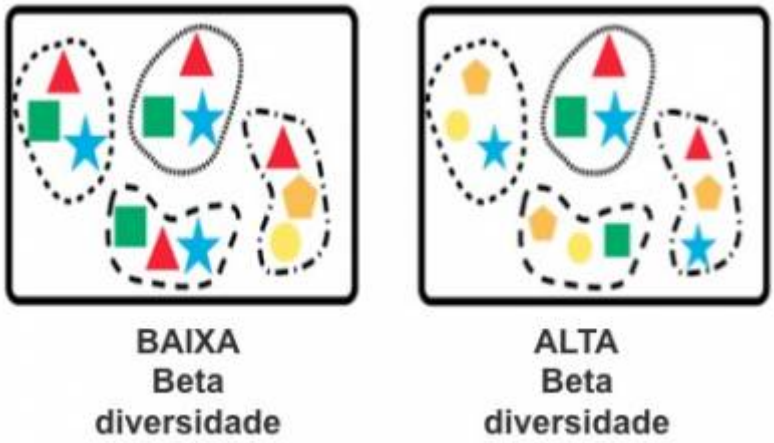
Partição de variação de comunidades

Comunidades vegetais podem apresentar diferentes conjuntos de espécies mesmo em localidades espacialmente próximas e diferentes fatores podem determinar essas diferenças. Quantificar a importância relativa dos diferentes fatores é essencial para entender a estruturação e o funcionamento das comunidades e tem sido um grande desafio. Uma das técnicas muito utilizadas atualmente é a chamada **Partição da Variação** e aqui você vai aprender a aplicar essa técnica a dados de natureza multivariada, que são os conjuntos de dados que possuem muitas variáveis de resposta. Esse é o caso de estudos sobre a **variação na composição de espécies entre comunidades ou entre amostras de comunidades**. Essa variação é também conhecida como **beta diversidade**.



Partição da Variação para Dados Multivariados

Agora que você entendeu como funciona a técnica de partição da variação usando apenas uma espécie (Roteiro de partição de variação univariada), vamos voltar ao interesse central dessa prática, que é entender quais fatores e processos são mais importantes para explicar a **variação na composição de espécies entre diferentes comunidades ou entre diferentes amostras de comunidades**. Como nosso interesse é analisar a estrutura da comunidade como um todo, ao invés de analisarmos cada espécie separadamente, vamos analisar os padrões de abundância de muitas espécies em conjunto.

Considerando que cada espécie representa uma variável importante para a definição das diferenças entre as localidades e que serão analisadas várias espécies simultaneamente, é necessário o uso de técnicas de **Análises Multivariadas**. Existem muitos procedimentos analíticos diferentes e, em função do pouco tempo, teremos que focar em alguns procedimentos específicos. Entretanto, alguns textos e links estão disponíveis no final desse roteiro para quem se interessar mais pelo assunto.

Inicialmente vamos utilizar os chamados **Métodos de Ordenação Irrestrita** para identificar quais localidades são mais similares e quais são mais diferentes entre si e para identificar quais espécies

são mais importantes na determinação dessas diferenças.

Depois, vamos propor alguns fatores que poderiam explicar os padrões observados e vamos aplicar os **Métodos de Ordenação Restrita** para quantificar a importância de cada fator na explicação do padrão observado, seguindo a mesma lógica aplicada aos dados univariados trabalhados no (roteiro de partição de variação para uma espécie).

Por fim, vamos utilizar os valores obtidos a partir dos métodos de ordenação para construir a tabela de **Partição da Variação** e quantificar a importância relativa de cada fator.

Preparação para o exercício

Para começar, escolha uma pasta para você guardar os arquivos e resultados deste roteiro em seu computador. Copie para essa pasta os arquivos com os dados que vamos usar:

- especies.csv
- ambientais.csv
- coordenadas.csv

Em seguida, abra o programa R, clicando no ícone que está na área de trabalho do seu computador.

Já com a janela do programa R aberto, o próximo passo será mudar o diretório de trabalho para aquela pasta que você acabou de criar. Com isso será mais fácil importar os dados dos arquivos ".csv" para dentro do ambiente R.

A mudança de diretório deve ser feita da seguinte forma:

- Abra o Menu "Arquivo" (ou "File");
- Selecione "Mudar dir" (ou "Change dir");
- Escolha a sua pasta na janela que abrir.

[Obs. Para Mac, essa opção está no Menu "Misc" e a opção é "Change working dir"]

Para checar se você está na pasta correta, copie e cole o comando abaixo na linha de comando do R. Atenção: O comando deve ser colado na frente do símbolo ">". Este símbolo indica o início da linha de comando ou "prompt", onde você deve escrever os comandos para o R.

```
getwd()
```

Após colar, aperte a tecla "enter" e veja se o R retorna o nome da sua pasta. Se sim, ótimo. Se não, chame um monitor ou professor.

Instalando e carregando o pacote vegan

Para fazer as análises vamos usar um pacote muito legal para quem quer trabalhar com análises de

vegetação. O pacote se chama *vegan* e para instalá-lo, use o comando:

```
install.packages("vegan")
```

Se o programa perguntar se você quer instalar as dependências, selecione **sim/yes**!

E agora, carregue o pacote no R, usando o comando library:

```
library(vegan)
```

Conhecendo os dados da comunidade vegetal hipotética

Imagine que amostramos uma área de 1km x 1km, dividindo-a em 100 parcelas de 100 x 100m. Identificamos todas as espécies arbóreas acima de 20 cm de DAP e anotamos o número de indivíduos de cada espécie em cada parcela.

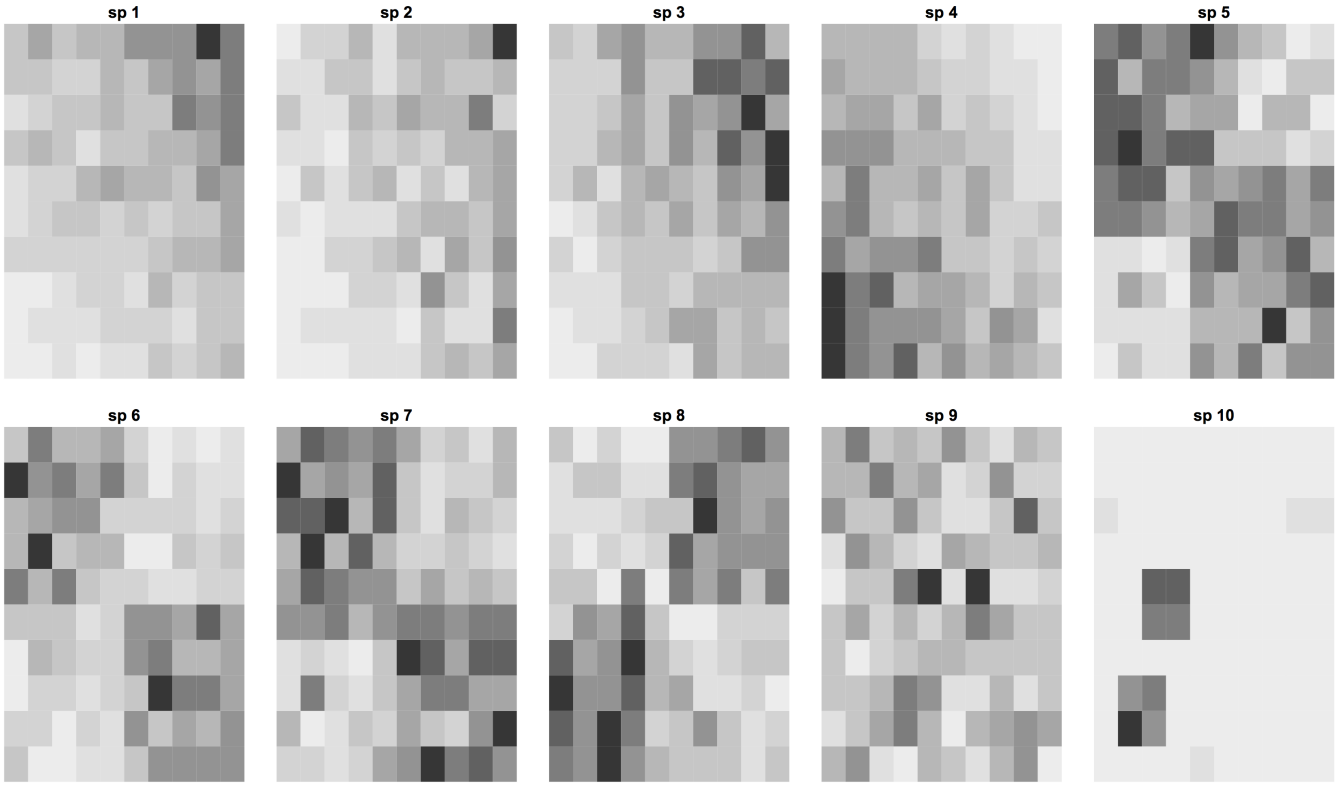
Tabulamos os dados em uma planilha (especies.csv) e agora vamos importar essa planilha para o R e visualizá-la dentro do próprio R. Para isso copie e cole o comando abaixo no R:

```
(sp<-read.csv2 ("especies.csv"))
```

Suba a tela do R de volta até o início da planilha para ver o cabeçalho.

Nessa planilha as linhas se referem às parcelas e as colunas se referem às espécies. Cada célula da planilha contém o número de indivíduos (ou seja, a abundância) de uma espécie em uma dada parcela.

Agora vamos visualizar figuras que mostram a variação de abundância de cada uma das espécies na nossa área. Nessa figura, quanto mais escuro for o quadrado, maior a abundância da espécie naquela amostra.

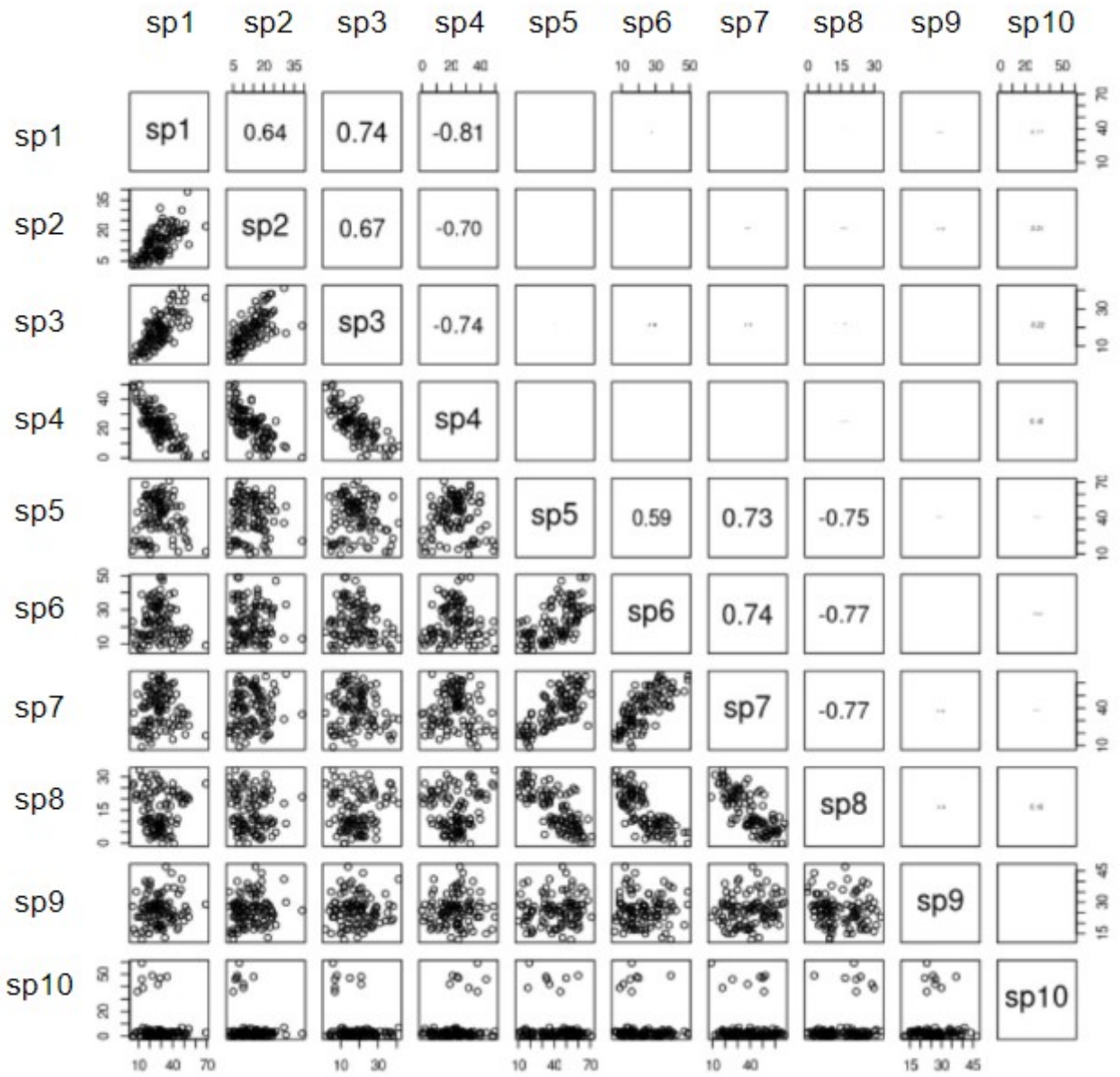


rior à esquerda é a parcela número 1 da planilha e a parcela no vértice superior à direita é a parcela numero 100 da planilha. Dê uma conferida rápida se os valores de abundância das 10 espécies nessas duas parcelas (1 e 100) faz sentido com o que está sendo visualizado nos mapas acima.

Visualizando correlações entre as espécies

A partir dos mapas acima podemos observar que algumas espécies têm distribuições bem parecidas na área. É possível identificar uns três ou quatro padrões principais de distribuição das espécies.

Uma forma de tentar entender como as espécies estão estruturadas em uma comunidade é analisando a associação entre cada par de espécie presente na comunidade. Podemos quantificar isso por meio de uma **Análise de Correlação** entre as abundâncias das espécies. Quanto maior for a correlação, mais associada será a distribuição de um dado par de espécies. O sinal da correlação indicará se a associação é positiva (ou seja, nas parcelas em que uma espécie é mais abundante a outra também será) ou negativa (ou seja, nas parcelas em que uma espécie é abundante a outra será rara). Alguns pares de espécies podem não estar associados (valores pequenos de correlação), ou seja, as abundâncias das duas espécies variam independentemente. A figura abaixo sintetiza estas correlações em uma matriz de gráficos e respectivos coeficientes de correlação.



Na diagonal principal da figura temos os códigos das espécies, que também estão indicados em cada linha e coluna

Na parte inferior e à esquerda da diagonal principal, temos os gráficos que mostram as relações entre as abundâncias de cada par de espécies. Cada ponto nesses gráficos representa uma das 100 parcelas da nossa área. Para entender qual espécie está representada no eixo X e qual está no eixo Y de cada um dos gráficos, basta encontrar os códigos das espécies nas linhas e colunas correspondentes. Por exemplo, no cruzamento da coluna 3 com a linha 4 tem o gráfico mostrando a relação das abundâncias da espécie 3 (Sp3) no eixo X, e da espécie 4(Sp4) no eixo Y.

Na parte superior e à direita da diagonal principal da figura temos os **valores das correlações** entre as abundâncias das espécies nas parcelas. Esses valores são referentes aos gráficos que se encontram na posição espelhada na parte inferior da diagonal principal. Por exemplo, o valor do coeficiente de correlação entre as abundâncias da espécie 3 e 4 está no cruzamento da linha da espécie 3 com a coluna da espécie 4, que está acima da diagonal. Coeficientes de correlação estão sempre dentro do intervalo entre -1,0 e +1,0. Valores dos coeficientes em módulo (“sem o sinal”) próximos de um indicam os pares de espécies que estão mais fortemente associadas. O sinal indica se a correlação é positiva (no caso, se a abundância de uma espécie é alta nas parcelas em que a abundância da outra também é alta), ou negativa (no caso, se a abundância de uma espécie é alta

nas parcelas onde a outra é pouco abundante). Para facilitar a visualização na figura, quanto maior for o valor da correlação, maior será tamanho dos caracteres usados para escrever o valor da correlação. Correlações muito fracas (valores próximos de zero) sequer aparecem na figura.

Preste atenção aos pares de espécies com correlações mais fortes, avalie se as correlações entre elas são positivas ou negativas e anote um par para o qual não há nenhuma evidência de correlação.

Agora volte à figura anterior, e localize os mapas de abundâncias por parcelas das espécies do par que você escolheu. A relação entre os mapas das duas espécies são coerentes com a correlação calculada entre as abundâncias das espécies?

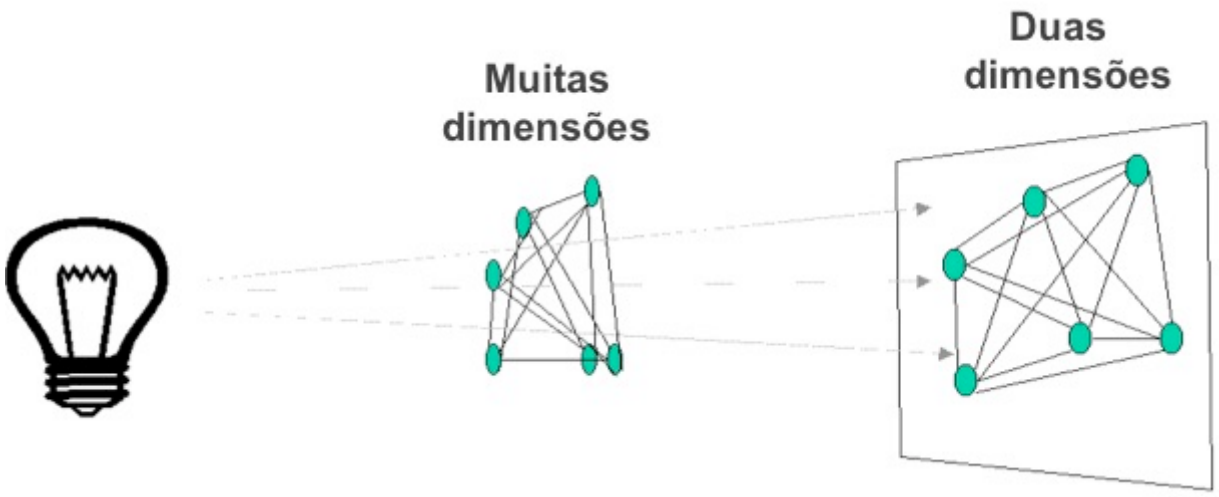
Repare que como são apenas 10 espécies não é difícil visualizar as correlações par-a-par desta maneira, pois temos apenas 45 gráficos. Mas imagine se tivéssemos as mais de 1000 espécies da parcela permanente de Manaus? Só para terem uma ideia, para 100 espécies, teríamos 4.950 gráficos... Então, nesses casos, precisamos utilizar técnicas multivariadas, como, por exemplo, os **Métodos de Ordenação**.

Método de Ordenação Irrestrita - PCA (Análise de Componentes Principais)

Os métodos de ordenação irrestrita ("*unconstrained ordination analysis*") são aplicados em ecologia de comunidades vegetais quando desejamos, por exemplo, analisar padrões de similaridade de composição de espécies entre parcelas mas temos um conjunto muito grande de espécies.

Se quiséssemos visualizar a similaridade entre as 100 parcelas da nossa comunidade usando diretamente os dados das abundâncias das nossas 10 espécies, precisaríamos de um gráfico com 10 eixos (dimensões), cada eixo representando a abundância de uma espécie. Porém, não teríamos a capacidade de visualizá-lo, pois só conseguimos visualizar até 3 dimensões (sim, é triste admitir, mas nós humanos somos muito limitados...).

Entretanto, se existem correlações fortes entre algumas das espécies, podemos reduzir as dimensões (eixos) a um número menor de **eixos derivados** (também chamados de "componentes") nos quais a maior parte das relações entre as parcelas estará representada. Matematicamente, isso será obtido por meio de álgebra matricial[1)], mas não entraremos em detalhes aqui.



A Análise de Componentes Principais (PCA) vai “procurar”, a partir da nuvem de pontos formada pelas abundâncias das 10 espécies nas 100 parcelas, um eixo imaginário que capta a maior parte da variação nos dados. Esse eixo, conhecido como **Componente Principal 1** (ou pela sigla **PCA1**), expressará as similaridades e as diferenças mais marcantes entre as parcelas. Depois da definição do PCA1, em geral, ainda haverá uma certa quantidade de correlação não explicada. A Análise de Componentes Principais vai então procurar um segundo eixo (PCA2) que vai explicar o máximo possível da variação remanescente e depois um terceiro eixo e assim por diante [2)].

Vamos, então, aplicar o procedimento de análise para a nossa comunidade de 10 espécies e 100 parcelas!

Para realizar a Análise de Componentes Principais para a nossa comunidade vamos aplicar a função *rda* (existente no pacote *vegan* do R) à nossa planilha de espécies (sp) e vamos gerar um novo objeto, que vamos chamar de “pca.sp”. Esse objeto “pca.sp” vai conter todos os resultados da análise, incluindo todas as informações necessárias para construirmos um gráfico com os dois primeiros eixos. Copie e cole no R os comandos abaixo[3)] e você produzirá o gráfico:

```
pca.sp <-rda(sp, scale = TRUE)
biplot(pca.sp, scaling=2, type = c("text", "text"))
```

No gráfico que foi produzido, os pontos representam as parcelas, os vetores vermelhos (setas) indicam em que direção aumenta a abundância de cada espécie e o comprimento dos vetores indica a importância da espécie na definição de cada eixo.

Em um extremo do novo eixo derivado 1 (PCA1) estarão projetadas as parcelas mais parecidas entre si, enquanto no extremo oposto estarão as parcelas mais diferentes dessas (porém, similares entre si). Também podemos identificar para esse eixo 1 as espécies mais importantes na definição das similaridades e diferenças entre parcelas. A mesma interpretação se aplica ao eixo 2 (PCA2)

Você consegue relacionar esse gráfico com os mapas de distribuição das espécies e com as figuras de correlação apresentadas anteriormente? Quais espécies têm abundâncias correlacionadas, ou seja, quais são as associações de espécies? Quais parcelas têm cada associação?

Muito bem, depois de entendermos a lógica básica de funcionamento de uma PCA e interpretarmos a ordenação produzida pelos dois primeiros eixos, a pergunta que surge naturalmente é:

“Por que algumas parcelas são mais similares, com conjuntos de espécies ocorrendo predominantemente e em maior abundância do que em outras parcelas?”

Tempo para pensar!

Uma das respostas mais rápidas que vêm à mente de um(a) ecólogo(a) é que parcelas com maior similaridade de espécies podem estar sujeitas a variáveis ambientais mais parecidas e que as diferenciam do restante.

Como podemos avaliar e quantificar isso? É aqui que entram os **Métodos de Ordenação Restrita**

Método de Ordenação Restrita - RDA (Análise de Redundância)

Um método de ordenação restrita (“*constrained ordination analysis*”) permite quantificar qual proporção da variação na composição de espécies entre as parcelas pode ser explicada por um dado conjunto de variáveis explanatórias. Essas variáveis explanatórias podem ser variáveis ambientais, espaciais, atributos biológicos, etc. A proporção de variação explicada é similar ao R^2 que usamos para uma espécie no roteiro sobre partição univariada. Agora vamos generalizar o conceito de variação explicada para dados multivariados, ou seja, para muitas espécies.

Uma das análises que faz esses cálculos é a chamada de Análise de Redundância (*RDA - Redundancy Analysis*). Esse tipo de análise é chamado de ordenação “restrita” exatamente porque o conjunto de dados que tem as variáveis resposta (espécies, no nosso caso) vai ser condicionado (*constrained*) pelo conjunto de dados que tem as variáveis explanatórias (por exemplo, variáveis ambientais, no nosso primeiro caso abaixo).

Variáveis ambientais

Podemos ir a campo e medir variáveis ambientais em todas as parcelas, como características físicas do solo (granulometria, porosidade, etc), características químicas do solo (N, P, K, Ca, Mg, pH, Matéria Orgânica, Capacidade de Troca Catiônica, etc), topografia e luminosidade, etc. Com esses dados

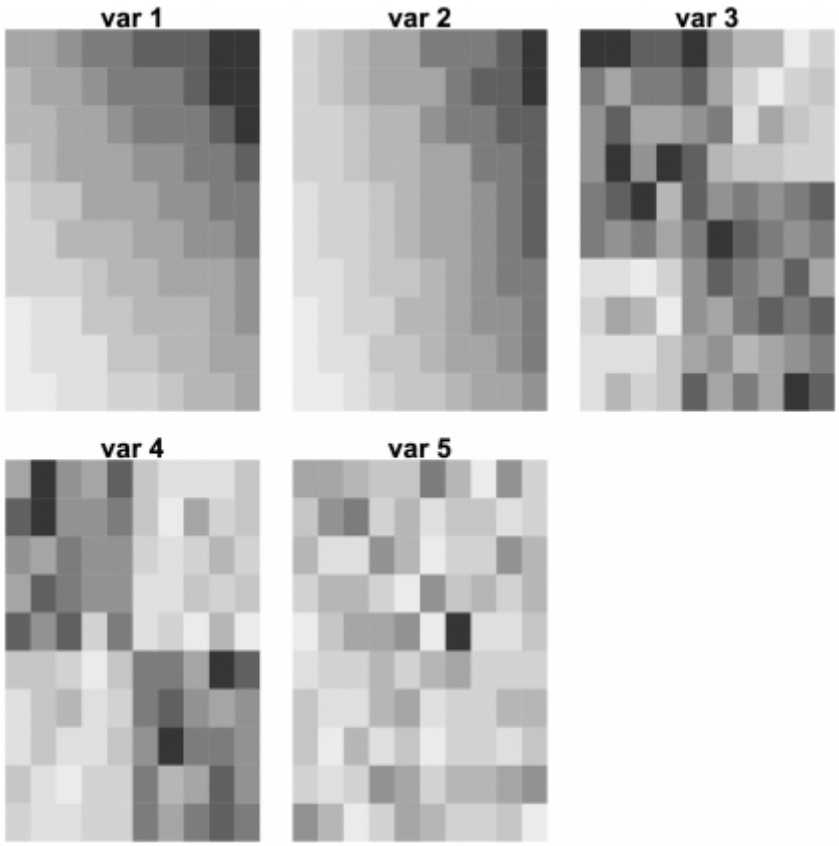
podemos construir uma tabela em que as linhas são as parcelas e as colunas são as variáveis ambientais.

Vamos conhecer então as variáveis ambientais hipotéticas desse nosso exercício prático. Importe a planilha para o R e visualize:

```
(amb<-read.csv2 ("ambientais.csv"))
```

Suba a tela de volta ao início da planilha. Nela, temos as 5 variáveis ambientais hipotéticas (x1 a x5) e os valores de cada uma delas em cada parcela.

Agora, vamos visualizar como essas variáveis ambientais estão distribuídas espacialmente na nossa área. Nessa figura, quanto mais escuro for o quadrado, maior o valor da variável ambiental naquela parcela. Dê um conferida se os valores das variáveis nas parcelas #1 e #100 fazem sentido com os mapas abaixo.

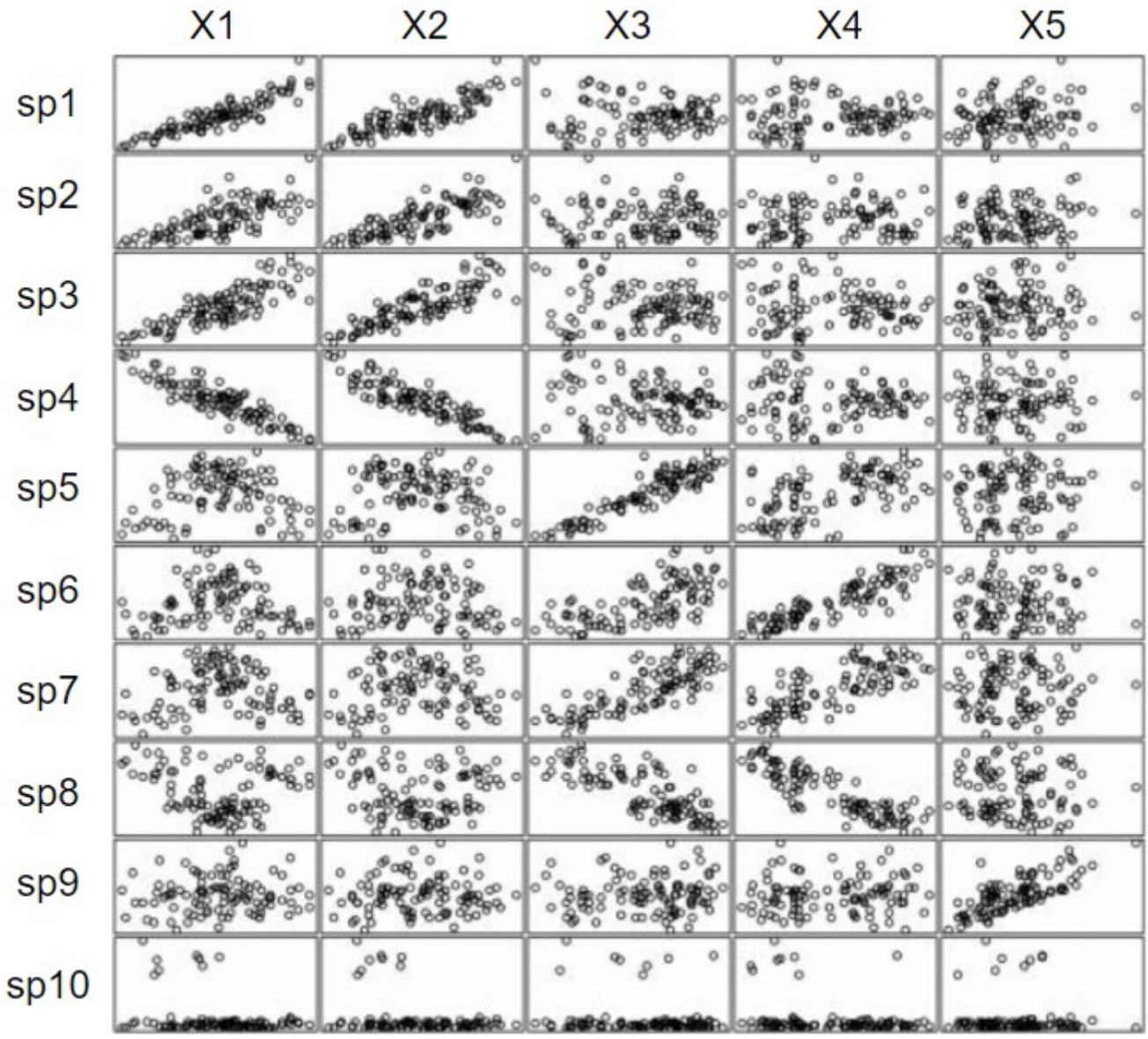


Quais variáveis parecem estar mais correlacionadas entre si?

Relações entre espécies e variáveis ambientais

Queremos entender se as variáveis ambientais ajudam a explicar as similaridades entre as parcelas. Um primeiro passo pode ser inspecionar visualmente as correlações entre as abundâncias das espécies e os valores das variáveis ambientais em cada parcela.

Essa figura abaixo é diferente da figura anterior de correlações. Nessa, cada espécie está representada em uma linha e cada variável ambiental em uma coluna. Os gráficos apresentam as relações entre os valores de abundância de uma espécie (linhas) e os valores de uma variável ambiental (colunas) para todas as parcelas. Assim o primeiro gráfico em cima e à esquerda apresenta a relação entre a abundância da espécie 1 (*sp1*) e os valores da variável ambiental 1 (*X1*), nas parcelas. O gráfico logo abaixo apresenta a relação entre a espécie 2 (*sp2*) e a variável 1 (*X1*), e assim por diante.



Preste atenção em quais variáveis ambientais apresentam correlações mais fortes com quais espécies.

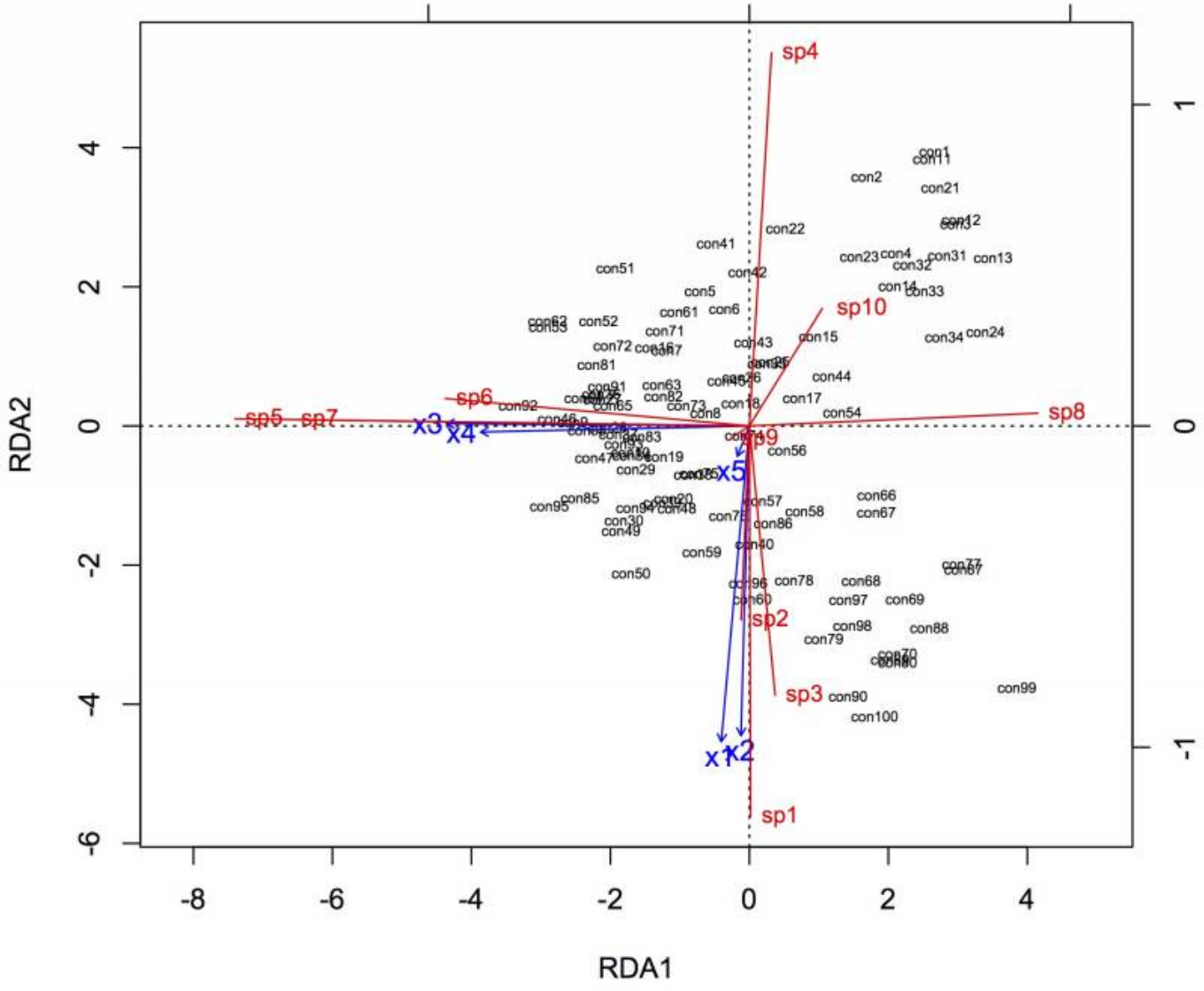
Calculando RDA para espécies x variáveis ambientais

Em uma análise de RDA, quanto maior o número de espécies correlacionadas com as variáveis ambientais e quanto mais forte forem essas correlações, maior será a importância das variáveis ambientais na **explicação da variação da similaridade entre as parcelas** e portanto, maior será o valor de proporção explicada da estrutura da comunidade.

Vamos então fazer a RDA para os nossos dados, ajustando os dados das abundâncias das espécies (a planilha "sp") em função das variáveis ambientais (planilha "amb") [4)]:

```
rda.sp.amb<-rda(sp, amb)
```

Como a elaboração do gráfico resultante dessa análise é um pouco mais complicada, já vamos mostrar o gráfico pronto para você. Esse gráfico mostra a ordenação das parcelas, das espécies e das variáveis ambientais nos dois primeiros eixos (RDA1 e RDA2). As parcelas estão representadas pelos pontos pretos, as variáveis ambientais estão representadas pelas setas azuis e as espécies pelas linhas em vermelho.



Você consegue entender quais são as variáveis ambientais que determinam mais fortemente a distribuição das espécies e que, por sua vez, determinam a ordenação das parcelas?

Agora vamos olhar e interpretar os resultados numéricos obtidos nessa análise e que nos interessam para estimar a Partição da Variação. Para isso, copie e cole o comando abaixo:

```
rda.sp.amb
```

Na primeira tabela que aparece, na linha referente ao item "*Constrained*", o valor apresentado na coluna "*Proportion*" representa a proporção da variação das abundâncias das espécies nas parcelas que é explicada pelas variáveis ambientais que utilizamos nessa análise de ordenação restrita. Essa proporção tem a mesma interpretação do R^2 do Roteiro 2.

O valor obtido é equivalente ao X do Roteiro de partição univariada, ou seja, a fração total explicada pelas variáveis ambientais.

Para salvar o valor de X dentro do R, use o comando abaixo:

```
(X <- (rda.sp.amb$CCA$tot.chi)/(rda.sp.amb$tot.chi))
```

E então, você acha que nossas variáveis ambientais explicam pouco ou muito da distribuição das espécies?

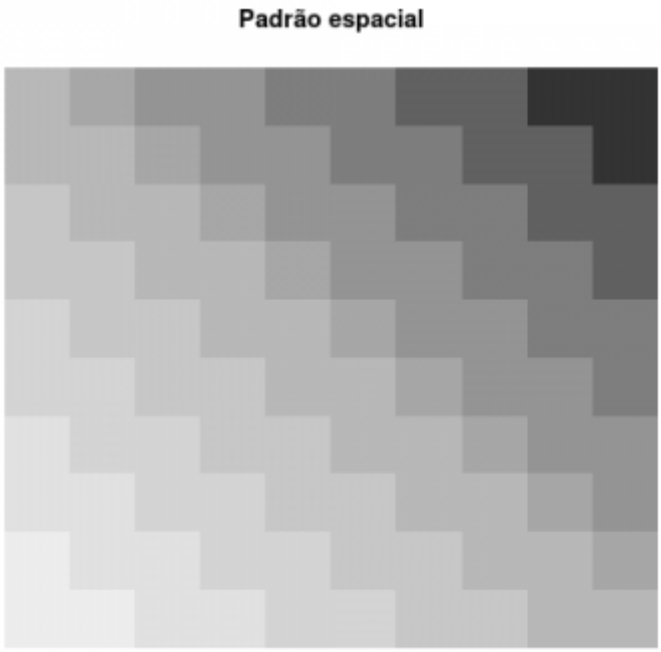
Estrutura espacial

Ok, até aqui, então, temos uma possível explicação para a distribuição das espécies nas parcelas. Porém, existe ainda uma outra explicação que os(as) ecólogos(as) sempre precisam avaliar antes de afirmar que são as variáveis ambientais que determinam a similaridade na composição de espécies entre determinadas parcelas.

As espécies podem estar distribuídas de forma **espacialmente agregada** em algumas parcelas e isso pode ser, por exemplo, um reflexo da **limitação da dispersão** das espécies.

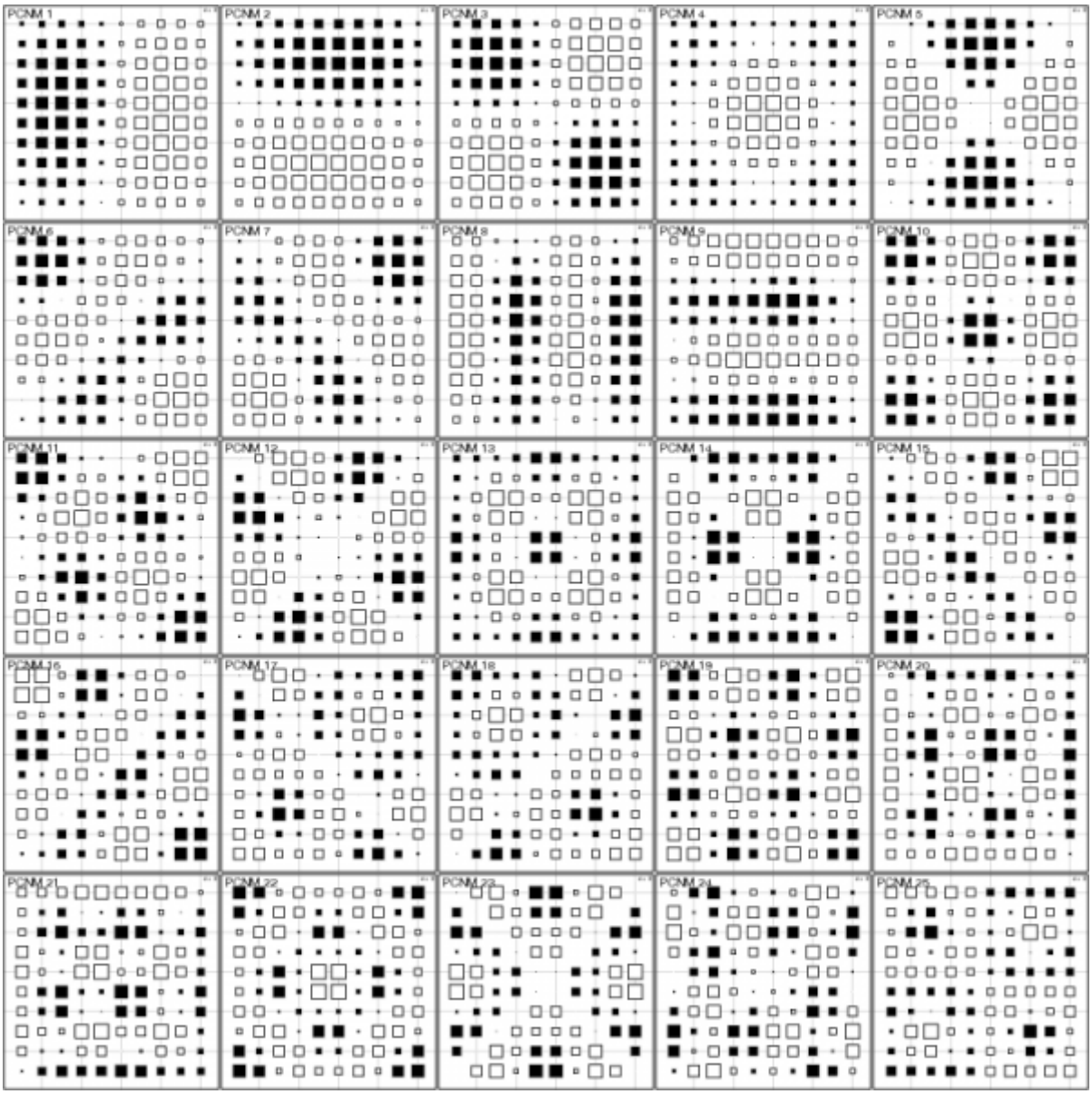
Cuidado! Aqui, no contexto de comunidades, estamos falando de agregação **de espécies**, ou seja, um conjunto formado pelas mesmas espécies e com altas abundâncias ocorrendo em parcelas espacialmente próximas. Isso é um pouco diferente da agregação **dos indivíduos de uma população de apenas uma espécie**, em algumas parcelas. No caso de comunidades, se uma agregação for identificada, isso significa que parcelas mais próximas devem apresentar maior similaridade de composição e abundância de espécies.

De forma similar ao que foi explicado no roteiro de partição para uma espécie, podemos ter uma hipótese ecológica para uma determinada estrutura espacial das espécies. Por exemplo, podemos imaginar que todas as espécies da comunidade começaram a colonizar a área a partir de um dos vértices e que algumas espécies ainda não conseguiram colonizar o vértice oposto da parcela. Dessa forma, o vértice inicial teria parcelas com composições mais similares e as parcelas do vértice oposto teriam composições mais diferentes daquelas.



Porém, poderiam existir muitas outras estruturas de distribuições espaciais, para as quais não temos hipóteses iniciais, mas que poderiam explicar a similaridade entre as parcelas independentemente das condições ambientais.

Existe um método que gera um enorme conjunto de estruturas espaciais que podem ser utilizadas para comparar com os dados reais de uma certa área. Cada estrutura espacial simulada é chamada de **PCNM (Principal Coordinates of Neighbour Matrices)** ou **dbMEM (distance-based Moran Eigenvector Maps)**. Nas figuras abaixo [5)] são mostrados 25 exemplos de diferentes estruturas espaciais possíveis para uma área quadrada dividida em 100 parcelas (como é o caso da nossa área):



Vamos entender essas figuras acima já no contexto dos nossos dados. No nosso caso, os dados reais para os quais queremos avaliar a estrutura espacial são referentes à **similaridade na composição de espécies nas parcelas**. Então, dentro de cada uma das 25 figuras (lembrando que cada uma representa uma estrutura espacial diferente, indicada por um PCNM diferente) os maiores quadrados pretos representariam as parcelas mais similares entre si e os maiores quadrados brancos seriam as parcelas mais diferentes das pretas e mais parecidas entre si. Podemos expressar este código de cores fazendo com que quadrados pretos tenham valores positivos, e os quadrados brancos valores negativos. Os tamanhos dos quadrados expressam o módulo destes valores (ou seja, quadrados pretos grande são valores altos positivos,e os brancos são os valores muito negativos).

IMPORTANTE: Note que PCNMs com números mais baixos (PCNM1, PCNM2, PCNM3) representam estruturas espaciais em escalas mais amplas, ou seja, aquelas que representam grandes manchas dividindo a área ao meio, por exemplo. Por outro lado, os PCNMs com números mais altos (PCNM24, PCNM25) indicam manchas menores espalhadas pela área, representando estruturas espaciais em escalas mais finas.

Entretanto, não vale a pena incluir nas análises todas as estruturas espaciais possíveis. É melhor fazer uma pré-seleção das estruturas mais plausíveis. Para isso é usado um procedimento chamado de ***forward selection***.

A criação dos PCNMs e a seleção dos principais PCNMs (por meio da ***forward selection***) são procedimentos muito usados em ecologia de comunidades e vale a pena conhecer como funcionam. Mas neste exercício não vamos entrar em detalhes sobre esses métodos e vamos usar uma tabela com as variáveis espaciais já trabalhadas e selecionadas para o nosso conjunto de dados.

Então, vamos pedir que acredite em nós!

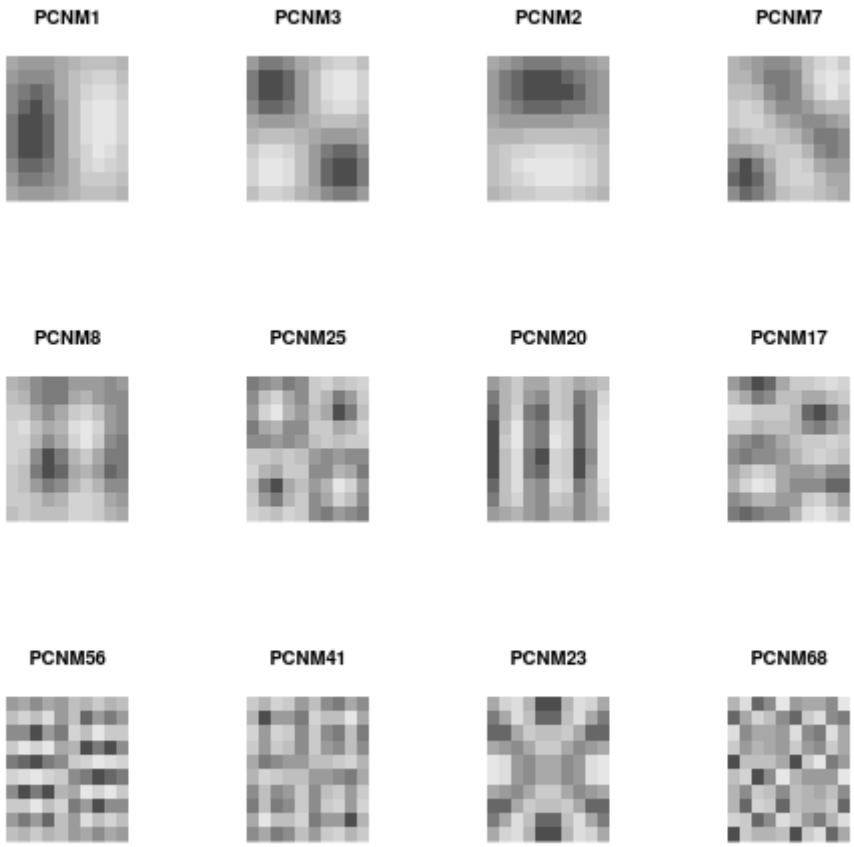
Se você acreditar, copie essa tabela para a sua pasta:

tab_coords_sel.csv

Agora, importe a tabela para o R:

```
tab.coords.sel<-read.csv2("tab_coords_sel.csv")
```

Essa tabela importada (“tab.coords.sel”) contém as informações que serão usadas para representar a estrutura espacial dos dados. Abaixo você vê o mapa dos 12 primeiros PCNMs selecionados, na mesma notação de tons de cinza dos mapas das abundâncias de espécies e variáveis de hábitat. Nestas representações, os tons expressam um certo padrão espacial, com gradientes, manchas, etc.



Calculando RDA para espécies x variáveis espaciais

Agora, aplicaremos uma RDA na qual ordenaremos as parcelas de acordo com a composição de espécies, mas condicionadas pelas estruturas espaciais. Vamos direto para os resultados numéricos:

```
(rda.espacial<-rda(sp, tab.coords.sel))
```

O valor da proporção da variação "*Constrained*" nesses resultados representa a proporção de variação na similaridade de composição de espécies entre as parcelas que é explicada pela estruturas espaciais.

O valor obtido é equivalente ao W do Roteiro de partição univariada, ou seja a fração total explicada pelas variáveis espaciais

Para salvar esse valor dentro do R, use o comando abaixo:

```
(W <- (rda.espacial$CCA$tot.chi)/(rda.espacial$tot.chi))
```

Variáveis ambientais + Estrutura espacial

Por fim, vamos aplicar uma RDA[6)] na qual ordenaremos as parcelas de acordo com a composição de espécies condicionadas pelo conjunto completo de variáveis, incluindo **tanto as variáveis ambientais quanto as estruturas espaciais**. Note que ambas as planilhas de variáveis explanatórias (ambientais e espaciais selecionadas) estão incluídas nessa análise. Dessa forma calcularemos a proporção de variação explicada por todas as variáveis disponíveis.

Calculando RDA para espécies x variáveis ambientais + espaciais

Novamente vamos direto para os resultados numéricos:

```
(rda.amb.espacial<-rda(as.matrix(sp)~ as.matrix(tab.coords.sel) +
as.matrix(amb)))
```

O valor da proporção da variação "*Constrained*" nesses resultados representa o total de variação na similaridade de composição de espécies entre as parcelas que é explicado pelo conjunto total de variáveis disponíveis.

Esse valor é equivalente ao XW ou $a+b+c$ do Roteiro de partição univariada, ou seja a fração total explicada

Para salvar esse valor dentro do R, use o comando abaixo:

```
(XW<-(rda.amb.espacial$CCA$tot.chi)/(rda.amb.espacial$tot.chi))
```

O valor da proporção da variação "*Unconstrained*" nesses resultados representa a variação não explicada pelo conjunto total de variáveis disponíveis.

Esse valor é equivalente ao d do Roteiro de partição univariada, ou seja, o Resíduo ou a variação não explicada.

Para salvar esse valor dentro do R, use o comando abaixo:

```
(d<-(rda.amb.espacial$CA$tot.chi)/(rda.amb.espacial$tot.chi))
```

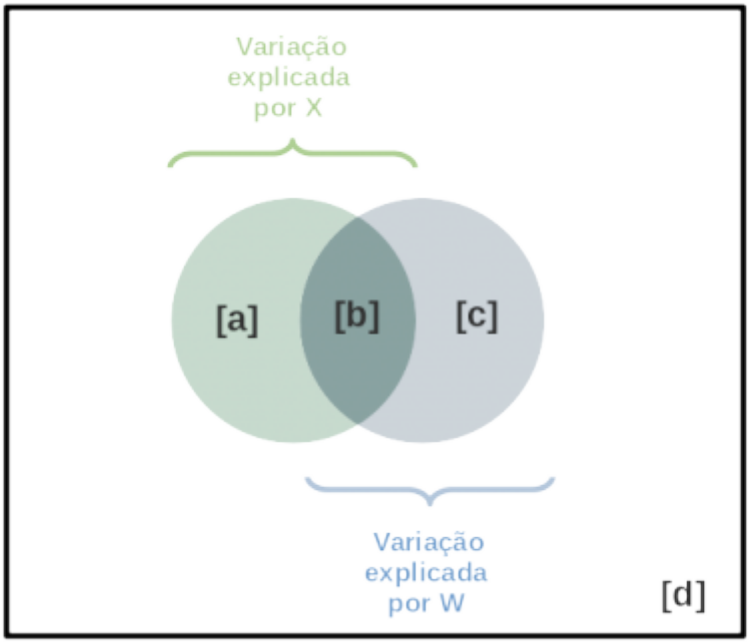
Tabela de Partição de Variação de Dados Multivariados

Agora estamos interessados em calcular qual a proporção da variação da similaridade em composição de espécies das parcelas que é explicada:

- exclusivamente pelas variáveis ambientais (a)
- exclusivamente pela estrutura espacial (c)
- pelos dois fatores em conjunto (b)

E também queremos saber a proporção de variação que não é explicada por nenhuma das variáveis utilizadas. Esse é o Resíduo (d)

Relembrando a figura que mostra toda a estrutura de variação:



Juntando as peças

Agora temos todas as peças que precisamos para calcular como a variação da similaridade na composição de espécies das parcelas pode ser particionada entre os componentes do diagrama acima:

$$ \left\{ \begin{array}{l} a + b = X\\ c + b = W\\ a + b + c = XW\\ d = 1- XW\\ b = d + X + W - 1\\ a = X - b\\ c = W - b \end{array} \right. $$

Nota: O termo $a+b+c$ pode ser obtido tanto por "XW" quanto por "$1-d$".

Já temos os valores de X, W, XW e d salvos no R. Vamos retomá-los aqui, digitando no R:

```
X
W
XW
d
```

Então, podemos agora, calcular os valores de a, b, e c. Para isso use os comandos abaixo

```
(b <- d + X + W - 1)
(a <- X - b)
(c <- W - b)
```

Pronto! Agora sua Tabela de Partição da Variação já está completa e você pode pensar em uma forma de expressar os valores de a, b, c e d graficamente. Existem formas diferentes. Crie a sua!

Nota: Lembre que o valor (b), que representa o efeito conjunto das variáveis ambientais e da estrutura espacial, revela a relação entre essas duas matrizes de dados. Valores altos para os "dois

fatores em conjunto" indicam que parcelas mais próximas espacialmente têm também variáveis ambientais mais similares, ou, em outras palavras, isso significa que **as variáveis ambientais estão espacialmente estruturadas**.

E então, como você interpreta esses resultados obtidos, no contexto de processos baseados em nicho e processos baseados em limitação de dispersão?

OPCIONAL: Fazendo a partição da variação diretamente no R

Calculamos a partição da variação passo-a-passo, a partir das diferentes RDAs, mas claro que alguma boa alma já fez uma função no R que calcula a partição da variação diretamente, a partir das planilhas base. Vamos então aplicar a função "varpart" usando as nossas três planilhas base ("sp", "amb" e "tab.coords.sel"), vamos criar um objeto ("particao") que contém os resultados dessa análise e vamos comparar com os resultados obtidos na nossa análise passo-a-passo, feita anteriormente:

```
(particao <- varpart(sp, amb , tab.coords.sel))
```

A tabela resultante mostra a variação total, decomposta em:

- $a+b+c =$ variação total
- $a+b =$ variação total devido ao ambiente ($X1$, que é igual ao X anterior)
- $b+c =$ variação total devido ao espaço ($X2$, que é igual ao W anterior)

Note que:

- $a = $ variação exclusivamente devido ao ambiente ($X1|X2$ ou $X|W$)[7)]
- $c =$ variação exclusivamente devido à estrutura espacial ($X2|X1$ ou $W|X$) [8)]

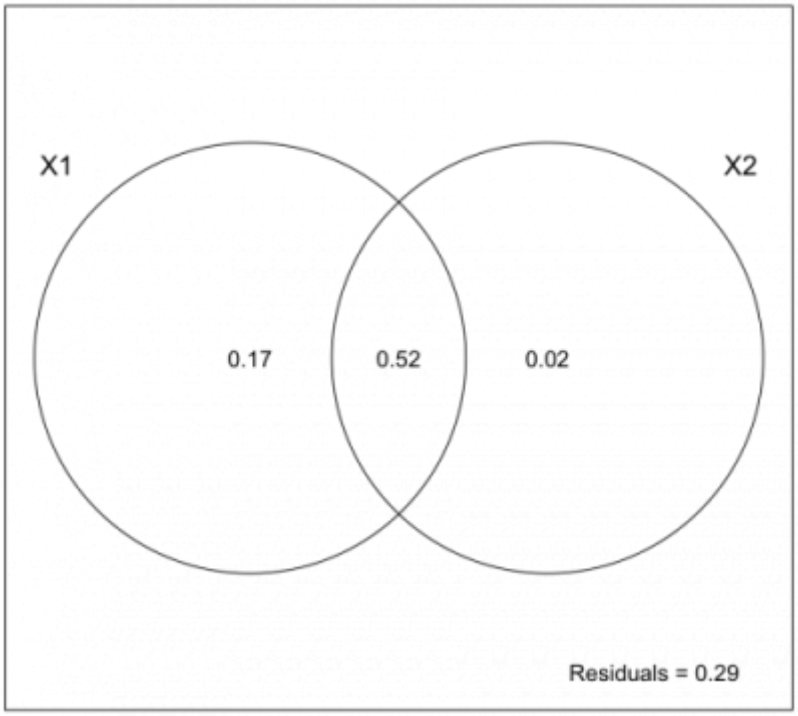
Além disso, temos também:

- $d = $ variação não explicada (resíduo)

*Obs. A Tabela de Partição obtida por essa função tem duas colunas com valores de R^2. Uma "*R.Squared*", que mostra os valores iguais aos que calculamos e outra "*Adj.R.Squared*",com valores levemente diferentes, pois utiliza valores de R^2 ajustados. Não se preocupe com isso agora.*

Um gráfico (Diagrama de Venn) que ajuda a entender essa partição:

```
plot(particao)
```



Obs.: Não leve em consideração o tamanho dos círculos nem a proporção de sobreposição entre eles, mas sim os valores numéricos indicados. O valores usados nessa figura são os de R^2 ajustado

E então, como você interpreta esses resultados obtidos, no contexto de processos determinísticos baseados em nicho e processos baseados em limitação de dispersão?

Quer saber mais?

- Borcard et al. (2011). Numerical Ecology with R.
- Legendre & Legendre (2003). Numerical Ecology.
- Sobre os PCNMs: Borcard, D. and P. Legendre. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecological Modelling 153(1-2): 51-68.
- Ótima página de Michael Palmer
- Ótima página de David Zeleny
- Tutorial do vegan
- Introdução a análise multivariada do vegan

Se você gostar de programação e quiser entender como os valores dessa prática foram produzidos, clique aqui e veja os códigos em R que usamos.

1)

Esses eixos derivados são definidos a partir da matriz de correlações entre as espécies, ou seja, daqueles valores que estavam na parte direita e acima da diagonal da figura de correlações que

visualizamos anteriormente. A partir dessa matriz de correlação serão obtidos os *autovalores* e os *autovetores* que definirão esses novos eixos derivados. **Cuidado!** A matriz da qual os autovalores são obtidos aqui na PCA é muito diferente daquelas matrizes da Análise de Projeção Matricial que aprendemos no módulo de Dinâmica de Populações. Apesar da matemática por trás do cálculo dos autovalores ser a mesma, as interpretações são diferentes!

2)

O número de eixos que podem ser obtidos é igual ao número de variáveis originais (no nosso caso seriam 10), porém, esperamos que uma PCA consiga explicar a maior parte da variação em poucos eixos.

3)

Não se preocupe com o termo “scale=TRUE” dentro dos parênteses. É uma particularidade do pacote *vegan* para esse tipo de análise.

4)

Note que estamos usando a mesma função “rda”, mas agora temos **duas planilhas** dentro dos parênteses (“sp” e “amb”

5)

obtida na página do *California Soil Resource Lab* da University of California, Davis

6)

levemente modificada a forma de escrever, mas é basicamente a mesma coisa

7)

ou seja, variação que pode ser atribuída ao ambiente, descontado o efeito da estrutura espacial.

8)

ou seja, variação que pode ser atribuída à estrutura espacial, descontado o efeito das variáveis ambientais.

From:
http://ecovirtual.ib.usp.br/ -

Permanent link:
http://ecovirtual.ib.usp.br/doku.php?id=ecovirt:roteiro:particao:particao_multivar

Last update: **2023/11/20 06:10**