Questions 1 to 10: Use the PhysicalData.txt file (linked to the website) for these questions. The data consist of physical measurements for n=55 college students. Measurements were made by the students during a class activity. (Source: William Harkness, Pennsylvania State University.)

There are 11 columns of data:
Column Name Description
1. Height Self-reported height, inches
2. LeftArm Length of left forearm, cm
3. RtArm Length of right forearm, cm
4. LeftFoot Length of left foot, cm
5. RtFoot Length of right foot, cm
6. LeftHand Width of left palm, cm
7. RtHand Width of right palm, cm
8. HeadCirc Head circumference, cm
9. Nose Nose length, cm
10. Female 1 if Female, 0 if Male
11. Male 1 if Male, 0 if Female

The goal is to see whether Y = Head circumference can be predicted using the following X variables LeftFoot, RtFoot and Male, and for Question 2, Female.

1. Find the correlation between these sets of variables:
   a. Male and Female
   b. RtFoot and LeftFoot
   c. HeadCirc and RtFoot
   d. HeadCirc and LeftFoot
   e. HeadCirc and Male

   NOTE: You can find the correlation matrix for all of the variables and restrict them to 3 decimal places using
   > options(digits=3)
   > cor(PhysicalData)
   assuming you kept that name for the dataset; otherwise use the name you gave it.

   f. Using these correlations, which single variable – LeftFoot, RtFoot, or Male, would be the best single variable to use to predict head circumference? Explain your choice.

2. Try fitting the model with only the two predictors Male and Female. Look at the summary of the model in R. What kind of message does R give? Explain in your own words what the problem is with trying to use both Male and Female in the model.
For questions 3 to 8, fit the “Full” model by using Y = HeadCirc and the three X variables in this order: LeftFoot, RtFoot and Male. Call the model “Full”. (You do not need to go through the step of checking conditions for any of the requested tests.)

3. Test the coefficient corresponding to each of the three X variables. I.e., test the hypotheses $H_0: \beta_j = 0$ versus $H_a: \beta_j \neq 0$ for $j = 1, 2, 3$. Give a test statistic and $p$-value and make a conclusion for each test. Use $\alpha= 0.05$. Are any of them statistically significantly different from 0?

4. Get the Anova table for the full model using the command `> anova(Full)`. Use the Anova table to test the null hypothesis $H_0: \beta_1 = 0$ versus $H_a: \beta_1 \neq 0$, i.e., to test whether LeftFoot should be in the model. Use $\alpha= 0.05$.

5. Test the overall hypothesis $H_0: \beta_1 = \beta_2 = \beta_3 = 0$. Give a test statistic and $p$-value. What is your conclusion? Use $\alpha= 0.05$.

6. Compare the results of your test of $H_0: \beta_1 = 0$ in Questions 3 and 4. Did you reach the same conclusion? If so, what was it? If not, explain why not.

7. Compare the results in Questions 3 and 5. In both cases, you tested all 3 of the coefficients. Did you reach the same conclusion in both questions? If so, what is the conclusion? If not, explain why not.

8. Find the Variance Inflation Factor for each of the 3 X variables. Use the command `>vif(Full)` where “Full” is what you called the full model when you fit it. You will first need to use the command `>library(car)` and if you don’t have the package “car” you will need to install it first. Based on these VIF values, what do you suggest about what variable(s) to include in your model?

9. Fit two more models, call the first one Left and use the variables LeftFoot and Male. Call the other one Right and use the variables RtFoot and Male. Are any of the variables statistically significant predictors based on the tests of the individual coefficients? Use $\alpha= 0.05$.

10. Which of the 3 models (Full, Left, Right) would you recommend be used to predict Head circumference? Give a statistical reason for choosing that model.

11. Do Problem 3.8 in the book. For each part, first write the model using $\beta$ multiplied by variable names. (For example, $Active = \beta_0 + \beta_1 (Rest) + \beta_2 (Gender) + \beta_3 (Rest)(Gender) + \epsilon$.) Then write the code that would go in the `lm` command to fit that model. (You do not need to include the dataset name as part of the code.) If you use Xs instead of variable names, define what each X represents first.