

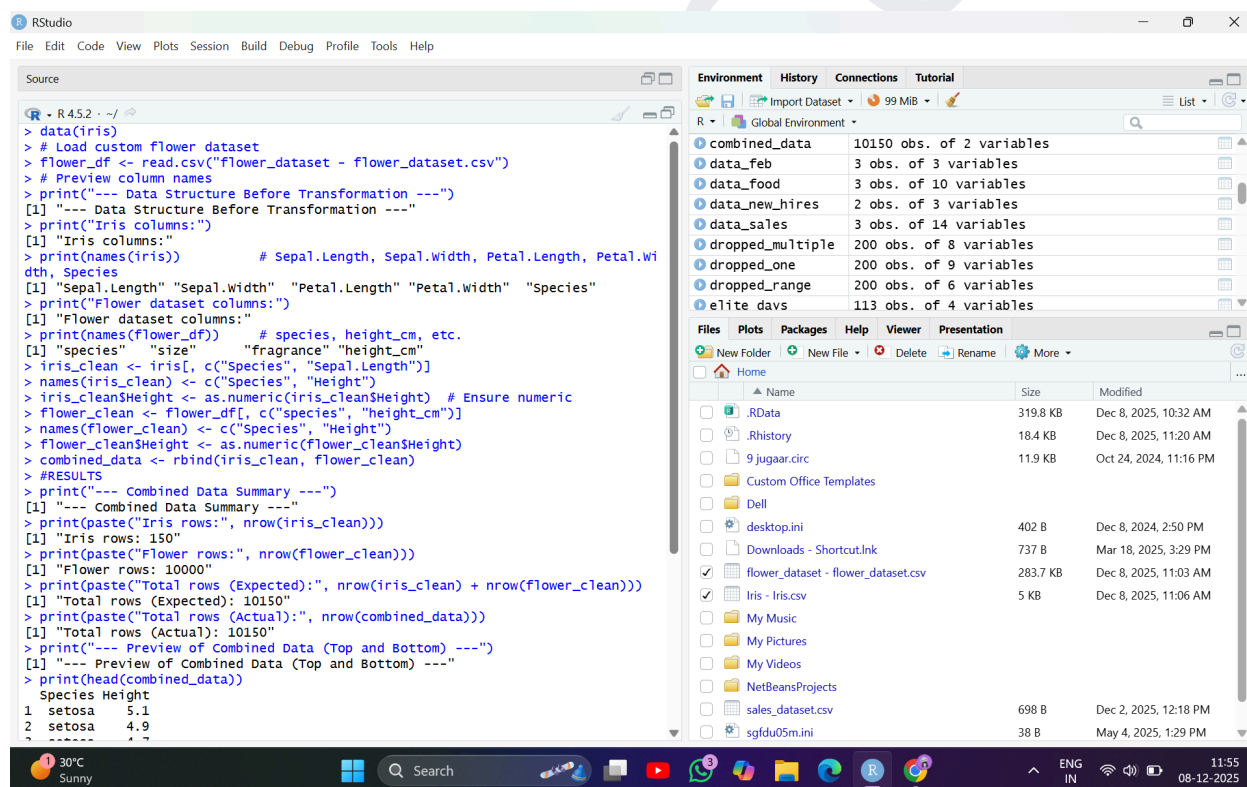
SHETH L.U.J AND SIR M.V COLLEGE

Subject: Data Analysis with SAS / SPSS /R

Practical no. 12

Aim: Combining datasets vertically
(concatenation) using rbind() (R).

Outputs→

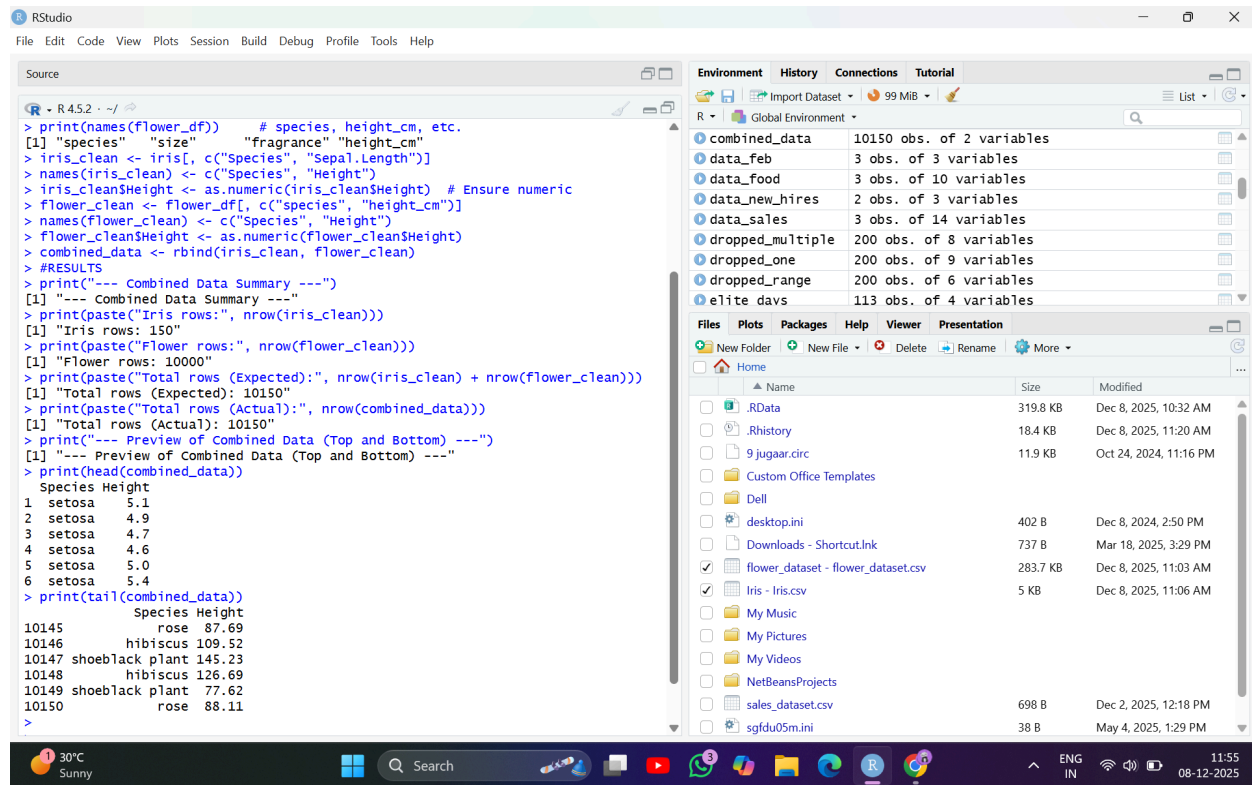


The screenshot displays the RStudio interface. The Source pane on the left contains R code that loads two datasets, 'iris' and 'flower_dataset.csv', cleans them, and combines them into a single dataset 'combined_data' using the 'rbind()' function. The Environment pane on the right shows the objects in the Global Environment, including 'combined_data' (10150 observations, 2 variables), 'data_feb' (3 observations, 3 variables), 'data_food' (3 observations, 10 variables), 'data_new_hires' (2 observations, 3 variables), 'data_sales' (3 observations, 14 variables), 'dropped_multiple' (200 observations, 8 variables), 'dropped_one' (200 observations, 9 variables), 'dropped_range' (200 observations, 6 variables), and 'elite_davs' (113 observations, 4 variables). The Files pane at the bottom shows the file explorer with various files and folders.

```
> data(iris)
> # Load custom flower dataset
> flower_df <- read.csv("flower_dataset - flower_dataset.csv")
> # Preview column names
> print("--- Data Structure Before Transformation ---")
[1] "--- Data Structure Before Transformation ---"
> print("Iris columns:")
[1] "Iris columns:"
> print(names(iris))      # Sepal.Length, Sepal.Width, Petal.Length, Petal.Wi
dth, Species
[1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width" "Species"
> print("Flower dataset columns:")
[1] "Flower dataset columns:"
> print(names(flower_df)) # species, height_cm, etc.
[1] "species" "size" "fragrance" "height_cm"
> iris_clean <- iris[, c("Species", "Sepal.Length")]
> names(iris_clean) <- c("Species", "Height")
> iris_clean$Height <- as.numeric(iris_clean$Height) # Ensure numeric
> flower_clean <- flower_df[, c("species", "height_cm")]
> names(flower_clean) <- c("Species", "Height")
> flower_clean$Height <- as.numeric(flower_clean$Height)
> combined_data <- rbind(iris_clean, flower_clean)
> #RESULTS
> print("--- Combined Data Summary ---")
[1] "--- Combined Data Summary ---"
> print(paste("Iris rows:", nrow(iris_clean)))
[1] "Iris rows: 150"
> print(paste("Flower rows:", nrow(flower_clean)))
[1] "Flower rows: 10000"
> print(paste("Total rows (Expected):", nrow(iris_clean) + nrow(flower_clean)))
[1] "Total rows (Expected): 10150"
> print(paste("Total rows (Actual):", nrow(combined_data)))
[1] "Total rows (Actual): 10150"
> print("--- Preview of Combined Data (Top and Bottom) ---")
[1] "--- Preview of Combined Data (Top and Bottom) ---"
> print(head(combined_data))
Species Height
1 setosa 5.1
2 setosa 4.9
```

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

Subject: Data Analysis with SAS / SPSS /R



The screenshot displays the RStudio environment with the following components:

- Source Pane:** Contains R code for data cleaning and summarization. The code includes comments, variable assignments, and print statements to verify the data.
- Environment Pane:** Lists objects in the Global Environment, including 'combined_data' (10150 obs. of 2 variables), 'data_feb' (3 obs. of 3 variables), 'data_food' (3 obs. of 10 variables), 'data_new_hires' (2 obs. of 3 variables), 'data_sales' (3 obs. of 14 variables), 'dropped_multiple' (200 obs. of 8 variables), 'dropped_one' (200 obs. of 9 variables), 'dropped_range' (200 obs. of 6 variables), and 'elite_davs' (113 obs. of 4 variables).
- Files Pane:** Shows the file explorer with various files and folders, including 'RData', 'Rhistory', '9 jugaar.circ', 'Custom Office Templates', 'Dell', 'desktop.ini', 'Downloads - Shortcut.lnk', 'flower_dataset - flower_dataset.csv', 'Iris - Iris.csv', 'My Music', 'My Pictures', 'My Videos', 'NetBeansProjects', 'sales_dataset.csv', and 'sgfdu05m.ini'.
- Console/Output:** Displays the output of the R code, showing the 'Combined Data Summary' and a preview of the combined data.

```
> print(names(flower_df)) # species, height_cm, etc.
[1] "species" "size" "fragrance" "height_cm"
> iris_clean <- iris[, c("Species", "Sepal.Length")]
> names(iris_clean) <- c("Species", "Height")
> iris_clean$Height <- as.numeric(iris_clean$Height) # Ensure numeric
> flower_clean <- flower_df[, c("Species", "height_cm")]
> names(flower_clean) <- c("Species", "Height")
> flower_clean$Height <- as.numeric(flower_clean$Height)
> combined_data <- rbind(iris_clean, flower_clean)
> #RESULTS
> print("---- Combined Data Summary ----")
[1] "---- Combined Data Summary ----"
> print(paste("Iris rows:", nrow(iris_clean)))
[1] "Iris rows: 150"
> print(paste("Flower rows:", nrow(flower_clean)))
[1] "Flower rows: 10000"
> print(paste("Total rows (Expected):", nrow(iris_clean) + nrow(flower_clean)))
[1] "Total rows (Expected): 10150"
> print(paste("Total rows (Actual):", nrow(combined_data)))
[1] "Total rows (Actual): 10150"
> print("---- Preview of Combined Data (Top and Bottom) ----")
[1] "---- Preview of Combined Data (Top and Bottom) ----"
> print(head(combined_data))
  Species Height
1  setosa   5.1
2  setosa   4.9
3  setosa   4.7
4  setosa   4.6
5  setosa   5.0
6  setosa   5.4
> print(tail(combined_data))
  Species Height
10145  rose   87.69
10146 hibiscus 109.52
10147 shoeblack plant 145.23
10148 hibiscus 126.69
10149 shoeblack plant  77.62
10150  rose   88.11
```