

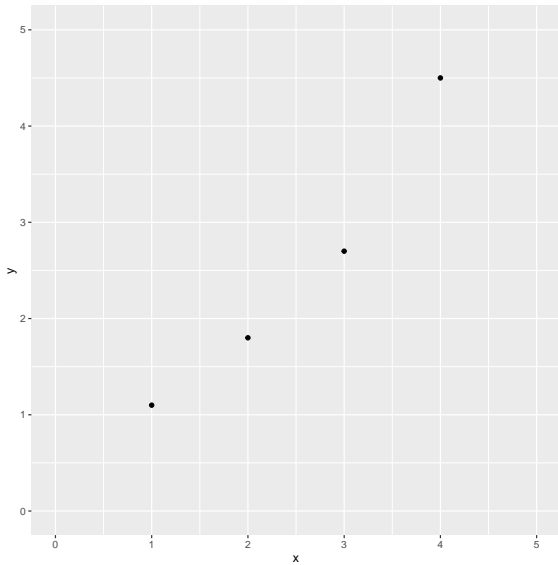
Principal Component Analysis

Simple Linear Model

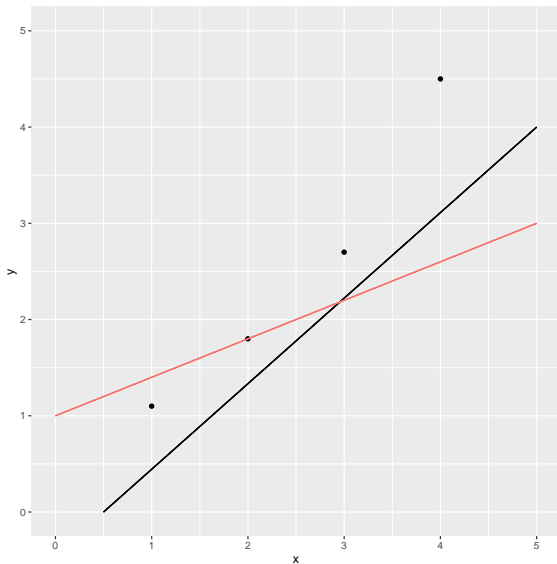
► Given the data

x_1	x_2	y
1	0	-2
2	1	0
3	-2	-1
4	3	1

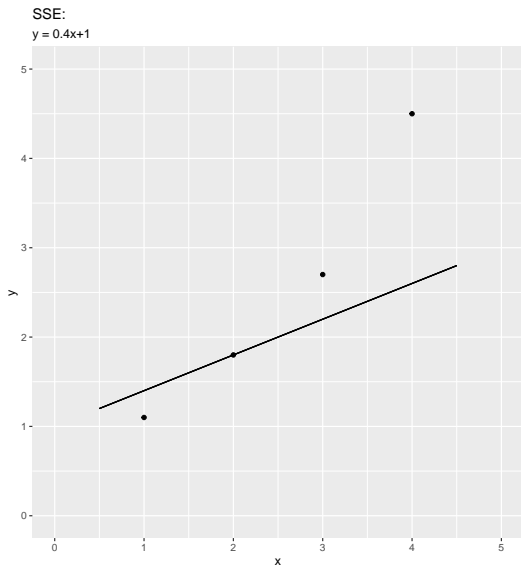
Scatter plot



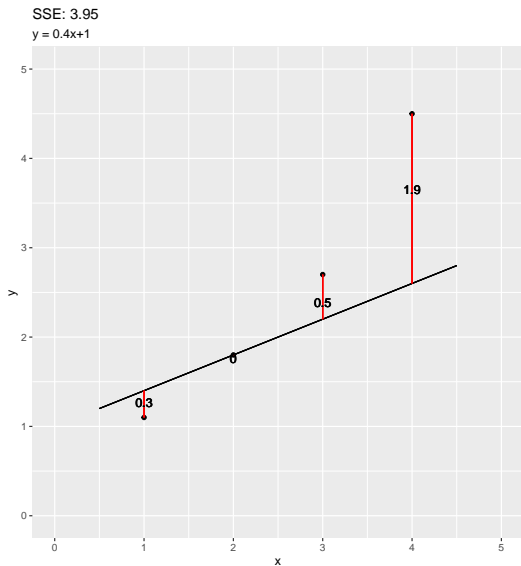
Which line is closer to the points?



Line 1



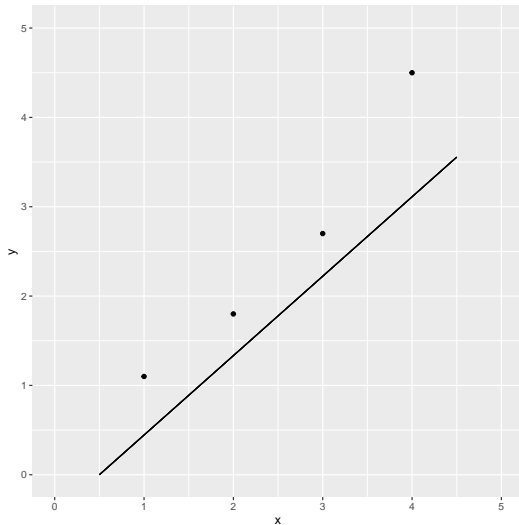
Line 1



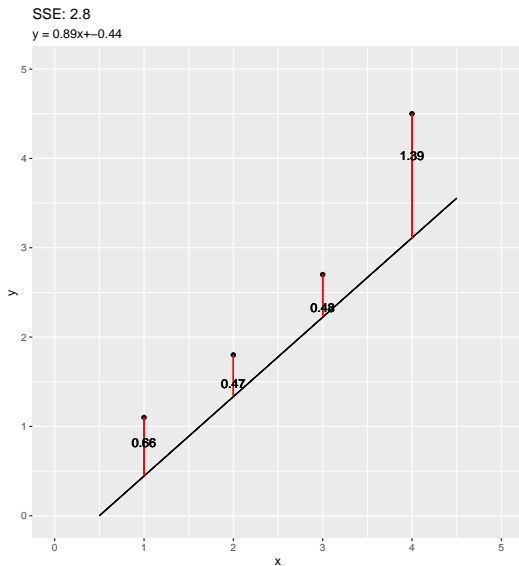
Line 2

SSE:

$$y = 0.89x + -0.44$$



Line 2



What is the closest line to the points?

- ▶ The least squared methods give us the formula for the closest line:
- ▶ $y = \hat{\beta}_1 x + \hat{\beta}_0$
- ▶
$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$$
- ▶
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$
- ▶ This line is also called the best fitted line

Calculation

x	y
1	1.1
2	1.8
3	2.7
4	4.5

Calculation

x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$	$(x - \bar{x})^2$
1	1.1				
2	1.8				
3	2.7				
4	4.5				

Calculation

x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$	$(x - \bar{x})^2$
1	1.1				
2	1.8				
3	2.7				
4	4.5				

► $\bar{x} = \frac{1+2+3+4}{4} = 2.5$

► $\bar{y} = \frac{1.1+1.8+2.4+4.5}{4} = 1.11$

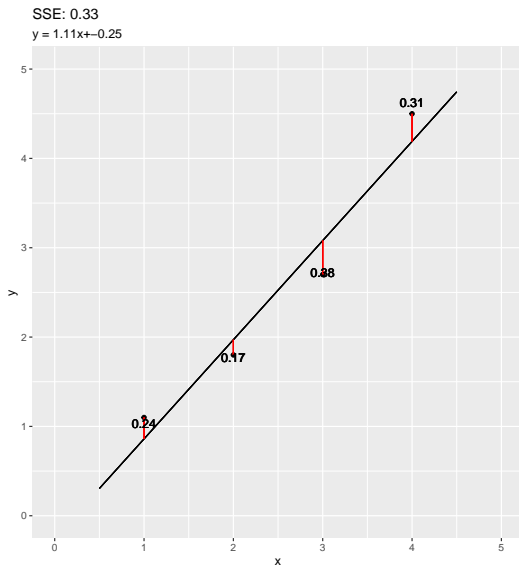
Calculation

x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$	$(x - \bar{x})^2$
1	1.1	-1.5	-1.425	2.1375	2.25
2	1.8	-0.5	-0.725	0.3625	0.25
3	2.7	0.5	0.175	0.0875	0.25
4	4.5	1.5	1.975	2.9625	2.25
				5.55	5

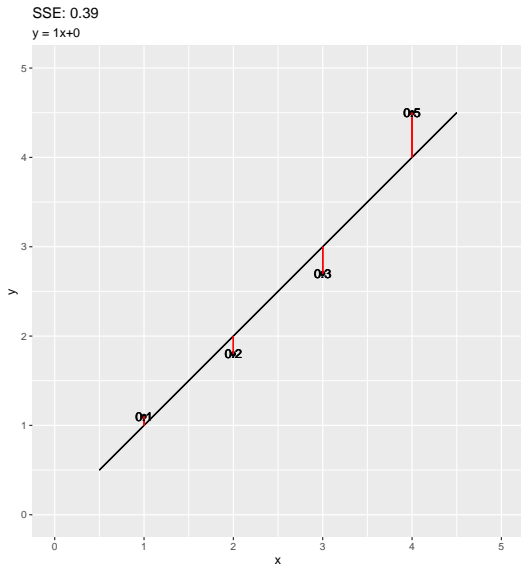
► $\hat{\beta}_1 = \frac{5.55}{5} = 1.11$

► $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = -0.25$

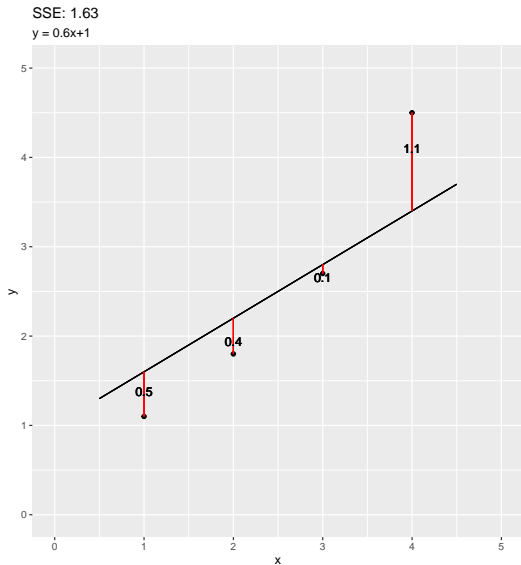
Graph



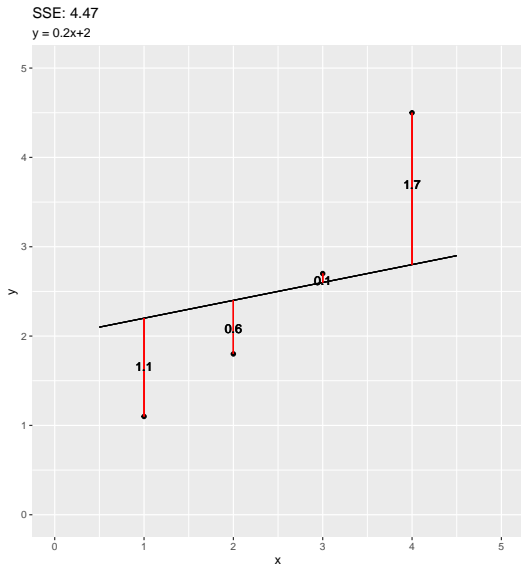
Some other lines



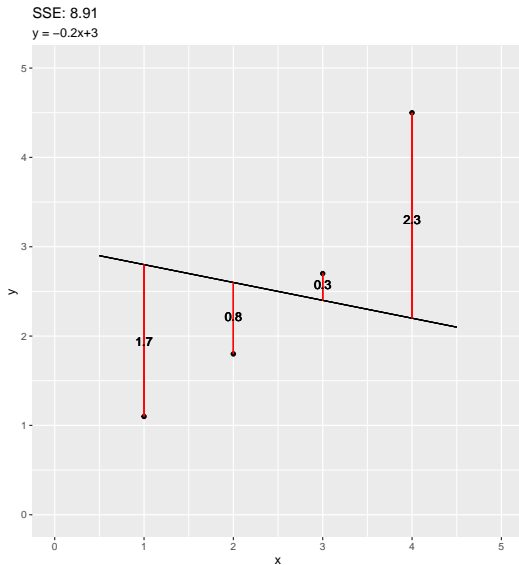
Some other lines



Some other lines



Some other lines



Sum Up

- ▶ The best fitted line or the least squared line is the line the closest to the data point.