

```
library(readr)

start_time <- Sys.time()

unzip('flights.zip')
```

Warning in unzip("flights.zip"): erro 1 na extração a partir de arquivo zip

```
dadosreadr<- 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...
##
## # Use `spec()` to retrieve the full column specification for this data.
## # Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
end_time <- Sys.time()

tempo_carregamento <- end_time - start_time

print(tempo_carregamento)
```

Time difference of 38.09677 secs

```
head(dadosreadr)
```

```
## # A tibble: 6 × 31
##   YEAR MONTH   DAY DAY_OF_WEEK AIRLINE FLIGHT_NUMBER TAIL_NUMBER ORIGIN_AIRPORT
##   <dbl> <dbl> <dbl>         <dbl> <chr>         <dbl> <chr>         <chr>
## 1  2015     1     1           4 AS             98 N407AS      ANC
## 2  2015     1     1           4 AA            2336 N3KUAA      LAX
## 3  2015     1     1           4 US             840 N171US      SFO
## 4  2015     1     1           4 AA             258 N3HYAA      LAX
## 5  2015     1     1           4 AS             135 N527AS      SEA
## 6  2015     1     1           4 DL             806 N3730B      SFO
## # # 23 more variables: DESTINATION_AIRPORT <chr>, SCHEDULED_DEPARTURE <chr>,
## # DEPARTURE_TIME <chr>, DEPARTURE_DELAY <dbl>, TAXI_OUT <dbl>,
## # WHEELS_OFF <chr>, SCHEDULED_TIME <dbl>, ELAPSED_TIME <dbl>, AIR_TIME <dbl>,
## # DISTANCE <dbl>, WHEELS_ON <chr>, TAXI_IN <dbl>, SCHEDULED_ARRIVAL <chr>,
## # ARRIVAL_TIME <chr>, ARRIVAL_DELAY <dbl>, DIVERTED <dbl>, CANCELLED <dbl>,
## # CANCELLATION_REASON <chr>, AIR_SYSTEM_DELAY <dbl>, SECURITY_DELAY <dbl>,
## # AIRLINE_DELAY <dbl>, LATE_AIRCRAFT_DELAY <dbl>, WEATHER_DELAY <dbl>
```

Time difference of 19.12769 secs

```
start_time <- Sys.time()

unzip('flights.zip')
```

Warning in unzip("flights.zip"): erro 1 na extração a partir de arquivo zip

```
dadosbase <- 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, :
linha final incompleta encontrada por readTableHeader em 'flights.csv.zip'

```
end_time <- Sys.time()

tempo_carregamento <- end_time - start_time

print(tempo_carregamento)
```

Time difference of 0.010252 secs

```
head(dadosbase)
```

```
## [1] PK...
## <0 linhas> (ou row.names de comprimento 0)
```

Time difference of 57.79526 secs

```
import time
import pandas as pd
ini = time.time()
dados = pd.read_csv("flights.csv.zip", compression="zip", header=0, sep=",", low_memory=False)
fim = time.time()
fim - ini
```

65.40469217300415