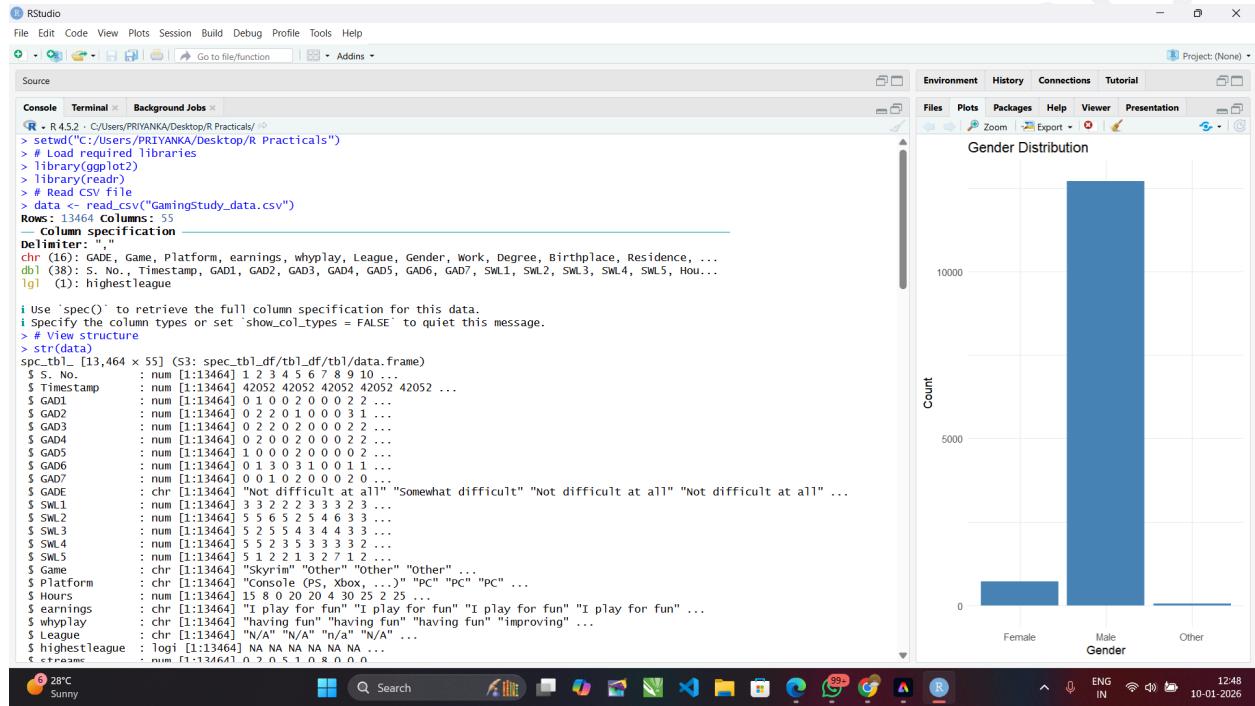


SHETH L.U.J AND SIR M.V. COLLEGE
SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R

Module 2 Practical 10-12

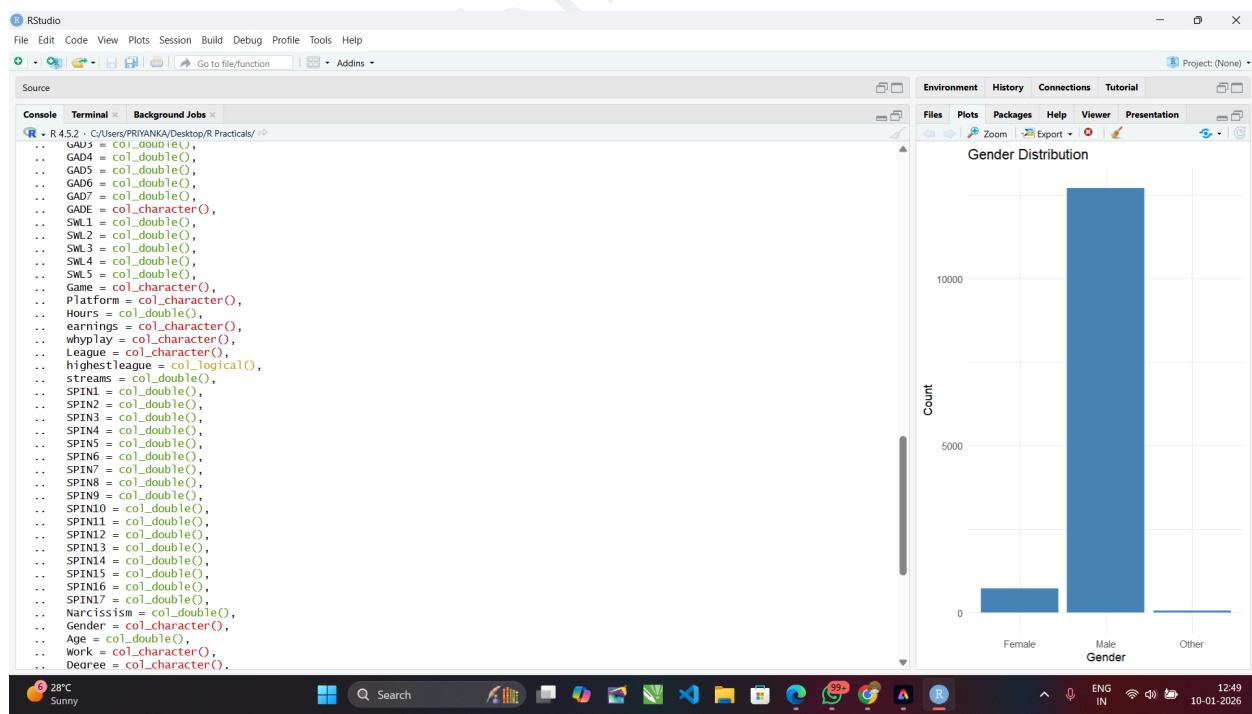
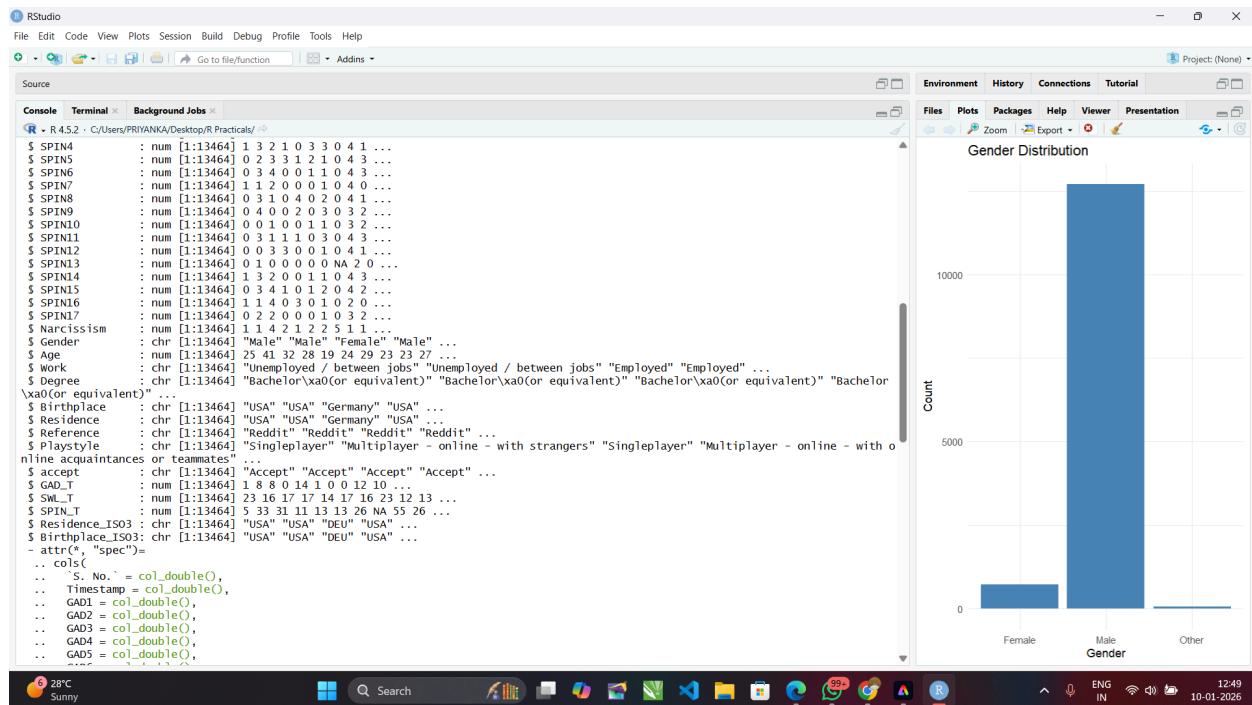
Aim: Creating graphical reports using ,ggplot2 (R)

OUTPUT:



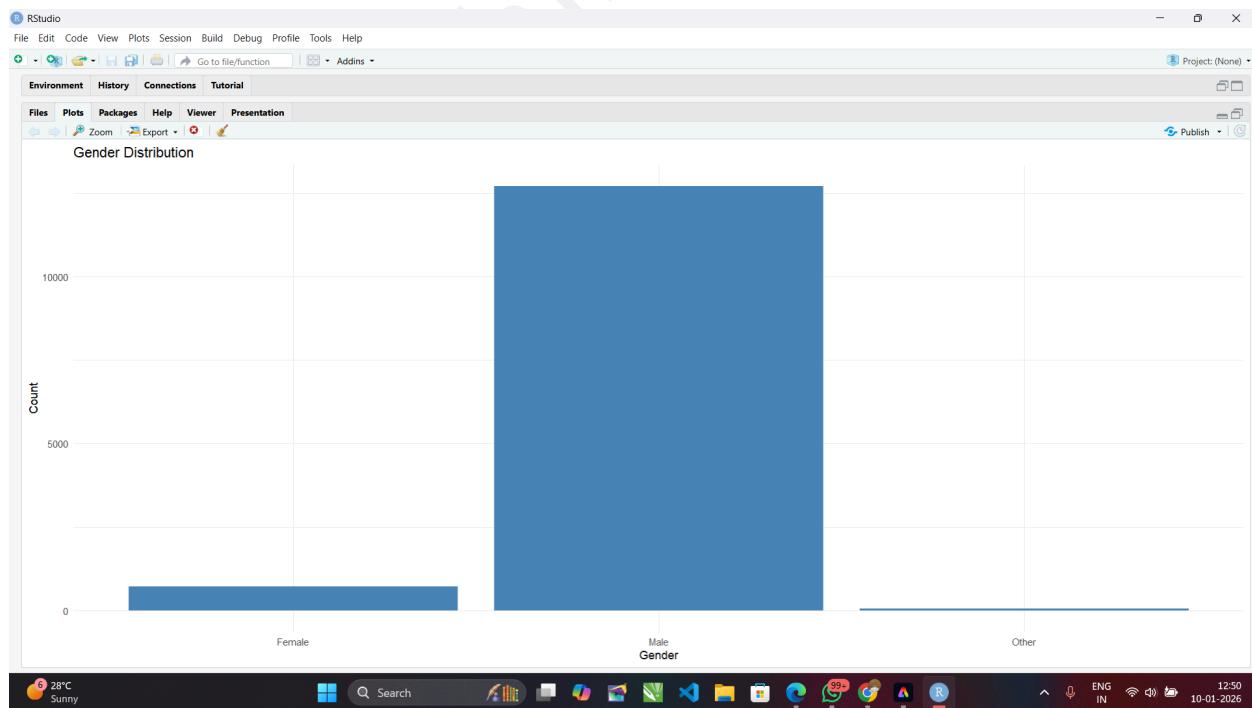
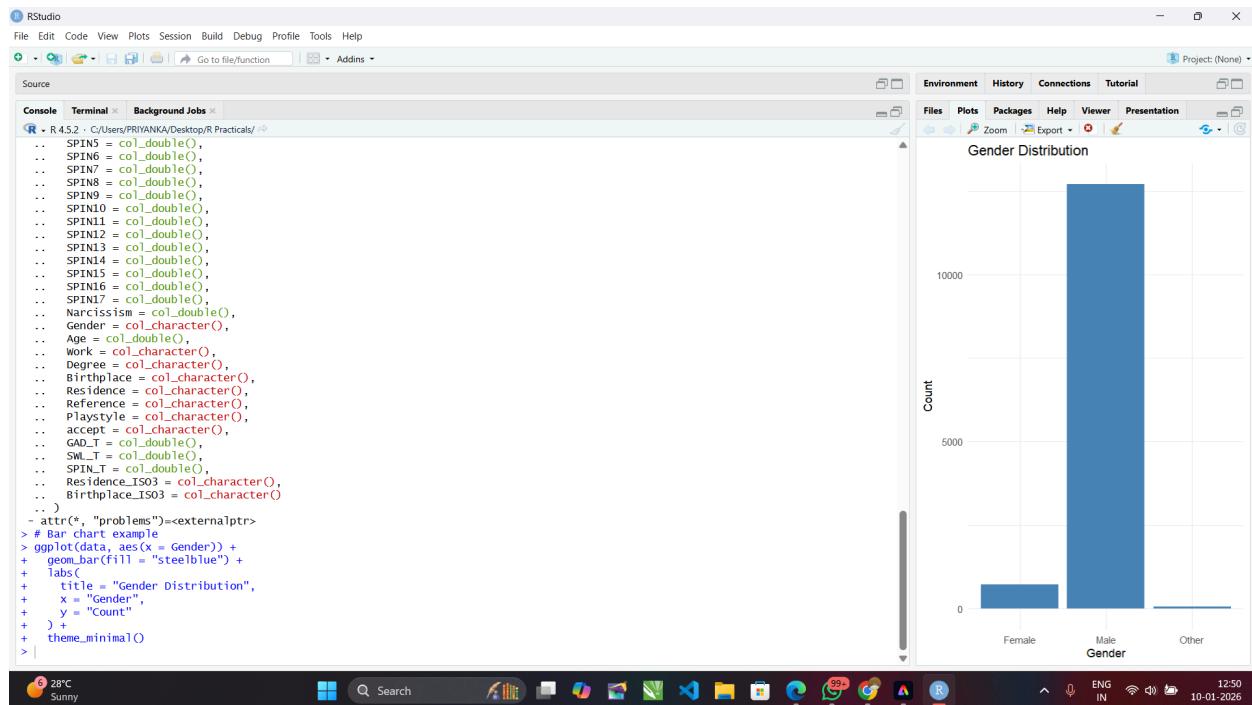
SHETH L.U.J AND SIR M.V. COLLEGE

SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R



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

SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R



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

Aim: Generating histograms and box plots using ggplot2 (R).

OUTPUT:

The screenshot shows an RStudio interface with the following details:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Source Editor:** Displays R code for reading data and creating a tibble, followed by a warning about column specification and types.
- Environment View:** Shows objects like `data` (13464 obs.), `Flight...` (10683 obs.), and `India...` (10683 obs.).
- Plots:** A box plot titled "Box Plot of Flight Tick" showing the distribution of "Ticket Price". The y-axis ranges from 0 to 80,000 with major ticks at 0, 20,000, 40,000, 60,000, and 80,000. The x-axis ranges from -0.4 to 0.4 with major ticks at -0.4, -0.2, 0, 0.2, and 0.4. The plot shows a median around 15,000, a box spanning approximately 10,000 to 25,000, and several outliers reaching up to 80,000.
- Bottom Status Bar:** Shows system icons for battery (99%), signal strength, and network.

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source

Console Terminal Background Jobs

Project: (None)

```
R 4.3.2 - /Users/PRIVANNA/Desktop/R Practicals/
#> Flight_Data : 'tbl_df' num [1:10683] 22:20:00 05:50:00 09:25:00 18:05:00 ...
#>   .attr("units")= chr "secs"
#> Arrival_Time : chr [1:10683] "01:10 22 Mar" "13:15" "04:25 10 Jun" "23:30" ...
#> Duration : chr [1:10683] "2h 50m" "7h 25m" "10h" "5h 25m" ...
#> Total_Stops : chr [1:10683] "non-stop" "2 stops" "2 stops" "1 stop" ...
#> Additional_Info : chr [1:10683] "No info" "No info" "No info" "No info" ...
#> Price : num [1:10683] 3897 7662 13882 6218 13302 ...
- attr(*, "spec")=
.. cols(
..   Airline = col_character(),
..   Date_of_Journey = col_character(),
..   Source = col_character(),
..   Destination = col_character(),
..   Route = col_character(),
..   Dep_Time = col_time(format = ""),
..   Arrival_Time = col_character(),
..   Duration = col_character(),
..   Total_Stops = col_character(),
..   Additional_Info = col_character(),
..   Price = col_double()
.. )
- attr(*, "problems")=<externalptr>
> ggplot(Flight_Data, aes(x = Price)) +
+   geom_histogram(binwidth = 1000, fill = "skyblue", color = "black") +
+   labs(
+     title = "Histogram of Flight Ticket Prices",
+     x = "Ticket Price",
+     y = "Frequency"
+   ) +
+   theme_minimal()
> ggplot(Flight_Data, aes(y = Price)) +
+   geom_boxplot(fill = "lightgreen") +
+   labs(
+     title = "Box Plot of Flight Ticket Prices",
+     y = "Ticket Price"
+   ) +
+   theme_minimal()
> |
```

USD/INR
+0.37%

Search

ENG IN 10-01-2026

Environment History Conn

Global Environment

Data

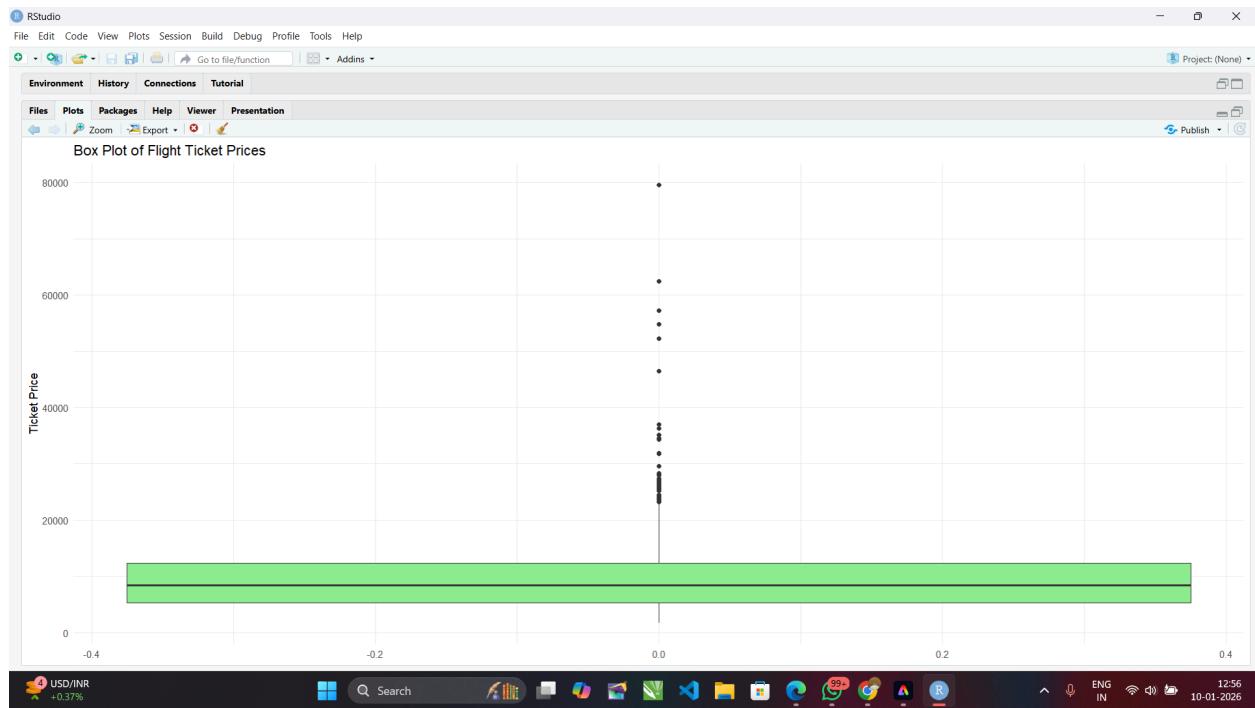
data 13464 obs. o...
flight 10683 obs. o...
India 10683 obs. o...

Files Plots Packages Help

Box Plot of Flight Ticket Prices

Ticket Price

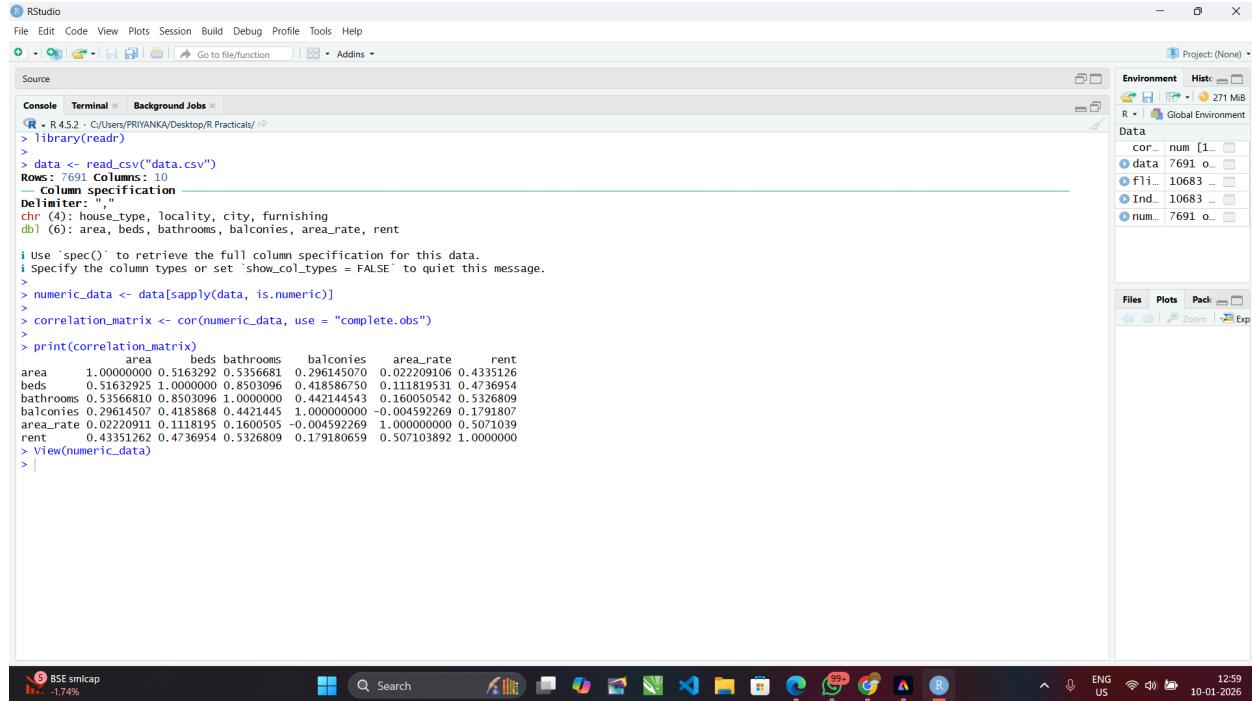
SHETH L.U.J AND SIR M.V. COLLEGE
SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R



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

Aim: Generating correlation matrices using cor() (R).

OUTPUT:



The screenshot shows the RStudio interface with the following details:

- Console:** Displays R code and its output. The code reads a CSV file named "data.csv" and calculates a correlation matrix for columns "area", "beds", "bathrooms", "balconies", "area_rate", and "rent".
- Data View:** Shows a preview of the "data" dataset with 7691 rows and 6 columns.
- Environment View:** Shows variables: cor_ (correlation matrix), data (original dataset), F1_ (first column), Ind_ (index vector), and num_ (numerical data).
- Plots View:** Empty.
- Session View:** Shows the current R version (4.5.2) and path (C:/Users/PRIYANKA/Desktop/R Practicals).
- Task View:** Shows various packages loaded.
- File Bar:** Includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- System Tray:** Shows battery level (17%), signal strength, and date/time (10-01-2026).

```
R - R 4.5.2 · C:/Users/PRIYANKA/Desktop/R Practicals/ >
> library(readr)
> 
> data <- read_csv("data.csv")
Rows: 7691 Columns: 10
--- Column specification ---
Delimiter: ","
chr (4): house_type, locality, city, furnishing
dbl (6): area, beds, bathrooms, balconies, area_rate, rent
i use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
> 
> numeric_data <- data[sapply(data, is.numeric)]
> 
> correlation_matrix <- cor(numeric_data, use = "complete.obs")
> 
> print(correlation_matrix)
      area   beds bathrooms   balconies   area_rate   rent
area  1.00000000 0.5163292 0.5356681 0.296145070 0.022209106 0.4335126
beds  0.51632925 1.0000000 0.8503096 0.418586750 0.111819531 0.4736954
bathrooms 0.53566810 0.8503096 1.0000000 0.442144543 0.160050542 0.5326809
balconies 0.29614507 0.4185868 0.4421445 1.00000000 -0.004592269 0.1791807
area_rate 0.02220911 0.1118195 0.1600505 -0.004592269 1.000000000 0.5071039
rent   0.43351262 0.4736954 0.5326809 0.179180659 0.507103892 1.0000000
> 
> View(numeric_data)
> |
```