Draft Manuscript

Emma Menchions

2022-09-23

Title: Can gradients in annual rainfall & forest type predict the distribution & abundance of 4 Ericaceous shrubs? - BIOL 548T LDP

Emma Menchions[1]*

1. Dept. of Botany, University of British Columbia, Vancouver, BC, Canada

*Corresponding Author

email: emma.menchions@botany.ubc.ca

Author Contributions: EM conceived of the study, conducted the analyses, and wrote the original, and revised drafts of the manuscript.

Data Availability: The data and code that support the findings of this study are openly available on GitHub at https://github.com/emench/Menchions-BIOL548T.

Conflict of Interest statement

No conflicts of interest

Abstract

1. Ecologists often have lots of questions about lots of stuff
2. We evaluated a bunch of things using sophisticated methods and carried out complicated statistical tests
3. We discovered a bunch of things that we didn't already know but suspected

4. Our research has greatly advanced out knowledge about stuff and will make a significant contribution to something and someone

Key-words: Ecology, Ericaceae, LDP, Plant Ecology, Vegetation Mapping, Vegetation Patterns

Introduction

Four species of Ericaceae dominate the understory in the northern boreal forests of Quebec, Canada. Two of these species *Rhododenron groenlandicum* and *Chamaedaphne caluculata* are said to prefer acidic soils and wetter habitats whereas the other dominant species *Kalmia angustifolia*, *Vaccinium angustifolium*, and *Vaccinium myrtillodes* are known to prefer drier, but also acidic soils. Many factors contribute to soil acidity, but a few variables in particular, rainfall and forest type have been observed to drive broad patterns in soil pH in the United States (2). Therefore, I ask, can annual rainfall and forest type (conifer, deciduous, or mixed) predict patterns of abundance for these species in the boreal zone of Northern Quebec between 1985 and 2000?

Acidic soils are often associated with conifer forests and higher rainfall (1,2). Therefore, species with a preference for more acidic soils should be more often found in conifer forest types with higher precipitation. Therefore, I predict that *R. groenlandicum* and *C. calyculata*, will have the highest mean percent cover in areas with conifer forest types and higher annual precipitation. In contrast, I predict that *K. angustifolia* and *Vaccinium sp.*, will have the highest mean percent cover in conifer forests with lower annual precipitation.

Specifically, I predict the following pattern of mean percent cover for *R. groenlandicum* and *C. calyculata*: (H = high precipitation, L = low precipitation, c= conifer forests, d=deciduous forests, m =mixed forest)

From highest to lowest predicted mean percent cover with high rainfall and conifer forests being indicative of higher abundance: H-c, L-c, H-m, H-d, L-m, L-d

For *K. angustifolium* and *Vaccinium sp.* I predict the following pattern, with less rainfall being of greater importance to abundance: L-c, L-m, L-d, H-c, H-m, H-d

Methods

Data collection

Data was retrieved from (3): https://datadryad.org/stash/dataset/doi:10.5061%2Fdryad.4767v. According to Thiffault et al. (2016), the data on abundance of the 4 species were collected from 15,339 circular plots across central and northern Quebec along 1.5 – 2 km long transects of which supported 5-7 plots. Forest type: was determined by visual estimations at the time that the plots were visited by assigning a type corresponding to the two most dominant tree species within the plot.This lead to a total of 15 total forest types. Climate data was collected by identifying the closest meteorological station (51 in the study region) and correcting for differences in latitude, longitude and elevation between the plot and the station using distance weighting and multiple linear regression.

Data download (script: “part-1_data_download”)

All data download, preparation and statistical analyses were carried out in R 4.2.1 (R Core Team 2022). All scripts used are available on Zenodo / GitHub https://github.com/emench/Menchions-BIOL548T. I first retrieved this data from Dryad using a script that automates the download from Dryad.

Data preparation (script: “part-2_data_preparation”)

I then prepared the data for analysis by altering the forest type classification and creating a variable indicating whether the plot was in a high or low rainfall area. For the forest type classifications, I used the metadata on forest type codes provided to create 4 categories of forest type: “c” (coniferous, if the two dominant tree species in the plot were both coniferous, n=1282), “d” (deciduous, if the two dominant species were both deciduous, n=45), “m” (mixed, if the two most dominant species contained a conifer and a deciduous species, n=69), and “n” (non-forested, n=26).

```r
# Loading data tables ----
X_mat <- read.csv(here::here("data","Matrice_X.csv"),
                  sep = ";") # explanatory variables - see metadata
Y_mat <- read.csv(here::here("data","Matrice_Y.csv"),
                  sep = ";") # response variables (mean ericaceous cover, in %)
clim <- read.csv(here::here("data","Climate_value.csv"),
                 sep = ";") # numerical codes used for clim vars in X_mat

# Classifying Forest Type ----
# Identifying two most dominant tree species in each plot and assigning forest type as
  # conifer (c)
  # deciduous (d)
  # mixed (m)
# looping through each row of X_mat dataframe

for (i in 1:dim(X_mat)[1]){
  most_abun <-  as.numeric(Rfast::nth(as.matrix(X_mat[i,2:14]),
                                      k=1, descending = T,index.return = T))
  # returning the index of the column with the most abundant tree species for each row
  sec_abun <- as.numeric(Rfast::nth(as.matrix(X_mat[i,2:14]),
                                    k=2, descending = T,index.return = T))
```

```r
# returning index of second most abundant species
dominant_sp <- c(most_abun, sec_abun)+1 # adding 1 to account for the
# first column which was left out of vector for previous function
dominant_sp <- colnames(X_mat)[dominant_sp] # creating character vector containing
# column names of most dominant species

if (dominant_sp[1] == "AB" || # if both species conifers
    dominant_sp[1] == "PB" ||
    dominant_sp[1] == "PM"||
    dominant_sp[1] == "PM_AB" ||
    dominant_sp[1] == "TO" &&
    dominant_sp[2] == "AB" ||
    dominant_sp[2] == "PB" ||
    dominant_sp[2] == "PM"||
    dominant_sp[2] == "PM_AM" ||
    dominant_sp[2] == "TO"){
    X_mat$Forest_Type[i] = "c" # forest type conifer

}else if (dominant_sp[1] == "BA" || # if both species deciduous
          dominant_sp[1] == "BP" ||
          dominant_sp[1] == "PT" &&
          dominant_sp[2] == "BA" ||
          dominant_sp[2] == "BP" ||
          dominant_sp[2] == "PT"){
  X_mat$Forest_Type[i] = "d" # forest type deciduous

}else if (dominant_sp[1] == "BP_AB" || # if either species is mixed
          dominant_sp[1] == "BP_PM" ||
          dominant_sp[1] == "PT_AB"||
          dominant_sp[1] == "PT_PM" ||
          dominant_sp[1] == "BP_AB" ||
          dominant_sp[1] == "BP_PM" ||
            dominant_sp[1] == "PT_AB"||
            dominant_sp[1] == "PT_PM"){
    X_mat$Forest_Type[i] = "m"  # forest type mixed

  }else if(dominant_sp[1] == "AB" || # if first species is conifer, second is deciduous
            dominant_sp[1] == "PB" ||
            dominant_sp[1] == "PM"||
            dominant_sp[1] == "PM_AB" ||
            dominant_sp[1] == "TO" &&
            dominant_sp[2] == "BA" ||
            dominant_sp[2] == "BP" ||
            dominant_sp[2] == "PT"){
    X_mat$Forest_Type[i] = "m" # forest type mixed

  }else if(dominant_sp[1] == "BA" || # if first species is deciduous, second is conifer
            dominant_sp[1] == "BP" ||
            dominant_sp[1] == "PT" &&
            dominant_sp[2] == "AB" ||
            dominant_sp[2] == "PB" ||
            dominant_sp[2] == "PM"||
            dominant_sp[2] == "PM_AB" ||
            dominant_sp[2] == "TO"){
            X_mat$Forest_Type[i] = "m" # forest type mixed
}else{
  X_mat$Forest_Type[i] = "n" # non-forested - other
}
}

# checking the dimensions of the dataframe when we select for specific cateogires
# tells us that the conifer type dominates
# this difference will be normalized later when we take the difference
```

```
dim(dplyr::filter(X_mat,Forest_Type=="c"))
```

```
## [1] 1282   49
```

```
dim(dplyr::filter(X_mat,Forest_Type=="d"))
```

```
## [1] 45 49
```

```
dim(dplyr::filter(X_mat,Forest_Type=="m"))
```

```
## [1] 69 49
```

```
dim(dplyr::filter(X_mat,Forest_Type=="n"))
```

```
## [1] 26 49
```

I then created a rainfall index. I calculated the median annual usable precipitation (cm) which was 5cm/ year. I used this as a threshold and defined any plot with annual usable precipitation > 5cm to a high precipitation category (H) and below this threshold to low annual rainfall (L).

```
# Classifying as high or low precipitation ----
# PRECU is the annual rainfall variable

med <- median(X_mat$PRECU)
# will use the median as the threshold

X_mat <- X_mat %>%
  mutate(., RainFall = case_when(PRECU < med ~ "L",
            PRECU >= med ~ "H"))
```

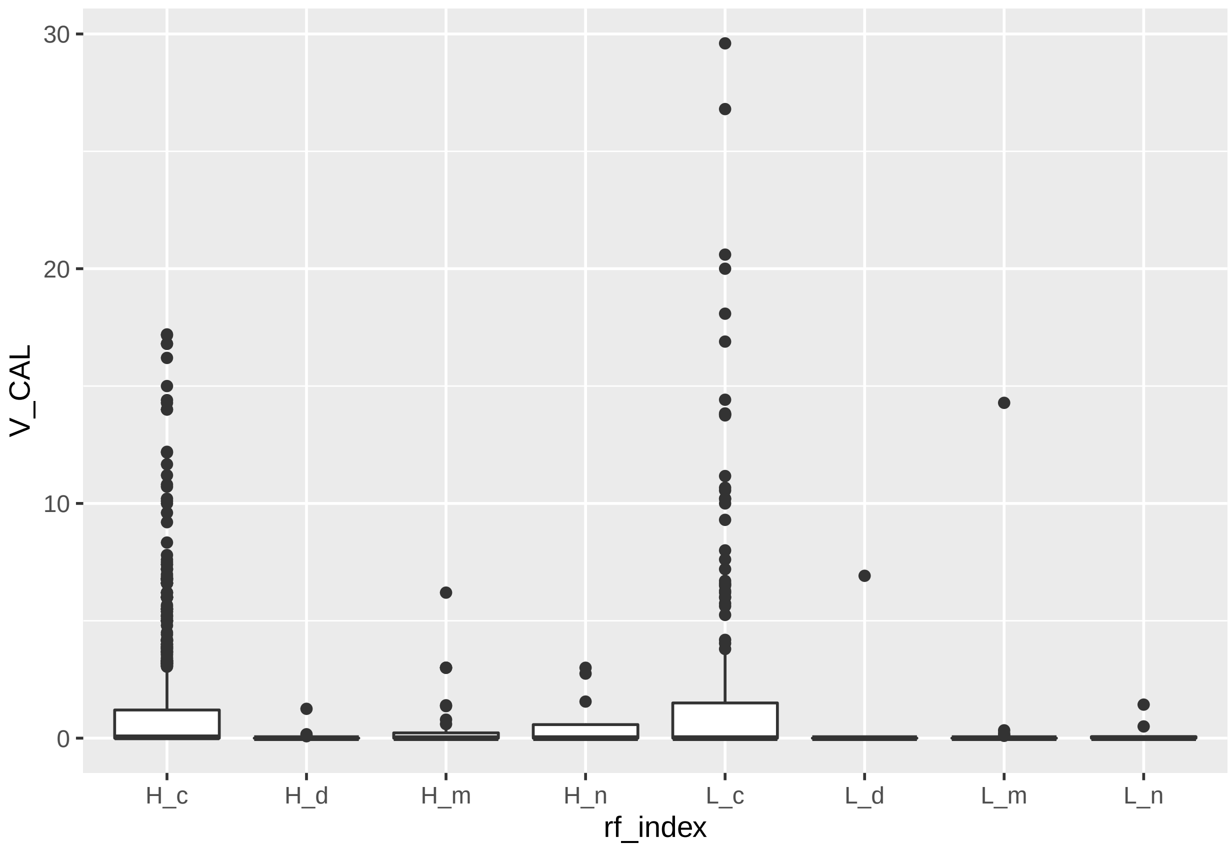
Finally, I combined the forest type and rainfall variables into an index (rf index) that described the rainfall category followed by the forest type (e.g. H_c = high rainfall, conifer forest).

```
# Combining and classifying as forest type - rainfall index
X_mat <-
  X_mat %>%
  unite(., rf_index, c("RainFall","Forest_Type"), sep = "_", remove =F, na.rm = FALSE)


# writing altered data as .csv ----
write.csv(X_mat, here::here("data","X_mat_processed.csv"))
```

Plotting (script: "part-3_data_analysis")

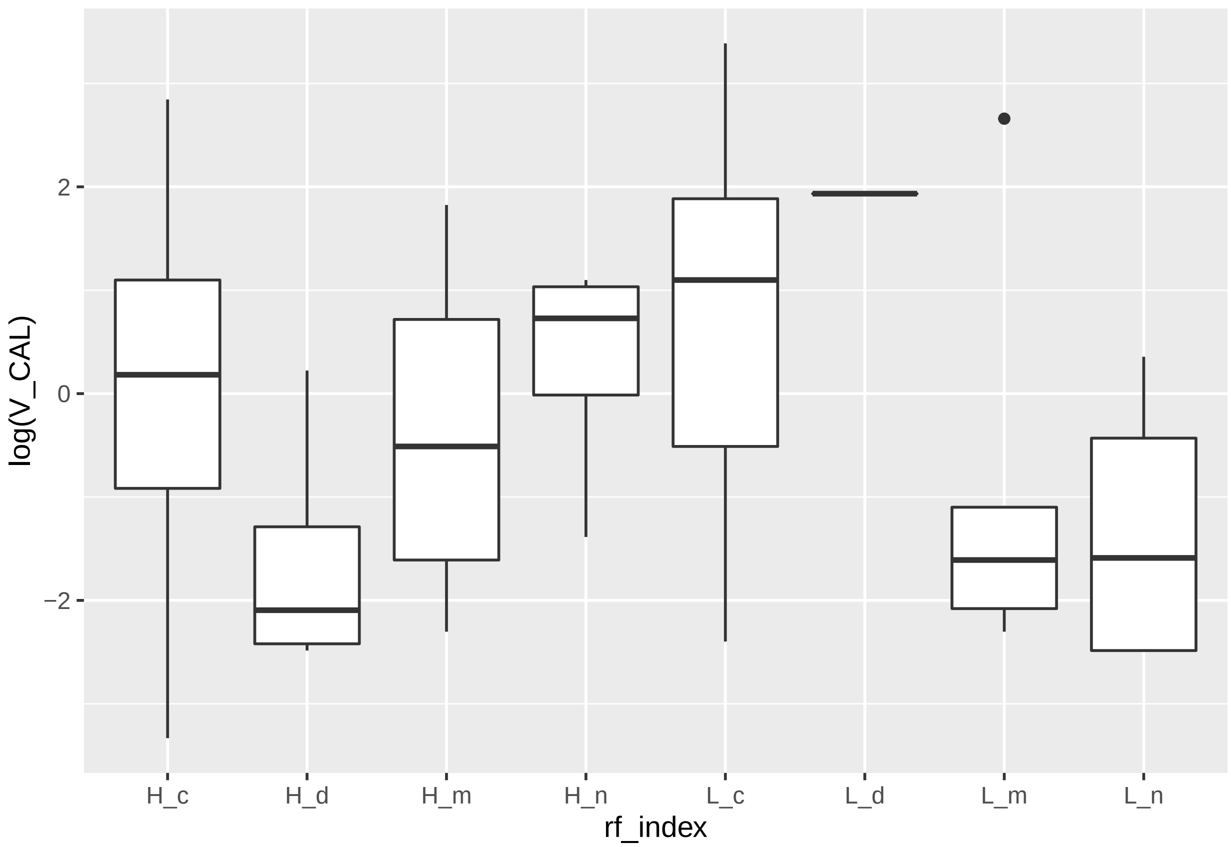
I constructed box plots displaying percent cover of each species as the response variable and rf (rainfall forest cover index) as the grouping variable using using ggplot2::ggplot. I decided to log-transform the response variable since the data was skewed with outliers making the display and comparison between groups difficult in all of the plots.

```
# loading in prepared data ----
  X_mat <- read.csv(here::here("data","X_mat_processed.csv"))
  X_mat <- X_mat %>%
    select(CLE, Forest_Type, PRECU, RainFall,rf_index)


  Y_mat <- read.csv(here::here("data","Matrice_Y.csv"),
                   sep = ";") # response variables (mean ericaceous cover, in %)
  clim <- read.csv(here::here("data","Climate_value.csv"),
                  sep = ";") # numerical codes used for clim vars in X_mat


# PLOTTING WITH BOX PLOTS----


  # For Chamaedaphne calyculata
  # creating dataframe with only this species
  CAL <- data.frame(Y_mat$V_CAL,X_mat$PRECU, X_mat$rf_index)
  colnames(CAL) <- c("V_CAL","PRECU","rf_index") # renaming columns
  ggplot(data=CAL, aes(x=rf_index, y=V_CAL)) +
    geom_boxplot() # plotting only this species with the rainfall-forest
```

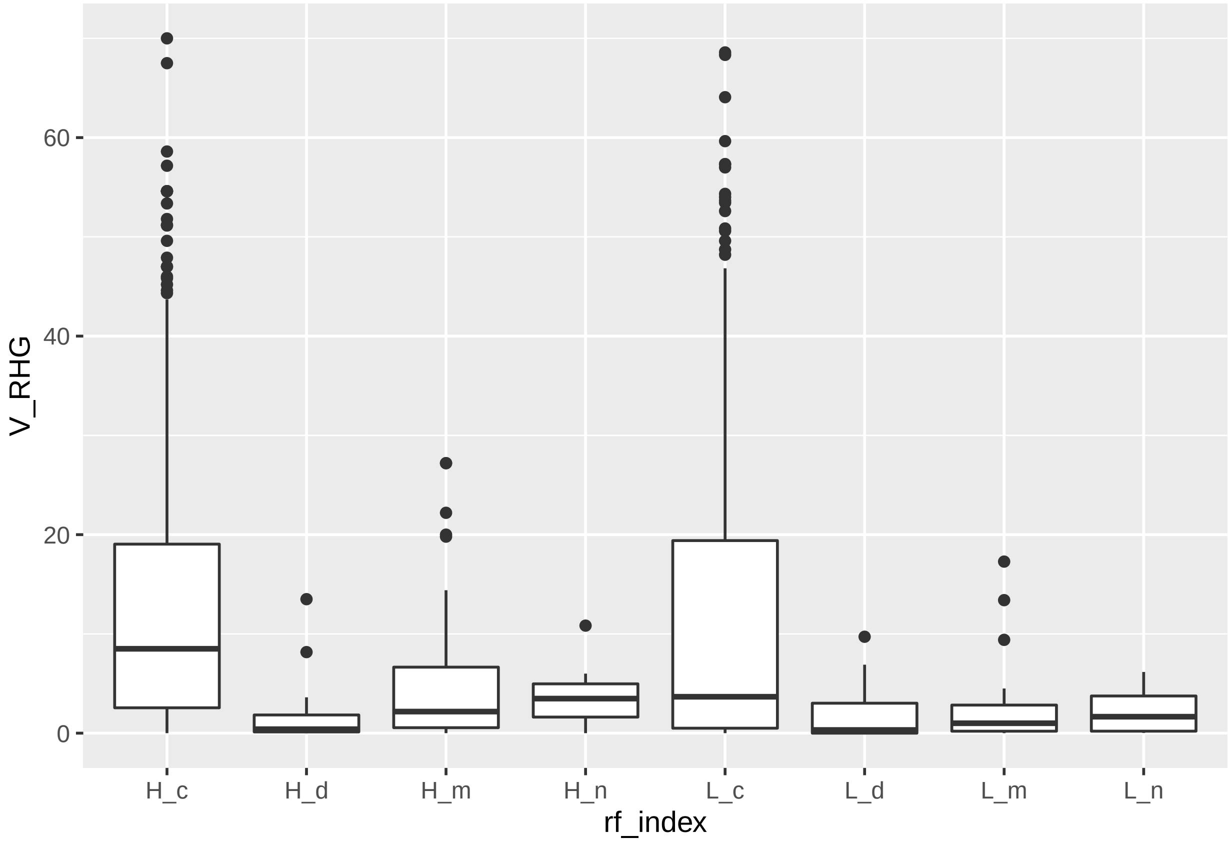


```
  # type index in box plots


ggplot(data=CAL, aes(x=rf_index, y=log(V_CAL))) + # log transforming the y axis
  geom_boxplot()  # automatically removes observations with 0 % cover - which is ok
```

```
## Warning: Removed 769 rows containing non-finite values (stat_boxplot).
```

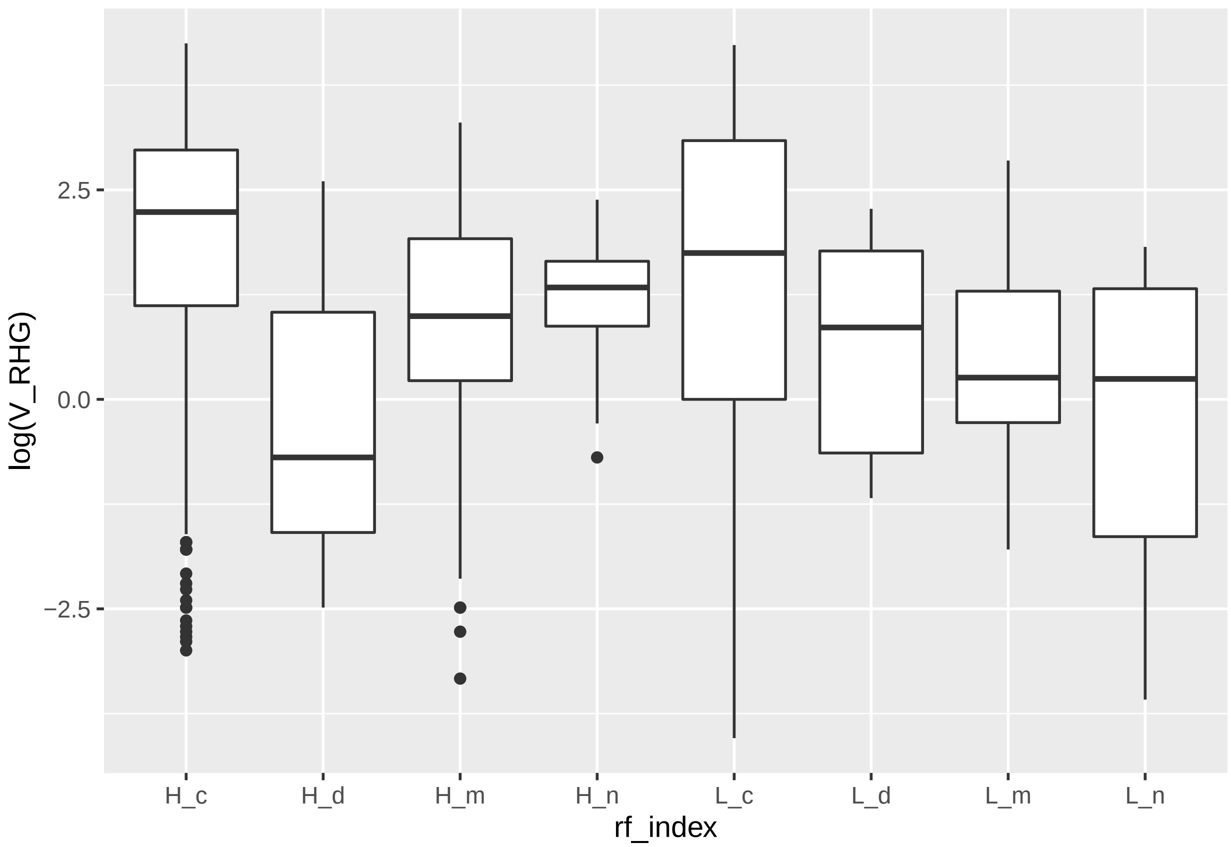


```
# Repeating for Rhododendron groenlandicum
RHG <- data.frame(Y_mat$V_RHG,X_mat$PRECU,X_mat$rf_index)
colnames(RHG) <- c("V_RHG","PRECU","rf_index")
ggplot(data=RHG, aes(x=rf_index, y=V_RHG)) +
  geom_boxplot()
```

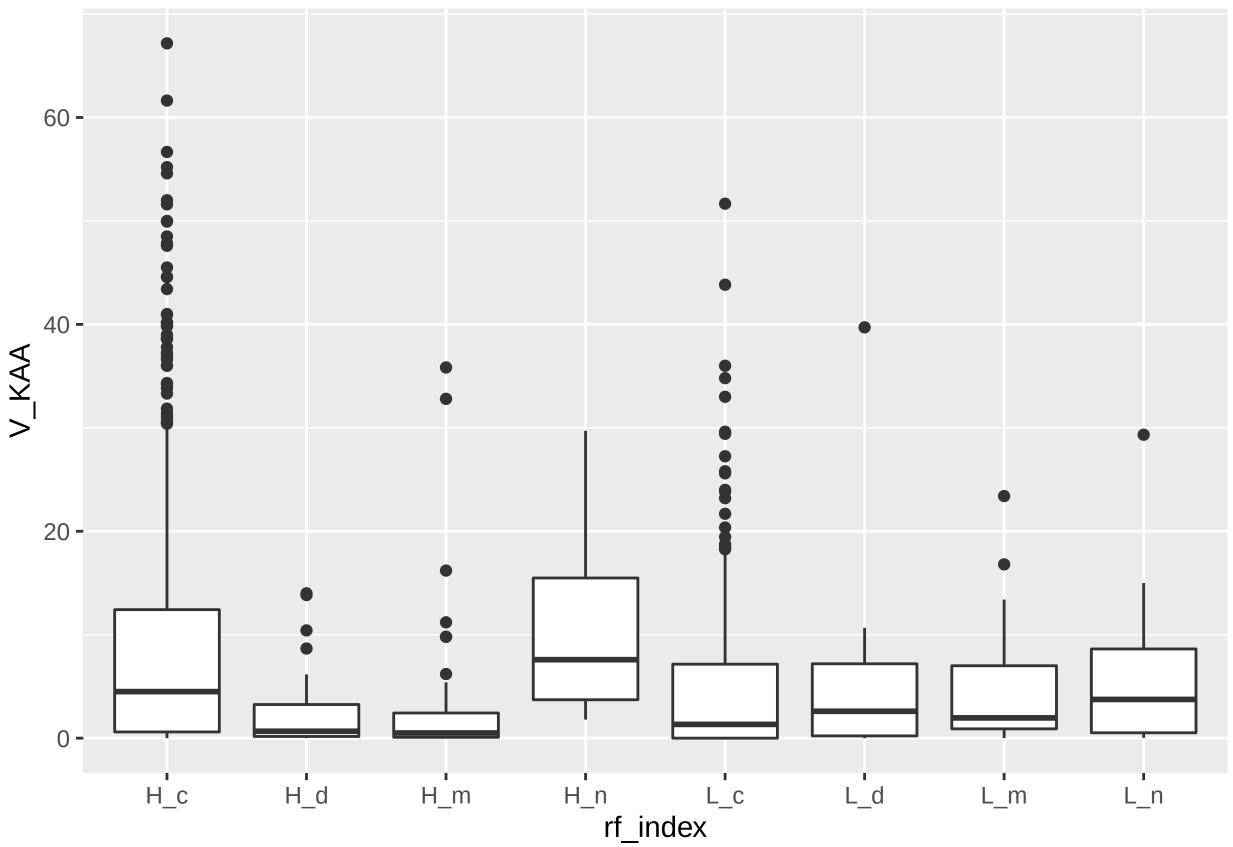


```
ggplot(data=RHG, aes(x=rf_index, y=log(V_RHG))) + # log transforming the y axis
  geom_boxplot()
```

```
## Warning: Removed 87 rows containing non-finite values (stat_boxplot).
```

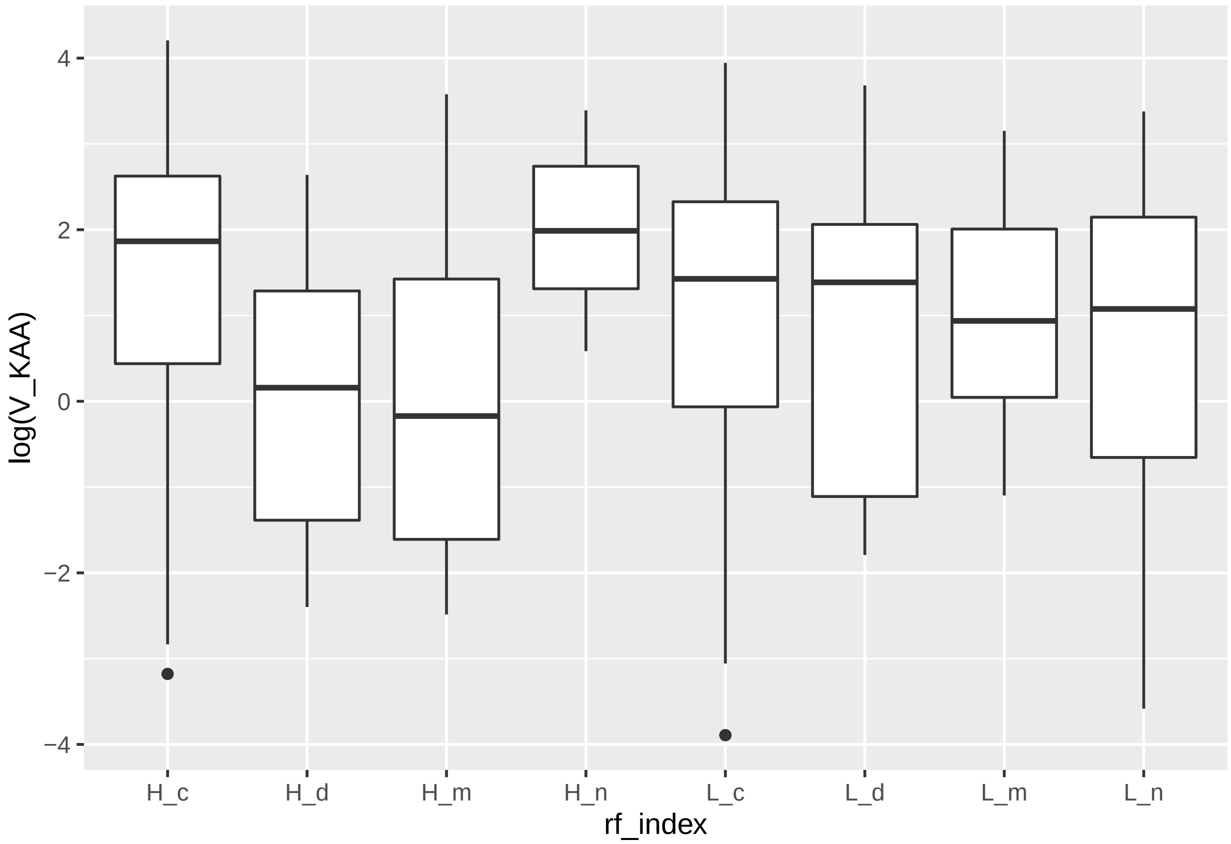


```r
## Repeating for Kalmia angustifolia

  # Plotting:

  KAA <- data.frame(Y_mat$V_KAA,X_mat$PRECU,X_mat$rf_index)

  colnames(KAA) <- c("V_KAA","PRECU","rf_index")


  ggplot(data=KAA, aes(x=rf_index, y=V_KAA)) +

    geom_boxplot()
```

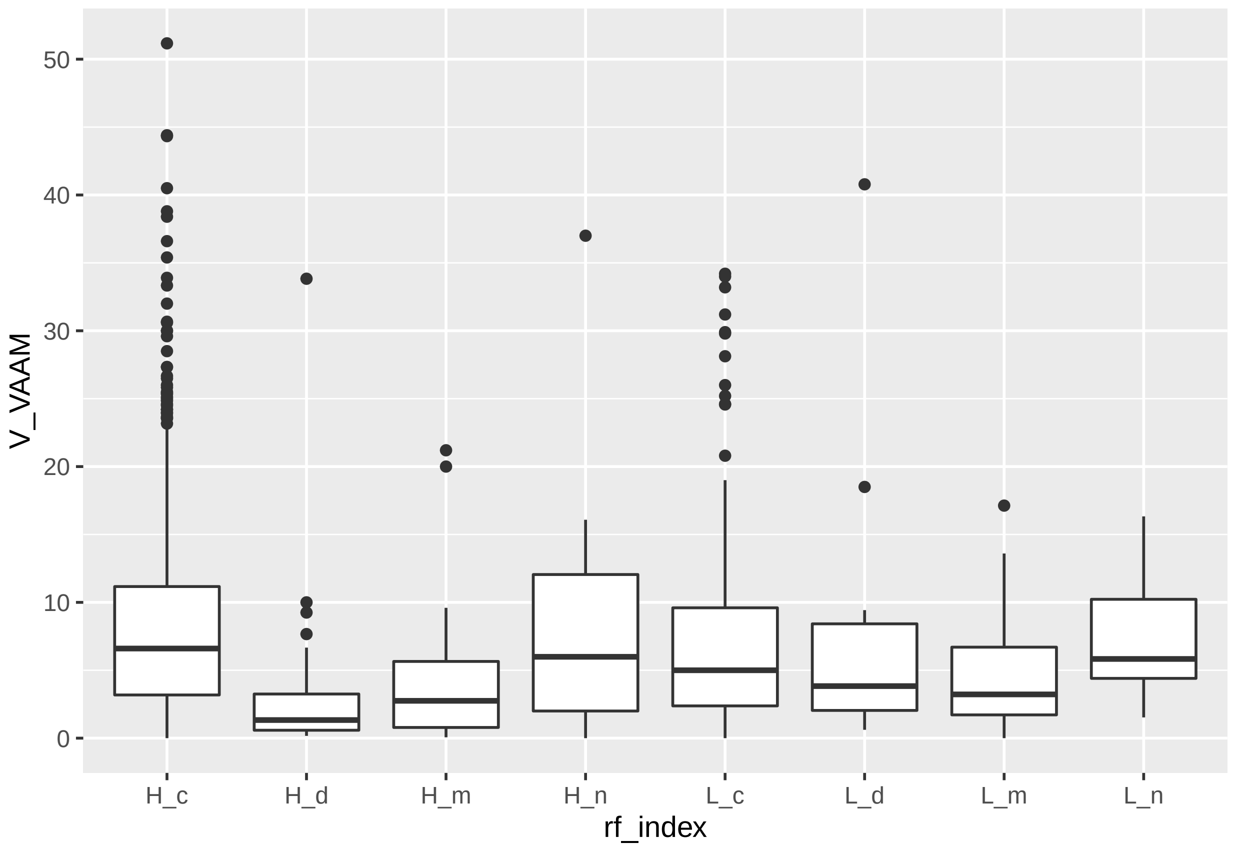


```r
ggplot(data=KAA, aes(x=rf_index, y=log(V_KAA))) + # log transforming the y axis
  geom_boxplot()
```

```
## Warning: Removed 220 rows containing non-finite values (stat_boxplot).
```

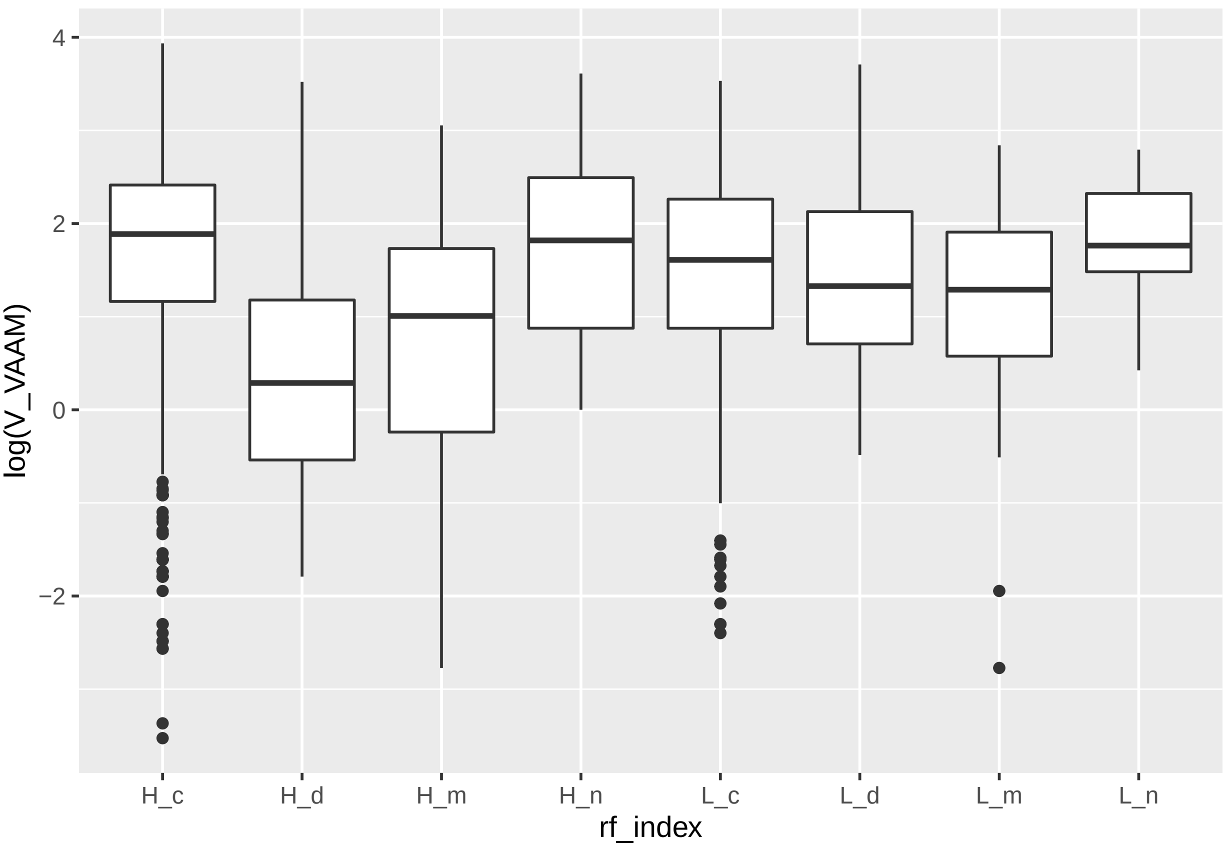


```
## Repeating for Vaccinium sp.
    VAAM <- data.frame(Y_mat$V_VAAM,X_mat$PRECU,X_mat$rf_index)
    colnames(VAAM) <- c("V_VAAM","PRECU","rf_index")


    ggplot(data=VAAM, aes(x=rf_index, y=V_VAAM)) +
      geom_boxplot()
```



```
ggplot(data=VAAM, aes(x=rf_index, y=log(V_VAAM))) + # log transforming the y axis
  geom_boxplot()
```

```
## Warning: Removed 8 rows containing non-finite values (stat_boxplot).
```



I then combined all of the above plots into one single chart for easier comparison.

Statistical analysis

I used a Kruksal-Wallis test to determine if any any groups were significantly different from the rest, or in other words, if at least one rainfall forest cateogry contains a significantly different percent cover of a given species than the other groups. For all species, p <0.05 indicating we can reject the null hypothesis that percent cover is the same across rainfall forest type categories for all species. To further assess the relationships, I performed a pairwise wilcox test to determine whether percent cover differed significantly from each other in pairs. This was not in the initial pre-registration but is an extended step to further test the hypothesis.

Results

# CAL significant differences (p< 0.05): H_d vs H_c ; H_m vs H_d ; L_c vs H_d ; L_m vs H_c ; L_m vs L_c

# KAA significant differences (P<0.05): H_c vs H_d ; H_c vs H_m ; H_c vs L_c ; H_d vs H_n ; H_m vs H_n ; H_m vs L_m ; H_n vs L_c ; H_n vs L_m

# RHGsignificant differences (P<0.05): all types with H_c; H_d vs H_m ; H_d vs H_n ; H_d vs L_c ; L_c vs L_d ; L_c vs L_m ; L_c vs L_n

# VAAM significant differences (P<0.05): H_c vs H_d; H_c vs H_m; H_c vs L_c ; H_c vs L_m; H_d vs L_c; H_d vs L_n ; H_m vs L_c; H_m vs L_n

Discussion

References

Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., *et al.* (2022). *Rmarkdown: Dynamic documents for r*.

Aust, F. & Spitzer, L. (2022). *Prereg: R markdown templates to preregister scientific studies*.

Müller, K. (2020). *Here: A simpler way to find your files*.

Papadakis, M., Tsagris, M., Dimitriadis, M., Fafalios, S., Tsamardinos, I., Fasiolo, M., *et al.* (2022). *Rfast: A collection of efficient and extremely fast r functions*.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rodríguez-Sánchez, F., Jackson, C.P. & Hutchins, S.D. (2022). *Grateful: Facilitate citation of r packages*.

Simonsohn, U. & Gruson, H. (2022). *Groundhog: Version-control for CRAN, GitHub, and GitLab packages*.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., *et al.* (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4, 1686.

Xie, Y. (2014). Knitr: A comprehensive tool for reproducible research in R. In: *Implementing reproducible computational research* (eds. Stodden, V., Leisch, F. & Peng, R.D.). Chapman; Hall/CRC.

Xie, Y. (2015). *Dynamic documents with R and knitr*. 2nd edn. Chapman; Hall/CRC, Boca Raton, Florida.

Xie, Y. (2022). *Knitr: A general-purpose package for dynamic report generation in r*.

Xie, Y., Allaire, J.J. & Grolemund, G. (2018). *R markdown: The definitive guide*. Chapman; Hall/CRC, Boca Raton, Florida.

Xie, Y., Dervieux, C. & Riederer, E. (2020). *R markdown cookbook*. Chapman; Hall/CRC, Boca Raton, Florida.

```r
grateful::cite_packages(output = "paragraph")
```

We used R version 4.2.1 (R Core Team 2022) and the following R packages: grateful v. 0.1.11 (Rodríguez-Sánchez *et al.* 2022), groundhog v. 2.0.1 (Simonsohn & Gruson 2022), here v. 1.0.1 (Müller 2020), knitr v. 1.40 (Xie 2014, 2015, 2022), prereg v. 0.6.0 (Aust & Spitzer 2022), Rfast v. 2.0.6 (Papadakis *et al.* 2022), rmarkdown v. 2.16 (Xie *et al.* 2018, 2020; Allaire *et al.* 2022), tidyverse v. 1.3.2 (Wickham *et al.* 2019).

Tables

Table 1. Mean body mass of penguins on different islands over time.

Note: for the word version the `kable` does not output correctly. You could try using `flextable` package instead: https://taehoonh.me/content/post/alternative-to-kable-function-when-knitting-to-ms-word.html

“‘{#{r table_01, message=FALSE, warning=FALSE, include=FALSE, paged.print=TRUE}

without the “hold_position” the table ends up at the top of the page

~~~
\newpage

# Figure Captions

**Figure 1**. Pretty coloured dots about penguins

**Figure 2**. Wow, even prettier plot about penguins that shows stuff

\newpage

# Figures

![]()

Figure 1.

\newpage

```{#{r figure_02, echo=FALSE, message=FALSE, warning=FALSE}
~~~

Figure 2.

Figure 3.

```
{#{r file="../scripts/figure_03.R"} # this code chunk calls an external script to generate
the plot. If you # want to load or run code not for a plot you could also use the following.
# Note that it does not seem to work for plots. # source("../scripts/figure_03.R", local
= knitr::knit_global())
```

Appendices