

CSV Format and Working Directory

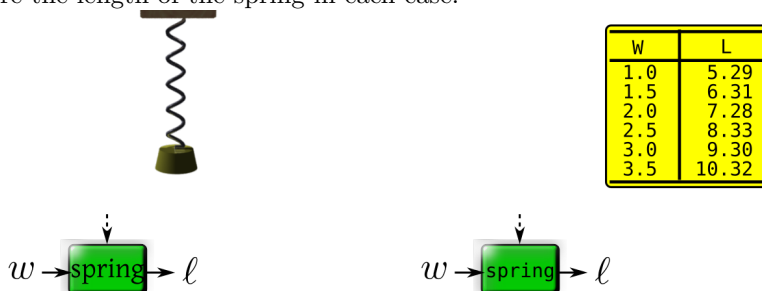
Detailed Example

Bhaskara Rahul Sanku

August 2022

In the previous section we have seen the proper way to store the data i.e, data layout. Now let us learn how we represent our data in a file. As we basically use R for analysing data, it is important to learn which format of a file is easy to handle in R. The **CSV(Comma Separated Values)** format is the most used format in R as it is easy to exchange data between softwares in this format. We can also convert files from various formats like excel to CSV format by appropriate steps.

Consider an example where we hang 6 different weights to a same spring and measure the length of the spring in each case.



In CSV format, The **variables and their values** are separated by **commas** and **each case/trial** are separated by writing them in **different lines**. Also the no of spaces after comma doesn't matter in a case and there will not be any comma at the end of the case.

So the data in CSV format looks like (let the file name be spring.csv)

```
W , L
1.0 , 5.29
1.5 , 6.31
2.0 , 7.28
2.5 , 8.33
3.0 , 9.30
3.5 , 10.32
```

Now let us learn how to read data from a file using R. R follows the concept of working directory i.e, whenever we try to read anything in the harddrive R will look for it in certain directory/folder. So we have to instruct R to look into the folder where our data file is present. If we give the wrong folder, it will not work.

We can get working directory using the command **getwd()**. If the data file is not present in our present working directory, we can change it using the command **setwd(" ")** and writing the required directory in between inverted commas.

Alternatively we can just use the command **dat=read.csv(" ")** (we are naming our data as **dat**), where in between the inverted commas we write the name of our data file which in our present example is **spring.csv**. To ensure we loaded the correct data we can check no of trials and variables using the command **dim(dat)** or else we can check values of variables in specific trials. This is how we read data in a file in R.