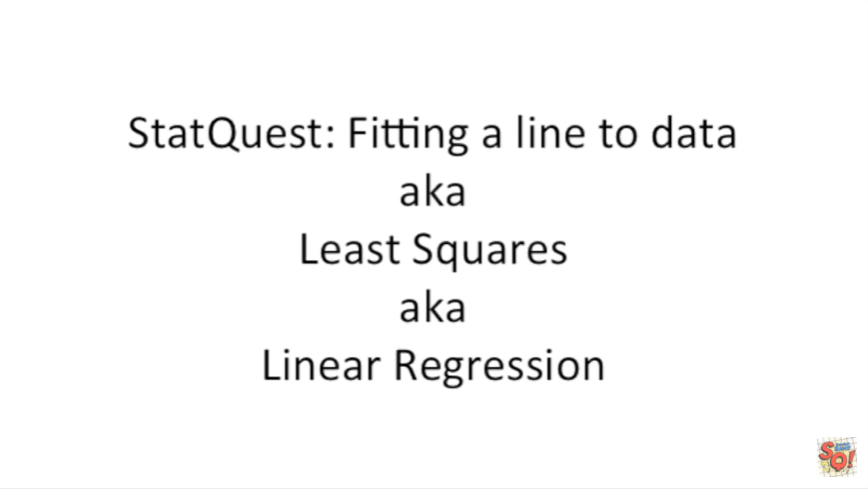
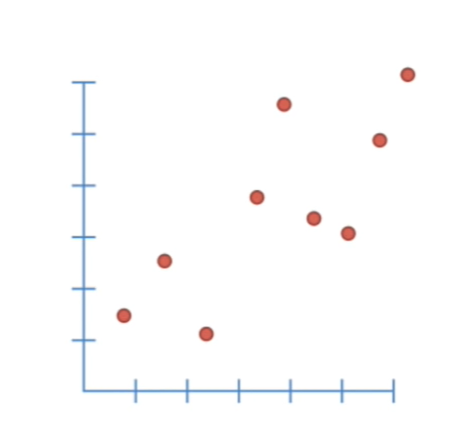
<https://www.youtube.com/watch?v=PaFPbb66DxQ&list=PLblh5JKOoLUIzaEkCLIUxQFjPIlapw8nU&index=1>



Hello and welcome to stat quest.

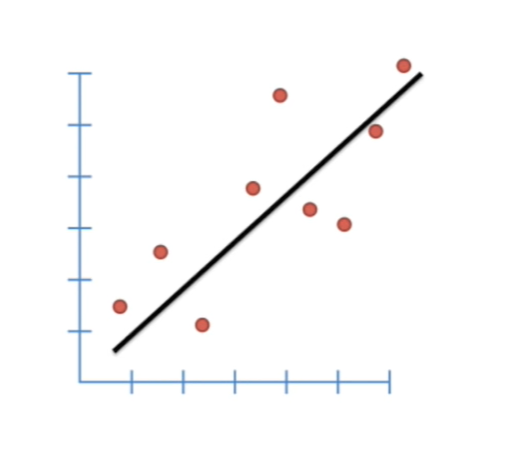
Stat quest is brought to you by the friendly folks in the genetics department at the University of North Carolina at Chapel Hill.

Today we're going to talk about fitting a line to data, aka least-squares, aka linear regression.



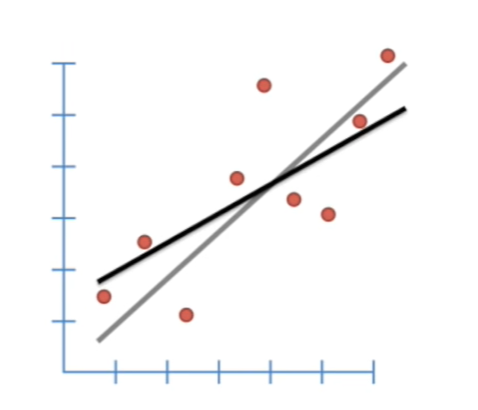
Now let's get to it okay worked really hard you did the experiment and now you got some data.

Here it is plotted on an XY graph.



We usually like to add a line to our data so we can see what the trend is.

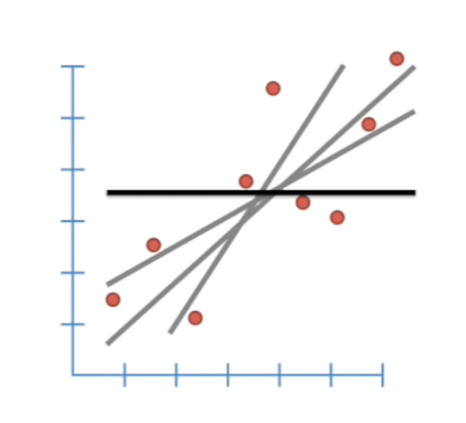
But is this the best line we should use ?



Or does this new line fit the data even better ?



Or what about this line is it better or worse than the other options ?



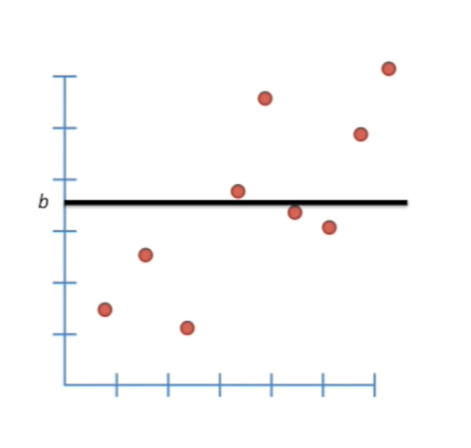
A horizontal line that cuts through the average Y value of our data is probably the worst fit of all.

However it gives us a good starting point for talking about how to find the optimal line to fit our data.



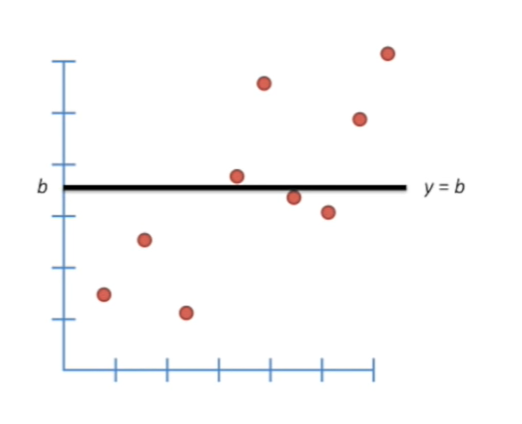
So now let's focus on this horizontal line.

It cuts through the average Y value which is 3.5.

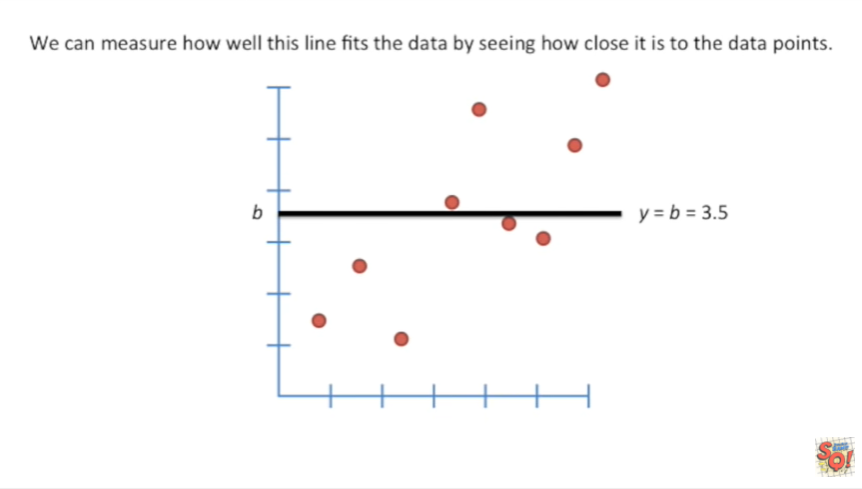


Let's just call this point B.

Because different data sets will have different average values on the y axis.

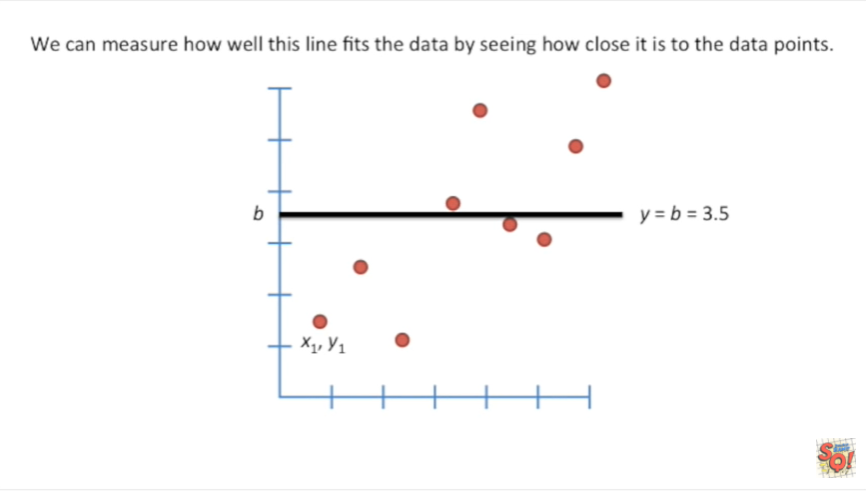


That is to say the Y value for this line is B.

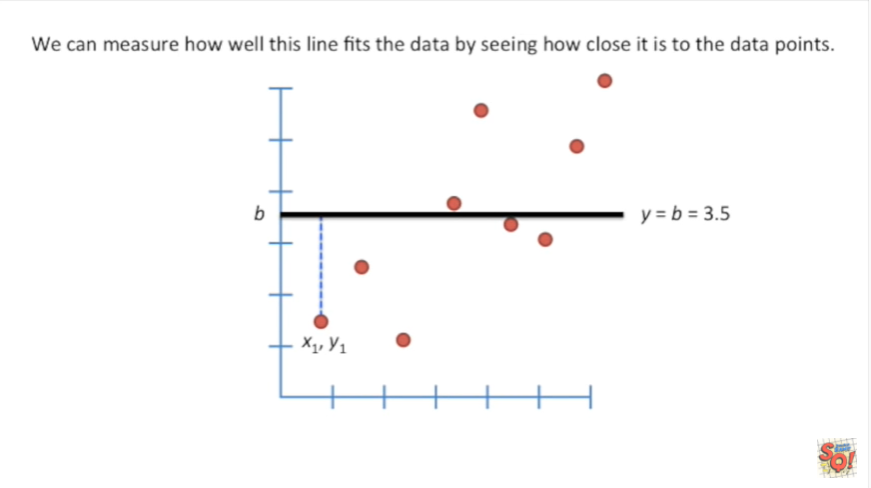


And for this particular data set B equals 3.5.

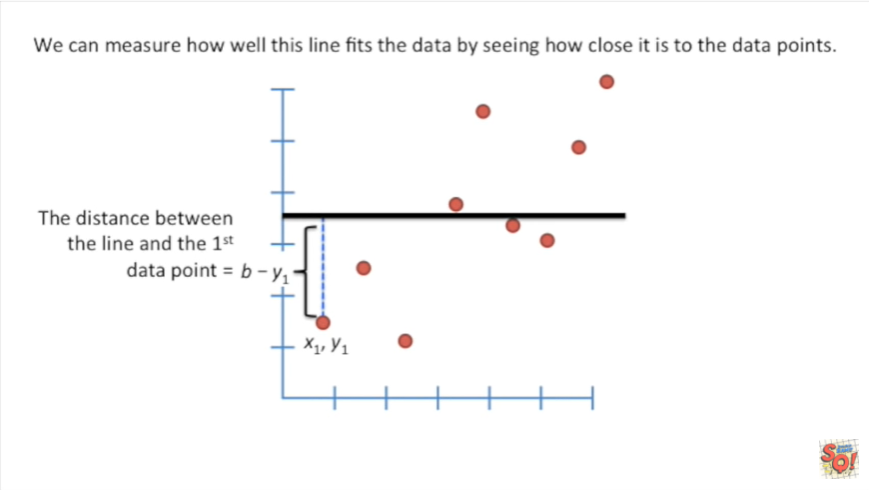
We can measure how well this line fits the data by seeing how close it is to the data points.



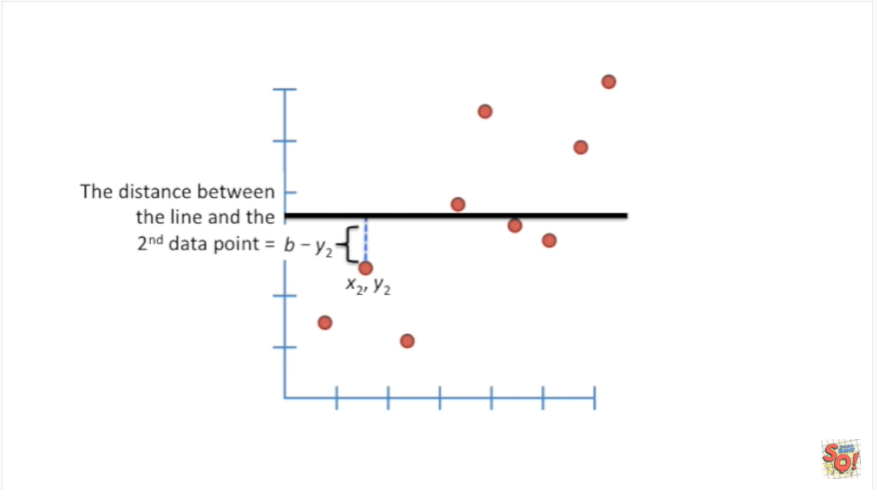
We'll start with the point in the lower left-hand corner of the graph with coordinates x-one y-one.



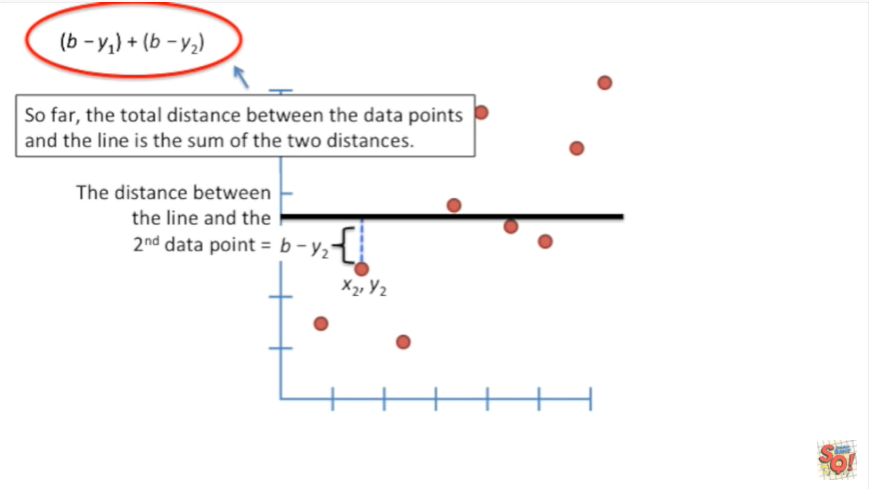
We can now draw a line from this point up to the line that cuts across the average Y value for this data set.



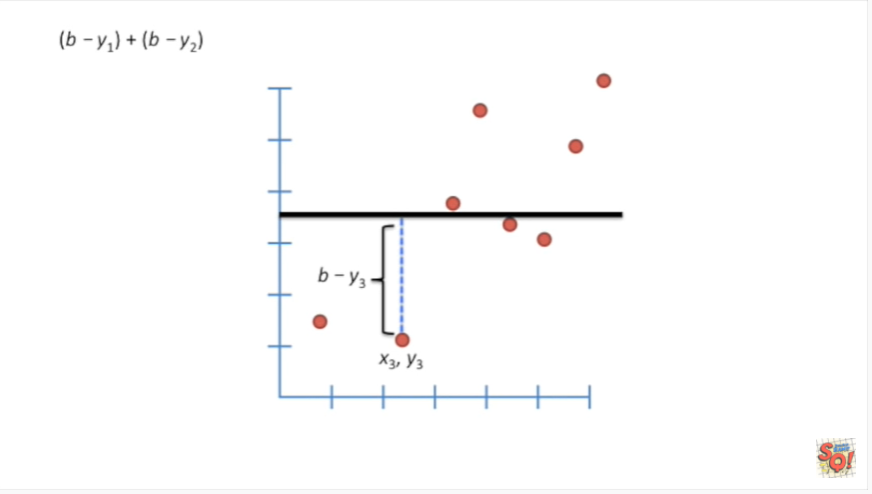
The distance between the line and the first data point equals B minus y1.



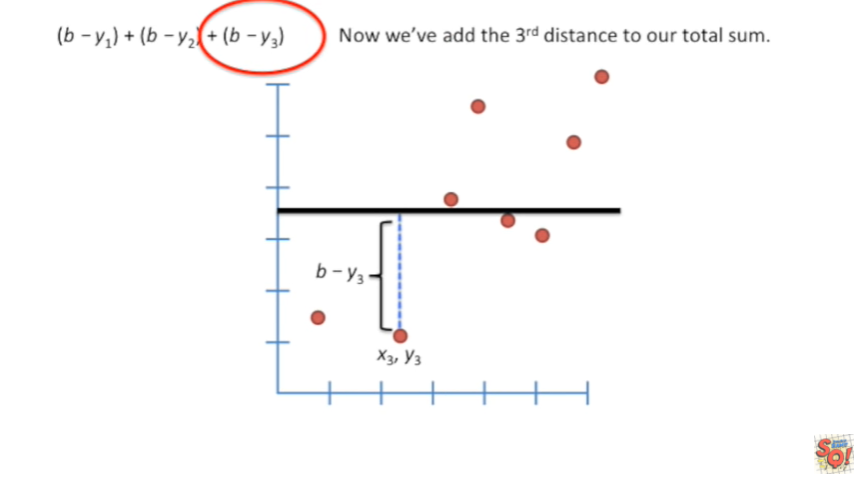
The distance between the line and the second data point is B minus y2.



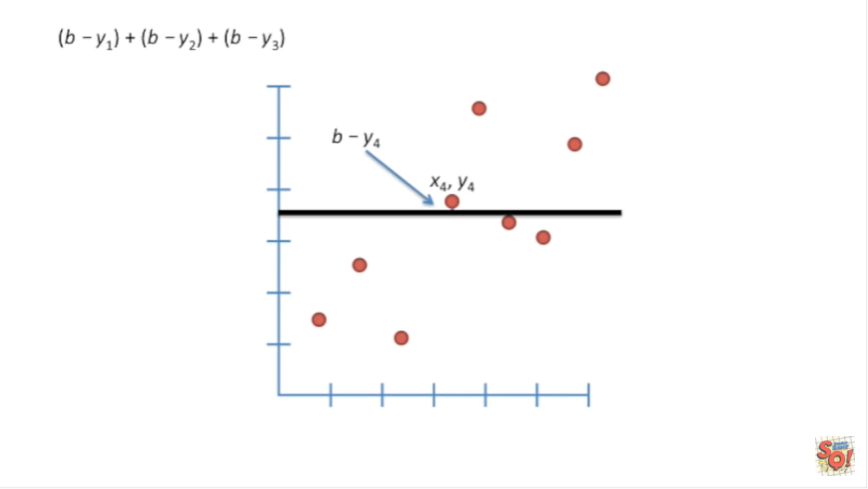
So far the total distance between the data points and the line is the sum of the two distances.



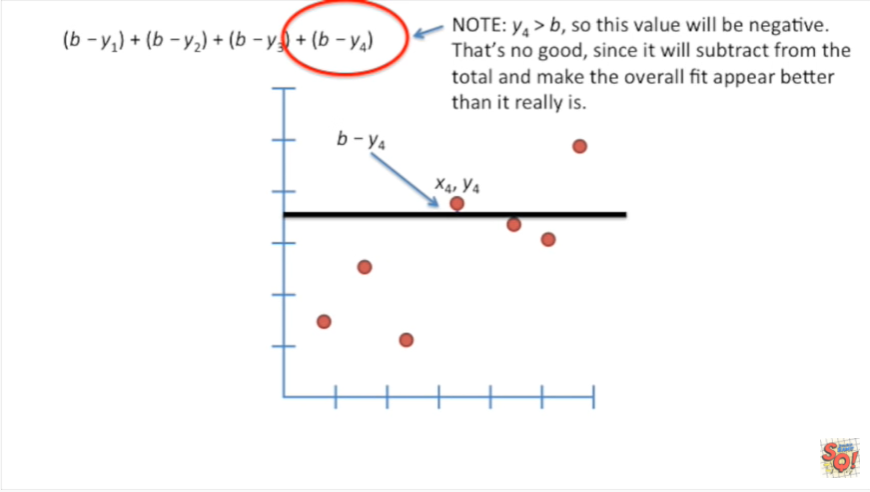
And we can calculate the distance between the line and the third point that equals B minus y3.



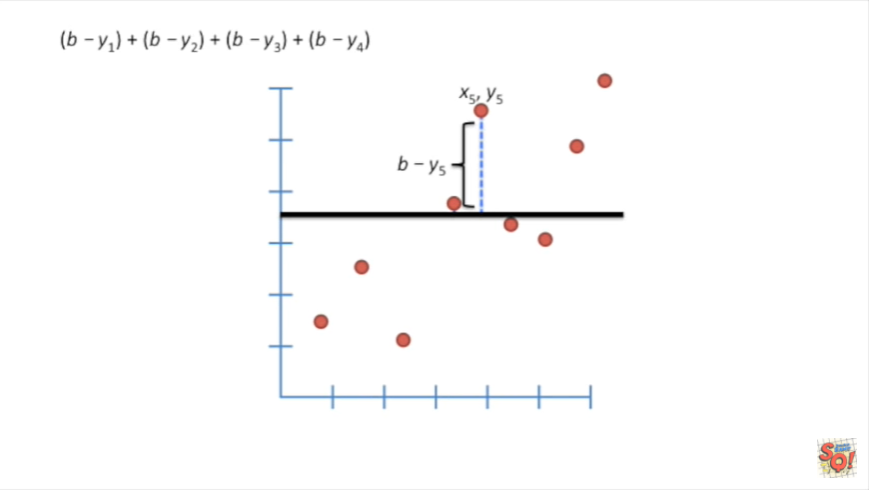
Now we've added the third distance to our total sum.



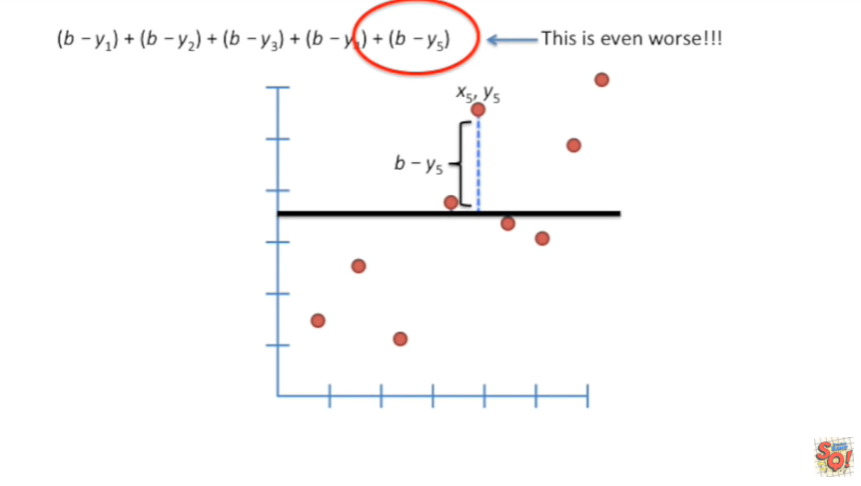
The distance for the fourth point is B minus y4.



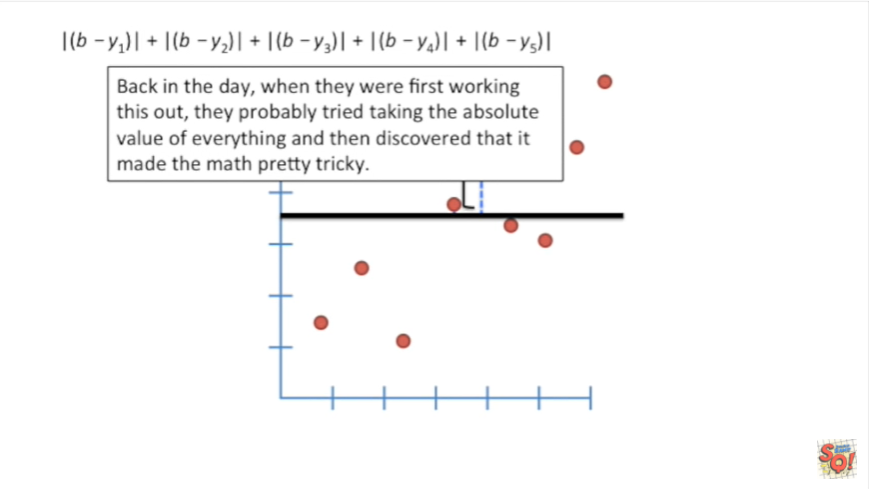
Note : y4 is greater than B because it's above the horizontal line so this value will be negative that's no good since it will subtract from the total and make the overall fit appear better than it really is.



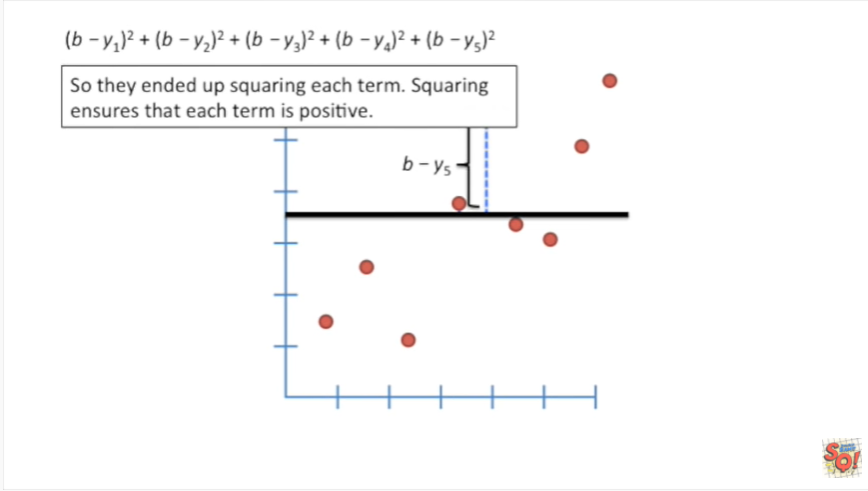
The fifth data point is even higher relative to the horizontal line.



This distance is going to be very negative.

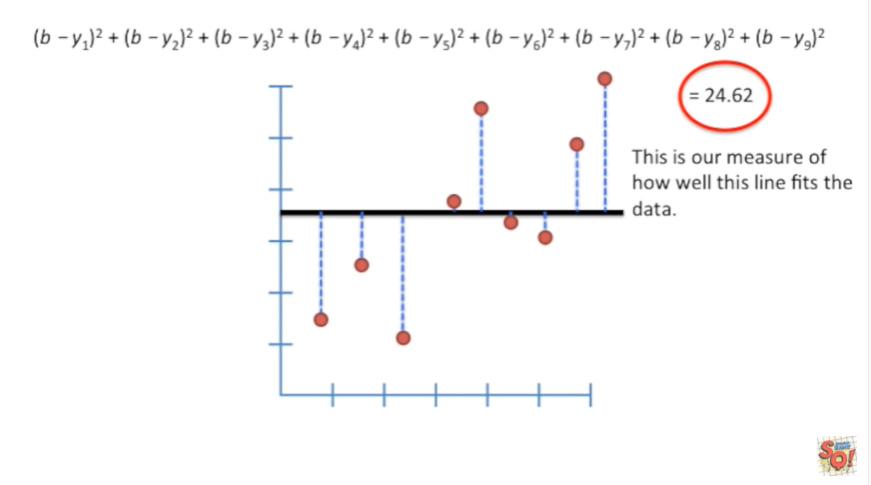


Back in the day when they were first working this out they probably tried taking the absolute value of everything and then discovered that it made the math pretty tricky.



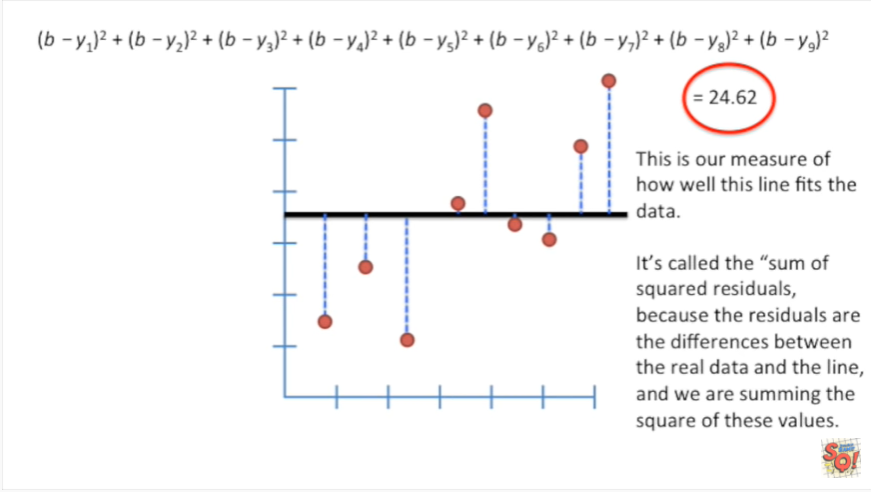
So they ended up squaring each term.

Squaring ensures that each term is positive.

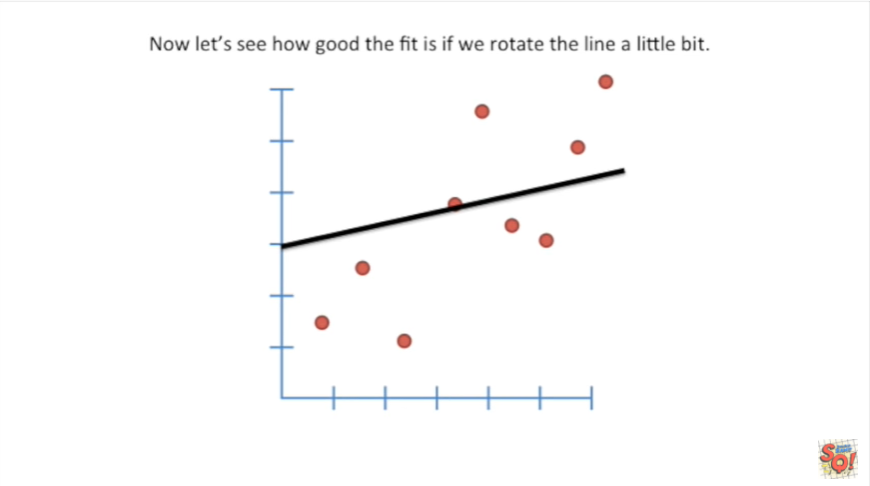


Here's the equation that shows the total distance the data points have from the horizontal line.

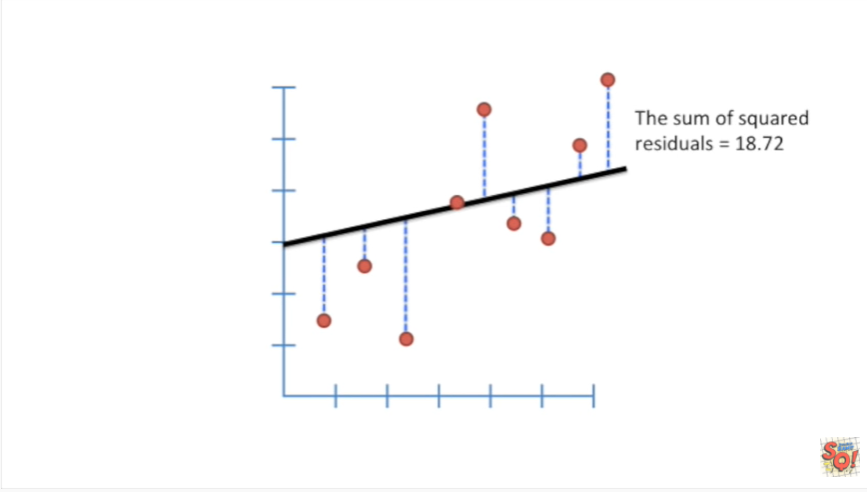
In this specific example twenty four point six two is our measure of how well this line fits the data.



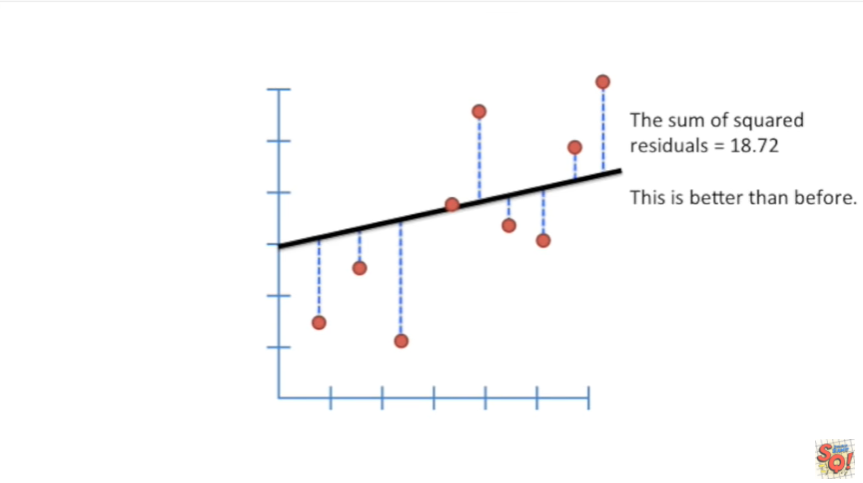
It's called the sum of squared residuals because the residuals are the differences between the real data and the line. And we are summing the square of these values.



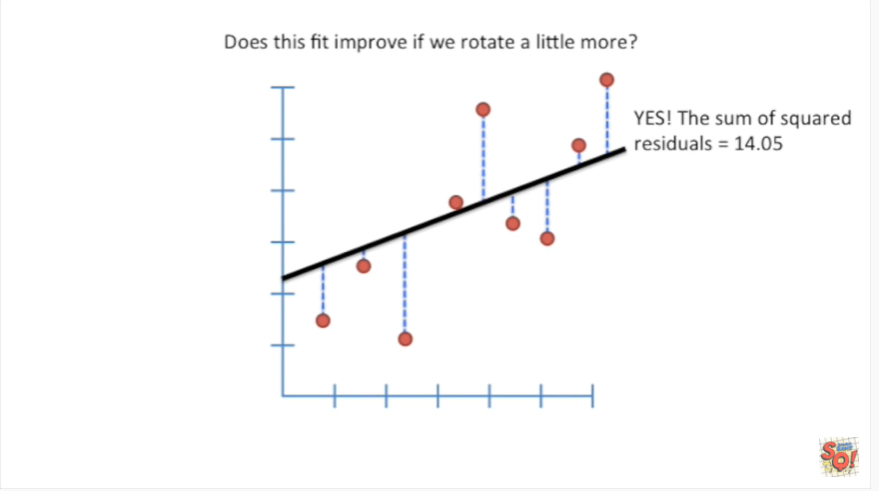
Now let's see how good the fit is if we rotate the line a little bit.



In this case the sum of squared residuals equals eighteen point seven two.



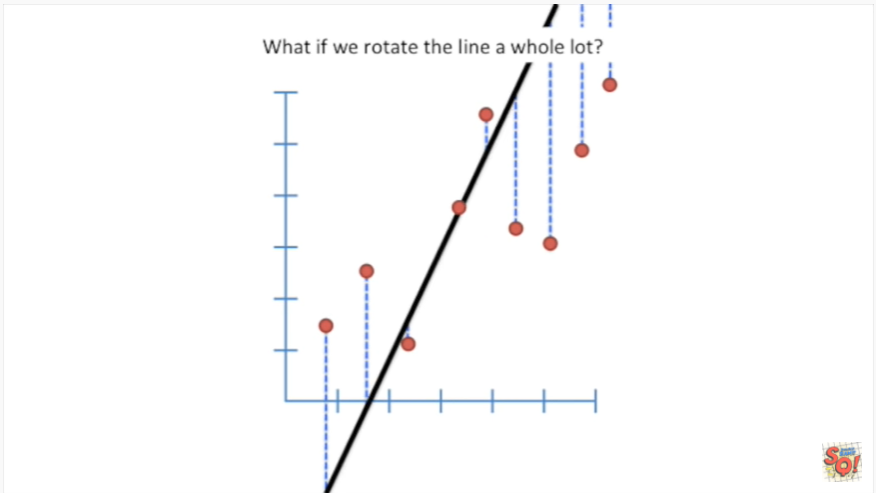
This is better than before.



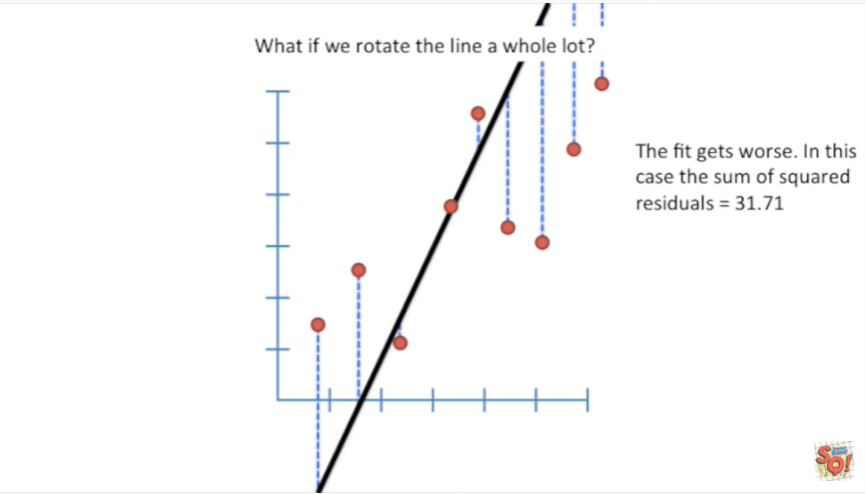
Does this fit improve if we rotate a little more ?

Yes the sum of squared residuals now equals 14 point O 5.

That value keeps going down the more we rotate the line.

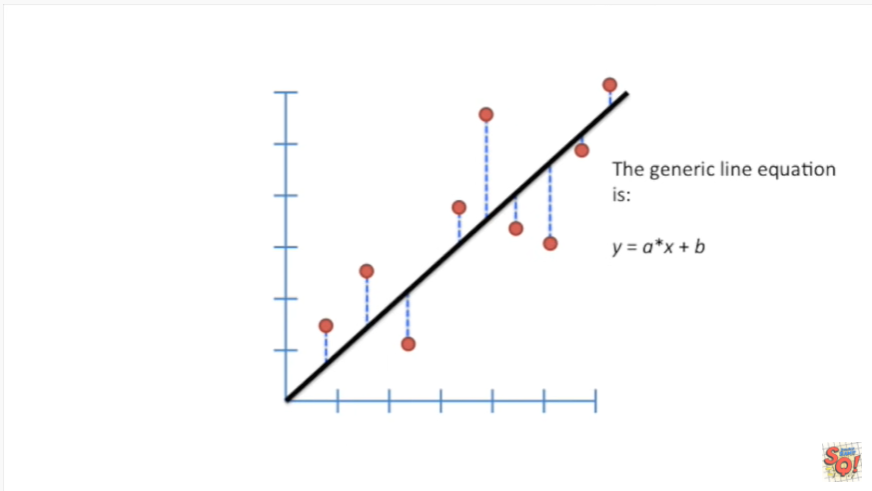


What if we rotate the line a whole lot ?



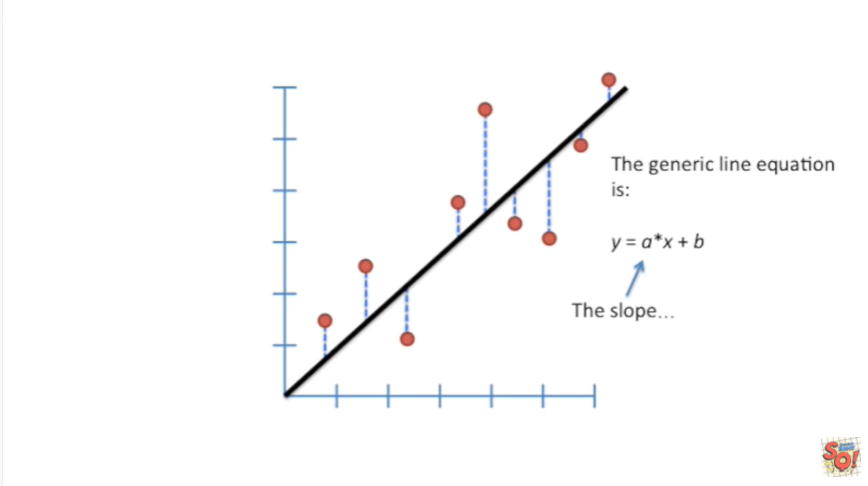
Well as you can see the fit gets worse in this case the sum of squared residuals is 31 point 7 1.

So there's a sweet spot in between horizontal and two vertical.

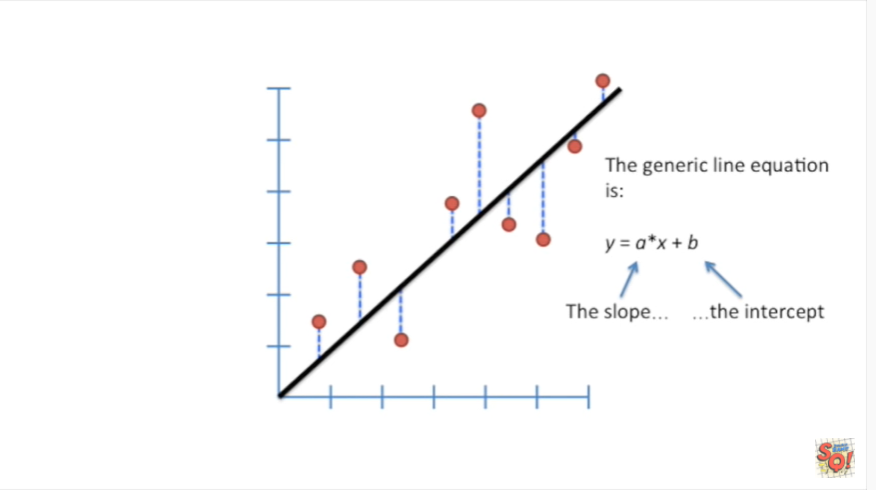


To find that sweet spot let's start with the generic line equation.

This is y equals ax or a times X plus B.

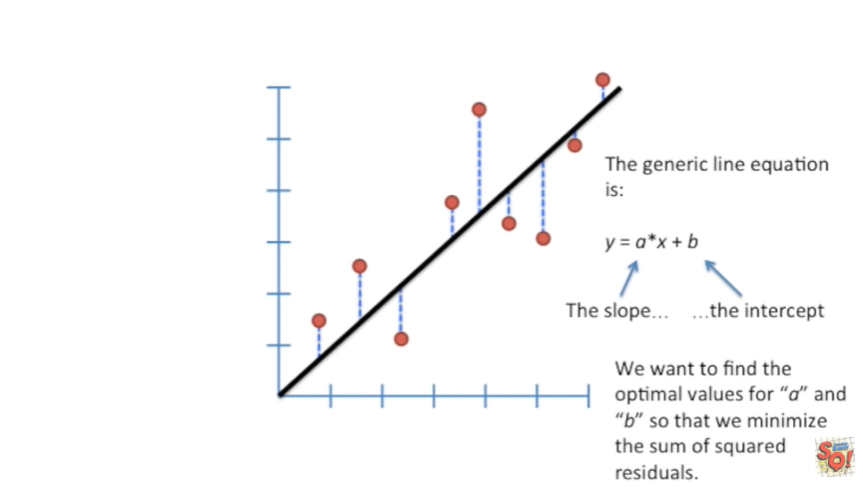


A is the slope of the line

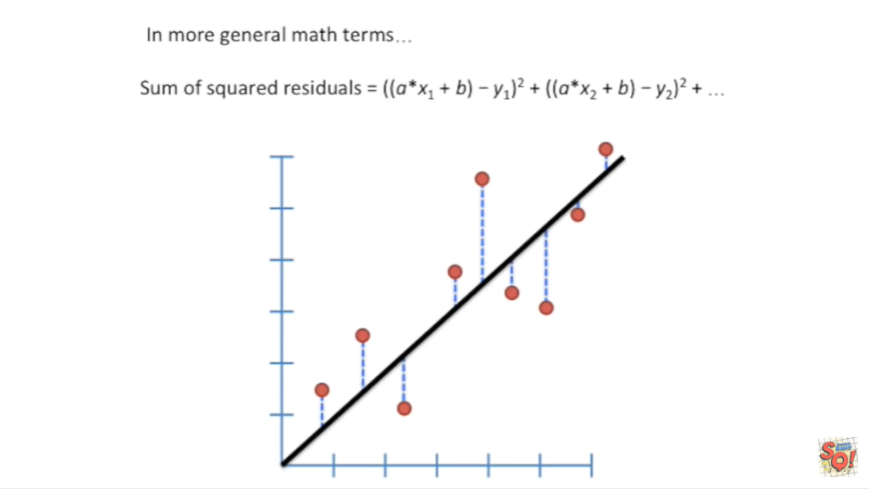


And B is the y-intercept of the line.

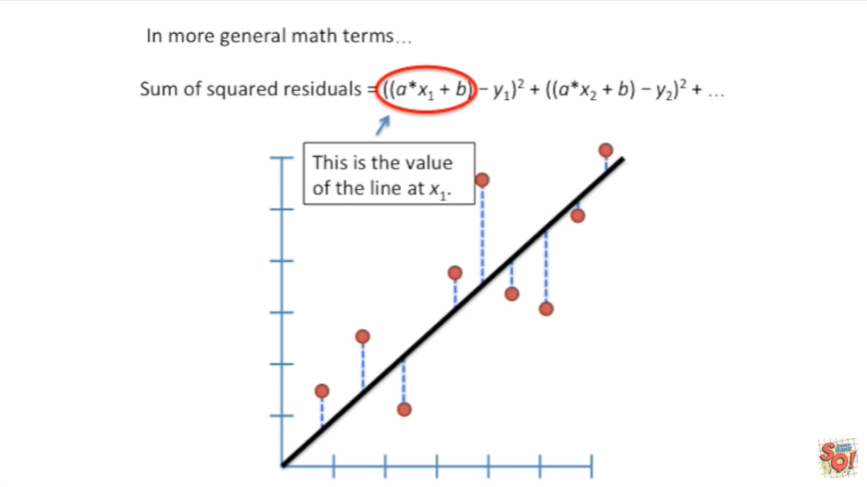
That's the location on the y axis that the line crosses when x equals 0.



We want to find the optimal values for a and B so that we minimize the sum of squared residuals.

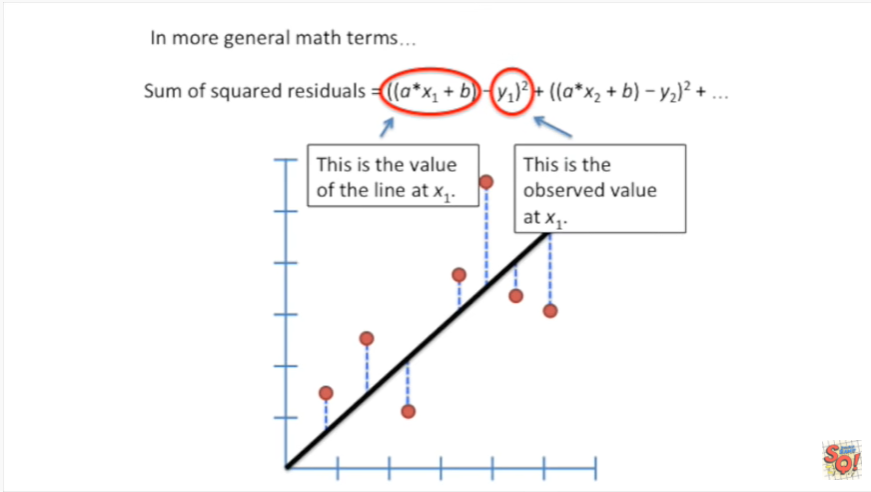


In more general math terms the sum of squared residuals is this complicated mathematical equation.



But it's actually not that complicated.

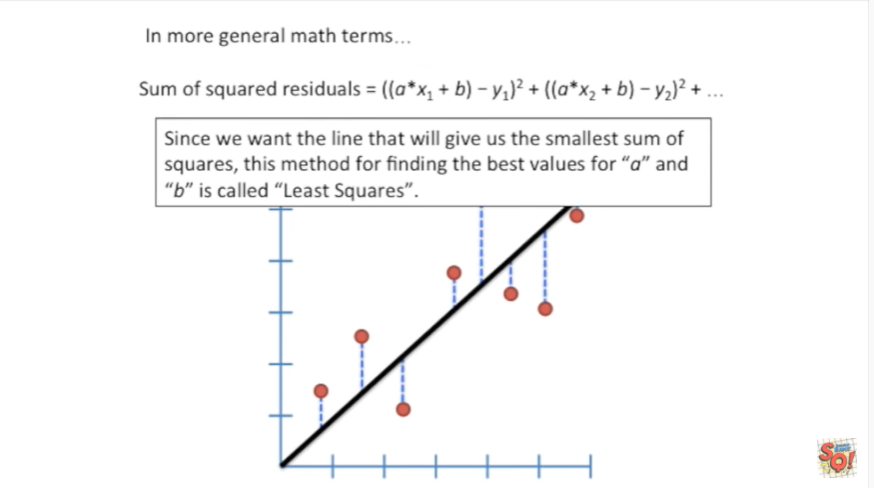
This first part is the value of the line at x1.



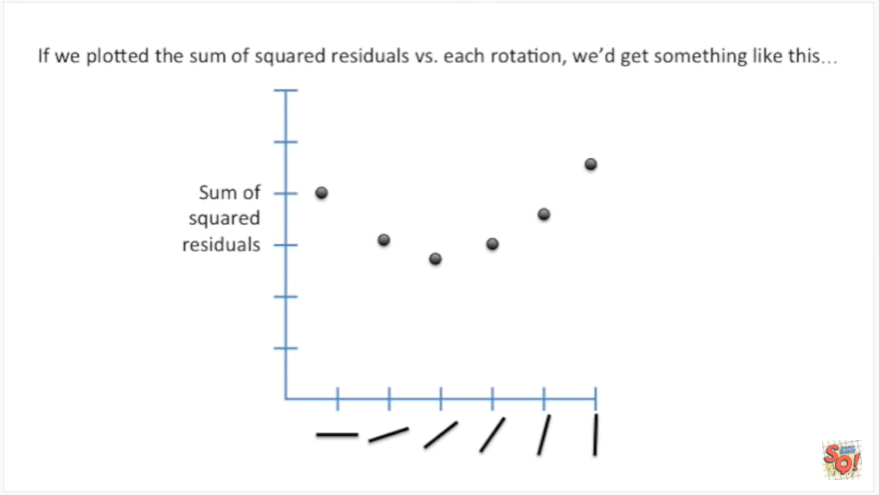
And this second part is the observed value at x1.

So really all we're doing in this part of the equation is calculating the distance between the line and the observed value.

So this is no big deal.



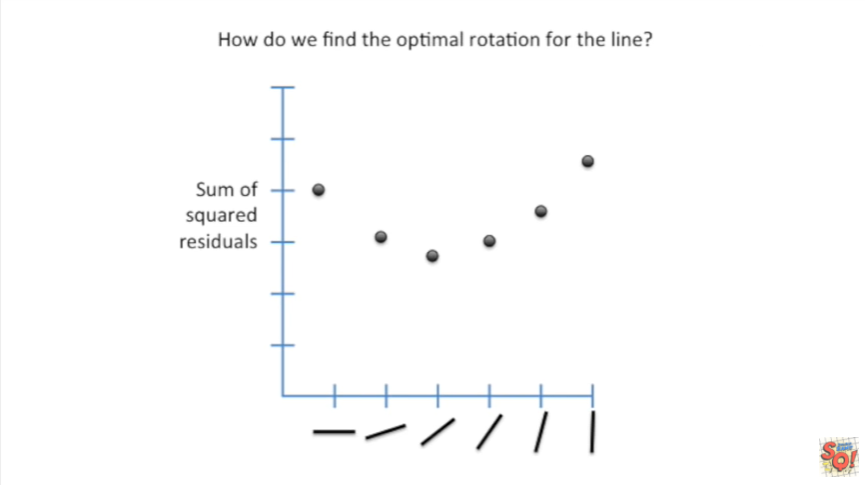
Since we want the line that will give us the smallest sum of squares, this method for finding the best values for a and B is called least squares.



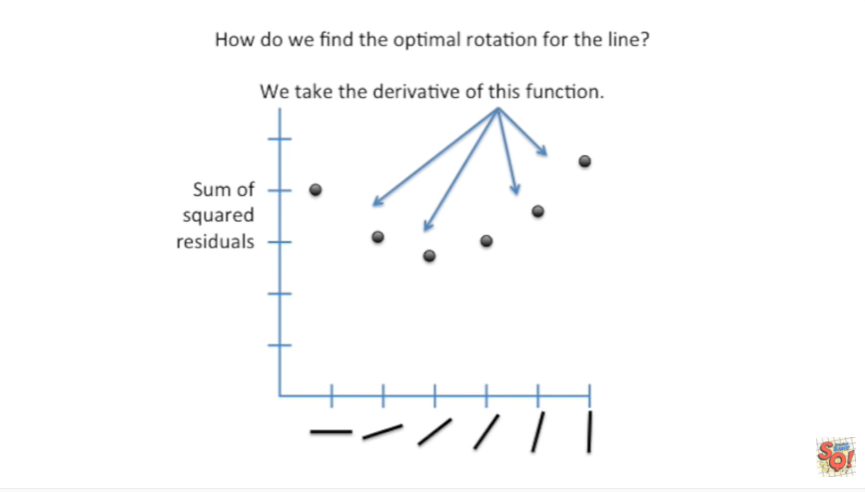
If we plotted the sum of squared residuals versus each rotation we get something like this.

Where on the y axis we have the sum of squared residuals and on the x axis we've got each different rotation of the line.

We see that the sum of squared residuals goes down when we start rotating the line, but that it's possible to rotate the line too far in the sum of squared residual starts going back up again.



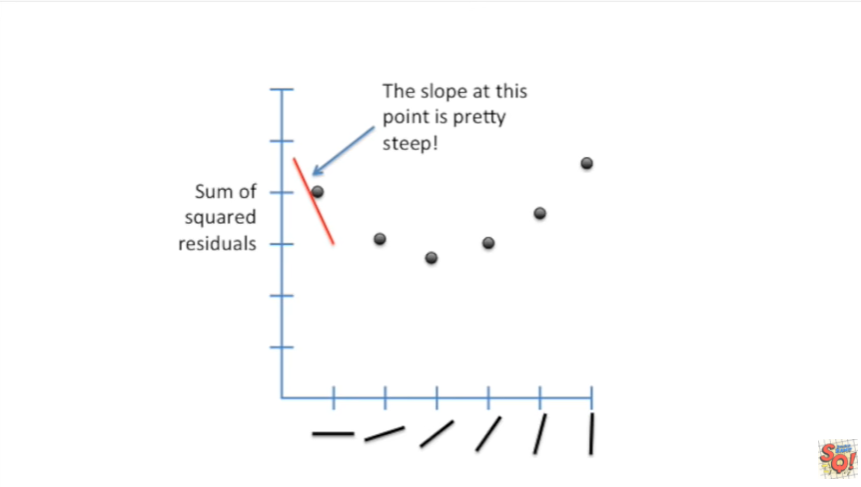
How do we find the optimal rotation for the line ?



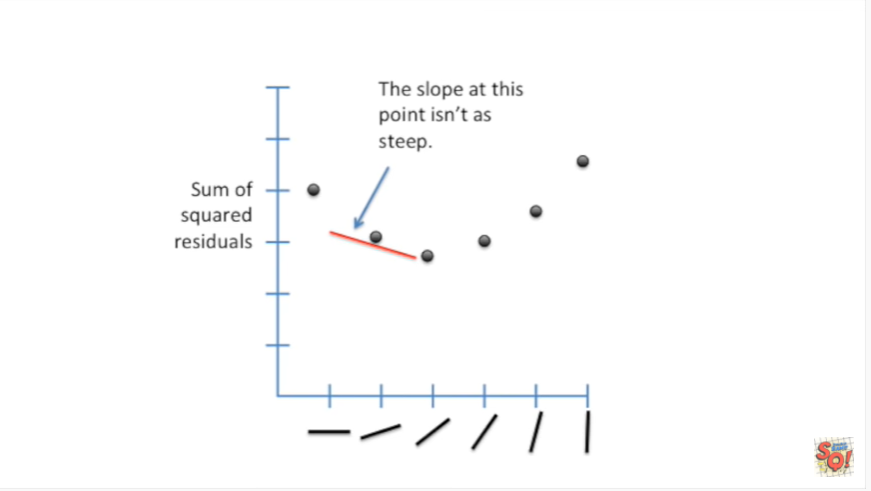
Well we take the derivative of this function.



The derivative tells us the slope of the function at every point.



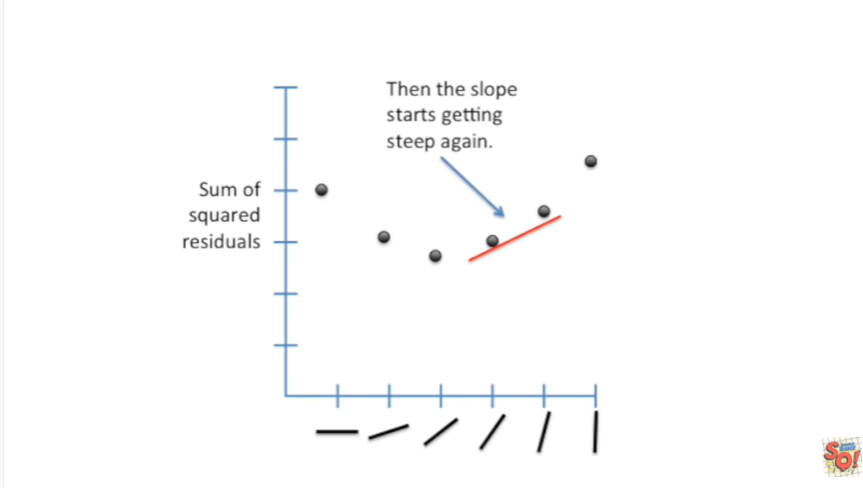
The slope at the point on the far left side is pretty steep.



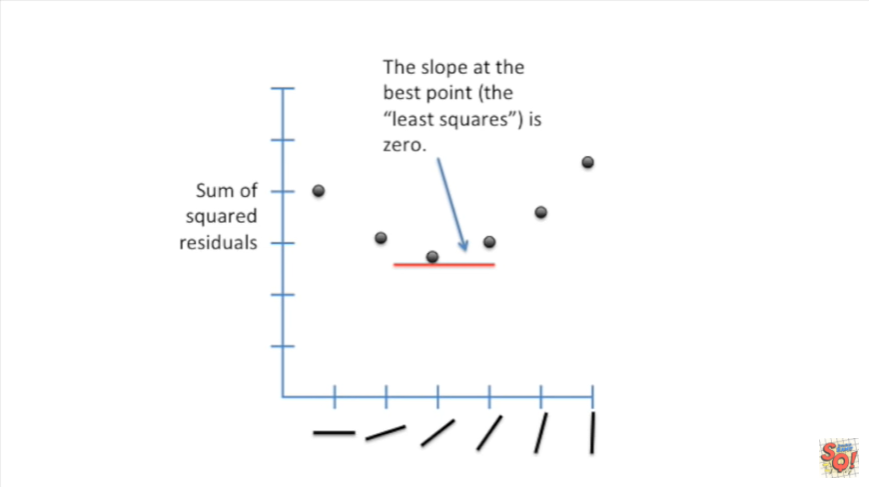
As we move to the right we see that the slope isn't as steep.



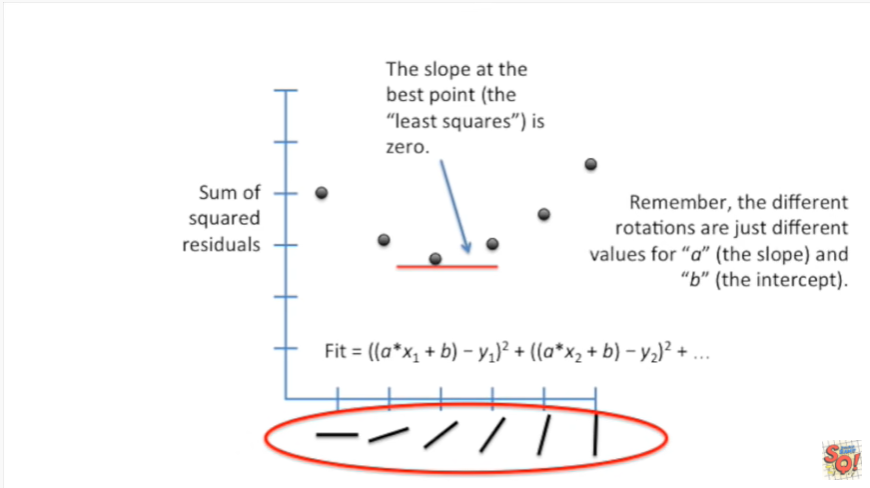
The slope at the best point where we have the least squares is zero.



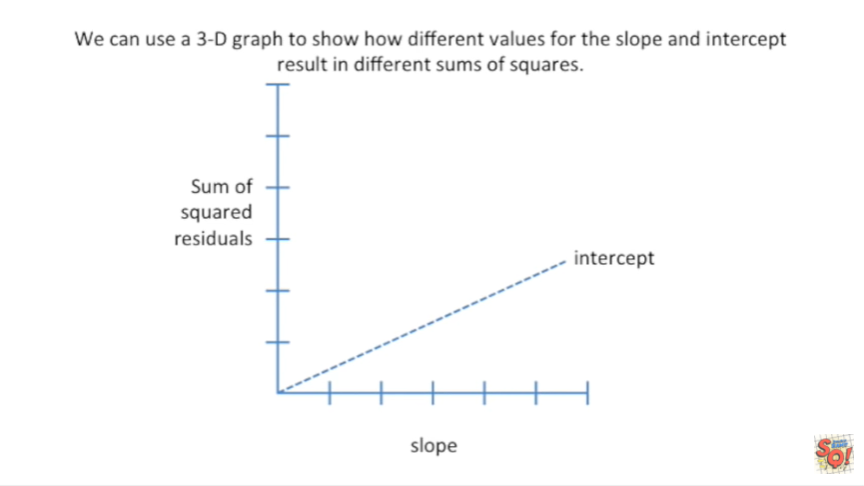
After that the slope starts getting steep again.



Let's go back to that middle point where we have the least squares value and the slope is zero.



Remember the different rotations are just different values for a the slope and B the intercept.



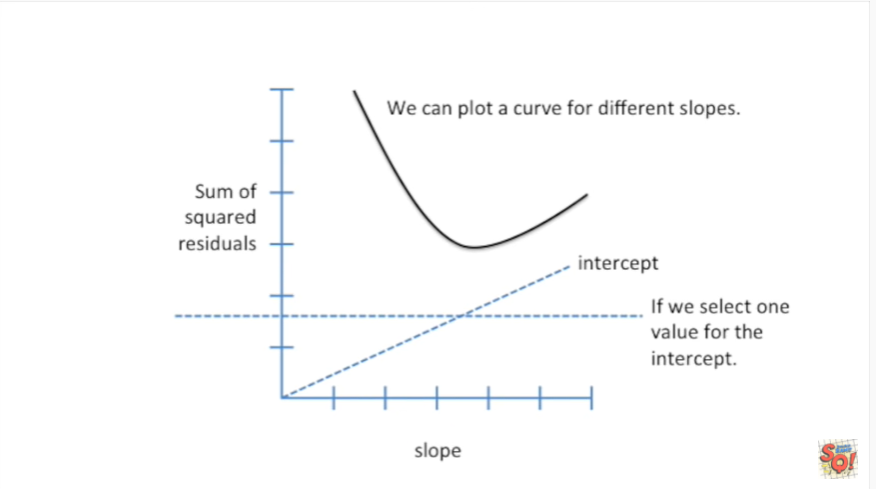
We can use a 3d graph to show how different values for the slope and intercept result in different sums of squares.



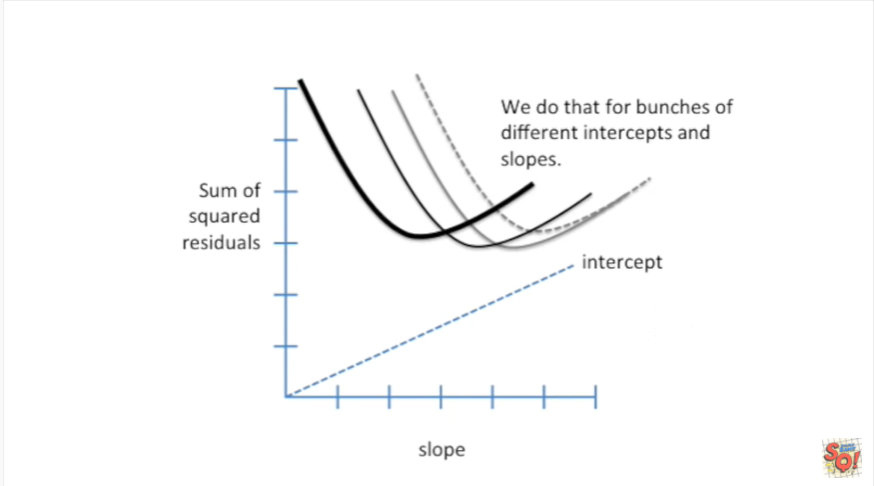
In this graph the intercept is the z axis so it's going back sort of deep into your computer screen.

And if we select one value for the intercept.

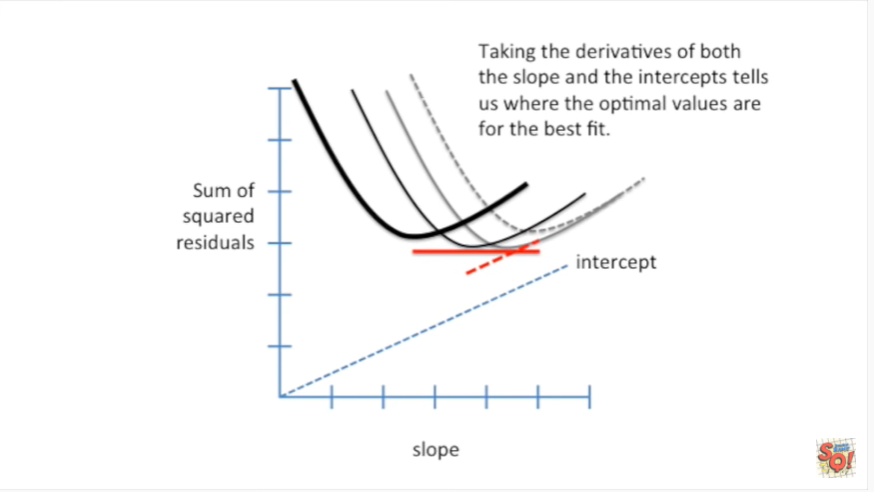
For example assume we set the intercept value to be three.



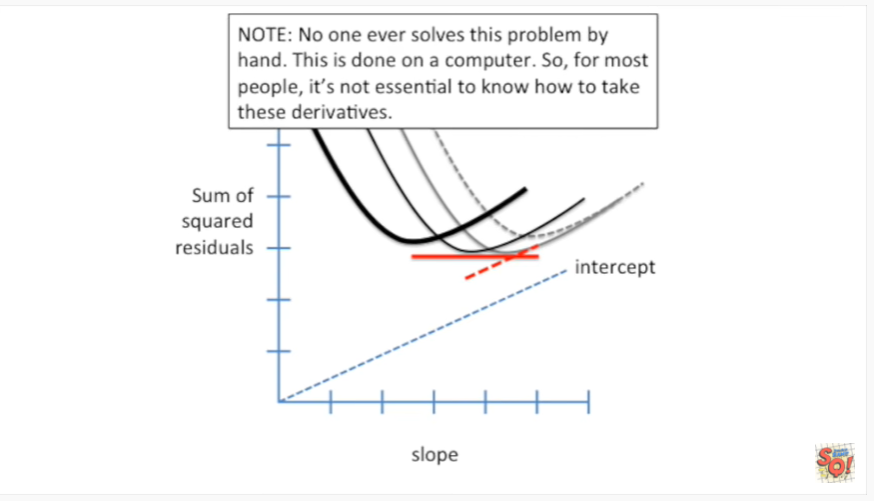
Then we could change values for the slope and see how an intercept of 3 plus different values for the slope would affect the sum of squared residuals.



Anyways we do that for bunches of different intercepts and slopes.



Taking the derivatives of both the slope and the intercepts tells us where the optimal values are for the best fit.

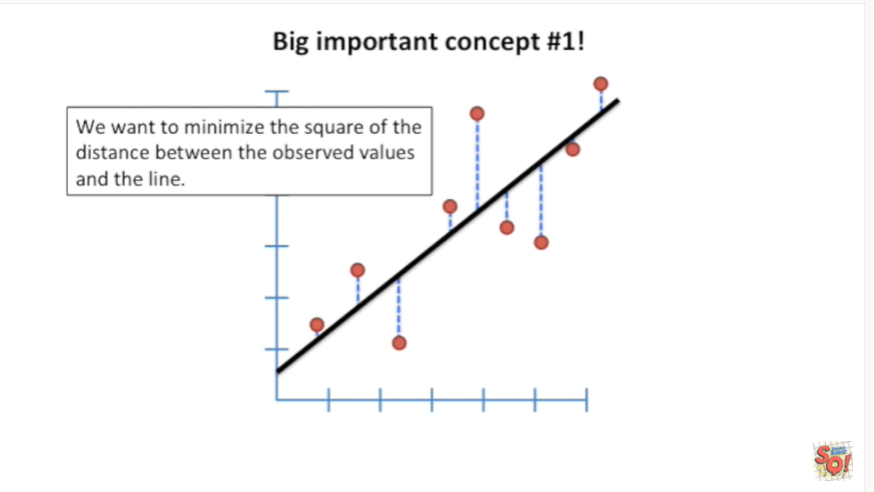


Note : no one ever solves this problem by hand this is done on a computer.

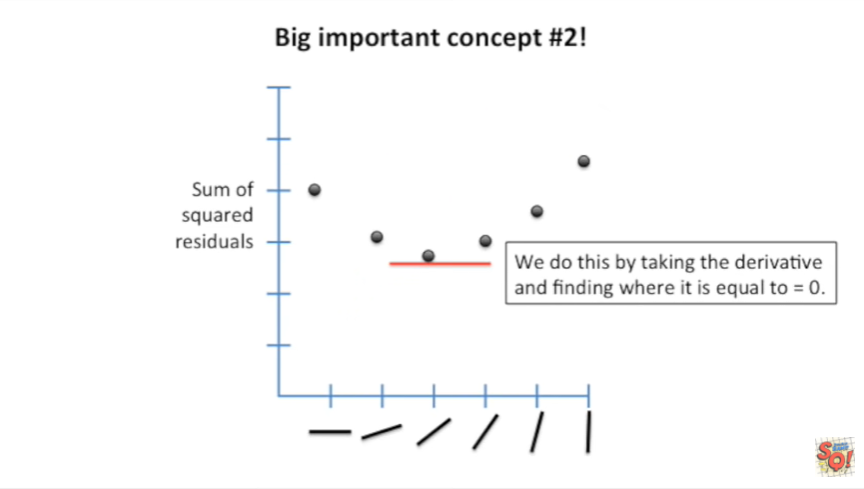
So for most people it's not essential to know how to take these derivatives.



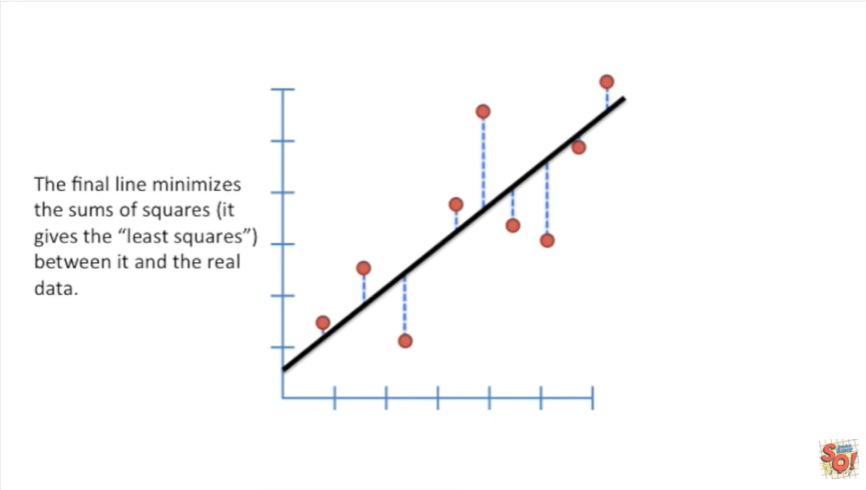
However it's essential to understand the concepts.



Big important concept number one : we want to minimize the square of the distance between the observed values and the line.

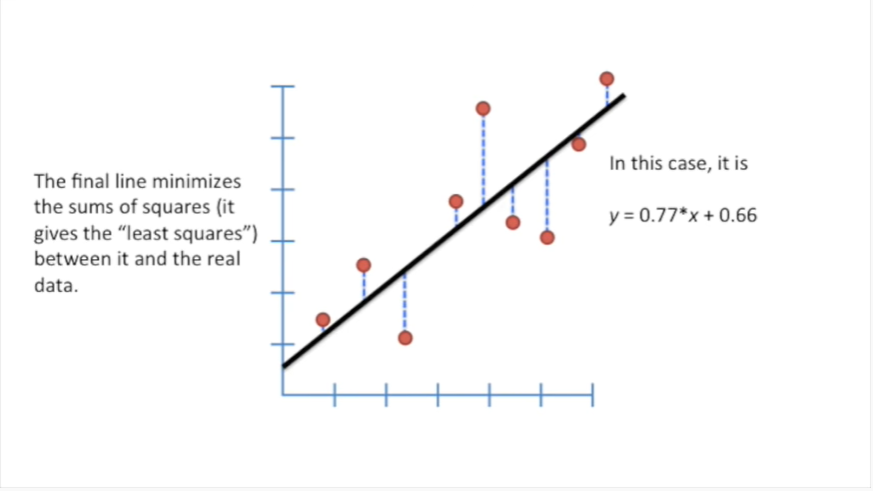


Big important concept number : two we do this by taking the derivative and finding where it is equal to zero.



The final line minimizes the sums of squares.

It gives the least squares between it and the real data.



In this case the line is defined by the following equation : y equals zero point seven seven times X plus zero point six six.

Hooray we've made it to the end of another stat quest.

Tune in next time for another exciting adventure in statistics land.