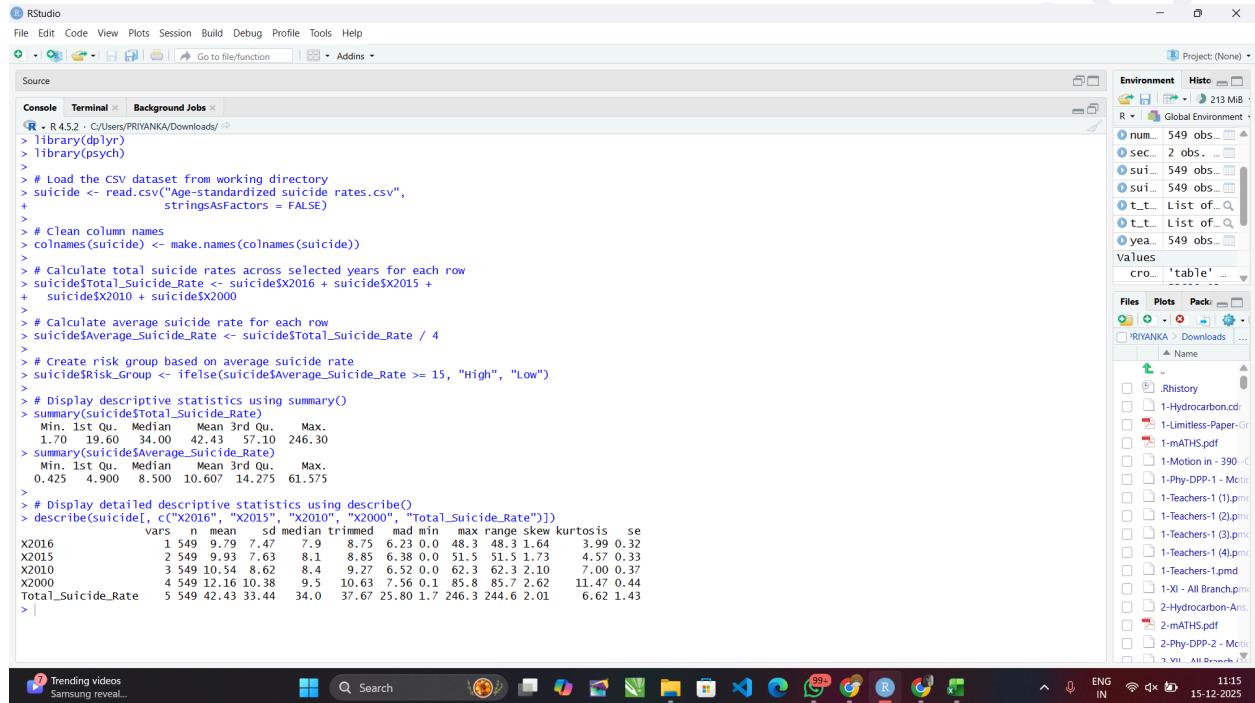


**SHETH L.U.J AND SIR M.V. COLLEGE**  
**SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R**

## Module 2 Practical 1-6

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

### OUTPUT:



The screenshot shows an RStudio interface with the following details:

- File menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Includes icons for file operations like Open, Save, Print, and a Go to file/function search bar.
- Console tab:** Displays R session history with the following code:

```
> library(dplyr)
> library(psych)
>
> # Load the CSV dataset from working directory
> suicide <- read.csv("age-standardized suicide rates.csv",
+                      stringsAsFactors = FALSE)
>
> # Clean column names
> colnames(suicide) <- make.names(colnames(suicide))
>
> # Calculate total suicide rates across selected years for each row
> suicide$total_Suicide_Rate <- suicide$X2016 + suicide$X2015 +
+   suicide$X2010 + suicide$X2000
>
> # Calculate average suicide rate for each row
> suicide$Average_Suicide_Rate <- suicide$total_Suicide_Rate / 4
>
> # Create risk group based on average suicide rate
> suicide$Risk_Group <- ifelse(suicide$Average_Suicide_Rate >= 15, "High", "Low")
>
> # Display descriptive statistics using summary()
> summary(suicide$Total_Suicide_Rate)
  Min. 1st Qu. Median Mean 3rd Qu. Max.
  1.70 19.60 34.00 42.43 57.10 246.30
> summary(suicide$Average_Suicide_Rate)
  Min. 1st Qu. Median Mean 3rd Qu. Max.
  0.425 4.900 8.500 10.607 14.275 61.575
>
> # Display detailed descriptive statistics using describe()
> describe(suicide[, c("X2016", "X2015", "X2010", "X2000", "Total_Suicide_Rate")])
   vars n mean sd median trimmed mad min max range skew kurtosis se
X2016    1 549  9.79  7.43   7.9  8.75  6.23  0.0  48.3  48.3  1.64   3.99  0.32
X2015    2 549  9.93  7.63   8.1  8.85  6.30  0.0  51.5  51.5  1.73   4.57  0.33
X2010    3 549 10.54  8.62   8.4  9.25  7.52  0.1  62.3  62.3  2.10   7.00  0.37
X2000    4 549 12.16 10.38   9.5 10.63  8.58  0.1  85.8  85.7  2.62  11.47  0.44
Total_Suicide_Rate 5 549 42.43 33.44  34.0 37.67 25.80  1.7 246.3 244.6  2.01   6.62  1.43
```
- Environment tab:** Shows the global environment with objects like R (213 MB), num\_ (549 obs.), sec\_ (2 obs.), sui\_ (549 obs.), sui\_ (549 obs.), t.t\_ (List of 1), t.t\_ (List of 1), yea\_ (549 obs.), Values (cro\_ 'table'), and Jhistory.
- Files tab:** Shows a file tree with various PDF and PPT files.
- Plots tab:** Shows a small preview of a plot.
- Packs tab:** Shows a list of packages.
- Bottom status bar:** Shows system information: Trending videos, Samsung reveal..., Search, ENG IN, 11:15, 15-12-2025.

# 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:

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source Terminal Background Jobs
R 4.5.2 · C:/Users/PRIYANKA/Downloads/
> library(dplyr)
>
> data <- read.csv("Business_sales_EDA.csv", stringsAsFactors = FALSE, sep = ";")
>
> table(data$Product.Position)
    Aisle      End-cap Front of Store
    7810       6791      5651
> table(data$Promotion)
    No Yes
11812 8440
> table(data$Product.Category)
clothing
20252
> table(data$Seasonal)
    No Yes
10136 10116
> table(data$brand)
Zara
20252
> table(data$section)
MAN WOMAN
6998 13254
> table(data$season)
Autumn Spring Summer Winter
7665 4537 2906 5144
> table(data$material)
Acrylic Cotton Denim Linen Linen Blend Polyester Satin Silk Viscose Wool Wool Blend
881 3851 1027 2573 807 2775 132 38 990 3805 3373
> table(data$origin)
Argentina Bangladesh Brazil Cambodia China India Morocco Pakistan Portugal Spain Turkey Vietnam
179 3617 795 981 4026 2033 1653 605 1420 1248 2475 1220

```

```

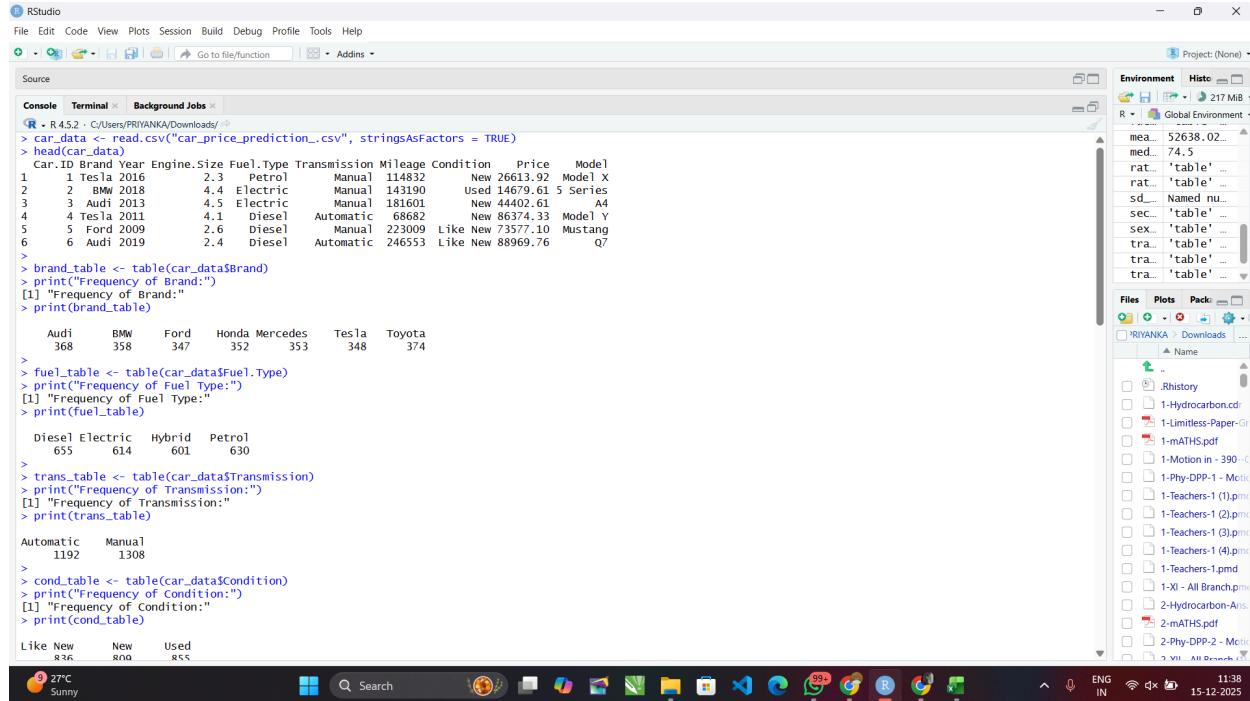
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source Terminal Background Jobs
R 4.5.2 · C:/Users/PRIYANKA/Downloads/
Argentina Bangladesh Brazil Cambodia China India Morocco Pakistan Portugal Spain Turkey Vietnam
179 3617 795 981 4026 2033 1653 605 1420 1248 2475 1220
>
> data %>% count(Product.Position)
Product.Position n
1 Aisle 7810
2 End-cap 6791
3 Front of Store 5651
> data %>% count(Promotion)
Promotion n
1 No 11812
2 Yes 8440
> data %>% count(Product.Category)
Product.Category n
1 clothing 20252
> data %>% count(Seasonal)
Seasonal n
1 No 10136
2 Yes 10116
> data %>% count(brand)
brand n
1 Zara 20252
> data %>% count(section)
section n
1 MAN 6998
2 WOMAN 13254
> data %>% count(season)
season n
1 Autumn 7665
2 Spring 4537
3 Summer 2906
4 Winter 5144
> data %>% count(material)
material n
1 Acrylic 881
2 Cotton 3851
3 Denim 1027
4 Linen 2573
5 Linen Blend 807
6 Polyester 2775

```

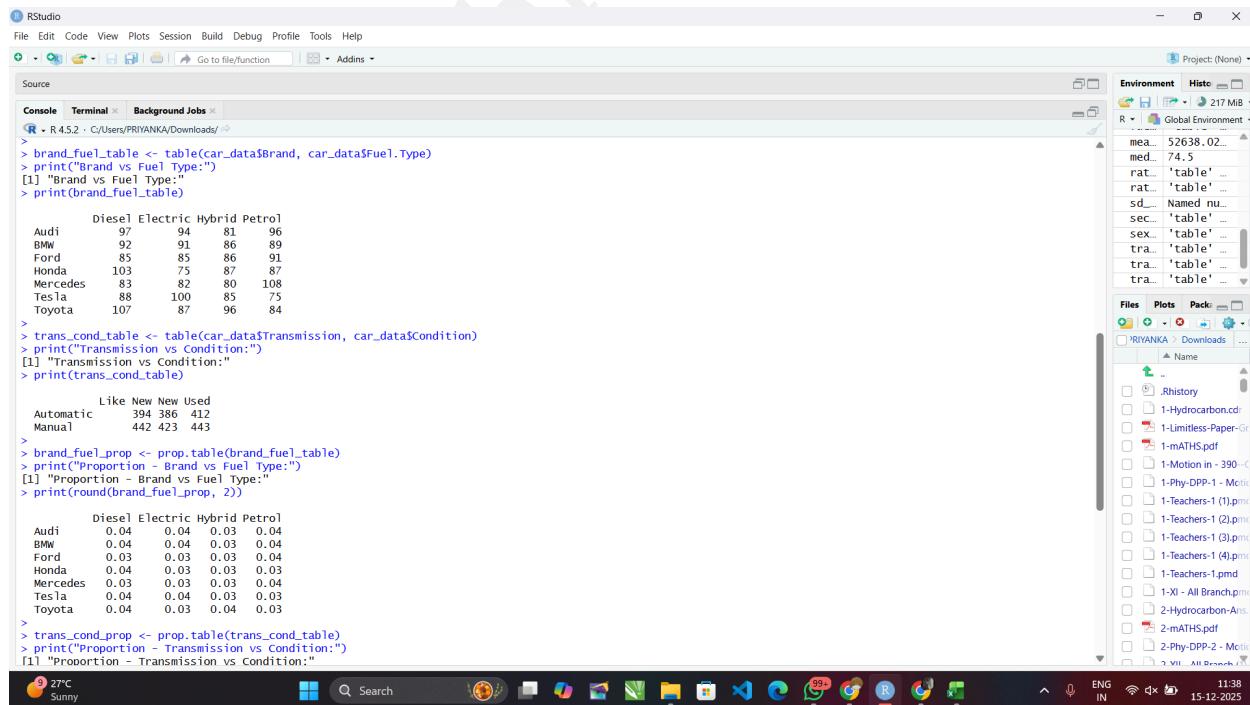
**SHETH L.U.J AND SIR M.V. COLLEGE**  
**SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R**

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

## OUTPUT:

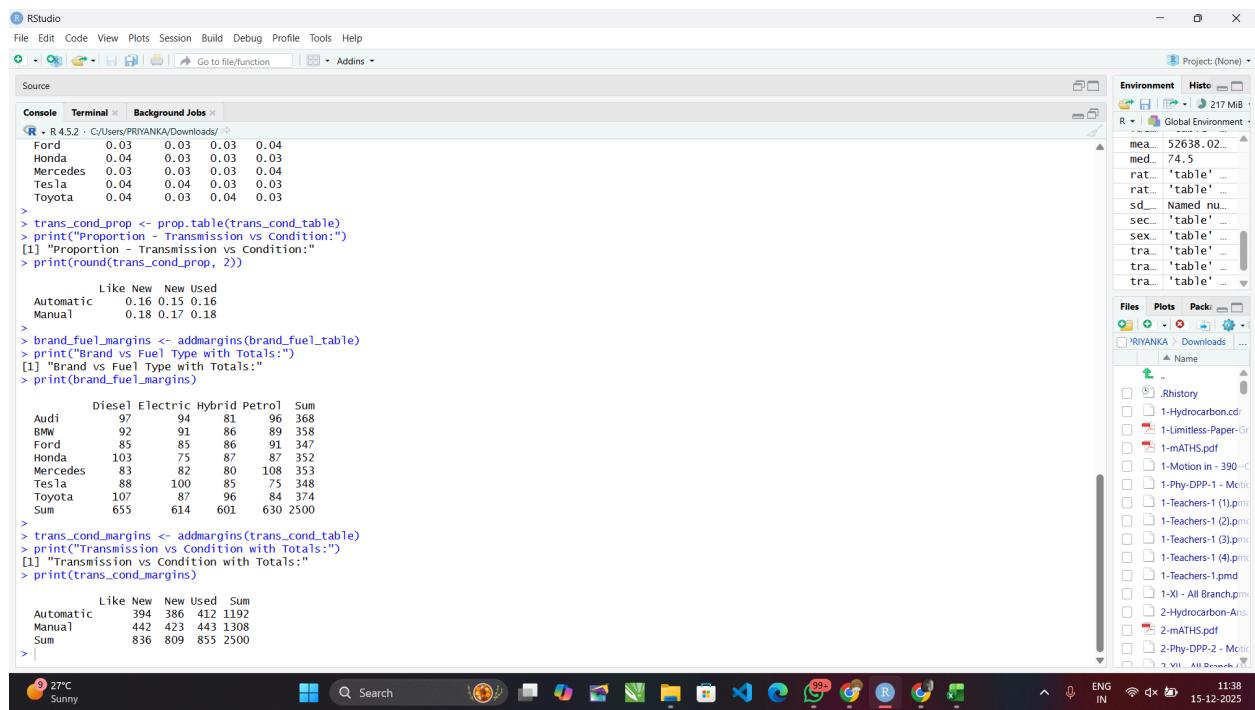


```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source Environment Histograms Global Environment
Console Terminal Background Jobs
[R - R 4.5.2 - C:/Users/PRIYANKA/Downloads/]
> car_data <- read.csv("car_price_prediction_.csv", stringsAsFactors = TRUE)
> head(car_data)
  Car.ID 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 Used 
  836  800 855
27°C Sunny
ENG IN 11:38 15-12-2025
```



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source Environment Histograms Global Environment
Console Terminal Background Jobs
[R - R 4.5.2 - C:/Users/PRIYANKA/Downloads/]
>
> 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     87     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 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:"
> print(round(trans_cond_prop, 2))
```

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



The screenshot shows the RStudio interface with the following R code in the Console tab:

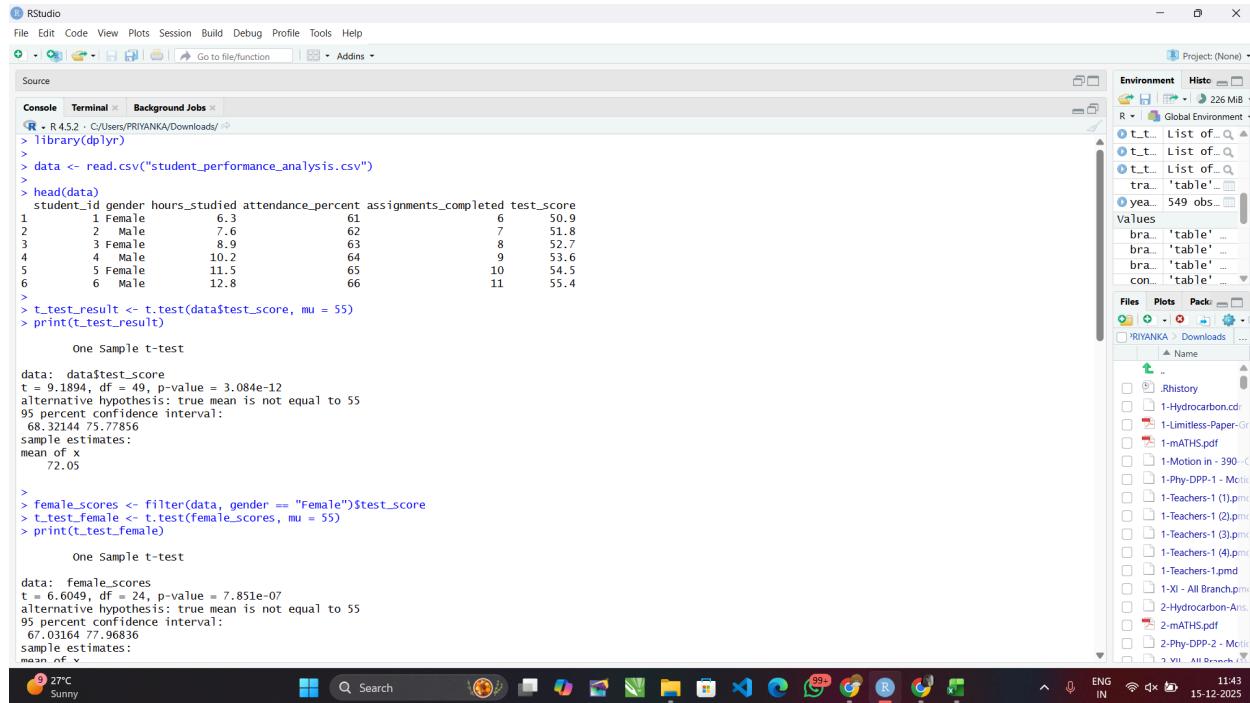
```
R - R 4.5.2 - C:/users/PRIYANKA/Downloads/~/  
> Ford      0.03   0.03   0.03   0.04  
Honda     0.04   0.03   0.03   0.03  
Mercedes  0.03   0.03   0.03   0.04  
Tesla     0.04   0.04   0.03   0.03  
Toyota    0.04   0.03   0.04   0.03  
>  
> trans_cond_prop <- prop.table(trans_cond_table)  
> print("Proportion - Transmission vs Condition:")  
[1] "Proportion - Transmission vs Condition:"  
> print(round(trans_cond_prop, 2))  
          Like New New Used  
Automatic  0.16 0.15 0.16  
Manual     0.18 0.17 0.18  
>  
> brand_fuel_margins <- addmargins(brand_fuel_table)  
> print("Brand vs Fuel Type with Totals:")  
[1] "Brand vs Fuel Type with Totals."  
> print(brand_fuel_margins)  
          Diesel Electric Hybrid Petrol Sum  
Audi       97     94     81     96   368  
BMW        92     91     86     89   358  
Ford       85     85     86     91   347  
Honda     103     75     87     87   352  
Mercedes   83     82     80     108  353  
Tesla      88     100    85     75   348  
Toyota     107     87     96     84   374  
Sum       655    614    601    630  2500  
>  
> trans_cond_margins <- addmargins(trans_cond_table)  
> print("Transmission vs Condition with Totals:")  
[1] "Transmission vs Condition with Totals."  
> print(trans_cond_margins)  
          Like New New Used Sum  
Automatic  394   386   412  1192  
Manual     442   423   443  1308  
Sum       836   809   855  2500  
>
```

The Environment pane shows various objects defined in the session, such as mea\_, med\_, rat\_, sd\_, sec\_, sex\_, tra\_, and tra\_. The Files pane shows a directory structure under 'PRIYANKA' containing various files like 'Rhistory', '1-Hydrocarbon.pdf', '1-Limits-Paper-Gr...', '1-MATHS.pdf', etc.

**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:



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source Console Terminal Background Jobs
R > R 4.5.2 - C:/Users/PRIYANKA/Downloads/
> 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)

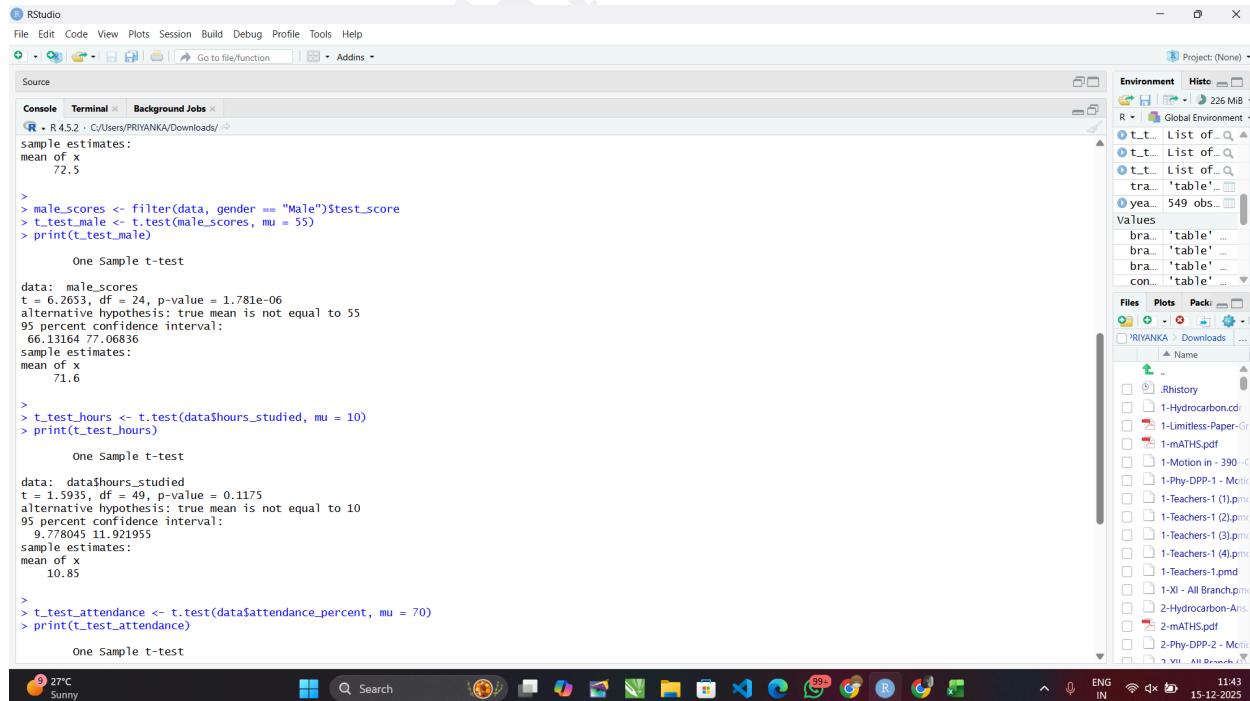
One Sample t-test

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:
mean of x
```



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source Console Terminal Background Jobs
R > R 4.5.2 - C:/Users/PRIYANKA/Downloads/
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)

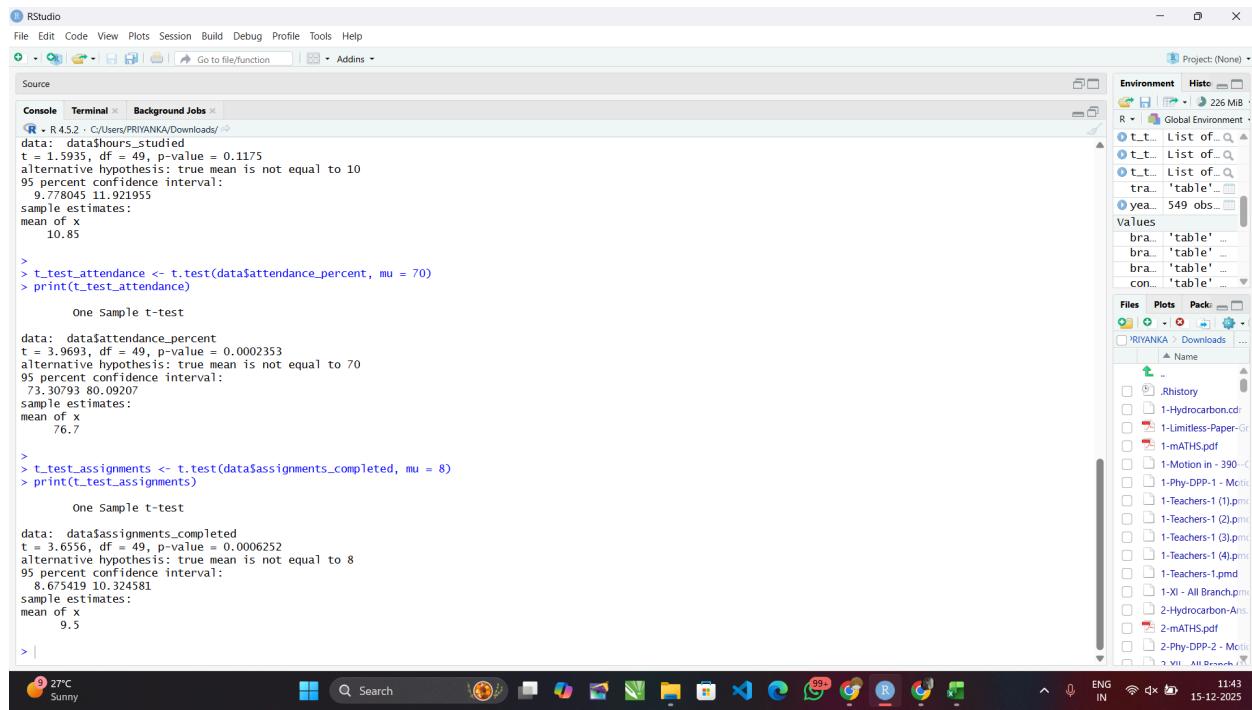
One Sample t-test

data: data$hours_studied
t = 1.5393, df = 49, p-value = 0.1175
alternative hypothesis: true mean is not equal to 10
95 percent confidence interval:
5.778045 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
```

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



The screenshot shows the RStudio interface with the following details:

- Console Tab:** Displays R code and its output. The code performs three t-tests on dataframes: `data$hours_studied`, `data$attendance_percent`, and `data$assignments_completed`, comparing them against hypothesized means of 10, 70, and 8 respectively.
- Environment Tab:** Shows the global environment with objects like `t.t...`, `yea...`, and `Values`.
- Files Tab:** Shows a file tree under the path `PRIYANKA > Downloads`, containing various PDF files related to hydrocarbons and teachers.
- Plots Tab:** Shows a histogram titled "225 MB".
- Bottom Bar:** Includes system icons for weather (27°C, Sunny), search, taskbar, and system status (ENG IN, 11:43, 15-12-2025).

# 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:

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source Terminal Background Jobs
R - R 4.5.2 - C:/Users/PRIYANKA/Downloads/
> # Load data
> data <- read.csv("Food_Delivery_Route_Efficiency_Dataset.csv")
>
> # Convert traffic_level to factor
> data$traffic_level <- factor(data$traffic_level, levels = c("Low", "Medium", "High"))
>
> # Independent two-sample t-test: High vs Low traffic
> high_low <- subset(data, traffic_level %in% c("High", "Low"))
> t.test(delivery_time_min ~ traffic_level, data = high_low)

Welch Two Sample t-test

data: delivery_time_min by traffic_level
t = 0.32837, df = 130.4, p-value = 0.7432
alternative hypothesis: true difference in means between group Low and group High is not equal to 0
95 percent confidence interval:
-7.178204 10.035353
sample estimates:
mean in group Low mean in group High
47.04769      45.61912

> # Independent two-sample t-test: Medium vs Low traffic
> medium_low <- subset(data, traffic_level %in% c("Medium", "Low"))
> t.test(delivery_time_min ~ traffic_level, data = medium_low)

Welch Two Sample t-test

data: delivery_time_min by traffic_level
t = 1.2332, df = 129.82, p-value = 0.2197
alternative hypothesis: true difference in means between group Low and group Medium is not equal to 0
95 percent confidence interval:
-3.278418 14.129027
sample estimates:
mean in group Low mean in group Medium
47.04769      41.62239

> # Independent two-sample t-test: High vs Medium traffic
> high_medium <- subset(data, traffic_level %in% c("High", "Medium"))
> t.test(delivery_time_min ~ traffic_level, data = high_medium)

```

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source Terminal Background Jobs
R - R 4.5.2 - C:/Users/PRIYANKA/Downloads/
alternative hypothesis: true difference in means between group Low and group High is not equal to 0
95 percent confidence interval:
-7.178204 10.035353
sample estimates:
mean in group Low mean in group High
47.04769      45.61912

> # Independent two-sample t-test: Medium vs Low traffic
> medium_low <- subset(data, traffic_level %in% c("Medium", "Low"))
> t.test(delivery_time_min ~ traffic_level, data = medium_low)

Welch Two Sample t-test

data: delivery_time_min by traffic_level
t = 1.2332, df = 129.82, p-value = 0.2197
alternative hypothesis: true difference in means between group Low and group Medium is not equal to 0
95 percent confidence interval:
-3.278418 14.129027
sample estimates:
mean in group Low mean in group Medium
47.04769      41.62239

> # Independent two-sample t-test: High vs Medium traffic
> high_medium <- subset(data, traffic_level %in% c("High", "Medium"))
> t.test(delivery_time_min ~ traffic_level, data = high_medium)

Welch Two Sample t-test

data: delivery_time_min by traffic_level
t = -0.92924, df = 132.88, p-value = 0.3544
alternative hypothesis: true difference in means between group Medium and group High is not equal to 0
95 percent confidence interval:
-32.504145 4.510686
sample estimates:
mean in group Medium mean in group High
41.62239      45.61912

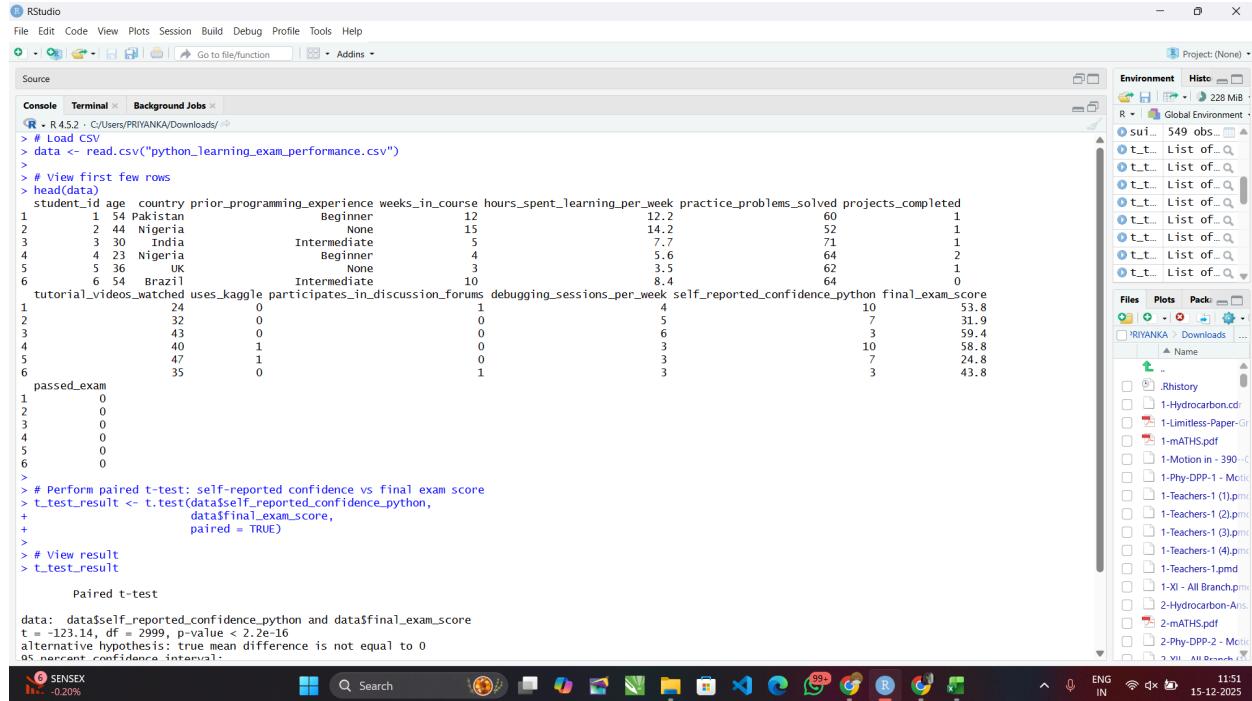
```

# 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).

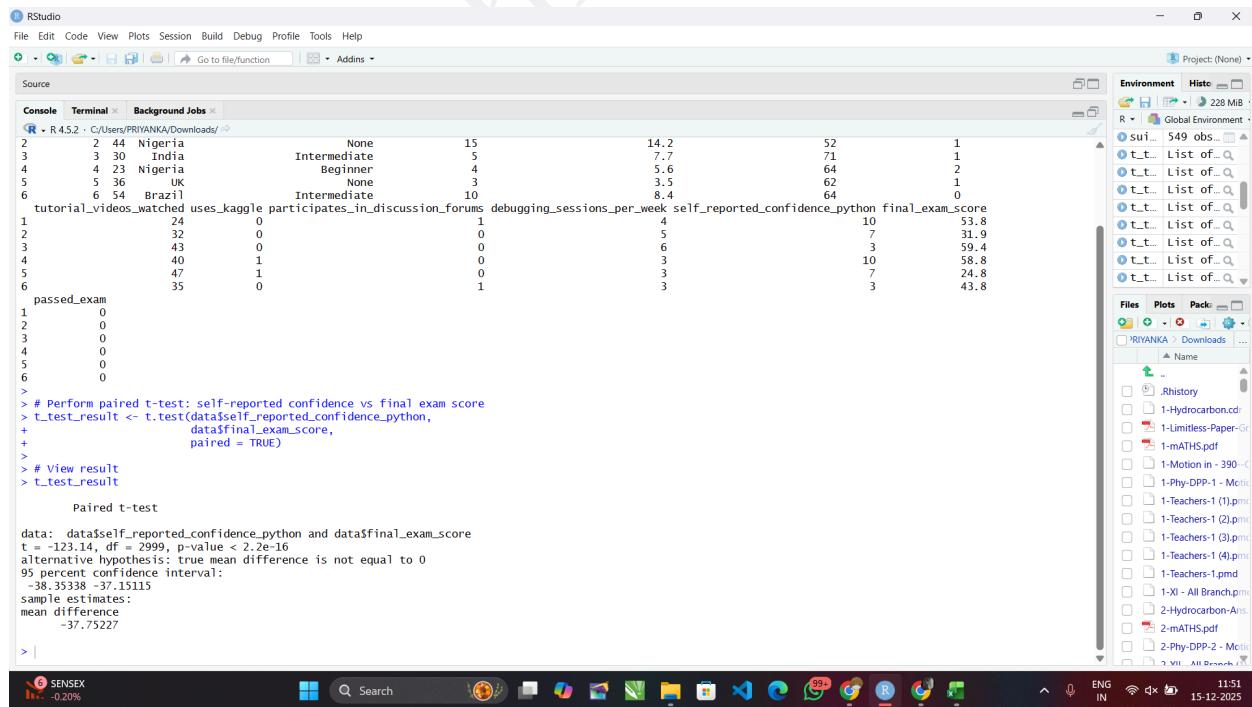
### OUTPUT:



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
[R - R 4.5.2 - C:/Users/PRIYANKA/Downloads/]
> # 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
> |
```

The screenshot shows the RStudio interface with the console tab active. The code in the console reads a CSV file named "python\_learning\_exam\_performance.csv" and displays its first six rows. It then performs a paired t-test comparing the "self\_reported\_confidence\_python" and "final\_exam\_score" columns. The output shows a t-value of -123.14, a degrees of freedom (df) of 2999, and a p-value less than 2.2e-16, indicating a significant difference. The alternative hypothesis is that the true mean difference is not equal to zero. The 95% confidence interval is given as [-38.35338, -37.15115]. The sample estimate for the mean difference is -37.75227.



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
[R - R 4.5.2 - C:/Users/PRIYANKA/Downloads/]
> # 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
> |
```

This screenshot is identical to the one above, showing the RStudio interface with the console tab active. It displays the same code and t-test results, indicating no changes between the two screenshots.