Oftentimes, statistical studies are done where data is collected on **two variables** instead of one in order to establish whether there is a **relationship** between the **two variables**. This is called a **bivariate data analysis**.

***Exercise* #1:** A survey was taken of 10 low and high temperatures, in Fahrenheit, in the month of April to try to establish a relationship between a day's low temperature and high temperatures.

Low Temperature, x	26	28	30	32	34	35	37	38	41	45
High Temperature, y	49	50	57	54	60	58	64	66	63	72

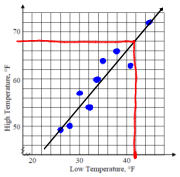
(a) Construct a scatter plot of this bivariate data set on the grid below.

(b) Draw a line of best fit through this data set.

(c) Calculate the slope of this line by picking off two points (not necessarily data points).

(26,49) (45,72)

$\frac{72-49}{45-26} = \boxed{\frac{23}{19}}$



(d) Use your line of best fit to estimate the high temperature for a day in April given that the low temperature was 42 degrees. Illustrate your answer on your graph.

68°F

(e) Would you characterize the relationship between the low and high temperature as a **positive correlation** or a **negative correlation?** Explain.

Positive Correlation

May 15-6:55 AM

Two variables can have a **strong relationship** with one another, as seen on a scatterplot, but might not have a **causal relationship**. A causal relationship exists when the **change in one variable** actually **causes** the **change** in the other (or is one of the primary causes).

***Exercise* #2:** In each of the following scenarios, two variables are given that if plotted would have a strong correlation (a scatterplot where the data falls nearly in a line). Determine if there exists a **causal** relationship between the two variables. If so, which variable causes the other?

(a) The high temperature in New York City and the number of bottles of water sold.

(b) A person's height and a person's shoe size.

(c) A person's weight loss and the number of hours a person spends in the gym per week.

(d) The years of education a person achieves and the salary that person starts at upon entering the work force.

May 15-6:59 AM

***Exercise* #3:** The table below shows the number of firefighters required to fight a given fire versus the dollar damage done to the house by the fire.

Number of firefighters	2	3	5	8	9	12	16
Damage done by fire (in dollars)	2,932	9,750	15,575	23,190	22,900	35,400	52,900

(a) Are the data positively or negatively correlated? How can you tell?

(b) Does the number of firefighters cause the damage done to the house? If not, what hidden variable is causing both variables to change?

May 15-7:01 AM