

PRACTICAL NO:4

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

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
R - R 4.1.2 - ~/
> 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/

Installing package into 'C:/Users/IT-03/Documents/R/win-library/4.1'
(as 'lib' is unspecified)
also installing the dependency 'vctrs'

There are binary versions available but the source versions are later:
binary source needs_compilation
vctrs 0.6.1 0.6.5 TRUE
dplyr 1.1.2 1.1.4 TRUE

Binaries will be installed

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/vctrs_0.6.1.zip'
Content type 'application/zip' length 1581321 bytes (1.5 MB)
downloaded 1.5 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/dplyr_1.1.2.zip'
Content type 'application/zip' length 1646641 bytes (1.6 MB)
downloaded 1.6 MB

package 'vctrs' successfully unpacked and MD5 sums checked
Warning: cannot remove prior installation of package 'vctrs'
Warning: restored 'vctrs'
package 'dplyr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:\Users\IT-03\AppData\Local\Temp\RtmpCSHkBC\downloaded_packages
Warning message:
In file.copy(savedcopy, lib, recursive = TRUE) :

Environment History Connections Tutorial
R - Global Environment
Data
furnished_subset 367 obs. of 13 variables
high_price_subset 8 obs. of 13 variables
housing 545 obs. of 13 variables
Housing 545 obs. of 13 variables
insurance 1338 obs. of 7 variables
large_highprice_s... 4 obs. of 13 variables
mainroad_basement 168 obs. of 13 variables
Files Plots Packages Help Viewer Presentation
New Folder New File Delete Rename More
Home My Web Sites
Name Size Modified
insurance.csv 54.3 KB Nov 24, 2025, 11:39 AM
WebSite1
Housing.csv 29.3 KB Nov 18, 2025, 11:51 AM

```

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
R - R 4.1.2 - ~/
> 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

Warning message:
package 'dplyr' was built under R version 4.1.3

> library(readr) # For efficient reading
> # Load your Housing.csv file
> housing <- read_csv("My web sites/Housing.csv")
Rows: 545 Columns: 13
-- Column specification -----
Delimiter: ","
chr (7): mainroad, guestroom, basement, hotwaterheating, airconditioning, pr...
dbl (6): price, area, bedrooms, bathrooms, stories, parking

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.
> # quick look at the data structure
> head(housing)
# A tibble: 6 x 13
  price area bedrooms bathrooms stories mainroad guestroom basement
<dbl> <dbl> <dbl> <dbl> <dbl> <chr> <chr> <chr>

```

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

RStudio interface showing the following R code in the console:

```
> head(housing)
# A tibble: 6 x 13
  price area bedrooms bathrooms stories mainroad guestroom basement
  <dbl> <dbl> <dbl> <dbl> <dbl> <chr> <chr> <chr>
1 13300000 7420 4 2 3 yes no no
2 12250000 8960 4 4 4 yes no no
3 12250000 9960 3 2 2 yes no yes
4 12215000 7500 4 2 2 yes no yes
5 11410000 7420 4 1 2 yes yes yes
6 10850000 7500 3 3 1 yes no yes

# i 5 more variables: hotwaterheating <chr>, airconditioning <chr>,
# parking <dbl>, prefarea <chr>, furnishingstatus <chr>
> high_price_subset <- subset(housing, price > 10000000)
> cat("Number of high-price houses (price > 10M):", nrow(high_price_subset), "\n")
Number of high-price houses (price > 10M): 8
> summary(high_price_subset$price)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
10150000 10675000 11812500 11571875 12250000 13300000
> large_highprice_subset <- subset(housing, price > 10000000 & area > 8000)
> cat("Number of large, high-price houses:", nrow(large_highprice_subset), "\n")
Number of large, high-price houses: 4
> head(large_highprice_subset)
# A tibble: 4 x 13
  price area bedrooms bathrooms stories mainroad guestroom basement
  <dbl> <dbl> <dbl> <dbl> <dbl> <chr> <chr> <chr>
1 12250000 8960 4 4 4 yes no no
2 12250000 9960 3 2 2 yes no yes
3 10150000 8580 4 3 4 yes no no
4 10150000 16200 5 3 2 yes no no

# i 5 more variables: hotwaterheating <chr>, airconditioning <chr>,
# parking <dbl>, prefarea <chr>, furnishingstatus <chr>
> special_houses_subset <- subset(housing, guestroom == "yes" | airconditioning == "yes")
> cat("Number of special houses (guestroom OR AC):", nrow(special_houses_subset), "\n")
Number of special houses (guestroom OR AC): 225
```

The Environment pane on the right shows the following objects:

Object	Size	Modified
furnished_subset	367 obs. of 13 variables	
high_price_subset	8 obs. of 13 variables	
housing	545 obs. of 13 variables	
Housing	545 obs. of 13 variables	
insurance	1338 obs. of 7 variables	
large_highprice_s...	4 obs. of 13 variables	
mainroad_basement	168 obs. of 13 variables	

RStudio interface showing the following R code in the console:

```
> head(special_houses_subset)
# A tibble: 6 x 13
  price area bedrooms bathrooms stories mainroad guestroom basement
  <dbl> <dbl> <dbl> <dbl> <dbl> <chr> <chr> <chr>
1 13300000 7420 4 2 3 yes no no
2 12250000 8960 4 4 4 yes no no
3 12215000 7500 4 2 2 yes no yes
4 11410000 7420 4 1 2 yes yes yes
5 10850000 7500 3 3 1 yes no yes
6 10150000 8580 4 3 4 yes no no

# i 5 more variables: hotwaterheating <chr>, airconditioning <chr>,
# parking <dbl>, prefarea <chr>, furnishingstatus <chr>
> many_bathrooms <- housing |>
+ filter(bathrooms > 3)
> cat("Number of houses with > 3 bathrooms:", nrow(many_bathrooms), "\n")
Number of houses with > 3 bathrooms: 1
> summary(many_bathrooms$bathrooms)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
     4         4         4         4         4         4
> mainroad_basement <- housing |>
+ filter(mainroad == "yes", basement == "yes")
> cat("Houses on main road with basement:", nrow(mainroad_basement), "\n")
Houses on main road with basement: 168
> head(mainroad_basement)
# A tibble: 6 x 13
  price area bedrooms bathrooms stories mainroad guestroom basement
  <dbl> <dbl> <dbl> <dbl> <dbl> <chr> <chr> <chr>
1 12250000 9960 3 2 2 yes no yes
2 12215000 7500 4 2 2 yes no yes
3 11410000 7420 4 1 2 yes yes yes
4 10850000 7500 3 3 1 yes no yes
5 9870000 8100 4 1 2 yes yes yes
6 9800000 13200 3 1 2 yes no yes

# i 5 more variables: hotwaterheating <chr>, airconditioning <chr>,
# parking <dbl>, prefarea <chr>, furnishingstatus <chr>
```

The Environment pane on the right shows the following objects:

Object	Size	Modified
furnished_subset	367 obs. of 13 variables	
high_price_subset	8 obs. of 13 variables	
housing	545 obs. of 13 variables	
Housing	545 obs. of 13 variables	
insurance	1338 obs. of 7 variables	
large_highprice_s...	4 obs. of 13 variables	
mainroad_basement	168 obs. of 13 variables	

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

The screenshot displays the RStudio environment with the following components:

- Source Pane:** Contains R code for data manipulation. The code includes comments, variable assignments, and filtering operations. The output shows a summary of houses with more than 3 bathrooms and a subset of houses based on mainroad and basement status.
- Console:** Displays the output of the R code, including a summary of houses with more than 3 bathrooms and a subset of houses based on mainroad and basement status.
- Environment Pane:** Lists the objects in the global environment, including 'furnished_subset', 'high_price_subset', 'housing', 'insurance', 'large_highprice_s...', and 'mainroad_basement'.
- Files Pane:** Shows the file structure, including 'insurance.csv', 'WebSite1', and 'Housing.csv'.

```
R - R 4.1.2 ~\> # i 5 more variables: hotwaterheating <chr>, airconditioning <chr>,  
# parking <dbl>, prefarea <chr>, furnishingstatus <chr>  
> many_bathrooms <- housing |>  
+ filter(bathrooms > 3)  
> cat("Number of houses with > 3 bathrooms:", nrow(many_bathrooms), "\n")  
Number of houses with > 3 bathrooms: 1  
> summary(many_bathrooms$bathrooms)  
Min. 1st Qu. Median Mean 3rd Qu. Max.  
4 4 4 4 4 4  
> mainroad_basement <- housing |>  
+ filter(mainroad == "yes", basement == "yes")  
> cat("Houses on main road with basement:", nrow(mainroad_basement), "\n")  
Houses on main road with basement: 168  
> head(mainroad_basement)  
# A tibble: 6 x 13  
  price area bedrooms bathrooms stories mainroad guestroom basement  
  <dbl> <dbl> <dbl> <dbl> <dbl> <chr> <chr> <chr>  
1 12250000 9960 3 2 2 yes no yes  
2 12215000 7500 4 2 2 yes no yes  
3 11410000 2420 4 1 2 yes yes yes  
4 10850000 7500 3 3 1 yes no yes  
5 9870000 8100 4 1 2 yes yes yes  
6 9800000 13200 3 1 2 yes no yes  
# i 5 more variables: hotwaterheating <chr>, airconditioning <chr>,  
# parking <dbl>, prefarea <chr>, furnishingstatus <chr>  
> furnished_subset <- housing |>  
+ filter(furnishingstatus %in% c("furnished", "semi-furnished"))  
> cat("Number of furnished or semi-furnished houses:", nrow(furnished_subset), "\n")  
Number of furnished or semi-furnished houses: 367  
> table(furnished_subset$furnishingstatus)  
  
furnished semi-furnished  
140 227  
>  
>
```