Download the data file "bpress.xls" from D2L. Import it into R. This data has four variables:

bp = blood pressure (diastolic measure)
weight = in kilograms
bsa = body surface area (sq meters)
dur = length of time person has been diagnosed with hypertension (in years)
pulse = beats per minute
stress = index of stress in individual's life
age = age of individuals (in years)

Use R to answer the following questions. You will need to turn in a printout of your work in R and answers for each of the questions listed above.

1. Estimate a regression with bp_i as the dependent variable and the other six variables as independent variables. Without doing formal t-tests, note which variables seem to be significantly different than zero. (Check to see if the t-stat is close to 2.)

2. Estimate correlations between the different variables:
   a. weight, bsa
   b. weight, dur
   c. weight, pulse
   d. weight, stress
   e. weight, age
   f. bsa, dur
   g. bsa, pulse
   h. bsa, stress
   i. bsa, age
   j. dur, pulse
   k. dur, stress
   l. dur, age
   m. pulse, stress
   n. pulse, age
   o. stress, age

   Are any of these independent variables highly correlated with one another? For those that have a correlation > 0.6, create scatterplots.

3. Calculate VIFs for each of the six independent variables.

4. Do you have any suggestions for how we should improve the model? Should we possibly drop a variable? If so, which one?

5. Re-estimate the model you suggested in (4). What happens to the coefficients of the remaining independent variables?