

# How can I find and count missing values in R, and what are some examples of missing value scenarios?

Authored by  
**stats writer**

May 12, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can I find and count missing values in R, and what are some examples of missing value scenarios?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=143791>

In R, missing values refer to any data points that are not available or could not be recorded. These values can significantly affect the accuracy and reliability of data analysis. To find and count missing values in R, the "is.na()" function can be used, which returns a logical vector indicating the presence of missing values. This can then be combined with the "sum()" function to calculate the total number of missing values.

Some examples of missing value scenarios in R include:

1. Incomplete data entry: This occurs when data is not entered for certain variables or observations.
2. Measurement errors: Sometimes, data may not be accurately recorded, leading to missing values in the dataset.
3. System failures: Technical issues or errors in data collection systems can result in missing values.
4. Non-response: In surveys or questionnaires, some individuals may choose not to answer certain questions, resulting in missing values.
5. Outliers: Extreme values that are not representative of the data can also be considered as missing values in certain cases.

It is important to identify and handle missing values appropriately in data analysis to avoid biased results. Various techniques such as imputation, deletion, or modeling can be used to address missing values in R.

## Find and Count Missing Values in R (With Examples)

**You can use the following methods to find and count missing values in R:**

### Method 1: Find Location of Missing Values

```
which(is.na(df$column_name))
```

## Method 2: Count Total Missing Values

```
sum(is.na(df$column_name))
```

The following examples show how to use these functions in practice.

Example 1: Find and Count Missing Values in One Column

Suppose we have the following data frame:

```
#create data frame
```

```
df <- data.frame(team=c('A', 'B', 'C', NA, 'E'),  
points=c(99, 90, 86, 88, 95),  
assists=c(NA, 28, NA, NA, 34),  
rebounds=c(30, 28, 24, 24, NA))
```

```
#view data frame
```

```
df
```

```
team points assists rebounds
```

```
1 A 99 NA 30
```

```
2 B 90 28 28
```

```
3 C 86 NA 24
```

```
4 NA 88 NA 24
```

```
5 E 95 34 NA
```

We can use the following code to identify which positions have missing values in the 'assists' column and find the total missing values in the 'assists' column:

```
#identify locations of missing values in 'assists' column  
which(is.na(df$assists))
```

```
1 3 4
```

```
#count total missing values in 'assists' column  
sum(is.na(df$assists))
```

```
3
```

From the output we can see that positions 1, 3, and 4 have missing values in the 'assists' column and there are a total of 3 missing values in the column.

Example 2: Count Missing Values in All Columns

The following code shows how to count the total missing values in every column of a data frame:

```
#create data frame  
df <- data.frame(team=c('A', 'B', 'C', NA, 'E'),  
points=c(99, 90, 86, 88, 95),
```

```
assists=c(NA, 28, NA, NA, 34),  
rebounds=c(30, 28, 24, 24, NA))
```

**#count total missing values in each column of data frame**

```
sapply(df, function(x) sum(is.na(x)))
```

```
team points assists rebounds  
1 0 3 1
```

**From the output we can see:**

The 'team' column has 1 missing value. The 'points' column has 0 missing values. The 'assists' column has 3 missing values. The 'rebounds' column has 1 missing value.

**Example 3: Count Missing Values in Entire Data Frame**

The following code shows how to count the total missing values in an entire data frame:

```
#create data frame  
df <- data.frame(team=c('A', 'B', 'C', NA, 'E'),  
points=c(99, 90, 86, 88, 95),  
assists=c(NA, 28, NA, NA, 34),
```

```
rebounds=c(30, 28, 24, 24, NA))
```

```
#count total missing values in entire data frame
```

```
sum(is.na(df))
```

5

From the output we can see that there are 5 total missing values in the entire data frame.

The following tutorials explain how to perform other common operations with missing values in R: