2.17 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.

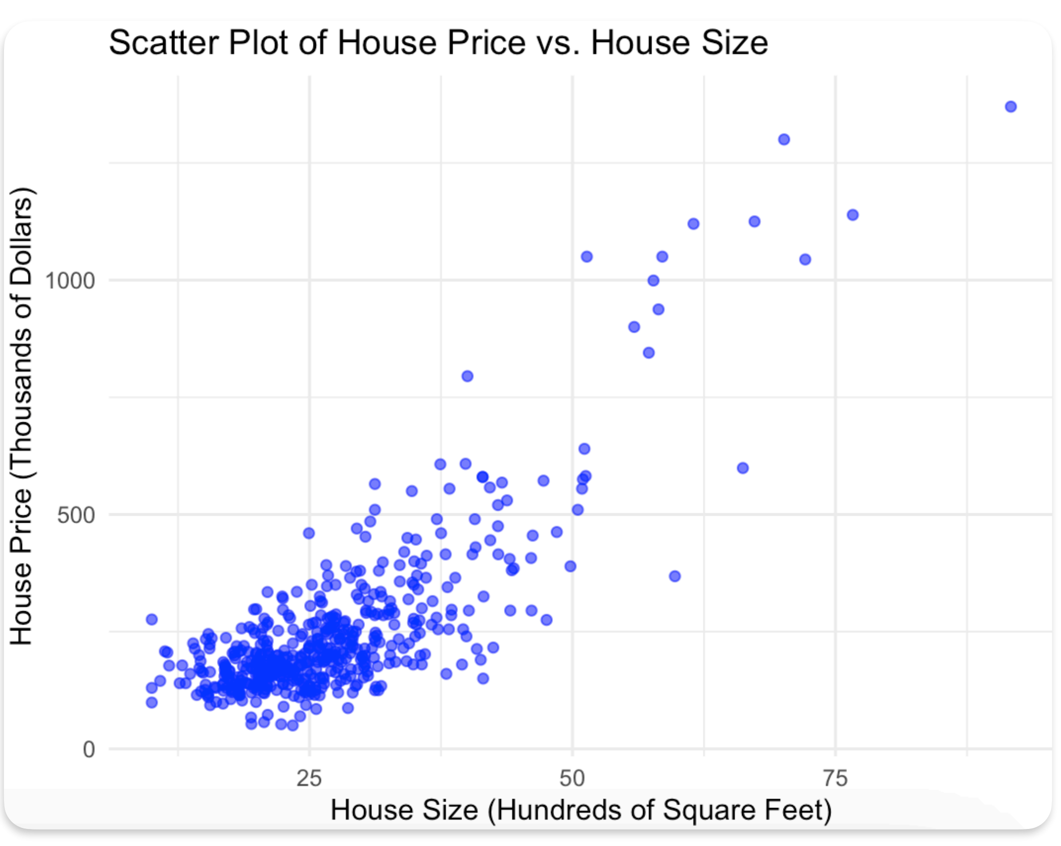
d. Graph the fitted curve for the model in part (c). On the graph, sketch the line that is tangent to the curve for a 2000-square-foot house.

e. For the model in part (c), compute the elasticity of *PRICE* with respect to *SQFT* for a home with 2000 square feet of living space.

f. For the regressions in (b) and (c), compute the least squares residuals and plot them against *SQFT*. Do any of our assumptions appear violated?

g. One basis for choosing between these two specifications is how well the data are fit by the model. Compare the sum of squared residuals (*SSE*) from the models in (b) and (c). Which model has a lower *SSE*? How does having a lower *SSE* indicate a "better-fitting" model?

a.



b. $\widehat{price} = -115.4236 + 13.4029\ sqft$

Intercept ($b_1 = -115.4236$):

a house with zero square feet has an expected price of \$ −115,423.6

Slope ($b_2 = 13.4029$):

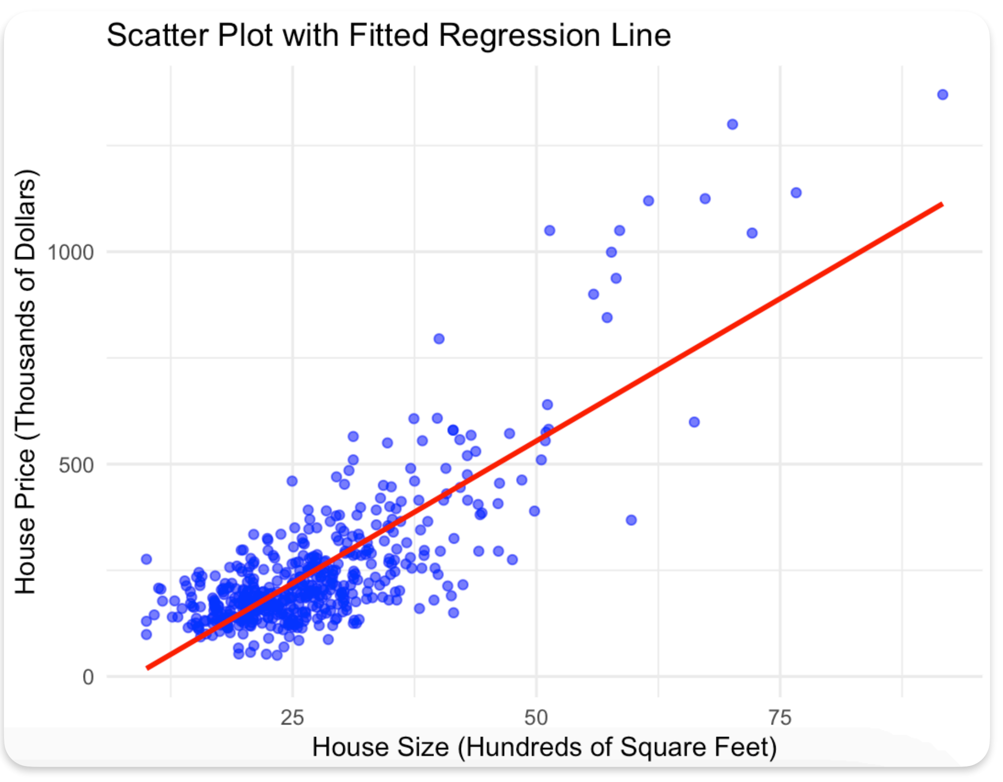
for every 100 square feet increase in house size, the expected house price increases by \$ 13,402.9

```
Call:
lm(formula = price ~ sqft, data = collegetown)

Residuals:
    Min      1Q  Median      3Q     Max
-316.93  -58.90   -3.81   47.94  477.05

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -115.4236    13.0882  -8.819   <2e-16 ***
sqft          13.4029     0.4492  29.840   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 102.8 on 498 degrees of freedom
Multiple R-squared:  0.6413,    Adjusted R-squared:  0.6406
F-statistic: 890.4 on 1 and 498 DF,  p-value: < 2.2e-16
```



c. $\widehat{price} = 93.565854 + 0.184519\ sqft^2$

Marginal Effect (ME): $\frac{d\,(price)}{d\,(sqft)} = 2 \times 0.184519\ sqft$

when sqft = 20 : $ME = 2 \times 0.184519 \times 20 = 7.38076$

Interpretation:

When the house size is 2000 square feet, increasing the size by 100 square feet is expected to increase the house price by approximately \$ 738.076

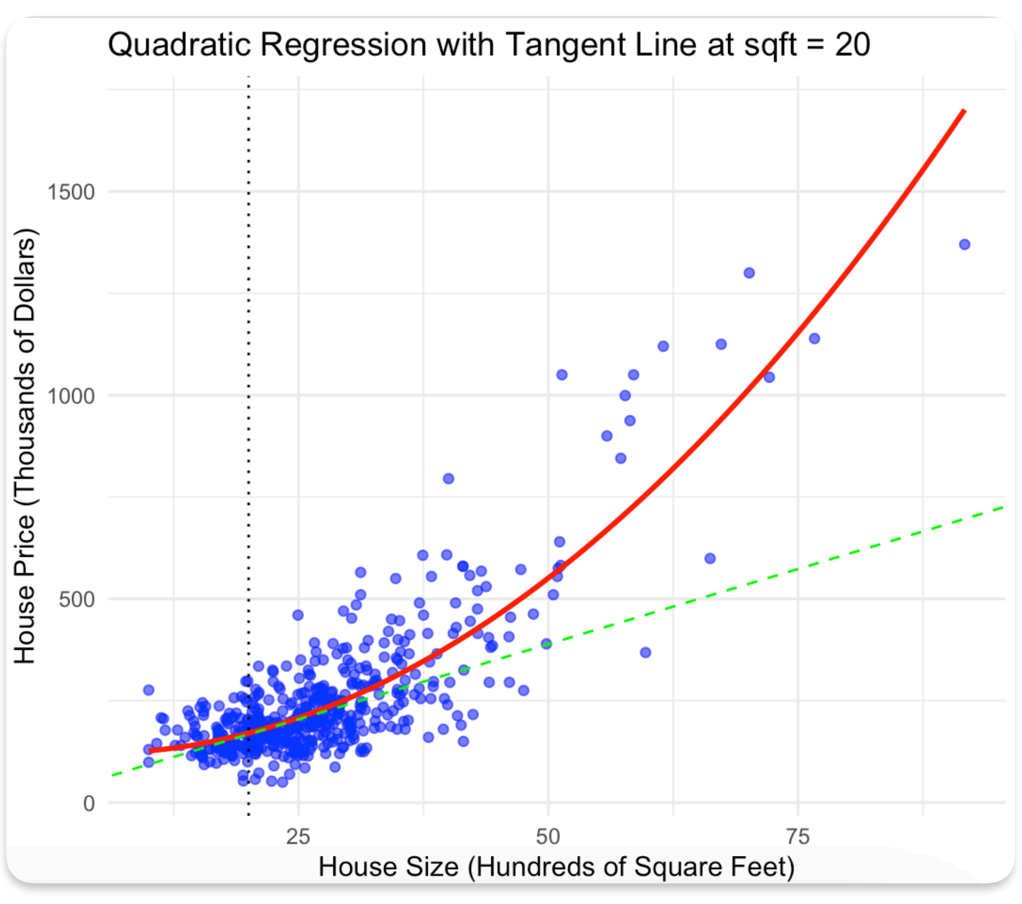
```
Call:
lm(formula = price ~ I(sqft^2), data = collegetown)

Residuals:
    Min      1Q  Median      3Q     Max
-383.67  -48.39   -7.50   38.75  469.70

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 93.565854   6.072226   15.41   <2e-16 ***
I(sqft^2)    0.184519   0.005256   35.11   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 92.08 on 498 degrees of freedom
Multiple R-squared:  0.7122,    Adjusted R-squared:  0.7117
F-statistic:  1233 on 1 and 498 DF,  p-value: < 2.2e-16
```

d.

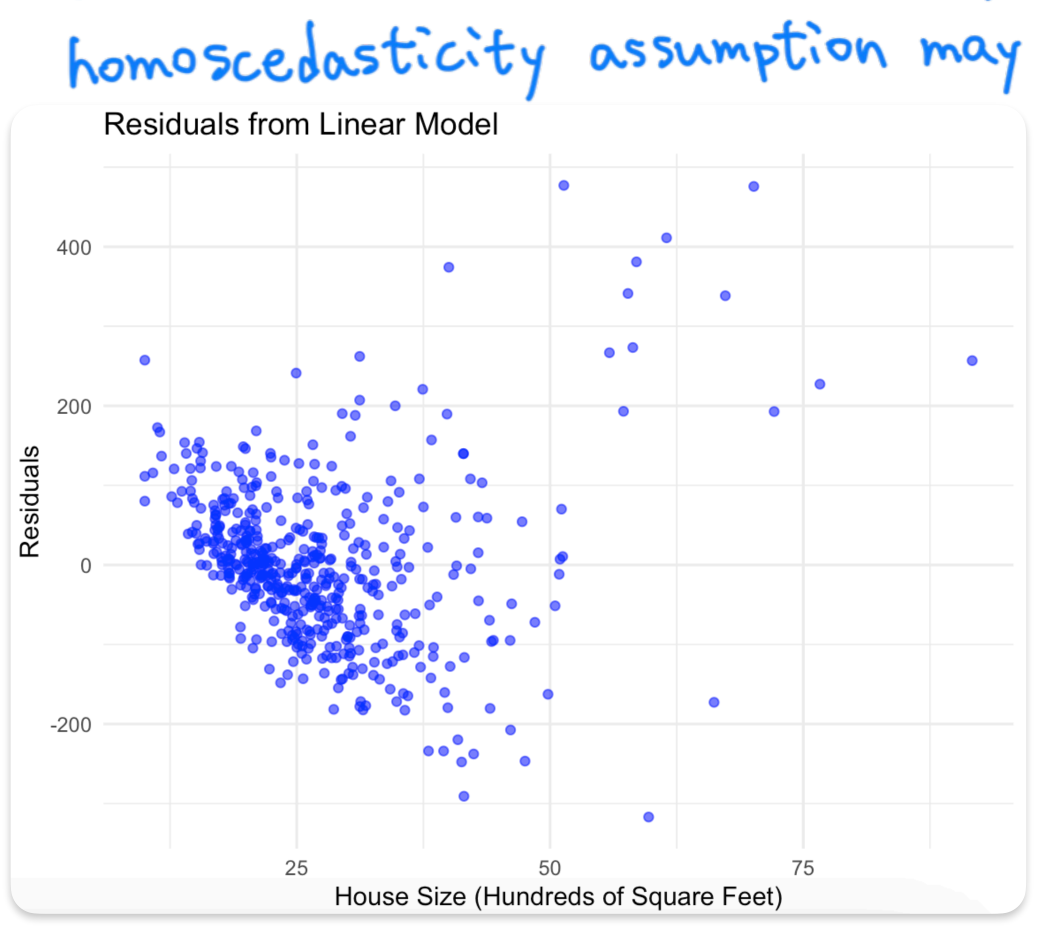


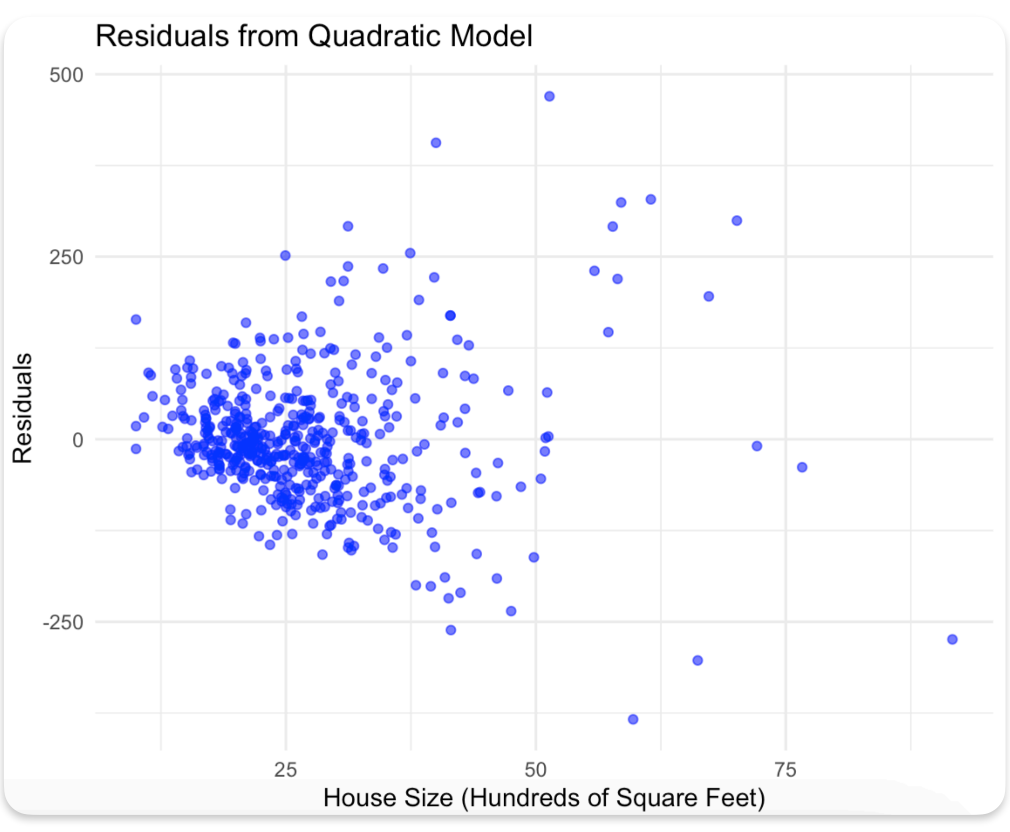
e. $E = \frac{d\,(price)}{d\,(sqft)} \times \frac{sqft}{price} = 2 \times 0.184519 \times sqft \times \frac{sqft}{\widehat{price}}$

$sqft = 20$, $\widehat{price} = 93.565854 + 0.184519\ sqft^2$

$E = 0.881951$

f. Both Residuals from Linear Model and Residuals from Quadratic Model are non-random, suggesting that the homoscedasticity assumption may be violated.





g. SSE (linear) = 5,262,847

SSE (quadratic) = 4,222,356

∴ SSE (quadratic) < SSE (linear)

The quadratic model has a lower SSE.

SSE表示模型對data的擬合程度，SSE越小，表示模型的預測值和實際值越接近，誤差值越小。