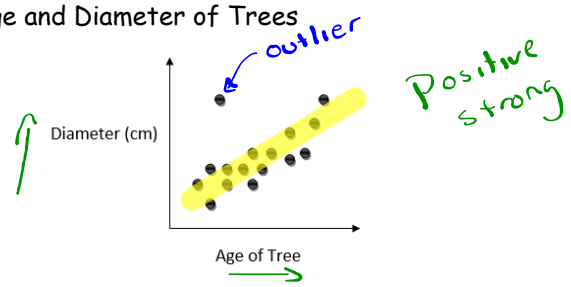
Correlation

A correlation between two variables indicates that there exists a relationship between them.

Ex:
- A persons weight and height
- Number of years of school and future income

We can illustrate a two variable distribution on the Cartesian plane by plotting data points (x and y coordinates). This is called a **Scatter Plot**.

Ex: Age and Diameter of Trees

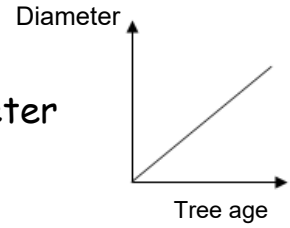


Nov 18-3:45 PM

A correlation can be positive or negative

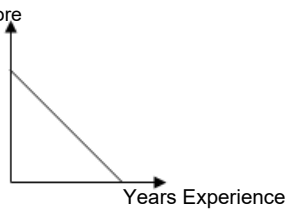
- **Positive** : when both x and y increase

  Ex: tree age increases, so does its diameter



- **Negative** : when x increases and y decreases

  Ex: your golf score decreases as years of experience increases



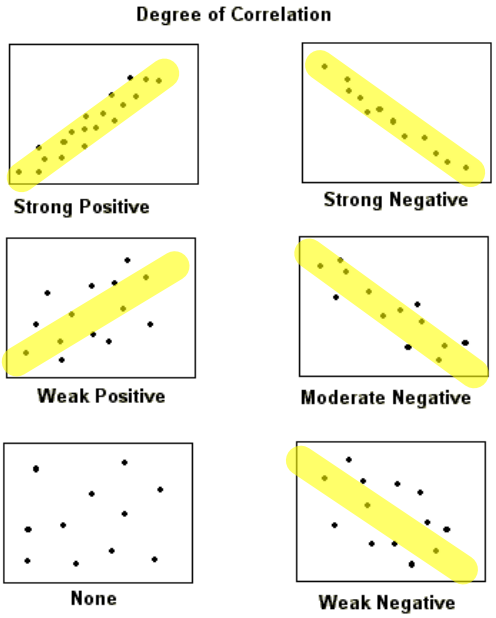
Nov 18-3:48 PM

Correlations are also characterized by their **strength**

Strong Correlation - If there is a strong correlation then the scatterplot graph will resemble a **line**

Weak correlation - If there are dots all over the place it is a weak or there is no correlation

Nov 18-4:27 PM



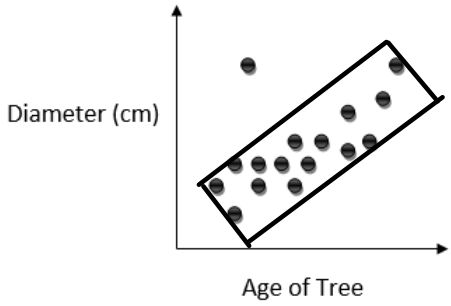
The strength is determined by how closely the scatter plot forms a line.

Nov 18-3:52 PM

Correlation Coefficient

To measure the strength of a correlation we need to determine a correlation coefficient (r)

Step 1 – Draw a rectangle around the points **(ignoring the outliers)**



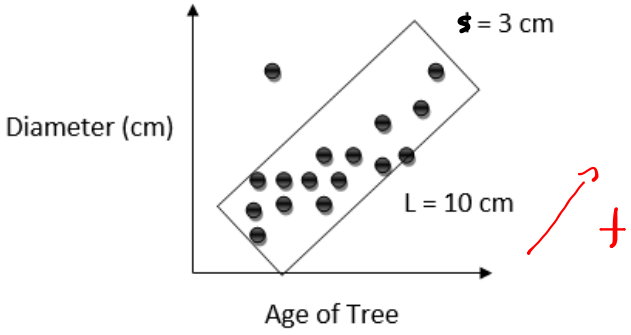
An OUTLIER is a point that indicates an abnormal piece of data or something out of the ordinary.

It is located far from main cloud of points in the scatter plot.

NB: Make the rectangle as tight a fit as possible.
The sides should be **parallel** to each other.

Nov 18-3:55 PM

Step 2 – Measure the long side (L) and the short side (s)



Nov 18-3:59 PM

Step 3 – Apply correlation formula

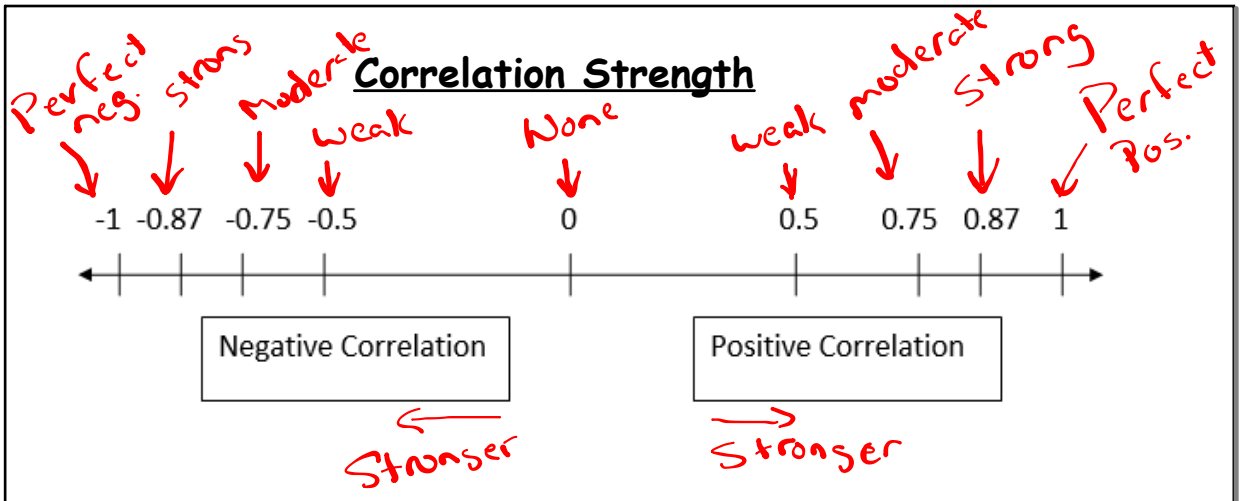
$$r \approx \pm \left(1 - \frac{\text{Length of short side}}{\text{Length of long side}}\right)$$

$$r \approx \left(1 - \frac{3}{10}\right)$$

$r \approx 0.7$ → this indicates a moderate positive correlation

Nov 18-4:00 PM

Correlation Strength



Correlation Coefficient		Meaning
Positive	Negative	
Near 0	Near 0	Zero
Near 0.5	Near -0.5	Weak
Near 0.75	Near -0.75	Moderate
Near 0.87	Near -0.87	Strong
Near 1	Near -1	Perfect

Nov 18-4:01 PM