Measures of Fit: Multiple R^2

- Several predictors in a multiple regression
- We can't just look at a simple R^2; there's an R^2 for each independent variable.
- Multiple R^2 is used instead.
- **Multiple R^2:** the square of the correlation between the observed Y values and the Y values predicted by the multiple regression model
- Result of calculation: one R^2 value that can be used for entire multiple regression model
- We interpret multiple R^2 just as we do simple R^2: the amount of variation in the outcome variable that is accounted for by the model.

The Problem With R^2 and Multiple R^2

- Multiple R^2 goes up as more variables are incorporated into the model, even if these variables don't add to the overall model.
- Therefore, a model with more predictor variables will appear to fit better, even though it may be less parsimonious than a simpler model with fewer strong predictors.
    - Example: Three strong predictors with R^2 = 0.8 (i.e., these predictors explain 80% of variation in Y).
    - Adding 15 more variables might push R^2 to 0.86.
    - Decision: Use model with three variables and R^2 of 0.8? Or model with 18 variables and R^2 of 0.86?
- Key principle: parsimony
- Determine what the strongest predictors are in order to understand linear relationship between predictors and dependent variables.