

SHETH LUJ AND SIR MV COLLEGE
Subject: Data Analysis with SAS / SPSS / R

Practical No: 6

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

Code:

```
library(dplyr)
```

1. SETUP: Load and Prepare Data

```
country_data <- read.csv("ESGCountry.csv")  
series_data <- read.csv("ESGCountry-Series.csv")
```

```
country_list_1 <- head(country_data, 5)  
country_list_2 <- country_data[6:10, ]
```

```
print("--- Country List 1 (First 5) ---")  
print(head(country_list_1[, c("Country.Code", "Short.Name")]))
```

```
print("--- Country List 2 (Next 5) ---")  
print(head(country_list_2[, c("Country.Code", "Short.Name")]))
```

2. MERGE (Joining Columns)

```
merged_data <- merge(  
  country_data,  
  series_data,  
  by.x = "Country.Code", # Column name in first dataset  
  by.y = "CountryCode" # Column name in second dataset  
)
```

```
print("--- Merged Data (Columns Added) ---")  
print(head(merged_data[, c("Country.Code", "Short.Name", "SeriesCode", "DESCRIPTION")]))
```

3. APPEND (Stacking Rows)

```
final_list <- bind_rows(country_list_1, country_list_2)
```

```
print("--- Appended Data (Rows Added) ---")  
print(final_list[, c("Country.Code", "Short.Name")])
```

Output:

SHETH LUJ AND SIR MV COLLEGE

Subject: Data Analysis with SAS / SPSS / R

The image displays two screenshots of the RStudio environment, showing the process of loading and merging datasets.

Top Screenshot:

- Console:** Shows the R version (4.5.2) and the loading of the 'dplyr' package. The following objects are masked from 'package:stats': filter, lag. The following objects are masked from 'package:base': intersect, setdiff, setequal, union.
- Environment:** Lists the loaded datasets: country_data (239 obs. of 31 variables), country_list_1 (5 obs. of 31 variables), country_list_2 (5 obs. of 31 variables), ESGCountry (240 obs. of 31 variables), ESGCountry.Seri (1349 obs. of 4 variables), final_list (10 obs. of 31 variables), merged_data (1349 obs. of 34 variables), and series_data (1349 obs. of 4 variables).

Bottom Screenshot:

- Console:** Shows the execution of R code to load datasets, create country lists, and merge them. The code includes:

```
> country_data <- read.csv("~/Data Analysis with SAS , SPSS, R/datasets/ESGCountry.csv", header=FALSE)
> View(ESGCountry)
> ESGCountry.Series <- read.csv("~/Data Analysis with SAS , SPSS, R/datasets/ESGCountry-Series.csv")
> View(ESGCountry.Series)
> library(dplyr)

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':
  filter, lag

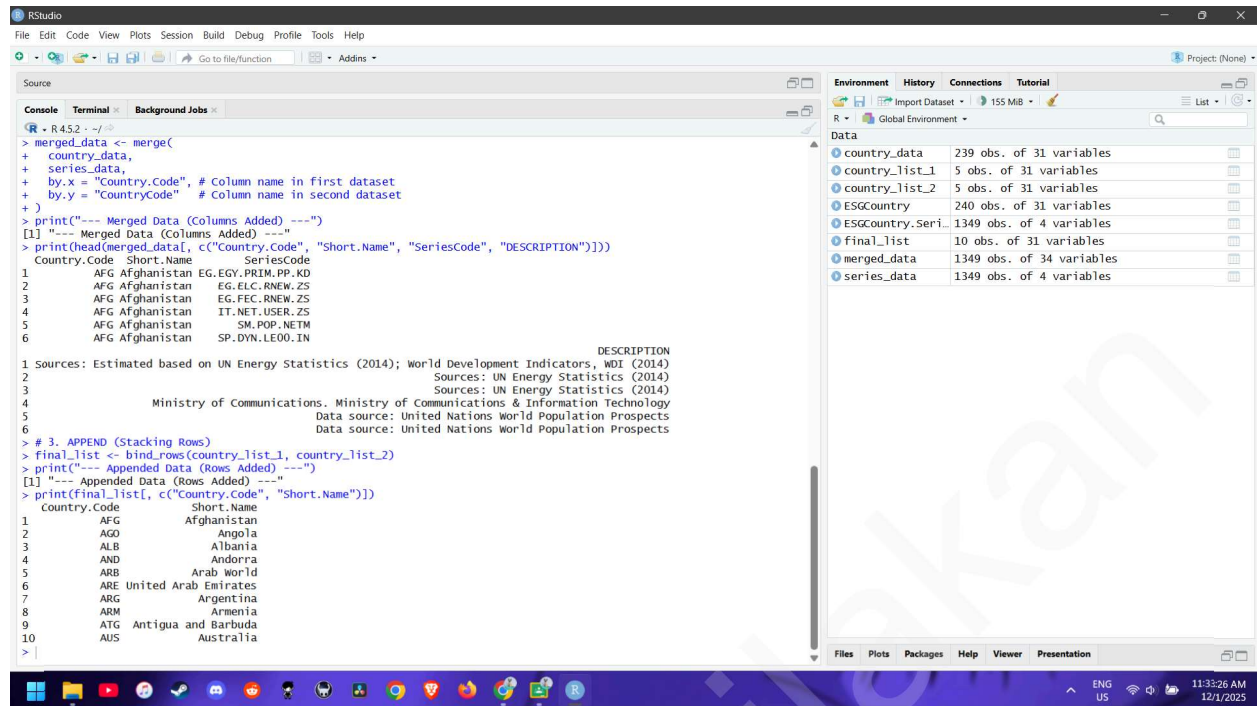
The following objects are masked from 'package:base':
  intersect, setdiff, setequal, union

> country_data <- read.csv("ESGCountry.csv")
> series_data <- read.csv("ESGCountry-Series.csv")
> country_list_1 <- head(country_data, 5)
> country_list_2 <- country_data[6:10, ]
> print("--- Country List 1 (First 5) ---")
[1] "--- Country List 1 (First 5) ---"
> print(head(country_list_1[, c("Country.Code", "Short.Name")]))
  Country.Code Short.Name
1          AFG Afghanistan
2          AGO  Angola
3          ALB  Albania
4          AND  Andorra
5          ARB  Arab World

> print("--- Country List 2 (Next 5) ---")
[1] "--- Country List 2 (Next 5) ---"
> print(head(country_list_2[, c("Country.Code", "Short.Name")]))
  Country.Code Short.Name
6          ARE United Arab Emirates
7          ARG  Argentina
8          ARM  Armenia
9          ATG Antigua and Barbuda
10         AUS  Australia

> merged_data <- merge(
+   country_data,
+   series_data,
+   by.x = "Country.Code", # Column name in first dataset
+   by.y = "CountryCode"  # Column name in second dataset
+ )
> merged_data <- merge(
+   country_data,
+   series_data,
+   by.x = "Country.Code", # Column name in first dataset
+   by.y = "CountryCode"  # Column name in second dataset
+ )
> merged_data <- merge(
+   country_data,
+   series_data,
+   by.x = "Country.Code", # Column name in first dataset
+   by.y = "CountryCode"  # Column name in second dataset
+ )
> print("--- Merged Data (Columns Added) ---")
[1] "--- Merged Data (Columns Added) ---"
> print(head(merged_data[, c("Country.Code", "Short.Name", "SeriesCode", "DESCRIPTION")]))
```
- Environment:** Lists the loaded datasets: country_data (239 obs. of 31 variables), country_list_1 (5 obs. of 31 variables), country_list_2 (5 obs. of 31 variables), ESGCountry (240 obs. of 31 variables), ESGCountry.Seri (1349 obs. of 4 variables), final_list (10 obs. of 31 variables), merged_data (1349 obs. of 34 variables), and series_data (1349 obs. of 4 variables).

SHETH LUJ AND SIR MV COLLEGE
Subject: Data Analysis with SAS / SPSS / R



```
> merged_data <- merge(
+   country_data,
+   series_data,
+   by.x = "Country.Code", # Column name in first dataset
+   by.y = "CountryCode"   # Column name in second dataset
+ )
> print("--- Merged Data (Columns Added) ---")
[1] "--- Merged Data (Columns Added) ---"
> print(head(merged_data[, c("Country.Code", "Short.Name", "SeriesCode", "DESCRIPTION")]))
  Country.Code Short.Name SeriesCode
1      AFG Afghanistan EG.EGY.PRIM.PP.KD
2      AFG Afghanistan EG.ELC.RNEW.ZS
3      AFG Afghanistan EG.FEC.RNEW.ZS
4      AFG Afghanistan IT.NET.USER.ZS
5      AFG Afghanistan SM.POP.NETM
6      AFG Afghanistan SP.DYN.LE00.IN

1 Sources: Estimated based on UN Energy Statistics (2014); World Development Indicators, WDI (2014)
2 Sources: UN Energy Statistics (2014)
3 Sources: UN Energy Statistics (2014)
4 Ministry of Communications. Ministry of Communications & Information Technology
5 Data source: United Nations World Population Prospects
6 Data source: United Nations World Population Prospects

> # 3. APPEND (Stacking Rows)
> final_list <- bind_rows(country_list_1, country_list_2)
> print("--- Appended Data (Rows Added) ---")
[1] "--- Appended Data (Rows Added) ---"
> print(final_list[, c("Country.Code", "Short.Name")])
  Country.Code Short.Name
1      AFG Afghanistan
2      AGO Angola
3      ALB Albania
4      AND Andorra
5      ARB Arab World
6      ARE United Arab Emirates
7      ARG Argentina
8      ARM Armenia
9      ATG Antigua and Barbuda
10     AUS Australia
> |
```

The Environment pane on the right shows the following objects:

Object	Size
country_data	239 obs. of 31 variables
country_list_1	5 obs. of 31 variables
country_list_2	5 obs. of 31 variables
ESGCountry	240 obs. of 31 variables
ESGCountry.Seri	1349 obs. of 4 variables
final_list	10 obs. of 31 variables
merged_data	1349 obs. of 34 variables
series_data	1349 obs. of 4 variables