Improving the Climate for Statistics

Tenth Annual

Vidya Jyothi Professor V K Samaranayake

Memorial Oration

Delivered by:

Roger Stern

Professor of Applied Statistics (Emeritus)
University of Reading

UCSC

University of Colombo School of Computing

06th September 2017

INTRODUCTION

I am very honoured to have been invited to deliver the 10th Vidya Jyothi Professor V. K. Samaranayake memorial oration. I thank the Director and the Academic staff of the U.C.S.C., and the University of Colombo for this. I believe I am the first non-Sri Lankan to have been invited in this way. I understand this is not only the 10th anniversary of this memorial series, but is also being celebrated as the 50th anniversary of computing at the University of Colombo.

Professor Samaranayake (Prof, from now on in this talk), influenced developments in Sri Lanka in many areas. I am a statistician and therefore I use this occasion mainly to discuss his influence in this area. In particular I consider some topics that combine statistics and computing. This influence between our organisations was partly via a link between the University of Colombo and the University of Reading. The formal link was more than 30 years ago, but some of the activities remain important today.

In 1967, 50 years ago, Prof first taught FORTRAN programming to University of Colombo students. This was the same year that I joined the Department of Applied Statistics at the University of Reading. Statistics has been a problem subject to many people, in many countries, for a long time. I describe this problem in Box 1 below.

Box 1: Common problems with statistics teaching

Service course training is often dominated by analysis, with relatively little on data organisation, or on design. Examples are usually relatively small, hence little time is devoted to important descriptive methods.

Sometimes a "recipe-book" approach is used which usually results in topics covered in their order of mathematical complexity, rather than their importance. It also often results in the overuse, and often irrelevant use, of significance tests. This approach provides little understanding of principles.

Training still usually emphasises formulae and even hand calculation. This is claimed to aid understanding, but it often does not. These more theoretical courses are often separated from the use of computers for good statistical practice.

For service courses the presentation is often too mathematical, rather than conceptual. It may be taught by a statistician who has little interest in the student's main subject, or by someone from the subject area, who has not been taught the key concepts earlier.

Most of these weaknesses are similar for the training of statisticians. They often lack practical skills for the analysis of large data sets (not even "big data", but just "medium data" poses problems for them). They also may lack skills in data organisation and of design, particularly of areas such as participatory methods. They may also lack a real understanding of concepts, though they are happy with the formulae.

These weaknesses results in graduates (who are non-statisticians) with a near universal dislike of statistics – and sometimes a fear of statisticians. It also results in a strong demand for in-service courses, because the importance of practical data handling skills has become clear.

And too many statisticians have become relatively inflexible in their advice. They are then feared and ignored by potential clients, who prefer to discuss issues with someone in their own subject area, though they understand far less in statistics.

I have been interested in how to improve the teaching of statistics for the whole of my career. As long ago as 1973 we had a paper on our teaching at Reading titled "The Use of a Computer in the Teaching of Statistics". You may notice in this title that it is "a computer" and not "computers". The idea was that the university as a whole might have a computer. Staff (or students) interact via a card reader or perhaps even a terminal. The idea

that a person might have their own computer was unthinkable.

The year 1973 was also when I arrived in Colombo. My job was to help the teaching of statistics to become more applied and also more relevant to the needs of the country. There were 8 computers in Sri Lanka at this time. The nearest to the University was at the Department of Census and Statistics. The University had card punches and part of my job was to cycle to the Department of Census to deliver cards and collect the printout. This was still an era when slide rules and tables of mathematical functions were used routinely. It was only 2 year previously that our department in Reading had its first electronic calculator. This was the size of a small desktop today. I remember being amazed that not only did it have 2 memories, but it also had a special square-root button!

Professor Samaranayake came back to the University in the following year. I worked under him on the teaching of statistics at undergraduate level. I was followed, 2-years later, by Ian Wilson, also from Reading, who started the MSc programme. On the Colombo side a key person, in these early years, was Savitri Abeyasekera. Box 2 describes some memories from Ian Wilson.

Box 2: Ian Wilson on Professor Samaranayake

Professor Samaranayake was my boss during the two years I spent in the Statistical Unit of the Mathematics Department at Colombo University (1975–7), but was always Sam to me – my sometimes frustrating, but always dear and admirable friend.

Sam met me at the airport when I arrived to join his Department. He immediately got me to draft the newspaper advertisement for the Diploma (later M.Sc.) I was to head. I asked him for the relevant University documents so I'd know what had been agreed, and got my first insight into Sam's ways of working. He had yet to tell not told the authorities yet, because that would waste too much time. There would be hold-ups –

and questions whether such an undertaking should be allowed at all! If they were told we had already recruited a good number of students, who would pay a considerable amount in fees, it would be much easier to accept than to undo the proposal: it worked! A few years later Roger Stern, a couple of other colleagues and I did exactly the same to Reading University when we set up unconventional summer courses: it worked again, so thank you Sam.

Sam was never afraid of being criticised. He had no time for rumour-mongering or back-biting. He was never idle, or lethargic, and didn't put off doing what needed to be done. His enormous energy and force of personality were combined with very astute understanding of the personalities, qualities and potential of those around him. He was enormously generous-hearted to others and did all manner of favours for huge numbers of those around him, so on occasion he could call in a favour with every reason to be sure he would be obliged. I'm sure that's how he knew Colombo University would let our Diploma course go ahead.

The time I have spent with Sam (and his lovely wife Sriya) was special, and remains one of the great gifts life has given me.

The link then started and continued for about 10 years. Reading University had two leaders in statistics, namely Professors Robert Curnow and Roger Mead. Robert writes as follows, in Box 3:

Box 3: The Reading Colombo link

The Department of Applied Statistics at the University of Reading has been involved for over fifty years with teaching and research in developing countries .The longest and most successful link was with the Statistics Unit within the Department of Mathematics at the University of Colombo. The establishment and the success of the link owed much to the enthusiasm and support of Professor Samaranayake and the

continuing involvement of Dr Savitri Abeyasekera.

The link involved the exchange of staff and research students. The aim was to develop the teaching of statistics in Colombo at undergraduate and postgraduate level and to provide statistical assistance to agricultural research institutes in Sri Lanka. Two year visits from the U.K. by Roger Stern and Ian Wilson were followed by shorter visits by many members of the Reading Department. All of us experienced the kindness and hospitality of our hosts in Colombo and gained both professionally and personally from these visits. I do hope the link will now continue although obviously in a more informal way.

Robert Curnow, Emeritus Professor
Department of Mathematics and Statistics,
University of Reading

THE LINK

I select initiatives from the link that still endure. More details are in the 1984 Biometrics article "Stimulating statistical progress In a developing country by a link between universities", by Ian Wilson et al.

Statistical Games

The influential GAISE (Guidelines for Assessment and Instruction in Statistics Education) college report in 2016 strongly recommended that (introductory) statistics course should "integrate real data with a context and purpose". They also notes that the "rapid increase in available data has made the field of statistics more salient".

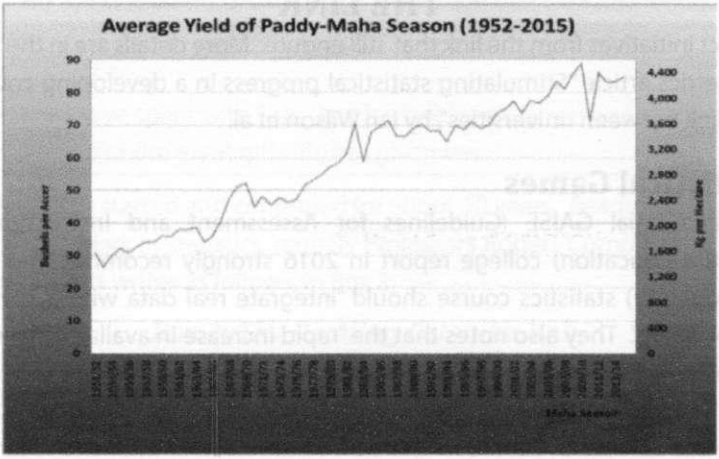
There are at least three challenges to be addressed in statistics teaching. The first is to include the whole process, i.e. design and data collection as well as data analysis. And this usually has to be done with large classes within the limited time allocated for lectures and practical work.

The second challenge is that most teaching of statistics is method-based. We teach a particular method, such as regression analysis, and then all examples simply illustrate this method. The real world is "the other way round". It is problem based, and part of the work of a statistician involves finding the appropriate methods.

The third is the need to have examples where there is no single right answer. Too often students are taught in ways that give the impression that the lecturer knows the answer to each problem. Once this information is "poured into the student" they can graduate and possibly then become lecturers themselves. Unfortunately the real world is not like that. Very few problems in statistics have a single right method, or right answer.

Fig. 1 shows the yield of paddy from Sri Lanka's crop cutting survey that has been conducted regularly since 1952.

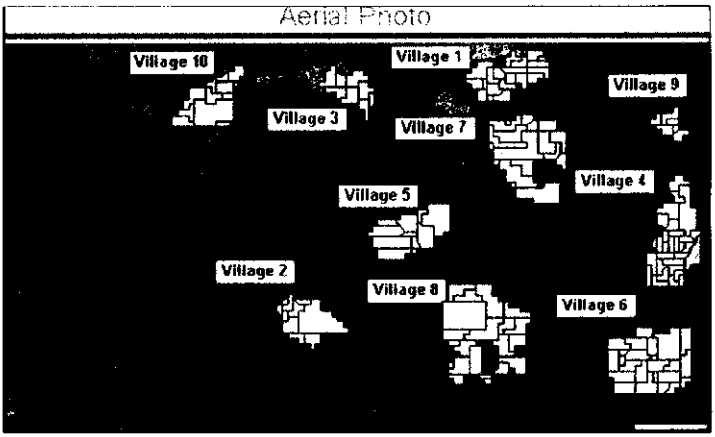
Fig. 1: Results from Sri Lanak's regular crop-cutting survey



Note: Average yield of Paddy 2010/11 Maha season should be updated as 3699 Kg/Hec.

Part of solution, both in Reading and Colombo, was for some courses to be "problem based". Some were real problems, often based on our consultancy work. Others were simulations, and we called them "statistical games". One of our most successful "games" was based on Sri Lanka's crop cutting survey. It was first produced in 1978 during a visit to Reading by Anila Wijesinha (now Dias Bandaranaike).

Fig. 2: The villages in the rice survey game



The initial task, for the students, was to choose a suitable sampling scheme from a subset of the 10 villages, shown in Fig. 2. Each field corresponded to an envelop. They then interviewed the farmer, i.e. recorded data from outside the envelop, on fertiliser application and variety used. They then took a random slip of paper from within the envelop to simulate the yield from the crop-cut.

The students then have to design a form for the data entry into a spreadsheet, or statistics package. They then analyse the data and write a short report.

Fig. 3 shows an example of data from a sample of 4 villages. The first column in Fig. 3 is the village name, which may not be immediately recognisable as typical Sri Lankan names. It does reflect four of the Colombo staff involved

in the link, namely SABEY = Savitri Abayasekera, KESEN = Kevin Seneviratne, NIKO = Prof. N D Kodikara and NANDA = Nandasara!

In the 1990s we decided, in Reading, to fully computerise this game. However we still found the envelop version to be popular. More recently, the software used in the 1990s is no longer available, but the hand version remains a valuable and much appreciated training resource. Its most recent use was to 54 MSc students in 2016. They found it both useful and very different from their other exercises.

Fig. 3 Example of data from the simulated rice survey

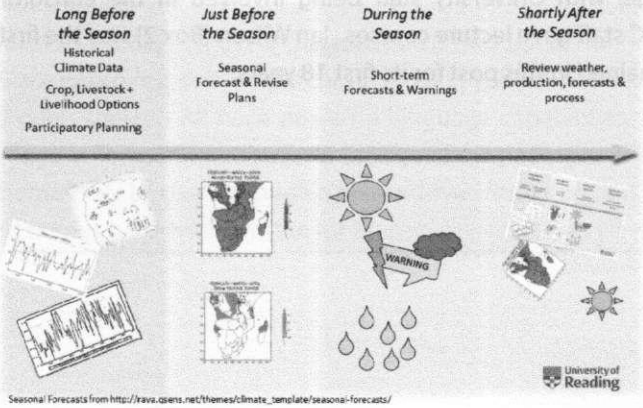
	Village (f)	Field	Size	Fert	Variety (f)	Yield	Fertgrp (f
1	SABEY	3	2.0	2.5	OLD	53.6	>2cwt
2	SABEY	13	5.0	1.5	OLD	44.6	.5-2cwt
3	SABEY	4	5.0	3.0	OLD	50.7	>2cwt
4	SABEY	20	1.5	1.0	TRAD	33.6	.5-2cwt
5	SABEY	19	5.0	2.5	NEW	62.1	>2cwt
6	SABEY	10	4.0	1.5	TRAD	30.6	.5-2cwt
7	SABEY	8	4.5	2.0	OLD	37.7	>2cwt
8	SABEY	11	3.5	0.0	TRAD	24.3	0cwt
9	SABEY	14	5.0	2.0	NEW	56.8	>2cwt
10	SABEY	12	4.5	2.5	OLD	59.3	>2cwt
11	KESEN	9	1.5	2.0	OLD	40.4	>2cwt
12	KESEN	8	7.0	0.0	TRAD	25.8	0cwt
13	KESEN	1	2.5	1.5	OLD	40.7	.5-2cwt
14	KESEN	4	8.0	0.0	TRAD	27.6	0cwt
15	KESEN	6	4.5	2.5	OLD	48.7	>2cwt
16	KESEN	3	2.0	0.5	TRAD	27.0	.5-2cwt
17	KESEN	5	4.0	0.0	TRAD	19.1	0cwt
18	NIKO	2	4.5	0.0	TRAD	26.3	0cwt
19	NIKO	8	2.5	0.0	TRAD	24.7	0cwt
20	NIKO	7	6.0	0.5	OLD	40.4	.5-2cwt
21	NIKO	3	6.0	0.0	OLD	31.8	0cwt
22	NIKO	4	8.0	0.0	TRAD	29.6	0cwt
23	NANDA	4	4.5	2.0	TRAD	36.6	>2cwt
24	NANDA	2	3.5	2.5	OLD	57.4	>2cwt
25	NANDA	13	20.0	3.0	TRAD	42.7	>2cwt

Climatic Data

In the early 1980s Reading had a team funded to investigate new methods of analysing climatic data. This came from work in Africa, but the ideas were equally applicable in Sri Lanka. This was long before the current interest in climate change. Various publications by staff from Colombo showed how rainfall data could be modelled, e.g. "The analysis of daily rainfall data for agricultural purposes", by Savitri Aberasekera et al, and this work now also involved staff from the Sri Lanka Department of Meteorology.

This area of work continues in Reading in ways that may again be applicable in Sri Lanka. In particular there is an initiative called PICSA (Participatory Integrated Climate Services for Agriculture) that has become popular in many countries in Africa and in the Caribbean. Part of the novelty in PICSA is the statistical component, i.e. the extensive use that is made of the historical climatic data, shown in the first panel in Fig. 4, below. Many other initiatives that provide climate information for farmers, concentrate primarily on the short and the seasonal forecasts. These components are also considered within PICSA. However initial plans are made by farmers, partly through their own analyses of the historical records.

Fig. 4 Participatory Integrated Climate Services for Agriculture (PICSA)



Seasonal Forecasts from http://rava.qsens.net/themes/climate_template/seasonal-forecasts/

Consulting Services

I remain of the view that it is difficult to teach applied statistics unless the staff are involved in genuine applications themselves. Reading University has always had a free statistical consulting service for staff and postgraduate students, which the University pays for, through a staff allocation.

In Colombo University a similar service was established which also offered its services outside as paid consultancy work. For example, a report in 1981 detailed work for Lever Brothers Ltd, the Industrial Development board, Ministry of Plan Implementation, Coconut Research Institute, Lady Ridgeway Hospital, Marga Institure, UNICEF, University Grants Commission, Instiute of Agriculture, Faculty of Medicine, State Cashew Plantation and Jaffna Training College, as well as support for students and staff from the Department of Zoology, Geography, Sociology, Meteorology and Chemistry.

In Reading the Statistical Services Centre (SSC) was established in 1983. This is a 100% self-financing centre that provided special training courses for international students, plus paid consultancy services, outside the University for clients, both in the UK, and internationally. There was also an exchange, with University staff being involved in the consulting work, whileSSC staff gave lecture courses. Ian Wilson (Box 2) was the first director and remained in this post for its first 18 years.

Statistical Software

This is not just the 50th anniversary of computing in Sri Lanka. Statistical packages, unlike most other software, has a similarly long history, see Box 4.

Box 4: History of Some Statistical Packages

Statistical software has a long history. SPSS is perhaps the most used package now and it was first released in 1968. The other giant of statistical packages, SAS, was first developed in 1966, while Genstat was released in 1971 and Minitab in 1972. This is a long way back in the history of computers that only really started in the 1940s.

These software systems have each added components as new methods appeared. However they have not always been quick to abandon methods developed earlier. They sometimes show their age, with components still available that are no longer as useful as 40 years ago. This provides users with a rich, but also sometimes confusing mix of methods for their data analysis.
Microcomputers appeared in the 1980s, together with many new statistics packages. By the 1990s, once the microcomputers were large enough, the original systems dominated again, joined by a few newcomers, particularly Stata and R.

For universities, access to most of this software is through annual licenses, except Stata which is a one-off purchase for each version and R which is free and open source. All have powerful language capabilities, though the language in some of them reflects the age of the package. All, except R, can alternatively be used through a menu-driven front end and various front-end menus have been written for R.

Until the 1980s most of these packages were inaccessible to students, staff or researchers in most developing countries. Indeed the Rothamsted research institute, in the UK, offered a service whereby agricultural researchers could send their data to the UK and would eventually be sent the results.

From 1982 Reading offered an annual short course called MASD (Management and Analysis of Statistical Data) which was designed for researchers in different developing countries. Prof argued that this should be given in the developing world instead, and that Colombo was the obvious place. This led to a 6 week regional course from December 1984, designed for agricultural statisticians.

One key issue was the statistical software to be used on the course. The University of Colombo had access to microcomputers, but they was not powerful enough for any of the software described in Box 4. Through the link, teams from Reading and Colombo worked on a new small package, designed for the recent BBC Micro-computer. These had been in Colombo University since the early 1980s, and had, most notably, been used for the 1982 Presidential Election, Box 5.

Box 5: The 1982 Presidential Election

Professor Stanlaey Wijesundera
Vice Chancellor,
Colombo University.

Dear Professor,

Presidential Election – 1982: Declaration of Result

At the Presidential Election just concluded, I had the priviledge of obtaining the assistance of the computer services from the Mathematics Department of the Colombo University.

I appreciate the initiative taken by maranayake of the Mathematics Department to offer the services of their computers to work out the results. The availability of this service made my task smooth, reliable and speedy.

Professor Samaranayake and his staff spent well over 30 hours of continuous monitoring of results apart from the time spent on preliminary srrangements at the Town Hall. I am sure you would have seen for yourself the live-telecast of results and the value of their contribution at the Centre for the Declaration of Results.

I am most grateful to Professor Samaranayake and his assistants whose wiling cooperation at a crucial stage of this election made my task very pleasant and easy. Please convey my sincere thanks to all of them. I look forward to similar cooperation in the future too.

Yours sincerely

(R. K. Chandrananda de Silva)
Commissioner of Elections

Reading and Colombo staff worked together on a suitable package for this course. It was later called Instat (for INteractive STATistics). The course went well.

Instat was then further developed and this included a special menu for the analysis of climatic data. It later was adapted for Windows, and remains in use, particularly for climatic analyses, though it is now many years past its "sell-by" date. We discuss, below, what might be used instead.

WHAT NOW?

This is the 10th memorial lecture. Hence I now consider what has been happening more recently. I wonder whether some of the topics below might have been even more prominent in Sri Lanka had Prof still been leading in his usual inspirational way.

Computer-based math

The TED talk by Conrad Wolfram
(https://www.ted.com/talks/conrad_wolfram_teaching_kids_real_math_with_computers) is now 7 years old, Fig. 5. He urges us, very persuasively, to stop just teaching calculating and calling it mathematics (including statistics). We should use computers more and adopt a more problem-solving approach. This applies both in schools and universities. Of course it requires access to technology, but this is now available almost everywhere.

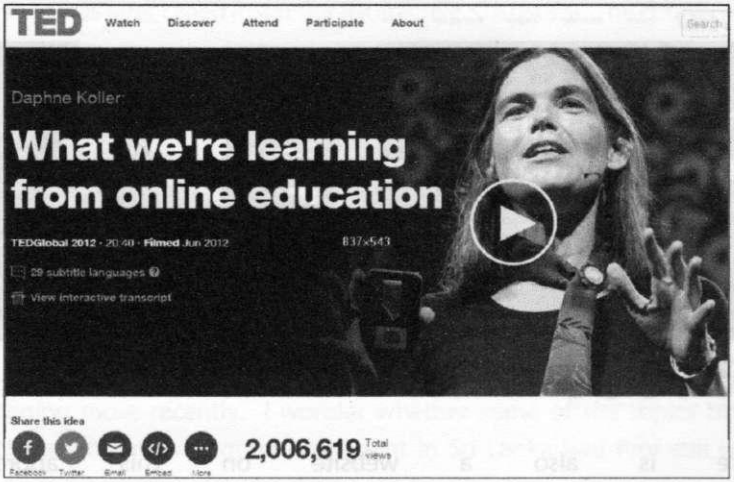
Fig. 5: Conrad Wolfram TED talk 2010

There is also a website on this approach (https://www.computerbasedmath.org/) that boldly proclaims "Let's Fix Maths Education". But there is, so far, only one real case study and that is largely to apply the ideas to statistics teaching in Estonia. This is perhaps an important area, where the broad ideas are sound but "the devil is in the detail".

MOOCs

MOOCs are Massive Open Online Courses. Many people from all over the world have taken Coursera online courses, and the TED talk illustrated in Fig. 6 makes the case very persuasively.

Fig. 6 Daphne Koller TED talk 2012



But is this a case to take MOOCS developed in the US or is it for each university to compete to produce their own courses? Reading University has been developing courses that are run through the FutureLearn consortium. The case at Reading seemed to be largely that it was better to be in, than out.

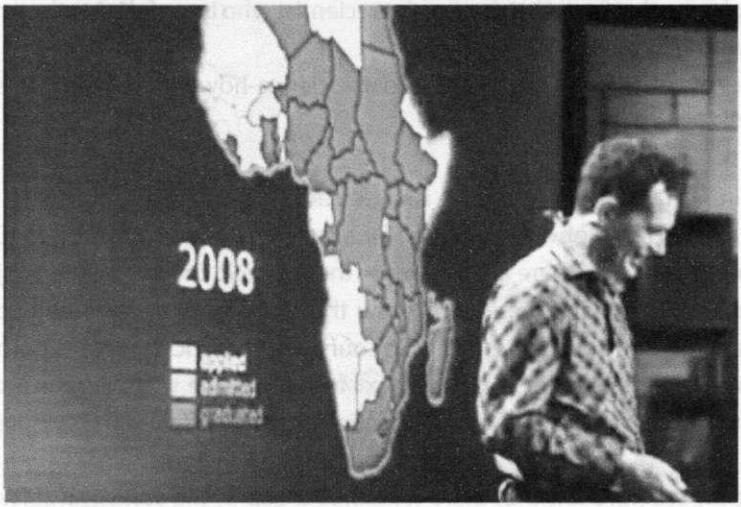
Should Colombo University be developing its own MOOCs? -perhaps it already does so? However, for all the hype, there are now some serious questions even on whether MOOCs are now dead!,

e.g. http://moreorlessbunk.net/technology/moocs/moocs-a-postmortem/

Mathematical Sciences

In 2008 there was the first of these memorial talks. It was also the year that Neil Turok won the TED prize for his innovative MSc teaching of Mathematical Sciences to African students, Fig. 7.

Fig. 7 Neil Turok TED talk 2008



I do not know whether he met Prof, but his broad interests in mathematics together with his professional background in physics mirrors Prof and I am sure they would have got on well together. I wonder what would have happened. In particular what of the style of the MSc teaching – now in 6 regional centres, and producing about 300 graduates per year, and also the other activities – could have been exported to Sri Lanka? This has been an exciting innovation, see https://www.nexteinstein.org/

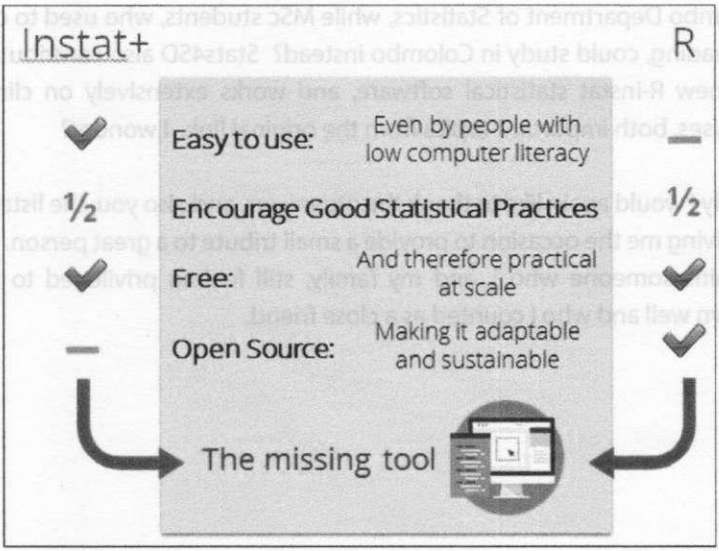
Statistical Software and Data Science

I was delighted to see that the Colombo Statistics Department has a new Centre for Data Science. Hadley Wickham, the author of many R packages including the excellent ggplot2, based on the grammar of graphics was asked recently of the difference between a statistician and a data scientist. His (cynical) answer was that "a data scientist is a statistician who is useful". A reply was that "a statistician is a data scientist who is useful". A rather more thoughtful discussion is here:
http://bulletin.imstat.org/2014/09/data-science-how-is-it-different-to-statistics%E2%80%89/

The importance of data science is undeniable. The statistical software, R, is well known as an obvious tool; see for example the book called "R for Data Science". In the UK there are now more than 50 MSc courses on the subject. What is of immediate interest is how this data science revolution could impact the undergraduate and MSc courses in statistics? What is perhaps particularly interesting, in our context today, are the ways in which data science integrates parts of computer science more closely with statistics.

Recently we have felt that there remained a gap in the statistical software packages available. This was at least in Africa, but the resulting software could perhaps be useful everywhere. We made the case in our first crowd-sourcing campaign in 2015, Fig. 8,
(https://chuffed.org/project/africandatainitiative). The case was for software that built on the Instat experience, that had started through the Colombo-Reading link. However this time it would also build on the power and inspiration of the free and open-source R statistical software. The resulting package is called R-Instat. Like the original Instat of 30 years ago it is designed to support improved teaching of statistics and also to replace the old Instat for the analysis of historical climatic data.

Fig. 8: The basis for the new software



IN CONCLUSION

In box 3 above, Professor Curnow queries whether the link between Colombo and Reading could continue? If so, it would be very different to the previous link. I have been delighted to hear of the continued strengths of the Statistics Department in the University of Colombo. Their website still mentions our Department of Applied Statistics at Reading. However, statistics at Reading is now, sadly, much small than before. The well-known MSc in Biometry has ceased, as has the single subject BSc degree in statistics. In addition, after 34 years, the Statistical Services Centre (SSC) will be closed from the end of 2017. This is, however, not the end of Applied Statistics at Reading. The staff of SSC have moved to form a small company called Statistics for Sustainable Development (Stats4SD, https://stats4sd.org/) with offices in the middle of Reading. Staff are also visiting scientists at the University of Southampton, one of the strongest universities in the UK for statistics. Stats4SD provides consultancy services in statistics worldwide on a not-for-profit basis.

Perhaps this group could again link with the consultancy work of the Colombo Department of Statistics, while MSc students, who used to come to Reading, could study in Colombo instead? Stats4SD also contributes to the new R-Instat statistical software, and works extensively on climatic analyses, both important topics from the original link. I wonder?

Finally I would again like to thank the organisers, and also you, the listeners, for giving me the occasion to provide a small tribute to a great person. Sam remains someone who I, and my family, still feel so privileged to have known well and who I counted as a close friend.

University of Colombo School of Computing,
No. 35, Reid Avenue,
Colombo 00700,
Sri Lanka.

Tel: +94 -11- 2581245/ 7
Fax: +94-11-2587239

Email: info@ucsc.cmb.ac.lk
Web: http: //ucsc.cmb.ac.lk