

Scatter Diagrams and Linear Correlation

Example 1:

Is there a connection between the average weight of watermelons a vine produces and the root depth of the vine?

Suspicion: vines with deeper roots have a better water supply, and thus larger average melons.

Large watermelon field, 15 vines are chosen at random. At the end of 8 weeks the watermelons are removed from each vine, weighed, and then the average weight (in pounds) is determined. Root depth of each vine is measured (in inches).

Plot a scatter diagram.

<i>Root depth:</i>	26	14	18	10	26	21	7	26	13	19	17	13	16	28	23
<i>Mean weight:</i>	20	10	13	9	19	17	8	15	9	13	12	7	9	17	14

1. What should be x and what should be y ?

x – explanatory variable

y – response variable

y depends on x

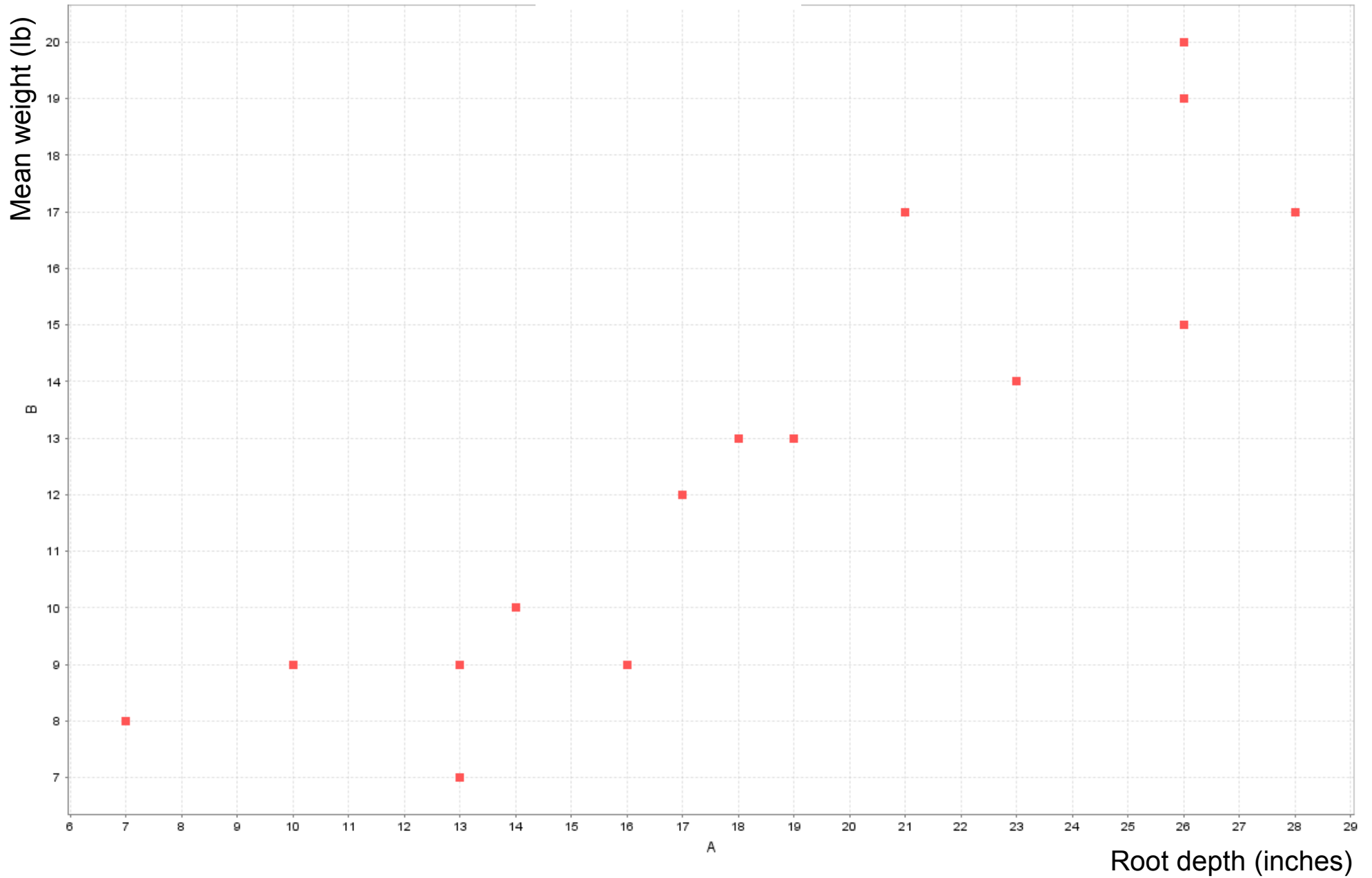
see suspicion: weight depends on root depth

thus x – *root depth*, y – *mean weight*

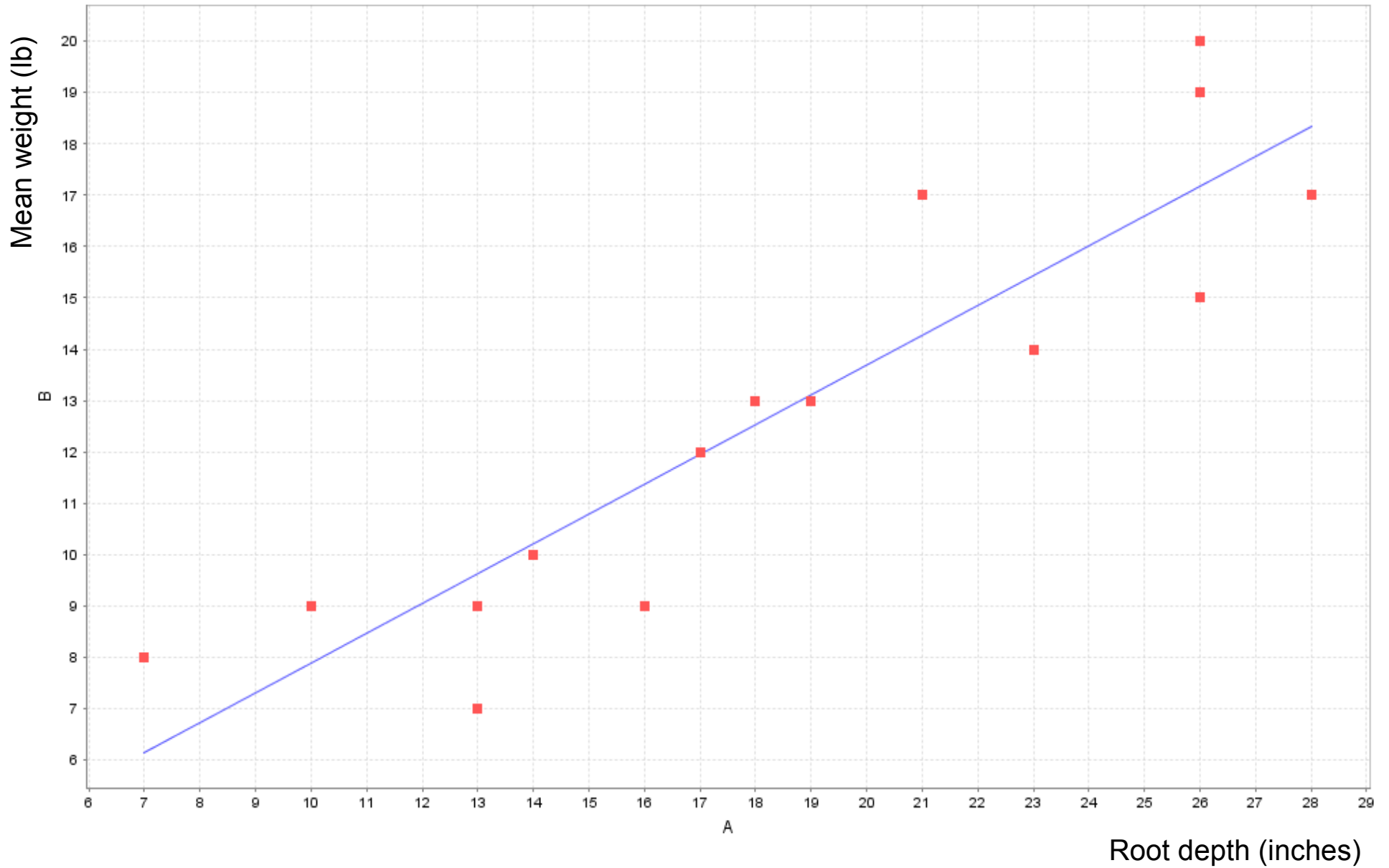
for each plant there is an ordered pair (x,y) : (*root depth*, *mean weight*)

2. Let's plot the points (plot scatter diagram):

Root depth (x): 26 14 18 10 26 21 7 26 13 19 17 13 16 28 23
Mean weight (y): 20 10 13 9 19 17 8 15 9 13 12 7 9 17 14



3. Can we draw a straight line, so that the points are close to it?



Scatter Diagrams and Linear Correlation

4. Let's find correlation coefficient r .

Root depth: 26 14 18 10 26 21 7 26 13 19 17 13 16 28 23
Mean weight: 20 10 13 9 19 17 8 15 9 13 12 7 9 17 14

X	Y	X ²	Y ²	XY
26	20	676	400	520
14	10	196	100	140
18	13	324	169	234
10	9	100	81	90
26	19	676	361	494
21	17	441	289	357
7	8	49	64	56
26	15	676	225	390
13	9	169	81	117
19	13	361	169	247
17	12	289	144	204
13	7	169	49	91
16	9	256	81	144
28	17	784	289	476
23	14	529	196	322
$\Sigma x=277$	$\Sigma y=192$	$\Sigma x^2=5695$	$\Sigma y^2=2698$	$\Sigma xy=3882$

Scatter Diagrams and Linear Correlation

4. Let's find correlation coefficient r :

X	Y	X ²	Y ²	XY
26	20	676	400	520
14	10	196	100	140
Skipped data values				
13	7	169	49	91
16	9	256	81	144
28	17	784	289	476
23	14	529	196	322
$\Sigma x=277$	$\Sigma y=192$	$\Sigma x^2=5695$	$\Sigma y^2=2698$	$\Sigma xy=3882$

$$\begin{aligned}
 r &= \frac{(n \sum xy - (\sum x)(\sum y))}{(\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2})} = \frac{(15 * 3882 - 277 * 192)}{(\sqrt{15 * 5695 - 277^2} \sqrt{15 * 2698 - 192^2})} = \\
 &= \frac{5046}{(\sqrt{8696} \sqrt{3606})} \approx \frac{5046}{(93.25 * 60.05)} \approx 0.901
 \end{aligned}$$

$r \approx 0.901$ – there is a strong positive linear correlation between the root length and the watermelon weight