Preamble:

Consider the snail data available in **Snail2.csv**. Take **Length** as the response variable and we would like to understand how it is related with other variables in the data.

Details about the data is available at [Morphometrics and phylogeography of the cave-obligate land snail Helicodiscus barri (Gastropoda, Stylommatophora, Helicodiscus)](.).

Problem statements:

a) Perform an exploratory analysis of data.

b) Is **Length** appropriate as a response variable or a transformation is necessary? In case a transformation of response is necessary, try the natural log transformation or some other simple transformation and use it for the rest of this problem.

c) For each predictor, fit a simple linear regression model to predict the response. Describe your results. In which of the models is there a statistically significant association between the predictor and the response? Create some plots to back up your assertions.

d) Fit a multiple regression model to predict the response using all of the predictors. Describe your results. For which predictors can we reject the null hypothesis $H_0 : \beta_j = 0$?

e) Build a reasonably "good" multiple regression model for these data. Be sure to explore interactions of ShellType with other predictors. Carefully justify all the choices you make in building the model and verify the model assumptions.

f) Write the final model in equation form, being careful to handle qualitative predictors and interactions (if any) properly.

g) Use the final model to predict the **Length** of a **Type1** snail with other predictors set equal to their sample means. Also provide a 95% prediction interval for the response and a 95% confidence interval for the mean response. Repeat for a **Type2** snail, and compare the answers.

h) Now split the data set into training (80%) and test set.

i) Use a bagging approach with B = 1000 to predict the Length in the given data. Compute the MSEs.

j) Repeat (i) with a random forest approach and with $m \approx \sqrt{p}$

k) Compare all your results and comment on your findings.