

pdf

Chase Clark

January 17, 2018

Load dendrogram cached file for testing with:

```
library("tidyverse")
```

```
## -- Attaching packages -----
## v ggplot2 2.2.1      v purrr   0.2.4
## v tibble  1.4.1      v dplyr   0.7.4
## v tidyr   0.7.2      v stringr 1.2.0
## v readr   1.1.1      v forcats 0.2.0

## Warning: package 'ggplot2' was built under R version 3.4.3
## Warning: package 'tibble' was built under R version 3.4.3
## Warning: package 'tidyr' was built under R version 3.4.3
## Warning: package 'readr' was built under R version 3.4.3
## Warning: package 'purrr' was built under R version 3.4.3
## Warning: package 'dplyr' was built under R version 3.4.3
## Warning: package 'stringr' was built under R version 3.4.3
## Warning: package 'forcats' was built under R version 3.4.3

## -- Conflicts -----
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library("dendextend")
```

```
## Warning: package 'dendextend' was built under R version 3.4.3

##
## -----
## Welcome to dendextend version 1.6.0
## Type citation('dendextend') for how to cite the package.
##
## Type browseVignettes(package = 'dendextend') for the package vignette.
## The github page is: https://github.com/talgalili/dendextend/
##
## Suggestions and bug-reports can be submitted at: https://github.com/talgalili/dendextend/issues
## Or contact: <tal.galili@gmail.com>
##
## To suppress this message use: suppressPackageStartupMessages(library(dendextend))
## -----

##
## Attaching package: 'dendextend'

## The following object is masked from 'package:stats':
##
##      cutree
```

```
dend<-readRDS("Data/Distance-cosineD_Clustering-ward.D2_UsedPresenceAbsence_SNR-4_PercentPresence-70_Lo")

# dend<-dend %>% set("labels_color", "white")

dend<-dend %>% set("labels_cex", .65)
```

Read csv of groups

```
groupFile<-read_csv("data/b.csv")
```

```
## Parsed with column specification:
## cols(
##   Sample_location = col_integer(),
##   Sample_type_1 = col_character(),
##   `Sample name` = col_character(),
##   Media_type = col_character(),
##   Treatment = col_character(),
##   Sample_type_2 = col_character()
## )
```

Load colored_Dots.R function

```
dendrogramLabels<-as_tibble(labels(dend))
names(dendrogramLabels)<-'Sample name'

#join but keep prder of dendrogram label
joinedData<-left_join(dendrogramLabels,groupFile,by="Sample name")

naReplaceValues<-as.list(sapply(names(joinedData),function(x)paste0("Missing ",x)))

joinedData<-joinedData %>% replace_na(replace=naReplaceValues)

colsel<-'Sample_type_2'

small<-bind_cols(joinedData[,1],joinedData[colsel])

#w<-small %>% group_by(.dots=paste0(colsel))

groupedList<-split(small,factor(small[colsel][[1]]))

bigList<-lapply(1:length(groupedList),function(x)left_join(dendrogramLabels,groupedList[[x]],by='Sample
labels(bigList)<-labels(groupedList)

for(x in 1:length(bigList)){
  bigList[[x]][colsel][!is.na(bigList[[x]][colsel])]<-"#000000"
  bigList[[x]][colsel][is.na(bigList[[x]][colsel])<-"#00000000"
}

a<-which(bigList$Marine_water$Sample_type_2 == "#000000")
b<-joinedData[a,]
```

Now we we have a list. Each list item represents a group/column from excel table. These are associated to the entire dendrogram, by label and are in the same order as the dendrogram from top to bottom. 0 means a label is not part of the group, 1 means it is part of the group

`/#000000` is black `/#00000000` is black but transparent in case we want color supprt later

```
bigMatrix<-NULL
for (i in 1:length(bigList)){
  bigMatrix<-bind_cols(bigMatrix,bigList[[i]][,2])}

names(bigMatrix)<-names(bigList)
```

The below chunk will only run when the rmd file is run with R (won't work if viewing static html).

```
par(mar = c(8,3,8,2))

plot(dend,hORIZ=T)

colored_dots(bigMatrix, dend,
  rowLabels = names(bigMatrix),horiz=T,sort_by_labels_order = FALSE)
```

