

**SETH L.U.J and SIR M.V COLLEGE
PRAC: 11 12 13 14 15**

AIM:- 11 Reshaping data using pivot_longer()/pivot_wider() (R).

PRANAV CHAUGULE

S075

**SETH L.U.J and SIR M.V COLLEGE
PRAC: 11 12 13 14 15**

The screenshot shows the RStudio interface with several panes:

- Code Editor:** Displays R code for data manipulation, including pivot operations and wide format conversion.
- Console:** Shows the output of the executed R code, including printed data frames and their structures.
- Environment:** Shows the global environment with objects like `category_pivot`, `df`, `long_df`, `student_marks`, and `wide_df`.
- Files:** A file browser showing various files and folders in the current directory, including JData, Rhistory, and several DS Prac files.

The screenshot shows an RStudio interface with the following details:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Go to file/function, Addins.
- Environment Tab:** Shows the global environment with objects like category_pivot, df, long_df, student_marks, and wide_df.
- Data Tab:** Shows data frames: category_pivot (100 obs. of 7 variables), df (100 obs. of 4 variables), long_df (300 obs. of 3 variables), student_marks (100 obs. of 3 variables), and wide_df (100 obs. of 4 variables).
- Console Tab:** Displays R code and its output. The code includes pivoting operations and advanced examples involving multiple columns and categories.
- File Explorer:** Shows the project structure with files like .JNData, .Rhistory, 6th CN 5075 PRAC 6.docx, 6th CN 5091 PRAC 6.docx, bats_information.csv, Cleanned_Student_Mental_Health.csv, CNpkt, Custom Office Templates, desktop.ini, DS Prac 1.py, DS 5075, from_pypy.py, gayatriplib, gayatriplibx, GIS DataBase, IISExpress, My Data Sources, and My Web Sites.

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Untitled1 Student_Marks
Source on Save Run Source Environment History Connections Tutorial
Project: None
19
20 print("---- 2. Long Format (pivot_longer) ---")
21 print(head(long_df, 9)) # show first 9 rows (3 metrics per student)
22
23 # 3. PIVOT_WIDER: Spread back to wide format
24 wide_df <- long_df %>%
25 pivot_wider(
26   names_from = Metric,
27   values_from = Value
28 )
29
30
34:28 [Top Level] R Script
Console Terminal Background Jobs
[R - R4.52 - ~]
> print("---- 3. wide Format (Back to original) ---")
[1] "---- 3. wide Format (Back to Original) ---"
> print(head(wide_df))
# A tibble: 6 x 4
  StudentID number_courses time_study Marks
    <dbl>           <dbl>      <dbl>    <dbl>
1        1             4     4.51    12.0
2        2             4     4.096   7.73 
3        3             4     3.13    13.8 
4        4             6     7.91    53.0 
5        5             8     7.81    55.3 
6        6             6     3.21    17.8 
> # 4. ADVANCED EXAMPLE! Pivot by number_courses to see Marks distribution
> category_pivot <- df %>%
+ select(StudentID, number_courses, Marks) %>%
+ pivot_wider(
+   names_from = number_courses,
+   values_from = Marks
+ )
>
> print("---- 4. Course Pivot (Spreading number_courses) ---")
[1] "---- 4. course Pivot (Spreading number_courses) ---"
> print(head(category_pivot))
# A tibble: 6 x 7
  StudentID `3` `4` `6` `8` `5` `7`
    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1        1  19.2 NA   NA   NA   NA
2        2 NA   NA   NA   NA   NA
```

PRANAV CHAUGULE

S075

**SETH L.U.J and SIR M.V COLLEGE
PRAC: 11 12 13 14 15**

The screenshot shows the RStudio interface with several panes:

- File Explorer:** Shows a tree view of files and folders, including JRDns, .Rhistory, 6th CN 5075 PRAC 6.docx, 6th CN 5091 PRAC 6.docx, bats_information.csv, Cleanned_Student_Mental_Health.csv, CN.pkt, Custom Office Templates, desktop.ini, DS Prac 1.py, DS 5075, flytrap.py, gaytrappix, GIS Database, IISExpress, My Data Sources, and My Web Sites.
- Code Editor:** Displays R script code for pivoting data frames. It includes sections for printing long format, spreading wide format, and pivoting by number_courses. The code uses dplyr and tidyverse packages.
- Console:** Shows the output of the R script, including the printed data frames and the resulting pivot operations.
- Environment:** Shows the global environment with objects like category_pivot, df, long_df, student_marks, and wide_df.
- Plots:** No plots are currently displayed.
- Packages:** No packages are currently displayed.
- Help:** No help pages are currently displayed.
- Viewer:** No files are currently displayed.
- Presentation:** No presentations are currently displayed.

12 Combining datasets vertically (concatenation) using rbind() (R).

The screenshot shows the RStudio interface with the following details:

- Code Editor:** The left pane displays R code for data cleaning and combining datasets. It includes creating `bats_clean` and `student_clean` data frames, and then merging them into `combined_data`.
- Console:** The bottom-left pane shows the output of the executed code, including the number of rows for each dataset and the total combined rows.
- Environment:** The top-right pane lists the objects in the global environment, such as `bats_clean` (30000 obs., 11 variables), `bats_df` (30000 obs., 11 variables), and `category_pivot` (100 obs., 7 variables).
- File Explorer:** The bottom-right pane shows the file structure, including CSV files like `bats_information.csv` and `Student_Marks.csv`, and various R and Python scripts.

AIM:- 13 Identifying and handling duplicates using distinct() (R).

OUTPUT:-

PRANAV CHAUGULE

S075

**SETH L.U.J and SIR M.V COLLEGE
PRAC: 11 12 13 14 15**

The screenshot shows the RStudio interface with the following details:

- Source** tab: Contains R code for reading a CSV file and creating a clean data frame.
- Console** tab: Shows the output of the R code, including a warning message about coercion.
- Environment** tab: Lists various objects in the global environment, such as `bats_df`, `combined_data`, and `df`.
- Files** tab: Shows the project structure with files like `RData`, `.Rhistory`, and `bats_information.csv`.

The screenshot shows the RStudio interface with the following details:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Source Editor:** Displays R code for reading a CSV file and creating a data frame. The code includes filtering for rows where n > 1 to identify duplicate entries.
- Environment Tab:** Shows the global environment with objects like `combined_data`, `duplicates_report`, and `student_clean`.
- Files Tab:** Lists files in the current directory, including RData, history files, and various CSV and Python files.

**SETH L.U.J and SIR M.V COLLEGE
PRAC: 11 12 13 14 15**

The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help, and Addins. The bottom status bar shows the date and time as 12:30 08-12-2025. The main area has tabs for Source, Environment, History, Connections, and Tutorial. The Source tab displays R code for cleaning a dataset of bat observations, removing duplicates, and printing a confirmation message. The Environment tab lists various objects in the global environment, such as 'cleaned_exacL' (18 obs., 2 variables), 'combined_data' (30100 obs., 2 variables), and 'Student_Marks' (100 obs., 3 variables). The Files tab shows a file tree with files like 'RData', '.Rhistory', '6th CN 5075 PRAC 6.docx', '6th CN 5091 PRAC 6.docx', 'bats_information.csv', 'Cleaned_Student_Mental_Health.csv', 'CN.pptx', 'Custom Office Templates', 'desktop.ini', 'DS Prac 1.py', 'DS S075', 'from pympput.py', 'gayathri.bix', 'GIS DataBase', and 'IISExpress'. The bottom taskbar includes icons for file operations, search, and system status.

The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help, and Addins. The left sidebar has Source, Console, Terminal, and Background jobs tabs. The main area displays the following R code and its output:

```
R > 4.5.2 .-/
8   yangochiroptera (Microbat); Example: Horseshoe Bat NA
9   Yinpterochiroptera (Megabat); Example: Little Brown Bat NA
10  Yangochiroptera (Microbat); Example: Bumblebee Bat NA
11  Yinpterochiroptera (Megabat); Example: Big Brown Bat NA
12  Yangochiroptera (Microbat); Example: Fringed Myotis NA
13  Yangochiroptera (Microbat); Example: Spectral Bat NA
14  Yinpterochiroptera (Megabat); Example: Bumblebee Bat NA
15  Yinpterochiroptera (Megabat); Example: Spectral Bat NA
16  Yangochiroptera (Microbat); Example: Egyptian Fruit Bat NA
17  Yangochiroptera (Microbat); Example: Vampire Bat NA
18 Yinpterochiroptera (Megabat); Example: Egyptian Fruit Bat NA
>
> unique_names <- bats_clean %>
+ distinct(Name, .keep_all = TRUE)
> print("---- Unique Names Only ----")
[1] "---- Unique Names Only ----"
> print(unique_names)
```

The output lists 18 unique bat names. The right side of the interface shows the Environment pane with variables like `combined_data`, `df`, and `student_clean`, and the Files pane displaying a list of files in the current directory.

14 Extracting date components using lubridate:: functions (R).

OUTPUT:-

PRANAV CHAUGULE

S075

**SETH L.U.J and SIR M.V COLLEGE
PRAC: 11 12 13 14 15**

R version 4.5.2 (2025-10-31 ucrt) -- "[Not] Part in a Rumble"
Copyright (C) 2025 The R Foundation for statistical Computing
Platform: x86_64-w64-mingw32/x64

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/RData]

```
> install.packages("lubridate")
```

WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

<https://cran.rstudio.com/bin/windows/Rtools/>

Installing package into 'c:/users/mvlui/appdata/local/r/win-library/4.5'
(as 'lib' is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/lubridate_1.9.4.zip'
Content type 'application/zip' length 996067 bytes (972 kB)
downloaded 972 kB

package 'lubridate' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
c:/users/mvlui/appdata/local/temp/rtmpgmgdul/downloaded_packages

```
> install.packages("lubridate")
```

WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

Environment History Connections Tutorial
Project: (None)

Data

Name	Value
dates_df	4 obs. of 2 variables
processed_d	0 obs. of 11 variables
current_time	2025-12-08 12:29:29 IST

Files Plots Packages Help Viewer Presentation
Home Name Size Modified
RData 49 B Dec 8, 2025
Rhistory 2.6 KB Dec 8, 2025
Abd Analytics Custom Office Templates DATA ANALYSIS PRAC 6 5084.pdf 194.9 KB Dec 1, 2025
GIS Database Global_Health (1) (1).csv 196 B Dec 1, 2025
Global_Health (1).csv 547 B Dec 1, 2025
Global_Health.csv 8.7 MB Nov 29, 2025
Global_Health.xls 5.9 KB Dec 1, 2025
IISExpress My Web Sites Power BI Desktop R Virtual Machines Visual Studio 2022

15 Generating basic summaries using str() or summary() (R).

OUTPUT:-

**SETH L.U.J and SIR M.V COLLEGE
PRAC: 11 12 13 14 15**

```
RStudio
File Edit Code View Plot Session Build Debug Profile Tools Help
Go to RStudio ... -> Jupyter ...
Source
Console Terminal Background Jobs ...
(R = 452)
> +
> print("... meta loaded ...")
[1] "... meta loaded ..."
> print("... output or xrd() ...")
[1] "... output or xrd() ..."
> xrd(retail_uf)
'data.frame': 6 obs. of 5 variables:
 $ id: num 1 2 3 4 5 6
 $ category: chr "Electronics" "None" "Electronics" "Clothing" ...
 $ price: num 500 45.998 na 380 ...
 $ ISBLOCK: logi TRUE TRUE FALSE TRUE FALSE TRUE
 $ Rating: num 4.53 4.94 3.54 2
 > print("... output of summary() (before factor conversion) ...")
[1] "... output of summary() (before factor conversion) ..."
> summary(retail_uf)
   id      category      price      ISBLOCK       Rating
Min. : 1.000   Category: 0   Min. : 21.8   FALSE:12   Min. :4.11
1st Qu.: 2.250   class: character   1st Qu.: 41.8   FALSE:12   1st Qu.:13.998
Median : 3.500   mode: character   median:380.0   TRUE: 14   median:14.198
Mean   : 4.100   nullempty:1     mean:14.150   nullempty:1   mean:14.150
3rd Qu.: 4.750   3rd Qu.:380.0   3rd Qu.:14.221   3rd Qu.:14.221   3rd Qu.:14.221
Max. : 5.000   max: 380.0   Max. :14.998   Max. :14.998   Max. :14.998
> retail_UFcategory <- as.factor(summary$category)
> print("... output of summary() (after factor conversion) ...")
[1] "... output of summary() (before factor conversion) ..."
> summary(retail_UF)
   id      category      price      ISBLOCK       Rating
Min. : 1.000   Clothing:2   Min. : 21.8   FALSE:12   Min. :4.100
1st Qu.: 2.250   Electronics:2  1st Qu.: 41.8   FALSE:12   1st Qu.:13.998
Median : 3.500   None:1      Median:380.0   TRUE: 14   Median:14.198
Mean   : 4.100   mean:1        mean:14.150   nullempty:1   mean:14.150
3rd Qu.: 4.750   3rd Qu.:380.0   3rd Qu.:14.221   3rd Qu.:14.221   3rd Qu.:14.221
Max. : 5.000   max: 380.0   Max. :14.998   Max. :14.998   Max. :14.998
> avg_rating <- mean(summary$Rating)
> max_price <- max(retail_UF$price, na.rm = TRUE)
> print(paste0("average rating", avg_rating))
[1] "average rating: 4.15"
> print(paste0("highest price", max_price))
[1] "highest price: 380"
> 
```