Bringing Data Literacy to UK Secondary Schools

Daniel Gooch[1], Annika Wolff[1], Jose Cavero[1]

[1] The Open University, Department of Computing and Communications, Milton Keynes, UK
{daniel.gooch, annika.wolff, jose.cavero}@open.ac.uk

ABSTRACT As data becomes an important part of everyday life, the ability for the average citizen to have some level of data literacy is increasingly important. With the rise of smart cities, citizens need to be able to understand and process the vast amounts of data being generated about them, their community and the city they live within. This paper describes an approach to teaching data skills in secondary schools using real life, complex, data sets collected as part of a smart city project. An inquiry-based methodology is applied to frame the activities. This paper describes the initial trial in a UK secondary school in which seventeen students aged 13-14 learnt about home energy consumption and the generation of solar energy from home solar PV, using smart meter data and satellite collected data that are currently part of smart city research. We discuss the appropriateness of our approach and what it means for developing web-based platforms designed to support teaching data skills in school.

Keywords: Urban Data · Data Literacy · Smart Cities.

1 Introduction

The term data literacy generally refers to the ability to consume for knowledge, produce coherently and think critically about data [2]. A large number of skills are involved in data literacy, including identifying sources of data, asking coherent questions of the data, analysing and interpreting data and using data as evidence to support an argument. What distinguishes data literacy from statistical literacy is the embedding of data in a context through which the data is interpreted, and the use of data to solve real problems [5].

Data literacy is becoming increasingly important as a necessary life skill as increasingly large datasets become commonplace. Data skills learnt against smaller, more traditionally sized datasets – such as those currently taught in United Kingdom (UK) secondary schools - cannot be easily extended and applied to larger, more abstract data. In fact a whole new set of technologies and techniques must be learned [4]. A further problem in the UK is that data analysis forms part of teaching across many subjects within the national curriculum, but does not exist as a subject in its own right.

In this paper we demonstrate a teaching approach to raise levels of data literacy in young learners by describing work in which real-life complex data sets were successfully used in a secondary school classroom as a teaching resource, to teach data literacy in the context of a task looking at energy consumption. We identify some of the barriers to teaching data literacy using complex data sets in the classroom and some approaches to overcome these problems. Our experience suggests that with the correct support, these complex skills are accessible to secondary school students.

2 An Approach to Teaching Data Literacy in the Classroom

As large, complex, data has not typically been available as a teaching resource, there is not yet a clear research base to assess the best approach to using them as a teaching resource in schools. We have had to draw upon existing learning methodologies to inform the initial design phase. Inquiry-based learning is a constructivist approach in which the learner is supported to take an active role in designing and running their own experiments [1]. An inquiry-based task generally follows a cycle of asking questions, hypothesis generation and experimentation, analysis of evidence, question answering and reflection. We have proposed a modification to a PPDAC cycle used for teaching statistical data analysis in New Zealand. This follows a cycle of **P**roblem, **P**lan, **D**ata collection, **A**nalysis and **C**onclusion, normally starting at the Problem specification [6]. We suggest that when using the cycle with abstract, non personally collected data, the starting point should be the analysis of a carefully chosen snap-shot of the dataset, leading to further questions and more in depth analysis across the whole set. Preliminary analysis of this approach has taken place in a primary school [7].

2.1 The Data

The data sets that we are using have been collected as part of the MK:Smart smart city project. Two data sets were chosen which were both related to the topic of energy. One data set is focused on smart meter energy data detailing home energy consumption of both the whole house and individual appliances across a period of time. This data set also includes solar PV generation data from the same homes in Milton Keynes. A second dataset contains satellite collected data that can be used to estimate the solar energy yield from existing or potential solar panels on the roofs of houses in Milton Keynes and which contains information relating to the cost of installation and expected return over different time periods. The examples we will use in this paper are based on real data, but for the purposes of publication have been completely anonymised. Both data sets are from external city-level infrastructure systems which, in the future, could be used in the classroom in real-time.

2.2 Teaching Approach

Our approach takes the interpretation of an existing visualisation as the first activity and for this to act as a seed for defining a new problem. We then use the idea of ‘telling stories’ from the data to help interpretation by encouraging learners to identify the setting (when and where) of the data, the main theme of the visualisation and the actors (who may or may not be a person, depending on the dataset). Learners then continue to formulate questions around this same data set. These questions are used to structure queries that will retrieve some part of the large data set that may then be used as evidence in answering the questions the students have developed.

This approach is designed to overcome what we foresee as a ‘cold start’ problem, where learners could find it difficult to define questions that can be answered from a large data set without first forming some view of the data by interpreting existing visualisations. This view is supported by the teachers who have assisted in developing the approach.

One possible problem is that students might find it hard to understand how the data was collected, which can make it harder to interpret [3]. We make a further suggestion that learning might be supported through a personal data collection task that is related to the original data inquiry. Essentially, students first work with an existing large data set, then undertake a second phase where they conduct an alternative but related inquiry from their personally collected data. The purpose is to consolidate the learning of data skills and also to provide additional narrative and context for understanding the larger, more abstract, data.

2.3 The activities

A set of activities have been devised from the selected data sets for use in classrooms. The first activity uses the smart meter data set to pose questions around how and when energy is consumed in an average household and when solar panels on a roof are producing energy. The second activity uses the satellite collected data to pose questions related to whether or not properties are suitable for having solar panels or not. The following section describes the process through which these activities were used in a secondary school classroom.

3 CLASSROOM TRIAL

The main aim of the trial was to elicit initial feedback on the developed materials and to gain an understanding of how teachers would use the resources in a real classroom setting. To allow the teacher full flexibility in choosing how to use the content, it was decided that evaluation should be through a mixture of classroom observation followed by more structured feedback by teachers and students at the end of the session. The focus of the evaluation is on a) how well students were able to interpret visualisations of complex data sets b) how easily the materials could be incorporated into regular classroom teaching c) the students engagement with and learning about the chosen topic ‘energy’. In this paper we are only reporting on the results of our classroom observations.

The classroom trial was conducted over two science classes at a publicly-funded secondary school with seventeen Year 9 children (aged 13-14). The classroom sessions were led by the teacher with two researchers present to observe and help as necessary.

3.1 Session 1

The first session incorporated both energy consumption and solar generation data. At the beginning of the session, students took part in a whole-class discussion of a graph that plotted overall electricity consumption of a single house across a single day. Following on from this, students had to:

1. Write a story hypothesising what activities might have been occurring on that day to explain the pattern shown in the graph
2. Create 5 questions related to the data set that could test the story hypotheses
3. Select two of these questions and construct semi-formal queries to extract data which could help the students answer that question

The students had to first work on these questions in groups with regards to the electricity consumption data. Students easily identified the salient points from the graphs including the maximum and minimum times of consumption and the pattern of the graph. They were able to come up with a wide variety of theories from the consumption graph, for example hypothesising that the data might be from a weekend because there was a large amount of consumption during the day, demonstrating a good understanding of what could cause electricity consumption to fluctuate.

At this point the teacher initiated a whole-class discussion that went beyond the teaching materials. Focussing on a period of zero electricity use, she prompted the students to consider whether it was it a power cut or whether there were alternative explanations? Students ventured that it might have been a fuse-box failure or that a pre-paid meter had run out of cash. Interestingly, no student raised the possibility that the data could simply be incorrect.

A wide range of questions were generated by the students including “Did the entire street have a power cut?”, “Does the house use the same amount of energy on weekends and week days?” and “How does the energy use change in winter?”. A number of questions went beyond surface-level interpretations. For example, “Do they have young children in the house?” – the students argued that by looking at microwave and kettle use that occurred overnight they could determine whether there was a baby that needed feeding. Such deep interpretations of the data were completely unanticipated. All 17 students managed to construct at least one semi-formal query that would provide data that could help answer the question they had posed.

Having worked through the electricity consumption data with teacher intervention, students had to work independently through the same set of questions using solar generation data for a single property in Milton Keynes over a 5-day period in June.

The solar data is less fine grain and has a clear repeating pattern of increasing during the morning, peaking and then decreasing across the afternoon to a trough at night. The nature of the data meant that while students were capable of generating a narrative and questions regarding the data, the queries were all based on two themes – either comparing across different houses (e.g. “did all the houses produce a similar amount of electricity?”) or looking across a wider time-span (e.g. “does the house hold the same pattern in the winter?”). All 17 students managed to construct at least one semi-formal query that would provide data that could help answer the question they had posed.

Informal post-session feedback from the teacher highlighted that the session had gone really well; she was surprised that one student in particular who wasn't normally engaged was very engaged and encouraging other students. She also highlighted that through providing real-world local data that was not accessible to the school we were providing a very useful resource that made the session more interesting for students.

3.2 Session 2

This session started with students making their own graphs from electricity consumption data based on their queries from session 1. This was done using Excel and a set of prepared data that had been curated based on the queries students had written. After 20 minutes students went through the same process using the solar generation data. Having gone through this process they were directed to explore an online tool that visualised 12 months of solar generation and electricity consumption data. Each plotted point was an aggregate of that days data. Students could select the house, type of data (consumption/generation) and date range.

When comparing the two approaches, students preferred the online tool because "it's easier" and "you can get more analysis done" but noted that it was limited in terms of forming more complex visualisations around plotting multiple houses on the same graph. Students queried the data much more rapidly using the online tool and constructed a wider array of questions beyond the scope of the immediate task (e.g. "Why are there consistent periods of no electricity consumption?"). This led to more in-depth interpretations (e.g. solar panels switching on when moving from the national grid vs. running out of money on a key meter).

The next task involved interpreting satellite images of an estate in Milton Keynes with an overlaid data set that predicted the yield for producing solar energy in 1 year based on roof size, pitch and direction of roofs as calculated from the aerial photography. Each aerial map was overlaid with yield data and was provided alongside a raw data table that contained several other attributes, including cost of installing the panel, and the estimated financial cost benefits. Students were asked to use the data to discover the relationship between roof area and yield and to understand how other factors such as pitch and direction influenced the yield. The students successfully identified which houses had the best yield and the relationship this house had with roof area, pitch and direction. One pair of students went further and noted that while one house had the highest yield, the up-front cost for the panel was much higher than a house that had a similar yield. This led to a discussion of how efficiency varies depending on the time period you are looking at.

The students had also been given homework to 'be a smart meter' and had been keeping data diaries during the week between sessions where they recorded what appliances were being used and how long for. The final task was for the students to design their own visualisation of the data they had collected for their homework. All of the students managed to construct a visualisation that accurately represented the data they had collected.

Overall, across the session students demonstrated a very good ability to understand the data, create their own visualisations, interpret various visualisations and use their interpretation to generate further questions.

4 DISCUSSION, CONCLUSIONS and FUTURE WORK

Our approach, enacted through a mix of teacher-led, interactive and independent work was very successful. Our students demonstrated that they were capable of achieving the tasks presented to them, with the level of help that would be normally expected in a classroom. They demonstrated an ability to conduct their own queries, create visualisations and make accurate interpretations using unmodified complex urban datasets. Therefore, the outcome of this trial might be taken as an encouraging indication of the capabilities of students

From our experiences it is clear that providing data with a local angle provides a great deal of interest to students. Using real-world datasets was considered to be interesting by both the teacher and students as it is a resource they do not typically use and answers the common student question of ‘why do I need to know this?’.

This trial is informing the development of a web-based platform called the Urban Data School which will provide teachers with access to big data sets and the lesson plans, tutorials and other resources required to use them in the classroom. We suggest that if, as we have demonstrated, young learners are capable of working with large data sets, this provides evidence for the argument to create a provision within the UK curriculum for this topic so that future generations will leave school with better levels of data literacy.

5 REFERENCES

1. Dewey, J. 1933. How we think: a restatement of the relation of reflective thinking to the educative process. Heath: Boston, MA
2. Gray, J., Chambers, L., and Bounegru, L., 2012. *The data journalism handbook*. O'Reilly Media, Inc.
3. Kastens, K. and Turrin, M. 2011. Geoscience data puzzles: developing students' ability to make meaning from data, *ED11C-04*, San Francisco, US
4. Maycotte, H. Data literacy - What it is and why none of us have it. 2014. Retrieved May 8, 2015 from Forbes: http://www.forbes.com/sites/homaycotte/2014/10/28/data-literacy-what-it-is-and-why-none-of-us-have-it/
5. Vahey, P., Yarnall, L., Patton, C., Zalles, D. and Swan, K. 2006. Mathematizing Middle School: Results From a Cross-Disciplinary Study of Data Literacy. *American Educational Research Association Annual Conference*, San Francisco, CA
6. Wild, C., and Pfannkuch, M. 1999. Statistical thinking in empirical enquiry (with discussion). *International Statistical Review*, 67(3), 223-265
7. Wolff, A., Kortuem, G., and Cavero, J. (2015) Urban Data in the primary classroom: bringing data literacy to the UK curriculum. Data Literacy Workshop, WebSci 15.