

MVLU COLLEGE

PRACTICAL NO. 7

AIM: Selecting and dropping variables using select() in R. import dataset.

The screenshot shows the RStudio interface with a script editor on the left and the Environment pane on the right. The script performs the following steps:

- 1. Load the `dplyr` library.
- 2. Read the CSV file `"C:/Users/itlab/OneDrive/Documents/S105/spotify_data_clean.csv"` into a data frame named `spotify`.
- 3. Print the first 3 rows of the dataset.
- 4. Select specific columns: `track_name`, `artist_name`, and `track_popularity`.
- 5. Print the first 3 rows of the selected columns.
- 6. Select a range of columns from `artist_name` to `album_name`.
- 7. Print the first 3 rows of the selected range.
- 8. Select columns starting with `'track'`.
- 9. Print the first 3 rows of the selected columns.
- 10. Drop the `explicit` column.
- 11. Print the names of the dropped column.
- 12. Drop multiple columns: `artist_followers` and `artist_genres`.
- 13. Print the names of the dropped columns.
- 14. Drop a range of columns from `album_id` to `album_total_tracks`.
- 15. Print the names of the dropped range.

The Environment pane on the right shows the following objects:

Object	Size
appended	298 obs. of 17 variables
Cleaned_BMW_Sales	99 obs. of 11 variables
Cleaned_Car_Price	199 obs. of 10 variables
df1	99 obs. of 11 variables
df2	199 obs. of 10 variables
dropped_multiple	4573 obs. of 13 variables
dropped_one	4573 obs. of 14 variables
dropped_range	4573 obs. of 11 variables
housing	4573 obs. of 15 variables
merged_full	294 obs. of 17 variables
merged_inner	4 obs. of 17 variables
merged_left	99 obs. of 17 variables
range_cols	4573 obs. of 6 variables
selected_cols	4573 obs. of 3 variables
spotify	4573 obs. of 15 variables
spotify_data_clean	8573 obs. of 15 variables
starts_with_track	4573 obs. of 5 variables

The screenshot shows the RStudio interface with a console window displaying an error message. The script is as follows:

```
> library(dplyr)
> spotify <- read.csv("C:/Users/itlab/OneDrive/Documents/S105/spotify_data_clean.csv")

Warning message:
In scan(file = file, what = what, sep = sep, quote = quote, dec = dec, :
  EOF within quoted string

> cat("\n--- Original Dataset (First 3 rows) ---\n")
--- Original Dataset (First 3 rows) ---
> print(head(spotify, 3))
  track_id track_name track_number track_popularity explicit artist_name
1 3E35Lyekim1Tf5rBfMz1 Trippy Mane (ft. Project Pat) 4 0 TRUE Diplo
2 1oqw6GZ1wMuq1ppP270B OMG! 1 0 TRUE Yelawolf
3 7ndkjp0ivf1r9c2p0mV Hard 2 Find 1 4 TRUE Riff Raff

  artist_popularity artist_followers artist_genres album_id album_name
1 77 2812821 moombahton 3QRfNGBmGcPBKf2XtZ5z d00mscrv11, Vol. 1
2 64 2363438 country hip hop, southern hip hop 4Summmv0xtJRcLdJczG2z2
3 48 193302 N/A 3E3ZEAL8gUvWALYB9L7gpb Hard 2 Find

  album_release_date album_total_tracks album_type track_duration_min
1 2025-10-31 9 album 1.55
2 2025-10-31 1 single 3.07
3 2025-10-31 1 single 2.55

> selected_cols <- spotify %>%
+ select(track, Artist, Streams)

Error in `select()`:
! Can't select columns that don't exist.
x column 'Track' doesn't exist.
Run `rlang::last_trace()` to see where the error occurred.

> selected_cols <- spotify %>%
+ selected_cols <- spotify %>%
+ select(track, Artist, Streams)

Error in `select()`:
! Can't select columns that don't exist.
x column 'Track' doesn't exist.
Run `rlang::last_trace()` to see where the error occurred.

> names(spotify)
```

The Environment pane on the right shows the following objects:

Object	Size
appended	298 obs. of 17 variables
Cleaned_BMW_Sales	99 obs. of 11 variables
Cleaned_Car_Price	199 obs. of 10 variables
df1	99 obs. of 11 variables
df2	199 obs. of 10 variables
dropped_multiple	4573 obs. of 13 variables
dropped_one	4573 obs. of 14 variables
dropped_range	4573 obs. of 11 variables
housing	4573 obs. of 15 variables
merged_full	294 obs. of 17 variables
merged_inner	4 obs. of 17 variables
merged_left	99 obs. of 17 variables
range_cols	4573 obs. of 6 variables
selected_cols	4573 obs. of 3 variables
spotify	4573 obs. of 15 variables
spotify_data_clean	8573 obs. of 15 variables
starts_with_track	4573 obs. of 5 variables

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

Go to file/function
Addins

Source
Console
Background Jobs

R - R452 - ~/r

```

> names(spotify)
[1] "track_id"           "track_name"           "track_number"         "track_popularity"     "explicit"
[6] "artist_name"        "artist_popularity"    "artist_followers"     "artist_genres"        "album_id"
[11] "album_name"         "album_release_date"   "album_total_tracks"   "album_type"           "track_duration_min"
> selected_cols <- spotify %>%
+ select(track_name, artist_name, track_popularity)
> print(head(selected_cols, 3))
  track_name artist_name track_popularity
1 Trippy Mane (ft. Project Pat) Diplo 0
2 OMG! Yelawolf 0
3 Hard 2 Find Riff Raff 4
> range_cols <- spotify %>%
+ select(artist_name,album_name)
> print(head(range_cols, 3))
  artist_name album_name
1 Diplo 77 2812821
2 Yelawolf 64 2363438 country hip hop, southern hip hop
3 Riff Raff 48 193302 N/A 3E32EAL8UYwALYB9L7gpb
> album_name
1 d0m5scrV1l, vol. 1
2 OMG!
3 Hard 2 Find
>
> starts_with_track <- spotify %>%
+ select(starts_with("track"))
> print(head(starts_with_track, 3))
  track_id track_name track_number track_popularity track_duration_min
1 3E355yek0Im1f5r6nc1 Trippy Mane (ft. Project Pat) 4 1.55
2 1oq6G62iWumq1pp2708 OMG! 1 0 3.07
3 7mdkjzoiY1f1rx9Etapmu Hard 2 Find 1 4 2.55
> dropped_one <- spotify %>%
+ select(-explicit)
> print(names(dropped_one))
[1] "track_id"           "track_name"           "track_number"         "track_popularity"     "artist_name"
[6] "artist_popularity"  "artist_followers"     "artist_genres"        "album_id"             "album_name"
[11] "album_release_date" "album_total_tracks"   "album_type"           "track_duration_min"
> dropped_multiple <- spotify %>%
+ select(-(artist_followers, ~artist_genres))
> print(names(dropped_multiple))
[1] "track_id"           "track_name"           "track_number"         "track_popularity"     "explicit"
[6] "artist_name"        "artist_popularity"    "album_id"            "album_name"           "album_release_date"
[11] "album_total_tracks" "album_type"           "track_duration_min"

```

Environment
History
Connections
Tutorial
R - Global Environment
Data

- appended 298 obs. of 17 variables
- cleaned_BMW_Sales 99 obs. of 11 variables
- cleaned_Car_Price 199 obs. of 10 variables
- df1 99 obs. of 11 variables
- df2 199 obs. of 10 variables
- dropped_multiple 4573 obs. of 17 variables
- dropped_one 4573 obs. of 14 variables
- dropped_range 4573 obs. of 11 variables
- housing 4573 obs. of 15 variables
- merged_full 294 obs. of 17 variables
- merged_inner 4 obs. of 17 variables
- merged_left 99 obs. of 17 variables
- range_cols 4573 obs. of 6 variables
- selected_cols 4573 obs. of 3 variables
- spotify 4573 obs. of 15 variables
- spotify_data_clean 8573 obs. of 15 variables
- starts_with_track 4573 obs. of 5 variables

values

- key "vehicle_id"
- keys chr [1:4] "model" "year" "fuel_type" "transmission"

Files
Plots
Packages
Help
Viewer
Presentation

27°C Sunny

11:56 01-12-2022

The screenshot displays the RStudio environment with the following components:

- Source Editor:** Contains R code for data manipulation using `dplyr` and `tidyr`. The code includes:


```
track_name artist_name track_popularity
1 Trippy Mane (ft. Project Pat) Diplo 0
2 OMG! Yelawolf 0
3 Hard 2 Find Riff Raff 4
> range_cols <- spotify %>%
+ select(artist_name:album_name)
+ print(head(range_cols, 3))
artist_name artist_popularity artist_followers artist_genres album_id
1 Diplo 77 2812821 moombaton SQRNGBw4GPaBkF2xt752
2 Yelawolf 64 2363438 country hip hop, southern hip hop 4USummv0XTjRcdJcz66g
3 Riff Raff 48 193302 N/A 3E2EA8gUYwalyB9L7gpb
album_name
1 d00mscrvll, vol. 1
2 OMG!
3 Hard 2 Find
>
> starts_with_track <- spotify %>%
+ select(starts_with("track"))
+ print(head(starts_with_track, 3))
track_id track_name track_number track_popularity track_duration_min
1 3E255lyek0im1Tf5rBfMz1 Trippy Mane (ft. Project Pat) 4 0 1.55
2 1oqW6GZ1wUmq1Pp27D08 "track_name" "track_number" "track_popularity" "artist_name"
[6] "artist_popularity" "artist_followers" "artist_genres" "album_name"
[1] "album_release_date" "album_total_tracks" "album_type" "track_duration_min"
> dropped_multiple <- spotify %>%
+ select(-artist_followers, -artist_genres)
+ print(names(dropped_multiple))
[1] "track_id" "track_name" "track_number" "track_popularity" "explicit"
[6] "artist_name" "artist_popularity" "album_id" "album_name" "album_release_date"
[11] "album_total_tracks" "album_type" "track_duration_min"
> dropped_range <- spotify %>%
+ select(-(album_id:album_total_tracks))
+ print(names(dropped_range))
[1] "track_id" "track_name" "track_number" "track_popularity" "explicit"
[6] "artist_name" "artist_popularity" "artist_followers" "artist_genres" "album_type"
[11] "track_duration_min"
>
```
- Console:** Shows the output of the R code, including the first few rows of the data frames and the names of the columns in the resulting data objects.
- Environment:** Lists the objects created in the R session, such as `spotify`, `range_cols`, `starts_with_track`, `dropped_multiple`, and `dropped_range`.
- Global Environment:** Shows the global environment with the `spotify` object.
- Files:** Displays the files in the current project directory.
- Plots:** Shows any plots generated by the code.
- Packages:** Lists the packages loaded in the R session.
- Help:** Provides access to R documentation.
- Viewer:** Displays any plots or output generated by the code.
- Presentation:** Provides a presentation view of the R code and output.