

MVLU COLLEGE R PRACTICAL 4

The image displays two screenshots of the RStudio IDE, showing the process of loading and manipulating a dataset in R.

Top Screenshot: The R script editor shows the following code:

```
## 2018 ##
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 # -----
8 *** CRITICAL: Set the working directory (optional, but highly recommended) ***
9 # you must ensure 'Fastest Fifties - 2018.csv' is in this folder,
10 # or use the full path.
11 # setwd("C:/Your/Path/Here")
12 # -----
13
14 # --- 1. Load your specific CSV file using Base R ---
15 # this is the most compatible way to load the file.
16 # Note: It creates column names x4s and x6s.
17 fastest_fifties <- read.csv("fastest_fifties - 2018.csv")
18
19 # FIX 1: Convert the data frame to a tibble for full dplyr compatibility
20 fastest_fifties_tbl <- as_tibble(fastest_fifties)
21
22 # Quick look to confirm the data loaded correctly
23 cat("--- Head of data ---\n")
24 print(head(fastest_fifties_tbl))
25 cat("\n--- Data Structure (Note x4s and x6s) ---\n")
26 print(str(fastest_fifties_tbl))
27
28 # -----
29 # Method 1: using subset() (Base R) - Adjusted for x4s/x6s
30 # -----
31
32 # Example 2: Multiple conditions (Runs > 50 AND 6s > 3)
33 low_sixes_high_score_subset <- subset(fastest_fifties_tbl, Runs > 50 & x6s > 3)
34 cat("\nMethod 1] Rows with Runs > 50 and 6s > 3:", nrow(low_sixes_high_score_subset), "\n")
35
36 # -----
37 # Method 2: using filter() (dplyr package) - FIX APPLIED
38 # -----
39
40 # Example 1: Single condition (Using pipe operator |>)
41 # Adapted for you: filter for 'BF' (balls Faced) less than 25 (a very fast fifty)
42 low_bf_filter <- fastest_fifties_tbl |>
43   filter(bf < 25)
44
45 cat("\nMethod 2] Number of rows with BF < 25:", nrow(low_bf_filter), "\n")
46 print(summary(low_bf_filter$bf))
47
48 # -----
49 # Example 2: Multiple conditions (Comma-separated & AND)
50 # Adapted for you: 'Runs' > 80 AND '4s' > 8
51 # FIX 2: Using the corrected column name x4s instead of 4s
52 high_runs_high_fours_filter <- fastest_fifties_tbl |>
53   filter(Runs > 80, x4s > 8)
54
55 cat("\nMethod 2] Rows where Runs > 80 and 4s > 8:", nrow(high_runs_high_fours_filter), "\n")
56 print(head(high_runs_high_fours_filter))
57
58 # -----
59 # Example 3: checking for values in a set (N%in%)
60 # Adapted for you: filter where 'Player' is exactly "KL Rahul" or "Ishan Kishan"
61 player_filter <- fastest_fifties_tbl |>
62   filter(player %in% c("KL Rahul", "Ishan Kishan"))
63
64 cat("\nMethod 2] Number of rows with specific players:", nrow(player_filter), "\n")
65 print(head(player_filter))
66
67 # -----
68 # Method 3: using filter() (dplyr package) - FIX APPLIED
69 # -----
```

The Environment pane on the right shows the loaded objects:

- data
- 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
- low_bf_filter
- low_sixes_high_score_sub...
- player_filter

Bottom Screenshot: The R script editor shows the same code as the top screenshot, but with additional code at the bottom:

```
68 # -----
69 # Method 3: using filter() (dplyr package) - FIX APPLIED
70 # -----
71
72 # Example 1: Single condition (Using pipe operator |>)
73 # Adapted for you: filter for 'BF' (balls Faced) less than 25 (a very fast fifty)
74 low_bf_filter <- fastest_fifties_tbl |>
75   filter(bf < 25)
76
77 cat("\nMethod 2] Number of rows with BF < 25:", nrow(low_bf_filter), "\n")
78 print(summary(low_bf_filter$bf))
79
80 # -----
81 # Example 2: Multiple conditions (Comma-separated & AND)
82 # Adapted for you: 'Runs' > 80 AND '4s' > 8
83 # FIX 2: Using the corrected column name x4s instead of 4s
84 high_runs_high_fours_filter <- fastest_fifties_tbl |>
85   filter(Runs > 80, x4s > 8)
86
87 cat("\nMethod 2] Rows where Runs > 80 and 4s > 8:", nrow(high_runs_high_fours_filter), "\n")
88 print(head(high_runs_high_fours_filter))
89
90 # -----
91 # Example 3: checking for values in a set (N%in%)
92 # Adapted for you: filter where 'Player' is exactly "KL Rahul" or "Ishan Kishan"
93 player_filter <- fastest_fifties_tbl |>
94   filter(player %in% c("KL Rahul", "Ishan Kishan"))
95
96 cat("\nMethod 2] Number of rows with specific players:", nrow(player_filter), "\n")
97 print(head(player_filter))
98
99 # -----
100 # Method 3: using filter() (dplyr package) - FIX APPLIED
101 # -----
```

The Environment pane on the right shows the same loaded objects as the top screenshot.

output:-

Hariprasad Vishwakarma
S126
SYCS

MVLU COLLEGE R PRACTICAL 4

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to File/Function Addins

Source
R432 - v2
# 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))
      Player Runs BP Xs6 Xs5 Against      Venue      Match Date
1      KL Rahul    51 14 6 4      DC      IS Bindra Stadium 08 April 2018
2      Ishan Kishan 62 17 5 6      KKR      Eden Gardens      09 May 2018
3      Sunil Narine 50 17 4 5      RCB      Eden Gardens 08 April 2018
4      Jos Buttler 67 18 4 7      DC      Arun Jaitley Stadium 02 May 2018
5      Sam Billings 56 21 2 5      KKR      Chidambaram 10 April 2018
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 ...
 $ bp        : int  14 17 17 18 21 22 22 22 22 22 ...
 $ xs6       : int  6 5 4 4 2 2 2 2 5 5 ...
 $ xs5       : 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("Number 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))
      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 & xs6 > 3) # Note: '6s' is renamed to 'xs6' by read.csv
> cat("Number 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))
      Player Runs BP Xs6 Xs5 Against      Venue      Match Date
1      KL Rahul    51 14 6 4      DC      IS Bindra Stadium 08 April 2018
2      Ishan Kishan 62 17 5 6      KKR      Eden Gardens      09 May 2018
4      Jos Buttler 67 18 4 7      DC      Arun Jaitley Stadium 02 May 2018
5      Sam Billings 56 21 2 5      KKR      Chidambaram 10 April 2018
6      KL Rahul    66 22 2 7      KKR      Holkar Cricket Stadium 12 May 2018
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 xs6 and xs5.
> fastest_fifties <- read.csv("Fastest Fifties - 2018.csv")
>
> # =====
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to File/Function Addins

Source
R432 - v2
> # Method 1: using subset() (base R)
> # =====
> # Example 1: Single Condition (Runs > 50)
> high_score_subset <- subset(fastest_fifties, runs > 50)
> cat("Number 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))
      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 & xs6 > 3)
> cat("Number 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))
      Player Runs BP Xs6 Xs5 Against      Venue      Match Date
1      KL Rahul    51 14 6 4      DC      IS Bindra Stadium 08 April 2018
2      Ishan Kishan 62 17 5 6      KKR      Eden Gardens      09 May 2018
4      Jos Buttler 67 18 4 7      DC      Arun Jaitley Stadium 02 May 2018
5      Sam Billings 56 21 2 5      KKR      Chidambaram 10 April 2018
6      KL Rahul    66 22 2 7      KKR      Holkar Cricket Stadium 12 May 2018
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 xs6 and xs5.
> fastest_fifties <- read.csv("Fastest Fifties - 2018.csv")
>
> # =====
```

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_bf_filter <- fastest_fifties_tbl |>
# filter(bf < 25)
# cat("\nMethod 2] Number of rows with bf < 25:", nrow(low_bf_filter), "\n")
# Method 2] Number of rows with bf < 25: 14
print(summary(low_bf_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 Savaii Mansingh stadium 11 May 20...
3 26 Jos buttler 82 27 9 1 PAKS Savaii Mansingh 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_bf_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 molkar 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 Savaii Mansingh 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 molkar 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 Savaii Mansingh 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"))
```