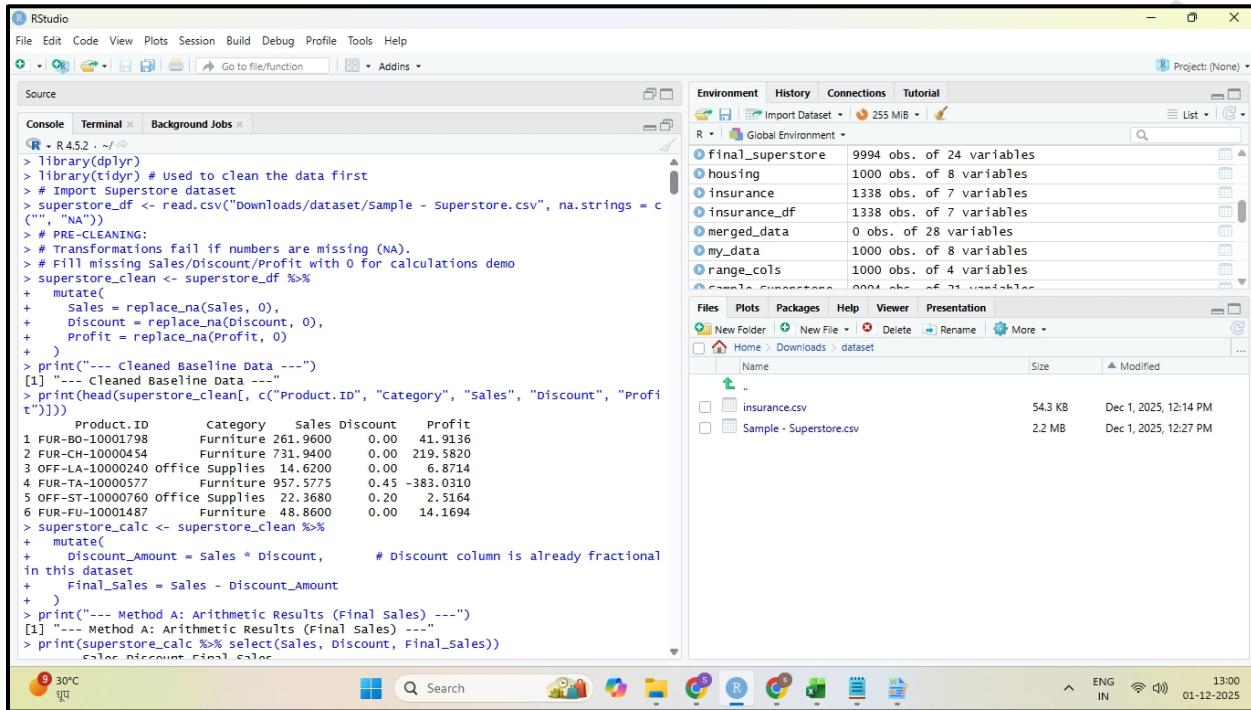


Sheth L.U.J. College of Arts & Sir M.V. College Of Science & Commerce
SUBJECT NAME: Data Analysis with SAS / SPSS/R

PRACTICAL NO : 10

AIM : Creating new variables using transformations and calculations in R. import dataset.



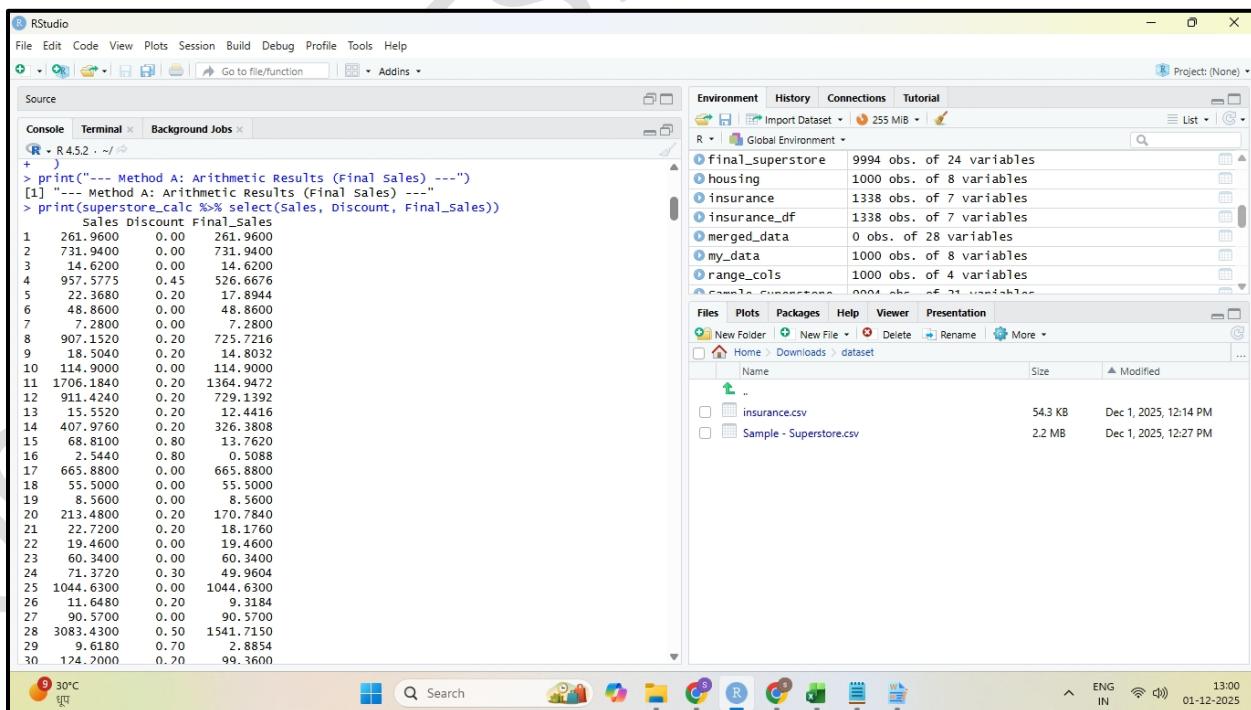
The screenshot shows the RStudio interface with the following details:

- Console:** Displays R code for reading a CSV file, cleaning it, and calculating a new column named "Final_Sales".
- Environment:** Shows the global environment with various datasets loaded, including "final_superstore", "housing", "insurance", etc.
- File Explorer:** Shows files in the "dataset" folder: "insurance.csv" (54.3 KB) and "Sample - Superstore.csv" (2.2 MB).
- System Tray:** Shows battery level (9%), temperature (30°C), and system status (ENG IN).

```

R RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source Environment History Connections Tutorial
Project: (None)
Console Terminal Background Jobs
> library(dplyr)
> library(tidyverse) # used to clean the data first
> # Import Superstore dataset
> superstore_df <- read.csv("Downloads/dataset/Sample - Superstore.csv", na.strings = c("", "NA"))
> # PRE-CLEANING:
> # Transformations fail if numbers are missing (NA).
> # Fill missing Sales/Discount/Profit with 0 for calculations demo
> superstore_clean <- superstore_df %>%
+   mutate(
+     Sales = replace_na(Sales, 0),
+     Discount = replace_na(Discount, 0),
+     Profit = replace_na(Profit, 0)
+   )
> print("--- Cleaned Baseline Data ---")
[1] "--- Cleaned Baseline Data ---"
> print(head(superstore_clean[, c("Product.ID", "Category", "Sales", "Discount", "Profit")]))
  Product.ID Category Sales Discount Profit
1 FUR-BO-10001798 Furniture 261.9600 0.00 41.9136
2 FUR-CH-10000454 Furniture 731.9400 0.00 219.5820
3 OFF-LA-10000240 Office Supplies 14.6200 0.00 6.8714
4 FUR-TA-10000577 Furniture 957.5775 0.45 -383.0310
5 OFF-ST-10000760 Office Supplies 22.3680 0.20 2.5164
6 FUR-FU-10001487 Furniture 48.8600 0.00 14.1694
> superstore_calc <- superstore_clean %>%
+   mutate(
+     Discount_Amount = Sales * Discount,      # Discount column is already fractional
+     Final_Sales = Sales - Discount_Amount
+   )
> print("--- Method A: Arithmetic Results (Final sales) ---")
[1] "--- Method A: Arithmetic Results (Final sales) ---"
> print(superstore_calc %>% select(Sales, Discount, Final_Sales))
  Sales Discount Final_Sales
1 261.9600 0.00 261.9600
2 731.9400 0.00 731.9400
3 14.6200 0.00 14.6200
4 957.5775 0.45 526.6676
5 22.3680 0.20 17.8944
6 48.8600 0.00 48.8600
7 7.2800 0.00 7.2800
8 907.1520 0.20 725.7216
9 18.5040 0.20 14.8032
10 114.9000 0.00 114.9000
11 1706.1840 0.20 1364.9472
12 911.4240 0.20 729.1392
13 15.5520 0.20 12.4416
14 407.9760 0.20 326.3808
15 68.8100 0.80 13.7620
16 2.5440 0.80 0.5088
17 665.8800 0.00 665.8800
18 55.5000 0.00 55.5000
19 8.5600 0.00 8.5600
20 213.4800 0.20 170.7840
21 22.7200 0.20 18.1760
22 19.4600 0.00 19.4600
23 60.3400 0.00 60.3400
24 71.3720 0.30 49.9604
25 1044.6300 0.00 1044.6300
26 11.6480 0.20 9.3184
27 90.5700 0.00 90.5700
28 3083.4300 0.50 1541.7150
29 9.6180 0.70 2.8854
30 174.2000 0.20 99.3600

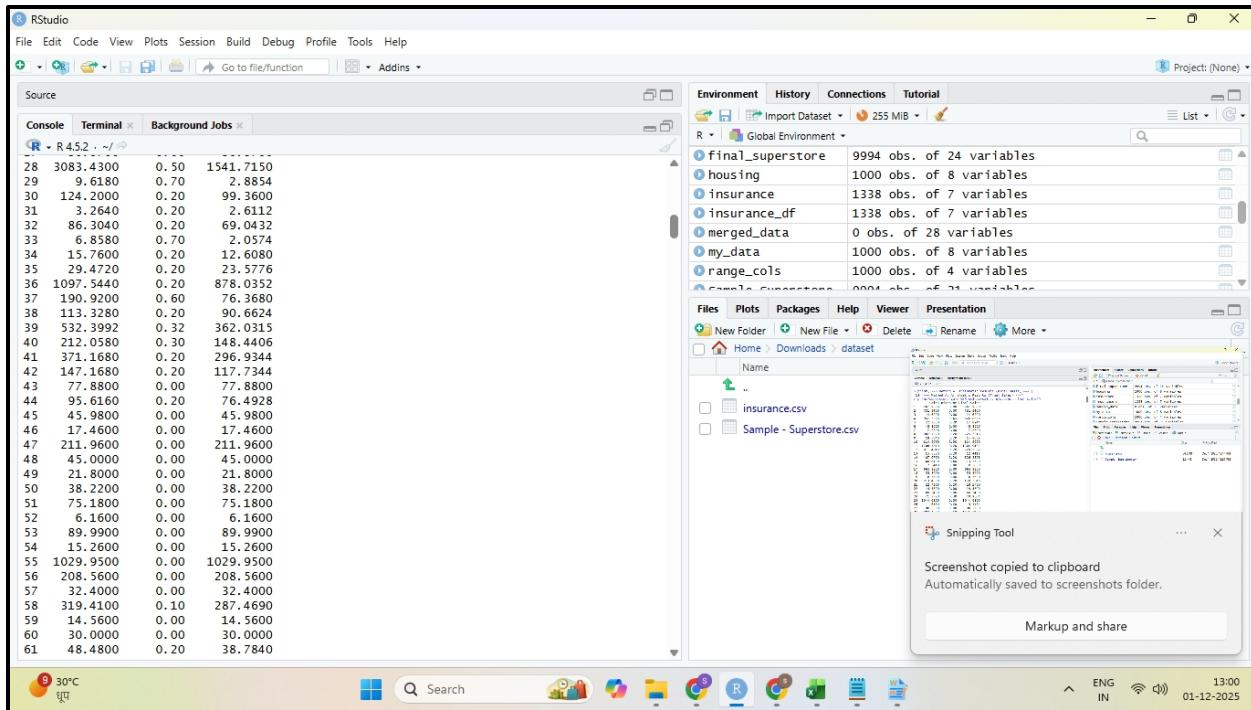
```



The screenshot shows the RStudio interface with the following details:

- Console:** Displays the same R code as the previous screenshot, showing the output of the arithmetic calculations.
- Environment:** Shows the global environment with various datasets loaded.
- File Explorer:** Shows files in the "dataset" folder: "insurance.csv" (54.3 KB) and "Sample - Superstore.csv" (2.2 MB).
- System Tray:** Shows battery level (9%), temperature (30°C), and system status (ENG IN).

Sheth L.U.J. College of Arts & Sir M.V. College Of Science & Commerce
SUBJECT NAME: Data Analysis with SAS / SPSS/R



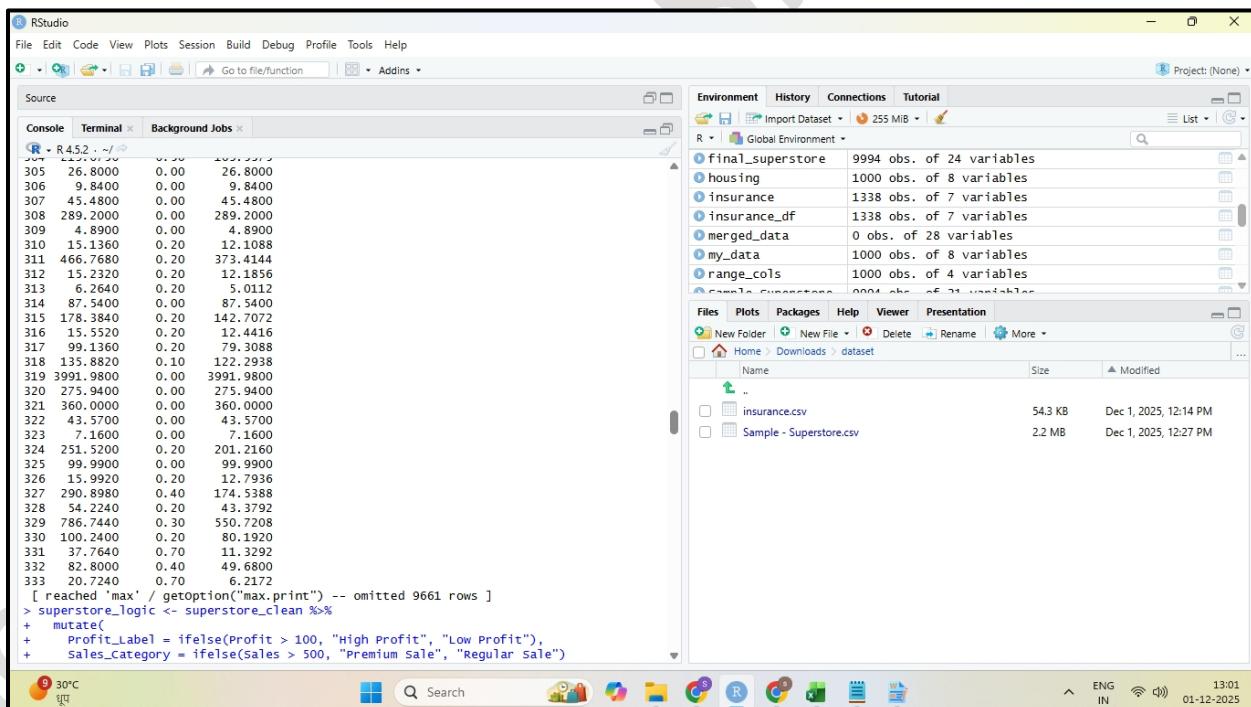
RStudio interface showing a console output of a data frame and the Global Environment pane.

```
R - R 4.5.2 - ~/Desktop
28 3083.4300 0.50 1541.7150
29 9.6180 0.70 2.8854
30 124.2000 0.20 99.3600
31 3.2640 0.20 2.6112
32 86.3040 0.20 69.0432
33 6.8580 0.70 2.0574
34 15.7600 0.20 12.6080
35 29.4720 0.20 23.5776
36 1097.5440 0.20 878.0352
37 190.9200 0.60 76.3680
38 113.3280 0.20 90.6624
39 532.3992 0.32 362.0315
40 212.0580 0.30 148.4406
41 371.1680 0.20 296.9344
42 147.1680 0.20 117.7344
43 77.8800 0.00 77.8800
44 95.6160 0.20 76.4928
45 45.9800 0.00 45.9800
46 17.4600 0.00 17.4600
47 211.9600 0.00 211.9600
48 45.0000 0.00 45.0000
49 21.8000 0.00 21.8000
50 38.2200 0.00 38.2200
51 75.1800 0.00 75.1800
52 6.1600 0.00 6.1600
53 89.9900 0.00 89.9900
54 15.2600 0.00 15.2600
55 1029.9500 0.00 1029.9500
56 208.5600 0.00 208.5600
57 32.4000 0.00 32.4000
58 319.4100 0.10 287.4690
59 14.5600 0.00 14.5600
60 30.0000 0.00 30.0000
61 48.4800 0.20 38.7840
```

The Global Environment pane shows the following objects:

- final_superstore: 9994 obs. of 24 variables
- housing: 1000 obs. of 8 variables
- insurance: 1338 obs. of 7 variables
- insurance_df: 1338 obs. of 7 variables
- merged_data: 0 obs. of 28 variables
- my_data: 1000 obs. of 8 variables
- range_cols: 1000 obs. of 4 variables
- sample_Superstore: 9994 obs. of 21 variables

A Snipping Tool window is open, showing a screenshot of the RStudio interface. The status bar at the bottom right indicates: ENG IN, 13:00, 01-12-2025.



RStudio interface showing a console output of a data frame and the Global Environment pane.

```
R - R 4.5.2 - ~/Desktop
305 26.8000 0.00 26.8000
306 9.8400 0.00 9.8400
307 45.4800 0.00 45.4800
308 289.2000 0.00 289.2000
309 4.8900 0.00 4.8900
310 15.1360 0.20 12.1088
311 466.7680 0.20 373.4144
312 15.2320 0.20 12.1856
313 6.2640 0.20 5.0112
314 87.5400 0.00 87.5400
315 178.3840 0.20 142.7072
316 15.5520 0.20 12.4416
317 99.1360 0.20 79.3088
318 135.8820 0.10 122.2938
319 3991.9800 0.00 3991.9800
320 275.9400 0.00 275.9400
321 360.0000 0.00 360.0000
322 43.5700 0.00 43.5700
323 7.1600 0.00 7.1600
324 251.5200 0.20 201.2160
325 99.9900 0.00 99.9900
326 15.9920 0.20 12.7936
327 290.8980 0.40 174.5388
328 54.2240 0.20 43.3792
329 786.7440 0.30 550.7208
330 100.2400 0.20 80.1920
331 37.7640 0.70 11.3292
332 82.8000 0.40 49.6800
333 20.7240 0.70 6.2172
[ reached 'max' / getoptoption("max.print") -- omitted 9661 rows ]
> superstore_logic <- superstore_clean %>
+ mutate(
+   Profit_Label = ifelse(Profit > 100, "High Profit", "Low Profit"),
+   Sales_Category = ifelse(Sales > 500, "Premium Sale", "Regular Sale")
```

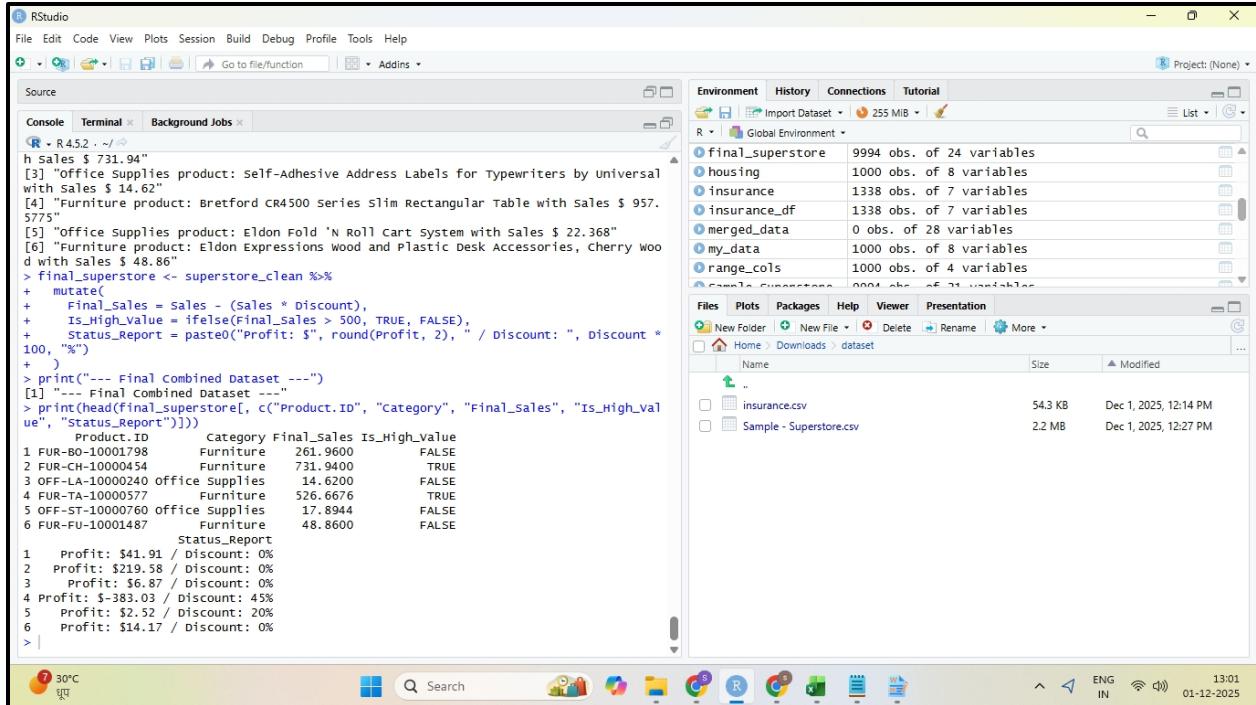
The Global Environment pane shows the same objects as the first screenshot.

The status bar at the bottom right indicates: ENG IN, 13:01, 01-12-2025.

Sheth L.U.J. College of Arts & Sir M.V. College Of Science & Commerce
SUBJECT NAME: Data Analysis with SAS / SPSS/R

The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help, and a Project dropdown set to '(None)'. The main area has tabs for Source, Console, Terminal, and Background Jobs. The Console tab displays a series of R commands and their outputs, including a large dataset named 'superstore_text' and its transformation into 'final_superstore'. The right side features a sidebar with tabs for Environment, History, Connections, and Tutorial. Below the sidebar is a file browser with sections for Files, Plots, Packages, Help, Viewer, and Presentation. It shows a list of files including 'insurance.csv' and 'Sample - Superstore.csv'. The bottom of the screen shows the Windows taskbar with various pinned icons.

Sheth L.U.J. College of Arts & Sir M.V. College Of Science & Commerce
SUBJECT NAME: Data Analysis with SAS / SPSS/R



The screenshot shows the RStudio interface. The left pane displays the R console output:

```
R - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source Terminal Background Jobs
Console Terminal x Background Jobs x
R - R 4.5.2 - ~
[1] "Sales $ 731.94"
[3] "Office supplies product: self-Adhesive Address Labels for Typewriters by universal with sales $ 14.62"
[4] "Furniture product: Bretford CR4500 Series Slim Rectangular Table with Sales $ 957.575"
[5] "Office supplies product: Eldon Fold 'N Roll Cart System with Sales $ 22.368"
[6] "Furniture product: Eldon Expressions Wood and Plastic Desk Accessories, Cherry wood with Sales $ 48.86"
> final_superstore <- superstore_clean %>
+   mutate(
+     Final_Sales = Sales - (Sales * Discount),
+     Is_High_Value = ifelse(Final_Sales > 500, TRUE, FALSE),
+     Status_Report = paste0("Profit: $", round(Profit, 2), " / Discount: ", round(Discount * 100, "%"))
+   )
> print("--- Final Combined Dataset ---")
[1] "--- Final Combined Dataset ---"
> print(head(final_superstore[, c("Product.ID", "Category", "Final_sales", "Is_High_Value", "Status_Report")]))
  Product.ID Category Final_sales Is_High_Value
1 FUR-B0-10001798 Furniture    261.9600    FALSE
2 FUR-CH-10000454 Furniture    731.9400    TRUE
3 OFF-LA-10000240 office_supplies 14.6200    FALSE
4 FUR-TA-10000577 Furniture    526.6676    TRUE
5 OFF-ST-10000760 office_supplies 17.8944    FALSE
6 FUR-FU-10001487 Furniture    48.8600    FALSE
  Status_Report
1   Profit: $41.91 / Discount: 0%
2   Profit: $219.58 / Discount: 0%
3   Profit: $6.87 / Discount: 0%
4 Profit: $-383.03 / Discount: 45%
5   Profit: $2.52 / Discount: 20%
6   Profit: $14.17 / Discount: 0%
```

The right pane shows the RStudio environment:

Object	Description
final_superstore	9994 obs. of 24 variables
housing	1000 obs. of 8 variables
insurance	1338 obs. of 7 variables
insurance_df	1338 obs. of 7 variables
merged_data	0 obs. of 28 variables
my_data	1000 obs. of 8 variables
range_cols	1000 obs. of 4 variables
Sample_Superstore	2000 obs. of 21 variables

Below the environment pane is a file browser showing two CSV files:

Name	Size	Modified
insurance.csv	54.3 KB	Dec 1, 2025, 12:14 PM
Sample - Superstore.csv	2.2 MB	Dec 1, 2025, 12:27 PM