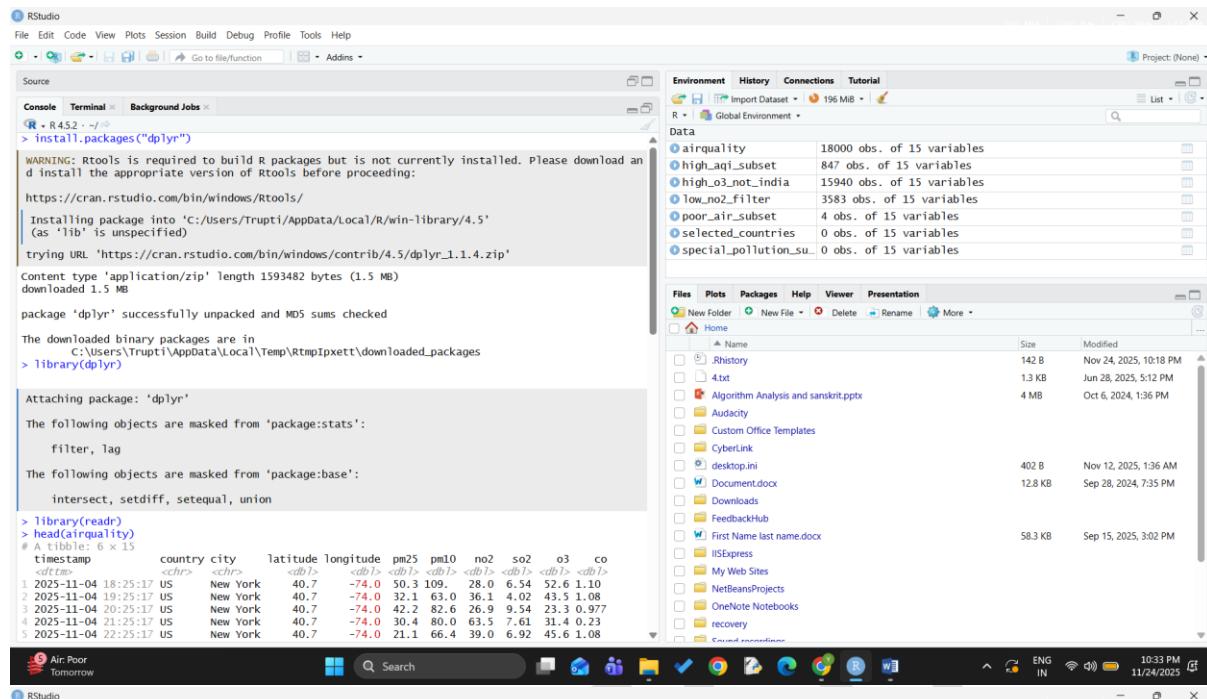


# SHETH L.U.J AND SIR M.V. COLLEGE

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

### PRACTICAL – 4



```
R 4.5.2 - ~/R
> 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/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)

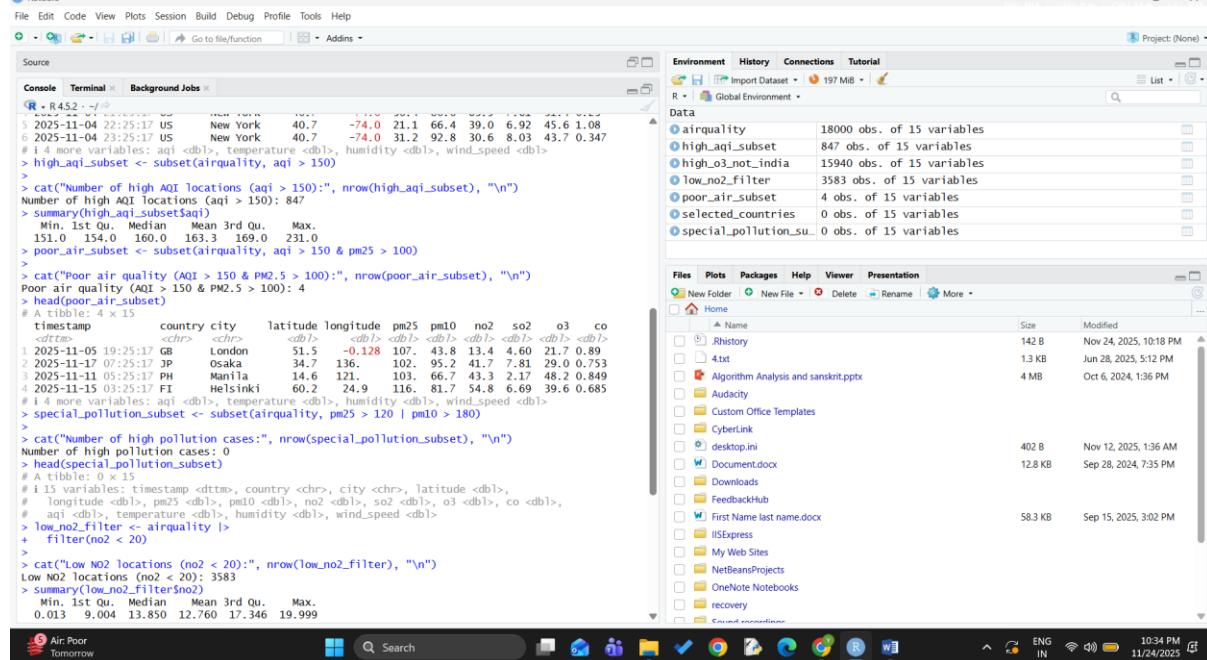
Attaching package: 'dplyr'

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 × 15
#>   timestamp    country city  latitude longitude pm25 pm10 no2 so2 o3 co
#>   <dttm>      <chr>  <chr>     <dbl>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
#> 1 2025-11-04 18:25:17 US  New York  40.7    -74.0  50.3  100.0  26.0  9.54  43.5  1.10
#> 2 2025-11-04 19:25:17 US  New York  40.7    -74.0  50.3  100.0  26.0  9.54  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
#> # ... with 13 more variables: aqi <dbl>, temperature <dbl>, humidity <dbl>, wind_speed <dbl>
```

The screenshot shows the RStudio interface with the console tab active. The user has run the command `install.packages("dplyr")` and attached the `dplyr` package. They then load the `airquality` dataset and use the `head` function to view the first six rows. The data includes columns for timestamp, country, city, latitude, longitude, and various air quality parameters like pm25, pm10, no2, so2, o3, and co.

```
R 4.5.2 - ~/R
> cat("Number of high AQI locations (aqi > 150):", nrow(high_aqi_subset), "\n")
Number of high AQI locations (aqi > 150): 847
> summary(high_aqi_subset$aqi)
  Min. 1st Qu. Median  Mean 3rd Qu. Max.
151.0 154.0 160.0 163.3 169.0 231.0
> poor_air_subset <- subset(airquality, aqi > 150 & pm25 > 100)
> head(poor_air_subset)
#> # A tibble: 4 × 15
#>   timestamp    country city  latitude longitude pm25 pm10 no2 so2 o3 co
#>   <dttm>      <chr>  <chr>     <dbl>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
#> 1 2025-11-17 09:25:17 GB  London  51.5    -0.128 107. 43.8 13.4  4.60 21.7  0.89
#> 2 2025-11-17 10:25:17 JP  Osaka   34.7    136. 102. 95.2  41.7  7.81 29.0  0.753
#> 3 2025-11-17 11:25:17 PH  Manila   14.6   121. 103. 66.7  43.3  2.17 48.2  0.849
#> 4 2025-11-17 12:25:17 FI  Helsinki 60.2   21.9 116. 81.7  3.8  6.69 39.6  0.685
#> # ... with 13 more variables: aqi <dbl>, temperature <dbl>, humidity <dbl>, wind_speed <dbl>
> special_pollution_subset <- subset(airquality, pm25 > 120 | pm10 > 180)
> head(special_pollution_subset)
#> # A tibble: 0 × 15
#> # i 15 variables: timestamp <dttm>, country <chr>, city <chr>, latitude <dbl>,
#> #   longitude <dbl>, pm25 <dbl>, pm10 <dbl>, no2 <dbl>, so2 <dbl>, o3 <dbl>, co <dbl>,
#> #   aqi <dbl>, temperature <dbl>, humidity <dbl>, wind_speed <dbl>
> low_no2_filter <- airquality[no2 < 20]
> filter(no2 < 20)
>
> cat("Low NO2 locations (no2 < 20):", nrow(low_no2_filter), "\n")
Low NO2 locations (no2 < 20): 3583
> summary(low_no2_filter$no2)
  Min. 1st Qu. Median  Mean 3rd Qu. Max.
0.013  9.004 13.850 12.760 17.346 19.999
```

This screenshot continues the data analysis. It creates two subsets: `high\_aqi\_subset` for locations where AQI is greater than 150, and `poor\_air\_subset` for locations where both AQI and PM2.5 are above their respective thresholds. It also creates a subset for locations with low NO2 levels. The code uses logical operators and the `filter` function to manipulate the data frame.

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

**PRACTICAL – 4**

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

Source

```
> cat("Number of high pollution cases:", nrow(special_pollution_subset), "\n")
Number of high pollution cases: 0
> head(special_pollution_subset)
# A tibble: 0 × 15
#> # ... with 15 variables: timestamp <dttm>, 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$no2)
   Min. 1st Qu. Median  Mean 3rd Qu.  Max.
0.013  9.004 13.850 12.760 17.346 19.999
> high_o3_not_india <- airquality[>
+   filter(country != "India", o3 > 30)
>
> cat("High O3 locations outside India:", nrow(high_o3_not_india), "\n")
High O3 locations outside India: 15940
> head(high_o3_not_india)
# A tibble: 6 × 15
  timestamp    country    city    latitude    longitude    pm25    pm10    no2    so2    o3    co
<dttm> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 2025-11-04 18:25:17 US New York  40.7  -74.0  50.3  109.  28.0  6.54  52.6  1.10
2 2025-11-04 19:25:17 US New York  40.7  -74.0  32.1  63.0  36.1  4.02  43.5  1.08
3 2025-11-04 21:25:17 US New York  40.7  -74.0  30.4  80.0  63.5  7.61  31.4  0.23
4 2025-11-04 22:25:17 US New York  40.7  -74.0  21.1  66.4  39.0  6.92  45.6  1.08
5 2025-11-04 23:25:17 US New York  40.7  -74.0  31.2  92.8  30.6  8.03  43.7  0.347
6 2025-11-04 23:25:17 US New York  40.7  -74.0  27.7  65.2  20.3  7.64  62.7  0.734
# i 4 more variables: aqi <dbl>, temperature <dbl>, humidity <dbl>, wind_speed <dbl>
> selected_countries <- airquality[>
+   filter(country %in% c("India", "China", "USA"))
>
> cat("Locations in India/China/USA: ", nrow(selected_countries), "\n")
Locations in India/China/USA: 0
> table(selected_countries$country)
# table of extent 0 >
```

Environment History Connections Tutorial

R - Global Environment

Data

Name	Number of observations	Number of variables
airquality	18000	15
high_aqi_subset	847	15
high_o3_not_india	15940	15
low_no2_filter	3583	15
poor_air_subset	4	15
selected_countries	0	15
special_pollution_su	0	15

Files Plots Packages Help Viewer Presentation

New Folder New File Delete Rename More

Home

- 📁 Rhistory 142 B Nov 24, 2025, 10:18 PM
- 📄 4.txt 1.3 KB Jun 28, 2025, 5:12 PM
- 📅 Algorithm Analysis and sanskrit.pptx 4 MB Oct 6, 2024, 1:36 PM
- 📁 Audacity
- 📁 Custom Office Templates
- 📁 CyberLink
- .desktop.ini 402 B Nov 12, 2025, 1:36 AM
- .Document.docx 12.8 KB Sep 28, 2024, 7:35 PM
- Downloads
- 📁 FeedbackHub
- First Name last.name.docx 58.3 KB Sep 15, 2025, 3:02 PM
- IFIExpress
- My Web Sites
- NetBeansProjects
- OneNote Notebooks
- recovery

Air Pollution Tomorrow

Search

10:34 PM ENG IN 11/24/2025

The screenshot shows the RStudio interface with the following details:

- Code Editor:** The left pane displays the script `prac4.R` containing R code for data analysis. The code includes installing packages, loading a dataset, and performing filtering operations.
- Environment View:** The top right pane shows the global environment with objects like `airquality`, `high_aqi_subset`, and `low_no2_filter`.
- File Explorer:** The bottom right pane shows the file system structure, including files like `J.history`, `4.txt`, and `Algorithm Analysis and sanskrit.pptx`.
- Console:** The bottom left pane shows the R console output, detailing the execution of the R code.

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

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

**PRACTICAL – 4**

**Yash - S072**