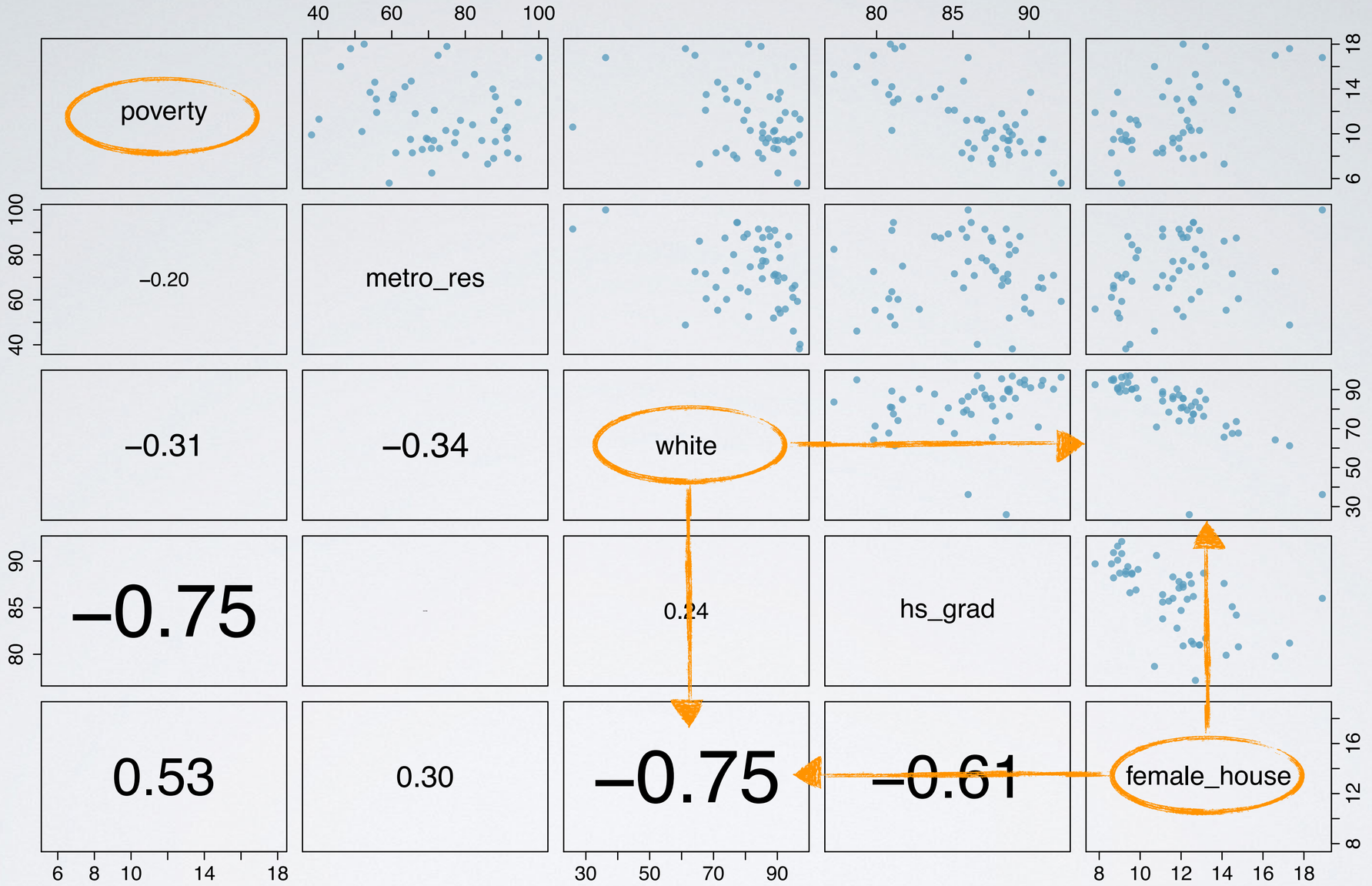
collinearity and parsimony

Dr. Mine Çetinkaya-Rundel
Duke University

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.


poverty
metro_res
white
hs_grad
female_house
−0.20
−0.31
−0.34
−0.75
0.24
0.53
0.30
−0.75
−0.61
40 60 80 100
80 85 90
6 10 14 18
40 60 80 100
30 50 70 90
80 85 90
6 8 10 14 18
8 10 12 14 16 18

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