How can we count the number of non-NA values in R?

Authored by

stats writer

June 29, 2024

To count the number of non-NA values in R, we can use the function "sum(!is.na())". This function will first check if each value is NA or not, and then sum up the number of non-NA values. This method is useful for handling missing data and obtaining an accurate count of the available values in a dataset. It can be applied to any data type, such as vectors, data frames, or matrices, making it a versatile tool for data analysis in R. By using this function, we can efficiently and accurately determine the number of non-NA values in our dataset, which is crucial for various statistical analyses and modeling processes.

Count Non-NA Values in R (3 Examples)

You can use the following methods to count non-NA values in R:

Method 1: Count Non-NA Values in Entire Data Frame

sum(!is.na(df))

Method 2: Count Non-NA Values in Each Column of Data Frame

colSums(!is.na(df))

Method 3: Count Non-NA Values by Group in Data Frame

library(dplyr)

df %>%

```
group_by(var1) %>%
summarise(total_non_na = sum(!is.na(var2)))
```

The following example shows how to use each of these methods in practice with the following data frame:

```
#create data frame
df <- data.frame(team=c('A', 'A', 'A', 'A', 'B', 'B', 'B', 'B'),
points=c(12, NA, 30, 32, 20, 22, 17, NA),
rebounds=c(10, 8, 9, 13, NA, 20, 8, 7))

#view data frame
df

team points rebounds
1 A 12 10
2 A NA 8
3 A 30 9
4 A 32 13
5 B 20 NA
6 B 22 20
7 B 17 8
8 B NA 7
```

Method 1: Count Non-NA Values in Entire Data Frame

The following code shows how to count the total non-NA values in the entire data frame:

#count non-NA values in entire data frame
sum(!is.na(df))

21

From the output we can see that there are 21 non-NA values in the entire data frame.

Method 2: Count Non-NA Values in Each Column of Data Frame

The following code shows how to count the total non-NA values in each column of the data frame:

#count non-NA values in each column
colSums(!is.na(df))

team points rebounds
8 6 7

From the output we can see:

There are 8 non-NA values in the team column.There

are 6 non-NA values in the points column.There are 7 non-NA values in the rebounds column.

Method 3: Count Non-NA Values by Group

The following code shows how to count the total non-NA values in the points column, grouped by the team column:

```
library(dplyr)
df %>%
group_by(team) %>%
summarise(total_non_na = sum(!is.na(points)))

# A tibble: 2 x 2
team total_non_na

1 A 3
2 B 3
```

From the output we can see:

There are 3 non-NA values in the points column for team A.There are 3 non-NA values in the points column for team B.

Additional Resources

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