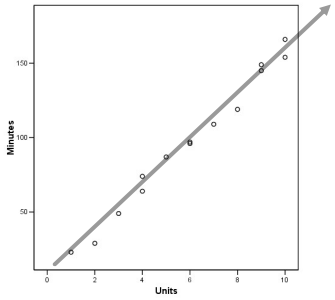
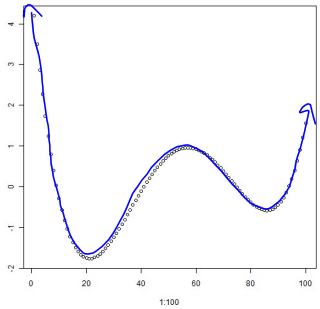
3.4 Correlation in Linear and Non-linear Data

Can you draw a LoBF for the following scatter plots?

A



B



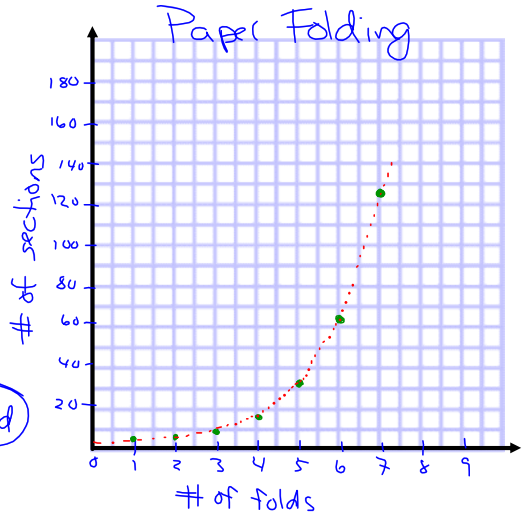
Which graph do you think has the strongest **correlation**? Why?

B - the points are ON the curve.

Smooth curve represents the trend or pattern in a scatter plot better than a line. This curve is called a **curve-of-best-fit**. These relationships are called **non-linear**.

Example 1: When a paper is folded in half, one crease line and two sections are created. When the paper is folded in half again and again, more crease lines and sections are created. Complete the following table and plot the data.

# of Folds	# of Sections
1	2
2	4
3	8
4	16
5	32
6	64
7	128



Solid or Dashed

Can't have half a fold

a) Is the relationship between the number of folds and the number of sections linear or non-linear?

Curve

b) Use your graph to predict how many sections there would be after 7 folds. Is this **interpolation** or **extrapolation**?

c) Do you think it is possible to fold one sheet of paper 10 times?

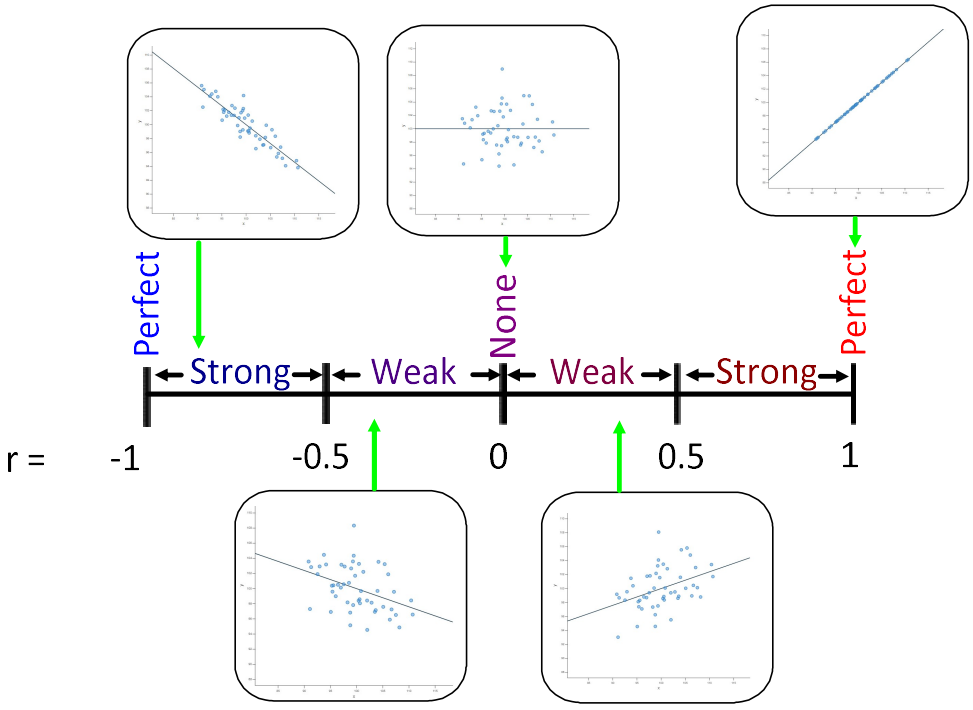
d) How could we check our answer for c)? (click!) Mythbusters

Correlation in Linear Data

For **linear data**, we can measure the strength of the correlation between two variables using the **correlation coefficient** (r).

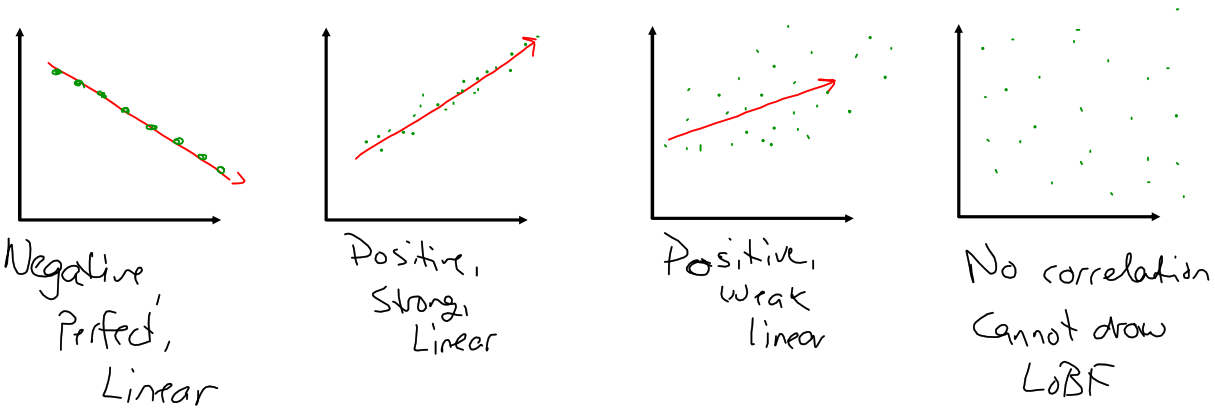
This value represents how far the points are, on average, from the LoBF.

- The closer the value is to 1 or -1, the closer the fit.
- The **sign** of r tells you if the **linear** data is increasing (positive) or decreasing (negative).



Example 3: Sketch the general appearance of a scatter plot and **describe the correlation** associated with each value of r. Include a LoBF.

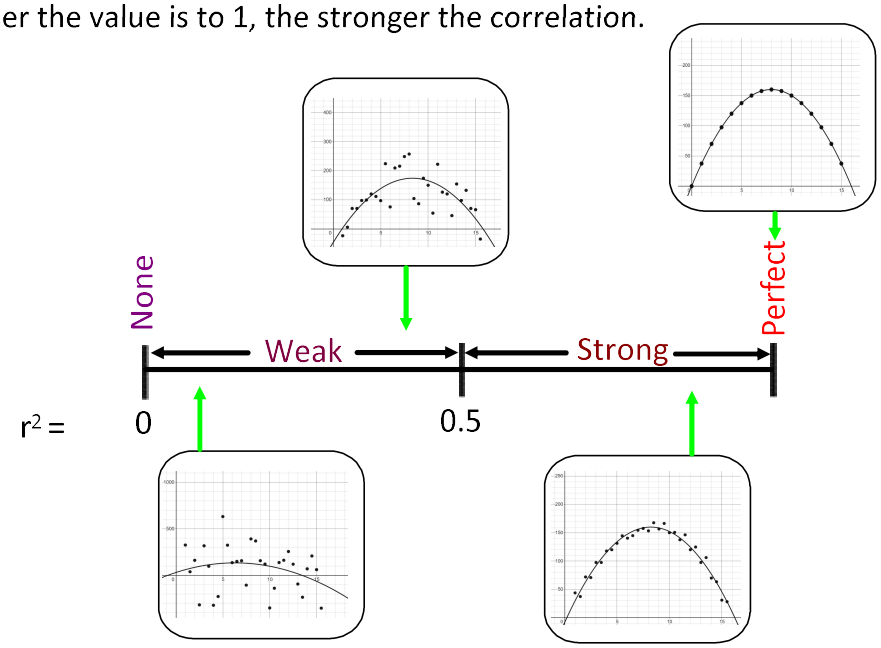
a) r = -1 b) r = 0.92 c) r = 0.45 d) r = 0



Correlation in Non-linear Data

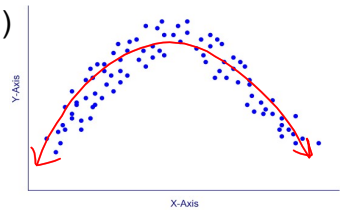
For **non-linear data**, we can measure the strength of the correlation between two variables using the square of the correlation coefficient (r^2).

This value represents how far the points are, on average, from the Curve-of-Best-Fit.
→ The closer the value is to 1, the stronger the correlation.



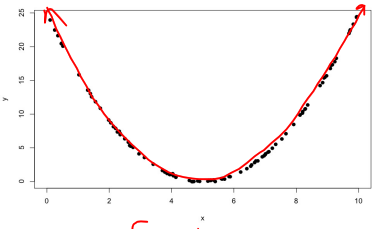
Example 4: Describe the relationships below.

a)



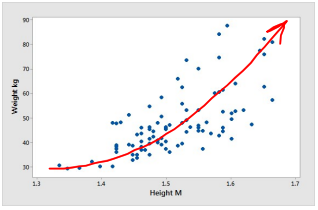
Strong
Non-linear
Positive to negative

b)



Perfect non-linear
Negative to Positive

c)



Weak
Non linear
Positive

Example 5: The value of vehicles **depreciate** (decrease) over time. Husniyah purchased a new car 10 years ago for $16,500. Every two years, they checked the value of their car and recorded the data in the table below.

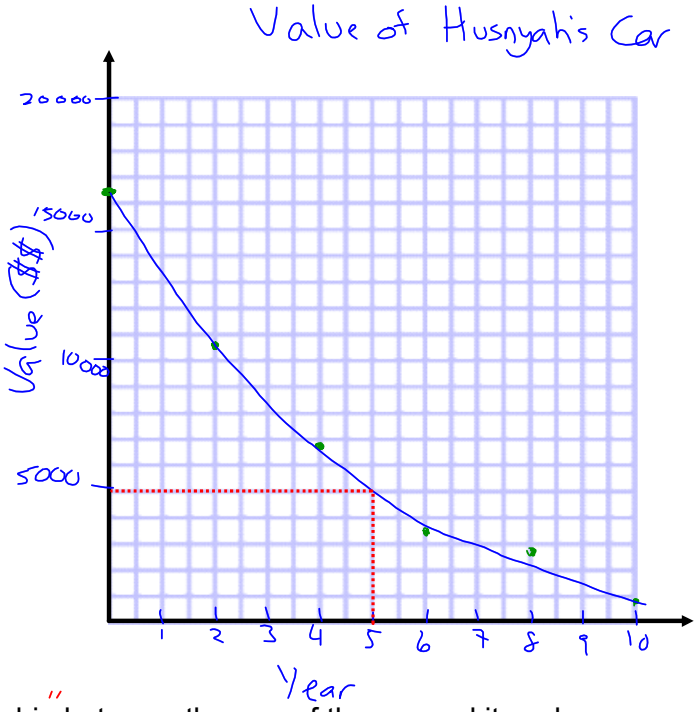
Year	Value ($)
0	16,500
2	10,560
4	6,760
6	4,330
8	2,770
10	1,770

Solid or Dashed?

a) Independent variable? Year

Dependent variable? Value

b) Graph the scatter plot and draw a curve-of-best-fit.



Words!

c) Describe the relationship between the age of the car and its value.

As the year goes up, the price goes down

"correlation"? Perfect non-linear

d) Estimate the price of Husniyah's car when it is 5 years old. **Show traces of your work on the graph!**

About $5000

e) Will the value of the car keep decreasing forever? Will it ever be zero?

Practically, can always sell for SOMETHING

But the curve will keep getting closer and closer to zero.

- Value will never be negative.