

Forum article

Emma Rand

Contents

1 A blended-learning design for data skills modules in Biology	1
1.1 Overcoming the challenges of teaching data analysis and programming to early career and sometimes wary biologists. Emma Rand	1
1.2 Supporting Material	6

1 A blended-learning design for data skills modules in Biology

1.1 Overcoming the challenges of teaching data analysis and programming to early career and sometimes wary biologists. Emma Rand

1.1.1 Introduction

The analysis and presentation of data are core learning outcomes of Biology-led undergraduate programmes and a requirement for conducting biological research. However, the combination of a stage one biology student's perception of their subject and their previous experience can lead them to be wary of data analysis and coding. Furthermore, the cognitive processes required to learn these skills are towards the apex Bloom's taxonomy and this is in contrast to those normally required to perform well in pre-university and stage 1 assessments. This article discusses some challenges in teaching data skills to students for whom it is taught as a skill to do the research in their chosen field, and describes a blended learning approach to meet some of those challenges.

1.1.2 Data Skills as a core part of the Biologist's training

Underlying all biological discoveries are data! The ability to generate reliable measures of biological phenomena then analyse them and communicate the results, is essential for a biologist. This has long been the case (see, for example: Royal Statistical Society (1947); Finney (1968)), but more recently an explosion of large-scale and complex biological data has made the acquisition of data skills even more crucial (National Research Council and others 2003). This is reflected in the Programme Learning Outcomes for Biology-led programmes. See table 1.

Opportunities to develop and apply these skills are embedded across entire programmes but we also include modules with a specific focus on data analytics. One these is a stage one module called "Laboratory and Professional Skills for the Bioscientist" in which a term, carrying 10 credits, is devoted to Data Analysis in R. There are 25 hours of contact time (nine one-hour lectures and eight two-hour workshops) and 75 hours of independent study and assessment. The module is taken by 250-280 stage 1 students.

1.1.3 How do the challenges arise?

A stage 1 biology student may expect to spend the most of their time in the laboratory or the field as portrayed on Admissions pages (See figure 1).

In addition, they can be surprised by the importance placed on data handling and analysis in a biology degree. These subjects are recognised as among the more difficult to teach (e.g., Boulay (1986)), require

Table 1: BSc Biology Programme Learning Outcomes.

Provide thorough explanations that demonstrate a deep understanding of the principles, concepts and theories on the origin, evolution, structure, function, development, and distribution of living organisms, through critical evaluations of the primary scientific literature in Biology
Formulate hypotheses, **design and execute experiments for the collection, analysis and modelling of biological data, that tests biological systems to produce figures, graphs and tables** that are explained in comprehensive laboratory report
Thoroughly evaluate experimental, analytical and quantitative techniques and methodologies, and first-hand practical experience and training in laboratories or the field, to demonstrate an awareness and appreciation of the application of these approaches in tackling the major global challenges in Biology of the 21st century
Work effectively as an individual, in teams and laboratory groups to solve biological problems by applying logical reasoning and lateral thinking to develop safe, ethical and socially responsible solutions that may benefit humankind
Communicate and interpret complex information with clarity and precision through critical reviews in written, oral and other explanations, questioning dogma and demonstrating impact at the forefront of research in Biology to real-world and global issues for expert, professional, business, industrial and lay audiences
Demonstrate independence, originality, and a deep understanding of cutting-edge practice and technology in Biology, **apply numerical, quantitative, and computer-based transferable skills** to a range of working environments including laboratories, fieldwork, education, industry, business, health services, policy, government, and media

more and harder work from the student and may be relatively unpopular (e.g., Uttl and Smibert (2017)). As consequence at the start of Data Analysis in R, students often do not expect to it with 70% disagreeing with the statement “I will enjoy data analysis.” (See Figure 2).

The education they have primarily been exposed to is at the foundational levels of the revised Bloom’s taxonomy of learning (Anderson and Bloom (2001)) where as the acquisition data analytic skills occurs at the levels of “applying”, “analysing” and “evaluating” (left side of Figure 3). Memorising does little to help a student analyse and represent their data. Every situation is new and they must analyse it, recognise how to apply complex skills and evaluate the results.

Associated with increasing levels of Bloom’s taxonomy is an increase in the number of mistakes made (right side of Figure 3). Research is characterised repeated rounds of ‘trial and error’ with successes being infrequent. This is in contrast one-time correction of one’s factual knowledge and tends to be more uncomfortable. One way to characterise Bloom’s taxonomy is that it is process of becoming more comfortable with failure as a learning opportunity.

1.1.4 What’s required to meet the challenges

Learning to programme requires a lot of practice. Novice programmers see many aspects of programming syntax such as the use of brackets, quotes, commas and spaces as arbitrary and need substantial practice and exposure to start to see code as experienced practitioners do (Sorva 2018). This activity is well-suited to directed independent learning but it can be difficult to motivate students to do sufficient amounts (A’Brook and Weyers 1996; Carey and Papin 2018). Using contact time to partially achieve this can increase engagement but does not maximise the value of that contact time which is better spent engaging with students to develop their conceptual understanding and strengthen the staff-student partnership.



Figure 1: Image taken from Biology Undergraduate Admissions pages

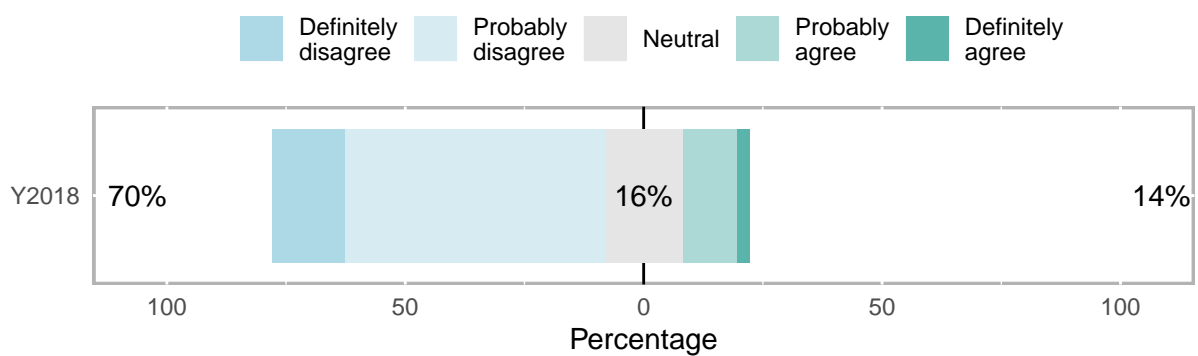


Figure 2: Expectations of Stage 1 Biologists about of Data Analysis in 2018: Responses to the question '**I will enjoy data analysis.**' given as the first slide of contact time slot.

Previous Learning Experience

Bloom's Taxonomy

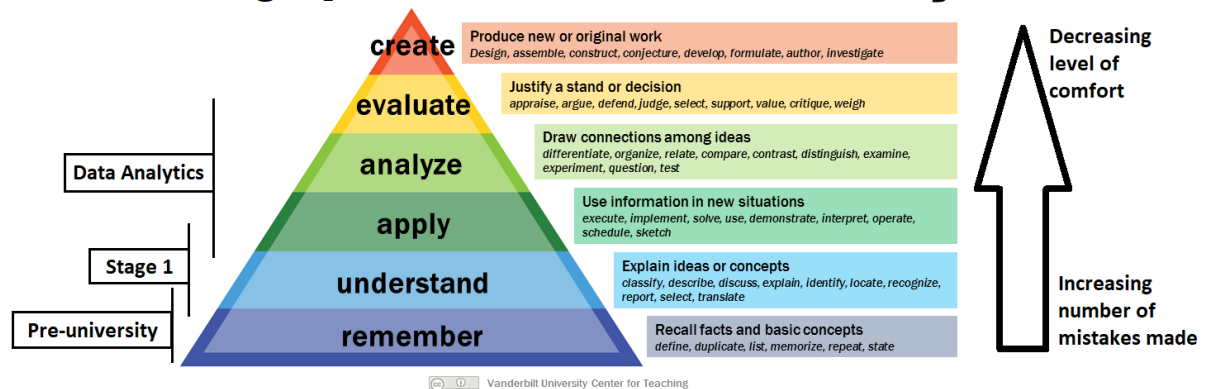


Figure 3: An annotated Revised Bloom's taxonomy adapted from Vanderbilt University Center for Teaching. Data analysis typically requires learning over a wider range of levels with a greater number of mistakes than a stage 1 student maybe used to.

Structure of a week

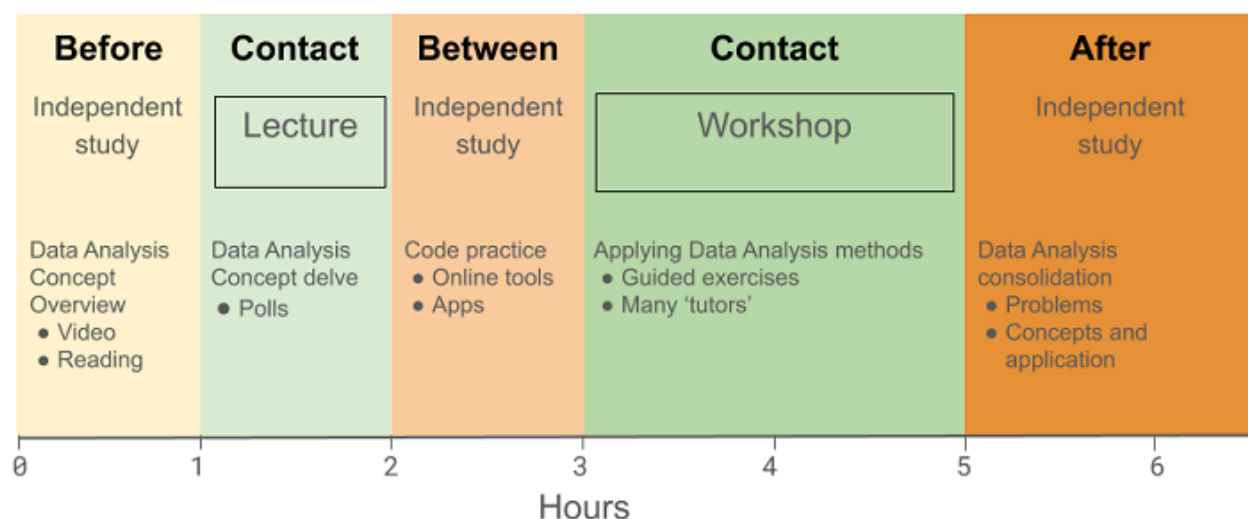


Figure 4: The structure of a 'Data Analysis in R' week. There are two periods of contact: a lecture covering statistical concepts and a workshop putting those concepts in to practice in R. The contact periods are interleaved with guided independent study which focusses on the conceptual overview, code practice and the independent application of both to complex problems.

1.1.5 Blended learning solutions

Flipped-learning is employed during Data Analysis in R to help students meet the challenge of acquiring data skills. The module has a repeated weekly structure comprising periods of guided independent study interleaved with active and problem based learning placed within contact time. These are followed up with formative work. See Figure 4.

Independent study issued before the first contact point gives an overview of the statistical concepts covered during that week. The exercises are designed to take an hour or less and vary in format. They include videos, accessible articles and material from online data analysis books. The lecture provides a deeper coverage of the topic concepts and an introduction to code that will be used in the workshop. An audience response system is used to facilitate active learning which has been shown to increase student performance in STEM subjects (Freeman et al. 2014). Between contact points, the independent study is focused on code 'mechanics' using an online tool, which gamifies code practice (<https://www.datacamp.com/home>) and allows students to practice the R language syntax without having to simultaneously consider statistical concepts. The second period of weekly contact consists of a two-hour workshop of structured data analysis problems which require students to apply their understanding of statistical concepts using R to generate analyses of biological data in forms suitable for a report. The problems are broken down into the steps required to take the problem scenario and data through to conclusion and reporting. This activity is well supported with graduate demonstrators. The independent study exercises following the workshop give scenario with data for which students design an appropriate analysis workflow.

In 2019 I introduced a discrete unit of independent study designed to be carried out before the start of the

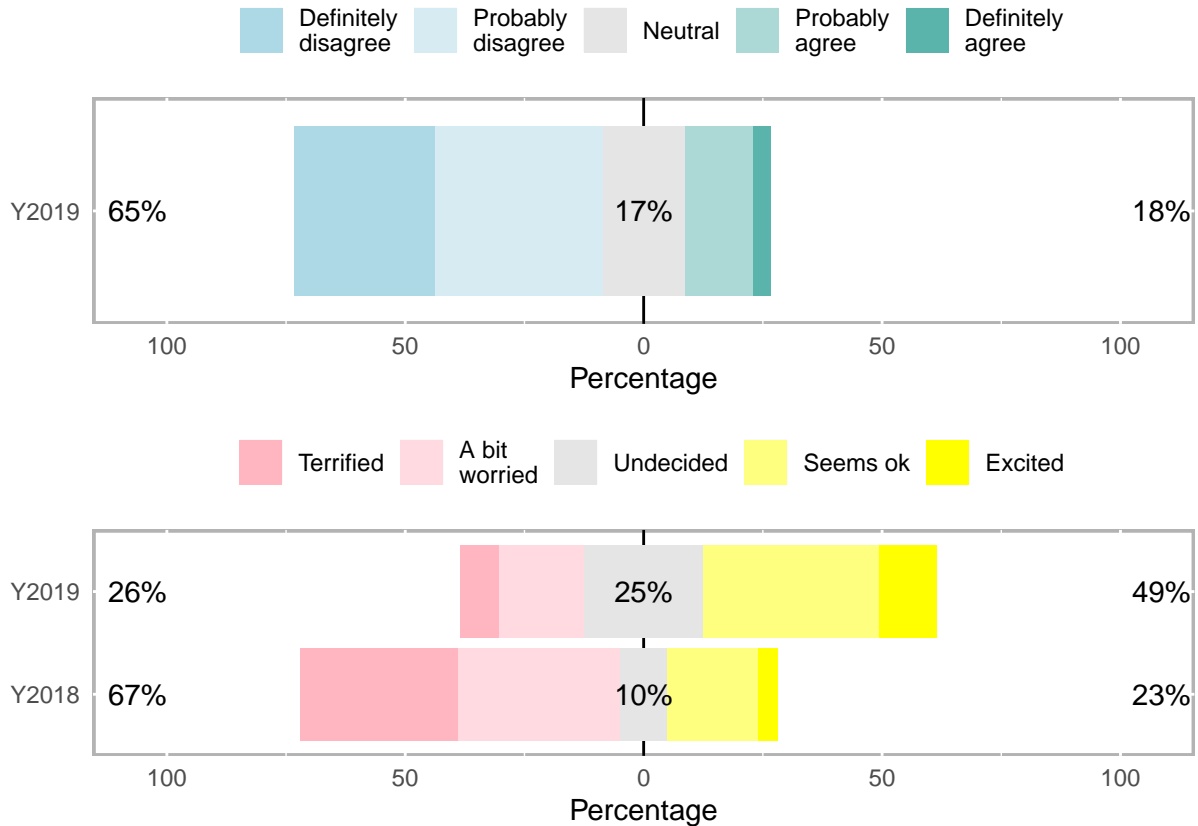


Figure 5: Expectations of Stage 1 Biologists about of Data Analysis in 2019. Responses to the question ‘**I will enjoy data analysis**’ follow the same pattern as in 2018 (upper panel). However, when asked ‘**How do you feel about R?**’ anxiety was significantly reduced following the introduction of an independent study activity prior to any contact time. (lower panel?)

module. This was specifically aimed at training students in R language core concepts. The aim was to make less steep the learning curve involved in acquiring data analytic skills which requires students to assimilate both statistical concepts and coding concepts. This one, relatively short piece of work significantly reduced the level of fear expressed about learning R. See Figure 5. Whilst students’ expectations of data analysis were unchanged from 2018, the percentage of students terrified of, or worried about, R dropped from 67% to 26% and that of students unconcerned or excited about R rose from 23% to 49%. Since fear of a subject can influence an individual’s capacity to initiate and complete tasks (Onwuegbuzie 2004) we might predict an increase in learning achieved. This will be the subject of future investigation.

In summary, module teaching comprises online, independent and face-to-face learning activities in which the independent study is guided and fully integrated by interleaving it with contact time.

1.2 Supporting Material

1.2.1 Biography

Emma Rand is a Teaching & Scholarship lecturer in the department of Biology, the University of York. She specialises in experimental design and reproducible data science from pre-processing and exploration to statistical analysis and visualisation. She was an invited tutor at useR!2019, the International R Conference,



Figure 6: photograph of emma rand

and a trainer for the Royal Society of Biology and The Biochemical Society. She is twice the winner of YUSU's Teacher of the Year Excellence Award

Email: emma.rand@york.ac.uk

A'Brook, Richard, and Jonathan DB Weyers. 1996. "Teaching of Statistics to Uk Undergraduate Biology Students in 1995." *Journal of Biological Education* 30 (4): 281–88.

Anderson, Lorin W, and 1913- Bloom Benjamin Samuel. 2001. *A Taxonomy for Learning, Teaching, and Assessing : A Revision of Bloom's Taxonomy of Educational Objectives*. Book; Book/Illustrated. Abridged ed. New York : Longman.

Boulay, Benedict Du. 1986. "Some Difficulties of Learning to Program." *Journal of Educational Computing Research* 2 (1): 57–73. <https://doi.org/10.2190/3LFX-9RRF-67T8-UVK9>.

Carey, Maureen A., and Jason A. Papin. 2018. "Ten Simple Rules for Biologists Learning to Program." *PLOS Computational Biology* 14 (1): 1–11. <https://doi.org/10.1371/journal.pcbi.1005871>.

Finney, D. J. 1968. "Teaching Biometry in the University." *Biometrics* 24 (1): 1–12. <http://www.jstor.org/stable/2528456>.

Freeman, Scott, Sarah L. Eddy, Miles McDonough, Michelle K. Smith, Nnadozie Okoroafor, Hannah Jordt, and Mary Pat Wenderoth. 2014. "Active Learning Increases Student Performance in Science, Engineering, and Mathematics." *Proceedings of the National Academy of Sciences* 111 (23): 8410–5. <https://doi.org/10.1073/pnas.1319030111>.

- National Research Council and others. 2003. “NRC. 2003. Bio 2010: Transforming Undergraduate Education for Future Research Biologists.” National Academies Press, Washington, DC, USA.
- Onwuegbuzie, Anthony J. 2004. “Academic Procrastination and Statistics Anxiety.” *Assessment & Evaluation in Higher Education* 29 (1): 3–19. <https://doi.org/10.1080/0260293042000160384>.
- Royal Statistical Society. 1947. “Report on the Teaching of Statistics in Universities and University Colleges.” *Journal of the Royal Statistical Society* 110 (1): 51–57. <http://www.jstor.org/stable/2981262>.
- Sorva, Juha. 2018. “Misconceptions and the Beginner Programmer.” *Computer Science Education: Perspectives on Teaching and Learning in School*, 171.
- Uttl, Bob, and Dylan Smibert. 2017. “Student Evaluations of Teaching: Teaching Quantitative Courses Can Be Hazardous to One’s Career.” *PeerJ* 5 (May): e3299.