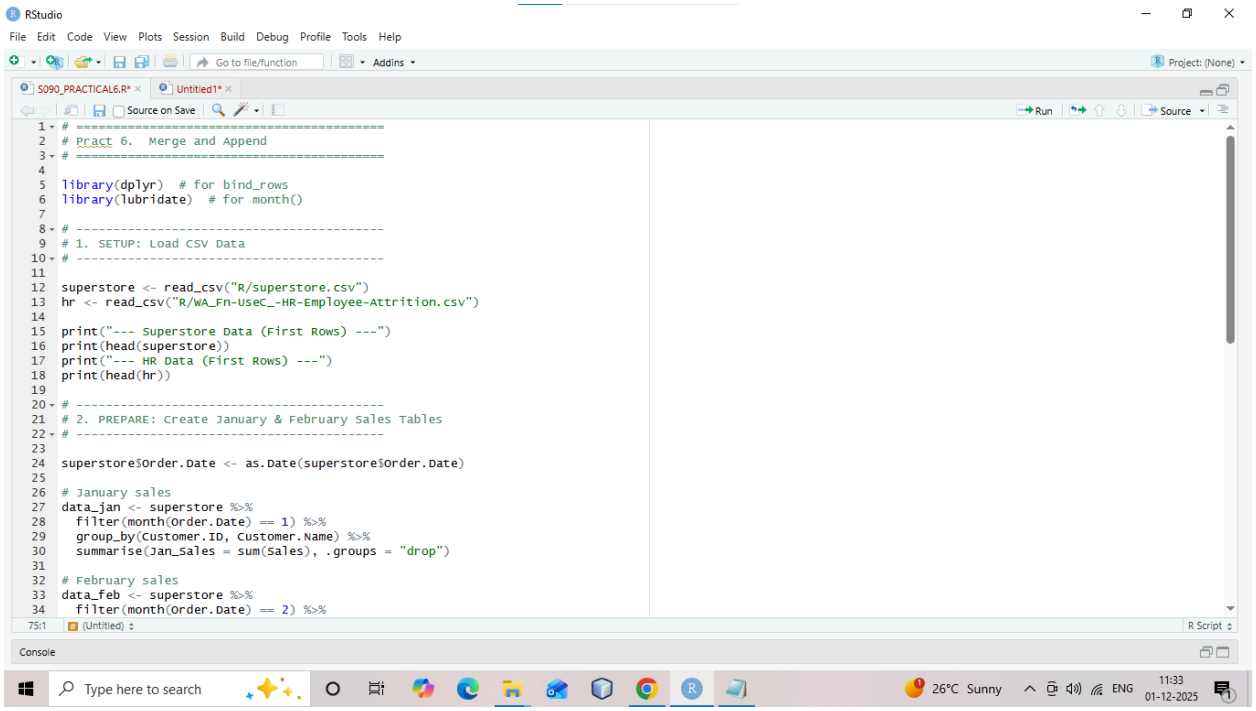
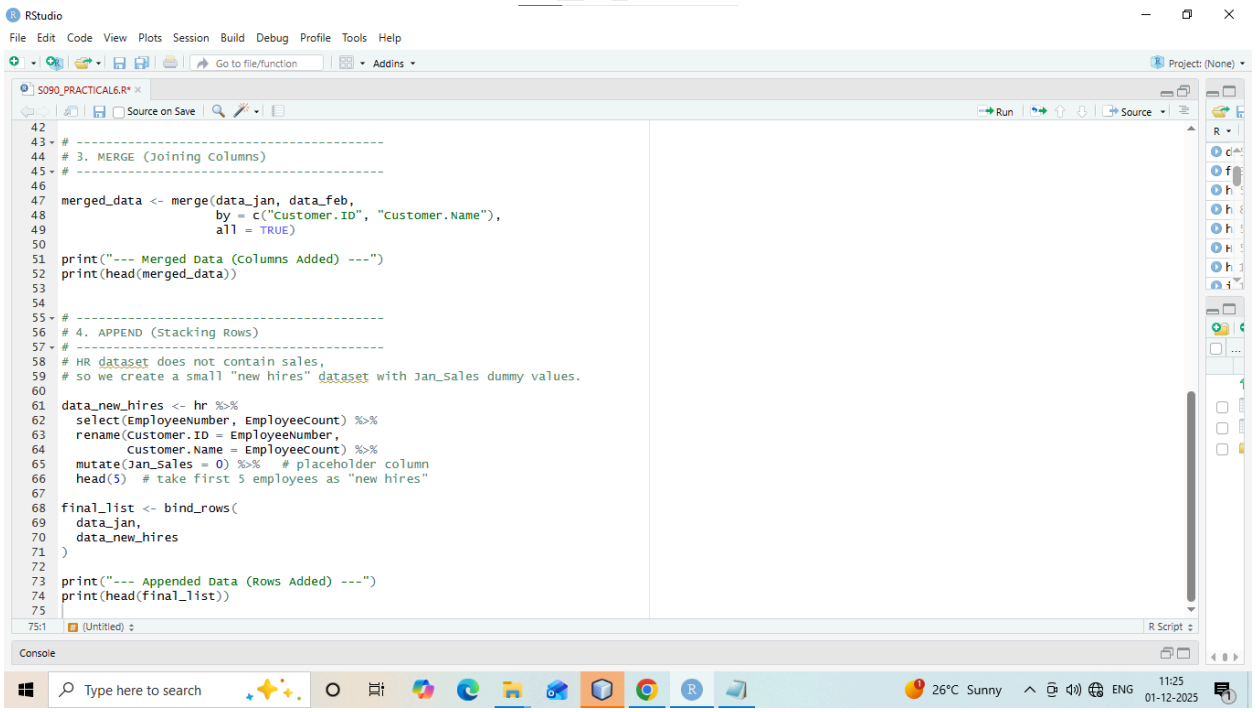
PRACTICAL NO 6

AIM: Combining and appending datasets using merge() or bind_rows() in R.

CODE:

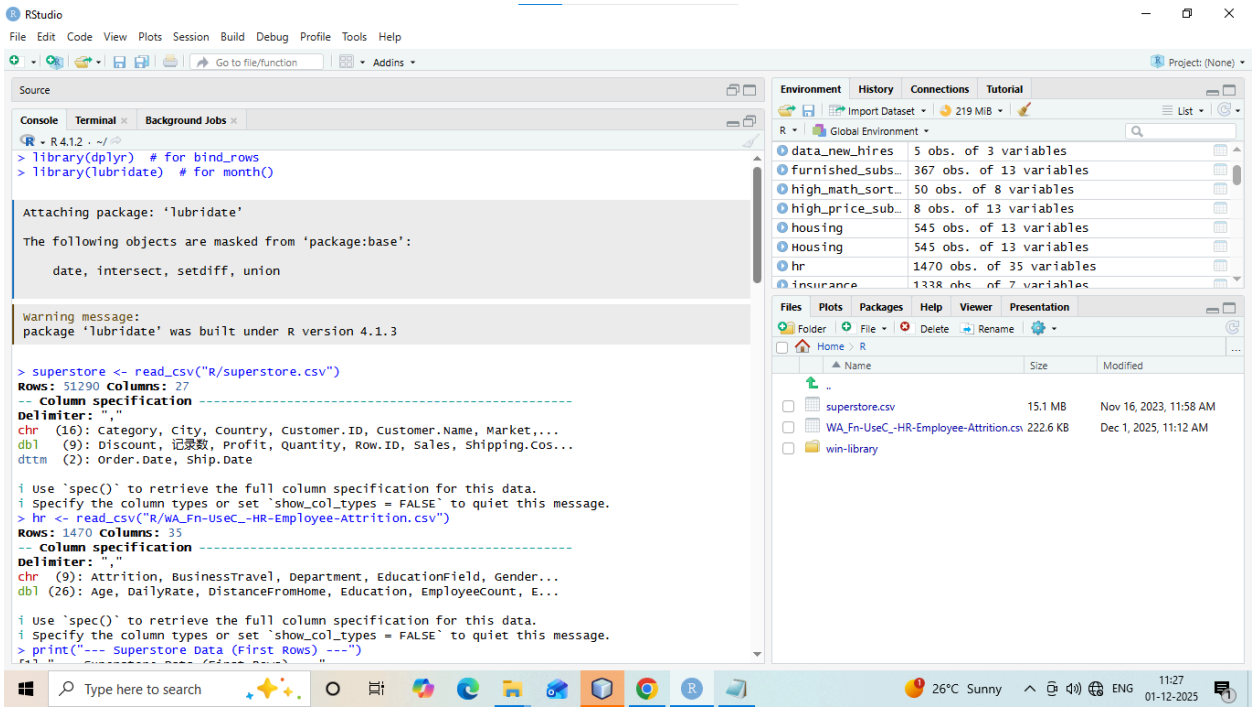


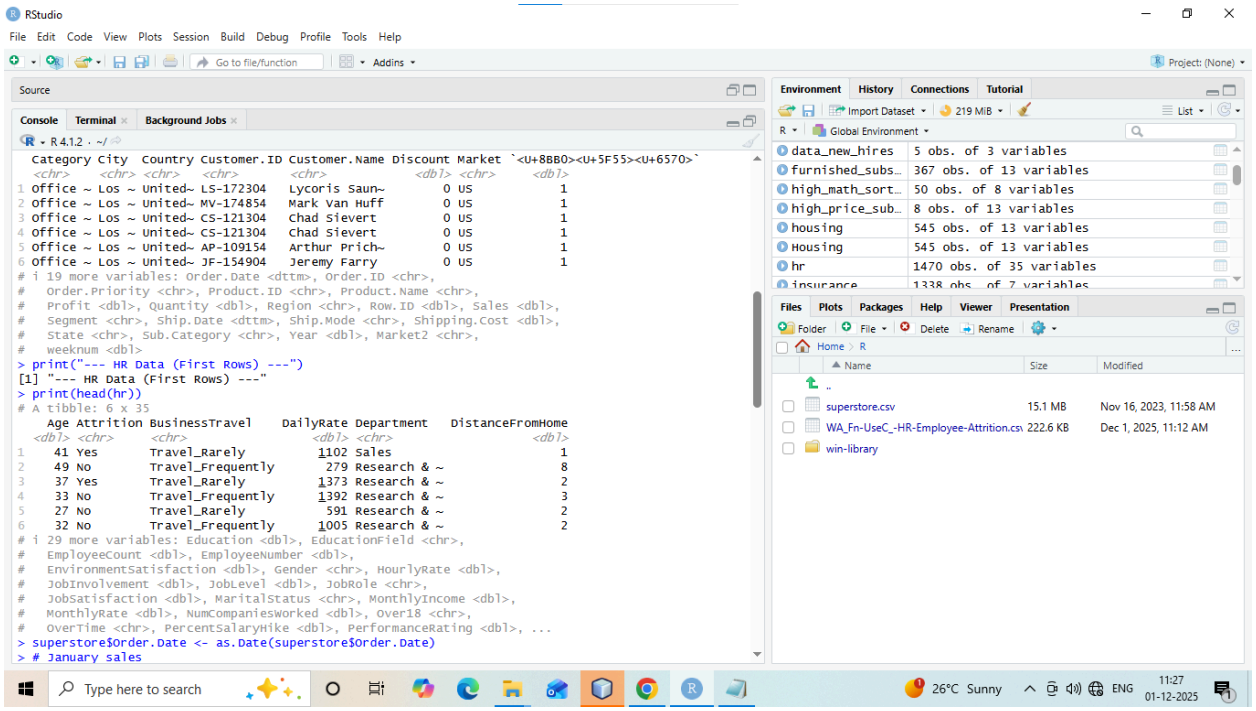
```
# ==========================================
# Pract 6.  Merge and Append
# ==========================================

library(dplyr)  # for bind_rows
library(lubridate)  # for month()

# ------------------------------------------
# 1. SETUP: Load CSV Data
# ------------------------------------------

superstore <- read_csv("R/superstore.csv")
hr <- read_csv("R/WA_Fn-UseC_-HR-Employee-Attrition.csv")

print("--- Superstore Data (First Rows) ---")
print(head(superstore))
print("--- HR Data (First Rows) ---")
print(head(hr))

# ------------------------------------------
# 2. PREPARE: Create January & February Sales Tables
# ------------------------------------------

superstore$Order.Date <- as.Date(superstore$Order.Date)

# January sales
data_jan <- superstore %>%
  filter(month(Order.Date) == 1) %>%
  group_by(Customer.ID, Customer.Name) %>%
  summarise(Jan_Sales = sum(Sales), .groups = "drop")

# February sales
data_feb <- superstore %>%
  filter(month(Order.Date) == 2) %>%
```

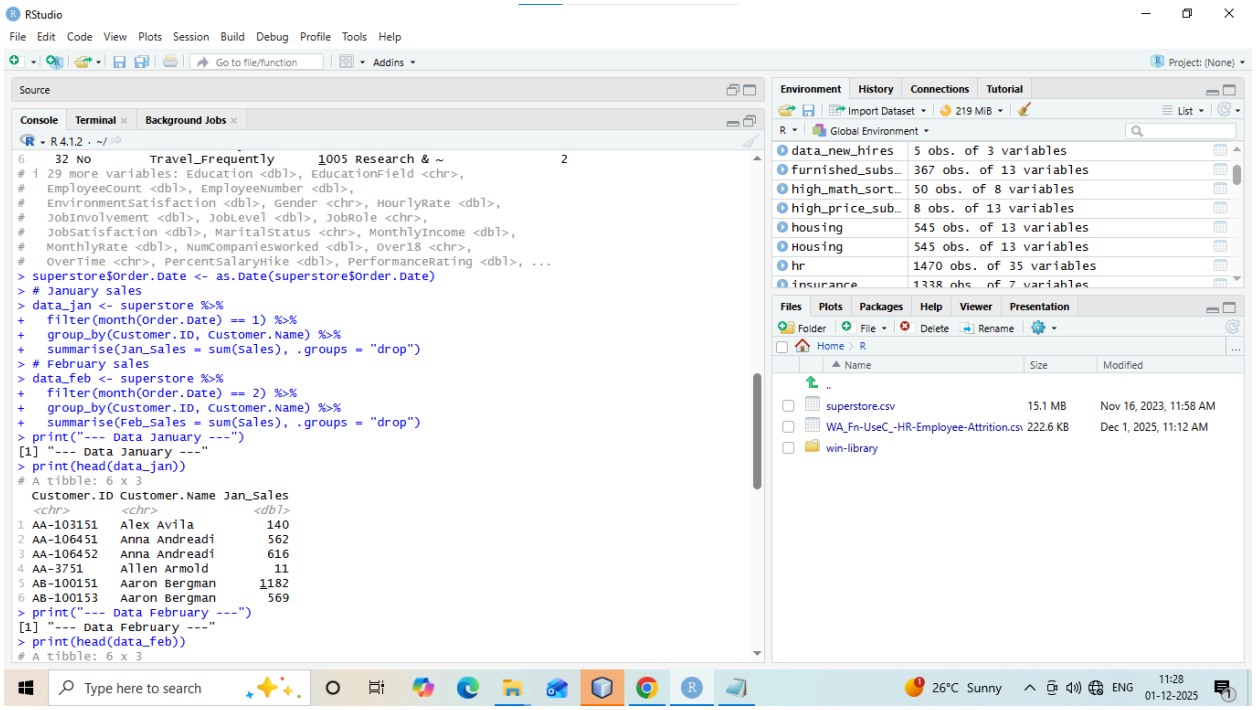


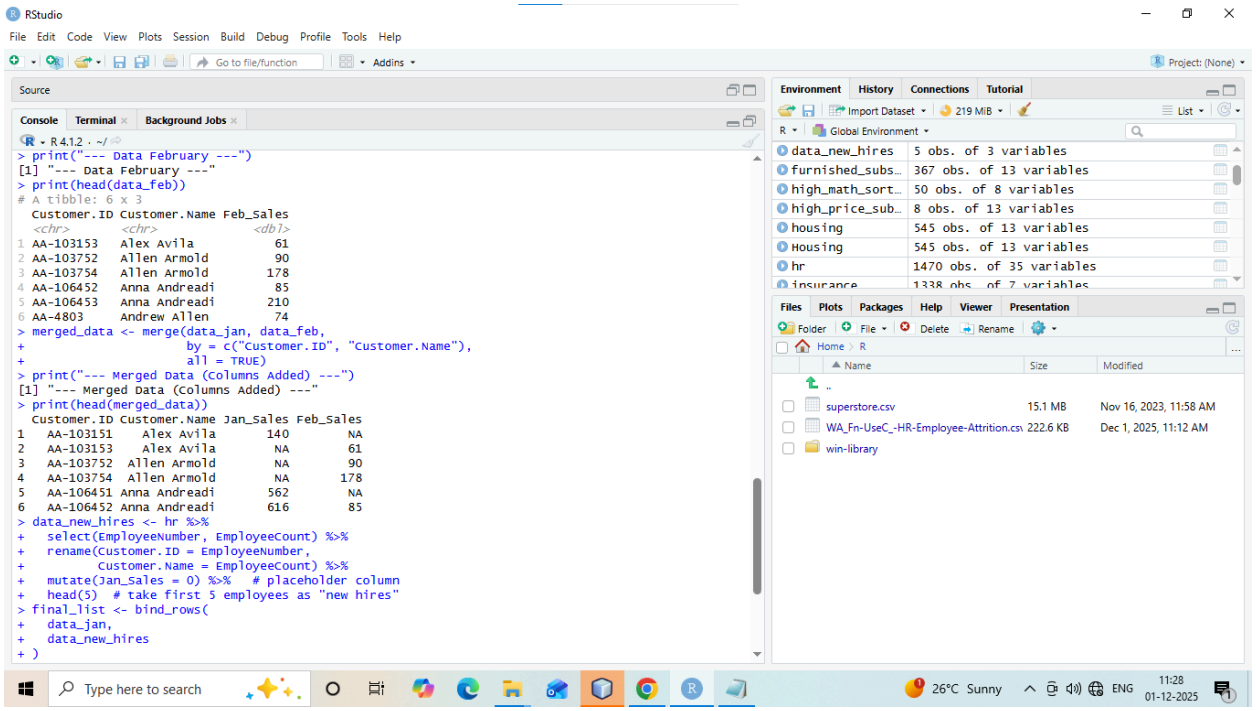
```

# ------------------------------------------
# 3. MERGE (Joining Columns)
# ------------------------------------------

merged_data <- merge(data_jan, data_feb,
                     by = c("Customer.ID", "Customer.Name"),
                     all = TRUE)

print("--- Merged Data (Columns Added) ---")
print(head(merged_data))


# ------------------------------------------
# 4. APPEND (Stacking Rows)
# ------------------------------------------
# HR dataset does not contain sales,
# so we create a small "new hires" dataset with Jan_Sales dummy values.

data_new_hires <- hr %>%
  select(EmployeeNumber, EmployeeCount) %>%
  rename(Customer.ID = EmployeeNumber,
         Customer.Name = EmployeeCount) %>%
  mutate(Jan_Sales = 0) %>%   # placeholder column
  head(5)  # take first 5 employees as "new hires"

final_list <- bind_rows(
  data_jan,
  data_new_hires
)

print("--- Appended Data (Rows Added) ---")
print(head(final_list))
```

OUTPUT:


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
R 4.1.2 · ~/
> library(dplyr) # for bind_rows
> library(lubridate) # for month()
Attaching package: 'lubridate'
The following objects are masked from 'package:base':
date, intersect, setdiff, union
Warning message:
package 'lubridate' was built under R version 4.1.3
> superstore <- read_csv("R/superstore.csv")
Rows: 51290 Columns: 27
-- Column specification --
Delimiter: ","
chr (16): Category, City, Country, Customer.ID, Customer.Name, Market,...
dbl (9): Discount, 记录数, Profit, Quantity, Row.ID, Sales, Shipping.Cos...
dttm (2): Order.Date, Ship.Date
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.
> hr <- read_csv("R/WA_Fn-UseC_-HR-Employee-Attrition.csv")
Rows: 1470 Columns: 35
-- Column specification --
Delimiter: ","
chr (9): Attrition, BusinessTravel, Department, EducationField, Gender...
dbl (26): Age, DailyRate, DistanceFromHome, Education, EmployeeCount, E...
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.
> print("--- Superstore Data (First Rows) ---")
Environment History Connections Tutorial
Import Dataset 219 MiB
Global Environment
data_new_hires 5 obs. of 3 variables
furnished_subs... 367 obs. of 13 variables
high_math_sort... 50 obs. of 8 variables
high_price_sub... 8 obs. of 13 variables
housing 545 obs. of 13 variables
Housing 545 obs. of 13 variables
hr 1470 obs. of 35 variables
Files Plots Packages Help Viewer Presentation
Folder File Delete Rename
Home R
Name Size Modified
superstore.csv 15.1 MB Nov 16, 2023, 11:58 AM
WA_Fn-UseC_-HR-Employee-Attrition.csv 222.6 KB Dec 1, 2025, 11:12 AM
win-library
Type here to search
26°C Sunny
ENG
11:27
01-12-2025


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
R 4.1.2 · ~/
Category City Country Customer.ID Customer.Name Discount Market `<U+8BB0><U+5F55><U+6570>`
1 Office ~ Los ~ United~ LS-172304 Lycoris Saun~ 0 US 1
2 Office ~ Los ~ United~ MV-174854 Mark Van Huff 0 US 1
3 Office ~ Los ~ United~ CS-121304 Chad Sievert 0 US 1
4 Office ~ Los ~ United~ CS-121304 Chad Sievert 0 US 1
5 Office ~ Los ~ United~ AP-109154 Arthur Prich~ 0 US 1
6 Office ~ Los ~ United~ JF-154904 Jeremy Farry 0 US 1
i 19 more variables: Order.Date <dttm>, Order.ID <chr>,
Order.Priority <chr>, Product.ID <chr>, Product.Name <chr>,
Profit <dbl>, Quantity <dbl>, Region <chr>, Row.ID <dbl>, Sales <dbl>,
Segment <chr>, Ship.Date <dttm>, Ship.Mode <chr>, Shipping.Cost <dbl>,
State <chr>, Sub.Category <chr>, Year <dbl>, Market2 <chr>,
weeknum <dbl>
> print("--- HR Data (First Rows) ---")
[1] "--- HR Data (First Rows) ---"
> print(head(hr))
A tibble: 6 x 35
Age Attrition BusinessTravel DailyRate Department DistanceFromHome
41 Yes Travel_Rarely 1102 Sales 1
49 No Travel_Frequently 279 Research & ~ 8
37 Yes Travel_Rarely 1373 Research & ~ 2
33 No Travel_Frequently 1392 Research & ~ 3
27 No Travel_Rarely 591 Research & ~ 2
32 No Travel_Frequently 1005 Research & ~ 2
i 29 more variables: Education <dbl>, EducationField <chr>,
EmployeeCount <dbl>, EmployeeNumber <dbl>,
EnvironmentSatisfaction <dbl>, Gender <chr>, HourlyRate <dbl>,
JobInvolvement <dbl>, JobLevel <dbl>, JobRole <chr>,
JobSatisfaction <dbl>, MaritalStatus <chr>, MonthlyIncome <dbl>,
MonthlyRate <dbl>, NumCompaniesWorked <dbl>, Over18 <chr>,
OverTime <chr>, PercentSalaryHike <dbl>, PerformanceRating <dbl>, ...
> superstore$Order.Date <- as.Date(superstore$Order.Date)
> # January sales
Environment History Connections Tutorial
Import Dataset 219 MiB
Global Environment
data_new_hires 5 obs. of 3 variables
furnished_subs... 367 obs. of 13 variables
high_math_sort... 50 obs. of 8 variables
high_price_sub... 8 obs. of 13 variables
housing 545 obs. of 13 variables
Housing 545 obs. of 13 variables
hr 1470 obs. of 35 variables
Files Plots Packages Help Viewer Presentation
Folder File Delete Rename
Home R
Name Size Modified
superstore.csv 15.1 MB Nov 16, 2023, 11:58 AM
WA_Fn-UseC_-HR-Employee-Attrition.csv 222.6 KB Dec 1, 2025, 11:12 AM
win-library
Type here to search
26°C Sunny
ENG
11:27
01-12-2025


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
R 4.1.2 · ~/
6 32 No Travel_Frequently 1005 Research & ~ 2
i 29 more variables: Education <dbl>, EducationField <chr>,
EmployeeCount <dbl>, EmployeeNumber <dbl>,
EnvironmentSatisfaction <dbl>, Gender <chr>, HourlyRate <dbl>,
JobInvolvement <dbl>, JobLevel <dbl>, JobRole <chr>,
JobSatisfaction <dbl>, MaritalStatus <chr>, MonthlyIncome <dbl>,
MonthlyRate <dbl>, NumCompaniesWorked <dbl>, Over18 <chr>,
OverTime <chr>, PercentSalaryHike <dbl>, PerformanceRating <dbl>, ...
> superstore$Order.Date <- as.Date(superstore$Order.Date)
> # January sales
> data_jan <- superstore %>%
+ filter(month(Order.Date) == 1) %>%
+ group_by(Customer.ID, Customer.Name) %>%
+ summarise(Jan_Sales = sum(Sales), .groups = "drop")
> # February sales
> data_feb <- superstore %>%
+ filter(month(Order.Date) == 2) %>%
+ group_by(Customer.ID, Customer.Name) %>%
+ summarise(Feb_Sales = sum(Sales), .groups = "drop")
> print("--- Data January ---")
[1] "--- Data January ---"
> print(head(data_jan))
A tibble: 6 x 3
Customer.ID Customer.Name Jan_Sales
<chr> <chr> <dbl>
1 AA-103151 Alex Avila 140
2 AA-106451 Anna Andreadi 562
3 AA-106452 Anna Andreadi 616
4 AA-3751 Allen Arnold 11
5 AB-100151 Aaron Bergman 1182
6 AB-100153 Aaron Bergman 569
> print("--- Data February ---")
[1] "--- Data February ---"
> print(head(data_feb))
A tibble: 6 x 3
Environment History Connections Tutorial
Import Dataset 219 MiB
Global Environment
data_new_hires 5 obs. of 3 variables
furnished_subs… 367 obs. of 13 variables
high_math_sort… 50 obs. of 8 variables
high_price_sub… 8 obs. of 13 variables
housing 545 obs. of 13 variables
Housing 545 obs. of 13 variables
hr 1470 obs. of 35 variables
insurance 1338 obs. of 7 variables
Files Plots Packages Help Viewer Presentation
Folder File Delete Rename
Home R
Name Size Modified
superstore.csv 15.1 MB Nov 16, 2023, 11:58 AM
WA_Fn-UseC_-HR-Employee-Attrition.csv 222.6 KB Dec 1, 2025, 11:12 AM
win-library
Type here to search
26°C Sunny
ENG
11:28
01-12-2025


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
R 4.1.2 · ~/
> print("--- Data February ---")
[1] "--- Data February ---"
> print(head(data_feb))
A tibble: 6 x 3
Customer.ID Customer.Name Feb_Sales
<chr> <chr> <dbl>
1 AA-103153 Alex Avila 61
2 AA-103752 Allen Arnold 90
3 AA-103754 Allen Arnold 178
4 AA-106452 Anna Andreadi 85
5 AA-106453 Anna Andreadi 210
6 AA-4803 Andrew Allen 74
> merged_data <- merge(data_jan, data_feb,
+ by = c("Customer.ID", "Customer.Name"),
+ all = TRUE)
> print("--- Merged Data (Columns Added) ---")
[1] "--- Merged Data (Columns Added) ---"
> print(head(merged_data))
Customer.ID Customer.Name Jan_Sales Feb_Sales
1 AA-103151 Alex Avila 140 NA
2 AA-103153 Alex Avila NA 61
3 AA-103752 Allen Arnold NA 90
4 AA-103754 Allen Arnold NA 178
5 AA-106451 Anna Andreadi 562 NA
6 AA-106452 Anna Andreadi 616 85
> data_new_hires <- hr %>%
+ select(EmployeeNumber, EmployeeCount) %>%
+ rename(Customer.ID = EmployeeNumber,
+ Customer.Name = EmployeeCount) %>%
+ mutate(Jan_Sales = 0) %>% # placeholder column
+ head(5) # take first 5 employees as "new hires"
> final_list <- bind_rows(
+ data_jan,
+ data_new_hires
+)
Environment History Connections Tutorial
Import Dataset 219 MiB
Global Environment
data_new_hires 5 obs. of 3 variables
furnished_subs… 367 obs. of 13 variables
high_math_sort… 50 obs. of 8 variables
high_price_sub… 8 obs. of 13 variables
housing 545 obs. of 13 variables
Housing 545 obs. of 13 variables
hr 1470 obs. of 35 variables
insurance 1338 obs. of 7 variables
Files Plots Packages Help Viewer Presentation
Folder File Delete Rename
Home R
Name Size Modified
superstore.csv 15.1 MB Nov 16, 2023, 11:58 AM
WA_Fn-UseC_-HR-Employee-Attrition.csv 222.6 KB Dec 1, 2025, 11:12 AM
win-library
Type here to search
26°C Sunny
ENG
11:28
01-12-2025