

# **Introduction to R Software**

## **Data Handling**

**:::**

## **Importing CSV and Tabular Data Files**

**Shalabh**

**Department of Mathematics and Statistics**

**Indian Institute of Technology Kanpur**

## Setting up directories

❑ We can change the current working directory as follows:

```
> setwd("<location of the dataset>")
```

**Example:**

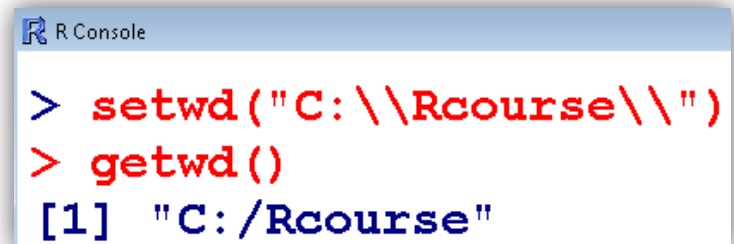
```
> setwd("C:/RCourse/")
```

or

```
> setwd("C:\\Rcourse\\")
```

❑ The following command returns the current working directory:

```
> getwd()  
[1] "C:/RCourse/"
```



```
R Console  
> setwd("C:\\Rcourse\\")  
> getwd()  
[1] "C:/Rcourse"
```

# **Importing Data Files**

**Suppose we have some data on our computer and we want to import it in R.**

**Different formats of files can be read in R**

- **comma-separated values (CSV) data file,**
- **table file (TXT),**
- **Spreadsheet (e.g., MS Excel) file,**
- **files from other software like SPSS, Minitab etc.**

## Importing Data Files

One can also read or upload the file from Internet site.

We can read the file containing rent index data from website:

<http://home.iitk.ac.in/~shalab/Rcourse/munichdata.asc>

as follows

```
> datamunich <- read.table(file=  
"http://home.iitk.ac.in/~shalab/Rcourse/munichdata.asc", header=TRUE)
```

File name is `munichdata.asc`

# Importing Data Files

Comma-separated values (CSV) files

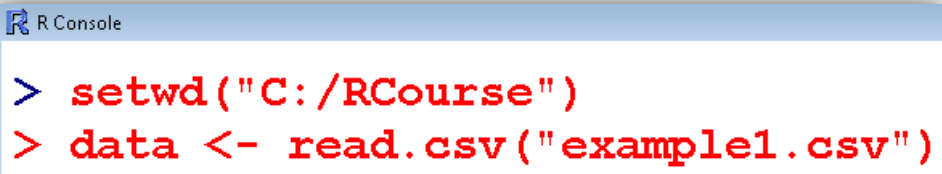
First set the working directory where the CSV file is located.

```
setwd("<location of your dataset>")
```

```
> setwd("C:/RCourse/")
```

To read a CSV file

Syntax: `read.csv("filename.csv")`

A screenshot of an R Console window. The title bar says "R Console". The console shows two lines of code being executed: the first line is "> setwd('C:/RCourse')", and the second line is "> data <- read.csv('example1.csv')". Both lines are highlighted in red.

```
> setwd("C:/RCourse")  
> data <- read.csv("example1.csv")
```

**Example:**

```
> data <- read.csv("example1.csv")
```

# Importing Data Files

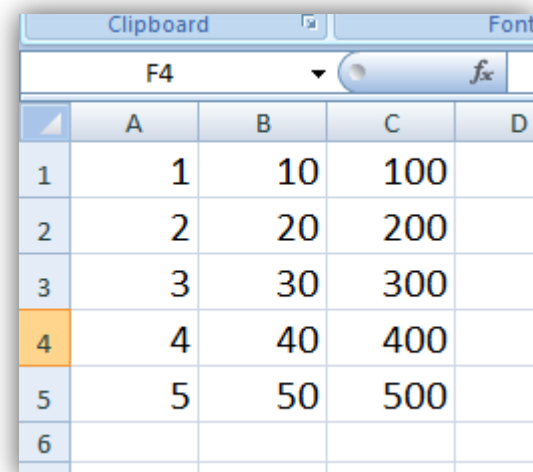
## Comma-separated values (CSV) files

### Example:

```
> data <- read.csv("example1.csv")
```

```
> data
```

```
      x1 x10 x100
1      2  20  200
2      3  30  300
3      4  40  400
4      5  50  500
```



	A	B	C	D
1	1	10	100	
2	2	20	200	
3	3	30	300	
4	4	40	400	
5	5	50	500	
6				

R Console

```
> setwd("C:/RCourse")
> data <- read.csv("example1.csv")
> data
      x1 x10 x100
1      2  20  200
2      3  30  300
3      4  40  400
4      5  50  500
```

Notice the difference in the first rows of excel file and output

# Importing Data Files

## Comma-separated values (CSV) files

Data files have many formats and accordingly we have options for loading them.

If the data file does not have headers in the first row, then use

```
data <- read.csv("datafile.csv", header=FALSE)
```

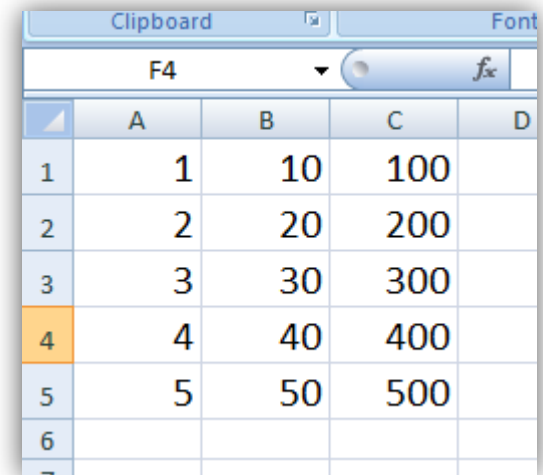
# Importing Data Files

Comma-separated values (CSV) data

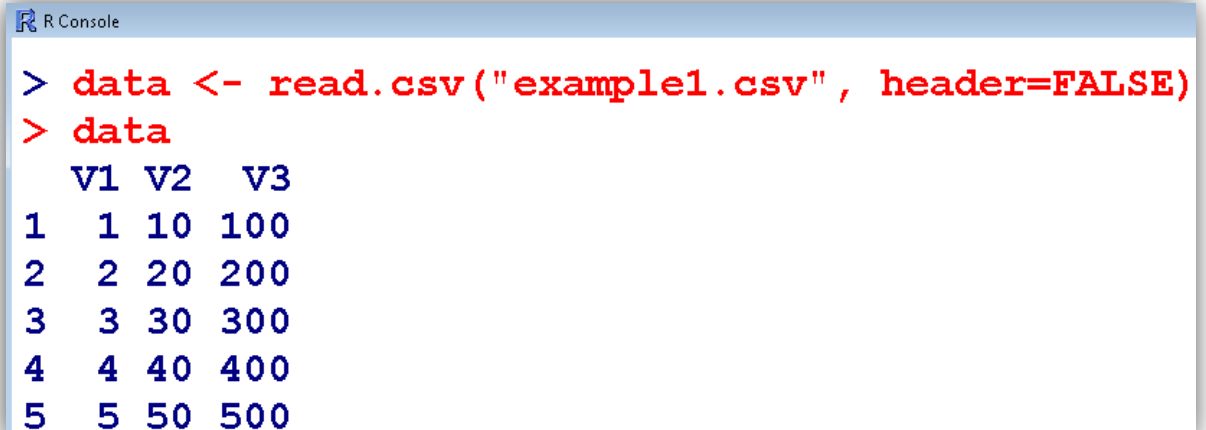
**Example:**

```
> data <- read.csv("example1.csv", header=FALSE)
```

```
> data
  V1 V2  V3
1  1 10 100
2  2 20 200
3  3 30 300
4  4 40 400
5  5 50 500
```



	A	B	C	D
1	1	10	100	
2	2	20	200	
3	3	30	300	
4	4	40	400	
5	5	50	500	
6				



```
> data <- read.csv("example1.csv", header=FALSE)
> data
  V1 V2  V3
1  1 10 100
2  2 20 200
3  3 30 300
4  4 40 400
5  5 50 500
```



# Importing Data Files

Comma-separated values (CSV) files

The resulting data frame will have columns named V1, V2, ...

We can rename the header names manually:

```
> names(data) <-c("Column1", "Column2", "Column3")
```

```
> data
```

	Column1	Column2	Column3
1	1	10	100
2	2	20	200
3	3	30	300
4	4	40	400
5	5	50	500

# Importing Data Files

## Comma-separated values (CSV) files

```
R Console
> data <- read.csv("example1.csv", header=FALSE)
> data
  V1 V2  V3
1  1 10 100
2  2 20 200
3  3 30 300
4  4 40 400
5  5 50 500
> names(data) <- c("Column1", "Column2", "Column3")
> data
  Column1 Column2 Column3
1         1      10      100
2         2      20      200
3         3      30      300
4         4      40      400
5         5      50      500
```

# Importing Data Files

Comma-separated values (CSV) files

We can set the delimiter with `sep`.

If it is tab delimited, use `sep="\t"`.

```
data <- read.csv("datafile.csv", sep="\t")
```

If it is space-delimited, use `sep=" "`.

```
data <- read.csv("datafile.csv", sep=" ")
```

# Importing Data Files

## Reading Tabular Data Files

Tabular data files are text files with a simple format:

- Each line contains one record.
- Within each record, fields (items) are separated by a one-character delimiter, such as a space, tab, colon, or comma.
- Each record contains the same number of fields.

We want to read a text file that contains a table of data.

`read.table` function is used and it returns a data frame.

```
read.table("FileName")
```

# Importing Data Files

## Reading Tabular Data Files

Data:

1 10 100

2 20 200

3 30 300

4 40 400

5 50 500

Saved in example3.txt

```
> data <- read.table("example3.txt", sep=" ")
```

```
> data
```

	V1	V2	V3
1	1	10	100
2	2	20	200
3	3	30	300
4	4	40	400
5	5	50	500

# Importing Data Files

## Reading Tabular Data Files

```
RGui (32-bit)
> data <- read.table("example3.txt", sep=" ")
> data
```

	V1	V2	V3
1	1	10	100
2	2	20	200
3	3	30	300
4	4	40	400
5	5	50	500