

Análises Multivariadas

(Ordenação de objetos)

Bioestatística | 1º semestre 2021

Pós-Graduações stricto sensu

Faculdade de Ciências Biológicas e Ambientais

Universidade Federal da Grande Dourados



Quantas
espécies?





Imagem de [Gordon Johnson](#) por [Pixabay](#)



Imagem de [Gordon Johnson](#) por [Pixabay](#)



Número de espécies



Número de espécies

Variável UNIdimensional



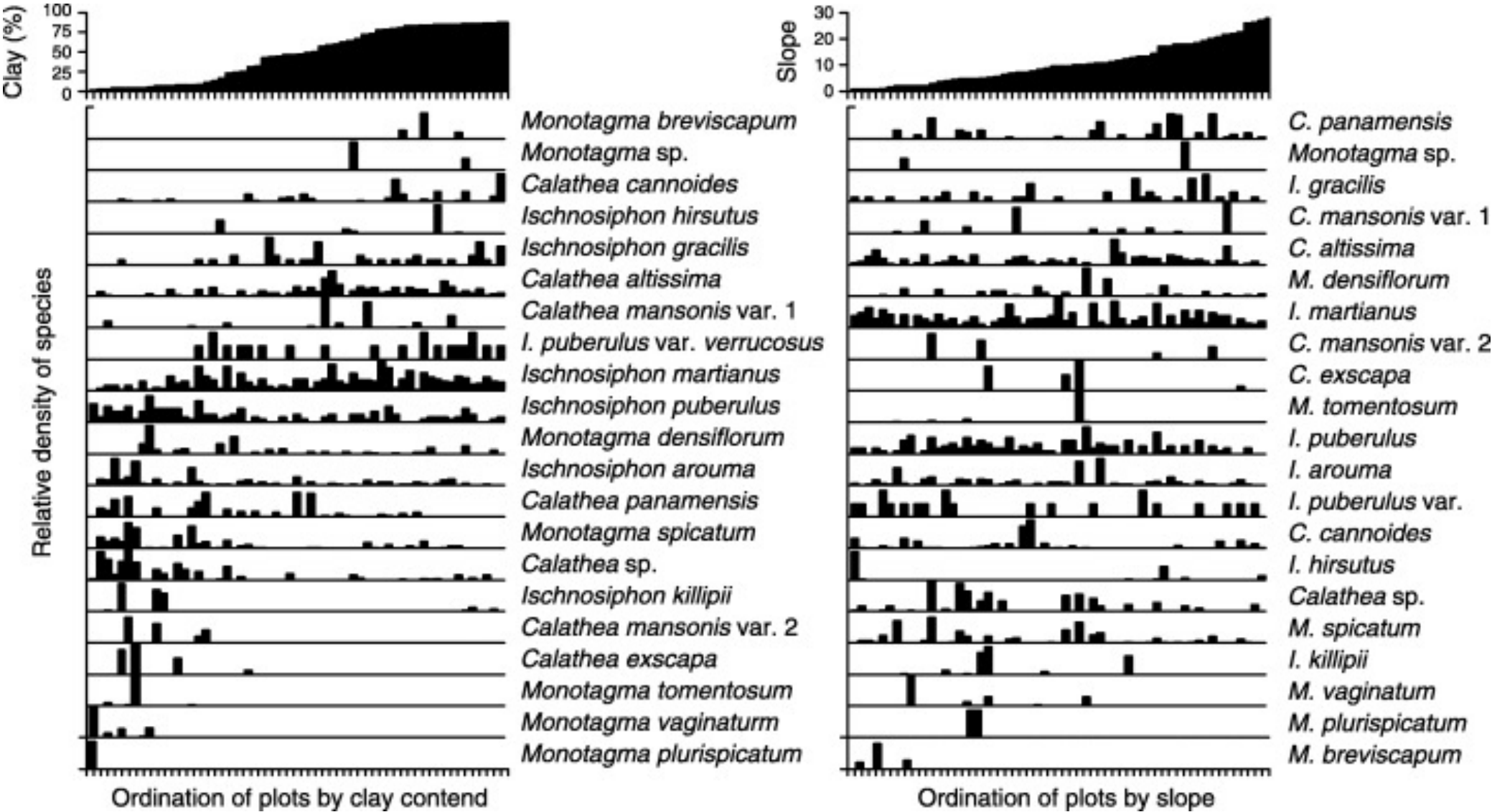
Composição de espécies



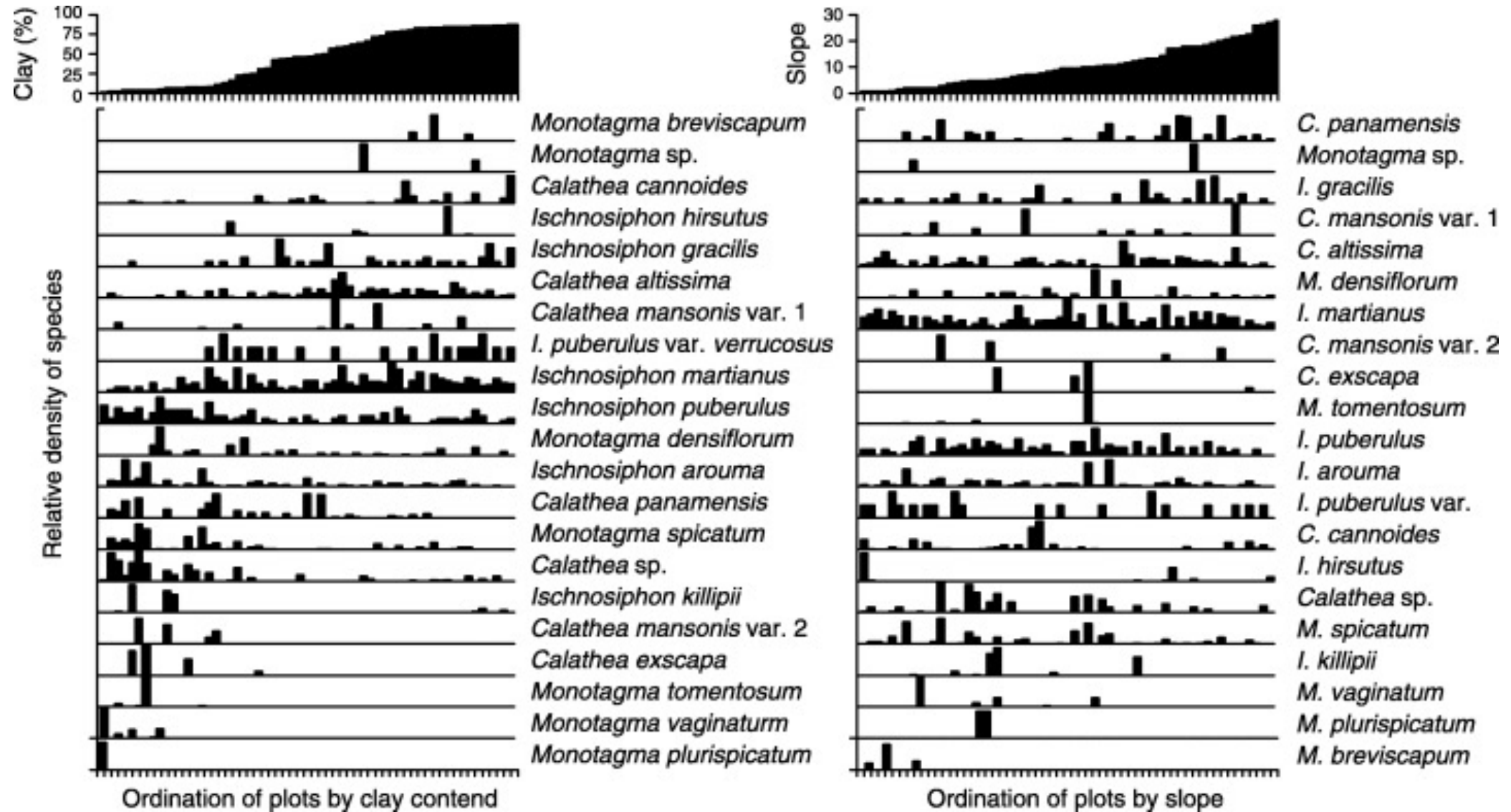
Composição de espécies

Variável
MULTIdimensional

Mesoscale distribution patterns of Amazonian understorey herbs in relation to topography, soil and watersheds



Ordenação direta



Mas sem qualquer variável externa,
como a composição de espécies varia?



Mas sem qualquer variável externa,
como a composição de espécies varia?

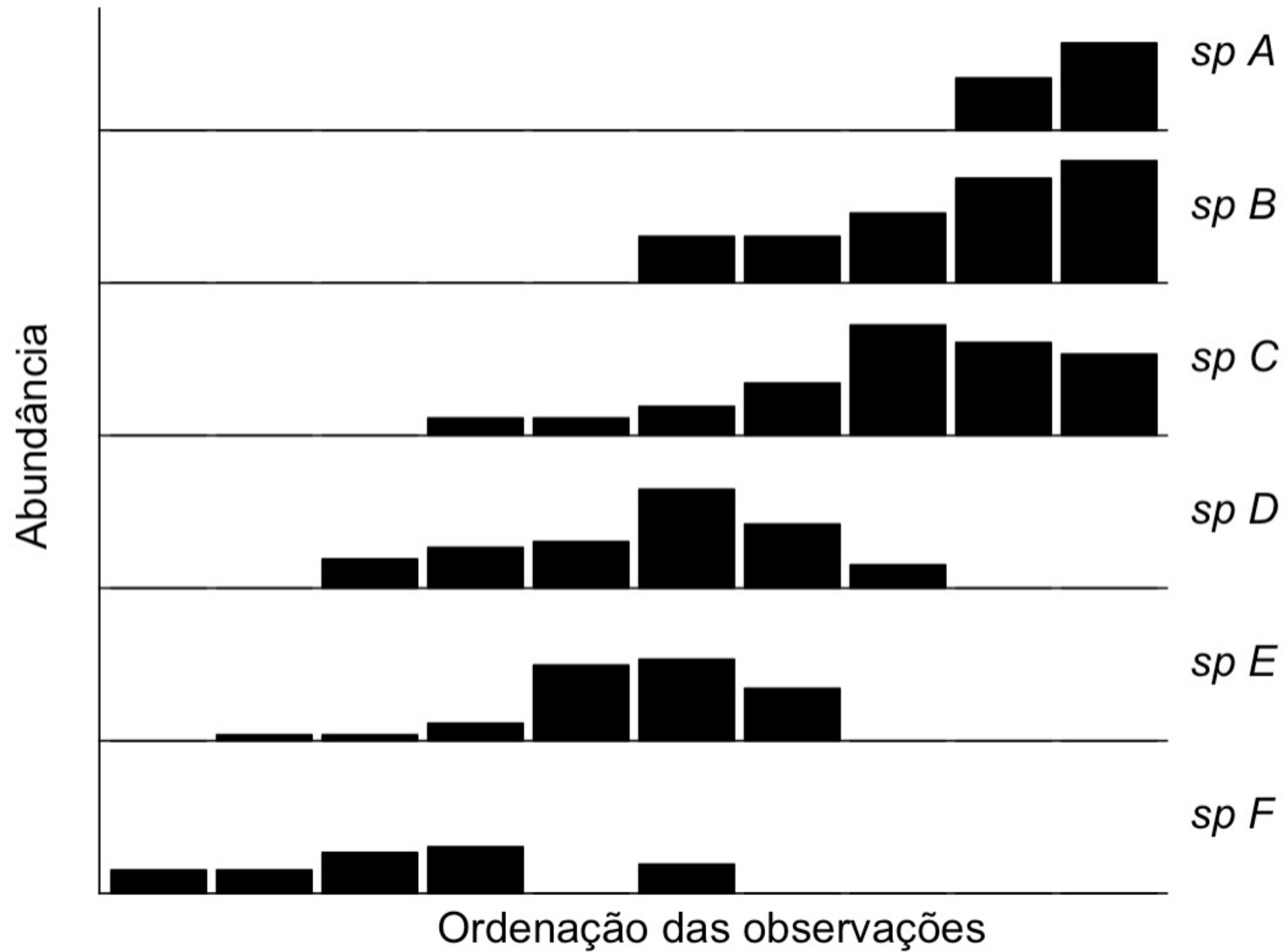


Qual o principal padrão?

	sp_A	sp_B	sp_C	sp_D	sp_E	sp_F
1	0	0	3	8	13	0
2	0	0	0	0	1	4
3	0	12	19	4	0	0
4	0	8	5	17	14	5
5	15	21	14	0	0	0
6	0	0	3	7	3	8
7	9	18	16	0	0	0
8	0	0	0	0	0	4
9	0	0	0	5	1	7
10	0	8	9	11	9	0

	sp_A	sp_B	sp_C	sp_D	sp_E	sp_F
5	15	21	14	0	0	0
7	9	18	16	0	0	0
3	0	12	19	4	0	0
10	0	8	9	11	9	0
4	0	8	5	17	14	5
1	0	0	3	8	13	0
6	0	0	3	7	3	8
9	0	0	0	5	1	7
2	0	0	0	0	1	4
8	0	0	0	0	0	4

	sp_A	sp_B	sp_C	sp_D	sp_E	sp_F
5	15	21	14	0	0	0
7	9	18	16	0	0	0
3	0	12	19	4	0	0
10	0	8	9	11	9	0
4	0	8	5	17	14	5
1	0	0	3	8	13	0
6	0	0	3	7	3	8
9	0	0	0	5	1	7
2	0	0	0	0	1	4
8	0	0	0	0	0	4



	sp1	sp2
loc_A	1	5
loc_B	5	4

	sp1	sp2
loc_A	1	5
loc_B	5	4

Diferenças 4 1

	sp1	sp2
loc_A	1	5
loc_B	5	4

Diferenças

4

1

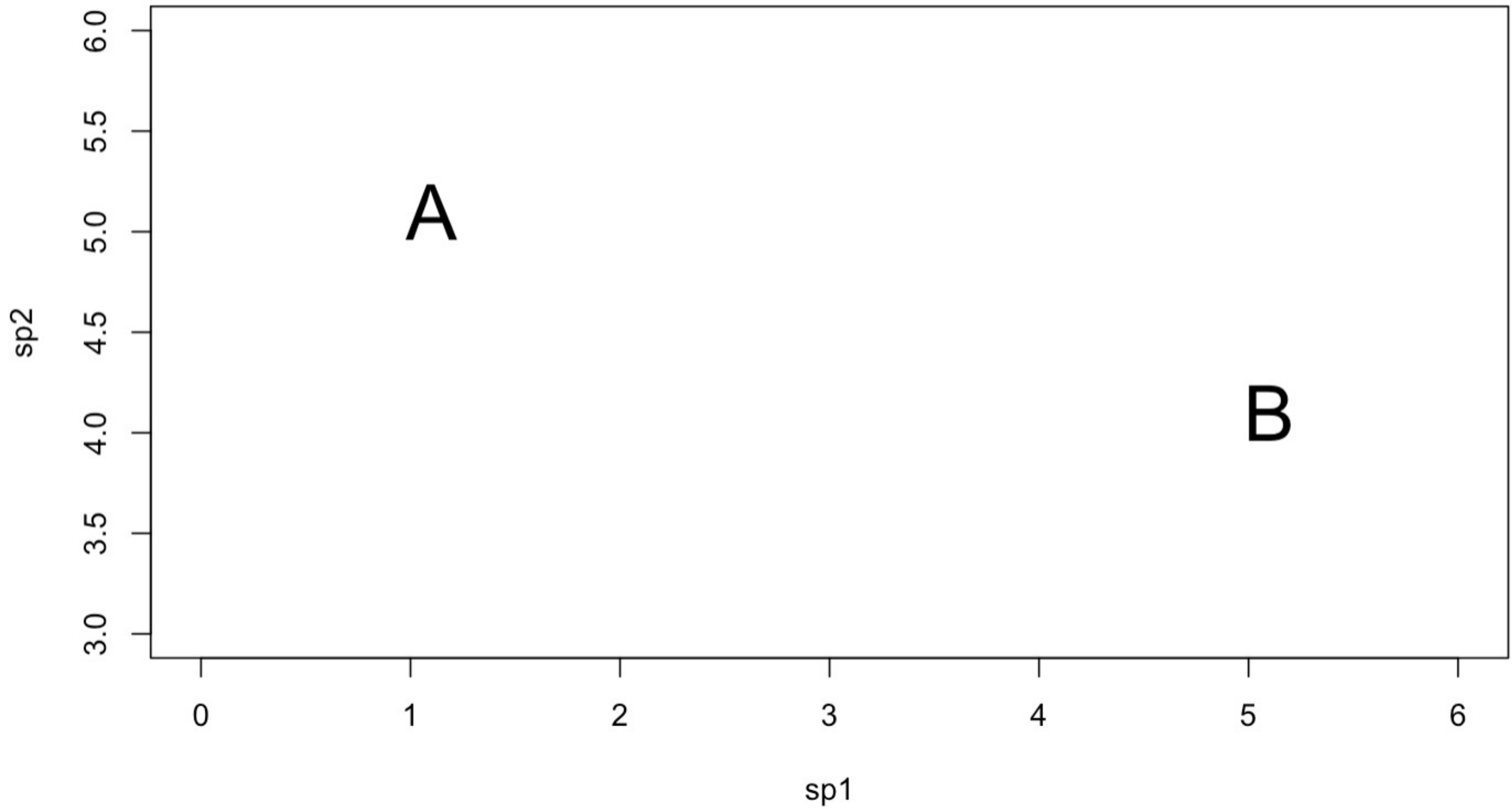
4

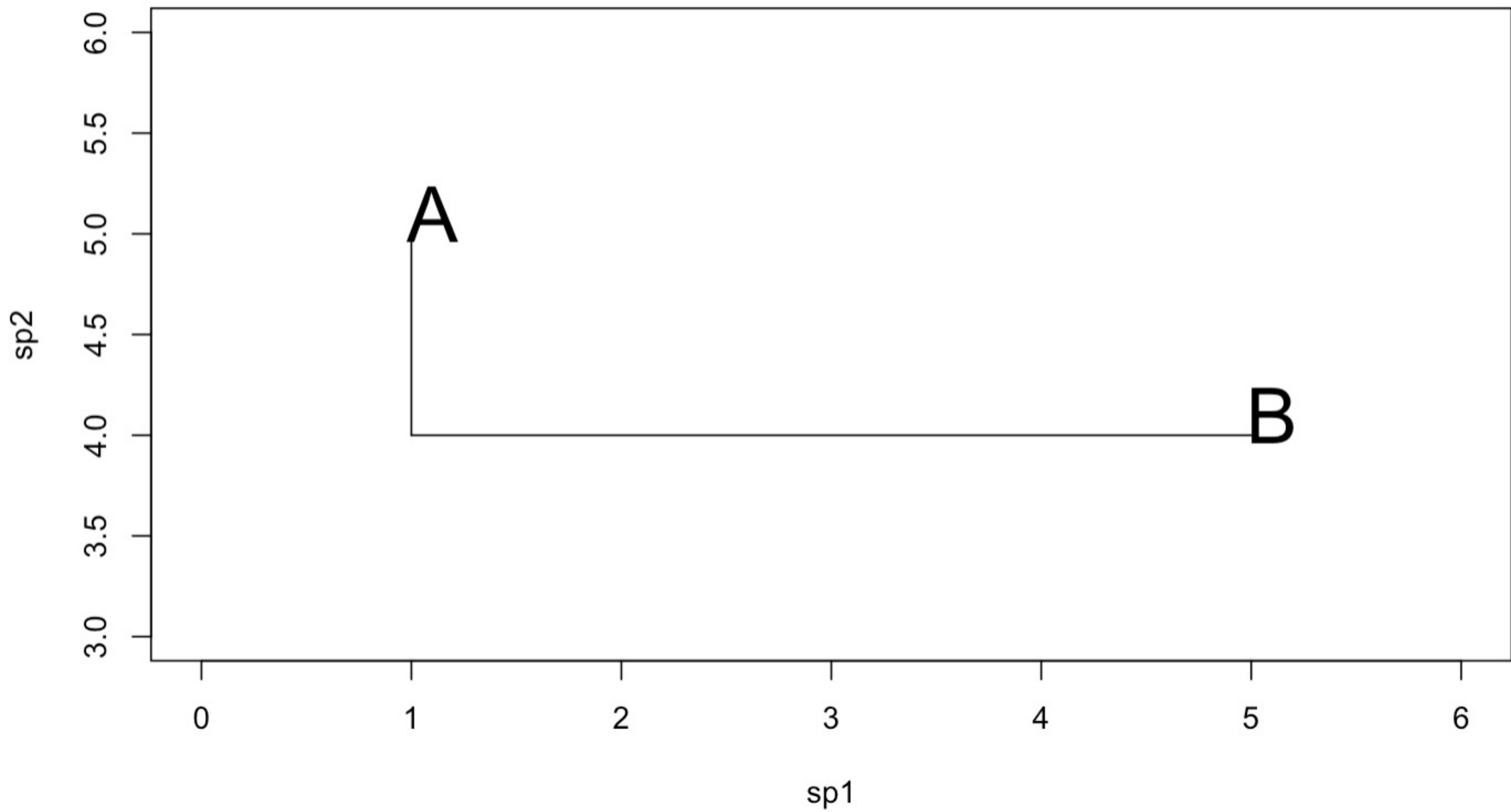
+

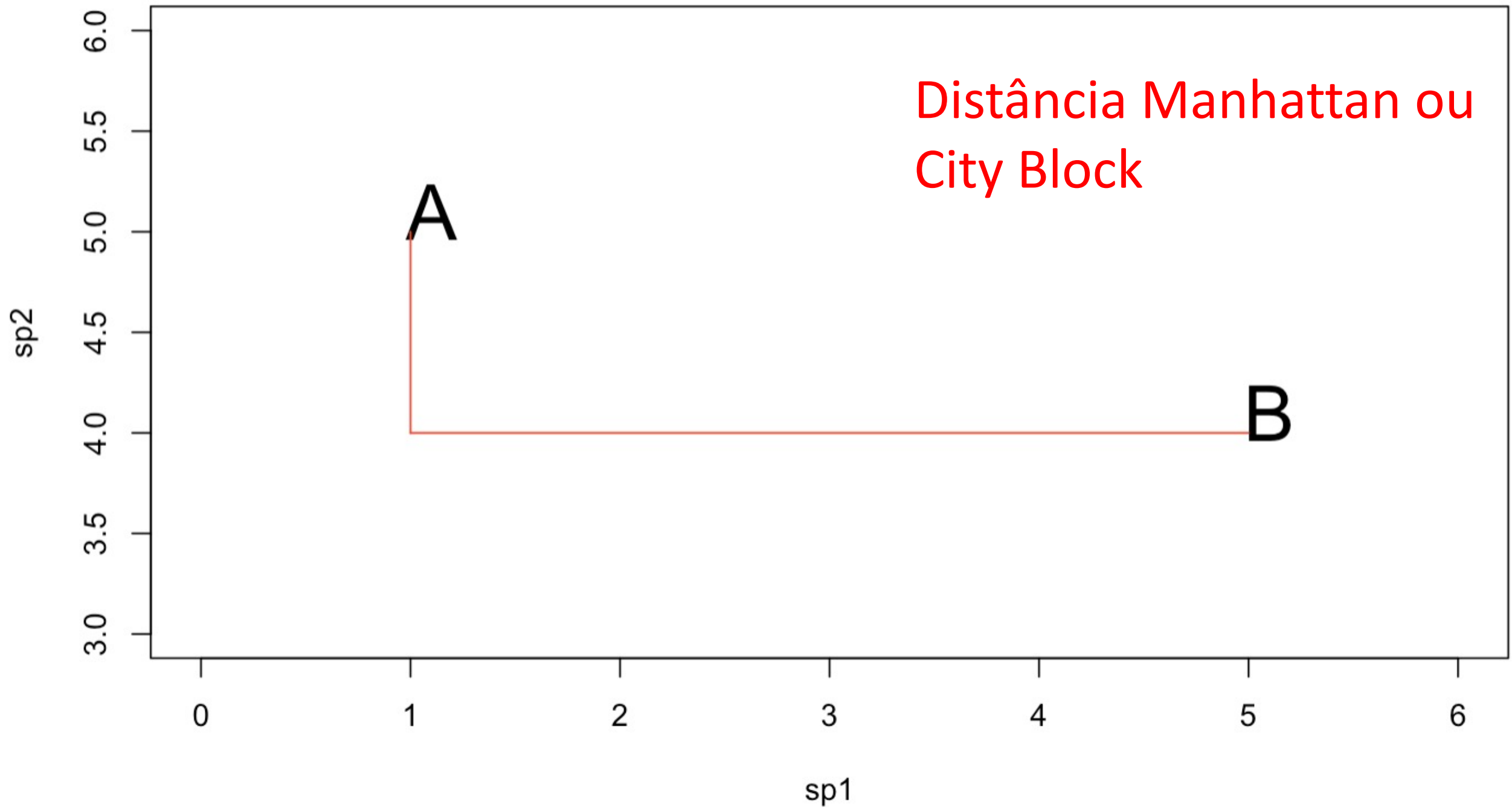
1

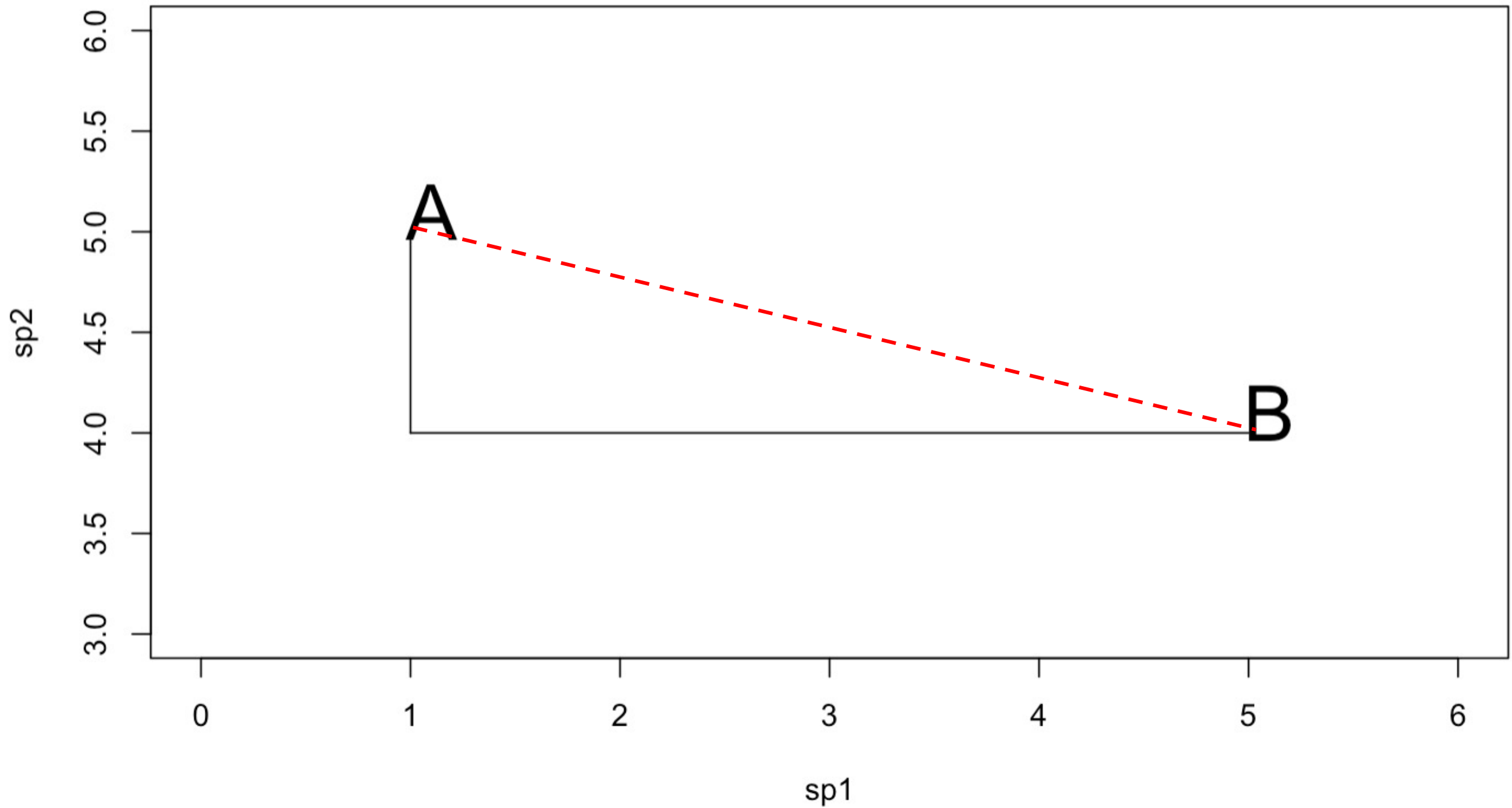
=

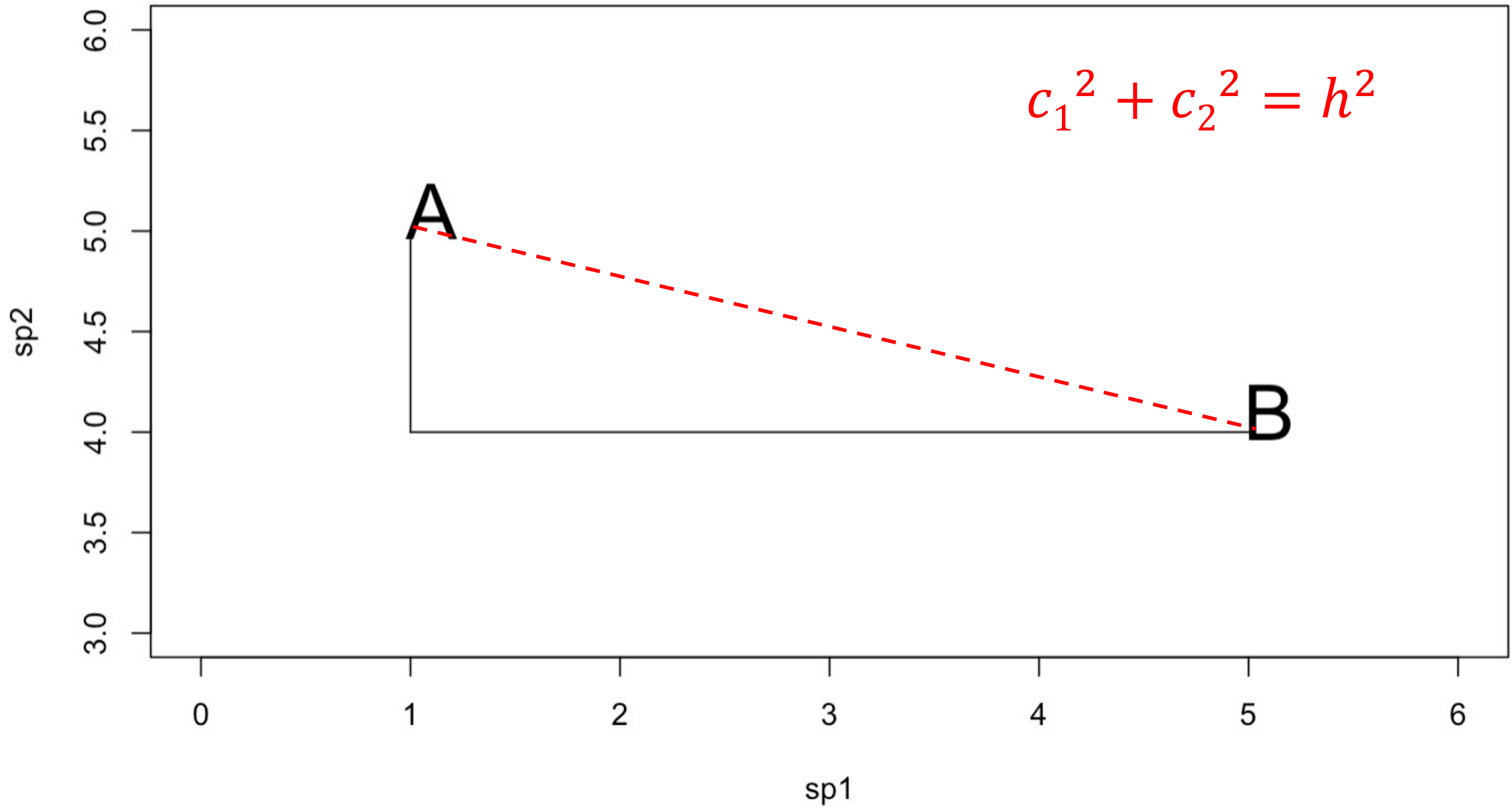
5

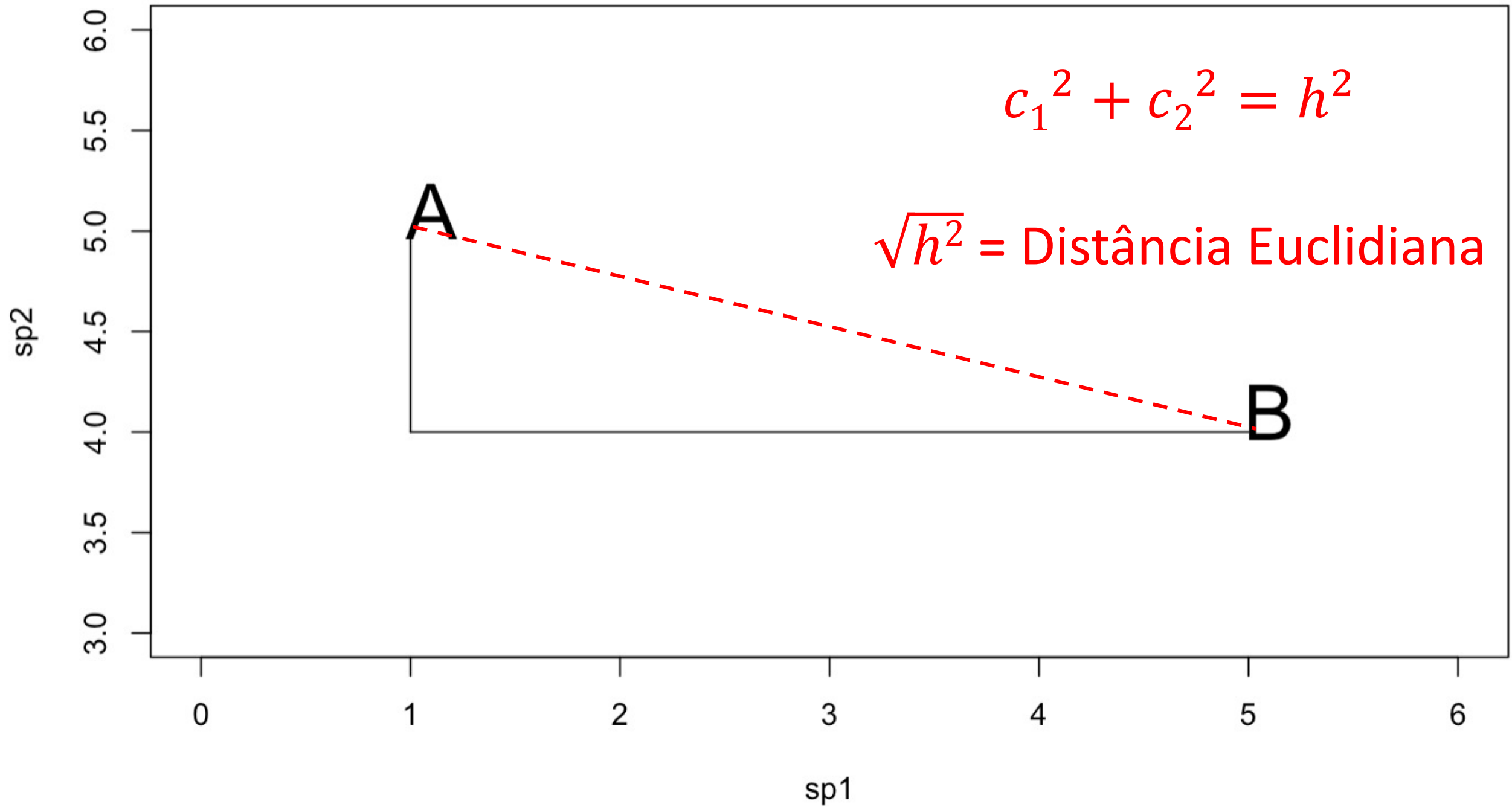












euclidean $d[jk] = \text{sqrt}(\text{sum}((x[ij]-x[ik])^2))$
 binary: $\text{sqrt}(A+B-2*J)$

manhattan $d[jk] = \text{sum}(\text{abs}(x[ij] - x[ik]))$
 binary: $A+B-2*J$

gower $d[jk] = (1/M) \text{sum}(\text{abs}(x[ij]-x[ik])/(max(x[i])-min(x[i])))$
 binary: $(A+B-2*J)/M$
 where M is the number of columns (excluding missing values)

altGower $d[jk] = (1/NZ) \text{sum}(\text{abs}(x[ij] - x[ik]))$
 where NZ is the number of non-zero columns excluding double-zeros (Anderson et al. 2006).
 binary: $(A+B-2*J)/(A+B-J)$

canberra $d[jk] = (1/NZ) \text{sum} (\text{abs}(x[ij]-x[ik])/(abs(x[ij])+abs(x[ik])))$
 where NZ is the number of non-zero entries.
 binary: $(A+B-2*J)/(A+B-J)$

clark $d[jk] = \text{sqrt} ((1/NZ) \text{sum} (((x[ij]-x[ik])/(x[ij]+x[ik]))^2))$
 where NZ is the number of non-zero entries.
 binary: $(A+B-2*J)/(A+B-J)$

bray $d[jk] = (\text{sum} \text{abs}(x[ij]-x[ik]))/(\text{sum} (x[ij]+x[ik]))$
 binary: $(A+B-2*J)/(A+B)$

kulczynski $d[jk] 1 - 0.5*(\text{sum}(\text{min}(x[ij],x[ik]))/(\text{sum} x[ij]) + \text{sum}(\text{min}(x[ij],x[ik]))/(\text{sum} x[ik]))$
 binary: $1-(J/A + J/B)/2$

vegan::vegdist

euclidean $d[jk] = \sqrt{\sum (x[ij] - x[ik])^2}$
binary: $\sqrt{A + B - 2 * J}$

manhattan $d[jk] = \sum (abs(x[ij] - x[ik]))$
binary: $A + B - 2 * J$

gower $d[jk] = (1/M) \sum (abs(x[ij] - x[ik]) / (max(x[i]) - min(x[i])))$
binary: $(A + B - 2 * J) / M$
where M is the number of columns (excluding missing values)

altGower $d[jk] = (1/NZ) \sum (abs(x[ij] - x[ik]))$
where NZ is the number of non-zero columns excluding double-zeros (Anderson et al. 2006).
binary: $(A + B - 2 * J) / (A + B - J)$

canberra $d[jk] = (1/NZ) \sum (abs(x[ij] - x[ik]) / (abs(x[ij]) + abs(x[ik])))$
where NZ is the number of non-zero entries.
binary: $(A + B - 2 * J) / (A + B - J)$

clark $d[jk] = \sqrt{(1/NZ) \sum (((x[ij] - x[ik]) / (x[ij] + x[ik]))^2)}$
where NZ is the number of non-zero entries.
binary: $(A + B - 2 * J) / (A + B - J)$

bray $d[jk] = (\sum abs(x[ij] - x[ik])) / (\sum (x[ij] + x[ik]))$
binary: $(A + B - 2 * J) / (A + B)$

kulczynski $d[jk] = 1 - 0.5 * (\sum (min(x[ij], x[ik])) / (\sum x[ij]) + \sum (min(x[ij], x[ik])) / (\sum x[ik]))$
binary: $1 - (J/A + J/B) / 2$

	sp_A	sp_B	sp_C	sp_D	sp_E	sp_F
1	0	0	3	8	13	0
2	0	0	0	0	1	4
3	0	12	19	4	0	0
4	0	8	5	17	14	5
5	15	21	14	0	0	0
6	0	0	3	7	3	8
7	9	18	16	0	0	0
8	0	0	0	0	0	4
9	0	0	0	5	1	7
10	0	8	9	11	9	0

```
> vegdist(plantas, "man")
```

	1	2	3	4	5	6	7	8	9
2	27								
3	45	40							
4	25	44	50						
5	68	55	33	73					
6	19	16	42	34	65				
7	61	48	22	66	11	58			
8	28	1	39	45	54	17	47		
9	25	8	40	40	63	8	56	9	
10	21	40	30	20	53	32	46	41	38

```
> vegdist(plantas, "man")
```

	1	2	3	4	5	6	7	8	9
1									
2	27								
3	45	40							
4	25	44	50						
5	68	55	33	73					
6	19	16	42	34	65				
7	61	48	22	66	11	58			
8	28	1	39	45	54	17	47		
9	25	8	40	40	63	8	56	9	
10	21	40	30	20	53	32	46	41	38



$$N_dist = n(n - 1) / 2$$



	1	2	3	4	5	6	7	8	9
2	27								
3	45	40							
4	25	44	50						
5	68	55	33	73					
6	19	16	42	34	65				
7	61	48	22	66	11	58			
8	28	1	39	45	54	17	47		
9	25	8	40	40	63	8	56	9	
10	21	40	30	20	53	32	46	41	38

?

	sp_A	sp_B	sp_C	sp_D	sp_E	sp_F
5	15	21	14	0	0	0
7	9	18	16	0	0	0
3	0	12	19	4	0	0
10	0	8	9	11	9	0
4	0	8	5	17	14	5
1	0	0	3	8	13	0
6	0	0	3	7	3	8
9	0	0	0	5	1	7
2	0	0	0	0	1	4
8	0	0	0	0	0	4

	1	2	3	4	5	6	7	8	9		sp_A	sp_B	sp_C	sp_D	sp_E	sp_F
2	27									5	15	21	14	0	0	0
3	45	40								7	9	18	16	0	0	0
4	25	44	50							3	0	12	19	4	0	0
5	68	55	33	73						10	0	8	9	11	9	0
6	19	16	42	34	65					4	0	8	5	17	14	5
7	61	48	22	66	11	58				1	0	0	3	8	13	0
8	28	1	39	45	54	17	47			6	0	0	3	7	3	8
9	25	8	40	40	63	8	56	9		9	0	0	0	5	1	7
10	21	40	30	20	53	32	46	41	38	2	0	0	0	0	1	4
										8	0	0	0	0	0	4

Iteração ou
Álgebra de matrizes

NMDS – Nonmetric Multidimensional Scaling

NMDS – Nonmetric Multidimensional Scaling

Método iterativo de ordenação em espaço reduzido:

Ordena os objetos em poucas dimensões tentando preservar as distâncias originais

NMDS – Nonmetric Multidimensional Scaling

Método iterativo de ordenação em espaço reduzido:

Ordena os objetos em poucas dimensões tentando preservar as distâncias originais

Ordene A, B e C em uma dimensão tentando preservar essas distâncias

A – B: 5

A – C: 10

B – C: 5

NMDS – Nonmetric Multidimensional Scaling

Método iterativo de ordenação em espaço reduzido:

Ordena os objetos em poucas dimensões tentando preservar as distâncias originais

Ordene A, B e C em uma dimensão tentando preservar essas distâncias

A-----B-----C

A – B: 5

A – C: 10

B – C: 5

NMDS – Nonmetric Multidimensional Scaling

Método iterativo de ordenação em espaço reduzido:

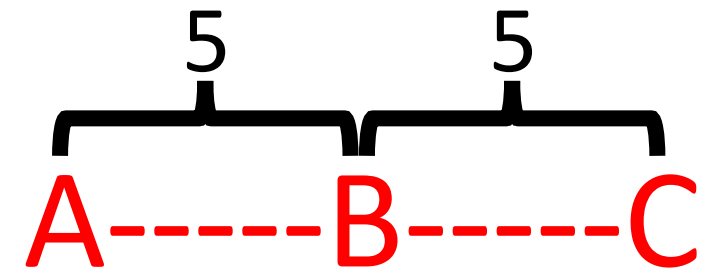
Ordena os objetos em poucas dimensões tentando preservar as distâncias originais

Ordene A, B e C em uma dimensão tentando preservar essas distâncias

A – B: 5

A – C: 10

B – C: 5



NMDS – Nonmetric Multidimensional Scaling

Método iterativo de ordenação em espaço reduzido:

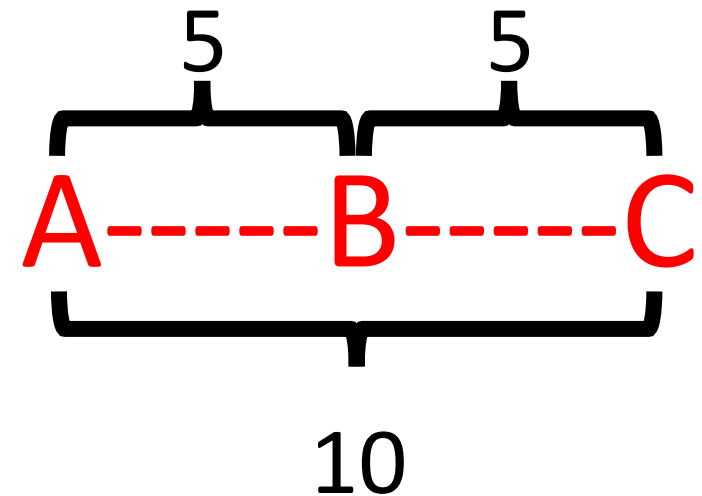
Ordena os objetos em poucas dimensões tentando preservar as distâncias originais

Ordene A, B e C em uma dimensão tentando preservar essas distâncias

A – B: 5

A – C: 10

B – C: 5



Ordene A, B e C em uma dimensão
tentando preservar essas distâncias

$$A - B: 5$$

$$A - C: 10$$

$$B - C: 5$$

Ordene A, B e C em uma dimensão
tentando preservar essas distâncias

A – B: 5

A – C: 10

B – C: 5

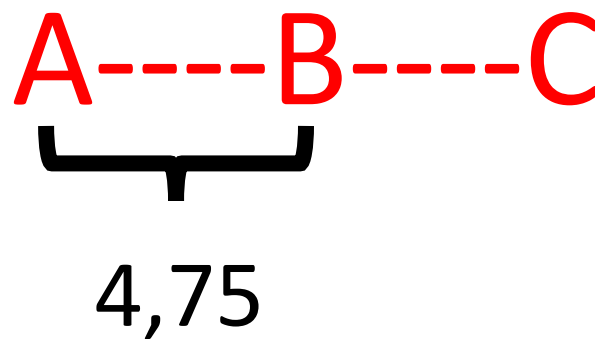
A-----B-----C

Ordene A, B e C em uma dimensão
tentando preservar essas distâncias

A – B: 5

A – C: 10

B – C: 5



Ordene A, B e C em uma dimensão
tentando preservar essas distâncias

$$A - B: 5$$

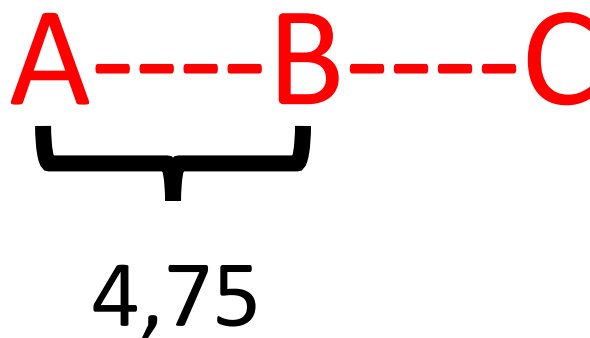
$$A - C: 10$$

$$B - C: 5$$

$$A - B: 4,75$$

$$A - C: 9,5$$

$$B - C: 4,75$$



Ordene A, B e C em uma dimensão
tentando preservar essas distâncias

A – B: 5

A – C: ~~10~~ 9

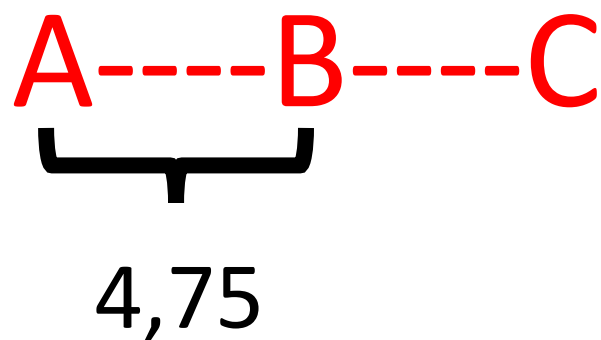
B – C: 5

A diferença entre as distâncias
originais e aquelas obtidas pela
ordenação é uma medida de
STRESS (dado entre 0 e 1)

A – B: 4,75

A – C: 9,5

B – C: 4,75



Com R

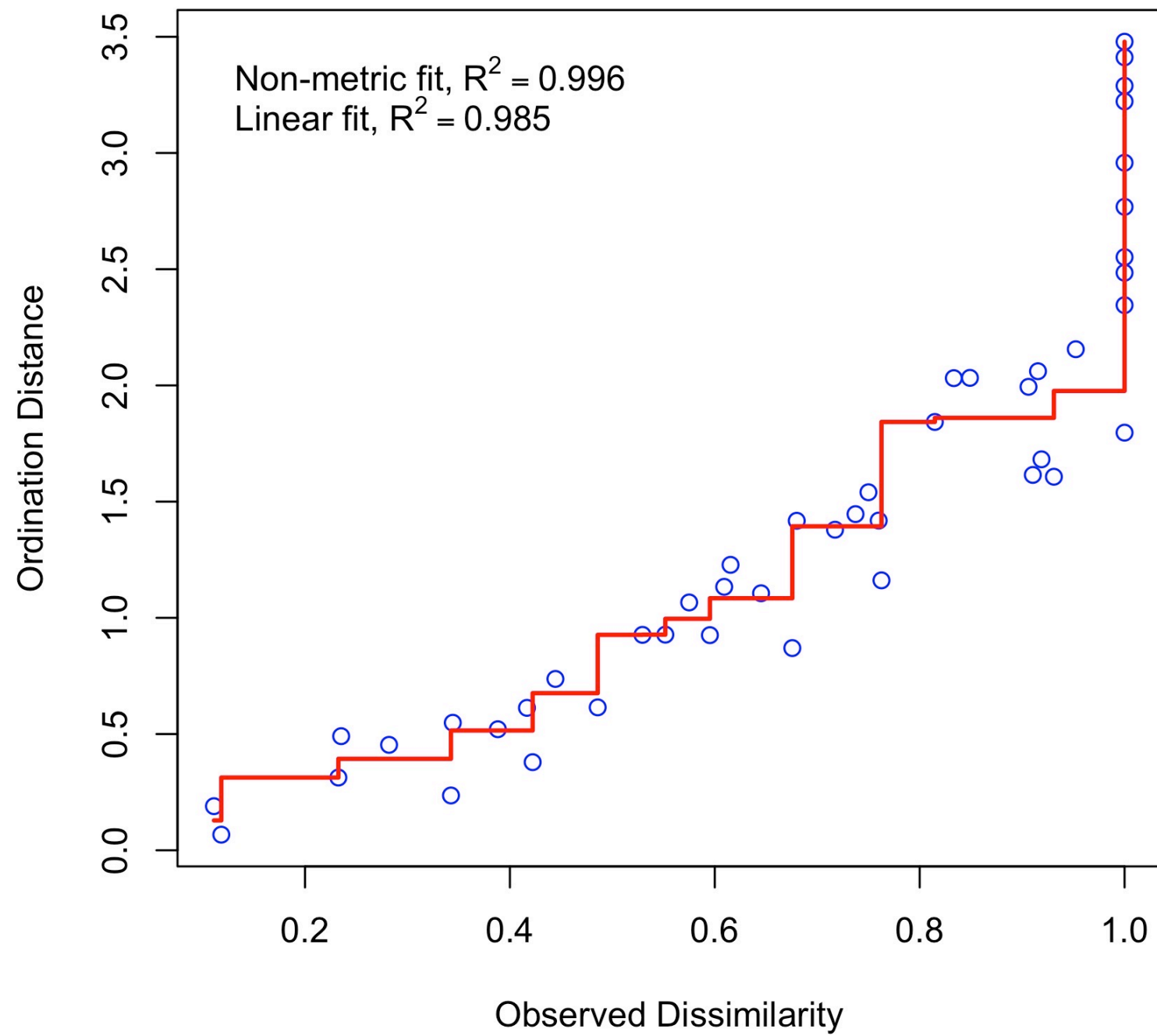
O pacote 'vegan' traz diversas funcionalidades para lidarmos com ordenações. Para fazer NMDS usamos a função 'metaMDS()'.

Funciona assim:

```
metaMDS(comm = plantas, distance = "bray", k = 1, autotransform = F)
```

Essa ordenação em 1 dimensão chegou ao

Stress: 0.06471916



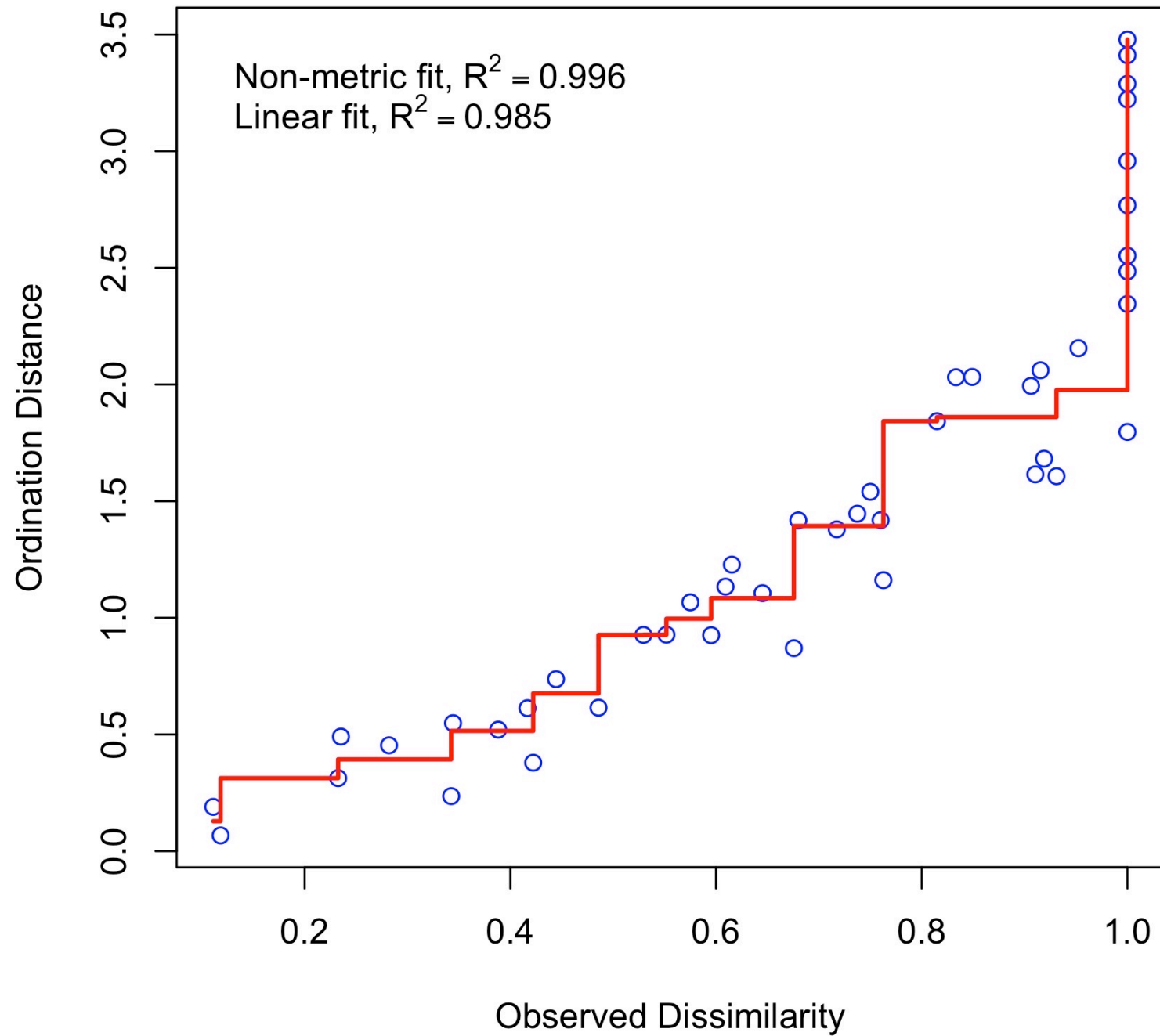


Diagrama de Shepard

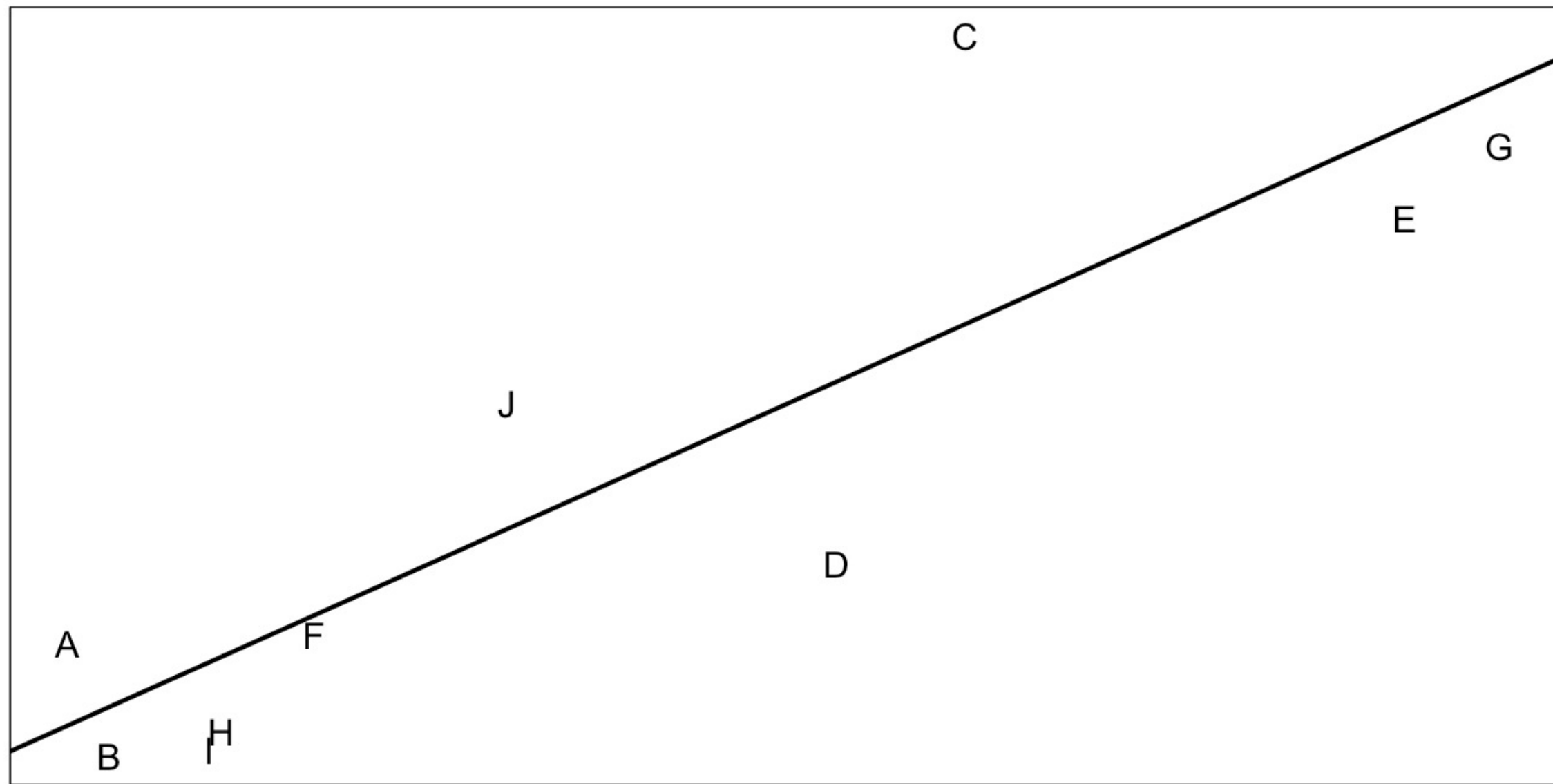
`vegan::stressplot`

PCoA – Principal Coordinate Analysis

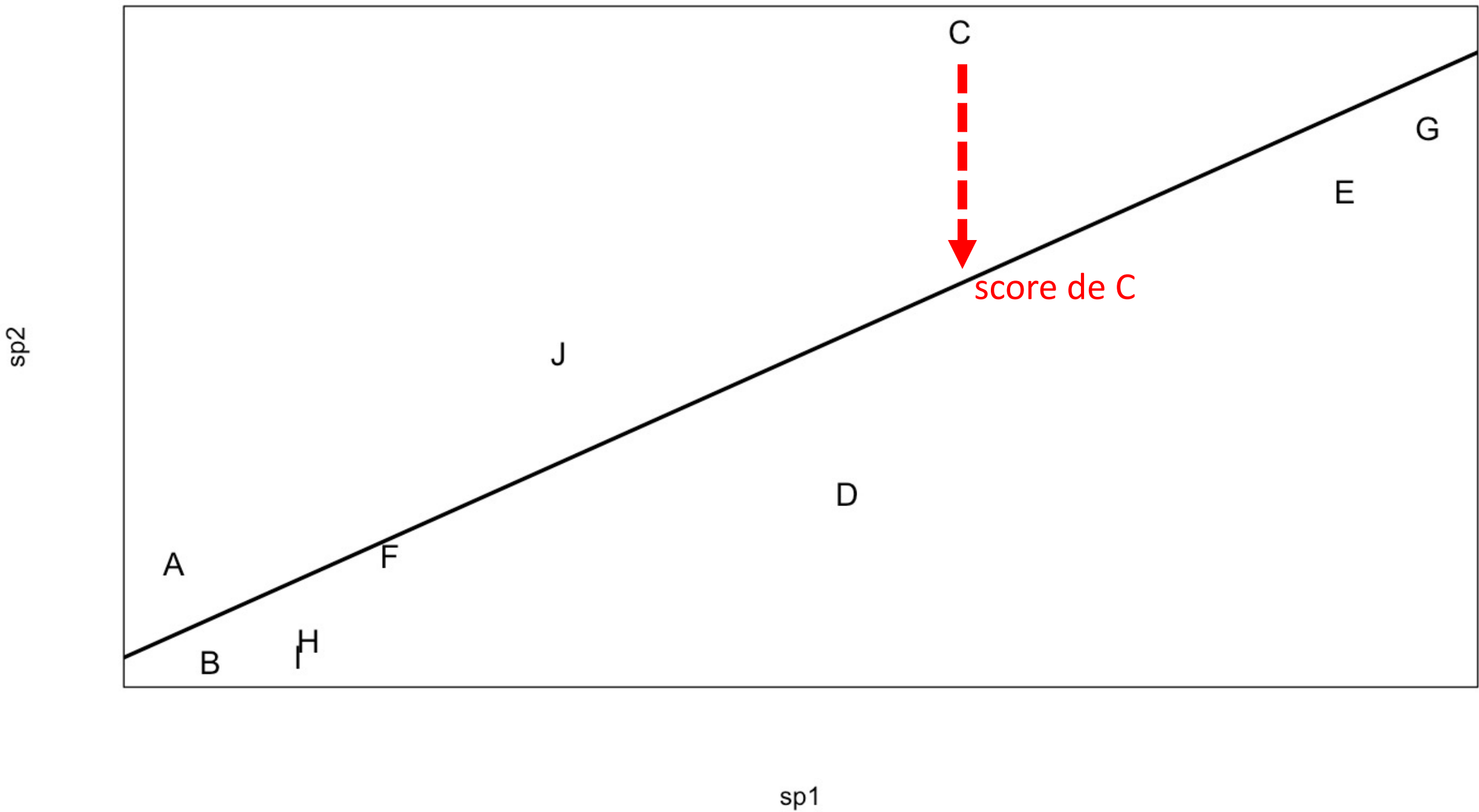
Método de ordenação com extração de eixos (“eigen”) por álgebra de matrizes:

Extraí eixos de regressão diretamente do espaço multidimensional

sp2



sp1



Com R

Para fazer PCoA usamos a função 'cmdscale()'.
Funciona assim:

```
cmdscale(d = vegdist(plantas, "bray"))
```

As análises 'eigen', tal como a PCoA, extraem todos os eixos de forma a recuperar 100% da variância dos dados no espaço multidimensional.

Cada eixo recupera uma parte da variância que é proporcional ao seu 'eigen value' ('autovalor'), que é seu comprimento.

As análises 'eigen', tal como a PCoA, extraem todos os eixos de forma a recuperar 100% da variância dos dados no espaço multidimensional.

Para ver os 'scores' (que são valores arbitrários que definem as posições dos objetos projetadas nos eixos) usamos:

```
scores(pcoa)
```


Dim1

site1	-0.02483937
site2	0.53054496
site3	-0.40633323
site4	-0.07460112
site5	-0.46020661
site6	0.24700393
site7	-0.47335545
site8	0.52317205
site9	0.41975027
site10	-0.28113542

ord

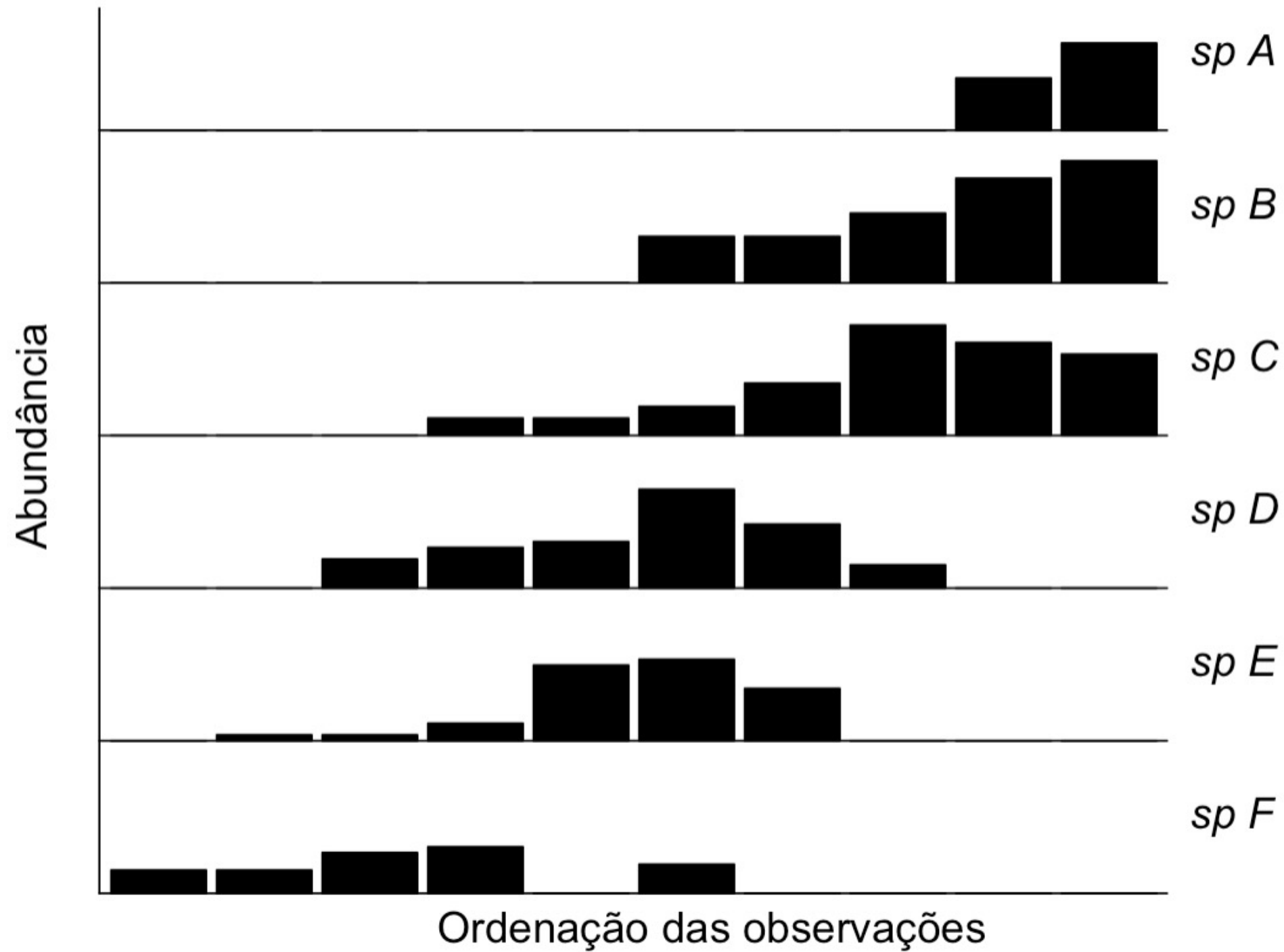
site7	-0.47335545
site5	-0.46020661
site3	-0.40633323
site10	-0.28113542
site4	-0.07460112
site1	-0.02483937
site6	0.24700393
site9	0.41975027
site8	0.52317205
site2	0.53054496

Com NMDS podemos usar o mesmo para obter os scores:

```
scores(nmdds)
NMDS1
site 1 -0.05896566
site 2 -1.66602098
site 3  1.10231332
site 4  0.17676708
site 5  1.62301004
site 6 -0.43803366
site 7  1.55589660
site 8 -1.85569908
site 9 -0.92889345
site 10 0.48962579
```

	ord_nmds
5	1.62301004
7	1.55589660
3	1.10231332
10	0.48962579
4	0.17676708
1	-0.05896566
6	-0.43803366
9	-0.92889345
2	-1.66602098
8	-1.85569908

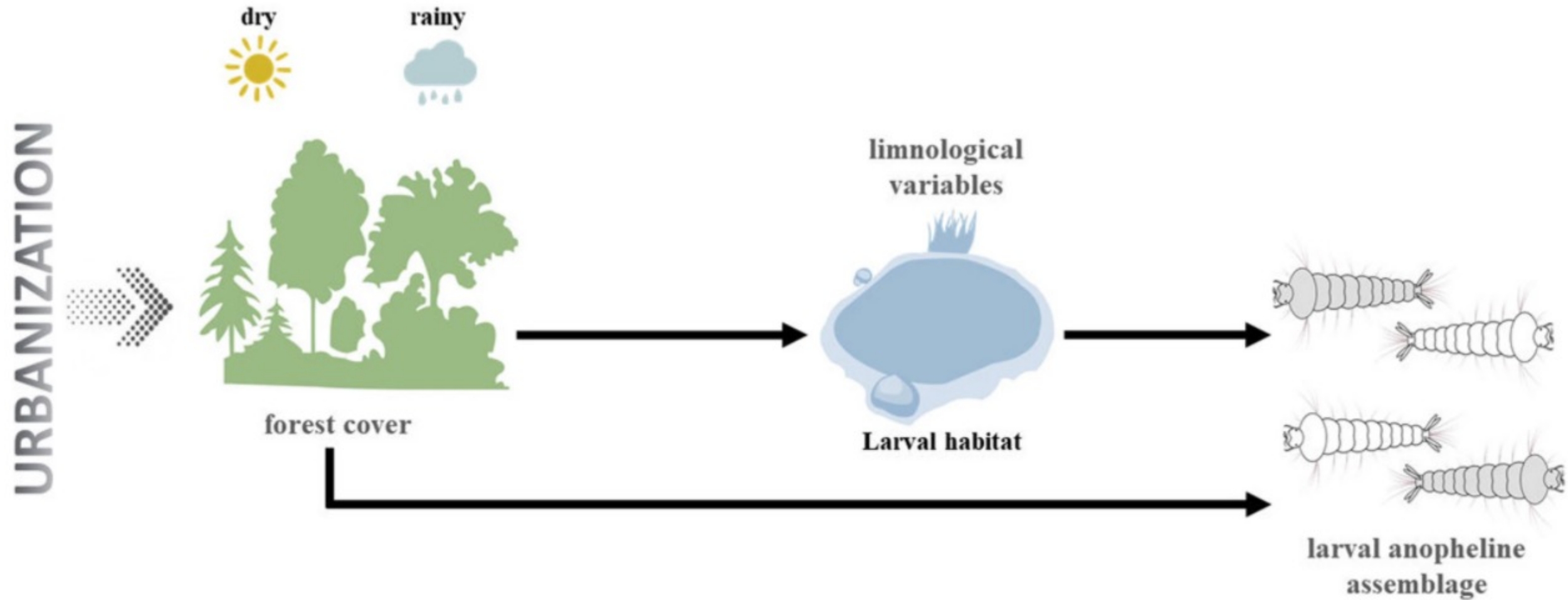
	sp_A	sp_B	sp_C	sp_D	sp_E	sp_F
5	15	21	14	0	0	0
7	9	18	16	0	0	0
3	0	12	19	4	0	0
10	0	8	9	11	9	0
4	0	8	5	17	14	5
1	0	0	3	8	13	0
6	0	0	3	7	3	8
9	0	0	0	5	1	7
2	0	0	0	0	1	4
8	0	0	0	0	0	4



Agora vamos pensar num exemplo real

Figure 1

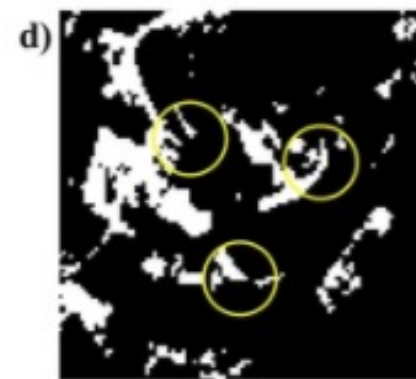
From: [Seasonality modulates the direct and indirect influences of forest cover on larval anopheline assemblages in western Amazônia](#)



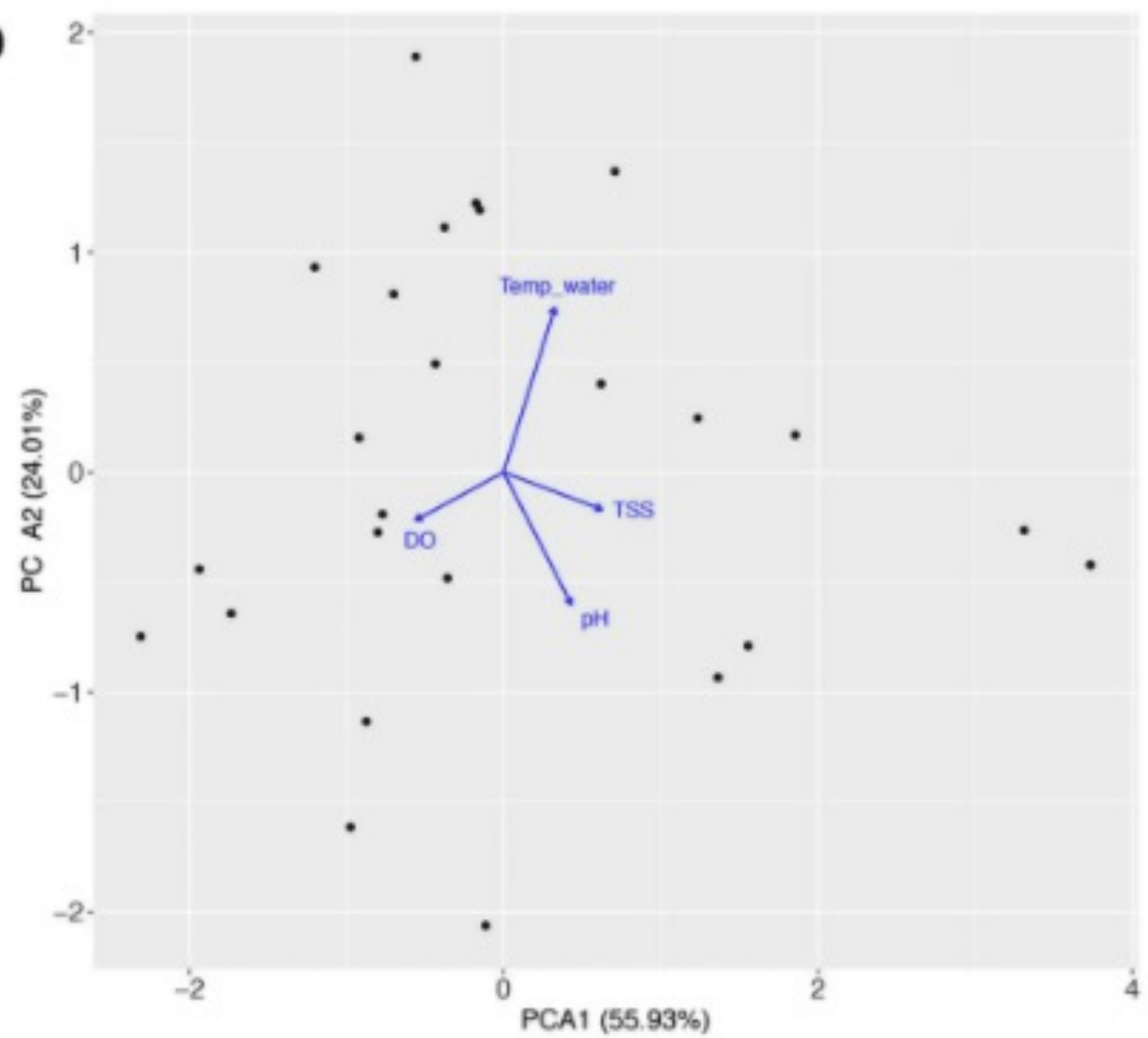
Conceptual structural equation model. The image was created using Adobe Illustrator 2020 (<https://www.adobe.com/br/products/illustrator/>).

[Scientific Reports](#) **volume 11**, Article number: 12721 (2021)

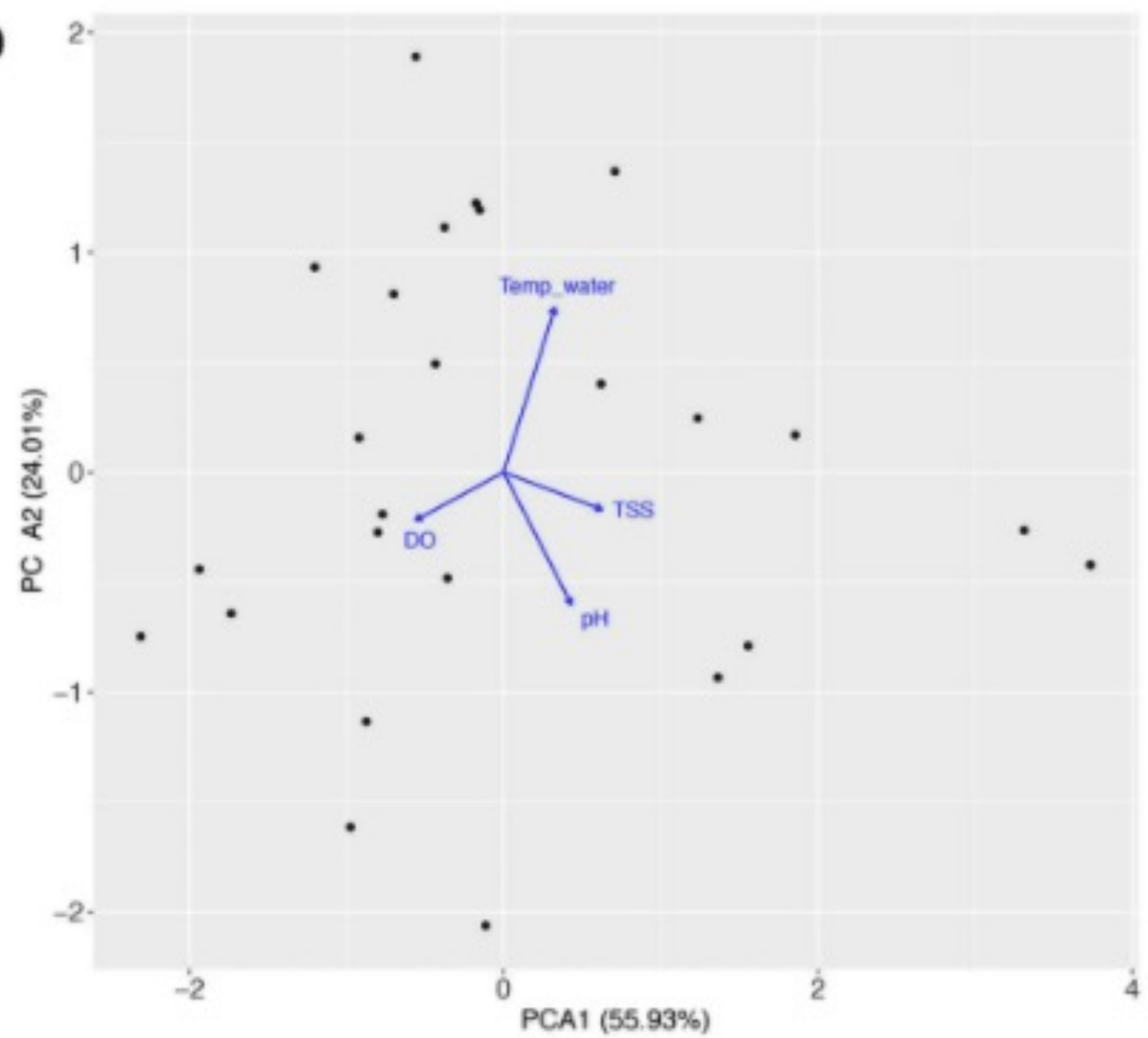
<https://www.nature.com/articles/s41598-021-92217-9>



b

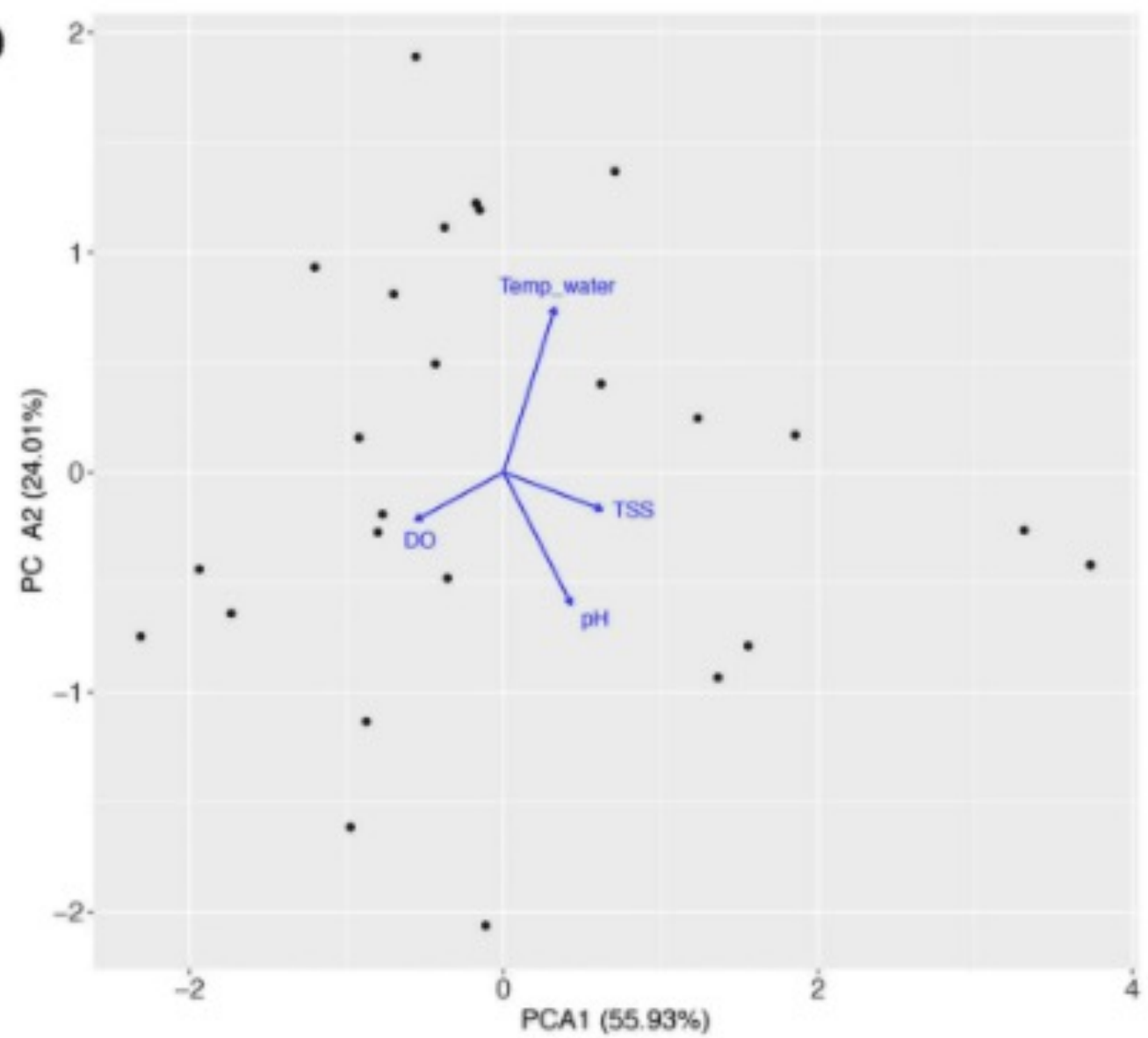


b

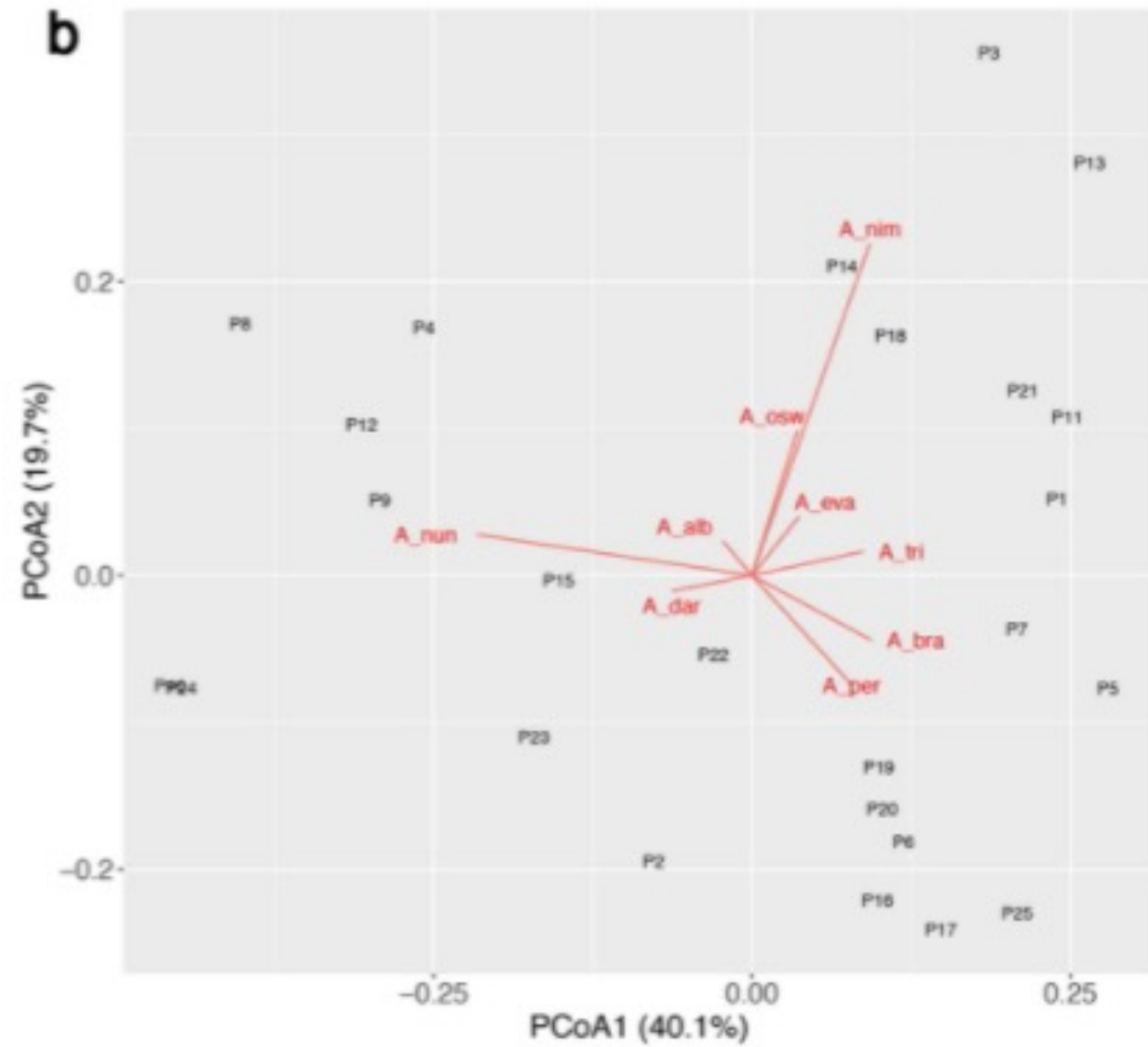


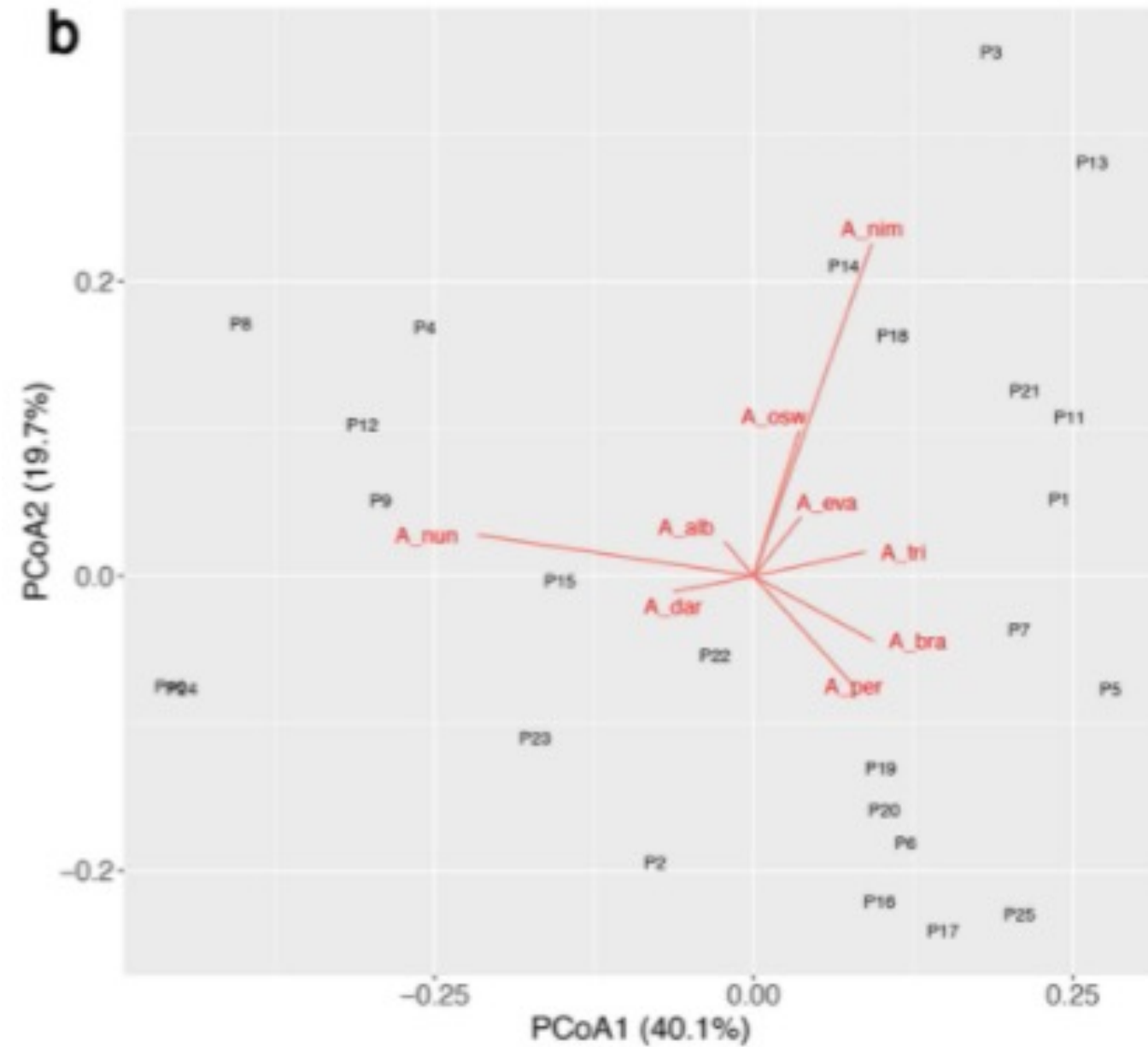
As setas indicam
as correlações
das variáveis
com os eixos da
PCA:
loadings

b

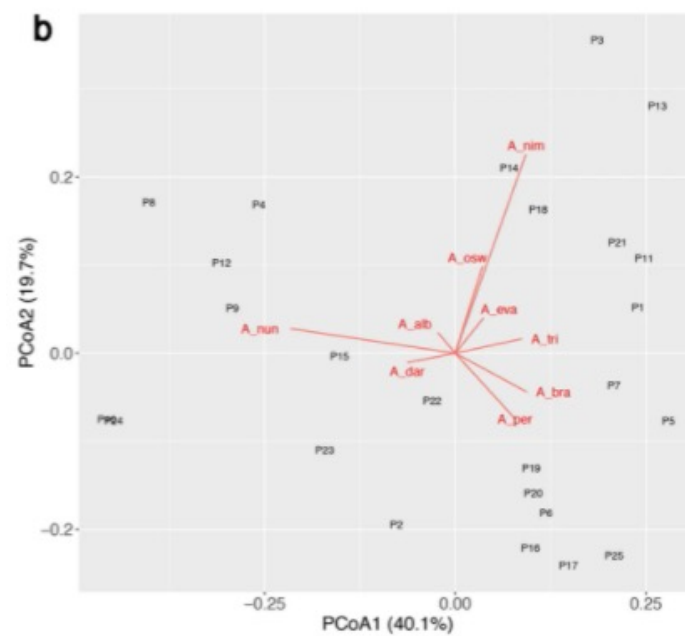
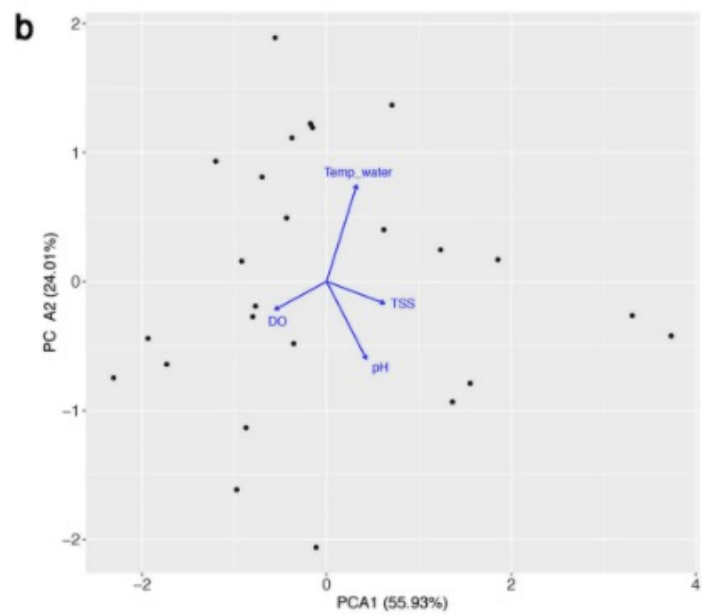
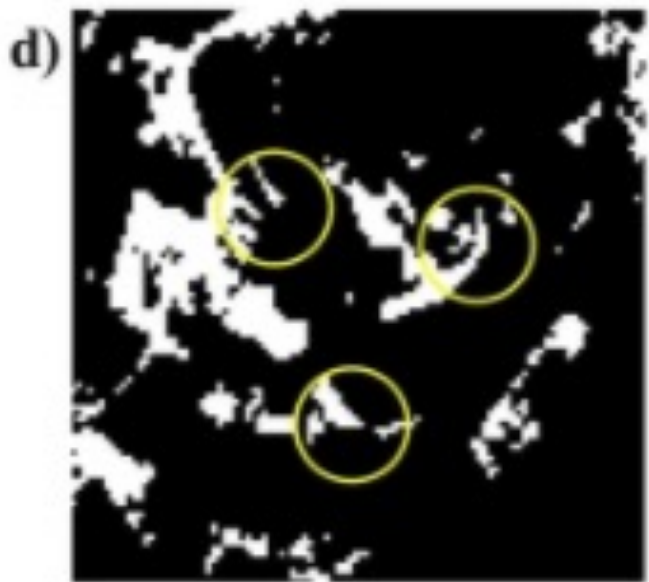


Qualidade da
água/habitat

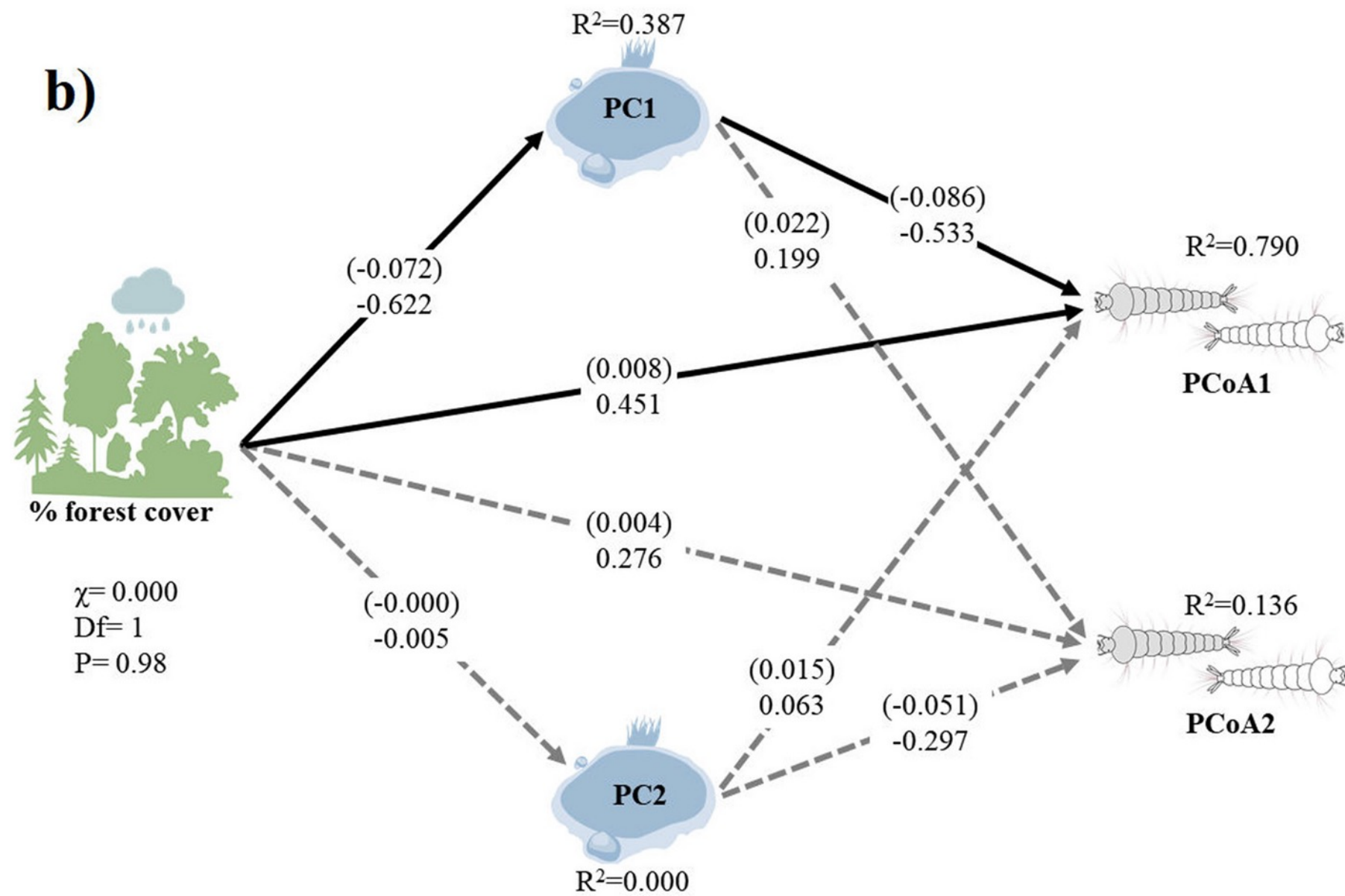
b

b

Composição de espécies



b)



Com R

Preparei um script com o passo a passo para aprender como fazer NMDS e PCoA.

'multivariate_analysis.R'

Para você entender o processo usando os dados simulados de plantas

Os arquivos necessários estão em nossa sala do Moodle:
<https://presencial.ead.ufgd.edu.br/course/view.php?id=535>