How can I read multiple CSV files in R?

Authored by

stats writer

June 24, 2024

In order to read multiple CSV files in R, one can use the "read.csv()" function along with the "list.files()" function to create a vector of the file names. This vector can then be used as an input to a for loop, allowing the user to read each file individually and combine them into a single data frame. Alternatively, the "lapply()" function can be used to apply the "read.csv()" function to each file in the vector, resulting in a list of data frames which can then be merged together using the "rbind()" function. Both methods provide efficient ways to read and combine multiple CSV files in R.

Using read.csv() is not a good option to import multiple large CSV files into an R data frame, however, R has several packages that provide a method to read large various CSV files into a single R DataFrame.

In my previous article, I discussed how to read a CSV file, In this article, I will demonstrate how to read multiple CSV files from a folder into a single data frame in R by using different packages.

1. Quick Examples of R Read Multiple CSV Files

The following are examples of importing multiple CSV files into a data frame in R using different packages.

```
# Quick examples

# Example 1 - Use data.table package
library(data.table)
df <-
list.files(path = "/Users/admin/apps/csv-courses/", pattern = "*.csv") %>%
map_df(~fread(.))
df

# Example 2 - Using tidyverse
library(tidyverse)
df <-
list.files(path = "/Users/admin/apps/csv-courses/", pattern = "*.csv") %>%
map_df(~read_csv(.))
df

# Example 3 - Using readr package
library(readr)
list_csv_files <- list.files(path = "/Users/admin/apps/csv-courses/")
df2 <- readr::read_csv(list_csv_files, id = "file_name")
df2
```

```
# Example 4 - Using read.csv()
list_csv_files <- list.files(path = "/Users/admin/apps/csv-courses/")
df2 = do.call(rbind, lapply(list_csv_files, function(x) read.csv(x, stringsAsFactors = FALSE)))
df2
```

2. Read Multiple CSV Files in R (The best approach)

To read multiple CSV files or all files from a folder in R, use data.table package. It is a third-party library hence, to use the data.table library, you need to first install it by using install.packages('data.table'). Once installation is completed, load the `data.table` library by using `library("data.table")`.

I am using a `fread()` version of `data.table` package as this is the efficient option in R to import multiple larger CSV files as it gives better performance compared with other packages.

```
# Use data.table package
library(data.table)
df <-
list.files(path = "/Users/admin/apps/csv-courses/", pattern = "*.csv") %>%
map_df(~fread(.))
df
```

Yields below output. This by default uses *`stringsAsFactors = FALSE`*. Here list.files() returns all CSV files from a specific path.

```
# Output
id name dob gender
1: 10 sai 1990-10-02 M
2: NA ram 1981-03-24
3: -1 <NA> 1987-06-14 F
4: 13 1985-08-16 <NA>
5: 10 sai 1990-10-02 M
6: NA ram 1981-03-24
7: -1 <NA> 1987-06-14 F
8: 13 1985-08-16 <NA>
```

3. Using tidyverse to Read Multiple CSV Files From a Folder

Using the tidyverse to read multiple CSV files into a single DataFrame in R is the second-best approach.

```
# Using tidyverse
library(tidyverse)
df <-
list.files(path = "/Users/admin/apps/csv-courses/", pattern = "*.csv") %>%
map_df(~read_csv(.))
df
```

Yields below output.

```
# Output
# A tibble: 8 × 4
id name dob gender
<dbl> <chr> <date> <chr>
1 10 sai 1990-10-02 M
2 NA ram 1981-03-24 NA
3 -1 <NA> 1987-06-14 F
4 13 NA 1985-08-16 <NA>
5 10 sai 1990-10-02 M
6 NA ram 1981-03-24 NA
7 -1 <NA> 1987-06-14 F
8 13 NA 1985-08-16 <NA>
```

4. Using readr Package

You can consider this as a third option for loading multiple CSV files into an R data frame. This method uses the `read_csv()` function from the `readr` package, which is a third-party library. To use the `readr` library, you need to install it first by running `install.packages('readr')`. After the installation is complete, load the `readr` library using `library('readr')` to access the `read_csv()` function.

```
# Using readr package
library(readr)
list_csv_files <- list.files(path = "/Users/admin/apps/csv-courses/")
```

```
df <- readr::read_csv(list_csv_files, id = "file_name")
df
```

Yields the same output as above.

5. Using R Base read.csv()

R base function provides read.csv() to import a CSV file into DataFrame. You can also use to this to import multiple CSV files at a time in R.

This is the slowest method of all hence it's not recommended to use on large files. If you have small files and you don't have the above packages installed then you could use this option.

```
# Using read.csv()
list_csv_files <- list.files(path = "/Users/admin/apps/csv-courses/")
df2  =  do.call(rbind,  lapply(list_csv_files,  function(x)  read.csv(x,
stringsAsFactors  =  FALSE)))
df2
```

Yields below output.

```
# Output
id name dob gender
1 10 sai 1990-10-02 M
2 NA ram 1981-03-24
3 -1 <NA> 1987-06-14 F
4 13 1985-08-16 <NA>
5 10 sai 1990-10-02 M
6 NA ram 1981-03-24
7 -1 <NA> 1987-06-14 F
8 13 1985-08-16 <NA>
```


Conclusion

In this article, you have learned how to read/import multiple CSV files from a folder into a single R DataFrame.

Related Articles

References