

SHETH L.U.J. AND SIR M.V. COLLEGE
DATA ANALYSIS WITH SAS/SPSS/RR

PRACTICAL NO: 11

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

CODE:

```
1 library(dplyr)
2 library(tidy)
3
4 df <- read.csv("JEE Mains 2013-25 Top Ranks.csv", na.strings = c("", "NA")) %>%
5   mutate(StudentID = row_number()) %>%
6   select(StudentID, Category, Maths_Marks, Physics_Marks, Chemistry_Marks)
7
8 print("~~~ 1. Original Wide Data ~~")
9 print(head(df))
10
11 long_df <- df %>%
12   pivot_longer(
13     cols = c(Maths_Marks, Physics_Marks, Chemistry_Marks),
14     names_to = "Subject",
15     values_to = "Marks"
16   )
17
18 print("~~~ 2. Long Format (pivot_longer) ~~")
19 print(head(long_df, 9))
20
21
22 wide_df <- long_df %>%
23   pivot_wider(
24     names_from = Subject,
25     values_from = Marks
26   )
27
28 print("~~~ 3. Wide Format (Back to Original) ~~")
29 print(head(wide_df))
30
31 df_clean <- df %>%
32   mutate(Category = ifelse(is.na(Category), "Unknown", Category))
33
34 category_pivot <- df_clean %>%
35   select(StudentID, Category, Maths_Marks) %>%
36   pivot_wider(
37     names_from = Category,
38     values_from = Maths_Marks
39   )
40
41 print("~~~ 4. Category Pivot (Spreading Categories) ~~")
42 print(head(category_pivot))
```

SHETH L.U.J. AND SIR M.V. COLLEGE DATA ANALYSIS WITH SAS/SPSS/RR

OUTPUT:

```
package 'tidyr' was built under R version 4.4.3

> df <- read.csv("JEE Mains 2013-25 Top Ranks.csv", na.strings = c("", "NA")) %>%
+   mutate(StudentID = row_number()) %>%
+   select(StudentID, Category, Maths_Marks, Physics_Marks, Chemistry_Marks)

> print("~~~ 1. Original Wide Data ~~~")
[1] "~~~ 1. Original Wide Data ~~~"
> print(head(df))
  StudentID Category Maths_Marks Physics_Marks Chemistry_Marks
1         1   General          75             83              48
2         2   General          71             85              62
3         3   General          67             94              73
4         4 Reserved          52             71              66
5         5   General          40             65              65
6         6   General          99             75              50

> long_df <- df %>%
+   pivot_longer(
+     cols = c(Maths_Marks, Physics_Marks, Chemistry_Marks),
+     names_to = "Subject",
+     values_to = "Marks"
+   )
>
> print("~~~ 2. Long Format (pivot_longer) ~~~")
[1] "~~~ 2. Long Format (pivot_longer) ~~~"
> print(head(long_df, 9))
# A tibble: 9 x 4
  StudentID Category Subject      Marks
  <int>   <chr>   <chr>   <int>
1         1   General Maths_Marks    75
2         1   General Physics_Marks  83
3         1   General Chemistry_Marks 48
```

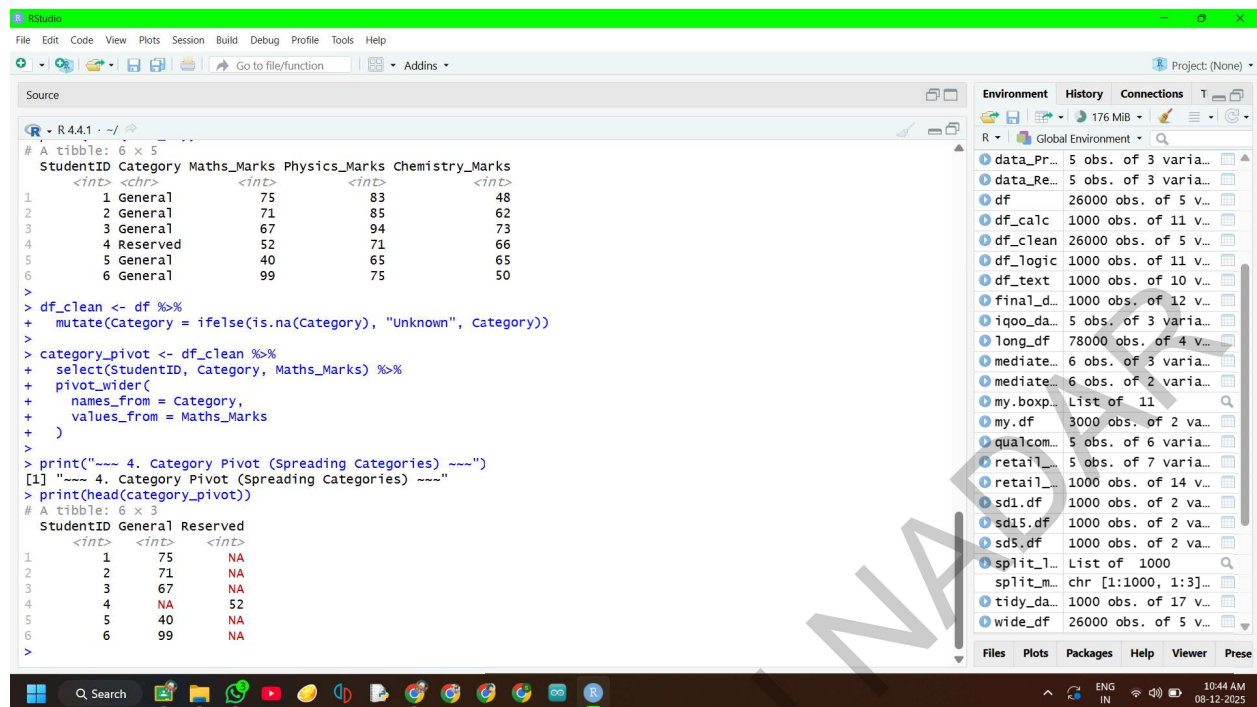
```
1         1   General Physics_Marks    83
2         1   General Chemistry_Marks  48
3         2   General Maths_Marks     71
4         2   General Physics_Marks    85
5         2   General Chemistry_Marks  62
6         3   General Maths_Marks     67
7         3   General Physics_Marks    94
8         3   General Chemistry_Marks  73
9         3   General Chemistry_Marks  73

> wide_df <- long_df %>%
+   pivot_wider(
+     names_from = Subject,
+     values_from = Marks
+   )
>
> print("~~~ 3. Wide Format (Back to Original) ~~~")
[1] "~~~ 3. Wide Format (Back to Original) ~~~"
> print(head(wide_df))
# A tibble: 6 x 5
  StudentID Category Maths_Marks Physics_Marks Chemistry_Marks
  <int>   <chr>   <int>   <int>   <int>
1         1   General          75             83              48
2         2   General          71             85              62
3         3   General          67             94              73
4         4 Reserved          52             71              66
5         5   General          40             65              65
6         6   General          99             75              50

> df_clean <- df %>%
+   mutate(Category = ifelse(is.na(Category), "Unknown", Category))
>
> category_pivot <- df_clean %>%
```

SELVESH NADAR

SHETH L.U.J. AND SIR M.V. COLLEGE DATA ANALYSIS WITH SAS/SPSS/RR



The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for data manipulation using dplyr. The code defines a tibble with 6 rows and 5 columns (StudentID, Category, Maths_Marks, Physics_Marks, Chemistry_Marks). It then performs a pivot_wider operation to spread the Category variable into columns.
- Environment Panel:** Lists the objects in the R environment, including data frames like df, df_clean, and the resulting category_pivot.
- Console:** Shows the output of the print statements, displaying the structure and first few rows of the category_pivot tibble.

```
# A tibble: 6 x 5
  StudentID Category Maths_Marks Physics_Marks Chemistry_Marks
  <int>    <chr>      <int>      <int>      <int>
1         1 General        75         83         48
2         2 General        71         85         62
3         3 General        67         94         73
4         4 Reserved       52         71         66
5         5 General        40         65         65
6         6 General        99         75         50

> df_clean <- df %>%
+   mutate(Category = ifelse(is.na(Category), "Unknown", Category))
>
> category_pivot <- df_clean %>%
+   select(StudentID, Category, Maths_Marks) %>%
+   pivot_wider(
+     names_from = Category,
+     values_from = Maths_Marks
+   )
>
> print("~~~ 4. Category Pivot (Spreading Categories) ~~~")
[1] "~~~ 4. Category Pivot (Spreading Categories) ~~~"
> print(head(category_pivot))
# A tibble: 6 x 3
  StudentID General Reserved
  <int>    <int>      <int>
1         1        75        NA
2         2        71        NA
3         3        67        NA
4         4         NA        52
5         5         40         NA
6         6         99         NA
```

