NM_Envr_Data

2024-03-21

Set working directory

```r
#set working directory to Rmd file location
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
```

```r
#read in appropriate packages
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(readr)
library(forcats)
library(RColorBrewer)
library(circlize)
```

```
## ========================================
## circlize version 0.4.16
## CRAN page: https://cran.r-project.org/package=circlize
## Github page: https://github.com/jokergoo/circlize
## Documentation: https://jokergoo.github.io/circlize_book/book/
##
## If you use it in published research, please cite:
## Gu, Z. circlize implements and enhances circular visualization
##   in R. Bioinformatics 2014.
##
## This message can be suppressed by:
##   suppressPackageStartupMessages(library(circlize))
## ========================================
```

```r
library(BiocManager)
```

```
## Bioconductor version '3.19' is out-of-date; the current release version '3.20'
##   is available with R version '4.4'; see https://bioconductor.org/install
```

```r
library(ComplexHeatmap)
```

```
## Loading required package: grid
```

```
## ========================================
## ComplexHeatmap version 2.20.0
## Bioconductor page: http://bioconductor.org/packages/ComplexHeatmap/
## Github page: https://github.com/jokergoo/ComplexHeatmap
## Documentation: http://jokergoo.github.io/ComplexHeatmap-reference
##
## If you use it in published research, please cite either one:
## - Gu, Z. Complex Heatmap Visualization. iMeta 2022.
## - Gu, Z. Complex heatmaps reveal patterns and correlations in multidimensional
##     genomic data. Bioinformatics 2016.
##
##
## The new InteractiveComplexHeatmap package can directly export static
## complex heatmaps into an interactive Shiny app with zero effort. Have a try!
##
## This message can be suppressed by:
##   suppressPackageStartupMessages(library(ComplexHeatmap))
## ========================================
```

```r
library(vegan)#for Mantel test
```

```
## Loading required package: permute
```

```
## Loading required package: lattice
```

```
## This is vegan 2.6-8
```

All the data for each site are in separate csv files. There is one temp and one salinity csv for each site. We need to combine those files into one salinity data frame and one temperature data frame, and then join the salinity and temperature data frames into one to conduct the distance analysis.

Combining temp and salinity csvs respectively from each site

```r
#set working directory to temperature files location
setwd("../../data/envr_of_origin/Temp")
# read in all temp files as csvs and combine into one file named "temp"
temp <- list.files(path=".") %>%
  lapply(read_csv) %>%
  bind_rows
```

```
## Rows: 1 Columns: 14
## -- Column specification ------------------------------------------------------
## Delimiter: ","
## chr  (4): site_name, download_date, source_description, collection_type
## dbl (10): lat, lon, firstyear, finalyear, Mean_Annual_Temperature_C, Mean_ma...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## New names:
## Rows: 1 Columns: 15
## -- Column specification ------------------------------------------------------
## Delimiter: ","
## chr  (4): site_name, download_date, source_description, collection_type
## dbl (11): ...1, lat, lon, firstyear, finalyear, Mean_Annual_Temperature_C, M...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 1 Columns: 14
## -- Column specification ------------------------------------------------------
## Delimiter: ","
## chr  (4): site_name, download_date, source_description, collection_type
## dbl (10): lat, lon, firstyear, finalyear, Mean_Annual_Temperature_C, Mean_ma...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## New names:
## Rows: 1 Columns: 15
## -- Column specification ------------------------------------------------------
## Delimiter: ","
## chr  (4): site_name, download_date, source_description, collection_type
## dbl (11): ...1, lat, lon, firstyear, finalyear, Mean_Annual_Temperature_C, M...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 1 Columns: 14
## -- Column specification ------------------------------------------------------
## Delimiter: ","
## chr  (4): site_name, download_date, source_description, collection_type
## dbl (10): lat, lon, firstyear, finalyear, Mean_Annual_Temperature_C, Mean_ma...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## New names:
## Rows: 1 Columns: 15
## -- Column specification ------------------------------------------------------
## Delimiter: ","
## chr  (4): site_name, download_date, source_description, collection_type
## dbl (11): ...1, lat, lon, firstyear, finalyear, Mean_Annual_Temperature_C, M...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## New names:
## Rows: 1 Columns: 15
## -- Column specification ------------------------------------------------------
```

```
## Delimiter: ","
## chr  (4): site_name, download_date, source_description, collection_type
## dbl (11): ...1, lat, lon, firstyear, finalyear, Mean_Annual_Temperature_C, M...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
## New names:
## Rows: 1 Columns: 15
## -- Column specification -------------------------------------------------------
## Delimiter: ","
## chr  (4): site_name, download_date, source_description, collection_type
## dbl (11): ...1, lat, lon, firstyear, finalyear, Mean_Annual_Temperature_C, M...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```r
#View temp file
View(temp)
#set working directory to Salinity folder
setwd("../Salinity")
# read in all salinity files as csvs and combine into one file named "sal"
sal <- list.files(path=".") %>%
  lapply(read_csv) %>%
  bind_rows
```

```
## Rows: 1 Columns: 14
## -- Column specification -------------------------------------------------------
## Delimiter: ","
## chr  (4): site_name, download_date, source_description, collection_type
## dbl (10): lat, lon, firstyear, finalyear, Mean_Annual_Salinity_ppt, Mean_max...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
## Rows: 1 Columns: 14
## -- Column specification -------------------------------------------------------
## Delimiter: ","
## chr  (4): site_name, download_date, source_description, collection_type
## dbl (10): lat, lon, firstyear, finalyear, Mean_Annual_Salinity_ppt, Mean_max...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
## Rows: 1 Columns: 14
## -- Column specification -------------------------------------------------------
## Delimiter: ","
## chr  (4): site_name, download_date, source_description, collection_type
## dbl (10): lat, lon, firstyear, finalyear, Mean_Annual_Salinity_ppt, Mean_max...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
## Rows: 1 Columns: 14
## -- Column specification -------------------------------------------------------
## Delimiter: ","
## chr  (4): site_name, download_date, source_description, collection_type
## dbl (10): lat, lon, firstyear, finalyear, Mean_Annual_Salinity_ppt, Mean_max...
```

```
## 
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
## Rows: 1 Columns: 14
## -- Column specification ------------------------------------------------------
## Delimiter: ","
## chr  (4): site_name, download_date, source_description, collection_type
## dbl (10): lat, lon, firstyear, finalyear, Mean_Annual_Salinity_ppt, Mean_max...
## 
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
## Rows: 1 Columns: 14
## -- Column specification ------------------------------------------------------
## Delimiter: ","
## chr  (4): site_name, download_date, source_description, collection_type
## dbl (10): lat, lon, firstyear, finalyear, Mean_Annual_Salinity_ppt, Mean_max...
## 
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
## New names:
## Rows: 1 Columns: 15
## -- Column specification ------------------------------------------------------
## Delimiter: ","
## chr  (4): site_name, download_date, source_description, collection_type
## dbl (11): ...1, lat, lon, firstyear, finalyear, Mean_Annual_Salinity_ppt, Me...
## 
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
## New names:
## Rows: 1 Columns: 15
## -- Column specification ------------------------------------------------------
## Delimiter: ","
## chr  (4): site_name, download_date, source_description, collection_type
## dbl (11): ...1, lat, lon, firstyear, finalyear, Mean_Annual_Salinity_ppt, Me...
## 
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```r
View(sal)

#set working directory to location where code is saved
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

#remove extra column at the end of each data frame
ncol(temp) #ncol = 15
```

```
## [1] 15
```

```r
temp <- temp[ , -c(15)]

ncol(sal) #ncol = 15
```

```
## [1] 15
```

```r
sal <- sal[ , -c(15)]
```

Join temp and sal data frames

```r
#left join temp and salinity dfs and save into an object called "envr"
envr <-temp %>%
          left_join(sal, by = "site_name")

#remove rows unnecessary for distance analysis and save into an object called "envr_sub"
envr_sub <- subset(envr, select = -c(download_date.x, firstyear.x, finalyear.x, source_description.x, la

# make envr_sub into a data frame and save as "envr_df"
envr.df <- data.frame(envr_sub)

#update site names with labeling scheme
envr.df["site_name"][envr.df["site_name"] == "TX2"] <- "W1-TX"
envr.df["site_name"][envr.df["site_name"] == "LA4"] <- "W2-LA"
envr.df["site_name"][envr.df["site_name"] == "FL3"] <- "W3-FL"
envr.df["site_name"][envr.df["site_name"] == "LOLA"] <- "S1-LOLA"
envr.df["site_name"][envr.df["site_name"] == "DEBY"] <- "S2-DEBY"
envr.df["site_name"][envr.df["site_name"] == "JR"] <- "W4-VA"
envr.df["site_name"][envr.df["site_name"] == "NH1"] <- "W5-NH"
envr.df["site_name"][envr.df["site_name"] == "ME2"] <- "W6-ME"

#reorder rows of data frame
envr.df <- envr.df %>%
  arrange(factor(site_name, levels = c("W1-TX", "W2-LA", "W3-FL", "S1-LOLA", "S2-DEBY", "W4-VA", "W5-NH

#save envr.df as a csv for future analyses
write.csv(envr.df, "../../data/envr_of_origin/envr_summary_stats.csv")

#rename rows with MVP population names
rownames(envr.df) <- c("W1-TX", "W2-LA", "W3-FL", "S1-LOLA", "S2-DEBY", "W4-VA", "W5-NH", "W6-ME")
#remove site names from data frame and save as object called "envr.df.dist"
envr.df.dist <- subset(envr.df, select = -c(site_name))
```

Run distance analysis and visualize with a heat map

```r
#make run distance analysis on envr.df.dist and save the results as a matrix named "envr.dist.mat"
envr.dist.mat <- as.matrix(dist(envr.df.dist))

#save envr distance matrix as csv
write.csv(envr.dist.mat, "../../data/envr_of_origin/envr_dist.csv")

#View distance matrix
View(envr.dist.mat)
```