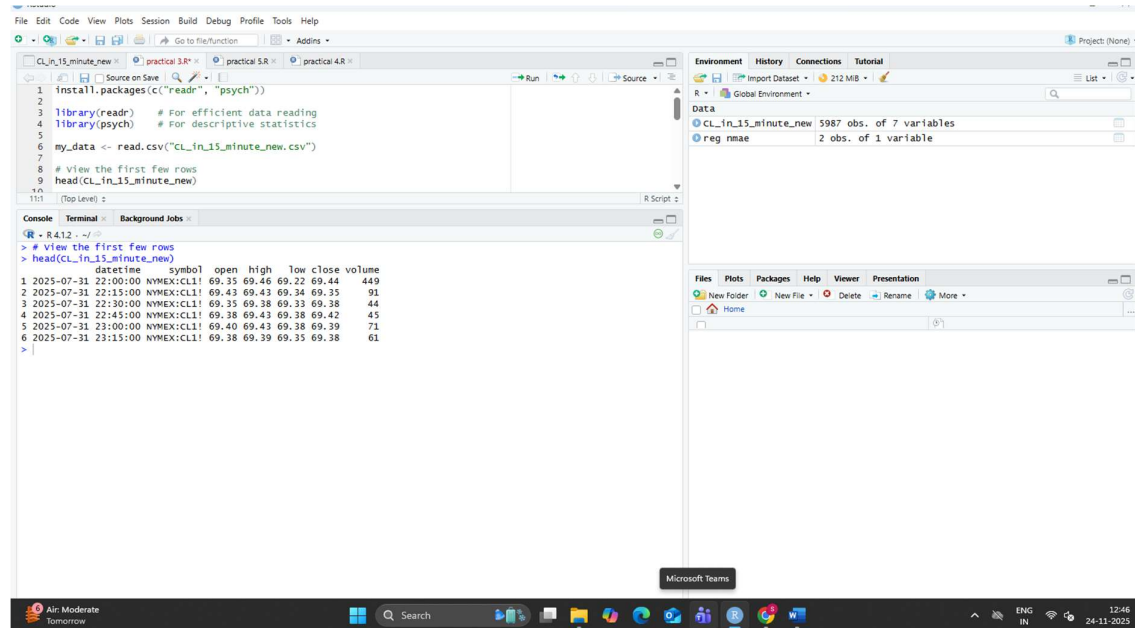


SHETH L.U.J AND SIR M.V. COLLEGE
SUBJECT NAME: Data Analysis with SAS / SPSS / R
Practical No. 3

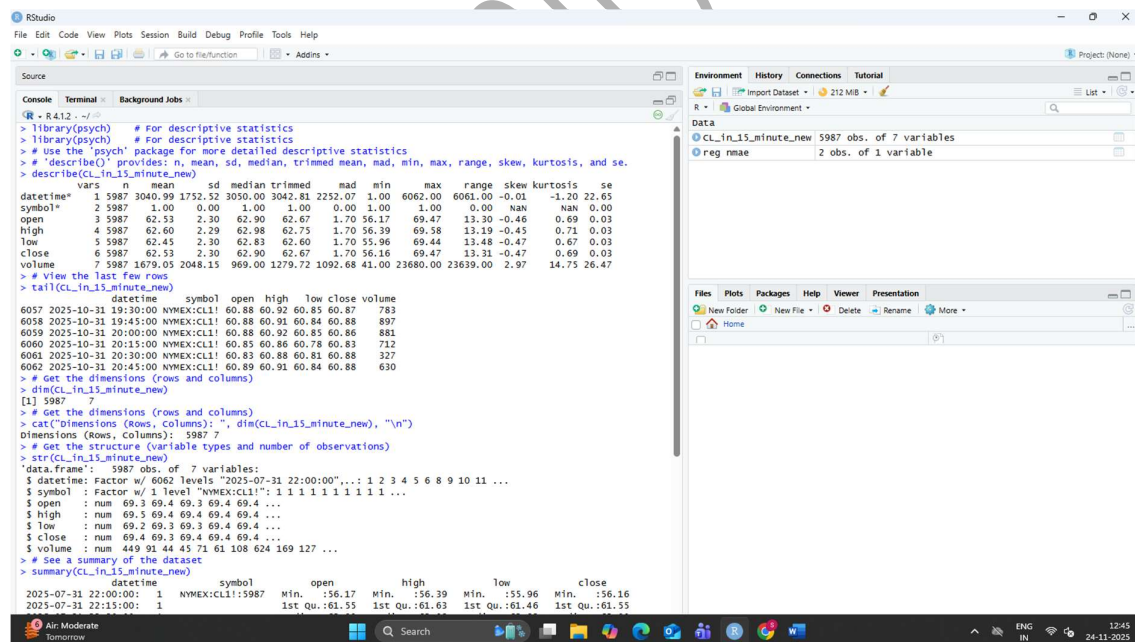
Aim: Exploring data:

Output-



```
1 install.packages(c("readr", "psych"))
2
3 library(readr) # For efficient data reading
4 library(psych) # For descriptive statistics
5
6 my_data <- read_csv("CL_in_15_minute_new.csv")
7
8 # View the first few rows
9 head(CL_in_15_minute_new)
```

```
R - R412.~ /~
> # View the first few rows
> head(CL_in_15_minute_new)
  datetime      symbol  open  high  low close volume
1 2025-07-31 22:00:00 NYMEX:CL1 69.35 69.46 69.22 69.44 449
2 2025-07-31 22:15:00 NYMEX:CL1 69.43 69.43 69.34 69.35 91
3 2025-07-31 22:30:00 NYMEX:CL1 69.35 69.38 69.33 69.38 44
4 2025-07-31 22:45:00 NYMEX:CL1 69.38 69.43 69.38 69.42 45
5 2025-07-31 23:00:00 NYMEX:CL1 69.40 69.43 69.38 69.39 71
6 2025-07-31 23:15:00 NYMEX:CL1 69.38 69.39 69.35 69.38 61
```



```
> library(psych) # For descriptive statistics
> library(psych) # For descriptive statistics
> # 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	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
volume	7	5987	1679.05	2048.15	969.00	1279.72	1092.68	41.00	23680.00	23639.00	2.97	14.75	26.47

```
> # View the last few rows
> tail(CL_in_15_minute_new)
```

datetime	symbol	open	high	low	close	volume
6057 2025-10-31 19:30:00	NYMEX:CL1	60.88	60.92	60.85	60.87	783
6058 2025-10-31 19:45:00	NYMEX:CL1	60.88	60.91	60.84	60.88	897
6059 2025-10-31 20:00:00	NYMEX:CL1	60.88	60.92	60.83	60.86	881
6060 2025-10-31 20:15:00	NYMEX:CL1	60.85	60.86	60.78	60.83	712
6061 2025-10-31 20:30:00	NYMEX:CL1	60.83	60.88	60.81	60.88	327
6062 2025-10-31 20:45:00	NYMEX:CL1	60.89	60.91	60.84	60.88	630

```
> # Get the dimensions (rows and columns)
> dim(CL_in_15_minute_new)
```

```
[1] 5987 7
```

```
> # get the dimensions (rows and columns)
> cat("Dimensions (Rows, Columns): ", dim(CL_in_15_minute_new), "\n")
```

```
Dimensions (Rows, Columns): 5987 7
```

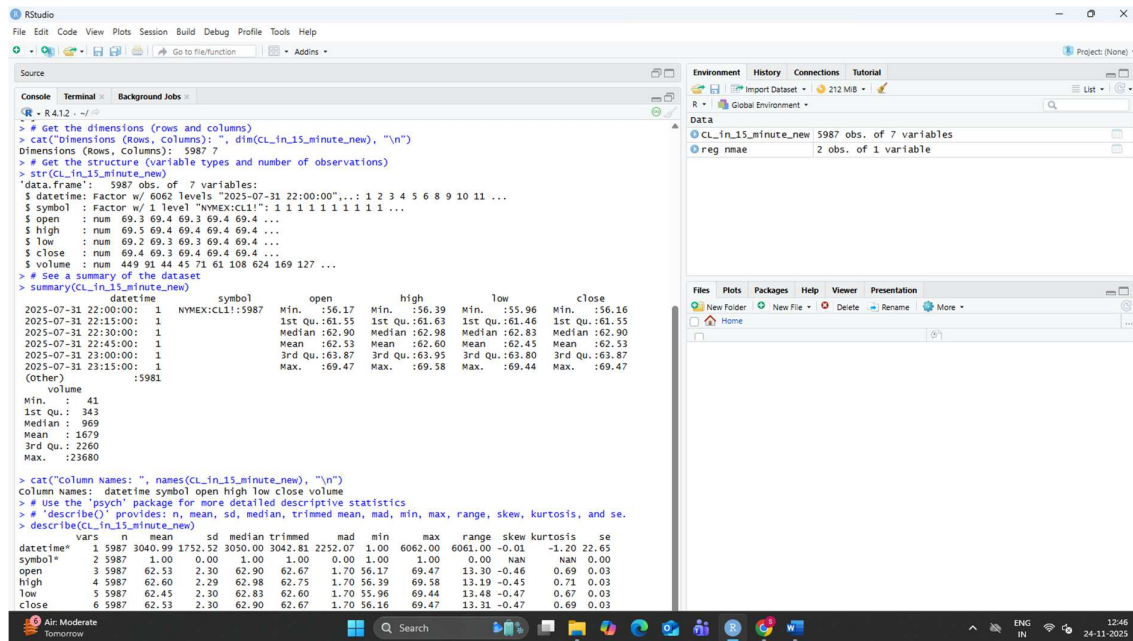
```
> # get the structure (variable types and number of observations)
> str(CL_in_15_minute_new)
```

```
'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

SHETH L.U.J AND SIR M.V. COLLEGE
SUBJECT NAME: **Data Analysis with SAS / SPSS / R**
Practical No. 3



The screenshot displays the RStudio environment with the following components:

- Source Editor:** Contains R code for loading and summarizing a dataset.
- Console:** Shows the output of the R commands.
- Environment:** Lists the objects in the global environment.
- Files:** Shows the file explorer.

R Code:

```
> # 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:CL11": 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)
```

Console Output:

```
summary(CL_in_15_minute_new)
  datetime      symbol      open      high      low      close
2025-07-31 22:00:00: 1 NYMEX:CL11:5987 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
  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)
```

Environment:

- CL_in_15_minute_new: 5987 obs. of 7 variables
- reg mae: 2 obs. of 1 variable