

# Probability with R - Exercise 1.1

Lucas Liona

## Table of contents

Problem . . . . .	1
<b>Approach 1: Two Separate Vectors</b>	<b>2</b>
<b>Approach 2: Single Vector with Factor</b>	<b>2</b>
<b>Part (b): Correcting Data Entry Error</b>	<b>3</b>
<b>Part (c): Saving the Workspace</b>	<b>4</b>
<b>Additional Analysis</b>	<b>4</b>
Summary Stats . . . . .	4
<b>Visualization</b>	<b>5</b>
<b>Conclusion</b>	<b>6</b>

## Problem

In a class of 50 students of computing, 23 are female and 27 are male. The results of their first year Java programming examination are as follows:

Females: 57, 59, 78, 79, 60, 65, 68, 71, 75, 48, 51, 55, 56, 41, 43, 44, 75, 78, 80, 81, 83, 83, 85

Males: 48, 49, 49, 30, 30, 31, 32, 35, 37, 41, 86, 42, 51, 53, 56, 42, 44, 50, 51, 65, 67, 51, 56, 58, 64, 64, 75 (a) Read these data into R by storing them in the following ways:

As two vectors, one for the females and one for the males. As one vector, with a factor vector designating the gender.

- (b) If it was discovered that the 34th was entered incorrectly and should have obtained the marks 46 instead of 86, use an appropriate editing procedure to change this.

(c) Save the workspace in a suitable directory for access later.

Part (a): Reading Data into R First, we'll read the data into R using the two approaches mentioned in the problem.

## Approach 1: Two Separate Vectors

```
# Create separate vectors for female and male scores
females <- c(57, 59, 78, 79, 60, 65, 68, 71, 75, 48, 51, 55, 56, 41, 43, 44, 75, 78, 80, 81,
males <- c(48, 49, 49, 30, 30, 31, 32, 35, 37, 41, 86, 42, 51, 53, 56, 42, 44, 50, 51, 65, 67,

# Verify the length of each vector
cat("Number of female scores:", length(females), "\n")
```

Number of female scores: 23

```
cat("Number of male scores:", length(males), "\n")
```

Number of male scores: 27

## Approach 2: Single Vector with Factor

```
# Create a single vector with all scores
marks <- c(57, 59, 78, 79, 60, 65, 68, 71, 75, 48, 51, 55, 56, 41, 43, 44, 75, 78,
          80, 81, 83, 83, 85, 48, 49, 49, 30, 30, 31, 32, 35, 37, 41, 86, 42, 51,
          53, 56, 42, 44, 50, 51, 65, 67, 51, 56, 58, 64, 64, 75)

# Create a factor for gender
gender <- rep(c("Female", "Male"), c(23, 27))

# Create a data frame for easier manipulation
student_data <- data.frame(marks = marks, gender = gender)

# Display the first few rows of the data frame
head(student_data)
```

```
marks gender
1    57 Female
2    59 Female
3    78 Female
4    79 Female
5    60 Female
6    65 Female
```

## Part (b): Correcting Data Entry Error

```
# Check the value before correction
cat("Value before correction (marks[34]):", marks[34], "\n")
```

Value before correction (marks[34]): 86

```
cat("Value before correction (student_data$marks[34]):", student_data$marks[34], "\n")
```

Value before correction (student\_data\$marks[34]): 86

```
cat("Value before correction in males vector (males[11]):", males[11], "\n")
```

Value before correction in males vector (males[11]): 86

```
# Make the corrections
marks[34] <- 46
student_data$marks[34] <- 46
males[11] <- 46 # This is the 11th element in males vector

# Verify the changes
cat("Value after correction (marks[34]):", marks[34], "\n")
```

Value after correction (marks[34]): 46

```
cat("Value after correction (student_data$marks[34]):", student_data$marks[34], "\n")
```

Value after correction (student\_data\$marks[34]): 46

```
cat("Value after correction in males vector (males[11]):", males[11], "\n")
```

Value after correction in males vector (males[11]): 46

```
# Show the specific row in the data frame
student_data[34,]
```

```
marks gender
34    46  Male
```

## Part (c): Saving the Workspace

```
# Save all objects in the current session
save.image("class_marks.RData")

# Alternatively, save specific objects
save(females, males, marks, gender, student_data, file = "class_marks.RData")
```

## Additional Analysis

### Summary Stats

```
# Summary statistics for female scores
cat("Female summary statistics:\n")
```

Female summary statistics:

```
summary(females)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
41.00	55.50	68.00	65.87	78.50	85.00

```
# Summary statistics for male scores
cat("\nMale summary statistics:\n")
```

Male summary statistics:

```
summary(males)
```

```
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
30.00   41.50   49.00   48.78   56.00   75.00
```

```
# Create a comparison table
stats_comparison <- data.frame(
  Gender = c("Female", "Male"),
  Min = c(min(females), min(males)),
  Mean = c(mean(females), mean(males)),
  Median = c(median(females), median(males)),
  Max = c(max(females), max(males))
)

# Display the comparison table
knitr::kable(stats_comparison, digits = 2, caption = "Comparison of Exam Scores by Gender")
```

Table 1: Comparison of Exam Scores by Gender

Gender	Min	Mean	Median	Max
Female	41	65.87	68	85
Male	30	48.78	49	75

## Visualization

```
# Set margins for the plot
par(mar = c(4, 4, 2, 1)) # Bottom, left, top, right margins

# Create the boxplot
boxplot(marks ~ gender, data = student_data,
        main = "Java Programming Exam Scores by Gender",
```

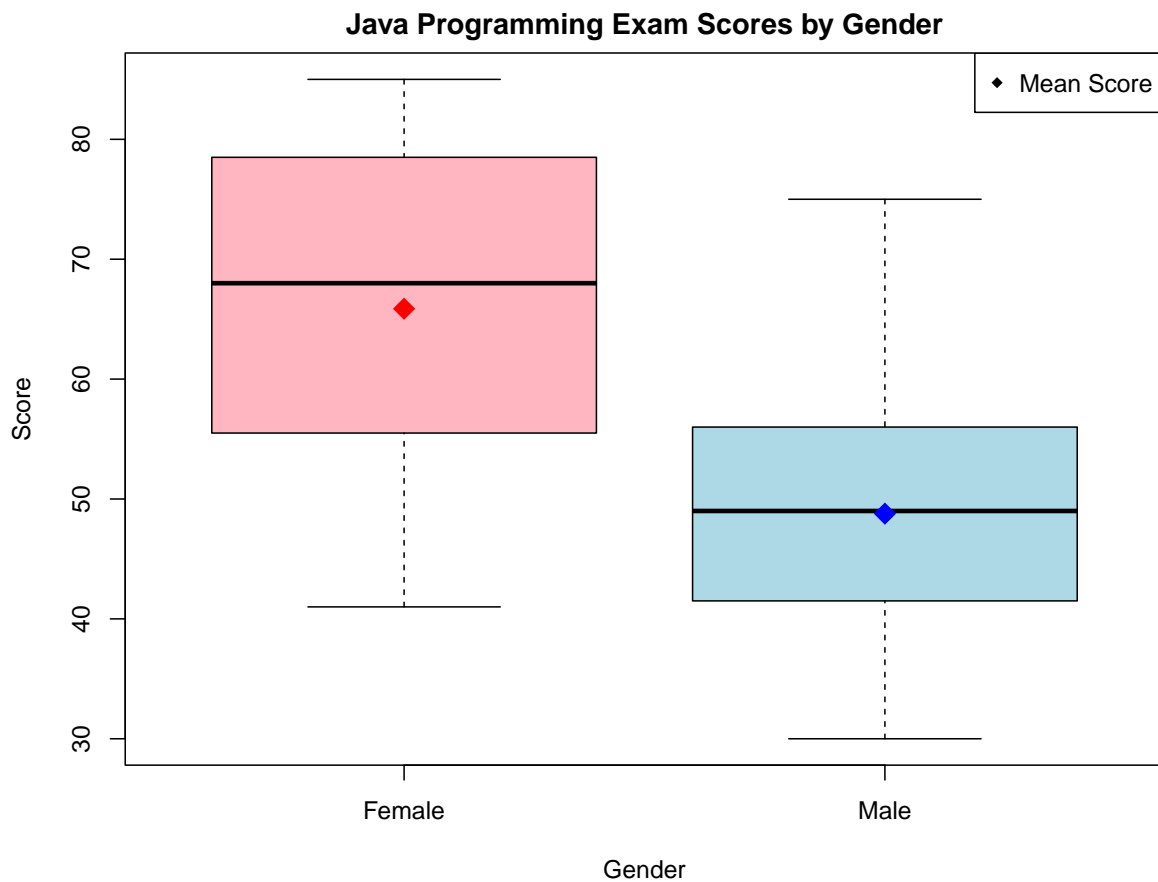
```

xlab = "Gender", ylab = "Score",
col = c("lightpink", "lightblue"))

# Add mean points to the boxplot
points(1, mean(females), pch = 18, col = "red", cex = 2)
points(2, mean(males), pch = 18, col = "blue", cex = 2)

# Add a legend
legend("topright", legend = "Mean Score", pch = 18, col = "black")

```



## Conclusion

From our analysis, we can observe that female students (mean =  $\text{round}(\text{mean}(\text{females}), 2)$ ) performed better on average than male students (mean =  $\text{round}(\text{mean}(\text{males}), 2)$ ) in this

Java programming examination. The boxplot shows not only the difference in central tendency but also the spread of scores within each group.