

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

PRACTICAL NO 12

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

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
R - R 4.1.2 . ~/ ...
> 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
Source
Console Terminal Background Jobs
R - R 4.1.2 . ~/ ...
> # 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     2420
2 furnished     8960
3 semi-furnished 9960
4 furnished     2500
5 furnished     2420
6 semi-furnished 2500
> 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 interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help, and Addins. The main area has tabs for Source, Terminal, and Background Jobs. The Source tab displays R code for combining datasets and printing their summaries. The right panel shows a file browser with several CSV files listed under the 'R' directory. The bottom taskbar includes a search bar, system icons, and a status bar showing the date (08-12-2025), time (10:58), and weather (32°C Sunny).

```
R - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source Terminal Background Jobs
> # ----- 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
```