

MVLU COLLEGE.

PRACTICAL NO :- 04

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

```
CODE —install.packages("dplyr")
```

```
library(dplyr)
```

```
library(readr) # For efficient reading
```

```
student <- read_csv("C:/Users/itlab/Downloads/S100/StudentPerformanceFactors.csv")
```

```
head(student)
```

```
# Example 1: Single Condition (Pipe Operator |>)
```

```
# Filter students with exam score above 80
```

```
high_exam_subset <- subset(student, Exam_Score > 80)
```

```
cat("Number of students with exam score > 80:", nrow(high_exam_subset), "\n")
```

```
head(high_exam_subset)
```

```
# Example 2: Multiple Conditions (AND)
```

```
# Students who studied > 20 hours AND had attendance > 90%
```

```
high_study_high_attendance <- subset(student,  
                                     Hours_Studied > 20 & Attendance > 90)
```

```
cat("High study + high attendance:",  
    nrow(high_study_high_attendance), "\n")
```

```
head(high_study_high_attendance)
```

```
# Example 3: Multiple Conditions (OR)
```

```
# Students who sleep more than 8 hours OR do extracurricular activities
```

```
sleep_or_extracurricular <- subset(student,  
                                   Sleep_Hours > 8 | Extracurricular_Activities == "Yes")
```

```
cat("Sleep > 8 hours OR extracurricular:",  
    nrow(sleep_or_extracurricular), "\n")
```

```
head(sleep_or_extracurricular)
```

```
# METHOD 2: Using dplyr::filter()
```

```
# -----
```

NANDINI PANDIT S100

Data Analysis with SAS / SPSS /R Practical

MVLU COLLEGE.

Example 1: Single Condition (Pipe Operator)

Students scoring above 85 marks in exams

```
high_exam_score <- student |>
  filter(Exam_Score > 80)
cat("Students with exam score > 80:",
    nrow(high_exam_score), "\n")
head(high_exam_score)
```

Example 2: Multiple Conditions (AND using commas)

Students with low sleep (<6 hours) AND low motivation

```
low_sleep_low_motivation <- student |>
  filter(Sleep_Hours < 8, Motivation_Level == "Low")

cat("Low sleep + low motivation:",
    nrow(low_sleep_low_motivation), "\n")
head(low_sleep_low_motivation)
```

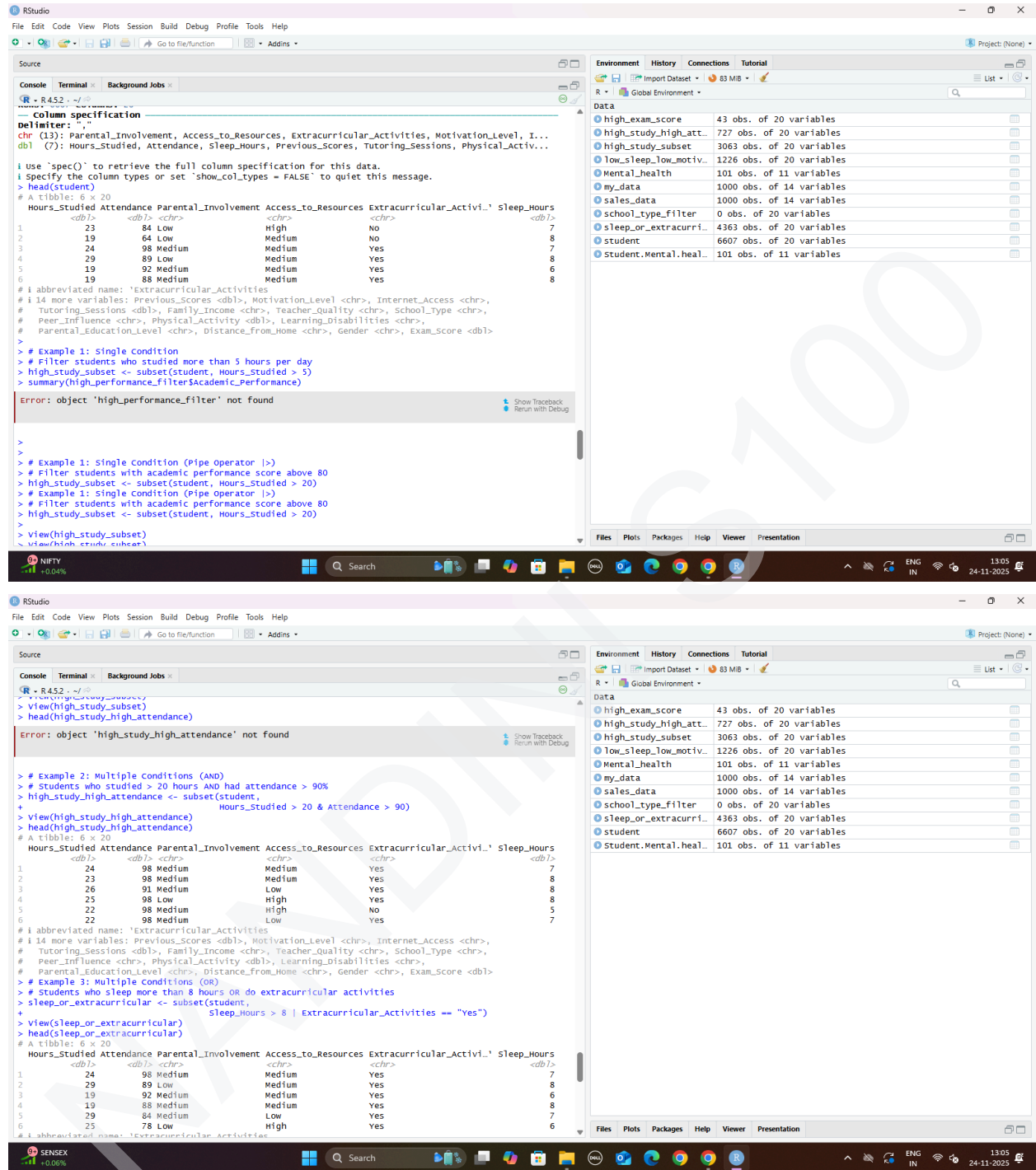
Example 3: Using %in% operator

Filter by school type: Public or Private

```
school_type_filter <- student |>
  filter(School_Type %in% c("Public_School", "Private_School"))

cat("School type Public or Private:",
    nrow(school_type_filter), "\n")
table(school_type_filter$School_Type)
```


MVLU COLLEGE.



The image displays two screenshots of the RStudio interface, showing data manipulation tasks in R.

Top Screenshot:

- Source:** Shows a script with a column specification for a data frame named 'student'. The columns are: Hours_Studied, Attendance, Parental_Involvement, Access_to_Resources, Extracurricular_Activities, Sleep_Hours, and Exam_Score.
- Console:** Displays the output of the column specification and a subset operation. The subset operation is: `high_performance_filter <- subset(student, Hours_Studied > 5)`. The output shows the first 6 rows of the subsetted data.
- Environment:** Lists the objects in the environment, including 'high_exam_score', 'high_study_high_att...', 'high_study_subset', 'low_sleep_low_motiv...', 'Mental_health', 'my_data', 'sales_data', 'school_type_filter', 'sleep_or_extracurri...', 'student', and 'Student.Mental.heal...'. Each object is associated with a specific number of observations and variables.

Bottom Screenshot:

- Source:** Shows a script with a subset operation based on multiple conditions (AND): `high_study_high_attendance <- subset(student, Hours_Studied > 20 & Attendance > 90)`.
- Console:** Displays the output of the subset operation. The output shows the first 6 rows of the subsetted data.
- Environment:** Lists the objects in the environment, including 'high_exam_score', 'high_study_high_att...', 'high_study_subset', 'low_sleep_low_motiv...', 'Mental_health', 'my_data', 'sales_data', 'school_type_filter', 'sleep_or_extracurri...', 'student', and 'Student.Mental.heal...'. Each object is associated with a specific number of observations and variables.

NANDINI PANDIT S100
Data Analysis with SAS / SPSS /R Practical

MVLU COLLEGE.

The image displays two screenshots of the RStudio environment, illustrating data manipulation tasks in R.

Top Screenshot: The console shows the execution of R code. The first example filters students with an exam score greater than 80. The second example filters students with low sleep (less than 8 hours) and low motivation, resulting in a tibble with 6 rows and 20 columns. The Environment pane on the right lists the objects created, including 'high_exam_score', 'low_sleep_low_motiv...', 'my_data', 'sales_data', 'school_type_filter', 'sleep_or_extracurri...', 'student', and 'Student.Mental.heal...'.

Bottom Screenshot: The console shows an attempt to filter by school type (Public or Private) using the 'filter()' function. This results in an error: "Error: object 'Private' not found". The Environment pane on the right shows the same objects as the top screenshot, plus 'school_type_filter'.

NANDINI PANDIT S100

Data Analysis with SAS / SPSS /R Practical

MVLU COLLEGE.

RStudio interface showing a data table with 19 rows and 8 columns. The columns are: Hours_Studied, Attendance, Parental_Involvement, Access_to_Resources, Extracurricular_Activities, Sleep_Hours, and Previous_Scores. The Environment pane on the right lists various data objects: high_exam_score (43 obs. of 20 variables), high_study_high_att... (727 obs. of 20 variables), high_study_subset (3063 obs. of 20 variables), low_sleep_low_motiv... (1226 obs. of 20 variables), Mental_health (101 obs. of 11 variables), my_data (1000 obs. of 14 variables), sales_data (1000 obs. of 14 variables), school_type_filter (0 obs. of 20 variables), sleep_or_extracurri... (4363 obs. of 20 variables), student (6607 obs. of 20 variables), and Student.Mental.heal... (101 obs. of 11 variables). The Console pane shows R code for filtering data by school type and viewing exam scores.

RStudio interface showing a data table with 19 rows and 8 columns. The columns are: Hours_Studied, Attendance, Parental_Involvement, Access_to_Resources, Extracurricular_Activities, Sleep_Hours, and Previous_Scores. The Environment pane on the right lists various data objects: high_exam_score (43 obs. of 20 variables), high_study_high_att... (727 obs. of 20 variables), high_study_subset (3063 obs. of 20 variables), low_sleep_low_motiv... (1226 obs. of 20 variables), Mental_health (101 obs. of 11 variables), my_data (1000 obs. of 14 variables), sales_data (1000 obs. of 14 variables), school_type_filter (0 obs. of 20 variables), sleep_or_extracurri... (4363 obs. of 20 variables), student (6607 obs. of 20 variables), and Student.Mental.heal... (101 obs. of 11 variables). The Console pane shows R code for filtering data by school type and viewing exam scores.

NANDINI PANDIT S100
Data Analysis with SAS / SPSS /R Practical

MVLU COLLEGE.

The screenshot displays the RStudio environment with a data table, an environment pane, and a console window.

Data Table:

Type	Peer_Influence	Physical_Activity	Learning_Disabilities	Parental_Education_Level	Distance_from_Home	Gender	Exam_Score
Neutral		4	No	Postgraduate	Near	Male	74
Negative		4	No	High School	Moderate	Male	71
Neutral		4	No	College	Near	Female	70
Positive		3	No	Postgraduate	Near	Male	71
Neutral		2	No	High School	Moderate	Male	67
Negative		2	No	High School	Far	Male	66
Positive		5	No	High School	Moderate	Male	72
Neutral		2	No	High School	Near	Male	71
Positive		4	No	High School	Near	Male	70
Positive		3	No	Postgraduate	Near	Male	66
Neutral		4	No	Postgraduate	Near	Male	65
Positive		3	No	College	Near	Male	60
Neutral		3	No	High School	Near	Female	65
Positive		2	No	College	Moderate	Female	67
Positive		4	No	College	Near	Male	66
Neutral		3	No	High School	Moderate	Male	66
Positive		3	No	High School	Near	Female	63
Positive		3	No	Postgraduate	Far	Male	64
Neutral		2	No	College	Moderate	Male	71

Environment Pane:

- high_exam_score: 43 obs. of 20 variables
- high_study_high_att...: 727 obs. of 20 variables
- high_study_subset: 3063 obs. of 20 variables
- low_sleep_low_motiv...: 1226 obs. of 20 variables
- Mental_health: 101 obs. of 11 variables
- my_data: 1000 obs. of 14 variables
- sales_data: 1000 obs. of 14 variables
- school_type_filter: 0 obs. of 20 variables
- sleep_or_extracurri...: 4363 obs. of 20 variables
- student: 6607 obs. of 20 variables
- Student.Mental.heal...: 101 obs. of 11 variables

Console:

```
> View(school_type_filter)
> View(school_type_filter)
> View(school_type_filter)
> table(school_type_filter$School_Type)
< table of extent 0 >
> View(high_exam_score)
> View(high_exam_score)
> |
```

NANDINI PANDIT S100

Data Analysis with SAS / SPSS /R Practical

MVLU COLLEGE.

The image displays two screenshots of the RStudio interface, showing data analysis steps.

Top Screenshot:

- Environment:** Lists data objects including `high_exam_score` (43 obs. of 20 variables), `high_study_high_att...` (727 obs. of 20 variables), `high_study_subset` (3063 obs. of 20 variables), `low_sleep_low_motiv...` (1226 obs. of 20 variables), `Mental_health` (101 obs. of 11 variables), `my_data` (1000 obs. of 14 variables), `sales_data` (1000 obs. of 14 variables), `school_type_filter` (0 obs. of 20 variables), `sleep_or_extracurri...` (4363 obs. of 20 variables), `student` (6607 obs. of 20 variables), and `Student.Mental.heal...` (101 obs. of 11 variables).
- Console:** Shows R commands: `view(school_type_filter)`, `view(school_type_filter)`, `view(school_type_filter)`, `table(school_type_filter$School_Type)`, `table of extent 0 >`, `view(high_exam_score)`, and `view(high_exam_score)`.

Bottom Screenshot:

- Environment:** Lists data objects including `high_exam_score` (43 obs. of 20 variables), `high_study_high_att...` (727 obs. of 20 variables), `high_study_subset` (3063 obs. of 20 variables), `low_sleep_low_motiv...` (1226 obs. of 20 variables), `Mental_health` (101 obs. of 11 variables), `my_data` (1000 obs. of 14 variables), `sales_data` (1000 obs. of 14 variables), `school_type_filter` (0 obs. of 20 variables), `sleep_or_extracurri...` (4363 obs. of 20 variables), `student` (6607 obs. of 20 variables), and `Student.Mental.heal...` (101 obs. of 11 variables).
- Console:** Shows R commands: `view(school_type_filter)`, `view(school_type_filter)`, `view(school_type_filter)`, `table(school_type_filter$School_Type)`, `table of extent 0 >`, `view(high_exam_score)`, and `view(high_exam_score)`.

NANDINI PANDIT S100

Data Analysis with SAS / SPSS /R Practical