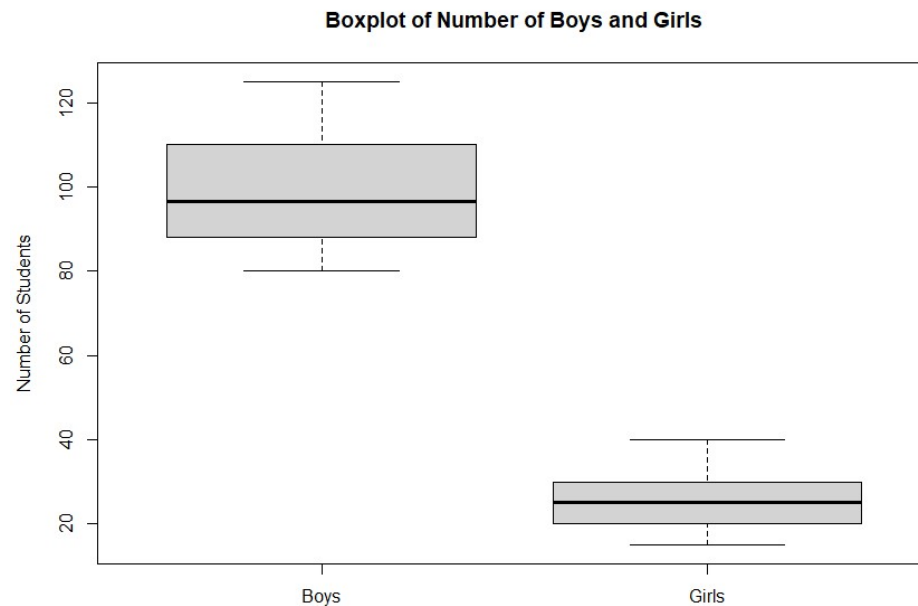Question 01

a) R commands:
   > number.of.boys <- c(80, 82, 85, 90, 88, 92, 95, 98, 100, 105, 110, 115, 120, 125)
   > number.of.girls <- c(20, 22, 25, 30, 35, 40, 26, 30, 20, 35, 19, 25, 20, 15)

b)

Boxplot of Number of Boys and Girls

R command: > boxplot(number.of.boys, number.of.girls, names = c("Boys", "Girls"), main = "Boxplot of Number of Boys and Girls", ylab = "Number of Students")

Comments:
The centre of the boys' data (median) is around 100, while the centre of the girls' data is around 25. The variability (interquartile range) of the boys' data is larger than that of the girls' data. The boys' data has a higher maximum value (125) compared to the girls' data (40).

Question 02

a) > jem_data <- c(1.71, 1.73, 1.81, 1.82, 1.80)
    > kimber_data <- c(171, 173, 181, 182, 180)
    > mean_jem <- mean(jem_data)
    > sd_jem <- sd(jem_data)
    > mean_kimber <- mean(kimber_data)
    > sd_kimber <- sd(kimber_data)
    > mean_jem
    [1] 1.774
    > sd_jem
    [1] 0.05029911
    > mean_kimber
    [1] 177.4
    > sd_kimber
    [1] 5.029911

    Mean Jem: 1.774
    Standard Deviation Jem: 0.05029911
    Mean Kimber: 177.4
    Standard Deviation Kimber: 5.029911

b) > cv_jem <- (sd_jem / mean_jem) * 100
    > cv_kimber <- (sd_kimber / mean_kimber) * 100
    >
    > cv_jem
    [1] 2.83535
    > cv_kimber
    [1] 2.83535

    Coefficient of variation Jem: 2.83535
    Coefficient of variation Kimber: 2.83535

Question 03

a)

Histogram of Simulated Data

Boxplot of Simulated Data
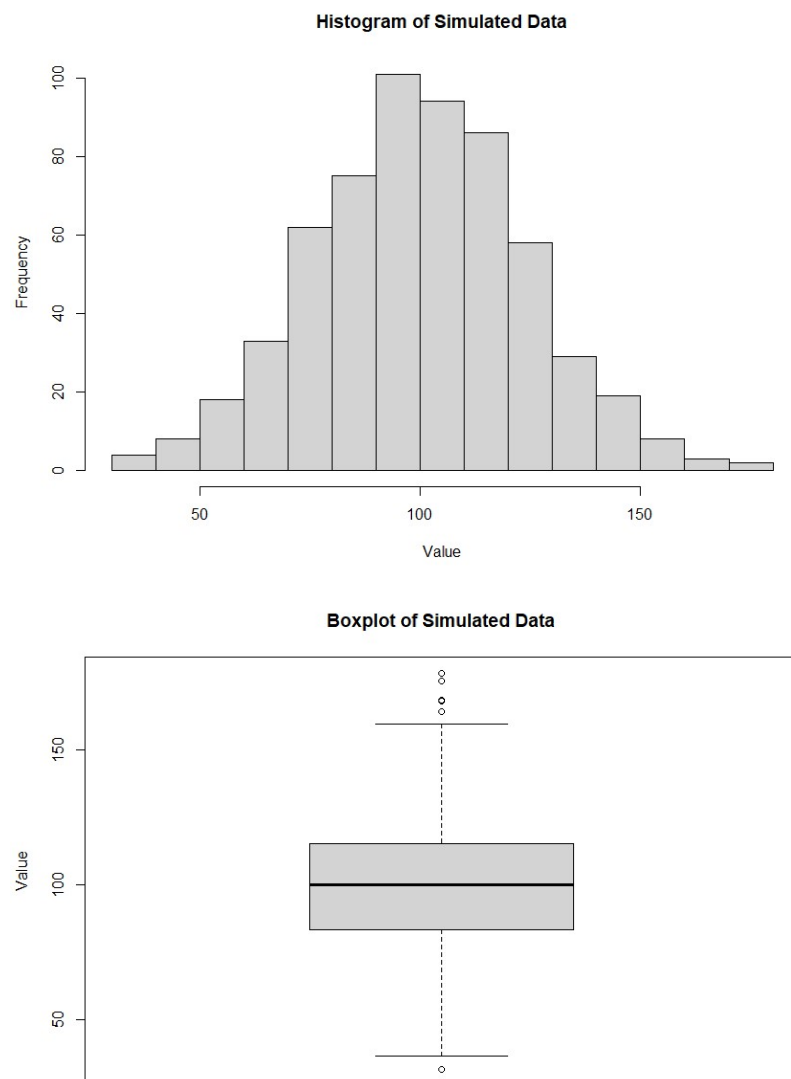
Commands in R:
set.seed(2025)
simvector <- rnorm(600, 100, 24)
hist(simvector, main = "Histogram of Simulated Data", xlab = "Value")
boxplot(simvector, main = "Boxplot of Simulated Data", ylab = "Value")

b) > iqr_simvector <- IQR(simvector)
   > iqr_simvector
   [1] 31.92888

   Interquartile Range: 31.92888

c) The histogram and boxplot suggest that the dataset is approximately normally distributed. There is one peak in the histogram, indicating unimodality. The boxplot shows no significant outliers, and the median is close to the mean, suggesting symmetry.

Question 04

a)

Scatterplot of Study Hours vs Exam Scores
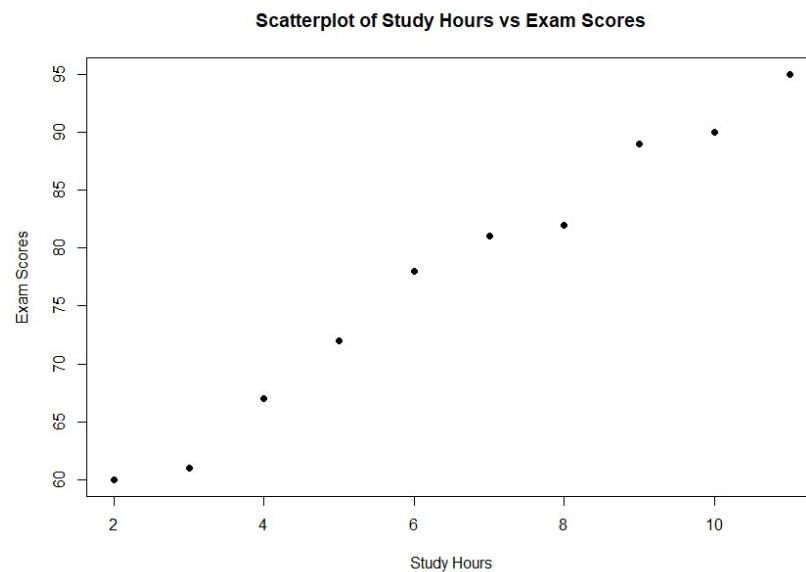
R commands:
x <- c(2, 3, 4, 5, 6, 7, 8, 9, 10, 11)
y <- c(60, 61, 67, 72, 78, 81, 82, 89, 90, 95)

plot(x, y, main = "Scatterplot of Study Hours vs Exam Scores", xlab = "Study Hours", ylab = "Exam Scores", pch = 19)

b) > correlation_coefficient <- cor(x, y)
> correlation_coefficient
[1] 0.9916109

Correlation coefficient: 0.9916109

c) The scatterplot shows a positive linear relationship between study hours and exam scores. The correlation coefficient is positive, indicating a positive linear relationship. The straight line would have a positive slope.