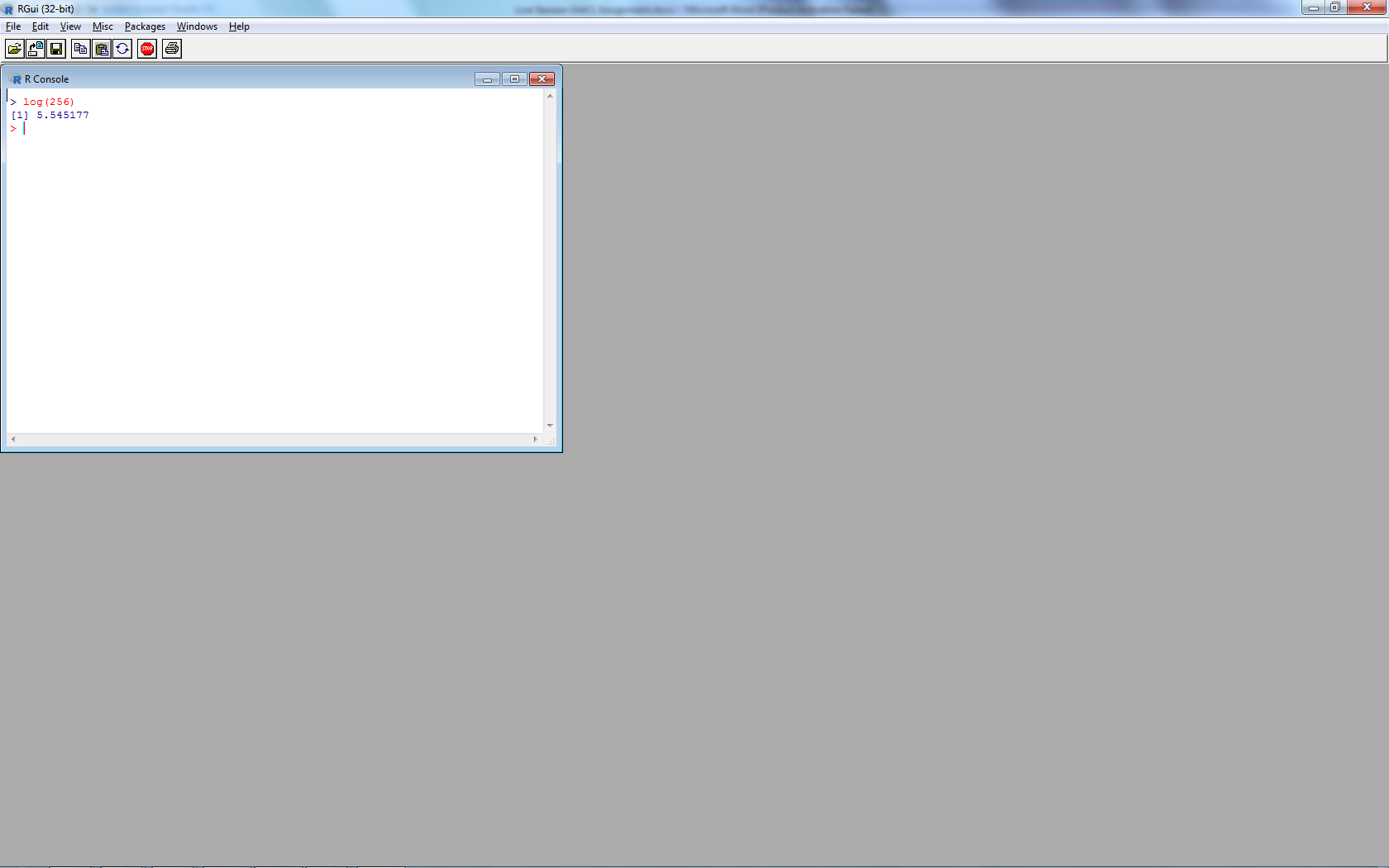
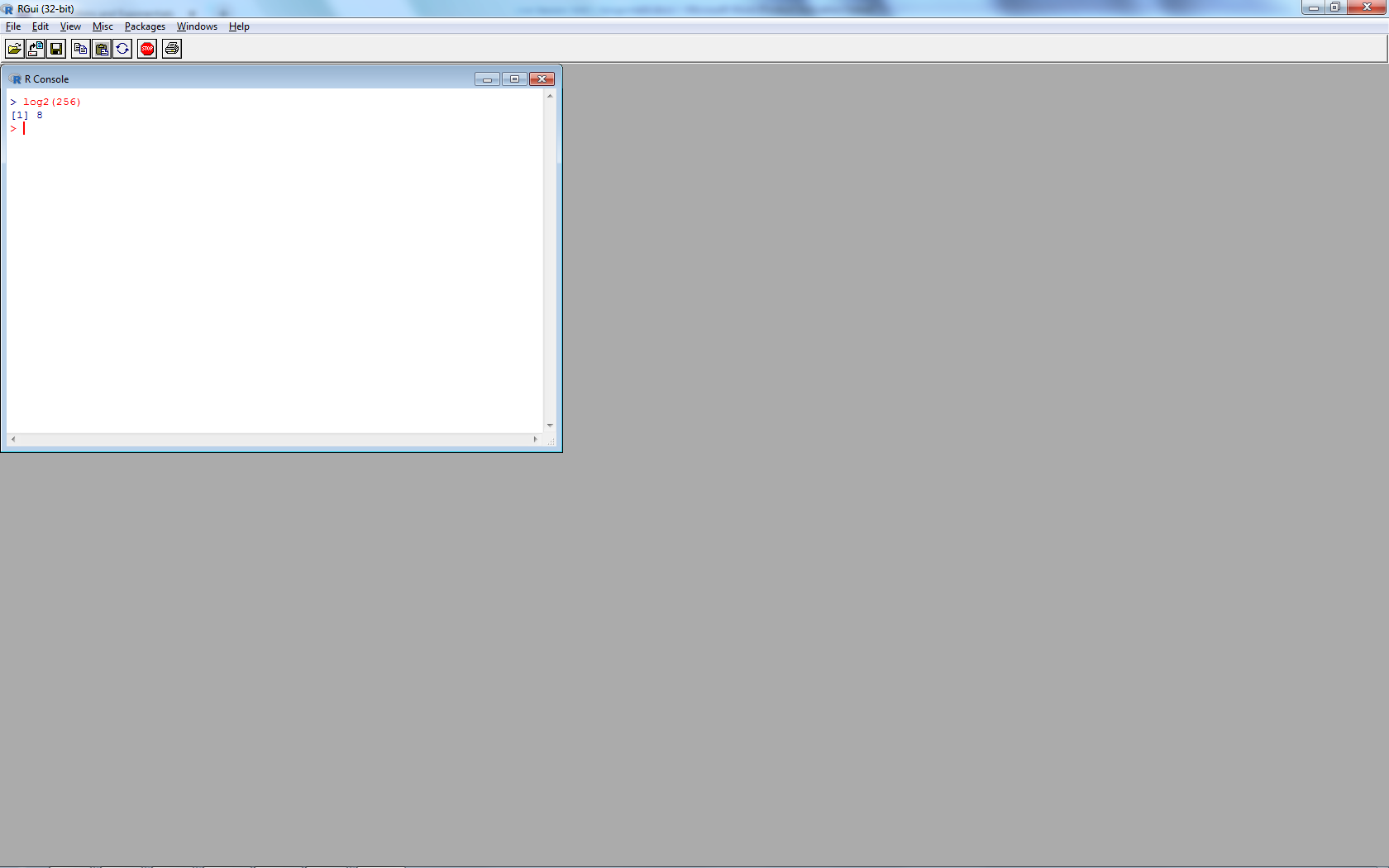
**Live Session Unit 1 Assignment**

**1.** (20 points) Basic Math – each question has 5% credits

**a.** The log of a positive number.

log(256)

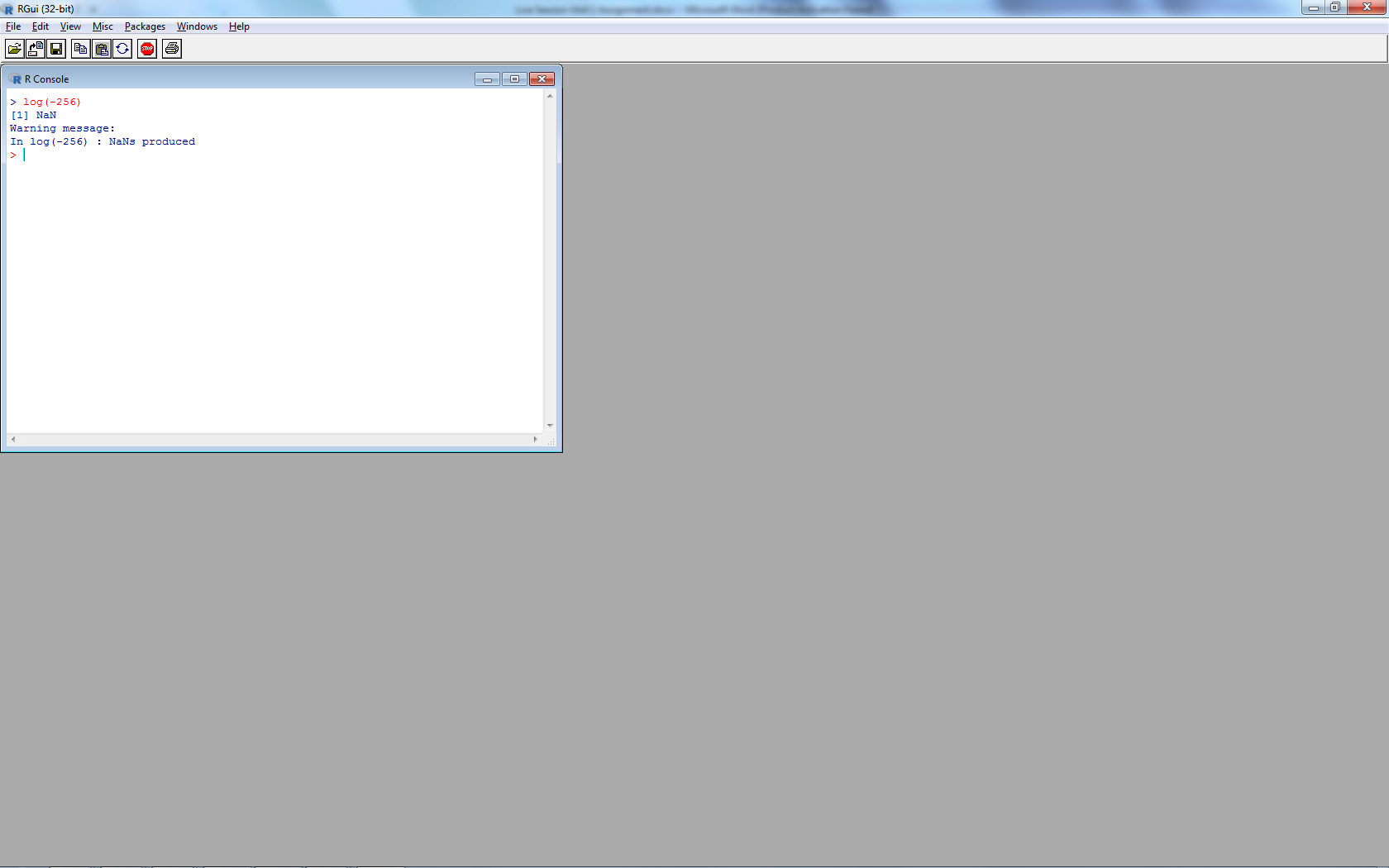
**b.** What is the default base for the log function?

Calculate the log of your previous number with a

different base.

log2(256)

**c.** The log of a negative number. (explain the answer)

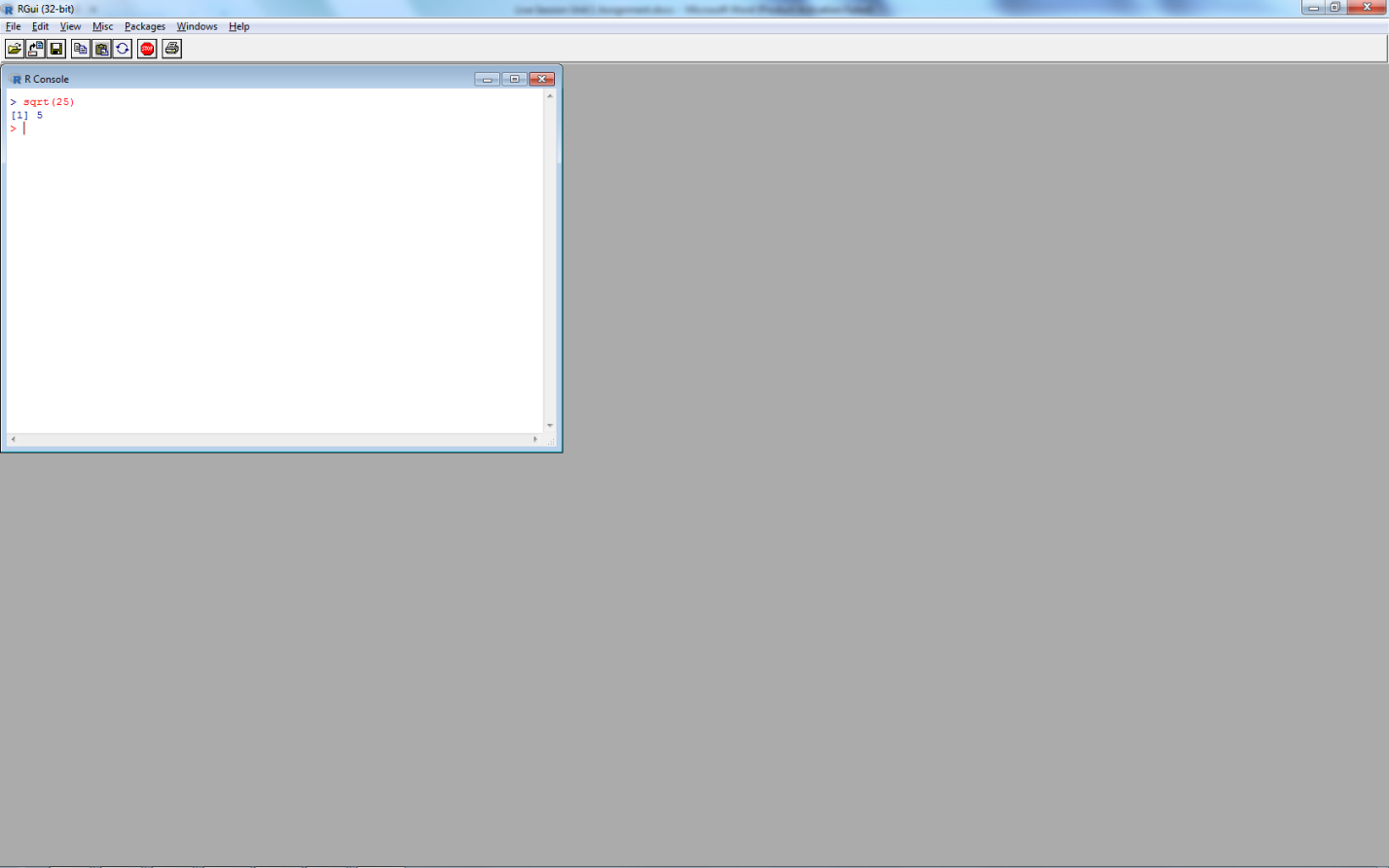


log(-256)

The logarithm of a negative number is not defined.

Therefore, R returns NAN (not a number) as the

result.



**d.** The square-root of a positive number.

sqrt(25)

**2.** (15 points) Random number generation.

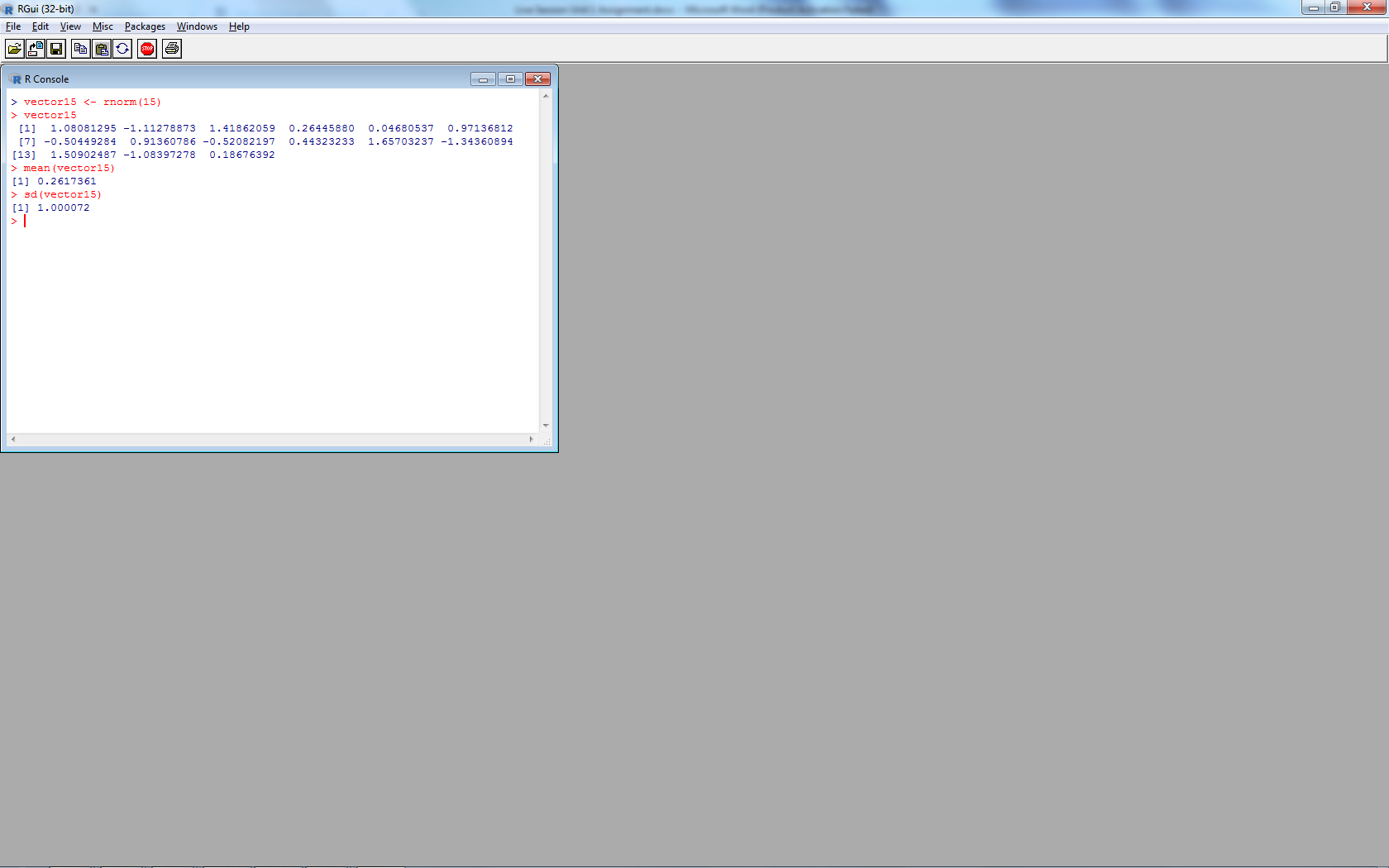
**a.** Create a vector of 15 standard normal random variables. Calculate its mean and

SD (Standard Deviation).

vector15 <- rnorm(15)

mean(vector15)

sd(vector15)



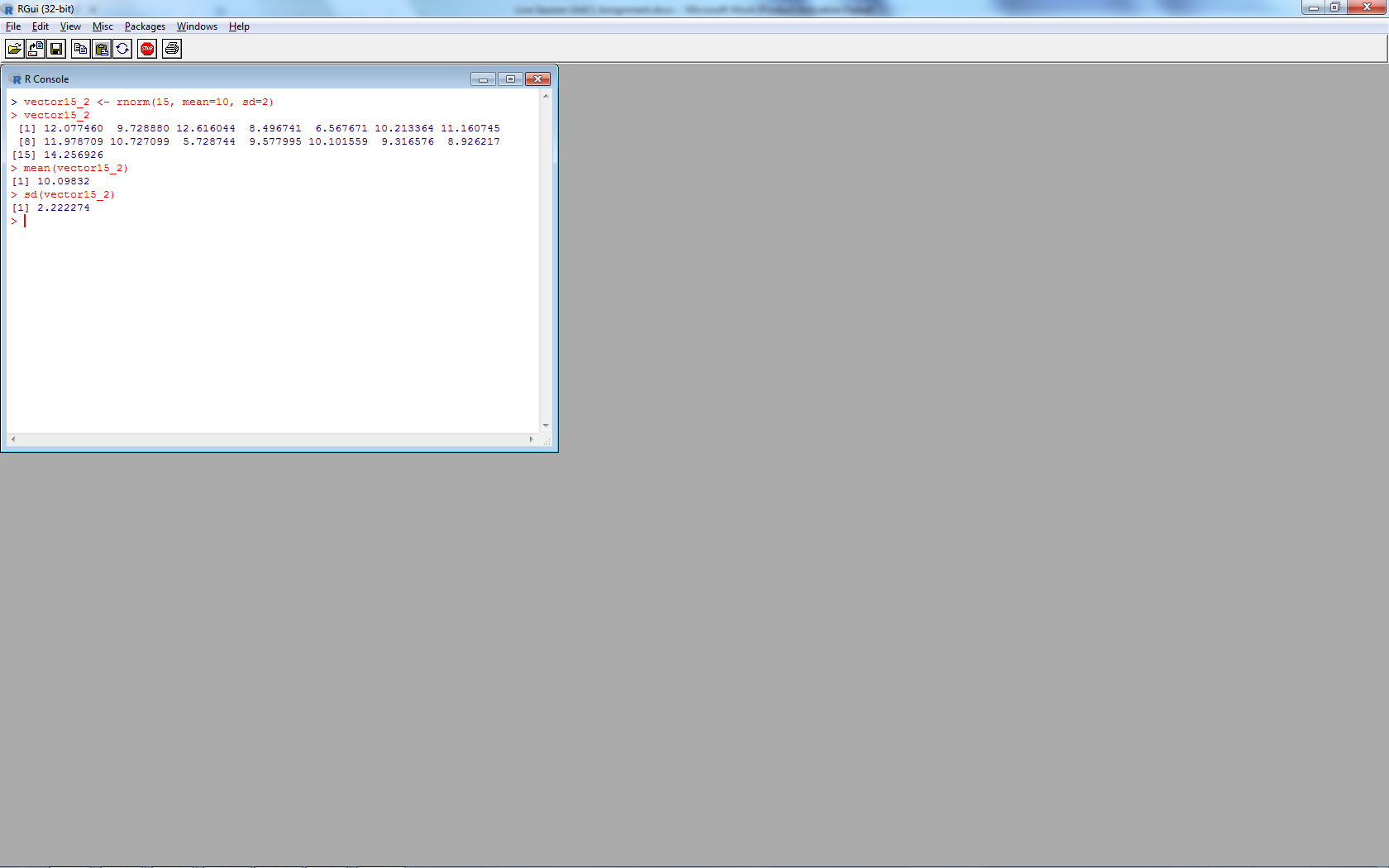
**b.** Change the mean to 10 and the SD to 2 and recalculate the vector of 15 random

normal variables. Calculate its mean and SD.

vector15\_2 <- rnorm(15, mean = 10, sd=2)

mean(vector15\_2)

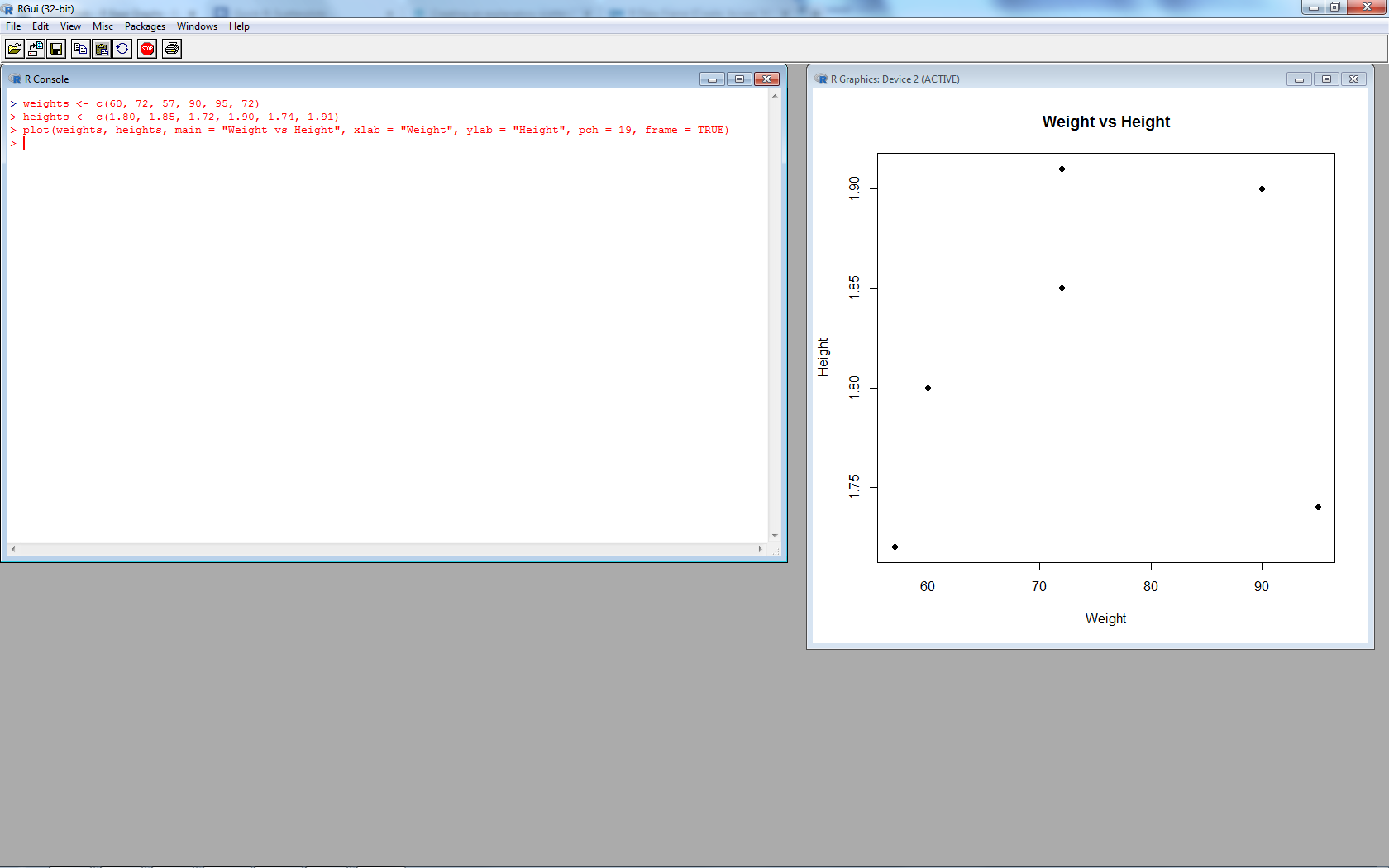
sd(vector15\_2)



**c.** Why are the means and SD not exactly the same as the means and SDs specified in

the function? The difference is most likely attributed to the rounding of significant digits.

**3.** (40 points) Vector Operations

**a.** The weights of 6 individuals in kg are 60, 72, 57, 90, 95, 72.

**b.** Their heights (in m) are 1.80, 1.85, 1.72, 1.90, 1.74, 1.91.

**c.** Enter these vectors into R.

weights <- c(60, 72, 57, 90, 95, 72)

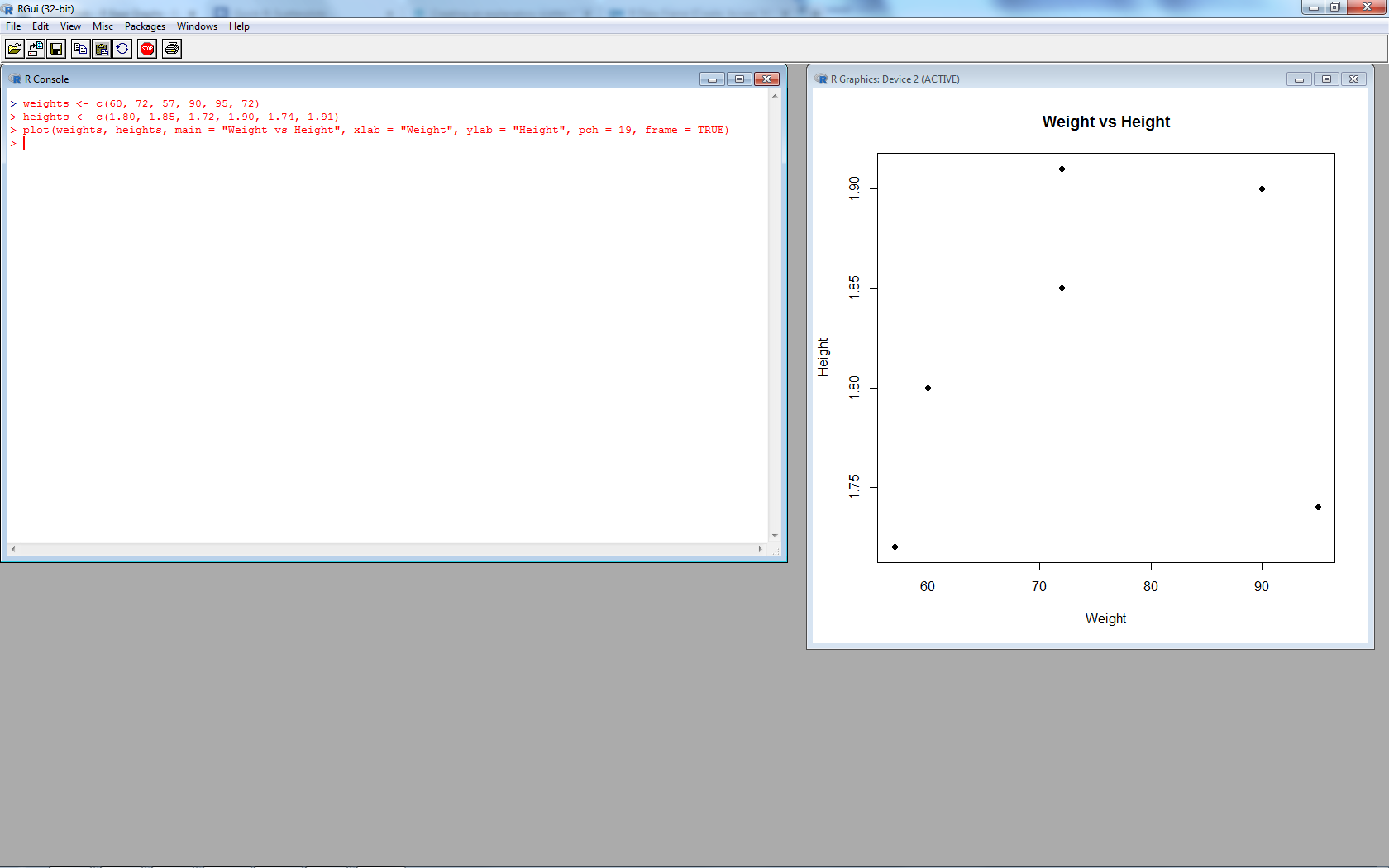
heigths <- c(1.80, 1.85, 1.72, 1.90, 1.74, 1.91)

**d.** Create a scatterplot of weight vs. height. Interpret the scatterplot.

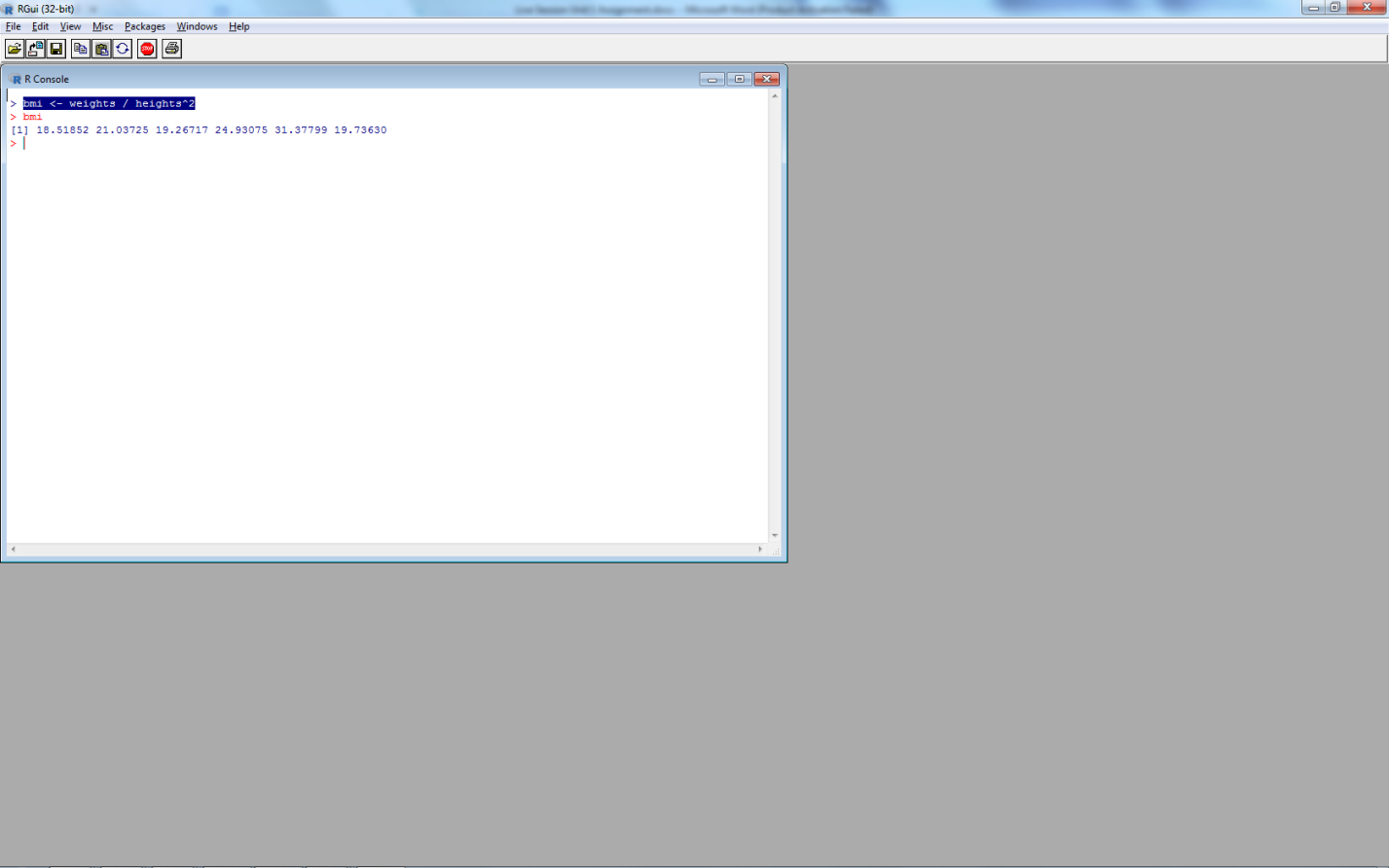
plot(weights, heights, main = "Weight vs Height",

xlab = "Weight", ylab = "Height", pch = 19, frame = TRUE)

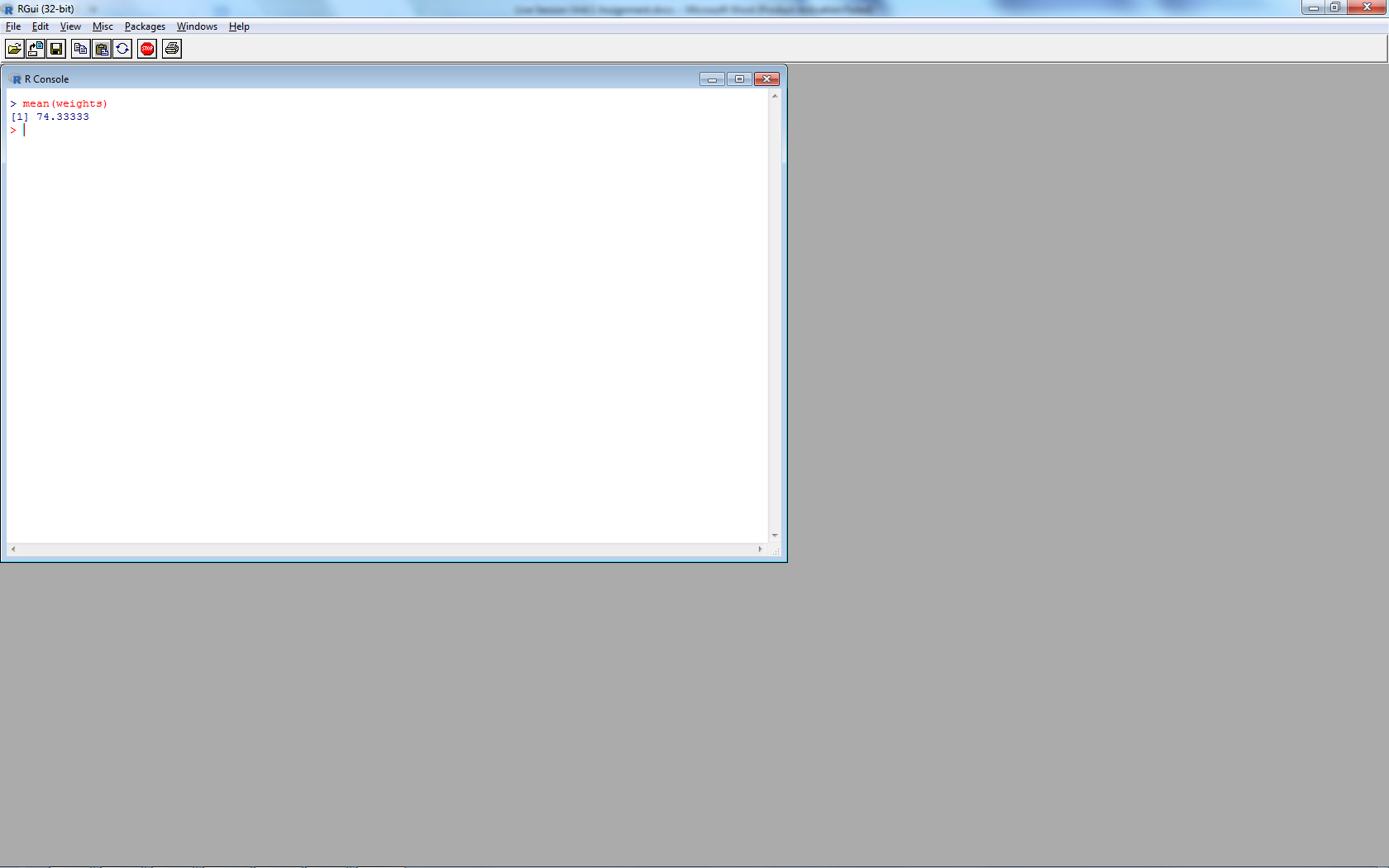
There seems to be a correlation between weight and height though subject 5 seems to be an outlier.



**e.** Calculate the BMI for each individual. (BMI = weight in kg divided by the square

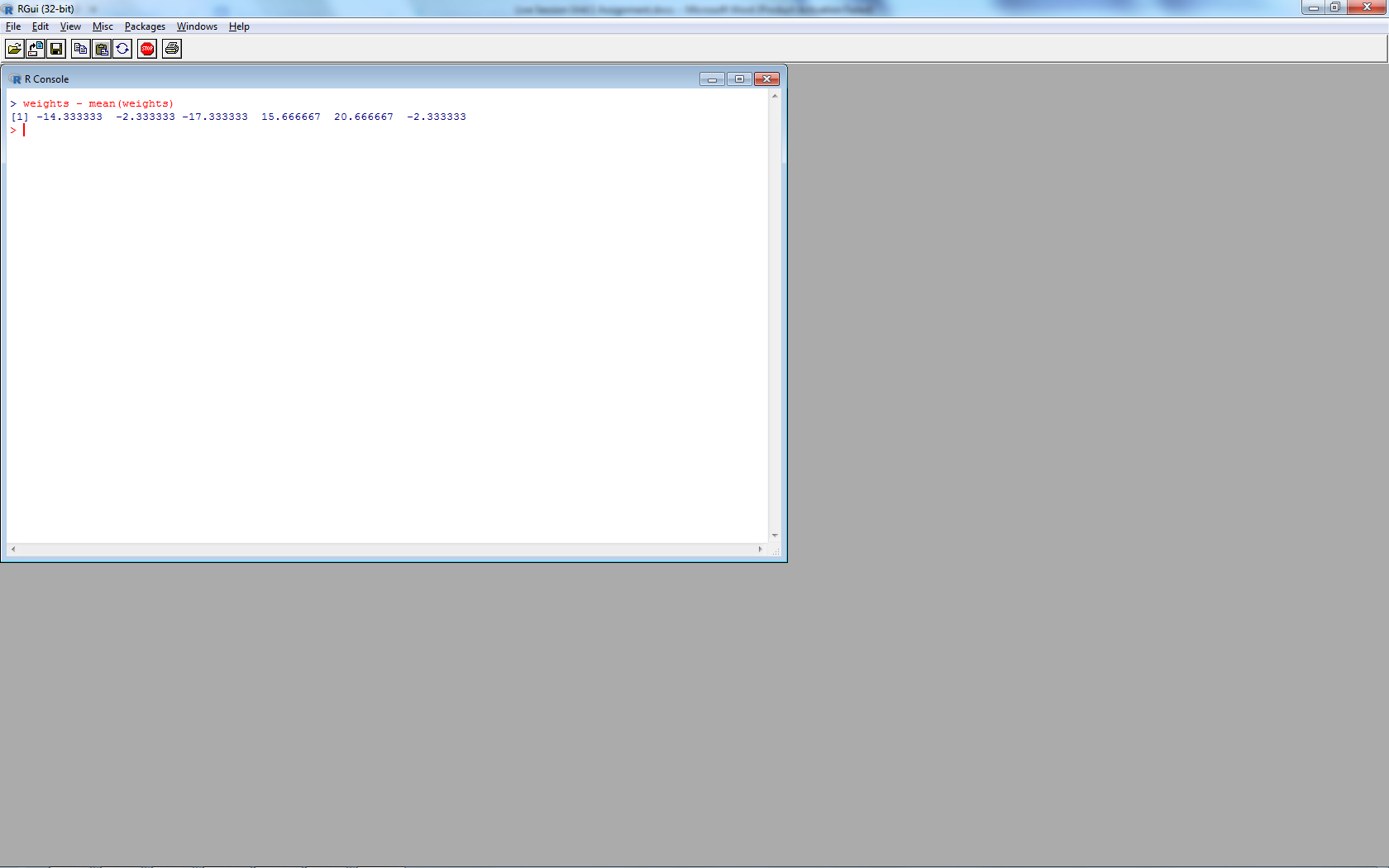
of the height in m)

bmi <- weights / heights^2



**f.** Calculate the mean for weight.

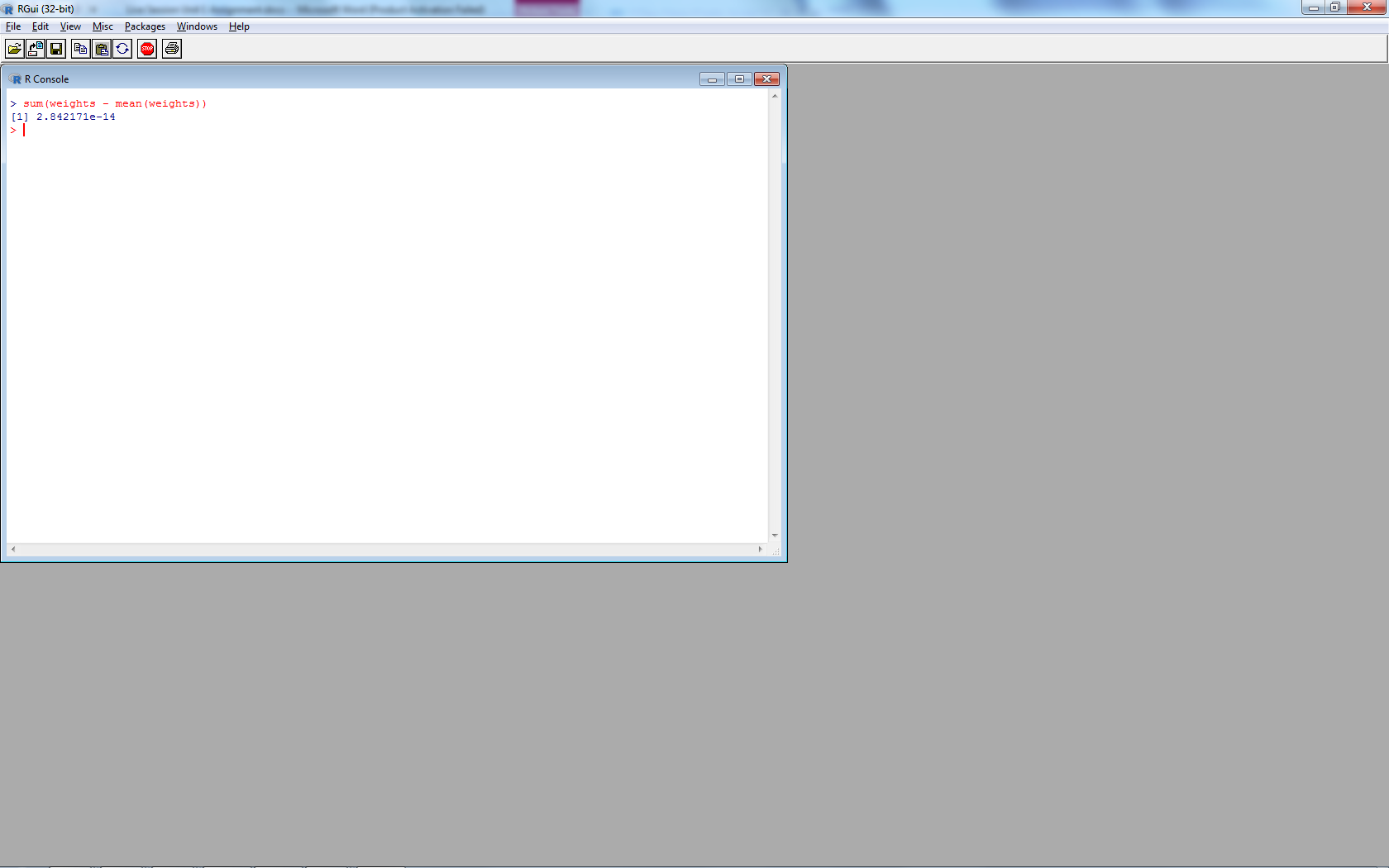
mean(weights)



**g.** Subtract the mean from each value of weight.

weights – mean(weights)

**h.** Sum the result. Now you know why we square the deviations from the mean to

calculate a standard deviation!

sum(weights - mean(weights))

**4.** (25 points) Your data science profile. Enter your data science profile into R as a data

frame with two columns. Call it by your first name. The categories are computer

programming, math, statistics, machine learning, domain expertise, communication and

presentation skills, and data visualization. Your ranking for each category 1-5, with 5 as

best. Create a bar graph of your data science profile. When you submit your work, please

submit your code, including the data entry piece.

> categories <- c("Programming", "Math", "Stats", "M.Learning", "Domain", "Presentation", "Data.Vis")

> scores <- c(1, 2, 2, 1, 2, 3, 2)

> samuel <- data.frame(categories, scores)

> p <- ggplot(samuel, aes(categories, scores))

> p +geom\_bar(stat = "identity")

