The Librarian's Toolkit for Hosting Biomedical Data Competitions

A Sample Long-Term, Team-Based Biomedical Data Competition Model

Bethany McGowan

Published on: Apr 27, 2021

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Session Specifics

This lesson plan is intended for a long-term 3-week extracurricular biomedical datathon. The approach described is face-to-face, however hybrid and virtual variations are offered in the Reflection section. This approach compliments the “Information as Creation” frame of the Association of College and Research Libraries Framework for Information Literacy for Higher Education and could be adapted for use in credit-bearing courses, especially in undergraduate education.

The promotion of this long-term, team-based hacking model is the result of an NIH/National Network of Libraries of Medicine-funded study which found that well-designed, team-based, and long-term hacking models where a team lead was assigned and that allowed participants to iteratively have their progress judged before final submissions were especially effective at attracting and retaining diverse participants.

The development of resources reported in this presentation are supported by the National Library of Medicine (NLM), National Institutes of Health (NIH) under cooperative agreement number 1UG4LM012346. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

Instructional Partners

This lesson plan was conceived as an extracurricular instructional experience taught by two information literacy instructors, faculty in the Libraries and School of Information Studies. The approach lends itself to collaboration with faculty in Health Sciences, Technology, and Engineering disciplines and with industry partners willing to share proprietary data.

Audience

Undergraduate college students, with strategies for recruiting and retaining participants typically underrepresented in data competitions, particularly women and people of color.

Curricular Context

Overview

This lesson engages students in the “Information as Creation” frame by asking learners to develop, by way of their own creation processes, an understanding that their choices and biases impact the purposes for how data and information will be used and the messages they convey. Asked to develop data literacy competencies that support the creative process we call ‘hacking’, students will gain a better understanding of the data-information-knowledge-wisdom continuum, and will practice using data to produce new knowledge.

This lesson may fit in as an extracurricular activity or as a curricular/co-curricular activity in a course dedicated to data management and data literacy concepts.

In an extracurricular setting, the activity would work well for students enrolled in a data science or health science-related Learning Community. If paired with an industry partner, the activity would also work well for students interested in entrepreneurship.

In a curricular or co-curricular setting, the activity could serve to introduce the concept of information as creation early in a course. It would also work particularly well later in a course, after lessons covering basic data literacy concepts, such as cleaning and formatting data; FAIR data use principles; and basic data analysis.

Rationale

Why hacking?

Biomedical data competitions, when guided by librarian instruction, support participant's understanding of research data management and health data literacy-related competencies. Participation in biomedical data challenges helps current and future health professionals identify, consider, and develop solutions to challenges associated with how big biomedical data influences patient well-being, and data competitions are particularly useful for better understanding issues associated with the underuse of uniform data standards and issues associated with navigating siloed data.

Learning Outcome

After this competition, students will be able to:

- develop, in their own creation processes, an understanding that their choices impact the purposes for which an information product will be used and the message it

conveys.

Preparation

Instructor Prep

- Prepare the data literacy assessment on familiarity with data literacy competencies. This can be used as a pre/post assessment.
- Select background readings and viewings on the data literacy competencies required for participation in the challenge—e.g. fair data use, data analysis, and data visualization. See the 'Resources for Data Literacy Training' section of this community for suggestions.
- Decide on a competition theme, rules, and/or guidelines. Use the resources available in the 'Hackathons' section of this community to guide your planning.
- Create a handout, presentation, or guide to describe the competitions rules and processes. See examples: [Purdue University Geography of Nutrition Datathon](#) and [NIH Women-led Biodata Science Hackathon](#)
- Develop a process for assigning individuals to teams, or decide if individual should self-select team members. If the instructor is assigning teams, use results from the data literacy pre-assessment to assign teams with a mix of strengths.
- Assign a team lead for each team. This person will be responsible for planning team meeting and guiding progress.
- Secure sponsorships for prizes. Consider campus and industry partners.

Student Prep

- Read and view selected assigned readings and viewings.
- Respond to pre-assessment questionnaire

Materials

Instructor

- Computer and internet access
- Data competition hosting platform, such as Kaggle (academic version available), DrivenData, or GitHub
- Data competition judges—consider campus industry, and local/national government agency partners who might also provide open or proprietary datasets. You will not need judges for predictive modeling competitions, if using a platform like Kaggle. However, you will need to develop a model to 'feed' the platform.

- Datasets: for beginners, consider providing cleaned and formatted data, i.e. [Kaggle's datasets for beginners](#). Carefully consider your target audience and avoid datasets that might be too complex.
- computer labs or classroom spaces where teams can meet

Students

- Computer and Internet access

Session Outline

Pre-Competition

1. Distribute pre-assessment.
2. Point participants to the wide range of online resources available for honing the skills needed to thrive in data competitions, for example all or select portions of the Coursera [MOOC course](#) on How to Win a Data Science Competition. Consider using results from the pre-assessment to lead live trainings on select topics.
3. Develop a scoring rubric
4. Develop a project timeline, ideally 3-4 weeks.
5. Review project timeline and scoring rubric with judges, or upload a predictive modeling project to the data competition hosting platform.
6. Host at least 1 live Q&A session
7. Assign teams and team leads.

During the Competition

1. Send out weekly reminders just before submissions are due to judges or due to be uploaded to the data competition hosting platform.
2. Meet weekly with each team to discuss progress and/or concerns.

At Competition End

1. Select winners! Consider multiple categories, such as 'most improved' or 'best collaborative spirit'.
2. Distribute post-assessment

Data Literacy Assessment

The following questionnaire is intended to be used as a pre and post assessment:

Score each of the 17 traits according to how proficient you feel you are on a scale of 0 to 4, and how important each trait is to your professional success.

We will use your responses to determine your Learning Priority Score for each trait, the Importance Score minus the Proficiency Score. This will range from -4 (very low learning priority) to +4 (very high learning priority).

Circle the appropriate response below, where 0 is the lowest response and 4 is the highest response.

Knowledge

1. I am knowledgeable of the data-information-knowledge-wisdom continuum

1. Perceived Proficiency: 0 1 2 3 4

2. Perceived Importance: 0 1 2 3 4

2. I am knowledgeable of the basic elements of data

1. Perceived Proficiency: 0 1 2 3 4

2. Perceived Importance: 0 1 2 3 4

3. I am knowledgeable of data storage methods

1. Perceived Proficiency: 0 1 2 3 4

2. Perceived Importance: 0 1 2 3 4

4. I am knowledgeable of data management methods

1. Perceived Proficiency: 0 1 2 3 4

2. Perceived Importance: 0 1 2 3 4

5. I am knowledgeable of basic data analysis methods

1. Perceived Proficiency: 0 1 2 3 4

2. Perceived Importance: 0 1 2 3 4

6. I am knowledgeable of data visualization principles

1. Perceived Proficiency: 0 1 2 3 4

2. Perceived Importance: 0 1 2 3 4

Skills

7. I am able to accurately read visual displays of data

1. Perceived Proficiency: 0 1 2 3 4

2. Perceived Importance: 0 1 2 3 4

8. I am able to prepare data for analysis

1. Perceived Proficiency: 0 1 2 3 4

2. Perceived Importance: 0 1 2 3 4

9. I am able to use data to answer research questions and make evidence-based decisions

1. Perceived Proficiency: 0 1 2 3 4

2. Perceived Importance: 0 1 2 3 4

10. I am able to create clear data visuals

1. Perceived Proficiency: 0 1 2 3 4

2. Perceived Importance: 0 1 2 3 4

11. I am able to effectively communicate findings from data analysis

1. Perceived Proficiency: 0 1 2 3 4

2. Perceived Importance: 0 1 2 3 4

Attitudes about Data

12. I am able to assess the inclusivity of data gathering methods

1. Perceived Proficiency: 0 1 2 3 4

2. Perceived Importance: 0 1 2 3 4

13. I am able to determine if data is being used ethically

1. Perceived Proficiency: 0 1 2 3 4

2. Perceived Importance: 0 1 2 3 4

14. I feel confident in my ability to use data to make decisions.

1. Perceived Proficiency: 0 1 2 3 4

2. Perceived Importance: 0 1 2 3 4

15. I feel confident in my awareness of best data management practices

1. Perceived Proficiency: 0 1 2 3 4

2. Perceived Importance: 0 1 2 3 4

Behaviors

16. I resourcefully identify, navigate, and use data to make decisions.

1. Perceived Proficiency: 0 1 2 3 4

2. Perceived Importance: 0 1 2 3 4

17. I continuously improve my data gathering, analysis, and visualization practices.

1. Perceived Proficiency: 0 1 2 3 4

2. Perceived Importance: 0 1 2 3 4

18. I effectively advocate for FAIR data principles.

1. Perceived Proficiency: 0 1 2 3 4

2. Perceived Importance: 0 1 2 3 4

19. I enthusiastically spread best data literacy practices in my research and publication practices.

1. Perceived Proficiency: 0 1 2 3 4
2. Perceived Importance: 0 1 2 3 4

Reflection

Lesson variations - Some elements of this lesson work best in person, namely, the team-based components are easiest to execute face-to-face. However, with robust internet connections and access to video conferencing software even these elements can be move online, allowing for the creation of a completely virtual experience.

Additional Instructional Materials

Explore this community for additional instructional materials.