Handling an Excel File in R

Microsoft Excel is the most widely used spreadsheet program which stores data in the .xls or .xlsx format. R can read directly from these files using some excel specific packages. Few such packages are - XLConnect, xlsx, gdata etc. We will be using xlsx package. R can also write into excel file using this package.

Install xlsx Package

You can use the following command in the R console to install the "xlsx" package. It may ask to install some additional packages on which this package is dependent. Follow the same command with required package name to install the additional packages.

install.packages("xlsx")

Verify and Load the "xlsx" Package

Use the following command to verify and load the "xlsx" package.

# Verify the package is installed.

any(grepl("xlsx",installed.packages()))

# Load the library into R workspace.

library("xlsx")

When the script is run we get the following output.

[1] TRUE

Loading required package: rJava

Loading required package: methods

Loading required package: xlsxjars

Input as xlsx File

Open Microsoft excel. Copy and paste the following data in the work sheet named as sheet1.

id name salary start\_date dept

1 Rick 623.3 1/1/2012 IT

2 Dan 515.2 9/23/2013 Operations

3 Michelle 611 11/15/2014 IT

4 Ryan 729 5/11/2014 HR

5 Gary 43.25 3/27/2015 Finance

6 Nina 578 5/21/2013 IT

7 Simon 632.8 7/30/2013 Operations

8 Guru 722.5 6/17/2014 Finance

Also copy and paste the following data to another worksheet and rename this worksheet to "city".

name city

Rick Seattle

Dan Tampa

Michelle Chicago

Ryan Seattle

Gary Houston

Nina Boston

Simon Mumbai

Guru Dallas

Save the Excel file as "input.xlsx". You should save it in the current working directory of the R workspace.

Reading the Excel File

The input.xlsx is read by using the **read.xlsx()** function as shown below. The result is stored as a data frame in the R environment.

# Read the first worksheet in the file input.xlsx.

data <- read.xlsx("input.xlsx", sheetIndex = 1)

print(data)

When we execute the above code, it produces the following result −

id, name, salary, start\_date, dept

1 1 Rick 623.30 2012-01-01 IT

2 2 Dan 515.20 2013-09-23 Operations

3 3 Michelle 611.00 2014-11-15 IT

4 4 Ryan 729.00 2014-05-11 HR

5 NA Gary 843.25 2015-03-27 Finance

6 6 Nina 578.00 2013-05-21 IT

7 7 Simon 632.80 2013-07-30 Operations

8 8 Guru 722.50 2014-06-17 Finance