

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

## SUBJECT: Data Analysis with R

Aim: Generating basic summaries using str() or summary() (R). ‘

The screenshot displays the R Studio interface with the following components:

- Console:** Shows the execution of R code to load the 'tidyverse' and 'readr' libraries, read the 'test.csv' file into 'titanic\_df', and generate summaries using 'str()' and 'summary()'.
- Environment:** Lists the loaded data frame 'titanic\_df' with 418 observations and 11 variables.
- Files:** A file explorer showing the project directory with various files and folders.

**Code Executed in Console:**

```
> # Load necessary library to read csv
> library(readr)
> # 1. LOAD DATA
> titanic_df <- read_csv("test.csv") # replace with your csv file path
Rows: 418 Columns: 11
Column specification
Delimiter: ","
chr (5): Name, Sex, Ticket, Cabin, Embarked
dbl (6): PassengerId, Pclass, Age, Sibsp, Parch, Fare

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 Loaded ----")
[1] "---- Data Loaded ----"
> # 2. USING str() (Structure)
> print("---- OUTPUT OF str() ----")
[1] "---- OUTPUT OF str() ----"
> str(titanic_df)
spc_tbl_ [418 x 11] (5): spec_tbl_df/tbl_df/tbl/data.frame
$ PassengerId: num [1:418] 892 893 894 895 896 897 898 899 900 901 ...
$ Pclass: num [1:418] 3 3 2 3 3 3 3 2 3 3 ...
$ Name: chr [1:418] "Kelly, Mr. James" "Wilkes, Mrs. James (Ellen Needs)" "Myles, Mr. Thomas Francis" "Wirtz, Mr. Albert" ...
$ Sex: chr [1:418] "male" "female" "male" "male" ...
$ Age: num [1:418] 34.5 47 62 27 22 14 30 26 18 21 ...
$ Sibsp: num [1:418] 0 1 0 0 1 0 0 1 0 2 ...
$ Parch: num [1:418] 0 0 0 0 1 0 0 1 0 0 ...
$ Ticket: chr [1:418] "330911" "363272" "240276" "315154" ...
$ Fare: num [1:418] 7.83 7.99 8.66 12.29 ...
$ Cabin: chr [1:418] NA NA NA NA ...
$ Embarked: chr [1:418] "Q" "S" "Q" "S" ...
- attr(*, "spec")=
.. cols(
.. .. PassengerId = col_double(),
.. .. Pclass = col_double(),
.. .. Name = col_character(),
.. .. Sex = col_character(),
.. .. Age = col_double(),
.. .. Sibsp = col_double(),
.. .. Parch = col_double(),
.. .. Ticket = col_character(),
.. .. Fare = col_double(),
.. .. Cabin = col_character(),
.. )
> # 3. USING summary() (Statistical Summary)
> print("---- OUTPUT OF summary() [Before Factor conversion] ----")
[1] "---- OUTPUT OF summary() [Before Factor conversion] ----"
> summary(titanic_df)
 PassengerId      Pclass      Name      Sex      Age      Sibsp      Parch      Ticket
Min.   : 892.0   Min.   :1.000   Length:418   Length:418   Min.   : 0.17   Min.   :0.0000   Min.   :0.0000   Length:418
1st Qu.: 996.2   1st Qu.:1.000   Class :character   Class :character   1st Qu.:21.00   1st Qu.:0.0000   1st Qu.:0.0000   Class :chara
Median :1100.5   Median :3.000   Mode  :character   Mode  :character   Median :27.00   Median :0.0000   Median :0.0000   Mode  :chara
Mean   :1100.5   Mean   :2.266                                     Mean   :30.27   Mean   :0.4474   Mean   :0.3923
3rd Qu.:1204.8   3rd Qu.:3.000                                     3rd Qu.:39.00   3rd Qu.:1.0000   3rd Qu.:0.0000
Max.   :1309.0   Max.   :3.000                                     Max.   :76.00   Max.   :8.0000   Max.   :9.0000
NA's   :86
Fare      Cabin      Embarked
Min.   : 0.000   Length:418   Length:418
1st Qu.: 7.896   Class :character   Class :character
Median :14.454   Mode  :character   Mode  :character
Mean   :35.627
3rd Qu.:31.500
Max.   :512.329
NA's   :1
> # 4. IMPROVING summary() WITH FACTORS
> # convert some character columns to factor for better summary
> titanic_df$Pclass <- as.factor(titanic_df$Pclass)
> titanic_df$Sex <- as.factor(titanic_df$Sex)
> titanic_df$Embarked <- as.factor(titanic_df$Embarked)
> print("---- OUTPUT OF summary() [After Factor conversion] ----")
[1] "---- OUTPUT OF summary() [After Factor conversion] ----"
> summary(titanic_df)
 PassengerId      Pclass      Name      Sex      Age      Sibsp      Parch      Ticket      Fa
Min.   : 892.0   1:107   Length:418   female:152   Min.   : 0.17   Min.   :0.0000   Min.   :0.0000   Length:418   Min.
: 0.000
1st Qu.: 996.2   2: 93   Class :character   male :266   1st Qu.:21.00   1st Qu.:0.0000   1st Qu.:0.0000   Class :character   1st Q
u.: 7.896
Median :1100.5   3:218   Mode  :character   Median :27.00   Median :0.0000   Median :0.0000   Mode  :character   Median
:14.454
Mean   :1100.5                                     Mean   :30.27   Mean   :0.4474   Mean   :0.3923
:35.627
3rd Qu.:1204.8                                     3rd Qu.:39.00   3rd Qu.:1.0000   3rd Qu.:0.0000
Max.   :1309.0                                     Max.   :76.00   Max.   :8.0000   Max.   :9.0000
NA's   :1
```

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

## SUBJECT: Data Analysis with R

```
> titanic_df$Embarked <- as.factor(titanic_df$Embarked)
> print("---- OUTPUT OF summary() [After Factor conversion] ----")
[1] "---- OUTPUT OF summary() [After Factor conversion] ----"
> summary(titanic_df)
  PassengerId   Pclass      Name      Sex      Age      SibSp      Parch      Ticket      Fa
Min.   : 892.0   1:107   Length:418   female:152   Min.   : 0.17   Min.   :0.0000   Min.   :0.0000   Length:418   Min.
: 0.000
1st Qu.: 996.2   2: 93   Class :character   male :266   1st Qu.:21.00   1st Qu.:0.0000   1st Qu.:0.0000   Class :character   1st Q
u.: 7.896
Median :1100.5   3:218   Mode  :character               Median :27.00   Median :0.0000   Median :0.0000   Mode :character   Median
: 14.454
Mean    :1100.5               Mean    :30.27   Mean    :0.4474   Mean    :0.3923               Mean
: 35.627
3rd Qu.:1204.8               3rd Qu.:39.00   3rd Qu.:1.0000   3rd Qu.:0.0000               3rd Q
u.: 31.500
Max.    :1309.0               Max.    :76.00   Max.    :8.0000   Max.    :9.0000               Max.
:512.329
NA's    :86
NA's
:1
  Cabin      Embarked
Length:418   C:102
Class :character   Q: 46
Mode :character   S:270

> # 5. Accessing Specific Summaries
> avg_age <- mean(titanic_df$Age, na.rm = TRUE)
> max_fare <- max(titanic_df$Fare, na.rm = TRUE)
> print(paste("Average Age:", round(avg_age, 2)))
[1] "Average Age: 30.27"
> print(paste("Highest Fare:", max_fare))
[1] "Highest Fare: 512.3292"
> view(titanic_df)
>
```

Environment

Object	Class	Attributes
titanic_df	data.frame	418 obs. of 11 variables

Files

Name	Size	Modified
NetBeansProjects		
Power BI Desktop		
scala for DS		
Sound Recordings		
T050_AnjaliT		
vgsales.csv	1.3 MB	Nov 13, 2025, 8:49 AM
Virtual Machines		
Visual Studio 18		
Visual Studio 2022		
WindowsPowerShell		
Student.csv	163 KB	Dec 8, 2025, 10:48 AM
car_price_prediction_csv	156.3 KB	Nov 15, 2025, 1:10 PM
disease_diagnosis.csv	189.1 KB	Dec 1, 2025, 11:33 AM
flavors_of_cacao.csv	126.5 KB	Nov 13, 2025, 8:06 AM
london_merged.csv	1010.6 KB	Dec 8, 2025, 11:55 AM
test.csv	28 KB	Dec 8, 2025, 12:11 PM
train.csv	59.8 KB	Dec 8, 2025, 12:11 PM