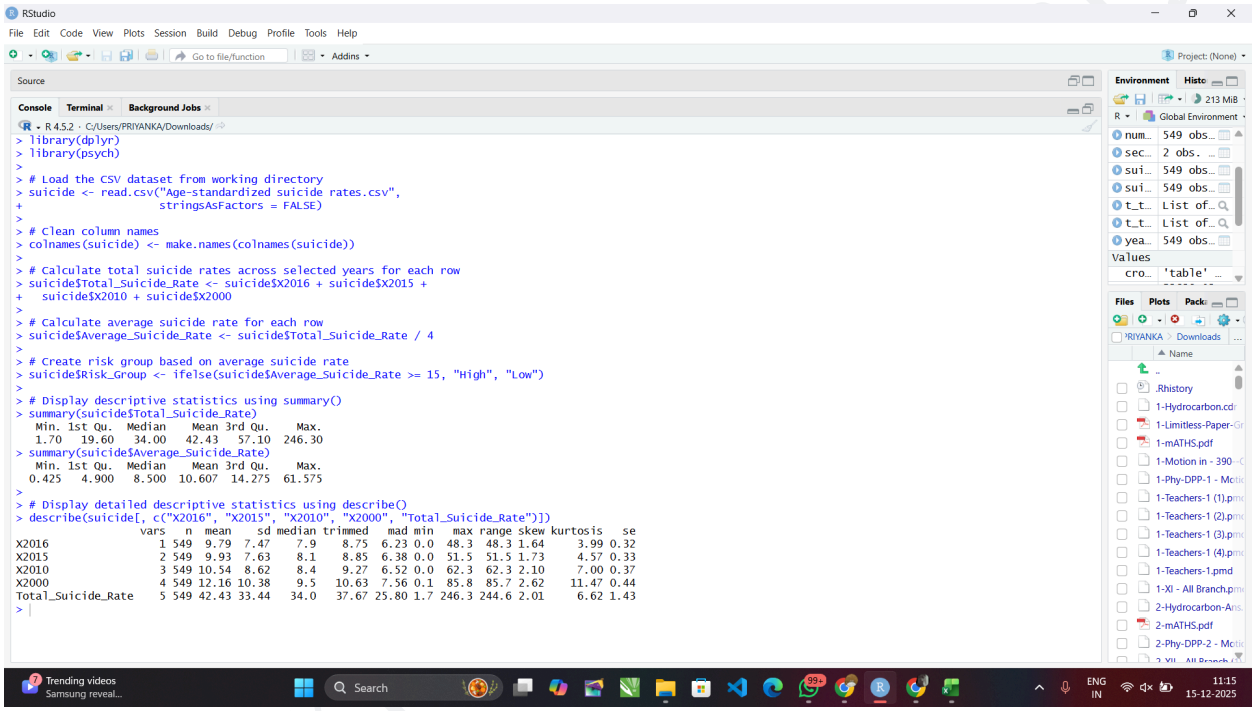
SHETH L.U.J AND SIR M.V. COLLEGE
SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R

Module 2 Practical 1-6

Aim: Generating descriptive statistics using summary() or describe() (R).

OUTPUT:



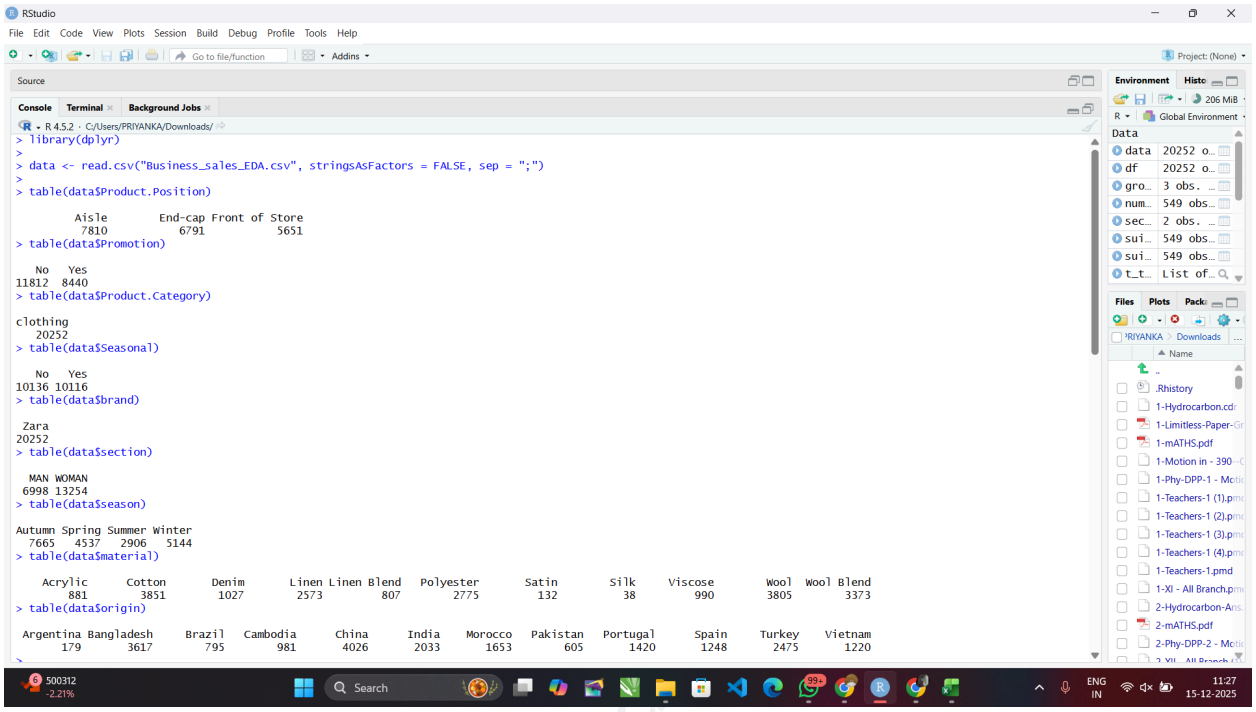
Krishna Nandkumar Chaudhary
S074

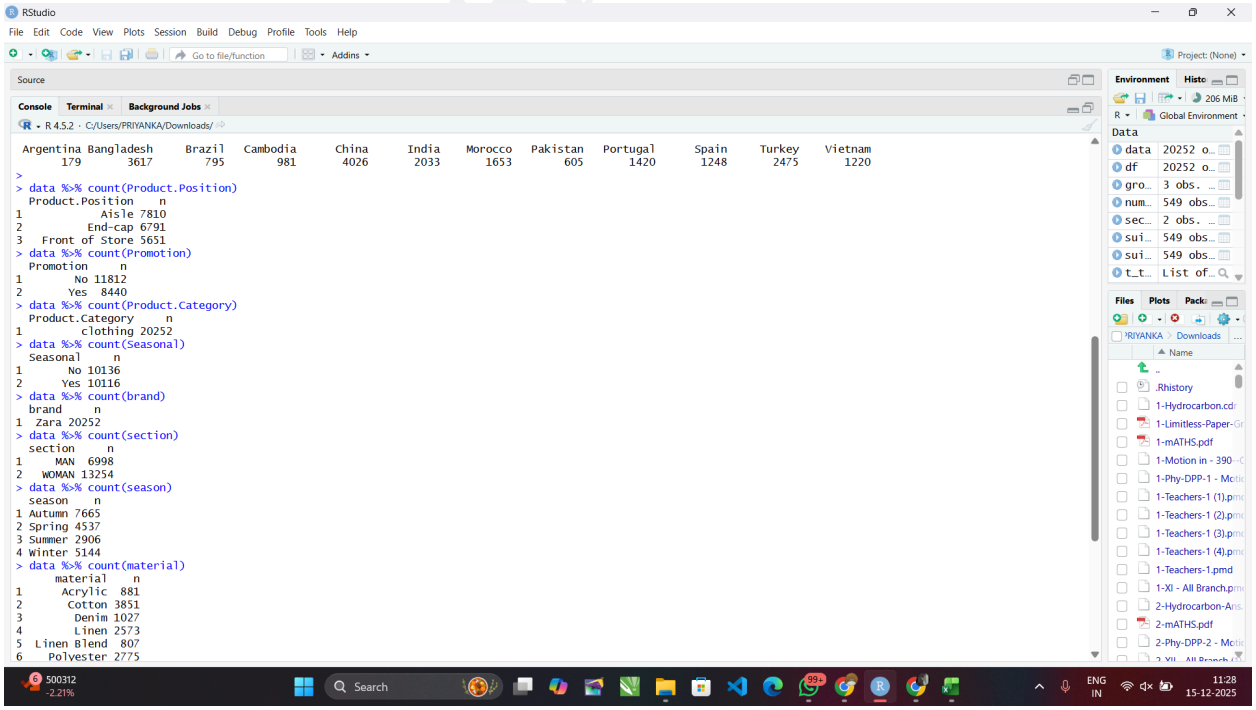
SHETH L.U.J AND SIR M.V. COLLEGE

SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R

Aim: Generating frequency tables using table() or count() (R).

OUTPUT:





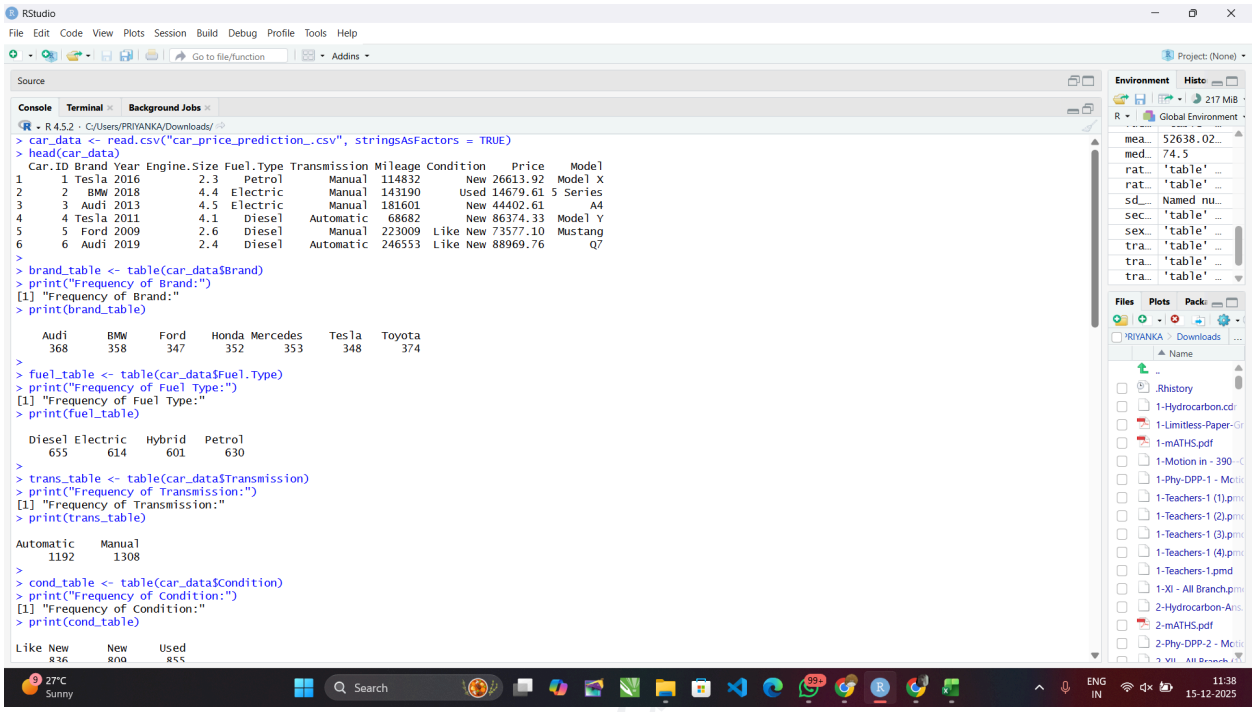
Krishna Nandkumar Chaudhary
S074

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

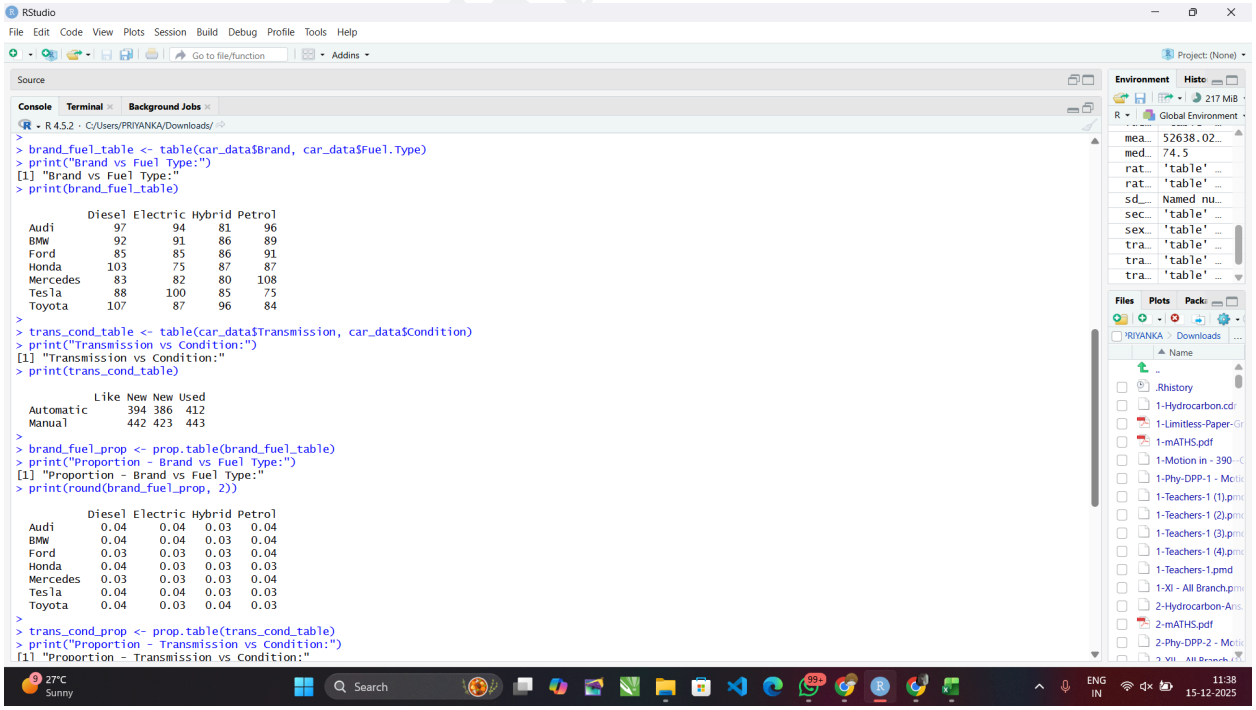
SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R

Aim: Creating cross-tabulations and two-way tables using table() (R).

OUTPUT:



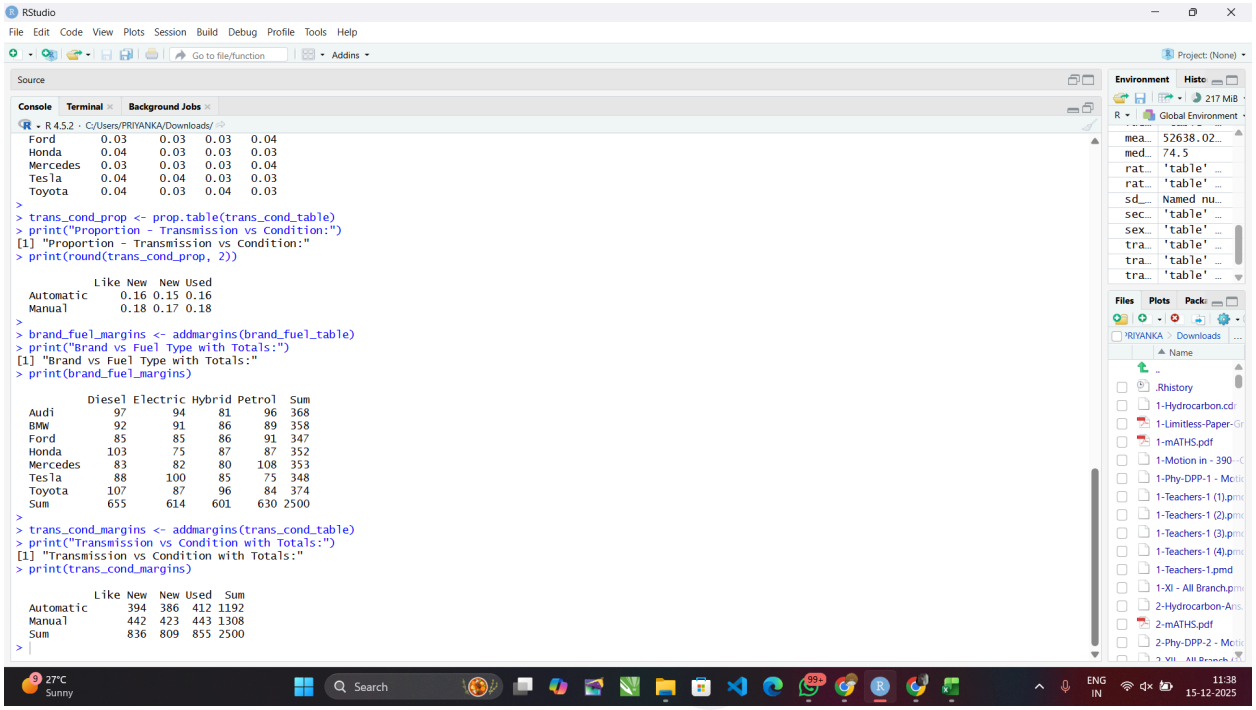
```
> car_data <- read.csv("car_price_prediction_.csv", stringsAsFactors = TRUE)
> head(car_data)
  Car.ID Brand Year Engine.Size Fuel.Type Transmission Mileage Condition    Price    Model
1      1 Tesla 2016         2.3    Petrol       Manual  114832       New 26613.92  Model X
2      2   BMW 2018         4.4  Electric       Manual  143190      Used 14679.61 5 Series
3      3  Audi 2013         4.5  Electric       Manual  181601       New 44402.61       A4
4      4 Tesla 2011         4.1    Diesel    Automatic   68682       New 86374.33  Model Y
5      5  Ford 2009         2.6    Diesel       Manual  223009  Like New 73577.10  Mustang
6      6  Audi 2019         2.4    Diesel    Automatic  246553  Like New 88969.76       Q7
>
> brand_table <- table(car_data$Brand)
> print("Frequency of Brand:")
[1] "Frequency of Brand:"
> print(brand_table)

   Audi     BMW    Ford   Honda Mercedes   Tesla  Toyota
    368     358     347     352     353     348     374
>
> fuel_table <- table(car_data$Fuel.Type)
> print("Frequency of Fuel Type:")
[1] "Frequency of Fuel Type:"
> print(fuel_table)

  Diesel Electric   Hybrid   Petrol
     655      614      601      630
>
> trans_table <- table(car_data$Transmission)
> print("Frequency of Transmission:")
[1] "Frequency of Transmission:"
> print(trans_table)

Automatic    Manual
     1192      1308
>
> cond_table <- table(car_data$Condition)
> print("Frequency of Condition:")
[1] "Frequency of Condition:"
> print(cond_table)

Like New      New     Used
     836      809      855
```



```
>
> brand_fuel_table <- table(car_data$Brand, car_data$Fuel.Type)
> print("Brand vs Fuel Type:")
[1] "Brand vs Fuel Type:"
> print(brand_fuel_table)

          Diesel Electric Hybrid Petrol
  Audi        97       94     81     96
  BMW         92       91     86     89
  Ford        85       85     86     91
  Honda      103       75     87     87
  Mercedes    83       82     80    108
  Tesla       88      100     85     75
  Toyota     107       87     96     84
>
> trans_cond_table <- table(car_data$Transmission, car_data$Condition)
> print("Transmission vs Condition:")
[1] "Transmission vs Condition:"
> print(trans_cond_table)

           Like New New Used
  Automatic      394 386  412
  Manual         442 423  443
>
> brand_fuel_prop <- prop.table(brand_fuel_table)
> print("Proportion - Brand vs Fuel Type:")
[1] "Proportion - Brand vs Fuel Type:"
> print(round(brand_fuel_prop, 2))

          Diesel Electric Hybrid Petrol
  Audi      0.04     0.04   0.03   0.04
  BMW       0.04     0.04   0.03   0.04
  Ford      0.03     0.03   0.03   0.04
  Honda     0.04     0.03   0.03   0.03
  Mercedes  0.03     0.03   0.03   0.04
  Tesla     0.04     0.04   0.03   0.03
  Toyota    0.04     0.03   0.04   0.03
>
> trans_cond_prop <- prop.table(trans_cond_table)
> print("Proportion - Transmission vs Condition:")
[1] "Proportion - Transmission vs Condition:"
```

Krishna Nandkumar Chaudhary
S074

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



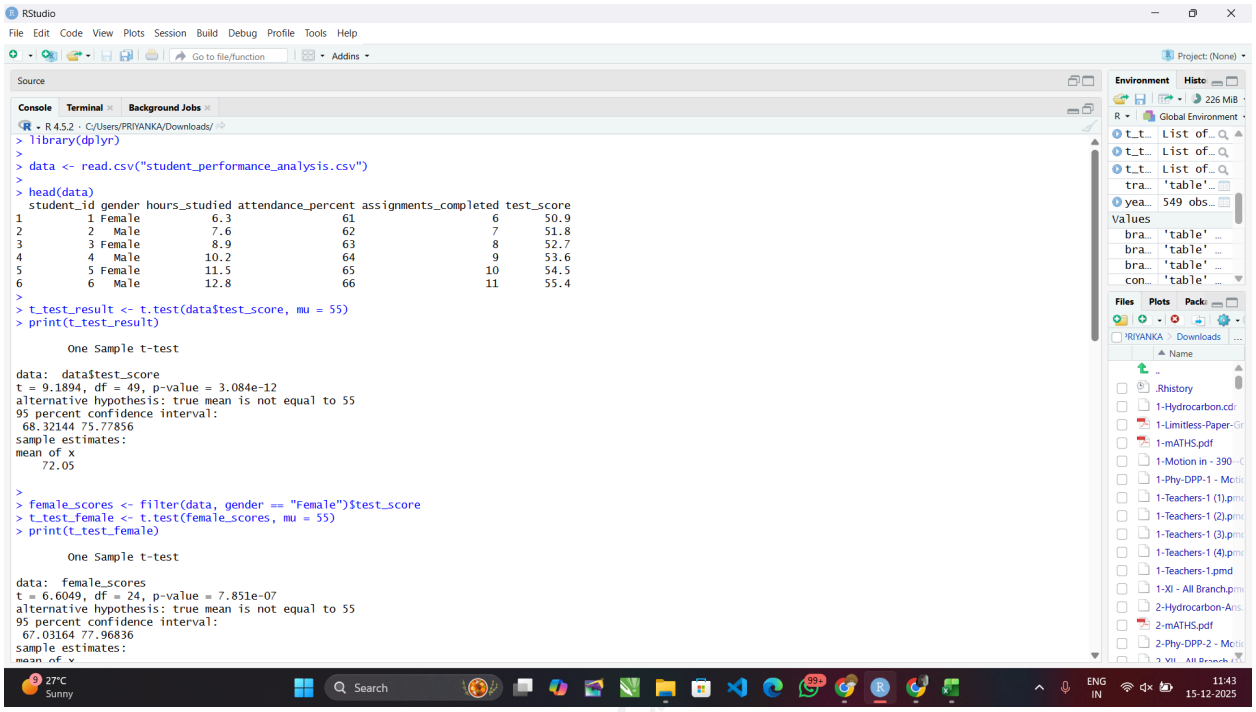
Krishna Nandkumar Chaudhary
S074

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

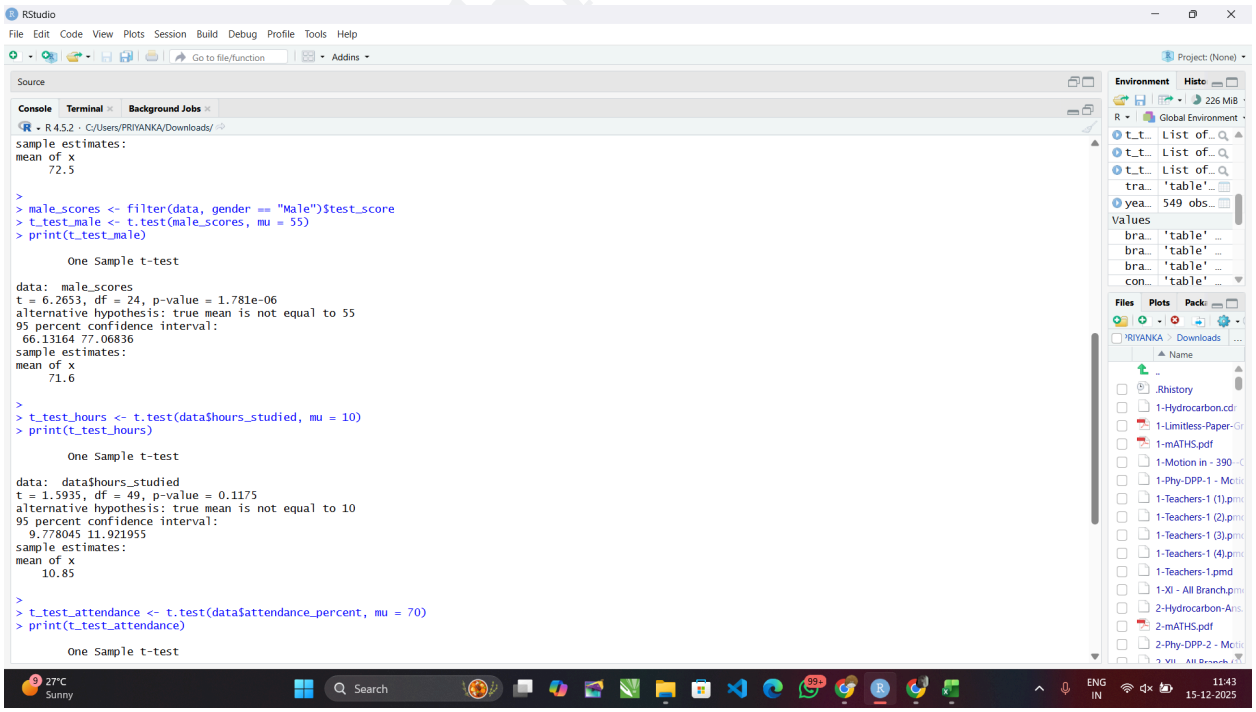
SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R

Aim: Performing one-sample t-tests using t.test() (R).

OUTPUT:



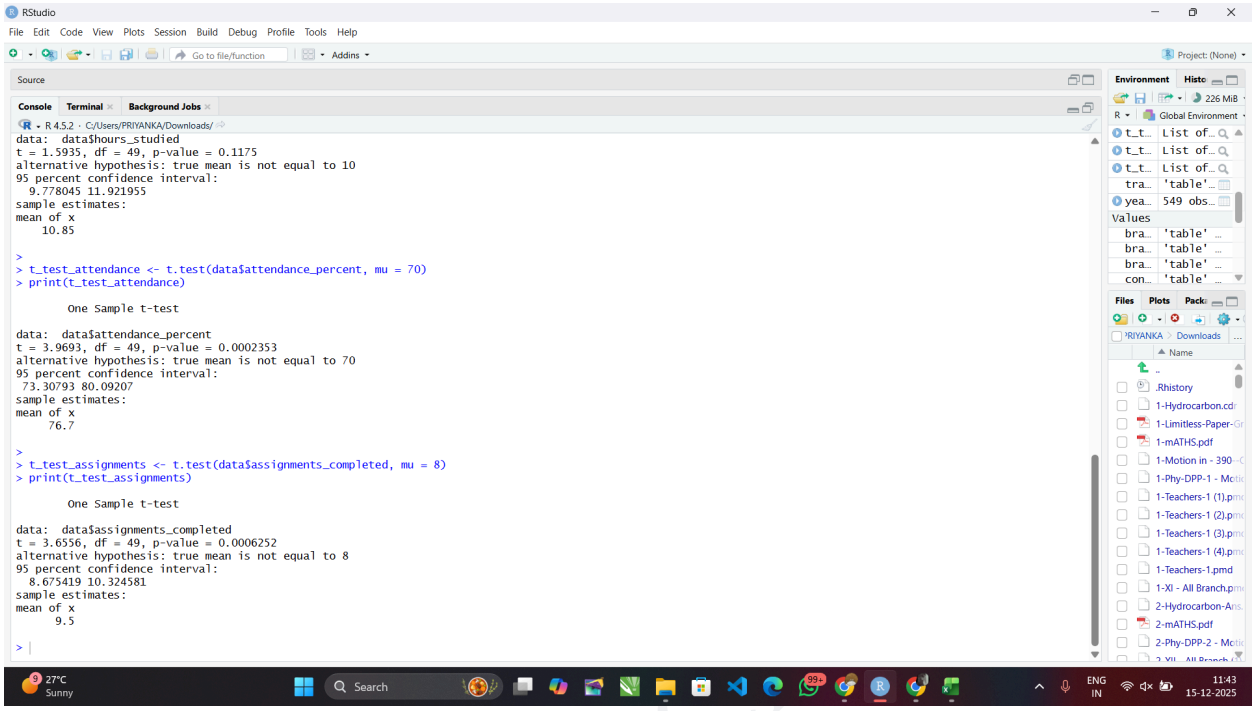
```
> library(dplyr)
>
> data <- read.csv("student_performance_analysis.csv")
>
> head(data)
  student_id gender hours_studied attendance_percent assignments_completed test_score
1          1 Female          6.3                 61                     6       50.9
2          2   Male          7.6                 62                     7       51.8
3          3 Female          8.9                 63                     8       52.7
4          4   Male         10.2                 64                     9       53.6
5          5 Female         11.5                 65                    10       54.5
6          6   Male         12.8                 66                    11       55.4
>
> t_test_result <- t.test(data$test_score, mu = 55)
> print(t_test_result)

        One Sample t-test

data:  data$test_score
t = 9.1894, df = 49, p-value = 3.084e-12
alternative hypothesis: true mean is not equal to 55
95 percent confidence interval:
 68.32144 75.77856
sample estimates:
mean of x
    72.05

>
> female_scores <- filter(data, gender == "Female")$test_score
> t_test_female <- t.test(female_scores, mu = 55)
> print(t_test_female)

        One Sample t-test

data:  female_scores
t = 6.6049, df = 24, p-value = 7.851e-07
alternative hypothesis: true mean is not equal to 55
95 percent confidence interval:
 67.03164 77.96836
sample estimates:
```



```
sample estimates:
mean of x
     72.5

>
> male_scores <- filter(data, gender == "Male")$test_score
> t_test_male <- t.test(male_scores, mu = 55)
> print(t_test_male)

        One Sample t-test

data:  male_scores
t = 6.2653, df = 24, p-value = 1.781e-06
alternative hypothesis: true mean is not equal to 55
95 percent confidence interval:
 66.13164 77.06836
sample estimates:
mean of x
     71.6

>
> t_test_hours <- t.test(data$hours_studied, mu = 10)
> print(t_test_hours)

        One Sample t-test

data:  data$hours_studied
t = 1.5935, df = 49, p-value = 0.1175
alternative hypothesis: true mean is not equal to 10
95 percent confidence interval:
  9.778045 11.921955
sample estimates:
mean of x
    10.85

>
> t_test_attendance <- t.test(data$attendance_percent, mu = 70)
> print(t_test_attendance)

        One Sample t-test
```

Krishna Nandkumar Chaudhary
S074

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



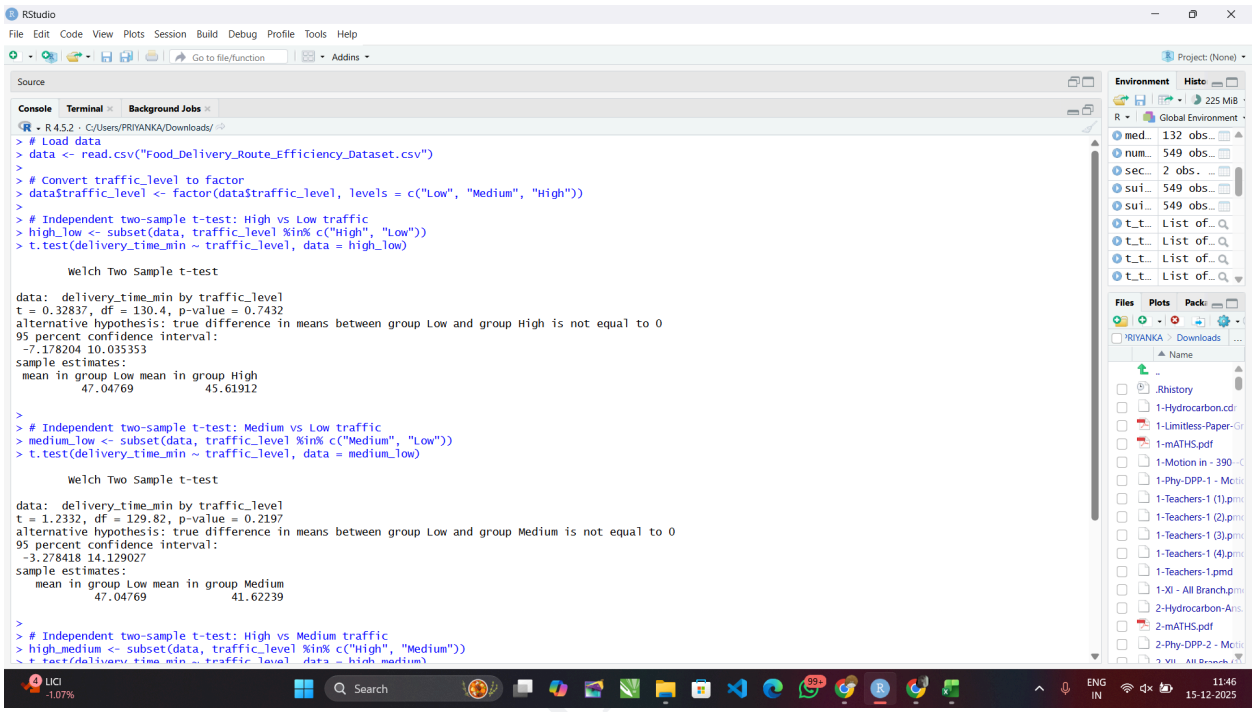
Krishna Nandkumar Chaudhary
S074

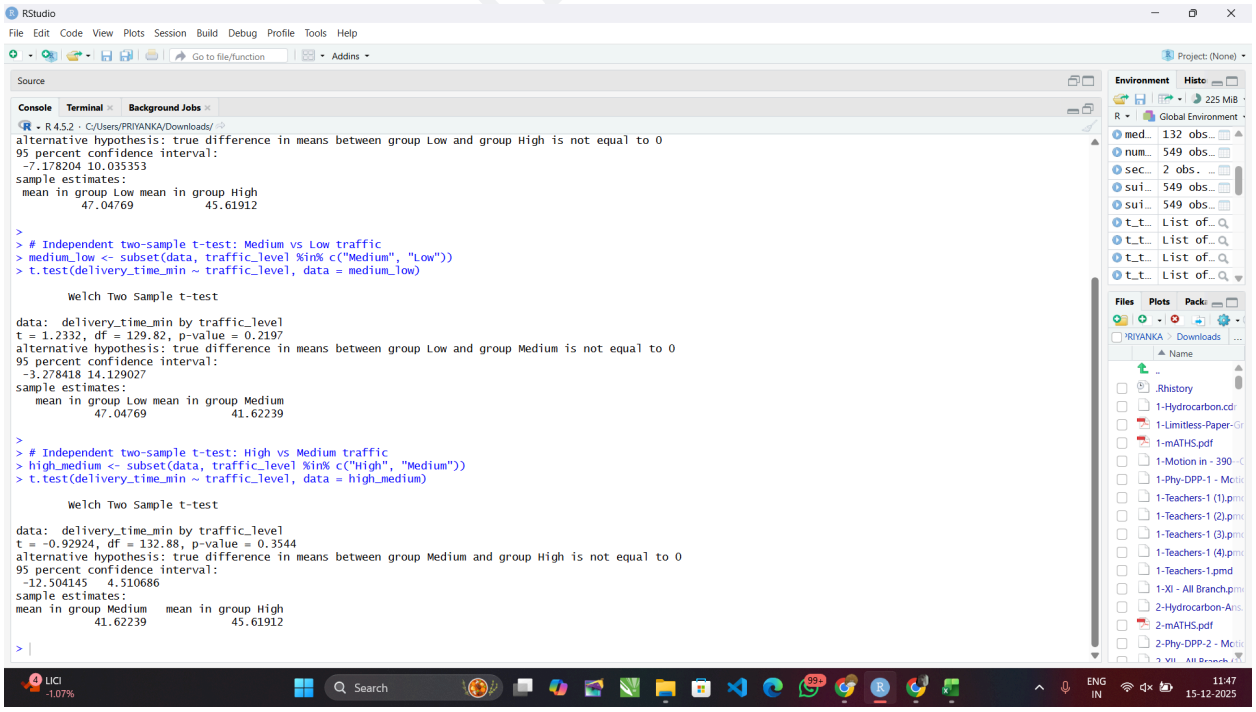
SHETH L.U.J AND SIR M.V. COLLEGE

SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R

Aim: Performing independent two-sample t-tests using t.test() with grouping (R).

OUTPUT:





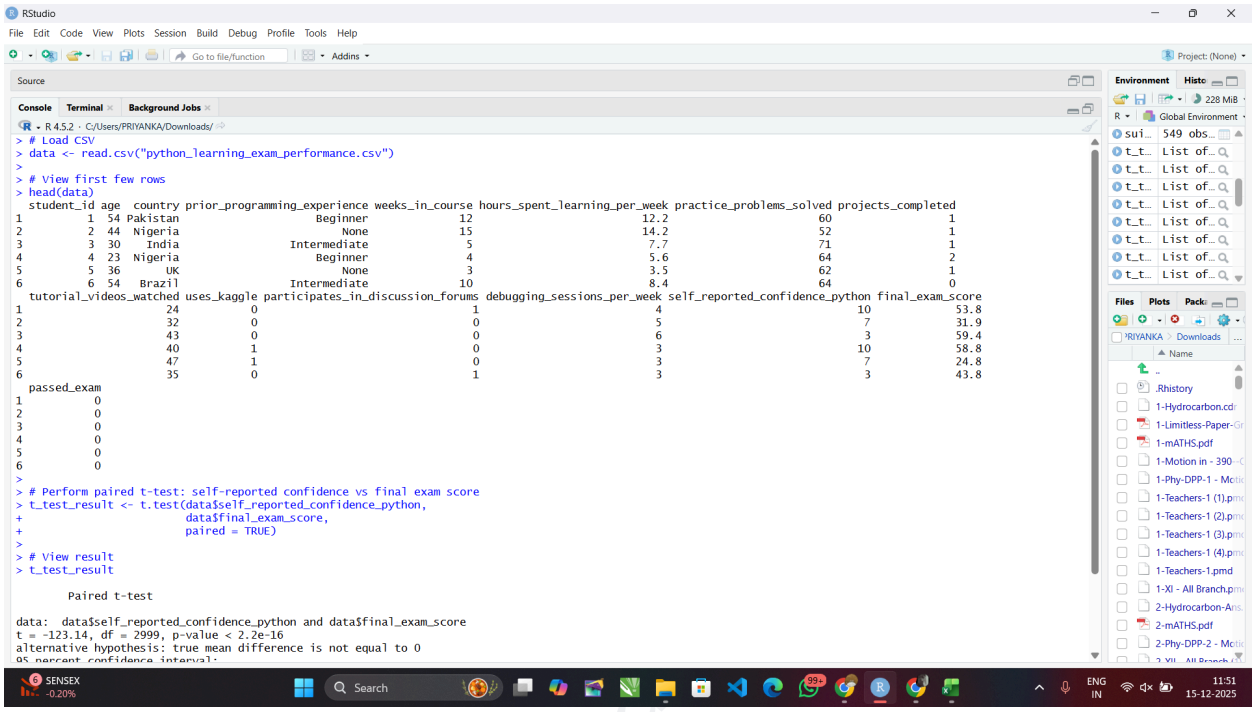
Krishna Nandkumar Chaudhary
S074

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

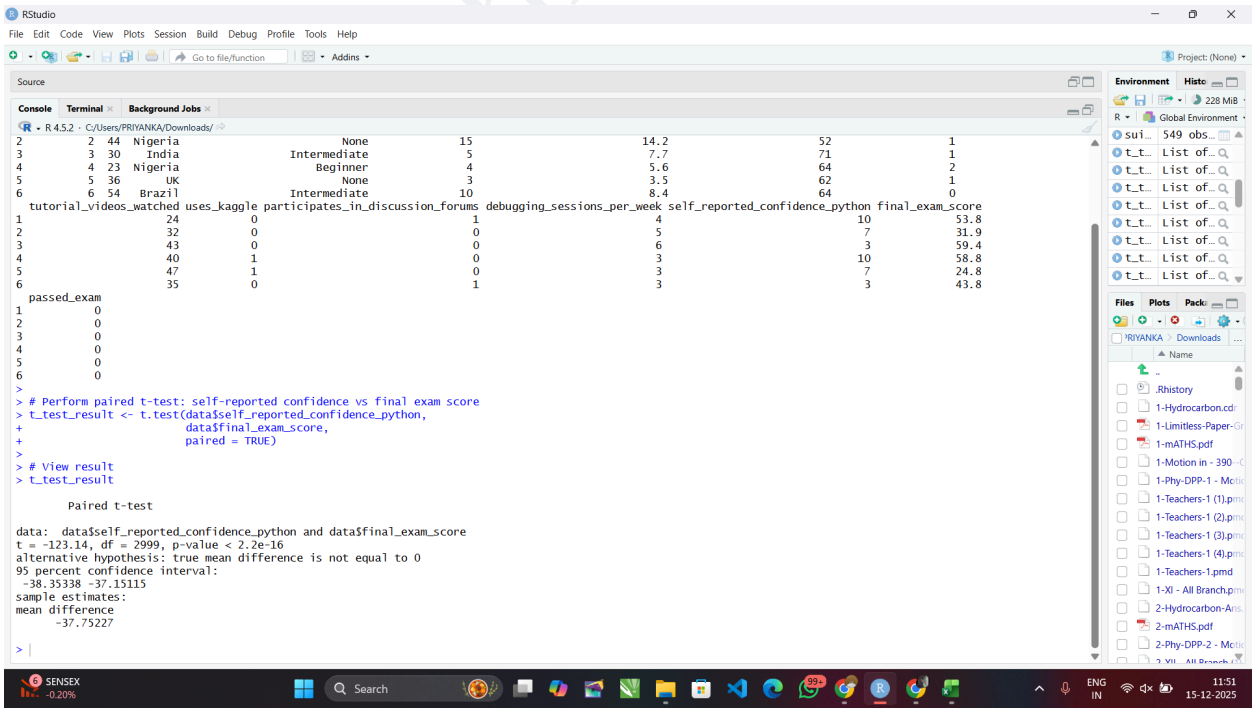
SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R

Aim: Performing paired t-tests using t.test(paired = TRUE) (R).

OUTPUT:



```
> # Load CSV
> data <- read.csv("python_learning_exam_performance.csv")
>
> # View first few rows
> head(data)
  student_id age  country prior_programming_experience weeks_in_course hours_spent_learning_per_week practice_problems_solved projects_completed
1          1  54 Pakistan                    Beginner              12                          12.2                        60                   1
2          2  44  Nigeria                        None              15                          14.2                        52                   1
3          3  30    India                Intermediate               5                           7.7                        71                   1
4          4  23  Nigeria                    Beginner               4                           5.6                        64                   2
5          5  36       UK                        None               3                           3.5                        62                   1
6          6  54   Brazil                Intermediate              10                           8.4                        64                   0
  tutorial_videos_watched uses_kaggle participates_in_discussion_forums debugging_sessions_per_week self_reported_confidence_python final_exam_score
1                      24          0                                  1                           4                           10                53.8
2                      32          0                                  0                           5                            7                31.9
3                      43          0                                  0                           6                            3                59.4
4                      40          1                                  0                           3                           10                58.8
5                      47          1                                  0                           3                            7                24.8
6                      35          0                                  1                           3                            3                43.8
  passed_exam
1           0
2           0
3           0
4           0
5           0
6           0
>
> # Perform paired t-test: self-reported confidence vs final exam score
> t_test_result <- t.test(data$self_reported_confidence_python,
+                         data$final_exam_score,
+                         paired = TRUE)
>
> # View result
> t_test_result

        Paired t-test

data:  data$self_reported_confidence_python and data$final_exam_score
t = -123.14, df = 2999, p-value < 2.2e-16
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
```



```
> # Load CSV
> data <- read.csv("python_learning_exam_performance.csv")
>
> # View first few rows
> head(data)
  student_id age  country prior_programming_experience weeks_in_course hours_spent_learning_per_week practice_problems_solved projects_completed
1          1  54 Pakistan                    Beginner              12                          12.2                        60                   1
2          2  44  Nigeria                        None              15                          14.2                        52                   1
3          3  30    India                Intermediate               5                           7.7                        71                   1
4          4  23  Nigeria                    Beginner               4                           5.6                        64                   2
5          5  36       UK                        None               3                           3.5                        62                   1
6          6  54   Brazil                Intermediate              10                           8.4                        64                   0
  tutorial_videos_watched uses_kaggle participates_in_discussion_forums debugging_sessions_per_week self_reported_confidence_python final_exam_score
1                      24          0                                  1                           4                           10                53.8
2                      32          0                                  0                           5                            7                31.9
3                      43          0                                  0                           6                            3                59.4
4                      40          1                                  0                           3                           10                58.8
5                      47          1                                  0                           3                            7                24.8
6                      35          0                                  1                           3                            3                43.8
  passed_exam
1           0
2           0
3           0
4           0
5           0
6           0
>
> # Perform paired t-test: self-reported confidence vs final exam score
> t_test_result <- t.test(data$self_reported_confidence_python,
+                         data$final_exam_score,
+                         paired = TRUE)
>
> # View result
> t_test_result

        Paired t-test

data:  data$self_reported_confidence_python and data$final_exam_score
t = -123.14, df = 2999, p-value < 2.2e-16
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -38.35338 -37.15115
sample estimates:
mean difference
      -37.75227

>
```

Krishna Nandkumar Chaudhary
S074