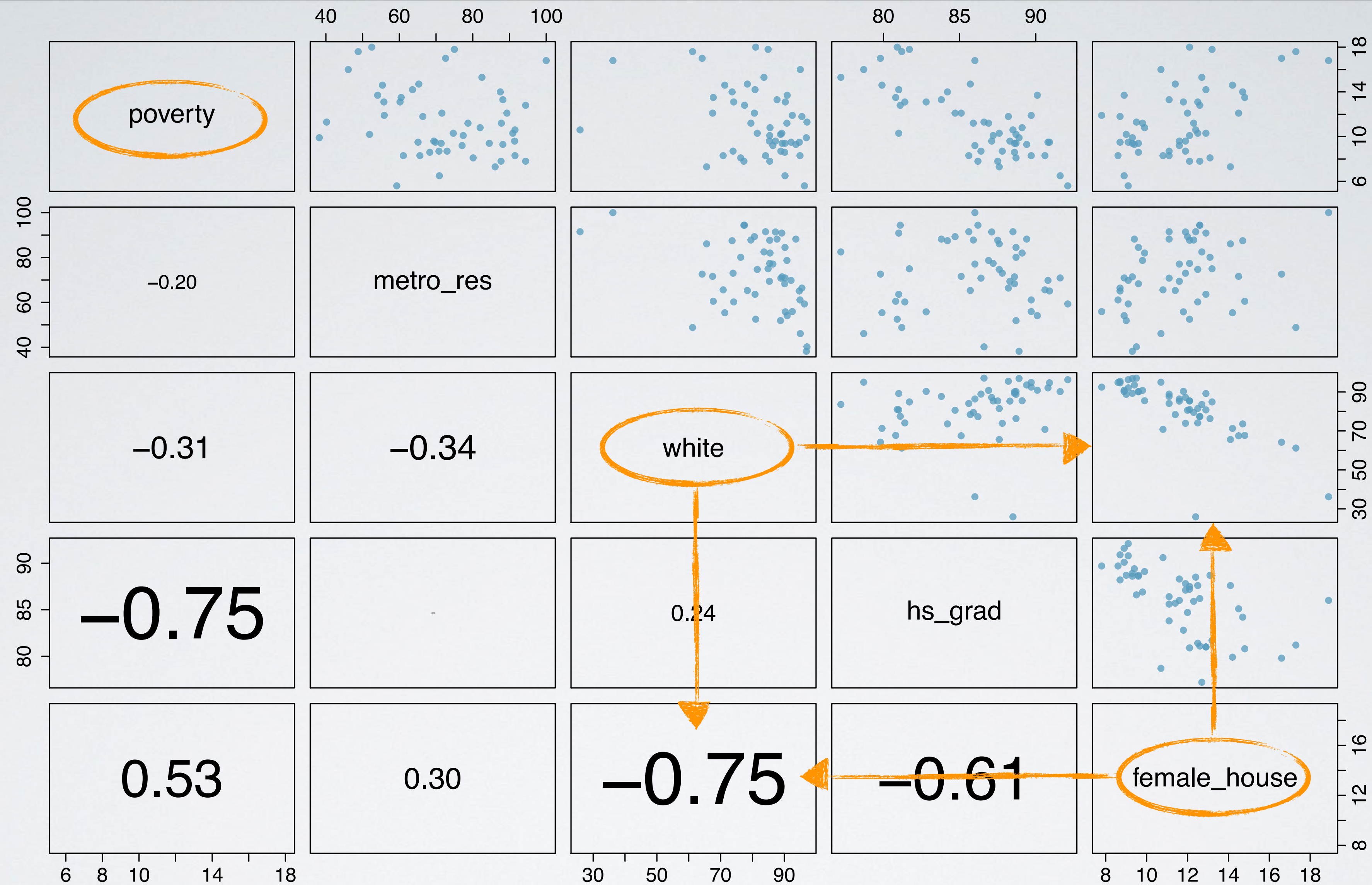


collinearity and parsimony

collinearity

- ▶ Two predictor variables are said to be **collinear** when they are correlated with each other.
- ▶ Remember: Predictors are also called independent variables, so they should be independent of each other.
- ▶ Inclusion of collinear predictors (also called **multicollinearity**) complicates model estimation.



parsimony

- ▶ Avoid adding predictors associated with each other because often times the addition of such variable brings nothing new to the table
- ▶ Prefer the simplest best model, i.e. the [parsimonious model](#)
 - ▶ Occam's razor: Among competing hypotheses, the one with the fewest assumptions should be selected
- ▶ Addition of collinear variables can result in biased estimates of the regression parameters
- ▶ While it's impossible to avoid collinearity from arising in observational data, experiments are usually designed to control for correlated predictors