Lucid Dreams about the Future

Chris Wild

"I have a dream!" said Martin Luther King in one of the greatest speeches ever delivered. I have taken this invitation to think about the future as an invitation to dream.

I have a dream of a multitude of students spellbound by the broad vistas of the data landscape. I have a dream of their flying on magic carpets that enable them to swoop effortlessly over this landscape exploring its nooks and crannies in search of its hidden treasures. I have a dream of students empowered to look at data, explore analysis systems and educational environments designed so that, like Alice in Wonderland, they keep crying *"Curiouser and curiouser!"* and have the ability and confidence to go where that curiosity leads. I have a dream of educational and analysis environments designed to leverage the power of "I wonder …?" to draw students in to learning more and more – the power of "I wonder why …?", the power of "I wonder what happens if …?", the power of "I wonder what that does?", the power of "I wonder what's around the next bend or just over the horizon?" I have a dream of software that finesses away the mundane, the mind-numbing and the soul-destroying difficulties. I have a dream of software that creates rich, virtual data-generating environments that mimic real real-world environments well enough to enable realistic experiential learning in accelerated time frames. I have a dream. But this is not drug-induced fantasy. We are already building systems that are converging on this utopian dreamscape. Working to fulfil these dreams makes it really worth getting up in the morning.

Complementing the dreams are realizations. We are all well aware that the accelerating onslaught of technology is having profound effects on our everyday and workaday lives. But in statistics education, the most profound effect of technology is its effect on what is actually worth learning and by whom. Ours is a fast-changing world of ever-expanding possibilities, where the limits of what-machines-can-do expand inexorably, leaving us ever freer to concentrate on the thinking that is necessarily human. Increasingly, everything that is purely procedural in statistics will be automated in software (a process that will be accelerated by the large numbers of computer scientists in "data science", their every sinew and fibre commanding, "Automate!"). Real investigators and data analysts won't need to know the details any more than they now know the details of what is happening under the hoods of their own cars. Teaching procedural details, i.e. having students learn to operate particular algorithms (or particular software menus), is teaching short-term, death-dated skills leaving them with little of long-term value for their lives. We need to start operating at much higher and more conceptual levels because only the big-picture concepts, the fundamental principles and questions, are of real enduring value.

We should be educating large numbers of students (I would argue virtually all students) to think with data and have some facility with conducting and critiquing real-world investigations, and much smaller numbers of people who can develop new methodologies and turn them into new tools. For the larger group, we need to get them fast to broad vistas and the big issues to create a sense of possibility and potential for their lives – to open eyes, quicken hearts, liberate the imagination, and empower – then temper this with proper caution. All of this should be facilitated by the best tools available, by software that is part of the solution (liberators, accelerators) not part of the problem (time-sinks, shackles, quicksand). If struggling to get the right stuff into and out of software chews up an appreciable proportion of these students' time then it's the wrong software.

The small group needs the big pictures of destinations fast too – to ground their thinking, but then back-filled by the technical computer and/or mathematical understandings and skills that will enable them to push boundaries and to improve or create new tools with even more powerful capabilities.

To moderate some of the statements above, "short-term" is not entirely a negative. Short-term skills can be crucial for that imminent dissertation or that first job. Additionally, I am not claiming that learning to operate an algorithm (follow instructions) is a not a useful skill. It is a very useful skill, but not for operating the algorithms you have been trained on, but for operating new algorithms in the time before they become shrink-wrapped in software. The same applies to learning how to write computer programs (e.g., for data wrangling). Today's complex programming task is tomorrow's mouse-click. These skills deliver their value when they been mastered to a level where they allow you to do new things that others have not already catered for – bridging the gap between the available and a desired "something more".

Teaching operating procedures or algorithms to foster the ability to operate as-yet-unseen procedures is subtly different from teaching the operation of procedures for their own sake and utility. This is obvious to teachers of programming but much less so to teachers of data analysis. How should it influence the ways in which we should teach? There, I think, lie some very important research questions.

There is very little work being done that consciously focusses on enabling students to experience much more of the data world much more quickly. I distinguish here between "what you can see in data" and "enabling professional skills for data analysis" on which there is more (and very valuable) work. The difference being emphasised is the contrast between providing visions of exciting holiday destinations and providing the ability "to work one's passage" (pay for a sea voyage by working on the ship). It is the former that arouses the desire to invest in the latter.

I think the key to "experience a lot quickly" is software and educational experiences developed in tandem. The software is designed to make possible capabilities that are educationally desirable. The educational experiences exploit the doors opened by the software. Comprehensive systems are much more desirable than sets of one-off applets. While the latter can be fine for illustrating particular points, they do not foster the forming of broad understandings facilitated by unifying frameworks. Additionally, every time someone reaches for a new system there is the time-sink of figuring out how the new system "thinks". Besides wasting precious time, this can trigger a "too-hard" response that inhibits even getting started.

We need to distinguish between software that accelerates the speed with which learners can discover new landscapes and is good for occasional users to dip in and out of, and software that professionals will immerse themselves in. The former will tend to prioritise areas where the inputs and outputs can be easily understood and not worry about comprehensive coverage, whereas the latter must provide almost everything a professional could want. This inevitably adds to complexity in human-software interactions and steeper learning curves. Software to accelerate breath of data-world awareness and empower occasional users, on the other hand, should minimise how many names you have to know before you can get value out of the software and should maximise how far default settings can take you (memories of the what and how of doing things dim surprisingly quickly when not constantly refreshed by regular use).

I have experimented with these ideas in building the iNZight system for data visualisation and analysis (www.stat.auckland.ac.nz/~wild/iNZight/) and VIT (Visual Inference Tools, www.stat.auckland.ac.nz/~wild/VIT/), a visualisation system for developing inferential concepts. iNZight is heavily used in New Zealand schools, particularly the last three years of high school, but has capabilities covering much of undergraduate statistics. The free online MOOC "Data to Insight" (see www.stat.auckland.ac.nz/~wild/d2i/4StatEducators/) was produced to prototype a getting-further-faster introductory-statistics course leveraging the capabilities for acceleration provided by iNZight and VIT. The comments left on the last page of the course were extremely enthusiastic about both the course and the software – this from a student body that ranged from members of a small high-school physics-honours class in upstate-New York to PhD researchers from many areas, and also from journalists, linguists, arts administrators, economists, data managers, marketers and scientists.

At one level, iNZight and VIT are prototypes to be learned from in building better getting-further-faster software for the future. But at another they are also very good systems for teaching and data exploration right here, right now. Although they do not quite usher in the opening dreamscape they come close enough to show that there is huge potential here, that this is something that is almost within our grasp.

There is one element of the dream that has not been touched on – virtual (data-generating) environments for the acceleration of experiential learning. Since the excitement of the projects referenced by Wild (2007), there has (sadly) been very little progress. The one glittering exception is Michael Bulmer's wonderful Island (see Bulmer and Haladyn, 2011) which continues to grow and become more sophisticated and realistic every year.

How can and should research relate to all of this?

The data world is expanding fast but the extent of what we convey has changed very little. If our breadth of view does not start to keep pace with the expansion of the data world, our educational offerings will illuminate an ever-shrinking segment of reality converging to irrelevance, and we will (deservedly) fade into oblivion. Our educational future needs to be very different from our educational present and our educational past. To chart our way into that future we need seers and dreamers, we need entrepreneurs and innovators, we need architects and builders, and we badly need research and researchers.

We need our researchers to take on problems that are bigger and more fundamental, but messier and less well-understood. There, even just crystallising key research questions constitutes making research contributions of fundamental value. I conclude with a list of some research challenges.

Future-facing areas in need of substantial research:

- Where is statistics now? What new fields are opening up?
- Of the potential new areas, which are most worth pursuing and why? What is essential versus what is peripheral and why? What don't we need to teach anymore and why?

- Who are the clients for this educational offering (including future employers)? What do they need? What skillsets are required for that? Where in the programme should we address that?
- What can machines do and what of statistical thinking is essentially human?
- What difficulties can we circumvent with new technology?

Elements to underpin research-informed innovation:

- Generating new ideas about how to do things differently
- Building new enabling technology
- Conducting student and teacher-facing research on student capabilities and how people learn in this newly built environment
- Devising enabling pedagogy that leverages the capabilities of the new technology
- Conducting student and teacher-facing research on effectiveness of implementations of new pedagogy
- Feeding what has been learned back into the technology and pedagogy's development cycles

References

Bulmer, M. and Haladyn, J. K. (2011). Life on an Island: A simulated population student projects in statistics. *Technology Innovations in Statistics Education*, **5**(1). Retrieved from http://escholarship.org/uc/item/2q0740hv

Wild, C.J. Virtual Environments and the Acceleration of Experiential Learning. *International Statistical Review*, **75**, 322-335, 2007.