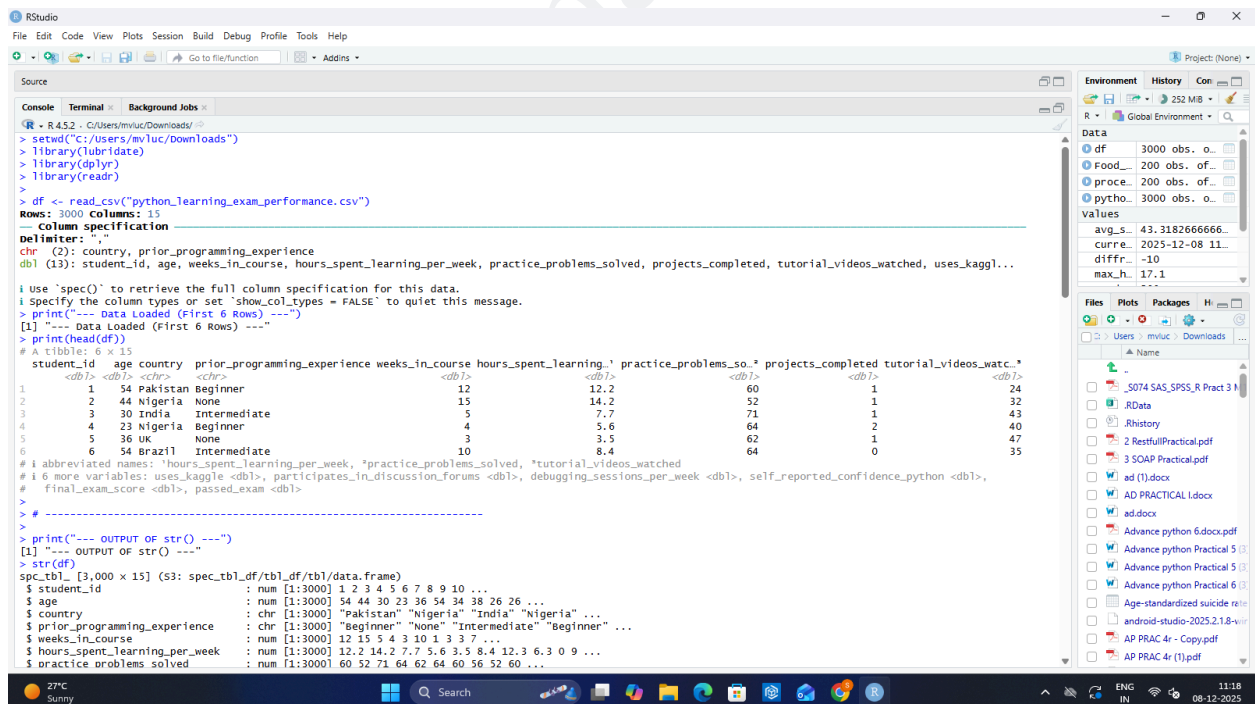


SHETH L.U.J AND SIR M.V. COLLEGE
SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R

Module 1 Practical 14

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

OUTPUT:



```
R • R 4.5.2 - C:/Users/mvlu/Downloads/
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins

Source
Console Terminal Background Jobs
> setwd("C:/Users/mvlu/Downloads")
> library(lubridate)
> library(dplyr)
> library(readr)
>
> df <- read_csv("python_learning_exam_performance.csv")
Rows: 3000 Columns: 15
Column specification
Delimiter: ","
chr (2): country, prior_programming_experience
dbl (13): student_id, age, weeks_in_course, hours_spent_learning_per_week, practice_problems_solved, projects_completed, tutorial_videos_watched, uses_kaggl...

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 (First 6 Rows) ----")
[1] "---- Data Loaded (First 6 Rows) ----"
> print(head(df))
# A tibble: 6 x 15
  student_id age country prior_programming_experience weeks_in_course hours_spent_learning_per_week practice_problems_solved projects_completed tutorial_videos_watched
  <dbl> <dbl> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>
1 1 54 Pakistan Beginner 12 12.2 60 1 24
2 2 44 Nigeria None 15 14.2 52 1 32
3 3 30 India Intermediate 5 7.7 71 1 43
4 4 23 Nigeria Beginner 4 5.6 64 2 40
5 5 36 UK None 3 3.5 62 1 47
6 6 54 Brazil Intermediate 10 8.4 64 0 35

# i abbreviated names: 'hours_spent_learning_per_week', 'practice_problems_solved', 'tutorial_videos_watched'
# i 6 more variables: uses_kaggle <dbl>, participates_in_discussion_forums <dbl>, debugging_sessions_per_week <dbl>, self_reported_confidence_python <dbl>,
# final_exam_score <dbl>, passed_exam <dbl>
>
> # -----
>
> print("---- OUTPUT OF str() ----")
[1] "---- OUTPUT OF str() ----"
> str(df)
'spc_tbl' [3,000 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ student_id      : num [1:3000] 1 2 3 4 5 6 7 8 9 10 ...
 $ age             : num [1:3000] 54 44 30 23 36 54 34 38 26 26 ...
 $ country         : chr [1:3000] "Pakistan" "Nigeria" "India" "Nigeria" ...
 $ prior_programming_experience : chr [1:3000] "Beginner" "None" "Intermediate" "Beginner" ...
 $ weeks_in_course : num [1:3000] 12 15 5 4 3 10 1 3 3 7 ...
 $ hours_spent_learning_per_week : num [1:3000] 12.2 14.2 7.7 5.6 3.5 8.4 12.3 6.3 0 9 ...
 $ practice_problems_solved : num [1:3000] 60 52 71 64 62 64 60 56 52 60 ...
```

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

SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
R - R 4.5.2 - C:/Users/mvuc/Downloads/
$ tutorial_videos_watched : num [1:3000] 24 32 43 40 47 35 35 41 26 42 ...
$ uses_kaggle : num [1:3000] 0 0 1 1 0 0 1 0 1 ...
$ participates_in_discussion_forums : num [1:3000] 1 0 0 0 1 0 1 0 1 ...
$ debugging_sessions_per_week : num [1:3000] 4 5 6 3 3 3 5 7 6 ...
$ self_reported_confidence_python : num [1:3000] 10 7 3 10 7 3 2 1 6 3 ...
$ final_exam_score : num [1:3000] 53.8 31.9 59.4 58.8 24.8 43.8 40.8 24.7 20.7 45.3 ...
$ passed_exam : num [1:3000] 0 0 0 0 0 0 0 0 0 ...
- attr(*, "spec")=
.. cols(
.. student_id = col_double(),
.. age = col_double(),
.. country = col_character(),
.. prior_programming_experience = col_character(),
.. weeks_in_course = col_double(),
.. hours_spent_learning_per_week = col_double(),
.. practice_problems_solved = col_double(),
.. projects_completed = col_double(),
.. tutorial_videos_watched = col_double(),
.. uses_kaggle = col_double(),
.. participates_in_discussion_forums = col_double(),
.. debugging_sessions_per_week = col_double(),
.. self_reported_confidence_python = col_double(),
.. final_exam_score = col_double(),
.. passed_exam = col_double()
.. )
- attr(*, "problems")=externalptr>
> # -----
> print("---- OUTPUT of summary() [before Factor conversion] ----")
[1] "---- OUTPUT of summary() [before Factor conversion] ----"
> summary(df)
  student_id      age      country      prior_programming_experience weeks_in_course hours_spent_learning_per_week practice_problems_solved
Min. : 1.0      Min. :16.00      Length:3000      Length:3000      Min. : 1.000      Min. : 0.000      Min. :35.00
1st Qu.: 750.8    1st Qu.:26.00      Class :character      Class :character      1st Qu.: 5.000      1st Qu.: 5.000      1st Qu.:55.00
Median :1500.5    Median :36.00      Mode :character      Mode :character      Median : 8.000      Median : 7.000      Median :60.00
Mean :1500.5      Mean :35.33                                     Mean : 8.121      Mean : 7.035      Mean :59.98
3rd Qu.:2250.2    3rd Qu.:45.00                                     3rd Qu.:12.000      3rd Qu.: 9.000      3rd Qu.:65.00
Max. :3000.0      Max. :54.00                                     Max. :15.000      Max. :17.100      Max. :87.00
projects_completed tutorial_videos_watched uses_kaggle participates_in_discussion_forums debugging_sessions_per_week self_reported_confidence_python
Min. :0.000      Min. :19.0      Min. :0.0000      Min. :0.0000      Min. : 0.000      Min. : 1.000
1st Qu.:1.000    1st Qu.:35.0    1st Qu.:0.0000    1st Qu.:0.0000    1st Qu.: 3.000    1st Qu.: 3.000
Mean :2.000      Mean :40.0      Mean :0.0007      Mean :0.4983      Mean : 4.973      Mean : 5.566
3rd Qu.:3.000    3rd Qu.:44.0    3rd Qu.:1.0000    3rd Qu.:1.0000    3rd Qu.: 6.000    3rd Qu.: 8.000
Max. :9.000      Max. :63.0      Max. :1.0000      Max. :1.0000      Max. :17.000      Max. :10.000
final_exam_score passed_exam
Min. : 0.00      Min. :0.0000
1st Qu.:30.90    1st Qu.:0.0000
Median :43.10    Median :0.0000
Mean :43.32      Mean :0.1773
3rd Qu.:55.60    3rd Qu.:0.0000
Max. :100.00     Max. :1.0000
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
R - R 4.5.2 - C:/Users/mvuc/Downloads/
> # -----
> df$prior_programming_experience <- as.factor(df$prior_programming_experience)
> df$country <- as.factor(df$country)
> print("---- OUTPUT of summary() [After Factor conversion] ----")
[1] "---- OUTPUT of summary() [After Factor conversion] ----"
> summary(df)
  student_id      age      country      prior_programming_experience weeks_in_course hours_spent_learning_per_week practice_problems_solved
Min. : 1.0      Min. :16.00      Brazil : 339      Beginner : 270      Min. : 1.000      Min. : 0.000      Min. :35.00
1st Qu.: 750.8    1st Qu.:26.00      India : 304      Intermediate: 634      1st Qu.: 5.000      1st Qu.: 5.000      1st Qu.:55.00
Median :1500.5    Median :36.00      USA : 298      None :1062      Median : 8.000      Median : 7.000      Median :60.00
Mean :1500.5      Mean :35.33      Max. :1.0000      Max. :1.0000      Mean : 8.121      Mean : 7.035      Mean :59.98
3rd Qu.:2250.2    3rd Qu.:45.00      Max. :1.0000      Max. :1.0000      3rd Qu.:12.000      3rd Qu.: 9.000      3rd Qu.:65.00
Max. :3000.0      Max. :54.00      Max. :1.0000      Max. :1.0000      Max. :15.000      Max. :17.100      Max. :87.00
(Other):1141
projects_completed tutorial_videos_watched uses_kaggle participates_in_discussion_forums debugging_sessions_per_week self_reported_confidence_python
Min. :0.000      Min. :19.0      Min. :0.0000      Min. :0.0000      Min. : 0.000      Min. : 1.000
1st Qu.:1.000    1st Qu.:35.0      1st Qu.:0.0000      1st Qu.:0.0000      1st Qu.: 3.000      1st Qu.: 3.000
Median :2.000      Median :40.0      Median :0.0000      Median :0.0000      Median : 5.000      Median : 6.000
Mean :2.001      Mean :39.9      Mean :0.4007      Mean :0.4983      Mean : 4.973      Mean : 5.566
3rd Qu.:3.000    3rd Qu.:44.0      3rd Qu.:1.0000      3rd Qu.:1.0000      3rd Qu.: 6.000      3rd Qu.: 8.000
Max. :9.000      Max. :63.0      Max. :1.0000      Max. :1.0000      Max. :17.000      Max. :10.000
final_exam_score passed_exam
Min. : 0.00      Min. :0.0000
1st Qu.:30.90    1st Qu.:0.0000
Median :43.10    Median :0.0000
Mean :43.32      Mean :0.1773
3rd Qu.:55.60    3rd Qu.:0.0000
Max. :100.00     Max. :1.0000
```

