

SHETH L.U.J AND SIR M.V. COLLEGE
SUBJECT NAME: **Data Analysis with SAS / SPSS / R**
Practical No. 4

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

Output-

The image displays two screenshots of the RStudio interface, showing R code being executed in the console and the resulting data structures in the Environment pane.

Top Screenshot:

```
# Quick look to confirm the data loaded correctly
> head(market_data)
# A tibble: 6 x 7
  datetime      symbol open  high  low close volume
<chr>         <dbl> <dbl> <dbl> <dbl> <dbl>
1 31-07-2025 22:00 NYMEX:CL11 69.4 69.5 69.2 69.4 449
2 31-07-2025 22:15 NYMEX:CL11 69.4 69.4 69.3 69.4 91
3 31-07-2025 22:30 NYMEX:CL11 69.4 69.4 69.3 69.4 44
4 31-07-2025 22:45 NYMEX:CL11 69.4 69.4 69.4 69.4 45
5 31-07-2025 23:00 NYMEX:CL11 69.4 69.4 69.4 69.4 71
6 31-07-2025 23:15 NYMEX:CL11 69.4 69.4 69.4 69.4 61

> # Example 1: Single Condition
> # Professor's logic: Filter for value > X
> # Adapted for you: Filter rows where 'close' price is greater than 69.40
> high_value_subset <- subset(market_data, close > 69.40)
> cat("Number of high close values (close > 69.40):", nrow(high_value_subset), "\n")
Number of high close values (close > 69.40): 8
> summary(high_value_subset$close)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 69.42  69.44  69.44  69.44  69.45  69.47

> cat("Number of rows with close > 69.40 and volume < 100:", nrow(low_vol_high_val_subset), "\n")
Number of rows with close > 69.40 and volume < 100: 1
> head(low_vol_high_val_subset)
# A tibble: 1 x 7
  datetime      symbol open  high  low close volume
<chr>         <dbl> <dbl> <dbl> <dbl> <dbl>
1 31-07-2025 22:45 NYMEX:CL11 69.4 69.4 69.4 69.4 45

> cat("Number of special rows (open < 69.20 OR volume > 500):", nrow(special_subset), "\n")
Number of special rows (open < 69.20 OR volume > 500): 6032
> head(special_subset)
# A tibble: 6 x 7
  datetime      symbol open  high  low close volume
<chr>         <dbl> <dbl> <dbl> <dbl> <dbl>
1 01-08-2025 00:00 NYMEX:CL11 69.4 69.4 69.2 69.3 624
2 01-08-2025 02:30 NYMEX:CL11 69.2 69.2 69.1 69.2 291
3 01-08-2025 02:45 NYMEX:CL11 69.2 69.2 69.1 69.2 131
4 01-08-2025 06:15 NYMEX:CL11 69.3 69.3 69.1 69.1 676
5 01-08-2025 06:30 NYMEX:CL11 69.1 69.2 69.0 69.1 420
6 01-08-2025 06:45 NYMEX:CL11 69.2 69.2 69.1 69.2 420

> # Example 1: Single Condition (Using Pipe Operator |>)
> # Adapted for you: Filter for 'volume' less than 200
> low_vol_filter <- market_data |>
+ filter(volume < 200)
> cat("Number of rows with volume < 200:", nrow(low_vol_filter), "\n")
Number of rows with volume < 200: 903
> summary(low_vol_filter$volume)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 6.0   81.5   121.0   117.6   158.0   199.0

> cat("Rows where high > 69.50 and volume > 300:", nrow(high_vol_high_price_filter), "\n")
Rows where high > 69.50 and volume > 300: 4
> head(high_vol_high_price_filter)
# A tibble: 4 x 7
  datetime      symbol open  high  low close volume
<chr>         <dbl> <dbl> <dbl> <dbl> <dbl>
1 01-08-2025 04:30 NYMEX:CL11 69.5 69.6 69.4 69.4 318
2 01-08-2025 08:15 NYMEX:CL11 69.4 69.5 69.4 69.4 1933
3 01-08-2025 13:00 NYMEX:CL11 69.3 69.6 69.2 69.4 6142
4 01-08-2025 13:15 NYMEX:CL11 69.4 69.6 69.0 69.3 3491

> cat("Number of rows with specific symbols:", nrow(symbol_filter), "\n")
Error: object 'symbol_filter' not found

> head(symbol_filter)
Error: object 'symbol_filter' not found

> # Example 3: Checking for Values in a Set (%in%)
> # Professor's logic: Filter where a column matches a specific list of values
> # Adapted for you: Filter where 'symbol' is exactly "NYMEX:CL" or "NYMEX:CL11"
> symbol_filter <- market_data |>
+ filter(symbol %in% c("NYMEX:CL", "NYMEX:CL11"))
> cat("Number of rows with specific symbols:", nrow(symbol_filter), "\n")
```

Bottom Screenshot:

```
# A tibble: 6 x 7
  datetime      symbol open  high  low close volume
<chr>         <dbl> <dbl> <dbl> <dbl> <dbl>
1 01-08-2025 00:00 NYMEX:CL11 69.4 69.4 69.2 69.3 624
2 01-08-2025 02:30 NYMEX:CL11 69.2 69.2 69.1 69.2 291
3 01-08-2025 02:45 NYMEX:CL11 69.2 69.2 69.1 69.2 131
4 01-08-2025 06:15 NYMEX:CL11 69.3 69.3 69.1 69.1 676
5 01-08-2025 06:30 NYMEX:CL11 69.1 69.2 69.0 69.1 420
6 01-08-2025 06:45 NYMEX:CL11 69.2 69.2 69.1 69.2 420

> # Example 1: Single Condition (Using Pipe Operator |>)
> # Adapted for you: Filter for 'volume' less than 200
> low_vol_filter <- market_data |>
+ filter(volume < 200)
> cat("Number of rows with volume < 200:", nrow(low_vol_filter), "\n")
Number of rows with volume < 200: 903
> summary(low_vol_filter$volume)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 6.0   81.5   121.0   117.6   158.0   199.0

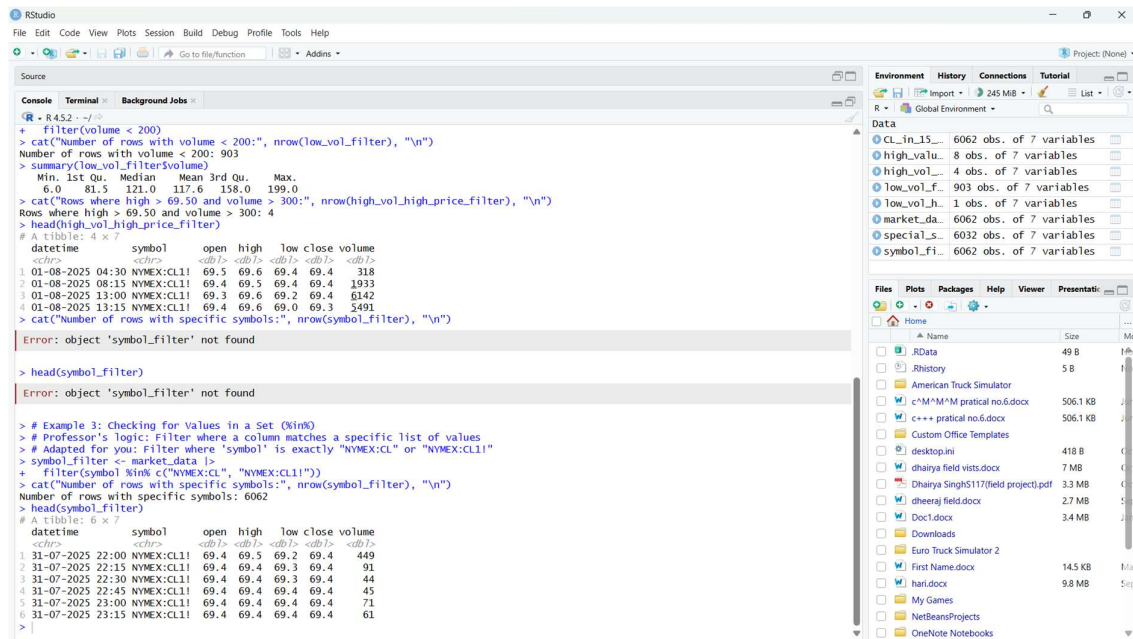
> cat("Rows where high > 69.50 and volume > 300:", nrow(high_vol_high_price_filter), "\n")
Rows where high > 69.50 and volume > 300: 4
> head(high_vol_high_price_filter)
# A tibble: 4 x 7
  datetime      symbol open  high  low close volume
<chr>         <dbl> <dbl> <dbl> <dbl> <dbl>
1 01-08-2025 04:30 NYMEX:CL11 69.5 69.6 69.4 69.4 318
2 01-08-2025 08:15 NYMEX:CL11 69.4 69.5 69.4 69.4 1933
3 01-08-2025 13:00 NYMEX:CL11 69.3 69.6 69.2 69.4 6142
4 01-08-2025 13:15 NYMEX:CL11 69.4 69.6 69.0 69.3 3491

> cat("Number of rows with specific symbols:", nrow(symbol_filter), "\n")
Error: object 'symbol_filter' not found

> head(symbol_filter)
Error: object 'symbol_filter' not found

> # Example 3: Checking for Values in a Set (%in%)
> # Professor's logic: Filter where a column matches a specific list of values
> # Adapted for you: Filter where 'symbol' is exactly "NYMEX:CL" or "NYMEX:CL11"
> symbol_filter <- market_data |>
+ filter(symbol %in% c("NYMEX:CL", "NYMEX:CL11"))
> cat("Number of rows with specific symbols:", nrow(symbol_filter), "\n")
```

SHETH L.U.J AND SIR M.V. COLLEGE
SUBJECT NAME: **Data Analysis with SAS / SPSS / R**
Practical No. 4



The screenshot shows the RStudio environment with the following components:

- Source Editor:** Contains R code for data filtering and summarization. The code includes comments in Hindi and English, and uses functions like `filter()`, `summary()`, `cat()`, and `head()`.
- Console:** Displays the output of the R code, including a summary of the `low_vol_filter` data frame and the number of rows with specific symbols.
- Environment Pane:** Lists the objects in the R environment, including `CL_in_15`, `high_valu`, `low_vol_f`, `market_da`, `special_s`, and `symbol_fi`.
- Files Pane:** Shows the file explorer with various files and folders, including `RData`, `Rhistory`, `American Truck Simulator`, `c++ practical no.6.docx`, `Custom Office Templates`, `desktop.ini`, `dhairya field vists.docx`, `Dhairya Singh's117 field project.pdf`, `dheeraj field.docx`, `Doc1.docx`, `Downloads`, `Euro Truck Simulator 2`, `First Name.docx`, `hari.docx`, `My Games`, `NetBeansProjects`, and `OneNote Notebooks`.

```
+ filter(volume < 200)
> cat("number of rows with volume < 200:", nrow(low_vol_filter), "\n")
Number of rows with volume < 200: 903
> summary(low_vol_filter$volume)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    6.0   81.5   121.0   117.6   158.0   199.0
> cat("Rows where high > 69.50 and volume > 300:", nrow(high_vol_high_price_filter), "\n")
Rows where high > 69.50 and volume > 300: 4
> head(high_vol_high_price_filter)
# A tibble: 4 x 7
  datetime      symbol open  high low close volume
  <chr>         <chr>   <dbl> <dbl> <dbl> <dbl>
1 01-08-2025 04:30 NYMEX:CL11 69.5 69.6 69.4 69.4 318
2 01-08-2025 08:15 NYMEX:CL11 69.4 69.5 69.4 69.4 1933
3 01-08-2025 13:00 NYMEX:CL11 69.3 69.6 69.2 69.4 6142
4 01-08-2025 13:15 NYMEX:CL11 69.4 69.6 69.0 69.3 5491
> cat("number of rows with specific symbols:", nrow(symbol_filter), "\n")
Error: object 'symbol_filter' not found

> head(symbol_filter)
Error: object 'symbol_filter' not found

> # Example 3: Checking for Values in a Set (Nin%)
> # Professor's logic: Filter where a column matches a specific list of values
> # Adapted for you: Filter where 'symbol' is exactly "NYMEX:CL" or "NYMEX:CL11"
> symbol_filter <- market_data |>
+ filter(symbol %in% c("NYMEX:CL", "NYMEX:CL11"))
> cat("number of rows with specific symbols:", nrow(symbol_filter), "\n")
Number of rows with specific symbols: 6062
> head(symbol_filter)
# A tibble: 6 x 7
  datetime      symbol open  high low close volume
  <chr>         <chr>   <dbl> <dbl> <dbl> <dbl>
1 31-07-2025 22:00 NYMEX:CL11 69.4 69.5 69.2 69.4 449
2 31-07-2025 22:15 NYMEX:CL11 69.4 69.4 69.3 69.4 91
3 31-07-2025 22:30 NYMEX:CL11 69.4 69.4 69.3 69.4 44
4 31-07-2025 22:45 NYMEX:CL11 69.4 69.4 69.4 69.4 45
5 31-07-2025 23:00 NYMEX:CL11 69.4 69.4 69.4 69.4 71
6 31-07-2025 23:15 NYMEX:CL11 69.4 69.4 69.4 69.4 61
```