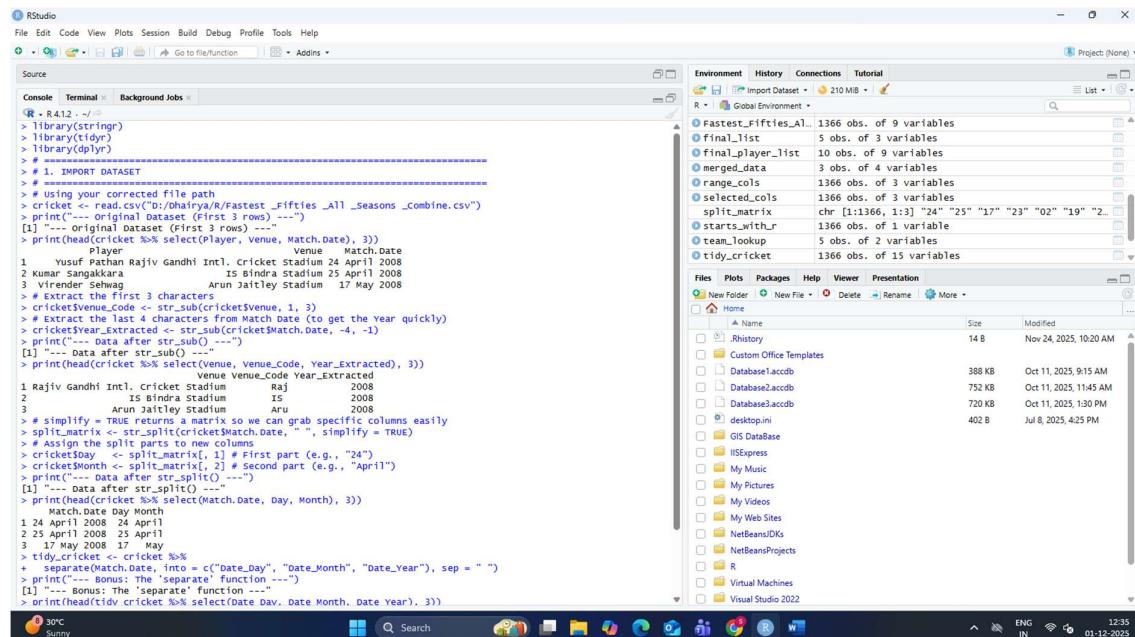


**SHETH L.U.J AND SIR M.V. COLLEGE**  
**SUBJECT NAME: Data Analysis with SAS / SPSS /R**  
**Practical No. 9**

**Aim-Performing text manipulation using str\_sub (), str\_split () (R). import dataset.**

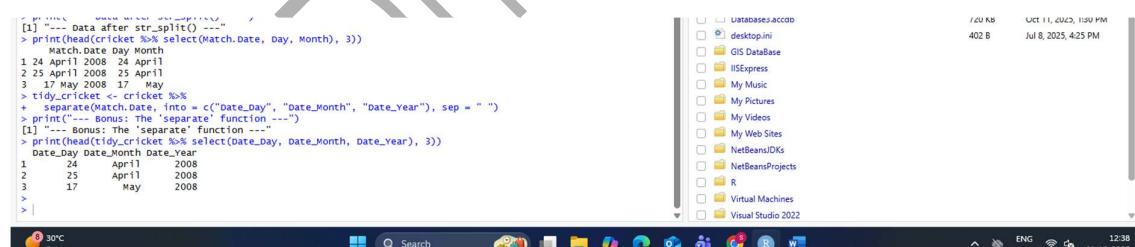
**Output-**



```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Console Terminal Background Jobs
(R - R4.12)
> library(stringr)
> library(tidyverse)
> library(dplyr)
> # =====
> # 1. IMPORT DATASET
> # =====
> # using your corrected file path
> cricket <- read.csv("D:/Dhairyaa/R/Fastest_Fifties_All_Seasons_Combine.csv")
> print("--- original Dataset (First 3 rows) ---")
[1] "--- original dataset (First 3 rows) ---"
> print(head(cricket %>% select(Player, Venue, Match.Date), 3))
#> #> #> Player           Venue   Match.Date
#> #> 1 Yusuf Pathan Rajiv Gandhi Intl. Cricket Stadium 24 April 2008
#> #> 2 Kumar Sangakkara IS Bindra Stadium 25 April 2008
#> #> 3 Virender Sehwag Arun Jaitley Stadium 17 May 2008
> # Extract the first 3 characters from the venue name
> cricket$Venue_code <- str_sub(cricket$Venue, 1, 3)
> # Extract the last 4 characters from Match Date (to get the year quickly)
> cricket$Year_Extracted <- str_sub(cricket$Match.date, -4, -1)
> print("--- data after str_sub() ---")
[1] "--- data after str_sub() ---"
> print(head(cricket %>% select(Venue, Venue_Code, Year_Extracted), 3))
#> #> #> Venue Venue_Code Year_Extracted
#> #> 1 Rajiv Gandhi Intl. Cricket Stadium Raj 2008
#> #> 2 IS Bindra Stadium IS 2008
#> #> 3 Arun Jaitley Stadium Aru 2008
> # simplify a few things here because we can grab specific columns easily
> split_matrix <- str_split(cricket$Match.date, " ", simplify = TRUE)
> # Assign the split parts to new columns
> cricket$Day <- split_matrix[, 1] # First part (e.g., "24")
> cricket$Month <- split_matrix[, 2] # Second part (e.g., "April")
> print("--- data after str_split() ---")
[1] "--- data after str_split() ---"
> print(head(cricket %>% select(Match.Date, Day, Month), 3))
#> #> #> Match.Date Day Month
#> #> 1 24 April 2008 24 April
#> #> 2 25 April 2008 25 April
#> #> 3 17 May 2008 17 May
> tidy_cricket <- cricket %>%
+   separate(Match.Date, into = c("Date_Day", "Date_Month", "Date_Year"), sep = " ")
> print("--- bonus: The 'separate' function ---")
[1] "--- bonus: The 'separate' function ---"
> print(head(tidy_cricket %>% select(Date_Day, Date_Month, Date_Year), 3))
#> #> #> Date_Day Date_Month Date_Year
#> #> 1 24 April 2008
#> #> 2 25 April 2008
#> #> 3 17 May 2008
> |

```



```

> print(head(cricket %>% select(Match.Date, Day, Month), 3))
#> #> #> Match.Date Day Month
#> #> 1 24 April 2008 24 April
#> #> 2 25 April 2008 25 April
#> #> 3 17 May 2008 17 May
> tidy_cricket <- cricket %>%
+   separate(Match.Date, into = c("Date_Day", "Date_Month", "Date_Year"), sep = " ")
> print("--- bonus: The 'separate' function ---")
[1] "--- bonus: The 'separate' function ---"
> print(head(tidy_cricket %>% select(Date_Day, Date_Month, Date_Year), 3))
#> #> #> Date_Day Date_Month Date_Year
#> #> 1 24 April 2008
#> #> 2 25 April 2008
#> #> 3 17 May 2008
> |

```