

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

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

RStudio interface showing R code for generating basic summaries:

```

R - R 4.5.2 - ~/r
> wh_df <- read.csv("C:/Users/mvluc/Downloads/weight-height.csv")
>
> print("---- Data Loaded From CSV ---")
[1] "---- Data Loaded From CSV ---"
> print(head(wh_df))
  Gender Height weight
1   Male    73.8 184.7
2   Male    74.2 182.9
3   Male    74.1 110.1
4   Male    71.3  90.8
5   Male    69.8 181.0
6   Male    67.2 253.0
>
>
> print("---- OUTPUT OF str() ---")
[1] "---- OUTPUT OF str() ---"
> str(wh_df)
'data.frame': 100000 obs. of 3 variables:
$ Gender: chr "Male" "Male" "Male" "Male" ...
$ Height: num 73.8 68.8 74.1 71.7 69.9 ...
$ weight: num 184.7 182.9 110.1 90.8 181.0 ...
>
>
> print("---- OUTPUT OF summary() [Before Factor Conversion] ---")
[1] "---- OUTPUT OF summary() [Before Factor Conversion] ---"
> summary(wh_df)
  Gender      Height      weight
Length:100000  Min. :54.26  Min. :64.7
Class:character 1st Qu.:63.51  1st Qu.:135.8
Mode :character  Median :66.32  Median :161.2
                  Mean  :66.37  Mean  :161.4
                  3rd Qu.:69.17  3rd Qu.:187.2
                  Max.  :79.00  Max.  :270.0
>
>
> # Convert Gender column to Factor (if it exists)
> wh_df$Gender <- as.factor(wh_df$Gender)
>
> print("---- OUTPUT OF summary() [After Factor conversion] ---")

```

RStudio interface showing the continuation of the code for generating basic summaries:

```

R - R 4.5.2 - ~/r
>
>
> print("---- OUTPUT OF summary() [Before Factor Conversion] ---")
[1] "---- OUTPUT OF summary() [Before Factor Conversion] ---"
> summary(wh_df)
  Gender      Height      weight
Length:100000  Min. :54.26  Min. :64.7
Class:character 1st Qu.:63.51  1st Qu.:135.8
Mode :character  Median :66.32  Median :161.2
                  Mean  :66.37  Mean  :161.4
                  3rd Qu.:69.17  3rd Qu.:187.2
                  Max.  :79.00  Max.  :270.0
>
>
> # Convert Gender column to Factor (if it exists)
> wh_df$Gender <- as.factor(wh_df$Gender)
>
> print("---- OUTPUT OF summary() [After Factor Conversion] ---")
[1] "---- OUTPUT OF summary() [After Factor Conversion] ---"
> summary(wh_df)
  Gender      Height      weight
Female:5000  Min. :54.26  Min. :64.7
Male :5000   1st Qu.:63.51  1st Qu.:135.8
             Median :66.32  Median :161.2
             Mean  :66.37  Mean  :161.4
             3rd Qu.:69.17  3rd Qu.:187.2
             Max.  :79.00  Max.  :270.0
>
>
> # Accessing only selected statistics
> avg_height <- mean(wh_df$height, na.rm = TRUE)
> max_weight <- max(wh_df$weight, na.rm = TRUE)
>
> print(paste("Average Height:", avg_height))
[1] "Average Height: 66.3675597548212"
> print(paste("Highest weight:", max_weight))
[1] "Highest weight: 269.989698505106"
>
> print("Yukta_Sonawane_S120")
[1] "Yukta_Sonawane_S120"
>

```