

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

PRACTICAL NO: 1

Generating descriptive statistics using summary () or describe()(R)

CODE:

```
#
=====
=====
# R Script: Descriptive Statistics
# Functions: summary(), describe()
# Dataset: globalAirQuality.csv
#
=====
=====

# -----
# 1. Install Required Packages
# -----
install.packages("psych")
install.packages("dplyr")

# -----
# 2. Load Libraries
# -----
library(psych)
library(dplyr)

# -----
# 3. Set Working Directory (CHANGE PATH IF NEEDED)
# Use forward slashes to avoid 'U' error

# -----
# 4. Import Dataset

air <- read.csv("C:/Users/vibro/Downloads/globalAirQuality.csv")

# -----
# 5. View Dataset Structure
# -----
str(air)

# -----
# 6. Descriptive Statistics using summary()
```

RIYA RANE
S107 SYCS

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

```
# -----  
summary(air)  
  
# -----  
# 7. Select Only Numeric Columns  
# -----  
air_numeric <- air %>% select(where(is.numeric))  
  
# -----  
# 8. Descriptive Statistics (Numeric Data Only)  
# -----  
summary(air_numeric)  
  
# -----  
# 9. Detailed Descriptive Statistics using describe()  
# -----  
describe(air_numeric)
```

OUTPUT:

The screenshot shows the RStudio interface. The console on the left displays the following code and output:

```
> air <- read.csv("C:/Users/vibro/downloads/globalAirQuality.csv")  
Error: '\U' used without hex digits in character string  
(<input>:1:21)  
  
> # 4. Import Dataset  
>  
> air <- read.csv("C:/Users/vibro/downloads/globalAirQuality.csv")  
> # 5. View Dataset Structure  
> # -----  
> str(air)  
'data.frame': 18000 obs. of 15 variables:  
 $ timestamp : chr "2025-11-04 18:25:17.554219" "2025-11-04 19:25:17.554219" "2025-11-04 20:25:17.554219" "2025-11-04 21:25:17.554219" ...  
 $ country : chr "US" "US" "US" "US" ...  
 $ city : chr "New York" "New York" "New York" "New York" ...  
 $ latitude : num 40.7 40.7 40.7 40.7 ...  
 $ longitude : num -74 -74 -74 -74 ...  
 $ pm25 : num 50.3 32.1 42.2 30.4 21.1 ...  
 $ pm10 : num 108.9 63.82 6.80 66.4 ...  
 $ no2 : num 28.36 1.26 9.63 5.39 ...  
 $ so2 : num 6.54 4.02 9.54 7.61 6.92 ...  
 $ o3 : num 52.6 43.5 23.3 31.4 45.6 ...  
 $ co : num 1.096 1.075 0.977 0.23 1.085 ...  
 $ aqi : int 108 90 84 158 97 92 155 115 121 76 ...  
 $ temperature : num 18.5 5.84 31.83 23.14 13.63 ...  
 $ humidity : num 70.2 80.1 62.8 89.2 76.5 ...  
 $ wind_speed : num 3.73 8.97 9.65 8.96 4.02 ...  
> # 6. Descriptive Statistics using summary()  
> # -----  
> summary(air)
```

The Environment pane on the right shows the following objects:

Object	Size	Modified
air	18000 obs. of 15 variables	
air_numeric	18000 obs. of 12 variables	
cat_filter	3891 obs. of 18 variables	
furniture_high	460 obs. of 18 variables	
globalAirQuality	18001 obs. of 15 variables	
high sales	1146 obs. of 18 variables	

The Files pane shows the following files:

Name	Size	Modified
.RData	1.2 MB	Dec 1, 2025, 7:20 PM
.Rhistory	3 KB	Dec 1, 2025, 7:20 PM
Custom Office Templates		
desktop.ini	418 B	Dec 8, 2025, 10:41 AM
JAVA PROJECT by S094 and S107 (Shivam Mane and)	11 MB	Oct 3, 2025, 6:52 PM
java project clip 1.mp4	5.4 MB	Oct 3, 2025, 7:36 PM
java project clip 2.mp4	7.2 MB	Oct 3, 2025, 7:36 PM
NetBeansProjects		
prac4r.R	799 B	Dec 1, 2025, 6:40 PM
prac8 java - Construct a GUI using Java Swing to acc 0 B		Sep 11, 2025, 7:08 PM
raneS107		
UltraViewer		

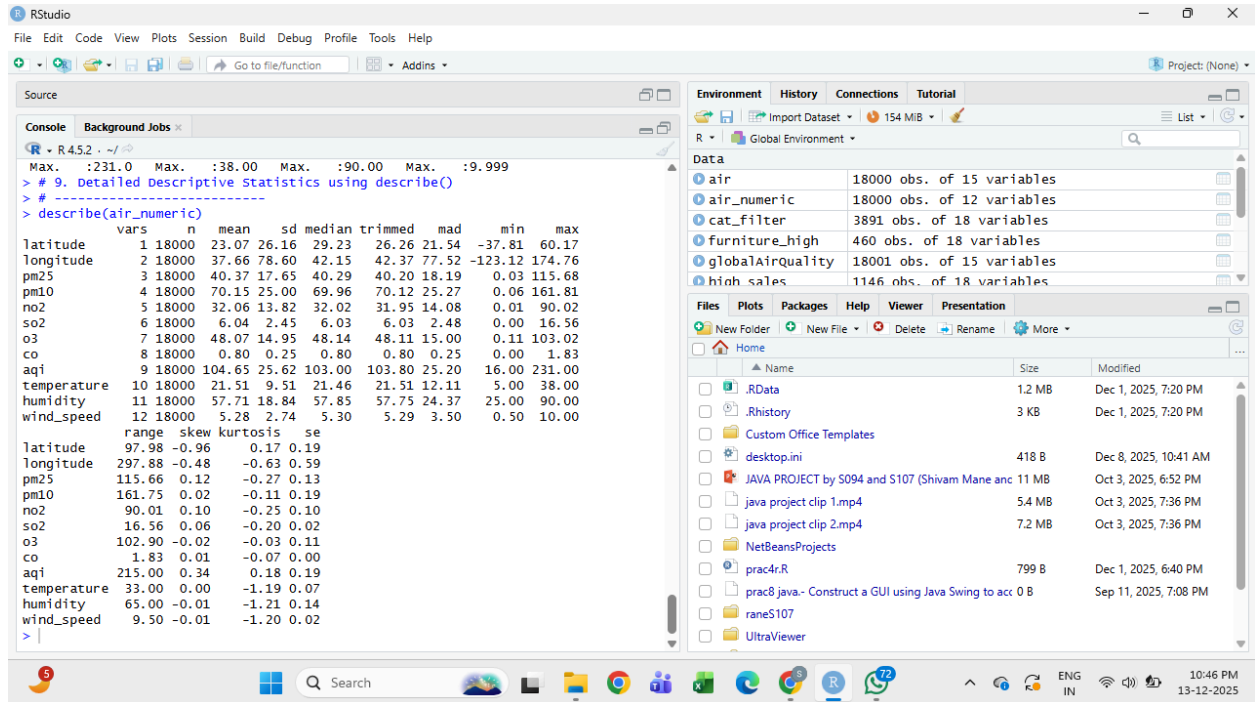
RIYA RANE
S107 SYCS

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

RStudio interface showing the console output of the `summary(air)` command. The output displays summary statistics for various variables in the `air` dataset, including `timestamp`, `country`, `city`, `latitude`, `longitude`, `pm25`, `pm10`, `no2`, `so2`, `o3`, `co`, `aqi`, `temperature`, `humidity`, and `wind_speed`. The Environment pane on the right shows the loaded datasets: `air` (18000 obs. of 15 variables), `air_numeric` (18000 obs. of 12 variables), `cat_filter` (3891 obs. of 18 variables), `furniture_high` (460 obs. of 18 variables), `globalAirQuality` (18001 obs. of 15 variables), and `hinh sales` (1146 obs. of 18 variables).

RStudio interface showing the console output of the `summary(air_numeric)` command. The output displays detailed summary statistics for the `air_numeric` dataset, including `latitude`, `longitude`, `pm25`, `pm10`, `no2`, `so2`, `o3`, `co`, `aqi`, `temperature`, `humidity`, and `wind_speed`. The Environment pane on the right shows the loaded datasets: `air` (18000 obs. of 15 variables), `air_numeric` (18000 obs. of 12 variables), `cat_filter` (3891 obs. of 18 variables), `furniture_high` (460 obs. of 18 variables), `globalAirQuality` (18001 obs. of 15 variables), and `hinh sales` (1146 obs. of 18 variables).

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



PRACTICAL NO:2

GENERATING FREQUENCY TABLE USING TABLE() OR COUNT().

CODE:

```
#  
=====
```

```
# R Script: Frequency Table Generation  
# Functions Used: table(), count()  
# Dataset: car_price_prediction_.csv  
#  
=====
```

```
# -----  
# 1. Install Required Package (Run Once)  
# -----  
install.packages("dplyr")  
  
# -----  
# 2. Load Library (Must Run Every Time)  
# -----  
library(dplyr)
```

RIYA RANE
S107 SYCS

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

```
# -----
# 3. Import Dataset
# Use forward slashes to avoid \"U\" error
# -----
car <- read.csv("C:/Users/vibro/Downloads/car_price_prediction_.csv")

# -----
# 4. Clean Column Names (IMPORTANT)
# Converts names like "Fuel Type" -> Fuel_Type
# -----
names(car) <- make.names(names(car))

# -----
# 5. Verify Column Names
# -----
names(car)

#
=====
=====
# FREQUENCY TABLE USING table() (Base R)
#
=====
=====

table(car$Fuel_Type)
table(car$Transmission)
table(car$Owner)

#
=====
=====
# FREQUENCY TABLE USING count() (dplyr)
#
=====
=====

count(car, Fuel_Type)
count(car, Transmission)
count(car, Owner)
```

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

#

=====

=====

SORTED FREQUENCY TABLE (Descending Order)

#

=====

=====

```
count(car, Fuel_Type, sort = TRUE)
count(car, Transmission, sort = TRUE)
count(car, Owner, sort = TRUE)
```

OUTPUT:

The screenshot displays the RStudio environment with the following components:

- Source Editor:** Contains R code for importing a dataset, cleaning column names, and generating frequency tables.
- Console:** Shows the execution of the R code, including the output of the `names(car)` function.
- Environment:** Lists the objects in the global environment, including `air`, `air_numeric`, `car`, `car_price_predic...`, `cat_filter`, and `furniture_high`.
- Files:** Shows a file explorer view of the user's home directory, listing various files and folders.

R Code in Source Editor:

```
> # 3. Import Dataset
> # Use forward slashes to avoid '\u' error
> # -----
> car <- read.csv("C:/Users/vibro/downloads/car_price_prediction_.csv")
> # -----
> # 4. Clean Column Names (IMPORTANT)
> # Converts names like "Fuel Type" -> Fuel_Type
> # -----
> names(car) <- make.names(names(car))
> # -----
> # 5. Verify Column Names
> # -----
> names(car)
[1] "Car.ID"      "Brand"      "Year"      "Engine.Size"
[5] "Fuel.Type"   "Transmission" "Mileage"    "Condition"
[9] "Price"      "Model"
> # -----
> # FREQUENCY TABLE USING table() (Base R)
> # -----
> # -----
> table(car$Fuel_Type)
< table of extent 0 >
> table(car$Transmission)
```

Environment Panel:

Object	Size
air	18000 obs. of 15 variables
air_numeric	18000 obs. of 12 variables
car	2500 obs. of 10 variables
car_price_predic...	2500 obs. of 10 variables
cat_filter	3891 obs. of 18 variables
furniture_high	460 obs. of 18 variables

Files Panel:

Name	Size	Modified
.RData	1.2 MB	Dec 1, 2025, 7:20 PM
.Rhistory	3 KB	Dec 1, 2025, 7:20 PM
Custom Office Templates		
desktop.ini	418 B	Dec 8, 2025, 10:41 AM
JAVA PROJECT by S094 and S107 (Shivam Mane anc	11 MB	Oct 3, 2025, 6:52 PM
java project clip 1.mp4	5.4 MB	Oct 3, 2025, 7:36 PM
java project clip 2.mp4	7.2 MB	Oct 3, 2025, 7:36 PM
NetBeansProjects		
prac4r.R	799 B	Dec 1, 2025, 6:40 PM
prac8 java.- Construct a GUI using Java Swing to acc 0 B		Sep 11, 2025, 7:08 PM
raneS107		
UltraViewer		

RIYA RANE
S107 SYCS

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

```
table(car$Fuel_Type)
table of extent 0 >
table(car$Transmission)

automatic    Manual
      1192      1308
table(car$Owner)
table of extent 0 >
# =====
# FREQUENCY TABLE USING count() (dplyr)
# =====
count(car, Fuel_Type)

Error in `count()`:
! Must group by variables found in `.data`.
✖ Column `Fuel_Type` is not found.
Run `rlang::last_trace()` to see where the error occurred.
```

PRACTICAL NO: 3

CREATING CROSS TABULATION AND TWO WAY TABLES USING TABLES()(R)

CODE:

```
#
=====
# R Script: Cross Tabulation using table()
#
=====

# 1. Read the dataset
data <- read.csv("C:/Users/vibro/Downloads/data_date.csv", stringsAsFactors = FALSE)

# 2. Check column names
names(data)

# 3. Create Two-Way Cross Tabulation
# Replace Column1 and Column2 with actual column names
cross_tab <- table(data$Column1, data$Column2)
```

RIYA RANE
S107 SYCS

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

4. Display Cross Tabulation

cross_tab

#

=====

=====

5. Row-wise Percentages

row_percent <- prop.table(cross_tab, 1) * 100

row_percent

#

=====

=====

6. Column-wise Percentages

col_percent <- prop.table(cross_tab, 2) * 100

col_percent

#

=====

=====

7. Overall Percentages

overall_percent <- prop.table(cross_tab) * 100

overall_percent

#

=====

=====

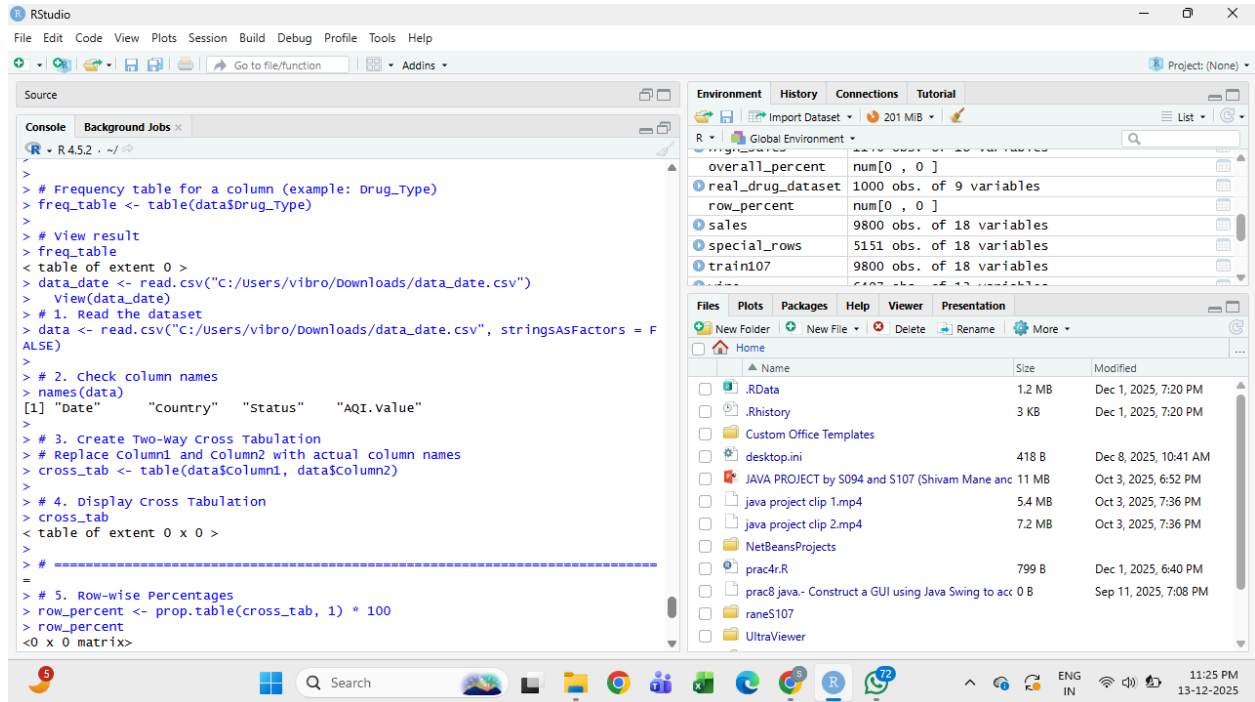
8. Cross Tabulation Including Missing Values

cross_tab_na <- table(data\$Column1, data\$Column2, useNA = "ifany")

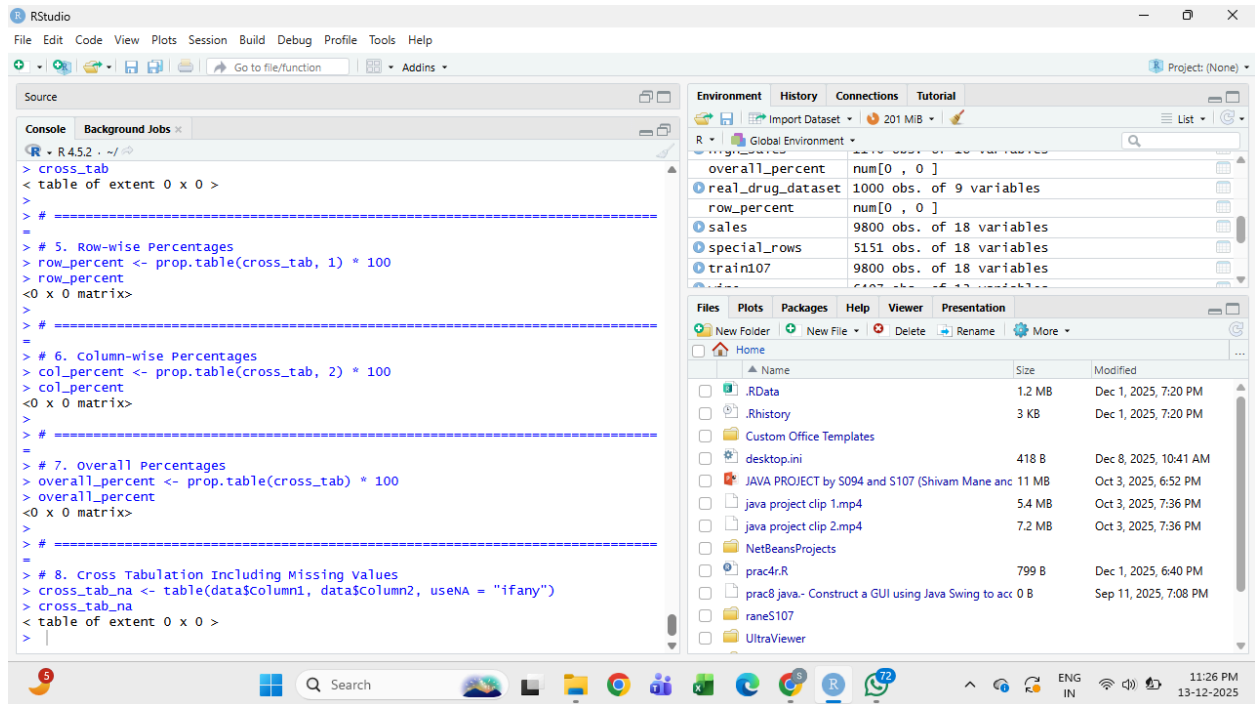
cross_tab_na

OUTPUT:

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



```
> # Frequency table for a column (example: Drug_Type)
> freq_table <- table(data$Drug_Type)
>
> # view result
> freq_table
<table of extent 0 >
> data_date <- read.csv("c:/users/vibro/downloads/data_date.csv")
> view(data_date)
> # 1. Read the dataset
> data <- read.csv("C:/Users/vibro/Downloads/data_date.csv", stringsAsFactors = F
ALSE)
>
> # 2. Check column names
> names(data)
[1] "Date"      "country"   "Status"    "AQI.Value"
>
> # 3. Create Two-way Cross Tabulation
> # Replace column1 and column2 with actual column names
> cross_tab <- table(data$column1, data$column2)
>
> # 4. Display Cross Tabulation
> cross_tab
<table of extent 0 x 0 >
>
> # =====
>
> # 5. Row-wise Percentages
> row_percent <- prop.table(cross_tab, 1) * 100
> row_percent
<0 x 0 matrix>
```



```
> # =====
>
> # 5. Row-wise Percentages
> row_percent <- prop.table(cross_tab, 1) * 100
> row_percent
<0 x 0 matrix>
>
> # =====
>
> # 6. column-wise Percentages
> col_percent <- prop.table(cross_tab, 2) * 100
> col_percent
<0 x 0 matrix>
>
> # =====
>
> # 7. overall Percentages
> overall_percent <- prop.table(cross_tab) * 100
> overall_percent
<0 x 0 matrix>
>
> # =====
>
> # 8. Cross Tabulation Including Missing values
> cross_tab_na <- table(data$column1, data$column2, useNA = "ifany")
> cross_tab_na
<table of extent 0 x 0 >
>
> |
```

PRACTICAL NO: 4

PERFORMING ONE SAMPLE T-TESTS USING T.TESTS()..GIVE CODE

RIYA RANE
S107 SYCS

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

CODE:

```
# -----  
# ONE-SAMPLE T-TEST IN R  
# -----  
  
# Clear workspace  
rm(list = ls())  
  
# -----  
# STEP 1: Load required libraries  
# -----  
# (No extra packages required for t.test)  
  
# -----  
# STEP 2: Import data (CSV file)  
# -----  
# Replace file path and column name as needed  
data <- read.csv("C:/Users/vibro/Downloads/HR_Analytics.csv.csv")  
  
# View structure  
str(data)  
summary(data)  
  
# -----  
# STEP 3: Select the variable for test  
# -----  
# Example: Monthly Income  
x <- data$MonthlyIncome  
  
# Remove missing values  
x <- na.omit(x)  
  
# -----  
  
# -----  
# STEP 5: One-Sample t-Test  
# -----  
# H0: Mean Monthly Income = 6500  
  
t_test_re
```

OUTPUT:

RIYA RANE
S107 SYCS

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

The screenshot shows the RStudio interface with the following components:

- Source Panel:** Contains R code for importing a CSV file and viewing its structure.

```
> # STEP 2: Import data (csv file)
> # Replace file path and column name as needed
> data <- read.csv("C:/Users/vibro/Downloads/HRE_Analytics.csv.csv")
>
> # View structure
> str(data)
'data.frame': 1470 obs. of 35 variables:
 $ Age: int 41 49 37 33 27 32 59 30 38 36 ...
 $ Attrition: chr "Yes" "No" "Yes" "No" ...
 $ BusinessTravel: chr "Travel_Rarely" "Travel_Frequently" "Travel_Rarely" ...
 $ DailyRate: int 1102 279 1373 1392 591 1005 1324 1358 216 1299 ...
 $ Department: chr "Sales" "Research & Development" "Research & Development" ...
 $ DistanceFromHome: int 1 8 2 3 2 2 3 24 23 27 ...
 $ Education: int 2 1 2 4 1 2 3 1 3 3 ...
 $ EducationField: chr "Life Sciences" "Life Sciences" "other" "Life Sciences" ...
 $ EmployeeCount: int 1 1 1 1 1 1 1 1 1 ...
 $ EmployeeNumber: int 1 2 4 5 7 8 10 11 12 13 ...
 $ EnvironmentSatisfaction: int 2 3 4 4 1 4 3 4 4 3 ...
 $ Gender: chr "Female" "Male" "Male" "Female" ...
 $ HourlyRate: int 94 61 92 56 40 79 81 67 44 94 ...
 $ JobInvolvement: int 3 2 2 3 3 3 4 3 2 3 ...
 $ JobLevel: int 2 1 1 1 1 1 1 3 2 ...
 $ JobRole: chr "Sales Executive" "Research Scientist" "Laboratory Technician" "Research Scientist" ...
 $ JobSatisfaction: int 4 2 3 3 2 4 1 3 3 3 ...
 $ MaritalStatus: chr "Single" "Married" "Single" "Married" ...
```
- Environment Panel:** Shows the loaded dataset 'wine_quality_mer...' with 6497 observations and 13 variables.
- Files Panel:** Displays the file explorer with various project files and folders.

The screenshot shows the RStudio interface with the following components:

- Source Panel:** Contains R code for summarizing the data frame.

```
> summary(data)
   Age      Attrition    BusinessTravel    DailyRate
Min.   :18.00  Length:1470  Length:1470  Min.   :102.0
1st Qu.:30.00  Class :character  Class :character  1st Qu.: 465.0
Median :36.00  Mode  :character  Mode  :character  Median : 802.0
Mean   :36.92                                     Mean   : 802.5
3rd Qu.:43.00                                     3rd Qu.:1157.0
Max.   :60.00                                     Max.   :1499.0

   Department    DistanceFromHome    Education    EducationField
Length:1470     Min.   : 1.000     Min.   :1.000     Length:1470
Class :character  1st Qu.: 2.000     1st Qu.:2.000     Class :character
Mode  :character  Mean   : 9.193     Mean   :2.913     Mode  :character
3rd Qu.:14.000    3rd Qu.:4.000
Max.   :29.000    Max.   :15.000

EmployeeCount EmployeeNumber EnvironmentSatisfaction Gender
Min.   :1      Min.   :1.0      Min.   :1.000      Length:1470
1st Qu.:1      1st Qu.: 491.2    1st Qu.:2.000      Class :character
Median :1      Median :1020.5    Median :3.000      Mode  :character
Mean   :1      Mean   :1024.9    Mean   :2.722
```
- Environment Panel:** Shows the loaded dataset 'wine_quality_mer...' with 6497 observations and 13 variables.
- Files Panel:** Displays the file explorer with various project files and folders.

RIYA RANE
S107 SYCS

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

The screenshot shows the RStudio interface with the following details:

- Source:** R 4.5.2 - ~/
- Console:** Displays summary statistics for 'Employee' data, including counts, means, medians, and standard deviations for variables like EmployeeNumber, Environmentsatisfaction, Gender, HourlyRate, JobInvolvement, JobLevel, JobRole, JobSatisfaction, MaritalStatus, MonthlyIncome, MonthlyRate, NumCompaniesworked, over18, overTime, PercentSalaryHike, PerformanceRating, Relationshipsatisfaction, standardHours, and stockoptionLevel.
- Environment:** Shows the 'wine_quality_mer...' dataset with 6497 observations and 13 variables.
- Files:** Lists files in the project directory, including .RData, .Rhistory, Custom Office Templates, desktop.ini, JAVA PROJECT by S094 and S107 (Shivam Mane anc), java project clip 1.mp4, java project clip 2.mp4, NetBeansProjects, prac4r.R, prac8 java- Construct a GUI using Java Swing to acc 0 B, raneS107, and UltraViewer.

The screenshot shows the RStudio interface with the following details:

- Source:** R 4.5.2 - ~/
- Console:** Displays the execution of a t-test. The code includes comments for 'STEP 3: Select the variable for test' and 'STEP 5: One-Sample t-Test'. The output shows the mean monthly income (6500) and the t-test results. An error message is displayed: 'Error: object 't_test_re' not found'.
- Environment:** Shows the 'wine_quality_mer...' dataset with 6497 observations and 13 variables.
- Files:** Lists files in the project directory, including .RData, .Rhistory, Custom Office Templates, desktop.ini, JAVA PROJECT by S094 and S107 (Shivam Mane anc), java project clip 1.mp4, java project clip 2.mp4, NetBeansProjects, prac4r.R, prac8 java- Construct a GUI using Java Swing to acc 0 B, raneS107, and UltraViewer.

PRACTICAL NO:5

PERFORMING INDEPENDENT TWO SAMPLE T-TESTS USING T.TESTS() WITH GROUPING

RIYA RANE
S107 SYCS

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

CODE:

```
# -----  
# INDEPENDENT TWO SAMPLE T-TEST USING t.test()  
# -----  
  
# 1. Create / Load Data  
# Example dataset  
data <- data.frame(  
  Score = c(45, 50, 48, 52, 47, 60, 62, 58, 65, 61),  
  Group = c("A", "A", "A", "A", "A", "B", "B", "B", "B", "B")  
)  
  
# View data  
print(data)  
  
# -----  
# 2. Check structure  
# -----  
str(data)  
  
# -----  
# 3. Descriptive statistics  
# -----  
aggregate(Score ~ Group, data = data,  
  FUN = function(x) c(  
    Mean = mean(x),  
    SD = sd(x),  
    N = length(x)  
  ))  
  
# -----  
# 4. Normality Test (Shapiro-Wilk)  
# -----  
shapiro.test(data$Score[data$Group == "A"])  
shapiro.test(data$Score[data$Group == "B"])  
  
# -----  
# 5. Equality of Variance Test (F-test)  
# -----  
var.test(Score ~ Group, data = data)  
  
# -----
```

RIYA RANE
S107 SYCS

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

6. Independent Two Sample t-test

6a. Welch t-test (Unequal variance) - DEFAULT

t_test_result <- t.test(Score ~ Group, data = data)

Print result

print(t_test_result)

6b. Equal variance assumed

t_test_equal_var <- t.test(Score ~ Group, data = data, var.equal = TRUE)

print(t_test_equal_var)

7. Manual Method (Subsetting)

group_A <- data\$Score[data\$Group == "A"]

group_B <- data\$Score[data\$Group == "B"]

t.test(group_A, group_B)

END OF CODE

OUTPUT:

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

```
> # Step 1: Load dataset
> # -----
> data <- read.csv("C:/Users/vibro/Downloads/top_rated_2000webseries.csv")
>
> # View structure and column names
> str(data)
'data.frame':   2000 obs. of  11 variables:
 $ id          : int  1 2 3 4 5 6 7 8 9 10 ...
 $ title       : chr  "Breaking Bad" "Avatar: The Last Airbender" "Arcane"
 $ original_title : chr  "Breaking Bad" "Avatar: The Last Airbender" "Arcane"
 $ overview    : chr  "Walter White, a New Mexico chemistry teacher, is diagnosed with Stage III cancer and given a prognosis of only '...' truncated. 'In a war-torn world of elemental magic, a young boy reawakens to undertake a dangerous mystic quest to fulfill '...' truncated. 'Amid the stark discord of twin cities Piltover and Zaun, two sisters fight on rival sides of a war between magi'...' truncated. 'In Jeju, a spirited girl and a steadfast boy's island story blossoms into a lifelong tale of setbacks and triumph'...' truncated. ...
 $ premiere_date : chr  "2008-01-20" "2005-02-21" "2021-11-06" "2025-03-07" ...
 $ popularity    : num  108.8 12.5 22.6 18.6 28.4 ...
 $ genre         : chr  "Drama, Crime" "Animation, Action & Adventure, Sci-Fi & Fantasy" "Animation, Sci-Fi & Fantasy, Drama, Action & Adventure" "Drama" ...
 $ country_origin : chr  "United States" "United States" "United States" "South Korea" ...
 $ original_language : chr  "English" "English" "English" "Korean" ...
 $ rating        : num  8.9 8.8 8.8 8.75 8.73 ...
 $ votes         : int  16556 4557 5481 423 565 5043 2291 689 5992 7275 ...
```

Environment: List of 10

- t_test_result: List of 10
- train107: 9800 obs. of 18 variables
- wine: 6497 obs. of 13 variables
- wine_numeric: 6497 obs. of 12 variables
- wine_quality_mer...: 6497 obs. of 13 variables

Files: .RData (1.2 MB), .Rhistory (3 KB), Custom Office Templates, desktop.ini (418 B), JAVA PROJECT by S094 and S107 (Shivam Mane anc... (11 MB), java project clip 1.mp4 (5.4 MB), java project clip 2.mp4 (7.2 MB), NetBeansProjects, prac4r.R (799 B), prac8 java- Construct a GUI using Java Swing to acc... (0 B), raneS107, UltraViewer

```
> colnames(data)
[1] "id"           "title"         "original_title"
[4] "overview"     "premiere_date" "popularity"
[7] "genre"        "country_origin" "original_language"
[10] "rating"       "votes"

> summary(data)
   id           title           original_title      overview
Min.   : 1.0   Length:2000   Length:2000   Length:2000
1st Qu.: 500.8 Class :character Class :character Class :character
Median :1000.5 Mode  :character Mode  :character Mode  :character
Mean   :1000.5
3rd Qu.:1500.2
Max.   :2000.0
premiere_date popularity      genre      country_origin
Length:2000   Min.   : 0.3921   Length:2000   Length:2000
Class :character 1st Qu.: 4.8211   Class :character Class :character
Mode  :character  Median : 8.2478   Mode  :character Mode  :character
                  Mean   :13.5572
                  3rd Qu.:15.1774
                  Max.   :338.2498

original_language rating      votes
Length:2000       Min.   :7.086   Min.   : 200.0
Class :character 1st Qu.:7.500   1st Qu.: 201.8
```

Environment: List of 10

- t_test_result: List of 10
- train107: 9800 obs. of 18 variables
- wine: 6497 obs. of 13 variables
- wine_numeric: 6497 obs. of 12 variables
- wine_quality_mer...: 6497 obs. of 13 variables

Files: .RData (1.2 MB), .Rhistory (3 KB), Custom Office Templates, desktop.ini (418 B), JAVA PROJECT by S094 and S107 (Shivam Mane anc... (11 MB), java project clip 1.mp4 (5.4 MB), java project clip 2.mp4 (7.2 MB), NetBeansProjects, prac4r.R (799 B), prac8 java- Construct a GUI using Java Swing to acc... (0 B), raneS107, UltraViewer

RIYA RANE
S107 SYCS

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

The screenshot shows the RStudio interface with the following components:

- Console:** Contains R code for creating a dataset, viewing it, and performing initial checks. The output shows a data frame with 10 observations and 2 variables (Score and Group).
- Environment:** Lists the objects in the global environment: t_test_result, train107, wine, wine_numeric, and wine_quality_mer....
- Files:** Shows the file explorer with various files and folders, including .RData, .Rhistory, and Custom Office Templates.

```
> # Example dataset
> data <- data.frame(
+   Score = c(45, 50, 48, 52, 47, 60, 62, 58, 65, 61),
+   Group = c("A", "A", "A", "A", "A", "B", "B", "B", "B", "B")
+ )
> # view data
> print(data)
  Score Group
1    45    A
2    50    A
3    48    A
4    52    A
5    47    A
6    60    B
7    62    B
8    58    B
9    65    B
10   61    B
> # -----
> # 2. check structure
> # -----
> str(data)
'data.frame':   10 obs. of  2 variables:
 $ Score: num   45 50 48 52 47 60 62 58 65 61
 $ Group: chr   "A" "A" "A" "A" "A" ...
> # -----
> # 3. Descriptive statistics
> # -----
```

The screenshot shows the RStudio interface with the following components:

- Console:** Contains R code for calculating descriptive statistics and performing normality tests. The output shows the mean, standard deviation, and sample size for each group, followed by Shapiro-Wilk test results.
- Environment:** Lists the objects in the global environment: t_test_result, train107, wine, wine_numeric, and wine_quality_mer....
- Files:** Shows the file explorer with various files and folders, including .RData, .Rhistory, and Custom Office Templates.

```
> # -----
> # 3. Descriptive statistics
> # -----
> aggregate(Score ~ Group, data = data,
+   FUN = function(x) c(
+     Mean = mean(x),
+     SD = sd(x),
+     N = length(x)
+   ))
  Group Score.Mean Score.SD Score.N
1     A  48.400000  2.701851  5.000000
2     B  61.200000  2.588436  5.000000
> # -----
> # 4. Normality Test (Shapiro-wilk)
> # -----
> shapiro.test(data$Score[data$Group == "A"])
      Shapiro-Wilk normality test
data:  data$Score[data$Group == "A"]
W = 0.98998, p-value = 0.9796
> shapiro.test(data$Score[data$Group == "B"])
      Shapiro-Wilk normality test
data:  data$Score[data$Group == "B"]
W = 0.98396, p-value = 0.9546
> # -----
```

RIYA RANE
S107 SYCS

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

The screenshot shows the RStudio interface with the following content:

```
data: data$Score[data$Group == "B"]
w = 0.98396, p-value = 0.9546

> # -----
> # 5. Equality of Variance Test (F-test)
> # -----
> var.test(Score ~ Group, data = data)

F test to compare two variances

data: Score by Group
F = 1.0896, num df = 4, denom df = 4, p-value = 0.9358
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1134415 10.4646370
sample estimates:
ratio of variances
 1.089552

> # -----
> # 6. Independent Two Sample t-test
> # -----
> # 6a. Welch t-test (Unequal variance) - DEFAULT
> t_test_result <- t.test(Score ~ Group, data = data)
> # Print result
> print(t_test_result)
```

The Environment pane on the right shows the following objects:

Object	Class	Attributes
t_test_result	List of 10	
train107	9800 obs. of 18 variables	
wine	6497 obs. of 13 variables	
wine_numeric	6497 obs. of 12 variables	
wine_quality_mer...	6497 obs. of 13 variables	

The Files pane shows the following files:

Name	Size	Modified
.RData	1.2 MB	Dec 1, 2025, 7:20 PM
.Rhistory	3 KB	Dec 1, 2025, 7:20 PM
Custom Office Templates		
desktop.ini	418 B	Dec 8, 2025, 10:41 AM
JAVA PROJECT by S094 and S107 (Shivam Mane anc	11 MB	Oct 3, 2025, 6:52 PM
java project clip 1.mp4	5.4 MB	Oct 3, 2025, 7:36 PM
java project clip 2.mp4	7.2 MB	Oct 3, 2025, 7:36 PM
NetBeansProjects		
prac4r.R	799 B	Dec 1, 2025, 6:40 PM
prac8 java.- Construct a GUI using Java Swing to acc 0 B	0 B	Sep 11, 2025, 7:08 PM
raneS107		
UltraViewer		

The screenshot shows the RStudio interface with the following content:

```
mean in group A mean in group B
48.4 61.2

> # -----
> # 6b. Equal variance assumed
> # -----
> t_test_equal_var <- t.test(Score ~ Group, data = data, var.equal = TRUE)
> print(t_test_equal_var)

Two Sample t-test

data: Score by Group
t = -7.6495, df = 8, p-value = 6.02e-05
alternative hypothesis: true difference in means between group A and group B is n
ot equal to 0
95 percent confidence interval:
 -16.658683 -8.941317
sample estimates:
mean in group A mean in group B
48.4 61.2

> # -----
> # 7. Manual Method (Subsetting)
> # -----
> group_A <- data$Score[data$Group == "A"]
> group_B <- data$Score[data$Group == "B"]
> t.test(group_A, group_B)
```

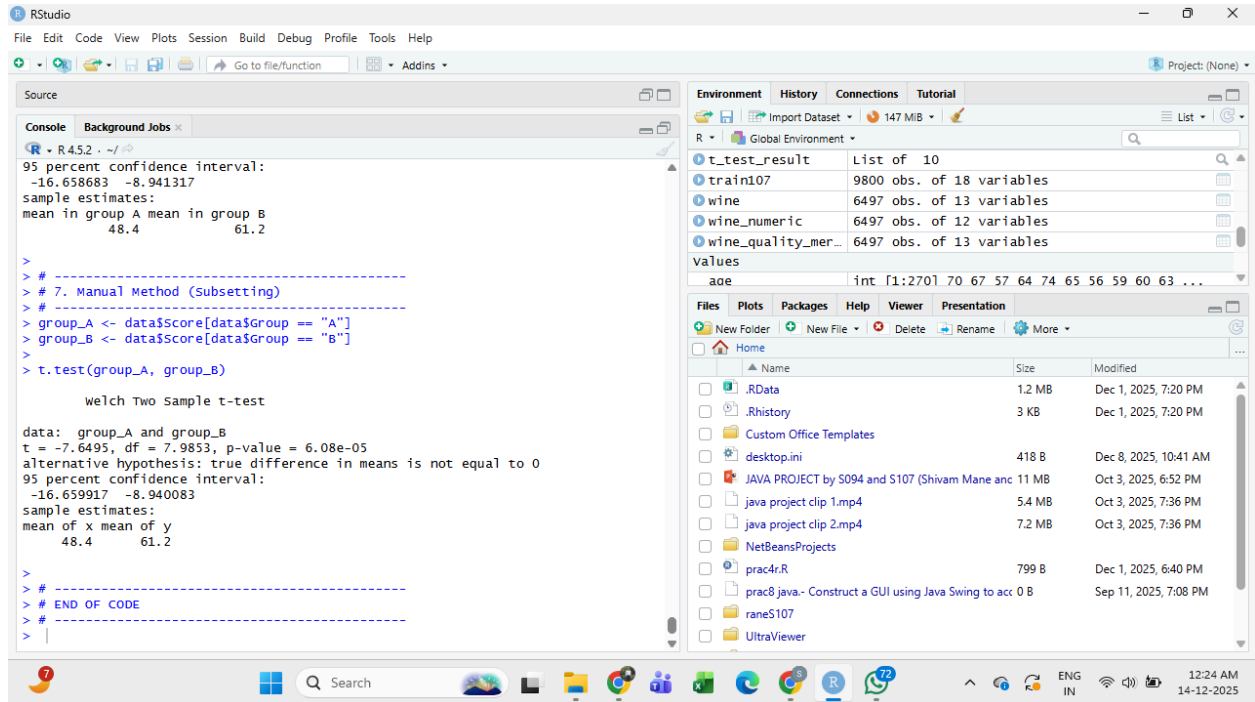
The Environment pane on the right shows the following objects:

Object	Class	Attributes
t_test_result	List of 10	
train107	9800 obs. of 18 variables	
wine	6497 obs. of 13 variables	
wine_numeric	6497 obs. of 12 variables	
wine_quality_mer...	6497 obs. of 13 variables	

The Files pane shows the following files:

Name	Size	Modified
.RData	1.2 MB	Dec 1, 2025, 7:20 PM
.Rhistory	3 KB	Dec 1, 2025, 7:20 PM
Custom Office Templates		
desktop.ini	418 B	Dec 8, 2025, 10:41 AM
JAVA PROJECT by S094 and S107 (Shivam Mane anc	11 MB	Oct 3, 2025, 6:52 PM
java project clip 1.mp4	5.4 MB	Oct 3, 2025, 7:36 PM
java project clip 2.mp4	7.2 MB	Oct 3, 2025, 7:36 PM
NetBeansProjects		
prac4r.R	799 B	Dec 1, 2025, 6:40 PM
prac8 java.- Construct a GUI using Java Swing to acc 0 B	0 B	Sep 11, 2025, 7:08 PM
raneS107		
UltraViewer		

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



PRACTICAL NO: 6

PERFORMING PAIRED T.TESTS USING T.TESTS(PAIRED=TRUE)

CODE:

```
# 1. Create Own Dataset (Before-After Example)
data <- data.frame(
  Subject = 1:10,
  Before = c(65, 70, 68, 72, 66, 75, 78, 74, 80, 77),
  After = c(70, 75, 72, 78, 71, 80, 82, 79, 85, 83)
)
```

```
# View data
print(data)
```

```
# -----
# 2. Check structure
# -----
str(data)
```

```
# -----
# 3. Descriptive statistics
# -----
```

RIYA RANE
S107 SYCS

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

```
mean(data$Before)
mean(data$After)

sd(data$Before)
sd(data$After)

# -----
# 4. Normality Test on Differences
# -----
difference <- data$After - data$Before
shapiro.test(difference)

# -----
# 5. Paired t-test
# -----
paired_t_test <- t.test(
  data$Before,
  data$After,
  paired = TRUE
)

# Print result
print(paired_t_test)

# -----
# 6. Alternative way (Formula method)
# -----
# Convert to long format
long_data <- data.frame(
  Score = c(data$Before, data$After),
  Time = rep(c("Before", "After"), each = nrow(data))
)

t.test(Score ~ Time, data = long_data, paired = TRUE)

# -----
# END OF CODE
# -----
```

OUTPUT:

RIYA RANE
S107 SYCS

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

This screenshot shows the RStudio interface with the following content:

- Console:**

```
R - R4.5.2 - ~/r
> # view data
> print(data)
  subject Before After
1       1     65    70
2       2     70    75
3       3     68    72
4       4     72    78
5       5     66    71
6       6     75    80
7       7     78    82
8       8     74    79
9       9     80    85
10      10     77    83

> # -----
> # 2. check structure
> # -----
> str(data)
'data.frame':   10 obs. of  3 variables:
 $ subject: int   1  2  3  4  5  6  7  8  9 10
 $ Before : num  65 70 68 72 66 75 78 74 80 77
 $ After  : num  70 75 72 78 71 80 82 79 85 83

> # -----
> # 3. Descriptive statistics
> # -----
> mean(data$Before)
[1] 72.5
> mean(data$After)
[1] 77.5
> sd(data$Before)
```
- Environment:** Lists objects: paired_t_test (List of 10), real_drug_dataset (1000 obs. of 9 variables), row_percent (num[0, 0]), sales (9800 obs. of 18 variables), special_rows (5151 obs. of 18 variables), t_test_equal_var (List of 10), t_test result (List of 10).
- Files:** Shows a file explorer with various system and project files.

This screenshot shows the RStudio interface with the following content:

- Console:**

```
R - R4.5.2 - ~/r
> # -----
> # 3. Descriptive statistics
> # -----
> mean(data$Before)
[1] 72.5
> mean(data$After)
[1] 77.5
> # -----
> sd(data$Before)
[1] 5.169354
> sd(data$After)
[1] 5.275731
> # -----
> # 4. Normality Test on Differences
> # -----
> difference <- data$After - data$Before
> shapiro.test(difference)

Shapiro-wilk normality test

data: difference
W = 0.81484, p-value = 0.02195

> # -----
> # 5. Paired t-test
> # -----
> paired_t_test <- t.test(
+   data$Before,
+   data$After,
```
- Environment:** Same as the previous screenshot.
- Files:** Same as the previous screenshot.

RIYA RANE
S107 SYCS

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

RStudio interface showing a script with the following code:

```
data: difference
w = 0.81484, p-value = 0.02195

> # -----
> # 5. Paired t-test
> # -----
> paired_t_test <- t.test(
+   data$Before,
+   data$After,
+   paired = TRUE
+ )
> # Print result
> print(paired_t_test)

Paired t-test

data: data$Before and data$After
t = -23.717, df = 9, p-value = 2.01e-09
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -5.476905 -4.523095
sample estimates:
mean difference
      -5

> # -----
> # 6. Alternative way (Formula method)
```

The Environment pane on the right shows the following objects:

Object	Type
paired_t_test	List of 10
real_drug_dataset	1000 obs. of 9 variables
row_percent	num[0, 0]
sales	9800 obs. of 18 variables
special_rows	5151 obs. of 18 variables
t_test_equal_var	List of 10
t_test_result	List of 10

RStudio interface showing a script with the following code:

```
> # Print result
> print(paired_t_test)

Paired t-test

data: data$Before and data$After
t = -23.717, df = 9, p-value = 2.01e-09
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -5.476905 -4.523095
sample estimates:
mean difference
      -5

> # -----
> # 6. Alternative way (Formula method)
> # -----
> # Convert to long format
> long_data <- data.frame(
+   Score = c(data$Before, data$After),
+   Time = rep(c("Before", "After"), each = nrow(data))
+ )
> t.test(Score ~ Time, data = long_data, paired = TRUE)

Error in t.test.formula(Score ~ Time, data = long_data, paired = TRUE) :
cannot use 'paired' in formula method
```

The Environment pane on the right shows the following objects:

Object	Type
paired_t_test	List of 10
real_drug_dataset	1000 obs. of 9 variables
row_percent	num[0, 0]
sales	9800 obs. of 18 variables
special_rows	5151 obs. of 18 variables
t_test_equal_var	List of 10
t_test_result	List of 10

RIYA RANE
S107 SYCS