

Minicurso de Exploração e Visualização de dados biológicos em R

M.Sc Amanda Araújo Serrão de Andrade
Laboratório de Bioinformática (LNCC/UFRJ)

Conteúdo programático

Dia 2: Pré-processamento e Análise exploratória dos dados (9h - 11h)

- Estatísticas descritivas para dados biológicos (Objetos DNAMStringSet e dataframes).
- Pré-processamento de dados: tratamento de dados ausentes, transformação de dados e controle de qualidade.
- Exercício prático: Pré-processamento e condução de uma análise básica em um conjunto de dados do mundo real.
- Introdução à visualização de dados no R (Pacote ggplot2 - Teórica).

Dados biológicos que podem ser analisados no R

O R é uma ferramenta muito versátil para análise de dados biológicos. Aqui estão alguns exemplos de tipos de dados biológicos que podem ser importados e analisados no R.

Sequências de DNA e RNA;

Para a leitura das sequências, o R oferece pacotes como Bioconductor, Biostrings e seqinr. Para a análise, os principais pacotes são: GenomicRanges e VariantAnnotation.

Você pode importar sequências em formato FASTA ou GenBank, realizar alinhamentos, calcular composições de bases, identificar genes, fazer anotações genômicas e identificar variações genéticas.

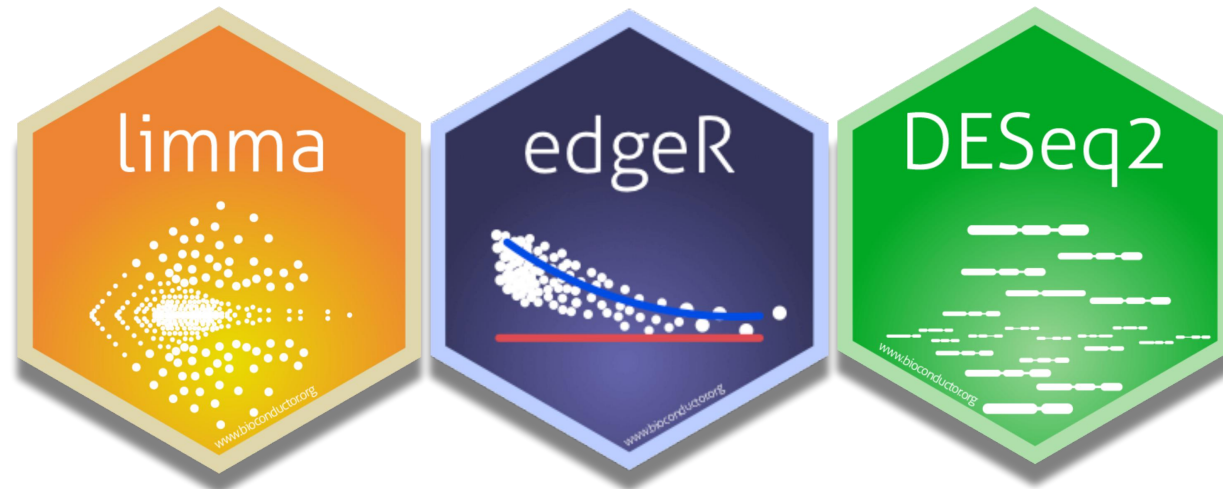


Dados biológicos que podem ser analisados no R

Dados de expressão gênica;

O R é amplamente utilizado na análise de dados de expressão gênica obtidos por meio de técnicas como microarrays e RNA-Seq.

Pacotes como DESeq2 e limma são comuns para análise de diferenciação de expressão gênica.



Dados biológicos que podem ser analisados no R

Dados de filogenia;

Para análise filogenética, você pode importar árvores filogenéticas em formato Newick ou Nexus e usar pacotes como ape, ggtree e phytools para realizar análises evolutivas.

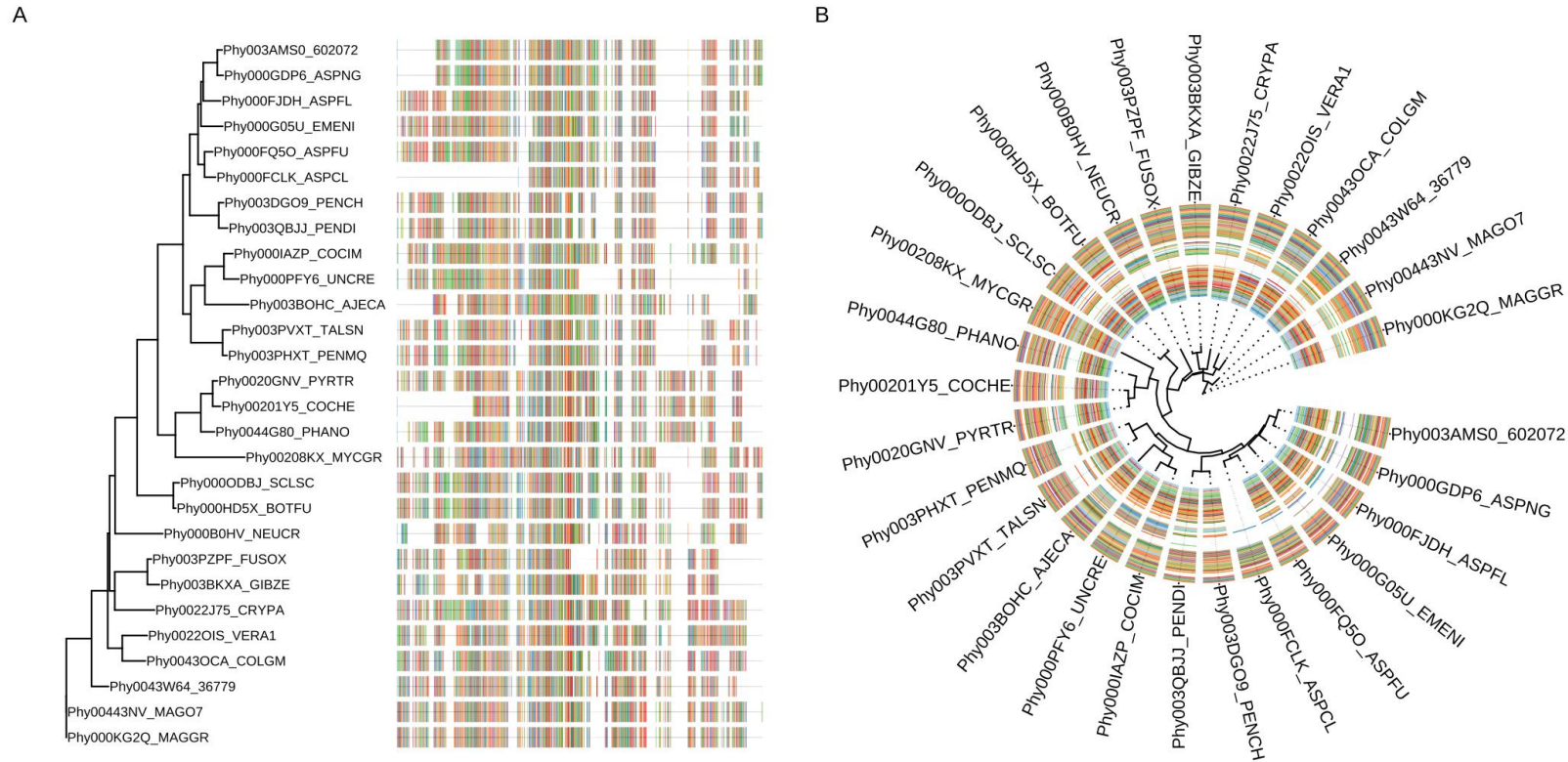


Fig 1. Exemplo de uma árvore gerada e customizada no R, utilizando o ggtree (<https://guangchuangyu.github.io/ggtree-book/chapter-ggtree.html>)

Dados biológicos que podem ser analisados no R

Dados de filogenia;

Para análise filogenética, você pode importar árvores filogenéticas em formato Newick ou Nexus e usar pacotes como ape, ggtree e phytools para realizar análises evolutivas.

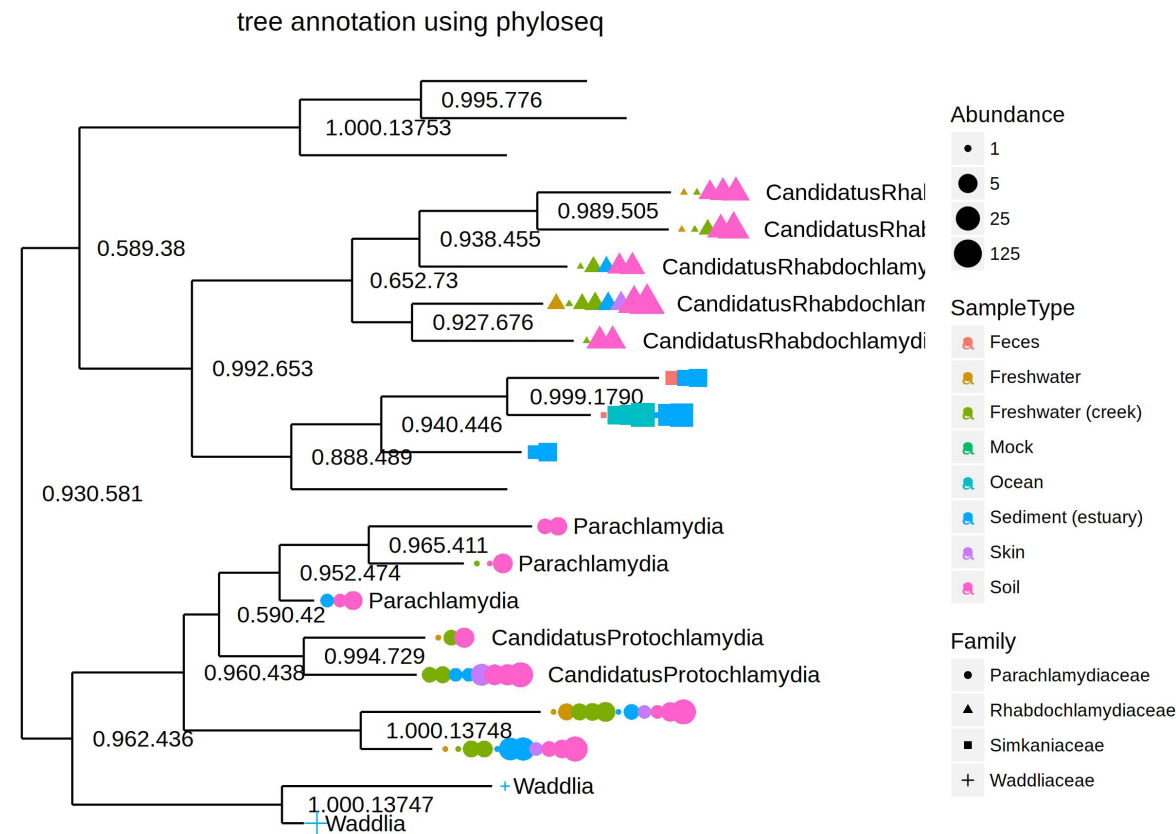


Fig 2. Exemplo de uma árvore customizada no R, utilizando o phyloseq (<https://guangchuangyu.github.io/ggtree-book/chapter-ggtree.html>)

Dados biológicos que podem ser analisados no R

Dados de microbiologia;

O R pode ser usado para análise de microbiomas, identificação de espécies microbianas e análise de diversidade microbiana (alfa e beta diversidade, por exemplo) e testes de abundância diferencial usando pacotes como phyloseq, iNEXT, vegan e microbiome.

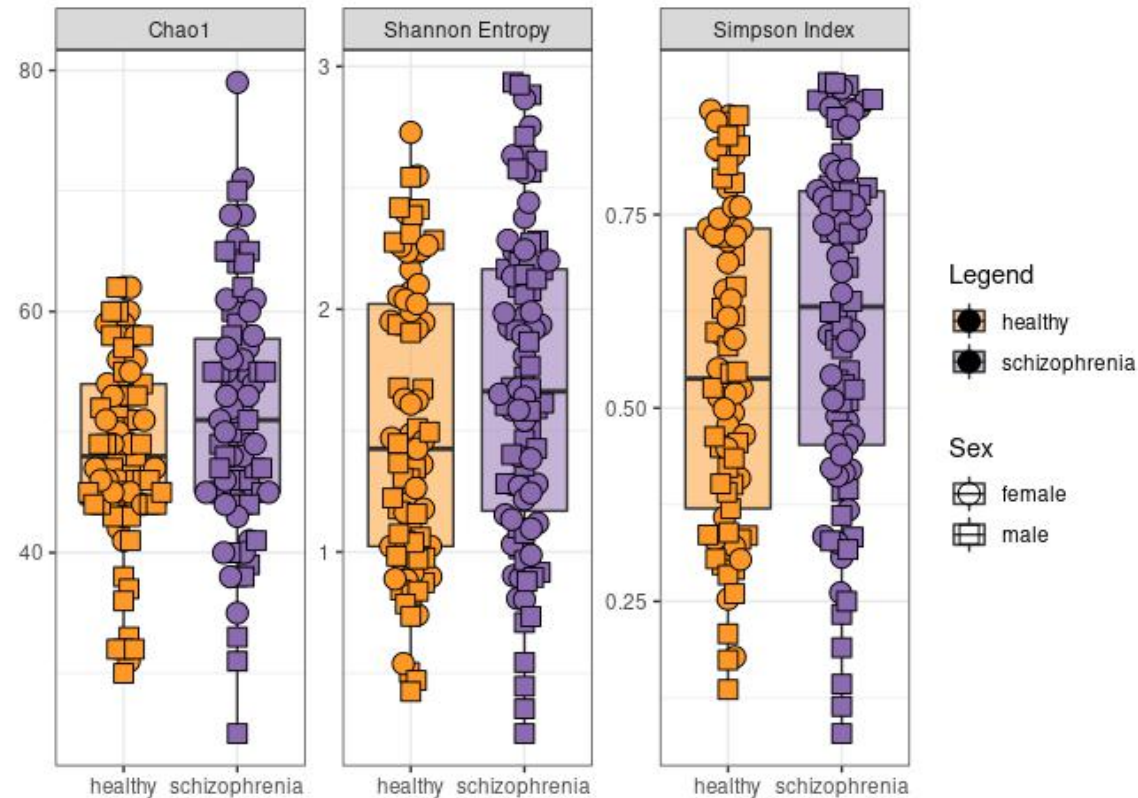


Fig 3. Exemplo de boxplot do resultado de uma análise de Alfa Diversidade. Fonte: <https://github.com/thomazbastiaanssen/Tjazi>

Dados biológicos que podem ser analisados no R

Dados de epidemiologia;

Dados de citometria de fluxo;

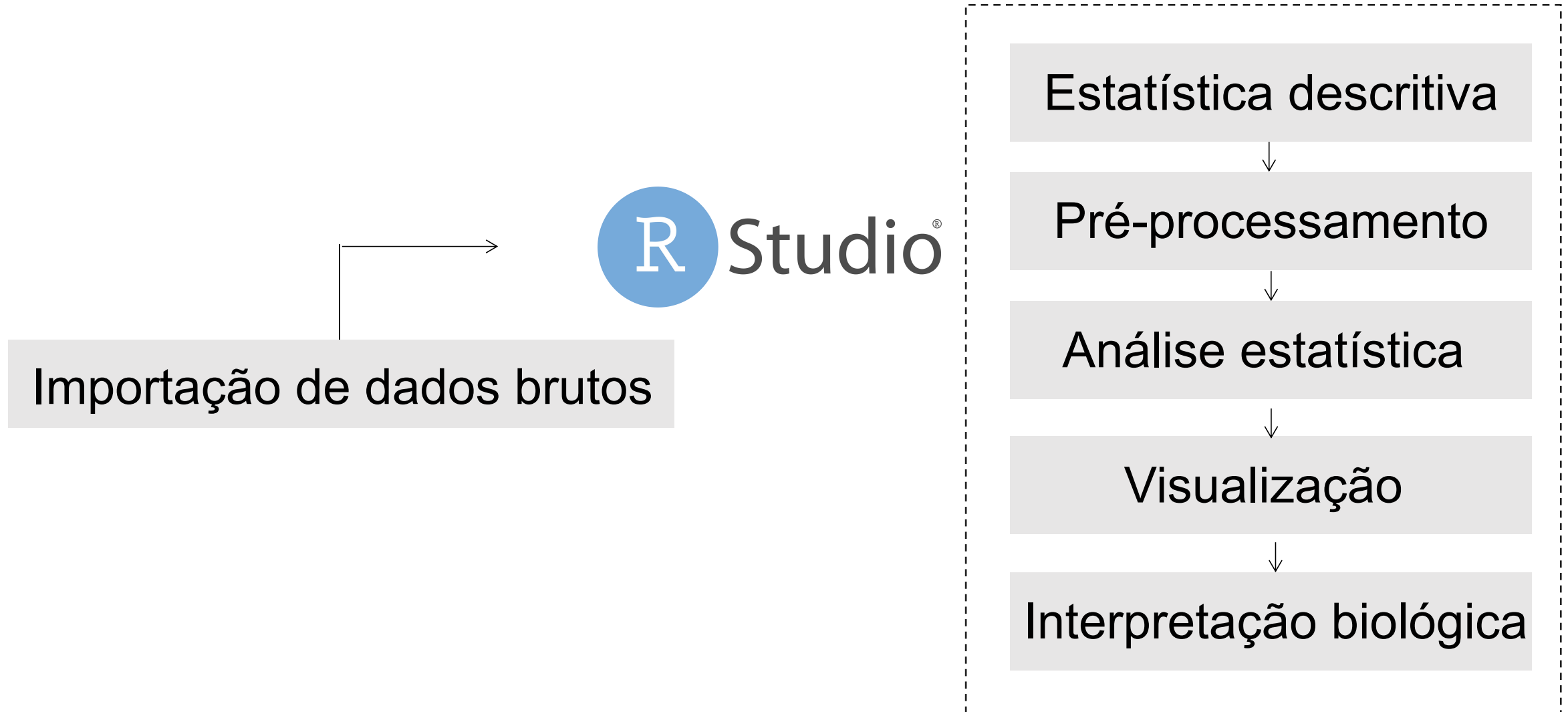
Dados de estrutura de proteínas;

Dados de ecologia;

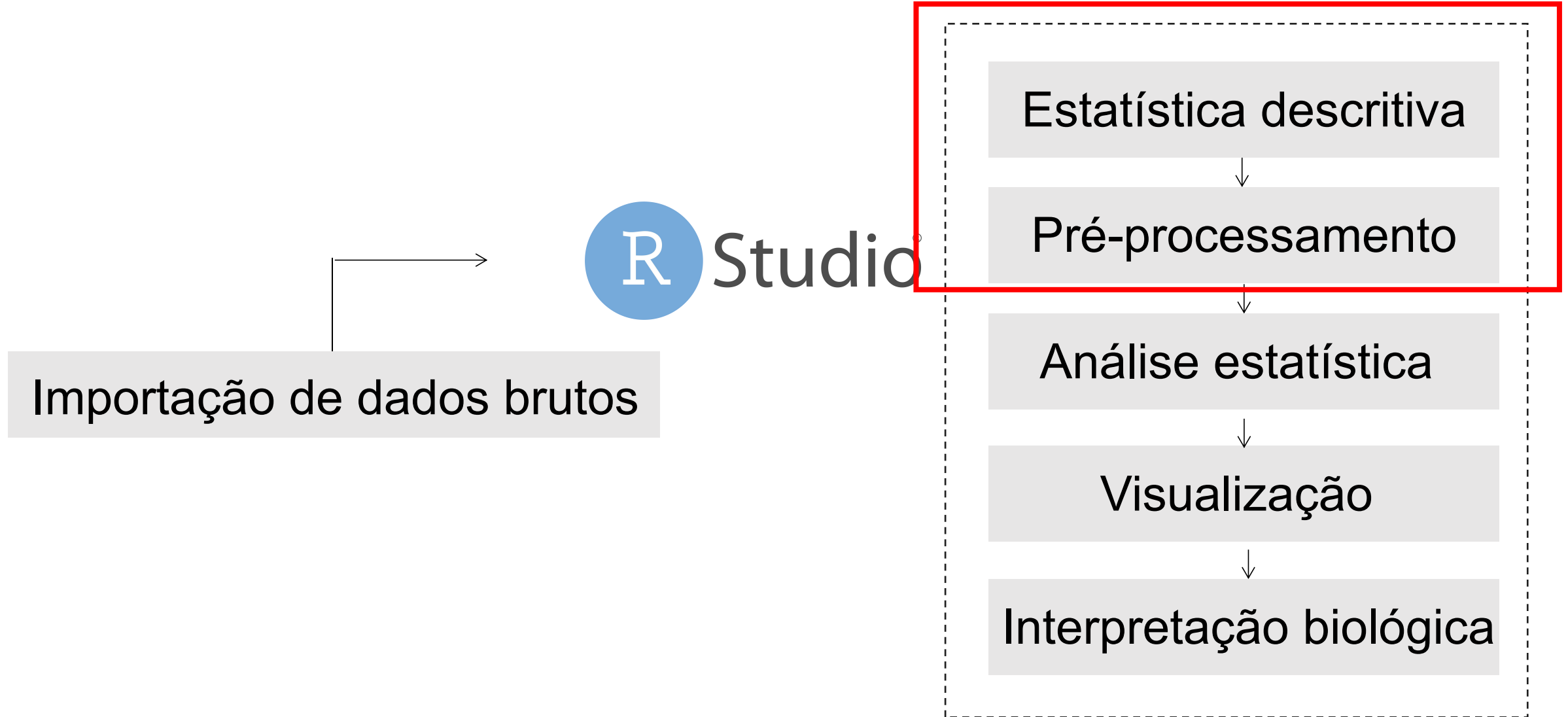
Esses são apenas alguns exemplos e o R é muito flexível para lidar com uma ampla gama de dados biológicos. A escolha dos pacotes específicos dependerá dos tipos de análises que você deseja realizar em seus dados.

Neste curso: Iremos analisar dados taxonômicos provenientes de trabalho de metagenômica.

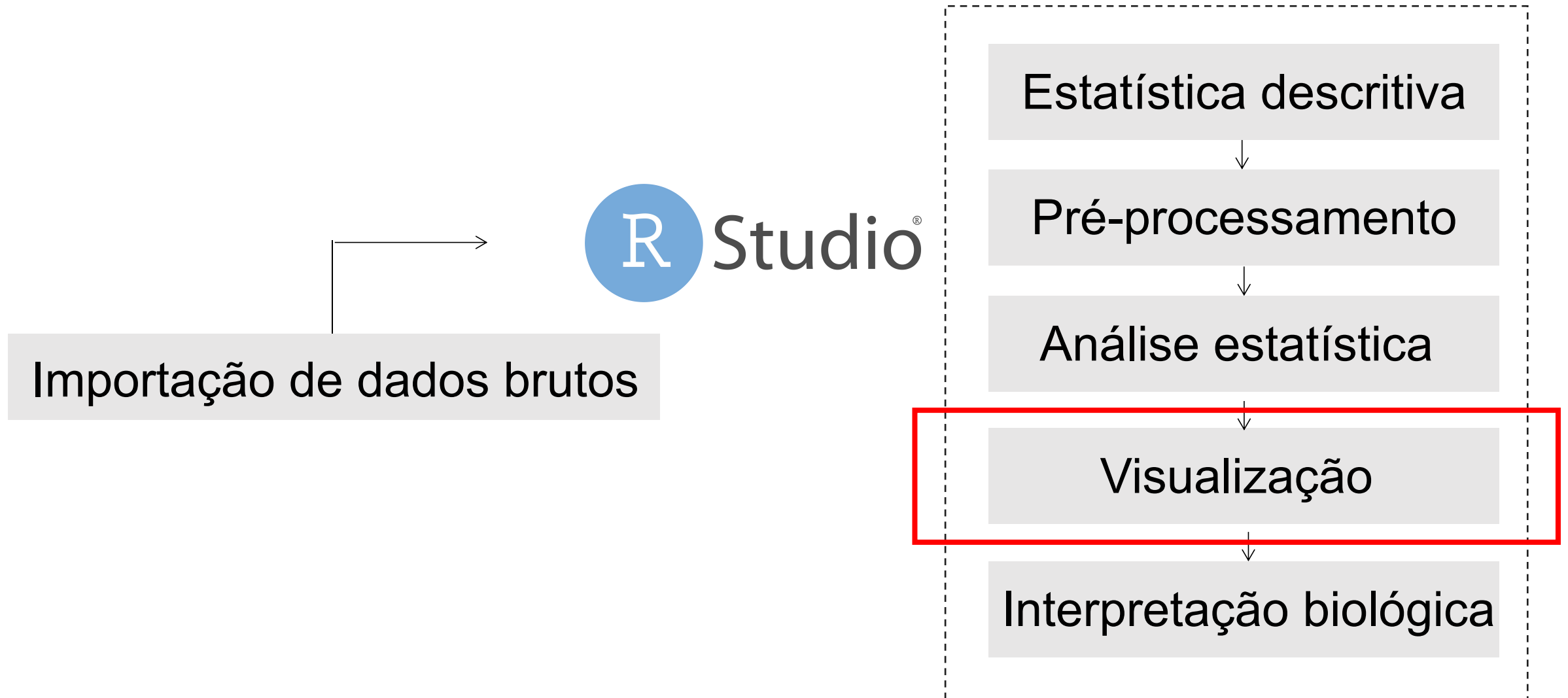
Fluxograma básico de análise



Fluxograma básico de análise



Fluxograma básico de análise



Vamos para a prática??

