

## In Class Exercise 3

This question should be answered using the **Weekly** data set, which is part of the **ISLR** package. This data is similar in nature to the **Smarket** data from the lab, except that it contains 1,089 weekly returns for 21 years, from the beginning of 1990 to the end of 2010.

- (a) Produce some numerical and graphical summaries of the **Weekly** data. Do there appear to be any patterns?
- (b) Use the full data set to perform a logistic regression with **Direction** as the response and the five lag variables plus **Volume** as predictors. Use the summary function to print the results. Do any of the predictors appear to be statistically significant? If so, which ones?
- (c) Compute the confusion matrix and overall fraction of correct predictions. Explain what the confusion matrix is telling you about the types of mistakes made by logistic regression.
- (d) Now fit the logistic regression model using a training data period from 1990 to 2008, with **Lag2** as the only predictor. Compute the confusion matrix and the overall fraction of correct predictions for the held out data (that is, the data from 2009 and 2010).