

# Data Manipulation

By Nilosree Sengupta,  
Coursera Project Network

**Data Manipulation** is the process of exploring given data and converting unstructured and raw data to structured data for easier analysis that include -filtering, sorting, modifying, creating new data from existing, summarising, arranging, deleting, slicing, inserting data, etc. Summarising data also gives clarity and insights.

We are working with the tidyverse package : **dplyr**

The following are R functions and the concepts used for Data manipulation :

## i. To extract data by rows(select particular rows):

R function	Definition
filter()	To extract particular rows depending on columns.
slice()	To select particular rows based on location
slice_head()	To select particular rows based on location from the top of the data
slice_tail()	To select particular rows based on location from the bottom of data
arrange()	To order the required rows in ascending or descending order.

## ii. To extract data by column:

R function	Definition
select()	To extract particular columns based on given data
mutate()	To compute new columns from existing columns based on given data.

## iii. To summarise data:

R function	Definition
summarise()	To summarise a data frame into

	a single row or value as output
--	---------------------------------

## iv. To group data :

R function	Definition
group_by()	To group data by particular cases

## v. With the dplyr functions of summarise() and group\_by(), we are going to do the Descriptive Statistics.

R function	The Descriptive Statistical Measure
sum()	Summation
mean()	Mean/Average
median()	Median
min()	Minimum
max()	Maximum
range()	Range
var()	variance
sd()	Standard deviation
summary()	Summary
quantile()	Quantile

## vi. Pipe operator: %>%

It is used to connect the result of the output of one function to another function. Hence it is time saving. We can directly perform multiple operations in one line with % operator.