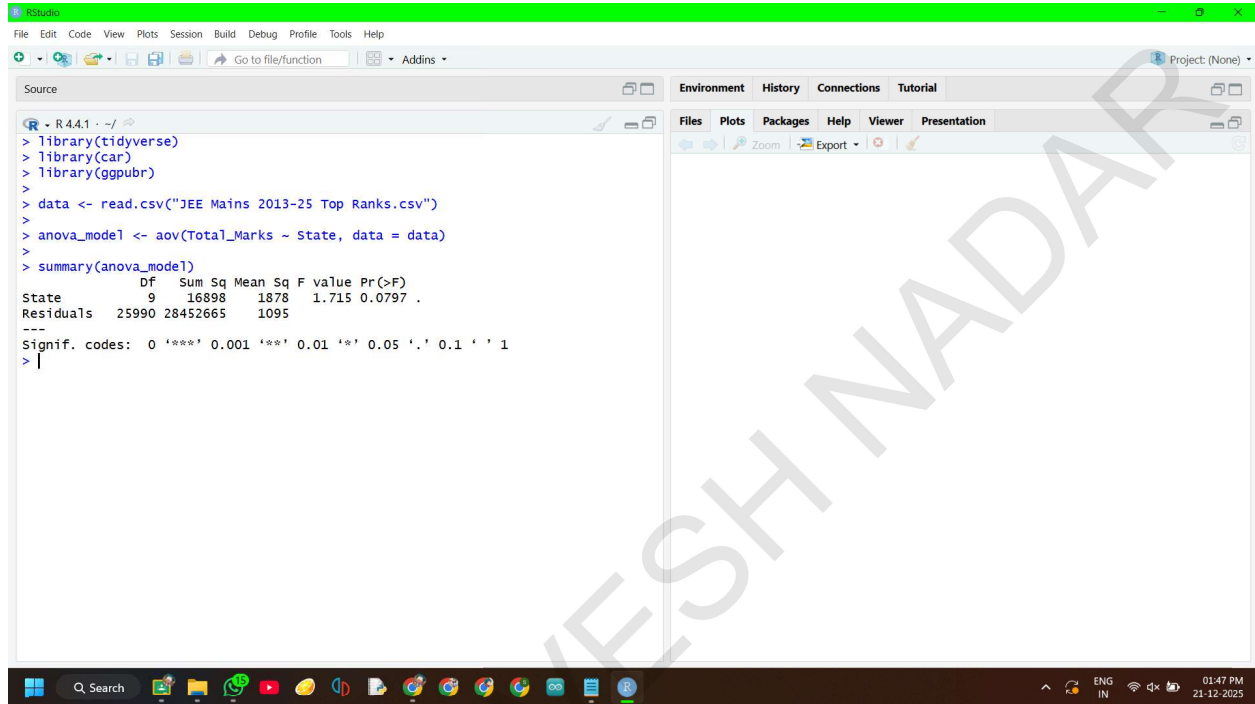


# SHETH L.U.J. AND SIR M.V. COLLEGE

## DATA ANALYSIS WITH R

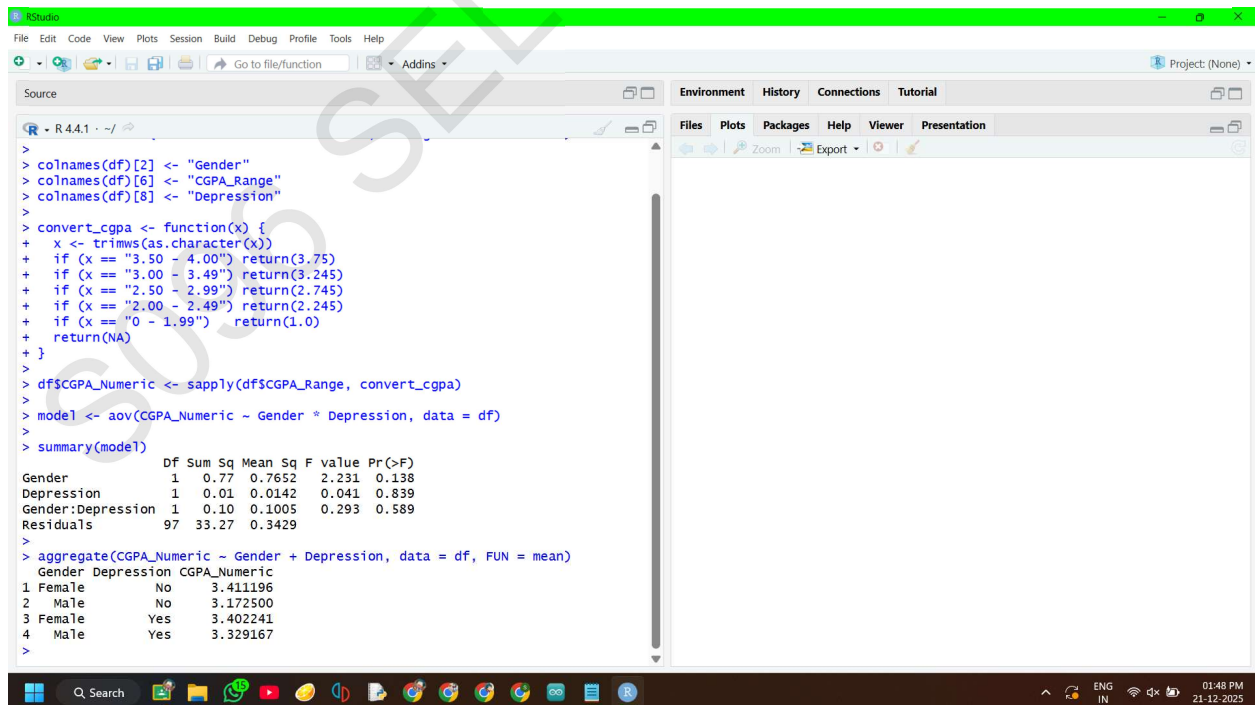
- AIMS: Performing one-way ANOVA using aov() (R).  
8 Performing two-way ANOVA using aov() (R).  
9 Conducting Chi-square tests using chisq.test() (R)

### PRAC 7 OUTPUT:



```
R - R 4.4.1 - ~/ -  
File Edit Code View Plots Session Build Debug Profile Tools Help  
Go to file/function Addins  
Project: (None)  
Source  
> library(tidyverse)  
> library(car)  
> library(ggpubr)  
>  
> data <- read.csv("JEE Mains 2013-25 Top Ranks.csv")  
>  
> anova_model <- aov(Total_Marks ~ State, data = data)  
> summary(anova_model)  
              Df Sum Sq Mean Sq F value Pr(>F)        
State          9  16898    1878   1.715 0.0797 .        
Residuals     25990 28452665    1095                
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
>
```

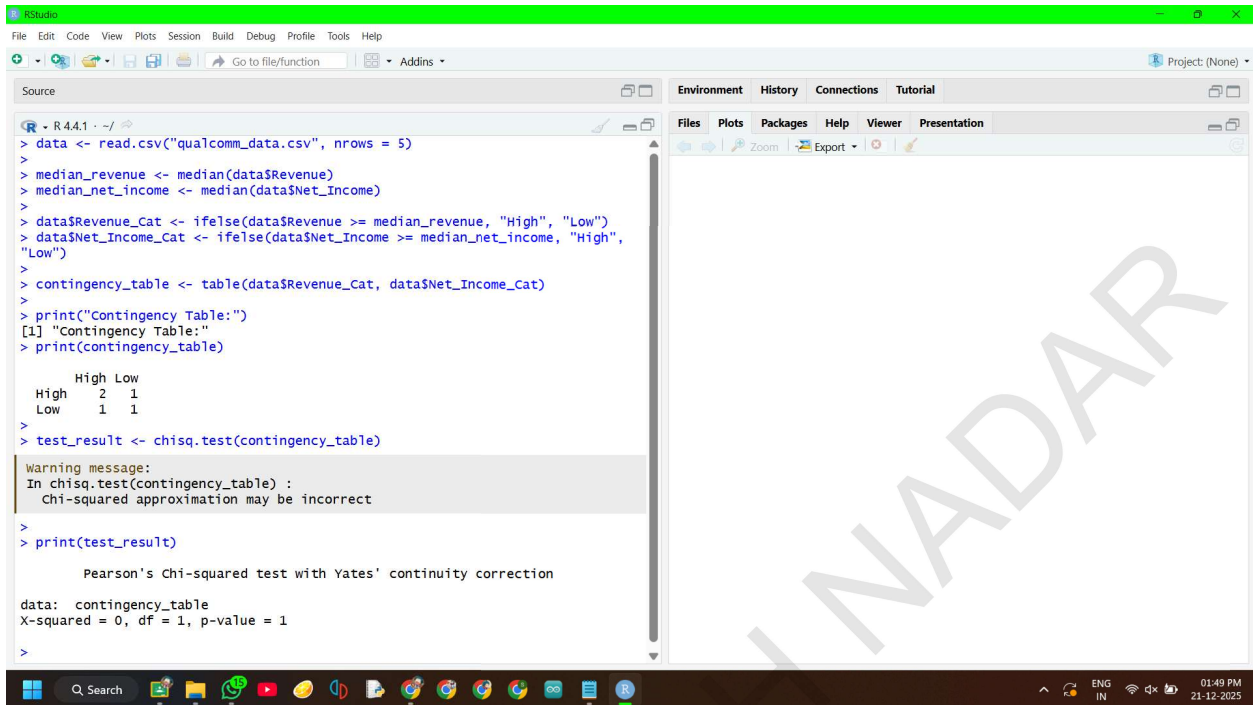
### PRAC 8 OUTPUT:



```
R - R 4.4.1 - ~/ -  
File Edit Code View Plots Session Build Debug Profile Tools Help  
Go to file/function Addins  
Project: (None)  
Source  
>  
> colnames(df)[2] <- "Gender"  
> colnames(df)[6] <- "CGPA_Range"  
> colnames(df)[8] <- "Depression"  
>  
> convert_cgpa <- function(x) {  
+   x <- trimws(as.character(x))  
+   if (x == "3.50 - 4.00") return(3.75)  
+   if (x == "3.00 - 3.49") return(3.245)  
+   if (x == "2.50 - 2.99") return(2.745)  
+   if (x == "2.00 - 2.49") return(2.245)  
+   if (x == "0 - 1.99") return(1.0)  
+   return(NA)  
+ }  
>  
> df$CGPA_Numeric <- sapply(df$CGPA_Range, convert_cgpa)  
>  
> model <- aov(CGPA_Numeric ~ Gender * Depression, data = df)  
> summary(model)  
              Df Sum Sq Mean Sq F value Pr(>F)        
Gender          1  0.77  0.7652   2.231 0.138        
Depression       1  0.01  0.0142   0.041 0.839        
Gender:Depression 1  0.10  0.1005   0.293 0.589        
Residuals       97 33.27  0.3429                
>  
> aggregate(CGPA_Numeric ~ Gender + Depression, data = df, FUN = mean)  
  Gender Depression CGPA_Numeric  
1 Female         No      3.411196  
2 Male          No      3.172500  
3 Female         Yes      3.402241  
4 Male          Yes      3.329167  
>
```

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

PRAC 9 OUTPUT:



The screenshot shows the RStudio interface with the following code in the Source pane:

```
> data <- read.csv("qualcomm_data.csv", nrow = 5)
>
> median_revenue <- median(data$Revenue)
> median_net_income <- median(data$Net_Income)
>
> data$Revenue_Cat <- ifelse(data$Revenue >= median_revenue, "High", "Low")
> data$Net_Income_Cat <- ifelse(data$Net_Income >= median_net_income, "High",
"Low")
>
> contingency_table <- table(data$Revenue_Cat, data$Net_Income_Cat)
>
> print("Contingency Table:")
[1] "Contingency Table:"
> print(contingency_table)
      High Low
High    2   1
Low     1   1
>
> test_result <- chisq.test(contingency_table)

Warning message:
In chisq.test(contingency_table) :
  Chi-squared approximation may be incorrect
>
> print(test_result)

Pearson's Chi-squared test with Yates' continuity correction

data: contingency_table
X-squared = 0, df = 1, p-value = 1
>
```

The Environment pane on the right is empty. The bottom status bar shows the system time as 01:49 PM on 21-12-2025.