Provost's Office Innovative Course Design Grant

Interactive simulations to support inquiry-based statistics instruction in Frontiers of Science

Nicholas Bock[1], Vincent FitzPatrick[2], Debora Monego[3]

1. Department of Earth and Environmental Sciences, nab2161@columbia.edu
2. Department of Biological Sciences, vdf2104@columbia.edu
3. Department of Chemistry, dlm2206@columbia.edu

Proposal Overview

Frontiers of Science is a required course for all first-year Columbia College students. The goal of the course is to provide an introduction to current research across four different scientific disciplines while helping students develop a framework for approaching, analyzing, and interpreting quantitative information (hereafter referred to as the scientific "habits of mind"). Students come to Frontiers of Science with a wide range of prior science coursework and attitudes/interests/experiences in science. Therefore, a central challenge in planning individual seminar sessions is identifying learning activities that address course objectives while serving the learning needs of all students in the classroom.

This challenge can be particularly acute when introducing statistical concepts. Frontiers of Science students are introduced to basic statistical concepts (e.g., measures of central tendency and dispersion) at the start of the semester. During the first three weeks of the semester, students are expected to develop an understanding of how these concepts can be used to formulate statistically supported inferences (e.g., the use of standard error to evaluate confidence intervals and to support hypothesis tests). However, based on an introductory survey completed by students in seminar sections during the spring 2022 semester, only 50 percent of students reported having completed previous coursework in statistics. Therefore, there is a clear need within Frontiers of Science for instructional tools that are appropriate and challenging for students both with and without prior experience in statistics.

With this grant, **we propose to develop a series of interactive simulations to support inquiry-based learning of basic statistical concepts in Frontiers of Science seminar sessions**. We additionally propose to work with SOLER Initiative faculty to investigate the impact of the proposed instructional materials on learning outcomes in Frontiers of Science seminar sections. Because all incoming first-year students enroll in Frontiers of Science, student interactions with these simulations would potentially impact learning outcomes across Columbia College. Moreover, because the statistical concepts to be introduced are foundational to data analysis across many different fields of study, these simulations would be of potential use to faculty across the university. We have already been in contact with Dr. Simon Mason at the Columbia

Climate School, who expressed interest in adapting the simulations to introduce Earth Science students to statistical analysis of environmental data. The simulations would additionally expand upon interactive simulations previously developed by faculty in the School of International and Public Affairs (SIPA) as part of the Stats Interactives website (http://stats-interactives.ctl.columbia.edu/).

Structure of Course

The primary goals of Frontiers of Science are to introduce students to cutting-edge research topics across a range of scientific disciplines and help students develop the critical thinking frameworks employed by scientists to make sense of the natural world. The course is taught through weekly 1.5-hour lectures led by senior faculty from across the natural science departments and 2-hour seminars taught by Frontiers of Science Fellows and other faculty. The lecture sessions introduce material relevant to discipline-specific content objectives. Discussion sessions support students in practicing course process objectives (e.g., the habits of mind) within the context of the topics introduced in lecture sessions.

Plan of Project

Key Learning Objectives

Based on the results of in-class surveys conducted during the spring 2022 semester, students enter Frontiers of Science with a wide range of previous experiences in quantitative data analysis. Even students who have completed previous coursework in statistics have limited first-hand experience using statistical concepts to analyze authentic data sets. Assessments conducted during the spring 2022 semester (e.g., class participation, responses on homework, and midterm exams) additionally highlight that students have difficulty identifying what distinguishes different measures of dispersion (e.g., standard deviation, standard error) and why the standard error specifically is used to perform statistical tests.

Current Course Structure

The existing Frontiers of Science curriculum includes four content objectives related to statistical concepts:

- Demonstrate an understanding of mean, standard deviation, and standard error
- Estimate the mean and standard deviation from a sample distribution
- Calculate confidence intervals and determine statistical significance from a bar graph
- Determine statistical significance from p-values

At present, these objectives are introduced and assessed during the first three weeks of the course via 1) a series of lessons via the SmartSparrow adaptive learning platform; 2) direct instruction via slide-based mini-lectures conducted during seminar sessions;

and 3) small-group and whole-class discussions in seminar sessions examining how these concepts are used to interpret and communicate results in peer-reviewed scientific research articles. Concepts targeted by these objectives are also drawn on in subsequent units and, depending on the seminar instructor, in students' end-of-semester projects.

Proposed redesign

The proposed redesign would develop 1) a series of online interactive statistical simulations to complement existing instructional materials, and 2) a set of instructional materials that would embed these simulations within an inquiry-based learning activity to be conducted in Frontiers of Science seminar sessions. Critically, inquiry-based learning pedagogies have been shown to improve learning outcomes in undergraduate STEM classrooms (Udovic et al., 2002; Knight and Wood, 2005) while simultaneously enhancing the development of transferable skills (Kuhn et al., 2016). This makes inquiry-based approaches especially relevant to courses like Frontiers of Science, where many students are non-STEM majors.

Statistical simulations will be based on a series of existing simulations developed by one of the PIs (Nicholas Bock) using the R-based Shiny platform during the spring 2022 semester. These simulations allow students to iteratively sample from a large, publicly-available database of Spotify audio data (https://nickbock.shinyapps.io/sdse/), thereby demonstrating:

1. The role of random variability within an individual sample (Figure 1A)
2. How measures like the standard error account for this random variability as a function of sample size (Figure 1B)
3. How standard error can be leveraged in hypothesis tests to infer whether differences between groups are statistically significant (Figure 1C)

Spotify data were chosen for these simulations because their interpretation does not require any discipline-specific content knowledge, thereby helping to minimize any additional barriers to accessing the statistics objectives outlined above.

While students were enthusiastic about the simulations and accessed the app independently outside of class (based on app usage data provided by shinyapps.io), implementation of these simulations was curtailed by only having a subscription to the free tier of the Shiny app hosting service. As a result, it was difficult for multiple students to use the app simultaneously, limiting its usefulness during in-class instruction beyond a visual aid during seminar mini-lectures. Therefore, the focus of the proposed redesign is to collaborate with Center for Teaching and Learning (CTL) staff to re-develop the existing simulations in JavaScript, thereby making it possible to host the simulations on university domains and resolving limitations on the number of users that can access the simulations simultaneously. As part of the redesign, we would also explore options for allowing users to select the dataset used in simulations, thereby making simulations useful to STEM faculty interested in using the simulations within the context of specific

disciplines (for instance, by using environmental data in place of the default Spotify data).

The proposed redesign would additionally develop a set of instructional materials to support Frontiers of Science instructors in utilizing these simulations within a process-oriented guided inquiry learning (POGIL) framework: an active learning approach demonstrated to improve learning outcomes for undergraduate students across a wide range of content areas (Eberlein et al., 2008). Students participating in the activity would follow the typical POGIL learning cycle: 1) working in small groups to interact with the simulations; 2) responding to a series of guiding questions to identify targeted statistical concepts demonstrated by simulations; and 3) applying these concepts to groups of data to evaluate student-defined research questions (for example, what musical attributes distinguish music from two different genres?). To ensure that the simulations are accessible to students with limited previous experience in graph interpretation, the activity would include a preliminary set of exercises where students would become familiarized with the graphs used in the simulations.

Technologies and media for proposed redesign

All simulations will be developed in JavaScript with in-kind support from the CTL learning design and software development teams. Simulations will be published on university domains, potentially as part of the existing Stats Interactives website (https://stats-interactives.ctl.columbia.edu/) developed by the School of International and Public Affairs (SIPA). If the JavaScript version of the app is not ready in time for the fall 2022 semester, simulations will be published on the Shinyapps.io hosting service, with funding requested to pay for an annual subscription to the service.

Testing and evaluation of redesign

At the start of the fall 2022 semester, seminar instructors will receive training in the use of the simulations and the corresponding POGIL activity. The effectiveness of the redesign will be evaluated based on two metrics: student responses on a pre and post assessment and on a Student Assessment of Learning (SALG) survey (Seymour et al., 2000). Pre and post assessments will consist of identical short-response questions corresponding to each of the course statistics objectives. The SALG survey will consist of a short series of questions interrogating the extent to which different instructional materials have contributed to students' perception of their learning of statistical concepts. Surveys will be conducted via Google Forms or the SALG website (https://salgsite.net/). Student responses, in addition to feedback from seminar instructors, will be used to further refine the proposed instructional materials.

Assessment of redesign as a pedagogical methodology

During the spring 2023 semester, A/B testing will be used to evaluate the extent to which inquiry-based learning activities facilitate mastery of statistical concepts amongst undergraduate students. Seminar instructors participating in the assessment will teach

one seminar section using existing instructional materials (e.g., SmartSparrow and slide-based mini-lectures) and one section using the redesign instructional materials. The effectiveness of either approach will be based on pre and post assessments, as described above. Simulations will be made available to all students following the assessment to ensure that all students are able to benefit from interacting with the redesign course materials. Approval of the proposed assessment will be sought from the Institutional Review Board (IRB) before the end of the fall 2022 semester, and permission will be obtained from all students participating in the assessment. We will submit a manuscript reporting the effectiveness of redesign (e.g., Journal of Statistics and Data Science Education, Technology Innovations in Statistics Education) upon completion of our study.

Budget and budget justification

The total budget for the proposed redesign is $5025, as itemized below:

Software and development

Shinyapps.io Standard subscription for one year – $1100

Implementation and testing

Eight former students for pre-testing of application, each receiving one gift card of $50 value – $400

Training

Registration fees for summer 2022 Fundamentals of POGIL virtual workshop for each PI – $175/instructor – $525

Publication costs

Fees to publish manuscript reporting effectiveness of redesign (e.g., Journal of Statistics and Data Science Education, Technology Innovations in Statistics Education) – $3000

In-kind support from CTL

For the completion and refinement of the proposed redesign, it is requested that the Center for Teaching and Learning will provide in-kind support for

- Redevelopment of existing simulations in JavaScript, allowing for simulations to be published on university domains
- Consultation in developing POGIL activities centered around the use of simulations during class sessions
- Consultation in designing and developing tools to assess the effectiveness of the proposed redesign

In-kind support from SOLER Initiative

For the assessment of the proposed redesign, it is requested that the SOLER Initiative provide in-kind support for

- Obtaining IRB approval for assessment of the redesign during the spring 2023 semester
- Developing rigorous pre and post questionnaires (or any additional measures of learning outcomes) to be used in assessing the effectiveness of redesign materials during the spring 2023 semester

Sustainability of Project

The authors of this proposal intend to use the redesign materials for multiple semesters. As part of this project, the authors will additionally develop detailed instructional guides to ensure that materials can be readily implemented by future Frontiers of Science instructors. Collaborations will also be sought out with faculty in other departments, including Dr. Simon Mason at the Columbia Climate School, to promote the adoption of redesign materials outside of Frontiers of Science. Finally, re-developing simulations in JavaScript, with the support of CTL software and development teams, will ensure long-term maintenance of simulations on university domains without incurring annual subscription fees for hosting services.

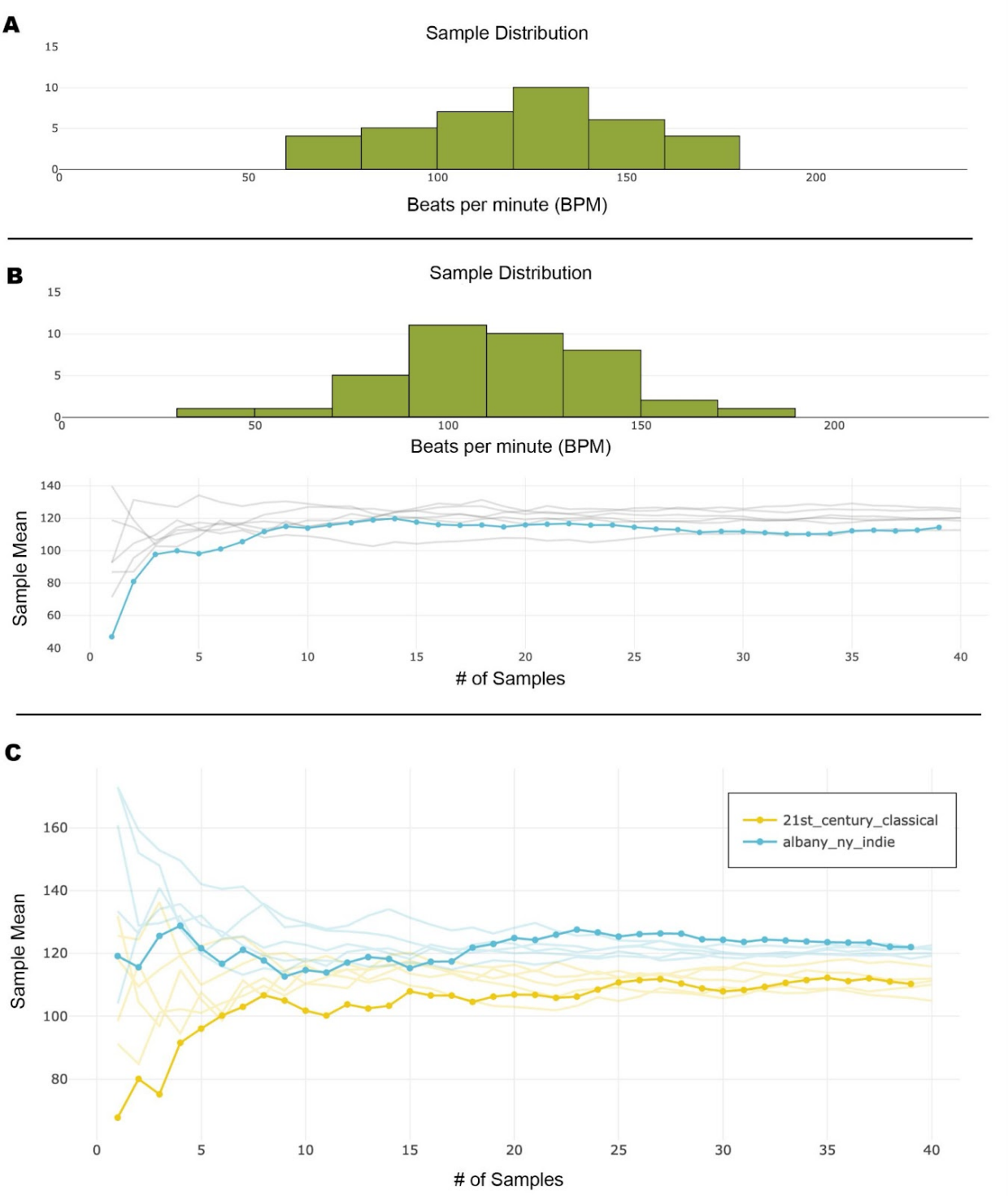


Figure 1: A) Screenshot of histogram developed in part 1 of the app. **B)** Screenshot of plots generated in part 2 of the app; histogram in green corresponds to the distribution of the current sample; blue line corresponds to cumulative mean for current sample; gray lines correspond to cumulative means for previous samples. **C)** Screenshot of plots

generates in part 3 of the app, comparing cumulative sample means for two different genres: 21st Century Classical music (yellow) and Albany, NY Indie music (blue).

References

Eberlein, T., Kampmeier, J., Minderhout, V., Moog, R. S., Platt, T., Varma-Nelson, P., et al. (2008). Pedagogies of engagement in science: A comparison of PBL, POGIL, and PLTL. *Biochem. Mol. Biol. Educ.* 36, 262–273. doi:10.1002/bmb.20204.

Knight, J. K., and Wood, W. B. (2005). I1536-7509-4-4-298. 4, 298–310. doi:10.1187/05.

Kuhn, D., Arvidsson, T. S., Lesperance, R., and Corprew, R. (2017). Can Engaging in Science Practices Promote Deep Understanding of Them? *Sci. Educ.* 101, 232–250. doi:10.1002/sce.21263.

Seymour, E., Wiese, D. J., Hunter, A.-B., and Daffinrud, S. (2000). Creating a Better Mousetrap: On-line student assessment of their learning gains. *Natl. Meet. Am. Chem. Soc.*, 1–40.

Udovic, D., Morris, D., Dickman, A., Postlethwait, J., and Wetherwax, P. (2002). Workshop biology: Demonstrating the effectiveness of active learning in an introductory biology course. *Bioscience* 52, 272–281. doi:10.1641/0006-3568(2002)052[0272:WBDTEO]2.0.CO;2.

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK
DEPARTMENT OF EARTH & ENVIRONMENTAL SCIENCES
LAMONT-DOHERTY EARTH OBSERVATORY

May 1, 2022

Columbia University
205 Low Memorial Library, MC 4313
535 West 116th Street
New York NY 10027
Email: ColumbiaCTL+hybrid@columbia.edu

Dear Colleagues:

I am writing to offer my enthusiastic endorsement of the application by Nicholas Bock, Vincent FitzPatrick and Debora Monego for funding from the CTL Hybrid Learning Grant competition for their project "Interactive simulations to support inquiry-based statistics instruction in Frontiers of Science." This is a very promising proposal to develop a series of interactive statistics simulations that would support and complement the existing instructional materials, and help students to achieve mastery of course objectives within an active learning framework. A greater understanding of statistics would enhance student learning across the sciences, and active learning approaches are especially relevant to Frontiers of Science, where the overwhelming majority of students are non-STEM majors. As someone who has taught in this Columbia Core course for over a decade, I believe that the proposed approach will contribute to the greater achievement of course learning objectives while enhancing the students' overall learning experience and retention. A successful implementation here also has the potential for wider applications beyond this individual course. I am therefore enthusiastic about both this approach and the proponents, and very much hope that CTL will see fit to support this proposal.

Sincerely,

Jerry F. McManus, Professor and Chair
Dept. of Earth & Environmental Sciences
Lamont-Doherty Earth Observatory

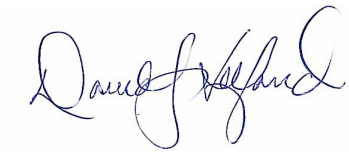
COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK
COLUMBIA ASTROPHYSICS LABORATORY

30 April 2022

To whom it may concern:

As chair of the Columbia College Core Curriculum course *Frontiers of Science*, I hereby approve recruitment of our students and seminar leaders for the 2022 Grant Proposal project entitled "Interactive Simulations to Support Inquiry-Based Statistics Instruction in *Frontiers of Science*". The results of this study, interesting in their own right, would likely be valuable as we continue to evolve the course content and delivery techniques.

Sincerely,



David J. Helfand
Professor of Astronomy, Columbia University
Chair, Frontiers of Science
Chair, American Institute of Physics & AIP Publishing
Legacy Fellow and Past President, American Astronomical Society
President Emeritus, Quest University Canada

550 West 120th Street Pupin Hall Mail Code 5247 New York, NY 10027 212-854-7815 Fax 212-854-8121