

# Análisis de la familia de plantas *Malvaceae* en la parcela de 50 Hectáreas de la Isla Barro Colorado, Panamá.

**Carolain Pérez Ureña**     *Estudiante, Universidad Autónoma de Santo Domingo (UASD)*

---

La Isla Barro Colorado más conocida como BCI (por sus siglas en inglés) se caracteriza por ser un lugar idóneo para la realización de estudios científicos. Tal es el caso de esta investigación donde se estudió la familia de planta *Malvaceae* perteneciente a la parcela de 50 hectárea de la Isla. Este tuvo como objetivo analizar la riqueza, abundancia, asociación, agrupamiento de datos, distribución, y autocorrelación entre las variables de la familia *Malvaceae*. Así como verificar si existe una relación entre los factores ambientales (elementos del suelo y relieve) con su patrón de distribución y asociación. También busca identificar las especies consideradas como indicadoras, las especies que más contribuyeron a la diversidad (alpha y beta) y distinguir los sitios con mayor y menor abundancia. Evaluar la colinealidad entre las variables del suelo y especies con técnicas de ordenación e identificar concretamente cuáles son las especies, elementos químicos de suelo y área geomorfológica más significativos en la distribución, diversidad, autocorrelación espacial, y en su homogeneidad. Mediante métricas y con datos de censos recolectados durante años se obtuvieron los resultados esperados. De este modo, se determinó el comportamiento de la familia de plantas *Malvaceae* en la parcela, así como los factores ambientales, del relieve y elementos del suelo que intervienen en dicho comportamiento.

**Keywords:** *Malvaceae*, ecología espacial, variables ambientales, especies, BCI, *Quararibea asterolepis*

---

## 1 Introducción

Una de las áreas naturales más estudiadas por los científicos durante años ha sido la isla Barro Colorado (BCI) localizada en el lago Gatún del canal de Panamá. Su diversidad de bosques y su fauna la han convertido en un centro de estudio, útil para hacer proyectos de exploración. Es considerada como uno de los sitios con más larga historia de investigación continua en los trópicos del Nuevo Mundo, lo que ha proporcionado una base de información científica inigualada en todo el planeta. Cada año, entre 200 y 400 científicos de todo el mundo visitan el Monumento Natural Barro Colorado. (Fisher, 2016)

La mitad de la isla se encuentra cubierta de bosque joven tropical húmedo semi-perenne de 100 o más años de edad, el resto está cubierto de bosque viejo el cual ha sufrido muy pocas perturbaciones en los últimos 400 años (Moreno, 2012). Se caracteriza por tener un promedio anual de temperatura de 27° C en áreas abiertas, con una variación diurna de 9° C. Tiene una precipitación promedio anual es de 2,600 mm, con una estación lluviosa que va de mayo a diciembre, y una estación seca que comprende los meses restantes. Está constituida por un total de 265 especies de plantas, y cada una pertenece a una familia. (Croat, 1978).

Dentro de su amplia variedad de familias de plantas se encuentran las *Malvaceae*. Esta planta pertenece a la familia de las *Malvales*, que reúne cerca de 250 géneros y 3929 especies distribuidas por las regiones templadas y cálidas de todo el mundo (Rondón, 2009).

Son plantas de hierbas, arbustos o árboles, generalmente con pelos estrellados. Los tallos son de fibra de líber robusta con cavidad de mucílago. Las hojas son simples, alternas, palmadamente

divididas, palmadas veteadas, con estípulas y pecioladas. Las flores son actinomorfas, solitarias, fasciculadas o dispuestas en cimas o panículas con sépalos de 3 a 5, libres o connatos y valvados. Los pétalos son cinco, libres, giratorios, adnados a la columna estaminal en la base. Los estambres son numerosos, filamentos connados en tubos, conocidos como adelfos (Xu & Deng, 2017)(ver figura 1).



Figure 1: Flor de la planta Malvaceae

Barro Colorado ofrece ventajas excepcionales para un estudio de ecología, y la morfología de ciertos grupos de plantas podría investigarse provechosamente (Standley, 1927). Razón, por la que el presente estudio se basó en analizar el comportamiento de la familia de plantas *Malvaceae* en BCI; en el cual abarcó un análisis de abundancia, riqueza, asociación, distribución, agrupamiento, autocorrelación espacial entre las variables de la familia y la influencia que tienen los factores ambientales y geomorfológicos en su patrón de distribución y asociación.

De igual forma, con el análisis de diversidad (alpha y beta) conocer la riqueza y equidad de las especies junto con los sitios que tienen mayor y menor diversidad. Además de, conocer las especies consideradas como indicadoras, y determinar la colinealidad existente entre variables del suelo especies.

Por último, identificar cuáles son las especies, elementos del suelo y áreas morfológicas más significativos en la distribución, diversidad, autocorrelación espacial, y en su homogeneidad. Por otro lado, se busca conocer el método más eficiente para agrupar y organizar los datos a través de dendogramas. Esta investigación podrá ser utilizada como base para estudios posteriores con esta importante familia de plantas.

## 2 Metodología

### *Área de estudio y descripción de metodología*

El estudio se realizó en la Isla Barro Colorado (BCI), localizada entre los 9° 09' N y 79° 51' W, que forma parte del Monumento Natural de Barro Colorado (5,500 ha, Leigh, 1999). Formada en 1914 cuando se represó el Río Gatún como parte del trabajo para la creación del Canal de Panamá (Moreno, 2012). Es una zona administrada por el Instituto de Investigaciones Tropicales del Smithsonian dedicada a investigaciones científicas.

Dentro de Isla se encuentra la parcela de 50 Hectáreas con 1,000 metros de largo y 500 metros de ancho lo que da un total de 50 ha que se subdivide en 1 ha. En la que se llevó a cabo el presente estudio (ver figura 2).

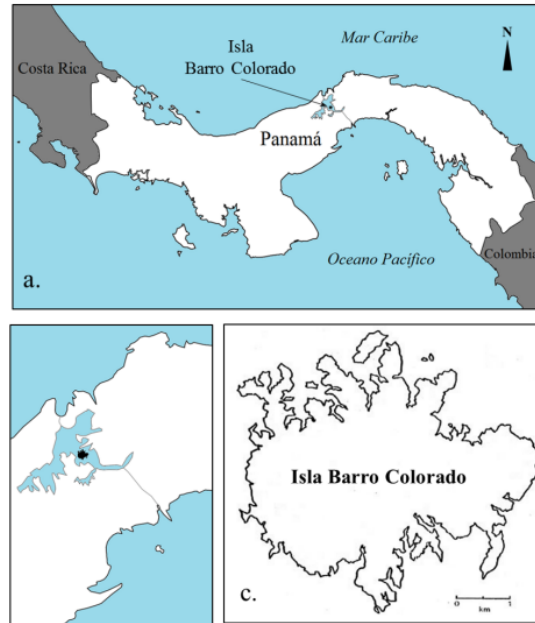


Figure 2: Mapa de la Isla Barro Colorado

La mayor parte del trabajo se desarrolló en entorno al lenguaje de programación R, en el cuál se efectuaron los análisis de agrupamientos, ambientales, ordenamiento y de ecología numérica, apoyandose con paquetes como *tidyverse* junto con *dplyr* y *vegan* con la cual se crearon las matrices de comunidad ambientales. Con el paquete *simple features* (*sf*) se crearon cuadros por hectáreas para así obtener la densidad de cada especie. En menor medida se emplearon los paquetes *adespatial*, *vegetarian*, *broom* y *cluster*. Paquetes como *ez* fueron usados para la correlación entre variables y *ggplot2*, *mapview* y *graphics* para la representación gráfica. Los datos se obtuvieron a partir del repositorio de ecología numérica de José Martínez Battle (Battle, 2020)(2020).

En la primera etapa del trabajo se efectuó un análisis ambiental de asociación estadística con los datos pre-censales de la parcela de BCI, usando una matriz de comunidad convertidas en columnas de hábitats para generar mapas de abundancia por especie, y riqueza numérica de toda la comunidad.

Para el análisis de asociación se usaron las métricas de modo Q y R; con el coeficiente de correlación de Pearson que mide la relación estadística entre dos variables. En cuanto a la medición de asociación de distancia entre sitios; se usó la transformación de la matriz de Hellinger con la metodología de la similaridad de Jaccard para obtener una matriz de distancia de comunidad transformada a la cuál se le calculó su distancia euclídea la cual indica que mientras mayor distancia menor similaridad, es decir, mientras más crece la distancia el parecido entre los sitios es cada vez menor.

La segunda fase del trabajo se basó en el análisis de agrupamiento. Con el fin de comprobar el método más adecuado para agrupar las especies en forma de dendrograma se compararon los métodos de por enlace simple, completo, de grupo de pares ponderados con media aritmética conocido como UPGMA (siglas de *unweighted pair group method with arithmetic mean*) y el método Ward teniendo como criterio la correlación cofenética.

Con los valores de abundancia de especie junto con el método de varianza mínima de agrupamiento de Ward se construyó un árbol dentríptico tomando el criterio de la técnica de anchura de silueta, que refleja los cortes del arbol en varios grupos usando la posición que ocupa el promedio

más alto. Para obtener resultados más fiables se usó el reemuestreo de bootstrap multiescalar que permite resolver problemas relacionados con la estimación de intervalos de confianza o la prueba de significación estadística (Ledesma, 2008).

Por medio de la prueba *t* de *Student*, y la prueba no paramétrica de la suma de rangos de Wilcoxon (medianas), se evaluaron la homogeneidad de medias y medianas para dos grupos usando como variable de agrupamiento los grupos establecidos en el agrupamiento Ward. Estas sirvieron para hacer una correlación con los resultados de abundancia la global y la de la especie.

Mediante el valor indicador (indVal) se detectaron las especies consideradas como indicadores y un análisis de especies con preferencia por hábitat por medio del coeficiente de correlación biserial puntual.

Utilizando la técnica del análisis de componentes principales (PCA) y técnica de ordenación restringida de análisis de abundancia o RDA (siglas de *Redundancy Analysis*) y canónica por la prueba de *Chi* cuadrado. Se comprobó la colinealidad entre las variables del suelo y tipos de especies y reconocer la especie más contribuidora, esto se realizó por el criterio de valores señalados con VIF.

Para la tercera etapa se midió la de diversidad (alpha y beta) donde se determinaron dos componentes principales; la riqueza y equidad. Por medio de la entropía de Shannon y la antrípia de Simpson se midió el índice de equidad de Pielou para la diversidad alpha. En el análisis beta se buscó la equidad usando la aproximación de Whittaker, asociada a los números de Hill y la ratio. Se identificaron las especie y sitios que contribuyen a la diversidad beta. También, se utilizó el método de la rarefacción para poder estimar combinaciones se utilizaron las métricas de la entropía de Renyi generaliza para obtienen los números de diversidad de Hill.

En la última etapa, basada fundamentalmente en ecología espacial se llevaron a cabo análisis de autocorrelación espacial mediante correlograma de puntos y correlación. Así las técnicas aplicadas fueron la de mantel para determinar la correlación entre dos matrices de distancia y determinar autocorrelación mediante la prueba de permutación para *I* de *morán*, utilizando los denominados Lisa.

### 3 Resultados

La familia de plantas *Malvaceae* cuenta con una cantidad de 3,792 individuos dentro de la parcela (Ver Tabla 1). En promedio la cantidad de especies por hectárea ronda en torno a unas 8 individuos.

A simple vista en la figura presentada más abajo la especie más abundante del conjunto es *Quararibea asterolepis* guardando una similitud con el resto de las especies (Ver figura 3).

Especies de planta	Cantidad
Quararibea asterolepis	2171
Herrania purpurea	542
Apeiba membranacea	308
Luehea seemannii	215
Hampea appendiculata	191
Guazuma ulmifolia	74
Ceiba pentandra	62
Sterculia apetala	53
Apeiba tibourbou	50
Pseudobombax septenatum	42

Especies de planta	Cantidad
Cavanillesia platanifolia	36
Pachira sessilis	18
Theobroma cacao	16
Ochroma pyramidale	11
Trichospermum galeottii	2
Pachira quinata	1

Tabla 1. Abundancia por especies

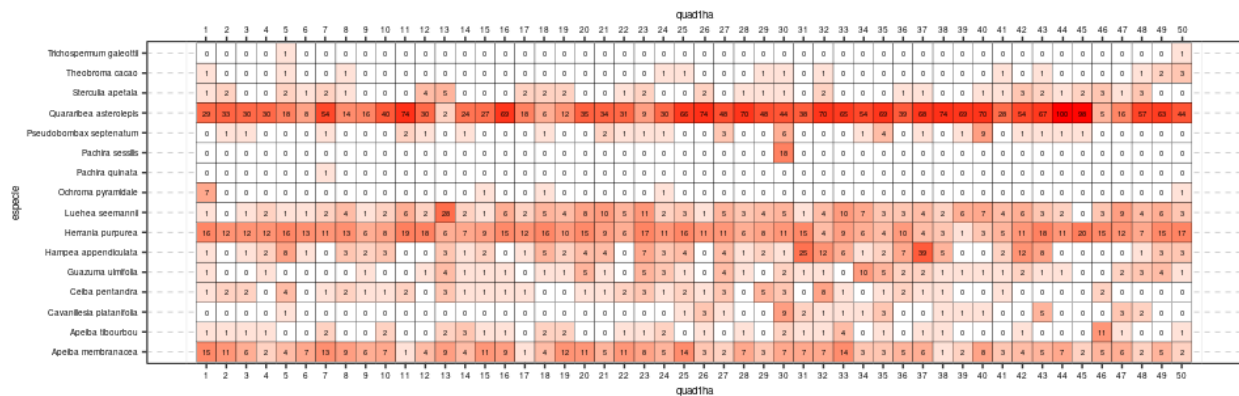


Figure 3: Gráfico de mosaico de abundancia por especie

Al evaluar la primera etapa del estudio basado en el análisis ambiental se refleja que en la parte oriental de la parcela existe una mayor abundancia de la familia *Malvaceae* y una distribución de riquezas máximas concentrada en el borde superior central (Ver figura 4 y 5).

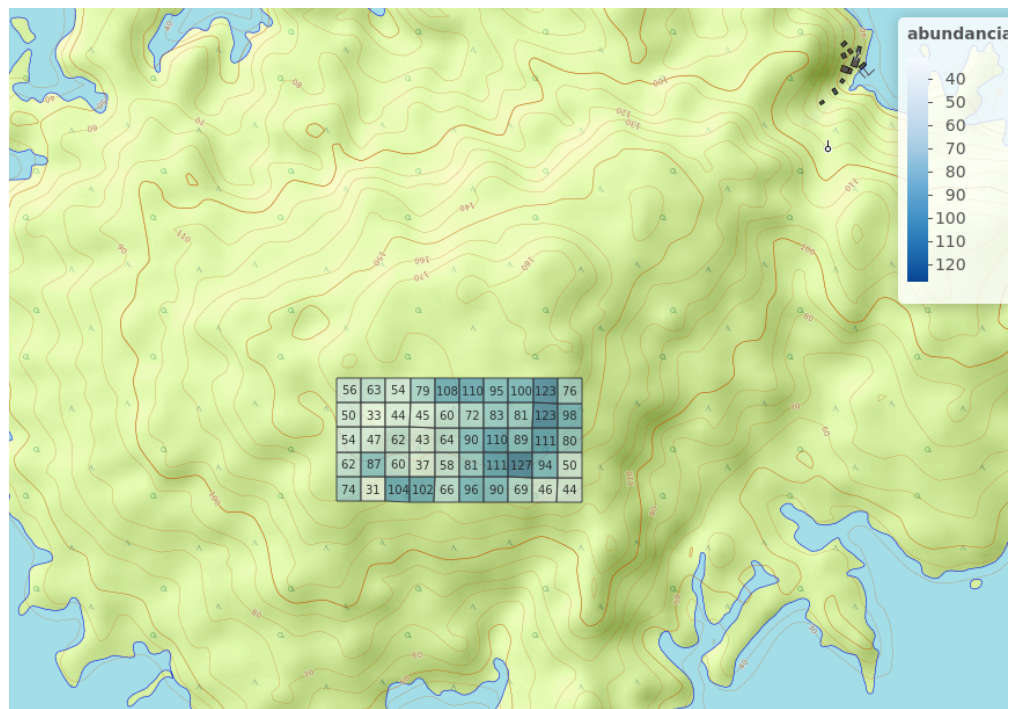


Figure 4: Distribución de la abundancia por cuadros de la familia Malvaceae por Ha

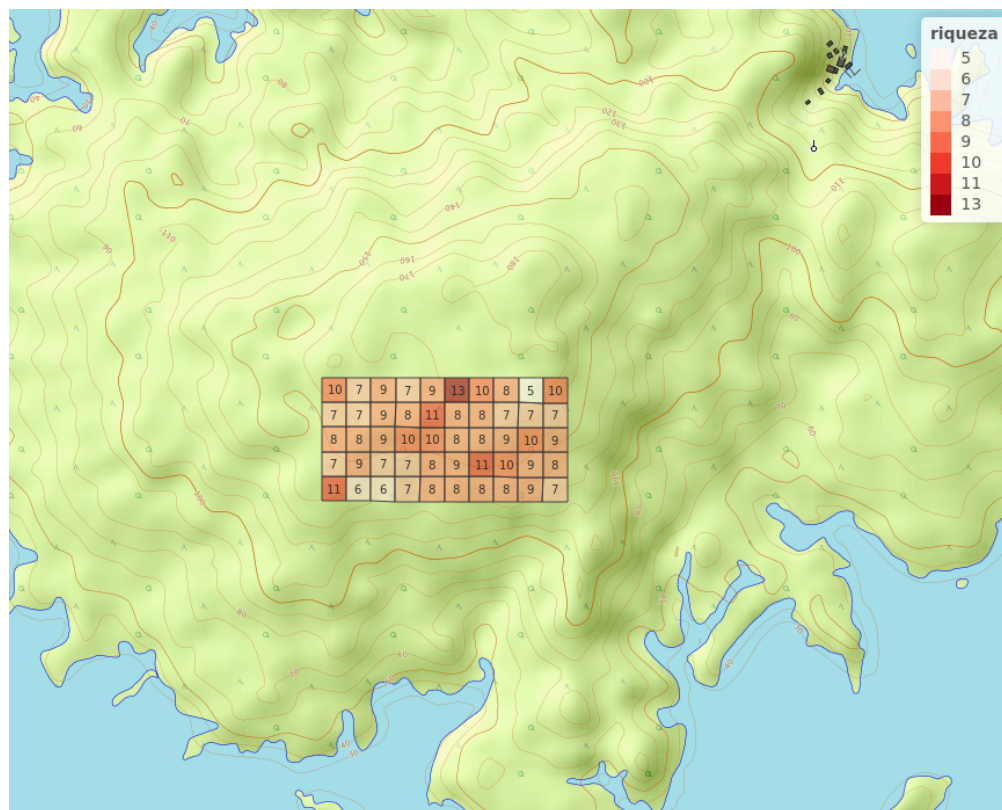


Figure 5: Distribución de las riquezas por cuadros de 1 Ha



### 3.1 Análisis de asociación

En el mapa de calor ordenado (a la derecha), presenta un clúster gigante en el centro que indica un patrón ordenado de dependencia entre las especies relacionadas. En la diagonal desde *Pseudobombax septenatum* hasta *Apeiba tibourbou* (cuadros de color rosa centrales). También se observan las especies que no parecen asociarse con otras, situadas en los extremos de la diagonal, y relacionadas con otras por medio de valores pequeños de distancia (cuadros azules), como *Theobroma cacao* y *Pachira sessilis*. El color rosado indica distancia corta y mientras más cortas los cluster se parecen entre sí (Ver figura 6).

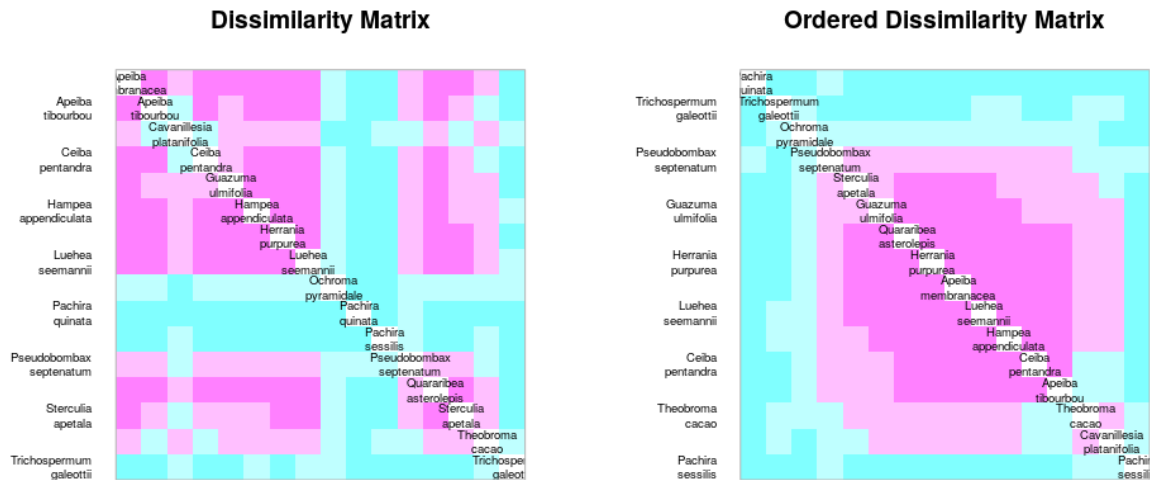


Figure 6: Matriz de asociación entre especies

### 3.2 Análisis de agrupamiento

Los dendogramas fueron generados por los métodos de enlaces simple, completos, UGMA y Ward (Ver figura 7). Estos mostraron cortes desiguales en los método UPGMA, completo y simple produciendo agrupamientos integrados por un gran número de sitios o por un sitio único. Se detectaron elementos que no forman grupos, es decir, sitios que aparecen aislados del resto como fue el caso del método por enlace simple. No obstante, con el método Ward de anchura de silueta se generó un dendograma más legible. Dicho método sugirió dividir el dendograma en 2 grupos; un grupo pequeño integrado por 8 sitios y otro grupo compuesto por los 42 sitios restantes (Ver figura 8).

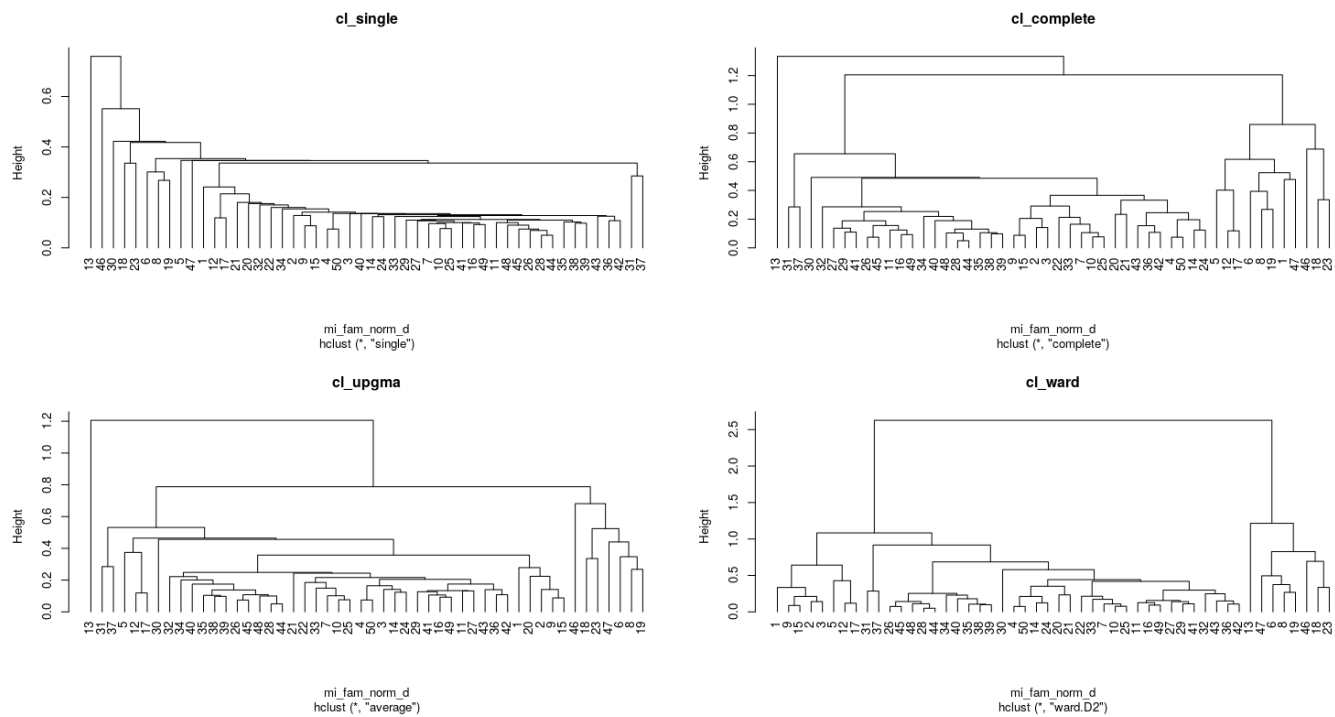


Figure 7: Dendrogramas por los cuatro métodos

## Cluster Dendrogram

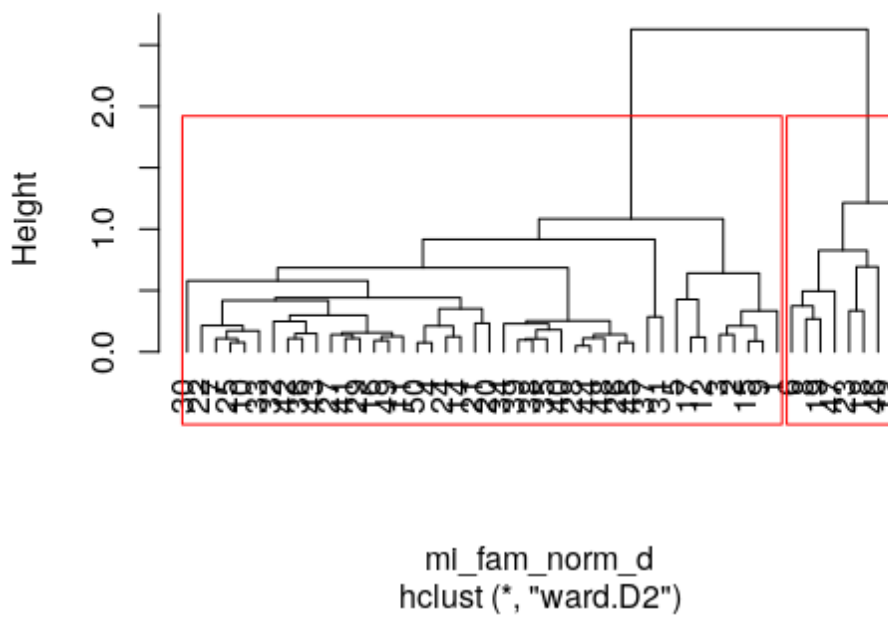


Figure 8: Dendrograma por el método Ward



Para la homogeneidad de promedios se evaluaron mediante las pruebas *T* de *Student* y la suma de rango de Wilcoxon por el métodos de Ward divididas en dos grupos. Estos paneles muestran el promedio entre las variables geomorfológicas y elementos del suelo. El nitrógeno (N), pH, cobre (Cu), y magnesio (Mg) y zinc (Zn) resultaron ser significativamente diferentes en media y mediana; en el caso del relieve el la pendiente media resultó tener el promedio más diferente, en cambio los elementos con mayor homegeneidad de medianas fueron el boro (B), manganeso (Mn), y potasio (K) y en las áreas de relieve estuvieron la elevación media, espolón, e interfluvio (ver figura 9).

Por medio del método del “valor indicador” (Indval), se encontraron en total 3 especies que pueden ser consideradas como indicadoras con preferencia de habitats. Especificamente estas especies fueron *Quararibea asterolepis* como especie asociada al grupo 1, *Luehea seemannii* y *Sterculia apetala* pertenecientes al grupo 2, lo significa que estas especies son extremadamente importantes en la prueba de permutación.

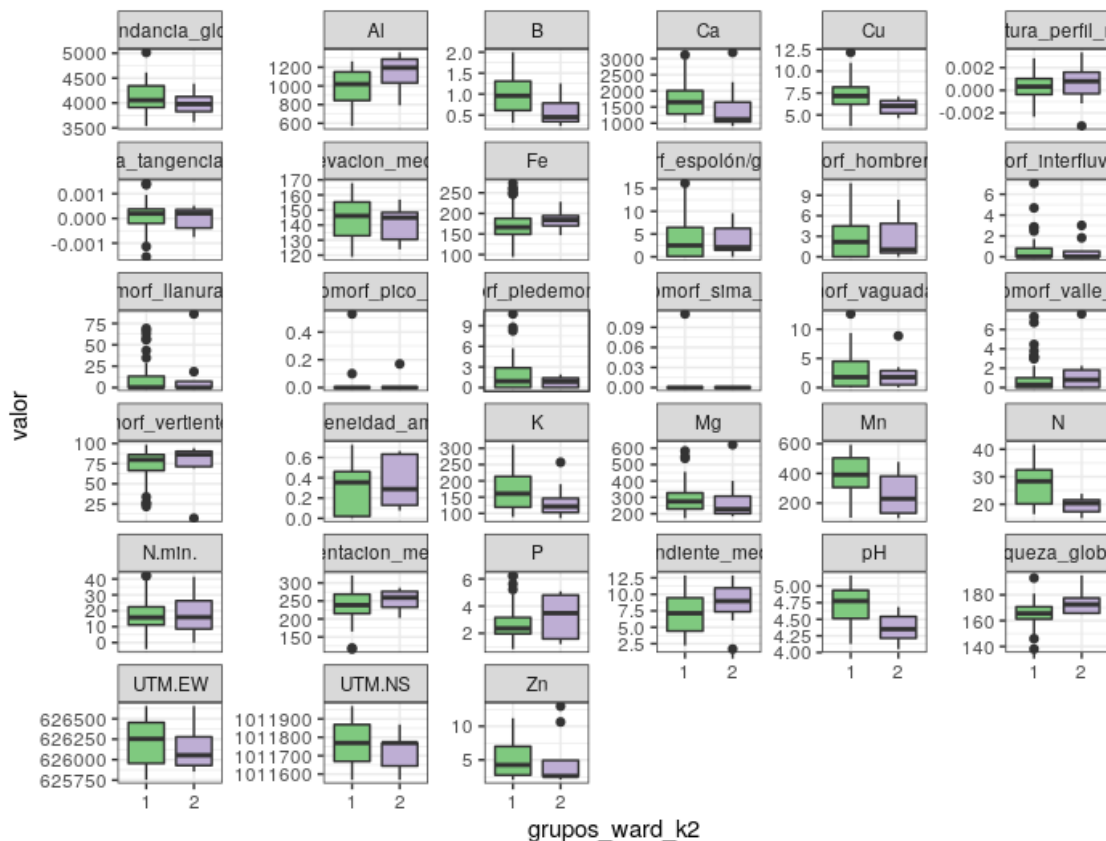


Figure 9: Pruebas de igualdad de promedios entre las variables

### 3.3 Técnicas de ordenación

En el análisis de correspondencia por el método PCA, al ajustarlo a la matriz de comunidad y usando la distancia *Chi* cuadrado, se encuentra que muchos de los componentes del suelo se encuentran asociados en las variables de comunidad, por lo que, presentan algún grado de asociación entre las especies. (Ver figura 10).

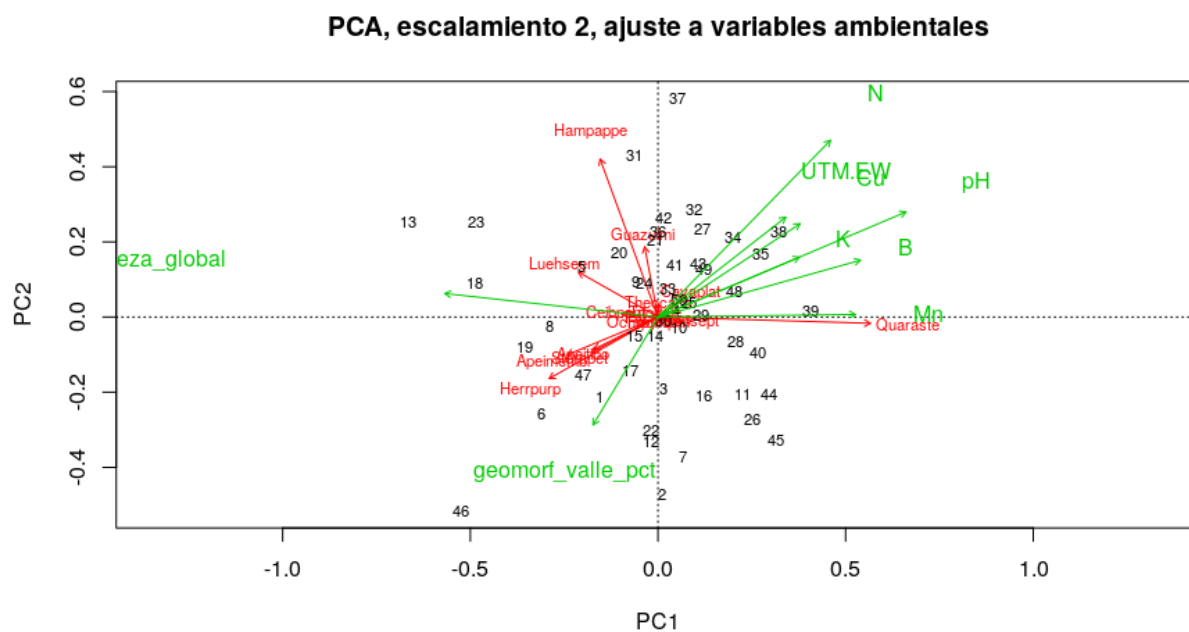


Figure 10: Biplot de variables ambientales por el método PCA

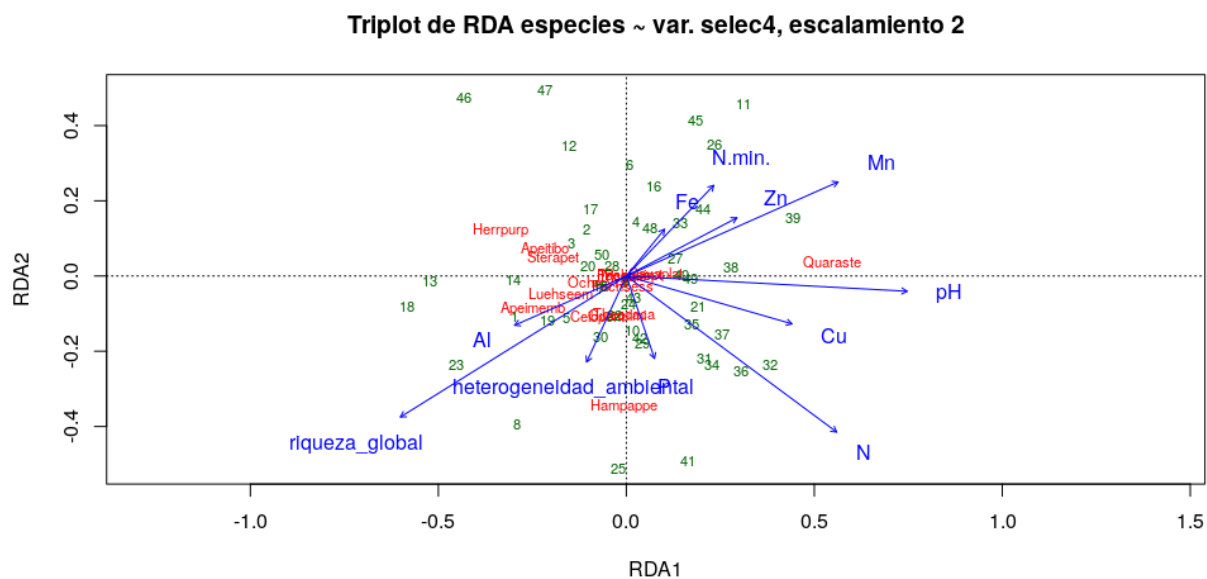


Figure 11: Triplot de especies y variables por el método RDA

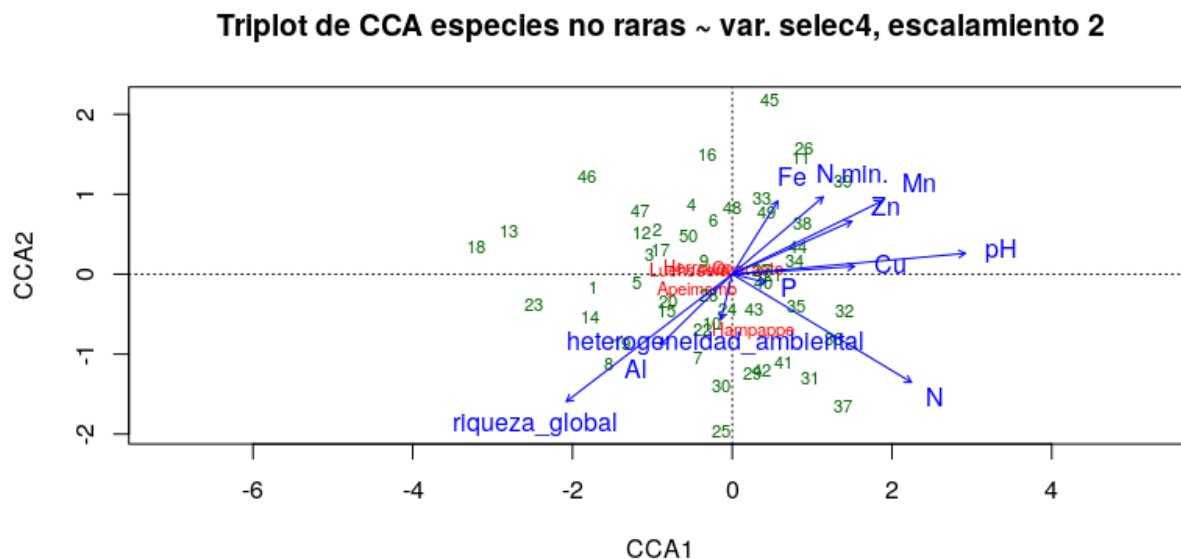


Figure 12: Triplot por el método CCA de especies no raras

En el análisis de redundancia (RDA) anterior muestra la coalelidad que existe entre las diferentes variables de especies y elementos del suelo. Las especies *Hampea appendiculata*, *Ceiba pentandra*, tienen una gran contribución en diferentes sitios, que se relacionan con el Fósforo (P), un poco con el nitrógeno (N), en el caso de *Apeimemb* tiene mucha asociación con el Aluminio (Al). *Quaraste* es la que más contribuye al conjunto de sitios con más elementos. Por otro lado, Fueron excluidas algunas variables por tener el valor *VIF* por encima de 10 como es el caso del magnesio (Mg), calcio(Ca) y las coordenadas UTM. A pesar de, habían variables con un alto valor *VIF* por lo que se optó en conservarlas por razones biogeoquímicas y de asociación (Ver gráfico 11).

De esta misma manera, en el análisis de correspondencia canónica (CCA) fueron excluidas las especies con menos de 100 individuos (especies raras) de la matriz de comunidad, se conservaron 5 en total, y se excluyeron 11 (Ver gráfico 12). En ambos gráficos las variables aparecen en el mismo lugar salvo algunas especies que desaparecieron en el “triplot de CCA” por tener menos de 100 individuos.

### 3.4 Análisis de diversidad alpha y beta

Se usaron diferentes métricas para medir los principales componentes de la diversidad: abundancia y equidad. Se hicieron mediante la equidad de Shanon (equidad) y entropía de Simpson (abundancia), con los ratios de Hill, junto con los números de entropía de Renyi y la equidad de Pielu. En el panel de correlación las especies tuvieron una fuerte asociación en presencia Magnesio (Mg), calcio (K), zinc (Zn), y una altísima correlación con el pH sobre todo en la zona de hombrera, perfil de curvatura media con mayor equidad hacia el Este. Los rojos representan alta correlación significativa los azules baja (ver figura 13 y 14).

En el modelo beta las especies *Hampea appendiculata* y *Quararibea asterolepis* fueron las más contribuyentes a la diversidad con 0.18 y 0.14 % (ver tabla 2), mientras que los sitios con mayor

aportación son el 13 con 0.12 y 46 con 0.86 % de especies.

En el análisis de rarefacción establece los sitio de mayor y menor diversidad. En los modelos de abundancia de alpha, el sitio con mayor riqueza es el 30 con 13 especies y el de menor riqueza es el 45 con 5 especies, la abundancia máxima y mínima fueron en los sitios 6 con 127 y 37 con 31, la abundancia en el sitio más pobre fue 123 en el sitio 45, y la abundancia en el sitio más rico fue 110 en el sitio 30.



Figure 13: Representación de la distribución de especies por unidad geomorfológica

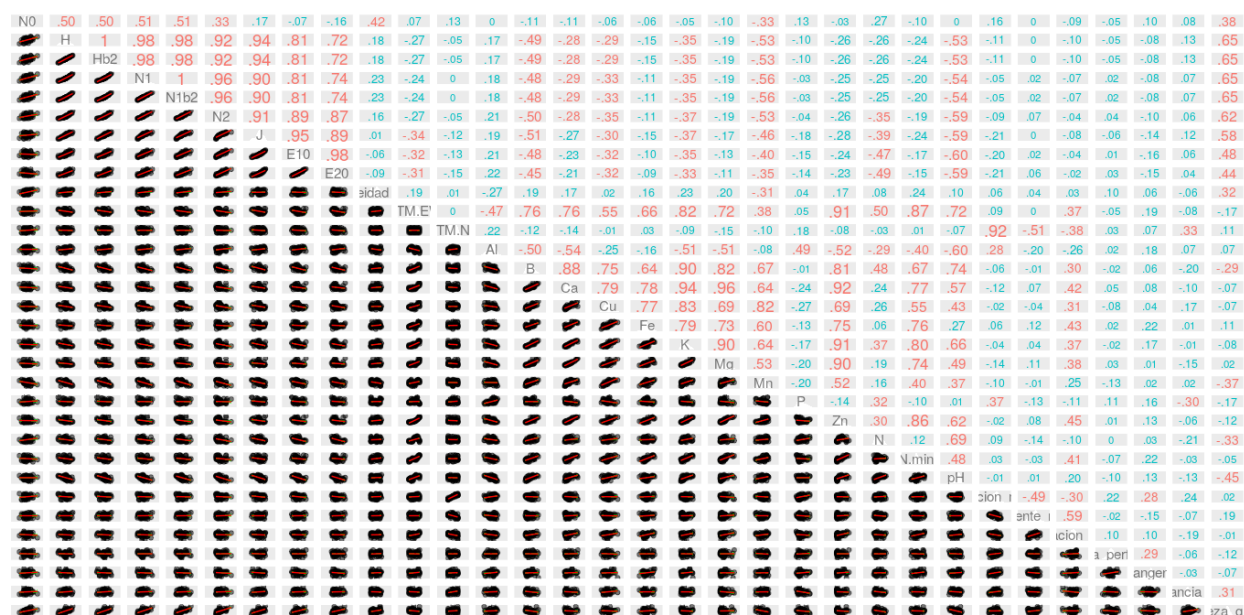


Figure 14: Representación de especies por variables ambientales

[Contribución de las especies en la diversidad Beta ]

Especies de planta	valor
Apeiba membranacea	0.08
Apeiba tibourbou	0.07
Hampea appendiculata	0.14
Herrania purpurea	0.11
Luehea seemannii	0.09
Quararibea asterolepis	0.18

### 3.5 Ecología espacial

Mediante la prueba de permutación para del *I de Moran* se probó la autocorrelación entre cada especie de plantas. De acuerdo con el resumen estadístico las especies con mayor autocorrelación fueron *Apeiba membranacea*, *Herrania purpurea* y *Quararibea asterolepis* (ver figura 15). Asimismo, los elementos químicos del suelos más representativos en la autocorrelación fueron el zinc (Zn) con 0.85%, el potasio (K) 0.74%, el calcio (Ca) 0.69 y el pH con 0.72%. En cuanto a las zonas geomorfológicas se encuentra una alta correlación en la llanura, en el espolón, en la vertiente y en la vaguada (ver figura 16).

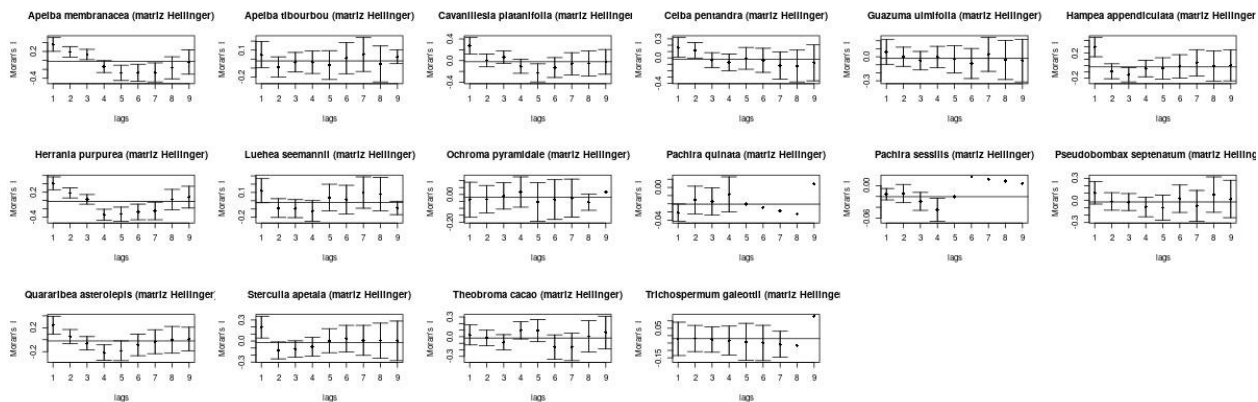


Figure 15: Correlograma de matriz de comunidad transformada por especies

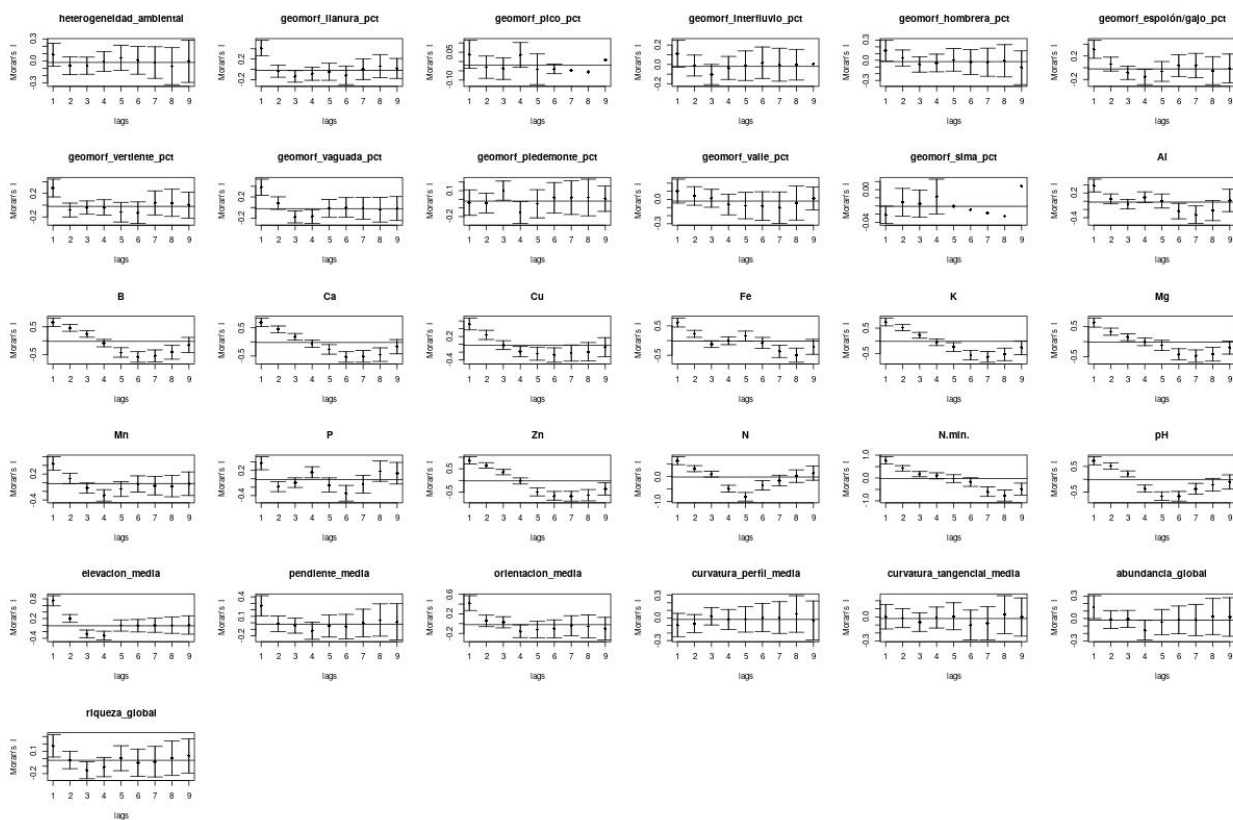


Figure 16: Correlación entre variables geomorfológicas y elementos del suelo

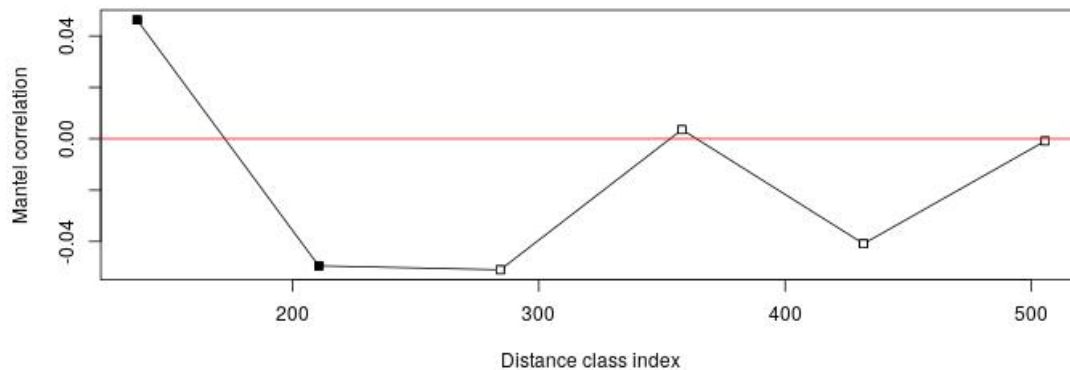


Figure 17: Correlograma Mantel de las especies de la familia

El correlograma presenta que para el nivel de significancia 0.04 en los 200 metros existe autocorrelación espacial en la matriz de comunidad. Esto significa que la posición 1 y 2 son las que se encuentran más autocorrelacionadas mientras que los otros órdenes (parte de los residuos) no existe una relación entre especies (Ver figura 17).

Los clúster LISA mostraron que el pH es el elemento químico con más autocorrelación espacial, específicamente en la zona de elevación media de BCI localizada al norte (Ver figura 18). En ese mismo orden, indicó que *Quararibea asterolepis* tiene valores de grandes de abundancias y autocorrelación seguido de *Sterculia apetala*, *Theobroma cacao* y *Pseudobombax septenatum* que presentan un patrón similar (Ver figura 19). Los cuadros rojos representan la autocorrelación espacial con valores altos y el azul con valores bajos.

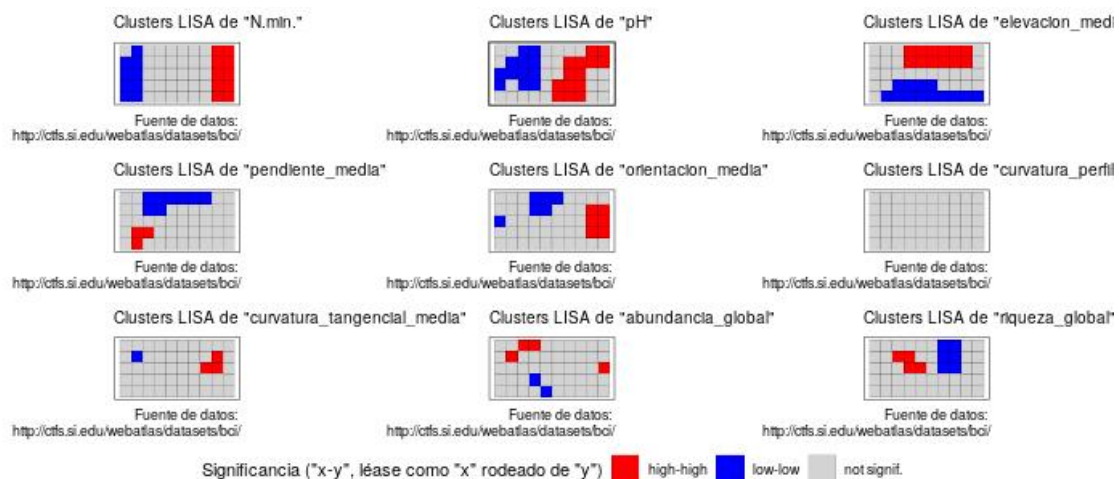


Figure 18: Clúster LISA aplicado a variables ambientales



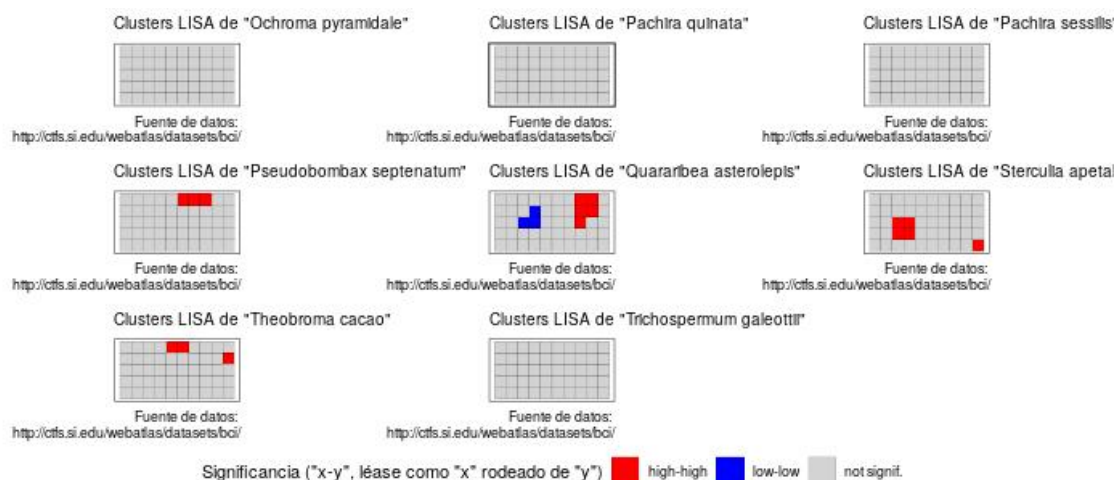


Figure 19: Clúster LISA aplicado especies transformadas

#### 4 Discusión

Los resultados mostraron que la familia de plantas *Malvaceae* posee una abundancia de 3,792 especies agrupadas mayormente en la parte oriental de manera dispersa en la parcela, con riquezas máximas situadas en el borde superior central. Esta distribución de especies en grupos aleatorios depende de diferentes factores abióticos (no vivo) y bióticos (vivos).

Se determinó que en la mayoría de las especies hay una gran asociación entre ellas, y desde el punto de vista de la distancia de Jaccard están muy próximas. Adicionalmente, el agrupamiento de especies en forma de dendrograma se generó mediante el método Ward por ser el más fácil de interpretar.

Al analizar por separado los gráficos de escalonamiento de ordenación RDA y CCA, los resultados indicaron que solo algunos compuestos del suelo están asociados con especies de la familia, como son el magnesio (Mg), calcio (Ca) siendo el pH el más representativo del área estudiada.

Se destaca la especie *Quararibea asterolepis* por encima de las otras de la familia tanto en la diversidad, distribución como en contribución de los sitios. De igual manera fue la especie con mayor autocorrelación espacial, seguida de *Apeiba membranacea* y *Herrania purpurea* las cuales también contribuyeron a la diversidad de sitios. Las zonas donde se produjo esta autocorrelación fueron en la llanura, espolón y vertiente en presencia de pH, zinc (Zn), boro (B) y calcio (Ca). En cambio, las zonas de hombrera y elevación media presentaron mayor diversidad en presencia de magnesio (Mg), calcio (Ca), zinc (Zn) que son conocidas por ser zonas húmedas.

La topografía puede verse modificada por los compuestos del suelo en la vegetación. La cantidad de agua disponible en el suelo, el pH, la cantidad de nutrientes y la textura del suelo son factores que modifican la pendiente (relieve) y por lo tanto, influyen en la distribución de las plantas en los bosques. (Clark, 2002) En BCI en áreas de tierra firme, el nivel freático puede subir ocasionalmente, y causar periodos excepcionalmente adversos y hasta letales para la vegetación (Clark, 2002). Por tanto, las condiciones climáticas, factores edáficos y la topografía del terreno influyen directamente en la distribución y modificación de las plantas en BCI. En el caso de los compuestos donde hay escasa correlación puede deberse a un exceso de precipitación o largos periodos de sequía.

Asimismo, la alta presencia de pH en el suelo genera una acidificación que provoca una dis-

minución en la disponibilidad de ciertos elementos nutritivos como son el fósforo, magnesio y calcio en aquellos suelos donde suelen ser absorbidos por las plantas, por lo tanto una letalidad más rápida en las plantas.(G.J., 2019).

A partir de los datos estudiados y comparándolos con estudios anteriores se distingue que las plantas son propensas a desarrollarse en áreas húmedas y con disponibilidad de agua esto se comprueba por patrones discontinuo por aglomeración en presencia en las zonas de vaguada, la llanura, el espolón y la vertiente.

Según investigadores, los patrones de lluvia están cambiando, con períodos secos extremos más frecuentes en toda la región. En el pasado, estos períodos secos pudieron haber conducido a un aumento de las tasas de extinción de fauna y flora y alteración el los compuesto del suelo (Armuelles, 2021). En este aspecto, si los patrones de lluvias continúan cambiando con periodos más secos, en dos décadas la Isla Barro Colorado podría perder especies de plantas como de animales, y esto podrían ocasionar un aumento de la tasa de extinción de las clases de plantas de la familia *Malvaceae*.

## 5 Agradecimientos

Agradezco al profesor de Biogeografía José Ramón Martínez Battle del área de las ciencias geográficas de la Universidad Autónoma de Santo Domingo (UASD) por tener la iniciativa, propiciar las investigaciones y facilitar las herramientas para este estudio. De igual forma a la escuela de geografía de la UASD por ser fuente de formación de profesionales en el área de geografía.

De igual forma, al Instituto Smithsonian de Instigaciones Tropicales por facilitar los datos recogidos de años en la Isla Barro Colorado. Por último a la estudiante de Ciencias Geográficas Ana Valera por ayudarme a la redacción de esta investigación.

## 6 Información de soporte

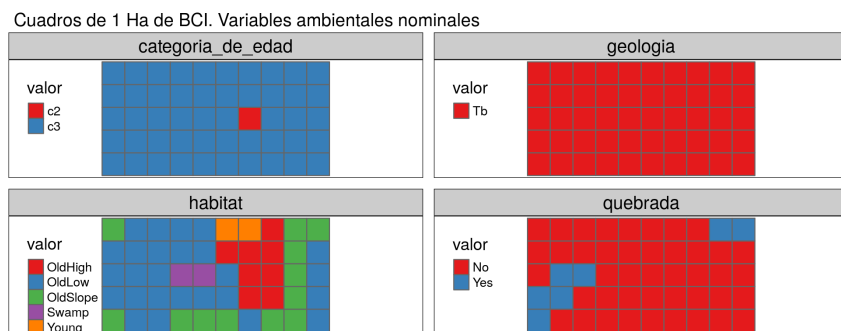


Figure 20: Representación de variables ambientales nominales

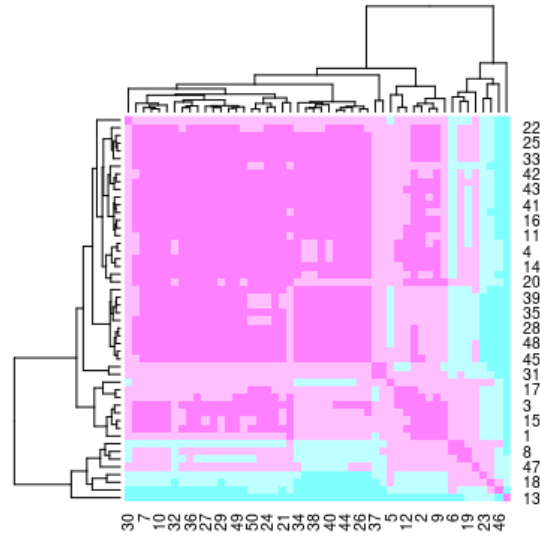


Figure 21: Mapa 4. Matriz anchura de siluetas por el método Ward

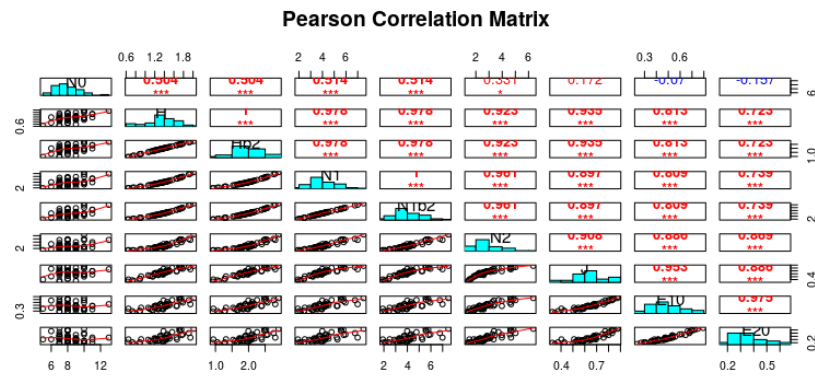


Figure 22: Índice de correlación de pearsons

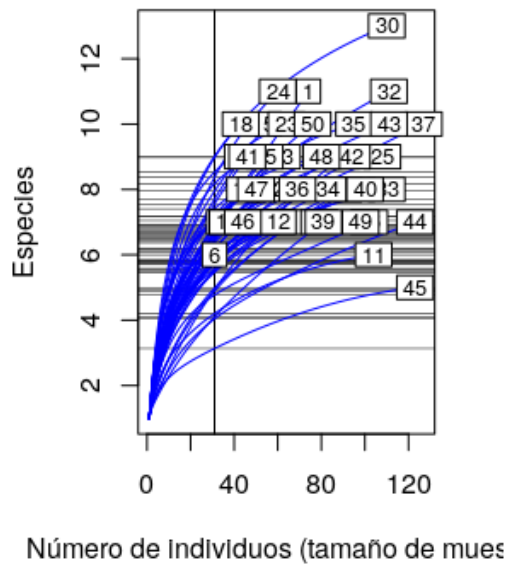


Figure 23: Gráfico de rarefacción

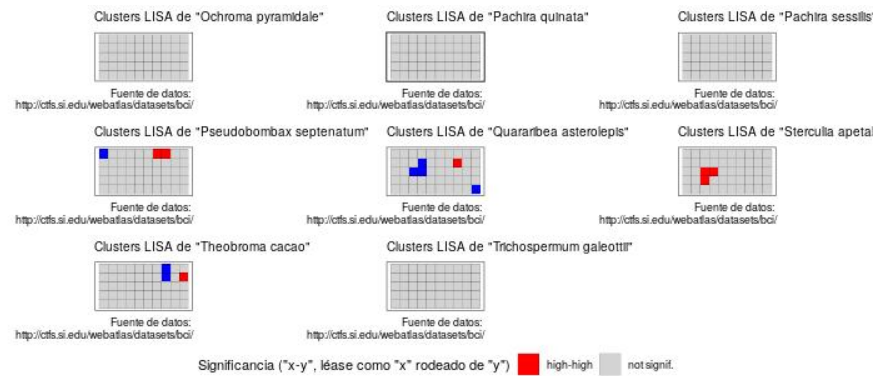


Figure 24: Cluster LISA aplicado a especies (sin tendencia)

...

## 7 Script reproducible

### 7.1 Análisis exploratorio de datos. Riqueza y abundancia

```
#' ---
#' title: "Análisis exploratorio de datos. Riqueza y abundancia"
#' author: "JR"
#' date: "13 de octubre, 2020"
#' output: github_document
```

```

#' ---

#' ### Área de cargar paquetes
library(vegan)
library(tidyverse)
library(sf)
source('biodata/funciones.R')

#' ### Área de cargar datos
#' Censo (el objeto se carga con prefijo "censo") y matriz de comunidad (prefijo "mc")
load('biodata/Malvaceae.Rdata')
load('biodata/matriz_ambiental.Rdata') #Matriz ambiental, se carga como "bci_env_grid"

#' ### Imprimir datos en pantalla (impresiones parciales con head)
head(censo_malvc)
head(mc_malvc)
bci_env_grid # No necesita imprimirse parcialmente

#' ### También podemos usar
#' Requiere que se haya cargado ya la colección tidyverse
censo_malvc %>% tibble
mc_malvc %>% tibble

#' ### Lista de especies
sort(colnames(mc_malvc))

#' ### Número de sitios, tanto en matriz de comunidad como en ambiental
#' Verifica que coinciden
nrow(mc_malvc) #En la matriz de comunidad
nrow(bci_env_grid) #En la matriz ambiental

#' ### Riqueza numérica de especies (usando matriz de comunidad) por quadrat
#' Nota: cargar paquete vegan arriba, en el área de paquetes
specnumber(mc_malvc)
sort(specnumber(mc_malvc)) # Ordenados ascendentemente
summary(specnumber(mc_malvc)) # Resumen estadístico

#' ### Abundancia de especies por quadrat
sort(rowSums(mc_malvc))
summary(rowSums(mc_malvc)) # Resumen estadístico

#' ### Abundancia por especie
sort(colSums(mc_malvc))
summary(colSums(mc_malvc)) # Resumen estadístico

#' ### Riqueza numérica de toda la "comunidad"
specnumber(colSums(mc_malvc))

```

```

#' ### Abundancia de toda la comunidad
sum(colSums(mc_malvc))

#' ### Una tabla para el manuscrito, es necesario asignarle nombre
#' Para esto, usaré la colección "tidyverse"
abun_sp <- censo_malvc %>%
  group_by(Latin) %>%
  count() %>%
  arrange(desc(n))
abun_sp

#' ### Un gráfico para el manuscrito
#' Gráfico de mosaicos de la abundancia por especie por cuadros

abun_sp_<- crear_grafico_mosaico_de_mc(mc_malvc, tam_rotulo = 6)
abun_sp_

```

## 7.2 *Análisis exploratorio de datos. Mapas de riqueza y abundancia global de Malvaceae*

```

#' ---
#' title: "Análisis exploratorio de datos. Mapas de riqueza y abundancia global y de mi familia"
#' author: "JR"
#' date: "25 de octubre, 2020"
#' output: github_document
#' ---

#' ### Cargar paquetes
library(mapview)
library(tidyverse)
library(vegan)
library(sf)
library(RColorBrewer)

#' ### Cargar datos
load('biodata/matriz_ambiental.Rdata')
load('biodata/Malvaceae.Rdata')

#' ### Explorar el objeto de matriz ambiental
bci_env_grid

#' ### Generar mapa de cuadros sin simbología
mapa_cuadros <- mapView(
  bci_env_grid,
  col.regions = 'grey80',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',

```

```

legend = F, zoom = 14,
zcol = 'id') %>% addStaticLabels() %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)
mapa_cuadros
mapa_cuadros %>% mapshot(file = 'mapa_cuadros.png') #Genera archivo

#' ### Paletas
Azul <- colorRampPalette(brewer.pal(8, "Blues"))
rojo <- colorRampPalette(brewer.pal(8, "Reds"))

#' ### Mapa de cuadros, simbología por abundancia global
mapa_cuadros_abun_global <- mapView(
  bci_env_grid,
  layer.name = 'abundancia',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T, zoom = 14,
  col.regions = azul,
  zcol = 'abundancia_global') %>%
  addStaticLabels(label = bci_env_grid$abundancia_global, textsize = "7pt") %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_cuadros_abun_global
mapa_cuadros_abun_global %>% mapshot(file = 'mapa_cuadros_abun_global.png')

#' ### Mapa de cuadros, simbología por riqueza global
mapa_cuadros_riq_global <- mapView(
  bci_env_grid,
  layer.name = 'riqueza',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T, zoom = 14,
  col.regions = rojo,
  zcol = 'riqueza_global') %>%
  addStaticLabels(label = bci_env_grid$riqueza_global, textsize = "7pt") %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_cuadros_riq_global
mapa_cuadros_riq_global %>% mapshot(file = 'mapa_cuadros_riq_global.png')

#' ### Mapa de cuadros, simbología por abundancia de mi familia

```



```

mapa_cuadros_abun_mi_familia <- mapView(
  bci_env_grid %>% mutate(abun = rowSums(mc_malvc)),
  layer.name = 'abundancia',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T, zoom = 14,
  col.regions = azul,
  zcol = 'abun') %>%
  addStaticLabels(label = rowSums(mc_malvc)) %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_cuadros_abun_mi_familia
mapa_cuadros_abun_mi_familia %>% mapshot(file = 'mapa_cuadros_abun_mi_familia.png')
# La mayor abundancia de la familia Malvaceae se concentra en el cluster de la parte oriental, m

#' ### Mapa de cuadros, simbología por riqueza de mi familia
mapa_cuadros_riq_mi_familia <- mapView(
  bci_env_grid %>% mutate(riq = specnumber(mc_malvc)),
  layer.name = 'riqueza',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T, zoom = 14,
  col.regions = rojo,
  zcol = 'riq') %>%
  addStaticLabels(label = specnumber(mc_malvc)) %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_cuadros_riq_mi_familia
mapa_cuadros_riq_mi_familia %>% mapshot(file = 'mapa_cuadros_riq_mi_familia.png')
#Las riquezas máximas en la familia de plantas Malvaceae se encuentran en el borde superior cent

```

### 7.3 *Análisis exploratorio de datos. Mapas de variables ambientales*

```

#' ---
#' title: "Análisis exploratorio de datos. Mapas de variables ambientales"
#' author: "JR"
#' date: "25 de octubre, 2020"
#' output: github_document
#' ---

#' ### Cargar paquetes
library(mapview)

```

```

library(tidyverse)
library(sf)
library(RColorBrewer)

#' ### Cargar datos
load('biodata/matriz_ambiental.Rdata')

#' ### Paletas
azul <- colorRampPalette(brewer.pal(8, "Blues"))
rojo <- colorRampPalette(brewer.pal(8, "Reds"))
rojo_inv <- colorRampPalette(rev(brewer.pal(8, "Reds")))

#' ### Mapa de cuadros, simbología por pendiente
mapa_cuadros_pendiente <- mapView(
  bci_env_grid,
  layer.name = 'pendiente',
  alpha.regions = 0.4,
  map.types = 'OpenTopoMap',
  legend = T, zoom = 14,
  col.regions = rojo,
  zcol = 'pendiente_media') %>%
  addStaticLabels(label = round(bci_env_grid$pendiente_media, 1)) %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_cuadros_pendiente
mapa_cuadros_pendiente %>% mapshot(file = 'mapa_cuadros_pendiente.png') #Genera archivo

#' ### Mapa de cuadros, simbología por Nitrógeno
mapa_cuadros_nit <- mapView(
  bci_env_grid,
  layer.name = 'N (mg/kg)',
  alpha.regions = 0.4,
  map.types = 'OpenTopoMap',
  legend = T, zoom = 14,
  col.regions = rojo,
  zcol = 'N') %>%
  addStaticLabels(label = round(bci_env_grid$N, 1)) %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_cuadros_nit
mapa_cuadros_nit %>% mapshot(file = 'mapa_cuadros_nit.png')

#' ### Mapa de cuadros, simbología por pH
mapa_cuadros_ph <- mapView(

```

```

bci_env_grid,
layer.name = 'pH',
alpha.regions = 0.4,
map.types = 'OpenTopoMap',
legend = T, zoom = 14,
col.regions = rojo_inv,
zcol = 'pH') %>%
addStaticLabels(label = round(bci_env_grid$pH, 1)) %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)
mapa_cuadros_ph
mapa_cuadros_ph %>% mapshot(file = 'mapa_cuadros_ph.png')
#'La concentración de PH se encuentra en la parte oriental los porcentaje mas rojos son mas acid

```

#### 7.4 *Análisis exploratorio de datos. Correlaciones entre variables ambientales*

```

#' ---
#' title: "Análisis exploratorio de datos. Correlaciones entre variables ambientales"
#' author: "JR"
#' date: "25 de octubre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ### Cargar paquetes
library(tidyverse)
library(sf)
library(ez)
library(psych)
library(vegan)

#' ### Cargar datos
load('biodata/matriz_ambiental.Rdata')
load('biodata/Malvaceae.Rdata')

#' ### Una correlación simple
cor(bci_env_grid$pendiente_media, bci_env_grid$geomorf_vertiente_pct)
plot(bci_env_grid$pendiente_media, bci_env_grid$geomorf_vertiente_pct)
cor.test(bci_env_grid$pendiente_media, bci_env_grid$geomorf_vertiente_pct)

#' ### Generar objeto de columnas numéricas
#' El objeto que generaré, denominado 'env_num', no tendrá las columnas 'id' y las de coordenada
env_num <- bci_env_grid %>%
  dplyr::select_if(is.numeric) %>%

```

```

dplyr::select(-id, -matches('^U.*')) %>%
st_drop_geometry %>%
mutate(
  riqueza_mifam = specnumber(mc_malvc),
  abundancia_mifam = rowSums(mc_malvc)) %>%
rename_all(gsub, pattern = '_pct$', replacement = '') %>%
rename_all(gsub, pattern = '_| ', replacement = '\\n')
env_num %>% tibble

#' ### Panel de correlaciones con herramientas del paquete 'graphics' y 'psych'
cor(env_num)
ncol(env_num)
pairs(env_num[,sample(1:33, 15)]) # paquete graphics
env_num[,sample(1:33, 15)] %>% pairs.panels #paquete psych

#' ### Panel de correlaciones con 'ez'
#'
#' ##### Todas las variables (se empasta). Comentado, sólo mostrado para fines didácticos
# p_cor_todos <- env_num %>%
#   ezCor(r_size_lims = c(4,8), label_size = 4)
# p_cor_todos

#' ##### Sólo suelo (elementos y pH), abundancia/riqueza
p_cor_suelo_ar <- env_num %>%
  dplyr::select(matches('^([A-T,Z]|abundancia|riqueza|^pH$'), ignore.case = F)) %>%
  ezCor(r_size_lims = c(4,8), label_size = 3)
p_cor_suelo_ar

#' ##### Sólo heterogeneidad, geomorfología, abundancia/riqueza
p_cor_geomorf_ar <- env_num %>%
  dplyr::select(-matches('^([A-T,Z]|pH$'), ignore.case = F)) %>%
  ezCor(r_size_lims = c(4,8), label_size = 3)
p_cor_geomorf_ar

#' ##### Matriz de comunidad
p_cor_mc <- mc_malvc %>%
  rename_all(gsub, pattern = '_| ', replacement = '\\n') %>%
  ezCor(r_size_lims = c(4,8), label_size = 3)
p_cor_mc

```

## 7.5 *Análisis exploratorio de datos. Mapas de variables ambientales por lotes*

```

#' ---
#' title: "Análisis exploratorio de datos. Mapas de variables ambientales por lotes"
#' author: "JR"
#' date: "3 de diciembre, 2020"
#' output: github_document

```

```

#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo
#'
#' ### Cargar paquetes
#'
library(tmap)
library(sf)
library(tidyverse)
library(RColorBrewer)
#'
#' ### Cargar datos
#'
load('biodata/matriz_ambiental.Rdata')
#'
#' ## Convertir a KML
#'
st_write(
  bci_env_grid %>% rename(Name = id),
  driver = 'KML',
  dsn = 'matriz_ambiental.kml')
st_write(
  bci_env_grid %>% rename(Name = id) %>% st_centroid(),
  driver = 'KML',
  dsn = 'matriz_ambiental_puntos.kml')
#'
#' Uní los dos archivos anteriores en un único KML nombrado como 'mapa_cuadros_1ha_para_google_e
#'
#' ## Generar mapas por lotes
#'
#' ### Variables ambientales numéricas con 'ggplot2'
#'
mapas_var_amb_num_gg <- bci_env_grid %>%
  select_if(is.numeric) %>%
  gather(variable, valor, -geometry) %>%
  group_by(variable) %>%
  mutate(
    valor = scales::rescale(valor, to = c(0, 1)),
    id = rep(1:50)) %>%
  ggplot +
  aes(geometry = geometry, fill = valor) +
  theme(axis.text = element_blank()) +
  geom_sf(lwd = 0.1, color = 'grey50', alpha = 0.8) + coord_sf() +
  scale_fill_gradientn(colours = brewer.pal(11, 'BrBG')) +
  geom_sf_text(aes(label = id, color = between(valor, 0.3, 0.7)), size = 1.75) +
  scale_color_manual(guide = FALSE, values = c("white", "black")) +

```

```

    facet_wrap(~ variable, ncol = 6) +
    ggtitle('Cuadros de 1 Ha de BCI. Variables ambientales numéricas escaladas de 0 a 1')
mapas_var_amb_num_gg
#'
#' PNG
#'
png(
  filename = 'mapas_variables_ambientales_numericas.png',
  width = 1700, height = 1080, res = 150)
mapas_var_amb_num_gg
dev.off()
#'
#' ### Variables ambientales numéricas con 'tmap'
#'
mapas_var_amb_num_tmap <- bci_env_grid %>%
  select_if(is.numeric) %>%
  gather(variable, valor, -geometry) %>%
  group_by(variable) %>%
  mutate(
    valor = scales::rescale(valor, to = c(0, 1)),
    id = rep(1:50)) %>%
  tm_shape() +
  tm_polygons(col = 'valor',
    palette = brewer.pal(11, 'BrBG'),
    style = 'cont',
    legend.is.portrait = FALSE) +
  tm_facets(by = 'variable', ncol = 6, nrow = 6) +
  tm_layout(main.title="Cuadros de 1 Ha de BCI. Variables ambientales numéricas escaladas de 0 a 1",
    main.title.size = 0.7,
    legend.outside.position="bottom",
    legend.outside=TRUE,
    legend.width = 0.2,
    legend.text.size = 0.5,
    legend.stack="horizontal",
    outer.margins=0)
mapas_var_amb_num_tmap
#'
#' PNG
#'
png(
  filename = 'mapas_variables_ambientales_numericas_tmap.png',
  width = 1800, height = 1400, res = 350, pointsize = 12)
mapas_var_amb_num_tmap
dev.off()
#'
#' ### Variables ambientales nominales con 'tmap'
#'
mapas_var_amb_nom_tmap <- bci_env_grid %>%

```

```

select_if(negate(is.numeric)) %>%
gather(variable, valor, -geometry) %>%
tm_shape() +
tm_polygons(col = 'valor',
             palette = brewer.pal(8, 'Set1'),
             legend.show = T) +
tm_facets(by = 'variable', ncol = 2, free.scales = T, free.coords = T) +
tm_layout(main.title="Cuadros de 1 Ha de BCI. Variables ambientales nominales",
           main.title.size = 0.7,
           asp = 3.5,
           legend.text.size = 0.7)
mapas_var_amb_nom_tmap
#'
#' PNG
#'
png(
  filename = 'mapas_variables_ambientales_nominales_tmap.png',
  width = 2000, height = 1200, res = 350, pointsize = 12)
mapas_var_amb_nom_tmap
dev.off()

```

## 7.6 Medición de asociación. Introducción a los modos de análisis Q y R. Modo Q aplicado a la paradoja de Orlóci

```

#' ---
#' title: "Medición de asociación. Introducción a los modos de análisis Q y R. Modo Q aplicado a
#' author: "JR"
#' date: "3 de noviembre, 2020"
#' output: github_document
#' ---
#'
knitr::opts_chunk$set(fig.width=8, fig.height=5)
#'
#' ## Preámbulo
#'
#' ### Cargar paquetes
library(vegan)
library(adespatial)
library(tidyverse)
library(gridExtra)
source('biodata/funciones.R')
#'
#' ## Modos Q y R
#'
#' En modo Q mides asociación entre pares de objetos, como por ejemplo, entre dos sitios de mues
#'
#' En modo R mides asociación entre pares de descriptores, como por ejemplo, entre dos variables

```



```

#'
#' ## Modo Q: matrices de disimilaridad entre objetos
#'
#' ### Modo Q para datos cuantitativos de especies (abundancia). La paradoja de Orlóci
#'
#' La paradoja de Orlóci (1978) plantea que la distancia euclídea es más pequeña entre dos sitios
#'
#' Esta paradoja se explica por la presencia de "ceros" (especies ausentes) en la matriz de comu
#'
#' Te muestro la paradoja con un ejemplo y, posteriormente, te explico cómo solucionar el proble
#'
(mc_orloci <- tibble(
  sp1 = c(1, 0, 4),
  sp2 = c(1, 0, 8),
  sitio = paste0('sit', 1:3)) %>%
  column_to_rownames('sitio'))
#'
#' ...donde ambas especies están ausentes en 'sit2', en 'sit1' presentes con poca abundancia y e
#'
#' En modo Q, calcularé la "distancia" o "disimilaridad" entre sitios según las especies que los
#'
(dist_ldc(mc_orloci, "euclidean", silent = T))
#'
#' Te muestro un gráfico de dispersión de los sitios según la abundancia de especies (los ejes r
#'
mc_orloci %>% rownames_to_column('id') %>%
  ggplot() +
  aes(x = sp1, y = sp2, label = id) +
  geom_point(size = 3) +
  geom_text(vjust="inward",hjust="inward", size = 5, color = 'grey40') +
  coord_equal() +
  theme_bw() +
  theme(text = element_text(size = 16))
#'
#' Para facilitar la lectura de las distancias, en esta explicación ordenaré las matrices de dis
#'
(d_euc <- dist_ldc(mc_orloci, "euclidean", silent = T) %>%
  organizar_matriz_distancia(func_dist = 'Euclidean'))
#'
#' Siendo los sitios 1 y 2 tan diferentes en cuanto a las especies que los componen (de hecho, n
#'
#' Existen distintas maneras de solucionar el problema planteado en la paradoja, normalmente rec
#'
#' - *Chord*:
#'
d_cho <- dist_ldc(mc_orloci, "chord", silent = T) %>%
  organizar_matriz_distancia(func_dist = 'Chord')
#'

```

```

#' - *Ji*-cuadrado:
#'
d_chi <- dist.ldc(mc_orloci, "chisquare", silent = T) %>%
  organizar_matriz_distancia(func_dist = 'chi-square distance')
#'
#' - *Hellinger* (valores primero divididos por abundancia total > sqrt)
#'
d_hel <- dist.ldc(mc_orloci, "hellinger", silent = T) %>%
  organizar_matriz_distancia(func_dist = 'Hellinger')
#'
#' - Uniendo y comparando
(d_todas <- bind_rows(d_euc, d_cho, d_chi, d_hel))
#'
#' Verás que el par 'sit1|sit3' tiene corta distancia, es decir, son muy parecidos (0.17 en *Hel
#'
#' Nota igualmente que, tanto los pares 'sit1'|'sit2' y 'sit2|sit3' están distantes (distancia 1
#'
mc_orloci
#'
#' A continuación, generaré la matriz transformada según el método *chord*. Esta matriz se calcula
#'
mc_orloci_norm <- sqrt(mc_orloci^2/rowSums(mc_orloci^2)) %>%
  replace(is.na(.), 0)
mc_orloci_norm
#'
#' La matriz de comunidad se dice que está "normalizada". Lo anterior se puede hacer más fácilmente
#'
(mc_orloci_norm <- decostand(mc_orloci, "normalize"))
#'
#' Al graficar los sitios sobre un espacio bidimensional, cada eje representando una especie, se
#'
p1 <- mc_orloci %>%
  rownames_to_column('id') %>%
  ggplot() +
  aes(x = sp1, y = sp2, label = id) +
  geom_point(size = 3) +
  geom_text(vjust="inward",hjust="inward", size = 5, color = 'grey40') +
  coord_equal() +
  theme_bw() +
  theme(text = element_text(size = 16)) +
  ggtitle('mc original')
p2 <- mc_orloci_norm %>%
  rownames_to_column('id') %>%
  ggplot() +
  aes(x = sp1, y = sp2, label = id) +
  geom_point(size = 3) +
  geom_text(vjust="inward",hjust="inward", size = 5, color = 'grey40') +
  coord_equal() +

```

```

  theme_bw() +
  theme(text = element_text(size = 16)) +
  ggtitle('mc transformada')
grid.arrange(p1, p2, nrow = 1)
#'
#' - Por último, para completar el proceso realizado por 'dist.ldc', debes calcular la distancia
#'
(d_cho_2_pasos <- dist(mc_orloci_norm, method = 'euclidean')) %>%
  organizar_matriz_distancia(func_dist = 'Chord en dos pasos')
#'
#' Compara la matriz anterior con la generada por 'dist.ldc', objeto 'd_cho':
d_cho

```

## 7.7 Edición de asociación. Modo Q aplicado a mi familia asignada

```

#' ---
#' title: "Medición de asociación. Modo Q aplicado a mi familia asignada"
#' author: "JR"
#' date: "9 de noviembre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo

#' ### Cargar paquetes
library(vegan)
library(adespatial)
library(broom)
library(tidyverse)
library(sf)
library(cluster)
library(gclus)
source('biodata/funciones.R')

#' ### Cargar datos
#'
load('biodata/matriz_ambiental.Rdata')
load('biodata/Malvaceae.Rdata')
#'
#' ## Modo Q: matrices de disimilaridad entre objetos
#'
#' ### Modo Q para datos cuantitativos de especies (abundancia). Datos de mi familia asignada
#'
#' Aplicado a mi familia asignada de BCI, en la forma de matriz de distancia euclídea, utilizando
#'

```

```

mi_fam_d_hel <- dist.ldc(mc_malvc, "hellinger", silent = T)
mi_fam_d_hel %>% tidy # Para evitar desbordar la consola
#’
#’ Para interpretar esta matriz, es necesario representarla gráficamente. En la representación e
#’
coldiss(mi_fam_d_hel, diag = T)
#’
#’ Mejorable el gráfico, quizá este es más explícito:
#’
coldissgg(mi_fam_d_hel, ordered = T, nc = 4, fsz = 0)
#’
#’ Con valores de distancia sobreimpresos (se empastan un poco)
#’
coldissgg(mi_fam_d_hel, ordered = T, nc = 4, fsz = 1.5)
#’
#’ Puedes guardar el gráfico usando el botón ‘Export’ de la pestaña ‘Plots’
#’
#’ Una forma alterna de guardar el gráfico es mediante funciones de R. La calidad de gráficos ex
#’
png(
  filename = 'matriz_disimilaridad_hellinger.png',
  width = 2400, height = 1200, pointsize = 32
)
coldiss(mi_fam_d_hel, diag = T)
dev.off()
#’
#’ MUY IMPORTANTE. La última función, ‘dev.off()’, es necesaria para cerrar el dispositivo. Si n
#’
#’ ### Modo Q para datos binarios (presencia/ausencia)
#’
#’ Habitualmente, sólo dispones de datos de presencia/ausencia. En tales casos, existe un conjun
#’
#’ Un error común consiste en referirse a los índices de Jaccard y de Sorensen "a secas", sin es
#’
#’ Si alguna vez te enfrentas a textos donde no se especifica qué tipo de métrica se usa, te sug
#’
#’ Recalco: es imprescindible declarar qué tipo de métrica estás usando. Ejemplos de redacci
#’
#’ - Correcto: "índice de disimilaridad de Jaccard", "índice de similaridad de Sorensen"
#’
#’ - Incorrecto: "índice de Jaccard", "índice de Sorensen".
#’
#’ A continuación, muestro cómo calcular la distancia de Jaccard ( $D_{J}$ ) en un ú
#’
mi_fam_jac <- vegdist(mc_malvc, method = 'jac', binary = T)
mi_fam_jac %>% tidy # Mostrando sólo las primeras 10 combinaciones en modo data.frame
#’
#’ El argumento ‘binary=T’ en ‘vegdist’ "ordena" que se realice primero ‘decostand(mc_apcyn_meli

```

```

#'
#' En esta matriz de disimilaridad, al igual que en la anterior, un valor pequeño (rosa) signifi
#'
coldiss(mi_fam_jac, diag = T)
#'
#' La distancia de Jaccard ( $D_{J}$ ) se puede expresar como "la proporción de especies
#'
#' Como la distancia de Jaccard ( $D_{J}$ ) es el complemento a 1 de la similaridad de J
#'
(1 - mi_fam_jac) %>% tidy %>% rename(similaridad=distance) #Similaridad
#'
#' Dado que este resultado muestra la similaridad, podemos leerlo como "el sitio 1 y el 2 compar
#'
#' La fórmula de la similaridad de Jaccard es  $S_{J} = a/(a+b+c)$ , donde  $a$  es el núm
#'
#' Para obtener las variables  $a$ ,  $b$  y  $c$ , usaré La función 'betadiver' del paquete 've
#'
mi_fam_abc <- betadiver(mc_malvc)
mi_fam_abc %>%
  map(tidy) %>%
  map(slice, 1) %>%
  map_df(I, .id = 'tipo') %>%
  dplyr::select(tipo, n_especies=distance)
#'
#' Puedes notar que ambos sitios comparten 11 especies ( $a$ ), que el sitio 2 no tiene especies
#'
round(11/12*100,2) #Porcentaje de especies compartidas = similaridad
#'
#' Con 'betadiver' también puedes calcular índices de similaridad. Por ejemplo, el Jaccard se ca
#'
betadiver(mc_malvc, method = 'j') %>% tidy
#'
#' No obstante, usaremos esta función en los análisis de diversidad beta más adelante.
#'
#' Además de la distancia de Jaccard, otra distancia muy utilizada es la de Sorensen o Bray-Curt
#'
mi_fam_sor <- vegdist(mc_malvc, method = 'bray', binary = T)
mi_fam_sor %>% tidy
coldiss(mi_fam_sor, diag = T)
#'
#' ### Modo Q para datos cuantitativos, NO de abundancia de especies (variables ambientales)
#'
#' En este ejemplo, usaré sólo variables de suelo, todas cuantitativas, puedes combinar con otra
#'
env_suelo_punt_z <- bci_env_grid %>%
  st_drop_geometry() %>%
  dplyr::select(matches('^[A-T,Z]|^pH$', ignore.case = F)) %>%
  scale()

```

```

env_suelo_punt_z_d <- dist(env_suelo_punt_z)
env_suelo_punt_z_d %>% tidy
coldiss(env_suelo_punt_z_d, diag = T)
#'
#' ### Modo Q para datos cualitativos y cuantitativos (mixtos), NO de abundancia de especies (va
#'
#' En este ejemplo, usaré las siguientes variables mixtas (funciona igualmente para datos cualit
#'
#' - 'heterogeneidad_ambiental'. Índice cuantitativo calculado como la diversidad de Simpson a
#'
#' - 'habitat'. Tipo de hábitat. Asume los siguientes valores posibles: *OldHigh*, *OldLow* y *O
#'
#' - 'quebrada'. Informa sobre si hay o no quebrada. Los valores posibles son *Yes* o *No*.
#'
env_mix <- bci_env_grid %>%
  st_drop_geometry() %>%
  dplyr::select(heterogeneidad_ambiental, habitat, quebrada)
env_mix_d <- daisy(x = env_mix, metric = 'gower')
env_mix_d %>% as.dist %>% tidy
env_mix_d %>% coldiss(diag = T)

```

## 7.8 Edición de asociación. Modo R aplicado a mi familia asignada

```
r paste (readLines('ma_3:Medicion_de_asociación.R'), collapse = '\n')
```

## 7.9 Análisis de agrupamiento (cluster analysis). Parte 2: Interpretación y comparación de resultados

```

#' ---
#' title: "Análisis de agrupamiento (cluster analysis). <br> Parte 2: Interpretación y comparaci
#' author: "JR"
#' date: "11 de noviembre, 2020"
#' output: github_document
#' ---

```

```
knitr::opts_chunk$set(fig.width=12, fig.height=8)
```

```

#' ## Preámbulo
#'
#' ### Cargar paquetes
#'
library(vegan)
library(tidyverse)
library(broom)
library(cluster)
library(gclus)
library(pvclust)

```

```

library(sf)
source('biodata/funciones.R')
#'
#' ### Cargar datos
#'
load('biodata/Malvaceae.Rdata')
mi_fam <- mc_malvc
load('biodata/matriz_ambiental.Rdata')
mi_fam %>% tibble
bci_env_grid %>% tibble
#'
#' ### Generar matriz de distancias de cuerdas
#'
mi_fam_norm <- decostand(mi_fam, "normalize")
mi_fam_norm_d <- vegdist(mi_fam_norm, "euc")
mi_fam_norm_d %>% tidy
#'
#' ## Interpretación visual de dendrogramas
#'
#' [En el script anterior](aa_analisis_de_agrupamiento_1_jerarquico.md) realicé los dendrogramas
#'
#' La primera evaluación de los dendrogramas NO debe venir de la mano de sofisticados análisis n
#'
#' Evita concentrar tu vista en grupos extremadamente pequeños; comienza analizando el árbol des
#'
#' Anota tus impresiones, para que las compares con los resultados que posteriormente obtendrás;
#'
#' Para la exploración visual, generaré los objetos de cluster dentro de una lista:
#'
lista_cl <- list(
  cl_single = hclust(mi_fam_norm_d, method = 'single'),
  cl_complete = hclust(mi_fam_norm_d, method = 'complete'),
  cl_upgma = hclust(mi_fam_norm_d, method = 'average'),
  cl_ward = hclust(mi_fam_norm_d, method = 'ward.D2')
)
#'
#' Un plot en panel 2x2 ayuda a visualizarlos todos de manera conjunta. En tu caso, observa y co
#'
par(mfrow = c(2,2))
invisible(map(names(lista_cl), function(x) plot(lista_cl[[x]], main = x, hang = -1)))
par(mfrow = c(1,1))
#'
#' En mi caso, exceptuando el dendrograma generado por medio del enlace simple, detecto al menos
#'
#' - Un grupo pequeño, compuesto por los sitios 1, 42, 12, 21, 11, 2 y 16.
#' - Un "grupo" heterogéneo y grande, conformado por 25, 31,..., 26,..., 35,..., 34,...,32, 17,..
#'
#' Además de los grupos anteriores, detecto elementos que no forman grupos, es decir, sitios que

```

```

#’
#’ ## Elegir método y número de clústers
#’
#’ Existen varios criterios para elegir un dendrograma idóneo, como por ejemplo, los gráficos ti
#’
#’ ### Seleccionar método de agrupamiento por correlación cofenética
#’
#’ La correlación cofenética implica conocer la distancia cofenética, y esta última se entiende m
#’
#’ Usando la lista de objetos de clústers, calcularé la correlación cofenética dentro de un ‘map
#’
map_df(lista_cl, function(x) {
  coph_d <- cophenetic(x)
  corr <- cor(mi_fam_norm_d, coph_d)
  return(corr)
})
#’
#’ Habrás notado que, tanto UPGMA como enlace completo, tienen valores altos de correlación cofe
#’
#’ ### Elegir número de clústers
#’
#’ Elegiré UPGMA como método de agrupamiento y determinaré cuántos grupos son idóneos de acuerdo
#’
#’ El objetivo de la función ‘calcular_anchuras_siluetas’ está implícito en su nombre, y requiere
#’
#’ 1. Las anchuras promedio para cada partición, excepto para las particiones ‘i=1’ y ‘i=50’, po
#’
#’ 2. Número óptimo de grupos. Haré los cálculos para UPGMA y Ward, y luego explico en qué consi
#’
#’ Para UPGMA:
#’
anch_sil_ward <- calcular_anchuras_siluetas(
  mc_orig = mi_fam,
  distancias = mi_fam_norm_d,
  cluster = lista_cl$cl_ward)
anch_sil_ward
#’
#’ El objeto ‘anchuras_siluetas’ de la lista ‘anch_sil_upgma’ te muestra un vector con los prome
#’
#’ Igualmente, el objeto ‘n_grupos_optimo’ te indica cuál es el número óptimo de clústers a crea
#’
#’ Haré el gráfico de dendrograma, aunque nota que en este caso primero reordenaré los sitios co
#’
u_dend_reord <- reorder.hclust(lista_cl$cl_ward, mi_fam_norm_d)
plot(u_dend_reord, hang = -2)
rect.hclust(
  tree = u_dend_reord,
  k = anch_sil_ward$n_grupos_optimo)

```



```

#’
#’ Ahora compararé el dendrograma con el mapa de calor en un mismo gráfico, colocando los dendro
#’
heatmap(
  as.matrix(mi_fam_norm_d),
  Rowv = as.dendrogram(u_dend_reord),
  symm = TRUE,
  margin = c(3, 3),
  col = rev(cm.colors(4))
)
#’
#’ En general, hay dos grupos, uno grande y otro pequeño, y parece haber un tercero en el mapa d
#’
#’ Mostraré el resultado para Ward:
#’
anch_sil_ward <- calcular anchuras siluetas(
  mc_orig = mi_fam,
  distancias = mi_fam_norm_d,
  cluster = lista_cl$cl_ward)
anch_sil_ward
#’
#’ En este caso, el valor máximo, que ocupa la posición número 2, no se diferencia mucho del de
#’
w_dend_reord <- reorder.hclust(lista_cl$cl_ward, mi_fam_norm_d)
plot(w_dend_reord, hang = -1)
rect.hclust(
  tree = w_dend_reord,
  k = anch_sil_ward$n_grupos_optimo)
plot(w_dend_reord, hang = -1)
rect.hclust(
  tree = w_dend_reord,
  k = anch_sil_ward$n_grupos_optimo + 1)
#’
#’ Comparando el dendrograma con el mapa de calor. Verificar si el número de grupos hace sentido
#’
heatmap(
  as.matrix(mi_fam_norm_d),
  Rowv = as.dendrogram(w_dend_reord),
  symm = TRUE,
  margin = c(3, 3),
  col = rev(cm.colors(4))
)
#’
#’ Nótese que este dendrograma hace más sentido que el sugerido por UPGMA. En cualquier casos, c
#’
#’ ### Evaluación mediante remuestreo por *bootstrap* multiescalar
#’
#’ Con suerte, un agrupamiento aplicado a datos muestrales reflejará los patrones naturales de o

```

```

#’
#’ Sin embargo, los datos de BCI también tienen sesgo, pues se usa un DAP de corte para decidir
#’
#’ No obstante, aun con todas sus bondades, los datos censales carecen de una fortaleza: no refl
#’
#’ Por estas razones, los análisis de agrupamientos realizados hasta este punto, reflejan tanto
#’
#’ El método de remuestreo por *bootstrap* multiescalar está implementado en el paquete ‘pvclust’
#’
#’ La función ‘pvclust’ devolverá un dendrograma enriquecido, que incluirá los valores de AU y BP
#’
#’ - Rectángulos de borde azul, para todos aquellos grupos que resulten con valores de AU>0.91 e
#’
#’ - Líneas inferiores rojas, que resaltan aquellos grupos (o subgrupos) que obtuvieron AU>0.95.
#’
#’ Ten presente que, al realizar remuestreo por *bootstrap* multiescalar, cada corrida puede arr
#’
#’ ##### UPGMA
#’
cl_pvclust_ward<-
  pvclust(t(mi_fam_norm),
    method.hclust = "average",
    method.dist = "euc",
    iseed = 91, # Resultado reproducible
    parallel = TRUE)
# Añadir los valores de p
plot(cl_pvclust_ward, hang = -1)
# Añadir rectángulos a los grupos significativos
lines(cl_pvclust_ward)
pvrect(cl_pvclust_ward, alpha = 0.91, border = 4)
#’
#’ ##### Ward
#’
cl_pvclust_ward <-
  pvclust(t(mi_fam_norm),
    method.hclust = "ward.D2",
    method.dist = "euc",
    iseed = 191, # Resultado reproducible
    parallel = TRUE)
# Añadir los valores de p
plot(cl_pvclust_ward, hang = -1)
# Añadir rectángulos a los grupos significativos
lines(cl_pvclust_ward)
pvrect(cl_pvclust_ward, alpha = 0.91, border = 4)
#’
#’ ### Recapitulando los grupos de sitios.
#’
#’ ##### Patrones comunes y dispares

```

```

#’
#’ Detecto algunos patrones consistentes en cuanto a grupos de sitios según composición de las e
#’
#’ - Tanto en UPGMA como en Ward, detecté al menos dos o tres grandes grupos. Con el primer méto
#’
#’ - En el caso específico del dendrograma Ward, ‘pvclust’ atomizó los sitios en demasiados grup
#’
#’ ##### ¿Cómo declaro los grupos de sitios?
#’
#’ Para conservar las clasificaciones de grupos de sitios anteriores, crearé un vector con el id
#’
#’ UPGMA:
(grupos_ward_k2 <- as.factor(cutree(lista_cl$cl_ward, k = 2)))
#’
#’ En este caso, los sitios 1 y 2 pertenecen al grupo 1, los sitios 3 al 6 pertenecen al grupo 2
#’
table(grupos_ward_k2)
#’
#’ Nota lo desiguales que son estos grupos, un efecto esperado dado el alto grado de autocorrela
#’
#’ Ward:
#’
(grupos_ward_k2 <- as.factor(cutree(lista_cl$cl_ward, k = 2)))
table(grupos_ward_k2)
#’
#’ Guardaré estos vectores en archivos para reutilizarlos en *scripts* posteriores:
#’
saveRDS(grupos_ward_k2, 'grupos_ward_k2.RDS')
saveRDS(grupos_ward_k2, 'grupos_ward_k2.RDS')
#’
#’ Evita usar este, y cualquier otro procedimiento, de manera mecánica. En tu caso, quizá tengas
#’
#’ En el próximo *script*, aprenderás a comparar este resultado con las variables ambientales. T

```

## 7.10 *Análisis de agrupamiento (cluster analysis). Parte 3: Grupos (clústers), variables ambientales y mapas*

```

#’ ---
#’ title: "Análisis de agrupamiento (cluster analysis). <br> Parte 3: Grupos (clústers), variabl
#’ author: "JR"
#’ date: "15 de noviembre, 2020"
#’ output: github_document
#’ ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#’ ## Preámbulo

```

```

#'
#' ### Cargar paquetes
#'
library(mapview)
library(tidyverse)
library(sf)
library(RColorBrewer)
source('biodata/funciones.R')
#'
#' ### Cargar datos
#'
load('biodata/Malvaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
grupos_ward_k2 <- readRDS('grupos_ward_k2.RDS')
table(grupos_ward_k2) #Importante, tener en cuenta los desiguales tamaños de los grupos
grupos_ward_k2 <- readRDS('grupos_ward_k2.RDS')
table(grupos_ward_k2)
#'
#' ### Paletas
#'
rojo <- colorRampPalette(brewer.pal(8, "Reds"))
rojo_inv <- colorRampPalette(rev(brewer.pal(8, "Reds")))
colores_grupos <- brewer.pal(8, "Accent")
#'
#' ## Explorar efectos
#'
#' ### Pruebas de igualdad de promedios de las variables entre 2 grupos
#'
#' Para evaluar homogeneidad de promedios usaré las pruebas *t* (medias), basada en la distribución
#'
#' Primero crearé un objeto que permita realizar tanto las pruebas como los diagramas de cajas.
#'
(m_amb_ward_k2 <- bci_env_grid %>%
  select_if(is.numeric) %>% select(-id) %>%
  mutate(grupos_ward_k2) %>%
  st_drop_geometry() %>%
  pivot_longer(-grupos_ward_k2, names_to = "variable", values_to = "valor"))
#'
#' A continuación, las pruebas:
#'
m_amb_ward_k2 %>%
  group_by(variable) %>%
  summarise(
    p_valor_t = t.test(valor ~ grupos_ward_k2)$p.value,
    p_valor_w = wilcox.test(valor ~ grupos_ward_k2, exact = F)$p.value) %>%
  arrange(p_valor_t) %>%
  print(n=Inf)
#'

```

```

#’ Interesa observar las variables que obtuvieron valores de  $p < 0.01$ . Reitero que, en mi caso, mi
#’
#’ Por otra parte, este es un buen momento para "revisitar" tus análisis exploratorios de datos
#’
#’ Los gráficos:
#’
m_amb_ward_k2 %>%
  group_by(variable) %>%
  ggplot() + aes(x = grupos_ward_k2, y = valor, fill = grupos_ward_k2) +
  geom_boxplot() +
  scale_fill_brewer(palette = 'Accent') +
  theme_bw() +
  theme(legend.position="none") +
  facet_wrap(~ variable, scales = 'free_y')
#’
#’ Mapas:
#’
mapa_upgma_k2 <- mapView(
  bci_env_grid %>% mutate(grupos_upgma_k2),
  layer.name = 'Grupos (2) UPGMA',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,
  col.regions = colores_grupos[1:2],
  zcol = 'grupos_upgma_k2') %>%
  addStaticLabels(label = bci_env_grid$id) %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_upgma_k2
mapa_upgma_k2 %>% mapshot(
  file = 'mapa_upgma_k2.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#’
#’ Mapa de una de las variables donde se presentó efecto de su promedio ( $p < 0.01$ ), en este caso,
#’
mapa_zn <- mapView(
  bci_env_grid,
  layer.name = 'Zinc',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,
  col.regions = rojo,
  zcol = 'Zn') %>%
  addStaticLabels(label = bci_env_grid$id) %>%
  leaflet::setView(

```

```

    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_zn
mapa_zn %>% mapshot(
  file = 'mapa_zinc.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
#' ### Pruebas de igualdad de promedios de las variables entre 3 grupos o más
#'
#' Objeto común:
#'
(m_amb_ward_k3 <- bci_env_grid %>%
  select_if(is.numeric) %>% select(-id) %>%
  mutate(grupos_ward_k3) %>%
  st_drop_geometry() %>%
  pivot_longer(-grupos_ward_k3, names_to = "variable", values_to = "valor"))
#'
#' Pruebas, en este caso ANOVA (evalúa homogeneidad de medias; no se cumplen muchos de los supue
#'
m_amb_ward_k3 %>%
  group_by(variable) %>%
  summarise(
    p_valor_a = oneway.test(valor ~ grupos_ward_k3)$p.value,
    p_valor_k = kruskal.test(valor ~ grupos_ward_k3)$p.value) %>%
  arrange(p_valor_k) %>%
  print(n=Inf)
#'
#' Gráficos:
#'
m_amb_ward_k3 %>%
  group_by(variable) %>%
  ggplot() + aes(x = grupos_ward_k3, y = valor, fill = grupos_ward_k3) +
  geom_boxplot() +
  scale_fill_brewer(palette = 'Accent') +
  theme_bw() +
  theme(legend.position="none") +
  facet_wrap(~ variable, scales = 'free_y')
#'
#' Mapas:
#'
mapa_ward_k3 <- mapView(
  bci_env_grid %>% mutate(grupos_ward_k3),
  layer.name = 'Grupos (3) Ward',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,

```

```

col.regions = colores_grupos[1:3],
zcol = 'grupos_ward_k3') %>%
addStaticLabels(label = bci_env_grid$id) %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)
mapa_ward_k3
mapa_ward_k3 %>% mapshot(
  file = 'mapa_ward_k3.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
#' Mapa de una de las variables donde se presentó efecto de su promedio (p<0.01), en este caso,
#'
mapa_ph <- mapView(
  bci_env_grid,
  layer.name = 'pH',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,
  col.regions = rojo_inv,
  zcol = 'pH') %>%
addStaticLabels(label = bci_env_grid$id) %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)
mapa_ph
mapa_ph %>% mapshot(
  file = 'mapa_ph.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'

```

### 7.11 *Análisis de agrupamiento (cluster analysis). Parte 4: Especies indicadoras, especies con preferencia por hábitats*

```

#' ---
#' title: "Análisis de agrupamiento (cluster analysis). <br> Parte 4: Especies indicadoras, espe
#' author: "JR"
#' date: "15 de noviembre, 2020"
#' output: github_document
#' ---

```

```
knitr::opts_chunk$set(fig.width=12, fig.height=8)
```

```

#' ## Preámbulo
#'
#' ### Cargar paquetes
#'
library(indicspecies)
source('biodata/funciones.R')
#'
#' ### Cargar datos
#'
load('biodata/Malvaceae.Rdata')
mi_fam <- mc_malvc
grupos_upgma_k2 <- readRDS('grupos_upgma_k2.RDS')
table(grupos_upgma_k2)
grupos_ward_k2 <- readRDS('grupos_ward_k2.RDS')
table(grupos_ward_k2)
#'
#' ## Análisis de especies indicadoras mediante IndVal
#'
#' ### UPGMA
#'
iva_upgma_k2 <- multipatt(
  x = mi_fam,
  cluster = grupos_upgma_k2,
  func = 'IndVal.g',
  max.order = 1,
  control = how(nperm = 999))
summary(iva_upgma_k2, indvalcomp = TRUE)
colSums(mi_fam)
(p_upgma_adj <- p.adjust(iva_upgma_k2$sign$p.value))
(iva_upgma_boot <- strassoc(
  X = mi_fam,
  cluster = grupos_upgma_k2,
  func = "IndVal.g",
  nboot = 1000))
#'
#' Ward
#'
iva_ward_k2 <- multipatt(
  x = mi_fam,
  cluster = grupos_ward_k2,
  func = 'IndVal.g',
  max.order = 2,
  control = how(nperm = 999))
summary(iva_ward_k2, indvalcomp = TRUE)
colSums(mi_fam)
(p_ward_adj <- p.adjust(iva_ward_k2$sign$p.value))
(iva_ward_boot <- strassoc(
  X = mi_fam,

```



```

    cluster = grupos_ward_k2,
    func = "IndVal.g",
    nboot = 1000))
#'
#' ## Análisis de especies con preferencia por hábitat mediante el coeficiente de correlación bi
#'
#' ### UPGMA
#'
phi_upgma_k2 <- multipatt(
  mi_fam,
  grupos_upgma_k2,
  func = "r.g",
  max.order = 1,
  control = how(nperm = 999))
summary(phi_upgma_k2)
colSums(mi_fam)
(phi_upgma_boot <- strassoc(
  X = mi_fam,
  cluster = grupos_upgma_k2,
  func = "r.g",
  nboot = 1000))
#'
#' Ward
#'
phi_ward_k2 <- multipatt(
  mi_fam,
  grupos_ward_k2,
  func = "r.g",
  max.order = 2,
  control = how(nperm = 999))
summary(phi_ward_k2)
colSums(mi_fam)
(phi_ward_boot <- strassoc(
  X = mi_fam,
  cluster = grupos_ward_k2,
  func = "r.g",
  nboot = 1000))
#'

```

## 7.12 Técnicas de ordenación. Parte 1: Ordenación no restringida. PCA, CA y PCoA

```

#' ---
#' title: "Técnicas de ordenación. <br> Parte 1: Ordenación no restringida. <br> PCA, CA y PCoA"
#' author: "JR"
#' date: "21 de noviembre, 2020"
#' output: github_document
#'

```

```

#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo
#'
#' ### Cargar paquetes
#'
library(vegan)
library(tidyverse)
library(sf)
library(mapview)
source('biodata/funciones.R')
#'
#' ### Cargar datos
#'
load('biodata/Malvaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
mi_fam <- mc_malvc
(colnames(mi_fam) <- make.cepnames(colnames(mi_fam)))
(df_equivalecias <- data.frame(
  nombre_original = colnames(mc_malvc),
  colnames(mi_fam)))
bci_env_grid %>% tibble
grupos_upgma_k2 <- readRDS('grupos_upgma_k2.RDS')
table(grupos_upgma_k2)
grupos_ward_k3 <- readRDS('grupos_ward_k3.RDS')
table(grupos_ward_k3)
#'
#' ## Ordenación
#'
#' La ordenación se basa en los mismos principios que la medición de asociación (similaridad) y
#'
#' A diferencia del análisis de agrupamiento, o como complemento de éste, el análisis de ordenación
#'
#' El análisis de ordenación puede ser no restringido (o simple) y restringido (o 'canónico'). En
#'
#' Las principales técnicas de ordenación no restringida son análisis de componentes principales
#'
#' ### Análisis de componentes principales (PCA)
#'
#' Es el método tradicional basado en vectores propios que comúnmente se aplica a datos cuantitativos
#'
#' #### PCA aplicado a datos ambientales
#'
#' Para aplicar PCA a datos ambientales, es necesario que todas las variables sean numéricas y
#'
#' Dado que se requiere que las variables de entrada sean exclusivamente numéricas, el primer pa

```

```

#’
#’ ¡IMPORTANTE! Haré esta demostración sólo con las variables de suelo, **pero puedes (y debes)
#’
#’ A partir de los datos de suelo, la función ‘rda’, de ‘vegan’ realizará los siguientes pasos:
#’
env_suelo <- bci_env_grid %>%
  st_drop_geometry %>%
  dplyr::select(matches('^[A-T,Z]|^pH$', ignore.case = F))
env_suelo %>% tibble
env_suelo_pca <- rda(env_suelo, scale = TRUE)
env_suelo_pca
summary(env_suelo_pca)
#’
#’ Para agilizar la producción de scripts analíticos de referencia, trasladaré las explicaciones
#’
#’ En el vídeo asociado, explico el significado de:
#’
#’ - Inercia, *Inertia*
#’ - Valores propios, autovalores, *Eigenvalues*
#’ - Escalamiento, *Scaling*
#’ - Puntuaciones de "especies", *Species scores*
#’ - Puntuaciones de "sitios", *Site scores*
#’
screeplot(env_suelo_pca, bstick = TRUE)
#’
#’ Usando función ‘cleanplot.pca’
#’
par(mfrow = c(1, 2))
cleanplot.pca(env_suelo_pca, scaling = 1, mar.percent = 0.08, cex.char1 = 1.5)
cleanplot.pca(env_suelo_pca, scaling = 2, mar.percent = 0.04, cex.char1 = 1.5)
par(mfrow = c(1, 1))
#’
#’ Comparar distribución de los sitios en biplots con distribución real en el mapa:
#’
#’ ### Generar mapa de cuadros sin simbología
#’
mapa_cuadros <- mapView(
  bci_env_grid,
  col.regions = 'grey80',
  alpha.regions = 0.3,
  map.types = 'OpenTopoMap',
  legend = F, zoom = 14,
  zcol = 'id') %>% addStaticLabels() %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_cuadros

```

```

#’
#’ Comparar con resultados de un análisis de agrupamiento del mismo conjunto de datos. Primero a
#’
(env_agrupamiento <- hclust(dist(scale(env_suelo)), 'ward.D'))
(env_grupos <- cutree(env_agrupamiento, k = 3))
(mi_cluster <- factor(env_grupos))
(mi_cluster_l <- levels(mi_cluster))
(mi_cluster_l_seq <- 1:length(mi_cluster_l))
#’
#’ Observa que estoy generando un agrupamiento basado en los datos de suelo. No estoy comparando
#’
#’ Luego calculo las puntuaciones de los sitios para usarlas luego como coordenadas de los puntos
#’
(puntuaciones <- scores(env_suelo_pca, display = 'wa', scaling = 1))
#’
#’ Luego creo el gráfico base, coloco los puntos sobre el gráfico usando las puntuaciones, les c
#’
grafico_base <- plot(
  env_suelo_pca,
  display = "wa",
  scaling = 1,
  type = "n",
  main = "PCA y grupos"
)
abline(v = 0, lty = "dotted")
abline(h = 0, lty = "dotted")
for (i in mi_cluster_l_seq) {
  points(puntuaciones[mi_cluster == i, ],
    pch = (14 + i),
    cex = 2,
    col = i + 1)
}
text(puntuaciones, row.names(env_suelo), cex = 1, pos = 3)
legend(
  "topright", # Otras alternativas: "bottomleft", "bottomright" y "topleft"
  paste("Grupo", c(mi_cluster_l_seq)),
  pch = 14 + c(mi_cluster_l_seq),
  col = 1 + c(mi_cluster_l_seq),
  pt.cex = 2
)
#’
#’ Es razonable que el análisis cluster y el biplot muestren patrones consistentes, puesto que s
#’
#’ Si hago lo mismo, pero usando mi análisis de agrupamiento anterior (*scripts* "aa_analisis_de
#’
# (mi_cluster_anterior <- grupos_upgma_k2)
(mi_cluster_anterior <- grupos_ward_k3)
(mi_cluster_anterior_l <- levels(mi_cluster_anterior))

```

```

(mi_cluster_anterior_l_seq <- 1:length(mi_cluster_anterior_l))
grafico_base <- plot(
  env_suelo_pca,
  display = "wa",
  scaling = 1,
  type = "n",
  main = "PCA y grupos"
)
abline(v = 0, lty = "dotted")
abline(h = 0, lty = "dotted")
for (i in mi_cluster_anterior_l_seq) {
  points(puntuaciones[mi_cluster_anterior == i, ],
    pch = (14 + i),
    cex = 2,
    col = i + 1)
}
text(puntuaciones, row.names(env_suelo), cex = 1, pos = 3)
legend(
  "topright", # Otras alternativas: "bottomleft", "bottomright" y "topleft"
  paste("Grupo", c(mi_cluster_anterior_l_seq)),
  pch = 14 + c(mi_cluster_anterior_l_seq),
  col = 1 + c(mi_cluster_anterior_l_seq),
  pt.cex = 2
)
#
#' Esto podría significar que las tendencias/patrones de mi matriz de comunidad (cuadros de 1 Ha
#
#' ##### PCA aplicado a datos de comunidad transformados
#
mi_fam_hel <- decostand(mi_fam, method = 'hellinger')
mi_fam_hel %>% tibble
mi_fam_hel_pca <- rda(mi_fam_hel)
summary(mi_fam_hel_pca)
screeplot(
  mi_fam_hel_pca,
  bstick = TRUE,
  npcs = length(mi_fam_hel_pca$CA$eig)
)
mi_fam_hel_pca_sc1 <- scores(mi_fam_hel_pca,
  display = "species", scaling = 1)
mi_fam_hel_pca_sc2 <- scores(mi_fam_hel_pca,
  display = "species", scaling = 2)

par(mfrow = c(1, 2))
cleanplot.pca(mi_fam_hel_pca, scaling = 1, mar.percent = 0.06, cex.char1 = 0.7)
cleanplot.pca(mi_fam_hel_pca, scaling = 2, mar.percent = 0.06, cex.char1 = 0.7)
par(mfrow = c(1, 1))
#
#' Si intentáramos realizar el PCA a datos de comunidad no transformados, no recogeríamos apropi

```

```

#’
#’ Las especies que contribuyen mucho a los ejes 1 y 2 del PCA (aquellas cuyos vectores sobresalen)
#’
#’ Evaluaré el ajuste del PCA de datos de comunidad a datos ambientales, mediante la función ‘envfit’
#’
biplot(
  mi_fam_hel_pca,
  main = "PCA, escalamiento 2, ajuste a variables ambientales")
(mi_fam_hel_pca_envfit <- envfit(mi_fam_hel_pca, env_suelo, scaling = 2))
plot(mi_fam_hel_pca_envfit, p.max = 0.05 , col = 3)
#’
#’ Comento los resultados en el vídeo asociado. También probaré ajuste con todas las numéricas de la comunidad
#’
#’ NOTA: te recomiendo probar otros métodos de selección de variables, como por ejemplo, usando stepAIC
#’
env_num <- bci_env_grid %>%
  select_if(is.numeric) %>%
  select(-id) %>%
  st_drop_geometry
(mi_fam_hel_pca_envfit_num <- envfit(mi_fam_hel_pca, env_num, scaling = 2))
biplot(
  mi_fam_hel_pca,
  main = "PCA, escalamiento 2, ajuste a variables ambientales")
plot(mi_fam_hel_pca_envfit_num, p.max = 0.05 , col = 3)
biplot(
  mi_fam_hel_pca,
  main = "PCA, escalamiento 2, ajuste a variables ambientales")
plot(mi_fam_hel_pca_envfit_num, p.max = 0.1 , col = 3)
#’
#’ Comento los resultados en el vídeo asociado.
#’
#’ ¿Cuándo o a qué datos aplicar PCA?
#’
#’ - PCA no es especialmente sensible a datos muy desviados de la normalidad.
#’ - Como toda técnica, PCA tiene limitaciones.
#’ - Las variables deben ser dimensionalmente homogéneas (unidades comparables o adimensionales)
#’ - No usar en matriz transpuestas (no hace sentido la covarianza entre objetos).
#’ - Es posible usar PCA con dato de presencia/ausencia, en cuyo caso, la matriz de comunidad debe ser de tipo
#’ - Las relaciones entre variables se miden por ángulos, no por proximidad de las puntas de los vectores
#’
#’ ### Análisis de correspondencia (CA)
#’
mi_fam_ca <- cca(mi_fam)
summary(mi_fam_ca)
summary(mi_fam_ca, scaling = 1)
#’
#’ Screeplot
#’

```

```

screplot(mi_fam_ca, bstick = TRUE, npcs = length(mi_fam_ca$CA$eig))
#'
#' Biplots
#'
par(mfrow = c(1, 2))
plot(mi_fam_ca,
     scaling = 1,
     main = "Análisis de correspondencia, escalamiento 1"
)
plot(mi_fam_ca,
     scaling = 2, # Por defecto scaling=2, lo escribo sólo para fines didáticos
     main = "Análisis de correspondencia, escalamiento 2")
par(mfrow = c(1, 1))
#'
#' Excluyendo especie *Thevetia ahouai*, abreviada como *Thevahou*.
#'
mi_fam_ca <- cca(mi_fam[, -grep('Pachsess', colnames(mi_fam))])
summary(mi_fam_ca)
summary(mi_fam_ca, scaling = 1)
screplot(mi_fam_ca, bstick = TRUE, npcs = length(mi_fam_ca$CA$eig))
par(mfrow = c(1, 2))
plot(mi_fam_ca,
     scaling = 1,
     main = "CA, escalamiento 1, sin Thevetia ahouai"
)
plot(mi_fam_ca,
     scaling = 2,
     main = "CA, escalamiento 2, sin Thevetia ahouai")
par(mfrow = c(1, 1))
#'
#' Análisis de coordenadas principales (PCoA)
#'
#' Las técnicas de ordenación anteriores preservan la distancia euclídea entre los objetos. Si n
#'
#' La función que realiza el PCoA en '{vegan}' es 'cmdscale' (de *Classical (Metric) Multidimens
#'
mi_fam_d_bray <- vegdist(mi_fam, method = 'bray') # En realidad, 'bray' es la opción por defecto
mi_fam_d_bray_pcoa <- cmdscale(
  mi_fam_d_bray,
  k = (nrow(mi_fam) - 1),
  add = T,
  eig = TRUE)
round(mi_fam_d_bray_pcoa$eig, 2)
round(sum(mi_fam_d_bray_pcoa$eig[mi_fam_d_bray_pcoa$eig<0]),2)
round(sum(mi_fam_d_bray_pcoa$eig[mi_fam_d_bray_pcoa$eig>=0]),2)
ordiplot(scores(mi_fam_d_bray_pcoa, choices = c(1, 2)),
         type = "t",
         main = "PCoA con promedios ponderados de especies")

```

```

abline(h = 0, lty = 3)
abline(v = 0, lty = 3)
mi_fam_d_bray_pcoa_wa <- wascores(mi_fam_d_bray_pcoa$points[, 1:2], mi_fam)
text(
  mi_fam_d_bray_pcoa_wa,
  rownames(mi_fam_d_bray_pcoa_wa),
  cex = 0.7, col = "red")
(mi_fam_d_bray_pcoa_env <- envfit(mi_fam_d_bray_pcoa, env_num))
plot(mi_fam_d_bray_pcoa_env, p.max = 0.05, col = 3)

```

### 7.13 Técnicas de ordenación. Parte 2: Ordenación restringida o 'canónica'. RDA, CCA

```

#' ---
#' title: "Técnicas de ordenación. <br> Parte 2: Ordenación restringida o 'canónica'. <br> RDA,
#' author: "JR"
#' date: "21 de noviembre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo
#'
#' ### Cargar paquetes
#'
library(vegan)
library(tidyverse)
library(sf)
source('biodata/funciones.R')
#'
#' ### Cargar datos
#'
load('biodata/Malvaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
mi_fam <- mc_malvc
(colnames(mi_fam) <- make.cepnames(colnames(mi_fam)))
(df_equivalencias <- data.frame(
  nombre_original = colnames(mc_malvc),
  colnames(mi_fam)))
bci_env_grid %>% tibble
#'
#' ## Ordenación restringida
#'
#' Con este conjunto de técnicas de ordenación, las tendencias detectadas en un conjunto de dato
#'
#' Las principales técnicas de ordenación restringida son análisis de redundancia o RDA (siglas
#'

```



```

#' ### Análisis de redundancia (RDA)
#'
#' En el análisis de ordenación simple o no restringida, la matriz de datos (de comunidad o ambi
#'
#' RDA combina la regresión y el análisis de componentes principales. El RDA es una regresión li
#'
#' - Cada variable de respuesta (e.g. matriz de comunidad) es ajustada (por regresión) a la matr
#'
#' - Se prueba significancia estadística de la relación entre las variables de respuesta y expli
#'
#' - Si la prueba resulta significativa, se realiza un PCA de la matriz de valores ajustados. Es
#'
#' - Luego se usa la matriz de vectores propios canónicos para calcular dos tipos de puntuacione
#'
#'     - Usando la matriz de valores ajustados para obtener una ordenación en el espacio de las
#'
#'     - Usando la matriz de valores centrados de las variables de respuesta para obtener una or
#'
#' Mostraré a continuación ejemplos aplicados a mi familia asignada. Para agilizar, trasladaré l
#'
#' Nota (de nuevo): recuerda que los datos de BCI están autocorrelacionados espacialmente, por l
#'
#' #### Ejemplo usando las matriz ambiental de variables suelo:
#'
mi_fam_hel <- decostand(mi_fam, method = 'hellinger')
mi_fam_hel %>% tibble
env_suelo <- bci_env_grid %>%
  st_drop_geometry %>%
  dplyr::select(matches('^[A-T,Z]|~pH$', ignore.case = F))
env_suelo %>% tibble
mi_fam_hel_rda_suelo <- rda(mi_fam_hel ~ ., env_suelo)
summary(mi_fam_hel_rda_suelo)
#'
#' ¿Qué partes del resumen debes mirar?
#'
#' - La varianza, particionada, de la cual se muestra tanto la inercia (absoluta) como la propor
#'
#' - Los valores propios (*eigenvalues*) y su contribución a la varianza. Existen *eigenvalues*
#'
#' - Las puntuaciones de sitios (*site scores*) y de especies (*species scores*) tienen el mismo
#'
#' - Las puntuaciones restringidas de sitio (en '{vegan}' se usa la etiqueta *Site constraints (
#'
#' - Finalmente, el resumen presenta las puntuaciones para el *biplot*, que más adelante usaré e
#'
#' Tal como comenté arriba, es importante calcular un  $R^2$  insesgado, siendo este más ú
#'
RsquareAdj(mi_fam_hel_rda_suelo)$adj.r.squared

```

```

#'
#' Normalmente, el  $R^2$  insesgado es mucho más bajo que el sesgado, porque se relativiza
#'
#' Otro diagnóstico importante es la determinación de los factores de inflación de la varianza (
#'
vif.cca(mi_fam_hel_rda_suelo)
#'
#' Variables con valores por encima de 10 deben ser examinadas y, desde una posición conservadora
#'
#' Finalmente, la representación del modelo se realiza en un *triplot*, que es un gráfico enriquecido
#'
#' Escalamiento 1:
#'
plot(mi_fam_hel_rda_suelo,
     scaling = 1,
     display = c("sp", "lc", "cn"),
     main = "Triplot de RDA especies ~ var. suelo, escalamiento 1"
)
mi_fam_hel_rda_suelo_sc1 <-
  scores(mi_fam_hel_rda_suelo,
        choices = 1:2,
        scaling = 1,
        display = "sp"
  )
arrows(0, 0,
       mi_fam_hel_rda_suelo_sc1[, 1] * 0.9,
       mi_fam_hel_rda_suelo_sc1[, 2] * 0.9,
       length = 0,
       lty = 1,
       col = "red"
)
#'
#' Paso la interpretación de este gráfico y la del siguiente al vídeo correspondiente.
#'
#' Escalamiento 2
#'
plot(mi_fam_hel_rda_suelo,
     scaling = 2,
     display = c("sp", "lc", "cn"),
     main = "Triplot de RDA especies ~ var. suelo, escalamiento 2"
)
mi_fam_hel_rda_suelo_sc2 <-
  scores(mi_fam_hel_rda_suelo,
        scaling = 2,
        choices = 1:2,
        display = "sp"
  )
arrows(0, 0,

```

```

        mi_fam_hel_rda_suelo_sc2[, 1] * 0.9,
        mi_fam_hel_rda_suelo_sc2[, 2] * 0.9,
        length = 0,
        lty = 1,
        col = "red"
    )
#'
#' ##### Ejemplo usando las matriz ambiental con variables seleccionadas
#'
#' El RDA anterior mostró que las variables de suelo son útiles para predecir la matriz de comun
#'
#' - Hay mucha colinealidad entre ellas.
#'
#' - No se probó mejorar el modelo añadiendo otras variables, además de las de suelo.
#'
#' Crearé una matriz ambiental con las variables que resultaron significativas en el ajuste *pos
#'
env_selec <- bci_env_grid %>%
  select(
    heterogeneidad_ambiental,
    riqueza_global,
    UTM.EW,
    Al, B, Ca, Cu, Fe, K, Mg, Mn, P, Zn, N, N.min., pH) %>%
  st_drop_geometry
mi_fam_hel_rda_selec <- rda(mi_fam_hel ~ ., env_selec)
#'
vif.cca(mi_fam_hel_rda_selec)
#'
#' Haré el plot del escalamiento 2 para comprobar gráficamente asociación entre variables sin la
#'
plot(mi_fam_hel_rda_selec,
     scaling = 2,
     display = c("sp", "lc", "cn"),
     main = "Triplot de RDA especies ~ var. selec, escalamiento 2"
)
#'
#' Tal como comenté arriba, variables con valores VIF por encima de 10, deben ser examinadas. En
#'
env_selec2 <- bci_env_grid %>%
  select(
    heterogeneidad_ambiental,
    riqueza_global,
    UTM.EW,
    Al, B, Cu, Fe, Mg, Mn, P, Zn, N, N.min., pH) %>%
  st_drop_geometry
mi_fam_hel_rda_selec2 <- rda(mi_fam_hel ~ ., env_selec2)
vif.cca(mi_fam_hel_rda_selec2)
plot(mi_fam_hel_rda_selec2,

```

```

    scaling = 2,
    display = c("sp", "lc", "cn"),
    main = "Triplot de RDA especies ~ var. selec2, escalamiento 2"
)
#'
#' Nota que las posiciones rotaron, no así la asociación entre la mayoría de las variables. 'B'
#'
env_selec3 <- bci_env_grid %>%
  select(
    heterogeneidad_ambiental,
    riqueza_global,
    UTM.EW,
    Al, Cu, Fe, Mn, P, Zn, N, N.min., pH) %>%
  st_drop_geometry
mi_fam_hel_rda_selec3 <- rda(mi_fam_hel ~ ., env_selec3)
vif.cca(mi_fam_hel_rda_selec3)
plot(mi_fam_hel_rda_selec3,
     scaling = 2,
     display = c("sp", "lc", "cn"),
     main = "Triplot de RDA especies ~ var. selec3, escalamiento 2"
)
#'
#' Finalmente, la coordenada 'UTM.EW' tiene un alto valor VIF, por lo que es preferible excluirla
#'
env_selec4 <- bci_env_grid %>%
  select(
    heterogeneidad_ambiental,
    riqueza_global,
    Al, Cu, Fe, Mn, P, Zn, N, N.min., pH) %>%
  st_drop_geometry
mi_fam_hel_rda_selec4 <- rda(mi_fam_hel ~ ., env_selec4)
vif.cca(mi_fam_hel_rda_selec4)
plot(mi_fam_hel_rda_selec4,
     scaling = 2,
     display = c("sp", "lc", "cn"),
     main = "Triplot de RDA especies ~ var. selec4, escalamiento 2"
)
#'
#' Las variables 'N.min.' y 'Zn' ahora presentan valores VIF aceptables. Habrás notado que es to
#'
summary(mi_fam_hel_rda_selec4)
RsquareAdj(mi_fam_hel_rda_selec4)$adj.r.squared
#'
#' Triplot, pero ahora con las flechas para las especies.
#'
#' Escalamiento 1:
#'
plot(mi_fam_hel_rda_selec4,

```

```

        scaling = 1,
        display = c("sp", "lc", "cn"),
        main = "Triplot de RDA especies ~ var. selec4, escalamiento 1"
    )
mi_fam_hel_rda_selec4_sc1 <-
    scores(mi_fam_hel_rda_selec4,
           choices = 1:2,
           scaling = 1,
           display = "sp"
    )
arrows(0, 0,
       mi_fam_hel_rda_selec4_sc1[, 1] * 0.9,
       mi_fam_hel_rda_selec4_sc1[, 2] * 0.9,
       length = 0,
       lty = 1,
       col = "red"
    )
#'
#' Escalamiento 2
#'
plot(mi_fam_hel_rda_selec4,
     scaling = 2,
     display = c("sp", "lc", "cn"),
     main = "Triplot de RDA especies ~ var. selec4, escalamiento 2"
    )
mi_fam_hel_rda_selec4_sc2 <-
    scores(mi_fam_hel_rda_selec4,
           scaling = 2,
           choices = 1:2,
           display = "sp"
    )
arrows(0, 0,
       mi_fam_hel_rda_selec4_sc2[, 1] * 0.9,
       mi_fam_hel_rda_selec4_sc2[, 2] * 0.9,
       length = 0,
       lty = 1,
       col = "red"
    )
#'
#' ### Análisis de correspondencia canónica (CCA)
#'
#' ##### Ejemplo usando las matriz ambiental con variables seleccionadas:
#'
mi_fam_cca_selec4 <- cca(mi_fam ~ ., env_selec4)
summary(mi_fam_cca_selec4)
RsquareAdj(mi_fam_cca_selec4)
#'
#' Escalamiento 1

```

```

#'
plot(mi_fam_cca_selec4,
     scaling = 1,
     display = c("sp", "lc", "cn"),
     main = "Triplot de CCA especies ~ var. selec4, escalamiento 1"
)
#'
#' Escalamiento 2
#'
plot(mi_fam_cca_selec4,
     scaling = 2,
     display = c("sp", "lc", "cn"),
     main = "Triplot de CCA especies ~ var. selec4, escalamiento 2"
)
#'
#' Excluyendo especies con abundancia menor a 100 individuos
#'
colSums(mi_fam) %>% sort
mi_fam_no_raras <- mi_fam %>% select_if(colSums(.) > 100)
intersect(colnames(mi_fam), colnames(mi_fam_no_raras))
setdiff(colnames(mi_fam), colnames(mi_fam_no_raras))
mi_fam_no_raras_cca_selec4 <- cca(mi_fam_no_raras ~ ., env_selec4)
summary(mi_fam_no_raras_cca_selec4)
RsquareAdj(mi_fam_no_raras_cca_selec4)
#'
#' Escalamiento 1
#'
plot(mi_fam_no_raras_cca_selec4,
     scaling = 1,
     display = c("sp", "lc", "cn"),
     main = "Triplot de CCA especies no raras ~ var. selec4, escalamiento 1"
)
#'
#' Escalamiento 2
#'
plot(mi_fam_no_raras_cca_selec4,
     scaling = 2,
     display = c("sp", "lc", "cn"),
     main = "Triplot de CCA especies no raras ~ var. selec4, escalamiento 2"
)

```

#### 7.14 *Análisis de diversidad. Parte 1: Diversidad alpha*

```

#' ---
#' title: "Análisis de diversidad. <br> Parte 1: Diversidad alpha"
#' author: "JR"
#' date: "29 de noviembre, 2020"

```

```

#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo
#'
#' ### Cargar paquetes
#'
library(vegan)
library(adespatial)
library(plyr)
library(RColorBrewer)
library(tidyverse)
library(sf)
library(SpadeR)
library(iNEXT)
source('biodata/funciones.R')
#'
#' ### Cargar datos
#'
load('biodata/Malvaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
mi_fam <- mc_malvc
bci_env_grid %>% tibble
grupos_upgma_k2 <- readRDS('grupos_upgma_k2.RDS')
table(grupos_upgma_k2)
grupos_ward_k3 <- readRDS('grupos_ward_k3.RDS')
table(grupos_ward_k3)
#'
#' ## Diversidad alpha
#'
#' **La tentación de medir la diversidad mediante un único número, ha atrapado a muchos investigadores
#'
```

**La riqueza de especies**, una cifra de uso común en estudios ecológicos, es un elemento muy sensible a la escala de observación.

El término **"diversidad biológica"** se asume que fue acuñado a principios de los 80, y **se refiere a la diversidad de especies**.

**Biodiversidad** es de factura más reciente, y es una contracción de "biológica-diversidad".

La definición de biodiversidad del PNUMA es: **"variabilidad entre organismos vivos de todo tipo dentro de un sistema ecológico"**.

**Hubbell** (2001) ofrece una definición más adaptada a la práctica actual y mucho más restrictiva.

En estas dos últimas acepciones (Hubbell y Magurran), la diversidad biológica puede dividirse en tres componentes:

- La diversidad de especies como un único número**

```

#’ Usaré la notación *q* para designar el número de especies o riqueza de especies, y *n* para n
#’
#’ ##### Riqueza de especies y rarefacción
#’
#’ Magurran (2004) distingue entre **densidad de especies**, que equivale al número de especies
#’
#’ Para asegurar la comparabilidad entre sitios, se han propuesto distintos métodos. Uno es la r
#’
#’ 
#’
#’ ##### Componentes de la diversidad de especies basada en abundancia: riqueza y equidad
#’
#’ Asumamos que un sitio de muestreo es una variable cualitativa, y cada especie es un "estado".
#’
#’ 
#’
#’ Desde el punto de vista ecológico, la **H** tiene dos propiedades importantes:
#’
#’ - Crece al aumentar la riqueza de especies *q*.
#’ - Crece con la uniformidad (=equidad o equitabilidad, es decir, qué tan bien repartida se enc
#’
#’ Para una *q* dada, *H* asume su valor máximo (*Hmax*) cuando todas las especies es
#’
#’ 
#’
#’ Un ejemplo ilustra lo anterior:
#’
#’ - Cinco especies, 55 individuos, abundancias desiguales:
#’
foo1 <- c(25, 16, 9, 4, 1)
diversity(foo1)
#’
#’ - Cinco especies, 55 individuos, abundancias homogéneas
#’
foo2 <- c(11, 11, 11, 11, 11)
diversity(foo2)
log(5)
#’
#’ Otros índices miden exclusivamente la homogeneidad de la abundancia, a lo que se denomina equ
#’
#’ Uno de los más usados es la **equidad de Pielou** o *J* (1966), que es la razón entre la entr
#’
#’ 
#’
#’ Teóricamente, **este índice sólo debería aumentar cuando aumenta la homogeneidad de la abunda
#’
#’ Todos los **índices de equidad se relacionan con la forma de los modelos de abundancia de esp
#’

```



```

#’ Otra medida común en trabajos de ecología numérica, pero que es inversa o contraria a la equi
#’
#’ 
#’
#’ Este valor aumenta con la dominancia, por lo que realmente no mide diversidad, sino más bien
#’
#’ La riqueza de especies (*q*), la entropía de Shannon (*H*), la equidad de Pielou (*J*) y el i
#’
#’ 
#’
#’ donde *a* es el orden de la medida de entropía (*a=0,1,2...*), la cual cuantifica la importan
#’
#’ 
#’
#’ Las tres primeras entropías de Renyi (*H<sub>a</sub>*), donde *a=0, 1 y 2*, y los correspondi
#’
#’ Por otra parte, los tres primeros números de diversidad de Hill tienen significados que conoc
#’
#’  <br>
#’ > Según Borcard et al., 2018.
#’
#’ Bajo esta notación:
#’
#’ - La equidad de Pielou (denominada también equidad de Shannon) equivale a  $J = H_{1} / H_{0}$ 
#’
#’ Usando la misma aproximación, Hill propuso también otras ratios que, a diferencia de lo que o
#’
#’ -  $E_{1} = N_{1} / N_{0}$  a la cual el propio Hill denominó como su versi
#’
#’ -  $E_{2} = N_{2} / N_{0}$ , a la cual Hill denominó como su versión de la
#’
#’ Los números de diversidad y las ratios de Hill son menos sensibles a las matrices de comunida
#’
#’ **Estas afirmaciones tienen implicaciones muy importantes desde el punto de vista ecológico,
#’
#’ **Índices, entropías, equidades, ratios**
#’
(indices <- alpha_div(mi_fam))
pairs(indices,
      lower.panel = panel.smooth,
      upper.panel = panel.cor,
      diag.panel = panel.hist,
      main = "Pearson Correlation Matrix")
indices_env <- bind_cols(
  indices,
  bci_env_grid %>%
    select_if(is.numeric) %>%
    st_drop_geometry %>%

```

```

    select(-id) %>%
    select(-matches('^geom.*pct$'))
indices_env %>% tibble
ezCorM(indices_env, r_size_lims = c(3,5), label_size = 4)
#'
#' **Modelos de abundancia de especies**
#'
mi_fam_mae <- radfit(mi_fam)
plot(mi_fam_mae)
#'
#' **Rarefacción**
#'
#' Riqueza por sitios
#'
riqueza <- specnumber(mi_fam)
riqueza %>% sort
#'
#' Sitios con riqueza mínima y máxima
#'
riqueza[riqueza == min(riqueza)]
riqueza[riqueza == max(riqueza)]
range(riqueza)
#'
#' Abundancia por sitios
#'
abundancia <- rowSums(mi_fam)
abundancia %>% sort
#'
#' Sitios con abundancias mínima y máxima
#'
abundancia[abundancia == min(abundancia)]
abundancia[abundancia == max(abundancia)]
(rango_abun <- range(abundancia))
#'
#' Abundancia en el sitio más pobre
#'
abundancia[riqueza == min(riqueza)]
#'
#' Abundancia en el sitio más rico
#'
abundancia[riqueza == max(riqueza)]
#'
#' Riqueza en el sitio con menor abundancia
#'
riqueza[abundancia == min(abundancia)]
#'
#' Riqueza en el sitio con mayor abundancia
#'

```

```

riqueza[abundancia == max(abundancia)]
#'
#' Rarefacción a la abundancia más pequeña encontrada
#'
riqueza_menor_abun <- rarefy(mi_fam, sample = rango_abun[1])
sort(riqueza)
sort(round(riqueza_menor_abun))
rarecurve(
  mi_fam,
  step = 1,
  sample = rango_abun[1],
  xlab = "Número de individuos (tamaño de muestra)",
  ylab = "Especies",
  label = TRUE,
  col = "blue"
)
#'
#' ### Riqueza de especies, estimación y comparación, "completitud de muestra" (existe en el dic
#'
#' Aproximación básica:
#'
specpool(mi_fam)
specpool(mi_fam)[[1]]/specpool(mi_fam)*100
#'
#' Lista comprensiva de métodos (incluyendo recientes):
#'
#' - **Enfoques asintóticos. Estiman la riqueza de especies**:
#'   - Paramétricos:
#'     - Modelo homogéneo (estándar y MLE)
#'   - No paramétricos:
#'     - Chao1 y Chao1-bc
#'     - iChao1
#'     - Basados en "cobertura" o "completitud de muestra". ACE para datos de abundancia
#'     - Estimadores Jackknife (de primer y segundo órdenes)
#' - **Enfoques no asintóticos. Se utilizan para hacer rarefacción y extrapolación**:
#'   - Basados en tamaño de la muestra
#'   - Basados en "cobertura" o "completitud de muestra"
#'
#' #### Enfoques asintóticos y no asintóticos aplicados a la matriz de comunidad combinada
#'
#' Generar la matriz de comunidad combinada, en la que todos los sitios forman uno.
#'
mi_fam_combinada <- colSums(mi_fam)
mi_fam_combinada %>% sort
mi_fam_combinada_chao <- estimacion_riqueza_chao(
  mc = mi_fam_combinada,
  n_raras = 10)
mi_fam_combinada_chao$asintoticos_estimacion

```

```

mi_fam_combinada_chao$no_asintoticos_rarefaccion_extrapolacion
mi_fam_combinada_chao$no_asintoticos_rarefaccion_extrapolacion_grafico
#’
#’ ##### Enfoques asintóticos y no asintóticos aplicados a una matriz de comunidad agrupada
#’
#’ #’ Generar matriz de comunidad agrupada según el método de Ward (tres grupos), procedente de
#’
mi_fam_k3 <- mi_fam %>%
  mutate(g=grupos_ward_k3) %>%
  group_by(g) %>%
  summarise_all(sum) %>%
  select(-g) %>%
  data.frame
mi_fam_k3 %>% rowSums %>% sort
mi_fam_k3_chao <- estimacion_riqueza_chao(
  mc = mi_fam_k3,
  n_raras = 10)
mi_fam_k3_chao$asintoticos_estimacion
mi_fam_k3_chao$no_asintoticos_rarefaccion_extrapolacion
mi_fam_k3_chao$no_asintoticos_rarefaccion_extrapolacion_grafico

```

### 7.15 *Análisis de diversidad. Parte 2: Diversidad Beta*

```

#’ ---
#’ title: "Análisis de diversidad. <br> Parte 2: Diversidad beta"
#’ author: "JR"
#’ date: "2 de diciembre, 2020"
#’ output: github_document
#’ ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#’ ## Preámbulo
#’
#’ ### Cargar paquetes
#’
library(vegan)
library(adespatial)
library(plyr)
library(tidyverse)
library(sf)
library(vegetarian)
library(mapview)
source('biodata/funciones.R')
#’
#’ ### Cargar datos
#’

```

```

load('biodata/Malvaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
mi_fam <- mc_malvc
bci_env_grid %>% tibble
#'
#' ## Diversidad beta
#'
#' En la concepción de Whittaker, la diversidad alfa es la local o de sitio, la diversidad beta
#'
#' La diversidad alfa la medimos usando métricas que ponderan comúnmente la riqueza y la equidad
#'
#' ### Diversidad beta con un único número
#'
beta_multiplicativa <- calcular_beta_multiplicativa(
  mc = mi_fam,
  orden = 0:20)
beta_multiplicativa
#'
#' Notar que, para esta matriz de comunidad, la diversidad beta multiplicativa disminuye en esce
#'
#' ### Contribución de especies a la diversidad beta (SCBD, *species contribution to beta divers
#'
determinar_contrib_local_y_especie(
  mc = mi_fam,
  alpha = 0.05,
  nperm = 9999,
  metodo = 'hellinger')
#'
mapa_cuadros <- mapView(
  bci_env_grid,
  col.regions = 'grey80',
  alpha.regions = 0.3,
  map.types = 'OpenTopoMap',
  legend = F, zoom = 14,
  zcol = 'id') %>% addStaticLabels() %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
#'
#' Utilizar el mapa de cuadros para identificar aquellos que contribuyen más a la diversidad bet
#'
mapa_cuadros

```

## 7.16 Análisis espacial de datos ecológicos. Autocorrelación

```
#' ---
```

```

#' title: "Análisis espacial de datos ecológicos. <br> Autocorrelación"
#' author: "JR"
#' date: "5 de diciembre, 2020"
#' output: github_document
#' ---
#'
knitr::opts_chunk$set(fig.width=12, fig.height=8)
#'
#' ## Preámbulo
#'
#' ### Cargar paquetes
#'
library(ape)
library(spdep)
library(ade4)
library(adegraphics)
library(adespatial)
library(vegan)
library(tidyverse)
library(sf)
library(gridExtra)
library(grid)
library(gtable)
source('biodata/funciones.R')
source('https://raw.githubusercontent.com/maestria-geotel-master/unidad-3-asignacion-1-vecindad-
#'
#' ### Cargar datos
#'
load('biodata/Malvaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
mi_fam <- mc_malvc
mi_fam %>% tibble
bci_env_grid %>% tibble
#'
#' ## Preparar datos
#'
#' ### Generar matriz Hellinger
#'
mi_fam_hel <- decostand (mi_fam, "hellinger")
#'
#' ### Transformar matriz ambiental en objeto 'sp', generar vecindad
#'
bci_env_grid_sp <- bci_env_grid %>% as_Spatial
centroides <- bci_env_grid %>% st_centroid
bci_xy <- centroides %>% st_coordinates %>% as.data.frame
(vecindad <- bci_env_grid_sp %>% poly2nb)
(pesos_b <- nb2listw(vecindad, style = 'B'))
#+ fig.width=12, fig.height=6

```

```

plot(bci_env_grid_sp)
plot(vecindad, coords = bci_xy, add=T, col = 'red')
#'
#' ## Autocorrelación espacial mediante correlograma
#'
#' ### Una y solo una variable ambiental
#'
var_ph <- bci_env_grid %>% st_drop_geometry %>% pull(pH)
ph_correl <- sp.correlogram(vecindad,
                           var_ph,
                           order = 9,
                           method = "I",
                           zero.policy = TRUE)
print(ph_correl, digits = 2, p.adj.method = 'holm')
#+ fig.width=12, fig.height=6
plot(ph_correl)
#'
#' ### Múltiples variables
#'
#' ##### Abundancias de especies (matriz de comunidad transformada)
#'
suppressWarnings(auto_spp_hel <- calcular_autocorrelacion(
  df_fuente = mi_fam_hel,
  orden = 9,
  obj_vecindad = vecindad,
  pos_var = '(matriz Hellinger)'))
print(auto_spp_hel, digits = 2, p.adj.method = 'holm')
dim_panel <- rev(n2mfrow(ncol(mi_fam_hel)))
#+ fig.width=12, fig.height=10
par(mfrow = dim_panel)
suppressWarnings(invisible(lapply(auto_spp_hel, function(x) plot(x, main = x$var))))

jpeg('correlograma_matriz_de_comunidad_transformda.jpg', width = 1200, height = 800)
par(mfrow = dim_panel)
suppressWarnings(invisible(lapply(auto_spp_hel, function(x) plot(x, main = x$var))))
dev.off()

#'
#' ### Variables ambientales (matriz ambiental)
#'
bci_env_grid_num <- bci_env_grid %>%
  st_drop_geometry %>%
  select_if(is.numeric) %>%
  select(-id, -UTM.EW, -UTM.NS)
suppressWarnings(auto_amb <- calcular_autocorrelacion(
  df_fuente = bci_env_grid_num,
  orden = 9,

```

```

    obj_vecindad = vecindad))
print(auto_amb, digits = 2, p.adj.method = 'holm')
dim_panel <- rev(n2mfrow(ncol(bci_env_grid_num)))
#+ fig.width=12, fig.height=14
par(mfrow = dim_panel)
suppressWarnings(invisible(lapply(auto_amb, function(x) plot(x, main = x$var))))

jpeg('correlacion_matriz_comunidad_geomorofologicas_variables_suelo.jpg', width = 1200, height = 
par(mfrow = dim_panel)
suppressWarnings(invisible(lapply(auto_amb, function(x) plot(x, main = x$var))))
dev.off()

#'
#' ## Autocorrelación espacial mediante prueba Mantel (matrices de distancia)
#'
#' Para aplicar la prueba Mantel a datos de comunidad, es necesario quitar la tendencia espacial
#'
mi_fam_sin_tendencia <- resid(
  lm(as.matrix(mi_fam_hel) ~ .,
    data = bci_xy))
mi_fam_sin_tendencia_d <- dist(mi_fam_sin_tendencia)
(mi_fam_correlograma <- mantel.correlog(
  mi_fam_sin_tendencia_d,
  XY = bci_xy,
  nperm = 999))
#+ fig.width=12, fig.height=6
plot(mi_fam_correlograma)

jpeg('mi_fam_correlograma.jpg', width = 700, height = 300)
plot(mi_fam_correlograma)
dev.off()

#'
#' ## Autocorrelación espacial por medio de pruebas de permutación para el I de Moran
#'
#' ### I de Moran global aplicado a abundancia de especies transformadas sin tendencia
#'
(autocor_global_residuos <- sapply(
  dimnames(mi_fam_sin_tendencia)[[2]],
  function(x)
    moran.mc(
      x = mi_fam_sin_tendencia[,x],
      listw = pesos_b,
      zero.policy = T,
      nsim = 9999),
  simplify = F))
#'
```



```

#' ### I de Moran local
#'
#' ##### Aplicado a variables ambientales
#'
jpeg('cluster_LISA_variables_ambientales.jpg', width = 700, height = 300)
bci_env_grid_num_sf <- bci_env_grid %>%
  select_if(is.numeric) %>%
  select(-id, -UTM.EW, -UTM.NS)
bci_env_grid_num_sf %>% tibble
lisamaps_amb <- sapply(grep('geometry', names(bci_env_grid_num_sf), invert = T, value = T),
  function(x) {
    m <- lisamap(objesp = bci_env_grid_num_sf[x],
      var = x,
      pesos = pesos_b,
      titulo_leyenda = 'Significancia ("x-y", léase como "x" rodeado de "y")',
      leyenda = F,
      anchura_titulo = 50,
      tamaño_titulo = 10,
      fuente_datos = '\nhttp://ctfs.si.edu/webatlas/datasets/bci/',
      titulo_mapa = paste0('Clusters LISA de "', x, '"'))
    return(m$grafico)
  }, simplify = F
)
lisamaps_amb$leyenda <- gtable_filter(ggplot_gtable(ggplot_build(lisamaps_amb[[1]] + theme(legend.position = "top",
  grid.arrange(do.call('arrangeGrob', c(lisamaps_amb[1:12], nrow = 3)), lisamaps_amb$leyenda, height = 100),
  grid.arrange(do.call('arrangeGrob', c(lisamaps_amb[13:22], nrow = 3)), lisamaps_amb$leyenda, height = 100),
  grid.arrange(do.call('arrangeGrob', c(lisamaps_amb[23:31], nrow = 3)), lisamaps_amb$leyenda, height = 100),
  dev.off()

#'
#' ##### Aplicado a abundancias de especies transformadas
#'
jpeg('cluster_LISA_aplicado_especies_transformadas.jpg', width = 700, height = 300)
mi_fam_hel_sf <- bci_env_grid %>% select %>% bind_cols(mi_fam_hel)
lisamaps_mifam <- sapply(
  grep('geometry', names(mi_fam_hel_sf), invert = T, value = T),
  function(x) {
    m <- lisamap(objesp = mi_fam_hel_sf[x],
      var = x,
      pesos = pesos_b,
      titulo_leyenda = 'Significancia ("x-y", léase como "x" rodeado de "y")',
      leyenda = F,
      anchura_titulo = 50,
      tamaño_titulo = 10,
      fuente_datos = '\nhttp://ctfs.si.edu/webatlas/datasets/bci/',
      titulo_mapa = paste0('Clusters LISA de "', x, '"'))
    # dev.new(); print(m$grafico)
    return(m$grafico)
  }
)

```

```

    }, simplify = F
  )
  lisamaps_mifam$leyenda <- gtable_filter(ggplot_gtable(ggplot_build(lisamaps_mifam[[1]]) + theme(1
  grid.arrange(do.call('arrangeGrob', c(lisamaps_mifam[1:8], nrow = 3)), lisamaps_mifam$leyenda, h
  grid.arrange(do.call('arrangeGrob', c(lisamaps_mifam[9:16], nrow = 3)), lisamaps_mifam$leyenda,
  dev.off()

#'
#' ##### Aplicado a abundancias de especies transformadas sin tendencia
#'
jpeg('cluster_LISA_aplicado_especies_transformadas_sin_tendencia.jpg', width = 700, height = 300
mi_fam_sintendencia_sf <- bci_env_grid %>% select %>% bind_cols(mi_fam_sin_tendencia %>% as.data
lisamaps_mifam_sintendencia <- apply(
  grep('geometry', names(mi_fam_sintendencia_sf), invert = T, value = T),
  function(x) {
    m <- lisamap(objesp = mi_fam_sintendencia_sf[x],
      var = x,
      pesos = pesos_b,
      titulo_leyenda = 'Significancia ("x-y", léase como "x" rodeado de "y")',
      leyenda = F,
      anchura_titulo = 50,
      tamanotitulo = 10,
      fuentedatos = '\nhttp://ctfs.si.edu/webatlas/datasets/bci/',
      titulo_mapa = paste0('Clusters LISA de "', x, '"'))
    # dev.new(); print(m$grafico)
    return(m$grafico)
  }, simplify = F
)
lisamaps_mifam_sintendencia$leyenda <- gtable_filter(ggplot_gtable(ggplot_build(lisamaps_mifam_s
  grid.arrange(do.call('arrangeGrob', c(lisamaps_mifam_sintendencia[1:8], nrow = 3)), lisamaps_mif
  grid.arrange(do.call('arrangeGrob', c(lisamaps_mifam_sintendencia[9:16], nrow = 3)), lisamaps_mi
  dev.off()

```

Armuelles, L. G. (2021). Cambio climático podría reducir la biodiversidad de la isla barro colorado. Retrieved August 13, 2021, from <https://www.laestrella.com.pa/nacional/210813/cambio-climatico-reducir-biodiversidad-isla-barro-colorado>

Battle, J. R. M. (2020). biogeografia-master/scripts-de-analisis-BCI: Long coding sessions (Version v0.0.0.9000). <https://doi.org/10.5281/zenodo.4402362>

Clark, D. B. (2002). Los factores edáficos y la distribución de las plantas. *Ecología Y Conservación de Bosques Neotropicales*. LUR, Cartago, Costa Rica, 193–221.

Croat, T. B. (1978). *Flora of barro colorado island*. Stanford University Press.

Fisher, J. (2016). Monumento natural barro colorado. Retrieved November 8, 2016, from <https://es.slideshare.net/JoelyF15/monumento-natural-barro-colorado-68359938.pdf>

G.J. (2019). La acidificación de los suelos. Retrieved October 10, 2019, from <https://www.fertibox.net/single-post/acidificacion>

Ledesma, R. (2008). Introducción al bootstrap. desarrollo de un ejemplo acompañado de software de aplicación. *Tutorials in Quantitative Methods for Psychology*, 4(2), 51–60.

Moreno, R. (2012). Ámbito de hogar y actividad circadiana del ocelote (*leopardus pardalis*) en

la isla de barro colorado, panamá. *Mesoamericana*, 16(3), 30–39.

Rondón, J. B. (2009). La subfamilia malvoideae (malvaceae s.l.) en el occidente del estado sucre, venezuela. Retrieved April 20, 2009, from <http://www.bioline.org.br/pdf/cg09076>

Standley, P. C. (1927). The flora of barro colorado island, panama. *Smithsonian Miscellaneous Collections*.

Xu, Z., & Deng, M. (2017). Malvaceae. In *Identification and control of common weeds: Volume 2* (pp. 717–735). Springer.