**Handling Missing Values in R Programming**

As the name indicates, Missing values are those elements that are not known. NA or NaN are reserved words that indicate a missing value in [R Programming language](https://www.geeksforgeeks.org/introduction-to-r-programming-language/) for q arithmetical operations that are undefined.

**R – Handling Missing Values**

Missing values are practical in life. For example, some cells in spreadsheets are empty. If an insensible or impossible arithmetic operation is tried then NAs occur.

**Dealing Missing Values in R**

Missing Values in R, are handled with the use of some pre-defined functions:

**is.na() Function for Finding Missing values:**

A logical vector is returned by this function that indicates all the NA values present. It returns a Boolean value. If NA is present in a vector it returns TRUE else FALSE.

|  |
| --- |
| x<- c(**NA**, 3, 4, **NA**, **NA**, **NA**)  is.na(x) |

**Output:**

[1] TRUE FALSE FALSE TRUE TRUE TRUE

**Properties of Missing Values:**

* For testing objects that are NA use is.na()
* For testing objects that are NaN use is.nan()
* There are classes under which NA comes. Hence integer class has integer type NA, the character class has character type NA, etc.
* A NaN value is counted in NA but the reverse is not valid.

The creation of a vector with one or multiple NAs is also possible.

* R

|  |
| --- |
| x<- c(**NA**, 3, 4, **NA**, **NA**, **NA**)  x |

**Output:**

[1] NA 3 4 NA NA NA

**Removing NA or NaN values**

There are two ways to remove missing values:

**Extracting values except for NA or NaN values:**

**Example 1:**

* R

|  |
| --- |
| x <- c(1, 2, **NA**, 3, **NA**, 4)  d <- is.na(x)  x[! d] |

**Output:**

[1] 1 2 3 4

**Example 2:**

* R

|  |
| --- |
| x <- c(1, 2, 0 / 0, 3, **NA**, 4, 0 / 0)  x  x[! is.na(x)] |

**Output:**

[1] 1 2 NaN 3 NA 4 NaN  
  
[1] 1 2 3 4

A function called **complete.cases()** can also be used. This function also works on data frames.

**Missing Value Filter Functions**

The modeling functions in R language acknowledge a **na.action** argument which provides instructions to the function regarding its response if NA comes in its way.

And hence this way the function calls one of the missing value filter functions. Missing Value Filter Functions alter the data set and in the new data set the value of NAs has been changed. The default Missing Value Filter Function is **na.omit**. It omits every row containing even one NA. Some other Missing Value Filter Functions are:

* **na.omit**– omits every row containing even one NA
* **na.fail**– halts and does not proceed if NA is encountered
* **na.exclude**– excludes every row containing even one NA but keeps a record of their original position
* **na.pass**– it just ignores NA and passes through it
* R

|  |
| --- |
| # Creating a data frame  df <- data.frame (c1 = 1:8,                    c2 = factor (c("B", "A", "B", "C", "A", "C", "B", "A")))    # Filling some NA in data frame  df[4, 1] <- df[6, 2] <- **NA**    # Printing all the levels(NA is not considered one)  levels(df$c2)    # fails if NA is encountered  na.fail (df)    # excludes every row containing even one NA  na.exclude (df) |

**Output:**

[1] "A" "B" "C"  
Error in na.fail.default(df) : missing values in object  
Calls: na.fail -> na.fail.default  
Execution halted

**Find and Remove NA or NaN values from a dataset**

In R we can remove and find missing values from the entire dataset. there are some main functions we can use and perform the tasks.  
First, we will create one data frame and then we will find and remove all the missing values which are present in the data.

* R

|  |
| --- |
| # Create a data frame with 5 rows and 3 columns  data <- data.frame(    A = c(1, 2, **NA**, 4, 5),    B = c(**NA**, 2, 3, **NA**, 5),    C = c(1, 2, 3, **NA**, **NA**)  )    # View the resulting data frame  Data |

**Output:**

A B C  
1 1 NA 1  
2 2 2 2  
3 NA 3 3  
4 4 NA NA  
5 5 5 NA

**Find all the missing values in the data**

* R

|  |
| --- |
| # Finding missing values in data.  sum(is.na(data)) |

**Output:**

[1] 5

**Find all the missing values in the columns**

* R

|  |
| --- |
| # Finding missing values column wise  colSums(is.na(data)) |

**Output:**

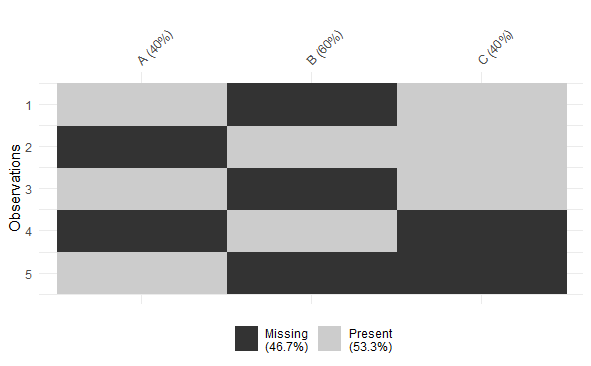
A B C   
1 2 2

**Visualization of missing values of a dataset**

* R

|  |
| --- |
| # Install and load the 'visdat' package  install.packages("visdat")  library(visdat)    # Create a data frame with missing values  data <- data.frame(    A = c(1, **NA**, 3, **NA**, 5),    B = c(**NA**, 2, **NA**, 4, **NA**),    C = c(1, 2, 3, **NA**, **NA**)  )    # Plot the missing value diagram  vis\_miss(data) |

**Output:**



*Handling missing values in R*