

SHETH L.U.J. AND SIR M.V. COLLEGE

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

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

Source
Console Terminal Background Jobs
R - R4.5.2 - ~/ -

> library(dplyr)
> library(readr)
> attendance <- read_csv("Attendance-Prediction.csv")
Rows: 20000 Columns: 15
Column specification
Delimiter: ","
chr (9): gender, course, year, parent_education, internet_access, hostel_resident, class_ty...
dbl (6): student_id, age, study_hours, sleep_hours, travel_time_minutes, attendance

i Use 'spec()' to retrieve the full column specification for this data.
i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
> print("--- Original Dataset (First 3 rows) ---")
[1] "--- Original Dataset (First 3 rows) ---"
> print(head(attendance, 3))
# A tibble: 3 x 15
  student_id age gender course year parent_education internet_access hostel_resident
  <dbl> <dbl> <chr> <chr> <chr> <chr> <chr> <chr>
1 201. 17 male pharmacy 1st year high school no yes
2 201. 23 other diploma 2nd year high school yes yes
3 201. 22 male b.sc 3rd year no formal yes yes
# i 7 more variables: class_type <chr>, weather <chr>, study_hours <dbl>, sleep_hours <dbl>,
# travel_time_minutes <dbl>, attendance <dbl>, absence_reason <chr>
> selected_cols <- attendance %>%
+ select(student_id, age, attendance, study_hours)
> print("--- Selected Specific Columns ---")
[1] "--- Selected Specific Columns ---"
> print(head(selected_cols, 3))
# A tibble: 3 x 4
  student_id age attendance study_hours
  <dbl> <dbl> <dbl> <dbl>
1 201. 17 1 3.63
2 201. 23 0 4.39
3 201. 22 1 5.43
> range_cols <- attendance %>%
+ select(age:study_hours)
> print("--- Selected Range of Columns (age to study_hours) ---")
[1] "--- Selected Range of Columns (age to study_hours) ---"
> print(head(range_cols, 3))
# A tibble: 3 x 4
  age gender course year parent_education internet_access hostel_resident
  <dbl> <chr> <chr> <chr> <chr> <chr> <chr>
1 17 male pharmacy 1st year high school no yes
2 23 other diploma 2nd year high school yes yes
3 22 male b.sc 3rd year no formal yes yes
# i 1 more variable: study_hours <dbl>
> starts_with_s <- attendance %>%
+ select(starts_with("s"))
> print("--- Selected Columns Starting with 'S' ---")
[1] "--- Selected Columns Starting with 'S' ---"
> print(names(dropped_one))
Error: object 'dropped_one' not found
Show Traceback
Rerun with Debug

> print(head(starts_with_s, 3))
# A tibble: 3 x 3
  student_id study_hours sleep_hours
  <dbl> <dbl> <dbl>
1 201. 3.63 7.24
2 201. 4.39 7.98
3 201. 5.43 6.53
> dropped_one <- attendance %>%
+ select(-gender)
> print("--- Dataset with 'gender' Dropped ---")
[1] "--- Dataset with 'gender' Dropped ---"
> dropped_multiple <- attendance %>%
+ select(-attendance, -study_hours)
> print("--- Dataset with Multiple Columns Dropped ---")
[1] "--- Dataset with Multiple Columns Dropped ---"
> print(names(dropped_multiple))
[1] "student_id" "age" "parent_education" "internet_access" "course" "hostel_resident"
[5] "year"
```

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

Source
Console Terminal Background Jobs
R - R4.5.2 - ~/ -

> range_cols <- attendance %>%
+ select(age:study_hours)
> print("--- Selected Range of Columns (age to study_hours) ---")
[1] "--- Selected Range of Columns (age to study_hours) ---"
> print(head(range_cols, 3))
# A tibble: 3 x 10
  age gender course year parent_education internet_access hostel_resident class_type weather
  <dbl> <chr> <chr> <chr> <chr> <chr> <chr> <chr>
1 17 male pharm_1st high school no yes offline sunny
2 23 other diplo_2nd high school yes yes offline cloudy
3 22 male b.sc 3rd no formal yes yes offline rainy
# i 1 more variable: study_hours <dbl>
> starts_with_s <- attendance %>%
+ select(starts_with("s"))
> print("--- Selected Columns Starting with 'S' ---")
[1] "--- Selected Columns Starting with 'S' ---"
> print(names(dropped_one))
Error: object 'dropped_one' not found
Show Traceback
Rerun with Debug

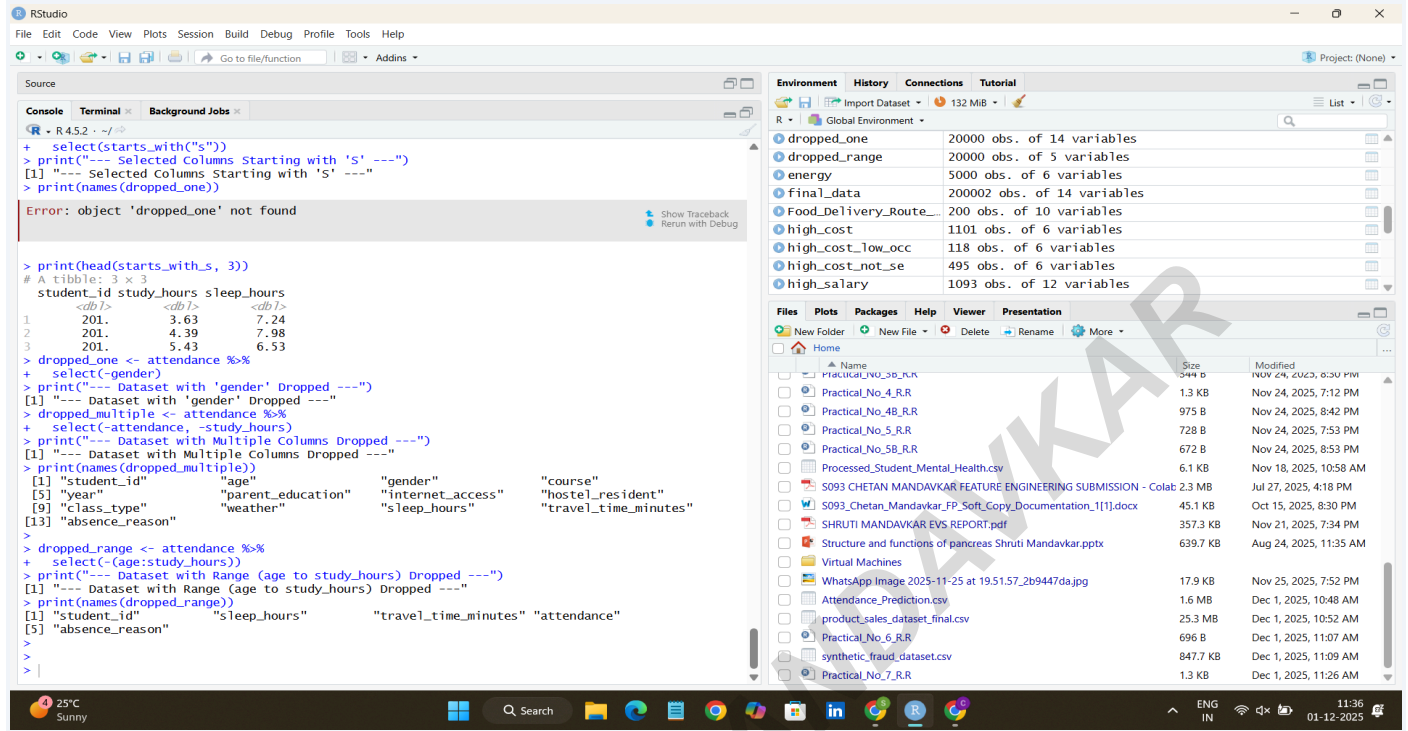
> print(head(starts_with_s, 3))
# A tibble: 3 x 3
  student_id study_hours sleep_hours
  <dbl> <dbl> <dbl>
1 201. 3.63 7.24
2 201. 4.39 7.98
3 201. 5.43 6.53
> dropped_one <- attendance %>%
+ select(-gender)
> print("--- Dataset with 'gender' Dropped ---")
[1] "--- Dataset with 'gender' Dropped ---"
> dropped_multiple <- attendance %>%
+ select(-attendance, -study_hours)
> print("--- Dataset with Multiple Columns Dropped ---")
[1] "--- Dataset with Multiple Columns Dropped ---"
> print(names(dropped_multiple))
[1] "student_id" "age" "parent_education" "internet_access" "course" "hostel_resident"
[5] "year"
```

NAME:- CHETAN MANDAVKAR

ROLL NO. S093

SUBJECT:- Data Analysis with SAS / SPSS / R

SHETH L.U.J. AND SIR M.V. COLLEGE



NAME:- CHETAN MANDAVKAR

ROLL NO. S093

SUBJECT:- Data Analysis with SAS / SPSS /R