

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

PRACTICAL NO 15

AIM: Generating basic summaries using str() or summary() (R).

RStudio Environment View showing the creation of a sample dataset named `retail_df`. The dataset contains columns: ID, Category, Price, In_Stock, and Rating. The code then demonstrates the use of `str()` to show the internal structure of the `retail_df` object, which is a data frame with 6 observations and 5 variables.

```

> # 1. SETUP: Create Sample Data
> # we create a data frame with mixed data types (Numeric, Character, Logical, NA)
> retail_df <- data.frame(
+   ID = 1:6,
+   Category = c("Electronics", "Home", "Electronics", "Clothing", "Home", "Clothing"),
+   Price = c(500.50, 45.0, 900.00, NA, 300.00, 25.00), # Note the NA
+   In_Stock = c(TRUE, TRUE, FALSE, TRUE, FALSE, TRUE),
+   Rating = c(4.5, 3.8, 4.9, 4.0, 3.5, 4.2),
+   stringsAsFactors = FALSE
+ )
> print("---- Data Loaded ---")
[1] "---- Data Loaded ---"
> print(retail_df)
#> # 2. USING str() (Structure)
#> # Purpose: Compactly display the internal structure of the R object.
> print("---- OUTPUT OF str() ---")
[1] "---- OUTPUT OF str() ---"
> str(retail_df)
'data.frame': 6 obs. of 5 variables:
 $ ID : int 1 2 3 4 5 6
 $ Category: chr "Electronics" "Home" "Electronics" "Clothing" ...
 $ Price : num 500.5 45 900.0 NA 300 ...
 $ In_Stock: logi TRUE TRUE FALSE TRUE FALSE
 $ Rating : num 4.5 3.8 4.9 4.0 3.5 4.2
> # 3. USING summary() (Statistical Summary)
> # Purpose: detailed summary statistics for each column.
> print("---- OUTPUT OF summary() [Before Factor Conversion] ---")
[1] "---- OUTPUT OF summary() [Before Factor Conversion] ---"

```

RStudio Environment View showing statistical summaries for the `retail_df` dataset. It first displays the output of `summary(retail_df)`, which provides descriptive statistics for each column. Then, it shows the output of `summary(retail_df$Category)` after converting the `Category` column to factors, and finally the output of `summary(retail_df)` again to show the updated statistics.

```

> print("---- OUTPUT OF summary() [Before Factor Conversion] ---")
[1] "---- OUTPUT OF summary() [Before Factor Conversion] ---"
> summary(retail_df)
#> # 4. IMPROVING summary() WITH FACTORS
#> # by default category is character here; convert to factor to get counts per level.
> print("---- category counts (before factor conversion) ---")
[1] "---- category counts (before factor conversion) ---"
> print(table(retail_df$category))
Clothing Electronics Home
2 2 2
> retail_df$category <- as.factor(retail_df$category)
> print("---- OUTPUT OF summary() [After Factor Conversion] ---")
[1] "---- OUTPUT OF summary() [After Factor Conversion] ---"
> summary(retail_df)
#> # 5. Accessing Specific Summaries
#> # Sometimes you only want single values; use na.rm = TRUE when needed.
> avg_rating <- mean(retail_df$Rating, na.rm = TRUE)
> max_price <- max(retail_df$Price, na.rm = TRUE) # na.rm ignores the missing value

```

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

The screenshot shows the RStudio interface with the following details:

- Console Tab:** Displays R code and its output. The code demonstrates various summary statistics (Mean, Median, Min, Max) for different categories (Clothing, Electronics, Home) before and after converting them into factors. It also shows how to access specific values using `na.rm = TRUE`.
- File Browser:** Shows a list of files in the current directory:
 - data.csv (43.5 MB)
 - flipkart_com-ecommerce_sample.csv (36.3 MB)
 - netflix_titles.csv (3.2 MB)
 - superstore.csv (15.1 MB)
 - WA_Fn-UseC_HR-Employee-Attrition (222.6 KB)
 - win-library
 - Housing.csv (29.3 KB)
- System Tray:** Shows system status including temperature (30°C), weather (Sunny), battery level (ENG), and date/time (12:02, 08-12-2025).