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
```

```
#READ IN DATA AS DATAFRAME
?read_excel
```

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

```
data<-read_excel(path="data/TransformGuaduaSet.xlsx", col_names = TRUE, na="x")
```

TODO

Imputation to add in missing data –> Mice package , logistic regression

Distance matrix for binary is simple matching coefficient

Example

```
data(sipoo)



## Some useful measures

#http://pbil.univ-lyon1.fr/ADE-4/ade4-html/dist.binary.html



# Simple Matching Coefficient

# Shared presence and shared absence (as proportion of total)

dist.binary(sipoo,method = 2)
```

```
##           Svartholm L.Hogholm Ledholmen S.Hogholm Torrvedsh  SkataLed Flakaskar
## L.Hogholm 0.2828427
## Ledholmen 0.2449490 0.2449490
## S.Hogholm 0.2828427 0.2828427 0.3162278
## Torrvedsh 0.2828427 0.2828427 0.2449490 0.2828427
## SkataLed  0.4000000 0.2828427 0.3741657 0.2828427 0.2828427
## Flakaskar 0.3464102 0.3464102 0.3741657 0.2000000 0.2828427 0.2828427
## S.farholm 0.4000000 0.4000000 0.3741657 0.4000000 0.4000000 0.4000000 0.4472136
## Farholmn  0.4000000 0.3464102 0.4242641 0.2828427 0.4000000 0.2828427 0.3464102
## Asplandet 0.4472136 0.4000000 0.4690416 0.3464102 0.4000000 0.3464102 0.3464102
## Granlndet 0.5099020 0.4690416 0.4898979 0.4690416 0.4242641 0.3741657 0.4690416
## Hanskholm 0.4000000 0.4000000 0.3741657 0.4000000 0.2828427 0.4000000 0.4000000
## Ragskar   0.4472136 0.4472136 0.4242641 0.4472136 0.3464102 0.4000000 0.4472136
## Trutland  0.4898979 0.4898979 0.4242641 0.5291503 0.4472136 0.5291503 0.5291503
## Kaivokari 0.5656854 0.5291503 0.5477226 0.5291503 0.4898979 0.4472136 0.5291503
## Mustahevo 0.5830952 0.5830952 0.5656854 0.5477226 0.5099020 0.5099020 0.5477226
## Kaunissri 0.7483315 0.7211103 0.7071068 0.7483315 0.6928203 0.6928203 0.7483315
## Onas      0.8000000 0.7745967 0.7874008 0.8000000 0.7483315 0.7483315 0.7745967
##           S.farholm  Farholmn Asplandet Granlndet Hanskholm   Ragskar  Trutland
## L.Hogholm
## Ledholmen
## S.Hogholm
## Torrvedsh
## SkataLed
```

```
## Flakaskar
## S.farholm
## Farholmn  0.3464102
## Asplandet 0.4472136 0.3464102
## Granlndet 0.4242641 0.4242641 0.4242641
## Hanskholm 0.4000000 0.4000000 0.4472136 0.4690416
## Ragskar   0.4472136 0.4000000 0.4898979 0.4690416 0.3464102
## Trutland  0.4898979 0.5656854 0.5291503 0.5477226 0.4472136 0.4898979
## Kaivokari 0.5291503 0.4472136 0.5291503 0.5099020 0.4898979 0.4000000 0.5656854
## Mustahevo 0.5830952 0.5099020 0.5830952 0.5291503 0.5099020 0.3741657 0.5830952
## Kaunissri 0.6633250 0.6928203 0.7211103 0.6782330 0.6324555 0.6324555 0.6633250
## Onas      0.7745967 0.7483315 0.8000000 0.7615773 0.7211103 0.6633250 0.7745967
##           Kaivokari Mustahevo Kaunissri
## L.Hogholm
## Ledholmen
## S.Hogholm
## Torrvedsh
## SkataLed
## Flakaskar
## S.farholm
## Farholmn
## Asplandet
## Granlndet
## Hanskholm
## Ragskar
## Trutland
## Kaivokari
## Mustahevo 0.4690416
## Kaunissri 0.6324555 0.5830952
## Onas      0.6324555 0.6480741 0.6324555
```

```
#Jaccard distance

# Uses shared presence only (as proportion of total)

dist.binary(sipoo,method = 1)
```

```
##           Svartholm L.Hogholm Ledholmen S.Hogholm Torrvedsh  SkataLed Flakaskar
## L.Hogholm 0.8164966
## Ledholmen 0.7071068 0.7071068
## S.Hogholm 0.8944272 0.8944272 0.9128709
## Torrvedsh 0.7559289 0.7559289 0.6546537 0.8164966
## SkataLed  0.9428090 0.7559289 0.8819171 0.8164966 0.7071068
## Flakaskar 0.9258201 0.9258201 0.9354143 0.7071068 0.7559289 0.7559289
## S.farholm 0.8528029 0.8528029 0.7977240 0.8944272 0.8164966 0.8164966 0.9128709
## Farholmn  0.9428090 0.8660254 0.9486833 0.8164966 0.8944272 0.7071068 0.8660254
## Asplandet 0.9534626 0.8944272 0.9574271 0.8660254 0.8528029 0.7745967 0.8164966
## Granlndet 0.9309493 0.8864053 0.8944272 0.9198662 0.8017837 0.7337994 0.8864053
## Hanskholm 0.8528029 0.8528029 0.7977240 0.8944272 0.6324555 0.8164966 0.8528029
## Ragskar   0.8770580 0.8770580 0.8320503 0.9128709 0.7071068 0.7844645 0.8770580
## Trutland  0.8944272 0.8944272 0.8017837 0.9660918 0.8164966 0.9074852 0.9354143
## Kaivokari 0.9428090 0.9074852 0.9128709 0.9354143 0.8401681 0.7905694 0.9074852
## Mustahevo 0.9459053 0.9459053 0.9176629 0.9393364 0.8498366 0.8498366 0.9128709
## Kaunissri 0.9503819 0.9309493 0.9128709 0.9660918 0.8944272 0.8944272 0.9503819
## Onas      0.9561829 0.9393364 0.9411239 0.9701425 0.9074852 0.9074852 0.9393364
```

```
##           S.farholm  Farholmn Asplandet Granlndet Hanskholm   Ragskar  Trutland
## L.Hogholm
## Ledholmen
## S.Hogholm
## Torrvedsh
## SkataLed
## Flakaskar
## S.farholm
## Farholmn  0.7385489
## Asplandet 0.8451543 0.7745967
## Granlndet 0.7500000 0.8017837 0.7745967
## Hanskholm 0.7559289 0.8164966 0.8451543 0.8043997
## Ragskar   0.7905694 0.7844645 0.8660254 0.7817360 0.6546537
## Trutland  0.8164966 0.9428090 0.8819171 0.8451543 0.7669650 0.7947194
## Kaivokari 0.8366600 0.7905694 0.8583951 0.7867958 0.7947194 0.6666667 0.8340577
## Mustahevo 0.8790491 0.8498366 0.8997354 0.7977240 0.8062258 0.6236096 0.8416254
## Kaunissri 0.8424235 0.8944272 0.9013878 0.8348471 0.8164966 0.8032193 0.8164966
## Onas      0.9004503 0.9074852 0.9299811 0.8735891 0.8618916 0.8043997 0.8770580
##           Kaivokari Mustahevo Kaunissri
## L.Hogholm
## Ledholmen
## S.Hogholm
## Torrvedsh
## SkataLed
## Flakaskar
## S.farholm
## Farholmn
## Asplandet
## Granlndet
## Hanskholm
## Ragskar
## Trutland
## Kaivokari
## Mustahevo 0.7071068
## Kaunissri 0.7784989 0.7288690
## Onas      0.7559289 0.7637626 0.6900656
```

Remove Column V and Y

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. #### Fill in empty cells with NA