Exercise: Two-way ANOVA in R

Stephan Huber

2024-06-04

This is an exercise where data management with the `dplyr` functions `pivot_longer`, `rename`, and `bind_rows` is practiced. Moreover, I exemplify how an ANOVA analysis can be executed with R. Doing so I refer to the content of Childs et al. (2021, Chapter 27). All data to this exercise can be found in Huber (2024).

Table of contents

1 Read

Read Childs et al. (2021):

Our goal is to learn how to work with two-way ANOVA models in R, using an example from a plant competition experiment. The work flow is very similar to one-way ANOVA in R. We'll start with the problem and the data, and then work through model fitting, evaluating assumptions, significance testing, and finally, presenting the results.

2 Set up the R session

Download and open the script that you find here. This script contains all the code that is shown below. The solutions to the exercises can be found here.

```
rm(list = ls())

if (!require(pacman)) install.packages("pacman")
pacman::p_load(tidyverse, rstatix, ggpubr, agricolae)
```

3 Create a data frame

Plants have an optimal soil pH for growth, and this varies between species. Consequently we would expect that if we grow two plants in competition with each other at different pH values the effect of competition might vary according to the soil pH. In a recent study the growth of the grass Festuca ovina (Sheep's Fescue) in competition with the heather Calluna vulgaris (Ling) was investigated in soils with different pH. Calluna is well adapted to grow on very acidic soils such as on the Millstone grit and blanket bogs around Sheffield. Festuca grows on soils with a much wider range of pH. We might hypothesise that Calluna will be a better competitor of Festuca in very acid soils than in moderately acid soils. Here are the data: The column (Weight) contains the Festuca dry weights, (pH) contains the codes for the pH treatment (levels: pH3.5, pH5.5), the column (Calluna) contains the codes for the presence or absence of Calluna (levels: Present, Absent).

```
data_present <- data.frame(
  Condition = rep(c("Calluna Present"), each = 5),
  `pH 3.5` = c(2.76, 2.39, 3.54, 3.71, 2.49),
  `pH 5.5` = c(3.21, 4.10, 3.04, 4.13, 5.21),
  check.names = FALSE
)
data_present
```

```
        Condition pH 3.5 pH 5.5
1 Calluna Present    2.76    3.21
2 Calluna Present    2.39    4.10
3 Calluna Present    3.54    3.04
4 Calluna Present    3.71    4.13
5 Calluna Present    2.49    5.21
```

```
data_absent <- data.frame(
  Condition = rep(c("Calluna Absent"), each = 5),
  `pH 3.5` = c(4.10, 2.72, 2.28, 4.43, 3.31),
  `pH 5.5` = c(5.92, 7.31, 6.10, 5.25, 7.45),
  check.names = FALSE
)
data_absent
```

```
        Condition pH 3.5 pH 5.5
1 Calluna Absent    4.10    5.92
2 Calluna Absent    2.72    7.31
3 Calluna Absent    2.28    6.10
4 Calluna Absent    4.43    5.25
5 Calluna Absent    3.31    7.45
```

4 Make the data "long"

Read the R documentation of the function `bind_rows` which is part of the `dplyr` package.

Combine the objects `data_present` and `data_absent`.

```
         Condition pH 3.5 pH 5.5
1  Calluna Present    2.76    3.21
2  Calluna Present    2.39    4.10
3  Calluna Present    3.54    3.04
4  Calluna Present    3.71    4.13
5  Calluna Present    2.49    5.21
6   Calluna Absent    4.10    5.92
7   Calluna Absent    2.72    7.31
8   Calluna Absent    2.28    6.10
9   Calluna Absent    4.43    5.25
10  Calluna Absent    3.31    7.45
```

Read Wickham & Grolemund (2023): 5.3 Lengthening data.

Rearrange the data so that they look like this:

```
festuca
```

```
# A tibble: 20 x 3
   Calluna         pH      Weight
   <chr>           <chr>   <dbl>
 1 Calluna Present pH 3.5   2.76
 2 Calluna Present pH 5.5   3.21
 3 Calluna Present pH 3.5   2.39
 4 Calluna Present pH 5.5   4.1
 5 Calluna Present pH 3.5   3.54
 6 Calluna Present pH 5.5   3.04
 7 Calluna Present pH 3.5   3.71
 8 Calluna Present pH 5.5   4.13
 9 Calluna Present pH 3.5   2.49
10 Calluna Present pH 5.5   5.21
11 Calluna Absent  pH 3.5   4.1
12 Calluna Absent  pH 5.5   5.92
13 Calluna Absent  pH 3.5   2.72
14 Calluna Absent  pH 5.5   7.31
15 Calluna Absent  pH 3.5   2.28
16 Calluna Absent  pH 5.5   6.1
17 Calluna Absent  pH 3.5   4.43
18 Calluna Absent  pH 5.5   5.25
19 Calluna Absent  pH 3.5   3.31
20 Calluna Absent  pH 5.5   7.45
```

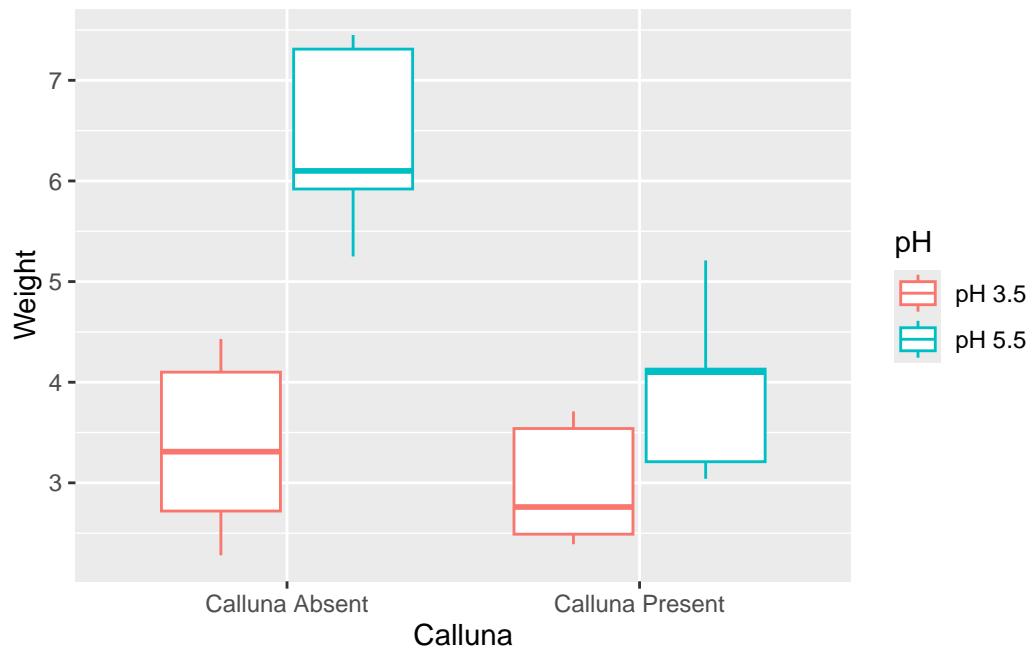# 5 Descriptive statistics

Calculate the following statistics and graphs:

```
summary_stats
```

```
# A tibble: 4 x 4
  Calluna         pH      mean    var
  <chr>           <chr>   <dbl>  <dbl>
1 Calluna Absent  pH 3.5  3.37  0.818
2 Calluna Absent  pH 5.5  6.41  0.893
```

```
3 Calluna Present pH 3.5  2.98 0.371
4 Calluna Present pH 5.5  3.94 0.754
```


6 ANOVA

Use this model to calculate the ANOVA: `Weight ~ pH + Calluna + pH:Calluna`

```
Analysis of Variance Table

Response: Weight
          Df  Sum Sq Mean Sq F value     Pr(>F)
pH         1 19.9800 19.9800 28.1792 7.065e-05 ***
Calluna    1 10.2102 10.2102 14.4001   0.00159 **
pH:Calluna 1  5.3976  5.3976  7.6126   0.01397 *
Residuals 16 11.3446  0.7090
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
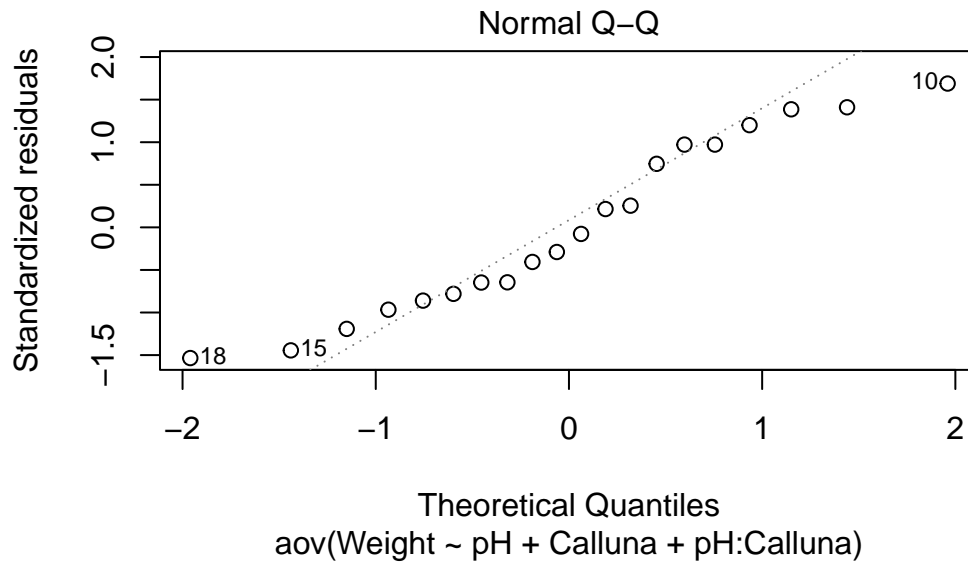
7 Diagnostics

Read Childs et al. (2021): 27.5 Diagnostics. Moreover, this page is worth a look. You find some alternatives R functions that you may find helpful for ANOVA diagnostics.

```
plot(festuca_model, which = 2, add.smooth = FALSE)
```

Normal Q–Q

aov(Weight ~ pH + Calluna + pH:Calluna)

```
plot(festuca_model, which = 3, add.smooth = FALSE)
```

Scale–Location

√|Standardized residuals|

Fitted values
aov(Weight ~ pH + Calluna + pH:Calluna)
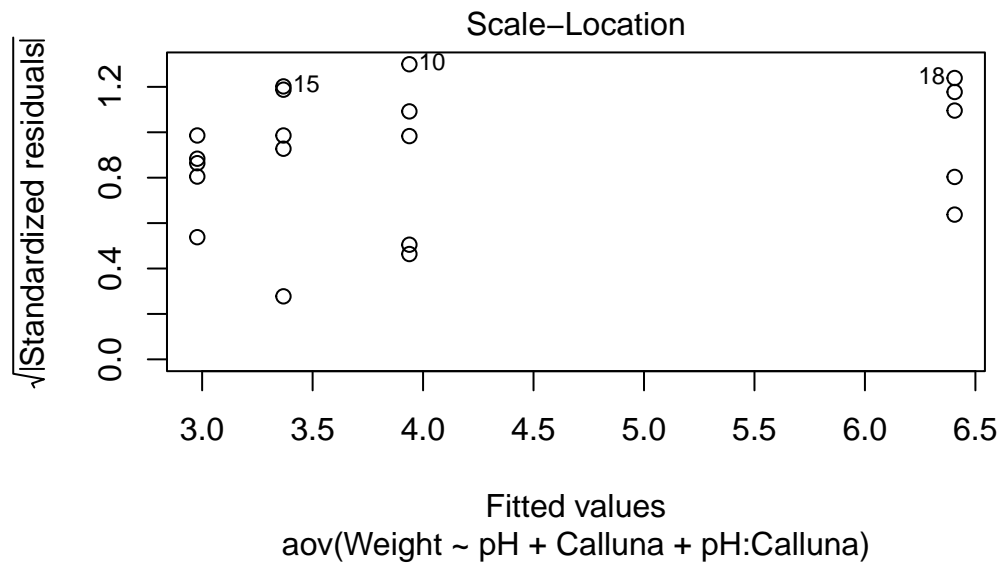
8 Interaction diagram

Can you use the function `interaction.plot` to create the following:
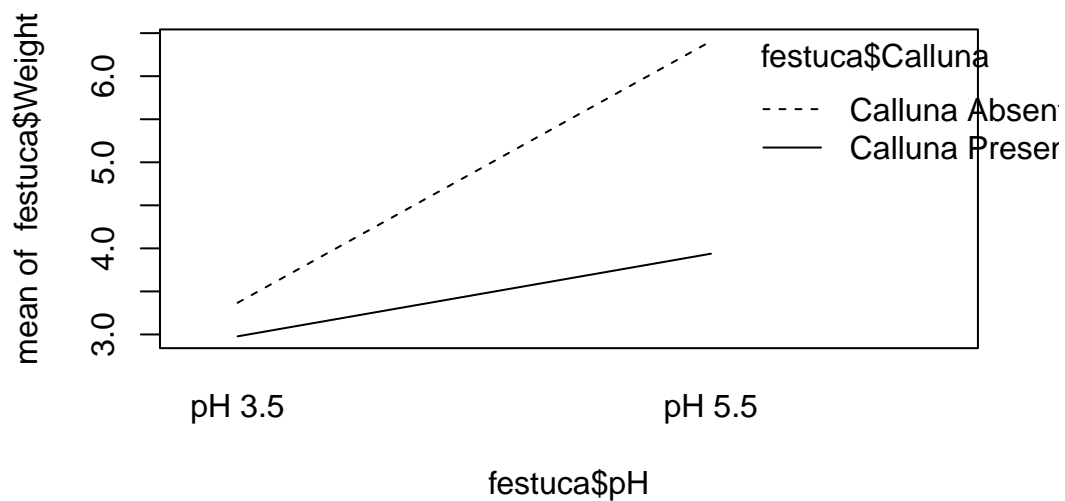
Here is a much nicer and more flexible approach to make interaction plots using `tidyverse` functions:

```
# step 1. calculate means for each treatment combination
festuca_means <-
  festuca %>%
  group_by(Calluna, pH) %>% # <- remember to group by *both* factors
  summarise(Means = mean(Weight))
```

```
# step 2. plot these as an interaction plot
ggplot(festuca_means,
       aes(x = Calluna, y = Means, colour = pH, group = pH)) +
  geom_point(size = 4) + geom_line()
```
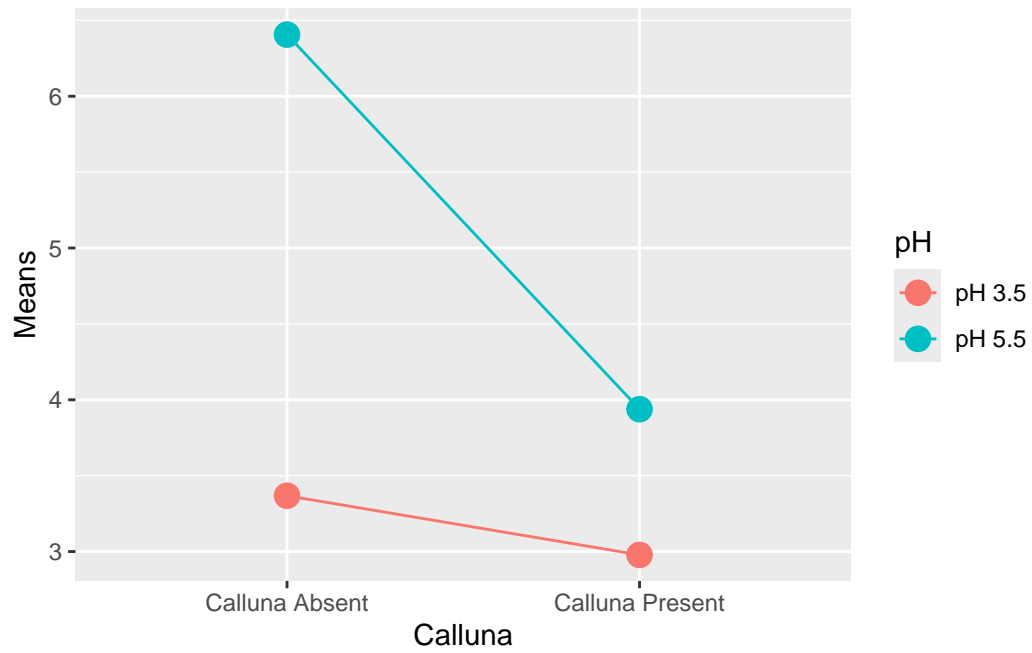
Read Childs et al. (2021): and consider the following figure:

With *Calluna*
Without *Calluna*

Festuca dry weight

- Significant effect of soil pH
- No significant effect of *Calluna*
- No significant interaction

Festuca grows better at higher pH regardless of whether *Calluna* is present or not.

pH 3.5   pH 5.5

Festuca dry weight
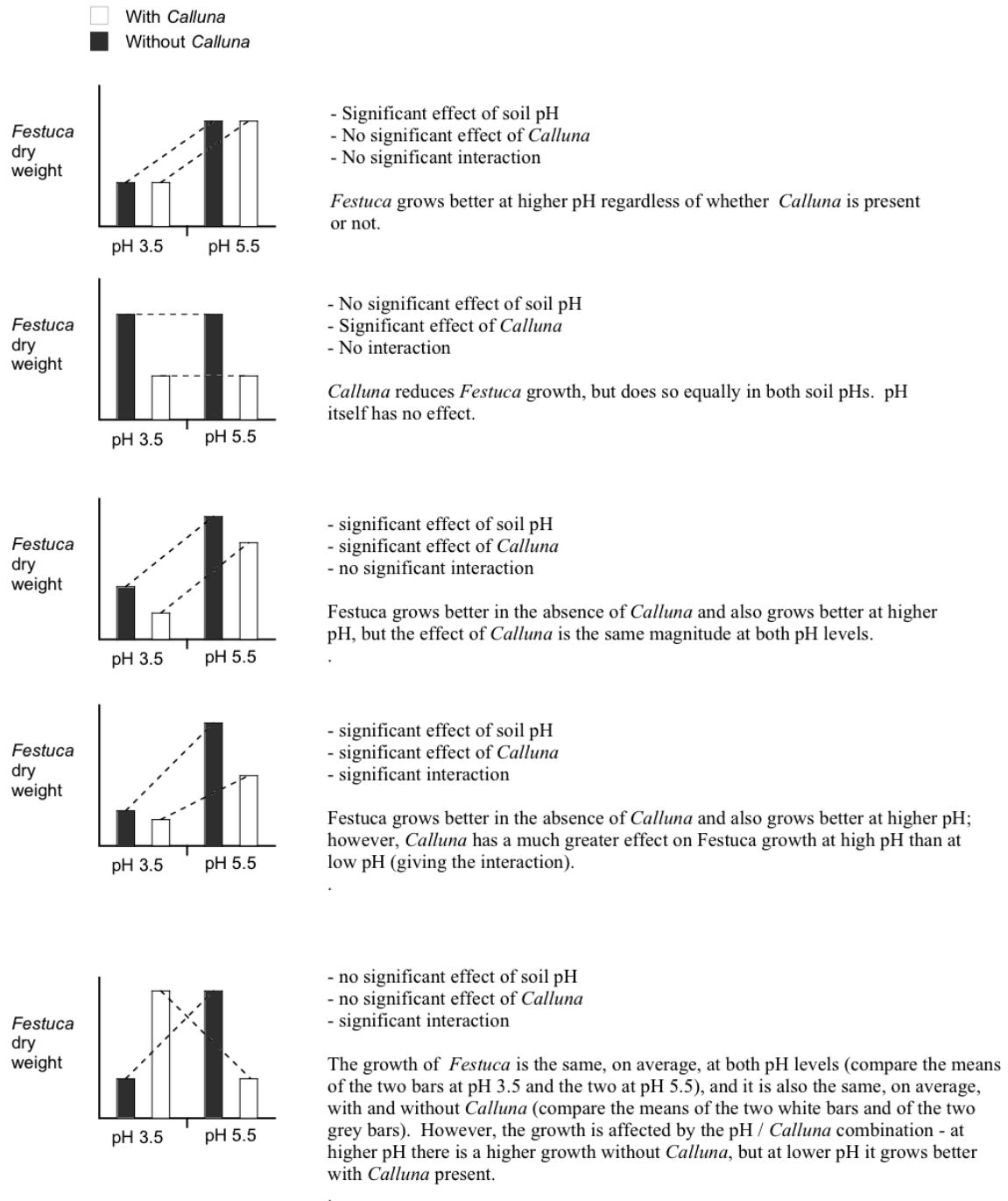
- No significant effect of soil pH
- Significant effect of *Calluna*
- No interaction

Calluna reduces *Festuca* growth, but does so equally in both soil pHs. pH itself has no effect.

pH 3.5   pH 5.5

Festuca dry weight

- significant effect of soil pH
- significant effect of *Calluna*
- no significant interaction

Festuca grows better in the absence of *Calluna* and also grows better at higher pH, but the effect of *Calluna* is the same magnitude at both pH levels.
.

pH 3.5   pH 5.5

Festuca dry weight

- significant effect of soil pH
- significant effect of *Calluna*
- significant interaction

Festuca grows better in the absence of *Calluna* and also grows better at higher pH; however, *Calluna* has a much greater effect on Festuca growth at high pH than at low pH (giving the interaction).
.

pH 3.5   pH 5.5

Festuca dry weight

- no significant effect of soil pH
- no significant effect of *Calluna*
- significant interaction

The growth of *Festuca* is the same, on average, at both pH levels (compare the means of the two bars at pH 3.5 and the two at pH 5.5), and it is also the same, on average, with and without *Calluna* (compare the means of the two white bars and of the two grey bars). However, the growth is affected by the pH / *Calluna* combination - at higher pH there is a higher growth without *Calluna*, but at lower pH it grows better with *Calluna* present.
.

pH 3.5   pH 5.5

Figure 1: Understanding the model graphically

9 Multiple comparison tests

```r
TukeyHSD(festuca_model, which = 'pH:Calluna')
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Weight ~ pH + Calluna + pH:Calluna, data = festuca)

$`pH:Calluna`
                                                diff       lwr        upr
pH 5.5:Calluna Absent-pH 3.5:Calluna Absent    3.038  1.5143518  4.5616482
pH 3.5:Calluna Present-pH 3.5:Calluna Absent  -0.390 -1.9136482  1.1336482
pH 5.5:Calluna Present-pH 3.5:Calluna Absent   0.570 -0.9536482  2.0936482
pH 3.5:Calluna Present-pH 5.5:Calluna Absent  -3.428 -4.9516482 -1.9043518
pH 5.5:Calluna Present-pH 5.5:Calluna Absent  -2.468 -3.9916482 -0.9443518
pH 5.5:Calluna Present-pH 3.5:Calluna Present  0.960 -0.5636482  2.4836482
                                                  p adj
pH 5.5:Calluna Absent-pH 3.5:Calluna Absent    0.0001731
pH 3.5:Calluna Present-pH 3.5:Calluna Absent   0.8826936
pH 5.5:Calluna Present-pH 3.5:Calluna Absent   0.7117913
pH 3.5:Calluna Present-pH 5.5:Calluna Absent   0.0000443
pH 5.5:Calluna Present-pH 5.5:Calluna Absent   0.0014155
pH 5.5:Calluna Present-pH 3.5:Calluna Present  0.3079685
```

::: {.cell}

```` ```{.r .cell-code} ````
```r
HSD.test(festuca_model, trt = c("pH", "Calluna"), console = TRUE)
```

```
Study: festuca_model ~ c("pH", "Calluna")

HSD Test for Weight

Mean Square Error:  0.709035

pH:Calluna,  means
```

```
                   Weight       std r        se  Min  Max  Q25  Q50  Q75
pH 3.5:Calluna Absent    3.368 0.9042511 5 0.3765727 2.28 4.43 2.72 3.31 4.10
pH 3.5:Calluna Present   2.978 0.6089089 5 0.3765727 2.39 3.71 2.49 2.76 3.54
pH 5.5:Calluna Absent    6.406 0.9451614 5 0.3765727 5.25 7.45 5.92 6.10 7.31
pH 5.5:Calluna Present   3.938 0.8685448 5 0.3765727 3.04 5.21 3.21 4.10 4.13


Alpha: 0.05 ; DF Error: 16
Critical Value of Studentized Range: 4.046093


Minimun Significant Difference: 1.523648


Treatments with the same letter are not significantly different.


                   Weight groups
pH 5.5:Calluna Absent    6.406     a
pH 5.5:Calluna Present   3.938     b
pH 3.5:Calluna Absent    3.368     b
pH 3.5:Calluna Present   2.978     b
```

:::

10 Drawing conclusions and presenting results

Here are some code examples how the shown graphs above could be made much nicer.

```
# step 1. calculate means for each treatment combination
festuca_stats <-
  festuca %>%
  group_by(Calluna, pH) %>% # <- remember to group by the two factors
  summarise(Means = mean(Weight), SEs = sd(Weight)/sqrt(n()))
```

`summarise()` has grouped output by 'Calluna'. You can override using the `.groups` argument.

```
# step 1. calculate means for each treatment combination
festuca_stats <-
  festuca %>%
  group_by(Calluna, pH) %>% # <- remember to group by the two factors
  summarise(Means = mean(Weight), SEs = sd(Weight)/sqrt(n()))
```

```
# step 2. plot these as an interaction plot
ggplot(festuca_stats,
       aes(x = Calluna, y = Means, colour = pH,
           ymin = Means - SEs, ymax = Means + SEs)) +
  # this adds the mean
  geom_point(size = 3) +
  # this adds the error bars
  geom_errorbar(width = 0.1) +
  # controlling the appearance
  scale_y_continuous(limits = c(2, 7)) +
  xlab("Calluna") + ylab("Festuca yield (g dry weight)") +
  # use a more professional theme
  theme_bw()
```
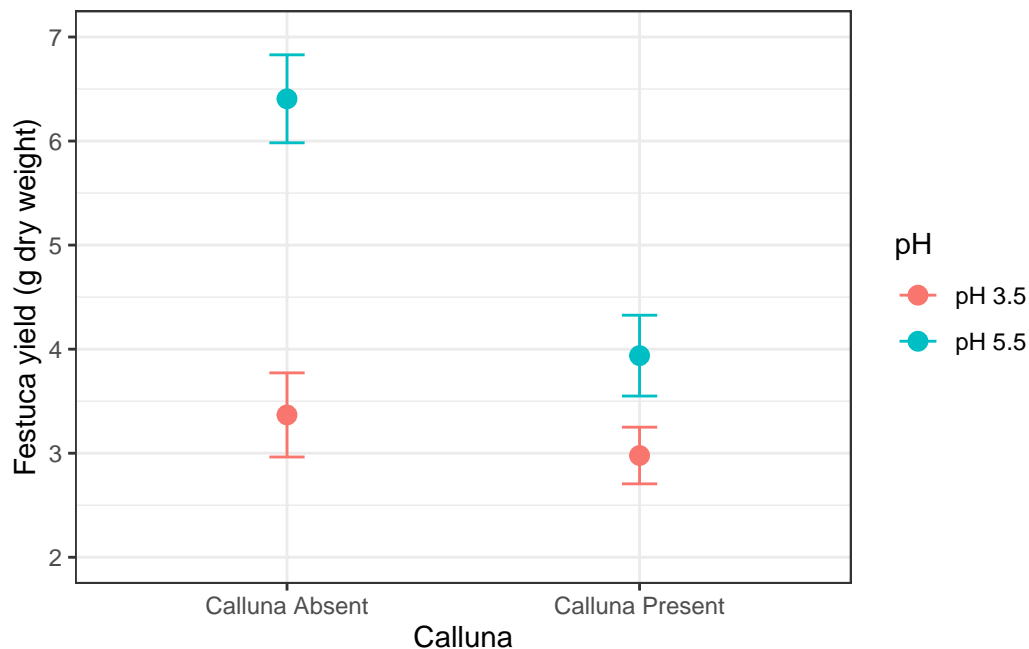


```
# define a position adjustment
pos <- position_dodge(0.15)
# make the plot
ggplot(festuca_stats,
       aes(x = Calluna, y = Means, colour = pH,
           ymin = Means - SEs, ymax = Means + SEs)) +
  # this adds the mean (shift positions with 'position =')
  geom_point(size = 3, position = pos) +
```

```
# this adds the error bars (shift positions with 'position =')
geom_errorbar(width = 0.1, position = pos) +
# controlling the appearance
scale_y_continuous(limits = c(2, 7)) +
xlab("Calluna") + ylab("Festuca yield (g dry weight)") +
# use a more professional theme
theme_bw()
```
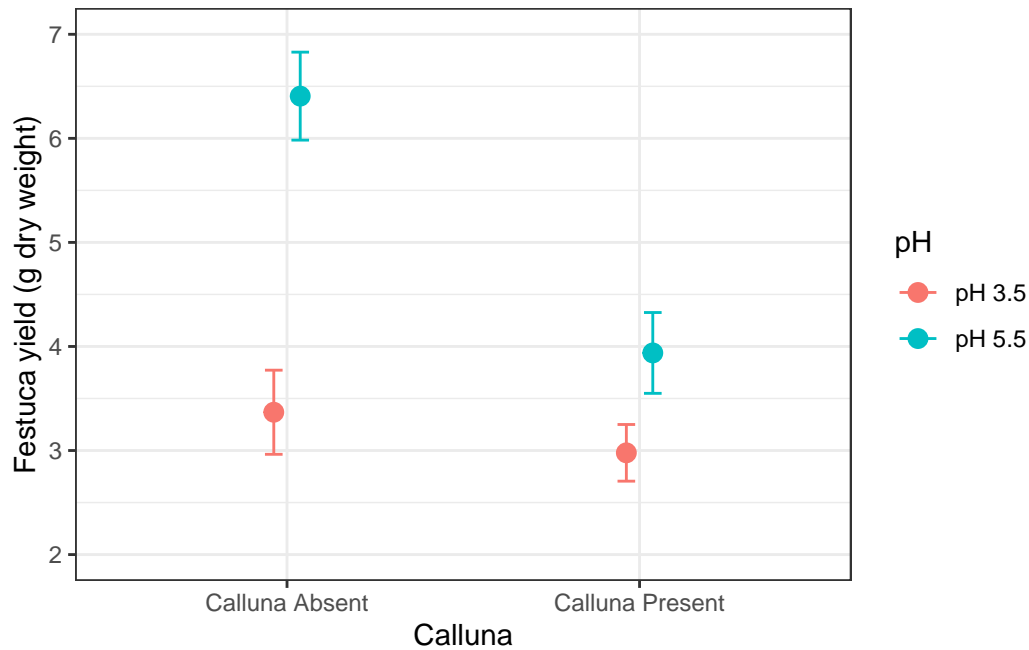


```
ggplot(festuca_stats,
       aes(x = Calluna, y = Means, fill = pH,
           ymin = Means - SEs, ymax = Means + SEs)) +
  # this adds the mean
  geom_col(position = position_dodge()) +
  # this adds the error bars
  geom_errorbar(position = position_dodge(0.9), width=.2) +
  # controlling the appearance
  xlab("Calluna") + ylab("Festuca yield (g dry weight)")
```
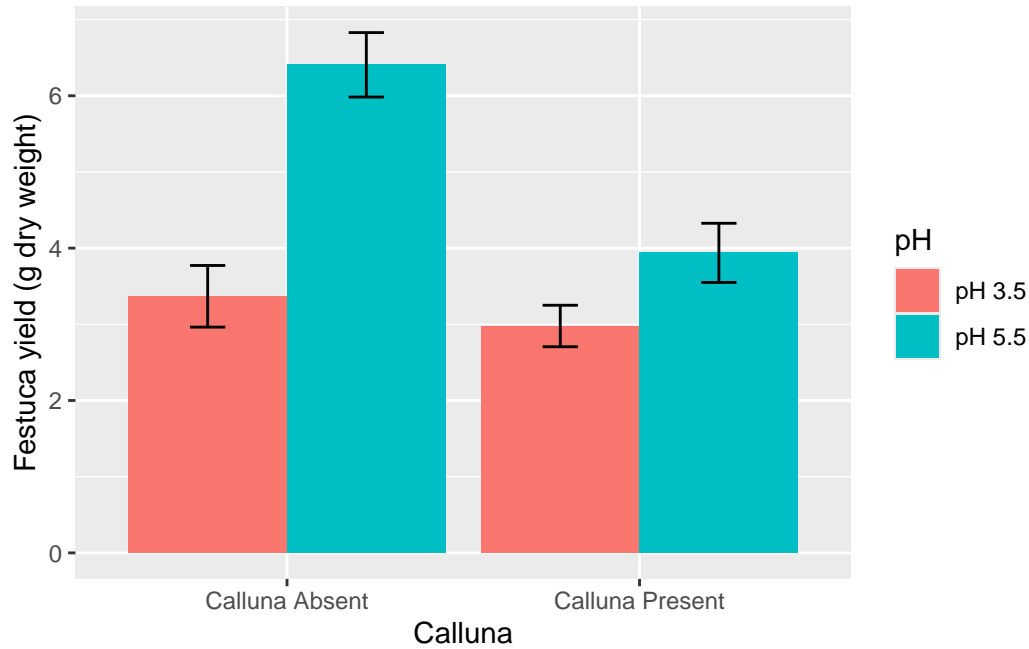
References

Childs, D. Z., Hindle, B. J., & Warren, P. H. (2021). *APS 240: Data analysis and statistics with r.* online. https://dzchilds.github.io/stats-for-bio

Huber, S. (2024). *Empirisch-wissenschaftlich arbeiten (ewa).* GitHub repository. https://github.com/hubchev/ewa

Wickham, H., & Grolemund, G. (2023). *R for data science (2e).* Accessed January 30, 2023. https://r4ds.hadley.nz/