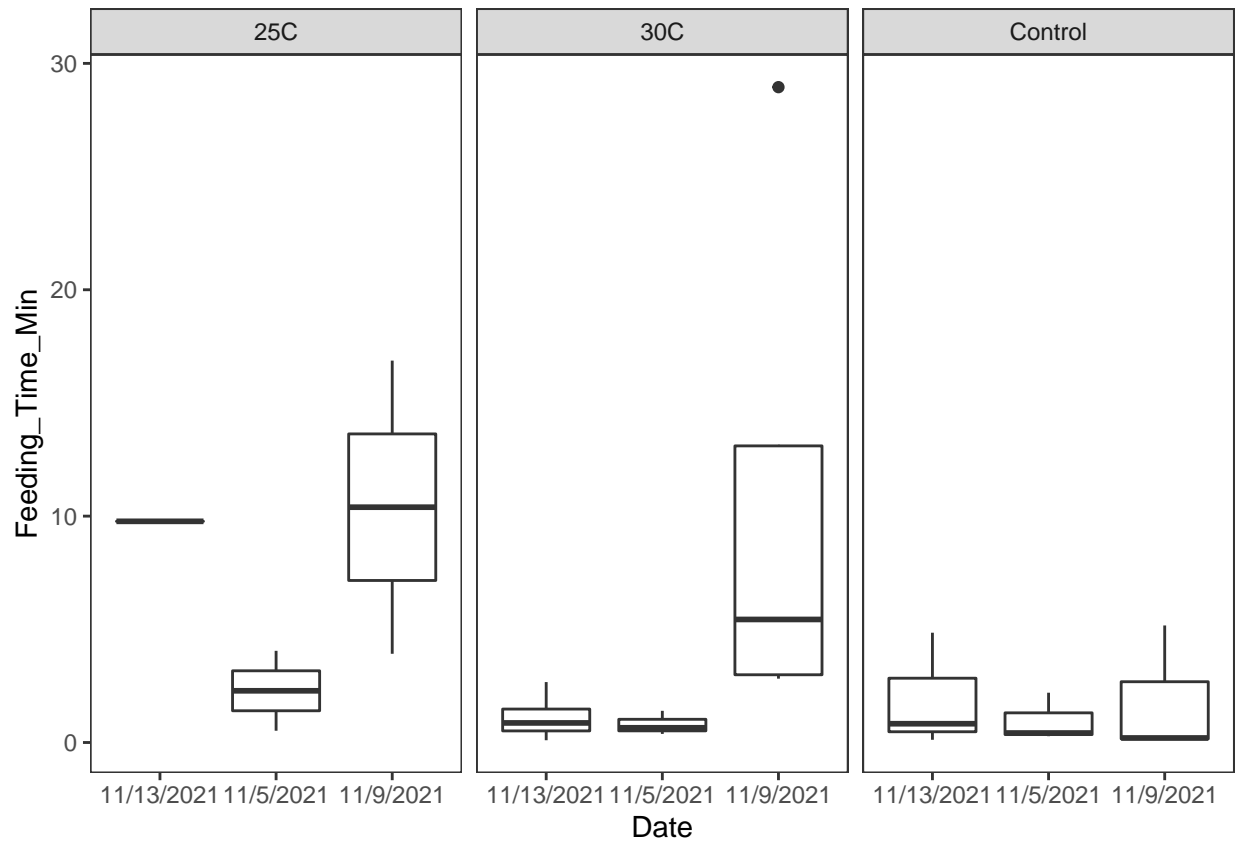
```
shapiro.test(filtered_dates$Feeding_Time_Min)      #p-value < 0.05, not normal distribution
```

```
##
##  Shapiro-Wilk normality test
##
## data:  filtered_dates$Feeding_Time_Min
## W = 0.61603, p-value = 6.509e-07
```

```
bartlett.test(Feeding_Time_Min ~ Treatment, filtered_dates)      #p-value < 0.01, does not meet assumption
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  Feeding_Time_Min by Treatment
## Bartlett's K-squared = 11.934, df = 2, p-value = 0.002562
```

```
ggplot(filtered_dates, aes(x = Date, y = Feeding_Time_Min)) +
  geom_boxplot() +
  facet_grid(. ~ Treatment) +
  theme_test()
```



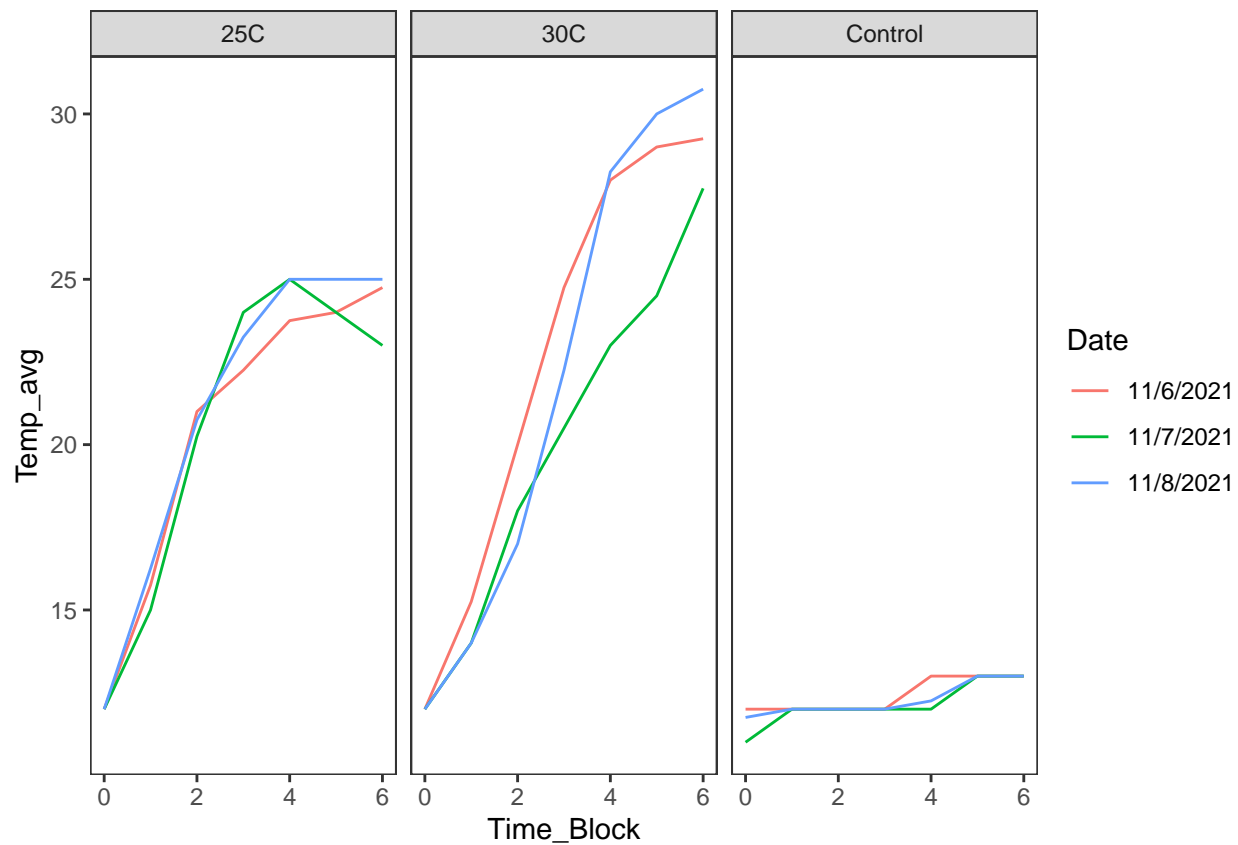
Heat Data

Includes temperature during heatwave and behavioral responses (open vs closed)

```
heat_data <- heat_data %>%
  mutate(Bucket = as.factor(Bucket), Treatment = as.factor(Treatment), Field_Site = as.factor(Field_Site))
  group_by(Date, Treatment, Time_Block) %>%
  mutate(Temp_avg = mean(Bucket_Temp))
```

Temperature over time Graph

```
ggplot(data = heat_data, aes(x = Time_Block, y = Temp_avg, color = Date)) +
  geom_line() +
  facet_grid(. ~ Treatment) +
  theme_test()
```



Open/Closed Graph

```
open_closed_data <- heat_data %>%
  group_by(Day, Treatment, Time_Block) %>%
  count(Open_Closed)

ggplot(data = open_closed_data, aes(x = Time_Block, y = n, fill = Open_Closed)) +
  geom_bar(position="fill", stat="identity") +
  facet_grid(. ~ Treatment) +
  scale_fill_brewer(palette = "Reds") +
  theme_cowplot()
```

