

Asociación y composición florística de la familia Sapotaceae en la parcela permanente de 50h, Isla Barro Colorado

Merali Rosario *Estudiante, Universidad Autónoma de Santo Domingo (UASD)*

La diversidad y estructura de los bosques miden los recursos y la abundancia en un área geográfica, por ejemplo, los bosques de la familia Sapotaceae son importantes para proporcionar alimentos a las especies de vida silvestre. La Isla Barro Colorado es una reserva natural ubicada en el lago Gatún del Canal de Panamá. Debido a su capacidad de investigación, es una de las regiones tropicales más conocidas en materia de biología y ecología tropical. La distribución actual de la abundancia de especies en BCI se atribuye principalmente a los mecanismos que definen comunidades específicas, como la prevalencia de especies dominantes. Sin embargo, la distribución de la abundancia relativa se ve afectada por interacciones que aún no se han determinado plenamente, ni en qué grado inciden en la estructura de la comunidad. El objetivo de este trabajo es determinar la asociación, composición florística y distribución de la familia Sapotaceae en la parcela permanente de 50h de la isla Barro Colorado. Los datos de cada uno de los cuadrantes de una hectárea que componen BCI fueron procesados en R, teniendo en cuenta la matriz ambiental y la matriz de comunidad, los cuales contienen datos de las variables ambientales, tales como condiciones edáficas, tipo de hábitat, topografía del lugar, clasificación etaria del bosque, datos demográficos y geofrenciación espacial de todos los individuos censados. Se adaptaron scripts reproducibles, utilizando colecciones de paquetes multifuncionales. Todos los datos fueron examinados utilizando análisis de ecología numérica. Se registraron un total de 5 especies y 2 géneros distribuidos en 2029 individuos en toda la parcela. La riqueza por cuadro fue de 4 especies y la mediana de la abundancia por cuadro fue de 39 individuos. La especie más abundante fue *Pouteria reticulata*, con 1084 individuos, y la menos abundante fue *Pouteria fossicola* con 3 individuos. La riqueza de la familia Sapotaceae presenta correlación con la presencia de cobre y nitrógeno en el suelo. Las variables geomorfológicas presentan asociación con la abundancia y riqueza. La riqueza de la familia Sapotaceae aumenta en función del contenido de hidrógeno, nitrógeno y cobre, Además, aumenta con la equidad. Los índices de diversidad alfa arrojaron valores muy bajos. Las especies que contribuyen de manera significativa a la diversidad beta fueron: *Chrysophyllum argenteum*, *Chrysophyllum cainito* y *Pouteria stipitata*, de las cuales la que más contribuyó a la diversidad beta fue *Chrysophyllum cainito*. De acuerdo con estos resultados, se concluye que la riqueza de la familia Sapotaceae no presenta mucha diversidad en BCI, lo cual se estima que la riqueza seguiría constante o no aumentaría significativamente aunque se hiciera un mayor esfuerzo de muestreo.

Keywords: Diversidad, Sapotaceae, Riqueza, Abundancia, Asociación

1 Introducción

La diversidad y estructura de los bosques miden los recursos y la abundancia en un área geográfica, y cumplen funciones elementales para mantener la vida, por ejemplo, los bosques tropicales poseen una gran diversidad de especies y complejidad ecológica. Además, cubren un 10% de

la superficie terrestre y son de gran importancia para el planeta, debido a que capturan y procesan inmensas cantidades de carbono (Campos-Pineda, Moreno, & Mendieta, 2017; Martínez-Sovero et al., 2021).

La familia Sapotaceae está ampliamente distribuida en las zonas tropicales (Smedmark, 2007). Producen frutas tropicales y algunas especies producen látex, y madera de alta calidad, siendo una familia de plantas de importancia ecológica y económica (Martínez-Sovero, Iglesias-Osores, & Villena-Velásquez, 2020). Los bosques de la familia Sapotaceae son importantes para proporcionar alimentos a las especies de vida silvestre (Martínez-Sovero et al., 2021; Wan, 2020).

La Isla Barro Colorado es una reserva natural ubicada en el lago Gatún del Canal de Panamá. Debido a su capacidad de investigación, es una de las regiones tropicales más conocidas en materia de biología y ecología tropical ("Isla barro colorado y biología tropical," 1990). La isla exhibe características importantes, tres de las cuales son la estabilidad ambiental, su ubicación geográfica (en un área de importancia internacional) y la capacidad para investigar grupos específicos de organismos (Rodríguez-Flores & Barrios, 2020). Cabe señalar, que en trabajos anteriores (R. Condit et al., 2002) sobre los bosques tropicales de Panamá y el grado de diversidad beta entre especies en diferentes comunidades, indican que la disimilaridad aumenta con la distancia a la que están separadas en el espacio. Sin embargo, estos estudios no minimizan la importancia de la variabilidad del hábitat, y se toma en cuenta en este estudio, ya que un acercamiento inicial a los datos de abundancia de las diferentes especies de la familia Sapotaceae en Barro Colorado arrojaron indicaciones de posibles patrones de su distribución y se plantea la posibilidad de que hay especies con algún grado de asociación con las variables ambientales que allí prevalecen. Por otra parte, aunque la distribución actual de la abundancia de especies se atribuye principalmente a los mecanismos que definen comunidades específicas, como la prevalencia de especies dominantes, que son relativamente más abundantes que las especies raras. La distribución de la abundancia relativa se ve afectada por interacciones que aún no se han determinado plenamente, ni en qué grado inciden en la estructura de la comunidad (Horvát, Derzsi, Nédá, & Balog, 2010).

El objetivo de este trabajo es determinar la asociación, composición florística y distribución de la familia Sapotaceae en la parcela permanente de 50h de la isla Barro Colorado. Además, analizar la organización de las especies en los cuadros de 1 hectárea e identificar si existe algún patrón con alguna variable ambiental, así como también, explicar si hay especies indicadoras o con preferencia por determinadas condiciones ambientales. Por otra parte, evaluar si la familia Sapotaceae está suficientemente representada según los análisis de estimación de riqueza, determinar cuáles son las variables ambientales que presentan asociación con la diversidad alpha y mostrar cuáles son las especies que contribuyen a la diversidad beta. Por último, pero no menos importante, examinar la autocorrelación espacial de las especies.

2 Metodología

2.1 Área de Estudio

La isla de Barro Colorado es una colina de 1,500 hectáreas ubicada a 137 msnm en el lago Gatún. La parte superior de la isla es ancha y plana, y se asienta sobre un lecho de roca de basalto, de la cual irradian colinas empinadas y valles tallados en rocas sedimentarias que contienen gran cantidad de restos volcánicos. El suelo es arcilloso y la profundidad varía de 50 cm a un metro. El clima es típico de las áreas tropicales (Pérez et al., 2005; Windsor et al., 1990).

La parcela permanente de árboles de 50 hectáreas se estableció en 1980 en el bosque húmedo tropical. El sitio es un rectángulo de 1,000 m de largo por 500 m de ancho, ubicado en la meseta central de la isla. Está dividido en 1,250 cuadrantes de 20x20 m, en el cual se han contabilizado

todos los árboles con más de 10 mm de diámetro a la altura del pecho cada cinco años desde 1985 (R. A. y H. Condit Richard y Chisholm, 2012; R. y L. Condit Richard y Pérez, 2017; Pérez et al., 2005) (ver mapa 1).

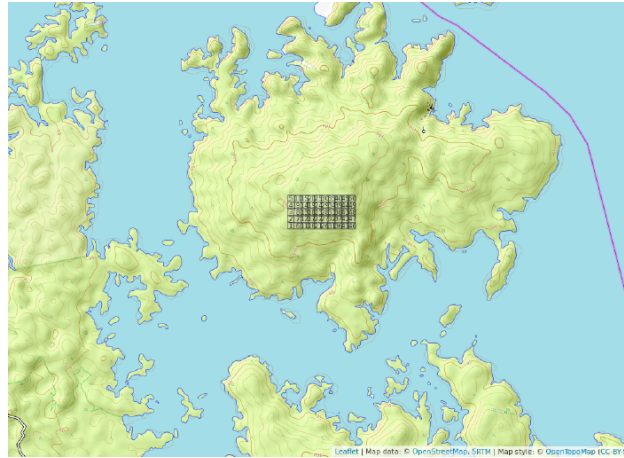


Figure 1: Área de estudio, parcela de 50ha, Isla Barro Colorado

2.2 Materiales y Métodos

La información de los cuadrantes que componen BCI fue procesada en R (R Core Team, 2020), teniendo en cuenta la matriz ambiental y la matriz de comunidad, los cuales contienen datos de las variables ambientales, tales como condiciones edáficas, tipo de hábitat, topografía del lugar, clasificación etaria del bosque, y datos demográficos y geofrenciación espacial de todos los individuos censados. Se adaptaron scripts reproducibles recuperados de Batlle (2020), utilizando la colección de paquetes multifuncionales: *vegan* (Oksanen et al., 2019), *Tidyverse* (Wickham, 2017), *BiodiversityR* (R. Kindt & Coe, 2005) y *indicspecies* (De Caceres & Legendre, 2009).

Para conocer las características de los datos almacenados de la matriz de comunidad y ambiental, se realizó un análisis exploratorio que incluyó visualización de gráficos, tablas, mapas de los cuadrantes de una hectárea y tablas de correlación lineal entre las dos variables de la matriz, lo que permitió una vista común y ayudó a determinar procedimientos más detallados a continuación.

2.3 Medición de asociación (*ma*)

Para realizar las pruebas de medición de asociación, se calculó la distancia euclidiana entre los cuadrados considerados objetos. Para ello, se requirió una transformación de la matriz de comunidad mediante el método Hellinger, que incluye elevar al cuadrado la abundancia relativa y_{ij} (cociente resultante de cada valor de abundancia entre la suma de los sitios) (Legendre & Gallagher, 2001). Además, se evaluó la distancia euclidiana entre los cuadrantes en términos de ocurrencia de especies. Se utilizó el índice de disimilitud de Jaccard de la matriz normalizada para convertir el valor de abundancia en un valor binario (Brocard, Gillet, & Legendre, 2018). del mismo modo, se empleó la métrica de Jaccard para aplicar la transposición de la matriz de la comunidad y convertir a datos Presencia / ausencia para medir el grado de asociación entre especies.

Para poder comparar la relación entre especies en función de su abundancia, se utilizó es-

tandarización *ji*-cuadrado de la matriz de comunidad transpuesta (Legendre & Gallagher, 2001). Se examinó la ocurrencia entre especies y su distribución en BCI por el coeficiente de correlación entre rangos de Spearman, para medir el grado de correlación entre las variables riqueza numérica de especies y la abundancia con las variables ambientales geomorfológicas, y la composición química del suelo (Brocard et al., 2018).

2.4 *Análisis de agrupamiento*

El método jerárquico aglomerativo de asociación entre pares de cuadrantes (según la composición de especies) bajo el estándar de enlace completo, y el método de Ward basado en la varianza mínima, se utilizó como método preliminar para el análisis de agrupamiento, con el fin de probar su efectividad en lograr un grupo consistente de importancia ecológica (Brocard et al., 2018). Luego, estos generaron dendrogramas que posteriormente son comparados con la matriz de distancia de cuerdas (Legendre & Gallagher, 2001). Usando correlación cofenética entre los dos para determinar el número ideal de grupos. Además, se utilizó remuestreo bootstrap y bootstrap multiescalar para conocer la probabilidad de éxito de la inferencia del número de grupos y la identidad de sus componentes (Brocard et al., 2018). Las distribuciones se basaban en una probabilidad de 91% o más de acierto para el método bootstrap y de un 95% para bootstrap multiescalar.

Para conocer las especies distintas o asociadas a cada grupo, se utilizó el valor del indicador o índice IndVal ("Conjuntos de especies y especies indicadoras," 1997), basado en permutaciones aleatorias de los sitios según la presencia de especies y la abundancia de estos. De manera similar, el grado de asociación de una especie con una preferencia particular por el grupo de cuadrantes considerado grupo en estudio, expresado como el coeficiente de correlación biserial puntual (Brocard et al., 2018). Se adoptó un enfoque similar al anterior a lo largo de las pruebas estadísticas de la hipótesis nula, basada en las especies presentes en los cuadrantes pertenecientes a un determinado grupo realizado al azar. Esta prueba se hizo reordenando aleatoriamente los valores de abundancia y comparando sus distribuciones con las obtenidas previamente (Cáceres & Legendre, 2009).

2.5 *Análisis de diversidad alpha*

La diversidad alpha representa la diversidad de especies a lo largo de todas las subunidades locales relevantes, y por definición abarca dos variables importantes: (1) la riqueza de especies, y (2) la abundancia relativa de especies (Carmona-Galindo & Carmona, 2013). Con el fin de determinar la diversidad alpha se utilizaron métodos como la Entropía de Shannon H_1 , que calcula el grado de desorden en la muestra, el índice de concentración de Simpson, que calcula la probabilidad de que dos individuos seleccionados aleatoriamente puedan ser de la misma especie. Además, se empleó la serie de números de diversidad de Hill, la fórmula de la entropía de Rengi y el índice de equidad de Pielou.

2.6 *Análisis de diversidad beta*

La diversidad beta, de acuerdo con (Whittaker, 1960), se define como el diferencial entre la diversidad de un hábitat y la diversidad total de un paisaje de hábitats. Teniendo en cuenta lo anterior, se utilizó el método hellinger para determinar cuáles son las especies que contribuyen a la diversidad beta, es decir, las especies que no se encuentran compartidas entre los cuadrantes.

2.7 Análisis de ordenación simple (no restringida) y canónica (restringida)

Se aplicaron los análisis de componentes principales o PCA a los datos de las variables ambientales para determinar la ordenación no restringida, donde se tomó en cuenta específicamente las variables del suelo. y para determinar la ordenación restringida se exploró de manera explícita las relaciones entre una matriz de respuesta y una matriz explicativa con los análisis de redundancia o RDA, lo cual combina la regresión y el análisis de componentes principales (PCA), por ejemplo, busca tendencias en la matriz de comunidad restringiéndolas a la matriz ambiental.

2.8 Ecología espacial

En ecología espacial se utilizó la matriz transformada de hellinger y la matriz ambiental para crear un cuadro de vecindad y ver como se autocorrelacionan los sitios. Se genera un correlograma para la variable que queremos estudiar mediante la función 'sp.correlogram' y para varias variables como la abundancia de especies y las variables ambientales. También, se utilizaron otros métodos como la prueba Mantel con matrices de distancia para autocorrelación espacial con y sin tendencia, y el I de Moran con una matriz de abundancia de especies transformada sin tendencia, lo cual se aplica a variables ambientales para obtener los datos de autocorrelación, distribución de especies y variables en los sitios de muestreo.

3 Resultados

Se registraron un total de 5 especies y 2 generos distribuidos en 2029 individuos en toda la parcela. La riqueza por cuadro fue de 4 especies y la mediana de la abundancia por cuadro fue de 39 individuos. La especie más abundante fue *Pouteria reticulata*, con 1084 individuos, y la menos abundante fue *Pouteria fossicola* con 3 individuos. La tabla 1 y la figura 2 resume estos resultados. La distribución de la riqueza numérica de especies de la familia Sapotaceae sigue un patrón homogéneo, lo cual los agregados de riqueza máxima están distribuidos en casi todo el área (ver Figura 3). Cabe resaltar, que la distribución de las variables ambientales PH y pendiente media siguen un patrón relativamente similar a la distribución de la riqueza numérica de la familia Sapotaceae.

Table 1: Abundancia por especie de la familia Sapotaceae

Latin	n
<i>Pouteria reticulata</i>	1084
<i>Chrysophyllum argenteum</i>	711
<i>Chrysophyllum cainito</i>	171
<i>Pouteria stipitata</i>	60
<i>Pouteria fossicola</i>	3

		quad1ha																																																				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50			
especie	Pouteria stipitata	0	0	1	4	1	2	1	3	1	2	1	1	1	2	0	1	5	2	5	2	0	0	0	0	0	1	0	1	1	1	4	0	1	2	0	1	0	0	0	0	0	0	0	0	1	0	1	1	5	4	0	1	0
	Pouteria reticulata	12	23	29	18	16	19	31	17	7	25	24	22	20	20	39	18	15	36	14	42	23	20	23	12	29	20	24	10	14	16	20	24	35	23	16	32	20	21	21	34	22	16	11	28	13	32	10	34	15	19			
	Pouteria fossicola	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
	Chrysophyllum cainito	2	1	1	2	0	0	3	0	0	1	1	3	5	3	3	3	1	5	3	8	4	4	12	3	4	0	3	5	3	7	2	4	6	12	7	0	4	1	7	8	1	3	4	10	2	3	2	1	2	2			
	Chrysophyllum argenteum	21	11	19	38	21	18	9	14	25	17	18	11	8	22	19	13	5	9	6	18	19	5	10	9	12	6	8	10	5	8	13	12	5	13	13	18	16	16	17	3	18	29	19	23	9	11	15	25	11	11			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50			
		quad1ha																																																				

Figure 2: Abundancia por especie por quadrat

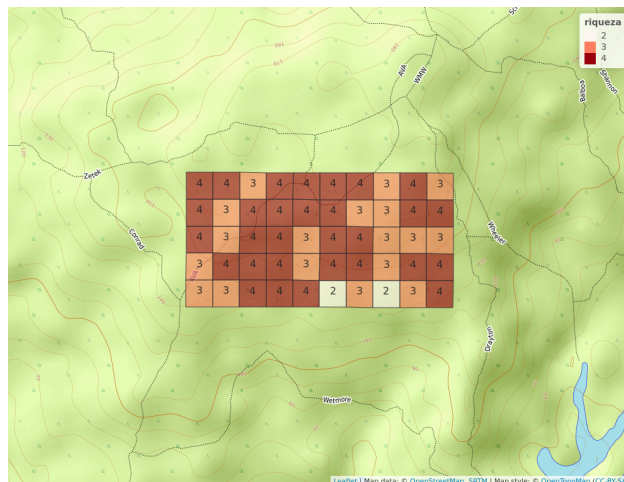


Figure 3: Distribucion de la riqueza de la familia Sapotaceae

Los valores para el coeficiente de Spearman presentados en el panel de correlación de la figura 4, mostraron que la abundancia de la familia sapotaceae solo presenta correlación con la abundancia global, mientras que la riqueza tiene correlación con la presencia de cobre y nitrógeno en el suelo, lo que sugiere que mientras mas concentración de cobre y nitrógeno hay, mayor será la riqueza de especies.



Figure 4: correlacion de las variables del suelo

El índice de similaridad de Jaccard muestra que el sitio 1 y 2 comparten un 100% de sus especies, por lo que ambos sitios comparten 3 especies y no tienen especies exclusivas (ver figura 5). Las variables geomorfológicas presentan asociación con la abundancia y riqueza, lo cual la figura 6 muestra que hay mucha similaridad entre estas variables. Las pruebas de correlación entre los grupos 1 y 2 formulados por upgma resultaron significativas respecto a la variable fósforo. Por otro lado, el contenido de cobre y la abundancia global promedio, es decir, la media correspondiente a todas las plantas en BCI, son significativamente diferentes entre los sitios de ambos grupos, para un nivel de significancia de $\alpha = 0.1$ (ver figura 7).

El grupo 1 generado por enlace upgm está conformado por 5 cuadrante y el grupo 2 por 46 cuadrantes (ver figura 8). El grupo 2 contiene los sitios con tendencia a presentar valores altos de zinc y contenido de cobre. Es probable que las especies indicadoras del grupo con un mayor contenido de cobre estén mostrando preferencia por estas condiciones ambientales. No obstante, la mayoría de componentes del suelo en BCI tienen valores bastante homogéneos, y más bien se presentan pequeños gradientes entre los cuadrantes, lo cual evita que este tipo de acercamiento sea concluyente. La especie *Chrysophyllum argenteum* fue la que obtuvo un valor alto de confianza al examinar su potencial como especies indicadoras del grupo 1. Para el caso del grupo 2, la especie indicadora fue *Pouteria reticulata*.

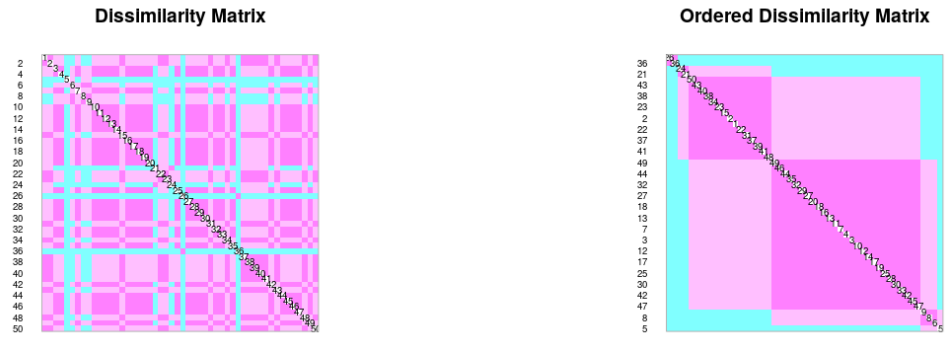


Figure 5: Similaridad de Jaccard (color fucsia (magenta, rosa) significa “corta distancia=muy similares”, y cian (celeste) significa “gran distancia=poco similares”)

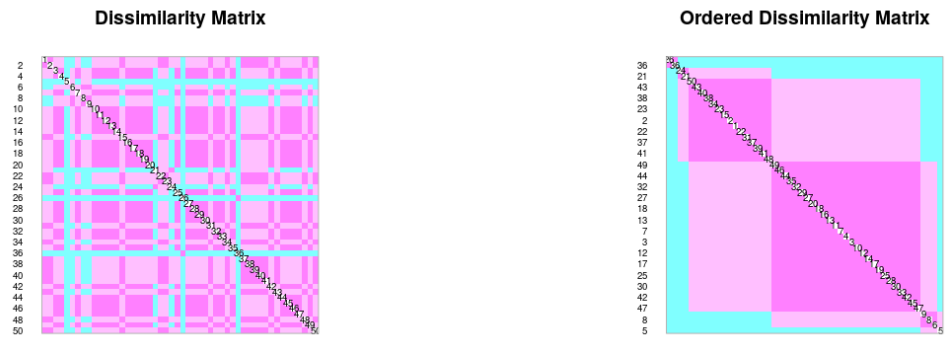


Figure 6: Panel de correlacion de Spearman entre los datos de la comunidad y las variables geomorfológicas (color fucsia (magenta, rosa) significa “corta distancia=muy similares”, y cian (celeste) significa “gran distancia=poco similares”)

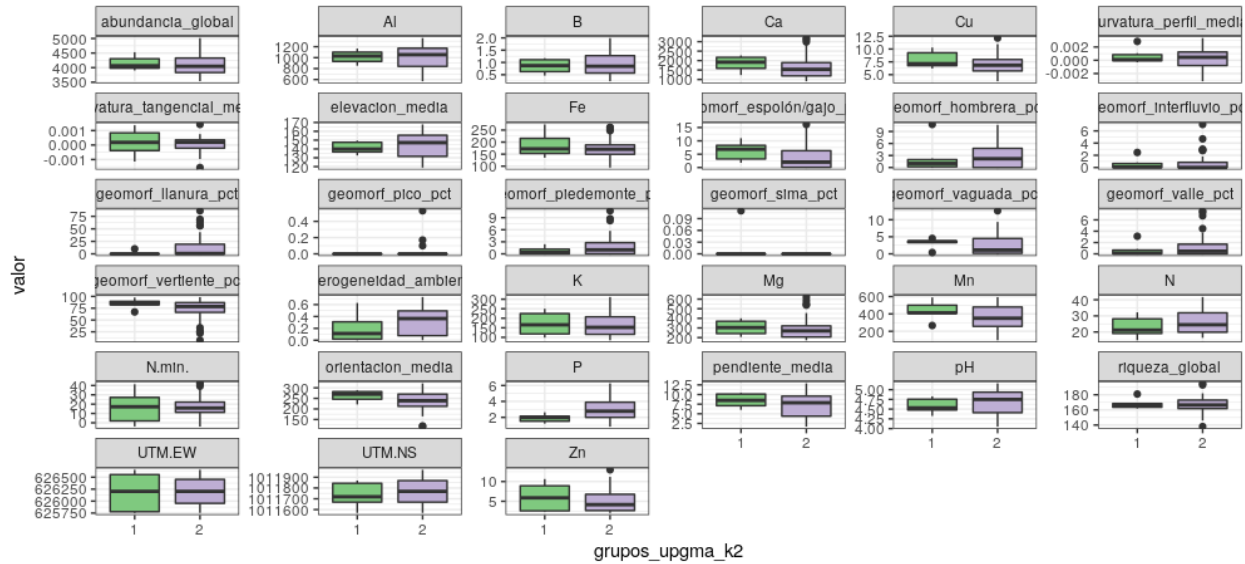


Figure 7: Diagramas de caja de las variables que tuvieron un efecto, segun las pruebas de igualdad de medias

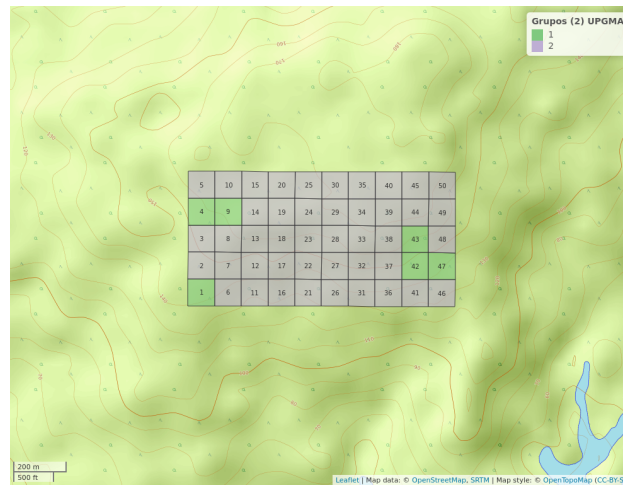


Figure 8: Mapa en el que se presenta la distribución de sitios en los grupos formulados por enlace upgma

La riqueza de la familia Sapotaceae aumenta en función del contenido de hidrógeno, nitrógeno y cobre, además, aumenta con la equidad. Cabe destacar, que algunos sitios de BCI tienen valores altos de equidad (ver figura 9). La Curva de rarefacción de los sitios, muestra como va aumentando la cantidad de individuos y de especies de los cuadrantes (ver figura 10), donde la mayor concentración de individuos está entre los 20 a 50 individuos y la abundancia máxima es de 70 individuos. Los valores de los índices de diversidad alfa fueron: Riqueza de especies (0.22), Shannon (0.05) y Simpson (0.04).

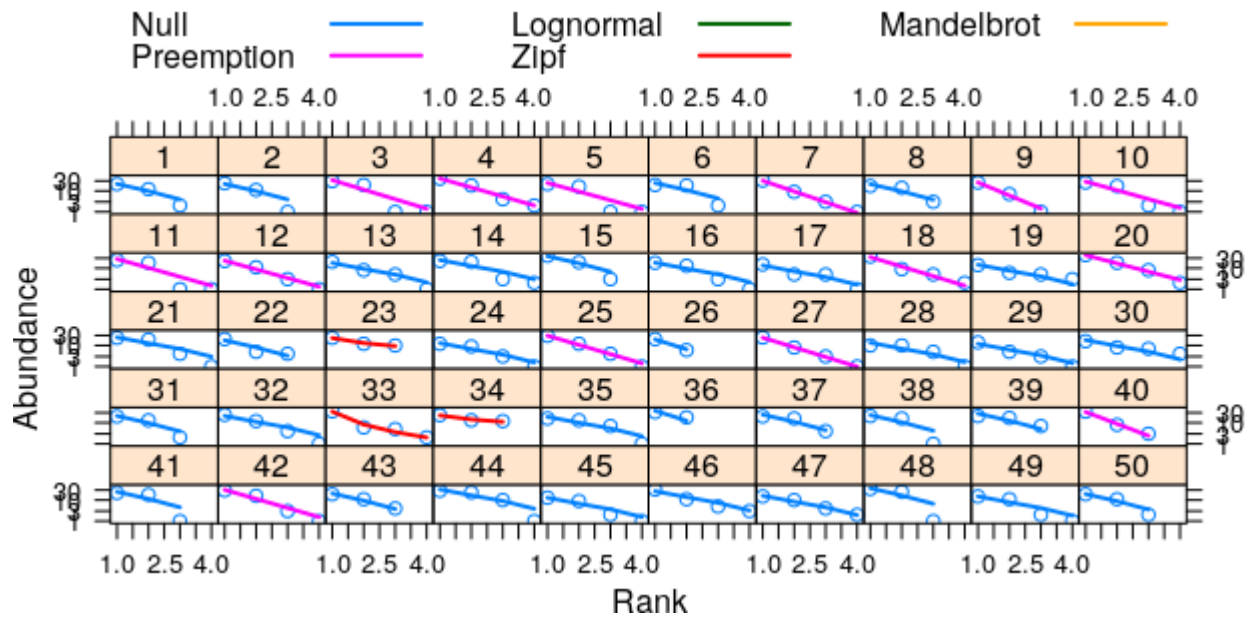


Figure 9: Grafico dque presenta los valores de equidad por sitios

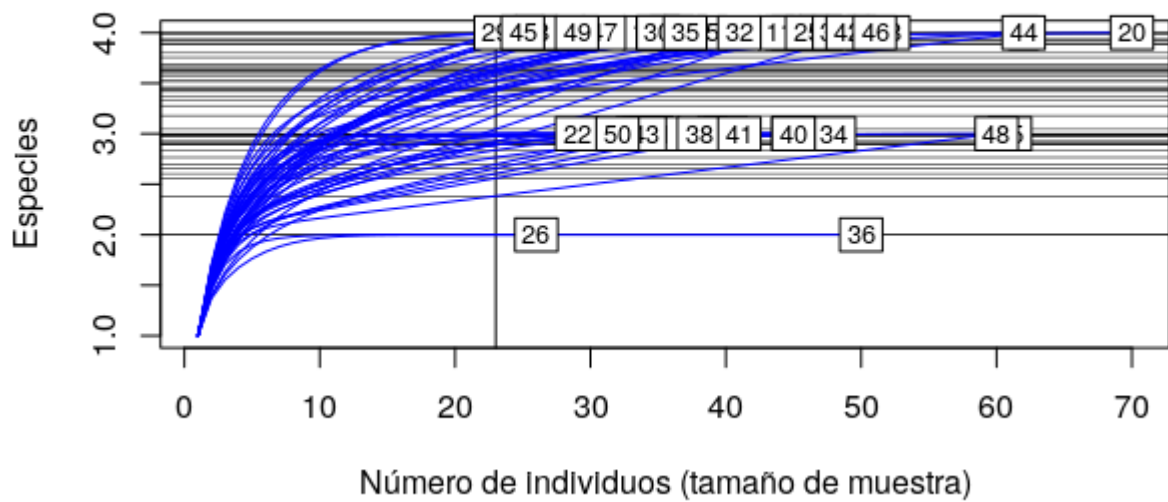


Figure 10: Curva de rarefaccion de los sitios

Las especies que contribuyen de manera significativa a la diversidad beta fueron: *Chrysophyllum argenteum* (0.25), *Chrysophyllum cainito* (0.31) y *Pouteria stipitata* (0.27), de las cuales la que mas contribuyó fué *Chrysophyllum cainito*. No obstante, los sitios que contribuyen a la diversidad

beta son los cuadrantes 9 y 40, lo cual presentan contribución debido a la incidencia de algunas variables ambientales.

El gráfico 11 incluye el comportamiento de los componentes principales de la varianza en las variables suelo y geomorfología en BCI, predecido por el modelo de barra quebrada, representado por la línea roja formando la curva (La escala denominada “Inertia” representa la suma de los cuadrados de toda la varianza). En el diagrama rotulado como escalamiento 1 de la figura 12, se observan tres grupos de cuadrantes diferenciados entre sí. Un grupo de sitios con un alto grado de acidez y contenido en aluminio, otro grupo caracterizado por la presencia de elementos metálicos, y un tercero, con una cantidad de fósforo, nitrógeno y valor de pH mayor. Las variables nitrógeno, fósforo y pH aportan la mayor parte de la varianza explicada. La relación entre las variables se encuentra debidamente representada en el recuadro del escalamiento 2, por medio de los ángulos que forman sus vectores (ver figura 12). Los resultados del PCA de los datos de la matriz de comunidad se encuentran resumidos en los diagramas de la figura 13. El escalamiento 1, muestra muchos de los cuadrantes dispuestos alrededor del origen formado por los ejes, lo que indica una contribución a la varianza relativamente equitativa por parte de las especies.

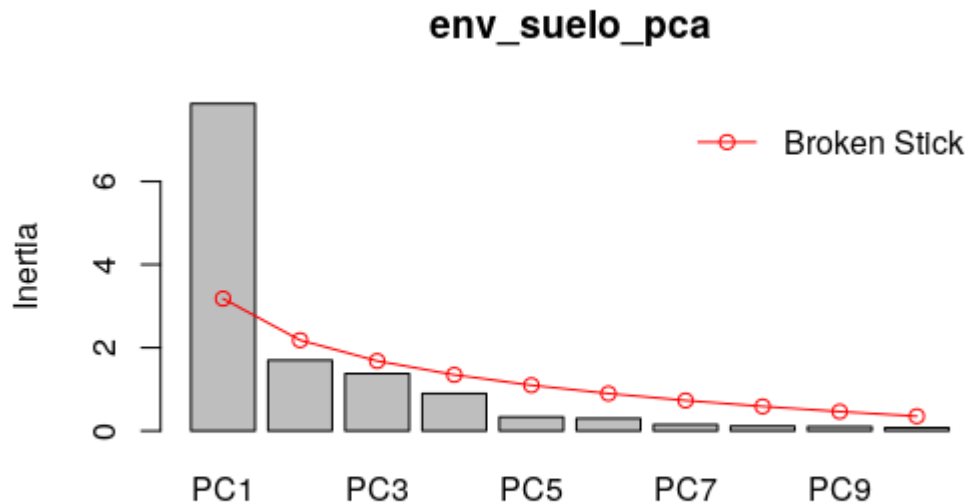


Figure 11: grafico de los componentes principales de la varianza en las variables suelo y geomorfología en BCI

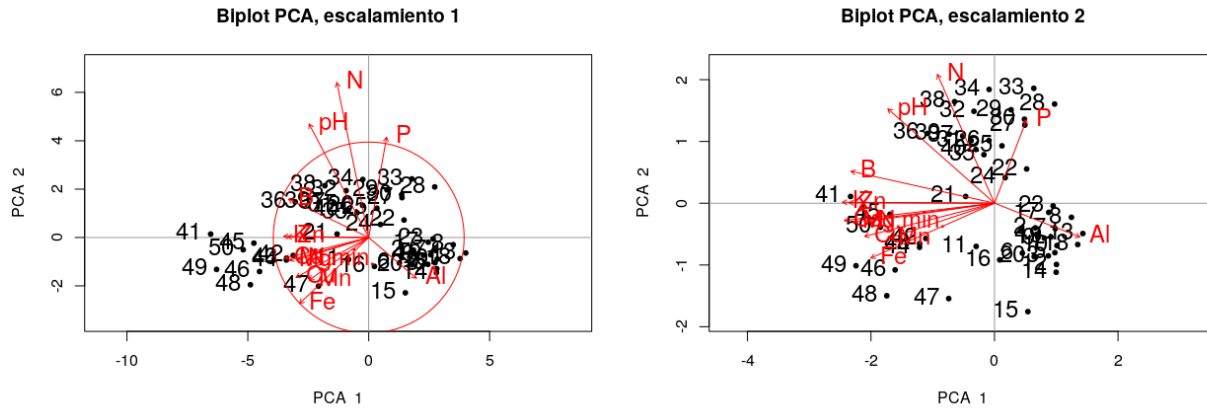


Figure 12: Biplots generados en el PCA de las viariables de suelo

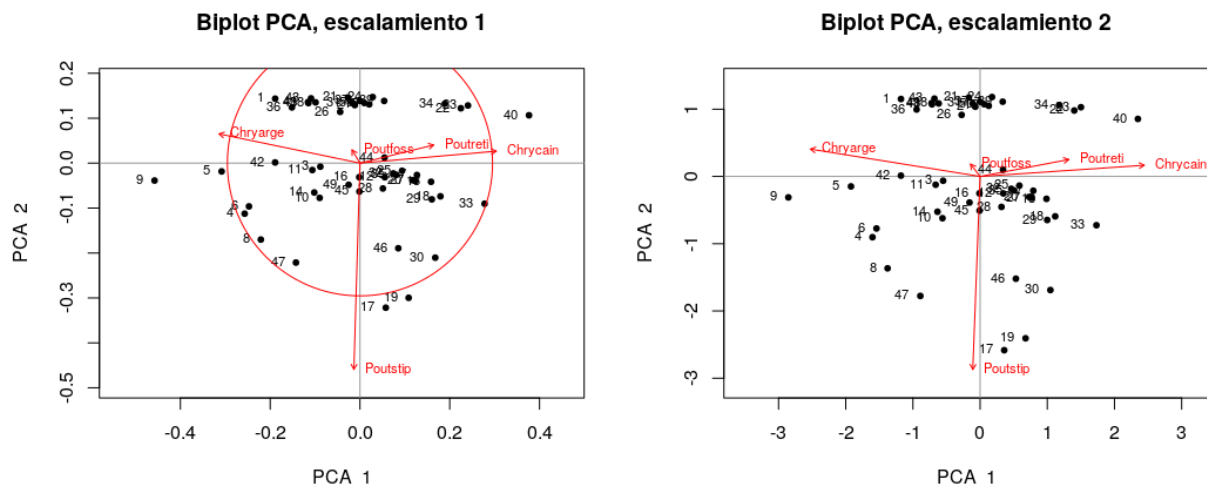


Figure 13: Biplots generados en el PCA de las viariables de suelo

El escalamiento 2 de la figura 14 en el análisis de correspondencia mostró que las especies *Pouteria reticulata*, *Chrysophyllum argenteum* y *Chrysophyllum cainito* se encuentran asociadas. Las especies restantes tienen una abundancia reducida, y en consecuencia, aparecen cercanas a los pocos cuadrantes en los que se encuentran representadas. La disparidad en la incidencia de las especies se refleja en su disposición en el diagrama. Sin embargo, estos resultados no coinciden del todo con los arrojados por el PCA de la matriz de distancias.

Los primero ordenes de *Chrysophyllum argenteum* y *Chrysophyllum cainito* presentan valores de autocorrelación alta o positiva, mientras que los ordenes de las demas especies presentan mayormente valores de autocorrelación negativa (Ver figura 15).

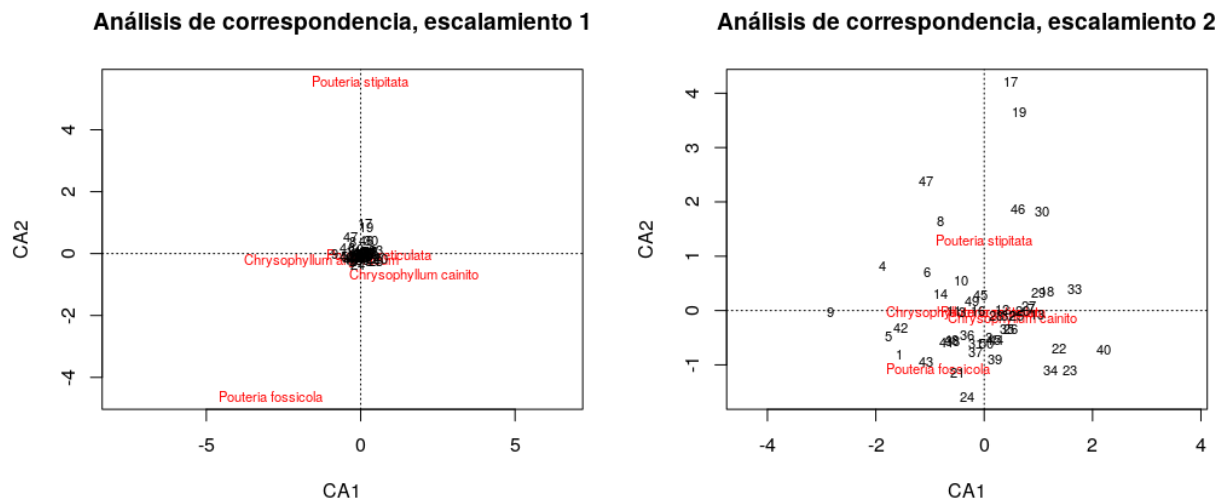


Figure 14: Biplot del analisis de correspondencia de los datos de abundancia de las especies de Sapotaceae

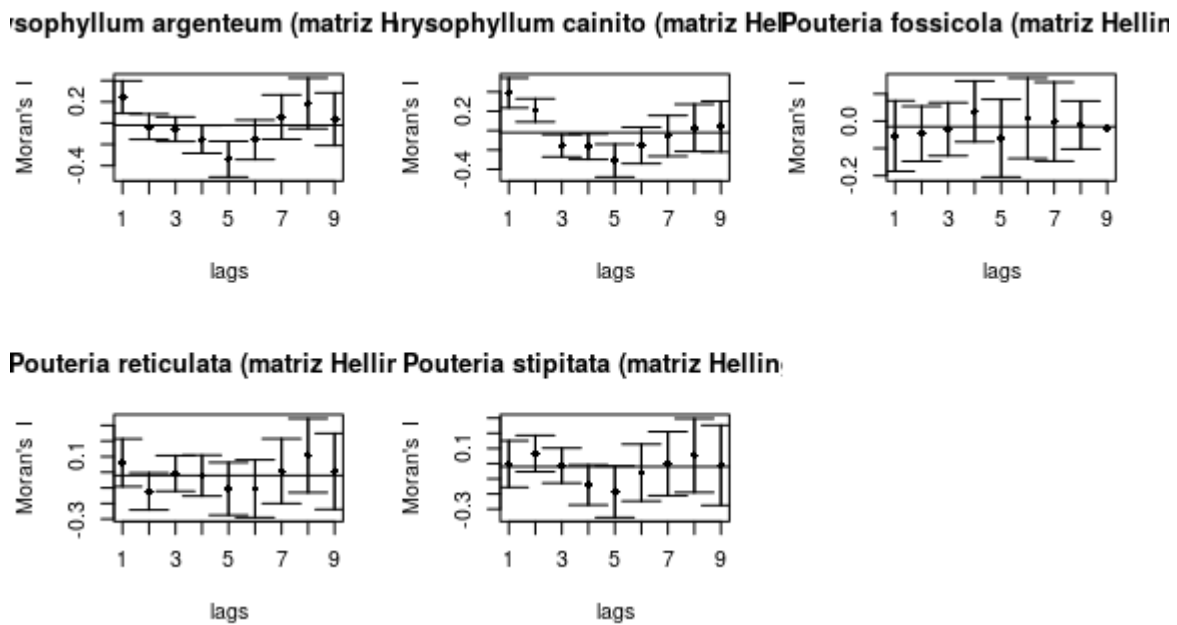


Figure 15: Autocorrelación espacial de las especies

4 Discusión

Estudios de la familia Sapotaceae también reportan que *Pouteria* es un género que parece siempre presentar una cantidad significativa de individuos (Martínez-Sovero et al., 2021). La riqueza

de la familia Sapotaceae aumenta en función del contenido de hidrógeno, nitrógeno y cobre, los cuales son algunos de los nutrientes que más se correlacionan con la diversidad de especies de plantas en el neotrópico (Doblado Amador, 2011). Además, la riqueza aumenta con la equidad, debido a que las especies están distribuidas en casi todos los cuadrantes.

Los valores de los índices de diversidad alfa fueron muy bajos, por consiguiente, la familia Sapotaceae no presenta mucha diversidad en BCI. Se estima que la riqueza seguiría constante o no aumentaría significativamente aunque se hiciera mayor esfuerzo de muestreo (ver figura 16).

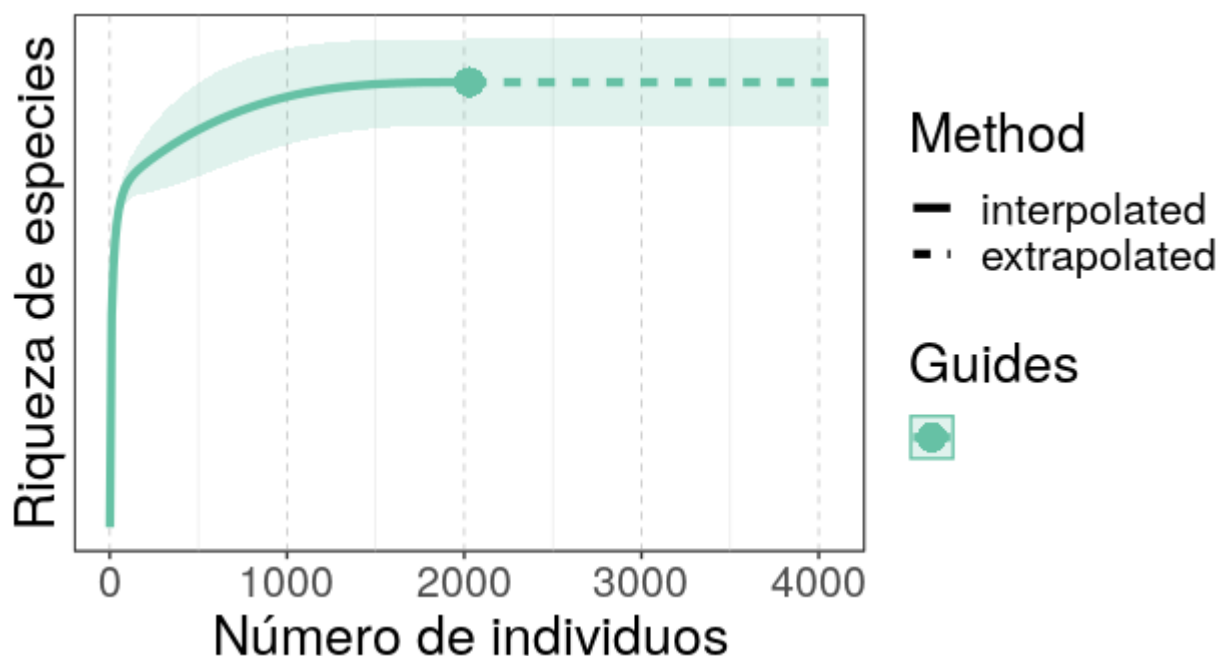


Figure 16: Grafico de acumulacion de especies en funcion de numeros de individuos

Las especies que contribuyen de manera significativa a la diversidad beta fueron: *Chrysophyllum argenteum*, *Chrysophyllum cainito* y *Pouteria stipitata*, las cuales fueron las que obtuvieron valores intermedios de individuos. No obstante, los sitios que contribuyen a la diversidad beta son los cuadrantes 9 y 40, ya presentan contribución por la incidencia de algunas variables ambientales, tales como PH, aluminio, boro, manganeso, etc. Lo anterior coincide con (Horvát et al., 2010), lo cual explica que la distribución de la abundancia relativa se ve afectada por interacciones, que en este caso se debe a la incidencia de algunas variables ambientales. Además, (R. Condit et al., 2002) señala que la diversidad beta aumenta dependiendo la distancia que están separadas las especies o los sitios donde se encuentran estas.

Las especies *Pouteria reticulata*, *Chrysophyllum argenteum* y *Chrysophyllum cainito* se encuentran asociadas, debido a que tienen los valores mas altos de abundancia dentro de la comunidad (1084, 711 y 171 individuos, respectivamente). Finalmente, los primeros ordenes de *Chrysophyllum argenteum* y *Chrysophyllum cainito* presentan valores de autocorrelación positiva, mientras que los ordenes de las demás especies presentan mayormente valores de autocorrelación negativa, teniendo en cuenta que el orden 5 de todas las especies presenta autocorrelacion negativa, lo que sugiere que el orden 5 de las especies está autocorrelacionado espacialmente negativo.

5 Agradecimientos

Agradezco al maestro José Ramón Martínez, por su motivación y ayuda en todos los aspectos para que este trabajo salga bien.

6 Información de soporte

...

7 Script reproducible

7.1 *Análisis de 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/Sapotaceae.Rdata')
load('biodata/matriz_ambiental.Rdata') #Matriz ambiental, se carga como "bci_env_grid"

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

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

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

#' ### Número de sitios, tanto en matriz de comunidad como en ambiental
#' Verifica que coinciden
nrow(mc_saptc) #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_saptc)
sort(specnumber(mc_saptc)) # Ordenados ascendentemente
summary(specnumber(mc_saptc)) # Resumen estadístico

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

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

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

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

#' ### Una tabla para el manuscrito, es necesario asignarle nombre
#' Para esto, usaré la colección "tidyverse"
abun_sp <- censo_saptc %>%
  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_q <- crear_grafico_mosaico_de_mc(mc_saptc, tam_rotulo = 12)
abun_sp_q

```

7.2 *Análisis con la colección tidyverse*

```

#' ---
#' title: "Análisis exploratorio de datos. Colección tidyverse"
#' author: "JR"
#' date: "18 de octubre, 2020"
#' output: github_document
#' ---

#' # ¿Qué es tidyverse?
#'

```



```

#’ Es una colección de paquetes con los que podrás importar, transformar, visualizar, modelar y
#’
#’ Todos estos paquetes comparten estructuras comunes. Una de las herramientas que incorpora la
#’
#’ ‘objeto1 %>% funcion1()’ es equivalente a ‘funcion1(argumento1 = objeto1)’
#’
#’ > La idea del *pipe* pertenece a la tradición de sistemas tipo Unix y, en origen, su función
#’
#’ Su ventaja radica en que, si necesitaras continuar procesando los datos, no tendrás que anidar
#’
#’ ‘resultado <- objeto1 %>% funcion1() %>% funcion2() %>% funcion3()’
#’
#’ Puedes leer lo anterior como *"objeto1 pasa como primer argumento de funcion1, **luego** el r
#’
#’ Para replicar esta operación sin la pipa, podrías realizarlo de, por ejemplo, dos maneras dis
#’
#’ * Opción 1, anidar:
#’
#’ ‘resultado <- funcion3(funcion2(funcion1(objeto1)))’
#’
#’ Opción 2, crear objetos intermedios:
#’
#’ ‘tmp1 <- funcion1(objeto1)’
#’ ‘tmp2 <- funcion2(tmp1)’
#’ ‘resultado <- funcion3(tmp2)’
#’
#’ Notarás que, comparada con estas dos últimas opciones, la tubería es más limpia que estas dos
#’
#’ ¡Comencemos!
#’
#’ ## Paquetes
#’
library(tidyverse)
library(sf)
#’
#’ > ‘sf’ te ayudará a leer el objeto ‘bci_env_grid’ como un *simple feature*, el cual se encuen
#’
#’ ## Cargar datos
#’
load('biodata/matriz_ambiental.Rdata')
load('biodata/Sapotaceae.Rdata')
#’
#’ ## Paquete ‘dplyr’
#’
#’ Te servirá para manipular datos mediante verbos. Los verbos de ‘dplyr’ que conocerás son (hay
#’
#’ ### Verbo ‘select’
#’

```

```

#’ Comúnmente, necesitas seleccionar una o varias columnas de una tabla. Para esto existe el verbo
bci_env_grid %>%
  select(id, pH)
#’ > Importante: el objeto ‘bci_env_grid’ permanece intacto, a menos que se use dicho nombre para crear otro
#’ Fíjate en la clase del objeto ‘bci_env_grid’. Para ello usaré la función de R ‘class’. No sólo
bci_env_grid %>%
  class
#’ El objeto ‘bci_env_grid’ es a la vez de clase ‘sf’ (*simple feature*) y ‘data.frame’, es decir, es un objeto de clase ‘sf’
bci_env_grid %>%
  select(id, pH) %>%
  st_drop_geometry
#’ Fíjate ahora en la clase de ‘bci_env_grid %>% select(id, pH) %>% st_drop_geometry’, que en este caso es ‘sf’
bci_env_grid %>%
  select(id, pH) %>%
  st_drop_geometry %>%
  class
#’ > Al introducir un ‘<enter>’ después de la pipa, el código puede continuar en la línea siguiente
#’
#’ Seleccionaré, y a la vez renombraré, dos columnas con ‘select’ (recuerda: no estoy modificando los datos)
bci_env_grid %>%
  select(id_de_quadrat = id, pH_del_suelo = pH) %>%
  st_drop_geometry %>%
  head
#’
#’ Otra funcionalidad de ‘select’ es poder seleccionar columnas según patrones. Por ejemplo, si quiero seleccionar
#’ columnas que contengan ‘geomorf’
bci_env_grid %>%
  select(contains(‘geomorf’)) %>%
  st_drop_geometry
#’ ...y también usando expresiones regulares con ‘matches’, usando por ejemplo dos cadenas de caracteres
bci_env_grid %>%
  select(matches(‘geomorf|habit’, ignore.case = F)) %>%
  st_drop_geometry
#’ ...o pidiendo todas las columnas que comienzan por mayúsculas, excepto las que comienzan por ‘geomorf’
bci_env_grid %>%
  select(matches(‘^[A-T,Z]’, ignore.case = F)) %>%
  st_drop_geometry
#’ ### Verbo ‘filter’
#’
#’ Ahora mostraré sólo los elementos con ‘pH’ mayor que 5, usando el verbo ‘filter’
bci_env_grid %>%
  select(id, pH) %>%
  st_drop_geometry %>%
  filter(pH>5)
#’ O filtro por aquellos con ‘id’ 31 y 50:
bci_env_grid %>%
  select(id, pH) %>%
  st_drop_geometry %>%

```

```

    filter(id == c(31, 50))
#'
#' ### Verbo 'arrange'
#'
#' Pruebo también con la matriz de comunidad. Por ejemplo, introduzco en la tubería la función '
mc_saptc %>%
  colSums
#' Y también obtengo la abundancia por quadrat.
mc_saptc %>%
  rowSums
#' Uso a continuación el verbo 'arrange' para mostrar los registros de la matriz ambiental orden
bci_env_grid %>%
  select(id, pH) %>%
  st_drop_geometry %>%
  arrange(pH)
#' Ahora usaré 'arrange' para mostrar los registros de la matriz ambiental ordenados DESCendente
bci_env_grid %>%
  select(id, pH) %>%
  st_drop_geometry %>%
  arrange(desc(pH))
#'
#' ### Verbo 'mutate'
#'
#' Usaré el verbo 'mutate' para crear una nueva columna. Por ejemplo, creo una columna que conte
bci_env_grid %>%
  st_drop_geometry %>%
  select(habitat, quebrada) %>%
  mutate(habitat_quebrada = paste(habitat, quebrada, sep = ', '))
#' Ahora 'mutate', pero con números: creo una columna de área de cada cuadro (necesitas también
bci_env_grid %>%
  mutate(area = st_area(geometry)) %>%
  select(id, area) %>%
  st_drop_geometry %>%
  head
#' ...y ahora más complejo: obtengo la densidad de individuos por metro cuadrado, ordenados desc
bci_env_grid %>%
  mutate(area = st_area(geometry), densidad_indiv = abundancia_global/area) %>%
  select(id, densidad_indiv) %>%
  st_drop_geometry %>%
  arrange(desc(densidad_indiv)) %>%
  head
#'
#' ### Verbos 'group_by' y 'summarise'
#'
#' Los verbos 'group_by' y 'summarise' son útiles para producir resúmenes por grupos.
#'
#' Agruparé la matriz ambiental por la columna 'habitat', dejando sólo las variables numericas q
agrupado_por_habitat <- bci_env_grid %>%

```

```

    st_drop_geometry %>%
    group_by(habitat) %>%
    select_if(is.numeric) %>%
    select(-id, -UTM.EW, -UTM.NS)
agrupado_por_habitat
#' Observa el encabezado: el objeto es 'A tibble: 50 x 32' y hay 5 grupos ('Groups:   habitat [5
agrupado_por_habitat %>% summarise(n = n())
#' ...y también algunos estadísticos de las columnas 'pH', 'abundancia_global' y 'riqueza_global'
agrupado_por_habitat %>%
  summarise(
    n = n(),
    media_pH = mean(pH),
    media_abundancia = mean(abundancia_global),
    media_riqueza = mean(riqueza_global)
  )
#' ...o la media de todas las variables numéricas
agrupado_por_habitat %>%
  summarise_all(mean)
#' ...no caben, mejor por partes
agrupado_por_habitat %>%
  summarise_all(mean) %>%
  select(1:6) %>%
  print(width=300)
agrupado_por_habitat %>%
  summarise_all(mean) %>%
  select(1,7:12) %>%
  print(width=300)
agrupado_por_habitat %>%
  summarise_all(mean) %>%
  select(1, 13:25) %>%
  print(width=300)
agrupado_por_habitat %>%
  summarise_all(mean) %>%
  select(1, 26:32) %>%
  print(width=300)
#' ...y no sólo un estadístico, sino varios:
agrupado_por_habitat %>%
  summarise_all(
    list(
      media = mean,
      mediana = median,
      varianza = var,
      minimo = min,
      maximo = max
    )
  )
#' Ejecuto también un ANOVA de una vía, de la 'riqueza_global' respecto de 'habitat' de tipo '01
agrupado_por_habitat %>%

```

```

    filter(str_detect(habitat, 'Old*')) %>%
    oneway.test(formula = riqueza_global ~ habitat)
#' El resultado sugiere que "existen 'diferencias significativas' de 'riqueza_global' entre 'hab
#'
#' Finalmente, te muestro 'join'. Más que una función, 'join' es una función genérica con varios
#'
#' Obtendré una tabla con dos columnas: código identificador de quadrat de 1 ha (le llamaré 'id'
id_abundancia_fam <- mc_saptc %>%
  mutate(abundancia_mi_familia = rowSums(.)) %>%
  rownames_to_column(var = 'id') %>%
  mutate(id = as.numeric(id)) %>% #Numérico, garantiza compatibilidad con id de bci_env_grid
  select(id, abundancia_mi_familia)
id_abundancia_fam %>% tibble
#' Dado que 'id_abundancia_fam' y 'bci_env_grid' comparten el campo 'id', a través de éste se pu
bci_env_grid %>%
  inner_join(y = id_abundancia_fam, by = 'id')
#' El resultado muestra la 'bci_env_grid', ahora con los datos de mi familia como parte de la ma
#'
#' ## 'tidyr'
#'
#' Te ayudará a reordenar datos, mediante transformación de su estructura, para organizarlos de
#'
#' ### Verbo 'pivot_longer'
#'
#' Cuando necesitas reunir varias columnas, o lo que es lo mismo, hacerlas que pivoten a lo larg
#'
#' 
#' *Tomado de: UC Business Analytics R Programming Guide. Reshaping Your Data with tidyr. https://uc-r.github.io/public/images/dataWrangling/gather1.png
#'
#' Es común realizar "reunión" columnas cuando nos interesa aplicar análisis masivos a múltiples
#'
#' Pongo un ejemplo. Por tipo de hábitat, ¿cuánto es el promedio de los porcentajes de cada uno
pivotpaso1 <- bci_env_grid %>%
  st_drop_geometry %>%
  select(matches('geomorf|habitat'))
pivotpaso1 %>% tibble
#' ...luego reunir todas las columnas de geomorfología pivotando en torno a la columna 'habitat'
pivotpaso2 <- pivotpaso1 %>%
  pivot_longer(
    cols = contains('geomorf'),
    names_to = 'variable',
    values_to = 'valor')
pivotpaso2 %>% tibble
#' ...y finalmente obtener las medias de porcentajes de geomorfología por cada grupo de habitat,
pivotpaso3 <- pivotpaso2 %>%
  group_by(habitat, variable) %>%
  summarise(media = mean(valor))
pivotpaso3 %>% arrange(habitat, desc(media)) %>% print(n=Inf)

```

```

# 'pivot_longer' también es útil para realizar paneles de gráficos de muchas variables, como ve
#
# La operación contraria a 'pivot_longer' se realiza con 'pivot_wider'. Supongamos que ahora ne
#
pivotpaso3 %>%
  ungroup() %>%
  pivot_wider(
    id_cols = habitat,
    names_from = variable,
    values_from = media)
#
# ## 'ggplot2'
#
# Te ayudará en la visualización de tus datos, utilizando gramática de gráficos.
#
# Un gráfico 'ggplot' utiliza capas para mostrar la información. Los objetos fuente son 'data.f
#
# Explicaré su uso con ejemplos, descomponiendo las partes de una sentencia 'ggplot' para fines
#
# Primero incluiré la función 'ggplot', para crear un espacio de coordenadas según los datos di
p0 <- ggplot(bci_env_grid)
p0
# A continuación, definiré las variables estéticas sobre las que construiré la simbología, añad
p1 <- p0 + aes(x = abundancia_global, y = riqueza_global)
p1
# El espacio de coordenadas ya está creado, y 'ggplot2' está preparado para aceptar geometrías.
p2 <- p1 + geom_point()
p2
# Dado que en 'p1' definí las coordenadas de mapeo 'aes(x = abundancia_global, y = riqueza_glob
p3 <- p2 + geom_smooth(formula = y ~ x, method = 'lm')
p3
# En 'p3', tanto 'geom_point' como 'geom_smooth' aprovechan las coordenadas del mapeo definido
#
# Una forma alterna permite definir la capa estética dentro de la geometría con resultado idént
p4 <- p0 +
  geom_point(mapping = aes(x = abundancia_global, y = riqueza_global))
p4
# Esta forma tiene la ventaja de ser más corta, pero tiene la desventaja de que impide reutiliz
#
# También definiré propiedades globales del gráfico mediante temas.
p5 <- p3 + theme_bw()
p5
p6 <- p3 + theme_classic()
p6
p7 <- p3 + theme_minimal()
p7
# Con una variable categórica, se pueden estilizar los elementos del gráfico. Por ejemplo, haré
p8 <- p0 +

```

```

geom_point(
  mapping = aes(
    x = abundancia_global,
    y = riqueza_global,
    color = habitat))
p8
#’
#’ Ahora mostraré cómo construir el último gráfico con una sentencia de conjunto, sin reaprovechar
p9 <- ggplot(bci_env_grid) +
  geom_point(
    mapping = aes(
      x = abundancia_global,
      y = riqueza_global,
      color = habitat))
p9
#’ Las posibilidades de personalización de gráficos de ‘ggplot2’ son enormes y superan el cometido
p10 <- p0 +
  geom_boxplot(mapping = aes(x = habitat, y = abundancia_global))
p10
#’ Y ejemplifico también ‘riqueza_global’:
p11 <- p0 +
  geom_boxplot(mapping = aes(x = habitat, y = riqueza_global))
p11
#’ ...la cual muestra efectos más marcados que ‘abundancia_global’.
#’
#’ Los dos gráficos anteriores son muy informativos, pero tienen la desventaja de que para poder
#’
#’ Necesitamos tres columnas, una con los nombres de los hábitats, otra con los nombres de las variables
habitat_riqueza_abundancia <- bci_env_grid %>% st_drop_geometry %>%
  select(habitat, abundancia_global, riqueza_global) %>%
  pivot_longer(
    cols = c(abundancia_global, riqueza_global),
    names_to = 'variable',
    values_to = 'valor')
habitat_riqueza_abundancia
#’ Construiré el gráfico definiendo a ‘habitat’ en el eje ‘x’, y valor en ‘y’, mientras que usaré
p13<-habitat_riqueza_abundancia %>%
  ggplot() +
  aes(x = habitat, y = valor) +
  geom_boxplot() +
  facet_wrap( ~ variable, scal = 'free_y')
p13
#’ En resumen, usa ‘tidyverse’ para sacar el máximo provecho de tus datos. El paquete ‘dplyr’ te

```

7.3 Mapas de riqueza y abundancia global y de mi familia

```

#' ---
#' 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/Sapotaceae.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 = 0.3,
  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_saptc)),
  layer.name = 'abundancia',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T, zoom = 14,
  col.regions = azul,
  zcol = 'abun') %>%
addStaticLabels(label = rowSums(mc_saptc)) %>%
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')

#' ### Mapa de cuadros, simbología por riqueza de mi familia
mapa_cuadros_riq_mi_familia <- mapView(
  bci_env_grid %>% mutate(riq = specnumber(mc_saptc)),
  layer.name = 'riqueza',
  alpha.regions = 0.6,

```

```

map.types = 'OpenTopoMap',
legend = T, zoom = 14,
col.regions = rojo,
zcol = 'riq') %>%
addStaticLabels(label = specnumber(mc_saptc), textsize = "12pt") %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 16)
mapa_cuadros_riq_mi_familia
mapa_cuadros_riq_mi_familia %>% mapshot(file = 'mapa_cuadros_riq_mi_familia.png')

```

7.4 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 = 16,
  col.regions = rojo,
  zcol = 'pendiente_media') %>%
addStaticLabels(label = round(bci_env_grid$pendiente_media, 1), textsize = '10pt') %>%
leaflet::setView(

```

```

    lng = -79.85136,
    lat = 9.15097,
    zoom = 16)
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 = 16,
  col.regions = rojo_inv,
  zcol = 'pH') %>%
  addStaticLabels(label = round(bci_env_grid$pH, 1), textsize = '10pt') %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 16)
mapa_cuadros_ph
mapa_cuadros_ph %>% mapshot(file = 'mapa_cuadros_ph.png')

```

7.5 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/Sapotaceae.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_saptc),
    abundancia_mifam = rowSums(mc_saptc)) %>%
  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_saptec %>%
  rename_all(gsub, pattern = '_| ', replacement = '\\n') %>%
  ezCor(r_size_lims = c(4,8), label_size = 3)
p_cor_mc

```

7.6 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
            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()

```

7.7 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 sitio
#'
#' 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 calcu
#’
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
#'

```

7.8 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 sitio
#'
#' Esta paradoja se explica por la presencia de "ceros" (especies ausentes) en la matriz de comun
#'
#' 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 calcu
#'
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 fácilme
#'
(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
#’

```

7.9 Modo Q aplicado a mi familia

```

#’ ---
#’ 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/Sapotaceae.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_saptec, "hellinger", silent = T)
mi_fam_d_hel %>% tidy # Para evitar desbordar la consola
#'
#' Para interpretar esta matriz, es necesario representarla gráficamente. En la representación
#'
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 conjunto de índices de Jaccard y de Sorensen "a secas", sin especificar el tipo de métrica.
#' Un error común consiste en referirse a los índices de Jaccard y de Sorensen "a secas", sin especificar el tipo de métrica.
#' Si alguna vez te enfrentas a textos donde no se especifica qué tipo de métrica se usa, te sugiero que uses los índices de Jaccard y de Sorensen "a secas", sin especificar el tipo de métrica.
#' Recalco: es imprescindible declarar qué tipo de métrica estás usando. Ejemplos de redacción:
#' - Correcto: "índice de disimilitud de Jaccard", "índice de similitud de Sorensen".
#' - Incorrecto: "índice de Jaccard", "índice de Sorensen".
#'
#' A continuación, muestro cómo calcular la distancia de Jaccard ( $D_J$ ) en un conjunto de datos binarios.
#'
mi_fam_jac <- vegdist(mc_saptc, 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_saptc, method = 'jac', binary = T)'.
#' En esta matriz de disimilitud, al igual que en la anterior, un valor pequeño (rosa) significa alta similitud.
#'
coldiss(mi_fam_jac, diag = T)
#'
#' La distancia de Jaccard ( $D_J$ ) se puede expresar como "la proporción de especies que no comparten entre los dos sitios".
#' Como la distancia de Jaccard ( $D_J$ ) es el complemento a 1 de la similitud de Jaccard ( $S_J$ ), podemos calcular la similitud de Jaccard como:
#'
(1 - mi_fam_jac) %>% tidy %>% rename(similitud=distance) #Similitud
#'
#' Dado que este resultado muestra la similitud, podemos leerlo como "el sitio 1 y el 2 comparten X especies".
#' La fórmula de la similitud de Jaccard es  $S_J = a / (a + b + c)$ , donde a es el número de especies compartidas, b es el número de especies que sólo están en el sitio 1, y c es el número de especies que sólo están en el sitio 2.
#' Para obtener las variables a, b y c, usaré la función 'betadiver' del paquete 'vegan'.
#'
mi_fam_abc <- betadiver(mc_saptc)
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 (c), y que el sitio 1 tiene 19 especies (b).
#'
round(3/3*100,2) #Porcentaje de especies compartidas = similitud
#'

```

```

#’ Con ‘betadiver’ también puedes calcular índices de similaridad. Por ejemplo, el Jaccard se ca
#’
betadiver(mc_saptc, 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_saptc, 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 *0
#’
#’ - ‘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)
#’

```

7.10 Análisis de agrupamiento jerárquico

```

#’ ---
#’ title: "Análisis de agrupamiento (cluster analysis). <br> Parte 1: agrupamiento jerárquico"

```



```

#' 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(magrittr)
library(broom)
source('biodata/funciones.R')

#' ### Cargar datos
#'
load('biodata/-Sapotaceae.Rdata')
mi_fam <- mc_saptc
#'
#' ## Características de las técnicas de agrupamiento
#'
#' Las técnicas de agrupamiento se clasifican según los algoritmos que emplean y el orden de eje
#'
#' - Secuenciales o simultáneos.
#' - Por aglomeración o por división. En referencias en español encontrarás "aglomerativos" y "d
#' - Monotéticos o politéticos.
#' - Jerárquicos o no jerárquicos.
#' - Probabilísticos o no probabilísticos.
#' - Restringidos o no restringidos.
#'
#' ## Agrupamiento jerárquico
#'
#' El agrupamiento jerárquico (AJ) es una técnica de agrupamiento secuencial que consiste en la
#'
#' Dentro del AJ es frecuente usar un enfoque aglomerativo, lo cual implica aplicar algoritmos s
#'
#' El AJ aglomerativo dispone de varios algoritmos de resolución del agrupamiento por pares, que
#'
#' Normalmente, en el análisis de agrupamiento nos interesa agrupar sitios en función de sus des
#'
#' Dado que los cuadros en BCI están autocorrelacionados espacialmente, violamos el supuesto de
#'
#' ### Agrupamiento "aglomerativo" por enlace simple
#'
#' Este método utiliza, como criterio de enlace para agrupar sucesivamente pares de objetos, la
#'
#' Para aplicar este método, debes transformar la matriz de comunidad utilizando alguno de los m

```

```

#’
mi_fam_norm <- decostand(mi_fam, "normalize")
mi_fam_norm_d <- vegdist(mi_fam_norm, "euc")
mi_fam_norm_d %>% tidy
#’
#’ Es importante, para garantizar consistencia a lo largo del agrupamiento, asignar los nombres
#’
attr(mi_fam_norm_d, "labels") <- rownames(mi_fam)
#’
#’ Posteriormente, el agrupamiento jerárquico lo realizaré con la función ‘hclust’ del paquete ‘
#’
(cl_single <- hclust(mi_fam_norm_d, method = 'single'))
#’
#’ Finalmente, el dendrograma a continuación:
plot(cl_single, labels = rownames(mi_fam), hang = -1,
     main = "Sitios de BCI según composición de especies de Sapotaceae\nEnlace simple a partir d
     xlab = 'Sitios', ylab = 'Altura')
#’
#’ ### Agrupamiento "aglomerativo" por enlace completo
#’
#’ En este caso, el criterio de enlace para agrupar sucesivamente pares de objetos es la menor s
#’
(cl_complete <- hclust(mi_fam_norm_d, method = 'complete'))
plot(cl_complete, labels = rownames(mi_fam), hang = -1,
     main = "Sitios de BCI según composición de especies de Sapotaceae\nEnlace completo a partir
     xlab = 'Sitios', ylab = 'Altura')
#’
#’ ### Agrupamiento "aglomerativo" por enlace promedio
#’
#’ En este caso, el criterio de enlace para agrupar sucesivamente pares de objetos es el promedi
#’
#’ Así, dependiendo de si se media o centroide, o si se ponderan o no las distancias originales,
#’
#’ Sólo crearé el dendrograma del método UPGMA.
#’
(cl_upgma <- hclust(mi_fam_norm_d, method = 'average'))
plot(cl_upgma, labels = rownames(mi_fam), hang = -1,
     main = "Sitios de BCI según composición de especies de Sapotaceae\nUPGMA a partir de matriz
     xlab = 'Sitios', ylab = 'Altura')
#’
#’ ### Agrupamiento por el método de Ward de varianza mínima
#’
#’ Se basa en los mismos supuestos y criterios de la regresión lineal por mínimos cuadrados, sim
#’
(cl_ward <- hclust(mi_fam_norm_d, method = 'ward.D2'))
plot(cl_ward, labels = rownames(mi_fam), hang = -1,
     main = "Sitios de BCI según composición de especies de Sapotaceae\nMétodo de Ward a partir
     xlab = 'Sitios', ylab = 'Altura')

```

7.11 Análisis de agrupamiento (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/Sapotaceae.Rdata')
mi_fam <- mc_saptc
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 requier
#’
#’ 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_upgma <- calcular_anchuras_siluetas(
  mc_orig = mi_fam,
  distancias = mi_fam_norm_d,
  cluster = lista_cl$cl_upgma)
anch_sil_upgma
#’
#’ El objeto ‘anchuras_siluetas’ de la lista ‘anch_sil_upgma’ te muestra un vector con los prome
#’
#’ Igualmente, el objeto ‘n_grupos_optimo’ te indica cuál es el número óptimo de clústers a crea
#’
#’ Haré el gráfico de dendrograma, aunque nota que en este caso primero reordenaré los sitios co
#’
u_dend_reord <- reorder.hclust(lista_cl$cl_upgma, mi_fam_norm_d)
plot(u_dend_reord, hang = -1)
rect.hclust(
  tree = u_dend_reord,
  k = anch_sil_upgma$n_grupos_optimo)
#’
#’ Ahora compararé el dendrograma con el mapa de calor en un mismo gráfico, colocando los dendro
#’
heatmap(
  as.matrix(mi_fam_norm_d),
  Rowv = as.dendrogram(u_dend_reord),
  symm = TRUE,
  margin = c(3, 3),
  col = rev(cm.colors(4))
)
#’
#’ En general, hay dos grupos, uno grande y otro pequeño, y parece haber un tercero en el mapa d
#’
#’ Mostraré el resultado para Ward:
#’
anch_sil_ward <- calcular_anchuras_siluetas(
  mc_orig = mi_fam,
  distancias = mi_fam_norm_d,
  cluster = lista_cl$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_upgma <-
  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_upgma, hang = -1)

```

```

# Añadir rectángulos a los grupos significativos
lines(cl_pvclust_upgma)
pvrect(cl_pvclust_upgma, 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_upgma_k2 <- as.factor(cutree(lista_cl$cl_upgma, k = 2)))
#'
#' En este caso, los sitios 1 y 2 pertenecen al grupo 1, los sitios 3 al 6 pertenecen al grupo 2
#'
table(grupos_upgma_k2)
#'
#' Nota lo desiguales que son estos grupos, un efecto esperado dado el alto grado de autocorrela
#'
#' Ward:
#'
(grupos_ward_k3 <- as.factor(cutree(lista_cl$cl_ward, k = 3)))
table(grupos_ward_k3)
#'
#' Guardaré estos vectores en archivos para reutilizarlos en *scripts* posteriores:
#'

```

```

saveRDS(grupos_upgma_k2, 'grupos_upgma_k2.RDS')
saveRDS(grupos_ward_k3, 'grupos_ward_k3.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

```

7.12 Análisis de agrupamiento (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/Sapotaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
grupos_upgma_k2 <- readRDS('grupos_upgma_k2.RDS')
table(grupos_upgma_k2) #Importante, tener en cuenta los desiguales tamaños de los grupos
grupos_ward_k3 <- readRDS('grupos_ward_k3.RDS')
table(grupos_ward_k3)
#'
#' ### 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 distribuc
#’
#’ Primero crearé un objeto que permita realizar tanto las pruebas como los diagramas de cajas.
#’
(m_amb_upgma_k2 <- bci_env_grid %>%
  select_if(is.numeric) %>% select(-id) %>%
  mutate(grupos_upgma_k2) %>%
  st_drop_geometry() %>%
  pivot_longer(-grupos_upgma_k2, names_to = "variable", values_to = "valor"))
#’
#’ A continuación, las pruebas:
#’
m_amb_upgma_k2 %>%
  group_by(variable) %>%
  summarise(
    p_valor_t = t.test(valor ~ grupos_upgma_k2)$p.value,
    p_valor_w = wilcox.test(valor ~ grupos_upgma_k2, exact = F)$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_upgma_k2 %>%
  group_by(variable) %>%
  ggplot() + aes(x = grupos_upgma_k2, y = valor, fill = grupos_upgma_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 = 16)
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")
)
#'
```

7.13 Análisis de agrupamiento (*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/Sapotaceae.Rdata')
mi_fam <- mc_saptc
grupos_upgma_k2 <- readRDS('grupos_upgma_k2.RDS')
table(grupos_upgma_k2)
grupos_ward_k3 <- readRDS('grupos_ward_k3.RDS')
table(grupos_ward_k3)
#'
#' ## 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_k3 <- multipatt(
  x = mi_fam,
  cluster = grupos_ward_k3,
  func = 'IndVal.g',
  max.order = 2,
  control = how(nperm = 999))
summary(iva_ward_k3, indvalcomp = TRUE)
colSums(mi_fam)
(p_ward_adj <- p.adjust(iva_ward_k3$sign$p.value))
(iva_ward_boot <- strassoc(
  X = mi_fam,
  cluster = grupos_ward_k3,
  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_k3 <- multipatt(
  mi_fam,
  grupos_ward_k3,

```

```

func = "r.g",
max.order = 2,
control = how(nperm = 999))
summary(phi_ward_k3)
colSums(mi_fam)
(phi_ward_boot <- strassoc(
  X = mi_fam,
  cluster = grupos_ward_k3,
  func = "r.g",
  nboot = 1000))

```

7.14 *Análisis de 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/Sapotaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
mi_fam <- mc_saptec
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
```

#'

```

#' El término "diversidad biológica" se asume que fue acuñado a principios de los 80, y se refiere a la
```

#'

```

#' 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 todos los niveles
```

#'

```

#' 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 dos 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 el número de individuos
```

#'

```

#' #### Riqueza de especies y rarefacción
```

#'

```

#' Magurran (2004) distingue entre densidad de especies, que equivale al número de especies presentes en un sitio
```

#'

```

#' Para asegurar la comparabilidad entre sitios, se han propuesto distintos métodos. Uno es la rarefacción
```

#'

```

#' 
```

#'

```

#' #### 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 encuentran las especies)
```

#'

```

#' Para una q dada, H asume su valor máximo ( $H_{\max}$ ) cuando todas las especies presentes en el sitio tienen la misma
```

#'

```

#' 
```

#'

```

#' 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 equi
#'
#' 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 ( $H_{a=0,1,2}$ ), 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  $J = H_{a=1} / H_{a=0}$ 
#'
#' Usando la misma aproximación, Hill propuso también otras ratios que, a diferencia de lo que o
#'
#' -  $E_{a=1} = N_{a=1} / N_{a=0}$  a la cual el propio Hill denominó como su versi
#'

```



```

#' -  $E_{2/N}^2$ , 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(2,4), label_size = 2)
#'
#' **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 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/Sapotaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
mi_fam <- mc_saptc
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 de ordenación no restringida*

```

#' ---
#' 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/Sapotaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
mi_fam <- mc_saptc
(colnames(mi_fam) <- make.cepnames(colnames(mi_fam)))
(df_equivalencias <- data.frame(
  nombre_original = colnames(mc_saptc),

```

```

  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 paso es
#'
#' ¡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*

```

```

#’
screepplot(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 sobresal
  #'
  #' Evaluaré el ajuste del PCA de datos de comunidad a datos ambientales, mediante la función 'en
  #'
  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 d
  #'
  #' NOTA: te recomiendo probar otros métodos de selección de variables, como por ejemplo, usando
  #'
  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 de
#' - Las relaciones entre variables se miden por ángulos, no por proximidad de las puntas de los
#'
#' ### Análisis de correspondencia (CA)
#'
mi_fam_ca <- cca(mi_fam)
summary(mi_fam_ca)
summary(mi_fam_ca, scaling = 1)
#'
#' Screeplot
#'
screeplot(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('Thevahou', colnames(mi_fam))])
summary(mi_fam_ca)
summary(mi_fam_ca, scaling = 1)
screeplot(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)

```

7.17 Análisis de ordenación restringida o canónica

```

#' ---
#' 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/Sapotaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
mi_fam <- mc_saptc
(colnames(mi_fam) <- make.cepnames(colnames(mi_fam)))
(df_equivencias <- data.frame(
  nombre_original = colnames(mc_saptc),
  colnames(mi_fam)))
bci_env_grid %>% tibble
#'
#' ## Ordenación restringida
#'
#' Con este conjunto de técnicas de ordenación, las tendencias detectadas en un conjunto de datos
#'
#' 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 ambiente)
#'
#' RDA combina la regresión y el análisis de componentes principales. El RDA es una regresión lineal
#'
#' - Cada variable de respuesta (e.g. matriz de comunidad) es ajustada (por regresión) a la matriz de
#'
#' - Se prueba significancia estadística de la relación entre las variables de respuesta y explicativas
#'
#' - Si la prueba resulta significativa, se realiza un PCA de la matriz de valores ajustados. Es decir, se
#'
#' - Luego se usa la matriz de vectores propios canónicos para calcular dos tipos de puntuaciones
#'
#'   - Usando la matriz de valores ajustados para obtener una ordenación en el espacio de las variables
#'
#'   - Usando la matriz de valores centrados de las variables de respuesta para obtener una ordenación
#'
#' Mostraré a continuación ejemplos aplicados a mi familia asignada. Para agilizar, trasladaré los datos
#'
#' Nota (de nuevo): recuerda que los datos de BCI están autocorrelacionados espacialmente, por lo que

```

```

#'
#' ##### 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 relativiz
#'
#' 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 conservador
#'
#' Finalmente, la representación del modelo se realiza en un *triplot*, que es un gráfico enriqu
#'
#' 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"
)

```

7.18 *Análisis espacial de datos ecológicos*

```

#' ---
#' 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/Sapotaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
mi_fam <- mc_saptec
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$var))))
#'
#' ### Variables ambientales (matriz ambiental)
#'
bci_env_grid_num <- bci_env_grid %>%
  st_drop_geometry %>%
  select_if(is.numeric) %>%
  select(-id, -UTM.EW, -UTM.NS)
suppressWarnings(auto_amb <- calcular_autocorrelacion(
  df_fuente = bci_env_grid_num,
  orden = 9,
  obj_vecindad = vecindad))
print(auto_amb, digits = 2, p.adj.method = 'holm')
dim_panel <- rev(n2mfrow(ncol(bci_env_grid_num)))
#+ fig.width=12, fig.height=14
par(mfrow = dim_panel)
suppressWarnings(invisible(lapply(auto_amb, function(x) plot(x, main = x$var))))
#'
#' ## 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)
#'
#' ## 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
#'
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))
#'
#' #### Aplicado a abundancias de especies transformadas
#'
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$leyenda,
#
#' ##### Aplicado a abundancias de especies transformadas sin tendencia
#
mi_fam_sintendencia_sf <- bci_env_grid %>% select %>% bind_cols(mi_fam_sin_tendencia %>% as.data
lisamaps_mifam_sintendencia <- sapply(
  grep('geometry', names(mi_fam_sintendencia_sf), invert = T, value = T),
  function(x) {
    m <- lisamap(objesp = mi_fam_sintendencia_sf[x],
      var = x,
      pesos = pesos_b,
      titulo_leyenda = 'Significancia ("x-y", léase como "x" rodeado de "y")',
      leyenda = F,
      anchuratitulo = 50,
      tamanotitulo = 10,
      fuentedatos = '\nhttp://ctfs.si.edu/webatlas/datasets/bci/',
      titulumapa = paste0('Clusters LISA de "', x, '"'))
    # dev.new();print(m$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

```

Referencias

- 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>
- Brocard, D., Gillet, F., & Legendre, P. (2018). Numerical ecology with r. *Springer Nature, Second Edition*, 52–66. Retrieved from <https://doi.org/10.1007/978-3-319-71404-2>
- Campos-Pineda, E. G., Moreno, J., & Mendieta, J. (2017). Análisis florístico de la vegetación arbórea de una parcela de bosque en el parque natural metropolitano, provincia de panamá. *Scientia*, 27(1), 7–24.
- Carmona-Galindo, V. D., & Carmona, T. V. (2013). La diversidad de los análisis de diversidad la diversidad de los analisis de diversidad [the diversity of diversity analyses]. *Bioma*.
- Cáceres, M. D., & Legendre, P. (2009). Associations between species and groups of sites: Indices and statistical inference. *Ecology*, 90(12), 3566–3574.
- Condit, R. A. y H., Richard y Chisholm. (2012). Treinta años de censo forestal en barro colorado y la importancia de la inmigración para mantener la diversidad. *PloS One*, 7(11), e49826.
- Condit, R. y L., Richard y Pérez. (2017). Tendencias demográficas y clima durante 35 años en

la parcela de 50 ha de barro colorado. *Forest Ecosystems*, 4(1), 1–13.

Condit, R., Pitman, N., Leigh, E. G., Chave, J., Terborgh, J., Foster, R. B., ... others. (2002). Beta-diversity in tropical forest trees. *Science*, 295(5555), 666–669.

Conjuntos de especies y especies indicadoras: La necesidad de un enfoque asimétrico flexible. (1997). *Monografías Ecológicas*, 67(3), 345–366.

De Caceres, M., & Legendre, P. (2009). Associations between species and groups of sites: Indices and statistical inference. In *Ecology*. Retrieved from <http://sites.google.com/site/miqueldecaceres/>

Doblado Amador, L. S. (2011). Identificación y caracterización de tipos de bosque y su relación con variables ambientales, en un paisaje fragmentado al norte de honduras. *Proyecto Finnfor I Y Finnfor II-CATIE*.

Horvát, S., Derzsi, A., Nédá, Z., & Balog, A. (2010). A spatially explicit model for tropical tree diversity patterns. *Journal of Theoretical Biology*, 265(4), 517–523.

Isla barro colorado y biología tropical. (1990). *Cuatro Bosques Neotropicales*, 28–47.

Kindt, R., & Coe, R. (2005). *Tree diversity analysis. a manual and software for common statistical methods for ecological and biodiversity studies*. Retrieved from <http://www.worldagroforestry.org/output/tree-diversity-analysis>

Legendre, P., & Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129(2), 271–280.

Martínez-Sovero, G., Iglesias-Osores, S., & Villena-Velásquez, J. J. (2020). Importancia de la familia sapotaceae en madre de dios, Perú. *Manglar*, 17(4), 287.

Martínez-Sovero, G., Iglesias-Osores, S., Muñoz-Chavarry, P., Seminario-Cunya, A., Alva-Mendoza, D., & Villena-Velásquez, J. (2021). Diversidad y estructura de sapotaceae en bosques amazónicos de madre de dios, Perú. *Ciencia Amazónica (Iquitos)*, 9(1), 59–72.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... Wagner, H. (2019). *Vegan: Community ecology package*. Retrieved from <https://CRAN.R-project.org/package=vegan>

Pérez, R., Aguilar, S., Condit, R., Foster, R., Hubbell, S., & Lao, S. (2005). Metodología empleada en los censos de la parcela de 50 hectareas de la isla de barro colorado, Panamá. *Centro de Ciencias Forestales Del Tropico (CTFS) Y Instituto Smithsonian de Investigaciones Tropicales (STRI)*, 1–24.

R Core Team. (2020). *R: A language and environment for statistical computing*. Retrieved from <https://www.R-project.org/>

Rodríguez-Flores, W., & Barrios, H. (2020). Scolytinae y platypodinae (coleoptera: Curculionidae) de la isla barro colorado, Panamá. *Scientia*, 30(1), 15–52.

Smedmark, A. A., Jenny EE y Anderberg. (2007). La migración boreotropical explica la hibridación entre linajes geográficamente distantes en el clado pantropical sideroxyloae (sapotaceae). *American Journal of Botany*, 94(9), 1491–1505.

Wan, D. (2020). IMPORTANCIA de los bosques y estado de los bosques en Guatemala. *Revista Ingeniería Y Ciencia*, 2(12).

Whittaker, R. H. (1960). Vegetation of the siskiyou mountains, Oregon and California. *Ecological Monographs*, 30(3), 279–338.

Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'*. Retrieved from <https://CRAN.R-project.org/package=tidyverse>

Windsor, D., FOSTER, R., BROKAW, N., Leigh, E., Rand, A., & others. (1990). *Estructura e historia de la vegetación de la isla barro colorado ecología de un bosque tropical: Ciclos estacionales y cambios a largo plazo* (pp. 113–127).