

Homework 5

Due on Tuesday 6 June (Week 10) at 11:59 PM

In this assignment, you will build on the work we did in workshop 8 to compare models and create a full report of your findings. Read the instructions before starting your work. A checklist of components that your assignment must have *at minimum* (i.e. to earn at least partial credit) is at the end.

Use the *Sarracenia* physiology data that we cleaned and subset in class to answer the question:

How do *Sarracenia* characteristics predict biomass?

Download the data [here](#). Read the metadata to understand where the data came from.

You will be graded on the logic and execution of your statistical methods, writing, data visualization, the appearance of your final rendered/knitted HTML document, and the context you set for your code/data (i.e. your GitHub repo organization and README).

Set up

1. Create a public repo (if you haven't already) called "ENVS-193DS_homework-05".
2. Enable GitHub pages.
3. Organize your repository using the [{here}](#) package.
4. Write an informative README (see [resource](#) from the library).
5. Commit/push at least 5 changes to the repository.

Assignment

6. Write a 5-10 sentence introduction to the analysis you conduct. Some useful information could include (but is not limited to):
 - a. What are *Sarracenia*?
 - b. Why might people care about them?
 - c. Why do you think it could be useful and/or important to predict individual biomass from morphological, physiological, or taxonomic characteristics?
 - d. What questions are you addressing?
 - e. What hypotheses are you testing?
 - f. Include 3-5 in-text citations of peer-reviewed work.

Note 1: need help finding peer-reviewed papers? One way to start digging into scientific literature is to use [Google Scholar](#) or [Web of Science](#) and search for whatever you're interested in (*Sarracenia* or otherwise). As with other reading assignments, be sure to use the [library log in](#) or a campus VPN (*do not pay for papers*).

Note 2: not sure how to cite things in-text? Start with MLA formatting (here are some [tips](#) from the Purdue Online Writing Lab). If you want to use another format, go for it! Just make sure it's consistent.

Note 3: if you are interested in using a citation manager to construct your bibliography, I recommend [Zotero](#). If you want to manage your citations within RStudio (with or without Zotero),

you can start by reading this [guide from RStudio](#) or follow along with [this guide using a class example](#). Neither of these tools are *required*; however, you will need to list a full bibliography no matter what.

7. Write a methods section that includes
 - a. Sampling methods (these should be summarized from the metadata and cited)
 - b. Description of data organization and processing (1-3 sentences)
 - c. Visualization of missing observations with an accompanying caption and discussion (caption: 1-3 sentences, discussion: 1-3 sentences)
 - d. Visualization of Pearson's correlation between variables with an accompanying caption and discussion (caption: 1-3 sentences, discussion: 1-3 sentences)
 - e. Visualization of relationships between variables (not correlation) with an accompanying caption and discussion (caption: 1-3 sentences, discussion: 1-3 sentences)
 - f. Creating the null and full model with accompanying discussion (1-3 sentences)
 - g. Visual and statistical assumption checks for the full model with accompanying discussion (1-3 sentences)
 - h. Explanation of any transformation and justification (1-3 sentences)
 - i. Model construction with visual and statistical assumption checks for *three* additional models with justification (i.e. why did you choose those predictor variables?) (justification: 1-3 sentences per model)
 - j. Variance inflation factor check for the full model with accompanying discussion (discussion: 1-3 sentences)
 - k. Model comparison using the null model, full model, and additional models from part 7g
8. Write a results section that includes
 - a. A description of the best model and how you chose it, with a summary of model results (parenthetical and/or table, 1-3 sentences)
 - b. Interpretation of your chosen model (1-3 sentences)
 - c. Visualization of model predictions for biomass as a function of the predictor variable of *your choice* with an accompanying caption (caption: 1-3 sentences)
 - d. Discussion of results: what does this model mean *biologically*? (1-3 sentences)
9. Include a bibliography for the sources you cited.

Checklist

Your rendered/knitted HTML document should include

- ☐ Your name, a title, and the date
- ☐ A link to your GitHub repository (wherever you want to put it)
- ☐ Annotated code without messages or warnings
- ☐ Folded code
- ☐ All tables and figures displayed *correctly* (for example, are your column names spread out? Are your figure labels clear?)

- ☐ Text for parts 6-8 in *paragraph* form organized with headers for each section (Introduction, Methods, Results)
- ☐ A bibliography/works cited section with a header

On Canvas by 11:59 PM on Tuesday 6 June, you will submit

- ☐ A link to the rendered/knitted HTML document hosted on your GitHub repository
- ☐ A link to your GitHub repository