# Homework 3

2024-09-27

Read the data in the chunk below. The data come from of a study that investigated the flood tolerance of three different clover species.

The data can be interpreted as arising from an experiment with a 5×2 factorial treatment structure in a completely randomized design with 24 repetitions. The experimental units were pots in the greenhouse. The researchers cultivated the pots under two growing conditions: one was normally irrigated (control treatment), and one was left 12 days under flooded conditions. After the flooding, the researchers measured the aboveground biomass (in grams). The objective of the study is to find out if the flooding treatment affects the plant biomass, and if this effect is different for different species.

```r
library(tidyverse)
```

```r
url <- "https://raw.githubusercontent.com/jlacasa/stat705_fall2024/main/classes/data/lotus_hw3.csv"
dd <- read.csv(url)
```

```r
dd %>%
  ggplot(aes(species, agb_g))+
  geom_boxplot(aes(fill = trt))+
  scale_fill_manual(values = c("#43AA8B", "#DB504A"))+
  labs(x = "Species",
       y = "Aboveground biomass (g)",
       fill = "Species")+
  theme_classic()+
  theme(aspect.ratio = 1)
```
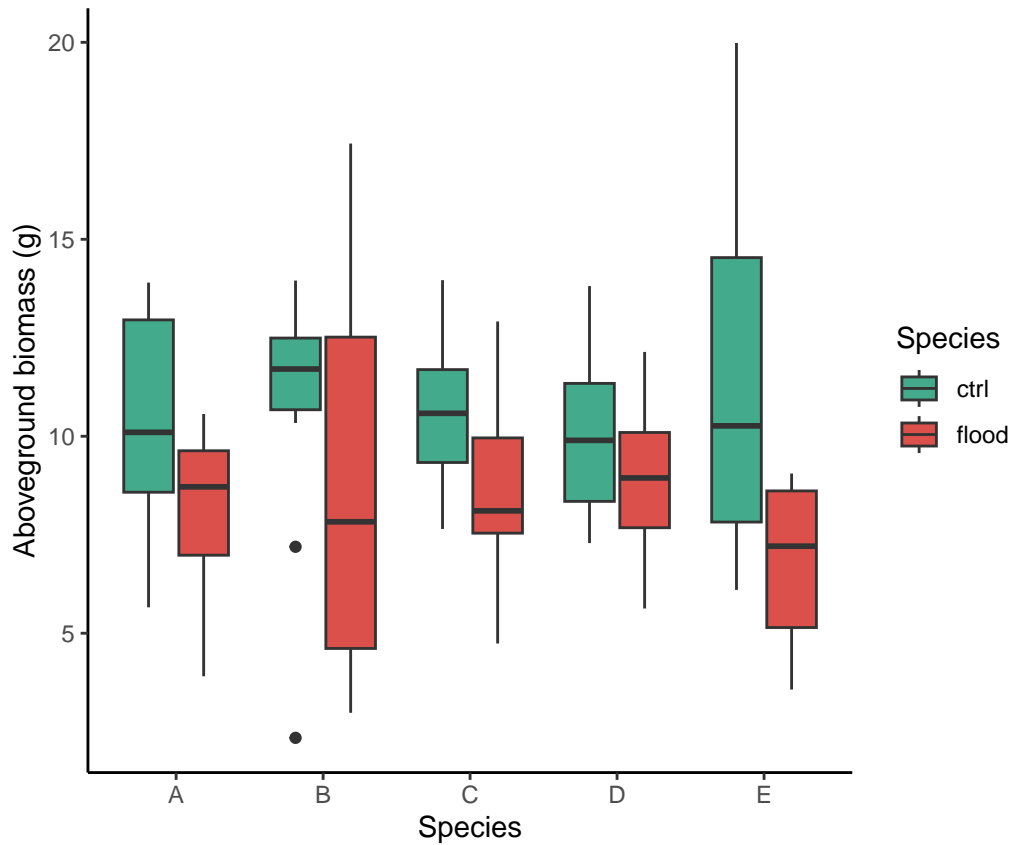
a. Propose a statistical model (using mathematical notation) to describe the relationship between aboveground biomass, legume species and treatment. Fit that model to the data.

b. Mention the assumptions in your model and check them. If you believe one of the assuptions is badly wrong and should be changed, name a possible alternative. (Note: you only have to state if you think they are approximately appropriate or not, and name a possible alternative. You do not need to do it!)

c. Provide point estimates and 95% confidence intervals of the different species under both treated and control conditions.