Some people argue that Americans "vote their pocketbooks" in presidential elections. If this is the case, then the outcome of a presidential election should reflect the state of the economy immediately prior to that election. Alternatively, some people say that voters think mainly about how well the incumbent president has carried out his/her responsibilities in office. Therefore, presidential election outcomes should be related to public approval of the president in the months immediately preceding the election. Information contained in a dataset called Elcoll can be used to test these two theories of presidential election outcomes.

The Elcoll dataset contains information about every American presidential election from 1948 through 2000 (hence, \( n = 14 \)). The dataset is stored in the file "Elcoll.dta" and it is available on the course web site. You will need to use the read.dta function from the foreign library to read the data into R. There are four variables:

- Year of the presidential election (Variable name: year).
- Percentage of Electoral College vote received by candidate from incumbent party (Variable name: vote).
- Percent change in Gross National Product during the preceding year (Variable name: gnp).
- Public approval of president’s job performance during July of election year (Variable name: approval).

The exercises below should be completed using R. You will need to use functions from the car and MASS packages in order to do so. Please generate, print, and explain (briefly) the graphs and regression output for each question.

1. Construct scatterplots of vote against gnp and against approval. Does the initial graphical evidence suggest that a linear model would be appropriate for these data? Go ahead and estimate a linear equation, regressing Electoral College vote on GNP change and presidential approval. Interpret the estimates in terms of the arguments described at the beginning of this assignment.

2. Use Added-variable plots and component-plus-residual plots to assess problems in the model. Also, use an appropriate plot of the residuals against the predicted values from the model in order to determine whether heteroskedastic errors might be present.

3. Use trial-and-error, based on Tukey and Mosteller’s "Bulging Rule" to determine the appropriate transformation(s) for the data. Re-fit the model, if necessary.

4. Try using Box-Cox and Box-Tidwell transformations to deal with any problems that may exist in the data (note, however, that the small sample size will make the significance tests a bit suspect). Do the results from these transformation strategies agree with your conclusions based upon the trial-and-error approach? Does one or the other of these transformation strategies seem to work more effectively with these data?