How can I use is.na in R to identify missing values?

Authored by

stats writer

May 2, 2024

The function "is.na" in R is used to identify missing values in a dataset. It returns a logical vector indicating which elements in the dataset are NA (Not Available) or NULL. This function is commonly used in data cleaning and analysis to detect and handle missing data. By identifying missing values, users can make informed decisions on how to handle them, such as imputing or removing them from the dataset. Overall, the "is.na" function in R is a powerful tool for efficiently managing and analyzing data with missing values.

Use is.na in R (With Examples)

You can use the is.na() function in R to check for missing values in vectors and data frames.

#check if each individual value is NA
is.na(x)

#count total NA values
sum(is.na(x))

#identify positions of NA values
which(is.na(x))

The following examples show how to use this function in practice.

Example 1: Use is.na() with Vectors

The following code shows how to use the is.na() function to check for missing values in a vector:

```
#define vector with some missing values
x <- c(3, 5, 5, NA, 7, NA, 12, 16)

#check if each individual value is NA
is.na(x)
```

```
FALSE FALSE FALSE TRUE FALSE TRUE FALSE FALSE
```

```
#count total NA values
sum(is.na(x))
```

```
2
```

```
#identify positions of NA values
which(is.na(x))
```

```
4 6
```

From the output we can see:

There are 2 missing values in the vector.The missing values are located in position 4 and 6.

Example 2: Use is.na() with Data Frames

The following code shows how to use the is.na()

function to check for missing values in a data frame:

```
#create data frame
df <- data.frame(var1=c(1, 3, 3, 4, 5),
var2=c(7, NA, NA, 3, 2),
var3=c(3, 3, 6, NA, 8),
var4=c(NA, 1, 2, 8, 9))

#view data frame
df

var1 var2 var3 var4
1 1 7 3 NA
2 3 NA 3 1
3 3 NA 6 2
4 4 3 NA 8
5 5 2 8 9

#find total NA values in data frame
sum(is.na(df))

4

#find total NA values by column
sapply(df, function(x) sum(is.na(x)))
```

var1 var2 var3 var4
0 2 1 1

From the output we can see that there are 4 total NA values in the data frame.

We can also see:

There are 0 NA values in the 'var1' column.There are 2 NA values in the 'var2' column.There are 1 NA values in the 'var3' column.There are 1 NA values in the 'var4' column.

The following tutorials explain other useful functions that can be used to handle missing values in R.

How to Use is.null in R