

# desafio 1

Sophia ra277230

2025-09-25

carregue os dados utilizando as funções: read.csv() do pacote base do

```
# medir o tempo
```

```
start_time <- Sys.time()
```

```
getwd()
```

```
## [1] "C:/Users/Shophia/Documents/me315"
```

```
setwd("C:/Users/Shophia/Documents/me315")
```

```
flights_base <- read.csv("flights.csv.zip")
```

```
## Warning in read.table(file = file, header = header, sep = sep, quote = quote, :  
## line 1 appears to contain embedded nulls
```

```
## Warning in read.table(file = file, header = header, sep = sep, quote = quote, :  
## line 2 appears to contain embedded nulls
```

```
## Warning in scan(file = file, what = what, sep = sep, quote = quote, dec = dec,  
## : EOF within quoted string  
## Warning in scan(file = file, what = what, sep = sep, quote = quote, dec = dec,  
## : EOF within quoted string
```

```
## Warning in scan(file = file, what = what, sep = sep, quote = quote, dec = dec,  
## : embedded nul(s) found in input
```

```
end_time <- Sys.time()
```

```
tempo_base <- end_time - start_time
```

```
print(tempo_base)
```

```
## Time difference of 0.008541107 secs
```

read\_csv() do pacote readr do

```
library(readr)
```

```
start_time <- Sys.time()
```

```
flights_readr <- read_csv("flights.csv.zip")
```

```
## Rows: 5819079 Columns: 31
## -- Column specification -----
## Delimiter: ","
## chr (11): AIRLINE, TAIL_NUMBER, ORIGIN_AIRPORT, DESTINATION_AIRPORT, SCHEDUL...
## dbl (20): YEAR, MONTH, DAY, DAY_OF_WEEK, FLIGHT_NUMBER, DEPARTURE_DELAY, TAX...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
end_time <- Sys.time()
tempo_readr <- end_time - start_time
print(tempo_readr)
```

```
## Time difference of 18.01815 secs
```

```
read_csv() do pandas no
```

```
import pandas as pd
import time

start_time = time.time()

flights_pd = pd.read_csv("flights.csv.zip")
```

```
## <string>:2: DtypeWarning: Columns (7,8) have mixed types. Specify dtype option on import or set low_
```

```
end_time = time.time()
tempo_pd = end_time - start_time
print("Tempo pandas:", tempo_pd, "segundos")
```

```
## Tempo pandas: 18.383955478668213 segundos
```