

SHETH L.U.J. AND SIR M.V. COLLEGE

SUBJECT: Data Analysis with R

Aim: Combining datasets vertically (concatenation) using rbind() (R).

```
R - R452 - ~/
> # 1.1 Load the Student Performance dataset
> students_df <- read.csv("Student.csv")
> # 1.2 Load the Heart Disease dataset
> heart_df <- read.csv("heart_disease_ucl.csv")
> print("---- Data Structure Before Transformation ----")
[1] "---- Data Structure Before Transformation ----"
> print(names(students_df)) # StudentID, Age, Gender, GPA, etc.
[1] "StudentID" "Age" "Gender" "Ethnicity" "ParentalEducation" "StudyTimeWeekly"
[7] "Absences" "tutoring" "ParentalSupport" "Extracurricular" "Sports" "Music"
[13] "Volunteering" "GPA" "GradeClass"
> print(names(heart_df)) # id, age, sex, chol, etc.
[1] "id" "age" "sex" "dataset" "cp" "trestbps" "chol" "fbs" "restecg" "thalch" "exang" "oldpeak"
[13] "slope" "ca" "num"
> #-----
> # 2.1 Prepare Student Data
> # Use 'GPA' as the numeric value
> #-----
> students_clean <- data.frame(
+ dataset = "Student",
+ value = students_df$GPA # GPA column
+ )
> #-----
> # 2.2 Prepare Heart Disease Data
> # Use 'chol' as the numeric value
> #-----
> heart_clean <- data.frame(
+ dataset = "Heart",
+ value = heart_df$chol # cholesterol column
+ )
> # Ensure both 'value' columns are numeric (Good practice)
> students_clean$value <- as.numeric(students_clean$value)
> heart_clean$value <- as.numeric(heart_clean$value)
> combined_data <- rbind(students_clean, heart_clean)
> print("---- Combined Data Summary ----")
[1] "---- Combined Data Summary ----"
> print(paste("Student rows:", nrow(students_clean)))
[1] "Student rows: 2392"
> print(paste("Heart rows:", nrow(heart_clean)))
[1] "Heart rows: 920"
> print(paste("Total rows (Expected):",
+ nrow(students_clean) + nrow(heart_clean)))
[1] "Total rows (Expected): 3312"
> print(paste("Total rows (Actual):", nrow(combined_data)))
[1] "Total rows (Actual): 3312"

+ value = heart_df$chol # cholesterol column
+ )
> # Ensure both 'value' columns are numeric (Good practice)
> students_clean$value <- as.numeric(students_clean$value)
> heart_clean$value <- as.numeric(heart_clean$value)
> combined_data <- rbind(students_clean, heart_clean)
> print("---- Combined Data Summary ----")
[1] "---- Combined Data Summary ----"
> print(paste("Student rows:", nrow(students_clean)))
[1] "Student rows: 2392"
> print(paste("Heart rows:", nrow(heart_clean)))
[1] "Heart rows: 920"
> print(paste("Total rows (Expected):",
+ nrow(students_clean) + nrow(heart_clean)))
[1] "Total rows (Expected): 3312"
> print(paste("Total rows (Actual):", nrow(combined_data)))
[1] "Total rows (Actual): 3312"
> print("---- Preview of Combined Data (Top and Bottom) ----")
[1] "---- Preview of Combined Data (Top and Bottom) ----"
> print(head(combined_data)) # Shows Student GPA values
  dataset value
1 Student 2.9291956
2 Student 3.0429148
3 Student 0.1126023
4 Student 2.0542181
5 Student 1.2880612
6 Student 3.0841836
> print(tail(combined_data)) # Shows Heart chol values
  dataset value
3307 Heart 310
3308 Heart 333
3309 Heart 139
3310 Heart 223
3311 Heart 385
3312 Heart 254
> view(combined_data)
> |
```

SHETH L.U.J. AND SIR M.V. COLLEGE

SUBJECT: Data Analysis with R

The screenshot displays the RStudio interface with two data frames loaded: `heart_df` and `students_df`.

heart_df Data:

id	age	sex	dataset	cp	trestbps	chol	fbs	restecg	thaich	exang	oldpeak	slope	ca	thal	num	
1	1	63	Male	Cleveland	typical angina	145	233	TRUE	lv hypertrophy	150	FALSE	2.3	downsloping	0	fixed defect	0
2	2	67	Male	Cleveland	asymptomatic	160	286	FALSE	lv hypertrophy	108	TRUE	1.5	flat	3	normal	2
3	3	67	Male	Cleveland	asymptomatic	120	229	FALSE	lv hypertrophy	129	TRUE	2.6	flat	2	reversible defect	1
4	4	37	Male	Cleveland	non-anginal	130	250	FALSE	normal	187	FALSE	3.5	downsloping	0	normal	0
5	5	41	Female	Cleveland	atypical angina	130	204	FALSE	lv hypertrophy	172	FALSE	1.4	upslloping	0	normal	0
6	6	56	Male	Cleveland	atypical angina	120	236	FALSE	normal	178	FALSE	0.8	upslloping	0	normal	0
7	7	62	Female	Cleveland	asymptomatic	140	268	FALSE	lv hypertrophy	160	FALSE	3.6	downsloping	2	normal	3
8	8	57	Female	Cleveland	asymptomatic	120	354	FALSE	normal	163	TRUE	0.6	upslloping	0	normal	0
9	9	63	Male	Cleveland	asymptomatic	130	254	FALSE	lv hypertrophy	147	FALSE	1.4	flat	1	reversible defect	2
10	10	53	Male	Cleveland	asymptomatic	140	203	TRUE	lv hypertrophy	155	TRUE	3.1	downsloping	0	reversible defect	1
11	11	57	Male	Cleveland	asymptomatic	140	192	FALSE	normal	148	FALSE	0.4	flat	0	fixed defect	0
12	12	56	Female	Cleveland	atypical angina	140	294	FALSE	lv hypertrophy	153	FALSE	1.3	flat	0	normal	0
13	13	56	Male	Cleveland	non-anginal	130	256	TRUE	lv hypertrophy	142	TRUE	0.6	flat	1	fixed defect	2
14	14	44	Male	Cleveland	atypical angina	120	263	FALSE	normal	173	FALSE	0.0	upslloping	0	reversible defect	0
15	15	52	Male	Cleveland	non-anginal	172	199	TRUE	normal	162	FALSE	0.5	upslloping	0	reversible defect	0
16	16	57	Male	Cleveland	non-anginal	150	168	FALSE	normal	174	FALSE	1.6	upslloping	0	normal	0
17	17	48	Male	Cleveland	atypical angina	110	229	FALSE	normal	168	FALSE	1.0	downsloping	0	reversible defect	1
18	18	54	Male	Cleveland	asymptomatic	140	239	FALSE	normal	160	FALSE	1.2	upslloping	0	normal	0
19	19	48	Female	Cleveland	non-anginal	130	275	FALSE	normal	139	FALSE	0.2	upslloping	0	normal	0
20	20	49	Male	Cleveland	atypical angina	130	266	FALSE	normal	171	FALSE	0.6	upslloping	0	normal	0
21	21	64	Male	Cleveland	typical angina	110	211	FALSE	lv hypertrophy	144	TRUE	1.8	flat	0	normal	0
22	22	58	Female	Cleveland	typical angina	150	283	TRUE	lv hypertrophy	162	TRUE	1.0	upslloping	0	normal	0
23	23	58	Male	Cleveland	atypical angina	120	284	FALSE	lv hypertrophy	160	FALSE	1.8	flat	0	normal	1
24	24	58	Male	Cleveland	non-anginal	132	224	FALSE	lv hypertrophy	173	FALSE	3.2	upslloping	2	reversible defect	3
25	25	60	Male	Cleveland	asymptomatic	130	206	FALSE	lv hypertrophy	132	TRUE	2.4	flat	2	reversible defect	4
26	26	50	Female	Cleveland	non-anginal	120	219	FALSE	normal	158	FALSE	1.6	flat	0	normal	0

students_df Data:

StudentID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Extracurricular	Sports	Music	Volunteering	GPA	GradeClass
1	1001	17	1	0	2	19.833722808	7	1	2	0	0	0	2.92919559	2
2	1002	18	0	0	1	15.408756056	0	0	1	0	0	0	3.04291483	1
3	1003	15	0	2	3	4.210569769	26	0	2	0	0	0	0.11260225	4
4	1004	17	1	0	3	10.028029474	14	0	3	1	0	0	2.05421814	3
5	1005	17	1	0	2	4.672495273	17	1	3	0	0	0	1.28806118	4
6	1006	18	0	0	1	8.191218545	0	0	1	1	0	0	3.06418361	1
7	1007	15	0	1	1	15.601680475	10	0	3	0	1	0	2.74832741	2
8	1008	15	1	1	4	15.424496306	22	1	1	1	0	0	1.36014271	4
9	1009	17	0	0	0	4.562007558	1	0	2	0	1	0	2.89681919	2
10	1010	16	1	0	1	16.444466363	0	0	3	1	0	0	3.57347421	0
11	1011	17	0	0	1	11.851363655	11	0	1	0	0	0	2.14717163	3
12	1012	17	0	0	1	7.598485819	15	0	2	0	0	0	1.65959452	4
13	1013	17	0	1	1	10.038711616	21	0	3	1	0	0	1.52007781	4
14	1014	17	0	1	2	12.101425069	21	0	4	0	1	0	1.75158096	4
15	1015	18	1	0	1	11.197810637	9	1	2	0	0	0	2.39678812	3
16	1016	15	0	0	2	9.728100711	17	1	0	0	1	0	1.34152072	4
17	1017	18	0	3	1	10.098656082	14	0	2	1	1	0	2.23217528	3
18	1018	18	1	0	0	3.528338209	16	1	2	0	0	0	1.38440418	4
19	1019	18	0	1	3	16.254658086	29	0	2	1	0	0	0.46955332	4
20	1020	17	0	0	1	10.835206399	9	0	2	0	0	1	2.39578409	3
21	1021	16	1	0	3	2.621597234	2	0	3	0	0	0	1.277841130	2
22	1022	15	0	0	2	15.323142032	25	0	1	1	0	0	0.34689404	4
23	1023	16	1	1	0	18.648879568	29	1	1	0	0	0	0.31254623	4
24	1024	18	1	3	4	18.946137985	20	0	2	1	0	0	1.77013188	4
25	1025	18	1	0	1	7.380354648	15	0	2	0	0	0	1.50515562	4
26	1026	16	1	0	3	2.710337471	5	0	4	0	0	1	2.97785192	2

SHETH L.U.J. AND SIR M.V. COLLEGE

SUBJECT: Data Analysis with R

The top screenshot shows the RStudio interface with a dataset of student IDs. The console displays the following data:

Dataset	Value
1 Student	2.92919559
2 Student	3.04291483
3 Student	0.11260225
4 Student	2.05421814
5 Student	1.28806118
6 Student	3.08418361
7 Student	2.74823741
8 Student	1.36014271
9 Student	2.89681919
10 Student	3.57374241
11 Student	2.14717163
12 Student	1.55959452
13 Student	1.52007781
14 Student	1.75158096
15 Student	2.39678812
16 Student	1.34152072
17 Student	2.23217528
18 Student	1.38440418
19 Student	0.46955332
20 Student	2.39578409
21 Student	2.77841130
22 Student	0.34689404
23 Student	0.31254623
24 Student	1.77013188
25 Student	1.50515562
26 Student	2.97785192

The bottom screenshot shows the RStudio interface with a dataset of heart rates. The console displays the following data:

Dataset	Value
1 Heart	233
2 Heart	286
3 Heart	229
4 Heart	250
5 Heart	204
6 Heart	236
7 Heart	268
8 Heart	354
9 Heart	254
10 Heart	203
11 Heart	192
12 Heart	294
13 Heart	256
14 Heart	263
15 Heart	199
16 Heart	168
17 Heart	229
18 Heart	239
19 Heart	275
20 Heart	266
21 Heart	211
22 Heart	283
23 Heart	284
24 Heart	224
25 Heart	206
26 Heart	219

SHETH L.U.J. AND SIR M.V. COLLEGE

SUBJECT: Data Analysis with R

The screenshot displays the R Studio environment. The main window shows a data frame with 27 rows and 2 columns. The first column is labeled 'Student' and the second is labeled 'Value'. The data is as follows:

Student	Value
1	2.92919559
2	3.04291483
3	0.11260225
4	2.05421814
5	1.28806118
6	3.08418361
7	2.74823741
8	1.36014271
9	2.89681919
10	3.57347421
11	2.14717163
12	1.55999452
13	1.52007781
14	1.75158096
15	2.39678812
16	1.34152072
17	2.23217528
18	1.38440418
19	0.46955332
20	2.39578409
21	2.77841130
22	0.34689404
23	0.31254623
24	1.77013188
25	1.50515562
26	2.97785192

The Environment pane on the right shows the following data objects:

Object	Size
combined_data	3312 obs. of 2 variables
heart_clean	920 obs. of 2 variables
heart_df	920 obs. of 16 variables
students_clean	2392 obs. of 2 variables
students_df	2392 obs. of 15 variables

The Files pane on the right shows the following files:

File	Size	Modified
loan_approval.csv	113.1 KB	Nov 24, 2025, 12:06 PM
vg-sales.csv	1.3 MB	Nov 13, 2025, 8:49 AM
Student.csv	163 KB	Dec 8, 2025, 10:48 AM

The console at the bottom shows the following output:

```
Showing 1 to 27 of 3,312 entries, 2 total columns
```