**Problem Posing Template for Individual Activity**

**Module Overview**: Question brainstorming for a group project (~4 weeks); final project for the class

**Setting**: An upper-level, elective undergraduate course in “data carpentry,” or data science, in a wildlife ecology major. A mixture of lecture and lab-like work time. The class is typically about 30 students.

Learning Outcomes:

* Students will demonstrate ability to organize a research project
* Students will be able to write and annotate scripts for analysis
* Students will create appropriate data visualizations given the data and question asked
* Students will use version control and be able to collaborate on a project through interfacing with GitHub (i.e., using pull requests, etc.)

How does data acumen align with this learning outcome? Place an “X” in the column next to the skills practiced in this activity

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Quantitative Pillars** |  | **Data Life Cycle** |  | **Social/Pedagogical Concepts** |  |
| Mathematical |  | Data import | X | Communication | X |
| Computational | X | Management | X | Equity, Diversity, Inclusivity |  |
| Statistical thinking | X | Curation | X | Universal Design for Learning | X |
| Reproducibility | X | Analysis | X | Ethics |  |
|  |  | Sharing/ Reporting | X |  |  |

**Activity/Module**:

This is the final project for the undergraduate data carpentry class, which is a flipped class; students read through materials going over new coding techniques before they come to class, and class consists of small bits of “lecture” and working through exercises.

Students use the teaching dataset from the Portal Project (<https://github.com/weecology/portal-teachingdb>). The product is a collaborative group project which includes data analysis, data visualization, and version control, all through R, Rstudio, and the git/GitHub integration with Rstudio. The final product should be an R project that, when sourced, will run without any issues. The project should include an Rmarkdown file (and potentially additional scripts for analysis) that goes through the analytical steps used in the analysis, the data visualization to show the results, and the conclusions. Students should also be using git/GitHub collaboratively (i.e., using pull requests to merge changes to files).

The problem posing will be used to help students and groups narrow down what questions they want to ask using the Portal data. Beforehand, students will be told about the structure and content of the dataset and given an opportunity to look through the data for a few minutes to get a sense of what the data includes. Once this quick exploratory phase is over, the question focus will be introduced: “population or community ecology principles in the dataset.”

**Assessment**:

Sound ecological questions will be encouraged, but the ecological “validity” of the question will not have bearing on the grade. Learning outcomes will be assessed via a rubric. Points will be determined by whether the analyses are reproducible, the code is clean and annotated, the explanations for the analyses done are sound, the data visualization is appropriate given the data types and the questions asked, how well students followed good practice in version control, etc.

Since this is the final project, students will not be using this skill again. Since this is an elective, however, most students in the class are considering careers in ecology and will likely use this project format many times down the road in their careers (and potentially other upper-level classes).

**Extra information**:

Students will need to know the structure of the database, what data is included in the dataset, and what the expectations are for the final project. These resulting questions from this problem posing aren’t necessarily to determine the final question but get people thinking.

Still very much a work in progress!