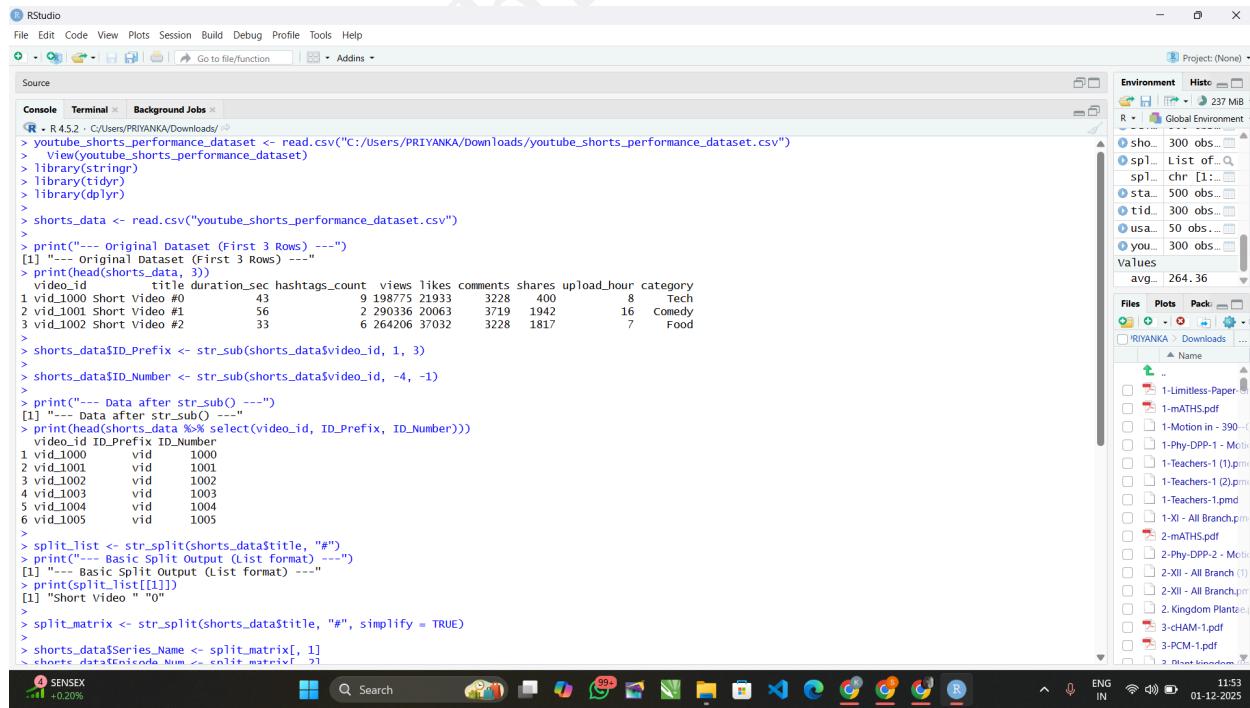


SHETH L.U.J AND SIR M.V. COLLEGE
SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R

Module 1 Practical 9

Aim: Performing text manipulation using str_sub(), str_split() (R).
import dataset.

OUTPUT:

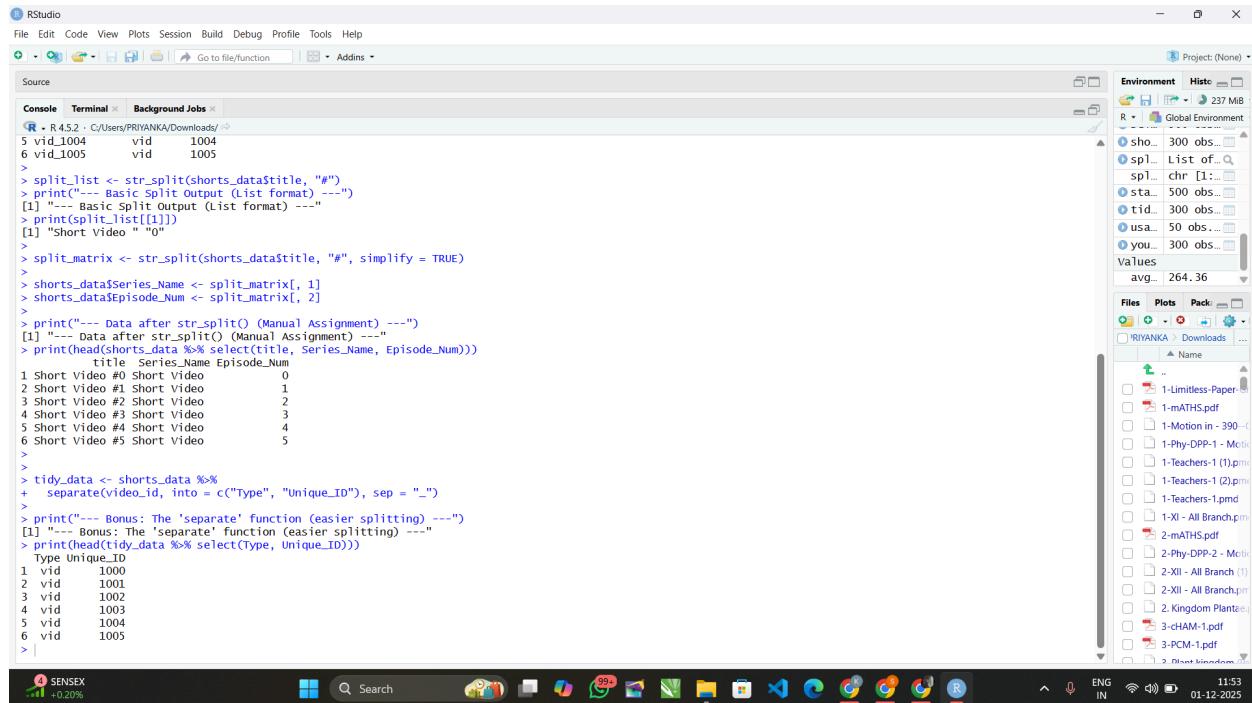


The screenshot shows the RStudio interface with the following details:

- Console Tab:** Displays the R code and its execution results. The code reads a CSV file, prints the first three rows, and then performs text manipulation using str_sub() and str_split() to extract video IDs from the title column.
- Environment Tab:** Shows the global environment with variables like shorts_data, shor..., sp1..., sta..., tid..., usa..., you..., and Values.
- Files Tab:** Shows the project structure under PRIYANKA > Downloads, including various PDF files and RMD files.
- Plots Tab:** Shows a histogram titled "1-Limitless-Paper" with a red arrow pointing to it.
- Session Tab:** Shows the session information including the operating system (Windows 10), R version (4.5.2), and the date (01-12-2025).

```
R - R 4.5.2 · C:/users/PRIYANKA/Downloads/ ◊
> youtube_shorts_performance_dataset <- read.csv("C:/Users/PRIYANKA/Downloads/youtube_shorts_performance_dataset.csv")
> View(youtube_shorts_performance_dataset)
> library(stringr)
> library(tidyverse)
> library(dplyr)
>
> shorts_data <- read.csv("youtube_shorts_performance_dataset.csv")
>
> print("--- Original Dataset (First 3 Rows) ---")
[1] "--- original Dataset (First 3 Rows) ---"
> print(head(shorts_data, 3))
#> #> #> #> #> #>
video_id          title duration_sec hashtags_count   views likes comments shares upload_hour category
1 vid_1000 Short Video #0           43        9 198775 21933    3228   400       8     Tech
2 vid_1001 Short Video #1           56        2 290336 20063    3719   1942      16 Comedy
3 vid_1002 Short Video #2           33        6 264206 37032    3228   1817       7 Food
>
> shorts_data$ID_Prefix <- str_sub(shorts_data$title, 1, 3)
>
> shorts_data$ID_Number <- str_sub(shorts_data$title, -4, -1)
>
> print("--- Data after str_sub() ---")
[1] "--- Data after str_sub() ---"
> print(head(shorts_data %>% select(video_id, ID_Prefix, ID_Number)))
#> #> #> #> #> #>
video_id ID_Prefix ID_Number
1 vid_1000    vid    1000
2 vid_1001    vid    1001
3 vid_1002    vid    1002
4 vid_1003    vid    1003
5 vid_1004    vid    1004
6 vid_1005    vid    1005
>
> split_list <- str_split(shorts_data$title, "#")
> print("--- Basic Split Output (List format) ---")
[1] "--- Basic Split Output (List format) ---"
> print(split_list[[1]])
[1] "Short Video #0"
>
> split_matrix <- str_split(shorts_data$title, "#", simplify = TRUE)
>
> shorts_data$Series_Name <- split_matrix[, 1]
> shorts_data$enclata_Num <- enclata_matrix[, 1]
```

SHETH L.U.J AND SIR M.V. COLLEGE
SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R



The screenshot shows the RStudio interface with the following details:

- Console:** Displays R code and its output. The code involves reading a file, splitting it into a list, and then into a matrix. It then filters the data by title and episode number, and finally uses the 'tidy' function to separate columns into rows.
- Environment:** Shows the global environment with objects like `sho_`, `sp1_`, `sta_`, `usa_`, `you_`, and `Values`.
- Files:** Shows a folder structure with various files and sub-folders, including PDFs and PPTs related to ATHS assignments.
- Plots:** Shows a small preview of a plot.
- Pack:** Shows a list of packages installed.

The R code in the console is as follows:

```
R 4.5.2 · C:/Users/PRIYANKA/Downloads/ 
5 vid_1004     vid    1004
6 vid_1005     vid    1005
> 
> split_list <- str_split(shorts_data$title, "#")
> print("--- Basic Split Output (List format) ---")
[1] "--- Basic Split Output (List format) ---"
> print(split_list[[1]])
[1] "Short Video " "0"
> 
> split_matrix <- str_split(shorts_data$title, "#", simplify = TRUE)
> 
> shorts_data$Series_Name <- split_matrix[, 1]
> shorts_data$Episode_Num <- split_matrix[, 2]
> 
> print(" --- Data after str_split( Manual Assignment) ---")
[1] " --- Data after str_split( Manual Assignment) ---"
> print(head(shorts_data %>% select(title, Series_Name, Episode_Num)))
#> #>   title Series_Name Episode_Num
1 Short Video #0 Short Video      0
2 Short Video #1 Short Video      1
3 Short Video #2 Short Video      2
4 Short Video #3 Short Video      3
5 Short Video #4 Short Video      4
6 Short Video #5 Short Video      5
> 
> tidy_data <- shorts_data %>%
+   separate(video_id, into = c("Type", "Unique_ID"), sep = "_")
> 
> print(" --- Bonus: The 'separate' function (easier splitting) ---")
[1] " --- Bonus: The 'separate' function (easier splitting) ---"
> print(head(tidy_data %>% select(Type, Unique_ID)))
#> #>   Type Unique_ID
1 vid      1000
2 vid      1001
3 vid      1002
4 vid      1003
5 vid      1004
6 vid      1005
> |
```