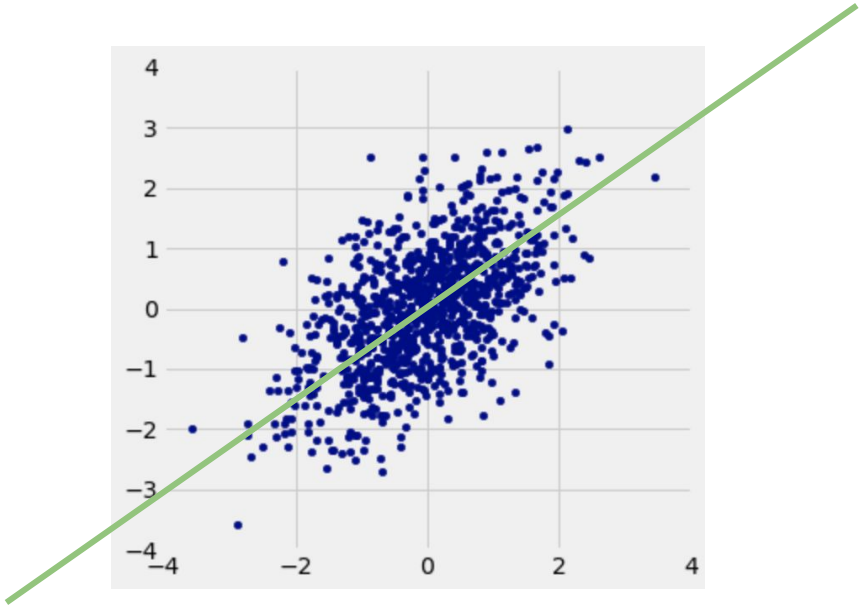
Least Squares Regression

The Regression Line



The Regression Line:

r = Correlation Coefficient

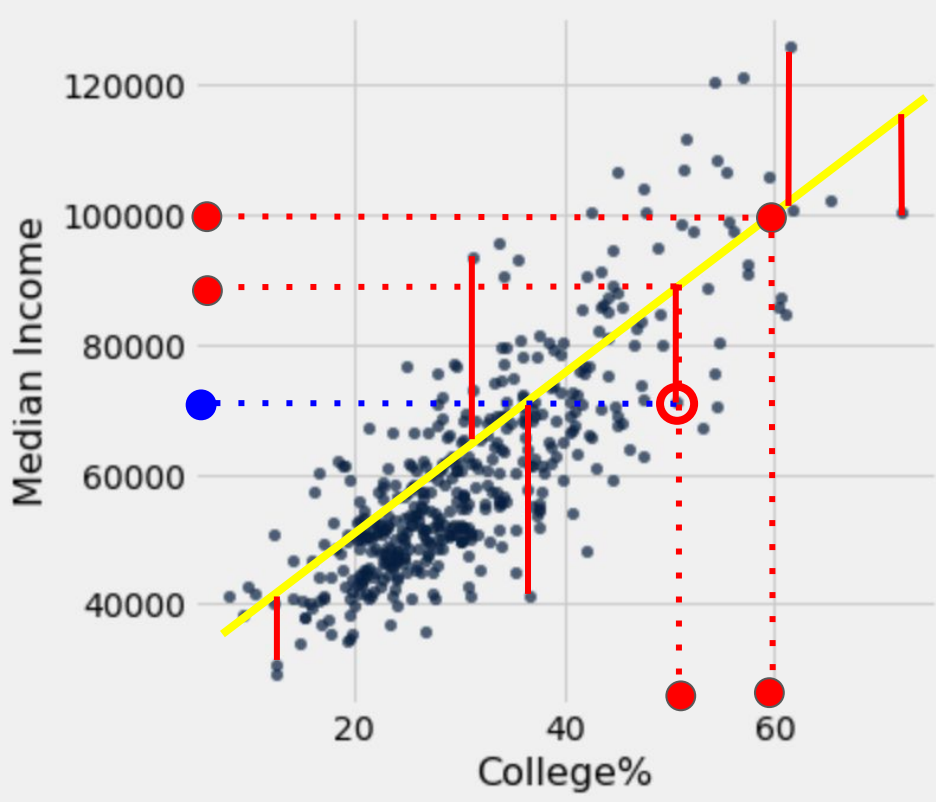
= average of products of x, y values (in standard units)

$y = r\,x$ (standard units)

Is the regression line the ‘best’ line to make predictions?

Do our formulas for slope and intercept work for any scatter plot?

What's the 'best' line?



The best line is the one with the *least total error.*

Mean Squared Error (MSE) =
Sum of squares of all errors=
sum of all (length of red line)^2

Root Mean Squared Error =
Square root of MSE

The best line is the one with the *least RMSE.*

The Regression Line = The Least Squares Line

The regression line is the unique straight line that minimizes the mean squared error of estimation among all straight lines.

$$\textbf{Slope of regression line} = r \cdot \frac{\text{SD of } y}{\text{SD of } x}$$

$$\textbf{Intercept of regression line} = \text{Average of } y - \text{slope} \cdot \text{average of } x$$