

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

Aim:4 Applying conditional filters subset() or filter() in R.

The screenshot displays the RStudio interface with the following components:

- Console:** Shows the execution of R code to load the 'dplyr' package, read a CSV file, and apply filters. The code includes:

```
> 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

> library(readr)
> library(readr)
> supermarket_sales <- read_csv("C:/Users/yukta/R Prog_S120/supermarket_sales.csv")
Rows: 1000 Columns: 17
Column specification:
  Delimiter: ","
  chr (8): Invoice ID, Branch, City, Customer type, Gender, Product line, Date, Payment
  dbl (8): Unit price, Quantity, Tax 5%, Total, cogs, gross margin percentage, gross income, ...
  time (1): Time

I use 'spec()' to retrieve the full column specification for this data.
I specify the column types or set 'show_col_types = FALSE' to quiet this message.
> view(supermarket_sales)
> member_or_normal <- sales |>
+ filter(Customer type %in% c("Member", "Normal"))

Error: object 'sales' not found

> sales <- read_csv("C:/Users/yukta/R Prog_S120/supermarket_sales.csv")
Error: '\u' used without hex digits in character string (<input>:1:23)

> sales <- read_csv("C:/Users/yukta/R Prog_S120/supermarket_sales.csv")
Error: '\u' used without hex digits in character string (<input>:1:23)
```
- Environment:** Lists the loaded objects and their dimensions:
 - animal: 7 obs. of 1 variable
 - creditcard: 284807 obs. of 31 variables
 - female_high_quantity: 260 obs. of 17 variables
 - high_total_high_qty: 158 obs. of 17 variables
 - high_total_subset: 227 obs. of 17 variables
 - low_price_filter: 439 obs. of 17 variables
 - member_or_normal: 1000 obs. of 17 variables
 - sales: 1000 obs. of 17 variables
 - special_sales: 516 obs. of 17 variables
- Files:** Shows the project files, including 'supermarket_sales.csv'.
- Source:** Displays the R code being executed, including the following steps:

```
> sales <- read_csv("C:/Users/yukta/R Prog_S120/supermarket_sales.csv")
Rows: 1000 Columns: 17
Column specification:
  Delimiter: ","
  chr (8): Invoice ID, Branch, City, Customer type, Gender, Product line, Date, Payment
  dbl (8): Unit price, Quantity, Tax 5%, Total, cogs, gross margin percentage, gross income, ...
  time (1): Time

I use 'spec()' to retrieve the full column specification for this data.
I specify the column types or set 'show_col_types = FALSE' to quiet this message.
> head(sales)
# A tibble: 6 x 17
  Invoice ID Branch City Customer type Gender Product line Unit price Quantity Tax 5%
  <chr>      <chr> <chr> <chr>      <chr> <chr>      <dbl>    <dbl> <dbl>
1 750-67-8428 A Yang Member Female Health and be... 74.7 7 26.1
2 226-31-3081 C Nayp Normal Female Electronic ac... 15.3 5 3.82
3 631-41-3108 A Yang Normal Male Home and life... 46.3 7 16.2
4 123-19-1176 A Yang Member Male Health and be... 58.2 8 23.3
5 373-73-7910 A Yang Normal Male Sports and tr... 86.3 7 30.2
6 699-14-3026 C Nayp Normal Male Electronic ac... 85.4 7 29.9

# I 8 more variables: Total <dbl>, Date <chr>, Time <time>, Payment <chr>, cogs <dbl>,
# gross margin percentage <dbl>, gross income <dbl>, Rating <dbl>
> high_total_subset <- subset(sales, Total > 500)
>
> cat("Number of high-value sales (Total > 500):", nrow(high_total_subset), "\n")
Number of high-value sales (Total > 500): 227
> summary(high_total_subset$Total)
Min. 1st Qu. Median Mean 3rd Qu. Max.
503.6 580.1 688.7 702.9 803.1 1042.7
>
> high_total_high_qty <- subset(sales, Total > 500 & Quantity > 7)
>
> cat("High Total & High Quantity:", nrow(high_total_high_qty), "\n")
High Total & High Quantity: 158
> head(high_total_high_qty)
# A tibble: 6 x 17
  Invoice ID Branch City Customer type Gender Product line Unit price Quantity Tax 5%
  <chr>      <chr> <chr> <chr>      <chr> <chr>      <dbl>    <dbl> <dbl>
1 315-22-5665 C Nayp Normal Female Home and life... 73.6 10 36.8
2 820-34-3010 A Yang Normal Female Health and be... 71.4 10 35.7
```

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

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for data manipulation. It starts with a data frame of 8 more variables: Total, Date, Time, Payment, cogs, and Rating. The code uses `subset` to filter data based on 'Credit card' payment and 'Rating > 8'. It then creates a new variable 'low_price_filter' based on 'Unit price' and calculates the number of low-price items.
- Console:** Shows the execution of the R code, including the creation of the 'low_price_filter' variable and the calculation of the number of low-price items.
- Environment:** Lists the objects in the environment, including 'animal', 'creditcard', 'Female_high_quantity', 'high_total_high_qty', 'high_total_subset', 'low_price_filter', 'member_or_normal', 'sales', and 'special_sales'.
- Files:** Shows the file explorer, listing files such as 'animal.xlsx', 'creditcard.csv', 'creditcard.xlsx', 'pt_S120R', 'pt.R', 'P2_S120R', 'R-4.5.2-win.exe', 'RStudio.2025.09.2-418.exe', 'StudentsPerformance (1).csv', 'StudentsPerformance (1).xlsx', and 'supermarket_sales.csv'.

The screenshot displays the RStudio IDE with the following components:

- Source Editor:** Contains R code for filtering data based on price and gender.
- Console:** Shows the output of the executed code, including summary statistics for filtered datasets.
- Environment Pane:** Lists objects created in the global environment, such as animal, creditcard, female_high_quantity, etc., along with their dimensions.

```
R - R4.5.2 ~\j>
> low_price_filter <- sales |>
+   filter('Unit price' <= 50)
> cat("Number of low-price items (<= 50):", nrow(low_price_filter), "\n")
Number of low-price items (<= 50): 439
> summary(low_price_filter$Unit price)
    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 10.08   20.57   29.22   29.92   39.75   49.92 
>
> female_high_quantity <- sales |>
+   filter(Gender == "Female", Quantity > 5)
> cat("Female customers with quantity > 5:", nrow(female_high_quantity), "\n")
Female customers with quantity > 5: 260
> head(female_high_quantity)
# A tibble: 6 x 17
      Invoice ID Branch City Customer type Gender Product line Unit price Quantity Tax 5%
      <chr>     <chr>   <chr>       <chr>   <chr>   <chr>         <dbl>    <dbl>   <dbl>
1 750-67-8428 A Yang Member Female Health and be... 74.7      7    26.1
2 315-53-5943 A Yang Member Female Electronic ac... 68.8      6    20.7
3 315-22-5665 C Nayp Normal Female Home and life... 73.6     10    36.8
4 829-34-3910 A Yang Normal Female Health and be... 71.4     10    35.7
5 299-46-1805 B Mand Member Female Sports and tr... 93.7      6    28.1
6 656-95-9349 A Yang Member Female Health and be... 68.9      7    24.1
# 18 more variables: Total <dbl>, Date <chr>, Time <times>, Payment <chr>, cogs <dbl>,
# gross margin percentage <dbl>, gross income <dbl>, Rating <dbl>
>
> member_or_normal <- sales |>
+   filter(Customer type %in% c("Member", "Normal"))
> cat("Total customers (Member or Normal):", nrow(member_or_normal), "\n")
Total customers (Member or Normal): 1000
> table(member_or_normal$Customer type)
```

Customer type	Count
Member	499
Normal	501

The Environment pane shows the following objects:

- animal: 7 obs. of 1 variable
- creditcard: 284807 obs. of 31 variables
- female_high_quantity: 260 obs. of 17 variables
- high_total_high_qty: 158 obs. of 17 variables
- high_total_subset: 227 obs. of 17 variables
- low_price_filter: 439 obs. of 17 variables
- member_or_normal: 1000 obs. of 17 variables
- sales: 1000 obs. of 17 variables
- special_sales: 516 obs. of 17 variables