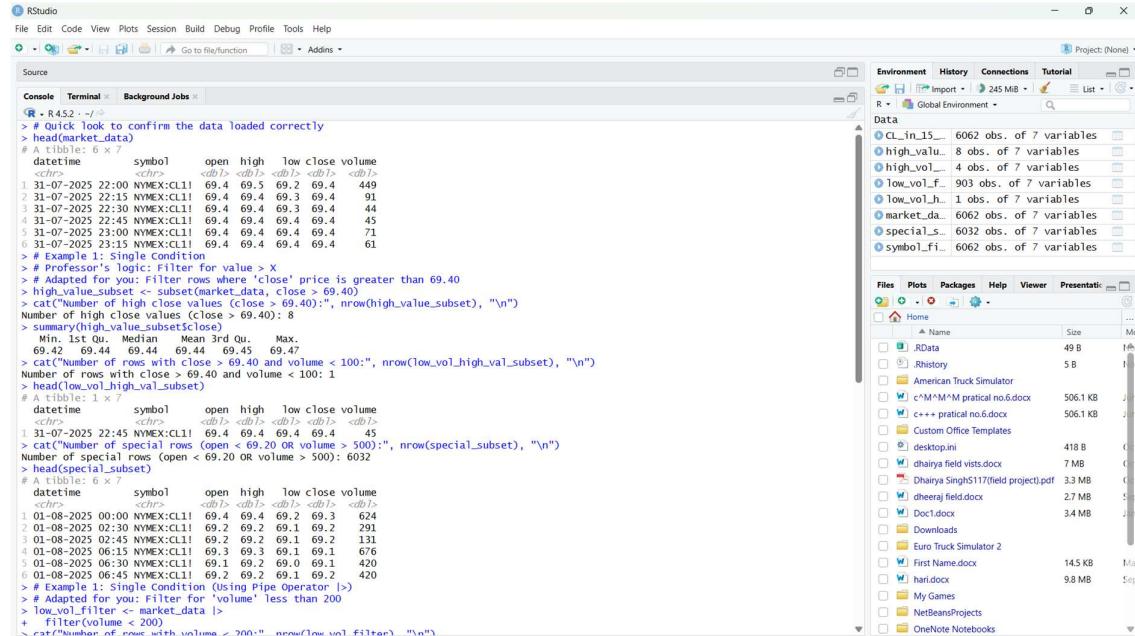


SHETH L.U.J AND SIR M.V. COLLEGE
SUBJECT NAME: Data Analysis with SAS / SPSS /R
Practical No. 4

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

Output-

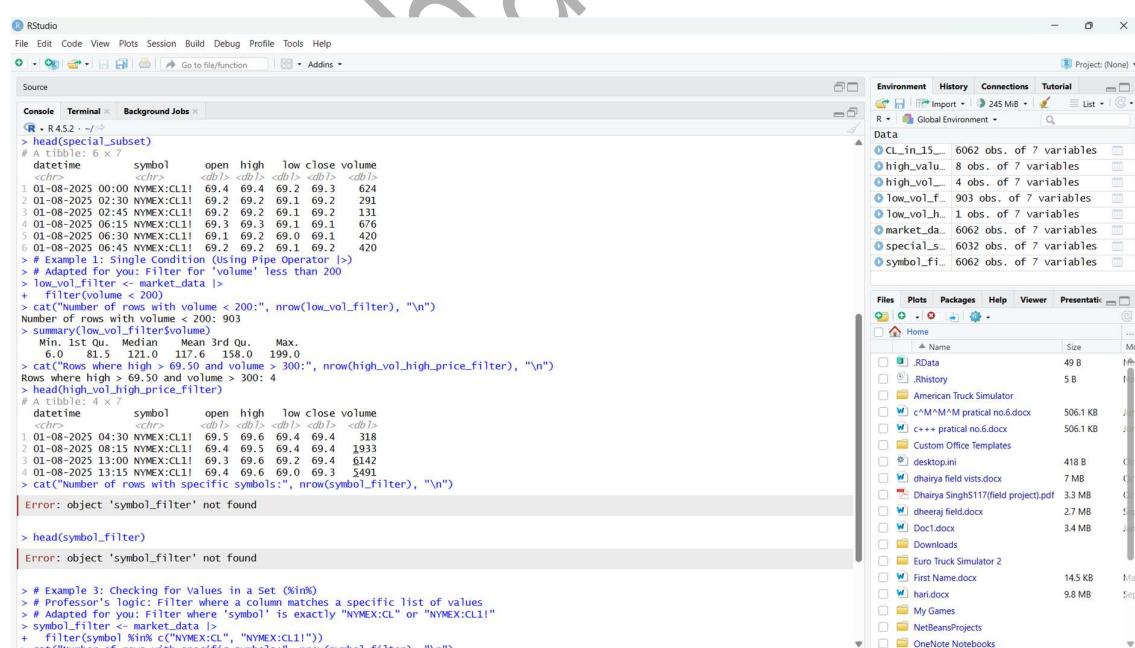


RStudio interface showing R code in the Source pane and the Global Environment pane. The code demonstrates how to filter data based on specific conditions like price and volume.

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Console Terminal Background Jobs
R > R452 ->
> # Quick look to confirm the data loaded correctly
> head(market_data)
# A tibble: 6 × 7
  datetime symbol   open   high   low close volume
  <chr>    <chr>  <dbl>  <dbl>  <dbl> <dbl>  <dbl>
1 31-07-2025 22:00 NYMEX:CL1 69.4 69.5 69.2 69.4 449
2 31-07-2025 22:15 NYMEX:CL1 69.4 69.4 69.3 69.4 91
3 31-07-2025 22:30 NYMEX:CL1 69.4 69.4 69.3 69.4 44
4 31-07-2025 22:45 NYMEX:CL1 69.4 69.4 69.3 69.4 45
5 31-07-2025 23:00 NYMEX:CL1 69.4 69.4 69.4 69.4 71
6 31-07-2025 23:15 NYMEX:CL1 69.4 69.4 69.4 69.4 61
# Example 1: Single Condition (Using Pipe Operator |>)
# 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)
Min. 1st Qu. Median Mean 3rd Qu. Max.
69.42 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 × 7
  datetime symbol   open   high   low close volume
  <chr>    <chr>  <dbl>  <dbl>  <dbl> <dbl>  <dbl>
1 31-07-2025 22:45 NYMEX:CL1 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 × 7
  datetime symbol   open   high   low close volume
  <chr>    <chr>  <dbl>  <dbl>  <dbl> <dbl>  <dbl>
1 01-08-2025 00:00 NYMEX:CL1 69.4 69.4 69.2 69.3 624
2 01-08-2025 02:30 NYMEX:CL1 69.2 69.2 69.1 69.2 291
3 01-08-2025 02:45 NYMEX:CL1 69.2 69.2 69.1 69.2 131
4 01-08-2025 02:51 NYMEX:CL1 69.2 69.2 69.1 69.1 676
5 01-08-2025 06:30 NYMEX:CL1 69.1 69.2 69.0 69.1 420
6 01-08-2025 06:45 NYMEX:CL1 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")

```



RStudio interface showing R code in the Source pane and the Global Environment pane. This example focuses on filtering by specific symbols.

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Console Terminal Background Jobs
R > R452 ->
> head(special_subset)
# A tibble: 6 × 7
  datetime symbol   open   high   low close volume
  <chr>    <chr>  <dbl>  <dbl>  <dbl> <dbl>  <dbl>
1 01-08-2025 00:00 NYMEX:CL1 69.4 69.4 69.2 69.3 624
2 01-08-2025 02:30 NYMEX:CL1 69.2 69.2 69.1 69.2 291
3 01-08-2025 02:45 NYMEX:CL1 69.2 69.2 69.1 69.2 131
4 01-08-2025 02:51 NYMEX:CL1 69.2 69.2 69.1 69.1 676
5 01-08-2025 06:30 NYMEX:CL1 69.1 69.2 69.0 69.1 420
6 01-08-2025 06:45 NYMEX:CL1 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 × 7
  datetime symbol   open   high   low close volume
  <chr>    <chr>  <dbl>  <dbl>  <dbl> <dbl>  <dbl>
1 01-08-2025 04:30 NYMEX:CL1 69.5 69.6 69.4 69.4 318
2 01-08-2025 08:30 NYMEX:CL1 69.3 69.4 69.2 69.4 1933
3 01-08-2025 13:00 NYMEX:CL1 69.3 69.6 69.2 69.3 6342
4 01-08-2025 13:15 NYMEX:CL1 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 (%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:CL1"
> symbol_filter <- market_data |
+ filter(symbol %in% c("NYMEX:CL", "NYMEX:CL1"))
> 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



RStudio interface showing R code execution in the console and file navigation in the sidebar.

```

R - R4.2.2 - - -
> 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 3rd Qu. Max.
  0.000 138.00 199.00 318.00 318.00
> 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 × 7
  datetime     symbol   open   high   low close volume
  <chr>       <chr>    <dbl>  <dbl>  <dbl> <dbl>  <dbl>
1 01-08-2025 04:30 NYMEX:CL1! 69.5   69.6   69.4   69.4   318
2 01-08-2025 08:15 NYMEX:CL1! 69.4   69.5   69.4   69.4   193
3 01-08-2025 13:00 NYMEX:CL1! 69.3   69.6   69.2   69.4   6142
4 01-08-2025 13:15 NYMEX:CL1! 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 (%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:CL1"
> symbol.filter <- market_data |>
+   filter(symbol %in% c("NYMEX:CL", "NYMEX:CL1"))
> cat("Number of rows with specific symbols:", nrow(symbol.filter), "\n")
Number of rows with specific symbols: 6062
> head(symbol.filter)
# A tibble: 6 × 7
  datetime     symbol   open   high   low close volume
  <chr>       <chr>    <dbl>  <dbl>  <dbl> <dbl>  <dbl>
1 31-07-2025 22:00 NYMEX:CL1! 69.4   69.4   69.3   69.4   449
2 31-07-2025 22:15 NYMEX:CL1! 69.4   69.4   69.3   69.4   91
3 31-07-2025 22:30 NYMEX:CL1! 69.4   69.4   69.3   69.4   44
4 31-07-2025 22:45 NYMEX:CL1! 69.4   69.4   69.4   69.4   45
5 31-07-2025 23:00 NYMEX:CL1! 69.4   69.4   69.4   69.4   71
6 31-07-2025 23:15 NYMEX:CL1! 69.4   69.4   69.4   69.4   61
>

```