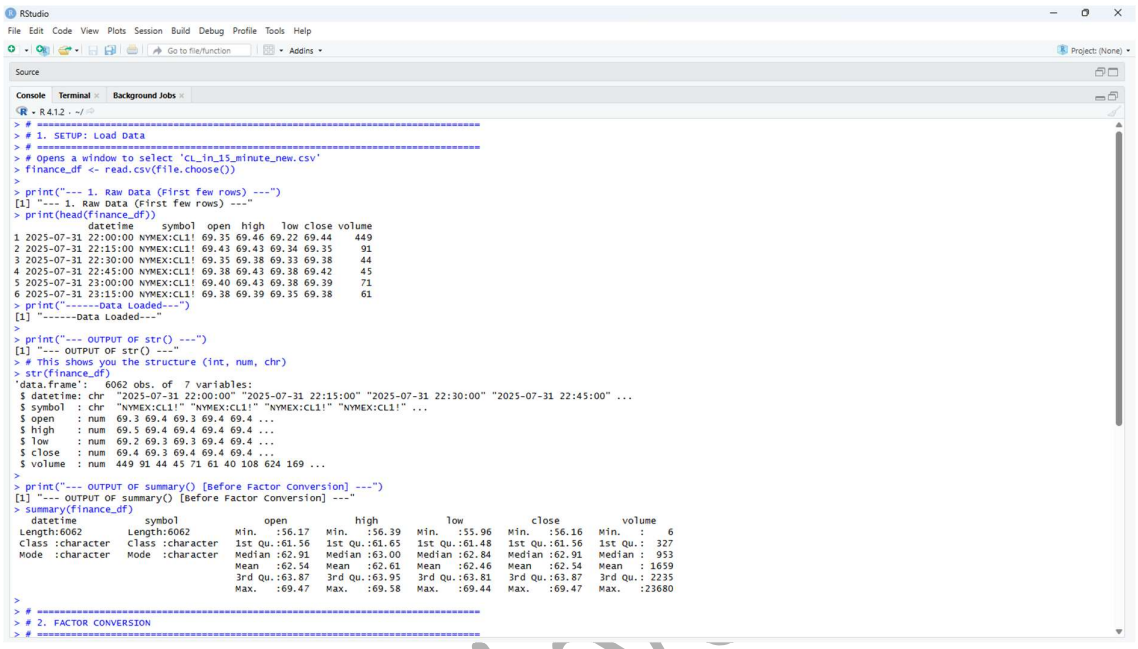
SHETH L.U.J AND SIR M.V. COLLEGE

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

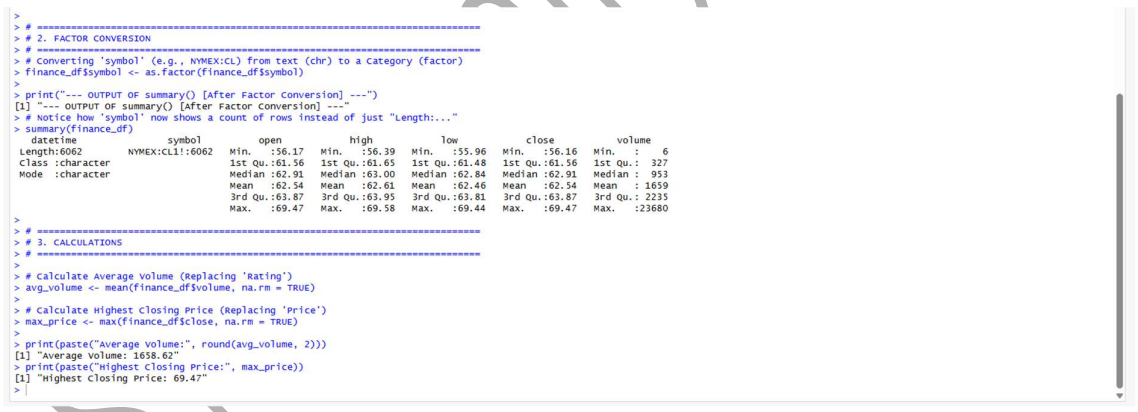
Practical No. 15

Aim: Generating basic summaries using str() or summary () (R).

Output:



```
> # ================================================================================
> # 1. SETUP: Load Data
> # ================================================================================
> # Opens a window to select 'CL_in_15_minute_new.csv'
> finance_df <- read.csv(file.choose())
>
> print("--- 1. Raw Data (First few rows) ---")
[1] "--- 1. Raw Data (First few rows) ---"
> print(head(finance_df))
             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
> print("------Data Loaded---")
[1] "------Data Loaded---"
>
> print("--- OUTPUT OF str() ---")
[1] "--- OUTPUT OF str() ---"
> # This shows you the structure (int, num, chr)
> str(finance_df)
'data.frame':   6062 obs. of  7 variables:
 $ datetime: chr  "2025-07-31 22:00:00" "2025-07-31 22:15:00" "2025-07-31 22:30:00" "2025-07-31 22:45:00" ...
 $ symbol  : chr  "NYMEX:CL1!" "NYMEX:CL1!" "NYMEX:CL1!" "NYMEX:CL1!" ...
 $ 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 40 108 624 169 ...
>
> print("--- OUTPUT OF summary() [Before Factor Conversion] ---")
[1] "--- OUTPUT OF summary() [Before Factor Conversion] ---"
> summary(finance_df)
   datetime            symbol              open            high            low             close           volume
 Length:6062        Length:6062        Min.   :56.17   Min.   :56.39   Min.   :55.96   Min.   :56.16   Min.   :   6
 Class :character   Class :character   1st Qu.:61.56   1st Qu.:61.65   1st Qu.:61.48   1st Qu.:61.56   1st Qu.:  327
 Mode  :character   Mode  :character   Median :62.91   Median :63.00   Median :62.84   Median :62.91   Median :  953
                                       Mean   :62.54   Mean   :62.61   Mean   :62.46   Mean   :62.54   Mean   : 1659
                                       3rd Qu.:63.87   3rd Qu.:63.95   3rd Qu.:63.81   3rd Qu.:63.87   3rd Qu.: 2235
                                       Max.   :69.47   Max.   :69.58   Max.   :69.44   Max.   :69.47   Max.   :23680
>
> # ================================================================================
> # 2. FACTOR CONVERSION
> # ================================================================================
```



```
>
> # ================================================================================
> # 2. FACTOR CONVERSION
> # ================================================================================
> # Converting 'symbol' (e.g., NYMEX:CL) from text (chr) to a Category (factor)
> finance_df$symbol <- as.factor(finance_df$symbol)
>
> print("--- OUTPUT OF summary() [After Factor Conversion] ---")
[1] "--- OUTPUT OF summary() [After Factor Conversion] ---"
> # Notice how 'symbol' now shows a count of rows instead of just "Length:..."
> summary(finance_df)
   datetime                symbol          open            high            low             close           volume
 Length:6062        NYMEX:CL1!:6062   Min.   :56.17   Min.   :56.39   Min.   :55.96   Min.   :56.16   Min.   :   6
 Class :character                     1st Qu.:61.56   1st Qu.:61.65   1st Qu.:61.48   1st Qu.:61.56   1st Qu.:  327
 Mode  :character                     Median :62.91   Median :63.00   Median :62.84   Median :62.91   Median :  953
                                      Mean   :62.54   Mean   :62.61   Mean   :62.46   Mean   :62.54   Mean   : 1659
                                      3rd Qu.:63.87   3rd Qu.:63.95   3rd Qu.:63.81   3rd Qu.:63.87   3rd Qu.: 2235
                                      Max.   :69.47   Max.   :69.58   Max.   :69.44   Max.   :69.47   Max.   :23680
>
> # ================================================================================
> # 3. CALCULATIONS
> # ================================================================================
>
> # Calculate Average Volume (Replacing 'Rating')
> avg_volume <- mean(finance_df$volume, na.rm = TRUE)
>
> # Calculate Highest Closing Price (Replacing 'Price')
> max_price <- max(finance_df$close, na.rm = TRUE)
>
> print(paste("Average Volume:", round(avg_volume, 2)))
[1] "Average Volume: 1658.62"
> print(paste("Highest Closing Price:", max_price))
[1] "Highest Closing Price: 69.47"
>
```