

How can I use dplyr to join multiple data frames together?

Authored by
stats writer

April 18, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I use dplyr to join multiple data frames together?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=136664>

Dplyr is a popular R package that provides efficient tools for data manipulation and analysis. It includes a function called "join" which allows users to combine multiple data frames based on a common variable or set of variables. This can be helpful when working with large datasets or when trying to combine different sources of data. By using dplyr's join function, users can easily merge data frames and perform various operations such as filtering, sorting, and summarizing the data. This feature makes data management tasks more streamlined and organized, ultimately improving the overall data analysis process.

Join Multiple Data Frames Using dplyr

Often you may be interested in joining multiple data frames in R. Fortunately this is easy to do using the `left_join()` function from the `dplyr` package.

```
library(dplyr)
```

For example, suppose we have the following three data frames:

```
#create data frame
```

```
df1 <- data.frame(a = c('a', 'b', 'c', 'd', 'e', 'f'),  
b = c(12, 14, 14, 18, 22, 23))
```

```
df2 <- data.frame(a = c('a', 'a', 'a', 'b', 'b', 'b'),  
c = c(23, 24, 33, 34, 37, 41))
```

```
df3 <- data.frame(a = c('d', 'e', 'f', 'g', 'h', 'i'),  
d = c(23, 24, 33, 34, 37, 41))
```

To join all three data frames together, we can simply perform two left joins, one after the other:

```
#join the three data frames
```

```
df1 %>%
```

```
left_join(df2, by='a') %>%
```

```
left_join(df3, by='a')
```

```
a b c d
```

```
1 a 12 23 NA
```

```
2 a 12 24 NA
```

```
3 a 12 33 NA
```

```
4 b 14 34 NA
```

```
5 b 14 37 NA
```

```
6 b 14 41 NA
```

```
7 c 14 NA NA
```

```
8 d 18 NA 23
```

```
9 e 22 NA 24
```

```
10 f 23 NA 33
```

Note that you can also save the result of this join as a data frame:

```
#join the three data frames and save result as new data  
frame named all_data
```

```
all_data <- df1 %>%  
left_join(df2, by='a') %>%  
left_join(df3, by='a')
```

```
#view summary of resulting data frame  
glimpse(all_data)
```

Observations: 10

Variables: 4

\$ a <chr> "a", "a", "a", "b", "b", "b", "c", "d", "e", "f"

\$ b <dbl> 12, 12, 12, 14, 14, 14, 14, 18, 22, 23

\$ c <dbl> 23, 24, 33, 34, 37, 41, NA, NA, NA, NA

\$ d <dbl> NA, NA, NA, NA, NA, NA, NA, NA, 23, 24, 33

How to Filter Rows in R

How to Remove Duplicate Rows in R

How to Group & Summarize Data in R