# Correlation: Finding a Relationship between Two Variables

# Directions for Instructors

# Introduction:

This activity can be used by both Biology and Math instructors, for Math students, to help them see the connections between how the material taught in Math can be used in Biology and for Biology students to see how Biology data can be analyzed and visualized by using Math skills.

# Learning Objectives:

After completing this activity, students will be able to:

Explore the correlation of two variables to detect possible causal relationships

Plot the data and the regression trendline

Determine if the relationship is positive, negative, or nonexistent

Determine how well the regression fits the data with an r-squared value

Determine an appropriate independent variable

Identify independent and dependent variables

Locate a variable that may be influencing a variable of interest

Hypothesize about the relationship between these variables

Distinguish between correlation and causation

Recognize that plotting a relationship and the related regressions only show correlation

Explain that causation can only be determined through controlled experiments

Describe a possible study that could determine if the correlation is actually causation

## Goal 1. To analyze biological data by plotting two variables against each other and determining if there is a correlation between the two variables.

# Choosing an Independent Variable

1. Summarize what students have accomplished so far: choosing a variable of interest and summarizing it to better understand the dat.
2. Explain that the next question they may have is “Why?” is the data what the data is.
3. Explain that by choosing an independent variable, they are choosing to explore whether another variable is influencing their original variable.
4. Note: If you want to ensure a clear relationship, you can choose to insert a list of possible variable pairs to explore. This will allow you to know ahead of time what the relationships should look like and appropriate interpretation. Another option is to have them create some quick bubble graphs in GapMinder to make sure there will be a strong relationship between the variables.


# Plot the Data to Look for a Relationship

1. Explain that when you make a graph with points of two sets of continuous data, you are plotting the relationship between those datasets and can summarize the relationship with a trendline.
2. The r-squared for the trendline tells you how well the line explains the variation in the data. The closer the points to the line, the higher r-squared; i.e. the more variation that is explained by the line.
3. Direct students to use the worksheet and Excel workbook to explore their own variables.
4. Note: You can choose to have students explore nonlinear relationships by choosing nonlinear options in the trendline formatting window.

## Goal 2: The most important part of this activity is that students understand how to interpret the results in context of the biology problem and understand the role of graphing in mathematics and science contexts.

1. Explain to students that a positive slope (m in y=mx+b equation of trendline), then there is a positive relationship. As the independent variable goes up, the dependent variable goes up. If the slope is negative, then it is a negative relationship. If the slope is near 0, then there appears to be no relationship between the variables.
2. Explain that regardless of the outcome, we can only know the correlation of the variables. We can not know whether the independent variable causes the dependent variable to change. Experiments are required for this.
3. Guide students as they work on the worksheet to think about how they will explain the relationship in a biological context. If you are using the poster presentation portion of this activity, have them keep digital copies of their graphs for use on the poster. Have them think about how they will explain the relationship and help them think about how they could run an experiment to test for a causal relationship.
4. You may want to have students share their work with you when they have chosen their independent variable, when they add the sample of data to the worksheet, and when they paste their practice figure. These are good points to detect problems before they complete the work.

----------------------------------------------------------------------------------------------------

### Some Helpful Resources

Here’s a video that describes trendline interpretation: <https://youtu.be/mfX_yUvwJho>

Here is a long (12’) video on determining R-squared of a regression line: <https://youtu.be/lng4ZgConCM>

Here is a video on how to make scatterplots in Excel: <https://youtu.be/gBbGBrHTMrM>

Here is a video on correlation vs. causation: <https://www.khanacademy.org/math/statistics-probability/designing-studies/sampling-and-surveys/v/correlation-and-causality>

This material is based upon work supported by the National Science Foundation under Grant No.1919613. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.