

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

Subject: Data Analysis with SAS / SPSS /R

Practical no. 4

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

Outputs→

The screenshot displays the RStudio environment with the following components:

- Source:** Contains R code for three examples of conditional filtering.
- Console:** Shows the execution output of the code, including summary statistics and data frames.
- Environment:** Lists the objects created in the global environment: `adj_summary` (1 obs. of 4 variables) and `aubank` (122 obs. of 7 variables).
- Files:** A file explorer showing the local file system.

Code in Source:

```
> # Example 1 - Single Condition
> # Days where Closing Price > 900
> close_gt_900 <- subset(aubank, Close > 900)
> cat("Days with Close > 900:", nrow(close_gt_900), "\n")
Days with Close > 900: 113
> summary(close_gt_900$Close)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 904.8  988.5 1053.4 1072.0 1148.8 1298.9
> # Example 2 - Multiple Conditions (AND)
> # Days where Close > 900 AND Volume > 1,000,000
> close_vol_subset <- subset(aubank, Close > 900 & Volume > 1000000)
> cat("Days with Close > 900 & Volume > 1,000,000:", nrow(close_vol_subset),
"\n")
Days with Close > 900 & Volume > 1,000,000: 62
> head(close_vol_subset)
# A tibble: 6 x 7
  Date       Open High Low Close `Adj Close` Volume
<date>    <dbl> <dbl> <dbl> <dbl>    <dbl>    <dbl>
1 2021-01-14  896.  932.  896.  914      914  1015700
2 2021-01-15  914  922.  907.  920.     920.  1411827
3 2021-02-01  891  950.  891.  942.     942.  2308622
4 2021-02-02  950.  999.  935.  961.     961.  1130196
5 2021-02-10 1074. 1115. 1058. 1099.    1099.  1145055
6 2021-02-18 1098. 1154. 1091. 1135.    1135.  1356719
> # Example 3 - Multiple Conditions (OR)
> # Days where High > 930 OR Low < 880
> volatile_days <- subset(aubank, High > 930 | Low < 880)
> cat("Days with High > 930 OR Low < 880:", nrow(volatile_days), "\n")
Days with High > 930 OR Low < 880: 114
> head(volatile_days)
# A tibble: 6 x 7
  Date       Open High Low Close `Adj Close` Volume
<date>    <dbl> <dbl> <dbl> <dbl>    <dbl>    <dbl>
1 2021-01-11  901.  904.  874.  899.     899.   564393
2 2021-01-14  896.  932.  896.  914      914  1015700
3 2021-01-20  905.  930.  892.  920.     920.   946234
4 2021-01-27  885.  891.  842.  853.     853.   563921
5 2021-01-28  866.  877.  845.  870.     870.   306074
```

Output in Console:

```
Days with Close > 900: 113
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 904.8  988.5 1053.4 1072.0 1148.8 1298.9
Days with Close > 900 & Volume > 1,000,000: 62
# A tibble: 6 x 7
  Date       Open High Low Close `Adj Close` Volume
<date>    <dbl> <dbl> <dbl> <dbl>    <dbl>    <dbl>
1 2021-01-14  896.  932.  896.  914      914  1015700
2 2021-01-15  914  922.  907.  920.     920.  1411827
3 2021-02-01  891  950.  891.  942.     942.  2308622
4 2021-02-02  950.  999.  935.  961.     961.  1130196
5 2021-02-10 1074. 1115. 1058. 1099.    1099.  1145055
6 2021-02-18 1098. 1154. 1091. 1135.    1135.  1356719
Days with High > 930 OR Low < 880: 114
# A tibble: 6 x 7
  Date       Open High Low Close `Adj Close` Volume
<date>    <dbl> <dbl> <dbl> <dbl>    <dbl>    <dbl>
1 2021-01-11  901.  904.  874.  899.     899.   564393
2 2021-01-14  896.  932.  896.  914      914  1015700
3 2021-01-20  905.  930.  892.  920.     920.   946234
4 2021-01-27  885.  891.  842.  853.     853.   563921
5 2021-01-28  866.  877.  845.  870.     870.   306074
```

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

Subject: Data Analysis with SAS / SPSS /R

The screenshot shows the RStudio interface with the 'aubank' dataset loaded. The console displays the following R code and output:

```
R - R4.5.2 - ~/
> # Example 1 - Single Condition (Pipe Operator)
> # Days where Open < 900
> open_lt_900 <- aubank |>
+ filter(Open < 900)
> cat("Days with Open < 900:", nrow(open_lt_900), "\n")
Days with Open < 900: 7
> summary(open_lt_900$Open)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 866.0   882.5   891.0   887.0  895.1   896.4
> # Example 2 - Multiple Conditions (Comma = AND)
> # Days where Adj Close > 910 AND Volume < 600,000
> adj_high_low_vol <- aubank |>
+ filter(Adj Close > 910, Volume < 600000)
> cat("Days with Adj Close > 910 & Volume < 600,000:", nrow(adj_high_low_vol),
"\n")
Days with Adj Close > 910 & Volume < 600,000: 11
> head(adj_high_low_vol)
# A tibble: 6 x 7
   Date       Open High   Low Close `Adj Close` Volume
<date>     <dbl> <dbl> <dbl> <dbl>   <dbl>   <dbl>
1 2021-02-15 1110 1115. 1061. 1092.   1092. 591093
2 2021-02-16 1089 1128 1075. 1093.   1093. 455426
3 2021-02-17 1093. 1112. 1081 1098.   1098. 394405
4 2021-02-19 1124. 1132. 1084. 1109   1109 553747
5 2021-02-22 1109 1128 1048 1075.   1075. 462383
6 2021-02-24 1146. 1198. 1140. 1186.   1186. 571815
> # Example 3 - Values in a Set (%in%)
> # Days where High is either 908 or 920
> high_specific <- aubank |>
+ filter(High %in% c(908, 920))
> cat("Days where High = 908 or 920:", nrow(high_specific), "\n")
Days where High = 908 or 920: 1
> table(high_specific$High)
```

The Environment pane shows the 'aubank' dataset with 122 observations and 7 variables. The Files pane shows the file explorer.

The screenshot shows the RStudio interface with the 'aubank' dataset loaded. The console displays the following R code and output:

```
R - R4.5.2 - ~/
> Days where High = 908 or 920: 1
> table(high_specific$High)
908
1
> # Example 1 - Filter + Select specific columns
> high_close_selected <- aubank |>
+ filter(Close > 900) |>
+ select(Date, High, Low, Close, Volume)
> cat("\nHigh Close days (selected columns):\n")
High Close days (selected columns):
> head(high_close_selected)
# A tibble: 6 x 5
   Date       High   Low Close Volume
<date>     <dbl> <dbl> <dbl> <dbl>
1 2021-01-12 908   895. 905. 472697
2 2021-01-14 932. 896. 914 1015700
3 2021-01-15 922. 907. 920 1411827
4 2021-01-20 930. 892. 920 946234
5 2021-01-21 925. 905 913. 977221
6 2021-02-01 950. 891. 942. 2308622
> # Example 2 - Sort data by descending Volume
> sorted_by_volume <- aubank |>
+ arrange(desc(Volume))
> cat("\nHighest Volume Days:\n")
Highest Volume Days:
> head(sorted_by_volume)
# A tibble: 6 x 7
   Date       Open High   Low Close `Adj Close` Volume
<date>     <dbl> <dbl> <dbl> <dbl>   <dbl>   <dbl>
1 2021-07-06 1056 1143. 1056 1114.   1114. 8527690
2 2021-03-30 1200. 1354. 1165 1299.   1299. 7650626
3 2021-05-03 949   950 914. 924.    924. 6959261
4 2021-04-30 1072 1084 995. 1004.   1004. 6959375
5 2021-05-05 930   952. 910 942.    942. 4528774
6 2021-06-11 996 1064. 996 1046.   1046. 3552844
```

The Environment pane shows the 'aubank' dataset with 122 observations and 7 variables. The Files pane shows the file explorer.

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

Subject: Data Analysis with SAS / SPSS /R

The screenshot displays the RStudio environment with the following components:

- Source:** Contains the R script being executed.
- Console:** Shows the output of the R commands.
- Environment:** Lists the objects in the global environment.
- Files:** Shows the file explorer.

R Script Content:

```
# Example 3 - Combined: Filter + Sort + Select
> elite_days <- aubank |>
+ filter(High > 920) |>
+ arrange(desc(Close)) |>
+ select(Date, High, Close, Volume)
> cat("\nElite days (High > 920 sorted by Close):\n")

Elite days (High > 920 sorted by Close):
> head(elite_days)
# A tibble: 6 x 4
  Date       High Close Volume
<date>     <dbl> <dbl>   <dbl>
1 2021-03-30 1354. 1299. 7650626
2 2021-03-04 1290. 1268. 2247523
3 2021-04-01 1278. 1268. 1087053
4 2021-03-08 1290. 1264. 925121
5 2021-03-05 1294. 1261. 1383020
6 2021-03-12 1275. 1255. 803163

# Example 4 - Summary for conditions
> adj_summary <- aubank |>
+ filter(Adj Close > 900) |>
+ summarise(
+   count = n(),
+   avg_close = mean(Close),
+   max_close = max(Close),
+   min_close = min(Close)
+ )
> cat("\nSummary of days with Adj Close > 900:\n")

Summary of days with Adj Close > 900:
> print(adj_summary)
# A tibble: 1 x 4
  count avg_close max_close min_close
<int>   <dbl>     <dbl>     <dbl>
1    113    1072.    1299.     905.
```

Environment Pane:

- Global Environment:** Lists objects in the global environment.
- adj_summary:** 1 obs. of 4 variables.
- aubank:** 122 obs. of 7 variables.

Files Pane: Shows the file explorer with a list of files and folders.