

PRACTICAL NO 15

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

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
R - R4.1.2 - ~/
> # 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),
+   stringsAsFactors = FALSE
+ )
> print("--- Data Loaded ---")
[1] "--- Data Loaded ---"
> print(retail_df)
  ID Category Price In_Stock Rating
1 1 Electronics 500.5    TRUE   4.5
2 2      Home  45.0    TRUE   3.8
3 3 Electronics 900.0   FALSE   4.9
4 4    Clothing    NA    TRUE   4.0
5 5      Home  300.0   FALSE   3.5
6 6    Clothing  25.0    TRUE   4.2
> # 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 NA 300 ...
 $ In_Stock : logi  TRUE TRUE FALSE TRUE FALSE TRUE
 $ Rating   : num  4.5 3.8 4.9 4 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
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
R - R4.1.2 - ~/
> 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 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$Category)
> print("--- OUTPUT OF summary() [After Factor Conversion] ---")
[1] "--- OUTPUT OF summary() [After Factor Conversion] ---"
> summary(retail_df)
  ID      Category      Price      In_Stock      Rating
Min.   :1.00  clothing :2   Min.   : 25.0   Mode :logical Min.   :3.500
1st Qu.:2.25  Electronics:2 1st Qu.: 45.0   FALSE:2   1st Qu.:3.850
Median :3.50      Home   :2 Median :300.0  TRUE :4    Median :4.100
Mean   :3.50                                Mean   :354.1   Mean   :4.150
3rd Qu.:4.75                                3rd Qu.:500.5   3rd Qu.:4.425
Max.   :6.00                                Max.   :900.0   Max.   :4.900
NA's   :1
> # 5. Accessing Specific Summaries
> # Sometimes you only want single values; use na.rm = TRUE when needed.
> avg_rating <- mean(retail_df$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 displays the RStudio environment with the following components:

- Source Editor:** Contains R code for data analysis, including summary statistics and factor conversion.
- Console:** Shows the output of the R code, including summary statistics and category counts.
- Environment:** Lists the loaded packages and data objects.
- Files:** Shows the file explorer with a list of CSV files and their sizes.

R Code in Source Editor:

```
> # 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)

  ID          category      Price      In_Stock      Rating
Min. :1.00  Clothing :2   Min. : 25.0   Mode :logical   Min. :3.500
1st Qu.:2.25  Electronics:2 1st Qu.: 45.0   FALSE:2        1st Qu.:3.850
Median :3.50      Home :2   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

> # 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
> print(sprintf("Average Rating: %.2f", avg_rating))
[1] "Average Rating: 4.15"
> print(sprintf("Highest Price: %.2f", max_price))
[1] "Highest Price: 900.00"
>
```

Files in Environment:

Name	Size
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