R Exercises Part 2: Imputation Methods

1. Calculate a correlation matrix using pair-wise deletion.
2. Perform an unconditional mean imputation
3. Perform a conditional mean imputation on a continuous variable with a single fully observed categorical variable as the conditioning variable.
4. Perform a regression imputation on a single variable.
5. Perform a stochastic regression imputation on a single variable.
6. Perform a set of regression imputations on a data set with a monotone missing pattern.
7. Perform a stochastic conditional means imputation.
8. Perform an unconditional hot deck imputation.
9. Perform a conditional hot deck imputation.

Hints:

For 2, use apply(x,2,mean,rm.na=TRUE) to get the means, then the results of the last set of R puzzles to set the values.

For 3, 7 and 9, you may find the R function split() which splits a data set based on the value of another variable useful.

For 4, 5, 6 and 7 you might find the predict() function (used on the results of lm()) useful.

For 5 and 7, the "arm" package has a useful function called sim() which will simulate from the slope and intercepts (see Gelman and Hill).

For 8 and 9, make a set of appropriate complete cases row indexes and then sample() from them.