WORKING WITH A COMMUNITY ORGANISATION TO CREATE OPEN DATASETS FROM RECORDS AND FILES

Ian Hay and Rosemary Callingham
University of Tasmania, Australia
Ian.Hay@utas.edu.au

Many community organisations maintain historical and other records and files that are potentially of interest to the wider community, including schools. They may lack the expertise, however, to make this knowledge available in suitable statistical formats. In this project a local group researched the history of 165 World War 1 servicemen. In addition to a published book, the data were used to create a spreadsheet that could be used by schools, and a website accessible to the general public. A local school investigated the data using Tinkerplots. Students were highly engaged and motivated by a real dataset that they could relate to through family connections. This case study paper provides suggestions for community organisations and statistics educators about working together for successful outcomes.

INTRODUCTION

Statistical literacy is regarded as an essential component in any educational program designed to prepare students and adults for decision-making in the 21st Century (Gal, 2004; Wallman, 1993). There is the claim that statistical literacy needs to be better taught and students gain more from their statistics courses, if they understand the content, and have teachers who actively link the statistics to students' own experiences and interests (English & Watson, 2015; Shaughnessy, Chance, & Kranendonk, 2009). One to way to facilitate students' interest is to use data sets that are relevant and interesting to students (Hay, Callingham, & Carmichael, 2015; Watson & Neal, 2012). This paper provides a case study about generating such a data set.

Background to the community project

Schools are located within communities and within each community there are data sets that have the potential to be developed so that they can be incorporated into the school program. Ideally statistical literacy should be embedded across all school subjects and so there is a need for data sets that have relevance within the different curriculums. This case study considers the development of a community-based data set based on the centenary of the First World War (1914 to 1918). This war had a profound effect on Australia, which had only become a federated country in 1901. There were large numbers of young men volunteering to go to war and very high casualty rates. The national grief at the loss of so many men resulted in a proliferation of war memorials at the end of the war in every town across Australia to record those who fought and died. The events of the war are widely regarded as pivotal in terms of Australia identity (Fitzsimons, 2015) and every year there is a public holiday to mark ANZAC (Australian New Zealand Army Corps) Day, which commemorates the first landing of the Australian and New Zealand troops at Gallipoli. The First World War is a required topic of study in the primary and secondary schools' history curriculum. There are, however, few relevant data sets that have been designed for students that relate to the First World War. To mark the centenary of the First World War, the Australian Government set up a competitive community fund for local initiatives to which organisations could apply. A church in Launceston, Tasmania, Australia decided to seek funds to investigate the names of 165 Northern Tasmanians whose names were listed on a First World War memorial board within the church. This provided an opportunity to work with a community group to facilitate the development of a relevant statistical literacy resource of Australians who fought in the First World War.

THE PROJECT TAKES FORM

A project committee consisted of members of the church interested in the project, including the two authors associated with the University. The funding application was developed and the University experience helped the committee to identify the steps in the overall project, outline a realistic timeline and budget, and identify the deliverables. The aim of developing a data set based on the 165 names that could be provided to schools via a website, along with a book, was a component of the application that had not been considered until suggested by the authors. The

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

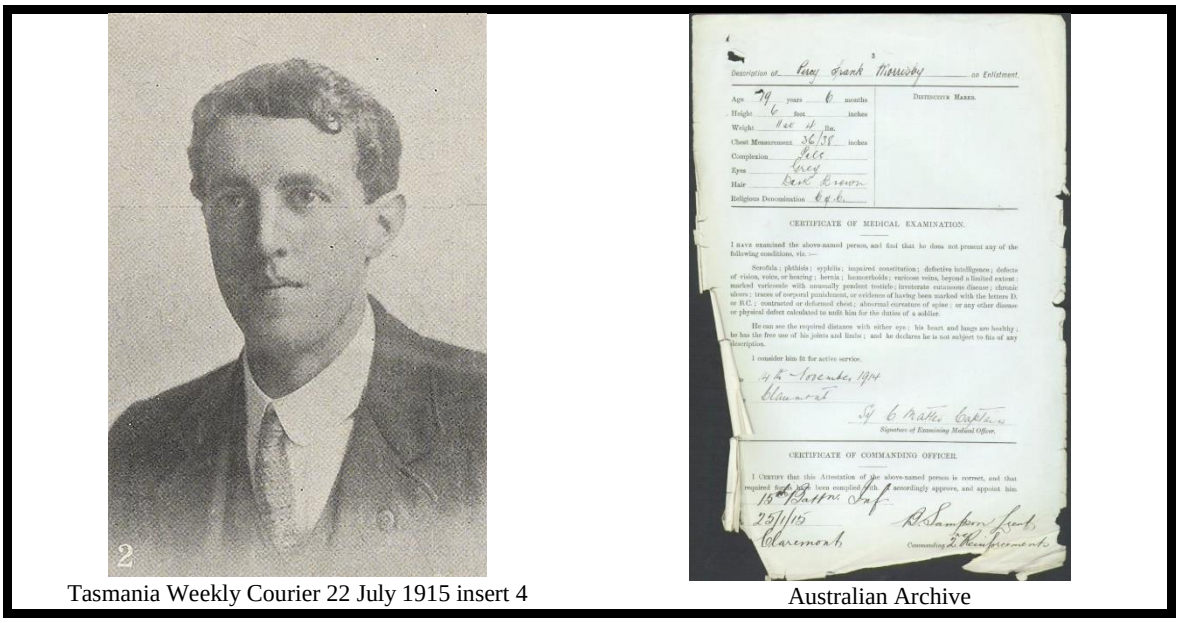
funding application was modest as it was recognised that much of the researching and development of the data set and associated information would be conducted by unpaid volunteers.

Identifying the records

The first task was to gain an understanding of the availability of records associated with each of the men listed on the memorial. This process of data mining for information is dependent on the quality of accessible historical records. In Australia, First World War records have been digitised and are available electronically through the Australian War Memorial. The digitised resource consists of some 420,000 100-year-old records that were hand written and completed by many different people in diverse locations. Hence the quality and quantity of information varied. The records can be searched by name. Although this may seem like a straightforward process, often there were several people with similar names, and identifying the correct serviceman required additional research using next of kin and parish records. The names on the memorial did not always match the name on the records. Substitution names were common, such as Jack for John, or there were different spellings of names on the memorial compared with those in the records. For example, the name Algerian on the memorial board was actually Algernon in the records. Many names on the memorial were of men associated with Tasmania and the local community but who had moved to other places. The service records of those who enlisted in Australia could be relatively easily located, but there were some who served with the armed forces of New Zealand and South Africa, and one who signed up in Fiji. The memorial was also inaccurate, with names included twice, and the year of enlistment often incorrect. The volunteers undertaking the task of identifying the correct records spent hundreds of hours searching different data sources, and were also well supported by the local library. These challenges are important to acknowledge and recognise.

Working from a record to a spreadsheet

Working with original and historical records is not a typical activity of statisticians. Similarly, those interested in history are unfamiliar with the needs of statisticians. Statistics educators at the school level also have unique needs, requiring a data set that is of intrinsic interest to students but that has sufficient information to be able to develop statistical ideas. The idea of selecting suitable variables from the information available required specialist input to identify potentially interesting but also usable data. The enlistment records usually included information such as age, height and weight, eye, hair and complexion colour, and next of kin. They also listed the various battalions and places to which the servicemen were posted, details of injuries and medical conditions, and the ships in which they travelled. Although all of this is interesting in itself, it may be of limited statistical use. Figure 1 shows the photograph and part of the records of one of the names listed on the memorial.



Tasmania Weekly Courier 22 July 1915 insert 4 Australian Archive

Figure 1. Photograph sourced from local library records and part of the digitised record for Percy F. Morrisby.

From a number of national records a profile was created. Figure 2 is an example of part of the summary profile for Percy Morrisby, produced from the records. Percy Morrisby was one of the many men who were killed in action or died of their wounds.

1895	07 May	Born Hobart. Parents Percy Inkerman Morrisby and Harriet nee Scott. 1st of 3 children
1912		Completed Hobart Technical School. Results: gained 1st class in mechanical drawing.
1914	10 December	Enlisted in Claremont as Private in the 15th Battalion, infantry
		Regimental number 1625
	Enlistment Details:	
		Age: 19 yrs 6 mths Height: 6 ft Weight: 11 st 4 lbs Chest measurement: 36 – 38 ins Complexion: pale; grey eyes; dark brown hair Religious denomination: Church of England Trade or calling: Engineer Not married Former service/Cadets: N/A
	Next of kin	Father, P I Morrisby, Bryan Street, Invermay, Launceston
1915	2 February	Embarked from Melbourne on *HMAT Clan McGillivray* A46
	13 April	Proceeded to join the MEF
	8 May	Killed in action in Gallipoli aged 20

Figure 2. Summary record for Percy F. Morrisby.

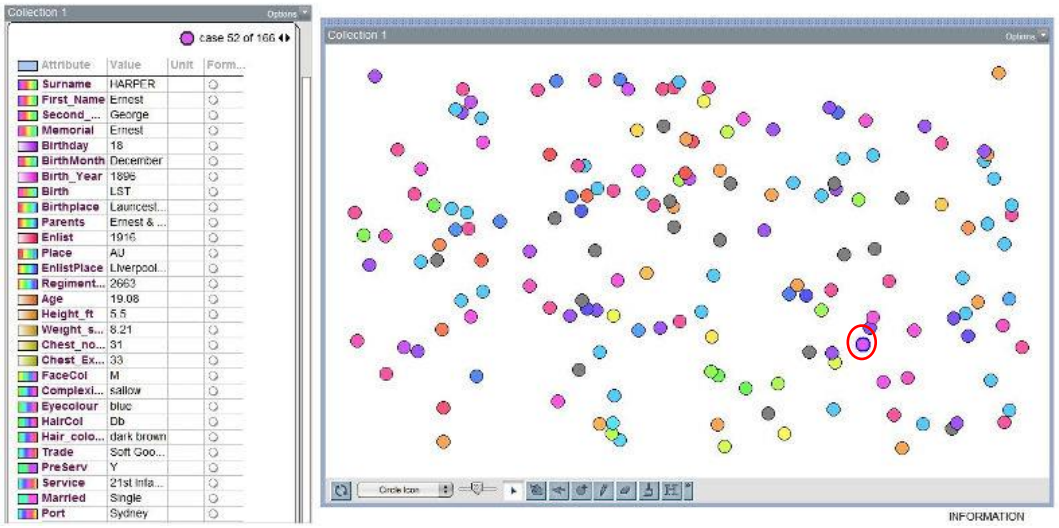
From information sheet to a data file

The data obtained for each of the names on the memorial board were entered into an Excel spreadsheet. Decisions had to be made about which fields to include. Eventually 26 fields were covered, including height and weight, hair and eye colour, date of birth, year and place of enlistment, killed in action and died of wounds, date of return to Australia, and military honours and awards. These fields were chosen partly because of their intrinsic interest for students but also because they could be used to develop statistical ideas. As can be seen in Figure 2, however, these were not yet in a condition to be easily accessed by Australian school students. Height and weight, for example, were in Imperial measures (feet and inches; stones and pounds)—completely unknown to a generation of students brought up with metric units. Hair colour, complexion, and eye colour had no standard recording. For example, eye colour was recorded as blue and brown but also light grey and yellowish! In addition, because volunteers entered the data, there were inconsistencies in capitalisation and date records, spelling, and inclusion of more than one name in a single cell. All of these matters had to be dealt with to prepare the file for use by students. The challenge was to create a data file that had the 'feel' of the original data but was useful for students to explore. Rather than change all the Imperial measures to metric it was decided to decimalise the original units (e.g., 5 foot 6 inches became 5.50 feet). Categories were also created for hair, eye and skin colour, and for place of birth and enlistment. The original forms of the data were retained as separate fields, however, because that provided a potential discussion topic for teachers about the nature of useful statistical data.

As well as a standard Excel spreadsheet the file was also saved as a Tinkerplots (Konold & Miller, 2005) file (www.tinkerplots.com). Tinkerplots is a data visualisation tool designed for students from Year (Grade) 4 upwards. Using this tool gives students the opportunity to investigate and display a range of issues of interest to themselves by working with the data file.

TRIALING THE DATA FILE WITH STUDENTS

A local high school agreed to trial the data set in a Year 8 mathematics class. Because of the intuitive nature of the Tinkerplots interface, the school was provided with a class set of licences for Tinkerplots that the students downloaded to their laptops, along with the 165 ANZACs data file. Starter activities were provided for the students to familarise them with both the program and the data file. An early activity was "Get to Know an ANZAC". Students were asked to choose a serviceman at random by clicking on one of the dots. They then created a story about their ANZAC. An example of one of these is shown in Figure 3. The aim of the activity was to familiarise students with the variables.



Ernest George Harper was born on 18th December 1896 in Launceston. He enlisted in 1916 in Liverpool, New South Wales. He was just over 19 when he signed up. He was 5.5 feet tall, and his weight was 8.21 stone. He had a sallow complexion with blue eyes and dark brown hair. He was killed in action.

Figure 3. Get to Know an ANZAC activity.

Initially the students focused on the descriptive information such as complexion and name. They did not go beyond the obvious. Other activities were intended to take them outside these limitations. For example they were asked to find an ANZAC who was tall and thin, and to justify their choice. To do this they had to associate the height and weight variables, as shown in Figure 4. Some students simply chose the tallest man (marked A in the Figure). Others were confused about height and weight and chose the heaviest man (marked B in the Figure). The better responses chose one of the taller ANZACs, such as C in Figure 4 making statements like "he is tall but a bit underweight". None of these Year 8 students thought to sketch a line of best fit or calculate an average height and weight in order to justify their choice.

Over two lessons the students became more familiar with the data set and with Tinkerplots. They began to explore questions of interest to them, such as frequency of surnames. Because the data set was local from an Australian State with a very stable population, many students recognised some of the surnames as belonging to friends. They also noted that some names occurred more frequently, because brothers and cousins tended to enroll together. This kind of activity brought the data set alive, with one girl in the class asking, "Are these real people?"

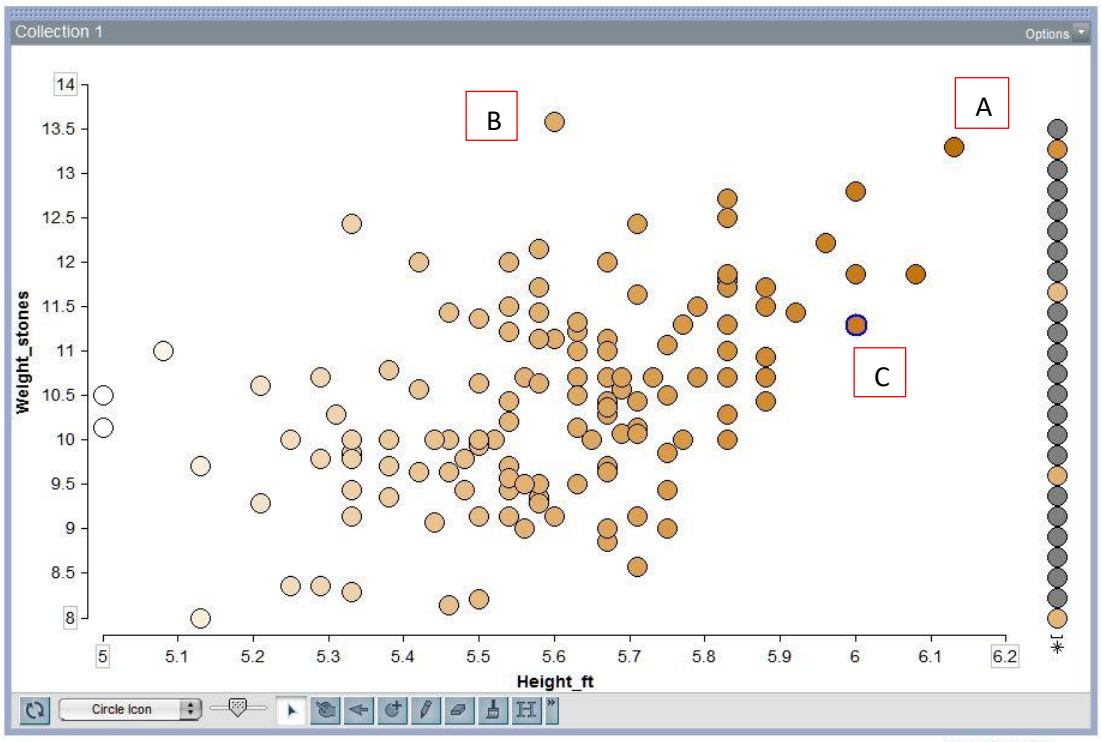


Figure 4. Association of height and weight of the 164 ANZACs.

The data set is sufficiently complex to allow for the introduction of a number of statistical ideas. For example, the mean age on enlistment was 22.9 years but the median age was 24.7 years, providing an opportunity to discuss the different meanings of these measures, and the idea of a skewed data set. Association can be introduced by considering the ratio of height to weight (see Figure 4) or through the chest measurements, which were both normal and expanded. Handling categorical data can be discussed by considering the eye, hair and complexion colour. At the same time, the size of the data set, at 165 entries, is not too overwhelming for students.

COMMUNITY ENGAGEMENT AND STATISTICAL LITERACY

There are several important ideas that arise from this small, local project.

- Specialist input

Projects of this nature are important and doable but they require specialist input. The community volunteers were willing but could not initially go beyond creating profiles of the servicemen. The authors' and others' specialist knowledge aided the grant application, the decisions about the format of the data, the nature of the useful variables for inclusion, and the publication of the final product, a book (Box et al., 2017) and a website (www.165anzacs.org). The undertaking was a genuine joint venture drawing on a range of skills.

- Students' interest

Once the students realised that these were real soldiers, they became more deeply involved with analysing the data available. The Tinkerplots program helped students to visualise the data and the intuitive interface assisted the students' explorations of the data. The students reported that working with data about real people was an aspect that they found engaging. This is not a new finding as research by Fizallen and Watson (2010) identified a similar response from students. Having students work with meaningful data sets enhances students' interest, engagement and competence in statistical literacy (Hay et al., 2015). Teachers have also identified that having access to meaningful data sets was an important aspect in engaging students in statistical literacy activities (Hay, 2010). Because the 165 ANZACs data set is authentic and diverse, it provides teachers with the opportunity to use and revisit the data set at different stages of the students' developing reasoning in statistical literacy (Watson & Callingham, 2017).

- Community involvement

Typically, there is a wealth of knowledge and expertise available in the community but support was needed for those who wanted to participate. In this project, for example, people used high level research skills to track down the correct set of data for each serviceman, and photographs for many of them. Given the state of the records, this was a considerable effort. Without specialist input, however, the project might not have produced the rich outcomes that were achieved. In addition to the local project team, many others became interested in the information. Specialist input built on existing skills, such as building the website where the person involved was familiar with placing information on an existing website, but did not have the skills to build a new website from scratch. The specialists became an integral part of the team rather than being seen as outsiders. The funds that were available, were mainly used to buy in web design services and to have the book produced by a reputable printer, for its distribution to schools and libraries.

CONCLUSION

This kind of small, local project could be replicated anywhere. The importance goes beyond the immediate product to capture a slice of local history. Using specialists who became part of the team was productive and ensured that the project achieved far more than its original modest aims. The final data set has proved to be engaging for students and useful for teachers.

REFERENCES

Box, K., Callingham, R., Hay, I., MacDonald, J., Mann, D., & Mann, M. (2017). *Telling Part of the Story of 165 Northern Tasmanian World War 1 ANZACs.* Launceston: Holy Trinity Anglican Church.

English, L.D. & Watson, J, (2015). Exploring variation in measurement as a foundation for statistical thinking in the elementary school. *International Journal of STEM Education***,** *2*(3) 1-20.

Fitzallen, N. & Watson, J. (2010). Developing statistical reasoning facilitated by TinkerPlots. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society* (Proceedings of the 8th International Conference on Teaching Statistics, Ljubljana, Slovenia).

Fitzsimons, P. (2015). *Gallipoli*. North Sydney, NSW: Random House.

Gal, I. (2004). Adults' statistical literacy: Meanings, components, responsibilities. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 47-78). Dordrecht, The Netherlands: Kluwer.

Hay, I. (2010). Teachers' perceptions of best practice in statistical literacy education. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society* (Proceedings of the 8th International Conference on Teaching Statistics, Ljubljana, Slovenia).

Hay, I., Callingham, R., & Carmichael, C. (2015). Interest, self-efficacy, and academic achievement in a statistics lesson. In, K.A., Renniger, M. Nieswandt, & S. Hidi (Eds.), *Interest in mathematics and science learning* (pp. 173-188). American Educational Research Association: Washington DC.

Konold, C., & Miller, C.D. (2005). *TinkerPlots: Dynamic data exploration*. [Computer software] Emeryville, CA: Key Curriculum Press.

Shaughnessy, J.M., Chance, B., & Kranendonk, H. (2009), *Focus in high school mathematics: Reasoning and sense making in statistics and probability*. Reston, VA: NCTM.

Wallman, K.K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association, 88*(421), 1-8.

Watson, J.M. & Callingham, R. (2017). Developing learning progressions to support mathematical reasoning in the middle years – statistical reasoning. In A. Downton, S. Livy, & J. Hall (Eds.), *40 years on: We are still learning!* :*Proceedings of the 40th annual conference of the Mathematics Education Research Group of Australasia* (pp. 663-666). Melbourne: MERGA.

Watson, J., & Neal, D. (2012). Preparing students for decision-making in the 21st century – statistics and probability in the Australian curriculum: Mathematics. In B. Atweh, M. Goos, R. Jorgensen & D. Siemon, (Eds.). (2012). E*ngaging the Australian National Curriculum: Mathematics – Perspectives from the Field* (pp. 89-113). Online Publication: Mathematics Education Research Group of Australasia.