

ME115 - Linguagem R

Atividade Prática 06 - Gabarito

1º semestre de 2023

Introdução

Nessa atividade, exploraremos:

1. Importação e exportação de arquivos
2. Caminhos e diretórios no R
3. Manipulação de banco de dados

Atividade

Antes de iniciar essa atividade, baixe o arquivo **iris.zip** disponível no Moodle e descompacte esse arquivo em algum diretório de sua escolha. Esses arquivos serão usados nessa atividade.

- (1) Crie o diretório “Pratica06” no seu diretório local e dentro dele crie os seguintes subdiretórios:

- csv,
- xls.

Solução: Vocês podem realizar isso manualmente dentro do Explorador de Arquivos do seu computador, mas caso queiram criar uma pasta nova pelo R, utilizamos a função `dir.create()`:

```
dir.create('csv')
dir.create('xls')
```

2. A seguir, leia o arquivo **iris.csv** fornecido no Moodle usando a função `read.table()` da base do R e guarde-o no objeto `iris.table`. A seguir exporte o data frame `iris.table` para o diretório “csv” que você criou, especificando o caminho para chegar até lá.

Solução:

```
iris.table <- read.table("../Dados/iris.csv", sep=",", header=TRUE)
write.table(iris.table, "csv/iris.table.csv", col.names=FALSE, row.names=FALSE,
            quote=FALSE)
```

Cuidado: o caminho do arquivo `iris.csv` pode ser diferente no seu computador.

3. A seguir, leia o arquivo **iris.txt** fornecido no Moodle usando a função `read_delim()` do pacote `readr` e guarde-o no objeto `iris_delim`. A seguir, usando a função `write_delim()`, exporte o objeto `iris_delim` para o diretório “csv” que você criou, especificando o caminho para chegar até lá.

Solução:

```
library(readr)
iris_delim <- read_delim("../Dados/iris.txt", delim=" ", show_col_types = FALSE)
write_delim(iris_delim, "csv/iris_delim.csv", delim = ",")
```

4. A seguir, leia o arquivo **iris.csv** fornecido no Moodle usando a função `read_csv()` do pacote `readr` e guarde-o no objeto `iris_csv`. A seguir exporte o objeto `iris_csv` para o diretório “csv” que você criou, especificando o caminho para chegar até lá.

Solução:

```
iris_csv <- read_csv("../Dados/iris.csv", show_col_types = FALSE)
write_csv(iris_csv, "csv/iris_csv.csv")
```

5. Qual o ganho em se usar as funções `read_table()` e `read_csv()` do pacote `readr` em relação à função `read.table()` da base do R?

Solução: O primeiro ganho é que as funções do `readr` já leem o conjunto de dados no formato `tibble`. Elas também permitem nomes de colunas com nomes não usuais e até espaços em branco, e as colunas de caracteres nunca são convertidas automaticamente como fatores.

6. Usando o pacote `readxl`, leia o arquivo `iris.xlsx` fornecido no Moodle usando a função `read_excel()` ou `read_xlsx()` e guarde-o no objeto `iris_xls`. A seguir, usando o pacote `writexl`, exporte o objeto `iris_xls` para o diretório “xls” que você criou, especificando o caminho para chegar até lá.

Solução:

```
library(readxl)
library(writexl)
iris_xls <- read_xlsx("../Dados/iris.xlsx")
write_xlsx(iris_xls, "xls/iris_xls.xls")
```

7. Usando os comandos abaixo para calcular tempo de execução de função, verifique se existe diferença entre os tempos de execução de leitura nos itens (2), (3), (4) e (6). Qual deles é mais rápido?

```
start_time <- Sys.time() ## dispara o cronômetro

## função a ser executada

end_time <- Sys.time() ## para o cronômetro
end_time - start_time ## calcula a diferença de tempos
```

Solução:

```
library(readxl)
## 0 código abaixo calcula os tempos para cada item e mostra numa lista
list(
  ## Tempo do item (2)
  '2.' = { start_time <- Sys.time();
    read.table("../Dados/iris.csv", sep=";", header=TRUE);
    end_time <- Sys.time();
    end_time - start_time },
  ## Tempo do item (3)
  '3.' = { start_time <- Sys.time();
    read_delim("../Dados/iris.txt", delim=" ", show_col_types = FALSE);
    end_time <- Sys.time();
    end_time - start_time },
  ## Tempo do item (4)
  '4.' = { start_time <- Sys.time();
    read_csv("../Dados/iris.csv", show_col_types = FALSE);
    end_time <- Sys.time();
    end_time - start_time },
  ## Tempo do item (6)
  '6.' = { start_time <- Sys.time();
    read_xlsx("../Dados/iris.xlsx");
    end_time <- Sys.time();
    end_time - start_time }
```

```
)

## $`2.`
## Time difference of 0.0009279251 secs
##
## $`3.`
## Time difference of 0.006614923 secs
##
## $`4.`
## Time difference of 0.006567001 secs
##
## $`6.`
## Time difference of 0.005300045 secs
```

8. Usando a função `file.info()`, verifique se existe diferença entre os tamanhos dos arquivos resultantes das exportações feitas nos itens (2), (3), (4) e (6). Qual deles é menor?

Solução:

```
file.info("csv/iris.table.csv")$size ## item (2)

## [1] 3658
file.info("csv/iris_delim.csv")$size ## item (3)

## [1] 3716
file.info("csv/iris_csv.csv")$size ## item (4)

## [1] 3716
file.info("xls/iris_xls.xls")$size ## item (6)

## [1] 8497
```

Desafio

Leia o arquivo `iris.csv` usando o caminho apropriado e guarde no objeto `iris.frame`. Considere a variável `Sepal.length`. A seguir faça:

- a) Crie uma função que ordene o vetor `Sepal.length` e guarde o resultado num data frame chamado `sepal.1.ordenado`. **Dica:** use `for` duplo. Compare seu resultado com o obtido pela função `sort()`.

Solução:

```
iris.frame <- read_csv("../Dados/iris.csv", show_col_types = FALSE)
v <- iris.frame$Sepal.Length
n <- nrow(iris_csv)
for (i in 1:(n-1)){
  for (j in (i+1):n){
    if (v[i] > v[j]){
      v.temp <- v[i]
      v[i] <- v[j]
      v[j] <- v.temp
    }
  }
}
sepal.1.ordenado <- v
sepal.1.ordenado <- tibble::as_tibble(cbind( Meu= sepal.1.ordenado,Sort=sort(v)))
all(sepal.1.ordenado == sort(iris.frame$Sepal.Length))
```

```
## [1] TRUE
```

- b) Exporte o data frame ordenado em (a) com o nome `sepal.1.ordenado` e formato xls para o diretório “xls” criado por você.

Solução:

```
write_xlsx(sepal.1.ordenado, "xls/sepal.1.ordenado.xls")
```

- c) Exporte o data frame ordenado em (a) com o nome `sepal.1.ordenado` e formato csv para o diretório “csv” criado por você.

Solução:

```
write_csv(iris.xls, "csv/sepal.1.ordenado.csv")
```

Agradecimento

O material foi produzido pela Profa. Tatiana Benaglia para o curso de ME115.