**IST687 Linear Modeling Homework – Making Predictions**

The chapter on linear models (“Lining Up Our Models”) introduces linear predictive modeling using the tool known as multiple regression. The term “multiple regression” has an odd history, dating back to an early scientific observation of a phenomenon called “regression to the mean.” These days, multiple regression is just an interesting name for using a simple linear modeling technique to measuring the connection between one or more predictor variables and an outcome variable.

**Part A: Load and condition the data**

1. The data is available on blackboard (hotelSurveySherison), as a JSON file.  
    *Hint: Don’t forget to use setwd() to make sure that R is looking in the right folder for your text file.*
2. Use the str command to make sure you can see the following attributes

overallCustSat: Overall customer satisfaction (based on survey response)

hotelSize: The size of the hotel (number of rooms in the hotel)

gender: gender of the customer

guestAge: age of the guest (in years)

lengthOfStay: nights at the hotel

whenBookedTrip: number of days before the trip the room was reserved

checkInSat: The customer;s survey response to their checkin experience

hotelState: The state (location) of the hotel – in the united states

hotelClean: The customer’s survey response on the cleanliness of the hotel

hotelFriendly: The customer’s survey response on the friendliness of the staff

freeText: free form customer comments

**Part B: Explore the data**

1. Create bivariate plots for each of the attributes.

Your code should produce nine separate plots. Make sure the Y-axis and X-axis are labeled. Keeping in mind that the overall customer satisfaction is the outcome (or dependent) variable, which axis should it go on in your plots?  
*Hint: use the jitter command, so you can see all the surveys (something such as)  
 jitter(hotelSurvey$checkInSat)*

1. What do you observe from the plots? Note via a block comment.

**Part B: Generate a linear model**

1. Next, create one regression model predicting the overall customer satisfaction from the other variables (but not the freeText response). Refer to page 202 in the text for syntax and explanations of lm( ). Make sure to include all predictors in one model – NOT different models each with one predictor.
2. Report the R-Squared in a comment. Which of the predictors are statistically significant in the model? In a comment, report the coefficients (AKA slopes or B-weights) for each predictor that is statistically significant.
3. Write a block comment that explains in a narrative your overall interpretation of the model. Make sure to refer to each variable (one dependent and three independent) by a descriptive name (i.e., not X1, X2, etc.).

**Part C: Generate a different linear model**

1. Next, create a different regression model predicting the overall customer satisfaction from the one variable you think is best.
2. Write a block comment comparing the two lm models.