Homework 3

Odile Gabbiani

Table of contents

## Quarto

Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see <https://quarto.org>.

## Homework Setup

### Loading in Libraries

library(tidyverse)   
library(readxl)  
library(here)  
library(janitor)  
library(GGally)  
library(MuMIn)  
library(ggeffects)  
library(gtsummary)  
library(flextable)  
library(modelsummary)  
  
drought\_exp <- read\_xlsx(path = here("data",   
 "Valliere\_etal\_EcoApps\_Data.xlsx"), #reading in data from excel sheet   
 sheet = "First Harvest") #choosing which sheet to use  
  
str(drought\_exp) #taking a look at data

tibble [70 × 13] (S3: tbl\_df/tbl/data.frame)  
 $ Species : chr [1:70] "ENCCAL" "ENCCAL" "ENCCAL" "ENCCAL" ...  
 $ Water : chr [1:70] "WW" "WW" "WW" "WW" ...  
 $ Rep # : num [1:70] 1 2 3 4 5 1 2 3 4 5 ...  
 $ Height (cm) : num [1:70] 5.8 4.9 8.4 6.5 7.1 3.2 4.4 4.2 4.5 3.9 ...  
 $ Leaf # : num [1:70] 11 8 11 12 10 7 7 10 8 6 ...  
 $ Leaf dry weight (g): num [1:70] 0.0294 0.0185 0.0177 0.0178 0.0164 0.017 0.0193 0.0153 0.0159 0.0133 ...  
 $ Leaf area (cm2) : num [1:70] 5.01 3.98 3.69 3.84 3.63 3.06 3.1 2.94 2.73 2.61 ...  
 $ SLA : num [1:70] 170 215 209 216 222 ...  
 $ Total LA : num [1:70] 55.1 31.8 40.6 46.1 36.3 ...  
 $ Shoot (g) : num [1:70] 0.253 0.164 0.241 0.213 0.232 ...  
 $ Root (g) : num [1:70] 0.202 0.165 0.209 0.146 0.12 ...  
 $ Total (g) : num [1:70] 0.455 0.329 0.45 0.359 0.352 ...  
 $ R:S : num [1:70] 0.8 1 0.9 0.7 0.5 0.8 1.2 3.1 0.9 1.2 ...

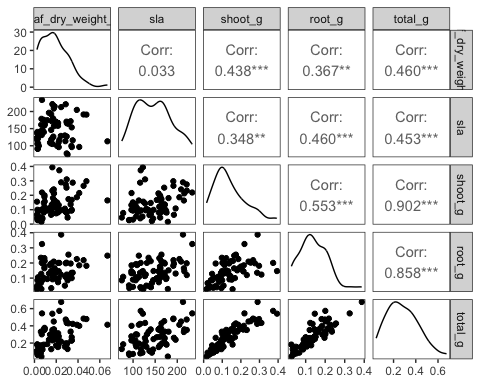
class(drought\_exp)

[1] "tbl\_df" "tbl" "data.frame"

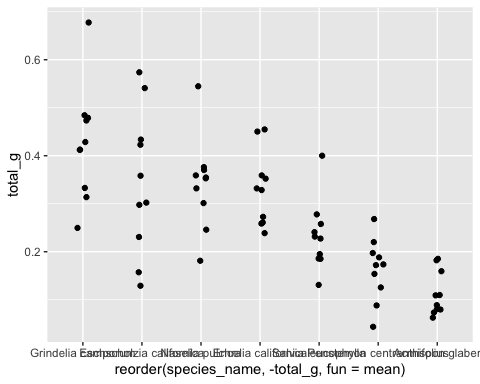
You can add options to executable code like this

drought\_exp\_clean <- drought\_exp %>%   
 clean\_names() %>% # nicer column names  
 mutate(species\_name = case\_when( # adding column with species scientific names  
 species == "ENCCAL" ~ "Encelia californica", # bush sunflower  
 species == "ESCCAL" ~ "Eschscholzia californica", # California poppy  
 species == "PENCEN" ~ "Penstemon centranthifolius", # Scarlet bugler  
 species == "GRICAM" ~ "Grindelia camporum", # great valley gumweed  
 species == "SALLEU" ~ "Salvia leucophylla", # Purple sage  
 species == "STIPUL" ~ "Nasella pulchra", # Purple needlegrass  
 species == "LOTSCO" ~ "Acmispon glaber" # deerweed  
 )) %>%   
 relocate(species\_name, .after = species) %>% # moving species\_name column after species  
 mutate(water\_treatment = case\_when( # adding column with full treatment names  
 water == "WW" ~ "Well watered",  
 water == "DS" ~ "Drought stressed"  
 )) %>%   
 relocate(water\_treatment, .after = water) # moving water\_treatment column after water

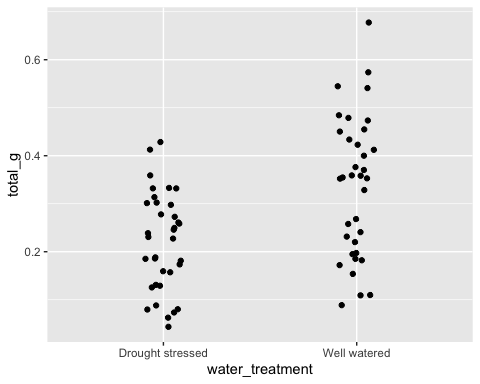
ggpairs(drought\_exp\_clean, # data frame  
 columns = c("leaf\_dry\_weight\_g", # columns to visualize  
 "sla",   
 "shoot\_g",   
 "root\_g",   
 "total\_g"),   
 upper = list(method = "pearson")) + # calculating Pearson correlation coefficient  
 theme\_bw() + # cleaner theme  
 theme(panel.grid = element\_blank()) # getting rid of gridlines



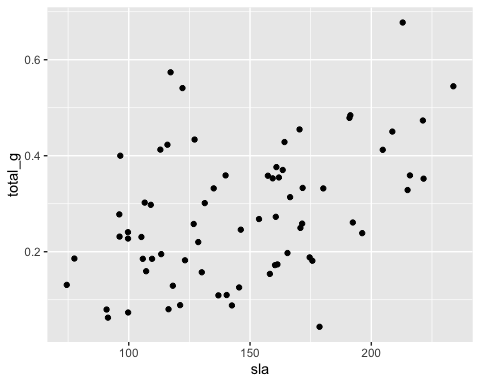
ggplot(data = drought\_exp\_clean, # data frame  
 aes(x = reorder(species\_name, # reordering x-axis  
 -total\_g, # in reverse order of mean total mass  
 fun = mean), # calculating mean to reorder  
 y = total\_g)) + # y-axis  
 geom\_jitter(width = 0.1, # narrow jitter  
 height = 0) # not jittering points up and down



ggplot(data = drought\_exp\_clean, # data frame  
 aes(x = water\_treatment, # x-axis  
 y = total\_g)) + # y-axis  
 geom\_jitter(width = 0.1, # narrow jitter  
 height = 0) # not jittering points up and down

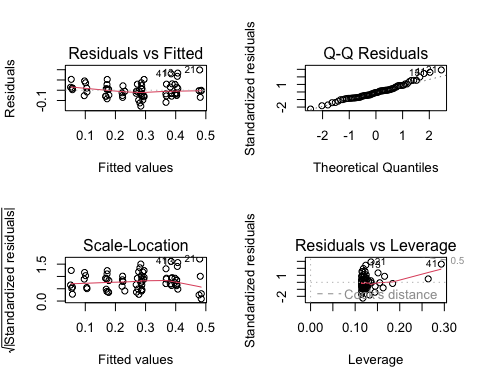


ggplot(data = drought\_exp\_clean, # data frame  
 aes(x = sla, # x-axis  
 y = total\_g)) + # y-axis  
 geom\_point() # scatterplot

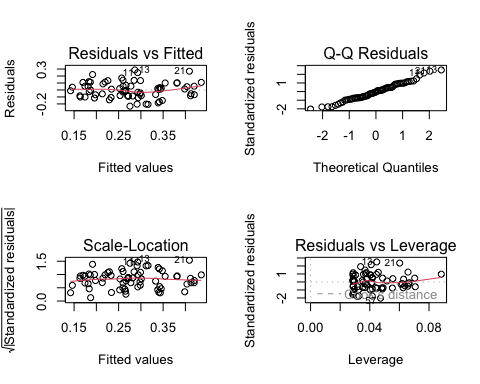


model0 <- lm(total\_g ~ 1, # formula  
 data = drought\_exp\_clean) # data frame

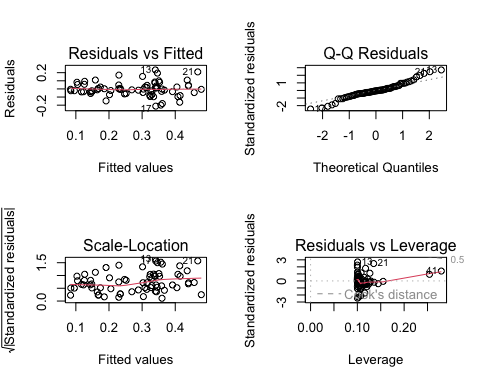
model1 <- lm(total\_g ~ sla + water\_treatment + species\_name, #formula   
 data = drought\_exp\_clean) #data frame   
  
#diagnsotics for model   
par(mfrow = c(2, 2))   
plot(model1)



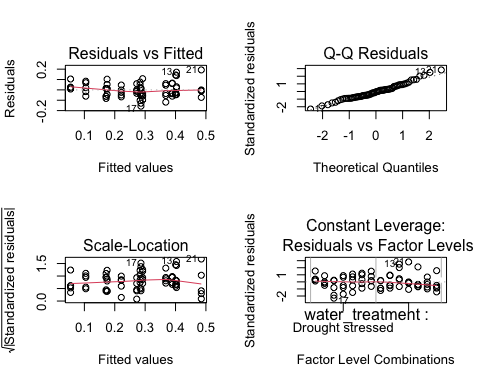
model2 <- lm(total\_g ~ sla + water\_treatment, #formula   
 data = drought\_exp\_clean) #data frame  
#model diagnsotics   
par(mfrow = c(2,2))  
plot(model2)



model3 <- lm(total\_g ~ sla + species\_name, #formula   
 data = drought\_exp\_clean) #data frame   
#model diagnsotics   
par(mfrow = c(2,2))  
plot(model3)



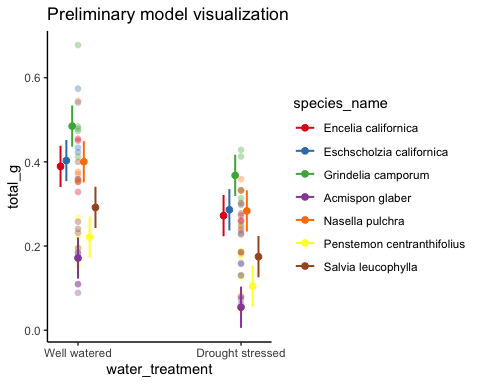
model4 <- lm(total\_g ~ water\_treatment + species\_name, #formula   
 data = drought\_exp\_clean) #data frame  
#model diagnostics   
par(mfrow = c(2,2))  
plot(model4)



#model selection for all four models   
model.sel <- model.sel(model0,  
 model1,   
 model2,   
 model3,   
 model4)  
#creating a flextable   
flextable::as\_flextable(model.sel)

| (Intercept) | sla | species\_name | water\_treatment | df | logLik | AICc | delta | weight |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| numeric | numeric | factor | factor | integer | numeric | numeric | numeric | model.weights |
| 0.1 |  | + | + | 9 | 88.6 | -156.2 | 0.0 | 0.8 |
| 0.1 | -0.0 | + | + | 10 | 88.7 | -153.8 | 2.4 | 0.2 |
| -0.0 | 0.0 | + |  | 9 | 72.5 | -124.1 | 32.1 | 0.0 |
| 0.0 | 0.0 |  | + | 4 | 52.2 | -95.8 | 60.4 | 0.0 |
| 0.3 |  |  |  | 2 | 39.6 | -75.0 | 81.2 | 0.0 |
| n: 5 | | | | | | | | |

model\_preds <- ggpredict(model4,   
 terms = c("water\_treatment",   
 "species\_name"))  
plot(model\_preds, # model predictions  
 limit\_range = TRUE, # limit the range of predictions to the range of predictor values  
 show\_data = TRUE) + # show the underlying data  
 # everything below this is ggplot() stuff  
 theme\_classic() + # classic theme  
 labs(title = "Preliminary model visualization") + # plot title  
 theme(panel.grid = element\_blank()) # getting rid of gridlines

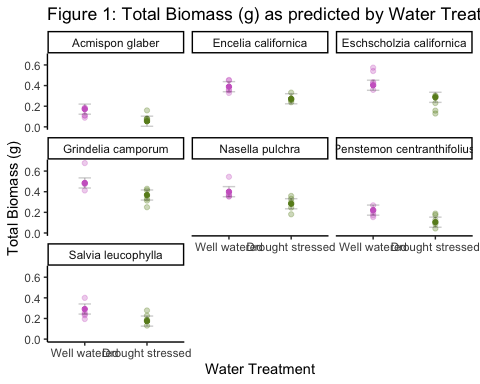


# Problem 1

## a. Table 1: Model Selection: This table lists the different types of models and the predictors associated with each one. It also lists the AIC and Delta AIC values for each model.

## b. To address the central question of whether biomass is influenced by specific leaf area, plant species, and/or water treatment, different models were assigned to a combination of variables to determine which model contained the best predictor variables for total biomass. Model 0 included none of the variables, model 1 included all of them, and models 2 through 4 contained a combination of two out of the three variables. The model selection function was used to determine which model was most representative or best explained the data collected over the course of the experiment. Through model selection, model 4, which contained the species type and water treatment, was determined to be the most precise model (AIC = -156.2, delta AIC = 0.00). Because the AIC value was the largest and the delta AIC was 0.00, model 4 was determined to best explain the data collected during the experiment. Diagnostics were taken to evaluate normality, homoscedasticity, and outliers.

model\_preds\_for\_plotting <- model\_preds %>% #assigning to new object   
 rename(water\_treatment = x, #renaming   
 species\_name = group)  
  
ggplot() +  
 # underlying data  
 geom\_point(data = drought\_exp\_clean, #adding points   
 alpha = 0.3, #tranparency   
 aes(x = water\_treatment, #assigning x axis   
 y = total\_g, #assigning y axis   
 color = water\_treatment)) + #assigning color   
geom\_errorbar(data = model\_preds\_for\_plotting, #adding errorbars   
 aes(x = water\_treatment, #assigning x   
 y = predicted, #assigning y   
 ymin = conf.low, #assigning y min   
 ymax = conf.high, #assigning y max   
 fill = water\_treatment), #assigning fill   
 alpha = 0.2, width = 0.2) +  
 geom\_point(data = model\_preds\_for\_plotting, #adding points   
 aes(x = water\_treatment, #assigning aesthetics   
 y = predicted,  
 color = water\_treatment)) +  
 # cleaner theme  
 theme\_classic() +  
 # creating different panels for species  
 facet\_wrap(~species\_name) +   
 theme(legend.position = "none") + #removing legend   
 labs(x = "Water Treatment", #adding labels   
 y = "Total Biomass (g)",   
 title = "Figure 1: Total Biomass (g) as predicted by Water Treatment and Plant Species") + #adding title   
 scale\_color\_manual(values = c("Well watered" = "orchid3", #assigning color  
 "Drought stressed" = "olivedrab4"))



## d. Figure 1: Total Biomass (g) as predicted by Water Treatment and Plant Species represents the response of different plant species to different water treatments (well watered and drough stressed). Bolded points represent model predictions by showing the relationship between different variables. Valliere, Justin; Zhang, Jacqueline; Sharifi, M.; Rundel, Philip (2019). Data from: Can we condition native plants to increase drought tolerance and improve restoration success? [Dataset]. Dryad. https://doi.org/10.5061/dryad.v0861f7

## e.

# Problem 2

## a. The data I took over the course of this class describes the number and species of birds seen perching on an artificial floating wetland island (FWI) on the UCSB Lagoon. An affectively visualization of my data could consist of birds sitting on pilings of different heights to represent how often each bird species was seen. Each piling would correspond to a different bird and that would be represented by having that specific species sitting atop it. While the pilings don’t represent exactly where the birds were seen perching, they do give an idea of how many birds of each species were seen.

## b.

## c.

## d. UCSB Lagoon Birds: this piece shows the number of bird species seen perching on an artificial floating wetland island (FWI) at the UCSB Lagoon over a period of five months, from January to May 2024. Each piling represents a different observed bird species and the piling’s heights represent how many of those birds were seen. This piece is a Caran D’ache color pencil drawing that also incorporates Archival Ink pens to outline certain aspects of the visualization. To make this, I first made a sketch of what I wanted it to look like, which I transferred onto paper. I found images of the six bird species seen on the UCSB Lagoon and I drew them on top of the pilings. I did everything in pencil first, then added color, and finished by adding touches of ink.

# Problem 3

## a. To address their main research question (are CFWs effective in removing N and P from aquatic systems?), the authors used a variety of tests including Levene’s Test, Kruskall Wallis Test, Spearman’s R Correlation Coefficient, and Linear Regression. The authors did not include a visualization of their data in the article but they do have tables listing data points and showing the relationship between different variables. This is the table they included for the Spearman’s R Correlation Coefficient.

## b. To visualize their data, I would suggest the authors use