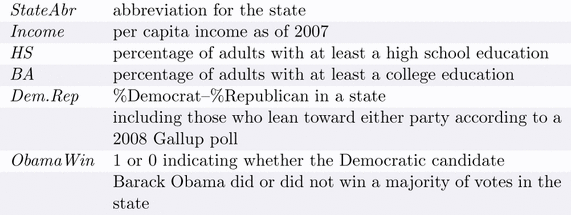
Can we use state-level variables to predict whether a state votes for the Democratic versus the Republican presidential nominee? The file **Election08** contains data from 50 states plus the District of Columbia. The variables recorded are:



See the attached R output to help answer the questions below.

1. Consider the separate logistic regression models that predict *ObamaWin* using each of the predictors *Income, HS, BA*, and *Dem.Rep* (model1, model2, model3, and model4). Which of these variables does the most effective job of predicting this response? Which is the least effective? Explain the criteria you use to make these decisions.
2. What additional information do you need to determine which of the four variables is the most appropriate predictor of *ObamaWin*?
3. Using model1, use the estimated slope from the logistic regression to compute an estimated odds ratio and write a sentence that interprets this value in the context of this problem.
4. Interpret the 95% confidence interval for the odds ratio in #3.
5. The units of the *Income* variable are dollars, with values ranging from $28,845 (Mississippi) to $61,092 (District of Columbia). The odds ratio and interval in #3 and #4 are awkward to interpret since they deal with the change in the odds when state income changes by $1, a very trivial amount! To get an odds ratio that may be more meaningful, we created a new variable (call it *IncomeTh*) using *Income*/1000 to express the state per capita incomes in $1000s. The logistic regression using *IncomeTh* as the predictor of *ObamaWin* is model5. How does the fitted prediction equation change?
6. Using model5, use the estimated slope from the logistic regression to compute an estimated odds ratio and write a sentence that interprets this value in the context of this problem.
7. Using model5, calculate the probability of Obama winning a state whose per capita income is $35000.
8. The logistic model with *ObamaWin* as the response and *Dem.Rep, HS, BA*, and *Income* as the predictors is model6. Which predictor has the strongest relationship with the response in this model? Which predictors (if any) are not significantly related to *ObamaWin* in that model?
9. The insignificance of certain variables in model6 seems to conflict with the significance of the variables in models 1,2,3,4 in #1. Explain this apparent contradiction.
10. Perform a nested drop-in-deviance test to compare model6 to model4.
    1. Write down the hypotheses for this test.
    2. Write down the test statistic, and how you would find the p-value (you may write the R code needed to find the p-value; or explain in words, but you must be explicit!).
    3. The p-value from part (b) is 0.000573. Make a conclusion in context.
11. You want to find the best 1-, 2-, 3-, or 4-variable model to predict *ObamaWin*. If you were working with this data, what would your next step be?
12. We talked about the relationship between chi-squared tests and logistic regression. Could a chi-square test be performed to tell us something about the relationship between *ObamaWin* and any of these 4 variables? If your answer is No, explain why not. If your answer is Yes, explain what hypotheses the chi-square test would test.