

Sheth L.U.J. & Sir M.V. College

6. Combining and appending datasets using merge() or bind_rows() in R.

The screenshot shows the RStudio interface with two data frames loaded:

- study_performance**: A data frame with 26 rows and 8 columns. Columns include gender, race_ethnicity, parental_level_of_education, lunch, test_preparation_course, math_score, reading_score, and writing_score. The data shows various student characteristics and their corresponding scores.
- StudentPerformanceFactors**: A data frame with 26 rows and 10 columns. Columns include Hours_Studied, Attendance, Parental_Involvement, Access_to_Resources, Extracurricular_Activities, Sleep_Hours, Previous_Scores, Motivation_Level, Internet_Access, Tutoring_Sessions, Family_Income, and Teacher_Qualification. This data likely represents factors influencing student performance.

The two data frames are displayed side-by-side in the main RStudio window, allowing for comparison and merging.

The screenshot shows the RStudio interface with the **StudentPerformanceFactors** data frame loaded. The data frame has 26 rows and 10 columns, with the following structure:

	Hours_Studied	Attendance	Parental_Involvement	Access_to_Resources	Extracurricular_Activities	Sleep_Hours	Previous_Scores	Motivation_Level	Internet_Access	Tutoring_Sessions	Family_Income	Teacher_Qualification
1	23	84	Low	High	No	7	73	Low	Yes	0	Low	Medium
2	19	64	Low	Medium	No	8	59	Low	Yes	2	Medium	Medium
3	24	98	Medium	Medium	Yes	7	91	Medium	Yes	2	Medium	Medium
4	29	89	Low	Medium	Yes	8	98	Medium	Yes	1	Medium	Medium
5	19	92	Medium	Medium	Yes	6	65	Medium	Yes	3	Medium	High
6	19	88	Medium	Medium	Yes	8	89	Medium	Yes	3	Medium	Medium
7	29	84	Medium	Low	Yes	7	68	Low	Yes	1	Low	Medium
8	25	78	Low	High	Yes	6	50	Medium	Yes	1	High	High
9	17	94	Medium	High	No	6	80	High	Yes	0	Medium	Low
10	23	98	Medium	Medium	Yes	8	71	Medium	Yes	0	High	High
11	17	80	Low	High	No	8	88	Medium	No	4	Medium	High
12	17	97	Medium	High	Yes	6	87	Low	Yes	2	Low	High
13	21	83	Medium	Medium	Yes	8	97	Low	Yes	2	Medium	Medium
14	9	82	Medium	Medium	Yes	8	72	Medium	Yes	2	Medium	Medium
15	10	78	Medium	High	Yes	8	74	Medium	Yes	1	Low	Medium
16	17	68	Medium	Medium	No	8	70	Medium	Yes	2	Medium	Medium
17	14	60	Medium	Low	Yes	10	65	Low	Yes	0	High	Medium
18	22	70	Low	Medium	Yes	6	82	Medium	Yes	1	Low	High
19	15	80	Medium	Medium	Yes	9	91	Low	Yes	3	Low	Medium
20	12	75	Medium	High	Yes	7	58	Medium	Yes	3	Medium	Medium
21	29	78	Medium	Medium	No	5	99	High	Yes	0	High	Medium
22	19	99	Medium	High	No	6	84	Medium	Yes	1	Medium	High
23	20	74	Medium	High	No	8	89	Low	Yes	1	Medium	Medium
24	11	78	High	Medium	Yes	8	100	High	Yes	1	Low	Medium
25	17	65	Low	High	Yes	5	75	Medium	Yes	2	Low	Medium
26	21	62	High	Low	Yes	6	54	High	Yes	0	High	High

The data shows various student characteristics and their corresponding scores.

Name :- Priya Gupta
Roll no. :- S081

Sheth L.U.J. & Sir M.V. College

```
1 # Practical 6: Merge and Append
2 install.packages("dplyr")
3 library(dplyr)
4
5 # Create data sets
6 data_jan <- data.frame(
7   ID = c(1, 2, 3),
8   Name = c("Alice", "Bob", "Charlie"),
9   Jan_sales = c(100, 150, 200)
10 )
11
12 data_feb <- data.frame(
13   ID = c(1, 2, 3),
14   Name = c("Alice", "Bob", "Charlie"),
15   Feb_sales = c(120, 160, 210)
16 )
17
18 data_new_hires <- data.frame(
19   ID = c(4, 5),
20   Name = c("David", "Eva"),
21   Jan_sales = c(50, 60)
22 )
23
24 print("--- Data January ---")
25 print(data_jan)
26 print("--- Data February ---")
27 print(data_feb)
28
29 # Merge data sets (add columns)
30 merged_data <- merge(data_jan, data_feb, by = c("ID", "Name"))
31 print("--- Merged Data ---")
32 print(merged_data)
33
34 # Append data sets (add rows)
35 final_list <- bind_rows(data_jan, data_new_hires)
36 print("--- Appended Data ---")
37 print(final_list)
38
39
```

```
R - 4.5.2 - ~
> # Practical 6: Merge and Append
warning message:
In file.copy(savedcopy, lib, recursive = TRUE) :
  problem copying c:/users/itlab/appdata/local/r/win-library/4.5\00LOCK\dplyr\x64\dplyr.dll to c:/users/itlab/appdata/local/r/win-library/4.5\dplyr\x64\dplyr.dll: Permission denied

> install.packages("dplyr")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of R tools before proceeding:
https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'c:/users/itlab/appdata/local/r/win-library/4.5'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/dplyr_1.1.4.zip'
content type 'application/zip' length 1593482 bytes (1.5 MB)
downloaded 1.5 MB

package 'dplyr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  c:/users/itlab/appdata/local/temp/rtmpujkyma/downloaded_packages
> library(dplyr)

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':
  filter, lag

The following objects are masked from 'package:base':
  intersect, setdiff, setequal, union

> # Create data sets
> data_jan <- data.frame(
+   ID = c(1, 2, 3),
+   Name = c("Alice", "Bob", "Charlie"),
+   Jan_sales = c(100, 150, 200)
```

Name :- Priya Gupta
Roll no. :- S081

Sheth L.U.J. & Sir M.V. College

The screenshot shows the RStudio interface with the following code in the console:

```
> # Create data sets
> data_jan <- data.frame(
+   ID = c(1, 2, 3),
+   Name = c("Alice", "Bob", "Charlie"),
+   Jan_Sales = c(100, 150, 200)
+ )
>
> data_feb <- data.frame(
+   ID = c(1, 2, 3),
+   Name = c("Alice", "Bob", "Charlie"),
+   Feb_Sales = c(120, 160, 210)
+ )
>
> data_new_hires <- data.frame(
+   ID = c(4, 5),
+   Name = c("David", "Eva"),
+   Jan_Sales = c(50, 60)
+ )
>
> print("--- Data January ---")
[1] "--- Data January ---"
> print(data_jan)
#> ID   Name Jan_Sales
#> 1   Alice    100
#> 2   Bob      150
#> 3 Charlie    200
>
> print("--- Data February ---")
[1] "--- Data February ---"
> print(data_feb)
#> ID   Name Feb_Sales
#> 1   Alice    120
#> 2   Bob      160
#> 3 Charlie    210
>
> # Merge data sets (add columns)
> merged_data <- merge(data_jan, data_feb, by = c("ID", "Name"))
> print("--- Merged Data ---")
[1] "--- Merged Data ---"
> print(merged_data)
#> ID   Name Jan_Sales Feb_Sales
#> 1   Alice    100      120
#> 2   Bob      150      160
#> 3 Charlie    200      210
```

The Environment pane shows the following objects:

- data_feb: 3 obs. of 3 variables
- data_jan: 3 obs. of 3 variables
- data_new_hires: 2 obs. of 3 variables
- final_list: 5 obs. of 3 variables
- merged_data: 3 obs. of 4 variables

The screenshot shows the RStudio interface with the following code in the console:

```
> # Create data sets
> data_jan <- data.frame(
+   ID = c(1, 2, 3),
+   Name = c("Alice", "Bob", "Charlie"),
+   Jan_Sales = c(100, 150, 200)
+ )
>
> print("--- Data January ---")
[1] "--- Data January ---"
> print(data_jan)
#> ID   Name Jan_Sales
#> 1   Alice    100
#> 2   Bob      150
#> 3 Charlie    200
>
> print("--- Data February ---")
[1] "--- Data February ---"
> print(data_feb)
#> ID   Name Feb_Sales
#> 1   Alice    120
#> 2   Bob      160
#> 3 Charlie    210
>
> # Merge data sets (add columns)
> merged_data <- merge(data_jan, data_feb, by = c("ID", "Name"))
> print("--- Merged Data ---")
[1] "--- Merged Data ---"
> print(merged_data)
#> ID   Name Jan_Sales Feb_Sales
#> 1   Alice    100      120
#> 2   Bob      150      160
#> 3 Charlie    200      210
>
> # Append data sets (add rows)
> final_list <- bind_rows(data_jan, data_new_hires)
> print("--- Appended Data ---")
[1] "--- Appended Data ---"
> print(final_list)
#> ID   Name Jan_Sales
#> 1   Alice    100
#> 2   Bob      150
#> 3 Charlie    200
#> 4   David     50
#> 5   Eva      60
```

The Environment pane shows the following objects:

- data_feb: 3 obs. of 3 variables
- data_jan: 3 obs. of 3 variables
- data_new_hires: 2 obs. of 3 variables
- final_list: 5 obs. of 3 variables
- merged_data: 3 obs. of 4 variables

With Dataset

Name :- Priya Gupta
Roll no. :- S081

Sheth L.U.J. & Sir M.V. College

RStudio interface showing R code for merging datasets. The code reads two CSV files, prints their structures and columns, finds common columns, and performs four types of merges (Inner, Left, Right, Full Outer) before printing previews of each result.

```
41 # PRACT-6 : MERGE + APPEND on your Data sets with Data sets
42
43
44 # 1. LOAD BOTH DATASETS
45 library(dplyr)
46
47 # -----#
48 # 2. FIND COMMON COLUMNS
49 data1 <- read.csv("C:/users/itlab/onedrive/desktop/5081_R_Studio/study_performance.csv")
50 data2 <- read.csv("C:/users/itlab/onedrive/desktop\\5081_R_Studio\\study_performance_predition.csv")
51
52 cat("\n===== DATASET 1 STRUCTURE =====\n")
53 str(data1)
54
55 cat("\n===== DATASET 2 STRUCTURE =====\n")
56 str(data2)
57
58 cat("\n===== DATASET 1 COLUMNS =====\n")
59 print(names(data1))
60
61 cat("\n===== DATASET 2 COLUMNS =====\n")
62 print(names(data2))
63
64
65
66
67 # -----#
68 # 3. DIFFERENT TYPES OF MERGE
69 # -----
70 common_cols <- intersect(names(data1), names(data2))
71
72 cat("\n===== COMMON COLUMNS =====\n")
73 print(common_cols)
74
75
76 # -----#
77 # 4. APPEND (STACK ROWS)
78 # -----
79
80 # 1 INNER JOIN (only matching rows)
81 inner_merge <- merge(data1, data2, by = common_cols)
82
83 # 2 LEFT JOIN (all rows from data1)
84 left_merge <- merge(data1, data2, by = common_cols, all.x = TRUE)
85
86 # 3 RIGHT JOIN (all rows from data2)
87 right_merge <- merge(data1, data2, by = common_cols, all.y = TRUE)
88
89 # 4 FULL OUTER JOIN (all rows from both datasets)
90 full_merge <- merge(data1, data2, by = common_cols, all = TRUE)
91
92
93 # -----#
94 # 5. PRINT PREVIEWS
95 # -----
96 appended_data <- bind_rows(data1, data2)
97
98
99 # -----#
100 # 5. PRINT PREVIEWS
101
102 cat("\n===== INNER MERGE PREVIEW =====\n")
103 print(head(inner_merge))
104
105 cat("\n===== LEFT MERGE PREVIEW =====\n")
106 print(head(left_merge))
107
108 cat("\n===== RIGHT MERGE PREVIEW =====\n")
109 print(head(right_merge))
110
111 cat("\n===== FULL MERGE PREVIEW =====\n")
112 print(head(full_merge))
113
114 cat("\n===== APPENDED DATA PREVIEW =====\n")
115 print(head(appended_data))
116
117
```

RStudio interface showing R code for merging datasets with preview output. This version of the script includes the preview command after each merge type to show the resulting data frames.

```
77 # 3. DIFFERENT TYPES OF MERGE
78 #
79
80 # 1 INNER JOIN (only matching rows)
81 inner_merge <- merge(data1, data2, by = common_cols)
82
83 # 2 LEFT JOIN (all rows from data1)
84 left_merge <- merge(data1, data2, by = common_cols, all.x = TRUE)
85
86 # 3 RIGHT JOIN (all rows from data2)
87 right_merge <- merge(data1, data2, by = common_cols, all.y = TRUE)
88
89 # 4 FULL OUTER JOIN (all rows from both datasets)
90 full_merge <- merge(data1, data2, by = common_cols, all = TRUE)
91
92
93 # -----#
94 # 4. APPEND (STACK ROWS)
95 # -----
96 appended_data <- bind_rows(data1, data2)
97
98
99 # -----#
100 # 5. PRINT PREVIEWS
101
102 cat("\n===== INNER MERGE PREVIEW =====\n")
103 print(head(inner_merge))
104
105 cat("\n===== LEFT MERGE PREVIEW =====\n")
106 print(head(left_merge))
107
108 cat("\n===== RIGHT MERGE PREVIEW =====\n")
109 print(head(right_merge))
110
111 cat("\n===== FULL MERGE PREVIEW =====\n")
112 print(head(full_merge))
113
114 cat("\n===== APPENDED DATA PREVIEW =====\n")
115 print(head(appended_data))
116
117
```

Name :- Priya Gupta
Roll no. :- S081

Sheth L.U.J. & Sir M.V. College

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Console Background Jobs
R 4.5.2 - ~/
> # ===== with uploaded Data sets=====
>
> # PRACT-6 : MERGE + APPEND on Your Data sets with Data sets
>
> library(dplyr)
>
> # -----
> # 1. LOAD BOTH DATASETS
> #-----
> data1 <- read.csv("C:/users/itlab/onedrive/Desktop/S081_R_Studio/study_performance.csv")
> data2 <- read.csv("C:/users/itlab/onedrive/Desktop/S081_R_Studio/study_performance_prediton.csv")
>
> cat("\n===== DATASET 1 STRUCTURE =====\n")
===== DATASET 1 STRUCTURE =====
> str(data1)
'data.frame': 1000 obs. of 8 variables:
 \$ gender : chr "Female" "Female" "Female" "Male" ...
 \$ race.ethnicity : chr "group A" "group C" "group C" "group A" ...
 \$ parental.level.of.education: chr "Bachelor's degree" "Some college" "Master's degree" "Associate's degree" ...
 \$ lunch : chr "standard" "standard" "standard" "free/reduced" ...
 \$ test.preparation.course : chr "none" "completed" "none" "none" ...
 \$ math.score : int 72 69 90 47 76 71 88 40 64 38 ...
 \$ reading.score : int 72 70 95 57 78 83 95 43 64 60 ...
 \$ writing.score : int 74 88 93 44 75 78 92 39 67 50 ...
>
> cat("\n===== DATASET 2 STRUCTURE =====\n")
===== DATASET 2 STRUCTURE =====
> str(data2)
'data.frame': 1000 obs. of 8 variables:
 \$ gender : chr "Female" "Male" "Female" "Male" ...
 \$ race.ethnicity : chr "group D" "group D" "group B" "group B" ...
 \$ parental.level.of.education: chr "Some college" "Associate's degree" "Some college" "Some college" ...
 \$ lunch : chr "standard" "standard" "free/reduced" "free/reduced" ...
 \$ test.preparation.course : chr "none" "completed" "none" "none" ...
 \$ math.score : int 59 96 57 70 83 68 82 46 80 57 ...
 \$ reading.score : int 70 93 76 70 85 57 83 61 75 69 ...
 \$ writing.score : int 78 87 77 63 86 54 80 58 73 77 ...
>
> cat("\n===== DATASET 1 COLUMNS =====\n")
===== DATASET 1 COLUMNS =====
26°C Sunny 11:34:59 AM 01-12-2025

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Console Background Jobs
R 4.5.2 - ~/
> math.score : int 72 69 90 47 76 71 88 40 64 38 ...
\$ reading.score : int 70 93 76 70 85 57 83 61 75 69 ...
\$ writing.score : int 78 87 77 63 86 54 80 58 73 77 ...
>
> cat("\n===== DATASET 1 COLUMNS =====\n")
===== DATASET 1 COLUMNS =====
> print(names(data1))
[1] "gender" "race.ethnicity" "parental.level.of.education" "lunch" "test.preparation.course"
[6] "math.score" "reading.score" "writing.score"
>
> cat("\n===== DATASET 2 COLUMNS =====\n")
===== DATASET 2 COLUMNS =====
> print(names(data2))
[1] "gender" "race.ethnicity" "parental.level.of.education" "lunch" "test.preparation.course"
[6] "math.score" "reading.score" "writing.score"
>
> # -----
> # 2. FIND COMMON COLUMNS
> #-----
> common_cols <- intersect(names(data1), names(data2))
>
> cat("\n===== COMMON COLUMNS =====\n")
===== COMMON COLUMNS =====
> print(common_cols)
[1] "gender" "lunch"
>
> # -----
> # 3. DIFFERENT TYPES OF MERGE
> #-----
>
> # 1 INNER JOIN (only matching rows)
> inner_merge <- merge(data1, data2, by = common_cols)
>
> # 2 LEFT JOIN (all rows from data1)
> left_merge <- merge(data1, data2, by = common_cols, all.x = TRUE)
>
> # 3 RIGHT JOIN (all rows from data2)
> right_merge <- merge(data1, data2, by = common_cols, all.y = TRUE)
26°C Sunny 11:35:17 AM 01-12-2025

Name :- Priya Gupta
Roll no. :- S081

Sheth L.U.J. & Sir M.V. College

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source

R # R4.5.2: ~

```
> # 4. FULL OUTER JOIN (all rows from both datasets)
> full_merge <- merge(data1, data2, by = common_cols, all = TRUE)
>
>
> #
> # 4. APPEND (STACK ROWS)
> #
> appended_data <- bind_rows(data1, data2)
>
>
> #
> # 5. PRINT PREVIEWS
> #
>
> cat("\n----- INNER MERGE PREVIEW -----")
----- INNER MERGE PREVIEW -----
> print(head(inner_merge))
gender      lunch race_ethnicity parental_level_of_education test_preparation_course math_score reading_score writing_score race.ethnicity
1 female free/reduced group A          high school completed        77       88       85       85       group C
2 female free/reduced group A          high school completed        77       88       85       85       group B
3 female free/reduced group A          high school completed        77       88       85       85       group C
4 female free/reduced group A          high school completed        77       88       85       85       group D
5 female free/reduced group A          high school completed        77       88       85       85       group B
6 female free/reduced group A          high school completed        77       88       85       85       group C
parental.level.of.education test.preparation.course math.score reading.score writing.score
1 bachelor's degree           none            69       78       78
2          high school completed        69       78       75
3          some high school completed        52       59       58
4          master's degree   none            93       98       94
5          high school      none            39       60       52
6          high school      none            76       90       84
>
> cat("\n----- LEFT MERGE PREVIEW -----")
----- LEFT MERGE PREVIEW -----
> print(head(left_merge))
gender      lunch race_ethnicity parental_level_of_education test_preparation_course math_score reading_score writing_score race.ethnicity
1 female free/reduced group A          high school completed        77       88       85       85       group C
2 female free/reduced group A          high school completed        77       88       85       85       group B
3 female free/reduced group A          high school completed        77       88       85       85       group C
4 female free/reduced group A          high school completed        77       88       85       85       group D
5 female free/reduced group A          high school completed        77       88       85       85       group B
Project (None)
```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help

Source

R - R4.5.2 - ~\

```
1 female free/reduced group A high school completed 77 88 85 group C
2 female free/reduced group A high school completed 77 88 85 group B
3 female free/reduced group A high school completed 77 88 85 group C
4 female free/reduced group A high school completed 77 88 85 group D
5 female free/reduced group A high school completed 77 88 85 group B
6 female free/reduced group A high school completed 77 88 85 group C
parental.level.of.education test.preparation.course math.score reading.score writing.score
1 bachelor's degree none 69 78 78
2 high school completed 69 78 75
3 some high school completed 52 59 58
4 master's degree none 93 98 94
5 high school none 39 60 52
6 high school none 76 90 84
> cat("\n===== RIGHT MERGE PREVIEW =====\n")
===== RIGHT MERGE PREVIEW =====
> print(head(right_merge))
gender lunch race.ethnicity parental.level.of.education test.preparation.course math.score reading.score writing.score race.ethnicity
1 female free/reduced group A high school completed 77 88 85 group C
2 female free/reduced group A high school completed 77 88 85 group B
3 female free/reduced group A high school completed 77 88 85 group C
4 female free/reduced group A high school completed 77 88 85 group D
5 female free/reduced group A high school completed 77 88 85 group B
6 female free/reduced group A high school completed 77 88 85 group C
parental.level.of.education test.preparation.course math.score reading.score writing.score
1 bachelor's degree none 69 78 78
2 high school completed 69 78 75
3 some high school completed 52 59 58
4 master's degree none 93 98 94
5 high school none 39 60 52
6 high school none 76 90 84
> cat("\n===== FULL MERGE PREVIEW =====\n")
===== FULL MERGE PREVIEW =====
> print(head(full_merge))
gender lunch race.ethnicity parental.level.of.education test.preparation.course math.score reading.score writing.score race.ethnicity
1 female free/reduced group A high school completed 77 88 85 group C
2 female free/reduced group A high school completed 77 88 85 group B
3 female free/reduced group A high school completed 77 88 85 group C
4 female free/reduced group A high school completed 77 88 85 group D
5 female free/reduced group A high school completed 77 88 85 group E
6 female free/reduced group A high school completed 77 88 85 group F
7 female free/reduced group A high school completed 77 88 85 group G
8 female free/reduced group A high school completed 77 88 85 group H
9 female free/reduced group A high school completed 77 88 85 group I
10 female free/reduced group A high school completed 77 88 85 group J
```

Name :- Priya Gupta
Roll no. :- S081

Sheth L.U.J. & Sir M.V. College

The screenshot shows the RStudio interface with the following details:

- Console:** Displays R code and its output. The code includes `cat` statements for "FULL MERGE PREVIEW" and "APPENDED DATA PREVIEW". The previewed data frames show columns such as gender, race, ethnicity, parental level of education, lunch, test preparation course, math score, reading score, writing score, and race.ethnicity.
- Environment:** Shows a list of objects in the current session, including `app_`, `dat_`, `fin_`, `ful_`, `inn_`, `lef_`, `mer_`, `rig_`, `stu_`, and `values`.
- Plots:** No plots are visible.
- Session:** Shows the R version (R 4.5.2), memory usage (1.74 GB), and the date and time (11:35:53 AM 01/12/2025).

Name :- Priya Gupta
Roll no. :- S081