Instructions: Your instructor will supply a CSV file. Assign the imported data set to a data frame called “hiredata.” This data set contains n=295 survey responses from raters who participated in a hiring process. The dependent variable, “hired,” is a binary variable with 0 for a candidate who was subsequently not hired and 1 for a candidate who was. The “recommend” variable is each participant’s recommendation of whether to hire, with 1 = “Definitely Hire,” 2 = “Possibly Hire,” and 3 = “Do Not Hire.” In addition, there are six belief questions, all on 1 to 4 scales (with 1 as most favorable and 4 as least favorable) with assessments on issues like leadership and collaboration. The research question is to understand the connection between survey responses and the hiring decisions. Can survey responses predict who will be hired? The exercise has three phases: 1) create an initial logistic regression model using the *recommend* variable as a predictor; 2) run a Bayesian model of logistic regression; and 3) find additional predictor(s) that may improve the model.

**Phase 1**: In this first phase, read the data into R, inspect the data, and develop a basic logistic regression model with one predictor – the hire recommendation.

1. Run summary(hiredata). Take special note of the min and max of the survey variables: recommend, vision, issues, trends, consult, lead, and collab. Run histograms on each of the numeric variables. Comment on any anomalies and what you will do about them.
2. Create a correlation matrix of the data. The cor() procedure will not work on text or factor data, so you need to select numerical values: *cor(hiredata[,c(2,4:10)])*. Add comments about the values in the correlation matrix.
3. Run and interpret a basic logistic regression model using glm:  
   *glmOut <-   
   glm(formula = hired ~ recommend, family = binomial(link="logit"), data = hiredata)*  
   Run summary() on the glmOut object and write a brief statement summarizing the results. Is the predictor statistically significant?
4. Use exp() and confint() commands as described on page 225 of the textbook to convert log odds for the coefficient on the predictor into regular odds.
5. The plain odds version of the coefficient on the predictor is fractional. This can make interpretation of the results more difficult, particularly for non-statisticians to whom you may wish to communicate your results. Invert the recommend variable:  
   *hiredata$recInv <- (4 - hiredata$recommend)*  
   Note that this command adds a new variable to your existing data set. To cross check your results, correlate the new variable with the old one.
6. Rerun the code for Questions 4 and 5 using *recInv*. Interpet the plain odds ratio, based on the output you get from applying exp() and confint() to the coefficient.
7. Produce and interpret a Nagelkerke pseudo-R-squared using this code:  
   *install.packages("BaylorEdPsych")*

*library(BaylorEdPsych)*

*PseudoR2(glmOut)*

**Phase 2**: Conduct a Bayesian logistic regression analysis, using the MCMCpack package.

1. Run the MCMClogit() function using the same model as for Question 7 above. You will need to install MCMCpack and library() it. The following code should work:  
   *bayesLogitOut <- MCMClogit(formula = hired ~ recInv, data = hiredata)*  
   Run summary() and comment on how the Bayesian (MCMC) mean of the coefficient parameter on the predictor compares with the corresponding result from the conventional glm() analysis.
2. Create a plot of the MCMC output by running plot(bayesLogitOut). Take note of any anomalies in the trace plot. Does the distribution of parameter estimates on the predictor overlap with zero?
3. We can improve our view of the parameter estimates of the coefficient by converting the distribution from log odds to plain odds. The following code develops a histogram of the posterior distribution of plain odds:  
   *recLogOdds <- as.matrix(bayesLogitOut[,"recInv"])  
   recOdds <- apply(recLogOdds,1,exp)   
   hist(recOdds, main=NULL)*
4. Write a block comment with an interpretation of the Bayesian output. Obtain and report specific values for the mean of the posterior distribution of plain odds as well as the upper and lower bounds of its HDI.
5. **Answer the original research questions: Can survey respondents accurately assess who will be hired? What variables predict hiring decisions?**