

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

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

OUTPUT:-

The screenshot shows the RStudio interface. The console window displays the following commands and output:

```
> 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_br1
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 first rows
> head(energy)
# A tibble: 6 x 6
  customer_id customer_type regions building_size_m2 occupants energy_cost_br1
  <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
```

The Environment pane on the right shows the loaded data objects:

Object	Size	Modified
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	

The screenshot shows the RStudio interface. The console window displays the following commands and output:

```
> # View first rows
> head(energy)
# A tibble: 6 x 6
  customer_id customer_type regions building_size_m2 occupants energy_cost_br1
  <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

> install.packages("MASS")
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 'MASS' is in use and will not be installed

> library(MASS)
> head(energy)
# A tibble: 6 x 6
  customer_id customer_type regions building_size_m2 occupants energy_cost_br1
  <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

> 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")
```

The Environment pane on the right shows the loaded data objects:

Object	Size	Modified
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	

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
Go to file/function Addins

Source
Console Terminal Background Jobs
R - R452 - ~/ -
> 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: ","
chr (3): customer_id, customer_type, regions
dbl (3): building_size_m2, occupants, energy_cost_br1

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(energy)
# A tibble: 6 x 6
  customer_id customer_type regions building_size_m2 occupants energy_cost_br1
  <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_br1 > 100)
> cat("Number of high-cost customers (energy_cost_br1 > 100):",
+     nrow(high_cost), "\n")
Number of high-cost customers (energy_cost_br1 > 100): 1101
> summary(high_cost$energy_cost_br1)
  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_br1 > 100 & occupants < 2)
> cat("High cost + low occupants customers:",
+     nrow(high_cost_low_occ), "\n")
High cost + low occupants customers: 118
```

Environment History Connections Tutorial

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.html	273 KB	Oct 17, 2023, 2:33 PM
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.txt	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 18:44 24-11-2025

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins

Source
Console Terminal Background Jobs
R - R452 - ~/ -
+ nrow(high_cost), "\n")
Number of high-cost customers (energy_cost_br1 > 100): 1101
> summary(high_cost$energy_cost_br1)
  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_br1 > 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_br1
  <chr> <chr> <chr> <dbl> <dbl> <dbl>
1 CUSTOMER_0017 residential North 17 1 103.
2 CUSTOMER_0026 commercial Midwest 24 1 101.
3 CUSTOMER_0038 residential South 24 1 101.
4 CUSTOMER_0058 residential Northeast 45 1 101.
5 CUSTOMER_0071 commercial Midwest 77 1 100.
6 CUSTOMER_0123 residential North 52 1 101.

> special_buildings <- subset(energy, building_size_m2 > 40 | occupants > 4)
> cat("Number of special buildings:", nrow(special_buildings), "\n")
Number of special buildings: 3007
> head(special_buildings)
# A tibble: 6 x 6
  customer_id customer_type regions building_size_m2 occupants energy_cost_br1
  <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_occupants_filter <- energy |>
+   filter(occupants < 2)
> cat("Number of low-occupant customers:",
+     nrow(low_occupants_filter), "\n")
Number of low-occupant customers: 1386
> summary(low_occupants_filter$occupants)
  Min. 1st Qu. Median Mean 3rd Qu. Max.
1 1 1 1 1 1 1
> high_cost_not_se <- energy |>
+   filter(energy_cost_br1 > 100 & !high_cost_low_occ)
> cat("Number of high-cost customers (not high cost + low occupants):",
+     nrow(high_cost_not_se), "\n")
Number of high-cost customers (not high cost + low occupants): 483
```

Environment History Connections Tutorial

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.html	273 KB	Oct 17, 2023, 2:33 PM
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.txt	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

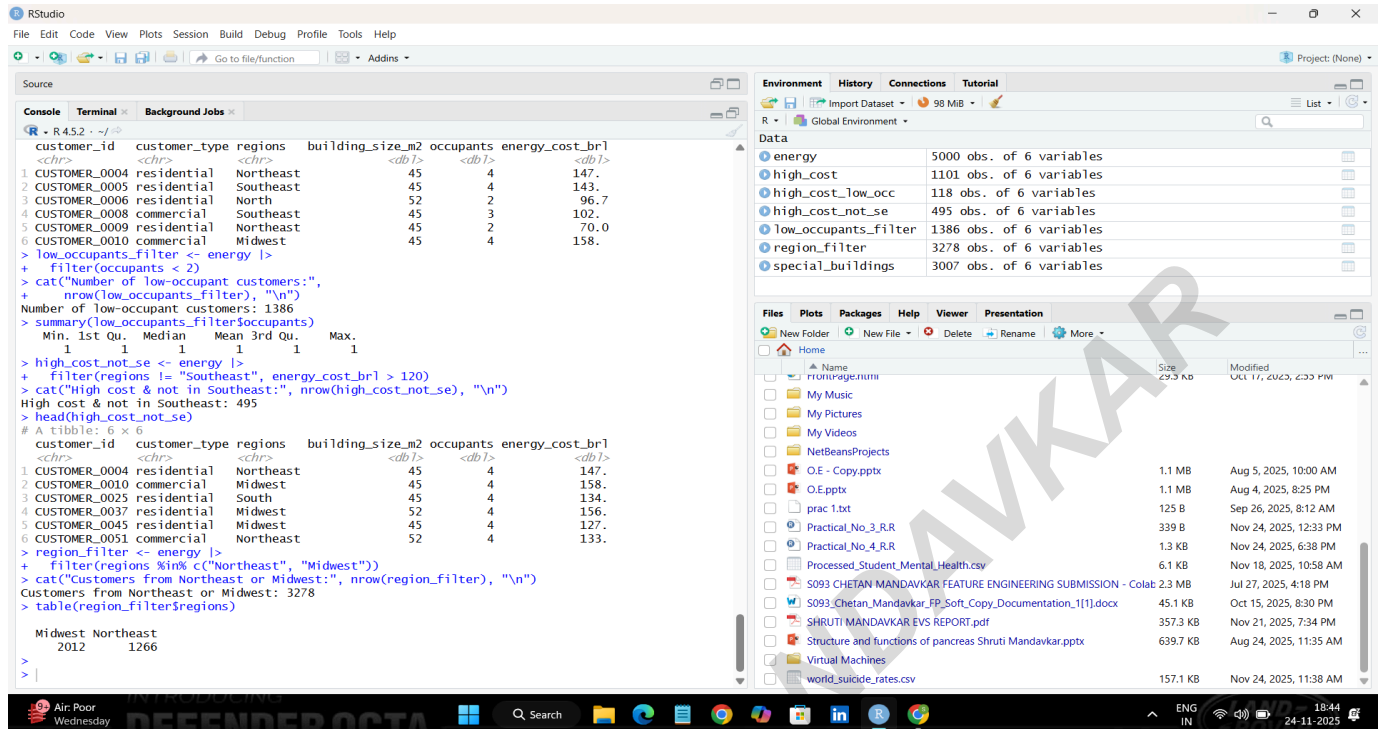
Air Poor Wednesday 18:44 24-11-2025

NAME:- CHETAN MANDAVKAR

ROLL NO. S093

SUBJECT:- Data Analysis with SAS / SPSS / R

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



The screenshot displays the RStudio interface. The console on the left shows the following R code and its output:

```
R - R4.3.2 - ~/ -  
customer_id customer_type regions building_size_m2 occupants energy_cost_br1  
<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_occupants_filter <- energy |>  
+ filter(occupants < 2)  
> cat("Number of low-occupant customers:",  
+ nrow(low_occupants_filter), "\n")  
Number of low-occupant customers: 1386  
> summary(low_occupants_filter$occupants)  
Min. 1st Qu. Median Mean 3rd Qu. Max.  
1 1 1 1 1 1  
> high_cost_not_se <- energy |>  
+ filter(regions != "Southeast", energy_cost_br1 > 120)  
> cat("High cost & not in Southeast:", nrow(high_cost_not_se), "\n")  
High cost & not in Southeast: 495  
> head(high_cost_not_se)  
# A tibble: 6 x 6  
customer_id customer_type regions building_size_m2 occupants energy_cost_br1  
<chr> <chr> <chr> <dbl> <dbl> <dbl>  
1 CUSTOMER_0004 residential Northeast 45 4 147.  
2 CUSTOMER_0010 commercial Midwest 45 4 158.  
3 CUSTOMER_0025 residential South 45 4 134.  
4 CUSTOMER_0037 residential Midwest 52 4 156.  
5 CUSTOMER_0045 residential Midwest 45 4 127.  
6 CUSTOMER_0051 commercial Northeast 52 4 133.  
> region_filter <- energy |>  
+ filter(regions %in% c("Northeast", "Midwest"))  
> cat("Customers from Northeast or Midwest:", nrow(region_filter), "\n")  
Customers from Northeast or Midwest: 3278  
> table(region_filter$regions)  
  
Midwest Northeast  
2012 1266  
>  
>
```

The Environment pane on the right lists the following objects:

- 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

The Files pane shows the file explorer with various files and folders, including 'Home', 'My Music', 'My Pictures', 'My Videos', 'NetBeansProjects', 'O.E - Copy.pptx', 'O.E.pptx', 'prac1.txt', 'Practical_No_3_RR', 'Practical_No_4_RR', 'Processed_Student_Mental_Health.csv', 'S093 CHETAN MANDAVKAR FEATURE ENGINEERING SUBMISSION - Colab', 'S093_Chetan_Mandavkar_FP_Soft_Copy_Documentation_1[1].docx', 'SHRUTI MANDAVKAR EVS REPORT.pdf', 'Structure and functions of pancreas Shruti Mandavkar.pptx', 'Virtual Machines', and 'world_suicide_rates.csv'.

NAME:- CHETAN MANDAVKAR

ROLL NO. S093

SUBJECT:- Data Analysis with SAS / SPSS /R