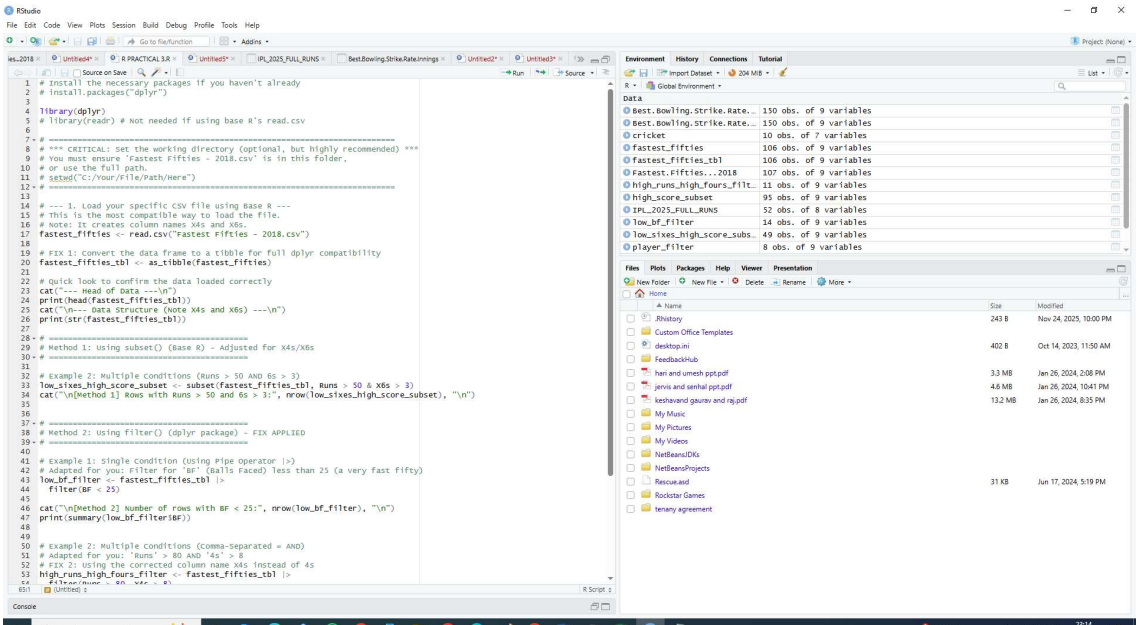
MVLU COLLEGE
R PRACTICAL 4

Aim: Applying conditional filters subset() or filter() in R.

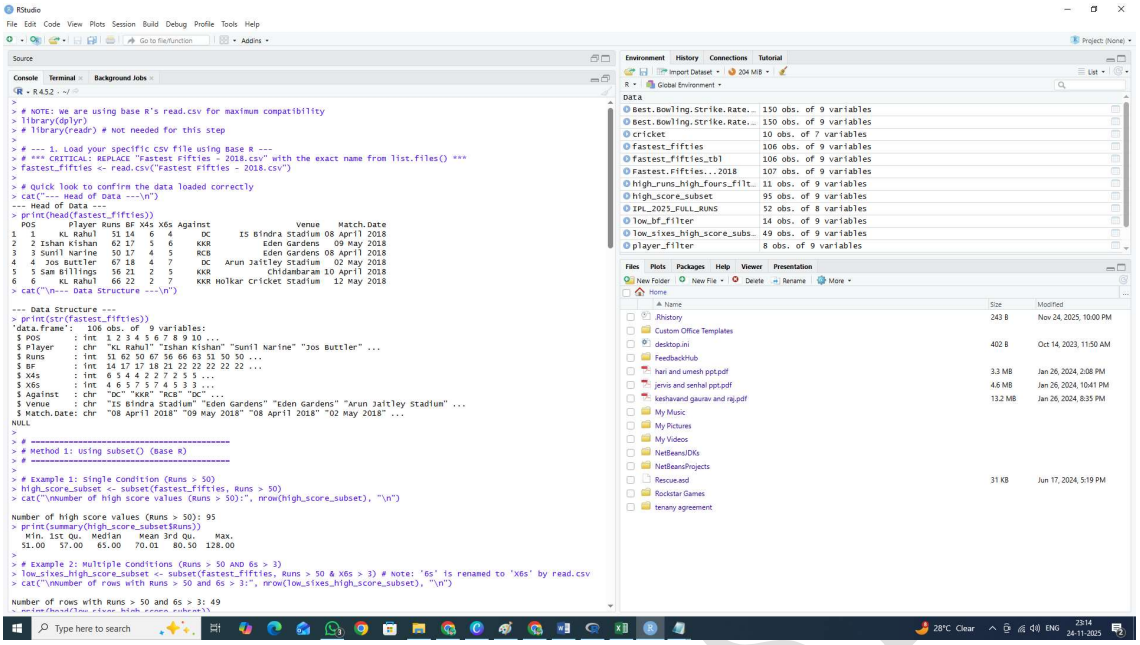


```
# Install the necessary packages if you haven't already
# install.packages("dplyr")

library(dplyr)
# library(readr) # Not needed if using base R's read.csv

# ==========================================================================
# *** CRITICAL: Set the working directory (optional, but highly recommended) ***
# You must ensure 'Fastest Fifties - 2018.csv' is in this folder.
# Or use the full path.
# setwd("C:/Your/File/Path/Here")
# ==========================================================================

# --- 1. Load your specific CSV file using Base R ---
# This is the most compatible way to load the file.
# Note: It creates column names X4s and X6s.
fastest_fifties <- read.csv("Fastest Fifties - 2018.csv")

# FIX 1: Convert the data frame to a tibble for full dplyr compatibility
fastest_fifties_tbl <- as_tibble(fastest_fifties)

# Quick look to confirm the data loaded correctly
cat("--- Head of Data ---\n")
print(head(fastest_fifties_tbl))
cat("\n--- Data Structure (Note X4s and X6s) ---\n")
print(str(fastest_fifties_tbl))

# ========================================
# Method 1: Using subset() (Base R) - Adjusted for X4s/X6s
# ========================================

# Example 2: Multiple Conditions (Runs > 50 AND 6s > 3)
low_sixes_high_score_subset <- subset(fastest_fifties_tbl, Runs > 50 & X6s > 3)
cat("\n[Method 1] Rows with Runs > 50 and 6s > 3:", nrow(low_sixes_high_score_subset), "\n")


# ========================================
# Method 2: Using filter() (dplyr package) - FIX APPLIED
# ========================================

# Example 1: Single Condition (Using Pipe Operator |>)
# Adapted for you: Filter for 'BF' (Balls Faced) less than 25 (a very fast fifty)
low_bf_filter <- fastest_fifties_tbl |>
  filter(BF < 25)

cat("\n[Method 2] Number of rows with BF < 25:", nrow(low_bf_filter), "\n")
print(summary(low_bf_filter$BF))


# Example 2: Multiple Conditions (Comma-Separated = AND)
# Adapted for you: 'Runs' > 80 AND '4s' > 8
# FIX 2: Using the corrected column name X4s instead of 4s
high_runs_high_fours_filter <- fastest_fifties_tbl |>
  filter(Runs > 80, X4s > 8)

cat("\n[Method 2] Rows where Runs > 80 and 4s > 8:", nrow(high_runs_high_fours_filter), "\n")
print(head(high_runs_high_fours_filter))


# Example 3: Checking for Values in a Set (%in%)
# Adapted for you: Filter where 'Player' is exactly "KL Rahul" or "Ishan Kishan"
player_filter <- fastest_fifties_tbl |>
  filter(Player %in% c("KL Rahul", "Ishan Kishan"))

cat("\n[Method 2] Number of rows with specific players:", nrow(player_filter), "\n")
print(head(player_filter))
```

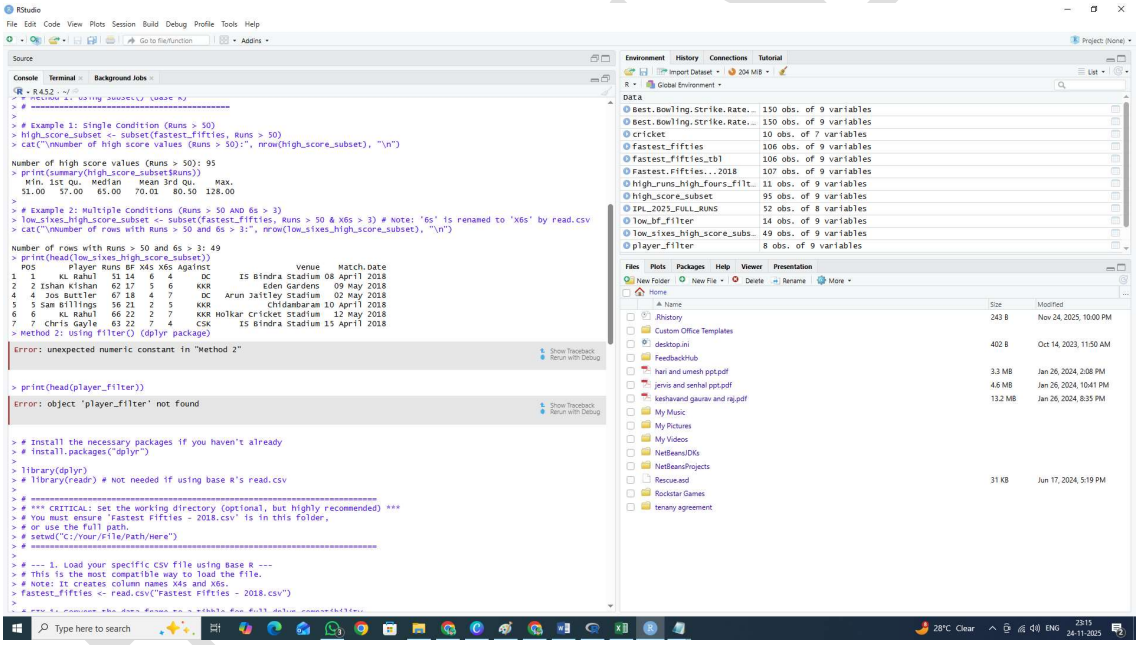
output:-

Hariprasad Vishwakarma
S126
SYCS

MVLU COLLEGE
R PRACTICAL 4



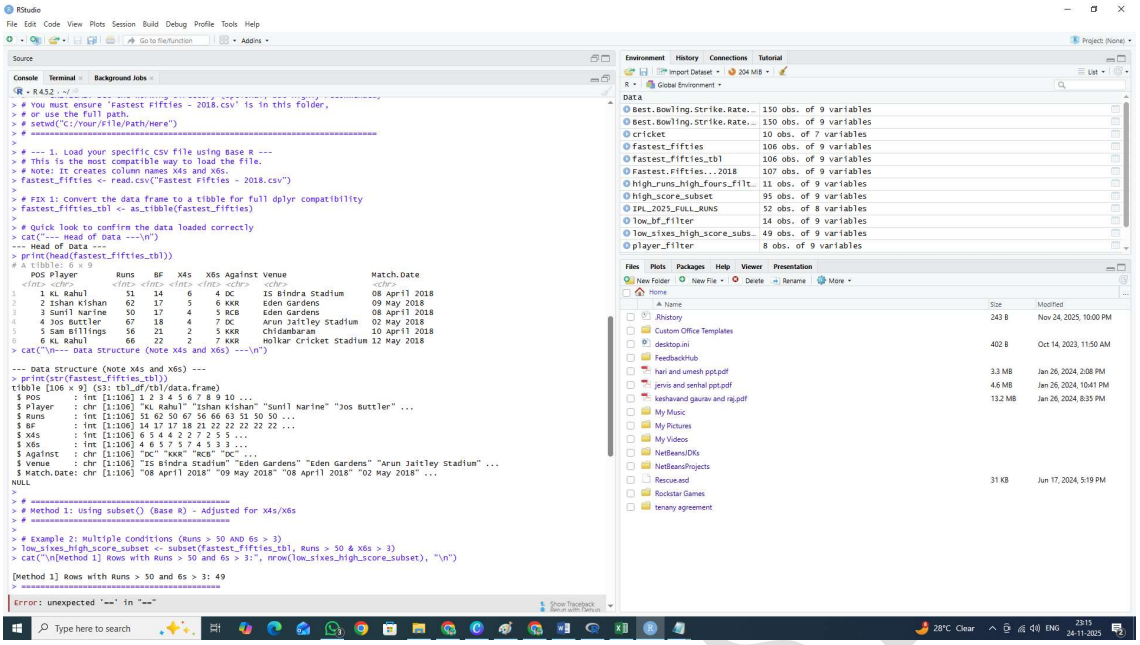
```
> # NOTE: We are using base R's read.csv for maximum compatibility
> library(dplyr)
> # library(readr) # Not needed for this step
>
> # --- 1. Load your specific CSV file using Base R ---
> # *** CRITICAL: REPLACE "Fastest Fifties - 2018.csv" with the exact name from list.files() ***
> fastest_fifties <- read.csv("Fastest Fifties - 2018.csv")
>
> # Quick look to confirm the data loaded correctly
> cat("--- Head of Data ---\n")
--- Head of Data ---
> print(head(fastest_fifties))
  POS       Player Runs BF X4s X6s Against                 Venue    Match.Date
1   1      KL Rahul   51 14   6   4      DC       IS Bindra Stadium 08 April 2018
2   2  Ishan Kishan   62 17   5   6     KKR            Eden Gardens   09 May 2018
3   3  Sunil Narine   50 17   4   5     RCB            Eden Gardens 08 April 2018
4   4   Jos Buttler   67 18   4   7      DC   Arun Jaitley Stadium   02 May 2018
5   5  Sam Billings   56 21   2   5     KKR             Chidambaram 10 April 2018
6   6      KL Rahul   66 22   2   7     KKR Holkar Cricket Stadium   12 May 2018
> cat("\n--- Data Structure ---\n")

--- Data Structure ---
> print(str(fastest_fifties))
'data.frame':   106 obs. of  9 variables:
 $ POS       : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Player    : chr  "KL Rahul" "Ishan Kishan" "Sunil Narine" "Jos Buttler" ...
 $ Runs      : int  51 62 50 67 56 66 63 51 50 50 ...
 $ BF        : int  14 17 17 18 21 22 22 22 22 22 ...
 $ X4s       : int  6 5 4 4 2 2 7 2 5 5 ...
 $ X6s       : int  4 6 5 7 5 7 4 5 3 3 ...
 $ Against   : chr  "DC" "KKR" "RCB" "DC" ...
 $ Venue     : chr  "IS Bindra Stadium" "Eden Gardens" "Eden Gardens" "Arun Jaitley Stadium" ...
 $ Match.Date: chr  "08 April 2018" "09 May 2018" "08 April 2018" "02 May 2018" ...
NULL
>
> # ----------------------------------------------
> # Method 1: Using subset() (Base R)
> # ----------------------------------------------
>
> # Example 1: Single Condition (Runs > 50)
> high_score_subset <- subset(fastest_fifties, Runs > 50)
> cat("\nNumber of high score values (Runs > 50):", nrow(high_score_subset), "\n")

Number of high score values (Runs > 50): 95
> print(summary(high_score_subset$Runs))
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  51.00   57.00   65.00   70.01   80.50  128.00
>
> # Example 2: Multiple Conditions (Runs > 50 AND 6s > 3)
> low_sixes_high_score_subset <- subset(fastest_fifties, Runs > 50 & X6s > 3) # Note: '6s' is renamed to 'X6s' by read.csv
> cat("\nNumber of rows with Runs > 50 and 6s > 3:", nrow(low_sixes_high_score_subset), "\n")

Number of rows with Runs > 50 and 6s > 3: 49
```

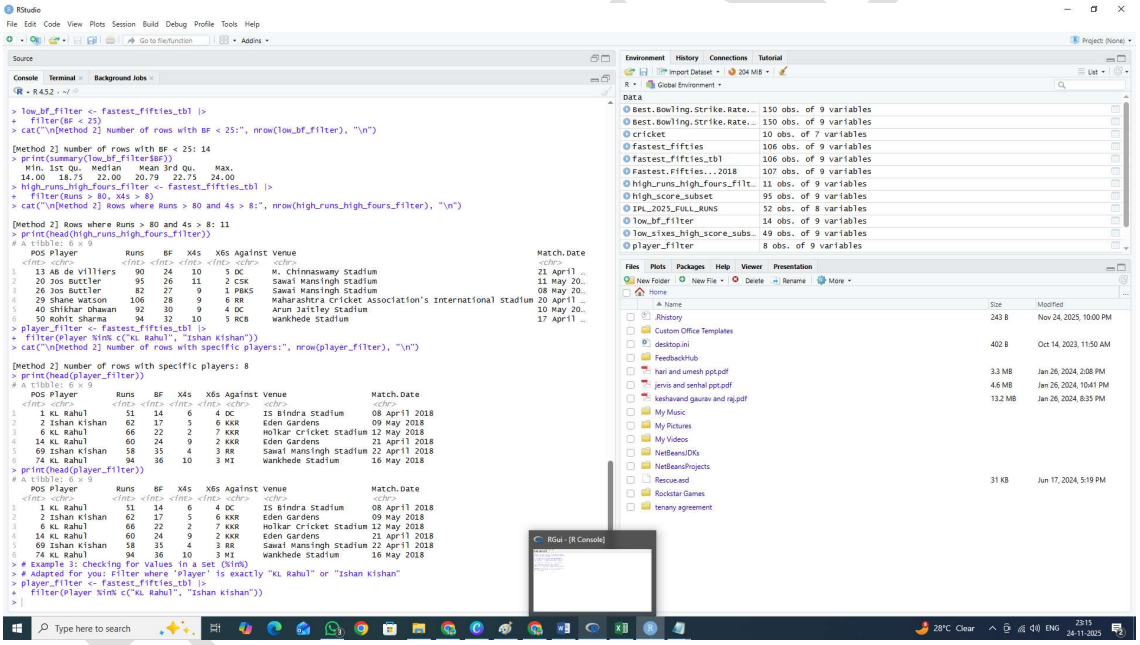


```
> # ----------------------------------------------
>
> # Example 1: Single Condition (Runs > 50)
> high_score_subset <- subset(fastest_fifties, Runs > 50)
> cat("\nNumber of high score values (Runs > 50):", nrow(high_score_subset), "\n")

Number of high score values (Runs > 50): 95
> print(summary(high_score_subset$Runs))
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  51.00   57.00   65.00   70.01   80.50  128.00
>
> # Example 2: Multiple Conditions (Runs > 50 AND 6s > 3)
> low_sixes_high_score_subset <- subset(fastest_fifties, Runs > 50 & X6s > 3) # Note: '6s' is renamed to 'X6s' by read.csv
> cat("\nNumber of rows with Runs > 50 and 6s > 3:", nrow(low_sixes_high_score_subset), "\n")

Number of rows with Runs > 50 and 6s > 3: 49
> print(head(low_sixes_high_score_subset))
  POS       Player Runs BF X4s X6s Against                 Venue    Match.Date
1   1      KL Rahul   51 14   6   4      DC       IS Bindra Stadium 08 April 2018
2   2  Ishan Kishan   62 17   5   6     KKR            Eden Gardens   09 May 2018
4   4   Jos Buttler   67 18   4   7      DC   Arun Jaitley Stadium   02 May 2018
5   5  Sam Billings   56 21   2   5     KKR             Chidambaram 10 April 2018
6   6      KL Rahul   66 22   2   7     KKR Holkar Cricket Stadium   12 May 2018
7   7   Chris Gayle   63 22   7   4     CSK       IS Bindra Stadium 15 April 2018
> Method 2: Using filter() (dplyr package)
Error: unexpected numeric constant in "Method 2"

> print(head(player_filter))
Error: object 'player_filter' not found

> # Install the necessary packages if you haven't already
> # install.packages("dplyr")
>
> library(dplyr)
> # library(readr) # Not needed if using base R's read.csv
>
> # ======================================================================
> # *** CRITICAL: Set the working directory (optional, but highly recommended) ***
> # You must ensure 'Fastest Fifties - 2018.csv' is in this folder,
> # or use the full path.
> # setwd("C:/Your/File/Path/Here")
> # ======================================================================
>
> # --- 1. Load your specific CSV file using Base R ---
> # This is the most compatible way to load the file.
> # Note: It creates column names X4s and X6s.
> fastest_fifties <- read.csv("Fastest Fifties - 2018.csv")
>
```

Hariprasad Vishwakarma
S126
SYCS

MVLU COLLEGE
R PRACTICAL 4





Hariprasad Vishwakarma
S126
SYCS