Developing statistical and data science skills in interdisciplinary scientists

Chris Brignell
University of Nottingham, UK
chris.brignell@nottingham.ac.uk

The University of Nottingham's Natural Sciences course is an interdisciplinary science undergraduate degree. A new module was developed to teach introductory statistics based on the recommendations and goals of the GAISE report (ASA Revision Committee, 2016) with a focus on statistical thinking, conceptual understanding, real contexts, active learning, use of technology, and assessment which drives learning. In the pilot, second year undergraduate Natural Sciences students participated, with qualitative feedback on module design and delivery collected through survey instruments. Student feedback was positive, with them acknowledging that interactive classes and computer classes were engaging ways to learn using real-world contexts and that continuous assessment added depth to their learning. This case study demonstrates the potential to teach statistics to scientists through context-relevant scenarios, ultimately enabling students to apply their knowledge to projects involving their own scientific data.

INTRODUCTION AND CONTEXT

The University of Nottingham's Natural Sciences course is an interdisciplinary science undergraduate degree, with students studying different combinations of archaeology, biology, chemistry, environmental science, geography, mathematics, physics and psychology. These science, technology, engineering and mathematics (STEM) students need to develop an understanding of statistics and data analysis. The aim of this study was to reform the existing provision of introductory statistics education designed for mathematicians to this interdisciplinary context.

Students at the University of Nottingham, UK, study 120 credits per academic year. The existing introductory statistics provision for mathematics students was two 10 credit semester-long modules, taken sequentially, on probability and statistics respectively. Both modules consisted of 20 1-hour lectures, accompanied by 5 hours of support classes where students worked on problems, and the statistics module had an additional 2 hours of computer classes where students learned R. The primary method of assessment in each module was a 2-hour exam worth 80% or more of the module mark with the remaining assessment being multiple-choice questions in a mid-term test (for probability) or coursework (for statistics). The curriculum for both modules is set out in Table 1.

Table 1. The curriculum for mathematics students.

Probability	Statistics
Sample spaces and events	Summary statistics
Counting problems	Plots and graphs
Conditional probability	Unbiased estimators
Random variables	Tests and confidence intervals of means
Expectation of random variables	Tests and confidence intervals of variances
Discrete distributions	Hypothesis tests of correlation
Continuous distributions	Tests and confidence intervals of proportions
Transformations of random variables	Simple linear regression
Bivariate random variables	Chi-squared tests of independence
Probability generating functions	Use of statistics tables
Central limit theorem	Report writing and R

Before developing new provision for interdisciplinary STEM students, this existing provision was evaluated against the principles set out in the next section.

In: Birk, L., Loth, G., Jotzo, L., Binder, K., & Frischemeier, D. (Eds.) (2026). *Statistics and Data Science Education in STEAM. Proceedings of the Satellite Conference of the International Association for Statistical Education (IASE), September/October 2025, Münster, Germany*.

PRINCIPLES FOR CURRICULUM, TEACHING AND ASSESSMENT

The American Statistical Association published the Guidelines for Assessment and Instruction in Statistics Education (GAISE) (ASA Revision Committee, 2016). Their six recommendations were:

- Teach *statistical thinking*: as an investigative process of problem-solving and decision-making and give experience of multivariable thinking.
- Focus on *conceptual* understanding.
- Integrate *real data* with a context and purpose.
- Foster *active learning*.
- Use *technology* to explore concepts and analyze data.
- Use *assessments* to improve and evaluate student learning.

The existing provision of introductory statistics education, designed for mathematics students, set out in the previous section did not conform to the GAISE recommendations for several reasons. Firstly, although students had to submit a report, the scope for the analysis was tightly constrained and did not encourage open-ended *investigation*. There was also limited multivariate thinking aside from bivariate cases such as linear correlation, simple linear regression and test of independence. Secondly, the emphasis was on procedural tasks, such as hypothesis tests, which were manually calculated and could be correctly calculated without understanding of the *concepts*. Thirdly, using closed-book exams as the primary assessment method placed an undue emphasis on manual calculations. This had consequences for the use of *real data*, as small datasets were required which lacked real-world context and purpose. It also reduced the role of *technology* with some students passing the modules without engaging with the software. Overall, the *assessment* strategy was largely summative and did not improve student learning. Fourthly, the primary method of teaching was lectures which encouraged passive engagement, with *active engagement* limited to support classes.

Therefore, it was concluded that the existing provision did not meet the GAISE recommendations and should be reformed for developing provision for interdisciplinary science students. Furthermore, for interdisciplinary science students, the existing provision placed too much emphasis on probability theory rather than application, and the division of probability and statistics into separate modules was unhelpful for linking theory and application. It was also felt manual graph plotting and the use of statistics tables were unnecessary in a statistics module for an interdisciplinary applied context when computational tools were available.

In designing the new introductory statistics provision for interdisciplinary students, it was helpful to consider the goals of statistics education. Table 2 shows a summary of "what a student should know and understand at the conclusion of a first course in statistics" according to the GAISE report (ASA Revision Committee, 2016) and the earlier goals for students learning statistics according to Garfield and Chance (2000).

Table 2. Goals of statistical education.

GAISE (2016)	Garfield and Chance (2000)
Become critical consumers of popular media.	Understand the purpose of statistical investigations.
Answer questions using an investigative process.	Understand the process of statistical investigations.
Interpret graphs and numerical summaries.	Learn statistical skills.
Understand the importance of variability.	Understand probability and chance.
Understand the importance of randomness.	Develop statistical literacy.
Gain experience of statistical models.	Develop a statistical disposition.
Understand statistical inference.	Develop statistical reasoning.
Interpret output from statistical software.	
Be aware of ethical issues.	

DESIGN OF THE NEW MODULE

The new module for interdisciplinary science students was designed using the principles and goals set out in the previous section, with adaptation for an interdisciplinary science context.

Curriculum

The new module removed the unhelpful division of probability and statistics by creating a single 20 credit year-long module. The curriculum was modified by removing some elements of probability, such as counting problems, transformations of random variables and probability generating functions, and the use of statistical tables. The theme for each week in the redesigned module is shown in Table 3.

Table 3. Curriculum for interdisciplinary scientists.

Week	Semester 1	Semester 2
1	Univariate distributions and histograms	Joint distributions and conditional probability
2	Robust measures and correlations	Chi-squared tests
3	Probability, random variables and expectation	Data collection and sampling variability
4	Discrete distributions	Simple linear regression
5	Continuous distributions	Multiple linear regression
6	Estimates and confidence intervals	Analysis of Variance (ANOVA)
7	Significance testing	Design of experiments
8	Tests of means	Logistic regression
9	Sample size calculations	Individual project
10	Multiple comparisons problem	Individual project

The curriculum set out in Table 3 contains topics relevant for interdisciplinary scientists who will be using statistics in an applied context. For example, sample size calculations take the concepts of Type I errors, Type II errors and statistical power that the mathematicians learned but also applies them to the context of designing experiments. This also motivated discussion of the difference between statistical significance and meaningful effect size. Similarly, introducing the multiple comparisons problem not only builds understanding of an important issue for applied scientists conducting hundreds of experiments, but simulation of the issue also reinforces understanding of the concept of Type I errors introduced the previous week, before exploring correction methods such as Bonferroni and Benjamini-Hochberg algorithms for controlling the family-wise error rate and false discovery rate, respectively. In the second semester topics essential for applied scientists such as data collection principles (e.g. randomization, measurement precision) and design of experiments concepts (e.g. blocking, Latin square) were introduced. In alignment with the GAISE goals, scientific principles (e.g. repeatability, reproducibility) were introduced to motivate discussion of data ethics and the validity of statistical findings. Similarly, discussion of scientific findings as reported in the media and science journals enabled students to be critical consumers of statistics and increase their statistical literacy and reasoning as suggested by Garfield and Chance (2000).

Teaching

The overall pedagogic approach was to encourage *active learning* as recommended by GAISE. For this context, engagement consisted of a 1-hour interactive class and a 2-hour computer class each week. The interactive class was teacher-led but included activities which introduced the main concepts for the week. The computer class then applied these concepts through self-paced learning using the R statistical software.

The interactive classes made use of activities suggested by Gelman and Nolan (2017) as well as those conceived by the author. Examples of activities which worked well included:

- Using students' own data for in-class examples. In an early session students shared data such as their height, mobile phone usage and number of siblings. The collection and use of this data in successive examples demonstrated some principles regarding data collection, distributions, outliers and spurious correlations with no causal relationship.
- Sampling chocolates from a box. Using a box of chocolates that contained several variety of chocolate, students were able to test the manufacturer's claim for the total weight of the box or

whether one variety of chocolate weighed more than another. This can illustrate one- and two-sample hypothesis tests and, later, analysis of variance with multiple categories.

- Sampling random points on the earth's surface. Students were able to estimate the proportion of the earth's surface covered by water by successively sampling points and noting their estimate converged, and the confidence interval narrowed, as the sample size increased. (As an extension, it also illustrates that sampling uniformly on a sphere is non-trivial.)
- Measuring the time for paper helicopters to fall. Multiple teaching points were illustrated through this example. Firstly, the scenario motivates questions around different sources of sampling variability (e.g. natural variation in air movement, measurement error from imprecise recordings) and how these can be controlled or minimized by careful experiment design. Secondly, when students can vary the design of the paper helicopter, it becomes an investigative cycle with students fitting models of the time to fall which can then be used to make predictions about other potential designs that can then be tested.

Although the demonstrations above are mostly artificial, they were useful for illustrating statistical *concepts* and encouraged *active learning* in a classroom setting. The other GAISE principles of using *real data* and *technology* were implemented using the weekly 2-hour computer classes. In each computer class, students were given an R Markdown file which contained a summary of the main learning outcomes from the 1-hour class, plus a series of exercises for them to complete using R. The R Markdown file enables students to integrate text, R code and their computer output (e.g. graphs, analysis) into a single file. It was pre-populated with instructions and R code which read in data and performed simple examples. Students followed the instructions at their own pace but were invited to work with friends and ask the instructor questions. The advantages of this approach are:

- Students are active learners.
- Students build up knowledge of R functions each week.
- Students could add annotate their file with their own explanations, interpretations and notes to create their own 'textbook' during the year.
- Students could analyze large datasets or simulate data with particular characteristics.
- Students could analyze real datasets with interesting contexts and conclusions.

Some of the computer exercises were motivated by Downey (2014) but adapted for R. In particular, the US National Survey of Family Growth and the Behavioral Risk Factor Surveillance System provided large real datasets. Using real datasets exposed students to practical issues such as missing data. Some of the original datasets were so large that each student was given a different subset to work with. This had the added benefits of reducing collusion between students and illustrating how results are dependent on sampling variability.

Assessment

Each student's module mark was determined through four different assessments:

- Portfolio (20%). Students submitted a portfolio at the end of each half-semester consisting of their completed R Markdown files from the computer classes (see above). This format enabled students to gain credit for actively engaging with the curriculum and combined the processes of learning and assessment.
- Presentation (20%). Each week one student would give a 20-minute oral presentation on a more advanced statistical topic not covered in the main syllabus. Examples of such topics included survey sampling, non-parametric testing and cluster analysis. These were topics which were deemed not essential for inclusion in the main syllabus, but knowledge of the topic's existence might be useful for students in their future scientific work. It also developed statistical communication skills as each student summarized the topic for their peers.
- Project (20%). For the final two weeks of the module, students were invited to write a report analyzing data on a topic of their choice. This was designed to further develop their statistical investigation skills and be a more authentic form of assessment similar to reports required of them as scientists. Students were credited for formulating research questions, choosing appropriate statistical methods or models, and applying statistical techniques to answer their questions.

Examples of datasets chosen by students included archaeological glass composition, breast cancer markers, climate change and traffic fatalities.

- Exam (40%). University regulations required the module to have an exam. It consisted of 16 short-answer questions designed to test knowledge and understanding of key statistical principles, methods and concepts, as well as interpretation of R output. There were only a few questions where students were required to perform calculations.

EVALUATION

The design and implementation of the module was evaluated via a student survey distributed in the final class of the module. Eight out of the ten students who participated in the pilot responded.

Teaching and ethos

Students were positive of the pedagogic approach noting it was "a more interactive and engaging way of learning that allows practice and help". The computer classes were praised as they "helped me to consolidate the lectures and made the concepts less abstract", "allowed me to test my understanding and ask questions early in the term" and "made learning the material so much easier and having examples to use is very useful". Students sometimes struggled with R coding, and it was noted that "maybe [the teacher] could have gone through what each part of the code in the examples has done" and "I have had to learn [R] by myself".

The overall approach of the module received positive feedback. One student reported "The workload for this module has been the most consistent and manageable of any other, though still challenging" while another said, "I've really enjoyed this module, and [it] has improved me to think about a career in something stats based".

Assessment

The survey was conducted before the project and exam had been completed so some feedback focused on the presentations. Students had mixed opinions on the presentations with some noting the "[Presentations were] really interesting. Nice to see others present" and "Research into topics improved techniques of presenting and working out how to do R code on your own". Whereas others thought more guidance was necessary, "Maybe [give us] references to textbooks or a checklist of things we should look out for in each topic" and "Guidelines on good presentation practices would have been useful".

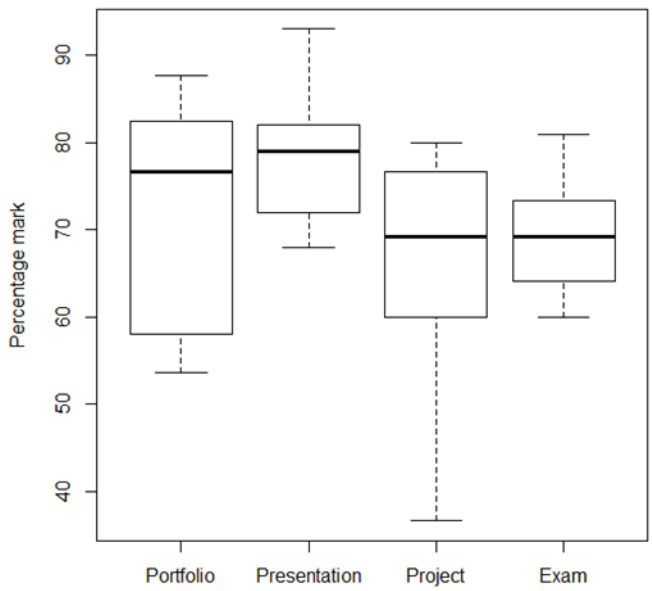

Figure 1. Distribution of student marks.

The variety of assessment generally drew praise with comments including, "I really like how it is assessed in different ways, as it allows us to build on all kinds of skills", "You can build up your grade throughout the year so therefore the exam doesn't decide everything" and "I much prefer regular assessment as it keeps me up to date… and I can see where I'm going and how to improve". One student commented that the portfolio and project helped towards exam revision whereas another noted they would prefer more examples of exam questions.

Figure 1 shows the distribution of student marks across the different assessments. The portfolio, presentation and project were all positively correlated ($r = 0.61$ to $r = 0.76$), with the exam only weakly correlated with the project ($r = 0.25$), presentation ($r = 0.29$) and portfolio ($r = 0.48$).

CONCLUSION

This study set out to apply the principles and goals of introductory statistical education, as set out in the GAISE report (ASA Revision Committee, 2016), to the design of a new statistics module for interdisciplinary STEM students. Overall, students welcomed the approach compared to most of their other modules which used a traditional lecture and exam format. However, as a pilot, the class size was small so caution must be exercised with extrapolating conclusions too far. For future iterations of the module, there are some issues which should be considered:

- Coding skills. In some computer class exercises, students placed too much focus on R coding which detracted from their focus on understanding statistical concepts. It is necessary to balance developing statistical skills versus coding skills, and exercises could be adapted to achieve optimal balance. R Shiny apps, for example, could teach statistical concepts without requiring R coding.
- Assessment load. The advantage of using multiple assessments is it presents the opportunity to develop multiple skills. Although the students in the pilot responded to this positively, there is the danger of over-assessing. The marks from the different assessments were correlated, which suggests some are redundant for forming the module mark and could be made formative.
- Presentations and projects. Students requested more support and guidance with these open-ended tasks. The quality of the presentations were generally very good, with students demonstrating good research skills. However, there was little opportunity for students to feed forward their learning or for the audience to engage. Uniting the theme of each student's presentation and project could be beneficial.
- Depth and breadth. The Natural Sciences course has one of the highest entry requirements in the university so, although not mathematical specialists, students taking the module were very strong academically. If a similar approach was to be taken for more mixed-ability students, then breadth and/or depth may need to be compromised.
- Class size. The pilot involved a small number of students, providing good opportunities for teacher-student instruction and peer-to-peer learning. For a larger class size some adaptations would be necessary such as group presentations instead of individual presentations, or automated assessment of computing portfolios.

This case study demonstrates the potential to teach statistics to scientists through context-relevant scenarios, ultimately enabling students to apply their knowledge to projects involving their own scientific data. Through an 'apprentices of skills' rather than 'learners of knowledge' approach, the case study illustrates one way to implement the GAISE recommendations to positive effect, with further work needed to demonstrate the findings are widely applicable.

REFERENCES

ASA Revision Committee (2016). *Guidelines for Assessment and Instruction in Statistics Education College Report 2016.* American Statistical Association.

Downey, A. B. (2014). *Think stats: Exploratory data analysis in Python* (2nd edition). Green Tea Press.

Garfield, J. & Chance, B. (2000). Assessment in Statistics Education: Issues and Challenges. *Mathematical thinking and learning, 2*(1–2), 99–125. https://doi.org/10.1207/S15327833MTL0202_5

Gelman, A. & Nolan, D. (2017). *Teaching statistics: A bag of tricks* (2nd edition). Oxford University Press.