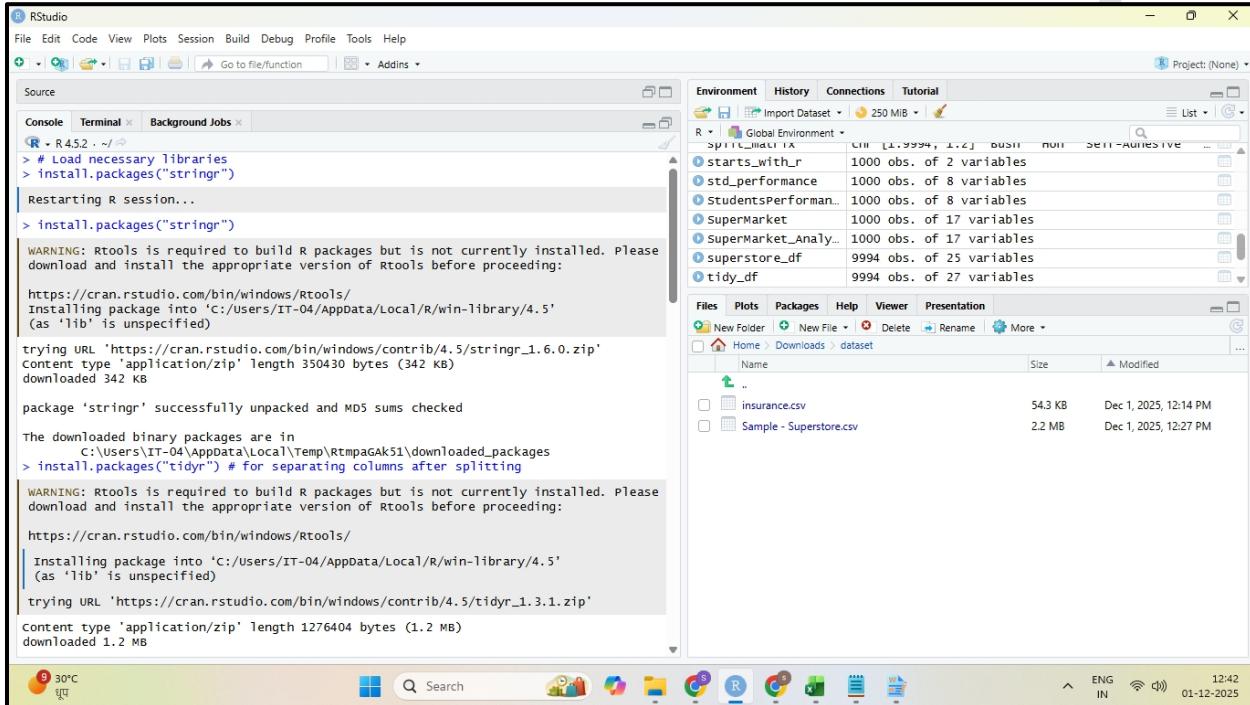


Sheth L.U.J. College of Arts & Sir M.V. College Of Science & Commerce  
SUBJECT NAME: Data Analysis with SAS / SPSS/R

PRACTICAL NO : 9

**AIM :** Performing text manipulation using `str_sub()`, `str_split()` (R). import dataset.



```
R + R 4.5.2 - ~/ ~
> # Load necessary libraries
> install.packages("stringr")
Restarting R session...
> install.packages("stringr")

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/IT-04/AppData/Local/R/win-library/4.5'
(as 'lib' is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/stringr_1.6.0.zip'
Content type 'application/zip' length 350430 bytes (342 KB)
downloaded 342 KB

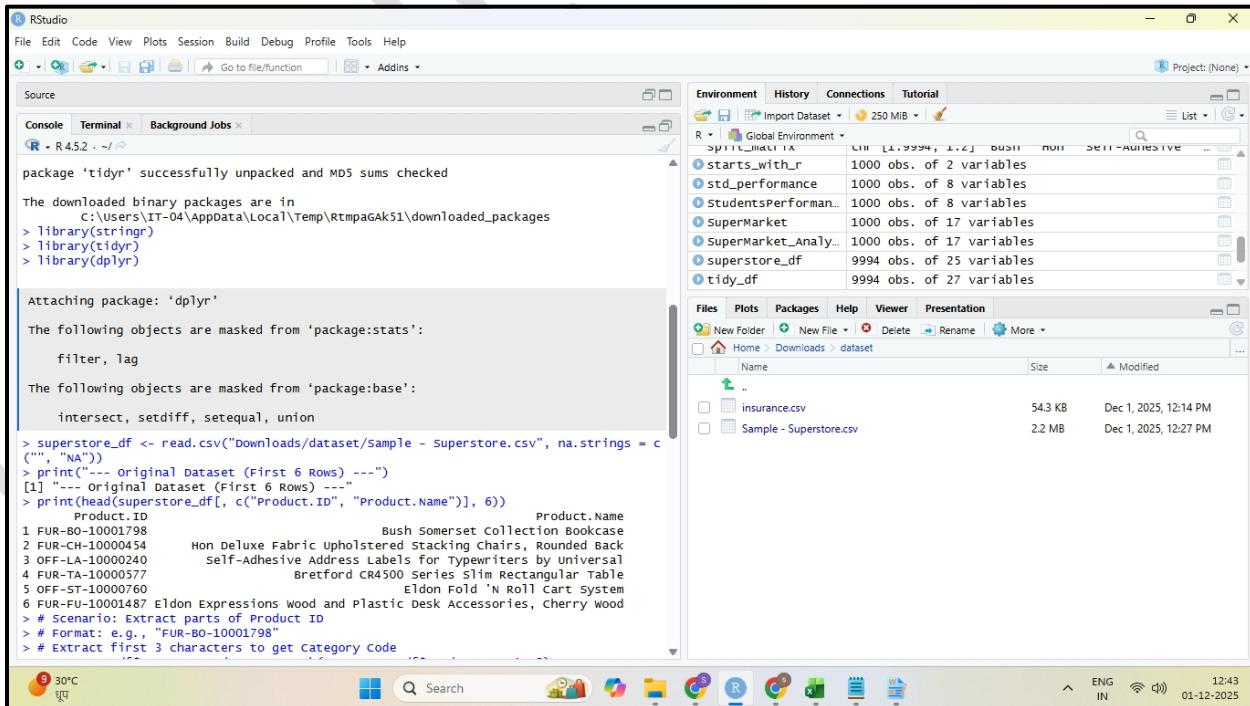
package 'stringr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:/users/IT-04/Appdata/Local/Temp/RtmpaGAKs1/downloaded_packages
> install.packages("tidyverse") # for separating columns after splitting

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/IT-04/AppData/Local/R/win-library/4.5'
(as 'lib' is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/tidyr_1.3.1.zip'
Content type 'application/zip' length 1276404 bytes (1.2 MB)
downloaded 1.2 MB
```



```
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

> superstore_df <- read.csv("Downloads/dataset/Sample - Superstore.csv", na.strings = c(
  "", "NA"))
> print("--- original dataset (First 6 Rows) ---")
[1] "--- original dataset (First 6 Rows) ---"
> print(head(superstore_df[, c("Product.ID", "Product.Name")], 6))
   Product.ID          Product.Name
1 FUR-B0-10001798    Bush Somerset Collection Bookcase
2 FUR-CH-10000454    Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back
3 OFF-LA-10000240    Self-Adhesive Address Labels for Typewriters by Universal
4 FUR-TA-10000577    Bretford CR4500 Series Slim Rectangular Table
5 OFF-ST-10000760    Elton Fold 'N Roll Cart System
6 FUR-FU-10001487    Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood
> # Scenario: Extract parts of Product ID
> # Format: e.g., "FUR-B0-10001798"
> # Extract first 3 characters to get Category Code
```

Sheth L.U.J. College of Arts & Sir M.V. College Of Science & Commerce  
SUBJECT NAME: Data Analysis with SAS / SPSS/R

The screenshot shows the RStudio interface with the following details:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Addins:** Go to file/function, Addins.
- Source:** Global Environment, History, Connections, Tutorial.
- Console:** Displays R code and its output. The code involves extracting category codes and product numbers from a dataset, splitting product names into main name and extra description, and then selecting specific columns.
- Environment:** Shows the global environment with various datasets loaded, such as starts\_with\_r, std\_performance, StudentsPerformance, SuperMarket, SuperMarket\_Analytic, superstore\_df, and tidy\_df.
- Files:** Shows files in the current directory: insurance.csv (543 KB, Dec 1, 2025) and Sample - Superstore.csv (2.2 MB, Dec 1, 2025).
- Bottom Bar:** Includes icons for search, help, and various RStudio features like file operations and project status.

The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help, and Help. Below the menu is a toolbar with various icons. The main area has tabs for Source, Console, Terminal, and Background Jobs. The Console tab is active, displaying R code and its output. The code involves reading a dataset, splitting it into main and sub names, and then printing the results. It also demonstrates the use of the 'separate' function from the 'tidyverse' package to separate Product.ID into Dept, Type, and Prod\_Num. The output shows the original dataset and the resulting tidy dataset. To the right of the console is the Environment pane, which lists global variables and their details. Below the Environment pane is a file browser with tabs for Files, Plots, Packages, Help, Viewer, and Presentation. The file browser shows a directory structure with 'dataset' selected, and files like 'insurance.csv' and 'Sample - Superstore.csv' listed.

Sheth L.U.J. College of Arts & Sir M.V. College Of Science & Commerce  
SUBJECT NAME: Data Analysis with SAS / SPSS/R



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Console Terminal Background Jobs

```
> superstore_df$Main_Name <- split_matrix[, 1] # First word
> superstore_df$Sub_Name <- split_matrix[, 2] # Remaining text
> print(" --- Data after str_split() ---")
[1] " --- Data after str_split() ---"
> print(superstore_df %>% select(Product.Name, Main.Name, Sub.Name) %>% head(6))
#> # A tibble: 6 × 3
#>   Product.Name Main.Name Sub.Name
#>   <chr>        <chr>    <chr>
#> 1 Bush Somerset Collection Bookcase Bush
#> 2 Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back Hon
#> 3 Self-Adhesive Address Labels for Typewriters by Universal self-Adhesive
#> 4 Bretford CR4500 Series Slim Rectangular Table Bretford
#> 5 Eldon Fold 'N Roll Cart System Eldon
#> 6 Eldon Expressions Wood and Plastic Desk Accessories, Cherry wood Eldon
#> 
#> # A tibble: 6 × 3
#>   Product.Name Main.Name Sub.Name
#>   <chr>        <chr>    <chr>
#> 1 Somerset Collection Bookcase
#> 2 Deluxe Fabric Upholstered Stacking Chairs, Rounded Back
#> 3 Address Labels for Typewriters by Universal
#> 4 CR4500 Series Slim Rectangular Table
#> 5 Fold 'N Roll Cart System
#> 6 Expressions Wood and Plastic Desk Accessories, Cherry wood
> # we can also separate Product.ID into Category, Type, and Number
> tidy_df <- superstore_df %>%
+   separate(Product.ID, into = c("Dept", "Type", "Prod_Num"), sep = "-")
> print(" --- Bonus: The 'separate' function ---")
[1] " --- Bonus: The 'separate' function ---"
> print(tidy_df %>% select(Dept, Type, Prod_Num) %>% head(6))
#> # A tibble: 6 × 3
#>   Dept Type Prod_Num
#>   <chr> <chr>   <dbl>
#> 1 FUR   BU  100001798
#> 2 FUR   CH  10000454
#> 3 OFF   LA  10000240
#> 4 FUR   TA  10000577
#> 5 OFF   ST  10000760
#> 6 FUR   FU  10001487
```

Environment History Connections Tutorial

Project: (None)

Imports: Superstore (1.9394, 1.2.1) RStudio (1.4.1.1) SETI - AVILES (1.0.0)

Global Environment

- starts\_with\_r 1000 obs. of 2 variables
- std\_performance 1000 obs. of 8 variables
- StudentsPerformance 1000 obs. of 8 variables
- SuperMarket 1000 obs. of 17 variables
- SuperMarket\_Analy... 1000 obs. of 17 variables
- superstore\_df 9994 obs. of 25 variables
- tidy\_df 9994 obs. of 27 variables

Files Plots Packages Help Viewer Presentation

New Folder New File Delete Rename More

Home Downloads dataset

Name	Size	Modified
..		
insurance.csv	54.3 KB	Dec 1, 2025, 12:14 PM
Sample - Superstore.csv	2.2 MB	Dec 1, 2025, 12:27 PM

ENG IN 12:43 01-12-2025