

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

SUBJECT :- Data analysis of SAS/SPSS/R

PRACTICAL – 5

The image shows two separate instances of the RStudio IDE running on a Windows operating system. Both instances have the title bar 'RStudio' and are connected to the same project, indicated by the 'Project (None)' dropdown.

Session 1 (Top Window):

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Console Terminal < Background Jobs
R + R 4.5.2 - r
> source("~/active-rstudio-document", echo = TRUE)
> # Load dplyr package
> library(dplyr)
> # Load the Global Air Quality dataset
> airquality <- read.csv("C:\\Users\\Trupti\\Downloads\\globalAirQuality.csv")
Rows: 18000 Columns: 15
-- Column specification --
Delimiter: ","
chr (2): country, city
dbl (12): latitude, longitude, pm25, pm10, no2, so2, o3, co, aqi, temperature, humidity, w...
dttm (1): timestamp
# 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.

# Example 1: Sorting by a Single Variable (Ascending)
# ... [TRUNCATED]

# Show lowest AQI values
head(airquality_sorted_aqi, 5)
# A tibble: 5 × 15
  timestamp    country city   latitude longitude pm25 pm10 no2   so2   o3   co
  <dttm>      <chr>   <chr>     <dbl>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 2015-11-14 18:25:17 GB London 51.5 -0.128 8.12 11.8 6.11 4.21 1.6 0.58
2 2025-11-11 09:25:17 US San Fran... 37.8 -120.2 13.2 17.78 8.54 3.5 5.1 0.294
3 2025-11-11 23:25:17 US Los Angel... 34.0 118.2 18.7 22.4 10.35 3.06 42.1 1.17
4 2025-11-12 20:25:17 CN Shanghai 31.2 121. 1.47 16.1 11.1 3.26 40.5 1.09
5 2025-11-12 23:25:17 TH Bangkok 13.8 101. 8.85 27.7 6.86 4.29 28.0 0.67
# i 4 more variables: aqi <dbl>, temperature <dbl>, humidity <dbl>, wind_speed <dbl>

# Example 2: Sorting by a Single Variable (Descending)
# ... [TRUNCATED]

# Show highest PM2.5 cities
head(airquality_sorted_pm25_desc, 5)
# A tibble: 5 × 15
  timestamp    country city   latitude longitude pm25 pm10 no2   so2   o3   co
  <dttm>      <chr>   <chr>     <dbl>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 2025-11-11 09:25:17 US New York 40.7 -74.0 33.8 12.0 12.0 1.4 4.8 0.58
2 2025-11-11 23:25:17 CN Shanghai 31.2 121. 1.47 16.1 11.1 3.26 40.5 1.09
3 2025-11-12 20:25:17 CN Beijing 39.9 116.3 35.2 24.3 20.0 1.4 4.8 0.58
4 2025-11-12 23:25:17 TH Bangkok 13.8 101. 8.85 27.7 6.86 4.29 28.0 0.67
5 2025-11-12 23:25:17 TH Bangkok 13.8 101. 8.85 27.7 6.86 4.29 28.0 0.67
# i 4 more variables: aqi <dbl>, temperature <dbl>, humidity <dbl>, wind_speed <dbl>

# Example 3: Sorting by Multiple Columns
# ... [TRUNCATED]

# First 10 rows show countries alphabetically,
# and within each country, highest AQI first
head(airquality_multi_sort, 10)
# A tibble: 10 × 15
  timestamp    country city   latitude longitude pm25 pm10 no2   so2   o3   co
  <dttm>      <chr>   <chr>     <dbl>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 2025-11-15 03:25:17 FI Helsinki 60.2 24.9 116. 81.7 54.8 6.69 39.6 0.685
2 2025-11-15 09:25:17 GB London 51.5 -0.128 107. 43.8 13.4 4.60 21.7 0.89
3 2025-11-15 05:25:17 PH Manila 14.6 121. 103. 66.7 43.3 2.17 48.2 0.849
4 2025-11-17 07:25:17 JP Osaka 34.7 136. 102. 95.2 41.7 7.81 29.0 0.753
5 2025-11-15 23:25:17 RU Moscow 55.8 37.6 98.9 5.46 36.5 5.95 46.5 0.886
# i 4 more variables: aqi <dbl>, temperature <dbl>, humidity <dbl>, wind_speed <dbl>

# Example 4: Combined Filter + Sort
# ... [TRUNCATED]

cat("Top 5 polluted locations (PM10 > 100) with lowest CO levels:\n")
Top 5 polluted locations (PM10 > 100) with lowest CO levels:
# i 4 more variables: aqi <dbl>, temperature <dbl>, humidity <dbl>, wind_speed <dbl>

```

Session 2 (Bottom Window):

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Console Terminal < Background Jobs
R + R 4.5.2 - r
> source("~/active-rstudio-document", echo = TRUE)
> # Load dplyr package
> library(dplyr)
> # Load the Global Air Quality dataset
> airquality <- read.csv("C:\\Users\\Trupti\\Downloads\\globalAirQuality.csv")
Rows: 18000 Columns: 15
-- Column specification --
Delimiter: ","
chr (2): country, city
dbl (12): latitude, longitude, pm25, pm10, no2, so2, o3, co, aqi, temperature, humidity, w...
dttm (1): timestamp
# 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.

# Example 1: Sorting by a Single Variable (Ascending)
# ... [TRUNCATED]

# Show lowest AQI values
head(airquality_sorted_aqi, 5)
# A tibble: 5 × 15
  timestamp    country city   latitude longitude pm25 pm10 no2   so2   o3   co
  <dttm>      <chr>   <chr>     <dbl>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 2015-11-14 18:25:17 GB London 51.5 -0.128 8.12 11.8 6.11 4.21 1.6 0.58
2 2025-11-11 09:25:17 US San Fran... 37.8 -120.2 13.2 17.78 8.54 3.5 5.1 0.294
3 2025-11-11 23:25:17 US Los Angel... 34.0 118.2 18.7 22.4 10.35 3.06 42.1 1.17
4 2025-11-12 20:25:17 CN Shanghai 31.2 121. 1.47 16.1 11.1 3.26 40.5 1.09
5 2025-11-12 23:25:17 TH Bangkok 13.8 101. 8.85 27.7 6.86 4.29 28.0 0.67
# i 4 more variables: aqi <dbl>, temperature <dbl>, humidity <dbl>, wind_speed <dbl>

# Example 2: Sorting by a Single Variable (Descending)
# ... [TRUNCATED]

# Show highest PM2.5 cities
head(airquality_sorted_pm25_desc, 5)
# A tibble: 5 × 15
  timestamp    country city   latitude longitude pm25 pm10 no2   so2   o3   co
  <dttm>      <chr>   <chr>     <dbl>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 2025-11-11 09:25:17 US New York 40.7 -74.0 33.8 12.0 12.0 1.4 4.8 0.58
2 2025-11-11 23:25:17 CN Shanghai 31.2 121. 1.47 16.1 11.1 3.26 40.5 1.09
3 2025-11-12 20:25:17 CN Beijing 39.9 116.3 35.2 24.3 20.0 1.4 4.8 0.58
4 2025-11-12 23:25:17 TH Bangkok 13.8 101. 8.85 27.7 6.86 4.29 28.0 0.67
5 2025-11-12 23:25:17 TH Bangkok 13.8 101. 8.85 27.7 6.86 4.29 28.0 0.67
# i 4 more variables: aqi <dbl>, temperature <dbl>, humidity <dbl>, wind_speed <dbl>

# Example 3: Sorting by Multiple Columns
# ... [TRUNCATED]

# First 10 rows show countries alphabetically,
# and within each country, highest AQI first
head(airquality_multi_sort, 10)
# A tibble: 10 × 15
  timestamp    country city   latitude longitude pm25 pm10 no2   so2   o3   co
  <dttm>      <chr>   <chr>     <dbl>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 2025-11-15 03:25:17 FI Helsinki 60.2 24.9 116. 81.7 54.8 6.69 39.6 0.685
2 2025-11-15 09:25:17 GB London 51.5 -0.128 107. 43.8 13.4 4.60 21.7 0.89
3 2025-11-15 05:25:17 PH Manila 14.6 121. 103. 66.7 43.3 2.17 48.2 0.849
4 2025-11-17 07:25:17 JP Osaka 34.7 136. 102. 95.2 41.7 7.81 29.0 0.753
5 2025-11-15 23:25:17 RU Moscow 55.8 37.6 98.9 5.46 36.5 5.95 46.5 0.886
# i 4 more variables: aqi <dbl>, temperature <dbl>, humidity <dbl>, wind_speed <dbl>

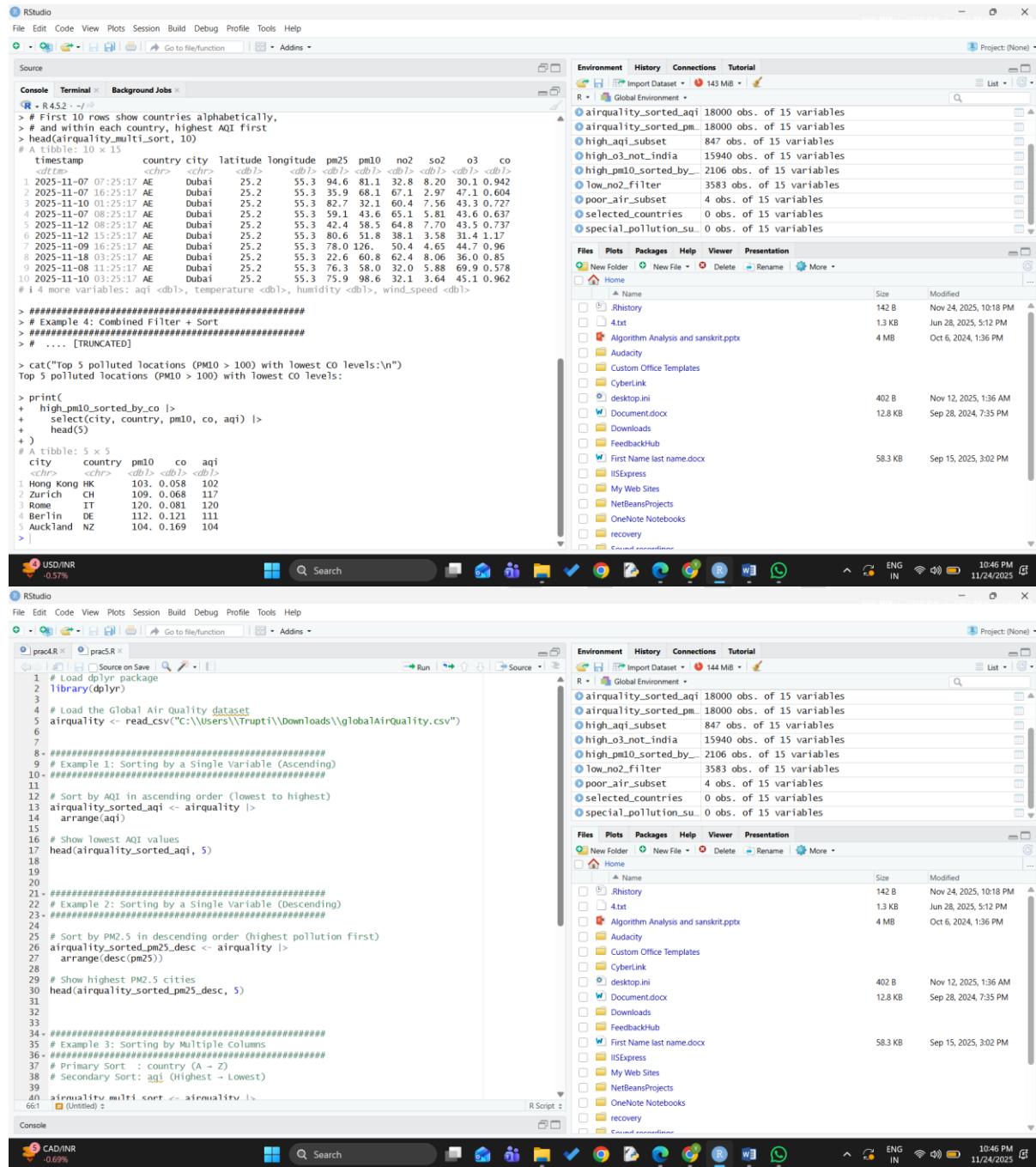
# Example 4: Combined Filter + Sort
# ... [TRUNCATED]

cat("Top 5 polluted locations (PM10 > 100) with lowest CO levels:\n")
Top 5 polluted locations (PM10 > 100) with lowest CO levels:
# i 4 more variables: aqi <dbl>, temperature <dbl>, humidity <dbl>, wind_speed <dbl>

```

SUBJECT :- Data analysis of SAS/SPSS/R

PRACTICAL – 5



Yash Bhagat
S072