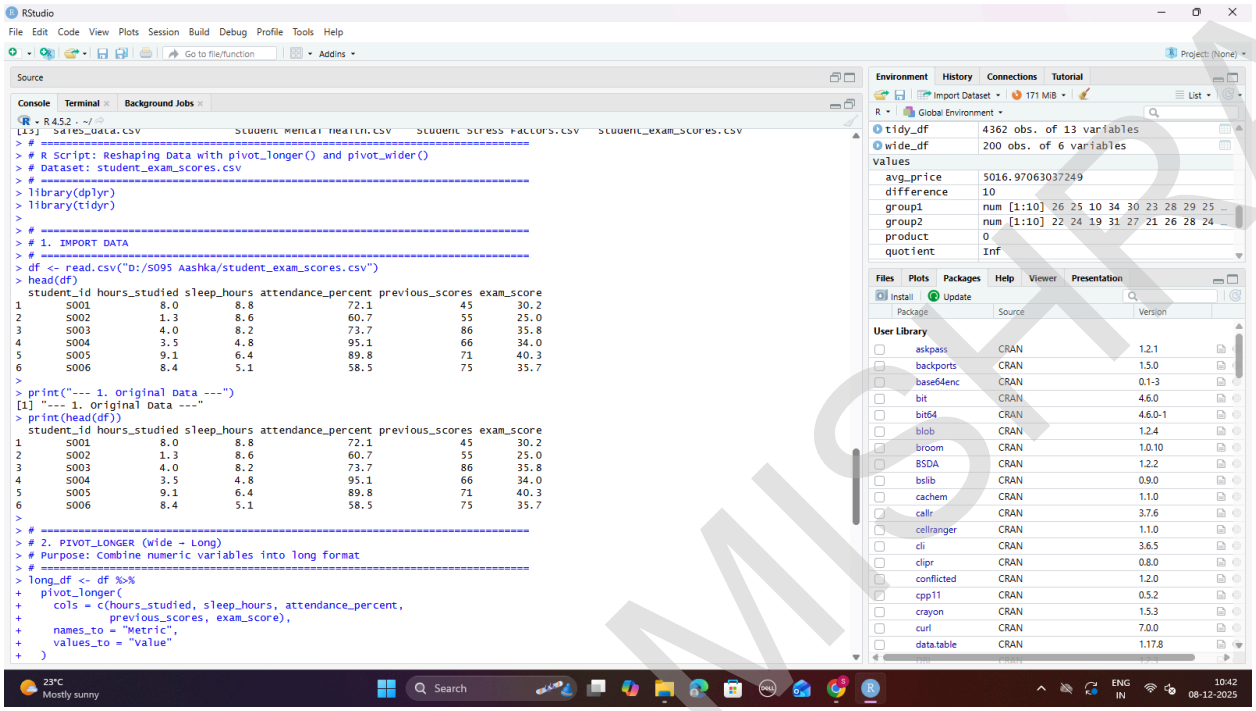
SHETH L.U.J. AND SIR M.V. COLLEGE
SUBJECT: DATA ANALYSIS WITH R

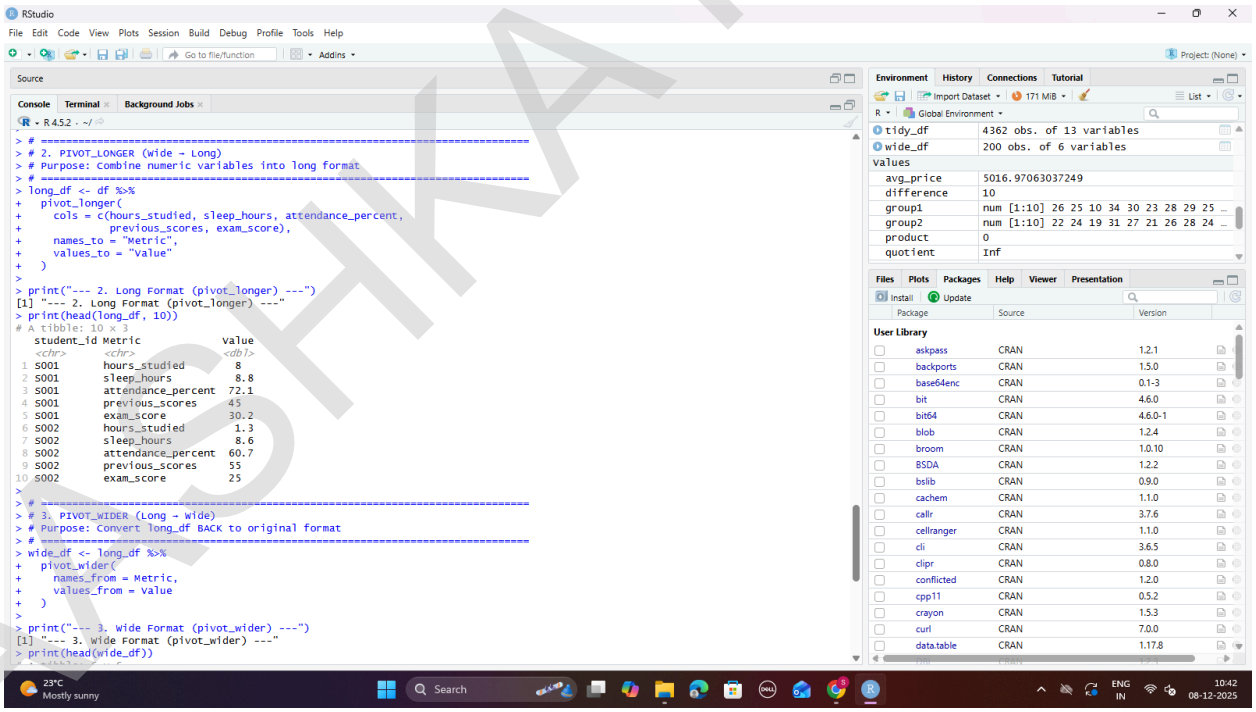
PRACTICAL NO. 11

AIM: Reshaping data using pivot_longer()/pivot_wider() (R).

OUTPUT:



```
> # ================================================================
> # R Script: Reshaping Data with pivot_longer() and pivot_wider()
> # Dataset: student_exam_scores.csv
> # ================================================================
> library(dplyr)
> library(tidyr)
>
> # ================================================================
> # 1. IMPORT DATA
> # ================================================================
> df <- read.csv("D:/S095 Aashka/student_exam_scores.csv")
> head(df)
  student_id hours_studied sleep_hours attendance_percent previous_scores exam_score
1       S001           8.0         8.8              72.1              45       30.2
2       S002           1.3         8.6              60.7              55       25.0
3       S003           4.0         8.2              73.7              86       35.8
4       S004           3.5         4.8              95.1              66       34.0
5       S005           9.1         6.4              89.8              71       40.3
6       S006           8.4         5.1              58.5              75       35.7
>
> print("--- 1. Original Data ---")
[1] "--- 1. Original Data ---"
> print(head(df))
  student_id hours_studied sleep_hours attendance_percent previous_scores exam_score
1       S001           8.0         8.8              72.1              45       30.2
2       S002           1.3         8.6              60.7              55       25.0
3       S003           4.0         8.2              73.7              86       35.8
4       S004           3.5         4.8              95.1              66       34.0
5       S005           9.1         6.4              89.8              71       40.3
6       S006           8.4         5.1              58.5              75       35.7
>
> # ================================================================
> # 2. PIVOT_LONGER (Wide → Long)
> # Purpose: Combine numeric variables into long format
> # ================================================================
> long_df <- df %>%
+   pivot_longer(
+     cols = c(hours_studied, sleep_hours, attendance_percent,
+              previous_scores, exam_score),
+     names_to = "Metric",
+     values_to = "Value"
+   )
```

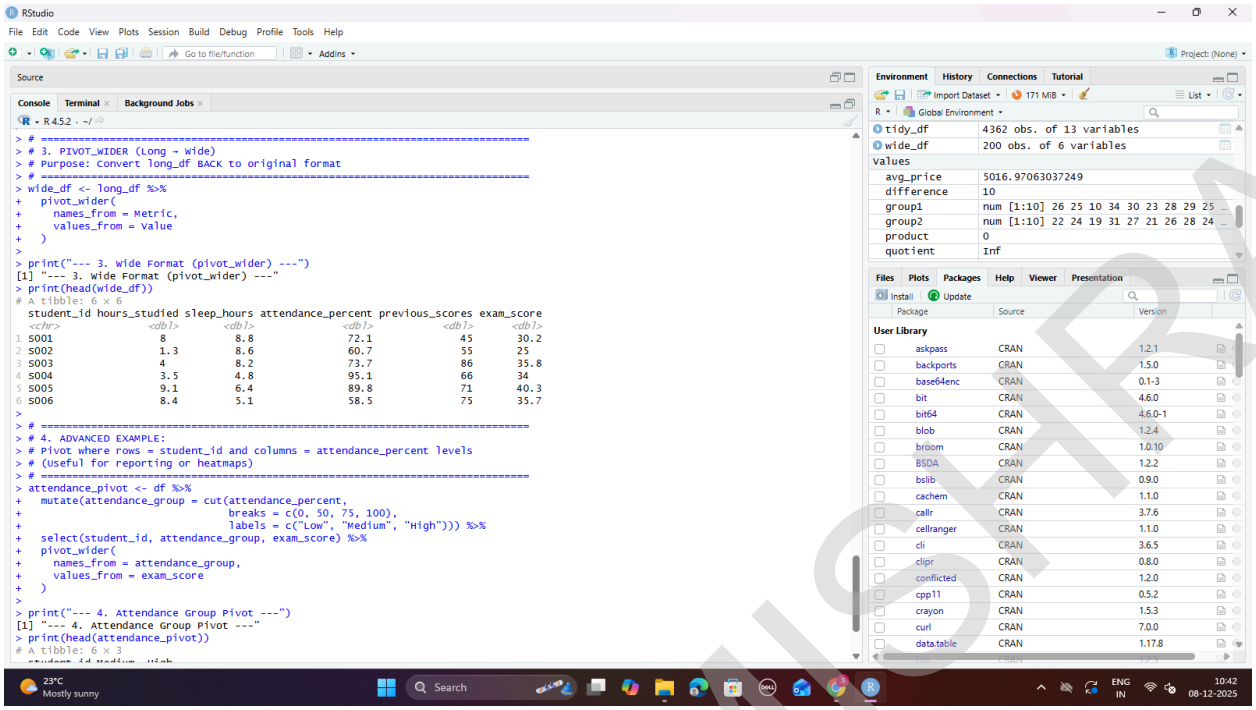


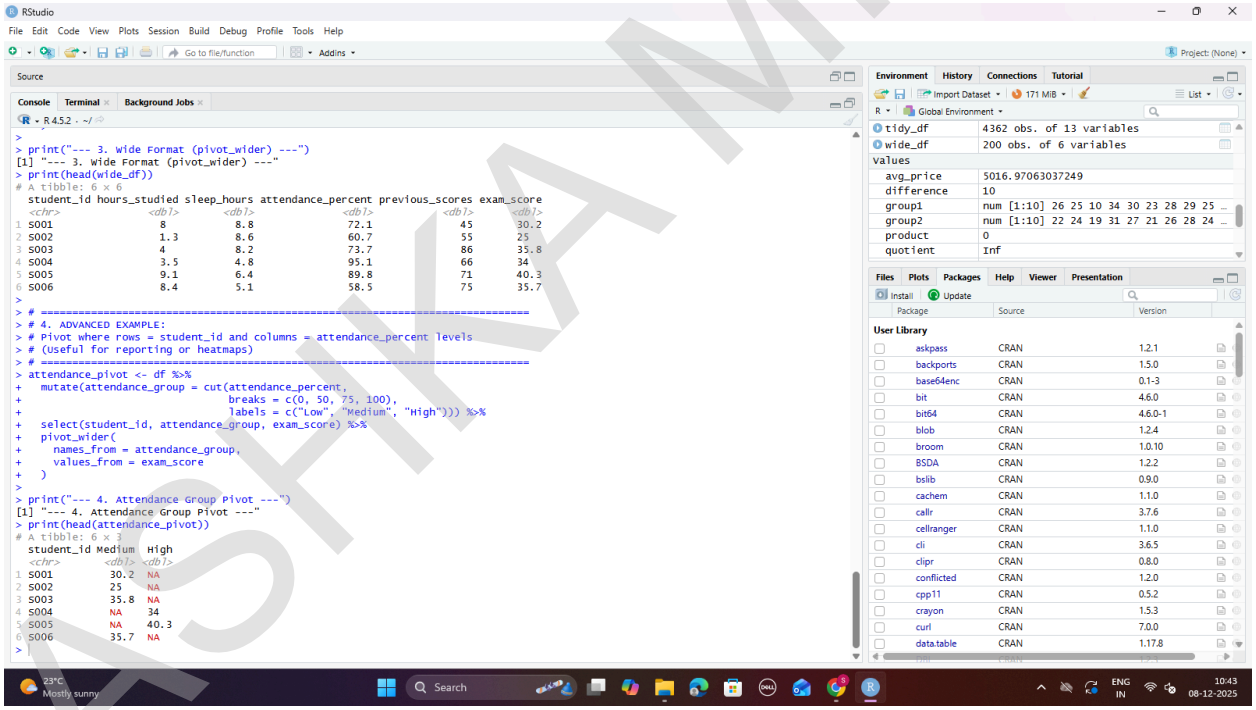
```
> # ================================================================
> # 2. PIVOT_LONGER (Wide → Long)
> # Purpose: Combine numeric variables into long format
> # ================================================================
> long_df <- df %>%
+   pivot_longer(
+     cols = c(hours_studied, sleep_hours, attendance_percent,
+              previous_scores, exam_score),
+     names_to = "Metric",
+     values_to = "Value"
+   )
>
> print("--- 2. Long Format (pivot_longer) ---")
[1] "--- 2. Long Format (pivot_longer) ---"
> print(head(long_df, 10))
# A tibble: 10 × 3
   student_id Metric           Value
   <chr>      <chr>            <dbl>
 1 S001       hours_studied      8
 2 S001       sleep_hours        8.8
 3 S001       attendance_percent 72.1
 4 S001       previous_scores    45
 5 S001       exam_score         30.2
 6 S002       hours_studied      1.3
 7 S002       sleep_hours        8.6
 8 S002       attendance_percent 60.7
 9 S002       previous_scores    55
10 S002       exam_score         25
>
> # ================================================================
> # 3. PIVOT_WIDER (Long → Wide)
> # Purpose: Convert long_df BACK to original format
> # ================================================================
> wide_df <- long_df %>%
+   pivot_wider(
+     names_from = Metric,
+     values_from = Value
+   )
>
> print("--- 3. Wide Format (pivot_wider) ---")
[1] "--- 3. Wide Format (pivot_wider) ---"
> print(head(wide_df))
```

NAME: AASHKA MISHRA

ROLL NO: S095

SHETH L.U.J. AND SIR M.V. COLLEGE
SUBJECT: DATA ANALYSIS WITH R



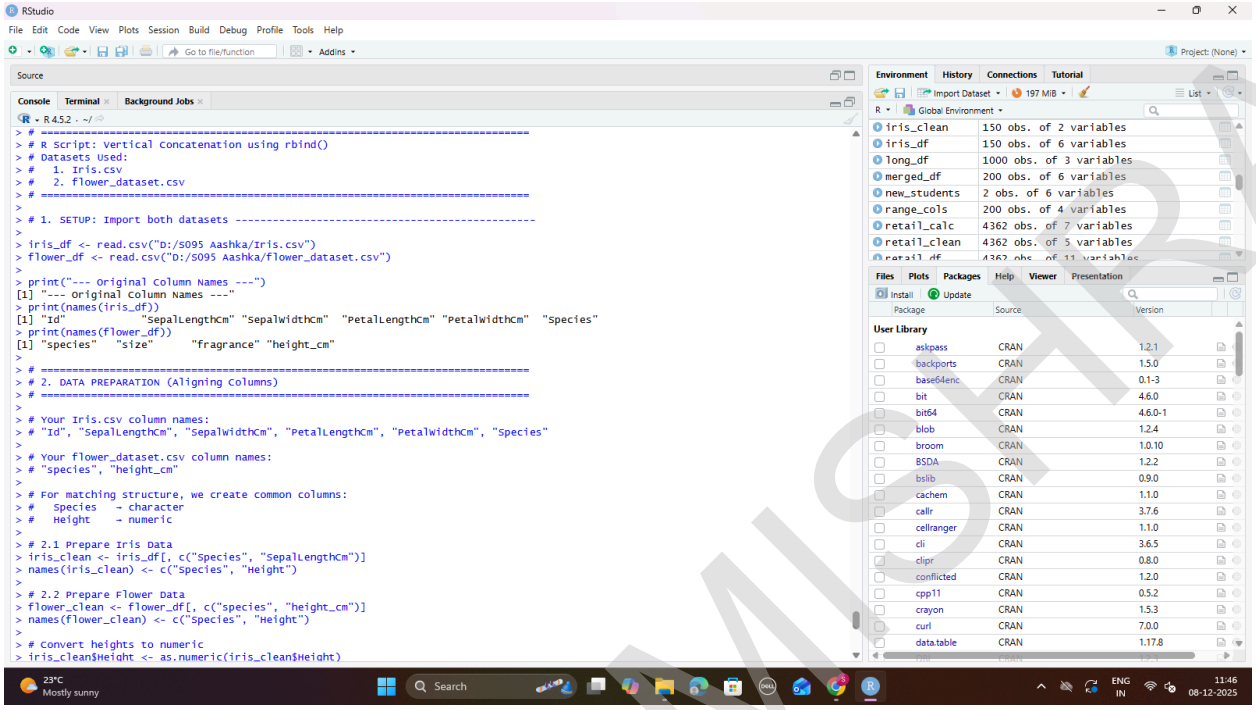


NAME: AASHKA MISHRA
ROLL NO: S095

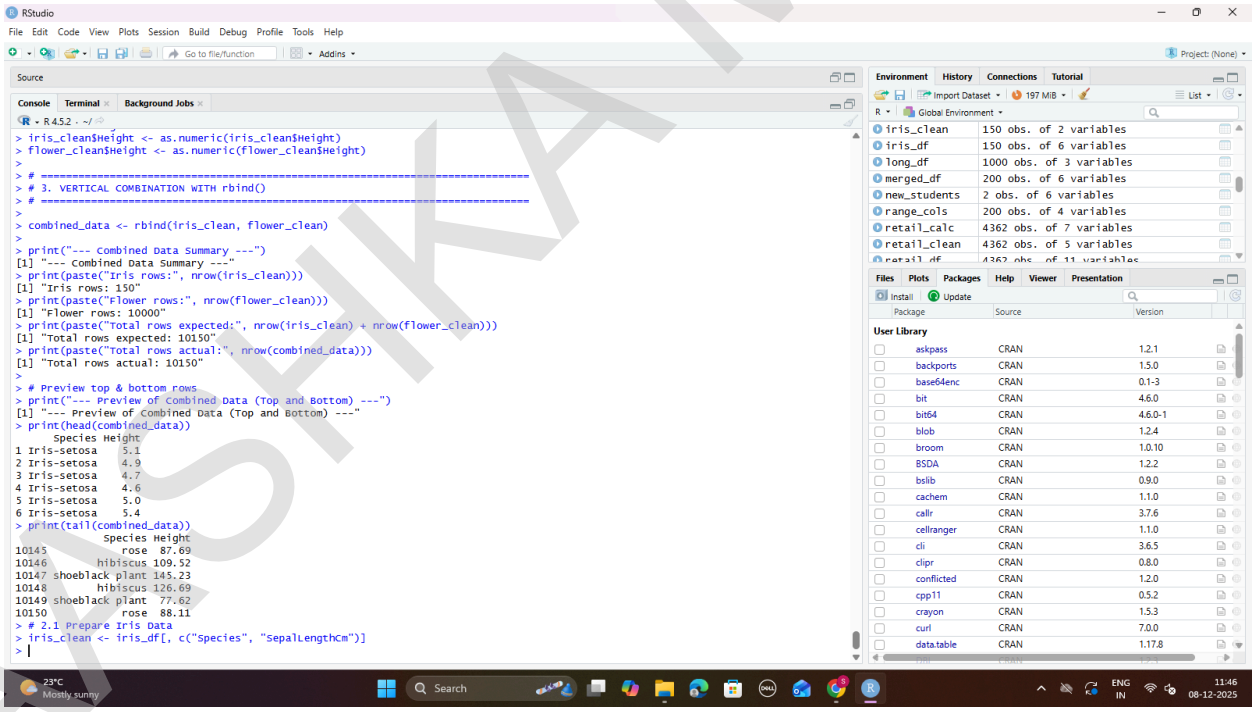
SHETH L.U.J. AND SIR M.V. COLLEGE
SUBJECT: DATA ANALYSIS WITH R

PRACTICAL NO. 12

AIM: Combining datasets vertically (concatenation) using rbind() (R).



```
> # ================================================================================
> # R Script: Vertical Concatenation using rbind()
> # Datasets Used:
> #   1. Iris.csv
> #   2. flower_dataset.csv
> # ================================================================================
>
> # 1. SETUP: Import both datasets ------------------------------------------------
>
> iris_df <- read.csv("D:/S095 Aashka/Iris.csv")
> flower_df <- read.csv("D:/S095 Aashka/flower_dataset.csv")
>
> print("--- Original Column Names ---")
[1] "--- Original Column Names ---"
> print(names(iris_df))
[1] "Id"            "SepalLengthCm" "SepalWidthCm"  "PetalLengthCm" "PetalWidthCm"  "Species"
> print(names(flower_df))
[1] "species"   "size"      "fragrance" "height_cm"
>
> # ================================================================================
> # 2. DATA PREPARATION (Aligning Columns)
> # ================================================================================
>
> # Your Iris.csv column names:
> # "Id", "SepalLengthCm", "SepalWidthCm", "PetalLengthCm", "PetalWidthCm", "Species"
>
> # Your flower_dataset.csv column names:
> # "species", "height_cm"
>
> # For matching structure, we create common columns:
> #   Species   → character
> #   Height    → numeric
>
> # 2.1 Prepare Iris Data
> iris_clean <- iris_df[, c("Species", "SepalLengthCm")]
> names(iris_clean) <- c("Species", "Height")
>
> # 2.2 Prepare Flower Data
> flower_clean <- flower_df[, c("species", "height_cm")]
> names(flower_clean) <- c("Species", "Height")
>
> # Convert heights to numeric
> iris_clean$Height <- as.numeric(iris_clean$Height)
```



```
> iris_clean$Height <- as.numeric(iris_clean$Height)
> flower_clean$Height <- as.numeric(flower_clean$Height)
>
> # ================================================================================
> # 3. VERTICAL COMBINATION WITH rbind()
> # ================================================================================
>
> combined_data <- rbind(iris_clean, flower_clean)
>
> print("--- Combined Data Summary ---")
[1] "--- Combined Data Summary ---"
> print(paste("Iris rows:", nrow(iris_clean)))
[1] "Iris rows: 150"
> print(paste("Flower rows:", nrow(flower_clean)))
[1] "Flower rows: 10000"
> print(paste("Total rows expected:", nrow(iris_clean) + nrow(flower_clean)))
[1] "Total rows expected: 10150"
> print(paste("Total rows actual:", nrow(combined_data)))
[1] "Total rows actual: 10150"
>
> # Preview top & bottom rows
> print("--- Preview of Combined Data (Top and Bottom) ---")
[1] "--- Preview of Combined Data (Top and Bottom) ---"
> print(head(combined_data))
    Species Height
1 Iris-setosa    5.1
2 Iris-setosa    4.9
3 Iris-setosa    4.7
4 Iris-setosa    4.6
5 Iris-setosa    5.0
6 Iris-setosa    5.4
> print(tail(combined_data))
             Species Height
10145           rose  87.69
10146      hibiscus 109.52
10147 shoeblack plant 145.23
10148      hibiscus 126.69
10149 shoeblack plant  77.62
10150           rose  88.11
> # 2.1 Prepare Iris Data
> iris_clean <- iris_df[, c("Species", "SepalLengthCm")]
>
```

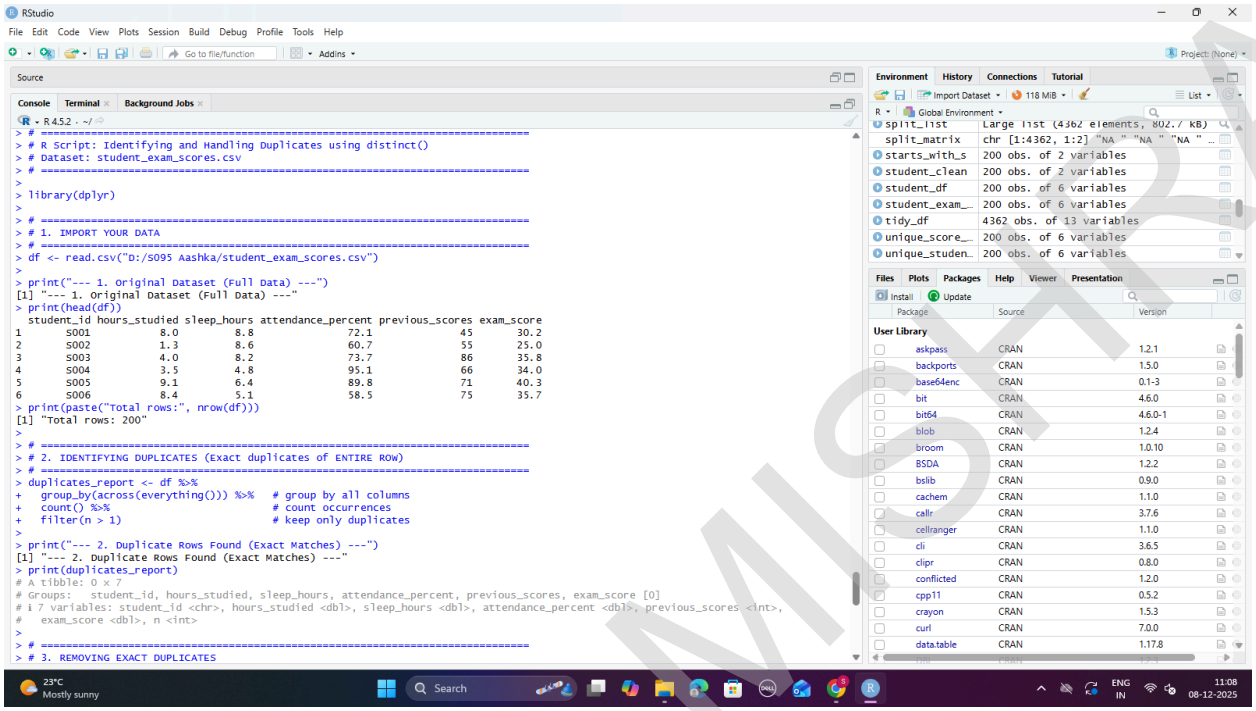

NAME: AASHKA MISHRA
ROLL NO: S095

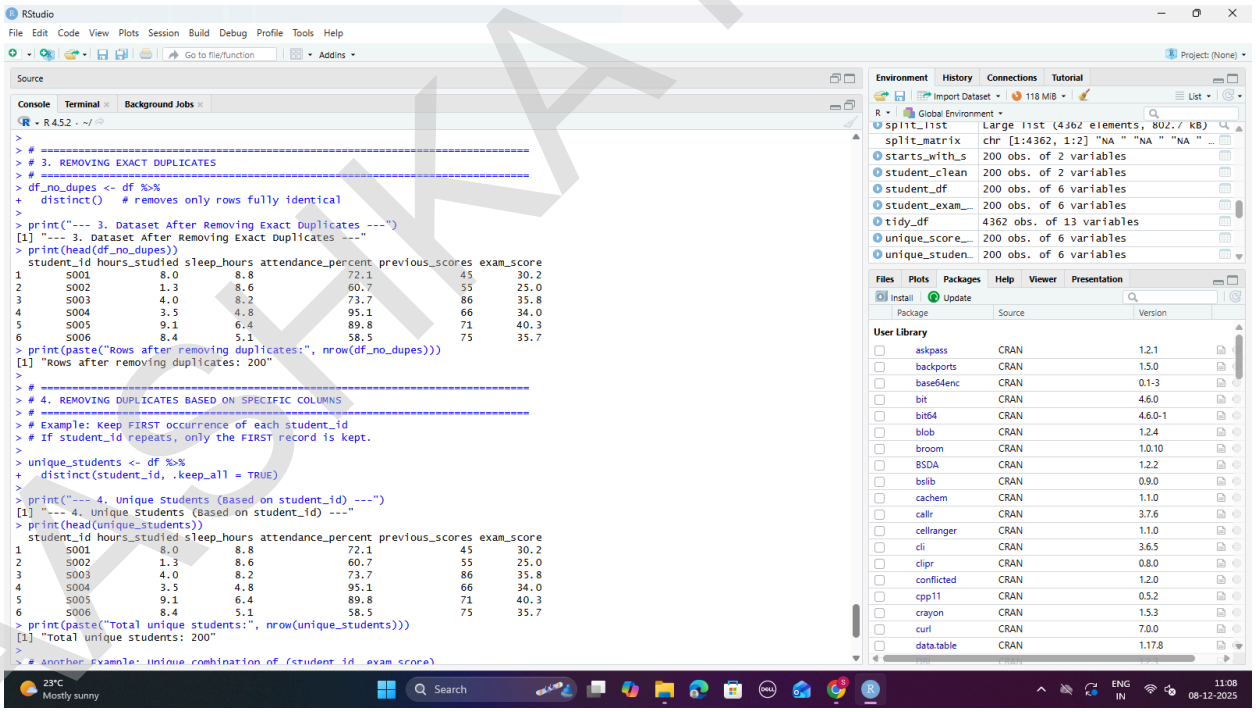
SHETH L.U.J. AND SIR M.V. COLLEGE
SUBJECT: DATA ANALYSIS WITH R

PRACTICAL NO. 13

AIM: Identifying and handling duplicates using distinct() (R).

OUTPUT:

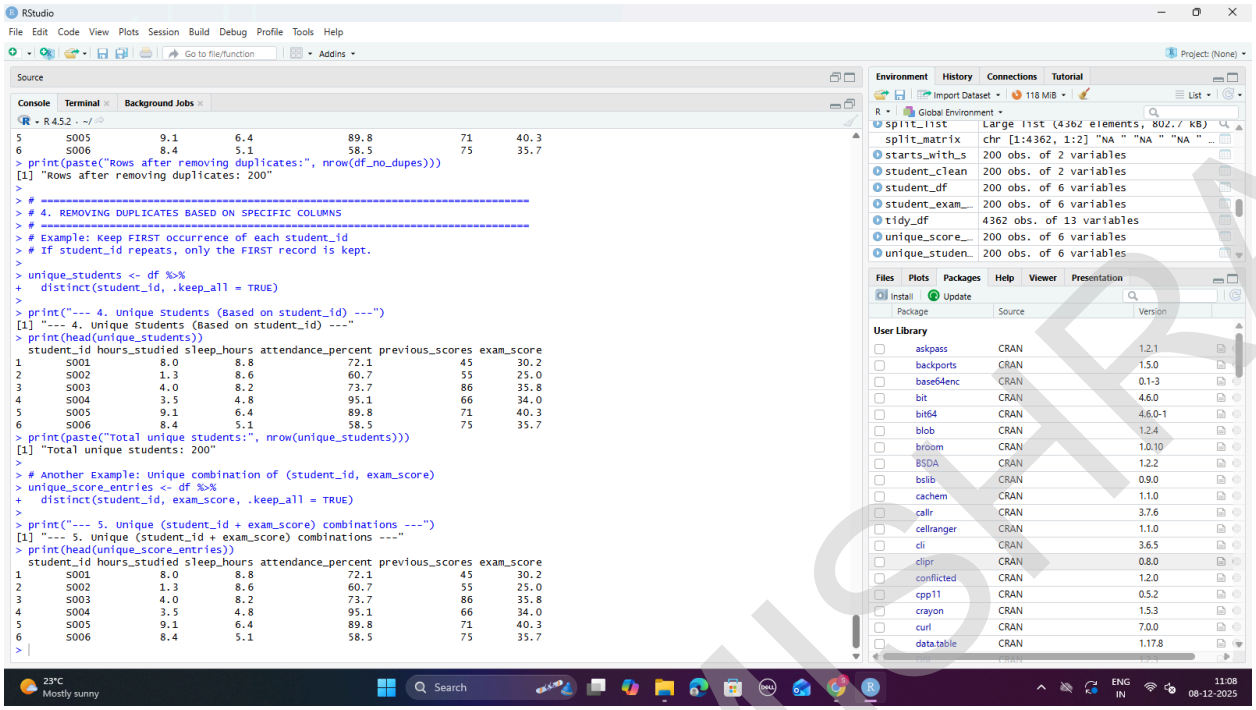




NAME: AASHKA MISHRA

ROLL NO: S095

SHETH L.U.J. AND SIR M.V. COLLEGE
SUBJECT: DATA ANALYSIS WITH R



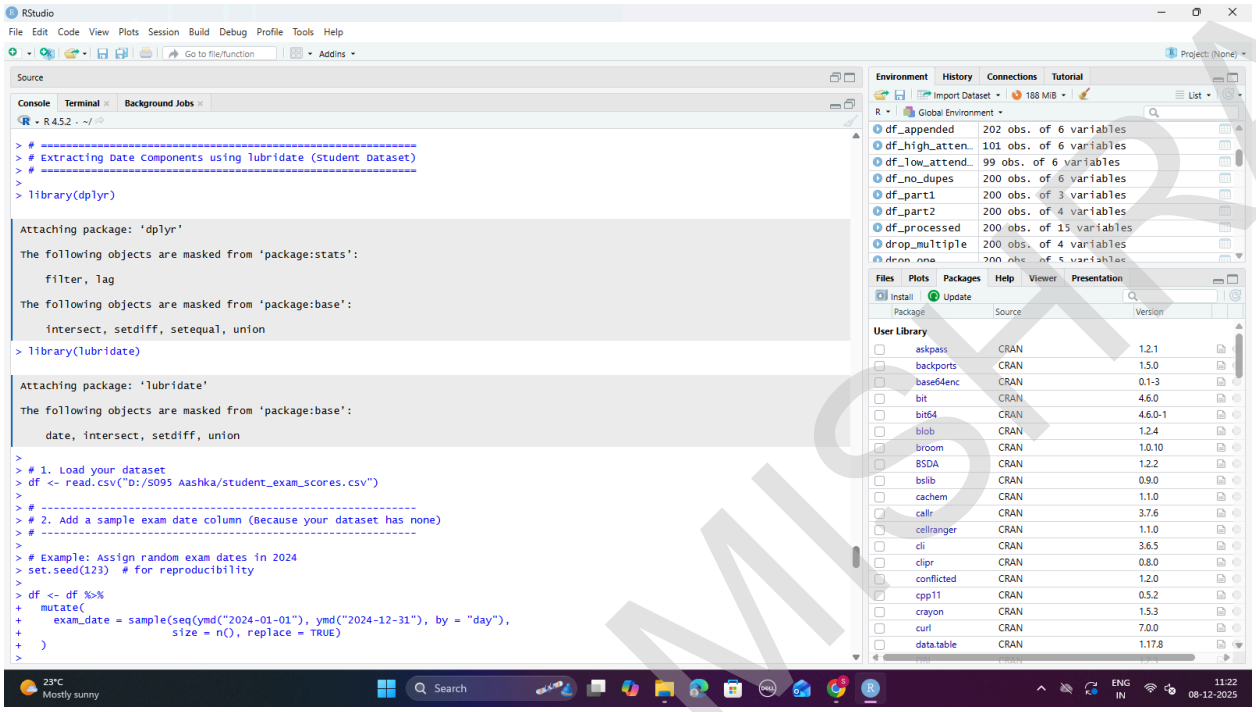
NAME: AASHKA MISHRA
ROLL NO: S095

SHETH L.U.J. AND SIR M.V. COLLEGE
SUBJECT: DATA ANALYSIS WITH R

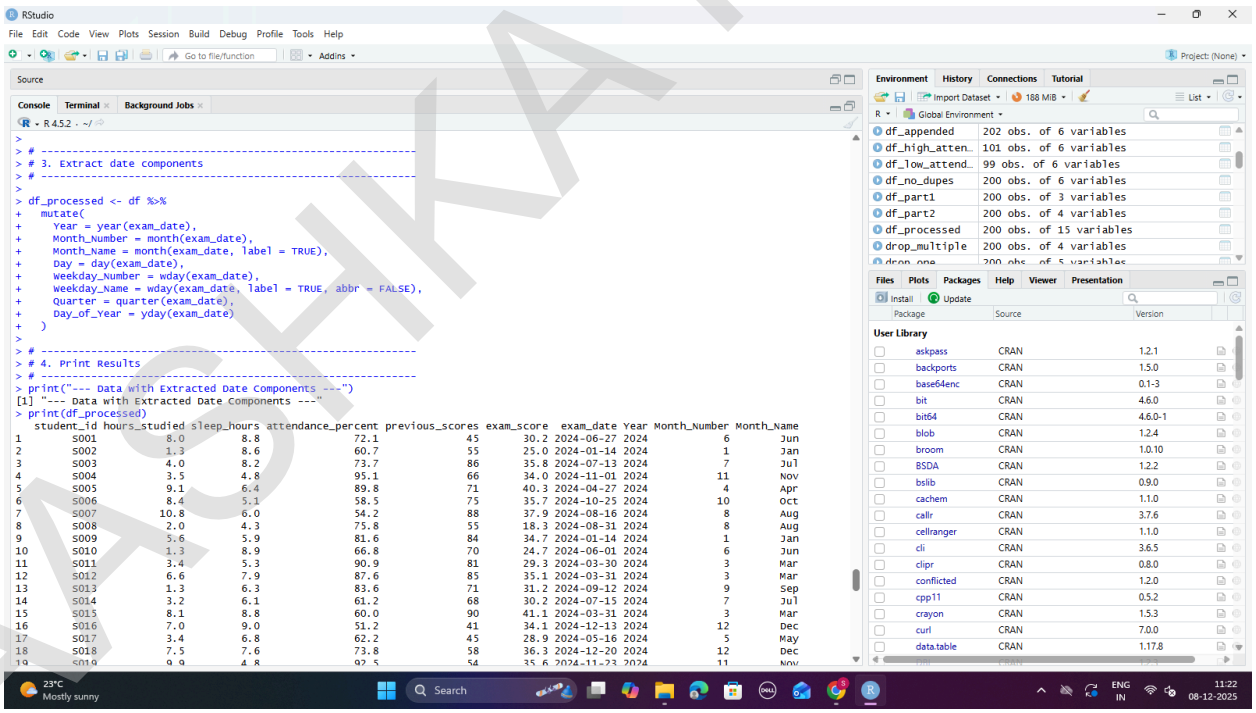
PRACTICAL NO. 14

AIM: Extracting date components using lubridate:: functions (R).

OUTPUT:



```
> # ============================================================
> # Extracting Date Components using lubridate (Student Dataset)
> # ============================================================
>
> library(dplyr)

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union

> library(lubridate)

Attaching package: 'lubridate'

The following objects are masked from 'package:base':

    date, intersect, setdiff, union

>
> # 1. Load your dataset
> df <- read.csv("D:/S095 Aashka/student_exam_scores.csv")
>
> # ------------------------------------------------------------
> # 2. Add a sample exam date column (Because your dataset has none)
> # ------------------------------------------------------------
>
> # Example: Assign random exam dates in 2024
> set.seed(123)  # for reproducibility
>
> df <- df %>%
+   mutate(
+     exam_date = sample(seq(ymd("2024-01-01"), ymd("2024-12-31"), by = "day"),
+                        size = n(), replace = TRUE)
+   )
>
```

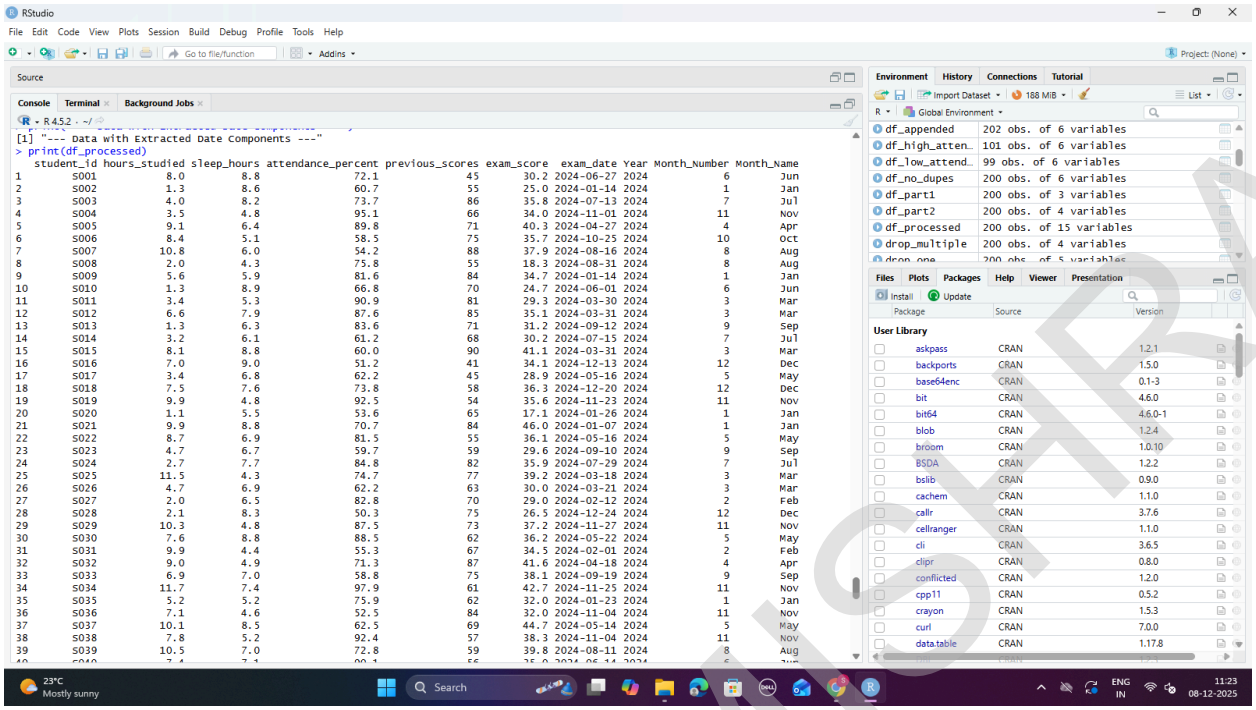


```
>
> # ------------------------------------------------------------
> # 3. Extract date components
> # ------------------------------------------------------------
>
> df_processed <- df %>%
+   mutate(
+     Year = year(exam_date),
+     Month_Number = month(exam_date),
+     Month_Name = month(exam_date, label = TRUE),
+     Day = day(exam_date),
+     Weekday_Number = wday(exam_date),
+     Weekday_Name = wday(exam_date, label = TRUE, abbr = FALSE),
+     Quarter = quarter(exam_date),
+     Day_of_Year = yday(exam_date)
+   )
>
> # ------------------------------------------------------------
> # 4. Print Results
> # ------------------------------------------------------------
> print("--- Data with Extracted Date Components ---")
[1] "--- Data with Extracted Date Components ---"
> print(df_processed)
   student_id hours_studied sleep_hours attendance_percent previous_scores exam_score  exam_date Year Month_Number Month_Name
1        S001           8.0         8.8              72.1              45        30.2 2024-06-27 2024            6        Jun
2        S002           1.3         8.6              60.7              55        25.0 2024-01-14 2024            1        Jan
3        S003           4.0         8.2              73.7              86        35.8 2024-07-13 2024            7        Jul
4        S004           3.5         4.8              95.1              66        34.0 2024-11-01 2024           11        Nov
5        S005           9.1         6.4              89.8              71        40.3 2024-04-27 2024            4        Apr
6        S006           8.4         5.1              58.5              75        35.7 2024-10-25 2024           10        Oct
7        S007          10.8         6.0              54.2              88        37.9 2024-08-16 2024            8        Aug
8        S008           2.0         4.3              75.8              55        18.3 2024-08-31 2024            8        Aug
9        S009           5.6         5.9              81.6              84        34.7 2024-01-14 2024            1        Jan
10       S010           1.3         8.9              66.8              70        24.7 2024-06-01 2024            6        Jun
11       S011           3.4         5.3              90.9              81        29.3 2024-03-30 2024            3        Mar
12       S012           6.6         7.9              87.6              85        35.1 2024-03-31 2024            3        Mar
13       S013           1.3         6.3              83.6              71        31.2 2024-09-12 2024            9        Sep
14       S014           3.2         6.1              61.2              68        30.2 2024-07-15 2024            7        Jul
15       S015           8.1         8.8              60.0              90        41.1 2024-03-31 2024            3        Mar
16       S016           7.0         9.0              51.2              41        34.1 2024-12-13 2024           12        Dec
17       S017           3.4         6.8              62.2              45        28.9 2024-05-16 2024            5        May
18       S018           7.5         7.6              73.8              58        36.3 2024-12-20 2024           12        Dec
19       S019           9.9         4.8              92.5              54        35.6 2024-11-23 2024           11        Nov
```

NAME: AASHKA MISHRA

ROLL NO: S095

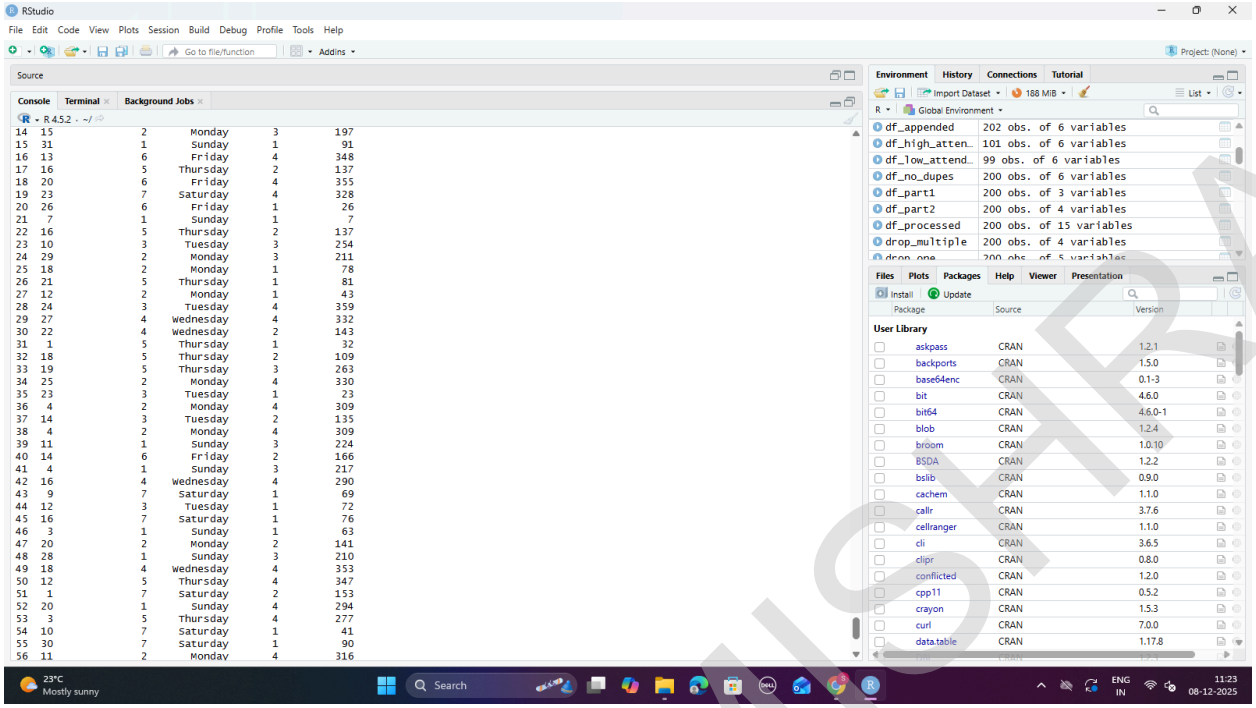
SHETH L.U.J. AND SIR M.V. COLLEGE
SUBJECT: DATA ANALYSIS WITH R

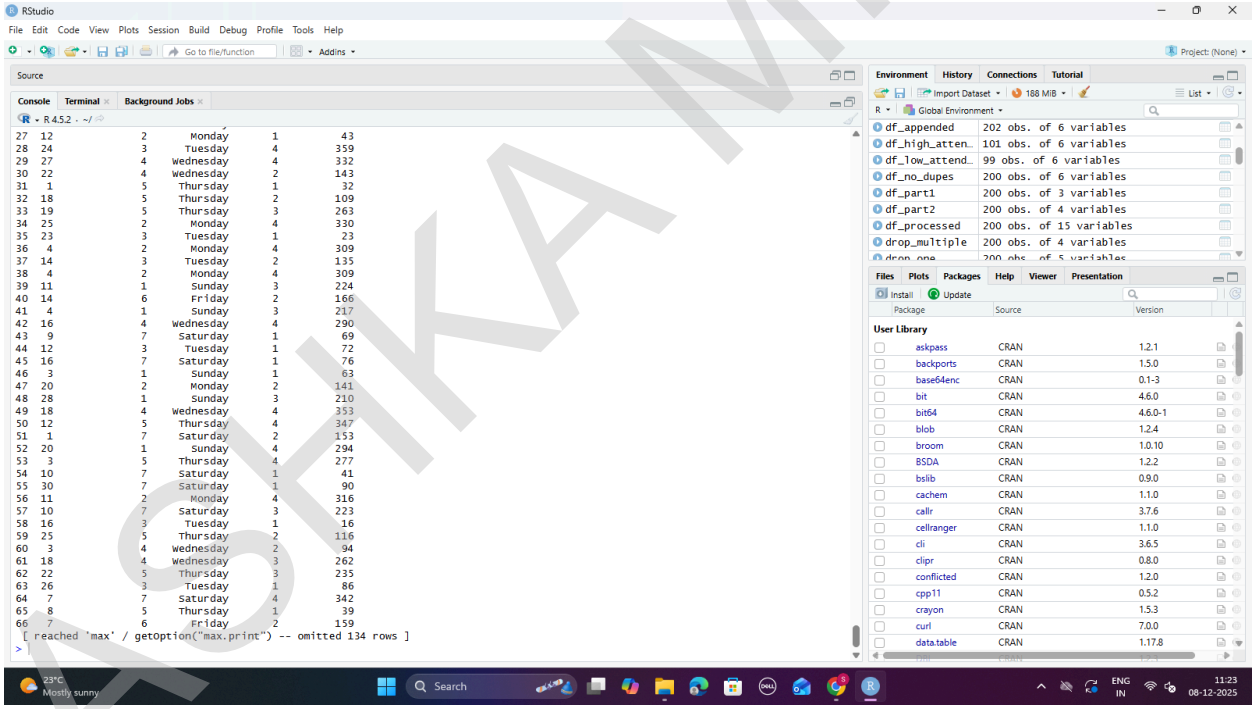


```
[1] "--- Data with Extracted Date Components ---"
> print(df_processed)
   student_id hours_studied sleep_hours attendance_percent previous_scores exam_score  exam_date Year Month_Number Month_Name
1        S001           8.0         8.8                72.1             45       30.2 2024-06-27 2024            6        Jun
2        S002           1.3         8.6                60.7             55       25.0 2024-01-14 2024            1        Jan
3        S003           4.0         8.2                73.7             86       35.8 2024-07-13 2024            7        Jul
4        S004           3.5         4.8                95.1             66       34.0 2024-11-01 2024           11        Nov
5        S005           9.1         6.4                89.8             71       40.3 2024-04-27 2024            4        Apr
6        S006           8.4         5.1                58.5             75       35.7 2024-10-25 2024           10        Oct
7        S007          10.8         6.0                54.2             88       37.9 2024-08-16 2024            8        Aug
8        S008           2.0         4.3                75.8             55       18.3 2024-08-31 2024            8        Aug
9        S009           5.6         5.9                81.6             84       34.7 2024-01-14 2024            1        Jan
10       S010           1.3         8.9                66.8             70       24.7 2024-06-01 2024            6        Jun
11       S011           3.4         5.3                90.9             81       29.3 2024-03-30 2024            3        Mar
12       S012           6.6         7.9                87.6             85       35.1 2024-03-31 2024            3        Mar
13       S013           1.3         6.3                83.6             71       31.2 2024-09-12 2024            9        Sep
14       S014           3.2         6.1                61.2             68       30.2 2024-07-15 2024            7        Jul
15       S015           8.1         8.8                60.0             90       41.1 2024-03-31 2024            3        Mar
16       S016           7.0         9.0                51.2             41       34.1 2024-12-13 2024           12        Dec
17       S017           3.4         6.8                62.2             45       28.9 2024-05-16 2024            5        May
18       S018           7.5         7.6                73.8             58       36.3 2024-12-20 2024           12        Dec
19       S019           9.9         4.8                92.5             54       35.6 2024-11-23 2024           11        Nov
20       S020           1.1         5.5                53.6             65       17.1 2024-01-26 2024            1        Jan
21       S021           9.9         8.8                70.7             84       46.0 2024-01-07 2024            1        Jan
22       S022           8.7         6.9                81.5             55       36.1 2024-05-16 2024            5        May
23       S023           4.7         6.7                59.7             59       29.6 2024-09-10 2024            9        Sep
24       S024           2.7         7.7                84.8             82       35.9 2024-07-29 2024            7        Jul
25       S025          11.5         4.3                74.7             77       39.2 2024-03-18 2024            3        Mar
26       S026           4.7         6.9                62.2             63       30.0 2024-03-21 2024            3        Mar
27       S027           2.0         6.5                82.8             70       29.0 2024-02-12 2024            2        Feb
28       S028           2.1         8.3                50.3             75       26.5 2024-12-24 2024           12        Dec
29       S029          10.3         4.8                87.5             73       37.2 2024-11-27 2024           11        Nov
30       S030           7.6         8.8                88.5             62       36.2 2024-05-22 2024            5        May
31       S031           9.9         4.4                55.3             67       34.5 2024-02-01 2024            2        Feb
32       S032           9.0         4.9                71.3             87       41.6 2024-04-18 2024            4        Apr
33       S033           6.9         7.0                58.8             75       38.1 2024-09-19 2024            9        Sep
34       S034          11.7         7.4                97.9             61       42.7 2024-11-25 2024           11        Nov
35       S035           5.2         5.2                75.9             62       32.0 2024-01-23 2024            1        Jan
36       S036           7.1         4.6                52.5             84       32.0 2024-11-04 2024           11        Nov
37       S037          10.1         8.5                62.5             69       44.7 2024-05-14 2024            5        May
38       S038           7.8         5.2                92.4             57       38.3 2024-11-04 2024           11        Nov
39       S039          10.5         7.0                72.8             59       39.8 2024-08-11 2024            8        Aug
40       S040           7.4         7.1                90.1             56       35.0 2024-06-14 2024            6        Jun
41       S041           8.8         6.1                83.4             54       34.2 2024-08-04 2024            8        Aug
42       S042           1.5         6.9                99.4             47       23.1 2024-10-16 2024           10        Oct
43       S043           3.5         6.6                79.8             86       37.2 2024-03-09 2024            3        Mar
44       S044           4.2         8.7                97.5             52       30.3 2024-03-12 2024            3        Mar
45       S045           1.9         5.0                94.6             60       26.1 2024-03-16 2024            3        Mar
46       S046           3.6         7.6                80.6             47       31.8 2024-03-03 2024            3        Mar
47       S047           2.1         5.2                86.0             87       31.6 2024-05-20 2024            5        May
48       S048           4.1         6.0                75.2             74       31.1 2024-07-28 2024            7        Jul
49       S049           8.0         7.4                91.5             88       41.2 2024-12-18 2024           12        Dec
50       S050           5.0         5.5                77.4             84       30.6 2024-12-12 2024           12        Dec
51       S051           5.1         5.6                94.9             51       28.8 2024-06-01 2024            6        Jun
52       S052           3.3         7.8                87.2             52       30.7 2024-10-20 2024           10        Oct
53       S053           3.9         4.4                73.7             53       26.9 2024-10-03 2024           10        Oct
54       S054          11.3         6.3                63.0             87       46.4 2024-02-10 2024            2        Feb
55       S055           8.1         9.0                62.4             70       34.4 2024-03-30 2024            3        Mar
56       S056           7.7         9.0                81.9             57       37.0 2024-11-11 2024           11        Nov
57       S057           2.9         4.4                88.3             86       35.2 2024-08-10 2024            8        Aug
58       S058           9.0         5.1                76.1             77       38.2 2024-01-16 2024            1        Jan
59       S059           2.8         5.3                81.3             88       33.0 2024-04-25 2024            4        Apr
60       S060           5.2         8.7                63.7             73       29.2 2024-04-03 2024            4        Apr
61       S061          11.9         8.4                53.9             78       48.6 2024-09-18 2024            9        Sep
62       S062           8.0         8.4                64.3             58       36.1 2024-08-22 2024            8        Aug
63       S063           7.1         5.8                63.6             46       27.1 2024-03-26 2024            3        Mar
64       S064           8.5         4.8                66.0             93       36.1 2024-12-07 2024           12        Dec
65       S065          10.3         8.2                77.0             52       39.5 2024-02-08 2024            2        Feb
66       S066           9.5         7.5                56.9             58       36.7 2024-06-07 2024            6        Jun
   Day Weekday_Number Weekday_Name Quarter Day_of_Year
1   27              5     Thursday       2         179
2   14              1       Sunday       1          14
3   13              7     Saturday       3         195
4    1              6       Friday       4         306
5   27              7     Saturday       2         118
6   25              6       Friday       4         299
7   16              6       Friday       3         229
8   31              7     Saturday       3         244
9   14              1       Sunday       1          14
10   1              7     Saturday       2         153
11  30              7     Saturday       1          90
12  31              1       Sunday       1          91
13  12              5     Thursday       3         256
14  15              2       Monday       3         197
```

NAME: AASHKA MISHRA
ROLL NO: S095

SHETH L.U.J. AND SIR M.V. COLLEGE
SUBJECT: DATA ANALYSIS WITH R





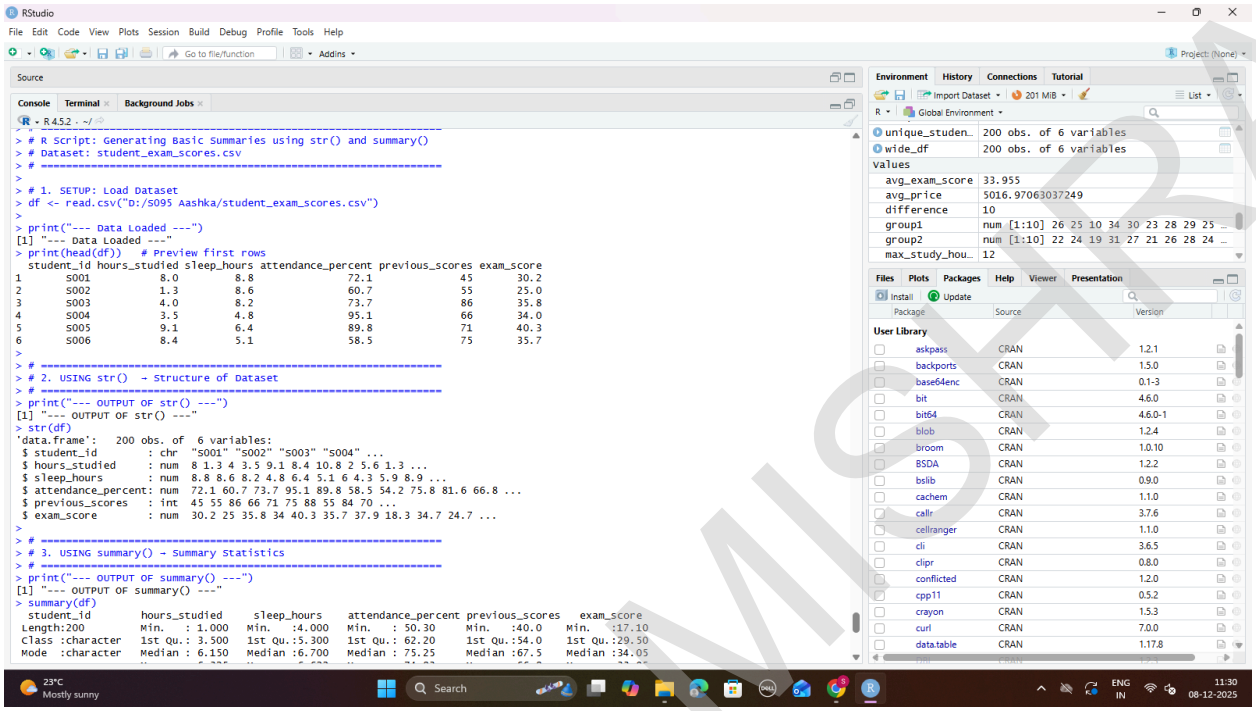
NAME: AASHKA MISHRA
ROLL NO: S095

SHETH L.U.J. AND SIR M.V. COLLEGE
SUBJECT: DATA ANALYSIS WITH R

PRACTICAL NO. 15

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

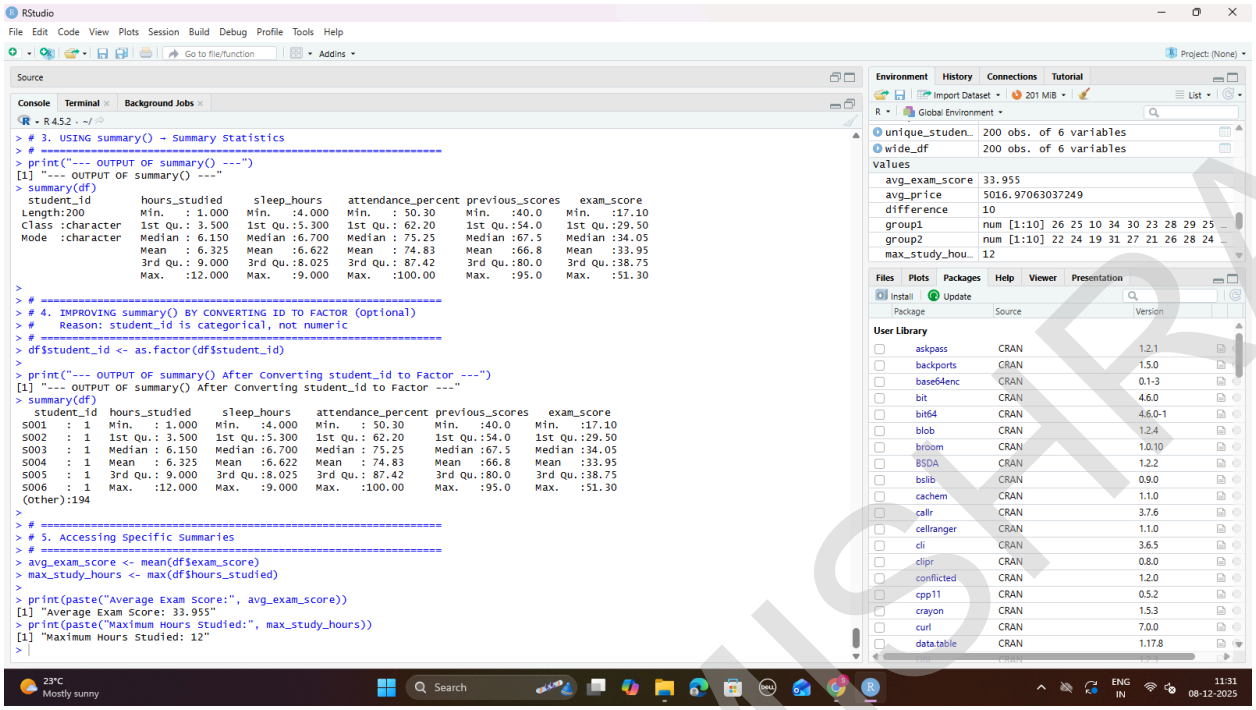
OUTPUT:



```
> # R Script: Generating Basic Summaries using str() and summary()
> # Dataset: student_exam_scores.csv
> # ================================================================
>
> # 1. SETUP: Load Dataset
> df <- read.csv("D:/S095 Aashka/student_exam_scores.csv")
>
> print("--- Data Loaded ---")
[1] "--- Data Loaded ---"
> print(head(df))   # Preview first rows
  student_id hours_studied sleep_hours attendance_percent previous_scores exam_score
1       S001           8.0         8.8                72.1              45       30.2
2       S002           1.3         8.6                60.7              55       25.0
3       S003           4.0         8.2                73.7              86       35.8
4       S004           3.5         4.8                95.1              66       34.0
5       S005           9.1         6.4                89.8              71       40.3
6       S006           8.4         5.1                58.5              75       35.7
>
> # ================================================================
> # 2. USING str()  → Structure of Dataset
> # ================================================================
> print("--- OUTPUT OF str() ---")
[1] "--- OUTPUT OF str() ---"
> str(df)
'data.frame':   200 obs. of  6 variables:
 $ student_id        : chr  "S001" "S002" "S003" "S004" ...
 $ hours_studied     : num  8 1.3 4 3.5 9.1 8.4 10.8 2 5.6 1.3 ...
 $ sleep_hours       : num  8.8 8.6 8.2 4.8 6.4 5.1 6 4.3 5.9 8.9 ...
 $ attendance_percent: num  72.1 60.7 73.7 95.1 89.8 58.5 54.2 75.8 81.6 66.8 ...
 $ previous_scores   : int  45 55 86 66 71 75 88 55 84 70 ...
 $ exam_score        : num  30.2 25 35.8 34 40.3 35.7 37.9 18.3 34.7 24.7 ...
>
> # ================================================================
> # 3. USING summary()  → Summary Statistics
> # ================================================================
> print("--- OUTPUT OF summary() ---")
[1] "--- OUTPUT OF summary() ---"
> summary(df)
  student_id        hours_studied     sleep_hours    attendance_percent previous_scores   exam_score
 Length:200         Min.   : 1.000   Min.   :4.000   Min.   : 50.30     Min.   :40.0    Min.   :17.10
 Class :character   1st Qu.: 3.500   1st Qu.:5.300   1st Qu.: 62.20     1st Qu.:54.0    1st Qu.:29.50
 Mode  :character   Median : 6.150   Median :6.700   Median : 75.25     Median :67.5    Median :34.05
                    Mean   : 6.325   Mean   :6.622   Mean   : 74.83     Mean   :66.8    Mean   :33.95
                    3rd Qu.: 9.000   3rd Qu.:8.025   3rd Qu.: 87.42     3rd Qu.:80.0    3rd Qu.:38.75
                    Max.   :12.000   Max.   :9.000   Max.   :100.00     Max.   :95.0    Max.   :51.30
>
> # ================================================================
> # 4. IMPROVING summary() BY CONVERTING ID TO FACTOR (Optional)
> #    Reason: student_id is categorical, not numeric
> # ================================================================
> df$student_id <- as.factor(df$student_id)
>
> print("--- OUTPUT OF summary() After Converting student_id to Factor ---")
[1] "--- OUTPUT OF summary() After Converting student_id to Factor ---"
> summary(df)
  student_id  hours_studied     sleep_hours    attendance_percent previous_scores   exam_score
 S001   :  1   Min.   : 1.000   Min.   :4.000   Min.   : 50.30     Min.   :40.0    Min.   :17.10
 S002   :  1   1st Qu.: 3.500   1st Qu.:5.300   1st Qu.: 62.20     1st Qu.:54.0    1st Qu.:29.50
 S003   :  1   Median : 6.150   Median :6.700   Median : 75.25     Median :67.5    Median :34.05
 S004   :  1   Mean   : 6.325   Mean   :6.622   Mean   : 74.83     Mean   :66.8    Mean   :33.95
 S005   :  1   3rd Qu.: 9.000   3rd Qu.:8.025   3rd Qu.: 87.42     3rd Qu.:80.0    3rd Qu.:38.75
 S006   :  1   Max.   :12.000   Max.   :9.000   Max.   :100.00     Max.   :95.0    Max.   :51.30
 (Other):194
>
> # ================================================================
> # 5. Accessing Specific Summaries
> # ================================================================
> avg_exam_score <- mean(df$exam_score)
> max_study_hours <- max(df$hours_studied)
>
> print(paste("Average Exam Score:", avg_exam_score))
[1] "Average Exam Score: 33.955"
> print(paste("Maximum Hours Studied:", max_study_hours))
[1] "Maximum Hours Studied: 12"
```

NAME: AASHKA MISHRA
ROLL NO: S095

SHETH L.U.J. AND SIR M.V. COLLEGE
SUBJECT: DATA ANALYSIS WITH R



```
> # 3. USING summary() → Summary Statistics
> # ================================================================
> print("--- OUTPUT OF summary() ---")
[1] "--- OUTPUT OF summary() ---"
> summary(df)
  student_id         hours_studied     sleep_hours    attendance_percent previous_scores   exam_score
 Length:200         Min.   : 1.000   Min.   :4.000   Min.   : 50.30     Min.   :40.0    Min.   :17.10
 Class :character   1st Qu.: 3.500   1st Qu.:5.300   1st Qu.: 62.20     1st Qu.:54.0    1st Qu.:29.50
 Mode  :character   Median : 6.150   Median :6.700   Median : 75.25     Median :67.5    Median :34.05
                    Mean   : 6.325   Mean   :6.622   Mean   : 74.83     Mean   :66.8    Mean   :33.95
                    3rd Qu.: 9.000   3rd Qu.:8.025   3rd Qu.: 87.42     3rd Qu.:80.0    3rd Qu.:38.75
                    Max.   :12.000   Max.   :9.000   Max.   :100.00     Max.   :95.0    Max.   :51.30
>
> # ================================================================
> # 4. IMPROVING summary() BY CONVERTING ID TO FACTOR (Optional)
> #    Reason: student_id is categorical, not numeric
> # ================================================================
> df$student_id <- as.factor(df$student_id)
>
> print("--- OUTPUT OF summary() After Converting student_id to Factor ---")
[1] "--- OUTPUT OF summary() After Converting student_id to Factor ---"
> summary(df)
   student_id  hours_studied     sleep_hours    attendance_percent previous_scores   exam_score
 S001   :  1   Min.   : 1.000   Min.   :4.000   Min.   : 50.30     Min.   :40.0    Min.   :17.10
 S002   :  1   1st Qu.: 3.500   1st Qu.:5.300   1st Qu.: 62.20     1st Qu.:54.0    1st Qu.:29.50
 S003   :  1   Median : 6.150   Median :6.700   Median : 75.25     Median :67.5    Median :34.05
 S004   :  1   Mean   : 6.325   Mean   :6.622   Mean   : 74.83     Mean   :66.8    Mean   :33.95
 S005   :  1   3rd Qu.: 9.000   3rd Qu.:8.025   3rd Qu.: 87.42     3rd Qu.:80.0    3rd Qu.:38.75
 S006   :  1   Max.   :12.000   Max.   :9.000   Max.   :100.00     Max.   :95.0    Max.   :51.30
 (other):194
>
> # ================================================================
> # 5. Accessing Specific Summaries
> # ================================================================
> avg_exam_score <- mean(df$exam_score)
> max_study_hours <- max(df$hours_studied)
>
> print(paste("Average Exam Score:", avg_exam_score))
[1] "Average Exam Score: 33.955"
> print(paste("Maximum Hours Studied:", max_study_hours))
[1] "Maximum Hours Studied: 12"
>
```

NAME: AASHKA MISHRA
ROLL NO: S095