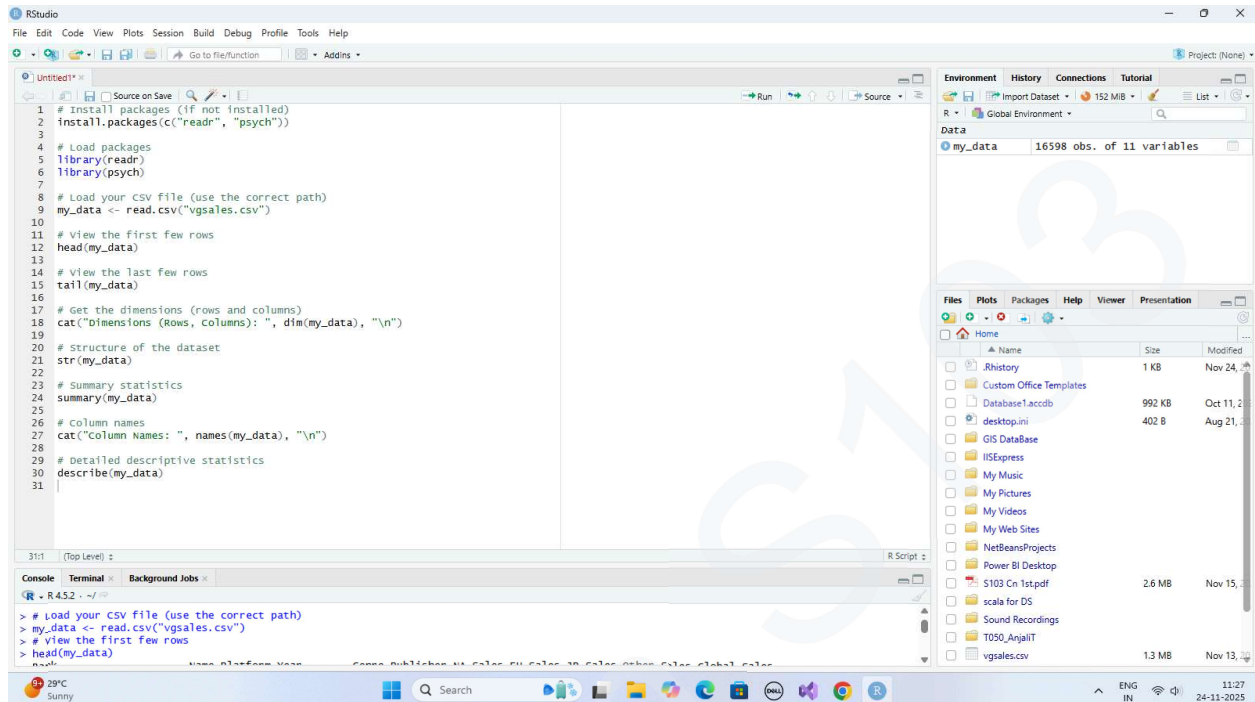


# SHETH L.U.J. AND SIR M.V. COLLEGE

## SUBJECT: Data Analysis with R

Aim: Exploring data: View() or print() (R).

### Input

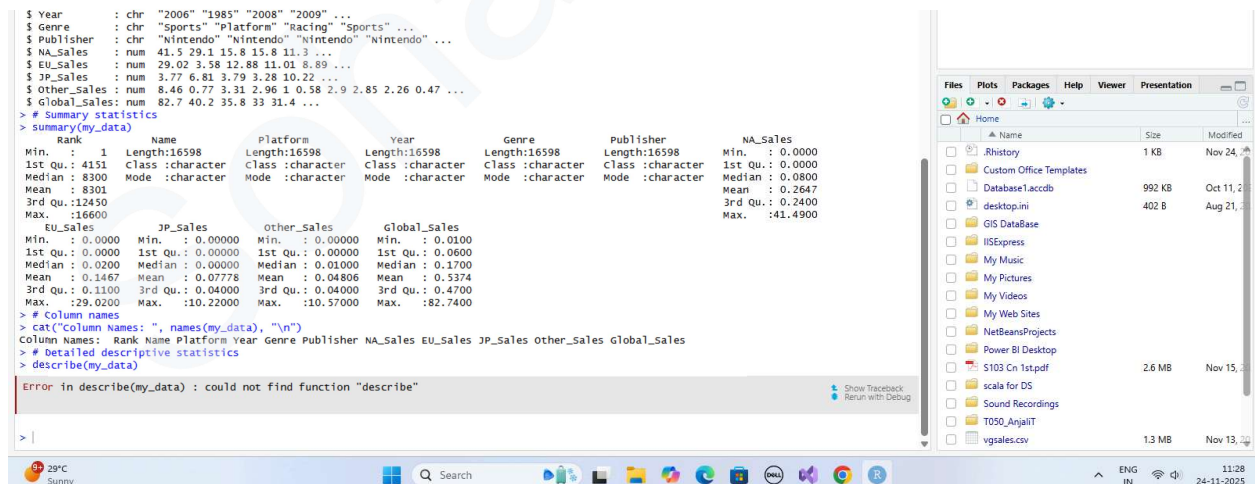


The screenshot shows the RStudio interface with the following code in the script editor:

```
1 # Install packages (if not installed)
2 install.packages(c("readr", "psych"))
3
4 # Load packages
5 library(readr)
6 library(psych)
7
8 # Load your csv file (use the correct path)
9 my_data <- read.csv("vgsales.csv")
10
11 # View the first few rows
12 head(my_data)
13
14 # View the last few rows
15 tail(my_data)
16
17 # Get the dimensions (rows and columns)
18 cat("Dimensions (Rows, columns): ", dim(my_data), "\n")
19
20 # Structure of the dataset
21 str(my_data)
22
23 # Summary statistics
24 summary(my_data)
25
26 # Column names
27 cat("Column Names: ", names(my_data), "\n")
28
29 # Detailed descriptive statistics
30 describe(my_data)
31
```

The Environment pane on the right shows the variable `my_data` with 16598 observations and 11 variables. The Files pane shows the location of `vgsales.csv` in the `Downloads` folder.

### Output



The screenshot shows the RStudio interface with the following output in the console:

```
$ Year      : chr "2006" "1985" "2008" "2009" ...
$ Genre     : chr "Sports" "Platform" "Racing" "Sports" ...
$ Publisher  : chr "Nintendo" "Nintendo" "Nintendo" "Nintendo" ...
$ NA_Sales  : num 41.5 29.1 15.8 15.8 11.3 ...
$ EU_Sales  : num 29.02 3.58 12.88 11.01 8.89 ...
$ JP_Sales  : num 3.77 6.81 3.79 3.28 10.22 ...
$ Other_Sales : num 8.46 0.77 3.31 2.96 1 0.58 2.9 2.85 2.26 0.47 ...
$ Global_Sales : num 82.7 40.2 35.8 33 31.4 ...

> # Summary statistics
> summary(my_data)
      Rank      Name      Platform      Year      Genre      Publisher      NA_Sales
Min.   : 1      Length:16598      Length:16598      Length:16598      Length:16598      Length:16598      Min.   : 0.0000
1st Qu.: 4151      Class :character      Class :character      Class :character      Class :character      Class :character      1st Qu.: 0.0000
Median : 8300      Mode  :character      Mode  :character      Mode  :character      Mode  :character      Mode  :character      Median : 0.0800
Mean   : 8301
3rd Qu.:12450
Max.   :16600

      EU_Sales      JP_Sales      other_Sales      Global_Sales
Min.   : 0.0000      Min.   : 0.00000      Min.   : 0.00000      Min.   : 0.0100
1st Qu.: 0.0000      1st Qu.: 0.00000      1st Qu.: 0.00000      1st Qu.: 0.0600
Median : 0.0200      Median : 0.00000      Median : 0.01000      Median : 0.1700
Mean   : 0.1467      Mean   : 0.07778      Mean   : 0.04806      Mean   : 0.5374
3rd Qu.: 0.1100      3rd Qu.: 0.04000      3rd Qu.: 0.04000      3rd Qu.: 0.4700
Max.   :29.0200      Max.   :10.22000      Max.   :10.57000      Max.   :82.7400

> # Column names
> cat("Column Names: ", names(my_data), "\n")
Column Names: Rank Name Platform Year Genre Publisher NA_Sales EU_Sales JP_Sales Other_Sales Global_Sales

> # Detailed descriptive statistics
> describe(my_data)
Error in describe(my_data) : could not find function "describe"
```

The Files pane on the right shows the location of `vgsales.csv` in the `Downloads` folder.

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

## SUBJECT: Data Analysis with R

```
Column Names: Rank Name Platform Year Genre Publisher NA_Sales EU_Sales JP_Sales Other_Sales Global_Sales
> describe(my_data)
```

|              | vars | n     | mean    | sd      | median  | trimmed | mad     | min  | max      | range    | skew  | kurtosis | se    |
|--------------|------|-------|---------|---------|---------|---------|---------|------|----------|----------|-------|----------|-------|
| Rank         | 1    | 16598 | 8300.61 | 4791.85 | 8300.50 | 8300.56 | 6152.05 | 1.00 | 16600.00 | 16599.00 | 0.00  | -1.20    | 37.19 |
| Name*        | 2    | 16598 | 5795.87 | 3324.01 | 5864.50 | 5810.23 | 4270.63 | 1.00 | 11493.00 | 11492.00 | -0.03 | -1.21    | 25.80 |
| Platform*    | 3    | 16598 | 16.71   | 8.29    | 17.00   | 16.67   | 10.38   | 1.00 | 31.00    | 30.00    | -0.05 | -1.00    | 0.06  |
| Year*        | 4    | 16598 | 27.61   | 6.00    | 28.00   | 28.00   | 5.93    | 1.00 | 40.00    | 39.00    | -0.86 | 1.68     | 0.05  |
| Genre*       | 5    | 16598 | 5.93    | 3.76    | 6.00    | 5.86    | 5.93    | 1.00 | 12.00    | 11.00    | 0.07  | -1.43    | 0.03  |
| Publisher*   | 6    | 16598 | 299.40  | 181.98  | 329.00  | 303.97  | 272.80  | 1.00 | 579.00   | 578.00   | -0.15 | -1.40    | 1.41  |
| NA_Sales     | 7    | 16598 | 0.26    | 0.82    | 0.08    | 0.13    | 0.12    | 0.00 | 41.49    | 41.49    | 18.80 | 648.86   | 0.01  |
| EU_Sales     | 8    | 16598 | 0.15    | 0.51    | 0.02    | 0.06    | 0.03    | 0.00 | 29.02    | 29.02    | 18.87 | 755.71   | 0.00  |
| JP_Sales     | 9    | 16598 | 0.08    | 0.31    | 0.00    | 0.02    | 0.00    | 0.00 | 10.22    | 10.22    | 11.20 | 194.15   | 0.00  |
| other_sales  | 10   | 16598 | 0.05    | 0.19    | 0.01    | 0.02    | 0.01    | 0.00 | 10.57    | 10.57    | 24.23 | 1024.92  | 0.00  |
| Global_Sales | 11   | 16598 | 0.54    | 1.56    | 0.17    | 0.27    | 0.21    | 0.01 | 82.74    | 82.73    | 17.40 | 603.68   | 0.01  |