

MVLU COLLEGE

R PRACTICAL 15

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

Output :

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
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", "food", "Computer", "bike accessories", "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(4.7, 4.0, 5.0, 3.1, 2.9, 3.9),
+   stringsAsFactors = FALSE
+ )
> print("--- Data Loaded ---")
[1] "--- Data Loaded ---"
> print(retail_df)
  ID Category Price In_Stock Rating
1  1 Electronics 500.5    TRUE  4.7
2  2    food    45.0    TRUE  4.0
3  3 Computer  900.0   FALSE  5.0
4  4 bike accessories NA    TRUE  3.1
5  5    Home   300.0   FALSE  2.9
6  6  Clothing    25.0    TRUE  3.9
> # 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" "food" "Computer" "bike accessories" ...
 $ Price   : num  500.5 45 900 NA 300 ...
 $ In_Stock: logi  TRUE TRUE FALSE TRUE FALSE TRUE
 $ Rating  : num  4.7 4.5 3.1 2.9 3.9
> # 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] ---"
summary(retail_df)
  ID      Category      Price      In_Stock      Rating
Min.   :1.00   Length:6   Min.   : 25.0   Mode :logical Min.   :2.900
1st Qu.:2.25   Class :character 1st Qu.: 45.0   FALSE:2   1st Qu.:3.300
Median :3.50   Mode  :character   Median :300.0   TRUE :4    Median :3.950
Mean   :3.50           Mean   :354.1           Mean :3.933
3rd Qu.:4.75           3rd Qu.:500.5           3rd Qu.:4.525
Max.   :6.00           Max.   :900.0           Max.   :5.000
NA's   :1
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
R - R4.1.2 - ~/
'data.frame': 6 obs. of 5 variables:
 $ ID      : int  1 2 3 4 5 6
 $ Category: chr  "Electronics" "food" "Computer" "bike accessories" ...
 $ Price   : num  500.5 45 900 NA 300 ...
 $ In_Stock: logi  TRUE TRUE FALSE TRUE FALSE TRUE
 $ Rating  : num  4.7 4.5 3.1 2.9 3.9
> # 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] ---"
> summary(retail_df)
  ID      Category      Price      In_Stock      Rating
Min.   :1.00   Length:6   Min.   : 25.0   Mode :logical Min.   :2.900
1st Qu.:2.25   Class :character 1st Qu.: 45.0   FALSE:2   1st Qu.:3.300
Median :3.50   Mode  :character   Median :300.0   TRUE :4    Median :3.950
Mean   :3.50           Mean   :354.1           Mean :3.933
3rd Qu.:4.75           3rd Qu.:500.5           3rd Qu.:4.525
Max.   :6.00           Max.   :900.0           Max.   :5.000
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$Category))
bike accessories    clothing    computer    electronics    food
               1              1              1              1              1
Home
1
> retail_df$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   bike accessories:1   Min.   : 25.0   Mode :logical Min.   :2.900
```

MVLU COLLEGE

R PRACTICAL 15

The screenshot shows the RStudio interface with the following components:

- Source Panel:** Contains R code for data analysis, including summary statistics, factor conversion, and specific value extraction.
- Console Panel:** Displays the output of the R code, showing summary statistics for the 'retail_df' dataset.
- Environment Panel:** Lists the objects in the global environment, including 'runs_by_opponent', 'selected_cols', 'starts_with_p', 'team_lookup', and 'tidy_cricket'.
- Files Panel:** Shows the file explorer with various files and folders, including 'tue1_sales.csv', 'GIS DataBase', 'IISExpress', 'My Music', 'My Pictures', 'My Videos', 'My Web Sites', 'R', 'SM PRACTICAL - II 20.txt', 'SM PRACTICAL - II.txt', 'total_bills.csv', 'Virtual Machines', 'Visual Studio 2022', and 'WindowsPowerShell'.

R Code in Source Panel:

```
R - R 4.1.2 - ~/
3rd Qu.:4.75      3rd Qu.:500.5      3rd Qu.:4.525
Max.:6.00        Max.:900.0        Max.:5.000
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))

bike accessories      clothing      computer      Electronics      food
1                    1                    1                    1                    1
Home
1

> 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 bike accessories:1 Min.:25.0 Mode:logical Min.:2.900
1st Qu.:2.25 clothing:1 1st Qu.:45.0 FALSE:2 1st Qu.:3.300
Median:3.50 computer:1 Median:300.0 TRUE:4 Median:3.950
Mean:3.50 Electronics:1 Mean:354.1 Mean:3.933
3rd Qu.:4.75 food:1 3rd Qu.:500.5 3rd Qu.:4.525
Max.:6.00 Home:1 Max.:900.0 Max.:5.000
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: 3.93"
> print(sprintf("Highest Price: %.2f", max_price))
[1] "Highest Price: 900.00"
>
```

Console Output:

```
3rd Qu.:4.75      3rd Qu.:500.5      3rd Qu.:4.525
Max.:6.00        Max.:900.0        Max.:5.000
NA's:1

[1] "--- Category counts (before factor conversion) ---"
[1] "--- Category counts (before factor conversion) ---"
bike accessories      clothing      computer      Electronics      food
1                    1                    1                    1                    1
Home
1

[1] "--- OUTPUT OF summary() [After Factor conversion] ---"
[1] "--- OUTPUT OF summary() [After Factor conversion] ---"
   ID      Category      Price      In_Stock      Rating
Min.:1.00 bike accessories:1 Min.:25.0 Mode:logical Min.:2.900
1st Qu.:2.25 clothing:1 1st Qu.:45.0 FALSE:2 1st Qu.:3.300
Median:3.50 computer:1 Median:300.0 TRUE:4 Median:3.950
Mean:3.50 Electronics:1 Mean:354.1 Mean:3.933
3rd Qu.:4.75 food:1 3rd Qu.:500.5 3rd Qu.:4.525
Max.:6.00 Home:1 Max.:900.0 Max.:5.000
NA's:1

[1] "Average Rating: 3.93"
[1] "Highest Price: 900.00"
>
```

Environment Panel:

Object	Class	Attributes
runs_by_opponent	data.frame	8 obs. of 5 variables
selected_cols	data.frame	106 obs. of 9 variables
starts_with_p	data.frame	106 obs. of 9 variables
team_lookup	data.frame	8 obs. of 2 variables
tidy_cricket	data.frame	106 obs. of 11 variables

Files Panel:

Name	Size	Modified
tue1_sales.csv	700 B	Sep 18, 2023, 8:33 AM
GIS DataBase		
IISExpress		
My Music		
My Pictures		
My Videos		
My Web Sites		
R		
SM PRACTICAL - II 20.txt	1.2 KB	Aug 22, 2025, 3:02 PM
SM PRACTICAL - II.txt	1.2 KB	Aug 22, 2025, 3:01 PM
total_bills.csv	1.2 KB	Sep 18, 2025, 8:47 AM
Virtual Machines		
Visual Studio 2022		
WindowsPowerShell		