***Exploring how climate will impact plant-insect distributions***

***and interactions using open data and informatics***

**SP-2: Species Distribution Maps (SDMs) and Hypothesis**

**- Instructions -**

**I. Brief Summary**: Naturalists are engaging in biodiversity research more than ever. Web-based biodiversity records have been widely adopted, data science has become increasingly sophisticated, and biodiversity technologies have become mobile. In particular, the creation of massive online citizen science web platforms and smartphone applications has resulted in real-time biodiversity data collection across continental and global scales by citizen scientists. These data acquisition of the world’s biodiversity is the type of data needed for precision conservation. With this data, mapping and data science converge to tell the up-to-date narratives of each of Earth’s inhabitants and to signal when to intervene and how well our conservation efforts are helping. Here we will map species distributions of butterflies and an important host plant using iNaturalist open source data and basic data science tools such as R, R Studio, and Git. From these maps, we will make testable hypotheses about the effects of climate change on the distributions and interactions between this butterfly and its larval host.

**II. Learning Goals & Objectives:**

1. Use data science tools (Citizen science crowd sourced data, R programming language, GitHub collaborative web platform, data visualization) to study a butterfly-host plant interaction.
	1. Apply common research computing tools including RStudio and GitHub to visualize butterfly-host plant distributions
	2. Search, filter and download butterfly and host-plant distribution data from biodiversity citizen science web platforms, specifically iNaturalist
	3. Differentiate descriptive and predictive species distribution models Create maps for species distribution models using R from open source biodiversity data for hypothesis generation
	4. Develop a hypothesis regarding the effects of climate change on a plant-insect interaction 50 years from now

**III. Before beginning:** Prior to class you should have downloaded R, RStudio, and GitHub. If you had any difficulty or concerns, please let your instructor know before beginning.

**IV. Instructions**

***How To Make a Species Observation Map:***

1. Find your taxa on iNaturalist (<https://www.inaturalist.org/taxa/>)
2. Print screen of species observation map
3. Crop as appropriate for Part 2 (see Canvas for details) making sure to provide the link to the iNaturalist site and the date of data visualization

 ***How to Make a Species Distribution Model:***

*Clone a Git Repository* ([https://github.com/jcoliver/biodiversity-sdm-lesson.git)](https://github.com/jcoliver/biodiversity-sdm-lesson.git%29). Recall the troubleshooting document HelpDocumentforCommonErrorsandHelpfulWebsite.docx if you run into any difficulties.

1. Start RStudio
2. Make a new project via File -> New Project…
3. Select **Version Control** (the third option) in the dialog that pops up
4. Select **Git** (the first option) in the dialog that pops up
5. In the Clone Git Repository dialog:
	1. In the **Repository URL** field, enter the GitHub URL: https://github.com/jcoliver/biodiversity-sdm-lesson.git (Note the .git extension on the URL)
	2. The **Project directory name** field should be automatically filled in for you (biodiversity-sdm-lesson)
	3. In the **Create project as subdirectory or field**, use the Browse... button to navigate to whatever folder you would like to put the project in. Select that folder by clicking on it once and click Create Project.
	4. Click **'Create Project'**. The files will download from GitHub and be available on your machine!

***Data Setup – recommendation: only one person in the group downloads the data but all members of the group discuss and decide on appropriate filters (see below).***

1. You will need to download occurrence data for both your butterfly and host plant. To do so, go to iNaturalist and search for one of the two species.
2. Click on the **Filters** button to the right of the search bar near the top of the screen. Here you can change the filters to sort your observations to only include verifiable or research grade (note these will affect the number of observations you have to work with). Be sure to record any filters you place on your search.
3. Click the **Download** button in the lower right hand area of this pop up window. This will bring you to a new screen with many different options. At the very least you should see the species you searched for in the **Taxon** box about 1/3 of the way down the screen. Scroll through this page and select the options you would like. Importantly, the more options you choose in **3. Choose Columns**, the longer it will take to download your data. You should look carefully and think about the data that might be useful to you when considering your results. At the very least, you must have the **latitude** and **longitude** columns checked as these are necessary for generating maps in the next part of the project.
4. Once you have made your selections, click Create export. This may take a few minutes depending on how much data you downloaded.
5. Save the file as a csv file in the **data** folder within the **biodiversity-sdm-lesson** folder you downloaded from Git through RStudio. Rename the file **genus\_species\_data.csv** (replace genus and species with the appropriate names). **Note: there can be no spaces in your file names – use an underscore ( \_ ) any place you would otherwise want to put a space.**
6. Repeat for the second species of interest.

*Running Analyses*

1. Go to the **biodiversity-sdm-lesson** folder you downloaded from Git via RStudio.
2. **Generating script files**: First, we need to make a few changes to prepare to run script files in R.
	1. In the **scripts** folder, find the file named **run-sdm-single.R** and make a duplicate copy. Rename this copy **<species>-sdm-single.R**, replacing **<species>** with the name of your butterfly. Use underscores instead of spaces; so for example, the butterfly species *Adelpha eulalia*, the file name would be **Adelpha\_eulalia-sdm-single.R**. If when copying the file any words such as ‘copy’ were added to the file name, delete them so your file name appears exactly as it should.
	2. Make a second copy of the file named **run-sdm-single.R.** Rename this copy **<species>-sdm-single.R**, replacing **<species>** with the name of your host plant. Use underscores instead of spaces; so for example, the plant species *Quercus gambelli*, the file name would be **Quercus\_gambelli-sdm-single.R**. If when copying the file any words such as ‘copy’ were added to the file name, delete them so your file name appears exactly as it should.
	3. Now find the file named **run-sdm-pairwise.R** and make a duplicate copy. Rename this copy **<species>-sdm-pairwise.R**, replacing <species> with the name of your butterfly. Use underscores instead of spaces; so for example, the butterfly species *Adelpha eulalia*, the file name would be **Adelpha\_eulalia-sdm-pairwise.R**. If when copying the file any words such as ‘copy’ were added to the file name, delete them so your file name appears exactly as it should.
3. **Making changes to script files**: Open these new files in RStudio by going to File -> Open File. You should see the file appear in the upper left hand window.
	1. For the butterfly ‘single’ file:
		1. On line 14 you should see:

**infile <- “data/MY\_SPECIES.csv"**  Change **"MY\_SPECIES.csv"** so it matches the file of butterfly data you saved above. Using an example, the final line would appear as follows:

**infile <- “data/Adelpha\_eulalia\_data.csv"**

* + 1. On line 15 you should see:

**outprefix <- “MY\_SPECIES”** Change **MY\_SPECIES** to have the name of your butterfly (e.g. Adelpha\_eulalia)

* + 1. Save the file with these updates
	1. For the host plant ‘single’ file:
		1. On line 14 you should see:

**infile <- “data/MY\_SPECIES.csv"** Change **"MY\_SPECIES.csv"** so it matches the file of host plant data you saved above (e.g. Quercus\_gambelli\_data.csv)

* + 1. On line 15 you should see:

**outprefix <- “MY\_SPECIES”**

Change **“MY\_SPECIES”** to have the name of your host plant (e.g. Quercus\_gambelli)

* + 1. Save the file with these updates
	1. For the butterfly ‘pairwise’ file:
		1. On line 14 you should see:

**butterfly.data.file <- "data/BUTTERFLY\_DATA.csv”**  Change **"BUTTERFLY\_DATA.csv"** so it matches the filename of butterfly data you saved above (e.g. Adelpha\_eulalia\_data.csv)

* + 1. On line 15 you should see:

**plant.data.file <- "data/PLANT\_DATA.csv"**  Change **"PLANT\_DATA.csv"** so it matches the file of plant data you saved above (e.g. Quercus\_gambelli\_data.csv)

* + 1. On line 16 you should see:

**outprefix <- “MY\_SPECIES”**

Change **“MY\_SPECIES”** to have the genus of the butterfly and host plant (e.g. Adelpha\_Quercus)

* + 1. Save the file with these updates
1. **Running script files:**
	1. In RStudio, change your working directory to the **biodiversity-sdm-lesson** folder. To do this, in the lower right panel of the RStudio window, navigate to the **biodiversity-sdm-lesson** folder. Once you are there, click the **More** button at the top of the pane and choose **Set as working directory**.
	2. Type the following command in the Console window:

**source(file = “scripts/setup.R”)**

* 1. To generate a map for the butterfly distribution:
		1. Run the analyses for the butterfly species distribution model by typing the following command in the Console:

**source(file = "scripts/<species>-sdm-single.R")**

replacing **<species>** with your butterfly species name (e.g. **scripts/Adelpha\_eulalia-sdm-single.R**)

* + 1. Your map should appear as a PDF in the **output** folder in the **biodiversity-sdm-lesson** folder.
	1. To generate a map for the host plant distribution:
		1. Run the analyses for the butterfly species distribution model by typing the following command in the Console:

**source(file = "scripts/<species>-sdm-single.R")**

replacing **<species>** with your host plant species name (e.g. **scripts/Quercus\_gambelli-sdm-single.R**)

* + 1. Your map should appear as a PDF in the **output** folder in the **biodiversity-sdm-lesson** folder.
	1. To generate a map for the butterfly and host plant distributions:
		1. Run the analyses for the butterfly species distribution model by typing the following command in the Console:

**source(file = "scripts/<species>-sdm-pairwise.R")**

replacing **<species>** with your butterfly species name (e.g. **scripts/Adelpha\_eulalia-sdm-pairwise.R**)

* + 1. Your map should appear as a PDF in the **output** folder in the **biodiversity-sdm-lesson** folder.
		2. Record the percent overlap that is reported in the console window.
	1. More information about what the scripts are doing can be found in SP2-AnalysisScriptInformation.

***Next Steps***: Use your maps to complete **SP-2 Assignment: Species Distribution Maps and Hypothesis** to report your findings and develop testable hypotheses on how climate change may affect your butterfly species distribution.