

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

## SUBJECT :- Data analysis of SAS/SPSS/R

### PRACTICAL – 4

The image displays two screenshots of the RStudio interface, showing the execution of R code for data analysis. The top screenshot shows the installation of the 'dplyr' package and the loading of the 'airquality' dataset. The bottom screenshot shows the execution of various data manipulation functions like 'subset', 'filter', and 'summarize' on the 'airquality' dataset.

**Top Screenshot:**

```
R - R452 - ~/R
> install.packages("dplyr")
WARNING: Rtools is required to build a 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/Trupti/AppData/Local/R/win-library/4.5'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/dplyr_1.1.4.zip'
Content type 'application/zip' length 1593482 bytes (1.5 MB)
downloaded 1.5 MB
package 'dplyr' successfully unpacked and MD5 sums checked
The downloaded binary packages are in
C:/Users/Trupti/AppData/Local/Temp/RtmpIpxett/downloaded_packages
> 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)
> head(airquality)
# A tibble: 6 x 15
  timestamp          country city    latitude longitude  pm25 pm10  no2  so2  o3  co
<dtm>          <chr>    <chr>    <dbl>    <dbl>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 2025-11-04 18:25:17 US      New York  40.7    -74.0    50.3 109.  28.0  6.54 52.6 1.10
2 2025-11-04 19:25:17 US      New York  40.7    -74.0    32.1  63.0  36.1  4.02 43.5 1.08
3 2025-11-04 20:25:17 US      New York  40.7    -74.0    42.2  82.6  26.9  9.54 23.3 0.977
4 2025-11-04 21:25:17 US      New York  40.7    -74.0    30.4  80.0  63.5  7.61 31.4 0.23
5 2025-11-04 22:25:17 US      New York  40.7    -74.0    21.1  66.4  39.0  6.92 45.6 1.08
```

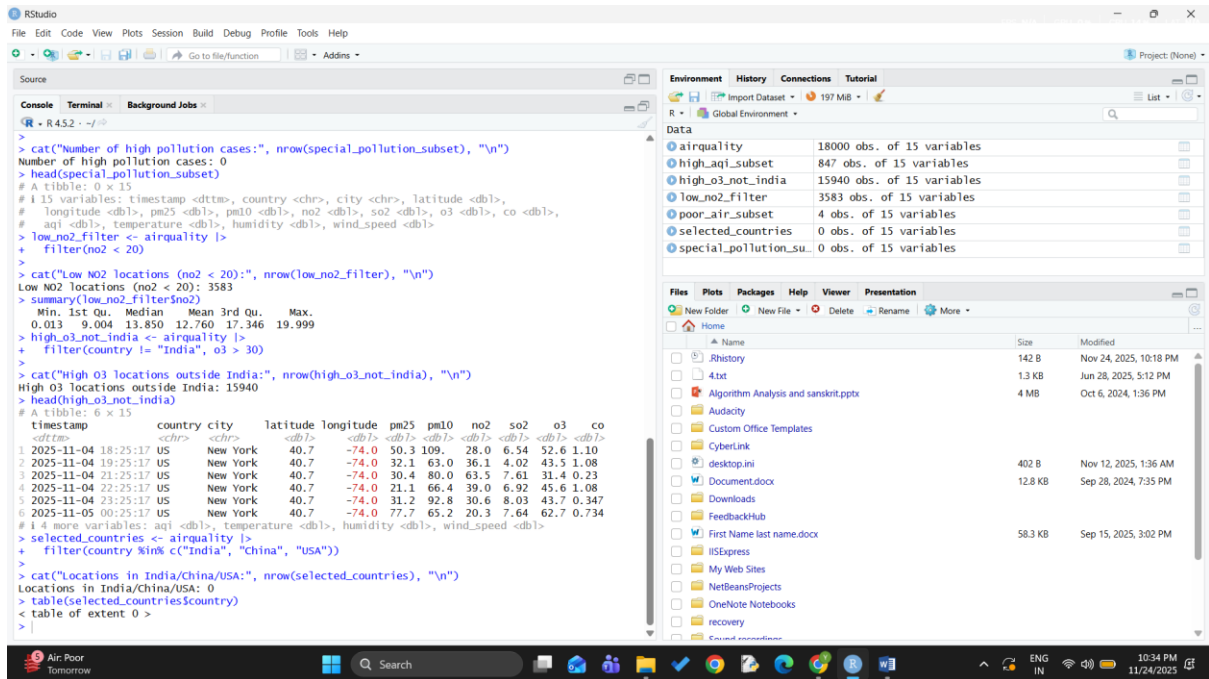
**Bottom Screenshot:**

```
R - R452 - ~/R
> subset(airquality, aqi > 150)
# A tibble: 4 x 15
  timestamp          country city    latitude longitude  pm25 pm10  no2  so2  o3  co
<dtm>          <chr>    <chr>    <dbl>    <dbl>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 2025-11-05 19:25:17 GB      London  51.5    -0.128  107.  43.8  13.4  4.60 21.7 0.89
2 2025-11-17 07:25:17 JP      Osaka   34.7    136.  102.  95.2  41.7  7.81 29.0 0.753
3 2025-11-11 05:25:17 PH      Manila  14.6    121.  103.  66.7  43.3  2.17 48.2 0.849
4 2025-11-15 03:25:17 FI      Helsinki 60.2    24.9  116.  81.7  54.8  6.69 39.6 0.685
# 4 more variables: aqi <dbl>, temperature <dbl>, humidity <dbl>, wind_speed <dbl>
> special_pollution_subset <- subset(airquality, pm25 > 120 | pm10 > 180)
> cat("Number of high pollution cases:", nrow(special_pollution_subset), "\n")
Number of high pollution cases: 0
> head(special_pollution_subset)
# A tibble: 0 x 15
# 15 variables: timestamp <dtm>, country <chr>, city <chr>, latitude <dbl>,
# longitude <dbl>, pm25 <dbl>, pm10 <dbl>, no2 <dbl>, so2 <dbl>, o3 <dbl>, co <dbl>,
# aqi <dbl>, temperature <dbl>, humidity <dbl>, wind_speed <dbl>
> low_no2_filter <- airquality |>
+ filter(no2 < 20)
> cat("Low NO2 locations (no2 < 20):", nrow(low_no2_filter), "\n")
Low NO2 locations (no2 < 20): 3583
> summary(low_no2_filter$no2)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.013  9.004  13.850  12.760  17.346  19.999
```

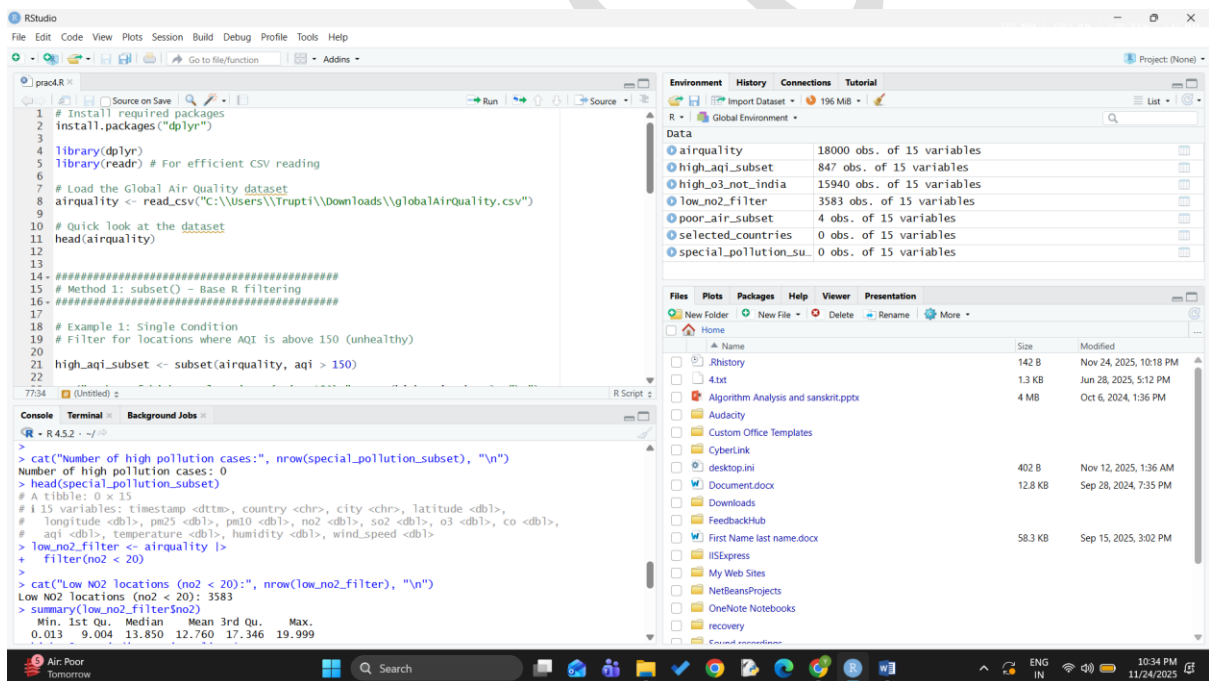
# SHETH L.U.J AND SIR M.V. COLLEGE

## SUBJECT :- Data analysis of SAS/SPSS/R

### PRACTICAL – 4



```
> cat("Number of high pollution cases:", nrow(special_pollution_subset), "\n")
Number of high pollution cases: 0
> head(special_pollution_subset)
# A tibble: 0 x 15
#   timestamp <dtm> country <chr> city <chr> latitude <dbl>
#   longitude <dbl> pm25 <dbl> pm10 <dbl> no2 <dbl> so2 <dbl>
#   aqi <dbl> temperature <dbl> humidity <dbl> wind_speed <dbl>
#   low_no2_filter <airquality>
#   filter(no2 < 20)
>
> cat("Low NO2 locations (no2 < 20):", nrow(low_no2_filter), "\n")
Low NO2 locations (no2 < 20): 3583
> summary(low_no2_filter$no2)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.013  9.004  13.850  12.760  17.346  19.999
> high_o3_not_india <- airquality |>
+ filter(country != "India", o3 > 30)
>
> cat("High O3 locations outside India:", nrow(high_o3_not_india), "\n")
High O3 locations outside India: 15940
> head(high_o3_not_india)
# A tibble: 6 x 15
#   timestamp <dtm> country <chr> city <chr> latitude <dbl>
#   longitude <dbl> pm25 <dbl> pm10 <dbl> no2 <dbl> so2 <dbl>
#   o3 <dbl> co <dbl>
1 2025-11-04 18:25:17 US New York 40.7 -74.0 30.3 109. 28.0 6.54 52.6 1.10
2 2025-11-04 19:25:17 US New York 40.7 -74.0 32.1 63.0 36.1 4.02 43.5 1.08
3 2025-11-04 21:25:17 US New York 40.7 -74.0 30.4 80.0 63.5 7.61 31.4 0.23
4 2025-11-04 22:25:17 US New York 40.7 -74.0 21.1 66.4 39.0 6.92 45.6 1.08
5 2025-11-04 23:25:17 US New York 40.7 -74.0 31.2 92.8 30.6 8.03 43.7 0.347
6 2025-11-05 00:25:17 US New York 40.7 -74.0 77.7 65.2 20.3 7.64 62.7 0.734
# 4 more variables: aqi <dbl>, temperature <dbl>, humidity <dbl>, wind_speed <dbl>
> selected_countries <- airquality |>
+ filter(country %in% c("India", "China", "USA"))
>
> cat("Locations in India/China/USA:", nrow(selected_countries), "\n")
Locations in India/China/USA: 0
> table(selected_countries$country)
<table of extent 0>
> |
```



```
1 # Install required packages
2 install.packages("dplyr")
3
4 library(dplyr)
5 library(readr) # For efficient CSV reading
6
7 # Load the Global Air Quality dataset
8 airquality <- read_csv("C:\\Users\\Trupti\\Downloads\\GlobalAirQuality.csv")
9
10 # Quick look at the dataset
11 head(airquality)
12
13
14 #####
15 # Method 1: subset() - Base R Filtering
16 #####
17
18 # Example 1: Single Condition
19 # Filter for locations where AQI is above 150 (unhealthy)
20
21 high_aqi_subset <- subset(airquality, aqi > 150)
22
```

```
> cat("Number of high pollution cases:", nrow(special_pollution_subset), "\n")
Number of high pollution cases: 0
> head(special_pollution_subset)
# A tibble: 0 x 15
#   timestamp <dtm> country <chr> city <chr> latitude <dbl>
#   longitude <dbl> pm25 <dbl> pm10 <dbl> no2 <dbl> so2 <dbl>
#   aqi <dbl> temperature <dbl> humidity <dbl> wind_speed <dbl>
#   low_no2_filter <airquality>
#   filter(no2 < 20)
>
> cat("Low NO2 locations (no2 < 20):", nrow(low_no2_filter), "\n")
Low NO2 locations (no2 < 20): 3583
> summary(low_no2_filter$no2)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.013  9.004  13.850  12.760  17.346  19.999
```

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

**SUBJECT :- Data analysis of SAS/SPSS/R**  
**PRACTICAL – 4**

Yash S072