1. The following values are the number of pencils available in the different boxes. Create a vector and find out the mean, median and mode values of set of pencils in the given data.

| Box1 | Box2 | Box3 | Box4 | Box5 | Box6 | Box7 | Box8 | Box9 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 25 | 23 | 12 | 11 | 6 | 7 | 8 | 9 | 10 |

**CODE**:

pencils <- c(25, 23, 12, 11, 6, 7, 8, 9, 10)

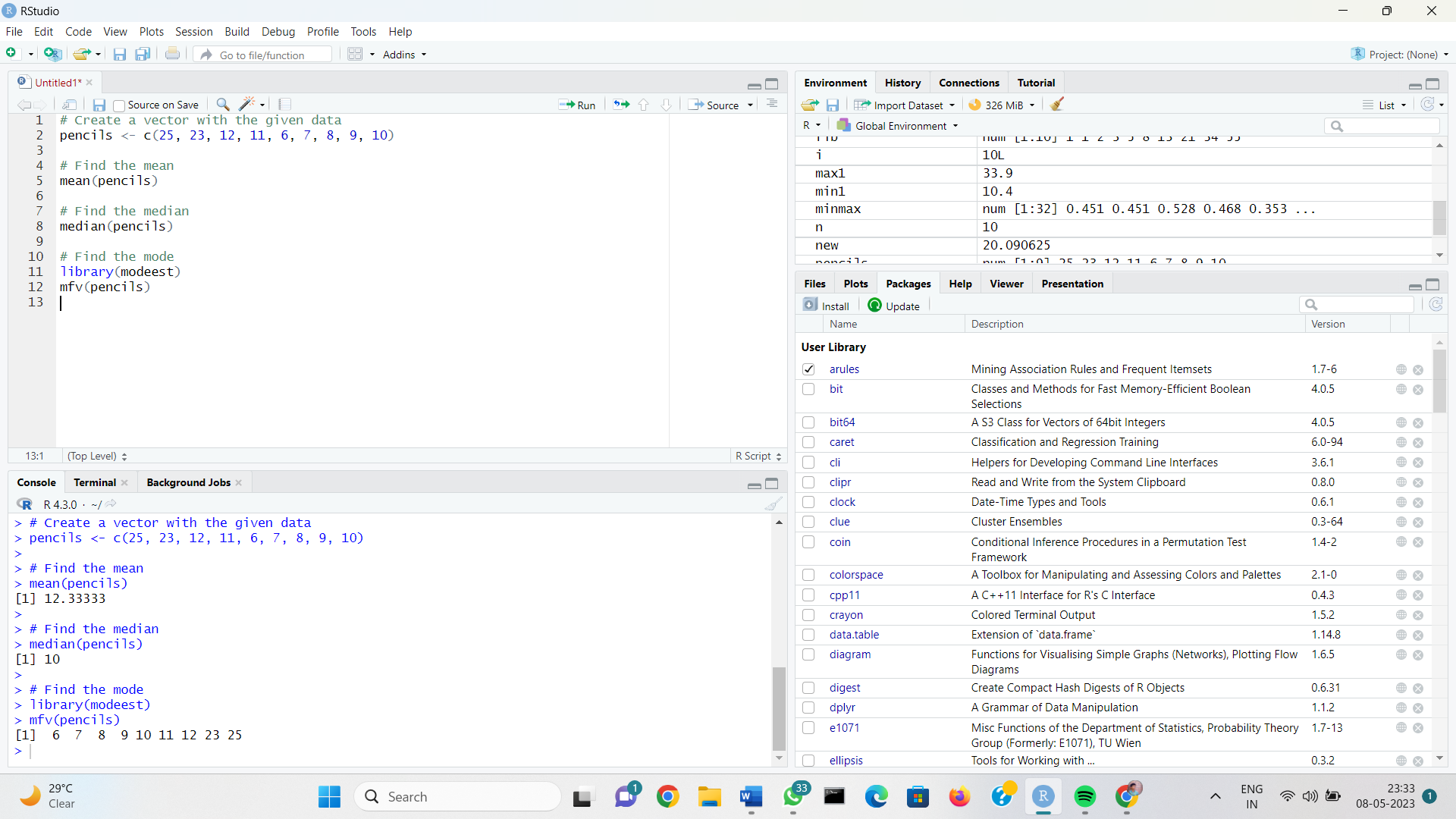
mean(pencils)

median(pencils)

library(modeest)

mfv(pencils)

**OUTPUT:**

****

**02. Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.**

**Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data?**

**CODE:**

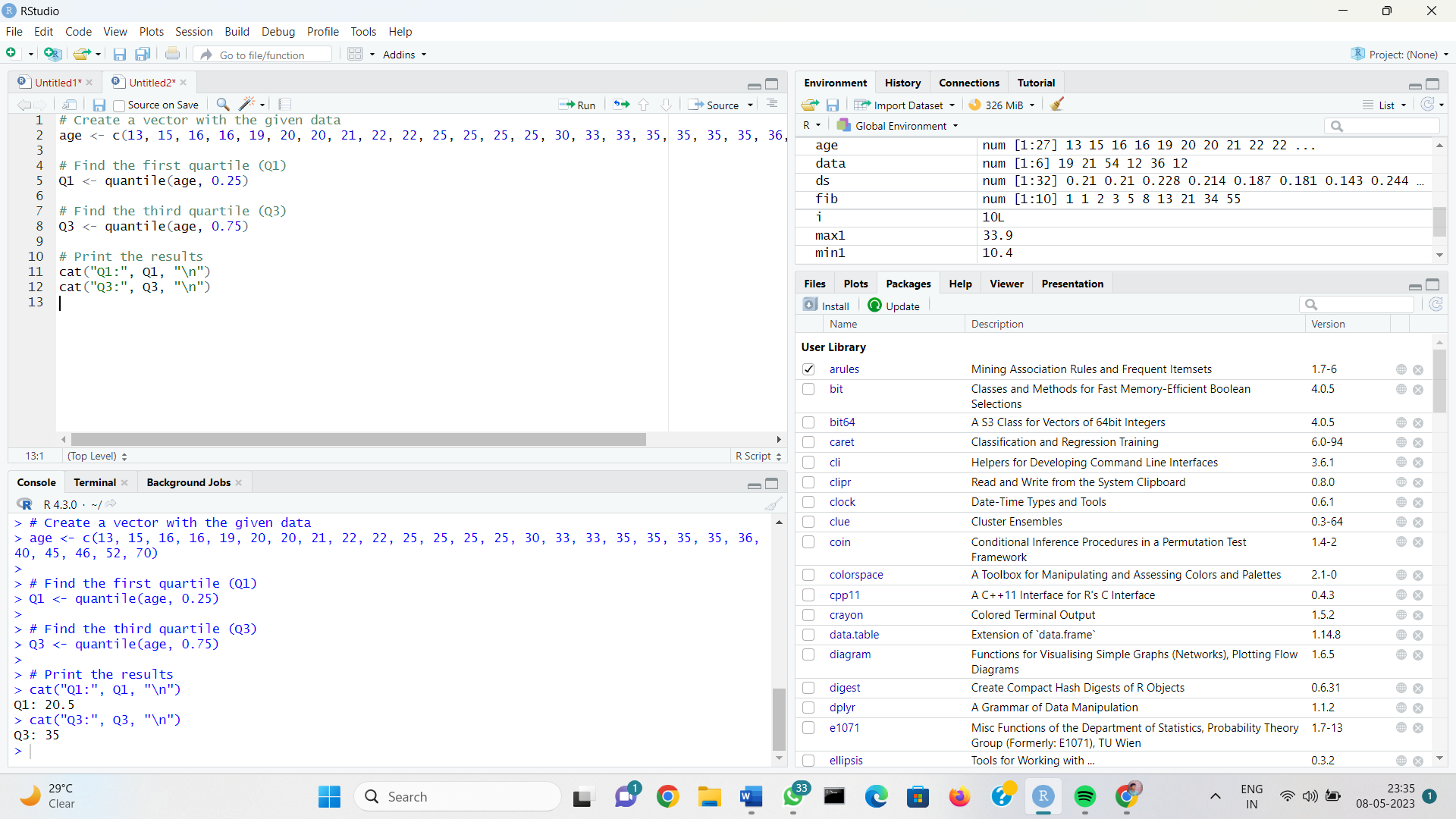
**age <- c(13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70)**

**Q1 <- quantile(age, 0.25)**

**Q3 <- quantile(age, 0.75)**

**cat("Q1:", Q1, "\n")**

**cat("Q3:", Q3, "\n")**

****

**03. Two Maths teachers are comparing how their Year 9 classes performed in the end of year exams. Their results are as follows:**

**Class A: 76, 35, 47, 64, 95, 66, 89, 36, 8476,35,47,64,95,66,89,36,84**

**Class B: 51, 56, 84, 60, 59, 70, 63, 66, 5051,56,84,60,59,70,63,66,50**

**(i) Find which class had scored higher mean, median and range.**

**(ii) Plot above in boxplot and give the inferences**

**CODE:**

classA <- c(76, 35, 47, 64, 95, 66, 89, 36, 84)

classB <- c(51, 56, 84, 60, 59, 70, 63, 66, 50)

meanA <- mean(classA)

meanB <- mean(classB)

medianA <- median(classA)

medianB <- median(classB)

rangeA <- max(classA) - min(classA)

rangeB <- max(classB) - min(classB)

cat("Class A Mean:", meanA, "\n")

cat("Class B Mean:", meanB, "\n")

cat("Class A Median:", medianA, "\n")

cat("Class B Median:", medianB, "\n")

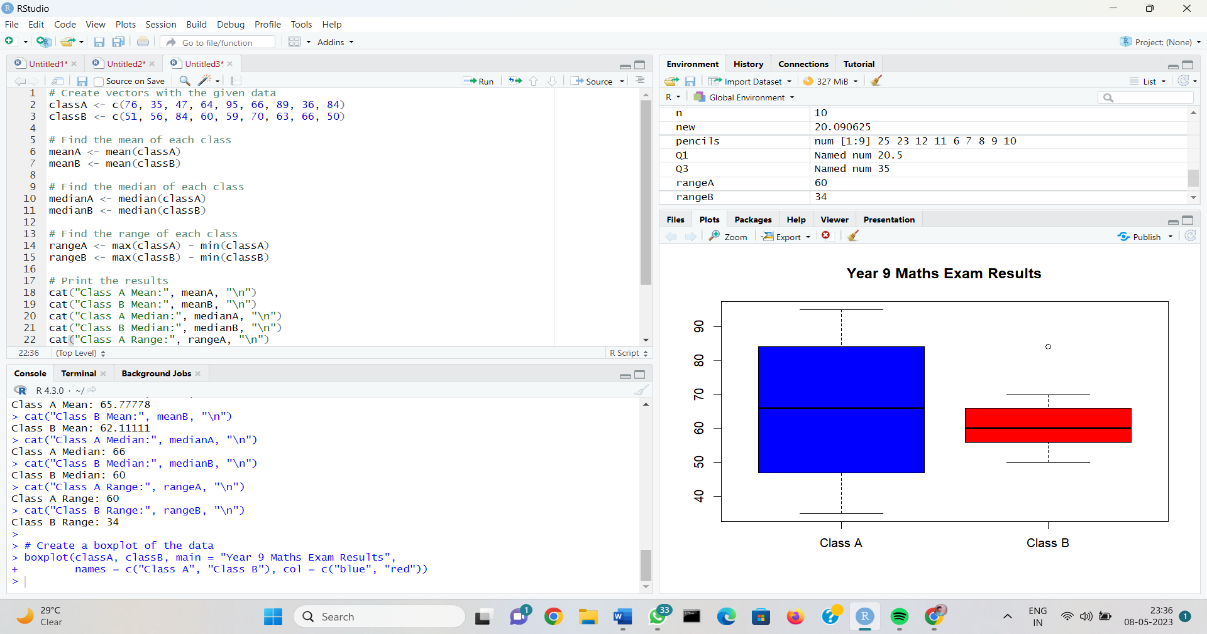
cat("Class A Range:", rangeA, "\n")

cat("Class B Range:", rangeB, "\n")

boxplot(classA, classB, main = "Year 9 Maths Exam Results",

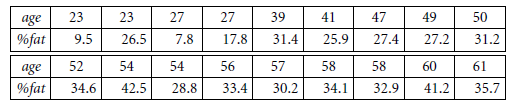
names = c("Class A", "Class B"), col = c("blue", "red"))

**OUTPUT:**

****

**04. (a) Suppose that a hospital tested the age and body fat data for 18 randomly selected adults**

**with the following results:**

****

**(a) Calculate the mean, median, and standard deviation of *age* and *%fat*.**

**(b) Draw the boxplots for *age* and *%fat*.**

**(c) Draw a *scatter plot* and a *q-q plot* based on these two variables**

**CODE:**

**# Create a data frame with the given data**

**data <- data.frame(**

**age = c(23, 33, 48, 30, 52, 42, 37, 27, 53, 39, 32, 41, 45, 29, 25, 37, 46, 51),**

**fat\_pct = c(30.3, 25.3, 41.3, 32.1, 34.6, 39.7, 29.5, 30.1, 35.7, 39.3, 22.8, 36.8, 38.1, 27.2, 28.8, 33.4, 40.6, 38.8)**

**)**

**# Calculate the mean, median, and standard deviation of age and %fat**

**age\_mean <- mean(data$age)**

**age\_median <- median(data$age)**

**age\_sd <- sd(data$age)**

**fat\_mean <- mean(data$fat\_pct)**

**fat\_median <- median(data$fat\_pct)**

**fat\_sd <- sd(data$fat\_pct)**

**# Print the results**

**cat("Age Mean:", age\_mean, "\n")**

**cat("Age Median:", age\_median, "\n")**

**cat("Age SD:", age\_sd, "\n")**

**cat("%Fat Mean:", fat\_mean, "\n")**

**cat("%Fat Median:", fat\_median, "\n")**

**cat("%Fat SD:", fat\_sd, "\n")**

**# Draw boxplots for age and %fat**

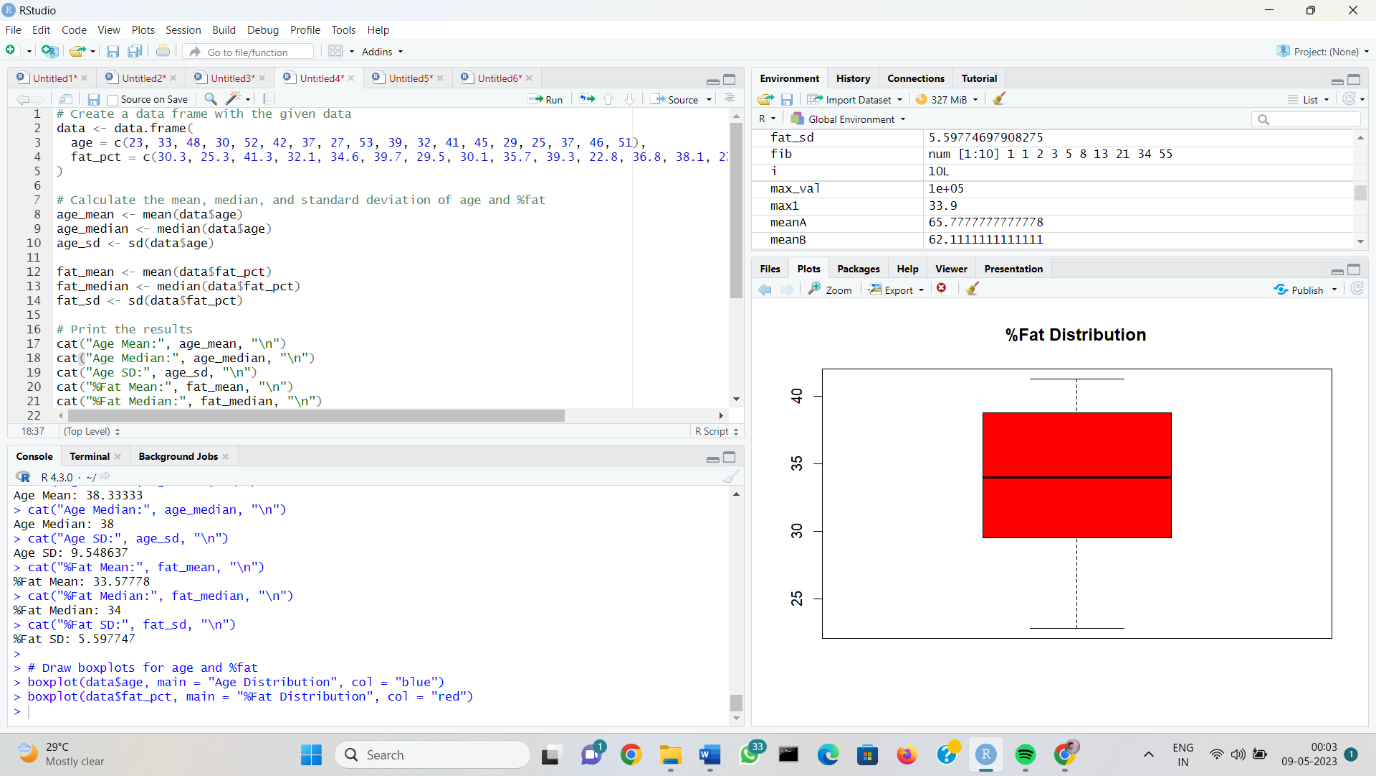
**boxplot(data$age, main = "Age Distribution", col = "blue")**

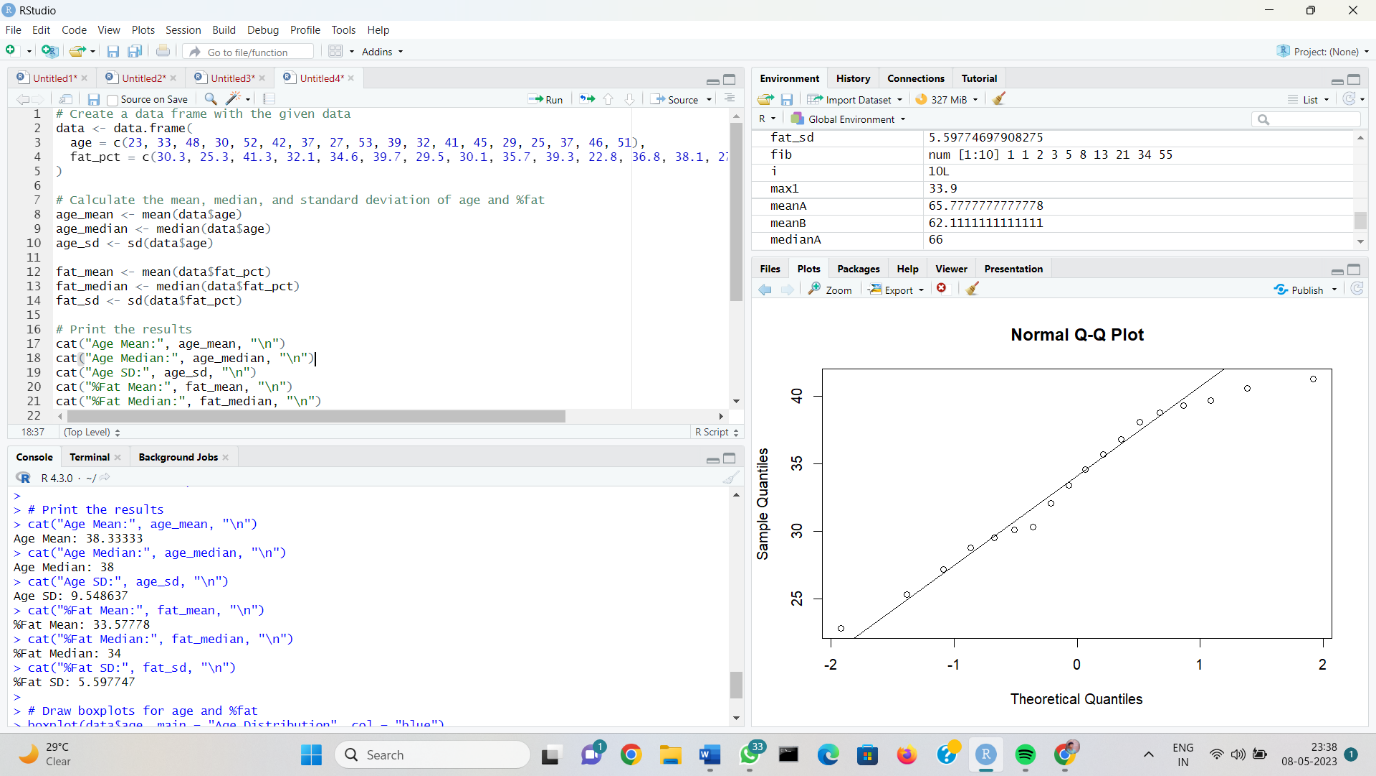
**boxplot(data$fat\_pct, main = "%Fat Distribution", col = "red")**

**# Create a scatter plot and q-q plot**

**plot(data$age, data$fat\_pct, main = "Scatter Plot of Age vs %Fat", xlab = "Age", ylab = "%Fat")**

**qqnorm(data$fat\_pct)**

**qqline(data$fat\_pct)**

****

**05. Create the Confusion matrix using this scenario:**

**A shepherd boy gets bored tending the town's flock. To have some fun, he cries out, "Wolf!" even though no wolf is in sight. The villagers run to protect the flock, but then get really mad when they realize the boy was playing a joke on them.One night, the shepherd boy sees a real wolf approaching the flock and calls out, "Wolf!" The villagers refuse to be fooled again and stay in their houses. The hungry wolf turns the flock into lamb chops. The town goes hungry. Panic ensues.**

**CODE:**

actual <- c("No Wolf", "Wolf")

predicted <- c("No Wolf", "Wolf")

true\_neg <- 1 # The boy cried wolf but there was no wolf, and the villagers didn't react

false\_pos <- 1 # The boy cried wolf but there was no wolf, and the villagers did react

false\_neg <- 1 # The boy saw a wolf but the villagers didn't react

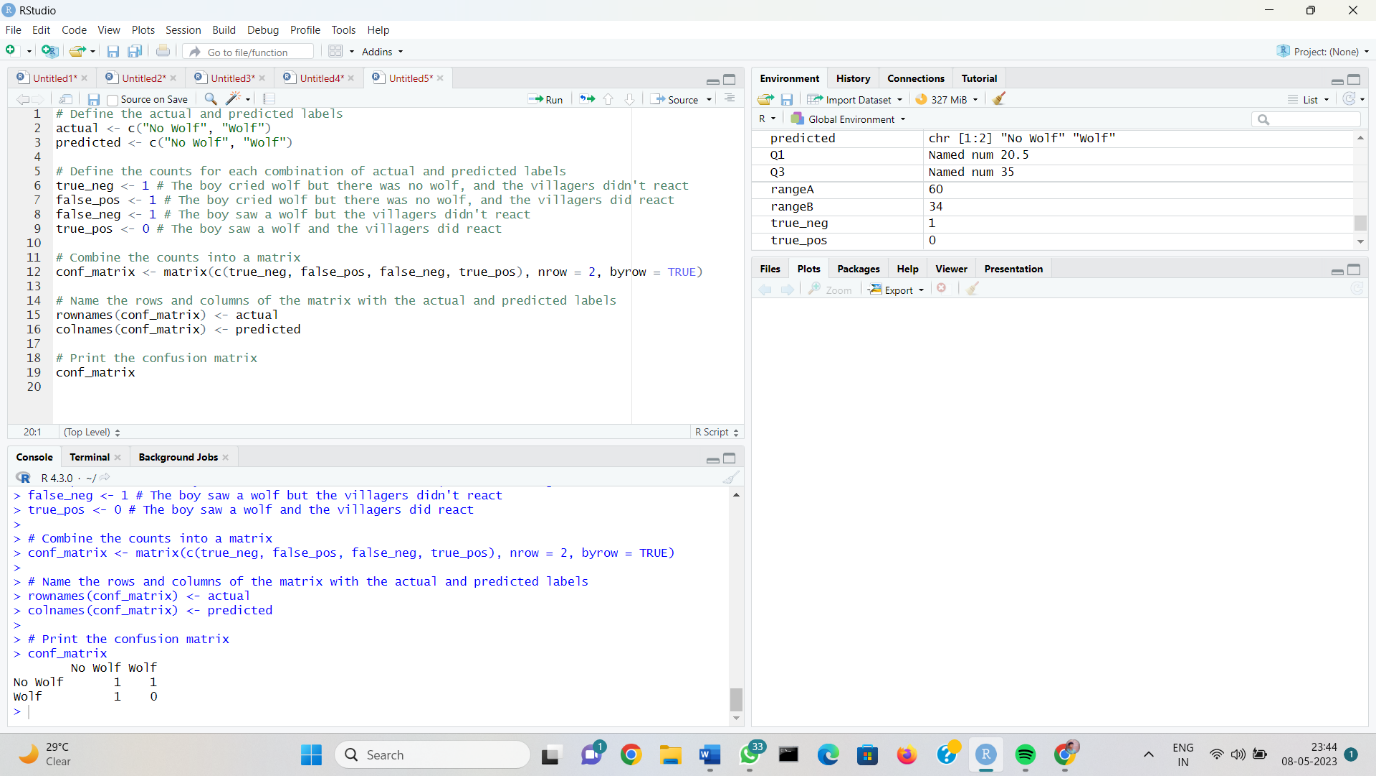
true\_pos <- 0 # The boy saw a wolf and the villagers did react

conf\_matrix <- matrix(c(true\_neg, false\_pos, false\_neg, true\_pos), nrow = 2, byrow = TRUE)

rownames(conf\_matrix) <- actual

colnames(conf\_matrix) <- predicted

conf\_matrix



**06. a)Let us consider one example to make the calculation method clear. Assume that the minimum and maximum values for the feature F are $50,000 and $100,000 correspondingly. It needs to range F from 0 to 1. In accordance with min-max normalization, v = $80,**

**b) Use the two methods below to normalize the following group of data: 200, 300, 400, 600, 1000**

**(a) min-max normalization by setting min = 0 and max = 1**

**(b) z-score normalization**

**CODE:**

**min\_val <- 50000**

**max\_val <- 100000**

**v <- 80000**

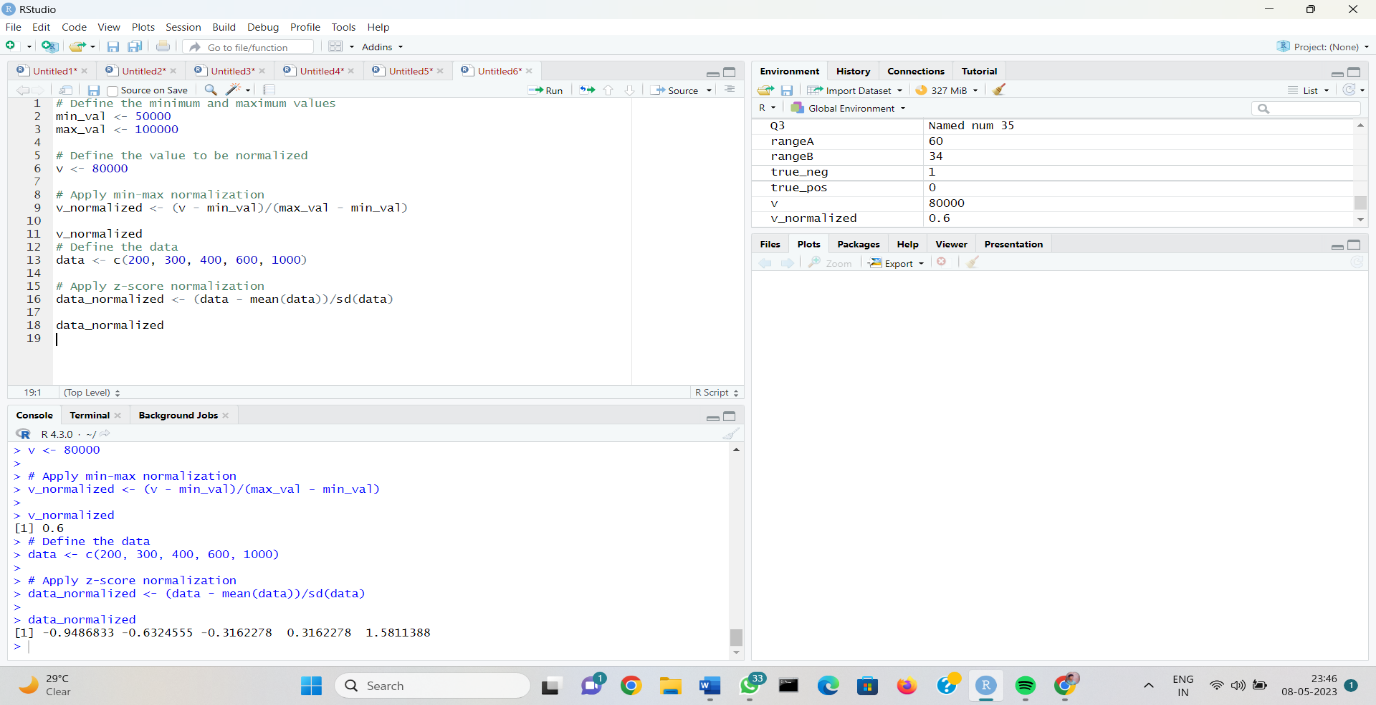
**v\_normalized <- (v - min\_val)/(max\_val - min\_val)**

**v\_normalized**

**data <- c(200, 300, 400, 600, 1000)**

**data\_normalized <- (data - mean(data))/sd(data)**

**data\_normalized**

****