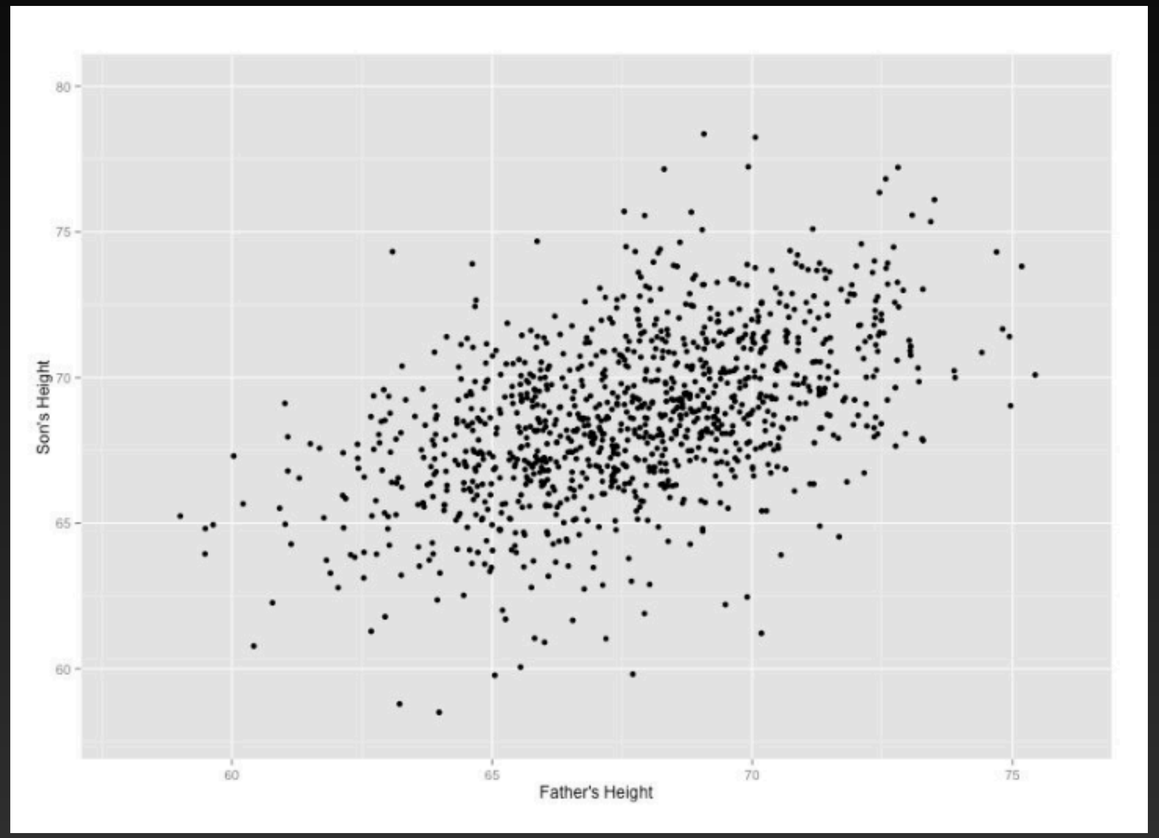
Introduction to Linear Regression

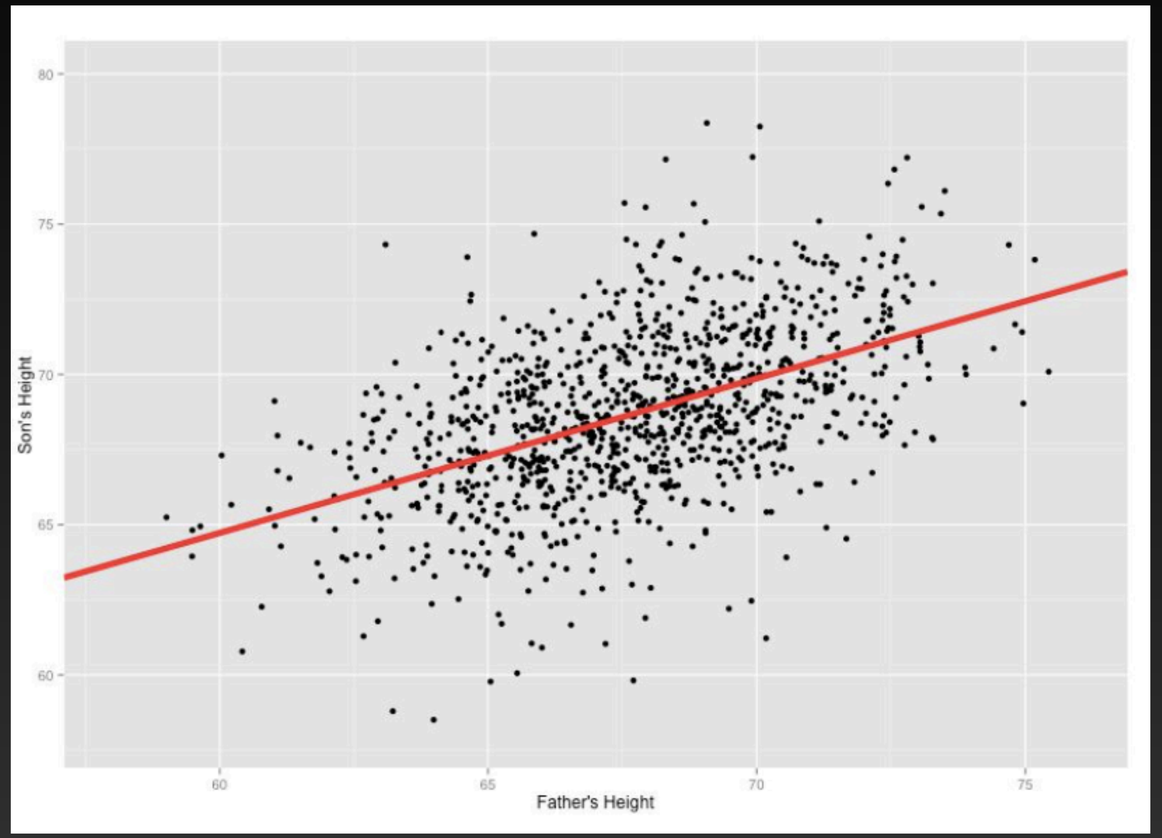
The Goal with Linear Regression

For example, we are plotting some data points here with father’s height on X-axis and son’s height on Y-axis and we want to predict the son’s height on the basis of the father’s height



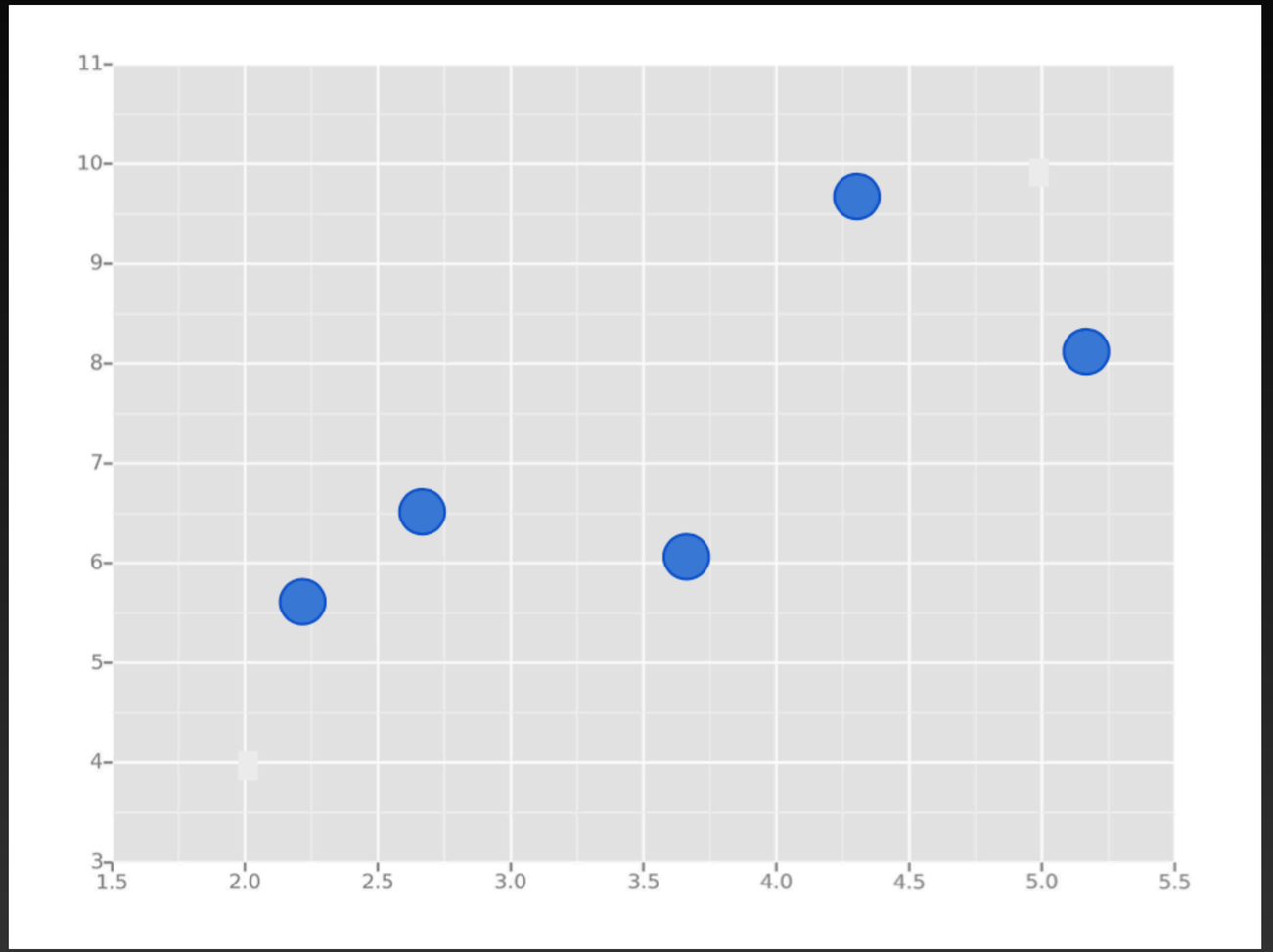
Our goal is to minimise the vertical distance between all the data points and our line

So, in determining the best line, we are attempting to minimise the distance between all the data points and our linear regression line



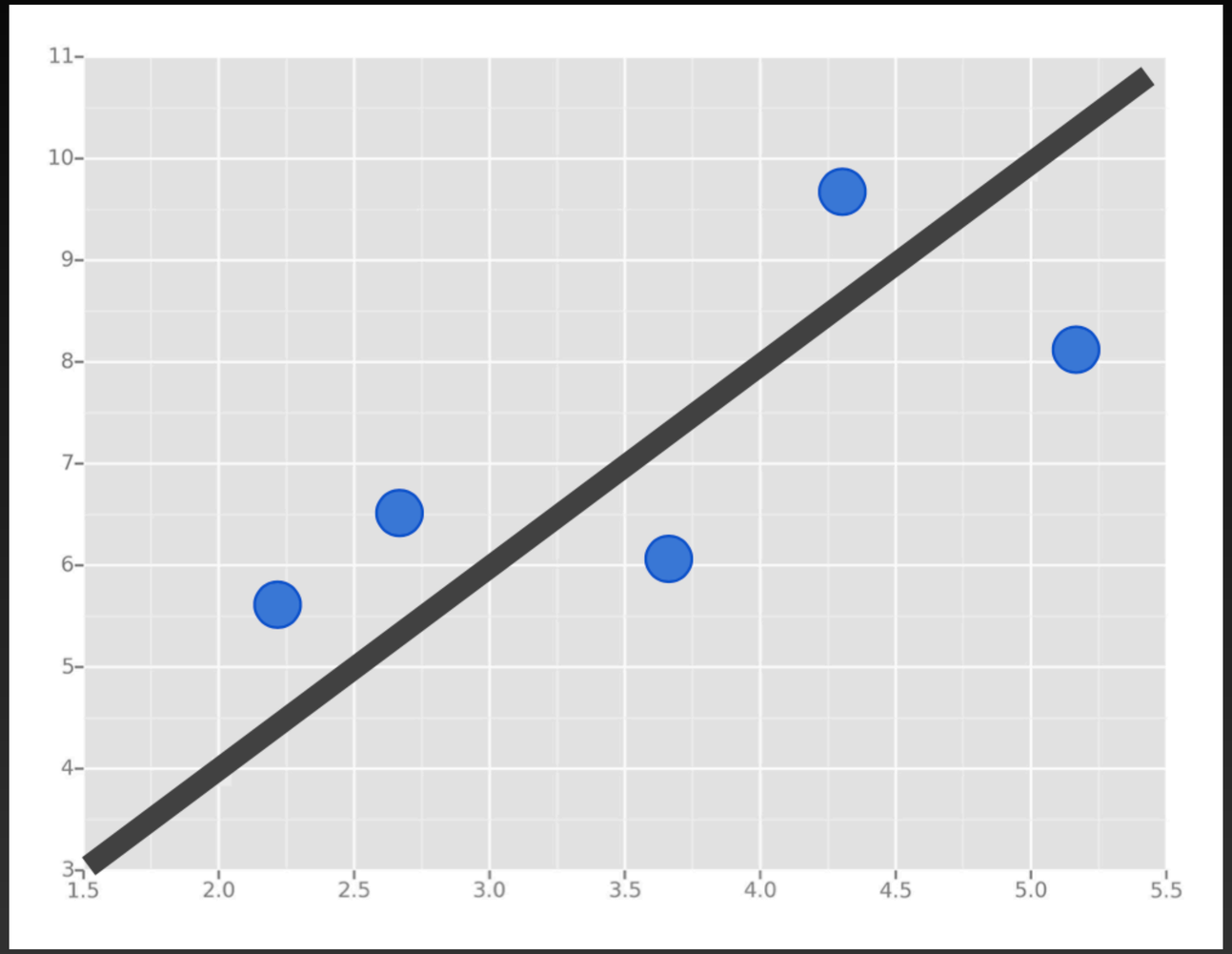
A Simpler Example

Here, we have four data points

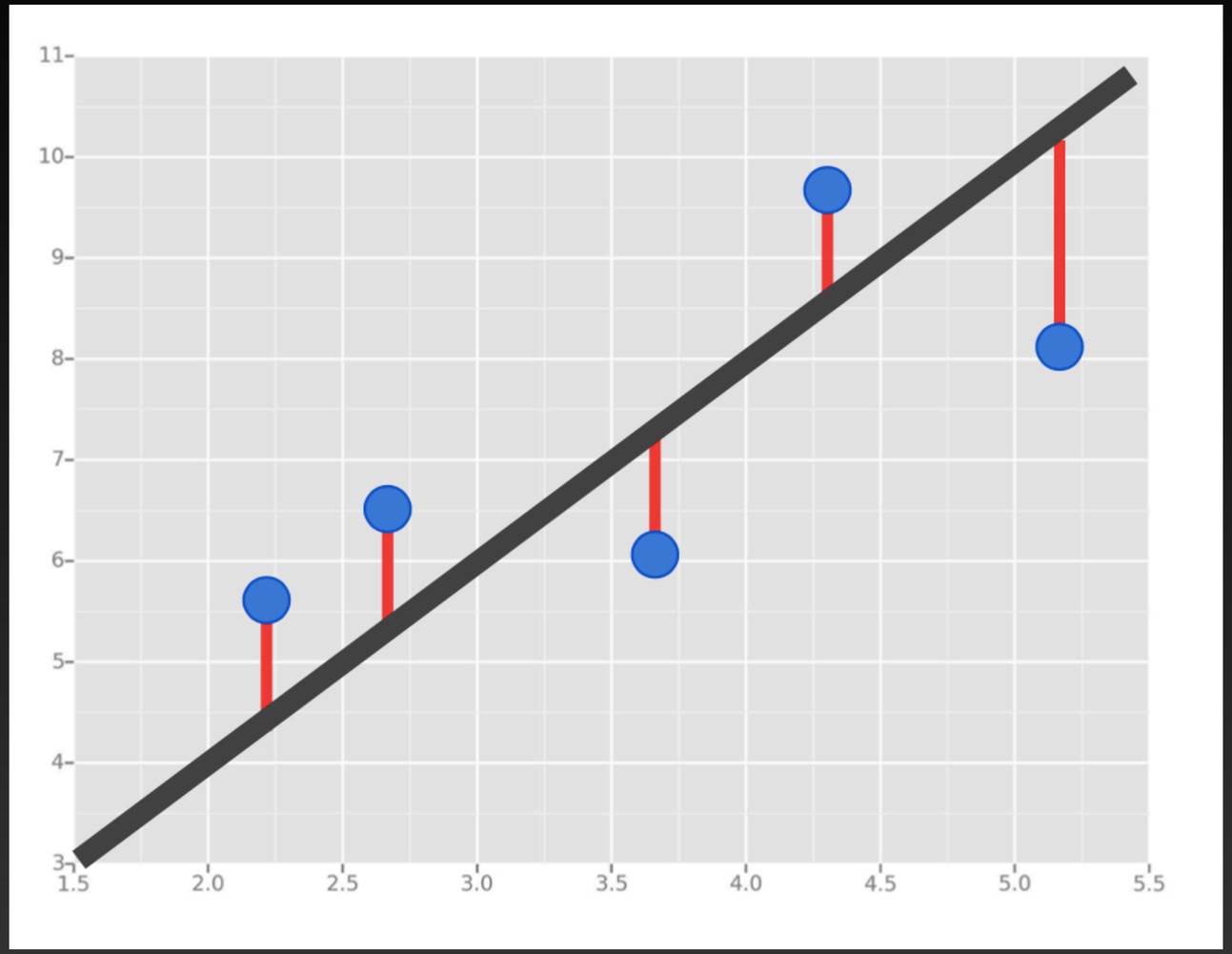


Now we want to fit a linear regression here

How do we decide which line is the most suitable?



Here we use the Least Squares method which is fitted by minimising the sum of squares of the residuals



Linear Equation

Linear Regression Equation

$y = \alpha_0 + \alpha_1 x_1$

OR

$y = mx + c$

where:

x is the independent variable

m is the slope of the line

c is the intercept

and y is the dependent variable