**Student Assignment (complete out of class)**

For the last two weeks we have been exploring evolution and ecology, and also focusing on data collection and interpretation. This assignment will allow you do explore both the topic of phenology and use scatterplots to explore relationships between variables.

**Guiding questions**:

* Based on observations of bumble bee phenology, are bumblebees in the western United States behaving differently from 2011 to 2019?
* What climate variables may help explain bumblebee activity?

**Data Source**: All data for this activity were collected by community scientists and are available from the National Phenology Network website: <https://data.usanpn.org/observations/get-started>

**Objectives**: For this assignment, please answer the questions below in each of the 4 sections. You will use a website called CODAP to explore and interpret scatterplot graphs and regressions about the phenology of bumblebees and other variables. You will make decisions about which data to use and evaluate your confidence in your conclusions given the nature of the available data. You will include images of your scatterplots.

**Part A. Phenology (3 questions)**

1. What is phenology?
2. Why do people care about phenology?
3. What is most interesting to you about phenology?

**Part B.** **Are bumblebees behaving differently from 2011 to 2019? (8 questions)**

To explore this question, you will use a website called CODAP:

* Go to the Explanation Paper #4 assignment link on Moodle and download the file named “Explanation Paper 4 dataset\_bumblebeeactivity.csv”. Be sure to save it to your desktop or somewhere that you can find the file. (Note: I generated this data file from the National Pollinator Network website!).
* Go to the CODAP website: <https://codap.concord.org/>
* Click on “Try CODAP” in the middle of the page.
* The message, “What would you like to do?” will pop up – select “Create New Document”. You will see a relatively empty workspace.
* Locate the explanation paper 4 data set file that you downloaded from Moodle. Click and drag the file anywhere onto the workspace. The datafile of bumblebee activity should now be a table on the workspace. (Note: The original data file format needs to be .csv; do not change the file format when you download the file from Moodle.)
* If you would like, you can resize the data table in your workspace by clicking and dragging on the bottom right corner.
* In the top left corner of the workspace, there are multiple icons (e.g. Tables, Graph, Map, etc.). Click on “Map”.
	+ Because the dataset from the National Pollinator Network has the coordinates of each observation, you can open a map to see where all the observations came from!
* Next, click on “Graph”. A graph with all of the data points appears. The data points are floating randomly in space because we have not specified the axes yet.
* Click on the Y axis where it says “Click here, or drag an attribute here”.
* From the menu of variables, select “First\_Yes\_DoY”. This is the “Day of Year” when the first bumblebee activity was recorded at that site.
	+ Notice that when you specify which variable you would like on the Y axis that the data points organize into a “data distribution” on the Y axis. Think to yourself: *How would you describe the center and the spread of this data distribution?*
* Now specify the X axis by clicking on “Click here, or drag an attribute here”.
* From the menu of variables, select “First\_Yes\_Year”. This is the year of the measurement.
	+ You now have a scatterplot that shows the relationship between time (year) and first day of bumblebee activity!
* Now, click on the “ruler” icon from the menu on the right side of the scatterplot. Place a checkmark next to the “Least Squares Line”. Click the “ruler” icon again to hide the menu.
	+ Remember that the equation for a linear regression is Y = α + b X, where Y is the dependent variable; X is the independent variable; α is the Y intercept; and b is the slope of the line.
	+ Remember that R-squared (R2) is the coefficient of determination that tells us the proportion of variation in the dependent variable explained by the independent variable. When R2 ~1, the data form a perfectly straight line. As the data become more scattered from the line, R2 decreases toward 0. Higher R-squared values indicate a stronger relationship between the two variables.
* Click on the camera icon from the menu on the right side of the scatterplot to save your graph as a .png file. Save it as a local file and give the file an appropriate name, e.g. “bee activity by year”. Click “Download” and note where the file is saved on the computer.

Now use your scatter plot and linear regression to answer the following questions:

1. Insert the graph you just exported into this document (or you can upload it as a separate file with your assignment).
2. Use the map or the data table to list the states where the observations were recorded.
3. What is the slope of the relationship (b) and what does that value tell you?
4. What is the R2 of the relationship and what does that value tell you?
5. Describe the relationship between time (year) and first day of bumblebee activity in your own words.
6. Based on these data, how would you answer the question: ***Are bumblebees in the western U.S. behaving differently from 2011 to 2019?***
7. How confident are you in your answer to the question above? (Pick a number 1 through 5)

*Not at all Confident 1 2 3 4 5 Very Confident*

1. What would help you become more confident in your answer?

**Part C. What site variables best predict bumblebee emergence phenology? (5 questions)**

We explored change in first day of bumblebee activity (emergence) over time (year), but what variables (if any) could help us understand variation in bumblebee activity? We can use environmental data recorded for each site to identify how much variation in bumblebee emergence phenology is explained by temperature or other climate-related variables.

Below are the variables in your dataset:

|  |  |
| --- | --- |
| **Variable** | **Description** |
| Latitude | Latitude of site; strongly relates to temperature and growing season length |
| Longitude | Longitude of site |
| Elevation\_in\_Meters | Elevation of site |
| State | State where observation was made |
| Phenophase\_Description | Type of observation: This dataset includes the date of the first adult activity observed and the date for the first flower visitation observed for each site reported in a year. |
| First\_Yes\_Year | Year of measurement |
| First\_Yes\_DOY | Day of Year when the first bumblebee activity was recorded. (Jan 1 is 1 and Dec 31 is 365) |
| Tmax\_Winter | Maximum winter temperature during the year |
| Tmax\_Spring | Maximum spring temperature during the year |
| Tmin\_Winter | Minimum winter temperature during the year |
| Tmin\_Spring | Minimum spring temperature during the year |
| Prcp\_Winter | Precipitation in winter in mm |
| Prcp\_Spring | Precipitation in spring in mm |
| Accum\_Prcp | Total precipitation in mm |
| Daylength | Length of the day in seconds (s) |

**! Before exploring the data, answer the following question:**

1. Which of the variables in the dataset do you think will be the best predictor of bumblebee emergence phenology? Explain why you chose this variable.

Explore how these variables are related to bumblebee emergence (First\_Yes\_DOY) or how they are related to year (First\_Yes\_Year) using the following steps:

* Return to your CODAP workspace. It should be the same as how you left it.
* Click on the blue label of the X axis and the list of variables will pop up. Select the a variable you are curious about.
	+ Think to yourself: *How did the scatterplot graph change? How did the linear regression line change? How did the slope change? How did the R2 change?*
* Continue to explore other variables you are curious about how they relate to bumblebee emergence.
	+ Remember you can use the camera icon to export a picture of the scatterplots you want to save!
* You can also explore how the climate variables have changed over time. Click on the blue label of the X-axis and select “First\_Yes\_Year” to put time (year) on the X-axis again. Then click on the blue label of the Y-axis and select one of the climate variables.
* After exploring the relationships between the variables, identify which variables you think are relatively good predictors of bumblebee emergence phenology and be sure to export images of those scatterplots.

Now use your scatter plots and linear regressions to answer the following questions:

1. Of the variables you tested, which are the best predictors of bumblebee emergence phenology? Insert the images of the scatterplots
2. How did you decide these were the best? Describe in detail the elements of the scatterplots and linear regression lines that helped you make your decisions.
3. What ideas do you have about bumblebees that could possibly explain the relationships between bumblebee emergence and these variables (you do not need to look anything up, just brainstorm some ideas)?
4. What other kinds of data would you like to test your ideas in #4, if they were available?

**Part D. Phenological mismatch (2 questions)**

Because species often interact with each other in ways that are critical for their survival and/or reproduction, if interacting species show different sensitivities to change in climate variables, they could end up with **phenological mismatches**.

1. Explain what a phenological mismatch is, in your own words.
2. Give an example (from class or another idea) of a possible phenological mismatch that could occur. Be sure to explain why the phenological mismatch.