Final Project - EEB590C

Devin Molnau and Elizabeth McMurchie

April 29, 2021

Homework 3

This assignment is due prior to the last day class. You are to self-select and work in groups: 2-3 in a group. For the assignment below submit one R-script. Annotations via comments are highly encouraged. The script should run!

Assignment:

1: Obtain a dataset. This may be one of your own, a dataset from DRYAD, or some other source. Identify hypotheses for this dataset and analyze patterns in the data. You may use any methods learned during the semester, but at least one analysis must come from material learned in weeks 11-13.

USE COMMENTS IN THE R CODE to describe what the patterns you find represent.

```
## Loading required package: rgl

## Loading required package: Matrix

## -- Attaching packages -------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.6     v dplyr   1.0.4
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x tidyr::expand() masks Matrix::expand()
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x tidyr::pack()   masks Matrix::pack()
## x tidyr::unpack() masks Matrix::unpack()

## Loading required package: permute

## Loading required package: lattice

## This is vegan 2.5-7

##
## Attaching package: 'mice'

## The following object is masked from 'package:stats':
##
##     filter

## The following objects are masked from 'package:base':
##
##     cbind, rbind
```

TODO

INTRO

```
#READ IN DATA AS DATAFRAME
mydata<-read_excel(path="data/TransformGuaduaSet.xlsx", col_names = TRUE, na="x")
```

Preprocessing

Remove Column V and Y

Remove column V and Y, because there is missing data and we cannot impute the data because it just doesn't make sense to attempt to predict these columns.

V. Adaxial: Frequency of stomates if present on the adaxial surface of foliage leaf blades: 0 = common; 1 = infrequent. Y. Adaxial: Papillae on long cells of the intercostal zone adjacent to the stomates: 0 = not overarching the stomates; 1 = overarching the stomates.

```
to_drop<-c("Adaxial: Papillae on long cells of the intercostal zone adjacent to the stomates:  0 = not

df=data.frame(mydata[,!( names(mydata) %in% to_drop)],stringsAsFactors = TRUE)
```

Make sure that the columns are what they should be

```
df<-data.frame(lapply(df,as.factor))
str(df)
```

```
## 'data.frame':    55 obs. of  36 variables:
## $ Sample
## $ Country
## $ Species
## $ Group_1
## $ Group_2
## $ Habitat
## $ Abaxial.papillae.on.the.long.cells.in.the.stomatal.zone.0...absent.1...present
## $ Abaxial.papillae.on.the.long.cells.in.the.interstomatal.zone...0...absent..1...present.
## $ Abaxial..Prickle.hairs..0...absent..1...present
## $ Abaxial..Bicellular.microhairs..0...absent..1...present
## $ Abaxial..Macrohairs..0...absent..1...present
## $ Abaxial..Dumbbell.shaped.silica.bodies.in.the.intercostal.zone...0...present..1...absent
## $ Abaxial..Saddle.shaped.silica.bodies.in.the.intercostal.zone...0...present..1...absent
## $ Abaxial..Crenate..elongated..silica.bodies.in.the.intercostal.zone...0...present..1...absent.
## $ Abaxial..Dumbbell.shaped.silica.cells.in.the.costal.zone...0...present..1...absent.
## $ Abaxial..Saddle.shaped.silica.cells.in.the.costal.zone...0...present..1...absent.
## $ Abaxial..Crenate..elongated..silica.cells.in.the.costal.zone...0...present..1...absent.
## $ Abaxial..Triangular.stomata.subsidiary.cells..0...absent..1...present
## $ Abaxial..Dome.shaped.stomata.subsidiary.cells..0...absent..1...present
## $ Abaxial..Parallel.sided.stomata.subsidiary.cells..0...absent..1...present
## $ Adaxial..stomates.on.the.adaxial.surface.of.foliage.leaf.blades...0...absent..1...present.
## $ Adaxial.papillae.on.the.long.cells.in.the.stomatal.zone.exclusive.of.bulliform.cells.0...absent.1
## $ Adaxial.papillae.on.the.long.cells.in.the.interstomatal.zone...0...absent..1...present.
## $ Adaxial.papillae.on.the.bulliform.cells...0...absent..1...present.
## $ Adaxial..Prickle.hairs..0...absent..1...present
## $ Adaxial..Bicellular.microhairs..0...absent..1...present
## $ Adaxial..Macrohairs..0...absent..1...present
## $ Adaxial..Dumbbell.shaped.silica.bodies.in.the.intercostal.zone...0...present..1...absent
```

```
##  $ Adaxial..Saddle.shaped.silica.bodies.in.the.intercostal.zone...0...present..1...absent
##  $ Adaxial..Crenate..elongated..silica.bodies.in.the.intercostal.zone...0...present..1...absent.
##  $ Adaxial..Dumbbell.shaped.silica.cells.in.the.costal.zone...0...present..1...absent.
##  $ Adaxial..Saddle.shaped.silica.cells.in.the.costal.zone...0...present..1...absent.
##  $ Adaxial..Crenate..elongated..silica.cells.in.the.costal.zone...0...present..1...absent.
##  $ Adaxial..Triangular.stomata.subsidiary.cells..0...absent..1...present
##  $ Adaxial..Dome.shaped.stomata.subsidiary.cells..0...absent..1...present
##  $ Adaxial..Parallel.sided.stomata.subsidiary.cells..0...absent..1...present
```

Imputation of missing data

In the last three columns, in some species the papillae of the cells adjacent to the stomata obscured theshape of the subsidiary cells. Therefore, the shapes of these subsidiary cells resulted in missing data that we imputed using the MICE package's logistic regression method.

```r
init = mice(df,maxit=0)
meth = init$method
predM=init$predictorMatrix
meth[c(colnames(df[,7:length(df)]))]<-"logreg"
set.seed(183)
imputed<-mice(df,method = meth, predictorMatrix = predM, m=5,printFlag = FALSE)
```

```
## Warning: Number of logged events: 90
```

```r
imputed<-complete(imputed)
```

```r
sapply(imputed,function(x) sum(is.na(x))) #double check to make sure there are no more NA
```

```
##
##
##                                                                                            (
##
##                                                                                            S
##
##                                                                                            (
##
##                                                                                            (
##
##                                                                                            N
##
##                          Abaxial.papillae.on.the.long.cells.in.the.stomatal.zone.0...absent.1...p
##
##                Abaxial.papillae.on.the.long.cells.in.the.interstomatal.zone...0...absent..1...pr
##
##                                                      Abaxial..Prickle.hairs..0...absent..1...p
##
##                                                  Abaxial..Bicellular.microhairs..0...absent..1...p
##
##                                                        Abaxial..Macrohairs..0...absent..1...p
##
##              Abaxial..Dumbbell.shaped.silica.bodies.in.the.intercostal.zone...0...present..1..
##
##                Abaxial..Saddle.shaped.silica.bodies.in.the.intercostal.zone...0...present..1..
##
##            Abaxial..Crenate..elongated..silica.bodies.in.the.intercostal.zone...0...present..1...a
##
```

```
##                               Abaxial..Dumbbell.shaped.silica.cells.in.the.costal.zone...0...present..1...a
##
##                               Abaxial..Saddle.shaped.silica.cells.in.the.costal.zone...0...present..1...a
##
##                          Abaxial..Crenate..elongated..silica.cells.in.the.costal.zone...0...present..1...a
##
##                                    Abaxial..Triangular.stomata.subsidiary.cells..0...absent..1...p
##
##                                     Abaxial..Dome.shaped.stomata.subsidiary.cells..0...absent..1...p
##
##                                 Abaxial..Parallel.sided.stomata.subsidiary.cells..0...absent..1...p
##
##              Adaxial..stomates.on.the.adaxial.surface.of.foliage.leaf.blades...0...absent..1...pr
##
## Adaxial.papillae.on.the.long.cells.in.the.stomatal.zone.exclusive.of.bulliform.cells.0...absent.1...p
##
##                  Adaxial.papillae.on.the.long.cells.in.the.interstomatal.zone...0...absent..1...p
##
##                                       Adaxial.papillae.on.the.bulliform.cells...0...absent..1...p
##
##                                                    Adaxial..Prickle.hairs..0...absent..1...p
##
##                                          Adaxial..Bicellular.microhairs..0...absent..1...p
##
##                                                  Adaxial..Macrohairs..0...absent..1...p
##
##              Adaxial..Dumbbell.shaped.silica.bodies.in.the.intercostal.zone...0...present..1..
##
##                Adaxial..Saddle.shaped.silica.bodies.in.the.intercostal.zone...0...present..1..
##
##           Adaxial..Crenate..elongated..silica.bodies.in.the.intercostal.zone...0...present..1...a
##
##                  Adaxial..Dumbbell.shaped.silica.cells.in.the.costal.zone...0...present..1...a
##
##                    Adaxial..Saddle.shaped.silica.cells.in.the.costal.zone...0...present..1...a
##
##              Adaxial..Crenate..elongated..silica.cells.in.the.costal.zone...0...present..1...a
##
##                                   Adaxial..Triangular.stomata.subsidiary.cells..0...absent..1...p
##
##                                    Adaxial..Dome.shaped.stomata.subsidiary.cells..0...absent..1...p
##
##                                Adaxial..Parallel.sided.stomata.subsidiary.cells..0...absent..1...p
##
```

Analysis

PCOA

```
#TODO
?dist.binary() #Distance matrix for binary is simple matching coefficient
```

```
## starting httpd help server ... done
```

STICK PLOT of PCOA

```
#TODO
#example
#screeplot(pca.bumpus,bstick = TRUE)
```

FACTORIAL ANOVA

```
#TODO
```

INTERACTIONS

```
#TODO
#### Group VS country
#### Group vs habitat
```

MODEL COMPARISON USING LIKELIHOOD RATIO TEST (LTR)

```
#TODO
```