

# How can I perform an inner join in R, and what are some examples of using this function in data analysis?

Authored by  
**stats writer**

June 29, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can I perform an inner join in R, and what are some examples of using this function in data analysis?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=159176>

An inner join in R is a function that combines two data sets by matching and merging rows that have a common value in a specified column. This allows for the identification of data relationships and can provide valuable insights in data analysis.

To perform an inner join in R, the "merge" function is used, where the first argument is the data frame on the left side of the join, the second argument is the data frame on the right side, and the "by" argument specifies the column to match on. The resulting data frame will contain only rows that have matching values in the specified column.

Some examples of using inner join in data analysis include combining customer data with purchase history to analyze buying behavior, merging employee data with performance metrics to identify top performers, and merging sales data with demographic information to target specific markets. Overall, inner join in R is a powerful tool for combining and analyzing data sets to gain deeper insights and make informed decisions.

## Do an Inner Join in R (With Examples)

There are two common ways to perform an inner join in R:

### Method 1: Use Base R

```
merge(df1, df2, by='column_to_join_on')
```

### Method 2: Use dplyr

```
library(dplyr)
```

```
inner_join(df1, df2, by='column_to_join_on')
```

Both methods will produce the same result, but the

**dplyr method will tend to work faster on extremely large datasets.**

**The following examples show how to use each of these functions in practice with the following data frames:**

**#define first data frame**

```
df1 = data.frame(team=c('A', 'B', 'C', 'D', 'E', 'F', 'G', 'H'),  
points=c(18, 22, 19, 14, 14, 11, 20, 28))
```

**df1**

**team points**

**1 A 18**

**2 B 22**

**3 C 19**

**4 D 14**

**5 E 14**

**6 F 11**

**7 G 20**

**8 H 28**

**#define second data frame**

```
df2 = data.frame(team=c('A', 'B', 'C', 'D', 'G', 'H'),  
assists=c(4, 9, 14, 13, 10, 8))
```

**df2**

**team assists**

**1 A 4**

**2 B 9**

**3 C 14**

**4 D 13**

**5 G 10**

**6 H 8**

**Example 1: Inner Join Using Base R**

**We can use the merge() function in base R to perform an inner join, using the 'team' column as the column to join on:**

```
#perform inner join using base R
```

```
df3 <- merge(df1, df2, by='team')
```

```
#view result
```

```
df3
```

**team points assists**

**1 A 18 4**

**2 B 22 9**

**3 C 19 14**

**4 D 14 13**

**5 G 20 10**

**6 H 28 8**

**Notice that only the teams that were in both datasets are kept in the final dataset.**

**Example 2: Inner Join Using dplyr**

**We can use the `inner_join()` function from the package to perform an inner join, using the 'team' column as the column to join on:**

```
library(dplyr)
```

```
#perform inner join using dplyr
```

```
df3 <- inner_join(df1, df2, by='team')
```

```
#view result
```

```
df3
```

```
team points assists
```

```
1 A 18 4
```

```
2 B 22 9
```

```
3 C 19 14
```

```
4 D 14 13
```

**5 G 20 10**

**6 H 28 8**

**Notice that this matches the result we obtained from using the merge() function in base R.**

**Additional Resources**

ARABPSYCHOLOGY.COM