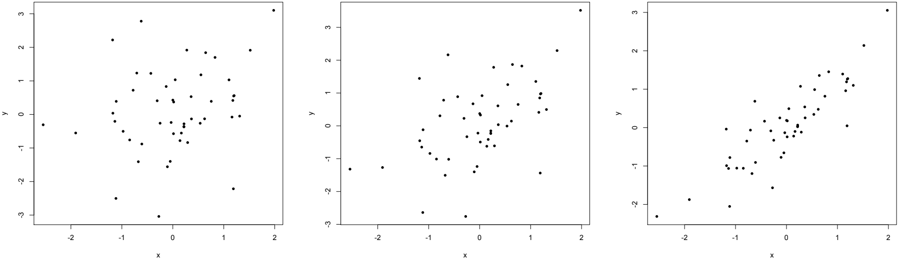
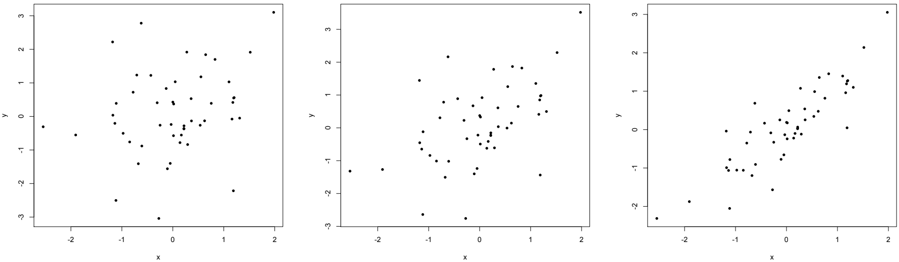
How linear?

How linear?



How linear?



correlation coefficient (r): a number between -1 and 1; it measures **linear association,** that is, how tightly the points are clustered about a straight line.

Example

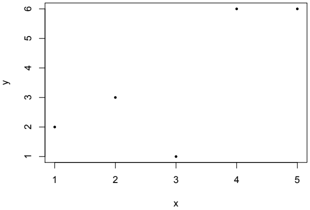
Data:

(1, 2)
(2, 3)
(3, 1)
(4, 6)
(5, 6)

Example

Data:

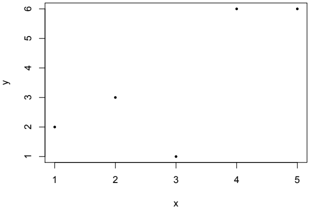
(1, 2)
(2, 3)
(3, 1)
(4, 6)
(5, 6)



Example

Data:

(1, 2)
(2, 3)
(3, 1)
(4, 6)
(5, 6)



Expect r to be positive but not 1.

Calculating r

Calculating r

x	y	x in std. units	y in std. units	**product of std. units**
1	2	-1.41	-0.78	1.10
2	3	-0.71	-0.29	0.21
3	1	0	-1.26	0
4	6	0.71	1.16	0.82
5	6	1.41	1.16	1.64
mean = 3	mean = 3.6			**mean = 0.75**
SD = 1.41	SD = 2.06			**= r**

The formula in two languages

Formula for r

1. Convert both lists to standard units.
2. Multiply corresponding pairs of standard units.
3. r is the average of the products.

The formula in two languages

Formula for r

1. Convert both lists to standard units.
2. Multiply corresponding pairs of standard units.
3. r is the average of the products.

For those who like math notation and have read the algebra supplement:

If the data are (x_i, y_i), $1 \leq i \leq n$, then

$$r = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i - \mu_x}{\sigma_x} \right) \left(\frac{y_i - \mu_y}{\sigma_y} \right)$$

Properties of r

1. The calculation uses only standard units. So r **is a pure number** with no units.

Properties of r

1. The calculation uses only standard units. So r **is a pure number** with no units.

2. $\mathbf{-1 \leq r \leq 1}$. Trust me.

Properties of r

1. The calculation uses only standard units. So r **is a pure number** with no units.

2. $\mathbf{-1 \leq r \leq 1}$. Trust me.
The extreme cases: $r = -1$ is when the scatter is a perfect straight line sloping down; $r = 1$ is when the scatter diagram is a perfect straight line sloping up.

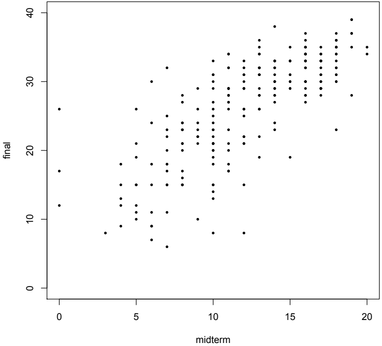
Properties of r

1. The calculation uses only standard units. So r **is a pure number** with no units.

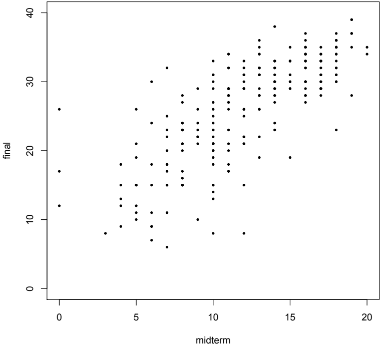
2. $\mathbf{-1 \leq r \leq 1}$. Trust me.
The extreme cases: $r = -1$ is when the scatter is a perfect straight line sloping down; $r = 1$ is when the scatter diagram is a perfect straight line sloping up.

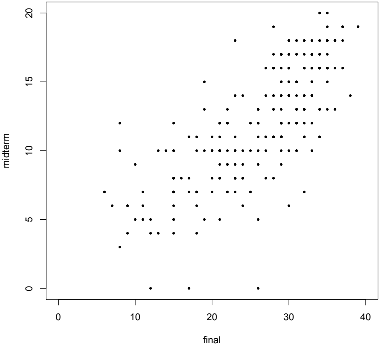
3. It doesn't matter if you **switch the variables x and y; r stays the same.**

Switching axes doesn't affect linearity



Switching axes doesn't affect linearity





Linear transformations

4. **Adding a constant** to one of the lists just slides the scatter diagram, so r **stays the same.**

Linear transformations

4. **Adding a constant** to one of the lists just slides the scatter diagram, so r **stays the same.**

5. **Multiplying one the lists by a positive constant** does not change standard units, so r **stays the same.**

Linear transformations

4. **Adding a constant** to one of the lists just slides the scatter diagram, so r **stays the same.**

5. **Multiplying one the lists by a positive constant** does not change standard units, so r **stays the same.**

6. **Multiplying just one (not both) of the lists by a negative constant** switches the signs of the standard units of that variable, so r **has the same absolute value but its sign gets switched.**