

# JULIA

## 1. Criar as bases de exemplo

```
using CSV, DataFrames, XLSX, Random
```

```
# 1) CSV pequeno
df_peq = DataFrame(
    id = 1:10,
    nome =
    ["Ana", "Bruno", "Carla", "Diego", "Eva", "Fernando", "Gabi", "Hugo", "Iris", "João"],
    idade = rand(18:60, 10),
    salario = rand(2500.0:50.0:9000.0, 10)
)
CSV.write("dados_pequenos.csv", df_peq)
```

```
# 2) TXT com separador \t
df_tab = DataFrame(
    produto = ["A", "B", "C", "D", "E"],
    preco = rand(20.0:1.0:100.0, 5),
    estoque = rand(1:200, 5)
)
CSV.write("dados_tab.txt", df_tab; delim = '\t')
```

```
# 3) CSV grande (100 mil linhas)
Random.seed!(123)
N = 100_000
df_grande = DataFrame(
    id = 1:N,
    categoria = rand(["A", "B", "C", "D"], N),
    valor = rand(1000.0:0.1:5000.0, N),
    quantidade = rand(1:100, N)
)
CSV.write("dados_grandes.csv", df_grande)
```

```
# 4) Arquivo Excel XLSX com aba "Planilha1"
XLSX.writetable(
    "dados.xlsx",
    collect(eachcol(df_peq)),
    String.(names(df_peq));
    sheetname = "Planilha1",
    overwrite = true
)
```

---

## 2. Importar as bases em Julia

```
using CSV, DataFrames, XLSX

# CSV pequeno
df_csv = CSV.read("dados_pequenos.csv", DataFrame)

# TXT com tab
df_txt = CSV.read("dados_tab.txt", DataFrame; delim = '\t')

# CSV grande
df_big = CSV.read("dados_grandes.csv", DataFrame)

# XLSX - aba "Planilha1"
tbl = XLSX.readtable("dados.xlsx", "Planilha1") # XLSX.DataTable
df_xlsx = DataFrame(tbl)

# espiadinhas
first(df_csv, 5)
first(df_txt, 5)
size(df_big)
first(df_xlsx, 5)
```

---

## 3. Benchmark de importação em Julia

```
using BenchmarkTools, CSV, DataFrames

# aquece o compilador
CSV.read("dados_grandes.csv", DataFrame);

# benchmark principal
@btime CSV.read("dados_grandes.csv", DataFrame);
```

---



## 1. Criar as mesmas bases em R

```
install.packages(c("readr", "writexl")) # se precisar

library(readr)
```

```

library(writexl)

set.seed(123)

# 1) CSV pequeno
df_peq <- data.frame(
  id      = 1:10,
  nome   = c("Ana", "Bruno", "Carla", "Diego", "Eva",
            "Fernando", "Gabi", "Hugo", "Iris", "Joao"),
  idade = sample(18:60, 10, replace = TRUE),
  salario = sample(seq(2500, 9000, by = 50), 10, replace = TRUE)
)
write_csv(df_peq, "dados_pequenos.csv")

# 2) TXT com \t
df_tab <- data.frame(
  produto = c("A", "B", "C", "D", "E"),
  preco   = sample(20:100, 5, replace = TRUE),
  estoque = sample(1:200, 5, replace = TRUE)
)
write_tsv(df_tab, "dados_tab.txt")

# 3) CSV grande (100k linhas)
N <- 100000
df_grande <- data.frame(
  id        = 1:N,
  categoria = sample(c("A", "B", "C", "D"), N, replace = TRUE),
  valor     = runif(N, min = 1000, max = 5000),
  quantidade = sample(1:100, N, replace = TRUE)
)
write_csv(df_grande, "dados_grandes.csv")

# 4) XLSX com aba "Planilha1"
write_xlsx(list(Planilha1 = df_peq), "dados.xlsx")

```

---

## 2. Importar as bases em R

```

library(readr)
library(data.table)
library(readxl)

# CSV pequeno
df_csv <- read_csv("dados_pequenos.csv")

```

```
# TXT com tab
df_txt <- read_tsv("dados_tab.txt")

# CSV grande
df_big <- fread("dados_grandes.csv")

# XLSX
df_xlsx <- read_excel("dados.xlsx", sheet = "Planilha1")

head(df_csv)
head(df_txt)
dim(df_big)
head(df_xlsx)
```

---

### 3. Benchmark de importação em R

```
install.packages(c("data.table", "microbenchmark", "dplyr")) # se precisar

library(data.table)
library(readr)
library(microbenchmark)
library(dplyr)

# warmup
tmp1 <- fread("dados_grandes.csv")
tmp2 <- read_csv("dados_grandes.csv", show_col_types = FALSE)

# benchmark
res <- microbenchmark(
  fread = fread("dados_grandes.csv"),
  readr = read_csv("dados_grandes.csv", show_col_types = FALSE),
  times = 5
)

# resumo em ms
summary(res) %>%
  select(expr, median) %>%
  mutate(median_ms = median / 1e6)
```

---

## 1. Criar as bases em Python (opcional, se quiser tudo em Py)

```
import pandas as pd
import numpy as np

# 1) CSV pequeno
df_peq = pd.DataFrame({
    "id": range(1, 11),
    "nome": ["Ana", "Bruno", "Carla", "Diego", "Eva",
              "Fernando", "Gabi", "Hugo", "Iris", "João"],
    "idade": np.random.randint(18, 61, size=10),
    "salario": np.random.choice(range(2500, 9001, 50), size=10)
})
df_peq.to_csv("dados_pequenos.csv", index=False)

# 2) TXT com \t
df_tab = pd.DataFrame({
    "produto": ["A", "B", "C", "D", "E"],
    "preco": np.random.randint(20, 101, size=5),
    "estoque": np.random.randint(1, 201, size=5)
})
df_tab.to_csv("dados_tab.txt", sep="\t", index=False)

# 3) CSV grande (100k linhas)
N = 100_000
np.random.seed(123)
df_grande = pd.DataFrame({
    "id": np.arange(1, N+1),
    "categoria": np.random.choice(list("ABCD"), size=N),
    "valor": np.random.uniform(1000.0, 5000.0, size=N),
    "quantidade": np.random.randint(1, 101, size=N)
})
df_grande.to_csv("dados_grandes.csv", index=False)

# 4) XLSX com aba "Planilha1"
df_peq.to_excel("dados.xlsx", sheet_name="Planilha1", index=False)
```

---

## 2. Importar as bases em Python

```
import pandas as pd

# CSV pequeno
df_csv = pd.read_csv("dados_pequenos.csv")
```

```
# TXT com \t
df_txt = pd.read_csv("dados_tab.txt", sep="\t")

# CSV grande
df_big = pd.read_csv("dados_grandes.csv")

# XLSX
df_xlsx = pd.read_excel("dados.xlsx", sheet_name="Planilha1")

print(df_csv.head())
print(df_txt.head())
print(df_big.shape)
print(df_xlsx.head())
```

---

### 3. Benchmark de importação em Python

```
import pandas as pd
import time

def load_with_pandas():
    pd.read_csv("dados_grandes.csv")

# warmup
load_with_pandas()

tempos = []
for _ in range(5):
    inicio = time.perf_counter()
    load_with_pandas()
    fim = time.perf_counter()
    tempos.append(fim - inicio)

print("Tempos (s):", tempos)
print("Tempo médio (s):", sum(tempos) / len(tempos))
print("Tempo mínimo (s):", min(tempos))
print("Tempo máximo (s):", max(tempos))
```