The attached who.csv dataset contains real-world data from 2008. The variables included follow.

Country: name of the country
LifeExp: average life expectancy for the country in years
InfantSurvival: proportion of those surviving to one year or more
Under5Survival: proportion of those surviving to five years or more
TBFree: proportion of the population without TB.
PropMD: proportion of the population who are MDs
PropRN: proportion of the population who are RNs
PersExp: mean personal expenditures on healthcare in US dollars at average exchange rate
GovtExp: mean government expenditures per capita on healthcare, US dollars at average exchange rate
TotExp: sum of personal and government expenditures.

1. Provide a *scatterplot* of LifeExp~TotExp, and run simple linear regression. Do not transform the variables. Provide and interpret the F statistics, R^2, standard error,and p-values only. Discuss whether the assumptions of simple linear regression met.

2. Raise life expectancy to the 4.6 power (i.e., LifeExp^4.6). Raise total expenditures to the 0.06 power (nearly a log transform, TotExp^.06). Plot LifeExp^4.6 as a function of TotExp^.06, and r re-run the simple regression model using the transformed variables. Provide and interpret the F statistics, R^2, standard error, and p-values. Which model is "better?"

3. Using the results from 3, forecast life expectancy when TotExp^.06 =1.5. Then forecast life expectancy when TotExp^.06=2.5.

4. Build the following multiple regression model and interpret the F Statistics, R^2, standard error, and p-values. How good is the model?

LifeExp = b0+b1 x PropMd + b2 x TotExp +b3 x PropMD x TotExp

5. Forecast LifeExp when PropMD=.03 and TotExp = 14. Does this forecast seem realistic? Why or why not?