

## PRACTICAL NO 12

AIM: Combining datasets vertically (concatenation) using rbind() (R).

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
R - R4.1.2 - ~/
> library(readr)
> library(dplyr)
> # ---- 1. Load files (adjust paths if necessary) ----
> housing_df <- read_csv("R/Housing.csv", show_col_types = FALSE)
> store_df <- read_csv("R/superstore.csv", show_col_types = FALSE)
> cat("--- Column Names ---\n")
--- Column Names
> cat("Housing columns:\n"); print(names(housing_df))
Housing columns:
[1] "price" "area" "bedrooms" "bathrooms" "stories" "mainroad"
[7] "guestroom" "basement" "hotwaterheating" "airconditioning" "parking" "prefarea"
[13] "furnishingstatus"
> cat("Store columns:\n"); print(names(store_df))
Store columns:
[1] "Category" "City" "Country" "Customer.ID" "Customer.Name" "Discount"
[7] "Market" "记录数" "Order.Date" "Order.ID" "Order.Priority" "Product.ID"
[13] "Product.Name" "Profit" "Quantity" "Region" "Row.ID" "Sales"
[19] "Segment" "Ship.Date" "Ship.Mode" "Shipping.Cost" "State" "Sub.category"
[25] "Year" "Market2" "weeknum"
> # ---- 2. Prepare housing_clean ----
> # Choose a categorical column for 'Species' and numeric for 'Height'
> # Based on your housing columns: we'll use 'furnishingstatus' as Species and 'area' as Height
> if (!all(c("furnishingstatus", "area") %in% names(housing_df))) {
+ stop("Expected columns 'furnishingstatus' and/or 'area' not found in housing_df. Check column names.")
+ }
> housing_clean <- housing_df %>%
+ select(Species = furnishingstatus, Height = area) %>%
+ mutate(
+ Species = as.character(Species),
+ Height = as.numeric(Height)
+ )
> # ---- 3. Prepare store_clean ----
> # Use Category as Species and Sales as Height (you already used these)

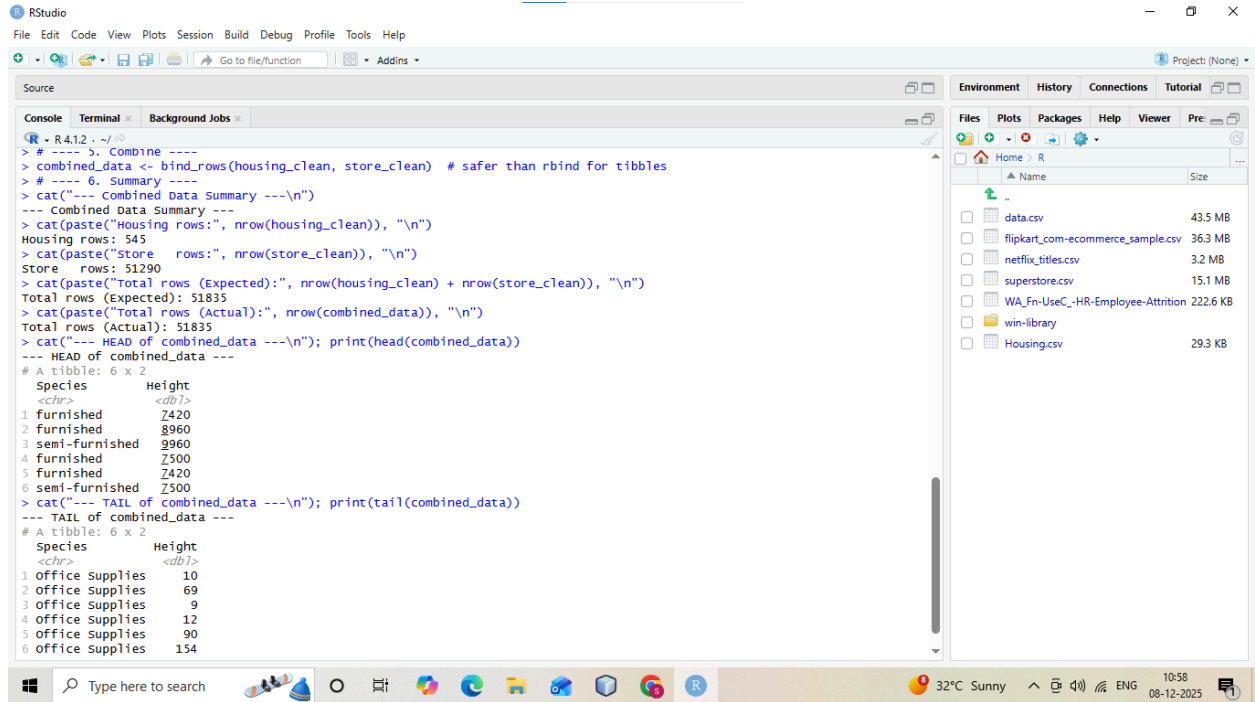
```

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
R - R4.1.2 - ~/
> # Use Category as Species and Sales as Height (you already used these)
> if (!all(c("Category", "Sales") %in% names(store_df))) {
+ stop("Expected columns 'Category' and/or 'Sales' not found in store_df. Check column names.")
+ }
> store_clean <- store_df %>%
+ select(Species = Category, Height = Sales) %>%
+ mutate(
+ Species = as.character(Species),
+ Height = as.numeric(Height)
+ )
> # ---- 4. optional: preview and clean NAs ----
> cat("--- Housing preview ---\n"); print(head(housing_clean))
--- Housing preview ---
# A tibble: 6 x 2
Species Height
<chr> <dbl>
1 furnished 7420
2 furnished 8960
3 semi-furnished 9960
4 furnished 7500
5 furnished 7420
6 semi-furnished 7500
> cat("--- Store preview ---\n"); print(head(store_clean))
--- Store preview ---
# A tibble: 6 x 2
Species Height
<chr> <dbl>
1 Office Supplies 19
2 Office Supplies 19
3 Office Supplies 21
4 Office Supplies 111
5 Office Supplies 6
6 Office Supplies 13
> # Remove rows where Height is NA (if any)
> housing_clean <- housing_clean %>% filter(!is.na(Height))

```

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



The screenshot shows the RStudio environment with the following components:

- Source Editor:** Contains R code for combining data from 'housing\_clean' and 'store\_clean' into 'combined\_data' using `bind_rows()`. It includes summary statistics and head/tail views of the resulting tibble.
- Console:** Displays the output of the R code, showing row counts and the structure of the combined data tibble.
- Environment:** Lists the objects in the environment, including 'data.csv', 'flipkart\_com-e-commerce\_sample.csv', 'netflix\_titles.csv', 'superstore.csv', 'WA\_Fn-UseC\_HR-Employee-Attrition', 'win-library', and 'Housing.csv'.
- Files:** Shows the file explorer with the same list of files as the Environment pane.

```
> # ---- 5. Combine ----
> combined_data <- bind_rows(housing_clean, store_clean) # safer than rbind for tibbles
> # ---- 6. Summary ----
> cat("--- Combined Data Summary ---\n")
--- Combined Data Summary ---
> cat(paste("Housing rows:", nrow(housing_clean)), "\n")
Housing rows: 545
> cat(paste("Store rows:", nrow(store_clean)), "\n")
Store rows: 51290
> cat(paste("Total rows (Expected):", nrow(housing_clean) + nrow(store_clean)), "\n")
Total rows (Expected): 51835
> cat(paste("Total rows (Actual):", nrow(combined_data)), "\n")
Total rows (Actual): 51835
> cat("--- HEAD of combined_data ---\n"); print(head(combined_data))
--- HEAD of combined_data ---
# A tibble: 6 x 2
  Species      Height
  <chr>      <dbl>
1 furnished    2420
2 furnished    8960
3 semi-furnished 9960
4 furnished    2500
5 furnished    2420
6 semi-furnished 2500
> cat("--- TAIL of combined_data ---\n"); print(tail(combined_data))
--- TAIL of combined_data ---
# A tibble: 6 x 2
  Species      Height
  <chr>      <dbl>
1 office Supplies 10
2 office Supplies 69
3 office Supplies 9
4 office Supplies 12
5 office Supplies 90
6 office Supplies 154
```