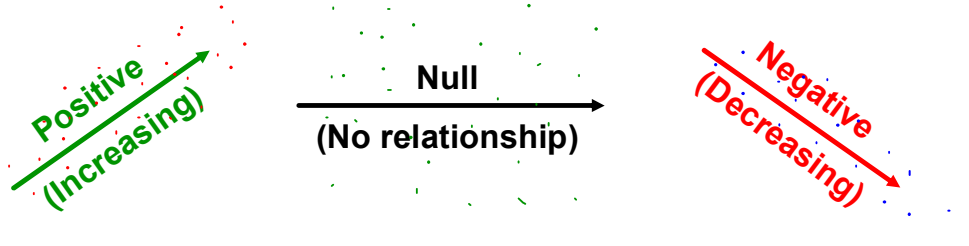


3.2 - Scatter Plots and Correlation

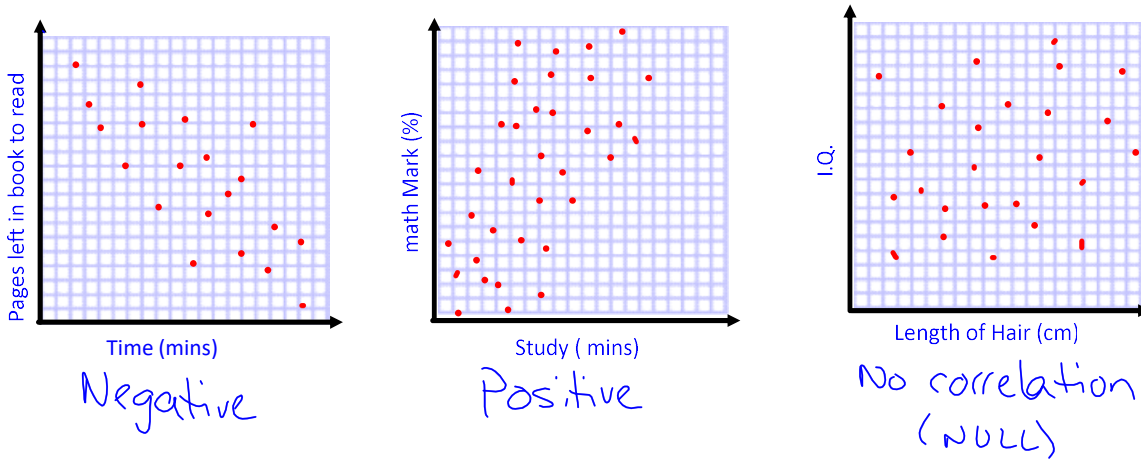
A **scatter plot** is a graph that is made up of a **set** of points between two variables.

- They help us visualize trends in graphs
- They help us make inferences (conclusions) from our data

Scatter plots help us see if two variables are **correlated** - might have some sort of relationship. The correlation could be...



Example 1: Determine whether the scatter plots below show a positive, negative or null correlation.



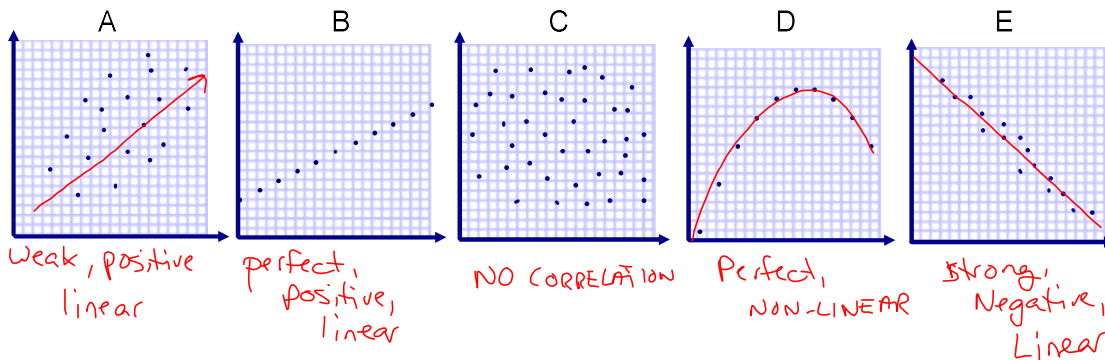
Example 2: Describe the following as positive, negative or no correlation.

- The amount of popcorn left in a bowl and the time into the movie.
Negative (Handwritten blue text with a red arrow pointing down-right)
- The size of your hand and the number of rings you own.
No correlation (Handwritten blue text)
- The outside temperature and the number of people swimming.
Positive (Handwritten blue text with a red arrow pointing up-right)

We can also describe the **strength** of the correlation as weak, strong or perfect.
 The closer the points are to following a line (**linear**) or another pattern (**non-linear**), the stronger the correlation.

Example 3: Using the word bank below, describe the correlations for each of the following scatter plots.

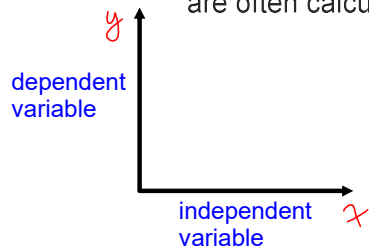
[weak, strong, perfect, positive, negative, linear, non-linear]



Variables

Independent Variable: a variable that affects the value of another variable; we often choose the values.

Dependent Variable: a variable that is affected by another variable; these values are often calculated.



Example 4: Determine which is the independent variable, and which is the dependent variable.

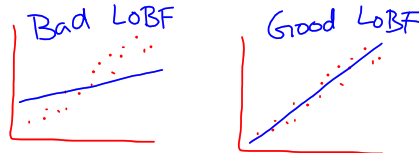
- a) how much studying you do and your test mark
I D
- b) your height and your age
D I
- c) how long you invest your money and how much you earn in interest
I D
- d) your speed and the distance you have traveled
I D

Using a Graph to Make Predictions:

Line of Best Fit: a line that best fits the data. Can be used for predicting trends in your scatter plot.

Inside the data
Interpolation: If you predict a value that is between points on your scatter plot.

Extending outside the data
Extrapolation: If you predict a value that is beyond the points on your scatter plot.

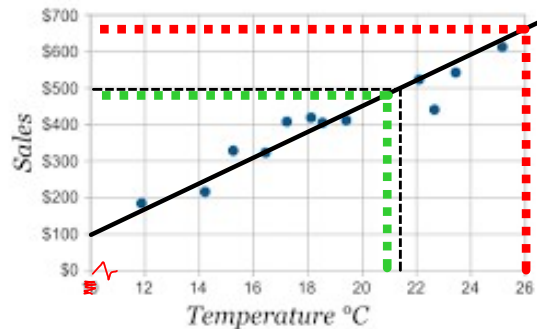


LoBF by Hand:
 Same number of points above & below the line
 spread across entire line

Example 5: Draw a Line of Best Fit (LoBF) for this data and answer the questions that follow.

a) What were the sales when the temperature was 21°C? Did you interpolate or extrapolate to find your answer? *About \$480*

Interpolation -> since it was inside our data



b) What will the sales be when the temperature is 26°C? Did you interpolate or extrapolate to find your answer? *About \$670*

Extrapolation

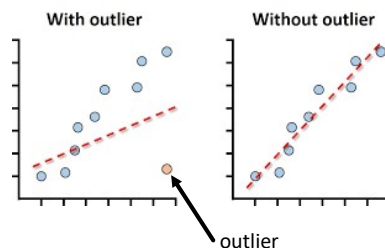
Since it was past our last data pt.

c) What temperature would result in sales of \$500? Did you interpolate or extrapolate to find your answer? *About 21.5°C*

Interpolation

An outlier: a point that doesn't follow the same trend as the other points.

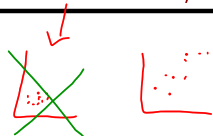
- Sometimes it is caused by a measurement error.
- It can "distort" the trend/line of best fit.
- Removing the outlier can result in a Line of Best Fit that more accurately reflects the trend of the data.



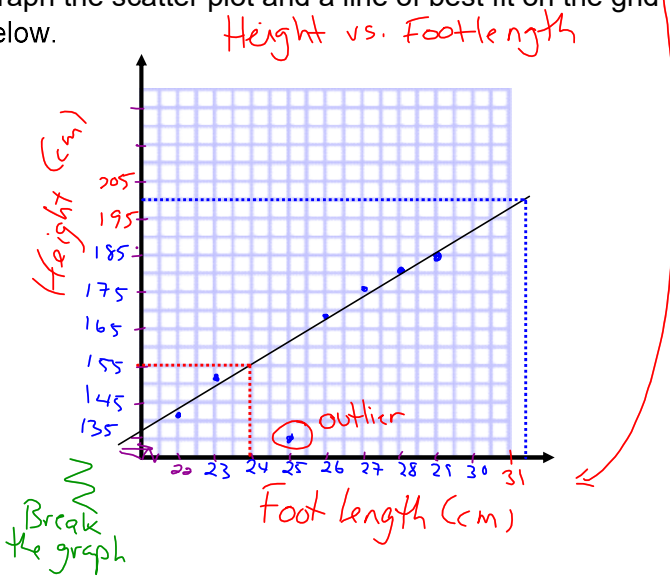
Example 6: Simon is investigating to see if there is a relationship between a person's height (in cm) and the length of their foot (in cm). They recorded the data in the table below.

Foot length (cm)	Height (cm)
28	180
23	152
25	135
22	142
26	169
27	176
29	185

- Graphing Requirements**
- title, labels, units
 - scales can be different on x and y
 - arrows on axis
 - use a ruler
 - do not crowd your work



- a) Independent variable? Foot length
 Dependent variable? Height
- b) Graph the scatter plot and a line of best fit on the grid below.



- c) Describe the correlation between foot length and height.
 Strong, positive, linear
- d) Estimate the height of someone who has a 24cm foot. Did you use interpolation or extrapolation?
 ~ 155 cm
- e) Estimate the foot length of someone who is 200cm tall. Did you use interpolation or extrapolation?
 ~ 31.5 cm
- f) Is there an outlier? Should it be removed?
 Yes Yes, because without it there is a strong correlation.