Homework 5

Luke Todd

2023-06-05

Loading Packages

```r
# should haves (from last week)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ---------------------- tidyverse 2.0.0 --
## v dplyr     1.1.1     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts --------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(here)
```

```
## here() starts at /Users/luketodd/Desktop/Rprojects/ES 193DS/ES 193DS Homework 5
```

```r
library(janitor)
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```r
library(ggeffects)
library(performance)
library(naniar) # or equivalent
library(flextable) # or equivalent
```

```
##
## Attaching package: 'flextable'
##
## The following object is masked from 'package:purrr':
##
##     compose
```

```
library(car)
```

```
## Loading required package: carData
##
## Attaching package: 'car'
##
## The following object is masked from 'package:dplyr':
##
##     recode
##
## The following object is masked from 'package:purrr':
##
##     some
```

```
library(broom)
# would be nice to have
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
library(AICcmodavg)
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

Loading Data

```
sar <- read.csv(here("Data/hf109-01-sarracenia.csv")) %>%
  clean_names() %>%
  select(totmass, species, feedlevel, sla, chlorophyll, amass, num_lvs, num_phylls)
```

7a

7b

7c

```
gg_miss_var(sar)
```

```
# creating a dataset without the missing values
sar_nona <- sar %>%
  drop_na(sla, chlorophyll, amass, num_lvs, num_phylls)
```
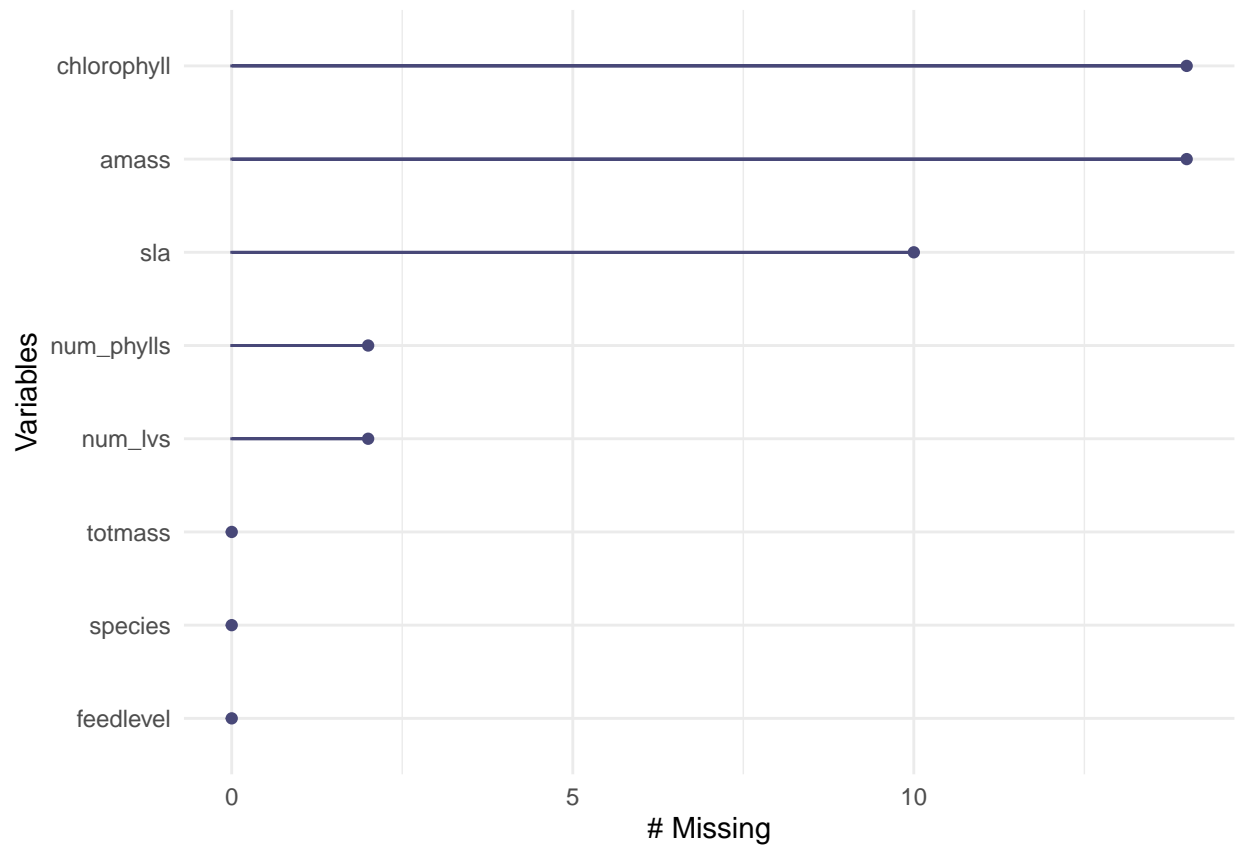
Figure 1: This figure displays the number of missing values for each variable in the sar data frame. Based on this figure, we can see that chlorophyll (Chlorophyll content), amass (Photosynthetic Rate), and sla (Specific Leaf Area) are missing the most values with 14, 14, and 10 missing values, respectively. Num_phylls and num_lvs are missing 2 values, and the rest are missing none.

7d

```
# calc Pearson's r for numerical values
sar_cor <- sar_nona %>%
  select(feedlevel:num_phylls) %>%
  cor(method = "pearson")

# plot correlation values
corrplot(sar_cor,
         method = "ellipse",
         addCoef.col = "black")
```
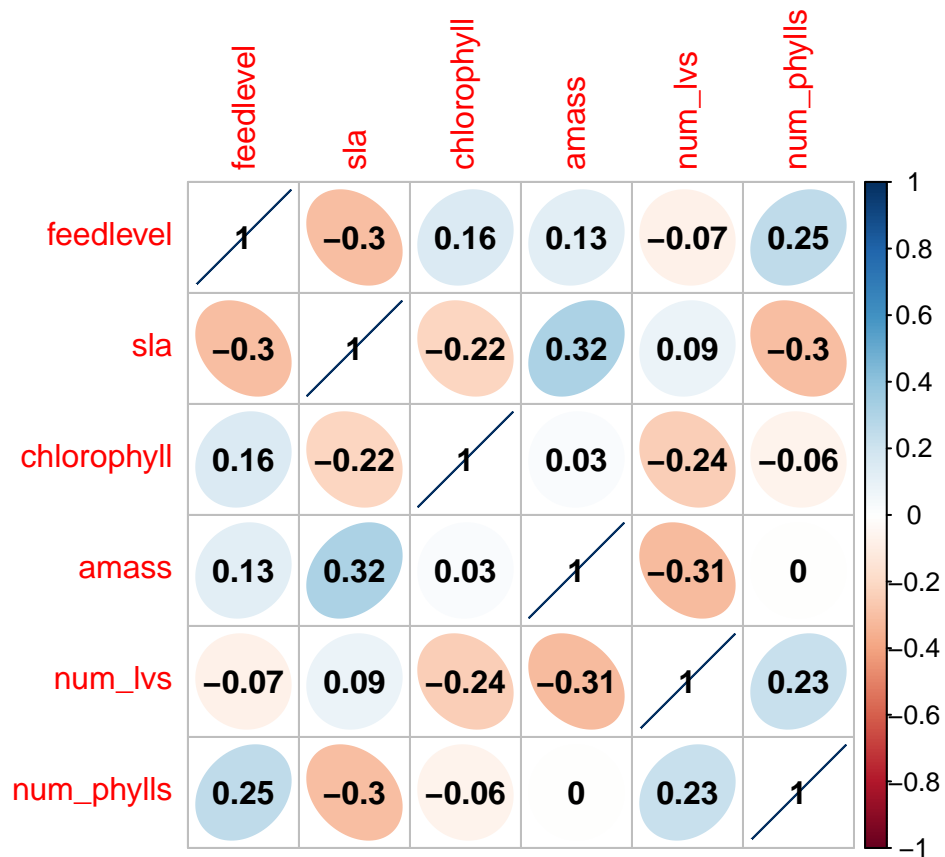

Figure 2: This figure displays the Person's correlation values between different variables. High absolute values means that there is a greater correlation between the variables. For example, sla and amass have the greatest positive correlation between each other with a value of 0.32. On the otherhand, num_lvs and amass have the greatest negative correlation between each other with a value of -0.31.

7e

```
sar_nona %>%
  select(species:num_phylls) %>%
  ggpairs()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
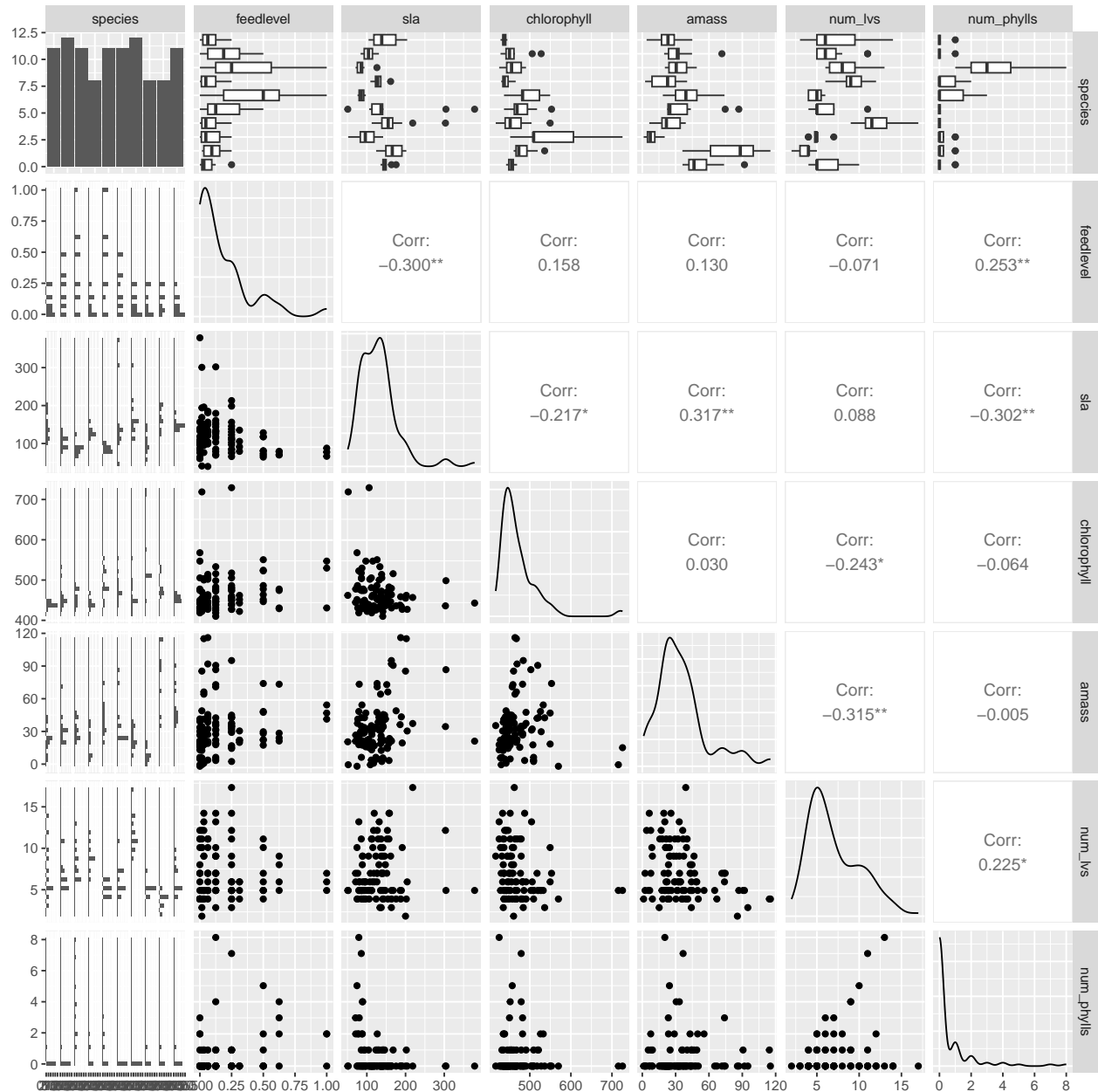

7f

As we are trying to predict totmass, the y-value is set to totmass. For the null model, we set the x-value to 1, as this selects for just the intercept. For the full model, we select every variable that we are interested in.

5

```
null <- lm(totmass ~ 1, data = sar_nona)
full <- lm(totmass ~ species + feedlevel +
             sla + chlorophyll + amass +
             num_lvs + num_phylls,
          data = sar_nona)
```

7g