

# Introdução à programação com R

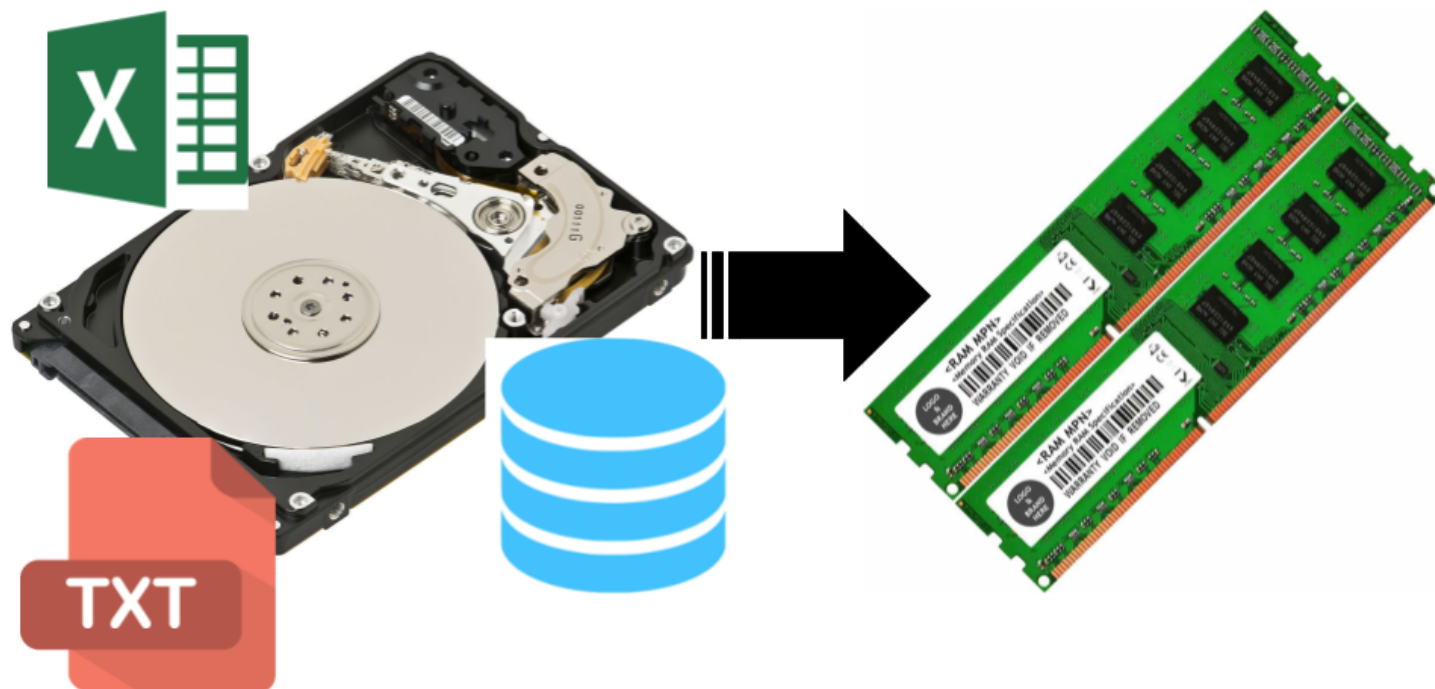
## Introdução à Importação de dados



# Importação

# O que é?

Importar uma base de dados para o R significa levar a informação contida no disco rígido (HD) para a memória RAM.



# Caminhos

Um passo importante na tarefa de importação de dados para o R é saber onde está o arquivo que queremos importar.

Toda função de importação vai exigir um **caminho**, uma string que representa o endereço do arquivo no computador.

Há duas formas de passarmos o caminho de arquivo: usar o **caminho absoluto** ou usar o **caminho relativo**.

Antes de falarmos sobre a diferença dos dois, precisamos definir o que é o **diretório de trabalho**.

# Diretório de trabalho

O diretório de trabalho (*working directory*) é a pasta em que o R vai procurar arquivos na hora de ler informações ou gravar arquivos na hora de salvar objetos.

Se você está usando um projeto, o diretório de trabalho da sua sessão será, por padrão, a pasta raiz do seu projeto (é a pasta que contém o arquivo com extensão `.Rproj`).

Se você não estiver usando um projeto ou não souber qual é o seu diretório de trabalho, você pode descobri-lo usando a seguinte função `getwd()`.

Ela vai devolver uma string com o caminho do seu diretório de trabalho.

A função `setwd()` pode ser utilizada para mudar o diretório de trabalho. Como argumento, ela recebe o caminho para o novo diretório.

# Caminhos absolutos

Caminhos absolutos são aqueles que tem início na pasta raiz do seu computador/usuário. Por exemplo:

```
/Users/beatrizmilz/Documents/Curso-R/main-intro-programacao/slides
```

Esse é o caminho absoluto para a pasta onde esses slides foram produzidos.

Na grande maioria dos casos, caminhos absolutos são uma **má prática**, pois deixam o código irreprodutível. Se você trocar de computador ou passar o script para outra pessoa rodar, o código não vai funcionar, pois o caminho absoluto para o arquivo muito provavelmente será diferente.

# Caminhos relativos

Caminhos relativos são aqueles que tem início no diretório de trabalho da sua sessão.

O diretório de trabalho da sessão utilizada para produzir esses slides é a pasta `intro-programacao-em-r-mestre`. Veja o caminho absoluto no slide anterior. Então, o caminho relativo para a pasta onde esses slides foram produzidos seria apenas `slides/`.

**Trabalhar com projetos no RStudio ajuda bastante o uso de caminhos relativos**, pois nos incentiva a colocar todos os arquivos da análise dentro da pasta do projeto.

Assim, se você usar apenas caminhos relativos e compartilhar a pasta do projeto com alguém, todos os caminhos existentes nos códigos continuarão a funcionar em qualquer computador!

# Lendo tabelas

Para ler tabelas, como arquivos `.csv`, utilizaremos funções do pacote `readr`.

Para isso, utilizamos a função `read_csv()` ou `read_csv2()`. Se o arquivo estiver bem formatado, a função só precisa do caminho até o arquivo para funcionar.



A mensagem devolvida pela função indica qual classe foi atribuída para cada coluna da base.

```
## i Using ',' as decimal and '.' as grouping mark. Use `read_delim()` for more control.
```

```
##
```

```
## — Column specification —————
```

```
## cols(
```

```
##   ano = col_double(),
```

```
##   mes = col_double(),
```

```
##   dia = col_double(),
```

```
##   horario_saida = col_double(),
```

```
##   saida_programada = col_double(),
```

```
##   atraso_saida = col_double(),
```

```
##   horario_chegada = col_double(),
```

```
##   chegada_prevista = col_double(),
```

```
##   atraso_chegada = col_double(),
```

```
##   companhia_aerea = col_character(),
```

```
##   voo = col_double(),
```

```
##   cauda = col_character(),
```

```
##   origem = col_character(),
```

```
##   destino = col_character(),
```

```
##   tempo_voo = col_double(),
```

```
##   distancia = col_double(),
```

```
##   hora = col_double(),
```

```
##   minuto = col_double(),
```

```
##   data_hora = col_datetime(format = "")
```

```
## )
```

- Em alguns países, como o Brasil, as vírgulas são utilizadas para separar as casas decimais dos números, inviabilizando os arquivos `.csv`. Nesses casos, os arquivos `.csv` são na verdade separados por ponto-e-vírgula. Para ler bases separadas por ponto-e-vírgula no R, utilize a função `read_csv2()`.

```
voos_csv <- readr::read_csv2("../dados/voos_de_janeiro.csv")
```

- Arquivos `.txt` podem ser lidos com a função `read_delim()`. Além do caminho até o arquivo, você também precisa indicar qual é o caractere utilizado para separar as colunas da base. Um arquivo separado por tabulação, por exemplo, pode ser lido utilizando a o código abaixo. O código `\t` é uma forma textual de representar a tecla TAB.
- Para ler planilhas do Excel (arquivos `.xlsx` ou `.xls`), basta utilizarmos a função `read_excel()` do pacote `readxl`.