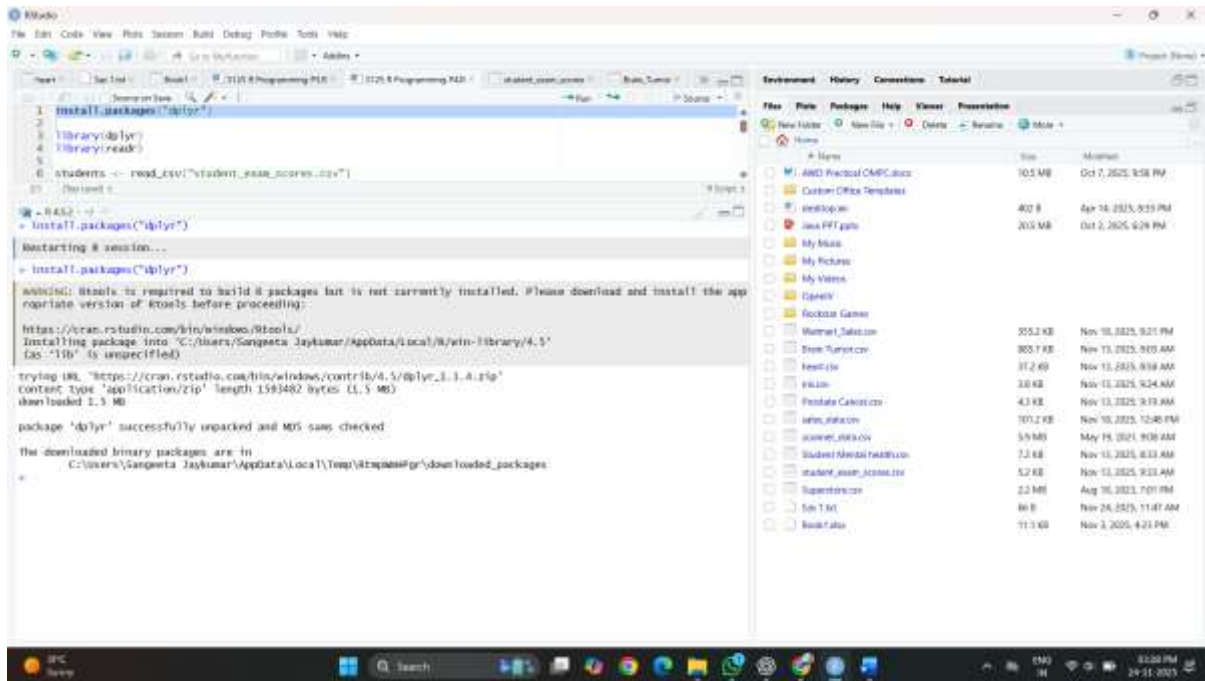


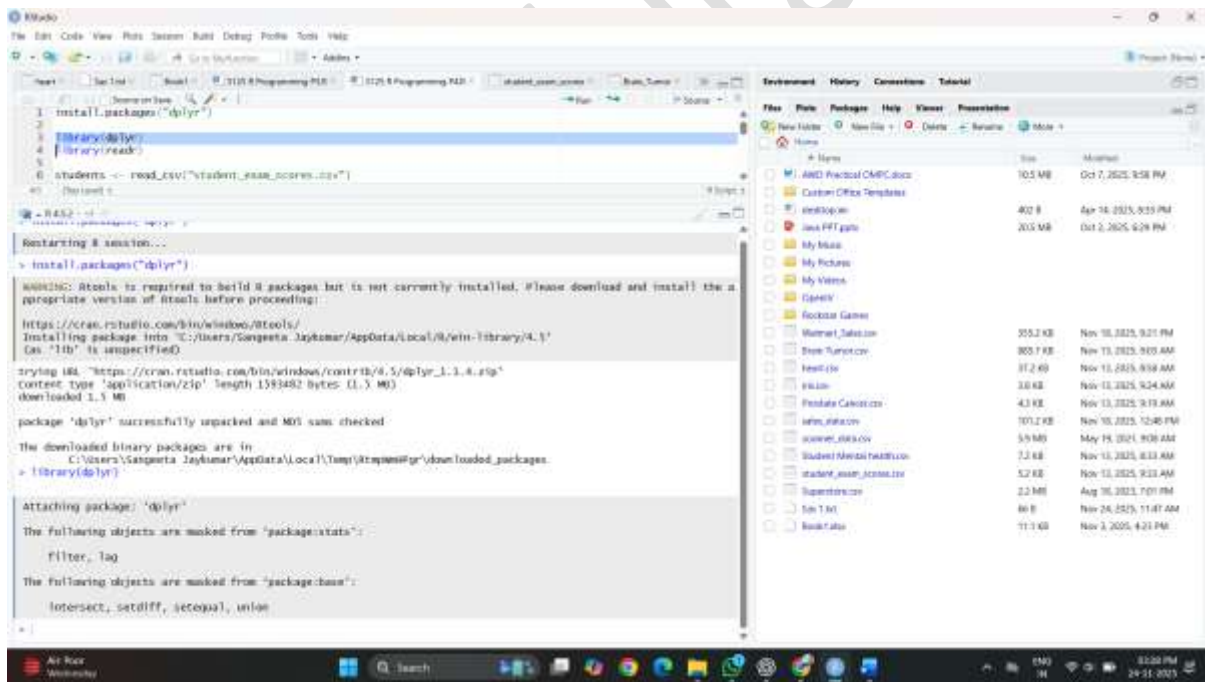
Sheth L.U.J & Sir M.V College  
SAS/SPSS/R  
Practical No.4

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

Output:



```
1 install.packages("dplyr")
2
3 library(dplyr)
4 library(readr)
5
6 students <- read_csv("student_exam_scores.csv")
7
8 # R6.5.2 ->
9 > install.packages("dplyr")
Restarting R session...
> install.packages("dplyr")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:
https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/Sangeeta Jaykumar/AppData/Local/R/win-library/4.5'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/dplyr_1.3.4.zip'
content type 'application/zip' length 1593482 bytes (1.5 MB)
downloaded 1.5 MB
package 'dplyr' successfully unpacked and MD5 sums checked
The downloaded binary packages are in
C:/Users/Sangeeta Jaykumar/AppData/Local/Temp/Rtmp8t8fgr/downloaded_packages
```



```
1 install.packages("dplyr")
2
3 library(dplyr)
4 library(readr)
5
6 students <- read_csv("student_exam_scores.csv")
7
8 # R6.5.2 ->
9 > library(dplyr)
Restarting R session...
> library(dplyr)
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:
https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/Sangeeta Jaykumar/AppData/Local/R/win-library/4.5'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/dplyr_1.3.4.zip'
content type 'application/zip' length 1593482 bytes (1.5 MB)
downloaded 1.5 MB
package 'dplyr' successfully unpacked and MD5 sums checked
The downloaded binary packages are in
C:/Users/Sangeeta Jaykumar/AppData/Local/Temp/Rtmp8t8fgr/downloaded_packages.
> library(dplyr)
Attaching package: 'dplyr'
The following objects are masked from 'package:stats':
  filter, lag
The following objects are masked from 'package:base':
  intersect, setdiff, setequal, union
```

Sheth L.U.J & Sir M.V College  
SAS/SPSS/R  
Practical No.4

The screenshot shows the RStudio interface with the following code in the script editor:

```
1 install.packages("dplyr")
2
3 library(dplyr)
4 library(readr)
5
6 students <- read_csv("student_exam_scores.csv")
7
8 # R432 ->
9
10 # Restarting & session...
11
12 # install.packages("dplyr")
13
14 ##[INFO] Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:
15 ##[INFO] https://cran.rstudio.com/bin/windows/Rtools/
16 ##[INFO] Installing package into 'C:/Users/Gangeta Jaykumar/AppData/Local/R/win-library/4.5'
17 ##[INFO] (as 'lib' is unspecified)
18
19 ##[INFO] trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/dplyr_1.3.4.zip'
20 ##[INFO] content type 'application/zip' length 139482 bytes (13.5 MB)
21 ##[INFO] downloaded 1.3 MB
22
23 ##[INFO] package 'dplyr' successfully unpacked and MD5 sums checked
24
25 ##[INFO] the downloaded binary packages are in
26 ##[INFO] C:/Users/Gangeta Jaykumar/AppData/Local/Temp/Rtmpw@pr/downloaded_packages
27
28 # library(dplyr)
29
30 ##[INFO] attaching package: 'dplyr'
31
32 ##[INFO] The following objects are masked from 'package:stats':
33 ##[INFO] filter, lag
34
35 ##[INFO] The following objects are masked from 'package:base':
36 ##[INFO] intersect, setdiff, setequal, union
37
38 # library(readr)
39
40 #
```

The Environment pane on the right shows the loaded packages: dplyr (1.3.4), readr (2.1.5), and R6 (2.5.1).

The screenshot shows the RStudio interface with the following code in the script editor:

```
1 # R432 ->
2
3 # Restarting & session...
4
5 # install.packages("dplyr")
6
7 ##[INFO] Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:
8 ##[INFO] https://cran.rstudio.com/bin/windows/Rtools/
9 ##[INFO] Installing package into 'C:/Users/Gangeta Jaykumar/AppData/Local/R/win-library/4.5'
10 ##[INFO] (as 'lib' is unspecified)
11
12 ##[INFO] trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/dplyr_1.3.4.zip'
13 ##[INFO] content type 'application/zip' length 139482 bytes (13.5 MB)
14 ##[INFO] downloaded 1.3 MB
15
16 ##[INFO] package 'dplyr' successfully unpacked and MD5 sums checked
17
18 ##[INFO] the downloaded binary packages are in
19 ##[INFO] C:/Users/Gangeta Jaykumar/AppData/Local/Temp/Rtmpw@pr/downloaded_packages
20
21 # library(dplyr)
22
23 ##[INFO] attaching package: 'dplyr'
24
25 ##[INFO] The following objects are masked from 'package:stats':
26 ##[INFO] filter, lag
27
28 ##[INFO] The following objects are masked from 'package:base':
29 ##[INFO] intersect, setdiff, setequal, union
30
31 # library(readr)
32
33 #
34
35 # head(students)
36
37 # high_hours <- subset(students, hours_studied > 7)
38
39 # cat("Students studying > 7 hours:", nrow(high_hours), "\n")
40
41 # summary(high_hours$hours_studied)
42
43 # high_attendance_high_score <- subset(students, attendance_percent > 80 & exam_score > 35)
44
45 # cat("High attendance & high score:", nrow(high_attendance_high_score), "\n")
46
47 # head(high_attendance_high_score)
48
49 #
50
51 # special_students <- subset(students, sleep_hours < 6 | previous_scores > 80)
52
53 # cat("Special Filter:", nrow(special_students), "\n")
54
55 # head(special_students)
56
57 #
58
59 # low_sleep_filter <- students %>%
60 #   filter(sleep_hours < 6)
61
62 #
63
64 # library(readr)
65
66 # students <- read_csv("student_exam_scores.csv")
67
68 # rows: 500 Columns: 6
69 # Column specification
70 # Delimiter: ","
71 # chr (1): student_id
72 # dbl (3): hours_studied, sleep_hours, attendance_percent
73 # num (2): previous_scores, exam_score
74
75 # Use 'spec()' to retrieve the full column specification for this data.
76 # Specify the column types or set 'show_col_types = FALSE' to quiet this message.
77 # head(students)
78
79 # # A tibble: 5 x 6
80 #   student_id hours_studied sleep_hours attendance_percent previous_scores exam_score
81 #   <chr>         <dbl>         <dbl>         <dbl>         <dbl>         <dbl>
82 # 1 2001           8           8.8           72.1          45          30.2
83 # 2 2002           1.1           8.6           80.7          55          25
84 # 3 2003           4           8.2           71.7          60          35.8
85 # 4 2004           3.5           4.8           95.1          60          34
86 # 5 2005           9.1           6.4           88.8          71          40.5
87 # 6 2006           8.4           5.1           18.5          75          35.7
88
89 # high_hours <- subset(students, hours_studied > 7)
90
91 #
```

The Environment pane on the right shows the loaded packages: dplyr (1.3.4), readr (2.1.5), and R6 (2.5.1).

[illegible]

The screenshot shows the RStudio interface. The main editor contains R code for data manipulation. The code defines a data frame 'students' with columns 'previous\_scores', 'low\_sleep', and 'hours\_studied'. It then filters the data to show students with 'previous\_scores' between 50 and 60, and 'low\_sleep' greater than 5. The code also calculates the 'exam\_score' for each student based on their 'previous\_scores' and 'hours\_studied'. The final output is a table of the filtered data.

```

11 selected_scores <- students[, ]
12 filter(previous_scores >= 50, <= 60, 70))
13 cat("Previous scores 50, 60, 70:", nrow(selected_scores), "\n")
14 table(selected_scores$previous_scores)
15
16 #> # A tibble: 8 x 8
17   student_id hours_studied sleep_hours attendance_percent previous_scores exam_score
18   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
19 1 50007 9.1 6.4 88.0 71 40.3
20 2 50012 8.4 7.9 87.0 85 35.1
21 3 50019 9.9 4.8 92.5 54 35.6
22 4 50029 10.1 4.8 87.5 73 37.2
23 5 50030 7.8 8.8 88.5 62 38.2
24 6 50034 11.7 7.4 87.9 61 42.7
25
26 #> selected_scores <- students[, ]
27 #> filter(previous_scores >= 50, <= 60, 70))
28 #> cat("Previous scores 50, 60, 70:", nrow(selected_scores), "\n")
29 #> previous_scores 50, 60, 70: 11
30 #> table(selected_scores$previous_scores)
31
32 #> # A tibble: 8 x 8
33   student_id hours_studied sleep_hours attendance_percent previous_scores exam_score
34   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
35 1 50007 9.1 6.4 88.0 71 40.3
36 2 50012 8.4 7.9 87.0 85 35.1
37 3 50019 9.9 4.8 92.5 54 35.6
38 4 50029 10.1 4.8 87.5 73 37.2
39 5 50030 7.8 8.8 88.5 62 38.2
40 6 50034 11.7 7.4 87.9 61 42.7
41
42 #> selected_scores <- students[, ]
43 #> filter(previous_scores >= 50, <= 60, 70))
44 #> cat("Previous scores 50, 60, 70:", nrow(selected_scores), "\n")
45 #> previous_scores 50, 60, 70: 11
46 #> table(selected_scores$previous_scores)
47
48 #> # A tibble: 8 x 8
49   student_id hours_studied sleep_hours attendance_percent previous_scores exam_score
50   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
51 1 50007 9.1 6.4 88.0 71 40.3
52 2 50012 8.4 7.9 87.0 85 35.1
53 3 50019 9.9 4.8 92.5 54 35.6
54 4 50029 10.1 4.8 87.5 73 37.2
55 5 50030 7.8 8.8 88.5 62 38.2
56 6 50034 11.7 7.4 87.9 61 42.7
57
58 #> selected_scores <- students[, ]
59 #> filter(previous_scores >= 50, <= 60, 70))
60 #> cat("Previous scores 50, 60, 70:", nrow(selected_scores), "\n")
61 #> previous_scores 50, 60, 70: 11
62 #> table(selected_scores$previous_scores)
63
64 #> # A tibble: 8 x 8
65   student_id hours_studied sleep_hours attendance_percent previous_scores exam_score
66   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
67 1 50007 9.1 6.4 88.0 71 40.3
68 2 50012 8.4 7.9 87.0 85 35.1
69 3 50019 9.9 4.8 92.5 54 35.6
70 4 50029 10.1 4.8 87.5 73 37.2
71 5 50030 7.8 8.8 88.5 62 38.2
72 6 50034 11.7 7.4 87.9 61 42.7
73
74 #> selected_scores <- students[, ]
75 #> filter(previous_scores >= 50, <= 60, 70))
76 #> cat("Previous scores 50, 60, 70:", nrow(selected_scores), "\n")
77 #> previous_scores 50, 60, 70: 11
78 #> table(selected_scores$previous_scores)
79
80 #> # A tibble: 8 x 8
81   student_id hours_studied sleep_hours attendance_percent previous_scores exam_score
82   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
83 1 50007 9.1 6.4 88.0 71 40.3
84 2 50012 8.4 7.9 87.0 85 35.1
85 3 50019 9.9 4.8 92.5 54 35.6
86 4 50029 10.1 4.8 87.5 73 37.2
87 5 50030 7.8 8.8 88.5 62 38.2
88 6 50034 11.7 7.4 87.9 61 42.7
89
90 #> selected_scores <- students[, ]
91 #> filter(previous_scores >= 50, <= 60, 70))
92 #> cat("Previous scores 50, 60, 70:", nrow(selected_scores), "\n")
93 #> previous_scores 50, 60, 70: 11
94 #> table(selected_scores$previous_scores)
95
96 #> # A tibble: 8 x 8
97   student_id hours_studied sleep_hours attendance_percent previous_scores exam_score
98   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
99 1 50007 9.1 6.4 88.0 71 40.3
100 2 50012 8.4 7.9 87.0 85 35.1
101 3 50019 9.9 4.8 92.5 54 35.6
102 4 50029 10.1 4.8 87.5 73 37.2
103 5 50030 7.8 8.8 88.5 62 38.2
104 6 50034 11.7 7.4 87.9 61 42.7
105
106 #> selected_scores <- students[, ]
107 #> filter(previous_scores >= 50, <= 60, 70))
108 #> cat("Previous scores 50, 60, 70:", nrow(selected_scores), "\n")
109 #> previous_scores 50, 60, 70: 11
110 #> table(selected_scores$previous_scores)
111
112 #> # A tibble: 8 x 8
113   student_id hours_studied sleep_hours attendance_percent previous_scores exam_score
114   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
115 1 50007 9.1 6.4 88.0 71 40.3
116 2 50012 8.4 7.9 87.0 85 35.1
117 3 50019 9.9 4.8 92.5 54 35.6
118 4 50029 10.1 4.8 87.5 73 37.2
119 5 50030 7.8 8.8 88.5 62 38.2
120 6 50034 11.7 7.4 87.9 61 42.7
121
122 #> selected_scores <- students[, ]
123 #> filter(previous_scores >= 50, <= 60, 70))
124 #> cat("Previous scores 50, 60, 70:", nrow(selected_scores), "\n")
125 #> previous_scores 50, 60, 70: 11
126 #> table(selected_scores$previous_scores)
127
128 #> # A tibble: 8 x 8
129   student_id hours_studied sleep_hours attendance_percent previous_scores exam_score
130   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
131 1 50007 9.1 6.4 88.0 71 40.3
132 2 50012 8.4 7.9 87.0 85 35.1
133 3 50019 9.9 4.8 92.5 54 35.6
134 4 50029 10.1 4.8 87.5 73 37.2
135 5 50030 7.8 8.8 88.5 62 38.2
136 6 50034 11.7 7.4 87.9 61 42.7
137
138 #> selected_scores <- students[, ]
139 #> filter(previous_scores >= 50, <= 60, 70))
140 #> cat("Previous scores 50, 60, 70:", nrow(selected_scores), "\n")
141 #> previous_scores 50, 60, 70: 11
142 #> table(selected_scores$previous_scores)
143
144 #> # A tibble: 8 x 8
145   student_id hours_studied sleep_hours attendance_percent previous_scores exam_score
146   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
147 1 50007 9.1 6.4 88.0 71 40.3
148 2 50012 8.4 7.9 87.0 85 35.1
149 3 50019 9.9 4.8 92.5 54 35.6
150 4 50029 10.1 4.8 87.5 73 37.2
151 5 50030 7.8 8.8 88.5 62 38.2
152 6 50034 11.7 7.4 87.9 61 42.7
153
154 #> selected_scores <- students[, ]
155 #> filter(previous_scores >= 50, <= 60, 70))
156 #> cat("Previous scores 50, 60, 70:", nrow(selected_scores), "\n")
157 #> previous_scores 50, 60, 70: 11
158 #> table(selected_scores$previous_scores)
159
160 #> # A tibble: 8 x 8
161   student_id hours_studied sleep_hours attendance_percent previous_scores exam_score
162   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
163 1 50007 9.1 6.4 88.0 71 40.3
164 2 50012 8.4 7.9 87.0 85 35.1
165 3 50019 9.9 4.8 92.5 54 35.6
166 4 50029 10.1 4.8 87.5 73 37.2
167 5 50030 7.8 8.8 88.5 62 38.2
168 6 50034 11.7 7.4 87.9 61 42.7
169
170 #&gt
```