Question 01

1. 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)

1. **A graph of a number of boys and girls

   Description automatically generated**  
     
   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

1. > 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

1. > 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

1. A graph of a graph of a graph

   Description automatically generated with medium confidenceA diagram of a box plot

   Description automatically generated  
     
   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")

1. > iqr\_simvector <- IQR(simvector)

> iqr\_simvector

[1] 31.92888  
  
**Interquartile Range:** 31.92888

1. 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

1. A graph with black dots

   Description automatically generated  
   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)

1. > correlation\_coefficient <- cor(x, y)

> correlation\_coefficient

[1] 0.9916109  
  
**Correlation coefficient:** 0.9916109

1. 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.