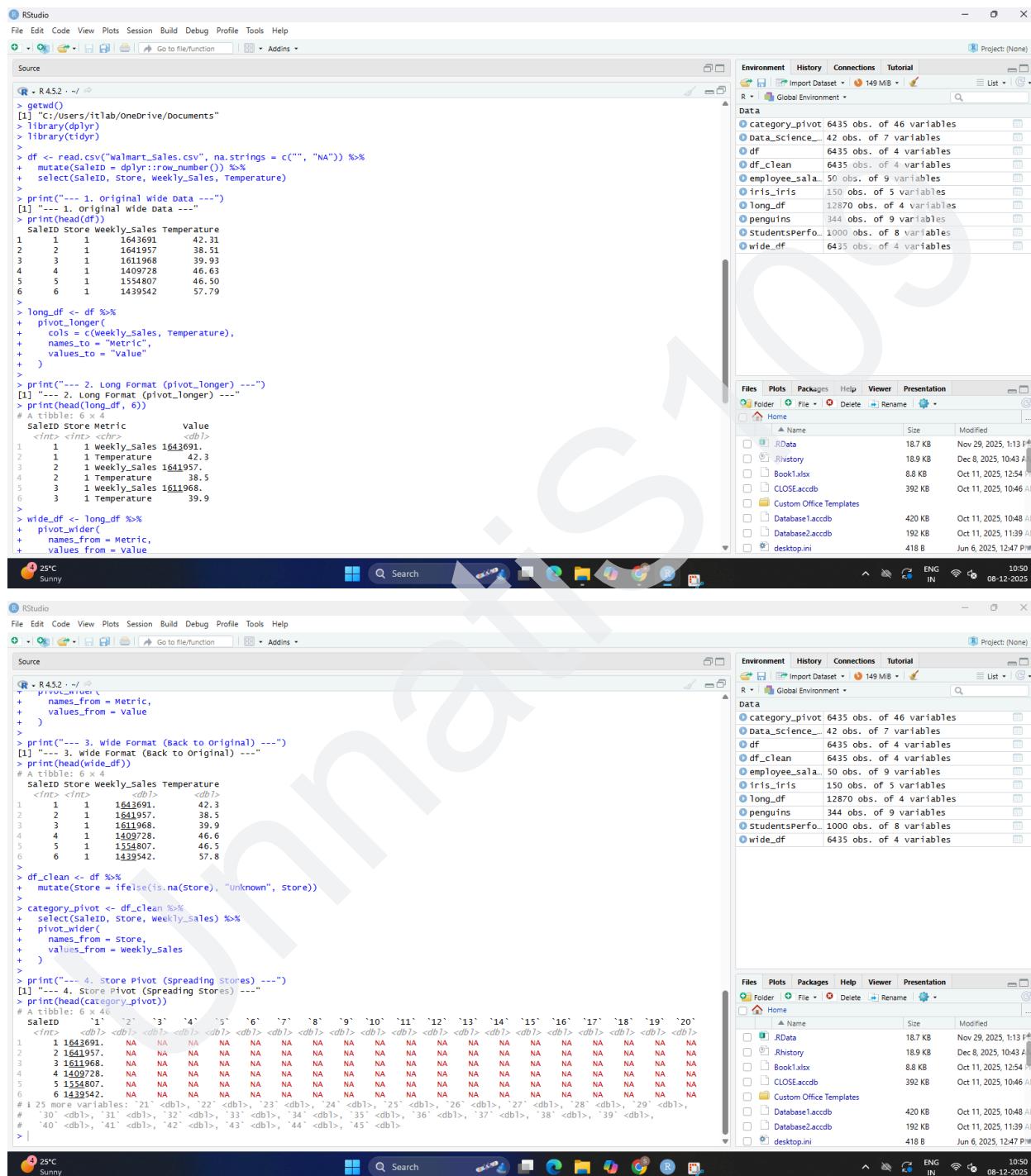


**SHETH L.U.J AND M.V COLLEGE**  
**PRACTICAL NO .11,12,13,14,15**  
**SUBJECT - DATA ANALYSIS**

## 11 Reshaping data using pivot\_longer()/pivot\_wider() (R).

### OUTPUT



```

# RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
File Edit Code View Plots Session Build Debug Profile Tools Help
Project: (None) Environment History Connections Tutorial
R + Global Environment
Data
category_pivot 6435 obs. of 46 variables
data_science_... 42 obs. of 7 variables
df 6435 obs. of 4 variables
df_clean 6435 obs. of 4 variables
employee_sala... 50 obs. of 9 variables
iris_iris 150 obs. of 5 variables
long_df 12870 obs. of 4 variables
penguins 344 obs. of 9 variables
studentsPerfo... 1000 obs. of 8 variables
wide_df 6435 obs. of 4 variables

Files Plots Packages Help Viewer Presentation
Folder File Delete Rename ...
Home Name Size Modified
RData 18.7 KB Nov 29, 2025, 1:13 PM
Rhistory 18.9 KB Dec 8, 2025, 10:43 AM
Book1.xlsx 8.8 KB Oct 11, 2025, 1:25 PM
CLOSE.accdb 392 KB Oct 11, 2025, 10:46 AM
Custom Office Templates
Database1.accdb 420 KB Oct 11, 2025, 10:48 AM
Database2.accdb 192 KB Oct 11, 2025, 1:19 PM
desktop.ini 418 B Jun 6, 2025, 12:47 PM

25°C Sunny 10:50 ENG IN 08-12-2025

```

```

# RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
File Edit Code View Plots Session Build Debug Profile Tools Help
Project: (None) Environment History Connections Tutorial
R + Global Environment
Data
category_pivot 6435 obs. of 46 variables
data_science_... 42 obs. of 7 variables
df 6435 obs. of 4 variables
df_clean 6435 obs. of 4 variables
employee_sala... 50 obs. of 9 variables
iris_iris 150 obs. of 5 variables
long_df 12870 obs. of 4 variables
penguins 344 obs. of 9 variables
studentsPerfo... 1000 obs. of 8 variables
wide_df 6435 obs. of 4 variables

Files Plots Packages Help Viewer Presentation
Folder File Delete Rename ...
Home Name Size Modified
RData 18.7 KB Nov 29, 2025, 1:13 PM
Rhistory 18.9 KB Dec 8, 2025, 10:43 AM
Book1.xlsx 8.8 KB Oct 11, 2025, 1:25 PM
CLOSE.accdb 392 KB Oct 11, 2025, 10:46 AM
Custom Office Templates
Database1.accdb 420 KB Oct 11, 2025, 10:48 AM
Database2.accdb 192 KB Oct 11, 2025, 1:19 PM
desktop.ini 418 B Jun 6, 2025, 12:47 PM

25°C Sunny 10:50 ENG IN 08-12-2025

```

```

# R 4.5.2 -/-
> getwd()
[1] "C:/users/itlab/oneDrive/Documents"
> library(dplyr)
> library(tidyr)
>
> df <- read.csv("walmart_Sales.csv", na.strings = c("", "NA")) %>%
+   mutate(SaleID = dplyr::row_number()) %>%
+   select(-SaleID, -store, weekly_sales, Temperature)
>
> print("--- 1. original wide data ---")
[1] "--- 1. original wide data ---"
> print(head(df))
# A tibble: 6 × 4
  SaleID store weekly_sales Temperature
  1     1      1       1643691    42.31
  2     2      1       1641957    38.51
  3     3      1       1611968    39.93
  4     4      1       1409728    46.63
  5     5      1       1554807    46.50
  6     6      1       1439542    57.79

> long_df <- df %>%
+   pivot_longer(
+     cols = c(weekly_sales, Temperature),
+     names_to = "Metric",
+     values_to = "Value"
+   )
>
> print("--- 2. Long Format (pivot_longer) ---")
[1] "--- 2. Long Format (pivot_longer) ---"
> print(head(long_df, 6))
# A tibble: 6 × 4
  SaleID store Metric      Value
  <int> <chr> <chr>    <dbl>
1     1      1 Weekly_sales 1643691.
2     1      1 Temperature   42.3
3     2      1 Weekly_sales 1641957.
4     2      1 Temperature   38.5
5     3      1 Weekly_sales 1611968.
6     3      1 Temperature   39.9

> wide_df <- long_df %>%
+   pivot_wider(
+     names_from = Metric,
+     values_from = Value
+   )
# R 4.5.2 -/-
> print("---- 3. wide Format (Back to original) ----")
[1] "---- 3. wide Format (Back to original) ----"
> print(head(wide_df))
# A tibble: 6 × 4
  SaleID store weekly_sales Temperature
  <int> <chr> <dbl>    <dbl>
1     1      1       1643691    42.3
2     2      1       1641957    38.5
3     3      1       1611968    39.9
4     4      1       1409728    46.6
5     5      1       1554807    46.5
6     6      1       1439542    57.8

> df_clean <- df %>%
+   mutate(store = ifelse(is.na(store), "unknown", store))

> Category_pivot <- df_clean %>%
+   select(-SaleID, -store, weekly_sales) %>%
+   pivot_wider(
+     names_from = store,
+     values_from = weekly_sales
+   )

> print("---- 4. Store Pivot (Spreading Stores) ----")
[1] "---- 4. Store Pivot (Spreading Stores) ----"
> print(head(category_pivot))
# A tibble: 6 × 46
  SaleID `1` `2` `3` `4` `5` `6` `7` `8` `9` `10` `11` `12` `13` `14` `15` `16` `17` `18` `19` `20` 
  <int> <dbl> 
1 1643691. NA 
2 1641957. NA 
3 1611968. NA 
4 1409728. NA 
5 1554807. NA 
6 1439542. NA 

# 25 more variables: `21` <dbl>, `22` <dbl>, `23` <dbl>, `24` <dbl>, `25` <dbl>, `26` <dbl>, `27` <dbl>, `28` <dbl>, `29` <dbl>, 
# `30` <dbl>, `31` <dbl>, `32` <dbl>, `33` <dbl>, `34` <dbl>, `35` <dbl>, `36` <dbl>, `37` <dbl>, `38` <dbl>, `39` <dbl>, 
# `40` <dbl>, `41` <dbl>, `42` <dbl>, `43` <dbl>, `44` <dbl>, `45` <dbl>
> |

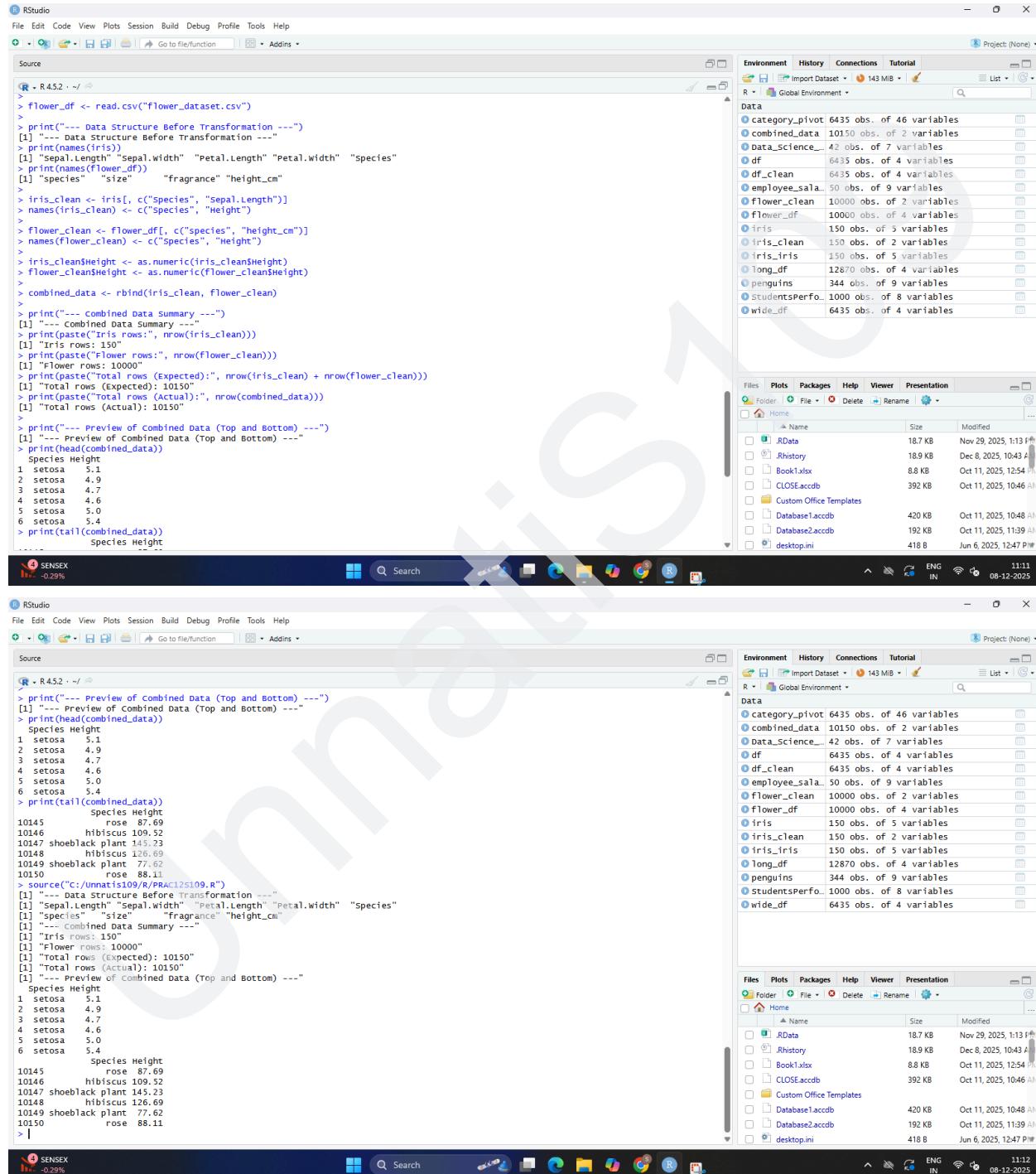
```

## 12 Combining datasets vertically (concatenation) using rbind() (R).

NAME - UNNATI RATHOD  
ROLL NO - S109

**SHETH L.U.J AND M.V COLLEGE**  
**PRACTICAL NO .11,12,13,14,15**  
**SUBJECT - DATA ANALYSIS**

**OUTPUT -**



The screenshot shows two instances of the RStudio interface running on a Windows desktop. Both instances show the same R session output, which is as follows:

```

R > flower_df <- read.csv("flower_dataset.csv")
> 
> print("--- Data Structure Before Transformation ---")
[1] "--- Data Structure Before Transformation ---"
> print(names(iris))
[1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width" "Species"
> print(names(flower_df))
[1] "species"      "size"        "fragrance"   "height_cm"
> 
> iris_clean <- iris[, c("Species", "Sepal.Length")]
> names(iris_clean) <- c("Species", "Height")
> 
> flower_clean <- flower_df[, c("species", "height_cm")]
> names(flower_clean) <- c("Species", "Height")
> 
> iris_clean$height <- as.numeric(iris_clean$height)
> flower_clean$height <- as.numeric(flower_clean$height)
> 
> combined_data <- rbind(iris_clean, flower_clean)
> 
> print("--- Combined Data Summary ---")
[1] "--- Combined Data Summary ---"
> print(paste("Iris rows:", nrow(iris_clean)))
[1] "Iris rows: 150"
> print(paste("Flower rows:", nrow(flower_clean)))
[1] "Flower rows: 10000"
> print(paste("Total rows (Expected):", nrow(iris_clean) + nrow(flower_clean)))
[1] "Total rows (Expected): 10150"
> print(paste("Total rows (Actual):", nrow(combined_data)))
[1] "Total rows (Actual): 10150"
> 
> print("--- Preview of Combined Data (Top and Bottom) ---")
[1] "--- Preview of Combined Data (Top and Bottom) ---"
> print(head(combined_data))
Species Height
1 setosa  5.1
2 setosa  4.9
3 setosa  4.7
4 setosa  4.6
5 setosa  5.0
6 setosa  5.4
> print(tail(combined_data))
Species Height
10145    rose  87.69
10146   hibiscus 109.52
10147 shoebblack plant 145.23
10148   hibiscus 126.69
10149 shoebblack plant 77.62
10150    rose  88.11
> source("C:/unniati109/R/PRAC12S109.R")
[1] "--- Data Structure Before Transformation ---"
[1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width" "Species"
[1] "species"      "size"        "fragrance"   "height_cm"
[1] "--- Combined Data Summary ---"
[1] "Iris rows: 150"
[1] "Flower rows: 10000"
[1] "Total rows (Expected): 10150"
[1] "Total rows (Actual): 10150"
[1] "--- Preview of Combined Data (Top and Bottom) ---"
Species Height
1 setosa  5.1
2 setosa  4.9
3 setosa  4.7
4 setosa  4.6
5 setosa  5.0
6 setosa  5.4
Species Height
10145    rose  87.69
10146   hibiscus 109.52
10147 shoebblack plant 145.23
10148   hibiscus 126.69
10149 shoebblack plant 77.62
10150    rose  88.11
  
```

The RStudio interface includes a top menu bar, a source code editor, a global environment viewer, a file browser, and a system tray at the bottom.

**13 Identifying and handling duplicates using distinct() (R).**

**NAME - UNNATI RATHOD**  
**ROLL NO - S109**

**SHETH L.U.J AND M.V COLLEGE  
PRACTICAL NO .11,12,13,14,15  
SUBJECT - DATA ANALYSIS**

## OUTPUT

## 14 Extracting date components using lubridate:: functions (R). OUTPUT

NAME - UNNATI RATHOD  
ROLL NO - S109

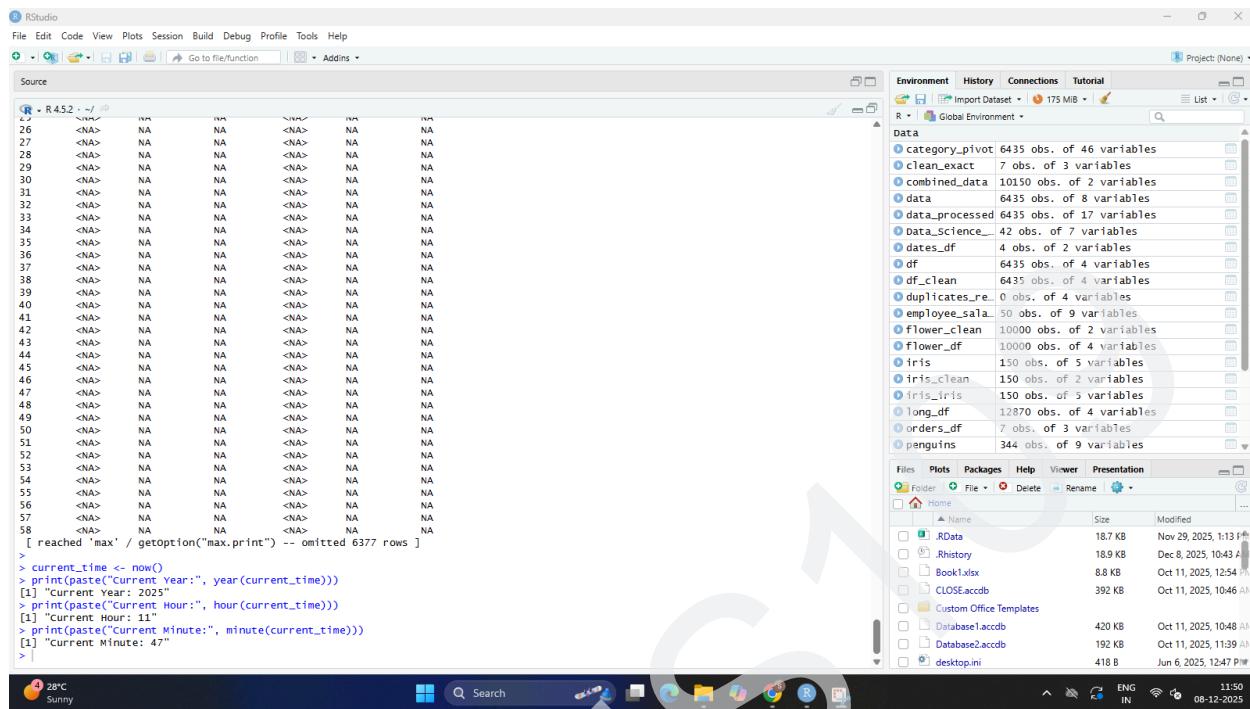
**SHETH L.U.J AND M.V COLLEGE  
PRACTICAL NO .11,12,13,14,15  
SUBJECT - DATA ANALYSIS**

The screenshot shows the RStudio interface with the following details:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Addins:** Go to file/function, Addins.
- Environment Tab:** Shows the global environment with various objects listed, such as category\_pivot, clean\_exact, combined\_data, data\_processed, data\_science\_, dates\_df, df, duplicates\_re, employee\_sala, flower\_clean, flower\_df, iris, iris\_clean, iris\_iris, long\_df, orders\_df, and penguins.
- Source Tab:** Displays a large dataset (R-452) with columns: Month, Name, Day, Num, weekday, Num, weekday, Quarter, Day\_of\_Year.
- Search Bar:** Located at the bottom center of the interface.

NAME - UNNATI RATHOD  
ROLL NO - S109

**SHETH L.U.J AND M.V COLLEGE**  
**PRACTICAL NO .11,12,13,14,15**  
**SUBJECT - DATA ANALYSIS**



```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins ...
Source
R > R 4.5.2 ~/ ...
26 <NA> NA NA <NA> NA NA
27 <NA> NA NA <NA> NA NA
28 <NA> NA NA <NA> NA NA
29 <NA> NA NA <NA> NA NA
30 <NA> NA NA <NA> NA NA
31 <NA> NA NA <NA> NA NA
32 <NA> NA NA <NA> NA NA
33 <NA> NA NA <NA> NA NA
34 <NA> NA NA <NA> NA NA
35 <NA> NA NA <NA> NA NA
36 <NA> NA NA <NA> NA NA
37 <NA> NA NA <NA> NA NA
38 <NA> NA NA <NA> NA NA
39 <NA> NA NA <NA> NA NA
40 <NA> NA NA <NA> NA NA
41 <NA> NA NA <NA> NA NA
42 <NA> NA NA <NA> NA NA
43 <NA> NA NA <NA> NA NA
44 <NA> NA NA <NA> NA NA
45 <NA> NA NA <NA> NA NA
46 <NA> NA NA <NA> NA NA
47 <NA> NA NA <NA> NA NA
48 <NA> NA NA <NA> NA NA
49 <NA> NA NA <NA> NA NA
50 <NA> NA NA <NA> NA NA
51 <NA> NA NA <NA> NA NA
52 <NA> NA NA <NA> NA NA
53 <NA> NA NA <NA> NA NA
54 <NA> NA NA <NA> NA NA
55 <NA> NA NA <NA> NA NA
56 <NA> NA NA <NA> NA NA
57 <NA> NA NA <NA> NA NA
58 <NA> NA NA <NA> NA NA
[ reached 'max' / getoption("max.print") -- omitted 6377 rows ]
> current_time <- now()
> print(paste("Current Year:", year(current_time)))
[1] "Current Year: 2025"
> print(paste("Current Hour:", hour(current_time)))
[1] "Current Hour: 11"
> print(paste("Current Minute:", minute(current_time)))
[1] "Current Minute: 47"
>

```

Files Plots Packages Help Viewer Presentation

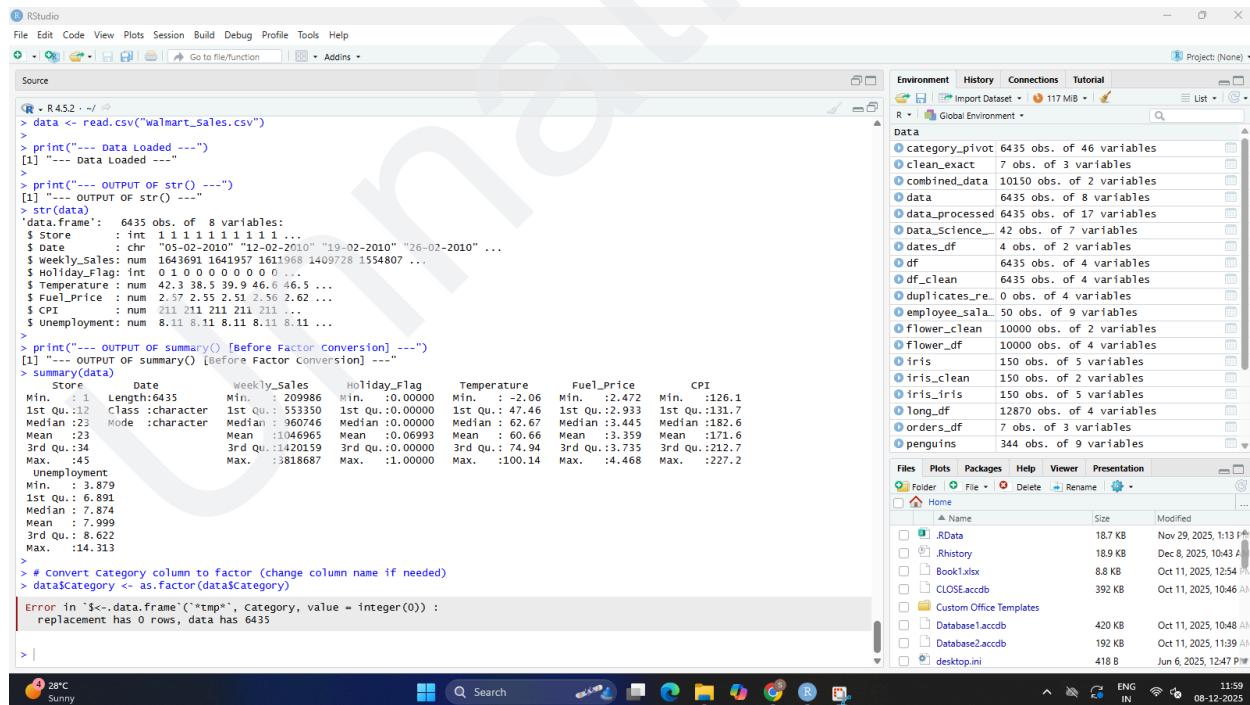
Home Name Size Modified

- RData 18.7 KB Nov 29, 2025, 1:13 PM
- Rhistory 18.9 KB Dec 8, 2025, 10:43 AM
- Book1.xlsx 8.8 KB Oct 11, 2025, 12:54 PM
- CLOSE.accdb 392 KB Oct 11, 2025, 10:46 AM
- Custom Office Templates
- Database1.accdb 420 KB Oct 11, 2025, 10:48 AM
- Database2.accdb 192 KB Oct 11, 2025, 11:39 AM
- desktop.ini 418 B Jun 6, 2025, 12:47 PM

28°C Sunny 11:50 ENG IN 08-12-2025

## 15 Generating basic summaries using str() or summary() (R).

### OUTPUT



```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins ...
Source
R > R 4.5.2 ~/ ...
> data <- read.csv("walmart_sales.csv")
> print("--- Data Loaded ---")
[1] "--- Data Loaded ---"
> print("--- OUTPUT OF str() ---")
[1] "--- OUTPUT OF str() ---"
> str(data)
'data.frame': 6435 obs. of 8 variables:
 $ Store      : int 1 1 1 1 1 1 1 1 1 ...
 $ Date       : chr "05-02-2010" "12-02-2010" "19-02-2010" "26-02-2010" ...
 $ Weekly_Sales: num 164097 1641957 1611968 1409728 1554807 ...
 $ StoreDayFlag: num 0 0 0 0 0 ...
 $ Temperature: num -12.3 38.5 39.9 46.6 46.5 ...
 $ Fuel_Price  : num 2.57 2.55 2.51 2.56 2.62 ...
 $ CPI        : num 211.211 211.211 211.211 ...
 $ Unemployment: num 8.11 8.11 8.11 8.11 8.11 ...
>
> print("--- OUTPUT OF summary() (Before Factor Conversion) ---")
[1] "--- OUTPUT OF summary() (Before Factor Conversion) ---"
> summary(data)
Store      Date      weekly_sales   holiday_flag   Temperature   Fuel_Price   Min.CPI
Min.   : 1   Length:6435   Min.   :20986   Min.   :0.00000   Min.   :-2.06   Min.   :12.472   Min.   :126.1
1st Qu.:12   Class :character 1st Qu.:553350  1st Qu.:0.00000  1st Qu.:47.46   1st Qu.:2.933   1st Qu.:131.7
Median :23   Mode  :character Median :960746  Median :0.00000  Median :62.67   Median :3.445   Median :182.6
Mean   :23   Mean   :1046965  Mean   :0.06993  Mean   :60.66   Mean   :3.359   Mean   :171.6
3rd Qu.:34   3rd Qu.:1420159 3rd Qu.:0.00000  3rd Qu.:74.94   3rd Qu.:3.735   3rd Qu.:212.7
Max.   :45   Max.   :3818687  Max.   :1.00000  Max.   :100.14   Max.   :4.468   Max.   :227.2
Unemployment
Min.   : 3.879
1st Qu.: 6.124
Median : 8.74
Mean   : 7.999
3rd Qu.: 8.622
Max.   :14.313
> # convert Category column to factor (change column name if needed)
> data$Category <- as.factor(data$Category)

Error in `<- data.frame`(*tmp*, Category, value = integer(0)) :
  replacement has 0 rows, data has 6435
>

```

Files Plots Packages Help Viewer Presentation

Home Name Size Modified

- RData 18.7 KB Nov 29, 2025, 1:13 PM
- Rhistory 18.9 KB Dec 8, 2025, 10:43 AM
- Book1.xlsx 8.8 KB Oct 11, 2025, 12:54 PM
- CLOSE.accdb 392 KB Oct 11, 2025, 10:46 AM
- Custom Office Templates
- Database1.accdb 420 KB Oct 11, 2025, 10:48 AM
- Database2.accdb 192 KB Oct 11, 2025, 11:39 AM
- desktop.ini 418 B Jun 6, 2025, 12:47 PM

28°C Sunny 11:59 ENG IN 08-12-2025