Growing Virtual Plants for Teaching and Learning Statistics

Michael Bulmer

Department of Mathematics
University of Queensland
m.bulmer@uq.edu.au

Abstract We describe the creation of a collection of virtual plant movies as an educational tool for teaching and learning statistics in large classes. The motivation for this is discussed, along with experiences in the delivery and evaluation of the virtual plants in a computer laboratory setting. The development was carried out very cheaply and details of the production process used to achieve this are also given.

Introduction

Teaching large service courses in statistics is a difficult challenge, requiring the motivation of students with a range of interests and mathematical backgrounds. Additionally a current concern for many students is time pressure and the need for efficiency. Lectures and laboratories based on virtual plants aim to address these issues. They motivate theories and methods by providing a concrete setting and emphasise qualitative aspects of the course, allowing students with weaker mathematical backgrounds to gain confidence. They achieve these aims without adding to student workloads, an important efficiency for large classes.

The virtual plants discussed in this paper are a collection of QuickTime animations of fields of plants, together with a simple interface for incorporating them into a statistical computer laboratory class. The animations arose from models based on so-called L-systems, an algorithmic language that has been successfully used to capture the dynamic structure of plants (Prusinkiewicz and Lindenmayer, 1990). We describe how these were generated in Section 2.

Most introductory statistics textbooks are rich with real data sets, allowing students to relate the results of their explorations and analyses back to real scenarios. This is certainly desirable; Cobb and Moore (1997) suggest that "statistics requires a different *kind* of thinking [to mathematics], because data are not just numbers, they are numbers with a context." However, it is still a somewhat passive experience because the students have to take the context for granted. They have not been involved in obtaining the data and so lack ownership of the setting. Mackisack (1994) gives an overview of the other benefits of experimental work. For instance, the students also get an appreciation of the practical issues involved in carrying out experiments and collecting data, an outcome encouraged by Higgins (1999). The first aim of using the virtual plants was to engage the students in thinking about the design of the experiment and the origin of the variability in the data, while not allowing this to be so time consuming that the rest of their learning suffers. We discuss how this was achieved in Section 3.

The second aim of the plants was to make students aware of the role of the measurement process. The way the measurements are carried out, with possible errors and biases, can have profound effects on the subsequent statistical analyses. The plants aimed to provide the students with a difficult setting for carrying out measurements to make them think about these issues. In Section 4 we discuss the measurement task and summarise the student evaluations of the project.

This project would not have been possible without the L-systems expertise of Dr Jim Hanan in the Centre for Plant Architecture Informatics at the University of Queensland. The use of the virtual plants in teaching and learning was tested by Lesley Neely, a mathematics teacher at the West Moreton Anglican College in Brisbane.

Growing the Plants

At the time the virtual plants were created there was no simple technology for delivering them to large numbers of students. The ideal was to have a Java-based plug-in for a web browser that could produce images and allow interaction, but no such software existed and it was too expensive to develop this ourselves. The plants could be exported as VRML, but no reliable plug-in for Macintosh was found and it was conceptually difficult to incorporate growth over time in this framework. Purchasing stand-alone copies of the L-Studio software was undesirable, since they would have to be run under Virtual PC, and also too expensive, being priced at roughly US$300 per copy and with 120 computers in the laboratory needing licences[1]. The full software was also inappropriate; while some students might take the opportunity to explore the L-systems further, for the vast majority of students whose focus was on the statistics this would have caused confusion.

The solution was to create a number of QuickTime animations of fields of plants grown under a variety of conditions. (The way in which these were later used is discussed in Section 3.) This was a cheap solution since the production process could be easily automated and no special software was needed to view the movies, making the laboratory manager particularly happy. It also gave the opportunity, in a revised version of the plants, to include a ray-tracing step in the process to give better quality than could have been obtained in any real-time implementation.

Each animation showed a field of 12 plants in 4 rows, with 2 rows having one growth rate and the other 2 rows having a (possibly) different growth rate. The production process was as follows:

1. A Perl script was used to generate 400 files describing L-systems. These files were nearly identical except that a pair of growth parameters, each ranging from 1 to 20, were changed from file to file.

2. L-Studio, running under Virtual PC, was used to animate the growth of each L-system over 20 time steps. The output from this was a set of 20 files in the input format for Rayshade, a freeware ray-tracing program.

3. Rayshade, running on Unix, was then used to process each of the 8000 ray files, producing RLE image files.

[1] Volume discounts for L-Studio were available so US$36000 (120 x $300) is an overestimate of this cost. However, the project was carried out on a shoe-string budget and even the cost of a single license used most of the funds available.

4. GraphicConverter was used to batch convert the 8000 RLE images to TIFF, since unfortunately QuickTime did not recognise Rayshade's RLE format.

5. Finally, AppleScript was used to automate to process of telling QuickTime Player to open each set of 20 image files as an image sequence and to then export the resulting movie as a QuickTime animation.

All but one of these steps were carried out on a PowerBook G3 (400Mhz) running Mac OS X. Step 3, the ray-tracing of 8000 images, was very time consuming and so, with the laptop needed for other work, this task was given to an Sun UltraSPARC-II Unix box (2 x 450Mhz).

The ray-tracing took just over two days on this shared machine, with the other steps taking under two hours. An example of one of the 400 QuickTime animations, scaled to fit on the page, is shown in Figure 1.[2]

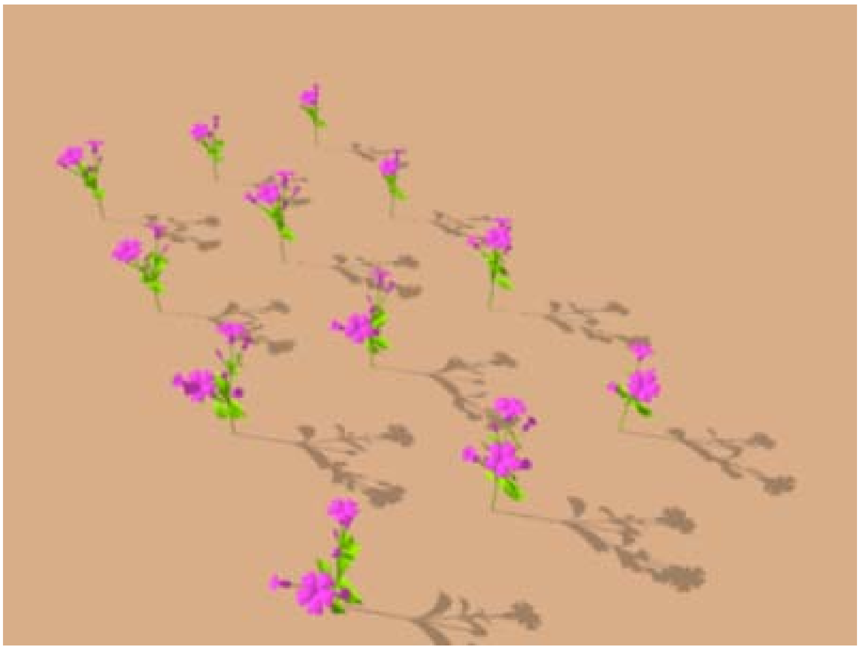

Figure 1

It is worth emphasising the cost-effectiveness of this approach. The combination of tools that could be brought together under Mac OS provided a straightforward and efficient means of developing these teaching materials. Given the limited funds available, this education solution may well have been unrealisable in another environment.

Delivering the Plants

The plant movies were placed on a web server and an interface was developed to link these movies to the course objectives. This interface was an HTML page using forms and JavaScript. JavaScript was chosen over a server-based system so that the plants could also be distributed on CD-ROM to secondary schools that might not have their own web server. As discussed below, it is also straightforward to modify the statistical model used by the interface, and this could be done by a computer novice by modifying two JavaScript

[2] *Editors note:* The final frame of the animation is shown here. The QuickTime movie can be found on the Proceedings CD as Bulmer.mov

formulas. In our university setting, the class of 550 students accessed the movies through a cluster of computer laboratories containing a total of 138 iMacs. This was also where they sat their final examination, so that their work with the plants was integrated with later formal assessment.

The first aim of using the plants was to give students the opportunity to have some input into the design of the experiment. The background given to the students was that they had a uniform plot of land in which they can grow 12 plants, 6 plants each in two subplots, to compare two treatments. There are two factors under consideration, the level of nitrogen fertilizer used (None, Low, Moderate, or High) and the irrigation level (None, Some, or Lots). Figure 2 shows the interface that students used to set their levels.

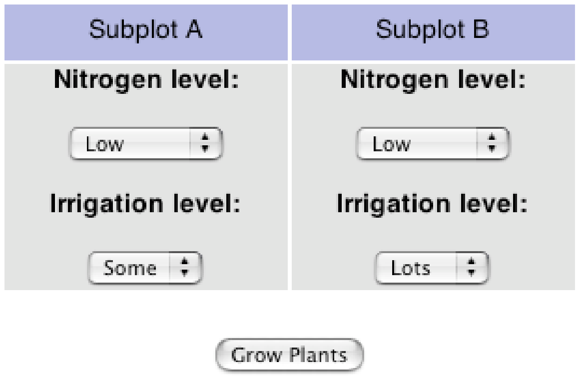

Figure 2. Form interface for setting experimental conditions.

If they hadn't realised it before, students quickly discover that to test for the effect of one factor they should keep the other factor fixed (at least in this simple setting). The choices in Figure 2 looks to compare the effects of extra irrigation in the presence of low levels of nitrogen fertilizer. Clicking on "Grow Plants" presents a movie of the resulting growth, as illustrated in Figure 1.

With 4 nitrogen levels and 3 irrigation levels, there are 12 possible treatments that could be applied to each subplot. As discussed in Section 3, there were 20 growth parameters used to make the plant movies (so movies for 8 of the parameter values are never used here). All the JavaScript code does when the user clicks "Grow Plants" is to map the specified treatment to a growth parameter, once for each subplot, and then link to the appropriate movie file. The map used here involved a quadratic response for fertilizer (so that High actually gives lower growth than Moderate) and a monotonic response for irrigation. For this example, Subplot A mapped to a growth parameter of 15 while Subplot B mapped to a growth of 17, quite a small difference to detect from 6 plants each.

By keeping the movies separate from this interface, new scenarios can be invented by providing a new interface and functional map between the inputs and the 20 available growth parameters. Other factors, such as available sunlight or the level of a second fertilizer ingredient, can be included by adding extra variables to the function.

The focus on this particular laboratory class was on comparing the total growth between the two treatments. There are also other issues to explore, such as the pattern of growth over time, that give students the freedom to formulate and test new hypotheses. They can also carry out more complicated factorial experiments by growing a new movie with different factor levels. This allows the discussion of more sophisticated statistical concepts, such as interaction effects.

Evaluating the Plants

Technological approaches to teaching such as this are often exciting for the developer but are ultimately worthless unless students enjoy them and, more importantly, can see that they will help their learning. At the end of the laboratory session the students were surveyed to obtain feedback on their use of the virtual plants. The survey question was very general: "Please comment on the use of virtual plants in this practical".

Overall the students were positive about the graphics. One student noted that this was better than a similar laboratory without graphics (involving virtual rats and blood pressure; see Bulmer, 2002) since you "could actually see the results in picture form which gave a better idea of what happened."

However, having set their factor levels, the students then have to get the data they needed for analysis. Rather than being given a measure of growth for each plant, the students have to deal with this measurement task. There are two main problems for them here. Firstly, they need to start by thinking about why they might be measuring plant growth. They could then measure the heights of plants if they wanted to see which treatment gave taller plants, or they could count branches, leaves, or flowers if they wanted to see which gave higher yields.

Measurements of height can be made from the screen using a plastic ruler. It is quite a satisfying experience to walk into a computer laboratory and see a room full of students with rulers up to the screens. They are physically engaging with the setting, rather than passively taking a set of mysterious numbers from a textbook exercise, or even from someone else's real study.

Measuring height with a ruler is difficult because it is hard to know where the top of a plant is; they branch outwards after an initial vertical growth. Counting flowers or leaves is difficult because it is hard to know you've seen them all, just as in real life. Furthermore, the plants drop flowers and leaves as they grow, so this counting must be over time. The measurement process should be difficult and students should have think about what simplifications or estimates they are making. One student complained that "it would have been more helpful and easier to use if some kind of variable was presented with the movie as a result (rather than making student measure it off the screen)" but most saw the purpose of the exercise in that it "made it more interesting than copying information out of a data set." Similarly, another student commented "I think this is really good, because it simulates real life situations, rather than 'ideal values'".

One of the reasons for not pursuing a VRML solution was to keep this measurement process non-trivial. Being able to move the plants around would have made it too easy! Ray tracing introduced a perspective projection that may actually make the measurement process too difficult, but it also brought a range of visual cues, such as shadows, that provide a richer environment to study.

In all the use of the virtual plants has been very effective, both in terms of time and in terms of engaging student interest. "It was efficient (non time consuming) and allowed for an understanding (better) through experiencing it (working it out)", as one student wrote, with another echoing that it was "a useful, enjoyable interactive learning experience". That students see the virtual laboratory as a "learning experience" is particularly significant as a sign of success.

Conclusions

This project met a number of aims. In addition to the main points discussed above, it also exposed first-year students to an active research area at the University. Virtual plants are being used for a wide range of applications, from examining the effectiveness of insecticide sprays to modelling biochemical regulation in plants (Hanan and Room, 1997).

Giving students the flexibility to set their own factor levels also means that students each have different data sets, which helps encourage cooperative learning (Magel, 1998) and broadens their learning experiences. Moreover, even if students chose the same factor levels, the measurements they obtained from the virtual plants may differ, again encouraging discussion about the measurement process and how this might affect statistical conclusions.

For the course in question, the students are also given projects in which they are asked to design and carry out their own experiments, followed by statistical analysis and the writing of a mock journal article. This is a rich form of assessment, but in a class of 550 it is difficult to have one-on-one discussions about each student's intended experiment and the issues they may face. The use of the virtual plants in the laboratory setting allows students to discuss experimentation aspects in an immediate way, using the simulation to highlight important points. In a similar way, the activity-based statistics of Schaeffer *et al.* (1996) works very well with small classes (Bulmer, 1999) but is impractical to implement in large classes. Tools such as the virtual plants can be used in this role, underlying a constructivist approach to learning statistics.

A number of other "virtual worlds" have been developed, including virtual rats and virtual goldfish, with others in preparation (Bulmer, 2002). Almost any statistical textbook exercise could be converted into such an experiment by first modelling the data and then using that model and random number generation to recreate the data "live". It gives students ownership of their work by putting the data into a context. The ability to add rich graphics helps further engage students while also adding new dimensions, such as the importance of measurement issues. This has been made possible by a cost-effective production process that arises naturally from an integrated computing environment.

References

[1] BULMER M. (1999) *Introductory Statistics with a Qualitative Emphasis* Effective Courses/Effective Teaching at University Conference. The University of Queensland. http://www.tedi.uq.edu.au/conferences/teach_conference99/papers/Bulmer.html

[2] BULMER M. (2002) *Teaching And Learning Mathematics With Virtual Worlds* Proceedings of the 2nd International Conference on the Teaching of Mathematics (at the undergraduate level). Wiley.

[3] COBB G.W. AND MOORE D.S. (1997) *Mathematics, statistics, and teaching* American Mathematical Monthly **104** (9), 801-823.

[4] HANAN J. AND ROOM P. (1997) *Practical aspects of virtual plant research* in *Plants to ecosystems*. Advances in Computational Life Sciences, Vol 1 (ed. Michalewicz, M.) 28-44. Melbourne : CSIRO Publishing.

[5] HIGGINS J.J. (1999) *Nonmathematical statistics: a new direction for the undergraduate discipline* The American Statistician **53** (1) 1-6.

[6] MACKISACK M. (1994) *What is the use of experiments conducted by statistics students?*, Journal of Statistics Education **2** (1) http://www.amstat.org/publications/jse/

[7] MAGEL R.C. (1998) *Using cooperative learning in a large introductory statistics class* Journal of Statistics *Educatio*n **6** (3) http://www.amstat.org/publications/jse/

[8] PRUSINKIEWICZ P. AND LINDENMAYER A. (1990) *The Algorithmic Beauty of Plants*. New York: Springer-Verlag.

[9] SCHAEFFER R.L., GNANADESIKAN M., WATKINS A., AND WITMER J.A. (1996) *Activity-Based Statistics*. New York: Springer-Verlag.