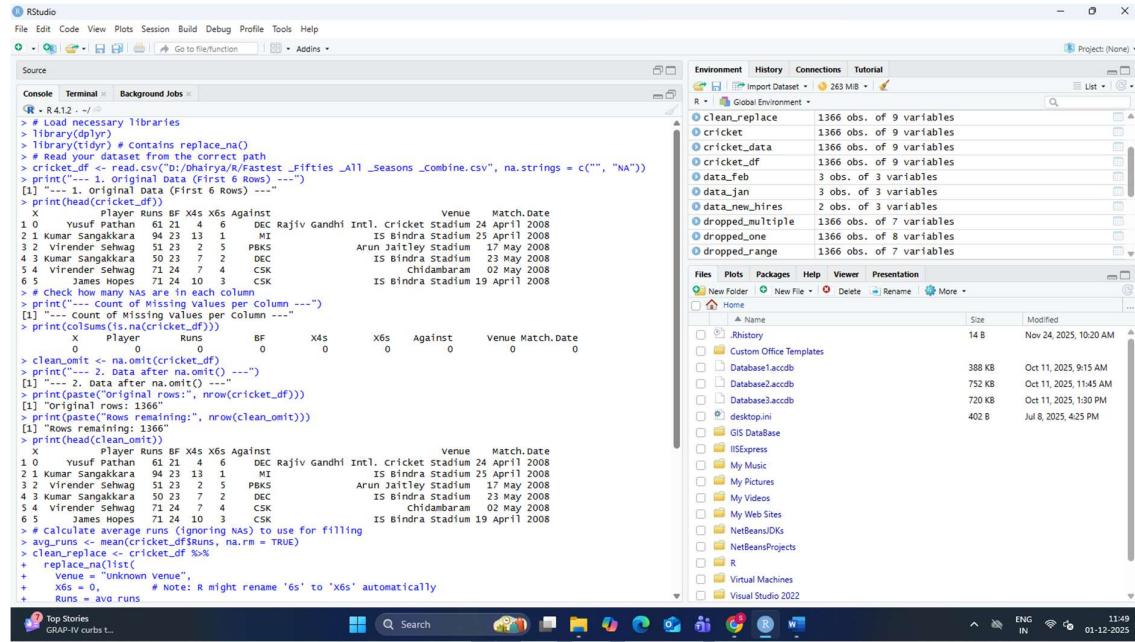


**SHETH L.U.J AND SIR M.V. COLLEGE**  
**SUBJECT NAME: Data Analysis with SAS / SPSS /R**  
**Practical No. 8**

**Aim**-Applying basic data cleaning functions: handling missing values using

Na. omit ()/replace\_na () in R. import dataset.

**Output-**

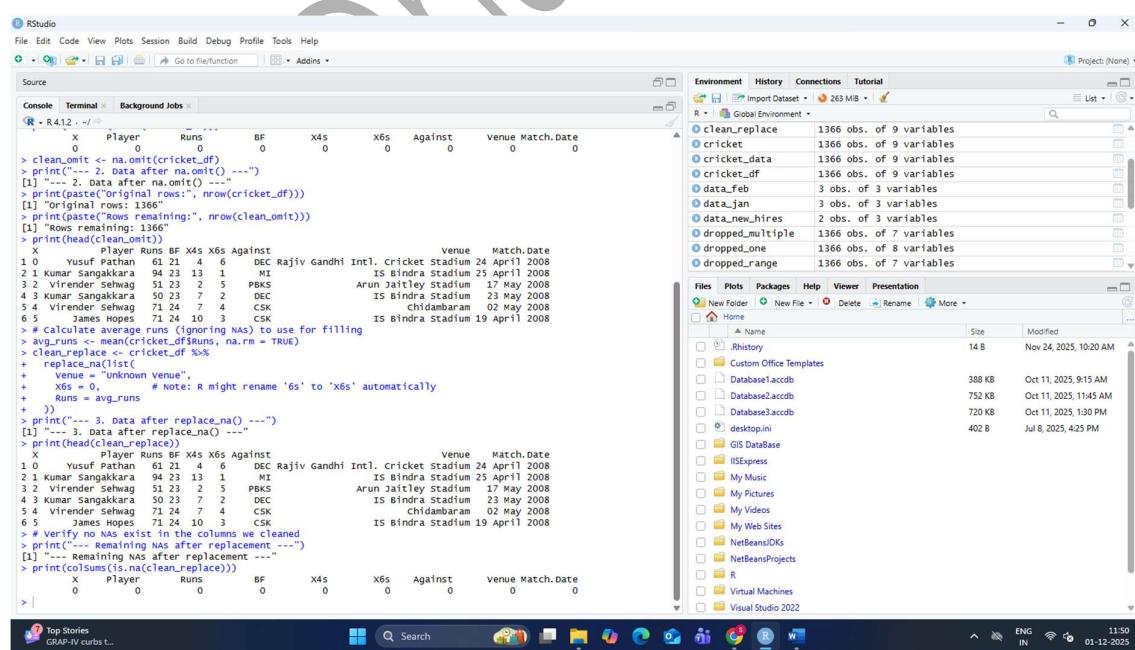


The screenshot shows the RStudio interface with two panes. The left pane (Console) displays R code and its output. The right pane (Environment) shows the global environment with various objects and their details. The code demonstrates the use of `na.omit` and `replace\_na` functions to handle missing values in a cricket dataset.

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Console Terminal Background Jobs
Source
R - R4.12 - /-
> # Load necessary libraries
> library(dplyr)
> library(tidyverse) # contains replace_na()
> # Read your dataset from the correct path
> cricket_df <- read.csv("D:/Dhairya/R/Fastest_Fifties_All_seasons_combine.csv", na.strings = c("", "NA"))
> print("--- 1. Original data (First 6 Rows) ---")
[1] "--- 1. Original data (First 6 Rows) ---"
> print(head(cricket_df))
#> #> #> #> #> #>
X Player Runs BF X45 X6s Against Venue Match.Date
1 0 Yusuf Pathan 61 21 4 6 DEC Rajiv Gandhi Intl. Cricket Stadium 24 April 2008
2 1 Kumar Sangakkara 93 23 13 1 MI IS Bindra Stadium 25 April 2008
3 2 Virender Sehwag 51 23 2 5 PBKS Arun Jaitley Stadium 17 May 2008
4 3 Kumar Sangakkara 50 23 7 2 DEC IS Bindra Stadium 23 May 2008
5 4 Virender Sehwag 51 24 7 4 CSK Chidambaram 02 May 2008
6 5 James Hopes 71 24 10 3 CSK IS Bindra Stadium 19 April 2008
#> # Check how many NAs are present in each column
> print("--- Count of Missing values per column ---")
[1] "--- Count of Missing values per column ---"
> print(crosssums(is.na(cricket_df)))
#> #> #> #> #> #>
X Player Runs BF X45 X6s Against Venue Match.Date
0 0 0 0 0 0 0 0 0 0
> clean omit <- na.omit(cricket_df)
> print("--- 2. Data after na.omit() ---")
[1] "--- 2. Data after na.omit() ---"
> print(paste("Original rows:", nrow(cricket_df)))
[1] "Original rows: 1366"
> print(paste("Rows remaining:", nrow(clean omit)))
[1] "Rows remaining: 1366"
> print(head(clean omit))
#> #> #> #> #> #>
X Player Runs BF X45 X6s Against Venue Match.Date
1 0 Yusuf Pathan 61 21 4 6 DEC Rajiv Gandhi Intl. Cricket Stadium 24 April 2008
2 1 Kumar Sangakkara 93 23 13 1 MI IS Bindra Stadium 25 April 2008
3 2 Virender Sehwag 51 23 2 5 PBKS Arun Jaitley Stadium 17 May 2008
4 3 Kumar Sangakkara 50 23 7 2 DEC IS Bindra Stadium 23 May 2008
5 4 Virender Sehwag 51 24 7 4 CSK Chidambaram 02 May 2008
6 5 James Hopes 71 24 10 3 CSK IS Bindra Stadium 19 April 2008
#> # Calculate average runs (ignoring NAs) to use for filling
> avg_runs <- mean(cricket_df$runs, na.rm = TRUE)
> clean replace <- cricket_df %>%
+ replace_na(list(
+   venue = "Unknown venue",
+   X6s = 0, # Note: R might rename '6s' to 'x6s' automatically
+   runs = avg_runs
+ ))
> print("--- 3. Data after replace_na() ---")
[1] "--- 3. Data after replace_na() ---"
> print(head(clean replace))
#> #> #> #> #> #>
X Player Runs BF X45 X6s Against Venue Match.Date
1 0 Yusuf Pathan 61 21 4 6 DEC Rajiv Gandhi Intl. Cricket Stadium 24 April 2008
2 1 Kumar Sangakkara 93 23 13 1 MI IS Bindra Stadium 25 April 2008
3 2 Virender Sehwag 51 23 2 5 PBKS Arun Jaitley Stadium 17 May 2008
4 3 Kumar Sangakkara 50 23 7 2 DEC IS Bindra Stadium 23 May 2008
5 4 Virender Sehwag 51 24 7 4 CSK Chidambaram 02 May 2008
6 5 James Hopes 71 24 10 3 CSK IS Bindra Stadium 19 April 2008
#> # Verify no NAs exist in the columns we cleaned
> print("--- Remaining NAs after replacement ---")
[1] "--- Remaining NAs after replacement ---"
> print(crosssums(is.na(clean replace)))
#> #> #> #> #> #>
X Player Runs BF X45 X6s Against Venue Match.Date
0 0 0 0 0 0 0 0 0

```



This screenshot is identical to the one above, showing the RStudio interface with the same code and environment. It demonstrates the use of `na.omit` and `replace\_na` functions to handle missing values in a cricket dataset.