TEACHING STATISTICAL LITERACY BY GETTING STUDENTS TO USE REAL WORLD DATA: 40 YEARS WORTH OF EXPERIENCE IN 40 MINUTES

Jackie Carter* and James Nicholson**
***University of Manchester, UK JackieCarter@manchester.ac.uk**
****SMART Centre, University of Durham, UK j.r.nicholson@durham.ac.uk**

The data deluge over the past twenty years has resulted in an explosion in volumes of available data. Access to data is increasingly easy; technology advances have resulted in increasingly sophisticated ways to represent and analyse these data. Citizens are confronted with statistics and numbers in a multitude of ways, so the imperative for improving statistical literacy is strong if we want a well-informed and data-literate population. Social sciences are embracing quantitative methods as demand grows, in the private and public sectors, for evidence-informed policy and a greater sophistication in approaching difficult to measure constructs, such as global sustainability, is emerging. The Sustainable Development Goals set out by the UN[1] in September 2015, and the data requirements associated with them, may accelerate all of these trends. This paper will reflect on the authors' experiences of working with real data in the context of schools, and university social science courses, over the past twenty years, and consider how this could inform discussions in developing statistical education.

INTRODUCTION

Teachers and lecturers are extremely busy – if we want to encourage a greater emphasis on embedding real data in substantive course content in the social sciences then we need to both convince the teachers and lecturers of the value of doing so, and make it much easier for them to assimilate it naturally into their courses. We also know that students who thought they had left maths behind at secondary school are loathe to pick up quantitative methods at university. The challenge to teachers is therefore to motivate learners, and if they are not maths or statistics teachers to minimize the effort required to incorporate numbers into their lessons.

To some extent the problem is in which part of the curriculum statistics is taught. Should it be in the maths or stats class, or embedded throughout other subjects? There are significant efforts underway, at least in the UK, to develop a 'statistically-literate population' for those who dropped maths at the age of 16, especially in the social sciences where quantitative skills are regarded to be in short supply (see Core Maths in England –see http://www.core-maths.org/)

Very often in the social sciences theory and aspects of the data landscape are introduced together – with data visualizations offering the opportunity for students to grasp the big picture – the "stories in the data" of the observations, before embarking on learning disciplinary theories which seek to explain the observed behaviour, or indeed working with multivariate (MV) analytical techniques. We aim here to draw on the authors' collective experience of teaching and supporting statistics and quantitative methods teaching at school and university level, to provide examples of how this learning can be applied in the workplace, and to contribute to the debate on what it means to develop statistically literate citizens. Attention will be given to the learning of statistical techniques and concepts, together with the application of that learning through practice especially in social science courses.

BACKGROUND AND MOTIVATION

This paper recounts two individual journeys, from two different perspectives. Nonetheless, the experiences covered here converge in ways that inform the development of statistical literacy education.

Both authors started their professional careers teaching mathematics in secondary schools (for ages 11 – 18). The first author (JC) has spent the last 20 years working at university level though the motivation for working with real data and applied statistics originated in experiences gained from teaching at school level. The second author (JN) worked in schools for a quarter century and for the past 12 years has been developing resources, especially data visualisations, to

[1] http://www.un.org/sustainabledevelopment/sustainable-development-goals/

In: J. Engel (Ed.), *Promoting understanding of statistics about society.*
Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE),
July 2016, Berlin, Germany. iase-web.org/Conference_Proceedings.php

allow secondary level students to engage with real data – a theme which had already been a focus for him for some 8 years prior to leaving teaching. We therefore start by each outlining our routes into statistical education, and proceed to provide a narrative of our personal journeys in the last two decades. The intention is to draw out common themes, and challenges, in teaching students to think statistically.

JC drifted into maths teaching, after completing a degree in mathematics with its applications (she actually wanted to study social psychology but, having completed A-levels in science, had forgotten how to write good essays, and so stayed within her comfort zone). Statistics were her favourite classes at university, not least because they were so well taught, highly visual and required interpretation of the results in a way that made sense and mattered to social issues of interest to her. As a young secondary school (11-18) maths teacher, she was asked once to teach probability and statistics to a lower ability class, a group of 13-14 year olds. She was fresh out of university, determined to capture their attention and prove she could rise to the challenge. She spent significant time collecting games, puzzles, spinners, coloured buttons and dice and designed multiple hands-on exercises to keep them busy over the course of the lessons. Barely any teaching was done at the blackboard (this was 1985). The class's response was very positive and the approach captured their attention and enthusiasm for practical learning – but the examples chosen were too 'text book' in style, and not sufficiently in tune with their daily lives. They weren't captivated by the statistical concepts of chance or probability demonstrated by counters or die falling on a particular value or colour or side-up, as a result of repeated throws or spins. She felt that she had missed a trick. Whilst they did enjoy the practical approach, they needed a context that was relevant to their experiences. The applied learning worked to a degree, and engaged their attention, but context was missing. It wasn't until years later that she reflected fully on this and thought how she could have used examples that were much more in tune with where they lived, and the daily challenges they faced (a former steel-producing town in South Wales), or the family circumstances from which they came (e.g., numbers of siblings). This 'adding context' however was to return in her next career.

In a parallel world, as statistics became a more central component of mathematics education in the UK in the 1990s JN began to notice how uncomfortable many teachers felt in teaching material that they had not studied in depth themselves at university level, and for which the teaching and assessment resources in many cases used data which the author or examiner had simply made up. The focus was primarily on procedural competence with little reward in high-stakes assessments for interpretation or modelling (Nicholson, Ridgway, & McCusker, 2006); textbooks used in schools often contained answers to every (part) question requiring a numerical answer, and to none of the (part) questions which asked for an interpretation; moreover, especially at the younger end of secondary education there would be whole exercises which contained no question to which anyone in the world would want to know the answer. It seemed to have the effect of saying to teachers and to students – *'the mathematics' is all that is important*. Teachers did not feel comfortable that they understood what was going on, something that JN felt had a lot to do with being asked to work with data someone had dreamt up to have particular features.

Hence, both authors have backgrounds that included time in the classroom. Both recognized the need to 'make the numbers matter' to the learner. This paper proceeds to explore the contributions of each author to the teaching of statistical literacy, and unpacks some of the challenges, changes, and opportunities afforded through access to large real-world data resources to support statistical literacy.

JACKIE CARTER'S JOURNEY

Teaching and Applied Learning: Exploratory Data Analysis

I started my career as a secondary school teacher, having graduated in Mathematics with its Applications, taking all the statistics options in my final year. My intention had been to work in the Office for National Statistics upon graduating, but instead I spent four years teaching at high school, before returning to higher education to undertake a Master's then a PhD in Computing. My PhD investigated the application of geostatistics to environmental radiation data, with the research contributing to the development of an emergency response system, using a Geographical

Information System (GIS) with multiple sources of data – geographic, socio-demographic and scientific. In a post-Chernobyl world, the research was driven by real-world needs; to predict radiation deposited on the ground following nuclear fallout from airborne particles, in order to enable movement of people and livestock from contaminated areas (see Carter, McLaren and Higgins, 1997)

In the late-1990s, I was appointed to a new role as 'data visualisation officer' at the University of Manchester. A software tool, called CDV or the Cartographic Data Visualiser, had been developed by a social geographer (Jason Dykes) and my remit was to promote this in universities as a means of visually exploring the 1991 UK Census of Population data. Working together we developed 'canned applications' around variables of interest to social science students. For example, we chose car ownership (as a proxy for income), long-term limiting illness (self-reported ill health) and employment rates (economic activity), to enable exploration of differences in social characteristics collected through the Census statistics at various geographic levels. Our express aim was to bring the data to life to those not trained in using GIS software.

Not all the websites and resources we developed persist (although the web archive retains some pages, see Web Archive, 2016), but we were successful in raising the profile of CDV as a data visualisation too across a number of institutions. We presented several papers at international conferences (e.g. Carter & Dykes, 1998) and connected with colleagues in Germany who were developing an online data visualization system of their own, called Descartes, which later we deployed in delivering census aggregate statistics and international time series data to students (see Andrienko et al., 1999; Andrienko *et al.*, 2000; Cole & Carter, 2001).

Back in the 1990s online mapping was just getting started. The 'Technology Trigger' (Gartner, 2015) supported browser-based visualisations, but the barriers to interactive online data tools were that browsers required plugins to be enabled, and teaching using these tools required computer labs to be available, as laptops were scarce. Multivariate data analysis tools like CDV and Descartes were ahead of their time, and uptake of their use was slower than we might have liked. Nonetheless, students loved these interactive tools, the ability to produce multiple, linked graphics and the exploratory data analysis supported by them. They provided a very quick route into visualizing multivariate complex data, spatially, and a starting point for analysis based on real world statistical data. They enabled concepts like the ecological fallacy (how not to draw individual level inferences from aggregate level data) to be demonstrated and explored, and showed how the way we bin or classify data can affect the conclusions we draw (Dykes, 2003).

I moved to working with a newly formed national service, as Digimap support officer, helping students use the then still very new Digimap Service (digimap.edina.ac.uk). We designed a course on mapping Census aggregate statistics in GIS, which we delivered UK-wide. Essentially this taught how to combine aggregate multivariate Census statistics with digitized boundary data, to produce choropleth maps. The barrier to social science undergraduates in combining data sources back in the early 2000s was high; data were held in difference services (e.g., the digitized boundary data in one service, the aggregate data in another) requiring data discovery as well as skills to combine data to produce a map, or in other words the need for data management skills, prior to analysis even beginning. Today the Digimap service (https://census.edina.ac.uk/thematic/map) provides an online interface to show how we taught through manual processes, demonstrating well how technology and online interfaces have developed.

Developing and sharing teaching resources

Services developed around data resources focused initially predominantly on researchers' rather than undergraduates' needs. Between 1999 and 2003 I led a collaborative project across four UK universities to increase the use of digital census data collections in teaching (see Carter, 2002; 2003). It was at this time that Mark Brown (University of Manchester social sciences lecturer) and I first worked together, and I learned how survey and census data were used in undergraduate courses. This project was hugely ambitious in its scope but with the single aim of supporting better use of population census data in teaching.

There was little evidence of sharing of practice in teaching communities. When I moved to my next role with the UK Data Service (then called the Economic and Social Data Service) we

started to collect case studies in teaching with real world data (see https://www.ukdataservice.ac.uk/use-data/data-in-use/case-studies), to share knowledge and information about how the socioeconomic data, provided through the service, were used in undergraduate and postgraduate classrooms. Concurrently, a teaching fellowship with the UK Open University enabled me to gather examples from social science educators to illustrate how socioeconomic data were used in teaching undergraduates to engage with the substantive issues they come to university to study (Carter, 2012). A key finding from this work was the huge reluctance of university lecturers to share their teaching resources (e.g see 'Learning to share, and sharing to learn' case study at http://www.jorum.ac.uk/wp-content/uploads/2015/01/Learning-to-share-and-Sharing-to-Learn.pdf). Even today those repositories set up for sharing teaching resources could probably be far more highly populated with teaching content than they are; e.g. ESRC's ReStore (http://www.restore.ac.uk/index.php), NCRM's Quantitative Methods Initiative website (http://www.quantitativemethods.ac.uk) and in the US, ICPSR's Teaching with Data Guides (https://www.icpsr.umich.edu/icpsrweb/instructors/biblio/resources).

Teaching social scientists applied data analysis

In the UK, although social science graduates are highly employable (Campaign for Social Science, 2013 at https://campaignforsocialscience.org.uk/wp-content/uploads/2013/10/Graduate-report-2013.pdf), there has been mounting concern over the lack of acquisition of quantitative skills in the social sciences at university level, captured through a series of high profile reports from government, Research Councils and employers (British Academy, 2015; ESRC, 1987; HEFCE, 2005; MacInnes, 2009; Mills et al., 2006). In UK education, it has been possible until very recently to drop the study of mathematics at sixteen; as a result we have students entering university without having studied maths for (at least) two years. Initiatives are ongoing to address this (e.g., Core Maths in England –see http://www.core-maths.org/ for details of the rationale for the new qualifications and links to resources), but students confronted with numbers at university could well have thought that they had left this aspect of their education behind them, and if a recent survey is to be believed this is a cause for concern with increasing volumes of statistical data (http://www.theguardian.com/education/2016/mar/07/a-fifth-of-uk-adults-have-forgotten-how-to-do-fractions-or-percentages-mathematics-english-science).

The low levels of mathematical skills in the UK labour market has been a driver for change. A 2015 British Academy report, 'Count Us In: Quantitative Skills for a New Generation[2]', set out a vision of 'A generation of citizens, consumers, students and workers as comfortable with numbers as they are with words' (British Academy, 2015). Prior to this, in 2013, three UK funding bodies joined forces to lead a national programme of activities, known as Q-Step (Quantitative Step Change) the aim of which was to encourage universities to train undergraduate social science and humanities students in quantitative data skills. Fifteen centres, and three affiliate centres, hosted in UK universities were funded for five years, with a commitment to embed the step-change in their institutions for a further five years (with the expectation that this would then continue). The centres were encouraged to introduce innovative, experimental approaches, to make numbers a normal part of the social science curriculum and turn the tide on 'teaching quants.'

The Q-Step Centre at Manchester[3], led by Mark Brown and myself, is enabling us to build on work undertaken since the 1999 Census learning and teaching project together. Manchester was in a strong position to develop a quantitative training centre for undergraduates, having a strong track record in teaching quantitative research methods at postgraduate level, through the research institute, CMIST (Cathie Marsh Institute for Social Research). Manchester's reputation for strong quantitative social science provided a firm foundation on which to build. Moreover, through a series of teaching grants Manchester has developed a new model of training for undergraduates that focuses on student engagement through the use of exciting real world datasets and a problem-based approach to learning statistics and data analysis (Buckley *et al.,* 2015; Carter *et al.,* 2014; Brown, 2013; Carter *et al.,* 2011). This enquiry-led, data-driven approach is being further developed at Manchester, and is the subject of a paper in the IASE Roundtable (Brown, 2016). At

[2] http://www.britac.ac.uk/policy/count_us_in_report.cfm
[3] http://www.humanities.manchester.ac.uk/q-step/

Manchester we have used the Q-Step grant to bring in lecturers in Sociology, Politics & International Relations, Linguistics, and Criminology, to deliver more quantitative courses in the curriculum, as well as introduced new 'with quantitative methods' pathways in our BA in Social Sciences degree programme.

Building on this strong teaching background, Manchester Q-Step is now, in addition, focusing strategically on applied data analysis, through the introduction of paid summer internships to enable students to undertake social research based on quantitative real-world data. Successfully placed interns are given real-world experiences to put their understanding of statistics taught in the classroom and computer lab, into practice in a work environment. With strong drivers for universities to demonstrate the acquisition of employability skills, the opportunity to undertake an applied, paid internship as part of a degree course is attractive to both current and prospective undergraduates. The Manchester Q-Step internships, designed around real world social research projects, provides an excellent opportunity for undergraduates to stretch their learning, and differentiate themselves from their peers. Interviews undertaken with students after completing their internships are starting to uncover a rich picture of the skills they develop, showing how basic statistical knowledge is often all that is required to enable a student to undertake interesting and motivating quantitative social research (Carter *et al*, 2016). To date we have placed a hundred students in some fifty organisations, including The World Bank, local and national Government departments, charities, media organisations, research institutes and consultancies, polling organisations and market research companies. As an indication of what can result from these experiences, one successful former intern is now delivering social research outputs in relation to the refugee crisis into the hands of Kofi Annan and senior advisers to Barack Obama. Other examples will be presented at the Roundtable event.

Reflections on this journey

The practice of statistics and data analysis and the *substantive context* in which these are taught underpins all of the experiences described in my 20+-year journey. A secondary driver for this work has been a deep belief in the need for exploratory (and sometimes spatial) data analysis as a starting point; indeed I would say this characterizes my interest in, and provides a first step into, the fascinating world of statistics. Students are curious about the social world. They are confronted with theories in their studies. Introducing empirical data in their substantive classes enables them to test those theories, challenge perceived wisdom, or headlines in the news, or anecdotal evidence in their peer groups. Our work (Mark Brown (2016), Carter *et al.* (2016), Buckley *et al*. (2015)) and others (e.g. Adeney and Carey, 2011) shows that students are more likely to engage with numbers and statistics if they are introduced in a way that resonates with their studies.

We are extending this approach by developing innovative opportunities through the Q-Step Centre at Manchester, placing students into public, private and voluntary sector organizations where they can practise their learning in the workplace. The results are highly encouraging, and already forming the basis for publications on the link between students learning in the classroom, and their application of their understanding in the workplace (Carter *et al.*, 2016, under submission). I will highlight the lessons taken from the experiential learning as part of the Roundtable discussions.

JAMES NICHOLSON'S JOURNEY:

The backstory

I grew up in Belfast, and went to Cambridge to study mathematics, where there was a completely free choice of courses in the second half of the degree – and I chose a broad range of courses including a substantial number with a stochastical basis. In contrast, I found when I returned to Belfast as a Head of Department later in my career, a very high proportion of the teachers in post had studied at Queen's University Belfast, where the degree structure was much more rigid, and very few of them had any substantial statistical background in their degree.

After Cambridge I taught for 12 years at a school in London which had a very unusual timetable structure – mathematics had 1.5 times the standard period allocation for subjects taken in

the sixth form enabling all students to fully cover both the mechanics and the statistics options. So for the first 12 years of my teaching career I taught both mechanics and statistics – giving me a very unusual breadth of experience in teaching applied mathematics in a short period of time.

When I returned to Northern Ireland I found that within my own department, and within schools generally, there was a scarcity of teachers who were confident in their ability to teach the statistical part of the mathematics courses. Teaching resources available did little to develop confidence in teachers – material was almost always presented as a mathematical procedure with little or no attempt to explain the rationale for different techniques, and examples were seen purely as a vehicle for practising the technique to develop procedural competence.

The high-stakes assessment of statistics exacerbated this situation further – the focus within qualifications at ages 16 and 19 was primarily on procedural competence: examiners set a lot of questions with no context, and where contexts were used, they predominantly set questions with data the examiner made up to have the properties they wanted to assess, irrespective of whether it was realistic in the context – I remember one question where data claiming to be the numbers of letters in the words of a paragraph in a children's book did not contain any words with less than 4 letters – the question certainly assessed the mathematics of calculating a mean from a frequency table of discrete values but did nothing to encourage a view that statistics is a relevant and useful discipline in the real world. Regrettably, little has changed in terms of the emphasis on procedural competence over the past 20 years.

A much more severe impediment to developing a strong teaching workforce for statistics education was the number of times that questions were set which contained errors. In May 2000 CCEA (the examination body in Northern Ireland) sent guidance to schools alerting them to errors or difficulties with 36 questions on 18 statistics papers set between February 1996 and February 2000 – more than one quarter of all the questions which appeared on those papers. I had acted as part of the examination team for some of those papers but resigned when errors I identified in advance of students taking the high-stakes assessment were not corrected. The board did not publicly acknowledge the errors on any of these papers during that 4 year period: this meant that teachers, who were preparing candidates to sit examinations, were using past paper questions which were not fit for purpose.

That was the back story to where I started to think more explicitly, and formally, about the conceptual understanding of my students in statistics classes, and started to become involved in national and international statistics education initiatives – talking about the things which had always informed my personal classroom practice.

Using real data and simulations to teach traditional statistics

Between 1995 and 2002 my focus was primarily on using simulations to support the development of key statistical concepts, see for example Nicholson & Mulhern, (2000) and Nicholson, Mulhern & Hunt (2002). The materials on Regression and Correlation can be seen at http://www.icse.xyz/discuss/regression/ and on Sampling and Estimation at http://www.icse.xyz/discuss/estimation/

At a time when almost all graph construction and calculation of summary statistics in the real world was already automated, the UK curriculum meant that students spent a disproportionate amount of time, drawing graphs by hand and repeating basic calculations – and doing nothing with the outcomes in either case.

Nicholson and Darnton (2003) looked at the type of reasoning required in national tests at age 14 and compared them with the requirements of the higher stakes assessments at ages 16 and 19. The hope was that as students worked through the system having had experience of assessments which genuinely required statistical reasoning, that the GCSE and GCE assessments at ages 16 and 19 would evolve to provide similar requirements of students. Sadly, this did not happen and the national tests in that format were lost because of the expense involved in trialling materials.

None of this involved multivariate data so far, but was very much concerned with using real data, and the use of technology to build robust conceptual understanding of the role of variation across the whole range of univariate and bivariate statistical techniques which appeared in the curriculum at that time. The mathematics qualification at age 16 did have a non-examination

assessment component which required students to carry out a statistical investigation. Nicholson (2003) encouraged the use of technology to partition multi-dimensional data by some characteristic to enable comparisons to be made, and there was an opportunity for some substantive statistical reasoning to take place in those investigations. Regrettably, in both the statistical and the mathematical investigations being experienced by the vast majority of students, the approach taken by teachers was very formulaic, and the same investigations were used repeatedly, meaning that there was no way to ensure that what a student handed in was actually their own work. The use of coursework was discontinued amid concerns about the reliability and validity of the assessments.

Engaging with multivariate data and social sciences

At about that time I got to know Jim Ridgway and Sean McCusker at Durham University, who shared a similar world view about statistics education and who were interested in exploring ways of making multivariate data accessible to secondary level students through data visualisations. So in 2003 I took a leave of absence from school to work part-time with them at Durham to develop data visualisations, a leave of absence which subsequently became permanent.

We developed materials over the next few years addressing real social issues – for example, the use of alcohol by young people, drugs, smoking, sexually transmitted infections (STIs), inequalities in educational performance, and crime data from the public disorder in the UK, see for example Ridgway, Nicholson and McCusker (2008, 2012). Ridgway, McCusker and Nicholson (2006) used Rasch scaling to show that items involving multivariate data, supported by appropriate use of technology, which are cognitively complex are often more accessible to children than the 'simpler' items they are presented with in mathematics lessons. One of the key strategies in the development of the interfaces was to observe pupils working with the materials and discuss with them in detail what they found easy or difficult to interpret, and broader issues regarding engagement.

We worked both with academically gifted students and with those who were in the lowest quartile of the ability spectrum – whose engagement with mathematics was often almost non-existent and who experienced almost no success with it. While the lower ability students would not be as articulate as the gifted students in describing stories in the data, they were able to identify and describe the key features, encompassing some quite sophisticated ideas such as trends in groups changing over time, and they described a feeling of accomplishment in making sense of what they could easily see was not the sort of trivial exercises they were often fed with in order for them to 'get something correct'.

None of the development work could happen of course without teacher engagement. Many teachers of other subjects wanted to make clear that they were not sure of their ability to handle mathematical ideas – but once they saw students working with the interfaces they realized that the students really did not recognise they were 'doing any maths' because they did not do any actual calculations or construct any graphs themselves – and what was happening was that students were asking questions about the context – *why does this or that happen*? Mathematics teachers tended to be more anxious as to whether they could handle the discussion aspects because that is often problematic with many textbook examples of data interpretation: somewhat perversely, the extra complexity of the data seems to liberate students – where in many cases with mathematics textbook questions they feel there is one particular thing they are being asked for and are not sure what it is, with these multivariate data sets, there is transparently not a single statement which can cover the 'stories in the data' and they are liberated enough to be able to start discussing it – and that teachers were happier to mediate such discussions (having been supplied with some support materials about the data interpretation) because of the readiness of students to engage with the data. Nicholson, Ridgway and McCusker (2006, 2009) and Ridgway, Nicholson and McCusker (2007 a, b, c, 2008 a, b, 2009) report on these matters much more fully. The curriculum materials produced in an early project for the Northern Ireland curriculum authority can be found at https://www.dur.ac.uk/smart.centre/curriculum_material/.

Nicholson, Ridgway and McCusker (2006) and Ridgway, Nicholson and McCusker (2007) argued that reasoning with data was already pervasive in our society and that the curriculum needed to change to prepare our young people adequately to take their place as fully

informed citizens. Social sciences and other subjects were trying to deal with complex situations where the observed data could not be reduced to univariate and bivariate (straight line) data without grossly oversimplifying what was going on. However, students only met univariate and bivariate data in their mathematics courses, and teachers in social sciences (and the textbooks used) would describe single factor headline statistics in looking at educational performance inequalities for example *girls do better than boys – 7.3% more girls get 5 good GCSEs including English and Mathematics* and *less than half as many children known to be eligible for free school meals get 5 good GCSEs including English and Mathematics than in the rest of the population (26.6% compared with 54.2%)*. They would then give the theory (theories) in the discipline which provides explanations for why this happens – but intertwined with this, they also needed to identify subsets of the population for which these statements were not true, for example within the Chinese ethnic group there was less than 1% difference between those known to be eligible for free school meals and those who were not – so the text accompanying the graphs of single variable or two variable outcomes was a mixture of describing patterns in data, introducing theory from the discipline to explain why these occur, and detailing special demographic groups who did not behave in a manner consistent with the main theory and trying to provide modifications to the theory to account for the observed behaviour. Nicholson, Ridgway and McCusker (20011a, b) describe data visualisations and associated classroom activities related to education and health inequalities which were developed as part of a Nuffield funded project (EDU 37713: Reasoning from Evidence) – materials are available at https://www.dur.ac.uk/smart.centre/nuffield/. The use of these data visualisations allows students to have a grasp of what the data landscape is – and what the theory needs to provide an explanation for.

Developing data visualisations for the 2011 Census revealed significant tensions between user requirements and technical development. The Census contains significant amounts of multivariate data, available at country, regional and Local Authority geographies and variables often have a large number of values – for example 'ethnic group' has 5 main categories and a further 18 subcategories. Researchers (quite reasonably) wanted to be able to have complete freedom to choose whatever geographies they wanted to make comparisons with – which includes geographies at different levels – and what values they wanted to compare within each variable. We currently need to curate a data set with specified levels of each variable in order to import it into the data visualisation so our Census visualisations were limited to national pictures for most data sets with some exemplars of what could be done with more detailed data sets. We hope to be able to import data sets through the Nomis interface with ONS data into the Smartplotter in the not too distant future, which would make the tool much more useful for both research and teaching purposes.

The need for understanding complex data is much wider than in education and we have recently collaborated with others including the House of Commons Library to support public understanding of civic statistics by the general public as well as politicians and policy makers. Sutherland *et al*. (2015) provide an account of some of the range of design issues encountered in the development of the Constituency Explorer interface (see www.constituencyexplorer.org.uk). One of the critical difficulties that the public, and teachers at both secondary and tertiary levels, face is the plethora of data based resources that exist – and the time and effort which would be required in order to evaluate their suitability or reliability.

Reflections on this journey

I learnt my own mathematics and statistics by making sense of it, not just rote learning techniques – I would naturally, unconsciously, think about how things worked when mathematics was being applied to a context. When I started teaching I realized that this is not a universal attribute in students naturally, but that it could be nurtured, although with varying degrees of success: indeed some students were very reluctant to engage with it at all. In the late 20th century, education, in common with other areas of public life, became subject to an accountability regime involving targets and indicators, many of which started as good indicators of what they purported to measure, but once funding was attached to performance against such measures, schools were able to find ways of improving performance against the measure without it actually being a real improvement in the wider context of school performance. Regrettably, this emphasis on

qualification outcomes seemed to result in a competition between qualification providers to grab an increased share of very lucrative markets, and a plethora of textbook resources authored, in many cases, by senior examiners, promised schools the best chances of students passing. The reality was that the resources were often extremely focused on drill and practice in particular types of questions that the provider commonly used in examinations, and offered little in the way of insights into how the real world works or the relevance of data in solving important real world problems.

The latest round of curriculum changes have resulted in an increased focus on the relevance of mathematics and statistics to the real world, and new rules for publishers and authors offer hope of improving the quality of textbook and other teaching resources. However, Nicholson (2015) argues that the continuing status of exclusively univariate and bivariate data techniques means that the scope for really improving statistical education through the use of real data is much more limited than it should be and increasing the emphasis on hypothesis testing is perverse at a time where the need for it is diminishing.

The opportunities for supporting social sciences directly through data visualisations of multivariate data are proving to be much more fruitful than attempts to move statistics education *per se* at school level to reflect the changing world of data.

DISCUSSION

Changes, challenges and opportunities

The individual journeys described have taken two educators into remarkably similar territory, albeit at different educational levels, in teaching statistical literacy. It is apparent from the two narratives described that a number of factors have changed across the years. Here we set out the changes observed, and draw out the challenges faced, as well as the opportunities now offered to statistical educators. We frame this discussion in the context of formal education, but with an appreciation that the lessons we draw out from these combined experiences have resonance with statistics education and literacy more broadly, in a world of increasing volumes of numeric data.

Changes

Beginning with changes, across the two decades, representing forty combined years, covered, there are perhaps three main areas to highlight. First is the availability of real-world data sources; both authors gave numerous examples of socioeconomic data that were used to develop new interfaces, new teaching resources, or exam questions. Twenty years or more ago this was not the case; today services making data available for research and teaching are significantly more user-friendly, and student-friendly. Hand-in-hand with this goes the inevitable progression of technology; whereas in the early nineties computers were lab-based, expensive and software applications required technical know-how (via plugins, or software written for specific browsers, and the like), in the 2010s technology has progressed to provide powerful analysis capability at the touch of a button on a handheld device, and the rise of social media allowing students to share insights about to get things to work. Coupled with these two factors is a change in teaching statistics, at least in the social sciences, from the mechanics of manipulating data to increased emphasis on the practice or application of statistics, together with an increased focus on the interpretation of results. The combination of more real-world data, better applications on mobile devices, and a move towards applied data analysis provides both challenges and opportunities for us as educators.

Challenges

These changes open up challenges to those involved in increasing statistical literacy. Whilst there has been an increase in data sources, including government statistics and administrative data that can be used in teaching, the data alone are not enough. Teaching needs to incorporate critical assessment of these data, their sources, their purpose, their representativeness and so on. It's not simply a case of crunching numbers. Technology advances can help in this regard (which we discuss below), but the challenge we all face is to whet students' appetites with the stories in the data, often delivered through visualizations or graphics, or headlines, in order to

proceed to critically evaluate these findings. In a teaching context the challenge often boils down to having lab space and time to do this, as well as suitable software.

The time required to teach statistics with a focus on its application is an obstacle. However we think there are parallels with doing some practical work in school statistics – it takes time, but it is time well spent in developing sound conceptual understanding of key ideas by experiencing variability – and saves time later on in the course unpicking difficulties which occur when that foundation is not secure. At this level, working with multivariate data, students in the past have had no mental schema available for what things like interactions mean, or look like, in complex data and so the interpretation of computer output of sophisticated and nuanced multivariate analyses has commonly been done in a formulaic and simplistic fashion. We feel that the new visualisations of multivariate data can help enormously to support students' understanding of the data landscape that their discipline theory is seeking to explain, and in the long term the stronger conceptual understanding of data structures will prove to be time well spent. There is a major tension in providing new visualisations between the desire to be able to import new data automatically and ensuring that the visual display accurately depicts the data. There is also a need for careful thought to be given to the scaffolding needed in a data visualisation for different levels of students – and probably in different disciplines.

Perhaps one of the greatest challenges, and obstacles to statistical literacy, is 'fear of maths'. Attitudes to learning matter: we know at university level, those coming to study social science mostly think they left maths behind at school, and some strongly resist being taught statistics. The challenge this poses to educators is not to be underestimated, but nor should it provide a self-fulfilling prophecy. Both authors have experience of transformational learning experiences for students, especially when they see the relevance of numbers, why they matter, and why thinking statistically helps them ask and answer important substantive social research questions. The British Academy has funded a special project to investigate 'maths anxiety', with Carter as co-investigator. Its purpose is to 'provide an overview of existing research and its implications for policy and practice, identifying potential interventions and gaps in our knowledge …to deepen awareness and demonstrate the importance of …quantitative skills. (Ref British Academy http://www.britac.ac.uk/news/news.cfm/newsid/1422). It is expected that the meta analysis and literature review undertaken under this project will inform the understanding of statistical literacy as it relates to anxiety in the population about maths.

A third challenge is that of sharing good practice. How can we better encourage or incentivize teachers to share their material that they labour so diligently over? Good examples – why the numbers matter – are central to engaging students meaningfully in working with complex data, but they can take considerable effort on the part of the instructor to get them into a usable form. The Q-Step centres are attempting to do this, for example the University of Exeter has developed a series of videos 'Methods Matter' to 'offer quantitative methods modules in an innovative and accessible way' (http://socialsciences.exeter.ac.uk/news/college/title_513153_en.html). Four videos are currently released on Youtube (https://www.youtube.com/channel/UCWA-QbolpCJS3zZQzDSKXGg/) covering Ordered Logit Regression, Linear Regression, Association Deductive Block Modelling and Network Analysis. Through interviews with experts in social research methods exploring substantive research matters, such as covert networks during the suffragette movement, the videos uncover how the quantitative methods are applied. Although challenging to develop, sharing resources online in this way can help us to better understand from each other how statistical methods are applied in research. Summer Schools are a conventional way to develop and practice research skills, but these are costly and time-consuming. Nonetheless, many challenges can also be seen as opportunities, and we will return to the challenge of sharing resources to support statistical literacy in the next section as an opportunity.

Opportunities

There has been considerable improvement in the past 20 years in the tools available to work with complex data. There are further opportunities for improving ways into data for students, for example through apps, and in particular for increasing our understanding of the ways that students interact with complex data and ensuring that data interfaces do not provide incorrect or

misleading representations. For example, it is disturbing how often the first view of a data set in an interface uses autoscaling. The willingness and enthusiasm of organisations like the UK Data Service and others like ONS, World Bank, ICPSR, OECD, and IPUMS to provide access to data in an intelligible form offers opportunities for teachers to build in the use of data relevant to their subject content rather than trying to build a content program around the available data sets.

Part of the aim of the Q-step initiative is that the expertise and resources generated by the funding provided should be shared across the higher education sector in the UK through a support programme. There is an opportunity for the community developing quantitative skills training to act in a much more collaborative spirit than has perhaps been evident up to now, for the most part. The example of Exeter's "Methods Matter' videos on YouTube provides an example of this starting to happen due to the Q-Step programme. The ProCivicStat project is a further example of such collaborations producing open access resources for the wider community to use. A real opportunity in that collaborative spirit is for interdisciplinary communities to grow – the Q-step centres normally involve a number of social science disciplines working together in ways they did not previously do and engaging more directly with mathematics and statistics specialists than they would naturally do. Contributors to this conference come from a much wider range of backgrounds than is common for a statistical education meeting – including sociology, psychology, food science, epidemiology, political science, law, business, and software design. It is to be hoped that lasting links can be made between these disciplines and the statistical education community, which will be mutually beneficial.

Little has been said here about online learning, with Moocs (Massive Open Online Courses) providing an obvious opportunity for sharing good pedagogic approaches to teaching quants. The University of Auckland's 'Data to Insight' (see https://www.futurelearn.com/courses/data-to-insight) and University of Sheffield's 'Making Sense of Data in the Media' (see https://www.futurelearn.com/courses/media-dataprovide two examples of recent Moocs that have developed to support such shared approaches. Both courses provide opportunity for novice learners to improve their statistical understanding, based on real world data examples. It is encouraging to see the development of online learning platforms in statistical education; the opportunities afforded by this approach remain to be evaluated, and it is hoped that these can be discussed by the Roundtable participants. A final question this raises however is in the challenge of not only the creation of these free resources, but their discoverability and use.

We hope that participants at the roundtable will share their experiences and views on what they see as the opportunities in this move towards using real data in classrooms and lecture theatres.

ACKNOWLEDGMENT:

The work reported in this paper was supported in part by ProCivicStat project, a strategic partnership of the Universities of Durham, Haifa, Ludwigsburg, Paderborn, Porto and Szeged, funded by the ERASMUS+ program of the European Commission. However the views and opinions expressed in this paper are those of the authors and do not necessarily reflect those of the funding agency.

REFERENCES

Adeney, K., & Carey, S. (2011). How to teach the reluctant and terrified to love statistics: the importance of context in teaching quantitative methods in the social sciences. In: G. Payne and M. Williams (Eds.), *Teaching Quantitative Methods: Getting the Basics Right*, 85-98. Sage Publishers.

Andrienko, G., Andrienko, N., Voss, H., & Carter, J. (2000). *Internet mapping for dissemination of statistical information*, Computers, Environment and Urban Systems, Vol. 23, 425-441.

British Academy (2015) British Academy urges UK government to address numeracy crisis. Count us in: quantitative skills for a new generation http://www.britac.ac.uk/news/news.cfm/newsid/1285(accessed 19 Feb 2016)

Buckley, J., Brown, M., Thomson, S., Olsen W., & Carter, J. (2015). Embedding quantitative skills into the Social Science Curriculum: case studies from Manchester, *International Journal of Social Research Methodology*, 495-510, DOI: 10.1080/13645579.2015.1062624.

Carter, J. (2002). Enhancing use of JISC Data Services *VINE: The Journal of Information and Knowledge Management Systems*, 126(1), 40-47.

Carter, J. (2003). *Learning and Teaching with the UK Census* ASSIGNation 20(4) ISSN: 0265-2587, 15-18.

Carter, J. (2004). *The JORUM+ Project* ASSIGNation 21(2) ISSN: 0265-2587, 43-48.

Carter, J. (2010). Real Data; Real World; Real Stories: A Case Study Approach to Demonstrating Impact on the Student Experience, *EDULEARN10 Proceedings* ISBN: 978-84-614-2439-9, 4792-4802.

Carter, J. (2012). *How OERs can help a Strategically Important and Vulnerable Subject Area - Quantitative Social Science* JIME Journal of Interactive Media in Education http://jime.open.ac.uk/jime/article/view/2012-16.

Carter, J., & Dykes, J. (1998). Developing a Visualization Gateway for the UK Census of Population Real *Spatial Data Handling Proceedings*, *8th International Conference,* 543-555.

Carter, J., Brown, M., & Simpson, K. (2016). (Under submission). Making it real: from the classroom to the workplace. Special Edition of *Statistics Education Research Journal.*

Carter, J., Noble, S., Russell, A., & Swanson, E. (2011). Developing Statistical Literacy using Real-world Data: Investigating Socioeconomic Secondary Data Resources used in Research and Teaching. *International Journal of Research & Method in Education* 34(3).

Carter, J., Rice, R., & Cole, K. (2005). JISC's Learning and Teaching Programme: the Data Cluster, *VINE: The Journal of Information and Knowledge Management Systems*, 35(1/2) ISSN: 0305-5728, 43-48.

Cole, K., & Carter, J. (2001). http://slideplayer.com/slide/796405/ (accessed 8 March 2016)

ESRC (1987), *Horizons and Opportunities in the Social Sciences*, London: ESRC.

HEFCE (2005) *Strategically Important and Vulnerable Subjects: Final Report of the Advisory Group*. HEFCE Policy Development, June 2005/24

MacInnes, J. (2009). *Proposals to support and improve the teaching of quantitative research methods at undergraduate level in the UK*. ESRC

Mills, D., Jepson, Al, Coxon, T., Easterby-Smith, M, Hawkins, P., & Spencer, J. (2006). *Demographic Review of the UK Social Sciences*. London: ESRC.

Nicholson, J. (2003). GCSE statistics coursework, and dealing with multi-dimensional data. *Mathematics in Schools, 32(3),* 8-14.

Nicholson, J. (2015). Reasoning with data or mathematical statistics: Is the UK moving in the right direction? *Proceedings of the IASE /ISI Satellite, 2015*.

Nicholson, J., & Darnton, C. (2003). Mathematics teachers teaching statistics: What are the challenges for the classroom teacher? *Proceedings of the Proceedings of the 54th Session of the International Statistical Institute.* Berlin, Germany

Nicholson, J., & Mulhern, G. (2000). Data interpretation in the 21st century: Issues in the classroom. In: A. Rogerson (Ed.), *Mathematics for Living Conference Proceedings. Amman, Jordan: The Mathematics Education into the 21st Century Project*.

Nicholson, J., Mulhern, G., & Hunt, N. (2002). Wizadry or pedagogy? What is the driving force in the use of the new technology in teaching statistics. In *Proceedings of the Sixth International Conference on Teaching Statistics.* ISI, Voorburg: Netherlands.

Nicholson, J., Ridgway, J., & McCusker, S. (2006). Reasoning with data – time for a rethink? *Teaching Statistics, 28(1),* 2-9.

Nicholson, J., Ridgway, J., & McCusker, S. (2011). Visualise then conceptualise. *Social Science Teacher* 40(3) 8-13.

Nicholson, J., Ridgway J., & McCusker, S. (2012). The summer riots 2011... lessons to be learnt? T*eaching Citizenship,* 34, 24 – 27.

Nicholson, J., Ridgway J., & McCusker, S. (2013). Health, wealth and lifestyle choices - Provoking discussion on public spending. *Teaching Citizenship 36, 23 – 27.*

Ridgway, J., McCusker, S., & Nicholson, J. (2006). Reasoning with evidence. Development of a scale. Paper presented at the 32nd Annual Conference of the International Association for Educational Assessment Conference, Singapore, 2006.

Ridgway, J., Nicholson, J., & McCusker, S. (2007). Teaching Statistics – Despite its Applications. *Teaching Statistics. 29(2),* 44-48.

Ridgway, J., Nicholson, J, & McCusker, S. (2008). *Alcohol and a Mash-up: Understanding Student Understanding*. ENAC biannual conference, Berlin.

Russell, C., Noble, S., Carter, J., & Wiseman, R. (2011). Real World, Real Stories: Teaching Quantitative Methods with Real Life Data'. *EDULEARN11 Proceedings* ISBN: 978-84-615-0441-1, 5748-5754.

Sutherland, S., Hedger, S., Ireland, M., & Ridgway, J. (2015). Designing interactive displays to promote effective use of evidence. *Proceedings of the IASE /ISI Satellite, 2015.*

Web Archive (2016) http://www.webarchive.org.uk/wayback/archive/20070613120000/http://www.mimas.ac.uk/argus/Tutorials/CartoViz/DataDetails/index-2.html (accessed 8 March 2016)

APPENDIX – LIST OF RESOURCES

Core Maths in England (see http://www.core-maths.org/) are new qualifications for 16 – 19 year old pupils not going on to study A-level Maths

Digimap service (digimap.edina.ac.uk and https://census.edina.ac.uk/thematic/map)

'Learning to share, and sharing to learn' case study at http://www.jorum.ac.uk/wp-content/uploads/2015/01/Learning-to-share-and-Sharing-to-Learn.pdf).

ReStore (http://www.restore.ac.uk/index.php),

NCRM's Quantitative Methods Initiative website (http://www.quantitativemethods.ac.uk)

In the US, ICPSR's Teaching with Data Guides (https://www.icpsr.umich.edu/icpsrweb/instructors/biblio/resources).

Materials on Regression and Correlation: http://www.icse.xyz/discuss/regression/http://www.icse.xyz/discuss/regression/

Materials on Sampling and Estimation: http://www.icse.xyz/discuss/estimation/

Durham SMART Centre materials addressing real social issues:
use of alcohol by young people: https://www.dur.ac.uk/smart.centre/becta/
sexually transmitted infections (STIs):
https://www.dur.ac.uk/resources/smart.centre/Freeware/STI_GUM_update1.swf
inequalities in educational performance: https://www.dur.ac.uk/smart.centre/nuffield/education/
public disorder in UK 2011L https://www.dur.ac.uk/smart.centre/nuffield/public_disorder/
curriculum materials developed for CCEA (on alcohol drugs, obesity, pensions and savings, poverty, smoking and STIs: https://www.dur.ac.uk/smart.centre/curriculum_material
constituency data (in collaboration with The House of Commons Library)
http://www.constituencyexplorer.org.uk

'Methods Matter' http://socialsciences.exeter.ac.uk/news/college/title_513153_en.html

Videos covering Ordered Logit Regression, Linear Regression, Association Deductive Block Modelling and Network Analysis: https://www.youtube.com/channel/UCWA-QbolpCJS3zZQzDSKXGg/

The University of Auckland's 'Data to Insight' (see http://www.education.auckland.ac.nz/en/about/notices/notices-2014/2014/09/mooc_data-to-insight.html)

The University of Sheffield's 'Making Sense of Data in the Media' (see https://www.futurelearn.com/courses/media-data)