

### Exercise Session 3

The data file *collegetown* contains observations on 500 single-family houses sold in Baton Rouge, Louisiana, during 2009–2013. The data include sale price (in thousands of dollars), *PRICE*, and total interior area of the house in hundreds of square feet, *SQFT*.

- a. Plot house price against house size in a scatter diagram
- b. Estimate the linear regression model  $PRICE = \beta_1 + \beta_2 SQFT + e$ . Interpret the estimates. Draw a sketch of the fitted line.
- c. Estimate the quadratic regression model  $PRICE = \alpha_1 + \alpha_2 SQFT^2 + e$ . Compute the marginal effect of an additional 100 square feet of living area in a home with 2000 square feet of living space.
- d. For the regressions in (b) and (c), compute the least squares residuals and plot them against *SQFT*. Do any of our assumptions appear violated?
- e. One basis for choosing between these two specifications is how well the data are fit by the model. Compare the sum of squared residuals (*SSR*) from the models in (b) and (c). Which model has a lower *SSR*? How does having a lower *SSR* indicate a “better-fitting” model?