

MVLU COLLEGE R PRACTICAL 4

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

```
1 # Install the necessary packages if you haven't already
2 # install.packages("dplyr")
3
4 library(dplyr)
5 # library(readr) # Not needed if using base R's read.csv
6
7 # *** CRITICAL: Set the working directory (optional, but highly recommended) ***
8 # You must ensure 'fastest_fifties - 2018.csv' is in this folder.
9 # or use the full path.
10 # setwd("C:/Your/File/Path/Here")
11 # =====
12
13 # --- 1. Load your specific csv file using Base R ---
14 # This is the most compatible way to load the file.
15 # Note: It creates column names X4s and X6s.
16 fastest_fifties <- read.csv("fastest_fifties - 2018.csv")
17
18 # FIX 1: Convert the data frame to a tibble for full dplyr compatibility
19 fastest_fifties_tbl <- as_tibble(fastest_fifties)
20
21 # Quick look to confirm the data loaded correctly
22 cat("--- Head of data ---\n")
23 print(head(fastest_fifties_tbl))
24 cat("\n--- Data structure (note X4s and X6s) ---\n")
25 print(str(fastest_fifties_tbl))
26
27 # =====
28 # Method 1: Using subset() (Base R) - Adjusted for X4s/X6s
29 # =====
30
31 # Example 2: Multiple conditions (Runs > 50 AND 6s > 3)
32 low_sixes_high_score_subset <- subset(fastest_fifties_tbl, Runs > 50 & X6s > 3)
33 cat("\nMethod 1] Rows with Runs > 50 and 6s > 3:", nrow(low_sixes_high_score_subset), "\n")
34
35 # =====
36 # Method 2: Using filter() (dplyr package) - FIX APPLIED
37 # =====
38
39 # Example 1: Single condition (using pipe operator |>)
40 # Adapted for you: filter for "sr" (balls faced) less than 25 (a very fast Fifty)
41 low_bf_filter <- fastest_fifties_tbl |>
42   filter(bf < 25)
43
44 cat("\nMethod 2] Number of rows with bf < 25:", nrow(low_bf_filter), "\n")
45 print(summary(low_bf_filter))
46
47 # =====
48 # Example 2: Multiple conditions (Comma-separated = AND)
49 # Adapted for you: Runs > 80 AND '4s' > 8
50 # FIX 2: Using the corrected column name X4s instead of 4s
51 high_runs_high_fours_filter <- fastest_fifties_tbl |>
52   filter(Runs > 80, X4s > 8)
53
54 cat("\nMethod 2] Rows where Runs > 80 and 4s > 8:", nrow(high_runs_high_fours_filter), "\n")
55 print(head(high_runs_high_fours_filter))
56
57 # Example 3: checking for values in a set %in%
58 # Adapted for you: filter where "player" is exactly "KL Rahul" or "Ishan Kishan"
59 player_filter <- fastest_fifties_tbl |>
60   filter(player %in% c("KL Rahul", "Ishan Kishan"))
61
62 cat("\nMethod 2] Number of rows with specific players:", nrow(player_filter), "\n")
63 print(head(player_filter))
```

output:-

MVLU COLLEGE R PRACTICAL 4

The image displays two screenshots of an RStudio interface, showing R code execution and the environment panel.

Top Screenshot:

- Source Panel:** Contains R code for loading data, inspecting it, and performing subsetting operations. The code includes comments and function calls like `read.csv()`, `subset()`, and `filter()`.
- Environment Panel:** Lists objects in the global environment, including `best.bowling.strike.rate`, `best.bowling.strike.rate...`, `cricket`, `fastest.fifties`, `fastest.fifties.tbl`, `fastest.fifties...2018`, `high.runs.high.fours.filt`, `high.score_subset`, `ipl_2023_full_runs`, `low.bf.filter`, `low.sixes.high.score_sub`, and `player.filter`.

Bottom Screenshot:

- Source Panel:** Continues the R code, showing errors and further data manipulation steps. It includes comments about installing packages and ensuring the working directory.
- Environment Panel:** Lists objects in the global environment, including `best.bowling.strike.rate`, `best.bowling.strike.rate...`, `cricket`, `fastest.fifties`, `fastest.fifties.tbl`, `fastest.fifties...2018`, `high.runs.high.fours.filt`, `high.score_subset`, `ipl_2023_full_runs`, `low.bf.filter`, `low.sixes.high.score_sub`, and `player.filter`.

MVLU COLLEGE R PRACTICAL 4

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
R - R4.2.2
Source Console Terminal Background Jobs
# 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 x5s.
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))
# A tibble: 8 x 9
  pos Player Runs BF x4s x5s Against Venue Match.Date
<ints> <chr> <ints> <ints> <ints> <chr> <chr>
1 1 kl.rahul 51 14 6 4 DC 15 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 3 KKR Chidambaram 10 April 2018
6 6 kl.rahul 66 22 2 7 KKR molkar cricket stadium 12 May 2018
# cat("---- data structure (note x4s and x5s) ----\n")
# data structure (note x4s and x5s) ----
print(str(fastest_fifties_tbl))
tibble [106 x 9] (S3: tbl_df/tbl/data.frame)
 $ pos      : chr [1:106] 1 2 3 4 5 6 7 8 9 10 ...
 $ Player   : chr [1:106] "kl.rahul" "Ishan kishan" "Sunil narine" "Jos buttler" ...
 $ Runs     : int [1:106] 51 62 50 67 56 66 61 51 50 50 ...
 $ BF       : int [1:106] 14 17 17 18 21 22 22 22 22 ...
 $ x4s      : int [1:106] 6 5 4 4 2 2 2 5 5 ...
 $ x5s      : int [1:106] 4 6 5 7 5 7 4 5 1 3 ...
 $ Against  : chr [1:106] "DC" "KKR" "RCB" "DC" ...
 $ Venue    : chr [1:106] "15 Bindra stadium" "eden gardens" "eden gardens" "Arun Jaitley stadium" ...
 $ Match.Date: chr [1:106] "08 April 2018" "09 May 2018" "08 April 2018" "02 May 2018" ...
 NULL
# -----
# Method 1: using subset() (base R) - Adjusted for x4s/x5s
# -----
# Example 2: Multiple conditions (Runs > 50 AND 6s > 3)
low_sixes_high_score_subset <- subset(fastest_fifties_tbl, Runs > 50 & x5s > 3)
cat("\nMethod 1] Rows with Runs > 50 and 6s > 3", nrow(low_sixes_high_score_subset), "\n")
# Method 1] Rows with Runs > 50 and 6s > 3: 49
# -----
Error: unexpected '---' in '---'
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
R - R4.2.2
Source Console Terminal Background Jobs
# low_sixes_filter <- fastest_fifties_tbl |>
# filter(BF < 25)
# cat("\nMethod 2] Number of rows with BF < 25:", nrow(low_sixes_filter), "\n")
# Method 2] Number of rows with BF < 25: 14
print(summary(low_sixes_filter))
# Min. 1st Qu. Median Mean 3rd Qu. Max.
# 14.00 18.75 22.00 20.79 22.75 24.00
high_runs_high_fours_filter <- fastest_fifties_tbl |>
# filter(Runs > 80, x4s > 8)
# cat("\nMethod 2] Rows where Runs > 80 and 4s > 8:", nrow(high_runs_high_fours_filter), "\n")
# Method 2] Rows where Runs > 80 and 4s > 8: 11
print(head(high_runs_high_fours_filter))
# A tibble: 6 x 9
  pos Player Runs BF x4s x5s Against Venue Match.Date
<ints> <chr> <ints> <ints> <ints> <chr> <chr>
1 13 AB de villiers 90 24 10 5 DC W. Chinnaswamy stadium 21 April ...
2 20 Jos buttler 95 26 11 2 CSK Savai Wansingh stadium 11 May 20...
3 26 Jos buttler 82 27 9 1 PAKS Savai Wansingh stadium 08 May 20...
4 29 Shane watson 106 28 9 6 RR Maharashtra Cricket Association's International stadium 20 April ...
5 40 Shikhar dhawan 92 30 9 4 DC Arun Jaitley stadium 10 May 20...
6 50 Rohit sharma 94 32 10 5 RCB Wankhede stadium 17 April ...
# low_sixes_filter <- fastest_fifties_tbl |>
# filter(player %in% c("kl.rahul", "Ishan kishan"))
# cat("\nMethod 2] Number of rows with specific players:", nrow(player_filter), "\n")
# Method 2] Number of rows with specific players: 8
print(head(player_filter))
# A tibble: 6 x 9
  pos Player Runs BF x4s x5s Against Venue Match.Date
<ints> <chr> <ints> <ints> <ints> <chr> <chr>
1 1 kl.rahul 51 14 6 4 DC 15 Bindra stadium 08 April 2018
2 2 Ishan kishan 62 17 5 6 KKR eden gardens 09 May 2018
3 6 kl.rahul 66 22 2 7 KKR holkar cricket stadium 12 May 2018
4 14 kl.rahul 60 24 9 2 KKR eden gardens 21 April 2018
5 69 Ishan kishan 58 35 4 3 RR Savai Wansingh stadium 22 April 2018
6 74 kl.rahul 94 36 10 3 MI Wankhede stadium 16 May 2018
# print(head(player_filter))
# A tibble: 6 x 9
  pos Player Runs BF x4s x5s Against Venue Match.Date
<ints> <chr> <ints> <ints> <ints> <chr> <chr>
1 1 kl.rahul 51 14 6 4 DC 15 Bindra stadium 08 April 2018
2 2 Ishan kishan 62 17 5 6 KKR eden gardens 09 May 2018
3 6 kl.rahul 66 22 2 7 KKR holkar cricket stadium 12 May 2018
4 14 kl.rahul 60 24 9 2 KKR eden gardens 21 April 2018
5 69 Ishan kishan 58 35 4 3 RR Savai Wansingh stadium 22 April 2018
6 74 kl.rahul 94 36 10 3 MI Wankhede stadium 16 May 2018
# 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"))
```