

PRACTICAL NO : 15

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

```

R - R4.5.2 - ~/
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
> # R Script: Generating Basic Summaries
> # Functions: str() and summary()
> # Dataset: Retail Product Data (simulated)
> #=====
> # 1. SETUP: Create Sample Data
> #=====
> # We create a dataframe 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, 45.00, 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)
+ )
> print("--- Data Loaded ---")
[1] "--- Data Loaded ---"
> #=====
> # 2. USING str() (Structure)
> #=====
> # Purpose: compactly display the internal structure of the R object.
> # What it tells you:
> # - The class (data.frame)
> # - Number of observations (rows) and variables (columns)
> # - The data type of each column (num, int, chr, logi)
> # - The first few values
> 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 45.0 900.0 NA 300.0 25.0
 $ In_Stock: logi  TRUE TRUE FALSE TRUE FALSE TRUE
 $ Rating  : num  4.5 3.8 4.9 4.0 3.5 4.2

```

```

R - R4.5.2 - ~/
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
[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 45.0 900.0 NA 300.0 25.0
 $ In_Stock: logi  TRUE TRUE FALSE TRUE FALSE TRUE
 $ 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.
> # What it tells you:
> # - Numeric Cols: Min, 1st Qu, Median, Mean, 3rd Qu, Max, and NA count.
> # - Logical Cols: Count of TRUE vs FALSE.
> # - Character Cols: Length and class (Usually not very useful yet).
> print("--- OUTPUT OF summary() [Before Factor conversion] ---")
[1] "--- OUTPUT OF summary() [Before Factor conversion] ---"
> summary(retail_df)
      ID      Category      Price      In_Stock      Rating
Min.   :1.00   Length:6     Min.   : 25.0   Mode :logical Min.   :3.500
1st Qu.:2.25   class :character 1st Qu.: 45.0   FALSE:2   1st Qu.:3.850
Median :3.50   Mode  :character   Median :300.0   TRUE :4    Median :4.100
Mean   :3.50                                Mean   :354.1   Mean   :4.150
3rd Qu.:4.75                                3rd Qu.:500.5   3rd Qu.:4.425
Max.   :6.00                                Max.   :900.0   Max.   :4.900
      NA's :1
> #=====
> # 4. IMPROVING summary() WITH FACTORS
> #=====
> # By default, R treats text as 'character' and summary() just counts the length.
> # If we convert 'category' to a 'factor', summary() will calculate counts per category

```

