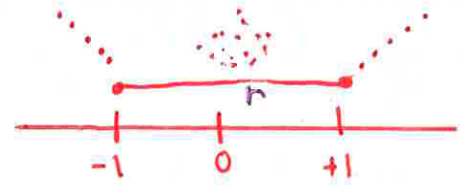


M9

# Data Trends

5.4

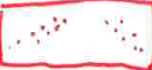




**A Correlation Coefficient** is a number that is used to show how closely or loosely a data set or a scatter plot follow a trend; this number is always between  $\pm 1$ .

Symbol for a correlation coefficient:      $r$     

$$-1 \leq r \leq 1$$

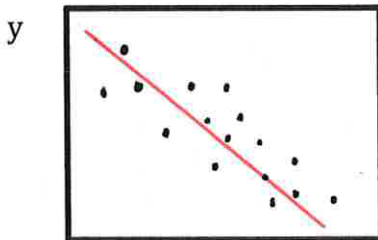
### Q: How does it work?

- When  $r$  is close to  $\pm 1$ , data follows a trend very closely = it is possible to draw a best-fit line. 
- When  $r$  is close to 0, data does not follow a trend at all = it is not possible to draw a best-fit line. 
- When  $r$  is close to  $\pm 0.5$ , data follows a trend rather loosely = there is a trend but data points are quite spread = it is possible to draw a best-fit line. 
- When  $r$  is close to -1, data shows negative correlation = line of best fit is decreasing.
- When  $r$  is close to +1, data shows positive correlation = line of best fit is increasing.

Example 1: A) Estimate the value of the correlation coefficient.

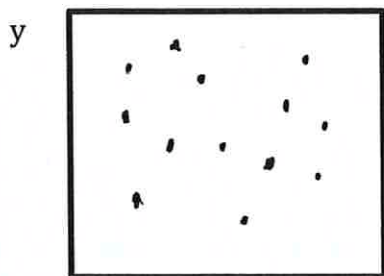
B) Determine whether data shows positive, negative, or no correlation.

I.



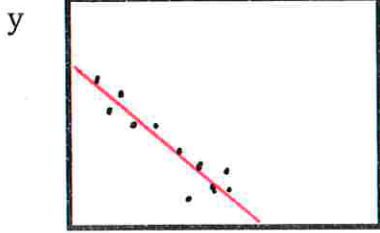
$r = -0.8$   
 Negative Correlation  
 Negative because it's decreasing

II.



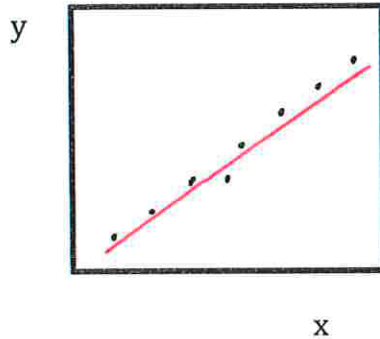
$r = 0$   
 No correlation

III.



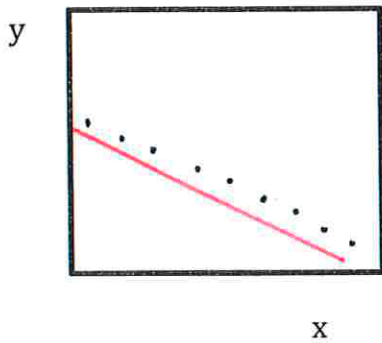
$r = -0.9$  or  $(-0.85)$   
Negative Correlation ↓

IV.



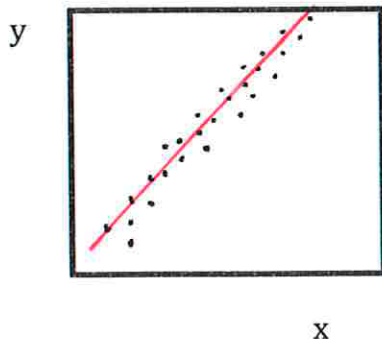
$r = 0.99$   
Positive Correlation ↑

V.



$r = -1$   
Negative correlation ↓

VI.



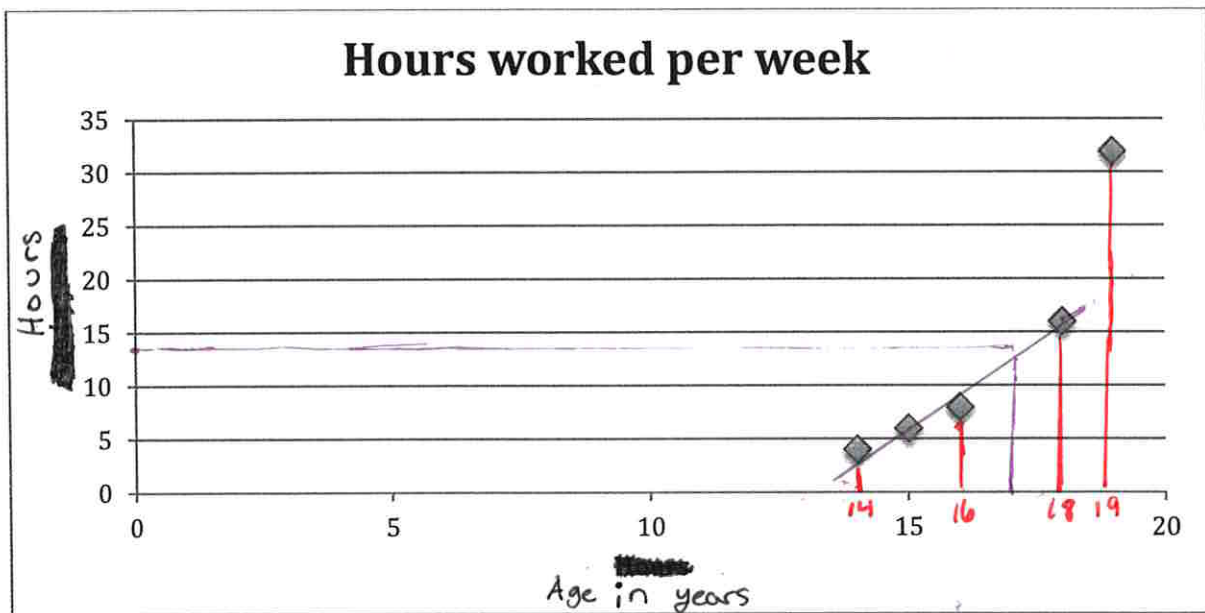
$r = 0.85$   
Positive correlation ↑

## Interpolation

- Interpolation is a process of estimating values off a scatter plot and the line of best fit.
- During interpolation you look for a value that is between the smallest and the largest value on the graph.
- Generally, interpolation results in a high-confidence predictions or estimates.

Examples:

1. Do you think the point associated with 19 years is an outlier? Why or why not?
2. Draw the line of best fit.
3. Estimate the number of hours a 17-year old works per week. Approx 15 hours

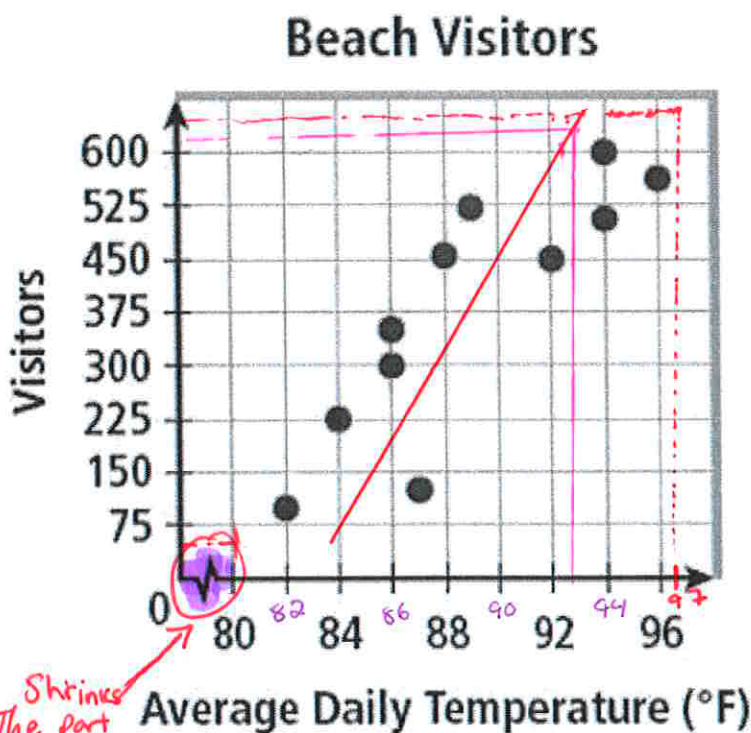


1. yes, because they are on the outside and don't follow the best fit line

## Extrapolation

➤ Extrapolation is a process of estimating values off a scatter plot and the line of best fit. During extrapolation you look for a value that is not between the smallest and the largest value on the graph.

- The value you look at is either smaller than the smallest value on the graph or it is bigger than the largest value on the graph.
- It is very important to exercise caution when extrapolation because you cannot assume that a pattern or a trend exists for values that are not on the graph.
- This means that when extrapolating, you are allowed only in a very close neighbourhood of the smallest and the largest value on the graph.
- The further away from points on the scatter plot, the lower the confidence of your prediction or estimate.



- Draw the best-fit line
- Estimate the following:

Number of visitors when  $T = 97^\circ\text{F}$

~ 600 people

Number of visitors when  $T = 80^\circ\text{F}$

~ 75

Temperature when there are 625 visitors

~ 93°F

Temperature when there are 60 visitors

~ 79°F

Shrinks  
The part  
That isn't  
Used.

Scale:  $\text{H} = 2^\circ\text{F}$

Scale:  $\text{I} = 75 \text{ people}$