

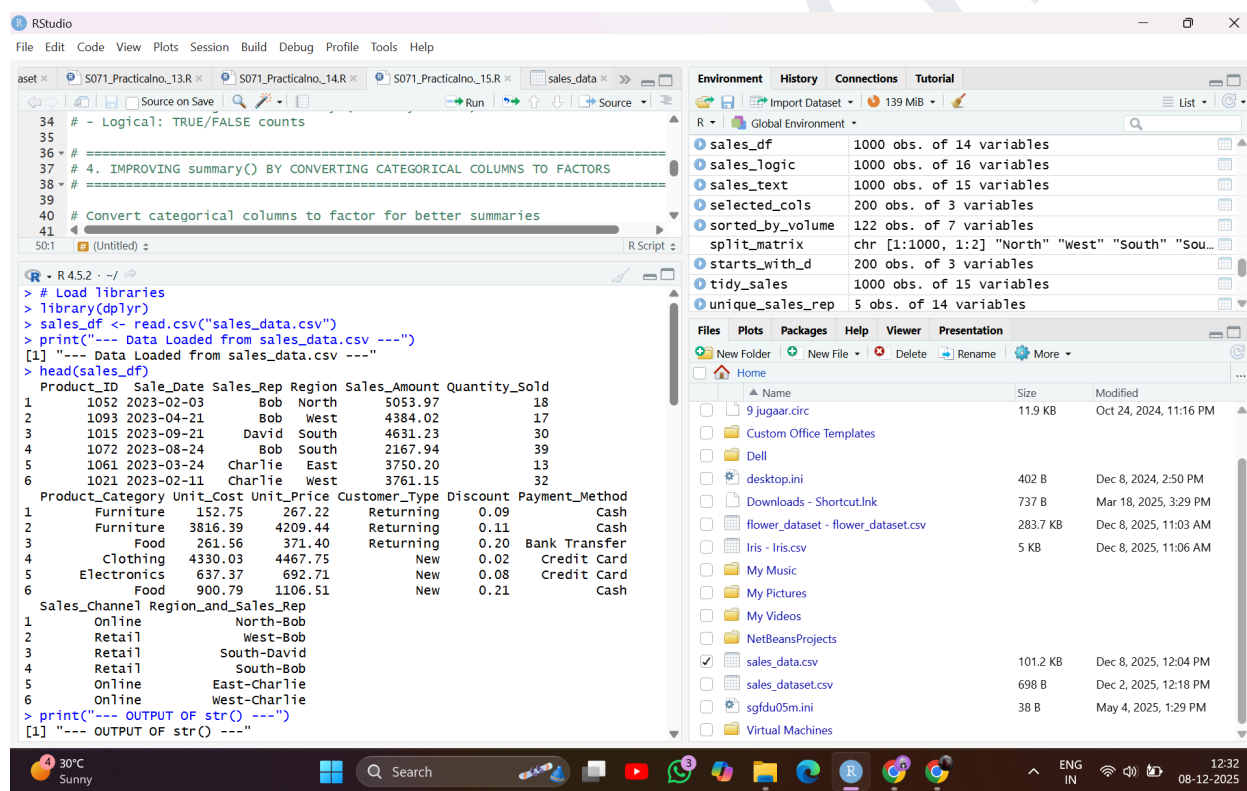
# SHETH L.U.J AND SIR M.V COLLEGE

Subject: Data Analysis with SAS / SPSS /R

## Practical no. 15

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

**Outputs**→



The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for loading libraries, reading a CSV file, and converting categorical columns to factors.
- Environment:** Lists objects in the Global Environment, including sales\_df, sales\_logic, sales\_text, selected\_cols, sorted\_by\_volume, split\_matrix, starts\_with\_d, tidy\_sales, and unique\_sales\_rep.
- Files:** Shows the file explorer with various files and folders, including sales\_data.csv.
- Console:** Displays the output of the R code, including the data loaded from sales\_data.csv and the output of the str() function.

```
# Load libraries
library(dplyr)
sales_df <- read.csv("sales_data.csv")
print("--- Data Loaded from sales_data.csv ---")
[1] "--- Data Loaded from sales_data.csv ---"
> head(sales_df)
  Product_ID Sale_Date Sales_Rep Region Sales_Amount Quantity_Sold
1      1052 2023-02-03      Bob North      5053.97             18
2      1093 2023-04-21      Bob West       4384.02             17
3      1015 2023-09-21    David South      4631.23             30
4      1072 2023-08-24      Bob South      2167.94             39
5      1061 2023-03-24    Charlie East       3750.20             13
6      1021 2023-02-11    Charlie West       3761.15             32
  Product_Category Unit_Cost Unit_Price Customer_Type Discount Payment_Method
1      Furniture      152.75      267.22      Returning    0.09          Cash
2      Furniture      3816.39     4209.44      Returning    0.11          Cash
3      Food          261.56       371.40      Returning    0.20 Bank Transfer
4      Clothing     4330.03     4467.75        New        0.02      Credit Card
5      Electronics    637.37     692.71        New        0.08      Credit Card
6      Food          900.79     1106.51        New        0.21          Cash
  Sales_Channel Region_and_Sales_Rep
1      Online      North-Bob
2      Retail      West-Bob
3      Retail      South-David
4      Retail      South-Bob
5      Online      East-Charlie
6      Online      West-Charlie
> print("--- OUTPUT OF str() ---")
[1] "--- OUTPUT OF str() ---"
```

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

Subject: Data Analysis with SAS / SPSS /R

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for loading and summarizing a dataset. The code includes comments and a `summary(sales_df)` command.
- Console:** Displays the output of `summary(sales_df)`, showing the structure of the data frame with 1000 observations and 14 variables. The variables are: `Product_ID` (integer), `Sale_Date` (character), `Sales_Rep` (character), `Region` (character), `Sales_Amount` (numeric), `Quantity_Sold` (integer), `Product_Category` (character), `Unit_Cost` (numeric), `Unit_Price` (numeric), `Customer_Type` (character), `Discount` (numeric), `Payment_Method` (character), `Sales_Channel` (character), and `Region_and_Sales_Rep` (character).
- Environment:** Lists the objects in the global environment, including `sales_df` (1000 obs. of 14 variables), `sales_logic` (1000 obs. of 16 variables), `sales_text` (1000 obs. of 15 variables), `selected_cols` (200 obs. of 3 variables), `sorted_by_volume` (122 obs. of 7 variables), `split_matrix` (chr [1:1000, 1:2] "North" "West" "South" "Sou..."), `starts_with_d` (200 obs. of 3 variables), `tidy_sales` (1000 obs. of 15 variables), and `unique_sales_rep` (5 obs. of 14 variables).
- Files:** Shows the file explorer with various files and folders, including `9 jugaar.circ`, `Custom Office Templates`, `Dell`, `desktop.ini`, `Downloads - ShortcutLink`, `flower_dataset - flower_dataset.csv`, `Iris - Iris.csv`, `My Music`, `My Pictures`, `My Videos`, `NetBeansProjects`, `sales_data.csv` (101.2 KB), `sales_dataset.csv` (698 B), `sgfdu05m.ini` (38 B), and `Virtual Machines`.

The screenshot shows the RStudio interface after converting categorical variables to factors. The `summary(sales_df)` output is as follows:

```
> print("---- OUTPUT OF summary() [Before Factor Conversion] ----")
[1] "---- OUTPUT OF summary() [Before Factor Conversion] ----"
> summary(sales_df)
  Product_ID  Sale_Date    Sales_Rep    Region
Min.   :1001  Length:1000  Length:1000  Length:1000
1st Qu.:1024  Class :character  Class :character  Class :character
Median :1051  Mode  :character  Mode  :character  Mode  :character
Mean   :1050
3rd Qu.:1075
Max.   :1100

  Sales_Amount  Quantity_Sold  Product_Category  Unit_Cost
Min.   : 100.1  Min.   : 1.00  Length:1000  Min.   : 60.28
1st Qu.:2550.3  1st Qu.:13.00  Class :character  1st Qu.:1238.38
Median :5019.3  Median :25.00  Mode  :character  Median :2467.24
Mean   :5019.3  Mean   :25.36  Mean   :2475.30
3rd Qu.:7507.4  3rd Qu.:38.00  Max.   :4995.30
Max.   :9989.0  Max.   :49.00

  Unit_Price  Customer_Type  Discount  Payment_Method
Min.   : 167.1  Length:1000  Min.   :0.0000  Length:1000
1st Qu.:1509.1  Class :character  1st Qu.:0.0800  Class :character
Median :2696.4  Mode  :character  Median :0.1500  Mode  :character
Mean   :2728.4  Mean   :0.1524
3rd Qu.:3958.0  3rd Qu.:0.2300
Max.   :5442.1  Max.   :0.3000

  Sales_Channel  Region_and_Sales_Rep
Length:1000
Class :character
Mode  :character
```

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

Subject: Data Analysis with SAS / SPSS /R

The screenshot shows the RStudio environment with the following components:

- Source Editor:** Contains R code for converting categorical columns to factors and a summary of the resulting data frame.
- Environment:** Lists the objects in the global environment, including sales\_df, sales\_logic, sales\_text, selected\_cols, sorted\_by\_volume, split\_matrix, starts\_with\_d, tidy\_sales, and unique\_sales\_rep.
- Files:** Shows the file explorer with various files and folders, including sales\_data.csv.

```
# - Logical: TRUE/FALSE counts
# 4. IMPROVING summary() BY CONVERTING CATEGORICAL COLUMNS TO FACTORS
# Convert categorical columns to factor for better summaries

> # Convert categorical columns to factor for better summaries
> sales_df$Region <- as.factor(sales_df$Region)
> sales_df$Product_Category <- as.factor(sales_df$Product_Category)
> sales_df$Sales_Rep <- as.factor(sales_df$Sales_Rep)
> sales_df$Customer_Type <- as.factor(sales_df$Customer_Type)
> sales_df$Payment_Method <- as.factor(sales_df$Payment_Method)
> sales_df$Sales_Channel <- as.factor(sales_df$Sales_Channel)
> print("---- OUTPUT OF summary() [After Factor Conversion] ----")
[1] "---- OUTPUT OF summary() [After Factor Conversion] ----"
> summary(sales_df)
```

Product_ID	Sale_Date	Sales_Rep	Region	Sales_Amount
Min. :1001	Length:1000	Alice :192	East :263	Min. : 100.1
1st Qu.:1024	Class :character	Bob :208	North:267	1st Qu.:2550.3
Median :1051	Mode :character	Charlie:169	South:226	Median :5019.3
Mean :1050		David :222	West :244	Mean :5019.3
3rd Qu.:1075		Eve :209		3rd Qu.:7507.4
Max. :1100				Max. :9989.0

Quantity_Sold	Product_Category	Unit_Cost	Unit_Price
Min. : 1.00	Clothing :268	Min. : 60.28	Min. : 167.1
1st Qu.:13.00	Electronics:246	1st Qu.:1238.38	1st Qu.:1509.1
Median :25.00	Food :226	Median :2467.24	Median :2696.4
Mean :25.36	Furniture :260	Mean :2475.30	Mean :2728.4
3rd Qu.:38.00		3rd Qu.:3702.86	3rd Qu.:3958.0
Max. :49.00		Max. :4995.30	Max. :5442.1

Customer_Type	Discount	Payment_Method	Sales_Channel
New :504	Min. :0.0000	Bank Transfer:342	Online:488
Returning:496	1st Qu.:0.0800	Cash :313	Retail:512
	Median :0.1500	Credit Card :345	

The screenshot shows the RStudio environment with the following components:

- Source Editor:** Contains R code for converting categorical columns to factors and a summary of the resulting data frame.
- Environment:** Lists the objects in the global environment, including sales\_df, sales\_logic, sales\_text, selected\_cols, sorted\_by\_volume, split\_matrix, starts\_with\_d, tidy\_sales, and unique\_sales\_rep.
- Files:** Shows the file explorer with various files and folders, including sales\_data.csv.

```
[1] "---- OUTPUT OF summary() [After Factor Conversion] ----"
> summary(sales_df)
```

Product_ID	Sale_Date	Sales_Rep	Region	Sales_Amount
Min. :1001	Length:1000	Alice :192	East :263	Min. : 100.1
1st Qu.:1024	Class :character	Bob :208	North:267	1st Qu.:2550.3
Median :1051	Mode :character	Charlie:169	South:226	Median :5019.3
Mean :1050		David :222	West :244	Mean :5019.3
3rd Qu.:1075		Eve :209		3rd Qu.:7507.4
Max. :1100				Max. :9989.0

Quantity_Sold	Product_Category	Unit_Cost	Unit_Price
Min. : 1.00	Clothing :268	Min. : 60.28	Min. : 167.1
1st Qu.:13.00	Electronics:246	1st Qu.:1238.38	1st Qu.:1509.1
Median :25.00	Food :226	Median :2467.24	Median :2696.4
Mean :25.36	Furniture :260	Mean :2475.30	Mean :2728.4
3rd Qu.:38.00		3rd Qu.:3702.86	3rd Qu.:3958.0
Max. :49.00		Max. :4995.30	Max. :5442.1

Customer_Type	Discount	Payment_Method	Sales_Channel
New :504	Min. :0.0000	Bank Transfer:342	Online:488
Returning:496	1st Qu.:0.0800	Cash :313	Retail:512
	Median :0.1500	Credit Card :345	

Region_and_Sales_Rep
Length:1000
Class :character
Mode :character

# 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 Editor:** Contains R code for calculating average and maximum sales, and total quantity sold. The code is as follows:

```
63 total_quantity <- sum(sales_df$Quantity_Sold, na.rm = TRUE)
64
65 print(paste("Average Sales Amount:", avg_sales))
66 print(paste("Maximum Sales Amount:", max_sales))
67 print(paste("Total Quantity Sold:", total_quantity))
68
69
```
- Environment Panel:** Shows the Global Environment with two objects: `unique_sales_rep` (5 obs. of 14 variables) and `volatile_days` (114 obs. of 7 variables). Below this, the `values` table is displayed:

Variable	Value
avg_dist	6.62475
avg_sales	5019.26523
avg_time	44.7445
current_time	2025-12-08 12:25:22 IST
max_sales	9989.04
total_quantity	25355L
- Files Panel:** Shows the file explorer with various files and folders, including `9 jugaar.circ`, `Custom Office Templates`, `Desktop.ini`, `Downloads - Shortcut.lnk`, `flower_dataset - flower_dataset.csv`, `Iris - Iris.csv`, `My Music`, `My Pictures`, `My Videos`, `NetBeansProjects`, `sales_data.csv` (101.2 KB), `sales_dataset.csv` (698 B), `sgfdu05m.ini` (38 B), and `Virtual Machines`.
- Console:** Displays the output of the R code execution, showing the average sales amount (5019.26523), maximum sales amount (9989.04), and total quantity sold (25355).