

# How can I handle “undefined columns selected” in R?

Authored by  
**stats writer**

April 20, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can I handle “undefined columns selected” in R?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=137396>

The issue of "undefined columns selected" in R refers to a common error that occurs when attempting to select columns that do not exist in a given dataset. This error can be handled by first ensuring that the columns being selected are present in the dataset and correctly spelled. If the issue persists, it may be necessary to check for any missing data or inconsistencies in the dataset. In some cases, using the "drop = FALSE" argument in the selection statement can also help to avoid this error. Additionally, utilizing the "tryCatch" function can be useful in identifying the specific source of the error and implementing appropriate measures to handle it. By following these steps, one can effectively handle the issue of "undefined columns selected" in R and ensure smooth execution of their data analysis.

## Handle "undefined columns selected" in R

**One of the most common errors that you'll encounter in R is:**

**undefined columns selected**

**This error occurs when you try to select a subset of a data frame and forget to add a comma.**

**For example, suppose we have the following data frame in R:**

```
#create data frame with three variables
```

```
data <- data.frame(var1 = c(0, 4, 2, 2, 5),
```

```
var2 = c(5, 5, 7, 8, 9),
```

```
var3 = c(2, 7, 9, 9, 7))
```

```
#view DataFrame
```

**data**

**var1 var2 var3**

**1 0 5 2**

**2 4 5 7**

**3 2 7 9**

**4 2 8 9**

**5 5 9 7**

**Now suppose we attempt to select all rows where var1 is greater than 3:**

**data**

**Error in `**

**var1 var2 var3**

**2 4 5 7**

**5 5 9 7**

**The reason you need to add a comma is because R uses the following syntax for subsetting data frames:**

**data**

If you only type `data`, then you're telling R to return the rows where `var1>3`, but you're not telling R which columns to return.

By using `data[,3:5]`, you're telling R to return the rows where `var1>3` and *all* of the columns in the data frame. An equivalent command would be `data`.

```
data
```

```
var1 var2 var3
```

```
2 4 5 7
```

```
5 5 9 7
```

Notice that this command returns the same subset of data as before.

*You can find more R tutorials [here](#).*