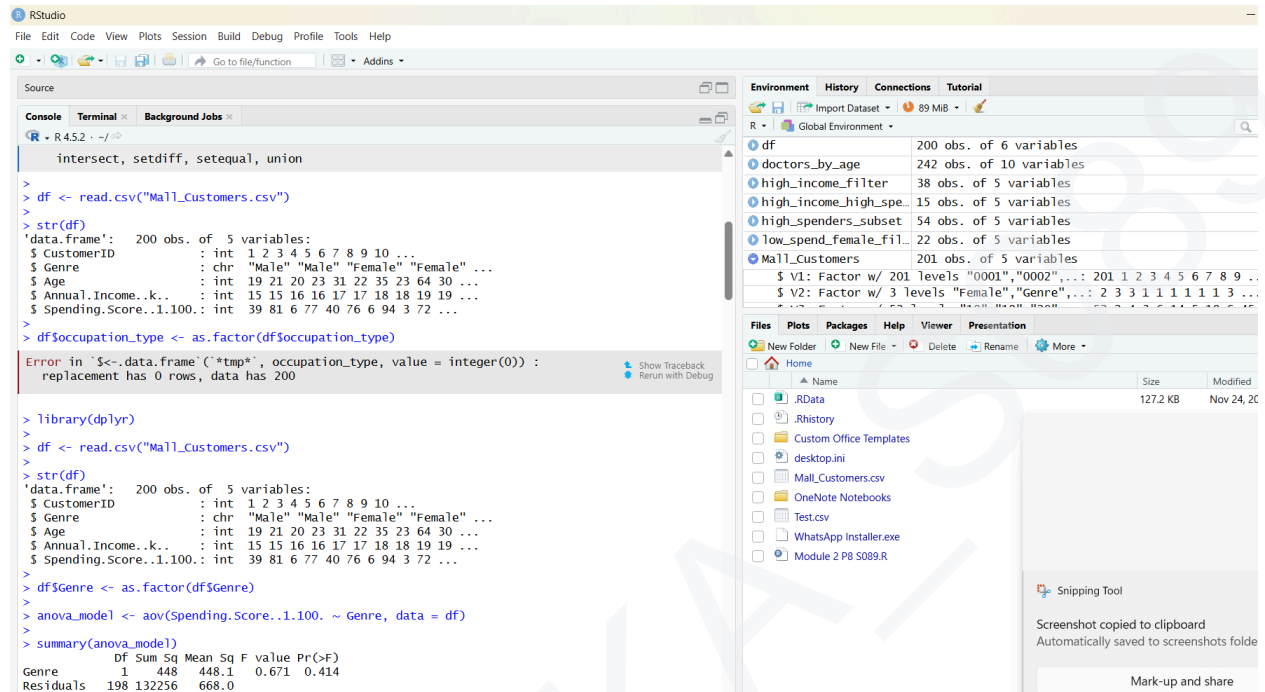


SHETH L.U.J AND SIR M.V COLLEGE
SUBJECT : Data Analysis with SAS / SPSS /R
PRACTICAL NO: 7,8,9

OUTPUTS:

AIM:

7 Performing one-way ANOVA using aov()



```
Source
Console Terminal Background Jobs
R 4.5.2 ~ /
intersect, setdiff, setequal, union

>
> df <- read.csv("Mall_Customers.csv")
> str(df)
'data.frame': 200 obs. of 5 variables:
 $ CustomerID : int 1 2 3 4 5 6 7 8 9 10 ...
 $ Genre : chr "Male" "Male" "Female" "Female" ...
 $ Age : int 19 21 20 23 31 22 35 23 64 30 ...
 $ Annual.Income..k.. : int 15 15 16 16 17 17 18 18 19 19 ...
 $ Spending.Score..1.100.: int 39 81 6 77 40 76 6 94 3 72 ...
> df$occupation_type <- as.factor(df$occupation_type)
Error in '$<-data.frame'(*tmp*, occupation_type, value = integer(0)) :
replacement has 0 rows, data has 200

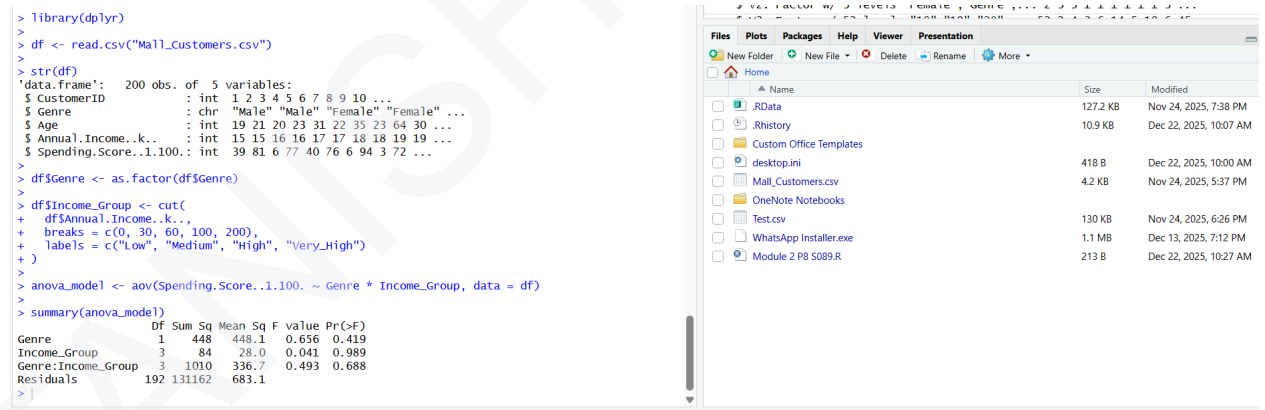
> library(dplyr)
>
> df <- read.csv("Mall_Customers.csv")
> str(df)
'data.frame': 200 obs. of 5 variables:
 $ CustomerID : int 1 2 3 4 5 6 7 8 9 10 ...
 $ Genre : chr "Male" "Male" "Female" "Female" ...
 $ Age : int 19 21 20 23 31 22 35 23 64 30 ...
 $ Annual.Income..k.. : int 15 15 16 16 17 17 18 18 19 19 ...
 $ Spending.Score..1.100.: int 39 81 6 77 40 76 6 94 3 72 ...
> df$Genre <- as.factor(df$Genre)
>
> anova_model <- aov(Spending.Score..1.100. ~ Genre, data = df)
> summary(anova_model)
              Df Sum Sq Mean Sq F value Pr(>F)
Genre          1    448    448.1    0.671  0.414
Residuals     198 132256    668.0
>
```

Environment History Connections Tutorial
R Global Environment
df 200 obs. of 6 variables
doctors_by_age 242 obs. of 10 variables
high_income_filter 38 obs. of 5 variables
high_income_high_spe 15 obs. of 5 variables
high_spenders_subset 54 obs. of 5 variables
low_spend_female_fil 22 obs. of 5 variables
Mall_Customers 201 obs. of 5 variables
\$ V1: Factor w/ 201 levels "0001","0002",...: 201 1 2 3 4 5 6 7 8 9 ...
\$ V2: Factor w/ 3 levels "Female","Genre",...: 2 3 3 1 1 1 1 1 1 3 ...

Files Plots Packages Help Viewer Presentation
New Folder New File Delete Rename More
Home
Name Size Modified
RData 127.2 KB Nov 24, 2025
Rhistory 10.9 KB Dec 22, 2025, 10:07 AM
Custom Office Templates
desktop.ini 418 B Dec 22, 2025, 10:00 AM
Mall_Customers.csv 4.2 KB Nov 24, 2025, 5:37 PM
OneNote Notebooks
Test.csv 130 KB Nov 24, 2025, 6:26 PM
WhatsApp Installer.exe 1.1 MB Dec 13, 2025, 7:12 PM
Module 2 P8 S089.R 213 B Dec 22, 2025, 10:27 AM

Snipping Tool
Screenshot copied to clipboard
Automatically saved to screenshots folder
Mark-up and share

8 Performing two-way ANOVA using aov() (R).



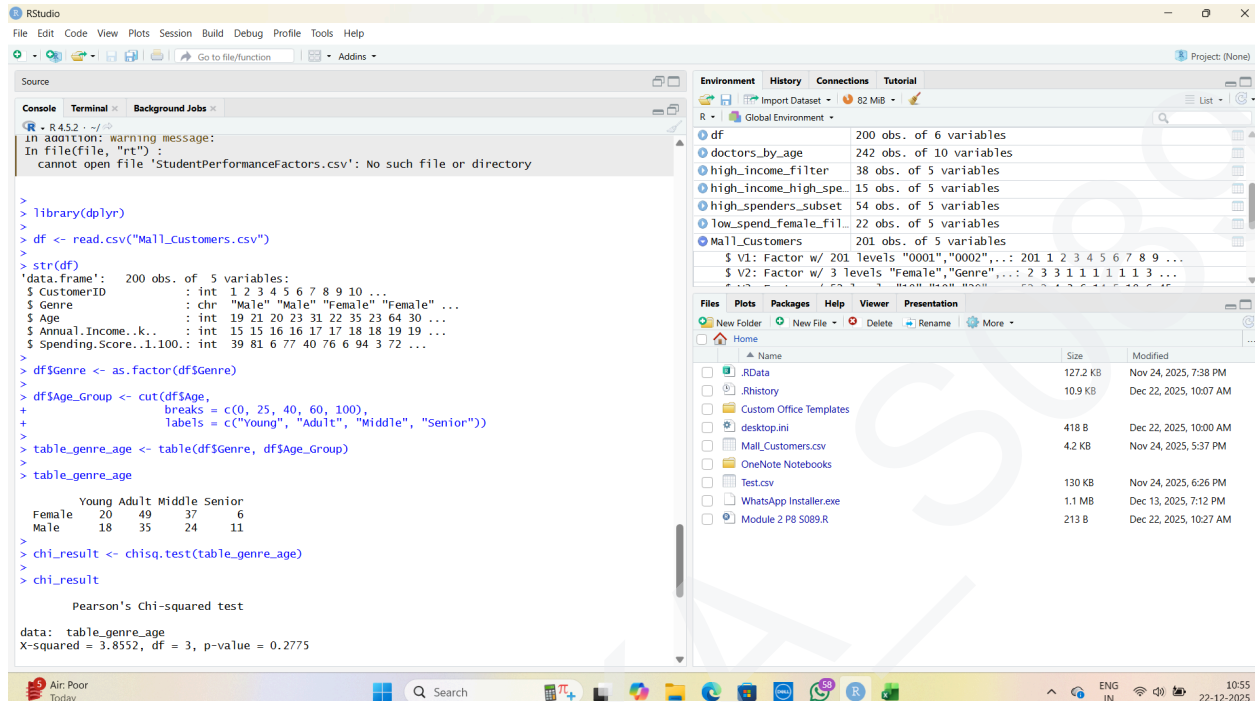
```
library(dplyr)
df <- read.csv("Mall_Customers.csv")
str(df)
'data.frame': 200 obs. of 5 variables:
 $ CustomerID : int 1 2 3 4 5 6 7 8 9 10 ...
 $ Genre : chr "Male" "Male" "Female" "Female" ...
 $ Age : int 19 21 20 23 31 22 35 23 64 30 ...
 $ Annual.Income..k.. : int 15 15 16 16 17 17 18 18 19 19 ...
 $ Spending.Score..1.100.: int 39 81 6 77 40 76 6 94 3 72 ...
> df$Genre <- as.factor(df$Genre)
> df$Income_Group <- cut(
+ df$Annual.Income..k..,
+ breaks = c(0, 30, 60, 100, 200),
+ labels = c("Low", "Medium", "High", "Very_High")
+ )
> anova_model <- aov(Spending.Score..1.100. ~ Genre * Income_Group, data = df)
> summary(anova_model)
              Df Sum Sq Mean Sq F value Pr(>F)
Genre          1    448    448.1    0.656  0.419
Income_Group    3     84     28.0    0.041  0.989
Genre:Income_Group 3    1010    336.7    0.493  0.688
Residuals     192 131162    683.1
>
```

Files Plots Packages Help Viewer Presentation
New Folder New File Delete Rename More
Home
Name Size Modified
RData 127.2 KB Nov 24, 2025, 7:38 PM
Rhistory 10.9 KB Dec 22, 2025, 10:07 AM
Custom Office Templates
desktop.ini 418 B Dec 22, 2025, 10:00 AM
Mall_Customers.csv 4.2 KB Nov 24, 2025, 5:37 PM
OneNote Notebooks
Test.csv 130 KB Nov 24, 2025, 6:26 PM
WhatsApp Installer.exe 1.1 MB Dec 13, 2025, 7:12 PM
Module 2 P8 S089.R 213 B Dec 22, 2025, 10:27 AM

NAME: TANISHKA K KUTWAL
ROLL NO: S089

SHETH L.U.J AND SIR M.V COLLEGE
SUBJECT : Data Analysis with SAS / SPSS / R
PRACTICAL NO: 7,8,9

9 Conducting Chi-square tests using chisq.test() (R)



```
R - R452 - ~/r
in addition: warning message:
In file(file, "rt") :
  cannot open file 'StudentPerformanceFactors.csv': No such file or directory

> library(dplyr)
> df <- read.csv("Mall_Customers.csv")
> str(df)
'data.frame': 200 obs. of 5 variables:
 $ CustomerID : int 1 2 3 4 5 6 7 8 9 10 ...
 $ Genre : chr "Male" "Male" "Female" "Female" ...
 $ Age : int 19 21 20 23 31 22 35 23 64 30 ...
 $ Annual.Income..k.. : int 15 15 16 16 17 17 18 18 19 19 ...
 $ Spending.Score..1.100.: int 39 81 6 77 40 76 6 94 3 72 ...
> df$Genre <- as.factor(df$Genre)
> df$Age_Group <- cut(df$Age,
+ breaks = c(0, 25, 40, 60, 100),
+ labels = c("Young", "Adult", "Middle", "Senior"))
> table_genre_age <- table(df$Genre, df$Age_Group)
> table_genre_age
      Young Adult Middle Senior
Female    20    49     37      6
Male     18    35     24     11
> chi_result <- chisq.test(table_genre_age)
> chi_result

Pearson's Chi-squared test

data: table_genre_age
X-squared = 3.8552, df = 3, p-value = 0.2775
```

NAME: TANISHKA K KUTWAL
ROLL NO: S089