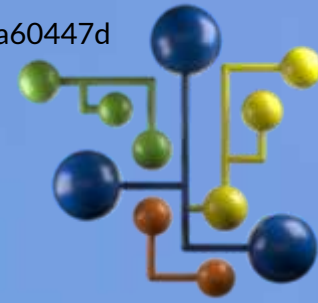




Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d



Big Data Analytics com R e Microsoft Azure Machine Learning



Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

Big Data Analytics com R e Microsoft Azure Machine Learning

Manipulação de Arquivos TXT, CSV e Planilhas Excel em R

Seja Bem-Vindo(a)!



Manipulação de Arquivos TXT, CSV e Planilhas Excel em R

Qual o objetivo em usar Linguagem R?

Analisar dados.



Manipulação de Arquivos TXT, CSV e Planilhas Excel em R

De onde importamos os dados para o R?

De diversas fontes, suportadas pela Linguagem R.

De onde importamos os dados para o R?





Manipulação de Arquivos TXT, CSV e Planilhas Excel em R

De onde importamos os dados para o R?

- Arquivos Texto - flat files (txt, csv)
- Arquivos Excel (xls, xlsx)
- Bancos de Dados (Oracle, SQL Server, MySQL, PostgreSQL, SQLite)
- Softwares Estatísticos (SAS, SPSS, Stata)
- Dados da Internet (Web Crawling, Web Scraping)
- Data Lakes (Apache Hadoop)



Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

Big Data Analytics com R e Microsoft Azure Machine Learning

Importação e Manipulação de Arquivos em R

Seja Bem-Vindo(a)!



Importação e Manipulação de Arquivos em R

A maior parte do seu tempo em tarefas de Ciência de Dados será usada na preparação dos dados.





Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

Importação e Manipulação de Arquivos em R

Importação e Manipulação de Arquivos





Importação e Manipulação de Arquivos em R

Importação e Manipulação de Arquivos

txt

csv

Excel



Importação e Manipulação de Arquivos em R

Importação e Manipulação de Arquivos Texto Flat Files (txt, csv)



Importação e Manipulação de Arquivos em R

Importação e Manipulação de Arquivos

A Linguagem R apresenta 3 pacotes principais para carga de dados a partir de arquivos texto:

`utils`

`readr`

`data.table`

Importação e Manipulação de Arquivos

Pacote utils	Pacote readr
<code>read.table()</code>	<code>read_delim()</code>
<code>read.csv()</code>	<code>read_csv()</code>
<code>read.delim()</code>	<code>read_tsv()</code>

Pacote data.table

`fread()`



Importação e Manipulação de Arquivos em R

Pacote utils

O pacote utils, que é automaticamente carregado na sua sessão R, pode importar arquivos simples em diferentes formas, através das funções:

Pacote utils

read.csv	Para valores separados por vírgula e ponto como separador decimal
read.csv2	Para valores separados por ponto e vírgula e vírgula como separador decimal
read.delim	Para valores separados por tab e ponto como separador decimal
read.delim2	Para valores separados por tab e vírgula e vírgula como separador decimal
read.fwf	Para valores com número exato de bytes por coluna
read.table	Faz a leitura de um arquivo em formato de tabela e carrega como dataframe



Importação e Manipulação de Arquivos em R

`read.table()`

Muito útil quando se está fazendo a leitura de arquivos ASCII, que contém dados em formato retangular.

```
read.table("arquivo.txt", header = TRUE, sep = ",", stringsAsFactors = FALSE)  
read.table("arquivo.txt", header = TRUE, sep = "\t", stringsAsFactors = FALSE)
```




Importação e Manipulação de Arquivos em R

`read.csv()`

Muito útil para importar arquivos com separadores de coluna.

```
read.csv("arquivo.csv", stringsAsFactors = FALSE)
```

```
read.csv2("arquivo.csv", sep = ";", dec = ",", stringsAsFactors = FALSE)
```



Importação e Manipulação de Arquivos em R

```
read.delim()
```

Muito útil para importar arquivos com qualquer tipo de separador de colunas.

```
read.delim("arquivo.txt")  
read.delim2("arquivo.txt")
```



Importação e Manipulação de Arquivos em R

Parâmetros

header
col.names
na.string
colClasses
sep
stringsAsFactors

Pacote readr





Importação e Manipulação de Arquivos em R

Pacote readr

Lançado em Abril/2015 pelos desenvolvedores do RStudio

```
install.packages("readr")
```

```
read_table()
```

```
read_csv ()
```

```
read_delim ()
```



Importação e Manipulação de Arquivos em R

Pacote readr

```
arq1 <- read_table("bigdatafile.txt", col_names = c("DAY", "MONTH", "YEAR", "TEMP"))
```

```
arq2 <- read.table("bigdatafile.txt", col.names = c("DAY", "MONTH", "YEAR", "TEMP"))
```



Importação e Manipulação de Arquivos em R

Pacote data.table

Usado para carga e manipulação de arquivos na ordem de 100 GB em memória RAM.



Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

Big Data Analytics com R e Microsoft Azure Machine Learning

Importação e Manipulação de Planilhas Excel em R

Seja Bem-Vindo(a)!



Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

Importação e Manipulação de Planilhas Excel em R



Excel



Importação e Manipulação de Planilhas Excel em R

Pacote XLConnect

```
loadWorkbook()  
getSheets()  
readWorksheet()  
createsheet()  
writeWorksheet()
```



Importação e Manipulação de Planilhas Excel em R

Pacote xlsx

```
read.xlsx(file, sheetIndex, header=TRUE, colClasses=NA)
```

```
read.xlsx2(file, sheetIndex, header=TRUE, colClasses="character")
```



Importação e Manipulação de Planilhas Excel em R

Pacote readxl

```
read_excel()  
Excel_sheets()
```



Importação e Manipulação de Planilhas Excel em R

Pacote gdata

`read.xls()`



Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

Big Data Analytics com R e Microsoft Azure Machine Learning

Dicas Para Importação e Manipulação de Arquivos em Linguagem R

Seja Bem-Vindo(a)!



Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

Dicas Para Importação e Manipulação de Arquivos em Linguagem R

- Em seus arquivos, evite espaços em branco e números como título para as colunas.



Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

Dicas Para Importação e Manipulação de Arquivos em Linguagem R

When you try to choose
a meaningful variable name.





Dicas Para Importação e Manipulação de Arquivos em Linguagem R

- Em seus arquivos, evite espaços em branco e números como título para as colunas.
- Normalmente, a primeira linha de cada arquivo é o cabeçalho, a lista de nomes para cada coluna.
- Para a concatenação de palavras, use . ou _
- Use nomes curtos como título de coluna.
- Evite o uso de caracteres especiais.
- Dados NA podem existir no seu conjunto de dados e isso será tratado no processo de limpeza.



Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

Dicas Para Importação e Manipulação de Arquivos em Linguagem R

Atenção aos Detalhes

- Encoding
- Linha de cabeçalho
- Separador de colunas
- Quoting (aspas)
- Missing values
- Linhas em branco
- Espaços em branco em campos do tipo caracter
- Comentários



Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

Dicas Para Importação e Manipulação de Arquivos em Linguagem R

Outros Pacotes para Importação de Arquivos:

- Pacote **rjson** - Leitura de arquivos JSON para o R
- Pacote **XML** - Leitura de arquivos xml
- Pacote **httr** - Leitura de páginas html para o R
- Pacote **Rcurl** - Web Crawling
- Pacote **foreign** - Leitura de arquivos do SPSS, SAS
- Pacote **sas7bdat** - Leitura de arquivos SAS



Muito Obrigado por Participar!



Tenha uma Excelente Jornada de Aprendizagem.

Equipe Data Science Academy