

Minicurso R

Motivação e Boas Vindas

Lineu Alberto Cavazani de Freitas
Nilton da Silva Reis Filho
Willian Henrique de Paula Ramos

PET - Estatística

O Curso

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

- 20/10.
- 10/11.
- 24/11.

Ementa

- Aula 1: comandos básicos, classes e objetos.
- Aula 2: indexação e seleção condicional;
- Aula 3: entrada e saída de dados e Análise exploratória.

Método de Aula

- As aulas serão todas ministradas no R e intercalando conceitos apresentados pelos ministrantes com atividades práticas em aula.
- Todo o material usado será disponibilizado online para consulta no (**link**).
- 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.

Quem nós somos

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.

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.

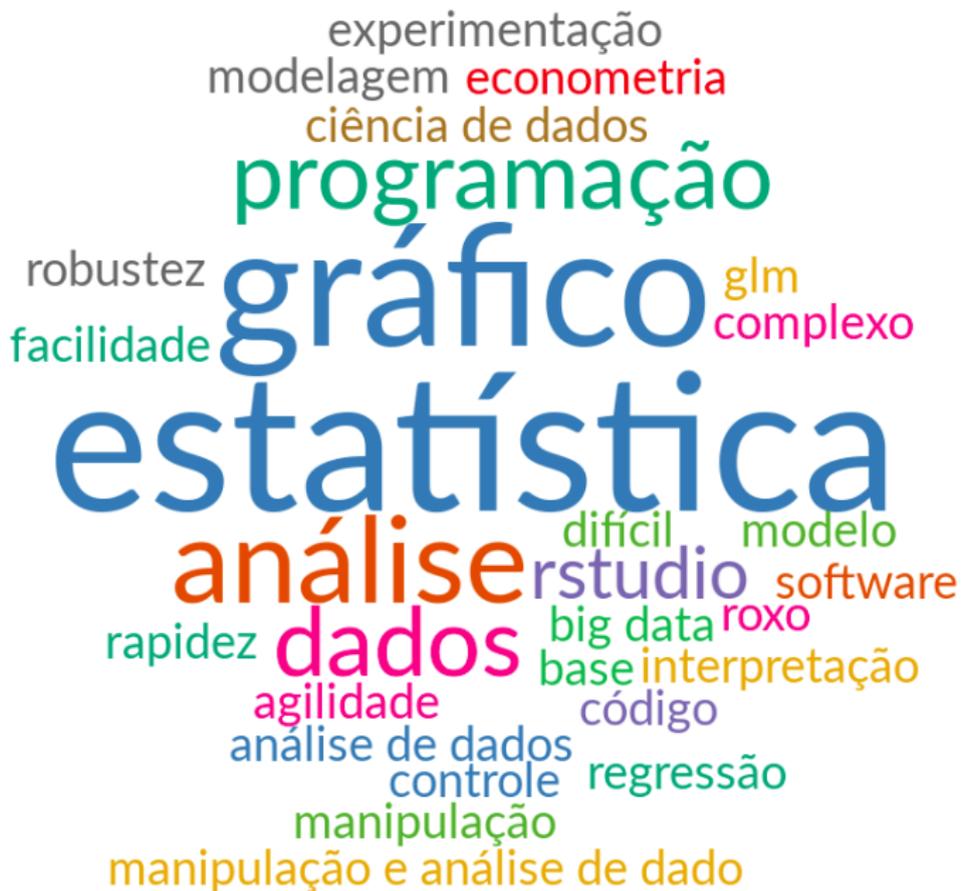
Ministrantes

- Lineu Alberto Cavazani de Freitas.
- Nilton da Silva Reis Filho.
- Willian Henrique de Paula Ramos.

Antes de começarmos. . .

Antes de começarmos. . .

- 1 Acessem o link:
<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.



O R

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

- O R Core Team foi formado em 1997.
- 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.

O R

- 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.
- 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.

O R

- 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.
- Por ser de código aberto, fornece acesso completo aos algoritmos e sua implementação. Nada de caixas pretas.

O R

- 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.
- Uma das mais famosas e de fácil aprendizado é o Rstudio.

Pacotes

O programa R é composto de 3 partes básicas:

- 1** O R-base, que contém as principais funções disponíveis e. Não há necessidade de carregar estas funções.
- 2** Os pacotes recomendados, que são automaticamente instalados junto com o R-base mas não são carregados quando iniciamos o programa.
- 3** Os pacotes contribuídos não são instalados junto com o R-base, o usuário deve instalá-los. Estes pacotes são disponibilizados na página do R e são chamados de pacotes oficiais.

Mas o que é um pacote?

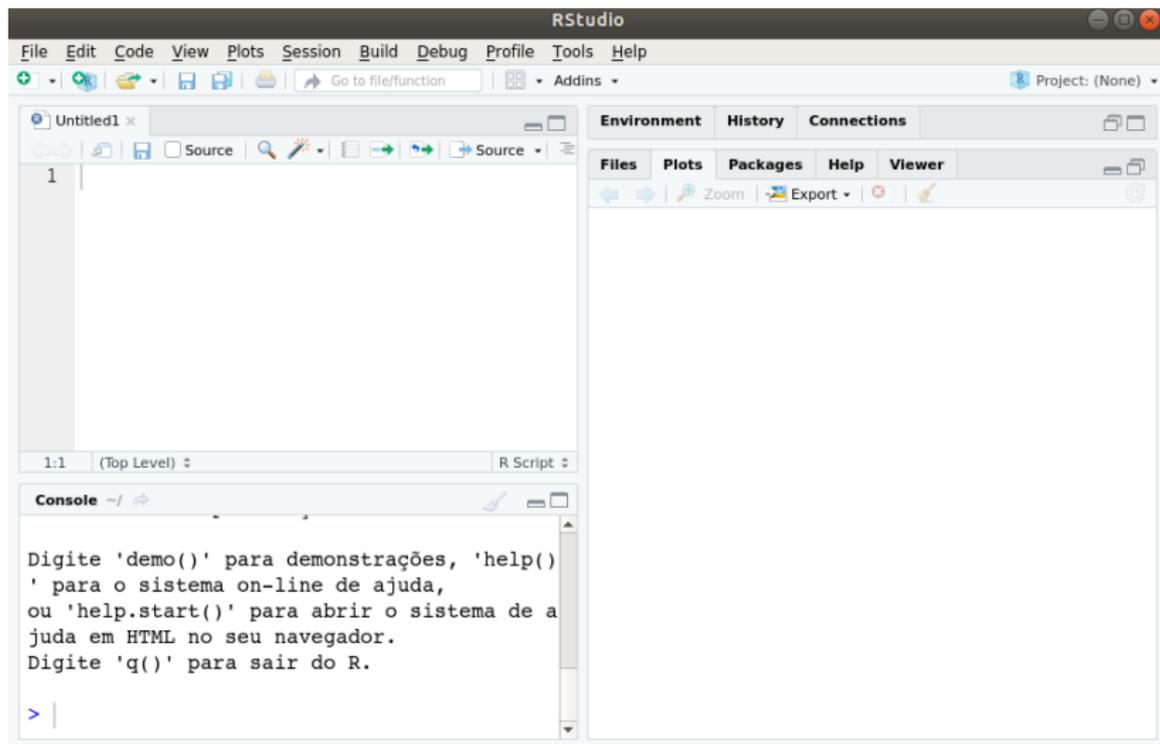
- Pacotes em R são bibliotecas contendo funções e dados, úteis para alguma finalidade específica.
- Estes pacotes adicionais fornecem funcionalidades e podem ser copiados, instalados e carregados.
- Portanto, um pacote é um conjunto de funções implementadas para realizar determinadas tarefas.

Exemplos de pacote:

- ggplot2: pacote específico para geração de gráficos.
- gamlss: pacote específico para ajuste Modelos Generalizados Aditivos para Localização, Escala e Forma.
- dplyr: pacote com funções para manipulação de dados.
- readxl: pacote para leitura de planilhas excel.
- Entre outros.

A Cara do Rstudio

Rstudio



Geração de Relatórios em R: o Rmarkdown

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.
- $\text{Rmarkdown} = \text{R} + \text{Markdown}$.

Geração de Relatórios em R: o Rmarkdown

- 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:
 - HTML
 - PDF
 - WORD

Entenda mais

Links Úteis

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

Avaliação (sem caráter avaliativo)

Descreva o que os fragmentos de código abaixo fazem:

1

```
setwd("~/pc/desktop/minicursoR")
```

2

```
read.table("aula.txt", header = TRUE, sep = ";", dec = ".")
```

3

```
read.csv2("DataCont.csv", header = T, sep = ',')
```

Descreva o que os fragmentos de código abaixo fazem:

4

```
x <- rnorm(100, 0, 1)
y <- rpois(100, 10)

plot(y ~ x)
```

5

```
hist(dados)
```

Descreva o que os fragmentos de código abaixo fazem:

6

```
vetor <- c('1', '2', '3', '4', '5')  
  
is.numeric(vetor)
```

7

```
summary(dados)
```

Descreva o que os fragmentos de código abaixo fazem:

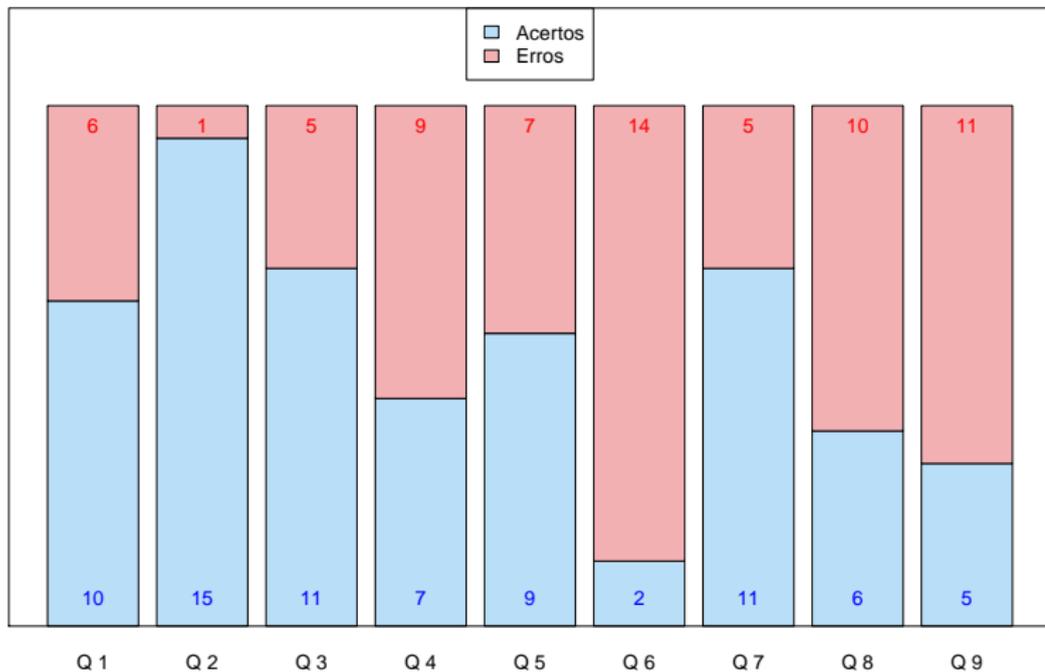
8

```
m0 <- lm(y ~ x1 + x2 + x3, data = dados)
```

9

```
m1 <- glm(y ~ x1 + x2 + x3, data = dados, family = 'poisson')
```

Número de Acertos e Erros por Questão



Obrigado!

Lineu Alberto Cavazani de Freitas

Nilton da Silva Reis Filho

Willian Henrique de Paula Ramos

PET-Estatística UFPR