ON AND OFF-LINE DYNAMIC DATA INTERROGATION

Mark Crowley[1], Kate Richards[2] and Neville Davies[2]
[1]Centre for Effective Learning in Science, Nottingham Trent University, United Kingdom
[2]Royal Statistical Society Centre for Statistical Education, University of Plymouth, United Kingdom
mark.crowley@ntu.ac.uk

In this paper we describe an interactive Flash®-based data interrogation tool which interacts with large datasets of real data collected from learners around the world. The dynamic tool can be used online to access in real-time databases stored on the Royal Statistical Society Centre for Statistical Education web server. It can also be used offline as a stand-alone facility to access local databases stored on the user's PC. The tool has been developed to help improve learning and understanding of statistics and to demonstrate the importance of looking at data to help make evidence based decisions. We shall illustrate with examples of data interrogation aimed at helping students make trustworthy decisions from those data.

BACKGROUND

The *CensusAtSchool* project was devised in 1999 through partnership formed between the Royal Statistical Society Centre for Statistical Education (RSSCSE), Office for National Statistics (ONS) and Maths Year 2000 project. It was designed to raise awareness of the 2001 UK population census and to improve statistical literacy. With over 62,000 learners sending in data, this highly successful project has developed into a dynamic, ongoing and exciting initiative running in the UK and a number of countries including Australia, Canada, New Zealand and South Africa. Other countries, such as the USA and Japan, intend to adopt it.

Student responses are collected throughout the school year and then compiled to provide summary data tables for students and teachers participating in the project. Individual countries provide additional web-based resources and access to samples of the data collected from their online questionnaires. Teaching resources and ideas are also provided which help students learn to work with data and draw conclusions from it. Davies, Connor & Spencer (2003) report that *CensusAtSchool* can help students realize the importance of being statistically literate, and help develop skills needed to make evidence-based decisions.

SAMPLING 'ATSCHOOL' DATA

According to Davies (2008), "a key hook for learners of all ages appears to be the motivation of knowing that their data contribute to part of a large database of 'similar' learners". From ten years of experience which the RSSCSE has had in running *CensusAtSchool*, it is clear that students are keen to discover how they compare with other participants: How does my class/school/country compare with others?

In order to investigate these questions, students and teachers need access to the responses for various questionnaires. Databases of the responses are maintained by individual countries and backed up on a web server provided by the RSSCSE which contains over 1.2 million responses. Each database can be sampled over the Internet, using the RSSCSE's Random Data Selector (RDS), with all returns being anonymised so that no individual learner can be identified. To help learners to interrogate and manipulate the data returned by the RDS, both coding sheets and electronic copies of the original questionnaires have been made available for download.

Summary data tables for the *CensusAtSchool* (New Zealand) project carried out during 2005 (www.censusatschool.org.nz/2005) can be accessed in chart form. Once a user has selected various options with a series of drop-down menus, the webpage loads the appropriate chart from a database of prepared images (Figure 1). All 15,982 possible graphs were generated using R-code once all returned questionnaires had been finished and analysed. The graphs for this 'fixed dataset' were then stored on the NZ team's web server.

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php [© 2010 ISI/IASE]

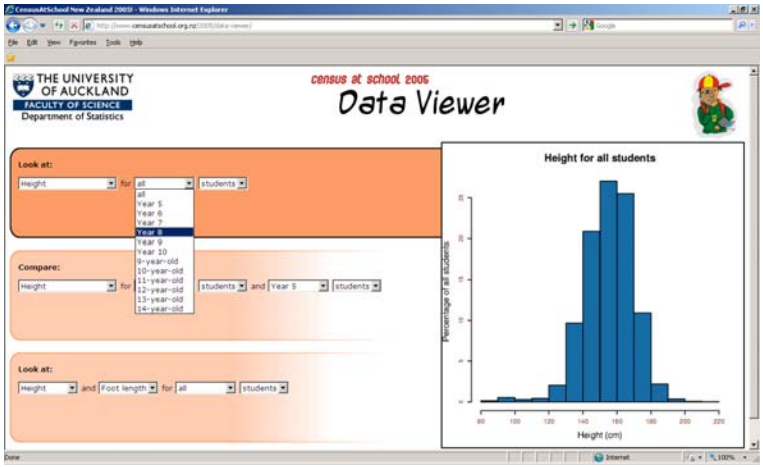

Figure 1. CensusAtSchool (NZ) 2005 Data Viewer

DYNAMIC ACCESS

Although the New Zealand tool was seen to be very useful, the *AtSchool* project team in the UK needed a method of interacting with the databases on-the-fly and create charts from dynamically changing data; in addition we wanted to allow students to compare their own classroom data with random samples of responses from their peers. Jacobs (2005) outlines some of the features supported by Macromedia Flash MX (now Adobe® Flash®) which our online learning environment required. Jacobs notes its potential to create "graphs which allow the user to interactively vary parameters".

At the outset of the project, only a small number of commercial packages existed which could create on-the-fly charts from online databases; one of these packages was the Adobe® Flash® driven XML/SWF Charts (www.maani.us/xml_charts). This package uses Extensible Markup Language (XML) to describe a chart in much the same way as Hypertext Markup Language (HTML) is used to create a webpage. The Adobe® Flash® content receives the XML file from a server and then renders the chart accordingly. Our first development was a simple online tool which could interrogate just one of *AtSchool* databases using a combination of ASP script, XML/SWF charts and webpage with a form. The form consisted of elements such as radio buttons and drop down menus which enabled the user to select the names of variables and to filter results by gender, age or region; the choices selected on the form were passed to the server, converted into a database query (SQL) and the resulting output was then converted into XML format. Although the process was user-friendly, the web page was considered cluttered and overly complex.

DRAG AND DROP

To simplify the user interface, we considered using an approach described by Britt & Gabrys (2004). A system they designed, to collect responses from participants, used 'drag-and-drop' options on a web page based on JavaScript. Their system enabled users to make decisions in a fairly natural way by moving a restricted number of elements around the screen - an interface familiar to many students. Unfortunately it was found that different browser capabilities and security permissions meant that JavaScript based drag-and-drop was not universally supported. Instead Adobe® Flash® was seen as more convenient route for creating drag-and-drop elements.

A bespoke approach was required to create a suitable application using Adobe® Flash® Professional software. The drag-and-drop interface was programmed using ActionScript™, the language used to create almost all of the interactivity (buttons, text entry fields) seen in many Flash® applications. A series of screens were designed through which users navigate to decide which database to interrogate, which names of variables to investigate and the size of random sample to collect. Optional screens were included so that users could add filters to obtain data from particular groups or categories such as boys or girls only. Colour coding was incorporated to help the user identify the types of data (continuous and discrete) stored in the *AtSchool* databases.

Furthermore, the colour coding was used to give the users a visual clue as to which charts continuous and discrete may be used to plot (Figure 2).

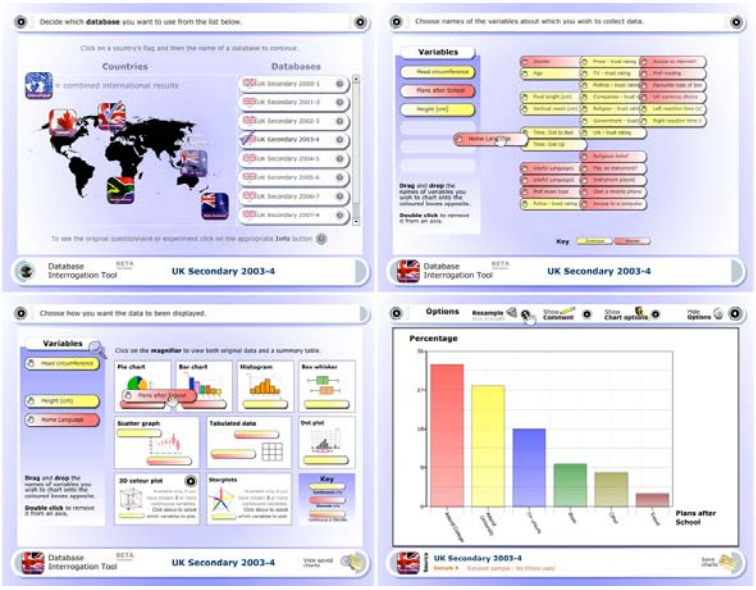

Figure 2. Screenshots of the fully functional version of Data Interrogation Tool (http://datatool.censusatschool.org.uk/datatool.swf)

INTERROGATING THE DATABASES

Pea (1987) suggests that providing students with the technology to produce many graphs quickly leaves students both the time and inclination to examine multiple graphs and different representations. While developing our Data Interrogation Tool, the RSSCSE team considered which graphical and visualization techniques would aid students in exploring and analyzing data quickly. Before the Data Interrogation Tool or *Datatool* was developed, students and teachers could obtain random samples from *AtSchool* databases as Microsoft Office Excel™ spreadsheets. These spreadsheets could contain up to seventy columns/headers and require a coding sheet to interpret the returned data. Rather than displaying raw data, once a random data sample has been returned by the server, the *Datatool* was set up to offer the user a choice to charts; the user needs only drag-and-drop the name of a variable on to an appropriate axis. The user may still view the raw data and export this via the clipboard into a spreadsheet.

As pupils are often interested in how they compare with their peers, an important function added to the *Datatool* was the ability to obtain and compare two samples. In Figure 3, the vertical reach of two random samples ($n_A=n_B=100$) have been plotted for (A) Males aged 12-13 and (B) Females aged 12-13. In addition to using random samples from the *AtSchool* databases users can add or overwrite their own data to the raw data collected by the *Datatool*.

Gibson et al (2007) describe the results of a project commissioned by the Qualifications and Curriculum Authority (QCA) in the UK. Resources developed for the RSSCSE/QCA project exemplify how the problem solving cycle can be used in conjunction with *CensusAtSchool* databases. The teaching materials and pupil activities developed were designed to reinforce the plan, collect, process, discuss stages of the problem solving cycle or Problem Solving Approach (PSA). In the resource 'Crime Scene Investigation' students are asked to consider whether or not a footprint left by a thief can be used to estimate the age/height/gender of teenaged suspect. As part of a preliminary investigation, learners can use the database interrogation tool to explore the data from, for example, the UK Secondary 2005 questionnaire.

Students can quickly create scatter plots to test potential relationships or hypotheses – the ability to quickly re-sample data using the tool provides them with valuable insights into the data. Students can then use their own class to gather data based on their initial findings or extend their investigation with secondary data from other countries via the *Datatool*.

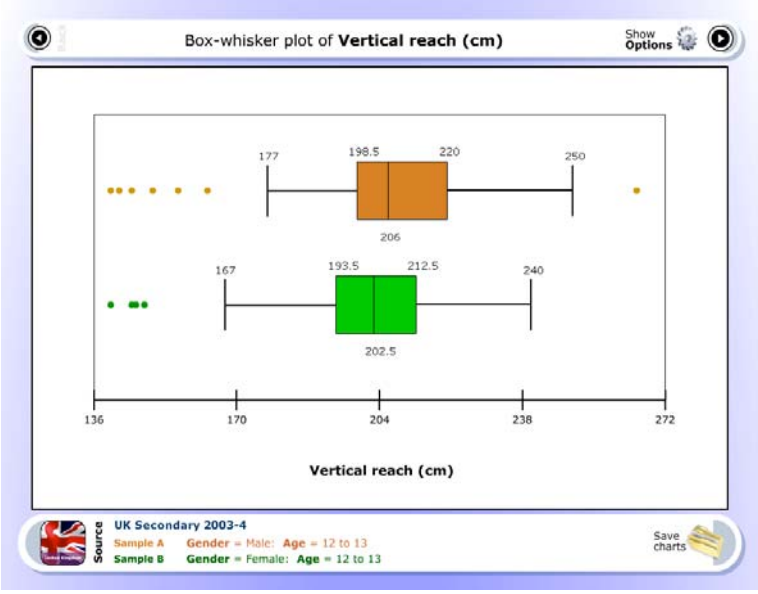

Figure 3. Example of chart comparing two random samples

CONCLUSION

Craig, Mehrens and Clarizio (1975), argue that learning tools that are perceived by students as enjoyable to use stimulate interest and create a more active learning environment. Clearly, learners enjoy web pages which contain interactive multimedia content such as games and videos. Adobe® Flash® is commonly used to embed such content in web pages and has proven very suitable for developing an interrogation tool to communicate dynamically with online databases. Chance et al. (2007) report how technology allows students to focus on interpretation of results and understanding concepts rather than on computational mechanics. Our work to design and build an interrogative tool allows learners and teachers to exploit the rich source of real data from a truly international and collaborative project.

REFERENCES

Britt, M. A., & Gabrys, G. (2004). Collecting responses through Web page drag and drop. *Behavior Research Methods, Instruments, and Computers, 36*(1), 52-68.

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The Role of Technology in Improving Student Learning in Statistics. *Technology Innovations in Statistics Education*, 1(1). Online: http://escholarship.org/uc/item/8sd2t4rr

Craig, R. C., Mehrens, W. H., & Clarizio, H. F. (1975). *Contemporary Educational Psychology: Concepts, Issues, Applications*. New York: John Wiley & Sons.

Davies, N. (2008), Developments of AtSchool Projects for Improving Collaborative Teaching and Learning in Statistics. July 29 [presentation] *Second CensusAtSchool International Workshop.* Los Angeles: Department of Statistics, UCLA.

Davies, N., Connor, D., & Spencer N. M. (2003). An International Project for the Development of Data Handling skills of Teachers and Pupils. *Journal of Applied Mathematics and Decision Sciences, 7*, 75-83.

Marriott, J. M., Davies, N., & Gibson, E. (2009). Teaching, Learning and Assessing Statistical Problem Solving. *Journal of Statistics Education, 17*(1).

Jacobs, K. L. (2005), Investigation of interactive online visual tools for the learning of mathematics. *International Journal of Mathematical Education in Science & Technology, 36* (7), 761-768.

Pea, R. D. (1987). Cognitive technologies for mathematics education. In A. H. Schoenfeld (Ed.), Cognitive science and mathematics education (pp. 89-122). Hillsdale, NJ: Lawrence Erlbaum Associates