

SHETH L.U.J & SIR M.V COLLEGE OF SCIENCE

SUBJECT : Data Analysis with SAS / SPSS /R

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 - /]
> 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)
> spotify <- read_csv("spotify_data.csv")
Rows: 8582 Columns: 15
#> # Column specification 
#> Delimiter: ","
#> chr (8): track_id, track_name, artist_name, artist_genres, album_id, album_name, album_release...
#> dbl (6): track_number, track_popularity, artist_popularity, artist_followers, album_total_tracks...
#> lgl (1): explicit
#> 
#> 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(spotify)
# A tibble: 6 × 15
  track_id   track_name track_number track_popularity explicit artist_name artist_popularity
    <chr>        <chr>       <dbl>           <dbl> <lg>      <chr>          <dbl>
1 3E35LyekDim... Trippy Ma...        4         0 TRUE Diplo            77
2 1oqW6G2z1mM... OMG!           1         0 TRUE Yelawolf        64
3 7mdkjz0t7B3d... Hell! 2 Fi...       1         0 TRUE Riff Raff        48
4 67w0t70B3d... Still Get...       8         30 TRUE Diplo            77
5 1xptTFRBrij... ride me 1...       2         0 TRUE Rumeilis        48
6 4ccpcCZyseq8... BLEED          1         2 FALSE Minzie          46
#> # 8 more variables: artist_followers <dbl>, artist_genres <chr>, album_id <chr>,
#> # album_name <chr>, album_release_date <chr>, album_total_tracks <dbl>, album_type <chr>,
#> # track_duration_min <dbl>
> high_popularity_subset <- subset(spotify, track_popularity > 80)

```

RStudio

```

File Edit Code View Plots Session Build Debug Profile Tools Help
Console Terminal Background Jobs
[R - R 4.5.2 - /]
> high_popularity_subset <- subset(spotify, track_popularity > 80)
> cat("Number of highly popular tracks (popularity > 80):", nrow(high_popularity_subset), "\n")
Number of highly popular tracks (popularity > 80): 592
> summary(high_popularity_subset$track_popularity)
Min. 1st Qu. Median Mean 3rd Qu. Max.
81.00 82.00 83.00 84.22 86.00 99.00
> 
> explicit_and_famous_subset <- subset(spotify, explicit == TRUE & artist_followers > 1000000)
> cat("Number of explicit tracks by artists with >1M followers:", nrow(explicit_and_famous_subset), "\n")
Number of explicit tracks by artists with >1M followers: 1798
> head(explicit_and_famous_subset)
# A tibble: 6 × 15
  track_id   track_name track_number track_popularity explicit artist_name artist_popularity
    <chr>        <chr>       <dbl>           <dbl> <lg>      <chr>          <dbl>
1 3E35LyekDim... Trippy Ma...        4         0 TRUE Diplo            77
2 1oqW6G2z1mM... OMG!           1         0 TRUE Yelawolf        64
3 67w0t70B3d... Still Get...       8         30 TRUE Diplo            77
4 2zg3tProy3GG... superstar        2         59 TRUE Artemas        74
5 1NxnbEAcPvY3... priscilla       3         80 TRUE Tame Impala      84
6 0gbmHk8SBV... PCYC HARD...       1         51 TRUE Diplo            77
#> # 8 more variables: artist_followers <dbl>, artist_genres <chr>, album_id <chr>,
#> # album_name <chr>, album_release_date <chr>, album_total_tracks <dbl>, album_type <chr>,
#> # track_duration_min <dbl>
> 
> special_tracks_subset <- subset(spotify, grep("dark r&b", tolower(artist_genres)) | track_duration_min > 4)
> cat("Number of special tracks (dark r&b OR duration > 4 mins):", nrow(special_tracks_subset), "\n")
Number of special tracks (dark r&b OR duration > 4 mins): 2205
> head(special_tracks_subset)
# A tibble: 6 × 15
  track_id   track_name track_number track_popularity explicit artist_name artist_popularity
    <chr>        <chr>       <dbl>           <dbl> <lg>      <chr>          <dbl>
1 1xptTFRBrij... ride me 1...       2         0 TRUE Rumeilis        48
2 4ccpcCZyseq8... BLEED          1         2 FALSE Minzie          46
3 3oQ3HQxtAjg... Te Procur...       1         20 FALSE AZEROK        30
4 1YEzbdt417sf... come clos...       1         27 FALSE Rumeilis        49
5 0NZKfCuA68wu... Root of a...       5         65 FALSE Daniel Cae...      85
6 66f6yfny6c... MELODY - ...      11         27 FALSE A2O             46

```

**SHETH L.U.J & SIR M.V COLLEGE OF SCIENCE
SUBJECT : Data Analysis with SAS / SPSS /R**

The screenshot shows the RStudio interface with a large watermark 'SAMPLE' diagonally across the center. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help, and Addins. The left sidebar has tabs for Source, Console, Terminal, and Background Jobs. The Source tab displays a long R script. The Environment tab shows a list of objects in the Global Environment, including various filters and datasets like album_or_single_filter, explicit_and_famous, and spotify. The Files tab lists files in the Home directory, such as RData, .Rhistory, desktop.ini, and spotify_data.csv. The bottom status bar shows the date and time as 24-11-2023 15:54, along with icons for battery, signal, and network.

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Addins ▾

Source
Console Terminal × Background Jobs ×
R - R 4.5.2 - ~/...
# 66f0yfrny6c... MELODY - ... 11 27 FALSE A20 46
# i 8 more variables: artist_followers <dbl>, artist_genres <chr>, album_id <chr>,
# album_name <chr>, album_release_date <chr>, album_total_tracks <dbl>, album_type <chr>,
# track_duration_min <dbl>
> 
> low_popularity_filter <- spotify[>
+   filter(track_popularity < 30)
> cat("Number of low-popularity tracks (< 30):", nrow(low_popularity_filter), "\n")
Number of low-popularity tracks (< 30): 1481
> summary(low_popularity_filter$track_popularity)
  Min. 1st Qu. Median 3rd Qu. Max.
  0.00  0.00  4.00 10.47 22.00 29.00
> 
> long_single_filter <- spotify[>
+   filter(album_type == "single", track_duration_min > 3.5)
+   cat("Number of singles longer than 3.5 minutes:", nrow(long_single_filter), "\n")
Number of singles longer than 3.5 minutes: 607
> head(long_single_filter)
# A tibble: 6 x 15
  track_id track_name track_number track_popularity explicit artist_name artist_popularity
  <chr>     <chr>        <dbl>      <dbl>    <chr>       <chr>        <dbl>
1 300q3HQTAjg... Te Procur... 1          20 FALSE AZERDK 30
2 2zzjpwG8w3s... Eggshell 1           52 FALSE The All-Am... 67
3 2MbcsG0tkey... Eyes on Y... 1           22 FALSE Smith & Li... 17
4 0C42YiaQy0b0... Hold On (... 1           34 FALSE Suark 32
5 1YKASTzrIp... Kay 1            2 FALSE Vince Bello 6
6 0rii4kdd4tP... Talk To Y... 1           72 TRUE Sam Fender 73
# i 18 more variables: artist_followers <dbl>, artist_genres <chr>, album_id <chr>,
# album_name <chr>, album_release_date <chr>, album_total_tracks <dbl>, album_type <chr>,
# track_duration_min <dbl>
> 
> album_or_single_filter <- spotify[>
+   filter(album_type %in% c("album", "single"))
> cat("Number of rows with album_type album or single:", nrow(album_or_single_filter), "\n")
Number of rows with album_type album or single: 8075
> table(album_or_single_filter$album_type)

album single
5856 2219
> 
```

Environment History Connections Tutorial

Import Dataset 195 MB Global Environment

Data

- album_or_single_filter 8075 obs. of 15 variables
- explicit_and_famous 1798 obs. of 15 variables
- high_popularity_sub 592 obs. of 15 variables
- long_single_filter 607 obs. of 15 variables
- low_popularity_filter 1481 obs. of 15 variables
- special_tracks_subset 2205 obs. of 15 variables
- spotify 8582 obs. of 15 variables

Files Plots Packages Help Viewer Presentation

New Folder New File Delete Rename More

Name	Size	Modified
RData	2.5 KB	Nov 24, 2025, 3:42 PM
.Rhistory	316 B	Nov 24, 2025, 3:42 PM
Custom Office Templates		
Dell		
desktop.ini	418 B	Jun 11, 2025, 2:45 AM
My Music		
My Videos		
NetBeansProjects		
Shortcut to Documents (OneDrive - Personal).lnk	1.7 KB	Jun 11, 2025, 2:45 AM
spotify_data.csv	1.3 MB	Nov 24, 2025, 3:42 PM

ENG IN 15:54 24-11-2023

