

**SHETH L.U.J. AND SIR M.V. COLLEGE
SUBJECT: DATA ANALYSIS WITH R**

PRACTICAL NO.: M2 1 - 6

AIM: Practical 1 to 6

1. Generating descriptive statistics using `summary()` or `describe()` (R).

The screenshot shows two RStudio sessions running simultaneously. The top session, titled 'R2Studio', contains R code for reading CSV files, calculating descriptive statistics, and displaying the sales data frame. The bottom session, also titled 'RStudio', shows a similar workflow for another dataset. A browser window is open in the background, displaying a search results page for 'DEEPMALA CHAVAN'.

R2Studio Session:

```
1 list.files()
2 sales_data <- read.csv("D:/S076/sales_data.csv")
3 head(sales_data)
4
5 # Load the csv file
6 sales_data <- read.csv("D:/S076/sales_data.csv")
7
8 # View dataset structure
9 str(sales_data)
10
11 # Descriptive statistics using summary()
12 summary(sales_data)
13
14 # Descriptive statistics using describe()
15 install.packages("psych") # run only once
16 library(psych)
17 describe(sales_data)
18
```

18:1 [Top Level] ↴

Console Terminal Background Jobs

R 4.5.2 - ~

```
$ 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" ...
>
```

25°C Sunny

Search

10:23 15-12-2025

Session 2 (Bottom):

```
> list.files()
[1] "Custom Office Templates" "desktop.ini"           "GIS Database"
[3] "My Pictures"             "My Videos"          "My web sites"
[5] "Power BI Desktop"        "Virtual Machines"
> sales_data <- read.csv("D:/S076/sales_data.csv")
> head(sales_data)
```

Product_ID	Sale_Date	Sales_Rep	Region	Sales_Amount	Quantity_Sold	Product_Category	Unit_Cost	Unit_Price	Customer_Type	Discount
1	1092 2023-02-03	Bob	North	5053.97	18	Furniture	152.75	267.22	Returning	0.09
2	1093 2023-02-21	Bob	West	4326.20	17	Furniture	388.59	420.44	Returning	0.11
3	1051 2023-09-21	David	North	4631.23	30	Food	261.56	311.40	Returning	0.0
4	1072 2023-08-24	Bob	South	2167.94	39	Clothing	4330.03	4467.75	New	0.02
5	1061 2023-03-24	Charlie	East	3750.20	23	Electronics	637.37	692.71	New	0.08
6	1021 2023-02-11	charlie	West	3761.15	32	Food	900.79	1106.51	New	0.21

```
Payment_Method Sales_Channel Region_and_Sales_Rep
1 Cash online North-Bob
2 Cash Retail West-Bob
3 Bank Transfer Retail South-David
4 Credit Card Retail South-Bob
5 Credit Card online East-Charlie
6 Cash Online West-Charlie
>
# Load the CSV file
> sales_data <- read.csv("D:/S076/sales_data.csv")
>
# View dataset structure
> str(sales_data)
```

'data.frame': 1000 obs. of 14 variables:

Product_ID	Sale_Date	Sales_Rep	Region	Sales_Amount	quantity_sold	Product_category	Unit_cost	Unit_Price	Customer_Type	Discount	Payment_Method	Sales_Channel	Region_and_Sales_Rep
1	1092 2023-02-03	Bob	North	5053.97	18	Furniture	152.75	267.22	Returning	0.09	Cash	online	North-Bob
2	1093 2023-02-21	Bob	West	4326.20	17	Furniture	388.59	420.44	Returning	0.11	Cash	Retail	West-Bob
3	1051 2023-09-21	David	North	4631.23	30	Food	261.56	311.40	Returning	0.0	Bank Transfer	Retail	South-David
4	1072 2023-08-24	Bob	South	2167.94	39	Clothing	4330.03	4467.75	New	0.02	Credit Card	Retail	South-Bob
5	1061 2023-03-24	Charlie	East	3750.20	23	Electronics	637.37	692.71	New	0.08	Credit Card	online	East-Charlie
6	1021 2023-02-11	charlie	West	3761.15	32	Food	900.79	1106.51	New	0.21	Cash	Online	West-Charlie

```
>
# Descriptive statistics using summary()
> summary(sales_data)
```

Product_ID Sale_Date Sales_Rep Region Sales_Amount quantity_sold Product_category

Environment Tab (Top Right):

Global Environment	1 obs. of 1 variable
BostonHouse1	1000 obs. of 14 variables
BostonHouse2	506 obs. of 14 variables
sales_data	1000 obs. of 14 variables

Files Tab (Bottom Right):

Source	Version
User Library	
askpass	https://cran.rstudio.com 1.2.1
backports	https://cran.rstudio.com 1.5.0
base64enc	https://cran.rstudio.com 0.1-3
bit	https://cran.rstudio.com 4.6.0
bit64	https://cran.rstudio.com 4.6.0-1
blob	https://cran.rstudio.com 1.2.4
broom	https://cran.rstudio.com 1.0.10
bslib	https://cran.rstudio.com 0.9.0
cachem	https://cran.rstudio.com 1.1.0
cellranger	https://cran.rstudio.com 3.7.6
cli	https://cran.rstudio.com 3.6.5
clipr	https://cran.rstudio.com 0.8.0
conflicted	https://cran.rstudio.com 1.2.0
cpp11	https://cran.rstudio.com 0.5.2
crayon	https://cran.rstudio.com 1.5.3
curl	https://cran.rstudio.com 7.0.0
data.table	https://cran.rstudio.com 1.17.8
DBI	https://cran.rstudio.com 1.2.3
dplyr	https://cran.rstudio.com 5.5.1

10:23 15-12-2025

NAME: DEESHA CHAVAN
ROLL NO.: S076

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

SUBJECT: DATA ANALYSIS WITH R

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help

Source
Console Terminal x Background Jobs x

R - 4.5.2 - ~ / R

$ 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" ...
>
> # Descriptive statistics using summary()
> summary(sales_data)
#> Product_ID     Sale_Date       Sales_Rep      Region      Sales_Amount   Quantity_Sold Product_Category
#> Min. :1000    Length:1000    Length:1000    Min. :100.1    Min. :1.00    Length:1000
#> 1st Qu.:1024   Class :character  Class :character  1st Qu.:2550.3  1st Qu.:13.00  Class :character
#> Median :1051  Mode  :character  Mode  :character  Median :5019.3   Median :25.00  Mode  :character
#> Mean  :1050   Mean  :1050      Mean  :1050      Mean  :5019.3   Mean  :25.36  Mean  :1050
#> 3rd Qu.:1075  Max.  :1075      Max.  :1075      3rd Qu.:7507.4  3rd Qu.:38.00  Max.  :1075
#> Max.  :1090   unit_cost      unit_price    Customer_Type  Discount      Payment_Method Sales_Channel  Region_and_Sales_Rep
#> Min. :60.28   Min. :167.1    Length:1000    Min. :0.0000  Length:1000  Length:1000  Length:1000
#> 1st Qu.:1238.38  1st Qu.:1509.1  Class :character  1st Qu.:0.0800  Class :character  Class :character  Class :character
#> Median :2467.24  Median :2696.4  Mode  :character  Median :0.1500  Mode  :character  Mode  :character  Mode  :character
#> Mean  :2475.30  Mean  :2728.4  Mean  :0.1524  Mean  :0.1524  Mean  :0.1524  Mean  :0.1524  Mean  :0.1524
#> 3rd Qu.:3702.86  3rd Qu.:3958.0  Max.  :0.2300  3rd Qu.:0.2300  3rd Qu.:0.2300  3rd Qu.:0.2300  3rd Qu.:0.2300
#> Max. :4995.30  Max. :5442.1  Max.  :0.3000  Max.  :0.3000  Max.  :0.3000  Max.  :0.3000  Max.  :0.3000
>
> # Descriptive statistics using describe()
> install.packages("psych") # run only once
Restarting R session...
> install.packages("psych")

WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/itlab/AppData/Local/R/win-library/4.5'
(as 'lib' is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/psych_2.5.6.zip'
content type 'application/zip' length 3594552 bytes (3.4 MB)
downloaded 3.4 MB

package 'psych' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:/Users/itlab/AppData/Local/Temp/Rtmppacut4C/downloaded_packages
> |
```

25°C Sunny 10:24 15-12-2025

2. Generating frequency tables using table() or count() (R).

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help

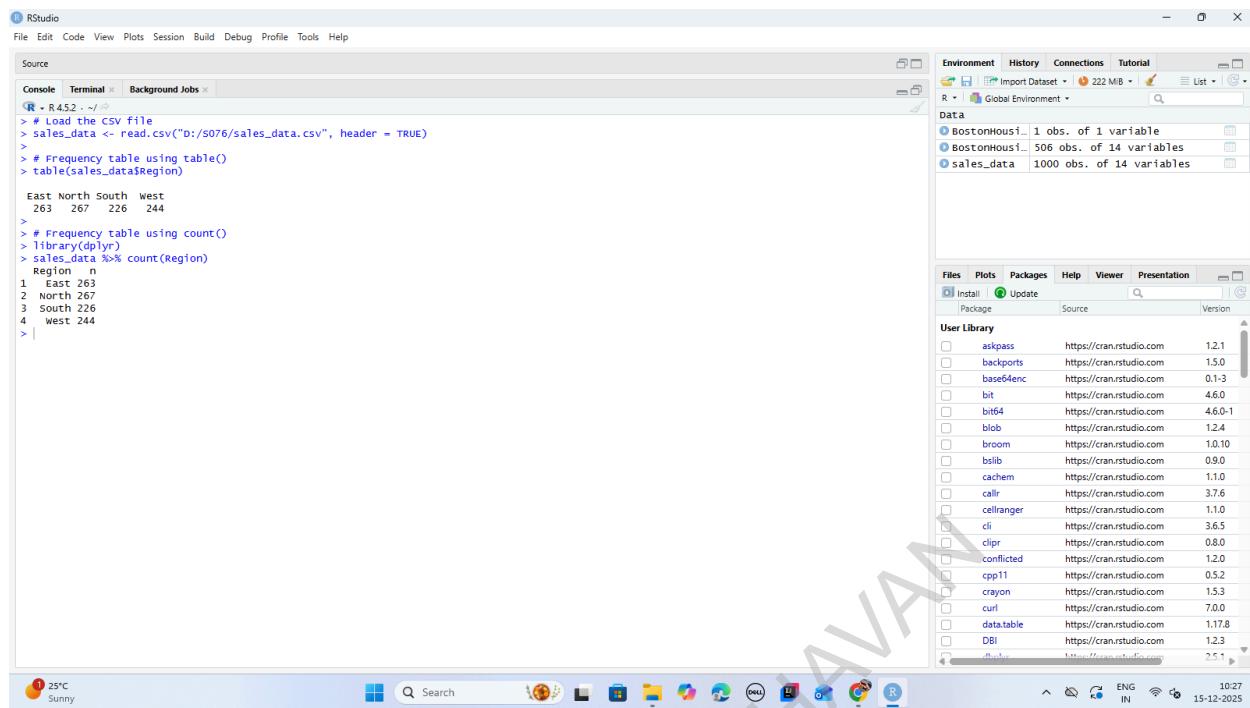
M2 PRAC 1.R x sales_data x

1 # Load the CSV file
2 sales_data <- read.csv("D:/S076/sales_data.csv", header = TRUE)
3
4 # Frequency table using table()
5 table(sales_data$Region)
6
7 # Frequency table using count()
8 library(dplyr)
9 sales_data %>% count(Region)
10

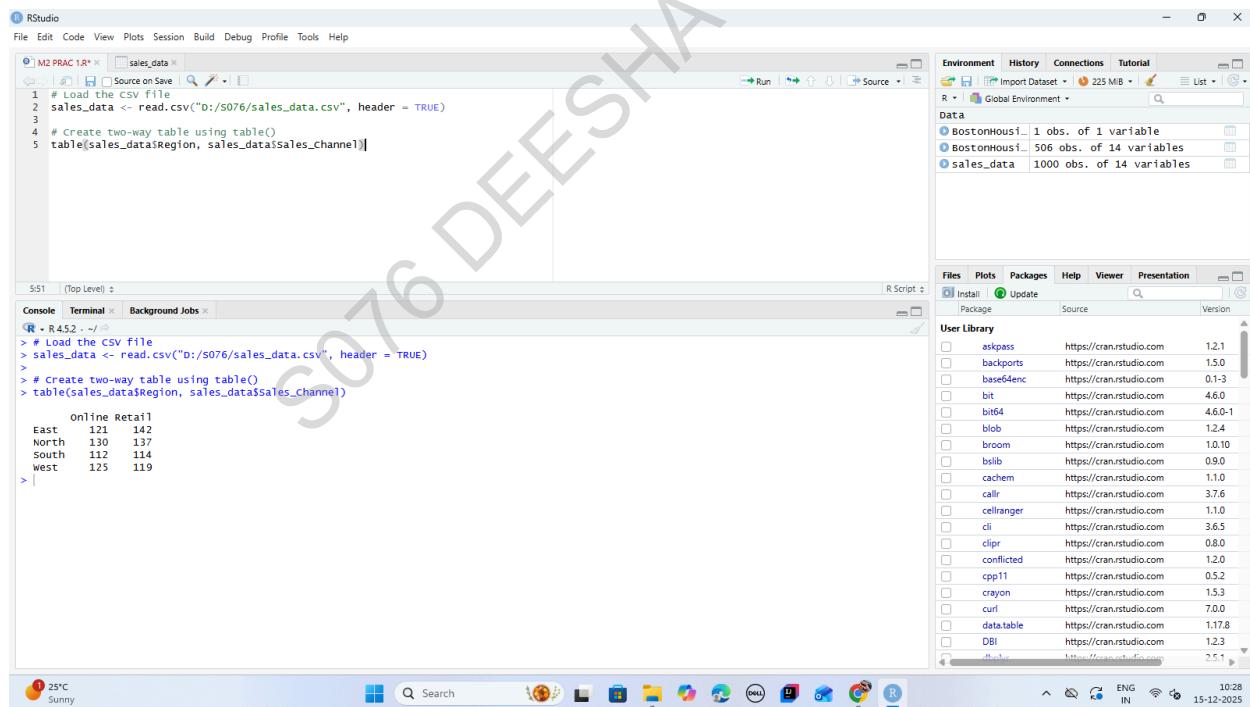
10:1 (Top Level) R Script

Console
25°C Sunny 10:27 15-12-2025
```

SHETH L.U.J. AND SIR M.V. COLLEGE
SUBJECT: DATA ANALYSIS WITH R

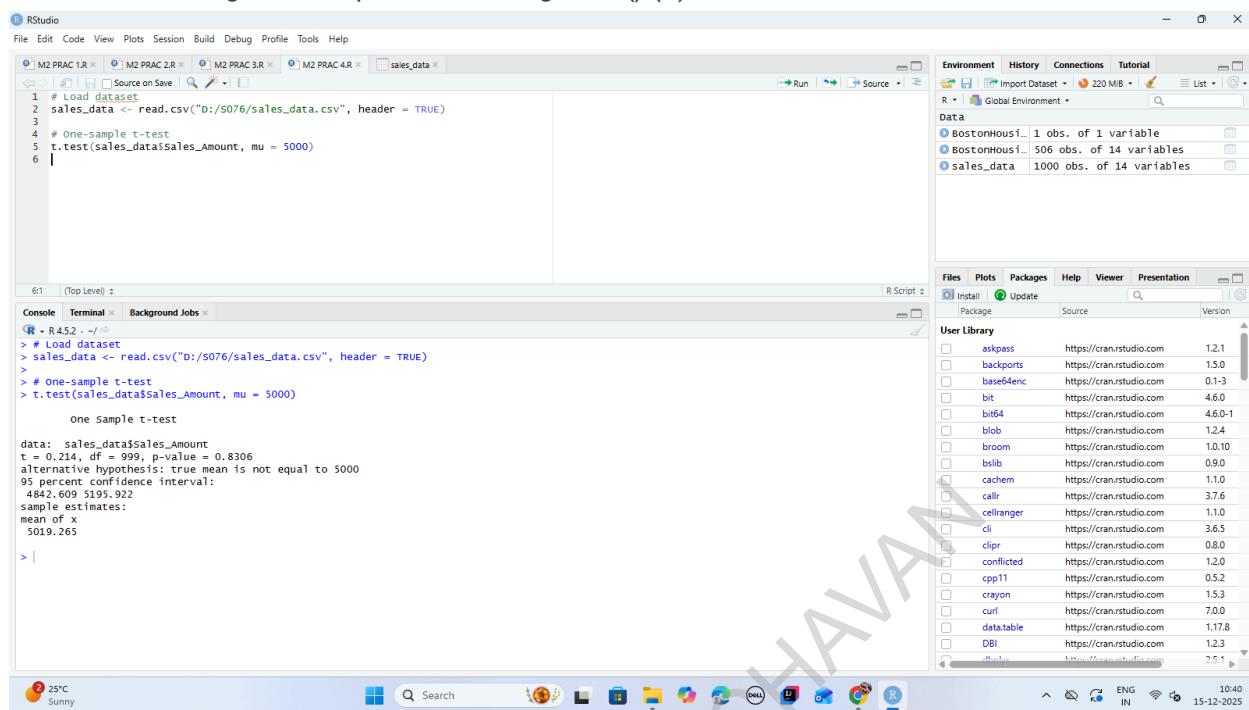


3. Creating cross-tabulations and two-way tables using table() (R).



SHETH L.U.J. AND SIR M.V. COLLEGE
SUBJECT: DATA ANALYSIS WITH R

4. Performing one-sample t-tests using t.test() (R).



The screenshot shows the RStudio interface with the following details:

- File Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Source Editor:** Shows R code for loading a dataset and performing a one-sample t-test.
- Console:** Displays the output of the R code, including the t-test results.
- Environment:** Shows objects in the global environment: BostonHouse (1 obs. of 1 variable), Bostonhouse (506 obs. of 14 variables), and sales_data (1000 obs. of 14 variables).
- Packages:** Shows installed packages and their versions.
- System Status:** Shows system information like temperature (25°C), battery level (Sunny), and system icons.

```

1 # Load dataset
2 sales_data <- read.csv("D:/S076/sales_data.csv", header = TRUE)
3 
4 # One-sample t-test
5 t.test(sales_data$Sales_Amount, mu = 5000)
6

```

```

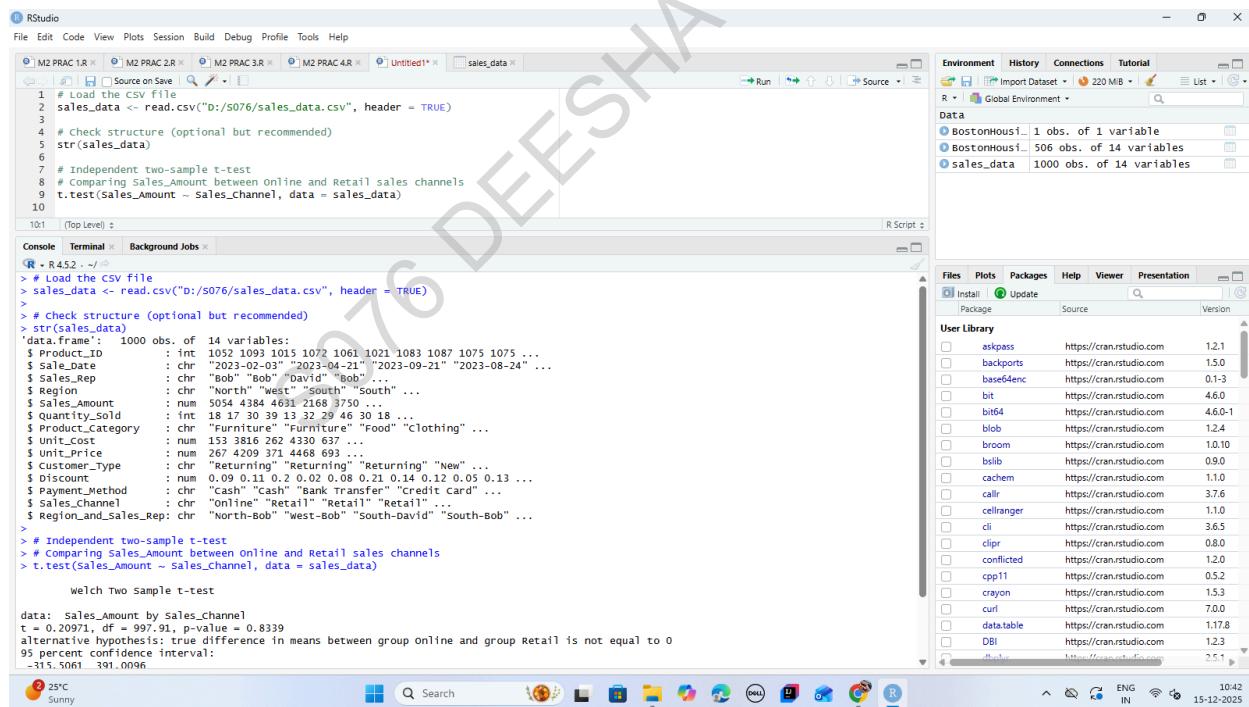
> # Load dataset
> sales_data <- read.csv("D:/S076/sales_data.csv", header = TRUE)
>
> # One-sample t-test
> t.test(sales_data$Sales_Amount, mu = 5000)

One Sample t-test

data: sales_data$Sales_Amount
t = 0.214, df = 999, p-value = 0.8306
alternative hypothesis: true mean is not equal to 5000
95 percent confidence interval:
4842.609 5195.922
sample estimates:
mean of x
5019.265

```

5. Performing independent two-sample t-tests using t.test() with grouping (R).



The screenshot shows the RStudio interface with the following details:

- File Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Source Editor:** Shows R code for loading a dataset and performing an independent two-sample t-test comparing Sales_Amount by Sales_Channel.
- Console:** Displays the output of the R code, including the t-test results.
- Environment:** Shows objects in the global environment: BostonHouse (1 obs. of 1 variable), Bostonhouse (506 obs. of 14 variables), and sales_data (1000 obs. of 14 variables).
- Packages:** Shows installed packages and their versions.
- System Status:** Shows system information like temperature (25°C), battery level (Sunny), and system icons.

```

1 # Load the CSV file
2 sales_data <- read.csv("D:/S076/sales_data.csv", header = TRUE)
3 
4 # Check structure (optional but recommended)
5 str(sales_data)
6 
7 # Independent two-sample t-test
8 # Comparing Sales_Amount between Online and Retail sales channels
9 t.test(Sales_Amount ~ sales_channel, data = sales_data)
10

```

```

> # Load the CSV file
> sales_data <- read.csv("D:/S076/sales_data.csv", header = TRUE)
>
> # check structure (optional but recommended)
> str(sales_data)
'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 2163 3750 ...
 $ Quantity_Sold: int 18 17 39 13 32 29 46 30 18 ...
 $ Product_Catagory: chr "Food" "Clothing" ...
 $ Unit_Cost    : num 153.3816 262.4330 631. ...
 $ Unit_Price   : num 267.4209 371.4468 693. ...
 $ Customer_Type: chr "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" ...
> # Independent two-sample t-test
> # Comparing Sales_Amount between Online and Retail sales channels
> t.test(Sales_Amount ~ Sales_Channel, data = sales_data)

Welch Two Sample t-test

data: Sales_Amount by Sales_Channel
t = 0.20971, df = 997.91, p-value = 0.8339
alternative hypothesis: true difference in means between group Online and group Retail is not equal to 0
95 percent confidence interval:
-15.5061 391.0096

```

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

SUBJECT: DATA ANALYSIS WITH R

The screenshot shows the RStudio interface with the following code in the script pane:

```

1 # Load the CSV file
2 sales_data <- read.csv("D:/S076/sales_data.csv", header = TRUE)
3
4 # Check structure (optional but recommended)
5 str(sales_data)
6
7 # Independent two-sample t-test
8 # Comparing Sales_Amount between online and retail sales channels
9 t.test(Sales_Amount ~ Sales_Channel, data = sales_data)
10

```

The console output shows the results of the Welch Two Sample t-test:

```

data: Sales_Amount by Sales_Channel
t = 0.20971, df = 997.91, p-value = 0.8339
alternative hypothesis: true difference in means between group online and group Retail is not equal to 0
95 percent confidence interval:
-315.5061 391.0096
sample estimates:
mean in group online mean in group Retail
      5038.594        5000.842

```

The environment pane shows the global variables:

- Bostonhouse1_1 obs. of 1 variable
- Bostonhouse1_506 obs. of 14 variables
- sales_data 1000 obs. of 14 variables

The packages pane lists the installed packages:

- askpass
- backports
- base64enc
- bit
- bit64
- blob
- broom
- bslib
- cachem
- callr
- cellranger
- cli
- clipr
- conflicted
- cpp11
- crayon
- curl
- data.table
- DBI
- httr

6. Performing paired t-tests using t.test(paired=TRUE) (R).

The screenshot shows the RStudio interface with the following code in the script pane:

```

1 # Load the CSV file
2 sales_data <- read.csv("D:/S076/sales_data.csv", header = TRUE)
3
4 # Select first 100 observations to form pairs
5 unit_cost <- sales_data$unit_cost[1:100]
6 unit_price <- sales_data$unit_price[1:100]
7
8 # Perform paired t-test
9 t.test(unit_cost, unit_price, paired = TRUE)
10

```

The console output shows the results of the Paired t-test:

```

Paired t-test

data: unit_cost and unit_price
t = -17.19, df = 99, p-value < 2.2e-16
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
-272.3805 -215.9735
sample estimates:
mean difference
-244.177

```

The environment pane shows the global variables:

- BOSTONHOUSE1_1 obs. of 1 variable
- BOSTONHOUSE1_506 obs. of 14 variables
- sales_data 1000 obs. of 14 variables

The packages pane lists the installed packages:

- askpass
- backports
- base64enc
- bit
- bit64
- blob
- broom
- bslib
- cachem
- callr
- cellranger
- cli
- clipr
- conflicted
- cpp11
- crayon
- curl
- data.table
- DBI
- httr