

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

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

## OUTPUT:-

RStudio  
File Edit Code View Plots Session Build Debug Profile Tools Help  
Console Terminal Background Jobs  
R - R 4.5.2 ->  
> `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/>  
Warning: package 'dplyr' is in use and will not be installed  
  
> `install.packages("readr")`  
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/>  
Warning: package 'readr' is in use and will not be installed  
  
> # Load Libraries  
> library(dplyr)  
> library(readr)  
> # LOAD YOUR CSV FILE ----  
> energy <- read\_csv("energy\_consumption.csv")  
Rows: 5000 Columns: 6  
# Column specification ---  
Delimiter:   
chr (3): customer\_id, customer\_type, regions  
dbl (3): building\_size\_m2, occupants, energy\_cost\_brl  
# Use `spec()` to retrieve the full column specification for this data.  
# Specify the column types or set `show_col_types = FALSE` to quiet this message.  
> view(first\_rows  
> head(energy)  
# A tibble: 6 × 6  
customer\_id customer\_type regions building\_size\_m2 occupants energy\_cost\_brl  
1 CUSTOMER\_0001 residential Northeast 24 2 64.5  
2 CUSTOMER\_0002 commercial Midwest 24 1 55.3  
3 CUSTOMER\_0003 commercial Southeast 24 1 74.5  
4 CUSTOMER\_0004 residential Northeast 45 4 147.  
5 CUSTOMER\_0005 residential Southeast 45 4 142.  
6 CUSTOMER\_0006 residential North 52 2 96.7  
  
Environment History Connections Tutorial  
Import Dataset 97 MB  
Global Environment  
Data  
● energy 5000 obs. of 6 variables  
● high\_cost 1101 obs. of 6 variables  
● high\_cost\_low\_occ 118 obs. of 6 variables  
● high\_cost\_not\_se 495 obs. of 6 variables  
● low\_occupants\_filter 1386 obs. of 6 variables  
● region\_filter 3278 obs. of 6 variables  
● special\_buildings 3007 obs. of 6 variables  
  
Files Plots Packages Help Viewer Presentation  
New Folder New File Delete Rename More  
Home ▲ Name Size Modified  
FrontPage.htm 29.5 KB Oct 17, 2023, 10:00 AM  
My Music  
My Pictures  
My Videos  
NetBeansProjects  
O.E - Copy.pptx 1.1 MB Aug 5, 2025, 10:00 AM  
O.E.pptx 1.1 MB Aug 4, 2025, 8:25 PM  
prac 1.ttx 125 B Sep 26, 2025, 8:12 AM  
Practical\_No\_3.R.R 339 B Nov 24, 2025, 12:33 PM  
Practical\_No\_4.R.R 1.3 KB Nov 24, 2025, 6:38 PM  
Processed\_Student\_Mental\_Health.csv 6.1 KB Nov 18, 2025, 10:58 AM  
S093 CHETAN MANDAVKAR FEATURE ENGINEERING SUBMISSION - Colab 2.3 MB Jul 27, 2025, 4:18 PM  
S093\_Chetan\_Mandavkar\_FP\_Soft\_Copy\_Documentation\_1[1].docx 45.1 KB Oct 15, 2025, 8:30 PM  
SHRUTI MANDAVKAR EVS REPORT.pdf 357.3 KB Nov 21, 2025, 7:34 PM  
Structure and functions of pancreas Shruti Mandavkar.pptx 639.7 KB Aug 24, 2025, 11:35 AM  
Virtual Machines  
world\_suicide\_rates.csv 157.1 KB Nov 24, 2025, 11:38 AM  
  
Air Poor Wednesday DECODED OCTA 18:42 IN 24-11-2025

RStudio  
File Edit Code View Plots Session Build Debug Profile Tools Help  
Console Terminal Background Jobs  
R - R 4.5.2 ->  
> # View first rows  
> head(energy)  
# A tibble: 6 × 6  
customer\_id customer\_type regions building\_size\_m2 occupants energy\_cost\_brl  
1 CUSTOMER\_0001 residential Northeast 24 2 64.5  
2 CUSTOMER\_0002 commercial Midwest 24 1 55.3  
3 CUSTOMER\_0003 commercial Southeast 24 1 74.5  
4 CUSTOMER\_0004 residential Northeast 45 4 147.  
5 CUSTOMER\_0005 residential Southeast 45 4 142.  
6 CUSTOMER\_0006 residential North 52 2 96.7  
  
Environment History Connections Tutorial  
Import Dataset 97 MB  
Global Environment  
Data  
● energy 5000 obs. of 6 variables  
● high\_cost 1101 obs. of 6 variables  
● high\_cost\_low\_occ 118 obs. of 6 variables  
● high\_cost\_not\_se 495 obs. of 6 variables  
● low\_occupants\_filter 1386 obs. of 6 variables  
● region\_filter 3278 obs. of 6 variables  
● special\_buildings 3007 obs. of 6 variables  
  
Files Plots Packages Help Viewer Presentation  
New Folder New File Delete Rename More  
Home ▲ Name Size Modified  
FrontPage.htm 29.5 KB Oct 17, 2025, 10:00 AM  
My Music  
My Pictures  
My Videos  
NetBeansProjects  
O.E - Copy.pptx 1.1 MB Aug 5, 2025, 10:00 AM  
O.E.pptx 1.1 MB Aug 4, 2025, 8:25 PM  
prac 1.ttx 125 B Sep 26, 2025, 8:12 AM  
Practical\_No\_3.R.R 339 B Nov 24, 2025, 12:33 PM  
Practical\_No\_4.R.R 1.3 KB Nov 24, 2025, 6:38 PM  
Processed\_Student\_Mental\_Health.csv 6.1 KB Nov 18, 2025, 10:58 AM  
S093 CHETAN MANDAVKAR FEATURE ENGINEERING SUBMISSION - Colab 2.3 MB Jul 27, 2025, 4:18 PM  
S093\_Chetan\_Mandavkar\_FP\_Soft\_Copy\_Documentation\_1[1].docx 45.1 KB Oct 15, 2025, 8:30 PM  
SHRUTI MANDAVKAR EVS REPORT.pdf 357.3 KB Nov 21, 2025, 7:34 PM  
Structure and functions of pancreas Shruti Mandavkar.pptx 639.7 KB Aug 24, 2025, 11:35 AM  
Virtual Machines  
world\_suicide\_rates.csv 157.1 KB Nov 24, 2025, 11:38 AM  
  
Air Poor Wednesday DECODED OCTA 18:43 IN 24-11-2025

NAME:- CHETAN MANDAVKAR  
ROLL NO. S093  
SUBJECT:- Data Analysis with SAS / SPSS /R

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

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
R > install.packages("readr")
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/
Warning: package 'readr' is in use and will not be installed

> library(dplyr)
> library(readr)
> energy <- read_csv("energy_consumption.csv")
Rows: 5000 Columns: 6
# Column specification:
# Delimiter: ","
# col1 (3): customer_id, customer_type, regions
# col2 (3): building_size_m2, occupants, energy_cost_brl
Use `spec()` to retrieve the full column specification for this data.
Specify the column types or set `show_col_types = FALSE` to quiet this message.
> head(energy)
# A tibble: 6 x 6
  customer_id customer_type regions  building_size_m2 occupants energy_cost_brl
  <chr>        <chr>      <chr>          <dbl>    <dbl>        <dbl>
1 CUSTOMER_0001 residential Northeast     24         2       64.5
2 CUSTOMER_0002 commercial Midwest      24         1       55.3
3 CUSTOMER_0003 commercial Southeast    24         1       74.5
4 CUSTOMER_0004 residential Northeast     45         4       147.
5 CUSTOMER_0005 residential Southeast    45         4       143.
6 CUSTOMER_0006 residential North        52         2       96.7
> high_cost <- subset(energy, energy_cost_brl > 100)
> cat("Number of high-cost customers (energy_cost_brl > 100): ", nrow(high_cost), "\n")
Number of high-cost customers (energy_cost_brl > 100): 1101
> summary(high_cost$energy_cost_brl)
   Min. 1st Qu. Median  Mean 3rd Qu. Max.
100.0 103.1 119.9 122.8 140.4 158.6
> high_cost_low_occ <- subset(energy, energy_cost_brl > 100 & occupants < 2)
> cat("High cost + low occupants customers:", nrow(high_cost_low_occ), "\n")
High cost + low occupants customers: 118
> head(high_cost_low_occ)
# A tibble: 6 x 6
  customer_id customer_type regions  building_size_m2 occupants energy_cost_brl
  <chr>        <chr>      <chr>          <dbl>    <dbl>        <dbl>
1 CUSTOMER_0017 residential North       17         1       103.
2 CUSTOMER_0026 commercial Midwest     24         1       101.
3 CUSTOMER_0036 residential South      24         1       101.
4 CUSTOMER_0054 residential Northeast    45         1       101.
5 CUSTOMER_0071 commercial Midwest     77         1       100.
6 CUSTOMER_0121 residential North       52         1       101.
> spatial_buildings <- subset(energy, building_size_m2 > 40 | occupants > 4)
> cat("Number of spatial buildings:", nrow(spatial_buildings), "\n")
Number of spatial buildings: 3007
> head(spatial_buildings)
# A tibble: 6 x 6
  customer_id customer_type regions  building_size_m2 occupants energy_cost_brl
  <chr>        <chr>      <chr>          <dbl>    <dbl>        <dbl>
1 CUSTOMER_0004 residential Northeast    45         4       147.
2 CUSTOMER_0005 residential Southeast    45         4       143.
3 CUSTOMER_0006 residential North        52         2       96.7
4 CUSTOMER_0008 commercial Southeast    45         3       102.
5 CUSTOMER_0009 residential Northeast    45         2       70.0
6 CUSTOMER_0010 commercial Midwest      45         4       158.
> low_occurrences <- filter(energy, energy < 2)
> cat("Number of low-occupant customers:", nrow(low_occurrences), "\n")
Number of low-occupant customers: 1386
> summary(low_occurrences$occupants)
   Min. 1st Qu. Median  Mean 3rd Qu. Max.
1           1           1           1           1           1

```

Air Poor Wednesday 18:44 24-11-2025

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
R > nrow(high_cost), "\n")
Number of high-cost customers (energy_cost_brl > 100): 1101
> summary(high_cost$energy_cost_brl)
   Min. 1st Qu. Median  Mean 3rd Qu. Max.
100.0 103.1 119.9 122.8 140.4 158.6
> high_cost_low_occ <- subset(energy, energy_cost_brl > 100 & occupants < 2)
> cat("High cost + low occupants customers:", nrow(high_cost_low_occ), "\n")
High cost + low occupants customers: 118
> head(high_cost_low_occ)
# A tibble: 6 x 6
  customer_id customer_type regions  building_size_m2 occupants energy_cost_brl
  <chr>        <chr>      <chr>          <dbl>    <dbl>        <dbl>
1 CUSTOMER_0017 residential North       17         1       103.
2 CUSTOMER_0026 commercial Midwest     24         1       101.
3 CUSTOMER_0036 residential South      24         1       101.
4 CUSTOMER_0054 residential Northeast    45         1       101.
5 CUSTOMER_0071 commercial Midwest     77         1       100.
6 CUSTOMER_0121 residential North       52         1       101.
> spatial_buildings <- subset(energy, building_size_m2 > 40 | occupants > 4)
> cat("Number of spatial buildings:", nrow(spatial_buildings), "\n")
Number of spatial buildings: 3007
> head(spatial_buildings)
# A tibble: 6 x 6
  customer_id customer_type regions  building_size_m2 occupants energy_cost_brl
  <chr>        <chr>      <chr>          <dbl>    <dbl>        <dbl>
1 CUSTOMER_0004 residential Northeast    45         4       147.
2 CUSTOMER_0005 residential Southeast    45         4       143.
3 CUSTOMER_0006 residential North        52         2       96.7
4 CUSTOMER_0008 commercial Southeast    45         3       102.
5 CUSTOMER_0009 residential Northeast    45         2       70.0
6 CUSTOMER_0010 commercial Midwest      45         4       158.
> low_occurrences <- filter(energy, energy < 2)
> cat("Number of low-occupant customers:", nrow(low_occurrences), "\n")
Number of low-occupant customers: 1386
> summary(low_occurrences$occupants)
   Min. 1st Qu. Median  Mean 3rd Qu. Max.
1           1           1           1           1           1

```

Air Poor Wednesday 18:44 24-11-2025

**NAME:- CHETAN MANDAVKAR**  
**ROLL NO. S093**  
**SUBJECT:- Data Analysis with SAS / SPSS /R**

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

The screenshot shows an RStudio interface with the following details:

- Console Tab:** Displays a command-line session with R code and its output. The code includes filtering data by regions like Northeast and Midwest, calculating summary statistics, and listing specific files.
- Environment Tab:** Shows the global environment with objects like `energy`, `high_cost`, and `region_filter`.
- Files Tab:** Browsing the local directory, showing files such as `Homework.Rmd`, `O.E-Copy.pptx`, `O.E.pptx`, `prac.1.R`, `Practical_No_3.R_R`, `Practical_No_4.R_R`, `Processed_Student_Mental_Health.csv`, `S093 CHETAN MANDAVKAR FEATURE ENGINEERING SUBMISSION - Colat`, `S093_Chetan_Mandavkar_FP_Soft_Copy_Documentation_[1].docx`, `SHRUTI MANDAVKAR EVS REPORT.pdf`, `Structure and functions of pancreas Shriti Mandavkar.pptx`, and `Virtual Machines`.

**NAME:- CHETAN MANDAVKAR**  
**ROLL NO. S093**  
**SUBJECT:- Data Analysis with SAS / SPSS /R**