

MVLU COLLEGE

PRACTICAL NO.7

Aim: Selecting and dropping variables using select() in R. import dataset.

Code:-

The screenshot shows the RStudio interface with two panes. The left pane displays an R script titled 'Untitled 1' containing code for selecting and dropping variables. The right pane shows the 'Environment' tab of the global environment, listing various datasets like 'dropped_one', 'Fastest.Fifties..2018', and 'Retail.Product'. Below the environment is a file browser showing local files including 'rue_sales.csv' and several R-related files.

```

# 7. Selecting and dropping variables using select() in R
# 1. IMPORT DATASET
# library(dplyr)
# *** IMPORTANT: UPDATE FILE PATH HERE ***
# Placeholder: Use the file name from your screenshot, but you must replace
# "path/to/your/file/" with your actual directory path, e.g.,
# "D:/Hariprasad/Fastest Fifties - 2018.csv"
cricket <- read.csv("path/to/your/file/Fastest Fifties - 2018.csv")
print("--- original dataset (First 3 rows) ---")
print(head(cricket, 3))

# 2. SELECTING VARIABLES (Keeping Columns)
# Method A: Select specific columns by name
# Scenario: we only want the Player Name, Runs scored, and the opponent (Against)
selected_cols <- cricket
select(Player, Runs, Against)
print("--- Selected specific columns ---")
print(head(selected_cols, 3))

# Method B: Select a range of adjacent columns
# Scenario: select everything from 'Player' to '4s'
range_cols <- cricket
select(Player:4)
# Note: The column name '4s' is used exactly as it appears in the Excel data.
print("--- selected Range of columns (Player through 4s) ---")
print(head(range_cols, 3))

# Method C: Select using helper functions
# Scenario: select columns that start with "P" (POS, Player)
starts_with_p <- cricket
select(starts_with("P"))
print("--- Selected columns starting with 'P' (POS, Player) ---")
print(head(starts_with_p, 3))

# 3. DROPPING VARIABLES (Removing columns)
# we use the minus sign (-) to remove variables
# Method A: Drop a single specific column
# Scenario: Remove the 'venue' column
dropped_venue <- cricket
select(-venue)
print("--- dataset with 'venue' dropped ---")
print(names(dropped_venue)) # Printing names to verify it's gone

# Method B: drop multiple columns
# Scenario: Remove '6s' and 'BF' (Balls Faced)
dropped_multiple <- cricket

```

This screenshot shows the same RStudio session as above, but the code has been modified to demonstrate dropping variables. The left pane now contains code for dropping 'venue' and '6s/BF' columns. The right pane remains the same, showing the environment and file browser.

```

# Method A: Drop a single specific column
# Scenario: Remove the 'venue' column
dropped_venue <- cricket
select(-venue)
print("--- dataset with 'venue' dropped ---")
print(names(dropped_venue)) # Printing names to verify it's gone

# Method B: drop multiple columns
# Scenario: Remove '6s' and 'BF' (Balls Faced)
dropped_multiple <- cricket

```

MVLU COLLEGE

PRACTICAL NO.7

The screenshot shows the RStudio interface. The left pane displays an R script with code for dataset manipulation, including dropping columns and rows. The right pane shows the Environment browser with various datasets loaded, such as 'Global Environment', 'dropped_one', 'Fastest.Fifties..2018', 'merged_data', 'range_cols', 'Retail.Product', and 'starts_with_p'. Below the browser is a file browser showing local files like 'fuel_sales.csv' and 'GIS DataBase'.

```

# Source on Save
# Selected columns starting with 'r' (row, player, ...)

# 3. DROPPING VARIABLES (Removing columns)
# we use the minus sign (-) to remove variables

# Method A: Drop a single specific column
# Scenario: Remove the 'venue' column
dropped_one <- cricket
select(-venue)

print("... dataset with 'venue' dropped ...")
print(names(dropped_one)) # Printing names to verify it's gone

# Method B: Drop multiple columns
# Scenario: Remove '6s' and 'BF' (Balls Faced)
dropped_multiple <- cricket
select(-6, -BF)

# Note: The column name '6s' is used exactly as it appears in the Excel data.

print("... dataset with '6s' and 'BF' dropped ...")
print(names(dropped_multiple))

# Method C: Drop a range of columns
# Scenario: Remove everything from '4s' to 'Match.Date'
dropped_range <- cricket
select(-4:Match.Date)

print("... dataset with range '4s' to 'Match.Date' dropped ...")
print(names(dropped_range))

```

Output:

The screenshot shows the RStudio interface. The left pane displays the R console with error messages and command-line output. The right pane shows the Environment browser and file browser.

```

> # *** IMPORTANT: UPDATE FILE PATH HERE ***
> # Placeholder: use the file name from your screenshot, but you must replace
> # "path/to/your/file/" with your actual directory path, e.g.,
> # "D:/Haris126/Fastest Fifties - 2018.csv"
> cricket <- read.csv("path/to/your/file/Fastest Fifties - 2018.csv")

Error in file(file, "rt") : cannot open the connection

In addition: warning message:
In file(file, "rt") :
  cannot open file 'path/to/your/file/Fastest Fifties - 2018.csv': No such file or directory

> print("--- original dataset (First 3 rows) ---")
[1] "--- original dataset (First 3 rows) ---"
> print(head(cricket, 3))

   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

> # Method A: Select specific columns by name
> # Scenario: We only want the Player Name, Runs scored, and the Opponent (Against)
> selected_cols <- cricket
> select(Player, Runs, Against)

Error in select(Player, Runs, Against) : object 'Player' not found

> print("--- selected specific Columns ---")
[1] "--- Selected Specific Columns ---"
> print(head(selected_cols, 3))

   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

```

MVLU COLLEGE

PRACTICAL NO.7

The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help, and Addins. The title bar says "RStudio". The left pane contains a "Console" tab with R code and its output. The right pane contains an "Environment" browser showing various global variables and their details. A large watermark "Hariprasad" is diagonally across the image.

```
R - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function | Addins |
Source
Console Terminal x Background Jobs x
R - R 4.1.2 - ~ / ...
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
> select(Player:4)
Error in select(Player:4) : object 'Player' not found

> print(" --- Selected Range of Columns (Player through 4s) --- ")
[1] " --- Selected Range of Columns (Player through 4s) --- "
> print(head(range_cols, 3))
   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
> # Method C: select using helper functions
> # Scenario: select columns that start with "P" (POS, Player)
> starts_with_p <- cricket
> select(starts_with("P"))

Error:
! 'starts_with()' must be used within a *selecting* function.
i See ?tidyselect::faq-selection-context for details.
Run `rlang::last_trace()` to see where the error occurred.

> print(" --- Selected columns starting with 'P' (POS, Player) --- ")
[1] " --- Selected columns starting with 'P' (POS, Player) --- "
> print(head(starts_with_p, 3))
   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
> # Method A: Drop a single specific column
> # Scenario: Remove the 'venue' column
> dropped_one <- cricket
> select(-venue)

Environment History Connections Tutorial
Project: (None) ▾
Import Dataset 213 MB
Global Environment
dropped_range 106 obs. of 9 variables
Fastest.Fiftie.. 107 obs. of 9 variables
merged_data 106 obs. of 10 variables
range_cols 106 obs. of 9 variables
Retail.Product 4362 obs. of 5 variables
selected_cols 106 obs. of 9 variables
starts_with_p 106 obs. of 9 variables
Files Plots Packages Help Viewer Presentation
Folder File Delete Rename ...
Home
Name Size Modified
tue_sales.csv 766 B Sep 18, 2025, 8:35 AM
GIS DataBase
IISExpress
My Music
My Pictures
My Videos
My Web Sites
R
SM PRACTICAL - II 20.txt 1.2 KB Aug 22, 2025, 3:02 PM
SM PRACTICAL - II.txt 1.2 KB Aug 22, 2025, 3:01 PM
total_bills.csv 1.2 KB Sep 18, 2025, 8:47 AM
Virtual Machines
Visual Studio 2022
WindowsPowerShell
12:23
ENG IN 01-12-2025
```

MVLU COLLEGE

PRACTICAL NO.7

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source

```

Console Terminal Background Jobs
R > R4.1.2 - ~
> select(-venue)

Error in select(-venue) : object 'venue' not found

> print("--- dataset with 'venue' dropped ---")
[1] "--- Dataset with 'Venue' dropped ---"
> print(names(dropped_one)) # Printing names to verify it's gone
[1] "POS"      "Player"    "Runs"     "BF"       "X4s"      "X6s"      "Against"
[8] "Venue"    "Match.Date"
> # Method B: Drop multiple columns
> # Scenario: Remove '6s' and 'BF' (Balls Faced)
> dropped_multiple <- cricket
> select(-6, -BF)

Error in useMethod("select") :
no applicable method for 'select' applied to an object of class "c('double', 'numeric')"

> print("--- dataset with '6s' and 'BF' dropped ---")
[1] "--- Dataset with '6s' and 'BF' dropped ---"
> print(names(dropped_multiple))
[1] "POS"      "Player"    "Runs"     "BF"       "X4s"      "X6s"      "Against"
[8] "Venue"    "Match.Date"
> # Method C: Drop a range of columns
> # Scenario: Remove everything from '4s' to 'Match_date'
> dropped_range <- cricket
> select(-(4:matchdate))

Error in select(-(4:MatchDate)) : object 'MatchDate' not found

> print("--- Dataset with range '4s' to 'Match_date' dropped ---")
[1] "--- Dataset with range '4s' to 'Match.Date' dropped ---"
> print(names(dropped_range))
[1] "POS"      "Player"    "Runs"     "BF"       "X4s"      "X6s"      "Against"
[8] "Venue"    "Match.Date"

```

Environment History Connections Tutorial

Project: (None)

Import Dataset 213 MB

- Global Environment
- dropped_one 100 obs. of 9 variables
- dropped_range 106 obs. of 9 variables
- Fastest.Fiftie.. 107 obs. of 9 variables
- merged_data 106 obs. of 10 variables
- range_cols 106 obs. of 9 variables
- Retail.Product 4362 obs. of 5 variables
- selected_cols 106 obs. of 9 variables
- starts_with_p 106 obs. of 9 variables

Files Plots Packages Help Viewer Presentation

Folder File Delete Rename

Home

Name	Size	Modified
tue_sales.csv	706 b	Sep 18, 2025, 8:35 AM
GIS DataBase		
IISExpress		
My Music		
My Pictures		
My Videos		
My Web Sites		
R		
SM PRACTICAL - II 20.txt	1.2 KB	Aug 22, 2025, 3:02 PM
SM PRACTICAL - II.txt	1.2 KB	Aug 22, 2025, 3:01 PM
total_bills.csv	1.2 KB	Sep 18, 2025, 8:47 AM
Virtual Machines		
Visual Studio 2022		
WindowsPowerShell		

ENG IN 12:23 01-12-2025

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source

```

Console Terminal Background Jobs
R > R4.1.2 - ~

> print("--- dataset with 'Venue' dropped ---")
[1] "--- Dataset with 'Venue' dropped ---"
> print(names(dropped_one)) # Printing names to verify it's gone
[1] "POS"      "Player"    "Runs"     "BF"       "X4s"      "X6s"      "Against"
[8] "Venue"    "Match.Date"
> # Method B: Drop multiple columns
> # Scenario: Remove '6s' and 'BF' (Balls Faced)
> dropped_multiple <- cricket
> select(-6, -BF)

Error in useMethod("select") :
no applicable method for 'select' applied to an object of class "c('double', 'numeric')"

> print("--- dataset with '6s' and 'BF' dropped ---")
[1] "--- Dataset with '6s' and 'BF' dropped ---"
> print(names(dropped_multiple))
[1] "POS"      "Player"    "Runs"     "BF"       "X4s"      "X6s"      "Against"
[8] "Venue"    "Match.Date"
> # Method C: Drop a range of columns
> # Scenario: Remove everything from '4s' to 'Match_date'
> dropped_range <- cricket
> select(-(4:matchdate))

Error in select(-(4:MatchDate)) : object 'MatchDate' not found

> print("--- Dataset with range '4s' to 'Match_date' dropped ---")
[1] "--- Dataset with range '4s' to 'Match.Date' dropped ---"
> print(names(dropped_range))
[1] "POS"      "Player"    "Runs"     "BF"       "X4s"      "X6s"      "Against"
[8] "Venue"    "Match.Date"

```

Environment History Connections Tutorial

Project: (None)

Import Dataset 213 MB

- Global Environment
- dropped_one 100 obs. of 9 variables
- dropped_range 106 obs. of 9 variables
- Fastest.Fiftie.. 107 obs. of 9 variables
- merged_data 106 obs. of 10 variables
- range_cols 106 obs. of 9 variables
- Retail.Product 4362 obs. of 5 variables
- selected_cols 106 obs. of 9 variables
- starts_with_p 106 obs. of 9 variables

Files Plots Packages Help Viewer Presentation

Folder File Delete Rename

Home

Name	Size	Modified
tue_sales.csv	706 b	Sep 18, 2025, 8:35 AM
GIS DataBase		
IISExpress		
My Music		
My Pictures		
My Videos		
My Web Sites		
R		
SM PRACTICAL - II 20.txt	1.2 KB	Aug 22, 2025, 3:02 PM
SM PRACTICAL - II.txt	1.2 KB	Aug 22, 2025, 3:01 PM
total_bills.csv	1.2 KB	Sep 18, 2025, 8:47 AM
Virtual Machines		
Visual Studio 2022		
WindowsPowerShell		

ENG IN 12:23 01-12-2025