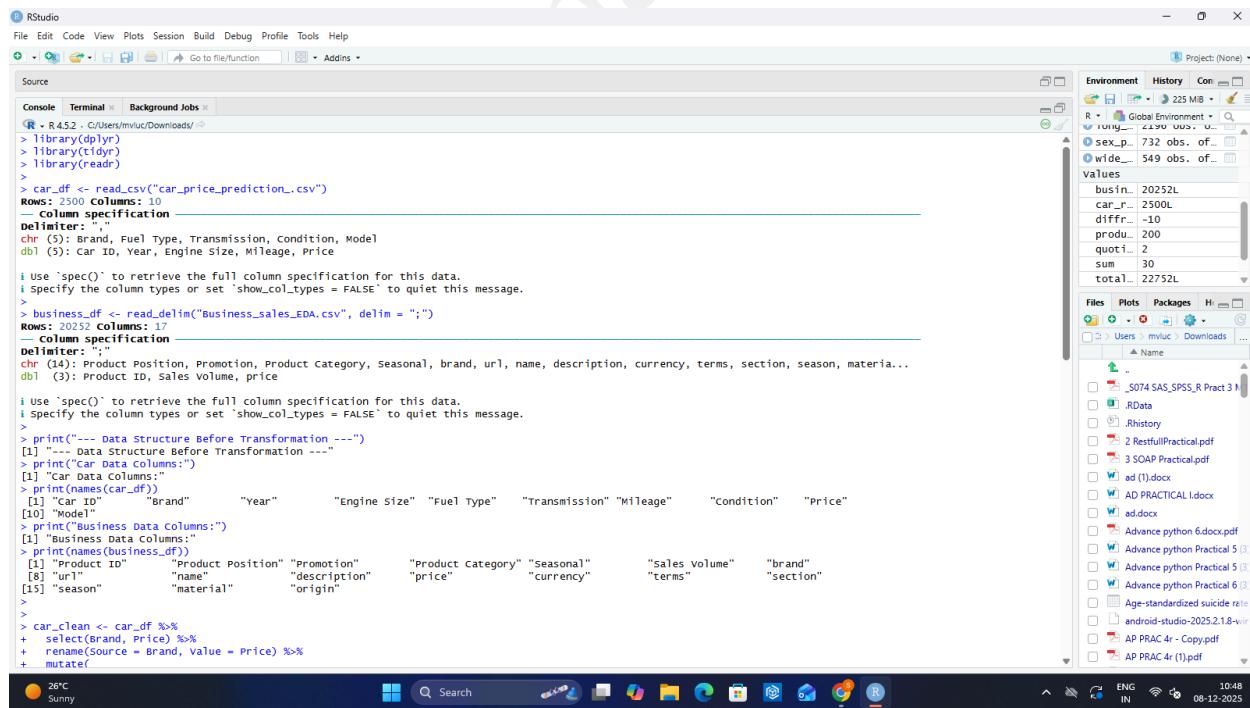


SHETH L.U.J AND SIR M.V. COLLEGE
SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R

Module 1 Practical 12

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

OUTPUT:



The screenshot shows the RStudio interface with the following details:

- Console:** Displays R code for reading CSV files and combining them using the `rbind` function.
- Environment:** Shows variables defined in the session, including `busin_` (2025L), `car_r_` (2500L), `differ_` (-10), `produ_` (200), `quoti_` (2), `sum` (30), and `total_` (22752L).
- Files:** Shows a list of files in the current directory, including various PDFs and Word documents related to practical sessions.
- Plots:** No plots are visible in this screenshot.
- Packages:** No packages are visible in this screenshot.
- Help:** No help documentation is visible in this screenshot.

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source Terminal Background Jobs
> library(dplyr)
> library(tidyverse)
> library(readr)
>
> car_df <- read_csv("car_price_prediction_.csv")
Rows: 2500 Columns: 10
--- column specification ---
Delimiter: ","
chr (5): Brand, Fuel Type, Transmission, Condition, Model
dbl (5): Car ID, Year, Engine Size, Mileage, Price
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.
>
> business_df <- read_delim("Business_sales_EDA.csv", delim = ";")
Rows: 20252 Columns: 17
--- column specification ---
Delimiter: ","
chr (14): Product Position, Promotion, Product Category, Seasonal, brand, url, name, description, currency, terms, section, season, material...
dbl (3): Product ID, Sales Volume, Price
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("--- Data Structure Before Transformation ---")
[1] "--- Data Structure Before Transformation ---"
> print("Car Data columns:")
[1] "Car data columns:"
> print(names(car_df))
[1] "Car ID"      "Brand"       "Year"        "Engine Size"  "Fuel Type"   "Transmission" "Mileage"     "Condition"   "Price"
[10] "Model"
> print("Business Data columns:")
[1] "Business Data columns:"
> print(names(business_df))
[1] "Product ID"    "Product Position" "Promotion"      "Product Category" "Seasonal"      "Sales Volume"  "Brand"        "Section"
[8] "URL"          "Name"          "Description"   "Price"          "Currency"     "Terms"        "Material"     "Origin"
[15] "Season"        "Material"      "Origin"
>
> car_clean <- car_df %>%
+   select(Brand, Price) %>%
+   rename(Source = Brand, Value = Price) %>%
+   mutate(
+     ...
```

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

SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R

The screenshot shows the RStudio interface with the following R code in the console:

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Console Terminal Background Jobs
R 4.5.2 - C:/Users/mvlic/Downloads/
>
> car_clean <- car_df %>%
+   select(-Brand, -Price) %>%
+   rename(source = Brand, value = Price) %>%
+   mutate(
+     dataset = "Car",
+     value = as.numeric(value)
+   ) %>%
+   filter(!is.na(value))
>
> business_clean <- business_df %>%
+   select(-Brand, -Price) %>%
+   rename(source = Brand, value = Price) %>%
+   mutate(
+     dataset = "Business",
+     value = as.numeric(value)
+   ) %>%
+   filter(!is.na(value))
>
> combined_data <- rbind(car_clean, business_clean)
>
> car_rows <- nrow(car_clean)
> business_rows <- nrow(business_clean)
> total_rows <- nrow(combined_data)
>
> print("---- Combined Data Summary ----")
[1] "---- Combined Data Summary ----"
> print(paste("Car rows:", car_rows))
[1] "Car rows: 2500"
> print(paste("Business rows:", business_rows))
[1] "Business rows: 20252"
> print(paste("Total rows (Expected):", car_rows + business_rows))
[1] "Total rows (Expected): 22752"
> print(paste("Total rows (Actual):", total_rows))
[1] "Total rows (Actual): 22752"
>
> print("---- Preview of combined Data (Top - Car Data) ----")
[1] "---- Preview of Combined Data (Top - Car Data) ----"
> print(head(combined_data))
# A tibble: 6 x 3
  Source Value Dataset
  <chr>  <dbl> <chr>
1 Tesla  26614. Car
2 BMW    14680. Car
3 Audi   44403. Car
4 Tesla  86374. Car
5 Ford   72357. Car
6 Audi   88970. Car
>
> print("---- Preview of Combined Data (Bottom - Business Data) ----")
[1] "---- Preview of Combined Data (Bottom - Business Data) ----"
> print(tail(combined_data))
# A tibble: 6 x 3
  Source Value Dataset
  <chr>  <dbl> <chr>
1 Zara   49.0 Business
2 Zara   32.0 Business
3 Zara   50.0 Business
4 Zara   21.0 Business
5 Zara   65.0 Business
6 Zara   65.0 Business

```

The RStudio environment pane shows variables like `busin_` (20252L), `car_r_` (2500L), `diffrr_-10`, `produ_` (200), `quoti_` (2), `sum` (30), and `total_` (22752L). The file browser pane shows various files including PDFs and Word documents related to practical sessions.

This screenshot shows the RStudio interface with the same R code as the previous one, but the output is truncated at the bottom. The visible output is identical to the first screenshot.