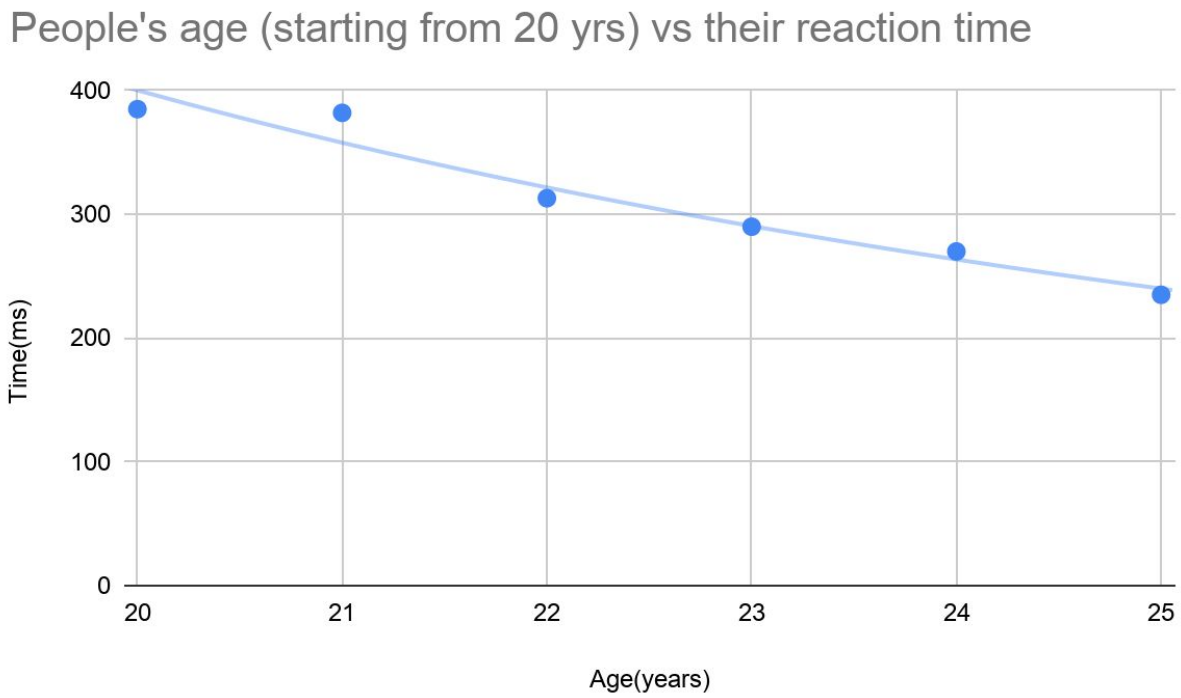
MDM4U Statistics of Two Variables Assignment

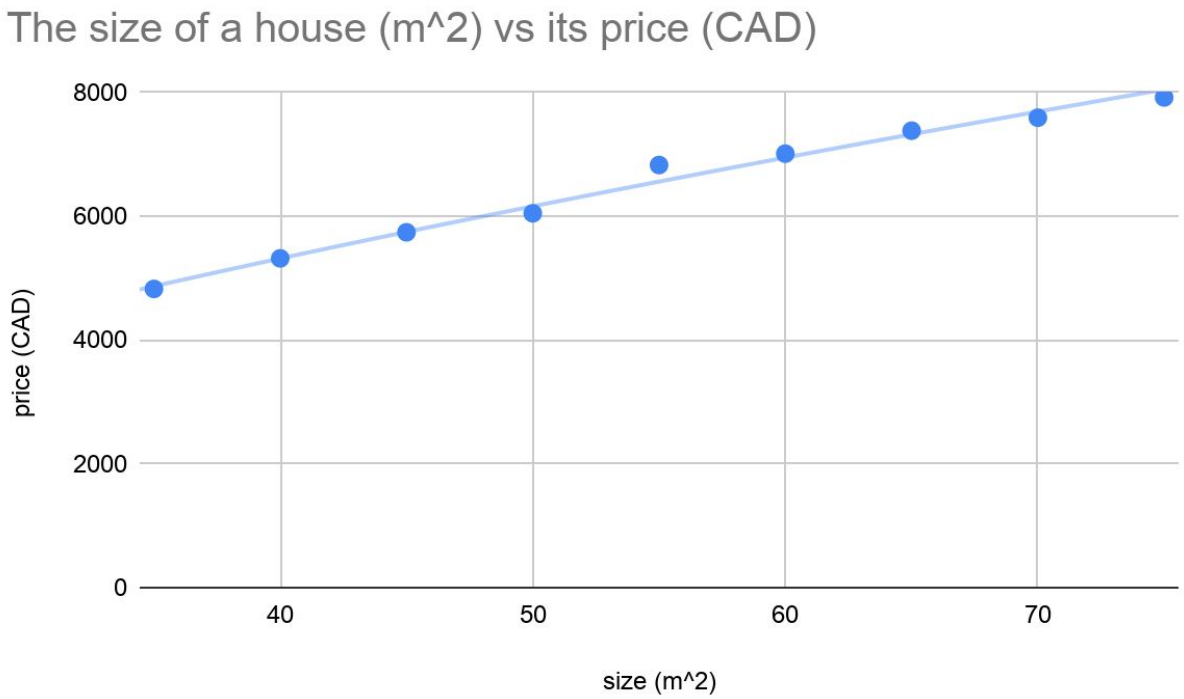
1. **Sketch a scatter plot that could represent data from each of the following. Label the axes to indicate the independent and dependent variables.**

a) people's ages (starting at 20) and their reaction times



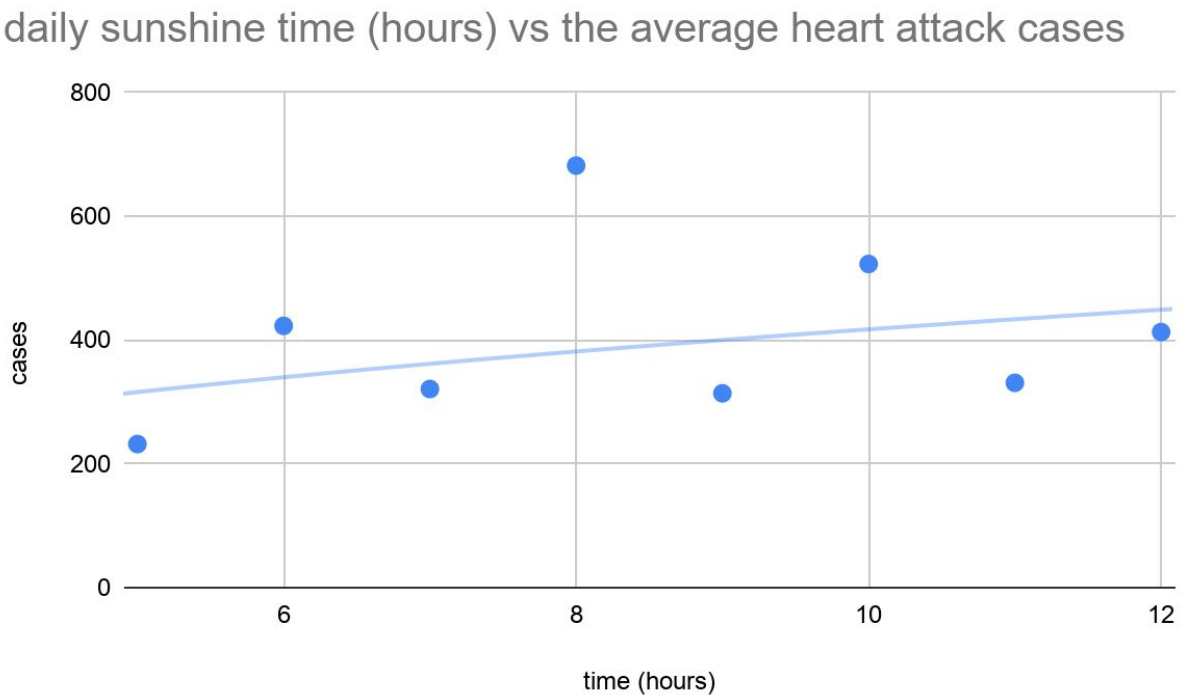
In the following situations, an individual's age is an independent variable, and the response time (in milliseconds) is a dependent variable: The trend in this scatterplot will follow a moderate negative linear correlation. This is because the more a person is in younger childhood, the faster the reaction time is.

b) size of a house and its price



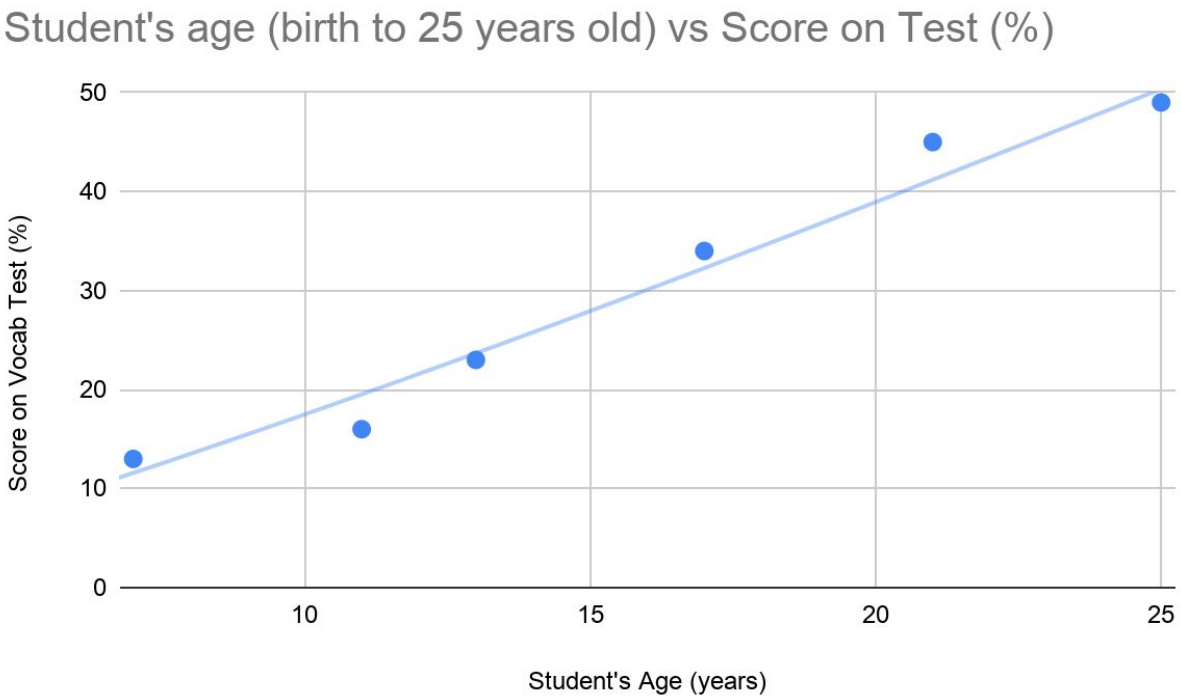
In the following circumstances, the housing size in square meters becomes an independent variable, and the housing price in Canadian dollars becomes a dependent variable. Assuming that the geographical conditions of all houses are the same, the trend of this scatterplot follows a strong positive linear correlation. The bigger the house, the more expensive it will be.

c) exposure to sunlight and the risk of heart attacks



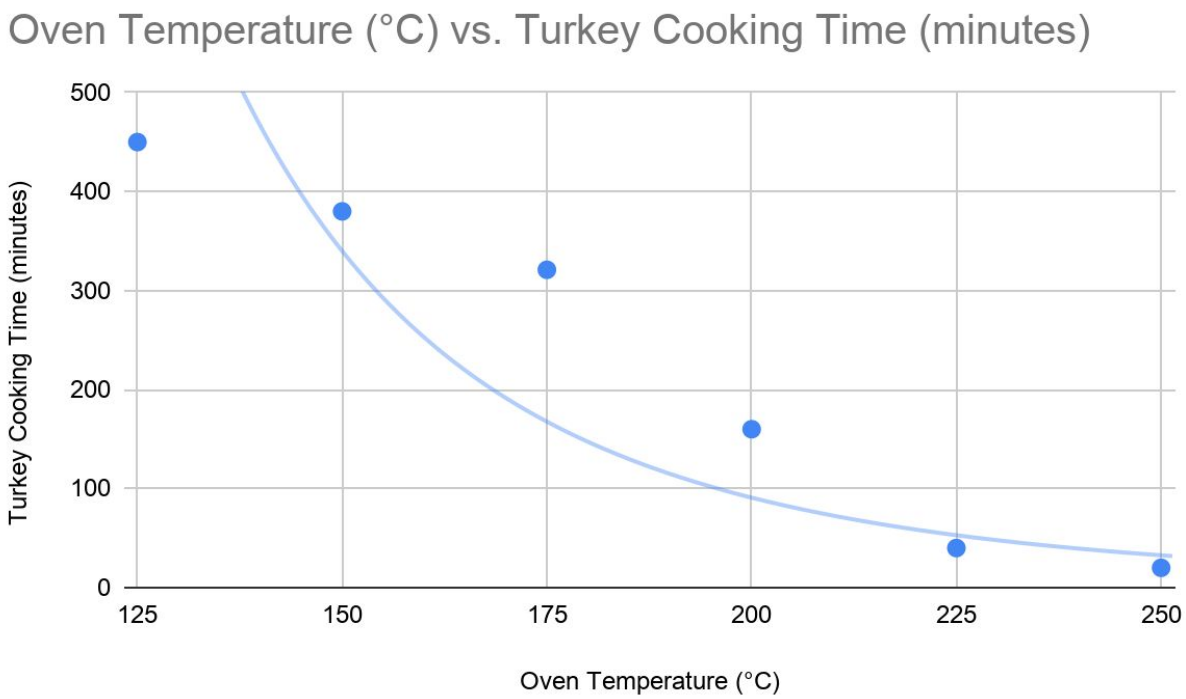
In the following situations, the sunshine time of day becomes an independent variable and the number of heart attack patients becomes a dependent variable: Because there is no clear pattern or trend between two variables, there is no correlation between them.

d) size of vocabulary and age (birth to 25 years old)



In the following situations, the student's age (from birth to 25) will be an independent variable, and the score obtained in the vocabulary test will be a dependent variable. The trend in this scatterplot will follow a strong positive linear correlation because the older a person gets, the more educated and the stronger the vocabulary that they appreciate.

e) oven temperature and the cooking time for a turkey



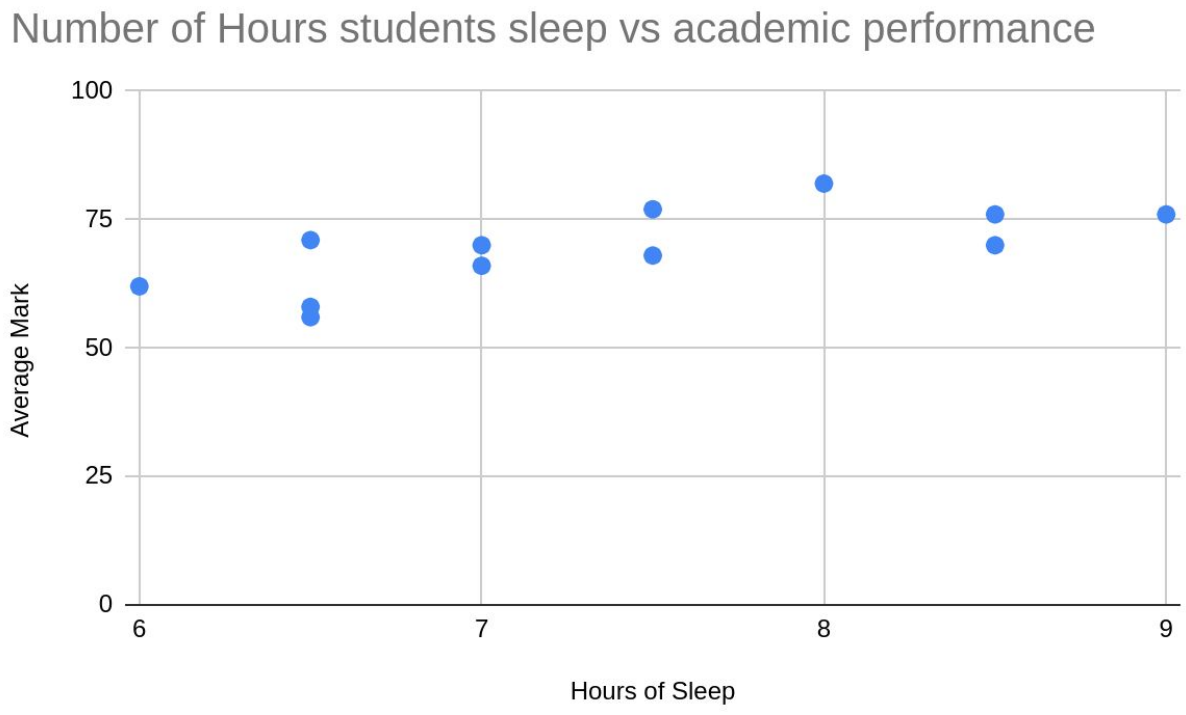
In the following situations, the temperature of the oven at Celsius temperature becomes an independent variable and the cooking time of the turkey is dependent within a few minutes. The trend in this scatterplot follows a strong negative correlation. However, unlike the previous four scatterplots shown above, the trendline for the scatterplot becomes much more curved.

2. A study was done with a group of university students to determine if there was a correlation between the amounts of sleep they got and their academic performance. The following table lists some data from the study.

Student	A	B	C	D	E	F
Hours of Sleep	6.0	6.5	7.0	6.5	8.5	8.0
Average Mark	62	58	66	71	76	82

Student	G	H	I	J	K	L
Hours of Sleep	9.0	8.5	7.0	7.5	6.5	7.5
Average Mark	76	75	70	68	56	77

a) Make a scatter plot of the data.



b) Using a spreadsheet, determine the PPMCC.

The Pearson product-moment correlation coefficient (PPMCC) of the data is available on Google Sheets by using CORREL formula. Using the formula, the PPMCC of the hours of sleep

(which is data_x, the range of values or a reference to the range of cells that consists of INDEPENDENT data) vs Average Mark data in percentage(%) form (which is data_y, the range of values or a reference to the range of cells that consists of DEPENDENT data).

The PPMCC according to the spreadsheet is 0.7622939709, which is about 0.76.

c) What would you conclude about the relationship between the two variables?

There is a moderate positive correlation between the amount of sleep a group of college students receives and their average academic performance. While a moderate positive correlation implies a causal relationship, a more logical conclusion is that the two factors are under common causes. These relationships may be common cause relationships. This is because diligent students tend to finish their work early and go to bed earlier because of their excellent time control. Therefore, it raises their academic scores and gives them more time to sleep.

3. A recent study found that people who illegally download music spend more on music than those who don't. These findings are often cited as evidence that piracy does not harm music sales, and can even enhance them.

Explain why the commonly reached conclusion to these findings may not be correct.

Because the study prefers causal relationships that incorrectly assume the common causes of the two variables (the frequency of illegally downloading music and the money spent on music), the conclusions that are generally reached may not be accurate. People who often copy music illegally spend less money on music because they copy illegally. However, people who want to listen to music while doing illegal copying may prefer music to people who don't do illegal copying. These people can rather spend more money on music.

Therefore, the discovery cannot be used to support claims that piracy does not harm music sales. Despite the fact that online streaming services such as YouTube Music and Spotify have become more popular, music sales are still high. If the results of a given study were true, music sales would have plummeted by now.

4. Give an example of each of the following types of causal relationships, fully explaining your suggestion:

a) common-cause factor

Consider a case in which one external variable changes two variables in the same way. For example, suppose a city has found that the revenue from parking fees on public beaches every summer is correlated with local tomato yields. It is very likely that the cars parked on the beach will not have any effect on the tomato harvest. Instead, good weather is a common cause of increased tomato harvests and the number of people parking on the beach.

b) Accidental relationship

The number of students enrolled in Python programming courses increases and the number of Tesla increases on the road. However, these two variables do not have a positive correlation that is purely coincidental. There is no causal relationship between one correlation variable. Because of the accidental relationship, there is no relationship between students' choice of classes and the type of driving they choose. So it's purely coincidental that they've all increased.

c) Reverse cause-and-effect relationship

In this case, consider the reversal of dependent and independent variables. For example, a researcher observes a positive linear correlation between the amount of coffee consumed by a group of medical students and their levels of anxiety. The researcher theorized that drinking coffee causes tension, but instead found that neurotic people are likely to drink more coffee.