

CORRELATION vs CAUSATION

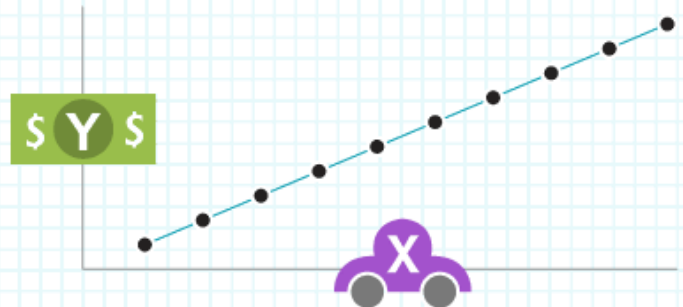
A MATHEMATOGRAPHIC

A **Correlation** measures the relationship between any 2 variables (**X** & **Y**).

A **Variable** is just about anything that (1) changes and (2) you can measure.

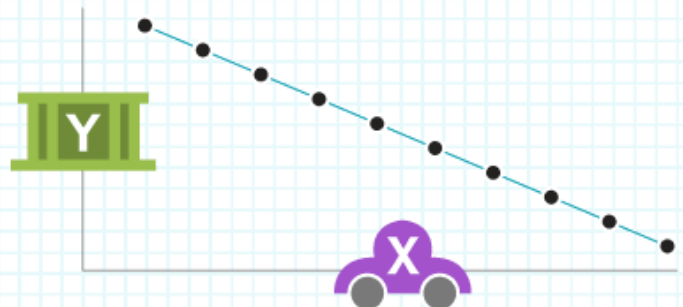
Example 1: Let's say **X** is the number of cars you buy, and **Y** is the amount of money you spend (assuming all the cars cost the same). The more cars you buy, the more money you spend.

We call this a **Positive Correlation**.



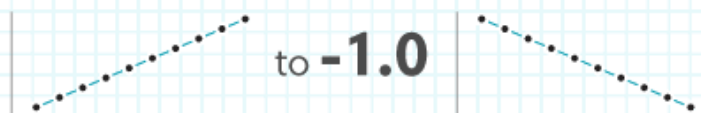
Example 2: Let's say **X** is still cars, but now **Y** is your bank account balance. With each car you buy, your bank account gets smaller and smaller. As **X** goes up, **Y** goes down.

We call this a **Negative Correlation**.

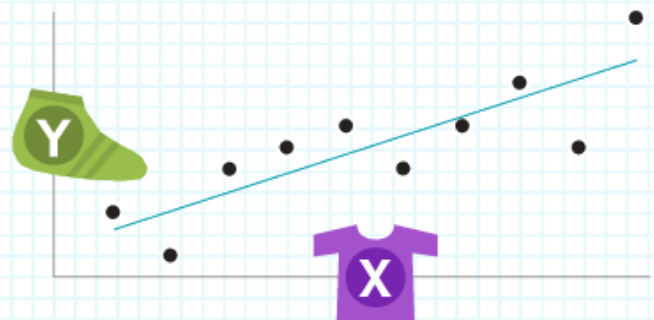


The **Correlation Coefficient (r)** measures the strength of the correlation.

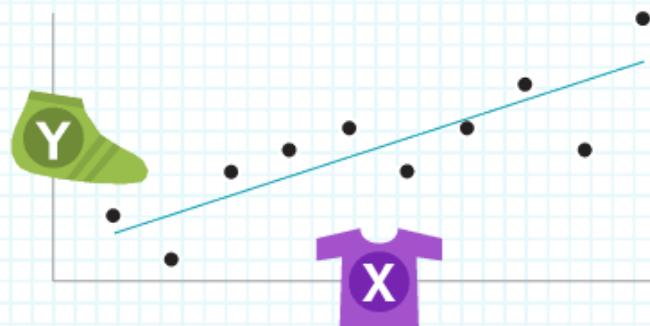
Values for (**r**) range from **+1.0**



Example 3: Let's try a less idealized example. Here, **X** is shirt size, and **Y** is shoe size. As one size goes up, so does the other, but the relationship varies from person to person. The correlation here is still positive, but it's not perfect (**r = +0.83**).



Example 3: Let's try a less idealized example. Here, **X** is shirt size, and **Y** is shoe size. As one size goes up, so does the other, but the relationship varies from person to person. The correlation here is still positive, but it's not perfect ($r = +0.83$).



CORRELATION DOES NOT IMPLY CAUSATION

In the 1st example, there's a clear cause at work. Buying cars causes you to spend money. So, if shirt and shoe size are correlated, does that mean that wearing bigger shirts causes you to wear bigger shoes? Of course not, and here's where things gets tricky.

There are **3 possible explanations**. First, **X** could cause **Y** or the other way around:



There could also be a hidden **Z Factor** at work, causing both **X** and **Y**. In this case, **Z** might be the person's height. Larger people usually wear bigger shirts **and** shoes:



What if the cause isn't obvious? You could run an experiment: manipulate **X** for a group and measure how **Y** changes. Unfortunately, if **Z** is involved, causation is tough to crack.

Those are the basics. The rest can fill books (and has). Let's review:

- A **Correlation** measures the relationship between **X** & **Y**.
- In a **Positive Correlation**, as **X** goes up, **Y** goes up.
- In a **Negative Correlation**, as **X** goes up, **Y** goes down.
- Correlation is expressed as (**r**), which goes from **-1.0** to **+1.0**.
- **Correlation does not imply causation.**