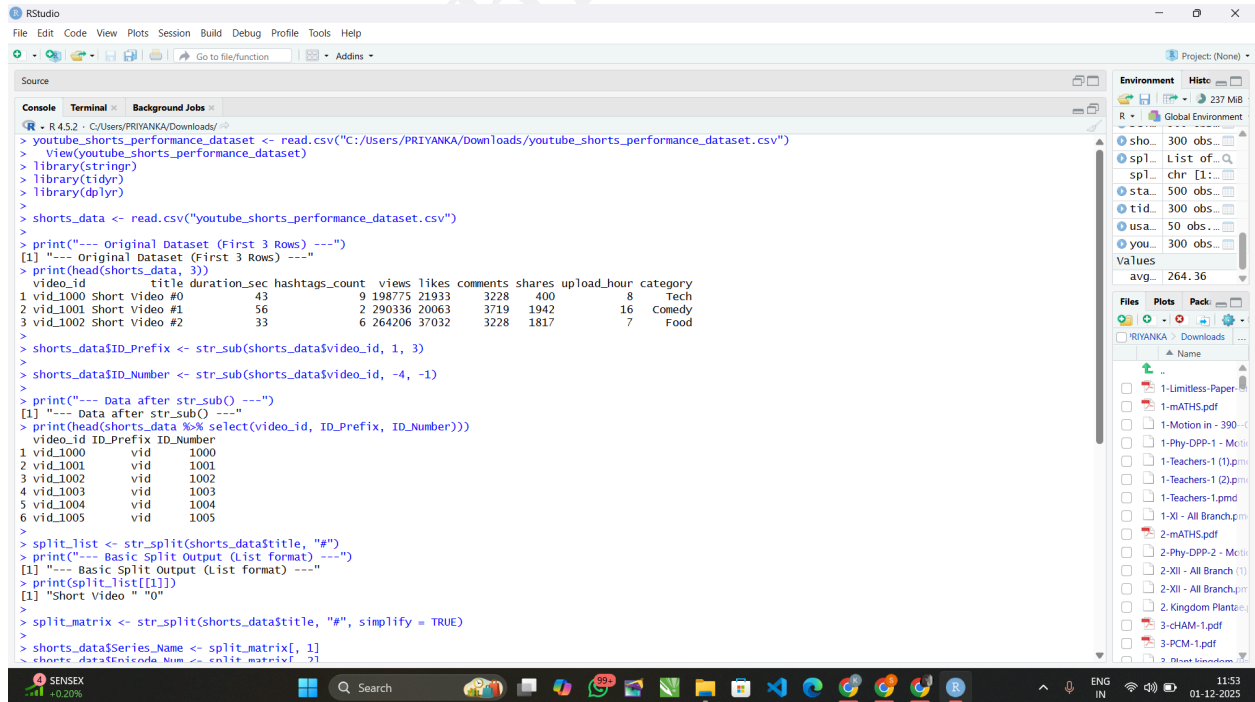


SHETH L.U.J AND SIR M.V. COLLEGE
SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R

Module 1 Practical 9

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

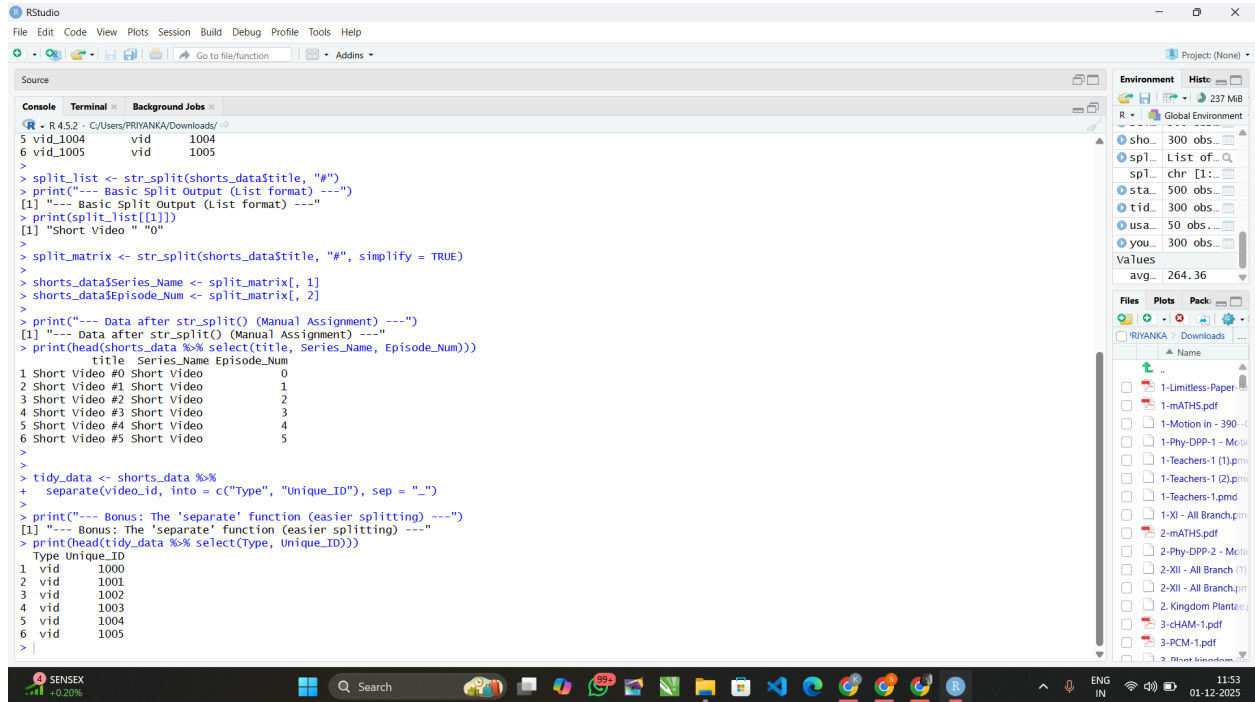
OUTPUT:



```
R • R 4.5.2 • C:/Users/PRIYANKA/Downloads/
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
> youtube_shorts_performance_dataset <- read.csv("C:/Users/PRIYANKA/downloads/youtube_shorts_performance_dataset.csv")
> View(youtube_shorts_performance_dataset)
> library(stringr)
> library(tidyr)
> 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$ID_Prefix <- str_sub(shorts_data$video_id, 1, 3)
>
> shorts_data$ID_Number <- str_sub(shorts_data$video_id, -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$Series_Name <- split_matrix[, 1]
> shorts_data$Series_Name <- col1str_matrix[, 1]
```

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

SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R



The screenshot displays the RStudio environment with the following components:

- Source Editor:** Contains R code for data manipulation.
- Console:** Shows the output of the executed code.
- Environment:** Lists objects in the workspace, including 'sho...', 'spl...', 'sta...', 'tid...', 'usa...', 'you...', and 'avg...'. The 'avg...' object has a value of 264.36.
- Files:** Shows a list of files in the 'Downloads' folder, including '1-Limitless-Paper...', '1-mATHS.pdf', '1-Motion in - 390-(...', '1-Phy-DPP-1 - Moti...', '1-Teachers-1 (1).pm...', '1-Teachers-1 (2).pm...', '1-XI - All Branch.p...', '2-mATHS.pdf', '2-Phy-DPP-2 - Moti...', '2-XII - All Branch (1)', '2-XII - All Branch.p...', '2. Kingdom Plantae...', '3-CHAM-1.pdf', and '3-PCM-1.pdf'.

R Code in Source Editor:

```
R 4.5.2 - C:/Users/PRIVANKA/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  
> |
```

Console Output:

```
R 4.5.2 - C:/Users/PRIVANKA/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  
> |
```