LEARNING ANALYTICS: A CASE STUDY OF THE PROCESS OF DESIGN OF VISUALIZATIONS

Martin Olmos

Linda Corrin

Graduate School of Medicine, University of Wollongong, Australia

ABSTRACT

The ability to visualize student engagement and experience data provides valuable opportunities for learning support and curriculum design. With the rise of the use of learning analytics to provide "actionable intelligence" [1] on students' learning, the challenge is to create visualizations of the data, which are clear and useful to the intended audience. This process of finding the best way to visually represent data is often iterative, with many different designs being trialled before the final design is settled upon. This paper presents a case study of the process of refining a visualization of students' learning experience data. In this case the aim was to create a visual representation of the continuity of care students were exposed to during a longitudinal placement as part of a medical degree. The process of visualization refinement is outlined as well as the lessons learned along the way.

KEYWORDS

learning analytics, medical education, visualizations

I. INTRODUCTION

Producing a visualization from a given dataset which brings light to specific questions involves a difficult process of design. As the field of learning analytics is still in its relative infancy, a greater understanding of the challenges and opportunities afforded by the creation of visualizations of complex datasets is needed in order to help develop higher analytical skill sets. Whilst there have been helpful case studies published where a visual designer walks through the process of design [2], more examples specific to the learning analytics field are needed. This paper describes such a design process in the learning analytics context. A case study of the visualization of students' patient care experiences in a medical degree is used to highlight the challenges and lessons learned in the iterative design and development process employed to visualize this data.

II. CONTINUITY IN THE MEDICAL CURRICULUM

The Graduate School of Medicine at the University of Wollongong was established in 2006 to help address the shortage of medical practitioners in regional, rural and remote areas of Australia. An innovative element of the four-year graduate MBBS degree is a longitudinal integrated community-based clerkship that students start during their third year. Students spend 12 months in a regional or rural setting participating in general practice and emergency department placements with additional education sessions run by regional academic leaders and other specialist clinicians. The design of this approach was informed by communities of practice theory and modelled on the Parallel Rural Community Curriculum Model developed at Flinders University in Australia [3,4].

One of the most important elements of the longitudinal placement design is the exposure students have to continuity of patient care. This involves the care of patients over time who present with a variety of clinical presentations as well as patients who present multiple times with the same presentation (e.g.,

pregnancy, chronic disease care, etc.). Repeated experience with patients allows students to build a relationship with the patient giving insight into their social context and other health factors that will inform the care given. The students can follow the care of the patient through various clinical settings (e.g., GP, hospital, specialist) and observe how a patient's diagnosis evolves. Dealing with a patient over time also gives the student an opportunity to see the effect of their prescribed treatment plan and to oversee the ongoing management of their treatment to obtain the best outcome for the patient.

Research at Harvard Medical School [5] and University of Alberta [6] found that students benefited greatly from longitudinal exposure to patients resulting in positive learning outcomes and a greater sense of professional identity. Supervising doctors involved in the GSM programme also recognized the authentic learning benefits of this approach, not only from a patient care perspective, but also from the view of making the student feel at home in the community, which will hopefully influence their decision to practice medicine in regional or rural locations in the future [3].

III. RECORDING CLINICAL EXPERIENCES

Students are required to keep a record of their patient experiences whilst on placements throughout their entire degree programme. To facilitate this process the GSM designed the Clinical Log, an online system that can be accessed on any web-enabled device. The Clinical Log gives students the ability to monitor their clinical involvement to ensure they are meeting curriculum requirements through exposure to an appropriate range of patient presentations. Students' clinical experiences are also monitored by the GSM to determine if interventions are necessary to ensure students get adequate coverage of the curriculum [7,8]. Correlating the Clinical Log data with data from other systems, such as the Placements Management System, allows the school to identify areas in need of development and helps to monitor equivalence of experience across the various geographical locations. Engagement with the Clinical Log is formally assessed as part of the Objective Structured Clinical Examination (OSCE), which the students complete at the end of the second and third phase of their degree.

For each patient consultation the students input the details of the case (avoiding any identifiable information) and map the patient's condition to a set of 93 core clinical presentations, which form the basis of the curriculum. Students are also encouraged to reflect on their learning needs and strategies to address these learning needs in relation to each patient experience. This feature allows them to revisit their identified learning needs at a later time to plan their study revision schedule.

For each Clinical Log entry there is a patient number field. Students are encouraged to enter in a code for each patient which can assist them to identify patients that they see more than once, but which does not relate to any identifiable information (i.e. Medicare number, phone number, etc.) to ensure patient confidentiality. This field allows the student to track all the consultations they have with a particular patient and provides the major data element for generating continuity reports.

IV. VISUALIZING CONTINUITY

The Community-Based Health Education team, who coordinate the longitudinal placement, sought a way to monitor the continuity of patient care experienced by the students on their placement. The audience for such reporting was placement facilitators and teachers as well as the individual students. A learning analytics approach was employed to generate visualizations of patterns and identify areas in need of improvement in regards to ongoing patient care. Learning analytics has been implemented at several points throughout the medical degree to aid in curriculum development and to ensure coverage of learning outcomes and clinical presentations whilst students are on placements throughout the four years [7].

Whilst the area of learning analytics is still relatively new, it is experiencing a rapid growth in interest and implementation, especially in the higher education arena. The 2012 NMC Horizon Report forecasted that the time to adoption of learning analytics is only two to three years and this timeframe has been boosted by the creation of a number of large-scale projects, learning analytics professional societies and conferences [9]. Unlike academic analytics which seeks to provide analytics at an institutional or national

level, the level of analysis of learning analytics is at the course or faculty level [10]. Learning analytics creates the potential to optimize student learning experiences and outcomes through monitoring students' experiences/performance within their degree courses. The definition of learning analytics used in this case study is: "the measurement, collection, analysis and reporting of data about learning and their contexts, for purposes of understanding and optimising learning and the environment in which it occurs" [11]. In recent studies learning analytics have been used to observe enrollment trends, measure student retention, identify 'at-risk' students, and visualize social network connections, to name only a few of the potential applications [12].

As the field of learning analytics has evolved, an emphasis on action analytics has emerged. Action analytics refers to the process of moving beyond the simple reporting of students' learning progress towards developing and implementing defined actions [13]. In particular, some forms of learning analytics provide the ability to generate data reports in real-time, minimising the time delay between when the data is captured and when actions can be implemented [14]. In the case study presented in this paper, the timeliness of reporting is crucial to the development of interventions to improve the students' learning experience. If students are not experiencing continuity of care or in only a limited number of clinical situations then opportunities need to be created for the student. If the visualizations were only available at the end of the placement then it would be too late to make changes to benefit the current cohort of students resulting in a less than optimal learning experience.

V. PROCESS OF REFINEMENT

This project began with the objective to explore the continuity of care being experienced by students on their longitudinal placement. The main reporting tool used for the creation of visualizations was the Business Intelligence and Reporting Tools (BIRT) which is an open source reporting tool designed for web applications. One of the main advantages of using BIRT is that it allows data to be obtained from multiple online data sources to create reports and visualizations in real time.

A. Design goals

The first step in the design process was the identification of design goals for the reporting. In particular this involved the identification of questions to be answered by the available data. The questions that needed to be answered included:

- Which students are consulting patients repeatedly?
- How many consultations do students have with each patient?
- How long is the period of care (from first to last consultation) for each patient?
- How many patients does each student see?
- How do patients' continuity of care patterns compare across the cohort?
- How do the consultations with each patient vary as to clinical presentation and clinical setting?
- How do patterns of care vary across placements, towns, and regions?

B. Data

The relevant dataset consisted of Clinical Log entries, each including the following data:

- Consultation date
- Student number
- Patient number
- Age group
- Gender
- Clinical setting (e.g. Hospital, General Practice, etc)

- Primary clinical presentation

Additionally, a secondary dataset was available showing each student's clinical placement, its town and region. This hierarchy afforded analysis of patterns of care at each level. For example, patterns could be compared between regions as well as students.

C. Challenges

During the initial planning stages some challenges were identified which impacted the way in which answers to the questions above would be addressed with the available data. These challenges included:

- There is a large set of details to be rendered in an understandable way. The challenge is to present the data in such a way that the canvas can be perceived as a whole – so that the audience can see the data as a complete set. There are a considerable number of students, each of which can consult many patients over a 12 month period. High data density and clarity are both needed.

- There are a number of elements in the dataset that are secondary to the central analytical questions, yet nevertheless provide important context to the consultation. These include age group, gender, clinical setting, and clinical presentation. There is a tension between showing these and the risk of crowding out the critical variables.

- There are multiple levels of abstraction at which the dataset can be analysed: patient, student, clinical placement, placement town, or region. For example, it might be important to see that a particular placement or region is mostly exposing students to patients with fever who are typically seen twice within a month in the general practice. Remedial action could then be taken to expose students to patients with other conditions, who are cared for over longer periods in a variety of clinical settings. However, this raises further questions of how to fit yet more data in a concise and understandable way. Should data be somehow aggregated and summarized per town or region, or should the raw data be displayed with some grouping?

- There is some uncertainty around student engagement and reliability in logging their continuity of care. Due to legal requirements of health information systems, students aren't allowed to store any identifiable information of patients in the Clinical Log. Although a couple of interface designs were implemented to assist students in entering a series of consultations with a patient, a number of consultation series may have not been logged due to the difficulty in remembering that a particular log entry is associated to another one in the past. On the other hand, this does not invalidate visualizing the data. Although perfect data is rarely available to a designer, visualizing the existing data can highlight and locate its gaps, advancing its improvement.

D. Process

Designing a visualization of continuity of care involved an iterative process of design and experimentation. Three visual designs were developed with the aim of answering the above questions with the available data. Each is described below, with a summary of advantages and disadvantages. Lastly, an ideal - but yet to be implemented - design is presented.

1. Initial Design: Table

The first design was a table showing the number of consultations for each patient per month (Figure 1).

Region	Town	Placement	Student	Patient	2010-07	2010-08	2010-09	2010-10	2010-11	2010-12	2011-01	2011-02	2011-03	Total
Region 1	Town 1	Placement 1	Student 1	4044	1	2								3
				4045	1	1								2
				4047	1	1								2
				4131		2								2
				4224		2	1							3
				4891			1	1						2
				4892				2						2
				4893				2						2
				6141				1	1					2
			TOTAL		3	8	2	6	1					20
		Placement 1 TOTAL			3	8	2	6	1					20
	Town 1 TOTAL				3	8	2	6	1					20
Region 1 TOTAL					3	8	2	6	1					20
Region 2	Town 2	Placement 2	Student 1	3716	2									2
				3720	1		1							2
				3781	3									3
				4185		1	1							2
				4561			1					1		2
				4982						2				2
			TOTAL		6	1	3			2		1		13
		Placement 2 TOTAL			6	1	3			2		1		13
	Town 2 TOTAL				6	1	3			2		1		13
	Town 3	Placement 3	Student 1	3867	2									2
				3868	2									2
				3965	1	1								2
				4053		2								2
				4143		2								2
				5863								2		2
			TOTAL		5	5						2		12
		Placement 3 TOTAL			5	5						2		12

Figure 1. Number of consultations per patient and per month, with group totals

The data was prepared by first joining Clinical Log records with placement data for each student (so as to include the student, their clinical placement, its town, and region). We then created a data cube with patients as rows (with groupings on student, placement, town, and regions), and months as columns. Each cell thus shows the number of times a student had a consultation with a specific patient in each month. Additional rows provide totals for each grouping. A greyscale highlighting is used to add a visual component, with cells with a higher number of consultations having a darker background.

This format has some advantages. It provides a lot of detail, some of which is unavailable in the other formats, including the placement, town, and region where a patient was consulted may be critical in identifying issues and patterns wider than the individual student. This highlights the value of reporting tools which allow joining data from different sources (e.g. Clinical Log and a placements spreadsheet). The highlighting of cells gives a visual sense of length and frequency of patient interactions.

Its main disadvantage is the amount of space it requires. For example, much space is taken up by labels, while significant areas display no data. This makes it harder to see the whole picture easily, which is a key reason to visualize data. Further, the large amount of detail might obscure the overall pattern. Although the group totals could be removed to vertically compress the table, this would be a loss of important information. Lastly, there is some loss in aggregating data to a monthly frequency, as it is unclear how close consultations in a month are to each other. This may be a potentially important detail.

This format gives the most amount of detail while retaining a visual representation of the larger patterns. However, this comes at the expense of size which in most situations it's likely to be impractically large.

2.   Revised Design: Gantt Chart

The second design (Figure 2) involved using a Gantt chart to represent consultations.

Figure 2. Consultations represented in a Gantt chart

Each consultation chain is represented by a 'task bar', starting at the first consultation and ending at the last one. The total number of consultations in each chain is shown to the right of each bar. This design effectively 'hacks' an existing chart type, the Gantt chart, by extending its use beyond its original design.

This design has a number of advantages compared to the table above. Most importantly, it displays the number of patients and span of consultations visually, making patterns and exceptions easy to identify. It groups and colour-codes bars by student, which is very helpful to an academic keen to identify at-risk students. For example, it is clear that some students had repeated consultations with more patients than others. Additionally, some students seemed to have shorter spans of care with their patients, while other students generally saw their patients across longer periods.

Its major disadvantage is that it only shows the first and last consultation for each patient, ignoring the timing of intermediate visits. For example, it may be important to see that a series consists of four consultations within a fortnight, with an additional one eight months afterwards. Additionally, there is a degree of overplotting with some bars overlapping, although their width and border help in minimising the impact. The chart could be made less crowded by showing series for a student at a time, but this would make it difficult to compare patterns between students, which is one of the central analytical questions.

Overall, this design visualizes continuity of care well, allowing comparison between both patients and students. The lack of a representation of individual consultations is its main disadvantage.

3. Revised Design: Line Chart

The third design uses a standard line chart, with each consultation chain as a series, and each consultation as a data point in that series (Figure 3).
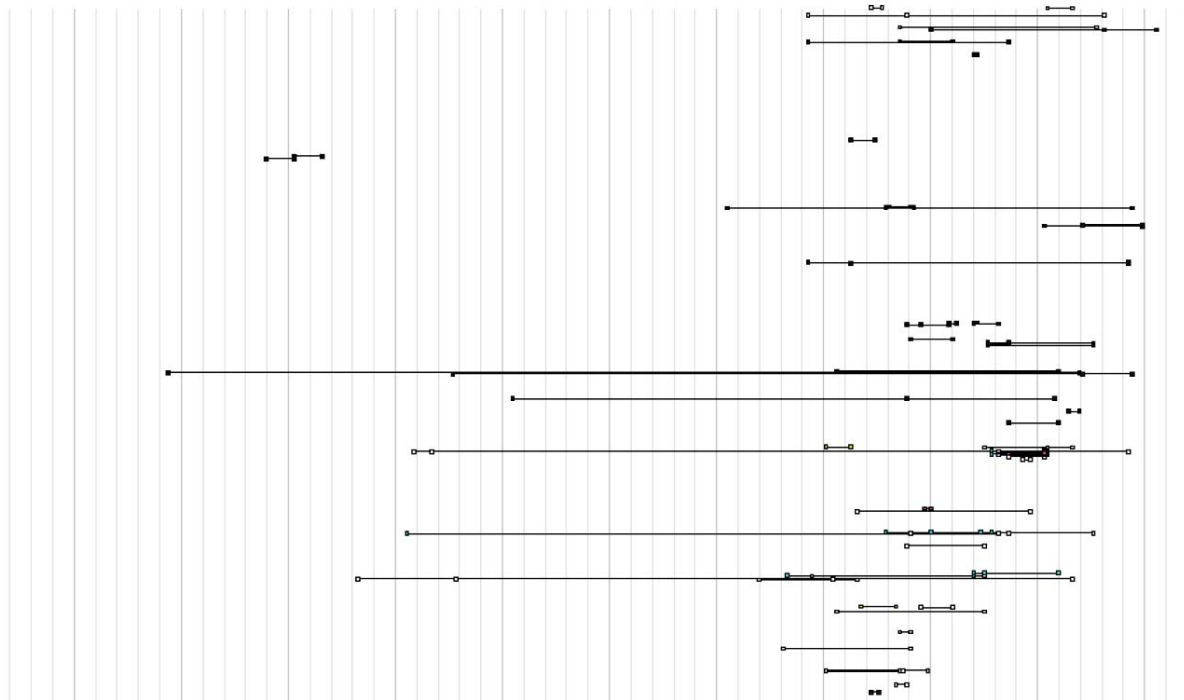

Figure 3. Line chart

This was an attempt to improve on the Gantt design, by plotting each consultation. Each horizontal line represents a series of consultations by a student with a patient. Each series is made up of date/patient number pairs, with each pair representing a consultation. Each consultation is thus plotted by a square marker, where the patient number is plotted against the y-axis. Ideally, the lines would be grouped and colour-coded by student, and sorted chronologically. However, the reporting tool used (BIRT) would not allow a categorical y-axis for a line chart. Therefore the patient number had to be mapped against a linear y-axis. This is another example of chart-hacking. A line chart would typically show change on the y-axis across time, with the y-values being the analytical variable. In this case, however, we're using the y-axis value to categorize the series. The y-value is not the critical variable, but rather the x-value: *when* each consultation occurred.

This design's main advantage over the Gantt format is its display of each consultation. This is helpful in showing the spread of consultations with each patient across time. Further, it is more vertically compact than both the table and the Gantt chart, which helps in representing the whole dataset in a conveniently sized canvas.

Unfortunately, mapping the patient number against a linear y-axis caused awkward placement of the lines. The lines are arranged by entry sequence across all students (since the patient number is assigned sequentially in the system), and therefore not grouped by student. This results in gaps at times, and in overlayed lines at other times. Perhaps more seriously, the design obscures how students might differ in their continuity of care. Although using lines rather than bars made for a more compact chart, it also involved the loss of colour as a way of categorising series by student. The markers are coloured, but their small size makes this less obvious. Thus, overall, neither placement nor colour could be used to categorize series by student, an important dimension to illustrate.

4. Future Design

Figure 4 represents a mocked up ideal design arrived at after reflection on the previous design iterations.
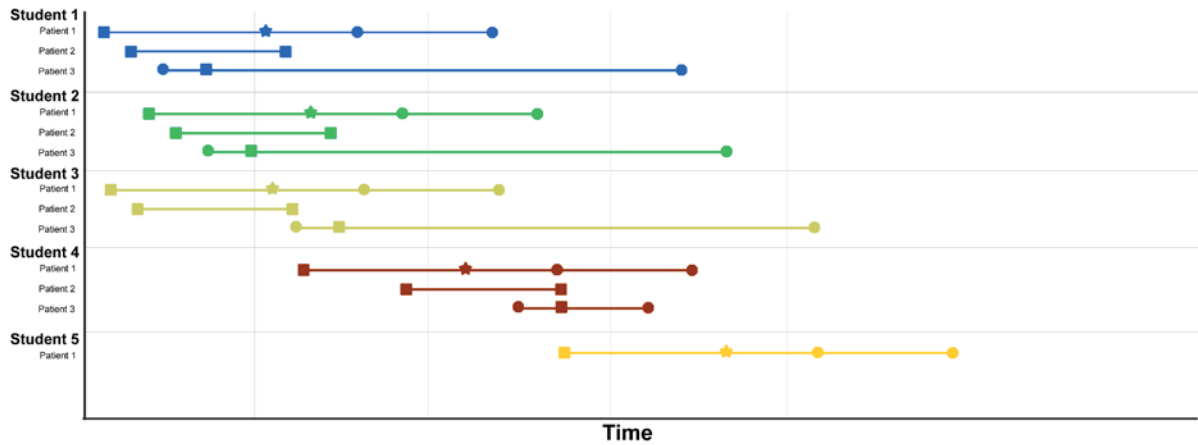

Figure 4. Ideal design mock up

Each series of consultations by a student with a patient is drawn as a horizontal line. Individual consultations are plotted as a marker in each series. Additionally, the clinical setting of each consultation is represented by the marker's shape. This is a dimension in the data not visualized in any of the previous designs, yet it's an important one as it shows the student's care of a patient across the health system. As with the Gantt chart, these bars are grouped and colour coded by student. Additionally, they are sorted within each group by the date of the initial consultation. However, the series are drawn as a line rather than a bar, so as to use vertical space more efficiently and avoid overplotting. The student and patient number are noted on the y-axis, which in this case is a categorical axis. Time, one of the critical analytical variables, is mapped on the x-axis. This is a natural and common way to represent the passing of time.

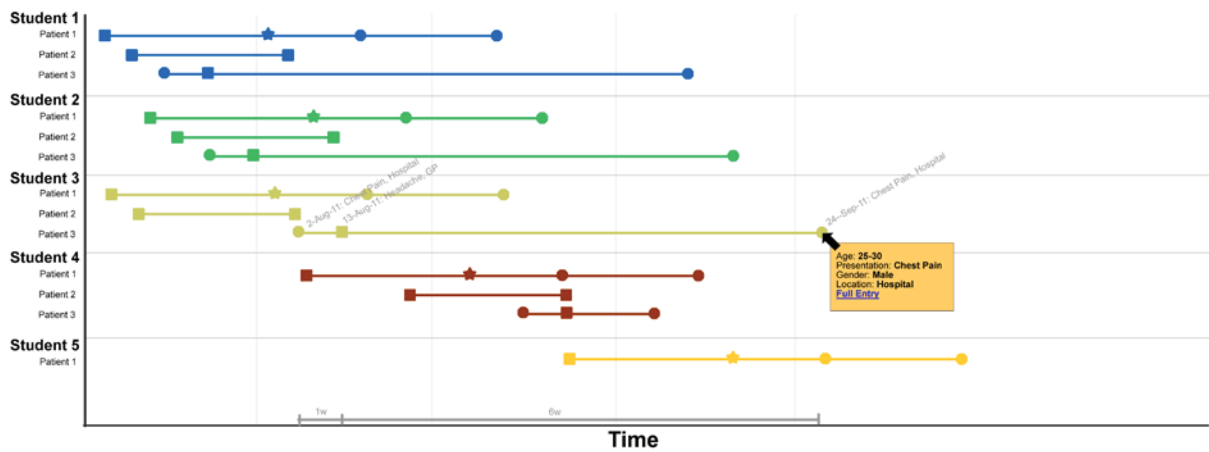Further, computer-rendered visualizations afford interactive functions (Figure 5).

Figure 5. Interactive features

Placing the mouse pointer on a marker could show extra details on the consultation, such as the patient's age, clinical presentation, gender, and location, as well as a link to the full entry in the Clinical Log. Further, the date, clinical presentation, and clinical setting would appear next to each consultation's marker in that series. On the x-axis, the time elapsed between consultations would be shown. Filtering would enable focus on a subset of series, such as those with a specific range of consultations as well as those containing a consultation with a particular clinical setting or presentation.

Such an interactive visualization would come closer to Schneiderman's visualization mantra: "overview first, zoom and filter, then details on demand" [15]. It would allow an academic to see the whole dataset represented and then focus on a subset, with additional details displayed as needed.

Unfortunately, no simple way has yet been found to implement this design, although an initial investigation into the Processing visualization language has been done. Clearly, this involves much more effort, time, and risk.

VI. LESSONS LEARNED

Various insights have been gained through this process. Although they will not all have universal relevance, considering the following may prove helpful when developing visualizations of learning analytics data.

A clear understanding of the *questions* to be answered and *data* available is critical. Indeed, it is important to start with the questions, so as not to be limited by the data available. A myopic focus on the data can lead to answering questions for which we have easy answers but which do not matter. Focusing on the most needed insights can motivate creative data collection and representation, even if it involves more work and time.

The process of representing the available data in a way that provides insight into the analytical questions identified is one of *visual analysis and design* that requires specific skills. As Few notes regarding analytical skills, "The ability to do this is not intuitive; it must be learned, and the good news is that we can learn these skills with relative ease… Unfortunately, few people have learned these simple skills, and most who have done so followed the hard road, as I did, making individual small discoveries here and there over many years" [16].

It is helpful to develop a clear design of a visualization, even if it's simply mocked up with a graphics package or a sketch. This can provide early feedback from the target audience, before significant development effort and time is wasted. It also provides a clear ideal to aim for, even if it cannot be readily implemented and compromise solutions have to be used in the meantime.

Designing a visualization involves navigating through trade-offs. It's critical to identify these and make choices based on clear goals. One way to deal with the inevitable trade-offs between different visual designs is to supplement. Rather than choosing a single design, use two together taking advantage of the strengths of each. For example, the Gantt design may be used to give a concise overview while the table provides a detailed view of students' clinical experiences.

One of the critical trade-offs that sometimes needs to be resolved is whether to use an existing and standard chart type, which isn't ideal but can be readily used, or to develop a new visualization which would be better but costlier. This is essentially an economic decision, pitting the marginal cost of time needed to develop a chart type from scratch versus the marginal benefit of the ideal visualization over the stock-standard, suboptimal but available option. A practical way forward is to start with the standard chart type while developing a new visualization as needed. However, the situations where this is critical are probably rare: standard chart types are usable for the vast majority of scenarios. Often, a standard chart type can be 'stretched' or 'hacked' by using it in a way beyond its original design purpose. This can often provide a good visualization of less-common scenarios without requiring programming.

VII. CONCLUSION

Work continues on the development of visualizations of continuity of care as we work towards something that resembles the ideal solution presented in this paper. The process of developing the visualization has highlighted a number of important lessons which will assist in future learning analytics projects. A closer look at the data being recorded in the system through the reports and visualizations has also informed further system design of the Clinical Log as we endeavour to create a more user-friendly method of tracking patient numbers. The benefit that learning analytics has afforded the GSM in terms of curriculum monitoring and development has been positively acknowledged by the faculty and we are working with a

number of academics to identify other projects and opportunities where learning analytics have the potential to improve the student learning experience. Foremost we are working to increase the knowledge and analytical skill set of educational technology staff within the faculty so they are able to contribute to the design and implementation of learning analytics initiatives into the future.

VIII.   ABOUT THE AUTHORS

Martin Olmos is an educational technology consultant, assisting the Graduate School of Medicine with analytics. He led the implementation of the school's educational technology leading up to its opening in 2007.

Linda Corrin is Lecturer: Educational Development in the Graduate School of Medicine at the University of Wollongong. Linda's research focuses on the visualization of learning analytics and the investigation of students' use of technology in everyday life and academic study contexts with a view to informing teaching strategies and policy in higher education.

IX.   REFERENCES

1. **Campbell, J.P., DeBlois, P.B., and Oblinger, D.G.** Academic Analytics: A New Tool for a New Era. *EDUCAUSE Review* 42(4): 40–57 (July/August 2007). http://www.educause.edu/library/erm0742.
2. **Steele, J., and Iliinsky, N.** (Eds.), *Beautiful Visualization: Looking at Data through the Eyes of Experts*, 1st ed., Stebastopal, CA :O'Reilly Media, 2010.
3. **Hudson, J.N., Weston, K.M., and Farmer, E.** Engaging rural preceptors in new longitudinal community clerkships during workforce shortage: a qualitative study. *BMC Family Practice* 12:103 (2011). http://www.biomedcentral.com/1471-2296/12/103.
4. **Worley, P., Silagy, C., Prideaux, D., Newble, D., and Jones, A.** The Parallel Rural Community Curriculum: an integrated clinical curriculum based in rural general practice. *Medical Education* 34(7): 558-565 (2000).
5. **Ogur, B., and Hirsh, D.** Learning Through Longitudinal Patient Care–Narratives From the Harvard Medical School–Cambridge Integrated Clerkship. *Academic Medicine* 84(7): 844-850 (July 2009).
6. **Konkin, J., and Suddards, C.** Creating stories to live by: caring and professional identity formation in a longitudinal integrated clerkship. *Advances in Health Sciences Education* (November 2011). DOI 10.1007/s10459-011-9335-y.
7. **Olmos, M., and Corrin, L.** Academic analytics in a medical curriculum: enabling educational excellence. *Australasian Journal of Educational Technology* 28 (1): 1-15 (2012).
8. **Corrin, L., and Olmos, M.,** Capturing clinical experiences: supporting medical education through the implementation of an online Clinical Log. In: Steel, C., Keppell, M. J., Gerbic, P., and Housego, S. (Eds.), *Curriculum, technology & transformation for an unknown future: Proceedings of ascilite Sydney 2010,* University of Queensland, 231-235, 2010.
9. **Johnson, L., Adams, S., and Cummins, M.,** *The NMC Horizon Report: 2012 Higher Education Edition*, Austin, Texas: The New Media Consortium, 2012.
10. **Long, P. and Siemens, G.** "Penetrating the Fog: Analytics in Learning and Education." *Educause Review* 31-40 (Sept/Oct 2011).
11. **Siemens, G.** 1st International Conference on Learning Analytics & Knowledge–About, 2011 https://tekri.athabascau.ca/analytics/.
12. **Goldstein, P.J.** Academic analytics: The uses of management information and technology in higher education. ECAR Key Findings, EDUCAUSE (December 2005). http://net.educause.edu/ir/library/pdf/EKF/EKF0508.pdf.
13. **Norris, D., Baer, L., Leonard, J., Pugliese, L., and Lefrere, P.** Action Analytics: Measuring and Improving Performance That Matters in Higher Education. *Educause Review* 43-67 (January/February 2008).

14. **Elias, T.** Learning Analytics: Definitions, Processes, and Potential (January 2011). http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf.
15. **Shneiderman**, **B.,** The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In: *Proceedings of the IEEE Symposium on Visual Languages*, 336-343, Washington: IEEE Computer Society Press, 1996.
16. **Few, S.,** Now you see it: simple visualization techniques for quantitative analysis. In: Mackinlay, S.K., and Shneiderman, B (Eds.), *Readings in Information Visualization: Using Vision to Think,* Massachusetts: Morgan Kaufmann, 1999.