   

# **Project EDDIE: Biomes, Vegetation Structure, and Canopy Height**

# **Student Handout (Activities B & C):**

# Testing hypotheses about canopy height and climatic factors

This module was initially developed by Mulcahy, M. 10 February 2022. Project EDDIE: Biomes, Vegetation Structure, and Canopy Height.

# Learning objectives:

##### Activity B (with instructor guidance)

* To construct a hypothesis about the relationship between the height of vegetation and growing season length
* *To test the hypothesis through visualizations of the data*
* *To draw a conclusion supporting or refuting a hypothesis based on evidence and statistics*

##### Activity C (independently or in small groups; instructor helping as needed)

* To construct hypotheses about the relationship among height of vegetation, temperature, and precipitation
* *To test the hypothesis through visualizations of the data*
* *To draw a conclusion supporting or refuting a hypothesis based on evidence and statistics*
* *To investigate the relationship between two variables of choice related to biomes, biotic, and abiotic factors*

# Overview

Many scientists consider vegetation structure as the most important biotic feature defining a biome. Without trees, for example, a plant community isn’t a tropical forest, even if it has lots of rain and year-long warm weather. One of the goals of today’s activity is think deeply about “vegetation structure.” What is vegetation structure? A helpful analogy might be the structures of buildings in a neighborhood: Does a subdivision contain many small single-story houses or many tall skyscrapers? How close together are the buildings? Are they built with wood or brick or another building material? Can you walk between them, or are they row houses, filling all the space? Do they have basements or are they entirely above-ground? Vegetation structure can refer to many aspects of a plant’s form. Two important features (often correlated) are the height of the tallest plants and whether the plants contain any wood. Some plants are tall and woody (trees), other are short and woody with many stems (shrubs), and other plants have no wood at all. Herbaceous (non-woody) plants can have broad leaves (forbs) or have thin narrow leaves like grasses and sedges (graminoids). These groups of plants have adaptations that may make it more or less likely that they will be found in particular abiotic conditions, especially dry (arid) conditions. In this exercise, we will focus on canopy height, keeping in mind that canopy height is only one component of vegetation structure. Canopy height may be defined as the height of the highest vegetation in that location or, alternatively, as the average height of the vegetation.

Today’s activity explores vegetation structure using the hypothesis-testing scientific approach. You will construct hypotheses about canopy height and its relationship to abiotic factors. You will test the hypotheses using data from the open-science National Ecological Observatory Network (NEON) project. Lastly, you will draw a conclusion that is supported by the visualization of the data and statistics.

# Outline:

1. Activity B: Constructing and testing a hypothesis about how canopy height may vary by growing season length
2. Activity C: Constructing and testing a hypothesis about how canopy height may vary by temperature and precipitation
3. Activity C: Develop and test your own hypothesis using variables of your choice.

# NEON field site data

We will be working with data collected from all 20 core terrestrial field sites from the open-science National Ecological Observatory Network (NEON) project. We will be using a free open-source graphing program called the Common Online Data Analysis Platform or “CODAP.”

**Activity B: How closely does canopy height vary correlate with growing season length?**

# Directions for Activity B

1. *CONSTRUCTING YOUR HYPOTHESIS*

Our guiding question for this activity is: Does canopy height vary with growing season length?

**In question 1 of the Question Set,** **state your hypothesis about your prediction of the relationship between canopy height and growing season length.** Make sure your hypothesis clarifies whether you expect canopy height to be positively correlated (with canopy height increasing with increasing number of growing season days) or negatively correlated (with canopy height decreasing with increasing number of growing season days).

**Respond to question 2 of the Question Set: Explain the biological reasoning behind your prediction.**

Your hypothesis is only an educated guess, and you will need to examine the data before drawing a conclusion about the truth of your statement.

1. *TESTING YOUR HYPOTHESIS*

Open the data. The file should open automatically in Codap as soon as you click on the link below –[EDDIE Biome Module Field Site Data Codap File](https://codap.concord.org/app/static/dg/en/cert/index.html#shared=https%3A%2F%2Fcfm-shared.concord.org%2FXsC4sVyBX41a53KhC0Ha%2Ffile.json). A reminder that these are actual data that are publicly available to anyone, and which were gathered at the 20 NEON core terrestrial field sites.

You will start by evaluating your hypothesis by visualizing the data with a scatter plot. To create a scatterplot, click on the graph symbol in Codap as shown in the following screenshots, and then click and choose the variables (attributes) on the axes.Which variables will you use? The NEON dataset has quite a lot of variables, and they have highly standardized names that were chosen by the NEON scientists! There is no variable called “growing season length,” but there is a variable called “field\_avg\_number\_of\_green\_days,” which is what we will use. For canopy height, we will use the variable called “field\_mean\_canopy\_height\_m.” The independent variable should be selected on the horizontal X-axis, and the dependent variable should be selected on the vertical Y-axis.

**Answer question 3 in the Question Set: Which variable is the independent variable? Which variable is the dependent variable?**

Graphical user interface, diagram, application

Description automatically generated

Graphical user interface

Description automatically generated with low confidence

1. It may be tempting to think you see a pattern in the scatterplot. Instead of just informally looking for a pattern, we will use statistics to help assess the presence of a pattern in a quantitative and precise way. To do this, we will construct a linear equation (a “best fit” line) and calculate a statistic called the r-squared value, which will give us a way to measure the nature and strength of the relationship between the variables. The r-squared statistic gives us a numerical value for how well the equation of the line (also called the “model”) approximate the real data points. Visit this [simple description](https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/regression-and-correlation/coefficient-of-determination-r-squared.html#:~:text=%C2%AFy)2.-,R%202%20%3D%201%20%E2%88%92%20sum%20squared%20regression%20(SSR)%20total,from%20the%20mean%20all%20squared.) to learn more about the statistic and how it is calculated. We can also examine the slope of the linear equation to see if there is a positive or negative association between the variables. You can add the best fit line by clicking on the ruler icon to the right of the graph as shown in the following screenshot (click on the ruler & then make sure the “Least Squares Line” is checked).

Chart, scatter chart

Description automatically generated

1. **In question of 4 of the Question Set, paste or insert a copy of the graph you made with the equation and r-squared value.** You can take a screen shot and paste it into your answers for the question set or you can click on the camera icon in the Codap menu to save the graph. If you save the graph, you will need to insert the picture into your answer sheet. To take a screenshot, use the “print screen” button or snipping tool on a Windows computer or Shift, Command, and 3 on a Macintosh computer.
2. How do you interpret the r-squared statistic?

R-squared values vary from zero to one. If an r-squared value is close to zero, the model does a poor job at predicting the real data. If an r-squared value is close to one, the model does a good job predicting the real data accurately.

R-squared values can also be interpreted as proportions, and by multiplying them by 100, can be restated as percentages. If your r-squared value is 0.01, then 1% of the variation in the dependent variable is explained by the variation in the independent variable, using the line equation. If your r-squared value is 0.85, then 85% of the variation in the dependent variable is explained by the independent variable, using the line equation.

**Respond to Question 5 of the Question Set by completing the following statement:**

The value of the slope of the best fit linear equation is \_\_\_\_\_\_\_\_\_(enter value). An r-squared value of \_\_\_\_\_\_\_ suggests that there is \_\_\_\_\_\_\_\_(no relationship, a weak relationship, or a strong relationship) between the two variables. \_\_\_\_\_\_% of the variation in the dependent variable, \_\_\_\_\_\_\_\_\_\_\_\_, can be explained by variation in the independent variable, \_\_\_\_\_\_\_\_\_\_\_\_, using the line equation.

# **Question Set for Activity B**

1. State your prediction or hypothesis about whether you expect the relationship between canopy height and growing season length to be positively correlated (with canopy height increasing with growing season length) or negatively correlated (with canopy height decreasing with growing season length).
2. Explain the reasoning behind your prediction.
3. To test your prediction/hypothesis, you created a scatter plot. Which variable is the independent variable? Which variable is the dependent variable?
4. Draw a conclusion about the accuracy of your hypothesis using the linear equation and r-squared value. To facilitate communicating your conclusion, please fill in the blanks for the following sentence. Notice how you are using evidence and statistics to draw a conclusion about the accuracy of your hypothesis:

The value of the slope of the best-fit linear equation is \_\_\_\_\_\_\_\_(enter value). An r-squared value of \_\_\_\_\_\_\_ suggests that there is \_\_\_\_\_\_\_\_(choose one: no relationship, a weak relationship, or a strong relationship) between the two variables. \_\_\_\_\_\_% of the variation in the dependent variable, \_\_\_\_\_\_\_\_\_\_\_\_, can be explained by variation in the independent variable, \_\_\_\_\_\_\_\_\_\_\_\_, using the line equation.

1. Paste a screenshot or insert a picture of your graph with the equation and least squares line here**.** To save the graph you can take a screen shot, use a clipping tool, or click the camera button to below the rule tool which will allow you to save an image file that can be inserted into the Question Set for Activity B.

**Activity C: How closely does canopy height correlate with temperature and precipitation?**

Now that you have practiced the hypothesis-testing approach in Activity B, you will repeat the procedure in Activity C for canopy height in relation to additional abiotic factors. The ecologist Robert Whittaker proposed that the two most important abiotic factors that might affect biome are temperature and precipitation. In Activity C, you will use the NEON field site data and r-squared values to evaluate Whittaker’s decision to focus on just temperature and precipitation.

**Directions for Activity C**

1. Test if canopy height is associated with temperature using the NEON field site dataset.
2. Construct a hypothesis about the relationship between canopy height and temperature.
3. Discuss with your group what independent and what dependent variable you will use to evaluate the above statement.
4. Test your hypothesis through visualization of the data and the r-squared statistics. In other words, create a single graph. If your group has second thoughts after making one graph, feel free to change the axes until you feel like you have graphed the best possible pair of dependent and independent variables. As with Activity B, add the best fit line and the r-squared value to the graph.
5. Save a picture or screenshot of just one of the graphs that your group decides is the best one to evaluate the statement.
6. Work with your group to write a concluding statement which includes the r-squared value.

**Document your work for Activity C part 1 a-e in the Question Set.**

1. Test if canopy height is associated with precipitation using the NEON field site dataset.
2. Construct a hypothesis about the relationship between canopy height and precipitation.
3. Discuss with your group what independent and what dependent variable you will use to evaluate the above statement.
4. Test your hypothesis through visualization of the data and the r-squared statistics. In other words, create a single graph. If your group has second thoughts after making one graph, feel free to change the axes until you feel like you have graphed the best possible pair of dependent and independent variables. As with Activity B, add the best fit line and the r-squared value to the graph.
5. Save a picture or screenshot of just one of the graphs that your group decides is the best one to evaluate the statement.
6. Work with your group to write a concluding statement which includes the r-squared value.

**Document your work for Activity C part 2 a-e in the Question Set, and respond to question 3 in the Question Set: Based on your graphed data, is canopy height more strongly correlated with temperature or precipitation? How can you tell?**

1. Test a hypothesis of a relationship between any two variables of your choice.
2. Explore the dataset and construct a hypothesis about the relationship between any two variables in the dataset. Try to choose variables that you think ought to have a relationship to each other for scientific reasons.
3. Test your hypothesis by creating an appropriate visualization of the data and calculating the r-squared statistic.
4. Draw a conclusion based on the evidence and statistics.

**Document your work for Activity C part 3 a-c in the Question Set.**

# **Question Set for Activity C**

1. Document your work for Activity C part 1 a through e.
2. Document your work for Activity C part 2 a through e.
3. Based on your graphed data, is canopy height more strongly correlated with temperature or precipitation? How can you tell?
4. Document your work to Activity C part 3 a through c.