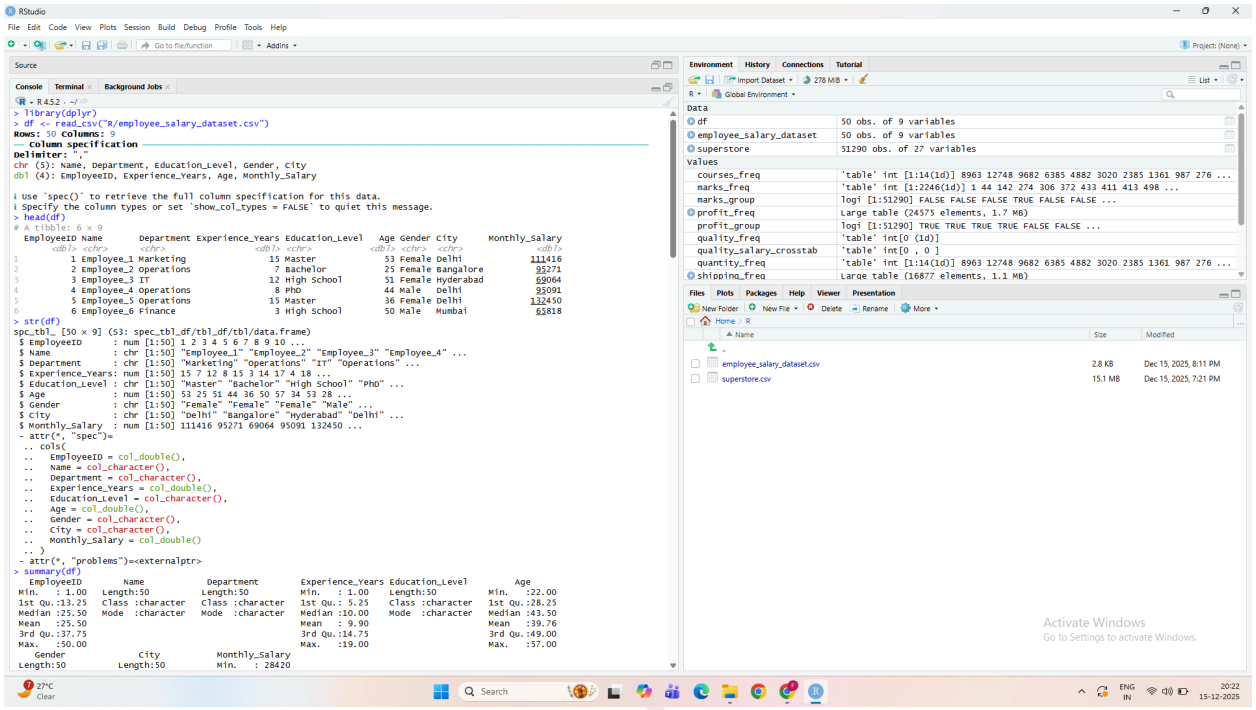
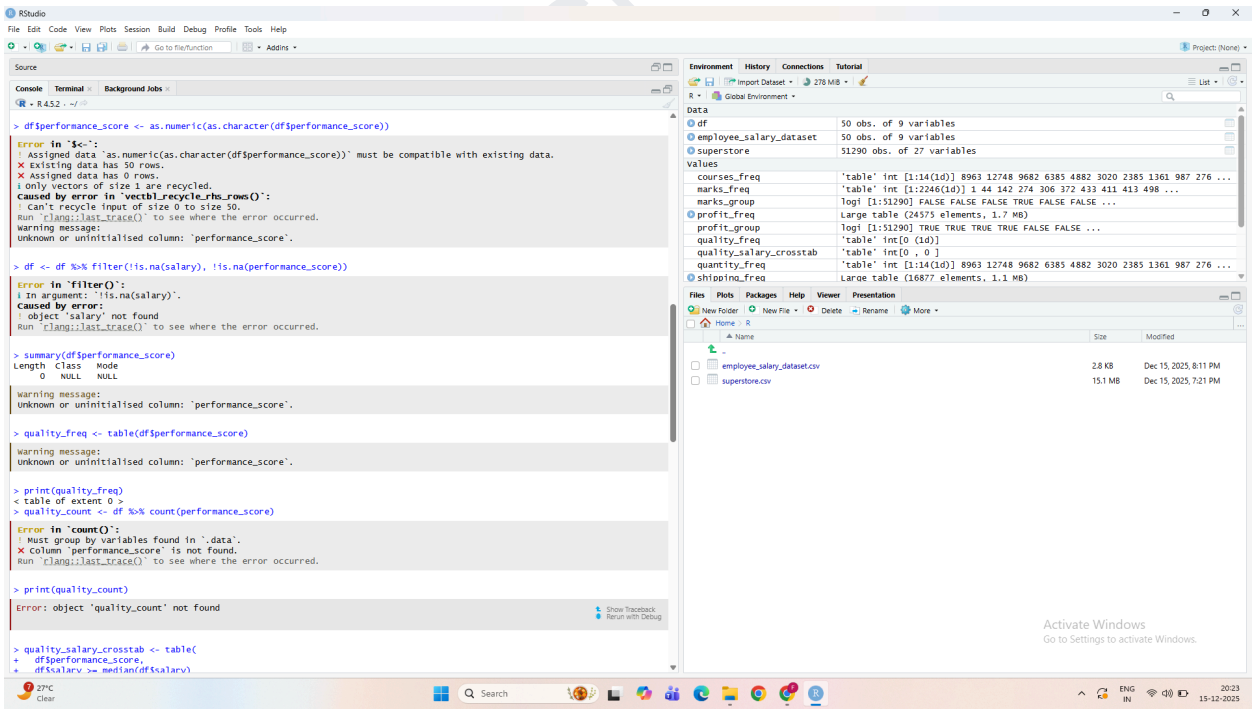
SHETH L.U.J. AND SIR M.V. COLLEGE

PRACTICAL NO 5

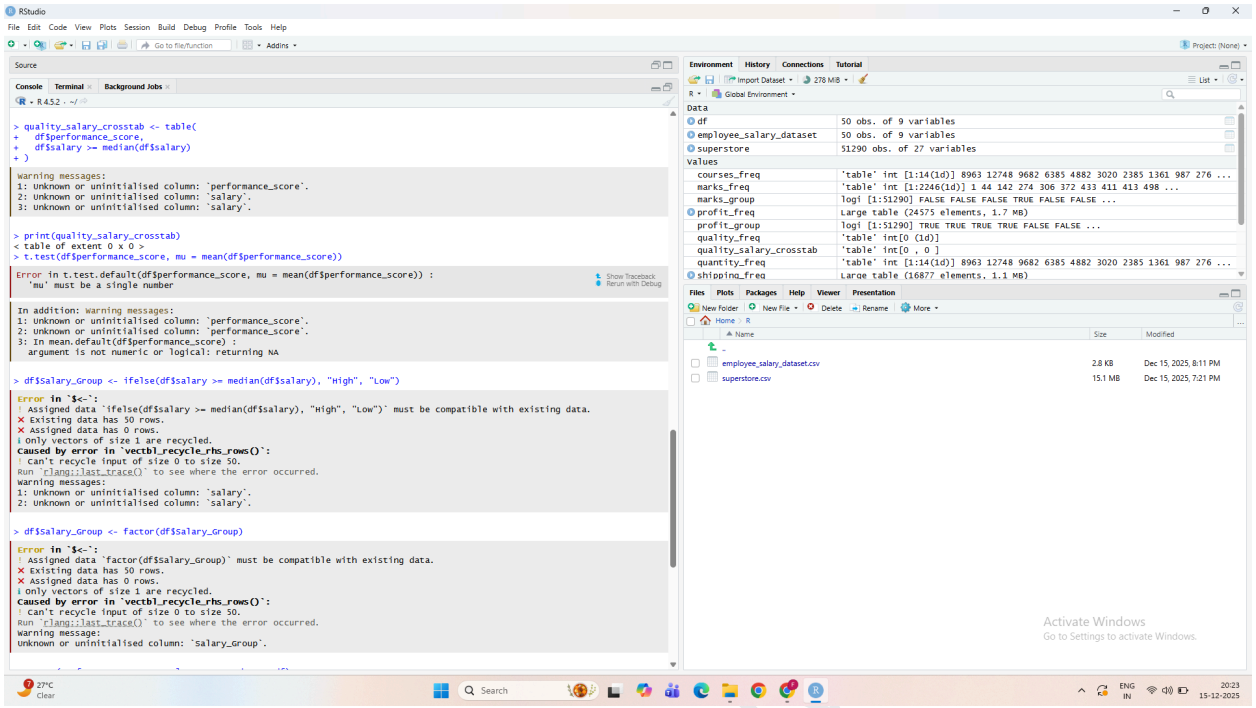
AIM: Performing independent two-sample t-tests using t.test() with grouping (R).

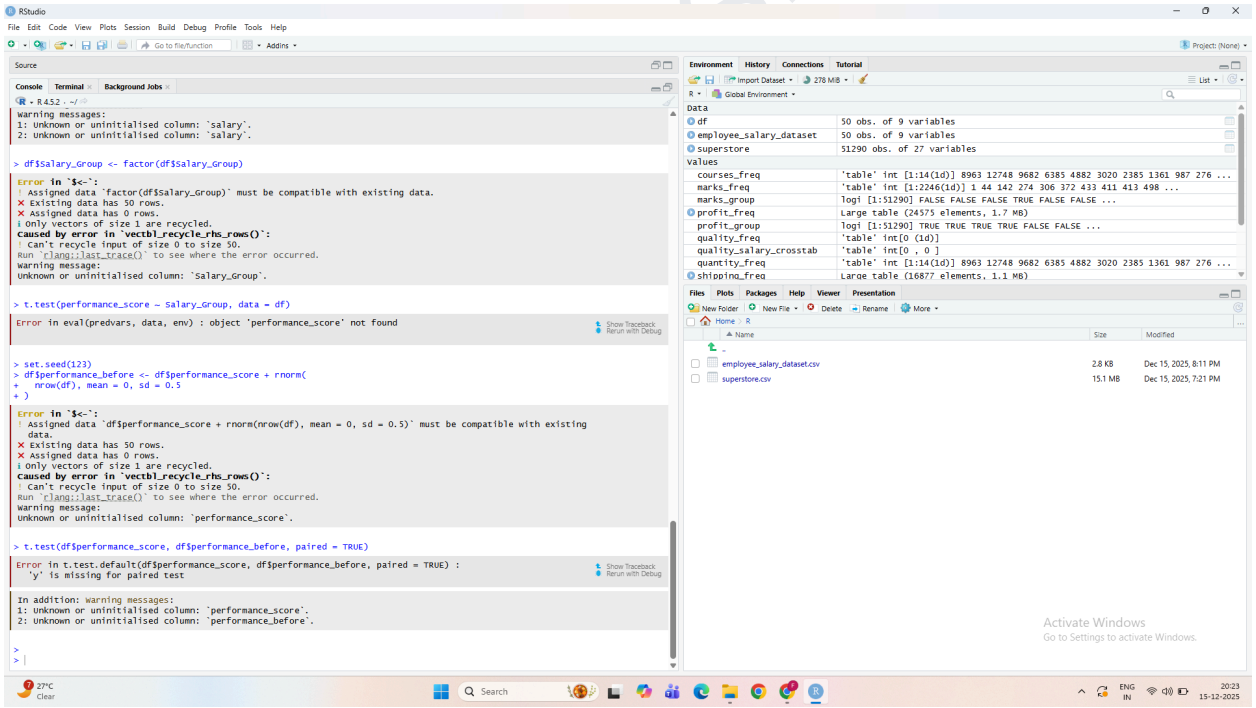


```
> library(dplyr)
> df <- read_csv("R/employee_salary_dataset.csv")
Rows: 50 Columns: 9
── Column specification ──
Delimiter: ","
chr (5): Name, Department, Education_Level, Gender, City
dbl (4): EmployeeID, Experience_Years, Age, Monthly_Salary

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
> head(df)
# A tibble: 6 × 9
  EmployeeID Name       Department Experience_Years Education_Level   Age Gender City      Monthly_Salary
       <dbl> <chr>      <chr>                 <dbl> <chr>           <dbl> <chr>  <chr>             <dbl>
1          1 Employee_1 Marketing                15 Master             53 Female Delhi            111416
2          2 Employee_2 Operations                7 Bachelor           25 Female Bangalore         95271
3          3 Employee_3 IT                       12 High School        51 Female Hyderabad         69064
4          4 Employee_4 Operations                8 PhD                44 Male   Delhi             95091
5          5 Employee_5 Operations               15 Master             36 Female Delhi            132450
6          6 Employee_6 Finance                   3 High School        50 Male   Mumbai            65818
> str(df)
spc_tbl_ [50 × 9] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ EmployeeID      : num [1:50] 1 2 3 4 5 6 7 8 9 10 ...
 $ Name            : chr [1:50] "Employee_1" "Employee_2" "Employee_3" "Employee_4" ...
 $ Department      : chr [1:50] "Marketing" "Operations" "IT" "Operations" ...
 $ Experience_Years: num [1:50] 15 7 12 8 15 3 14 17 4 18 ...
 $ Education_Level : chr [1:50] "Master" "Bachelor" "High School" "PhD" ...
 $ Age             : num [1:50] 53 25 51 44 36 50 57 34 53 28 ...
 $ Gender          : chr [1:50] "Female" "Female" "Female" "Male" ...
 $ City            : chr [1:50] "Delhi" "Bangalore" "Hyderabad" "Delhi" ...
 $ Monthly_Salary  : num [1:50] 111416 95271 69064 95091 132450 ...
 - attr(*, "spec")=
  .. cols(
  ..   EmployeeID = col_double(),
  ..   Name = col_character(),
  ..   Department = col_character(),
  ..   Experience_Years = col_double(),
  ..   Education_Level = col_character(),
  ..   Age = col_double(),
  ..   Gender = col_character(),
  ..   City = col_character(),
  ..   Monthly_Salary = col_double()
  .. )
 - attr(*, "problems")=<externalptr>
> summary(df)
   EmployeeID        Name            Department        Experience_Years Education_Level        Age
 Min.   : 1.00   Length:50          Length:50          Min.   : 1.00    Length:50          Min.   :22.00
 1st Qu.:13.25   Class :character   Class :character   1st Qu.: 5.25    Class :character   1st Qu.:28.25
 Median :25.50   Mode  :character   Mode  :character   Median :10.00    Mode  :character   Median :43.50
 Mean   :25.50                                         Mean   : 9.90                       Mean   :39.76
 3rd Qu.:37.75                                         3rd Qu.:14.75                       3rd Qu.:49.00
 Max.   :50.00                                         Max.   :19.00                       Max.   :57.00
    Gender              City           Monthly_Salary
 Length:50          Length:50          Min.   : 28420
```



```
> df$performance_score <- as.numeric(as.character(df$performance_score))
Error in `$<-`:
! Assigned data `as.numeric(as.character(df$performance_score))` must be compatible with existing data.
✖ Existing data has 50 rows.
✖ Assigned data has 0 rows.
ℹ Only vectors of size 1 are recycled.
Caused by error in `vectbl_recycle_rhs_rows()`:
! Can't recycle input of size 0 to size 50.
Run `rlang::last_trace()` to see where the error occurred.
Warning message:
Unknown or uninitialised column: `performance_score`.

> df <- df %>% filter(!is.na(salary), !is.na(performance_score))
Error in `filter()`:
ℹ In argument: `!is.na(salary)`.
Caused by error:
! object 'salary' not found
Run `rlang::last_trace()` to see where the error occurred.

> summary(df$performance_score)
Length  Class   Mode
     0   NULL   NULL
Warning message:
Unknown or uninitialised column: `performance_score`.

> quality_freq <- table(df$performance_score)
Warning message:
Unknown or uninitialised column: `performance_score`.

> print(quality_freq)
< table of extent 0 >
> quality_count <- df %>% count(performance_score)
Error in `count()`:
! Must group by variables found in `.data`.
✖ Column `performance_score` is not found.
Run `rlang::last_trace()` to see where the error occurred.

> print(quality_count)
Error: object 'quality_count' not found

> quality_salary_crosstab <- table(
+   df$performance_score,
+   df$salary >= median(df$salary)
```


RStudio
> quality_salary_crosstab <- table(
+ df$performance_score,
+ df$salary >= median(df$salary)
+)
Warning messages:
1: Unknown or uninitialised column: `performance_score`.
2: Unknown or uninitialised column: `salary`.
3: Unknown or uninitialised column: `salary`.
> print(quality_salary_crosstab)
< table of extent 0 x 0 >
> t.test(df$performance_score, mu = mean(df$performance_score))
Error in t.test.default(df$performance_score, mu = mean(df$performance_score)) :
'mu' must be a single number
In addition: Warning messages:
1: Unknown or uninitialised column: `performance_score`.
2: Unknown or uninitialised column: `performance_score`.
3: In mean.default(df$performance_score) :
argument is not numeric or logical: returning NA
> df$Salary_Group <- ifelse(df$salary >= median(df$salary), "High", "Low")
Error in `$<-`:
! Assigned data `ifelse(df$salary >= median(df$salary), "High", "Low")` must be compatible with existing data.
✖ Existing data has 50 rows.
✖ Assigned data has 0 rows.
ℹ Only vectors of size 1 are recycled.
Caused by error in `vectbl_recycle_rhs_rows()`:
! Can't recycle input of size 0 to size 50.
Run `rlang::last_trace()` to see where the error occurred.
Warning messages:
1: Unknown or uninitialised column: `salary`.
2: Unknown or uninitialised column: `salary`.
> df$Salary_Group <- factor(df$Salary_Group)
Error in `$<-`:
! Assigned data `factor(df$Salary_Group)` must be compatible with existing data.
✖ Existing data has 50 rows.
✖ Assigned data has 0 rows.
ℹ Only vectors of size 1 are recycled.
Caused by error in `vectbl_recycle_rhs_rows()`:
! Can't recycle input of size 0 to size 50.
Run `rlang::last_trace()` to see where the error occurred.
Warning message:
Unknown or uninitialised column: `Salary_Group`.
Environment History Connections Tutorial
Data
df 50 obs. of 9 variables
employee_salary_dataset 50 obs. of 9 variables
superstore 51290 obs. of 27 variables
Values
courses_freq 'table' int [1:14(1d)] 8963 12748 9682 6385 4882 3020 2385 1361 987 276 ...
marks_freq 'table' int [1:2246(1d)] 1 44 142 274 306 372 433 411 413 498 ...
marks_group 'table' int [1:51290] FALSE FALSE FALSE TRUE FALSE FALSE ...
profit_freq Large table (24575 elements, 1.7 MB)
profit_group logi [1:51290] TRUE TRUE TRUE TRUE FALSE FALSE ...
quality_freq 'table' int[0 (1d)]
quality_salary_crosstab 'table' int[0 , 0]
quantity_freq 'table' int [1:14(1d)] 8963 12748 9682 6385 4882 3020 2385 1361 987 276 ...
shipping_freq Large table (16877 elements, 1.1 MB)
Files Plots Packages Help Viewer Presentation
employee_salary_dataset.csv 2.8 KB Dec 15, 2025, 8:11 PM
superstore.csv 15.1 MB Dec 15, 2025, 7:21 PM
Activate Windows
Go to Settings to activate Windows.


RStudio
Warning messages:
1: Unknown or uninitialised column: `salary`.
2: Unknown or uninitialised column: `salary`.
> df$Salary_Group <- factor(df$Salary_Group)
Error in `$<-`:
! Assigned data `factor(df$Salary_Group)` must be compatible with existing data.
✖ Existing data has 50 rows.
✖ Assigned data has 0 rows.
ℹ Only vectors of size 1 are recycled.
Caused by error in `vectbl_recycle_rhs_rows()`:
! Can't recycle input of size 0 to size 50.
Run `rlang::last_trace()` to see where the error occurred.
Warning message:
Unknown or uninitialised column: `Salary_Group`.
> t.test(performance_score ~ Salary_Group, data = df)
Error in eval(predvars, data, env) : object 'performance_score' not found
> set.seed(123)
> df$performance_before <- df$performance_score + rnorm(
+ nrow(df), mean = 0, sd = 0.5
+)
Error in `$<-`:
! Assigned data `df$performance_score + rnorm(nrow(df), mean = 0, sd = 0.5)` must be compatible with existing data.
✖ Existing data has 50 rows.
✖ Assigned data has 0 rows.
ℹ Only vectors of size 1 are recycled.
Caused by error in `vectbl_recycle_rhs_rows()`:
! Can't recycle input of size 0 to size 50.
Run `rlang::last_trace()` to see where the error occurred.
Warning message:
Unknown or uninitialised column: `performance_score`.
> t.test(df$performance_score, df$performance_before, paired = TRUE)
Error in t.test.default(df$performance_score, df$performance_before, paired = TRUE) :
'y' is missing for paired test
In addition: Warning messages:
1: Unknown or uninitialised column: `performance_score`.
2: Unknown or uninitialised column: `performance_before`.
Environment History Connections Tutorial
Data
df 50 obs. of 9 variables
employee_salary_dataset 50 obs. of 9 variables
superstore 51290 obs. of 27 variables
Values
courses_freq 'table' int [1:14(1d)] 8963 12748 9682 6385 4882 3020 2385 1361 987 276 ...
marks_freq 'table' int [1:2246(1d)] 1 44 142 274 306 372 433 411 413 498 ...
marks_group 'table' int [1:51290] FALSE FALSE FALSE TRUE FALSE FALSE ...
profit_freq Large table (24575 elements, 1.7 MB)
profit_group logi [1:51290] TRUE TRUE TRUE TRUE FALSE FALSE ...
quality_freq 'table' int[0 (1d)]
quality_salary_crosstab 'table' int[0 , 0]
quantity_freq 'table' int [1:14(1d)] 8963 12748 9682 6385 4882 3020 2385 1361 987 276 ...
shipping_freq Large table (16877 elements, 1.1 MB)
Files Plots Packages Help Viewer Presentation
employee_salary_dataset.csv 2.8 KB Dec 15, 2025, 8:11 PM
superstore.csv 15.1 MB Dec 15, 2025, 7:21 PM
Activate Windows
Go to Settings to activate Windows.