**Assignment 2 Output**

Statistics and Data Science Major, Parami University

CS 251: Statistical Programming with R

Prof. Mohamed Megheib

Htet Khant Linn

October 4, 2025

> library(readr)

> library(dplyr)

>

>

> # ================= Part 1 =====================

>

> students\_scores\_txt <- read\_csv("data/students\_scores.txt")

**Rows:** 11 **Columns:** 3

── **Column specification** ──────────────────────────────────────────────────────────────────────────────────────────────────────────

**Delimiter:** ","

chr (1): Name

dbl (2): StudentID, Score

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.

>

> str(students\_scores\_txt)

spc\_tbl\_ [11 × 3] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)

$ StudentID: num [1:11] 1 2 3 4 5 6 7 8 9 10 ...

$ Name : chr [1:11] "John" "Sarah" "Alex" "Emily" ...

$ Score : num [1:11] 85 78 92 65 49 54 88 67 54 56 ...

- attr(\*, "spec")=

.. cols(

.. StudentID = col\_double(),

.. Name = col\_character(),

.. Score = col\_double()

.. )

- attr(\*, "problems")=<externalptr>

>

> # saving students\_scores data as .csv for .txt format

>

> write\_csv(students\_scores\_txt, "data/students\_scores.csv")

>

>

> # ================= Part 2 =====================

>

> # A. Data Exploration:

>

> # Reading students\_scores csv file

> students\_scores <- read\_csv("data/students\_scores.csv")

**Rows:** 11 **Columns:** 3

── **Column specification** ──────────────────────────────────────────────────────────────────────────────────────────────────────────

**Delimiter:** ","

chr (1): Name

dbl (2): StudentID, Score

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.

>

>

> # displaying the first 6 rows

> head(students\_scores)

# A tibble: 6 × 3

StudentID Name Score

*<dbl>* *<chr>* *<dbl>*

1 1 John 85

2 2 Sarah 78

3 3 Alex 92

4 4 Emily 65

5 5 Daniel 49

6 6 Anna 54

>

> # Summarizing the data set

>

> # no. of rows

> nrow(students\_scores)

[1] 11

>

> # checking missing values

> sum(is.na(students\_scores))

[1] 0

>

> summary(students\_scores$Score)

Min. 1st Qu. Median Mean 3rd Qu. Max.

49.00 54.50 65.00 67.55 81.50 92.00

>

> # ==============

>

> # B. Data Manipulation:

>

>

> # Creating new column "Pass"

> students\_scores\_2 <- students\_scores %>%

+ mutate(Pass = Score >= 50)

>

>

> # Filtering Passed Students

> passed\_students <- students\_scores\_2 %>%

+ filter(Pass == TRUE)

>

> passed\_students %>% select(Name)

# A tibble: 10 × 1

Name

*<chr>*

1 John

2 Sarah

3 Alex

4 Emily

5 Anna

6 Adam

7 Anna

8 Ahmed

9 Nada

10 Dina

>

> # ==============

>

> # C. Simple Statistics:

>

> # Calculating the average score of all students

> average\_score\_all <- mean(students\_scores$Score)

> average\_score\_all

[1] 67.54545

>

> # The highest and lowest scores

> highest\_score <- max(students\_scores$Score)

> highest\_score

[1] 92

> lowest\_score <- min(students\_scores$Score)

> lowest\_score

[1] 49

>

> # the average score of students who passed

> avarage\_score\_passed <- mean(passed\_students$Score)

> avarage\_score\_passed

[1] 69.4

>

>

> # ==============

>

> # D. Data Visualization:

>

> # Creating scores hist

> ?hist

> hist(students\_scores$Score,

+ main = "Histogram of Students Scores",

+ xlab = "Score")

>

> # Creating barplot showing the number of students who passed and failed

> args(barplot) # checking arguemnts

function (height, ...)

NULL

>

> pass\_fail\_counts <- table(students\_scores\_2$Pass)

> names(pass\_fail\_counts) <- c("Failed", "Passed")

>

> ?barplot

> barplot(pass\_fail\_counts,

+ main = "Number of Students who Passed and Failed",

+ xlab = "Result",

+ ylab = "Number of Students")

>

>

>

> # ==============

>

> # E. Output:

>

> write\_csv(passed\_students, "data/passed\_students.csv")



