

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

SUBJECT :- Data analysis of SAS/SPSS/R

PRACTICAL – 5

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
R - R4.5.2 - ~/
> source("~/active-studio-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...
dtm (1): timestamp

I use 'spec()' to retrieve the full column specification for this data.
I 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 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 2025-11-14 18:25:17 GB London 51.5 -0.128 8.12 11.8 6.11 4.21 47.6 0.58
2 2025-11-11 09:25:17 US San Franc... 37.8 -122. 13.1 7.78 8.64 3.5 51.7 0.294
3 2025-11-11 23:25:17 US Los Ange... 34.1 -118. 13.8 23.0 9.35 5.96 42.1 1.12
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 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 2025-11-15 03:25:17 FI Helsinki 60.2 24.9 116. 81.7 54.8 6.69 39.6 0.685
2 2025-11-05 19:25:17 GB London 51.5 -0.128 107. 43.8 13.4 4.60 21.7 0.89
3 2025-11-11 05:25:17 PH Manila 14.6 121. 103. 66.7 43.3 2.17 48.2 0.849
4 2025-11-17 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 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 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 2025-11-07 07:25:17 AE Dubai 25.2 55.3 94.6 81.1 32.8 8.20 30.1 0.942
2 2025-11-07 16:25:17 AE Dubai 25.2 55.3 35.9 68.1 67.1 2.97 47.1 0.604
3 2025-11-10 01:25:17 AE Dubai 25.2 55.3 82.7 32.1 60.4 7.56 43.3 0.727
4 2025-11-07 08:25:17 AE Dubai 25.2 55.3 59.1 43.6 65.1 5.81 43.6 0.637
5 2025-11-12 08:25:17 AE Dubai 25.2 55.3 42.4 58.5 64.8 7.70 43.5 0.737
6 2025-11-12 15:25:17 AE Dubai 25.2 55.3 80.6 51.8 38.1 3.58 31.4 1.17
7 2025-11-09 16:25:17 AE Dubai 25.2 55.3 78.0 126. 50.4 4.65 44.7 0.96
8 2025-11-18 03:25:17 AE Dubai 25.2 55.3 22.6 60.8 62.4 8.06 36.0 0.85
9 2025-11-08 11:25:17 AE Dubai 25.2 55.3 76.3 58.0 32.0 5.88 69.9 0.578
10 2025-11-10 03:25:17 AE Dubai 25.2 55.3 75.9 98.6 32.1 3.64 45.1 0.962
# I 4 more variables: aqi <dbl>, temperature <dbl>, humidity <dbl>, wind_speed <dbl>

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

> # .... [TRUNCATED]

> cat("Top 5 polluted locations (PM10 > 100) with lowest CO levels:\n")
Top 5 polluted locations (PM10 > 100) with lowest CO levels:
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
R - R4.5.2 - ~/
# A tibble: 5 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 2025-11-15 03:25:17 FI Helsinki 60.2 24.9 116. 81.7 54.8 6.69 39.6 0.685
2 2025-11-05 19:25:17 GB London 51.5 -0.128 107. 43.8 13.4 4.60 21.7 0.89
3 2025-11-11 05:25:17 PH Manila 14.6 121. 103. 66.7 43.3 2.17 48.2 0.849
4 2025-11-17 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 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 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 2025-11-07 07:25:17 AE Dubai 25.2 55.3 94.6 81.1 32.8 8.20 30.1 0.942
2 2025-11-07 16:25:17 AE Dubai 25.2 55.3 35.9 68.1 67.1 2.97 47.1 0.604
3 2025-11-10 01:25:17 AE Dubai 25.2 55.3 82.7 32.1 60.4 7.56 43.3 0.727
4 2025-11-07 08:25:17 AE Dubai 25.2 55.3 59.1 43.6 65.1 5.81 43.6 0.637
5 2025-11-12 08:25:17 AE Dubai 25.2 55.3 42.4 58.5 64.8 7.70 43.5 0.737
6 2025-11-12 15:25:17 AE Dubai 25.2 55.3 80.6 51.8 38.1 3.58 31.4 1.17
7 2025-11-09 16:25:17 AE Dubai 25.2 55.3 78.0 126. 50.4 4.65 44.7 0.96
8 2025-11-18 03:25:17 AE Dubai 25.2 55.3 22.6 60.8 62.4 8.06 36.0 0.85
9 2025-11-08 11:25:17 AE Dubai 25.2 55.3 76.3 58.0 32.0 5.88 69.9 0.578
10 2025-11-10 03:25:17 AE Dubai 25.2 55.3 75.9 98.6 32.1 3.64 45.1 0.962
# I 4 more variables: aqi <dbl>, temperature <dbl>, humidity <dbl>, wind_speed <dbl>

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

> # .... [TRUNCATED]

> cat("Top 5 polluted locations (PM10 > 100) with lowest CO levels:\n")
Top 5 polluted locations (PM10 > 100) with lowest CO levels:
```

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

SUBJECT :- Data analysis of SAS/SPSS/R PRACTICAL – 5

The image displays two screenshots of the RStudio interface, showing R code for data analysis of air quality data.

Top Screenshot: The console shows the first 10 rows of the 'airquality_multi_sort' dataset, sorted by country and then by AQI. The code includes comments and a print statement to display the first 5 rows of the sorted data.

```
# First 10 rows show countries alphabetically,
# and within each country, highest AQI first
head(airquality_multi_sort, 10)
# A tibble: 10 x 15
  timestamp          country city latitude longitude pm25 pm10 no2 so2 o3 co
<dtm>          <chr>    <chr>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
1 2025-11-07 07:25:17 AE Dubai 25.2 55.3 94.6 81.1 32.8 8.20 30.1 0.942
2 2025-11-07 16:25:17 AE Dubai 25.2 55.3 35.9 68.1 67.1 2.97 47.1 0.604
3 2025-11-10 01:25:17 AE Dubai 25.2 55.3 82.7 32.1 60.4 7.56 43.3 0.727
4 2025-11-07 08:25:17 AE Dubai 25.2 55.3 59.1 43.6 65.1 5.81 43.6 0.637
5 2025-11-12 08:25:17 AE Dubai 25.2 55.3 42.4 58.5 64.8 7.70 43.5 0.737
6 2025-11-12 15:25:17 AE Dubai 25.2 55.3 80.6 51.8 38.1 3.58 31.4 1.17
7 2025-11-09 16:25:17 AE Dubai 25.2 55.3 78.0 126. 50.4 4.65 44.7 0.96
8 2025-11-18 03:25:17 AE Dubai 25.2 55.3 22.6 60.8 62.4 8.06 36.0 0.85
9 2025-11-08 11:25:17 AE Dubai 25.2 55.3 76.3 58.0 32.0 5.88 69.9 0.578
10 2025-11-10 03:25:17 AE Dubai 25.2 55.3 75.9 98.6 32.1 3.64 45.1 0.962
# 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:

> print(
+   high_pm10_sorted_by_co |>
+   select(city, country, pm10, co, aqi) |>
+   head(5)
+ )
# A tibble: 5 x 5
  city          country pm10 co aqi
<chr>          <chr>    <dbl> <dbl> <dbl>
1 Hong Kong HK 103. 0.058 102
2 Zurich CH 109. 0.068 117
3 Rome IT 120. 0.081 120
4 Berlin DE 112. 0.121 111
5 Auckland NZ 104. 0.169 104
```

Bottom Screenshot: The console shows the loading of the 'dplyr' package and the 'airquality' dataset. The code demonstrates sorting the dataset by AQI (ascending) and PM2.5 (descending), and then selecting the top 5 cities for each sort.

```
1 # Load dplyr package
2 library(dplyr)
3
4 # Load the Global Air Quality dataset
5 airquality <- read_csv("C:\\Users\\Trupti\\Downloads\\GlobalAirQuality.csv")
6
7
8 #####
9 # Example 1: Sorting by a Single Variable (Ascending)
10 #####
11
12 # Sort by AQI in ascending order (lowest to highest)
13 airquality_sorted_aqi <- airquality |>
14   arrange(aqi)
15
16 # Show lowest AQI values
17 head(airquality_sorted_aqi, 5)
18
19
20
21 #####
22 # Example 2: Sorting by a Single Variable (Descending)
23 #####
24
25 # Sort by PM2.5 in descending order (highest pollution first)
26 airquality_sorted_pm25_desc <- airquality |>
27   arrange(desc(pm25))
28
29 # Show highest PM2.5 cities
30 head(airquality_sorted_pm25_desc, 5)
31
32
33
34 #####
35 # Example 3: Sorting by Multiple Columns
36 #####
37 # Primary Sort : country (A - Z)
38 # Secondary Sort: aqi (Highest - Lowest)
39
40 airquality_multi_sort <- airquality |>
41   arrange(country, aqi)
```