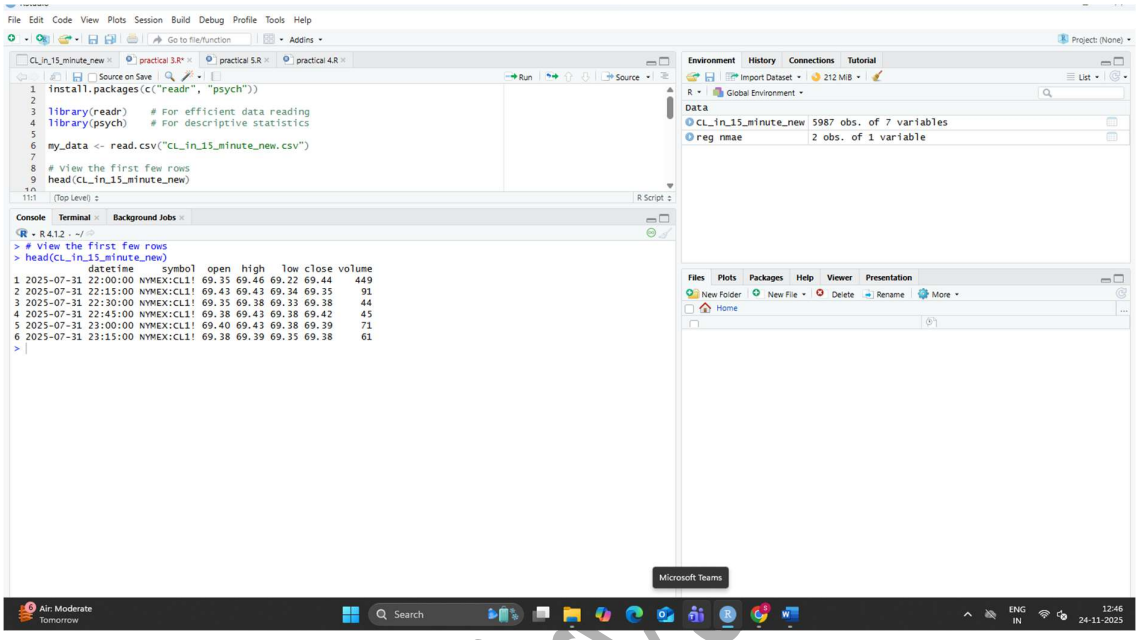
SHETH L.U.J AND SIR M.V. COLLEGE

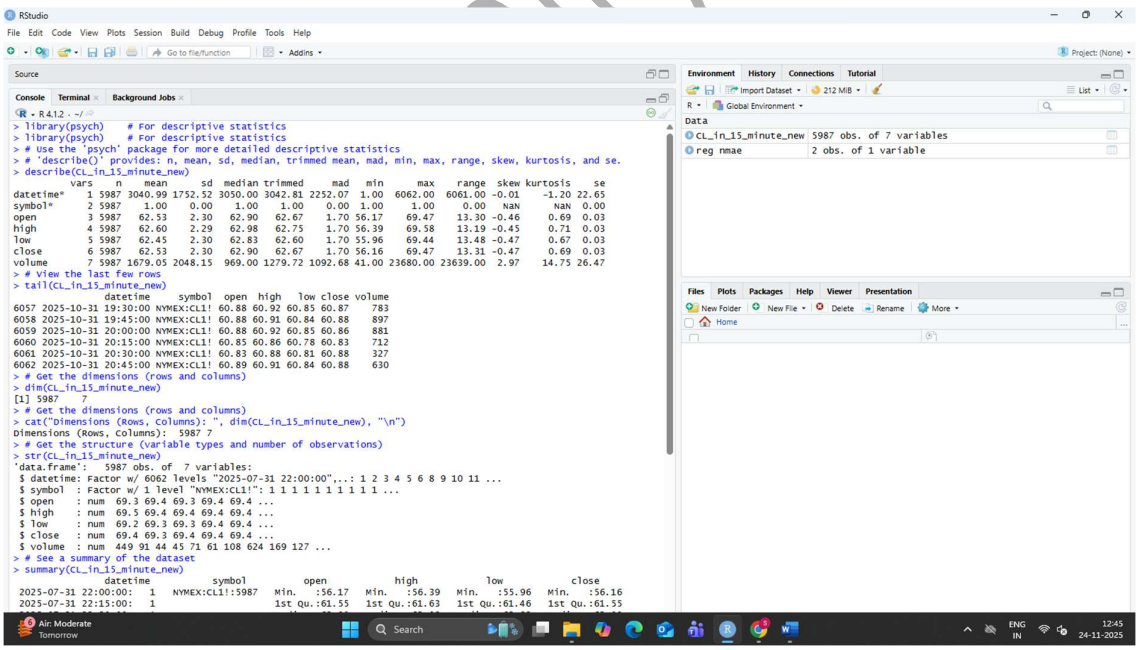
SUBJECT NAME: **Data Analysis with SAS / SPSS /R**

Practical No. 3

Aim: Exploring data:

Output-

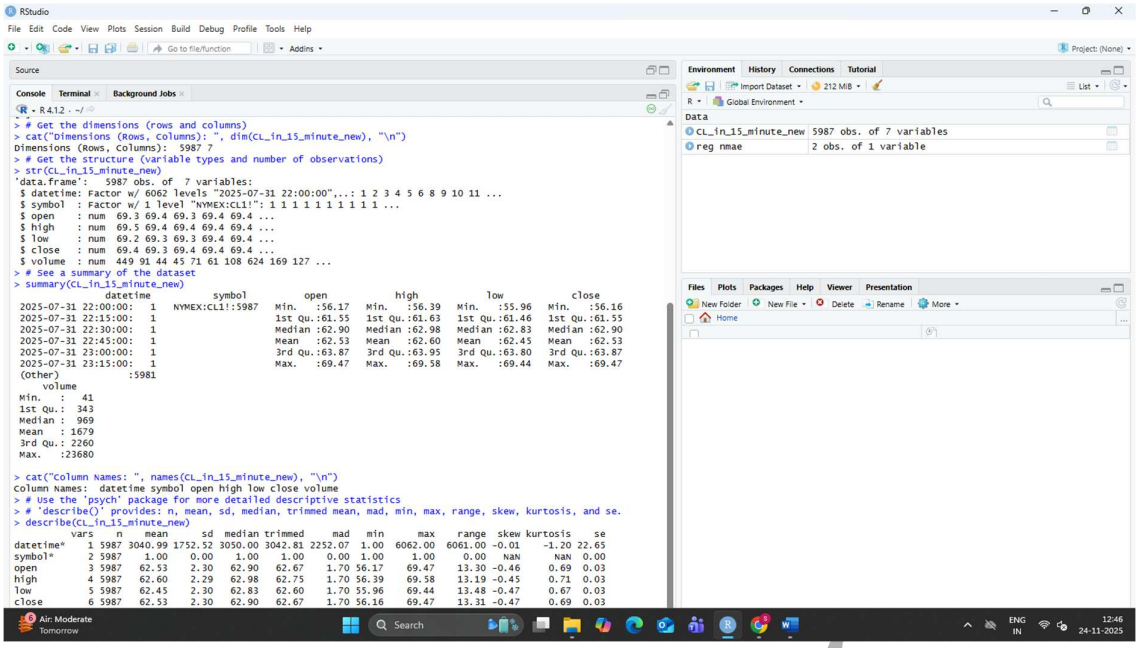




SHETH L.U.J AND SIR M.V. COLLEGE

SUBJECT NAME: **Data Analysis with SAS / SPSS /R**

Practical No. 3



```
> # Get the dimensions (rows and columns)
> cat("Dimensions (Rows, Columns): ", dim(CL_in_15_minute_new), "\n")
Dimensions (Rows, Columns):  5987 7
> # Get the structure (variable types and number of observations)
> str(CL_in_15_minute_new)
'data.frame':   5987 obs. of  7 variables:
 $ datetime: Factor w/ 6062 levels "2025-07-31 22:00:00",..: 1 2 3 4 5 6 8 9 10 11 ...
 $ symbol  : Factor w/ 1 level "NYMEX:CL1!": 1 1 1 1 1 1 1 1 1 1 ...
 $ open    : num  69.3 69.4 69.3 69.4 69.4 ...
 $ high    : num  69.5 69.4 69.4 69.4 69.4 ...
 $ low     : num  69.2 69.3 69.3 69.4 69.4 ...
 $ close   : num  69.4 69.3 69.4 69.4 69.4 ...
 $ volume  : num  449 91 44 45 71 61 108 624 169 127 ...
> # See a summary of the dataset
> summary(CL_in_15_minute_new)
              datetime          symbol          open            high            low             close
 2025-07-31 22:00:00:   1   NYMEX:CL1!:5987   Min.   :56.17   Min.   :56.39   Min.   :55.96   Min.   :56.16
 2025-07-31 22:15:00:   1                     1st Qu.:61.55   1st Qu.:61.63   1st Qu.:61.46   1st Qu.:61.55
 2025-07-31 22:30:00:   1                     Median :62.90   Median :62.98   Median :62.83   Median :62.90
 2025-07-31 22:45:00:   1                     Mean   :62.53   Mean   :62.60   Mean   :62.45   Mean   :62.53
 2025-07-31 23:00:00:   1                     3rd Qu.:63.87   3rd Qu.:63.95   3rd Qu.:63.80   3rd Qu.:63.87
 2025-07-31 23:15:00:   1                     Max.   :69.47   Max.   :69.58   Max.   :69.44   Max.   :69.47
 (Other)            :5981
     volume
 Min.   :   41
 1st Qu.:  343
 Median :  969
 Mean   : 1679
 3rd Qu.: 2260
 Max.   :23680

> cat("Column Names: ", names(CL_in_15_minute_new), "\n")
Column Names:  datetime symbol open high low close volume
> # Use the 'psych' package for more detailed descriptive statistics
> # 'describe()' provides: n, mean, sd, median, trimmed mean, mad, min, max, range, skew, kurtosis, and se.
> describe(CL_in_15_minute_new)
          vars    n    mean    sd  median trimmed     mad   min      max    range  skew kurtosis   se
datetime*    1 5987 3040.99 1752.52 3050.00 3042.81 2252.07  1.00  6062.00 6061.00 -0.01    -1.20 22.65
symbol*      2 5987    1.00    0.00    1.00    1.00    0.00  1.00     1.00    0.00   NaN      NaN  0.00
open         3 5987   62.53    2.30   62.90   62.67    1.70 56.17    69.47   13.30 -0.46     0.69  0.03
high         4 5987   62.60    2.29   62.98   62.75    1.70 56.39    69.58   13.19 -0.45     0.71  0.03
low          5 5987   62.45    2.30   62.83   62.60    1.70 55.96    69.44   13.48 -0.47     0.67  0.03
close        6 5987   62.53    2.30   62.90   62.67    1.70 56.16    69.47   13.31 -0.47     0.69  0.03
```