Open Research Online

The Open University's repository of research publications
and other research outputs

When one size does fit all: Simultaneous delivery of statistics teaching to multiple audiences

Journal Item

oro.open.ac.uk

When one size does fit all: Simultaneous delivery of statistics teaching to multiple audiences

Rachel Hilliam
The Open University, England, rachel.hilliam@open.ac.uk

Karen Vines
The Open University, England, karen.vines@open.ac.uk

Recommended Citation

When one size does fit all: Simultaneous delivery of statistics teaching to multiple audiences

Abstract

In 2019, The Open University celebrated its 50th anniversary. For most of its fifty years it has been delivering statistics education to both specialists and non-specialists, simultaneously and at scale. A single module must deliver statistics teaching across multiple qualifications in many different disciplines which means it is not possible to tailor module material and examples for each individual qualification. This is contrary to the way many service teaching modules are designed. For example, the introductory statistics module at The Open University delivers basic statistics and probability to around 1500 students a year. By structuring the material around topics to which everyone can relate; money, education and health, the module is able to engage students across a range of different qualifications from mathematics to economics, business studies, computing and natural sciences. Feedback from students has been overwhelmingly positive. Students are provided with printed material, which is supplemented with a variety of interactive elements, which will be described in this paper. Online forums provide a place for peer support. Small group tutorials, both face to face and online, are provided by tutors. Each tutor will typically support 20 students and provide feedback to those students on their assessable work. The students in an individual tutor's group could be studying one of many different qualifications. All the module material, including all the interactive elements, are embedded in the university's virtual learning environment which has made it possible to monitor how students engage with the material. This knowledge, and qualitative feedback from students, has provided support for the idea that by teaching through a narrative approach, with examples based on scenarios which have a broad interest to everyone, it is possible to engage all students with statistics regardless of their qualification. By taking this same narrative approach a new module is being created which will be studied concurrently by students in the final stages of their qualifications in mathematics, statistics, data science and economics. This paper will outline how the module will utilize Jupyter notebooks to deliver practical modelling with real data sets chosen to be of interest to all the different cohorts of students. During the last 50 years The Open University has always had to deliver modules at scale to students across many disciplines. With the increased use of data across qualifications, and the need for people to upskill in data analysis, it is essential to explore ways in which statistical training can be delivered at scale.

Keywords
Statistics, Simultaneous delivery, multiple disciplines

Introduction

Data is everywhere, driving business decisions and health, medical and social advancements. Employers across all organisations use data to make decisions. Therefore, the need for people with the required skills to collect, process, analyse and interpret data is growing at a rapid rate. This puts increasing pressure on universities to incorporate data analysis skills into qualifications in many disciplines, for example Economics, Psychology, Criminology, Computing, Social Science, Education, Business and Engineering. Whilst not all graduates from these different disciplines will eventually use statistics to analyse data, many students will need statistical literacy skills to interpret and critically evaluate statistical information. In a review of statistical provision, Tishkovskaya & Lancaster (2012), emphasise the increased importance of a statistically literate society and outline many application-based approaches where students are taught to think critically about the social and political situations in which data are used. Statistics teaching must therefore ensure students are equipped with the necessary skills to meet their individual aspirations and future challenges as graduates.

The world of statistical education is an evolving landscape. The increased use of data, both structured and unstructured, has led to many employers recruiting people to the relatively new role of 'data scientist' (Davenport & Patil, 2012). Whilst there is no agreed definition of a data scientist, the skills needed to process, analyse and interpret statistical findings are common in all definitions (Fayyad & Hamutcu, 2020). Data science is usually regarded as a blend of computer science, statistics and applied mathematics (Blei & Smyth, 2017). The increased demand for data scientists across a wide range of employment areas provides both opportunities and challenges for statistics educators. As recently as May 2019 a report by the Royal Society (2019) stated:

> We believe that data science can be an exciting and fulfilling career, that also addresses society's needs. That requires the right higher education and training to be made available. More broadly, users, analysts and citizens of the future will need to be comfortable with the application of data science to societally pressing questions.

There is also a need for society in general to have a greater understanding of data and the ability to use it to make decisions. One recent framework, produced by the International Data Science in Schools Project (IDSSP) (2019), aims to address these challenges by transforming the way teaching and learning about data science is carried out in the last two years of schooling. In the IDDSP (2019) framework for Introductory Data Science two objectives are stated:

> To ensure that school students acquire a sufficient understanding and appreciation of how data can be acquired and used to make decisions so that they can make informed judgments in their daily lives, as students and then as adults. In particular, we envisage future generations of lawyers, journalists, historians, and many others, leaving school with a basic understanding of how to work with data to make decisions in the presence of uncertainty, and how to interpret quantitative information presented to them in the course of their professional and personal activities.

> To instil in more scientifically able school students sufficient interest and enthusiasm for Data Science that they will seek to pursue tertiary studies in Data Science with a view of making a career in the area."

So, at the same time as ensuring students studying statistics in their qualifications have the necessary statistical literacy skills needed for their chosen professions, we also need to ensure that an equally

large group of students has the skills required to collect, analyse, interpret and communicate statistical information. The challenge, then, for University educators is how to deliver statistical training to large numbers of students.

Like all universities, The Open University (OU) has an employability framework which seeks to ensure that skills in problem-solving, communication, collaboration, numeracy, digital and information literacy, taking initiative, self-management and resilience, self-awareness, commercial and/or sector awareness and global citizenship are embedded into the curriculum. Dealing with, and analysing, data underpins many of these skills. Indeed, applying statistical techniques to real-world situations will be important to students in wide range of disciplines (Garfield, 1995).

Of course, statistics has been taught in large numbers of university qualifications for many years. How the statistics teaching is delivered has taken, and continues to take, many forms ranging from the a few afternoons of statistical training, defined modules in statistics, through to embedding statistics within the qualification being studied. However, these courses, which we shall term 'statistics services courses', pose challenges as students are often less motivated to learn statistical techniques and the amount of statistics covered in these courses can often be vast and daunting (Jaki, 2009). There is a danger that students emerge from these courses having been taught large numbers of statistical tests but are unable to decide which test to use when analysing their own data. Further problems arise as students often perceive statistics courses to be mathematical (Gordon, 2004). Students who are studying non-mathematical qualifications do not necessarily have a strong mathematical background. Careful thought needs to be given as to how to construct courses which will build up confidence in such students (Georgakis et al., 2011) as studying a statistics course which they consider to be mathematical can lead to them developing mathematics anxieties; that is negative emotional reactions when faced with a mathematical problem (Onwuegbuzie & Wilson, 2003). This means that a major challenge for teachers of statistics service courses is to ensure there are connections between the statistics curriculum and the curriculum within the student's own qualification. One way of achieving this is to use examples from the student's own qualification (Gordon & Nicholas, 2010). However, as Garfield (1995) points out, students will bring their own ideas and background into the classroom environment and only understand material after they have interpreted the new information in terms of their own acquired knowledge. Therefore, an alternative approach to using examples from a student's qualification, is to use examples that all students can relate to from their everyday life. This then gives students the tools to be able to apply the same techniques within their own qualification.

The OU has delivered statistics modules at scale for many years. These modules are written to serve multiple audiences; both in terms of different disciplines but also due to the students' wide range of prior educational knowledge. Given the diverse nature of the students it is not possible to deliver multiple statistics modules tailored to the student's chosen qualification. Unlike modules delivered by many Higher Education Institutions, modules at the OU must concurrently serve both mathematics and statistics students and act as service courses for other qualifications, with the same material.

The scale of statistics teaching at The Open University

The OU is one of the largest universities in Europe, providing distance learning education to over 150,000 students. It is also the largest provider of higher education to students in the UK who have a declared disability with 27,237 such students in 2018/19 (The Open University [OU], 2020). Having spent 50 years developing and delivering distance learning modules, the OU has a world-

wide reputation built on its distance learning methodology. The delivery of learning combines high quality learning material (both printed and online) and correspondence tuition together with online and face-to-face support provided by a network of about 6,000 tutors (officially designated Associate Lecturers). Since its inception in 1969, the OU has continually updated its practices to incorporate developments in both pedagogic theory and technology.

The OU uses three levels (numbered 1-3) for modules forming undergraduate degrees. Each level effectively matches the corresponding year of full-time study at an English university and hence corresponds to levels 4, 5 and 6 of the Framework for Higher Education Qualifications in England, Wales and Northern Ireland (QAA, 2014). At the OU every module earns the student a specified number of credits measured in CATS (Credit Accumulation and Transfer Scheme points) which are used by many universities in the UK. One credit or CAT point represents 10 learning hours. The statistics modules are all 30 CATS points, corresponding to the equivalent of a quarter year of full-time study. Modules can also be studied separately instead of as part of a qualification. Statistics modules are often taken by students as part of their continual professional development.

The size and scale of the OU presents challenges when designing modules as many modules on offer are simultaneously studied by students on multiple different qualifications. For example, on the most recent presentation of a second level module, *Analysing data*, there were roughly 300 students registered but only about half of these students were studying for a qualification in mathematics, statistics or data science. The other half were studying the module as part of another qualification or studying the module in isolation. This poses challenges as to how to make the module relevant to all these different audiences.

At level 1 the statistics module, *Introducing statistics*, is studied by students on an even greater number of qualifications. The module is large with roughly 700 students on each presentation, with two presentation running each year. As with the level two module *Analysing data*, only half of the cohort are studying *Introducing statistics* as part of a mathematics, statistics or data science qualification. The demographics of the students are equally diverse; for example, in age terms, a small group of students take this module whilst still at school and study alongside students over the age of 65. Approximately a quarter of the students are aged between 30 and 39. Generally, around 50% of the students are in full-time work, 20% part-time work, 20% not in paid work or unemployed, 2% retired; the remaining 8% have not declared their employment status to the university. Many of the students have not studied mathematics or statistics for many years. Over 20% of the students start the module with their highest previous qualification being less than A-level or equivalent. Some students start the module having already attained mathematical success at HE level. This means the module must cater for students with little mathematical background as well as simultaneously keeping the interest of more mathematically experienced students.

As already outlined, one method of generating interest in statistics from students outside of mathematics and statistics is to ensure that the examples are related to the student's qualification. However, for a module which needs to concurrently meet the needs of students from a multitude of disciplines and backgrounds such an approach is not possible. Instead the focus is on finding examples that relate to students' lives and interests outside of study.

The Introducing statistics module material

Alongside the students coming to *Introducing statistics* who have not studied statistics before, many others have had unsuccessful experiences with mathematics and/or statistics whilst at school. The module was therefore designed to increase the confidence of students whilst providing a rich experience of statistics that students can apply to many qualifications. To service the multiple qualifications and prior experiences of students the module highlights the importance of statistics and statistical thinking in everyday life rather than trying to use examples from multiple qualifications. Examples are drawn from three areas which were thought to be of importance to the general public: money, education and health. These areas were chosen because, at the time of writing the module, they touched on issues of national debate such as the gender pay gap, 'What is a good school?' and 'Is a new drug safe and effective?' Furthermore, it was felt that these areas would remain relevant to students over the lifetime of the module as they had already stood the test of time having been first used on a predecessor module developed in the 1980s.

In common with all OU modules, a week by week calendar is provided on the virtual learning environment (VLE) which guides the students through the material by indicating what should be studied each week. The core teaching material is provided in text format and can be accessed as pdfs on the VLE or in a printed format which is posted out to the student at the start of the module. The printed material is divided into 11 units of new material, together with a final 12th unit which brings all the material together. The title of each unit reflects a topic based in one of the three areas of money, education and health (Table 1). Each unit uses real data and examples to form a narrative about a topic in one of these three areas. Emphasis is placed on the statistical method being used and the way in which conclusions from the analysis are communicated rather than theoretical derivations. For example, Unit 2 focusses on the price of goods in order to teach the concept of statistical summaries. Examples of data used in Unit 2 range from a small data set which gives the price, in 15 different shops in Milton Keynes, of a 100g jar of instant coffee, to large data sets from the Office of National Statistics on the Consumer Prices Index (CPI) and the Retail Prices Index (RPI).

The core text for each unit contains a teaching narrative interspersed with examples, activities and exercises. Both the activities and exercises provide a measure of active learning. Activities are designed to help with initial learning and exercises act as a check on understanding and as a revision tool. In both cases full solutions are provided so that students can instantly access feedback, albeit of a generic nature. Within the core text the students are also directed to other resources such as use of the software and screencasts. Much of the material in Unit 2 is available for anyone to freely study and can be viewed at https://www.open.edu/openlearn/science-maths-technology/prices-location-and-spread/content-section-0?intro=1. The work involving the module software (Minitab) is chunked so that students are not tied to desktop or laptop machines all the time whilst they are studying. The students are also given some flexibility over when they complete activities involving the module software to enable them to make best use of when and where they can study on a week-to-week basis.

Table 1

Outline of topics and unit titles for Introducing statistics.

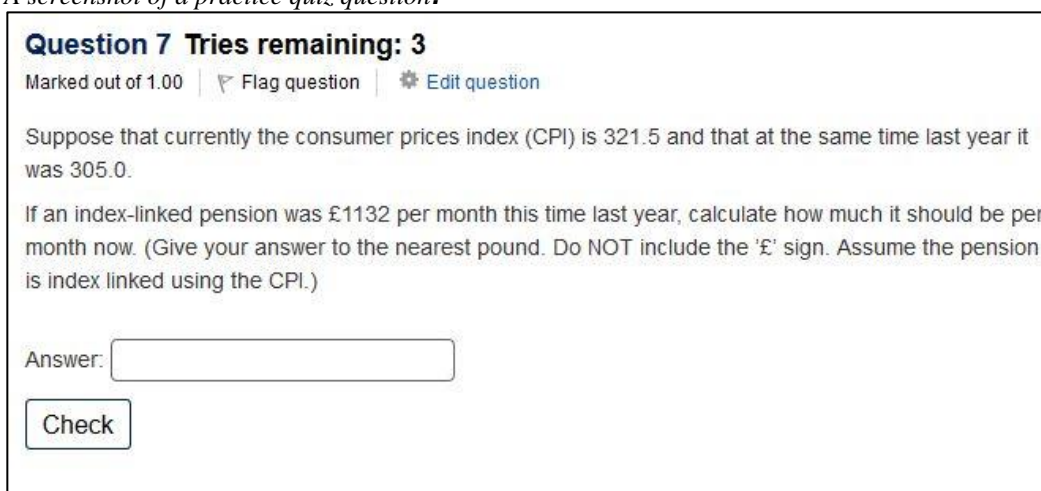| Example area	Unit title	Statistics covered
Money	Unit 1 Looking for patterns	The basic idea of statistics modelling and the modelling diagram
		Stem plots
		The shape (skewness, modes) of data sets
		Median and range
	Unit 2 Prices	Mean, weighted mean, quartiles, interquartile range
		Five-figure summary
		Simple ideas of index numbers
		UK consumer price indices (CPI, RPI)
	Unit 3 Earnings	Earnings ratio
		Percentile and deciles
		Boxplots
		Deviations, variance and standard deviation
		Average Weekly Earnings index and comparing changes in prices and earnings
	Unit 4 Surveys	Basic ideas of survey sampling
		Simple random sampling, systematic sampling, general ideas of stratification and clustering, quota sampling
		Sampling errors
	Unit 5 Relationships	Relationships, scatterplots, response and explanatory variables
		Describing relationships
		Lines and residuals
		Least squares regression
Education	Unit 6 Truancy	Basic ideas of probability
		Steps in a hypothesis test
		The sign test
		p-values and interpreting significance test results
	Unit 7 Factors affecting reading	The normal distribution
		One and two sample z-tests
	Unit 8 Teaching how to read	Contingency tables
		The chi-squared test
		Type 1 and type 2 errors
	Unit 9 Comparing schools	Causality and association
		Correlation
		Confidence intervals and prediction intervals
Health	Unit 10 Experiments	Basic ideas of scientific experimentation
		One and two sample t-test
		Matched pairs t-test
		Calculating confidence intervals
	Unit 11 Testing new drugs	Drug testing and clinical trials
		Type of design for trials (group comparative, matched pair, crossover)
		Phases of drug trials, post-marketing surveillance

During the last 50 years The Open University has transformed the way learning material is presented as technology has advanced. Daisy (Digital Accessible Information System) 'Talking Book' versions of the material are available to all students. These allow learners with visual impairments and dyslexia the option of hearing the material in a natural spoken way and include navigation options together with the ability to bookmark sections. Of course, by its nature, statistics is a highly visual subject involving not just mathematical notation but also many diagrams and plots. Each of the figures in the units are described in text format. The main plots are also available as tactile diagrams and as sonifications, providing visually impaired students the ability to interrogate and understand these plots (Vines et al., 2019). In addition to the core text material there are several interactive activities which are designed to keep students engaged in a variety of ways.

Practice quizzes

It is well established that when students are encouraged to engage with corrective activities, they are helped to overcome their own misconceptions. Feedback needs to be timely so that students can engage with it (Mevarech, 1983). By using interactive quizzes students can test their own understanding and are more likely to engage with the material (Dietz-Uhler et al., 2007). The study calendar is therefore punctuated with computer-marked practice quizzes, designed to test students' understanding on each unit and provide graded instantaneous feedback in order to assist with learning (Figure 1). The results of these quizzes do not count towards the students' overall assessment grade but are designed to enable students to self-diagnose areas in the module where they feel they may need more help and to clarify key concepts in the material. To meet these aims, the practice quizzes allow three attempts at each question, the level of support increasing with each attempt. After a first incorrect attempt students are given a hint of which section of the module material the question is testing. With a second incorrect answer the students receive a hint on how their answer was incorrect and a third incorrect answer results in a worked solution provided for the student. Students can repeat each practice quiz as and when they feel appropriate. As each question is chosen at random from a question bank of similar questions; this provides multiple opportunities for the student to practice a topic. Diagnostics on students attempts can be interrogated by members of the teaching team so that individual support can be offered to students who are struggling with areas of the module.

Figure 1
A screenshot of a practice quiz question.

Question 7 Tries remaining: 3

Marked out of 1.00 | ⚐ Flag question | ⚙ Edit question

Suppose that currently the consumer prices index (CPI) is 321.5 and that at the same time last year it was 305.0.

If an index-linked pension was £1132 per month this time last year, calculate how much it should be per month now. (Give your answer to the nearest pound. Do NOT include the '£' sign. Assume the pension is index linked using the CPI.)

Answer: [   ]

[Check]

Screencasts

There are several studies which have shown that videos of lectures help students, particularly students on service courses, assimilate material (Jaki, 2009). Short videos, roughly 5 minutes in length, known as screencasts, are made available on demand via the VLE (Figure 2). There are over 70 screencasts in this module which talk through examples to help explain key concepts. Each of the screencasts has subtitles and transcripts available. In the screencasts students hear from statisticians with different accents to help convey the message that successful statisticians come from a variety of backgrounds. By keeping the videos short, students are more likely to access the material multiple times and therefore this medium is particularly suited to helping students understanding key statistical concepts (Moore, 1993).

Figure 2
A screenshot of a screencast.
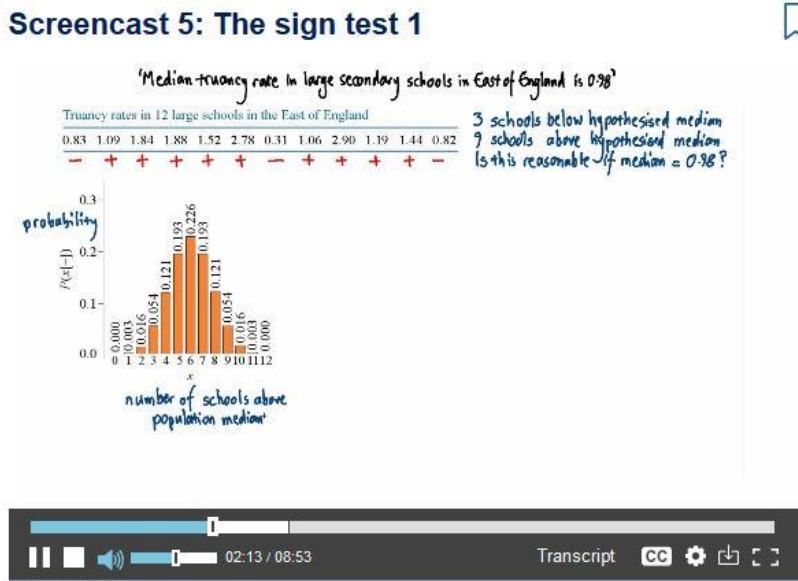
Computer simulations

For many years' computer simulations have been used to aid students with probability and statistical learning (Simon et al., 1976). Not only are animations useful when explaining difficult concepts but they also provide a change of study pace and media which helps to improve the student's attention and concentration (Freeman et al., 2008). In total there are ten interactive computer animations, and these have been key resources for many students in overcoming barriers to learning concepts. For example, the first interactive computer animation is used to give students the opportunity to explore the sensitivity of the mean, standard deviation and range and compare these with the robustness of the median and quartiles (Figure 3).

Figure 3
A screenshot of an interactive computer animation.
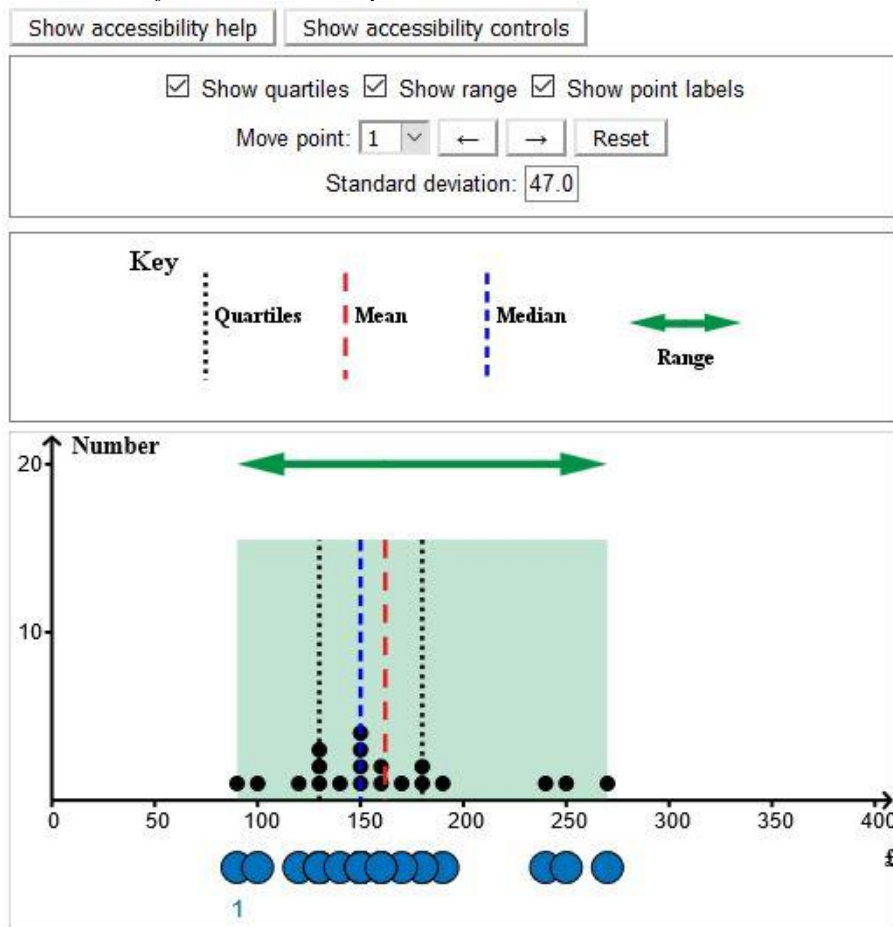

Software

Students are given a variety of real-life datasets and use software to carry out a range of different analyses. When real data is used students develop the ability to link statistics with real world situations (Mustafa, 1996). Whilst the ideal might be to use data linked to the student's qualification, as has seen to be the case for medical students (MacDougall et al., 2020), this is not possible when the module must serve students on multiple qualifications. By using real data, which is of interest to all citizens, the students have the opportunity to both visualise and interact with the data in a meaningful way, which is widely held as a way of improving a student's understanding of data (Weissglass & Cummings, 1991). Software use allows students to experience data analysis without the drudgery of numerical computation. In this module the chosen statistical package is Minitab. The package was chosen for its ease of use and straightforward output. Students install this software on the computer they use for study. Whilst this removes the need for students to access a computer lab, it means that support needs to be in place to assist students who run into installation problems. This support is provided by OU associate lecturers through VLE forums and a centralised OU helpdesk.

Experiment

Several different studies have shown that students learning is improved when they engage with activities that allow them to compare their beliefs with empirical results, (Shaughnessy 1977). One of many successful studies in using Problem-Based Learning (PBL) is in the Faculty of Science and Technology at Lancaster University on a doctoral development programme; students on a research methods course engage in PBL to develop research methods and statistics skills which will be needed in their doctoral studies in other disciplines (Jaki & Autin 2009). PBL is not easy to achieve within a distance-based environment. However, *Introducing statistics* replicates the idea by asking student to grow mustard seeds under different experimental conditions and analyse the root lengths (Figure 4). This forms part of the assessment for the module. This idea of comparing experimental conditions is then linked to clinical trials.

Figure 4
A photograph of the mustard seed experiment.
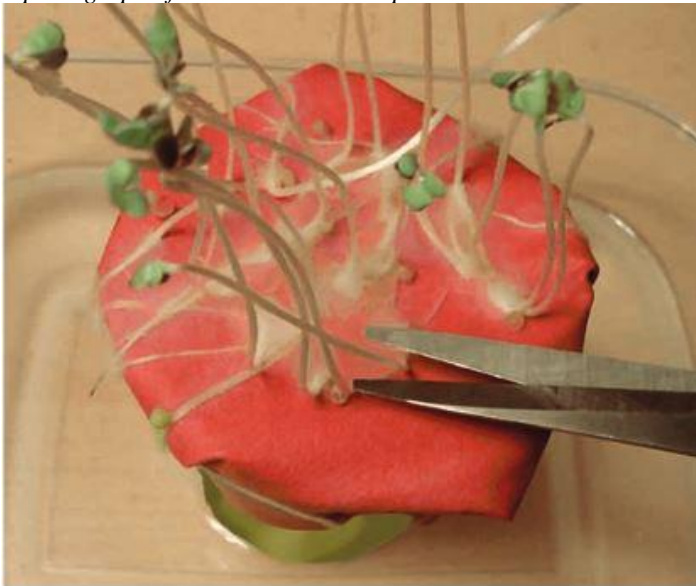
Each of these interactive activities are embedded into the VLE and students are pointed towards the appropriate unit text and activities which need to be studied each week. In this regard the OU could be described as the ultimate 'flipped classroom'. In a flipped classroom the students are introduced to new topics outside the classroom; classroom time is then used to explore the topics. At the OU the equivalent classroom interactions take place with associate lecturers in tutorials, which take place both face to face and online, and through correspondence tuition via assessment feedback. Online forums are embedded into the VLE which enable both peer to peer learning, together with input from the academics running the module. In a study Khan & Watson (2018), showed how the exam performance was improved by using a variety of flipped classroom techniques on a large level 1 statistics module.

Assessment

The module has no final examination but is assessed throughout the module presentation in a variety of different ways. There are four tutor marked assignments (TMAs) which are marked by an associate lecturer who whilst doing so provides personal and comprehensive correspondence tuition to the student. One of the TMAs asks the student to write a report about their mustard seed

experiment. In addition to the TMAs, there are three interactive computer marked assignments (iCMAs). These iCMAs have the same format as the practice quizzes but students are only allowed one attempt at each question. Feedback is provided to the student promptly after the iCMA submission date. As each iCMA is generated from a bank of questions each student receives a different, but equivalent, set of iCMA questions to prevent plagiarism. An end of module assessment (EMA) is comprised of another set of iCMA questions and a written assessment. The latter is always focussed on a single research topic which tries to capture the interest of all students. Previous topics have included data on topics as diverse as orienteering, ornithology, hurricanes and snails.

Evaluation of the module

Introducing statistics was presented for the first time in 2013. In the first presentation the module did not show the usual, marked attainment gaps for Black and Asian students, disabled students and students whose qualification prior to OU study was less than two A-levels, or equivalent. Each of these groups had a pass rate of around 70% compared to the overall pass rate of 76%. This has been sustained in subsequent presentations (Hilliam & Calvert, 2019) until the most recent presentation, which started in October 2019 and ended in June 2020, where there was a slight drop in the pass rate for the module to 73.4%. This drop may have been because of the Covid19 pandemic as many OU students have jobs and caring responsibilities which may have disrupted their studies more than usual in this difficult period. The attainment gap for the more educationally vulnerable groups detailed above remained in line with previous presentations. However, the pass rate for the module still exceeded the average pass rate of all OU level one modules which was 71.9%.

Students give feedback at the end of each presentation by filling in a module evaluation survey. In the first presentation in 2013, 89% of students were satisfied with the quality of the *Introducing statistics* module. Of the students who declared a disability, 68% of them said that they *definitely agreed*, and a further 19% *agreed*, that they could work with the material and learning activities. The remaining 13% *neither agreed nor disagreed.* Positive satisfaction amongst disabled students studying the module has been indicated in the module evaluation survey by comments such as:

> "*I found it helpful due to my hearing difficulty to receive transcripts of the tutorials.*"

The pass rates for students studying qualifications outside of mathematics, statistics and data science remain high. For example, in the latest presentation, the pass rate for economics students was 76% and 78% for students studying the combined STEM degree (a degree where the student can take any module in science, technology, engineering and mathematics subject to the correct number of credits being studied at each level). The module was also taken by 81 students studying a new degree in Business & Management Science. These students had a much lower pass rate of 61%; more evaluation is needed to understand why their attainment was not higher.

Students studying qualifications outside of mathematics, statistics and data science commented:

> "*I am studying for an open degree, mostly in the field of Social Science, and this module provides a good foundation for understanding the statistical methods used by Social Scientists to develop social theory.*"

The gentle introduction to the subject by means of real-world experiences has meant that all students can engage with the material, including those coming to the module with a prior mathematics and/or statistics 'phobia' as the following feedback exemplifies:

> *"As a student at school, I used to hate statistics. I studied S1 at A-level and decided to go no further than that, because statistics seemed like a horrible side of mathematics to me. However, upon starting to study this module, I found myself intrigued by statistics. ...The way that the information is presented to the student portrays statistics as an intriguing science, giving examples of how it is used in real-life situations. After seeing how useful statistics actually is to society in practice, I have definitely changed my mind about the science."*

The use of real data in areas of interest to all students also appeals to students who already have some statistical knowledge:

> *"I have some previous experience of statistics in the workplace, and was keen to increase my basic foundation in the subject. This was an ideal course with which to start that process. It was appropriately challenging and yet well grounded in reality. Indeed, the subjects chosen to illustrate the various statistical techniques were drawn from contemporary debate and, from my perspective, shone fresh light upon sometimes contentious matters. A genuinely interesting course!"*

Looking ahead to a new module: Applied statistical modelling

For many years it has been possible, in the OU, to evaluate and reflect on how students engage with different aspects of the module resources. Extensive feedback from students provided by responses to the module evaluation survey and gleaned from their forum postings has provided a rich and comprehensive resource for module and qualification developers. More recently, as all the module material has been embedded in the VLE, it has become possible, with the aid of analytics, to track how students use the various online resources. Information, for example, gained from online quizzes can be particularly helpful in identifying concepts with which some students struggle.

Information gleaned from student engagement with elements of *Introducing statistics*, is helping ascertain how students use the different assets. Just one example would be to see if students from different disciplines use certain screencasts more than others. Information gained from online quizzes can be particularly helpful in identifying concepts that some students might find difficult.

This experience is being put to good effect when designing the new module *Applied statistical modelling* which will run for the first time in 2022. *Applied statistical modelling* will be a compulsory module in BSc (Hons) in Mathematics and Statistics, BSc (Hons) in Data Science and the BA (Hons) in Economics as well as forming an optional module in many other qualifications. This again means that a module which can concurrently serve multiple audiences, including statistics specialists and non-specialists, needs to be constructed.

The main statistical theme running through the module will be linear and generalised linear modelling. These techniques are widely used by statisticians and others as they are of utmost importance to anyone modelling and analysing large and complex datasets. The module aims to explicitly develop students' general modelling skills such as describing and interpreting models using non-technical language, selecting appropriate models from a wide range of possibilities, model checking and model comparison. More general data issues such as handling missing data and the ethics surrounding data and its use will be explored. There will also be an emphasis on developing students' communication skills. These are skills that are relevant to students independently of the qualification in which they are interested.

In *Introducing statistics*, computer animations are available for six out of the 12 units. It has been identified that these are the least used of all the resources with less than half of the students viewing them, (Figure 5). More popular resources are the screencasts and practice quizzes, although use of both resources decline throughout the presentation. This pattern of decline is common in OU modules as many students have other responsibilities and can easily fall behind the recommended study calendar. In order to catch up, the advice given to such students is to concentrate on the core texts provided. The increased use of the screencasts and, to a lesser extent, the practice quiz for Unit 6 in the module is worth noting. This unit introduces the idea of hypothesis testing. Students often find this concept difficult so make use of the available resources to help them understand. In the new module, *Applied statistical modelling,* screencasts will be used to reinforce the more difficult content in the units by allowing the material to be re-expressed in a dynamic and informal way. In addition, it is intended that further screencasts will be produced to introduce the contexts surrounding the data analysis. It is hoped that these screencasts will help students develop the skill of translating the requirements on a non-statistical specialism into statistical terms as well as enthusing students about the contribution that data analysis can play in a multidisciplinary context.

Figure 5
Graph showing the percentage of students on the 2019 presentation of Introducing statistics who used the resources.
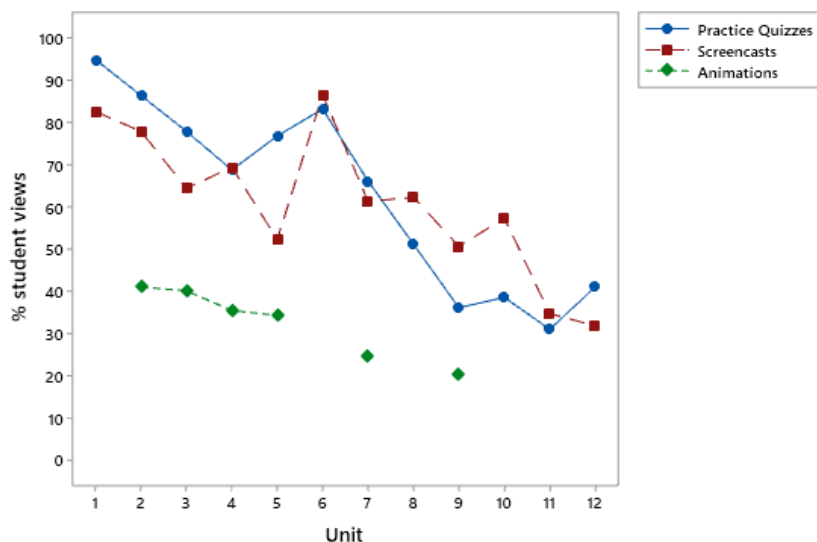
One of the few negative pieces of feedback that *Introducing statistics* receives each year regards the use of the computer package Minitab. There are several students who study the module who express a desire to use opensource software so that they can retain the use of the software beyond the end of their module. The new module, *Applied statistical modelling,* will therefore use the statistical computing environment R (R Core Team, 2020) to implement the statistical techniques they are taught. R is open source and multi-platform. However, the greater use of the command line in R compared to other statistical packages means that it is likely to be more daunting to students who are less confident with software. To overcome this the teaching of R will be delivered through Jupyter notebooks (Kluyver et al., 2016). The name Jupyter is derived from the combination of Julia, Python, and R. Jupyter notebooks are an opensource web application in which it is possible to share documents that contain live code, equations, visualisations and narrative text, https://jupyter.org/. A

complete analysis can be recorded and reported in a Jupyter notebook in the same way other sciences use a lab notebook to record tests, progress, results and conclusions (Toomey, 2018). These electronic documents allow R executable code (which can be modified by students) to be mixed in with explanatory text. This will enable the learning of R to be broken down into small chunks allowing a scaffolded approach to be taken. As animations can be embedded in the notebooks, it is hoped that using this approach will increase the use of animations from the disappointing uptake which is seen in *Introducing statistics*. The notebooks also provide a way in which students can present results and conclusions, alongside the code used to generate the results, using Markdown to format their text. Markdown is a straightforward mark-up language which, experience elsewhere in the university suggests, students should readily get to grips with. The notebooks also provide students with a means of providing an 'audit trail' of their data analyses, something that is regarded as good working practice.

It anticipated, when used as part of the teaching, that students will be presented with two notebooks. The first notebook will take the form explanatory text and example code along with questions and spaces for answers to be added (both in the form of code and in the form of prose providing interpretation of results). The second notebook will then be a model answer. The expectation is for students to submit answers to assignments in the form of Jupyter notebooks. Associate lecturers will then access these notebooks and see the results and conclusions alongside the code used to generate the result. This will enable them to provide feedback on all aspects of the students' work. This is particularly important given students' differing qualification goals, and hence different likely strengths. Jupyter notebooks are already used on other, non-statistical, modules, including some that students might study alongside *Applied statistical modelling*. Therefore, by choosing Jupyter notebooks, rather than alternatives such as RMarkdown, we will be using an environment that some students will already find familiar. Furthermore, it allows us to draw on experience and expertise that exists elsewhere in the OU to support students and associate lecturers in the use of the notebook software.

Based on the experience of *Introducing statistics*, this new module, *Applied statistical modelling* will motivate and illustrate statistical techniques using real-life datasets which should be of interest all students. For example, the intention is to include data from the citizen science project Treezilla (www.treezilla.org). This project, based on the idea to map every tree in Britain, taps into exploring the benefits that trees bring to the environment, particularly the urban environment. Thus, it deals with issues relevant to students' everyday lives as well as being a project which students can contribute to with the data they collect. Using this type of data means that it also possible to explore challenges with data collected from citizen science sites; such as the variable skill level and abilities which contributors to the site might have and locational dependence in the data. Other datasets that are being considered include household survey data from the Philippines obtained from https://www.kaggle.com/grosvenpaul/family-income-and-expenditure and performance data of national teams at the Olympics. The Olympics example was inspired by a BBC News website article: BBC News, 2nd August 2016 entitled 'Predicting the Rio Olympic medal table', https://www.bbc.co.uk/magazine-3655132 which led to an article by Bredtmann et al., (2016). The use of such datasets allows an international dimension to be brought to examples. Finding suitable data is not an easy task. Tactics such as searching repositories of datasets, searching journal articles, asking academic colleagues in other disciplines and using personal knowledge and enthusiasms are all used.

The new module will introduce students to one of two statistical specialisms: econometrics and data science. The econometrics teaching will provide students with the specialist knowledge about the application of statistics to economics expected for that discipline. Similarly, the data science

teaching will provide knowledge in two key areas: 'big data' and cluster analysis. In a departure from the design of previous modules these strands will run in parallel and be integrated with the main module material allowing students to exert choice over which specialism they study. Thus, it enables a small amount of tailoring of material within the module to suit different audiences whilst keeping a consistent core.

There is an expectation that practice quizzes will also be incorporated in the new module and embedded in the VLE. The quizzes are likely to be a combination of multiple-choice type questions and question using STACK (System for Teaching and Assessment using a Computer algebra Kernal). The latter allows students to input algebraic answers which STACK then checks to ensure the answer provided is algebraically equivalent to the correct answer and is in an appropriate form. STACK is freely available and can be explored from the webpage https://www.ed.ac.uk/maths/stack. STACK has shown to be of value in many modules at the OU and elsewhere (Pelkola et al., 2018). In addition CodeRunner, a free opensource plug in for VLE Moodle which allows students to check computer code https://coderunner.org.nz/, is being explored.

Discussion

This paper has argued that there is a growing need to provide statistical education to students in a wide range of disciplines in order to increase the statistical literacy in society (Lancaster et al., 2009). However, the challenge of providing statistical education through so called 'service courses' is problematic as the attitude of these students and their statistical confidence and ability is often quite different from those students on designated statistics qualifications.

One way of addressing the challenge has been for modules to embed examples from within the students' own discipline areas. This paper has outlined an alternative and successful, approach of incorporating real life data, of interest, and accessible to all citizens, into modules. Very importantly this approach has also enabled the OU to ensure that large cohorts of students, studying a range of qualifications from a range of backgrounds, can be taught simultaneously successfully.

Feedback data has been gathered from students on a level one statistics module, *Introducing statistics*, since its introduction in 2013. The module takes a narrative approach and uses examples in the areas of money, education and health to explain statistical techniques. In this way the module is designed to be of interest to all students regardless of their qualification and career aspiration. Typical feedback from students show that the module engages students by thinking about how statistics are used outside of their qualification:

> *"I wasn't sure what to expect of this course, I was interested in the statistics as much to see what statistics was all about. Certainly an eye-opener!! Most people tend to think that politicians, news and advertising media come up with figures to back up what they want to come across in their spiel, so you tend to ignore stats figures as just figures that generally are thrown about. Well, now I know that this is not strictly true, there are good statistics that are very definitely necessary and helpful and interesting. I do know that statistics are used in every walk of life, you need them"*

The quote also re-emphasises the idea that by choosing examples of general interest the module has fulfilled its aim of giving students the skills needed to reinterpret statistics in a personally meaningful way rather than as a meaningless tool from a potentially incomprehensible toolbox. Indeed, the module aspires to fulfil the aim of Garfield (1995) for students to apply ideas in new situations and

contexts in order to think critically and communicate their ideas. By using real data students can see the relevance of statistics and this increases their understanding and motivation (Neumann et al., 2013) whilst also providing a mechanism for them to feel that their statistical knowledge is a connected, rather than an isolated, technique which is important for students (Biggs & Tang, 2011). The idea of connecting statistics with situations that students can relate to has also helped students who have exhibited an anxiety towards statistics as this comment from a student show:

> *"Prior to starting this module, I had a fear of statistics that stemmed from when I was doing it as an AS level student and used to dread seeing the words standard deviation. This module has made more sense of the area of maths I have avoided for almost a decade, by applying it to real world situations and has made me determined to finish my degree."*

By evaluating the analytics on the VLE it has proved possible to identify which resources students utilise and value the most. Screencasts covering more difficult concepts have relatively high viewing figures. Whilst the interactive computer animations are used less by the students, qualitative feedback has nevertheless shown that students who do use these resources benefit from interacting with the animations.

> "*I have worked with statistics for many years, but this module opened my eyes to many aspects of statistics I had not thought about. For instance, I really enjoyed the interactive activities designed to help understand the difference between the mean and median! I also enjoyed watching the screencasts."*

The relatively high use of screencasts has meant that, in preparation for the new module, screencasts will be prioritised over animations. Ways to embed animations into the Jupyter notebooks will be explored and further evaluation will indicate whether this increases the number of students using these resources.

This idea of using a narrative based on real data which is of interest to everyone is currently being used to write the new level 3 statistics module. Students will use Jupyter notebooks both in their learning and for assessment. Sowey (1995) shows how assignments based on real-world problems can help students embed their understanding of the statistical method. It is hoped that by using real data, which students will analyse and report through the notebooks on this course, this will embed their statistical understanding and enable them to use these skills in their chosen career.

This paper has aimed to outline ways in which providers can use the same module to serve diverse student cohorts across a range of disciplines by using real data which is of interest to everyone. Given the data-driven society in which we now live, greater numbers of statistically literate employees will be required. The need to ensure employers are provided with graduates who have 'fit for employment' data analytical skills is the responsibility of all educators. In order to deliver this provision at scale universities may need to move away from individual bespoke service courses and re-imagine these courses for a broad audience, whilst doing so in a way that engages students and provides them with the ability to communicate in a statistically literate way.

References

Biggs, J., & Tang, C. (2011) *Teaching for quality learning at university.* McGraw-Hill.

Blei, D. M., & Smyth P. (2017). Science and data science. *Proceedings of the National Academy of Sciences*, 114(33), 8689-5692. https://doi.org/10.1073/pnas.1702076114

Bredtmann, J., Crede, C. J., & Otten, S. (2016). Olympic medals: Does the past predict the future? *Significance*, 13(3), 22-25. https:// doi.org/10.1111/j.1740-9713.2016.00915.x

Davenport, T. H., & Patil, D. J., (2012). Data scientist: the sexiest job of the 21st century. *Harvard Business Review,* 90(10), 70-76.

Dietz-Uhler, B., Fisher, A., & Han A. (2007). Designing online courses to promote student retention. *Journal of Educational Technology Systems,* 36(1), 105-112. https://doi.org/10.2190/ET.36.1.g

Fayyad, U., & Hamutcu, H. (2020). Toward foundations for data science and analytics: A knowledge framework for professional standards. *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.1a99e67a

Freeman, J. V., Collier, S., Staniforth, D., & Smith K. J. (2008). Innovations in curriculum design: A multi-disciplinary approach to teaching statistics to undergraduate medical students *BMC Medical Education,* 8(28). https://doi.org/10.1186/1472-6920-8-28

Garfield, J. (1995). How students learn statistics. *International Statistical Review*, 63(1), 25-34.

Georgakis, S., Wilson, R., & Hu, X. (2011). Addressing the teaching and learning challenges of 'service courses': a case study of pedagogical reform and innovation, *The International Journal of Learning,* 18(1), 233-244. https://doi.org/10.18848/1447-9494/CGP/v18i01/47421

Gordon, S. (2004). Understanding students' experiences of statistics in a service course. *Statistics Education Research Journal*, 3(1), 40-59.

Gordon, S., & Nicholas, J. (2010). Teaching with examples and statistical literacy: views from teachers in statistics service courses. *International Journal of Innovation in Science and Mathematics Education*, 18(1), 14-25.

Hilliam, R., & Calvert, C. (2019). Interactive statistics for a diverse student population. *Open Learning: The Journal of Open, Distance and e-Learning.* 34(2), 203-210. https://doi.org/10.1080/02680513.2017.1386095

International Data Science in Schools Project. (2019). *Curriculum frameworks for introductory data science.* http://www.idssp.org/files/IDSSP_Frameworks_1.0.pdf

Jaki, T. (2009). Recording lectures as a service in a service course. *Journal of Statistics Education,* 17(3). https://doi.org/10.1080/10691898.2009.11889534

Jaki, T., & Autin, M. (2009). Using a problem-based approach to teach statistics to postgraduate science students: a case study. *MSOR Connections,* 9(2), 40-47.

Kahn, R. N., & Watson, R. (2018). The flipped classroom with tutor support: an experience in a level one statistics unit. *Journal of University Teaching & Learning Practice,* 15(3). https://ro.uow.edu.au/jutlp/vol15/iss3/3

Kluyver, T., Ragan-Kelley, B., Perez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., & Willing, C. (2016). Jupyter Notebooks - a publishing format for reproducible computational workflows. In F. Loizides, & B. Schmidt (Eds.), *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (pp.87-90). IOS Press.

Lancaster, G., Francis, B., & Allen, R. (2009). Lancaster Postgraduate Statistics Centre – creating enterprise and innovation in teaching statistics across disciplines. *MSOR Connections*. 9(1), 41-46.

MacDougall, M., Cameron, H. S., & Maxwell, R. J. (2020). Medical graduate views on statistical learning needs for clinical practice: a comprehensive survey. *BMC Medical Education*, 20(1). https://doi.org/10.1186/s12909-019-1842-1

Moore, D. S. (1993). The place of video in new styles of teaching and learning statistics. *The American Statistician.* 47(3), 172-176. https://doi.org/10.1080/00031305.1993.10475970

Mustafa, R. Y. (1996). The challenge of teaching statistics to non-specialists. *Journal of Statistics Education.* 4(1). https://doi.org/10.1080/10691898.1996.11910504

Mevarech, Z. (1983). A deep structure model of students' statistical misconceptions. *Educational Studies in Mathematics*, 14(4), 415-429.

Neumann, D. L., Hood, M., & Neumann, M. M. (2013) Using real-life data when teaching statistics: student perceptions of this strategy in an introductory statistics course. *Statistics Education Research Journal.* 12(2), 59-70.

Onwuegbuzie, A., & Wilson, V. (2003). Statistics anxiety: Nature, aetiology, antecedents, effects and treatments – a comprehensive review of the literature. *Teaching in Higher Education*, 8(2), 195-209. https://doi.org/10.1080/1356251032000052447

Pelkola T., Rasila A., & Sangwin, C. J. (2018). Investigating Bloom's learning for mastery in mathematics with online assessment. *Informatics in Education.* 17(2), 363-380. https://doi.org/10.15388/infedu.2018.19

QAA. (2014). *The frameworks for HE Qualifications of UK degree-awarding bodies.* https://www.qaa.ac.uk/quality-code/qualifications-and-credit-frameworks

Royal Society. (2019). *Dynamics of data science skills: How can all sectors benefit from data science talent?* https://royalsociety.org/-/media/policy/projects/dynamics-of-data-science/dynamics-of-data-science-skills-report.pdf

R Core Team, (2020). *The R Project for Statistical Computing* https://www.R-project.org

Shaughnessy, J. M. (1977). Misconception of probability: An experiment with a small-group activity-based, model building approach to introductory probability at the college level. *Education Studies in Mathematics.* 8, 285-316. https://doi.org/10.1007/BF00385927

Simon, J. L., Atkinson, D. T., & Shevokas, C. (1976). Probability and statistics: Experimental results of a radically different teaching method. *American Mathematical Monthly.* 83(9), 733-739.

Sowey, E.R. (1995) Teaching statistics: Making it memorable. *Journal of Statistics Education.* 3(2). https://doi.org/10.1080/10691898.1995.11910487

The Open University (OU). (2020, September). *Facts and figures.* http://www.open.ac.uk/about/main/strategy-and-policies/facts-and-figures

Tishkovskaya, S., & Lancaster, G. A. (2012). Statistical education in the 21st century: a review of challenges, teaching innovations and strategies for reform. *Journal of Statistics Education*, 20(2). https://doi.org/10.1080/10691898.2012.11889641

Toomey, D. (2018). *Jupyter Cookbook.* Packt Publishing, Birmingham.

Vines, K., Hughes, C., Alexander, L., Calvert, C., Colwell, C., Holmes, H., Kotecki, C., Parks, K., & Pearson, V. (2019). Sonification of numerical data for education. *Open Learning: The Journal of Open, Distance and e-Learning*, 34(1), 19-39. https://doi.org/10.1080/02680513.2018.1553707

Weissglass, J., & Cummings, D. (1991). Dynamic visual experiments with random phenomena. In W. Zimmermann, & S. Cunningham (Eds.), *Visualization in mathematics*, (pp. 215-223). Mathematical Association of America.