

How can I resolve the error “Duplicate identifiers for rows” when trying to use the “Fix in R” function?

Authored by
stats writer

June 26, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I resolve the error “Duplicate identifiers for rows” when trying to use the “Fix in R” function?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=153521>

The "Fix in R" function is a tool used for resolving errors in data analysis using the R programming language. However, sometimes when attempting to use this function, an error may occur stating "Duplicate identifiers for rows", which can hinder the effectiveness of the tool. This error typically occurs when there are duplicate rows in the data, causing confusion and difficulty in identifying unique data points. To resolve this error, it is necessary to review the data and remove any duplicate rows before attempting to use the "Fix in R" function again. By doing so, the tool will be able to accurately identify and fix any errors in the data, ensuring a successful data analysis process.

Fix in R: Error: Duplicate identifiers for rows

One error you may encounter in R is:

Error: Duplicate identifiers for rows

This error occurs when you attempt to use the `spread()` function to spread the values in one or more columns in a data frame into their own columns.

However, an error can occur if there is no unique ID for each row so there is no way to determine which values belong with which observations when performing the spread.

The following example shows how to fix this error in practice.

Example: How to Fix the Error

Suppose we have the following data frame in R that contains information about various basketball players:

```
#create data frame
```

```
df <- data.frame(player=rep(c('A', 'B'), each=4),  
year=rep(1:4, times=2),  
assists=c(4, 10, 4, 4, 3, 7, 7, 6),  
points=c(14, 6, 18, 7, 22, 9, 38, 4))
```

```
#view data frame
```

```
df
```

```
player year assists points
```

```
1 A 1 4 14
```

```
2 A 2 10 6
```

```
3 A 3 4 18
```

```
4 A 4 4 7
```

```
5 B 1 3 22
```

```
6 B 2 7 9
```

```
7 B 3 7 38
```

```
8 B 4 6 4
```

Now suppose we would like to transform the data frame so that we have the year column as the id column and create new columns called `assists_A`, `assists_B`,

points_A, and points_B to represent the assists and points values for players A and B during each year.

Since the values in the year column will not be unique (there will be two 1's, two 2's, etc.) the spread() function will produce an error.

However, we can use the pivot_wider() function with the following syntax to produce the desired data frame:

```
library(tidyr)
```

```
#spread the values in the points and assists columns
```

```
pivot_wider(data = df,
```

```
id_cols = year,
```

```
names_from = player,
```

```
values_from = c('assists', 'points'))
```

```
# A tibble: 4 x 5
```

```
year assists_A assists_B points_A points_B
```

```
1 1 4 3 14 22
```

```
2 2 10 7 6 9
```

```
3 3 4 7 18 38
```

```
4 4 4 6 7 4
```

Notice that we don't receive any error and we're able to successfully create the new columns that display the points and assists values for players A and B during each of the four years.

The following tutorials explain how to fix other common errors in R:

ARABPSYCHOLOGY.COM