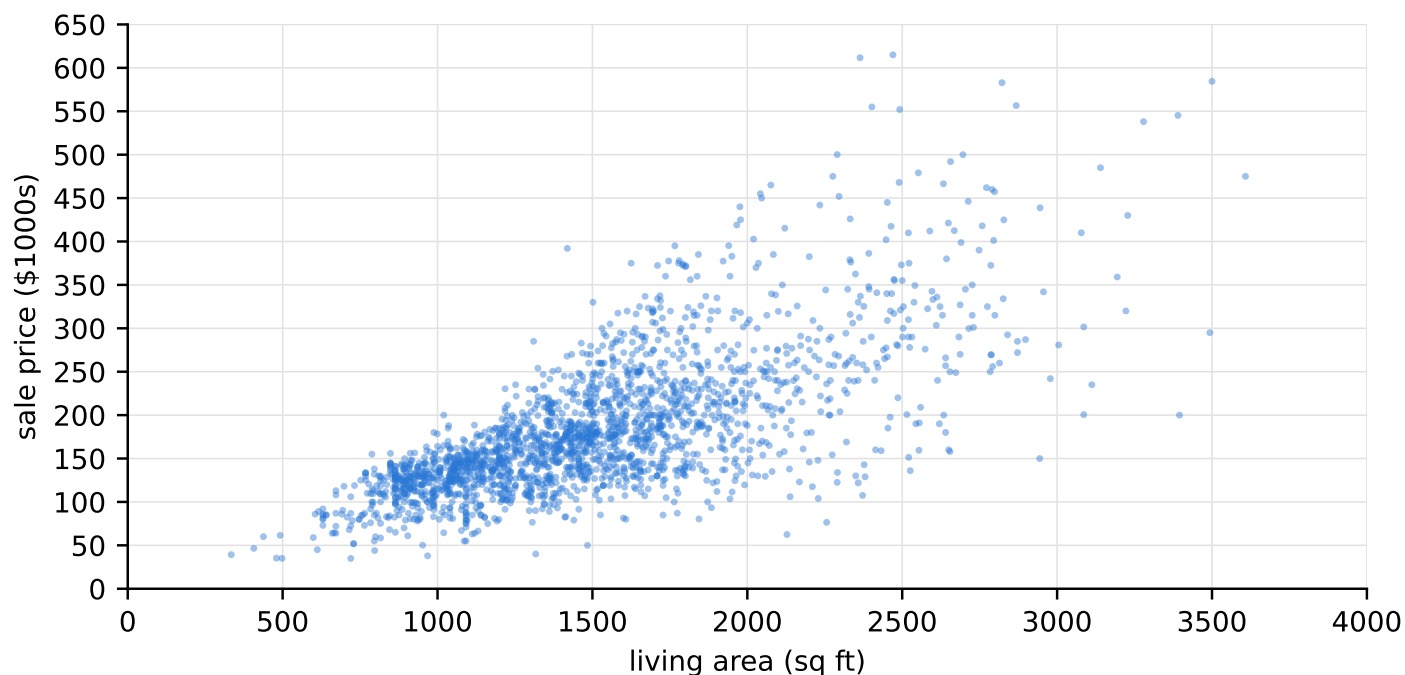


## DATA 202 — Linear Regression · Week 10 Monday

Same houses as today's slides: 2,925 home sales in Ames, Iowa, 2006–2010. Your professor will tell you when to do each part.

### 1 · Draw the line



1. With a ruler, draw the straight line that best follows the 2,193 training houses.
2. Your line's price for a 2,000 sq ft house: \$ \_\_\_\_\_
3. Your line's slope: about \$ \_\_\_\_\_ per square foot.

### 2 · Five test houses

The computer's line:  $\text{price} = \$9,194 + \$114.97 \times \text{living area}$ . Prices in **thousands of dollars**.

| house              | living area | actual | predicted | residual (actual – predicted) | residual <sup>2</sup> |
|--------------------|-------------|--------|-----------|-------------------------------|-----------------------|
| Somerset           | 1,861       | 257    | 223       | _____                         | _____                 |
| Brookside          | 2,068       | 210    | 247       | _____                         | _____                 |
| Northridge Heights | 1,358       | 224    | 165       | _____                         | _____                 |
| North Ames         | 1,142       | 114    | 140       | _____                         | _____                 |
| Northridge         | 2,202       | 263    | 262       | _____                         | _____                 |

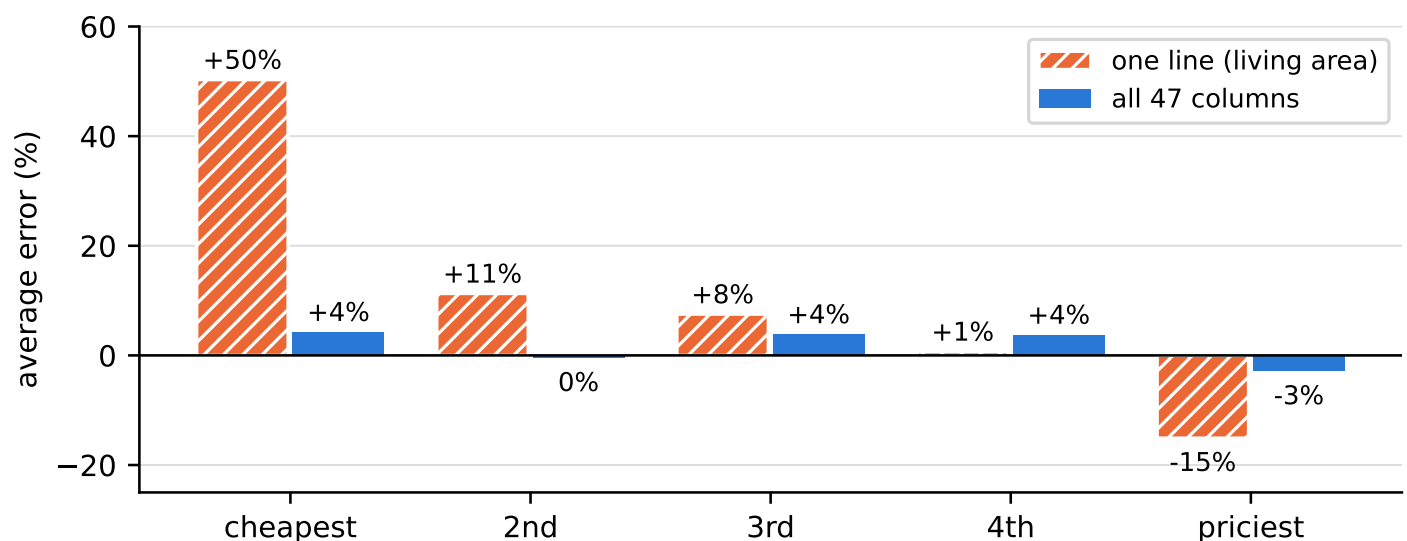
1. **MAE** = mean of the |residuals| = \_\_\_\_\_ thousand dollars
2. **RMSE** =  $\sqrt{\text{mean of the residuals}^2}$  = \_\_\_\_\_ thousand dollars
3. Why is the RMSE bigger than the MAE? \_\_\_\_\_

### 3 · Same size, one more bedroom

| model                          | coefficient of bedrooms |
|--------------------------------|-------------------------|
| price ~ bedrooms               | +\$14,077 per bedroom   |
| price ~ living area + bedrooms | -\$31,152 per bedroom   |

How can the same column add \$14,077 in one model and take away \$31,152 in the other?

### 4 · Over or under?



Average percentage error for the test houses, from the cheapest fifth to the priciest fifth. Above 0 = the model values the house **higher** than it sold for.

1. **Circle** the group the one-line model overvalues most, and the group it undervalues.
2. If these values set property taxes, who would pay too much? \_\_\_\_\_