

PRACTICAL NO 7

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

OUTPUT:-

1.IMPORT DATASET:

The screenshot displays the RStudio interface. The script editor on the left contains the following R code:

```
1- # 1. IMPORT DATASET
2- library(dplyr) # select() is part of dplyr
3- # Import the NYC flights CSV
4- flights <- read.csv("nyc_flights.csv")
5- cat("\n--- Original Dataset (First 3 rows) ---\n")
6- print(head(flights, 3))
7-
8- # 2. SELECTING VARIABLES (Keeping Columns)
```

The console on the bottom left shows the output of the code:

```
> library(dplyr)
> flights <- read.csv("nyc_flights.csv")
> cat("\n--- Original Dataset (First 3 rows) ---\n")
--- Original Dataset (First 3 rows) ---
> print(head(flights, 3))
  year month day dep_time arr_time carrier flight tailnum origin dest air_time distance dep_delay arr_delay
1 2013     7   4      875      222    AS   7380 N31689   JFK   MIA       48        313       116       169
2 2013     4  18     2344     1271    AA   6124 N38566   LGA   SFO       339       2370       103       239
3 2013    11  22     1774     2301    DL   9909 N23838   LGA   SFO       359       4192        66       103
```

The Environment pane on the right shows the loaded objects:

Object	Size
flights	120000 obs. of 14 variables
jan	10 obs. of 3 variables
merged	10 obs. of 5 variables
merged_data	10 obs. of 5 variables
newp	4 obs. of 3 variables
nyc_flights	120000 obs. of 14 variables
range_cols	120000 obs. of 3 variables
selected_cols	120000 obs. of 3 variables
starts_with_d	120000 obs. of 5 variables

The Files pane at the bottom shows the project directory structure, including the nyc_flights.csv file.

2. SELECTING VARIABLES (Keeping Columns)

The screenshot shows the RStudio interface with the following content:

Console:

```
R - R 4.5.2 - C:/Users/as993/DATA SET/
> selected_cols <- flights %>%
+   select(carrier, flight, distance)
> cat("\n--- Selected Specific Columns (carrier, flight, distance) ---\n")

--- Selected Specific Columns (carrier, flight, distance) ---
> print(head(selected_cols, 3))
  carrier flight distance
1     AS    7380     313
2     AA   6124    2370
3     DL   9909    4192

> range_cols <- flights %>%
+   select(year:day)
> cat("\n--- Selected Range of Columns (year to day) ---\n")

--- Selected Range of Columns (year to day) ---
> print(head(range_cols, 3))
  year month day
1 2013     7   4
2 2013     4  18
3 2013    11  22

> starts_with_d <- flights %>%
+   select(starts_with("d"))
> cat("\n--- Selected Columns Starting with 'd' ---\n")

--- Selected Columns Starting with 'd' ---
> print(head(starts_with_d, 3))
  day dep_time dest distance dep_delay
1   4      875  MIA      313         116
2  18     2344  SFO     2370         103
3  22     1774  SFO     4192          66
```

Environment:

Name	Size	Modified
flights	120000 obs. of 14 variables	
jan	10 obs. of 3 variables	
merged	10 obs. of 5 variables	
merged_data	10 obs. of 5 variables	
newp	4 obs. of 3 variables	
nyc_flights	120000 obs. of 14 variables	
range_cols	120000 obs. of 3 variables	
selected_cols	120000 obs. of 3 variables	
starts_with_d	120000 obs. of 5 variables	

3. DROPPING VARIABLES (Removing Columns)

The screenshot shows the RStudio interface with the following content:

Console:

```
PRACTICAL NO 7.R
# Example: day, dep_time, dep_delay, dest, distance
> starts_with_d <- flights %>%
+   select(starts_with("d"))
> cat("\n--- Dataset with 'tailnum' dropped ---\n")

--- Dataset with 'tailnum' dropped ---
> print(names(dropped_one))
[1] "year"      "month"     "day"       "dep_time"  "arr_time"  "carrier"   "flight"   "origin"
[9] "dest"      "air_time"  "distance"  "dep_delay" "arr_delay"

> dropped_multiple <- flights %>%
+   select(-dep_delay, -arr_delay)
> cat("\n--- Dataset with 'dep_delay' and 'arr_delay' dropped ---\n")

--- Dataset with 'dep_delay' and 'arr_delay' dropped ---
> print(names(dropped_multiple))
[1] "year"      "month"     "day"       "dep_time"  "arr_time"  "carrier"   "flight"   "tailnum"  "origin"
[10] "dest"      "air_time"  "distance"

> dropped_range <- flights %>%
+   select(-(air_time:arr_delay))
> cat("\n--- Dataset with range 'air_time' to 'arr_delay' dropped ---\n")

--- Dataset with range 'air_time' to 'arr_delay' dropped ---
> print(names(dropped_range))
[1] "year"      "month"     "day"       "dep_time"  "arr_time"  "carrier"   "flight"   "tailnum"  "origin"
[10] "dest"
```

Environment:

Name	Size	Modified
dropped_one	120000 obs. of 13 variables	
dropped_range	120000 obs. of 10 variables	
feb	10 obs. of 3 variables	
final	14 obs. of 3 variables	
final_list	14 obs. of 3 variables	
flights	120000 obs. of 14 variables	
jan	10 obs. of 3 variables	
merged	10 obs. of 5 variables	
merged_data	10 obs. of 5 variables	