

# Minicurso

## R Básico

João Matheus  
Lineu Alberto

PET - Estatística  
Universidade Federal do Paraná

2019

# Sumário

- 1 O Curso
- 2 Dinâmica
- 3 O R
- 4 O Rstudio
- 5 Funções e Pacotes
- 6 Cheatsheet
- 7 Rmarkdown
- 8 Links úteis

# O PET

O Programa de Educação Tutorial (PET) é desenvolvido por grupos de estudantes, com tutoria de um docente orientados pelo princípio da indissociabilidade entre ensino, pesquisa e extensão.

São objetivos do PET:

- Desenvolver atividades acadêmicas em padrões de qualidade de excelência.
- Contribuir para a elevação da qualidade da formação acadêmica dos alunos de graduação.
- Estimular a formação de profissionais e docentes de elevada qualificação técnica, científica, tecnológica e acadêmica.
- Formular novas estratégias de desenvolvimento e modernização do ensino superior no país.
- Estimular o espírito crítico, bem como a atuação profissional pautada pela ética, pela cidadania e pela função social da educação superior.



# Ministrantes

## 1 João Matheus S. K. T. Hneda

- Aluno do curso de Estatística.
- Membro do PET - Estatística.
- Estagiário no Dpto de Gestão de Dados e Algoritmos do Banco Bradesco S.A.

## 2 Lineu Alberto C. de Freitas

- Aluno do curso de Estatística.
- Membro do PET - Estatística.
- Estagiário no Dpto de Crédito do Banco Bradesco S.A.



# Objetivos

- Apresentar o software R como ferramenta para manipulação de dados e análises estatísticas.
- Proporcionar aos participantes um primeiro contato com o R de forma gradual com foco em noções básicas da linguagem.
- Tornar os participantes aptos a conduzir por si só suas análises em R.



## Datas e Horários

- 1 As aulas serão ministradas na Sala Multimídia do Departamento de Estatística da UFPR, localizada no bloco PA.
- 2 O curso terá carga horária de 12 horas, divididas em 4 aulas.
- 3 As aulas serão nos dias 27/04, 04/05, 11/05 e 28/05 das 09:00 até 12:00.



# Conteúdo Programado

## Aula 1

### Parte 1

- 1 O curso
- 2 O R
- 3 O Rstudio
- 4 Funções e Pacotes
- 5 Rmarkdown

### Parte 2

- 1 Conceito de funções
- 2 Funções básicas
- 3 Objetos e classes
- 4 Estruturas
  - Vetores
  - Matrizes
  - Dataframes
  - Listas
- 5 Leitura de dados
- 6 Gráficos simples



# Conteúdo Programado

## Aula 2

### Parte 1

- 1 if, else, for, while
- 2 Família apply
- 3 Criação de funções

### Parte 2

- 1 Introdução ao Tidyverse
- 2 %>%
- 3 readr
- 4 tibble



# Conteúdo Programado

## Aula 3

### Parte 1

- 1 XXXXXXXXXXXXXXXXXXXX
- 2 XXXXXXXXXXXXXXXXXXXX
- 3 XXXXXXXXXXXXXXXXXXXX
- 4 XXXXXXXXXXXXXXXXXXXX
- 5 XXXXXXXXXXXXXXXXXXXX

### Parte 2

- 1 XXXXXXXXXXXXXXXXXXXX
- 2 XXXXXXXXXXXXXXXXXXXX
- 3 XXXXXXXXXXXXXXXXXXXX
- 4 XXXXXXXXXXXXXXXXXXXX
- 5 XXXXXXXXXXXXXXXXXXXX



# Aulas

- As aulas serão todas ministradas no R, intercalando conceitos apresentados pelos ministrantes com atividades práticas em aula.
- Todo o material usado será disponibilizado online para consulta no [link](#).
- Ao final das 3 primeiras aulas serão realizadas pequenas avaliações com o conteúdo visto no dia.
- Haverá ainda um trabalho final varrendo todo o conteúdo do curso.
- Os participantes devidamente inscritos receberão um certificado de participação com o número de horas de aula assistidas a ser emitido pelo PET Estatística com assinatura do tutor do grupo: professor [Walmes Marques Zeviani](#).



## Antes de começarmos...

- 1 Acessem o <http://shiny.leg.ufpr.br/walmes/brainstorm/>
- 2 Digite as 3 primeiras palavras que vem à sua mente quando você pensa em R.
- 3 Vamos criar uma nuvem de palavras.



# Resultado



# O R

- O R é uma linguagem de programação voltada para análises estatísticas.
- Foi criado por Ross Ihaka e Robert Gentleman na Universidade de Auckland, Nova Zelândia.
- Começou como uma ferramenta de ensino para substituir o S-Plus.
- O R Core Team foi formado em 1997.



## O R

- Tem sido desenvolvido por um esforço colaborativo de usuários em vários locais do mundo.
- O R é um software open-source e, por isso, bastante popular no meio acadêmico e cada vez mais disseminado no mundo do Data Science.
- R é um software livre, gratuito e está disponível sob os termos da função de software livre 'GNU General Public License'.
- Pode ser instalados em computadores Linux, Windows e MacOS.



## O R

- Possui diversas técnicas estatísticas implementadas que vão desde de técnicas gráficas para análise exploratória até técnicas de modelagem uni e multivariada, séries temporais, dados longitudinais, etc.
- Além de uma série de implementações voltadas para manipulação de dados.
- A linguagem R é indicada para manipulação de conjuntos de dados de tamanho médio, análises estatísticas e produção de documentos e apresentações centradas em dados.
- O R, muitas vezes, não é considerada como uma linguagem de programação, sendo mais comparada a um produto estatístico especializado, como é o caso do SAS.



# O R

- Por ser de código aberto, fornece acesso completo aos algoritmos e sua implementação. Nada de caixas pretas.
- Por ser livre e gratuito, existem milhares de programadores e desenvolvedores com diversos fóruns e páginas de discussão sobre técnicas; desde básicas até avançadas.
- Permite Pesquisa Reproduzível, isto é, análises feitas em R podem ser reproduzidas por outros usuários com acesso ao código e aos dados
- Não é usual utilizar o R sem uma interface.



# O Rstudio

- O RStudio é uma das mais famosas interfaces R.
- É considerada como uma das mais amigáveis.
- Utilizaremos o Rstudio do começo ao fim do curso.



# O Rstudio

The screenshot displays the RStudio environment with the following components:

- Menu Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Includes icons for file operations, running code, and source navigation.
- Source Editor:** Shows a file named 'Untitled1.R' with a single line of code: `1`.
- Environment Panel:** Shows the 'Global Environment' with a table header: Name, Type, Length, Size, Value. The content below the header is 'Environment is empty'.
- Console Panel:** Displays the R startup output:
 

```
Platform: x86_64-pc-linux-gnu (64-bit)

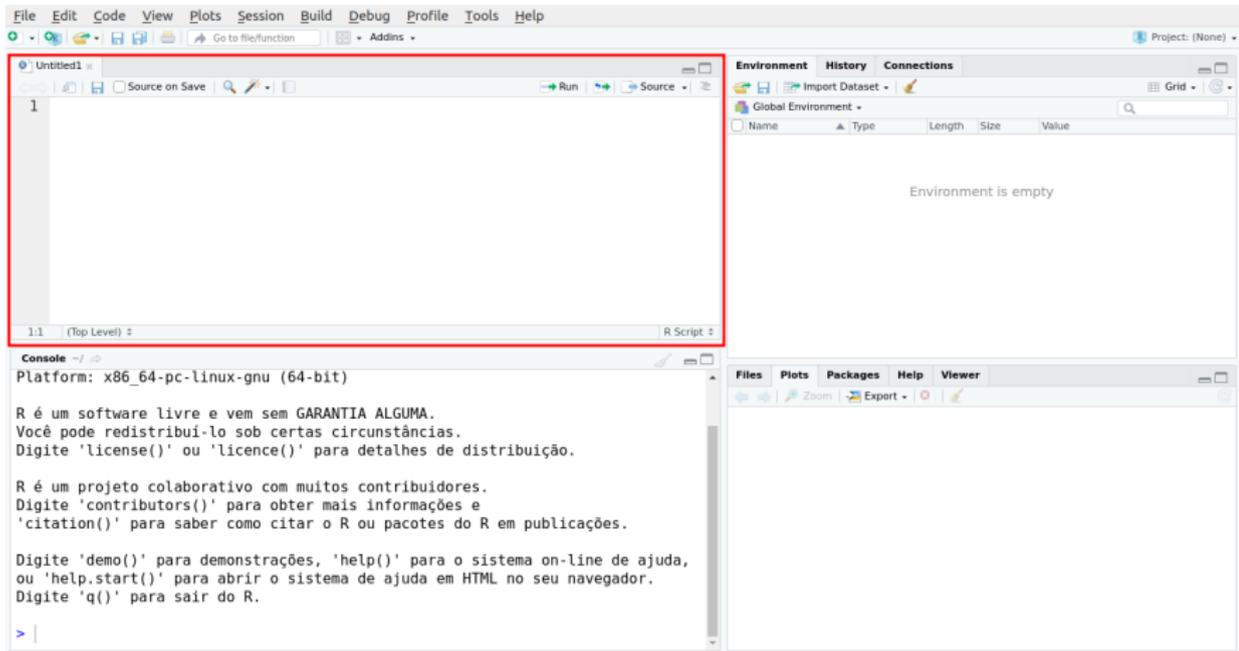
R é um software livre e vem sem GARANTIA ALGUMA.
Você pode redistribuí-lo sob certas circunstâncias.
Digite 'license()' ou 'licence()' para detalhes de distribuição.

R é um projeto colaborativo com muitos contribuidores.
Digite 'contributors()' para obter mais informações e
'citation()' para saber como citar o R ou pacotes do R em publicações.

Digite 'demo()' para demonstrações, 'help()' para o sistema on-line de ajuda,
ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.
Digite 'q()' para sair do R.

> |
```
- Bottom Panel:** Includes tabs for Files, Plots, Packages, Help, and Viewer.

# Editor



# Console

The screenshot shows the RStudio interface with the console window highlighted by a red border. The console displays the following text:

```

Platform: x86_64-pc-linux-gnu (64-bit)

R é um software livre e vem sem GARANTIA ALGUMA.
Você pode redistribuí-lo sob certas circunstâncias.
Digite 'license()' ou 'licence()' para detalhes de distribuição.

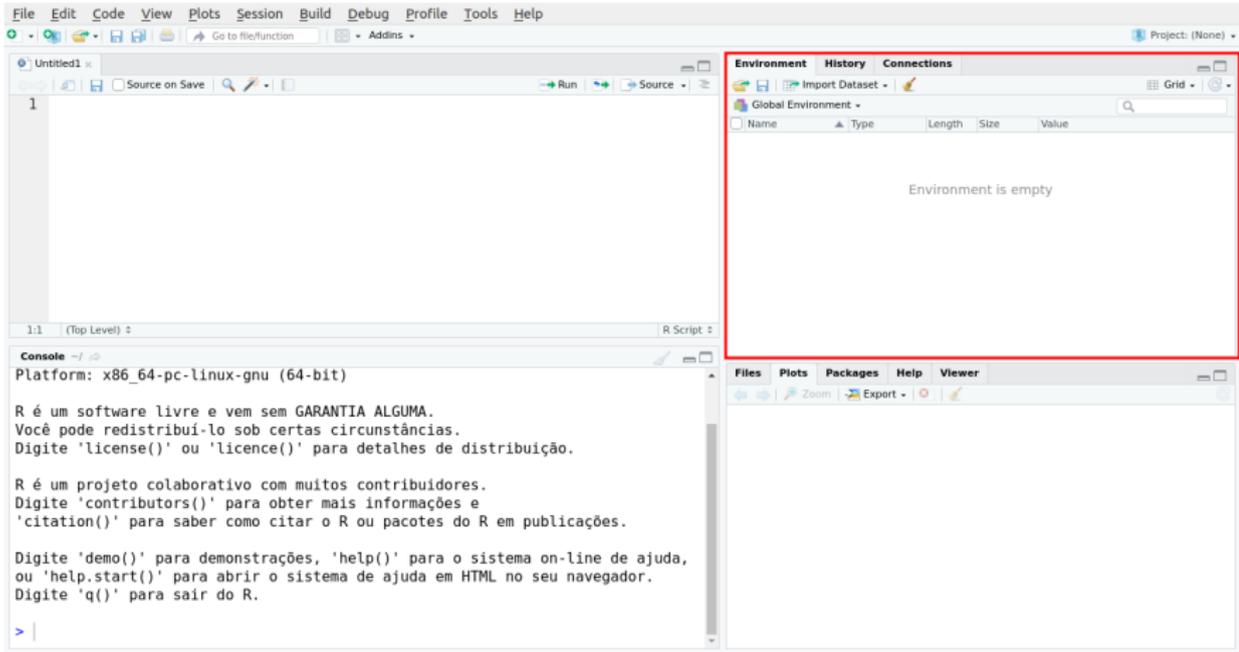
R é um projeto colaborativo com muitos contribuidores.
Digite 'contributors()' para obter mais informações e
'citation()' para saber como citar o R ou pacotes do R em publicações.

Digite 'demo()' para demonstrações, 'help()' para o sistema on-line de ajuda,
ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.
Digite 'q()' para sair do R.

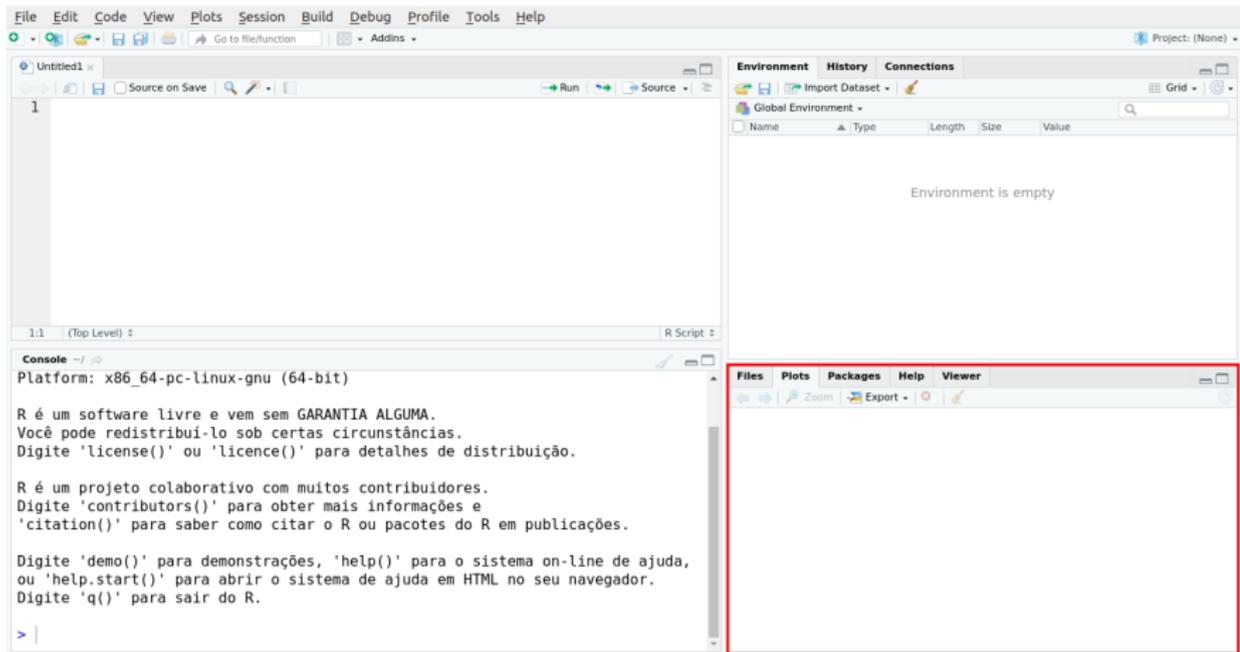
> |
  
```

The Environment pane on the right shows "Global Environment" with a table header: Name, Type, Length, Size, Value. The content below the header is "Environment is empty".

# Environment/ History



# Files/ Plots/ Packages/ Help



# Resumo

**Editor:** Onde escrevemos os códigos (executamos com CTRL+ENTER).

**Console:** Onde os resultados são printados.

**Environment:** Mostra todos os objetos criados.

**History:** Mostra todos os códigos executados.

**Files:** Mostra os arquivos no diretório atual.

**Plots:** Mostra os outputs de códigos que geram gráficos.

**Packages :** Mostra os pacotes instalados.

**Help:** Mostra a documentação de funções e pacotes.

**Atalhos:** ALT+SHIFT+K (janela com todos os atalhos disponíveis.)



# Funções

- Funções são códigos encapsulados para realização de uma tarefa.
- Utilizando uma função faz com que não seja necessário escrever e copiar o código todas as vezes que precisar executar aquela operação.
- Imagine um caso simples:
  - 1 você tem 20 vetores e deseja calcular a média de cada um destes vetores.
  - 2 Sem uma função seria necessário somar os valores de cada elemento do vetor, dividir pelo número de elementos e replicar isto 20 vezes.
  - 3 Se tivéssemos uma função em que fornecemos o vetor e output é a média do vetor, facilitaria muito nossa vida.
  - 4 Este exemplo parece simples, mas pense em casos em que temos tarefas mais complexas que uma média em que há necessidade de replicar mais de 20 vezes; se torna inviável não utilizar uma função.



# Pacotes

- Pacotes em R são bibliotecas contendo funções e dados, úteis para alguma finalidade específica.
- Estes pacotes adicionais fornecem funcionalidades e podem ser copiados, instalados e carregados.
- Portanto, um pacote é um conjunto de funções implementadas para realizar determinadas tarefas.



## Exemplos de pacote

- **ggplot2**: pacote para geração de gráficos.
- **readr**: Leitura de dados tabulares: csv, tsv, fwf.
- **dplyr**: pacote com funções para manipulação de dados.
- **forcats**: manipulação de variáveis categóricas.
- Entre outros.





## Geração de Relatórios em R: o Rmarkdown

- Markdown é uma linguagem de marcação leve e simples de se usar.
- Documentos markdown podem ser convertidos para diversos outros formatos.
- Rmarkdown = R + Markdown.
- Com Rmarkdown é possível gerar relatórios que combinam texto e código.
- É uma forma elegante de entregar análises estatísticas em um único material.
- Alguns formatos possíveis são:
  - 1 HTML
  - 2 PDF
  - 3 WORD
- Entenda mais ([link](#))
- Neste curso não veremos Rmarkdown, mas estamos planejando um minicurso para o segundo semestre.

- R ([link](#)).
- Rstudio ([link](#)).
- PET ([link](#)).
- LEG ([link](#)).
- Blog do LEG ([link](#)).
- Stack Overflow ([link](#)).
- R-bloggers ([link](#)).
- Página do professor Fernando Mayer ([link](#)).
- Curso-R ([link](#)).



# PET-Estatística UFPR