



## GETTING STARTED WITH **R**

# Contents

Downloading R from the CRAN server.....	1
Installing R by executing the R-3.2.2.exe file.....	3
Getting Started.....	8
Commands and Syntax.....	8
Data Structure.....	9
Graphics.....	12
Data Analysis.....	14



# Downloading R from the CRAN server

## Step1

Go to <http://cran.r-project.org> website



## Step2

The Comprehensive R Archive Network

---

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

- Here we can download latest R version for 'Linux', 'Mac', and 'Windows' operating System
- Click on [Download R for Windows](#).



## Step3

R for Windows

Subdirectories:

<a href="#">base</a>	Binaries for base distribution (managed by Duncan Murdoch). This is what you want to <a href="#">install R for the first time</a> .
<a href="#">contrib</a>	Binaries of contributed packages (managed by Uwe Ligges). There is also information on <a href="#">third party software</a> available for CRAN Windows services and corresponding environment and make variables.
<a href="#">Rtools</a>	Tools to build R and R packages (managed by Duncan Murdoch). This is what you want to build your own packages on Windows, or to build R itself.

- All the information and file required for R intallation are available in the subdirectories by name 'base'.
- Now click on [base](#) .

## Step4

### R-3.2.2 for Windows (32/64 bit)

[Download R 3.2.2 for Windows](#) (62 megabytes, 32/64 bit)

[Installation and other instructions](#)

[New features in this version](#)

If you want to double-check that the package you have downloaded exactly matches the package distributed by R, you can compare the [md5sum](#) of the .exe to the [true fingerprint](#). You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

- Here we can download R installation file and get other instruction and features of R.
- Click on [Download R.3.2.2 for Windows](#) to download the R installation file.



## Step5

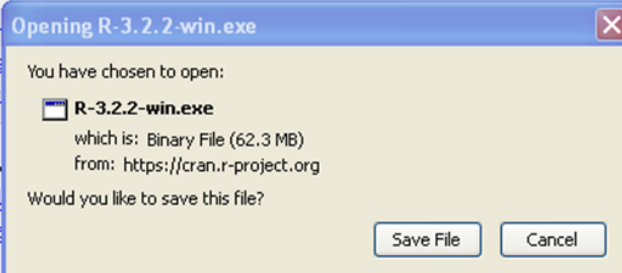
### R-3.2.2 for Windows (32/64 bit)


[Download](#)

[Installation and other instructions](#)

[New features in this version](#)

If you want to double-check that the package you have downloaded exactly matches the package distributed by R, you can compare the [md5sum](#) of the .exe to the [true fingerprint](#). You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

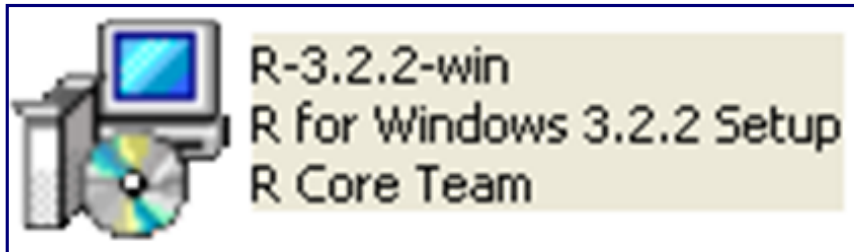


- The single file 'R-3.2.2-win.exe' contains components files both for 32bit and 64bit Windows OS.
- Click  tab to start downloading.

## Installing R by executing the R-3.2.2.exe file

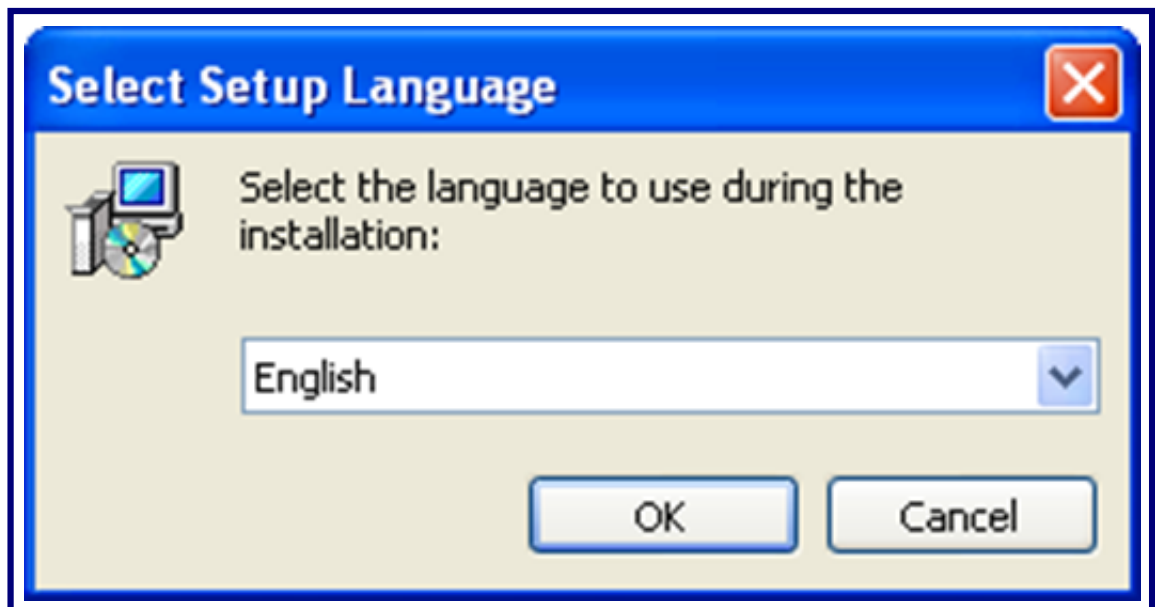
### Step1

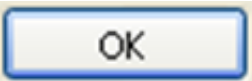
Double click on the downloaded file.



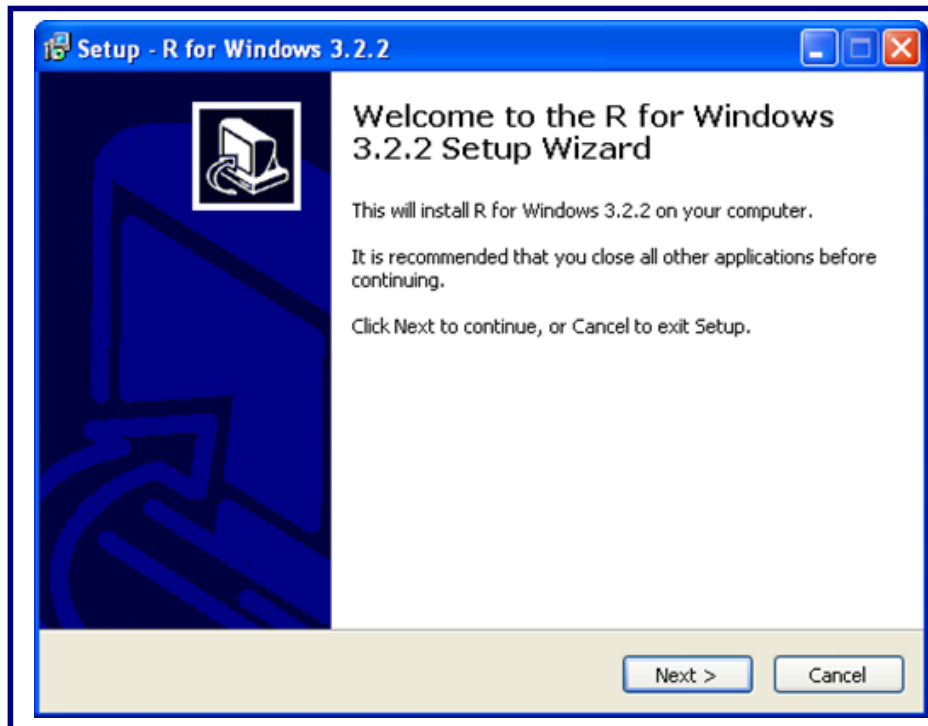
### Step2

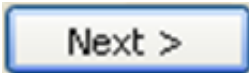
- a. Here we can choose any of the 24 languages. Let us take the default one which is English.



- b. Click on  tab.

## Step3

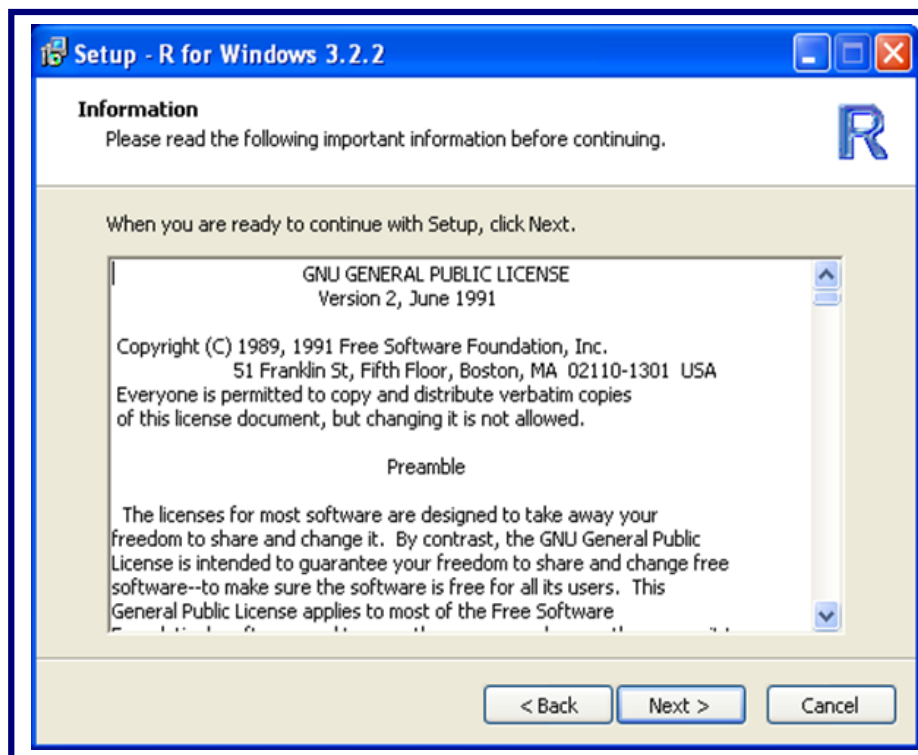



Click on  tab for proceeding.



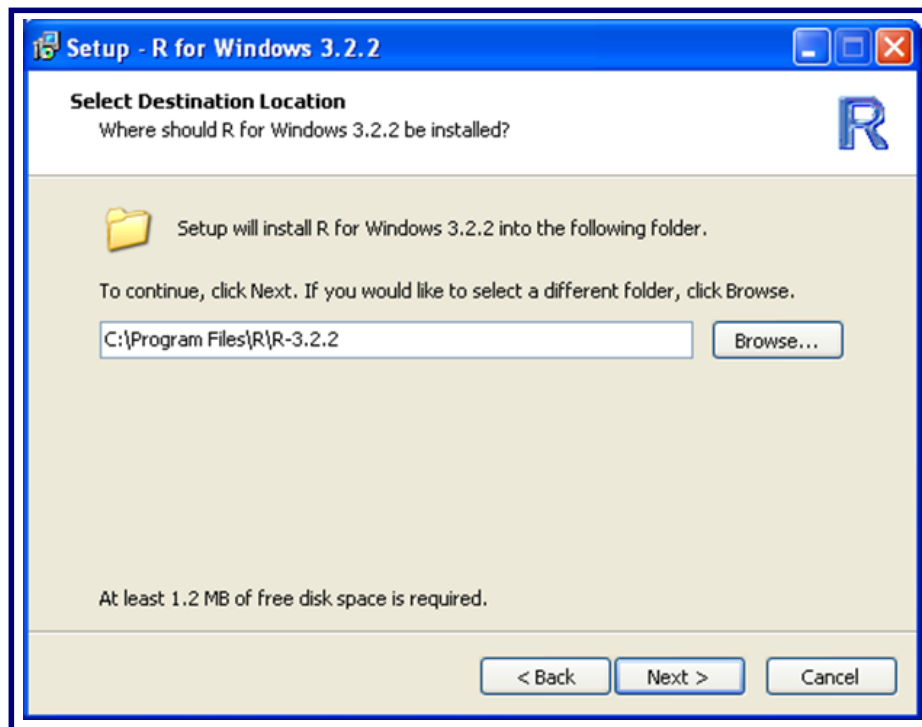
## Step4

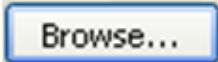
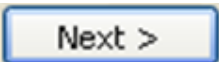
- Now read the license agreement carefully and continue.



- Click on  tab.

## Step5

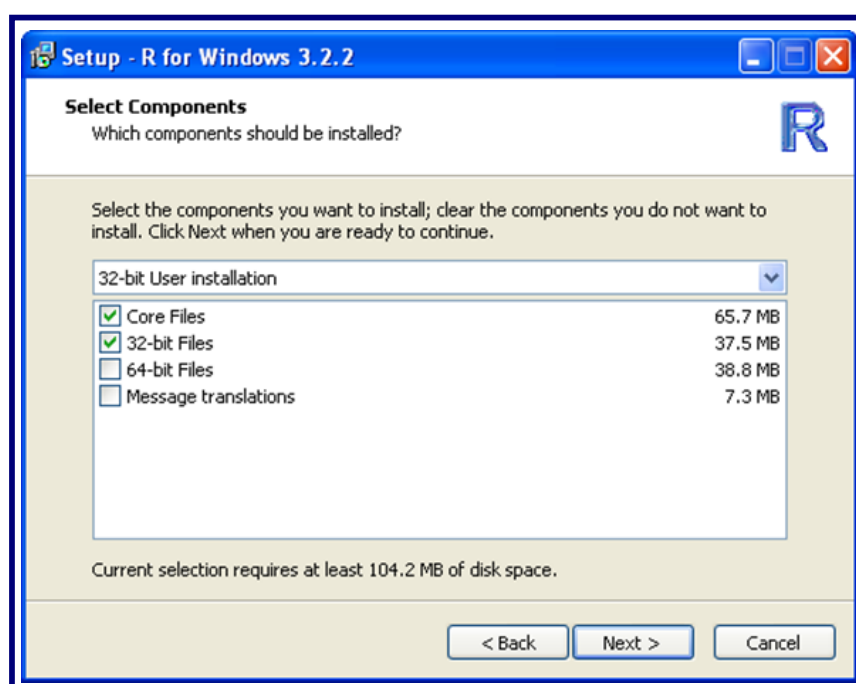


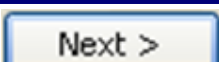
- Here we can change the default path by clicking  tab, if we wish to.
- Click on  tab.



## Step6

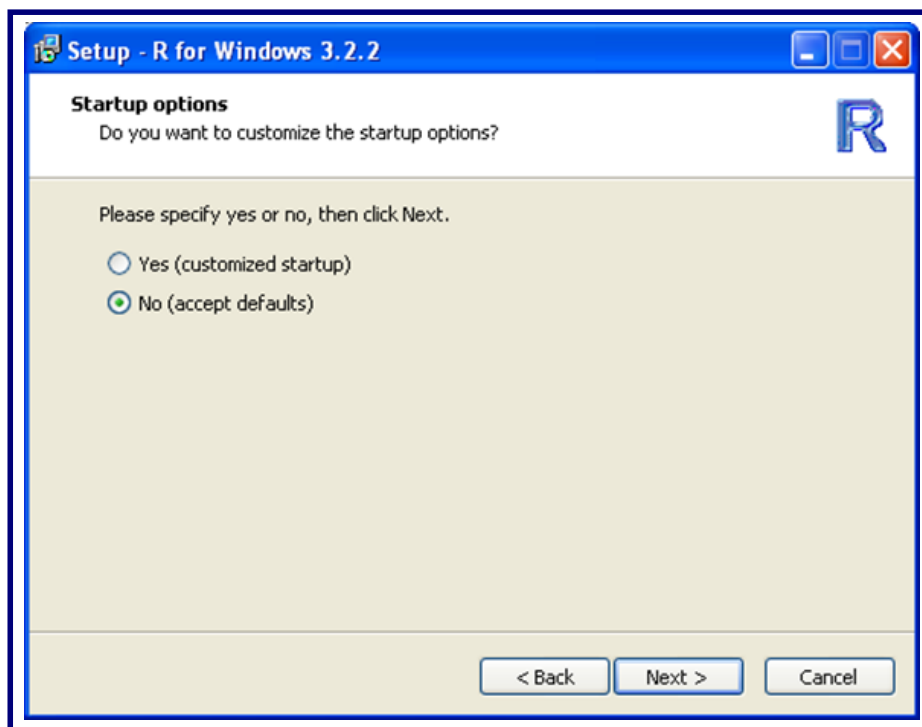
- Select the components depending on whether you are using 32bit or 64bit Windows OS.



- Click on  tab.

## Step7

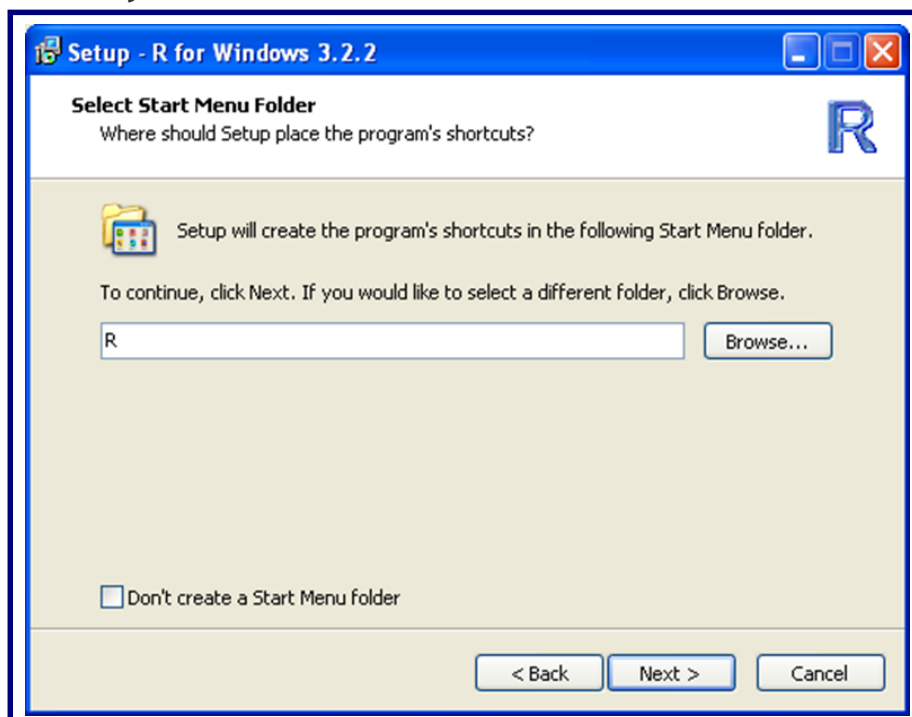
- a. Customize installation and Default installation are applicable here. We will go with the default setup.

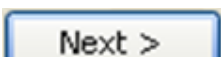


- b. Click on  tab.

## Step8

- a. Here we can create or select an older Start Menu folder. By default it is R.

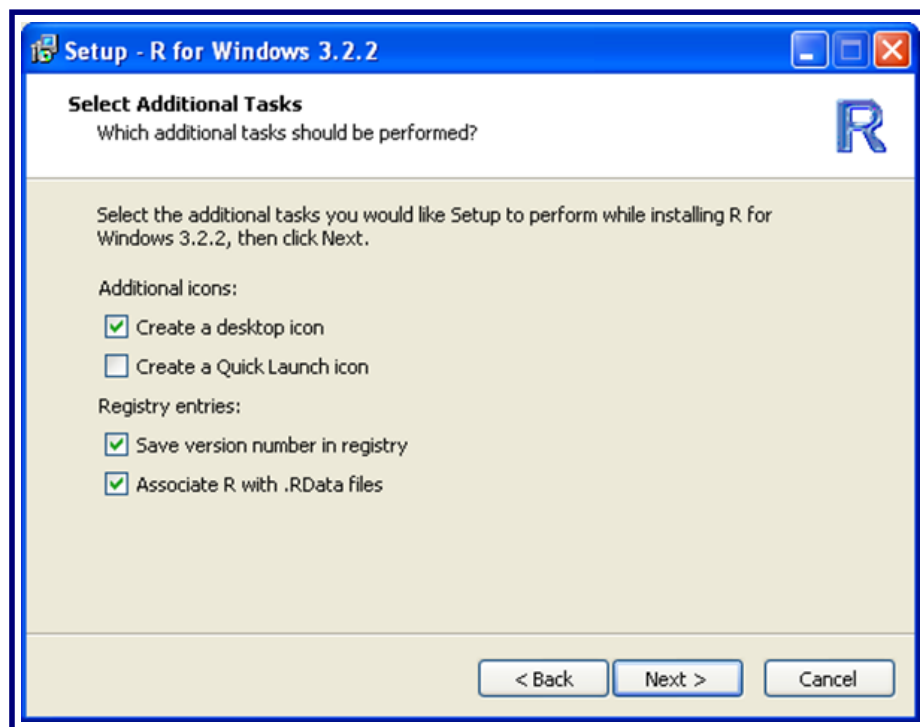


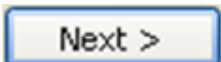
- b. Click on  tab.



## Step9

- a. Now select the additional task needed to be performed with the installation. Here the default options are selected.



- b. Click on  tab.

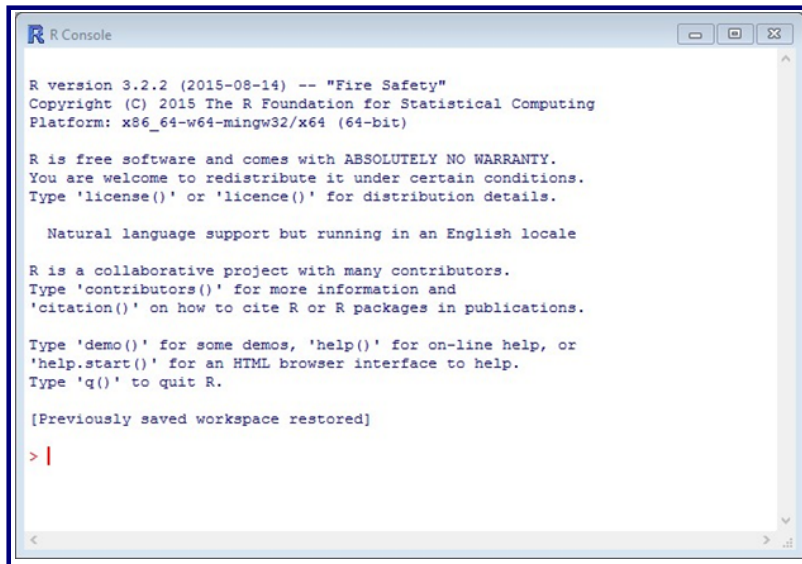
## Step10



- Click on  tab to finish the intallation.

## Getting started

When R starts you will see a window called the R Console. This is where you type your commands and see the text results. R Console looks as follows.



```
R Console

R version 3.2.2 (2015-08-14) -- "Fire Safety"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> |
```

All commands to be typed here at the '**>**' prompt.

## Commands and Syntax

Commands are executed at the command prompt. To add 2+2, type 2+2 in the console and hit the Enter Key to see the results.

```
> 2+2
```

This returns

```
[1] 4
```

We can assign a value to a variable and retrieve it by the variable name. Values can be assigned to a variable with '**=**', '**<-**' or '**->**' assign operators. Let us assign 4 to the variable **x**, 5 to the variable **y** and 6 to the variable **z**.

```
> x = 4
```

```
> y <- 5
```

```
> 6 -> z
```



R has operators for addition (+), subtraction (-), multiplication (\*) etc. To add x, y and z execute the following command.

```
> x + y + z  
[1] 15
```

To multiply x and y use x\*y:

```
> x*y  
[1] 20
```

Mathematical operations can be done with the mathematical functions such as exponential (exp()), logarithm(log()) etc.

Let us find the log value of x.

```
> log(x)  
[1] 1.386294
```

## Data Structure

Data structures help us store different types of data in different formats.

The basic important data structures in R are:

- Vectors
- Matrices
- Arrays
- Factors
- Lists
- Dataframes

## Vectors

Vectors are the simplest data structures in R. They are the collection of values of same data type stored inside a single entity. It is akin to a single row or column of values stored inside a variable. Vectors are of different types, namely

- i) Numeric vector
- ii) Character vector
- iii) Logical vector

Numerical vectors: { 12,15, 19, 23, 25, 27, 30 }

Character vectors: { "Red", "Green", "Blue", "Orange", "Yellow" }

Logical vectors: { FALSE, TRUE, TRUE, FALSE, TRUE, TRUE }

A vector can be created using `c()` command. Let us create a vector with the elements being numbers, namely 5, 25, 30, 20, 15 and store it in `x`.

```
> x=c(5, 25, 30, 20, 15)
```

Now the vector `x` is created with the given elements. Type `x` in R console and hit Enter to retrieve the elements of `x`.

```
> x
[1] 5 25 30 20 15
```

We can apply several useful functions on a vector. Let us sort the elements of `x` using the `sort()` function.

```
> sort(x)
[1] 5 15 20 25 30
```

The `sum()` function is used to get the sum of elements of the given vector.

```
> sum(x)
[1] 95
```

## Matrices

Matrices are another basic important data structure in R that can be used to store two dimensional data.

A matrix can be created with the `matrix()` function. Let us create a matrix with 3 rows and 3 columns which will be containing {1,2,3,4,5,6,7,8,9} as elements and assign it in `x`.

```
> x=matrix(data=c(1,2,3,4,5,6,7,8,9),nrow=3,ncol=3)
```

```
> x
      [,1] [,2] [,3]
[1,]    1    4    7
[2,]    2    5    8
[3,]    3    6    9
```

Each element in a matrix has a unique position which is (row number, column number). We may extract any element from a matrix by typing that position in '[ ]' after the matrix.

To extract the element which is at 3rd row, 3rd column of **x** execute the following command.

```
> x[3,3]
[1] 9
```

The 2nd row of x can be extracted by executing the following command.

```
> x[2, ]
[1] 2 5 8
```

Similarly, to extract 1st column of x execute the following command.

```
> x[,1]
[1] 1 2 3
```

Similarly, to extract 1st column of x execute the following command.

```
> x[,1]
[1] 1 2 3
```

## R Datasets

R has lots of inbuilt datasets in the form of vector, matrix, dataframe, list etc. These data can be loaded in R with the **data()** function.

Let us load the '**pressure**' dataset for further use.

```
> data(pressure)
```

The variables can be seen with the function **names()**. Let us see the list of variables present in the '**pressure**' dataset.

```
> names(pressure)
[1] "temperature" "pressure"
```

The values of variables can be accessed with the '\$' operator. To see the variable 'temperature' from the dataset 'pressure' use '**pressure\$temperature**'.

```
> pressure$temperature  
[1] 0 20 40 60 80 100 120 140 160 180 200 220 240
```

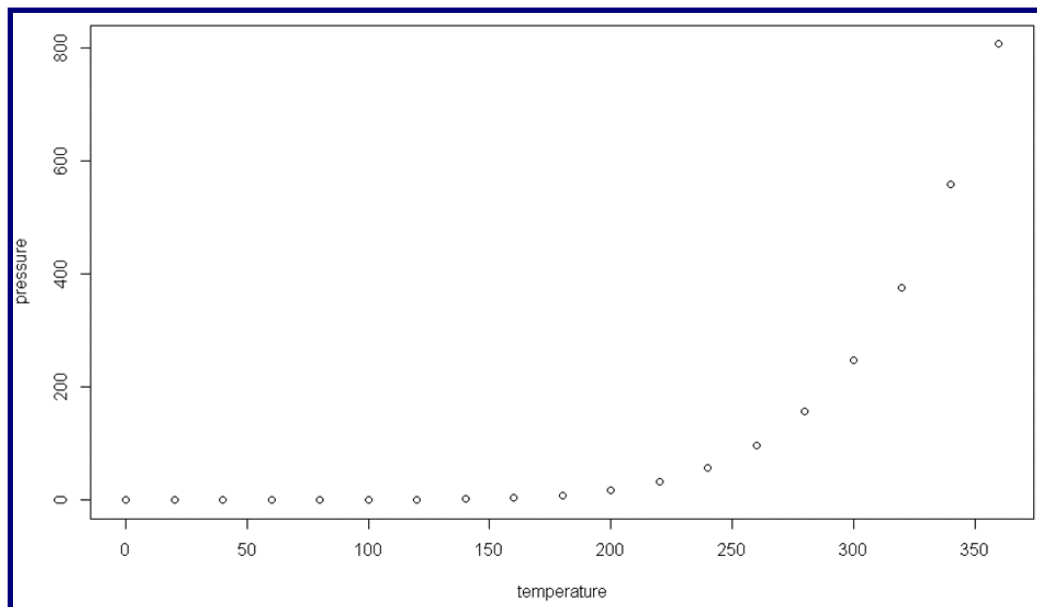
## Graphics

### Plot

Plots can be drawn using the **plot()** function.

Let us now create a scatter plot of pressure versus temperature.

```
> plot(pressure$temperature, pressure$pressure)
```



## Histogram

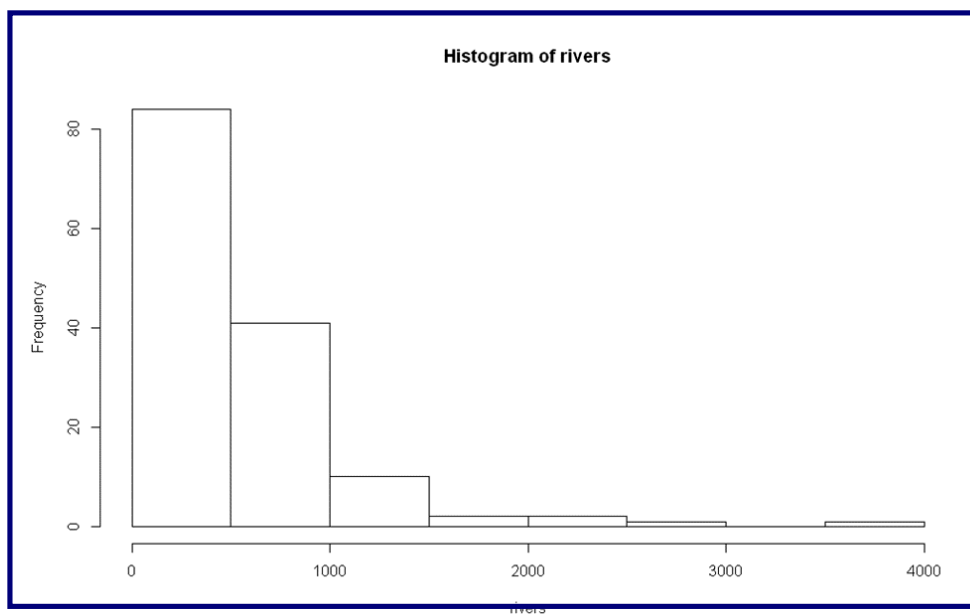
A histogram helps us to get an idea about the frequency distribution of a numeric variable. The `hist()` function can be used to build a histogram.

First load the e dataset 'rivers'.

```
> data(rivers)
```

Now draw a histogram of the 'rivers' dataset.

```
> hist(rivers)
```



## Frequency

Frequencies are the number of occurrences of individual observations in a dataset. A frequency table can be created using the `table()` function.

First load the e dataset 'chickwts'.

```
> data(chickwts)
```

Let us examine internal structure of 'chickwts'.

```
> str(chickwts)
```

```
'data.frame':  71 obs. of  2 variables:  
 $ weight: num  179 160 136 227 217 168 108 124 143 140 ...  
 $ feed  : Factor w/ 6 levels "casein",...: 2 2 2 2 2 2 ...
```

From the output it is clear that there are two variables 'feed' (factor or categorical) and 'weight' (numeric).

Now let us create the frequency table of the categorical variable 'feed'.

```
> table(chickwts$feed)
```

```
feed  
  casein horsebean  linseed  meatmeal  soybean sunflower  
      12         10       12        11        14         12
```

The output table is showing categories of 'feed' and corresponding frequencies.



## Summary Statistics

There are several statistics such as mean, median, mode, variance etc. can be computed for a numerical variable. The summary statistics shows few of them. The `summary()` function can be used to compute summary statistics (mean and five point summary statistics).

Here we will use the same 'chickwts' dataset.

Let us calculate mean and five point summary statistics of the numeric variable 'weight'.

```
> summary(chickwts$weight)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
108.0	204.5	258.0	261.3	323.5	423.0

.

-----



# Learn R Online

## Self paced Course

---

### Key Features

- \* Self paced learning program on an unique simulated environment
- \* Teaches concepts, execution as well as interpretation of results
- \* Quizzes and assignments

### Coverage

- |                                 |                                     |
|---------------------------------|-------------------------------------|
| 1. Introduction to R            | 10. Descriptive Statistics          |
| 2. Installing R for Windows     | 11. Cross Tabulation                |
| 3. R interface                  | 12. T-test and Non Parametric tests |
| 4. R Environment                | 13. One way and Two way ANOVA       |
| 5. Data structures in R         | 14. Correlation                     |
| 6. Importing and Exporting Data | 15. Linear Regression               |
| 7. Data types                   | 16. Time Series                     |
| 8. Simple plots and graphs      | 17. Formatting R output             |
| 9. Frequencies                  |                                     |

[CLICK HERE TO REGISTER](#)

### Contact Us

---

Predictive Analytics Solution Pvt. Ltd. #5, 1st Floor, 1st Cross, P&T Layout, Horamavu, Bangalore 560043, Phone: +91-80-41129655, 9986764814  
[www.letslearnanalytics.com](http://www.letslearnanalytics.com), Mail: [know.more@letslearnanalytics.com](mailto:know.more@letslearnanalytics.com)

