

CEN5807 - Bioinformática aplicada ao estudo de comunidades microbianas

Análise e visualização de dados no R



Anderson Freitas, MSc.

Piracicaba-SP
Setembro de 2023



Por que usar o R?

- Linguagem livre e amplamente difundida
- Menor exigência computacional
- Versatilidade nas análises
- Figuras bonitas e versáteis



O que vamos estudar hoje?

- Tipos de dados e conjuntos de dados
- ggplot2, phyloseq, microeco
- Testes de hipóteses para dados moleculares ou não

Tipos de Dados

Dados Estruturados

Mês	Valor da cota (em R\$)
Janeiro	R\$ 635,00
Fevereiro	R\$ 635,00
Março	R\$ 635,00
Abril	R\$ 630,00
Maio	R\$ 625,00
Junho	R\$ 595,00
Julho	R\$ 550,00
Agosto	R\$ 490,00
Setembro	R\$ 470,00
Outubro	R\$ 480,00
Novembro	R\$ 480,00
Dezembro	R\$ 480,00

Dados Semi-Estruturados

```
.home-page {
  background-color: #f0f0f0;
  padding: 13px;
}
.service-item {
  float: left;
  position: relative;
  a {
    text-decoration: none;
  }
}
.block-container {
  margin: 7px;
  padding: 5px;
  background-color: rgba(0, 0, 0, 0.6);
  border-radius: 4px;
  color: $blank;
  height: 90%;
  height: calc(100% - 24px);
  position: relative;
}
.service-icon {
  width: 85px;
  height: 85px;
  margin-left: calc(50% - 42.5px);
}
h4.service-title {
  font-size: 24px;
  text-align: center;
  margin: 0;
}
p.service-text {
  font-size: 16px;
  margin: 0;
}
```

Dados Não Estruturados

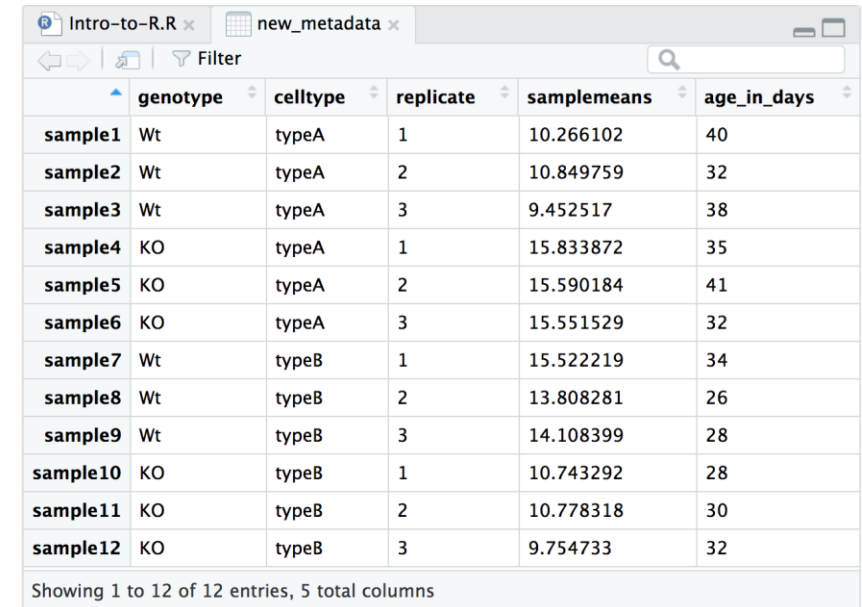


Tipos de Objetos que carregam dados

- Matriz

```
> A
      [,1] [,2] [,3] [,4] [,5]
[1,]    1    5    9   13   17
[2,]    2    6   10   14   18
[3,]    3    7   11   15   19
[4,]    4    8   12   16   20
```

- Dataframe

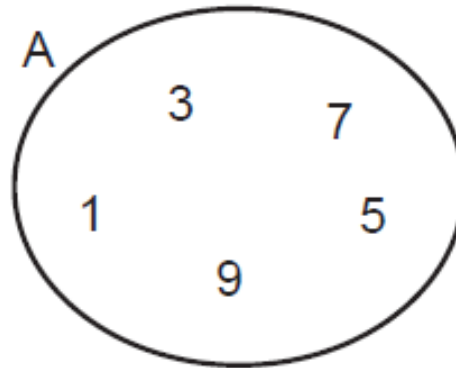


The screenshot shows an RStudio window with a tab titled 'new_metadata'. Below the tab is a search bar and a 'Filter' button. The main area displays a dataframe with the following columns: 'genotype', 'celltype', 'replicate', 'samplemeans', and 'age_in_days'. The rows are labeled 'sample1' through 'sample12'. The data is as follows:

	genotype	celltype	replicate	samplemeans	age_in_days
sample1	Wt	typeA	1	10.266102	40
sample2	Wt	typeA	2	10.849759	32
sample3	Wt	typeA	3	9.452517	38
sample4	KO	typeA	1	15.833872	35
sample5	KO	typeA	2	15.590184	41
sample6	KO	typeA	3	15.551529	32
sample7	Wt	typeB	1	15.522219	34
sample8	Wt	typeB	2	13.808281	26
sample9	Wt	typeB	3	14.108399	28
sample10	KO	typeB	1	10.743292	28
sample11	KO	typeB	2	10.778318	30
sample12	KO	typeB	3	9.754733	32

Showing 1 to 12 of 12 entries, 5 total columns

- Lista



Quais arquivos vamos trabalhar

- OTU/ASV table
- Tax table
- Mapping file



OTU/ASV table

#OTU	ID	F3D0	F3D141	F3D142	F3D143	F3D144	F3D145	F3D146	F3D147
OTU_6		749	535	313	372	607	849	493	2025
OTU_25		29	57	14	2	14	22	16	127
OTU_1		613	497	312	247	472	719	349	1720
OTU_8		426	378	255	237	382	627	330	1417
OTU_31		149	38	10	19	25	21	43	31
OTU_2		366	392	327	185	313	542	248	1367
OTU_7		196	370	92	107	48	155	74	105
OTU_10		46	169	87	109	171	209	120	864
OTU_80		26	6	0	1	4	8	18	11

Taxonomy (tax) table

	Kingdom	Phylum	Class	Order	Family	Genus
GCGGTAAGACGAACCTTGCTAACGTTA...	Bacteria	Firmicutes	Bacilli	Erysipelotrichales	Erysipelotrichaceae	[Anaerorhabdus] furcosa group
GCTGTAATACGTATGGAGCAAGCGTTAT...	Bacteria	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	[Eubacterium] eligens group
GCGGTAATACGTAGGTGGCGAGCGTTG...	Bacteria	Firmicutes	Clostridia	Oscillospirales	Ruminococcaceae	[Eubacterium] siraeum group
GCGGTAATACGTAGGTGGCGAGCGTTG...	Bacteria	Firmicutes	Clostridia	Oscillospirales	Ruminococcaceae	[Eubacterium] siraeum group
GCGGTAATACGTAGGTGGCAAGCGTTG...	Bacteria	Firmicutes	Clostridia	Oscillospirales	Ruminococcaceae	[Eubacterium] siraeum group
GCGGTAATACGTAGGTGGCGAGCGTTG...	Bacteria	Firmicutes	Clostridia	Oscillospirales	Ruminococcaceae	[Eubacterium] siraeum group
GCGGTAATACGTAGGGAGCAAGCGTTG...	Bacteria	Firmicutes	Bacilli	Bacillales	Planococcaceae	[Renibacterium] salmoninarum group
GCGGTAATACGTAGGTGGCAAGCGTTG...	Bacteria	Firmicutes	Bacilli	Bacillales	Planococcaceae	[Renibacterium] salmoninarum group
GCGGTAATACGTAGGGAGCAAGCGTTG...	Bacteria	Firmicutes	Bacilli	Bacillales	Planococcaceae	[Renibacterium] salmoninarum group
GCGGTAAAACGTAGGGGGCAAGCGTT...	Bacteria	Firmicutes	Bacilli	Bacillales	Planococcaceae	[Renibacterium] salmoninarum group
GCGGTAAGACAGAGGGGGCAAGCGTT...	Bacteria	Firmicutes	Bacilli	Bacillales	Planococcaceae	[Renibacterium] salmoninarum group
GCGGTAATACGAAGGGTGCAAACGTTG...	Bacteria	Firmicutes	Bacilli	Bacillales	Planococcaceae	[Renibacterium] salmoninarum group
GCGGTAATACGTAGGGGGCTAGCGTTG...	Bacteria	Firmicutes	Bacilli	Bacillales	Planococcaceae	[Renibacterium] salmoninarum group
GCGGTAATACGTAGGGGCGCAAGCGTTG...	Bacteria	Actinobacteriota	Actinobacteria	Streptosporangiales	Thermomonosporaceae	[Streptomyces] thermoautotrophicus
GCGGTAATACGTAGGGGCGCAAGCGTTG...	Bacteria	Actinobacteriota	Actinobacteria	Streptosporangiales	Thermomonosporaceae	[Streptomyces] thermoautotrophicus
GCGGTAATACGTAGGGGCGCAAGCGTTG...	Bacteria	Actinobacteriota	Actinobacteria	Streptosporangiales	Thermomonosporaceae	[Streptomyces] thermoautotrophicus
GCGGTAATACGTAGGGGCGCCAGCGTTG...	Bacteria	Actinobacteriota	Actinobacteria	Streptosporangiales	Thermomonosporaceae	[Streptomyces] thermoautotrophicus
GCGGTAATACGTAGGGGCGCAAGCGTTG...	Bacteria	Actinobacteriota	Actinobacteria	Streptosporangiales	Thermomonosporaceae	[Streptomyces] thermoautotrophicus
GCGGTAATACGTAGGGGCGCGAGCGTTG...	Bacteria	Actinobacteriota	Actinobacteria	Streptosporangiales	Thermomonosporaceae	[Streptomyces] thermoautotrophicus
GCGGTAATACGTAGGGGCGCGAGCGTTG...	Bacteria	Actinobacteriota	Actinobacteria	Streptosporangiales	Thermomonosporaceae	[Streptomyces] thermoautotrophicus

Mapping file

	Sample_ID	Substrate	Tree	U_brizantha_weight	Period
E19	E19	100%ADE	C. pachystachya	5.978	T1
E20	E20	100%ADE	C. pachystachya	7.868	T1
E21	E21	100%ADE	C. pachystachya	3.688	T1
E36	E36	100%ADE	C. pachystachya	4.878	T1
E25	E25	20%ADE	C. pachystachya	2.828	T1
E26	E26	20%ADE	C. pachystachya	3.748	T1
E27	E27	20%ADE	C. pachystachya	1.858	T1
E35	E35	20%ADE	C. pachystachya	4.098	T1
E22	E22	Control	C. pachystachya	1.148	T1
E23	E23	Control	C. pachystachya	0.408	T1
E24	E24	Control	C. pachystachya	1.358	T1
E34	E34	Control	C. pachystachya	1.138	T1

Bibliotecas

ggplot2 3.4.3


Reference

News ▾

Articles ▾

Extensions

Search for

 **ggplot2**

Overview

ggplot2 is a system for declaratively creating graphics, based on [The Grammar of Graphics](#). You provide the data, tell ggplot2 how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details.

Installation

```
# The easiest way to get ggplot2 is to install the whole tidyverse:  
install.packages\("tidyverse"\)  
  
# Alternatively, install just ggplot2:  
install.packages\("ggplot2"\)  
  
# Or the development version from GitHub:  
# install.packages("pak")  
pak::pak\("tidyverse/ggplot2"\)
```

LINKS

- [View on CRAN](#)
- [Browse source code](#)
- [Report a bug](#)
- [Learn more](#)
- [Extensions](#)

LICENSE

- [Full license](#)
- [MIT](#) + file [LICENSE](#)



COMMUNITY

- [Contributing guide](#)
- [Code of conduct](#)

CITATION

- [Citing ggplot2](#)

DEVELOPERS

- [Hadley Wickham](#)
Author 
- [Winston Chang](#)
Author 

<https://ggplot2.tidyverse.org>

Bibliotecas



<https://github.com/joey711/phyloseq>

Quick Install

In R terminal:

```
if(!requireNamespace("BiocManager")){\n  install.packages("BiocManager")\n}\nBiocManager::install("phyloseq")
```



See [the phyloseq installation page](#) for further details, examples.

Article on Improved Microbiome Analysis

McMurdie and Holmes (2014) [Waste Not, Want Not: Why Rarefying Microbiome Data is Statistically Inadmissible](#) *PLoS Computational Biology* 10(4): e1003531

Presubmission versions ahead of acceptance (2013): [PDF version 2](#), [PDF version 1](#)

Peer-reviewed articles about phyloseq

McMurdie and Holmes (2014) [Shiny-phyloseq: Web Application for Interactive Microbiome Analysis with Provenance Tracking](#). *Bioinformatics (Oxford, England)* 31(2), 282–283.

McMurdie and Holmes (2013) [phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data](#) *PLoS ONE* 8(4):e61217

Bibliotecas

microeco

An R package for data mining in microbial community ecology

CRAN 0.20.0 downloads 32K Release 0.20.0 Test 1.0.0



Background

With the development of high-throughput sequencing techniques, the increasing data amount and complexity make the microbiome data analysis and management a challenge. Though there has been a lot of R packages created for the microbiome profiling analysis, it is still difficult to perform data mining fast, efficiently and comprehensively. Therefore, we created R microeco package (abbreviated and pronounced as *[mi:kəu]*).

Main Features

- R6 Class to store and analyze data: flexible and modularized
- Taxonomic abundance analysis
- Venn diagram
- Alpha diversity
- Beta diversity
- Differential abundance test
- Machine learning
- Null model analysis
- Network analysis
- Environmental data analysis
- Functional prediction

<https://chiliubio.github.io/microeco/>

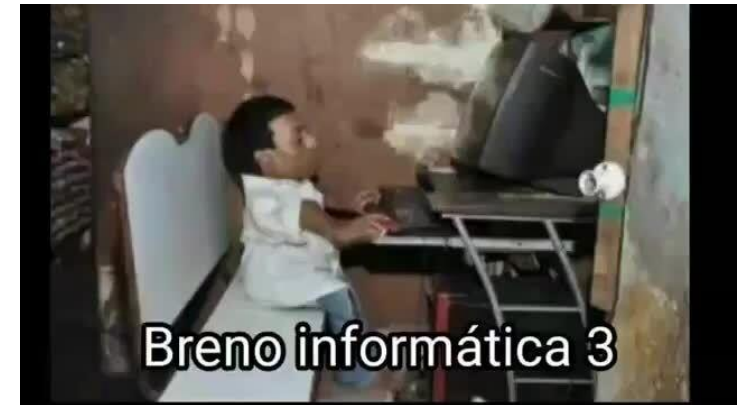
Install R/RStudio

If you do not already have R/RStudio installed, do as follows.

1. Install [R](#)
2. Install [RStudio](#)

Abrindo o R

- Ambiente do RStudio
- Importação de objetos
- Testes de Hipótese
- Gráficos
- Análises de *metabarcoding*



Acabou aqui, abram o R

