

INTRODUÇÃO AO



Monitoria I de Técnicas de Pesquisa em Economia

André Vitor Pereira de Melo

andre.melo@ufdpar.edu.br

Universidade Federal do Delta do Parnaíba

16 de maio de 2025



<https://github.com/drewmelo/introducao-R>



I - INTRODUÇÃO AO R

<https://github.com/drewmelo/introducao-R>



O QUE É A LINGUAGEM R?

- Ferramenta para análise de dados;
 - R é uma linguagem de programação orientada a objetos, criada para quem trabalha com **análise de dados e estatística** de forma estruturada.
- Versátil e com ampla aplicação;
 - Oferece ferramentas que permitem lidar com **grandes conjuntos de dados**, o que é muito útil em contextos atuais que envolvem *big data*.
- Tipagem fraca.
 - Permite operações entre diferentes tipos de dados sem a necessidade explícita de conversão.

POR QUE USAR O R?

- Exploração e visualização de dados;
 - A linguagem oferece recursos avançados para criar gráficos e visualizações que ajudam a entender melhor os dados e comunicar resultados de forma clara.
- *Open-source*.
 - Por ser um *software* de código aberto, possui uma comunidade ativa que **desenvolve e compartilha pacotes**, por isso, o R dispõe de uma vasta gama de aplicações e suporte colaborativo.

O QUE É O RSTUDIO?

- Ambiente integrado para programar em R;
 - O RStudio é um espaço pensado principalmente para quem usa R, reunindo tudo o que é necessário para escrever, rodar e revisar códigos com mais organização.
- Facilita o dia a dia com R.
 - Ele traz recursos como editor de scripts, gerenciador de projetos e ferramentas para depurar o código.

PASSOS DA INSTALAÇÃO

Visite [CRAN - The Comprehensive R Archive Network](#) e baixe a linguagem, através de um arquivo executável, para seu sistema operacional;

Visite [RStudio Desktop](#) e baixe o instalador para seu sistema operacional;

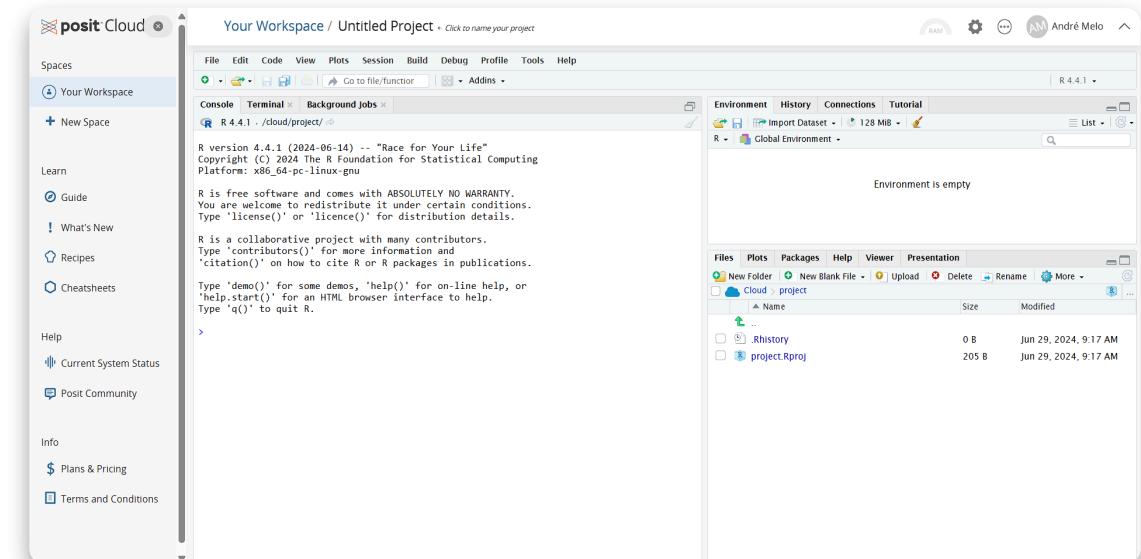
Siga as instruções fornecidas nos instaladores para completar o processo de instalação.



RSTUDIO CLOUD

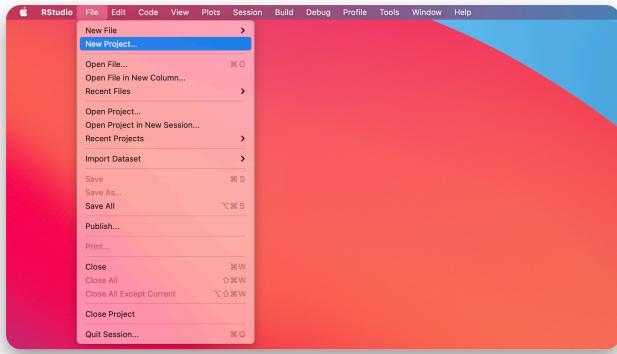
Visite a [Posit Cloud](#) para acessar e utilizar o RStudio sem a necessidade de instalação local;

Necessário realizar **login** (através de uma conta Google, Github, entre outros) para acessar o RStudio Cloud.

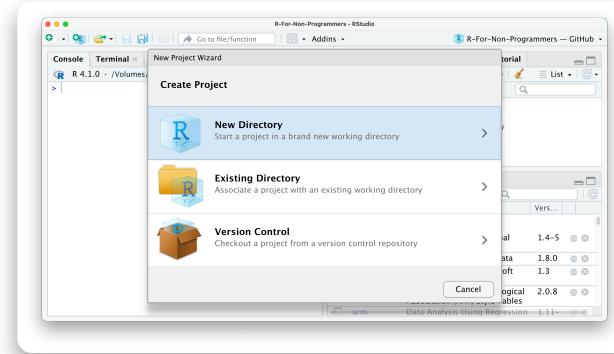


CRIAÇÃO DE UM PROJETO

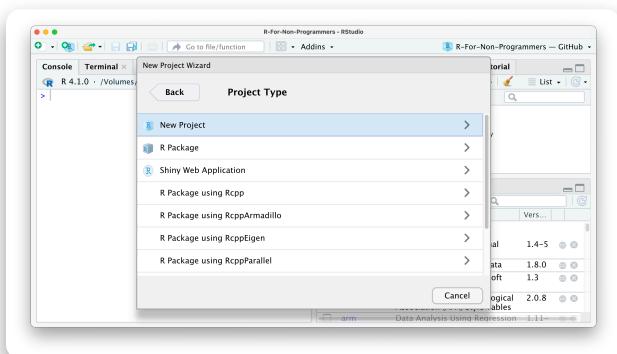
Passo 1: File > New Project.



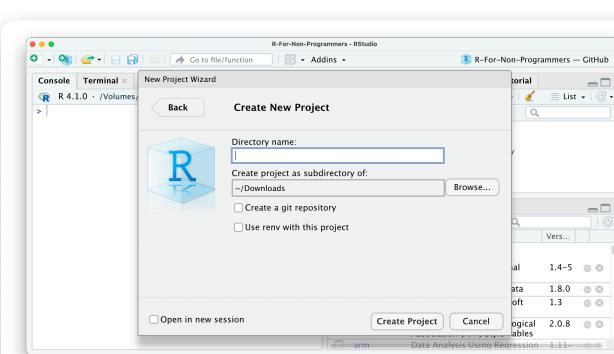
Passo 2: Escolha “New Directory”.



Passo 3: Clique em “New Project”.



Passo 4: Digite o nome da sua pasta em *Directory name*, marque “Open in new session” e clique em “Create Project”.



PACOTES

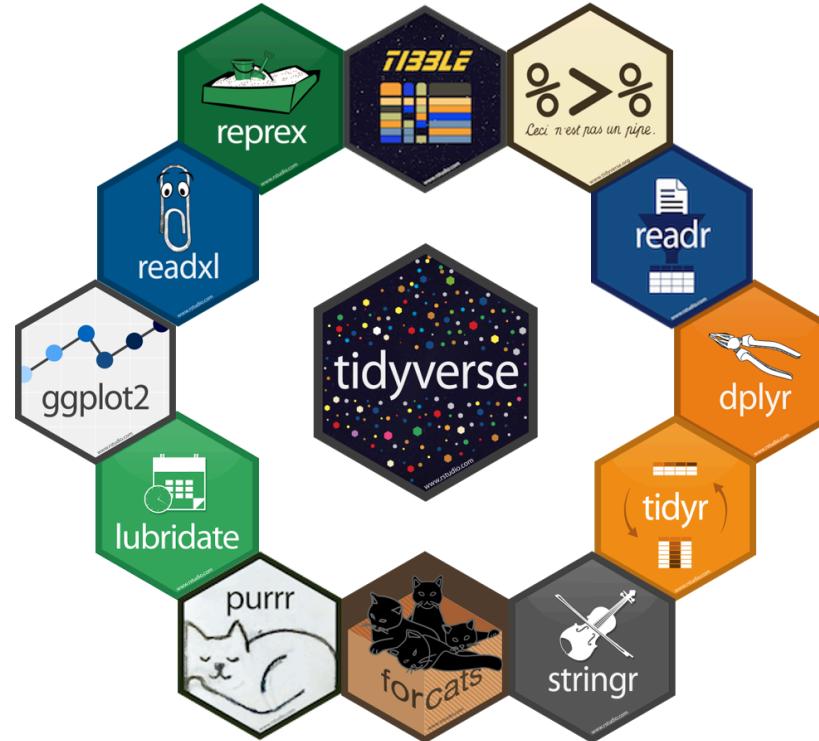
Os pacotes no R são coleções de funções e conjuntos de dados desenvolvidos pela comunidade.

```
1 # Instalando o pacote  
2 install.packages("tidyverse", dependencies = TRUE)  
3  
4 # Ativando o pacote  
5 library(tidyverse)
```

E para descarregar ou remover pacotes:

```
1 # Desativar pacotes  
2 detach("tidyverse")  
3  
4 # Desinstalar pacotes  
5 remove.packages("pacotes")
```

UM POUCO SOBRE TIDYVERSE



Fonte: Smith (2019).

UM POUCO SOBRE TIDYVERSE

O **tidyverse** oferece uma alternativa aos métodos base do R para análise de dados em dados organizados de forma “arrumada” (*tidy data*).

```
1 library(tidyverse)
— Attaching core tidyverse packages ————— tidyverse 2.0.0 —
✓ dplyr     1.1.4    ✓ readr     2.1.5
✓forcats   1.0.0    ✓ stringr   1.5.1
✓ ggplot2   3.5.1    ✓ tibble    3.2.1
✓ lubridate 1.9.3    ✓ tidyr     1.3.1
✓ purrr    1.0.2
— Conflicts ————— tidyverse_conflicts() —
✗ dplyr::filter() masks stats::filter()
✗ dplyr::lag()   masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
errors
```

OPERADOR PIPE

O operador pipe no R é uma ferramenta poderosa para encadear operações e facilitar a leitura e escrita de código. Existem dois operadores de pipe comuns:

- Operador `%>%` (pipe do pacote `magrittr`)
 - O operador `%>%` é usado para encadear funções, passando o resultado de uma função como argumento para a próxima função.
- Operador `|>` (pipe do pacote base)
 - Introduzido no R 4.1.0, o operador pipe nativo `|>` é uma alternativa ao operador do `magrittr`, oferecendo uma maneira semelhante de encadear operações.

R COMO CALCULADORA

A linguagem R nos permite realizar cálculos simples e complexos, fornecendo operações aritméticas como adição, subtração, multiplicação, divisão e outras operações matemáticas.

```
1 # Operações Combinadas  
2 (5 + 3) * 2 - 4 / 2
```

```
[1] 14
```

- Potência

```
1 2^3
```

```
[1] 8
```

R COMO CALCULADORA

- Módulo

```
1 10 %% 3
```

```
[1] 1
```

- Divisão inteira

```
1 10 %/% 3
```

```
[1] 3
```

- Raiz Quadrada

```
1 sqrt(16)
```

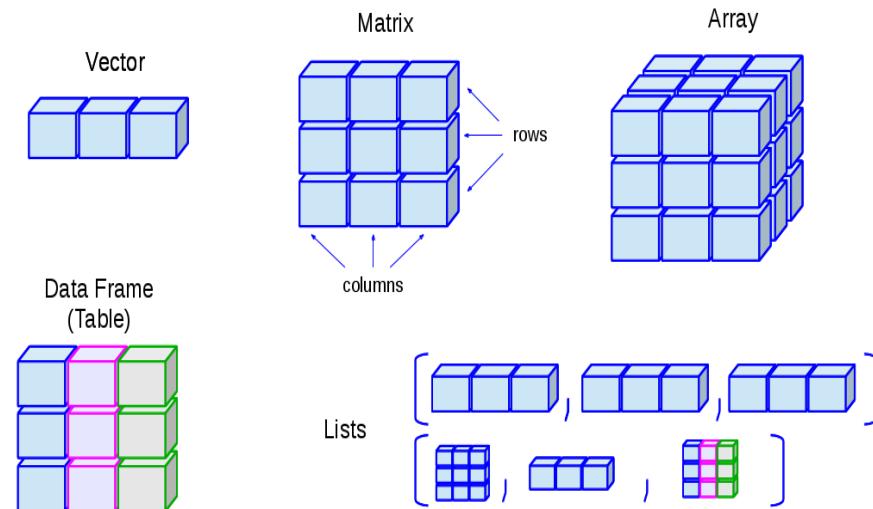
```
[1] 4
```

II - ESTRUTURAS E MANIPULAÇÕES DE DADOS

TIPOS DE ESTRUTURAS DE DADOS

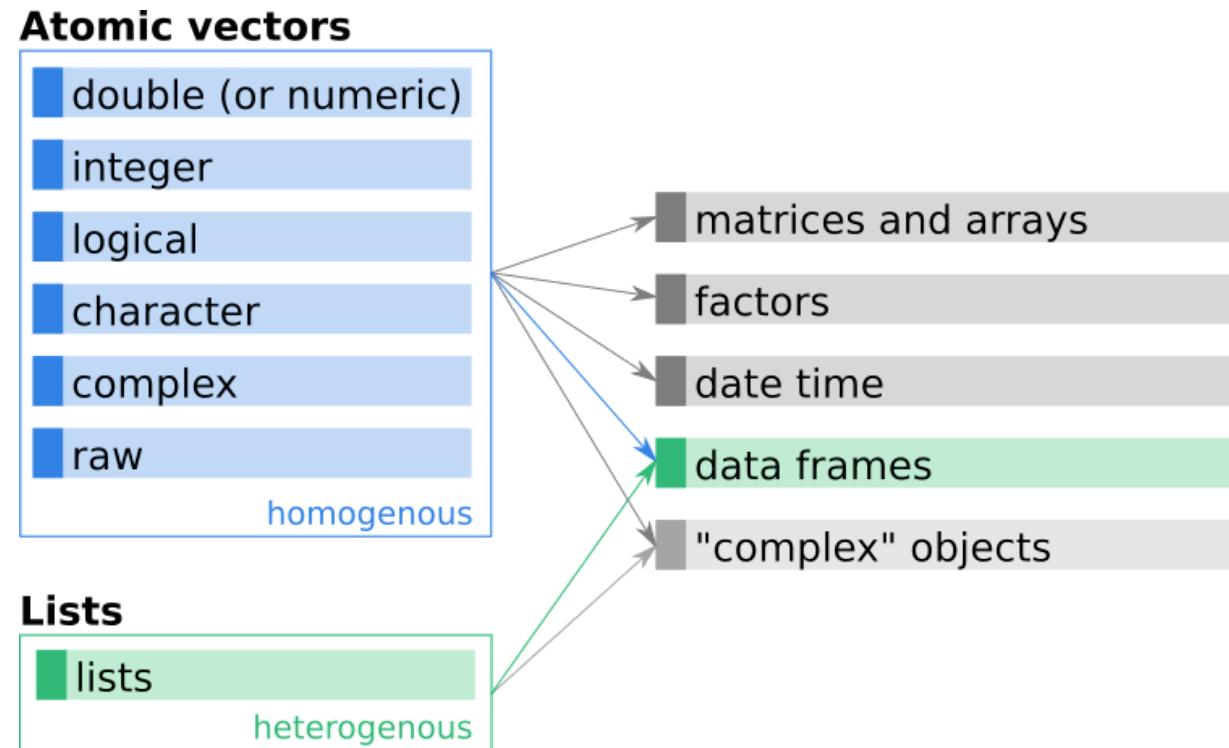
Na linguagem R, há várias estruturas de dados fundamentais que são usadas para armazenar e manipular informações de maneiras específicas.

- **Vetores:** Unidimensionais;
- **Matrizes:** Bidimensionais;
- **Data Frames:** Bidimensionais;
- **Arrays:** Multidimensionais;
- **Listas:** Flexíveis.



Fonte: Ceballos e Cardiel (2013).

TIPOS DE ESTRUTURAS DE DADOS



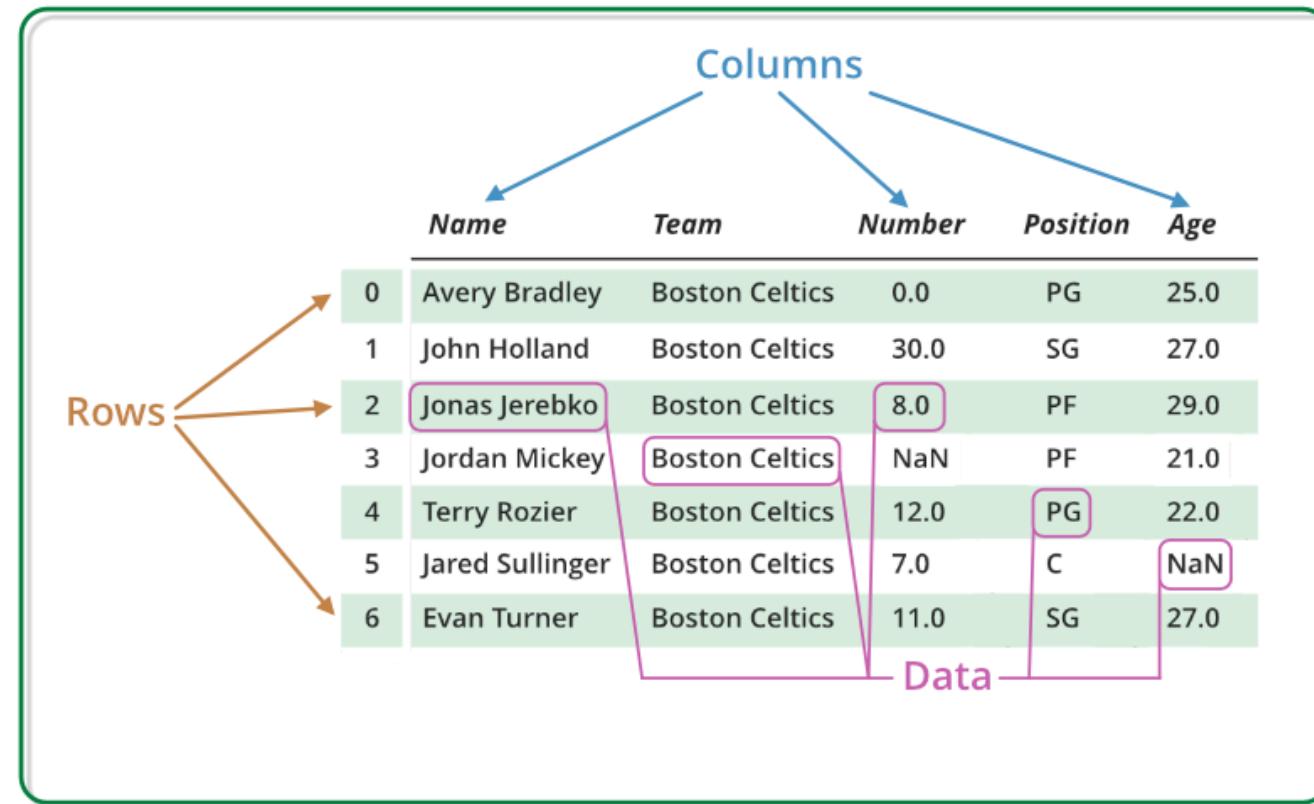
Fonte: Stauffer *et al.* (2025).

O QUE SÃO DATA FRAMES?

Os *data frames* são estruturas de dados fundamentais no R, muito utilizadas para **armazenar conjuntos de dados tabulares**, onde as colunas podem conter **diferentes tipos de classes** (numéricos, caracteres, lógicos, etc.).

- Quanto a sua estrutura, temos:
 - **Nome das Colunas (variáveis)**: Cada coluna em um *data frame* tem um nome que a identifica, representando seus respectivos atributos;
 - **Rótulos de Linhas (observações)**: As linhas podem ser rotuladas para identificar cada observação de maneira única ou significativa.

O QUE SÃO DATA FRAMES?



Fonte: GeeksforGeeks (2025).

O QUE SÃO VETORES?

Um vetor é uma estrutura de dados que contém elementos do mesmo tipo.
Pode ser numérico, de caracteres, lógico, etc.

```
1 # Vetor numérico (inteiro)
2 ano <- c(2022, 2023, 2024, 2025)
3
4 # Vetor de caracteres
5 revistas <- c("A", "B", "C", "B")
6
7 # Vetor numérico (números reais)
8 impacto <- c(2.75, 1.91, 5.23, 0.87)
```

CRIANDO DATA FRAMES

Podemos combinar os objetos `ano`, `revistas` e `impacto` em um *data frame*.

```
1 df <- data.frame(  
2   autor = c("Ana", "João", "Gabriel", "Ana"),  
3   revistas = revistas,  
4   ano = ano,  
5   impacto = impacto  
6 )  
7  
8 df
```

	autor	revistas	ano	impacto
1	Ana	A	2022	2.75
2	João	B	2023	1.91
3	Gabriel	C	2024	5.23
4	Ana	B	2025	0.87

DATA FRAMES MODERNIZADOS

Tibbles são *data frames* modernizados, introduzidos pelo pacote **tibble** no **tidyverse**. Eles foram projetados para corrigir algumas das limitações dos *data frames* tradicionais e para melhorar a usabilidade.

```
1 dados <- tibble::as_tibble(df)
2
3 dados
```

```
# A tibble: 4 × 4
  autor  revistas    ano  impacto
  <chr>   <chr>    <dbl>    <dbl>
1 Ana      A        2022     2.75
2 João     B        2023     1.91
3 Gabriel C        2024     5.23
4 Ana      B        2025     0.87
```

CRIANDO LISTAS

Uma lista é uma coleção ordenada de elementos que podem ser de diferentes tipos. Diferente de vetores, que só podem conter elementos de um único tipo, listas podem armazenar números, caracteres, vetores, *data frames* e até mesmo outras listas.

```
1 lista <- list(  
2     dados,  
3     hobbies = c("Leitura", "Esportes", "Música"),  
4     cars = head(cars, 5),  
5     mtcars = head(mtcars, 5)  
6 )
```

O QUE É O *DPLYR*?

- Parte do **tidyverse**, o **dplyr** é um pacote do R que incorpora uma gramática para transformação de dados, ou seja, fornece funções (verbos) para fazer operações com tabelas (isto é, *data frames* ou *tibbles*);
- Ele permite filtrar, selecionar, organizar, transformar e resumir dados de forma encadeada. E qual a finalidade disso? **um código limpo**.
- Principais funções (verbos)

filter(): filtra linhas (condições)

select(): seleciona colunas

mutate(): cria/edita colunas

arrange(): ordena linhas

summarise() ou **reframe()**:

resume dados

group_by(): agrupa dados para
sumarização

OPERADORES LÓGICOS

- Operadores Relacionais:
 - Igual a (`==`);
 - Diferente de (`!=`);
 - Maior que (`>`);
 - Menor que (`<`);
 - Maior ou igual a (`>=`);
 - Menor ou igual a (`<=`);
 - Pertence a (`%in%`).
- Também podemos utilizar operadores lógicos, tais como **AND**, **OR** e **NOT**, representados por `&`, `|` e `!`, respectivamente.

UTILIZANDO DPLYR NA PRÁTICA

Com os nossos **dados**, podemos filtrar artigos a partir de 2023, calcular a média de impacto por revista e ordenar do maior para o menor.

```
1 dados |>
2   filter(ano >= 2023) |>
3   group_by.revistas |>
4   summarise(media_impacto = mean(impacto)) |>
5   arrange(desc(media_impacto))
```

```
# A tibble: 2 × 2
  revistas media_impacto
  <chr>        <dbl>
1 C            5.23
2 B            1.39
```

EXPORTAÇÃO DE DADOS

Exportar dados é uma tarefa comum no R, especialmente quando se deseja salvar resultados de análises ou compartilhar informações com outros usuários.

1. Exportando para CSV

```
1 write.csv(dados, file = "dados/dados.csv", row.names = FALSE)
```

2. Exportando para Excel (xlsx)

```
1 # Ativando pacote
2 library(openxlsx) # Necessário instalar
3
4 write.xlsx(dados, file = "dados/dados.xlsx", rowNames = FALSE)
```

3. Exportando para Texto (txt)

```
1 write.table(dados, file = "dados/dados.txt", sep = "\t", row.names = FALSE)
```

IMPORTAÇÃO DE DADOS

Neste caso, ao importar dados no R, estaremos carregando conjuntos de dados externos para manipulação e/ou visualização.

1. Importando para CSV

```
1 readr::read_csv(file = "dados/dados.csv")
```

2. Importando para Excel (xlsx)

```
1 # Ativando o pacote
2 library(readxl) # Necessário instalar
3
4 readxl::read_excel(path = "dados/dados.xlsx")
```

3. Importando para Texto (txt)

```
1 read.table(file = "dados/dados.txt", header = TRUE, sep = "\t")
```

EXERCÍCIO DE FIXAÇÃO I

Hora de praticar!

Acesse o arquivo **main.pdf**, que está dentro da pasta *atividade*, localizada em “*tecnicas-pesquisa-economia*” no repositório *monitorias-ufdpar*, ou clique [aqui](#) para acessá-lo diretamente.

III - BIBLIOMETRIA COM O R

O QUE É BIBLIOMETRIA?

Bibliometria é um campo das áreas da biblioteconomia e da ciência da informação que emprega o estudo quantitativo da produção científica, através de métodos estatísticos e matemáticos para analisar publicações acadêmicas, como **artigos**, livros, teses e conferências.

- Exemplos de estudos bibliométricos:
 - **Contagem de publicações:** Quantos artigos sobre “sustentabilidade” foram publicados entre 2010 e 2020?
 - **Análise de citações:** Qual é o artigo **mais citado** sobre Inteligência Artificial?
 - **Análise de palavras-chave:** Quais são os termos **mais recorrentes** nas pesquisas sobre Mudanças Climáticas?

E QUAL A SUA CONTRIBUIÇÃO?

Aplicação	Descrição
Mapeamento de áreas de conhecimento	Identificar tendências, tópicos emergentes ou temas em declínio em uma disciplina.
Análise de tendências científicas	Ver como um tema ganha ou perde relevância ao longo do tempo.
Avaliação de desempenho acadêmico	Avaliar a produtividade de autores, universidades ou países (por número de publicações, citações etc.).
Identificação de redes de colaboração	Descobrir grupos de pesquisa que atuam juntos (análise de coautoria).
Tomada de decisão em políticas científicas	Ajudar agências de fomento e universidades a definir prioridades de investimento em ciência.

PACOTE BIBLIOMETRIX

Assim como o **dplyr**, mencionado anteriormente, o **bibliometrix** também é um pacote do R criado para realizar análises bibliométricas completas, desde a importação dos dados até a construção de gráficos e redes científicas.

- Criado por Aria e Cuccurullo (2017), além de outros colaboradores a partir de 2017.
 - Objetivo: fornecer uma solução *open-source* para análise bibliométrica e de ciência da ciência (*science mapping*);
- Fontes de dados compatíveis:

→ Web of Science (WoS)	PubMed	Coleta manual em formato BibTeX ou CSV
→ Scopus	CrossRef	
Dimensions.ai	OpenAlex (mais recente)	

ACESSO AS BASES DE DADOS ACADÊMICAS

Para realizar análises bibliométricas no R, primeiro é necessário baixar os dados de bases acadêmicas confiáveis.

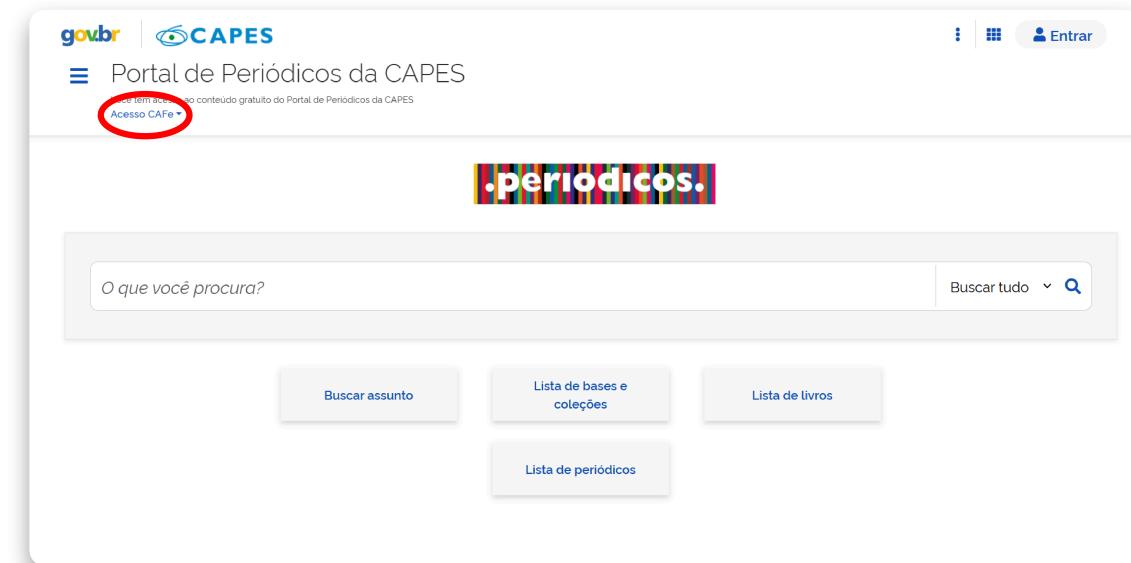
No Brasil, temos acesso gratuito a bases internacionais de alta qualidade por meio da:

- Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)
 - O Portal de Periódicos da CAPES oferece acesso gratuito para estudantes, professores e pesquisadores de instituições públicas e privadas conveniadas.
 - Por meio dele, é possível acessar bases como WoS, Scopus, ScienceDirect, SpringerLink, entre muitas outras.

COMO ACESSAR O PORTAL DE PERIÓDICOS DA CAPES?

Estando dentro da universidade (rede da instituição) ou

Pelo sistema de CAFé (Comunidade Acadêmica Federada), usando login e senha institucionais.

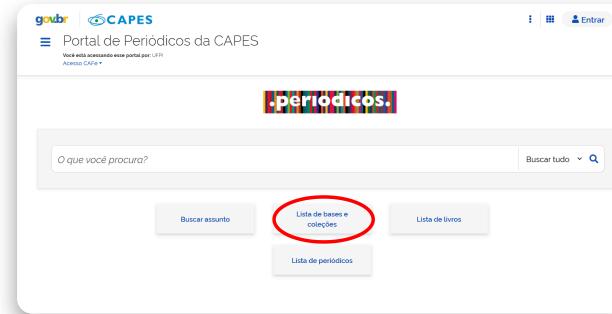


PASSOS PARA BAIXAR DADOS BIBLIOGRÁFICOS

Passo 1: Acesso CAFé (login) > UFPI.



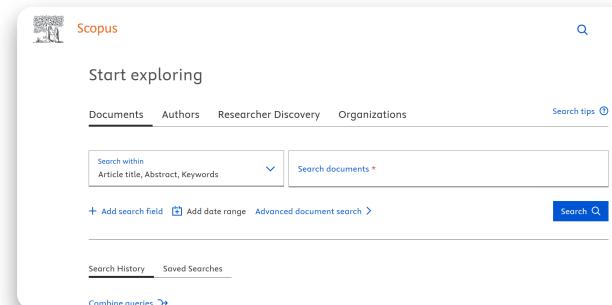
Passo 2: Clique em “Lista de bases e coleções”.



Passo 3: Após a busca pela base, clique em “Ver no editor”.

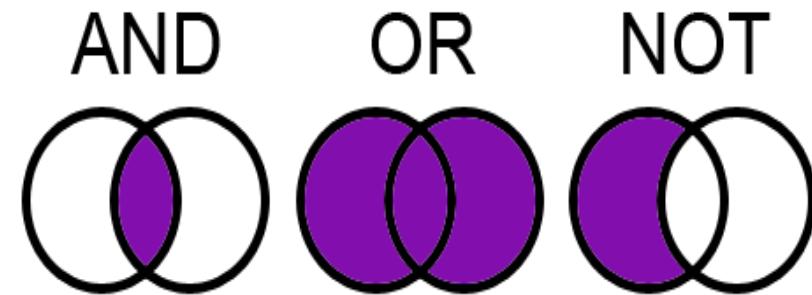


Passo 4: Interface da Scopus.



OPERADORES BOOLEANOS NAS PESQUISAS ACADÊMICA

- AND: inclui todos os termos;
- OR: inclui qualquer um dos termos;
- NOT: exclui termos indesejados



Ex.: Busca artigos que falem de Inflation e de pelo menos um dos temas citados.

1 (Modern Monetary Theory OR Post-Keynesian Economics) AND Inflation

PESQUISA AVANÇADA EM BASES ACADÊMICAS

1. Web of Science

Permite criar pesquisas mais específicas usando filtros como tema, autor, instituição, tipo de documento, ano de publicação etc.

```
1 TS=("Modern Monetary Theory") AND PY=(2015-2024)
```

2. Scopus

Assim como na Web of Science, permite construir consultas específicas, com vários critérios.

```
1 TITLE-ABS-KEY("Modern Monetary Theory") AND PUBYEAR > 2014 AND PUBYEAR < 2025
```

IMPORTAÇÃO DE DADOS BIBLIOGRÁFICOS

A função `convert2df()` é usada para ler os dados exportados das bases e transformá-los em data frames organizados, prontos para análise.

1. Importar dados da Web of Science (.txt)

```
1 library(bibliometrix)
2
3 df_wos <- convert2df(
4   file = "tecnicas-pesquisa-economia/dados/savedrecs.txt",
5   dbsource = "wos",
6   format = "plaintext"
7 )
```

2. Importar dados da Scopus (.csv)

```
1 df_scopus <- convert2df(
2   file = "tecnicas-pesquisa-economia/dados/scopus.csv",
3   dbsource = "scopus",
4   format = "csv"
5 )
```

IMPORTAÇÃO DE DADOS BIBLIOGRÁFICOS

Após importar os dados da Web of Science e Scopus separadamente, como demonstrado em 1. e 2., podemos combinar estes dados em um único data frame.

```
1 dados <- mergeDbSources(df_wos, df_scopus, remove.duplicated = TRUE)
```

SOBRE OS DADOS GERADOS PELO BIBLIOMETRIX

Quando importamos e transformamos os dados com `convert2df()`, o `bibliometrix` organiza as informações em uma tabela (data frame) com linhas e colunas.

Rows: 477

Columns: 8

```
$ PY    <dbl> 2024, 2023, 2024, 2024, 2017, 2023, 2018, 2024, 2015, 2020, 2022...
$ AU    <chr> "WILDOWICZ-SZUMARSKA A;OWSIAK K", "CEVIK S;JALLES J", "HUYNH C;H...
$ TI    <chr> "IMPACT OF CLIMATE CHANGE ON INCOME INEQUALITY IMPLICATIONS FOR ...
$ SO    <chr> "ECONOMICS AND ENVIRONMENT", "ENERGY POLICY", "JOURNAL OF THE AS...
$ DE    <chr> "CLIMATE CHANGE; INCOME INEQUALITY; ECONOMIC GROWTH", "INCOME IN...
$ TC    <dbl> 1, 29, 4, 11, 22, 2, 24, 2, 58, 4, 8, 21, 6, 5, 0, 11, 19, 1, 0, ...
$ DT    <chr> "ARTICLE", "ARTICLE", "ARTICLE; EARLY ACCESS", "ARTICLE", "ARTIC...
$ AU_UN <chr> "UNIVERSITY OF BIALYSTOK;CRACOW UNIVERSITY OF ECONOMICS", "INTER...
```

- **Esses são apenas alguns dos campos mais usados.** O glossário completo dos códigos de campos (Field Tags) pode ser acessado [aqui](#).

ANÁLISE BIBLIOMÉTRICA BÁSICA

A função `biblioAnalysis()` realiza uma análise exploratória automática dos dados bibliográficos importados.

- Esta função gera várias estatísticas descritivas, como:

→ Número de autores

→ Número de documentos

→ Revistas que mais publicaram

→ Autores mais produtivos

→ Citações recebidas

→ Palavras-chave mais usadas

```
1 resultado <- biblioAnalysis(dados)
```

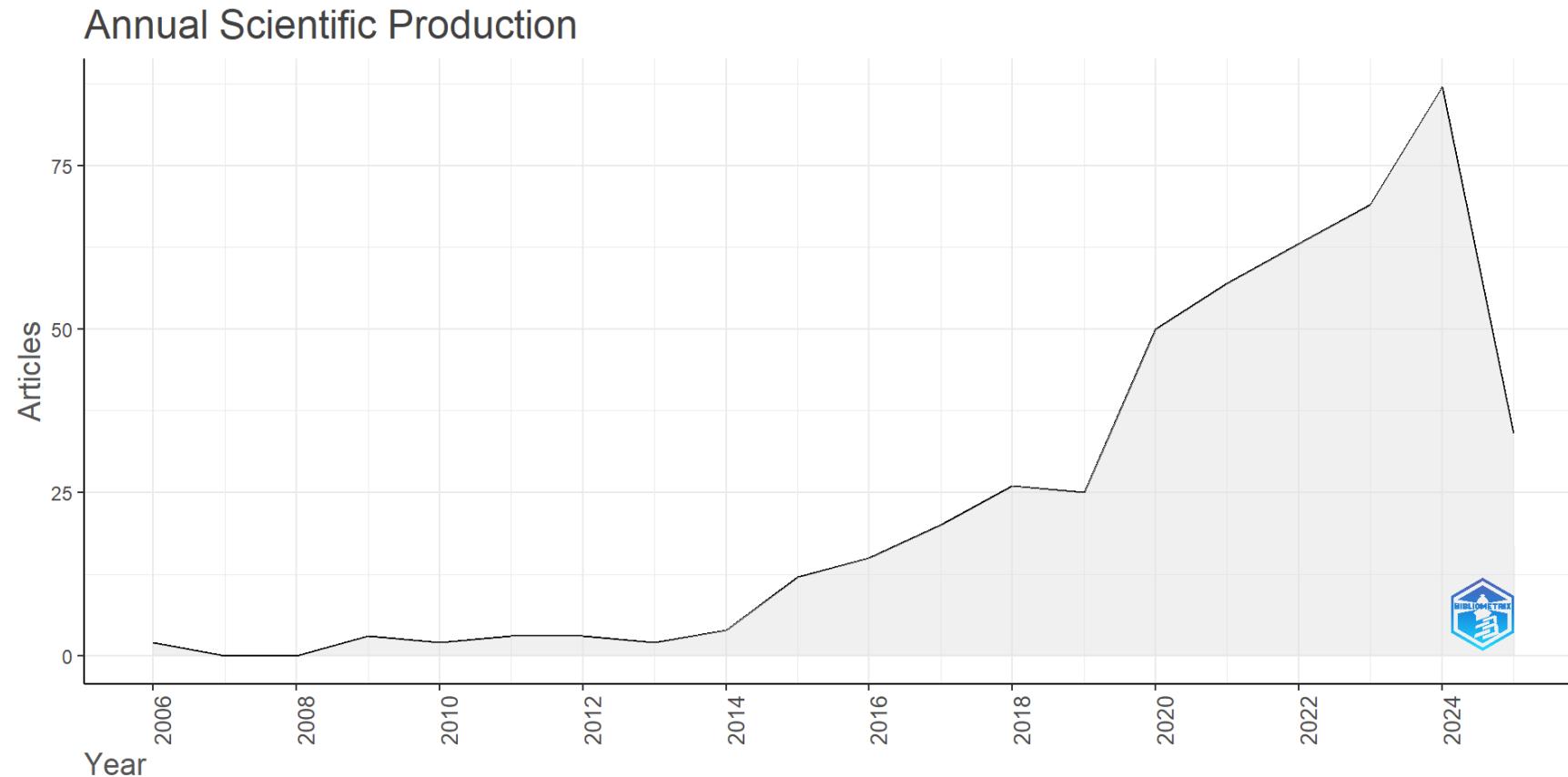
- Depois de fazer a análise, pode-se usar `summary()` para resumir e visualizar as principais informações.

```
1 summary(resultado, k = 10)
```

VISUALIZAÇÃO DOS RESULTADOS

Após realizar a análise bibliométrica básica, pode-se visualizar os principais resultados de forma gráfica usando a função `plot()`.

```
1 plot(resultado)
```



LEITURA NO BIBLIOSHINY

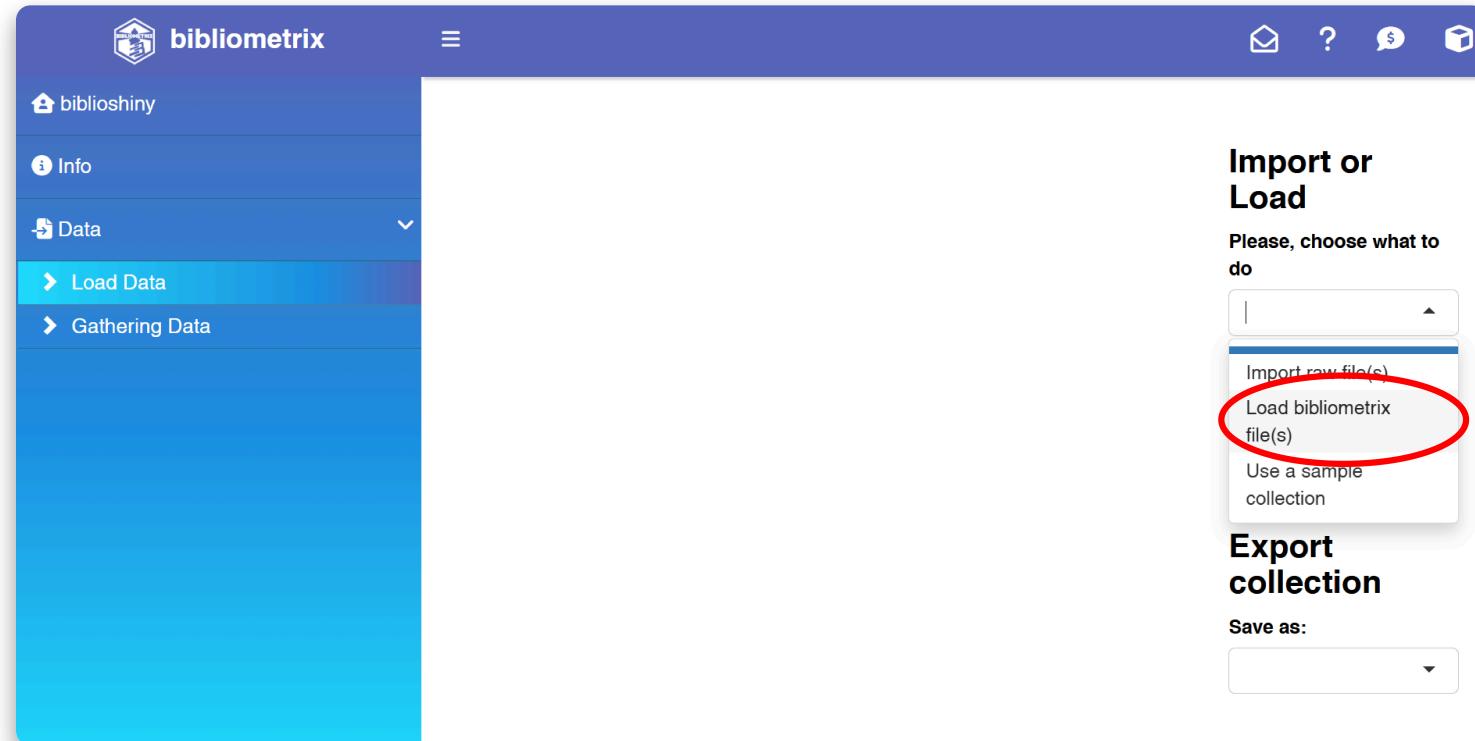
Por que exportar a base?

- O Biblioshiny (interface gráfica do bibliometrix) permite carregar dados manualmente a partir de um arquivo .xlsx.
- Assim, se quisermos analisar uma base já tratada no R dentro do Biblioshiny, precisamos exportá-la para o formato Excel.
 - Utilizamos a função `write.xlsx()` do pacote `openxlsx` para salvar o data frame como um arquivo .xlsx.

```
1 write.xlsx(x = dados,  
2             file = "dados/dados.xlsx",  
3             rowNames = FALSE)
```

LEITURA NO BIBLIOSHINY

Na interface do biblioshiny (biblioshiny) > Load Data > Load Bibliometrix file(s) > Arquivo exportado em .xlsx



EXERCÍCIO DE FIXAÇÃO II

Utilizando a palavra-chave “*Climate Change Policy*”, e considerando um escopo temporal dos últimos 10 anos (de 2015 a 2024), realize a atividade proposta a seguir.

- a. Acesse uma das bases de dados disponíveis, Web of Science ou Scopus, por meio do Portal de Periódicos da CAPES. Após obter os resultados da pesquisa, exporte o conjunto de dados em formato *Plaintext* (.txt), caso utilize a Web of Science, ou em formato CSV (.csv), caso utilize a Scopus.
- b. Após importar os dados para o R, utilize a função **biblioAnalysis()** para gerar uma análise bibliométrica exploratória básica. Em seguida, execute **biblioshiny()** para abrir a interface gráfica do bibliometrix.

REFERÊNCIAS

ARIA, M.; CUCCURULLO, C. bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of informetrics*, [s. l.], v. 11, n. 4, p. 959–975, 2017.

CEBALLOS, M.; CARDIEL, N. **Data structure**. [S. l.]: In: First Steps in R, 2013.

GEEKSFORGEEKS. DataFrame Operations in R. [S. l.: s. n.], 2025. Disponível em:
<https://www.geeksforgeeks.org/dataframe-operations-in-r/>.

SMITH, Z. M. **R in Practice**. [S. l.: s. n.], 2019.

STAUFFER, R.; CHIMIAK-OPOKA, J.; RODRÍGUEZ-R, L. M.; SIMON, T.; ZEILEIS, A. **Introduction to Programming with R**. [S. l.]: Digital Science Center, Universität Innsbruck, 2025. Disponível em:
<https://rpgdown.org/rprogramming/index.html>.