

Tutorial_diversidade_R

December 20, 2021

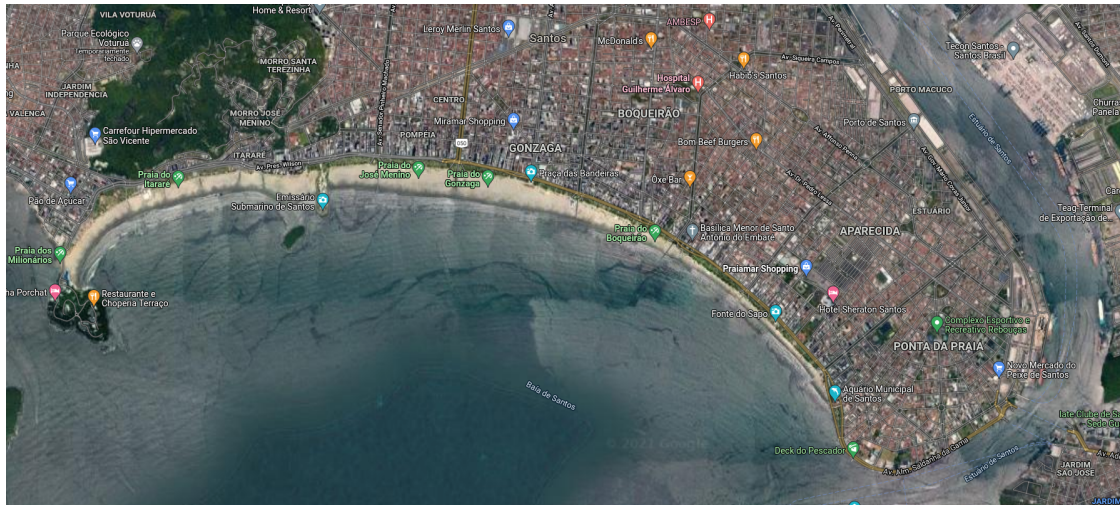
1 Medindo as diversidades α , β e γ : tutorial no R

Neste tutorial iremos utilizar uma base de dados denominada Baia_santos.xlsx, contendo a abundância de 10 espécies de peixes capturadas na zona de arrebentação da Baía de Santos em 2015. Cada linha da tabela representa uma amostra, isto é, um arrasto feito ao longo de 200 m na zona de arrebentação, seguindo a direção da linha da costa. No total foram 12 amostras. A primeira coluna identifica em que período do ano cada amostra foi obtida. As primeiras 6 amostras foram tomadas no VERAO de 2015 e as outras 6 no INVERNO de 2015. As demais colunas mostram o número de indivíduos de cada espécie nas amostras. Os nomes das espécies foram omitidos para facilitar a apresentação dos dados.

Vamos utilizar esta tabela como exemplo para calcularmos as diversidades α , β e γ .

1.1 Peixes de arrasto na zona de arrebentação da baía de Santos - SP





1.2 Carregando os pacotes do R

```
[1]: library(readxl)
library(tidyverse)
library(vegan)
library(patchwork)
library(iNEXT)
```

Attaching packages
1.3.1

tidyverse

ggplot2	3.3.5	purrr	0.3.4
tibble	3.1.6	dplyr	1.0.7
tidyr	1.1.4	stringr	1.4.0
readr	2.1.0	forcats	0.5.1

Conflicts

```
tidyverse_conflicts()
dplyr::filter() masks stats::filter()
dplyr::lag() masks stats::lag()
```

Loading required package: permute

Loading required package: lattice

This is vegan 2.5-7

1.3 Importante a base de dados

- 10 espécies (linhas)
- 12 amostras (6 verão de 2015; 6 inverno de 2015)

```
[2]: pe = read_excel('datasets/Baia_santos.xlsx')
```

```
[3]: pe
```

```

      Epoca    sp_1    sp_2    sp_3    sp_4    sp_5    sp_6    sp_7    sp_8    sp_9
      <chr>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
1 INVERNO      5        6       72      65       8        6        1        0        0
2 INVERNO      6        0        1      18       9       12        0        1        0
3 INVERNO      7        0        0      48       2        0        0        4        0
4 INVERNO      0        0        0      13       4        1        0        0        0
5 INVERNO      2        0      143      48       1        0        0        1        0
6 INVERNO      0        0       95      28       0        0        1        0        0
7 VERA0        1        0        0       8       0       27        0        0        1
8 VERA0        0        0        0       1       3        2        2        0        0
9 VERA0        8        0        0       1       4        6        0        0        4
10 VERA0       2        0        0       0       6        2        0        0        0
11 VERA0        1        0        0       3       0        2        0        0        1
12 VERA0        1        0        0       0       0        7        0        0        0

```

A tibble: 12 × 11

```
[4]: glimpse(pe)
```

```

Rows: 12
Columns: 11
$ Epoca <chr> "INVERNO", "INVERNO", "INVERNO", "INVERNO",
"INVERNO", "INVERNO"...
$ sp_1 <dbl> 5, 6, 7, 0, 2, 0, 1, 0, 8, 2, 1, 1
$ sp_2 <dbl> 6, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
$ sp_3 <dbl> 72, 1, 0, 0, 143, 95, 0, 0, 0, 0, 0, 0
$ sp_4 <dbl> 65, 18, 48, 13, 48, 28, 8, 1, 1, 0, 3, 0
$ sp_5 <dbl> 8, 9, 2, 4, 1, 0, 0, 3, 4, 6, 0, 0
$ sp_6 <dbl> 6, 12, 0, 1, 0, 0, 27, 2, 6, 2, 2, 7
$ sp_7 <dbl> 1, 0, 0, 0, 0, 1, 0, 2, 0, 0, 0, 0
$ sp_8 <dbl> 0, 1, 4, 0, 1, 0, 0, 0, 0, 0, 0, 0
$ sp_9 <dbl> 0, 0, 0, 0, 0, 0, 1, 0, 4, 0, 1, 0
$ sp_10 <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1

```

1.4 Diversidade α

A diversidade α se refere às diversidades registradas localmente, isto é, em cada uma das amostras.

Podemos mensurar a diversidade de espécies pelos seus componentes de **riqueza**, **equabilidade** ou por algum índice que combine estes componentes, isto é, um **índice de diversidade**.

```
[5]: pe
```

	Epoca <chr>	sp_1 <dbl>	sp_2 <dbl>	sp_3 <dbl>	sp_4 <dbl>	sp_5 <dbl>	sp_6 <dbl>	sp_7 <dbl>	sp_8 <dbl>	sp_9 <dbl>
A tibble: 12 × 11	INVERNO	5	6	72	65	8	6	1	0	0
	INVERNO	6	0	1	18	9	12	0	1	0
	INVERNO	7	0	0	48	2	0	0	4	0
	INVERNO	0	0	0	13	4	1	0	0	0
	INVERNO	2	0	143	48	1	0	0	1	0
	INVERNO	0	0	95	28	0	0	1	0	0
	VERAO	1	0	0	8	0	27	0	0	1
	VERAO	0	0	0	1	3	2	2	0	0
	VERAO	8	0	0	1	4	6	0	0	4
	VERAO	2	0	0	0	6	2	0	0	0
	VERAO	1	0	0	3	0	2	0	0	1
	VERAO	1	0	0	0	0	7	0	0	0

1.4.1 Riqueza de espécies

- refere-se simplesmente ao número de espécies em uma amostra
- função `specnumber()` do pacote `vegan`.

```
[6]: pe = pe %>%
      rowwise() %>%
      mutate(S = specnumber(c_across(sp_1:sp_10)))
```

```
[7]: pe
```

	Epoca <chr>	sp_1 <dbl>	sp_2 <dbl>	sp_3 <dbl>	sp_4 <dbl>	sp_5 <dbl>	sp_6 <dbl>	sp_7 <dbl>	sp_8 <dbl>	sp_9 <dbl>
A rowwise_df: 12 × 12	INVERNO	5	6	72	65	8	6	1	0	0
	INVERNO	6	0	1	18	9	12	0	1	0
	INVERNO	7	0	0	48	2	0	0	4	0
	INVERNO	0	0	0	13	4	1	0	0	0
	INVERNO	2	0	143	48	1	0	0	1	0
	INVERNO	0	0	95	28	0	0	1	0	0
	VERAO	1	0	0	8	0	27	0	0	1
	VERAO	0	0	0	1	3	2	2	0	0
	VERAO	8	0	0	1	4	6	0	0	4
	VERAO	2	0	0	0	6	2	0	0	0
	VERAO	1	0	0	3	0	2	0	0	1
	VERAO	1	0	0	0	0	7	0	0	0

1.4.2 Índice de diversidade de Simpson

- a probabilidade de que dois indivíduos retirados aleatoriamente de uma comunidade sejam da mesma espécie

$$D = \sum_{i=1}^S \left(\frac{n_i(n_i - 1)}{N(N - 1)} \right)$$

- função `diversity()` do pacote `vegan`
 - `index = 'invsimpson'` -> recíproca do índice $\frac{1}{D}$
 - `index = 'simpson'` -> complementar do índice $1 - D$

```
[8]: pe = pe %>%
      rowwise() %>%
      mutate(D = diversity(c_across(sp_1:sp_10),
                                index = 'invsimpson'))
```

```
[9]: pe
```

	Epoca	sp_1	sp_2	sp_3	sp_4	sp_5	sp_6	sp_7	sp_8	sp_9
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
A rowwise_df: 12 × 13	INVERNO	5	6	72	65	8	6	1	0	0
	INVERNO	6	0	1	18	9	12	0	1	0
	INVERNO	7	0	0	48	2	0	0	4	0
	INVERNO	0	0	0	13	4	1	0	0	0
	INVERNO	2	0	143	48	1	0	0	1	0
	INVERNO	0	0	95	28	0	0	1	0	0
	VERAO	1	0	0	8	0	27	0	0	1
	VERAO	0	0	0	1	3	2	2	0	0
	VERAO	8	0	0	1	4	6	0	0	4
	VERAO	2	0	0	0	6	2	0	0	0
	VERAO	1	0	0	3	0	2	0	0	1
	VERAO	1	0	0	0	0	7	0	0	0

1.4.3 Equabilidade de Simpson

- mede a uniformidade das abundâncias entre espécies

$$E_{1/D} = \frac{1/D}{S}$$

- varia de 0 (equabilidade mínima) a 1 se todas as espécies tiverem exatamente a mesma abundância.

```
[10]: pe = pe %>%
      rowwise() %>%
      mutate(E = D/S)
```

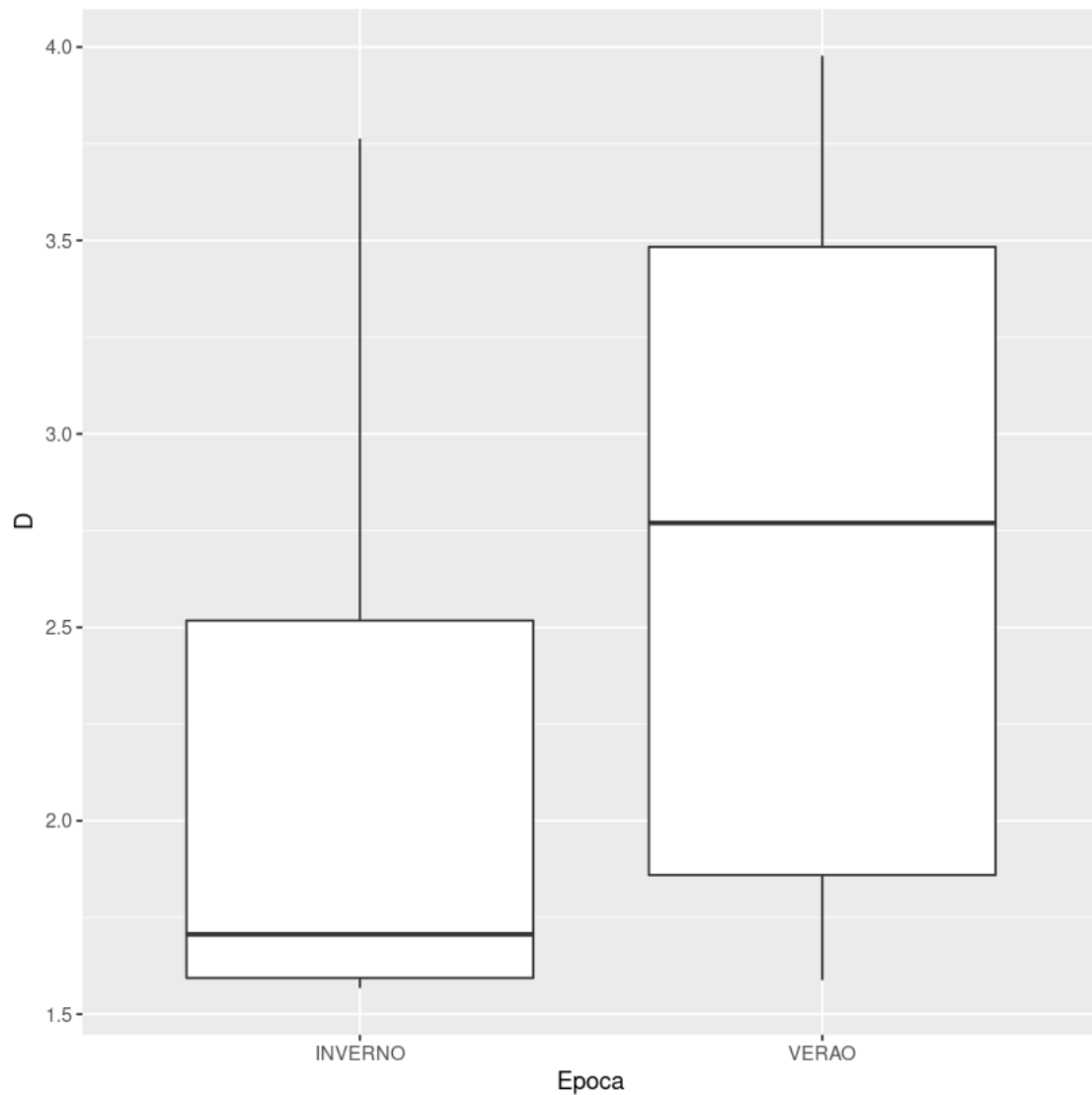
```
[11]: pe
```

	Epoca <chr>	sp_1 <dbl>	sp_2 <dbl>	sp_3 <dbl>	sp_4 <dbl>	sp_5 <dbl>	sp_6 <dbl>	sp_7 <dbl>	sp_8 <dbl>	sp_9 <dbl>
A rowwise_df: 12 × 14	INVERNO	5	6	72	65	8	6	1	0	0
	INVERNO	6	0	1	18	9	12	0	1	0
	INVERNO	7	0	0	48	2	0	0	4	0
	INVERNO	0	0	0	13	4	1	0	0	0
	INVERNO	2	0	143	48	1	0	0	1	0
	INVERNO	0	0	95	28	0	0	1	0	0
	VERAO	1	0	0	8	0	27	0	0	1
	VERAO	0	0	0	1	3	2	2	0	0
	VERAO	8	0	0	1	4	6	0	0	4
	VERAO	2	0	0	0	6	2	0	0	0
	VERAO	1	0	0	3	0	2	0	0	1
	VERAO	1	0	0	0	0	7	0	0	0

1.4.4 Representação gráfica da diversidade α : Boxplots

Como temos amostras de VERA0 e INVERNO, nos interessa saber qual a distribuição da riqueza, diversidade e equabilidade nestes dois períodos.

```
[12]: ggplot(pe) +
  aes(x = Epoca, y = D) +
  geom_boxplot()
```



```
[13]: plt_D = ggplot(pe) +
  aes(x = Epoca, y = D) +
  geom_boxplot(fill = 'lightblue', alpha = 0.5) +
  geom_jitter(width = 0.1, size = 3) +
  labs(y = 'Diversidade de Simpson (1/D)',
       x = '') +
  theme_classic()

plt_S = ggplot(pe) +
  aes(x = Epoca, y = S) +
  geom_boxplot(fill = 'lightblue', alpha = 0.5) +
  geom_jitter(width = 0.1, size = 3) +
  labs(y = 'Riqueza de Espécies (S)',
```

```

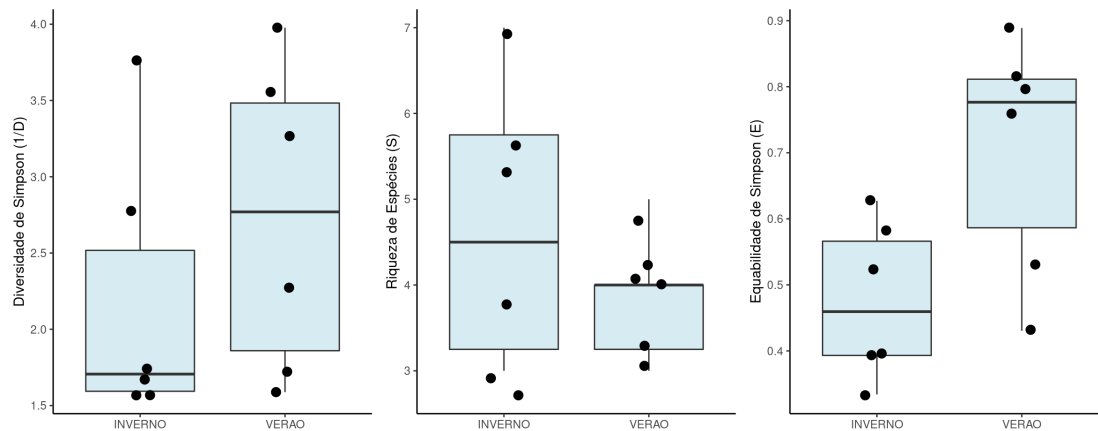
    x = '') +
  theme_classic()

plt_E = ggplot(pe) +
  aes(x = Epoca, y = E) +
  geom_boxplot(fill = 'lightblue', alpha = 0.5) +
  geom_jitter(width = 0.1, size = 3) +
  labs(y = 'Equabilidade de Simpson (E)',
       x = '') +
  theme_classic()

```

```
[14]: options(repr.plot.width = 12, repr.plot.height = 5, repr.plot.res = 200)
```

```
[15]: plt_D | plt_S | plt_E
```



1.5 Diversidade γ

- Curva de acumulação de espécies: como a riqueza total aumenta à medida que aumenta o esforço amostral, isto é, o número de indivíduos coletados?
- pacote iNEXT

```

[16]: pe_list = list()

pe_list$VERAO = pe %>%
  filter(Epoca == 'VERAO') %>%
  select(sp_1:sp_10) %>%
  colSums()

pe_list$INVERNO = pe %>%
  filter(Epoca == 'INVERNO') %>%
  select(sp_1:sp_10) %>%

```



```
colSums()
```

```
[17]: pe_list
```

```
$VERAO sp\_1 13 sp\_2 0 sp\_3 0 sp\_4 13 sp\_5 13 sp\_6 46 sp\_7 2 sp\_8 0  
      sp\_9      6 sp\_10      1  
$INVERNO sp\_1 20 sp\_2 6 sp\_3 311 sp\_4 220 sp\_5 24 sp\_6 19 sp\_7 2  
      sp\_8      6 sp\_9      0 sp\_10      0
```

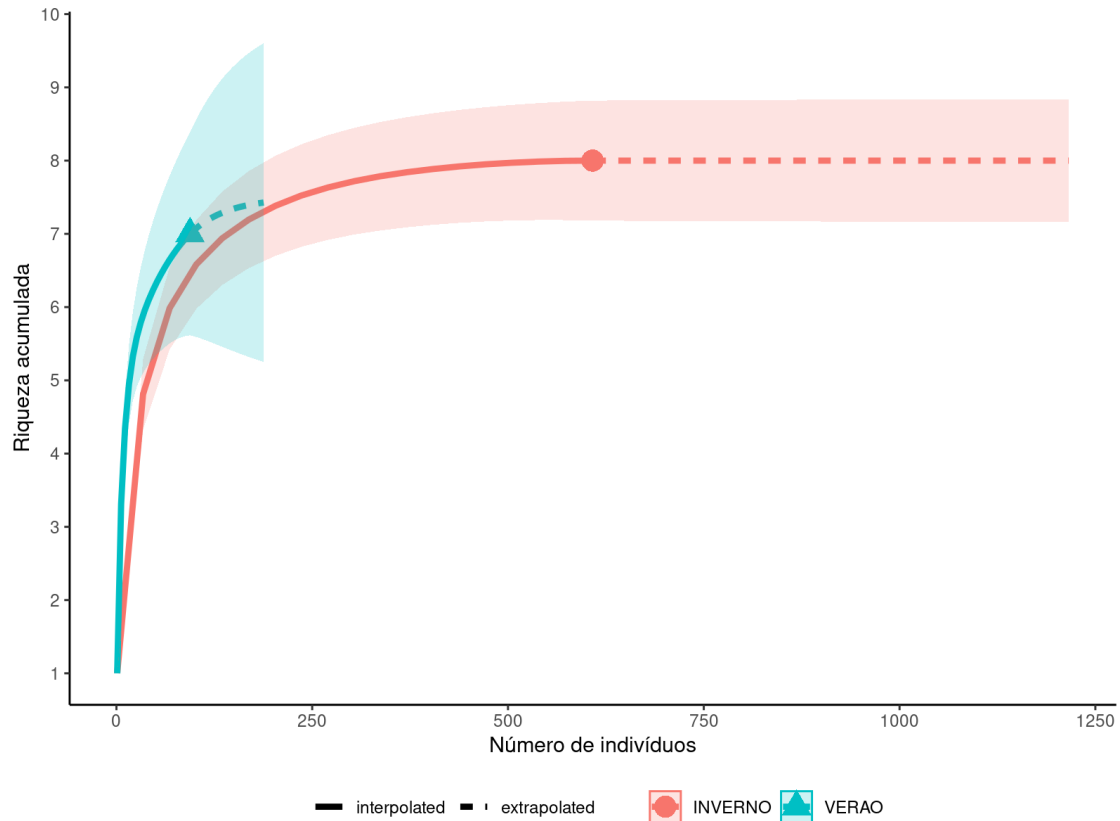
```
[18]: gama_ac = iNEXT(pe_list,  
                      datatype = 'abundance',  
                      q = 0) # q = 0 --> Riqueza de espécies
```

1.5.1 Representação gráfica da curva de acumulação

```
[19]: plt_gama = ggiNEXT(gama_ac) +  
      labs(y = 'Riqueza acumulada',  
           x = 'Número de indivíduos') +  
      theme_classic() +  
      scale_y_continuous(breaks = 0:10) +  
      theme(legend.position = "bottom",  
            legend.title=element_blank())
```

```
[20]: options(repr.plot.width = 8, repr.plot.height = 6, repr.plot.res = 200)
```

```
[21]: plt_gama
```



1.6 Diversidade β

- Índice de Jaccard calcula a similaridade par-a-par

$$J = \frac{a}{a + b + c}$$

+ a: número de espécies presentes nas duas amostras;

+ b: número de espécies presentes somente na Amostra 1;

+ c: número de espécies presentes somente na Amostra 2.

1.6.1 Matriz de distância

- a função `decostand(method = 'pa')` cria uma matriz de presença-ausência;
- função `vegdist()` do pacote `vegan` calcula a **distância** de Jaccard, isto é,

$$1 - J$$

```
[22]: pe_ocor = pe %>%
      select(sp_1:sp_10) %>%
      decostand(method = 'pa')

jac = vegdist(pe_ocor, method = 'jaccard', binary = TRUE)
```

1.6.2 Representando uma matriz de dissimilaridade

```
[23]: jac
```

	1	2	3	4	5	6	7
2	0.3750000						
3	0.6250000	0.3333333					
4	0.5714286	0.5000000	0.6000000				
5	0.5000000	0.1666667	0.2000000	0.6666667			
6	0.5714286	0.7142857	0.8333333	0.8000000	0.6666667		
7	0.6250000	0.5714286	0.6666667	0.6000000	0.7142857	0.8333333	
8	0.4285714	0.5714286	0.6666667	0.2500000	0.7142857	0.6000000	0.6666667
9	0.5000000	0.4285714	0.5000000	0.4000000	0.5714286	0.8571429	0.2000000
10	0.5714286	0.5000000	0.6000000	0.5000000	0.6666667	1.0000000	0.6000000
11	0.6250000	0.5714286	0.6666667	0.6000000	0.7142857	0.8333333	0.0000000
12	0.7500000	0.7142857	0.8333333	0.8000000	0.8571429	1.0000000	0.6000000
	8	9	10	11			
2							
3							
4							
5							
6							
7							
8							
9	0.5000000						
10	0.6000000	0.4000000					
11	0.6666667	0.2000000	0.6000000				
12	0.8333333	0.6666667	0.5000000	0.6000000			

1.6.3 Análise de Coordenadas Principais (PCoA)

- Representando as distâncias em dois eixos

```
[24]: # Fazendo a PCoA
pcoa = cmdscale(jac, eig = TRUE)

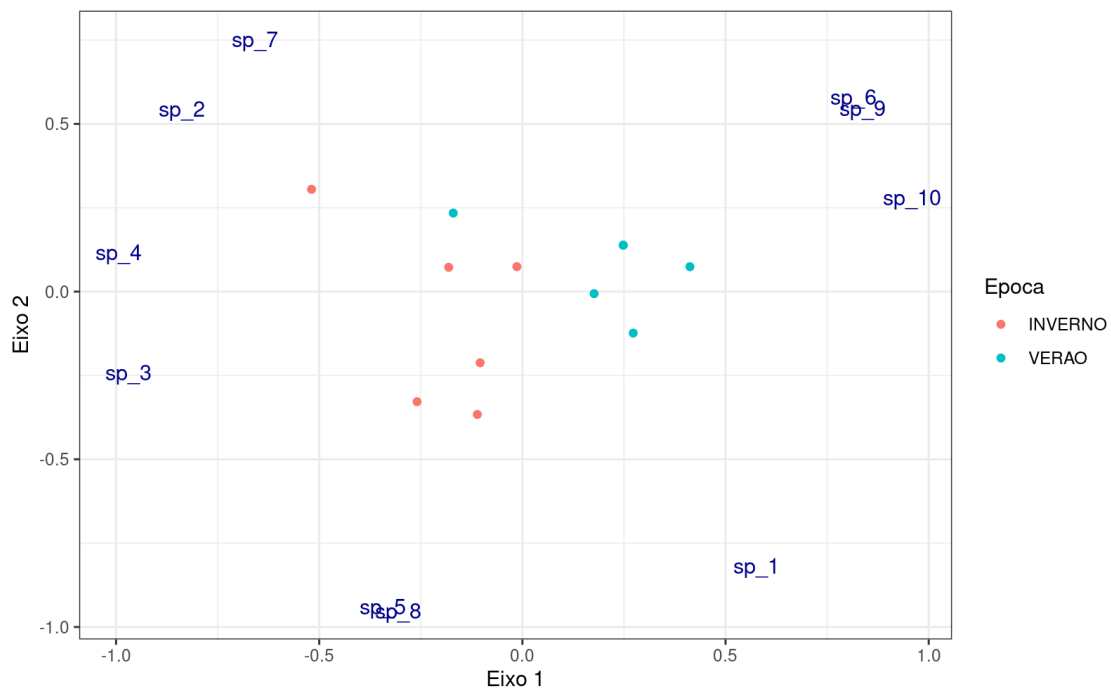
# Mapeando as amostras
mapa_amostras = pcoa$points %>%
  as.data.frame() %>%
  mutate(Epoca = pe$Epoca)
```

```
# Mapeando as espécies
mapa_especies = envfit(pcoa, pe_ocor)$vectors$arrows[,1:2] %>%
  as.data.frame() %>%
  rownames_to_column(var = 'Especies')
```

```
[25]: options(repr.plot.width = 8, repr.plot.height = 5, repr.plot.res = 200)
```

```
[26]: plt_beta = ggplot() +
  geom_point(data = mapa_amostras,
    aes(x = V1, y = V2, color = Epoca)) +
  geom_text(data = mapa_especies,
    aes(x = Dim1, y = Dim2, label = Especies),
    color = 'darkblue') +
  labs(x = 'Eixo 1', y = 'Eixo 2') +
  theme_bw()
```

plt_beta



1.7 Salvando as figuras deste tutorial

```
[27]: # Diversidade alfa
div_alfa = plt_D | plt_S | plt_E
ggsave(filename = 'Diversidade_alfa.png',
  plot = div_alfa,
```

```
width = 15, height = 5)

# Diversidade gama
div_gama = plt_gama
ggsave(filename = 'Diversidade_gama.png',
        plot = div_gama,
        width = 7, height = 5)

# Diversidade beta
div_beta = plt_beta
ggsave(filename = 'Diversidade_beta.png',
        plot = div_beta,
        width = 7, height = 5)
```