

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 shows the RStudio interface with the following details:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Code Editor:** An R script titled "sales_data.R" is open. The code includes comments for logical counts, improving summaries by converting categorical columns to factors, and converting categorical columns to factors for better summaries. It then loads the "sales_data.csv" file using dplyr, prints the first few rows, and prints the output of the str() function.
- Environment View:** Shows the global environment with objects like sales_df, sales_logic, sales_text, selected_col's, sorted_by_volume, split_matrix, starts_with_d, tidy_sales, and unique_sales_rep.
- File View:** Shows the file system structure under "Home".
- Bottom Bar:** Includes icons for search, file operations, and system status (ENG IN, 12:32, 08-12-2025).

```
34 # - Logical: TRUE/FALSE counts
35
36 # =====
37 # 4. IMPROVING summary() BY CONVERTING CATEGORICAL COLUMNS TO FACTORS
38 # =====
39
40 # Convert categorical columns to factor for better summaries
41
42 # R 4.5.2 - / 
> # 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 two panes. The left pane displays R code and its output. The right pane shows the file browser and environment.

Code and Output:

```
setwd(~)
# S071_Practicalno_13.R x S071_Practicalno_14.R x S071_Practicalno_15.R x sales_data x
34 # - Logical: TRUE/FALSE counts
35
36 # =====
37 # 4. IMPROVING summary() BY CONVERTING CATEGORICAL COLUMNS TO FACTORS
38 # =====
39
40 # Convert categorical columns to factor for better summaries
41
42 (Untitled) +
```

```
R > R 4.5.2 - / ~
> str(sales_df)
'data.frame': 1000 obs. of 14 variables:
 $ Product_ID    : int 1052 1093 1015 1072 1061 1021 1083 1087 1075 1075 ...
 $ Sale_Date      : chr "2023-02-03" "2023-04-21" "2023-09-21" "2023-08-24" ...
 ...
 $ Sales_Rep      : chr "Bob" "Bob" "David" "Bob" ...
 $ Region         : chr "North" "West" "South" "South" ...
 $ Sales_Amount   : num 5054 4384 4631 2168 3750 ...
 $ Quantity_Sold  : int 18 17 30 39 13 32 29 46 30 18 ...
 $ Product_Category: chr "Furniture" "Furniture" "Food" "Clothing" ...
 $ Unit_Cost       : num 153 3816 262 4330 637 ...
 $ Unit_Price      : num 267 4209 371 4468 693 ...
 $ Customer_Type   : chr "Returning" "Returning" "Returning" "New" ...
 $ Discount        : num 0.09 0.11 0.2 0.02 0.08 0.21 0.14 0.12 0.05 0.13 ...
 ...
 $ Payment_Method  : chr "Cash" "Cash" "Bank Transfer" "Credit Card" ...
 $ Sales_Channel   : chr "Online" "Retail" "Retail" "Retail" ...
 $ Region_and_Sales_Rep: chr "North-Bob" "West-Bob" "South-David" "South-Bob"
...
> 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
```

Environment:

Object	Type	Size	Variables
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		
unique_sales_rep	5 obs. of 14 variables		

File Browser:

- New Folder
- New File
- Delete
- Rename
- More

- Home
 - Name
 - Size
 - Modified
 - 9_jugar.circ
 - Custom Office Templates
 - Dell
 - desktop.ini
 - Downloads - Shortcut.lnk
 - flower_dataset - flower_dataset.csv
 - Iris - Iris.csv
 - My Music
 - My Pictures
 - My Videos
 - NetBeansProjects
 - sales_data.csv
 - sales_dataset.csv
 - sgfdud05.mni
 - Virtual Machines

The screenshot shows the RStudio interface with two panes. The left pane displays R code and its output. The right pane shows the Global Environment and File Explorer.

Code and Output:

```
set <- S071_Practicalno_13.R <- S071_Practicalno_14.R <- S071_Practicalno_15.R <- sales_data > Source > Run > Source > R Script > (Untitled) > R 4.5.2 > /> 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 3rd Qu.:3702.86 Max. :9989.0 Max. :49.00 Max. :4995.30 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 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 Length:1000 Class :character Class :character Mode :character
```

Environment:

- 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
- unique_sales_rep 5 obs. of 14 variables

File Explorer:

- Files Plots Packages Help Viewer Presentation
- New Folder New File Delete Rename More
- Home
 - Name Size Modified
 - 9 jugar.circ 11.9 KB Oct 24, 2024, 11:16 PM
 - Custom Office Templates
 - Dell
 - desktop.ini 402 B Dec 8, 2024, 2:50 PM
 - Downloads - Shortcut.lnk 737 B Mar 18, 2025, 3:29 PM
 - flower_dataset - flower_dataset.csv 283.7 KB Dec 8, 2025, 11:03 AM
 - Iris - Iris.csv 5 KB Dec 8, 2025, 11:06 AM
 - My Music
 - My Pictures
 - My Videos
 - NetBeansProjects
 - sales_data.csv 101.2 KB Dec 8, 2025, 12:04 PM
 - sales_dataset.csv 698 B Dec 2, 2025, 12:18 PM
 - sgfdu05m.ini 38 B May 4, 2025, 1:29 PM
 - Virtual Machines

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

Subject: Data Analysis with SAS / SPSS /R

RStudio interface showing R script execution and environment pane.

Code in R Script pane:

```
34 # - Logical: TRUE/FALSE counts
35
36 # =====
37 # 4. IMPROVING summary() BY CONVERTING CATEGORICAL COLUMNS TO FACTORS
38 # =====
39
40 # Convert categorical columns to factor for better summaries
41
```

Output of `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
1st Qu.:	:13.00	Electronics	:246	1st Qu.:1238.38
Median :	:25.00	Food	:226	Median :2467.24
Mean :	:25.36	Furniture	:260	Mean :2475.30
3rd Qu.:	:38.00			3rd Qu.:3702.86
Max. :	:49.00			Max. :3958.0

	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	
		Mean :0.1524		
		3rd Qu.:0.2300		
		Max. :0.3000		

RStudio interface showing R script execution and environment pane.

Code in R Script pane:

```
34 # - Logical: TRUE/FALSE counts
35
36 # =====
37 # 4. IMPROVING summary() BY CONVERTING CATEGORICAL COLUMNS TO FACTORS
38 # =====
39
40 # Convert categorical columns to factor for better summaries
41
```

Output of `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
1st Qu.:	:13.00	Electronics	:246	1st Qu.:1238.38
Median :	:25.00	Food	:226	Median :2467.24
Mean :	:25.36	Furniture	:260	Mean :2475.30
3rd Qu.:	:38.00			3rd Qu.:3702.86
Max. :	:49.00			Max. :3958.0

	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	
		Mean :0.1524		
		3rd Qu.:0.2300		
		Max. :0.3000		

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

Subject: Data Analysis with SAS / SPSS /R

The screenshot shows the RStudio interface. The left pane displays an R script with code for calculating average, maximum, and total sales. The right pane shows the R Environment pane with various variables and their values. Below the RStudio window is a taskbar with icons for weather, search, and other applications.

```
total_quantity <- sum(sales_df$Quantity_Sold, na.rm = TRUE)
print(paste("Average Sales Amount:", avg_sales))
print(paste("Maximum Sales Amount:", max_sales))
print(paste("Total Quantity Sold:", total_quantity))

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

> avg_sales <- mean(sales_df$Sales_Amount, na.rm = TRUE)
> max_sales <- max(sales_df$Sales_Amount, na.rm = TRUE)
> total_quantity <- sum(sales_df$Quantity_Sold, na.rm = TRUE)
> print(paste("Average Sales Amount:", avg_sales))
[1] "Average Sales Amount: 5019.26523"
> print(paste("Maximum Sales Amount:", max_sales))
[1] "Maximum Sales Amount: 9989.04"
> print(paste("Total Quantity Sold:", total_quantity))
[1] "Total Quantity Sold: 25355"
```

Environment pane variables:

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

File pane contents:

Name	Size	Modified
9 jugaar.circ	11.9 KB	Oct 24, 2024, 11:16 PM
Custom Office Templates		
Dell		
desktop.ini	402 B	Dec 8, 2024, 2:50 PM
Downloads - Shortcut.lnk	737 B	Mar 18, 2025, 3:29 PM
flower_dataset - flower_dataset.csv	283.7 KB	Dec 8, 2025, 11:03 AM
Iris - Iris.csv	5 KB	Dec 8, 2025, 11:06 AM
My Music		
My Pictures		
My Videos		
NetBeansProjects		
<input checked="" type="checkbox"/> sales_data.csv	101.2 KB	Dec 8, 2025, 12:04 PM
<input type="checkbox"/> sales_dataset.csv	698 B	Dec 2, 2025, 12:18 PM
<input type="checkbox"/> sgfdlu05m.ini	38 B	May 4, 2025, 1:29 PM
Virtual Machines		