

WORKSHOP INTRODUTÓRIO DE R

Bruna Kariny Fontes Rodrigues
Wallace Pinele Pantas Porto

Universidade Federal da Bahia
Grupo de Sistemas Elétricos de Potência Integrado
IEEE Power and Energy Society UFBA

Abril de 2021



G-SEPi
IEEE PES UFBA



Introdução

Este é o Workshop Introductório de R. Aqui você aprenderá a utilizar essa linguagem e a sua plataforma desde o mais básico até ser possível construir seus próprios gráficos.

O R é uma linguagem de programação de simples compreensão lagamente utilizada para análises estatísticas.

Uma das principais vantagens é a sua versatilidade devido à possibilidade de criação ou download de pacotes de extensão com funções úteis para casos específicos.

Para acompanhar o curso e aplicar as atividades você deve abrir a plataforma RStudio Cloud e se registrar com a sua conta Google.

A Plataforma

A plataforma consta de:

- Um menu principal superior;
- Quatro seções:
 - Script: onde é criado o código a ser executado;
 - Terminal: onde é possível executar comandos diretos e onde a execução do script é mostrada;
 - Variáveis salvas: onde é mostrada a lista de todas as variáveis salvas no projeto;
 - Histórico: onde é possível visualizar os arquivos na pasta do projeto, os gráficos feitos, os pacotes instalados e a janela de ajuda dos pacotes e funções.

A Plataforma

The screenshot shows the RStudio IDE interface. The top menu bar is highlighted with a red box and labeled "Menu". The main editor window is labeled "Script" and contains a red box around the "Run" button. The Environment pane on the right is labeled "Variáveis Salvas" and shows "Environment is empty". The Files pane on the right is labeled "Histórico" and shows a list of files. The Console/Terminal pane at the bottom is labeled "Terminal" and shows the R prompt and help text.

Menu

File Edit Code View Plots Session Build Debug Profile Tools Help

Run

Script

Environment History Connections Tutorial

Variáveis Salvas

Environment is empty

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

Cloud project

Name	Size	Modified
..		
.Rhistory	0 B	Apr 20, 2021, 6:38 PM
project.Rproj	205 B	Apr 20, 2021, 7:19 PM

Console Terminal Jobs

/cloud/project/

You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |

IDE da plataforma e suas seções

Funções Básicas

Operador	Função
+	Soma
-	Subtração
*	Multiplicação
/	Divisão
^	Potência
sqrt()	Raiz Quadrada
<-	Atribuição
?	Help
#	Comentário
Ctrl+Enter	Executa o código

- O que são
- Como instalar
- Como ativar
- Onde encontrar
- Ajuda com funções e pacotes
- Stack Overflow

O R é uma linguagem vetorial e matricial.

Um numero é entendido como um vetor de uma unica posição.

Atenção: letras maiúsculas e minúsculas são tratadas como letras distintas.

Classes de Variável

- Numeric: numérico que inclui os tipos:
 - Integer: números inteiros;
 - Double: números reais;
 - Complex: números complexos.
- Character: texto simples;
- Logical: binário dividido em True(T) ou False(F);
- Date: datas em diferentes modelos;
- Data.frame: um tipo de tabela de dados (em desuso, evitar formato).

OBS: Para determinar a classe da variável usa-se a função "class()"

Definição de Variáveis

- Vetor: estrutura bidimensional onde cada item deve ser da mesma classe. Para criar utiliza-se a função "c()";
- Matrizes: estrutura unidimensional onde cada item deve ser da mesma classe. Para criar utiliza-se a função "matrix(data, nrow = 1, ncol = 1, byrow = FALSE)" onde:
 - "data" corresponde aos dados que serão inseridos na matriz;
 - "nrow" corresponde ao número de linhas da matriz;
 - "ncol" corresponde ao número de colunas da matriz;
 - "byrow" seleciona a ordem de preenchimento da matriz (por colunas ou por linhas);
- Lista: estrutura unidimensional onde cada item pode ser uma estrutura de qualquer tipo. Para criar utiliza-se a função "list()".

Indexação de Variáveis

Para buscar um elemento específico a indexação depende do tipo de estrutura

- Vetor: "variável()"
- Matriz: "variável(,)"
- Lista: "variável[[]]"
- Tabela: "variável(,)" ou "variável\$nome_da_coluna"

Outras funções muito úteis no R são os laços.

Com eles é possível diminuir significativamente o "trabalho braçal" do usuário ou inserir condições no código.

Existem diversos tipos de laços no R, aqui mostraremos os laços "for", "if" e "apply".

- "for"
 - Laço iterativo;
 - Estrutura: "for("variável" in "vetor")ação";
 - Utilidade: muito útil para ações repetitivas com valores diferentes.
- "if"
 - Laço condicional;
 - Estrutura: "if("condição")ação";
 - Utilidade: muito útil quando se deseja que o código mude de acordo com valores a serem determinados no processo;
 - OBS: é também possível inserir ações alternativas utilizando "elseif" (para casos específicos) e "else" (para qualquer outro caso).
- "apply"
 - Laço iterativo;
 - Estrutura: "apply("variável", 1 ou 2, ação)";
 - Utilidade: muito usado para executar ações a cada linha ou coluna de uma matriz.

Aquisição e Manipulação de Dados

Um das principais utilidades do R é a aquisição e manipulação de dados.

Existem diversas funções de leituras de dados, dentre elas mostraremos o uso da "read_delim" do pacote "tidyverse" pela sua versatilidade e facilidade.

`read_delim(file =, delim =, ...)`, onde:

- `file` é onde deve ser colocado o nome do arquivo seguido do seu formato separado por ponto;
- `delim` é a separação entre os dados do arquivo e depende do arquivo que se quer ler.

OBS: existem mais especificações da função para casos especiais de leituras de dados, para mais informações leia o help da função (`?read_delim`)

Aquisição e Manipulação de Dados

Também existem diversas maneiras de manipular dados para extrair as informações desejadas.

Uma das funções mais úteis para isso é a função "subset" que seleciona uma parte do banco de dados de acordo com a condição especificada pelo usuário

`subset(variável, subset, select)`, onde:

- `variável` representa os dados a serem manipulados;
- `subset` é a condição que define as linhas que se deve manter;
- `select` representa as colunas a serem mantidas.

Aquisição e Manipulação de Dados

Outras funções úteis para manipulação de uma base de dados são:

- `"as_tibble("dados")"`: converte os dados para um formato mais aceito no ggplot2 (veremos a seguir);
- `"colnames("dados")"`: lista os nomes das colunas da base de dados. Caso o usuário atribua um vetor de mesmo tamanho contendo novos títulos, estes mudarão na base de dados desejada;
- `"mutate("dados", "mudança_desejada")"`: modifica os dados de diversas maneiras (ler help para mais detalhes).
- `"write.csv("dados", "nome_do_arquivo", ...)"`: salva o os dados como um arquivo na pasta do projeto para usos futuros (ler help para mais detalhes).

Criação de Gráficos

Finalizando esse workshop mostraremos um dos usos mais importantes do R, a criação de gráficos.

Com esta ferramenta é possível fazer uma análise visual muito simples e didática dos dados.

Aqui mostraremos a função "ggplot" do pacote "ggplot2" presente no conjunto "tidyverse".

Esta função é excelente, pois permite criar gráficos de diversos tipos e editá-los para que fiquem claros e bonitos.

A estrutura depende do tipo de gráfico que se deseja criar, porém todas se iniciam com "ggplot()", alguns exemplos serão dados no código.

Referências

<https://felipegalvao.com.br/pt/blog/basic-r-introduction-data-types-and-structures/>

<https://gilberto-sassi.netlify.app/pt/curso/curso-r/curso-r/>

<http://cursos.leg.ufpr.br/ecr/indexacao-e-selecao-condicional.html>

http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/carga_energia.aspx