

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
Go to file/function Addins
Source
R - R4.3.2 - ~/
> getwd()
[1] "C:/Users/itlab/OneDrive/Documents"
> library(dplyr)
> library(tidyverse)
>
> 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))
  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 x 4
  saleID Store Metric      Value
  <int> <int> <chr> <dbl>
1      1    1 weekly_Sales 1643691.
2      2    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
+   )
>
> print("--- 3. Wide Format (Back to original) ---")
[1] "--- 3. Wide Format (Back to original) ---"
> print(head(wide_df))
# A tibble: 6 x 4
  saleID Store weekly_Sales Temperature
  <int> <int> <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 x 46
  saleID   1'   2'   3'   4'   5'   6'   7'   8'   9'  10'  11'  12'  13'  14'  15'  16'  17'  18'  19'  20'
  <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1      1 1643691. NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
2      2 1641957. NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
3      3 1611968. NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
4      4 1409728. NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
5      5 1554807. NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
6      6 1439542. NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  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>
>
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
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.accd 392 KB Oct 11, 2025, 10:46 AM
Custom Office Templates
Database1.accd 420 KB Oct 11, 2025, 10:48 AM
Database2.accd 192 KB Oct 11, 2025, 11:39 AM
desktop.ini 418 B Jun 6, 2025, 12:47 PM
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
R - R4.3.2 - ~/
> pivot_wider(
+   names_from = Metric,
+   values_from = Value
+ )
>
> print("--- 3. Wide Format (Back to original) ---")
[1] "--- 3. Wide Format (Back to original) ---"
> print(head(wide_df))
# A tibble: 6 x 4
  saleID Store weekly_Sales Temperature
  <int> <int> <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 x 46
  saleID   1'   2'   3'   4'   5'   6'   7'   8'   9'  10'  11'  12'  13'  14'  15'  16'  17'  18'  19'  20'
  <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1      1 1643691. NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
2      2 1641957. NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
3      3 1611968. NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
4      4 1409728. NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
5      5 1554807. NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
6      6 1439542. NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  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>
>
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
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.accd 392 KB Oct 11, 2025, 10:46 AM
Custom Office Templates
Database1.accd 420 KB Oct 11, 2025, 10:48 AM
Database2.accd 192 KB Oct 11, 2025, 11:39 AM
desktop.ini 418 B Jun 6, 2025, 12:47 PM
```

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

```
R - R4.5.2 ~ /
> flower_df <- read.csv("flower_dataset.csv")
> print("---- 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 shoeblack plant 145.23
10148 hibiscus 126.69
10149 shoeblack plant 77.62
10150 rose    88.11
```

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

```
R - R4.5.2 ~ /
> 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 shoeblack plant 145.23
10148 hibiscus 126.69
10149 shoeblack plant 77.62
10150 rose    88.11
> source("C:/Unnatis109/R/PRACT12S109.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 shoeblack plant 145.23
10148 hibiscus 126.69
10149 shoeblack plant 77.62
10150 rose    88.11
> |
```

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

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins

Source
R - R4.5.2 - ~/
> library(dplyr)
> orders_df <- data.frame(
+   orderID = c(101, 102, 102, 103, 104, 101, 104),
+   customer = c("unnati", "kirti", "divya", "Geet", "Bhumika", "Aishwarya", "Parvati"),
+   Product = c("Laptop", "Phone", "Phone", "Tablet", "Monitor", "Laptop", "Mouse")
+ )
> print("---- 1. Original Dataset (Note 7 rows) ----")
[1] "---- 1. Original Dataset (Note 7 rows) ----"
> print(orders_df)
  orderID customer Product
1     101  unnati  Laptop
2     102   kirti   Phone
3     102   divya   Phone
4     103    Geet   Tablet
5     104  Bhumika  Monitor
6     101 Aishwarya  Laptop
7     104  Parvati   Mouse
>
> duplicates_report <- orders_df %>%
+   group_by(orderID, customer, Product) %>%
+   count() %>%
+   filter(n > 1)
>
> print("---- 2. Identification Report (Rows that are duplicated) ----")
[1] "---- 2. Identification Report (Rows that are duplicated) ----"
> print(duplicates_report)
# A tibble: 0 x 4
# Groups:   orderID, customer, Product [0]
# 4 variables: orderID <dbl>, customer <chr>, Product <chr>, n <int>
>
> clean_exact <- orders_df %>%
+   distinct()
>
> print("---- 3. Removed Exact Duplicates (distinct) ----")
[1] "---- 3. Removed Exact Duplicates (distinct) ----"
> print(clean_exact)
  orderID customer Product
1     101  unnati  Laptop
2     102   kirti   Phone
3     102   divya   Phone
4     103    Geet   Tablet
5     104  Bhumika  Monitor
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins

Source
R - R4.5.2 - ~/
>
> duplicates_report <- orders_df %>%
+   group_by(orderID, customer, Product) %>%
+   count() %>%
+   filter(n > 1)
>
> print("---- 2. Identification Report (Rows that are duplicated) ----")
[1] "---- 2. Identification Report (Rows that are duplicated) ----"
> print(duplicates_report)
# A tibble: 0 x 4
# Groups:   orderID, customer, Product [0]
# 4 variables: orderID <dbl>, customer <chr>, Product <chr>, n <int>
>
> clean_exact <- orders_df %>%
+   distinct()
>
> print("---- 3. Removed Exact Duplicates (distinct) ----")
[1] "---- 3. Removed Exact Duplicates (distinct) ----"
> print(clean_exact)
  orderID customer Product
1     101  unnati  Laptop
2     102   kirti   Phone
3     102   divya   Phone
4     103    Geet   Tablet
5     104  Bhumika  Monitor
6     101 Aishwarya  Laptop
7     104  Parvati   Mouse
>
> unique_customers <- orders_df %>%
+   distinct(customer, .keep_all = TRUE)
>
> print("---- 4. Unique Customers Only (Partial Duplicates removed) ----")
[1] "---- 4. Unique Customers Only (Partial Duplicates removed) ----"
> print(unique_customers)
  orderID customer Product
1     101  unnati  Laptop
2     102   kirti   Phone
3     102   divya   Phone
4     103    Geet   Tablet
5     104  Bhumika  Monitor
6     101 Aishwarya  Laptop
7     104  Parvati   Mouse
>
>
>
>
```

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 image displays two screenshots of the RStudio interface. The top screenshot shows the Source pane with R code for data processing, including a warning message about the 'mutate()' function. The bottom screenshot shows the same RStudio interface with the Environment pane open, displaying a list of data objects and their sizes.

Top Screenshot: Source Pane

```
R - R452 - ~/r
> # Ensure the column name matches the CSV (change if needed)
> # Suppose the date column is named "date"
> data_processed <- data %>%
+   mutate(
+     Actual_Date = ymd(Date),
+     Year_Num = year(Actual_Date),
+     Month_Num = month(Actual_Date),
+     Month_Name = month(Actual_Date, label = TRUE),
+     Day_Num = day(Actual_Date),
+     weekday_Num = wday(Actual_Date),
+     weekday_Name = wday(Actual_Date, label = TRUE, abbr = FALSE),
+     Quarter = quarter(Actual_Date),
+     Day_of_Year = yday(Actual_Date)
+   )
Warning message:
There was 1 warning in 'mutate()':
! in argument: 'Actual_Date = ymd(Date)'.
Caused by warning:
! All formats failed to parse. No formats found.

> print(data_processed)
```

Store	Date	weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment	Actual_Date	Year_Num	Month_Num
1	05-02-2010	1643691	0	42.31	2.572	211.0964	8.106	<NA>	NA	NA
2	12-02-2010	1641957	1	38.51	2.548	211.2422	8.106	<NA>	NA	NA
3	19-02-2010	1611968	0	39.93	2.514	211.2891	8.106	<NA>	NA	NA
4	26-02-2010	1409728	0	46.63	2.561	211.3196	8.106	<NA>	NA	NA
5	05-03-2010	1554807	0	46.50	2.625	211.3501	8.106	<NA>	NA	NA
6	12-03-2010	1439542	0	57.79	2.667	211.3806	8.106	<NA>	NA	NA
7	19-03-2010	1472516	0	54.58	2.720	211.2156	8.106	<NA>	NA	NA
8	26-03-2010	1404430	0	51.45	2.732	211.0180	8.106	<NA>	NA	NA
9	02-04-2010	1594968	0	62.27	2.719	210.8204	7.808	<NA>	NA	NA
10	09-04-2010	1545419	0	65.86	2.770	210.6229	7.808	<NA>	NA	NA
11	16-04-2010	1466058	0	66.32	2.808	210.4887	7.808	<NA>	NA	NA
12	23-04-2010	1391256	0	64.84	2.795	210.4391	7.808	<NA>	NA	NA
13	30-04-2010	1425101	0	67.41	2.780	210.3895	7.808	<NA>	NA	NA
14	07-05-2010	1603955	0	72.55	2.835	210.3400	7.808	<NA>	NA	NA
15	14-05-2010	1494252	0	74.78	2.854	210.3374	7.808	<NA>	NA	NA
16	21-05-2010	1399662	0	76.44	2.826	210.6171	7.808	<NA>	NA	NA
17	28-05-2010	1432070	0	80.44	2.759	210.8968	7.808	<NA>	NA	NA
18	04-06-2010	1615525	0	80.69	2.705	211.1764	7.808	<NA>	NA	NA
19	11-06-2010	1542561	0	80.43	2.668	211.4561	7.808	<NA>	NA	NA
20	18-06-2010	1503284	0	84.11	2.637	211.4538	7.808	<NA>	NA	NA
21	25-06-2010	1422712	0	84.34	2.653	211.3387	7.808	<NA>	NA	NA

Bottom Screenshot: Environment Pane

Object	Size	Modified
category_pivot	6435 obs. of 46 variables	
clean_exact	7 obs. of 3 variables	
combined_data	10150 obs. of 2 variables	
data	6435 obs. of 8 variables	
data_processed	6435 obs. of 17 variables	
data_science	42 obs. of 7 variables	
dates_df	4 obs. of 2 variables	
df	6435 obs. of 4 variables	
df_clean	6435 obs. of 4 variables	
duplicates_re	0 obs. of 4 variables	
employee_sal	50 obs. of 9 variables	
flower_clean	10000 obs. of 2 variables	
flower_df	10000 obs. of 4 variables	
iris	150 obs. of 5 variables	
iris_clean	150 obs. of 5 variables	
iris_iris	150 obs. of 5 variables	
long_df	12870 obs. of 4 variables	
orders_df	7 obs. of 3 variables	
penguins	344 obs. of 9 variables	

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
R452 ~ /
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"
> |
```

Environment History Connections Tutorial
R - Global Environment
Data
category_pivot 6435 obs. of 46 variables
clean_exact 7 obs. of 3 variables
combined_data 10150 obs. of 2 variables
data 6435 obs. of 8 variables
data_processed 6435 obs. of 17 variables
Data_Science_ 42 obs. of 7 variables
dates_df 4 obs. of 2 variables
df 6435 obs. of 4 variables
df_clean 6435 obs. of 4 variables
duplicates_re 0 obs. of 4 variables
employee_sala 50 obs. of 9 variables
flower_clean 10000 obs. of 2 variables
flower_df 10000 obs. of 4 variables
iris 150 obs. of 5 variables
iris_clean 150 obs. of 5 variables
iris_iris 150 obs. of 5 variables
long_df 12870 obs. of 4 variables
orders_df 7 obs. of 3 variables
penguins 344 obs. of 9 variables

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.accd 392 KB Oct 11, 2025, 10:46 AM
Custom Office Templates
Database1.accd 420 KB Oct 11, 2025, 10:48 AM
Database2.accd 192 KB Oct 11, 2025, 11:39 AM
desktop.ini 418 B Jun 6, 2025, 12:47 PM

28°C Sunny

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
R452 ~ /
> 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 1 ...
 $ Date : chr "05-02-2010" "12-02-2010" "19-02-2010" "26-02-2010" ...
 $ weekly_sales: num 1643691 1641957 1611968 1409728 1554807 ...
 $ Holiday_Flag: int 0 1 0 0 0 0 0 0 0 ...
 $ Temperature: num 42.3 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 ...
 $ 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 CPI
Min. :1 Length:6435 Min. : 209986 Min. :0.00000 Min. : -2.06 Min. :2.472 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.: 6.891
Median : 7.874
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
> |
```

Environment History Connections Tutorial
R - Global Environment
Data
category_pivot 6435 obs. of 46 variables
clean_exact 7 obs. of 3 variables
combined_data 10150 obs. of 2 variables
data 6435 obs. of 8 variables
data_processed 6435 obs. of 17 variables
Data_Science_ 42 obs. of 7 variables
dates_df 4 obs. of 2 variables
df 6435 obs. of 4 variables
df_clean 6435 obs. of 4 variables
duplicates_re 0 obs. of 4 variables
employee_sala 50 obs. of 9 variables
flower_clean 10000 obs. of 2 variables
flower_df 10000 obs. of 4 variables
iris 150 obs. of 5 variables
iris_clean 150 obs. of 5 variables
iris_iris 150 obs. of 5 variables
long_df 12870 obs. of 4 variables
orders_df 7 obs. of 3 variables
penguins 344 obs. of 9 variables

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.accd 392 KB Oct 11, 2025, 10:46 AM
Custom Office Templates
Database1.accd 420 KB Oct 11, 2025, 10:48 AM
Database2.accd 192 KB Oct 11, 2025, 11:39 AM
desktop.ini 418 B Jun 6, 2025, 12:47 PM

28°C Sunny

NAME - UNNATI RATHOD
ROLL NO - S109