

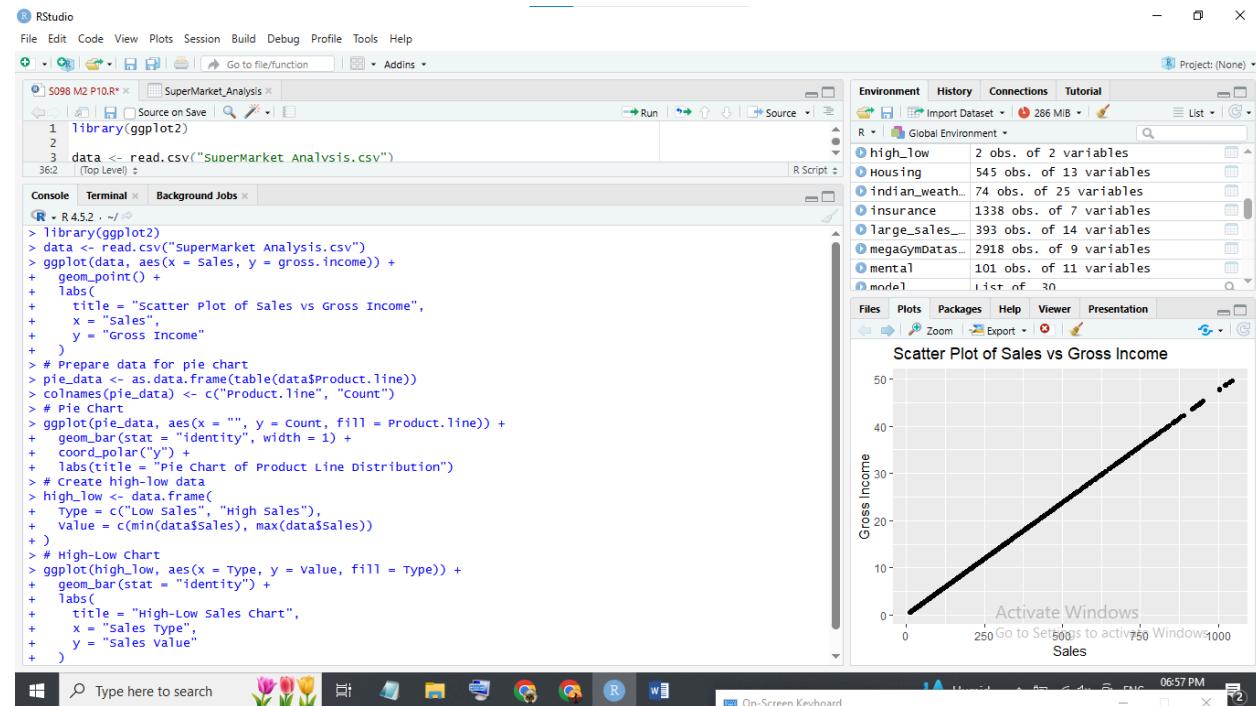
MVLU COLLEGE

Subject:-Data Analysis with SAS / SPSS /R

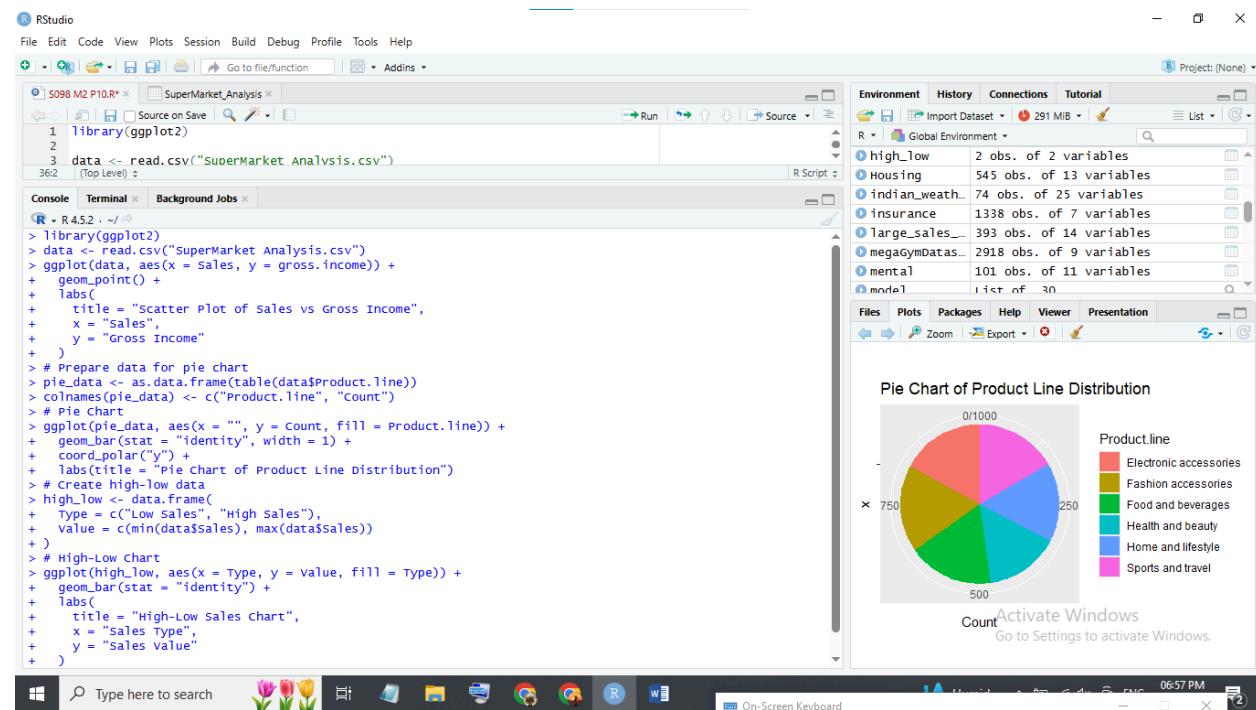
10 Creating graphical reports using ,ggplot2 (R).

Output:-

Scatter Plot:-



Pie Chart:-

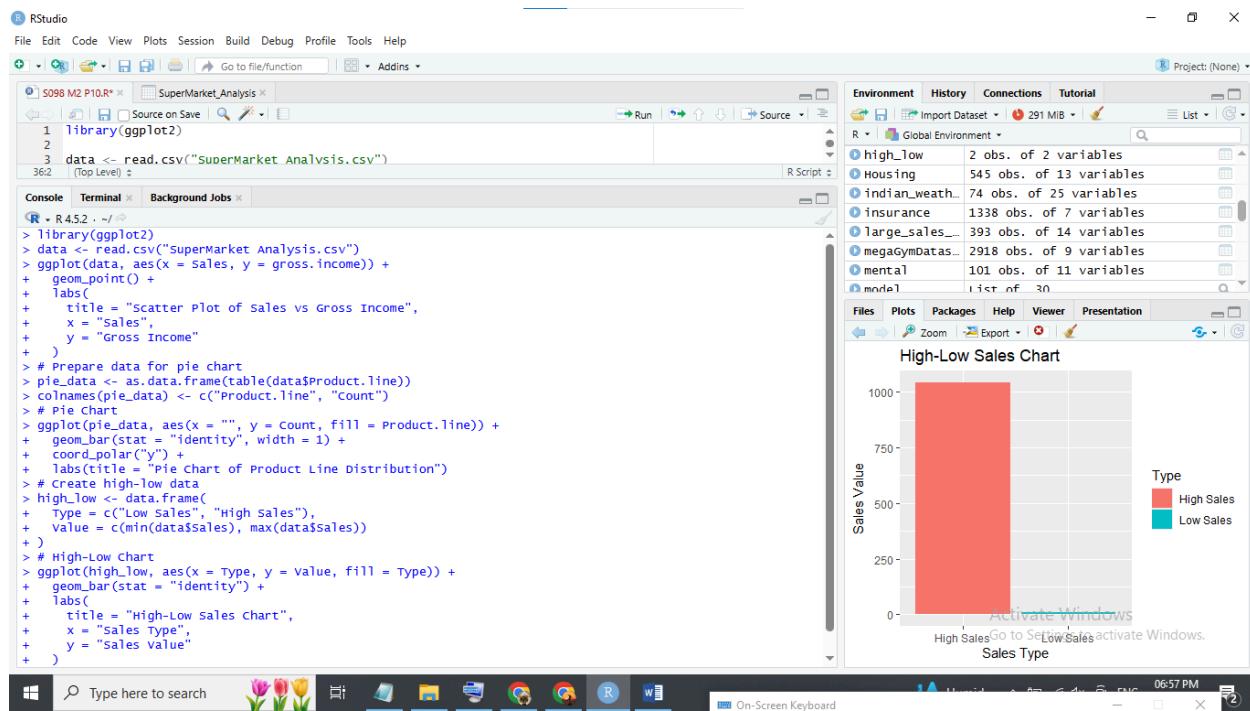


Vikas Pal
S098

MVLU COLLEGE

Subject:-Data Analysis with SAS / SPSS /R

High-low Chart:-

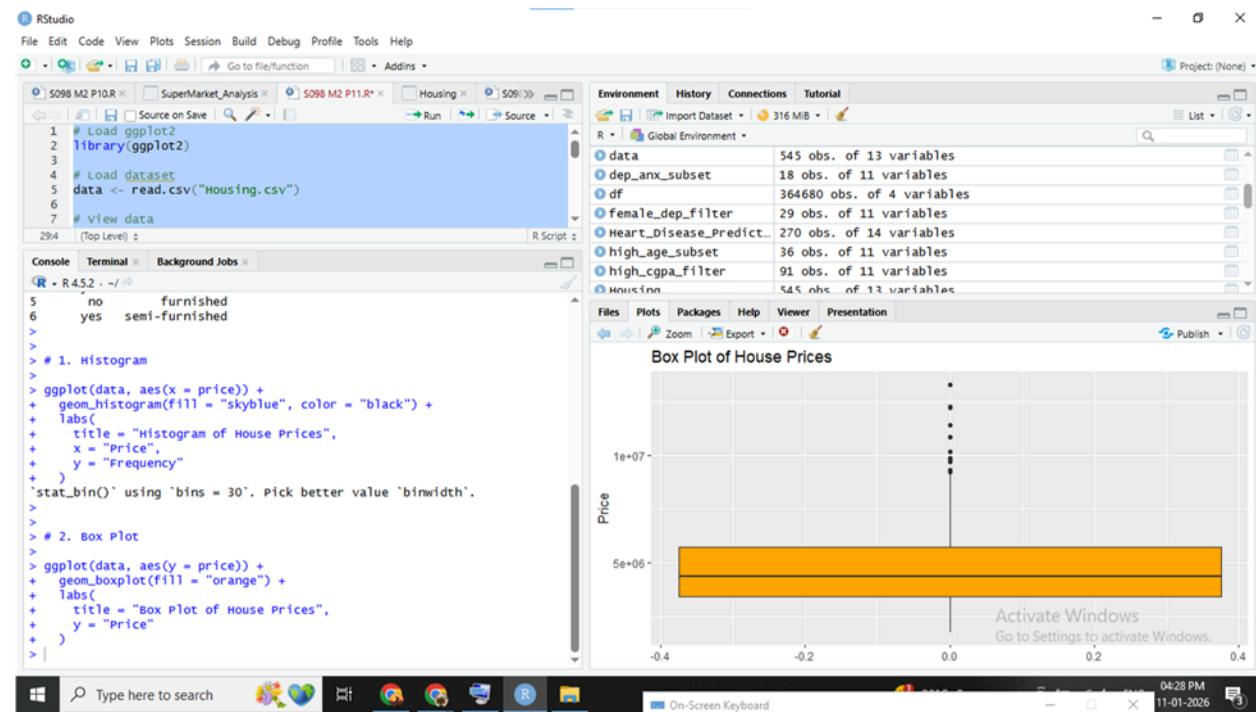
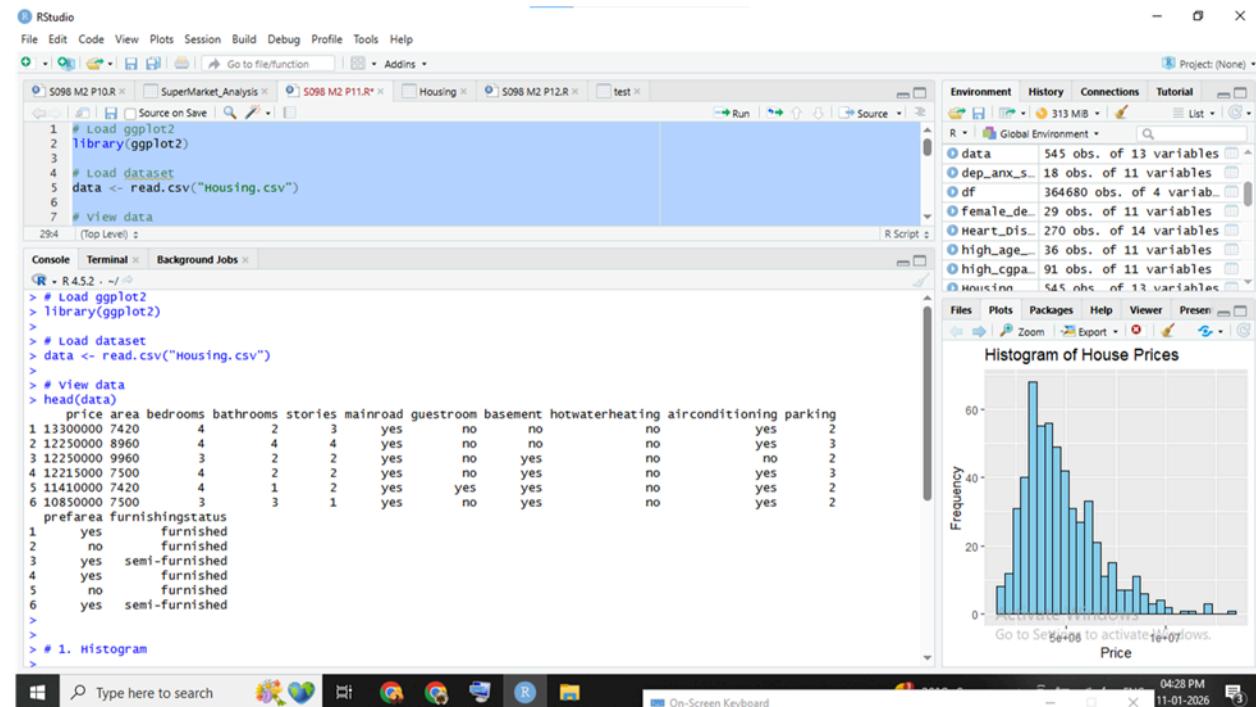


MVLU COLLEGE

Subject:-Data Analysis with SAS / SPSS /R

11 Generating histograms and box plots using ggplot2 (R).

Output:-

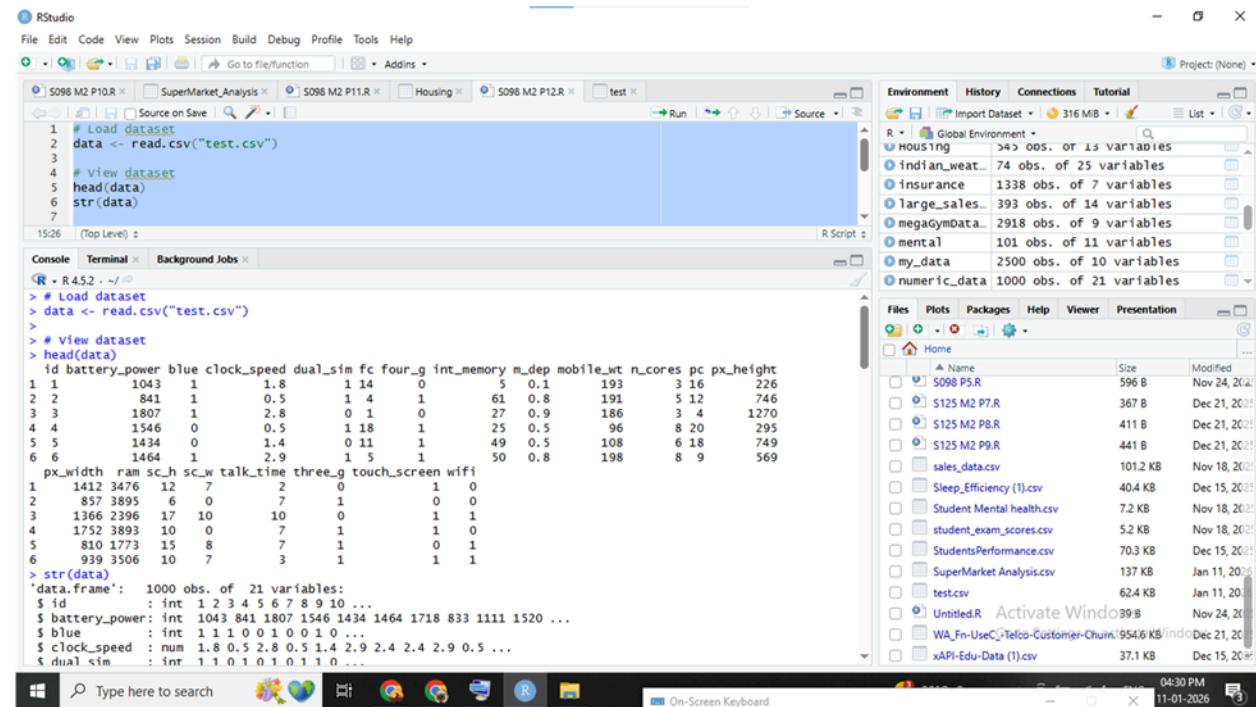


MVLU COLLEGE

Subject:-Data Analysis with SAS / SPSS /R

12 Generating correlation matrices using cor() (R).

Output:-



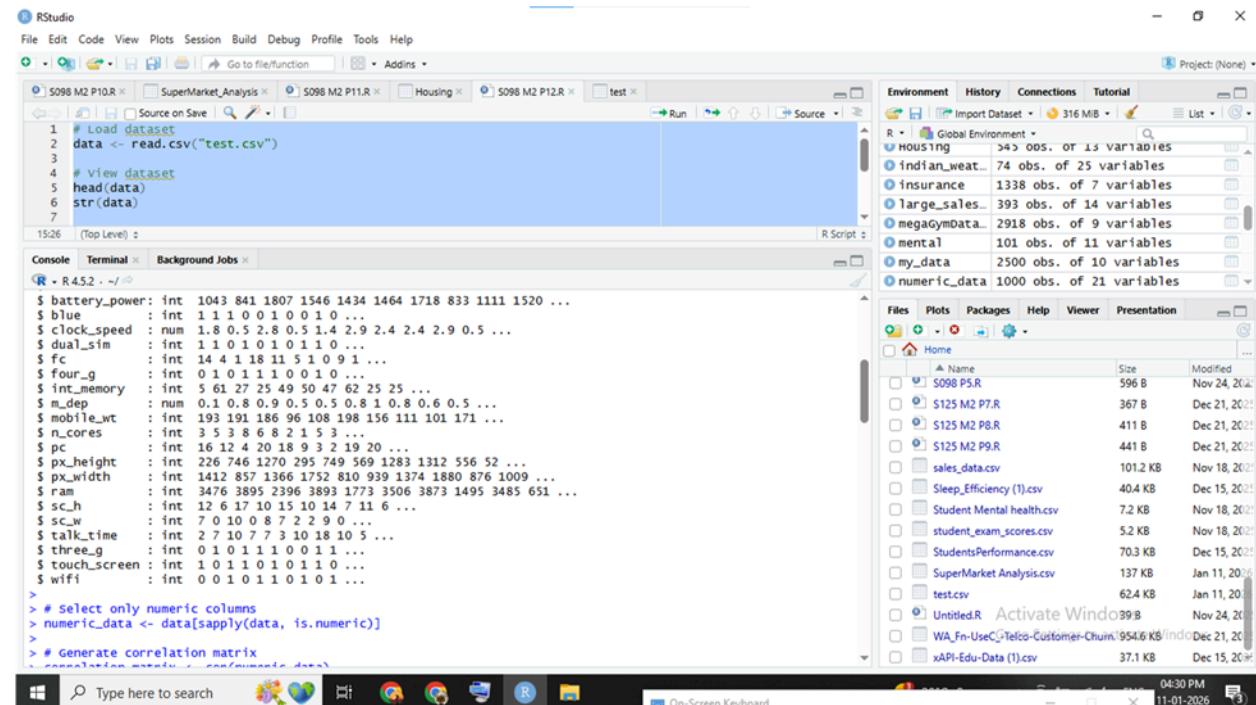
The screenshot shows the RStudio interface with the R console tab selected. The console displays the following R code and its output:

```

# Load dataset
> data <- read.csv("test.csv")
> # View dataset
> head(data)
> str(data)
#data.frame': 1000 obs. of 21 variables:
$ id : int 1 2 3 4 5 6 7 8 9 10 ...
$ battery_power : int 1043 841 1807 1546 1434 1464 1718 833 1111 1520 ...
$ blue : int 1 1 1 0 0 1 0 0 1 0 ...
$ clock_speed : num 1.8 0.5 2.8 0.5 1.4 2.9 2.4 2.4 2.9 0.5 ...
$ dual_sim : int 1 1 0 1 0 1 0 1 1 0 ...
$ fc : int 14 4 1 18 11 5 1 0 9 1 ...
$ four_g : int 0 1 0 1 1 0 0 1 0 ...
$ int_memory : int 5 61 27 25 49 50 47 62 25 25 ...
$ m_dep : num 0.1 0.8 0.9 0.5 0.5 0.8 1 0.8 0.6 0.5 ...
$ mobile_wt : int 193 198 96 108 198 156 111 101 171 ...
$ n_cores : int 3 5 3 8 6 8 2 1 5 3 ...
$ pc : int 16 12 4 20 18 9 3 2 19 20 ...
$ px_height : int 226 746 1270 295 749 569 1283 1312 556 52 ...
$ px_width : int 1412 857 1366 1752 810 939 1374 1880 876 1009 ...
$ ram : int 3476 3895 2396 3893 1773 3506 3873 1495 3485 651 ...
$ sc_h : int 12 6 17 10 15 10 14 7 11 6 ...
$ sc_w : int 7 0 10 0 8 7 2 2 9 0 ...
$ talk_time : int 2 7 10 7 7 3 10 18 10 5 ...
$ three_g : int 0 1 0 1 1 0 0 1 1 ...
$ touch_screen : int 1 0 1 1 0 1 0 1 1 0 ...
$ wifi : int 0 0 1 0 1 1 0 1 0 1 ...

```

The RStudio environment sidebar shows various datasets loaded, including `HOUSING`, `indian_wheat`, `insurance`, `large_sales`, `megadymData`, `mental`, `my_data`, and `numeric_data`. The file explorer sidebar shows the project structure and files.



The screenshot shows the RStudio interface with the R console tab selected. The console displays the following R code and its output:

```

# Load dataset
> data <- read.csv("test.csv")
> # View dataset
> head(data)
> str(data)
#data.frame': 1000 obs. of 21 variables:
$ battery_power : int 1043 841 1807 1546 1434 1464 1718 833 1111 1520 ...
$ blue : int 1 1 1 0 0 1 0 0 1 0 ...
$ clock_speed : num 1.8 0.5 2.8 0.5 1.4 2.9 2.4 2.4 2.9 0.5 ...
$ dual_sim : int 1 1 0 1 0 1 0 1 1 0 ...
$ fc : int 14 4 1 18 11 5 1 0 9 1 ...
$ four_g : int 0 1 0 1 1 0 0 1 0 ...
$ int_memory : int 5 61 27 25 49 50 47 62 25 25 ...
$ m_dep : num 0.1 0.8 0.9 0.5 0.5 0.8 1 0.8 0.6 0.5 ...
$ mobile_wt : int 193 198 96 108 198 156 111 101 171 ...
$ n_cores : int 3 5 3 8 6 8 2 1 5 3 ...
$ pc : int 16 12 4 20 18 9 3 2 19 20 ...
$ px_height : int 226 746 1270 295 749 569 1283 1312 556 52 ...
$ px_width : int 1412 857 1366 1752 810 939 1374 1880 876 1009 ...
$ ram : int 3476 3895 2396 3893 1773 3506 3873 1495 3485 651 ...
$ sc_h : int 12 6 17 10 15 10 14 7 11 6 ...
$ sc_w : int 7 0 10 0 8 7 2 2 9 0 ...
$ talk_time : int 2 7 10 7 7 3 10 18 10 5 ...
$ three_g : int 0 1 0 1 1 0 0 1 1 ...
$ touch_screen : int 1 0 1 1 0 1 0 1 1 0 ...
$ wifi : int 0 0 1 0 1 1 0 1 0 1 ...

```

The RStudio environment sidebar shows various datasets loaded, including `HOUSING`, `indian_wheat`, `insurance`, `large_sales`, `megadymData`, `mental`, `my_data`, and `numeric_data`. The file explorer sidebar shows the project structure and files.

MVLU COLLEGE

Subject:-Data Analysis with SAS / SPSS /R

RStudio interface showing R code and its output. The code reads a CSV file named 'test.csv' and prints a correlation matrix for numeric variables.

```

1 # Load dataset
2 data <- read.csv("test.csv")
3
4 # View dataset
5 head(data)
6 str(data)
7
15:26 (Top Level) : 

```

```

> correlation_matrix <- cor(numeric_data)
>
> # Display correlation matrix
> print(correlation_matrix)
      id battery_power    bluu   clock_speed dual_sim     fc
id  1.000000000 -0.0215114289  0.0004644277  0.0359170888 -0.0027208943  0.016933840
battery_power -0.0215114289  1.0000000000 -0.0466096315  -0.0390754297 -0.0611708963 -0.007846110
bluu   0.0004644277 -0.0466096315  1.0000000000  0.0347539970 -0.0111000991 -0.056063410
clock_speed  0.0359170888 -0.0390754297  0.0347539970  1.0000000000 -0.0124226539  0.010127218
dual_sim   -0.0027208943 -0.0611708963  -0.0111000991 -0.0124226539  1.0000000000  0.057606192
fc      0.0169338406  -0.0078461096 -0.0560634097  0.0101272178  0.0576061919  1.000000000
four_g    0.0309206471 -0.0425200594 -0.0011689937 -0.0246648231  0.0249068154  0.032832204
int_memory -0.0140234070  0.0037509249 -0.0124163270 -0.0304868400 -0.0121578310 -0.006564568
m_dep     -0.0027944047 -0.0090644893  0.0183194381  0.016946882  0.0217602256  0.020859314
mobile_wt  -0.0075411165 -0.0425200594 -0.0235133680 -0.0141069129 -0.0017341884  0.018352697
n_cores   -0.0159345009  0.0273115582  0.0032828200 -0.0122469187 -0.0031290360  0.020827702
pc       0.0019685860  -0.028467604 -0.0254274128  0.0474691494  0.0739364180  0.659338498
px_height -0.0250561723  0.0486468811 -0.0588100983  0.0172774418  0.0068419402 -0.017982400
px_width  -0.0121377560  0.0533655937 -0.0320535827  0.0705848351  0.0156095870  0.030549905
ram      -0.0434421532 -0.0232656392  0.0575699860 -0.006496637  0.0481710716 -0.0510996670
sc_h     -0.0119719018 -0.0556652209  0.0127796584 -0.0395030496  0.0062952549  0.045157825
sc_w     -0.0029181267 -0.0239047627  0.0042230804 -0.0271376160 -0.0020645979  0.006115145
talk_time 0.0308072019  0.0155460124 -0.0319952096 -0.0787968458  0.0043902542 -0.051458197
three_g   0.0495709892  0.015141655  0.0135298598 -0.0214064414  0.0006895838 -0.011121045
touch_screen 0.0397679064 -0.0101378408 -0.0600307436  0.0618927650  0.0340196691  0.015467061

```

RStudio interface showing R code and its output. The code reads a CSV file named 'test.csv' and prints a correlation matrix for numeric variables.

```

1 # Load dataset
2 data <- read.csv("test.csv")
3
4 # View dataset
5 head(data)
6 str(data)
7
15:26 (Top Level) : 

```

```

> correlation_matrix <- cor(numeric_data)
>
> # Display correlation matrix
> print(correlation_matrix)
      wifi   four_g int_memory m_dep mobile_wt n_cores pc px_height
id  0.0309206471 -0.0140234070 -0.0027944047  0.0037509249 -0.0124163270 -0.0304868400 -0.0121578310  0.006564568
battery_power -0.0425200594  0.003750925 -0.0090644894 -0.0470646578  0.02573115582  0.012846700  0.048646881
blue   -0.0011689937 -0.0214163270  0.0183194381  0.0235133680 -0.0141069129 -0.0017341884  0.018352697
clock_speed -0.0246648811  0.030486840 -0.016946882  0.019685860 -0.0122469187 -0.0031290360  0.020827702
dual_sim  -0.0249068154 -0.0121377560  0.0232656392  0.0575699860 -0.006496637  0.0481710716 -0.0510996670
fc      0.0328320236 -0.006564568  0.020859314  0.0183526973  0.020827702  0.659338498 -0.017982400
four_g   1.0000000000 -0.037488499  0.014806228 -0.0005086964  0.066715739  0.037669019  0.033655247
int_memory -0.0374884991  1.0000000000 -0.004386083 -0.0104473507  0.021601083  0.022682215 -0.009564431
m_dep     -0.0148062278  0.0148062278 -0.0005086964 -0.0419938426  0.010062154  0.012662551  0.062558682
mobile_wt  -0.0005086964 -0.010447351 -0.041993843  1.0000000000 -0.038907544  0.027342597  0.011157044
n_cores   0.0667157390  0.021601083  0.010062154 -0.0389075443  1.0000000000 0.014375511 -0.054432727
pc       0.0376690186  0.022682215  0.012662551  0.0273425965  0.014375511  1.0000000000  0.028909812
px_height  0.0336552474 -0.009564431  0.062558682  0.0111570445 -0.054432727  0.028909812  1.0000000000
px_width  0.0365445743 -0.003876535  0.034861055 -0.0145768285 -0.059388199  0.056397187  0.517649808
ram      0.0308209628 -0.007107274  0.018348503  0.0287859346 -0.042750444 -0.045987256  0.027945304
sc_h     -0.0150869147 -0.0092492935 -0.026159507 -0.0220521896 -0.034057185  0.019737958  0.011161867
sc_w     -0.0008931884  0.024521033 -0.023392943  0.0221475038 -0.012830091 -0.006315964  0.043486313
talk_time 0.0136919001  0.023759083  0.024124410 -0.0217044320 -0.005640231  0.038434209  0.052382641
three_g   0.5535262792 -0.015922391 -0.029278273  0.0062618805  0.050935894 -0.015220382 -0.011124604
touch_screen -0.0100033817  0.022185887  0.040253625  0.0445253109 -0.016615759  0.021995877 -0.019636564
wifi   -0.0356515451  0.01859684 -0.039705365  0.0697616072 -0.007255598 -0.054955025 -0.012459113
id     -0.012137756 -0.0434421522 -0.011971902  0.0029181267  0.030807202  0.0495709892  0.03976791

```

MVLU COLLEGE

Subject:-Data Analysis with SAS / SPSS /R

RStudio interface showing R code in the script pane and data frames in the environment browser.

```

1 # Load dataset
2 data <- read.csv("test.csv")
3
4 # View dataset
5 head(data)
6 str(data)

```

Console output (R 4.5.2) showing correlation matrix:

```

battery_power 0.053365398 -0.0323656392 -0.055665221 -0.023904782 0.015546012 0.0315141655 -0.01013784
blue -0.032053583 0.0575699860 0.012779658 0.0042230804 -0.031995210 0.0135298590 -0.06003074
clock_speed -0.070584835 -0.0064966367 -0.0271376160 -0.078796846 -0.0214064414 0.06189276
dual_sim 0.015609587 0.0481710716 0.006295255 -0.0020644597 0.004390254 0.0006895838 0.03401967
fc 0.030549905 -0.0519966705 0.045157825 0.0061151445 -0.051458197 -0.0111210450 0.01546706
four_g 0.036544574 0.0308209628 -0.015086915 -0.0008931884 0.013691908 0.553282792 -0.01000338
int_memory -0.003876535 -0.0071072737 -0.009249235 0.0245210335 0.023759083 -0.0159223913 0.02218589
m_dep 0.034861058 0.0183485033 -0.026159507 -0.0233929427 0.024124410 -0.0297282727 0.04025363
mobile_wt -0.014576829 0.0287859346 -0.020253190 0.02217475038 -0.021704432 0.0062618805 0.04452531
n_cores -0.059388199 -0.0427504441 -0.034057185 0.0128300914 -0.005640231 0.0509358945 -0.01661558
pc 0.056397187 -0.0459872564 0.019737958 -0.0063159642 -0.038434209 -0.0152203818 0.02199588
px_height 0.517649808 0.0279453038 0.011161867 0.0434863135 0.052382641 -0.0111246035 -0.01963656
px_width 1.000000000 -0.0263782762 -0.022610052 0.0046921000 0.053422954 0.0194385698 -0.03998625
ram -0.026378276 1.0000000000 0.022894408 0.0306779596 -0.003418840 0.0297124608 0.04365416
sc_h -0.022610052 0.0228944079 1.000000000 0.497154333 0.026061586 -0.0195283321 -0.01366232
sc_w 0.004692100 0.0306779596 0.497154331 1.000000000 0.063990199 0.0011559393 -0.04766466
talk_time 0.053422950 -0.034188404 0.026061584 0.0639901986 1.000000000 -0.0018050743 0.02893599
three_g 0.019438570 0.0297124608 -0.019528332 0.0011559393 -0.001805074 1.000000000 0.00000000
touch_screen -0.039986247 -0.0436541613 -0.013662322 -0.0476646559 0.028935995 0.0000000000 1.00000000
wifi -0.073996910 -0.0319042189 -0.002994426 0.0154750685 0.016709900 -0.0246454405 -0.02600255

```

Environment browser showing various datasets:

- HOUSING: 54 obs. or 15 variables
- indian_wheat: 74 obs. of 25 variables
- insurance: 1338 obs. of 7 variables
- large_sales: 393 obs. of 14 variables
- megagymdata: 2918 obs. of 9 variables
- mental: 101 obs. of 11 variables
- my_data: 2500 obs. of 10 variables
- numeric_data: 1000 obs. of 21 variables

RStudio interface showing R code in the script pane and data frames in the environment browser.

```

1 # Load dataset
2 data <- read.csv("test.csv")
3
4 # View dataset
5 head(data)
6 str(data)

```

Console output (R 4.5.2) showing correlation matrix:

```

touch_screen -0.039986247 -0.0436541613 -0.013662322 -0.0476646559 0.028935995 0.0000000000 1.00000000
wifi -0.073996910 -0.0319042189 -0.002994426 0.0154750685 0.016709900 -0.0246454405 -0.02600255
id -0.0366434123
battery_power -0.000414844
blue 0.0255675984
clock_speed -0.0485931165
dual_sim 0.0315453283
fc -0.0603727369
four_g -0.0356515451
int_memory 0.0118596843
m_dep -0.0397053653
mobile_wt 0.0697616072
n_cores -0.0072555976
pc -0.0549550254
px_height -0.0124591129
px_width -0.0739969103
ram -0.0319042189
sc_h -0.0029944261
sc_w 0.0154750685
talk_time 0.0167098998
three_g -0.0246454405
touch_screen -0.0260025484
wifi 1.0000000000

```

Environment browser showing various datasets:

- HOUSING: 54 obs. or 15 variables
- indian_wheat: 74 obs. of 25 variables
- insurance: 1338 obs. of 7 variables
- large_sales: 393 obs. of 14 variables
- megagymdata: 2918 obs. of 9 variables
- mental: 101 obs. of 11 variables
- my_data: 2500 obs. of 10 variables
- numeric_data: 1000 obs. of 21 variables