

Sheth L.U.J. & Sir M.V. College

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

Showing 1 to 26 of 6,607 entries, 20 total columns

Console

Gold +1.35% Search

11:55:16 AM 01-12-2025

```
1 # 1. IMPORT DATASET
2 # 2. SELECTING VARIABLES (keeping columns)
3 # Method A: Select specific columns by their names
4 # Scenario: we want to keep only Hours_Studied, Attendance, and Exam_Score
5 selected_cols <- student %>%
6   select(Hours_Studied, Attendance, Exam_Score)
7
8 print("... Original Dataset (First 3 rows) ...")
9 print(head(student, 3))
10
11
12 # Method B: select a continuous range of adjacent columns
13 # Scenario: Select from the first column to the 5th column
14 # Example: Selecting columns from Hours_Studied to Extracurricular_Activities
15 range_cols <- student %>%
16   select(Hours_Studied:Extracurricular_Activities)
17
18 print("... Selected Range of columns ...")
19 print(head(range_cols, 3))
20
21
22 # Method C: select using helper functions (starts_with)
23 # Scenario: Select all columns that start with the letter "p"
24 # Example: Parental_Involvement, Previous_Scores, Physical_Activity, etc.
25 starts_with_p <- student %>%
26   select(starts_with("p"))
27
28 print("... Selected columns starting with 'P' ...")
```

Console

NIFTY +0.17% Search

11:47:57 AM 01-12-2025

Name :- Priya Gupta
Roll no. :- S081

Sheth L.U.J. & Sir M.V. College

The screenshot shows the RStudio interface with the following details:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Project:** Project (None).
- Source Editor:** Contains R code for dataset manipulation, including methods for selecting columns starting with 'P', dropping variables, and dropping ranges of columns.
- Environment View:** Shows a list of objects in the global environment, such as appended_data, data_feb, data_jan, data_new_hir, data1, data2, dropped_mult, dropped_one, dropped_range, final_list, full_merge, inner_merge, left_merge, merged_data, range_cols, right_merge, selected_cols, starts_with_p, student, StudentPerfo..., study_perfor..., and common_cols.
- Console:** Displays the R command history and output.
- System Tray:** Shows battery level (100%), signal strength, and system date/time (11:48:46 AM 01-12-2025).

The screenshot shows the RStudio interface with the following details:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Project:** Project (None).
- Source Editor:** Contains R code for dataset manipulation, including importing a CSV file, printing the first 3 rows, and displaying the head of the student dataset.
- Environment View:** Shows a list of objects in the global environment, including append, data_1, data_2, data_3, data_4, data1, data2, dropp, dropp_, final, full, inner, left, merge, range, right, selec, start, stude, study, and common_.
- Console:** Displays the R command history and output.
- System Tray:** Shows battery level (+0.17%), signal strength, and system date/time (11:49:13 AM 01-12-2025).

Name :- Priya Gupta
Roll no. :- S081

Sheth L.U.J. & Sir M.V. College

The screenshot shows an RStudio interface with a script editor containing R code for dataset manipulation. The code demonstrates three methods to drop specific columns from a dataset:

- Method A:** Drop a single specific column ('Gender').
- Method B:** Drop multiple columns at once ('School_Type' and 'Peer_Influence').
- Method C:** Drop a range of columns ('Sleep_Hours' to 'Family_Income').

The script also includes print statements to show the datasets before and after each modification.

```
> # -----
> # Method A: Drop a single specific column
> # Scenario: Remove the column 'Gender'
> dropped_one <- student %>
+ select(-Gender)
>
> print("... Dataset with 'Gender' dropped ---")
[1] "... Dataset with 'Gender' dropped ..."
> print(names(dropped_one)) # checking if it is removed
[1] "Hours_Studied"           "Attendance"          "Parental_Involvement" "Access_to_Resources"
[2] "Sleep_Hours"              "Previous_Scores"     "Motivation_Level"    "Internet_Access"
[3] "Family_Income"            "Teacher_Quality"    "School_Type"         "Peer_Influence"
[4] "Learning_Disabilities"   "Parental_Education_Level" "Distance_from_Home" "Exam_Score"
>
> # Method B: Drop multiple columns at once
> # Scenario: Remove 'School_Type' and 'Peer_Influence'
> dropped_multiple <- student %>
+ select(-School_Type, -Peer_Influence)
>
> print("... Dataset with 'School_Type' and 'Peer_Influence' dropped ---")
[1] "... Dataset with 'School_Type' and 'Peer_Influence' dropped ..."
> print(names(dropped_multiple))
[1] "Hours_Studied"           "Attendance"          "Parental_Involvement" "Access_to_Resources"
[2] "Sleep_Hours"              "Previous_Scores"     "Motivation_Level"    "Internet_Access"
[3] "Family_Income"            "Teacher_Quality"    "Physical_Activity"   "Learning_Disabilities"
[4] "Gender"                  "Distance_from_Home" "Exam_Score"
>
> # Method C: Drop a range of columns
> # Scenario: Remove all columns from Sleep_Hours to Family_Income
> dropped_range <- student %>
+ select(-(Sleep_Hours:Family_Income))
>
> print("... Dataset with range 'Sleep_Hours' to 'Family_Income' dropped ---")
[1] "... Dataset with range 'Sleep_Hours' to 'Family_Income' dropped ..."
> print(names(dropped_range))
[1] "Hours_Studied"           "Attendance"          "Parental_Involvement" "Access_to_Resources"
[2] "Teacher_Quality"          "School_Type"         "Peer_Influence"       "Physical_Activity"
[3] "Parental_Education_Level" "Distance_from_Home" "Gender"             "Exam_Score"
```

Name :- Priya Gupta
Roll no. :- S081