Exercises for chapter 6

The AirCnC product manager for the “Mansions & Manors” category has just run an ad to encourage customers to “upgrade” and consider an M&M property for their next booking.

Exercise 1 (AirCnC M&M) - open. The AirCnC M&M dataset is missing some data for income. Can you diagnose the missingness of income and run the appropriate regression for the effect of income on the probability of booking an M&M property? *(if that’s too open-ended for you, the next exercise guides you through intermediary steps)*
Note: you’ll have to create the variable bkd_mm = bkd * mm if you haven’t already done the exercises for chapter 2.

Exercise 2 (AirCnC M&M) - guided.

1) Diagnosing the missingness
   a) How many records are missing for income?
   b) What is the correlation between the missingness of income and booking an M&M property?
   c) Run a logistic regression of bkd_mm on income. How much does the coefficient change if you replace the missing values for income by the minimum income? By the maximum income?
   d) Plot a scatterplot of the average income in a zip code vs. the percentage of missing values in that zip code. What do you conclude about the missingness of income? (is it likely to be MCAR? MAR? MNAR?)
2) Correcting for the missingness.
   a) Run a logistic regression of bkd_mm on income, imputing the missing values with MICE. What is the coefficient for income?