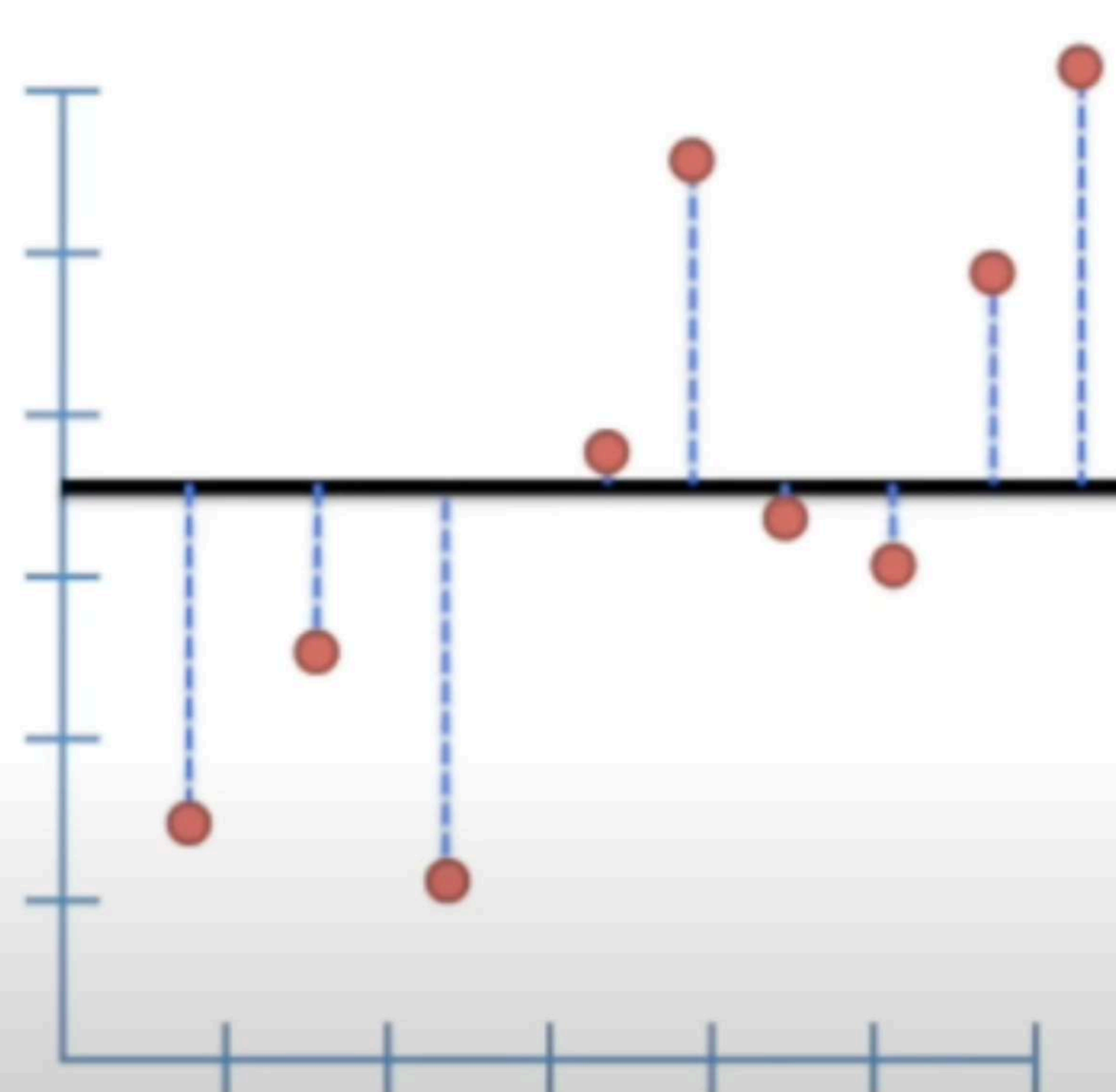


DATA ANALYTICS- UNIT 2

By
DEEPIKA KAMBOJ

Linear Regression

$$(b - y_1)^2 + (b - y_2)^2 + (b - y_3)^2 + (b - y_4)^2 + (b - y_5)^2 + (b - y_6)^2 + (b - y_7)^2 + (b - y_8)^2 + (b - y_9)^2$$



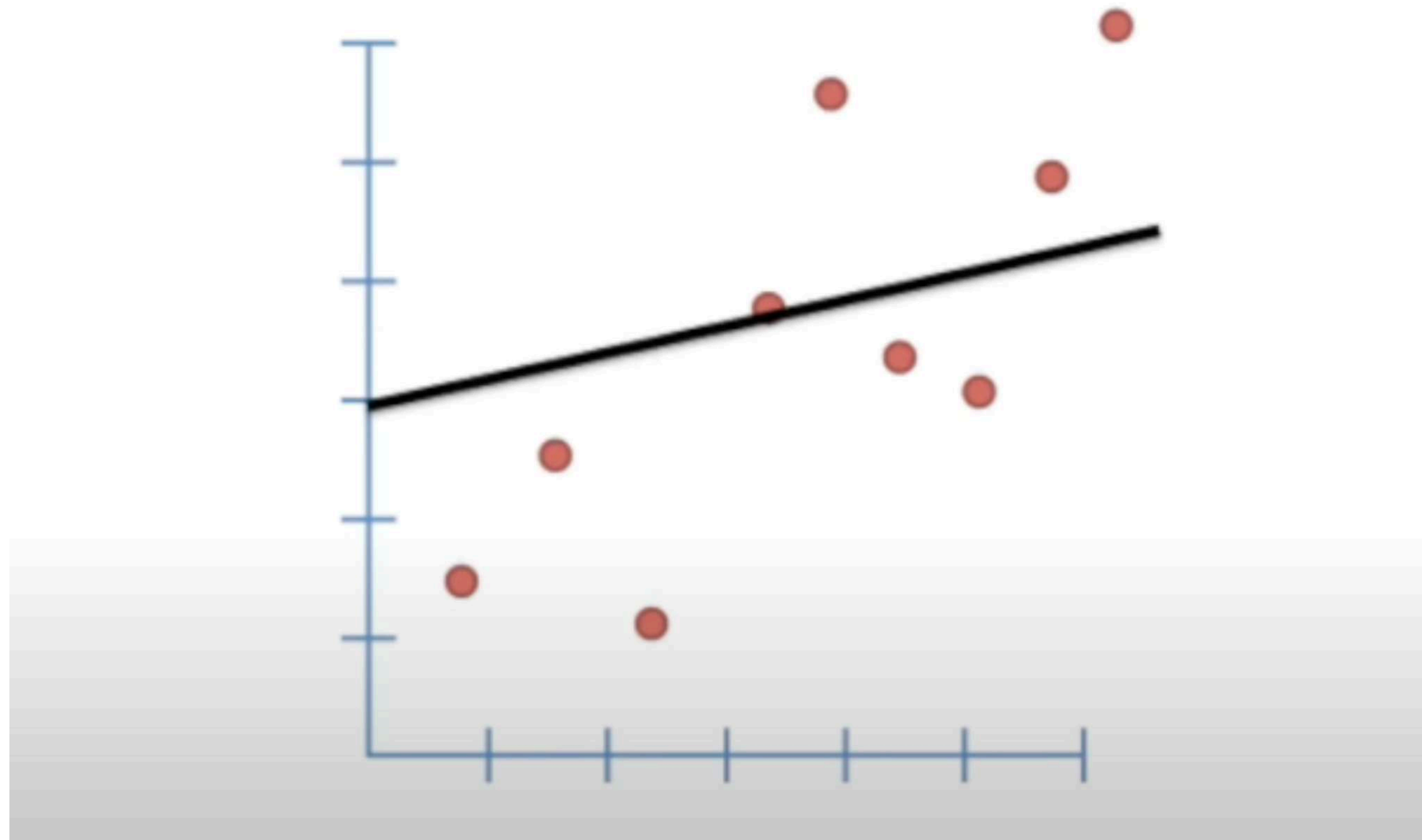
= 24.62

This is our measure of how well this line fits the data.

It's called the "sum of squared residuals," because the residuals are the differences between the real data and the line, and we are summing the square of these values.

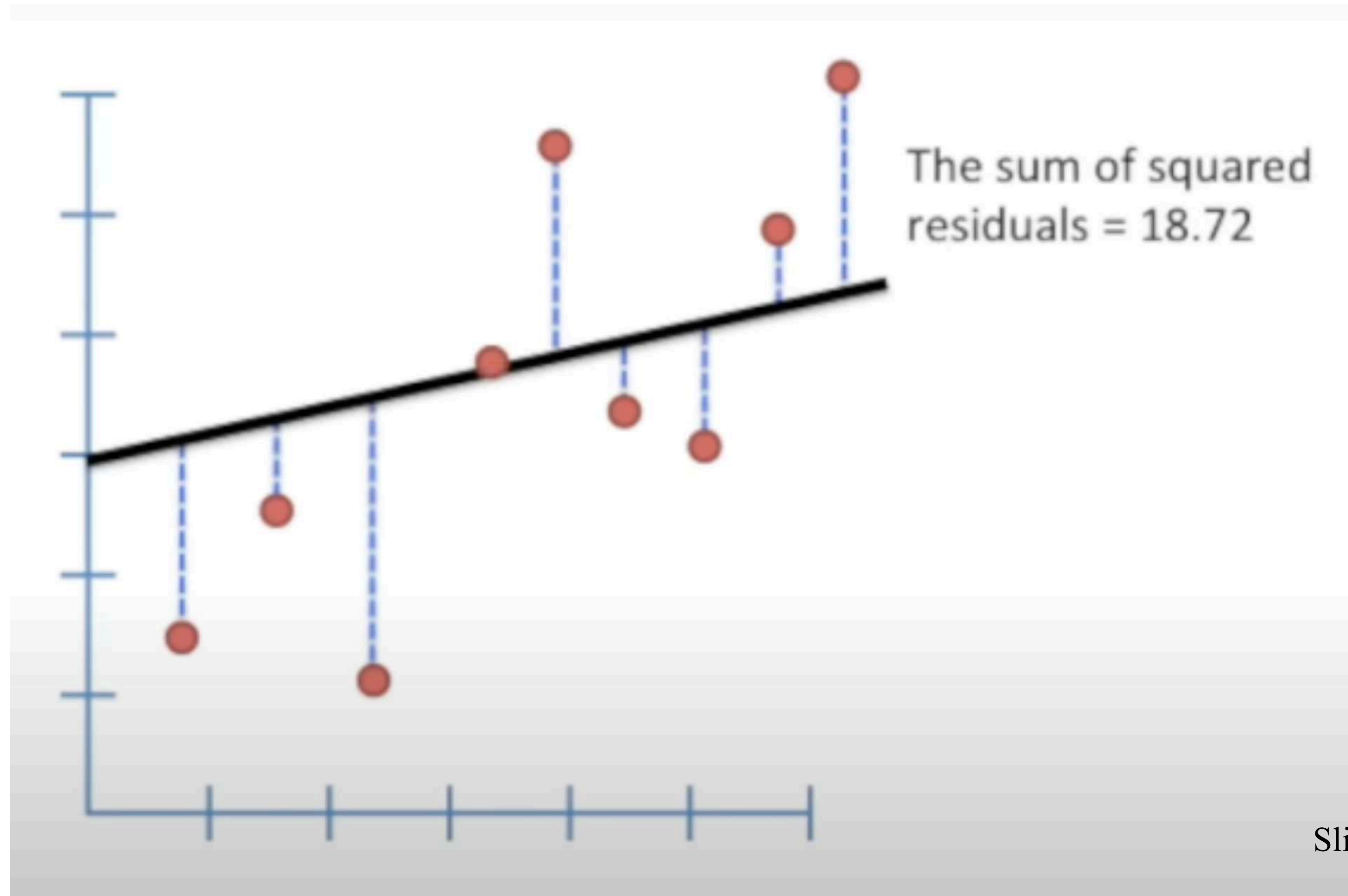
Linear Regression

Now let's see how good the fit is if we rotate the line a little bit.



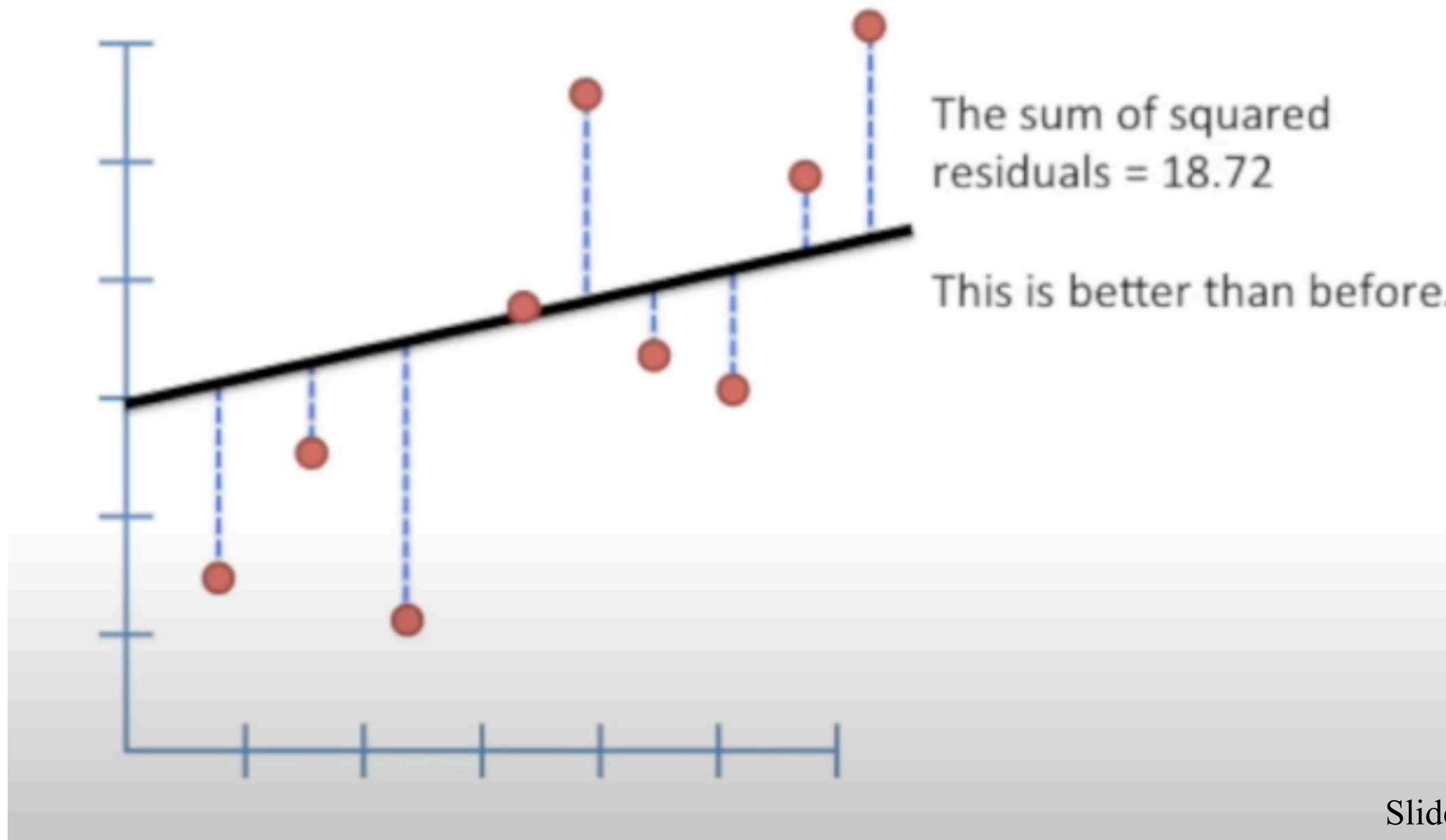
Slide Courtesy: StatQuest

Linear Regression



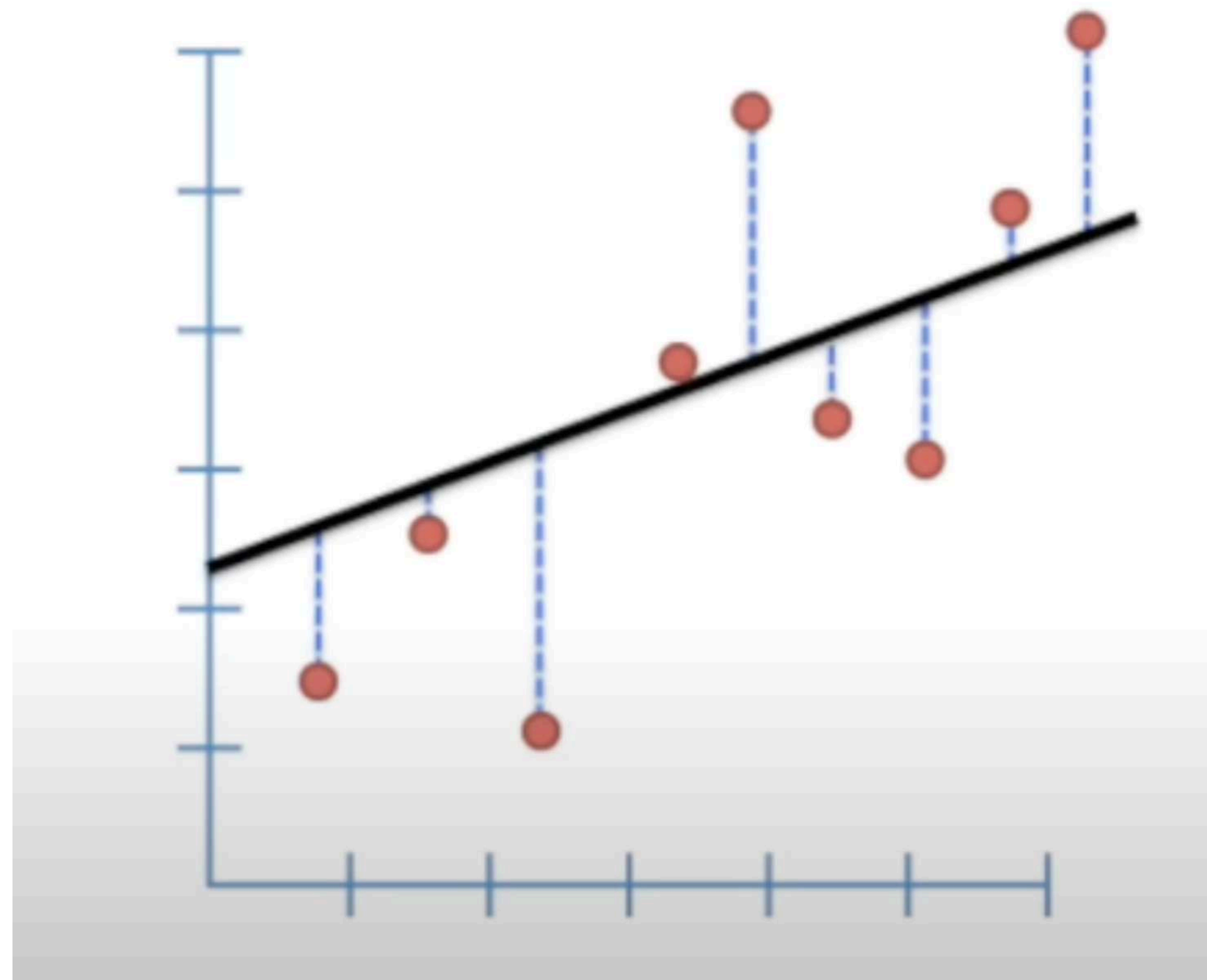
Slide Courtesy: StatQuest

Linear Regression



Linear Regression

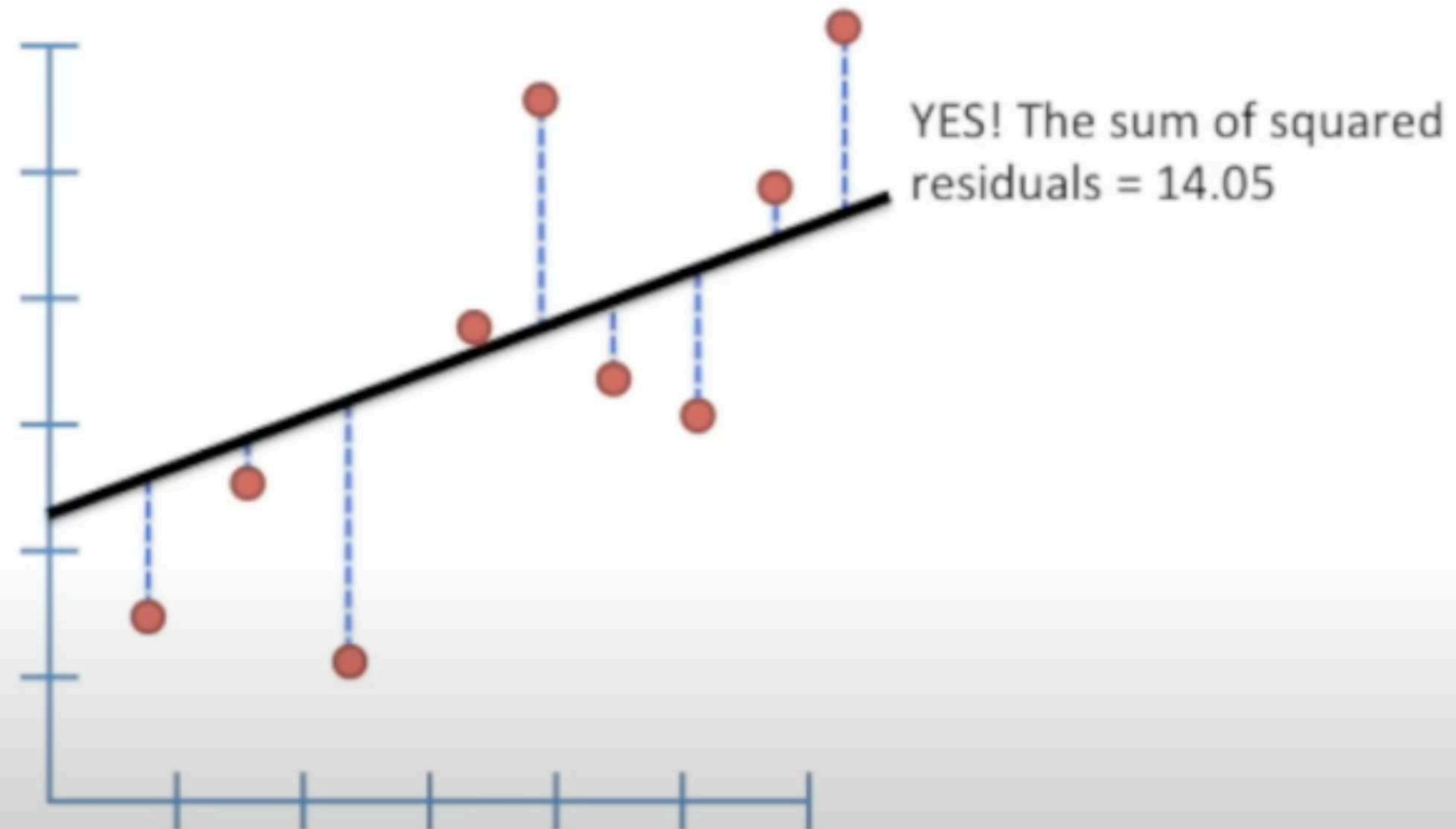
Does this fit improve if we rotate a little more?



Slide Courtesy: StatQuest

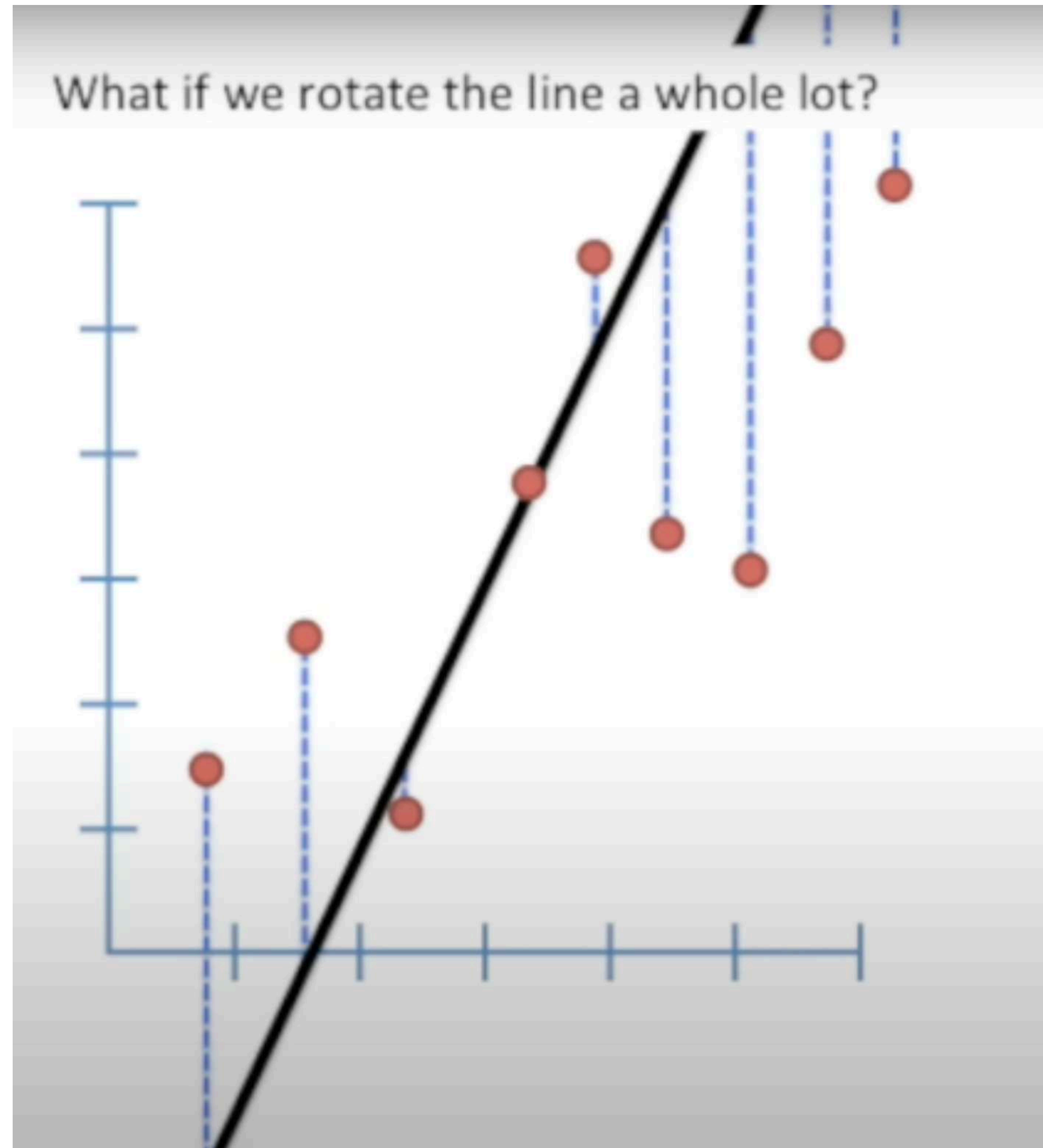
Linear Regression

Does this fit improve if we rotate a little more?



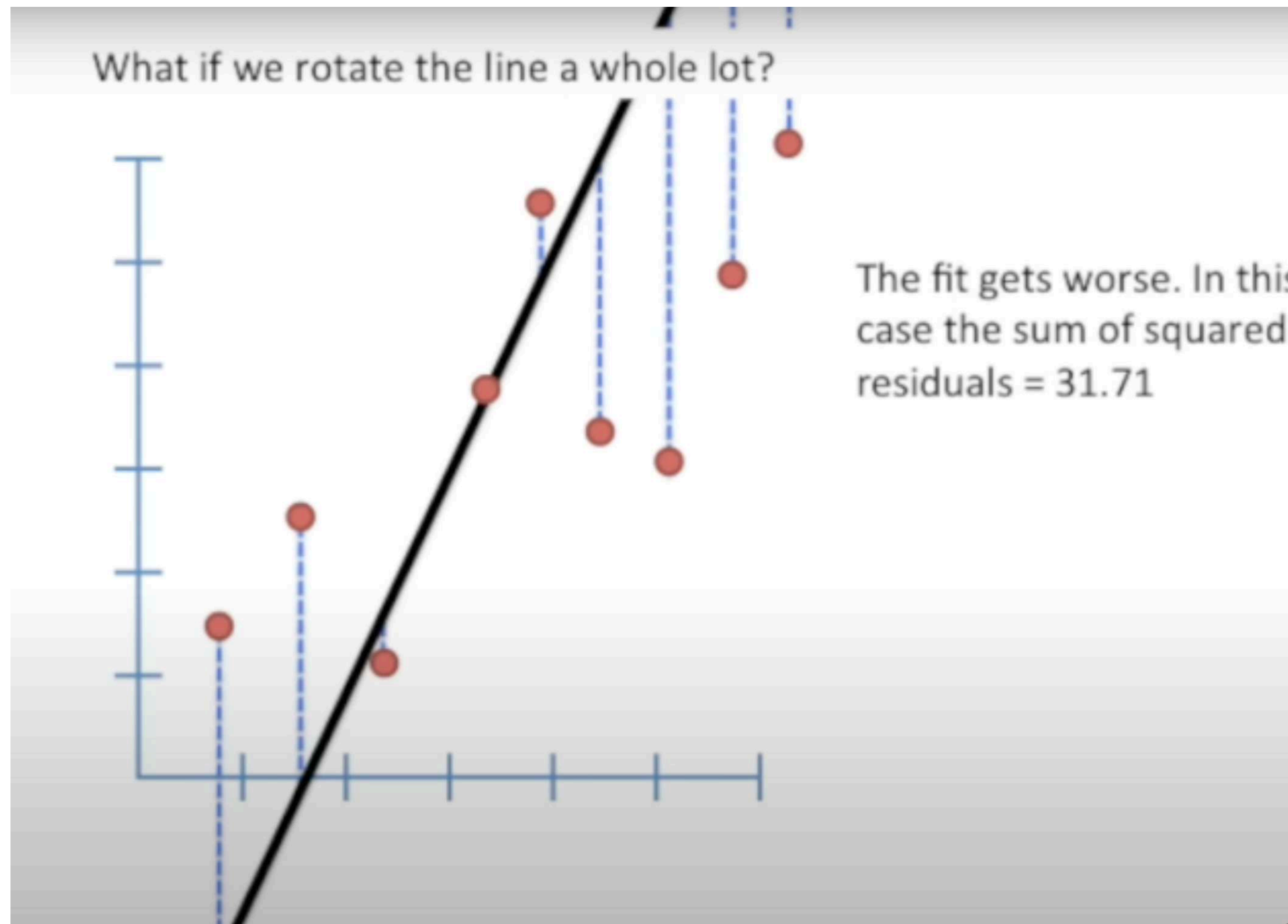
Slide Courtesy: StatQuest

Linear Regression



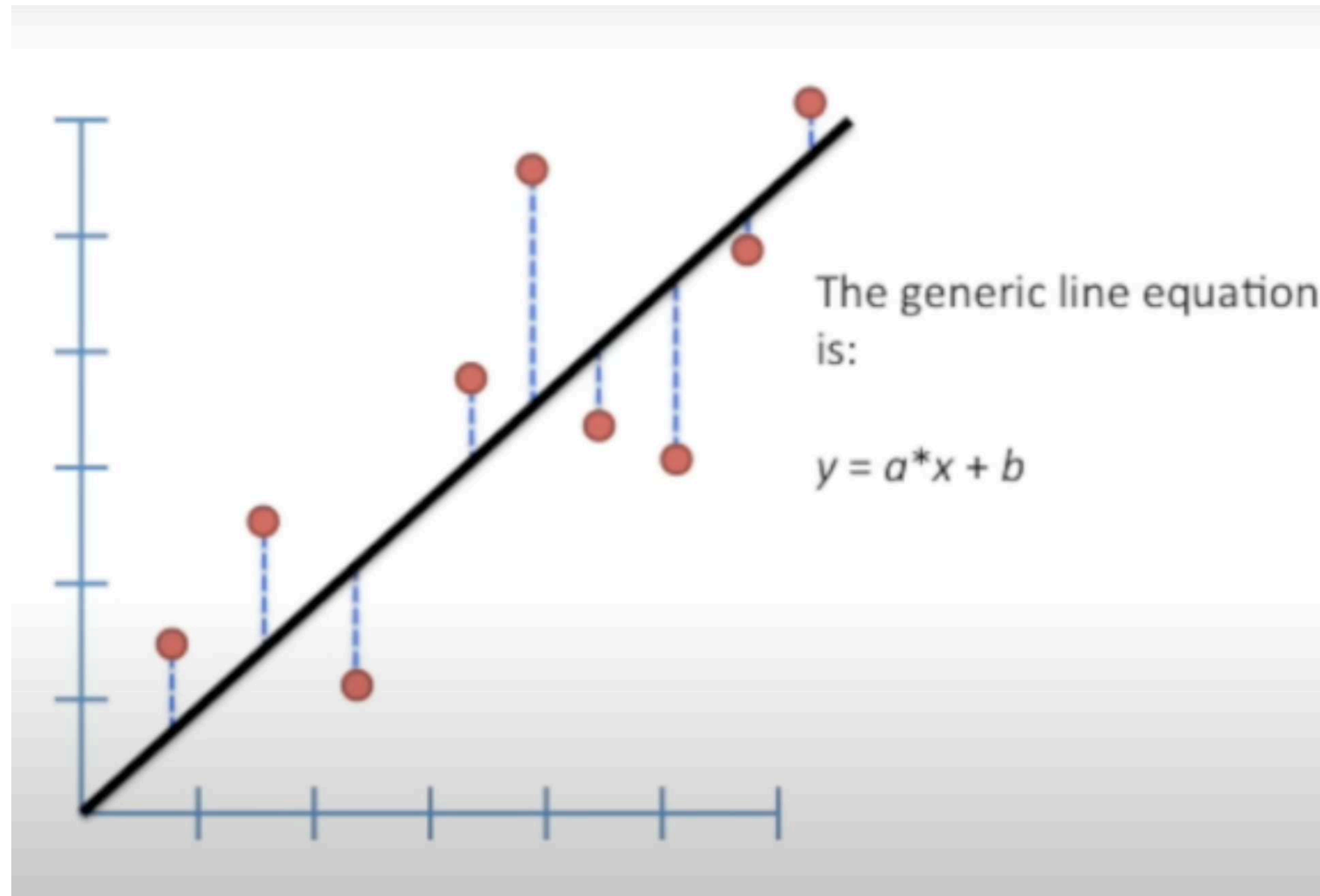
Slide Courtesy: StatQuest

Linear Regression



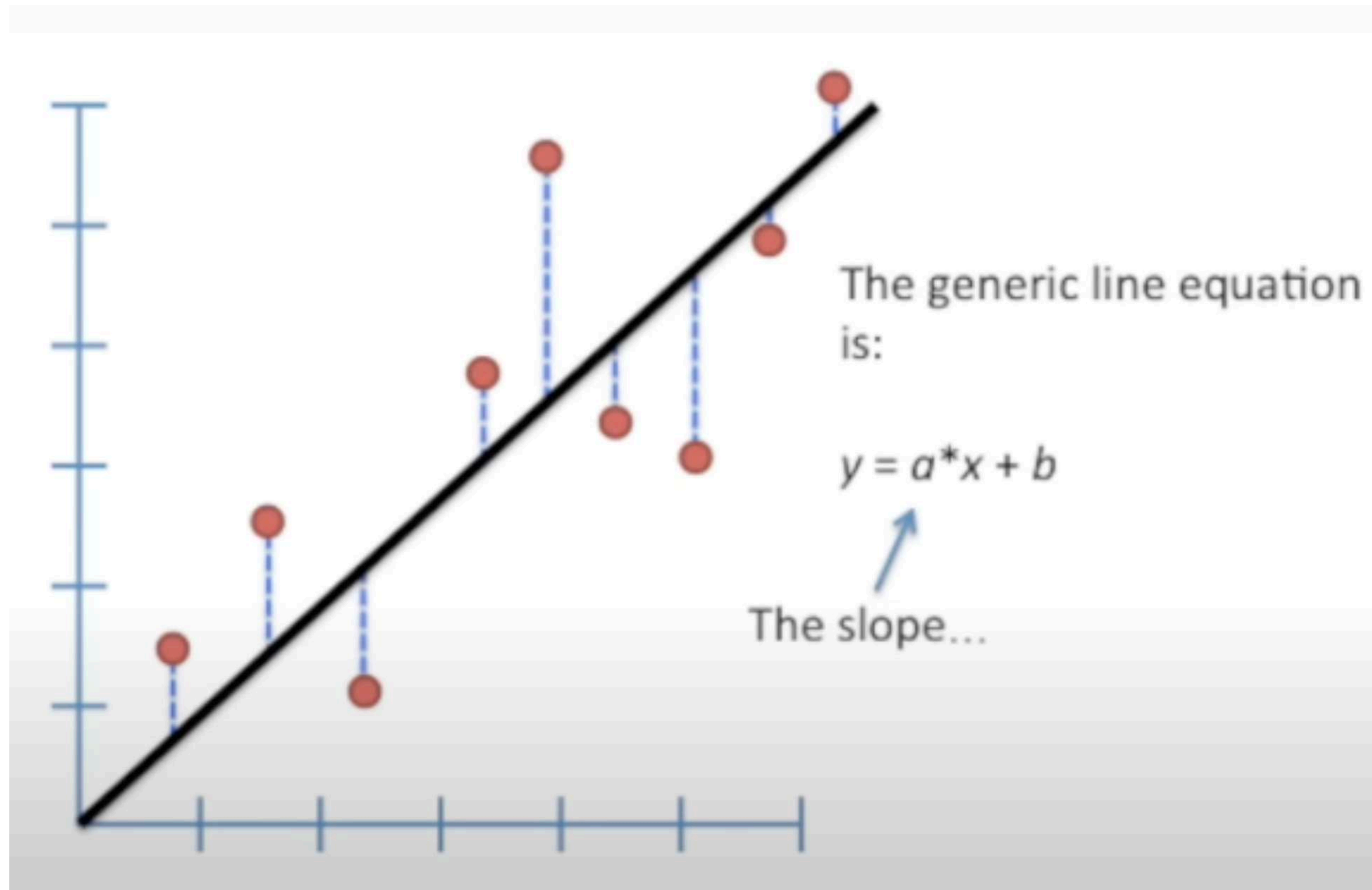
Slide Courtesy: StatQuest

Linear Regression

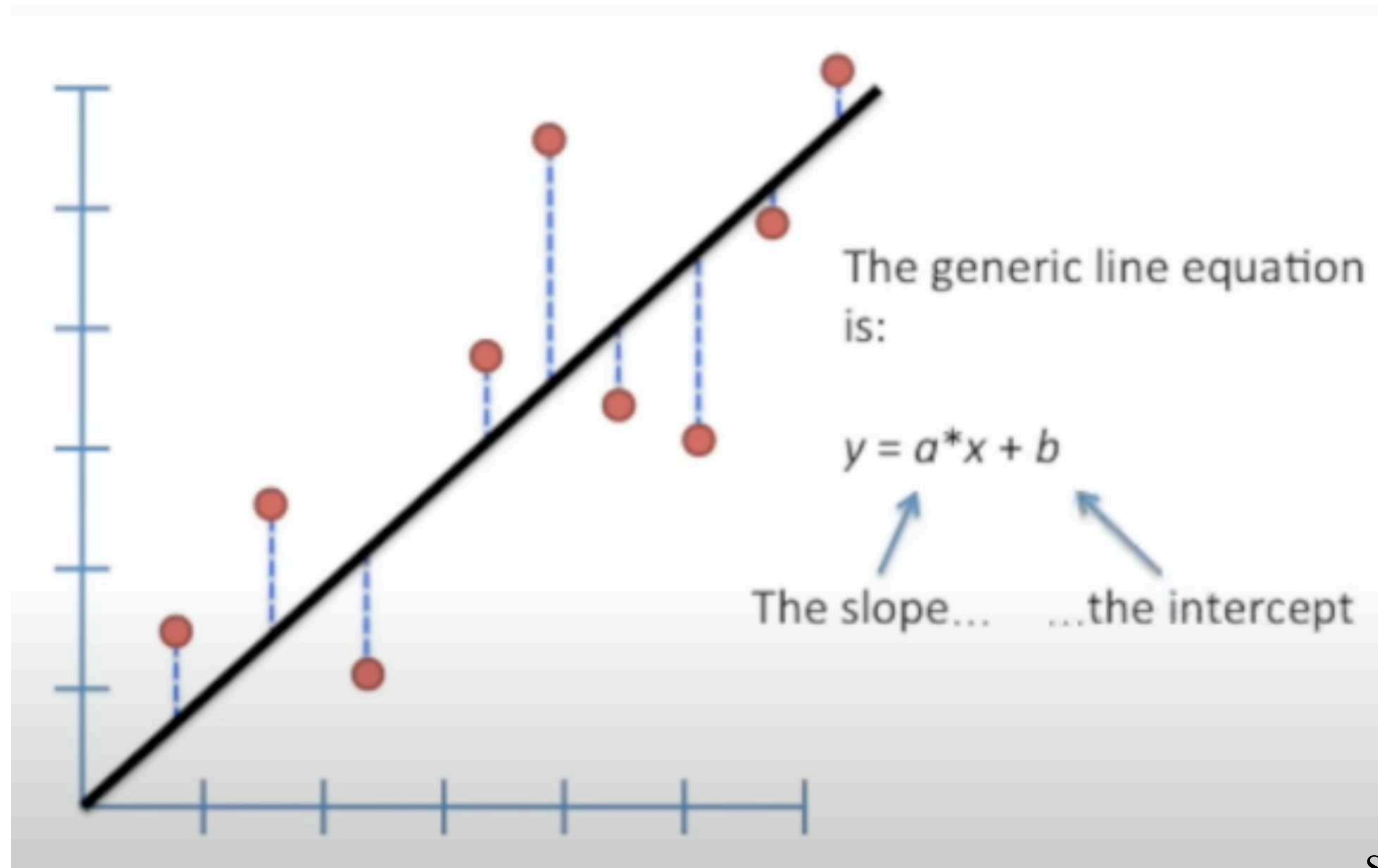


Slide Courtesy: StatQuest

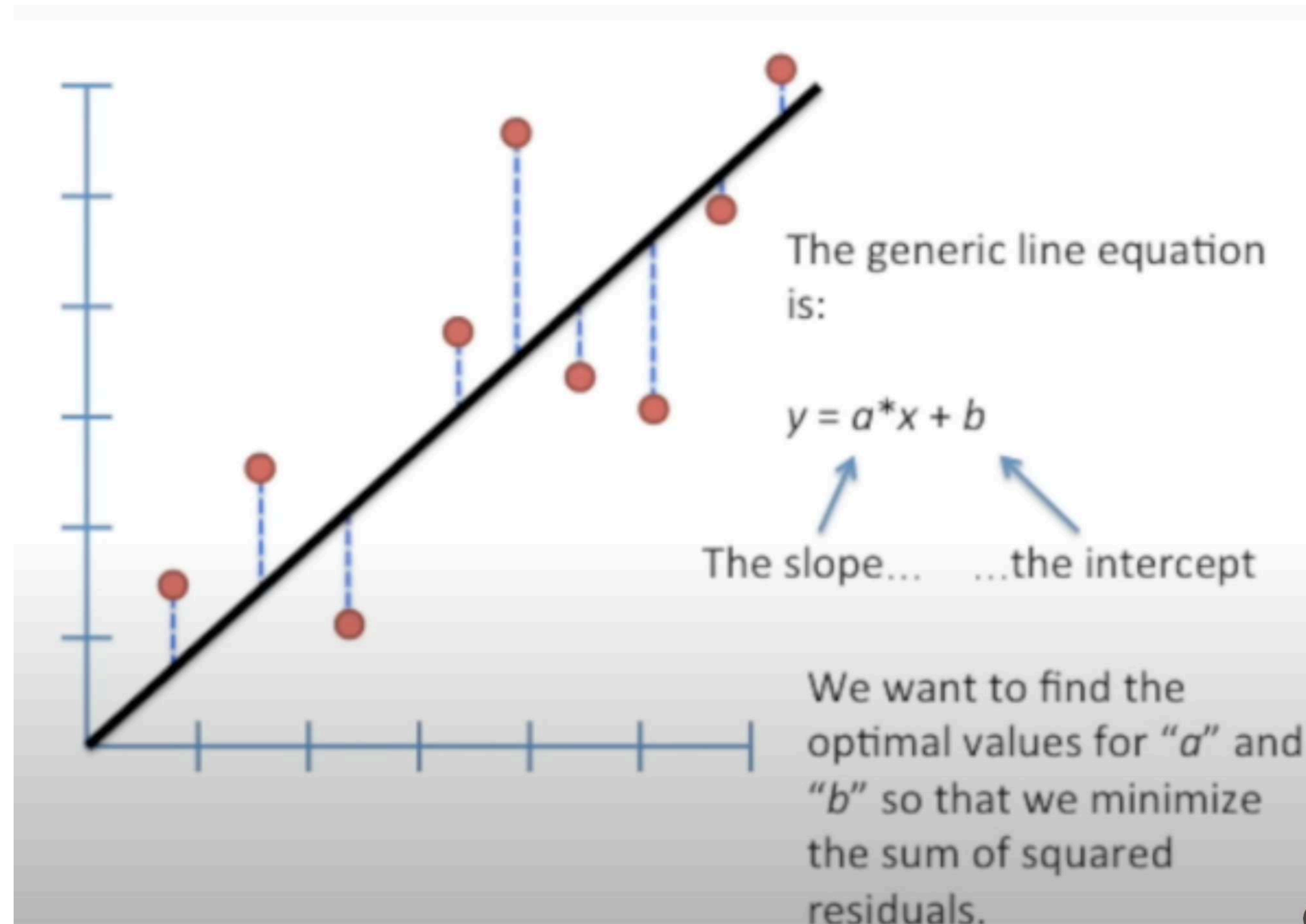
Linear Regression



Linear Regression



Linear Regression

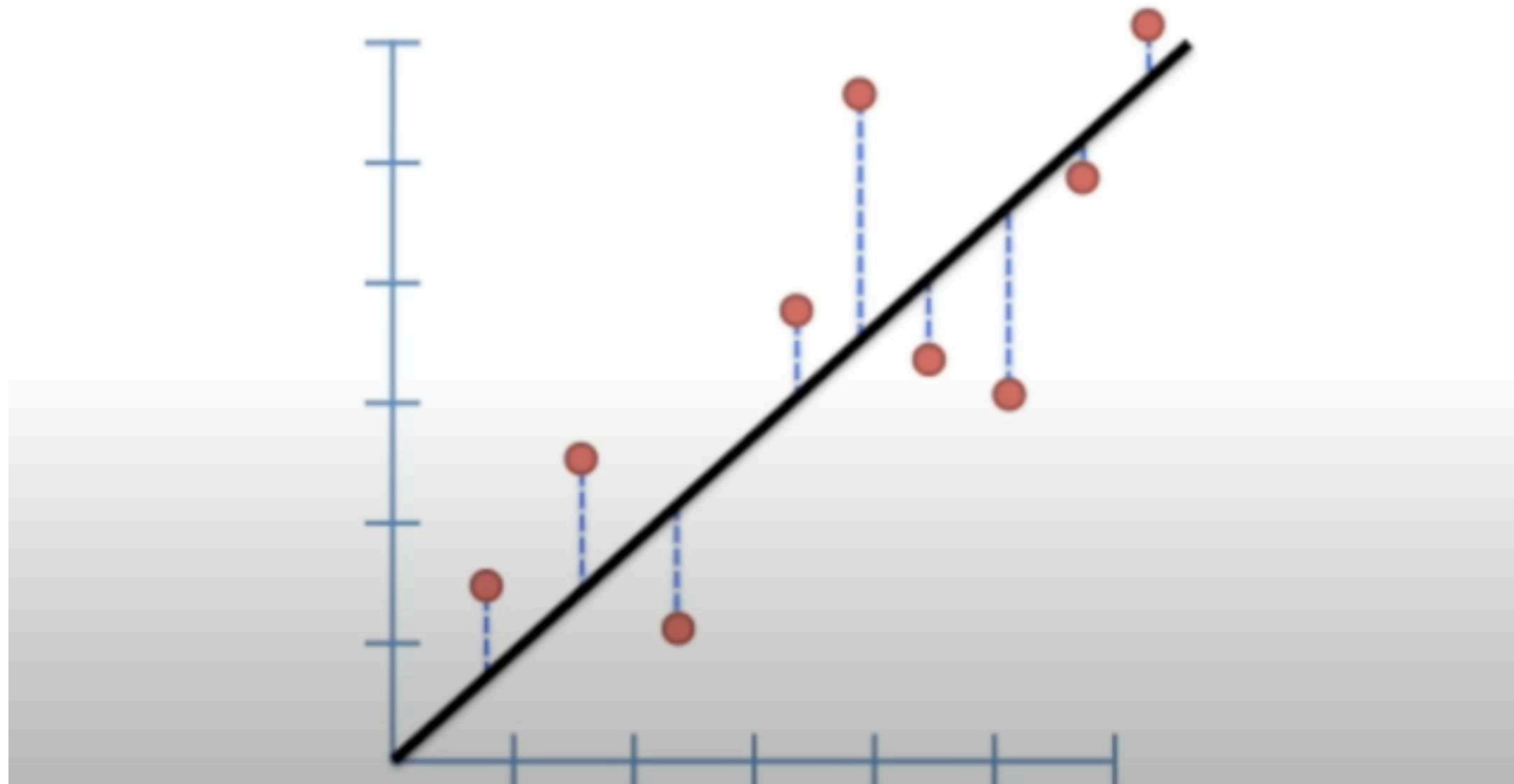


Slide Courtesy: StatQuest

Linear Regression

In more general math terms...

Sum of squared residuals = $((a \cdot x_1 + b) - y_1)^2 + ((a \cdot x_2 + b) - y_2)^2 + \dots$

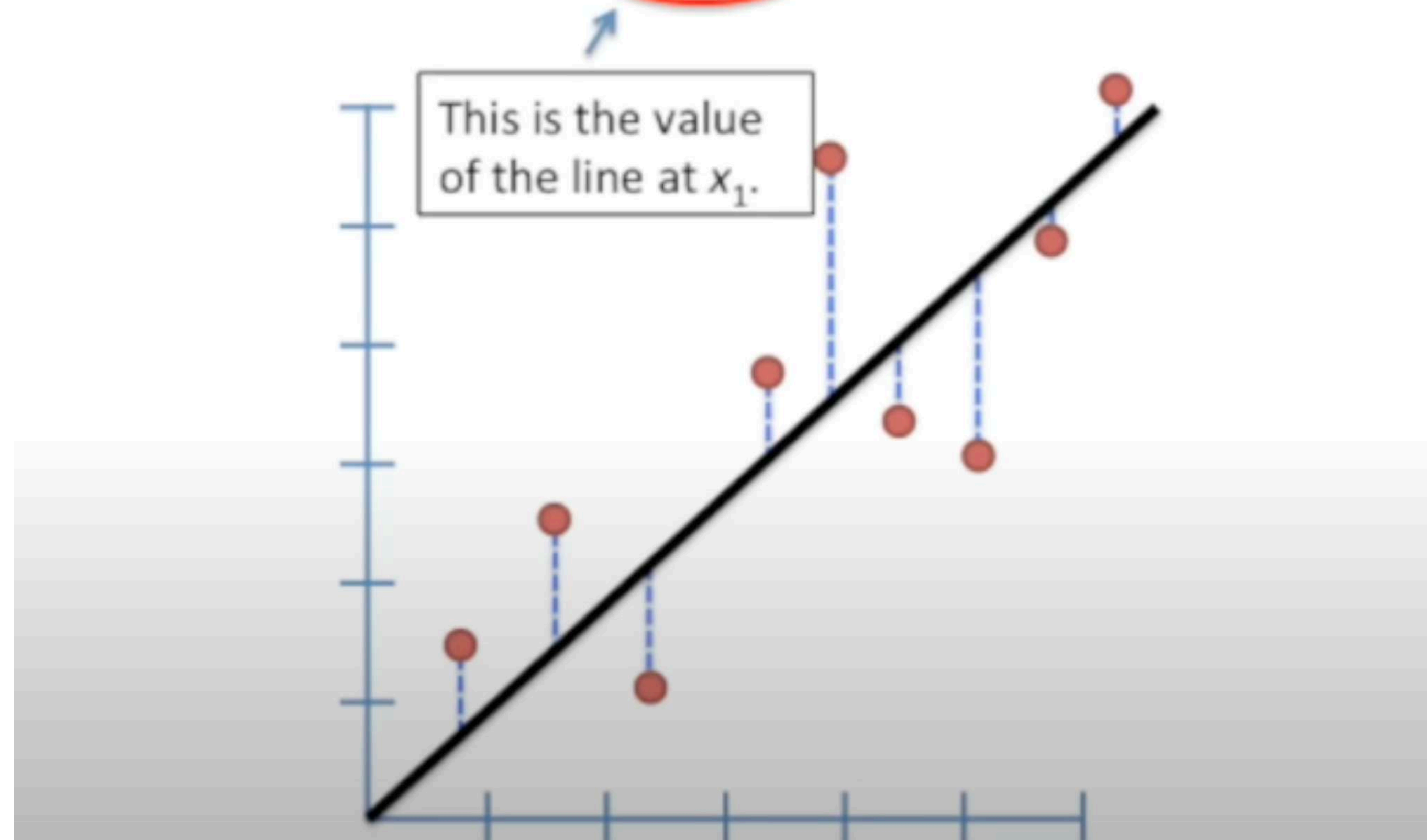


Slide Courtesy: StatQuest

Linear Regression

In more general math terms...

Sum of squared residuals = $((a \cdot x_1 + b) - y_1)^2 + ((a \cdot x_2 + b) - y_2)^2 + \dots$

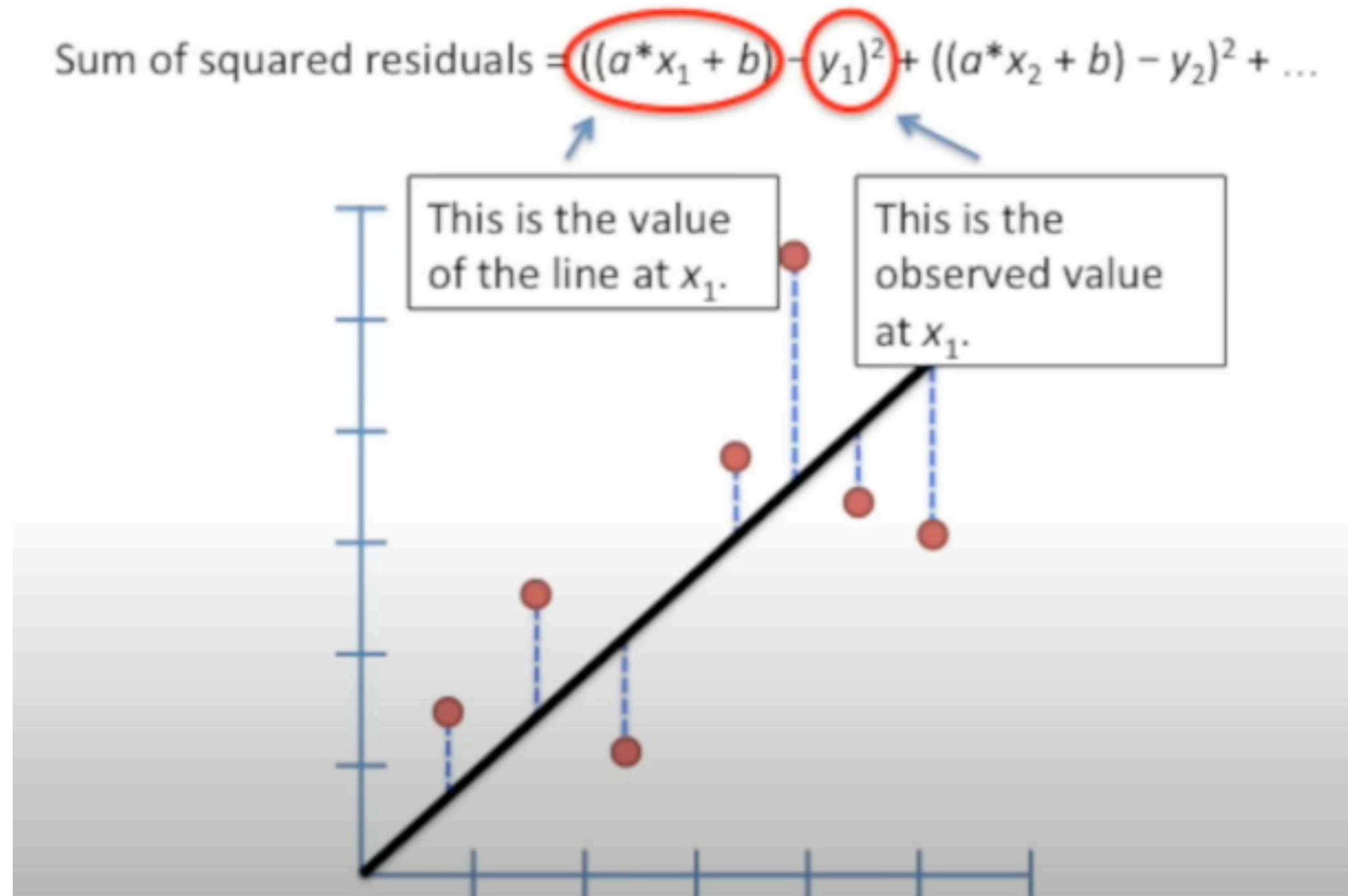


Slide Courtesy: StatQuest

Linear Regression

In more general math terms...

Sum of squared residuals = $((a \cdot x_1 + b) - y_1)^2 + ((a \cdot x_2 + b) - y_2)^2 + \dots$



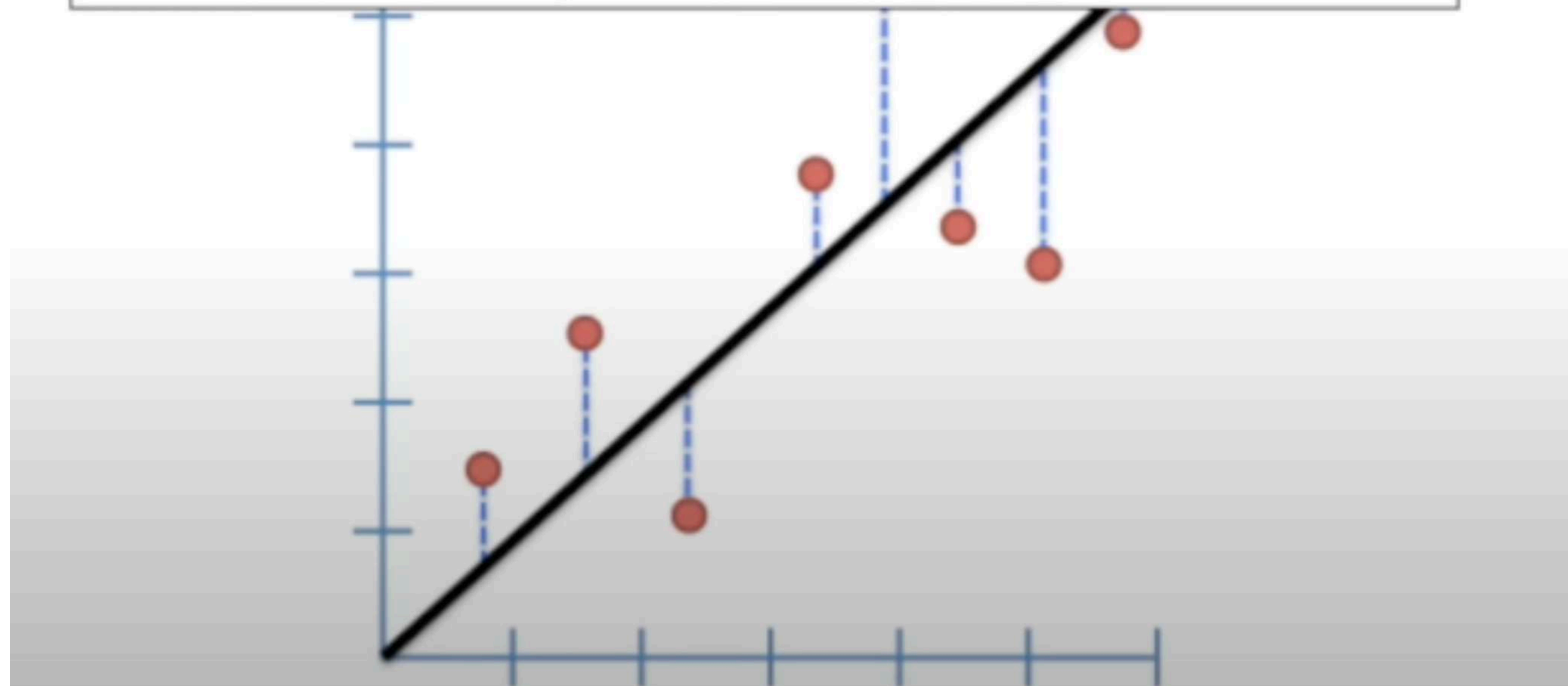
Slide Courtesy: StatQuest

Linear Regression

In more general math terms...

$$\text{Sum of squared residuals} = ((a \cdot x_1 + b) - y_1)^2 + ((a \cdot x_2 + b) - y_2)^2 + \dots$$

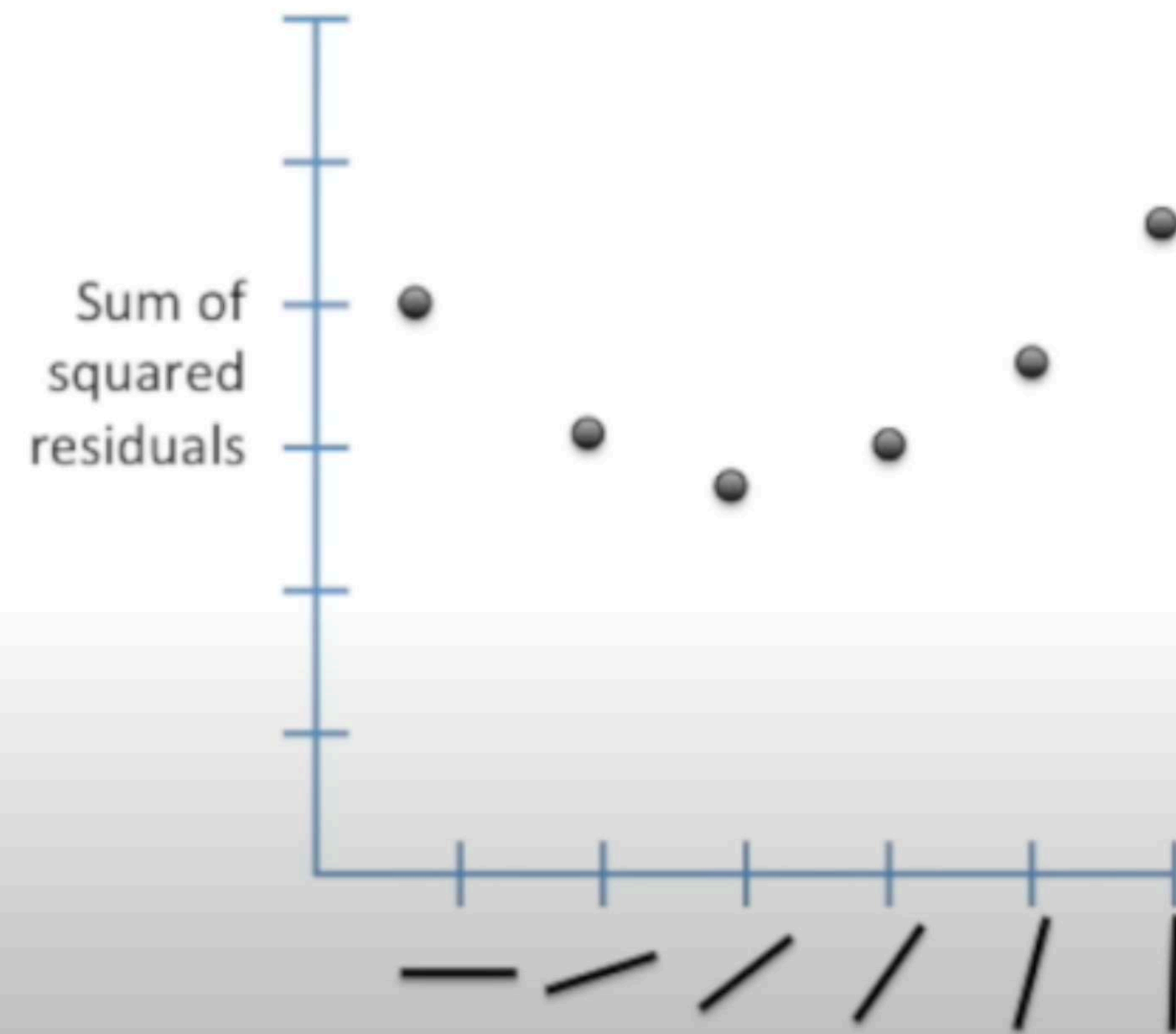
Since we want the line that will give us the smallest sum of squares, this method for finding the best values for “ a ” and “ b ” is called “Least Squares”.



Slide Courtesy: StatQuest

Linear Regression

If we plotted the sum of squared residuals vs. each rotation, we'd get something like this...

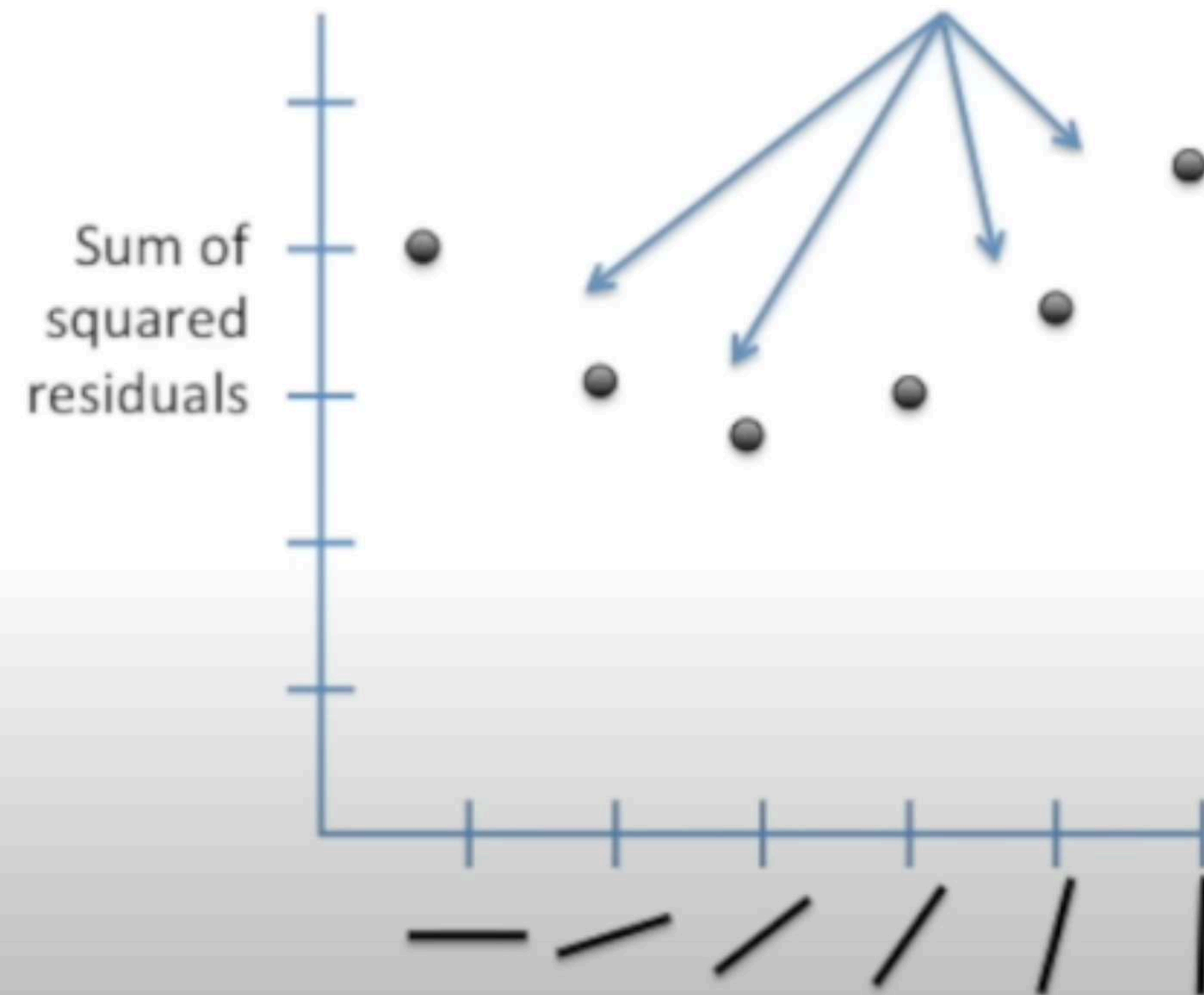


Slide Courtesy: StatQuest

Linear Regression

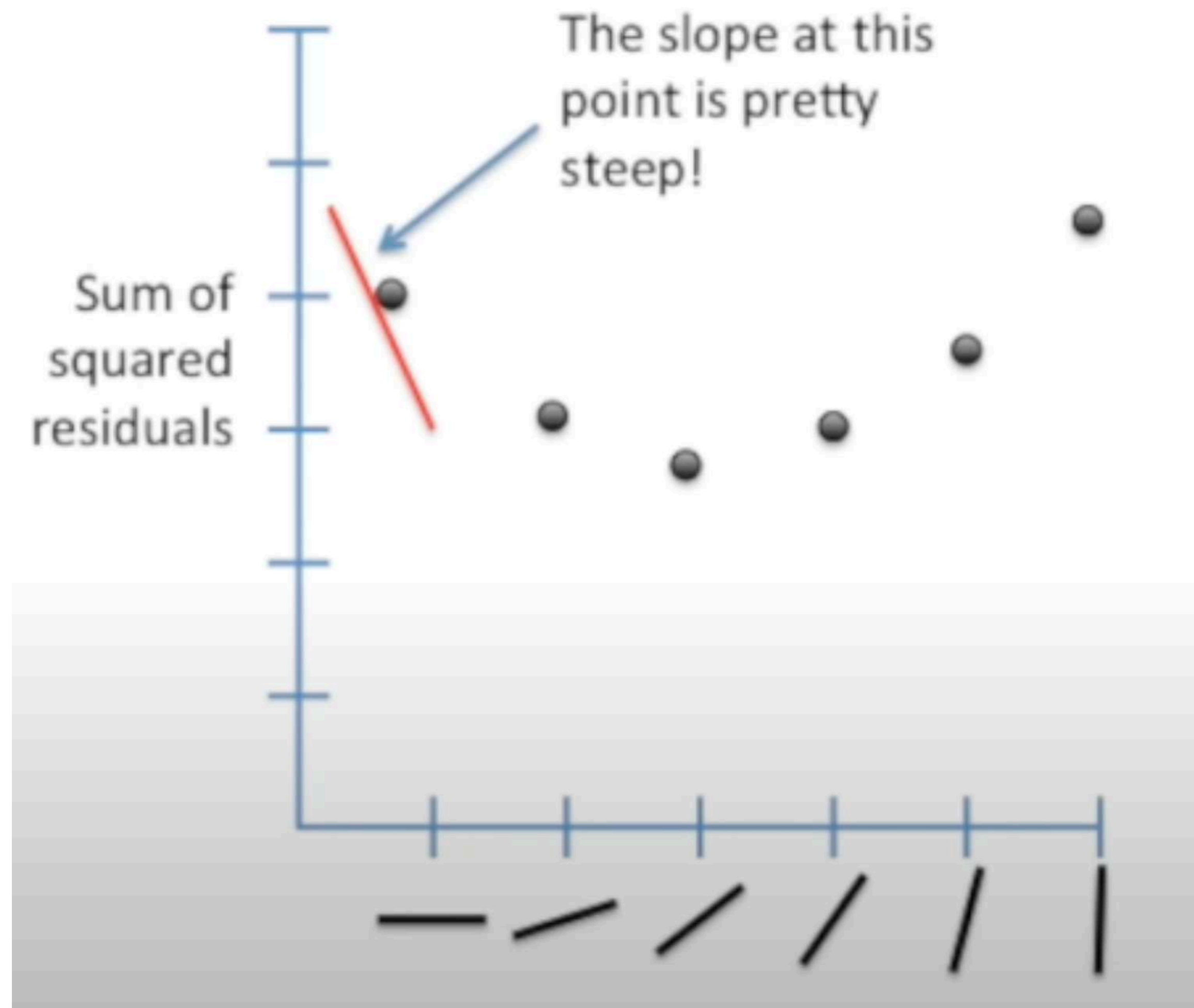
How do we find the optimal rotation for the line?

We take the derivative of this function.



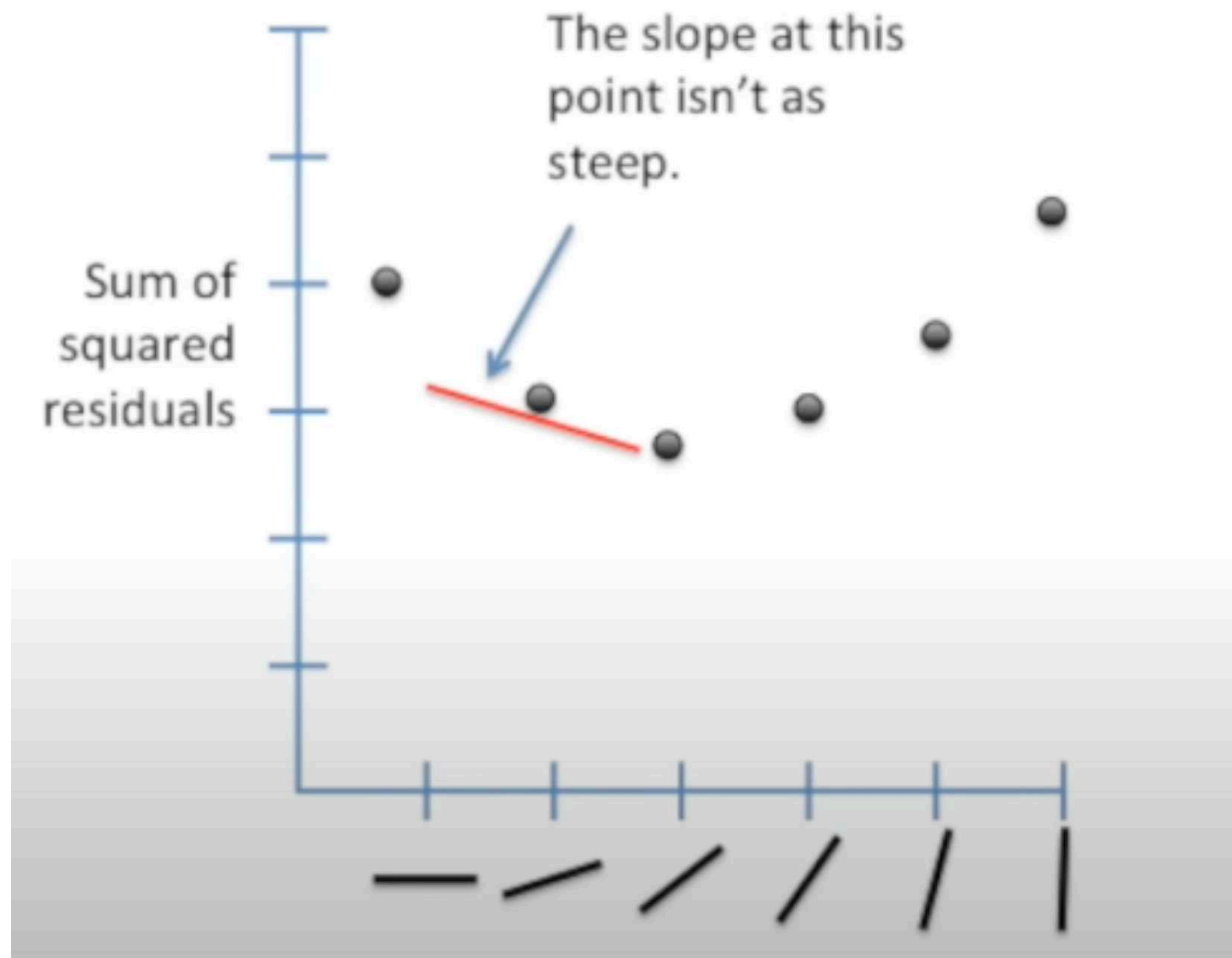
Slide Courtesy: StatQuest

Linear Regression



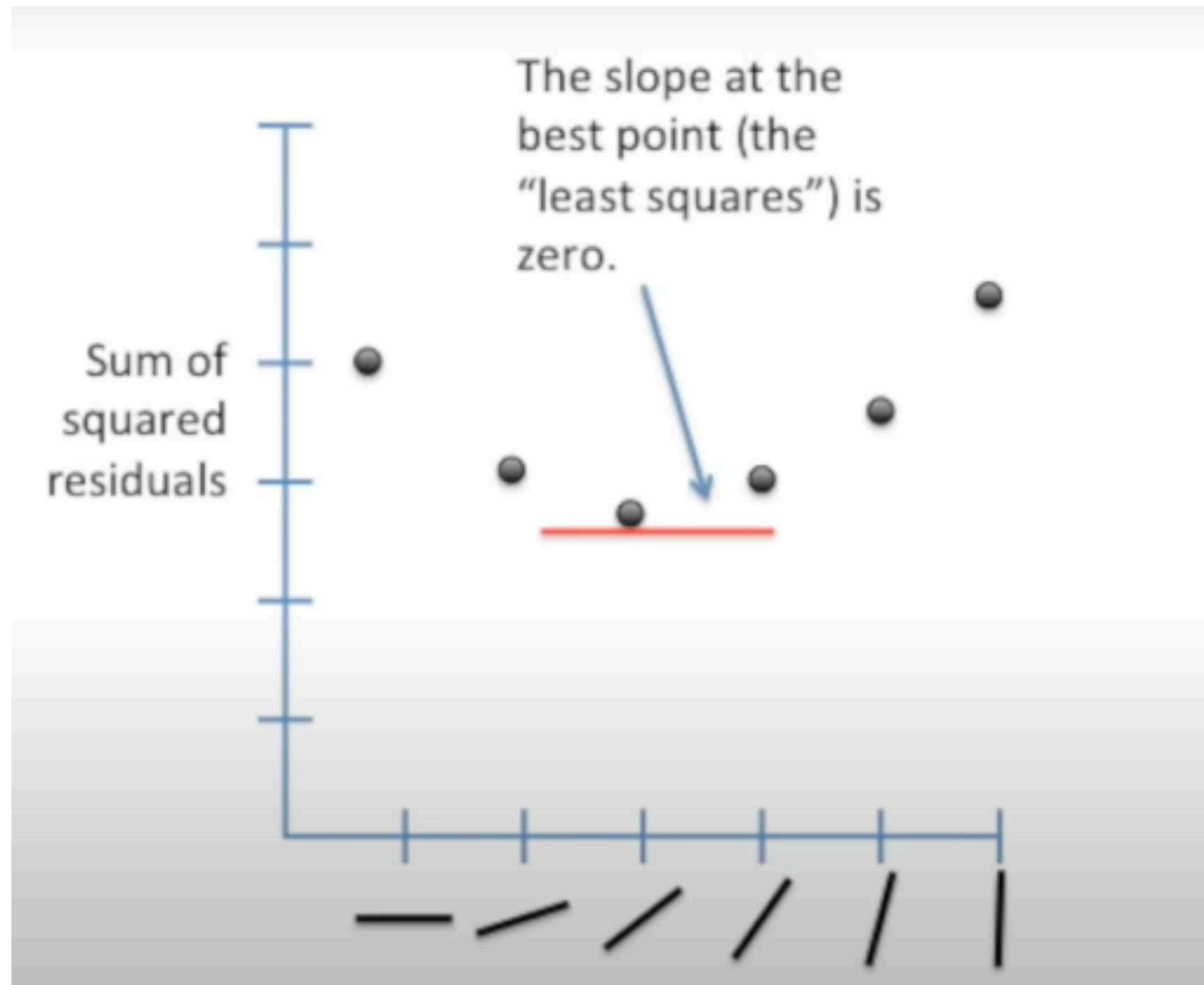
Slide Courtesy: StatQuest

Linear Regression



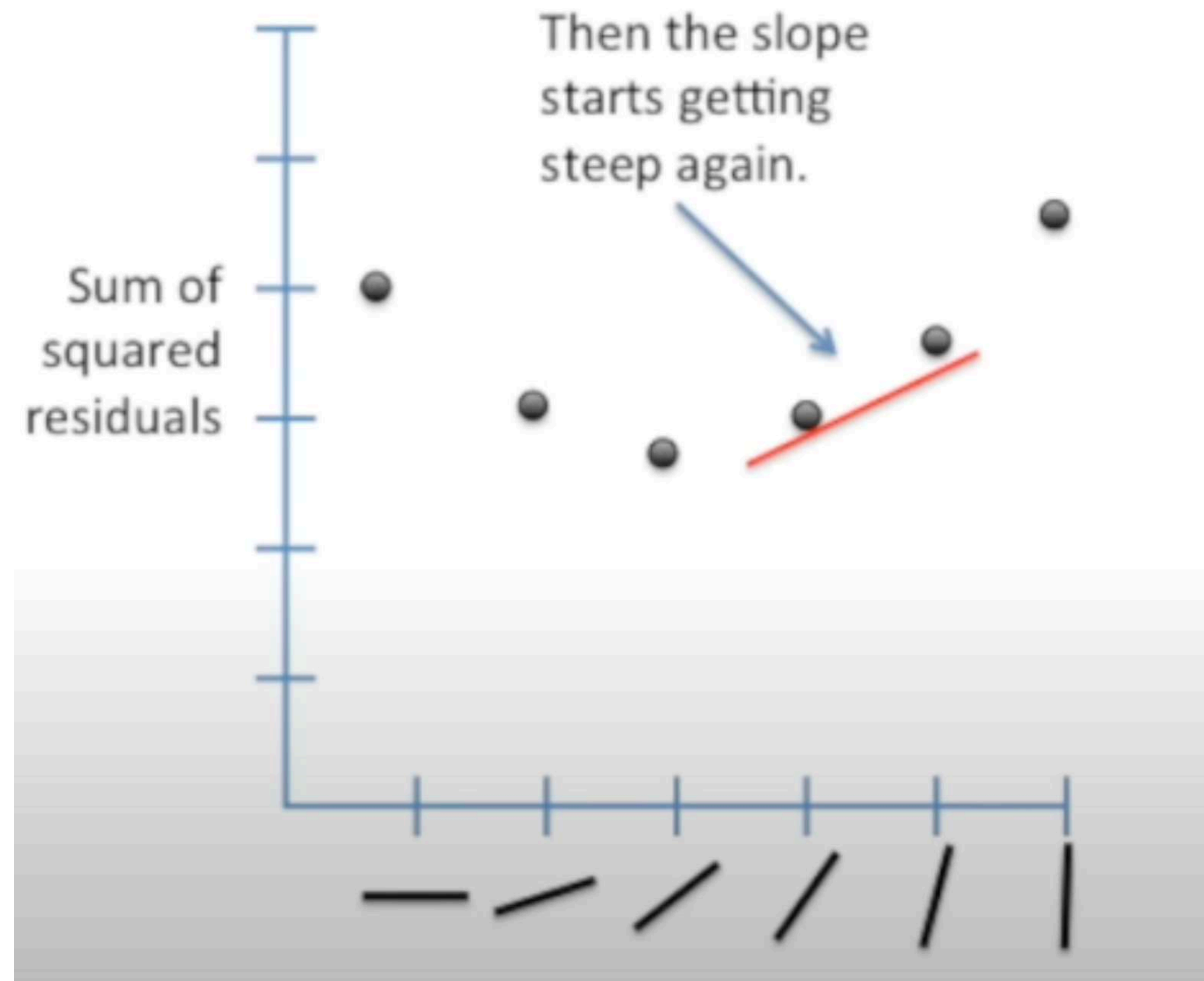
Slide Courtesy: StatQuest

Linear Regression

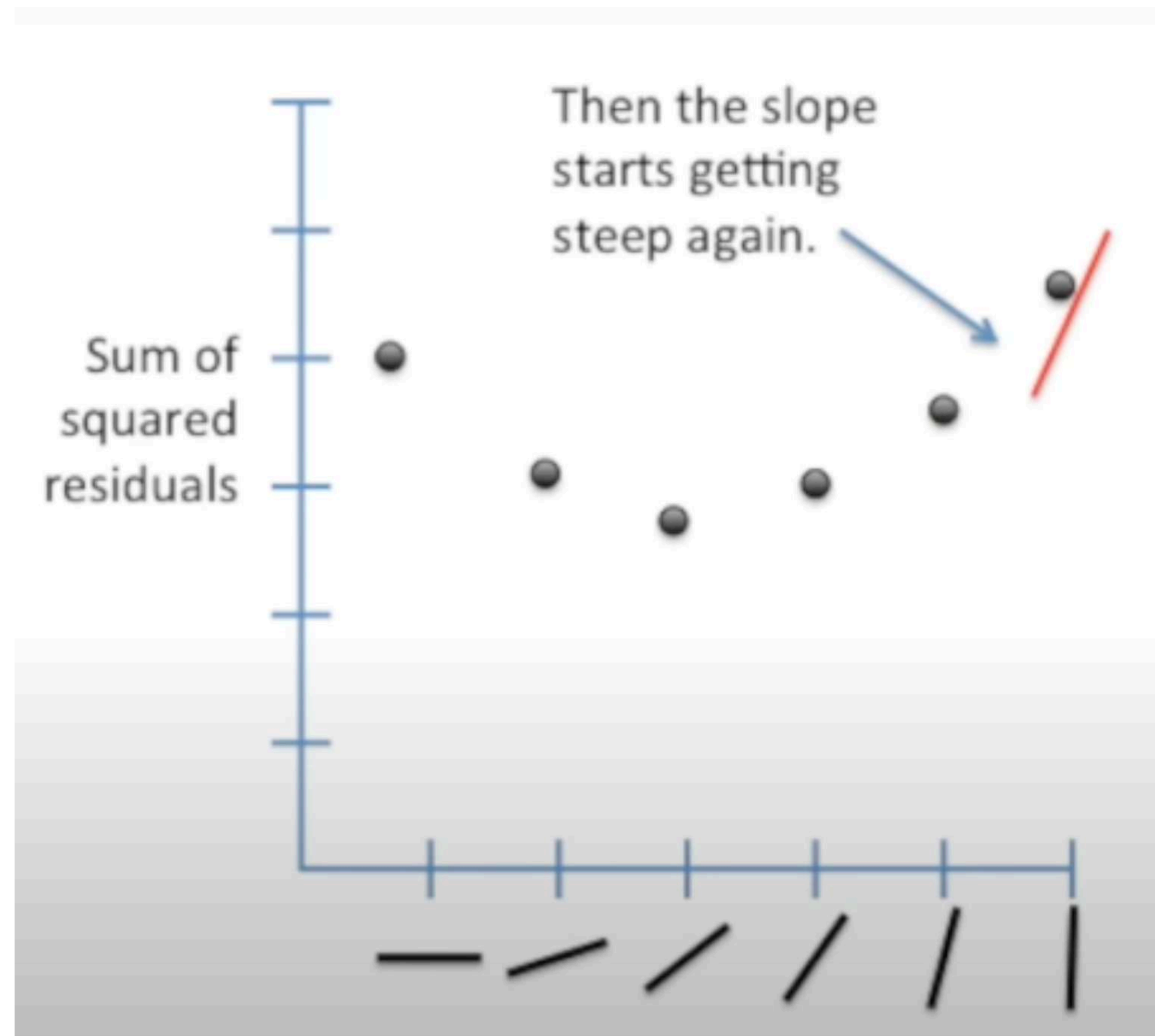


Slide Courtesy: StatQuest

Linear Regression

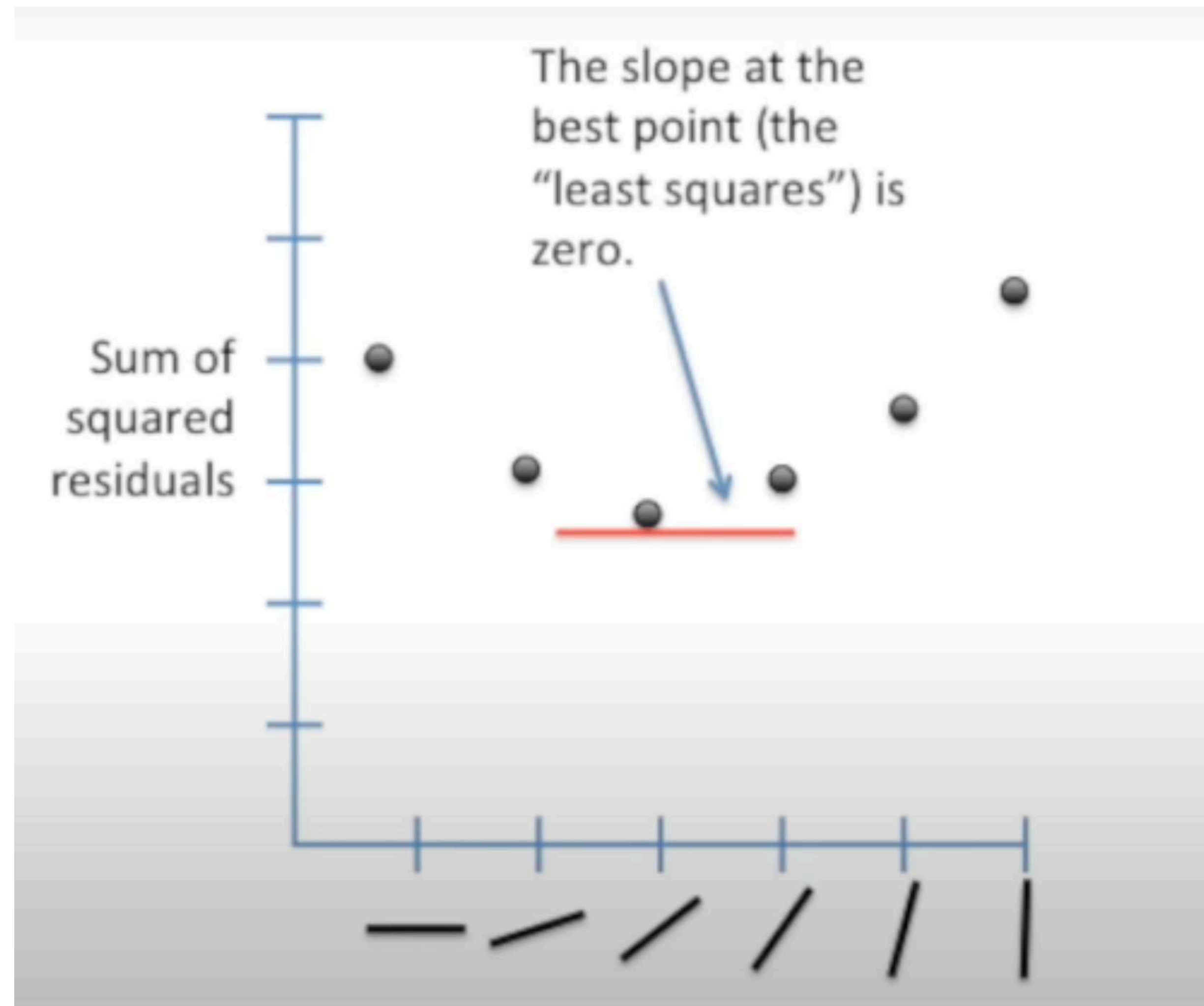


Linear Regression

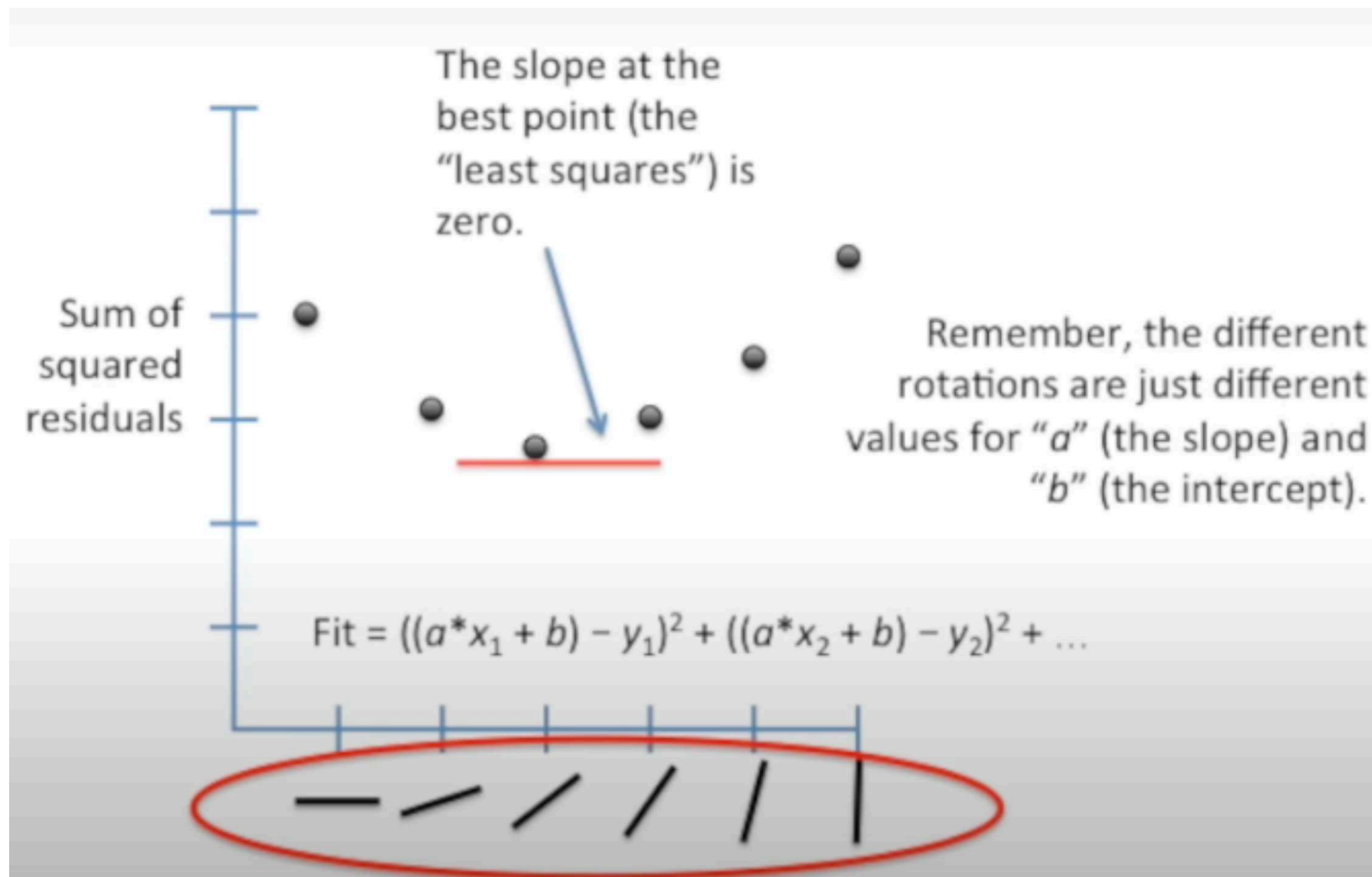


Slide Courtesy: StatQuest

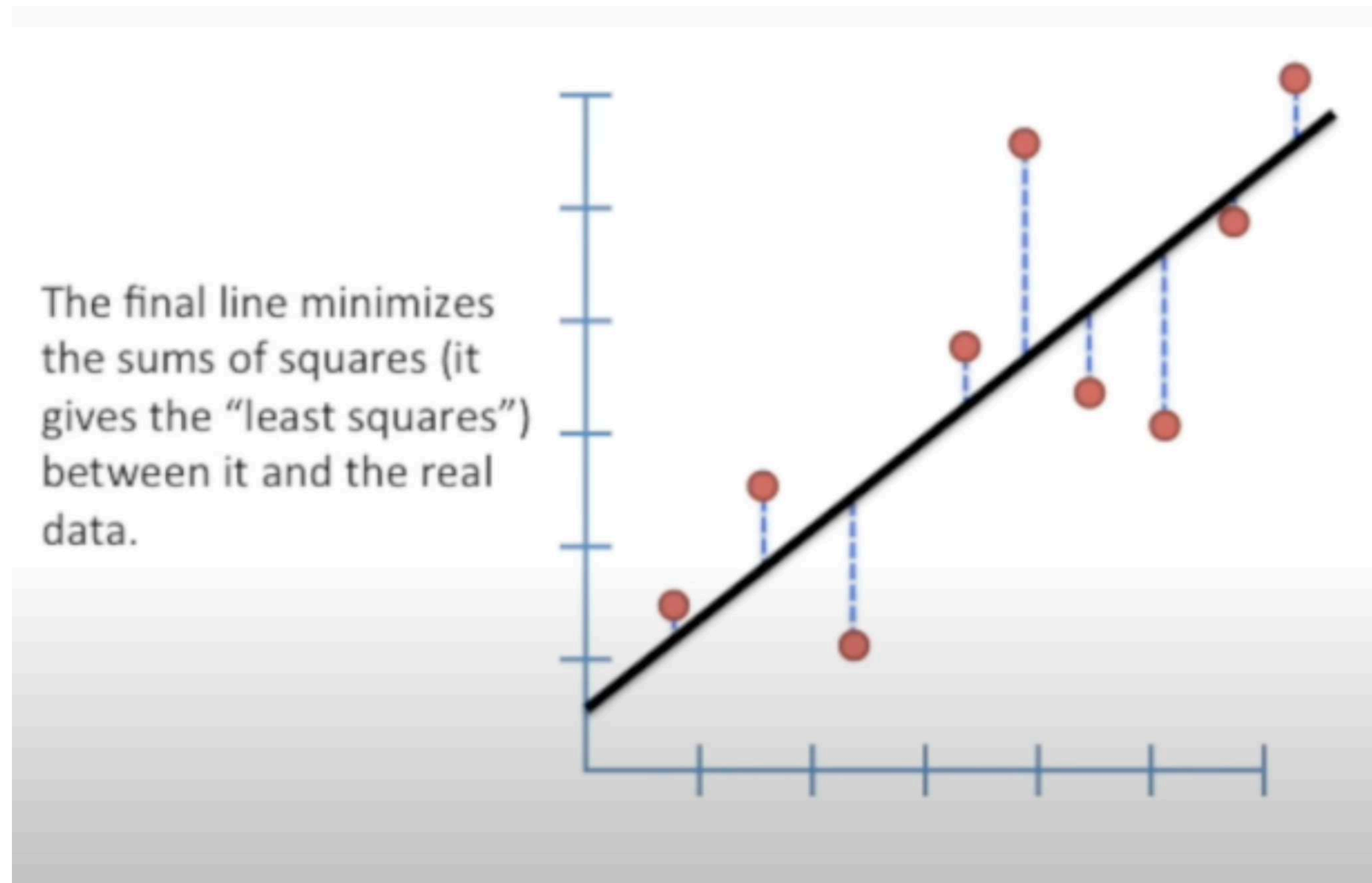
Linear Regression



Linear Regression



Linear Regression



Slide Courtesy: StatQuest

Linear Regression

