Enhancing statistical literacy – multivariate thinking in precollege social sciences courses

James Nicholson
Durham University, School of Education
Leazes Road
Durham, DH1 1TA, UK
j.r.nicholson@durham.ac.uk

Jim Ridgway
Durham University, School of Education
Leazes Road
Durham, DH1 1TA, UK
Jim.Ridgway@durham.ac.uk

Sean McCusker
Durham University, School of Education
Leazes Road
Durham, DH1 1TA, UK
Sean.McCusker@durham.ac.uk

Introduction

Social scientists theorise about issues such as inequalities in health, educational achievement, criminal behaviour or the distribution of wealth. The starting point for theory should be a clear understanding of phenomena. Teaching social sciences at precollege level is a challenge in part because social phenomena almost always involve a number of categorisations in groups, such as by social class, ethnicity or sex. Up to now the complexity of social data in schools has been a substantial obstacle, and the use of data has often been restricted to headline statistics based on aggregated data, and explorations of the behaviour of these groupings considered separately.

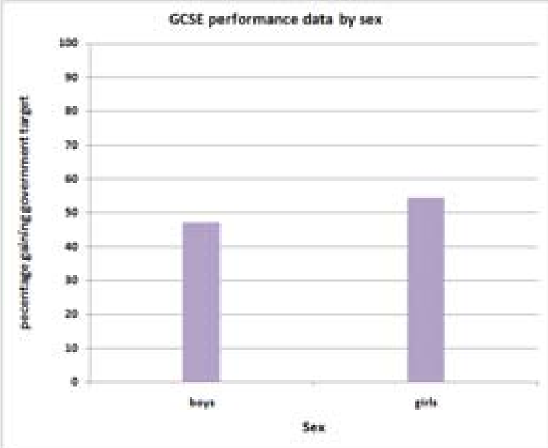
The Nuffield Foundation has funded a project, *Reasoning from Evidence*, which is developing some data visualisation tools to support the teaching of GCE A-level Sociology (student age typically 16 – 19), but the tools have much wider application across many subjects and wider age ranges. We will try to articulate the rationale for engaging with the interactions between multiple factors through an example in educational achievement where there are inequalities in sex, social class and ethnic groupings.

Case study: inequalities in educational performance in the UK

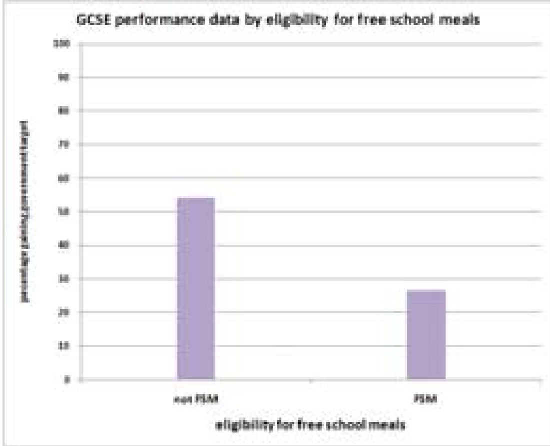
Figure 1 is a panel of graphics showing academic performance data for the UK in 2009, comparing the proportions of pupils who achieved 5 good GCSE grades including English and Maths (the UK government‟s target performance level at age 16). While these graphs are not taken from a textbook, they are similar to the graphs which do appear: they represent current data of the sort which is commonly presented over a number of pages in a textbook i.e. one of these graphs would be dealt with in a section of text before moving on to the next graph and discussion. A brief commentary is presented below each graph which summarises the story in the graph in a way the textbook commonly would.

Sometimes a pair of these will be combined into a comparative bar chart, so for example, the performances of boys and girls might be shown for a number of years, or ethnic groups compared for two years – see figure 2 for some examples of what these graphs would look like.

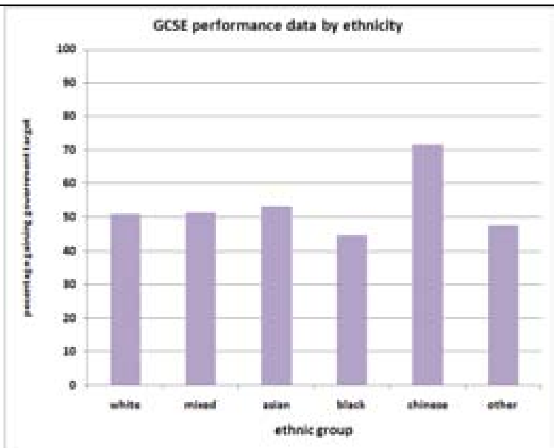
One feature of these graphs which would not be commonly shared by textbook versions is that the vertical axis goes to 100 (%), meaning there is a realistic view of the actual performance levels of all groups, not just an accurate view of their relative performance.



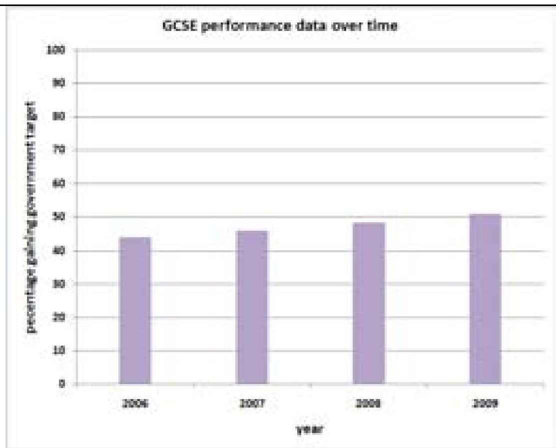
About 7% more girls than boys achieve government target levels.



Less than half as many pupils who are eligible for free school meals achieve the government target level compared with those who are not known to be eligible for free school meals.

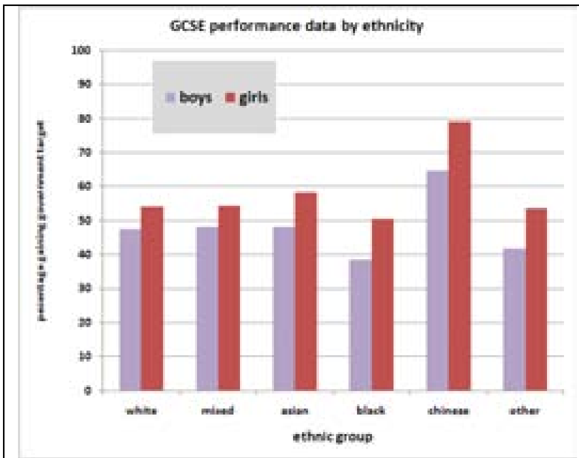


Chinese outperform other ethnic groups and Black have the lowest achievement levels.

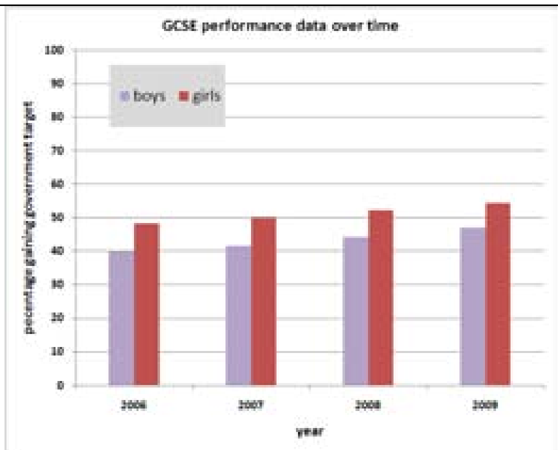


The proportion gaining the government target levels has increased steadily from 44% in 2006 to 50.7% in 2009.

Figure 1: single factor graphs showing inequalities in educational performance.



Chinese outperform other ethnic groups and Black have the lowest achievement levels. Girls do better than boys in all ethnic groups.



The proportion gaining the government target levels has increased steadily from 2006 to 2009 for both boys and girls, with girls always doing better.

Figure 2: comparative graphs showing inequalities in educational performance.

The use of technology offers the opportunity to let even young students get to grips (literally) with disaggregated data. Using drag and drop variable place-holders they can manipulate the data to explore which factors are associated with the biggest effects, and to describe patterns in the data. These patterns are often non-linear, and the patterns across values of one factor are sometimes different across values of another factor. Students can have the confidence to talk about the stories that the data tell, and can then judge the extent to which rival theories are consistent with the evidence.

The textbooks we have looked at provide quite a lot of text discussing the inequalities, and some headline statistics about them, so for example it is widely reported that girls do better than boys, and that students in higher socio-economic groups do better than those from lower groups. However, if one is constructing theories about educational attainment, it would be useful to know, for example, if the differences between boys‟ and girls‟ attainment is:

- the same across socio-economic groups;
- across all ethnic groups.

One difficulty with textbooks for any course is that any data references in them are fixed by the date of publication – teachers might look for material to provide more up to date information about the state of affairs, and there are many sources available through the internet. For example, The Equality and Human Rights Commission report (2010) section 10.4 *What we know about educational attainment at age 16* uses the main measure of educational attainment (5 A* - C at GCSE, including English and Maths). The underlying data being used appears in table 2 of *Key Stage 4 attainment by Pupil Characteristics, in England 2008/09* which is available from:
http://www.education.gov.uk/rsgateway/DB/SFR/s000900/index.shtml and is shown in figure 3.

Table 2: Achievements at GCSE[1] and Equivalents by ethnicity, free school meals and gender
Coverage: England[2]
Year: 2006-2009 [4]

Please select criteria below:	
Achievements:	5+ A*-C grades inc. English & Mathematics
Year:	2009

	Pupils not eligible for free school meals						Pupils known to be eligible for free school meals						All pupils[3]			
	Number of pupils			Percentage achieving 5+ A*-C grades inc. English & Mathematics			Number of pupils			Percentage achieving 5+ A*-C grades inc. English & Mathematics			Number of pupils			Percenta grad M
	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total	Boys
White	219,015	210,861	429,876	50.7	57.7	54.1	25,858	24,962	50,820	19.4	24.5	21.9	244,880	235,835	480,715	47.4
White British	210,556	202,958	413,514	50.8	57.8	54.3	24,397	23,518	47,915	19.0	24.0	21.5	234,960	226,485	461,445	47.5
Irish	860	792	1,652	62.9	65.3	64.0	161	131	292	22.4	26.0	24.0	1,021	923	1,944	56.5
Traveller of Irish Heritage	27	37	64	x	18.9	x	32	23	55	x	x	x	59	60	119	5.1
Gypsy/Roma	160	138	298	6.3	15.2	10.4	90	98	188	4.4	9.2	6.9	250	236	486	5.6
Any other White background	7,412	6,936	14,348	46.3	54.6	50.3	1,178	1,192	2,370	29.1	34.6	31.9	8,590	8,131	16,721	43.9
Mixed	6,780	6,791	13,571	53.3	59.3	56.3	1,624	1,691	3,315	26.6	34.7	30.7	8,404	8,483	16,887	48.2
White and Black Caribbean	2,347	2,448	4,795	42.9	50.5	46.8	700	744	1,444	23.7	30.9	27.4	3,047	3,192	6,239	38.5
White and Black African	610	649	1,259	50.8	62.2	56.7	149	178	327	26.2	31.5	29.1	759	827	1,586	46.0
White and Asian	1,433	1,329	2,762	65.2	68.8	67.0	257	231	488	29.2	43.3	35.9	1,690	1,560	3,250	59.8
Any other mixed background	2,390	2,365	4,755	57.1	62.3	59.7	518	538	1,056	29.3	37.4	33.4	2,908	2,904	5,812	52.1
Asian	16,181	15,182	31,363	52.4	62.9	57.5	5,175	4,974	10,149	35.1	44.3	39.6	21,358	20,156	41,514	48.2
Indian	6,128	5,862	11,990	64.8	73.7	69.2	674	690	1,364	43.3	52.6	48.0	6,803	6,552	13,355	62.7
Pakistani	5,736	5,283	11,019	41.9	52.4	46.9	2,628	2,397	5,025	30.4	38.5	34.2	8,365	7,680	16,045	38.3
Bangladeshi	1,639	1,583	3,222	48.7	57.0	52.8	1,314	1,453	2,767	37.7	47.8	43.0	2,953	3,036	5,989	43.8
Any other Asian background	2,678	2,454	5,132	48.7	63.8	55.9	559	434	993	41.9	51.2	45.9	3,237	2,888	6,125	47.5
Black	8,209	8,499	16,708	42.3	55.5	49.0	3,365	3,532	6,897	28.8	38.5	33.8	11,577	12,032	23,609	38.4
Black Caribbean	3,103	3,167	6,270	35.8	48.4	42.1	823	904	1,727	22.4	36.0	29.5	3,927	4,071	7,998	33.0
Black African	4,127	4,353	8,480	47.9	61.9	55.1	2,188	2,267	4,455	31.7	39.4	35.6	6,317	6,621	12,938	42.3
Any other Black background	979	979	1,958	38.9	49.7	44.3	354	361	715	26.3	39.1	32.7	1,333	1,340	2,673	35.6
Chinese	1,083	983	2,066	64.8	79.3	71.7	106	103	209	64.2	77.7	70.8	1,189	1,086	2,275	64.8
Any other ethnic group	2,225	1,874	4,099	44.8	56.1	50.0	1,041	960	2,001	35.8	49.1	42.2	3,266	2,834	6,100	41.9
All pupils[1]	**256,984**	**247,260**	**504,244**	**50.6**	**58.1**	**54.2**	**37,690**	**36,688**	**74,378**	**23.4**	**29.8**	**26.6**	**294,782**	**284,063**	**578,845**	**47.1**

Figure 3: Key Stage 4 attainment in England 2008/09

The report provides the commentary (see figure 4) and graphical summary (see figure 5), but does not reproduce the table of data:

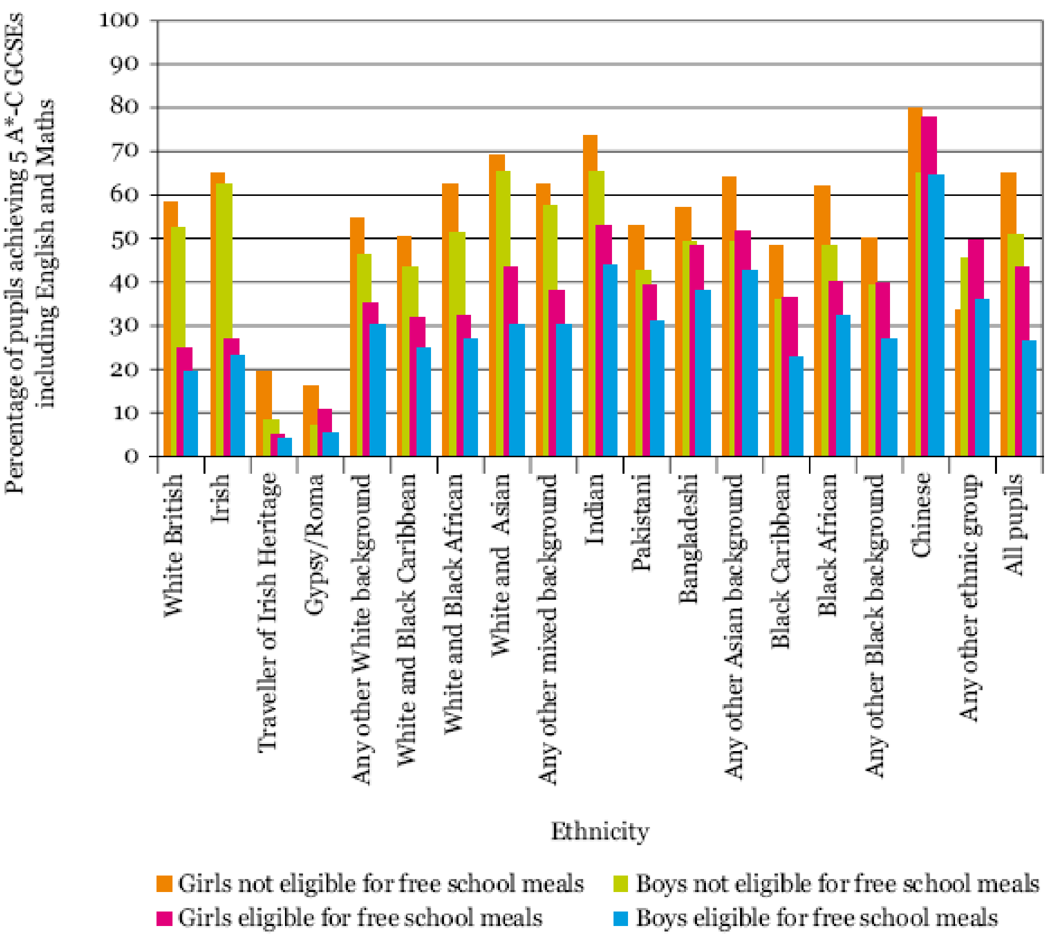
Socio-economic groups
FSM eligibility reveals one of the starkest differences in achievement at age 16.
Being eligible for FSM is associated with a much lower probability of achieving 5+ good GCSEs including English and Maths. In 2009 only 27% of students eligible for FSM achieved this compared to 54% for those not eligible.

Figure 10.4.1 below shows that inequalities in England vary considerably within different ethnic groups when combined with eligibility for FSM. It reveals, for example, that while both Chinese and Indian students do very well overall, performance of the latter varies much more by socio-economic status. So whereas Indian boys from low income families perform significantly below the overall average, their Chinese counterparts are well above average.

When analysing how ethnicity, gender and FSM eligibility affect GCSE performance, the National Equality Panel found that the FSM effect was larger than any of the other associations shown, with the exception of the higher performance of Chinese pupils, and the lower performance of Gypsy and Traveller children.

Figure 4: text on education inequalities in the How fair is Britain? report.

Figure 10.4.1 Percentage of pupils achieving 5 A*-C GCSEs including English and Maths 2009 by gender, FSM status and ethnic group in England, 2008-09[75]



Source: Key Stage 4 attainment by Pupil Characteristics, in England 2008/09.

Figure 5: graph on education inequalities in the How fair is Britain? report.

The key difficulty with using data in situations such as these is that traditional flat, static graphs are simply not equipped to communicate the requisite level of complexity and attempts to do so (like this one) are not usable in a classroom, nor is the table of data accessible to students at this level. The group of graphs shown earlier in figure 1 looking at single factors do not tell the whole story, but often politicians will look for „sound bites" – focusing on statements deriving from a single classification, or on the performance of a single group defined by multiple classifications – see figure 6.

- Children on free school meals do significantly worse than their peers at every stage of their education.
- They are just half as likely to get good GCSEs as the average.
- This vast gap between rich and poor is not pre-ordained. In Finland and Canada the gap is much smaller. Even in this country there are some groups – Chinese girls on free school meals for example – who significantly outperform the national average.

Figure 6: statements made in the foreword to the 2011 Education White Paper by the Prime Minister and Deputy Prime Minister.

The data interface we have been developing in Durham allows users to explore the different factors systematically – the published data provided 6 main ethnic headings of which „White", „Black", „Mixed" and „Asian" were then subdivided – the graph above just takes all the subdivisions. The opening tab of the data interface shows just the main groups – there are separate tabs which allow you to explore differences within each of these main groups and it separates out the behaviour of different groups in classifications by whether pupils are eligible for free school meals (FSM), ethnicity and gender so they can be understood more easily. The drag and drop facility with the position of variables allows the relationships to be explored quickly and reliably. The interactive data interface and guidance on how to use it, with classroom materials, can be found at www.dur.ac.uk/smart.centre/nuffield .

The data interface makes it possible for all the subgroups generated by these categorisations to be viewed, in an accessible and intuitive format, where anomalies can be identified and explored – in figure 7 the user is looking just at White students (the pointer on the slider below the graph is at „White") and can see that the performance of both boys and girls who are eligible for free school meals is much lower than the average of the whole group, but the performance of those who are not eligible for free school meals is not much higher than the average of the whole group.

In a number of the schools involved in the development process, teachers reported that the majority of their Sociology students had weak mathematical backgrounds. Very few were taking A-level Mathematics, many had only a grade C at GCSE having only entered at the foundation tier, and a substantial minority were retaking GCSE Mathematics because they had got lower than grade C. This graph opened up a discussion as to how this could be: figure 8 shows the comparable graph for the Black group and the average of the whole group is much closer to the average of the FSM and Not FSM subgroups, particularly for boys.

The knowledge that the proportion of Black students who are eligible for free school meals is much higher than the proportion of White students was there already from other topics they had studied, but they had not realised the greater effect this would have on the headline performance of Black students compared with groups with lower proportions eligible for free school meals.

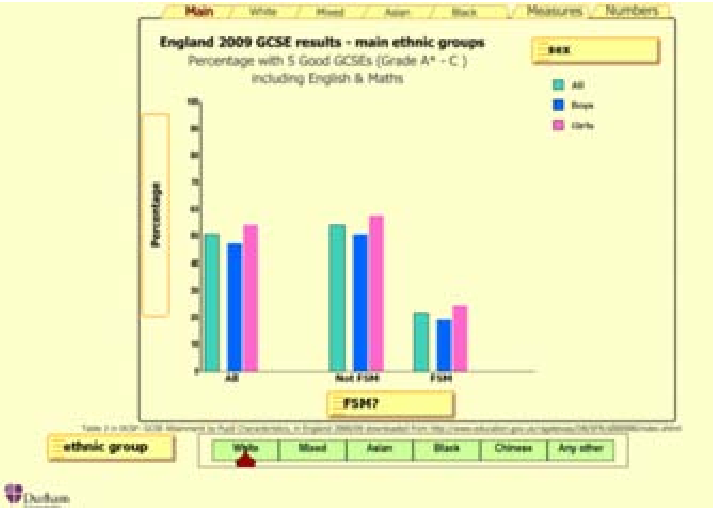


Figure 7: the performance of White pupils only.

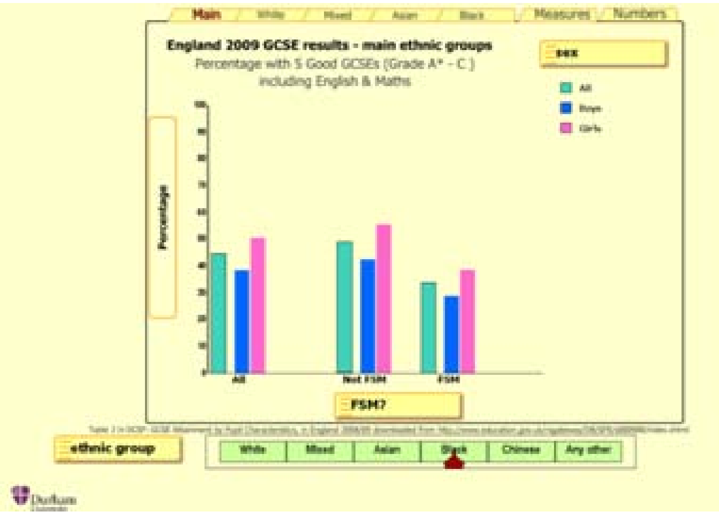


Figure 8: the performance of Black pupils only.

In figure 9 the user is comparing performance across the ethnic groups between those eligible for FSM and those who are not, and moving the slider between boys and girls, seeing that there are substantial differences in performance between White, Mixed, Asian and Black boys for those eligible for FSM, but for pupils who are not eligible for FSM the performance of these groups is very similar.

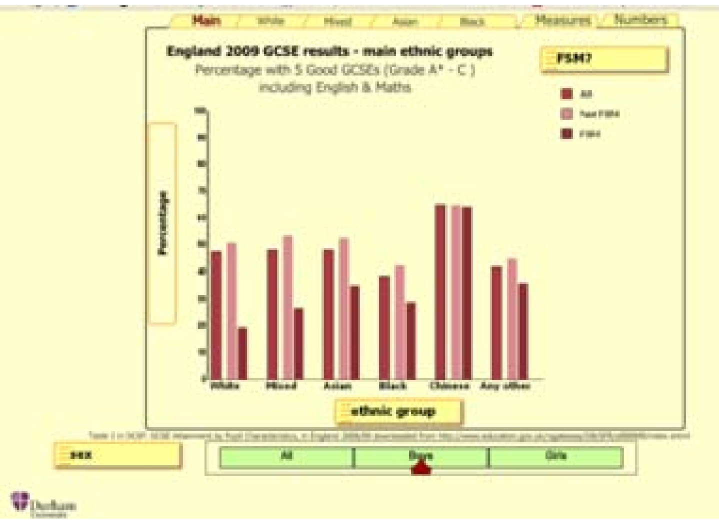


Figure 9: ethnic differences for boys only

So far we have only looked at the first page of the interface – the tab labelled Main which shows the data on the main ethnic groupings. However, the interface offers the opportunity to provide related data on other tabs – here, the tabs labelled White, Mixed, Asian and Black show the performance data by sex and eligibility for free school meals for the ethnic subgroups within those four main groups. Moreover, the tab labelled Measures shows performance data for the main ethnic groups by sex and gender as we have seen already but additionally allows the effect of including English and Maths in the performance measure to be explored – unsurprisingly it affects some ethnic groups more than others.

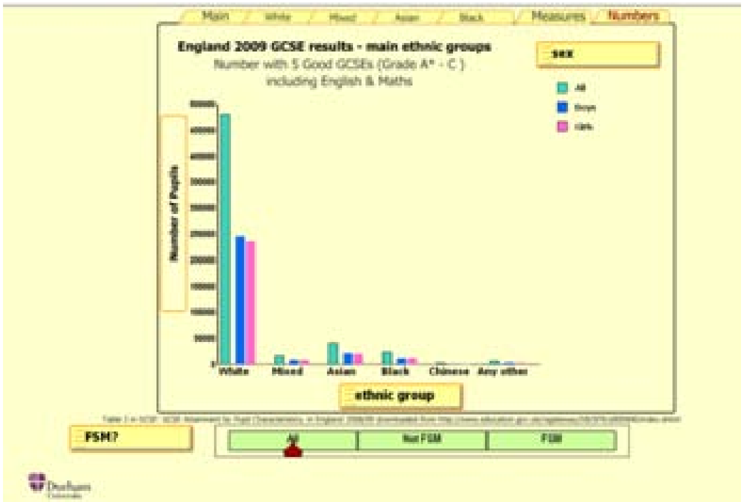


Figure 10: number of pupils in each of the classification groups

The last tab is labelled Numbers and it shows (see figure 10) the number of pupils in each of the classification groups – all the performance data is expressed in terms of proportions of the group achieving the government"s target level, but some of these groups have very small numbers in them. In figure 5, the statements made by the National Equality Panel refer to very few specific groups, but specific mention is made of the performance of Traveller children – only 119 of them in total and Gypsy – only 486 of them.

Sociology explicitly requires students to study research methods including strengths and limitations of quantitative approaches and of the data they generate. Apart from the contextual understanding we hope that viewing data in this way will be more accessible to students using these data interfaces we will be using survey based datasets as well as datasets such as these based on national official statistics. We will offer curriculum materials to illustrate and illuminate the principles in the analysis of strengths and limitations of data sources with reference to the datasets the students have been using. Such materials may be useful for other subjects with similar research methods requirements.

For example, in this dataset, the measurement criteria for the outcome is clear cut – has the pupil achieved 5 A* - C grades at GCSE, including English and Maths? One might expect that when the data is published by the Office for National Statistics it will not record errors in such a clear cut measure, but there will be pupils whose results are being appealed, and with several hundred thousand pupils each year it would be surprising if there were not occasional inaccuracies in recording any measurement. The categorisation as being eligible for free school meals has an option for „not known" into which a small minority of pupils fall – the categories FSM and not FSM are where the pupil is known to be eligible (whether they take them or not) or known not to eligible for free school meals. However, male / female is a strict dichotomy and some sociologists would argue that gender requires more complex categorisation than this. The ethnicity data are collected from parents when students first enrol at a school, with „missing" or „not known" recorded for only a small proportion, but the consistency of that self-reporting of ethnicity is an entirely reasonable question to raise. The data interfaces we produce to support Sociology and other social sciences will have the provenance clearly identified, and issues such as those above will be identified for discussion in class.

Quantitative reasoning in social sciences

The use of such interfaces opens up the possibility of developing curricula which have features of both quantitative and qualitative reasoning. Subjects like geography previously emphasised quantitative methods and required sophisticated statistical techniques because of the multivariate data involved. Current specifications across a range of curriculum areas have largely moved away from quantitative methods.

While we are not advocating a return to formal statistical techniques as a cornerstone of such courses, we believe that the lack of engagement with any quantitative aspects in these areas is unhelpful in the immediate understanding of the concepts under consideration, and, crucially, is a major contributing factor in inhibiting the development of skills in reasoning from evidence amongst those intending to work in social science areas. There is a further consideration: currently many good „numerate" students do not maintain interest in the social science disciplines because they don"t see any future in it for them and we believe that this represents a missed opportunity both for the students and for the social sciences community.

It is not just students for whom this is a problem: only a small proportion of teachers within the broad area of humanities and social sciences are really comfortable with using quantitative methods. In our classroom trials with interactive displays, teachers have commented that they feel more comfortable in mediating classroom discussion on informal inferences accessible through these interactive representations than working with the same data in tabular form.

One of our aspirations is to develop strategies in social science subjects which encourage and enhance students" capacity to reason with complex data. We intend to look at ways to extend students" activity beyond simple descriptive and inferential statistics which often relate to just two variables, to key statistical ideas such as effect size and interaction. We also want to develop ways to help students to understand what numerical data can tell us and what they do not say. We believe such skills are increasingly important for informed, educated future citizens, and especially for those who aspire to work in the areas of social science.

The relationships within the data in this interface are more complex than anything pupils have previously encountered, so building confidence in their ability to describe relationships is important – if they feel they can draw a sensible conclusion about something in the data, one test is whether the same conclusion holds when the display is reordered. Figure 11 shows the same data as in figure 7, just for White pupils, but with the position of the Sex and FSM? variables transposed so the comparison of performance of pupils eligible and not eligible for free school meals can be made even more directly than in figure 7, where the easiest comparison is between boys" and girls" performances.

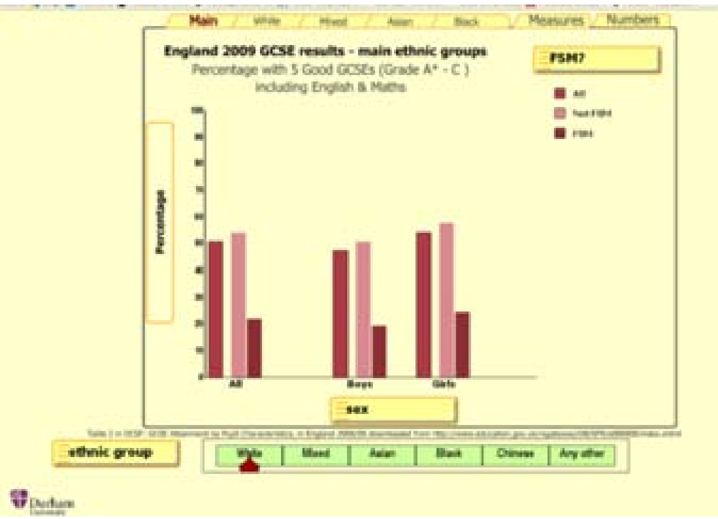


Figure 11: the performance of Chinese pupils with other variables transposed.

The data for almost all phenomena in social sciences involves multiple factors which have been hard to work with using traditional graphs and tables of data. Technology now offers the opportunity to visualise the relationships in complex data – in surprisingly accessible ways.

Discussion

There is widespread acceptance that developing understanding of a complex world in the citizens of today and of tomorrow is a worthwhile ambition. Organisations like OECD, WHO, UNESCO and the World bank as well as National Statistics Offices produce large amounts of readily available data. They, and others, are putting increasing amounts of efforts into communicating the stories behind the data (see, for example, Rosling (2010), Stenliden & Jern (2011), Riesch & Spiegelhalter (2011)) and emphasising the wider importance of engagement with the issues despite the complexity of data associated with them. The Royal Statistical Society launched the *getstats* 10 year campaign to improve statistical literacy in the UK on 20.10.2010. Gigerenzer (2010) identifies statistical illiteracy in health issues as both serious and systemic across most countries and argues that schools should teach the art of solving real-world problems instead of an emphasis on applying formulae to toy problems about coins and dice. Moreno (2010) argues convincingly that linking maths, science and social studies to formulate rich interesting questions would be better than isolating each into a curriculum „silo" where data analysis belongs in maths. Davies, Barnett & Marriott (2010), in reviewing 100 years of statistics teaching, lament that in the UK still textbooks and examinations pay scant regard to real data. McCusker, Nicholson & Ridgway (2010) provided examples of high stakes assessment where real world data was used inappropriately and others where the context was irrelevant or was unrealistic. Undoubtedly, there are serious issues that need to be addressed in the way statistics is taught within mathematics, but we also need to reach out into other subject areas and actively explore how we can support them in teaching their content through making complex data accessible.

We still know relatively little about the heuristics by which young people try to come to terms with understanding the stories behind data in the social sciences, and there is much to be done in finding ways that we can help them develop the reasoning skills that they need in order to work with this complexity. Stenliden & Jern (2011) report on the development of a teacher authoring tool which allows storytelling to use both standard data graphs and geographical data displays embedded within the story, with all the data accessible interactively. It will be critical to build up a library of stories at appropriate levels of complexity and for them to be accessible in ways that previous attempts: *Swivel* (now defunct) and *Many Eyes* have not managed to do – where a lot of material was created by public users, much of it trivial and often using inappropriate representations such as overlaying graphs with quite different quantities on the vertical axis, and where the capability to search for data was primitive at best. Nicholson, Ridgway & McCusker (2010) hypothesise that the quantum leap in working with increasing numbers of variables occurs quite early on: 1 and 2 variables can be easily represented graphically, moving to 3 or 4 variables is where the major conceptual hurdle occurs, and increasing to 5, 6 or more variables brings only an incremental increase in complexity. Yet even young children can cope with this level of complexity in data with visualisation tools embedding data of interest to them when it is in a context that they already have a vocabulary to use in describing relationships in the data. Almost without exception, students working with the SMART Centre visualisations focus exclusively on the information in context with any mention of anything identifiable as „mathematics" or „statistics" being rare; but they are engaged in relatively sophisticated statistical reasoning at times. Much remains to be done to understand the development of the reasoning processes, and in how visualising complex data at an early age might enhance formal statistical analysis at later stages.

REFERENCES:

Davies, N., Barnett, V. & Marriott, J. (2010) One hundred years of progress – Teaching statistics 1910 – 2010: What have we learned? Part I: It"s not mathematics but real data in context. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

Gigerenzer, G. (2010) Helping doctors and patients make sense of health statistics: towards an evidence-based society. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings* *of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

McCusker, S., Nicholson, J. & Ridgway, J (2010) Statistics Assessment: The Good, the Bad and the Ugly. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

Nicholson, J., McCusker, S. & Ridgway, J (2010) Luring non-quantitative majors into advanced statistical reasoning (and luring statistics educators into real statistics). In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

Riesch, H. &Spiegelhalter, D.J. 'Careless pork costs lives': risk stories from science to press release to media. *Health, Risk & Society* , **13**:47-64.

Rosling, H. (2010) What showbiz has to do with it. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

Stenliden, L. & Jern, M. (2011) “Visual Storytelling – Understanding and Knowledge in Education”, Awarded best paper, International Symposium on Academic Informing Science & Engineering, 2011 in the context of The 2nd International Multi-conference on Complexity, Informatics and Cybernetics: IMCIC 2011, Orlando, Florida, USA. ISBN-13: 978-1-936338-21-4.

Davies, N., Barnett, V. & Marriott, J. (2010) One hundred years of progress – Teaching statistics 1910 – 2010: What have we learned? Part I: It‟s not mathematics but real data in context. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

RÉSUMÉ (ABSTRACT)

Social sciences deal with complex phenomena where multiple factors are important. Increasingly there is access to disaggregated data where the effects of these factors and any interactions between them can be explored. However, because dealing with multivariate data has not been a realistic possibility until recently, much of the teaching in pre-college social sciences has been discursive, and based on headline statistics about aggregated data. There is also a tradition of ‘qualitative’ research, where data from large scale surveys is ignored, in preference to (for example) interview data from small samples. To overstate and oversimplify: there is a culture of theorising about causality without first understanding the phenomena in detail. Consider, for example, inequalities in educational attainment. It is widely reported that girls do better than boys, and that students in higher socio-economic groups do better than those from lower groups. However, if one is constructing theories about educational attainment, it is important to know if the differences between boys’ and girls’ attainment is: the same across socio-economic groups; in different regions; across all ethnic groups, and to explore changes over time.

Students studying social sciences at school in the UK rarely take advanced courses in mathematics. We suspect that statistical literacy levels among students (and perhaps teachers) are rather low. The use of visualisation tools facilitates data exploration, even at pre-college level. On a project funded by the Nuffield Foundation, we present multivariate data that is central to topics in a pre-college level Sociology course, via interactive displays. The paper will describe work with teachers and students on the development of their statistical literacy, and their understanding of key topics. The work has implications for both the definition of ‘statistical literacy’ and for approaches to fostering statistical literacy in a non-mathematically oriented student population, and in citizens in general.